UNIVERSIDADE FEDERAL FLUMINENSE

ALLAN RODRIGO DE SOUZA BRAGA

Impactos da Pandemia da COVID-19 nas Conexões das Redes de Computadores e na Cibersegurança

NITERÓI

2023

UNIVERSIDADE FEDERAL FLUMINENSE

ALLAN RODRIGO DE SOUZA BRAGA

Impactos da Pandemia da COVID-19 nas Conexões das Redes de Computadores e na Cibersegurança

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 Computação. Área de concentração: Ciência da Computação

Orientador:
DIEGO PASSOS

Co-orientador:
ANTONIO AUGUSTO DE ARAGÃO ROCHA

NITERÓI

2023

Ficha catalográfica automática - SDC/BEE
Gerada com informações fornecidas pelo autor

S719i Souza Braga, Allan Rodrigo de
Impactos da Pandemia da COVID-19 nas Conexões das Redes de Computadores e na Cibersegurança / Allan Rodrigo de Souza Braga. - 2023.
189 f.: il.

Orientador: Diego Gimenez Passos.
Coorientador: Antonio Augusto de Aragão Rocha.
Dissertação (mestrado)-Universidade Federal Fluminense, Instituto de Computação, Niterói, 2023.

1. Segurança. 2. Redes de computadores. 3. Aprendizado de máquina. 4. Produção intelectual. I. Gimenez Passos, Diego, orientador. II. Aragão Rocha, Antonio Augusto de, coorientador. III. Universidade Federal Fluminense. Instituto de Computação.IV. Título.

CDD - XXX

Bibliotecário responsável: Debora do Nascimento - CRB7/6368

ALLAN RODRIGO DE SOUZA BRAGA

Impactos da Pandemia da COVID-19 nas Conexões das Redes de Computadores e na Cibersegurança

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 Computação. Área de concentração: Ciência da Computação

Aprovada em JUNHO de 2023.

BANCA EXAMINADORA

Prof. Diego Gimenez Passos - Orientador, UFF

Prof. Antonio Augusto de Aragão Rocha - Co-orientador, UFF

Prof. Célio Vinicius Neves de Albuquerque, UFF

Prof. Pedro Braconnot Velloso, LIP6

Niterói

2023

À minha família, de onde vem minha força.

Agradecimentos

Primeiramente, agradeço a Deus por ter me sustentado nos momentos mais difíceis dessa jornada, me dando força para continuar quando parecia impossível.

Aos Professores Diego e Guto, pela orientação sempre precisa e competente, pelo tempo dedicado e pelo encorajamento. Suas experiências e habilidade foram fundamentais para a realização deste trabalho. Estou muito grato por tê-los como meus orientadores.

Gostaria de expressar, também, minha gratidão a todas as pessoas que me ajudaram em minha jornada acadêmica, em especial aos professores do Instituto de Computação, que transmitiram seus conhecimentos, os quais contribuíram sobremaneira para realizar esta pesquisa. Seus ensinamentos foram valiosos e me ajudaram a desenvolver minhas habilidades como pesquisador.

Gostaria de expressar meu profundo agradecimento aos meus pais, Márcia e Nilson, que sempre estiveram presentes em minha vida, me guiando com amor e paciência. Sem o seu apoio incondicional, eu não teria chegado tão longe. Eles são minha fonte de inspiração e exemplo de perseverança e dedicação. Mãe, em que pese o que aconteceu com a Sra., estarei sempre ao seu lado, hoje e sempre.

Aos meus maiores tesouros, Heitor e Alícia. Cada sorriso, abraço e momento que compartilhamos é precioso para mim. Agradeço a eles por me ensinarem a ver o mundo de uma forma diferente, com mais esperança e alegria. Ser pai é a melhor experiência que já tive, e sou grato por tê-los em minha vida.

Finalmente, à minha amada esposa, minha companheira de vida, minha parceira em todas as aventuras e minha melhor amiga. Ela é minha força e meu apoio, me encorajando a seguir meus sonhos e me ajudando a superar todos os desafios. Agradeço a ela por ser minha rocha e por me amar incondicionalmente. Sem ela, minha vida não estaria completa.

A todos vocês, meu sincero e profundo agradecimento.

Resumo

Considerando as restrições impostas em virtude da pandemia da COVID-19, em especial as relacionadas à mobilidade das pessoas, este trabalho tem como um de seus propósitos investigar os impactos dessas restrições sobre as conexões das redes de computadores e sobre o volume de tráfego. Para explorar essa temática, foram realizadas duas pesquisas de opinião, cujo objetivo principal foi avaliar a percepção do usuário sobre a estabilidade da sua conexão com a Internet. Porém, como a percepção do usuário pode não refletir, necessariamente, a realidade, realizou-se uma análise da mudança nos padrões de tráfego dos Pontos de Presença (PoPs) da RNP (Rede Nacional de Ensino e Pesquisa) e um acompanhamento pelo período de dezoito meses no tráfego de redes sem fio e conexões de dispositivos móveis em Estações Rádio Base (ERB), confrontando-o com informações sobre a mobilidade da comunidade. Observou-se que o tráfego da Internet cresceu. Já em outras redes, como aquelas das instituições de pesquisa, o volume de tráfego diminuiu. Somente no Brasil, o IX.br (Brasil Internet Exchange), infraestrutura para a interconexão entre os *Autonomous Systems* (ASs), registrou um aumento no tráfego na ordem de 60% entre dezembro de 2019 e dezembro de 2021. Além do consumo de banda, também houve uma mudança no tipo de tráfego. Nos PoPs da RNP, por exemplo, houve um aumento de até cinco vezes no consumo de banda para aplicações de videoconferência, embora o consumo total de banda tenha diminuído.

Ao verificar variações significativas tanto no tipo quanto no volume de tráfego, suscitou-se a possibilidade de impactos na cibersegurança, que é uma fonte de preocupação para a maioria das organizações devido ao crescimento das ameaças. Com o objetivo de subsidiar as equipes de mitigação de riscos com o máximo de informações da maneira mais precoce possível, este trabalho explora várias metodologias baseadas em aprendizado de máquina para classificar a severidade das vulnerabilidades e o tempo até que sejam corrigidas. A fim de explorar tais metodologias, utilizaram-se as descrições textuais e/ou métricas do vetor *Common Vulnerability Scoring System* (CVSS). Também foi analisado como o desequilíbrio de classes e o tipo de *software* influenciam na classificação de vulnerabilidades. Para isso, este trabalho avalia a eficiência dessas metodologias usando um conjunto de dados que compreende vulnerabilidades de vários aplicativos diferentes. No entanto, devido à particular relevância das aplicações de comunicação durante a pandemia da COVID-19, este trabalho avalia, também, a dificuldade de realizar as predições citadas para este grupo específico de aplicações.

Palavras-chave: Pandemia, COVID-19, Redes, Tráfego, Banda, Aplicações de Comunicação, Videoconferência, RNP, PoP, Mobilidade, Segurança, Cibersegurança, Vulnerabilidade, CVE, Aprendizado de Máquina, Severidade, Tempo de Correção.

Abstract

Considering the restrictions imposed due to the COVID-19 pandemic, in particular those related to people's mobility, this work has as one of its purposes to investigate the impacts of these restrictions on computer network connections and on the volume of traffic. In order to explore this topic, two opinion polls were conducted, whose main objective was to evaluate the user's perception of the stability of their Internet connection. However, as the user's perception may not necessarily reflect reality, an analysis of the change in traffic patterns of the Points of Presence (PoPs) of the *Rede Nacional de Ensino e Pesquisa* (RNP) and a monitoring of the traffic of wireless networks and mobile device connections in Radio Base Stations (RBS) for an 18-month period were performed, comparing them with information about community mobility. It was observed that Internet traffic grew. On other networks, such as those of research institutions, the volume of traffic has decreased. Only in Brazil, the IX.br (Brazil Internet Exchange), the infrastructure for interconnection between Autonomous Systems (ASs), registered an increase in traffic of around 60% between December 2019 and December 2021. In addition to bandwidth consumption, there was also a change in traffic type. For example, in the RNP's PoPs, there was an increase of up to five times in bandwidth consumption for videoconferencing applications, although the total bandwidth consumption decreased.

Upon observing significant variations in both type and volume of traffic, the possibility of impacts on cybersecurity, which is a source of concern for most organizations due to the growth of threats, was raised. In order to support risk mitigation teams with the maximum amount of information as early as possible, this work explores several methodologies based on machine learning to classify the severity of vulnerabilities and the time until they are corrected. To explore such methodologies, the textual descriptions and/or metrics of the Common Vulnerability Scoring System (CVSS) vector were used. The influence of class imbalance and software type on vulnerability classification was also analyzed. To that end, this work evaluates the efficiency of these methodologies using a dataset that comprises vulnerabilities of several different applications. However, due to the particular relevance of communication applications during the COVID-19 pandemic, this work also evaluates the difficulty of performing the predictions mentioned for this specific group of applications.

Keywords: Pandemic, COVID-19, Networks, Traffic, Bandwidth, Communication Applications, Videoconferencing, RNP, PoP, Mobility, Security, Cybersecurity, Vulnerability, CVE, Machine Learning, Severity, Patching Time.

Lista de Figuras

Lista de Tabelas

Lista de Abreviaturas e Siglas

AP	Access Point
ASs	Autonomous Systems
BI	Business Intelligence
BSR	Bad Session Rate
CNN	Convolutional Neural Networks
CSV	Comma-Separated Values
CVE	Common Vulnerabilities and Exposures
CVSS	Common Vulnerability Scoring System
D2V	Doc2Vec
DHCP	Dynamic Host Configuration Protocol
DNS	Domain Name System
ERB	Estação Rádio Base
ICMP	Internet Control Message Protocol
IPSec	IP Security Protocol
IX.br	Brasil Internet Exchange
L2TP	Layer 2 Tunnelling Protocol
LDA	Linear Discriminant Analysis
Mbps	Megabit por Segundo
ML	Machine Learning
MSE	Mean Squared Error

NIST	National Institute of Standards and Technology
NLP	Natural Language Processing
NVD	National Vulnerability Database
OMS	Organização Mundial da Saúde
PCA	Principal Component Analysis
PoPs	Pontos de Presença
QoE	Quality of Experience
RAM	Random Access Memory
RBS	Radio Base Stations
RIP	Routing Information Protocol
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Networks
RNP	Rede Nacional de Ensino e Pesquisa
RTT	Round Trip Time
SBRC	Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos
SNMP	Simple Network Management Protocol
TCP	Transmission Control Protocol
TF	Term Frequency
TF-IDF	Term Frequency-Inverse Document Frequency
UDP	User Datagram Protocol
VoLTE	Voice over Long Term Evolution
VPN	Virtual Private Network

Sumário

Capítulo 1

Introdução

1.1 Contextualização

O novo Coronavírus foi isolado, por autoridades Chinesas, em 7 de janeiro de 2020, de acordo com o *Situation Report* 1[1]. A COVID-19, doença causada pelo novo Coronavírus, foi caracterizada como pandêmica, pela Organização Mundial da Saúde (OMS), em 12 de março de 2020, em consonância com o *Situation Report* 52[1]. Desde então, considerando dados acumulados até janeiro de 2023, as infecções por esse vírus, em todo o mundo, ultrapassaram a marca dos 650 milhões de casos, enquanto mais de 6,5 milhões de mortes foram notificadas, segundo o consignado no *Weekly epidemiological update - 4 January 2023*[1].

Os alarmantes números da pandemia exigiram que os governos adotassem medidas emergenciais para tentar controlar a proliferação do vírus, reduzindo a contaminação. Não havendo tratamento ou vacina disponíveis, as medidas recomendadas, até então, eram através de intervenções não farmacológicas [29]. Uma dessas medidas foi o isolamento da população [4]. Cada país lidou de forma diferente com o isolamento. No Brasil, não houve diretrizes uniformes e cada estado / município estabeleceu seus indicadores e ações decorrentes, como o funcionamento de apenas serviços essenciais, estabelecido no Rio de Janeiro e São Paulo. China, Irã, Coréia do Sul, Estados Unidos da América e países europeus, como Itália, Espanha, França e Reino Unido, adotaram medidas de isolamento social mais rígidas.

As medidas de isolamento reduziram a circulação de pessoas e, por conseguinte, a transmissão do vírus. No entanto, efeitos colaterais puderam ser observados, em especial

[1]https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

nas características de tráfego das redes, como a Internet. Como exemplo, as chamadas de voz e vídeo no *WhatsApp* e no *Facebook Messenger* tiveram um acréscimo de mais de 100% [31]; Segundo o vice-presidente corporativo do *Microsoft* 365, a empresa registrou, no aplicativo *Microsoft Teams*, um aumento de 500% nas reuniões, chamadas e conferências, e um aumento de 200% no uso em dispositivos móveis [73]. Já a *Netflix* e o *YouTube* decidiram reduzir a qualidade do *streaming* na Europa para evitar o colapso das redes, devido à demanda sem precedentes [22].

No Brasil, o IX.br registrou em julho/2019 7 Tb/s de tráfego [50] e 8 Tb/s em dezembro/2019 [51]. Após a decretação da pandemia e a efetiva implementação de medidas restritivas, o crescimento do tráfego foi mais elevado, chegando a 10 Tb/s em março/2020 [47], 16 Tb/s em março/2021 [48] e 20 Tb/s em dezembro de 2021 [49]. Esse aumento de tráfego refletiu a mudança na rotina das pessoas, como a redução da mobilidade, a maior concentração das pessoas em casa e a maneira como elas se comunicam, trabalham, estudam, consomem serviços e realizam compras. Esta mudança pode ser vista na Figura 1.1, na qual o *Google* utiliza as seguintes definições de locais de concentração[2]:

- **Mercados e farmácias:** Tendências de mobilidade de lugares como mercados, armazéns de alimentos, feiras, lojas especializadas em alimentos, drogarias e farmácias;
- **Parques:** Tendências de mobilidade de lugares como parques locais e nacionais, praias públicas, marinas, parques para cães, praças e jardins públicos;
- **Estações de transp. público:** Tendências de mobilidade de lugares como terminais de transporte público, como estações de metrô, ônibus e trem;
- **Varejo e lazer:** Tendências de mobilidade de lugares como restaurantes, cafés, shopping centers, parques temáticos, museus, bibliotecas e cinemas;
- **Residencial:** Tendências de mobilidade de áreas residenciais; e
- **Locais de trabalho:** Tendências de mobilidade de locais de trabalho.

Considerando o expressivo aumento no tráfego e na utilização de aplicações de comunicação, uma questão fundamental que deve ser avaliada é a segurança. Isso se deve ao fato de que, repentinamente, essas aplicações ganharam um grande protagonismo e eventuais

[2] `https://www.google.com/covid19/mobility/data_documentation.html?hl=pt-BR`
[3] `https://www.google.com/covid19/mobility`

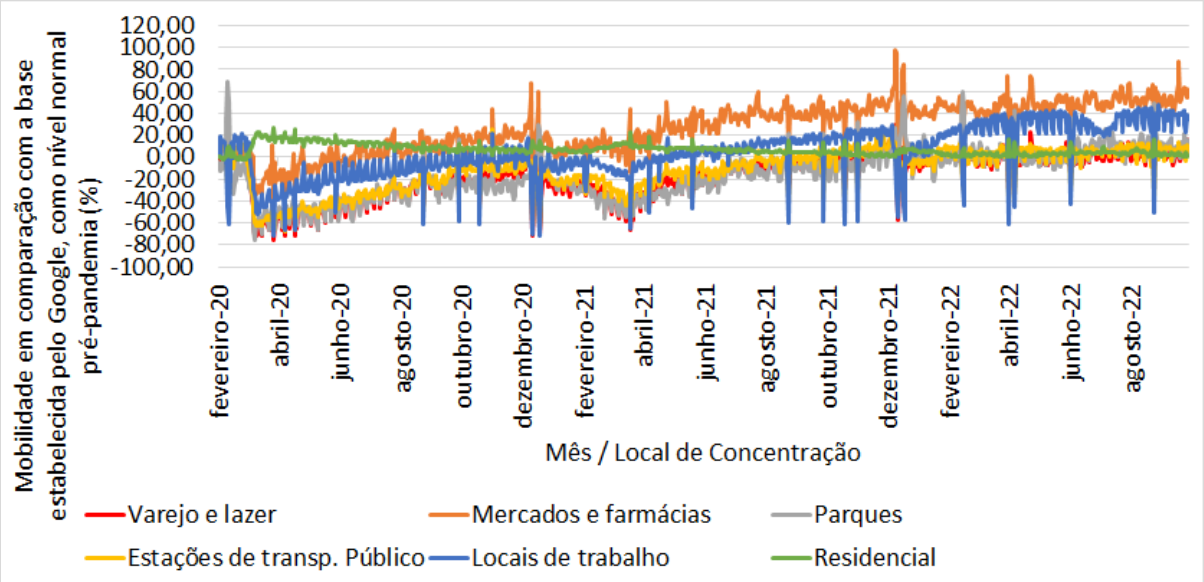


Figura 1.1: Variação da mobilidade no Brasil[3]

vulnerabilidades poderiam levar a exposição de muitos usuários. Uma vulnerabilidade de *software* é uma brecha em seu código que pode ser explorada para contornar mecanismos de segurança [68] visando causar danos ou obter acesso a informações para as quais não é permitido, entre outros objetivos. Para proteger seus sistemas, as organizações podem usar vulnerabilidades identificadas e catalogadas pelo *Common Vulnerabilities and Exposures (CVE)*. Essas vulnerabilidades são descobertas e publicadas por organizações em todo o mundo que têm parceria com o Programa CVE[4]. Depois de publicadas pela CVE, as vulnerabilidades são analisadas pelo *National Institute of Standards and Technology (NIST)* e publicadas no repositório de dados de gerenciamento de vulnerabilidades do governo dos EUA chamado *National Vulnerability Database (NVD)*[5]. Com base nessas informações, as equipes de segurança podem analisar e priorizar suas ações defensivas.

Entre outras informações, a análise do NIST resulta em métricas de impacto de associação conhecidas como *Common Vulnerability Scoring System (CVSS)* [58]. O CVSS resume informações como a facilidade de explorar a vulnerabilidade e a extensão dos danos que podem ser causados — o que, por sua vez, pode ser usado para avaliar a severidade da vulnerabilidade. Assim, quanto mais cedo o CVSS estiver disponível para uma determinada vulnerabilidade, melhores serão as chances de minimizar seus impactos (*e.g.*, conscientizando seus sobre seus riscos mais cedo e acionando respostas apropriadas). No entanto, o CVSS não é disponibilizado quando as vulnerabilidades são inicialmente divulgadas. Em vez disso, eles são definidos após o estudo da vulnerabilidade, o que leva

[4]https://www.cve.org/About/Overview
[5]https://nvd.nist.gov/general

tempo.

Para resolver esse problema, a predição automática da severidade com base em dados prontamente disponíveis quando a vulnerabilidade é divulgada pela primeira vez pode ser fundamental para priorizar o tratamento dos riscos cibernéticos. O *Machine Learning (ML)* e o *Natural Language Processing (NLP)* são capazes de auxiliar nesse processo. Embora a severidade seja o objetivo final, outras informações também podem ser preditas, de forma a detalhar ainda mais os riscos inerentes a uma determinada vulnerabilidade, quais sejam, o *base score* e as métricas do vetor CVSS. As predições podem ter como base a descrição textual da vulnerabilidade, disponível no banco de dados CVE. Essas descrições são fornecidas pelos parceiros CVE [6] e são escritas pelos desenvolvedores de software ou especialistas em segurança, resultando em uma estrutura um tanto consistente.

1.2 Objetivos

Em circunstâncias como a da pandemia do novo Coronavírus, em que há um elevado e repentino crescimento da demanda sobre as redes de dados, é possível que haja implicações na percepção da qualidade de serviço, além de eventuais indisponibilidades de redes e sistemas. Em virtude disso, esta dissertação tem o propósito de estudar os impactos que o cenário descrito é capaz de produzir. Dentre eles, elencam-se os impactos na percepção do desempenho da rede pelo usuário, os impactos nas redes de acesso, de *backbone*, sem fio e de telefonia celular e os impactos na segurança cibernética.

1.3 Questões de Pesquisa

De forma a cumprir os objetivos (Seção 1.2), e tomando como base o o período da pandemia do novo Coronavírus, é necessário responder a algumas questões, tais como:

1. Qual é a percepção do usuário sobre a estabilidade de sua conexão com a Internet durante o período da pandemia?

2. Qual foi o fator que mais influenciou a percepção do usuário sobre a estabilidade de sua conexão com a Internet durante o período da pandemia?

3. Qual foi o impacto sobre as aplicações de comunicação, como *Google Meet* e *WhatsApp*, e o comportamento de aplicações/protocolos de teletrabalho, como VPN e

[6] https://www.cve.org/PartnerInformation/ListofPartners

área de trabalho remota?

4. Houve, de fato, mudança no volume de tráfego nas redes em decorrência da pandemia?

5. As mudanças no volume de tráfego diferem dependendo da rede analisada, *e.g.*, Internet e redes corporativas ou de pesquisa?

6. É possível correlacionar as mudanças no volume de tráfego e nas conexões entre dispositivos móveis celulares e as ERB com a mobilidade da comunidade?

7. A fim de minimizar impactos na segurança cibernética, quais são, como implementar e avaliar diferentes metodologias, baseadas em inteligência artificial, para predizer a severidade de uma nova vulnerabilidade?

8. Quais são as vantagens e desvantagens de cada uma das metodologias para predizer a severidade de uma nova vulnerabilidade?

9. Qual o desempenho obtido pelas metodologias avaliadas para predizer a severidade de uma nova vulnerabilidade?

10. Qual o desempenho de um classificador treinado a partir de *datasets* com dados de aplicações genéricas ao realizar predições acerca de aplicações específicas, particularmente, no âmbito deste trabalho, sobre aplicações de comunicação?

11. Há viabilidade na implementação de algoritmos para predição do tempo de correção de vulnerabilidades?

1.4 Metodologia

Para responder às questões de pesquisa (Seção 1.3) e iniciar a avaliação sobre os impactos causados pela pandemia do novo Coronavírus, julgou-se importante ter uma visão sobre a percepção do usuário sobre a estabilidade de sua conexão com a Internet, correlacionando-a com informações sobre quais foram as aplicações mais utilizadas, situação laboral e plano de Internet contratada. Com esse fim, foi elaborada uma pesquisa de opinião.

A realização em larga escala do teletrabalho, decorrente da pandemia, culminou em impactos no volume de tráfego. Esses impactos originaram-se, mormente, pela utilização de aplicações de comunicação e ferramentas de acesso a ambientes computacionais remotos. Destarte, no que tange às aplicações de comunicação, especialmente aquelas

apontadas como mais utilizadas na pesquisa de opinião, é imprescindível conhecer seus comportamentos e o consumo de banda de cada uma, a fim de que as corporações tenham subsídios técnicos para selecionar a ferramenta que será adotada. Outrossim, no que diz respeito às ferramentas de acesso a ambientes computacionais remotos, faz-se mister analisar o desempenho e aspectos de segurança das opções disponíveis. No entanto, nesse caso, há que se salientar que o desempenho deve ser medido de acordo com a aplicação utilizada (*e.g.*, sistemas *desktop* ou sistemas *web*).

Ainda sobre o resultado da pesquisa de opinião, e considerando que a percepção dos respondentes era de uma piora na conexão com a Internet, surgiu outra questão a ser estudada, que diz respeito à origem desse problema, que poderia estar na infraestrutura pública ou na forma como os dispositivos se conectam à rede nas residências. Uma das avaliações necessárias para responder à tal questão é a possibilidade que a escolha na utilização de redes sem fio entre as faixas de frequência de 2,4 GHz e 5 GHz possa exercer influência sobre a percepção da estabilidade da conexão de Internet para o usuário. Houve, então, a demanda por analisar o tráfego de redes sem fio da faixa de frequência de 2,4 GHz, a fim de acompanhar sua evolução e correlação com os dados de mobilidade. Os dados foram coletados durante dezoito meses. Para complementar a análise em tela, também foram introduzidos os dados de telefonia móvel celular. Contudo, considerando que a percepção das pessoas não reflete, necessariamente, com precisão, o comportamento das redes, complementou-se esse estudo com a análise do tráfego de aplicações de comunicação e dos Pontos de Presença (PoPs) de uma rede de grande escala, a Rede Nacional de Ensino e Pesquisa (RNP). O objetivo foi entender como essas mudanças ocorreram e o impacto que elas causaram.

Na temática da segurança, o presente estudo se propõe a avaliar diversas metodologias para a predição da severidade de vulnerabilidades. Cabe ressaltar a proposição de um novo algoritmo de IA não é o objetivo deste trabalho, mas avaliar um conjunto de algoritmos já existentes. Também é examinada a efetividade da aplicação de algoritmo de redução de dimensionalidade, bem como da utilização da técnica de *voting*. Todas essas combinações são adotadas a fim de indicar a melhor forma de obter a severidade da vulnerabilidade, quais sejam, (i) severidade como saída do modelo de predição; (ii) severidade em função do *base score* predito; e (iii) severidade em função do *base score* calculado a partir das métricas do vetor CVSS. Para chegar a essa conclusão, é indispensável ponderar o consumo de recursos computacionais em cada situação. Até esse ponto, todos os modelos se baseiam em *datasets* com dados de aplicações genéricas. No entanto, considerando o contexto imposto pela pandemia, avaliou-se a capacidade de modelos treinados

com dados de aplicações genéricas fazerem predições acerca de aplicações específicas, particularmente, no âmbito deste trabalho, sobre aplicações de comunicação. Finalmente, traz-se à baila a viabilidade da predição do tempo de correção de vulnerabilidades.

1.5 Contribuições

Dentre as contribuições desta dissertação, encontra-se a demonstração da influência de aplicações de comunicação e ferramentas de acesso a ambientes computacionais remotos no impacto no volume de tráfego. Assim, em situações de excepcionalidade, é possível dimensionar os recursos computacionais necessários para a implementação de soluções corporativas de teletrabalho. Com o mesmo objetivo, apresenta-se a análise do funcionamento e do consumo de banda das aplicações de comunicação e das ferramentas de acesso a ambientes computacionais remotos. Para a última, levando em consideração o tipo de aplicação utilizada (*e.g.*, sistemas *desktop* ou sistemas *web*), e para ambas ponderando aspectos de segurança. Para as aplicações de comunicação, além de testes de desempenho locais, foi avaliada sua utilização em redes de grande porte, notadamente, a RNP.

Também figura entre as contribuições a demonstração da influência sobre a percepção do usuário acerca da estabilidade da conexão de Internet de acordo com a faixa de frequência utilizada na rede sem fio doméstica (*e.g.*, 2,4 GHz ou 5 GHz). Ainda na abordagem de redes, aduz a correlação entre o volume de tráfego em redes sem fio na faixa de frequência 2,4 GHz, dispositivos móveis conectados às ERB e a mobilidade da comunidade, publicada pelo *Google*. Assim, reconhecendo o forte impacto que a pandemia trouxe sobre as redes.

Concernente à interseção entre redes, *machine learning* e segurança, foram avaliados algoritmos, *datasets* e técnicas de ML e NLP para predição da severidade, *base score* e métricas do vetor CVSS de uma vulnerabilidade, a partir de sua descrição textual. O objetivo deste tópico é o apoio aos times de tratamento de riscos cibernéticos no direcionamento e na priorização das medidas protetivas contra eventuais incidentes advindos de vulnerabilidades ainda não analisadas. Para isso, são propostas três formas de obtenção da severidade de uma vulnerabilidade: (i) severidade como saída do modelo de predição; (ii) severidade em função do *base score* predito; e (iii) severidade em função do *base score* calculado a partir das métricas do vetor CVSS. Em todas propostas, são avaliados o impacto do desbalanceamento de classes no desempenho dos modelos e o consumo de recursos computacionais dos modelos de ML e NLP usados neste estudo. Também

confirmou-se a efetividade das predições acerca de aplicações específicas a partir de modelos treinados com dados de aplicações genéricas. Por fim, estudou-se a viabilidade da predição do tempo de correção de vulnerabilidades, para a qual não foram encontrados na literatura outros trabalhos com propositura similar.

Embora a ideia de usar ML e NLP para prever a severidade das vulnerabilidades com base em suas descrições textuais tenha sido explorada na literatura anteriormente, este trabalho tem diferenciais. Em primeiro lugar, considerou-se a predição do tempo necessário para corrigir as vulnerabilidades — uma informação que pode ser inestimável para os gerentes de rede quando eles decidem como lidar com uma vulnerabilidade. Também foi analisado como o tipo de aplicação influencia a capacidade de predizer informações sobre vulnerabilidades — mais especificamente, o quão difícil é predizer a severidade, o *base score* e as métricas do vetor CVSS para essas aplicações, em comparação com a população geral de vulnerabilidades disponíveis no NVD. Notadamente, avaliou-se a aplicabilidade de um modelo treinado para a população em geral às vulnerabilidades de aplicações de comunicação. Para isso, considerou-se o subconjunto de aplicações de comunicação, em virtude de terem se tornado potenciais alvos específicos para ataques durante a pandemia da COVID-19. Além disso, a influência da metodologia de extração de dados e o problema de desequilíbrio de classes no resultado final também são avaliados. Finalmente, enquanto a maioria dos trabalhos visa estudar um algoritmo específico para realizar predições de severidade de vulnerabilidades, adotou-se uma abordagem mais ampla, avaliando diversas metodologias, incluindo o uso de múltiplos sistemas classificadores.

Os resultados dos experimentos são encorajadores, pois geralmente mostram boa precisão. Os resultados também sugerem que classificadores treinados com vulnerabilidades genéricas podem fazer predições eficientes sobre vulnerabilidades em tipos específicos de aplicações.

1.6 Organização

Esta dissertação encontra-se organizada da seguinte forma. No Capítulo 2 são introduzidos alguns conceitos basilares para o desenvolvimento deste trabalho. O Capítulo 3 apresenta as duas pesquisas de opinião, discute os impactos da pandemia no volume de tráfego das aplicações de Internet, através da análise e tráfego e comparação entre as principais aplicações de comunicação e protocolos para teletrabalho, e aborda os impactos nas conexões das redes de acesso, por meio da análise de dados do *backbone* da RNP, da

análise de tráfego de redes sem fio e da análise de dados de telefonia celular. O Capítulo 4 apresenta possíveis abordagens para aplicação de *Machine Learning* na predição da severidade e do tempo de correção de vulnerabilidades, discutindo e avaliando os resultados de múltiplos modelos. O Capítulo 5 apresenta as conclusões da dissertação e a publicação derivada.

Capítulo 2

Referencial Teórico

Um dos maiores impactos do isolamento social é o relacionamento interpessoal. Até então, esse relacionamento era estimulado a ser presencial. Isso porque há estudos que revelam que, no relacionamento interpessoal virtual, um indivíduo adquire bons e maus hábitos, podendo culminar na dependência de ambientes virtuais ou, até mesmo em uma fobia social [35].

Após a chegada da pandemia e necessidade de distanciamento social, a comunicação passou a ser prementemente virtual. Com isso, diversos aplicativos, dentre eles os de comunicação, tiveram um expressivo incremento em sua utilização, sendo abordados, neste estudo, os de mensagens instantâneas, de chamadas de voz, de conferências, de chamadas de vídeo e de videoconferências.

Além da comunicação, em muitos casos, o teletrabalho exige que o usuário tenha acesso a recursos que só estão disponíveis no ambiente computacional corporativo, como: compartilhamento de arquivos e *softwares*. Há soluções disponíveis e largamente utilizadas, das quais descrevemos neste capítulo a área de trabalho remota e a VPN (*Virtual Private Network*).

2.1 Aplicativos de Mensagens Instantâneas

As mensagens de texto têm sido usadas com sucesso na área de saúde. Em 2009 as mensagens de texto foram usadas para promover a aplicação da segunda e terceira dose da vacina do Papilomavírus Humano (HPV) [34] e na prevenção contra o vírus Influenza [75]. Com o surgimento da COVID-19, as mensagens de texto [41] também foram utilizadas para alertar a população sobre os perigos da doença e formas de prevenção [18].

Nesse diapasão, a constante evolução das tecnologias permitiu uma comunicação muito mais rápida e sem custo adicional. As mensagens de texto tradicionais, como o *Short Message Service* (SMS), estão perdendo espaço, em especial entre os usuários de dispositivos móveis inteligentes, para os aplicativos de mensagens instantâneas, como o *WhatsApp* [12]. Alguns aplicativos de mensagens instantâneas permitem a criação de *chatbots* [27], que podem representar uma importante ferramenta em alguns tipos de serviço. Nesse contexto, em abril de 2020, a OMS criou um serviço interativo, utilizando um aplicativo de mensagens instantâneas, para que a população possa tirar dúvidas sobre o novo coronavírus [62]. Embora os aplicativos de mensagens instantâneas sejam utilizados na comunicação informal, essa categoria de aplicativos pode intermediar comunicações complexas de trabalho, como agendamentos e reuniões [30].

Um dos aplicativos de mensagens instantâneas mais usados é o *WhatsApp* [54]. É comum, atualmente, que esse tipo de aplicativo ofereça recursos para o compartilhamento de mídia, como imagens, vídeos, áudios e até a localização atual do dispositivo [54]. A eficiência desses aplicativos, no que se refere a utilização de banda, foi analisada [81] tomando-se como base o tráfego gerado por caractere enviado / recebido. Dentre os aplicativos comparados (*WeChat*, *WhatsApp*, *Facebook Messenger*, *Line* e *Viber*), o *Facebook Messenger* obteve o pior desempenho, enquanto os melhores foram o *WhatsApp* e o *Line*. Esse mesmo estudo demonstrou que o desempenho dos aplicativos de mensagens instantâneas é inferior à de outros tipos de aplicações, como e-mail e navegação web, se atendo às mensagens de texto e não considerando os recursos que utilizam maior largura de banda, como imagens, vídeos e áudios.

2.2 Chamadas de Voz/Vídeo e Videoconferências

Com o passar do tempo, as tecnologias que possibilitam o teletrabalho têm evoluído significativamente. Dentre essas tecnologias, as chamadas de voz e vídeo e as videoconferências tornam o trabalho remoto viável [32]. Em abril de 2020, o Governo Brasileiro publicou um documento para orientar trabalhadores e empregadores quanto aos cuidados que deveriam ser adotados no período da pandemia [62]. Dentre esses cuidados, cita-se o de evitar reuniões presenciais. Essa orientação pode ter contribuído para que as reuniões passassem a ser majoritariamente virtuais, utilizando recursos de videoconferência. Dentre as ferramentas disponíveis para videoconferências, citam-se: *WhatsApp*[1], *Skype*[2],

[1]https://www.whatsapp.com/
[2]http://www.skype.com/en/

Google Hangouts[3], *WebEx*[4] e *GoToMeeting*[5]. Todas as ferramentas apresentadas têm arquitetura cliente-servidor. Dessas ferramentas, apenas o *WhatsApp* e o *Skype* oferecem, paralelamente, suporte à arquitetura *Peer-to-Peer* (P2P) [25].

Outro fator atrativo dessas ferramentas é que as chamadas de voz são baseadas em VoIP (*Voice Over Internet Protocol*). Uma de suas vantagens é custo, pois as ligações entre usuários da mesma plataforma, em geral, são gratuitas, sendo necessária apenas uma conexão de Internet. Ao contrário da telefonia tradicional, o VoIP não exige a construção de uma infraestrutura exclusiva. A vantagem econômica é tão relevante, que há uma tendência para que as operadoras de telefonia substituam o atual sistema de comutação de circuito pelo VoIP [65]. Diversos aplicativos de mensagens instantâneas oferecem o serviço de VoIP, tornando-o ainda mais popular.

2.3 Área de Trabalho Remota

Nessa solução, um aplicativo instalado no dispositivo cliente é responsável por capturar e enviar ao servidor as entradas de teclado e *mouse* e receber e exibir para o usuário uma imagem que é resultado da execução do *software* ocorrida totalmente no servidor [80]. Os dispositivos clientes podem ser *smartphones*, *tablets* ou computadores, utilizando uma variedade de sistemas operacionais instalados. De acordo com as configurações e restrições impostas pelo administrador do sistema, essa tecnologia pode permitir acesso ao compartilhamento de arquivos e aos *softwares* corporativos.

Há uma grande variedade de protocolos de área de trabalho remota: o *Microsoft Remote Desktop Protocol (RDP)* [43], o *VMWare PC-over-IP (PCoIP)* [77] e o *Citrix's ICA / HDX* [74] são os mais comumente usados.

Dentre as vantagens dessa técnica estão o baixo custo, baixo tráfego de dados, alto nível de controle sobre os *softwares* utilizados, computação móvel e fácil gerenciamento.

2.4 VPN (*Virtual Private Network*)

As redes privadas virtuais são enlaces virtuais para estabelecer interconexão entre entidades. Esses enlaces são privados, o que significa dizer que somente têm acesso a eles

[3]https://www.google.com/hangouts/
[4]http://www.webex.com/
[5]http://www.gotomeeting.com/online/

os usuários autorizados, sendo formalmente definida como “um ambiente de comunicação construído pela segmentação controlada de uma infraestrutura de comunicação compartilhada para emular as características de uma rede privada.” [76].

A VPN possibilita que um dispositivo remoto, geograficamente distante da rede local, possa ingressar na rede e ter acesso aos seus recursos e serviços, como se nela estivesse fisicamente. Nesse ponto está a principal diferença entre a VPN e a área de trabalho remota. Assim, se houver a necessidade da utilização remota de um *software* que exige instalação no dispositivo, usando a VPN, ele deverá estar instalado no dispositivo remoto, o que não acontece com a área de trabalho remota. Isso implica em restrições quanto ao sistema operacional, por exemplo. Enquanto os *softwares* executados no servidor da área de trabalho remota podem ser visualizados e controlados a partir de diversos tipos de dispositivos e sistemas operacionais diferentes, a VPN por si só não garante essa característica.

Uma das grandes vantagens da VPN é o mecanismo de tunelamento, que pode ser criptografado para atribuir maior grau de segurança à rede virtual.

A desvantagem, em relação à área de trabalho remota é que, em virtude de o processamento ocorrer no dispositivo cliente, todas as informações necessárias a esse processamento trafegam entre a rede remota e o dispositivo cliente. Isso não acontece na utilização da área de trabalho remota, onde o tráfego, em geral, se restringe às entradas de teclado e mouse e às imagens que são resultado da execução do *software* ocorrida totalmente no servidor.

2.5 Aspectos de Segurança

Um debate importante sobre essas ferramentas diz respeito à segurança. Nesse quesito, sua utilização massiva, de forma repentina, não fugiu à regra. O aumento da utilização de meios eletrônicos e das redes, como a Internet, se traduz em oportunidade para os cibercriminosos. Tendo em vista que o usuário representa, normalmente, o elo mais fraco da segurança da informação digital [15] e que, com o tele trabalho, a utilização de dispositivos computacionais residenciais, livres do controle organizacional, aumentou no período da pandemia, o risco do comprometimento de informações tende a aumentar.

Embora não tenha havido aumento significativo no número de ataques, cujo nível foi considerado moderado, a popularização das videochamadas introduziu uma nova fonte de vulnerabilidades no dia-a-dia dos usuários, segundo um estudo realizado por especialistas em segurança da *Kaspersky* [63]. Ainda de acordo com o mesmo estudo, observou-se,

somente em relação ao *Skype*, um total de mais de 120 mil arquivos suspeitos disfarçados da aplicação original.

Na pesquisa realizada neste estudo, o aplicativo de videoconferência *Zoom* ocupa o segundo lugar entre os mais usados em sua categoria, com 56,5% de adesão. No entanto, em consulta à base de dados do *Common Vulnerabilities and Exposures* (CVE)[6], foram encontrados, considerando apenas entradas do ano de 2020, registros de vulnerabilidades da seguinte ordem: uma crítica, sete altas e uma baixa. Em face disso, a utilização do aplicativo foi proibida pela Agência Nacional de Vigilância Sanitária (Anvisa), pela fabricante de foguetes *SpaceX* e pelas Forças Armadas Australianas [64]. A procuradoria-geral de Nova York chegou a cobrar explicações a empresa desenvolvedora do aplicativo, citando, inclusive, um suposto compartilhamento de dados com o *Facebook* [82]. Contudo, utilizando o mesmo critério, destacam-se o *Cisco Webex*, cujos registros de vulnerabilidades são: uma crítica, oito altas, quatorze médias e quatro baixas e o *WhatsApp* com duas aguardando análise, três críticas, cinco altas, três médias e uma baixa. O *Skype* apresentou apenas uma vulnerabilidade crítica e duas médias. Não foram encontrados registros de vulnerabilidades em 2020 para o *Facebook Messenger*, *Google Meet* e para o *Microsoft Teams*. Com isso, observa-se que os aplicativos mais utilizados para videoconferência apresentam os maiores índices de registros de vulnerabilidades, o que leva à exposição de parcela significativa dos dispositivos.

As ferramentas de área de trabalho remota propiciam uma certa estanqueidade entre os ambientes computacionais domésticos e corporativos, ofertando ao administrador do sistema maior controle sobre as atividades permitidas a serem realizadas, uma vez que o processamento e o armazenamento de informações ocorre em um dispositivo localizado na empresa.

A VPN, no entanto, insere o dispositivo cliente na rede corporativa. Com isso, o fator de risco é significativamente majorado. Dentre as ameaças às quais a rede corporativa passa a ser exposta, pode-se citar o ataque negação de serviço, além do furto de dados. O dispositivo cliente, no caso de tele trabalho, é um dispositivo não gerenciado. Dessa forma, a corporação não tem controle sobre as ferramentas e políticas de segurança aplicadas ao dispositivo. Uma possível forma de contornar o problema é a criação e distribuição de máquinas virtuais com todas as políticas de segurança aplicadas e ferramentas de proteção instaladas. Porém, se o *hypervisor* for comprometido, todo o modelo de segurança é quebrado [60]. Além disso, há *malwares* que, através das características da virtualização,

[6]https://cve.mitre.org/

conseguem obter elevação de privilégios no sistema operacional da máquina virtual [70].

Desse modo, observa-se a necessidade de que o ambiente computacional para o *home office* seja adequadamente protegido através de ferramentas de segurança, tais como *firewall* e antivírus [16, 71]. A segurança deve ser ainda maior se a rede local for sem fio.

2.6 *Machine Learning*

Ao defrontar problemas de grande complexidade e difícil resolução, há a demanda pelo desenvolvimento de ferramentas computacionais que permitam solucioná-los. Como exemplo, pode-se citar o aprendizado de máquina. Como vantagem, o aprendizado de máquina possibilita que as decisões tomadas sejam continuamente aprimoradas a partir da introdução de novas informações, sem a necessidade de mudar o código da aplicação. No contexto do presente estudo, foi julgado pertinente comparar o desempenho de seis algoritmos diferentes para a construção de classificadores capazes de predizer a pontuação, a severidade, as métricas do vetor CVSS e o tempo de correção das vulnerabilidades, com base, essencialmente, em sua descrição textual. Destarte, os algoritmos avaliados foram: *Support Vector Machine* [55, 42], *Random Forest* [10], *Naive Bayes for Multinomial Models* [1], *Multi-layer Perceptron* [11], *Passive Aggressive* [14] e *Logistic Regression* [57].

2.6.1 Algoritmos

A escolha do algoritmo mais adequado para um problema específico de NLP depende de vários fatores. Dentre eles estão as características do problema, o porte e a complexidade dos dados ora analisados e a disponibilidade de recursos computacionais. Face ao exposto, é importante avaliar diferentes algoritmos e escolher aquele que melhor se adapta às necessidades do problema em questão, no caso em tela a predição da severidade de vulnerabilidades e, se viável, outros atributos da vulnerabilidade. Nesse contexto, esta subseção apresenta os algoritmos utilizados no presente trabalho, os quais foram selecionados por contarem com características que os indiquem para predição aplicada a problemas de NLP.

Support Vector Machine

O *Support Vector Machine* (SVM) é um algoritmo que usa o método de aprendizagem supervisionado e pode ser usado tanto para classificação quanto para regressão.

Considerando que a maioria dos problemas de categorização de texto são linearmente separáveis [33], optou-se por utilizar, neste trabalho, o *kernel* linear. Desse modo, o objetivo do algoritmo é encontrar o hiperplano que melhor diferencia duas classes. Tendo em vista que o *dataset* é multi-classe, a abordagem *one-vs-one* foi utilizada. Nesse caso, o total de classificadores (internamente construídos) é representado por: $\frac{n\_classes*(n\_classes-1)}{2}$. Cada um desses classificadores treina duas classes.

Random Forest

O algoritmo *Random Forest* usa o método de aprendizagem supervisionado e pode ser usado tanto para classificação quanto para regressão. Consiste em construir, durante o processo de treinamento, diversas árvores de decisão individuais. Cada árvore é construída usando um subconjunto aleatório dos dados de treinamento e um subconjunto aleatório das características (unigramas, bigramas ou trigramas) disponíveis. Isso permite que o modelo seja menos suscetível a *overfitting*, ou seja, melhor generalização do modelo para dados não vistos. O resultado da predição é a classe que obtiver mais votos após a avaliação da informação de entrada por cada uma das árvores existentes na floresta. O número de árvores selecionado após testes empíricos, especialmente considerando tempo de processamento e tamanho do classificador, foi de 400.

Naive Bayes for Multinomial Models

Também utilizando o método de aprendizagem supervisionada, o algoritmo *Naive Bayes for Multinomial Models*, diferente dos algoritmos anteriores, é utilizado em *datasets* cujas classes sejam discretas, ou seja, não é aplicável em problemas de regressão. Embora haja diversas variantes do modelo *Naive Bayes*, o *Multinomial* foi escolhido para compor o comparativo conduzido neste trabalho em virtude de ser frequentemente usado na classificação de textos [1]. Nesse algoritmo, a classificação é realizada de forma probabilística, com base no teorema de Bayes, assumindo que há independência entre pares de *features*.

Multi-layer Perceptron

O *Multi-layer Perceptron* (MLP) é uma rede neural, cujo algoritmo de aprendizado é supervisionado. Possui arquitetura formada por camadas interconectadas. Nele, há três tipos de camadas: a camada de entrada, a camada de saída e a camada oculta. A camada de entrada recebe a informação a ser processada e a camada de saída é responsável pela efetiva classificação. Entre a camada de entrada e de saída há um número arbitrário de camadas ocultas, nas quais seus neurônios transformam os valores oriundos da camada

anterior através de um peso. Devido a sua capacidade de aprender relações não lineares entre as entradas e as saídas, se torna adequado para uma variedade de tarefas, dentre elas a classificação de texto.

Passive-Aggressive

O *Passive-Aggressive* é um algoritmo de classificação, categorizado como *online learning*, *incremental* ou ainda *out-of-core learning*. Normalmente, ao treinar os modelos, eles não mudam mais. Caso haja novos dados, será necessário treinar um novo modelo. O problema dessa abordagem tradicional é que o modelo está quase sempre desatualizado. O *online learning* tenta resolver esse problema, possibilitando o aprendizado próximo ao tempo real. Ele considera que o ambiente de aprendizado é dinâmico e, dessa forma, o modelo pode mudar constantemente.

Quanto ao funcionamento, a cada iteração ele avalia o resultado da classificação. Caso ela esteja correta, não há alteração no modelo, motivo pelo qual parte do nome é *Passive*. No entanto, se a classificação estiver incorreta, o modelo é reajustado, motivo pelo qual parte do nome é *Aggressive*. A atualização do modelo à medida que novos dados são apresentados, demonstra que esse algoritmo é indicado para problemas de NLP, especialmente aqueles que envolvem grandes volumes de dados [69].

Logistic Regression

Apesar do nome, o *Logistic Regression* é um algoritmo de classificação supervisionado e não de regressão. Logo, é utilizado em *datasets* cujas classes sejam discretas. Esse algoritmo é comumente usado em classificação de e-mail SPAM, transações financeiras (classificar em fraudulenta ou não) e aplicações voltadas para a medicina (como análise de exames de imagem para sugerir patologias). O *Logistic Regression* pode ser do tipo binário, quando há duas classes, ou multinomial, quando há três ou mais classes possíveis. Seu funcionamento baseia-se na medição da relação entre a classe e uma ou mais *features*, estimando probabilidades, que são transformadas em valores binários.

Organicamente o *Logistic Regression* é limitado a problemas binários, ou seja, de duas classes. Para habilitar o algoritmo a resolver problemas multi-classe, pode ser adotada a técnica *one-vs-all* ou *one-vs-one*. No entanto, isso exigiria que a classificação fosse dividida em diversas outras classificações binárias. Neste estudo, foi utilizada a extensão multinomial do *Logistic Regression*, que altera função de perda para perda de entropia cruzada, possibilitando que o *Logistic Regression* suporte nativamente a resolução de problemas multi-classe.

2.6.2 Redução de Dimensionalidade

A redução de dimensionalidade é uma técnica usada para reduzir a quantidade de *features* (características) em um conjunto de dados. No entanto, essa redução é efetivada mantendo-se o máximo de informação possível. Em problemas de predição baseados em NLP, a redução de dimensionalidade é muito útil, pois os dados geralmente incluem muitos termos (*i.e.*, sequências de uma, duas ou três palavras), que podem tornar a análise e a modelagem mais complexas e demoradas.

Principal Component Analysis (PCA)

Devido a alta dimensionalidade presente no processamento do NLP, o PCA é um algoritmo muito útil para otimizar a solução de problemas de predição baseados em NLP [20]. Seu objetivo é reduzir a dimensionalidade de um conjunto de dados, preservando a maior quantidade possível de variância dos dados originais. Isso é realizado através da identificação dos principais recursos, ou seja, aqueles que são mais importantes para a tarefa de predição.

Linear Discriminant Analysis (LDA) on the Principal Component Analysis (PCA)

O LDA é um algoritmo supervisionado, que visa encontrar uma projeção linear que maximize a separação entre as classes [72]. Dessa forma são selecionadas novas dimensões, evitando o *overfitting* e reduzindo o custo computacional.

Uma técnica, usada neste estudo, é aplicar a redução de dimensionalidade primeiro através do algoritmo PCA e depois aplicar o algoritmo LDA para reduzir ainda mais a dimensionalidade. Assim, melhorando a precisão do modelo.

Capítulo 3

Impactos da Pandemia nas Redes de Computadores

3.1 Impactos da Pandemia na Percepção do Usuário

Esta seção tem por objetivo apresentar as duas pesquisas de opinião que incluíram a análise de variáveis quantitativas, no que tange, por exemplo, aos aplicativos utilizados e à banda disponível no plano de Internet fixa contratada, e variáveis qualitativas, no que se refere à estabilidade da conexão de Internet e à tecnologia utilizada na rede sem fio residencial. Em ambas as pesquisas, foi aplicada a estatística descritiva [24] a fim de identificar as informações mais relevantes acerca das variáveis pesquisadas, possibilitando o cruzamento de informações, objetivando a obtenção de inferência estatística.

As pesquisas foram aplicadas através de formulários eletrônicos. Os *links* para as pesquisas foram distribuídos em grupos de *WhatsApp* e *Telegram* da comunidade acadêmica da Universidade Federal Fluminense (UFF), grupos de trabalho, de amigos, da família e de moradores da cidade de Niterói-RJ, além de publicado em páginas de redes sociais.

3.1.1 Primeira Pesquisa de Opinião

Para entender melhor a mudança no perfil de utilização da Internet, em virtude da pandemia pelo novo Coronavírus, foi realizada a primeira pesquisa de opinião. Para viabilizar a pesquisa, foi elaborado um questionário eletrônico, cujas perguntas estão descritas no Apêndice A, utilizando a ferramenta gratuita *Google Forms*[1].

A pesquisa (Apêndice A) ficou disponível do dia 1º de junho de 2020 ao dia 13 de

[1]https://www.google.com/forms/

junho de 2020. No período em que a pesquisa ficou ativa, 379 pessoas, em 11 estados brasileiros, na faixa etária de 9 a 76 anos, responderam ao questionário.

3.1.1.1 População e Amostra

A única limitação do escopo populacional da pesquisa foi quanto ao país (Brasil). Foi utilizada uma amostra aleatória, sendo os participantes distribuídos conforme ilustrado nas Figuras 3.1, 3.2, 3.3 e 3.4. Destaca-se que, embora a pesquisa de opinião tenha registrado respondentes de 11 onze estados brasileiros, a maior parte dos respondentes eram do estado do Rio de Janeiro, no qual a cidade com maior destaque é a de Niterói.

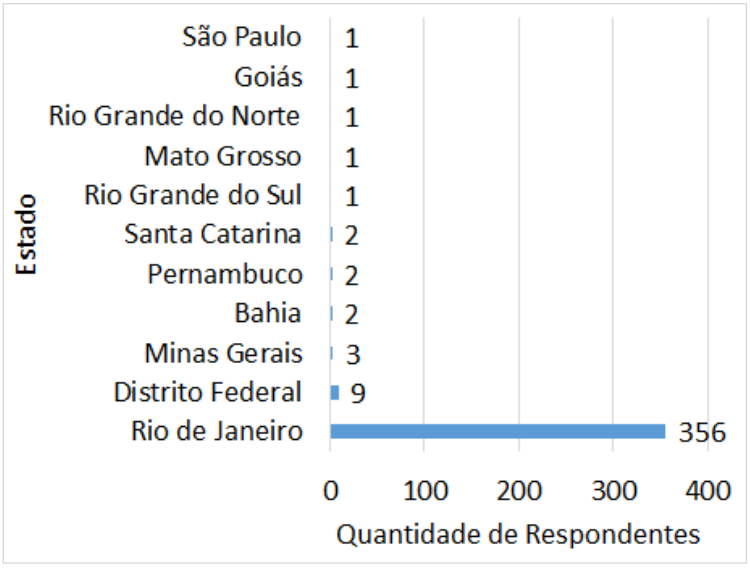


Figura 3.1: Respostas por estado, referentes à primeira pesquisa de opinião

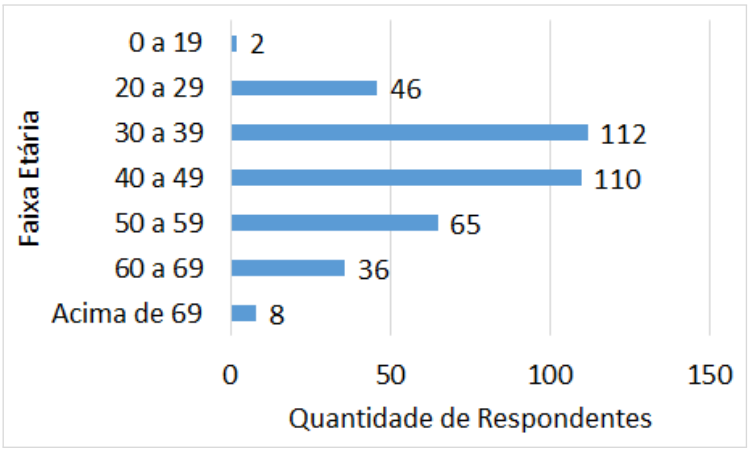


Figura 3.2: Respostas por faixa etária, referentes à primeira pesquisa de opinião

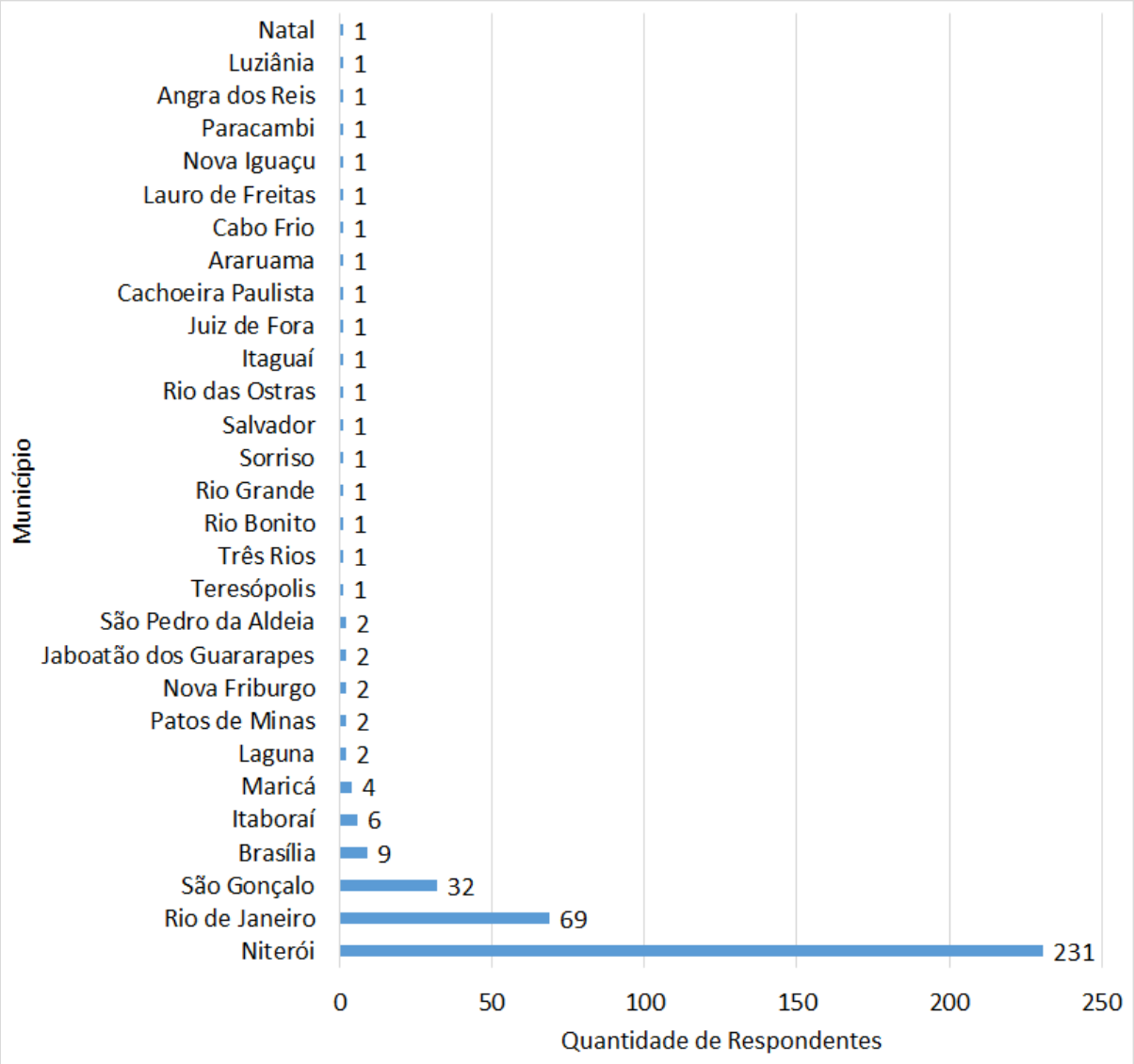


Figura 3.3: Respostas por município, referentes à primeira pesquisa de opinião

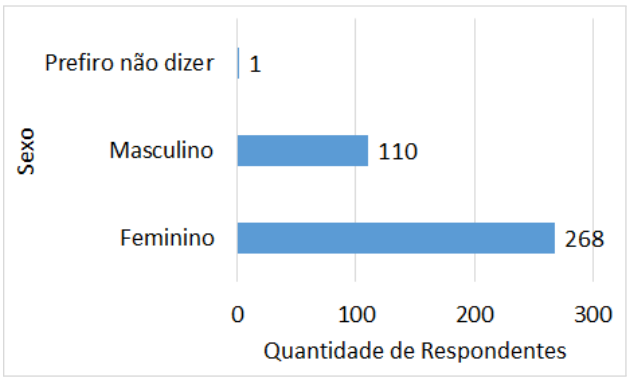


Figura 3.4: Respostas por sexo, referentes à primeira pesquisa de opinião

3.1.1.2 Análise de Dados

Os resultados da pesquisa, ilustrados na Figura 3.5, demonstram que 105 pessoas (27,7%) aumentaram a intensidade de uso da Internet com fins profissionais, enquanto 33 (8,7%)

reduziram e 241 (63,6%) mantiveram a intensidade. A quantidade de pessoas que utilizam a Internet com muita intensidade, para trabalho, passou de 155 (40,9%) para 230 (60,7%), durante a pandemia. Isso representa um aumento de 48,4%.

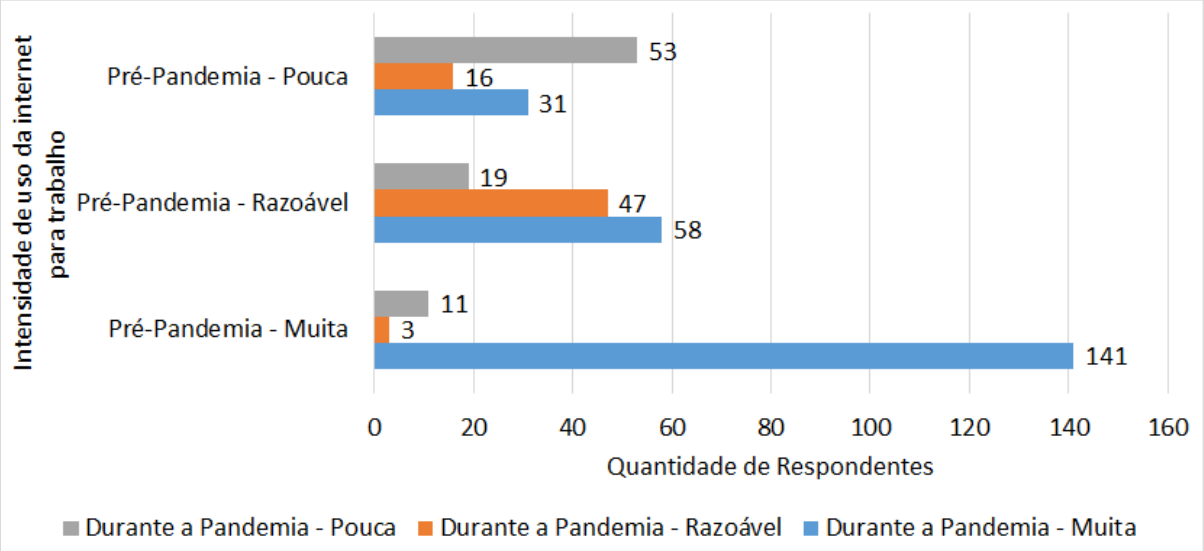


Figura 3.5: Comparativo da intensidade de uso da Internet para trabalho, antes e durante a pandemia, referente à primeira pesquisa de opinião

Observa-se, ainda, em consonância com a Figura 3.6, que 96 pessoas (25,3%) aumentaram a intensidade de uso da Internet para estudo, enquanto 34 (9%) reduziram e 249 (65,7%) mantiveram a intensidade. A quantidade de pessoas que utilizam a Internet com muita intensidade, para estudo, passou de 141 (37,2%) para 206 (54,4%), durante a pandemia, um aumento de 46,1%. Foi possível, também, com base na Figura 3.7, verificar que 98 pessoas (25,9%) aumentaram a intensidade de uso da Internet também para lazer e/ou entretenimento, enquanto 27 (7,1%) reduziram e 254 (67%) mantiveram a intensidade. A quantidade de pessoas que utilizam a Internet com muita intensidade, para lazer e/ou entretenimento, passou de 206 (54,4%) para 273 (72%), durante a pandemia, representando um aumento de 32,5%. Assim, identifica-se um expressivo aumento na intensidade de utilização da Internet, que é similar para todos os tipos de utilização (trabalho, estudo e lazer), variando entre 25,3% e 27,7%.

