UNIVERSIDADE FEDERAL FLUMINENSE

AUGUSTO PARISOT DE GUSMÃO NETO

Análise Híbrida de *Ransomware* para Sistema Operacional Windows

NITERÓI

2023

AUGUSTO PARISOT DE GUSMÃO NETO

Análise Híbrida de *Ransomware* para Sistema Operacional Windows

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal Fluminense como requisito parcial para a obtenção do Grau de Mestre em Ciência da Computação. Área de concentração: Sistemas de Computação.

Orientador:
Raphael Carlos Santos Machado

Coorientador:
Lucila Maria de Souza Bento

NITERÓI

2023

Ficha Catalográfica

Biblioteca da Escola de Engenharia e do Instituto de Computação

Para gerar a ficha catalográfica o usuário deverá preencher o seguinte formulário, somente a primeira letra maiúscula, a não ser nos casos de nomes e siglas. Após clicar em "Gerar ficha" será aberta uma nova página do navegador com a ficha catalográfica gerada. O usuário poderá realizar o download do arquivo, no formato PDF, e inserir no trabalho utilizando programas próprios para essa finalidade ou poderá tirar um "*print*" da imagem da ficha e em seguida colar no arquivo do trabalho.

Lembramos que toda informação inserida no formulário é de responsabilidade do usuário, portanto atenção no preenchimento dos campos, qualquer dúvida entre em contato com a biblioteca que atende o seu curso.

Augusto Parisot de Gusmão Neto

ANÁLISE HÍBRIDA DE RANSOMWARE PARA SISTEMA OPERACIONAL WINDOWS

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal Fluminense como requisito parcial para a obtenção do Grau de Mestre em Ciência Computação. Área de concentração: Sistemas de Computação

Aprovada em 05 de junho de 2023.

BANCA EXAMINADORA

Prof. Raphael Carlos Santos Machado - Orientador, IC - UFF

Prof. Lucila Maria de Souza Bento - Coorientadora, IME - UERJ

Prof. Nival Nunes de Almeida , Escola de Guerra Naval

Prof. Luiz Satoru Ochi, IC - UFF

Niterói

2023

Dedico este trabalho às vozes da minha cabeça, que são minhas amigas mais surpreendentes.

E à minha esposa e minha filha, que moram conosco.

Agradecimentos

Agradeço à minha esposa Renata Miguez, que me acompanhou de perto nessa jornada, que me apoiou para que este trabalho pudesse ser concebido e finalizado e por me aturar falando coisas que ninguém entende e, mesmo assim, mostrar interesse. À minha filha, que me encanta com seus sorrisos toda vez que a vejo. Aos meus pais Wanda e Tomás, que me apoiaram e incentivaram desde sempre. Aos meus orientadores, Raphael Machado e Lucila Bento que me mostraram os caminhos a serem seguidos e pela confiança depositada e ao meu grande amigo, Marcelo Rocha, pela parceria nas disciplinas e trabalhos.

Resumo

O crescimento do acesso a dispositivos computacionais aumentou sobremaneira desde o início dos anos 2000. A miniaturização de componentes eletrônicos, os avanços na tecnologia de baterias e telas barateou esses dispositivos, permitindo que uma mesma pessoa possua vários desses em uso (tablets, telefones, computadores e dispositivos domésticos inteligentes). Esse grande crescimento não é necessariamente acompanhado de aumento de mentalidade de segurança e ainda, a massa de dados gerada pela interação com esses dispositivos gera interesse de grupos com intenções maliciosas de lucro e todo tipo de software malicioso é criado diariamente para subverter e acessar esses dispositivos. Dentre esses muitos *softwares* maliciosos, temos os *ransomwares*: armas capazes de cifrar todos os arquivos da vítima para que esta se veja obrigada a pagar um resgate sob o risco de não conseguir recuperar seus dados. Neste trabalho, realizamos um conjunto de experimentos para avaliar dinamicamente técnicas de Aprendizado de Máquina para detecção de *malware* e sua classificação em suas respectivas famílias. Para executar os experimentos, coletamos um total de 989 amostras de *ransomwares* das oito famílias mais proeminentes em 2021 e 2022, baixadas de repositórios públicos : *Conti*, *Ryuk*, *Revil*, *Egregor*, *LockBit*, *Clop*, *Netwalker* e *MountLocker* além de 90 amostras de *software* benignos. Primeiro, montamos um ambiente controlado/isolado para registrar o comportamento do ransomware para avaliação de técnicas de Aprendizado de Máquina em termos de métricas de desempenho comumente usadas na literatura (*Accuracy*, *Precision*, *Recall* e Fi-*Mesure*). Para executar as análises utilizamos o *Cuckoo Sandbox*. Foram criadas ferramentas na linguagem Python para automatização de tarefas como busca das amostras nos repositórios públicos e mineração de dados para composição dos conjuntos de dados de detecção. A partir dos relatórios de execução salvos na forma de relatórios JSON, utilizamos técnicas de mineração de texto e de chamadas de API aplicadas em ferramentas que construímos especialmente para extrairmos um conjunto promissor de dados que representam o comportamento de uma amostra de *ransomware* e submetemos os conjuntos de dados à classificação utilizando seis algoritmos de Aprendizado de Máquina: *Decision Tree*, *Random Forest*, *K-Nearest Neighbors*, *Naive Bayes*, *Support Vector Machines* e *Multilayer Perceptron*. A principal motivação para elaboração dos experi-

mentos é que diferentes técnicas foram projetadas para otimizar diferentes critérios, que se comportam de maneira diferente, mesmo em condições semelhantes. Os resultados experimentais mostram que o métodos propostos podem alcançar um bom desempenho de classificação ao usar o algoritmos *Random Forest* e *Decision Tree*. Os melhores resultados de classificação foram alcançados com esses classificadores em três situações: a primeira e a segunda, utilizando-se o conjunto de dados minerados ao utilizar a técnica de mineração de texto TF-IDF nas seções *Signatures* e *Memory* dos relatórios de análise e a terceira, no conjunto de dados minerado a partir da contagem de chamadas de API. Além da classificação, revelamos as diretrizes utilizadas para proteção do ambiente de análise das ferramentas anti-VM, tanto para a configuração do Sistema Operacional quanto para a conectividade de rede utilizada.

Palavras-chave: *Ransomware*, *Cuckoo Sandbox*, Análise Dinâmica, Análise Estática, Aprendizado de Máquina, Detecção de *Malware*.

Abstract

The growth of access to computing devices has greatly increased since the early 2000s. The miniaturization of electronic components, advances in battery technology and screens have made these devices more affordable, allowing individuals to own multiple devices (such as tablets, phones, computers, and smart home devices). However, this rapid growth does not necessarily come with an increased security mindset. The massive amount of data generated by interacting with these devices has attracted the interest of groups with malicious intent, and all sorts of malicious software are created daily to exploit and gain access to these devices. Among these malicious software, ransomware stands out as a weapon capable of encrypting all of a victim's files, forcing them to pay a ransom in order to regain access to their data. In this work, we conducted a series of experiments to dynamically evaluate Machine Learning techniques for malware detection and classification into their respective families. To perform the experiments, we collected a total of 989 samples of ransomware from the eight most prominent families in 2021 and 2022, downloaded from public repositories: Conti, Ryuk, Revil, Egregor, LockBit, Clop, Netwalker, and MountLocker, in addition to 90 samples of benign software. First, we set up a controlled/isolated environment to record the behavior of the ransomware for evaluating Machine Learning techniques in terms of commonly used performance metrics such as Accuracy, Precision, Recall, and F1-Measure. We used the Cuckoo Sandbox to execute the analyses. We developed Python tools to automate tasks such as searching for samples in public repositories and data mining to compose the detection datasets. From the execution reports saved in the form of JSON reports, we employed text mining and API call techniques applied in tools we specifically built to extract a promising set of data representing the behavior of a ransomware sample. We then subjected the datasets to classification using six Machine Learning algorithms: Decision Tree, Random Forest, K-Nearest Neighbors, Naive Bayes, Support Vector Machines, and Multilayer Perceptron. The main motivation for conducting the experiments is that different techniques were designed to optimize different criteria, which behave differently even under similar conditions. The experimental results show that the proposed methods can achieve good classification performance when using the Random Forest and Decision Tree algorithms.

The best classification results were achieved with these classifiers in three situations: the first and second using the mined dataset by applying the TF-IDF text mining technique to the Signatures and Memory sections of the analysis reports, and the third using the mined dataset based on the API call count. In addition to classification, we revealed the guidelines used to protect the analysis environment from anti-VM tools, both for configuring the operating system and network connectivity used.

Keywords: Ransomware, Cuckoo Sandbox, Dynamic Analysis, Static analysis, Machine Learning, Malware Detection.

Lista de Figuras

Lista de Tabelas

Sumário

1 Introdução

Devido à natureza ubíqua que a computação tomou nos últimos anos, tem havido um crescimento acelerado na utilização de dispositivos computacionais conectados à internet (STATISTA, 2022). A projeção é que a quantidade de dispositivos conectados atinja 90,9 bilhões (IoT) e 10,3 bilhões (não IoT) em 2025. Esta grande quantidade de dispositivos gera uma imensa massa de dados, armazenada localmente ou em nuvem. Essa massa de dados tem um grande valor financeiro e, por isso, provoca a cobiça de agentes maliciosos interessados em acessá-los. Suas motivações, meios de acesso e utilização desses dados são os mais variados possíveis, dentre os quais podem ser citados ataques a instituições públicas e privadas, infraestruturas críticas, extorsão, chantagem, sequestro e/ou venda de dados, ativismo político e mineração de criptomoedas.

Para a consecução dos seus objetivos, os agentes maliciosos podem se utilizar de programas de computadores com algumas características especiais para conseguir ter acesso aos dados do alvo. Estes programas são chamados de *malwares*. Podemos citar algumas dessas características, como a capacidade autônoma de procurar outros dispositivos computacionais vulneráveis, auto replicação em mídias removíveis, evitar antivírus, camuflar sua execução para que o usuário não desconfie, enviar a si mesmos por e-mail, alterar o funcionamento do seu próprio código, utilizar vulnerabilidades *zero-day* [1] (STALLINGS; BRESSAN; BARBOSA, 2008), escravizar a máquina da vítima (bot) e/ou ter acesso total ao sistema (arquivos, câmera, o que é digitado no teclado e o que é mostrado na tela). Muitas vezes, a vítima nem percebe a ação maliciosa que ocorre em sua máquina, até que seja tarde. Os *ransomwares* foram criados a partir da necessidade dos agentes maliciosos de forçar a vítima a pagar uma quantia em dinheiro e ao mesmo tempo evitar exposição, pois anteriormente, quando ocorriam somente os roubos de dados, os operadores tinham que procurar ativamente possíveis compradores. Uma das abordagens desse modelo de negócio, adotada para evitar exposição, é o *double extortion* (KASPERSKY, 2021), que vai

[1] Vulnerabilidade de software descoberta por invasores antes que o fornecedor tome conhecimento. Como não são conhecidas, não existe correção, o que aumenta a probabilidade de a exploração ser bem-sucedida - https://www.kaspersky.com.br/resource-center/definitions/zero-day-exploit

além de extorquir as vítimas para descriptografar os dados: os agentes maliciosos copiam os dados da vítima e exigem outra quantia em dinheiro para que seus dados não sejam divulgados. Um dos grandes avanços na segurança de sistemas digitais, a criptografia teve um papel extremamente importante no desenvolvimento de *ransomwares*, tornando-se um instrumento forte nas mãos dos criminosos. Se implementado corretamente, seu impacto é irreversível: sem conhecer a chave de descriptografia, recuperar o conteúdo de um arquivo criptografado é computacionalmente inviável, fato bastante prejudicial para as vítimas. No entanto, implementar criptografia sem falhas é uma tarefa difícil, e os codificadores de *ransomwares* são desafiados pelos mesmos problemas que têm incomodado os engenheiros de segurança responsáveis pela implementação de aplicativos criptográficos dos quais um dos mais relevantes é gerar chaves de criptografia criptograficamente seguras e mantê-las dessa maneira durante todo o processo de cifração. Uma falha nesse processo torna a criptografia fraca, tornandor possível a recuperação das chaves e dos dados da vítima, o que comprometeria o modelo de negócios do *ransomware*.

1.1 Motivação

Os *ransomwares* têm causado volumosos prejuízos em todo o mundo (IBMSECURITY, 2022), não somente pelos valores exigidos pelo resgate dos dados e sua não divulgação, mas também por deixar as operações das vítimas paralisadas. Nos últimos anos, a quantidade de empresas e órgãos públicos que tiveram seus dados sequestrados cresceu exponencialmente. No ano de 2021, em média, ocorreu um ataque de *ransomware* a cada 11 segundos (KASPERSRKY, 2021b) e este fato, combinado com o crescimento de dispositivos conectados à internet (KASPERSRKY, 2021a), torna o *ransomware* uma forte ameaça e tópico relevante para a comunidade científica.

1.2 Justificativa

Para lidarmos com a ameaça imposta pelos *ransomwares*, precisamos saber como ocorre a infecção, como funcionam seus mecanismos de ocultação, o que o faz ativar-se e o que seus criadores ou utilizadores têm como motivação e objetivo, além de termos sistemas capazes de detectar e impedir sua ação. Nessa linha de raciocínio, é importante que tenhamos um delineamento da ameaça advinda de um *ransomware*, quais sejam: seus meios de propagação, infecção, reconhecimento, comunicação (Comando e Controle), pesquisa de arquivos, criptografia de dados e pedido de resgate, através de uma revisão da

literatura científica sobre o que há de mais atual sobre este assunto. Com o resultado do levantamento em mãos, estaremos mais preparados a evitar, detectar, reconhecer e recuperar sistemas afetados por um ataque de *ransomware*. A ideia é usar esse conjunto de conhecimentos em uma proposta de metodologia semi-automatizada, que permita a distinção entre *ransomwares* e aplicações não maliciosas, a partir da análise dinâmica de amostras coletadas.

1.3 Objetivos

O objetivo principal deste trabalho é realizar a detecção de ataques de *ransomwares* no Sistema Operacional Windows utilizando ferramentas de *sandboxing* para analisar as amostras selecionadas e Aprendizado de Máquina para efetuar a classificação.

Os Objetivos secundários deste trabalho são:

- Apresentar uma revisão dos trabalhos existentes na literatura que tratam de análise de *malware*, particularmente Análise Dinâmica com uso da ferramenta Cuckoo Sandbox.
- Revisar os conceitos que envolvem análise de *malware*, técnicas de detecção e características mais comuns encontradas nas amostras analisadas.
- Descrever as fases do ciclo de vida de um *ransomware* e suas características.
- Apresentar das famílias de *ransomwares* mais proeminentes nos anos de 2021 e 2022.
- Configurar um ambiente virtualizado capaz de evitar técnicas de evasão dos *ransomwares* analisados.
- Desenvolver uma técnica eficiente para extração das características selecionadas de arquivos de relatórios gerados pelo ambiente de análise.
- Analisar e selecionar as características relevantes para o processo de criação do conjuntos de dados das características dos *ransomwares* analisados.
- Construir uma metodologia que nos permita criar um conjunto de dados de características de *ransomwares* a partir da análise de amostras maliciosas em ambiente virtualizado.
- Submeter os conjuntos de dados a algoritmos classificadores de Aprendizado de Máquina.

- Comparar os resultados obtidos nas diversas técnicas aplicadas.

1.4 Metodologia

Este trabalho, a fim de colaborar com a defesa cibernética, iniciou com o levantamento dos problemas desse domínio para que pudéssemos definir o objetivo principal e os secundários da pesquisa. Uma vez delineado o escopo, foram buscados trabalhos relacionados para que as experiências pudessem servir de exemplo e guia para esta pesquisa. Com base nos exemplos e abordagens apresentadas nos trabalhos relacionados, foi realizado um levantamento do referencial teórico para identificar quais teorias e ferramentas seriam mais adequadas para criar a base cognitiva necessária para a compreensão dos conceitos e ferramentas utilizadas. Um levantamento das famílias de *ransomwares* foi realizado, a fim de identificar as famílias mais ativas em 2021 e 2022 para que essas fossem analisadas sob a ótica da abordagem proposta. Para obter as amostras a serem analisadas, foram realizadas buscas nos repositórios de *malware* mais comumente utilizados pelo mercado e pela comunidade científica. O *download* das amostras foi executado utilizando *scripts Python* confeccionados para esta tarefa. Além disso, os estudos realizados sobre o funcionamento e o comportamento dos *malwares* serviram como norte para a decisão de quais características seriam selecionadas para compor os diferentes conjuntos de dados construídos. Essa transformação também foi realizada por *scripts Python*. Posteriormente, os conjuntos de dados foram submetidos aos classificadores mais utilizados na literatura para este tipo de tarefa, informação também advinda do levantamento bibliográfico realizado. Ao final, apresentamos os resultados e as conclusões.

1.5 Contribuições

Nesta dissertação, montamos um ambiente de análise dinâmica utilizando o *Cuckoo Sandbox*, capaz de executar amostras maliciosas e emitir um relatório com diversas informações sobre o comportamento de tais amostras. Com base na análise das seções e características contidas nesses relatórios, confeccionamos os conjuntos de dados a serem submetidos à classificação. Descrevemos as famílias de *ransomwares* mais ativas em 2021 e 2022, as peculiaridades apresentadas em cada fase de seu ciclo de vida e como isso se relaciona com o trabalho desenvolvido. Além disso, estudamos métodos de ofuscação de Máquinas Virtuais para possibilitar a análise de *ransomwares* com artefatos anti Máquinas Virtuais (Anti-VM) e implementamos no ambiente experimental proposto, de modo que este

conseguisse simular um ambiente equivalente ao ambiente real. *Scripts Python* foram criados para interação com os repositórios de amostras e extração de características dos relatórios com sua transformação em conjunto de dados, focando em duas abordagens: mineração de texto e chamadas de API. Com as características convertidas em conjuntos de dados, utilizamos os algoritmos de Aprendizado de Máquina mais usados na literatura para classificarem as amostras, a fim de que pudéssemos verificar o desempenho de cada classificador sobre os mesmos dados.

Foram identificadas 8 famílias mais ativas nos anos de 2021 e 2022, das quais obtivemos 989 amostras em repositórios públicos. Os conjuntos de dados gerados a partir das análises de todas essas amostras foram submetidas a 8 classificadores e ferramentas de pré-processamento e redução de dimensionalidade. Além disso, utilizamos duas distribuições diferentes de treino e teste (*test size*) para a realização dos experimentos totalizando 1980 classificações, que foram avaliadas de acordo com as métricas *Accuracy*, *F1-Measure* e *Recall*.

1.6 Organização deste Trabalho

Este trabalho está dividido da seguinte forma:

- Capítulo 2: Introduzimos os conceitos básicos de *malware* e como são classificados pela literatura. Apresentamos também conceitos sobre técnicas de detecção de *malware* e Aprendizado de Máquina (classificação, padronização e redução de dimensionalidade).
- Capítulo 3: Complementa os conceitos e fundamentos apresentados no Capítulo 2, apresentando o que são os *ransomwares*, seu ciclo de vida, modelos de negócio adotados pelos agentes maliciosos e as famílias mais ativas em 2021 e 2022.
- Capítulo 4: Apresentamos os trabalhos relacionados encontrados na literatura, apresentados em grupos gerais de assuntos, como técnicas de detecção, abordagem de extração de características, algoritmos de classificação e métodos anti-evasão.
- Capítulo 5: Apresentamos a abordagem proposta para a execução deste trabalho, os passos seguidos para alcançarmos os objetivos propostos, as ferramentas utilizadas, as ferramentas construídas especificamente para esta pesquisa e os procedimentos executados no ambiente virtualizado de análise.

- Capítulo 6: Executamos os passos propostos no Capítulo 5. Apresentamos o ambiente de experimentação montado para realizarmos os experimentos, os detalhamentos das execuções de cada um deles e os resultados, que posteriormente foram analisados.

- Capítulo 7: Por fim, encerramos o trabalho apresentando as conclusões e algumas considerações finais.

2 Referencial Teórico

Neste Capítulo, apresentamos o arcabouço necessário para a compreensão dos conceitos contidos neste trabalho para que consigamos classificar corretamente os artefatos maliciosos e ferramentas utilizadas por pessoas mal-intencionadas ao infiltrarem-se nas máquinas de suas vítimas. Para este fim, será feita inicialmente uma definição de *malware* e onde, dentro das classificações mais usuais da literatura, situam-se os *ransomwares*. Além disso, uma análise do seu comportamento, meios de infecção e cifração dos dados das vítimas, persistência, comunicação com seus operadores, ofuscação e propagação, assim como montagem e operação de ambientes *sandbox*, geração de relatórios, processamento desses dados para transformação em um *dataset* que seja representativo desse domínio e que permita que algoritmos de Aprendizado de Máquina (AM) classificarem de forma satisfatória as amostras entre *ransomwares* (maliciosas) e benignas, com aplicação de ferramentas de redução de dimensionalidade e otimização de desempenho.

2.1 *Malware*

Existem na literatura muitas definições de *malware* (acrônimo de *malicious software* - *software* malicioso) (OR-MEIR et al., 2019). *Malware* é um *software* que se insere clandestinamente em outro programa com a intenção de destruir dados, executar outros programas ou de alguma forma comprometer a confidencialidade, integridade e disponibilidade dos dados da vítima, de suas aplicações ou de seu sistema operacional. É considerado a ameaça externa mais comum para a maioria dos computadores, causando danos extensivos que necessitam de grande esforço para serem recuperados (SOUPPAYA; SCARFONE, 2013).

A definição pode ser dividida em duas partes. A primeira parte considera o ponto de vista do usuário/administrador, dado que nenhum processo deve estar ativo em um sistema sem que estes tenham ciência disso. A segunda parte se refere a confidencialidade (*confidenciality*), integridade (*integrity*) e disponibilidade (*availability*), que são os três

pilares da segurança (também conhecido como *CIA Triad*)[1]. Quando um usuário acessa um sistema ou serviço, parte do princípio que este sistema está em consonância com os três aspectos mencionados, de modo que os dados estão corretos e completos, não serão acessados por um outro usuário que não tenha permissão e que estará acessível durante o tempo necessário para ser utilizado.

Na literatura, podemos encontrar diversos exemplos de *malwares* que causaram danos consideráveis nos seus alvos. Um desses exemplos é o *worm Stuxnet* (2010), criado para atacar usinas de enriquecimento de urânio que estivessem utilizando sistema SCADA desenvolvido pela Siemens. A usina de *Natanz* (no Irã), foi uma das mais afetadas, a despeito de seus computadores estarem fora da internet (*air gapped*). O *Stuxnet* era dotado com capacidade de explorar vulnerabilidades *zero-day* e se propagar por mídias removíveis. Uma vez alcançado a infra-estrutura de rede da usina, alterava a apresentação do centro de monitoramento das centrífugas de enriquecimento de urânio para parecer que tudo estava funcionando normalmente, quando na verdade, fazia com que as centrífugas girassem mais rápido, danificando o sistema e atrasando o programa nuclear daquele país (CLARKE; KNAKE, 2015).

Um outro exemplo é o *malware* conhecido como *I Love You* (UNIVERSITY, 2000), que era disseminado via um anexo de e-mail com o pretexto de ser uma carta de amor, causando prejuízo e interrompendo serviços, principalmente de jornais e revistas, pois atacava sobrescrevendo arquivos de imagem, vídeo e áudio. O vírus também era capaz de roubar senhas da máquina infectada e enviá-las a um servidor remoto.

Também usando e-mail para se propagar pela internet, o *Mydoom* (MICROSOFT, 2004) utilizava como subterfúgio mensagens com as palavras *Error*, *Mail delivery System*, *Test* ou *Mail Transaction Field* em várias línguas (incluindo inglês e francês) para enganar usuários incautos. Ao infectar uma máquina, se auto copiava para a "Pasta Compartilhada" do KaZaA[2], que também era usado como vetor para propagação. Uma vez executado, encaminhava cópias de si para todos os e-mails que encontrava nas agendas disponíveis na máquina da vítima. Curiosamente, evitava distribuição de suas cópias para domínios de algumas universidades como MIT, *Rutger*, *Berkeley* e *Stanford*, além de empresas como *Microsoft* e *Symantec*.

O *malware* DuQu (MCAFEE, 2012) é uma variedade de *softwares* agregados que conjuntamente compõem uma plataforma de provimento de serviços para agentes maliciosos.

[1] https://resources.infosecinstitute.com/topic/cia-triad/
[2] Programa de compartilhamento de arquivos bastante popular em 2004.

Pode ser personalizado para atacar alvos específicos e sua principal função, na época em que estava em atividade, era realizar roubo de informações e espionagem, explorando vulnerabilidades *zero-day* do *Windows*. Continha estrutura e capacidades semelhantes ao *Stuxnet* (TRENDMICRO, 2011). Uma vez infiltrado, realizava ações de reconhecimento e identificação da topologia da rede, utilizava um *exploit* que permitia escalada de privilégios, a partir do qual passava a ter acesso de administrador, utilizando a técnica *pass-the-hash*, em que o sistema permite a utilização do *hash* armazenado diretamente para fazer login no lugar das senhas em texto claro.

Outro *malware*, considerado sucessor do *Stuxnet* e que compartilha características com aquele e o DuQu, é o *Flamer* (GOSTEV, 2012). Arma de ataque sofisticada, este *malware* é mais complexo que o DuQu, possui capacidades de *backdoor*, *trojan* e características de *worm* (estes termos serão definidos na Seção 2.2). Uma vez que a máquina está infectada, inicia a varredura e o monitoramento da rede (*sniffing*), faz *screenshots* da tela, grava conversas em áudio, faz *keylogging*, entre outros. O *Flamer* foi considerado uma das ameaças mais complexas jamais vista até então. Seu objetivo principal é o roubo de dados e envio das informações roubadas para seu servidor de Comando e Controle (C&C).

A partir dos exemplos acima, podemos perceber que diferentes *malware* podem ter objetivos e *modus operandi* em comum (alguns tem até partes de códigos semelhantes) (BLEEPINGCOMPUTER, 2019), porém com algumas características diferentes, de forma que estas podem ser utilizadas para dividi-los em classes. Na próxima seção, será apresentado um resumo sobre as classes utilizadas na literatura, o que introduzirá os conceitos necessários para o avanço do trabalho para a classe de *malware* mais utilizada (KASPERSKY, 2020) para ataques maliciosos atualmente: o *ransomware*.

2.2 Classificação dos *Malwares*

Dependendo das considerações acerca do funcionamento e do comportamento do *malware*, podemos classificá-lo de acordo com o os itens descritos abaixo (OR-MEIR et al., 2019; STALLINGS; BRESSAN; BARBOSA, 2008; SIKORSKI; HONIG, 2012):

- ***Virus*:** se espalha para outras máquinas anexando a si mesmo em outros arquivos ou documentos (neste caso, os que suportam macros, como pdf, xls e doc). Causa funcionamento inesperado ou prejudicial, danificando o sistema e corrompendo ou destruindo dados. Vírus também é o termo usado genericamente para descrever qualquer tipo de *malware*.

- ***Remote Access Trojan* (RAT ou *Trojan Horse*):** *software* malicioso que tem como subterfúgio enganar o usuário, fazendo-o pensar que é inofensivo ou mesmo que tem alguma funcionalidade que lhe seja útil. Seu nome é proveniente da mitologia grega, onde um cavalo oco de madeira foi entregue aos troianos pelos gregos como presente de trégua na guerra de Troia (depois de 10 anos de cerco), mas que tinha em seu interior muitos soldados inimigos. Em sua maioria, este tipo de *malware* utiliza engenharia social, que é um conjunto de técnicas empregadas por criminosos virtuais com o intuito de induzir usuários desavisados a enviar dados confidenciais, infectar seus computadores com *malware* ou abrir links para sites infectados, para então começarem sua campanha maliciosa.

- ***Logic Bomb***: monitora várias condições do sistema da vítima e dispara uma ação quando ocorre um determinado alinhamento entre essas condições, que pode ser uma ação do usuário, uma data predeterminada ou um comando do servidor de C&C. Esta situação os torna praticamente indetectáveis até que sejam ativados e entrem em ação. Para ser considerado uma bomba lógica, o *payload*[3] deve ser caracterizado como indesejado (programas *trial* com bloqueio por período de uso instalados pelo usuário não entram nesta categoria).

- ***Downloaders***: sua única função é baixar e instalar outros *malwares* em uma máquina sob ataque e normalmente a vítima o recebe por e-mail. Como não carrega consigo a carga útil maliciosa (*payload*), muitas vezes consegue escapar da detecção. Adicionalmente, pode criar entradas no registro do sistema para que o *malware* baixado tenha persistência e seja carregado após reinício da máquina.

- ***Spyware*:** *software* malicioso que espiona a atividade do usuário. Instalado sem seu consentimento, pode passar desapercebido, pois praticamente não causa instabilidade ou mudanças de desempenho na máquina infectada, funcionando em segundo plano e em alguns casos nem aparece na lista de processos do Gerenciador de Tarefas[4] ou do Monitor de Recursos[5]. Seu objetivo principal é monitorar a atividade do usuário, como páginas visitadas, localização geográfica, senhas, dados pessoais e reportar as informações coletadas ao gente malicioso.

- ***Rootkit*:** código malicioso projetado para esconder a existência de outro código.

[3] A carga útil é a parte do *malware* que executa uma ação maliciosa.

[4] Ferramenta que monitora o desempenho de vários recursos do PC, como memória, uso do espaço de armazenamento, CPU, entre outros elementos de hardware.

[5] O Monitor de Recursos é usado para verificar a quantidade de recursos do sistema consumida por cada programa ou serviço em execução.

Normalmente são utilizados em conjunto com outros *malwares*, como um *backdoor*, para permitir acesso ao atacante e dificultar a detecção do código pela vítima.

- ***Worm*:** programa que se replica automaticamente, seja por meio da Internet, mensagens, conexões locais, dispositivos USB ou arquivos. Explora vulnerabilidades nos sistemas para roubar informações sensíveis, instala *backdoors* que podem ser usados para que os agentes maliciosos possam acessar diretamente as máquinas infectadas. Geralmente causam lentidão pois consomem muita memória e banda de rede. Seu objetivo geralmente é roubar dados de usuários ou empresas.

- ***Adware*:** este *software* indesejado é projetado para mostrar anúncios ao usuário, gerando receita para seu operador. O problema é que o *adware* mostra anúncios de maneira inconveniente e, muitas vezes, em lugares incomuns para o usuário. Geralmente não causa maiores problemas no sistema da vítima e apresenta como sintomas comuns mudança na página inicial do navegador, redirecionamentos inesperados de links e falhas no navegador.

- ***Bot*:** *software* malicioso que deixa a máquina da vítima completamente sujeita à vontade do agente malicioso, de forma que esta passa a fazer parte de um exército de zumbis. Como zumbis, as máquinas contaminadas podem visitar sites, replicar o *malware* em outras máquinas e realizar consultas (DNS, por exemplo). Os *bots* recebem comandos de um servidor de C&C ou de outros *bots* da rede e um grupo de *bots*, chamamos de *botnet*. Sua principal função é realizar ataques de negação de serviço (DoS), onde uma quantidade de requisições muito maior do que a capacidade computacional[6] ou de rede do fornecedor de um determinado serviço consegue atender, causando congestionamento e/ou indisponibilidade.

- ***Cryptominers*:** utiliza a capacidade computacional da vítima para mineração de criptomoedas em favor do agente malicioso. Geralmente são bastante sutis em sua tarefa para que a vítima não perceba o que acontece em sua máquina. Uma maneira bem comum de utilizar *cryptominers* é esconder um minerador no código de uma página web, potencializando a mineração de acordo com a quantidade de visitas à página. Uma outra maneira de utilizar esses *softwares* maliciosos é infectando servidores, devido a sua grande capacidade computacional e disponibilidade 24 horas por dia.

[6]Capacidade computacional refere-se à capacidade de um sistema de computador realizar operações de processamento de dados. Em outras palavras, é a medida da habilidade de um computador em executar tarefas complexas em um tempo razoável.

- ***Ransomware*:** é um tipo de *malware* que infecta o computador da vítima e sequestra seus dados, impedindo acesso e cobrando um resgate para recuperação dos arquivos, que é cobrado em criptomoedas.

2.3 Técnicas de Detecção de *Malware*

"Toda lei tem uma brecha", diz um velho provérbio, significando que uma vez que uma regra é conhecida, também se sabe como evitá-la. Isso vale também para a corrida *malware* versus anti-*malware*. Sabendo como o *malware* funciona, pode-se projetar defesas mais eficazes; sabendo como funcionam as defesas, pode-se projetar *malwares* mais furtivos (GENÇ; LENZINI; RYAN, 2018). O primeiro passo para que possamos elevar o nível de proteção e tentar quebrar esse ciclo é detectando e estudando o *malware*, entendendo como funciona, quais artefatos utiliza, como se comporta em determinados ambientes e os motivos pelos quais foi criado. Para isso é importante que tenhamos ferramentas e técnicas que nos permitam extrair informações do *malware* e, a partir desse conhecimento, criar sistemas mais inteligentes, aplicando novas técnicas ou melhorando as técnicas já existentes. Na literatura existem algumas técnicas básicas de detecção de *malware*: assinaturas e heurísticas (que são as técnicas mais usadas em sistemas antivírus). Além disso, temos os métodos de análise: Análise Estática e Análise Dinâmica e ambientes virtualizados de análise.

Nesta Seção, mostraremos os conceitos básicos sobre antivírus, Análise Estática, Análise Dinâmica e sobre ambientes *sandbox*.

2.3.1 Antivírus

Os antivírus são programas que protegem uma máquina contra ações de programas maliciosos que sejam alheias à vontade do usuário. Trabalham basicamente de duas maneiras: detecção baseada em assinaturas e detecção baseada em anomalias (ou heurísticas) (MANGIALARDO; DUARTE, s.d.).

No primeiro método, um arquivo executável é analisado de acordo com seus blocos de código, que em um arquivo benigno normalmente são os seguintes:

.text : contém as instruções que a CPU executa. Todas as outras seções são para armazenamento de dados e informação de suporte. De maneira geral, esta é a única seção que pode ser executada e é a única que deve conter código.

.rdata : seção que tipicamente contém informação de importação e exportação. Esta seção também pode ter dados somente leitura usado no programa. Em alguns casos, um arquivo pode ter seções como .idata e .edata, que armazenam informações de importação e exportação.

.data : contém os dados globais do programa, acessível de qualquer lugar dentro dele.

.rsrc : seção que inclui recursos usados pelo executável que não são consideradas partes dele, como ícones, imagens, menus e *strings*. Strings podem ser armazenadas nesta seção ou no programa principal.

Dessa maneira, cada bloco é comparado com a base de assinaturas que o antivírus possui e quando há coincidência entre a base e o arquivo analisado, uma alarme é mostrado ao usuário e o arquivo é deletado ou separado para quarentena. A principal desvantagem desse tipo de detecção, é que não é possível criar assinaturas em tempo real, o que obriga as fabricantes a gerarem diariamente atualizações do banco de assinaturas para tornar os *malwares* conhecidos pelos seus antivírus e como consequência, não consegue detectar ameaças desconhecidas. Este método não é efetivo contra ataques *zero-day*.

O segundo método é uma tentativa de evitar os efeitos da defasagem da análise e distribuição de atualizações dos bancos de *malwares* conhecidos. Neste método, o antivírus tem um banco com conjuntos e regras comportamentais que um *malware* pode executar na máquina da vítima que são considerados comportamentos maliciosos. Podemos citar como exemplo o envio de três arquivos idênticos em sequência para um determinado endereço. Esta atividade é considerada maliciosa e o programa é marcado como vírus. A maior desvantagem desse método é a dificuldade de se traduzir todos os possíveis comportamentos maliciosos na forma de regras que identifiquem atividades suspeitas. Por esses motivos, as detecções por antivírus sofrem de dois problemas: alta taxa de falsos positivos e falsos negativos na classificação, identificando *software* benigno como malicioso e falhando ao detectar o *malware*, respectivamente.

2.3.2 Métodos de Análise

A análise de *malware* é o estudo de amostras de *malware*, com o objetivo de determinar suas funcionalidades e extrair a maior quantidade de informações possíveis, dissecar o programa malicioso para entender como ele funciona, descobrir como o identificar, como evitar que se replique e como eliminá-lo (CALDAS, 2016). A análise de *malware* pode ser

executada basicamente de duas formas: Análise Estática e Análise Dinâmica, que serão mais detalhadas nas próximas Subseções.

2.3.2.1 Análise Estática

A Análise Estática se dá pela inspeção do código binário do programa com uma ferramenta *disassembler*, como o IDA[7], a fim de se verificar se existe alguma regra de segurança não conforme sem, no entanto, executar o programa. Pode ser aplicada no código fonte ou no arquivo binário, gerado após a compilação do programa (CALDAS, 2016). Este método de análise inicial envolve a extração de informações úteis do arquivo binário e, a partir daí, direciona a classificação e a realiza uma análise mais profunda da amostra (MONNAPPA, 2018). Esta tarefa requer habilidades de engenharia reversa e um conhecimento prévio do funcionamento de sistemas operacionais e, dependendo do objetivo desejado, pode ser bastante demorado e trabalhoso. Um dos grandes problemas da análise estática é o empacotamento do *malware*. Este artifício consiste em utilizar o binário do programa como um vetor de infecção, cujo único objetivo é descompactar a carga maliciosa e executá-la, evitando assim detecção de antivírus. Este método de evasão tem muitas vantagens, uma delas é fazer da Análise Estática uma técnica ineficiente, pois as tabelas de importação das funções utilizadas estão restritas somente a poucas funções que muitas vezes não denunciam o comportamento malicioso do *malware*. Além disso, ao analisar um código desmontado, não se sabe quais partes do código são efetivamente executadas e o estado da memória e dos registradores durante a execução. Existem algumas técnicas de empacotamento mais elaboradas do que simplesmente carregar o código malicioso em outro programa. Em alguns casos, o código malicioso é virtualizado e emula o comportamento de um processador para rodar o código original do *malware* (ROCCIA, 2017). Este código é traduzido para *bytecode* da máquina virtual e esta técnica é uma das mais difíceis de se fazer análise.

Como vantagem, na Análise Estática, o analista não fica limitado ao fluxo de execução como na Análise Dinâmica. Isto faz com que seja útil para descobrir sobre funções que seriam executadas apenas em condições específicas (analisam o *malware* de acordo com sua estrutura, fluxo de controle etc). Um bom exemplo é uma função que seria executada apenas depois de se ter realizado comunicação com o servidor de C&C. Adicionalmente, graças ao fato de que a Análise Estática não implica execução da amostra, este acaba sendo um método mais seguro de análise. Também há que se considerar que, neste tipo

[7] https://hex-rays.com/ida-pro/

de análise, não se está preso a executáveis restritos ao Sistema Operacional utilizado como plataforma. Por exemplo, seria perfeitamente possível fazer análise estática de binários ARM ou ELF em máquinas *Windows*, desde que as ferramentas utilizadas suportem o formato analisado (NDIBANJE et al., 2019).

2.3.2.2 Análise Dinâmica

A Análise Dinâmica, por outro lado, consiste em executar uma determinada amostra de *malware* em um ambiente virtualizado controlado e monitorar suas ações, com o objetivo de analisar seu comportamento. Ao ser executado, o *malware* altera chaves de registro e tenta mudar seu modo de execução para obter mais privilégios, quando então, consegue fazer alterações mais significativas no Sistema Operacional. Na Análise Dinâmica, o *malware* tem acesso total aos recursos do ambiente controlado, inclusive consegue executar em modo *debugger*. Ao final da execução do *malware*, o ambiente retorna a um estado anterior. O monitoramento é feito através da coleta de informações do sistema, como *logs*, arquivos alterados, processos criados, bibliotecas utilizadas pelo *malware*, comunicações de rede e protocolos de comunicação, entre outros (CALDAS, 2016).

A Análise Dinâmica pode ser manual ou automatizada. Ambas as abordagens possuem suas características, vantagens e desvantagens. A Análise Dinâmica manual apresenta o mesmo problema da análise Estática, no sentido de que poucas amostras podem ser analisadas por vez, inviabilizando a geração de conjunto de dados abrangente para aplicação em Aprendizado de Máquina, porém é uma ótima ferramenta quando se deseja estudar mais profundamente o funcionamento de determinado *software*. As ferramentas de Análise Dinâmica manual permitem que o *software* em análise seja executado passo a passo, o que dá ao analista a possibilidade de acompanhar o conteúdo da memória, ponteiros, requisições e as alterações na máquina da vítima. Já a análise automatizada permite capturar o funcionamento de muitas amostras de *software* ao mesmo tempo, ao custo de que o analista abdique um pouco do controle sobre os detalhes da execução. A implementação normalmente é feita através de *hooks* e DLL *injections* (SIKORSKI; HONIG, 2012), e apesar de mais complexa, traz bons resultados quando técnicas de polimorfismo e ofuscação são utilizadas pelas amostras de *malwares* para evasão (CALDAS, 2016).

Muitos sistemas foram propostos para observar o comportamento dinâmico de arqui-

vos executáveis, como o *CWSandbox*[8], *Anubis*[9] e *Cuckoo Sandbox*[10]. Estes ambientes permitem executar tanto programas maliciosos quanto benignos em um ambiente isolado, monitorar e analisar seu comportamento. Infelizmente, apenas o *Cuckoo Sandbox* é *open source*, e por isso foi escolhido para ser utilizado neste trabalho. Os detalhes da implementação e configuração do sistema de análise serão explorados no Capítulo 5.

2.3.3 Ambiente de Análise

Com o desenvolvimento da tecnologia de virtualização, a Análise Dinâmica de *malware* alcançou grande avanço ao permitir a construção de um ambiente seguro e controlável, facilitando o monitoramento do processo de execução e capturando as informações em tempo de execução. Este método é considerado o método mais eficaz para análise de *malware* (WANG et al., 2019). A ferramenta utilizada neste trabalho para virtualização foi o *VirtualBox*, controlado pelo *Cuckoo Sandbox* instalado em uma máquina hospedeira que realiza o monitoramento e compila as informações geradas pela análise.

2.3.3.1 *Cuckoo Sandbox*

O *Cuckoo Sandbox* foi criado por Guarnieri (GUARNIERI et al., 2012), como uma ferramenta aberta, flexível e escalável (MILLER et al., 2017). Esta ferramenta consiste em uma máquina hospedeira *Cuckoo* e uma ou mais máquinas virtualizadas. A máquina virtualizada se comunica com a máquina hospedeira por meio de uma rede virtual restrita, cuja estrutura na máquina *host* é um *hypervisor* que gerencia essas máquinas convidadas em vários ambientes diferentes, de acordo com a necessidade e os objetivos do analista. A máquina *host* é responsável por agendar, iniciar a análise, monitorar o comportamento malicioso e gerar relatórios de análise. A máquina convidada possui um ambiente virtualizado e restrito para execução, monitoramento e transmissão dos resultados da análise de volta ao *Cuckoo* por meio da rede virtual. Atualmente está na versão 2.0.7 e permite a submissão de arquivos que serão executados em ambiente isolado. Uma vez submetida uma amostra, o *Cuckoo* carrega na VM um *snapshot*, para retornar o sistema para um ponto livre de *malware*, para então transferir e executar o arquivo a ser analisado. Enquanto a amostra está sendo executada, o *Cuckoo* coleta informações da execução, como chamadas de API, arquivos baixados, tráfego de rede, *strings* contidas no arquivo, *packers*, *dump* da memória etc. O *Cuckoo Sandbox* tem uma API REST integrada que

[8] http://cwsandbox.org/
[9] https://www.anubisnetworks.com/
[10] https://cuckoosandbox.org/

permite realizar as operações de submissão de arquivos, acompanhamento das análises, monitoramento da máquina e *download* de relatórios, facilitando a integração com outros sistemas e a mineração dos dados gerados pelas análises.

Na tentativa de ajudar o analista a diferenciar a periculosidade das amostras analisadas, o *Cuckoo* oferece ao usuário um sistema de *score* que pontua o comportamento malicioso de cada amostra analisada. Esse *score* é baseado na compatibilidade do *malware* analisado com assinaturas conhecidas, que são disponibilizadas no repositório do *Guthub* da ferramenta. É calculado adicionando os níveis de severidade atribuído a cada assinatura compatível e, ao final, o valor é dividido por 5. Apesar de o cálculo do *score* ser de certa forma arbitrário, um alto *score* pode indicar que um arquivo é malicioso, mas para realmente ter uma avaliação precisa, deve-se conferir cada assinatura individualmente. O mesmo vale para um baixo *score*, o que não significa necessariamente que o arquivo não seja malicioso. O casamento de assinaturas é baseado em informações de comportamento coletadas. Um *score* de 1 a 3, quando se espera mais, pode significar que algum comportamento não foi coletado corretamente, devido a interrupção da amostra, a não execução da amostra ou algum outro problema (WANG et al., 2019). Na figura 1, apresentamos uma imagem com a interface web do *Cuckoo Sandbox* com as regras do *score* atribuído a uma amostra analisada.

Para evitar que as amostras de *malware* detectassem sua execução em ambiente virtualizado, utilizamos o *Paranoid Fish* (PaFish), que é uma ferramenta de código aberto que, ao ser executada, procura e relata ao usuário se há traços que denunciem a virtualização do sistema. As verificações são feitas em vários aspectos da máquina, como presença de *debuggers*, teste de *Turing* Reverso (FRENCH, 2000), informações de CPU, tamanho do disco, memória RAM, nome de usuário, número de processadores e outros específicos da ferramenta de virtualização, como entradas características no registro do *Windows* e endereços MAC padrão de interfaces de rede e instalação de adicionais de convidados (OKTAVIANTO; MUHARDIANTO, 2013).

Além de evitar a detecção do ambiente virtualizado, também tivemos o cuidado de oferecer às amostras analisadas um ambiente com *networking* funcional, no sentido de que nossa rede estivesse protegida da propagação dos *malwares* analisados, ao mesmo tempo que estes tivessem a sua disposição serviços de rede comuns na internet. Para a disponibilização desses serviços utilizamos o INETSIM. O INETSIM é uma ferramenta livre baseada em Linux que simula serviços comuns de internet, o que possibilita analisar o comportamento de rede de *malwares* desconhecidos, ao emular serviços como HTTP,

Figura 1: Seção da interface web do Cuckoo Sandbox mostrando as regras consideradas para composição do *score* de uma amostra.

HTTPS, FTP, IRC, DNS e SMTP. A opção em favor do uso desta ferramenta, em vez de deixar o *malware* livre para se comunicar com a internet, se dá por dois motivos: o primeiro é que haveria a possibilidade de contaminação de outros dispositivos ligados na mesma rede pelo *malware* em análise; e o segundo e principal motivo é que os operadores poderiam receber informações de muitas amostras do seu *ransomware* sendo executado em uma mesma máquina com um mesmo IP público, e isso poderia chamar atenção e desencadear ataques à rede local. Dentre outros artifícios usados pelo INETSIM, estão a resposta com *banner* do *Microsoft* IIS[11] *web server* quando é interrogado. Também responde com um arquivo falso quando requisitado, como por exemplo quando um *malware* requisita um arquivo de imagem JPEG de algum site para continuar executando, o INETSIM responde com um arquivo falso nesse formato, evitando uma resposta com código 404 (SIKORSKI; HONIG, 2012), o que potencialmente acusaria o ambiente de análise ao *malware*.

2.4 Aprendizado de Máquina

Aprendizado de Máquina é um sistema que pode modificar seu comportamento de maneira autônoma tendo como base a própria experiência, que é conseguida através de treinamento. Esta capacidade de aprendizado diminui drasticamente a necessidade de interferência humana. Esta modificação do comportamento consiste na construção de regras lógicas que visam melhorar o desempenho de uma tarefa ou tomada de decisão, dependendo do contexto. As regras lógicas são construídas a partir do reconhecimento de padrões dentro dos dados analisados (ZHAO et al., 2018).

Depois de cobrirmos alguns princípios básicos nesta seção, teremos uma boa ideia de como funcionam os algoritmos, porque foram escolhidos e como foram utilizados neste trabalho. O nível de detalhes sobre cada técnica utilizada será suficiente para o entendimento desse trabalho, visto que que não teremos como cobrir todas as nuances e complexidades desses algoritmos. As particularidades de cada algoritmo utilizado são apresentadas na Seção 2.4.1.

A grande vantagem do Aprendizado de Máquina é a capacidade de classificar padrões novos a partir de padrões previamente conhecidos. Neste caso, separamos os conjuntos de dados em duas partes: a primeira é utilizada para apresentar as características de cada amostra com seus respetivos rótulos ao algoritmo de detecção, o que chamamos de treinamento e, uma vez tendo o algoritmo treinado, apresentamos a segunda parte do

[11] https://www.iis.net/overview

conjunto de dados, que para o algoritmo são dados inéditos, e solicitamos então que sejam classificados. Depois da classificação, fazemos a comparação das classes com o gabarito, que são os rótulos daquelas amostras que tínhamos previamente e que não foram usadas no treinamento. A partir desta comparação, conseguimos verificar o acerto ou erro na classificação e medir o desempenho do classificador em relação àquela tarefa. Uma outra maneira de pensar o problema é que o Aprendizado de Máquina é o processo de usar dados históricos para criar um algoritmo de predição para dados futuros. A classificação pode ser binária, em que temos apenas duas classes ou multi classe, como quando se quer classificar um *malware* em *keylogger*, *ransomware* ou *remote access trojan* (FREEMAN; CHIO, 2018).

Aprendizado de Máquina também é usado para resolver problema de clusterização (agrupamento), onde, possuindo um conjunto de dados, em que esses dados são representados por pontos em um espaço n-dimensional, podemos tentar descobrir quais são similares entre si, como quando estamos tentando analisar um grande conjunto de dados de tráfego de internet para uma determinado site, podemos querer saber quais requisições podem ser agrupadas. Alguns *clusters* podem ser *botnets*, enquanto outros podem ser usuários legítimos. A tarefa considerada é apenas uma classificação: determinar a qual classe pertence um determinado ponto inédito.

O Aprendizado de Máquina tem ainda outras duas facetas com relação ao modo de classificar as amostras no conjunto de dados, a primeira delas é chamada de Aprendizado Supervisionado, no qual se tem as classificações de dados históricos e tentamos prever as classes de dados futuros. Este tipo de AM é o utilizado quando temos um grande conjunto de e-mails classificados como maliciosos ou legítimos, podemos treinar um classificador que tenta predizer se as novas mensagens recebidas são legitimas ou *spam*. Alternativamente, no Aprendizado de Máquina Não-Supervisionado os dados históricos não estão rotulados e o algoritmo aprende padrões de dados não rotulados sem nenhuma orientação ou supervisão explícita do usuário. Este tipo de AM é utilizado, por exemplo, quando temos um número desconhecido de ataques de *botnet* na rede e queremos distingui-los. Classificação e Regressão são exemplos de Aprendizado Supervisionado, e clusterização é um exemplo típico de Aprendizado não-Supervisionado. Neste trabalho, faremos uso das duas abordagens para classificação de *ransomware*.

2.4.1 Algoritmos de Classificação

Neste trabalho, foram aplicados alguns algoritmos de Aprendizado de Máquina (FACELI et al., 2021). Abaixo temos uma breve explicação do funcionamento de cada um deles:

Decision Tree (DT) é um método indutivo de Aprendizado de Máquina. Neste método, uma árvore é construída selecionando-se os atributos que melhor divide os exemplos de treino nas diferentes classes. A raiz é o inicio da classificação, e os nós representam as classes. Durante a classificação, a amostra desce a árvore através de cada nó, que representam cada atributo. Nas Árvores de Decisão, os efeitos podem ser melhorados principalmente pelo ajuste das métricas que avaliam a qualidade de uma características e ajustando a profundidade máxima da árvore.

Random Forest (RF) é um método de aprendizado *ensemble* e bem similar a Àrvore de Decisão. Em geral, as Árvores de Decisão aplicam todas as sequências de características para gerar o modelo. No entanto, nas *Random Forest*, apenas algumas características são selecionadas para produzir a Árvore de Decisão como um *base learner* (classificador básico), e vários *base learners* são combinados em um mecanismo de votação. Em *Random Forest*, podemos modificar o número de características a considerar para cada nó e a profundidade máxima de cada classificador para otimizar a classificação.

K-Nearest Neighbor (KNN) é um algoritmo de aprendizado baseado em instâncias. Neste algoritmo, o classificador distribui as isntâncias em um espaço n-dimensional e identifica os K vizinhos mais próximos de algumas instâncias de treinamento relativamente a nova a amostra e depois marca a classe de acordo com o maior número de vizinhos relativos à nova instância analisada. O agrupamento no qual a amostra de teste pertence é baseado na distância dos K pontos.

Naive Bayes (NB) é um método de aprendizado baseado na probabilidade. Ele calcula a probabilidade de uma amostra pertencer a uma determinada família utilizando o teorema de Bayes, o qual assume que todas as características são independentes entre si. Este método necessita quantidade pequena de dados para treinamento e é muito útil em categorização de textos.

Support Vector Machines (SVM) na sua forma mais simples é um classificador linear, o que significa que ele produz um hiperplano em um espaço de vetores que tenta separar os dados em classes ou grupos de dados dentro do conjunto utilizado. A SVM tenta encontrar o hiperplano com maior margem separando as duas classes, onde "margem" significa a distância do plano de separação até os pontos de dados mais próximos de cada

lado. Para os casos em que os dados não são linearmente separáveis, pontos dentro da margem são penalizados proporcionalmente a sua distância à margem.

Multilayer Perceptron (MLP) é um algoritmo de aprendizado supervisionado que aprende a função $f(\cdot) : R^m \rightarrow R^o$ a partir de um conjunto de dados, onde m é a quantidade de dimensões da entrada e o é a quantidade de dimensões da saída. Dado um conjunto de características $X = x_1, x_2, \ldots, x_m$ e um alvo, o MLP aprende um aproximador da função não linear para classificação ou regressão. Entre as camadas de entrada e saída, o MLP pode ter uma ou mais camadas não lineares, chamadas de camadas ocultas.

Neste trabalho, os algoritmos *GridSearchCV*, *StandardScaler* e PCA foram utilizados sobre o conjunto de dados para que pudéssemos verificar sua viabilidade tanto no que se refere a melhora no desempenho dos algoritmos de classificação quanto para redução no seu tempo de processamento.

2.4.2 Padronização dos dados

A padronização de conjuntos de dados é um requisito comum para muitos algoritmos de Aprendizado de Máquina implementados no *SciKit-Learn*[12], eles podem apresentar baixo desempenho se os dados não estiverem padronizados com distribuição normal: Gaussiano com média zero e variância unitária. Na prática, muitas vezes ignoramos a forma da distribuição e apenas transformamos os dados para centralizá-los, removendo o valor médio de cada valor e, em seguida, dimensionando-o dividindo os valores não constantes por seu desvio padrão. A centralização e o dimensionamento acontecem independentemente em cada característica, calculando as estatísticas relevantes nas amostras do conjunto de treinamento. A média e o desvio padrão são armazenados para serem usados em dados posteriores usando a mesma transformação.

No caso do conjunto de dados de chamadas de API (que será apresentado com maiores detalhes na Seção 6.2.1), podemos verificar grande variabilidade de valores, por exemplo, chamadas às API da biblioteca *Crypto* API do Windows (*WinAPI*[13]) como *CryptGenKey*, *CryptEncrypt*, *CryptDecrypt* e *CryptExportKey*, são algumas das principais utilizadas por *ransomwares* devido ao fato de estarem diretamente ligadas a sua finalidade, que é encriptar os arquivos do usuário. Por exemplo, muitos elementos usados na função objetivo de um algoritmo de aprendizado (como o *kernel* RBF da SVM) podem assumir que todos os valores são centrados em torno de zero ou têm variância na mesma ordem. Se

[12]https://scikit-learn.org/stable/
[13]https://learn.microsoft.com/en-us/windows/win32/apiindex/windows-api-list

Figura 2: Resultado da aplicação do *Standard Scaler* ao conjunto de dados de chamadas de API para classificação multi classe.

```
sc = StandardScaler()
table_sc = sc.fit_transform(table)
table_sc
```

```
array([[-0.70240463, -0.23207058, -0.05680101, ..., -0.1554926 ,
        -0.03074377, -0.03074377],
       [ 0.86081814, -0.14518093, -0.0409471 , ..., -0.1554926 ,
        -0.03074377, -0.03074377],
       [-0.70202958, -0.20803727, -0.05582963, ..., -0.1554926 ,
        -0.03074377, -0.03074377],
       ...,
       [-0.70240463, -0.22837315, -0.05680101, ..., -0.1554926 ,
        -0.03074377, -0.03074377],
       [-0.70202958, -0.20803727, -0.05582963, ..., -0.1554926 ,
        -0.03074377, -0.03074377],
       [ 0.85931793, -0.1525758 ,  0.02217675, ..., -0.1554926 ,
        -0.03074377, -0.03074377]])
```

uma característica tem uma variância que é ordem de grandeza maior do que outras, ela pode dominar a função objetivo e tornar o estimador incapaz de aprender com outras características corretamente. Na Figura 2, podemos observar o resultado descrito anteriormente na aplicação do *Standard Scaler* ao conjunto de dados de chamadas de API para classificação multiclasse.

2.4.3 Redução de Dimensionalidade

Dados com alta dimensionalidade são difíceis de treinar, precisam de mais poder computacional, demoram mais e são mais difíceis de visualizar. Além disso, a dimensionalidade está diretamente ligada à quantidade de memória necessária para carregar os dados. A redução de dimensionalidade pode levar a um melhor desempenho na classificação e menor tempo de processamento ao remover recursos redundantes, obsoletos e altamente correlacionados. Neste trabalho, escolhemos utilizar a Análise de Componentes Principais (PCA). O PCA é uma técnica popular para analisar conjuntos de dados contendo um grande número de características (dimensões) por observação, aumentando a interpretabilidade dos dados enquanto preserva a quantidade máxima de informações, permitindo visualização de dados multidimensionais. Formalmente, o PCA é uma técnica estatística para reduzir a dimensionalidade de um conjunto de dados, o que é feito pela transformação linear dos dados em um novo sistema de coordenadas onde a maior parte da variação nos dados pode ser descrita com menos dimensões do que os dados iniciais. A análise de componentes principais tem aplicações em muitos campos, como genética populacional, estudos de microbioma e ciência atmosférica. Para aplicarmos o PCA a um conjunto de dados, é obrigatório fazer a padronização das características antes de aplicá-lo se houver

uma diferença significativa na escala entre as características do conjunto de dados, por exemplo, quando uma característica varia em uma amostra entre 0 e 1 e em outra entre 100 e 1.000. Isso ocorre porque o PCA é muito sensível aos intervalos relativos das características. O PCA é usado para decompor um conjunto de dados multivariados em um conjunto de componentes ortogonais sucessivos que explicam uma quantidade máxima da variância. No *SciKit-Learn*, o PCA é implementado como um objeto transformador que aprende componentes em seu método de ajuste e pode ser usado em novos dados para projetá-los nesses componentes.

2.4.4 Mineração de texto

O acrônimo TF-IDF significa *Term Frequency* (LUHN, 1958) - *Inverse Document Frequency* (JONES, 1972) (Frequência de Termos, Frequência Inversa de Documento). É uma abordagem típica da área de Processamento de Linguagem Natural para identificação de palavras importantes e é baseada na teoria de modelagem de linguagem (ZHANG et al., 2019). Esta abordagem possui a heurística intuitiva de um termo que aparece em muitos documentos não é um discriminador muito bom e deve ter menos peso na classificação do que um que um termo que ocorre em poucos documentos. O TF-IDF dos termos de cada seção selecionada dos relatórios foi calculado utilizando-se as Fórmulas 2.1, 2.2 e 2.3, abaixo:

$$TF(i,j) = \frac{n(i,j)}{\sum_k n(k,j)} \tag{2.1}$$

onde $TF(i,j)$ representa a frequência do termo i em cada documento j, $n(i,j)$ representa a quantidade de vezes que o termo i aparece no documento j, e $\sum_k n(k,j)$ é a quantidade de termos no documento. Então,

$$IDF(i,j) = log_2 \frac{|D|}{|(j : i \in j | j \in D)|} \tag{2.2}$$

Onde o $IDF(i,j)$ descreve se o termo é raro nos relatórios. D é o conjunto de todos os relatórios e $|j : i \in j | j \in D|$ é o número de relatórios que contém aquele termo. Finalmente,

$$TF \cdot IDF(i,j_f) = TF(i,j_f) \times IDF(i) \tag{2.3}$$

Na Equação 2.3, combinamos todas as sequências de termos dos relatórios referentes a cada família f para formar uma longa sequencia de termos j_f. TF-IDF(i,j_f) representa o valor do TF-IDF do termo i em toda a sequência de termos j_f.