No entanto, a utilização da Internet com muita intensidade registrou, nesta pesquisa, aumento entre 32,5% e 48,4%. Esse aumento foi muito próximo para as atividades de trabalho e estudo, entre 46,1% e 48,4%. Para a atividade de lazer e entretenimento, o aumento foi menor, embora expressivo, atingindo o índice de 32,5%. O aumento menor se deve ao fato de que esta atividade já era muito demandada anteriormente. Assim, as atividades de lazer e/ou entretenimento continuam sendo as atividades mais demandadas,

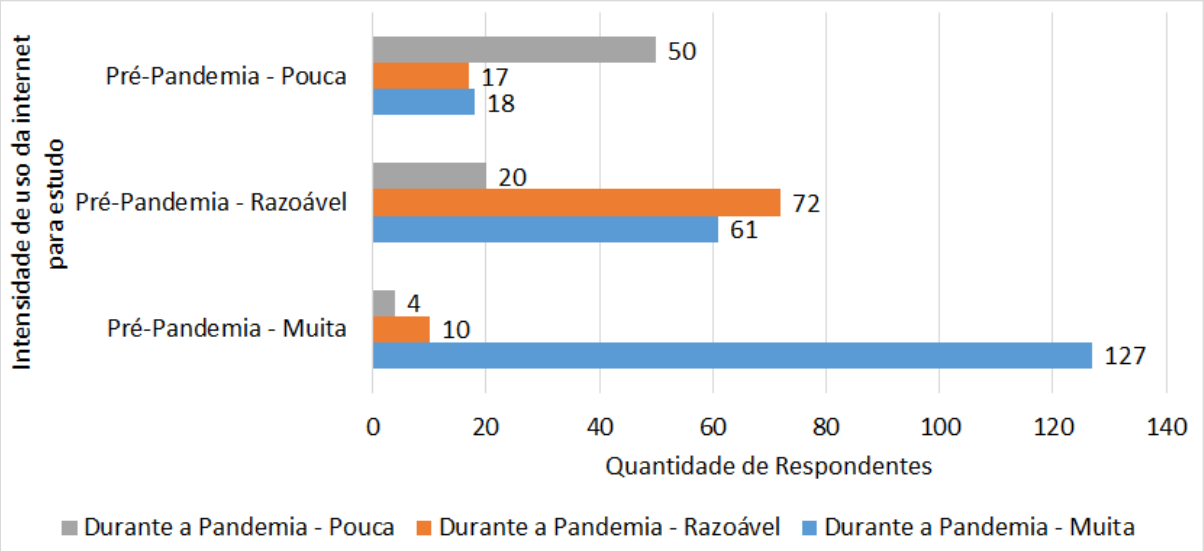


Figura 3.6: Comparativo da intensidade de uso da Internet para estudo, antes e durante a pandemia, referente à primeira pesquisa de opinião

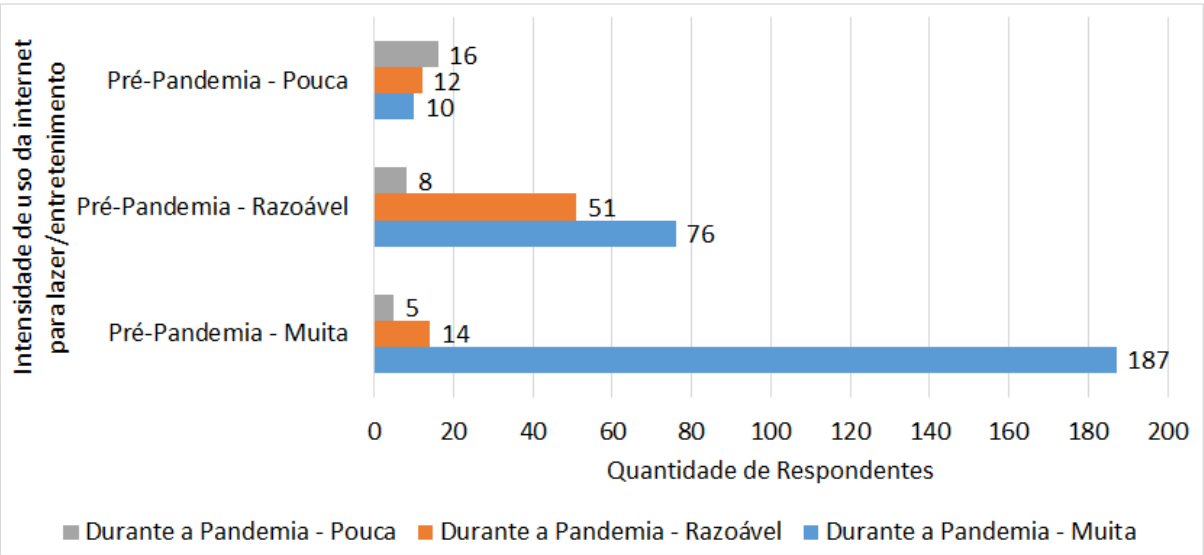


Figura 3.7: Comparativo da intensidade de uso da Internet para lazer/entretenimento, antes e durante a pandemia, referente à primeira pesquisa de opinião

com muita utilização relatada por 72% dos respondentes. Por outro lado, as atividades de trabalho e estudo registram os índices de 60,7% e 54,4%, respectivamente. Releva destacar que o tráfego relacionado ao lazer e/ou ao entretenimento, normalmente, é aquele com o maior consumo banda. Seu emprego, em geral, está associado aos tráfegos mais densos, como *streaming* de vídeo e jogos online. No entanto, esse tipo de tráfego é, normalmente, menos sensível às pequenas falhas nos enlaces de dados, já que podem contar um *buffer* para amenizá-las ou, até mesmo, suprimi-las.

Uma informação importante que pode ser obtida com a análise dos dados coletados

pela pesquisa é que, após a decretação da pandemia, 47% dos participantes passou a trabalhar em regime de tele trabalho, seja em regime integral ou parcial (Fig. 3.8), o que certamente colaborou para o aumento do tráfego [47]. O aumento do tráfego também pode ter influenciado a percepção sobre a estabilidade da conexão com a Internet. Dentre os que relataram uma percepção de uma conexão com a Internet instável, 64% disseram que a instabilidade iniciou após a decretação da pandemia. No entanto, ao somar a esse indicador à quantidade de pessoas que relataram que já utilizavam uma conexão com a Internet instável, mesmo antes da pandemia, tem-se a indicação de que 61,5% das pessoas enfrentam dificuldades para terem acesso à rede mundial de computadores (Fig. 3.9).

Embora uma grande parte dos participantes tenha relatado que a percepção de instabilidade na conexão com a Internet começou durante a pandemia, 24% deles afirmaram já enfrentar dificuldades com a conectividade antes mesmo da declaração oficial do estado pandêmico. Uma hipótese não confirmada, mas que pode ter influenciado a experiência do usuário, é que, ao contrário do *streaming* de vídeo, as aplicações de comunicação (*e.g.*, chamadas de voz e videoconferências), a área de trabalho remota e outras ferramentas e protocolos relacionados às atividade de teletrabalho são mais sensíveis às falhas de conexão. Nas aplicações de comunicação, pode haver atraso ou até perda de partes da conversa, enquanto nas ferramentas e protocolos de trabalho remoto, a falta de resposta imediata na tela após um clique do mouse ou pressionamento de uma tecla reduz a interatividade do usuário. Esses fatores também podem contribuir para uma percepção negativa da estabilidade da conexão com a Internet.

No que se refere à percepção sobre a estabilidade da conexão com a Internet, foi observada a ausência de correlação com a quantidade de dispositivos conectados à Internet (Fig. 3.10): a quantidade de respondentes que relataram conexão estável para residências onde há de 1 a 5 dispositivos conectados mantém equilíbrio com aquelas residências onde há de 6 a 10 dispositivos conectados. O mesmo equilíbrio foi constatado para o cenário de conexão instável.

De igual modo, as percepções sobre a estabilidade da conexão com a Internet foram comparadas com a quantidade de pessoas residentes no mesmo imóvel. É possível constatar que a proporção de residências com até três moradores, em relação às residências com mais de três moradores, é similar para todas as percepções, variando entre 77,7% e 85% (Fig. 3.11). Ao confrontar esse indicador com a tabela de domicílios particulares permanentes, por tipo do domicílio e número de moradores, do censo demográfico de 2010, do IBGE [28], verifica-se eles guardam uma correspondência entre si. Pois, as residências

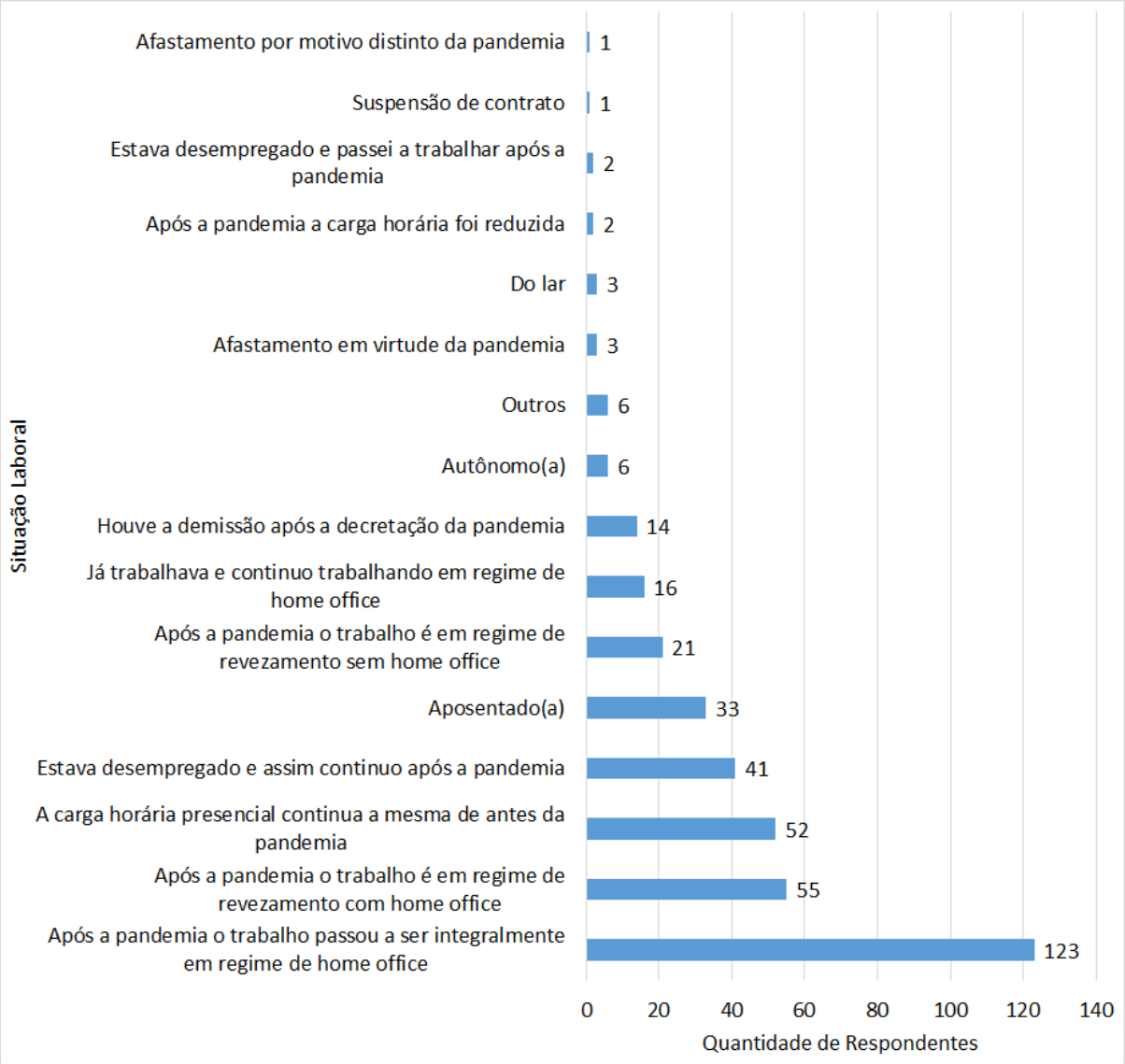


Figura 3.8: Situação laboral dos respondentes, referente à primeira pesquisa de opinião

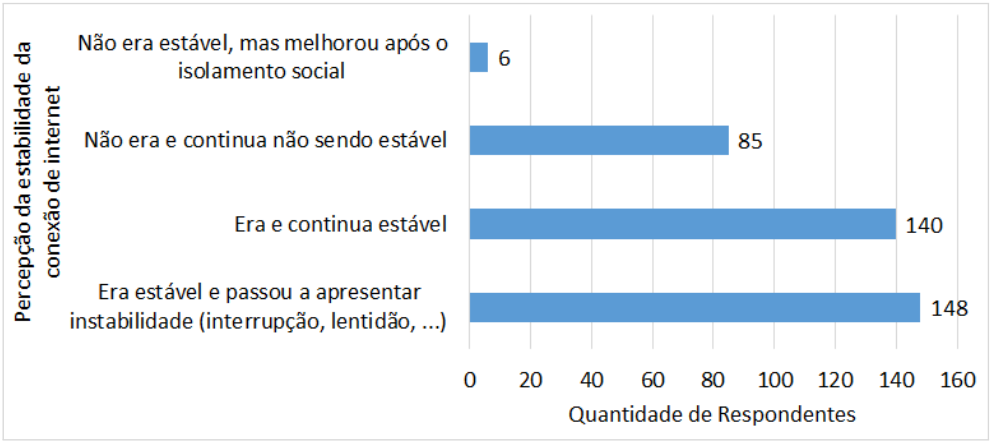


Figura 3.9: Percepção dos respondentes quanto à estabilidade da conexão com a Internet, referente à primeira pesquisa de opinião

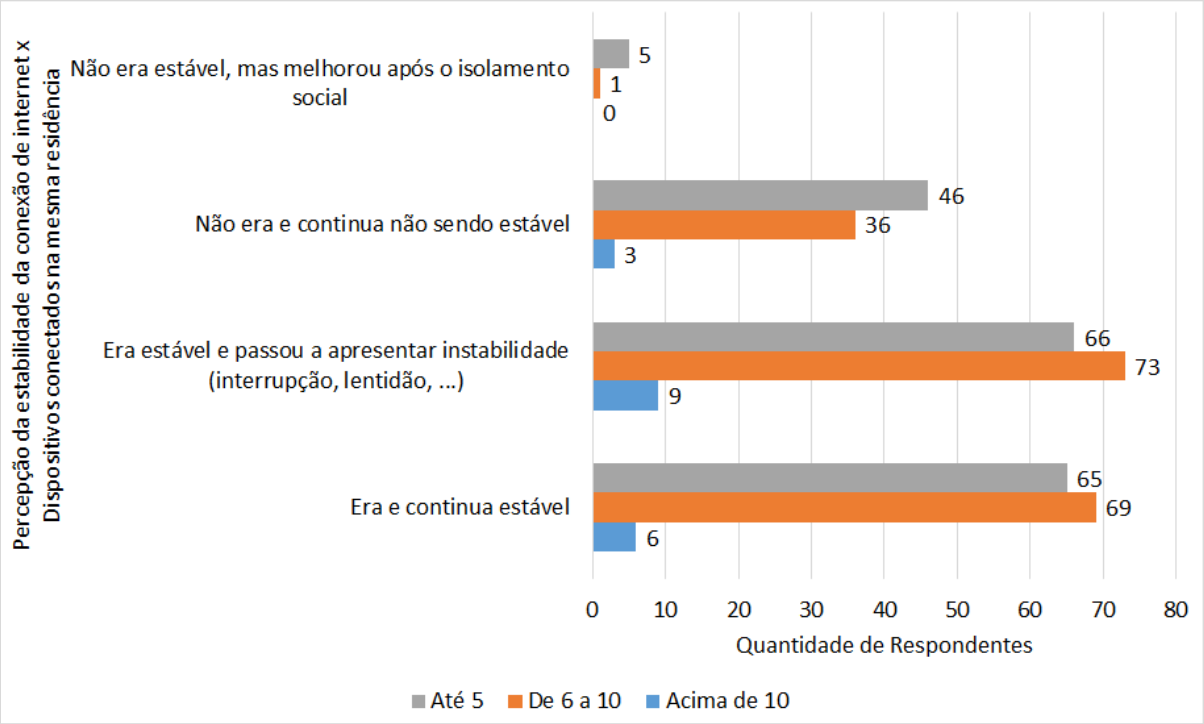


Figura 3.10: Correlação entre a percepção da estabilidade da conexão com a Internet e a quantidade de dispositivos conectados à Internet na residência, referente à primeira pesquisa de opinião

com até 3 moradores, naquele ano, somavam 59,28%.

Equitativamente, em todas as percepções sobre a estabilidade da conexão com a Internet, nota-se similitude na distribuição para os planos de Internet fixa contratados (Fig. 3.12). Em outro comparativo, para cada plano de Internet fixa contratado (Fig. 3.13), nota-se que os respondentes que consideraram que a estabilidade de sua conexão com a Internet piorou após o isolamento social somado àqueles que consideraram que a estabilidade de sua conexão com a Internet manteve-se igualmente instável é o grupo majoritário. Faz-se mister ressaltar que os planos de Internet fixa contratados que em os respondentes indicaram o menor nível de insatisfação foram aqueles até 5 Mbps. Assim, pode-se aduzir que o plano de Internet fixa contratado não representa fator decisivo na percepção dos respondentes sobre a estabilidade de sua conexão com a Internet.

As correlações estudadas e ilustradas nas Figuras 3.10, 3.11, 3.12 e 3.13 podem indicar que o problema na percepção da instabilidade da conexão com a Internet não tem, necessariamente, origem nas condições das redes de dados residenciais ou nos planos contratados. A maior parte das pessoas utiliza enlace com banda acima de 5 Mbps. Sendo assim, a maior parte das situações laborais encontram-se nessa faixa. Dentre os que utilizam enlaces de até 5 Mbps, 64,3% não tiveram alteração na situação laboral (Fig.

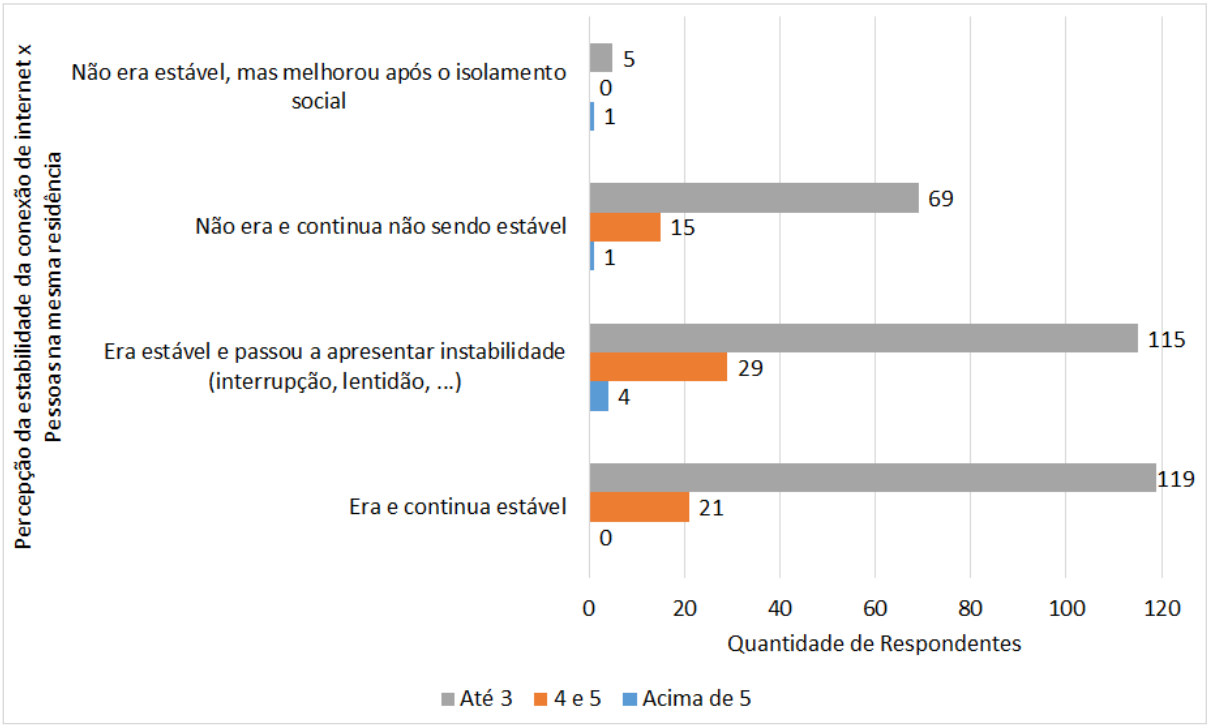


Figura 3.11: Correlação entre a percepção da estabilidade da conexão com a Internet e a quantidade de moradores na mesma residência, referente à primeira pesquisa de opinião

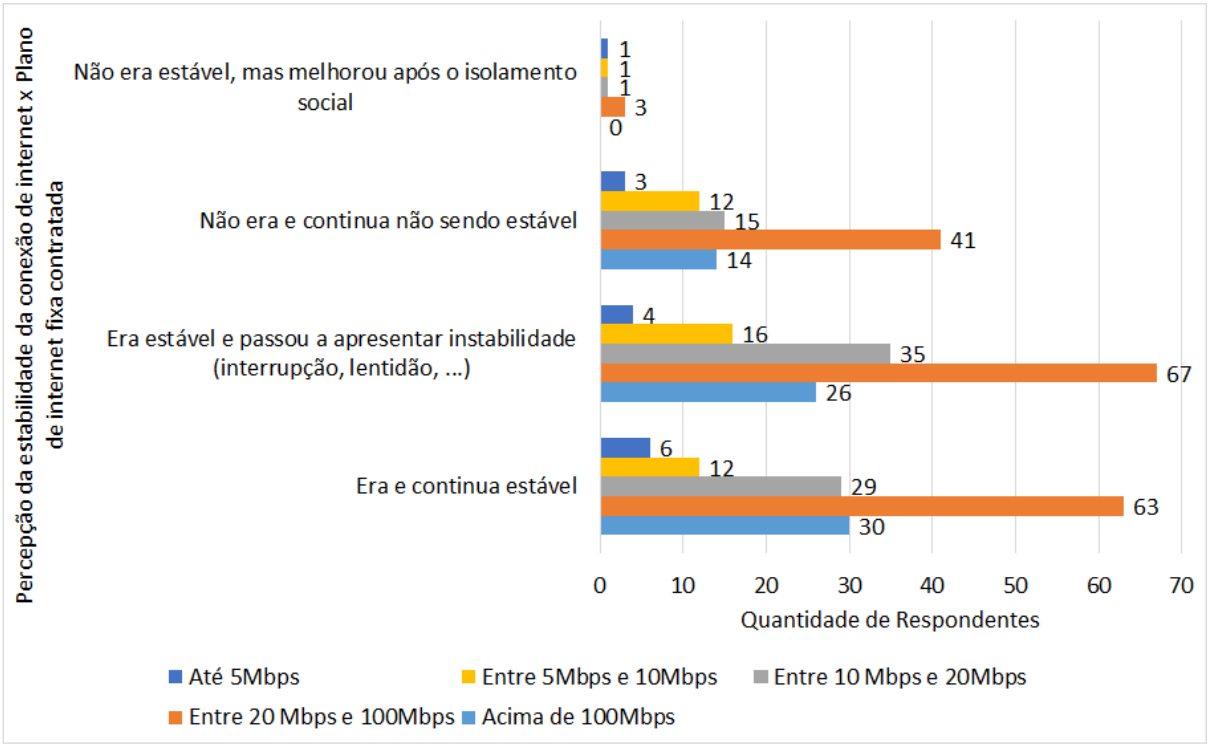


Figura 3.12: Correlação entre a percepção da estabilidade da conexão com a Internet e o plano de Internet fixa contratado, referente à primeira pesquisa de opinião

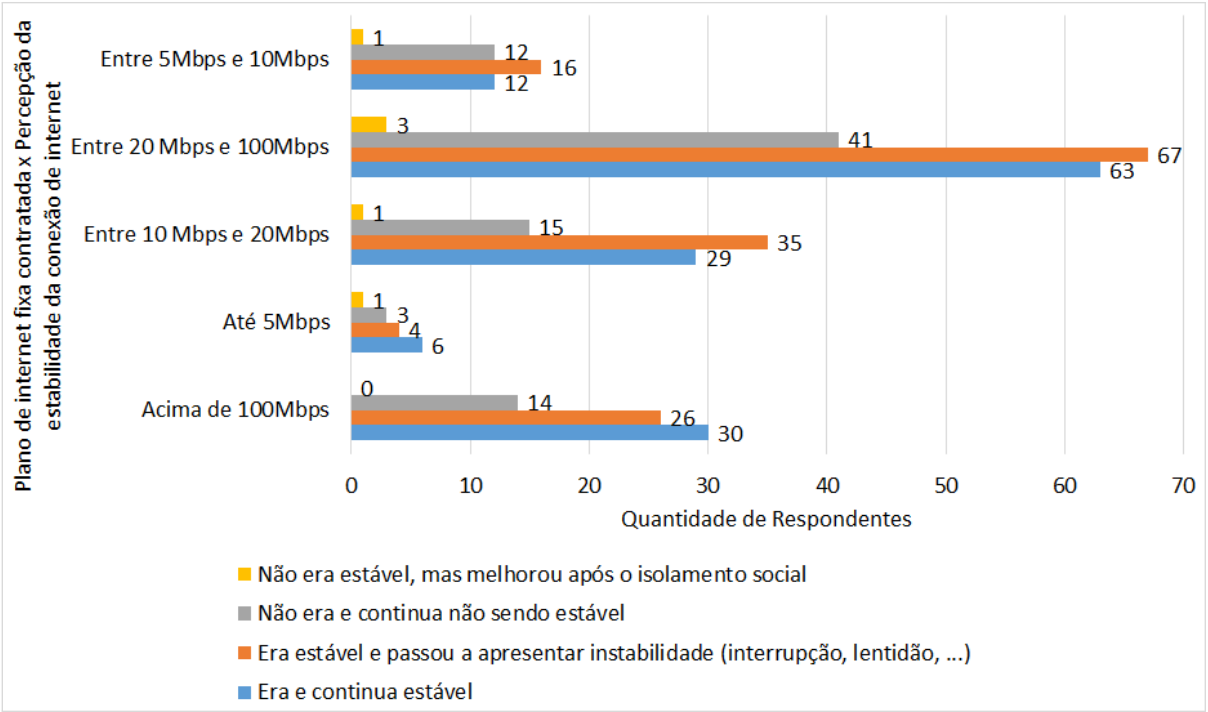


Figura 3.13: Correlação entre o plano de Internet fixa contratado e a percepção da estabilidade da conexão com a Internet, referente à primeira pesquisa de opinião

3.14).

Em relação aos aplicativos de mensagens instantâneas, a pesquisa revelou que quase a totalidade dos respondentes utiliza o *WhatsApp* (99,7%), conforme observa-se na Figura 3.15. No entanto, para a função de videoconferência/chamada de vídeo, essa relação cai para 81,5%, em que pese a manutenção da liderança e a alta taxa de utilização, seguido por: *Zoom* (56%), *Google Meet* (37%), *Skype* (28%), *Microsoft Teams* (19%), *Facebook Messenger* (13%) e *Cisco Webex* (8%), como ilustrado na Figura 3.16.

3.1.2 Segunda Pesquisa de Opinião

Tendo como base a análise das respostas da primeira pesquisa de opinião, particularmente no que concerne à percepção de instabilidade na conexão com a Internet por parte majoritária dos respondentes, criou-se a hipótese de que a origem desse problema pudesse estar relacionada à faixa de frequência usada nas redes sem fio domésticas. Outra questão investigada foi se houve mudança no plano de Internet fixa contratado, a fim de dirimir o problema da instabilidade na conexão com a Internet. Além dessas motivações, a segunda pesquisa de opinião também auxiliou na avaliação da evolução dos impactos causados pela pandemia. Dentre eles, além da própria percepção dos respondentes, está também a intensidade de uso de aplicações de comunicação. Assim como na primeira pes-

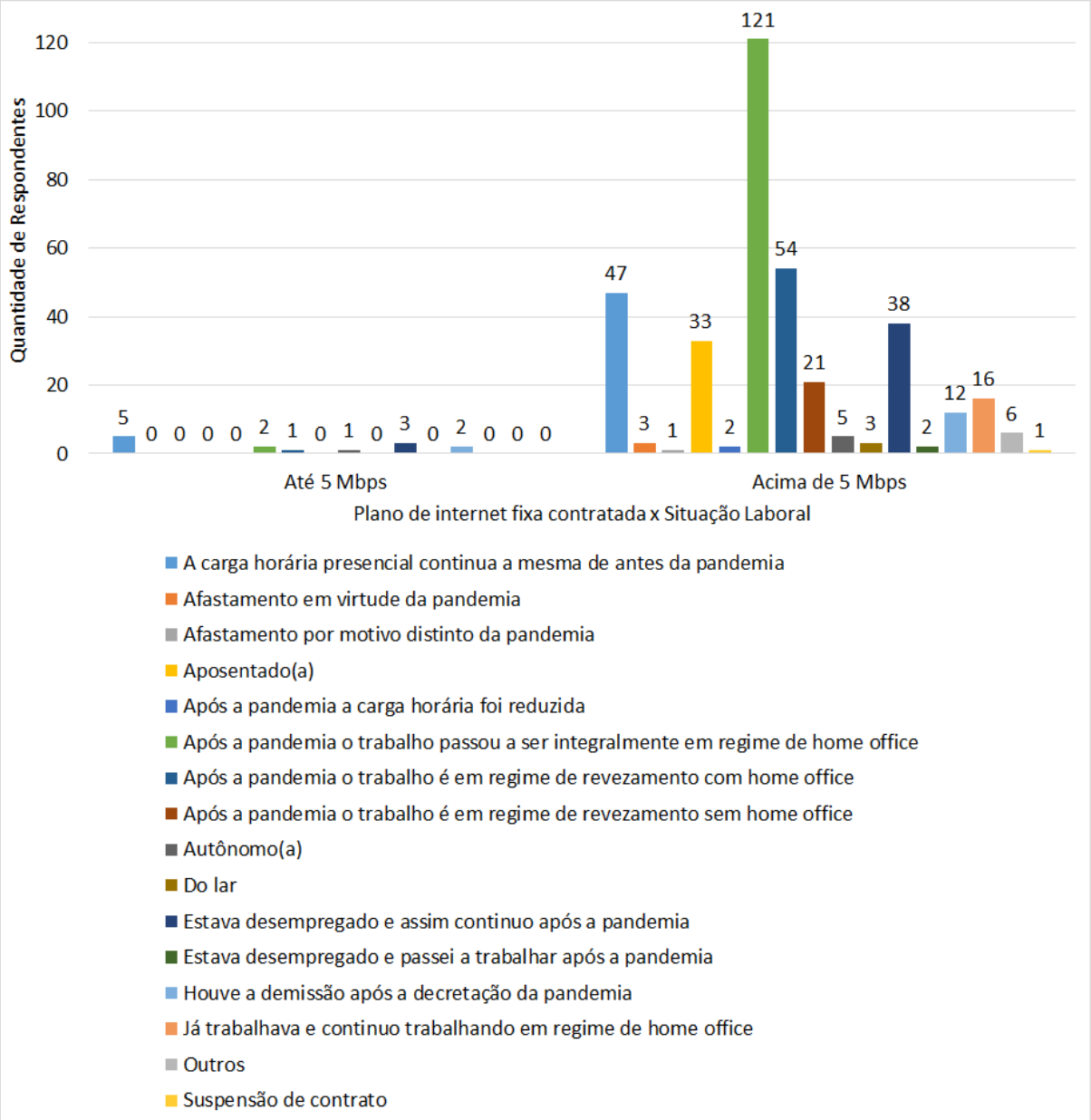


Figura 3.14: Correlação entre o plano de Internet fixa contratado e a situação laboral dos respondentes, referente à primeira pesquisa de opinião

quisa de opinião, foi elaborado um questionário eletrônico, cujas perguntas estão descritas no Apêndice B, utilizando a ferramenta gratuita *Google Forms*[2].

A pesquisa (Apêndice B) ficou disponível do dia 3 de janeiro de 2022 ao dia 14 de janeiro de 2022. No período em que a pesquisa ficou ativa, 68 pessoas, em 6 estados brasileiros, na faixa etária de 20 a 74 anos, responderam ao questionário.

[2]https://www.google.com/forms/

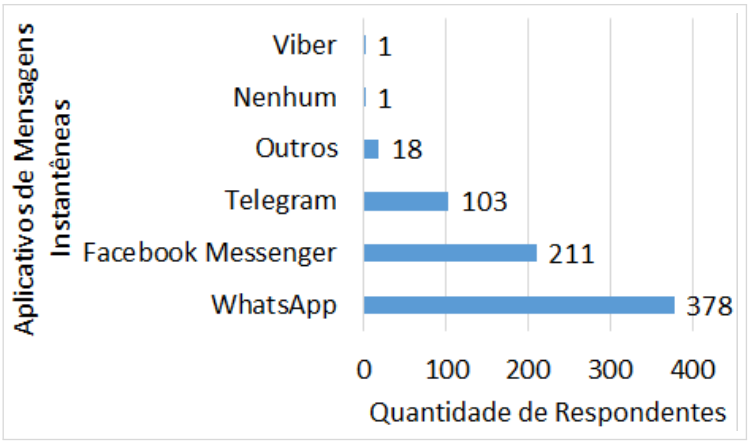


Figura 3.15: Respondentes que usam aplicativos de mensagens instantâneas, referentes à primeira pesquisa de opinião

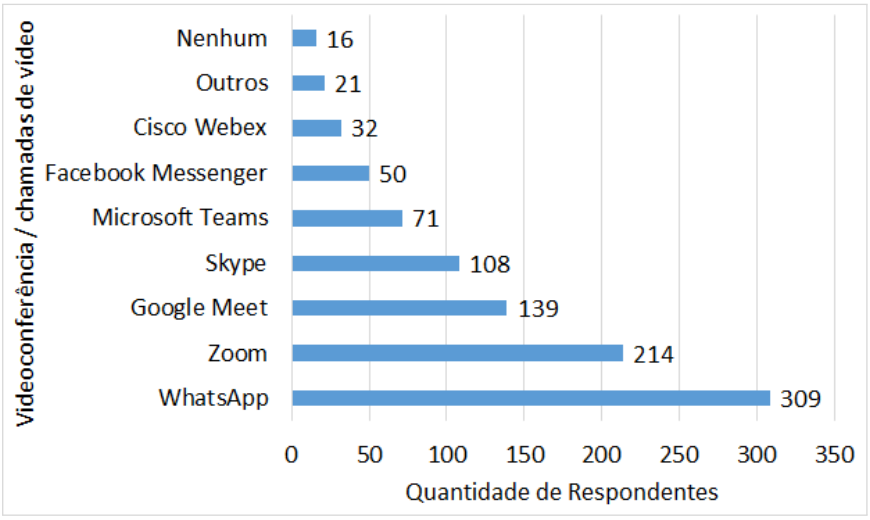


Figura 3.16: Respondentes que usam aplicativos de videoconferência/chamada de vídeo, referentes à primeira pesquisa de opinião

3.1.2.1 População e Amostra

Assim como na primeira pesquisa de opinião (Seção 3.1.1), não houve limitação do escopo populacional, exceto quanto ao país (Brasil). Também foi utilizada a amostra aleatória e os participantes da segunda pesquisa de opinião tinham entre 20 e 74 anos de idade, e estavam distribuídos conforme mostram as Figuras 3.17, 3.18, 3.19 e 3.20. Os respondentes tinham origem em 6 unidades da federação e 16 município, sendo majoritariamente da cidade de Niterói, no estado do Rio de Janeiro.

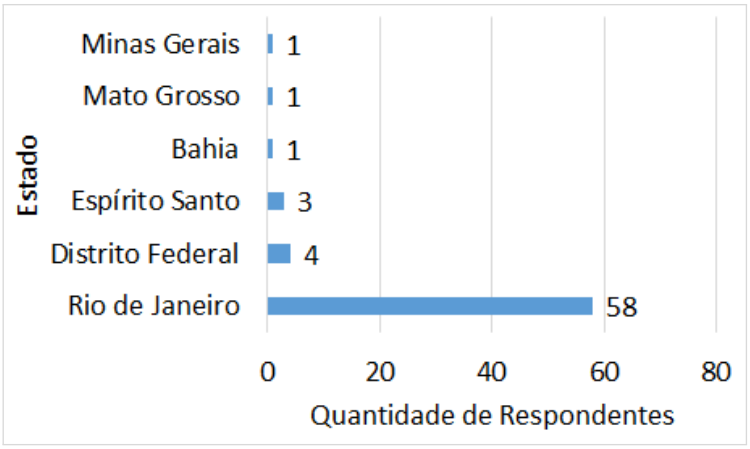


Figura 3.17: Respostas por estado, referentes à segunda pesquisa de opinião

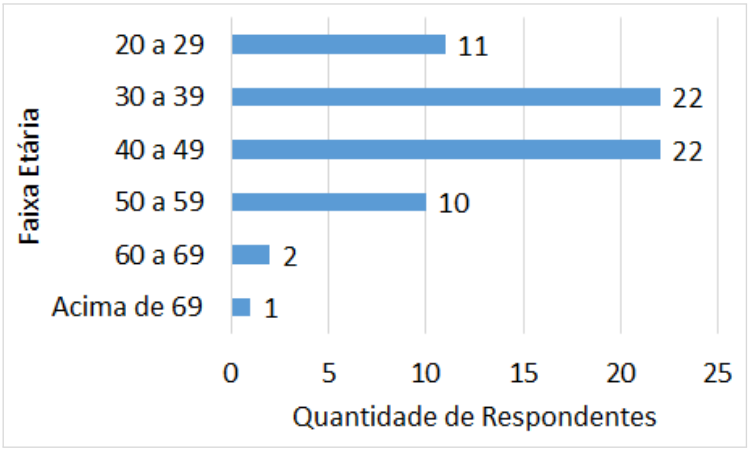


Figura 3.18: Respostas por faixa etária, referentes à segunda pesquisa de opinião

3.1.2.2 Análise de Dados

Relataram ter alterado o plano de Internet fixa contratado, 44% dos respondentes. Na totalidade dessas alterações houve incremento na largura de banda. Em média, a largura de banda saiu de 77 Mbps para 242 Mbps, representando um incremento médio de 215%, conforme demonstrado na Tabela 3.1. Dentre os respondentes que alteraram o plano de Internet, 37% relataram ter havido melhora na estabilidade da conexão de Internet (Fig. 3.21). Já entre os respondentes que não alteraram o plano de Internet, não houve relatos de melhora na estabilidade da conexão de Internet (Fig. 3.21). Assim, 83% dos respondentes que alteraram o plano de Internet consideram sua conexão com a Internet estável, contra 87% daqueles que não alteraram o plano de Internet.

Daqueles que consideram sua conexão com a Internet instável, observa-se que apenas 13% utiliza redes sem fio na faixa de 5 GHz (Fig. 3.22). Esse indicador sobe para 34% entre aqueles que consideram sua conexão com a Internet estável (Fig. 3.22).

Tabela 3.1: Alterações nos planos de Internet fixa, referentes à segunda pesquisa de opinião

Plano de Internet pré-pandemia	Plano de Internet pós-pandemia	Qtde.
300 Mb/s	400 Mb/s	1
200 Mb/s	500 Mb/s	1
150 Mb/s	500 Mb/s	1
100 Mb/s	450 Mb/s	1
100 Mb/s	350 Mb/s	1
100 Mb/s	300 Mb/s	3
100 Mb/s	200 Mb/s	5
99 Mb/s	300 Mb/s	1
99 Mb/s	200 Mb/s	1
90 Mb/s	300 Mb/s	1
70 Mb/s	70 Mb/s	1
60 Mb/s	500 Mb/s	1
50 Mb/s	200 Mb/s	2
25 Mb/s	150 Mb/s	1
20 Mb/s	35 Mb/s	1
15 Mb/s	309 Mb/s	1
15 Mb/s	200 Mb/s	1
15 Mb/s	60 Mb/s	1
15 Mb/s	50 Mb/s	1
10 Mb/s	200 Mb/s	1
6 Mb/s	15 Mb/s	1
5 Mb/s	200 Mb/s	1
3 Mb/s	4 Mb/s	1

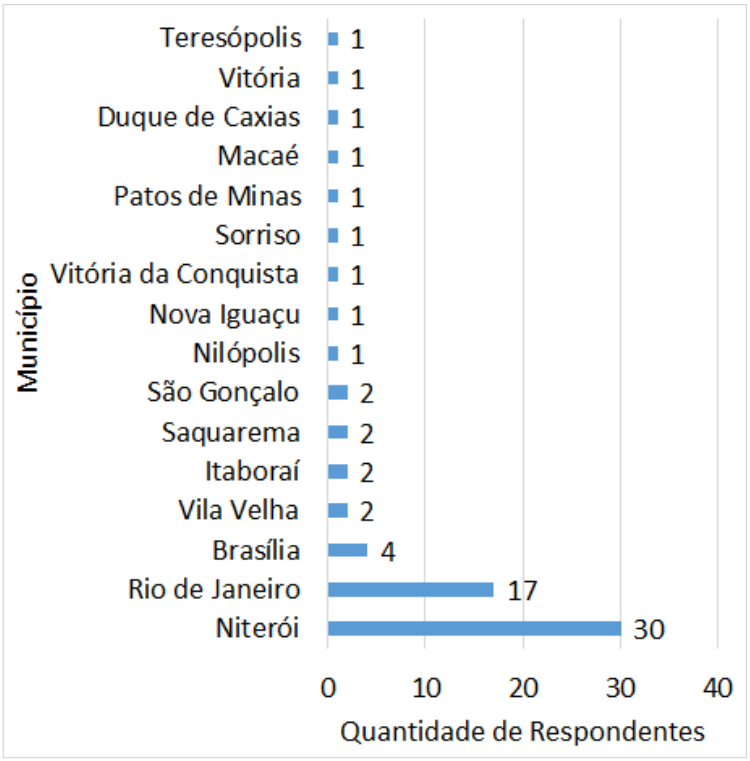


Figura 3.19: Respostas por município, referentes à segunda pesquisa de opinião

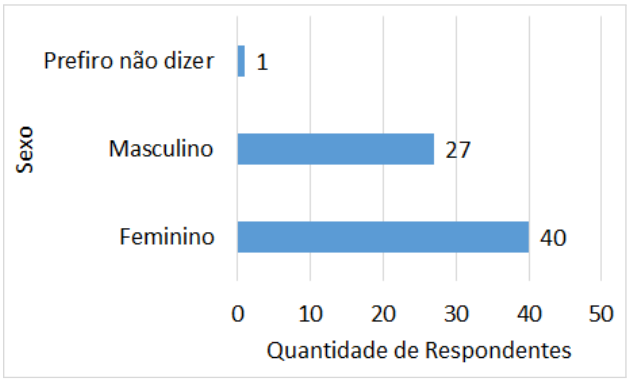


Figura 3.20: Respostas por sexo, referentes à segunda pesquisa de opinião

Apenas 25% dos respondentes relataram ter diminuído ou cessado o uso de ferramentas de videoconferência para trabalho/estudo e 50% relataram continuar usando com a mesma intensidade ou superior (Fig. 3.23). A redução não registrou números muito expressivos. No entanto, talvez esse índice seja o suficiente para ajudar a descongestionar as redes, em especial àquelas residenciais que operam na faixa de frequência de 2,4 GHz.

Dentre os respondentes, 74% disseram haver, na mesma residência, outras pessoas que fazem uso de ferramentas de videoconferência. Desses, 76% usam ferramentas de videoconferência para atividades laborais e educacionais e 68% para contato com amigos e/ou familiares (Fig. 3.24). É importante observar que 42% relataram diminuição de

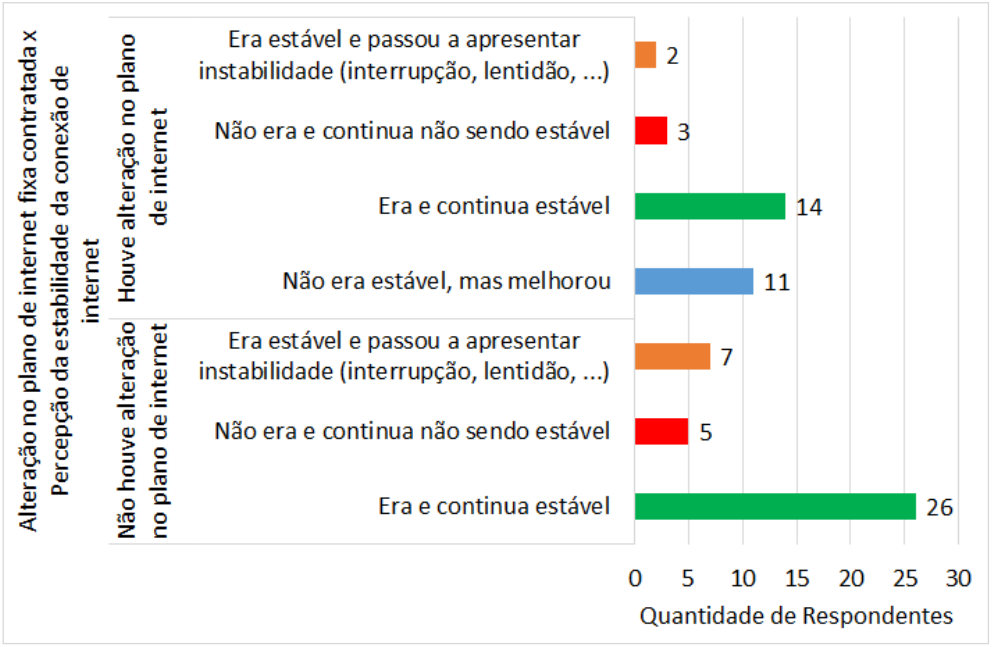


Figura 3.21: Correlação entre a alteração no plano de Internet fixa contratado e a percepção da estabilidade da conexão com a Internet, referente à segunda pesquisa de opinião

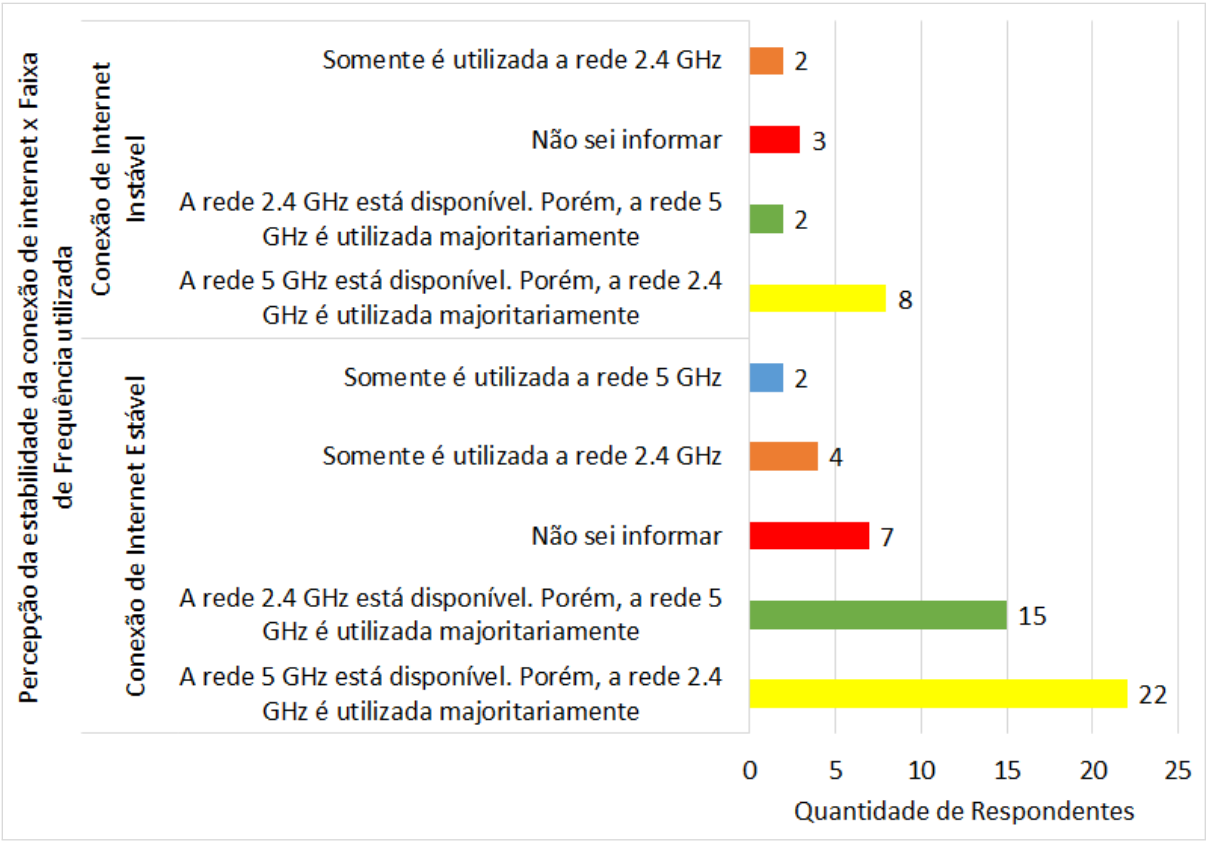


Figura 3.22: Correlação entre a percepção da estabilidade da conexão com a Internet e a faixa de frequência utilizada na rede sem fio doméstica, referente à segunda pesquisa de opinião

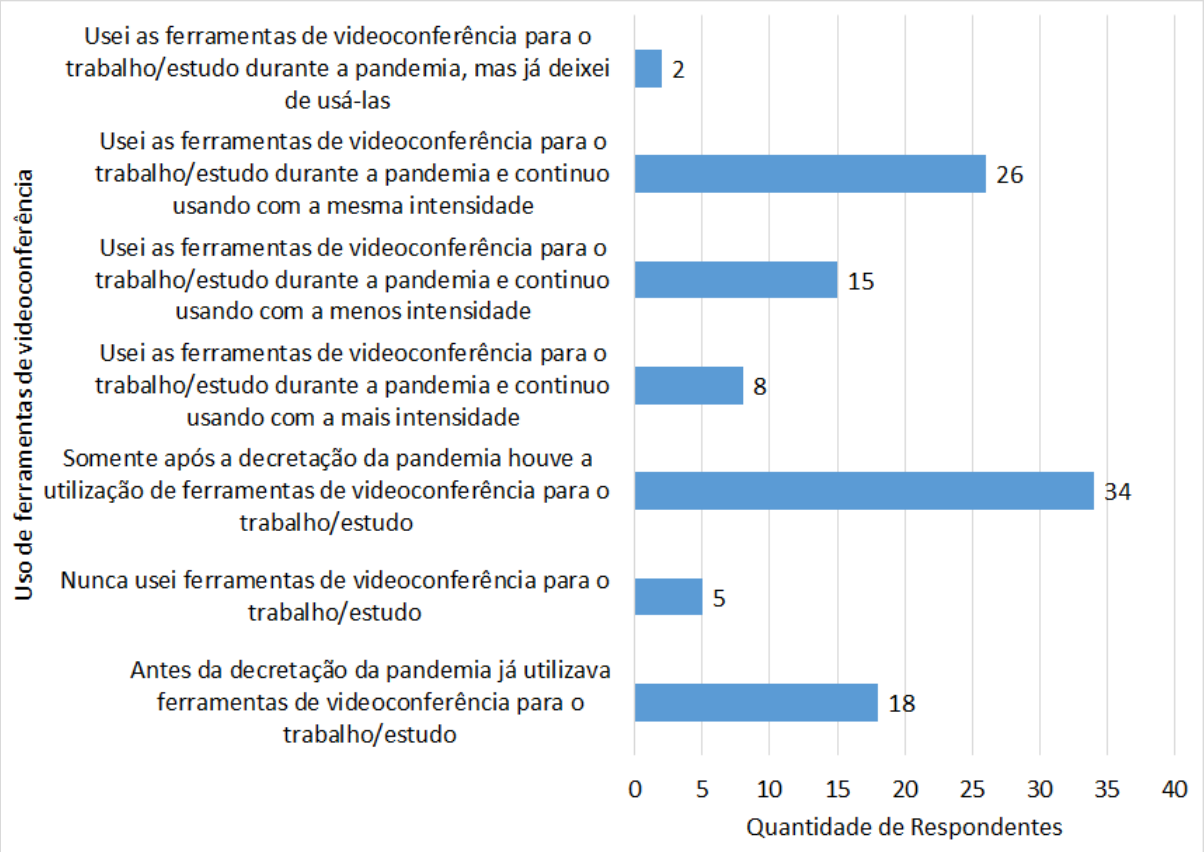


Figura 3.23: Utilização de ferramentas de videoconferência, referente à segunda pesquisa de opinião

utilização dessas ferramentas pelos demais moradores da mesma residência. Ao passo que 54% mantêm a utilização das ferramentas de videoconferência no mesmo nível ou superior, em comparação aos períodos mais restritivos da pandemia (Fig. 3.25).

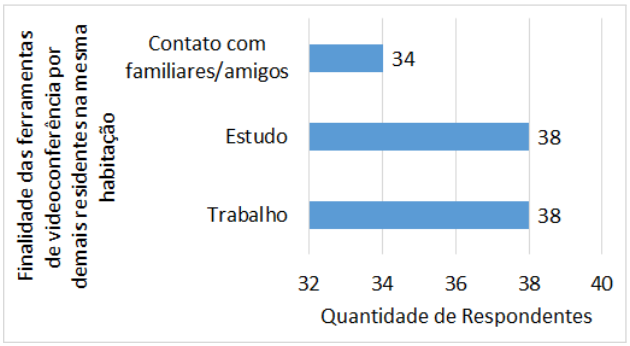


Figura 3.24: Atividades nas quais são utilizadas as ferramentas de videoconferência por demais residentes na mesma habitação, referente à segunda pesquisa de opinião

Diferente da primeira pesquisa de opinião, ainda durante um dos períodos mais restritivos da pandemia, 75% dos respondentes relataram uma percepção de estabilidade na conexão com a Internet (Fig. 3.26), contra 39% naquela ocasião (Fig. 3.9). É possível que essa melhora decorra da redução da utilização das ferramentas de videoconferência

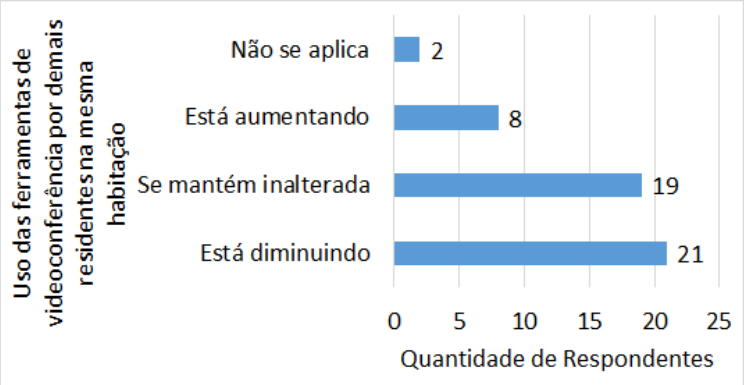


Figura 3.25: Utilização de ferramentas de videoconferência por demais residentes na mesma habitação, referente à segunda pesquisa de opinião

e a consequente redução no tráfego, melhoria no plano de Internet contratada, eventual aperfeiçoamento da infraestrutura de rede das operadoras, a fim de suportar o aumento da demanda durante a pandemia ou, ainda, que essa melhora seja apenas uma percepção do usuário, simplesmente porque passou a demandar menos os serviços na nuvem, decursivo da quase total volta à normalidade.

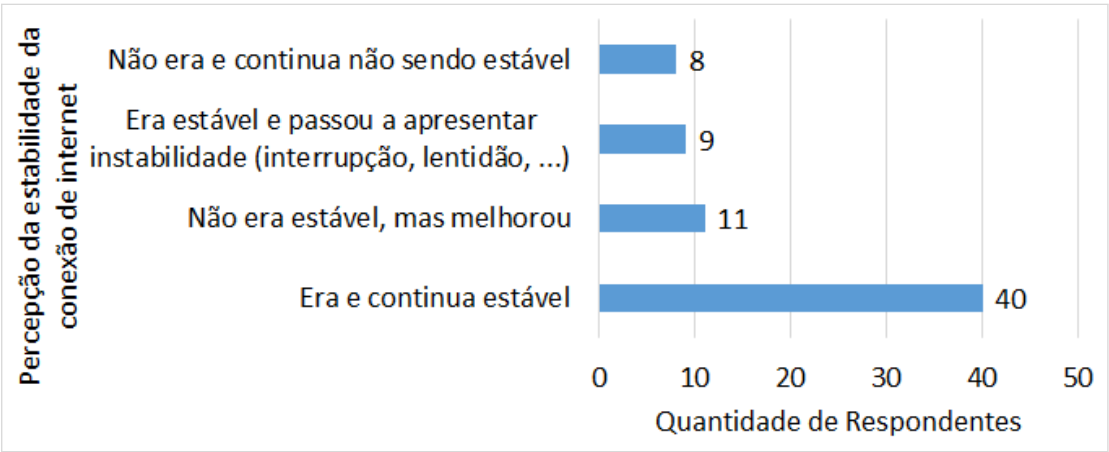


Figura 3.26: Percepção dos respondentes quanto à estabilidade da conexão com a Internet, comparada ao período mais restritivo da pandemia, referente à segunda pesquisa de opinião

Essa percepção de estabilidade na conexão com a Internet é desvelada mesmo em uma realidade em que mais da metade das pessoas afirmaram usar a faixa de frequência de 2,4 GHz. Essa frequência é usada por diversos tipos de dispositivos, como bluetooth, forno micro-ondas, babás eletrônicas, controles remotos de portão de garagem. telefones sem fio e outros. Além disso, as redes sem fio que utilizam a faixa de frequência de 2,4 GHz possuem um alcance maior. Por esses motivos, a faixa de frequência de 2,4 GHz é mais suscetível a problemas relacionados à interferência, se comparada à faixa de frequência de 5 GHz. Além da interferência, a largura de banda também tem a capacidade de impactar

o desempenho da rede e, por conseguinte, a percepção do usuário sobre sua estabilidade. A faixa de frequência de 2,4 GHz (entre 2.400 e 2.483,5 MHz) é muito mais estreita que faixa de frequência de 5 GHz (entre 5.150 e 5.350 MHz e entre 5.470 e 5.725 MHz) [2]. Dentre as pessoas que afirmaram usar a faixa de frequência de 2,4 GHz não estão incluídas aquelas que não souberam especificar a faixa de frequência usada em sua rede sem fios residencial. Dos respondentes que utilizam redes sem fio domésticas, ao menos 55% relataram usar, majoritariamente, a faixa de frequência de 2,4 GHz na rede sem fio residencial. Deve-se, ainda, considerar que 15% dos respondentes não souberam informar qual faixa de frequência usam na rede sem fio residencial. Dessa forma, o número de pessoas que utilizam a faixa de frequência de 2,4 GHz na rede sem fio residencial estaria entre 55% e 70%. Apenas 28% afirmaram usar a faixa de frequência de 5 GHz, que tem maior largura de banda (Fig. 3.27).

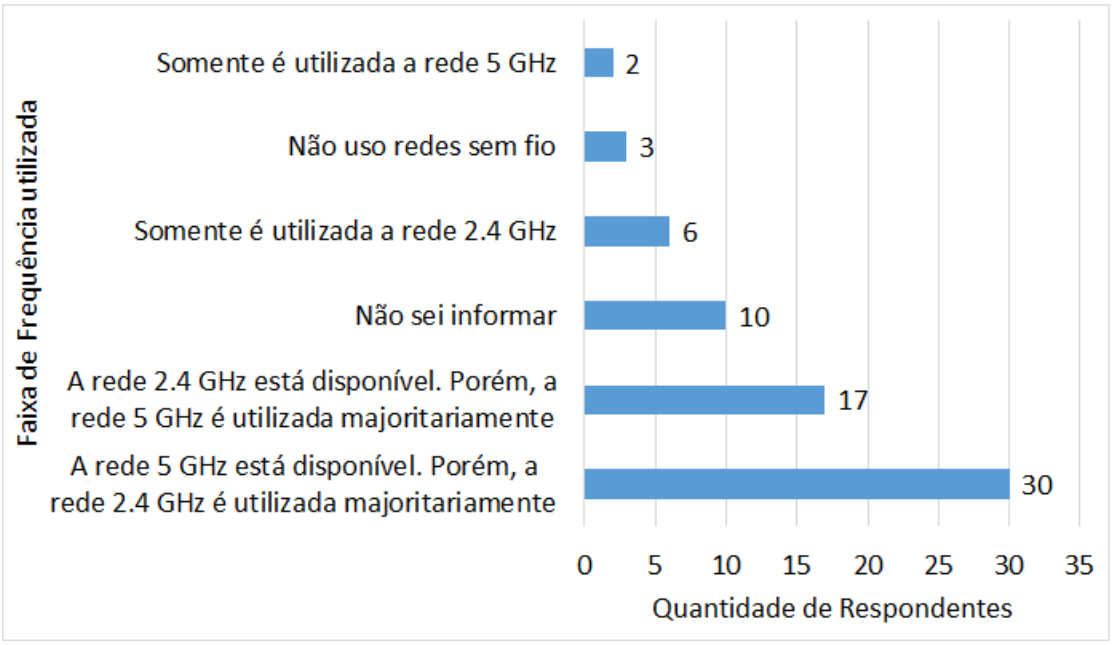


Figura 3.27: Perfil de rede residencial utilizada pelos respondentes, referente à segunda pesquisa de opinião

Em que pese todo o exposto, faz-se mister ressaltar que a maioria dos respondentes considera utilizar a Internet com muita intensidade para todos os propósitos pesquisados (Fig. 3.28).

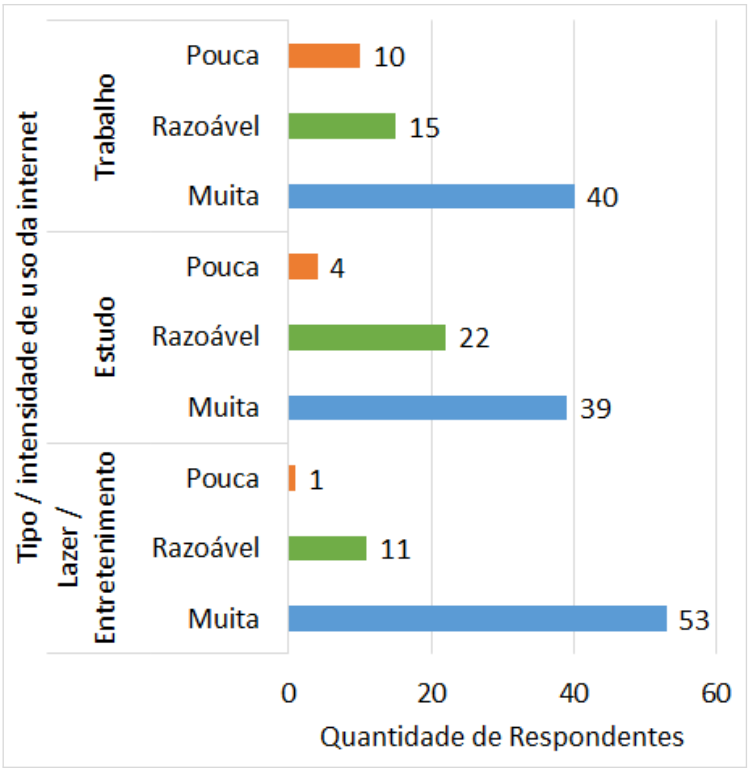


Figura 3.28: Correlação entre o tipo e a intensidade de uso da Internet, referente a segunda pesquisa de opinião

3.2 Impactos da Pandemia no Volume de Tráfego das Aplicações de Internet

Para melhor entender as ferramentas de conferência e videoconferência mais utilizadas, foi realizada uma análise de tráfego utilizando somente áudio e outra análise utilizando áudio e vídeo. A amostra coletada foi de, aproximadamente, dois minutos e as configurações dos participantes das chamadas, conferências ou videoconferências foram um subconjunto da topologia apresentada na Figura 3.29 (A e C; B e C; ou A, B e C).

O analisador de tráfego utilizado foi o *Wireshark*, instalado em notebook com sistema operacional Ubuntu Linux. A interface sem fio do notebook foi conectada ao roteador do ISP e a interface Ethernet conectada a um Ponto de Acesso (AP). A esse AP, conectou-se o dispositivo monitorado pelo analisador de tráfego.

3.2.1 Análise de Tráfego de Aplicações de Comunicação Multimídia Utilizando Somente Áudio

A análise de tráfego das ferramentas de comunicação, utilizando somente áudio, foi realizada com os dispositivos B e C, representados na Figura 3.29, para dois participantes, e

Figura 3.29: Topologia da rede de testes

utilizando os dispositivos A, B e C, ilustrados na mesma figura, para três participantes.

Consoante Figura 3.30 e Tabelas 3.2 e 3.3, observa-se que o incremento no consumo de banda na conferência com três participantes, tomando como base a chamada de voz com dois participantes, é mais expressivo no *Google Meet* e no *Facebook Messenger*, com um acréscimo de 106,13% e 64,41%, respectivamente. Já no *Microsoft Teams* houve um decremento de 8,68%. As demais ferramentas mantiveram um acréscimo entre 5% e 27%. O leve incremento no consumo de banda sugere que os servidores das respectivas aplicações mesclam os fluxos de áudio recebidos dos participantes [26]. Assim, cada participante recebe apenas um fluxo de áudio do servidor.

Tabela 3.2: Chamadas de voz com dois participantes

Aplicativo	**Chamada de Voz**						
	Tempo (s)	**Tráfego (MB)**	**Pacotes**	**Taxa de Transferência (Kbps)**	**Pacotes/s**	**Média do Tamanho dos Pacotes (B)**	**% Em Relação à Captura Completa**
Facebook Messenger	117,335	1,07	7.062	76,83	60	160	90,10
Google Meet	116,175	0,98	8.247	70,93	71	125	84,80
Microsoft Teams	119,285	2,01	10.875	141,48	91	194	88,00
Zoom	117,687	5,27	13.981	375,36	119	395	98,40
Webex	117,521	1,60	9.212	114,00	78	182	81,70
WhatsApp	120,124	0,70	41,52	49,08	35	177	89,80
Skype	117,853	2,03	12.143	114,36	103	175	93,90

A ferramenta que apresenta o maior consumo de banda é *Zoom*. Ela consome mais que o dobro que o *Skype*, segunda ferramenta que mais consome banda. A ferramenta que tem menor consumo de banda é o *WhatsApp*.

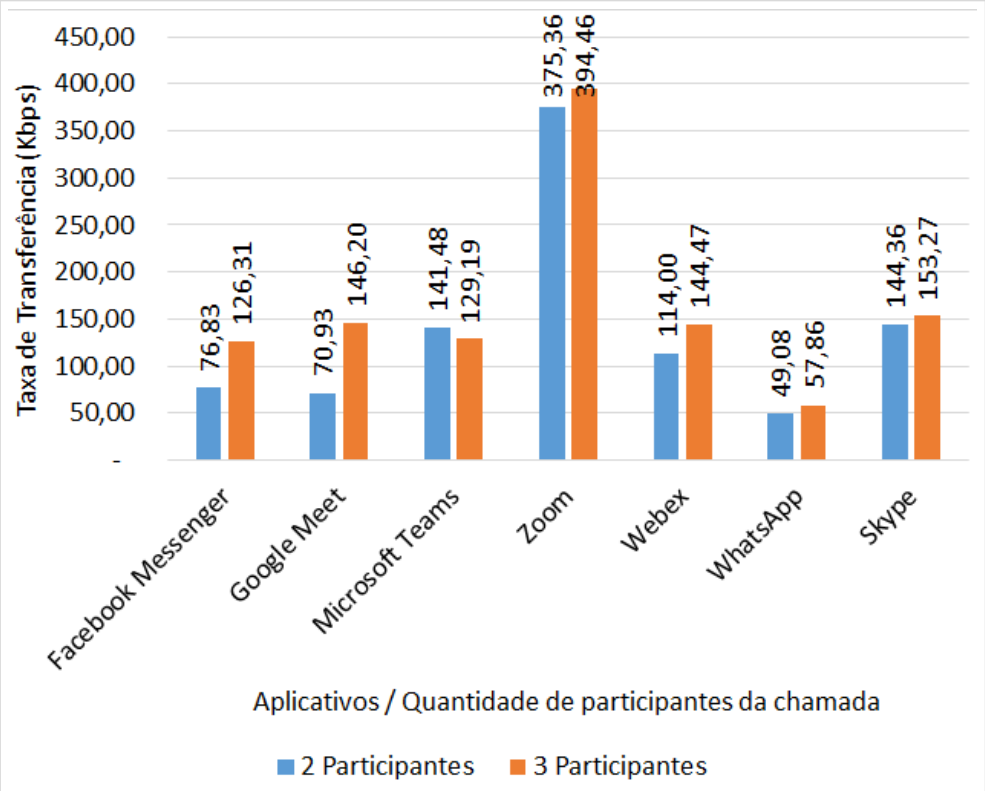


Figura 3.30: Gráficos comparativos de consumo de banda - Áudio

Tabela 3.3: Conferência com três participantes

Aplicativo	Conferência						
	Tempo (s)	Tráfego (MB)	Pacotes	Taxa de Transferência (Kbps)	Pacotes/s	Média do Tamanho dos Pacotes (B)	% Em Relação à Captura Completa
Facebook Messenger	170,733	1,67	10.079	126,31	91	173	90,90
Google Meet	114,920	2,00	14.665	146,20	128	143	92,00
Microsoft Teams	111,262	1,71	11.302	129,19	102	159	87,40
Zoom	118,952	5,59	15.711	394,46	132	367	98,40
Webex	113,466	1,95	11.290	144,47	100	181	58,10
WhatsApp	111,657	0,77	4.335	57,86	39	186	97,00
Skype	121,819	2,23	11.326	153,27	93	206	86,10

3.2.2 Análise de Tráfego de Aplicações de Comunicação Multimídia Utilizando Áudio e Vídeo

A análise de tráfego das ferramentas de comunicação, utilizando áudio e vídeo, foi realizada com os dispositivos B e C, representados na Figura 3.29, para dois participantes, e utilizando os dispositivos A, B e C, ilustrados na mesma figura, para três participantes.

Conforme ilustrado na Figura 3.31 e detalhado nas Tabelas 3.4 e 3.5, o *Facebook Messenger* se fixa com o maior incremento de consumo de banda, nessa modalidade, com 41,8%. As demais ferramentas mantiveram um acréscimo entre 18% e 33%. A ferramenta que tem o maior consumo de banda é *Google Meet* e a que tem o menor consumo de

banda é o *WhatsApp*, consumindo, com três participantes, menos da metade da banda consumida pelo *Skype*, segunda ferramenta de mais baixo consumo de banda.

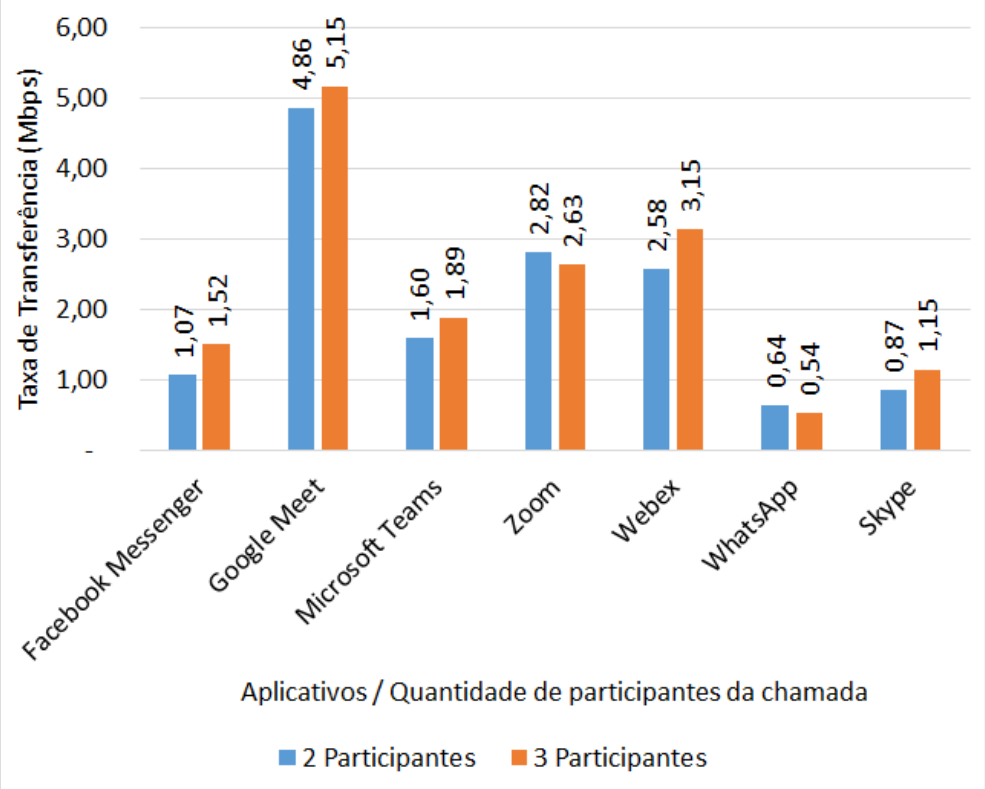


Figura 3.31: Gráficos comparativos de consumo de banda - Áudio e Vídeo

Tabela 3.4: Chamadas de vídeo com dois participantes

Aplicativo	**Chamada de Vídeo**						
	Tempo (s)	**Tráfego (MB)**	**Pacotes**	**Taxa de Transferência (Mbps)**	**Pacotes/s**	**Média do Tamanho dos Pacotes (B)**	**% Em Relação à Captura Completa**
Facebook Messenger	117,794	15,07	23.732	1,07	201	666	60,50
Google Meet	116,176	67,24	72.001	4,86	620	979	98,80
Microsoft Teams	112,251	21,41	35.802	1,60	319	127	99,00
Zoom	116,763	39,23	43.844	2,82	375	938	99,80
Webex	118,194	36,30	44.343	2,58	375	858	98,30
WhatsApp	119,333	9,05	11.954	0,64	100	794	99,20
Skype	121,055	12,55	27.878	0,87	230	472	99,80

Comparando os dados constantes das Figuras 3.30 e 3.31, verifica-se um comportamento distinto das ferramentas. Da chamada de vídeo, com dois participantes, para a videoconferência, com três participantes, o *Google Meet*, que registrou o maior aumento utilizando somente áudio, nesta modalidade é o que registra o menor aumento no consumo de banda, desconsiderando o *WhatsApp* e o *Zoom*, que registraram decréscimo de 15,8% e 6,64%, respectivamente. Salienta-se que, sob certas condições, as aplicações podem manipular parâmetros de codificação, como a quantidade de quadros por segundo, para reduzir a utilização de banda.

Tabela 3.5: Videoconferência com três participantes

Aplicativo	Videoconferência						
	Tempo (s)	Tráfego (MB)	Pacotes	Taxa de Transferência (Mbps)	Pacotes/s	Média do Tamanho dos Pacotes (B)	% Em Relação à Captura Completa
Facebook Messenger	101,769	18,46	26.938	1,52	265	718	99,00
Google Meet	115,866	71,19	78.914	5,15	681	946	99,60
Microsoft Teams	111,180	25,08	42.215	1,89	380	623	99,10
Zoom	119,248	37,40	47.059	2,63	395	833	99,80
Webex	109,908	41,21	51.643	3,15	470	837	98,50
WhatsApp	110,843	7,08	12.049	0,54	109	616	98,60
Skype	109,482	15,07	30.092	1,15	275	525	96,20

3.2.3 *Skype* e *WhatsApp*

Foi observada, nas ferramentas *Skype* e *WhatsApp*, durante a análise de tráfego, a capacidade de intercambiar do modo cliente-servidor para o modo par-a-par. Para chamadas com mais de dois participantes, as ferramentas utilizaram o modo cliente-servidor. O modo P2P está disponível apenas para chamadas com dois participantes.

Observou-se também que ambas as ferramentas conseguem detectar se a conexão P2P é local ou trafega pela Internet. Ao concluir que a conexão P2P é local, as aplicações transmitem os fluxos de dados diretamente ao outro dispositivo utilizando o IP privado da rede local. O consumo de banda em cada um desses tipos de conexão é significativamente diferente, como pode-se observar na Tabela 3.6 e na Figura 3.32. A taxa de transferência do *Skype* na rede local é 152% maior do que na Internet e a taxa de transferência do *WhatsApp* é 232% maior na rede local, se comparado ao da Internet.

Tabela 3.6: Comparação do tráfego, Internet e rede local, em chamadas de vídeo com dois participantes no *WhatsApp* e *Skype*

Aplicativo	Chamada de Vídeo						
	Tempo (s)	Tráfego (MB)	Pacotes	Taxa de Transferência (Mbps)	Pacotes/s	Média do Tamanho dos Pacotes (B)	% Em Relação à Captura Completa
WhatsApp Internet	119,333	9,05	11.954	0,64	100	794	99,20
WhatsApp Rede Local	119,231	21,02	24.884	1,48	209	886	97,00
Skype Internet	121,055	12,55	27.878	0,87	230	472	98,80
Skype Rede Local	120,024	18,94	32.728	1,32	273	607	99,20

3.2.4 *Round Trip Time* (RTT) das Aplicações de Comunicação Multimídia

Para realizar a medição do *Round Trip Time* (RTT), foi calculado o tempo para o recebimento do pacote *SYN-ACK* após a transmissão do pacote *SYN*, durante o *Three-way*

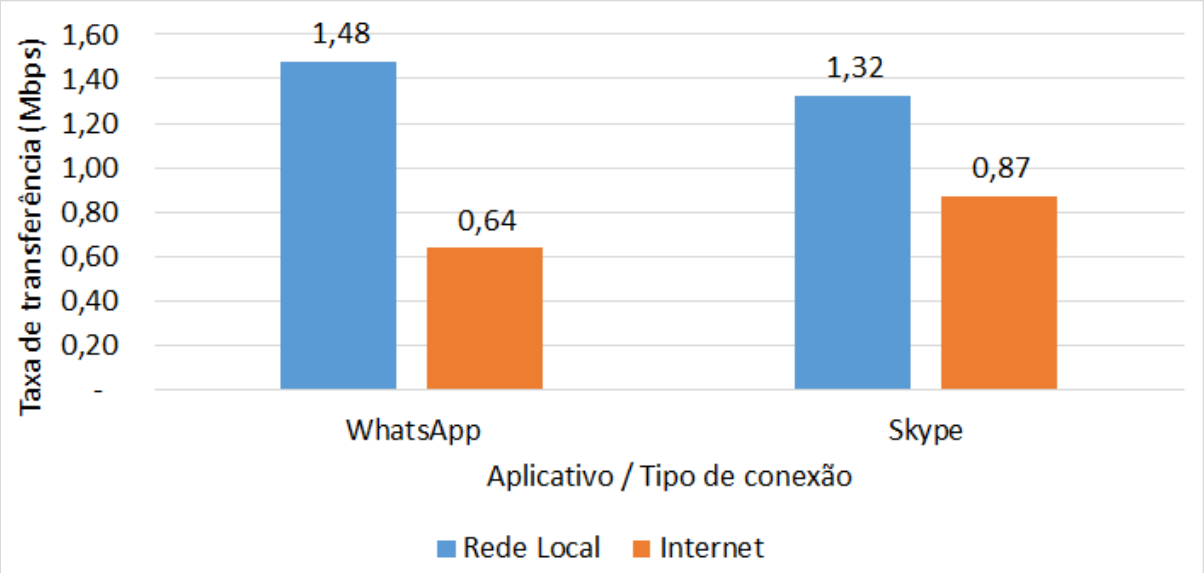


Figura 3.32: Gráfico comparativo do tráfego, Internet e rede local, em chamadas de vídeo com dois participantes no *WhatsApp* e *Skype*

Handshake da conexão TCP (*Transmission Control Protocol*), que é estabelecida pelas aplicações antes da transmissão dos fluxos de dados UDP (*User Datagram Protocol*). Foi calculada a média entre o valor obtido através do método descrito e do resultado de medições utilizando o *Internet Control Message Protocol* (ICMP), por intermédio do aplicativo *ping*.

Assim como descrito na Tabela 3.7, as ferramentas de comunicação analisadas apresentaram RTT entre 13 ms e 26 ms, com exceção do *Cisco Webex*, que apresentou RTT de 182 ms. Considerando apenas o atraso de propagação, esses servidores estariam a uma distância máxima da cidade de Niterói-RJ, onde os experimentos foram realizados, entre 1.950 km e 3.900 km. Por outro lado, ferramentas de geolocalização sugerem que os IPs desses servidores estejam localizados a distâncias muito maiores. Isso pode indicar que essas aplicações utilizam endereços IP *anycast* [40], ou seja, que implementam mecanismos para que os clientes possam selecionar o "melhor" servidor em um grupo de servidores. Já a distância máxima do servidor do *Cisco Webex* seria de 27.300 km. A recomendação de limite de tempo aceitável para uma transmissão VoIP unilateral é de 150 ms [23]. Destarte, o *Cisco Webex* é mais suscetível a problemas de interatividade para a localidade geográfica onde foram realizados os testes.

Embora não tenha sido possível determinar a quantidade exata de saltos para todos os servidores, encontrou-se um número mínimo de seis saltos.

Tabela 3.7: Tabela de Localizações, Saltos e Tempos de Resposta

Aplicativo	Localização **	Hops	Avg RTT
Facebook Messenger	Irlanda	8	13 ms
Google Meet	Estados Unidos	17	26 ms
Microsoft Teams	Estados Unidos	> 16	* 15 ms
Zoom	Estados Unidos	> 5	20 ms
Webex	Estados Unidos	12	182 ms
WhatsApp	Irlanda	8	18 ms
Skype	Estados Unidos	> 15	24 ms
* 52.113.194.132, IP que estabelece conexão TCP, calculado no *Wireshark*			
** Fonte: `https://www.ip-adress.com`			

3.2.5 Área de Trabalho Remota vs. *Virtual Private Network*

O desempenho dos protocolos de área de trabalho remota foram estudados em outros trabalhos [52] [61] [8] [36], assim como a VPN [59] [79] [37] [44]. Neste estudo, foi realizado um comparativo prático na utilização das duas soluções.

A topologia utilizada nos experimentos está representada na Figura 3.33. Os servidores de Área de Trabalho Remota, Banco de Dados, Domínio e Arquivos têm o *Microsoft Windows Server* 2019 como sistema operacional. O servidor de VPN (*Virtual Private Network*) utiliza o protocolo L2TP (*Layer 2 Tunnelling Protocol*) com IPSec (*IP Security Protocol*). O sistema operacional do dispositivo cliente é o *Microsoft Windows* 10. O ambiente corporativo e o dispositivo cliente estavam a 30 quilômetros de distância.

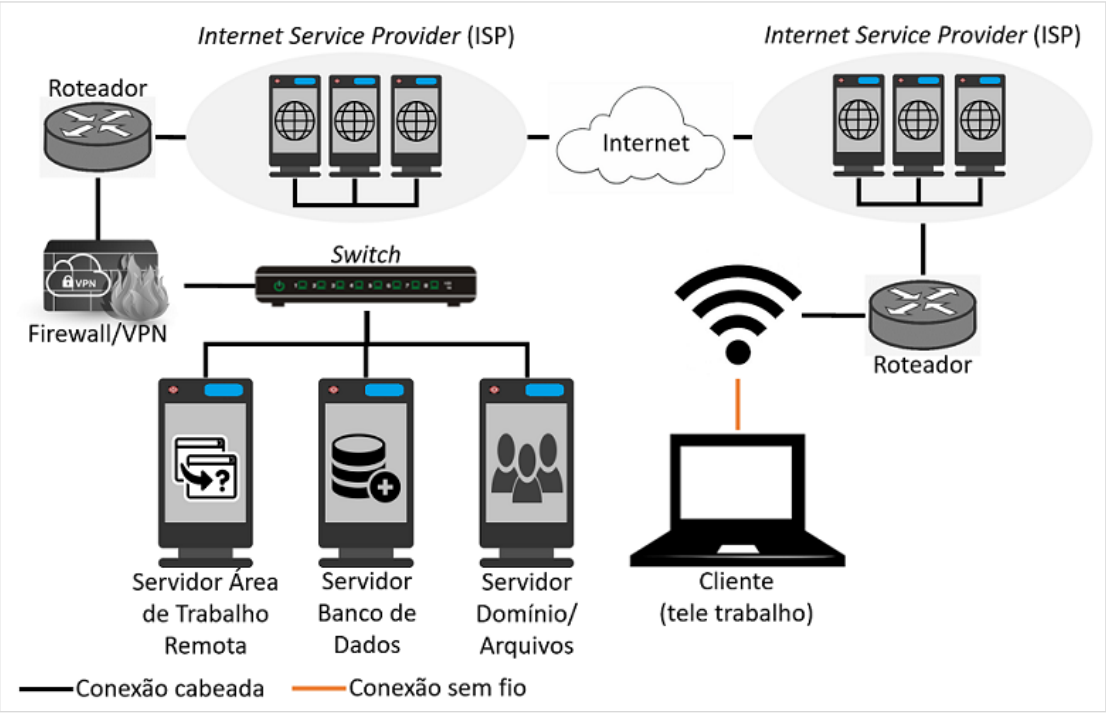


Figura 3.33: Topologia do Experimento da Área de Trabalho Remota / VPN

O primeiro teste consistiu em executar um *software Desktop*, fazer o *login* no sistema e

fazer uma pesquisa na base de dados do sistema, retornando 3 registros como resposta. A execução do sistema se deu de três formas: acesso pelo dispositivo cliente, através da área de trabalho remota, ao servidor hospedado fisicamente no ambiente corporativo (Tabela 3.8); através da execução do sistema, instalado no dispositivo cliente, com acesso à base de dados no servidor corporativo, por meio da VPN (Tabela 3.9); e executando o sistema no dispositivo cliente, cujo executável se encontra no servidor de arquivos corporativo e tem tamanho de 10,2 Megabytes, mantendo o acesso à base de dados, também, hospedada em um servidor corporativo, por meio da VPN (Tabela 3.10).

O segundo teste baseou-se na abertura de um arquivo de apresentação de slides, criado com o aplicativo *Microsoft PowerPoint*, com tamanho total de 14,2 Megabytes. O arquivo encontrava-se hospedado no servidor de arquivos corporativo e foi acessado, em primeiro momento, utilizando a área de trabalho remota (Tabela 3.11) e, em seguida, diretamente no dispositivo cliente por meio da VPN (Tabela 3.12). Em ambas as formas de acesso, o arquivo foi aberto e executado o modo de apresentação de slides, sendo realizadas as transições até o último slide.

O terceiro teste representou o acesso ao *PHPMyAdmin*, aplicação *web*. Após o *login*, foram exibidos os 122 registros de uma das tabelas do banco de dados. O detalhamento do acesso através da área de trabalho remota está descrito na Tabela 3.13, enquanto o detalhamento do acesso através da VPN está descrito na Tabela 3.14.

Os três testes foram repetidos três vezes e foi calculada a média aritmética dos resultados plotados nas tabelas. O tempo e o volume do tráfego foram capturados desde o estabelecimento da conexão até a desconexão da área de trabalho remota ou da VPN.

Tabela 3.8: Execução de Sistema *Desktop* pela Área de Trabalho Remota

Rodada	**Tempo (s)**	**Tráfego (MB)**	**Pacotes**	**Taxa de Transferência (Mbps)**	**Pacotes/s**	**Média do Tamanho dos Pacotes (B)**
1	18,16	0,55	1.824	0,25	100	317
2	19,63	0,55	1.795	0,23	91	319
3	21,40	0,57	1.869	0,22	87	319
Média	**19,73**	**0,56**	**1.829**	**0,24**	**93**	**318**

Tabela 3.9: Execução de Sistema *Desktop* pela VPN (Executável no Cliente)

Rodada	**Tempo (s)**	**Tráfego (MB)**	**Pacotes**	**Taxa de Transferência (Mbps)**	**Pacotes/s**	**Média do Tamanho dos Pacotes (B)**
1	42,54	0,93	2.506	0,18	59	391
2	39,90	1,00	2.506	0,21	63	401
3	64,78	0,94	2.563	0,12	40	386
Média	**49,07**	**0,96**	**2.525**	**0,17**	**54**	**392**

Tabela 3.10: Execução de Sistema *Desktop* pela VPN (Executável no Servidor de Arquivos)

Rodada	Tempo (s)	Tráfego (MB)	Pacotes	Taxa de Transferência (Mbps)	Pacotes/s	Média do Tamanho dos Pacotes (B)
1	85,80	37,30	46.896	3,65	547	834
2	125,51	33,09	39.734	2,21	317	873
3	93,02	34,05	41.469	3,07	446	861
Média	**101,44**	**34,81**	**42.700**	**2,98**	**436**	**856**

Tabela 3.11: Abertura de Apresentação PowerPoint pela Área de Trabalho Remota

Rodada	Tempo (s)	Tráfego (MB)	Pacotes	Taxa de Transferência (Mbps)	Pacotes/s	Média do Tamanho dos Pacotes (B)
1	57,21	7,17	10.453	1,05	183	719
2	57,84	6,72	10.139	0,97	175	695
3	55,76	5,88	9.260	0,88	166	665
Média	**56,94**	**6,59**	**9.951**	**0,97**	**175**	**693**

Tabela 3.12: Abertura de Apresentação PowerPoint pela VPN

Rodada	Tempo (s)	Tráfego (MB)	Pacotes	Taxa de Transferência (Mbps)	Pacotes/s	Média do Tamanho dos Pacotes (B)
1	185,48	84,47	95.526	3,82	515	927
2	181,59	83,47	91.462	3,86	504	957
3	156,90	83,74	93.577	4,48	596	938
Média	**174,66**	**83,89**	**93.522**	**4,05**	**538**	**940**

Tabela 3.13: Acesso à Sistema Web pela Área de Trabalho Remota

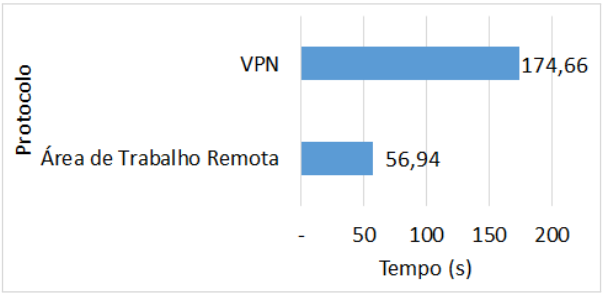
Rodada	Tempo (s)	Tráfego (MB)	Pacotes	Taxa de Transferência (Mbps)	Pacotes/s	Média do Tamanho dos Pacotes (B)
1	28,01	1,54	3.329	0,46	119	484
2	20,34	1,63	3.341	0,67	164	510
3	21,60	1,43	3.119	0,56	144	481
Média	**23,32**	**1,53**	**3.263**	**0,56**	**143**	**491**

Tabela 3.14: Acesso à Sistema Web pela VPN

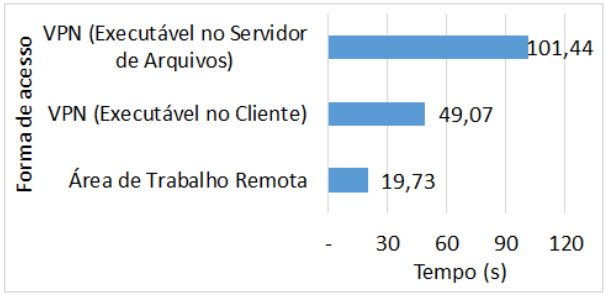
Rodada	Tempo (s)	Tráfego (MB)	Pacotes	Taxa de Transferência (Mbps)	Pacotes/s	Média do Tamanho dos Pacotes (B)
1	29,10	1,32	1.949	0,38	67	710
2	24,12	1,04	1.522	0,36	63	714
3	24,49	1,17	1.833	0,40	75	671
Média	**25,90**	**1,18**	**1.768**	**0,38**	**68**	**698**

Observando os gráficos do tempo em que a captura foi executada (Figuras 3.34a, 3.34b e 3.34c), que correspondem ao tempo que foi necessário para cumprir as tarefas, observa-se que utilizando a VPN há um maior dispêndio de tempo em relação à área de trabalho remota. É possível verificar que o aplicativo *Microsoft PowerPoint* é muito ineficiente para abrir um documento remoto, sendo necessário, em média, 207% a mais de

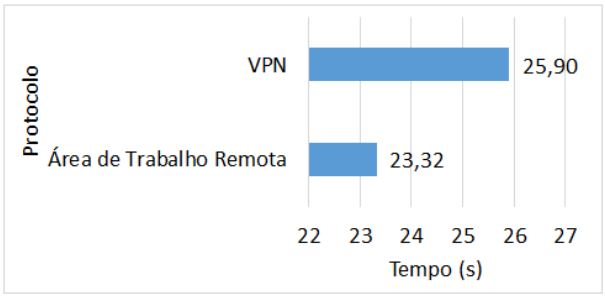
tempo para abrir o documento usando a VPN, em relação à área de trabalho remota. Em relação ao volume de dados trafegados, registra-se um acréscimo de 1.173% no consumo de banda da VPN, comparada à área de trabalho remota (Figura 3.35a).



(a) Tempo de Abertura de Apresentação *PowerPoint*



(b) Tempo de Execução de Sistema *Desktop*



(c) Tempo de Acesso de Sistema *Web*

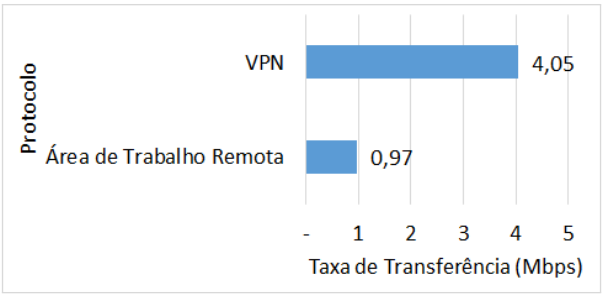
Figura 3.34: Comparativo de tempo demandado

Quando se trata da execução de sistema *Desktop*, utilizando a VPN com o sistema instalado no dispositivo cliente, o tempo gasto é 149% superior à área de trabalho remota e o consumo de banda é superior em 73% (Figura 3.35b). Se o executável do aplicativo estiver no servidor de arquivos remoto, o consumo de tempo é ainda maior, 414% a mais que a área de trabalho remota. O consumo de banda, nesse último caso, atinge a impressionante marca de 6.172% de acréscimo (Figura 3.35b).

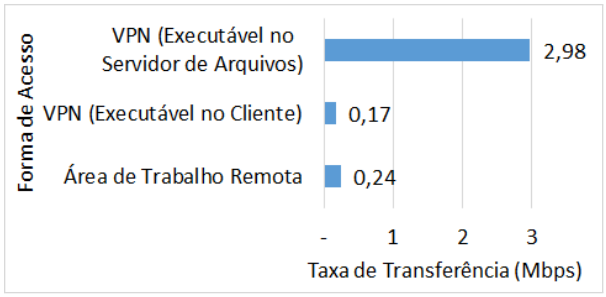
O teste em que a VPN demonstrou melhor desempenho foi no acesso à sistemas *Web*. Nessa modalidade, o volume trafegado na VPN foi 23% inferior ao registrado na área de trabalho remota (Figura 3.35c).

3.3 Impactos nas Conexões das Redes de Acesso

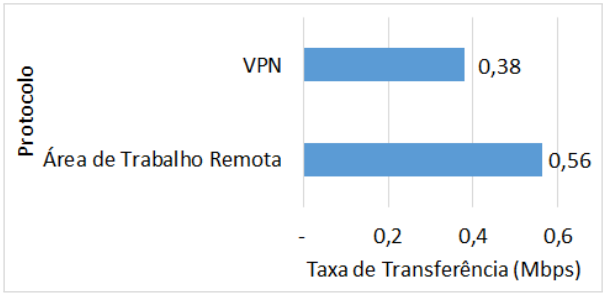
Esta seção apresenta as análises de captura de pacotes de redes sem fio e dados de conexões de telefonia celular, cotejando-os com as pesquisas de opinião e com os dados de mobilidade do *Google*, a fim de validá-los ou descartá-los. Ademais, faz-se uma análise do tráfego da Rede Nacional de Ensino e Pesquisa (RNP).



(a) Taxa de Transferência de Apresentação *PowerPoint*



(b) Taxa de Transferência de Sistema *Desktop*



(c) Taxa de Transferência de Sistema *Web*

Figura 3.35: Comparativo de taxa de transferência

No que diz respeito aos dados sobre mobilidade do *Google*, uma questão importante a ser considerada é que a *baseline* calculada pelo *Google* considera que um dia de referência representa um valor normal para esse mesmo dia da semana. O dia de referência é o valor mediano do período de cinco semanas (3 de janeiro a 6 de fevereiro de 2020). Para cada categoria de região, o valor base não é único: são sete valores individuais. O mesmo número de visitantes em dois dias diferentes da semana resulta em diferentes mudanças percentuais. Dessa forma, os dados de mobilidade representam mudanças relativas, não o número absoluto de pessoas em cada categoria de local de concentração. É importante ressaltar que os dados não incluem informações de identificação pessoal, como localização, contatos e deslocamento de indivíduos. As informações são derivadas de conjuntos de dados agregados e anônimos de usuários que optaram por ativar a função de Histórico de Localização.

O período usado pelo *Google* como referência é, normalmente, de férias escolares, nos quais muitos adultos costumam, também, gozar suas férias. Isso pode, de certa maneira, gerar algum tipo de distorção nos dados. Além disso, com o passar do tempo, pode haver uma variação populacional. Por exemplo, uma cidade pode ter uma quantidade maior ou menor de residentes ou de empresas. Isso pode impactar os dados, em especial se a análise abranger períodos mais longos. É, também, relevante o fato de que parte expressiva dos estudantes, a saber crianças na faixa dos 11 anos ou menores, não costuma levar o

telefone para a escola. Dessa forma, eles não interferem nas estatísticas de concentração residencial. No entanto, eles compõem um dos grupos que mais utilizaram as ferramentas de videoconferência durante a pandemia. Além disso, o grupo dessa faixa etária costuma utilizar largamente serviços de *streaming* de vídeo e jogos online.

3.3.1 Análise de Dados do *Backbone* da RNP

A Rede Nacional de Ensino e Pesquisa (RNP) é uma rede de abrangência nacional voltada para educação superior, pesquisa e inovação. Além de se conectar com 15 países da América Latina, com a rede europeia Géant e com os Estados Unidos, a RNP possui 27 Pontos de Presença (PoPs) no Brasil, um em cada estado da federação. Suas 50 redes comunitárias, mais de 4 milhões de usuários, 1.100 pontos e 800 organizações conectadas demonstram a relevância da RNP para nosso estudo.

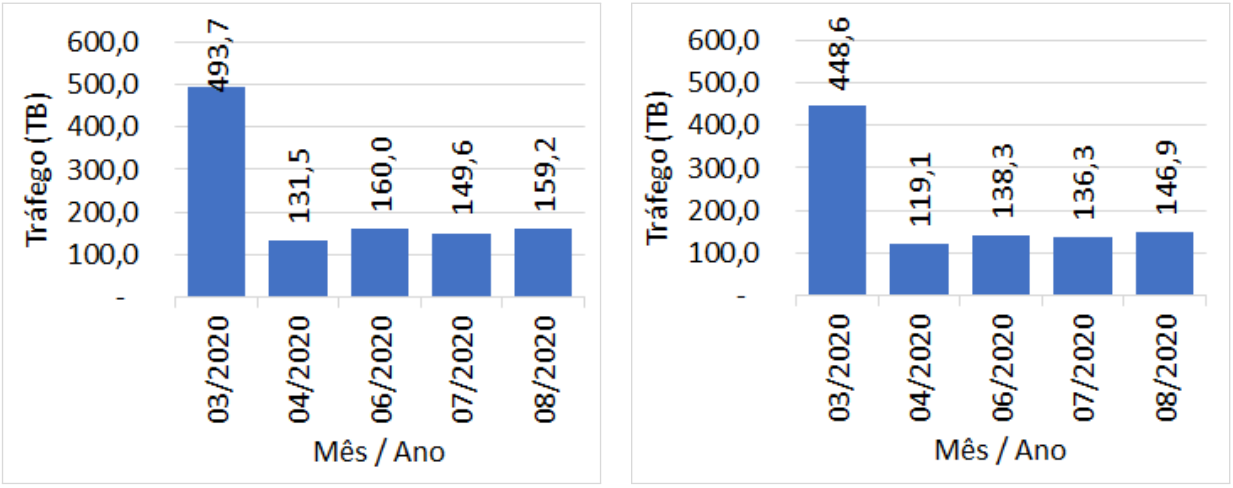
Extraiu-se, dessa rede, no formato CSV (*Comma-separated values*), registros *NetFlow* com volume de aproximadamente 7 PB. Os dados foram filtrados em duas fases, uma por protocolo e outra por protocolo/portas[3]. Após isso, os dados receberam tratamento para identificação do PoP e foram cadastrados em um banco de dados. Os dados foram compilados por uma ferramenta de *Business Intelligence* (BI).

Foram analisados os dados dos meses de março, abril, junho, julho e agosto de 2020 referentes a 15 estados brasileiros (AC, AL, AM, AP, ES, MA, MS, MT, PB, PI, RN, RO, RR, SE e TO). Os demais estados não fizeram parte da análise uma vez que não havia dados disponíveis desde março de 2020, o que inviabilizaria o comparativo do tráfego pré e pós decretação da pandemia.

Assim como a Internet, dados indicam que a Rede Nacional de Ensino e Pesquisa (RNP) também registrou uma mudança significativa no perfil do tráfego. No entanto, enquanto o tráfego da Internet foi majorado, o tráfego da RNP reduziu. Pode-se observar na Figura 3.36a que o tráfego total do protocolo TCP foi reduzido em 73% do mês de março para abril de 2020. Nos meses de abril, junho, julho e agosto, o volume de tráfego total se manteve estável. No que tange às portas 80 e 443 (Fig. 3.36b), principais portas utilizadas sobre o TCP, essas representam, em média, 90% do tráfego total. Uma possível justificativa para a redução do tráfego constatada é a redução e, em alguns casos, a suspensão das atividades presenciais nas Universidades. Há, também, uma hipótese de que tenha ocorrido um *offloading* de conteúdo didático para fora do ambiente acadêmico.

[3]TCP/UDP (40003, 19305, 3478, 3479, 3480, 3481, 9000, 8801, 80 e 443)

Nesse caso específico, o conteúdo que antes era predominantemente acessado no ambiente acadêmico passou a ser obtido em outros locais, como o *Google Classroom*, por exemplo. Porém, existe uma dificuldade em confirmar essa hipótese. Isso, em virtude de dois fatos: (i) devido ao *offloading*, em teoria, esse material não trafegaria mais na rede acadêmica; e (ii) mesmo que trafegasse na rede acadêmica seria difícil avaliar se e quais materiais estão sendo carregados nas plataformas, pois a maioria delas utiliza a conexão na porta 443. Isso significa que o tráfego dessas plataformas é criptografado e não é possível obter informações detalhadas apenas observando as conexões de rede.

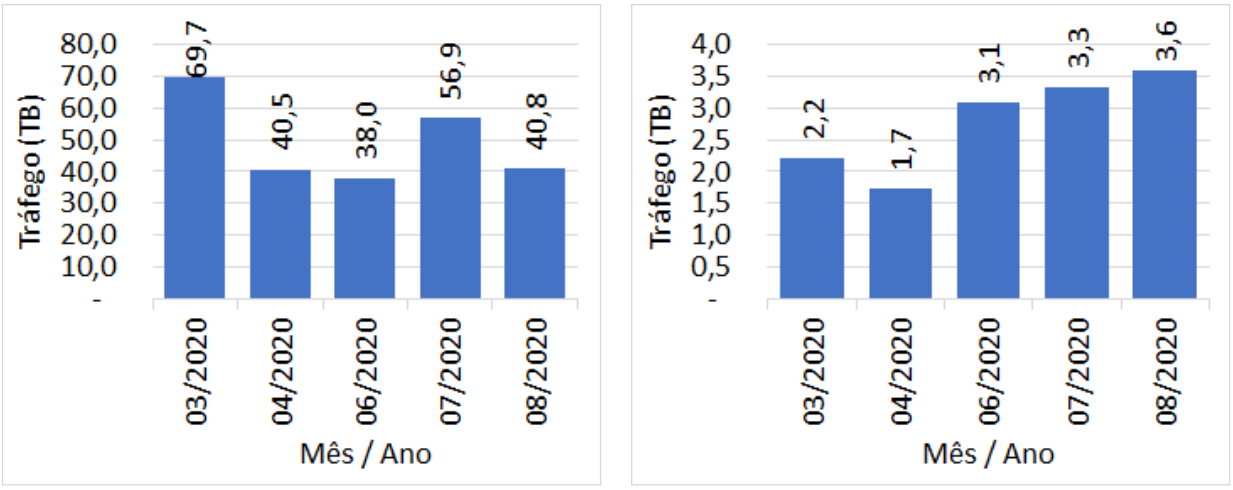


(a) Tráfego total (b) Tráfego das principais portas (80/443)

Figura 3.36: Gráficos comparativos de tráfego TCP

Como mostram as Figuras 3.36a e 3.37a, a redução do tráfego total do protocolo UDP foi menor que a do TCP, com o índice de 42%. Esse menor impacto pode estar relacionado ao fato de protocolos de gerência de redes, roteamento e outros, como o *Simple Network Management Protocol* (SNMP), *Routing Information Protocol* (RIP), *Dynamic Host Configuration Protocol* (DHCP) e *Domain Name System* (DNS), serem (majoritariamente) baseados no UDP. Além disso, a Figura 3.37b mostra que houve significativo acréscimo no tráfego das aplicações de comunicação analisadas neste estudo (ferramentas e portas usadas): *Facebook Messenger* (40003), *Google Meet* (19305), *MS Teams* (3478, 3479, 3480, 3481), *Skype* (3478, 3479, 3480), *Webex* (9000), *WhatsApp* (3478) e *Zoom* (8801). Destaca-se que o *Skype* e o *WhatsApp* usam portas altas aleatórias no caso de chamadas entre apenas dois participantes, se utilizando conexão par-a-par. Assim, esse cenário não foi incluído na análise.

A queda no tráfego referente às aplicações de comunicação, registrada apenas no mês posterior ao da decretação da pandemia, não foi proporcional à queda do tráfego total, sendo significativamente menor. O volume de tráfego dessas aplicações, nos meses



(a) Tráfego total

(b) Tráfego das portas de aplicações de comunicação (40003/19305/3478/3479/3480/3481/9000/8801)

Figura 3.37: Gráficos comparativos de tráfego UDP

posteriores, não acompanhou a tendência de estabilidade do volume de tráfego total do UDP, excetuado o mês de julho de 2020, onde houve um pico de tráfego 29% superior a média de tráfego dos meses de março a agosto de 2020 (Figura 3.37). Em vez disso, o tráfego referente às aplicações de videoconferência registrou expressivo aumento, com exceção do mês de abril de 2020, no qual houve uma readaptação da sociedade. Em agosto de 2020, o tráfego dessas portas foi quase o dobro daquele registrado antes da pandemia. O destaque no tráfego de videoconferência é registrado nas portas 8801 (utilizada pelo *Zoom*) e 19305 (utilizada pelo *Google Meet*), ambas chegando a quase 5 vezes de incremento no volume de tráfego, se comparados os meses março e agosto de 2020 (Figura 3.38).

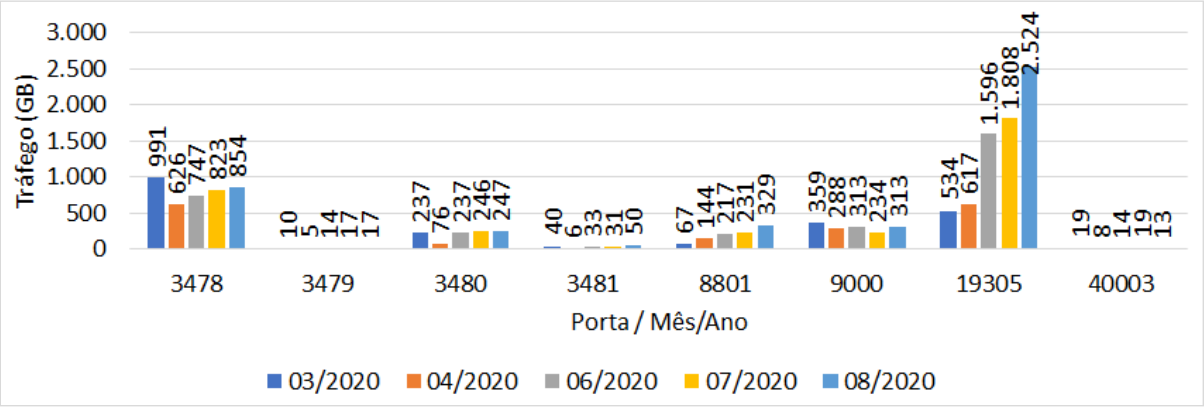


Figura 3.38: Gráfico de tráfego UDP das portas referentes a videoconferência (GB)

3.3.2 Análise de Tráfego de Redes sem Fio

No curso deste estudo foram conduzidas diversas fases de captura e análise de tráfego em redes sem fio. O objetivo foi ponderar se alguma característica dessas redes poderia influenciar a percepção dos usuários no que concerne à estabilidade da conexão de Internet. Outro objetivo foi avaliar a evolução do tráfego dessas redes e sua correlação com os efeitos causados pela pandemia. É imprescindível salientar que a obtenção dos dados de tráfego ocorreu em um ponto fixo situado no bairro de Santa Rosa, na cidade de Niterói, no estado do Rio de Janeiro (Fig. 3.39). Portanto, os dados fornecem uma visão do tráfego em uma área específica. Contudo, devido a origem dos diversos dados coletados no decorrer deste trabalho guardarem proximidade geográfica, se torna factível realizar comparações entre eles.

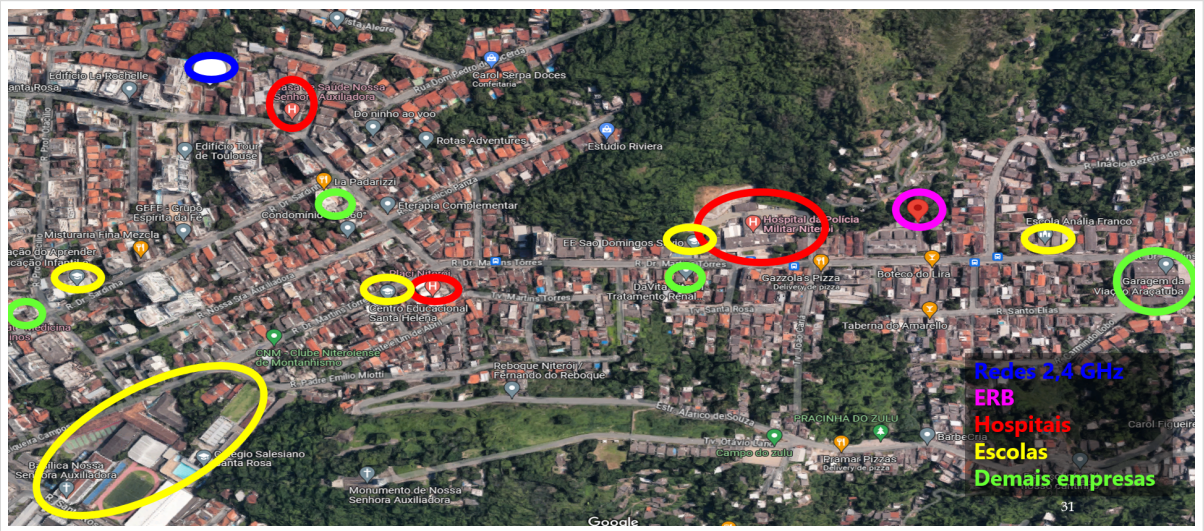


Figura 3.39: Mapa da região onde ocorreu a captura de tráfego e a coleta de dados da ERB da operadora TIM

Para analisar uma possível influência na percepção da estabilidade da conexão de Internet, foi conduzida uma captura e análise de tráfego nas faixas de frequência de 2,4 GHz e de 5 GHz. Considerando o resultado obtido e a necessidade de avaliar a evolução do tráfego das redes sem fio, correlacionando-a aos efeitos causados pela pandemia, foi iniciada uma série de capturas de tráfego da faixa de frequência de 2,4 GHz, e posterior análise dos dados obtidos, confrontando-os com as estatísticas de mobilidade do *Google*.

3.3.2.1 Metodologia de Captura de Tráfego

Foi utilizado um notebook com sistema operacional Linux, cuja interface de rede foi configurada no modo monitor[4]. Com isso, o filtro que descarta os quadros recebidos pelo meio físico que não sejam para o BSS ao qual essa interface está associada é desabilitado e todos os quadros recebidos no meio físico são repassados para o sistema. Adicionalmente, quando configurada em modo monitor, a interface inclui um pseudo-cabeçalho antes de entregar o quadro ao sistema. Esse pseudo-cabeçalho é chamado de *Radiotap*[5] e contém informações sobre os aspectos físicos do quadro, como por exemplo, a potência do sinal recebido, a taxa de transmissão utilizada e o tempo que a transmissão do quadro levou no meio físico.

Durante sete dias de cada mês, iniciando no domingo e finalizando no sábado, foram capturados vinte segundos de cada um dos treze canais da faixa de frequência de 2,4 GHz[6] a cada hora. Para possibilitar a automatização da captura, foi utilizada a ferramenta de linha de comando *tshark (Terminal wireSHARK)*[7]. Os campos armazenados de cada pacote foram: *Frame Time Epoch, Frame Number, Frame Length, Frame Protocols, WLAN Transmitter Addr, WLAN Receiver Addr, Radiotap Antenna, Radiotap Channel Frequency, WLAN Radio Channel, Radiotap Data Rate, Radiotap dbm Antenna Signal, Radiotap Flags, WLAN Radio PHY Type, WLAN Radio Duration, WLAN Radio IFS* e *WLAN Radio Preamble*.

Para obter as estatísticas do canal, por exemplo, a ocupação, foi usada a ferramenta de linha de comando *iw*[8]. Utilizando essa ferramenta, associada ao seu comando *dev <dev-name> survey dump*, foram extraídas as seguintes informações: *Date/Time, Frequency, Noise, Channel Active Time, Channel Busy Time, Channel Receive Time* e *Channel Transmit Time*.

Os dados foram, então, convertidos para o formato CSV [66] e posteriormente armazenados em um banco de dados. Assim, foi possível utilizar uma ferramenta de *Business Intelligence*, a fim de facilitar o tratamento dos dados.

As capturas ocorreram desde o mês de maio de 2021 até o mês de outubro de 2022,

[4]https://www.aircrack-ng.org/doku.php?id=airmon-ng

[5]https://www.radiotap.org/

[6]Canais: 01 - 2,412 GHz; 02 - 2,417 GHz; 03 - 2,422 GHz; 04 - 2,427 GHz; 05 - 2,432 GHz; 06 - 2,437 GHz; 07 - 2,442 GHz; 08 - 2,447 GHz; 09 - 2,452 GHz; 10 - 2,457 GHz; 11 - 2,462 GHz; 12 - 2,467 GHz; e 13 - 2,472 GHz.

[7]https://www.wireshark.org/docs/man-pages/tshark.html

[8]https://www.kali.org/tools/iw/

quando o *Google* deixou de atualizar os relatórios de mobilidade da comunidade[9], mais especificamente na data de 15/10/2022.

3.3.2.2 Faixa de Frequência 2,4 GHz vs. Faixa de Frequência 5 GHz

Considerando a percepção de 61,5% (Fig. 3.9, Subseção 3.1.1.2) dos respondentes da primeira pesquisa de opinião de terem enfrentado instabilidades no acesso à Internet e utilização maciça das ferramentas de tele trabalho e de comunicação multimídia, considerou-se importante avaliar a origem dessas instabilidades.

Essa avaliação iniciou-se através da análise da primeira pesquisa de opinião (Subseção 3.1.1.2). Foi notada fraca ou inexistente correlação entre a percepção da estabilidade da conexão com a Internet e a quantidade de dispositivos conectados à Internet na residência (Fig. 3.10), entre a percepção da estabilidade da conexão com a Internet e a quantidade de moradores na mesma residência (Fig. 3.11) e entre a percepção da estabilidade da conexão com a Internet e o plano de Internet fixa contratado (Fig. 3.12).

Esses dados podem indicar que o problema na percepção de instabilidade da conexão com a Internet não tem, necessariamente, origem nas condições das redes de dados residenciais ou nos planos contratados. Desse modo, qual seria a origem dessas instabilidades, uma vez que as correlações analisadas não demonstraram indícios de ingerência nesse quesito?

Além da possibilidade de problemas na infraestrutura de última milha, uma perspectiva é que uma eventual maior utilização da faixa de frequência de 2,4 GHz — que possui menor largura de banda e, portanto, menor capacidade de absorver o crescimento da demanda associado à pandemia —, em detrimento da faixa de 5 GHz, pode ser fator preponderante para elevar a percepção de instabilidade da conexão com a Internet. Para examinar essa possibilidade, realizou-se uma captura de tráfego em todos os canais das faixas de 2,4 GHz e 5 GHz, em abril de 2021. Observou-se que a utilização da faixa de 5 GHz é menor que a da faixa de 2,4 GHz. O volume de tráfego capturado na faixa de frequência de 5 GHz foi 73% menor do que o volume de tráfego capturado na faixa de frequência de 2,4 GHz (Fig. 3.40) e a quantidade de pacotes capturados na faixa de frequência de 5 GHz foi 28% menor do que a quantidade de pacotes capturados na faixa de frequência de 2,4 GHz (Fig. 3.41).

Cabe ressaltar que a diferença registrada no volume de tráfego e na quantidade de

[9]https://www.google.com/covid19/mobility/

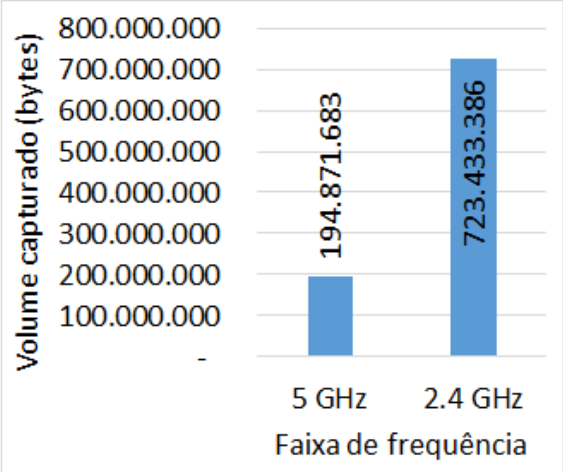


Figura 3.40: Volume do tráfego capturado nas faixas de 5 GHz e 2.4GHz

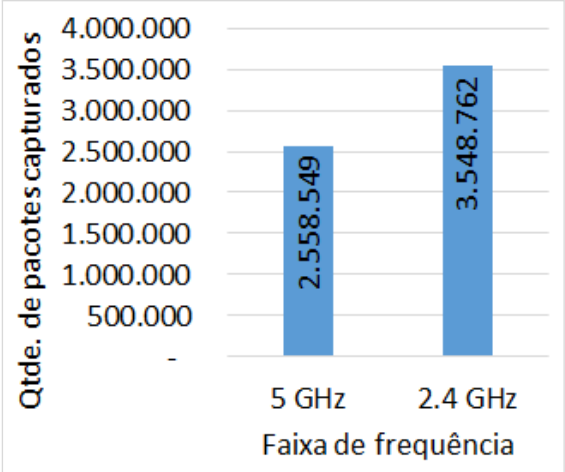


Figura 3.41: Quantidade de pacotes capturados nas faixas de 5 GHz e 2,4GHz

pacotes capturados pode sofrer influência de uma das características que diferencia as redes que operam nessas faixas de frequência. A área de cobertura das redes que operam na faixa de frequência de 2,4 GHz é maior do que a área de cobertura das redes que operam na faixa de frequência de 5 GHz. A segunda pesquisa de opinião (Seção 3.1.2) dirimiu essa dúvida, onde constatou-se que somente 28% dos respondentes utilizam a faixa de frequência de 5 GHz na rede sem fio doméstica (Fig. 3.27).

Considerando que a faixa de frequência mais utilizada é também aquela com maior área de cobertura e, por isso, mais suscetível a interferência de outras redes nas proximidades, elocubra-se a hipótese de que a maior utilização da faixa de frequência de 2,4 GHz em detrimento da faixa de frequência de 5 GHz pode ser fator preponderante para formar a percepção dos usuários de que a conexão com a Internet apresenta instabilidades. É importante ressaltar que a influência da localização geográfica dos participantes da pesquisa de opinião contribuiu para a comparação dos dados de tráfego capturados com os resultados dessa pesquisa.

3.3.2.3 Análise de Tráfego de Redes sem Fio na Faixa de Frequência 2,4 GHz

As capturas mensais do tráfego de redes sem fio, na faixa de 2,4 GHz, tiveram início em maio de 2021. Conforme ilustrado na Figura 3.42, a partir dessa data houve uma tendência de crescimento na média do volume de tráfego, que persistiu até setembro de 2021. Após esse período, se iniciou uma tendência de queda até dezembro de 2021. De janeiro a setembro de 2022, o volume médio de tráfego diário oscilou entre 68,5 MB e 78,4 MB, com picos de 85,6 MB em dias úteis de março de 2022 e 96 MB em finais de semana

de junho do mesmo ano. O volume médio de tráfego diário no mês de maio de 2021 foi de 56,4 MB, o menor de toda série. Embora em junho de 2022 as restrições, em virtude da pandemia, tenham sido muito menores, ou praticamente extintas, o volume médio de tráfego diário não retrocedeu ao patamar do ano anterior. Para exemplificar, os meses de maio e junho registraram um aumento nesse indicador da ordem de 39% e 24% (Fig. 3.43), respectivamente, quando comparados os anos de 2021 e 2022. A partir de julho de 2022 iniciou-se uma tendência de queda no tráfego, cuja média ficou em torno de 4% ao mês. Nos meses de agosto, setembro e outubro de 2022 houve queda mais expressiva na média diária de tráfego, se comparado ao mesmo período do ano anterior, da ordem de 25% em cada mês. A retração no tráfego iniciou-se com o período de férias e, associada à redução das restrições de mobilidade, o que pode ser verificado pela menor média[10] da concentração residencial em julho de 2022 (3% acima da base estabelecida pelo *Google*, como nível normal pré-pandemia) em relação à média da concentração residencial em julho de 2021 (9% acima da base estabelecida pelo *Google*, como nível normal pré-pandemia), continuou nos meses subsequentes.

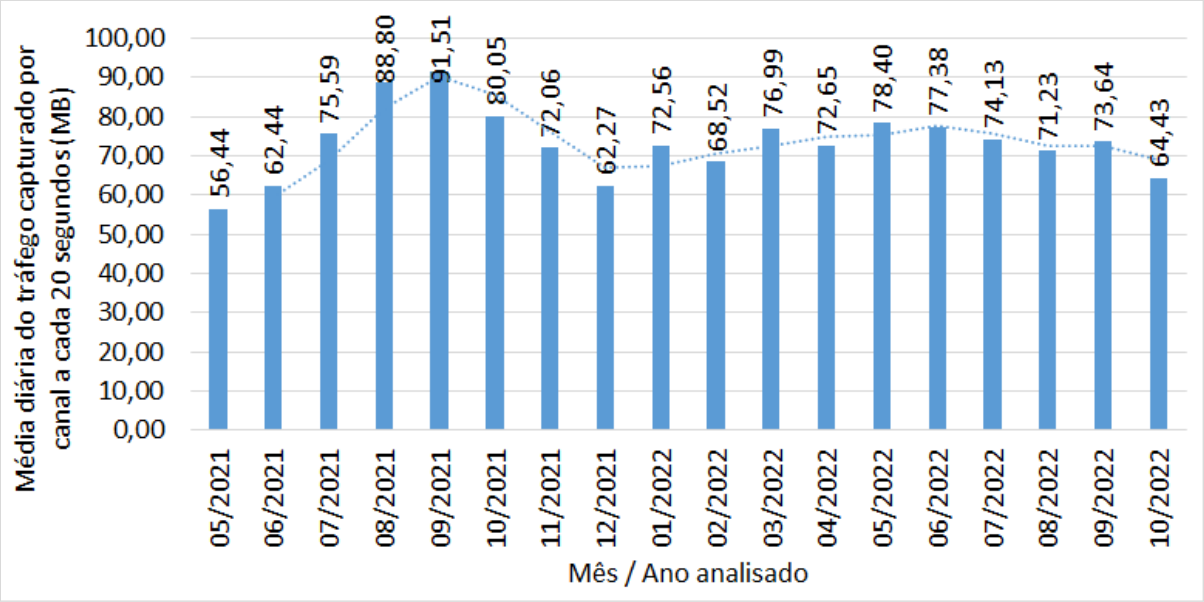


Figura 3.42: Média diária do tráfego capturado, por mês, em mega bytes, com linha de tendência do tipo média móvel

O volume médio de tráfego de todas as faixas horárias segue, aproximadamente, o mesmo padrão, tanto para finais de semana (Fig. 3.44) quanto para dias úteis (Fig. 3.45). Os picos de tráfego mais relevantes registrados em todo o período analisado foram: finais de semana de maio de 2021, entre 16:00 e 19:59; finais de semana de agosto de 2021, entre 20:00 e 23:59; finais de semana de janeiro de 2022, entre 12:00 e 15:59; e entre 04:00

[10]Média relativa à cidade de Niterói no período de captura de pacotes.