O TF-IDF avalia a importância de um determinado termo em um relatório, que é diretamente proporcional a frequência de sua ocorrência. Além disso, sua importância na discriminação entre os relatórios é inversamente proporcional a quantidade de relatórios em que determinado termo aparece. O TF-IDF pode distinguir cada família devido a particularidades na ocorrência de cada termo em determinada família. Depois de obtermos os vetores de características de cada amostra, foram utilizados os índices de análise gerados pelo *Cuckoo Sandbox*, para associar os respectivos rótulos. Ao usarmos o IDF, minimizamos o peso das frequências dos termos mais comuns enquanto termos de aparecem em poucos documentos tenham um maior impacto.

A principal vantagem do TF-IDF advém da simplicidade de ser calculado e da facilidade de uso. É computacionalmente barato e é um ponto de partida simples para cálculos de similaridade (VAJJALA et al., 2020). A grande desvantagem desta abordagem é que, quando se tem um vocabulário muito grande e/ou modelo esparso e torna-se custoso em termos de uso de memória. Além disso esta abordagem não considera o significado semântico, visto que consideramos cada termo isoladamente dos termo adjacentes no texto.

3 *Ransomware*

Os *ransomwares* têm mantido a liderança como maior ameaça cibernética desde 2005 (GROOT, 2020), porém os primeiros ataques deste tipo já haviam acontecido há bastante tempo. O primeiro ataque conhecido de *ransomware* é datado de 1989 (GRUSTNIY, 2021), quando Joseph L. Popp, biólogo pesquisador, criou o *trojan* AIDS e o distribuiu a pesquisadores dessa área em 20 mil disquetes, rotulados como um programa que poderia ajudar a analisar o risco de um indivíduo contrair AIDS (GANDHI et al., 2017). Nos discos, o conteúdo apenas disfarçava um programa que inicialmente permanecia dormente e se ativava somente após 90 partidas da máquina. Depois de ativado este gatilho, o *malware* criptografava os nomes dos arquivos no computador da vítima e exibia uma janela dizendo que o período de teste do *software* havia acabado e que o usuário deveria pagar 189 dólares por um ano ou 378 dólares pelo acesso vitalício. Este *ransomware* também ficou conhecido como *PC Cyborg*. Apesar de ser bastante rudimentar (considerando o que temos nos dias atuais), para a época foi uma ideia bastante inovadora: o autor utilizou um argumento legítimo para distribuir um *software* malicioso, servindo seu propósito de enganar usuários incautos e arrecadar dinheiro.

Os *ransomwares*, em geral, funcionam como uma ferramenta de pós exploração, o que significa que o agente malicioso já conseguiu acesso prévio ao sistema, possivelmente resultado de campanha de *phishing* ou exploração de alguma vulnerabilidade (AHLGREN, 2019). Quando executado, o *ransomware* inicialmente realiza um reconhecimento do sistema, vasculha programas instalados, verifica também se está rodando dentro de uma VM (neste caso, pode ficar inativo), recolhe informações sobre a máquina hospedeira e a rede onde ela funciona e lista os arquivos que serão criptografados. Após a fase de análise, o *ransomware* utiliza chaves criptográficas (que podem ter sido compiladas com o código, geradas localmente ou baixadas do servidor C&C) para encriptar os arquivos

locais, utilizando, em muitos casos, AES[1] e/ou RSA[2]. Uma das características dos *ransomwares* mais modernos é a capacidade de fazer varredura na rede onde se encontra a máquina infectada e utilizar vulnerabilidades, *como Eternal Blue*, presente no *Windows* (utilizada também pelo *Wannacry* e *NotPetya*) ou credenciais conhecidas (como senhas padrão de equipamentos ou serviços comuns), para infectar mais máquinas e se espalhar pela rede (GRUSTNIY, 2021).

Existem três tipos básicos de *ransomware*: *scareware*, *window blockers* e *crypto-ransomware* (KOK; ABDULLAH; JHANJHI; SUPRAMANIAM, 2019). O primeiro tipo, ***scareware***, tenta extorquir a vítima se apresentando como um antivírus, alegando que sua máquina está infectada e que a compra do *software* oferecido irá resolver seu problema de infecção, quando a função real é roubar informações da vítima. A infecção pode ocorrer por campanhas de *phishing* ou *popups* implantados maliciosamente em páginas comprometidas. Em outras situações, sua tática é assustar a vítima para que pague um resgate, fingindo ser uma autoridade (como o FBI) que encontrou material ilegal em sua máquina e sugerindo à vitima, que pagar uma multa evitará problemas judiciais. Em alguns casos ameaça expor as ilegalidades encontradas aos amigos e familiares da vítima (KOK; ABDULLAH; JHANJHI; SUPRAMANIAM, 2019).

O segundo e o terceiro tipo são bastante parecidos, mas têm uma diferença sutil e crucial: os ***blockers*** bloqueiam o navegador ou o Sistema Operacional com uma janela *pop-up*, impedindo o usuário de acessar o sistema. Este tipo de *ransomware* não faz encriptação nos arquivos da vítima e, por isto, pode ser facilmente reversível: basta reinstalar o sistema operacional na máquina infectada. Uma evolução no *modus operandi* dos *windows blockers* é a capacidade de infectar o MBR (*Master Boot Record*) de uma máquina vulnerável e impedir que o Sistema Operacional carregue. Para executar esta ação, o *malware* copia o MBR original e o substitui com código malicioso, forçando a máquina a fazer *reboot* para que sua alteração faça efeito, mostrando a notificação de extorsão quando o sistema carrega novamente (MICRO, 2021). Um exemplo de *ransomware* com este comportamento é o *Reveton* (MICRO, 2021), que apresentava ao usuário mensagens de agências de aplicação da lei do país em que a vítima se encontra, informando que a

[1] AES significa *Advanced Encryption Standard*, que é um algoritmo de criptografia amplamente utilizado para cifrar dados. É um algoritmo de criptografia de chave simétrica, o que significa que a mesma chave secreta é usada tanto para a criptografia quanto para a descriptografia dos dados, é considerado um algoritmo de criptografia altamente seguro e é amplamente utilizado em diversas aplicações, incluindo comunicação segura, gerenciamento de direitos digitais e armazenamento de dados.

[2] RSA é um algoritmo de criptografia de chave pública que foi desenvolvido por Ron Rivest, Adi Shamir e Leonard Adleman em 1977. É amplamente utilizado para proteger informações confidenciais, incluindo segurança de redes, criptografia de e-mails, autenticação de usuários, criptografia de senhas e transações financeiras.

vítima foi flagrada com atividades ilegais *online*. Para saber qual era agência a apresentar, utilizava a localização geográfica das vítimas, no caso de um cidadão localizado na América do Norte, a mensagem aparecia como sendo do FBI, para um cidadão localizado na França, a mensagem aparecia como sendo da *Gendarmerie Nationale*. O terceiro tipo, ***crypto-ransomware***, são bem mais complicados de lidar, pois encriptam os arquivos das vítimas, deixando-os inacessíveis sem uma chave de desencriptação válida e mesmo tentativas de ataques de força bruta demorariam milhares de anos, dependendo do tamanho da chave e do tipo de encriptação usados. Um *ransomware*, na maioria das vezes, permite a vítima desencriptar alguns arquivos como incentivo ao pagamento do resgate, como prova de "boa fé" (geralmente poucos arquivos e de tamanho pequeno). Este trabalho terá foco neste último tipo apresentado.

Diante do que foi apresentado, temos uma amostra de como um *ransomware* pode ser complexo. Seus desenvolvedores os aperfeiçoam e inserem novas funcionalidades para que atinjam seus objetivos de forma cada vez mais eficaz e permaneçam furtivos. Para conseguirem cumprir sua missão, utilizam artifícios que são características marcantes de outros *malwares* (STALLINGS; BRESSAN; BARBOSA, 2008), como a capacidade de se espalhar por uma rede como um *worm* e personificar a vítima e enviar cópias de si para a lista de contatos de e-mail, persistência, *keylogging*, restrições de recuperação do sistema da vítima, modo *stealth*, mapeamento de ambiente e elevação de privilégios (KOK; ABDULLAH; JHANJHI; SUPRAMANIAM, 2019).

Mesmo com todo esforço dos agentes maliciosos para desenvolvimento de *ransomwares* cada vez melhores, assim como acontece com softwares não-maliciosos, também estão sujeitos a falhas de projeto e/ou implementação e de ecossistema. Como exemplo de falha de ecossistema, temos a situação do *wannacry*, que ao infectar máquinas com *Windows* XP (SP1 e SP2), que tem problemas de baixa entropia do seu PRNG (PARISOT; BENTO; MACHADO, 2021; COMPUTERWORLD, 2007) e era usada pelo *ransomware* para executar a cifração dos dados da máquina da vítima, tornando possível rastrear a geração das chaves que foram ou que serão geradas, permitindo recuperar os arquivos cifrados (SYMANTEC, 2017). Como exemplo de problemas de implementação, temos o *LockBit*. Em algumas versões, tanto o encriptador (o próprio *ransomware*) quanto o decriptador possuem problemas que os impedem de funcionar corretamente e em consequência, torna o pagamento do resgate ineficaz. Apesar disso, o problema não é irreversível em nenhuma das duas situações, como mostra os relatórios de estudos realizados pela *Microsoft* (VELUZ, 2022a,b).

Na literatura, existem algumas formas de dividir as ações de um *ransomware* em uma máquina hospedeira, que podemos chamar de *ciclo de vida*. Neste trabalho, iremos utilizar a seguinte divisão (ZUHAIR; SELAMAT; KREJCAR, 2020), que será detalhada nas próximas seções:

- Infecção
- Levantamento de Informações
- Escalação de Privilégios
- Ofuscamento
- Comunicação com Servidor C&C
- Persistência
- Criptografia
- Prevenção de Recuperação
- Propagação
- Pagamento do Resgate

Neste trabalho, o termo *malware* será utilizado nas situações em que a capacidade explicitada não pertença somente aos *ransomwares*.

3.1 Infecção

A infecção por *malware* ocorre quando o há infiltração na máquina da vítima. As ações que os operadores de campanhas de *ransomware* utilizam para realizar a infecção são bastante variadas. Geralmente conseguem acesso em decorrência de campanhas de *phishing*, aproveitando-se de usuários distraídos. Suas vítimas, na maioria das vezes, são grandes empresas (IBMSECURITY, 2022). Os métodos de infecção incluem e-mails falsos com *scripts* maliciosos, VB macros dentro de documentos do *MS Office*, URL direcionadas para aplicações maliciosas, que são baixadas e executadas na máquina da vítima (CRACIUN; MOGAGE; SIMION, 2018), acessos RDP e uso de contas válidas. Os agentes maliciosos podem também ter uma abordagem mais direcionada usando RDP e *PsExec*[3]

[3]https://docs.microsoft.com/en-us/sysinternals/downloads/psexec

conjuntamente para tomar o controle da rede e depois implantar a carga maliciosa. Outra forma de acesso inicial vista mais recentemente é o comprometimento da cadeia de suprimento, como ocorrido no incidente da *Kaseya* (TRENDMICRO, 2021d).

As campanhas de *phishing* e-mail podem ser massivas. As listas de possíveis vítimas podem ser encontradas vazamentos na internet contendo de milhares ou até milhões de contatos. Um exemplo de grande vazamento de dados pessoais ocorreu em junho de 2022, onde supostamente, dados de 1 Bilhão de chineses foi roubado da Polícia Nacional de Xangai, explorando-se uma das vulnerabilidades mais básicas: o desenvolvedor inadvertidamente postou as credenciais de acesso aos dados em um código fonte numa publicação de artigo escrito para um blog de tecnologia chinês[4] (LU, 2022; ZHENG, 2022; XIONG; RITCHIE; GAN, 2022). Esta grande quantidade de dados pessoais pode ser utilizada em uma massiva campanha de *phishing*. Supondo-se que tenhamos uma taxa de retorno de 1%, em valores absolutos, seriam 10 milhões de pessoas. Além disso, os dados foram roubados da polícia e contêm nome, telefone, local de nascimento, número de identificação nacional[5] (número de identidade), ficha criminal, etc, e podem ser uados para para tentar enganar ou mesmo chantagear possíveis vítimas.

Em outras situações, os agentes maliciosos podem utilizar ferramentas chamadas *droppers*, como *TrickBot*, *Zloader* e *BazarLoader* (TRENDMICRO, 2021d; KASPERSKY, 2017). Os *droppers* podem instalar o *malware* diretamente, e, na maioria dos casos, não realizam outra ação maliciosa além dessa. Seu objetivo é instalar o *payload* na máquina da vítima sem ser notado. Diferentemente dos *loaders* (ou *launchers*), que recebem o componente malicioso do servidor C&C, um *dropper* já tem a carga útil embutida em seu código (normalmente possuem um arquivo executável ou uma DLL em sua própria seção de recursos) e quando executado, extrai o *payload* diretamente para a memória da máquina da vítima. Um *dropper* pode conter também instaladores de *malware*.

3.2 Levantamento de Informações

O levantamento de informações é uma fase importante das atividades dos agentes maliciosos, é nesta fase que os operadores buscam informações sobre onde se encontram na rede e que direitos e permissões eles possuem para então realizar a Movimentação Lateral e continuar sua campanha. Existem muitas ferramentas e métodos que podem fornecer essas informações como *Whoami*, *Nltest* e *Net* (utilizadas pelo grupo que opera

[4] https://www.csdn.net/
[5] http://www.gov.cn/banshi/2005-08/02/content_19457.htm

o *Conti*) (TRENDMICRO, 2021c), além de ferramentas como *Sharefinder*, usada para identificar os compartilhamentos necessários para realizar exfiltração de arquivos e implantação do *ransomware* ou *Mimikatz*[6] e *LaZagne*[7], usados para extrair credenciais das máquinas alvo. Já os operadores do *REvil* utilizam outras ferramentas como *AdFind* (também usadas pelos operadores do *LockBit*), *SharpSploit*[8], *BloodHound*[9], and *NBTScan* (TRENDMICRO, 2021d).

Arquivos *batch* também podem ser utilizados para desabilitar ferramentas de segurança, executados através de agendamento de tarefas. Os grupos também são conhecidos por usarem ferramentas de terceiros como *Atera* e *Anydesk* para controlar sistemas remotos, usam o *EternalBlue* para mover lateralmente na rede usando sistemas vulneráveis a esta falha, além de usar o *PsExec* para executar *scripts* remotamente, além do próprio *ransomware*.

3.3 Escalação de Privilégios

Na fase de escalada de privilégios, os agentes maliciosos usam as informações levantadas para conseguir acesso total ao sistema alvo. Para tal, usam alguns *exploits* como *ZeroLogon*, *PrintNightmare* e *Token Impersonation*[10] para fortalecer sua posição na rede. Existem outras ferramentas como *ProcDump*, que despeja os processos do sistema (normalmente *lssas.exe*) para que seja realizada análise, que pode ser combinado com o *Mimikatz* para conseguir as credenciais de acesso. O *lsass* também pode ser explorado com ferramentas nativas do *Windows*, como o próprio gerenciador de tarefas ou com a função *MiniDump* do arquivo DLL *comsvcs*. Outras formas de conseguir acesso a credenciais é utilizando o módulo *kerberoasting*[11] do *PowerShell* ou ferramentas como o *Rubeus* (TRENDMICRO, 2021c).

Para conseguir senhas em texto claro armazenadas nas preferências de política de grupo, os agentes podem usar ferramentas como o *Get-GPPPassword*. Para conseguir credenciais de navegadores e aplicações na nuvem podem usar ferramentas como *SharpChrome* e *SeatBelt*. Depois de conseguir credenciais suficientes, eles usam o *SMBAutoBrute* para automatizar a tarefa de realizar força bruta para descobrir senhas válidas. Com as

[6] https://github.com/ParrotSec/mimikatz
[7] https://github.com/AlessandroZ/LaZagne
[8] https://github.com/cobbr/SharpSploit
[9] https://bloodhound.readthedocs.io/en/latest/index.html
[10] https://capec.mitre.org/data/definitions/633.html
[11] https://attack.mitre.org/techniques/T1558/003/

informações das contas de domínio em mãos, os agentes fazem *dump* de credenciais de controle de domínio usando o *Ntdsutil*. Alternativamente, podem também usar o *Vssadmin* para criar *snapshots* do sistema e baixar *Ntds.dit*[12] para conseguir as credenciais.

3.4 Evasão

Métodos tradicionais de detecção de *malware* basicamente consistem no monitoramento do disco rígido pelo antivírus com o intuito de encontrar algum *malware* através da busca de padrões conhecidos (também chamados de assinaturas) e colocam os arquivos suspeitos em quarentena ou os deletam. Para evitar serem pegos por antivírus e outras ferramentas de detecção, os agentes maliciosos utilizam algumas técnicas para esconder as características maliciosas de seus programas, chamadas técnicas de ofuscação (SIKORSKI; HONIG, 2012, cap 18):

- **Polimorfismo**: O polimorfismo é uma técnica adotada para que o *malware* altere seu próprio código, dificultando sua detecção. Um dos métodos de mutabilidade polimórfica mais comuns é o da encriptação. Também conhecida como empacotamento, se dá através de encriptação/compactação do código binário da parte maliciosa do *malware*, restando intacta apenas a parte referente ao desempacotador (SIKORSKI; HONIG, 2012). Programas legítimos possuem em seu código muitas *strings*, porém os *malware* que foram ofuscados apresentam muito poucas. A encriptação embaralha o código, impedindo que seja estabelecido um padrão (assinatura) que possa ser correlacionado com os padrões conhecidos, devido a utilização de chaves diferentes em cada cópia do mesmo *malware*. O resultado do desempacotamento pode ser descarregado diretamente na memória ou no disco rígido. Quando o desempacotamento é terminado, a execução é então passada para o *malware* para que inicie suas ações no sistema (OR-MEIR et al., 2019). No nível de chamadas de funções da API do *Windows*, pode-se inferir que um software está empacotado se dentre as poucas importações apresentadas estão *LoadLibrary* e *GetProcAddress*. Outro indício importante é que arquivos compactados estão próximos de dados aleatórios e uma maneira eficiente de detectar quando isso acontece é o cálculo da entropia: pacotes não encriptados têm baixa entropia, enquanto os compactados têm entropia mais alta.

[12] https://docs.microsoft.com/pt-br/troubleshoot/windows-server/identity/use-ntdsutil-manage-ad-files

- **Metamorfismo**: para criar uma cópia diferente, os *malware* metamórficos fazem um rearranjo do código binário inserindo instruções supérfluas, alterando os registros utilizados ou mudando a ordem de instruções independentes. Essas ações conseguem mudar a assinatura mantendo as mesmas funcionalidades, dificultando a extração da assinatura para futuras detecções (BROWN; STALLINGS, 2017).

- **Redes Neurais Profundas**: recentemente, a IBM apresentou uma técnica de ofuscação utilizando redes neurais (STOECKLIN; JIYONG JANG, 2018), chamado de *Deeplocker*. A pesquisa foi desenvolvida com o intuito de estudar como modelos existentes de Inteligência Artificial combinados com as técnicas utilizadas em *malwares* poder ser utilizadas para evasão, como reconhecimento facial, de voz e geolocalização. Os pesquisadores utilizaram camadas intermediárias de uma rede neural para esconder as condições de ativação do *malware*. Dada a complexidade das redes neurais, atualmente não é possível realizar engenharia reversa nessas redes (por enumeração exaustiva das condições de ativação, por exemplo), o que faz desta ferramenta uma arma muito útil para ser utilizada em ofuscação. Mais recentemente, existem trabalhos que estudam a utilização de *Adversarial Neural Networks* para simular comportamento benigno e evitar detecção (CHEN, L. et al., 2018).

- **Algoritmo de Geração de Domínios**: muitas assinaturas são baseadas nas *strings* contidas no arquivo binário do *malware*. Essas *strings* podem ser facilmente extraídas utilizando ferramentas como *Strings2*[13] ou uma ferramenta de *debugging*. Dentre as *strings* encontradas em um *malware*, existem as URL que contém os endereços dos servidores de C&C. O algoritmo de geração de domínios gera nomes de domínios mediante requisição, que então podem ser comprados pelo atacante. Os dois benefícios dessa funcionalidade são evitar que os endereços para comunicação com o servidor sejam gravados nas *strings* do arquivo, já que são gerados dinamicamente e o bloqueio prévio de domínios comprometidos.

- ***Process Injection***: esta é uma das técnicas mais conhecidas pelos autores de *malware* para evitar *firewalls* e engenharia reversa executada por analistas pouco experientes, adicionando funcionalidades maliciosas ocultas em processos legítimos. Sua finalidade é injetar código em outro processo em memória ou DLL e executar este código dentro do espaço daquele processo. A partir do *Windows* 7, não é mais permitido injetar código em processos chave como *explorer.exe* ou em processos de outros usuários, porém ainda é permitido injeção na maioria dos navegadores e

[13] https://github.com/glmcdona/strings2

outros processos do próprio usuário. Essa técnica é usada legitimamente por várias aplicações de segurança para monitorar aplicativos, mas também é utilizada com finalidades maliciosas por autores de *malware* (HOSSEINI, 2017; KLEYMENOV; THABET, 2019).

- ***Hook Detection***: se o *malware* detecta um *debugger*, irá permanecer em *loop* infinito sem tomar nenhuma ação enquanto consome recursos da máquina infectada. Alguns *ransomwares*, como o *Maze* ainda procuram alguns processos específicos na máquina, mas ele não tem os nomes listados diretamente em seu código. Para tal, usa *hashcode* para fazer a busca e fechar, principalmente programas de análise como o *Frida* e *Ollydbg* (MCAFEE, 2020; KLEYMENOV; THABET, 2019).

- **Código Lixo**: é um código que não é necessário para o *malware* realizar suas ações maliciosas e tem a finalidade de manter os pesquisadores de antivírus ocupados lendo informações irrelevantes. Instruções de FPU (unidade de ponto flutuante) também são usadas para confundir emuladores de antivírus, dificultando a produção de informações descriptografadas ou legíveis.

- **Ofuscação**: Uma das formas de ofuscação é encriptar parte ou todo o código de um programa. Um ofuscador é uma ferramenta que converte código fonte simples em um programa que faz as mesmas coisas, porém mais difícil de ler de interpretar seu código ao esconder *strings* que poderiam denunciar as ações maliciosas do programa.

- **Empacotamento**: programas empacotados são um subconjunto de programas ofuscados nos quais o programa malicioso é compactado e não pode ser analisado. As técnicas de compactação e ofuscação limitarão as tentativas de analisar estaticamente o *malware*.

3.5 Comunicação com Servidor C&C

A comunicação do *malware* com seu servidor de C&C tem várias funções importantes e muito diferentes entre os tipos de *malware*. Tipicamente os *ransomwares* se comunicam com seus servidores para baixar chaves criptográficas, exfiltrar dados, acessar domínios *kill switch* ou distribuir atualizações.

O *WannaCry* inicia sua execução tentando se conectar a um domínio específico `www.iuqerfsodp9ifjaposdfjhgosurijfaewrwergwea.com`. Se a conexão não for estabelecida,

a exploração continua. No entanto, se a conexão for bem sucedida, o exploração é interrompida imediatamente. Na verdade, esta é uma forma de técnica de **evasão *Sandbox***. Assim que o *WannaCry* tiver certeza de que o domínio não está registrado, a exploração começa (POPLI; GIRDHAR, 2019)

Um artifício engenhoso utilizado por *malwares* para evadirem-se de monitoramento de tráfego da rede é o DGA (*Domain Generetor Algorythm*) (PLOHMANN et al., 2016). Os algoritmos de geração de nomes de domínio são algoritmos vistos em várias famílias de *malware*[14] e são usados para periodicamente gerar uma grande quantidade de nomes de domínio que são usados como ponto de encontro com seus servidores C&C. A grande quantidade de pontos de encontro geradas torna muito difícil autoridades policiais encontrarem os servidores e desligar as *botnets*, pois os computadores infectados tentarão contactar alguns desses domínios para receber atualizações e/ou comandos (YU et al., 2018). Por exemplo, um computador infectado poderia criar milhares de nomes de domínios tentar contactar apenas alguns deles (KÜHRER; ROSSOW; HOLZ, 2014).

A grande vantagem de usar o DGA é não precisar ter a lista de domínios a serem conectados previamente gravada diretamente no código do *malware* (com ou sem ofuscação), impedindo que um analista utilize esse conteúdo e crie uma lista negra de domínios, impedindo novos ataques por restrição de comunicação com seu servidor C&C. O DGA também pode combinar palavras em um dicionário para gerar nomes de domínios. Esses dicionários podem vir embutidos no código do *malware* ou baixados de alguma fonte acessível publicamente. Os domínios gerados pelo DGA com dicionário tendem a ser mais difíceis de detectar devido a sua similaridade com domínios conhecidos (SIDI; NADLER; SHABTAI, 2019). A maneira mais básica de mitigação para este artifício é o bloqueio usando listas negras, porém sua cobertura é bastante limitada ou inconsistente. Algumas abordagens usadas em detecção dependem de técnicas de agrupamento não supervisionadas e informações de contexto como respostas *NXDOMAIN*, *WHOIS* e DNS passivo para avaliar a legitimidade do nome de domínio. Tentativas recentes de detecção do DGA tem tido bastante sucesso através de *Deep Learning*, geralmente usando arquiteturas LSTM e CNN. Embora *deep word embeddings* tenha mostrado boa capacidade para detectar DGA de dicionário, as abordagens utilizando *Deep Learning* se mostraram bastante vulneráveis a técnicas adversariais (CURTIN et al., 2019; PEREIRA et al., 2018).

Assim como a comunicação digital do dia a dia utiliza criptografia assimétrica para evitar que um terceiro intercepte as mensagens trocadas pelo canal, os *malwares* também

[14] https://data.netlab.360.com/dga/

utilizam criptografia para tornar impossível para autoridades policiais imitar os comandos dados pelos controladores do *malware* na medida em que alguns *worms* rejeitam automaticamente atualizações que não foram assinadas pelos operadores. Na mesma linha de tentativa de evadir monitoramento da rede comprometida e evitar detecção, alguns *malwares* utilizam a rede TOR (*The Onion Router*). Em alguns casos, utilizam o serviço *Tor2web*[15], que permite que um usuário se conecte com uma página *onion* com um navegador convencional, servindo como um *gateway* para a rede TOR sobre HTTP ou HTTPS evitando que o *malware* tenha que carregar consigo um cliente TOR completo. Apesar da facilidade de utilizar o serviço *Tor2web*, existem casos que o *ransomware* determina que a vítima baixe e instale o navegador TOR manualmente (SINITSYN, 2014).

Em alguns casos a comunicação se faz necessária para a geração das chaves criptográficas que serão usadas nos arquivos das vítimas, apesar de a maioria dos *ransomwares* adotar a geração de chaves local, evitando trafegar as chaves pela rede e a necessidade de manter servidores ativos (CRACIUN; MOGAGE; SIMION, 2018), que podem chamar atenção de autoridades que podem tomá-los para investigação, além disso, não podem se arriscar a encriptar os arquivos dos usuários sem ter em mãos uma chave correlacionada àquela vítima. Se cada instância do vírus usasse uma chave de criptografia simétrica exclusiva para seu trabalho sujo e se destruísse essa chave depois de usá-la, a recuperação do arquivo seria inviabilizada. O único problema é que o operador deve saber qual é essa chave para liberar os arquivos da vítima. Assim, a máquina da vítima precisa conter algum dado que permita ao operador saber qual chave foi usada naquele caso.

Na tentativa de atrapalhar a detecção, alguns *malware* inserem atraso na comunicação com seu servidor. Por exemplo, um *downloader* pode ter deliberadamente um *script* C&C de resposta lenta, que usa um pequeno tamanho de janela TCP (cinco bytes, por exemplo). Isso atrasa o *download* do arquivo de configuração para que demore dez minutos para ser concluído. Este período de tempo não pode ser reduzido pela VM (como atrasos de *sleeping loop*) e, como resultado, pode gerar um problema de tempo limite na análise automatizada. Essencialmente, isso levará o sistema de análise automatizada a classificar erroneamente a ameaça como um arquivo não malicioso (WUEEST, 2014).

O *malware* pode iniciar a comunicação com o servidor C&C antes de iniciar a cifração dos arquivos ou quando terminar de criptografar. No primeiro caso, o *malware* entra em contato com o operador e recebe uma chave simétrica e um identificador de chave.

Em alguns casos raros, encontramos *malware* que não interrompe a execução do có-

[15] https://www.tor2web.org/

digo em uma VM, mas envia dados falsos. Esses *malware* podem fazer *ping* em servidores de C&C que não existem ou verificar entradas de registro aleatórias. Essas táticas destinam-se a confundir os pesquisadores de segurança ou induzir o processo de automação a acreditar que o *malware* é um aplicativo benigno (WUEEST, 2014).

3.6 Persistência

A persistência ocorre quando um *malware* mantém discretamente o acesso em sistemas comprometidos. Mesmo após interrupções como reinicializações, credenciais alteradas ou *logoff* e *logon*, o *malware* é acionado novamente para execução. Esse *malware* geralmente está oculto em pastas de inicialização legítimas ou em tarefas e serviços agendados, dificultando sua localização. A persistência é alcançada primordialmente utilizando-se tês artifícios: criação de *mutex* (*mutual exclusion*), criando entradas específicas no registro do *Windows* ou configurando *scripts* para que sejam executados automaticamente na inicialização. Por exemplo, o *Maze* cria um *mutex* com o nome *Global\x*, onde x é um valor único por máquina, com a finalidade de evitar duas ou mais execuções ao mesmo tempo (MCAFEE, 2020). Este *mutex* é verificado através de chamadas à função *GetLastError*, com a finalidade de fazer as seguintes verificações:

- 0x05 (ERROR_ACCESS_DENIED): se o *malware* receber este erro, significa que o *mutex* já existe no sistema, porém, por alguma razão o *malware* não pode acessá-lo (talvez por alguma política ou privilégio).
- 0xb7 (ERROR_ALREADY_EXISTS): se o *malware* receber este erro, significa que o *mutex* já existe no sistema e pode ser acessado.

Se qualquer das situações acima ocorrer, o *malware* permanece em execução porém não encripta nenhum arquivo nem usa recursos da máquina, o que significa que o processo aparece na lista com 0% do processador.

A Execução Automática em *Logon* ou *Boot* envolve violação de um processo legítimo do sistema operacional por um agente malicioso, por exemplo, uma reinicialização do sistema ou *logon*, obtendo persistência adicionando uma entrada às chaves de execução no Registro do *Windows* ou na pasta de inicialização. Como resultado, quaisquer programas referenciados serão executados quando um usuário efetuar *login*. Ao registrar *Scripts* de inicialização de logon ou *Boot*, agentes maliciosos tipicamente usam credenciais locais ou conta de administrador para rodar *scripts* que executem automaticamente em *boot*

ou em *login* para estabelecer persistência. Por sua vez, os invasores podem executar outros programas ou enviar informações para um servidor de *log* interno. É possível reduzir as chances de ser afetado por esse mecanismo de persistência se for garantido que as permissões adequadas sejam definidas e que o acesso de gravação a *scripts* de logon de administradores específicos sejam restritos. No entanto, esta não é uma medida de prevenção infalível (HAMMOND, 2021).

Abaixo, temos as duas chaves de registro mais comuns encontradas em *malware* para conseguir persistência (POPLI; GIRDHAR, 2019; HUNTING..., 2021):

- HKCU\SOFTWARE\Microsoft\Windows\CurrentVersion\Run\<*Random*> Valor: <*Fullpath*>\tasksche.exe. CU é a sigla de *Current User* e *Random* é um nome pseudo aleatório derivado do nome do computador. Esta é uma entrada com privilégios de usuário.
- HKLM\SOFTWARE\Microsoft\Windows\CurrentVersion\Run\<*Random*> Valor: <*Fullpath*>\tasksche.exe. LM é a sigla de *Local Machine* e *Fullpath* é o caminho completo para o arquivo. Esta é uma entrada com privilégios de sistema.

Ao registrar tarefas ou serviços agendados para manter persistência, um invasor subverte o recurso de agendamento de tarefas para realizar a execução inicial ou recorrente de código malicioso. Uma possibilidade é usar o Agendador de Tarefas do *Windows*, que pode ser usado para executar programas na inicialização do sistema ou de forma programada. Como exemplo, o *TrickBot*, um programa de *spyware trojan*, é conhecido por criar tarefas agendadas em sistemas comprometidos de uma forma que fornece persistência para o ataque. Como todos os principais sistemas operacionais apresentam utilitários para agendar programas ou *scripts* a serem executados, esse mecanismo de persistência é um risco para quase todos. A chave para detectar esse mecanismo comum de persistência de *malware* é revisar regularmente seu agendador de tarefas para eliminar quaisquer alterações nas tarefas que não se correlacionam com *software* conhecido, ciclos de *patch* e assim por diante. É importante notar que somente detectar o *malware* pode não resolver o problema da vítima, pois a detecção é apenas uma solução temporária e não resolve o problema maior: persistência. Se os gatilhos utilizados para persistência não forem mitigados no ambiente, os agentes maliciosos podem simplesmente recircular o *malware*, causando reinfecção. É por isso que é fundamental que a configuração da persistência seja encontrada e eliminada.

3.7 Criptografia

A criptografia é a ciência da escrita secreta com o objetivo de esconder o significado de uma mensagem (PAAR; PELZL, 2009). O primeiro registro de sua utilização data de 2000 AC, no Egito. Desde aquela época, a criptografia tem sido usada em várias civilizações, como os gregos e os romanos (a mais conhecida é a Cifra de César) (STALLINGS; BRESSAN; BARBOSA, 2008).

Trazendo a definição para a realidade tecnológica que temos atualmente, a escrita secreta passa a ser aplicada a dados, que uma vez criptografados, somente poderão ser processados após serem descriptografados (STALLINGS; BRESSAN; BARBOSA, 2008). Esta situação é que confere a segurança ao processo: um agente malicioso pode conseguir interceptar a comunicação, pode conhecer o algoritmo criptográfico utilizado, mas não será possível ler e interpretar os dados a menos que tenha a chave utilizada. Praticamente todos os dispositivos computacionais modernos utilizam algum tipo de criptografia, seja internamente para proteger dados sensíveis ou externamente, em sua comunicação com a internet.

As chaves criptográficas são valores matemáticos (números com vários bits) que controlam a operação de um algoritmo de criptografia, especificando a transformação de texto puro em texto cifrado, ou vice e versa. Para que possam desempenhar seu papel de forma ótima. Idealmente as chaves devem ser geradas a partir de um Gerador de Números Pseudoaleatórios Criptograficamente Seguro (CSPRNG) que lhes forneça entropia suficiente para frustrar tentativas de adivinhação e não permita vazamento dos dados (PARISOT; BENTO; MACHADO, 2021). Quase todas as cifras de fluxo modernas têm dois parâmetros de entrada: uma chave e um vetor de inicialização (IV). A primeira é a chave regular que é usada em todos os sistemas de criptografia simétricos. O IV serve como um randomizador e deve assumir um novo valor para cada sessão de criptografia. É importante notar que o IV não precisa ser mantido em segredo, apenas deve ser alterado a cada sessão. (PAAR; PELZL, 2009)

Para que tenham sucesso em sua empreitada, os *ransomwares* devem procurar seguir as mesmas diretrizes criptográficas de protocolos de segurança, para que da mesma forma que um invasor tenha dificuldade de burlar o esquema de segurança implementado, um pesquisador tenha dificuldade ao analisar um *ransomware* com a finalidade de desvendar seu trabalho criptográfico. Estas diretrizes podem ser traduzidas em três propriedades (LE GUERNIC; LEGAY, 2017):

- **Propriedade 1**: O código binário malicioso não deve conter nenhum segredo (por exemplo, chaves de decriptação). Pelo menos não de uma forma facilmente recuperável.
- **Propriedade 2**: Somente o autor do ataque deve ser capaz de descriptografar o dispositivo infectado.
- **Propriedade 3**: Descriptografar um dispositivo não pode fornecer nenhuma informação útil para outros dispositivos infectados, em particular a chave não deve ser compartilhada entre eles.

Tanto o Gerador de Números Pseudo Aleatórios (PRNG) quanto o algoritmo criptográfico podem ser implementados diretamente no código do *malware* ou utilizar a API criptográfica disponível no *Windows*. A *CryptoAPI* é um conjunto de bibliotecas vinculadas dinamicamente que fornecem uma camada de abstração que isola os programadores do código usado para criptografar os dados e suporta tanto criptografia simétrica quanto assimétrica e inclui funcionalidades de criptografar e descriptografar dados (*CryptEncrypt* e *CryptDecrypt*, respectivamente), geração de números pseudoaleatórios (*CryptGenRandom*), criação de *hashes* (*CryptCreateHash*) dentre outras primitivas criptográficas (MICROSOFT, 2022). O *ransomware* pode também ter implementado um código criptográfico personalizado. Considerando que um bom código criptográfico é o resultado de anos de dedicação de pesquisa e muitos testes (DWORKIN et al., 2001), um algoritmo criptográfico, implementado por desenvolvedores sem experiência, está altamente sujeito a erros (não cumprimento das 3 propriedades mencionadas) e pode causar falhas que permitam a vítima recuperar seus dados.

Para evitar esses problemas, os autores de *ransomware* procuram usar esquemas simples de cifração pela sua facilidade de implementação, que em muitos casos são suficientes, porém mesmo a simplicidade apresenta alguns inconvenientes (SIKORSKI; HONIG, 2012; ORMAN, 2016):

- Bibliotecas criptográficas podem ocupar espaço, então o *malware* deve integrar estaticamente ou *linkar* o código malicioso.
- *Linkar* o código que já existe no hospedeiro pode reduzir portabilidade.
- Bibliotecas criptográficas padrão são facilmente detectáveis (via importação de funções ou identificação de constantes criptográficas).

- A maioria dos *ransomwares* troca alguma segurança por desempenho, e isso lhes dá capacidade de criptografar mais dados de arquivo do usuário antes de ser detectado.
- Usuários de criptografia simétrica têm que se preocupar em como esconder a chave.
- Muitos algoritmos criptográficos padronizados dependem de uma chave forte para armazenar segredos.

A ideia geral é que o algoritmo de criptografia em si é bastante conhecido, mas sem a chave é virtualmente impossível desencriptar o texto cifrado, além disso, o tamanho da chave influencia no trabalho necessário para descobri-la, porém as chaves podem ser inadvertidamente expostas no *software*, as chaves públicas podem ter muitos bits, mas serem mal escolhidas, as codificações podem vazar dados, os geradores de chaves podem estar com defeito ou os servidores C&C podem ser hackeados.

Ao olharmos para os *ransomwares* à luz dessas premissas, vemos que muitos deles usam algoritmos padrão tanto para encriptar dados, quanto para *hashing* (AES, RSA, *Blowfish*, RC4). Outra situação interessante é que mais de 80% preferem bibliotecas padrão, como API do *Windows* ou *OpenSSL* à implementações personalizadas do *Salsa*, *Chacha* ou AES-512 (CRACIUN; MOGAGE; SIMION, 2018).

Na contramão do que foi exposto acima, alguns *ransomwares* possuem capacidades avançadas de encriptação. Dentre eles podemos mencionar o *Cryakl*, que apresenta procedimentos diferenciados para lidar com os formatos ZIP, 7z, TAR, CAB e RAR. Ele analisa cada um desses formatos e criptografa apenas as partes críticas do arquivo, oferecendo alto desempenho e evitando a recuperação de dados sem a chave (SINITSYN; ZINCHENKO, 2021). Outro exemplo é o *LockBit*, que em um teste realizado com outros *ransomwares*, incluindo *Ryuk*, *REvil* e *Conti* (TULAS, 2022), foi o que apresentou encriptação de arquivos mais rápida, com um desempenho de cifração de aproximadamente 25 mil arquivos por minuto. Essa façanha é conseguida pois o *LockBit* encripta somente os primeiros 4KB dos arquivos (TRENDMICRO, 2022b). O *CryptConsole* tem um esquema de encriptação ainda mais elaborado, pois inicia a encriptação inicial utilizando o AES, onde inicialmente, uma parte do arquivo é encriptada usando a chave e o IV, depois o *buffer* encriptado é revertido e encriptado novamente, desta vez utilizando-se uma segunda chave e um segundo IV, formando assim um esquema de dupla encriptação (SINITSYN; ZINCHENKO, 2021).

Nas próximas seções serão explorados dois assuntos dentro da criptografia, as cifras simétricas e as assimétricas e como os *ransomwares* as utilizam para sequestrar os arquivos

das vítimas.

3.7.1 Simétrica

A criptografia simétrica é o que a maioria das pessoas pensa como funciona a criptografia: dois agentes tem um mesmo método de encriptação e desencriptação para o qual eles compartilham uma chave secreta (PAAR; PELZL, 2009). Toda a criptografia até 1976 era exclusivamente baseada neste processo, que funciona com técnicas de substituição, transposição e embaralhamento (STALLINGS; BRESSAN; BARBOSA, 2008). Cifras simétricas ainda estão em uso atualmente e fazem parte do arsenal de ferramentas que os *ransomwares* possuem para chantagear suas vítimas.

Pelo fato de que os criadores de *ransomwares* normalmente melhoram suas habilidades com o tempo, esses *malwares* transformaram-se em versões melhoradas. Essas atualizações vão desde o nível de ofuscação até o nível de design de criptografia ou disseminação (CRACIUN; MOGAGE; SIMION, 2018), buscando sempre dificultar agentes de cibersegurança e pesquisadores a conseguirem quebrar a criptografia utilizada, mantendo assim a vantagem competitiva.

Em alguns casos, os *ransomwares* utilizam somente criptografia simétrica, em que algoritmos como o AES são utilizados. Sua boa velocidade de encriptação é uma vantagem, porém usar somente este tipo de criptografia de dados pode permitir ao usuário recuperar as chaves e por consequência os arquivos, pois as chaves usadas podem ficar gravadas em um arquivo local sem encriptação (MARINHO, 2018).

3.7.2 Assimétrica

Algoritmos de criptografia com chaves assimétricas (também chamado de chave pública) foram criados em 1976 por *Whitfield Diffie*, *Martin Hellman* e *Ralph Merkle* (DIFFIE; HELLMAN, 2019). Neste tipo de criptografia, baseado em funções matemáticas, ao invés de substituições e permutações, o usuário possui não somente uma chave secreta (como na criptografia simétrica) como também uma chave pública (STALLINGS; BRESSAN; BARBOSA, 2008). A utilização de de duas chaves tem grandes desdobramentos no que tange a confidencialidade, distribuição de chaves e autenticação, por exemplo, algoritmos de chaves assimétricas têm outras aplicações além de cifrar dados, como assinaturas digitais e combinação de chaves (PAAR; PELZL, 2009).

Como vantagem deste tipo de implementação, está a segurança conferida pela invia-

bilidade computacional de se calcular a chave privada a partir da chave pública. A grande desvantagem da criptografia assimétrica em relação a simétrica é a velocidade de encriptação, e este é o principal motivo de mesmo nos dias de hoje, a criptografia simétrica não ter sido totalmente substituída pela assimétrica. Além disso, muitos *ransomwares* utilizam as duas formas de encriptação em conjunto para sequestrar os dados das vítimas.

A implementação mais simples encontrada em *ransomwares* é a encriptação assimétrica com chave fornecida por servidor. Neste esquema, o servidor gera um par de chaves, a chave pública é gravada no código do *ransomware* e será usada para encriptar cada arquivo da vítima (MARINHO, 2018). A não ser que o direcionamento de alvos seja tal que cada vítima seja atacada por um executável com chaves diferentes, é possível que apenas uma vítima pague o valor exigido pelo sequestro dos dados e compartilhe a chave recebida com as outras, que poderão recuperar seus arquivos. Caso o *ransomware* não tenha a chave pública em seu código, ele terá que baixar uma chave gerada no servidor C&C, porém esta técnica se torna impraticável no caso de a vítima não ter acesso a internet. De toda maneira, esta prática é incomum, pois a velocidade sensivelmente menor de encriptação com chaves assimétricas dá a vítima maior tempo de reação ao incidente. O usuário pode detectar essa infecção antes que muitos arquivos sejam afetados. No entanto, os métodos de criptografia de chave pública não fornecerão informações úteis sobre como descriptografar os arquivos. Apenas a chave privada correspondente, mantida pelo operador, pode desfazer o dano.

3.7.3 Mista

Alguns *ransomwares* utilizam criptografia assimétrica e simétrica em conjunto para sequestrar os dados das vítimas, com o intuito de utilizar do melhor de cada uma para alcançar uma abordagem mais eficiente (simétrica - velocidade, assimétrica - segurança), certificando-se de que a vítima não consiga recuperar seus dados a menos que lhes seja fornecida a chave privada.

Apresentamos a seguir os esquemas de encriptação mais utilizados por *ransomwares* (PLOSZEK; ŠVEC; DEBNÁR, 2021; MARINHO, 2018) que fazem uso de chaves simétricas e assimétricas em conjunto:

- **Esquema 1:** o atacante gera um par de chaves para uma vítima específica e insere no código do *ransomware* a chave pública. Quando esta amostra é executada pela vítima, uma chave simétrica é gerada para cada arquivo que será encriptado.

Esta chave simétrica é depois encriptada com a chave pública que está gravada no *ransomware*.

- **Esquema 2:** a amostra maliciosa é distribuída com uma chave simétrica. Comparando com o esquema anterior, o par de chaves assimétricas é gerado diretamente na máquina da vítima. A chave privada gerada é encriptada com a chave simétrica original, enviada ao operador e então é apagada da memória. O processo de encriptação dos arquivos ocorre da mesma forma que o esquema 1.

- **Esquema 3:** é muito parecido com o primeiro esquema, porém é gerada uma chave simétrica única para cada vítima e esta mesma chave é usada para cifrar todos os arquivos.

- **Esquema 4:** o programa é compilado com uma chave pública. Ao ser executado, gera na vítima um par de chaves e a chave privada desse par gerado é encriptada com a chave pública do atacante. O processo de encriptação é o mesmo de 1 e 2, com uma chave simétrica utilizada para cifrar cada arquivo. Este esquema é utilizado pelo *wannacry* e pode ser considerado o esquema mais seguro encontrado em um *malware* deste tipo. A grande vantagem deste esquema é que as vítimas de um mesmo binário do *malware* não conseguem compartilhar suas chaves privadas, pois existe uma outra camada de encriptação assimétrica.

3.8 Prevenção de Recuperação

A eficácia de um ataque de *ransomware* depende, dentre outras situações, de que a vítima seja impedida de recuperar seu sistema para uma versão anterior. No *Windows*, esta possibilidade se dá a partir da deleção das *volume shadow copies*, utilizando ferramentas presentes no próprio SO, como *vssadmin.exe* e *WMIC.exe*.

Alguns *ransomwares*, como o *Maze*, para se certificarem de que a vítima não conseguirá recuperar o sistema, tenta excluir as cópias duas vezes, uma antes de encriptar os arquivos e outra após enviar cópias dos arquivos para o servidor C&C. Esta execução é feita através da função *CreateProcessW* (MCAFEE, 2020).

3.9 Propagação

A diferença principal entre a capacidade de propagação e infecção inicial é que a propagação é feita automaticamente pelo *malware* depois que já está inserido na rede da vítima (com pelo menos uma máquina infectada). Conforme mencionado anteriormente, alguns *ransomwares* (como *WannaCry*, *Petya* e o *SamSam*) tem esta capacidade. Para tal, fazem varreduras nas redes locais e verificam se as máquinas encontradas possuem alguma vulnerabilidade que ele tenha capacidade de atacar, como *EternalBlue*. Alguns *ransomwares* mais agressivos chegam a fazer varredura em IPs aleatórios na internet e verificam vulnerabilidades conhecidas nessas máquinas.

3.10 Pagamento do Resgate

O termo ***Big-game Hunting*** originalmente refere-se à caça de grandes animais para extração carne, chifres, peles, presas, ossos, gordura, taxidermia, ou simplesmente caça esportiva. O termo é frequentemente associado à caça dos animais *Big Five* da África (leão, elefante africano, búfalo do Cabo, leopardo africano e rinoceronte), além de tigres e rinocerontes no subcontinente indiano. O termo foi portado para a cibersegurança como um tipo de ataque cibernético que geralmente emprega o *ransomware* para atingir organizações grandes e de alto valor ou entidades de alto perfil. As vítimas são escolhidas com base em sua capacidade financeira de pagar um resgate, bem como na probabilidade de fazê-lo para retomar as operações comerciais e/ou evitar má publicidade.

Em 2020, vimos o surgimento de vários grupos *high-profile* no mundo do *ransomwares*. Os criminosos descobriram que as vítimas seriam mais propensas a pagar resgates se pudessem estabelecer de antemão alguma forma de reputação, para garantir que sua capacidade de restaurar arquivos criptografados nunca fosse questionada, cultivando presença online, escrevendo comunicados à imprensa e garantindo que seu nome fosse conhecido por todas as vítimas em potencial (GALOV; BEZVERSHENKO; KWIATKOWSKI, 2021).

Um *ransomware*, ao terminar de cifrar os arquivos da vítima, avisa sobre os danos causados em uma *ransom note*. Em algumas dessas notas, são exibidas apenas um e-mail de contato, enquanto em outros *ransomwares* mais sofisticados exibem links *onion*, que automatizam o pagamento e ajudam as vítimas com o processo de desencriptação (CRACIUN; MOGAGE; SIMION, 2018). Os ataques, no entanto, são 99% das vezes dirigidos a computadores que pertencem a empresas, pois estas tem a maior probabilidade de pagar

o resgate dos dados (CRACIUN; MOGAGE; SIMION, 2018).

Existe uma questão ética/prática que uma vítima de *ransomware* precisa lidar, que é a questão de pagar ou não o resgate (*ransom*). Apesar de parecer conveniente pagar o resgate e voltar aos negócios como de costume, a situação não é tão simples. Não é uma boa ideia pagar o resgate a menos que ofereça risco à vida humana, à segurança pública ou ameaça à sobrevivência da empresa (CYBEREASON, 2021). Em uma pesquisa realizada pelo *Cybereason* (TRAFIMCHUK; BUKHTEYEV; LADUTSKA, 2021), constatou-se que pagar o resgate não garante a recuperação dos dados, pois quase metade das organizações não conseguem recuperar os dados mesmo após pagamento e que 80% das empresas que admitiram ter pago o resgate, sofreram novos ataques desferidos pelos mesmos operadores. Um exemplo que pode ilustrar a situação descrita é uma falha na cifração do *Lockbit* 2.0 (VELUZ, 2022a), que por erro no tratamento os valores de retorno de uma da API de gravação de arquivo, causa um embaralhamento no arquivo cifrado, impedindo a vítima de recuperar seus dados mesmo depois de pagar o resgate (VELUZ, 2022b).

De acordo com estatísticas levantadas pela *Kaspersky*, as vítimas de *ransomware* raramente conseguem recuperar totalmente os dados depois de pagarem o resgate (GATELY, 2021). Esta situação sugere duas nuances: a primeira é que a vítima fica pressionada a pagar o resgate e ter seus arquivos de volta e a segunda é que o pagamento incentiva a ação maliciosa em prol de mais ataques. Além disso, os agentes maliciosos introduziram uma nova modalidade de chantagem, chamada de *double extortion*. Nessa abordagem, além de cifrar os dados da vítima, os agentes maliciosos copiam esses dados para seus servidores e ameaçam divulgá-los, o que pressiona ainda mais a vítima no sentido pagar o resgate.

O objetivo final de um *ransomware* é ganhar dinheiro. Inicialmente eram utilizados transferências bancárias e cartões pré pagos (cartões de presente *MoneyPak*, *Amazon* e *Apple*) ou SMS. Esses métodos são tranquilamente rastreados por agências policiais e nesse sentido, operações maliciosas de larga escala ficam limitadas para que não chamem atenção e é por este motivo que as novas gerações de *ransomware* utilizam *Bitcoin*. Seu lançamento acabou estimulando as operações deste tipo de *malware* graças a sua confidencialidade, velocidade de transferência e ausência de bancos centrais controlando a moeda. Em 2014, seu algoritmo foi estendido e comporta 80 bytes não relacionados com a transação. Uma variante do *CTB-locker* usa este campo como canal seguro e envia a chave de decriptação uma vez que a vítima tenha pago o resgate (LE GUERNIC; LEGAY, 2017).

Alguns governos estão empenhados na caça aos grupos de *ransomware*. O governo

dos EUA, por exemplo, está impondo sanções severas a corretoras de criptomoedas que viabilizam o pagamento de grupos de *ransomware* (EUA. . . , 2021; SATTER, 2020). Além disso, tem perseguido e sancionado os desenvolvedores do *Tornado Cash* (BURT, 2022), uma ferramenta de código aberto que efetua *mixing* de criptomoedas, impedindo agentes governamentais de rastrear o caminho seguido por determinado *bitcoin* (DE, 2022; REGUERA, 2022). O mundo do *ransomware* deve ser entendido como um ecossistema e tratado como tal: é um problema que só pode ser resolvido sistematicamente, por exemplo, impedindo que o dinheiro circule dentro dele – o que envolve, em primeiro lugar, que as vítimas não paguem resgate.

3.11 Ransomware as a Service (RaaS)

O *Ransomware as a Service* (RaaS) (TEAM, 2021b) é um modelo de negócio que tem sido adotado pelos grupos maliciosos. Neste modelo, os programadores do *malware* não se expõem realizando ataques diretamente, mas cedem seus *ransomwares* e a infraestrutura de suporte, como portais de contato e pagamento, a terceiros (chamados afiliados), que atacam as companhias e pagam a ferramenta com parte do lucro obtido. Outros grupos levam essa compartimentação de atividades mais a fundo, criando atividades de seleção e recrutando os vencedores para participarem dos grupos, como *hackers* que realizam a infiltração no sistema da vítima, dividindo as tarefas entre muitos participantes de acordo com suas especialidades: há os responsáveis pelo desenvolvimento, os administradores das *botnets*, que automatizam o processo de infecção, há os que vendem os acessos às redes corporativas e os que efetivamente operam o *ransomware* (TEAM, 2021a). . Dessa maneira, os agentes maliciosos conseguem maior eficácia em seus ataques, o que resulta em maiores ganhos. Devido a estes fatores o *RaaS* foi adotado pela maioria dos grupos em atividade atualmente.

Nos últimos anos, ocorreram algumas mudanças no paradigma de ataque e infecção dos grupos que se utilizam desse tipo de artefato. Um direcionamento mais profissional tem sido dado aos ataques de *ransomware*. De acordo com um levantamento realizado pela *Kaspersky*, os ataques direcionados aumentaram 700% (TEAM, 2021a). Como exemplo, podemos citar a mudança de ataques massivos para alvos específicos (grandes empresas) (REUTERS, 2021a), ativos de infraestruturas críticas, como portos (ADVISOR, 2021), gasodutos (REUTERS, 2021b) e organizações de saúde.

Alguns grupos, além de cobrarem o valor das suas vítimas para as chaves criptográficas

para desencriptar seus arquivos, exigem também valores para que não exponham os dados roubados, uma vez que uma das possibilidades dos *ransomwares* é permitir que os agentes maliciosos façam cópias dos dados das vítimas, o que é particularmente prejudicial para grandes empresas, principalmente as que possuem dados sensíveis de seus clientes como hospitais ou planos de saúde.

Contra-intuitivamente, as pessoas que obtêm o acesso inicial à rede da vítima não são as que implantam o *ransomware* posteriormente, e é útil pensar na coleta de acesso como um negócio totalmente separado. Para que seja viável, os vendedores precisam de um fluxo constante de "produto". Pode não fazer sentido financeiro passar semanas tentando violar um alvo difícil predeterminado, como uma empresa da Fortune 500, porque não há garantia de sucesso. Em vez disso, os vendedores de acesso vão atrás de vítimas mais vulneráveis. Existem duas fontes principais para tal acesso (GALOV; BEZVERSHENKO; KWIATKOWSKI, 2021):

Botnet owners. operadores de *malware* estão envolvidos nas maiores e mais abrangentes campanhas. Seu principal objetivo é criar redes de computadores infectados, embora a infecção esteja apenas latente neste momento. Os proprietários de *botnets* (os *botmasters*) vendem o acesso às máquinas das vítimas em massa, como um recurso que pode ser monetizado de várias maneiras, como organizar ataques DDoS, distribuir *spam* ou, no caso de *ransomware*, pegar carona nessa infecção inicial para se firmar em um alvo potencial.

Access sellers. *Hackers* que estão à procura de vulnerabilidades divulgadas publicamente em software voltado para a Internet, como dispositivos VPN ou *gateways* de e-mail. Assim que tal vulnerabilidade é divulgada, eles comprometem o maior número possível de servidores afetados antes que os defensores apliquem as atualizações correspondentes.

Em ambos os casos, é somente após o acesso inicial que os invasores dão um passo para trás e descobrem quem eles violaram e se essa infecção provavelmente levará ao pagamento de um resgate. Os atores no ecossistema de *ransomware* não fazem direcionamento, pois quase nunca optam por ir atrás de entidades específicas. Compreender esse fato destaca a importância de as empresas atualizarem os serviços voltados para a internet em tempo hábil e terem a capacidade de detectar infecções latentes antes que possam ser aproveitadas. Embora o número e a variedade de ofertas disponíveis na *darknet* certamente não sejam pequenos, os mercados não refletem todo o ecossistema de *ransomware*. Alguns grandes grupos de *ransomware* trabalham de forma independente ou encontram parceiros diretamente (por exemplo, o *Ryuk* conseguiu acessar alguns dos sistemas de suas vítimas

após uma infecção por *Trickbot*, o que sugere uma potencial parceria entre dois grupos). Portanto, os fóruns geralmente hospedam *players* menores – operadores de RaaS de médio porte, atores menores que vendem código-fonte e novatos.

Como exemplo da tentativa de aliciamento de funcionários de uma empresa vítima do *LockBit*, temos a Figura 3. Nela podemos ver o texto contido em um binário do *LockBit*, que contém uma mensagem que pode ser exibida a algum funcionário de uma empresa na tentativa de aliciá-lo a participar do programa e contribuir para infecção da própria empresa, fornecendo credenciais de acesso e em contrapartida receber algum dinheiro por isso. Este trecho foi encontrado dentro da seção Strings da amostra com *hash* SHA256 de valor *0dcbc979c995eeec6e76fa87dfc301228d230101b9acecba53289c1c085eadb1*, logo após a mensagem de ransom:

Figura 3: Imagens da seção de strings de um arquivo binário do LockBit. O texto está localizado logo após a nota de ransom

```
4873575         "Software\\Microsoft\\Windows NT\\CurrentVersion\\ICM\\Calibration",
4873576         "{D2E7041B-2927-42fb-8E9F-7CE93B6DC937}",
4873577         "Proxima Nova",
4873578         "All your files stolen and encrypted",
4873579         "for more information see",
4873580         "RESTORE-MY-FILES.TXT",
4873581         "that is located in every encrypted folder.",
4873582         "Would you like to earn millions of dollars?",
4873583         "Our company acquire access to networks of various companies, as well as insider information that can help you steal the most valuable data of any company.",
4873584         "You can provide us accounting data for the access to any company, for example, login and password to RDP, VPN, corporate email, etc. Open our letter at your
    ↪ email. Launch the provided virus on any computer in your company.",
4873585         "Companies pay us the foreclosure for the decryption of files and prevention of data leak.",
4873586         "You can communicate with us through the Tox messenger",
4873587         "https://tox.chat/download.html",
4873588         "Using Tox messenger, we will never know your real name, it means your privacy is guaranteed.",
4873589         "If you want to contact us, use ToxID: 3085B89A0C515D2FB124D645906F5D3DA5CB97CEBEA975959AE4F95302A04E1D709C3C4AE9B7",
4873590         "If this contact is expired, and we do not respond you, look for the relevant contact data on our website via Tor or Brave Browser",
4873591         "http://lockbitapt6vx57t3eeqjofwgcglmutr3a35nygvokja5uuccip4ykyd.onion",
4873592         "https://bigblog.at",
4873593         "%s.bmp",
```

3.12 Famílias mais Ativas (até 2022)

O conceito de família de *ransomware* está diretamente ligado ao RaaS. O que acontece é que os criadores de *ransomwares* preparam ferramentas de personalização que geram o *ransomware* de acordo com necessidades específicas do utilizador, como nome do operador, da vítima, meio de pagamento (carteira de *bitcoin* e o valor a ser pago), imagem a ser usada como papel de parede, imagem para ícone, chave de encriptação etc. Devido a esse fato, a maioria dos *malwares* descobertos nos dias de hoje são códigos reutilizados, utilizam funcionalidades conhecidas e não são escritos a partir do zero, o que também indica que os criadores de *malware* estão trabalhando para aperfeiçoar os códigos já existentes e reutilizam funcionalidades já consagradas (CALDAS, 2016). Nesse sentido, listamos a seguir algumas características de cada família de *ransomware* considerada neste trabalho. Mais adiante, na Seção 3.13, veremos um caso prático dessa situação ocorrendo entre versões diferentes de *ransomwares* de uma mesma família. A seleção das famílias a serem

analisadas foi feita com base em um relatório de *malware* publicado pela IBM (IBMSECURITY, 2022) com dados de ataques e estatísticas referentes ao ano de 2021 (inclusive).