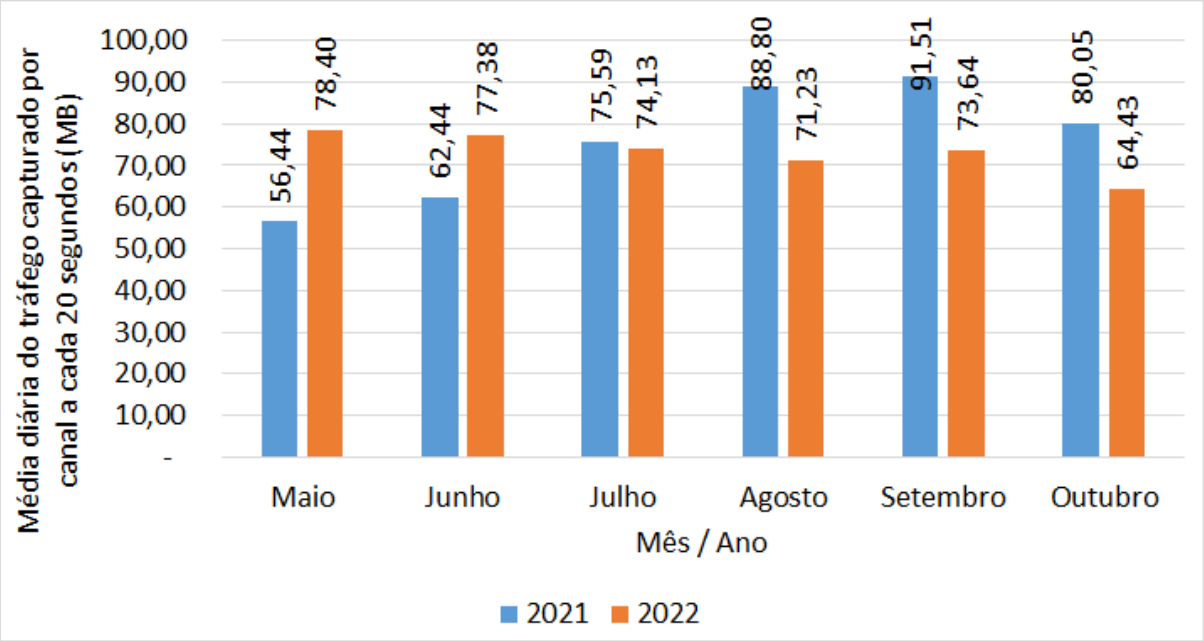


Figura 3.43: Comparativo da média diária do tráfego capturado, por mês, em megabytes

e 07:59 nos finais de semana dos meses de novembro de 2021 e maio de 2022.

Comparando o mês de maio dos anos de 2021 e 2022 (Fig. 3.46a), todas as faixas de horário em 2022 registraram tráfego superior ao de 2021, com exceção da faixa de 16h00 às 19h59, em virtude de um tráfego anômalo observado nessa faixa horária no ano de 2021. O mesmo ocorreu em junho (Fig. 3.46b), porém sem a exceção. Na comparação entre os anos de 2021 e 2022, referentes ao mês de julho (Fig. 3.46c), as faixas de 00h00 às 03h59, 04h00 às 07h59 e 12h00 às 15h59 registraram maior tráfego em 2022 e as faixas de 08h00 às 11h59, 16h00 às 16h59 e 20h00 às 23h59 registraram maior tráfego em 2021. Já para os meses de agosto, setembro e outubro, todas as faixas horárias registraram tráfego superior em 2021, em comparação ao ano de 2022, deixando mais clara a tendência de redução do tráfego, possivelmente em decorrência da volta à normalidade e cessação quase integral das restrições impostas no período da pandemia.

O gráfico constante da Figura 3.47 apresenta o acompanhamento da média mensal da ocupação do canal, representada pela análise de 20 segundos de cada canal por hora, conforme descrito na metodologia de captura de tráfego (Sub-subseção 3.3.2.1), observa-se que a ocupação do canal registrou crescimento entre maio e outubro de 2021, desconsiderando o mês de setembro do mesmo ano, que registrou uma ocupação do canal incomum. A partir de novembro de 2021, a ocupação do canal manteve-se estável, com índices pouco abaixo dos 3 segundos e picos de 4 segundos, em uma janela de 20 segundos. No mês de setembro de 2021 a ocupação do canal ficou, em média, na faixa dos 6 segundos, em

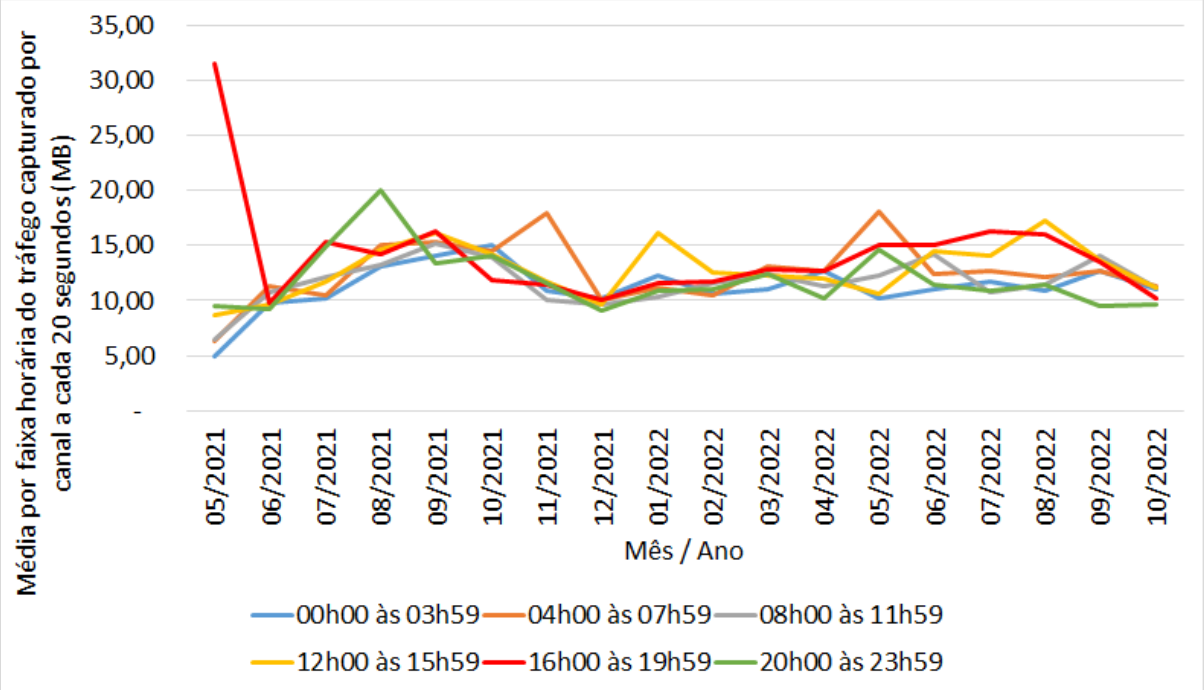


Figura 3.44: Média do tráfego capturado em finais de semana para cada faixa horária, por mês, em megabytes

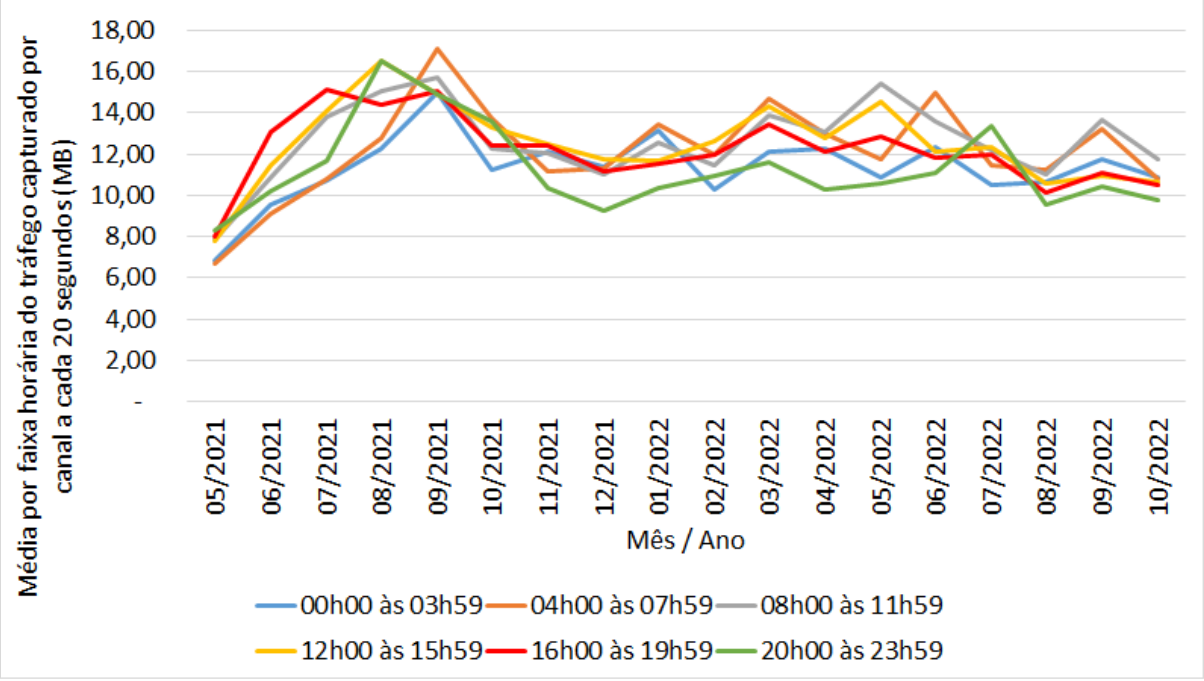
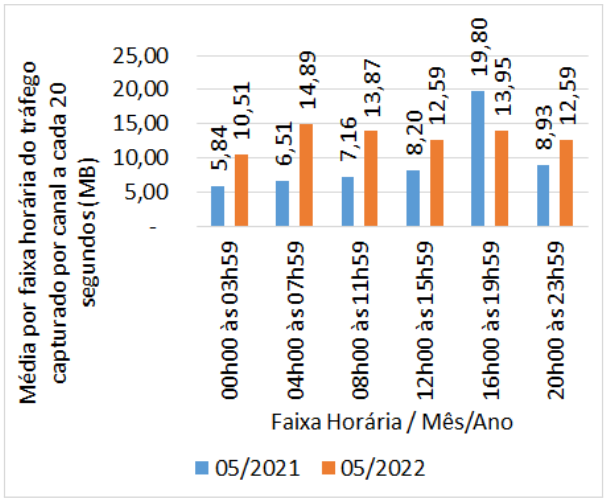
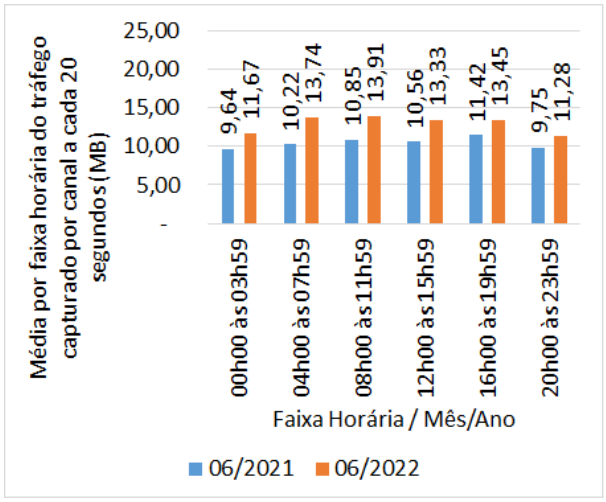


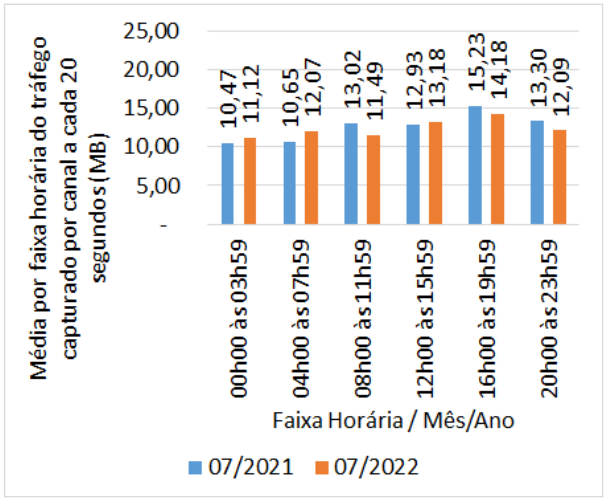
Figura 3.45: Média do tráfego capturado em dias úteis, por canal, por mês, em megabytes



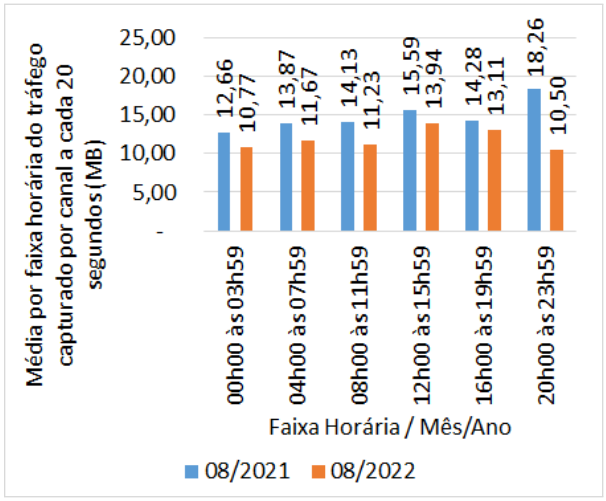
(a) Mês de maio



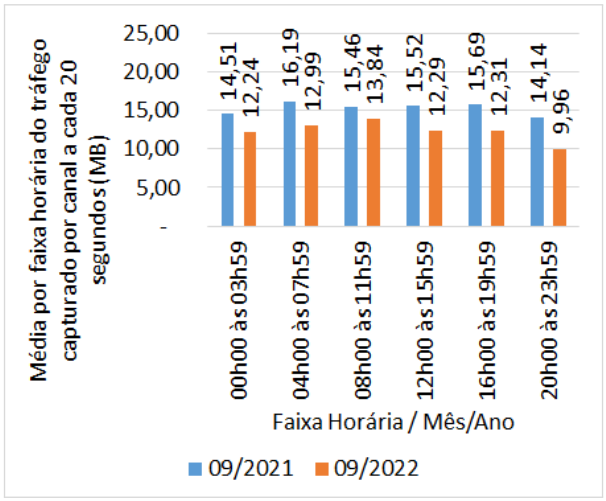
(b) Mês de junho



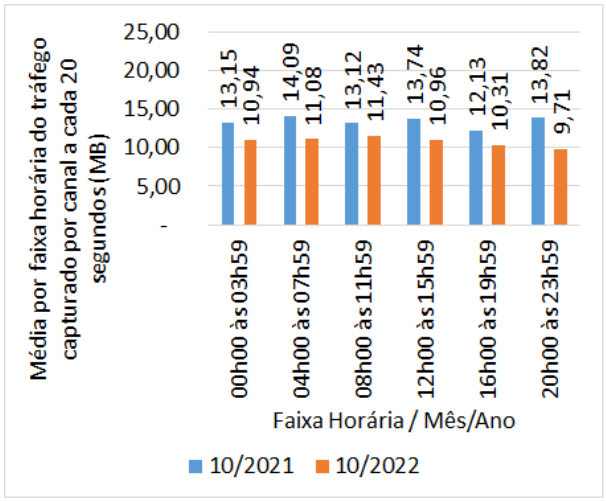
(c) Mês de julho



(d) Mês de agosto



(e) Mês de setembro



(f) Mês de outubro

Figura 3.46: Comparativo da média para cada faixa horária do tráfego capturado, em megabytes

uma janela de 20 segundos. A grosso modo, a média geral registrou ocupação do canal equivalente a aproximadamente $\frac{1}{7}$ do tempo total, ficando, em setembro, na ordem de $\frac{1}{3}$ do tempo total.

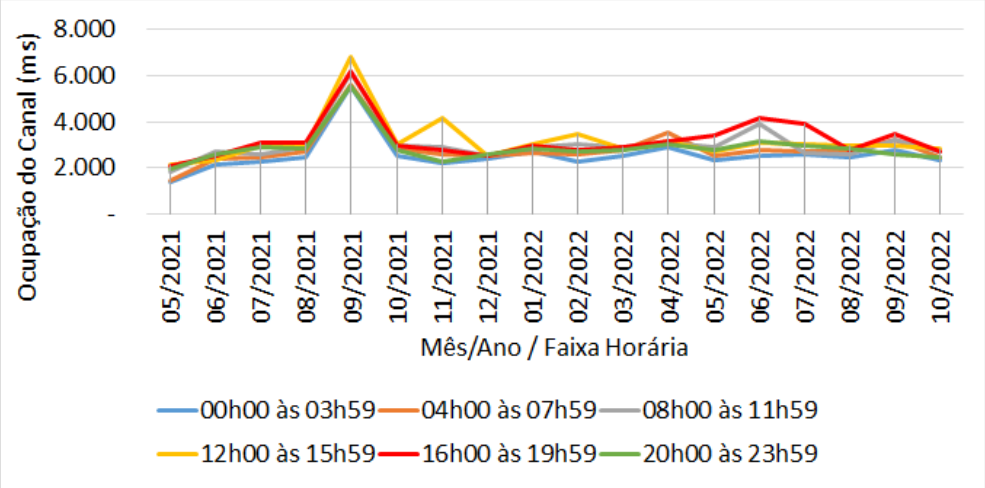


Figura 3.47: Média da ocupação de cada canal da faixa de 2,4 GHz, por mês

3.3.3 Análise de Dados de Telefonia Celular

Em concurso com a TIM Brasil, uma das maiores operadoras de telefonia móvel do Brasil, analisaram-se registros de conexão de usuários, compilados em intervalos de 5 minutos e anonimizados, em uma ERB, cuja localização é -22.898361º de latitude e -43.090417º de longitude, cidade de Niterói, no estado do Rio de Janeiro. O critério de seleção da ERB foi essencialmente pela região onde ela encontra-se instalada, que é a mais próxima do local onde foi realizada a análise de tráfego de redes sem fio. Esses registros de conexão são compostos por chamadas, mensagens de texto e conexões de dados 3G/4G. Os dados utilizados nesta Dissertação são da mesma fonte utilizada em [5]. No entanto, a proposta daquele trabalho é a análise da mobilidade e, neste trabalho, a proposta é a comparação dos dados de mobilidade com características de tráfego de redes sem fio, com o objetivo de avaliar a existência de correlação entre eles.

Conforme gráfico plotado na Figura 3.48, em outubro de 2021, iniciou-se um aumento sucessivo no quantitativo médio diário de dispositivos conectados na ERB, que durou até janeiro de 2022, a partir de quando o indicador começou a cair. De março a agosto de 2022 o quantitativo ficou estável, tendo um expressivo aumento em setembro de 2022, da ordem de 41%. Logo após, em outubro de 2022, caiu 19%, voltando ao patamar de janeiro do mesmo ano. Em setembro de 2022, a quantidade média diária de dispositivos conectados à ERB foi 54% acima da média histórica. Em que pese a queda em outubro de 2022, em

relação ao mês anterior, a quantidade média diária de dispositivos conectados à ERB foi 4% acima da média histórica. O aumento repentino no número de conexões pode ser em virtude da migração da carteira de clientes da operadora Oi para outras operadoras. Em todos os meses analisados, seja em dias úteis ou finais de semana, a quantidade média de dispositivos conectados à ERB segue um padrão, onde esse quantitativo aumenta a partir das 06h, segue estabilizado a partir das 10h e começa a cair às 20h. O período que registra o menor número de conexões é de 03h às 05h.

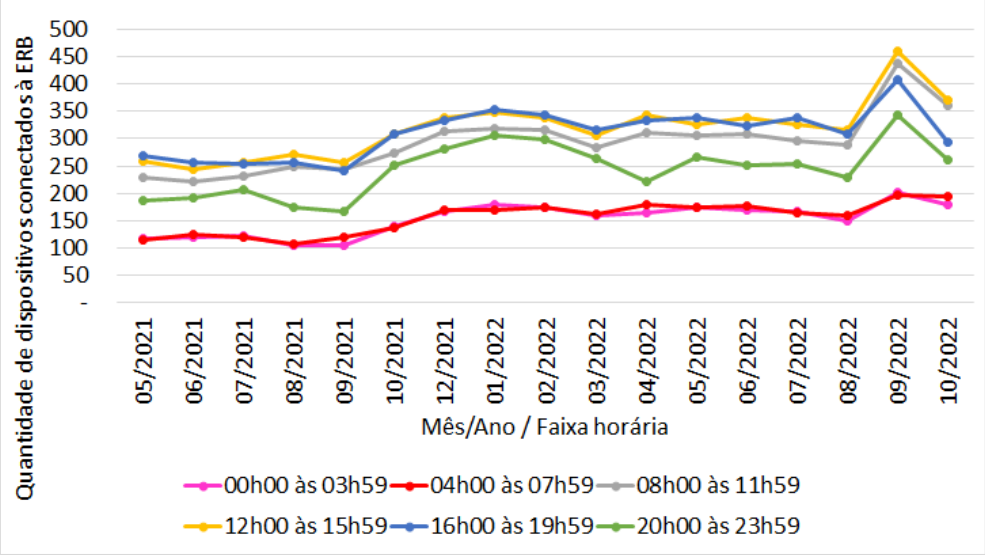


Figura 3.48: Média da quantidade de telefones conectados à ERB em cada faixa horária por mês

No que concerne à correlação entre a quantidade de conexões de dispositivos celulares e a média da concentração residencial mensal (Fig. 3.50), observa-se ser grande e negativa. Já a correlação entre a quantidade de conexões de dispositivos celulares e a média da concentração em locais de trabalho mensal (Fig. 3.50), observa-se, também, ser grande, porém positiva. Tendo em vista a maior área de cobertura de uma ERB em relação às redes sem fio da faixa de 2,4 GHz, é possível que essa correlação tenha sido influenciada pela proximidade de três hospitais, cinco escolas, uma garagem de uma empresa de ônibus e demais pequenos comércios próximos. Não foi possível constatar correlação significativa entre a quantidade de conexões de dispositivos celulares e o tráfego ou a quantidade de pacotes capturados na faixa de 2,4 GHz das redes sem fio.

3.4 Impactos na Mobilidade e Seus Reflexos nas Redes de Computadores

Conforme ilustrado na Figura 3.49, na cidade de Niterói, nos períodos em que foram realizadas as capturas de tráfego, a concentração em locais de trabalho, em comparação com a base estabelecida pelo *Google*, como nível normal pré-pandemia, registrou um padrão negativo (abaixo do normal) em dias úteis e feriados (nesse caso, o da independência) e positivo (acima do normal) em finais de semana. Em março de 2022, esse padrão começou a mudar e passou a ser positivo em todas as situações observadas. A tendência de crescimento na concentração em locais de trabalho esteve presente desde maio de 2021, com exceção do mês de setembro de 2021, supostamente em virtude do feriado da independência, e do mês de janeiro de 2022, possivelmente por ser um mês tradicionalmente de férias. Embora a média da semana analisada em setembro de 2022 não tenha sido negativa, o dia 7 de setembro, especificamente, registrou -63% de concentração em locais de trabalho. A concentração residencial, inversamente proporcional à concentração em locais de trabalho, registrou tendência de queda. A partir de março de 2022 houve uma estabilização desse indicador, cujo maior valor registrado sendo de 5%. Em outubro de 2022, houve registro de concentração residencial negativa (-1%), em comparação com a base estabelecida pelo *Google*, como nível normal pré-pandemia.

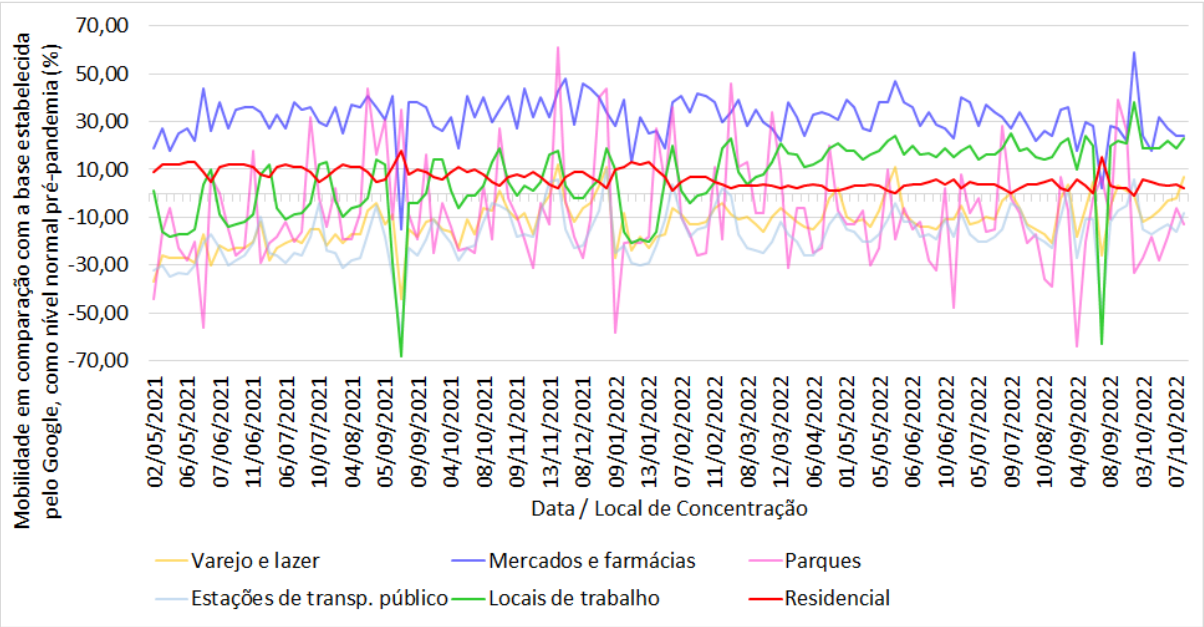


Figura 3.49: Dados da mobilidade da comunidade, disponibilizados pelo *Google*), nos dias em que foram realizadas as capturas de tráfego, referentes à cidade de Niterói

No mês de abril de 2022 a rotina já era praticamente igual àquela anterior à pande-

mia. Nesse momento, já era possível observar trânsito intenso em horários cada vez mais cedo, o carnaval foi realizado normalmente, as empresas e escolas já haviam retornado às atividades presenciais e a obrigatoriedade das medidas de proteção, como as máscaras, já haviam sido sustadas, como pode-se verificar na linha do tempo das ações da Prefeitura Municipal de Niterói para combate ao coronavírus [17]. Essa observação pôde ser confirmada no gráfico de mobilidade 3.49, em especial de concentração residencial e em locais de trabalhos, bem como no de volume de tráfego, que manteve certa estabilidade no ano de 2022.

Verifica-se no mapa de calor, representando a correlação entre o volume de tráfego e a quantidade de pacotes capturados, as estatísticas de mobilidade do *Google*[11], a quantidade de conexões de telefones celulares a uma Estação Rádio Base (ERB) (Seção 3.3.3) e as estatísticas sobre a COVID-19[12], utilizando a metodologia do Coeficiente de Correlação de Pearson (r), que há uma correlação positiva, considerada média[13] [13], entre a média da concentração residencial mensal e média mensal do volume de tráfego. Ao observar a correlação do volume de tráfego com a concentração em mercados e farmácias e com a concentração em áreas de varejo e lazer, há uma correlação negativa considerada grande. A correlação é sempre mais forte quando os valores de mobilidade são comparados com a quantidade de pacotes trafegados, em vez de comparados ao volume de tráfego. Também há uma correlação negativa, considerada média, entre a média da concentração em locais de trabalho mensal e média mensal do volume de tráfego. Uma das hipóteses para que essa correlação não seja mais forte é de que o tráfego de Internet já tem um crescimento natural. Isso pode ser observado na evolução do volume de tráfego registrado pelo IX.br (Fig. 3.51). Através de notícias veiculadas no site da instituição [46], [50], [51], [47], [48] e [49] foi possível calcular a derivada primeira, ou seja, a taxa de crescimento mensal do tráfego de Internet para o período antes da pandemia, conforme descrito nas equações 3.1 e 3.2. A derivada primeira também foi calculada para o período após a decretação da pandemia, conforme descrito nas equações 3.3, 3.4 e 3.5. O período entre 12/2019 e 03/2020 não foi calculado porque parte encontra-se antes da pandemia e parte durante.

$$\frac{7\ Tb/s - 5\ Tb/s}{07/2019 - 06/2018} = \frac{2}{13} = 0,15 \text{ Tb/s por mês} \tag{3.1}$$

[11] `https://www.google.com/covid19/mobility/`

[12] `https://covid.saude.gov.br/`

[13] Correlações: escores < 0,30 são consideradas pequenas; 0,30 >= escores < 0,50 são consideradas médias; e escores >= 0,50 são consideradas grandes.

$$\frac{8\ Tb/s - 7\ Tb/s}{12/2019 - 07/2019} = \frac{1}{5} = 0,2 \text{ Tb/s por mês} \tag{3.2}$$

$$\frac{14\ Tb/s - 10\ Tb/s}{12/2020 - 03/2020} = \frac{4}{9} = 0.44,\ \text{Tb/s por mês} \tag{3.3}$$

$$\frac{16\ Tb/s - 14\ Tb/s}{03/2021 - 12/2020} = \frac{2}{3} = 0.67,\ \text{Tb/s por mês} \tag{3.4}$$

$$\frac{20\ Tb/s - 16\ Tb/s}{12/2021 - 03/2021} = \frac{4}{9} = 0.44,\ \text{Tb/s por mês} \tag{3.5}$$

A partir desses resultados, foi possível calcular a derivada segunda, que representa a taxa de variação da taxa de crescimento. Para o período pré-pandemia, essa variação foi de 0,05 Tb/s ($0,2 - 0,15 = 0,05$). Durante a pandemia, a primeira variação registrada foi de 0,23 Tb/s ($0,67 - 0,44 = 0,23$). A segunda variação sofreu ligeira retração de 0,23 Tb/s ($0,44 - 0,67 = -0,23$). Assim, pode-se depreender que a taxa de crescimento do tráfego acelerou muito chegando a mais de 4 vezes a taxa pré-pandemia. Isso indica que a pandemia impactou direta e severamente o volume de tráfego na Internet.

3.5 Trabalhos Relacionados

Outros trabalhos também analisaram o impacto da pandemia sobre o tráfego das redes de computadores com perspectivas complementares às deste estudo.

Em um trabalho sobre as implicações da pandemia da COVID-19 no tráfego da Internet [21], foi analisado um conjunto de dados obtidos em um ISP da Europa Central, uma rede acadêmica metropolitana espanhola e 3 IXP — localizados na Europa central, no sul da Europa e na costa leste dos EUA. Os autores encontraram uma redução de 55% no volume de tráfego da rede acadêmica — uma redução maior que os 35% encontrados na análise realizada neste estudo sobre os dados da RNP. Dados daquele trabalho também registraram um aumento de até 30% no tráfego dos IXP, após as medidas de restrição de mobilidade. Além disso, também foi notado um aumento significativo do tráfego de aplicações de videoconferência, em especial no horário comercial. O mesmo aconteceu com a porta TCP/993, usada pelo IMAP sobre TLS, e com a porta TCP/8200, usada por serviço de *streaming* de canais de TV russos. Com isso, os autores concluíram que o aumento foi relacionado a trabalho e entretenimento. Tal conclusão coaduna com

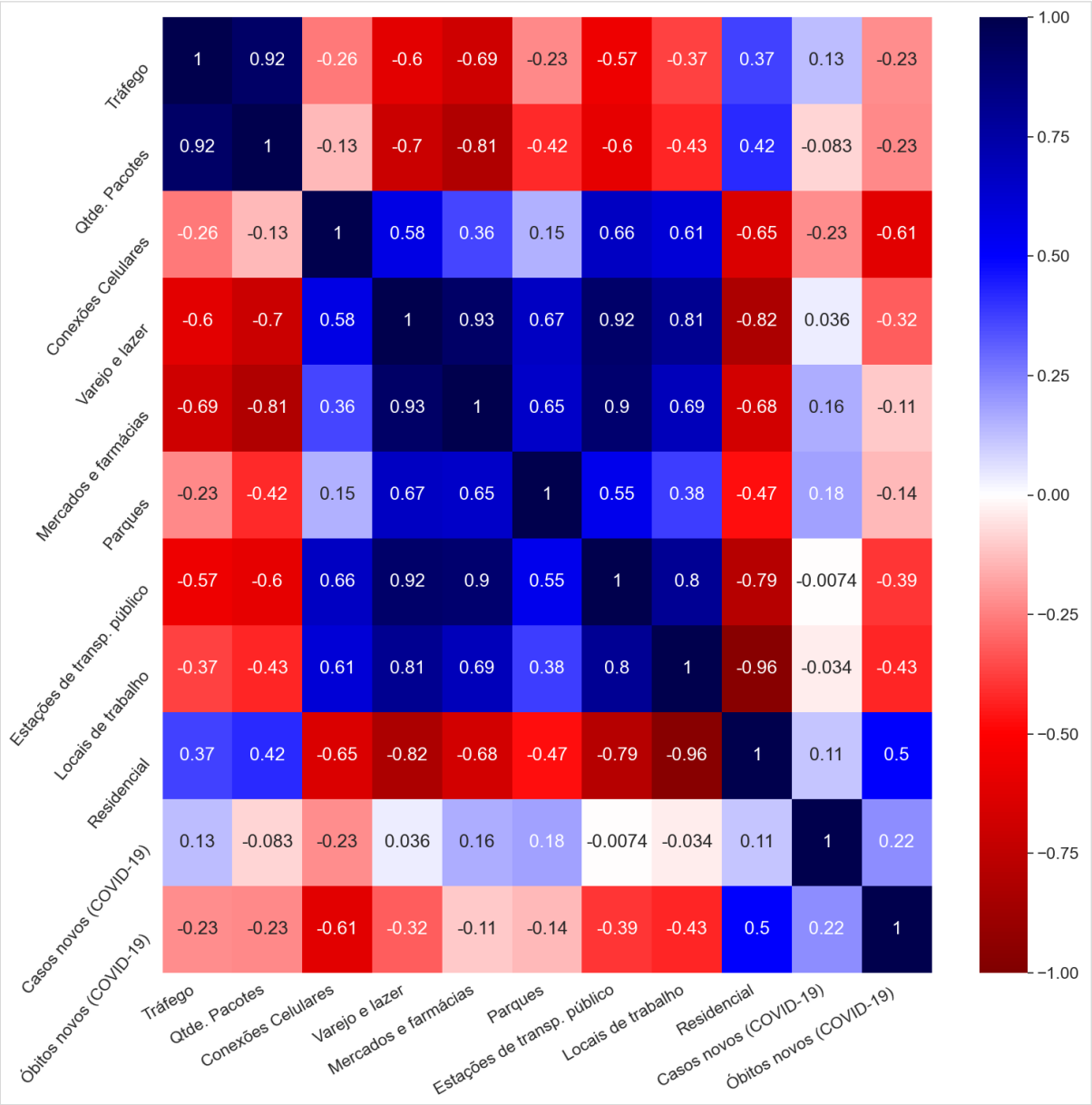


Figura 3.50: Mapa de calor com a correlação entre o tráfego e quantidade de pacotes capturados, estatísticas de mobilidade, quantidade de conexões de celulares e estatísticas sobre a COVID-19

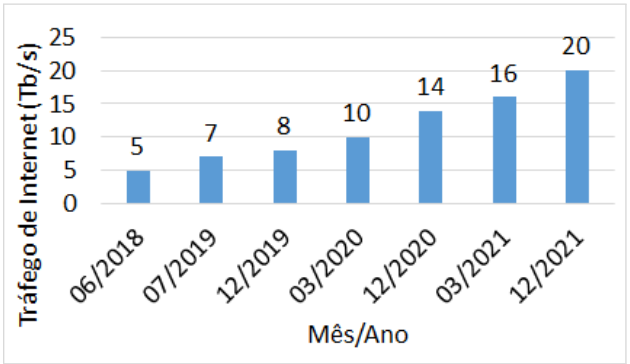


Figura 3.51: Evolução do tráfego registrado pelo IX.br

os resultados da primeira pesquisa de opinião, conduzida no âmbito deste estudo, onde verificou-se aumento na intensidade de uso da Internet na ordem de 27,7% para trabalho, 25,3% para estudo e 25,9% para lazer e/ou entretenimento.

Em artigo sobre a reação da Internet à COVID-19 [9], examinou-se a rede de borda do *Facebook*. É interessante observar o relato sobre a diferença na magnitude do impacto nas regiões do mundo. Para exemplificar, a América do Sul está elencada entre as regiões que experimentaram estresse em suas redes. Pudemos observar tal afirmação na primeira pesquisa de opinião, conduzida no âmbito deste estudo, na qual 39% dos respondentes relataram que passaram a observar uma instabilidade na conexão com a Internet, que não existia anteriormente. Assim como em outros estudos, o trabalho em tela relata que o aumento do tráfego se dá, principalmente, em decorrência de conteúdo multimídia. Para embasar suas conclusões, os autores também analisaram a QoE (*Quality of Experience*) de vídeos, utilizando a métrica BSR (*Bad Session Rate*). Eles encontraram uma degradação da experiência do usuário e a correlacionaram com métricas de rede, como RTT e redirecionamento do tráfego para enlaces indiretos. Ao observar que todos os fatores coincidiam, concluíram que há a indicação do congestionamento de rede.

Segundo um estudo produzido no Reino Unido [39], a redução na mobilidade foi 50%. Esse índice foi similar ao do Brasil[14] (em média, pouco mais de 50%). O conjunto de dados do estudo foi proveniente de uma operadora de rede móvel do Reino Unido com suporte à 2G, 3G e 4G. A informação curiosa apresentada pelo estudo é a redução do volume de tráfego de dados (*downlink*) da ordem de 25%. Correlacionando essa redução com o período de medidas restritivas decorrentes da COVID-19, os autores inferiram que uma causa possível seria uma maior utilização da banda larga residencial. Já o *uplink* sofreu alterações mais modestas, entre -7% e +1,5%, provavelmente em consequência do aumento do tráfego VoLTE (*Voice over Long Term Evolution*), que atingiu um pico de 150%, segundo os autores do artigo.

[14]https://www.google.com/covid19/mobility/

Capítulo 4

Aplicação de *Machine Learning* na predição da severidade e do tempo de correção de vulnerabilidades

4.1 Arcabouço Conceitual

Atualmente, a cibersegurança é uma fonte de preocupação para a maioria das organizações, devido ao crescimento das ameaças. Uma boa prática de segurança é auditar periodicamente os sistemas, verificando-os em busca de vulnerabilidades conhecidas. Para isso, existem bancos de dados de vulnerabilidades, como o NVD. Esses bancos de dados fornecem informações úteis sobre as vulnerabilidades, como severidade, explorabilidade e outras métricas. No entanto, o processo de geração dessas informações depende de uma análise manual de cada vulnerabilidade e, portanto, é possível que, para as vulnerabilidades mais recentes, tais informações ainda não estejam disponíveis no momento da consulta.

Para qualificar e avaliar as vulnerabilidades foi utilizado o padrão CVSS. Com o objetivo de maximizar a utilização dos dados disponíveis, possibilitando a análise de vulnerabilidades mais antigas, para as quais não há avaliação registrada na versão 3, foi empregada, neste estudo, a versão 2 desse padrão. A diferença entre as versões 2 e 3, no que se refere à severidade, é a inclusão de mais duas classificações, quais sejam *NONE* (*base score* = 0) e *CRITICAL* (*base score* >= 9). O CVSS, versão 2, possuí três grupos de métricas, quais sejam, base, temporal e ambiental. Dessas, apenas o grupo base é imutável sob a perspectiva do tempo e do ambiente do usuário. Nesse sentido, o grupo base é aquele que representa as características intrínsecas de uma vulnerabilidade. Por esse motivo, o grupo base foi o escolhido para ser discutido neste trabalho. O grupo base é

composto das métricas *Access Vector*, *Access Complexity*, *Authentication*, *Confidentiality*, *Availability* e a *Integrity*.

Para facilitar a avaliação e priorização no tratamento das vulnerabilidades, é utilizada uma gradação numérica de zero a dez, chamada de *base score*. O *base score* é obtido a partir das métricas CVSS, aplicando a seguinte Fórmula [58]:

$$BaseScore = round(((0,6 \times Impact) + (0,4 \times Exploitability) - 1,5) \times f(Impact)) \quad (4.1)$$

Nesta fórmula, a variável *Impact* é calculada como:

$$Impact = 10,41 \times (1 - (1 - ConfImpact) \times (1 - IntegImpact) \times (1 - AvailImpact)) \quad (4.2)$$

Já a função $f(Impact)$ usada na fórmula 4.1 retorna 0, se a variável *Impact* é igual a 0, ou 1,176, caso contrário. A variável *Exploitability* é definida como:

$$Exploitability = 20 \times AccessVector \times AccessComplexity \times Authentication \quad (4.3)$$

A Tabela 4.1 elenca as constantes presentes nas fórmulas.

Tabela 4.1: Definição de constantes presentes nas fórmulas para cálculo do *base score* a partir das métricas CVSS

Constante	**Condição/Impacto da Exploração da Vulnerabilidade**	**Valor**
AccessVector	Requer Acesso Local	0,395
	Requer Acesso à Rede Local	0,646
	Requer Acesso à Rede	1,000
AccessComplexity	Complexidade para Exploração Alta	0,350
	Complexidade para Exploração Média	0,610
	Complexidade para Exploração Baixa	0,710
Authentication	Requer Várias Instâncias de Autenticação	0,450
	Requer Instância Única de Autenticação	0,560
	Não Requer Autenticação	0,704
ConfImpact	Sem Impacto na Confiabilidade	0,000
	Impacta Parcialmente a Confiabilidade	0,275
	Impacta Totalmente a Confiabilidade	0,660
IntegImpact	Sem Impacto na Integridade	0,000
	Impacta Parcialmente a Integridade	0,275
	Impacta Totalmente a Integridade	0,660
AvailImpact	Sem Impacto na Disponibilidade	0,000
	Impacta Parcialmente a Disponibilidade	0,275
	Impacta Totalmente a Disponibilidade	0,660

O *base score* pode ser agrupado em três níveis distintos de severidade. A severidade

de uma vulnerabilidade é uma medida qualitativa[1]. Ela foi concebida com objetivo de auxiliar os desenvolvedores a priorizar a correção de vulnerabilidades. Considerando a versão 2 do CVSS, utilizada neste estudo, a severidade possui três níveis de gradação, obedecendo às seguintes regras:

- ***LOW***: *Base Score* < 4;
- ***MEDIUM***: $4 <=$ *Base Score* < 7; ou
- ***HIGH***: *Base Score* $>= 7$.

Nesse contexto, foram exploradas metodologias baseadas em aprendizado de máquina para classificar a severidade das vulnerabilidades e o tempo até que sejam corrigidas. Essas classificações são geradas a partir de suas descrições textuais e, possivelmente, a partir das métricas do vetor CVSS.

Os classificadores implementados no âmbito deste trabalho foram codificados na linguagem *Python*[2], baseados no módulo *scikit-learn* [56]. O *Python* é uma linguagem de programação multipropósito de código aberto e o módulo *scikit-learn* estende a linguagem, agregando diversos algoritmos de aprendizado de máquina. Neste estudo, foram utilizadas as ferramentas citadas focadas em problemas supervisionados.

A eficiência dessas metodologias foi avaliada usando um *dataset* composto por vulnerabilidades de diversas aplicações. No entanto, devido à particular relevância dos aplicativos de comunicação durante a pandemia do COVID-19, também foi avaliado o quão difícil é predizer essas informações para esse grupo específico de aplicativos a partir de modelos treinados com vulnerabilidades de aplicações que têm outras finalidades.

Abordagens para Predição da Severidade

Uma vez os classificadores treinados a partir do *dataset* de treinamento, eles podem ser utilizados para realizar predições para novos dados de entrada. Neste estudo em particular, técnicas de inteligência artificial foram empregadas para analisar as descrições textuais das vulnerabilidades e, com base nessa análise, gerar predições acerca de três importantes informações: severidade, *base score* e métricas do vetor CVSS, que medem o impacto e a explorabilidade das vulnerabilidades.

O objetivo final é a predição da severidade da vulnerabilidade. Para isso, foram consideradas três abordagens (Fig. 4.1). Em todas as abordagens, a entrada do classificador

[1]https://nvd.nist.gov/vuln-metrics/cvss
[2]https://www.python.org/

é a descrição textual da vulnerabilidade. Na primeira abordagem, a saída produzida pelo classificador é diretamente a severidade. Na segunda e na terceira abordagens, a severidade é obtida em função do *base score* da vulnerabilidade. Especificamente na segunda abordagem, foi empregado um classificador especializado para realizar a predição do *base score*. Já na terceira abordagem, o *base score* foi calculado usando os valores preditos das métricas CVSS, aplicando-se a fórmula descrita em [58].

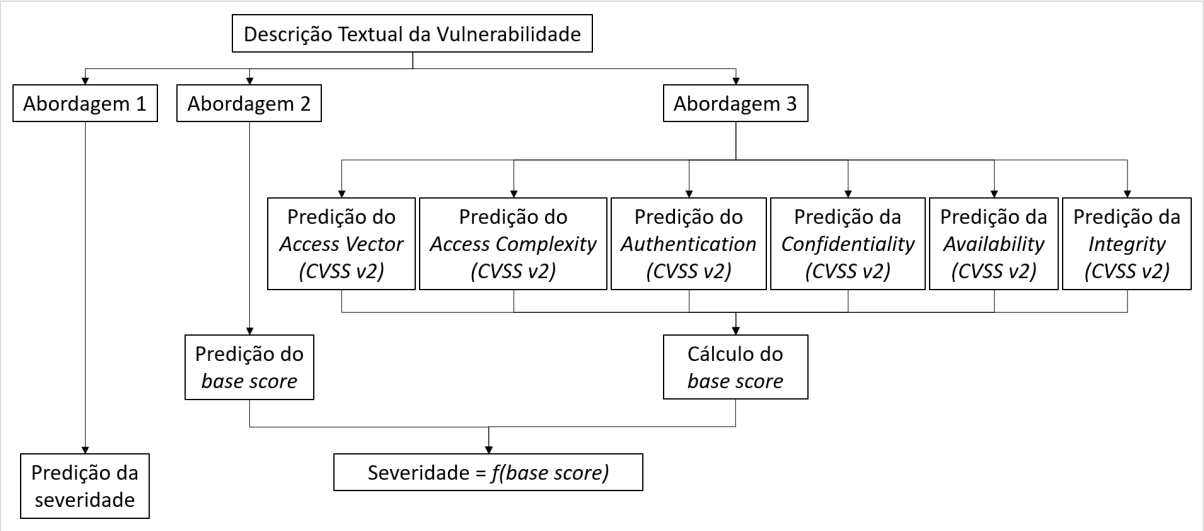


Figura 4.1: Fluxograma das abordagens para predição da severidade das vulnerabilidades

Extração de Dados

A base de dados CVE original[3] foi usada neste estudo como *dataset* completo. Essa base de dados contém a descrição textual das vulnerabilidades, mas não informações sobre as métricas do vetor CVSS das vulnerabilidades. Para complementar a base de dados CVE original com tais informações, foi construído um *script* para consultar a base de dados disponibilizada pelo NVD[4], para cada vulnerabilidade. Do *dataset* completo derivaram dois subconjuntos de dados, que foram utilizados nos experimentos: o *dataset* de treinamento/validação e o de teste. A segmentação entre esses dois subconjuntos é temporal, com o objetivo de examinar a estabilidade do modelo, em especial, avaliando seu desempenho em uma projeção de, aproximadamente, doze meses no futuro.

O *dataset* de treinamento/validação é composto por vulnerabilidades publicadas até março/2022. Com o objetivo de avaliar como a composição desse *dataset* afeta o desempenho dos classificadores, ele foi subdividido em dois tipos:

- **Balanceado**: a técnica de *under-sampling* aleatória [38] foi aplicada para que todas

[3] `https://www.cve.org/Downloads`
[4] `https://nvd.nist.gov/vuln/detail/[CVE-ID]`

as classes tivessem o mesmo número de amostras de vulnerabilidades. A classe com o menor número de vulnerabilidades no *dataset* de treinamento/validação foi totalmente extraída no *dataset* balanceado. As vulnerabilidades das demais classes foram selecionadas aleatoriamente até completar o quantitativo igual ao da classe com o menor número de vulnerabilidades; e

- **Proporcional**: as proporções das amostras de vulnerabilidades em cada classe são as mesmas do *dataset* de treinamento/validação. Para cada classe, as vulnerabilidades foram extraídas aleatoriamente do *dataset* de treinamento/validação até que o número desejado de amostras fosse obtido.

A quantidade total de amostras é a mesma para os dois tipos de *dataset* de treinamento/validação. Como a quantidade total de amostras do *dataset* balanceado é limitada pelo número de amostras da classe com menor quantidade de amostras, o tamanho desse *dataset* foi usado como limitador para o tamanho do *dataset* proporcional. No total, ambos os *datasets* têm 51.240 amostras. No *dataset* proporcional, há 5.260 vulnerabilidades de severidade baixa (10%), 29.528 médias (58%) e 16.452 altas (32%). Já no *dataset* balanceado são 17.080 amostras de cada uma das classes de severidade.

É importante ressaltar que o balanceamento e a proporcionalidade dos *datasets* de treinamento/validação são sempre baseados na severidade. Isso ocorre porque o produto final das três abordagens representadas na Figura 4.1 é a severidade. Outrossim, o balanceamento de classes para cada uma das métricas do vetor CVSS, por exemplo, inviabilizaria o cálculo do *base score*, conforme descrito na terceira abordagem.

O *dataset* de teste compreende as vulnerabilidades não incluídas no *dataset* de treinamento/validação, com exceção das vulnerabilidades de aplicações de comunicação, até o ano de 2022. Somente foram incluídas amostras cujos *base score* e métricas do vetor CVSS versão 2 existissem na base de dados NVD. O *dataset* de teste possui 12.726 amostras, sendo 1.856 vulnerabilidades de severidade baixa (14%), 7.722 médias (61%) e 16.452 altas (25%).

Com o objetivo de avaliar o quão difícil é predizer a severidade, o *base score* e as métricas do vetor CVSS para as aplicações de comunicação[5] (Seção 4.3), com base em modelos treinados a partir de aplicações genéricas, as vulnerabilidades das aplicações de comunicação foram suprimidas dos *datasets* de treinamento/validação e de teste. Essas vulnerabilidades compuseram os *datasets* específicos de aplicações de comunicação, para

[5] *Facebook Messenger*, *Google Meet*, *Microsoft Teams*, *Skype*, *Cisco Webex*, *WhatsApp* e *Zoom*

serem usado como validação e teste.

Não foi possível encontrar informações sobre o tempo de correção de todas as vulnerabilidades de aplicações de comunicação. Diante disso, o *dataset* de aplicações de comunicação foi complementado com informações oriundas dos sites dos desenvolvedores. As vulnerabilidades referentes às aplicações de comunicação para as quais não foram encontradas informações sobre o tempo de correção foram suprmidas desse novo *dataset*. O *dataset* de tempo de correção de aplicações de comunicação foi dividido em dois subconjuntos: o *dataset* de treinamento/validação e o de teste. Nesse caso, o critério de segmentação também foi temporal. O *dataset* de treinamento/validação é composto por vulnerabilidades publicadas até março/2022 e possui 157 amostras. O *dataset* de teste compreende as vulnerabilidades não incluídas no *dataset* de treinamento/validação até o ano de 2022, possuindo 19 amostras.

A Figura 4.2 ilustra o esquema de derivação dos *datasets* apresentados nesta seção. Para isso, são considerados os *datasets* externos, como o do CVE e NVD, os *datasets* intermediários, compilados a partir dos *datasets* externos, mas ainda serão processados, e aqueles efetivamente usados nos experimentos deste capítulo.

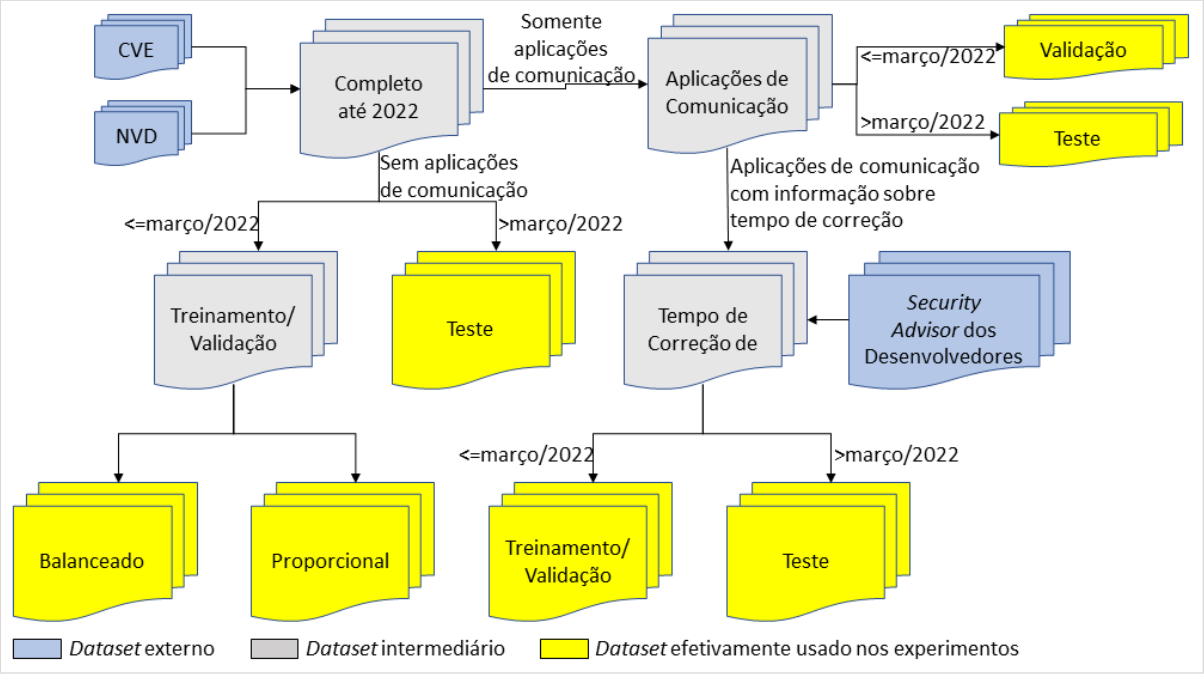


Figura 4.2: Esquema de derivação dos *datasets*

Extração de *Features*

Os classificadores treinados no âmbito deste estudo receberam como entrada as descrições textuais das vulnerabilidades. Para que essas descrições pudessem ser processadas

de maneira adequada pelos classificadores de ML, foram aplicados métodos de NLP. O processo de NLP empregado consistiu na remoção de números e símbolos presentes nas descrições, permanecendo apenas as letras do alfabeto (tanto maiúsculas quanto minúsculas). Posteriormente, as palavras foram convertidas para minúsculas e foi aplicada a técnica de lematização, cuja base é ignorar o tempo verbal, o gênero das palavras e suas formas plurais. Por fim, foram excluídas as palavras fechadas, tais como artigos, pronomes, preposições e conjunções, que não trazem informações relevantes para a análise das descrições das vulnerabilidades. O texto resultante do tratamento inicial da descrição é então processado, utilizando o *Term Frequency-Inverse Document Frequency (TF-IDF)* [7]. Assim, as features resultantes para uma certa vulnerabilidade são as informações binárias que indicam se unigramas, bigramas e trigramas (*i.e.*, sequências de uma, duas ou três palavras, denominadas, neste estudo, como termos) do corpus completo do *dataset* aparecem ou não na sua descrição textual.

Tendo em vista que nem sempre há uma correlação diretamente proporcional entre número de ocorrências de um determinado termo e sua relevância, foi usado escalonamento TF sublinear, que consiste na utilização do logaritmo da Frequência do Termo (TF), atribuindo um peso ao termo de acordo com a seguinte expressão: $(1 + \log TF)$. O objetivo é atribuir um peso menor aos termos mais frequentes e aumentar o peso de termos menos frequentes. Assim, é possível refletir com maior precisão a importância relativa de cada termo em relação ao conjunto de todas as vulnerabilidades. Também foi aplicada a normalização L2. Essa normalização consiste em igualar a um a soma dos quadrados dos valores de cada termo na descrição das vulnerabilidades. Isso permite que descrições de vulnerabilidades com tamanhos diferentes não influenciem em sua relevância.

Devido à elevada dimensionalidade do corpus, foram executados experimentos preliminares para avaliar o efeito da limitação do corpus, considerando o desempenho final dos algoritmos e os requisitos computacionais para executá-los. Em particular, a limitação do corpus que apresentou a melhor efetividade foi a dos 5.000 termos mais comuns. Portanto, este valor foi adotado para todos os experimentos reportados neste capítulo.

Além disso, foram testados o desempenho de dois algoritmos de redução de dimensionalidade, quais sejam, a *Linear Discriminant Analysis (LDA)* e redução de dimensionalidade *LDA on the Principal Component Analysis (PCA)*.

Métricas de Avaliação

Para avaliar as abordagens, vários algoritmos classificadores foram considerados, quais sejam: *Support Vector Machine* [42], *Random Forest* [10], *Naive Bayes for Multinomial*

Models [1], *Multi-layer Perceptron* [11], *Passive-Aggressive* [14] e *Logistic Regression* [19], conforme descrito na Seção 2.6.

Com o propósito de avaliar o desempenho dos classificadores que trabalham com classes discretas, as seguintes métricas foram utilizadas:

- **Acurácia**: para problemas de classificação em que existem múltiplas classes, representa percentual em que a predição das classes é correta, ou seja, é exatamente igual à real;
- ***Precision***: é a razão $\left(\frac{vp}{vp+fp}\right)$, onde vp é o número de verdadeiros positivos e fp o número de falsos positivos. *Precision* é a capacidade do classificador não rotular como positiva uma amostra que é negativa. Para problemas de classificação em que existem múltiplas classes, representa a média ponderada do *precision* de cada classe. O melhor valor é 1 e o pior valor é 0;
- ***Recall***: representa a razão $\left(\frac{vp}{vp+fn}\right)$, onde vp é o número de verdadeiros positivos e fn o número de falsos negativos. O *recall* é a habilidade do classificador em encontrar todas as amostras positivas. Para problemas de classificação em que existem múltiplas classes, representa a média ponderada do *recall* de cada classe. O melhor valor é 1 e o pior valor é 0; e
- ***F1 score***: é a média harmônica de *precision* e *recall*. No caso problemas de classificação em que existem múltiplas classes, representa a média ponderada do *F1 score* de cada classe. O melhor valor é 1 e o pior valor é 0.

Para as métricas *precision*, *recall* e *F1 score*, foram aplicadas três metodologias de cálculo:

- ***Micro***: calcula as métricas globalmente, contando o total de verdadeiros positivos, falsos negativos e falsos positivos. As métricas *precision*, *recall* e *F1 score* calculadas através desta metodologia resultaram em um valor exatamente igual a métrica acurácia. Assim, esta metodologia de cálculo não será citada no texto;
- ***Macro***: calcula as métricas para cada classe e encontra sua média não ponderada. Não levando em consideração o desequilíbrio entre as classes; e
- ***Weighted***: calcula as métricas para cada classe e encontra sua média ponderada pelo suporte (o número de instâncias verdadeiras para cada classe). Equivalente ao

'*Macro*', considerando o desequilíbrio entre as classes. O *recall* ponderado é igual à Acurácia. Utilizando esta metodologia de cálculo, apenas a métrica *precision* diferiu da métrica acurácia. Desse modo, *Weighted* será abordada apenas para a métrica *precision*.

Os modelos foram avaliados por meio da técnica de validação cruzada, utilizando o conjunto de dados de treinamento, com *k-fold* igual a dez. Para cada classificador cuja saída seja a severidade, o *base score* ou cada uma das métricas do vetor CVSS, foi calculada a diferença entre a menor e a maior acurácia, as médias das dez validações, com dados de vulnerabilidades publicadas até março/2022, e realizado um teste com dados de vulnerabilidades publicadas mais recentemente.

Como o *base score* é um valor numérico, para avaliar sua capacidade de predição por parte dos modelos, foi utilizada também a métrica de erro quadrático médio (*Mean Squared Error* - MSE). Essa métrica atribui um peso maior a erros maiores, uma vez que cada erro é elevado ao quadrado individualmente e, então, calcula-se a média desses erros quadráticos. Assim, seja $\hat{y}_i$ o valor predito da i-ésima amostra e y_i o valor verdadeiro correspondente, então o erro quadrático médio estimado sobre n amostras é definido pela Fórmula 4.4.

$$MSE(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2 \tag{4.4}$$

Uma desvantagem do MSE é a interpretabilidade. Isso ocorre porque a unidade de medida do erro é alterada ao elevar-se ao quadrado. Para resolver este problema, é possível aplicar a Fórmula 4.5, métrica denominada *Root Mean Squared Error* - RMSE.

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2} \tag{4.5}$$

Validação Cruzada

A validação cruzada é uma técnica largamente utilizada para avaliar a variância de um modelo [3]. Através dela, pode-se verificar a influência da aleatoriedade, ou seja, a capacidade de generalização de um modelo. Para melhor avaliar a variância e reduzir um possível viés, é recomendado utilizar um número de *folds* entre 5 e 10 [78]. Dessa forma, as avaliações conduzidas no âmbito deste estudo observam essa técnica com dez *folds*, como ilustrado na Figura 4.3. Nessa figura há uma representação didática da maneira como ocorre a divisão do *dataset* completo no subconjunto de treinamento e validação.

O *dataset* completo é dividido em dez partes, onde nove são usadas para treinamento e uma para validação, que varia cada *fold*. O teste é aplicado sobre o *fold* com melhor desempenho. Além das avaliações da variância (melhor e pior desempenho) e da média de desempenho dos dez *folds*, o classificador correspondente ao melhor dos dez *folds* foi usado para classificar o *dataset* de teste, como uma validação final do modelo.

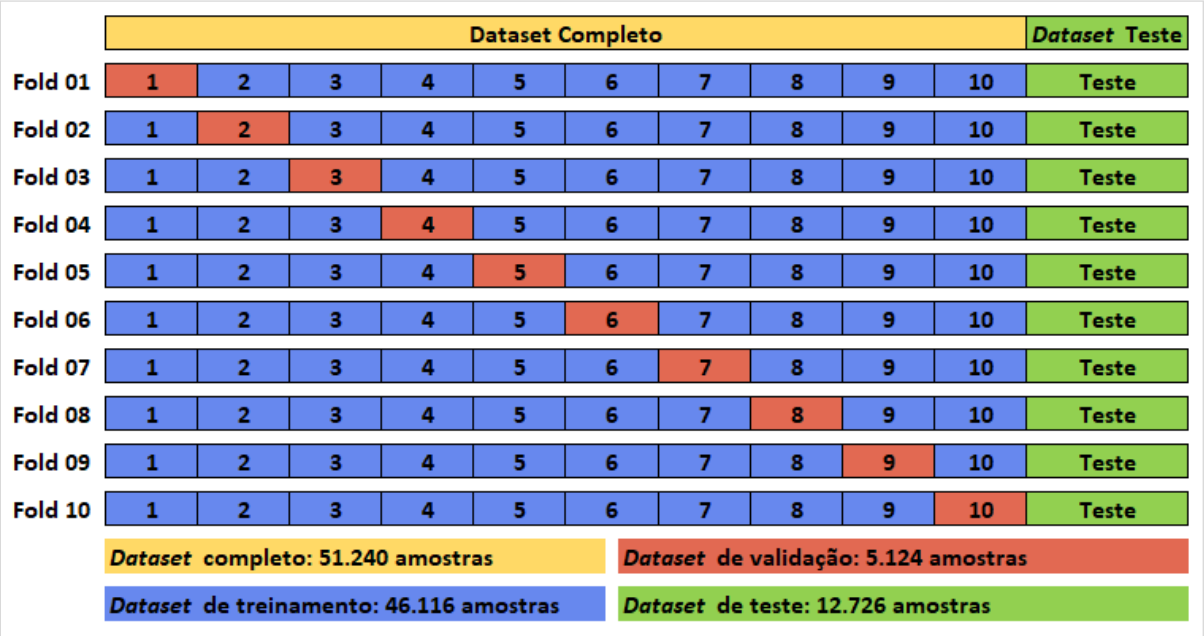


Figura 4.3: Validação cruzada com dez *folds* e teste

A divisão do *dataset* nos dez *folds* de treinamento e validação foi realizada através da função *Stratified K-Folds cross-validator*[6]. Essa função considera a proporcionalidade da distribuição de classes, ou a mais próxima possível, em relação ao *dataset* completo, ao efetivar a divisão entre os *folds*. Foi habilitado o parâmetro *shuffle*, de modo que, antes de dividir o *dataset* em *folds*, o algoritmo embaralha as amostras de cada classe. Diferente da função *Stratified ShuffleSplit cross-validator*[7], a *Stratified K-Folds cross-validator* garante que não haja sobreposição das amostras dos diferentes conjuntos de teste.

Em que pese a ilustração didática da Figura 4.3, a Figura 4.4 expõe a divisão efetiva entre treinamento e validação em cada um dos dez *folds*, tanto do *dataset* balanceado quanto do *dataset* proporcional. O que difere a divisão didática da efetiva, no que tange aos subconjuntos de treinamento e de validação, é que as amostras que efetivamente compõem esses subconjuntos não são, necessariamente, contíguas.

[6] `https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html`

[7] `https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedShuffleSplit.html`

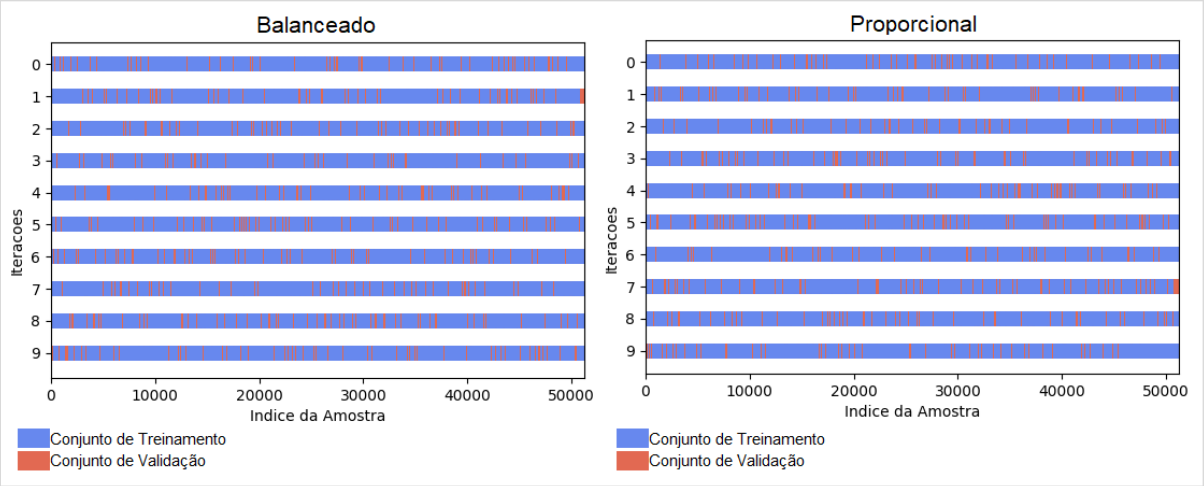


Figura 4.4: Comportamento da validação cruzada para os *datasets* balanceado e proporcional

4.2 Predição da Severidade

4.2.1 Abordagem 1 - A Partir da Descrição Textual da Vulnerabilidade

Metodologia de Predição

Nesta abordagem, foram implementados classificadores para fazer a predição da severidade diretamente a partir da descrição textual da vulnerabilidade. Eles são o produto da combinação entre os tipos de *datasets* e os algoritmos utilizados.

Avaliação dos Resultados

O desempenho da predição da severidade diretamente a partir da descrição textual da vulnerabilidade é ilustrado na Figura 4.5. A melhor média da acurácia das validações cruzadas foi de 78,6%. A *Precision Macro* e *Weighted* tiveram um resultado superior à acurácia em poucos centésimos de ponto percentual. No entanto, o *Recall* e o *F1 score*, ambos usando a metodologia de cálculo *Macro*, apresentaram, na média da validação cruzada, o resultado de 69%. Essa diferença se deve a dificuldade na predição de vulnerabilidades cujas severidades fossem altas e baixas. Dentre os cinco melhores resultados, avaliando a média de acurácia dos *folds* das validações cruzadas, o algoritmo *Random Forest* aparece na duas primeiras posições, em primeiro lugar com treinamento realizado a partir do *dataset* proporcional e em segundo lugar com o treinamento realizado a partir do *dataset* balanceado. Os demais três resultados, cada um com um algoritmo diferente, foram treinados a partir do *dataset* proporcional.

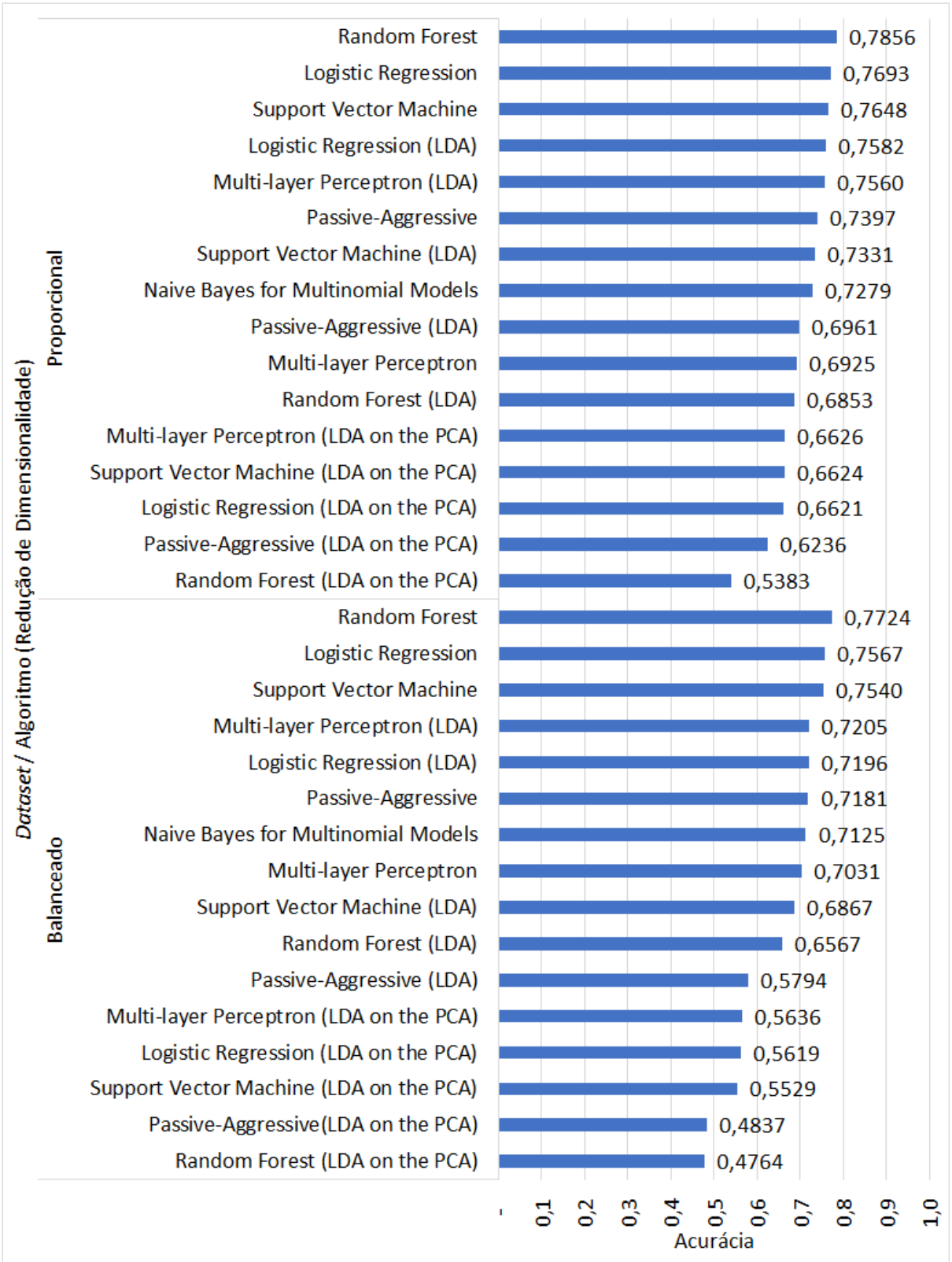


Figura 4.5: Médias de acurácia da validação cruzada da predição da severidade a partir da descrição textual da vulnerabilidade

O gráfico plotado na Figura 4.6 tem o objetivo de estimar a estabilidade e a capacidade de generalização do modelo. Nesse gráfico foi avaliada a amplitude no desempenho dos dez *folds*, bem como o desempenho da classificação do *dataset* de teste. O classificador que obteve o melhor resultado nos testes foi aquele treinado com o algoritmo *Random Forest* e o subconjunto, *fold*, F6 do *dataset* proporcional. Dentre os dez *folds* da validação cruzada, o pior desempenho aferido foi de 78% e o melhor de 79%. Dessa forma, a amplitude foi de, aproximadamente, 1 ponto percentual. Utilizando o classificador gerado no *fold* F6 na base de testes, foi obtido um desempenho de 72%, o que representa uma diferença de 7% para o *fold* com melhor desempenho na validação.

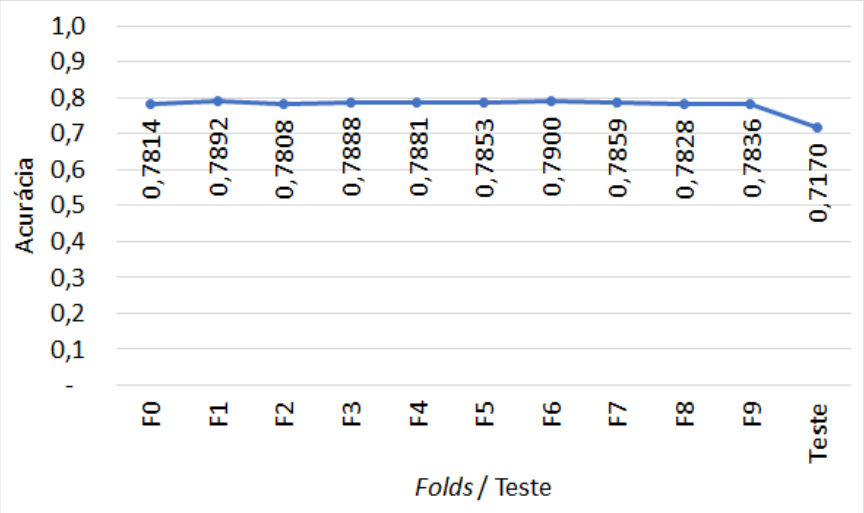


Figura 4.6: Desempenho do teste e de cada *fold* do classificador que apresentou a melhor média dos *folds* para predição da severidade a partir da descrição textual da vulnerabilidade (algoritmo *Random Forest*, *dataset* proporcional)

Ao considerar as dez melhores médias dos resultados das predições das severidades, diretamente a partir da descrição textual das vulnerabilidades, constata-se que a composição do *dataset* influencia o resultado final. Isso porque apenas três em dez desses melhores resultados utilizaram o *dataset* balanceado para treinamento do modelo. Nos demais 70%, os modelos foram treinados a partir do *dataset* proporcional. Chegamos à mesma conclusão ao analisar o desempenho das predições ao utilizar algoritmos de redução de dimensionalidade. O LDA está presente em 30% dos melhores dez resultados. Os 70% restantes não utilizam esse tipo de algoritmo. Entre os dez piores resultados, todos usam um dos dois algoritmos de redução de dimensionalidade abordados no presente estudo.

Com o objetivo de tentar refinar os resultados da previsão, os três melhores algoritmos (*Random Forest*, *Logistic Regression* e *Support Vector Machine*) foram combinados, utili-

zando o classificador *Voting*. A predição usando *Voting* foi feita no modo *hard*, utilizando os dois *datasets* para treinamento e aplicando a validação cruzada.

Conforme Figura 4.7, o algoritmo *voting* obteve uma acurácia (média dos dez *fold*) de 77% quando treinado a partir do *dataset* proporcional, e de 76%, quando treinado a partir do *dataset* balanceado. Sendo assim, o algoritmo *voting* não superou os resultados obtidos pelos modelos construídos a partir dos algoritmos de classificação de forma individual. A utilização do *dataset* balanceado implicou, em todos os casos, com o algoritmo *voting*, em uma acurácia inferior àquela obtida quando utilizado o *dataset* proporcional.

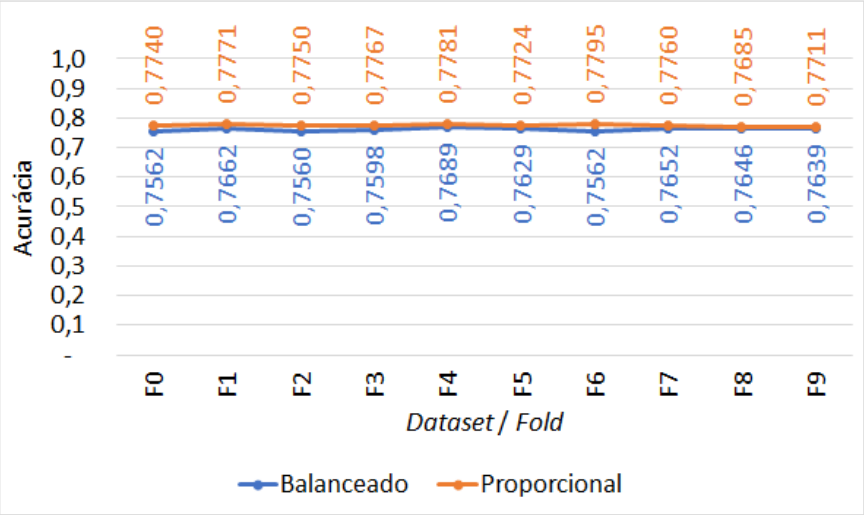


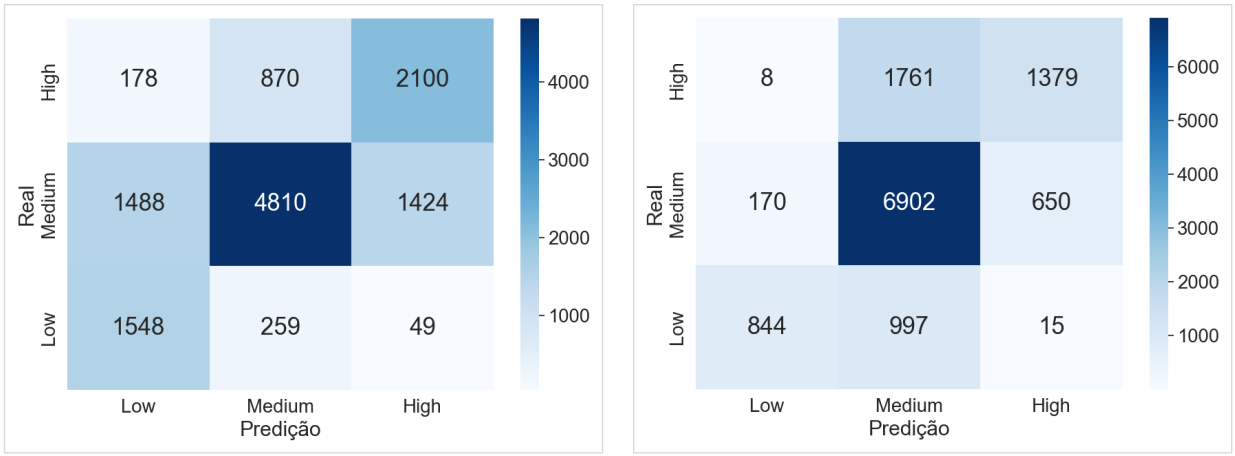
Figura 4.7: Desempenho dos *folds* da validação cruzada para predição da severidade a partir da descrição textual da vulnerabilidade utilizando o algoritmo *voting*

Com relação ao teste, analisando as matrizes de confusão (Fig. 4.8) conclui-se que o modelo treinado a partir do *dataset* proporcional encontrou maior dificuldade na predição de vulnerabilidades cujas severidades fossem altas e baixas. Os erros de classificação das vulnerabilidades cujas severidades são baixas e altas tendem a se concentrar em uma predição equivocada de severidade média. Já modelo treinado a partir do *dataset* balanceado encontrou maior dificuldade na predição de vulnerabilidades cujas severidades fossem médias. No entanto, os erros de classificação das vulnerabilidades cujas severidades são médias foram distribuídas equitativamente entre severidades baixas e altas.

4.2.2 Abordagem 2 - A Partir do *Base Score* Predito

Metodologia de Predição

Nesta abordagem, foram implementados classificadores para fazer a predição do *base score* diretamente a partir da descrição textual da vulnerabilidade, detalhados no Apên-



(a) Modelo gerado a partir do *dataset* balanceado

(b) Modelo gerado a partir do *dataset* proporcional

Figura 4.8: Matriz de confusão do teste de classificação da severidade a partir da descrição textual da vulnerabilidade

dice C. Com a obtenção do *base score*, procedeu-se à classificação da severidade.

Avaliação dos Resultados

Usando o *Base Score* predito diretamente a partir da descrição textual da vulnerabilidade, cujo detalhamento encontra-se no Apêndice C), foi aplicada uma função para obtenção da severidade. O desempenho na obtenção da severidade é ilustrado na Figura 4.9. A melhor média da acurácia das validações cruzadas foi de 78,1%. A média da *Precision Macro* foi de 77,6% e a *Precision Weighted* registrou um resultado inferior à acurácia em poucos centésimos de ponto percentual. Já, o *Recall* e o *F1 score*, ambos usando a metodologia de cálculo *Macro*, apresentaram, na média da validação cruzada, o resultado de 69,3%. A diferença de, aproximadamente, 8 pontos percentuais, decorre de um maior número de erros de classificação das vulnerabilidades cujas severidades são baixas e altas, em comparação com as vulnerabilidades cujas severidades são médias. Dentre os cinco melhores resultados, avaliando a média de acurácia dos *folds* das validações cruzadas, o algoritmo *Random Forest* aparece na primeira, terceira e quarta posições. Na primeira e na terceira posições, o modelo foi treinamento realizado a partir do *dataset* proporcional, sendo aplicado na terceira posição o algoritmo de redução de dimensionalidade LDA. O modelo que ocupa a quarta posição foi treinado a partir do *dataset* balanceado. Os demais dois resultados, cada um com um algoritmo diferente, foram treinados a partir do *dataset* proporcional.

O gráfico plotado na Figura 4.10 tem o objetivo de estimar a estabilidade e a capacidade de generalização do modelo. Nesse gráfico foi avaliada a amplitude no desempenho

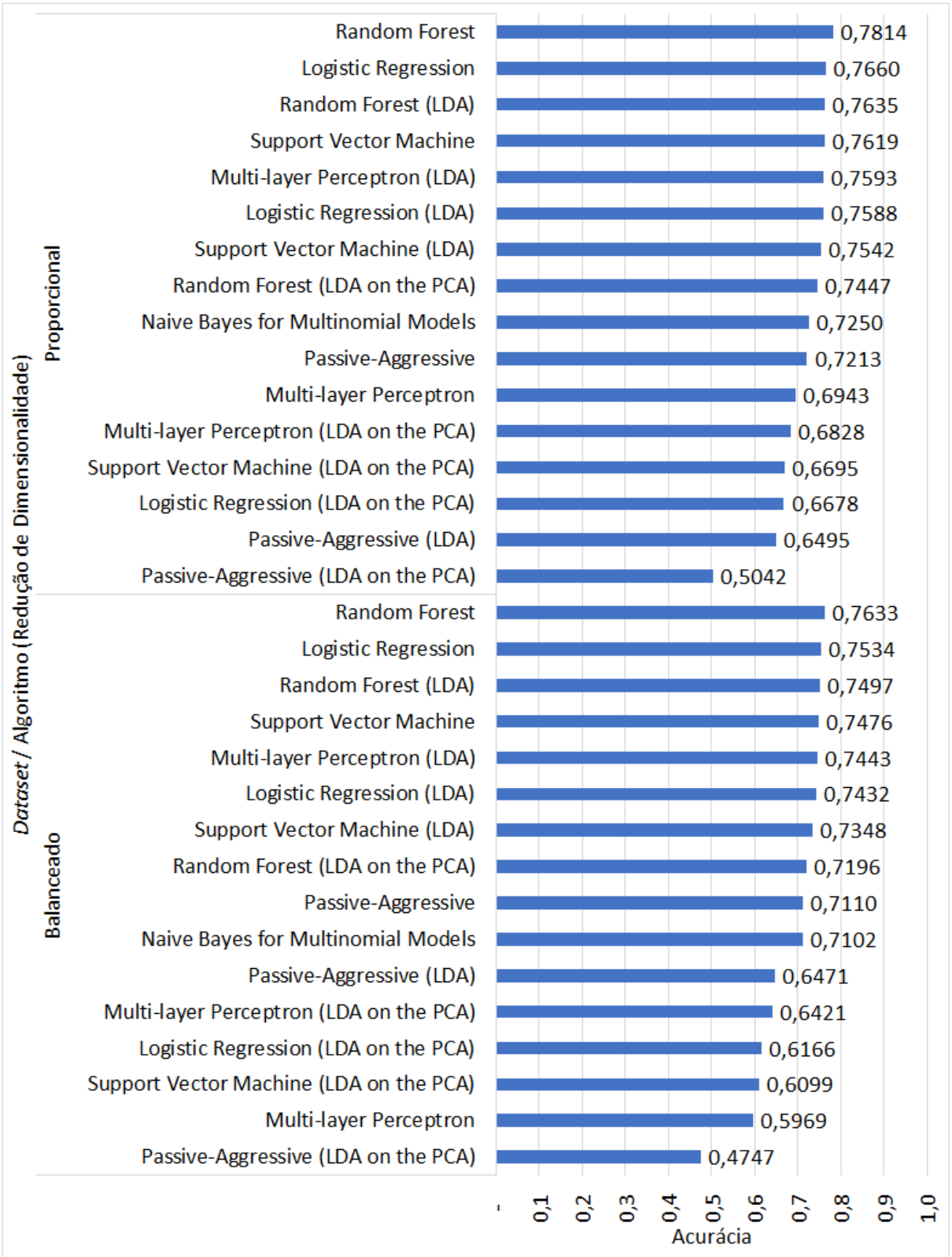


Figura 4.9: Médias de acurácia da validação cruzada da obtenção da severidade a partir do *base score* predito

dos dez *folds*, bem como o desempenho da classificação do *dataset* de teste. O classificador que obteve o melhor resultado nos testes foi aquele treinado com o algoritmo *Random Forest* e o subconjunto, *fold*, F1 do *dataset* proporcional. Dentre os dez *folds* da validação cruzada, o menor desempenho aferido foi de 77% e o maior de 79%. Dessa forma, a amplitude foi de, aproximadamente, 2 pontos percentuais. Já o desempenho verificado do teste foi de 72%, o que representa uma diferença de 7 pontos percentuais para o *fold* com melhor desempenho na validação.

Comparando a segunda abordagem, obtenção da severidade a partir do *base score* predito, com a primeira, predição da severidade a partir da descrição textual da vulnerabilidade, observa-se um desempenho muito similar. A diferença entre a melhor média da acurácia das duas abordagens é de 0,5 ponto percentual. Considerando, ainda, o classificador que apresentou a melhor média de acurácia, diferença de desempenho para o pior *fold* foi de um ponto percentual. Já o melhor *fold*, bem como o teste, registraram os mesmos resultados em ambas as abordagens.

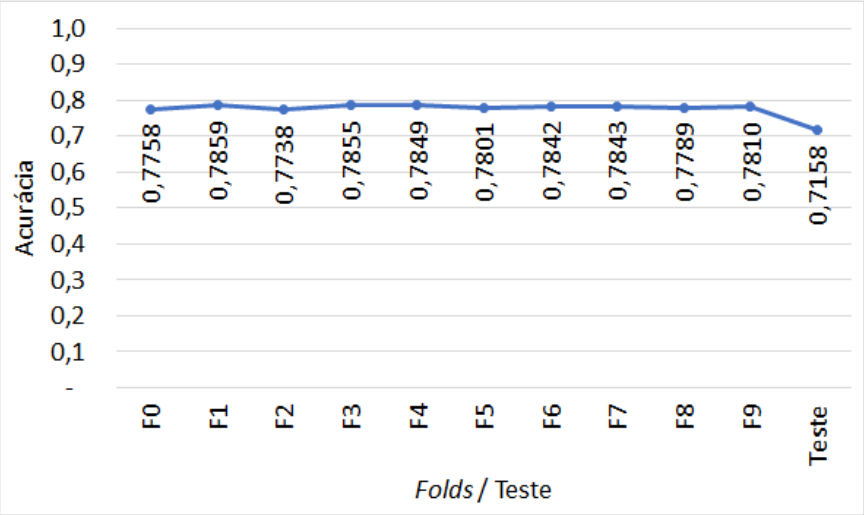


Figura 4.10: Desempenho do teste e de cada *fold* do classificador que apresentou a melhor média dos *folds* para obtenção da severidade a partir do *base score* predito (algoritmo *Random Forest*, *dataset* proporcional)

Ao considerar as dez melhores médias dos resultados da obtenção da severidade a partir do *base score* predito, constata-se que a composição do *dataset* influencia o resultado final. Isso porque apenas três em dez desses melhores resultados utilizaram o *dataset* balanceado para treinamento do modelo. Nos demais 70%, os modelos foram treinados a partir do *dataset* proporcional. O modelos que utilizam algoritmos de redução de dimensionalidade são cinco dos dez melhores resultados. Entre os dez piores resultados, apenas um não usa um dos dois algoritmos de redução de dimensionalidade abordados no

presente estudo.

Nesta abordagem o algoritmo *voting* registrou os mesmos resultados obtidos na primeira abordagem (Subseção 4.2.1), conforme Figura 4.11. Sendo assim, o algoritmo *voting* continuou não superando os resultados obtidos pelos modelos construídos a partir dos algoritmos de classificação de forma individual.

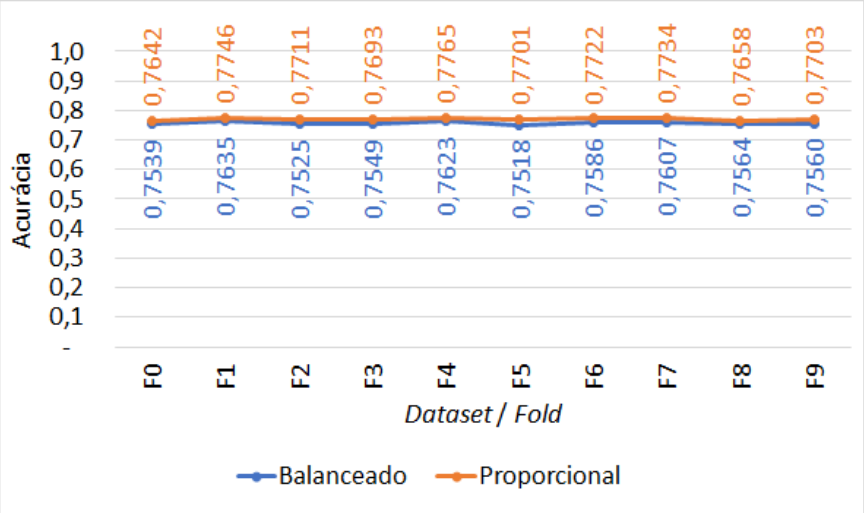
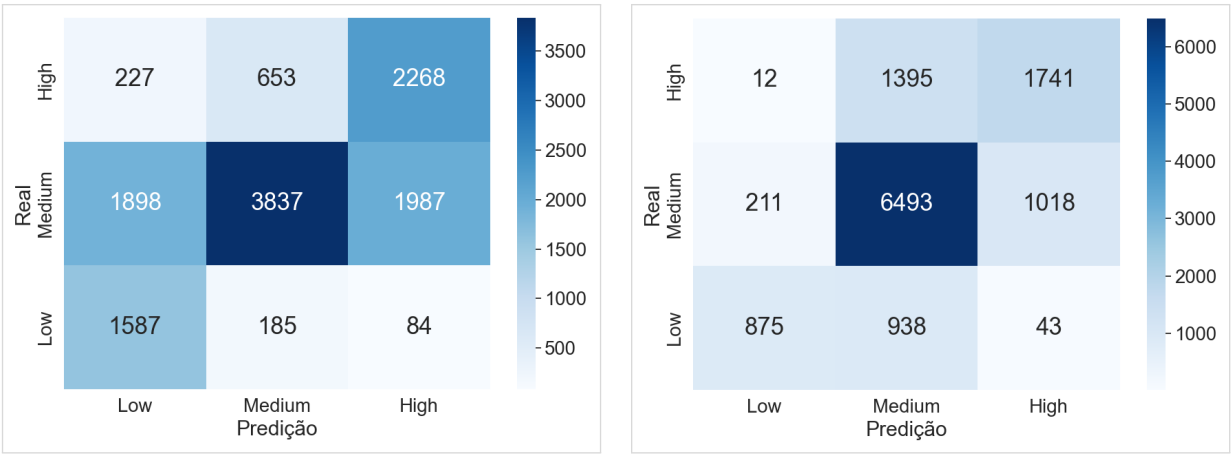


Figura 4.11: Desempenho dos *folds* da validação cruzada para obtenção da severidade a partir do *base score* predito utilizando o algoritmo *voting*

Analisando as matrizes de confusão do resultado do teste (Fig. 4.12), observa-se, no teste realizado a partir do modelo treinado com o *dataset* proporcional, uma maior dificuldade na predição de vulnerabilidades cujas severidades fossem altas e baixas. Os erros de classificação das vulnerabilidades cujas severidades são baixas e altas tendem a se concentrar em uma predição equivocada de severidade média. Já modelo treinado a partir do *dataset* balanceado encontrou maior dificuldade na predição de vulnerabilidades cujas severidades fossem médias. No entanto, os erros de classificação das vulnerabilidades cujas severidades são médias foram distribuídas equitativamente entre severidades baixas e altas.

Nesta abordagem, a severidade da vulnerabilidade é obtida em função do *Base Score* predito diretamente a partir da descrição textual da vulnerabilidade. Faz-se mister ressaltar que por exigência de alguns algoritmos de classificação, nesta abordagem, o *Base Score* foi truncado, resultando em classes inteiras de 0 a 10. Face ao exposto, com o objetivo de apresentar de forma sucinta uma noção da qualidade do classificador, cujos detalhes encontram-se no Apêndice C), a Figura 4.13 mostra a aplicação da métrica MSE para avaliar se as predições do incorretas do *Base Score* estão muito distantes do valor real. Com exceção do pior resultado, o MSE obtido não é alto. Isso significa que as predições



(a) Modelo gerado a partir do *dataset* balanceado

(b) Modelo gerado a partir do *dataset* proporcional

Figura 4.12: Matriz de confusão da obtenção da severidade a partir do *base score* predito (*dataset* de teste)

incorretas não são excessivamente discrepantes do valor real. Também é possível constatar que, embora o modelo treinado a partir do *dataset* balanceado tenha apresentado melhor acurácia, o modelo treinado a partir do *dataset* proporcional apresentou um MSE mais baixo. Ao aplicar a métrica RMSE ao melhor resultado de MSE, modelo treinado com algoritmo *Random Forest* e *dataset* proporcional, obtemos o resultado de 1,4. Assim, em média, o erro na predição do *base score* é de 1,4 para mais ou para menos.

4.2.3 Abordagem 3 - A Partir dos Valores Preditos das Métricas do Vetor CVSS

Metodologia de Predição

Nesta abordagem, foram implementados classificadores para fazer a predição de cada uma das métricas do vetor CVSS a partir da descrição textual da vulnerabilidade e, posteriormente, o cálculo do *base score*, conforme detalhado no Apêndice D. Sobre o *base score* calculado, aplica-se uma função para obtenção da severidade.

Avaliação dos Resultados

Realizado o cálculo do *Base Score* a partir das métricas do vetor CVSS, cujo detalhamento encontra-se no Apêndice D, foi aplicada uma função para obtenção da severidade. O desempenho na obtenção da severidade é ilustrado na Figura 4.14. A melhor média da acurácia das validações cruzadas foi de 78%. A média da *Precision Macro* foi de 76,1% e a *Precision Weighted* registrou um resultado inferior à acurácia em 0,2 ponto percentual.

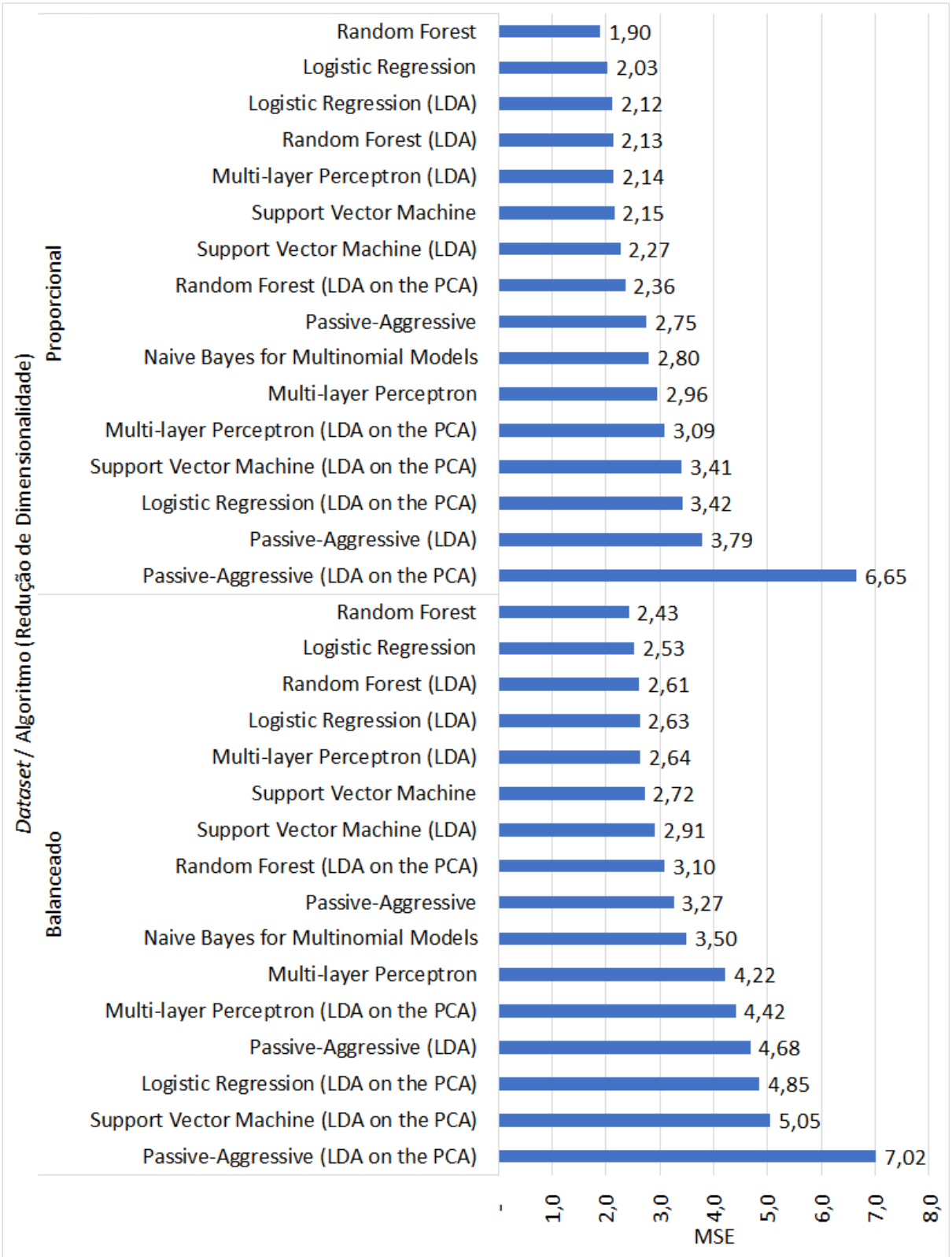


Figura 4.13: Médias do MSE na validação cruzada da predição do *base score* a partir da descrição textual da vulnerabilidade

Já, o *Recall* e o *F1 score*, ambos usando a metodologia de cálculo *Macro*, apresentaram, na média da validação cruzada, o resultado de 71,3%. Nesse caso, o desempenho mais baixo do *Recall Macro* e do *F1 score Macro* são consequência da maior dificuldade do algoritmo em classificar corretamente as vulnerabilidades cujas severidades são baixas. Dentre os cinco melhores resultados, avaliando a média de acurácia dos *folds* das validações cruzadas, o algoritmo *Random Forest* aparece na primeira posição e na quarta posições. O modelo que ocupa a primeira posição foi treinado a partir do *dataset* proporcional e o que ocupa a quarta posição treinado a partir do *dataset* balanceado. O algoritmo *Logistic Regression*, treinado com o *dataset* proporcional registrou o segundo melhor desempenho. O algoritmo *Support Vector Machine* também aparece duas vezes entre os cinco melhores modelos, na terceira posição treinado a partir do *dataset* proporcional e na quinta posição treinado a partir do *dataset* balanceado.

O gráfico plotado na Figura 4.15 tem o objetivo de estimar a estabilidade e a capacidade de generalização do modelo. Nesse gráfico foi avaliada a amplitude no desempenho dos dez *folds*, bem como o desempenho da classificação do *dataset* de teste. O classificador que obteve o melhor resultado nos testes foi aqule treinado com o algoritmo *Random Forest* e o subconjunto, *fold*, F6 do *dataset* proporcional. Dentre os dez *folds* da validação cruzada, o menor desempenho aferido foi de 77,4% e o maior de 78,5%. Dessa forma, a amplitude foi de, aproximadamente, 1 ponto percentual. Já o desempenho verificado do teste foi de 71,5%, o que representa uma diferença de 7 pontos percentuais para o *fold* com melhor desempenho na validação, utilizado para classificar o *dataset* de teste.

Ao confrontar os resultados alcançados a partir das três abordagens, em que pese a metodologia de obtenção da severidade seja muito diferente, os resultados são muito próximos. Nesse ponto, foram analisadas as principais medidas de desempenho, quais sejam melhor média da acurácia das validações cruzadas, acurácia do pior e do melhor *fold* do classificador que registrou a melhor média da acurácia das validações cruzadas e acurácia do teste. Dentre todas essas medidas, a maior diferença entre as abordagens foi de um ponto percentual. Dessa forma, o critério de seleção entre uma ou outra abordagem deve, mormente, considerar outros fatores em detrimento do desempenho.

Ao considerar as dez melhores médias dos resultados da obtenção da severidade a partir do *base score* predito, observa-se que 60% dos modelos utilizaram o *dataset* proporcional e 40% dos modelos utilizaram o *dataset* balanceado. Apenas dois modelos, oitava e décima posição, utilizaram o algoritmo de redução de dimensionalidade LDA. Entre os dez piores resultados, todos usaram um dos dois algoritmos de redução de dimensionalidade

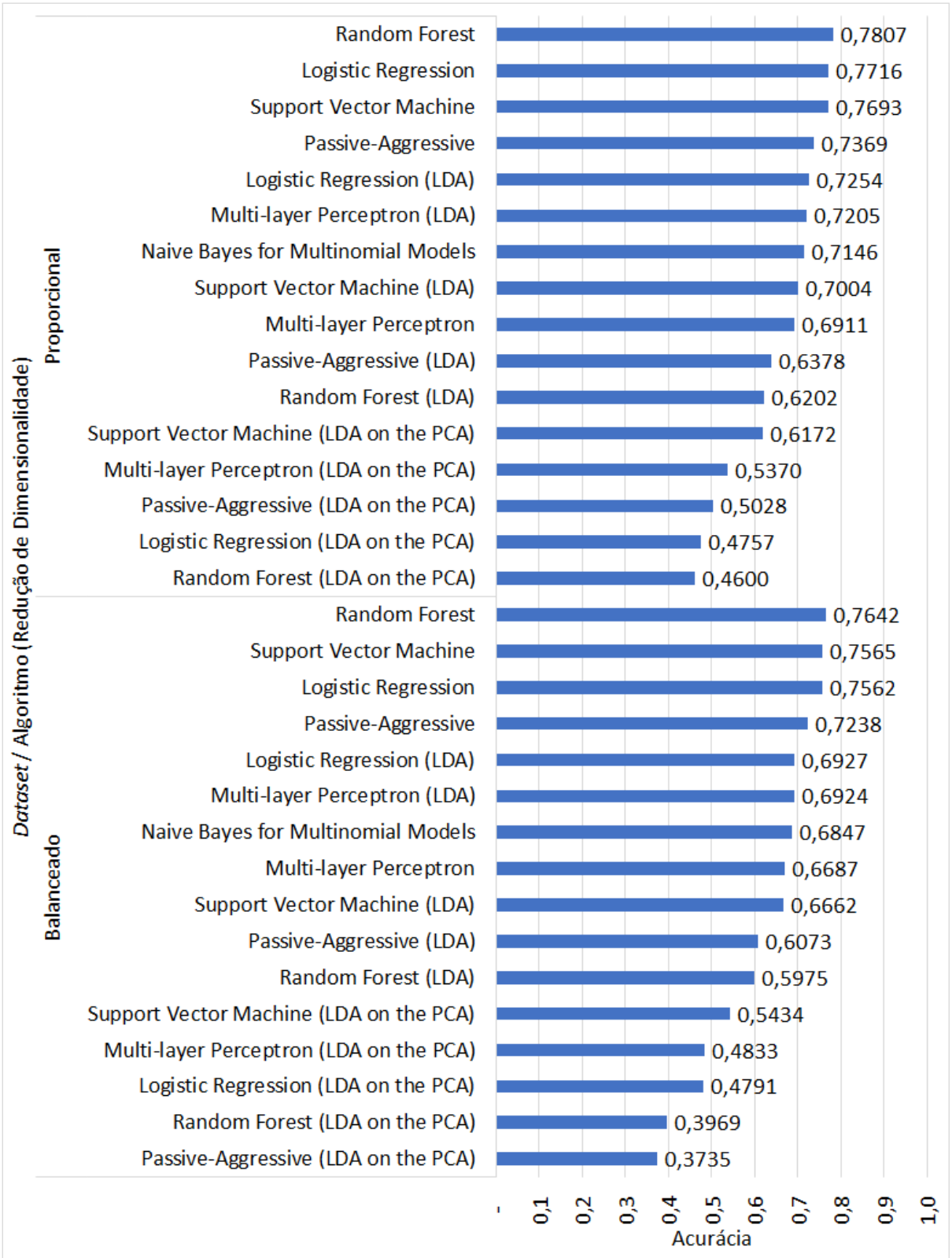


Figura 4.14: Médias de acurácia da validação cruzada da obtenção da severidade em função do *base score* calculado a partir dos valores preditos das métricas do vetor CVSS

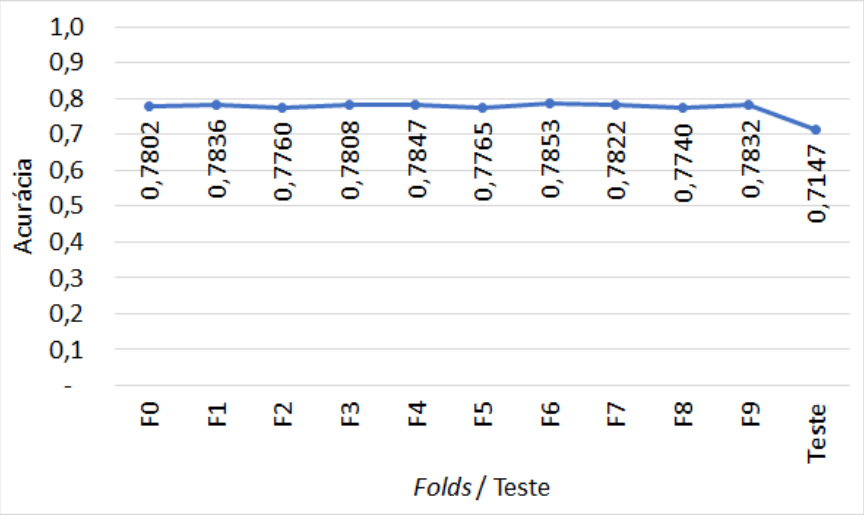


Figura 4.15: Desempenho do teste e de cada *fold* do classificador que apresentou a melhor média dos *folds* para obtenção da severidade em função do *base score* calculado a partir dos valores preditos das métricas do vetor CVSS (algoritmo *Random Forest*, *dataset* proporcional)

abordados no presente estudo.

Conforme Figura 4.16, o algoritmo *voting* obteve a acurácia (média dos dez *folds*), de 78% quando treinado a partir do *dataset* proporcional, e de 76%, quando treinado a partir do *dataset* balanceado. Sendo assim, o algoritmo *voting* não superou os resultados obtidos pelos modelos construídos a partir dos algoritmos de classificação de forma individual. A utilização do *dataset* balanceado implicou, em todos os casos, com o algoritmo *voting*, em uma acurácia inferior àquela obtida quando utilizado o *dataset* proporcional.

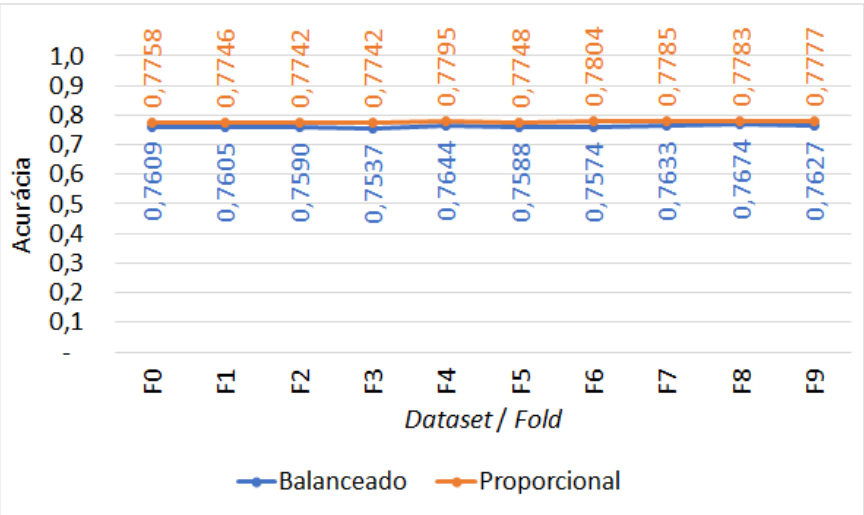
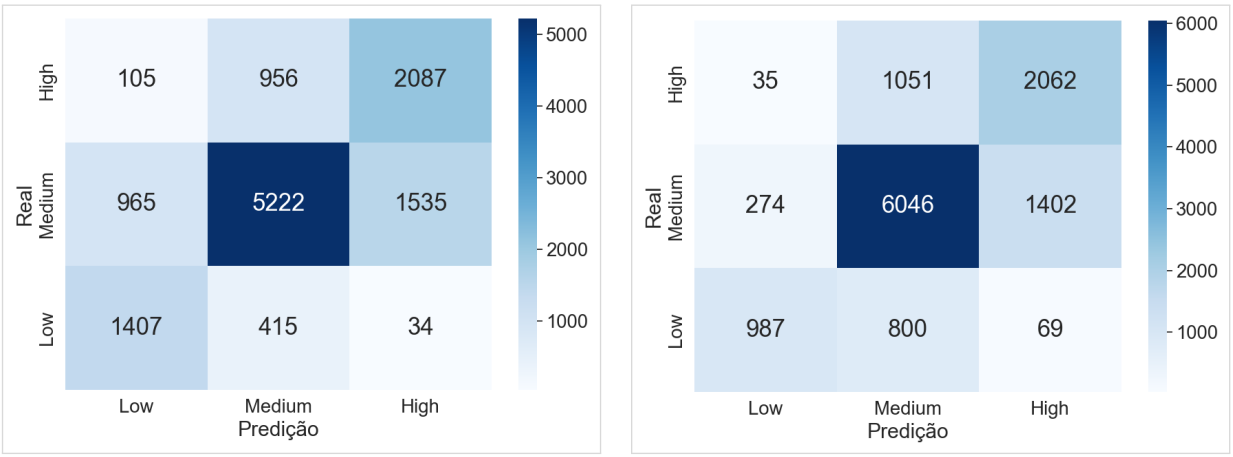


Figura 4.16: Desempenho dos *folds* da validação cruzada para obtenção da severidade em função do *base score* calculado a partir dos valores preditos das métricas do vetor CVSS utilizando o algoritmo *voting*

Com relação ao teste, comparando as matrizes de confusão (Fig. 4.17) observa-se que, no que tange à classe *LOW*, o modelo treinado com o *dataset* balanceado teve um desempenho, acurácia de 76%, muito superior ao registrado pelo modelo treinado com o *dataset* proporcional, acurácia de 53%. Esse comportamento foi o inverso da classe *MEDIUM*, onde o modelo treinado com o *dataset* balanceado teve um desempenho, acurácia de 68%, muito inferior ao registrado pelo modelo treinado com o *dataset* proporcional, acurácia de 78%. A classe *HIGH* registrou um desempenho similar com modelos treinados a partir de ambos os *datasets*, com acurácia de 66,3% no balanceado e acurácia de 65,5% no proporcional.



(a) Modelo gerado a partir do *dataset* balanceado

(b) Modelo gerado a partir do *dataset* proporcional

Figura 4.17: Matriz de confusão da obtenção da severidade em função do *base score* calculado a partir dos valores preditos das métricas do vetor CVSS (*dataset* de teste)

Nesta abordagem, a severidade da vulnerabilidade é obtida em função do *Base Score* calculado a partir dos valores preditos das métricas do vetor CVSS, as quais foram preditas diretamente a partir da descrição textual da vulnerabilidade. Diferente da Abordagem 2 (Subseção 4.2.2), nesta abordagem, o *Base Score* foi arredondado à primeira casa decimal. Considerando essas informações, com o propósito de apresentar uma breve ideia da qualidade do classificador, cujos detalhes encontram-se no Apêndice D, a Figura 4.18 mostra a aplicação da métrica MSE para avaliar se os resultados incorretos do cálculo do *Base Score* estão muito distantes do valor real. Entre os dez melhores MSE, o máximo medido foi 2,66. Isso significa que as predições incorretas não excessivamente discrepantes do valor real. Também é possível constatar que o modelo treinado a partir do *dataset* proporcional apresentou um MSE mais baixo em relação ao modelo treinado a partir do *dataset* balanceado. Ao aplicar a métrica RMSE sobre o melhor resultado de MSE, modelo treinado com algoritmo *Random Forest* e *dataset* proporcional, o resultado é 1,4.

Assim, em média, o erro na predição do *base score* é de 1,4 para mais ou para menos.

4.3 Predição para as aplicações de comunicação

4.3.1 Metodologia de Predição

As vulnerabilidades das aplicações de comunicação foram submetidas à classificação por meio dos modelos construídos e explicados nas Subseções 4.2.1, 4.2.2 e 4.2.3. Na avaliação desta seção, não foram usados os modelos criados a partir de algoritmos de redução de dimensionalidade, em virtude do inexpressivo desempenho nas demais avaliações. Ressalta-se que o modelo gerado em cada um dos dez *folds* da validação cruzada de aplicações genéricas foi utilizado para classificar todo o *dataset* de validação das aplicações de comunicação. O *dataset* de teste das aplicações de comunicação foi classificado com o mesmo classificador utilizado no teste de aplicações genéricas, ou seja, o melhor *fold* do classificador que obteve a melhor média de *folds*.

Extração de Dados

Com o objetivo de avaliar o quão difícil é predizer a vulnerabilidade de uma aplicação de comunicação a partir de modelos treinados com vulnerabilidades de aplicações que têm outras finalidades, os *datasets* não incluíram dados acerca de vulnerabilidades de aplicações de comunicação. Com a finalidade de avaliar a estabilidade do modelo e verificar se essa estabilidade se mantém ao longo do tempo, as vulnerabilidades de aplicações de comunicação foram distribuídas por dois *datasets* específicos desse tipo aplicação. O primeiro *dataset*, de validação, possui 188 amostras, sendo 15 vulnerabilidades de severidade baixa (8%), 105 médias (56%) e 68 altas (36%). O segundo *dataset*, de teste, conta com dezenove amostras.

4.3.2 Avaliação dos Resultados

Nas Subseções 4.2.1, 4.2.2 e 4.2.3, diversas informações são preditas ou computadas a partir destas predições. Nesta subseção, há um comparativo detalhado dos resultados. Quando os resultados da validação cruzada referente às aplicações genéricas, de emprego geral, são comparados aos resultados referentes às aplicações de comunicação (Fig. 4.19), todos os resultados obtidos na classificação do *dataset* de validação das aplicações de comunicação, a partir dos modelos treinados com *datasets* de aplicações genéricas, são superiores àqueles obtidos na classificação do próprio *dataset* de aplicações genéricas. A

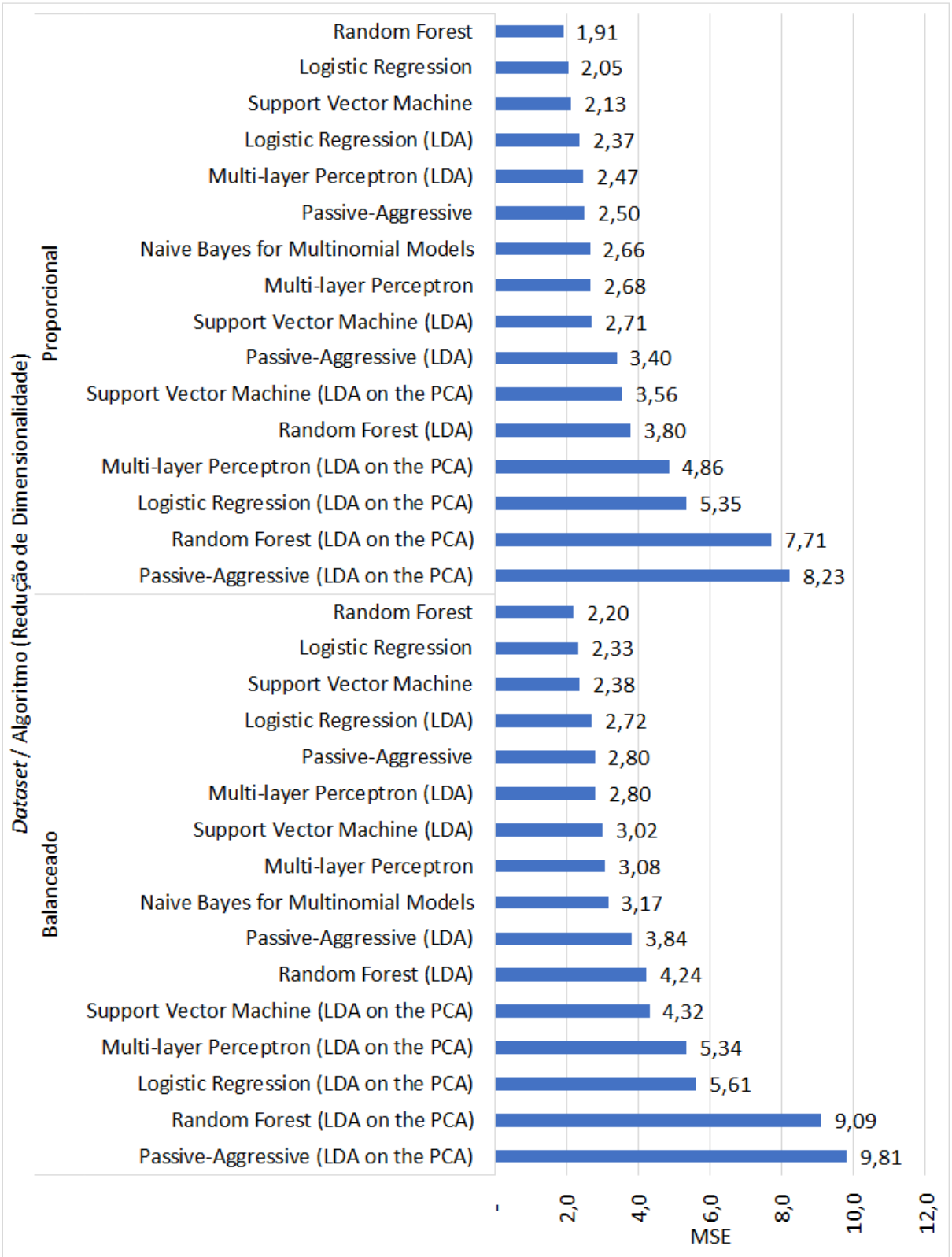


Figura 4.18: Médias do MSE na validação cruzada do *base score* calculado a partir dos valores preditos das métricas do vetor CVSS

única exceção foi a predição da métrica *Integrity* do Vetor CVSS. No entanto, a diferença foi inferior a 1 ponto percentual.

Ainda com base na Figura 4.19, encontram-se as informações acerca do comparativo entre os resultados dos testes referente às aplicações genéricas, de emprego geral, e os resultados dos testes referentes às aplicações de comunicação. Diferente da validação cruzada, nos testes nem sempre os resultados das classificações das aplicações de comunicação foram superiores. Porém, nas três abordagens para obtenção da severidade, a classificação do *dataset* de teste das aplicações de comunicação registraram resultados superiores. O destaque foi para a predição da severidade a partir da descrição textual da vulnerabilidade, com acurácia de 89% no teste de aplicações de comunicação. A Tabela 4.3 oferece uma visão detalhada, incluindo o tipo de *dataset* e algoritmo utilizados.

Outro fato que merece destaque é que o *base score* calculado a partir dos valores preditos das métricas do vetor CVSS e o *base score* predito a partir da descrição textual da vulnerabilidade, no teste a partir do *dataset* de aplicações de comunicação, tiveram um desempenho muito ruim e ficaram abaixo do resultado alcançado no teste a partir do *dataset* de aplicações genéricas, sendo a diferença 18,5 e 19,7 pontos percentuais respectivamente. No entanto, o MSE (Fig. 4.20) ao classificar o *dataset* de aplicações de comunicação é sempre menor do que ao classificar o *dataset* de aplicações genéricas, seja na validação cruzada ou no teste. Aplicando a raiz quadrada sobre o MSE, obtendo o RMSE, chega-se a uma margem de erro 1,4 no *base score*, para mais ou para menos, no *base score* calculado a partir dos valores preditos das métricas do vetor CVSS e no *base score* predito a partir da descrição textual da vulnerabilidade, tanto para o teste quanto para a validação cruzada a partir do *dataset* de aplicações de comunicação.

4.3.3 Síntese dos Resultados das Abordagens

Nesta seção foram exploradas três abordagens para obtenção da severidade de uma vulnerabilidade. Na primeira abordagem (Subseção 4.2.1), a severidade é a saída do modelo de predição. Mas apenas a severidade é entregue ao usuário. A segunda abordagem (Subseção 4.2.2) consiste em realizar a predição do *base score* e aplicar uma função para obter a severidade. Nesse caso, além da severidade, o algoritmo também retorna o *base score*, embora truncado. Na terceira abordagem (Subseção 4.2.3), a mais completa, é realizada a predição de cada uma das métricas do vetor CVSS, a partir delas é calculado o *base score* com uma casa decimal e, por fim, em função do *base score* é obtida a severidade.

[8]Detalhamento da Figura 4.19 encontra-se nas Tabelas 4.2 e 4.3.