3.12.1 *Conti*

O grupo teve suas atividades iniciadas em 2018, sob o nome *Ryuk*. O grupo é bem articulado e usa uma boa variedade de ferramentas, incluindo *malware* customizado e um modelo de intrusão em vários estágios (ESENTIRE, 2022), onde cada fase, como acesso inicial, movimento lateral e extorsão são executados por especialistas de cada área, o que confere bastante efetividade às atividades de ataque às vítimas. Inicialmente o *Trickbot* era utilizado como *backdoor*, porém recentemente passou a usar o *BazarLoader*. Utilizam também serviços de busca pública como *Shodan* e ferramentas de evasão como *Shelter Project*. Depois de contaminar a vítima, encripta seus arquivos utilizando o AES-256 (CISA, 2022; TULAS, 2022). Em outras versões, depois de realizar a exfiltração dos dados da vítima e distribuir o *ransomware* para as outras máquinas na rede, os arquivos são encriptados com *ChaCha20* e RSA 4096 para proteger as chaves e o *nonce* gerado pelo *ChaCha*.

Com ataques bem sucedidos a infraestruturas críticas como escolas, governos locais, redes de saúde e companhias de energia, coleciona uma boa lista de vítimas (ESENTIRE, 2022), das quais podemos citar *Aareon*, *Bank Indonesia*, *SEA-Invest*, *CS Energy*, *iTCo* e *Mabanaft Deutschland GmbH*, considerada a maior empresa de importação e venda de gás e petróleo na Alemanha e a *Panasonic* (GLOVER, 2022a).

3.12.2 *Ryuk*

Uma variante do antigo *ransomware Hermes*, *Ryuk* é um *ransomware* atribuído ao grupo *Wizard Spyder*. Quando infecta um sistema, primeiro fecha 180 serviços e 40 processos, de modo que não haja empecilhos ao seu funcionamento e, a partir deste ponto, começa a criptografar os dados (TRENDMICRO, 2021a).

Dentre as vítimas do *Ryuk* estão órgãos governamentais, companhias do ramo de academia, saúde, manufatura e organizações de tecnologia. Algumas delas são *Tribune Publishing*, gráfica responsável pelos impressos do *The New York Times* e do *Wall Street Journal*, *Universal Health Services* (UHS), uma empresa de saúde com hospitais nos EUA e no Reino Unido, o *Sky Lakes Medical Center* e o *Lawrence Health System*. Em 2019 teve a maior demanda de resgate, 12,5 milhões de dólares (TRENDMICRO, 2021a).

O que faz do *Ryuk* particularmente perigoso é sua capacidade de movimentar-se lateralmente no sistema. Usando tanto ferramentas maliciosas e vulnerabilidades como o *EternalBlue* e *Zerologon* para se propagar na rede. isto significa que ao invés de ter que infectar as outras máquinas individualmente, o *Ryuk* simplesmente precisa entrar na infraestrutura de TI para infectar várias máquinas.

3.12.3 *Revil*

Pesquisadores identificaram muitas semelhanças e reutilização de código entre o *REvil* e o *GrandCrab*. Tipicamente a infecção decorre de campanhas de *phishing*, ataques de força bruta ao RDP ou vulnerabilidades de software (REUTERS, 2021a).

Devido à sua natureza direcionada, o *REvil* usa uma variedade de ferramentas e *malwares*, dependendo da situação. Seus operadores parecem operar com alto nível de conhecimento sobre o ambiente de suas vítimas, como evidenciado pelo nível de personalização dos seus ataques. O *REvil* possui um bloco de configuração criptografado com muitos campos, que permitem que os invasores ajustem a carga útil (TRENDMICRO, 2021d). O executável pode encerrar processos na lista negra antes da criptografia, exfiltrar informações básicas da vítima, criptografar arquivos e pastas não incluídos na lista de permissões em dispositivos de armazenamento local e compartilhamentos de rede. Os afiliados são responsáveis por obter acesso inicial às redes corporativas e implantar o *locker* – uma prática padrão para o modelo RaaS. Deve-se notar que a gangue tem regras de recrutamento muito rígidas para novos afiliados: o *REvil* recruta apenas parceiros altamente qualificados, nativos de países de língua russa e com experiência em obter acesso a redes. (GALOV; BEZVERSHENKO; KWIATKOWSKI, 2021).

As vítimas desta campanha incluem empresas como *Travelex*, *Brown-Forman Corp.*, o grupo farmacêutico *Pierre Fabre*, JBS, *Harris Federation*, *Quanta Computer*, *Kaseya* e o famoso escritório de advocacia *Grubman Shire Meiselas & Sacks* (REUTERS, 2021a). Em março de 2021, a quadrilha invadiu a *Acer* e exigiu o maior resgate registrado de 50 milhões de dólares (o maior resgate de *ransomware* conhecido até aquele momento) (GALOV; BEZVERSHENKO; KWIATKOWSKI, 2021).

3.12.4 *Egregor*

Ativo desde 2020, é derivado das famílias de *ransomware Maze* e *Sekhmet*, que já foram desativadas (HEIMDALSECURITY, 2022). A divulgação dos dados de suas vítimas

ocorre através de um site na *deepweb* de nome *Egregor News* e suas vítimas costumam ter somente 72 horas para fazer contato e pagar o valor pelo sequestro de seus dados, sob pena de vazamento. No que tange à encriptação dos dados da vítima, utiliza um esquema de encriptação híbrida, onde emprega o *stream cipher ChaCha* e o RSA (FBI, 2021).

Algumas de suas vítimas foram a empresa *Barnes & Noble*, *Crytek*, *Ubisoft*, *Cencosud*, *Kmart*, *Translink* e *Randstad HR* (HEIMDALSECURITY, 2022). Curiosamente, caso a vítima faça o pagamento do resgate, o grupo oferece um tipo de consultoria, informando a vítima o que foi feito para conseguir acesso e meios de mitigar o problema (ENIGMASOFT, 2020).

3.12.5 *LockBit*

Sua primeira aparição foi em 2019 como *ransomware* ABCD, que foi aprimorado e se tornou uma das famílias mais lucrativas da atualidade. Seus criadores alegam que este é o *ransomware* com maior velocidade de encriptação dos arquivos das vítimas, o que consegue criptografando somente partes dos arquivos, enquanto outros cifram os arquivos por completo (TULAS, 2022), além disso, muda o papel de parede para exibir as instruções de pagamento do resgate.

O *payload* inicia a rotina de encriptação após a execução, que inclui criptografia local e de rede. Ele encripta os arquivos usando AES, cifrando sua chave com o RSA. A chave AES é gerada usando a função *BCryptGenRandom*. O *LockBit* também tem a capacidade de imprimir sua nota de ransom utilizando impressoras conectadas através da API *WinSpool* (TRENDMICRO, 2022b).

Podemos mencionar como vítimas do *LockBit* a empresa Atento (COMPTER, 2022), *Thales Group* (aeroespacial francesa), Ministério da Justiça da França, *Bridgestone Americas*, *Vivalia* (grupo belga de hospitais privados) e a Secretaria Municipal de Fazenda do Rio de janeiro (BLACKFOG, 2022).

3.12.6 *Clop*

Surgiu como uma evolução do *CryptoMix*. No início de 2019, pesquisadores descobriram o uso deste *ransomware* por um grupo chamado de TA505, quando lançou uma campanha de *spear phishing* por e-mail. De acordo com um relatório da *Trend Micro* (TRENDMICRO, 2022a) o ano de 2021 foi o de maior atividade do grupo, com alvos infectados pertencentes ao sistema de saúde (959 infecções), seguido do setor financeiro (150 infecções). Seis

suspeitos de participar da operação foram presos em 2021 na Ucrânia, acusados de somar 500 mil dólares em pagamentos de suas vítimas (IPDFORUM, 2021). No entanto, ao que tudo indica, estes não eram os principais operadores, visto que os ataques continuaram após uma breve pausa nas operações (CYBEREASON, 2021).

Dentre as vítimas do *Clop* estão a *E-Land*, a maior empresa de comércio eletrônico da Coreia do Sul, a *Swire Pacific Offshore*, *Maastricht University*, *Software AG IT*, *ExecuPharm* e *Indiabulls* (BLEEPINGCOMPTER, 2021).

3.12.7 *NetWalker*

Netwalker foi descoberto em 2019, como uma atualização do *Mailto* e nos seus seis primeiros meses de atividade, conseguiu mais de 25 milhões de dólares em resgate (UPGUARD, 2022) e é operado pelo grupo *Cicrcus Spider*. Para manter sua atividade imperceptível nas máquinas das vítimas, utiliza a técnica de *process hollowing*, onde um processo legítimo em execução é substituído por código malicioso, dessa maneira, camuflando-se em processos legítimos do sistema (MOHANTA; SALDANHA, 2020, p. 297).

Como exemplo de vítimas desse *malware*, podemos citar o sistema de saúde *Crozer-Keystone*, a empresa de transporte australiana *Toll Group*, o setor de pesquisas do COVID da Universidade da Califórnia, a cidade de *Wiz* (Áustria), *K-Eletric* (empresa paquistanesa de energia) e a agência oficial de Imigração da Argentina (UPGUARD, 2022).

3.12.8 *MountLocker*

Mountlocker está ativo desde julho de 2020. Os arquivos das vítimas são encriptados utilizando o *ChaCha20* e as chaves são encriptadas usando RSA 2048 (TRENDMICRO, 2021b). Sua implementação aparentemente é segura, no sentido de não apresentar nenhuma falha trivial que permita alguma facilidade de recuperação das chaves de encriptação, no entanto usa um método inseguro (API GetTickCount) para gerar as suas chaves, que podem ficar suscetíveis a ataques de força bruta (THREATPOST, 2021).

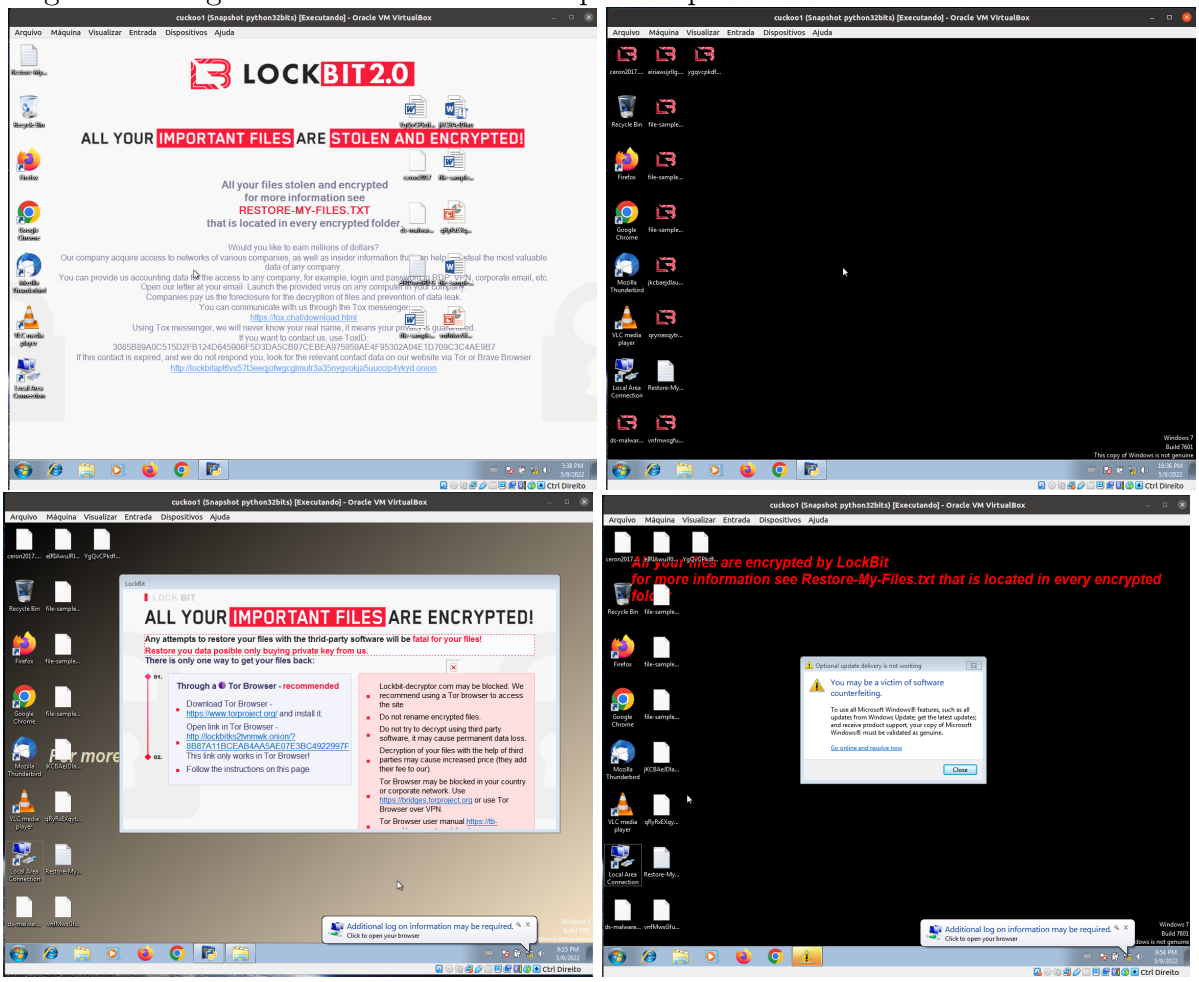
A maior parte de seus ataques são direcionados para o setor de biotecnologia e o maior exemplo de vítima deste *ransmoware* é a empresa *Miltenyi Biotec* (THREATPOST, 2021).

3.13 Análise de *Malware*

Conforme mencionado na Subseção 2.3.2, existem dois métodos de análise de *malware*: Análise Estática e Análise Dinâmica. A Analise Dinâmica observa o comportamento do *malware* e esta foi a abordagem escolhida como caminho par executarmos os experimentos deste trabalho, pois consegue capturar toda a interação do *malware* com o Sistema Operacional.

Durante a execução pudemos acompanhar a interação do *ransomware* com os arquivos da vítima. Tomando o *LockBit*, como exemplo, podemos notar que há razoável diferença entre o funcionamento das amostras analisadas para uma mesma família. Em alguns casos, os arquivos no *desktop* não são alterados, em alguns casos, os ícones dos arquivos criptografados são personalizados com as iniciais do *ransomware* (neste caso, LB). Há diferença no papel de parede: em alguns casos ele já contém as instruções para realização do pagamento, outras apenas o nome do *ransomware* e em outros casos, não ocorre troca. Estas especificidades das amostras do *LockBit* podem ser vistas na Figura 4.

Figura 4: Imagens de áreas de trabalho após ataques de diferentes versões do *LockBit*



A partir das imagens apresentadas, podemos ver algumas diferenças nos ataques realizados pelo *LockBit*. Seguindo a ordem das imagens por linha, da esquerda para a direita, na primeira imagem, vemos que os arquivos do *Desktop* não foram comprometidos e o papel de parede é utilizado para mostrar a nota de *ransom*. Na segunda imagem, os ícones dos arquivos criptografados são personalizados. Na terceira imagem, os arquivos do *Desktop* também são comprometidos, porém a nota de ransom é uma janela criada pelo *LockBit* e a imagem do papel de parede é alterada. Na última, o papel de parede indica que a vítima abra um arquivo de texto com as instruções. Cabe ressaltar que a despeito de haver janela ou papel de parede com instruções, nas quatro análises o *ransomware* grava um arquivo de texto no *Desktop* de nome *Restore-My-Files.txt* com as instruções de pagamento e recuperação dos dados.

4 Levantamento Bibliográfico

Neste capítulo será apresentado o estado da arte e como este trabalho se relaciona com a literatura. Para a realização do levantamento bibliográfico, foram realizadas buscas na literatura nos principais repositórios de artigos acadêmicos a saber: *ACM Digital Library*, *El Compendex*, *IEEE Digital Library*, *ISI Web of Science*, *Science Direct*, *Scopus* e *Springer Link*; a partir dos quais foram selecionados os artigos científicos mais afetos a esta pesquisa.

4.1 Trabalhos Relacionados

Nos últimos 5 anos, as campanhas de *ransomware* têm conseguido movimentar altas quantias de dinheiro em favor dos seus operadores (KASPERSRKY, 2021a), principalmente por causa da pandemia do COVID-19, que impulsionou o trabalho remoto em todo o mundo e abriu brechas antes não pensadas. Por mais que haja movimentação internacional, tanto de autoridades policiais e da academia, no sentido de investigar e coibir essas práticas de extorsão, elas ainda se mostram muito lucrativas. Esse embate, tem como consequência a dissolução e reagregação em novos grupos de agentes maliciosos, muitas vezes reutilizando partes do software anterior, que juntamente com a atualização para incorporação de novas capacidades no *malware*, resultam em uma grande quantidade de variantes do mesmo *ransomware* em atividade. Para acompanhar essa rápida evolução, há demanda por sistemas de análise cada vez mais inteligentes, no sentido de necessitarem de pouca ou nenhuma interação humana e que tenham alta efetividade na proteção dos sistemas. Ambientes *sandbox* são bastante utilizados em análise de *malware*, pois a automatização permite avaliar uma grande quantidade de amostras, facilitando a comparação de diferentes variantes e a construção de conjunto de dados para utilização em Aprendizado de Máquinas, conforme veremos a seguir.

A pesquisa de métodos de detecção de *malware* e, no caso particular dos *ransomwares*, tem uma grande diversidade de técnicas e abordagens que são aprimoradas e combinadas

na tentativa de que se consiga prevenir da melhor maneira possível este tipo de ameaça. O primeiro passo em direção ao estudo de análise de *malware* é a definição da abordagem utilizada para análise (Estática ou Dinâmica). Em seu trabalho, (HWANG et al., 2020) propõe a construção de um modelo híbrido de Análise Dinâmica em dois estágios. No primeiro, utiliza o modelo de *Markov* para modelar o comportamento extraído das sequências de chamadas de API e no segundo utiliza *Random Forest* para os dados restantes, a fim de controlar as taxas de falsos positivos e falsos negativos. Nessa linha de utilização de chamadas de API como base para modelagem do comportamento de *malware*, alguns trabalhos apresentam a proposta *Early Detection*, nos quais o sistema de monitoramento é projetado para detectar e impedir a ação do *ransomware* antes de sua fase de cifração de dados. O trabalho de (CHEN, Q. et al., 2019) é um desses exemplos. Nele os autores implementam um sistema automatizado de extração de padrão de *malware* e detecção precoce, testando três abordagens de Aprendizado de Máquina: TF-IDF (frequência de termo-frequência de documento inversa), LDA de *Fisher* (Análise Discriminante Linear) e ET (árvores extras/extremamente aleatórias) que podem analisar amostras de *malware* recém-descobertas em *sandboxes* e gerar relatórios de Análise Dinâmica (*logs* de *host*), extrair automaticamente a sequência de eventos induzidos por *malware*, dado um grande volume de *logs* de *host* de ambiente (não atacados) e os relativamente poucos *logs* de *hosts* infectados com *malware* potencialmente polimórfico, classificar os recursos mais discriminadores (padrões exclusivos) de *malware* e, a partir do comportamento aprendido, detectar atividades maliciosas e permitir que os operadores visualizem as características discriminantes e suas correlações para facilitar os esforços de analistas de *malware*.

Em alguns trabalhos, tenta-se conciliar os benefícios da Analise Estática com os benefícios da Análise Dinâmica. (SHAUKAT; RIBEIRO, 2018) emprega em seu trabalho esta premissa e apresenta o *RansomWall*, um sistema de defesa em camadas para proteção contra *ransomware* criptográfico que segue uma abordagem híbrida, combinadas para gerar um conjunto compacto de características do *ransomware*. A detecção a partir da Análise Estática considera características do arquivo *Portable Executable* (PE), como detalhes de cabeçalhos, recursos embutidos, detecção de empacotadores/cifradores, entropia da amostra, assinatura digital do arquivo PE, strings embutidas no arquivo e *hashes fuzzy*. Em contrapartida, a detecção da análise dinâmica considera o monitoramento das operações em sistemas de arquivos e modificação na entropia para monitoramento de atividades de encriptação massiva. O trabalho também implementa uma camada armadilha (que os autores chamam de *Honey Files & Trap Layer*), que ajuda na detecção precoce, na medida em que esses arquivos são monitorados pela aplicação e qualquer tentativa suspeita de

alteração alarma o sistema, que verifica se a atividade é autorizada. O sistema também emprega algoritmos de Aprendizado de Máquina para descobrir intrusões *zero-day*: *Logistic Regression*, *Support Vector Machines* (*Gaussian-Kernel*), *Artificial Neural Networks*, *Random Forests* e *Gradient Tree Boosting*. Em (BHAGWAT; PATIL, 2020), os autores apresentam o processo de extração de traços de comportamento de *malware* utilizando *Cuckoo Sandbox* e um *parser* para transformar os relatórios em *dataset*. Foram utilizados os dados relativos a acesso a arquivos, processos e mudanças de registro, porém o trabalho não entra em maiores detalhes sobre quais são exatamente as partes selecionadas e o que representam do conteúdo total dos relatórios. Adicionalmente, não há informação sobre o *parser* utilizado, quais características foram escolhidas após a aplicação dos métodos de redução propostos e nem há disponibilidade de acesso ao *dataset* produzido.

Os esforços da comunidade em tentar detectar e prevenir atividades maliciosas, juntamente com os avanços tecnológicos em criptografia e métodos de prevenção e segurança das aplicações, causam um movimento de avanço na complexidade dos *malwares*, que acabam adotando muitas dessas melhorias (de maneira subversiva) e criando outras para manterem sua furtividade e passarem sem serem detectados. Esse movimento obriga os analistas de *malware* a subirem o nível de seus ambientes de teste, visto que muitos *malwares* passaram a ser criados com a capacidade de monitorar o sistema em que estão sendo executados com a finalidade de evitarem análise. Na literatura, existem extensos trabalhos que abordam esse assunto, como por exemplo os trabalhos de (DARSHAN; KUMARA; JAIDHAR, 2016) e (MILLER et al., 2017). O primeiro propõe a construção de um sistema de Análise Dinâmica escalável, executa experimentos e compartilha as lições aprendidas no processo. A plataforma usa o *Cuckoo Sandbox* para Análise Dinâmica e é preparada para processar *malware* o mais rápido possível sem perda de informações e funciona sobre um *hardware* extremamente robusto, composto de 5 *nodes Cuckoo* gerenciando 100 VMs em virtualização aninhada, de modo a aumentar a proteção do sistema hospedeiro e evitar que os *malware* analisados escapem o ambiente virtualizado. Os autores utilizaram também o INETSIM e o *Paranoid Fish*. As lições aprendidas servem como base para desenvolvedores e utilizadores de sistemas de análise dinâmica semelhantes, como a escolha da plataforma apropriada e desabilitação de funcionalidades desnecessárias à análise. O segundo tem um trabalho exclusivamente voltado para *hardening* da VM, mostrando como um *malware* pode detectar o ambiente virtualizado e como podemos combatê-lo.

Em uma abordagem mais incisiva no que tange a detecção de *ransomware*, (ALSABEH et al., 2020) implementa um sistema que monitora as aplicações executadas em busca de sinais de reconhecimento do sistema em uma abordagem que monitora o comportamento

de um programa interceptando as chamadas APIs do *Windows*. por consequência, é possível determinar em tempo real se o programa está tentando inspecionar seus arredores antes do ataque e abortar sua atividade imediatamente antes do início de qualquer criptografia ou bloqueio malicioso. Em (POPLI; GIRDHAR, 2019) foi realizado um estudo analítico dos comportamentos dos *ransomwares* *WannaCry* e *Petya*. Seu foco principal, é descobrir quais técnicas podem ser aplicadas pelo *ransomware* para serem convertidos em *malwares* polimórficos e metamórficos. Esses tipos de *malware* prescindem de técnicas avançadas de detecção e mitigação para serem combatidos e, além disso, os autores realizam um extenso estudo do estrago que esse tipo de *ransomware* pode causar e o que pode ser feito para proteção. Foram executados amostras em ambiente simulado e seu processo de ataque foi analisado, juntamente com acessos ao sistema de arquivos, persistência e análise em nível de rede. A ferramenta utilizada para análise de comportamento foi o *Cuckoo Sandbox*. Depois disso, os autores tentam prever os futuros tipos de *ransomwares* que podem ser facilmente criados ao se utilizar kits de ferramentas disponíveis como *ADMmutate*, *Clet* e *Phatbot*, além do impacto e ameaça que eles podem causar e quão difícil seria detectá-los depois de empregar todas as técnicas furtivas mencionadas.

Quando um analista de *malware* tem um ambiente preparado para análise e uma abordagem de análise a ser empregada em sua pesquisa, seu próximo passo é determinar como serão extraídas as características dos *malwares* analisados e sua transformação em conjuntos de dados. Na literatura existem inúmeras abordagens que podem ser aplicadas nesse contexto. Algumas mais diretas, como a contagem de chamadas de API ((NDIBANJE et al., 2019; SHARMA; KANT, 2019; CHEN, Q. et al., 2019; TANG; QIAN, 2019)) e outras indiretas como o TF-IDF (ZHANG et al., 2019; AL-RIMY; MAAROF; SHAID, 2019) e N-Gram (ZHANG et al., 2019). Em seu trabalho, (AL-RIMY; MAAROF; SHAID, 2019) transforma as chamadas de API em documentos e aplica a técnica TF-IDF para classificação. Este trabalho também pertence ao grupo descrito no parágrafo anterior, pois também parte da premissa de detectar precocemente *ransomwares*, ou seja, antes que alguma chamada a API criptográfica seja executada. Apresenta também a abordagem *iBagging*, que cria conjuntos de dados com porcentagens de características do ataque até aquele momento organizados por temporalidade, desse modo o sucesso na classificação pode ser medido de acordo com o avanço das ações dos *ransomwares*. O trabalho também reporta que as melhores classificações ocorreram com 80% do total de chamadas de API. No trabalho de (TAKEUCHI; SAKAI; FUKUMOTO, 2018), são utilizadas sequências de chamadas de API, acessos a arquivos e árvores de processos, representadas através de um modelo de vetor. Dos trabalhos revisados, foi o único que mostrou algum detalhe da

configuração utilizada na VM utilizada para análise das amostras. Nesse trabalho, foram distribuídos aleatoriamente 10 pastas em 1000 subpastas dentro de pastas de trabalho do *Windows*, como "Meus Documentos" e "Desktop". Não houve nenhum compartilhamento de arquivos ou programas utilizados pelos autores. Em (BLACK et al., 2020) os autores propuseram um método baseado na observação de que *malwares* executam sequências de API únicas que podem ser utilizadas para distingui-los de outros programas. Os autores consideram a quantidade de chamadas de determinada API e, para cada amostra, as quantidades são transformadas em um vetor. As amostras utilizadas são de tês categorias de programas: maliciosos, benignos e benignos criptográficos. Apesar de mostrar as técnicas utilizadas para extrair as melhores características do conjunto formado, há um hiato em relação ao processo que transforma os relatórios do *sandbox* utilizado no *dataset* inicial. Além disso, os dados não são disponibilizados para acesso. O trabalho de (DINH et al., 2019) realiza três tipos de extração de dados de relatórios de análises do *Cuckoo Sandbox*. O primeiro deles considera o arquivo do relatório de cada a mostra como 1-gram. O segundo considera apenas as categorias *network*, *signatures*, *behaviours* e *others* (não especificadas) para extração de características para o dataset confeccionado e no terceiro, chamado pelos autores de *Behaviour Malware Extraction*, é criado um novo dicionário que recebe cada novo par chave-valor das seções descritas, exceto alguns casos, em que é inserida a quantidade de ocorrências de uma determinada chave. O trabalho de (ZHANG et al., 2019) aplica a abordagem TF-IDF em sequências N-gram extraídas dos códigos de operação de amostras de *ransomwares* baixadas do *VirusTotal* (VT), utilizando a ferramenta IDA. Os dados são submetidos a classificação por cinco algoritmos de ML: *Decisision Tree*, *Random Forest*, *K-Nearest Neighbor*, *Naive Bayes* e *Gradient Boosting* e *Decision Tree* (os dois últimos em conjunto).

Outros trabalhos, como o de (GOYAL et al., 2020) utilizam abordagem de monitoramento de comportamentos para classificação do *malware*. Nele, os autores fazem uma análise do comportamento dos *ransomwares* e delineiam alguns comportamentos específicos que podem ser monitoradas no sentido de permitir a classificação aplicações maliciosas e aplicações genuínas, inclusive aplicações criptográficas. Dentre os comportamentos monitorados o trabalho apresenta a alta taxa de geração de arquivos criptografados, altas operações de gravação de arquivos, alta utilização da CPU, exclusão de *shadow copies*, alteração de registros, renomeação de arquivos, aumento no tamanho dos arquivos, alteração do papel de parede e atividade de rede.

Uma vez que se tenha um conjunto de dados ou métricas para detecção, é necessário utilizar classificadores para identificar cada amostra entre maliciosa e não maliciosa.

Na literatura, existem muitos trabalhos que envolvem essa abordagem, uma vez que um sistema utilize algoritmos de Aprendizado de Máquina, necessita menos interação do operador e se torna mais efetivo, conseguindo manter vigilância em um grande volume de dados. Alguns trabalhos são mais focados na experimentação e melhor utilização de classificadores de Aprendizado de Máquina. Em seu trabalho, (KHAMMAS, 2020) investiga a técnica de Aprendizado de Máquina para a classificação de *ransomware* usando *Random Forest* e recursos extraídos diretamente dos bytes do arquivo. Diferentes tamanhos de semente e árvore foram testados experimentalmente para projetar o melhor classificador de *Random Forest* que pode detectar *ransomware* com precisão. As amostras foram baixadas do VT e a classificação foi executada na ferramenta WEKA.

Mostrando que as diversas técnicas apresentadas podem ser usadas complementarmente para compor um sistema eficiente de classificação de *ransomware*, (ZUHAIR; SELAMAT; KREJCAR, 2020) propõe um modelo de Aprendizado de Máquina híbrido, que é um modelo de análise de *streaming* multicamadas que classifica várias versões de *ransomwares* de 14 famílias, aprendendo 24 características estáticas e dinâmicas. O modelo proposto classifica as versões de *ransomwares* para suas famílias ancestrais numericamente e funde as de famílias multi-descendentes estatisticamente. Para a detecção de *ransomware*, foram criados no sistema de arquivos armadilha, que são monitorados quanto ao acesso não autorizado. O trabalho propõe também uma abordagem de classificação híbrida, combinando *Naive Bayes* e *Decision Tree*, sobre dados de amostras baixadas do VT e *Virus Share* (VS). Nesse trabalho, dois experimentos foram realizados: o primeiro, consistiu em comparar o desempenho do classificador hibrido proposto contra classificadores utilizados comumente na literatura, como LR, SVM, RF, DT, e NB, e o segundo, foi uma comparação entre a ferramenta proposta contra *BitDefender* (ferramenta textitanti-ransomware consagrada, baseada em assinatura), contra *R-Locker* (ferramentas *anti-ransomware* baseadas em uso indevido) e contra *ElderRan* e *RANDS* (ferramentas baseadas em Aprendizado de Máquina). Outro trabalho nessa linha de pensamento, é o desenvolvido por (BHAGWAT; PATIL, 2020), onde os autores utilizaram Análise Dinâmica e os seguintes classificadores: KNN, SVM, *Random Forest* e *Logistic Regression*. Para a seleção de características e redução de dimensionalidade do conjunto de dados, foi usado o *Recursive Feature Elimination* (RFE) e *Extra Trees*. Amostras enviadas a uma máquina *VirtualBox*, monitorado pelo *Cuckoo Sandbox*, que retorna os relatórios (utilizado a parte da análise de comportamento das amostras) em formato JSON. Foram utilizadas 262 características nas quais foram aplicados os métodos de seleção de características. Para seleção foi usado o RFE com validação cruzada. Este método gu-

loso, cria modelos repetidamente e remove a característica mais fraca em cada iteração, construindo o próximo modelo com as características restantes, até terminar com todas e então ranqueia por ordem de eliminação. Usando este método, a melhor quantidade de características encontrada foi 15. O segundo método foi utilizar o classificador *Extra Trees* com *Ginni Index* para medir a importância das características, que são organizadas em ordem decrescente de importância GINI. O usuário deve selecionar um número k de melhores características desejadas (neste trabalho, 15). Depois da seleção o conjunto de dados foi submetido aos seguintes classificadores: KNN, com k = 5, SVM, RF e LR, alcançando 96% de acurácia.

Os dados das amostras foram minerados do VT e submetidos á ferramenta WEKA para classificação em (EGUNJOBI; PARKINSON; CRAMPTON, 2019). A quantidade de amostras foi considerada muito pequena: 100 amostras benignas[1] e 100 de *ransomwares*), fato mencionado pelos próprios autores. Os classificadores utilizados foram SVM com *Sequential Optimization*, *instance-based* (IB1), *NB* e *RF*. Foi aplicada normalização nos dados. Foi conseguido *Accuracy* 99% com RF e SVM. Adicionalmente, os autores disponibilizaram a ferramenta confeccionada para interação com o VT e extração das características usadas para compor o conjunto de dados. Em (ZHAO et al., 2018), um conjunto de amostras reais de *malware* e programas benignos foram baixados do VT e executados em um ambiente virtualizado (*Cucko Sandbox*) para registrar o comportamento do *malware* para avaliação de técnicas de Aprendizado de Máquina em termos de métricas de desempenho comumente usadas. A partir dos relatórios de execução salvos na forma de relatórios JSON, que foram extraídos das amostras de *malware*, o conjunto identificado de recursos é empregado para classificação entre *malware* e amostras benignas, em submissão a 17 classificadores, divididos nas categorias *Ensemble*, *Probability*, *Instance*, *Function*, *Tree* e *Rule based*. A principal motivação deste trabalho é que diferentes técnicas foram projetadas para otimizar diferentes critérios. Então, eles se comportam de maneira diferente, mesmo em condições semelhantes. Além da classificação de *malware* os autores mostram diretrizes para os pesquisadores aplicarem técnicas de Aprendizado de Máquina para detectar *malware* dinamicamente e orientações para pesquisas adicionais na área. Com os resultados das classificações, foi realizada uma análise comparativa empírica das técnicas de ML utilizadas para identificar a que teve melhor desempenho para detecção de *malware*, elegendo-a como uma técnica candidata para o desenvolvimento de sistemas dinâmicos de detecção de *malware*.

Ao analisarmos os trabalhos relacionados em conjunto, vemos que ainda existem al-

[1]Programas não maliciosos, também chamados de *goodware*]

gumas lacunas. Por exemplo, as tarefas de extração de dados, seleção de características e transformação em *dataset* são bastante sucintas, dificultando a reprodutibilidade, além disso os trabalhos que utilizam técnicas de prevenção *anti-VM* não tratam o assunto com profundidade. Adicionalmente, os trabalhos utilizam principalmente as chamadas de API para representar o comportamento do *malware* analisado dinamicamente (poucos comparam técnicas de mineração de dados diferentes). Neste trabalho foram aplicadas duas abordagens distintas, chamadas de API e de mineração de texto,. Foram também aplicados os mesmos algoritmos de classificação para verificação e comparação do desempenho em cada situação proposta com a finalidade de dar uma abrangência tal que facilite a reprodutibilidade dos experimentos ao mesmo tempo evitar saltos de raciocínio na exposição das tarefas realizadas, tanto para as ferramentas que já existem quanto para as que foram confeccionadas especificamente para a realização deste trabalho, de modo que mesmo alguém com pouca experiência na área consiga acompanhar as ideias expostas e reproduzir os experimentos.

5 Procedimentos

Neste capitulo apresentaremos o plano de execução do trabalho a partir da utilização dos conceitos e ferramentas mostrados até agora. Em particular, na Seção 5.2, exploraremos alguns detalhes dos *sripts* que foram escritos para resolvermos alguns problemas, como a mineração de amostras dos repositórios, transformação dos relatórios das análises em um formato de fácil manipulação e extração de características a partir desse subproduto.

5.1 Abordagens Propostas

Foram escolhidas na literatura duas abordagens que se mostraram promissoras. A primeira abordagem é a representação dos *ransomwares* no conjunto de dados através da representação da contagem de chamadas a cada API durante o ataque. Esta técnica foi utilizada com sucesso em (BLACK et al., 2020), (TAKEUCHI; SAKAI; FUKUMOTO, 2018) e (MERCALDO, 2021) e por isso consideramos uma boa candidata a experimentação. A segunda abordagem é emprestada do campo do Processamento de Linguagem Natural, chamado TF-IDF *Term Frequency - Inverse Document Frequency*. Esta abordagem considera a quantidade de vezes que um termo aparece em determinado documento e a quantidade de documentos que contém aquele termo. Cabe ressaltar, que para efeitos deste trabalho, consideramos como documento um arquivo contendo relatório de análise de uma amostra analisada. Esta abordagem foi utilizada com sucesso em (ZHANG et al., 2019) e (AL-RIMY; MAAROF; SHAID, 2019).

5.2 Sequência de Atividades

A seguir será apresentado o passo a passo das atividades executadas para a realização deste trabalho, desde a seleção de amostras nos repositórios até a execução dos algoritmos classificadores e obtenção dos resultados.

5.2.1 Seleção e Download das Amostras

As amostras utilizadas foram selecionadas a partir da busca no banco de dados do VT (SHAUKAT; RIBEIRO, 2018; KOLODENKER et al., 2017; CHEN, L. et al., 2018; BHAGWAT; PATIL, 2020; EGUNJOBI; PARKINSON; CRAMPTON, 2019). De acordo com informações da própria página, o repositório possui um acervo com mais de 15 anos de histórico de arquivos maliciosos analisados e muitos artigos encontrados na literatura tomam como base seus serviços de análise (ZUHAIR; SELAMAT; KREJCAR, 2020; EGUNJOBI; PARKINSON; CRAMPTON, 2019; KHAMMAS, 2020; SHAUKAT; RIBEIRO, 2018). Em complemento ao VT, foram considerados os repositórios *MalwareBazaar* (MB), *Hybryd Analysis* (HB) e *VirusShare* (VS). O VT nos permite realizar buscas pelo nome do *malware* e das várias informações que o relatório de análise apresenta, porém não permite downloads de amostras (somente empresas assinantes do serviço *premium*). Para realizar o download das amostras de forma automática, foram criados scripts em *Python* para interação com as API disponibilizadas nesses repositórios. Dessa maneira, a partir dos resultados da busca de informações sobre as famílias selecionadas no VT (hashes), conseguimos baixar amostras dos outros repositórios, já que não permitem busca nominal. Como nosso interesse é analisar *malware* direcionado ao ambiente *Windows*, pela capacidade e configuração que utilizamos no *Cuckoo Sandbox*, foram baixados apenas arquivos EXE e DLL.

5.2.2 Análise das Amostras e Mineração de Dados

Uma vez que tenhamos as amostras armazenadas localmente, iniciamos o processo de análise e para tal, utilizamos as principais ferramentas apresentadas neste trabalho: *Cuckoo Sandbox* e *VirtualBox*. No ambiente com as configurações apresentadas na Seção 6.1 submeteremos as amostras para análise.

Primeiro, iniciamos o programa principal do *Cuckoo Sandbox* e a VM e em seguida a API e o INETSIM. O *Cuckoo Sandbox* injeta o arquivo de *script* de monitoramento no cliente e a amostra é executada no cliente. O *script* de monitoramento registra diversas informações durante a execução da amostra. Após a conclusão da execução, o *script* de monitoramento envia o resultado registrado de volta ao *Cuckoo* onde o *software* da VM está localizado através da rede virtual, e o componente de análise no *host* analisa e gera um arquivo de relatório. Finalmente, o *software* da VM usa o *snapshot* para restaurá-la ao seu estado inicial. Após todas as amostras analisadas, utilizamos *scripts* criados para extração dos relatórios do *Cuckoo* através da interação com a API disponibilizada.

Dessa forma, foram criados arquivos no formato JSON. Esse formato de arquivo é particularmente interessante de ser utilizado, permitindo fácil conversão em dicionário *Python*, pois utilizam o mesmo formato. A partir dessa transformação foram utilizados *scripts* que serviram de filtro para transformar os dados brutos dos relatórios em características em um conjunto de dados do Pandas[1]. Essa transformação ocorreu de duas maneiras, de acordo com as abordagens propostas (Chamadas de API e mineração de texto com TF-ID). A seleção de características é um passo importante, pois é usada para identificar as características mais significativas, permitindo a geração de modelos de Aprendizado de Máquina mais simples, reduzindo o tempo de treinamento e previsão e ajudando a combater o problema de *overfitting*.

A partir dessas duas abordagens, planejamos realizar classificações utilizando classificadores de Aprendizado de Máquina, a fim de verificar as métricas e avaliar a capacidade da utilização dessas abordagens em uma aplicação de detecção de *malware* em produção.

5.2.3 Pré-processamento e Classificação

Com os conjuntos de dados de cada abordagem em mãos, estamos aptos a aplicar técnicas de a otimização dos dados e realizar a classificação. Para que possamos ter parâmetros para comparar e verificar se realmente houve melhoria na classificação ao aplicarmos os otimizadores, faremos a comparação com os mesmos conjuntos de dados classificados pelos mesmos algoritmos de Aprendizado de Máquina sob os mesmos parâmetros. Neste trabalho utilizamos o termo otimizadores de maneira genérica ao nos referirmos a algoritmos de padronização (*StandardScaler*) e de redução de dimensionalidade (*Principal Component Analysis* - PCA). A aplicação desses métodos pode trazer mais desempenho tanto na questão do tempo de processamento quanto em relação ao acerto na classificação das amostras entre maliciosas e benignas. Ao final, os classificadores treinados conseguem predizer os rótulos de novas amostras na forma de vetores de características (WANG et al., 2019) e, a partir daí, a performance dos métodos podem ser validadas.

[1]Pandas é uma popular biblioteca de manipulação de dados de código aberto para a linguagem de programação *Python*. Ele é construído sobre a biblioteca *NumPy* e fornece estruturas de dados fáceis de usar e ferramentas de análise de dados para lidar e manipular dados tabulares. O Pandas é amplamente usado em ciência de dados, aprendizado de máquina e finanças para pré-processamento, limpeza e análise de dados.

5.3 Estrutura dos Arquivos Analisados

O resultado do monitoramento é um relatório completo do comportamento do *software* analisado, contendo sequência de chamada de função (lista de APIs) e valores de parâmetros chamados quando o *malware* é executado, conexões de rede com detalhamento completo (origem, destino, protocolo e conteúdo), análise de conteúdo da memória da máquina, com endereços de ponteiros, *strings* contidas no programa e arquivos e registros lidos, criados, modificados e apagados.

Os relatórios produzidos pelo *Cuckoo Sandbox* são gerados em formato JSON (*JavaScript Object Notation*) e foram extraídos por um *script*[2], através de interação com sua API. O arquivo JSON gerado pelo *Cuckoo Sandbox* pode ser bastante rico de informações sobre o *malware* analisado. Seu conteúdo é dividido nas seguintes seções:

info: armazena as configurações gerais referentes àquela análise, como data e hora de submissão para análise, início e fim da análise, *score* atribuído, o tipo de arquivo, plataforma, roteamento da rede e opções de configuração.

procmemory: lista com o resultado da análise do *dump* da memória pelo *Volatility*.

target: dicionário contendo dados como regras *Yara*, *hash* SHA256, tipo do arquivo, dentre outros.

extracted: dicionário contendo dados como regras *Yara*, *hash* SHA256, tipo do arquivo, de arquivos extraídos da amostra.

virustotal: dicionário com informações resultantes da consulta à ferramenta VT. Algumas de suas chaves se referem a quantidade de antivírus a que o *malware* foi submetido para análise e quantos indicaram a amostra como maliciosa, *hash* SHA1 e data da análise.

network: dicionário que contém as informações sobre os protocolos usados para comunicação durante a análise, como TLS, TCP, ICMP, DNS etc.

signatures: lista contendo dados sobre as assinaturas que tiveram correspondência durante a análise.

static: dicionário contendo dados de análise estática como data de compilação do programa, DLL importadas, seções do arquivo etc.

[2]Disponível no GitHub em https://github.com/aparisot84/Sandbox-Ransomware-Analysis-Dataset/tree/master/1 - Scripts

dropped: lista cujos itens são os detalhes sobre os *dropped Files*, como *hashes*, regras *Yara*, localização, e PID.

behavior: dicionário resultante da análise dinâmica do *malware* no *sandbox*. Contém a quantidade de chamadas de API dos processos e *treads* criados pelo *malware*, árvore de processos, DLL utilizadas, arquivos e registros criados e removidos etc.

debug: dicionário contendo os *logs* do *sandbox*.

screenshots: lista contendo o caminho das imagens da tela da VM durante a análise.

string: lista contendo todos os textos encontrados no arquivo submetido para análise que podem ser convertidos em *strings* ligíveis.

metadata: lista contendo informações dos arquivos referentes a análise que resultou naquele relatório, como *dumps* de memória, arquivos *packet capture* (PCAP) e *dropped files*.

Tabela 1: Tamanho dos arquivos de cada conjunto de dados produzidos pelo *Cuckoo Sandbox*

Nome do Arquivo	Tamanho
API	2MB
TFIDF Behavior	3.95GB
TFIDF Behaviorccel	190MB
TFIDF Behaviormnr	310MB
TFIDF Behaviorrevil	231MB
TFIDF Memory	223MB
TFIDF Network	31MB
TFIDF Signatures	2MB
TFIDF Strings	589MB

Devido ao tamanho dos relatórios gerados e a quantidade de amostras utilizadas, conforme podemos observar na Tabela 1, houve limitação da capacidade computacional para processamento dos relatórios para transformação nos conjuntos de dados e, por consequência, impossibilidade de utilização de todas as seções dos relatórios de uma só vez. Para sobrepor a este problema, a transformação dos relatórios em conjunto de dados foi realizada por seções. Os detalhes da implementação e das seções utilizadas serão detalhados no Capítulo 5. Ademais, para compor a parte dos conjuntos de dados referentes a amostras de programas não maliciosos foram utilizados nas análises 90 programas, dentre eles utilitários genéricos do Windows, como *drivers*, utilitários de rede, ferramentas de multimídia e instaladores das aplicações que foram instaladas na VM. Da parte componente dos *ransomwares*, foram usados um total de 989 amostras.

5.4 Principais Ferramentas Utilizadas

Como citado anteriormente, para que tenhamos um sistema virtualizado no qual os *ransomwares* efetivamente se manifestem, devemos ter a preocupação de preparar a VM que será usada para análise de modo a evitar que as amostras percebam o ambiente desta forma (MILLS; LEGG, 2020; MAFFIA et al., 2021; GARCIA; DECASTRO-GARCIA, 2021; TRAFIMCHUK; BUKHTEYEV; LADUTSKA, 2022). Assim, para verificação dos traços característicos de virtualização, utilizamos o *Paranoid Fish* (PaFish)[3]) (ORTEGA, 2021) e para convencer o *malware* em análise de que existem serviços de rede disponíveis (conexão com a internet), utilizamos o INETSIM[4].

Com relação aos scripts que foram construídos para interagir com os repositórios de *malware*, selecionar e baixar as amostras maliciosas, baixar os relatórios do *sandbox* em formato JSON e transformá-los em *Dataframe* Pandas foram utilizadas as seguintes bibliotecas *Python*:

- *requests*: esta biblioteca permite enviar requisições HTTP/1.1 com extrema facilidade, sem a necessidade de adicionar *strings* de consulta manualmente às URL ou codificar os dados PUT e POST, bastando utilizar o método JSON, embutido na ferramenta.

- *json*: JSON significa *JavaScript Object Notation* e é uma sintaxe para armazenar e trocar dados em forma de texto. Esta biblioteca do *Python* permite o *script* carregar e descarregar texto nesse formato.

- *pyzipper*: um substituto para o *pyzipper* que pode ler e gravar arquivos zip criptografados com AES, permitindo ao *script* ler e manipular arquivos zip, além de permitir extrair seu conteúdo.

- *Pandas*: Pandas é uma ferramenta de análise e manipulação de dados de código aberto rápida, poderosa, flexível e fácil de usar, construída sobre o *Python*. Neste trabalho, utilizamos esta ferramenta para organizar os dados utilizando o Pandas *Dataframe*, que permite organização de dados de forma tabular bidimensional, de tamanho variável e heterogênea, com estrutura de dados também eixos rotulados (linhas e colunas).

[3]Maiores informações em https://github.com/aparisot84/Sandbox-Ransomware-Analysis-Dataset/wiki

[4]https://www.inetsim.org/

- *math*: Este módulo dá acesso às funções matemáticas definidas pelo padrão C. Neste trabalho, utilizamos esta biblioteca para poder implementar os cálculos da abordagem de mineração de texto.

- *SciKit-Learn*: dispõe de ferramentas simples e eficientes para análise preditiva de dados, é reutilizável em diferentes situações, possui código aberto, sendo acessível a todos e foi construída sobre os pacotes *NumPy*, *SciPy* e *Matplotilib*. Contém uma gama de algoritmos utilizados em Aprendizado de Máquina que vão desde seleção de modelos, pré-processamento e redução de dimensionalidade, até classificação, clusterização e regressão.

5.5 Ferramentas Construídas

Conforme citado anteriormente, para automatização das tarefas de busca, filtragem, seleção, *download* e submissão das amostras, foram criados 4 *scripts* em *Python*. Estes *scripts* são modularizados para que os resultados intermediários pudessem ser avaliados antes de seguirmos para o próximo passo. O primeiro *script* recebe como informação o nome do *malware* a ser procurado no VT e retorna os *hashes* encontrados, divididos em arquivos separados pelo nome do *malware*. A partir desses arquivos contendo os *hashes*, o segundo *script* consulta a base do VT para informações como o tipo de arquivo (somente DLL e EXE serão baixados), campos do relatório do VT como type_description, *type_tag*, *popular_threat_classification*, *suggested_threat_label*, *sandbox_verdicts*, *malware_names* e *type_extension* e procura se existem as amostras para *download* no VS e no MB, respectivamente. O terceiro *script* tem como entrada os relatórios gerados pelo *Cuckoo Sandbox*, a partir da submissão das amostras de *ransomwares* pertencentes às famílias selecionadas (listadas na Seção 3.12). Seu processamento consiste em filtrar as informações selecionadas (conforme mencionado na Subseção 5.2.2) e ao final, retorna um *Dataframe* Pandas. Os *scripts* estão compartilhados no Github[5].

5.6 Ocultação do Ambiente de Teste

Na tentativa de evitar a análise e contornar os sistemas de segurança, os autores de *malware* geralmente projetam seu código para detectar ambientes virtualizados. Uma vez que tal ambiente é detectado, o mecanismo de evasão pode impedir a execução do código

[5]https://github.com/aparisot84/Sandbox-Ransomware-Analysis-Dataset/tree/master/1 - Scripts

malicioso ou alterar seu comportamento para evitar a exposição de atividades maliciosas ao analista. Alguns *malwares* são especificamente direcionados a atacar o ambiente virtualizado, como o *VMWare* ESXI *RansomExx* e *Yanluowang*). Dada a popularização dessa abordagem na concepção de sistemas de informação com a movimentação para serviços hospedados em nuvem (GLOVER, 2022b) e, além disso, a infecção desses sistema tem se mostrado vantajoso para os agentes maliciosos (IBMSECURITY, 2022). Um exemplo de *malware* que especificamente ataca VMs é o *W32.Crisis* através de uma ferramenta de montagem de arquivos VMDK presente no *VirtualBox*. Dessa maneira, consegue ter acesso ao armazenamento das VM e se insere na pasta de inicialização do *Windows*, onde é carregado cada vez que a VM é inicializada (WUEEST, 2014).

No presente trabalho, tivemos a preocupação de ocultar o ambiente virtualizado dos *ransomwares* analisados (OR-MEIR et al., 2019). Com isso, para correção dos indicadores de virtualização listados pelo *PaFish*, as entradas no Registro do *Windows* foram alteradas para os mesmos valores da máquina hospedeira e o MAC *address* padrão do *VirtualBox* foi substituído. Os indicadores relativos às informações da CPU foram mantidas pois teríamos que fazer um *patch* no *hypervisor* (VirtualBox) ou *hooking* nas funções de chamada de informações do CPUID (WUEEST, 2014), porém, como os testes preliminares mostraram que a configuração utilizada permitia aos *ransomwares* funcionarem normalmente, não foram realizadas interferências mais elaboradas na VM.

Os registros alterados para reduzir a assinatura de máquina virtual foram os listados abaixo:

- HKLM\HARDWARE\Description\System "SystemBiosVersion"
- HKLM\HARDWARE\Description\System "VideoBiosVersion"
- HKLM\HARDWARE\ACPI\DSDT\VBOX__
- HKLM\HARDWARE\ACPI\FADT\VBOX__
- HKLM\HARDWARE\ACPI\RSDT\VBOX__
- HKLM\SYSTEM\ControlSet001\Service\Vbox*
- HKLM\HARDWARE\DESCRIPTION\System "SystemBiosDate"

O pacote de adicionais de convidados, disponibilizados no *VirtualBox* não foi instalado, apesar de oferecer melhorias de desempenho e integração com o sistema hospedeiro.

Parte da preparação da VM foi utilizá-la como uma máquina de usuário final para baixar os arquivos e instalar os programas, de modo a popular o histórico de navegação, abertura de arquivos e instalação de programas (MOHANTA; SALDANHA, 2020, p. 34). Nesse sentido, foram instalados aplicações como *Mozilla Thunderbird*, *Mozilla Firefox*, MS *Ofice* 2016, *WinRAR*, VLC *Media Player* e *Java*. Além disso, foram gravados arquivos de vários formatos como documentos, texto, apresentação, som e vídeo em pastas de trabalho do *Windows*, na lixeira e no *Desktop*.

Para reduzir as barreiras de segurança oferecidas pelo próprio Sistema Operacional, foram feitas também algumas alterações na Política de Grupo (os valores entre parênteses são as opções selecionadas):

- *Behavior of the elevation prompt for administrators in Admin Approval Mode (Elevate without prompting)*
- *Detect application installations and prompt for elevation (Disabled)*
- *Run all administrators in Admin Approval Mode (Disabled)*
- *Configure Automatic Updates (Notify for download and notify for install)*
- *Protect all network connections (Disabled)*
- *Turn off Windows Defender Antivirus (Enabled)*

6 Experimentos

Neste capítulo mostraremos os passos seguidos para a execução deste trabalho, o *hardware* disponível para implementação e a configuração das ferramentas utilizadas. Mostraremos também o detalhamento dos experimentos à luz das duas abordagens de Mineração de Dados propostas a partir das análises do *Cuckoo Sandbox*, faremos a avaliação dos experimentos e análise dos resultados, comparando o desempenho dos classificadores para cada configuração proposta.

6.1 Montagem do Ambiente Experimental

A máquina usada no presente trabalho consiste de um Ubuntu 20.04.4 LTS (64 bits) rodando em um Intel® Core™ i7-8550U CPU @ 1.80GHz × 8 com 16 GB RAM DDR4. O S.O. da VM é o *Windows* 7 *Ultimate* SP1 x64 no *VirtualBox* versão 6.1.32 com 4 núcleos e 4 GB de RAM. Todas as implementações dos *scripts* de interação com os repositórios de *malware* e com o *Cuckoo Sandbox* foram feitas utilizando *Python* 3.9, porém a ferramenta *Cuckoo* funciona nativamente sobre a versão 2.7 do *python* (tanto na VM quanto na máquina hospedeira). A configuração do ambiente e das ferramentas utilizadas resultou na arquitetura apresentada na Figura 5.

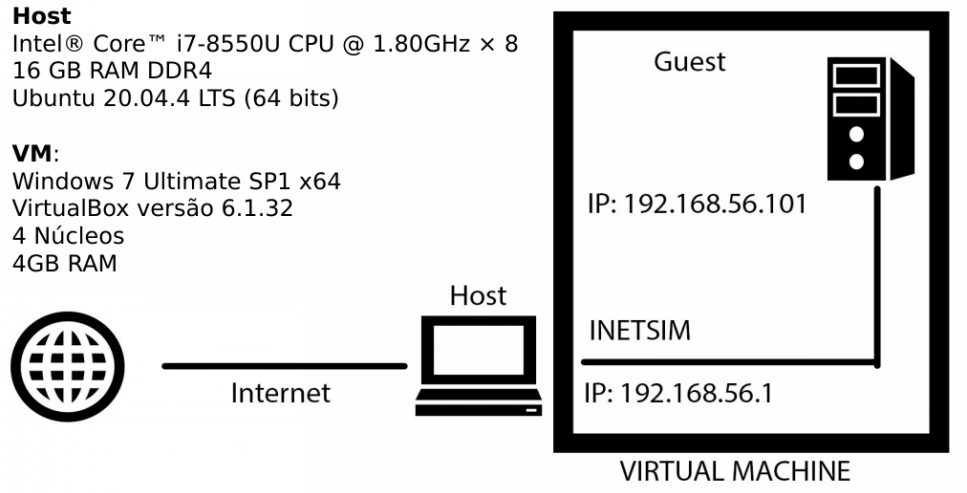
Durante as análises, foi mantido a configuração do *Cuckoo* para abrir a interface gráfica de modo que fosse possível acompanhar o comportamento das amostras.

6.2 Detalhamento dos Experimentos

As amostras usadas nos experimentos foram obtidas a partir dos repositórios *VirusShare*, *MalwareBazaar*, e *HybridAnalysis*, com base nos *hashes* SHA 256 pesquisados no *VirusTotal*, conforme foi discutido na Seção 5.2.1.

Após a busca nos repositórios, avaliação e processamento das listas de *hash* e download

Figura 5: Configuração do ambiente de teste



dos arquivos, cada família ficou com a quantidade de amostras conforme na Tabela 2.

Tabela 2: Quantidade de amostras por família

Nome	**Quantidade de Amostras**
Ryuk	52
Revil	629
NetWalker	78
MountLocker	17
LockBit	49
Egregor	45
Conti	104
Clop	15

Conforme mencionado na Seção 5.1, foram utilizadas duas abordagens para mineração de dados dos relatórios de análise: extração das chamadas de API e TF-IDF. No caso da segunda abordagem, devido a quantidade de amostras e o tamanho de cada relatório, *hardware* disponível para execução dos *scripts* estava aquém do necessário para a execução do trabalho, o processamento foi realizado considerando seções (*Behavior*, *Memory*, *Strings*, *Network* e *Signatures*) individualmente, onde cada uma foi transformada em um *dataset* separado. Para a divisão dos conjuntos de dados entre treino e teste (*test size*) foram utilizados dois valores: 1/3 (2/3 para treinamento e 1/3 para teste) e 1/2 (metade das amostras para treinamento e metade para teste), para que fosse possível verificar se como a diferença na quantidade de treino/teste afetaria os classificadores. Os resultados obtidos com os testes serão apresentados na Seção 6.4. Para que fosse capturado o máximo do comportamento das amostras analisadas, o tempo de execução de cada amostra foi configurado para 600 segundos.

Apesar de algumas famílias de *ransomware* consideradas terem poucas amostras (*Clop* e *Mountlocker*, com 15 e 17 amostras, respectivamente), relativamente a outras (*Revil* e *Conti*, com 629 e 104 amostras, respectivamente), nenhuma foi descartada, pois dessa maneira podemos verificar o desempenho dos mesmos classificadores em conjuntos de dados com muito mais amostras maliciosas do que benignas, quantidades aproximadamente iguais de ambas e com quantidades bem maiores de amostras de *ransomware* do que amostras benignas. Além do mais, tanto para a abordagem que utiliza TF-IDF quanto para a que utiliza as chamadas de API, foram aplicadas classificações multi-rótulo para que pudéssemos verificar a capacidade de diferenciação dentre todas as famílias e os programas benignos (classificação multiclasse) e classificações binárias, para diferenciar cada família individualmente versus as amostras benignas.