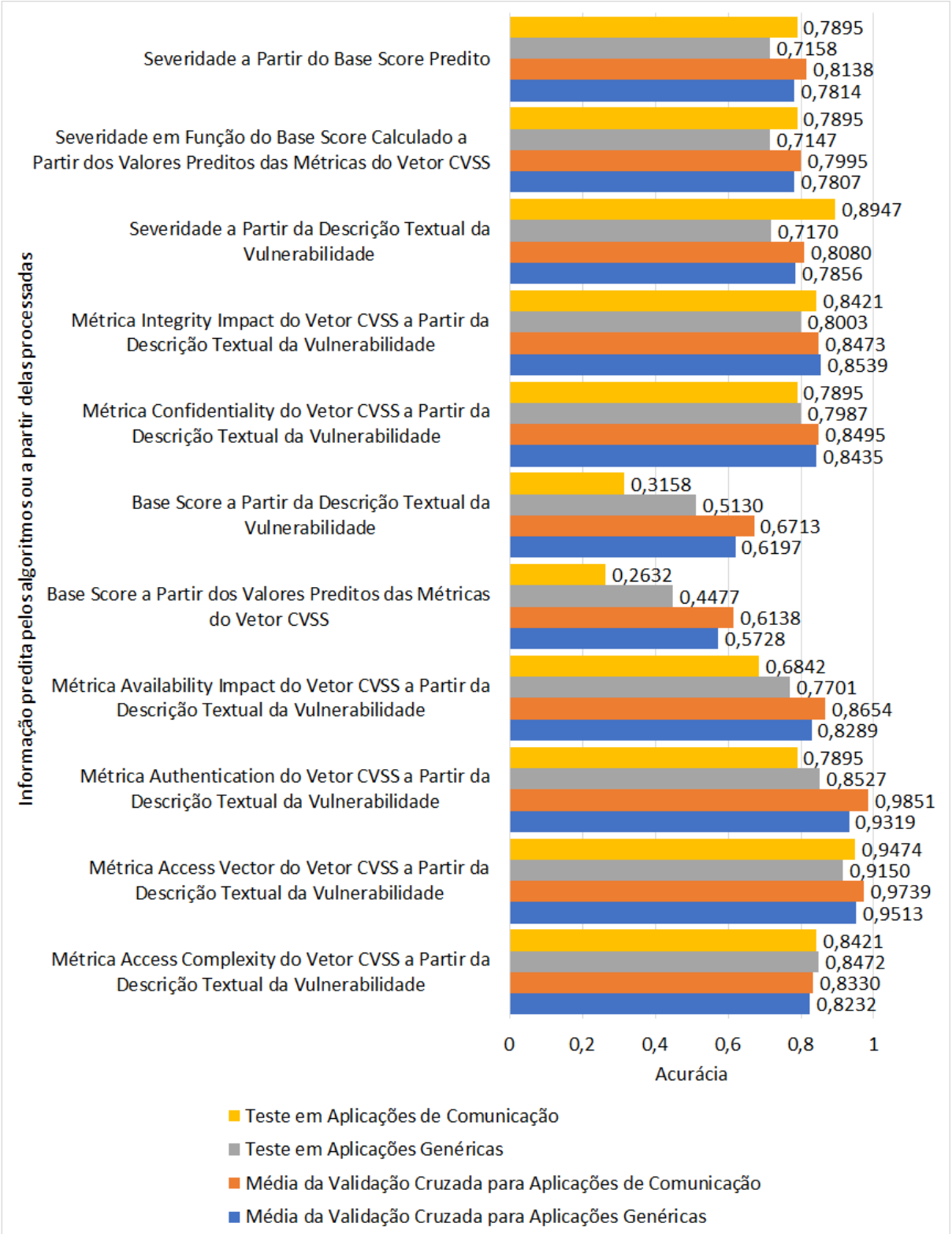


Figura 4.19: Comparativo das melhores médias de acurácia da validação cruzada e das melhores acurácias dos testes entre as aplicações genéricas as aplicações de comunicação[8]

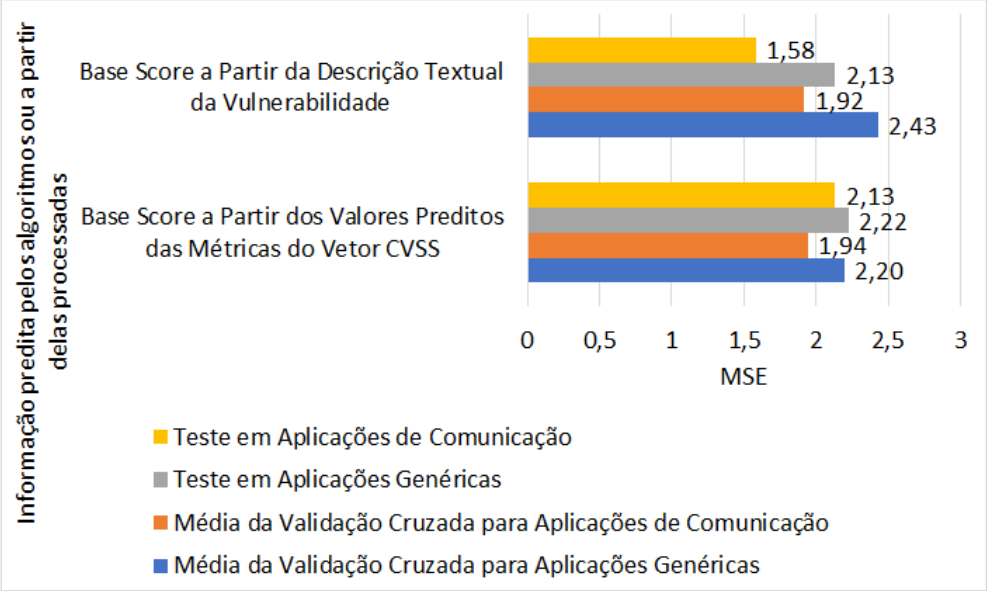


Figura 4.20: Comparativo das médias de MSE da validação cruzada e do teste de aplicações genéricas com as médias de MSE da validação cruzada e do teste de aplicações de comunicação

Tabela 4.2: Tabela dos melhores desempenhos para aplicações genéricas

Informação predita pelos algoritmos ou a partir delas processadas	**Validação Cruzada**		**Teste**	
	Dataset - Algoritmo	**Média de Acurácia dos Folds**	**Dataset - Algoritmo - Melhor Fold**	**Acurácia do Melhor Fold**
Métrica Access Complexity do Vetor CVSS a Partir da Descrição Textual da Vulnerabilidade	Proporcional - Random Forest	0,8232	Proporcional - Random Forest - F3	0,8472
Métrica Access Vector do Vetor CVSS a Partir da Descrição Textual da Vulnerabilidade	Proporcional - Support Vector Machine	0,9513	Proporcional - Support Vector Machine - F8	0,9150
Métrica Authentication do Vetor CVSS a Partir da Descrição Textual da Vulnerabilidade	Balanceado - Random Forest	0,9319	Proporcional - Logistic Regression - F4	0,8527
Métrica Availability Impact do Vetor CVSS a Partir da Descrição Textual da Vulnerabilidade	Balanceado - Random Forest	0,8289	Proporcional - Random Forest - F9	0,7701
Base Score a Partir dos Valores Preditos das Métricas do Vetor CVSS	Balanceado - Random Forest	0,5728	Proporcional - Random Forest - F3	0,4477
Base Score a Partir da Descrição Textual da Vulnerabilidade	Balanceado - Random Forest	0,6197	Proporcional - Random Forest - F3	0,5130
Métrica Confidentiality do Vetor CVSS a Partir da Descrição Textual da Vulnerabilidade	Balanceado - Random Forest	0,8435	Proporcional - Random Forest - F8	0,7987
Métrica Integrity Impact do Vetor CVSS a Partir da Descrição Textual da Vulnerabilidade	Balanceado - Random Forest	0,8539	Proporcional - Random Forest - F8	0,8003
Severidade a Partir da Descrição Textual da Vulnerabilidade	Proporcional - Random Forest	0,7856	Proporcional - Random Forest - F6	0,7170
Severidade em Função do Base Score Calculado a Partir dos Valores Preditos das Métricas do Vetor CVSS	Proporcional - Random Forest	0,7807	Proporcional - Random Forest - F6	0,7147
Severidade a Partir do Base Score Predito	Proporcional - Random Forest	0,7814	Proporcional - Random Forest - F1	0,7158

Tabela 4.3: Tabela dos melhores desempenhos para aplicações de comunicação

Informação predita pelos algoritmos ou a partir delas processadas	**Validação Cruazada**		**Teste**	
	Dataset - Algoritmo	**Média de Acurácia dos Folds**	**Dataset - Algoritmo - Melhor Fold**	**Acurácia do Melhor Fold**
Métrica Access Complexity do Vetor CVSS a Partir da Descrição Textual da Vulnerabilidade	Balanceado - Logistic Regression	0,8330	Balanceado - Logistic Regression - F2	0,8421
Métrica Access Vector do Vetor CVSS a Partir da Descrição Textual da Vulnerabilidade	Proporcional - Logistic Regression	0,9739	Proporcional - Support Vector Machine - F3	0,9474
Métrica Authentication do Vetor CVSS a Partir da Descrição Textual da Vulnerabilidade	Balanceado - Support Vector Machine	0,9851	Proporcional - Passive-Aggressive - F7	0,7895
Métrica Availability Impact do Vetor CVSS a Partir da Descrição Textual da Vulnerabilidade	Balanceado - Random Forest	0,8654	Balanceado - Random Forest - F8	0,6842
Base Score a Partir dos Valores Preditos das Métricas do Vetor CVSS	Proporcional - Logistic Regression	0,6138	Balanceado - Logistic Regression - F2	0,2632
Base Score a Partir da Descrição Textual da Vulnerabilidade	Proporcional - Random Forest	0,6713	Proporcional - Random Forest - F5	0,3158
Métrica Confidentiality do Vetor CVSS a Partir da Descrição Textual da Vulnerabilidade	Proporcional - Random Forest	0,8495	Proporcional - Random Forest - F3	0,7895
Métrica Integrity Impact do Vetor CVSS a Partir da Descrição Textual da Vulnerabilidade	Balanceado - Random Forest	0,8473	Proporcional - Random Forest - F4	0,8421
Severidade a Partir da Descrição Textual da Vulnerabilidade	Proporcional - Random Forest	0,8080	Proporcional - Random Forest - F6	0,8947
Severidade em Função do Base Score Calculado a Partir dos Valores Preditos das Métricas do Vetor CVSS	Proporcional - Logistic Regression	0,7995	Balanceado - Random Forest - F5	0,7895
Severidade a Partir do Base Score Predito	Proporcional - Random Forest	0,8138	Proporcional - Random Forest - F5	0,7895

O gráfico ilustrado na Figura 4.21 demonstra que os dois melhores modelos de cada uma das três abordagens para obtenção da severidade têm desempenho muito semelhantes. Assim, é essencial analisar os aspectos positivos e negativos de cada uma das metodologias. Isso significa que a escolha da abordagem e do algoritmo a serem usados pode, em detrimento do desempenho, ser baseada no detalhamento dos resultados gerados por cada abordagem/algoritmo ou nos recursos computacionais por eles consumidos.

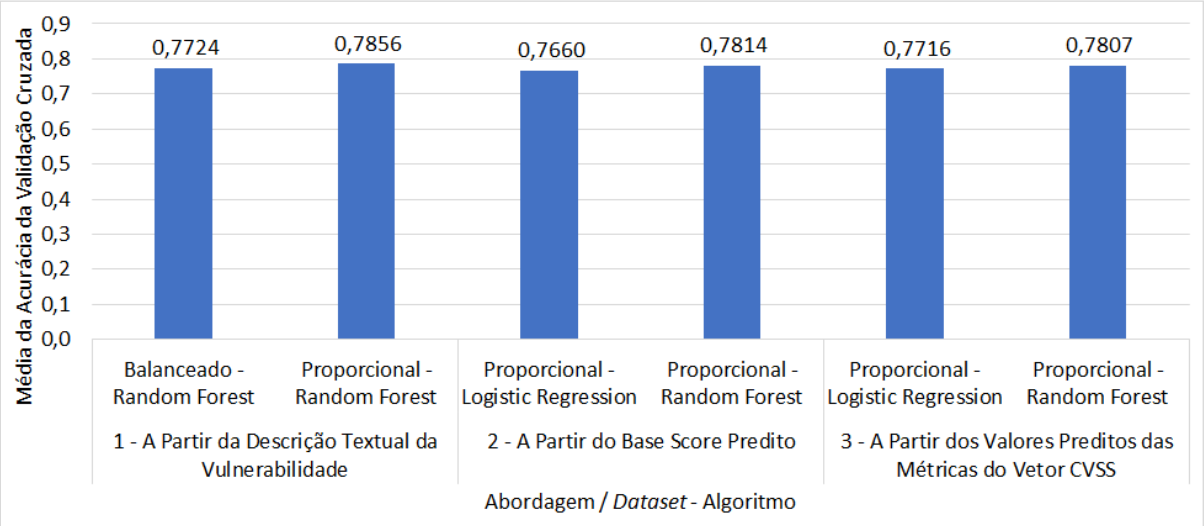


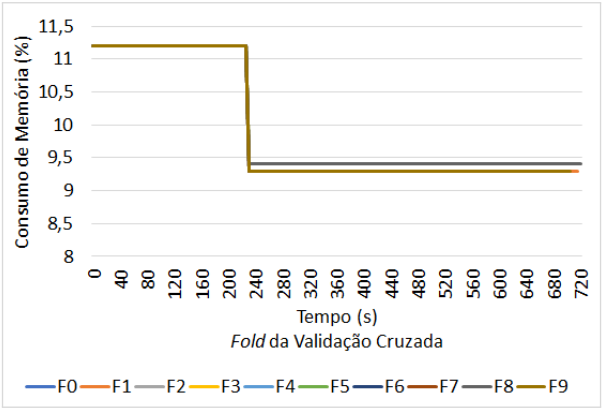
Figura 4.21: Comparativo da média da acurácia da validação cruzada dos dois melhores modelos de cada uma das três abordagens para obtenção da severidade

4.4 Avaliação da Demanda de Recursos Computacionais

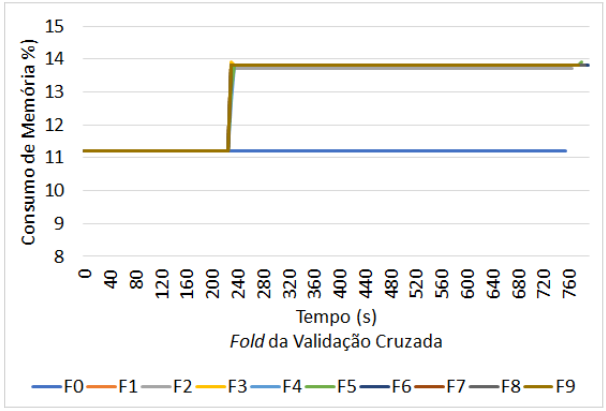
Considerando que a diferença no desempenho dos dois melhores modelos (Fig. 4.21) em todas as abordagens para obtenção da severidade de uma vulnerabilidade é muito pequena, foi avaliado o consumo de recursos computacionais para que esses modelos sejam construídos e também para que eles sejam sejam executados para classificação. Nesse contexto, é importante definir o ambiente em que os modelos foram processados. O servidor onde os modelos foram treinados tinha as seguintes caraterísticas de *hardware*: dois processadores *Intel(R) Xeon(R) Silver 4310 CPU @ 2.10GHz*, cada um com 12 núcleos, e 64 GB de memória RAM. O sistema operacional instalado foi o *VMWare ESXi* versão 7. Foi criada uma máquina virtual com 8 CPU, 60 GB de memória RAM e sistema operacional *Oracle Linux* na versão 7.9.

Na comparação da utilização de memória RAM durante o treinamento dos modelos (Fig. 4.22), utilizando o algoritmo *Random Forest*, a maior diferença entre a quantidade

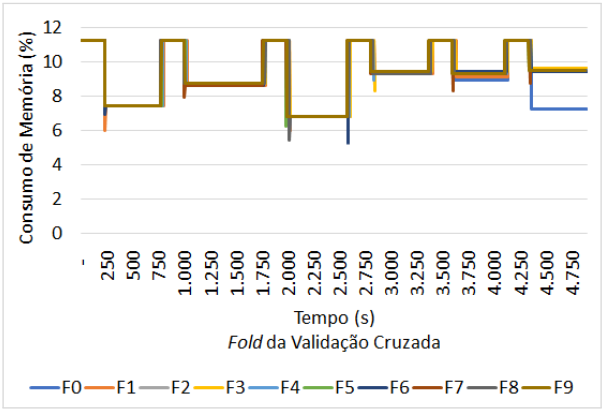
mínima e máxima demandada foi de 5,7% (Fig. 4.22c), o que equivale à 3,42 GB de memória. O maior gasto de memória foi registrado no modelo de obtenção da severidade em função do *base score* predito (Fig. 4.22b), chegando a 14%, ou seja, 8,4 GB. Os outros dois modelos (Figuras 4.22a e 4.22c) alcançaram o pico de 11%, equivalente a 6,6 GB. No gráfico plotado na Figura 4.22c é possível identificar a predição de cada métrica do vetor CVSS. Foi observada uma tendência de que os *folds* consumissem a mesma quantidade de memória RAM, seguindo os mesmos tempos ao variar o consumo, seja para maior ou menor.



(a) Predição da severidade a partir da descrição textual da vulnerabilidade



(b) Obtenção da severidade em função do *base score* predito

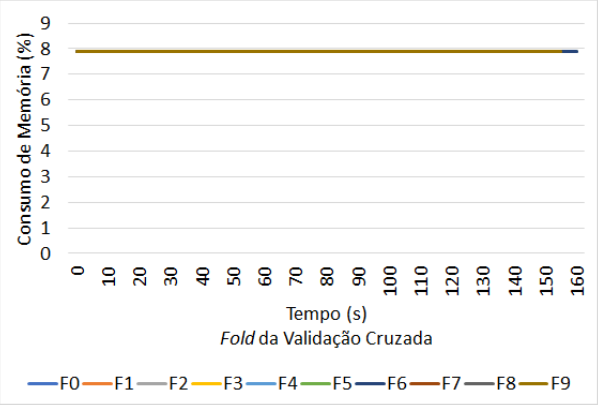


(c) Obtenção da severidade em função do *base score* calculado a partir das métricas do vetor CVSS preditas

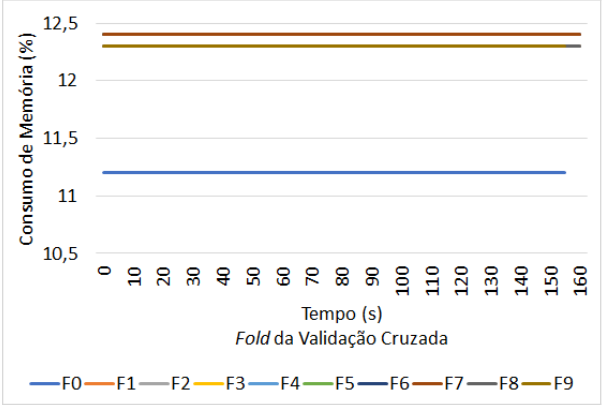
Figura 4.22: Gráficos de consumo de memória RAM para treinar os modelos com o algoritmo *Random Forest* a partir do *dataset* proporcional

Examinando a utilização de memória RAM no processo de classificação, utilizando os modelos treinados com o algoritmo *Random Forest* (Fig. 4.23), a maior diferença entre a quantidade mínima e máxima demandada foi menor do que a observada no treinamento, registrando 2,8% (Fig. 4.23c), o que equivale à 1,68 GB de memória. O maior uso de memória para classificação foi de 12%, equivalente a 7,2 GB, com o modelo obtenção

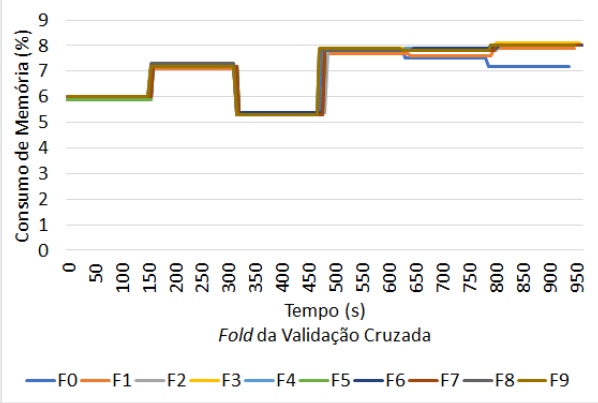
da severidade em função do *base score* predito (Fig. 4.23b). O modelo de predição da severidade a partir da descrição textual da vulnerabilidade (Fig. 4.23a) utilizou, de forma constante, 7,9% de memória RAM, ou seja, 4,74 GB. Já o modelo de obtenção da severidade em função do *base score* calculado a partir das métricas do vetor CVSS preditas (Fig. 4.23c) variou a utilização da memória RAM entre 5,3% (3,18 GB) e 8,1% (4,86 GB), mantendo uma média de 7% (4,2 GB). Embora haja diferença entre a utilização de memória no processo de treinamento e a utilização de memória no processo de classificação, ela é pequena e os recursos empregados são avultados, especialmente para a classificação.



(a) Predição da severidade a partir da descrição textual da vulnerabilidade



(b) Obtenção da severidade em função do *base score* predito

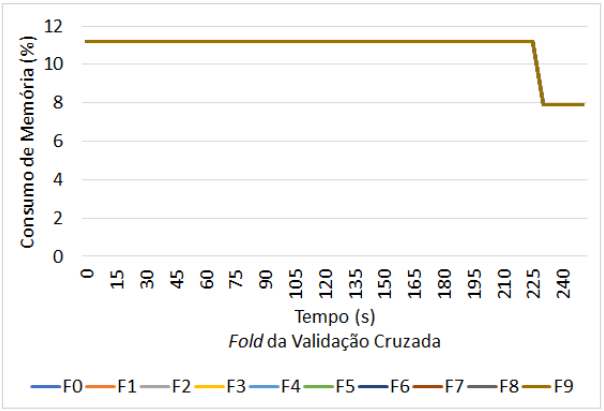


(c) Obtenção da severidade em função do *base score* calculado a partir das métricas do vetor CVSS preditas

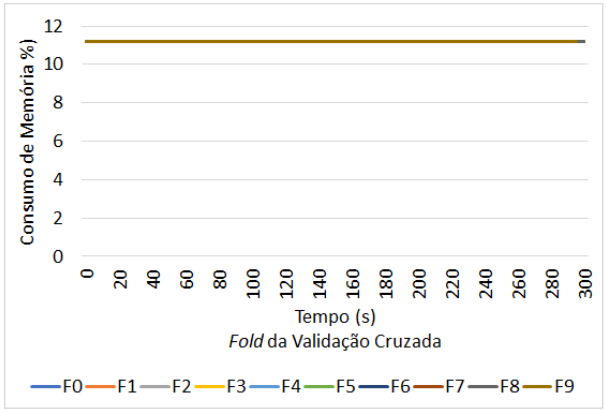
Figura 4.23: Gráficos de consumo de memória RAM para classificação utilizando os modelos com o algoritmo *Random Forest* a partir do *dataset* balanceado

Na comparação da utilização de memória RAM durante o treinamento dos modelos (Fig. 4.24), utilizando o algoritmo *Logistic Regression*, a maior diferença entre a quantidade mínima e máxima demandada foi de 5,7% (Fig. 4.24c), o que equivale à 3,42 GB de memória. Os três modelos atingiram o pico de 11,2%, ou 6,72 GB de memória RAM utilizada. O modelo de obtenção da severidade em função do *base score* predito (Fig. 4.24b)

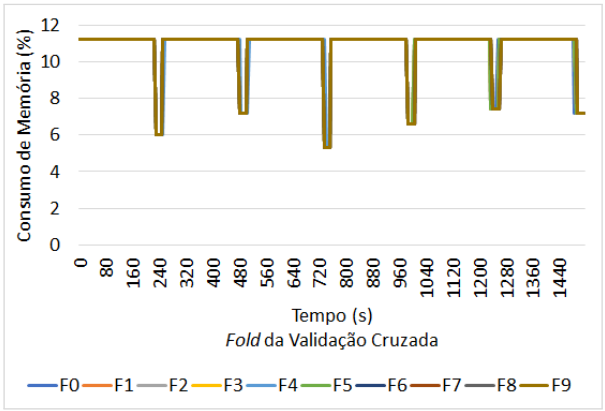
se manteve constante e não registrou variação em nenhum dos *folds*. O modelo predição da severidade a partir da descrição textual da vulnerabilidade (Fig. 4.24a) manteve os 11,2% de utilização de memória durante 3min45s em todos os *folds*. Após isso, a utilização da memória foi reduzida para 7,9%, ou 4,74 GB. O modelo de obtenção da severidade em função do *base score* calculado a partir das métricas do vetor CVSS preditas (Fig. 4.24c) manteve 11,2% de utilização de memória durante os primeiros 3min45s do treinamento de cada métrica do vetor CVSS, em todos os *folds*. Após esse tempo, o consumo da memória RAM caía para valores entre 5,3% (3,18 GB) e 7,4% (4,44 GB).



(a) Predição da severidade a partir da descrição textual da vulnerabilidade



(b) Obtenção da severidade em função do *base score* predito



(c) Obtenção da severidade em função do *base score* calculado a partir das métricas do vetor CVSS preditas

Figura 4.24: Gráficos de consumo de memória RAM para treinar os modelos com o algoritmo *Logistic Regression* a partir do *dataset* proporcional

Examinando o consumo de memória RAM no processo de classificação, utilizando os modelos treinados com o algoritmo *Logistic Regression*, só foi possível plotar o gráfico referente à obtenção da severidade em função do *base score* calculado a partir das métricas do vetor CVSS preditas (Fig. 4.25), pois, em virtude do baixo tempo necessário para a execução do processo, não foram registrados dados suficientes para plotar um gráfico. A

maior diferença entre a quantidade mínima e máxima demandada observada foi de 5,7%, o que equivale a 3,42 GB. O modelo de predição da severidade a partir da descrição textual da vulnerabilidade usou 7,9% (4,74 GB) de memória RAM e o modelo de obtenção da severidade em função do *base score* predito usou 11% (6,6 GB) de memória RAM. Já o modelo de obtenção da severidade em função do *base score* calculado a partir das métricas do vetor CVSS preditas (Fig. 4.25) fez uso de um quantitativo de memória RAM entre 5,3% (3,18) e 7,4% (4,44 GB), com média de 6,6% (3,96 GB).

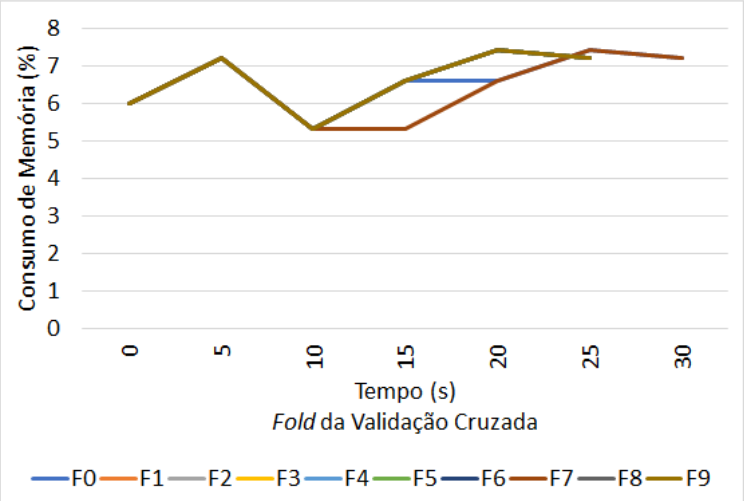


Figura 4.25: Gráfico do consumo de memória RAM durante a classificação para de obter a severidade em função do *base score* calculado a partir das métricas do vetor CVSS preditas utilizando modelos treinados com o algoritmo *Logistic Regression* a partir do *dataset* proporcional

A obtenção da severidade em função do *base score* calculado a partir das métricas do vetor CVSS preditas é a metodologia que retorna informações mais detalhadas sobre uma vulnerabilidade. No entanto, com base nas análises ora realizadas, é inconteste que ela apresenta o maior consumo de recursos. Ressalta-se que, em relação ao tempo de execução, com o emprego do algoritmo *Random Forest* (Fig. 4.26a), a assimetria entre as metodologias é notável, chegando a mais de seis vezes. O mesmo ocorre para a classificação. O espaço de armazenamento também é muito elevado, mais de 25 GB (Fig. 4.26b).

Fazendo uso do algoritmo *Logistic Regression* para a mesma metodologia, embora os recurso de processamento e memória empenhados sejam muito similares, em termos de tempo de execução (Fig. 4.26a), foi alcançado $\frac{1}{3}$ do tempo necessário de execução do algoritmo *Random Forest*. No que tange ao espaço de armazenamento (Fig. 4.26b) a diferença é ainda maior. Os modelos construídos com algoritmo *Logistic Regression* requerem até 120 vezes menos espaço de armazenamento. Se for considerado apenas o

espaço de armazenamento, a diferença entre as três metodologias, utilizando o algoritmo *Logistic Regression*, é praticamente inexistente.

Destarte, a decisão sobre qual modelo utilizar está mais associada à disponibilidade de recursos computacionais. Em último caso, a fim de manter o maior nível de detalhamento de informações entregue pelo algoritmo com a utilização de recursos computacionais mais limitados, é possível assumir uma margem de erro ligeiramente maior e substituir o algoritmo com melhor desempenho, *Random Forest*, pelo segundo melhor algoritmo, o *Logistic Regression*. Tendo em vista que as análises conduzidas neste estudo são a partir de *datasets*, foi desenvolvida uma prova de conceito para verificação prática da classificação de uma vulnerabilidade individualmente utilizando o algoritmo *Logistic Regression*[9].

4.5 Predição do Tempo de Correção de Vulnerabilidades de Aplicações de Comunicação

4.5.1 Metodologia de Predição

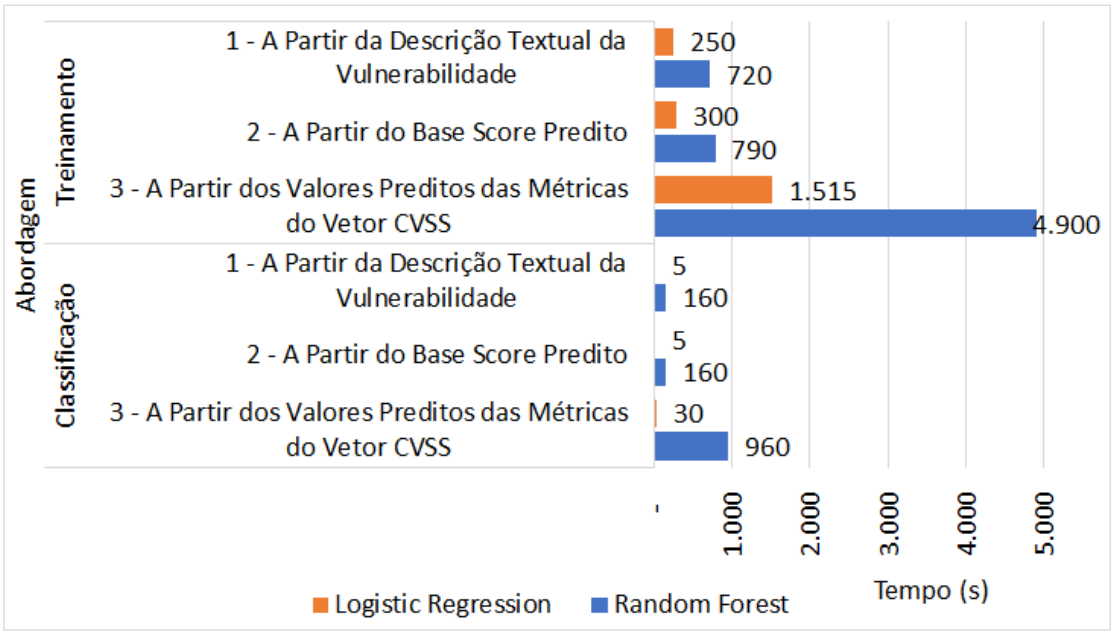
Para efetuar a predição do tempo necessário até que uma vulnerabilidade seja corrigida pelos desenvolvedores, foram implementados cinco tipos diferentes de classificadores, cuja diferença consiste nos parâmetros de entrada, quais sejam: (i) descrição textual da vulnerabilidade como parâmetro de entrada; (ii) descrição textual e métricas do vetor CVSS da vulnerabilidade como parâmetros de entrada; (iii) métricas do vetor CVSS da vulnerabilidade como parâmetros de entrada; (iv) descrição textual e severidade da vulnerabilidade como parâmetros de entrada; ou (v) severidade da vulnerabilidade como parâmetro de entrada. A predição consiste em uma classe que representa um intervalo de dias no qual supõe-se que a correção da vulnerabilidade será disponibilizada pelo desenvolvedor da aplicação.

Classes

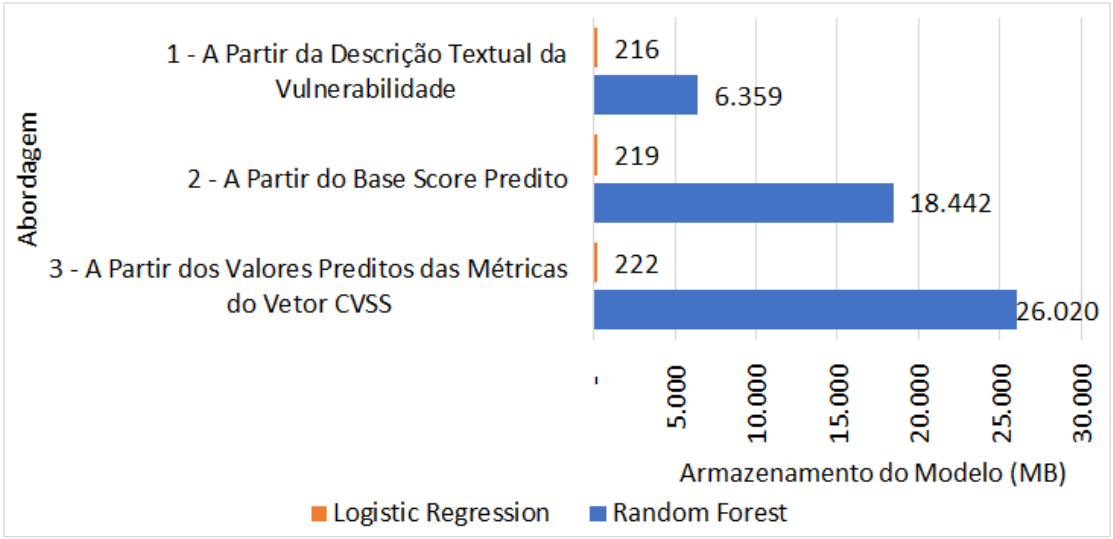
Considerando a grande variação na quantidade de dias decorridos entre a descoberta e a correção da vulnerabilidade, a discretização também foi aplicada ao tempo de correção para cada vulnerabilidade. O critério adotado foi criar classes que tivessem um número semelhante de vulnerabilidades. Em particular, foram definidas as seguintes classes:

- **A**: Tempo de Correção (dias) <= 40;

[9]https://github.com/arsbraga/CVSS-Predict



(a) Tempo de execução



(b) Espaço de armazenamento ocupados pelos modelos

Figura 4.26: Gráficos de tempo de execução e espaço de armazenamento ocupados pelos modelos treinados a partir do *dataset* proporcional

- **B**: 40 < Tempo de Correção (dias) <= 80;
- **C**: 80 < Tempo de Correção (dias) <= 120;
- **D**: 120 < Tempo de Correção (dias) <= 160; e
- **E**: Tempo de Correção (dias) > 160.

Extração de Dados

Os bancos de dados CVE ou NVD não têm dados suficientes sobre as datas em que as correções de vulnerabilidade foram disponibilizadas. Para obter tais informações, foi realizada uma busca nos sites dos desenvolvedores, com base nas informações relatadas nos *changelogs* dos *softwares* e nos avisos de segurança dos desenvolvedores. Ressalta-se que, no escopo do presente estudo, consideram-se como correções um pacote que corrija especificamente a vulnerabilidade ou a atualização de versão que seja capaz de corrigir a vulnerabilidade. A data da correção refere-se à data em que o pacote de correção ou a atualização da versão foi disponibilizada. O tempo necessário para correção da vulnerabilidade foi calculado entre a data de criação do registro na base de dados CVE e a data em que o desenvolvedor disponibilizou um pacote de correção ou uma nova versão que corrija a vulnerabilidade no *software* afetado. Destaca-se que a data de criação do registro na base de dados CVE refere-se à alocação ou reserva de um ID CVE. Essa data não representa, necessariamente, a data da descoberta ou da divulgação pública da vulnerabilidade[10].

Devido à natureza intrinsecamente manual desse processo, não seria viável realizar essa análise para todas as vulnerabilidades no CVE. Em virtude disso, um subconjunto do CVE foi selecionado para que essas informações de tempo de correção fossem acrescentadas. Dada a importância das aplicações de comunicação, com as restrições impostas pela COVID-19, esse processo foi realizado para as seguintes aplicações, com base em seus sites de *security advisory*: *Cisco Webex*[11], *Microsoft*[12] (*Skype* e *Teams*), *Facebook Messenger*[13], *Whatsapp*[14] e *Zoom*[15]. Esta pesquisa sugere que a *Cisco* é o desenvolvedor mais transparente em relação ao tempo entre a publicação da vulnerabilidade no banco de dados CVE e a disponibilização da correção para os aplicativos afetados. De todas as

[10]A explicação sobre a data de criação do registro consta em todas as páginas sobre vulnerabilidades em: https://cve.mitre.org/
[11]https://tools.cisco.com/security/center/
[12]https://msrc.microsoft.com/update-guide/vulnerability/
[13]https://www.facebook.com/security/advisories/[cve-year-id]
[14]https://www.whatsapp.com/security/advisories/
[15]https://explore.zoom.us/en/trust/security/security-bulletin/

vulnerabilidades *Cisco Webex* encontradas no CVE no momento da pesquisa, 97% foram detalhadas no site do desenvolvedor. Em seguida está o *Facebook*, desenvolvedor do *WhatsApp*, para o qual seu site continha, pelo menos, a informação do mês em que as correções de vulnerabilidades foram disponibilizadas, para 95% das vulnerabilidades relatadas.

O *dataset* de treinamento e validação contém 157 amostras. Em virtude da utilização da validação cruzada com dez *folds*, sete *folds* continham dezesseis amostras de validação e os três restantes quinze amostras. Considerando a metodologia de separação temporal, o *dataset* utilizado para teste na avaliação desta seção conta com dezenove amostras.

Extração de *Features*

Para os experimentos referentes à predição do tempo de correção de vulnerabilidades, além das *features* geradas a partir da descrição textual da vulnerabilidade, as *features* também podem incluir os valores reais das métricas do vetor CVSS ou as severidades reais das vulnerabilidades. Em certos experimentos, também considerou-se apenas as métricas de vetor CVSS ou a apenas a severidade como *features* de entrada para realizar a predição do tempo de correção de vulnerabilidades.

4.5.2 Avaliação dos Resultados

Ao verificar o resultado da validação cruzada, com base na média dos dez *folds*, observa-se que o melhor é 57,3% de acurácia, conforme gráfico plotado na Figura 4.27. Nesse gráfico pode-se observar que o desempenho do teste, em geral, foi significativamente inferior à média registrada na validação cruzada, ficando fora do desvio padrão. A melhor média registrada foi do modelo treinado com o algoritmo *Support Vector Machine*, cujas entradas foram a descrição textual e as métricas do vetor CVSS da vulnerabilidade. Dentre os cinco melhores resultados da validação cruzada, nenhum usou um dos dois algoritmos de redução de dimensionalidade abordados neste estudo.

Comparando os melhores resultados da validação cruzada (Fig. 4.27) com os melhores resultados do teste (Fig. 4.28), constata-se que apenas um dos cinco melhores resultados da validação cruzada, modelo treinado com o algoritmo *Support Vector Machine*, cujas entradas foram a descrição textual e as métricas do vetor CVSS da vulnerabilidade, aparece como um dos cinco melhores resultados do teste. Esse modelo foi o único, entre os cinco melhores desempenhos do teste, que não usou algoritmo de redução de dimensionalidade. O modelo com melhor desempenho no teste (36,8% de acurácia) foi treinado com o algoritmo *Passive-Aggressive* e utilização de algoritmo de redução de dimensionalidade

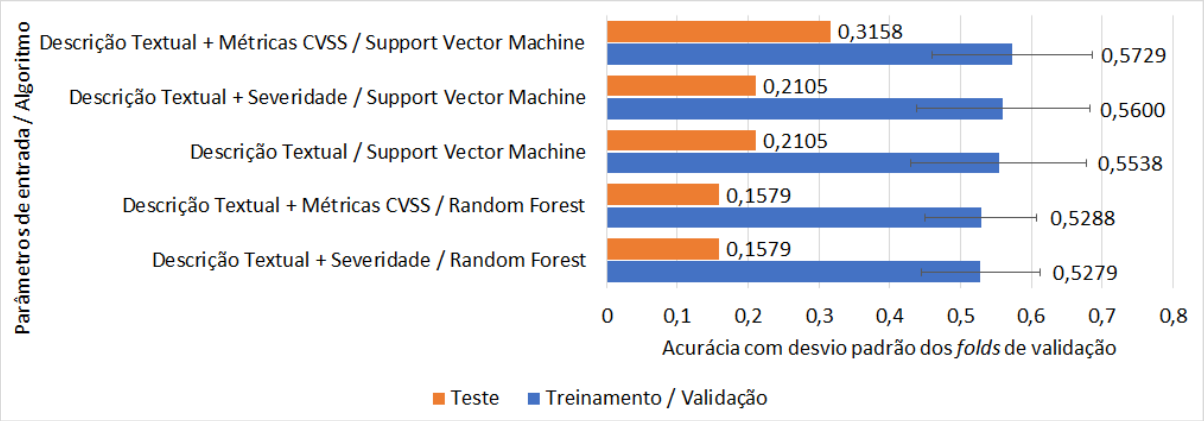


Figura 4.27: Cinco melhores médias de acurácia da validação cruzada da predição do tempo de correção de vulnerabilidades de aplicações de comunicação com desvio padrão médio dos *folds* de validação e comparativo com a acurácia da predição do teste

LDA on the PCA. A entrada do algoritmo foi apenas a descrição textual da vulnerabilidade. Esse resultado foi superior ao registrado na validação cruzada, cuja média foi de 50,8%, mas dentro do desvio padrão populacional (0,15) referente à acurácia dos *folds* da validação cruzada.

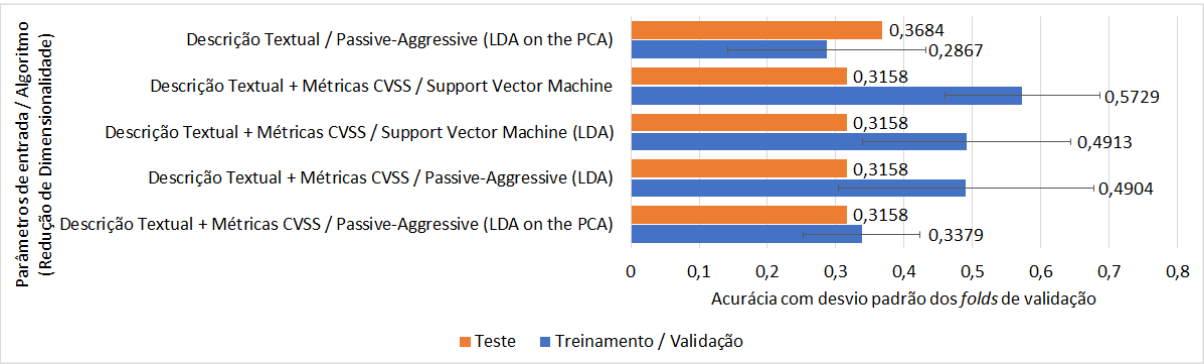


Figura 4.28: Cinco melhores médias de acurácia do teste da predição do tempo de correção de vulnerabilidades de aplicações de comunicação com desvio padrão médio dos *folds* de validação e comparativo com a acurácia da predição da validação cruzada

É possível examinar na Figura 4.29 que há uma instabilidade no modelo, verificada pela diferença no desempenho dos dez *folds* e do teste. No que concerne ao modelo treinado com o algoritmo *Passive-Aggressive*, com a utilização do algoritmo de redução de dimensionalidade LDA on the PCA e entrada sendo apenas a descrição textual da vulnerabilidade, o *fold* com o pior desempenho foi o F1, com 0% de acurácia e os melhores foram os *folds* F2 e F3, com 43,8% de acurácia cada um. O desvio padrão populacional, calculado com base nos dez *folds* desse modelo, foi de 0,15 e a amplitude, considerando o melhor e o pior desempenho, foi de 43,8 pontos percentuais. Já em relação ao modelo treinado com o algoritmo *Support Vector Machine*, cujas entradas foram a descrição textual

e as métricas do vetor CVSS da vulnerabilidade, o *fold* com o pior desempenho foi o F1, com 37,5% de acurácia e o melhor foi o *fold* F5, com 75% de acurácia. O desvio padrão populacional, calculado com base nos dez *folds* desse modelo, foi de 0,11 e a amplitude, considerando o melhor e o pior desempenho, foi de 37,5 pontos percentuais. Através dessa análise demonstra-se que o modelo que obteve melhor resultado no teste apresentou maior instabilidade durante a validação cruzada.

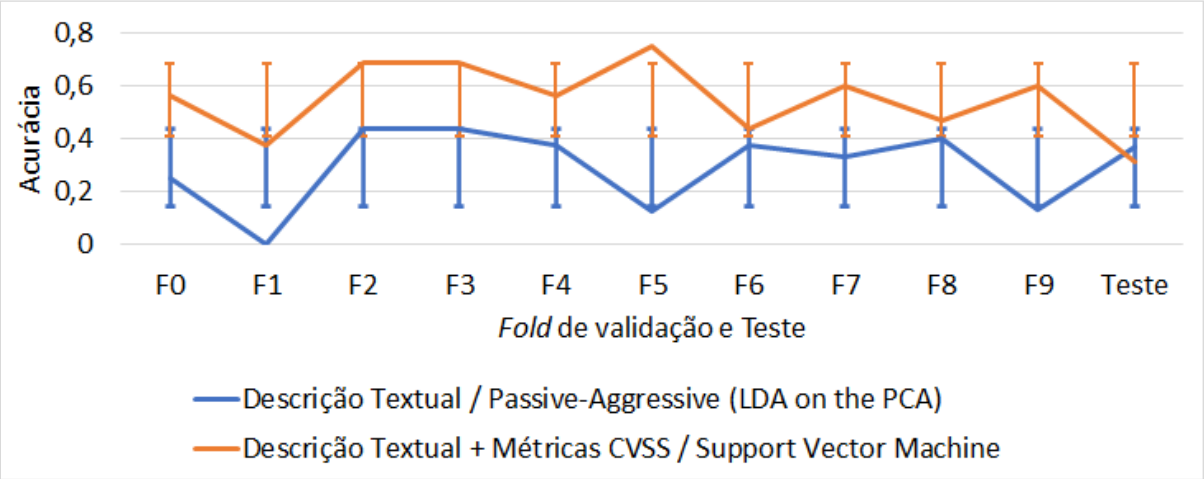


Figura 4.29: Desempenho dos *folds* da validação cruzada e do teste para predição do tempo de correção de vulnerabilidades de aplicações de comunicação e desvio padrão médio dos *folds* referente aos dois modelos com melhor desempenho no teste

Um fator que possivelmente influenciou no baixo desempenho é a quantidade reduzida de amostras disponíveis. Com uma quantidade maior de amostras, além da influência no algoritmo, seria possível efetuar um estudo mais refinado na definição das classes. Uma possível solução para o problema seria a concentração das informações sobre correção de vulnerabilidades na bases de dados do CVE ou do NIST.

4.6 Trabalhos Relacionados

Há outros trabalhos que versam sobre a aplicação de inteligência artificial para predição da severidade de vulnerabilidades.

Nikonov *et al.* [53] usaram um modelo de rede neural de aninhamento de vetores para documentos de texto *Doc2Vec* (D2V), que foi treinado a partir de descrições textuais de vulnerabilidades, gerando as *features*. Essas *features*, juntamente com CVE-ID e CVSS, foram usadas para construir modelos de Rede Neural. Esses modelos foram, então, usados para classificar e avaliar cada componente da métrica base.

Como neste trabalho, Shahid e Debar [67] aplicaram o processamento de linguagem natural na descrição das vulnerabilidades como ponto de partida para treinar classificadores de vulnerabilidades. Eles também estabeleceram uma metodologia para predizer cada uma das métricas do vetor CVSS. No entanto, o foco de seu trabalho foi o classificador BERT. A metodologia proposta pelos autores foi similar à terceira abordagem avaliada neste estudo. No entanto, o algoritmo utilizado por eles não estava entre os seis examinados no presente trabalho. Assim, naquele artigo o resultado foi de 55,3% de acurácia na predição do valor exato do *base score*. Por outro lado, nesta dissertação alcançou-se uma melhoraria desse resultado, chegando a uma acurácia média de 57%. Para alcançar esse resultado, foram utilizados seis algoritmos diferentes combinados com dois tipos distintos de *datasets*. Naquele trabalho também não foi realizada a validação cruzada, como neste. Dessa forma, se considerarmos apenas o desempenho do melhor *fold* do classificador que obteve a melhor média de acurácia, o desempenho sobe para 59%. Outra diferença é que Shahid e Debar analisaram a versão 3 do vetor CVSS.

Babalau *et al.* [6] desenvolveram métodos semelhantes aos usados neste trabalho, incluindo a predição de cada métrica do vetor CVSS. Predizer essas métricas é importante no sentido de oferecer uma visão mais ampla das peculiaridades de uma vulnerabilidade. Cinco métodos (*TF-IDF + MNB*, *BiLSTM + Word2Vec*, *smallBERT*, *CNN + Word2Vec*, *Multi-Task Learning*) diferentes foram usados por eles. A compensação do desequilíbrio de classe foi solucionado pelos autores daquele estudo aplicando pesos às classe e, em seguida, aumentando as classes *LOW* e *CRITICAL* usando o *framework TextAttack Python*. Mas apenas esse tipo de conjunto de dados foi usado. Assim como Shahid e Debar, Babalau *et al.* usaram a versão 3 do vetor CVSS como referência.

Não foram identificados trabalhos na literatura que propuseram uma metodologia para predizer o tempo necessário para que uma vulnerabilidade seja corrigida.

Capítulo 5

Conclusão

Nesta dissertação foi apresentada uma avaliação dos impactos da pandemia da COVID-19 na percepção do desempenho da rede pelo usuário, nas redes de acesso, de *backbone*, sem fio e de telefonia celular e na segurança cibernética.

Nesse sentido, utilizando recursos de pesquisa de opinião, observou-se que a percepção de 61,5% dos usuários era de instabilidade na conexão com a Internet, durante o período da pandemia, dirimindo a questão de pesquisa número 1. Do total de respondentes que relataram a percepção de uma conexão com a Internet instável, 64% disseram que essa percepção de instabilidade apareceu após o início da pandemia. É provável que tal fato tenha ocorrido em decorrência da mudança na forma como as pessoas se relacionam com a tecnologia. A utilização de aplicativos de videoconferência e de *streaming* vídeo aumentou significativamente, não só para entretenimento, mas também para atividades profissionais que, de certa forma, exigem um desempenho maior.

A fim de buscar o fator que teve maior influência nessa percepção de instabilidade da conexão com a Internet e responder a questão de pesquisa número 2, os dados da pesquisa de opinião foram cruzados entre si. Observou-se que 44% dos respondentes alteraram seus planos de Internet fixa por outro de maior capacidade. Porém, apenas 37% daqueles que alteraram o plano relataram melhora na percepção de estabilidade de conexão com a Internet. A quantidade de dispositivos conectados ou a quantidade de pessoas que moram na mesma residência também não revelaram influência sobre essa percepção. No entanto, daqueles que consideram sua conexão com a internet instável, apenas 13% utiliza redes sem fio na faixa de 5 GHz. Essa circunstância permite inferir que a maior popularização e área de cobertura das redes sem fio na faixa de 2,4 GHz em relação às redes sem fio na faixa de 5 GHz, e consequentemente uma maior possibilidade de interferência entre redes próximas, podem degradar significativamente o desempenho

da rede sem fio, impactando diretamente na percepção de estabilidade de conexão com a Internet. Uma possível simples solução para essa questão é a maior adoção do padrão IEEE 802.11ax [45].

Em atenção à questão de pesquisa número 3 foram conduzidos alguns experimentos e coletas de dados, a partir dos quais, restou evidenciado, como pela análise de tráfego da RNP, que os aplicativos de videoconferência tiveram seu uso largamente ampliado. O impacto dessa mudança na RNP, por exemplo, foi um aumento de tráfego na ordem de 1,4 TB após 5 meses de uso de ferramentas de videoconferência. Ainda nessa linha, ao examinar o comportamento dessas ferramentas constatou-se que o *Skype* e o *WhatsApp* se distinguem das demais ferramentas pela capacidade de intercambiar entre os modos cliente-servidor e par-a-par, dependendo do tipo de conexão dos dispositivos. Essa funcionalidade, se implementada pelos outros aplicativos, poderia ajudar a reduzir a demanda sobre seus respectivos servidores, bem como eventualmente reduzir o tráfego na internet. Ainda em relação ao teletrabalho, identificaram-se, na Subseção 3.2.5, as aplicações/protocolos mais adequados para cada tipo de atividade executada.

As avaliações conduzidas no âmbito desta pesquisa, em decorrência da questão de pesquisa número 4, apontaram para um expressivo aumento de tráfego na internet, conforme dados da análise de tráfego das redes sem fio do IX.br, que desvelou uma média de crescimento no tráfego, que no início da pandemia chegou a 680 Gb/s por mês, mais de 3 vezes a taxa de crescimento mensal pré-pandemia, 200 Gb/s por mês. Respondendo a questão de pesquisa número 5, observou-se que outras redes trilharam o caminho inverso, ou seja, a mudança de perfil no volume de tráfego difere a depender da rede analisada. A RNP, por exemplo, registrou uma redução de 73% no volume de tráfego total do protocolo TCP entre março e abril de 2020, possivelmente pela redução e, em alguns casos, pela suspensão das atividades presenciais nas Universidades. Em que pese tal redução no tráfego, enfatiza-se o relevante incremento no volume de tráfego medido nas portas utilizadas pelas aplicações de comunicação. Esses números demonstram diversas mudanças no volume de tráfego das redes em decorrência da pandemia, seja no maior ou menor volume, seja no tipo.

Essas mudanças no perfil de tráfego foram confrontadas com a quantidade de conexões de dispositivos celulares à uma ERB específica e com a mobilidade da comunidade, notadamente na cidade de Niterói-RJ. No que concerne a correlação entre a quantidade de conexões de dispositivos celulares e a mobilidade da comunidade, constatou-se ser grande para varejo e lazer, estações de transporte público e concentração residencial e em locais

de trabalho; média para mercados e farmácias; e pequena para parques. Já para a correlação entre a quantidade de conexões de dispositivos celulares e o tráfego ou a quantidade de pacotes capturados na faixa de 2,4 GHz das redes sem fio, essa correlação se mostrou pequena. Particularmente na avaliação da correlação entre as mudanças no perfil de tráfego e a mobilidade da comunidade, constatou-se ser grande para varejo e lazer, mercado e farmácia e estações de transporte público; média para concentração residencial e em locais de trabalho; e pequena para parques. Destarte a resposta para a questão de pesquisa número 6 é afirmativa.

À medida que foram encontradas evidências suficientes para concluir que o tráfego de dados e o uso de aplicações de comunicação aumentam drasticamente, a segurança é uma questão fundamental, que precisa ser destacada. Em virtude desse aumento, essas aplicações podem se tornar alvo preferencial de elementos adversos em possíveis explorações de vulnerabilidades, que podem colocar os usuários em risco. Face a isso e com o objetivo de solucionar as indagações abordadas na questão de pesquisa número 7, este trabalho avaliou três abordagens para predizer a severidade de uma vulnerabilidade, quais sejam:

1. **Predição da severidade:** a saída produzida pelo classificador é diretamente a severidade;
2. **Predição do *base score*:** a saída produzida pelo classificador é o *base score*, sobre o qual é aplicada uma função para obtenção da severidade; e
3. **Predição das métricas do vetor CVSS:** há seis classificadores, um para cada métrica a ser predita. Após obter todas as métricas o *base score* é calculado e, então, é aplicada sobre ele uma função para obtenção da severidade.

A diferença entre as três abordagens é o nível de detalhamento das informações obtidas a partir da descrição textual da vulnerabilidade. Quanto maior o nível de detalhamento, mais recursos computacionais (processador, memória e armazenamento dos modelos) são utilizados, seja na perspectiva da quantidade ou do tempo. Ao estudar tais diferenças e como essas diferenças podem impactar o produto final, elucida-se a questão de pesquisa número 8.

Considerando o melhor algoritmo, *Random Forest*, as três abordagens obtiveram resultados muito similares, no que concerne à acurácia. A primeira abordagem alcançou 78,6% e as demais 78,1%. O segundo melhor algoritmo, *Logistic Regression*, registrou a acurácia de 77% na primeira abordagem, 76,6 na segunda e 77,2 na terceira. Logo, a

decisão sobre qual abordagem utilizar dependerá da disponibilidade de recursos computacionais, pois a diferença de acurácia entre as abordagens é muito pequena. Enfatiza-se que a diferença no consumo de recursos computacionais entre os dois melhores algoritmos é significativa (Seção 4.4). Um aspecto relevante sobre o desempenho é o tipo de *dataset* utilizado. O *dataset* proporcional leva vantagem sobre o balanceado. A diferença média entre eles é de 1,6 ponto percentual para o algoritmo *Random Forest* e 1,4 ponto percentual para o algoritmo *Logistic Regression*. Também é importante assinalar que os algoritmos de redução de dimensionalidade e de *voting* não ajudaram a melhorar os resultados. Assim, argumenta-se que esses números são suficientes para permitir que os administradores de sistemas tenham uma visão mais ampla de uma vulnerabilidade que ainda não foi submetida ao processo de avaliação manual das equipes especializadas, como a do repositório de dados de gerenciamento de vulnerabilidades do governo dos EUA, o NVD. Nesse sentido, um administrador de sistemas pode avaliar a vulnerabilidade sob a ótica de cada métrica CVSS predita e, a partir destas, direcionar os esforços para mitigar os riscos para a equipe mais adequada, atribuindo a prioridade adequada à sua severidade. Sanada, então, a questão de pesquisa número 9, é relevante clarificar que as descrições textuais da vulnerabilidade, presentes na base de dados analisada, carregam um enorme grau de subjetividade. Esses textos são produzidos por pessoas e equipes diferentes, o que os faz não ter um padrão bem definido de redação. Além disso, o texto nem sempre contém um descritivo completo, que permita classificar a vulnerabilidade de acordo com cada métrica do vetor CVSS. Essa é uma hipótese que pode justificar a diferença no desempenho dos modelos de uma métrica para outra. Ademais, os resultados encontrados na literatura para aplicações similares não apresentaram uma diferença significativa.

Os resultados também sugerem que o aprendizado realizado a partir de um conjunto de dados genérico tem boa eficiência para predizer vulnerabilidades de aplicações específicas, em particular para aplicações de comunicação. De fato, quase todas as acurácias encontradas para vulnerabilidades de aplicações de comunicação foram superiores às obtidas nas bases de teste com vulnerabilidades genéricas. Assim, em réplica à questão de pesquisa número 10, os algoritmos revelaram uma excelente capacidade de generalização.

Em relação à questão de pesquisa número 11, concernente à predição do tempo de correção de vulnerabilidades de aplicações de comunicação, foi observada uma instabilidade no modelo. É possível chegar a essa conclusão em virtude da diferença no desempenho dos dez *folds* e do teste. A melhor média da validação cruzada foi de 57,3% de acurácia, obtendo 31,6% de acurácia no teste. A acurácia dos *folds* variou entre 37,5% e 75% para o modelo treinado com o algoritmo *Support Vector Machine*, cujas entradas foram a descri-

ção textual e as métricas do vetor CVSS da vulnerabilidade. Um dos fatores que pode ter contribuído para o baixo desempenho é a quantidade reduzida de amostras disponíveis. Um número maior de amostras não apenas influenciaria o algoritmo em si, mas também permitiria uma análise mais sofisticada para definição das classes. Uma possível solução para esse problema é centralizar as informações sobre a correção de vulnerabilidades no banco de dados CVE ou NVD.

Nota-se também a ausência de propostas de predição do tempo de correção de vulnerabilidades na literatura. Isso possivelmente se deve à falta de informações sobre a data de disponibilidade do pacote de correção nos bancos de dados do CVE e do NVD. Mesmo no site dos desenvolvedores, essas informações nem sempre estão disponíveis. Portanto, registrar sistematicamente essas informações pode ser importante para um maior desenvolvimento nessa área.

5.1 Publicações

A presente dissertação deu origem a seguinte publicação:

- Braga, Allan Rodrigo de Souza; Passos, Diego; Rocha, Antonio Augusto de Aragão . Mudança no perfil de tráfego de redes brasileiras em decorrência da pandemia pelo novo coronavírus. In: Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos, 2021, Brasil. Anais do XXXIX Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC 2021), 2021. p. 406.

Além do trabalho já publicado, mais um trabalho intitulado "*Machine Learning-Based Approach for Predicting Severity and Patching Time of Vulnerabilities*" está em produção com a colaboração dos professores Diego Passos e Antonio Augusto de Aragão Rocha.

5.2 Trabalhos Futuros

Constatada a percepção de instabilidade na conexão com a Internet, uma linha de trabalho futuro seria, além da rede sem fio, avaliar outros fatores, como o provedor de Internet, a localização geográfica e o tipo de infraestrutura de rede de última milha (fibra ótica, cabos metálicos, entre outros). Como foi observado que apenas 13% dos usuários que relataram que sua percepção da conexão com a Internet é instável utilizam redes sem fio na faixa de 5 GHz, seria interessante investigar as razões para a baixa adoção desse tipo

de rede. Isso poderia ajudar a desenvolver estratégias para aumentar a sua popularidade e melhorar a qualidade da conexão dos usuários.

Em um trabalho futuro também é possível aprofundar a compreensão das correlações entre a quantidade de conexões de dispositivos celulares e a mobilidade da comunidade. Por exemplo, pode-se investigar como essas correlações variam em diferentes regiões geográficas e em diferentes contextos sociais e culturais.

Acerca da aplicação de *Machine Learning* na predição da severidade e do tempo de correção de vulnerabilidades, a continuidade ao presente estudo pode se dar pela investigação do desempenho e da eficácia de outras abordagens de classificação e predição (*e.g.*, especialização de classificadores por fabricantes, por aplicações ou por tipo de aplicações). Além disso, explorar diferentes formas de representação de dados, como o uso de *embeddings* pré-treinados em NLP, podem ajudar a capturar melhor o significado semântico dos termos usados na descrição das vulnerabilidades.

Embora não tenha sido relatado formalmente neste trabalho, foram executados testes de aprendizado e validação das métricas do vetor CVSS utilizando algoritmos de *Deep Learning*. Dentre eles, o *Convolutional Neural Networks* (CNN) e *Recurrent Neural Networks* (RNN), do módulo *TensorFlow*. Os testes incluíram os algoritmos de redução de dimensionalidade. O desempenho dos algoritmos citados ficou muito próximo ao daqueles descritos no trabalho. Em alguns casos, os algoritmos de *Deep Learning* registraram desempenho inferior. Enfatiza-se que esses algoritmos têm uma demandam muito alta por recursos computacionais, bem como elevado tempo para convergência. Sugere-se, como um tema para trabalhos futuros, a implementação e a análise para ajustar os hiperparâmetros de modelos de *Deep Learning*.

Um ponto que merece maior atenção para elucidação, em um trabalho futuro, é o baixo desempenho dos algoritmos de redução de dimensionalidade. Uma sugestão é avaliar como lidar de forma mais otimizada com a esparsidade dos dados.

Realizar um estudo com um conjunto de dados de tempo de correção de vulnerabilidade mais amplo é fundamental para ter uma avaliação mais precisa do desempenho, estabilidade e capacidade de generalização do algoritmo. Esse estudo pode incluir vulnerabilidades de diferentes tipos de aplicações e sistemas operacionais, a fim de otimizar os modelos. Isso permitiria, inclusive, reavaliar a distribuição das classes de forma mais precisa. Por fim, uma outra linha de ação seria analisar a forma como os principais desenvolvedores divulgam informações sobre o acompanhamento das vulnerabilidades identificadas e suas respectivas correções. A partir desses dados seria viável propor um modelo

para centralizar essas informações em um único banco de dados.

Referências

[1] Abbas, M.; Ali, K.; Memon, S.; Jamali, A.; Memon, S.; Ahmed, A. Multinomial naive bayes classification model for sentiment analysis. 62–67.

[2] Anatel, B. Ato nº 14448, de 04 de dezembro de 2017. *Diário Oficial [da] República Federativa do Brasil* (2017).

[3] Anguita, D.; Ghelardoni, L.; Ghio, A.; Oneto, L.; Ridella, S. The'k'in k-fold cross validation. In *ESANN* (2012), pp. 441–446.

[4] Aquino, E. M. L.; Silveira, I. H.; Pescarini, J. M.; Aquino, R.; de Souza-Filho, J. A.; dos Santos Rocha, A.; Ferreira, A.; Victor, A.; Teixeira, C.; Machado, D. B.; Paixão, E.; Alves, F. J. O.; Pilecco, F.; Menezes, G.; Gabrielli, L.; Leite, L.; da Conceição Chagas de Almeida, M.; Ortelan, N.; Fernandes, Q. H. R. F.; Ortiz, R. J. F.; Palmeira, R. N.; Junior, E. P. P.; Aragão, E.; de Souza, L. E. P. F.; Netto, M. B.; Teixeira, M. G.; Barreto, M. L.; Ichihara, M. Y.; dos Reis Silva Lima, R. T. Medidas de distanciamento social no controle da pandemia de covid-19: potenciais impactos e desafios no brasil. 2423–2446. Acessado: 6 de Janeiro de 2023 `https://www.scielo.br/j/csc/a/4BHTCFF4bDqq4qT7WtPhvYr/?lang=pt`, Epub 05 Jun 2020. ISSN 1678-4561. https://doi.org/10.1590/1413-81232020256.1.10502020.

[5] Ayan, N.; Ramesh, A.; Seetharam, A.; de A. Rocha, A. A. Hierarchical models for detecting mobility clusters during covid-19. In *Proceedings of the 19th ACM International Symposium on Mobility Management and Wireless Access* (New York, NY, USA, 2021), MobiWac '21, Association for Computing Machinery, p. 43–51.

[6] Babalau, I.; Corlatescu, D.; Grigorescu, O.; Sandescu, C.; Dascalu, M. Severity prediction of software vulnerabilities based on their text description. In *23rd SYNASC* (2021), pp. 171–177.

[7] Bafna, P.; Pramod, D.; Vaidya, A. Document clustering: Tf-idf approach. In *ICEEOT* (2016), pp. 61–66.

[8] Berryman, A.; Calyam, P.; Honigford, M.; Lai, A. M. Vdbench: A benchmarking toolkit for thin-client based virtual desktop environments. In *2010 IEEE Second International Conference on Cloud Computing Technology and Science* (2010), pp. 480–487.

[9] Böttger, T.; Ibrahim, G.; Vallis, B. How the internet reacted to covid-19: A perspective from facebook's edge network. In *Proceedings of the ACM Internet Measurement Conference* (New York, NY, USA, 2020), IMC '20, Association for Computing Machinery, p. 34–41.

[10] Breiman, L. Random forests. *Machine Learning 45*, 1 (Oct 2001), 5–32.

[11] Chatterjee, A.; Saha, J.; Mukherjee, J. Clustering with multi-layered perceptron. *Pattern Recognition Letters 155* (2022), 92–99.

[12] Church, K.; de Oliveira, R. What's up with whatsapp? comparing mobile instant messaging behaviors with traditional sms. In *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services (MobileHCI '13)* (New York, NY, USA, 2013), pp. 352–361.

[13] Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, 1988.

[14] Crammer, K.; Dekel, O.; Keshet, J.; Shalev-Shwartz, S.; Singer, Y. Online passive-aggressive algorithms. *Journal of Machine Learning Research 7*, 19 (2006), 551–585.

[15] da Silva, D. R. P.; Stein, L. M. Segurança da informação: uma reflexão sobre o componente humano. *Ciências & Cognição 10* (abr. 2011), 46–53.

[16] da Silva, R. R. Home-officer: um surgimento bem-sucedido da profissão pós-fordista, uma alternativa positiva para os centros urbanos. *Revista Brasileira de Gestão Urbana 1*, 1 (2017), 85–94.

[17] de Niterói, P. M. Linha do tempo - ações da prefeitura para combate ao coronavírus - prefeitura municipal de niterói. Acessado: 11/10/2022 `http://www.niteroi.rj.gov.br/linha-do-tempo/`.

[18] do Governo do Estado de Santa Catarina, S. Coronavírus em sc: Governo do estado já emitiu mais de 120 mil alertas em sms à população sobre casos de covid-19. Acessado: 20/06/2020 `https://estado.sc.gov.br/noticias/coronavirus-em-sc-governo-do-estado-ja-emitiu-mais-de-120-mil-alertas-em-sms-a-populacao-sobre-casos-de-covid-19-2/`.

[19] Dreiseitl, S.; Ohno-Machado, L. Logistic regression and artificial neural network classification models: a methodology review. *Journal of Biomedical Informatics 35*, 5 (2002), 352–359.

[20] Drikvandi, R.; Lawal, O. Sparse principal component analysis for natural language processing. *Annals of Data Science 10*, 1 (Feb 2023), 25–41.

[21] Feldmann, A.; Gasser, O.; Lichtblau, F.; Pujol, E.; Poese, I.; Dietzel, C.; Wagner, D.; Wichtlhuber, M.; Tapiador, J.; Vallina-Rodriguez, N.; Hohlfeld, O.; Smaragdakis, G. The lockdown effect: Implications of the covid-19 pandemic on internet traffic. In *Proceedings of the ACM Internet Measurement Conference* (New York, NY, USA, 2020), IMC '20, Association for Computing Machinery, p. 1–18.

[22] Gold, H. Netflix and youtube are slowing down in europe to keep the internet from breaking. Acessado: 13/12/2020 `https://edition.cnn.com/2020/03/19/tech/netflix-Internet-overload-eu/index.html`.

[23] Goode, B. Voice over internet protocol (voip). *Proceedings of the IEEE 90*, 9 (2002), 1495–1517.

[24] Guimarães, P. R. B. *Métodos quantitativos estatísticos*, 1st ed. IESDE Brasil S.A., 2008.

[25] Gurung, S.; Kim, Y. Healthcare privacy: How secure are the voip/video-conferencing tools for phi data? In *2015 12th International Conference on Information Technology - New Generations* (2015), pp. 574–579.

[26] Hawwa, S. Audio mixing for centralized conferences in a sip environment. In *Proceedings. IEEE International Conference on Multimedia and Expo* (2002), vol. 2, pp. 269–272 vol.2.

[27] Hill, J.; Ford, W. R.; Farreras, I. G. Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior 49* (2015), 245–250.

[28] IBGE. Censo demográfico 2010 - domicílios particulares permanentes, por tipo do domicílio e número de moradores - resultados preliminares do universo, 2010. acessado em 20 de junho de 2020. https://sidra.ibge.gov.br/tabela/3152.

[29] Iezadi, S.; Azami-Aghdash, S.; Ghiasi, A.; Rezapour, A.; Pourasghari, H.; Pashazadeh, F.; Gholipour, K. Effectiveness of the non-pharmaceutical public health interventions against covid-19; a protocol of a systematic review and realist review. *PLOS ONE 15*, 9 (09 2020), 1–16.

[30] Isaacs, E.; Walendowski, A.; Whittaker, S.; Schiano, D. J.; Kamm, C. The character, functions, and styles of instant messaging in the workplace. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work (CSCW '02)* (New York, NY, USA, 2002), pp. 11–20.

[31] Iyengar, R. The coronavirus is stretching facebook to its limits. Acessado: 15/05/2021 `https://edition.cnn.com/2020/03/18/tech/zuckerberg-facebook-coronavirus-response/index.html`.

[32] Jackson, P.; van der Wielen, J. *Teleworking: New International Perspectives From Telecommuting to the Virtual Organisation.* Management of technology and innovation. Taylor & Francis, 2002.

[33] Joachims, T. Text categorization with support vector machines: Learning with many relevant features. In *European Conference on Machine Learning (ECML)* (Berlin, 1998), Springer, pp. 137–142.

[34] Kharbanda, E. O.; Stockwell, M.; Fox, H.; Andres, R.; Lara, M.; Rickert, V. Text messaging to promote hpv vaccination. *Journal of Adolescent Health 48* (2011), S4–S5.

[35] King, A.; Valença, A.; Silva, A.; Baczynski, T.; Carvalho, M.; Nardi, A. Nomophobia: Dependency on virtual environments or social phobia? *Computers in Human Behavior 29* (2013), 140–144.

[36] Lai, A.; Nieh, J. Limits of wide-area thin-client computing. *SIGMETRICS Perform. Eval. Rev. 30*, 1 (June 2002), 228–239.

[37] Lawas, J. B. R.; Vivero, A. C.; Sharma, A. Network performance evaluation of vpn protocols (sstp and ikev2). In *2016 Thirteenth International Conference on Wireless and Optical Communications Networks (WOCN)* (2016), pp. 1–5.

[38] Leevy, J. L.; Khoshgoftaar, T. M.; Bauder, R. A.; Seliya, N. A survey on addressing high-class imbalance in big data. *Journal of Big Data 5*, 1 (Nov 2018), 42.

[39] Lutu, A.; Perino, D.; Bagnulo, M.; Frias-Martinez, E.; Khangosstar, J. A characterization of the covid-19 pandemic impact on a mobile network operator traffic. In *Proceedings of the ACM Internet Measurement Conference* (New York, NY, USA, 2020), IMC '20, Association for Computing Machinery, p. 19–33.

[40] Ma, Z.; Zhou, J.; Zhang, L. A scalable framework for global application anycast. In *2009 Second International Conference on Information and Computing Science* (2009), vol. 1, pp. 250–253.

[41] Martins, V.; Müeller, P. Envio de sms para avisar quem mora perto de pessoas com coronavírus começa a funcionar em florianópolis. Acessado: 20/06/2020 `https://g1.globo.com/sc/santa-catarina/noticia/2020/04/01/envio-de-sms-para-avisar-quem-mora-perto-de-pessoas-com-coronavirus-comeca-a-funcionar-em-florianopolis.ghtml`.

[42] Meyer, D.; Leisch, F.; Hornik, K. The support vector machine under test. *Neurocomputing 55*, 1 (2003), 169–186. Support Vector Machines.

[43] Microsoft. [ms-rdpbcgr]: Remote desktop protocol: Basic connectivity and graphics remoting, 2021. acessado em 09 de outubro de 2021. https://docs.microsoft.com/en-us/openspecs/windows_protocols/ms-rdpbcgr/5073f4ed-1e93-45e1-b039-6e30c385867c?redirectedfrom=MSDN.

[44] Narayan, S.; Brooking, K.; de Vere, S. Network performance analysis of vpn protocols: An empirical comparison on different operating systems. In *2009 International Conference on Networks Security, Wireless Communications and Trusted Computing* (2009), vol. 1, pp. 645–648.

[45] Natkaniec, M.; Prasnal, L.; Szymakowski, M. A performance analysis of ieee 802.11ax networks. *International Journal of Electronics and Telecommunications vol. 66*, No 1 (2020), 225–230.

[46] NIC.br. Ix.br - brasil internet exchange - ix fórum regional - goiânia, go. Acessado: 17/12/2022 `https://regional.forum.ix.br/files/apresentacao/arquivo/262/IX.br%20-%20Brasil%20Internet%20Exchange%20-%20IX-Forum-Regional-Goiania-PMG%20-%2020180712.pdf`.

[47] NIC.br. Ix.br alcança marca de 10 tb/s de pico de tráfego internet. Acessado: 17/12/2022 `https://ix.br/noticia/releases/ix-br-alcanca-marca-de-10-tb-s-de-pico-de-trafego-Internet`.

[48] NIC.BR. Ix.br bate recorde histórico ao atingir 16 tbit/s de pico de tráfego internet. Acessado: 17/12/2022 https://ix.br/noticia/releases/ix-br-bate-recorde-historico-ao-atingir-16-tbit-s-de-pico-de-trafego-Internet.

[49] NIC.BR. Ix.br chega a 20 tbit/s de pico de tráfego, nova marca histórica. Acessado: 17/12/2022 https://ix.br/noticia/releases/ix-br-chega-a-20-tbit-s-de-pico-de-trafego-nova-marca-historica.

[50] NIC.BR. Ix.br completa 15 anos de operação, consolidado entre os maiores pontos de troca de tráfego internet do mundo. Acessado: 17/12/2022 https://ix.br/noticia/releases/ix-br-completa-15-anos-de-operacao-consolidado-entre-os-maiores-pontos-de-troca-de-trafego-Internet-do-mundo.

[51] NIC.BR. Ix.br ultrapassa marca de 8 tb/s de pico de tráfego internet. Acessado: 17/12/2022 https://ix.br/noticia/notas/ix-br-ultrapassa-marca-de-8-tb-s-de-pico-de-trafego-Internet.

[52] NIEH, J.; YANG, S. J.; NOVIK, N. Measuring thin-client performance using slow-motion benchmarking. 87–115.

[53] NIKONOV, A.; VULFIN, A.; VASILYEV, V.; KIRILLOVA, A.; MIKHAILOV, V. System for estimation cvss severity metrics of vulnerability based on text mining technology. In *ITNT* (2021), pp. 1–5.

[54] O'HARA, K. P.; MASSIMI, M.; HARPER, R.; RUBENS, S.; MORRIS, J. Everyday dwelling with whatsapp. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing (CSCW '14)* (New York, NY, USA, 2014), pp. 1131–1143.

[55] PAUL, S.; BOUTSIDIS, C.; MAGDON-ISMAIL, M.; DRINEAS, P. Random projections for linear support vector machines. *ACM Trans. Knowl. Discov. Data 8*, 4 (aug 2014).

[56] PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research 12* (2011), 2825–2830.

[57] PENG, C.-Y. J.; LEE, K. L.; INGERSOLL, G. M. An introduction to logistic regression analysis and reporting. *The Journal of Educational Research 96*, 1 (2002), 3–14.

[58] PETER MELL, KAREN SCARFONE, S. R. Cvss v2 complete documentation. Accessed: 2022-04-16 https://www.first.org/cvss/v2/guide.

[59] PUDELKO, M.; EMMERICH, P.; GALLENMÜLLER, S.; CARLE, G. Performance analysis of vpn gateways. In *2020 IFIP Networking Conference (Networking)* (2020), pp. 325–333.

[60] Reuben, J. S. A survey on virtual machine security. *Helsinki University of Technology 2*, 36 (2007).

[61] Rhee, J.; Kochut, A.; Beaty, K. Deskbench: Flexible virtual desktop benchmarking toolkit. In *2009 IFIP/IEEE International Symposium on Integrated Network Management* (2009), pp. 622–629.

[62] Rio, U. Oms disponibiliza número de whatsapp para esclarecer dúvidas sobre a covid-19. Acessado: 20/06/2020 https://unicrio.org.br/oms-disponibiliza-numero-de-whatsapp-para-esclarecer-duvidas-sobre-a-covid-19/.

[63] Rodrigues, R. Criminosos usam apps de reuniões online para infectar vítimas. Acessado: 15/08/2020 https://www.kaspersky.com.br/blog/online-reunioes-ataques-hackers/14724/.

[64] Rohr, A. Por que o zoom é alvo de desconfiança e quais são as alternativas para videoconferência? Acessado: 16/08/2020 https://g1.globo.com/economia/tecnologia/blog/altieres-rohr/post/2020/04/07/por-que-o-zoom-e-alvo-de-desconfianca-e-quais-sao-as-alternativas-para-videoconferencia.ghtml.

[65] Seo, S. Voip-telephone service: Economic efficiencies and policy implications. *Telematics and Informatics 25* (02 2008), 47–55.

[66] Shafranovich, Y. Rfc 4180 - common format and mime type for comma-separated values (csv) files. Acessado: 21/03/2022 https://datatracker.ietf.org/doc/html/rfc4180.

[67] Shahid, M. R.; Debar, H. CVSS-BERT: Explainable Natural Language Processing to Determine the Severity of a Computer Security Vulnerability from its Description. In *20th ICMLA* (Pasadena, United States, Dec. 2021).

[68] Shahzad, M.; Shafiq, M. Z.; Liu, A. X. A large scale exploratory analysis of software vulnerability life cycles. In *Proceedings of the 34th International Conference on Software Engineering* (2012), ICSE '12, IEEE Press, p. 771–781.

[69] Sharma, A.; Sharma, M.; Kr. Dwivedi, R. Exploratory data analysis and deception detection in news articles on social media using machine learning classifiers. *Ain Shams Engineering Journal* (2023), 102166.

[70] Sierra-Arriaga, F.; Branco, R.; Lee, B. Security issues and challenges for virtualization technologies. *ACM Comput. Surv. 53*, 2 (May 2020).

[71] Silva, M. A.; Ferreira, R.; Leite, O.; Macedo, S.; dos Santos, S. Segurança e confiabilidade para ambiente soho. *HOLOS 4*, 0 (2013), 66–76.

[72] Singh, A.; Prakash, B. S.; Chandrasekaran, K. A comparison of linear discriminant analysis and ridge classifier on twitter data. In *2016 International Conference on Computing, Communication and Automation (ICCCA)* (2016), pp. 133–138.

[73] Spataro, J. Our commitment to customers during covid-19. Acessado: 13/12/2020 https://www.microsoft.com/en-us/microsoft-365/blog/2020/03/05/our-commitment-to-customers-during-covid-19/.

[74] Systems, C. Ctrix hdx technologies, 2018. acessado em 09 de outubro de 2021. https://www.citrix.com/content/dam/citrix/en_us/documents/white-paper/citrix-hdx-technologies.pdf.

[75] Szilagyi, P. G.; Adams, W. G. Text Messaging: A New Tool for Improving Preventive Services. *JAMA 307*, 16 (04 2012), 1748–1749.

[76] Venkateswaran, R. Virtual private networks. *IEEE Potentials 20*, 1 (2001), 11–15.

[77] VMware. Vmware view 5 with pcoip: Network optimization guide, 2011. acessado em 09 de outubro de 2021. http://www.vmware.com/files/pdf/view/VMware-View-5-PCoIP-Network-Optimization-Guide.pdf.

[78] Wong, T.-T.; Yeh, P.-Y. Reliable accuracy estimates from k-fold cross validation. *IEEE Transactions on Knowledge and Data Engineering 32*, 8 (2020), 1586–1594.

[79] Wu, Z.; Xiao, M. Performance evaluation of vpn with different network topologies. In *2019 IEEE 2nd International Conference on Electronics Technology (ICET)* (2019), pp. 51–55.

[80] Yan, L. Development and application of desktop virtualization technology. In *2011 IEEE 3rd International Conference on Communication Software and Networks* (2011), pp. 326–329.

[81] Zhang, L.; Xu, C.; Pathak, P. H.; Mohapatra, P. Characterizing instant messaging apps on smartphones. In *Passive and Active Measurement* (Cham, 2015), J. Mirkovic and Y. Liu, Eds., Springer International Publishing, pp. 83–95.

[82] Época Negócios. Zoom entra na mira da justiça de nova york por problemas de segurança. Acessado: 16/08/2020 `https://epocanegocios.globo.com/Empresa/noticia/2020/04/zoom-entra-na-mira-da-justica-de-nova-york-por-problemas-de-seguranca.html`.

APÊNDICE A – Questionário da Primeira Pesquisa de Opinião

Seção 1

Idade [_____]

Sexo
[ ] Masculino
[ ] Feminino
[ ] Prefiro não dizer

Em relação ao trabalho, o que melhor se aplica ao seu caso?
[ ] A carga horária presencial continua a mesma de antes da pandemia
[ ] Após a pandemia o trabalho passou a ser integralmente em regime de home office
[ ] Após a pandemia o trabalho é em regime de revezamento com home office
[ ] Após a pandemia o trabalho é em regime de revezamento sem home office
[ ] Estava desempregado e assim continuo após a pandemia
[ ] Houve a demissão após a decretação da pandemia
[ ] Já trabalhava e continuo trabalhando em regime de home office
[ ] Aposentado(a)
[ ] Outros [_____]

Selecione seu estado

{Listagem de estados brasileiros}

Seções 2 a 28

Selecione seu município
{Listagem dos municípios para o estado brasileiro selecionado}

Seção 29

Qual é a quantidade de pessoas que moram com você? [_____]

Qual é a quantidade de dispositivos conectados à internet na sua residência (computadores, smartphones, tablets, smartTV, ...)? [_____]

Qual é o seu plano de internet fixa contratada?
[ ] Até 5Mbps
[ ] Entre 5Mbps e 10Mbps
[ ] Entre 10 Mbps e 20Mbps
[ ] Entre 20 Mbps e 100Mbps
[ ] Acima de 100Mbps

Quanto a estabilidade da conexão de internet (antes e após as medidas de isolamento social), você considera que:
[ ] Era e continua estável
[ ] Era estável e passou a apresentar instabilidade (interrupção, lentidão, ...)
[ ] Não era estável, mas melhorou após o isolamento social
[ ] Não era e continua não sendo estável

Quais são os aplicativos de mensagens instantâneas que você utiliza?
[ ] WhatsApp
[ ] Telegram
[ ] Facebook Messenger

[ ] Line
[ ] Viber
[ ] Nenhum
[ ] Outros [_____]

Quanto aos aplicativos de mensagens instantâneas, você considera que:
[ ] Não utilizava e continuo sem utilizar
[ ] Não utilizava e passei a utilizar
[ ] Já utilizava e não houve mudança na utilização
[ ] Já utilizava e houve um incremento na utilização
[ ] Já utilizava e houve um decremento na utilização
[ ] Já utilizava, mas não utilizo mais

Quais são os aplicativos de videoconferência / chamadas de vídeo que você utiliza?
[ ] Google Meet
[ ] Cisco Webex
[ ] Microsoft Teams
[ ] Zoom
[ ] Skype
[ ] Facebook Messenger
[ ] WhatsApp
[ ] Nenhum
[ ] Outros [_____]

Quanto aos aplicativos de videoconferência / chamadas de vídeo, você considera que:
[ ] Não utilizava e continuo sem utilizar
[ ] Não utilizava e passei a utilizar
[ ] Já utilizava e não houve mudança na utilização
[ ] Já utilizava e houve um incremento na utilização
[ ] Já utilizava e houve um decremento na utilização
[ ] Já utilizava, mas não utilizo mais

Antes das medidas de isolamento social, como você considera que era a sua intensidade na utilização da internet para os seguintes fins:

	Pouca	Razoável	Muita
Trabalho	[ ]	[ ]	[ ]
Estudo	[ ]	[ ]	[ ]
Lazer / Entretenimento	[ ]	[ ]	[ ]

Após as medidas de isolamento social, como você considera que é a sua intensidade na utilização da internet para os seguintes fins:

	Pouca	Razoável	Muita
Trabalho	[ ]	[ ]	[ ]
Estudo	[ ]	[ ]	[ ]
Lazer / Entretenimento	[ ]	[ ]	[ ]

APÊNDICE B – Questionário da Segunda Pesquisa de Opinião

Seção 1

Idade [_____]

Sexo
[ ] Masculino
[ ] Feminino
[ ] Prefiro não dizer

Selecione seu estado
{Listagem de estados brasileiros}

Seções 2 a 28

Selecione seu município
{Listagem dos municípios para o estado brasileiro selecionado}

Seção 29

Houve alteração no plano de internet fixa contratada?
[ ] Sim {Ir para a seção 30}
[ ] Não {Ir para a seção 31}

Seção 30

Qual era o plano de internet fixa contratada antes da decretação

da pandemia (Mbps)? [_____]

Qual é o novo plano de internet fixa contratada após a decretação da pandemia (Mbps)? [_____]

{Ir para a seção 32}

Seção 31

Qual é o seu plano de internet fixa contratada (Mbps)? [_____]

Seção 32

Quais são os aplicativos de videoconferência / chamadas de vídeo que você utiliza?
[ ] Google Meet
[ ] Cisco Webex
[ ] Microsoft Teams
[ ] Zoom
[ ] Skype
[ ] Facebook Messenger
[ ] WhatsApp
[ ] Nenhum
[ ] Outros [_____]

Em relação ao trabalho/estudo, marque as opções que melhor se aplicam à sua realidade:
[ ] Somente após a decretação da pandemia houve a utilização de ferramentas de videoconferência para o trabalho/estudo
[ ] Antes da decretação da pandemia já utilizava ferramentas de videoconferência para o trabalho/estudo
[ ] Usei as ferramentas de videoconferência para o trabalho/estudo durante a pandemia, mas já deixei de usá-las
[ ] Usei as ferramentas de videoconferência para o trabalho/estudo durante a pandemia e continuo usando com a mesma intensidade

[ ] Usei as ferramentas de videoconferência para o trabalho/ estudo durante a pandemia e continuo usando com a menos intensidade
[ ] Usei as ferramentas de videoconferência para o trabalho/ estudo durante a pandemia e continuo usando com a mais intensidade
[ ] Nunca usei ferramentas de videoconferência para o trabalho/ estudo

Comparando os períodos de mobilidade mais restritas da pandemia com o atual momento, a sua frequência/intensidade de uso de ferramentas de videoconferência para o trabalho/estudo:
[ ] Se mantém inalterada
[ ] Está aumentando
[ ] Está diminuindo
[ ] Não se aplica

Há, na mesma residência, outras pessoas que fazem uso de ferramentas de videoconferência?
[ ] Sim {Ir para a seção 33}
[ ] Não {Ir para a seção 34}

Seção 33

Para qual(is) atividade(s) as pessoas que residem com você utiliza(m) as ferramentas de videoconferência?
[ ] Trabalho
[ ] Estudo
[ ] Contato com família/amigos
[ ] Outros [_____]

Comparando os períodos de mobilidade mais restritas da pandemia com o atual momento, a frequência/intensidade de uso de ferramentas de videoconferência para dos demais moradores da mesma residência:
[ ] Se mantém inalterada
[ ] Está aumentando
[ ] Está diminuindo

[ ] Não se aplica

Seção 34

Comparando os períodos de mobilidade mais restritas da pandemia com o atual momento, você considera que sua conexão de internet:
[ ] Era e continua estável
[ ] Era estável e passou a apresentar instabilidade (interrupção, lentidão, ...)
[ ] Não era estável, mas melhorou
[ ] Não era e continua não sendo estável

Em sua residência há Wi-Fi (rede sem fio)?
[ ] Sim {Ir para a seção 35}
[ ] Não {Finalizar pesquisa}

Seção 35

Atualmente, como você classifica a intensidade na utilização da internet, considerando todos os moradores/frequentadores rotineiros com acesso ao Wi-Fi de sua residência, para os seguintes fins:

	Pouca	Razoável	Muita
Trabalho	[ ]	[ ]	[ ]
Estudo	[ ]	[ ]	[ ]
Lazer / Entretenimento	[ ]	[ ]	[ ]

Em relação ao Wi-Fi de sua residência, qual perfil mais se aplica?
[ ] A rede 2.4 GHz está disponível. Porém, a rede 5 GHz é utilizada majoritariamente
[ ] A rede 5 GHz está disponível. Porém, a rede 2.4 GHz é utilizada majoritariamente
[ ] Somente é utilizada a rede 2.4 GHz
[ ] Somente é utilizada a rede 5 GHz
[ ] Não sei informar

APÊNDICE C – Predição do *Base Score* a Partir da Descrição Textual da Vulnerabilidade

C.1 Metodologia de Predição

Nesta abordagem, foram implementados classificadores para fazer a predição do *base score* diretamente a partir da descrição textual da vulnerabilidade. Cada um desses classificadores foi modelado com base em um dos *datasets* e um dos algoritmos utilizados.

C.1.1 Classes

O *base score* é um valor discreto, arredondado à primeira casa decimal, que varia de 0,0 a 10,0. No entanto, alguns classificadores requerem classes com números inteiros. Assim, para essa abordagem, os valores do *base score* foram truncados, resultando em classes inteiras de 0 a 10.

C.2 Avaliação dos Resultados

O desempenho da predição do *base score* diretamente a partir da descrição textual da vulnerabilidade é ilustrado na Figura C.1. A melhor média da acurácia das validações cruzadas foi de 62%. Dentre os cinco melhores resultados, avaliando a média de acurácia dos *folds* das validações cruzadas, o algoritmo *Random Forest* aparece na duas primeiras posições, em primeiro lugar com treinamento realizado a partir do *dataset* balanceado e em segundo lugar com o treinamento realizado a partir do *dataset* proporcional. Os demais três resultados, foram treinados com o *dataset* balanceado, cada um com um algoritmo diferente. Na quarta posição o algoritmo *Random Forest* aparece novamente. Dessa vez, com a aplicação do algoritmo de redução de dimensionalidade LDA.

Tendo em vista a gradação numérica das classes (Subseção C.1.1), mesmo após o truncamento, é possível aplicar a métrica MSE para avaliar se as predições incorretas estão muito distantes do valor real (Fig. C.2). Com exceção do pior resultado, o MSE obtido não é alto. Isso significa que as predições incorretas não são excessivamente discrepantes do valor real. Também é possível constatar que, embora o modelo treinado a partir do *dataset* balanceado tenha apresentado melhor acurácia, o modelo treinado a partir do *dataset* proporcional apresentou um MSE mais baixo. Ao aplicar a métrica RMSE ao melhor resultado de MSE, modelo treinado com algoritmo *Random Forest* e *dataset* proporcional, obtemos o resultado de 1,4. Assim, em média, o erro na predição do *base score* é de 1,4 para mais ou para menos.

O gráfico plotado na Figura C.3 tem o objetivo de estimar a estabilidade e a capacidade de generalização do modelo. Nesse gráfico, foi avaliada a amplitude no desempenho dos dez *folds*, bem como o desempenho da classificação do *dataset* de teste. O classificador que obteve o melhor resultado nos testes foi aquele treinado com o algoritmo *Random Forest* e subconjunto, *fold*, F3 do *dataset* proporcional. Dentre os dez *folds* da validação cruzada, o pior desempenho aferido foi de 61% e o melhor de 63%. Dessa forma, a amplitude foi de, aproximadamente, 2 pontos percentuais. Já o desempenho verificado do teste foi de 51%, o que representa uma diferença de 12 pontos percentuais para o *fold* com melhor desempenho na validação.

Ao considerar as dez melhores médias dos resultados das predições dos *base scores*, diretamente a partir da descrição textual das vulnerabilidades, constata-se que a composição do *dataset* teve pouca influência no resultado final. Isso porque seis em dez desses melhores resultados utilizaram o *dataset* balanceado para treinamento do modelo. Nos demais 40%, os modelos foram treinados a partir do *dataset* proporcional. Chegamos à mesma conclusão ao analisar o desempenho das predições ao utilizar algoritmos de redução de dimensionalidade. O LDA está presente em 40% dos melhores dez resultados. Os 60% restantes não utilizam esse tipo de algoritmo. Entre os dez piores resultados, somente dois não usaram um dos dois algoritmos de redução de dimensionalidade abordados no presente estudo.

Conforme Figura C.4, o algoritmo *voting* obteve uma acurácia (média dos dez *folds*) de 61%, quando treinado com o *dataset* balanceado, e de 60%, quando treinado a partir do *dataset* proporcional. Sendo assim, o algoritmo *voting* não superou os resultados obtidos pelos modelos construídos a partir dos algoritmos de classificação de forma individual. A utilização do *dataset* balanceado implicou, exceto no *fold* F9, com o algoritmo *voting*, em

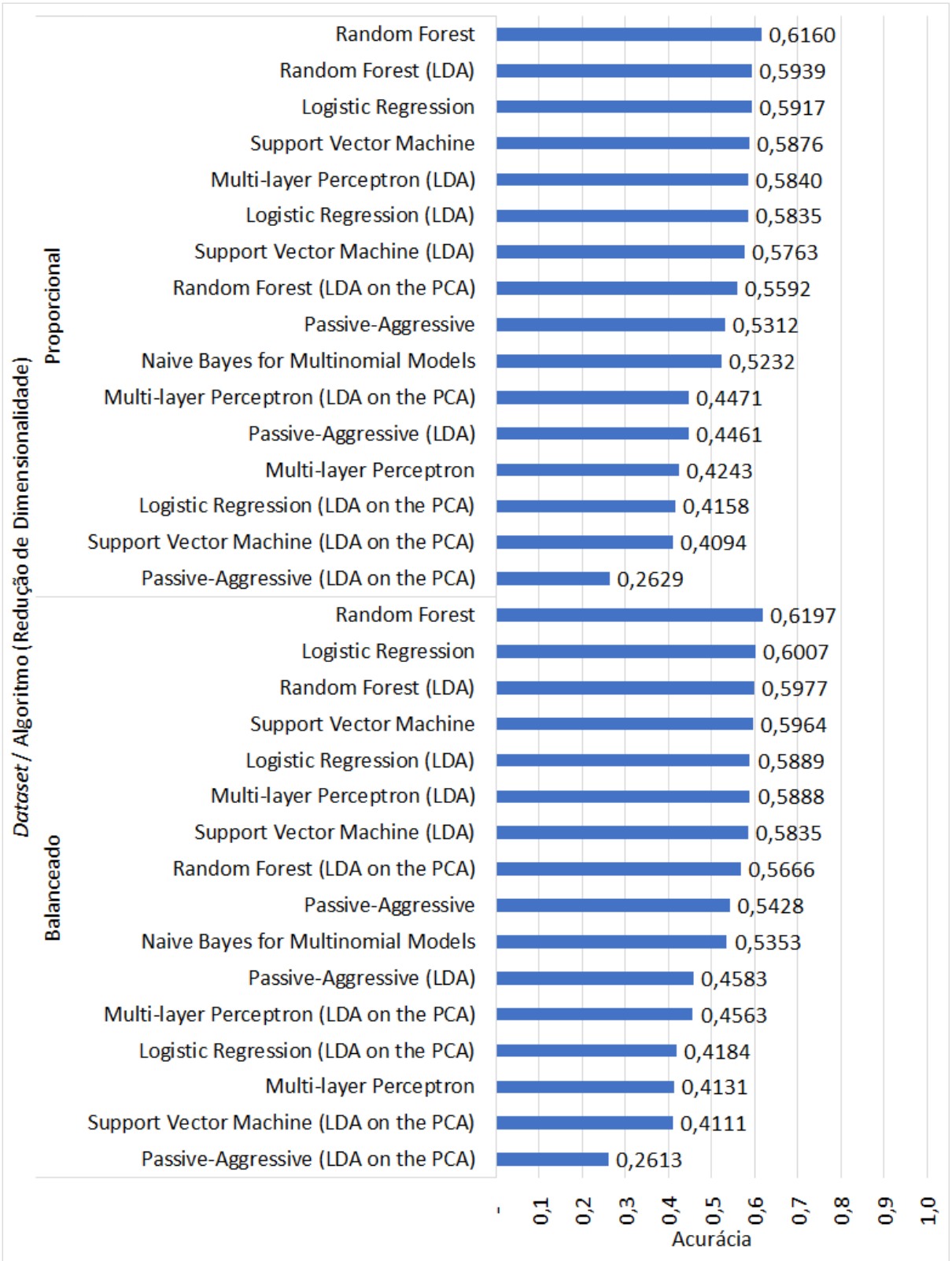


Figura C.1: Médias da acurácia na validação cruzada da predição do *base score* a partir da descrição textual da vulnerabilidade

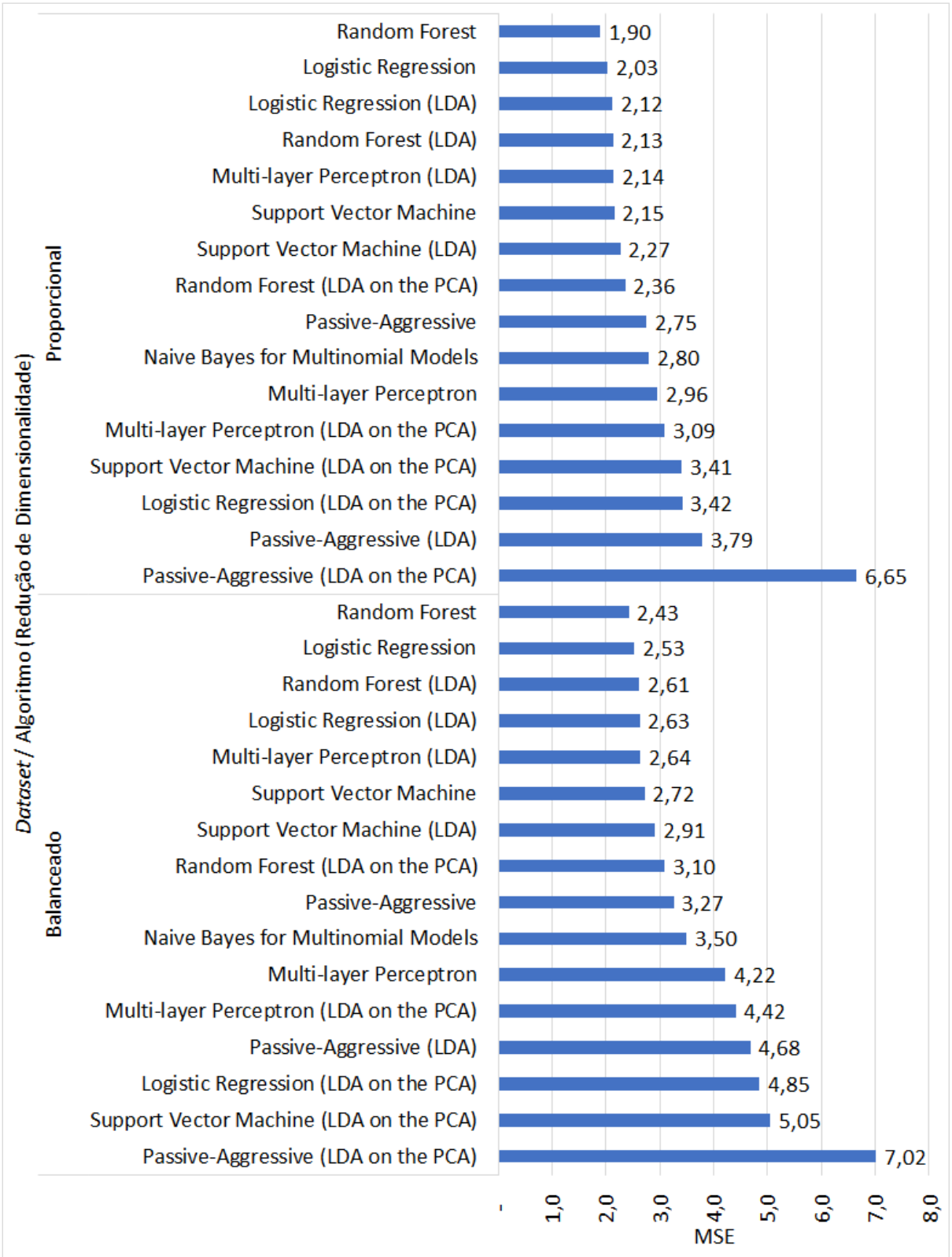


Figura C.2: Médias do MSE na validação cruzada da predição do *base score* a partir da descrição textual da vulnerabilidade

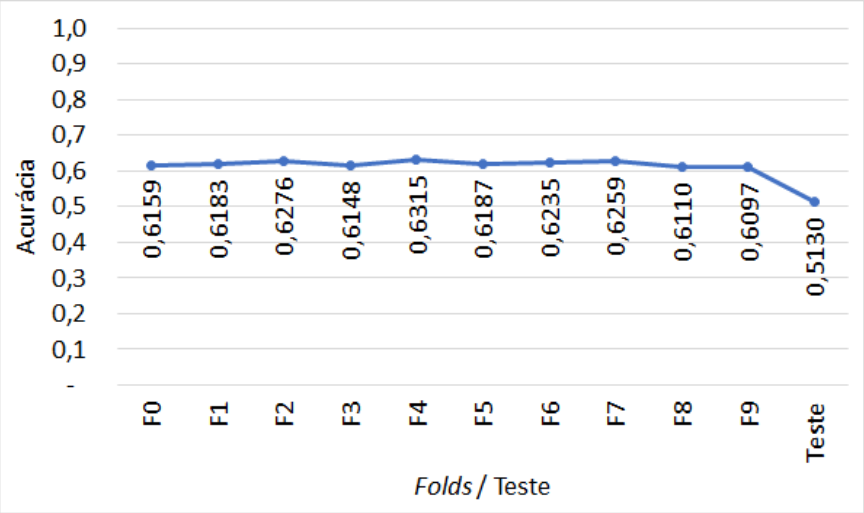


Figura C.3: Desempenho do teste e de cada *fold* do classificador que apresentou a melhor média dos *folds* para predição do *base score* a partir da descrição textual da vulnerabilidade (algoritmo *Random Forest*, *dataset* balanceado)

uma acurácia superior àquela obtida quando utilizado o *dataset* proporcional.

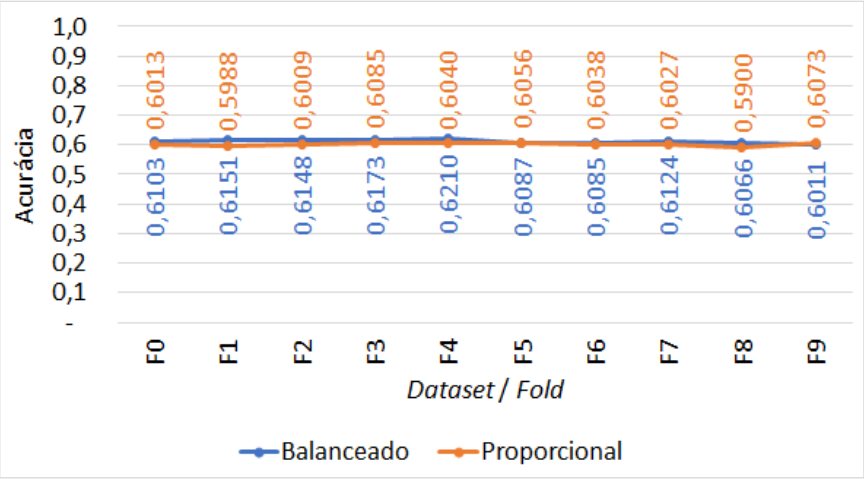
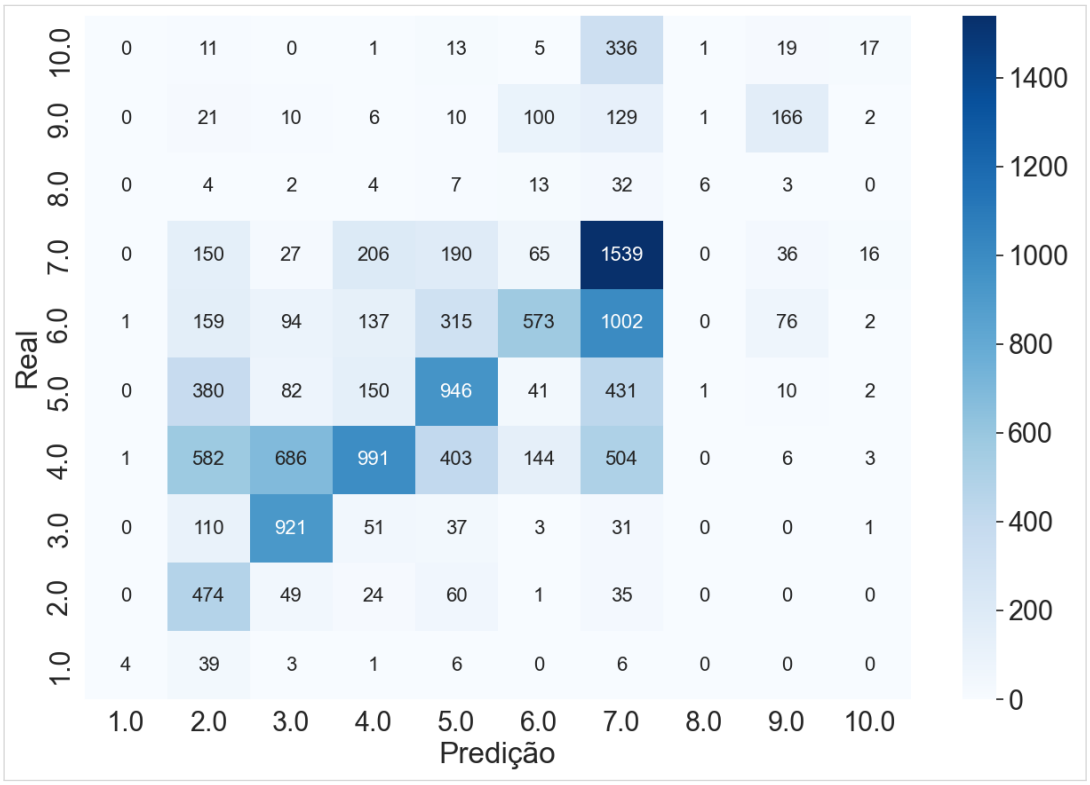
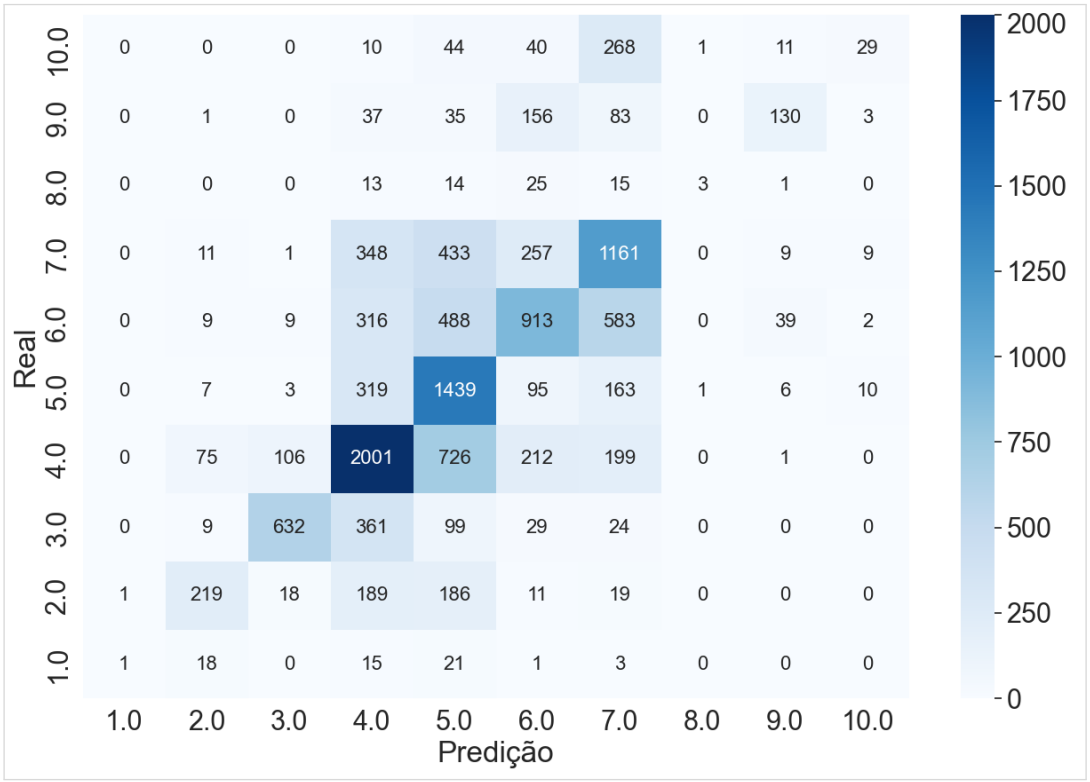


Figura C.4: Desempenho dos *folds* da validação cruzada para predição do *base score* a partir da descrição textual da vulnerabilidade utilizando o algoritmo *voting*

Analisando as matrizes de confusão do resultado do teste (Fig. C.5), observa-se, no teste realizado a partir do modelo treinado com o *dataset* proporcional, uma concentração mais próxima à diagonal, o que justifica um MSE de 2,1 e um RMSE de 1,4. Já no teste realizado a partir do modelo treinado com o *dataset* balanceado há uma dispersão um pouco maior. Em virtude disso, o MSE e o RMSE ficaram mais elevados, registrando os valores 3 e 1,7, respectivamente.



(a) Modelo gerado a partir do *dataset* balanceado



(b) Modelo gerado a partir do *dataset* proporcional

Figura C.5: Matriz de confusão do teste de classificação do *base score* a partir da descrição textual da vulnerabilidade

APÊNDICE D – Predição das Métricas do Vetor CVSS a Partir da Descrição Textual da Vulnerabilidade e Cálculo do *Base Score*

D.1 Metodologia de Predição

Nesta abordagem, foram implementados classificadores para fazer a predição de cada uma das métricas do vetor CVSS a partir da descrição textual da vulnerabilidade. De posse do valor predito de cada uma dessas métricas (*Access Vector*, *Access Complexity*, *Authentication*, *Confidentiality*, *Availability* e a *Integrity*), o *base score* da vulnerabilidade é calculado e arrendondado para uma casa decimal. Faz-se necessário relembrar que este estudo é baseado no CVSS versão 2 para permitir, também, a avaliação de vulnerabilidades mais antigas, para as quais não há avaliação registrada na versão 3 do CVSS. As principais diferenças entre as versões 2 e 3 são a alteração na pontuação de algumas métricas, como a *Confidentiality*, e a inclusão de novas métricas, como *User Interaction* e *Privileges Required*.

D.1.1 Classes

Para a predição das métricas do vetor CVSS, foram usados seis classificadores diferentes, um para cada uma das seis métricas — vários algoritmos são considerados para cada classificador. Cada métrica pode ter os seguintes valores categóricos (com valores numéricos correspondentes):

- ***Access Vector***: *Local* (L): 0; *Adjacent Network* (A): 1; e *Network* (N): 2;

- ***Access Complexity***: *High* (H): 0; *Medium* (M): 1; e *Low* (L): 2;
- ***Authentication***: *Multiple* (M): 0; *Single* (S): 1; e *None* (N): 2;
- ***Confidentiality***: *None* (N): 0; *Partial* (P): 1; e *Complete* (C): 2;
- ***Integrity***: *None* (N): 0; *Partial* (P): 1; e *Complete* (C): 2; e
- ***Availability***: *None* (N): 0; *Partial* (P): 1; e *Complete* (C): 2.

Tendo em vista que, nesse caso, o *base score* não é o produto dos algoritmos de classificação, mas obtido em função das métricas do vetor CVSS preditas, ele tem valor arredondado para uma casa decimal.

Considerando que o balanceamento e a proporcionalidade dos *datasets* de treinamento/validação são baseados nas classes da severidade, a Tabela D.1 demonstra a distribuição das classes de cada métrica do vetor CVSS para os *datasets* proporcional e balanceado.

D.2 Avaliação dos Resultados

D.2.1 Predição da Métrica *Access Vector* do Vetor CVSS Diretamente a Partir da Descrição Textual da Vulnerabilidade

O desempenho da predição da métrica *access vector* do vetor CVSS diretamente a partir da descrição textual da vulnerabilidade é ilustrado na Figura D.1. A melhor média da acurácia das validações cruzadas foi de 95%. Dentre os cinco melhores resultados, avaliando a média de acurácia dos *folds* das validações cruzadas, o algoritmo *Support Vector Machine* aparece na primeira e na quinta posições, sendo na última aplicado o algoritmo de redução de dimensionalidade. Os demais três resultados utilizaram algoritmos distintos. Todos os cinco melhores resultados foram treinados a partir do *dataset* proporcional.

O gráfico plotado na Figura D.2 tem o objetivo de estimar a estabilidade e a capacidade de generalização do modelo. Nesse gráfico foi avaliada a amplitude no desempenho dos dez *folds*, bem como o desempenho da classificação do *dataset* de teste. O classificador que obteve o melhor resultado nos testes foi aquele treinado com o algoritmo *Support Vector Machine* e o subconjunto, *fold*, F8 do *dataset* proporcional. Dentre os dez *folds* da validação cruzada, o pior desempenho aferido foi de 95% e o melhor de 96%. Dessa forma, a amplitude foi de, aproximadamente, 1 ponto percentual. Já o desempenho verificado

Tabela D.1: Distribuição das classes de cada métrica do vetor CVSS em cada tipo de *dataset*

Access Vector				
Classe	**Proporcional**		**Balanceado**	
	Qtde. de Amostras	**% de Amostras**	**Qtde. de Amostras**	**% de Amostras**
Local	7.365	14,4%	11.963	23,3%
Adjacent Network	1.227	2,4%	1.569	3,1%
Network	42.648	83,2%	37.708	73,6%
Access Complexity				
Classe	**Proporcional**		**Balanceado**	
	Qtde. de Amostras	**% de Amostras**	**Qtde. de Amostras**	**% de Amostras**
High	3.348	6,5%	4.210	8,2%
Medium	19.834	38,7%	19.288	37,6%
Low	28.058	54,8%	27.742	54,1%
Authentication				
Classe	**Proporcional**		**Balanceado**	
	Qtde. de Amostras	**% de Amostras**	**Qtde. de Amostras**	**% de Amostras**
Multiple	30	0,1%	43	0,1%
Single	7.746	15,1%	10.620	20,7%
None	43.464	84,8%	40.577	79,2%
Confidentiality				
Classe	**Proporcional**		**Balanceado**	
	Qtde. de Amostras	**% de Amostras**	**Qtde. de Amostras**	**% de Amostras**
None	16.688	32,6%	19.012	37,1%
Partial	25.356	49,5%	22.836	44,6%
Complete	9.196	17,9%	9.392	18,3%
Integrity				
Classe	**Proporcional**		**Balanceado**	
	Qtde. de Amostras	**% de Amostras**	**Qtde. de Amostras**	**% de Amostras**
None	15.178	29,6%	16.145	31,5%
Partial	27.127	52,9%	25.940	50,6%
Complete	8.935	17,4%	9.155	17,9%
Availability				
Classe	**Proporcional**		**Balanceado**	
	Qtde. de Amostras	**% de Amostras**	**Qtde. de Amostras**	**% de Amostras**
None	18.593	36,3%	22.222	43,4%
Partial	21.949	42,8%	18.452	36,0%
Complete	10.698	20,9%	10.566	20,6%

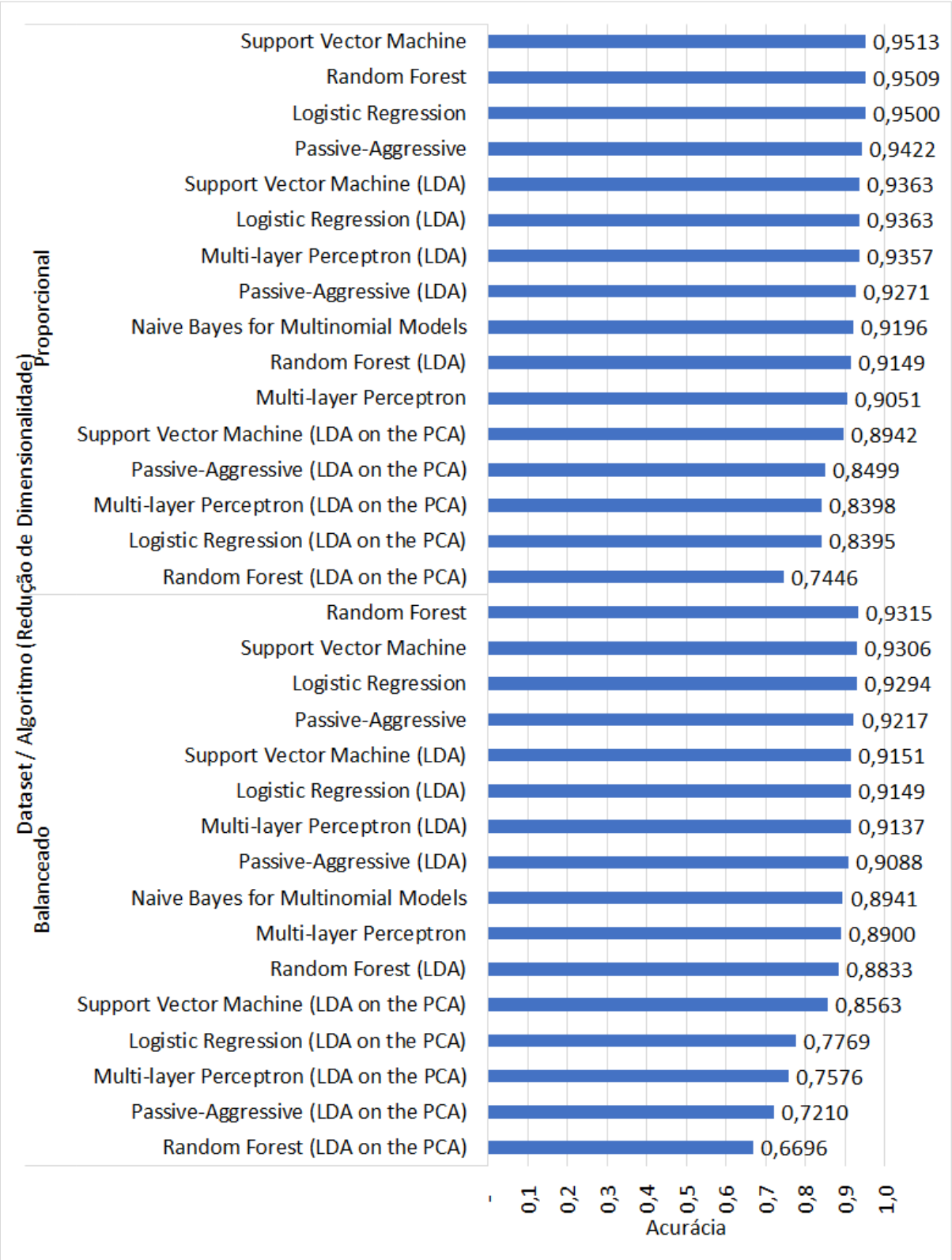


Figura D.1: Médias de acurácia da validação cruzada da predição da métrica *access vector* do vetor CVSS a partir da descrição textual da vulnerabilidade

do teste foi de 92%, o que representa uma diferença de 4 pontos percentuais para o *fold* com melhor desempenho na validação.

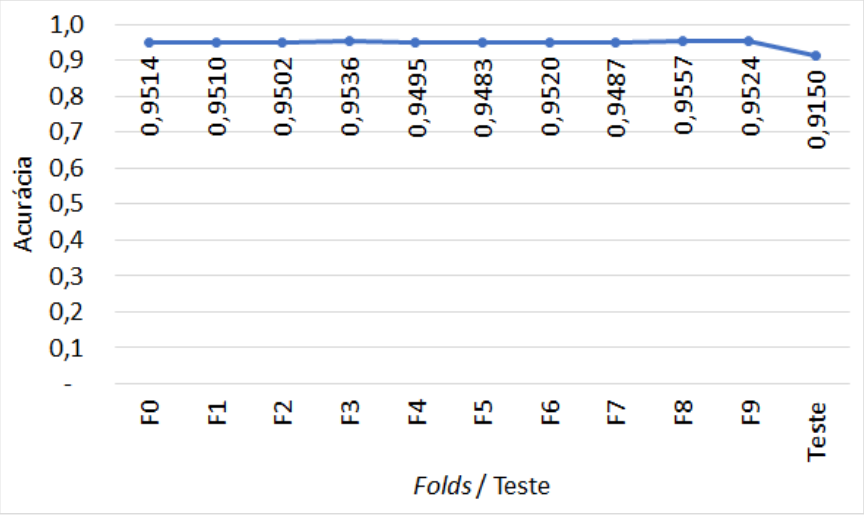


Figura D.2: Desempenho do teste e de cada *fold* do classificador que apresentou a melhor média para predição da métrica *access vector* do vetor CVSS a partir da descrição textual da vulnerabilidade (algoritmo *Support Vector Machine*, *dataset* proporcional)

Ao considerar as dez melhores médias dos resultados das predições da métrica *access vector* do vetor CVSS, diretamente a partir da descrição textual das vulnerabilidades, constata-se que a composição do *dataset* influencia o resultado final. Isso porque apenas três em dez desses melhores resultados utilizaram o *dataset* balanceado para treinamento do modelo. Nos demais 70%, os modelos foram treinados a partir do *dataset* proporcional. Chegamos à mesma conclusão ao analisar o desempenho das predições ao utilizar algoritmos de redução de dimensionalidade. O LDA está presente em 30% dos melhores dez resultados. Os 70% restantes não utilizam esse tipo de algoritmo. Entre os dez piores resultados, todos usam um dos dois algoritmos de redução de dimensionalidade abordados no presente estudo.