Para termos uma abordagem mais direta de quais seriam os parâmetros que resultariam em classificações mais eficientes para cada classificador utilizado, utilizamos o *GridSearchCV* (AL-RIMY; MAAROF; SHAID, 2019). Esta ferramenta realiza busca exaustiva nos parâmetros passados à função ao ser chamada. Um exemplo significativo deste algoritmo neste trabalho pode ser visto nas Figuras 6 e 7, nelas podemos ver que vários valores para estimadores e duas opções de critério foram oferecidas ao algoritmo para a busca da melhor opção. Os outros parâmetros além desses foram mantidos em seus valores padrão para evitar aumento excessivo no tempo de processamento. No caso dos estimadores, o intervalo oferecido (de 10 a 100, em intervalos de 10), não foi o suficiente para encontrar o parâmetro ótimo (o resultado foi 80) e, em uma nova busca, no intervalo de 80 a 87, foi então possível encontrar o parâmetro ótimo, no caso 82.

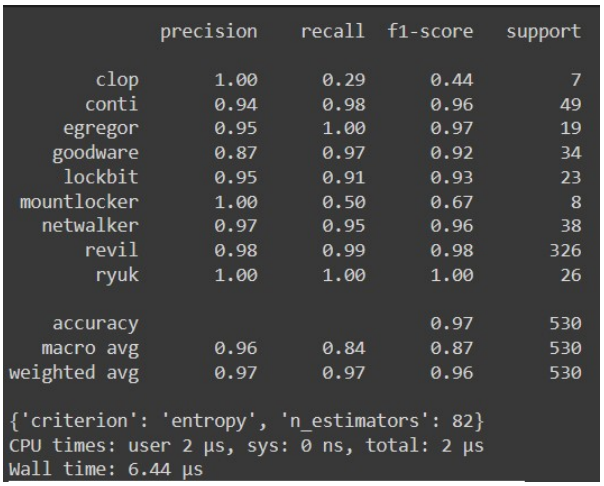
Figura 6: *Script* com a aplicação do *GridSearch* para busca de hiper-parâmetros ótimos para *Random Forest*.

```
from sklearn.ensemble import RandomForestClassifier

# RandomForest com GridSearch
#param_gridrf = {"n_estimators": [10,20,30,50,60,70,80,90,100], 'criterion':['gini', 'entropy']}
# Na primeira tentativa, a busca parou no 80. Para achar o parametro ótimo, fiz novo teste desde 80 até 90

param_gridrf = {"n_estimators": [80,81,82,83,84,85,86,87], 'criterion':['gini', 'entropy']}

modelrfgrid = GridSearchCV(RandomForestClassifier(), param_gridrf, n_jobs=-1, cv=5, error_score='raise')
modelrfgrid.fit(previsores_treinamento, classe_treinamento)
predictrf = modelrfgrid.predict(previsores_teste)
print(metrics.classification_report(classe_teste, predictrf))
print(modelrfgrid.best_params_)
cm = metrics.confusion_matrix(classe_teste, predictrf)
metrics.ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=modelrfgrid.classes_).plot(cmap='cividis')
%time
```

Os resultados do *GridSearchCV* podem ser errôneos, já que a busca dos parâmetros

está restrita aos valores inseridos pelo usuário em *param_grid* e nesse sentido, deve-se ter cuidado de se manter abrangência suficiente para que se direcione para o melhor valor ao mesmo tempo que se evite ter mutos valores, pois o algoritmo é executado com cada combinação e essa busca exaustiva de todos os parâmetros é uma tarefa incrivelmente demorada.

Figura 7: Resultado da aplicação do *GridSearch* para busca de hiper-parâmetros ótimos para *Random Forest*.



	precision	recall	f1-score	support
clop	1.00	0.29	0.44	7
conti	0.94	0.98	0.96	49
egregor	0.95	1.00	0.97	19
goodware	0.87	0.97	0.92	34
lockbit	0.95	0.91	0.93	23
mountlocker	1.00	0.50	0.67	8
netwalker	0.97	0.95	0.96	38
revil	0.98	0.99	0.98	326
ryuk	1.00	1.00	1.00	26
accuracy			0.97	530
macro avg	0.96	0.84	0.87	530
weighted avg	0.97	0.97	0.96	530

```
{'criterion': 'entropy', 'n_estimators': 82}
CPU times: user 2 µs, sys: 0 ns, total: 2 µs
Wall time: 6.44 µs
```

Em nosso exemplo, existem 10 valores para *n_estimators* e 2 para cada um para *criterion*. Isso produzirá um total de 20 combinações diferentes que, em princípio, pode não parecer muito, mas para gerar melhores resultados, também incluímos um valor de validação cruzada de 5 que traz o número total de *jobs* para 100. Cada valor adicionado ao dicionário de grade de parâmetros pode aumentar significativamente o tempo de execução total da função. Caso adicionemos mais algum parâmetro para busca, digamos, com 2 valores, o número total de *jobs* salta de 100 para 200. A espera pela conclusão da busca no gradiente não apenas leva algum tempo, mas assim que obtivermos seus resultados (melhores parâmetros e a respectiva classificação), ainda poderá haver mais alterações feitas em sua grade de parâmetros para possivelmente melhorar seus resultados (como aconteceu no exemplo da Figura 7). Este fato aconteceu nesse experimento e o melhor valor para *n_estimators* foi 80, porém verificamos se poderíamos ter maior desempenho em algum valor entre 80 e 90. A única maneira foi testá-lo utilizando novo intervalo e executando a pesquisa novamente. Mas acabamos de discutir como a adição de parâmetros aumenta significativamente o tempo de execução. Para evitar que a pesquisa demore muito para terminar, podemos retirar valores das extremidades ou substituir por novos valores (conforme fizemos). Dessa forma, o algoritmo continua realizando o mesmo número de *jobs*, mas com valores diferentes para os parâmetros.

Outros otimizadores são computacionalmente mais custosos e mais específicos. O *Stochastic Gradient Descent*[1], segundo a própria documentação, funciona melhor para classificadores lineares, e isso restringiria sua utilização à SVM, indo de encontro a uma das premissas que estabelecemos para o experimento, que é manter as mesmas condições para todos os conjuntos de dados, tanto quanto for possível. Por isso o *Stochastic Gradient Descent* não foi utilizado no nosso trabalho.

6.2.1 Chamadas de API

As chamadas de API representam a comunicação de um programa em execução com o sistema operacional no qual ele está inserido, de modo que as informações conseguidas através de seu monitoramento representam um traçado completo das ações daquele *software* (TAKEUCHI; SAKAI; FUKUMOTO, 2018). Isso permite a criação de perfis de API para detecção e classificação de *malware* (BLACK et al., 2020). Os perfis de frequência de chamadas da API são empregados para identificar o comportamento do *ransomware* em um ambiente controlado. Um exemplo da situação descrita é a diferenciação que podemos fazer de um *ransomware*, que apaga as cópias originais da máquina da vítima após a cifração dos arquivos e um programa de compactação de arquivos (WinZip ou WinRar), que apesar de ter a opção de apagar os arquivos originais depois de terminar a tarefa, não o faz por padrão e é incomum que os usuários a ativem. Na prática, isso se reflete diretamente na quantidade de chamadas a API de encriptação e deleção de arquivos: nos programas benignos as quantidades são muito diferentes entre si, nos *ransomwares*, os valores são próximos.

A seleção das chamadas de API é usada para identificar as chamadas mais significativas, permitindo a geração de modelos de Aprendizado de Máquina mais simples, reduzindo o tempo de treinamento e previsão e ajudando a combater o problema de *overfitting*. Foram consideradas todas as chamadas de API realizadas durante os 600 segundos de monitoramento dos *ransomwares* (BLACK et al., 2020). Apesar da grande dimensionalidade de algumas amostras (o conjunto de dados ficou com 446 colunas), o *overfitting* não será um problema, pois um dos experimentos realizados envolverá comparação do conjunto de dados sem e com redução de dimensionalidade.

Para esta abordagem, foram consideradas as diversas chamadas de API que o *ransomware* faz durante seu ataque. Adicionalmente, também foram incluídas as quantidades de chamadas aos protocolos de comunicação monitorados pelo INETSIM. Estas informa-

[1] https://scikit-learn.org/stable/modules/sgd.html

ções encontram-se nas seções *behavior*, subseção *apistats* e seção *network* dos relatórios. Após gerada a tabela do conjunto de dados, os campos que estavam preenchidos com *NaN*, devido a determinada amostra não ter feito chamada àquela API ou àquele determinado protocolo, foram substituídos por zeros.

Além da classificação, o conjunto de dados nos permite ter uma boa ideia do comportamento dos *ransomwares* em relação aos programas benignos analisados (KOK; ABDULLAH; JHANJHI, 2020), basta compararmos as chamadas de API presente nos *ransomwares* e ausentes dos *Goodware*. Temos abaixo a lista dessas chamadas no nosso conjunto de dados:

- ***CryptEncrypt***: função nativa de cifração de dados.
- ***CryptGenKey***: gera uma chave de sessão criptográfica aleatória ou um par de chaves pública/privada.
- ***CryptDecrypt***: descriptografa dados criptografados anteriormente usando a função *CryptEncrypt*.
- ***CryptExportKey***: exporta uma chave criptográfica ou um par de chaves de um CSP (Provedor de Serviços Criptográficos) de maneira segura.
- ***EnumServicesStatusW***: enumera serviços no banco de dados do gerenciador de controle de serviço especificado.
- ***StartServiceW***: é um identificador para um serviço.
- ***GetFileVersionInfoSizeExW***: determina se o SO pode retornar informações de versão de um determinado arquivo.
- ***GetFileVersionInfoExW***: retorna a informação de versão do arquivo especificado.
- ***InternetOpenUrlA***: abre um recurso especificado por uma URL, FTP ou HTTP completa.
- ***URLDownloadToFileW***: conecta-se à internet para baixar um arquivo, que será salvo em algum local.
- ***CreateRemoteThread***: cria uma *thread* que será executada no espaço virtual de outro processo.

- ***DeleteUrlCacheEntryW***: remove o arquivo associado ao nome de origem do cache, se o arquivo existir.
- ***DecryptMessage***: descriptografa uma mensagem.
- ***EncryptMessage***: criptografa uma mensagem para fornecer privacidade ou fornece um *hash* de integridade que pode ser verificado. Também permite que o aplicativo escolha entre os algoritmos criptográficos suportados pelo mecanismo escolhido.
- ***NtQueryMultipleValueKey***: recupera valores para a chave de múltiplos valores especificada.
- ***NtWriteVirtualMemory***: escreve dados em uma área da memória de um processo específico. Toda a área a ser escrita deve ser acessível ou a operação falha.
- ***NtQueueApcThread***: Abra o identificador para qualquer *Thread Object*, incluindo o *thread* do chamador.
- ***NtTerminateThread***: finaliza um *thread object*.
- ***NtSetContextThread***: identificador para o *Thread Object* aberto com sinalizador de acesso THREAD_SET_CONTEXT.
- ***FindWindowA***: recupera um identificador para a janela de nível superior cujo nome de classe e nome de janela correspondem às cadeias de caracteres especificadas.
- ***WSASend***: envia dados em um *socket* conectado.
- ***LoadStringA***: carrega um recurso de *string* do arquivo executável associado a um módulo especificado e copia a *string* em um *buffer* com um caractere nulo de terminação ou retorna um ponteiro somente leitura para o próprio recurso de *string*.
- ***FindResourceExA***: determina a localização do recurso com o tipo, nome e idioma especificados no módulo especificado.
- ***SendNotifyMessageA***: envia a mensagem especificada para uma janela ou janelas.
- ***WSASocketW***: cria um soquete vinculado a um provedor de serviço de transporte específico.
- ***MessageBoxTimeoutA***: mostra uma caixa de mensagem por um tempo especificado.

- ***gethostbyname***: recupera informações de *host* correspondentes a um nome de *host* de um banco de dados de *host*.
- ***DeleteService***: marca o serviço especificado para exclusão do banco de dados do gerenciador de controle de serviço.
- ***RegQueryInfoKeyA***: recupera informação sobre uma chave de registro específica.
- ***InternetReadFile***: lê dados de um identificador aberto pelas funções *InternetOpenUrl*, *FtpOpenFile* ou *HttpOpenRequest*.
- ***EnumServicesStatusA***: enumera os serviços no banco de dados do gerenciador de controle de serviço especificado. O nome e o status de cada serviço são fornecidos.
- ***SetStdHandle***: define o identificador para o dispositivo padrão especificado (entrada padrão, saída padrão ou erro padrão).
- ***RegisterHotKey***: define uma tecla de acesso para todo o sistema.

Ao analisar o conjunto das funções chamadas exclusivamente pelos *ransomwares* considerados, podemos sobrepor a ocorrência de um comportamento malicioso durante as execuções. Nesse sentido, temos algumas API do *CryptoAPI* do *Windows*, como *CryptEncrypt*, *CryptDecrypt*, *CryptGenKey* e *CryptExportKey*, que caracterizam o comportamento de geração de chaves e encriptação dos dados do usuário. Além desses, temos as funções *DecryptMessage* e *EncryptMessage* que são características de estabelecimento de canal seguro para comunicações, o que neste caso caracteriza envio e recebimento de informações do servidor C&C. No que tange a criação e destruição de *threads*, temos funções como *NtWriteVirtualMemory* *NtQueueApcThread*, *NtTerminateThread*, *NtSetContextThread*. A função *LoadStringA* carrega strings do arquivo executável para a memória, artifício que é particularmente utilizado em *ransomwares*, como o carregamento da *ransom note* e da nota tentando convencer o funcionário da empresa a ceder credenciais de acesso (Figura 3). Outras funções como*FindWindowA*, *SendNotifyMessageA* e *MessageBoxTimeoutA* são utilizadas para apresentar ao usuário a *ransom note* em uma janela temporária do *Windows Explorer*, como é o caso de uma das versões do *LockBit* (Figura 4). Outras funções listadas servem para baixar arquivos de URLs, criar e enviar dados por um *socket*, criar, verificar e desligar serviços etc.

6.2.2 TF-IDF

Com base na descrição apresentada na Seção 2.4.4, foi realizada a implementação do algoritmo do TF-IDF. Os testes de admissibilidade da implementação foram realizados com frações dos conjuntos de dados, de modo que o cálculo pudesse ser realizado manualmente e usado como base para avaliação. Considerando que em um relatório do *Cuckoo Sandbox* temos dicionários e listas aninhados e as informações contidas em cada item são sempre representados por um par chave-valor. Nas diversas seções selecionadas para comporem os conjuntos de dados, inicialmente transformamos os valores correspondentes às chaves de cada relatório em um arquivo de texto (somente com os valores das chaves) que posteriormente foi carregado pelo *script Python* para ter seu TF e IDF calculados e gravados na forma de um *Dataframe* do Pandas.

6.3 Avaliação dos Experimentos

Em cada avaliação a classificação multiclasse como família A é considerada como a classe positiva e as outras X famílias consideradas como a classe negativa. TP (*True Positive*) é o número de amostras que pertencem a família A e são classificadas como pertencendo a esta família. TN (*True Negative*) é a quantidade de amostras que não pertencem a família A e não são classificadas como família A. FP (*False Positive*) é o número de amostras que não pertencem a família A mas são classificadas como tal. FN (*False Negative*) é o número de amostras que pertencem a família A, mas não são reconhecidas como tal. Para avaliar os modelos propostos, os critérios de avaliação abrangem não apenas uma única família de *ransomware*, ou seja, classificação binária, mas também o desempenho dos modelos ao diferenciar as famílias de *ransomware* entre si. Para tal foram considerados as métricas *Precision*, *Recall*, *F1-measure* e *Accuracy*.

A métrica *Precision* mede a performance do classificador em prever uma família positivamente (dentre todas as classificações, quantas o modelo classificou corretamente) e é calculada utilizando-se a Equação 6.1:

$$Precision = \frac{TP}{TP + FP} \tag{6.1}$$

A métrica *Recall* calcula quanto dos positivos nosso modelo realmente capturou como positivo. Esta é a métrica que usamos quando temos um custo alto para Falsos Negativos e seu cálculo é feito através da Equação 6.2:

$$Recall = \frac{TP}{TP + FN} \tag{6.2}$$

A métrica F1 considera *Precision* e *Recall* em conjunto, avaliando a qualidade do classificador, em que quanto mais próximo de 1, melhor é a qualidade do modelo. Esta métrica é calculada utilizando-se a Equação 6.3:

$$F1score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{6.3}$$

Accuracy é a métrica que representa a correta classificação tanto positivamente quanto negativamente. Esta métrica é calculada através da Equação 6.4 :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6.4}$$

Para classificação multiclasse, *Accuracy* é o número de classificações corretas em todas as famílias dividido pelo número total de amostras e pode ser calculada utilizando-se a Equação 6.5:

$$Accuracy = \frac{amostras\ corretamente\ classificadas}{número\ total\ de\ amostras} \tag{6.5}$$

Accuracy é uma boa indicação geral de como o modelo performou. Porém, pode haver situações em que ela é enganosa. Por exemplo, na criação de um modelo de identificação de fraudes em cartões de crédito, o número de casos considerados como fraude pode ser bem pequeno em relação ao número de casos considerados legais. Para colocar em números, em uma situação hipotética de 280000 casos legais e 2000 casos fraudulentos, um modelo simplório que simplesmente classifica tudo como legal obteria uma acurácia de 99,3%. Ou seja, o sistema estaria validando como ótimo um modelo que falha em detectar fraudes.

Recall pode ser usado em uma situação em que os Falsos Negativos são considerados mais prejudiciais que os Falsos Positivos. Por exemplo, o modelo deve de qualquer maneira encontrar todos os pacientes doentes, mesmo que classifique alguns saudáveis como doentes (situação de Falso Positivo) no processo. Ou seja, o modelo deve ter alto *Recall*, pois classificar pacientes doentes como saudáveis pode ser uma tragédia.

F1-*Score* é simplesmente uma maneira de observar somente 1 métrica ao invés de duas (*Precision* e *Recall*) em alguma situação. É uma média harmônica entre as duas, que está muito mais próxima dos menores valores do que uma média aritmética simples. Ou seja,

quando tem-se um F1-*Score* baixo, é um indicativo de que ou *Precision* ou o *Recall* está baixo.

Neste trabalho levaremos em consideração principalmente a métrica *Accuracy*, pois esta é a métrica que corresponde às classificações corretas em relação ao total de classificações e como segundo nível de análise, será considerado o maior *Recall*, pois na situação de detecção de *ransomware*, é melhor classificar negativos como positivos do que positivos como negativos e deixar passar um ataque que possa causar enormes prejuízos. Nesse sentido, a ideia resumida do que significam as métricas *Precision* e *Recall* abaixo:

- ***Precision* alto**: quer dizer que a porcentagem de TP é muito próxima da porcentagem de todas as instâncias realmente positivas (TP + FP) e esta aproximação é possível quando o numero de FP é muito pequeno ou zero.
- ***Recall* alto:** quer dizer que a porcentagem de TP é muito próxima da porcentagem de todas as instâncias classificadas como positivas (TP + FN) e esta aproximação é possível quando o número de FN é muito pequeno ou zero.

6.4 Análise dos Resultados

Esta Seção apresenta a avaliação da efetividade dos métodos propostos para classificação das amostras de *ransomwares* das famílias selecionadas nas diversas condições apresentadas, utilizando algoritmos de Aprendizado de Máquina.

Seguindo a abordagem apresentada na Subseção 6.2.1, realizamos experimentos com o conjunto de dados gerado a partir da contagem de chamadas de API e a partir da classificação da relevância de termos presentes nos relatórios (TF-IDF) das amostras analisadas de *Goodware* e *ransomware* nas duas situações. Foram realizadas classificações multiclasse para verificar a capacidade dos classificadores diferenciarem dentre as famílias selecionadas e classificações binárias entre as amostras benignas e cada família individualmente, assim podemos verificar como um sistema real em produção conseguiria diferenciar um software normal de um *ransomware*. Ambas as classificações foram executadas em duas situações para divisão entre conjunto de dados e teste (test size) 1/3 e 1/2 e com o conjunto de dados sem otimização, após aplicação do *StandardScaler* para padronização dos dados e *Principal Component Analysis* (PCA) com $n = 100$ para redução de dimensionalidade.

Como temos dois tipos de classificação (binária e multiclasse), analisaremos cada uma dentro de suas particularidades, principalmente pela maneira que cada uma ficou organi-

zada. Para a avaliação das classificações multiclasse, consideraremos a capacidade de diferenciação conjunta dentre todas as famílias de *ransomware* e *Goodware* para aquela determinada configuração. Utilizando como exemplo a Tabela 5, o KNN realiza 3 classificações, que são a classificação sem aplicação do *StandardScaler* e sem PCA, somente com *StandardScaler* e com *StandardScaler* e PCA (lembrando que o PCA somente foi aplicado depois do *StandardScaler*), totalizando 18 classificações por tabela. Nesse caso, consideraremos uma classificação maior que 95% excelente, entre 80% e 95% ruins e menor que 80% muito ruins. No caso das classificações binárias, consideramos cada classificação um par *ransomware vs Goodware* como uma classificação. Tomando como exemplo a Tabela 9, consideramos uma classificação o KNN diferenciando entre *Clop* e *Goodware* sem otimização, outra classificação a diferenciação com aplicação do *StandardScaler* no conjunto e outra com a aplicação do PCA, totalizando 144 classificações por tabela. Nesse caso, consideramos baixo desempenho caso a classificação fique abaixo de 90% de *Accuracy*, muito ruim quando abaixo de 80%, boa classificação com *Accuracy* entre 90% e 95% e excelente para classificações que obtiveram 100%.

No intuito de facilitar a visualização dos melhores e piores resultados para as classificações, assim como as métricas *Accuracy* zeradas, utilizamos o seguinte esquema de cores: as células foram preenchidas com azul claro e vermelho claro para os melhores e piores resultados em *Accuracy*, respectivamente e os números em vermelho para zero *Accuracy*.

6.4.1 Resultados experimentais das Chamadas de API

Ao executamos os scripts com os classificadores sobre o conjunto de dados de chamadas de API, foram observadas algumas situações inusitadas:

A primeira delas é que em alguns casos, a métrica *Accuracy* ficou zerada pelo classificador, sem apresentar nenhum erro de processamento da biblioteca utilizada (*SciKit-Learn*). Essas situações estão listadas abaixo:

- No *Clop*, classificação multiclasse com *test size* 1/3 pelo KNN e MLP. No caso da divisão *test size* com 1/2, esta ocorrência apareceu somente para o MLP.
- Na classificação binária, esta mesma ocorrência apareceu em*test size*1/3 na classificação do *Clop*.
- Para classificação multiclasse,*test size*1/3, sem otimização, o *Clop* apresentou esse erro no KNN, NB e MLP. Nesta mesma situação, o *MountLocker* classificado pelo

KNN com *StandardScaler* apresentou o mesmo resultado.

- O *Clop*, com *test size* 1/2, sem otimização na execução do MLP, o *MountLocker* com *StandardScaler* no KNN e com PCA no RF.
- O *Clop* na classificação binária com *test size* 1/3, sem otimização na classificação do KNN e do MLP e o *MountLocker* no MLP apresentaram classificação zero.

Possivelmente este problema ocorre pela pequena quantidade de amostras dessas duas famílias (*Clop* com 15 amostras e *MountLocker* com 17 amostras) e pela dificuldade de generalização do NB, que é um classificador conhecido por apresentar dificuldades de generalização em alguns casos (CHEN, L. et al., 2018; ZHAO et al., 2018; HARAHSHEH; SHRAIDEH; SHARAEH, 2021). Podemos ver isso na prática ao observarmos as métricas, pois todas em que o valor foi zero ocorreram com este classificador. Uma outra consideração que deve ser feita, é que os métodos *ensemble*, apesar de serem mais custosos computacionalmente, geralmente têm melhor desempenho (SINGH; SINGH, 2021), conforme observamos nos resultados conseguidos na classificação pelo RF.

Tabela 3: Extrato das Tabelas 5 e 6 com os melhores resultados de classificação

	Malware	Test Size 0,33			Test Size 0,5		
		Normal			Normal		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score
RF	Clop	0.33	1.00	0.50	1.00	0.29	0.44
	Conti	0.94	0.97	0.95	0.94	0.98	0.96
	Egregor	1.00	1.00	1.00	0.95	1.00	0.97
	Goodware	1.00	0.95	0.98	0.87	0.97	0.92
	LockBit	0.94	0.94	0.94	0.95	0.91	0.93
	MountLocker	1.00	0.33	0.50	1.00	0.50	0.67
	NetWalker	1.00	0.89	0.94	0.97	0.95	0.96
	Revil	0.97	1.00	0.98	0.98	0.99	0.98
	Ryuk	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy	0.97			0.97		

Com relação ao desempenho dos classificadores nas diferentes situações, temos alguns resultados promissores. O RF tanto com*test size*1/3 quanto 1/2 foi o classificador que apresentou maior capacidade de discriminação entre as famílias de *ransomware* (Tabela 3), atingindo *Accuracy* geral de 97% em ambos os valores de *test size* e, individualmente, as métricas F1, *Recall* e *Precision* também apresentaram os melhores resultados, mesmo com a classificação do *Clop* tendo ficado com *Precision* 0,33 e F1 0,50. Podemos observar também que mesmo nas classificações com bom desempenho (*Accuracy* maior que 93%) temos pelo menos alguma família com métricas destoantes das demais (para baixo).

A Tabela 3 mostra um extrato das Tabelas 5 e 6, nela podemos ver que a classificação do *Clop* para *test size* 1/3 e 1/2: no primeiro caso, *Precision* e F1 ficaram em 0.33 e 0.5, respectivamente e para o segundo caso, os baixos resultados foram as métricas *Recall* e F1, com 0.29 e 0.44 respectivamente. Na primeira situação (1/3), as métricas sugerem que não houve falsos negativos na classificação (*Recall* 1.00), mas que tivemos muitos falsos positivo (*Precision* 0.33). Na segunda situação, os dois *ransomwares* ficaram com *Recall* e F1 muito baixos e *Precision* 1.00, sugerindo que não tivemos falsos positivos ou falsos negativos na classificação. As células da Tabela 3, referenciada neste parágrafo foram preenchidas com a cor cinza para melhor visualização.

Por outro lado, o pior desempenho na classificação, foi o obtido com NB otimizado com PCA, tanto para *test size* 1/3 quanto para *test size* 1/2. A Tabela 4 representa o extrato das melhores métricas de classificações encontradas nas Tabela 5 e 6. Nela podemos ver que a classificação realizada pelo NB não apresentou desempenho esperado, muitas métricas próximas ou abaixo de 0.50 (inclusive muito próximas de zero, como F1, *Recall* e *Precision* de *Goodware* para *test size* 1/2 e 1/3), exceto para o *Egregor*, que teve todas as métricas 1.00 nas classificações, ficando muito acima das métricas dos outros *ransomwares* e *Goodwares*. As células da Tabela 4 referenciada neste parágrafo foram preenchidas com a cor cinza para melhor visualização.

Tabela 4: Extrato das Tabelas 5 e 6 com os melhores resultados de classificação

	Malware	Test Size 0,33			Test Size 0,5		
		PCA			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score
NB	Clop	0.32	1.00	0.48	0.32	0.86	0.46
	Conti	0.33	0.82	0.47	0.31	0.86	0.46
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.07	0.12	0.09	0.07	0.12	0.09
	LockBit	0.27	0.48	0.34	0.29	0.52	0.37
	MountLocker	1.00	0.25	0.40	1.00	0.25	0.40
	NetWalker	0.30	0.63	0.40	0.18	0.50	0.26
	Revil	0.95	0.45	0.61	0.96	0.34	0.50
	Ryuk	0.44	0.58	0.50	0.42	0.54	0.47
	Accuracy	0.51			0.43		

As tabelas completas com os resultados de classificação de chamadas de API seguem abaixo:

Tabela 5: Tabela com os dados das classificações referente a abordagem de contagem de chamadas de API, com *test size* 0,33 e classificação multiclasse.

	Malware	Test Size 0,33								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
KNN	Clop	0.00	0.00	0.00	1.00	0.86	0.92	1.00	0.86	0.92
	Conti	0.78	0.94	0.85	0.83	0.88	0.85	0.93	0.86	0.89
	Egregor	0.83	0.91	0.87	0.73	1.00	0.84	0.79	1.00	0.88
	Goodware	0.59	0.62	0.60	0.89	0.91	0.90	0.86	0.91	0.89
	LockBit	0.93	0.78	0.85	0.95	0.91	0.93	0.91	0.91	0.91
	MountLocker	1.00	0.33	0.50	0.00	0.00	0.00	0.50	0.12	0.20
	NetWalker	1.00	0.75	0.86	0.92	0.92	0.92	0.81	0.89	0.85
	Revil	0.92	0.95	0.94	0.97	0.96	0.96	0.98	0.97	0.97
	Ryuk	1.00	0.94	0.97	0.96	0.96	0.96	0.93	0.96	0.94
	Accuracy		0.89			0.93			0.94	
SVM	Clop	0.11	1.00	0.20	1.00	0.86	0.92	1.00	0.86	0.92
	Conti	0.93	0.87	0.90	0.81	0.90	0.85	0.81	0.94	0.87
	Egregor	0.85	1.00	0.92	0.43	1.00	0.60	0.44	1.00	0.61
	Goodware	0.78	0.86	0.82	0.83	0.88	0.86	0.78	0.85	0.82
	LockBit	0.70	0.78	0.74	0.94	0.65	0.77	0.81	0.57	0.67
	MountLocker	0.40	0.33	0.36	1.00	0.12	0.22	0.67	0.25	0.36
	NetWalker	0.79	0.82	0.81	0.97	0.84	0.90	0.89	0.89	0.89
	Revil	0.97	0.94	0.96	0.97	0.93	0.95	0.99	0.92	0.95
	Ryuk	0.92	0.65	0.76	0.96	0.96	0.96	0.93	0.96	0.94
	Accuracy		0.89			0.90			0.89	
NB	Clop	0.00	0.00	0.00	0.19	1.00	0.33	0.32	1.00	0.48
	Conti	1.00	0.42	0.59	0.98	0.84	0.90	0.33	0.82	0.47
	Egregor	0.14	1.00	0.24	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.44	0.19	0.27	0.78	0.41	0.54	0.07	0.12	0.09
	LockBit	1.00	0.89	0.94	0.84	0.91	0.87	0.27	0.48	0.34
	MountLocker	0.40	0.33	0.36	0.27	0.75	0.40	1.00	0.25	0.40
	NetWalker	0.74	0.61	0.67	0.89	0.84	0.86	0.30	0.63	0.40
	Revil	0.98	0.82	0.89	1.00	0.94	0.97	0.95	0.45	0.61
	Ryuk	0.92	0.71	0.80	0.96	0.85	0.90	0.44	0.58	0.50
	Accuracy		0.72			0.89			0.51	
DT	Clop	0.50	1.00	0.67	0.67	0.86	0.75	0.67	0.86	0.75
	Conti	1.00	0.94	0.97	0.96	0.94	0.95	0.96	0.94	0.95
	Egregor	0.92	1.00	0.96	0.76	1.00	0.86	0.76	1.00	0.86
	Goodware	1.00	0.90	0.95	0.76	0.91	0.83	0.76	0.91	0.83
	LockBit	0.85	0.94	0.89	0.84	0.91	0.87	0.84	0.91	0.87
	MountLocker	0.40	0.67	0.50	0.40	0.50	0.44	0.40	0.50	0.44
	NetWalker	1.00	0.96	0.98	0.97	0.95	0.96	0.97	0.95	0.96
	Revil	0.98	0.97	0.97	1.00	0.94	0.97	1.00	0.94	0.97
	Ryuk	1.00	0.94	0.97	0.92	0.92	0.92	0.92	0.92	0.92
	Accuracy		0.95			0.93			0.93	
RF	Clop	0.33	1.00	0.50	1.00	0.14	0.25	1.00	0.86	0.92
	Conti	0.94	0.97	0.95	0.94	0.96	0.95	0.86	0.78	0.82
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	0.95	0.98	0.83	0.85	0.84	0.97	0.85	0.91
	LockBit	0.94	0.94	0.94	0.84	0.91	0.87	0.91	0.91	0.91
	MountLocker	1.00	0.33	0.50	1.00	0.38	0.55	0.50	0.12	0.20
	NetWalker	1.00	0.89	0.94	0.92	0.87	0.89	1.00	0.87	0.93
	Revil	0.97	1.00	0.98	0.97	0.99	0.98	0.92	0.98	0.95
	Ryuk	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.98
	Accuracy		0.97			0.95			0.93	
MLP	Clop	0.00	0.00	0.00	0.86	0.86	0.86	1.00	0.86	0.92
	Conti	0.84	0.68	0.75	0.85	0.92	0.88	0.80	0.92	0.86
	Egregor	0.48	1.00	0.65	0.86	1.00	0.93	0.95	1.00	0.97
	Goodware	0.60	0.14	0.23	0.81	0.88	0.85	0.86	0.91	0.89
	LockBit	0.93	0.78	0.85	0.87	0.87	0.87	0.90	0.83	0.86
	MountLocker	1.00	0.33	0.50	0.75	0.38	0.50	1.00	0.25	0.40
	NetWalker	0.69	0.71	0.70	0.88	0.92	0.90	0.97	0.84	0.90
	Revil	0.89	0.93	0.91	0.98	0.95	0.97	0.95	0.97	0.96
	Ryuk	0.57	0.76	0.65	0.93	0.96	0.94	1.00	0.92	0.96
	Accuracy		0.81			0.93			0.93	

Tabela 6: Tabela com os dados das classificações referente a abordagem de contagem de chamadas de API, com *test size* 0,5 e classificação multiclasse.

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
KNN	Clop	0.75	0.43	0.55	1.00	0.86	0.92	1.00	0.86	0.92
	Conti	0.84	0.88	0.86	0.83	0.88	0.85	0.90	0.88	0.89
	Egregor	0.86	0.95	0.90	0.73	1.00	0.84	0.76	1.00	0.86
	Goodware	0.53	0.62	0.57	0.89	0.91	0.90	0.86	0.91	0.89
	LockBit	0.79	0.83	0.81	0.95	0.91	0.93	0.95	0.91	0.93
	MountLocker	0.75	0.38	0.50	0.00	0.00	0.00	0.50	0.12	0.20
	NetWalker	0.91	0.84	0.88	0.92	0.92	0.92	0.89	0.89	0.89
	Revil	0.94	0.95	0.95	0.97	0.96	0.96	0.97	0.97	0.97

Continua na próxima página

Tabela 6 – continuação da página anterior

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Ryuk	0.95	0.77	0.85	0.96	0.96	0.96	0.93	0.96	0.94
	Accuracy		0.89			0.93			0.94	
SVM	Clop	0.50	1.00	0.67	1.00	0.86	0.92	0.83	0.71	0.77
	Conti	0.87	0.82	0.84	0.81	0.90	0.85	0.83	0.90	0.86
	Egregor	0.70	1.00	0.83	0.43	1.00	0.60	0.41	1.00	0.58
	Goodware	0.77	0.79	0.78	0.83	0.88	0.86	0.77	0.88	0.82
	LockBit	0.59	0.83	0.69	0.94	0.65	0.77	0.93	0.57	0.70
	MountLocker	0.20	0.25	0.22	1.00	0.12	0.22	0.25	0.12	0.17
	NetWalker	0.71	0.84	0.77	0.97	0.84	0.90	0.85	0.87	0.86
	Revil	0.96	0.90	0.93	0.97	0.93	0.95	0.97	0.90	0.93
	Ryuk	0.87	0.50	0.63	0.96	0.96	0.96	0.96	0.96	0.96
	Accuracy		0.86			0.90			0.87	
NB	Clop	0.10	0.71	0.18	0.19	1.00	0.33	0.32	0.86	0.46
	Conti	1.00	0.53	0.69	0.98	0.84	0.90	0.31	0.86	0.46
	Egregor	0.22	1.00	0.37	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.78	0.21	0.33	0.78	0.41	0.54	0.07	0.12	0.09
	LockBit	0.95	0.87	0.91	0.84	0.91	0.87	0.29	0.52	0.37
	MountLocker	0.33	0.25	0.29	0.27	0.75	0.40	1.00	0.25	0.40
	NetWalker	0.73	0.71	0.72	0.89	0.84	0.86	0.18	0.50	0.26
	Revil	0.99	0.84	0.91	1.00	0.94	0.97	0.96	0.34	0.50
	Ryuk	0.85	0.65	0.74	0.96	0.85	0.90	0.42	0.54	0.47
	Accuracy		0.75			0.89			0.43	
DT	Clop	0.67	0.86	0.75	0.67	0.86	0.75	0.67	0.86	0.75
	Conti	0.90	0.94	0.92	0.96	0.94	0.95	0.96	0.94	0.95
	Egregor	0.76	1.00	0.86	0.76	1.00	0.86	0.76	1.00	0.86
	Goodware	0.79	0.88	0.83	0.76	0.91	0.83	0.76	0.91	0.83
	LockBit	0.84	0.91	0.87	0.84	0.91	0.87	0.84	0.91	0.87
	MountLocker	0.40	0.50	0.44	0.40	0.50	0.44	0.40	0.50	0.44
	NetWalker	0.97	0.95	0.96	0.97	0.95	0.96	0.97	0.95	0.96
	Revil	1.00	0.94	0.97	1.00	0.94	0.97	1.00	0.94	0.97
	Ryuk	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92
	Accuracy		0.93			0.93			0.93	
RF	Clop	1.00	0.29	0.44	1.00	0.29	0.44	1.00	0.86	0.92
	Conti	0.94	0.98	0.96	0.92	0.98	0.95	0.89	0.80	0.84
	Egregor	0.95	1.00	0.97	0.90	1.00	0.95	1.00	1.00	1.00
	Goodware	0.87	0.97	0.92	0.89	0.91	0.90	0.88	0.85	0.87
	LockBit	0.95	0.91	0.93	0.88	0.91	0.89	0.88	0.91	0.89
	MountLocker	1.00	0.50	0.67	1.00	0.38	0.55	0.00	0.00	0.00
	NetWalker	0.97	0.95	0.96	0.89	0.89	0.89	0.94	0.87	0.90
	Revil	0.98	0.99	0.98	0.98	0.99	0.98	0.93	0.98	0.95
	Ryuk	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.98
	Accuracy		0.97			0.95			0.93	
MLP	Clop	0.00	0.00	0.00	0.86	0.86	0.86	1.00	0.86	0.92
	Conti	0.60	0.76	0.67	0.85	0.92	0.88	0.83	0.92	0.87
	Egregor	0.46	1.00	0.63	0.86	1.00	0.93	0.95	1.00	0.97
	Goodware	0.67	0.06	0.11	0.81	0.88	0.85	0.86	0.91	0.89
	LockBit	0.77	0.74	0.76	0.87	0.87	0.87	0.89	0.70	0.78
	MountLocker	0.50	0.12	0.20	0.75	0.38	0.50	1.00	0.38	0.55
	NetWalker	0.58	0.84	0.69	0.88	0.92	0.90	1.00	0.84	0.91
	Revil	0.94	0.90	0.92	0.98	0.95	0.97	0.93	0.97	0.95
	Ryuk	0.64	0.62	0.63	0.93	0.96	0.94	1.00	0.88	0.94
	Accuracy		0.79			0.93			0.92	

Os resultados da classificação binária do mesmo conjunto de dados apresentam algumas características semelhantes ao da classificação multiclasse. Por exemplo, podemos citar as ocorrências de métricas de classificação zero, que ocorreram com as mesmas famílias na classificação multiclasse (*Clop* e *Mounlocker*), com *test size* 1/3 nos classificadores MLP e KNN (Tabela 9), porém com uma recorrência menor. Outras duas observações interessantes são que essas ocorrências foram somente com o conjunto de dados sem padronização e que nas classificações com *test size* 1/2 esta situação não ocorreu (Tabela 10).

De modo geral, a aplicação do *StandardScaler* e do PCA melhoram razoavelmente o desempenho na classificação. Em alguns casos específicos os classificadores aumentaram

em 10 pontos percentuais a métrica *Accuracy*, com *test size* 1/3 no *LockBit* e *MountLocker* (classificados pelo KNN), *LockBit* (classificado pela SVM) e *Conti* e *Ryuk* (classificados pela MLP). Nas classificações com o *test size* 1/2 este mesmo fato ocorreu com o *Clop*, *LockBit* e *Mountlocker* (classificados pelo KNN), *Mounlocker* (classificado pela SVM) e *Clop* e *Egregor* (classificados pelo MLP). Em outros casos (principalmente quando a classificação teve pontuação muito baixa), foi capaz de aumentar a taxa de classificação em até 42 pontos percentuais (test size 1/2, *Egregor* classificado pelo NB), como pode ser visto nas Tabelas 7 e 8.

Em 27 classificações, os diversos algoritmos usados foram capazes de fazer classificação perfeita, onde todas as amostras de *ransomwares* são consideradas maliciosas e todas as amostras de *Goodware* são consideradas benignas. Em alguns desses isso aconteceu em todas as 3 situações do conjunto de dados (normal, padronizado e com PCA) tiveram *Accuracy* 100%: com *test size* 1/3, as classificações do *NetWalker* no KNN, do *Clop*, *NetWalker* e *Ryuk* na DT e *Clop* e *MountLocker* na RF e com *test size* 1/2, isso ocorreu nas classificações do *NetWalker* no KNN, *Clop*, *NetWalker* e *Ryuk* no DT, e *Clop* no RF.

As piores classificações ficaram por conta das classificações realizadas pelo NB, tanto com *test size* 1/3 quanto 1/2. Esses resultados foram extraídos para as Tabelas 7 e 8, para uma melhor visualização e as métricas consideradas muito baixas (menores que 0.70) estão com as células preenchidas com cinza.

Tabela 7: Extrato da Tabela 9 com os piores resultados de classificação

	Malware	Test Size 0,33								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
NB	Clop	0.08	1.00	0.14	0.11	1.00	0.19	0.20	1.00	0.33
	Goodware	1.00	0.11	0.20	1.00	0.37	0.54	0.20	0.70	0.83
	Accuracy		0.17			0.41			0.72	
	Conti	0.75	1.00	0.85	0.79	1.00	0.88	0.95	0.97	0.96
	Goodware	1.00	0.38	0.55	1.00	0.52	0.69	0.95	0.90	0.93
	Accuracy		0.78			0.83			0.95	
	Egregor	0.38	0.91	0.54	1.00	0.91	0.95	1.00	0.91	0.95
	Goodware	0.92	0.43	0.59	0.97	1.00	0.98	0.97	1.00	0.98
	Accuracy		0.56			0.97			0.97	
	Goodware	0.96	0.96	0.96	1.00	1.00	1.00	1.00	0.85	0.92
	LockBit	0.92	0.92	0.92	1.00	1.00	1.00	0.77	1.00	0.83
	Accuracy		0.95			1.00			0.85	
	Goodware	1.00	0.21	0.35	0.92	0.43	0.59	1.00	0.71	0.83
	MountLocker	0.08	1.00	0.15	0.06	0.5	0.11	0.20	1.00	0.33
	Accuracy		0.27			0.43			0.73	
	Goodware	0,50	0,09	0,15	0,90	0,41	0,56	0,95	0,86	0,90
	NetWalker	0,57	0,93	0,70	0,68	0,96	0,79	0,90	0,96	0,93
	Accuracy		0,56			0,72			0,92	
	Goodware	0.50	0.27	0.35	0.49	1.00	0.66	0.15	1.00	0.26
	Revil	0.93	0.97	0.95	1.00	0.89	0.94	1.00	0.41	0.58
	Accuracy		0.91			0.90			0.47	
	Goodware	0.88	1.00	0.93	0.93	0.46	0.62	1.00	0.93	0.96
	Ryuk	1.00	0.69	0.82	0.44	0.92	0.60	0.87	1.00	0.93
	Accuracy		0.90			0.61			0.95	

Tabela 8: Extrato da Tabela 10 com os piores resultados de classificação.

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
NB	Clop	0.05	0.67	0.10	0.10	1.00	0.19	0.11	0.67	0.19
	Goodware	0.86	0.15	0.25	1.00	0.37	0.54	0.96	0.61	0.75
	Accuracy		0.18			0.41			0.18	
	Conti	0.68	0.93	0.78	0.73	0.95	0.83	0.96	0.93	0.94
	Goodware	0.69	0.27	0.39	0.82	0.42	0.56	0.89	0.94	0.91
	Accuracy		0.68			0.75			0.93	
	Egregor	0.42	0.95	0.58	1.00	0.95	0.97	1.00	0.95	0.97
	Goodware	0.94	0.38	0.54	0.98	1.00	0.99	0.98	1.00	0.99
	Accuracy		0.56			0.98			0.98	
	Goodware	0.93	0.97	0.95	1.00	1.00	1.00	1.00	0.85	0.92
	LockBit	0.94	0.85	0.89	1.00	1.00	1.00	0.77	1.00	0.87
	Accuracy		0.93			1.00			0.90	
	Goodware	0.90	0.22	0.35	0.94	0.41	0.58	1.00	0.83	0.91
	MountLocker	0.09	0.75	0.15	0.11	0.75	0.19	0.36	1.00	0.53
	Accuracy		0.27			0.44			0.84	
	Goodware	0.57	0.11	0.18	0.87	0.35	0.50	0.97	0.89	0.93
	NetWalker	0.51	0.92	0.66	0.60	0.95	0.73	0.90	0.97	0.94
	Accuracy		0,52			0.65			0,93	
	Goodware	0.50	0.17	0.26	0.94	0.97	0.77	0.18	1.00	0.30
	Revil	0.91	0.98	0.95	1.00	0.94	0.97	1.00	0.49	0.66
	Accuracy		0.90			0.94			0.47	
	Goodware	0.87	1.00	0.93	0.94	0.42	0.59	1.00	0.90	0.95
	Ryuk	1.00	0.73	0.84	0.48	0.95	0.64	0.85	1.00	0.92
	Accuracy		0.90			0.61			0.94	

Na Figura 8, temos uma visualização sumarizada dos resultados das Tabelas 9 e 10. Nele estão dispostas as quantidade de classificação em determinada faixa de valores de *Accuracy*: Azul para 100%, laranja para valores entre 99% e 80% e cinza para valores menores que 80%, além disso, as classificações zero, quando acontecem, são representadas pelo topo amarelo desenhado nas barras. Assim, podemos observar diretamente a quantidade de classificações em cada faixa estipulada e comparar o desempenho de cada classificador:

Podemos observar que os classificadores apresentam desempenhos semelhantes para os dois *test size* escolhidos, porém com algumas diferenças:

- Com a mudança de *test size* de 1/3 para 1/2 o KNN e o MLP deixaram de apresentar classificação zero;
- SVM e DT tiveram menos classificações na faixa azul, mas não apresentaram nenhuma na faixa cinza;
- o NB foi o que apresentou as menores métricas, ou seja, as maiores quantidades de classificações na faixa cinza e na faixa laranja;
- Nessa configuração, os mais indicados para uso em detecção de *ransomwares* são as DT para qualquer tamanho de *test size* e em segundo lugar o KNN para *test size* 1/2.

Figura 8: Sumarização dos resultados das Tabelas 9 e 10.

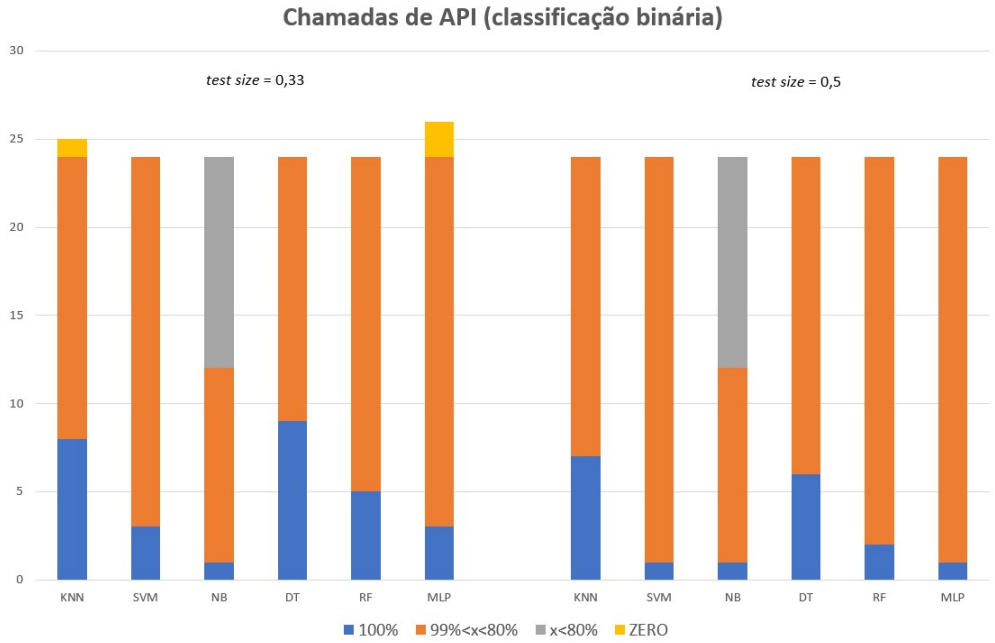


Tabela 9: Tabela com os dados das classificações referente a abordagem de contagem de chamadas de API, com *test size* 0,33 e classificação binária.

	Malware	Test Size 0,33								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
KNN	Clop	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.93	1.00	0.96	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0,93			1.00			1.00	
	Conti	0.94	0.89	0.92	0.93	1.00	0.96	0.93	1.00	0.96
	Goodware	0.83	0.9	0.86	1.00	0.86	0.92	1.00	0.86	0.92
	Accuracy		0.90			0.95			0.95	
	Egregor	1.00	1.00	1.00	0.85	1.00	0.92	0.85	1.00	0.92
	Goodware	1.00	1.00	1.00	1.00	0.93	0.96	1.00	0.95	0.96
	Accuracy		1.00			0.95			0.95	
	Goodware	0.96	0.82	0.88	1.00	0.93	0.96	0.97	0.97	0.97
	LockBit	0.69	0.92	0.79	0.86	1.00	0.92	0.95	0.95	0.95
	Accuracy		0.85			0.95			0.97	
	Goodware	0.96	0.93	0.95	1.00	1.00	1.00	1.00	1.00	1.00
	MountLocker	0.33	0.50	0.4	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.90			1.00			1.00	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	NetWalker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	0.76	0.86	0.81	0.87	0.91	0.89	0.91	0.91	0.91
	Revil	0.99	0.97	0.98	0.99	0.99	0.99	0.99	0.99	0.99
	Accuracy		0.96			0.98			0.98	
	Goodware	0.93	0.93	0.93	1.00	0.96	0.98	1.00	0.96	0.98
	Ryuk	0.85	0.85	0.85	0.93	1.00	0.96	0.93	1.00	0.98
	Accuracy		0.90			0.98			0.98	
SVM	Clop	1.00	1.00	1.00	0.67	1.00	0.8	1.00	1.00	1.00
	Goodware	1.00	1.00	1.00	1.00	0.96	0.98	1.00	1.00	1.00
	Accuracy		1.00			0.97			1.00	
	Conti	0.9	0.95	0.92	0.95	0.97	0.96	0.95	0.97	0.96
	Goodware	0.89	0.81	0.85	0.95	0.9	0.93	0.95	0.9	0.93
	Accuracy		0.9			0.95			0.95	
	Egregor	1.00	0.91	0.95	1.00	1.00	1.00	1.00	0.91	0.95
	Goodware	0.97	1.00	0.98	1.00	1.00	1.00	0.97	1.00	0.98
	Accuracy		0.97			1.00			0.97	
	Goodware	0.96	0.86	0.91	0.97	1.00	0.98	0.95	1.00	0.98
	LockBit	0.73	0.92	0.81	1.00	0.92	0.96	1.00	0.90	0.95
	Accuracy		0.88			0.97			0.97	
	Goodware	1.00	0.86	0.92	1.00	0.93	0.96	0.97	1.00	0.98
	MountLocker	0.33	1.00	0.5	0.5	1.00	0.67	1.00	0.5	0.67
	Accuracy		0.87			0.93			0.97	
	Goodware	0,88	0,95	0,91	0,85	1.00	0,92	0,95	0,95	0,95
	NetWalker	0,96	0,89	0,93	1.00	0,86	0,92	0,96	0,96	0,96
	Accuracy		0,92			0,92			0,96	
	Goodware	0.7	0.73	0.71	0.91	0.91	0.91	0.91	0.91	0.91
	Revil	0.97	0.97	0.97	0.99	0.99	0.99	0.99	0.99	0.99

Continua na próxima página

Tabela 9 – continuação da página anterior

	Malware	Test Size 0,33								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Accuracy		0.94			0.98			0.98	
	Goodware	1.00	0.89	0.94	1.00	0.96	0.98	1.00	0.96	0.98
	Ryuk	0.81	1.00	0.9	0.93	1.00	0.96	0.93	1.00	0.96
	Accuracy		0.93			0.98			0.98	
NB	Clop	0.08	1.00	0.14	0.11	1.00	0.19	0.20	1.00	0.33
	Goodware	1.00	0.11	0.20	1.00	0.37	0.54	0.20	0.7	0.83
	Accuracy		0.17			0.41			0.72	
	Conti	0.75	1.00	0.85	0.79	1.00	0.88	0.95	0.97	0.96
	Goodware	1.00	0.38	0.55	1.00	0.52	0.69	0.95	0.9	0.93
	Accuracy		0.78			0.83			0.95	
	Egregor	0.38	0.91	0.54	1.00	0.91	0.95	1.00	0.91	0.95
	Goodware	0.92	0.43	0.59	0.97	1.00	698	0.97	1.00	0.98
	Accuracy		0.56			0.97			0.97	
	Goodware	0.96	0.96	0.96	1.00	1.00	1.00	1.00	0.85	0.92
	LockBit	0.92	0.92	0.92	1.00	1.00	1.00	0.77	1.00	0.83
	Accuracy		0.95			1.00			0.85	
	Goodware	1.00	0.21	0.35	0.92	0.43	0.59	1.00	0.71	0.83
	MountLocker	0.08	1.00	0.15	0.06	0.5	0.11	0.2	1.00	0.33
	Accuracy		0.27			0.43			0.73	
	Goodware	0,5	0,09	0,15	0,9	0,41	0,56	0,95	0,86	0,9
	NetWalker	0,57	0,93	0,7	0,68	0,96	0,79	0,90	0,96	0,93
	Accuracy		0,56			0,72			0,92	
	Goodware	0.5	0.27	0.35	0.49	1.00	0.66	0.15	1.00	0.26
	Revil	0.93	0.97	0.95	1.00	0.89	0.94	1.00	0.41	0.58
	Accuracy		0.91			0.9			0.47	
	Goodware	0.88	1.00	0.93	0.93	0.46	0.62	1.00	0.93	0.96
	Ryuk	1.00	0.69	0.82	0.44	0.92	0.6	0.87	1.00	0.93
	Accuracy		0.9			0.61			0.95	
DT	Clop	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Conti	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
	Goodware	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
	Accuracy		0.93			0.93			0.93	
	Egregor	0.92	1.00	0.96	0.95	1.00	0.97	0.95	1.00	0.97
	Goodware	1.00	0.96	0.98	1.00	0.97	0.99	1.00	0.97	0.99
	Accuracy		0.97			0.98			0.98	
	Goodware	1.00	0.93	0.96	1.00	0.93	0.96	0.98	1.00	0.99
	LockBit	0.86	1.00	0.92	0.86	1.00	0.92	1.00	0.95	0.97
	Accuracy		0.95			0.95			0.95	
	Goodware	0.97	1.00	0.98	0.97	1.00	0.98	0.97	1.00	0.98
	MountLocker	1.00	0.5	0.67	1.00	0.5	0.67	1.00	0.5	0.67
	Accuracy		0.97			0.97			0.97	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	NetWalker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	0.76	0.86	0.81	0.76	0.86	0.81	0.76	0.86	0.81
	Revil	0.99	0.97	0.98	0.99	0.97	0.98	0.99	0.97	0.98
	Accuracy		0.96			0.96			0.96	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Ryuk	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
RF	Clop	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Conti	0.95	0.95	0.95	0.95	0.95	0.95	0.84	1.00	0.92
	Goodware	0.9	0.9	0.9	0.9	0.9	0.9	1.00	0.67	0.8
	Accuracy		0.93			0.93			0.88	
	Egregor	0.85	1.00	0.92	0.85	1.00	0.92	1.00	0.91	0.95
	Goodware	1.00	0.93	0.96	1.00	0.93	0.96	0.97	1.00	0.98
	Accuracy		0.95			0.95			0.97	
	Goodware	1.00	0.96	0.98	0.98	1.00	0.99	1.00	0.93	0.96
	LockBit	0.92	1.00	0.96	1.00	0.95	0.97	0.87	1.00	0.93
	Accuracy		0.97			0.98			0.95	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.95	0.97
	MountLocker	1.00	1.00	1.00	1.00	1.00	1.00	0.67	1.00	0.80
	Accuracy		1.00			1.00			0.96	
	Goodware	0.88	0.95	0.91	0.95	0.95	0.95	0.95	0.95	0.95
	NetWalker	0.96	0.89	0.93	0.96	0.96	0.96	0.96	0.96	0.96
	Accuracy		0.94			0.98			0.96	
	Goodware	0.95	0.95	0.95	0.91	0.95	0.93	1.00	0.82	0.90
	Revil	1.00	1.00	1.00	1.00	0.99	0.99	0.98	1.00	0.99
	Accuracy		0.99			0.99			0.98	
	Goodware	1.00	0.93	0.96	1.00	0.93	0.96	1.00	0.96	0.98
	Ryuk	0.87	1.00	0.93	0.87	1.00	0.93	0.93	1.00	0.96
	Accuracy		0.95			0.95			0.98	
	Clop	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.93	1.00	0.96	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.93			1.00			1.00	
	Conti	0.91	0.84	0.88	0.93	0.97	0.95	0.86	0.97	0.91
	Goodware	0.75	0.86	0.8	0.95	0.86	0.9	0.94	0.71	0.81
	Accuracy		0.85			0.93			0.88	

Continua na próxima página

MLP

Tabela 9 – continuação da página anterior

	Malware	Test Size 0,33								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Egregor	0.73	1.00	0.85	1.00	0.95	0.97	1.00	0.95	0.97
	Goodware	1.00	0.86	0.92	0.98	1.00	0.99	0.98	1.00	0.99
	Accuracy		0.90			0.98			0.98	
	Goodware	0.9	0.93	0.91	0.96	0.93	0.95	0.95	0.90	0.92
	LockBit	0.82	0.75	0.78	0.85	0.92	0.88	0.82	0.90	0.86
	Accuracy		0.88			0.93			0.90	
	Goodware	0.93	0.96	0.95	1.00	0.93	0.96	1.00	1.00	1.00
	MountLocker	0.00	0.00	0.00	0.5	1.00	0.67	1.00	1.00	1.00
	Accuracy		0.9			0.93			1.00	
	Goodware	0.85	1.00	0.92	0.95	0.95	0.95	0.95	0.95	0.95
	NetWalker	1.00	0.86	0.92	0.96	0.96	0.96	0.96	0.96	0.96
	Accuracy		0.92			0.96			0.96	
	Goodware	0.73	0.86	0.79	0.96	1.00	0.98	1.00	0.86	0.93
	Revil	0.99	0.97	0.98	1.00	1.00	1.00	0.99	1.00	0.99
	Accuracy		0.96			1.00			0.99	
	Goodware	0.87	0.96	0.92	1.00	0.96	0.98	1.00	0.93	0.96
	Ryuk	0.9	0.69	0.78	0.93	1.00	0.96	0.87	1.00	0.93
	Accuracy		0.88			0.98			0.95	

Tabela 10: Tabela com os dados das classificações referente a abordagem de contagem de chamadas de API, com *test size* 0,5 e classificação binária.

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
KNN	Clop	0.20	0.67	0.31	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.97	0.80	0.88	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0,89			1.00			1.00	
	Conti	0.94	0.93	0.94	0.95	0.98	0.96	0.95	0.98	0.96
	Goodware	0.88	0.91	0.90	0.97	0.91	0.94	0.97	0.91	0.94
	Accuracy		0.92			0.95			0.95	
	Egregor	0.76	1.00	0.86	0.90	1.00	0.95	0.90	1.00	0.95
	Goodware	1.00	0.85	0.92	1.00	0.95	0.97	1.00	0.95	0.97
	Accuracy		0.90			0.97			0.97	
	Goodware	0.94	0.80	0.86	0.97	0.97	0.97	0.97	0.97	0.97
	LockBit	0.69	0.90	0.78	0.95	0.95	0.95	0.95	0.95	0.95
	Accuracy		0.83			0.97			0.97	
	Goodware	0.95	0.85	0.90	1.00	1.00	1.00	1.00	1.00	1.00
	MountLocker	0.25	0.50	0.33	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.82			1.00			1.00	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	NetWalker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	0.76	0.89	0.82	0.87	0.97	0.92	0.94	0.97	0.96
	Revil	0.99	0.97	0.98	1.00	0.98	0.99	1.00	0.99	1.00
	Accuracy		0.96			0.98			0.99	
	Goodware	0.95	0.95	0.95	1.00	0.97	0.99	1.00	0.97	0.99
	Ryuk	0.91	0.91	0.91	0.96	1.00	0.98	0.96	1.00	0.98
	Accuracy		0.94			0.98			0.98	
SVM	Clop	0.2	0.67	0.31	0.67	0.67	0.67	1.00	0.67	0.8
	Goodware	0.97	0.8	0.88	0.98	0.98	0.98	0.98	1.00	0.99
	Accuracy		1.00			0.95			0.98	
	Conti	0.93	0.93	0.93	0.95	0.98	0.96	0.95	0.98	0.96
	Goodware	0.88	0.88	0.88	0.97	0.91	0.94	0.97	0.91	0.94
	Accuracy		0.91			0.95			0.95	
	Egregor	1.00	0.96	0.97	1.00	1.00	1.00	1.00	0.95	0.97
	Goodware	0.98	1.00	0.99	1.00	1.00	1.00	0.98	1.00	0.99
	Accuracy		0.98			1.00			0.98	
	Goodware	0.97	0.90	0.94	0.95	1.00	0.98	0.95	1.00	0.98
	LockBit	0.83	0.95	0.88	1.00	0.90	0.95	1.00	0.90	0.95
	Accuracy		0.92			0.97			0.97	
	Goodware	0.95	0.85	0.90	1.00	0.95	0.97	0.95	1.00	.98
	MountLocker	0.25	0.50	0.33	0.67	1.00	0.80	1.00	0.50	0.67
	Accuracy		0.82			0.96			0.96	
	Goodware	0.92	0.95	0.93	0.88	1.00	0.94	0.88	1.00	0.94
	NetWalker	0.95	0.92	0.93	1.00	0.87	0.93	1.00	0.87	0.93
	Accuracy		0.93			0.93			0.93	
	Goodware	0.76	0.83	0.79	0.87	0.94	0.90	0.85	0.94	0.89
	Revil	0.98	0.97	0.98	0.99	0.98	0.99	0.99	0.98	0.99
	Accuracy		0.96			0.98			0.99	
	Goodware	0.97	0.90	0.94	0.95	0.97	0.96	1.00	0.97	0.99
	Ryuk	0.84	0.95	0.89	0.95	0.91	0.93	0.96	1.00	0.98
	Accuracy		0.92			0.95			0.98	
NB	Clop	0.05	0.67	0.10	0.10	1.00	0.19	0.11	0.67	0.19
	Goodware	0.86	0.15	0.25	1.00	0.37	0.54	0.96	0.61	0.75
	Accuracy		0.18			0.41			0.18	
	Conti	0.68	0.93	0.78	0.73	0.95	0.83	0.96	0.93	0.94

Continua na próxima página

Tabela 10 – continuação da página anterior

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Goodware	0.69	0.27	0.39	0.82	0.42	0.56	0.89	0.94	0.91
	Accuracy		0.68			0.75			0.93	
	Egregor	0.42	0.95	0.58	1.00	0.95	0.97	1.00	0.95	0.97
	Goodware	0.94	0.38	0.54	0.98	1.00	0.99	0.98	1.00	0.99
	Accuracy		0.56			0.98			0.98	
	Goodware	0.93	0.97	0.95	1.00	1.00	1.00	1.00	0.85	0.92
	LockBit	0.94	0.85	0.89	1.00	1.00	1.00	0.77	1.00	0.87
	Accuracy		0.93			1.00			0.9	
	Goodware	0.90	0.22	0.35	0.94	0.41	0.58	1.00	0.83	0.91
	MountLocker	0.09	0.75	0.15	0.11	0.75	0.19	0.36	1.00	0.53
	Accuracy		0.27			0.44			0.84	
	Goodware	0.57	0.11	0.18	0.87	0.35	0.50	0.97	0.89	0.93
	NetWalker	0.51	0.92	0.66	0.60	0.95	0.73	0.90	0.97	0.94
	Accuracy		0,52			0.65			0,93	
	Goodware	0.50	0.17	0.26	0.94	0.97	0.77	0.18	1.00	0.30
	Revil	0.91	0.98	0.95	1.00	0.94	0.97	1.00	0.49	0.66
	Accuracy		0.90			0.94			0.47	
	Goodware	0.87	1.00	0.93	0.94	0.42	0.59	1.00	0.90	0.95
	Ryuk	1.00	0.73	0.84	0.48	0.95	0.64	0.85	1.00	0.92
	Accuracy		0.90			0.61			0.94	
DT	Clop	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Conti	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
	Goodware	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
	Accuracy		0.95			0.95			0.95	
	Egregor	0.95	1.00	0.97	0.95	1.00	0.97	0.95	1.00	0.97
	Goodware	1.00	0.97	0.99	1.00	0.97	0.99	1.00	0.97	0.99
	Accuracy		0.98			0.98			0.98	
	Goodware	0.98	1.00	0.99	0.98	1.00	0.99	0.98	1.00	0.99
	LockBit	1.00	0.95	0.97	1.00	0.95	0.97	1.00	0.95	0.97
	Accuracy		0.98			0.98			0.98	
	Goodware	0.98	1.00	0.99	0.98	1.00	0.99	0.98	1.00	0.99
	MountLocker	1.00	0.75	0.86	1.00	0.75	0.86	1.00	0.75	0.86
	Accuracy		0.98			0.98			0.98	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	NetWalker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	0.74	0.91	0.82	0.74	0.91	0.82	0.74	0.91	0.82
	Revil	0.99	0.97	0.98	0.99	0.97	0.98	0.99	0.97	0.98
	Accuracy		0.96			0.96			0.96	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Ryuk	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
RF	Clop	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.67	0.80
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	0.98	1.00	0.99
	Accuracy		1.00			1.00			0.98	
	Conti	0.98	0.89	0.93	0.96	0.96	0.96	0.91	0.96	0.94
	Goodware	0.84	0.97	0.90	0.94	0.94	0.94	0.93	0.85	0.89
	Accuracy		0.95			0.95			0.92	
	Egregor	0.90	1.00	0.95	0.90	1.00	0.95	1.00	0.95	0.97
	Goodware	1.00	0.95	0.97	1.00	0.95	0.97	0.98	1.00	0.99
	Accuracy		0.97			0.97			0.98	
	Goodware	0.97	0.97	0.97	0.98	1.00	0.99	0.97	0.95	0.96
	LockBit	0.95	0.95	0.95	1.00	0.95	0.97	0.90	0.95	0.93
	Accuracy		0.98			0.98			0.95	
	Goodware	1.00	0.95	0.97	1.00	0.95	0.97	0.98	0.98	0.98
	MountLocker	0.67	1.00	0.80	0.67	1.00	0.80	0.75	0.75	0.75
	Accuracy		0.96			0.96			0.96	
	Goodware	0.93	1.00	0.96	0.97	1.00	0.99	0.95	0.95	0.95
	NetWalker	1.00	0.92	0.96	1.00	0.97	0.99	0.95	0.95	0.95
	Accuracy		0.96			0.99			0.95	
	Goodware	0.97	0.94	0.96	0.87	0.97	0.92	1.00	0.91	0.96
	Revil	0.99	1.00	1.00	1.00	0.98	0.99	0.99	1.00	1.00
	Accuracy		0.99			0.99			0.99	
	Goodware	1.00	0.93	0.96	1.00	0.93	0.96	1.00	0.93	0.96
	Ryuk	0.88	1.00	0.94	0.88	1.00	0.94	0.88	1.00	0.94
	Accuracy		0.95			0.95			0.95	
MLP	Clop	0.2	0.67	0.31	1.00	1.00	1.00	1.00	0.33	0.5
	Goodware	0.97	0.8	0.88	1.00	1.00	1.00	0.95	1.00	0.98
	Accuracy		0.8			1.00			0.95	
	Conti	0.91	0.93	0.92	0.96	0.96	0.97	0.95	0.98	0.96
	Goodware	0.88	0.85	0.86	0.97	0.94	0.95	0.97	0.91	0.94
	Accuracy		0.9			0.97			0.92	
	Egregor	0.68	1.00	0.81	1.00	0.95	0.97	1.00	0.95	0.97
	Goodware	1.00	0.78	0.87	0.98	1.00	0.99	0.98	1.00	0.99
	Accuracy		0.85			0.98			0.98	
	Goodware	0.92	0.90	0.91	0.95	0.95	0.95	0.95	0.93	0.94
	LockBit	0.81	0.85	0.83	0.90	0.92	0.90	0.86	0.90	0.88
	Accuracy		0.88			0.93			0.92	
	Goodware	0.97	0.90	0.94	1.00	0.93	0.96	0.98	1.00	0.99
	MountLocker	0.43	0.75	0.55	0.57	1.00	0.73	1.00	0.75	0.86

Continua na próxima página

Tabela 10 – continuação da página anterior

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Accuracy		0.89			0.93			0.98	
	Goodware	0.86	1.00	0.92	0.97	0.97	0.97	0.97	1.00	0.99
	NetWalker	1.00	0.84	0.91	0.97	0.97	0.97	1.00	0.97	0.99
	Accuracy		0.92			0.97			0.99	
	Goodware	0.68	0.91	0.78	0.95	1.00	0.97	0.97	0.91	0.94
	Revil	0.99	0.95	0.97	1.00	0.99	1.00	0.99	1.00	0.99
	Accuracy		0.95			0.99			0.99	
	Goodware	0.97	0.97	0.97	1.00	0.93	0.96	1.00	0.93	0.96
	Ryuk	0.95	0.95	0.95	0.88	1.00	0.94	0.88	1.00	0.94
	Accuracy		0.97			0.95			0.98	

6.4.2 Resultados experimentais TF-IDF

Nesta abordagem pudemos experimentar outro escopo de utilização dos relatório gerados pelo *Cuckoo Sandbox*, dado que a conversão do seu conteúdo em termos textuais nos permite minerar mais informação do que a quantidade de chamadas de API permitiu. A intenção inicial nesta abordagem era utilizar todo o conjunto de relatórios como um grande *corpus* em que cada relatório seria um documento e então aplicar a técnica TF-IDF, a exemplo de como foi feito em (DINH et al., 2019). Entretanto, o tamanho total de todos os relatórios somados tornou impraticável a utilização desta abordagem, pois vários deles estão acima de 1GB e o processamento dos relatórios das mais de mil amostras seria computacionalmente inviável. A partir desse ponto foram feitas novas tentativas de criação de conjunto de dados a partir dos relatórios, porém fazendo a divisão dos conjuntos de dados utilizando cada seção individualmente, conforme mencionado no Capítulo 5.3. Mesmo depois disso, o conjunto de dados criado a partir da seção *Behavior* ficou com quase 4GB (Tabela 1), por este motivo, tivemos que proceder a divisão desse conjunto de dados em conjuntos menores, que pudessem ser processados no *test size* disponível (principalmente por causa da memória RAM). Sendo assim, a classificação multiclasse de todas as famílias de *ransomwares* em conjunto não foi possível diretamente. Para viabilizar a classificação multiclasse, o *corpus*[2] derivado da seção *Behavior* foi dividido em *MountLocker*, *NetWalker*, *Ryuk* e *Goodware* (Tabela 13) e *Clop*, *Conti*, *Egregor*, *LockBit* e *Goodware* (Tabela 12). Com relação ao *Revil*, como esta família é a que possui a maior quantidade de amostras, não foi possível realizar classificação multiclasse em conjunto com nenhuma outra família, além disso, a classificação binária dessas amostras foi apresentada diretamente com a classificação binária das outras famílias (Tabelas 15 e 16). Não obstante, esta abordagem foi a que apresentou menor índice *Accuracy* de todas as tabelas geradas nesta abordagem. Da mesma maneira que fizemos com o conjunto de dados de

[2]Em um conjunto de dados textuais cada unidade de informação é denominada documento; e um conjunto de certos documentos é denominado *corpus*.

Chamadas de API, foram realizadas classificações multiclasse e binária, cada uma com *test size* 1/3 e 1/2, sem otimização, com aplicação do *StandarScaler* para padronização dos dados e PCA com n = 100 para redução de dimensionalidade.

6.4.2.1 Seção *Behavior*

Esta seção corresponde aos dados minerados a partir da seção *Behavior* dos relatórios gerados pelo *Cuckoo Sandbox*, conforme mencionado anteriormente (Seção 6.4.2). Esta seção é a que contém a maior parte do comportamento capturado pelo *sandbox* e, apesar disso, também foi a seção que mais apresentou resultados abaixo de 80% de acurácia na classificação: os resultados obtidos na classificação multiclasse para *Clop*, *Conti*, *Egregor* e *LockBit* (Tabelas 11 e 12). Ao executar a classificação neste conjunto de dados, tivemos muitas classificações abaixo de 80% e nenhuma acima de 87% tanto para *test size* 1/3 quanto para 1/2, mesmo com aplicação do *StandardScaler* e PCA. Além disso, nas classificações com os dois valores de *test size*, houve resultados com zero nas métricas F1, *Recall* e *Precision*, principalmente *Clop* e *Egregor*, o que sugere que a quantidade de amostras, mesmo nesta abordagem, prejudica a diferenciação das famílias pelos classificadores. Há que se notar também o baixo desempenho do MLP ao classificar os famílias nesse conjunto de dados. No conjunto contendo as famílias *MountLocker*, *NetWalker*, *Ryuk* (Tabelas 13 e 14), os resultados foram sensivelmente melhores, no que se refere a classificação com *Accuracy* zerada. As métricas *Accuracy* se mantiveram entre 80% e 87% em ambos os valores de *test size*. Nesta situação, os classificadores RF e DT apresentaram os melhores valores de classificação. Pelo menor tamanho dos conjuntos de dados, não apresentaremos tabelas resumo dos resultados.

Tabela 11: Tabela com os dados das classificações referente a abordagem de TF-IDF (Behavior), com *test size* 0,33 e classificação multiclasse (Clop, Conti, Egregor e LockBit).

	Malware	Test Size 0,33								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
KNN	Clop	1.00	0.17	0.29	1.00	0.17	0.29	1.00	0.17	0.29
	Conti	0.60	0.91	0.72	0.49	0.59	0.54	0.80	0.75	0.77
	Egregor	0.00	0.00	0.00	0.00	0.00	0.00	0.94	0.88	0.91
	Goodware	0.84	0.88	0.86	0.57	0.96	0.72	0.71	0.83	0.77
	LockBit	0.68	0.87	0.76	0.93	0.87	0.90	0.68	0.87	0.76
	Accuracy		0.68			0.60			0.78	
SVM	Clop	0.00	0.00	0.00	1.00	0.17	0.29	0.00	0.00	0.00
	Conti	0.86	0.78	0.82	0.59	0.94	0.72	1.00	0.62	0.77
	Egregor	0.71	1.00	0.83	0.00	0.00	0.00	0.85	1.00	0.92
	Goodware	0.79	0.46	0.58	0.79	0.92	0.85	0.56	0.92	0.70
	LockBit	0.52	0.93	0.67	0.93	0.87	0.90	0.57	0.53	0.55
	Accuracy		0.71			0.70			0.71	
NB	Clop	0.05	0.17	0.08	1.00	0.17	0.29	0.12	0.17	0.14
	Conti	0.81	0.53	0.64	0.59	0.94	0.72	0.69	0.28	0.40
	Egregor	0.71	1.00	0.83	0.00	0.00	0.00	0.35	1.00	0.52
	Goodware	0.53	0.33	0.41	0.79	0.92	0.85	0.78	0.29	0.42
	LockBit	0.93	0.87	0.90	0.93	0.87	0.90	0.87	0.87	0.87
	Accuracy		0.60			0.56			0.50	
	Clop	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Conti	0.87	0.84	0.86	0.87	0.84	0.86	0.87	0.84	0.86
DT										

Continua na próxima página

Tabela 11 – continuação da página anterior

	Malware	Test Size 0,33								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Egregor	0.85	1.00	0.92	0.85	1.00	0.92	0.85	1.00	0.92
	Goodware	0.80	1.00	0.89	0.80	1.00	0.89	0.80	1.00	0.89
	LockBit	1.00	0.87	0.93	1.00	0.87	0.93	1.00	0.87	0.93
	Accuracy		0.86			0.86			0.86	
RF	Clop	0.00	0.00	0.00	1.00	0.17	0.29	1.00	0.17	0.29
	Conti	0.87	0.84	0.86	0.85	0.91	0.88	0.80	0.88	0.84
	Egregor	0.85	1.00	0.92	0.94	0.94	0.94	0.85	1.00	0.92
	Goodware	0.80	1.00	0.89	0.79	0.96	0.87	0.80	0.83	0.82
	LockBit	1.00	0.87	0.93	1.00	0.87	0.93	0.92	0.80	0.86
	Accuracy		0.87			0.87			0.83	
MLP	Clop	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Conti	0.34	1.00	0.51	0.43	1.00	0.60	0.60	0.56	0.58
	Egregor	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.59	0.74
	Goodware	0.00	0.00	0.00	1.00	0.12	0.22	0.39	0.67	0.49
	LockBit	0.00	0.00	0.00	0.88	0.93	0.90	1.00	0.87	0.93
	Accuracy		0.34			0.52			0.61	

Tabela 12: Tabela com os dados das classificações referente a abordagem de TF-IDF (Behavior), com *test size* 0,5 e classificação multiclasse (Clop, Conti, Egregor e LockBit).

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
KNN	Clop	1.00	0.29	0.44	1.00	0.29	0.44	1.00	0.29	0.44
	Conti	0.67	0.84	0.75	0.84	0.48	0.61	0.89	0.55	0.68
	Egregor	0.00	0.00	0.00	0.00	0.00	0.00	0.83	0.90	0.86
	Goodware	0.82	0.89	0.85	0.38	0.94	0.55	0.60	0.89	0.71
	LockBit	0.68	0.91	0.78	0.95	0.91	0.93	0.70	0.91	0.79
	Accuracy		0.71			0.58			0.73	
SVM	Clop	1.00	0.14	0.25	1.00	0.29	0.44	0.50	0.14	0.22
	Conti	0.94	0.79	0.85	0.81	0.84	0.82	0.94	0.55	0.70
	Egregor	0.71	0.95	0.82	0.00	0.00	0.00	0.87	0.95	0.91
	Goodware	0.74	0.49	0.59	0.55	0.94	0.69	0.56	0.94	0.70
	LockBit	0.53	1.00	0.70	0.95	0.91	0.93	0.68	0.74	0.71
	Accuracy		0.74			0.73			0.72	
NB	Clop	1.00	0.14	0.25	0.08	0.29	0.12	0.06	0.14	0.08
	Conti	0.94	0.79	0.85	0.87	0.59	0.70	0.77	0.30	0.44
	Egregor	0.71	0.95	0.82	0.61	0.90	0.73	0.30	0.95	0.45
	Goodware	0.74	0.49	0.59	0.48	0.31	0.38	0.78	0.20	0.32
	LockBit	0.53	1.00	0.70	0.84	0.91	0.87	0.78	0.91	0.84
	Accuracy		0.64			0.61			0.46	
DT	Clop	1.00	0.14	0.25	0.67	0.29	0.40	0.67	0.29	0.40
	Conti	0.94	0.79	0.85	0.92	0.80	0.86	0.92	0.80	0.86
	Egregor	0.71	0.95	0.82	0.83	0.95	0.89	0.83	0.95	0.89
	Goodware	0.74	0.49	0.59	0.74	0.91	0.82	0.74	0.91	0.82
	LockBit	0.53	1.00	0.70	0.91	0.91	0.91	0.91	0.91	0.91
	Accuracy		0.85			0.85			0.85	
RF	Clop	1.00	0.29	0.44	1.00	0.29	0.44	1.00	0.29	0.44
	Conti	0.89	0.86	0.87	0.87	0.86	0.86	0.82	0.84	0.83
	Egregor	0.95	0.90	0.93	0.95	0.90	0.93	0.87	0.95	0.91
	Goodware	0.75	0.94	0.84	0.75	0.94	0.84	0.76	0.80	0.78
	LockBit	0.95	0.91	0.93	1.00	0.91	0.95	0.87	0.87	0.87
	Accuracy		0.87			0.87			0.82	
MLP	Clop	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Conti	0.39	1.00	0.57	0.43	0.79	0.55	0.64	0.61	0.62
	Egregor	0.00	0.00	0.00	0.00	0.00	0.00	0.93	0.62	0.74
	Goodware	0.00	0.00	0.00	1.00	0.14	0.25	0.44	0.66	0.53
	LockBit	0.00	0.00	0.00	0.91	0.87	0.89	0.95	0.87	0.91
	Accuracy		0.39			0.49			0.63	

Tabela 13: Tabela com os dados das classificações referente a abordagem de TF-IDF (Behavior), com *test size* 0,33 e classificação multiclasse (Mountlocker, netwalker, Ryuk).

	Malware	Test Size 0,33								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
KNN	Goodware	0.81	0.96	0.88	0.45	1.00	0.62	0.62	0.96	0.76
	MountLocker	1.00	0.20	0.33	1.00	0.40	0.57	1.00	0.40	0.57
	NetWalker	1.00	0.67	0.80	1.00	0.19	0.31	0.83	0.70	0.76
	Ryuk	0.57	0.87	0.68	1.00	0.53	0.70	1.00	0.53	0.70
	Accuracy		0.78			0.56			0.74	

Continua na próxima página

Tabela 13 – continuação da página anterior

	Malware	Test Size 0,33								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
SVM	Goodware	1.00	0.50	0.67	0.59	1.00	0.74	0.48	0.96	0.64
	MountLocker	0.00	0.00	0.00	1.00	0.40	0.57	0.00	0.00	0.00
	NetWalker	0.45	1.00	0.62	1.00	0.67	0.80	1.00	0.30	0.46
	Ryuk	0.00	0.00	0.00	1.00	0.60	0.75	0.83	0.67	0.74
	Accuracy		0.55			0.75			0.59	
NB	Goodware	0.83	0.73	0.78	0.80	0.77	0.78	0.57	0.81	0.67
	MountLocker	0.29	1.00	0.45	0.38	1.00	0.56	0.33	0.60	0.43
	NetWalker	0.96	0.89	0.92	0.92	0.85	0.88	0.88	0.52	0.65
	Ryuk	1.00	0.53	0.70	1.00	0.67	0.80	0.91	0.67	0.77
	Accuracy		0.77			0.79			0.66	
DT	Goodware	0.86	0.92	0.89	0.86	0.92	0.89	0.86	0.92	0.89
	MountLocker	0.67	0.40	0.50	0.67	0.40	0.50	0.67	0.40	0.50
	NetWalker	0.79	0.96	0.87	0.79	0.96	0.87	0.79	0.96	0.87
	Ryuk	0.89	0.53	0.67	0.89	0.53	0.67	0.89	0.53	0.67
	Accuracy		0.82			0.82			0.82	
RF	Goodware	0.78	0.96	0.86	0.78	0.96	0.86	0.84	0.81	0.82
	MountLocker	0.80	0.80	0.80	0.80	0.80	0.80	0.57	0.80	0.67
	NetWalker	0.89	0.93	0.91	0.89	0.93	0.91	0.86	0.93	0.89
	Ryuk	1.00	0.53	0.70	1.00	0.53	0.70	0.83	0.67	0.74
	Accuracy		0.85			0.85			0.82	
MLP	Goodware	0.89	0.62	0.73	0.00	0.00	0.00	0.55	0.81	0.66
	MountLocker	0.00	0.00	0.00	0.40	0.40	0.40	0.33	0.20	0.25
	NetWalker	0.00	0.00	0.00	0.46	1.00	0.63	0.72	0.67	0.69
	Ryuk	0.25	0.93	0.40	1.00	0.60	0.75	1.00	0.47	0.64
	Accuracy		0.41			0.52			0.64	

Tabela 14: Tabela com os dados das classificações referente a abordagem de TF-IDF (Behavior), com *test size* 0,5 e classificação multiclasse (Mountlocker, netwalker, Ryuk).

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
KNN	Goodware	0.83	0.90	0.86	0.47	0.95	0.63	0.61	0.92	0.73
	MountLocker	1.00	0.12	0.22	1.00	0.38	0.55	1.00	0.38	0.55
	NetWalker	1.00	0.69	0.82	1.00	0.18	0.30	0.80	0.62	0.70
	Ryuk	0.55	0.92	0.69	0.86	0.75	0.80	0.89	0.67	0.76
	Accuracy		0.77			0.59			0.72	
SVM	Goodware	0.83	0.90	0.86	0.60	1.00	0.75	0.50	0.97	0.66
	MountLocker	1.00	0.12	0.22	1.00	0.12	0.22	0.00	0.00	0.00
	NetWalker	1.00	0.69	0.82	1.00	0.69	0.82	1.00	0.31	0.47
	Ryuk	0.55	0.92	0.69	1.00	0.71	0.83	0.86	0.75	0.80
	Accuracy		0.46			0.76			0.62	
NB	Goodware	0.91	0.74	0.82	0.91	0.77	0.83	0.25	0.03	0.05
	MountLocker	0.22	0.88	0.35	0.22	0.50	0.31	0.08	0.62	0.14
	NetWalker	0.97	0.85	0.90	0.94	0.82	0.88	0.88	0.54	0.67
	Ryuk	0.92	0.46	0.61	0.80	0.83	0.82	0.95	0.75	0.84
	Accuracy		0.73			0.78			0.41	
DT	Goodware	0.87	0.85	0.86	0.87	0.85	0.86	0.87	0.85	0.86
	MountLocker	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38
	NetWalker	0.81	0.90	0.85	0.81	0.90	0.85	0.81	0.90	0.85
	Ryuk	0.86	0.75	0.80	0.86	0.75	0.80	0.86	0.75	0.80
	Accuracy		0.81			0.81			0.81	
RF	Goodware	0.79	0.95	0.86	0.79	0.95	0.86	0.83	0.90	0.86
	MountLocker	0.75	0.38	0.50	1.00	0.38	0.55	0.60	0.38	0.46
	NetWalker	0.88	0.92	0.90	0.90	0.95	0.92	0.79	0.85	0.81
	Ryuk	0.89	0.67	0.76	0.89	0.71	0.79	0.90	0.79	0.84
	Accuracy		0.84			0.85			0.82	
MLP	Goodware	1.00	0.56	0.72	0.67	0.05	0.10	0.59	0.82	0.69
	MountLocker	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.25	0.33
	NetWalker	0.00	0.00	0.00	0.46	1.00	0.63	0.60	0.69	0.64
	Ryuk	0.27	1.00	0.43	0.78	0.75	0.77	1.00	0.29	0.45
	Accuracy		0.42			0.54			0.62	

Tabela 15: Tabela com os dados das classificações referente a abordagem de TF-IDF (Behavior), com *test size* 0,33 e classificação Binária.

	Malware	Test Size 0,33								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Clop	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Continua na próxima página

Tabela 15 – continuação da página anterior

	Malware	Test Size 0,33								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Accuracy		1.00			1.00			1.00	
	Conti	0.97	0.82	0.89	0.96	0.58	0.72	0.96	0.66	0.78
	Goodware	674	0.95	0.83	0.56	0.95	0.7	0.61	0.95	0.74
	Accuracy		0.86			0.71			0.76	
	Egregor	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Goodware	0.72	1.00	0.84	0.72	1.00	0.84	0.72	1.00	0.84
	Accuracy		0.72			0.72			0.72	
	Goodware	0.96	0.86	0.91	0.93	1.00	0.97	0.93	1.00	0.97
	LockBit	0.73	0.92	0.81	1.00	0.83	0.91	1.00	0.83	0.91
	Accuracy		0.88			0.95			0.95	
	Goodware	0.96	0.96	0.96	0.93	1.00	0.97	0.93	1.00	0.97
	MountLocker	0.50	0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00
	Accuracy		0.93			0.93			0.93	
	Goodware	0.95	0.95	0.95	0.50	1.00	0.67	0.51	0.91	0.66
	NetWalker	0.96	0.96	0.96	1.00	0.21	0.35	0.82	0.32	0.46
	Accuracy		0.96			0.56			0.58	
	Goodware	0.83	0.88	0.85	0.44	0.97	0.60	0.45	0.85	0.59
	Revil	0.91	0.88	0.90	0.89	0.16	0.28	0.75	0.31	0.43
	Accuracy		0.88			0.49			0.52	
	Goodware	0.92	0.43	0.59	0.85	1.00	0.92	0.88	1.00	0.93
	Ryuk	0.43	0.92	0.59	1.00	0.62	0.76	1.00	0.69	0.82
	Accuracy		0.77			0.88			0.90	
SVM	Clop	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.93	1.00	0.96	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.93			1.00			1.00	
	Conti	0.97	0.87	0.92	0.97	0.87	0.92	0.97	0.82	0.89
	Goodware	0.8	0.95	0.87	0.8	0.95	0.87	0.74	0.95	0.83
	Accuracy		0.9			0.9			0.86	
	Egregor	1.00	1.00	1.00	1.00	0.91	0.95	1.00	0.91	0.95
	Goodware	1.00	1.00	1.00	0.97	1.00	0.98	0.97	1.00	0.98
	Accuracy		1.00			0.97			0.97	
	Goodware	1.00	0.54	0.70	1.00	1.00	1.00	1.00	1.00	1.00
	LockBit	0.48	1.00	0.65	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.68			1.00			1.00	
	Goodware	0.96	0.93	0.95	0.93	1.00	0.97	0.93	1.00	0.97
	MountLocker	0.33	0.50	0.40	0.00	0.00	0.00	0.00	0.00	0.00
	Accuracy		0.9			0.93			0.93	
	Goodware	1.00	0.45	0.62	0.76	1.00	0.86	0.77	0.91	0.83
	NetWalker	0.70	1.00	0.82	1.00	0.75	0.86	0.92	0.79	0.85
	Accuracy		0.76			0.86			0.84	
	Goodware	1.00	0.39	0.57	0.56	0.97	0.71	0.95	0.64	0.76
	Revil	0.71	1.00	0.83	0.96	0.49	0.65	0.80	0.98	0.88
	Accuracy		0.76			0.68			0.84	
	Goodware	0.88	1.00	0.93	0.85	1.00	0.92	0.85	1.00	0.92
	Ryuk	1.00	0.69	0.82	1.00	0.62	0.76	1.00	0.62	0.76
	Accuracy		0.59			0.88			0.88	
NB	Clop	0.67	1.00	0.8	0.67	1.00	0.8	1.00	1.00	1.00
	Goodware	1.00	0.96	0.98	1.00	0.96	0.98	1.00	1.00	1.00
	Accuracy		0.97			0.97			1.00	
	Conti	0.96	0.71	0.82	0.95	0.55	0.7	0.94	0.87	0.9
	Goodware	0.65	0.95	0.77	0.54	0.95	0.69	0.79	0.9	0.84
	Accuracy		0.8			0.69			0.88	
	Egregor	0.71	0.91	0.8	0.71	0.91	0.8	1.00	0.91	0.95
	Goodware	0.96	0.86	0.91	0.96	0.86	0.91	0.97	1.00	0.98
	Accuracy		0.87			0.87			0.97	
	Goodware	0.96	0.96	0.96	0.96	0.89	0.93	0.96	0.93	0.95
	LockBit	0.92	0.92	0.92	0.79	0.92	0.85	0.85	0.92	0.88
	Accuracy		0.95			0.9			0.93	
	Goodware	0.96	0.93	0.95	0.95	0.68	0.79	0.96	0.82	0.88
	MountLocker	0.33	0.50	0.40	0.10	0.50	0.17	0.17	0.50	0.25
	Accuracy		0.9			0.67			0.8	
	Goodware	0.85	1.00	0.92	0.94	0.77	0.85	0.83	0.86	0.84
	NetWalker	1.00	0.86	0.92	0.84	0.96	0.90	0.89	0.86	0.87
	Accuracy		0.92			0.88			0.86	
	Goodware	0.85	1.00	0.92	0.56	0.97	0.71	0.97	0.91	0.94
	Revil	1.00	0.88	0.93	0.96	0.49	0.65	0.94	0.98	0.96
	Accuracy		0.93			0.93			0.95	
	Goodware	0.88	1.00	0.93	0.88	1.00	0.93	0.88	1.00	0.93
	Ryuk	1.00	0.69	0.82	1.00	0.69	0.82	1.00	0.69	0.82
	Accuracy		0.9			0.9			0.9	
DT	Clop	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Conti	0.95	0.92	0.93	0.95	0.92	0.93	0.95	0.92	0.93
	Goodware	0.86	0.9	0.88	0.86	0.9	0.88	0.86	0.9	0.88
	Accuracy		0.92			0.92			0.92	
	Egregor	1.00	0.91	0.95	1.00	0.91	0.95	1.00	0.91	0.95
	Goodware	0.97	1.00	0.98	0.97	1.00	0.98	0.97	1.00	0.98
	Accuracy		0.97			0.97			0.97	
	Goodware	1.00	0.93	0.96	1.00	0.93	0.96	1.00	0.93	0.96
	LockBit	0.86	1.00	0.92	0.86	1.00	0.92	0.86	1.00	0.92
	Accuracy		0.95			0.95			0.95	

Continua na próxima página

Tabela 15 – continuação da página anterior

	Malware	Test Size 0,33								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Goodware	0.96	0.89	0.93	0.96	0.89	0.93	0.96	0.89	0.93
	MountLocker	0.25	0.50	0.33	0.25	0.50	0.33	0.25	0.50	0.33
	Accuracy		0.87			0.87			0.87	
	Goodware	0.84	0.95	0.89	0.84	0.95	0.89	0.84	0.95	0.89
	NetWalker	0.96	0.86	0.91	0.96	0.86	0.91	0.96	0.86	0.91
	Accuracy		0.9			0.9			0.9	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Revil	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	0.85	1.00	0.92	0.85	1.00	0.92	0.85	1.00	0.92
	Ryuk	1.00	0.62	0.76	1.00	0.62	0.76	1.00	0.62	0.76
	Accuracy		0.88			0.88			0.88	
RF	Clop	1.00	1.00	1.00	1.00	1.00	1.00	0.5	0.5	0.5
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.96	0.96
	Accuracy		1.00			1.00			0.93	
	Conti	0.97	0.89	0.93	0.97	0.87	0.92	0.88	0.95	0.91
	Goodware	0.83	0.95	0.89	0.8	0.95	0.87	0.89	0.76	0.82
	Accuracy		0.92			0.9			0.88	
	Egregor	1.00	0.91	0.95	1.00	0.91	0.95	1.00	0.91	0.95
	Goodware	0.97	1.00	0.98	0.97	1.00	0.98	0.97	1.00	0.98
	Accuracy		0.97			0.97				
	Goodware	0.97	1.00	0.98	1.00	1.00	1.00	0.93	0.96	0.95
	LockBit	1.00	0.92	0.96	1.00	1.00	1.00	0.91	0.83	0.87
	Accuracy		0.97			1.00			0.93	
	Goodware	0.96	0.89	0.93	0.96	0.89	0.93	0.96	0.93	0.95
	MountLocker	0.25	0.50	0.33	0.25	0.50	0.33	0.33	0.50	0.40
	Accuracy		0.87			0.87			0.9	
	Goodware	0.85	1.00	0.92	0.84	0.95	0.89	0.82	0.82	0.82
	NetWalker	1.00	0.86	0.92	0.96	0.86	0.91	0.86	0.86	0.86
	Accuracy		0.92			0.9			0.84	
	Goodware	0.97	0.97	0.97	0.94	1.00	0.97	1.00	0.79	0.88
	Revil	0.98	0.98	0.98	1.00	0.96	0.98	0.88	1.00	0.93
	Accuracy		0.98			0.98			0.91	
	Goodware	0.88	1.00	0.93	0.85	1.00	0.92	0.85	1.00	0.92
	Ryuk	1.00	0.69	0.82	1.00	0.62	0.76	1.00	0.62	0.76
	Accuracy		0.9			0.88			0.88	
MLP	Clop	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.93	1.00	0.96	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.93			1.00			1.00	
	Conti	0.64	1.00	0.78	1.00	0.24	0.38	0.86	0.84	0.85
	Goodware	0.00	0.00	0.00	0.42	1.00	0.59	0.73	0.76	0.74
	Accuracy		0.64			0.51			0.81	
	Egregor	0.00	0.00	0.00	1.00	0.73	0.84	0.73	1.00	0.85
	Goodware	0.72	1.00	0.84	0.9	1.00	0.95	1.00	0.86	0.92
	Accuracy		0.72			0.92			0.9	
	Goodware	0.70	1.00	0.82	1.00	0.11	0.19	0.90	0.96	0.93
	LockBit	0.00	0.00	0.00	0.32	1.00	0.49	0.90	0.75	0.82
	Accuracy		0.7			0.38			0.9	
	Goodware	0.93	1.00	0.97	0.96	0.89	0.93	0.92	0.79	0.85
	MountLocker	0.00	0.00	0.00	0.25	0.50	0.33	0.00	0.00	0.00
	Accuracy		0.93			0.87			0.73	
	Goodware	0.94	0.68	0.79	0.69	1.00	0.81	0.85	0.77	0.81
	NetWalker	0.79	0.96	0.87	1.00	0.64	0.78	0.83	0.89	0.86
	Accuracy		0.84			0.8			0.84	
	Goodware	0.00	0.00	0.00	0.65	0.97	0.78	0.84	0.79	0.81
	Revil	0.60	1.00	0.75	0.97	0.65	0.78	0.86	0.90	0.88
	Accuracy		0.6			0.78			0.85	
	Goodware	0.68	1.00	0.81	0.85	1.00	0.92	0.83	0.71	0.77
	Ryuk	0.00	0.00	0.00	1.00	0.62	0.76	0.53	0.69	0.60
	Accuracy		0.68			0.88			0.71	

Tabela 16: Tabela com os dados das classificações referente a abordagem de TF-IDF (Behavior), com *test size* 0,5 e classificação Binária.

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Clop	1.00	1.00	1.00	1.00	0.67	0.8	1.00	0.67	0.8
	Goodware	1.00	1.00	1.00	0.98	1.00	0.99	0.98	1.00	0.99
	Accuracy		1.00			0.98			0.98	
	Conti	0.96	0.87	0.91	0.97	0.55	0.7	0.97	0.67	0.8
	Goodware	0.82	0.94	0.87	0.56	0.97	0.71	0.64	0.97	0.77
	Accuracy		0.9			0.7			0.78	
	Egregor	1.00	0.05	0.10	0.00	0.00	0.00	0.00	0.00	0.00
	Goodware	0.69	1.00	0.82	0.68	1.00	0.81	0.68	1.00	0.81
	Accuracy		0.69			0.68			0.68	
	Goodware	0.97	0.85	0.91	0.93	1.00	0.96	0.93	1.00	0.96

Continua na próxima página

KNN

Tabela 16 – continuação da página anterior

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	LockBit	0.76	0.95	0.84	1.00	0.85	0.92	1.00	0.85	0.92
	Accuracy		0.88			0.95			0.95	
	Goodware	0.95	0.95	0.95	0.93	1.00	0.96	0.93	1.00	0.96
	MountLocker	0.50	0.50	0.50	1.00	0.25	0.40	1.00	0.25	0.40
	Accuracy		0.91			0.93			0.93	
	Goodware	0.90	1.00	0.95	0.55	1.00	0.71	0.55	0.86	0.67
	NetWalker	1.00	0.89	0.94	1.00	0.21	0.35	0.71	0.32	0.44
	Accuracy		0.95			0.6			0.59	
	Goodware	0.83	0.88	0.85	0.44	0.97	0.60	0.45	0.85	0.59
	Revil	0.91	0.88	0.90	0.89	0.16	0.28	0.75	0.31	0.43
	Accuracy		0.88			0.49			0.52	
	Goodware	0.88	0.88	0.88	0.85	0.97	0.91	0.85	0.97	0.91
	Ryuk	0.77	0.77	0.77	0.94	0.68	0.79	0.94	0.68	0.79
	Accuracy		0.84			0.87			0.87	
SVM	Clop	0.00	0.00	0.00	1.00	0.67	0.8	1.00	0.67	0.8
	Goodware	0.93	1.00	0.96	0.98	1.00	0.99	0.98	1.00	0.99
	Accuracy		0.93			0.98			0.98	
	Conti	0.96	0.87	0.91	0.98	0.84	0.9	0.98	0.82	0.89
	Goodware	0.82	0.94	0.87	0.78	0.97	0.86	0.76	0.97	0.85
	Accuracy		0.89			0.89			0.88	
	Egregor	1.00	1.00	1.00	1.00	0.89	0.94	1.00	0.89	0.94
	Goodware	1.00	1.00	1.00	0.95	1.00	0.98	0.95	1.00	0.98
	Accuracy		1.00			0.97			0.97	
	Goodware	1.00	0.50	0.67	0.98	1.00	0.99	0.98	1.00	0.99
	LockBit	0.50	1.00	0.67	1.00	0.95	0.97	1.00	0.95	0.97
	Accuracy		0.67			0.98			0.98	
	Goodware	0.95	0.95	0.95	0.93	1.00	0.96	0.93	1.00	0.96
	MountLocker	0.50	0.50	0.50	1.00	0.25	0.40	1.00	0.25	0.40
	Accuracy		0.91			0.93			0.93	
	Goodware	1.00	0.51	0.68	0.80	1.00	0.89	0.82	0.86	0.84
	NetWalker	0.68	1.00	0.81	1.00	0.76	0.87	0.86	0.82	0.84
	Accuracy		0.76			0.88			0.84	
	Goodware	1.00	0.39	0.57	0.56	0.97	0.71	0.95	0.64	0.76
	Revil	0.71	1.00	0.83	0.96	0.49	0.65	0.80	0.98	0.88
	Accuracy		0.76			0.68			0.84	
	Goodware	0.89	0.40	0.55	0.85	0.97	0.91	0.85	0.97	0.91
	Ryuk	0.45	0.91	0.61	0.94	0.68	0.79	0.94	0.68	0.79
	Accuracy		0.58			0.87			0.87	
NB	Clop	0.75	1.00	0.86	0.6	1.00	0.75	0.6	1.00	0.75
	Goodware	1.00	0.98	0.99	1.00	0.95	0.97	1.00	0.95	0.97
	Accuracy		0.98			0.95			0.95	
	Conti	0.97	0.67	0.8	0.97	0.51	0.67	0.96	0.87	0.91
	Goodware	0.64	0.97	0.77	0.54	0.97	0.7	0.82	0.94	0.87
	Accuracy		0.78			0.68			0.9	
	Egregor	0.71	0.89	0.79	0.71	0.89	0.79	1.00	0.89	0.94
	Goodware	0.94	0.82	0.88	0.94	0.82	0.88	0.95	1.00	0.98
	Accuracy		0.85			0.85			0.97	
	Goodware	0.97	0.90	0.94	0.97	0.88	0.92	0.95	0.93	0.94
	LockBit	0.83	0.95	0.88	0.79	0.95	0.86	0.86	0.90	0.88
	Accuracy		0.92			0.9			0.92	
	Goodware	0.95	0.95	0.95	0.97	0.78	0.86	0.97	0.80	0.88
	MountLocker	0.50	0.50	0.50	0.25	0.75	0.38	0.27	0.75	0.40
	Accuracy		0.91			0.78			0.8	
	Goodware	0.88	1.00	0.94	0.80	1.00	0.89	0.86	0.81	0.83
	NetWalker	1.00	0.87	0.93	1.00	0.76	0.87	0.82	0.87	0.85
	Accuracy		0.93			0.85			0.84	
	Goodware	0.85	1.00	0.92	0.89	0.94	0.91	0.97	0.91	0.94
	Revil	1.00	0.88	0.93	0.96	0.92	0.94	0.94	0.98	0.96
	Accuracy		0.93			0.93			0.93	
	Goodware	0.85	0.97	0.91	0.87	0.97	0.92	0.86	0.93	0.89
	Ryuk	0.94	0.68	0.79	0.94	0.73	0.82	0.84	0.73	0.78
	Accuracy		0.87			0.89			0.85	
DT	Clop	0.5	1.00	0.67	0.5	1.00	0.67	0.5	1.00	0.67
	Goodware	1.00	0.93	0.96	1.00	0.93	0.96	1.00	0.93	0.96
	Accuracy		0.93			0.93			0.93	
	Conti	0.96	0.87	0.91	0.96	0.87	0.91	0.96	0.87	0.91
	Goodware	0.82	0.94	0.87	0.82	0.94	0.87	0.82	0.94	0.87
	Accuracy		0.9			0.9			0.9	
	Egregor	1.00	0.95	0.97	1.00	0.95	0.97	1.00	0.95	0.97
	Goodware	0.98	1.00	0.99	0.98	1.00	0.99	0.98	1.00	0.99
	Accuracy		0.98			0.98			0.98	
	Goodware	0.95	0.97	0.96	0.95	0.97	0.96	0.95	0.97	0.96
	LockBit	0.95	0.90	0.92	0.95	0.90	0.92	0.95	0.90	0.92
	Accuracy		0.95			0.95			0.95	
	Goodware	0.97	0.93	0.95	0.97	0.93	0.95	0.97	0.93	0.95
	MountLocker	0.50	0.75	0.60	0.50	0.75	0.60	0.50	0.75	0.60
	Accuracy		0.91			0.91			0.91	
	Goodware	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89
	NetWalker	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89
	Accuracy		0.89			0.89			0.89	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Revil	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Continua na próxima página

Tabela 16 – continuação da página anterior

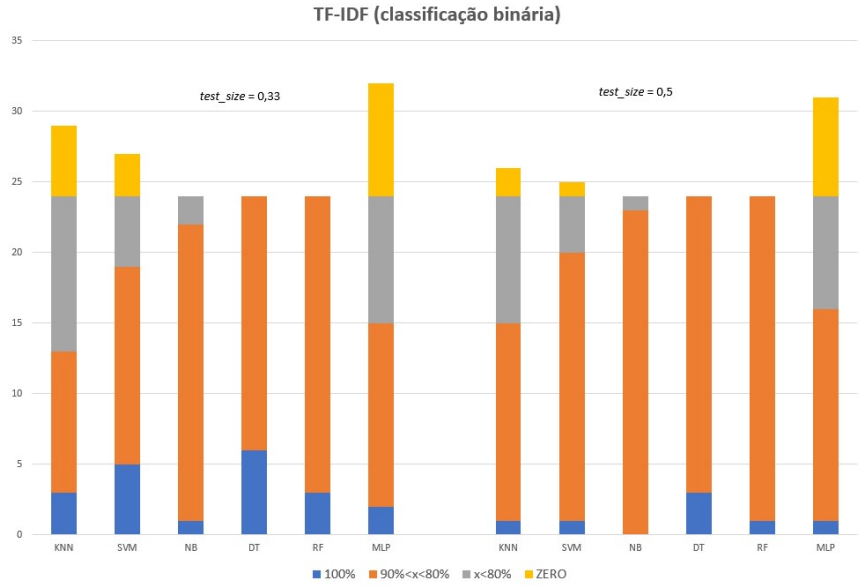
	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Accuracy		1.00			1.00			1.00	
	Goodware	0.85	1.00	0.92	0.85	1.00	0.92	0.85	1.00	0.92
	Ryuk	1.00	0.68	0.81	1.00	0.68	0.81	1.00	0.68	0.81
	Accuracy		0.89			0.89			0.89	
RF	Clop	1.00	0.67	0.8	1.00	1.00	1.00	0.67	0.67	0.8
	Goodware	0.98	1.00	0.99	1.00	1.00	1.00	0.98	1.00	0.99
	Accuracy		0.98			1.00			0.98	
	Conti	0.98	0.89	0.93	0.98	0.87	0.92	0.94	0.91	0.93
	Goodware	0.84	0.97	0.9	0.82	0.97	0.89	0.86	0.91	0.88
	Accuracy		0.92			0.91			0.91	
	Egregor	1.00	0.84	0.91	1.00	0.95	0.97	1.00	0.89	0.94
	Goodware	0.93	1.00	0.96	0.98	1.00	0.99	0.95	1.00	0.98
	Accuracy		0.95			0.98			0.97	
	Goodware	0.97	0.95	0.96	0.97	0.95	0.96	0.98	1.00	0.99
	LockBit	0.90	0.95	0.93	0.90	0.95	0.93	1.00	0.95	0.97
	Accuracy		0.97			0.95			0.98	
	Goodware	0.95	0.93	0.94	0.95	0.95	0.95	0.95	0.95	0.95
	MountLocker	0.40	0.50	0.44	0.50	0.50	0.50	0.50	0.50	0.50
	Accuracy		0.89			0.91			0.91	
	Goodware	0.86	0.97	0.91	0.88	1.00	0.94	0.86	0.81	0.83
	NetWalker	0.97	0.84	0.90	1.00	0.87	0.93	0.82	0.87	0.85
	Accuracy		0.91			0.93			0.84	
	Goodware	0.94	1.00	0.97	0.97	1.00	0.9	0.93	0.79	0.85
	Revil	1.00	0.96	0.98	1.00	0.98	0.99	0.87	0.96	0.91
	Accuracy		0.98			0.99			0.89	
	Goodware	0.83	1.00	0.91	0.83	0.97	0.90	0.84	0.90	0.87
	Ryuk	1.00	0.64	0.78	0.93	0.64	0.76	0.79	0.68	0.73
	Accuracy		0.87			0.85			0.82	
MLP	Clop	0.00	0.00	0.00	1.00	1.00	1.00	1.00	0.67	0.8
	Goodware	0.93	1.00	0.96	1.00	1.00	1.00	0.98	1.00	0.99
	Accuracy		0.93			1.00			0.98	
	Conti	0.62	1.00	0.77	0.98	0.82	0.89	0.93	0.73	0.82
	Goodware	0.00	0.00	0.00	0.76	0.97	0.85	0.67	0.91	0.77
	Accuracy		0.62			0.88			0.8	
	Egregor	0.00	0.00	0.00	0.39	1.00	0.56	0.77	0.89	0.83
	Goodware	0.68	1.00	0.81	1.00	0.25	0.4	0.95	0.88	0.91
	Accuracy		0.68			0.49			0.88	
	Goodware	0.67	1.00	0.80	1.00	0.07	0.14	0.89	0.97	0.93
	LockBit	0.00	0.00	0.00	0.35	1.00	0.52	0.94	0.75	0.83
	Accuracy		0.67			0.38			0.90	
	Goodware	0.91	1.00	0.95	0.95	0.95	0.95	0.92	0.88	0.90
	MountLocker	0.00	0.00	0.00	0.50	0.50	0.50	0.17	0.25	0.20
	Accuracy		0.91			0.93			0.82	
	Goodware	0.00	0.00	0.00	0.76	1.00	0.86	0.87	0.73	0.79
	NetWalker	0.51	1.00	0.67	1.00	0.68	0.81	0.77	0.89	0.83
	Accuracy		0.51			0.84			0.81	
	Goodware	0.00	0.00	0.00	0.65	0.97	0.78	0.86	0.76	0.81
	Revil	0.60	1.00	0.75	0.97	0.65	0.78	0.85	0.92	0.88
	Accuracy		0.60			0.78			0.85	
	Goodware	0.66	1.00	0.79	1.00	0.05	0.10	0.83	0.75	0.79
	Ryuk	1.00	0.05	0.09	0.37	1.00	0.54	0.62	0.73	0.67
	Accuracy		0.66			0.39			0.74	

A Figura 9, de maneira semelhante ao que fizemos na Tabela 8, apresenta a sumarização dos dados da classificação das Tabelas 15 e 16:

Podemos observar que os classificadores apresentam desempenhos distintos em cada *test size* escolhido:

- O KNN e a SVM tiveram suas quantidades de classificação na faixa laranja significativamente reduzidas e o MLP teve uma sensível redução nessa faixa;
- O NB passou a ter mais classificações na região laranja e algumas na região cinza e nenhuma na região azul, apresentando uma queda no desempenho da classificação;

Figura 9: Sumarização dos resultados das Tabelas 15 e 16.



- Os classificadores DT e RF tiveram uma redução nas classificações na faixa azul em prol da faixa laranja, mesmo com a mudança de parâmetros, esses classificadores não apresentaram faixa cinza;
- O NB, apesar de ter reduzido as classificações na faixa cinza, também apresentou redução na faixa azul, o que também significa redução no desempenho;
- Nessa configuração, os mais indicados para uso em detecção de *ransomwares* são as DT e RF para qualquer tamanho de *test size*.

6.4.2.2 Seção *Memory*

As classificações realizadas sobre os dados minerados da seção *Memory* dos relatórios de análise são apresentados nas Tabelas 17, 18, 19 e 20. Da mesma maneira que foi feito na classificação dos conjuntos de dados de chamadas de API, as tabelas desta seção estão divididas em classificação multiclasse e binária. Além disso, seguindo o modelo das classificações anteriores, usamos *test size* 1/2 e 1/3.

O conjunto de dados com *test size* 1/3 e classificação multiclasse (Tabela 17) apresentou bom desempenho: apenas 4 classificações com valores de *Accuracy* entre 0.84 e 0.86 e a maioria em 0.95 ou muito próximo deste valor. Para o mesmo conjunto de dados com *test size* 1/2 e multiclasse (Tabela 18), o MLP atingiu classificação 0.72 e o SVM atingiu classificação 0.75, ambos sem otimização no conjunto de dados. No primeiro caso, tivemos o *Clop* e o *MountLocker* com zero nas três métricas e no segundo caso, *Clop*, *Egregor*,

LockBit, *MountLocker* e *Ryuk* com métricas zero para o conjunto sem otimização. O MLP otimizado com PCA apresentou métricas zero para o *Clop*.

Curiosamente, o NB conseguiu pontuação satisfatória em todas as situações propostas: em nenhuma apresentou classificação zero e em várias apresentou *Accuracy* 0.95. No que se refere a classificação binária (Tabelas 19 e 19), tanto para *test size* 1/3 quanto 1/2, apenas duas classificações abaixo de 91%.

Na classificação binária, com *test size* 1/3, a menor classificação ficou com *Accuracy* 0.44, que ocorreu no MLP e na SVM, classificando o *NetWalker* (embora no SVM tenha ocorrido apenas com o *LockBit*). Os outros resultados de *Accuracy* abaixo de 0.88 ficaram concentrados em sua maioria por volta de 0.76, na classificação pelo MLP. De um total de 144 classificações, 58 conseguiram *Accuracy* 1.00, demonstrando que o conjunto de dados gerados a partir da seção *Memory* se mostra um ótimo conjunto de dados para compor um sistema de detecção de *ransomwares* em produção.

Com *test size* 1/2 na classificação binária, as menores classificações ficaram por conta do MLP sem otimização (4 delas em torno de 0.65 e uma 0.51), inclusive todas as métricas zero. As únicas ocorrências de *Accuracy* zero fora desse classificador foram as métricas de classificação do *Clop* na SVM, que mesmo assim teve *Accuracy* 0.93 devido a quantidade de *Goodware* ser muito maior que as amostras daquela família. Houve também algumas ocorrências de classificação perfeita, que colorimos de azul claro na Tabela 20.

Tabela 17: Tabela com os dados das classificações referente a abordagem de TF-IDF (Memory), com *test size* 0,33 e classificação multiclasse.

	Malware	Test Size 0,33								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
KNN	Clop	1.00	0.71	0.83	1.00	0.57	0.73	0.86	0.86	0.86
	Conti	0.90	0.92	0.91	0.73	0.95	0.82	0.90	0.95	0.92
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.81	0.81	0.81	1.00	0.76	0.86	0.79	0.90	0.84
	LockBit	0.78	0.78	0.78	1.00	0.06	0.11	0.71	0.28	0.40
	MountLocker	1.00	0.75	0.86	1.00	0.50	0.67	0.50	0.50	0.50
	NetWalker	1.00	0.96	0.98	1.00	0.96	0.98	0.86	0.96	0.91
	Revil	0.96	0.98	0.97	0.97	0.89	0.93	0.94	0.97	0.95
	Ryuk	1.00	0.94	0.97	0.27	0.81	0.40	1.00	0.81	0.90
	Accuracy		0.94			0.84			0.91	
SVM	Clop	1.00	0.71	0.83	1.00	0.57	0.73	0.67	0.57	0.62
	Conti	0.90	0.92	0.91	0.97	0.97	0.97	0.90	0.95	0.92
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.81	0.81	0.81	1.00	0.76	0.86	0.86	0.90	0.88
	LockBit	0.78	0.78	0.78	1.00	0.17	0.29	0.73	0.44	0.55
	MountLocker	1.00	0.75	0.86	1.00	0.50	0.67	0.40	0.50	0.44
	NetWalker	1.00	0.96	0.98	1.00	0.96	0.98	0.77	0.96	0.86
	Revil	0.96	0.98	0.97	0.88	1.00	0.94	0.96	0.96	0.96
	Ryuk	1.00	0.94	0.97	1.00	0.81	0.90	1.00	0.81	0.90
	Accuracy		0.83			0.91			0.91	
NB	Clop	1.00	1.00	1.00	1.00	1.00	1.00	0.67	0.57	0.62
	Conti	1.00	0.97	0.99	1.00	0.92	0.96	0.84	0.95	0.89
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	0.76	0.86	1.00	0.76	0.86	0.70	0.90	0.79
	LockBit	0.93	0.78	0.85	1.00	0.72	0.84	0.31	0.72	0.43
	MountLocker	1.00	0.50	0.67	1.00	0.50	0.67	1.00	0.50	0.67
	NetWalker	0.96	0.96	0.96	1.00	0.96	0.98	0.96	0.88	0.92
	Revil	0.94	0.99	0.96	0.92	1.00	0.96	0.97	0.83	0.89
	Ryuk	0.88	0.94	0.91	1.00	0.94	0.97	1.00	0.94	0.97

Continua na próxima página

Tabela 17 – continuação da página anterior

	Malware	Test Size 0,33								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Accuracy		0.95			0.95			0.85	
DT	Clop	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Conti	0.97	0.95	0.96	0.95	0.95	0.95	0.95	0.95	0.95
	Egregor	1.00	1.00	1.00	0.86	1.00	0.92	0.86	1.00	0.92
	Goodware	0.82	0.86	0.84	0.77	0.81	0.79	0.77	0.81	0.79
	LockBit	1.00	0.78	0.88	0.83	0.83	0.83	0.83	0.83	0.83
	MountLocker	0.67	0.50	0.57	1.00	0.50	0.67	1.00	0.50	0.67
	NetWalker	0.89	0.96	0.92	0.92	0.96	0.94	0.92	0.96	0.94
	Revil	0.95	0.98	0.97	0.97	0.97	0.97	0.97	0.97	0.97
	Ryuk	1.00	0.81	0.90	0.93	0.81	0.87	0.93	0.81	0.87
	Accuracy		0.95			0.94			0.94	
RF	Clop	1.00	0.71	0.83	1.00	0.71	0.83	1.00	0.57	0.73
	Conti	1.00	0.97	0.99	0.97	0.97	0.97	0.97	0.95	0.96
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	0.76	0.86	1.00	0.76	0.86	1.00	0.90	0.95
	LockBit	0.89	0.89	0.89	0.89	0.89	0.89	0.93	0.78	0.85
	MountLocker	1.00	0.50	0.67	1.00	0.50	0.67	1.00	0.50	0.67
	NetWalker	0.96	0.96	0.96	0.96	0.96	0.96	0.92	0.96	0.94
	Revil	0.94	1.00	0.97	0.94	1.00	0.97	0.93	0.99	0.96
	Ryuk	1.00	0.81	0.90	1.00	0.81	0.90	1.00	0.81	0.90
	Accuracy		0.95			0.95			0.95	
MLP	Clop	1.00	0.71	0.83	0.20	0.57	0.30	0.00	0.00	0.00
	Conti	1.00	0.97	0.99	0.91	0.82	0.86	0.97	0.95	0.96
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
	Goodware	1.00	0.76	0.86	0.77	0.81	0.79	1.00	0.90	0.95
	LockBit	0.89	0.89	0.89	1.00	0.11	0.20	0.60	0.17	0.26
	MountLocker	1.00	0.50	0.67	1.00	0.25	0.40	0.67	0.50	0.57
	NetWalker	0.96	0.96	0.96	1.00	0.96	0.98	0.75	0.96	0.84
	Revil	0.94	1.00	0.97	0.99	0.89	0.94	0.85	0.98	0.91
	Ryuk	1.00	0.81	0.90	0.33	1.00	0.50	0.93	0.81	0.87
	Accuracy		0.95			0.84			0.86	

Tabela 18: Tabela com os dados das classificações referente a abordagem de TF-IDF (Memory), com *test size* 0,5 e classificação multiclasse.

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
KNN	Clop	1.00	0.70	0.82	1.00	0.60	0.75	1.00	0.80	0.89
	Conti	0.88	0.95	0.91	0.74	0.93	0.83	0.83	0.97	0.89
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.78	0.81	0.79	1.00	0.77	0.87	0.87	0.87	0.87
	LockBit	0.85	0.71	0.77	1.00	0.04	0.08	0.38	0.25	0.30
	MountLocker	1.00	0.89	0.94	1.00	0.78	0.88	0.50	0.78	0.61
	NetWalker	0.98	0.98	0.98	1.00	0.98	0.99	0.90	0.90	0.90
	Revil	0.95	0.97	0.96	0.98	0.89	0.94	0.96	0.96	0.96
	Ryuk	1.00	0.87	0.93	0.28	0.87	0.42	1.00	0.87	0.93
	Accuracy		0.94			0.85			0.91	
SVM	Clop	0.00	0.00	0.00	1.00	0.60	0.75	1.00	0.60	0.75
	Conti	1.00	0.03	0.06	0.94	0.98	0.96	1.00	0.92	0.96
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.79	0.84	0.81	1.00	0.77	0.87	1.00	0.77	0.87
	LockBit	0.93	0.58	0.72	1.00	0.04	0.08	1.00	0.04	0.08
	MountLocker	0.00	0.00	0.00	1.00	0.78	0.88	1.00	0.78	0.88
	NetWalker	1.00	0.57	0.73	1.00	0.98	0.99	1.00	0.90	0.95
	Revil	0.71	0.99	0.83	0.89	1.00	0.94	0.87	1.00	0.93
	Ryuk	1.00	0.09	0.16	1.00	0.87	0.93	1.00	0.87	0.93
	Accuracy		0.75			0.92			0.91	
NB	Clop	0.82	0.90	0.86	1.00	1.00	1.00	1.00	0.60	0.75
	Conti	1.00	0.95	0.97	1.00	0.92	0.96	0.84	0.97	0.90
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	0.77	0.87	1.00	0.77	0.87	0.71	0.94	0.81
	LockBit	0.90	0.75	0.82	0.94	0.67	0.78	0.36	0.75	0.49
	MountLocker	1.00	0.78	0.88	1.00	0.78	0.88	0.78	0.78	0.78
	NetWalker	0.98	0.98	0.98	1.00	0.90	0.95	0.98	0.95	0.96
	Revil	0.95	0.99	0.97	0.92	1.00	0.96	0.98	0.86	0.91
	Ryuk	0.92	0.96	0.94	1.00	0.96	0.98	1.00	0.96	0.98
	Accuracy		0.96			0.95			0.88	
DT	Clop	0.73	0.80	0.76	0.73	0.80	0.76	0.73	0.80	0.76
	Conti	0.87	0.98	0.92	0.87	0.98	0.92	0.87	0.98	0.92
	Egregor	0.95	1.00	0.97	0.95	1.00	0.97	0.95	1.00	0.97
	Goodware	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81
	LockBit	0.76	0.67	0.71	0.76	0.67	0.71	0.76	0.67	0.71
	MountLocker	0.70	0.78	0.74	0.70	0.78	0.74	0.70	0.78	0.74
	NetWalker	1.00	0.83	0.91	1.00	0.83	0.91	1.00	0.83	0.91
	Revil	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
	Ryuk	1.00	0.96	0.98	1.00	0.96	0.98	1.00	0.96	0.98
	Accuracy		0.93			0.93			0.93	

Continua na próxima página

Tabela 18 – continuação da página anterior

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
RF	Clop	1.00	0.60	0.75	1.00	0.60	0.75	1.00	0.80	0.89
	Conti	0.97	0.98	0.98	0.95	0.98	0.97	0.98	0.97	0.98
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.89	0.77	0.83	0.92	0.77	0.84	1.00	0.81	0.89
	LockBit	0.87	0.83	0.85	0.95	0.83	0.89	0.86	0.79	0.83
	MountLocker	1.00	0.78	0.88	1.00	0.78	0.88	1.00	0.78	0.88
	NetWalker	0.93	0.98	0.95	1.00	0.98	0.99	0.95	0.98	0.96
	Revil	0.96	0.99	0.98	0.94	0.99	0.97	0.95	0.99	0.97
	Ryuk	1.00	0.96	0.98	1.00	0.87	0.93	1.00	0.87	0.93
	Accuracy		0.96			0.95			0.95	
MLP	Clop	0.00	0.00	0.00	0.29	0.60	0.39	0.00	0.00	0.00
	Conti	0.43	0.54	0.48	0.96	0.85	0.90	0.78	0.11	0.20
	Egregor	0.00	0.00	0.00	0.78	1.00	0.88	1.00	1.00	1.00
	Goodware	0.47	0.77	0.59	0.78	0.90	0.84	0.93	0.84	0.88
	LockBit	0.00	0.00	0.00	0.58	0.29	0.39	0.50	0.29	0.37
	MountLocker	0.00	0.00	0.00	1.00	0.56	0.71	0.70	0.78	0.74
	NetWalker	0.66	0.45	0.54	1.00	0.81	0.89	0.90	0.90	0.90
	Revil	0.82	0.98	0.89	1.00	0.89	0.94	0.79	0.98	0.87
	Ryuk	0.00	0.00	0.00	0.35	1.00	0.52	0.95	0.87	0.91
	Accuracy		0.72			0.85			0.81	

Tabela 19: Tabela com os dados das classificações referente a abordagem de TF-IDF (Memory), com *test size* 0,33 e classificação Binária.

	Malware	Test Size 0,33								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
KNN	Clop	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Conti	0.95	0.97	0.96	0.95	0.97	0.96	0.95	0.97	0.96
	Goodware	0.95	0.90	0.93	0.95	0.90	0.93	0.95	0.90	0.93
	Accuracy		0.95			0.95			0.95	
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	1.00	0.93	0.96	0.76	1.00	0.86	0.76	1.00	0.86
	LockBit	0.86	1.00	0.92	1.00	0.25	0.40	1.00	0.25	0.40
	Accuracy		0.95			0.78			0.78	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MountLocker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	NetWalker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	0.86	1.00	0.93	0.96	0.96	0.96	0.78	1.00	0.88
	Revil	1.00	0.98	0.99	1.00	1.00	1.00	1.00	0.97	0.98
	Accuracy		0.98			0.99			0.97	
	Goodware	1.00	0.96	0.98	0.96	0.93	0.95	0.96	0.93	0.95
	Ryuk	0.93	1.00	0.96	0.86	0.92	0.89	0.86	0.92	0.89
	Accuracy		9.98			0.93			0.93	
SVM	Clop	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.93	1.00	0.96	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.93			1.00			1.00	
	Conti	1.00	0.87	0.93	0.95	1.00	0.97	0.95	1.00	0.97
	Goodware	0.81	1.00	0.89	1.00	0.90	0.95	1.00	0.90	0.95
	Accuracy		0.92			0.97			0.97	
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	0.79	0.93	0.85	1.00	0.89	0.94	1.00	0.86	0.92
	LockBit	0.71	0.42	0.53	0.80	1.00	0.89	0.75	1.00	0.86
	Accuracy		0.78			0.93			0.9	
	Goodware	0.97	1.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00
	MountLocker	1.00	0.50	0.67	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.97			1.00			1.00	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	NetWalker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	1.00	0.76	0.86	1.00	0.96	0.98	0.92	0.96	0.94
	Revil	0.97	1.00	0.99	1.00	1.00	1.00	1.00	0.99	0.99
	Accuracy		0.97			1.00			0.99	
	Goodware	0.97	1.00	0.98	0.96	0.93	0.95	0.97	1.00	0.98
	Ryuk	1.00	0.92	0.96	0.86	0.92	0.89	1.00	0.92	0.96
	Accuracy		0.98			0.93			0.98	
	Clop	0.67	1.00	0.80	0.67	1.00	0.80	1.00	1.00	1.00
	Goodware	1.00	0.96	0.98	1.00	0.96	0.98	1.00	1.00	1.00

Continua na próxima página

Tabela 19 – continuação da página anterior

	Malware	Test Size 0,33								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Accuracy		0.97			0.97			1.00	
	Conti	0.97	1.00	0.99	0.90	1.00	0.95	0.97	0.97	0.97
	Goodware	1.00	0.95	0.98	1.00	0.81	0.89	0.95	0.95	0.95
	Accuracy		0.98			0.93			0.97	
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	1.00	0.86	0.92	1.00	0.86	0.92	0.96	0.93	0.95
	LockBit	0.75	1.00	0.86	0.75	1.00	0.86	0.85	0.92	0.88
	Accuracy		0.9			0.9			0.93	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.98
	MountLocker	1.00	1.00	1.00	1.00	1.00	1.00	0.67	1.00	0.80
	Accuracy		1.00			1.00			0.97	
	Goodware	1.00	1.00	1.00	1.00	0.95	0.98	0.91	0.95	0.93
	NetWalker	1.00	1.00	1.00	0.97	1.00	0.98	0.96	0.93	0.95
	Accuracy		1.00			0.98			0.94	
	Goodware	1.00	0.92	0.96	1.00	0.92	0.96	0.40	0.96	0.56
	Revil	0.99	1.00	1.00	0.99	1.00	1.00	0.99	0.83	0.90
	Accuracy		0.99			0.99			0.84	
	Goodware	1.00	0.93	0.96	1.00	0.93	0.96	1.00	1.00	1.00
	Ryuk	0.87	1.00	0.93	0.87	1.00	0.93	1.00	1.00	1.00
	Accuracy		0.95			0.95			1.00	
DT	Clop	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Conti	0.93	0.97	0.95	0.93	0.97	0.95	0.93	0.97	0.95
	Goodware	0.95	0.86	0.90	0.95	0.86	0.90	0.95	0.86	0.90
	Accuracy		0.93			0.93			0.93	
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	1.00	0.96	0.98	1.00	0.96	0.98	1.00	0.96	0.98
	LockBit	0.92	1.00	0.96	0.92	1.00	0.96	0.92	1.00	0.96
	Accuracy		0.97			0.97			0.97	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MountLocker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	NetWalker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	1.00	0.96	0.98	1.00	0.96	0.98	1.00	0.96	0.98
	Revil	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	0.97	1.00	0.98	0.97	1.00	0.98	0.97	1.00	0.98
	Ryuk	1.00	0.92	0.96	1.00	0.92	0.96	1.00	0.92	0.96
	Accuracy		0.98			0.98			0.98	
RF	Clop	1.00	1.00	1.00	1.00	1.00	1.00	0.50	0.50	0.50
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.96	0.96
	Accuracy		1.00			1.00			0.93	
	Conti	0.95	1.00	0.97	0.95	0.97	0.96	0.95	0.95	0.95
	Goodware	1.00	0.90	0.95	0.95	0.90	0.93	0.90	0.90	0.90
	Accuracy		0.97			0.97			0.93	
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	0.79	1.00	0.88
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.89	0.94
	Accuracy		1.00			1.00			0.92	
	Goodware	1.00	0.89	0.94	1.00	0.89	0.94	1.00	0.93	0.96
	LockBit	0.80	1.00	0.89	0.80	1.00	0.89	0.86	1.00	0.92
	Accuracy		0.93			0.93			0.95	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MountLocker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.95	0.98
	NetWalker	1.00	1.00	1.00	1.00	1.00	1.00	0.97	1.00	0.98
	Accuracy		1.00			1.00			0.98	
	Goodware	1.00	0.96	0.98	0.96	0.96	0.96	1.00	0.96	0.98
	Revil	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	1.00	0.93	0.96	1.00	0.96	0.98	0.96	0.96	0.96
	Ryuk	0.87	1.00	0.93	0.93	1.00	0.96	0.92	0.92	0.92
	Accuracy		0.95			0.98			0.95	
MLP	Clop	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.93	1.00	0.96	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.93			1.00			1.00	
	Conti	0.64	1.00	0.78	0.93	0.97	0.95	0.90	1.00	0.95
	Goodware	0.00	0.00	0.00	0.95	0.86	0.90	1.00	0.81	0.89
	Accuracy		0.64			0.93			0.93	
	Egregor	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.72	1.00	0.84	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.72			1.00			1.00	
	Goodware	0.70	1.00	0.82	1.00	0.89	0.94	1.00	0.86	0.92
	LockBit	0.00	0.00	0.00	0.80	1.00	0.89	0.75	1.00	0.86
	Accuracy		0.70			0.93			0.9	

Continua na próxima página

Tabela 19 – continuação da página anterior

	Malware	Test Size 0,33								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Goodware	0.93	1.00	0.97	1.00	1.00	1.00	1.00	1.00	1.00
	MountLocker	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.93			1.00			1.00	
	Goodware	0.44	1.00	0.61	1.00	0.95	0.98	1.00	0.95	0.98
	NetWalker	0.00	0.00	0.00	0.97	1.00	0.98	0.97	1.00	0.98
	Accuracy		0.44			0.98			0.98	
	Goodware	0.00	0.00	0.00	0.75	0.96	0.84	0.83	0.96	0.89
	Revil	0.89	1.00	0.94	0.99	0.96	0.98	1.00	0.98	0.99
	Accuracy		0.89			0.96			0.97	
	Goodware	0.68	1.00	0.81	1.00	0.89	0.94	1.00	1.00	1.00
	Ryuk	0.00	0.00	0.00	0.81	1.00	0.90	1.00	1.00	1.00
	Accuracy		0.68			0.93			1.00	

Tabela 20: Tabela com os dados das classificações referente a abordagem de TF-IDF (Memory), com *test size* 0,5 e classificação Binária.

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
KNN	Clop	1.00	1.00	1.00	1.00	0.67	0.80	1.00	0.67	0.80
	Goodware	1.00	1.00	1.00	0.98	1.00	0.99	0.98	1.00	0.99
	Accuracy		1.00			0.98			0.98	
	Conti	0.96	0.98	0.97	1.00	0.96	0.98	1.00	0.96	0.98
	Goodware	0.97	0.94	0.95	0.94	1.00	0.97	0.94	1.00	0.97
	Accuracy		0.97			0.98			0.98	
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	1.00	0.95	0.97	0.71	1.00	0.83	0.71	1.00	0.83
	LockBit	0.91	1.00	0.95	1.00	0.20	0.33	1.00	0.20	0.33
	Accuracy		0.97			0.73			0.73	
	Goodware	1.00	1.00	1.00	1.00	0.95	0.97	1.00	0.95	0.97
	MountLocker	1.00	1.00	1.00	0.67	1.00	0.80	0.67	1.00	0.80
	Accuracy		1.00			0.96			0.96	
	Goodware	1.00	1.00	1.00	1.00	0.97	0.99	1.00	0.97	0.99
	NetWalker	1.00	1.00	1.00	0.97	1.00	0.99	0.97	1.00	0.99
	Accuracy		1.00			0.99			0.99	
	Goodware	0.73	0.92	0.81	0.94	0.86	0.90	0.94	0.86	0.90
	Revil	0.99	0.96	0.98	0.98	0.99	0.99	0.98	0.99	0.99
	Accuracy		0.96			0.98			0.98	
	Goodware	0.97	0.97	0.97	0.92	0.90	0.91	0.92	0.90	0.91
	Ryuk	0.95	0.95	0.95	0.83	0.86	0.84	0.83	0.86	0.84
	Accuracy		0.97			0.89			0.89	
SVM	Clop	0.00	0.00	0.00	1.00	0.67	0.80	1.00	0.67	0.80
	Goodware	0.93	1.00	0.96	0.98	1.00	0.99	0.98	1.00	0.99
	Accuracy		0.93			0.98			0.98	
	Conti	1.00	0.69	0.82	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.66	1.00	0.80	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.81			1.00			1.00	
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	0.81	0.95	0.87	0.93	0.97	0.95	1.00	0.88	0.93
	LockBit	0.85	0.55	0.67	0.94	0.85	0.89	0.80	1.00	0.89
	Accuracy		0.82			0.93			0.92	
	Goodware	0.95	1.00	0.98	0.98	1.00	0.99	0.98	1.00	0.99
	MountLocker	1.00	0.50	0.67	1.00	0.75	0.86	1.00	0.75	0.86
	Accuracy		0.96			0.98			0.98	
	Goodware	1.00	1.00	1.00	0.92	0.97	0.95	0.93	1.00	0.96
	NetWalker	1.00	1.00	1.00	0.97	0.92	0.95	1.00	0.92	0.96
	Accuracy		1.00			0.95			0.96	
	Goodware	0.91	0.58	0.71	1.00	0.86	0.93	0.94	0.94	0.94
	Revil	0.95	0.99	0.97	0.98	1.00	0.99	0.99	0.99	0.99
	Accuracy		0.95			0.99			0.99	
	Goodware	0.98	1.00	0.99	0.93	0.93	0.93	0.93	1.00	0.96
	Ryuk	1.00	0.95	0.98	0.86	0.86	0.86	1.00	0.86	0.93
	Accuracy		0.98			0.9			0.95	
NB	Clop	0.75	1.00	0.86	0.60	1.00	0.75	0.60	1.00	0.75
	Goodware	1.00	0.98	0.99	1.00	0.95	0.97	1.00	0.95	0.97
	Accuracy		0.98			0.95			0.95	
	Conti	0.98	0.98	0.98	0.93	0.98	0.96	0.80	0.96	0.88
	Goodware	0.97	0.97	0.97	0.97	0.88	0.92	0.91	0.61	0.73
	Accuracy		0.98			0.94			0.83	
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	1.00	0.88	0.93	1.00	0.88	0.93	0.97	0.95	0.96

Continua na próxima página

Tabela 20 – continuação da página anterior

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	LockBit	0.80	1.00	0.89	0.80	1.00	0.89	0.90	0.95	0.93
	Accuracy		0.92			0.92			0.95	
	Goodware	1.00	0.98	0.99	1.00	0.90	0.95	1.00	0.98	0.99
	MountLocker	0.80	1.00	0.89	0.50	1.00	0.67	0.80	1.00	0.89
	Accuracy		0.98			0.91			0.98	
	Goodware	1.00	1.00	1.00	1.00	0.97	0.99	0.90	1.00	0.95
	NetWalker	1.00	1.00	1.00	0.97	1.00	0.99	1.00	0.89	0.94
	Accuracy		1.00			0.99			0.95	
	Goodware	1.00	0.89	0.94	1.00	0.83	0.91	0.38	0.89	0.53
	Revil	0.99	1.00	0.99	0.98	1.00	0.99	0.99	0.83	0.90
	Accuracy		0.99			0.98			0.84	
	Goodware	1.00	0.95	0.97	1.00	0.95	0.97	1.00	1.00	1.00
	Ryuk	0.92	1.00	0.96	0.92	1.00	0.96	1.00	1.00	1.00
	Accuracy		0.97			0.97			1.00	
DT	Clop	0.50	1.00	0.67	0.50	1.00	0.67	0.50	1.00	0.67
	Goodware	1.00	0.93	0.96	1.00	0.93	0.96	1.00	0.93	0.96
	Accuracy		0.93			0.93			0.93	
	Conti	0.96	0.91	0.93	0.96	0.91	0.93	0.96	0.91	0.93
	Goodware	0.86	0.94	0.90	0.86	0.94	0.90	0.86	0.94	0.90
	Accuracy		0.92			0.92			0.92	
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	0.98	1.00	0.99	0.98	1.00	0.99	0.98	1.00	0.99
	LockBit	1.00	0.95	0.97	1.00	0.95	0.97	1.00	0.95	0.97
	Accuracy		0.98			0.98			0.98	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MountLocker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	NetWalker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	1.00	0.89	0.94	1.00	0.89	0.94	1.00	0.89	0.94
	Revil	0.99	1.00	0.99	0.99	1.00	0.99	0.99	1.00	0.99
	Accuracy		0.99			0.99			0.99	
	Goodware	0.98	1.00	0.99	0.98	1.00	0.99	0.98	1.00	0.99
	Ryuk	1.00	0.95	0.98	1.00	0.95	0.98	1.00	0.95	0.98
	Accuracy		0.98			0.98			0.98	
RF	Clop	1.00	0.67	0.80	1.00	1.00	1.00	1.00	0.97	0.8
	Goodware	0.98	1.00	0.99	1.00	1.00	1.00	0.98	1.00	0.99
	Accuracy		0.98			1.00			0.98	
	Conti	0.95	0.98	0.96	1.00	1.00	1.00	0.96	0.91	0.93
	Goodware	0.97	0.91	0.94	1.00	1.00	1.00	0.86	0.94	0.90
	Accuracy		0.95			1.00			0.92	
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	1.00	0.95	0.97	1.00	0.93	0.96	1.00	0.97	0.99
	LockBit	0.91	1.00	0.95	0.87	1.00	0.93	0.95	1.00	0.98
	Accuracy		0.97			0.95			0.98	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	0.97	0.95	0.96
	MountLocker	1.00	1.00	1.00	1.00	1.00	1.00	0.60	0.75	0.67
	Accuracy		1.00			1.00			0.93	
	Goodware	0.93	1.00	0.96	1.00	1.00	1.00	0.95	0.97	0.96
	NetWalker	1.00	0.92	0.96	1.00	1.00	1.00	0.97	0.95	0.96
	Accuracy		0.99			1.00			0.96	
	Goodware	0.94	0.92	0.93	0.94	0.92	0.93	0.97	0.89	0.93
	Revil	0.99	0.99	0.99	0.99	0.99	0.99	0.99	1.00	0.99
	Accuracy		0.99			0.99			0.99	
	Goodware	1.00	0.95	0.97	1.00	0.95	0.97	0.97	0.97	0.97
	Ryuk	0.92	1.00	0.96	0.92	1.00	0.96	0.95	0.95	0.95
	Accuracy		0.97			0.97			0.97	
MLP	Clop	0.00	0.00	0.00	1.00	1.00	1.00	1.00	0.67	0.80
	Goodware	0.93	1.00	0.96	1.00	1.00	1.00	0.98	1.00	0.99
	Accuracy		0.93			1.00			0.98	
	Conti	0.62	1.00	0.77	1.00	0.87	0.93	0.92	1.00	0.96
	Goodware	0.00	0.00	0.00	0.82	1.00	0.90	1.00	0.85	0.92
	Accuracy		0.62			0.92			0.94	
	Egregor	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.68	1.00	0.81	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.68			1.00			1.00	
	Goodware	0.67	1.00	0.80	0.97	0.95	0.96	1.00	0.88	0.93
	LockBit	0.00	0.00	0.00	0.90	0.95	0.93	0.80	1.00	0.89
	Accuracy		0.67			0.95			0.92	
	Goodware	0.91	1.00	0.95	0.98	0.98	0.98	1.00	1.00	1.00
	MountLocker	0.00	0.00	0.00	0.75	0.75	0.75	1.00	1.00	1.00
	Accuracy		0.91			0.96			1.00	
	Goodware	0.00	0.00	0.00	1.00	0.97	0.99	1.00	0.97	0.99
	NetWalker	0.51	1.00	0.67	0.97	1.00	0.99	0.97	1.00	0.99
	Accuracy		0.51			0.99			0.99	
	Goodware	0.00	0.00	0.00	0.49	1.00	0.66	0.90	0.72	0.80
	Revil	0.90	1.00	0.95	1.00	0.88	0.94	0.97	0.99	0.98

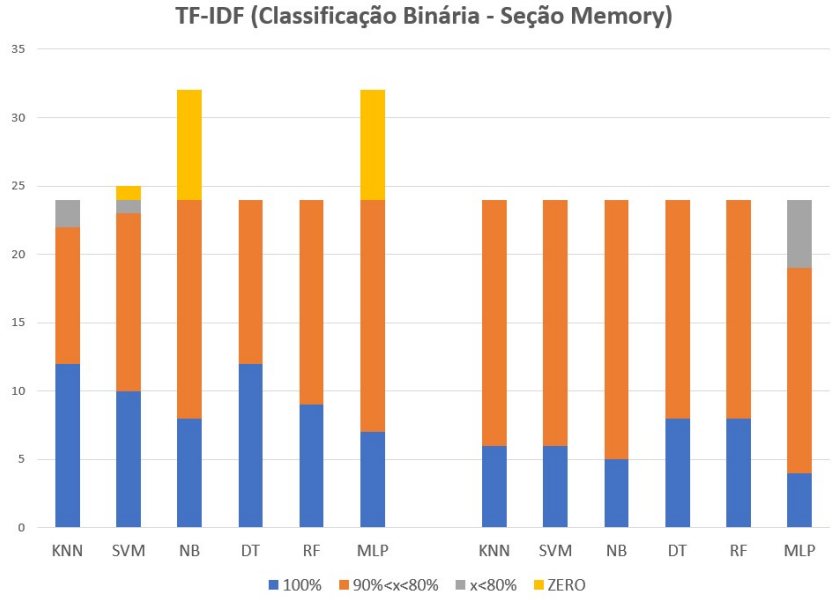
Continua na próxima página

Tabela 20 – continuação da página anterior

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Accuracy		0.9			0.89			0.96	
	Goodware	0.65	1.00	0.78	1.00	0.88	0.93	1.00	0.97	0.99
	Ryuk	0.00	0.00	0.00	0.81	1.00	0.90	0.96	1.00	0.98
	Accuracy		0.65			0.92			0.98	

A Figura 10, de maneira semelhante ao que fizemos nas Subseções anteriores, apresenta a sumarização dos dados da classificação das Tabelas 19 e 20:

Figura 10: Sumarização dos resultados das Tabelas 19 e 20.



Podemos observar que os classificadores apresentam desempenhos distintos em cada *test size* escolhido:

- Todos os classificadores tiveram redução de classificações na faixa azul, ao mesmo tempo, KNN, SVM, NB e MLP tiveram redução nas faixas cinza e laranja, o que demonstra que nesse caso, utilizar *test size* maior é mais vantajoso para obter melhor desempenho, exceto DT e RF;
- Da mesma maneira como ocorreu com os conjuntos de classificação anteriores, o DT e RF apresentaram o melhor desempenho, em ambos os *test size* utilizados;
- Nessa configuração, os mais indicados para uso em detecção de *ransomwares* são as DT e RF para qualquer tamanho de *test size*.

6.4.2.3 Seção *Strings*

Esta seção é a que mais se aproxima da ideia inicial de mineração de dados utilizando a técnica TF-IDF, pois realmente se trata da seção do arquivo PE[3]. Nesta abordagem foram analisadas as seções de strings extraídas pelo *sandbox* dos arquivos analisados.

As classificações realizadas se mostraram promissoras, com algumas poucas ressalvas. Nas classificações multiclasse, os classificadores apresentaram poucas métricas zero, que ficaram concentradas nas classificações na SVM e no MLP, com *test size* 1/3. Com *test size* 1/2, as classificações com métricas zero também ficaram concentradas com o MLP e SVM, porém também apareceram no *Clop* ao ser classificado com DT.

Tabela 21: Extrato da Tabela 23 para os classificadores KNN e NB

	Malware	TestSize 0,33		
		Normal		
		Precision	Recall	F1-Score
KNN	Clop	1.00	0.43	0.60
	Conti	0.15	1.00	0.27
	Egregor	1.00	0.75	0.86
	Goodware	1.00	0.14	0.25
	LockBit	1.00	0.72	0.84
	MountLocker	1.00	1.00	1.00
	NetWalker	1.00	0.96	0.98
	Revil	1.00	0.18	0.30
	Ryuk	1.00	0.31	0.48
	Accuracy	0.39		
NB		PCA		
	Clop	0.42	0.71	0.53
	Conti	0.33	0.33	0.05
	Egregor	1.00	1.00	1.00
	Goodware	0.67	0.57	0.62
	LockBit	0.70	0.78	0.74
	MountLocker	0.80	1.00	0.89
	NetWalker	0.89	0.96	0.92
	Revil	0.80	0.02	0.04
	Ryuk	0.05	0.75	0.09
	Accuracy	0.25		

As classificações de baixo desempenho ficaram entre 0.37 e 0.77 para *test size* 1/3 e entre 0.40 e 0.75 para *test size* 1/2, conforme Tabelas 21 e 22, que são os extratos das Tabelas 23 e 24 com os piores resultados de classificação.

[3] *Portable Executable* que possui textolegível. O PE é um formato de arquivo usado para arquivos executáveis, código de objeto e DLLs usados em versões de 32 bits e 64 bits dos sistemas operacionais *Windows*. Os arquivos *Portable Executable* são criados e usados pelo sistema operacional *Microsoft Windows* para executar e carregar aplicativos e DLLs.

Tabela 22: Extrato da Tabela 24 para o classificador Naive Bayes.

	Malware	TestSize 0,5		
		Normal		
		Precision	Recall	F1-Score
NB	Clop	0.44	0.70	0.54
	Conti	0.17	0.02	0.03
	Egregor	1.00	1.00	1.00
	Goodware	0.63	0.55	0.59
	LockBit	0.62	0.75	0.68
	MountLocker	0.67	0.22	0.33
	NetWalker	0.90	0.88	0.89
	Revil	0.80	0.01	0.03
	Ryuk	0.05	0.78	0.09
	Accuracy	0.23		

Para as classificações binárias, com *test size* 1/2 foram observadas algumas situações com *Accuracy* zero para *Clop* e *MountLocker*, na classificação com a SVM, conjunto de dados não otimizado. Este mesmo comportamento ocorreu com as classificações do *Conti*, *LockBit*, *MountLocker* e *Ryuk* ao serem submetidos ao MLP para classificação, porém a aplicação do *StandardScaler* e do PCA proporcionaram melhoras significativas, conforme podemos observar na Tabela 24. Surpreendentemente, o classificador que teve maior quantidade de classificações com *Accuracy* 1.00 foi o NB, ao mesmo tempo que não apresentou métricas zero. De 144 classificações, 25 ficaram com *Accuracy* 1.00. Ademais, para a situação de *test size* 1/2, tivemos apenas 4 situações com *Accuracy* zero: *Clop* e *MountLocker* com conjunto de dados não padronizado, classificados pela SVM e *LockBit*, *MountLocker* e *Ryuk*, com conjunto de dados não padronizado, classificados pelo MLP. De 144 classificações, 24 ficaram com *Accuracy* 1.00.

Tabela 23: Tabela com os dados das classificações referente a abordagem de TF-IDF (Strings), com *test size* 0,33 e classificação multiclasse.

	Malware	Test Size 0,33								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
KNN	Clop	1.00	0.57	0.73	1.00	0.43	0.60	0.83	0.71	0.77
	Conti	0.92	0.85	0.88	0.15	1.00	0.27	0.85	0.90	0.88
	Egregor	1.00	1.00	1.00	1.00	0.75	0.86	1.00	1.00	1.00
	Goodware	1.00	0.19	0.32	1.00	0.14	0.25	0.89	0.76	0.82
	LockBit	0.34	1.00	0.51	1.00	0.72	0.84	0.75	0.83	0.79
	MountLocker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	NetWalker	1.00	0.76	0.86	1.00	0.96	0.98	0.80	0.96	0.87
	Revil	0.97	0.96	0.96	1.00	0.18	0.30	1.00	1.00	1.00
	Ryuk	1.00	0.81	0.90	1.00	0.31	0.48	1.00	0.69	0.81
	Accuracy	0.87			0.39			0.94		
SVM	Clop	0.00	0.00	0.00	1.00	0.57	0.73	0.71	0.71	0.71
	Conti	0.72	0.85	0.78	0.89	0.82	0.85	0.80	0.90	0.84
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	0.29	0.44	1.00	0.67	0.80	1.00	0.67	0.80
	LockBit	0.00	0.00	0.00	1.00	0.72	0.84	0.75	0.83	0.79
	MountLocker	0.00	0.00	0.00	1.00	1.00	1.00	0.57	1.00	0.73
	NetWalker	1.00	0.96	0.98	0.96	0.96	0.96	0.80	0.96	0.87
	Revil	0.80	1.00	0.89	0.90	1.00	0.95	1.00	0.99	0.99
	Ryuk	1.00	0.06	0.12	1.00	0.69	0.81	1.00	0.69	0.81
	Accuracy	0.81			0.92			0.93		

Continua na próxima página

Tabela 23 – continuação da página anterior

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
NB	Clop	1.00	0.71	0.83	1.00	0.57	0.73	0.42	0.71	0.53
	Conti	0.92	0.92	0.92	0.95	0.95	0.95	0.33	0.03	0.05
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.80	0.95	0.87	0.84	1.00	0.91	0.67	0.57	0.62
	LockBit	0.94	0.83	0.88	1.00	0.89	0.94	0.70	0.78	0.74
	MountLocker	1.00	1.00	1.00	1.00	1.00	1.00	0.80	1.00	0.89
	NetWalker	0.96	0.96	0.96	0.92	0.96	0.94	0.89	0.96	0.92
	Revil	0.98	1.00	0.99	0.98	1.00	0.99	0.80	0.02	0.04
	Ryuk	1.00	0.75	0.86	1.00	0.75	0.86	0.05	0.75	0.09
	Accuracy		0.96			0.96			0.25	
DT	Clop	1.00	0.43	0.60	1.00	0.43	0.60	1.00	0.43	0.60
	Conti	0.85	0.87	0.86	0.85	0.87	0.86	0.85	0.87	0.86
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.87	0.95	0.91	0.87	0.95	0.91	0.87	0.95	0.91
	LockBit	0.94	0.89	0.91	0.94	0.89	0.91	0.94	0.89	0.91
	MountLocker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	NetWalker	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
	Revil	0.98	1.00	0.99	0.98	1.00	0.99	0.98	1.00	0.99
	Ryuk	0.86	0.75	0.80	0.86	0.75	0.80	0.86	0.75	0.80
	Accuracy		0.95			0.95			0.95	
RF	Clop	1.00	0.57	0.73	1.00	0.57	0.73	1.00	0.71	0.83
	Conti	0.89	0.79	0.84	0.87	0.87	0.87	0.82	0.79	0.81
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.69	0.95	0.80	0.77	0.95	0.85	0.81	0.81	0.81
	LockBit	0.88	0.83	0.86	0.79	0.83	0.81	0.88	0.83	0.86
	MountLocker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	NetWalker	0.96	0.96	0.96	0.96	0.96	0.96	0.89	0.96	0.92
	Revil	0.98	1.00	0.99	0.99	1.00	0.99	0.97	1.00	0.98
	Ryuk	0.85	0.69	0.76	0.92	0.69	0.79	1.00	0.69	0.81
	Accuracy		0.94			0.95			0.93	
MLP	Clop	1.00	0.43	0.60	0.50	0.14	0.22	0.00	0.00	0.00
	Conti	0.81	0.87	0.84	1.00	0.82	0.90	0.63	0.82	0.71
	Egregor	1.00	1.00	1.00	0.90	0.75	0.82	0.00	0.00	0.00
	Goodware	1.00	0.76	0.86	1.00	0.38	0.55	0.37	1.00	0.54
	LockBit	0.88	0.83	0.86	1.00	0.78	0.88	0.00	0.00	0.00
	MountLocker	1.00	1.00	1.00	0.44	1.00	0.62	0.50	1.00	0.67
	NetWalker	1.00	0.96	0.98	0.39	1.00	0.56	0.77	0.96	0.86
	Revil	0.92	1.00	0.96	1.00	0.96	0.98	1.00	0.96	0.98
	Ryuk	1.00	0.44	0.61	1.00	0.69	0.81	0.00	0.00	0.00
	Accuracy		0.92			0.87			0.80	

Tabela 24: Tabela com os dados das classificações referente a abordagem de TF-IDF (Strings), com *test size* 0,5 e classificação multiclasse.

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
KNN	Clop	1.00	0.60	0.75	1.00	0.40	0.57	0.83	0.71	0.77
	Conti	0.95	0.87	0.91	0.16	1.00	0.27	0.85	0.90	0.88
	Egregor	1.00	0.94	0.97	1.00	0.72	0.84	1.00	1.00	1.00
	Goodware	1.00	0.16	0.28	1.00	0.10	0.18	0.89	0.76	0.82
	LockBit	0.24	0.96	0.39	1.00	0.75	0.86	0.75	0.83	0.79
	MountLocker	1.00	0.89	0.94	1.00	0.89	0.94	1.00	1.00	1.00
	NetWalker	1.00	0.79	0.88	1.00	0.90	0.95	0.80	0.96	0.87
	Revil	0.98	0.91	0.94	1.00	0.15	0.27	1.00	1.00	1.00
	Ryuk	1.00	0.87	0.93	1.00	0.39	0.56	1.00	0.69	0.81
	Accuracy		0.85			0.38			0.94	
SVM	Clop	0.00	0.00	0.00	1.00	0.50	0.67	0.71	0.71	0.71
	Conti	0.75	0.85	0.80	0.93	0.84	0.88	0.80	0.90	0.84
	Egregor	1.00	0.94	0.97	1.00	0.94	0.97	1.00	1.00	1.00
	Goodware	1.00	0.23	0.37	0.93	0.45	0.61	1.00	0.67	0.80
	LockBit	0.00	0.00	0.00	1.00	0.75	0.86	0.75	0.83	0.79
	MountLocker	0.00	0.00	0.00	1.00	0.89	0.94	0.57	1.00	0.73
	NetWalker	1.00	0.90	0.95	0.97	0.90	0.94	0.80	0.96	0.87
	Revil	0.78	1.00	0.88	0.88	1.00	0.93	1.00	0.99	0.99
	Ryuk	0.00	0.00	0.00	1.00	0.78	0.88	1.00	0.69	0.81
	Accuracy		0.8			0.91			0.93	
NB	Clop	0.89	0.80	0.84	0.86	0.60	0.71	0.44	0.70	0.54
	Conti	0.94	0.95	0.94	0.91	0.95	0.93	0.17	0.02	0.03
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.86	0.97	0.91	0.84	1.00	0.91	0.63	0.55	0.59
	LockBit	0.91	0.88	0.89	0.95	0.88	0.91	0.62	0.75	0.68
	MountLocker	1.00	0.89	0.94	1.00	0.89	0.94	0.67	0.22	0.33
	NetWalker	1.00	0.88	0.94	0.97	0.88	0.93	0.90	0.88	0.89
	Revil	0.97	0.99	0.98	0.97	0.99	0.98	0.80	0.01	0.03
	Ryuk	0.95	0.83	0.88	0.95	0.83	0.88	0.05	0.78	0.09
	Accuracy		0.96			0.95			0.23	

Continua na próxima página

Tabela 24 – continuação da página anterior

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
DT	Clop	1.00	0.60	0.75	1.00	0.60	0.75	1.00	0.60	0.75
	Conti	0.81	0.89	0.84	0.81	0.89	0.84	0.81	0.89	0.84
	Egregor	0.95	1.00	0.97	0.95	1.00	0.97	0.95	1.00	0.97
	Goodware	0.88	0.90	0.89	0.88	0.90	0.89	0.88	0.90	0.89
	LockBit	0.91	0.83	0.87	0.91	0.83	0.87	0.91	0.83	0.87
	MountLocker	0.73	0.89	0.80	0.73	0.89	0.80	0.73	0.89	0.80
	NetWalker	1.00	0.90	0.95	1.00	0.90	0.95	1.00	0.90	0.95
	Revil	1.00	0.99	0.99	1.00	0.99	0.99	1.00	0.99	0.99
	Ryuk	0.76	0.83	0.79	0.76	0.83	0.79	0.76	0.83	0.79
	Accuracy		0.94			0.94			0.94	
RF	Clop	1.00	0.30	0.46	0.80	0.40	0.53	1.00	0.50	0.67
	Conti	0.92	0.79	0.85	0.91	0.80	0.85	0.89	0.82	0.85
	Egregor	0.86	1.00	0.92	0.86	1.00	0.92	1.00	1.00	1.00
	Goodware	0.69	1.00	0.82	0.62	0.97	0.76	0.69	0.81	0.75
	LockBit	0.83	0.83	0.83	0.91	0.83	0.87	0.69	0.83	0.75
	MountLocker	1.00	0.89	0.94	1.00	0.89	0.94	1.00	0.89	0.94
	NetWalker	0.86	0.90	0.88	0.93	0.90	0.92	0.84	0.90	0.87
	Revil	0.99	0.99	0.99	1.00	0.99	0.99	0.98	0.99	0.98
	Ryuk	0.82	0.78	0.80	0.86	0.78	0.82	0.94	0.74	0.83
	Accuracy		0.93			0.93			0.92	
MLP	Clop	0.00	0.00	0.00	1.00	0.40	0.57	1.00	0.30	0.46
	Conti	0.84	0.89	0.86	1.00	0.79	0.88	0.66	0.79	0.72
	Egregor	1.00	0.94	0.97	0.84	0.89	0.86	0.65	0.72	0.68
	Goodware	1.00	0.71	0.83	1.00	0.26	0.41	0.49	0.68	0.57
	LockBit	0.80	0.83	0.82	0.60	0.12	0.21	0.61	0.79	0.69
	MountLocker	1.00	0.89	0.94	0.08	0.78	0.14	0.89	0.89	0.89
	NetWalker	1.00	0.90	0.95	0.97	0.90	0.94	0.81	0.90	0.85
	Revil	0.91	1.00	0.95	0.99	0.94	0.96	1.00	0.97	0.99
	Ryuk	1.00	0.70	0.82	0.87	0.87	0.87	0.00	0.00	0.00
	Accuracy		0.92			0.82			0.86	

Tabela 25: Tabela com os dados das classificações referente a abordagem de TF-IDF (Strings), com *test size* 0,33 e classificação Binária.

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
KNN	Clop	1.00	0.50	0.67	1.00	0.50	0.67	1.00	0.50	0.67
	Goodware	0.96	1.00	0.98	0.96	1.00	0.98	0.96	1.00	0.98
	Accuracy		0.97			0.97			0.97	
	Conti	0.97	0.92	0.95	0.69	1.00	0.82	0.83	1.00	0.90
	Goodware	0.87	0.95	0.91	1.00	0.19	0.32	1.00	0.62	0.76
	Accuracy		0.93			0.71			0.86	
	Egregor	1.00	1.00	1.00	0.55	1.00	0.71	0.55	1.00	0.71
	Goodware	1.00	1.00	1.00	1.00	0.68	0.81	1.00	0.68	0.81
	Accuracy		1.00			0.77			0.77	
	Goodware	0.88	1.00	0.93	0.93	1.00	0.97	0.93	1.00	0.97
	LockBit	1.00	0.67	0.80	1.00	0.83	0.91	1.00	0.83	0.91
	Accuracy		0.65			0.95			0.95	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MountLocker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	0.81	1.00	0.90	1.00	0.32	0.48	1.00	0.32	0.48
	NetWalker	1.00	0.82	0.90	0.65	1.00	0.79	0.65	1.00	0.79
	Accuracy		0.90			0.70			0.70	
	Goodware	1.00	0.32	0.48	0.83	0.80	0.82	0.95	0.84	0.89
	Revil	0.65	1.00	0.79	0.98	0.98	0.98	0.98	1.00	0.99
	Accuracy		0.94			0.96			0.98	
	Goodware	0.96	0.96	0.96	1.00	0.29	0.44	1.00	0.82	0.90
	Ryuk	0.92	0.92	0.92	0.39	1.00	0.57	0.72	1.00	0.84
	Accuracy		0.95			0.51			0.88	
SVM	Clop	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.93	1.00	0.96	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.93			1.00			1.00	
	Conti	0.97	0.97	0.97	1.00	0.97	0.99	0.79	0.97	0.87
	Goodware	0.95	0.95	0.95	0.95	1.00	0.98	0.92	0.52	0.67
	Accuracy		0.97			0.98			0.81	
	Egregor	1.00	0.91	0.95	1.00	0.91	0.95	1.00	0.91	0.95
	Goodware	0.97	1.00	0.98	0.97	1.00	0.98	0.97	1.00	0.98
	Accuracy		0.97			0.97			0.97	
	Goodware	0.88	1.00	0.93	0.97	1.00	0.98	0.97	1.00	0.98
	LockBit	1.00	0.67	0.80	1.00	0.92	0.96	1.00	0.92	0.96
	Accuracy		0.9			0.97			0.97	
	Goodware	0.93	1.00	0.97	1.00	1.00	1.00	1.00	1.00	1.00
	MountLocker	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.93			1.00			1.00	
	Goodware	0.92	1.00	0.96	0.92	1.00	0.96	0.90	0.86	0.88

Continua na próxima página

Tabela 25 – continuação da página anterior

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	NetWalker	1.00	0.93	0.96	1.00	0.93	0.96	0.90	0.93	0.91
	Accuracy		0.96			0.96			0.9	
	Goodware	1.00	0.52	0.68	0.96	0.96	0.96	1.00	0.80	0.89
	Revil	0.95	1.00	0.97	1.00	1.00	1.00	0.98	1.00	0.99
	Accuracy		0.95			0.99			0.98	
	Goodware	1.00	1.00	1.00	1.00	0.93	0.96	1.00	0.93	0.96
	Ryuk	1.00	1.00	1.00	0.87	1.00	0.93	0.87	1.00	0.93
	Accuracy		1.00			0.95			0.95	
	Clop	1.00	1.00	1.00	0.50	0.50	0.50	0.25	1.00	0.40
	Goodware	1.00	1.00	1.00	0.96	0.96	0.96	1.00	0.78	0.88
	Accuracy		1.00			0.93			0.79	
	Conti	0.97	1.00	0.99	1.00	1.00	1.00	0.95	0.92	0.93
	Goodware	1.00	0.95	0.98	1.00	1.00	1.00	0.86	0.90	0.88
	Accuracy		0.98			1.00			0.92	
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.89	0.94
	LockBit	1.00	1.00	1.00	1.00	1.00	1.00	0.80	1.00	0.89
NB	Accuracy		1.00			1.00			0.93	
	Goodware	1.00	0.96	0.98	1.00	1.00	1.00	1.00	0.82	0.90
	MountLocker	0.67	1.00	0.80	1.00	1.00	1.00	0.29	1.00	0.44
	Accuracy		0.97			1.00			0.83	
	Goodware	0.92	1.00	0.96	0.92	1.00	0.96	1.00	1.00	1.00
	NetWalker	1.00	0.93	0.96	1.00	0.93	0.96	1.00	1.00	1.00
	Accuracy		0.96			0.96			1.00	
	Goodware	0.92	0.96	0.94	0.97	1.00	0.98	0.86	0.76	0.81
	Revil	1.00	0.99	0.99	1.00	0.92	0.96	0.97	0.99	0.98
	Accuracy		0.99			0.98			0.96	
	Goodware	0.93	0.93	0.93	0.97	1.00	0.98	0.93	0.93	0.93
	Ryuk	0.85	0.85	0.85	1.00	0.92	0.96	0.85	0.85	0.85
	Accuracy		1.00			0.98			0.93	
	Clop	0.67	1.00	0.80	0.67	1.00	0.80	0.67	1.00	0.80
	Goodware	1.00	0.96	0.98	1.00	0.96	0.98	1.00	0.96	0.98
	Accuracy		0.97			0.97			0.97	
	Conti	0.95	0.97	0.96	0.95	0.97	0.96	0.95	0.97	0.96
	Goodware	0.95	0.90	0.93	0.95	0.90	0.93	0.95	0.90	0.93
	Accuracy		0.95			0.95			0.95	
	Egregor	0.92	1.00	0.96	0.92	1.00	0.96	0.92	1.00	0.96
	Goodware	1.00	0.96	0.98	1.00	0.96	0.98	1.00	0.96	0.98
	Accuracy		0.97			0.97			0.97	
	Goodware	1.00	0.82	0.90	1.00	0.82	0.90	1.00	0.82	0.90
	LockBit	0.71	1.00	0.83	0.71	1.00	0.83	0.71	1.00	0.83
DT	Accuracy		0.88			0.88			0.88	
	Goodware	1.00	0.93	0.96	1.00	0.93	0.96	1.00	0.93	0.96
	MountLocker	0.50	1.00	0.67	0.50	1.00	0.67	0.50	1.00	0.67
	Accuracy		0.93			0.93			0.93	
	Goodware	0.92	1.00	0.96	0.92	1.00	0.96	0.92	1.00	0.96
	NetWalker	1.00	0.93	0.96	1.00	0.93	0.96	1.00	0.93	0.96
	Accuracy		0.96			0.96			0.96	
	Goodware	0.92	0.96	0.94	0.92	0.96	0.94	0.92	0.96	0.94
	Revil	1.00	0.99	0.99	1.00	0.99	0.99	1.00	0.99	0.99
	Accuracy		0.99			0.99			0.99	
	Goodware	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93
	Ryuk	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85
	Accuracy		0.90			0.90			0.90	
	Clop	0.67	1.00	0.80	0.67	1.00	0.80	0.67	1.00	0.80
	Goodware	1.00	0.96	0.98	1.00	0.96	0.98	1.00	0.96	0.98
	Accuracy		0.97			0.97			0.97	
	Conti	0.86	1.00	0.93	0.93	0.97	0.95	1.00	1.00	1.00
	Goodware	1.00	0.71	0.83	0.95	0.86	0.90	1.00	1.00	1.00
	Accuracy		0.9			0.93			1.00	
	Egregor	1.00	0.91	0.95	1.00	0.91	0.95	1.00	1.00	1.00
	Goodware	0.97	1.00	0.98	0.97	1.00	0.98	1.00	1.00	1.00
	Accuracy		0.97			0.97			1.00	
	Goodware	0.97	1.00	0.98	1.00	0.82	0.90	1.00	1.00	1.00
	LockBit	1.00	0.92	0.96	0.71	1.00	0.83	1.00	1.00	1.00
RF	Accuracy		0.97			1.00			1.00	
	Goodware	1.00	0.96	0.98	1.00	1.00	1.00	1.00	0.96	0.98
	MountLocker	0.67	1.00	0.80	1.00	1.00	1.00	0.67	1.00	0.80
	Accuracy		0.97			1.00			0.97	
	Goodware	0.92	1.00	0.96	0.91	0.95	0.93	0.92	1.00	0.96
	NetWalker	1.00	0.93	0.96	0.96	0.93	0.95	1.00	0.93	0.96
	Accuracy		0.96			0.94			0.96	
	Goodware	0.92	0.96	0.94	0.92	0.96	0.94	0.92	0.92	0.92
	Revil	1.00	0.99	0.99	1.00	0.99	0.99	0.99	0.99	0.99
	Accuracy		0.99			0.99			0.98	
	Goodware	0.96	0.96	0.96	1.00	1.00	1.00	0.96	0.89	0.93
	Ryuk	0.92	0.92	0.92	1.00	1.00	1.00	0.80	0.92	0.86
	Accuracy		0.95			1.00			0.90	
	Clop	1.00	0.50	0.67	0.67	1.00	0.80	0.00	0.00	0.00
	Goodware	0.96	1.00	0.98	1.00	0.96	0.98	0.93	1.00	0.96

Continua na próxima página

Tabela 25 – continuação da página anterior

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Accuracy		0.97			0.97			0.93	
	Conti	0.64	1.00	0.78	0.78	1.00	0.87	0.90	0.97	0.94
	Goodware	0.00	0.00	0.00	1.00	0.48	0.65	0.94	0.81	0.87
	Accuracy		0.64			0.81			0.92	
	Egregor	0.65	1.00	0.79	0.52	1.00	0.69	0.67	0.91	0.77
	Goodware	1.00	0.79	0.88	1.00	0.64	0.78	0.96	0.82	0.88
	Accuracy		0.85			0.74			0.85	
	Goodware	0.70	1.00	0.82	1.00	1.00	1.00	0.88	0.79	0.83
	LockBit	0.00	0.00	0.00	1.00	1.00	1.00	0.60	0.75	0.67
	Accuracy		0.70			1.00			0.78	
	Goodware	0.93	1.00	0.97	1.00	0.68	0.81	1.00	1.00	1.00
	MountLocker	0.00	0.00	0.00	0.18	1.00	0.31	1.00	1.00	1.00
	Accuracy		0.93			0.70			1.00	
	Goodware	0.63	1.00	0.77	1.00	0.95	0.98	0.93	0.64	0.76
	NetWalker	1.00	0.54	0.70	0.97	1.00	0.98	0.77	0.96	0.86
	Accuracy		0.74			0.98			0.82	
	Goodware	1.00	0.68	0.81	0.81	0.68	0.74	0.90	0.76	0.83
	Revil	0.96	1.00	0.98	0.96	0.98	0.97	0.97	0.99	0.98
	Accuracy		0.97			0.95			0.97	
	Goodware	0.68	1.00	0.81	0.97	1.00	0.98	0.93	0.50	0.65
	Ryuk	0.00	0.00	0.00	1.00	0.92	0.96	0.46	0.92	0.62
	Accuracy		0.68			0.98			0.63	

Tabela 26: Tabela com os dados das classificações referente a abordagem de TF-IDF (Strings), com *test size* 0,5 e classificação Binária.

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
KNN	Clop	1.00	0.67	0.80	1.00	0.67	0.80	1.00	0.67	0.80
	Goodware	0.98	1.00	0.99	0.98	1.00	0.99	0.98	1.00	0.99
	Accuracy		0.98			0.98			0.98	
	Conti	1.00	0.87	0.93	0.65	1.00	0.79	0.78	0.98	0.87
	Goodware	e	1.00	0.90	1.00	0.09	0.17	0.95	0.55	0.69
	Accuracy		0.92			0.66			0.82	
	Egregor	1.00	1.00	1.00	0.40	1.00	0.57	0.40	1.00	0.57
	Goodware	1.00	1.00	1.00	1.00	0.28	0.43	1.00	0.28	0.43
	Accuracy		1.00			0.51			0.51	
	Goodware	0.87	1.00	0.93	0.91	1.00	0.95	0.91	1.00	0.95
	LockBit	1.00	0.70	0.82	1.00	0.80	0.89	1.00	0.80	0.89
	Accuracy		0.60			0.93			0.93	
	Goodware	1.00	0.98	0.99	1.00	1.00	1.00	1.00	1.00	1.00
	MountLocker	0.80	1.00	0.89	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.98			1.00			1.00	
	Goodware	0.88	1.00	0.94	1.00	0.08	0.15	1.00	0.08	0.15
	NetWalker	1.00	0.87	0.93	0.53	1.00	0.69	0.53	1.00	0.69
	Accuracy		0.93			0.55			0.55	
	Goodware	1.00	0.44	0.62	1.00	0.17	0.29	0.88	0.81	0.84
	Revil	0.94	1.00	0.97	0.91	1.00	0.95	0.98	0.99	0.98
	Accuracy		0.94			0.91			0.97	
	Goodware	0.95	0.95	0.95	0.96	0.60	0.74	0.97	0.95	0.96
	Ryuk	0.91	0.91	0.91	0.57	0.95	0.71	0.91	0.95	0.93
	Accuracy		0.94			0.73			0.95	
SVM	Clop	0.00	0.00	0.00	1.00	0.67	0.80	1.00	1.00	1.00
	Goodware	0.93	1.00	0.96	0.98	1.00	0.99	1.00	1.00	1.00
	Accuracy		0.93			0.98			1.00	
	Conti	0.98	0.96	0.97	1.00	0.93	0.96	0.84	0.85	0.85
	Goodware	0.94	0.97	0.96	0.89	1.00	0.94	0.75	0.73	0.74
	Accuracy		0.97			0.95			0.81	
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.95	0.97
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	0.98	1.00	0.99
	Accuracy		1.00			1.00			0.98	
	Goodware	0.87	1.00	0.93	0.95	1.00	0.98	0.95	1.00	0.98
	LockBit	1.00	0.70	0.82	1.00	0.90	0.95	1.00	0.90	0.95
	Accuracy		0.9			0.97			0.97	
	Goodware	0.91	1.00	0.95	1.00	1.00	1.00	1.00	1.00	1.00
	MountLocker	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.91			1.00			1.00	
	Goodware	0.93	1.00	0.96	0.95	0.97	0.96	0.91	0.81	0.86
	NetWalker	1.00	0.92	0.96	0.97	0.95	0.96	0.83	0.92	0.88
	Accuracy		0.96			0.96			0.87	
	Goodware	1.00	0.39	0.56	0.82	0.86	0.84	1.00	0.67	0.80
	Revil	0.93	1.00	0.97	0.98	0.98	0.98	0.96	1.00	0.98
	Accuracy		0.94			0.97			0.97	
	Goodware	1.00	0.25	0.40	1.00	0.90	0.95	1.00	0.90	0.95
	Ryuk	0.42	1.00	0.59	0.85	1.00	0.92	0.85	1.00	0.92
	Accuracy		0.52			0.94			0.94	

Continua na próxima página

Tabela 26 – continuação da página anterior

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
NB	Clop	1.00	1.00	1.00	0.67	0.67	0.67	0.23	1.00	0.38
	Goodware	1.00	1.00	1.00	0.98	0.98	0.98	1.00	0.76	0.86
	Accuracy		1.00			0.95			0.77	
	Conti	1.00	0.91	0.95	1.00	0.91	0.95	0.96	0.91	0.93
	Goodware	0.87	1.00	0.93	0.87	1.00	0.93	0.86	0.94	0.90
	Accuracy		0.94			0.94			0.92	
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.78	0.87
	LockBit	1.00	1.00	1.00	1.00	1.00	1.00	0.69	1.00	0.82
	Accuracy		1.00			1.00			0.85	
	Goodware	1.00	0.93	0.96	1.00	1.00	1.00	1.00	0.95	0.97
	MountLocker	0.57	1.00	0.73	1.00	1.00	1.00	0.67	1.00	0.80
	Accuracy		0.93			1.00			0.84	
	Goodware	0.93	1.00	0.96	0.93	1.00	0.96	0.97	0.97	0.97
	NetWalker	1.00	0.92	0.96	1.00	0.92	0.96	0.97	0.97	0.97
	Accuracy		0.96			0.96			0.97	
	Goodware	0.94	0.89	0.91	0.97	0.89	0.93	0.84	0.89	0.86
	Revil	0.99	0.99	0.99	0.99	1.00	0.99	0.99	0.98	0.98
	Accuracy		0.98			0.99			0.97	
	Goodware	1.00	1.00	1.00	0.98	1.00	0.99	0.97	0.88	0.92
	Ryuk	1.00	1.00	1.00	1.00	0.95	0.98	0.81	0.95	0.88
	Accuracy		1.00			0.98			0.9	
DT	Clop	0.60	1.00	0.75	0.60	1.00	0.75	0.60	1.00	0.75
	Goodware	1.00	0.95	0.97	1.00	0.95	0.97	1.00	0.95	0.97
	Accuracy		0.95			0.95			0.95	
	Conti	0.96	0.93	0.94	0.96	0.93	0.94	0.96	0.93	0.94
	Goodware	0.89	0.94	0.91	0.89	0.94	0.91	0.89	0.94	0.91
	Accuracy		0.93			0.93			0.93	
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	1.00	0.80	0.89	1.00	0.80	0.89	1.00	0.80	0.89
	LockBit	0.71	1.00	0.83	0.71	1.00	0.83	0.71	1.00	0.83
	Accuracy		0.87			0.87			0.87	
	Goodware	1.00	0.95	0.97	1.00	0.95	0.97	1.00	0.95	0.97
	MountLocker	0.67	1.00	0.80	0.67	1.00	0.80	0.67	1.00	0.80
	Accuracy		0.96			0.96			0.96	
	Goodware	0.97	0.95	0.96	0.97	0.95	0.96	0.97	0.95	0.96
	NetWalker	0.95	0.97	0.96	0.95	0.97	0.96	0.95	0.97	0.96
	Accuracy		0.96			0.96			0.96	
	Goodware	0.90	1.00	0.95	0.90	1.00	0.95	0.90	1.00	0.95
	Revil	1.00	0.99	0.99	1.00	0.99	0.99	1.00	0.99	0.99
	Accuracy		0.99			0.99			0.99	
	Goodware	0.93	0.95	0.94	0.93	0.95	0.94	0.93	0.95	0.94
	Ryuk	0.90	0.86	0.88	0.90	0.86	0.88	0.90	0.86	0.88
	Accuracy		0.92			0.92			0.92	
RF	Clop	0.75	1.00	0.86	0.75	1.00	0.86	0.75	1.00	0.86
	Goodware	1.00	0.98	0.99	1.00	0.98	0.99	1.00	0.98	0.99
	Accuracy		0.98			0.98			0.98	
	Conti	0.94	0.93	0.94	0.96	0.93	0.94	1.00	0.96	0.98
	Goodware	0.88	0.91	0.90	0.89	0.94	0.91	0.94	1.00	0.97
	Accuracy		0.92			0.93			0.98	
	Egregor	0.95	1.00	0.97	0.95	1.00	0.97	1.00	1.00	1.00
	Goodware	1.00	0.97	0.99	1.00	0.97	0.99	1.00	1.00	1.00
	Accuracy		0.98			0.98			1.00	
	Goodware	0.93	1.00	0.96	0.95	1.00	0.98	0.95	1.00	0.98
	LockBit	1.00	0.85	0.92	1.00	0.90	0.95	1.00	0.90	0.95
	Accuracy		0.95			0.97			0.97	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MountLocker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	0.93	1.00	0.96	0.93	1.00	0.96	0.93	1.00	0.96
	NetWalker	1.00	0.92	0.96	1.00	0.92	0.96	1.00	0.92	0.96
	Accuracy		0.96			0.96			0.96	
	Goodware	0.88	1.00	0.94	0.88	1.00	0.94	0.90	1.00	0.95
	Revil	1.00	0.98	0.99	1.00	0.98	0.99	1.00	0.99	0.99
	Accuracy		0.99			0.99			0.99	
	Goodware	0.97	0.97	0.97	0.98	1.00	0.99	0.97	0.95	0.96
	Ryuk	0.95	0.95	0.95	1.00	0.95	0.98	0.91	0.95	0.93
	Accuracy		0.97			0.98			0.95	
MLP	Clop	0.25	1.00	0.40	0.43	1.00	0.60	0.22	0.67	0.33
	Goodware	1.00	0.78	0.88	1.00	0.90	0.95	0.97	0.83	0.89
	Accuracy		0.80			0.91			0.82	
	Conti	0.63	1.00	0.77	0.76	1.00	0.87	0.94	0.87	0.91
	Goodware	1.00	0.03	0.06	1.00	0.48	0.65	0.81	0.91	0.86
	Accuracy		0.64			0.81			0.89	
	Egregor	0.54	1.00	0.70	1.00	1.00	1.00	0.53	1.00	0.69
	Goodware	1.00	0.60	0.75	1.00	1.00	1.00	1.00	0.57	0.73
	Accuracy		0.73			1.00			0.71	
	Goodware	0.67	1.00	0.80	1.00	0.68	0.81	0.89	0.62	0.74

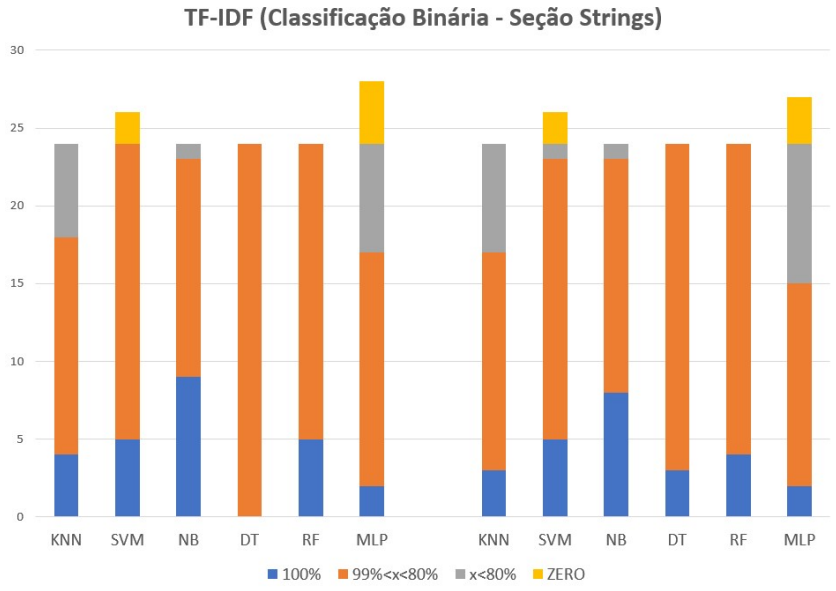
Continua na próxima página

Tabela 26 – continuação da página anterior

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	LockBit	0.00	0.00	0.00	0.61	1.00	0.75	0.53	0.85	0.65
	Accuracy		0.67			0.78			0.7	
	Goodware	0.91	1.00	0.95	1.00	0.68	0.81	1.00	1.00	1.00
	MountLocker	0.00	0.00	0.00	0.24	1.00	0.38	1.00	1.00	1.00
	Accuracy		0.91			0.71			1.00	
	Goodware	0.80	1.00	0.89	0.96	0.73	0.83	0.92	0.62	0.74
	NetWalker	1.00	0.76	0.87	0.79	0.97	0.87	0.72	0.95	0.82
	Accuracy		0.88			0.85			0.79	
	Goodware	1.00	0.44	0.62	0.73	0.75	0.74	0.85	0.92	0.88
	Revil	0.94	1.00	0.97	0.97	0.97	0.97	0.99	0.98	0.99
	Accuracy		0.94			0.95			0.97	
	Goodware	0.65	1.00	0.78	0.98	1.00	0.99	0.96	0.96	0.96
	Ryuk	0.00	0.00	0.00	1.00	0.95	0.98	0.92	0.92	0.92
	Accuracy		0.65			0.98			0.50	

A Figura 11, de maneira semelhante ao que fizemos nas Subseções anteriores, apresenta a sumarização dos dados da classificação das Tabelas 25 e 26:

Figura 11: Sumarização dos resultados das Tabelas 25 e 26.



Podemos observar que os classificadores apresentam desempenhos distintos em cada *test size* escolhido:

- De modo geral, nesta configuração houve pouca mudança no desempenho dos classificadores ao mudarmos o tamanho do *test size* de 1/3 para 1/2. SVM e MLP apresentaram classificações na faixa laranja em ambos os *test size* utilizados, indicando problemas nas classificações;
- KNN, NB e MLP classificações na faixa cinza, indicando desempenho abaixo do esperado em ambos os *test size* utilizados;

- O DT apresentou uma melhora na pontuação de desempenho na classificação, pois a faixa laranja diminuiu em prol de aumento da faixa azul
- O RF apresentaram o melhor desempenho, em ambos os *test size* utilizados, seguido pelo DT com *test size* 1/2;
- O MLP teve aproximadamente o mesmo desempenho em ambas os valores de *test size*, com faixa cinza e faixa laranja bastante proeminentes, o que indica desempenho abaixo do esperado e desqualificando este classificador para a configuração utilizada;
- Nessa configuração, os mais indicados para uso em detecção de *ransomwares* são as DT (*test size* 1/2) e RF (*test size* 1/3 e 1/2).

6.4.2.4 Seção *Network*

Conjunto de dados confeccionado a partir da seção *Network* dos relatórios, contendo as chamadas aos protocolos UDP, HTTP, DNS, TCP e aos Domínios foi transformada em um arquivo de texto, utilizando os mesmos processos das seções mencionadas anteriormente. Além disso, a utilização do INETSIM causou uma pequena alteração na construção dessa seção ao impor limitações de conexão dos *ransomwares* com seus servidores de C&C, pois responde as requisições diretamente do falso *Gateway* da rede, que neste caso é a máquina *host*. Apesar disso, ainda existem várias informações que podem servir como base para mineração de texto e extração de características, principalmente por causa da quantidade de tentativas de acesso e a quais nomes de domínios essas tentativas foram feitas.

Após o processamento das informações, extração de caraterísticas e otimização, o conjunto de dados foi submetido aos classificadores e os resultados estão nas Tabelas 27 e 28. Nessas tabelas, podemos observar muitas métricas zero, tanto na classificação multiclasse quanto em classificação binária. Em todas as situações a concentração principal fica por conta das classificações realizadas com o MLP, porém na classificação multiclasse com *test size* 1/3 e 1/2, há também uma concentração menor nas classificações realizadas pela SVM.

Tabela 27: Tabela com os dados das classificações referente a abordagem de TF-IDF (Network), com *test size* 0,33 e classificação multiclasse.

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
KNN	Clop	1.00	0.29	0.44	0.50	0.29	0.36	0.50	0.29	0.36
	Conti	0.71	0.56	0.63	0.67	0.41	0.51	0.67	0.41	0.51
	Egregor	0.31	0.33	0.32	0.18	0.17	0.17	0.18	0.17	0.17
	Goodware	0.54	0.62	0.58	0.35	0.71	0.47	0.35	0.71	0.47
	LockBit	1.00	0.83	0.91	1.00	0.83	0.91	1.00	0.83	0.91
	MountLocker	1.00	1.00	1.00	0.80	1.00	0.89	0.80	1.00	0.89

Continua na próxima página

Tabela 27 – continuação da página anterior

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	NetWalker	0.88	0.88	0.88	0.96	0.88	0.92	0.96	0.88	0.92
	Revil	0.89	0.95	0.92	0.88	0.88	0.88	0.88	0.88	0.88
	Ryuk	0.86	0.75	0.80	0.67	0.75	0.71	0.67	0.75	0.71
	Accuracy		0.83			0.68			0.77	
SVM	Clop	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Conti	0.65	0.62	0.63	0.65	0.56	0.60	0.90	0.23	0.37
	Egregor	0.75	0.75	0.75	1.00	0.08	0.15	0.00	0.00	0.00
	Goodware	0.77	0.48	0.59	0.82	0.43	0.56	0.75	0.43	0.55
	LockBit	1.00	0.89	0.94	1.00	0.94	0.97	1.00	0.72	0.84
	MountLocker	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
	NetWalker	1.00	0.88	0.94	1.00	0.88	0.94	1.00	0.44	0.61
	Revil	0.88	0.99	0.93	0.83	1.00	0.91	0.69	0.99	0.81
	Ryuk	0.91	0.62	0.74	1.00	0.62	0.77	0.67	0.25	0.36
	Accuracy		0.86			0.84			0.72	
NB	Clop	0.24	0.57	0.33	0.24	0.57	0.33	0.11	0.57	0.19
	Conti	0.62	0.38	0.48	0.34	0.38	0.36	0.32	0.23	0.27
	Egregor	0.12	1.00	0.21	0.12	1.00	0.21	0.09	0.92	0.16
	Goodware	0.25	0.19	0.22	0.25	0.19	0.22	0.53	0.48	0.50
	LockBit	0.84	0.89	0.86	0.70	0.39	0.50	0.67	0.78	0.72
	MountLocker	0.17	1.00	0.30	0.17	1.00	0.30	0.12	1.00	0.22
	NetWalker	0.96	0.88	0.92	0.95	0.80	0.87	0.58	0.28	0.38
	Revil	0.97	0.53	0.68	0.95	0.49	0.65	0.98	0.30	0.46
	Ryuk	0.83	0.62	0.71	0.62	0.31	0.42	0.67	0.62	0.65
	Accuracy		0.56			0.50			0.37	
DT	Clop	1.00	0.29	0.44	1.00	0.29	0.44	1.00	0.29	0.44
	Conti	0.55	0.54	0.55	0.55	0.54	0.55	0.55	0.54	0.55
	Egregor	0.29	0.50	0.36	0.29	0.50	0.36	0.29	0.50	0.36
	Goodware	0.53	0.48	0.50	0.53	0.48	0.50	0.53	0.48	0.50
	LockBit	1.00	0.89	0.94	1.00	0.89	0.94	1.00	0.89	0.94
	MountLocker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	NetWalker	1.00	0.88	0.94	1.00	0.88	0.94	1.00	0.88	0.94
	Revil	0.94	0.95	0.94	0.94	0.95	0.94	0.94	0.95	0.94
	Ryuk	0.72	0.81	0.76	0.72	0.81	0.76	0.72	0.81	0.76
	Accuracy		0.83			0.83			0.83	
RF	Clop	0.67	0.29	0.40	1.00	0.29	0.44	1.00	0.29	0.44
	Conti	0.69	0.56	0.62	0.67	0.56	0.61	0.69	0.46	0.55
	Egregor	0.23	0.25	0.24	0.29	0.33	0.31	0.22	0.17	0.19
	Goodware	0.56	0.48	0.51	0.68	0.62	0.65	0.59	0.62	0.60
	LockBit	0.84	0.89	0.86	0.85	0.94	0.89	0.82	0.78	0.80
	MountLocker	0.80	1.00	0.89	1.00	1.00	1.00	1.00	0.50	0.67
	NetWalker	1.00	0.88	0.94	1.00	0.88	0.94	0.96	0.88	0.92
	Revil	0.89	0.97	0.93	0.91	0.98	0.94	0.86	0.97	0.91
	Ryuk	0.92	0.69	0.79	0.92	0.75	0.83	0.73	0.69	0.71
	Accuracy		0.83			0.85			0.82	
MLP	Clop	0.00	0.00	0.00	1.00	0.14	0.25	0.00	0.00	0.00
	Conti	0.64	0.69	0.67	0.94	0.44	0.60	0.00	0.00	0.00
	Egregor	0.00	0.00	0.00	0.31	0.33	0.32	0.00	0.00	0.00
	Goodware	0.52	0.52	0.52	0.41	0.76	0.53	0.40	0.48	0.43
	LockBit	1.00	0.89	0.94	1.00	0.94	0.97	0.73	0.89	0.80
	MountLocker	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
	NetWalker	0.79	0.88	0.83	0.65	0.88	0.75	1.00	0.80	0.89
	Revil	0.89	0.97	0.93	0.89	0.94	0.92	0.74	0.93	0.82
	Ryuk	0.91	0.62	0.74	1.00	0.25	0.40	0.40	0.50	0.44
	Accuracy		0.83			0.8			0.71	

Tabela 28: Tabela com os dados das classificações referente a abordagem de TF-IDF (Network), com *test size* 0,5 e classificação multiclasse.

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
KNN	Clop	0.50	0.10	0.17	1.00	0.40	0.57	0.33	0.10	0.15
	Conti	0.65	0.51	0.57	0.41	0.48	0.44	0.57	0.41	0.48
	Egregor	0.31	0.44	0.36	0.21	0.39	0.27	0.12	0.22	0.16
	Goodware	0.71	0.48	0.58	0.15	0.35	0.22	0.34	0.52	0.41
	LockBit	1.00	0.83	0.91	1.00	0.92	0.96	0.82	0.75	0.78
	MountLocker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.89	0.94
	NetWalker	0.93	0.90	0.92	1.00	0.88	0.94	0.95	0.90	0.93
	Revil	0.87	0.96	0.91	0.86	0.73	0.79	0.86	0.86	0.86
	Ryuk	0.89	0.74	0.81	0.65	0.57	0.60	0.85	0.74	0.79
	Accuracy		0.83			0.68			0.75	
SVM	Clop	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Conti	0.65	0.49	0.56	0.69	0.48	0.56	0.84	0.34	0.49
	Egregor	0.48	0.67	0.56	1.00	0.11	0.20	0.00	0.00	0.00
	Goodware	0.74	0.45	0.56	0.81	0.42	0.55	0.78	0.45	0.57
	LockBit	1.00	0.88	0.93	1.00	0.92	0.96	0.95	0.75	0.84
	MountLocker	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00

Continua na próxima página

Tabela 28 – continuação da página anterior

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	NetWalker	1.00	0.88	0.94	1.00	0.88	0.94	1.00	0.48	0.65
	Revil	0.86	0.99	0.92	0.81	1.00	0.89	0.70	0.99	0.82
	Ryuk	0.94	0.65	0.77	1.00	0.65	0.79	0.78	0.30	0.44
	Accuracy		0.84			0.83			0.73	
NB	Clop	0.20	0.40	0.27	0.20	0.40	0.27	0.12	0.50	0.19
	Conti	0.73	0.36	0.48	0.40	0.44	0.42	0.46	0.31	0.37
	Egregor	0.12	0.94	0.21	0.12	0.94	0.21	0.09	0.89	0.16
	Goodware	0.21	0.16	0.18	0.21	0.16	0.18	0.36	0.48	0.41
	LockBit	0.84	0.88	0.86	0.73	0.46	0.56	0.56	0.79	0.66
	MountLocker	0.28	1.00	0.44	0.28	1.00	0.44	0.19	0.89	0.31
	NetWalker	1.00	0.88	0.94	1.00	0.83	0.91	0.70	0.45	0.55
	Revil	0.92	0.58	0.71	0.93	0.52	0.67	0.99	0.32	0.48
	Ryuk	0.89	0.70	0.78	0.71	0.43	0.54	0.75	0.65	0.70
	Accuracy		0.59			0.53			0.40	
DT	Clop	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Conti	0.60	0.46	0.52	0.60	0.46	0.52	0.60	0.46	0.52
	Egregor	0.23	0.39	0.29	0.23	0.39	0.29	0.23	0.39	0.29
	Goodware	0.67	0.45	0.54	0.67	0.45	0.54	0.67	0.45	0.54
	LockBit	0.85	0.92	0.88	0.85	0.92	0.88	0.85	0.92	0.88
	MountLocker	1.00	0.89	0.94	1.00	0.89	0.94	1.00	0.89	0.94
	NetWalker	1.00	0.88	0.94	1.00	0.88	0.94	1.00	0.88	0.94
	Revil	0.89	0.97	0.93	0.89	0.97	0.93	0.89	0.97	0.93
	Ryuk	0.95	0.78	0.86	0.95	0.78	0.86	0.95	0.78	0.86
	Accuracy		0.83			0.83			0.83	
RF	Clop	1.00	0.20	0.33	1.00	0.20	0.33	1.00	0.10	0.18
	Conti	0.66	0.51	0.57	0.65	0.52	0.58	0.64	0.48	0.55
	Egregor	0.26	0.39	0.31	0.25	0.39	0.30	0.26	0.33	0.29
	Goodware	0.71	0.48	0.58	0.73	0.52	0.60	0.71	0.48	0.58
	LockBit	1.00	0.88	0.93	1.00	0.92	0.96	0.95	0.83	0.89
	MountLocker	1.00	1.00	1.00	1.00	0.89	0.94	1.00	0.78	0.88
	NetWalker	1.00	0.88	0.94	0.97	0.88	0.93	0.97	0.88	0.93
	Revil	0.88	0.98	0.93	0.89	0.97	0.93	0.85	0.97	0.91
	Ryuk	0.95	0.78	0.86	0.90	0.78	0.84	0.84	0.70	0.76
	Accuracy		0.84			0.84			0.82	
MLP	Clop	0.00	0.00	0.00	1.00	0.40	0.57	0.00	0.00	0.00
	Conti	0.60	0.57	0.59	0.72	0.48	0.57	0.00	0.00	0.00
	Egregor	0.00	0.00	0.00	0.50	0.44	0.47	0.00	0.00	0.00
	Goodware	0.42	0.52	0.46	0.58	0.48	0.53	0.50	0.03	0.06
	LockBit	1.00	0.88	0.93	1.00	0.92	0.96	0.00	0.00	0.00
	MountLocker	0.00	0.00	0.00	1.00	0.89	0.94	0.00	0.00	0.00
	NetWalker	0.95	0.88	0.91	0.69	0.88	0.77	0.90	0.86	0.88
	Revil	0.87	0.99	0.92	0.85	0.96	0.90	0.69	0.99	0.81
	Ryuk	0.94	0.65	0.77	0.70	0.30	0.42	0.38	0.61	0.47
	Accuracy		0.82			0.81			0.68	

Tabela 29: Tabela com os dados das classificações referente a abordagem de TF-IDF (Network), com *test size* 0,33 e classificação Binária.

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
KNN	Clop	0.50	1.00	0.67	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	1.00	0.93	0.96	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.93			1.00			1.00	
	Conti	0.79	0.97	0.87	0.80	0.92	0.85	0.78	1.00	0.87
	Goodware	0.92	0.52	0.67	0.80	0.57	0.67	1.00	0.48	0.65
	Accuracy		0.81			0.80			0.81	
	Egregor	0.57	0.73	0.64	0.53	0.82	0.64	0.53	0.82	0.64
	Goodware	0.88	0.79	0.83	0.91	0.71	0.80	0.91	0.71	0.80
	Accuracy		0.77			0.74			0.74	
	Goodware	1.00	1.00	1.00	0.97	1.00	0.98	0.96	0.96	0.96
	LockBit	1.00	1.00	1.00	1.00	0.92	0.96	0.92	0.92	0.92
	Accuracy		1.00			0.97			0.95	
	Goodware	0.97	1.00	0.98	0.97	1.00	0.98	0.97	1.00	0.98
	MountLocker	1.00	0.50	0.67	1.00	0.50	0.67	1.00	0.50	0.67
	Accuracy		0.97			0.97			0.98	
	Goodware	1.00	1.00	1.00	0.96	1.00	0.98	0.96	1.00	0.98
	NetWalker	1.00	1.00	1.00	1.00	0.96	0.98	1.00	0.96	0.98
	Accuracy		1.00			0.98			0.98	
	Goodware	0.88	0.84	0.86	0.31	0.76	0.44	0.59	0.88	0.71
	Revil	0.98	0.99	0.98	0.96	0.79	0.87	0.98	0.93	0.96
	Accuracy		0.97			0.79			0.92	
	Goodware	1.00	0.96	0.98	0.82	1.00	0.90	0.81	0.89	0.85
	Ryuk	0.93	1.00	0.96	1.00	0.54	0.70	0.70	0.54	0.61
	Accuracy		0.98			0.85			0.78	
	Clop	1.00	1.00	1.00	0.67	1.00	0.80	1.00	1.00	1.00
	Goodware	1.00	1.00	1.00	1.00	0.96	0.98	1.00	1.00	1.00

Continua na próxima página

Tabela 29 – continuação da página anterior

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Accuracy		1.00			0.97			1.00	
	Conti	0.78	1.00	0.87	0.77	0.95	0.85	0.78	1.00	0.87
	Goodware	1.00	0.48	0.65	0.83	0.48	0.61	1.00	0.48	0.65
	Accuracy		0.81			0.78			0.81	
	Egregor	0.50	0.91	0.65	0.50	0.91	0.65	0.48	0.91	0.62
	Goodware	0.95	0.64	0.77	0.95	0.64	0.77	0.94	0.61	0.74
	Accuracy		0.72			0.72			0.69	
	Goodware	0.97	1.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00
	LockBit	1.00	0.92	0.96	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.97			1.00			1.00	
	Goodware	0.97	1.00	0.98	0.97	1.00	0.98	0.97	1.00	0.98
	MountLocker	1.00	0.50	0.67	1.00	0.50	0.67	1.00	0.50	0.67
	Accuracy		0.97			0.97			0.97	
	Goodware	0.92	1.00	0.96	1.00	1.00	1.00	1.00	1.00	1.00
	NetWalker	1.00	0.93	0.96	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.96			1.00			1.00	
	Goodware	1.00	0.84	0.91	1.00	0.84	0.91	0.91	0.84	0.87
	Revil	0.98	1.00	0.99	0.98	1.00	0.99	0.98	0.99	0.99
	Accuracy		0.98			0.98			0.97	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Ryuk	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
NB	Clop	0.08	1.00	0.15	0.08	1.00	0.15	0.50	1.00	0.67
	Goodware	1.00	0.15	0.26	1.00	0.15	0.26	1.00	0.93	0.96
	Accuracy		0.21			0.21			0.97	
	Conti	0.91	0.53	0.67	0.91	0.53	0.67	0.77	0.63	0.70
	Goodware	0.51	0.90	0.66	0.51	0.90	0.66	0.50	0.67	0.57
	Accuracy		0.66			0.66			0.64	
	Egregor	0.29	0.82	0.43	0.29	0.82	0.43	0.43	0.82	0.56
	Goodware	0.75	0.21	0.33	0.75	0.21	0.33	0.89	0.57	0.70
	Accuracy		0.38			0.38			0.64	
	Goodware	1.00	0.96	0.98	1.00	0.96	0.98	1.00	0.82	0.90
	LockBit	0.92	1.00	0.96	0.92	1.00	0.96	0.71	1.00	0.83
	Accuracy		0.97			0.97			0.88	
	Goodware	0.96	0.93	0.95	0.96	0.89	0.93	0.96	0.86	0.91
	MountLocker	0.33	0.50	0.40	0.25	0.50	0.33	0.20	0.50	0.29
	Accuracy		0.90			0.87			0.83	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	0.65	1.00	0.79
	NetWalker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.57	0.73
	Accuracy		1.00			1.00			0.76	
	Goodware	0.74	0.92	0.82	0.74	0.92	0.82	0.14	0.92	0.24
	Revil	0.99	0.96	0.98	0.99	0.96	0.98	0.97	0.32	0.48
	Accuracy		0.96			0.96			0.38	
	Goodware	1.00	0.89	0.94	1.00	0.89	0.94	0.90	1.00	0.95
	Ryuk	0.81	1.00	0.90	0.81	1.00	0.90	1.00	0.77	0.87
	Accuracy		0.93			0.93			0.93	
DT	Clop	0.50	1.00	0.67	0.50	1.00	0.67	0.50	1.00	0.67
	Goodware	1.00	0.93	0.96	1.00	0.93	0.96	1.00	0.93	0.96
	Accuracy		0.93			0.93			0.93	
	Conti	0.83	0.92	0.88	0.83	0.92	0.88	0.83	0.92	0.88
	Goodware	0.82	0.67	0.74	0.82	0.67	0.74	0.82	0.67	0.74
	Accuracy		0.83			0.83			0.83	
	Egregor	0.58	0.64	0.61	0.58	0.64	0.61	0.58	0.64	0.61
	Goodware	0.85	0.82	0.84	0.85	0.82	0.84	0.85	0.82	0.84
	Accuracy		0.77			0.77			0.77	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	LockBit	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	0.97	1.00	0.98	0.97	1.00	0.98	0.97	1.00	0.98
	MountLocker	1.00	0.50	0.67	1.00	0.50	0.67	1.00	0.50	0.67
	Accuracy		0.97			0.97			0.97	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	NetWalker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	0.95	0.84	0.89	0.95	0.84	0.89	0.95	0.84	0.89
	Revil	0.98	1.00	0.99	0.98	1.00	0.99	0.98	1.00	0.99
	Accuracy		0.98			0.98			0.98	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Ryuk	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
RF	Clop	0.67	1.00	0.80	0.67	1.00	0.80	0.40	1.00	0.57
	Goodware	1.00	0.96	0.98	1.00	0.96	0.98	1.00	0.89	0.94
	Accuracy		0.97			0.97			0.90	
	Conti	0.83	0.89	0.86	0.84	1.00	0.92	0.81	1.00	0.89
	Goodware	0.78	0.67	0.72	1.00	0.67	0.80	1.00	0.57	0.73
	Accuracy		0.81			0.88			0.85	
	Egregor	0.60	0.82	0.69	0.56	0.82	0.67	0.54	0.64	0.58
	Goodware	0.92	0.79	0.85	0.91	0.75	0.82	0.85	0.79	0.81
	Accuracy		0.79			0.77			0.74	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.89	0.94
	LockBit	1.00	1.00	1.00	1.00	1.00	1.00	0.80	1.00	0.89
	Accuracy		1.00			1.00			0.93	

Continua na próxima página

Tabela 29 – continuação da página anterior

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Goodware	0.97	1.00	0.98	0.97	1.00	0.98	0.92	0.86	0.89
	MountLocker	1.00	0.50	0.67	1.00	0.50	0.67	0.00	0.00	0.00
	Accuracy		0.97			0.97			0.80	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	NetWalker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	0.95	0.80	0.87	0.95	0.84	0.89	1.00	0.92	0.96
	Revil	0.98	1.00	0.99	0.98	1.00	0.99	0.99	1.00	1.00
	Accuracy		0.97			0.98			0.99	
	Goodware	1.00	0.93	0.96	1.00	1.00	1.00	1.00	0.93	0.96
	Ryuk	0.87	1.00	693	1.00	1.00	1.00	0.87	1.00	0.93
	Accuracy		0.95			1.00			0.95	
	Clop	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.93	1.00	0.96	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.93			1.00			1.00	
	Conti	0.64	1.00	0.78	0.86	0.97	0.91	0.77	0.95	0.85
	Goodware	0.00	0.00	0.00	0.94	0.71	0.81	0.83	0.48	0.61
	Accuracy		0.64			0.88			0.78	
	Egregor	0.00	0.00	0.00	0.57	0.73	0.64	0.57	0.73	0.64
	Goodware	0.72	1.00	0.84	0.88	0.79	0.83	0.88	0.79	0.83
	Accuracy		0.72			0.77			0.77	
	Goodware	0.70	1.00	0.82	1.00	0.89	0.94	1.00	0.86	0.92
	LockBit	0.00	0.00	0.00	0.80	1.00	0.89	0.75	1.00	0.86
MLP	Accuracy		0.70			0.93			0.90	
	Goodware	0.93	1.00	0.97	0.97	1.00	0.98	0.97	1.00	0.98
	MountLocker	0.00	0.00	0.00	1.00	0.50	0.67	1.00	0.50	0.67
	Accuracy		0.93			0.97			0.97	
	Goodware	0.00	0.00	0.00	1.00	1.00	1.00	0.85	1.00	0.92
	NetWalker	0.56	1.00	0.72	1.00	1.00	1.00	1.00	0.86	0.92
	Accuracy		0.56			1.00			0.92	
	Goodware	0.00	0.00	0.00	1.00	0.76	0.86	0.00	0.00	0.00
	Revil	0.89	1.00	0.94	0.97	1.00	0.99	0.89	1.00	0.94
	Accuracy		0.89			0.97			0.89	
	Goodware	0.70	1.00	0.82	1.00	0.96	0.98	1.00	0.96	0.98
	Ryuk	1.00	0.08	0.14	0.93	1.00	0.96	0.93	1.00	0.96
	Accuracy		0.71			0.98			0.98	

Tabela 30: Tabela com os dados das classificações referente a abordagem de TF-IDF (Network), com *test size* 0,5 e classificação Binária.

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Clop	0.60	1.00	0.75	0.67	0.67	0.67	0.67	0.67	0.67
	Goodware	1.00	0.95	0.97	0.98	0.98	0.98	0.98	0.98	0.98
	Accuracy		0.95			0.95			0.95	
	Conti	0.87	0.95	0.90	0.78	0.95	0.85	0.78	0.95	0.85
	Goodware	0.89	0.76	0.82	0.86	0.55	0.67	0.86	0.55	0.67
	Accuracy		0.88			0.80			0.78	
	Egregor	0.67	0.53	0.59	0.58	0.79	0.67	0.58	0.79	0.67
	Goodware	0.80	0.88	0.83	0.88	0.72	0.79	0.88	0.72	0.79
	Accuracy		0.76			0.76			0.75	
	Goodware	0.95	1.00	0.98	0.88	0.90	0.89	0.88	0.90	0.89
	LockBit	1.00	0.90	0.95	0.79	0.75	0.77	0.79	0.75	0.77
KNN	Accuracy		0,97			0.85			0.85	
	Goodware	0.98	1.00	0.99	0.98	1.00	0.99	0.98	1.00	0.99
	MountLocker	1.00	0.75	0.86	1.00	0.75	0.86	1.00	0.75	0.86
	Accuracy		0.98			0.98			0.98	
	Goodware	1.00	1.00	1.00	0.97	1.00	0.99	0.97	1.00	0.99
	NetWalker	1.00	1.00	1.00	1.00	0.97	0.99	1.00	0.97	0.99
	Accuracy		1.00			0.99			0.99	
	Goodware	0.84	0.72	0.78	0.28	0.72	0.4	0.59	0.89	0.71
	Revil	0.97	0.98	0.98	0.96	0.78	0.86	0.99	0.93	0.96
	Accuracy		0.96			0.78			0.93	
	Goodware	1.00	0.93	0.96	0.77	1.00	0.87	0.80	0.97	0.88
	Ryuk	0.88	1.00	0.94	1.00	0.45	0.62	0.92	0.55	0.69
	Accuracy		0.95			0.81			0.82	
	Clop	1.00	1.00	1.00	0.60	1.00	0.75	1.00	1.00	1.00
	Goodware	1.00	1.00	1.00	1.00	0.95	0.97	1.00	1.00	1.00
	Accuracy		1.00			0.95			1.00	
	Conti	0.79	0.96	0.87	0.77	0.93	0.84	0.83	0.95	0.88
	Goodware	0.90	0.58	0.70	0.82	0.55	0.65	0.88	0.67	0.76
	Accuracy		0.82			0.78			0.84	
	Egregor	0.64	0.95	0.77	0.64	0.95	0.77	0.72	0.68	0.70
	Goodware	0.97	0.75	0.85	0.97	0.75	0.85	0.85	0.88	0.86
	Accuracy		0.81			0.81			0.81	
	Goodware	0.95	1.00	0.98	0.98	1.00	0.99	0.98	1.00	0.99

Continua na próxima página

SVM

Tabela 30 – continuação da página anterior

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	LockBit	1.00	0.90	0.9	1.00	0.95	0.97	1.00	0.95	0.97
	Accuracy		0.97			0.98			0.98	
	Goodware	0.98	1.00	0.99	0.98	1.00	0.99	0.98	1.00	0.99
	MountLocker	1.00	0.75	0.86	1.00	0.75	0.86	1.00	0.75	0.86
	Accuracy		0.98			0.98			0.98	
	Goodware	0.95	1.00	0.97	1.00	1.00	1.00	1.00	1.00	1.00
	NetWalker	1.00	0.95	0.97	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.97			1.00			1.00	
	Goodware	1.00	0.75	0.86	1.00	0.78	0.88	0.95	0.56	0.7
	Revil	0.7	1.00	0.99	0.98	1.00	0.99	0.95	1.00	0.97
	Accuracy		0.97			0.98			0.95	
	Goodware	0.87	0.85	0.86	1.00	1.00	1.00	1.00	1.00	1.00
	Ryuk	0.74	0.77	0.76	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.82			1.00			1.00	
NB	Clop	0.08	1.00	0.15	0.08	1.00	0.15	0.38	1.00	0.55
	Goodware	1.00	0.15	0.26	1.00	0.15	0.26	1.00	0.88	0.94
	Accuracy		0.20			0.20			0.89	
	Conti	0.93	0.47	0.63	0.93	0.47	0.63	0.87	0.62	0.72
	Goodware	0.52	0.94	0.67	0.52	0.94	0.67	0.57	0.85	0.68
	Accuracy		0.65			0.65			0.70	
	Egregor	0.35	0.89	0.50	0.35	0.89	0.50	0.61	1.00	0.76
	Goodware	0.80	0.20	0.32	0.80	0.20	0.32	1.00	0.70	0.82
	Accuracy		0.42			0.42			0.8	
	Goodware	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.90	0.94
	LockBit	0.95	0.95	0.95	0.95	0.95	0.95	0.83	0.95	0.88
	Accuracy		0.97			0.98			0.92	
	Goodware	0.97	0.95	0.96	0.97	0.95	0.96	0.97	0.85	0.91
	MountLocker	0.60	0.75	0.67	0.60	0.75	0.67	0.33	0.75	0.46
	Accuracy		0.93			0.93			0.84	
	Goodware	1.00	0.92	0.96	1.00	0.92	0.96	0.66	1.00	0.80
	NetWalker	0.93	1.00	0.96	0.93	1.00	0.96	1.00	0.50	0.67
	Accuracy		0.96			0.96			0.75	
	Goodware	0.74	0.86	0.79	0.74	0.86	0.79	0.16	0.92	0.27
	Revil	0.98	0.97	0.97	0.98	0.97	0.97	0.98	0.44	0.61
	Accuracy		0.95			0.95			0.49	
	Goodware	1.00	0.93	0.96	1.00	0.90	0.95	0.89	1.00	0.94
	Ryuk	0.88	1.00	0.94	0.85	1.00	0.92	1.00	0.77	0.87
	Accuracy		0.95			0.94			0.92	
DT	Clop	0.38	1.00	0.55	0.38	1.00	0.55	0.38	1.00	0.55
	Goodware	1.00	0.88	0.94	1.00	0.88	0.94	1.00	0.88	0.94
	Accuracy		0.89			0.89			0.89	
	Conti	0.77	0.96	0.85	0.77	0.96	0.85	0.77	0.96	0.85
	Goodware	0.89	0.52	0.65	0.89	0.52	0.65	0.89	0.52	0.65
	Accuracy		0.8			0.8			0.8	
	Egregor	0.70	0.74	0.72	0.70	0.74	0.72	0.70	0.74	0.72
	Goodware	0.87	0.85	0.86	0.87	0.85	0.86	0.87	0.85	0.86
	Accuracy		0.81			0.81			0.81	
	Goodware	0.98	1.00	0.99	0.98	1.00	0.99	0.98	1.00	0.99
	LockBit	1.00	0.95	0.97	1.00	0.95	0.97	1.00	0.95	0.97
	Accuracy		0.98			0.98			0.98	
	Goodware	0.98	1.00	0.99	0.98	1.00	0.99	0.98	1.00	0.99
	MountLocker	1.00	0.75	0.86	1.00	0.75	0.86	1.00	0.75	0.86
	Accuracy		0.98			0.98			0.98	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	NetWalker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	0.96	0.69	0.81	0.96	0.69	0.81	0.96	0.69	0.81
	Revil	0.97	1.00	0.98	0.97	1.00	0.98	0.97	1.00	0.98
	Accuracy		0.97			0.97			0.97	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Ryuk	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
RF	Clop	0.50	1.00	0.67	0.75	1.00	0.86	0.40	0.67	0.50
	Goodware	1.00	0.93	0.96	1.00	0.98	0.99	0.97	0.93	0.95
	Accuracy		0.93			0.98			0.91	
	Conti	0.87	0.87	0.87	0.86	0.91	0.88	0.78	0.95	0.85
	Goodware	0.79	0.79	0.79	0.83	0.76	0.79	0.86	0.55	0.67
	Accuracy		0.84			0.85			0.78	
	Egregor	0.70	0.74	0.72	0.78	0.74	0.76	0.65	0.68	0.67
	Goodware	0.87	0.85	0.86	0.88	0.90	0.89	0.85	0.82	0.84
	Accuracy		0.81			0.85			0.78	
	Goodware	0.98	1.00	0.99	0.98	1.00	0.99	0.98	1.00	0.99
	LockBit	1.00	0.95	0.97	1.00	0.95	0.97	1.00	0.95	0.97
	Accuracy		0.98			0.98			0.98	
	Goodware	0.98	1.00	0.99	0.98	1.00	0.99	0.98	1.00	0.99
	MountLocker	1.00	0.75	0.86	1.00	0.75	0.86	1.00	0.75	0.86
	Accuracy		0.98			0.98			0.98	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	NetWalker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	0.96	0.69	0.81	0.96	0.69	0.81	0.93	0.75	0.83
	Revil	0.97	1.00	0.98	0.97	1.00	0.98	0.97	0.99	0.98

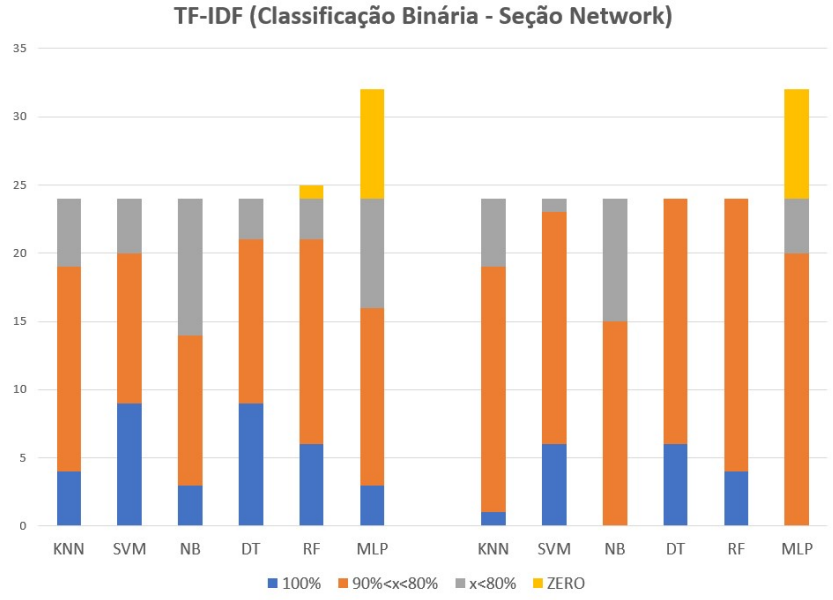
Continua na próxima página

Tabela 30 – continuação da página anterior

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Accuracy		0.97			0.97			0.97	
	Goodware	1.00	0.97	0.99	1.00	1.00	1.00	1.00	0.90	0.95
	Ryuk	0.96	1.00	0.98	1.00	1.00	1.00	0.85	1.00	0.92
	Accuracy		0.98			1.00			0.94	
MLP	Clop	0.00	0.00	0.00	0.50	1.00	0.67	0.60	1.00	0.75
	Goodware	0.93	1.00	0.96	1.00	0.93	0.96	1.00	0.95	0.97
	Accuracy		0.93			0.93			0.95	
	Conti	0.62	1.00	0.77	0.83	0.89	0.86	0.82	0.91	0.86
	Goodware	0.00	0.00	0.00	0.79	0.70	0.74	0.81	0.67	0.73
	Accuracy		0.62			0.82			0.82	
	Egregor	0.00	0.00	0.00	0.75	0.63	0.69	0.71	0.63	0.67
	Goodware	0.75	0.63	0.69	0.84	0.90	0.87	0.83	0.88	0.85
	Accuracy		0.68			0.81			0.8	
	Goodware	0.67	1.00	0.80	0.97	0.95	0.96	0.95	0.95	0.95
	LockBit	0.00	0.00	0.00	0.90	0.95	0.93	0.90	0.90	0.90
	Accuracy		0.67			0.95			0.93	
	Goodware	0.91	1.00	0.95	0.98	1.00	0.99	0.98	1.00	0.99
	MountLocker	0.00	0.00	0.00	1.00	0.75	0.86	1.00	0.75	0.86
	Accuracy		0.91			0.98			0.98	
	Goodware	0.00	0.00	0.00	0.97	1.00	0.99	0.77	1.00	0.87
	NetWalker	0.51	1.00	0.67	1.00	0.97	0.99	1.00	0.71	0.83
	Accuracy		0.51			1.00			0.85	
	Goodware	0.00	0.00	0.00	0.97	0.81	0.88	0.00	0.00	0.00
	Revil	0.90	1.00	0.95	0.98	1.00	0.99	0.90	1.00	0.95
	Accuracy		0.9			0.98			0.90	
	Goodware	0.80	1.00	0.89	1.00	0.95	0.97	1.00	0.93	0.96
	Ryuk	1.00	0.55	0.71	0.92	1.00	0.96	0.88	1.00	0.94
	Accuracy		0.84			0.97			0.95	

A Figura 12, de maneira semelhante ao que fizemos nas Subseções anteriores, apresenta a sumarização dos dados da classificação das Tabelas 29 e 30:

Figura 12: Sumarização dos resultados das Tabelas 29 e 30.



Podemos observar que os classificadores apresentam desempenhos distintos em cada *test size* escolhido:

- Nesta configuração houve mudança significativa no desempenho dos classificadores ao mudarmos o tamanho do *test size* de 1/3 para 1/2. Todos os classificadores apre-

sentaram faixa cinza significativa nas classificações com *test size* 1/3. Nas classificações com *test size* 1/2, os classificadores ainda mantiveram faixa cinza significativa, exceto DT e RF, que não tiveram nenhuma classificação na faixa cinza.

- RF apresentou classificações na faixa laranja com *test size* 1/3, contrariando o histórico com os outros conjuntos de dados, onde manteve bom desempenho. Já o MLP teve classificações significativas em ambos os *test size* utilizados;
- O DT apresentou uma melhora na pontuação de desempenho na classificação, pois a faixa cinza diminuiu ao mesmo tempo que houve um aumento da faixa laranja, apesar de a faixa azul também ter diminuído;
- Nessa configuração, os mais indicados para uso em detecção de *ransomwares* são as DT e RF para *test size* 1/2.

6.4.2.5 Seção *Signatures*

Nesta seção, faremos a avaliação dos resultados da análise da transformação da Seção *Signatures* em um conjunto de dados utilizando a técnica TF-IDF como método de extração de características. A seção *Signatures* é formada a partir de uma lista de comportamentos que o *Cuckoo Sandbox* pode identificar nos *malwares* analisados a partir dos quais calcula uma pontuação e classifica o quanto essa amostra se mostra maliciosa, conforme apresentado na Subseção 2.3.3.1. Os melhores resultados tanto para *test size* 1/3 (*Accuracy* 0.92) quanto para *test size* 1/2 (*Accuracy* 0.92), para a classificação multiclasse, foram alcançados a partir da submissão ao conjunto de dados a SVM e com aplicação do *StandardScaler*. As piores classificações deste conjunto de dados ficaram concentradas no NB, para *test size* 1/3 e 1/2. Ressalta-se que mesmo a aplicação de padronização e seleção de componentes não foi efetivo para essas classificações em particular. Além do NB, o MLP também apresentou classificações com muitas métricas *Precision*, F1 e *Recall* zero, conforme podemos visualizar nas Tabelas 31 e 32.

Podemos ver que esta abordagem pode ser empregada efetivamente para detecção de *malware* em ambiente real, a depender do classificador utilizado. O MLP (nas configurações utilizadas neste trabalho) e NB não são recomendados para tal tarefa.

Tabela 31: Tabela com os dados das classificações referente a abordagem de TF-IDF (Signatures), com *test size* 0,33 e classificação multiclasse.

	Malware	Test Size 0,33								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
KNN	Clop	0.40	0.29	0.33	0.50	0.29	0.36	0.50	0.29	0.36
	Conti	0.65	0.62	0.63	0.66	0.64	0.65	0.68	0.64	0.66
	Egregor	1.00	0.83	0.91	1.00	0.83	0.91	0.92	0.92	0.92
	Goodware	0.87	0.95	0.91	0.86	0.90	0.88	0.78	1.00	0.88
	LockBit	0.80	0.89	0.84	0.81	0.94	0.87	0.85	0.94	0.89
	MountLocker	0.67	1.00	0.80	1.00	1.00	1.00	1.00	0.75	0.86
	NetWalker	0.86	0.76	0.81	0.83	0.80	0.82	0.82	0.92	0.87
	Revil	0.92	0.95	0.93	0.93	0.96	0.94	0.95	0.97	0.96
	Ryuk	0.82	0.56	0.67	0.82	0.56	0.67	1.00	0.50	0.67
	Accuracy		0.86			0.87			0.89	
SVM	Clop	0.33	0.14	0.20	0.40	0.29	0.33	1.00	0.14	0.25
	Conti	0.76	0.82	0.79	0.78	0.90	0.83	0.75	0.85	0.80
	Egregor	1.00	0.92	0.96	1.00	0.92	0.96	1.00	0.92	0.96
	Goodware	0.88	1.00	0.93	0.88	1.00	0.93	0.74	0.95	0.83
	LockBit	1.00	0.94	0.97	0.94	0.94	0.94	0.94	0.89	0.91
	MountLocker	1.00	0.75	0.86	0.80	1.00	0.89	1.00	1.00	1.00
	NetWalker	0.88	0.92	0.90	0.96	0.92	0.94	0.92	0.88	0.90
	Revil	0.94	0.97	0.96	0.98	0.98	0.98	0.93	0.98	0.96
	Ryuk	0.80	0.50	0.62	0.80	0.50	0.62	1.00	0.19	0.32
	Accuracy		0.91			0.93			0.90	
NB	Clop	0.19	0.43	0.26	0.19	0.43	0.26	0.13	0.29	0.18
	Conti	0.52	0.36	0.42	0.52	0.36	0.42	0.67	0.41	0.51
	Egregor	0.14	0.83	0.24	0.14	0.83	0.24	0.32	0.83	0.47
	Goodware	0.88	0.33	0.48	1.00	0.33	0.50	0.22	0.24	0.23
	LockBit	0.40	0.89	0.55	0.41	0.89	0.56	0.58	0.78	0.67
	MountLocker	0.08	0.75	0.15	0.08	0.75	0.15	0.00	0.00	0.00
	NetWalker	0.65	0.60	0.63	0.60	0.60	0.60	0.00	0.00	0.00
	Revil	0.95	0.55	0.69	0.96	0.55	0.70	0.99	0.34	0.51
	Ryuk	0.75	0.38	0.50	0.67	0.38	0.48	0.20	0.31	0.24
	Accuracy		0.54			0.54			0.35	
DT	Clop	0.22	0.29	0.25	0.22	0.29	0.25	0.22	0.29	0.25
	Conti	0.73	0.85	0.79	0.73	0.85	0.79	0.73	0.85	0.79
	Egregor	1.00	0.92	0.96	1.00	0.92	0.96	1.00	0.92	0.96
	Goodware	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86
	LockBit	0.89	0.94	0.92	0.89	0.94	0.92	0.89	0.94	0.92
	MountLocker	0.80	1.00	0.89	0.80	1.00	0.89	0.80	1.00	0.89
	NetWalker	1.00	0.88	0.94	1.00	0.88	0.94	1.00	0.88	0.94
	Revil	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
	Ryuk	0.82	0.56	0.67	0.82	0.56	0.67	0.82	0.56	0.67
	Accuracy		0.91			0.91			0.91	
RF	Clop	0.67	0.29	0.40	0.40	0.29	0.33	0.67	0.29	0.40
	Conti	0.79	0.85	0.81	0.79	0.87	0.83	0.81	0.90	0.85
	Egregor	1.00	0.92	0.96	1.00	0.92	0.96	0.92	0.92	0.92
	Goodware	0.83	0.90	0.86	0.86	0.90	0.88	1.00	0.71	0.83
	LockBit	0.94	0.94	0.94	0.94	0.94	0.94	0.90	1.00	0.95
	MountLocker	0.80	1.00	0.89	0.80	1.00	0.89	1.00	0.75	0.86
	NetWalker	0.96	0.88	0.92	0.96	0.88	0.92	0.92	0.88	0.90
	Revil	0.96	0.99	0.97	0.96	0.99	0.97	0.93	0.99	0.96
	Ryuk	0.83	0.62	0.71	0.90	0.56	0.69	1.00	0.56	0.72
	Accuracy		0.92			0.92			0.91	
MLP	Clop	0.00	0.00	0.00	0.50	0.29	0.36	1.00	0.14	0.25
	Conti	0.27	0.44	0.33	0.79	0.77	0.78	0.74	0.79	0.77
	Egregor	0.00	0.00	0.00	0.79	0.92	0.85	0.85	0.92	0.88
	Goodware	0.00	0.00	0.00	0.95	0.90	0.93	0.68	1.00	0.81
	LockBit	0.00	0.00	0.00	1.00	0.83	0.91	1.00	0.83	0.91
	MountLocker	0.00	0.00	0.00	1.00	0.75	0.86	0.60	0.75	0.67
	NetWalker	0.28	0.52	0.37	0.88	0.88	0.88	0.88	0.88	0.88
	Revil	0.85	0.98	0.91	0.93	0.94	0.94	0.96	0.98	0.97
	Ryuk	1.00	0.06	0.12	0.45	0.56	0.50	1.00	0.38	0.55
	Accuracy		0.67			0.88			0.89	

Tabela 32: Tabela com os dados das classificações referente a abordagem de TF-IDF (Signatures), com *test size* 0,5 e classificação multiclasse.

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
KNN	Clop	0.33	0.30	0.32	0.38	0.30	0.33	0.38	0.30	0.33
	Conti	0.69	0.51	0.58	0.70	0.54	0.61	0.75	0.54	0.63
	Egregor	0.88	0.78	0.82	0.88	0.78	0.82	0.94	0.89	0.91
	Goodware	0.83	0.94	0.88	0.81	0.84	0.83	0.74	0.90	0.81
	LockBit	0.81	0.88	0.84	0.92	0.92	0.92	0.96	0.92	0.94
	MountLocker	0.83	0.56	0.67	1.00	0.56	0.71	0.67	0.44	0.53
	NetWalker	0.88	0.71	0.79	0.82	0.76	0.79	0.91	0.93	0.92
	Revil	0.89	0.97	0.93	0.89	0.97	0.93	0.91	0.98	0.94

Continua na próxima página

Tabela 32 – continuação da página anterior

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Ryuk	0.82	0.61	0.70	0.88	0.65	0.75	0.86	0.52	0.65
	Accuracy		0.85			0.86			0.87	
SVM	Clop	0.50	0.20	0.29	0.60	0.30	0.40	1.00	0.20	0.33
	Conti	0.86	0.79	0.82	0.80	0.85	0.83	0.80	0.66	0.72
	Egregor	0.94	0.89	0.91	0.94	0.89	0.91	1.00	0.89	0.94
	Goodware	0.85	0.90	0.88	0.86	0.97	0.91	0.63	0.87	0.73
	LockBit	1.00	0.92	0.96	1.00	0.92	0.96	0.88	0.88	0.88
	MountLocker	1.00	0.44	0.62	0.83	0.56	0.67	0.80	0.44	0.57
	NetWalker	0.98	0.95	0.96	0.95	0.95	0.95	0.93	0.93	0.93
	Revil	0.91	0.99	0.95	0.95	0.98	0.96	0.90	0.99	0.94
	Ryuk	0.87	0.57	0.68	0.88	0.61	0.72	1.00	0.22	0.36
	Accuracy		0.91			0.92			0.87	
NB	Clop	0.16	0.30	0.21	0.16	0.30	0.21	0.25	0.40	0.31
	Conti	0.29	0.48	0.36	0.28	0.44	0.34	0.77	0.61	0.68
	Egregor	0.42	0.78	0.55	0.42	0.78	0.55	0.44	0.78	0.56
	Goodware	0.95	0.68	0.79	1.00	0.68	0.81	0.17	0.13	0.15
	LockBit	0.38	0.88	0.53	0.38	0.88	0.53	0.31	0.79	0.44
	MountLocker	0.08	0.56	0.14	0.08	0.56	0.14	0.04	0.78	0.07
	NetWalker	0.67	0.62	0.64	0.62	0.62	0.62	0.07	0.02	0.04
	Revil	0.95	0.55	0.70	0.96	0.55	0.70	0.97	0.37	0.54
	Ryuk	0.74	0.61	0.67	0.78	0.61	0.68	0.21	0.35	0.26
	Accuracy		0.58			0.57			0.40	
DT	Clop	0.22	0.20	0.21	0.22	0.20	0.21	0.22	0.20	0.21
	Conti	0.70	0.90	0.79	0.70	0.90	0.79	0.70	0.90	0.79
	Egregor	1.00	0.89	0.94	1.00	0.89	0.94	1.00	0.89	0.94
	Goodware	0.86	0.77	0.81	0.86	0.77	0.81	0.86	0.77	0.81
	LockBit	0.94	0.71	0.81	0.94	0.71	0.81	0.94	0.71	0.81
	MountLocker	0.64	0.78	0.70	0.64	0.78	0.70	0.64	0.78	0.70
	NetWalker	0.97	0.81	0.88	0.97	0.81	0.88	0.97	0.81	0.88
	Revil	0.95	0.97	0.96	0.95	0.97	0.96	0.95	0.97	0.96
	Ryuk	0.69	0.48	0.56	0.69	0.48	0.56	0.69	0.48	0.56
	Accuracy		0.89			0.89			0.89	
RF	Clop	0.75	0.30	0.43	0.75	0.30	0.43	0.60	0.30	0.40
	Conti	0.72	0.84	0.77	0.76	0.89	0.82	0.81	0.72	0.77
	Egregor	0.94	0.89	0.91	0.94	0.89	0.91	0.94	0.89	0.91
	Goodware	0.86	0.81	0.83	0.86	0.81	0.83	0.83	0.65	0.73
	LockBit	0.96	0.92	0.94	0.96	0.92	0.94	0.96	0.92	0.94
	MountLocker	1.00	0.44	0.62	1.00	0.56	0.71	1.00	0.44	0.62
	NetWalker	0.95	0.93	0.94	1.00	0.90	0.95	0.95	0.88	0.91
	Revil	0.94	0.98	0.96	0.94	0.98	0.96	0.89	0.99	0.94
	Ryuk	0.93	0.57	0.70	0.83	0.65	0.73	0.87	0.57	0.68
	Accuracy		0.91			0.92			0.88	
MLP	Clop	0.00	0.00	0.00	0.60	0.30	0.40	0.40	0.20	0.27
	Conti	0.22	0.13	0.16	0.82	0.66	0.73	0.78	0.66	0.71
	Egregor	0.00	0.00	0.00	0.83	0.83	0.83	0.89	0.89	0.89
	Goodware	0.80	0.13	0.22	0.90	0.84	0.87	0.63	0.94	0.75
	LockBit	0.00	0.00	0.00	0.91	0.83	0.87	0.84	0.88	0.86
	MountLocker	0.00	0.00	0.00	1.00	0.44	0.62	0.67	0.44	0.53
	NetWalker	0.24	0.48	0.32	0.87	0.81	0.84	0.95	0.83	0.89
	Revil	0.79	0.99	0.88	0.91	0.95	0.93	0.93	0.98	0.95
	Ryuk	0.38	0.26	0.31	0.38	0.65	0.48	0.92	0.52	0.67
	Accuracy		0.66			0.85			0.88	

Tabela 33: Tabela com os dados das classificações referente a abordagem de TF-IDF (Signatures), com *test size* 0,33 e classificação Binária.

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
KNN	Clop	0.00	0.00	0.00	1.00	0.50	0.67	1.00	0.50	0.67
	Goodware	0.93	1.00	0.96	0.96	1.00	0.98	0.96	1.00	0.98
	Accuracy		0.97			0.97			0.97	
	Conti	1.00	0.82	0.90	0.97	0.79	0.87	0.97	0.79	0.87
	Goodware	0.75	1.00	0.86	0.71	0.95	0.82	0.71	0.95	0.82
	Accuracy		0.88			.085			0.85	
	Egregor	1.00	0.91	0.95	0.92	1.00	0.96	0.92	1.00	0.96
	Goodware	0.97	1.00	0.98	1.00	0.96	0.98	1.00	0.96	0.98
	Accuracy		0.97			0.97			0.97	
	LockBit	1.00	0.89	0.94	0.96	0.96	0.96	0.96	0.96	0.96
	Goodware	0.80	1.00	0.89	0.92	0.92	0.92	0.92	0.92	0.92
	Accuracy		0.93			0.95			0.95	
	MountLocker	0.93	1.00	0.97	0.93	1.00	0.97	0.93	1.00	0.97
	Goodware	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Accuracy		0.93			0.93			0.93	
	NetWalker	0.96	1.00	0.98	0.96	1.00	0.98	0.96	1.00	0.98
	Goodware	1.00	0.96	0.98	1.00	0.96	0.98	1.00	0.96	0.98
	Accuracy		0.98			0.98			0.98	

Continua na próxima página

Tabela 33 – continuação da página anterior

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Revil	1.00	1.00	1.00	0.96	0.96	0.96	1.00	0.96	0.98
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			0.99			1.00	
	Ryuk	0.93	1.00	0.97	0.88	1.00	0.93	0.88	1.00	0.93
	Goodware	1.00	0.85	0.92	1.00	0.69	0.82	1.00	0.69	0.82
	Accuracy		0.95			0.9			0.9	
SVM	Clop	0.08	1.00	0.15	1.00	0.50	0.67	1.00	0.50	0.67
	Goodware	1.00	0.19	0.31	0.96	1.00	0.98	0.96	1.00	0.98
	Accuracy		0.93			0.97			0.97	
	Conti	1.00	0.89	0.94	0.84	0.97	0.90	0.84	1.00	0.92
	Goodware	0.84	1.00	0.91	0.93	0.67	0.78	1.00	0.67	0.80
	Accuracy		0.93			0.86			0.88	
	Egregor	1.00	0.91	0.95	1.00	0.91	0.95	1.00	0.91	0.95
	Goodware	0.97	1.00	0.98	0.97	1.00	0.98	0.97	1.00	0.98
	Accuracy		0.97			0.97			0.97	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	LockBit	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	0.97	1.00	0.98	0.93	1.00	0.97	0.97	1.00	0.98
	MountLocker	1.00	0.50	0.67	0.00	0.00	0.00	1.00	0.50	0.67
	Accuracy		0.97			0.93			0.97	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	NetWalker	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	1.00	0.80	0.89	0.96	0.96	0.96	0.93	1.00	0.96
	Revil	0.98	1.00	0.99	1.00	1.00	1.00	1.00	0.99	1.00
	Accuracy		0.98			0.99			0.99	
	Goodware	0.90	1.00	0.95	0.97	1.00	0.98	0.96	0.82	0.88
	Ryuk	1.00	0.77	0.87	1.00	0.92	0.96	0.71	0.92	0.80
	Accuracy		0.93			0.98			0.85	
NB	Clop	0.50	0.50	0.50	0.08	1.00	0.15	0.40	1.00	0.57
	Goodware	0.96	0.96	0.96	1.00	0.19	0.31	1.00	0.89	0.94
	Accuracy		0.24			0.24			0.9	
	Conti	1.00	0.92	0.96	1.00	0.87	0.93	0.95	1.00	0.97
	Goodware	0.88	1.00	0.93	0.81	1.00	0.89	1.00	0.90	0.95
	Accuracy		0.95			0.92			0.97	
	Egregor	0.61	1.00	0.76	0.59	0.91	0.71	1.00	0.82	0.90
	Goodware	1.00	0.75	0.86	0.95	0.75	0.84	0.93	1.00	0.97
	Accuracy		0.82			0.79			0.95	
	Goodware	1.00	0.89	0.94	1.00	0.89	0.94	0.92	0.86	0.89
	LockBit	0.80	1.00	0.89	0.80	1.00	0.89	0.71	0.83	0.77
	Accuracy		0.93			0.93			0.85	
	Goodware	1.00	0.68	0.81	0.92	0.43	0.59	0.96	0.93	0.95
	MountLocker	0.18	1.00	0.31	0.06	0.50	0.11	0.33	0.50	0.40
	Accuracy		0.70			0.43			0.90	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.95	0.98
	NetWalker	1.00	1.00	1.00	1.00	1.00	1.00	0.97	1.00	0.98
	Accuracy		1.00			1.00			0.98	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	0.15	1.00	0.26
	Revil	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.31	0.47
	Accuracy		1.00			1.00			0.38	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.86	0.92
	Ryuk	1.00	1.00	1.00	1.00	1.00	1.00	0.76	1.00	0.87
	Accuracy		1.00			1.00				
DT	Clop	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
	Goodware	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
	Accuracy		0.93			0.93			0.93	
	Conti	0.95	0.97	0.96	0.95	0.97	0.96	0.95	0.97	0.96
	Goodware	0.95	0.90	0.93	0.95	0.90	0.93	0.95	0.90	0.93
	Accuracy		0.95			0.95			0.95	
	Egregor	1.00	0.91	0.95	1.00	0.91	0.95	1.00	0.91	0.95
	Goodware	0.97	1.00	0.98	0.97	1.00	0.98	0.97	1.00	0.98
	Accuracy		0.97			0.97			0.97	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	LockBit	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
	Goodware	0.97	1.00	0.98	0.97	1.00	0.98	0.97	1.00	0.98
	MountLocker	1.00	0.50	0.67	1.00	0.50	0.67	1.00	0.50	0.67
	Accuracy		0.97			0.97			0.97	
	Goodware	0.92	1.00	0.96	0.92	1.00	0.96	0.92	1.00	0.96
	NetWalker	1.00	0.93	0.96	1.00	0.93	0.96	1.00	0.93	0.96
	Accuracy		0.96			0.96			0.96	
	Goodware	0.89	0.96	0.92	0.89	0.96	0.92	0.89	0.96	0.92
	Revil	1.00	0.99	0.99	1.00	0.99	0.99	1.00	0.99	0.99
	Accuracy		0.98			0.98			0.98	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Ryuk	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		1.00			1.00			1.00	
RF	Clop	0.50	1.00	0.67	0.50	1.00	0.67	1.00	1.00	1.00
	Goodware	1.00	0.93	0.96	1.00	0.93	0.96	1.00	1.00	1.00
	Accuracy		0.93			0.93			1.00	
	Conti	1.00	0.95	0.97	1.00	0.95	0.97	0.88	1.00	0.94

Continua na próxima página

Tabela 33 – continuação da página anterior

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Goodware	0.91	1.00	0.95	0.91	1.00	0.95	1.00	0.76	0.86
	Accuracy		0.97			0.97			0.92	
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.91	0.95
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	0.97	1.00	0.98
	Accuracy		1.00			1.00			0.97	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.98
	LockBit	1.00	1.00	1.00	1.00	1.00	1.00	0.92	1.00	0.96
	Accuracy		1.00			1.00			0.97	
	Goodware	0.97	1.00	0.98	0.97	1.00	0.98	0.96	0.96	0.96
	MountLocker	1.00	0.50	0.67	1.00	0.50	0.67	0.50	0.50	0.50
	Accuracy		0.97			0.97			0.93	
	Goodware	0.92	1.00	0.96	0.96	1.00	0.98	1.00	1.00	1.00
	NetWalker	1.00	0.93	0.96	1.00	0.96	0.98	1.00	1.00	1.00
	Accuracy		0.96			0.98			1.00	
	Goodware	0.96	1.00	0.98	1.00	1.00	1.00	0.92	0.92	0.92
	Revil	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99
	Accuracy		1.00			1.00			0.98	
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.79	0.88
	Ryuk	1.00	1.00	1.00	1.00	1.00	1.00	0.68	1.00	0.81
	Accuracy		1.00			1.00			0.85	
MLP	Clop	0.00	0.00	0.00	0.25	0.50	0.33	0.67	1.00	0.80
	Goodware	0.93	1.00	0.96	0.96	0.89	0.92	1.00	0.96	0.98
	Accuracy		0.93			0.86			0.97	
	Conti	0.65	0.97	0.78	0.90	1.00	0.95	0.86	1.00	0.93
	Goodware	0.50	0.05	0.09	1.00	0.81	0.89	1.00	0.71	0.83
	Accuracy		0.64			0.93			0.9	
	Egregor	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.72	1.00	0.84	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.72			1.00			1.00	
	Goodware	0.96	0.96	0.96	1.00	1.00	1.00	1.00	1.00	1.00
	LockBit	0.92	0.92	0.92	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.95			1.00			1.00	
	Goodware	0.93	1.00	0.97	0.93	1.00	0.97	0.97	1.00	0.98
	MountLocker	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.50	0.67
	Accuracy		0.93			0.93			0.97	
	Goodware	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
	NetWalker	0.56	.00	0.72	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.56			1.00			1.00	
	Goodware	0.00	0.00	0.00	1.00	0.96	0.98	0.96	0.96	0.96
	Revil	0.89	1.00	0.94	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.89			1.00			0.99	
	Goodware	1.00	0.36	0.53	0.96	0.89	0.93	0.96	0.93	0.95
	Ryuk	0.42	1.00	0.59	0.80	0.92	0.86	0.86	0.92	0.89
	Accuracy		0.56			0.9			0.93	

Tabela 34: Tabela com os dados das classificações referente a abordagem de TF-IDF (Signatures), com *test size* 0,5 e classificação Binária.

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
KNN	Clop	1.00	0.67	0.80	1.00	0.67	0.80	1.00	0.67	0.80
	Goodware	0.98	1.00	0.99	0.98	1.00	0.99	0.98	1.00	0.99
	Accuracy		0.98			0.98			0.98	
	Conti	1.00	0.75	0.85	0.97	0.71	0.82	0.97	0.71	0.82
	Goodware	0.70	1.00	0.82	0.67	0.97	0.79	0.67	0.97	0.79
	Accuracy		0.84			0.81			0.81	
	Egregor	1.00	0.95	0.97	0.89	0.84	0.86	0.89	0.84	0.86
	Goodware	0.98	1.00	0.99	0.93	0.95	0.94	0.93	0.95	0.94
	Accuracy		0.98			0.92			0.92	
	LockBit	0.97	0.93	0.95	0.95	0.95	0.95	0.95	0.95	0.95
	Goodware	0.86	0.95	0.90	0.90	0.90	0.90	0.90	0.90	0.90
	Accuracy		0.95			0.93			0.93	
	MountLocker	0.95	1.00	0.98	0.95	1.00	0.98	0.95	1.00	0.98
	Goodware	1.00	0.50	0.67	1.00	0.50	0.67	1.00	0.50	0.67
	Accuracy		0.96			0.96			0.96	
	NetWalker	0.88	1.00	0.94	0.92	0.97	0.95	0.92	0.97	0.95
	Goodware	1.00	0.87	0.93	0.97	0.92	0.95	0.97	0.92	0.95
	Accuracy		0.93			0.95			0.95	
	Revil	1.00	0.92	0.96	0.96	0.72	0.83	0.96	0.67	0.79
	Goodware	0.99	1.00	1.00	0.97	1.00	0.98	0.96	1.00	0.98
	Accuracy		0.99			0.97			0.96	
	Ryuk	0.91	1.00	0.95	0.87	1.00	0.93	0.87	1.00	0.93
	Goodware	1.00	0.82	0.90	1.00	0.73	0.84	1.00	0.73	0.84
	Accuracy		0.94			0.9			0.9	
	Clop	1.00	0.33	0.50	1.00	0.67	0.80	1.00	0.67	0.80
	Goodware	0.95	1.00	0.98	0.98	1.00	0.99	0.98	1.00	0.99

continua na próxima página

Tabela 34 – continuação da página anterior

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Accuracy		0.95			0.98			0.98	
	Conti	1.00	0.91	0.95	0.88	0.91	0.89	0.89	1.00	0.94
	Goodware	0.87	1.00	0.93	0.84	0.79	0.81	1.00	0.79	0.88
	Accuracy		0.94			0.86			0.92	
	Egregor	1.00	0.95	0.97	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.98	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.98			1.00			1.00	
	Goodware	0.98	1.00	0.99	0.98	1.00	0.99	0.98	1.00	0.99
	LockBit	1.00	0.95	0.97	1.00	0.95	0.97	1.00	0.95	0.97
	Accuracy		0.98			0.98			0.98	
	Goodware	0.98	1.00	0.99	0.95	1.00	0.98	0.98	1.00	0.99
	MountLocker	1.00	0.75	0.86	1.00	0.50	0.67	1.00	0.75	0.86
	Accuracy		0.98			0.96			0.98	
	Goodware	0.97	1.00	0.99	0.97	1.00	0.99	0.97	1.00	0.99
	NetWalker	1.00	0.97	0.99	1.00	0.97	0.99	1.00	0.97	0.99
	Accuracy		0.99			0.99			0.99	
	Goodware	1.00	0.81	0.89	0.97	0.92	0.94	0.90	0.72	0.80
	Revil	0.98	1.00	0.99	0.99	1.00	0.99	0.97	0.99	0.98
	Accuracy		0.98			0.99			0.96	
	Goodware	0.98	1.00	0.99	0.93	1.00	0.96	0.97	0.82	0.89
	Ryuk	1.00	0.95	0.98	1.00	0.86	0.93	0.75	0.95	0.84
	Accuracy		0.92			0.95			0.87	
NB	Clop	0.09	1.00	0.16	0.09	1.00	0.16	0.20	0.33	0.25
	Goodware	1.00	0.22	0.36	1.00	0.22	0.36	0.95	0.90	0.92
	Accuracy		0.27			0.27			0.95	
	Conti	1.00	0.87	0.93	1.00	0.82	0.90	0.98	0.98	0.98
	Goodware	0.82	1.00	0.90	0.77	1.00	0.87	0.97	0.97	0.97
	Accuracy		0.92			0.89			0.98	
	Egregor	0.68	1.00	0.81	0.67	0.95	0.78	1.00	0.79	0.88
	Goodware	1.00	0.78	0.87	0.97	0.78	0.86	0.91	1.00	0.95
	Accuracy		0.85			0.83			0.93	
	Goodware	1.00	0.93	0.96	1.00	0.93	0.96	0.95	0.90	0.92
	LockBit	0.87	1.00	0.93	0.87	1.00	0.93	0.82	0.90	0.86
	Accuracy		0.95			0.95			0.9	
	Goodware	0.98	1.00	0.99	0.98	0.98	0.98	0.97	0.95	0.96
	MountLocker	1.00	0.75	0.86	0.75	0.75	0.75	0.60	0.75	0.67
	Accuracy		0.98			0.96			0.93	
	Goodware	0.97	0.97	0.97	0.93	0.70	0.80	0.97	0.97	0.97
	NetWalker	0.97	0.97	0.97	0.77	0.95	0.85	0.97	0.97	0.97
	Accuracy		0.97			0.83			0.97	
	Goodware	1.00	0.97	0.99	0.97	0.89	0.93	0.15	1.00	0.26
	Revil	1.00	1.00	1.00	0.99	1.00	0.99	1.00	0.36	0.53
	Accuracy		1.00			0.99			0.43	
	Goodware	0.98	1.00	0.99	0.98	1.00	0.99	1.00	0.85	0.92
	Ryuk	1.00	0.95	0.98	1.00	0.95	0.98	0.79	1.00	0.88
	Accuracy		0.98			0.98			0.90	
DT	Clop	0.20	0.33	0.25	0.20	0.33	0.25	0.20	0.33	0.25
	Goodware	0.95	0.90	0.92	0.95	0.90	0.92	0.95	0.90	0.92
	Accuracy		0.86			0.86			0.86	
	Conti	1.00	0.96	0.98	1.00	0.96	0.98	1.00	0.96	0.98
	Goodware	0.94	1.00	0.97	0.94	1.00	0.97	0.94	1.00	0.97
	Accuracy		0.98			0.98			0.98	
	Egregor	1.00	0.95	0.97	1.00	0.95	0.97	1.00	0.95	0.97
	Goodware	0.98	1.00	0.99	0.98	1.00	0.99	0.98	1.00	0.99
	Accuracy		0.98			0.98			0.98	
	Goodware	0.98	1.00	0.99	0.98	1.00	0.99	0.98	1.00	0.99
	LockBit	1.00	0.95	0.97	1.00	0.95	0.97	1.00	0.95	0.97
	Accuracy		0.98			0.98			0.98	
	Goodware	0.98	1.00	0.99	0.98	1.00	0.99	0.98	1.00	0.99
	MountLocker	1.00	0.75	0.86	1.00	0.75	0.86	1.00	0.75	0.86
	Accuracy		0.98			0.98			0.98	
	Goodware	0.94	0.89	0.92	0.94	0.89	0.92	0.94	0.89	0.92
	NetWalker	0.90	0.95	0.92	0.90	0.95	0.92	0.90	0.95	0.92
	Accuracy		0.92			0.92			0.92	
	Goodware	0.86	0.83	0.85	0.86	0.83	0.85	0.86	0.83	0.85
	Revil	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
	Accuracy		0.97			0.97			0.97	
	Goodware	0.98	1.00	0.99	0.98	1.00	0.99	0.98	1.00	0.99
	Ryuk	1.00	0.95	0.98	1.00	0.95	0.98	1.00	0.95	0.98
	Accuracy		0.98			0.98			0.98	
RF	Clop	0.33	0.67	0.44	0.38	1.00	0.55	0.75	1.00	0.86
	Goodware	0.97	0.90	0.94	1.00	0.88	0.94	1.00	0.98	0.99
	Accuracy		0.89			0.89			0.98	
	Conti	1.00	0.96	0.98	1.00	0.96	0.98	0.98	1.00	0.99
	Goodware	0.94	1.00	0.97	0.94	1.00	0.97	1.00	0.97	0.98
	Accuracy		0.98			0.98			0.99	
	Egregor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.91	0.95
	Goodware	1.00	1.00	1.00	1.00	1.00	1.00	0.97	1.00	0.98
	Accuracy		1.00			1.00			0.98	
	Goodware	0.98	1.00	0.99	0.98	1.00	0.99	0.97	0.97	0.97
	LockBit	1.00	0.95	0.97	1.00	0.95	0.97	0.95	0.95	0.95
	Accuracy		0.98			0.98			0.97	

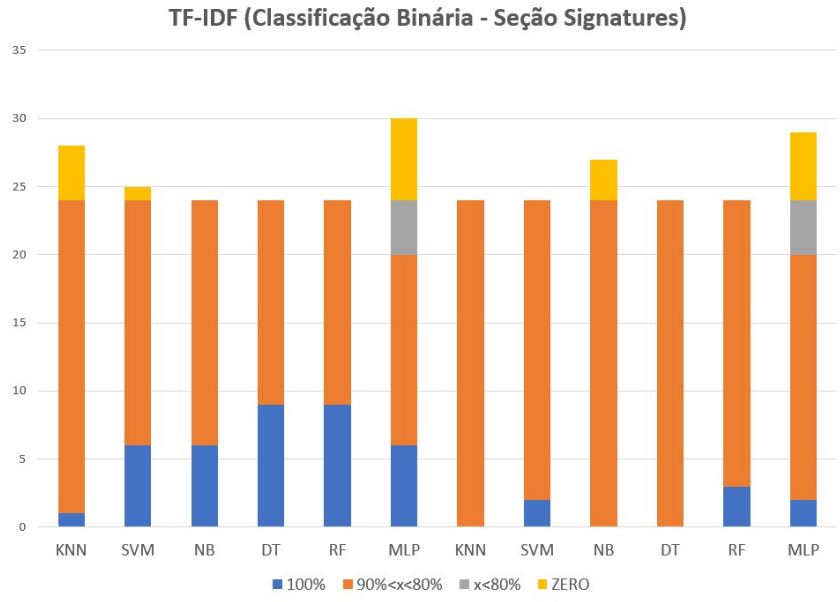
continua na próxima página

Tabela 34 – continuação da página anterior

	Malware	Test Size 0,5								
		Normal			StandardScaler			PCA		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	Goodware	0.98	1.00	0.99	1.00	1.00	1.00	0.95	0.98	0.96
	MountLocker	1.00	0.75	0.86	1.00	1.00	1.00	0.67	0.50	0.57
	Accuracy		0.98			1.00			0.93	
	Goodware	0.97	1.00	0.99	0.97	1.00	0.99	0.97	1.00	0.99
	NetWalker	1.00	0.97	0.99	1.00	0.97	0.99	1.00	0.97	0.99
	Accuracy		0.99			0.99			0.99	
	Goodware	0.97	0.92	0.94	0.97	0.86	0.91	0.96	0.69	0.81
	Revil	0.99	1.00	0.99	0.98	1.00	0.99	0.97	1.00	0.98
	Accuracy		0.99			0.98			0.97	
	Goodware	0.97	0.93	0.95	0.97	0.93	0.95	0.97	0.85	0.91
	Ryuk	0.88	0.95	0.91	0.88	0.95	0.91	0.78	0.95	0.86
	Accuracy		0.94			0.94			0.89	
	Clop	1.00	0.67	0.80	0.40	0.67	0.50	0.67	0.67	0.67
	Goodware	0.98	1.00	0.99	0.97	0.93	0.95	0.98	0.98	0.98
	Accuracy		0.93			0.91			0.95	
	Conti	0.64	0.98	0.78	0.93	0.96	0.95	0.90	1.00	0.95
	Goodware	0.75	0.09	0.16	0.94	0.88	0.91	1.00	0.82	0.90
	Accuracy		0.65			0.93			0.93	
	Egregor	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
	Goodware	0.68	1.00	0.81	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy		0.68			1.00			1.00	
	Goodware	0.95	0.88	0.91	0.98	1.00	0.99	0.95	1.00	0.98
	LockBit	0.78	0.90	0.84	1.00	0.95	0.97	1.00	0.90	0.95
MLP	Accuracy		0.88			0.98			0.97	
	Goodware	0.91	1.00	0.95	0.95	1.00	0.98	0.98	1.00	0.99
	MountLocker	0.00	0.00	0.00	1.00	0.50	0.67	1.00	0.75	0.86
	Accuracy		0.91			0.96			0.98	
	Goodware	0.00	0.00	0.00	0.97	1.00	0.99	0.97	1.00	0.99
	NetWalker	0.51	1.00	0.67	1.00	0.97	0.99	1.00	0.97	0.99
	Accuracy		0.51			0.99			0.99	
	Goodware	0.00	0.00	0.00	1.00	0.86	0.93	1.00	0.92	0.96
	Revil	0.90	1.00	0.95	0.98	1.00	0.99	0.99	1.00	1.00
	Accuracy		0.9			0.99			0.99	
	Goodware	0.00	0.00	0.00	0.95	0.93	0.94	0.93	0.93	0.93
	Ryuk	0.35	1.00	0.52	0.87	0.91	0.89	0.86	0.86	0.86
	Accuracy		0.35			0.92			0.90	

A Figura 12, de maneira semelhante ao que fizemos nas Subseções anteriores, apresenta a sumarização dos dados da classificação das Tabelas 33 e 34:

Figura 13: Sumarização dos resultados das Tabelas 33 e 34.



Podemos observar que os classificadores apresentam desempenhos distintos em cada

test size escolhido:

- Nesta configuração, o NB apresentou bom desempenho com *test size* 1/3, porém com *test size* 1/2 houve queda no desempenho (zero classificações na faixa azul e classificações na faixa amarela);
- O KNN teve melhora no desempenho, visto que com *test size* 1/3, apresentou algumas classificações na região amarela, apesar de ter passado de uma classificação na região azul para nenhuma na mudança de *test size*;
- O MLP apresentou desempenho abaixo do esperado para ambos os valores de *test size*, com faixas significativas tanto na região cinza quanto na amarela;
- Nessa configuração, os mais indicados para uso em detecção de *ransomwares* são as DT e RF para *test size* 1/2, seguidos pelo NB na mesma configuração.

Diante dos resultados, percebemos que os algoritmos de classificação baseados em Árvores (DT e RF) são os que apresentam melhor desempenho em todas as configurações propostas, tornando-se assim os melhores candidatos a comporem um sistema de tempo-real de detecção de *ransomware*.

7 Conclusão

O uso do *Cuckoo Sandbox* e ambientes virtualizados para análise de *malware* se provou como uma ferramenta valiosa no campo da segurança cibernética. Os relatórios detalhados gerados pelo *Cuckoo* fornecem uma riqueza de informações sobre o comportamento e as características do *malware*, que podem ser usadas para entender melhor e se defender contra ameaças. Ao construir um conjunto de dados a partir da análise desses relatórios, foi possível alimentar essas informações em classificadores de Aprendizado de Máquina. Neste trabalho, propusemos duas abordagens para a geração de conjuntos de dados para classificação de *ransomwares*. A partir dos conjuntos de dados gerados a partir de duas abordagens de extração de características (Quantidade de Chamadas de API e TF-IDF), utilizamos classificadores de Aprendizado de Máquina (DT, RF, KNN, NB e MLP). Ao usar esses classificadores, é possível identificar com rapidez e precisão padrões e tendências nos dados, que podem ser usados para entender e classificar melhor os diferentes tipos de *malware*. Isso pode ser particularmente útil na identificação de ameaças novas e emergentes, bem como no desenvolvimento de estratégias de defesa. Outro fato importante, é que durante a pesquisa, encontramos alguns trabalhos que versam sobre artefatos anti-VM em *malwares*, pois nos deparamos com o fato de que abordagens baseadas em Análise Dinâmica não são eficazes contra *ransomwares* com capacidade anti-VM. Este conhecimento se mostrou importante na construção do ambiente de análise cujo resultado foi apresentado na Seção 5.6.

As técnicas de classificação baseadas em estrutura de árvore (RF e DT) fora as que obtiveram os resultados mais promissores, considerando as métricas *Precision*, *Recall* e F1. Portanto, essas técnicas podem ser consideradas técnicas candidatas para a construção de um sistema dinâmico e eficaz de detecção de *ransomware* para amostras novas ou já conhecidas. Nos testes de classificação executados, o RF e o DT foram os que mais apresentaram classificação binária 100% e os que melhor conseguiram discernir entre as diversas famílias de *ransomwares* analisadas sob a ótica das duas abordagens utilizadas. Há que se considerar também que as abordagens utilizadas para extração de dados das

análises de *ransomware* e construção dos conjuntos de dados realizadas neste trabalho, permitiram que os classificadores diferenciassem tanto as famílias entre si quanto cada uma contra as amostras benignas.

7.1 Limitações do Trabalho

Como uma das principais limitações deste trabalho, temos a seleção das famílias a serem estudadas, onde foram considerados *ransomwares* mais ativos nos anos de 2021 e 2022. Como a atividade de um *ransomware* não está necessariamente relacionada com a quantidade de amostras geradas, algumas famílias desse grupo tiveram poucas amostras para estudo. O principal motivo de tal fato ter ocorrido é que o acesso a repositórios que permitam *download* de amostras sem custos e tenham grande quantidade de amostras (por exemplo, o VT somente permite que empresas associadas baixem amostras de seu repositório). Este fato nos obrigou a buscar amostras em outros repositórios, que não possuem o mesmo acervo do VT, reduzindo a quantidade de amostras baixadas e impactando no desempenho dos classificadores utilizados. Considerando esses dois fatos e olhando os resultados obtidos, vemos que as famílias com baixa quantidade de amostras foram as que apresentaram mais classificações zero (*Clop* e *MountLocker*, com 15 e 17 amostras, respectivamente). O que acontece é que nesses casos, mesmo errando completamente a classificação do ransomware, a métrica *Accuracy* permanece alta pois os *Goodwares* representam a maior parte dos dados utilizados na classificação. Apesar disso, pudemos transpor essa situação para o mundo real, em que há ocorrência de sistemas em que a maioria dos dados analisados são de amostras benignas e poucas são de amostras maliciosas. Além disso, como foi utilizado *GridSearchCV* para busca de melhoras parâmetros para cada classificador, o tempo de computação das classificações é o tempo geral das combinações de parâmetros utilizadas. Para que obtivéssemos o tempo de execução específico da melhor combinação de parâmetros teríamos que executar todas as classificações com esses parâmetros encontrados, tornando o processo muito mais demorado e que demandaria mais tempo de pesquisa. Além disso, consideramos que o tempo de execução seria um dado importante para o próximo passo dessa pesquisa, que será a implementação de um sistema de detecção de *ransomware* em tempo real em ambientes *Microsoft Windows*, por este motivo tal métrica não foi considerada.

7.2 Trabalhos Futuros

Como trabalhos futuros, sugere-se:

- Realizar uma análise da efetividade dos parâmetros utilizados nas classificações de modo a melhorar a taxa de acerto dos classificadores, principalmente o MLP.
- Implementar um sistema de detecção e prevenção de *ransomware* baseado nas abordagens descritas neste trabalho.
- Analisar o tempo de execução dos algoritmos no sistema de detecção implementado.
- Estender os estudos para classificação de *malwares* que não sejam *ransomwares*.

REFERÊNCIAS

ADVISOR, CISO. **Ransomware põe 4 portos da África do Sul em modo manual**. [S. l.: s. n.], 2021. Disponível em: <https://www.cisoadvisor.com.br/ransomware-poe-4-portos-da-africa-do-sul-em-operacao-manual/>. Acesso em: 2 ago. 2021.

AHLGREN, Filip. **Local And Network Ransomware Detection Comparison**. [S. l.: s. n.], 2019.

ALSABEH, Ali et al. Exploiting ransomware paranoia for execution prevention. In: IEEE. ICC 2020-2020 IEEE International Conference on Communications (ICC). [S. l.: s. n.], 2020. p. 1–6.

BHAGWAT, Laxmi B; PATIL, Balaji M. Detection of Ransomware on Windows System Using Machine Learning Technique: Experimental Results. In: SPRINGER. INTERNATIONAL Advanced Computing Conference. [S. l.: s. n.], 2020. p. 417–423.

BLACK, Paul et al. API based discrimination of ransomware and benign cryptographic programs. In: SPRINGER. INTERNATIONAL Conference on Neural Information Processing. [S. l.: s. n.], 2020. p. 177–188.

BLACKFOG. **The State of Ransomware in 2022**. [S. l.: s. n.], 2022. Disponível em: <https://www.blackfog.com/the-state-of-ransomware-in-2022/>. Acesso em: 7 jun. 2022.

BLEEPINGCOMPTER. **Clop ransomware is back in business after recent arrests**. [S. l.: s. n.], 2021. Disponível em: <https://www.bleepingcomputer.com/news/security/marine-services-provider-swire-pacific-offshore-hit-by-ransomware/>. Acesso em: 4 mai. 2022.

BLEEPINGCOMPUTER. **Shared Code Links Sodinokibi to GandCrab, Minus the Fun & Games**. [S. l.: s. n.], 2019. Disponível em: <https://www.bleepingcomputer.com/news/security/shared-code-links-sodinokibi-to-gandcrab-minus-the-fun-and-games/>. Acesso em: 8 jul. 2022.

BROWN, Lawrie; STALLINGS, William. **Segurança de computadores: princípios e práticas**. [S. l.]: Elsevier Brasil, 2017. v. 2.

BURT, Jeff. **US treasury whips up sanctions for crypto mixer Tornado Cash**. en. [S. l.]: The register, 2022. Disponível em: <https://www.theregister.com/2022/08/08/treasury_sanctions_tornado_cash_korea/>. Acesso em: 16 ago. 2022.

CALDAS, Daniel Mendes. **Análise e extração de características estruturais e comportamentais para perfis de malware**. [S. l.: s. n.], 2016.

CHEN, Li et al. Towards resilient machine learning for ransomware detection. **arXiv preprint arXiv:1812.09400**, 2018.

CHEN, Qian et al. Automated ransomware behavior analysis: Pattern extraction and early detection. In: SPRINGER. INTERNATIONAL Conference on Science of Cyber Security. [S. l.: s. n.], 2019. p. 199–214.

CISA. **Alert (AA21-265A) - Conti Ransomware**. [S. l.: s. n.], 2022. Disponível em: <https://www.cisa.gov/uscert/ncas/alerts/aa21-265a>. Acesso em: 30 mar. 2022.

CLARKE, Richard A; KNAKE, Robert K. **Guerra cibernética: a próxima ameaça à segurança e o que fazer a respeito**. [S. l.]: Brasport, 2015.

COMPTER, Bleeping. **LockBit victim estimates cost of ransomware attack to be $42 million**. [S. l.: s. n.], 2022. Disponível em: <https://www.bleepingcomputer.com/news/security/ten-notorious-ransomware-strains-put-to-the-encryption-speed-test/>. Acesso em: 7 jun. 2022.

COMPUTERWORLD. **Microsoft confirms that XP contains random number generator bug**. [S. l.: s. n.], 2007. Disponível em: <https://www.computerworld.com/article/2539986/microsoft-confirms-that-xp-contains-random-number-generator-bug.html>. Acesso em: 7 jul. 2022.

CRACIUN, Vlad Constantin; MOGAGE, Andrei; SIMION, Emil. Trends in design of ransomware viruses. In: SPRINGER. INTERNATIONAL Conference on Security for Information Technology and Communications. [S. l.: s. n.], 2018. p. 259–272.

CURTIN, Ryan R et al. Detecting DGA domains with recurrent neural networks and side information. In: PROCEEDINGS of the 14th International Conference on Availability, Reliability and Security. [S. l.: s. n.], 2019. p. 1–10.

CYBEREASON. **Cl0p Ransomware Gang Tries to Topple the House of Cards**. [S. l.: s. n.], 2021. Disponível em: <https://www.cybereason.com/blog/ceo-series/cl0p-ransomware-gang-tries-to-topple-the-house-of-cards>. Acesso em: 4 mai. 2022.

DARSHAN, SL Shiva; KUMARA, MA Ajay; JAIDHAR, CD. Windows malware detection based on cuckoo sandbox generated report using machine learning algorithm. In: IEEE. 2016 11th International Conference on Industrial and Information Systems (ICIIS). [S. l.: s. n.], 2016. p. 534–539.

DE, Nikhilesh. **Crypto-Mixing Service Tornado Cash Blacklisted by US Treasury**. en. [S. l.]: Coindesk, 2022. Disponível em: <https://www.coindesk.com/policy/2022/08/08/crypto-mixing-service-tornado-cash-blacklisted-by-us-treasury/>. Acesso em: 16 ago. 2022.

DIFFIE, Whitfield; HELLMAN, Martin E. New directions in cryptography. In: SECURE communications and asymmetric cryptosystems. [S. l.]: Routledge, 2019. p. 143–180.

DINH, Phai Vu et al. Behaviour-aware malware classification: Dynamic feature selection. In: IEEE. 2019 11th International Conference on Knowledge and Systems Engineering (KSE). [S. l.: s. n.], 2019. p. 1–5. DOI: 10.1109/KSE.2019.8919491.

DWORKIN, Morris et al. **Advanced Encryption Standard (AES)**. en. [S. l.]: Federal Inf. Process. Stds. (NIST FIPS), National Institute of Standards e Technology, Gaithersburg, MD, 2001. DOI: https://doi.org/10.6028/NIST.FIPS.197.

EGUNJOBI, Samuel; PARKINSON, Simon; CRAMPTON, Andrew. Classifying ransomware using machine learning algorithms. In: SPRINGER. INTERNATIONAL Conference on Intelligent Data Engineering and Automated Learning. [S. l.: s. n.], 2019. p. 45–52.

ENIGMASOFT. **Egregor Ransomware**. [S. l.: s. n.], 2020. Disponível em: <https://www.enigmasoftware.com/egregorransomware-removal/?gclid=CjwKCAjw0a-SBhBkEiwApljU0vivsrBHWmaEye7ZGZ-Cg05kk0MhXe7dtl0NhpXf1ngnCiGAP6TWihoCPygQAvD_BwE>. Acesso em: 5 abr. 2022.

ESENTIRE. **Conti Ransomware Gang Claims 50+ New Victims including Oil Terminal Operator Sea-Invest Disrupting Operations at 24 Seaports Across Europe and Africa**. [S. l.: s. n.], 2022. Disponível em: <https://www.esentire.com/security-advisories/conti-ransomware-gang-

claims-50-new-victims-including-oil-terminal-operator-sea-invest>. Acesso em: 30 mar. 2022.

EUA vão impor sanções a corretora de criptomoedas que viabilizou pagamentos de ransomware. en. [S. l.]: Forbes, 2021. Disponível em: <https://forbes.com.br/forbes-money/2021/09/eua-vao-impor-sancoes-a-corretora-de-criptomoedas-que-viabilizou-pagamentos-de-ransomware/>. Acesso em: 16 ago. 2022.

FACELI, Katti et al. Inteligência artificial: uma abordagem de aprendizado de máquina. LTC, 2021.

FBI. **FBI Private Industry Notification**. [S. l.: s. n.], 2021. Disponível em: <https://www.cisa.gov/sites/default/files/publications/Egregor%5C%20PIN.pdf>. Acesso em: 5 abr. 2022.

FREEMAN, D.; CHIO, C. **Machine Learning and Security: Protecting Systems with Data and Algorithms**. [S. l.]: O'Reilly Media, 2018. ISBN 9781491979891. Disponível em: <https://books.google.com.br/books?id=rl2ftgEACAAJ>.

FRENCH, Robert M. The Turing Test: the first 50 years. **Trends in Cognitive Sciences**, v. 4, n. 3, p. 115–122, 2000. ISSN 1364-6613. DOI: https://doi.org/10.1016/S1364-6613(00)01453-4. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1364661300014534>.

GALOV, Dmitry; BEZVERSHENKO, Leonid; KWIATKOWSKI, Ivan. **Ransomware world in 2021: who, how and why**. en. [S. l.]: SecureList, 2021. Disponível em: <https://securelist.com/ransomware-world-in-2021/102169/>. Acesso em: 15 ago. 2022.

GANDHI, Krunal A et al. Survey on ransomware: a new era of cyber attack. **International Journal of Computer Applications**, Foundation of Computer Science, v. 168, n. 3, 2017.

GARCIA, David Escudero; DECASTRO-GARCIA, Noemi. Optimal feature configuration for dynamic malware detection. **Computers & Security**, Elsevier, v. 105, p. 102250, 2021.

GATELY, Edward. **Kaspersky: Ransomware Victims Rarely Regain All Data After Paying Ransom**. [S. l.: s. n.], 2021. Disponível em: <https://www.channelfutures.com/mssp-insider/kaspersky-ransomware-victims-rarely-regain-all-data-after-paying-ransom>. Acesso em: 2 ago. 2021.

GENÇ, Ziya Alper; LENZINI, Gabriele; RYAN, Peter YA. Next generation cryptographic ransomware. In: SPRINGER. NORDIC Conference on Secure IT Systems. [S. l.: s. n.], 2018. p. 385–401.

GLOVER, Claudia. **Panasonic confirms cyberattack after Conti leaks data**. en. [S. l.]: TechMonitor, 2022. Disponível em: <https://techmonitor.ai/technology/cybersecurity/conti-breached-panasonic>. Acesso em: 18 jul. 2022.

_____. **Virtualisation platforms becoming a top target for ransomware gangs**. en. [S. l.]: TechMonitor, 2022. Disponível em: <https://techmonitor.ai/technology/cybersecurity/virtualisation-ransomware-yanluowang-kaspersky-mandiant>. Acesso em: 18 jul. 2022.

GOSTEV, Alexander. **The Flame: Questions and Answers**. [S. l.]: Securelist By Kaspersky, 2012. Disponível em: <https://securelist.com/the-flame-questions-and-answers/34344/>. Acesso em: 3 set. 2021.

GOYAL, Parth S et al. Crypto-ransomware detection using behavioural analysis. In: RELIABILITY, Safety and Hazard Assessment for Risk-Based Technologies. [S. l.]: Springer, 2020. p. 239–251.

GROOT, Juliana de. **Take a look at the history of ransomware, the most damaging ransomware attacks, and the future for this threat.** [S. l.: s. n.], 2020. Disponível em: <https://digitalguardian.com/blog/history-ransomware-attacks-biggest-and-worst-ransomware-attacks-all-time>. Acesso em: 24 jul. 2021.

GRUSTNIY, Leonid. **A saga do ransomware.** [S. l.: s. n.], 2021. Disponível em: <https://www.kaspersky.com.br/blog/history-of-ransomware/17280/>. Acesso em: 24 jul. 2021.

GUARNIERI, Claudio et al. The cuckoo sandbox. **Accessed: Dec**, v. 16, p. 2018, 2012.

HAMMOND, John. **Persistence in Cybersecurity**. en. [S. l.]: Huntress, 2021. Disponível em: <https://www.huntress.com/defenders-handbook/persistence-in-cybersecurity#what-is-persistence-in-cybersecurity>. Acesso em: 15 ago. 2022.

HARAHSHEH, Heba; SHRAIDEH, Mohammad; SHARAEH, Saleh. Performance of Malware Detection Classifier Using Genetic Programming in Feature Selection. **Informatica**, v. 45, n. 4, 2021.

HEIMDALSECURITY. **Egregor Ransomware Analysis: Origins, M.O., Victims**. [S. l.: s. n.], 2022. Disponível em: <https://heimdalsecurity.com/blog/egregor-ransomware/>. Acesso em: 5 abr. 2022.

HOSSEINI, Ashkan. **Ten process injection techniques: A technical survey of common and trending process injection techniques**. en. [S. l.]: Elastic, 2017. Disponível em: <https://www.elastic.co/pt/blog/ten-process-injection-techniques-technical-survey-common-and-trending-process>. Acesso em: 10 ago. 2022.

HUNTING for Persistence: Registry Run Keys / Startup Folder. en. [S. l.]: Cyborg Security, 2021. Disponível em: <https://www.cyborgsecurity.com/cyborg-labs/hunting-for-persistence-registry-run-keys-startup-folder/>. Acesso em: 15 ago. 2022.

HWANG, Jinsoo et al. Two-stage ransomware detection using dynamic analysis and machine learning techniques. **Wireless Personal Communications**, Springer, v. 112, n. 4, p. 2597–2609, 2020.

IBMSECURITY. **X-Force Threat Intelligence Index 2022**. [S. l.: s. n.], 2022. Disponível em: <https://www.ibm.com/downloads/cas/ADLMYLAZ>. Acesso em: 8 jun. 2022.

IPDFORUM. **South Korean police help take down cyber threat in Ukraine**. [S. l.: s. n.], 2021. Disponível em: <https://ipdefenseforum.com/2021/08/south-korean-police-help-take-down-cyber-threat-in-ukraine/>. Acesso em: 4 mai. 2022.

JONES, Karen Sparck. A statistical interpretation of term specificity and its application in retrieval. **Journal of documentation**, MCB UP Ltd, 1972.

KASPERSKY. **Kaspersky Security Bulletin 2020. Statistics**. [S. l.]: Securelist, 2020. Disponível em: <https://securelist.com/kaspersky-security-bulletin-2020-statistics/99804/>. Acesso em: 3 set. 2021.

_____. **Ransomware Double Extortion and Beyond: REvil, Clop, and Conti**. [S. l.: s. n.], 2021. Disponível em: <https://www.kaspersky.com/resource-center/threats/ransomware-attacks-and-types>. Acesso em: 7 jun. 2022.

_____. **Trojan Droppers**. en. [S. l.]: Kaspersky, 2017. Disponível em: <https://encyclopedia.kaspersky.com/glossary/trojan-droppers/>. Acesso em: 18 jul. 2022.

KASPERSRKY. **Ataques de ransomware direcionados crescem 700%**. [S. l.: s. n.], 2021. Disponível em: <https://www.kaspersky.com.br/blog/ataques-ransomware-direcionados-crescem-700/17470/>. Acesso em: 7 jun. 2022.

_____. **Principais ataques de Ransomware**. [S. l.: s. n.], 2021. Disponível em: <https://www.kaspersky.com.br/resource-center/threats/top-ransomware-2020>. Acesso em: 7 jun. 2022.

KHAMMAS, Ban Mohammed. Ransomware detection using random forest technique. **ICT Express**, Elsevier, v. 6, n. 4, p. 325–331, 2020.

KLEYMENOV, Alexey; THABET, Amr. **Mastering Malware Analysis: The complete malware analyst's guide to combating malicious software, APT, cybercrime, and IoT attacks**. [S. l.]: Packt Publishing Ltd, 2019.

KOK, SH; ABDULLAH, Azween; JHANJHI, NZ. Early detection of crypto-ransomware using pre-encryption detection algorithm. **Journal of King Saud University-Computer and Information Sciences**, Elsevier, 2020.

KOK, SH; ABDULLAH, Azween; JHANJHI, NZ; SUPRAMANIAM, Mahadevan. Prevention of crypto-ransomware using a pre-encryption detection algorithm. **Computers**, MDPI, v. 8, n. 4, p. 79, 2019.

KOLODENKER, Eugene et al. Paybreak: Defense against cryptographic ransomware. In: PROCEEDINGS of the 2017 ACM on Asia Conference on Computer and Communications Security. [S. l.: s. n.], 2017. p. 599–611.

KÜHRER, Marc; ROSSOW, Christian; HOLZ, Thorsten. Paint it black: Evaluating the effectiveness of malware blacklists. In: SPRINGER. INTERNATIONAL Workshop on Recent Advances in Intrusion Detection. [S. l.: s. n.], 2014. p. 1–21.

LE GUERNIC, Colas; LEGAY, Axel. Ransomware and the legacy crypto API. In: SPRINGER. RISKS and Security of Internet and Systems: 11th International Conference, CRiSIS 2016, Roscoff, France, September 5-7, 2016, Revised Selected Papers. [S. l.: s. n.], 2017. v. 10158, p. 11.

LU, Yiwen. **Hackers claim they breached data on 1 billion Chinese citizens**. en. [S. l.]: Washington Post, 2022. Disponível em: <https://www.washingtonpost.com/business/2022/07/06/china-hack-police/>. Acesso em: 18 jul. 2022.

LUHN, Hans Peter. The automatic creation of literature abstracts. **IBM Journal of research and development**, Ibm, v. 2, n. 2, p. 159–165, 1958.

MAFFIA, Lorenzo et al. Longitudinal Study of the Prevalence of Malware Evasive Techniques. **arXiv preprint arXiv:2112.11289**, 2021.

MANGIALARDO, Reinaldo José; DUARTE, Julio Cesar. Construindo uma base para experimentação de malwares utilizando as análises estática e dinâmica.

MARINHO, Tarcísio. **Ransomware encryption techniques**. Edição: BleepingComputer. [S. l.: s. n.], 2018. Disponível em: <https://medium.com/@tarcisioma/ransomware-encryption-techniques-696531d07bb9>. Acesso em: 12 jul. 2022.

MCAFEE. **McAfee Labs Consolidated Threat Report:Duqu**. [S. l.]: McAfee, 2012. Disponível em: <http://download.nai.com/products/mcafee-avert/dil/Duqu_CTR_v2.2f.pdf>. Acesso em: 2 nov. 2021.

_____. **Ransomware Maze**. [S. l.: s. n.], 2020. Disponível em: <https://www.mcafee.com/blogs/other-blogs/mcafee-labs/ransomware-maze/#_ftn4>. Acesso em: 15 jul. 2022.

OR-MEIR, Ori et al. Dynamic Malware Analysis in the Modern Era—A State of the Art Survey. **ACM Comput. Surv.**, Association for Computing Machinery, New York, NY, USA, v. 52, n. 5, set. 2019. ISSN 0360-0300. DOI: 10.1145/3329786. Disponível em: <https://doi.org/10.1145/3329786>.

MERCALDO, Francesco. A framework for supporting ransomware detection and prevention based on hybrid analysis. **Journal of Computer Virology and Hacking Techniques**, Springer, v. 17, n. 3, p. 221–227, 2021.

MICRO, TREND. **Ransomware**. [S. l.: s. n.], 2021. Disponível em: <https://www.trendmicro.com/vinfo/tw/security/definition/ransomware>. Acesso em: 2 ago. 2021.

MICROSOFT. **Primitivos criptográficos**. [S. l.]: Microsoft, 2022. Disponível em: <https://docs.microsoft.com/pt-br/windows/win32/seccng/cryptographic-primitives>. Acesso em: 13 jul. 2022.

_____. **Win32/Mydoom**. [S. l.: s. n.], 2004. Disponível em: <https://www.microsoft.com/en-us/wdsi/threats/malware-encyclopedia-description?Name=Win32%5C%2FMydoom>. Acesso em: 14 jul. 2022.

MILLER, Cody et al. Insights gained from constructing a large scale dynamic analysis platform. **Digital Investigation**, Elsevier, v. 22, s48–s56, 2017.

MILLS, Alan; LEGG, Phil. Investigating anti-evasion malware triggers using automated sandbox reconfiguration techniques. **Journal of Cybersecurity and Privacy**, MDPI, v. 1, n. 1, p. 19–39, 2020.

MOHANTA, Abhijit; SALDANHA, Anoop. **Malware Analysis and Detection Engineering: A Comprehensive Approach to Detect and Analyze Modern Malware**. [S. l.]: Springer, 2020.

MONNAPPA, KA. **Learning Malware Analysis: Explore the concepts, tools, and techniques to analyze and investigate Windows malware**. [S. l.]: Packt Publishing Ltd, 2018.

NDIBANJE, Bruce et al. Cross-method-based analysis and classification of malicious behavior by api calls extraction. **Applied Sciences**, MDPI, v. 9, n. 2, p. 239, 2019.

OKTAVIANTO, Digit; MUHARDIANTO, Iqbal. **Cuckoo malware analysis**. [S. l.]: Packt Publishing Ltd, 2013.

ORMAN, Hilarie. Evil offspring-ransomware and crypto technology. **IEEE Internet Computing**, IEEE, v. 20, n. 5, p. 89–94, 2016.

ORTEGA, Alberto. **Pafish**. [S. l.: s. n.], 2021. Disponível em: <https://github.com/a0rtega/pafish>. Acesso em: 31 mai. 2022.

PAAR, Christof; PELZL, Jan. **Understanding cryptography: a textbook for students and practitioners**. [S. l.]: Springer Science & Business Media, 2009.

PARISOT, Augusto; BENTO, Lucila MS; MACHADO, Raphael CS. Testing and selecting lightweight pseudo-random number generators for IoT devices. In: IEEE. 2021 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4. 0&IoT). [S. l.: s. n.], 2021. p. 715–720.

PEREIRA, Mayana et al. Dictionary extraction and detection of algorithmically generated domain names in passive DNS traffic. In: SPRINGER. INTERNATIONAL Symposium on Research in Attacks, Intrusions, and Defenses. [S. l.: s. n.], 2018. p. 295–314.

PLOHMANN, Daniel et al. A comprehensive measurement study of domain generating malware. In: 25TH USENIX Security Symposium (USENIX Security 16). [S. l.: s. n.], 2016. p. 263–278.

PLOSZEK, Roderik; ŠVEC, Peter; DEBNÁR, Patrik. Analysis of encryption schemes in modern ransomware. **Rad Hrvatske akademije znanosti i umjetnosti: Matematičke znanosti**, Hrvatska akademija znanosti i umjetnosti, p. 1–13, 2021.

POPLI, Navneet Kaur; GIRDHAR, Anup. Behavioural analysis of recent ransomwares and prediction of future attacks by polymorphic and metamorphic ransomware. In: COMPUTATIONAL Intelligence: Theories, Applications and Future Directions-Volume II. [S. l.]: Springer, 2019. p. 65–80.

REGUERA, Ezra. **Tornado Cash community fund multisignature wallet disbands amid sanctions**. en. [S. l.]: CoinTelegraph, 2022. Disponível em: <https://cointelegraph.com/news/tornado-cash-community-fund-multi-signature-wallet-disbands-amid-sanctions>. Acesso em: 16 ago. 2022.

REUTERS. **Meatpacker JBS says it paid equivalent of $11 mln in ransomware attack**. [S. l.: s. n.], 2021. Disponível em: <https://www.reuters.com/technology/jbs-paid-11-mln-response-ransomware-attack-2021-06-09/>. Acesso em: 2 ago. 2021.

_____. **U.S. government working to aid top fuel pipeline operator after cyberattack**. [S. l.: s. n.], 2021. Disponível em: <https://www.reuters.com/business/energy/top-us-fuel-pipeline-operator-pushes-recover-cyberattack-2021-05-09/>. Acesso em: 2 ago. 2021.

AL-RIMY, Bander Ali Saleh; MAAROF, Mohd Aizaini; SHAID, Syed Zainudeen Mohd. Crypto-ransomware early detection model using novel incremental bagging with enhanced semi-random subspace selection. **Future Generation Computer Systems**, Elsevier, v. 101, p. 476–491, 2019.

ROCCIA, Tomas. **Malware Packers Use Tricks to Avoid Analysis, Detection**. en. [S. l.]: McAfee, 2017. Disponível em: <https://www.mcafee.com/blogs/enterprise/malware-packers-use-tricks-avoid-analysis-detection/>. Acesso em: 13 jul. 2022.

SATTER, Raphael. **Companies may be punished for paying ransoms to sanctioned hackers - U.S. Treasury**. en. [S. l.]: Reuters, 2020. Disponível em: <https://www.theregister.com/2022/08/08/treasury_sanctions_tornado_cash_korea/>. Acesso em: 16 ago. 2022.

SHARMA, Harshit; KANT, Shri. Early detection of ransomware by indicator analysis and WinAPI call sequence pattern. In: INFORMATION and Communication Technology for Intelligent Systems. [S. l.]: Springer, 2019. p. 201–211.

SHAUKAT, Saiyed Kashif; RIBEIRO, Vinay J. RansomWall: A layered defense system against cryptographic ransomware attacks using machine learning. In: IEEE. 2018 10th International Conference on Communication Systems & Networks (COMSNETS). [S. l.: s. n.], 2018. p. 356–363.

SIDI, Lior; NADLER, Asaf; SHABTAI, Asaf. **MaskDGA: A Black-box Evasion Technique Against DGA Classifiers and Adversarial Defenses**. [S. l.]: arXiv, 2019. DOI: 10.48550/ARXIV.1902.08909. Disponível em: <https://arxiv.org/abs/1902.08909>.

SIKORSKI, Michael; HONIG, Andrew. **Practical malware analysis: the hands-on guide to dissecting malicious software**. [S. l.]: no starch press, 2012.

SINGH, Jagsir; SINGH, Jaswinder. A survey on machine learning-based malware detection in executable files. **Journal of Systems Architecture**, Elsevier, v. 112, p. 101861, 2021.

SINITSYN, Fedor. **A new generation of ransomware**. en. [S. l.]: SecureList, 2014. Disponível em: <https://securelist.com/a-new-generation-of-ransomware/64608/>. Acesso em: 13 ago. 2022.

SINITSYN, Fedor; ZINCHENKO, Yanis. **Ransomware in the CIS**. en. [S. l.]: SecureList, 2021. Disponível em: <https://securelist.com/cis-ransomware/104452/>. Acesso em: 15 ago. 2022.

SOUPPAYA, Murugiah; SCARFONE, Karen. **Nist special publication 800-83 revision 1, guide to malware incident prevention and handling for desktops and laptops**. [S. l.], 2013.

STALLINGS, William; BRESSAN, Graça; BARBOSA, Akio. **Criptografia e segurança de redes.** 4. ed. [S. l.]: Pearson Educacion, 2008.

STATISTA. **Internet of Things (IoT) and non-IoT active device connections worldwide from 2010 to 2025**. [S. l.: s. n.], 2022. Disponível em: <https://www.statista.com/statistics/1101442/iot-number-of-connected-devices-worldwide/>. Acesso em: 16 jun. 2022.

STOECKLIN, Marc Ph.; JIYONG JANG, Dhilung Kirat. **DeepLocker: How AI Can Power a Stealthy New Breed of Malware**. [S. l.: s. n.], 2018. Disponível em: <https://securityintelligence.com/deeplocker-how-ai-can-power-a-stealthy-new-breed-of-malware/>. Acesso em: 18 nov. 2021.

SYMANTEC. **Can files locked by WannaCry be decrypted: A technical analysis**. [S. l.: s. n.], 2017. Disponível em: <https://medium.com/threat-intel/wannacry-ransomware-decryption-821c7e3f0a2b>. Acesso em: 7 jul. 2022.

TAKEUCHI, Yuki; SAKAI, Kazuya; FUKUMOTO, Satoshi. Detecting ransomware using support vector machines. In: PROCEEDINGS of the 47th International Conference on Parallel Processing Companion. [S. l.: s. n.], 2018. p. 1–6.

TANG, Mingdong; QIAN, Quan. Dynamic API call sequence visualisation for malware classification. **IET Information Security**, Wiley Online Library, v. 13, n. 4, p. 367–377, 2019.

TEAM, Kaspersky. **Ataques de ransomware direcionados crescem 700%**. [S. l.: s. n.], 2021. Disponível em: <https://www.kaspersky.com.br/blog/ataques-ransomware-direcionados-crescem-700/17470/>. Acesso em: 2 ago. 2021.

_____. **Ransomware 2.0: extorsão e chantagem em nova estratégia**. [S. l.: s. n.], 2021. Disponível em: <https://www.kaspersky.com.br/blog/ransomware-extorsao-chantagem/16786/>. Acesso em: 7 jul. 2022.

THREATPOST. **Mount Locker Ransomware Aggressively Changes Up Tactics**. [S. l.: s. n.], 2021. Disponível em: <https://threatpost.com/mount-locker-ransomware-changes-tactics/165559/>. Acesso em: 7 jun. 2022.

TRAFIMCHUK, Aliaksandr; BUKHTEYEV, Alexey; LADUTSKA, Raman. **CheckPoint Research Evasion Techniques**. [S. l.: s. n.], 2022. Disponível em: <https://evasions.checkpoint.com/>. Acesso em: 31 mai. 2022.

_____. **Report: Ransomware Attacks and the True Cost to Business**. [S. l.: s. n.], 2021. Disponível em: <https://www.cybereason.com/blog/research/report-ransomware-attacks-and-the-true-cost-to-business>. Acesso em: 31 mai. 2022.

TRENDMICRO. **DUQU Uses STUXNET-Like Techniques to Conduct Information Theft**. [S. l.: s. n.], 2011. Disponível em: <https://www.trendmicro.com/vinfo/br/threat-encyclopedia/web-attack/90/duqu-uses-stuxnetlike-techniques-to-conduct-information-theft>. Acesso em: 8 ago. 2022.

_____. **O que é o Ransomware RYUK?** [S. l.: s. n.], 2021. Disponível em: <https://www.trendmicro.com/pt_br/what-is/ransomware/ryuk-ransomware.html>. Acesso em: 7 jun. 2022.

TRENDMICRO. **Ransom.Win64.MOUNTLOCKER.E**. [S. l.: s. n.], 2021. Disponível em: <https://www.trendmicro.com/vinfo/us/threat-encyclopedia/malware/Ransom.Win64.MOUNTLOCKER.E/>. Acesso em: 7 jun. 2022.

_____. **Ransomware Spotlight: Clop**. [S. l.: s. n.], 2022. Disponível em: <https://www.trendmicro.com/vinfo/us/security/news/ransomware-spotlight/ransomware-spotlight-clop>. Acesso em: 4 mai. 2022.

_____. **Ransomware spotlight: Conti**. en. [S. l.]: TrendMicro, 2021. Disponível em: <https://www.trendmicro.com/vinfo/us/security/news/ransomware-spotlight/ransomware-spotlight-conti>. Acesso em: 18 jul. 2022.

_____. **Ransomware Spotlight: LockBit**. [S. l.: s. n.], 2022. Disponível em: <https://www.trendmicro.com/vinfo/us/security/news/ransomware-spotlight/ransomware-spotlight-lockbit>. Acesso em: 6 abr. 2022.

_____. **Ransomware spotlight: REvil**. en. [S. l.]: TrendMicro, 2021. Disponível em: <https://www.trendmicro.com/vinfo/us/security/news/ransomware-spotlight/ransomware-spotlight-revil>. Acesso em: 18 jul. 2022.

TULAS, Bill. **Ten notorious ransomware strains put to the encryption speed test**. [S. l.]: Bleeping Computer, 2022. Disponível em: <https://www.bleepingcomputer.com/news/security/ten-notorious-ransomware-strains-put-to-the-encryption-speed-test/>. Acesso em: 7 jun. 2022.

UNIVERSITY, Carnegie Mellon. **CA-2000-04: CERT® Advisory CA-2000-04 Love Letter Worm**. [S. l.: s. n.], 2000. Disponível em: <https://resources.sei.cmu.edu/asset_files/WhitePaper/2000_019_001_496188.pdf>. Acesso em: 14 jul. 2022.

UPGUARD. **What is Netwalker Ransomware? Attack Methods & Protection Tips**. [S. l.: s. n.], 2022. Disponível em: <https://www.upguard.com/blog/what-is-netwalker-ransomware>. Acesso em: 7 jun. 2022.

VAJJALA, S. et al. **Practical Natural Language Processing: A Comprehensive Guide to Building Real-world NLP Systems**. [S. l.]: O'Reilly Media, 2020. ISBN 9781492054054. Disponível em: <https://books.google.com.br/books?id=G40jywEACAAJ>.

VELUZ, Danielle. **Part 1: LockBit 2.0 ransomware bugs and database recovery attempts**. [S. l.]: Microsoft, 2022. Disponível em: <https://techcommunity.microsoft.com/t5/security-compliance-and-identity/part-1-lockbit-2-0-ransomware-bugs-and-database-recovery/ba-p/3254354>. Acesso em: 1 jun. 2022.

_____. **Part 2: LockBit 2.0 ransomware bugs and database recovery attempts**. [S. l.]: Microsoft, 2022. Disponível em: <https://techcommunity.microsoft.com/t5/security-compliance-and-identity/part-2-lockbit-2-0-ransomware-bugs-and-database-recovery/ba-p/3254421>. Acesso em: 1 jun. 2022.

WANG, Lele et al. Cuckoo-based malware dynamic analysis. **International Journal of Performability Engineering**, v. 15, n. 3, p. 772, 2019.

WUEEST, Candid. Threats to virtual environments. **Symantec Security Response. Version**, v. 1, 2014.

XIONG, Yong; RITCHIE, Hannah; GAN, Nectar. **Nearly one billion people in China had their personal data leaked, and it's been online for more than a year**. en. [S. l.]: CNN, 2022. Disponível em: <https://edition.cnn.com/2022/07/05/china/china-billion-people-data-leak-intl-hnk/index.html>. Acesso em: 18 jul. 2022.

YU, Bin et al. Character level based detection of DGA domain names. In: IEEE. 2018 international joint conference on neural networks (IJCNN). [S. l.: s. n.], 2018. p. 1–8.

ZHANG, Hanqi et al. Classification of ransomware families with machine learning based on N-gram of opcodes. **Future Generation Computer Systems**, Elsevier, v. 90, p. 211–221, 2019.

ZHAO, Hongwei et al. Evaluation of supervised machine learning techniques for dynamic malware detection. **International Journal of Computational Intelligence Systems**, Atlantis Press, v. 11, n. 1, p. 1153–1169, 2018.

ZHENG, Sarah. **Why China's Massive Data Leak Is So Chilling**. en. [S. l.]: Bloomberg, 2022. Disponível em: <https://www.bloomberg.com/news/newsletters/2022-07-11/why-china-s-massive-data-leak-is-so-chilling>. Acesso em: 18 jul. 2022.

ZUHAIR, Hiba; SELAMAT, Ali; KREJCAR, Ondrej. A multi-tier streaming analytics model of 0-day ransomware detection using machine learning. **Applied Sciences**, Multidisciplinary Digital Publishing Institute, v. 10, n. 9, p. 3210, 2020.