Com o objetivo de tentar refinar os resultados da previsão, os três melhores algoritmos (*Random Forest*, *Logistic Regression* e *Support Vector Machine*) foram combinados, utilizando o classificador *Voting*. A predição usando *Voting* foi feita no modo *hard*, utilizando os dois *datasets* para treinamento e aplicando a validação cruzada.

Conforme Figura D.3, o algoritmo *voting* obteve a acurácia (média dos dez *folds*) de 95% quando treinado a partir do *dataset* proporcional, e de 93%, quando treinado a partir do *dataset* balanceado. Sendo assim, o algoritmo *voting* superou os resultados obtidos pelos modelos construídos a partir dos algoritmos de classificação de forma individual em menos de 1 ponto percentual. A utilização do *dataset* balanceado implicou, em todos os

casos, com o algoritmo *voting*, em uma acurácia inferior àquela obtida quando utilizado o *dataset* proporcional.

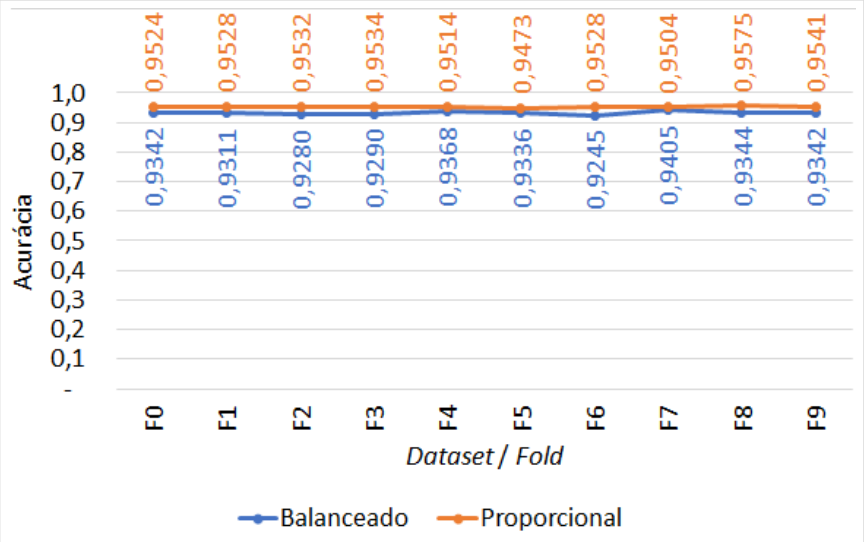
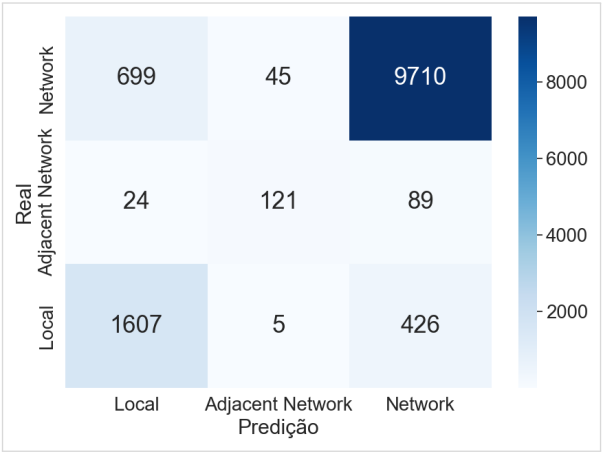
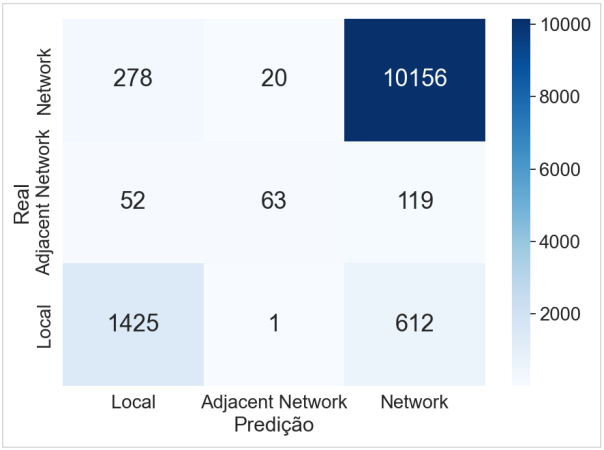


Figura D.3: Desempenho do teste e de cada *fold* do classificador que apresentou a melhor média dos *folds* para predição da métrica *access vector* do vetor CVSS a partir da descrição textual da vulnerabilidade utilizando o algoritmo *voting*

Analisando as matrizes de confusão do resultado do teste (Fig. D.4), conclui-se que tanto o modelo treinado a partir do *dataset* proporcional quanto o modelo treinado a partir do *dataset* balanceado tiveram mais dificuldade na predição da classe *adjacent network*. Porém, essa dificuldade foi maior no modelo treinado a partir do *dataset* proporcional.



(a) Modelo gerado a partir do *dataset* balanceado



(b) Modelo gerado a partir do *dataset* proporcional

Figura D.4: Matriz de confusão do teste de classificação da métrica *access vector* do vetor CVSS a partir da descrição textual da vulnerabilidade

D.2.2 Predição da Métrica *Access Complexity* do Vetor CVSS Diretamente a Partir da Descrição Textual da Vulnerabilidade

O desempenho da predição da métrica *access complexity* do vetor CVSS diretamente a partir da descrição textual da vulnerabilidade é ilustrado na Figura D.5. A melhor média da acurácia das validações cruzadas foi de 82%. Dentre os cinco melhores resultados, avaliando a média de acurácia dos *folds* das validações cruzadas, o algoritmo *Random Forest* ocupa as duas primeiras posições. Na primeira treinado a partir do *dataset* proporcional e na segunda a partir do *dataset* balanceado. O algoritmo *Support Vector Machine* também aparece duas vezes entre os cinco melhores: na terceira posição, treinado a partir do *dataset* proporcional e na quinta posição, treinado a partir do *dataset* balanceado. A quarta posição foi ocupada pelo algoritmo *Logistic Regression* treinado a partir do *dataset* proporcional do *dataset* proporcional.

Para estimar a estabilidade e a capacidade de generalização do modelo foi plotado o gráfico constante na Figura D.6. Nesse gráfico foi avaliada a amplitude no desempenho dos dez *folds*, bem como o desempenho da classificação do *dataset* de teste. O classificador que obteve o melhor resultado nos testes foi aquele treinado com o algoritmo *Random Forest* e o subconjunto, *fold*, F3 do *dataset* proporcional. Dentre os dez *folds* da validação cruzada, o menor desempenho aferido foi de 80% e o maior de 82%. Dessa forma, a amplitude foi de, aproximadamente, 2 pontos percentuais. O desempenho verificado do teste foi um caso particular, que superou o resultado da validação, registrando 85% de acurácia, o que representa uma diferença de aproximadamente 3 pontos percentuais a maior em relação ao *fold* com melhor desempenho na validação, utilizado para classificar o *dataset* de teste.

Ao considerar as dez melhores médias dos resultados das predições da métrica *access complexity* do vetor CVSS, diretamente a partir da descrição textual das vulnerabilidades, constata-se as seguintes proporções em relação ao *dataset*: 60% proporcional e 40% balanceado. Também foi possível observar que os seis melhores resultados não utilizaram algoritmos de redução de dimensionalidade, seguidos de quatro que usaram o algoritmo de redução de dimensionalidade LDA. Entre os dez piores resultados, todos usam o algoritmo de redução de dimensionalidade LDA on the PCA.

De acordo com a Figura D.7, o algoritmo *voting* alcançou a acurácia (média dos dez *folds*), de 81,7% quando treinado a partir do *dataset* proporcional, e de 81,5%, quando treinado a partir do *dataset* balanceado. Sendo assim, o algoritmo *voting* não superou os resultados obtidos pelos modelos construídos a partir dos algoritmos de classificação de

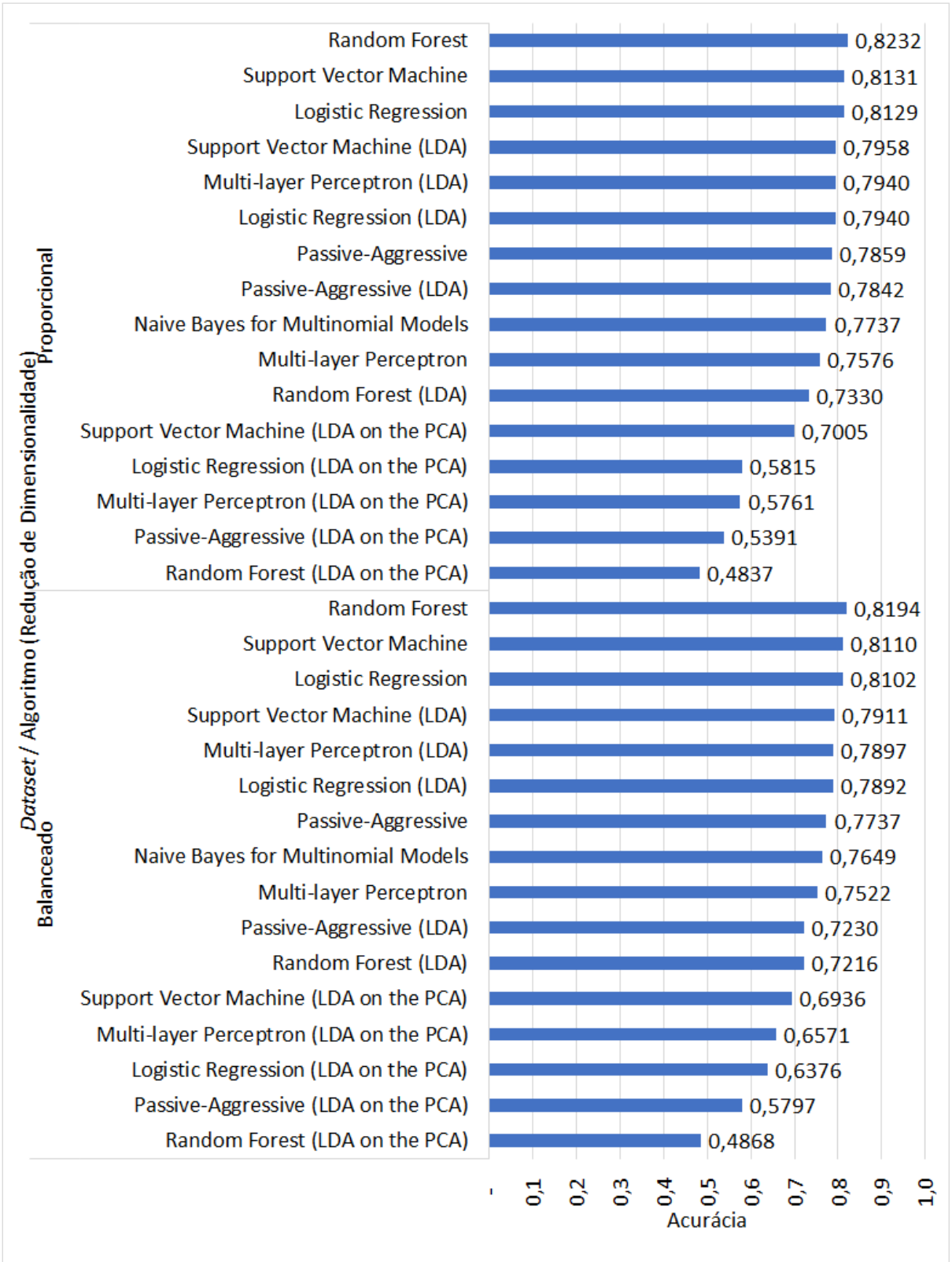


Figura D.5: Médias de acurácia da validação cruzada da predição da métrica *access complexity* do vetor CVSS a partir da descrição textual da vulnerabilidade

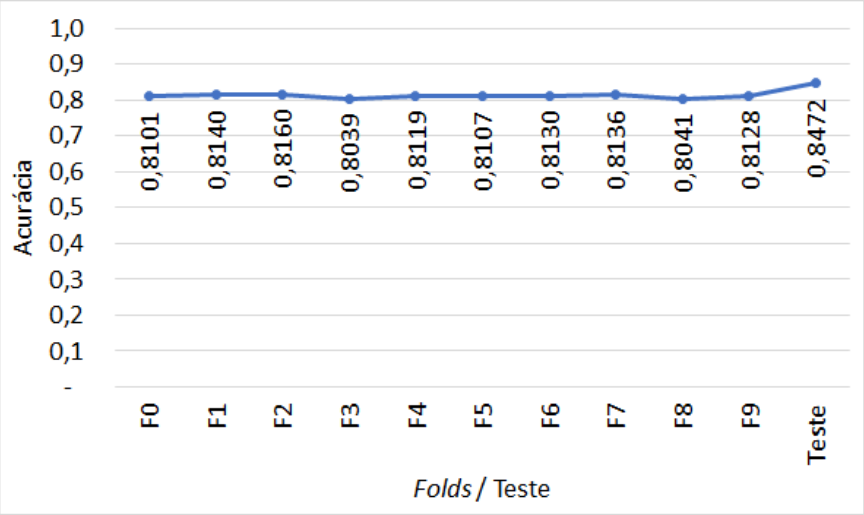


Figura D.6: Desempenho do teste e de cada *fold* do classificador que apresentou a melhor média dos *folds* para predição da métrica *access complexity* do vetor CVSS a partir da descrição textual da vulnerabilidade (algoritmo *Random Forest*, *dataset* proporcional)

forma individual. A utilização do *dataset* balanceado, com o algoritmo *voting*, somente superou a acurácia da utilização do *dataset* proporcional em 40% dos *folds*.

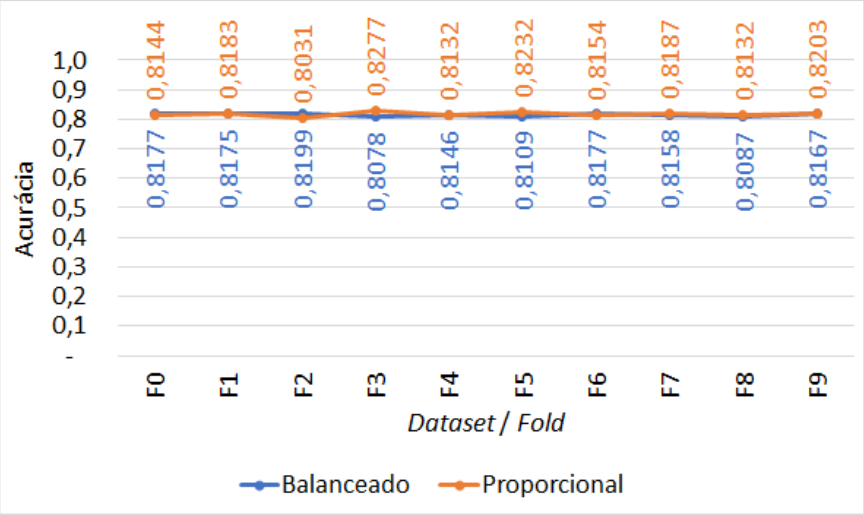
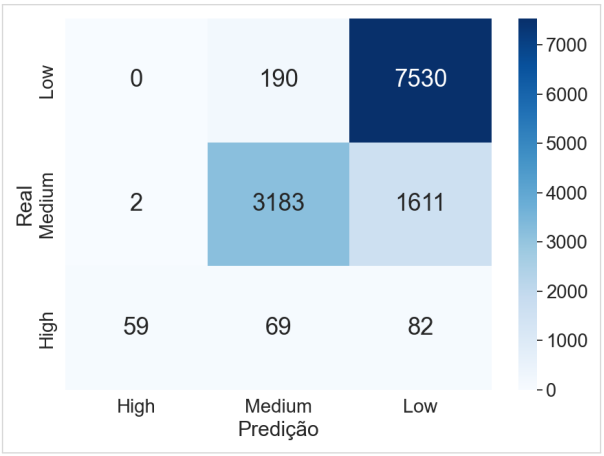
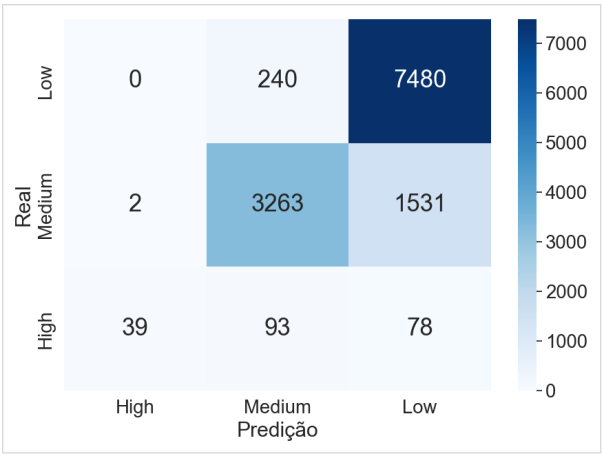


Figura D.7: Desempenho do teste e de cada *fold* do classificador que apresentou a melhor média dos *folds* para predição da métrica *access complexity* do vetor CVSS a partir da descrição textual da vulnerabilidade utilizando o algoritmo *voting*

Verificando as matrizes de confusão relativas ao resultado do teste (Fig. D.8) verifica-se que tanto o modelo treinado a partir do *dataset* proporcional quanto o modelo treinado a partir do *dataset* balanceado tiveram mais dificuldade na predição da classe *high*. As predições dessa classe foram distribuídas entre todas as classes de forma similar. A classe com o melhor desempenho foi a *low*, que tem o maior número de amostras, seguida da classe *medium*.



(a) Modelo gerado a partir do *dataset* balanceado



(b) Modelo gerado a partir do *dataset* proporcional

Figura D.8: Matriz de confusão do teste de classificação da métrica *access complexity* do vetor CVSS a partir da descrição textual da vulnerabilidade

D.2.3 Predição da Métrica *Authentication* do Vetor CVSS Diretamente a Partir da Descrição Textual da Vulnerabilidade

O desempenho da predição da métrica *authentication* do vetor CVSS diretamente a partir da descrição textual da vulnerabilidade é ilustrado na Figura D.9. A melhor média da acurácia das validações cruzadas foi de 93%. Dentre os cinco melhores resultados, avaliando a média de acurácia dos *folds* das validações cruzadas, o algoritmo *Random Forest* aparece na primeira e na terceira posição e o algoritmo *Logistic Regression* ocupa a segunda e a quarta posição. A quinta posição ficou para o algoritmo *Support Vector Machine*. O *dataset* proporcional foi utilizado no treinamento em três dos cinco melhores resultados, os dois demais foram treinados a partir do *dataset* balanceado. Porém, o melhor resultado teve o modelo treinado a partir do *dataset* balanceado, em que pese a diferença para o modelo treinado a partir do *dataset* proporcional ser inferior a 1 ponto percentual.

O gráfico plotado na Figura D.10 tem o objetivo de estimar a estabilidade e a capacidade de generalização do modelo. Nesse gráfico foi avaliada a amplitude no desempenho dos dez *folds*, bem como o desempenho da classificação do *dataset* de teste. O classificador que obteve o melhor resultado nos testes foi aquele treinado com o algoritmo *Logistic Regression* e o subconjunto, *fold*, F4 do *dataset* proporcional. Dentre os dez *folds* da validação cruzada, o menor desempenho aferido foi de 92,7% e o maior de 93,4%. Dessa forma, a amplitude foi inferior a 1 ponto percentual. Já o desempenho verificado do teste foi de 85,3%, o que representa uma diferença de aproximadamente 8 pontos percentuais

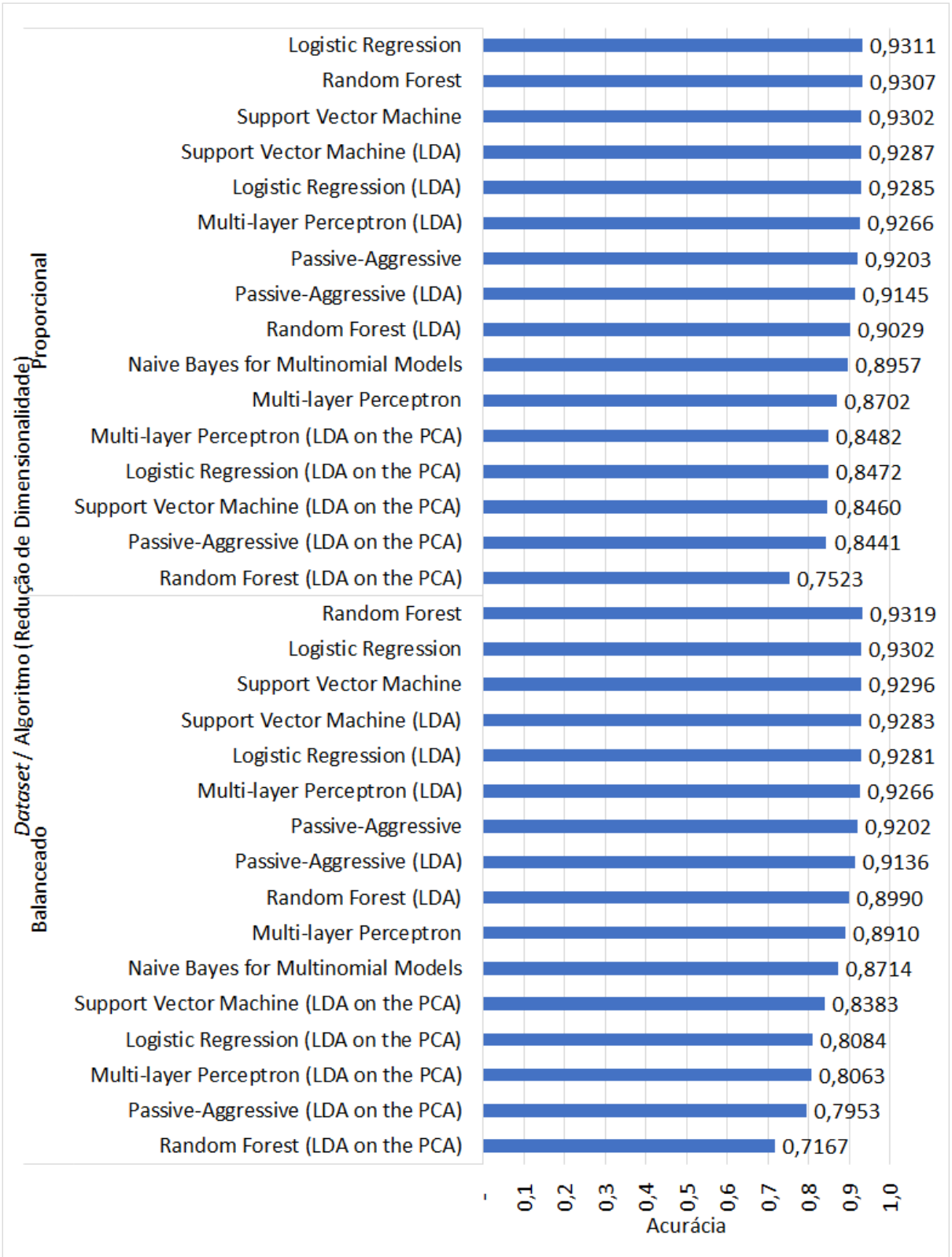


Figura D.9: Médias de acurácia da validação cruzada da predição da métrica *authentication* do vetor CVSS a partir da descrição textual da vulnerabilidade

para o *fold* com melhor desempenho na validação, utilizado para classificar o *dataset* de teste.

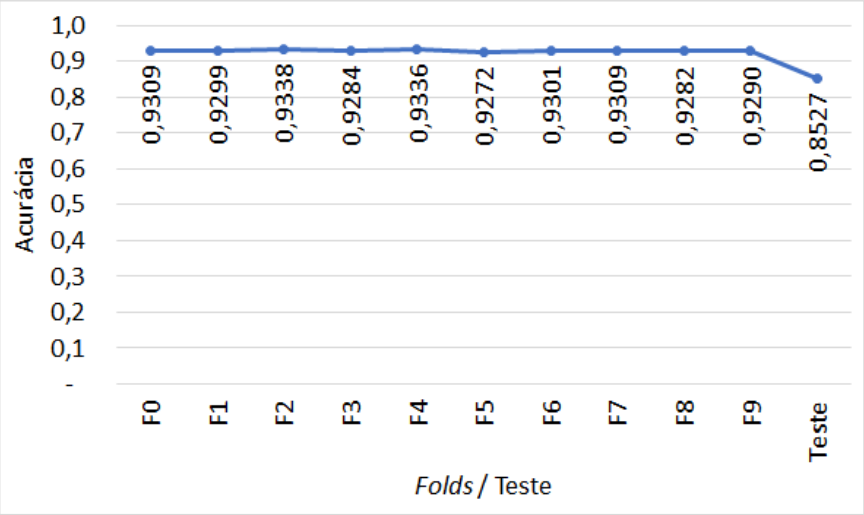


Figura D.10: Desempenho do teste e de cada *fold* do classificador que apresentou a melhor média dos *folds* para predição da métrica *authentication* do vetor CVSS a partir da descrição textual da vulnerabilidade (algoritmo *Random Forest*, *dataset* balanceado)

Ao considerar as dez melhores médias dos resultados das predições da métrica *authentication* do vetor CVSS, diretamente a partir da descrição textual das vulnerabilidades, constata-se que a composição do *dataset* não influencia o resultado final. Isso porque apenas cinco em dez desses melhores resultados utilizaram o *dataset* balanceado para treinamento do modelo. Nos demais 50%, os modelos foram treinados a partir do *dataset* proporcional. Além disso, a diferença no desempenho entre o melhor e o décimo melhor resultado foi de 0,4 ponto percentual. Quanto aos algoritmos de redução de dimensionalidade, O LDA foi usado do sétimo ao décimo melhores resultados. Entre os dez piores resultados, todos usam o algoritmo de redução de dimensionalidade LDA on the PCA.

Em consonância com a Figura D.11, o algoritmo *voting* registrou a acurácia (média dos dez *folds*), de 93% independentemente do *dataset* a partir do qual o algoritmo foi treinado. O algoritmo *voting* superou os resultados obtidos pelos modelos construídos a partir dos algoritmos de classificação de forma individual em 0,1 ponto percentual, especificamente utilizando o *dataset* proporcional. A utilização do *dataset* balanceado implicou, com o algoritmo *voting*, somente obteve uma acurácia melhor àquela obtida quando utilizado o *dataset* proporcional em 20% dos *folds*.

Examinando as matrizes de confusão (Fig. D.12) é possível verificar que tanto o modelo treinado a partir do *dataset* proporcional quanto o modelo treinado a partir do *dataset* balanceado tiveram mais dificuldade na predição da classe *single*. Quanto à classe *none*,

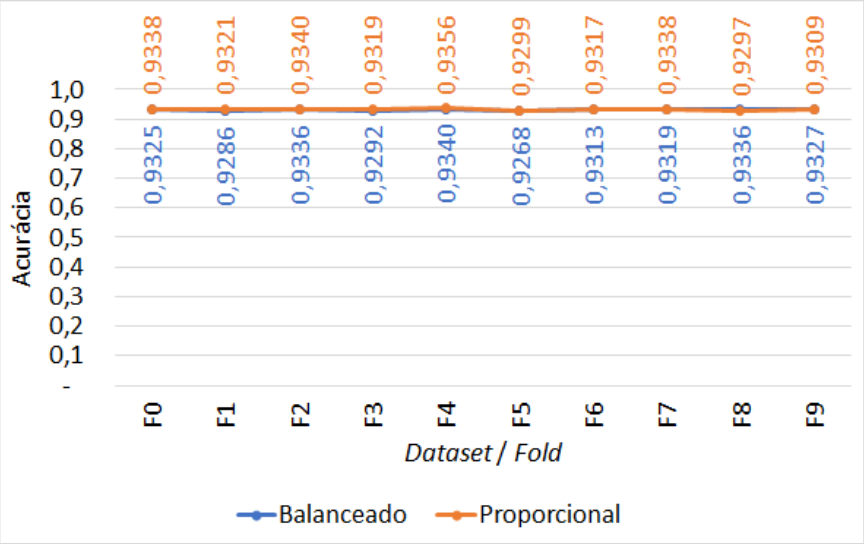
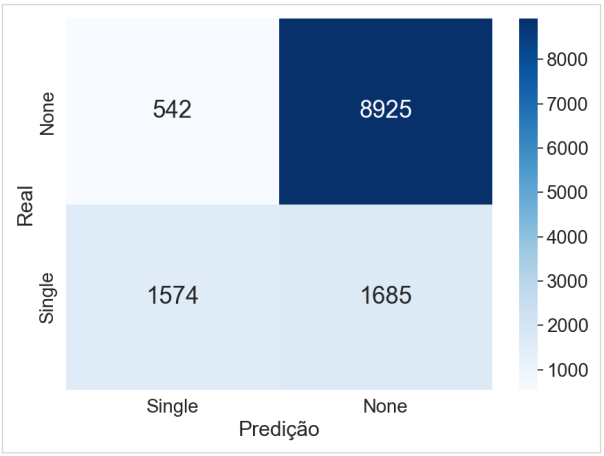
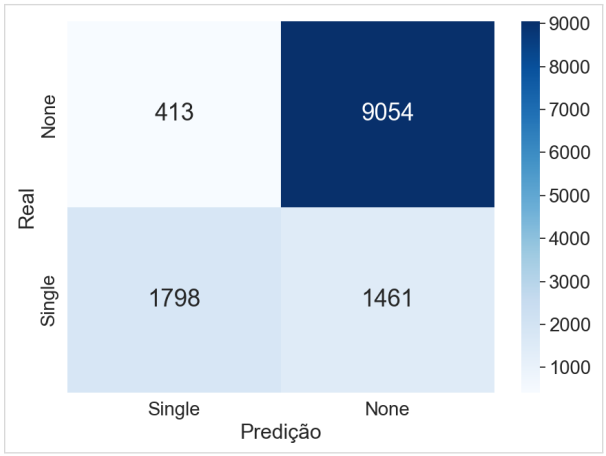


Figura D.11: Desempenho dos *folds* da validação cruzada para predição da métrica *authentication* do vetor CVSS a partir da descrição textual da vulnerabilidade utilizando o algoritmo *voting*

suas predições foram distribuídas equilibradamente entre as classes *single* e *none*.



(a) Modelo gerado a partir do *dataset* balanceado



(b) Modelo gerado a partir do *dataset* proporcional

Figura D.12: Matriz de confusão do teste de classificação da métrica *authentication* do vetor CVSS a partir da descrição textual da vulnerabilidade

D.2.4 Predição da Métrica *Confidentiality* do Vetor CVSS Diretamente a Partir da Descrição Textual da Vulnerabilidade

O desempenho da predição da métrica *confidentiality* do vetor CVSS diretamente a partir da descrição textual da vulnerabilidade é ilustrado na Figura D.13. A melhor média da acurácia das validações cruzadas foi de 84%. Dentre os cinco melhores resultados,

avaliando a média de acurácia dos *folds* das validações cruzadas, observa-se que os três melhores resultados foram treinados a partir do *dataset* balanceado e os dois seguintes treinados a partir do *dataset* proporcional. O algoritmo *Random Forest* ocupa a primeira e quarta posição, o algoritmo *Logistic Regression* a segunda e quinta posição e o algoritmo *Support Vector Machine* aparece na terceira posição.

O gráfico plotado na Figura D.14 tem o objetivo de estimar a estabilidade e a capacidade de generalização do modelo. Nesse gráfico foi avaliada a amplitude no desempenho dos dez *folds*, bem como o desempenho da classificação do *dataset* de teste. O classificador que obteve o melhor resultado nos testes foi aquele treinado com o algoritmo *Random Forest* e o subconjunto, *fold*, F8 do *dataset* proporcional. Dentre os dez *folds* da validação cruzada, o menor desempenho aferido foi de 83% e o maior de 85%. Dessa forma, a amplitude foi de, aproximadamente, 2 pontos percentuais. Já o desempenho verificado do teste foi de 80%, o que representa uma diferença de 5 pontos percentuais para o *fold* com melhor desempenho na validação, utilizado para classificar o *dataset* de teste.

Ao considerar as dez melhores médias dos resultados das predições da métrica *confidentiality* do vetor CVSS, diretamente a partir da descrição textual das vulnerabilidades, constata-se que as proporções em relação ao emprego de *datasets* são iguais, ou seja: 50% proporcional e 50% balanceado. Nenhum dos dez melhores resultados usa algoritmo de redução de dimensionalidade. No entanto, entre os dez piores resultados todos usam o algoritmo de redução de dimensionalidade LDA on the PCA.

Em conformidade com a Figura D.15, o algoritmo *voting* obteve a acurácia (média dos dez *folds*), de 84% quando treinado a partir do *dataset* balanceado, e de 83%, quando treinado a partir do *dataset* proporcional. O algoritmo *voting* superou os resultados obtidos pelos modelos construídos a partir dos algoritmos de classificação de forma individual em 0,04 ponto percentual. A utilização do *dataset* balanceado implicou, em 90% dos *folds*, com o algoritmo *voting*, em uma acurácia superior àquela obtida quando utilizado o *dataset* proporcional.

Consultando as matrizes de confusão (Fig. D.16) depreende-se que tanto o modelo treinado a partir do *dataset* proporcional quanto o modelo treinado a partir do *dataset* balanceado tiveram mais dificuldade na predição da classe *complete*. As predições incorretas dessa classe foram preditas, na maior parte dos casos, como *partial*.

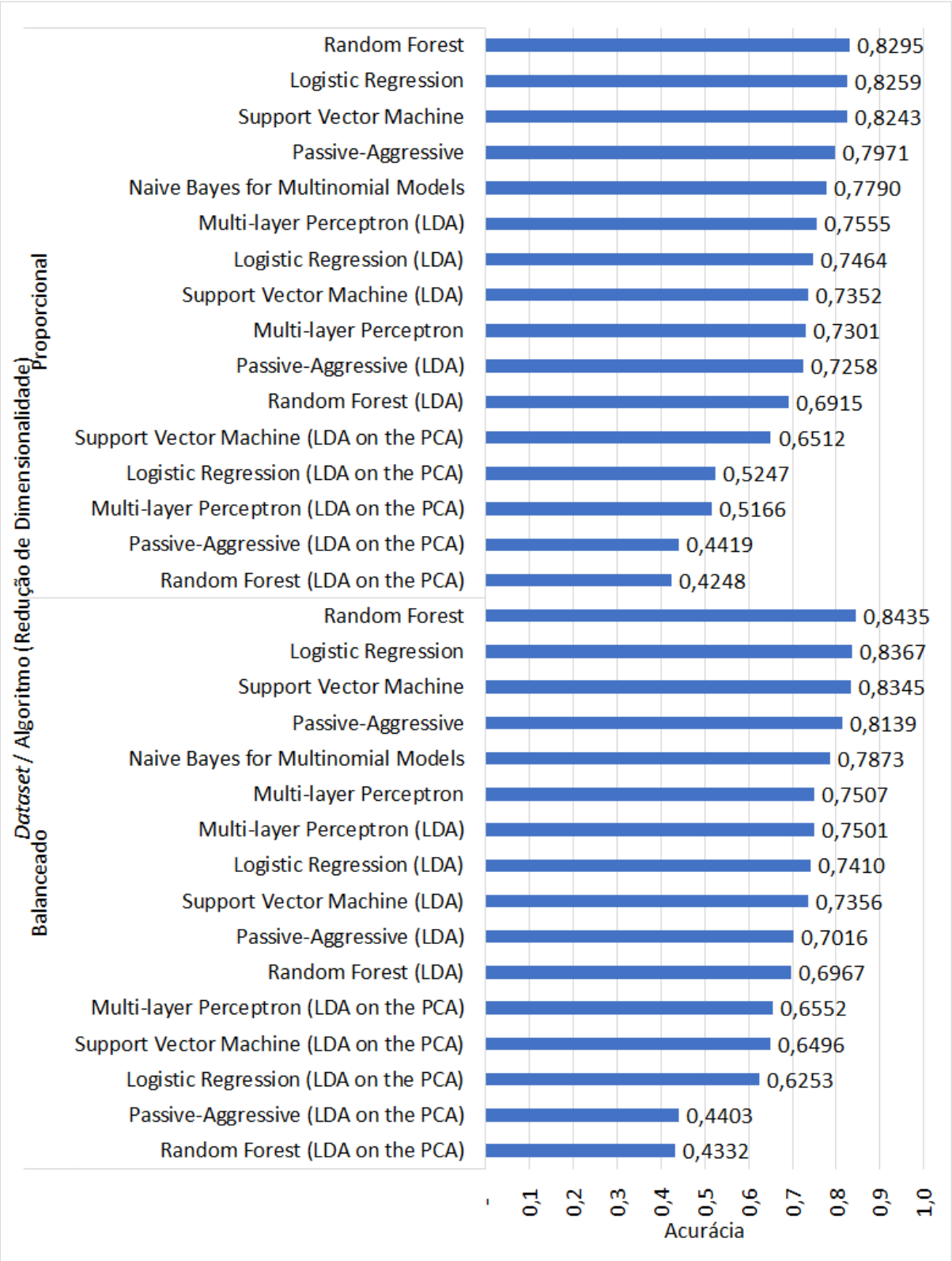


Figura D.13: Médias de acurácia da validação cruzada da predição da métrica *confidentiality* do vetor CVSS a partir da descrição textual da vulnerabilidade

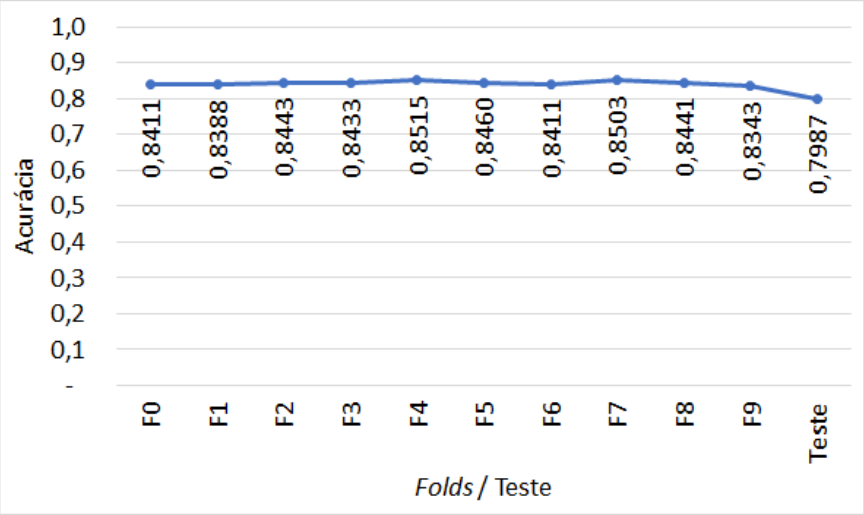


Figura D.14: Desempenho do teste e de cada *fold* do classificador que apresentou a melhor média dos *folds* para predição da métrica *confidentiality* do vetor CVSS a partir da descrição textual da vulnerabilidade (algoritmo *Random Forest*, *dataset* balanceado)

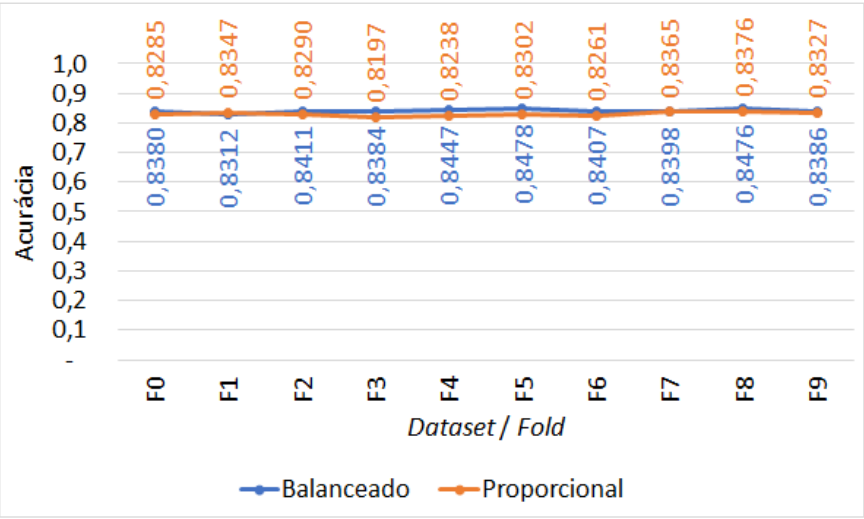
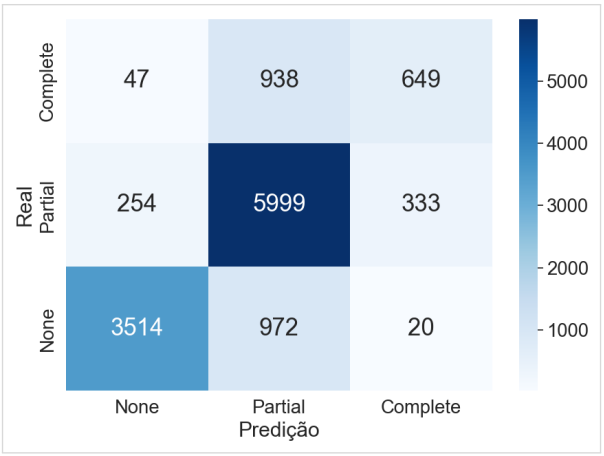
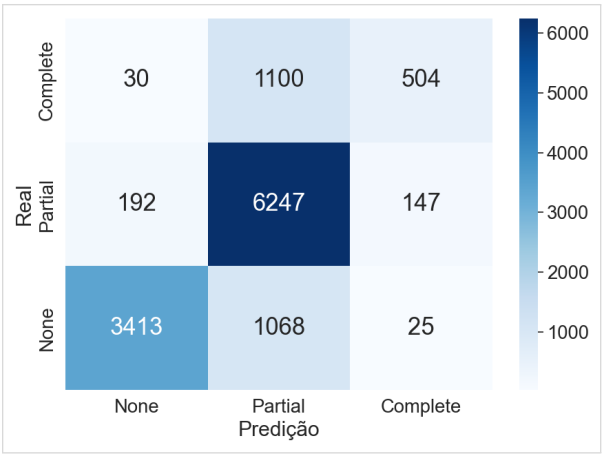


Figura D.15: Desempenho dos *folds* da validação cruzada para predição da métrica *confidentiality* do vetor CVSS a partir da descrição textual da vulnerabilidade utilizando o algoritmo *voting*



(a) Modelo gerado a partir do *dataset* balanceado



(b) Modelo gerado a partir do *dataset* proporcional

Figura D.16: Matriz de confusão do teste de classificação da métrica *confidentiality* do vetor CVSS a partir da descrição textual da vulnerabilidade

D.2.5 Predição da Métrica *Availability* do Vetor CVSS Diretamente a Partir da Descrição Textual da Vulnerabilidade

O desempenho da predição da métrica *availability* do vetor CVSS diretamente a partir da descrição textual da vulnerabilidade é ilustrado na Figura D.17. A melhor média da acurácia das validações cruzadas foi de 83%. Dentre os cinco melhores resultados, avaliando a média de acurácia dos *folds* das validações cruzadas, observa-se que os três melhores resultados foram treinados a partir do *dataset* balanceado e os dois seguintes treinados a partir do *dataset* proporcional. O algoritmo *Random Forest* ocupa a primeira e quarta posição, o algoritmo *Logistic Regression* a segunda e quinta posição e o algoritmo *Support Vector Machine* aparece na terceira posição.

O gráfico plotado na Figura D.18 tem o objetivo de estimar a estabilidade e a capacidade de generalização do modelo. Nesse gráfico foi avaliada a amplitude no desempenho dos dez *folds*, bem como o desempenho da classificação do *dataset* de teste. O classificador que obteve o melhor resultado nos testes foi aquele treinado com o algoritmo *Random Forest* e o subconjunto, *fold*, F4 do *dataset* balanceado. Dentre os dez *folds* da validação cruzada, o menor desempenho aferido foi de 82% e o maior de 84%. Dessa forma, a amplitude foi de, aproximadamente, 2 pontos percentuais. Já o desempenho verificado do teste foi de 77%, o que representa uma diferença de 7 pontos percentuais para o *fold* com melhor desempenho na validação, utilizado para classificar o *dataset* de teste.

Ao considerar as dez melhores médias dos resultados das predições da métrica *availability* do vetor CVSS, diretamente a partir da descrição textual das vulnerabilidades,

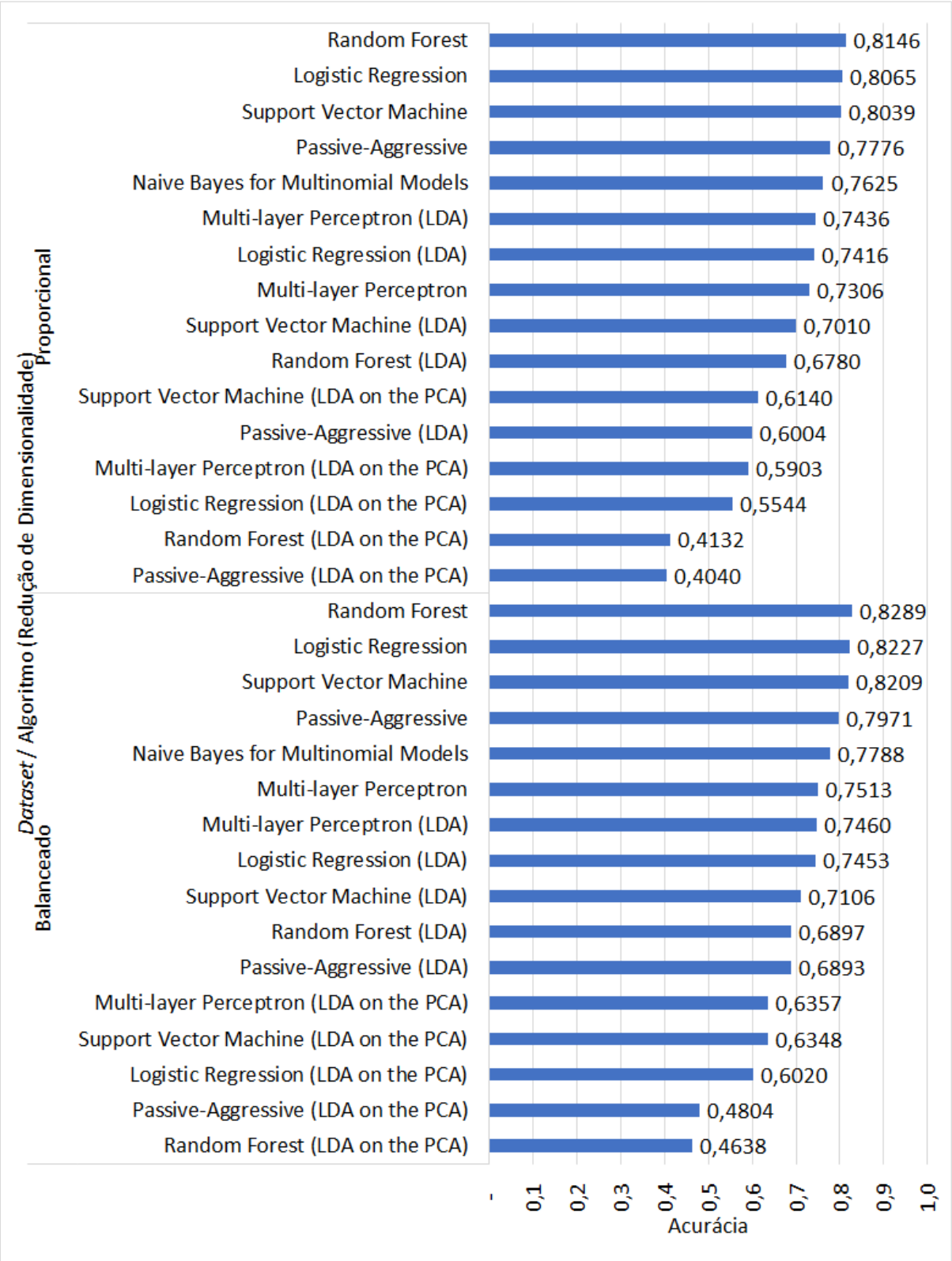


Figura D.17: Médias de acurácia da validação cruzada da predição da métrica *availability* do vetor CVSS a partir da descrição textual da vulnerabilidade

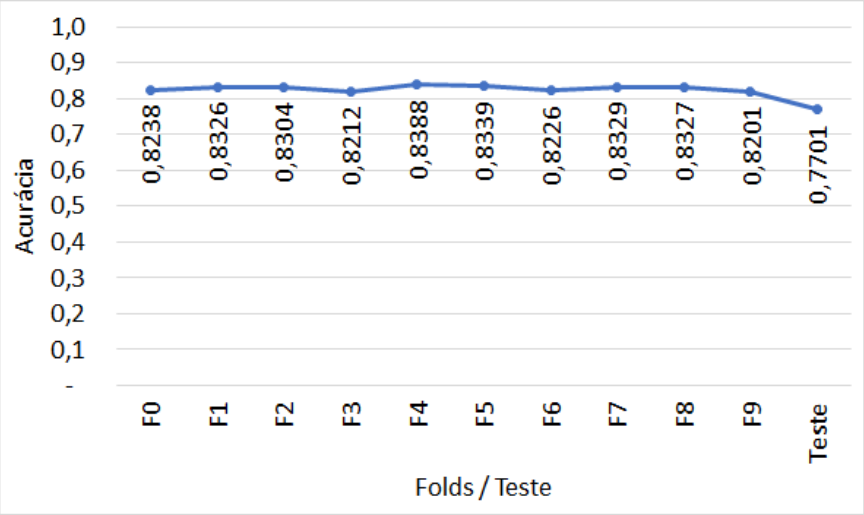


Figura D.18: Desempenho do teste e de cada *fold* do classificador que apresentou a melhor média dos *folds* para predição da métrica *availability* do vetor CVSS a partir da descrição textual da vulnerabilidade (algoritmo *Random Forest*, *dataset* balanceado)

constata-se que as proporções em relação ao emprego de *datasets* são iguais, ou seja: 50% proporcional e 50% balanceado. Nenhum dos dez melhores resultados usa algoritmo de redução de dimensionalidade. No entanto, Entre os dez piores resultados, todos usam um dos dois algoritmos de redução de dimensionalidade abordados no presente estudo.

Consoante Figura D.19, o algoritmo *voting* obteve a acurácia (média dos dez *folds*), de 83% quando treinado a partir do *dataset* balanceado, e de 81%, quando treinado a partir do *dataset* proporcional. Sendo assim, o algoritmo *voting* não superou os resultados obtidos pelos modelos construídos a partir dos algoritmos de classificação de forma individual, embora a diferença seja inferior à 1 ponto percentual. A utilização do *dataset* balanceado implicou, em todos os casos, com o algoritmo *voting*, em uma acurácia superior àquela obtida quando utilizado o *dataset* proporcional.

Estudando as matrizes de confusão (Fig. D.20) constata-se que tanto o modelo treinado a partir do *dataset* proporcional quanto o modelo treinado a partir do *dataset* balanceado tiveram mais dificuldade na predição da classe *complete*. As predições incorretas dessa classe foram preditas, na maior parte dos casos, como *partial*.

D.2.6 Predição da Métrica *Integrity* do Vetor CVSS Diretamente a Partir da Descrição Textual da Vulnerabilidade

O desempenho da predição da métrica *integrity* do vetor CVSS diretamente a partir da descrição textual da vulnerabilidade é ilustrado na Figura D.21. A melhor média

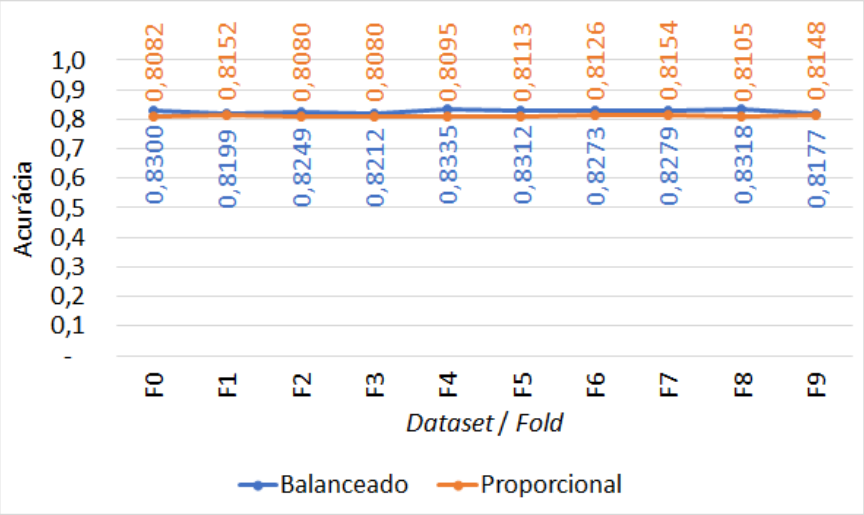
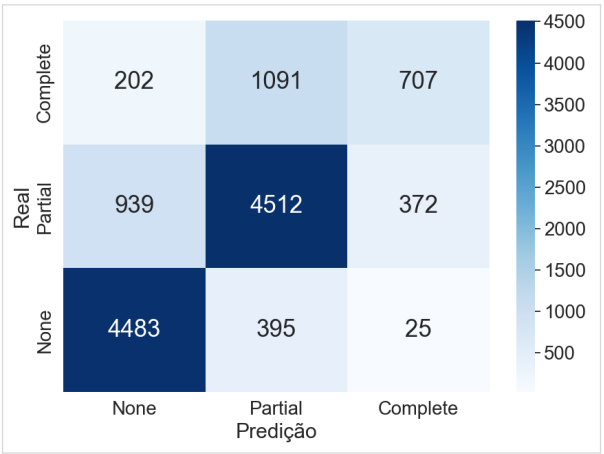
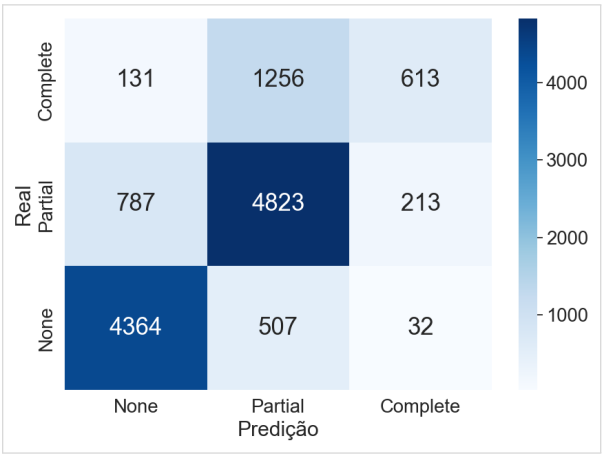


Figura D.19: Desempenho dos *folds* da validação cruzada para predição da métrica *availability* do vetor CVSS a partir da descrição textual da vulnerabilidade utilizando o algoritmo *voting*



(a) Modelo gerado a partir do *dataset* balanceado



(b) Modelo gerado a partir do *dataset* proporcional

Figura D.20: Matriz de confusão do teste de classificação da métrica *availability* do vetor CVSS a partir da descrição textual da vulnerabilidade

da acurácia das validações cruzadas foi de 85%. Dentre os cinco melhores resultados, avaliando a média de acurácia dos *folds* das validações cruzadas, observa-se que os três melhores resultados foram treinados a partir do *dataset* balanceado e os dois seguintes treinados a partir do *dataset* proporcional. O algoritmo *Random Forest* ocupa a primeira e quarta posição, o algoritmo *Logistic Regression* a segunda e quinta posição e o algoritmo *Support Vector Machine* aparece na terceira posição.

O gráfico plotado na Figura D.22 tem o objetivo de estimar a estabilidade e a capacidade de generalização do modelo. Nesse gráfico foi avaliada a amplitude no desempenho dos dez *folds*, bem como o desempenho da classificação do *dataset* de teste. O classifica-

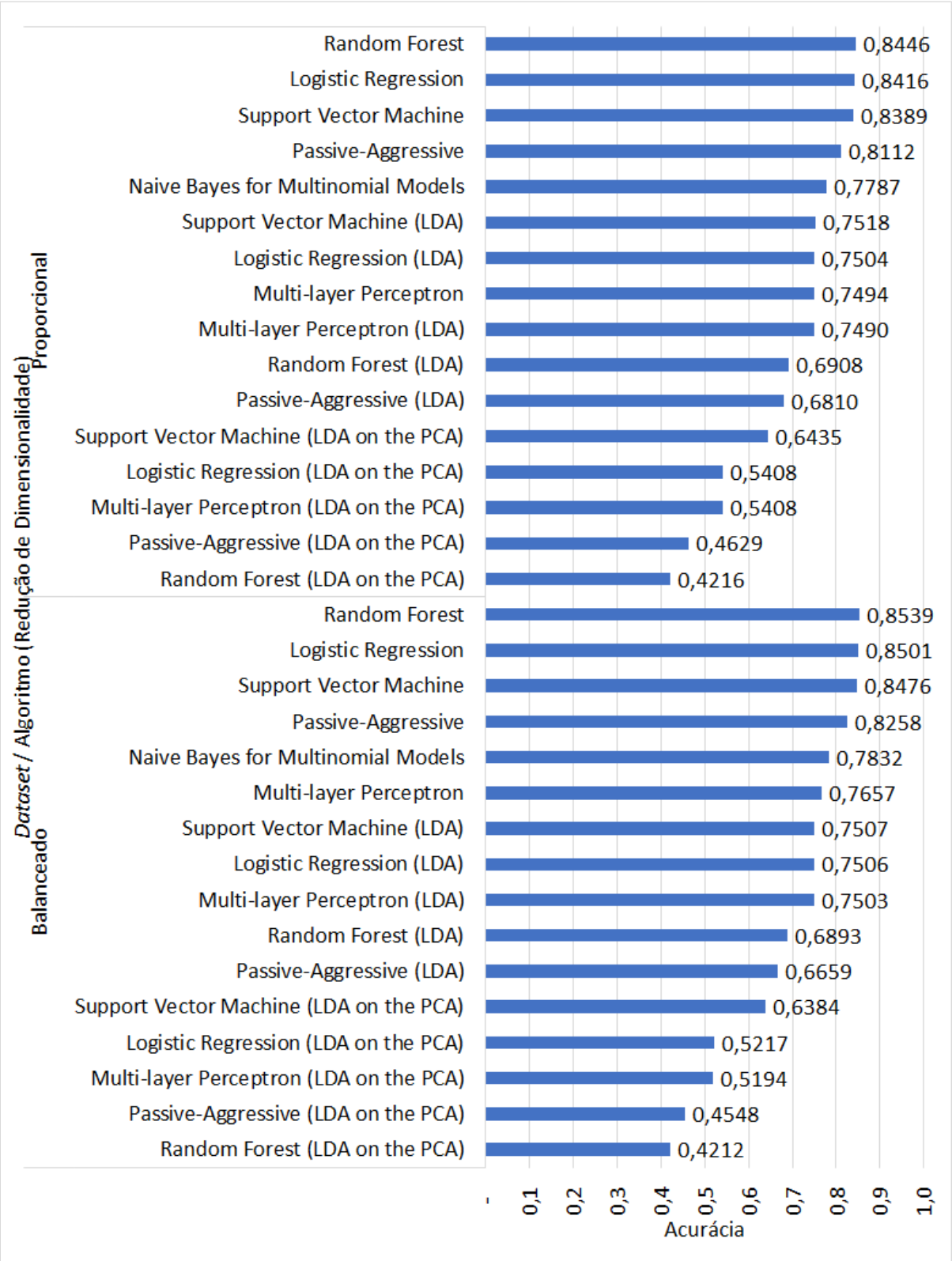


Figura D.21: Médias de acurácia da validação cruzada da predição da métrica *integrity* do vetor CVSS a partir da descrição textual da vulnerabilidade

dor que obteve o melhor resultado nos testes foi aquele treinado com o algoritmo *Random Forest* e o subconjunto, *fold*, F8 do *dataset* proporcional. Dentre os dez *folds* da validação cruzada, o menor desempenho aferido foi de 84% e o maior de 86%. Dessa forma, a amplitude foi de, aproximadamente, 2 pontos percentuais. Já o desempenho verificado do teste foi de 80%, o que representa uma diferença de 6 pontos percentuais para o *fold* com melhor desempenho na validação, utilizado para classificar o *dataset* de teste.

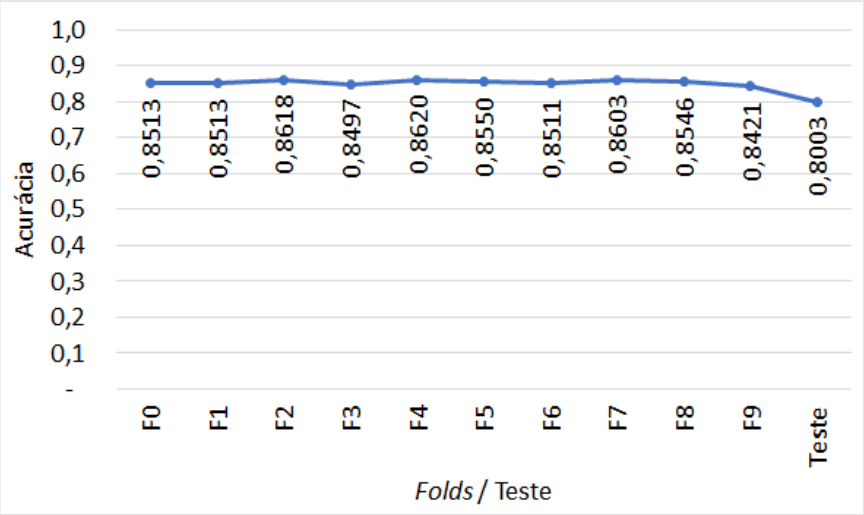


Figura D.22: Desempenho do teste e de cada *fold* do classificador que apresentou a melhor média dos *folds* para predição da métrica *integrity* do vetor CVSS a partir da descrição textual da vulnerabilidade (algoritmo *Random Forest*, *dataset* balanceado)

Ao considerar as dez melhores médias dos resultados das predições da métrica *integrity* do vetor CVSS, diretamente a partir da descrição textual das vulnerabilidades, constata-se que as proporções em relação ao emprego de *datasets* são iguais, ou seja: 50% proporcional e 50% balanceado. Nenhum dos dez melhores resultados usa algoritmo de redução de dimensionalidade. No entanto, entre os dez piores resultados todos usam o algoritmo de redução de dimensionalidade LDA on the PCA.

Conforme Figura D.23, o algoritmo *voting* foi registrou a acurácia (média dos dez *folds*), de 85,3% quando treinado a partir do *dataset* balanceado, e de 84,5%, quando treinado a partir do *dataset* proporcional. O algoritmo *voting* superou os resultados obtidos pelos modelos construídos a partir dos algoritmos de classificação de forma individual em 0,07 ponto percentual. A utilização do *dataset* balanceado implicou, em 80% dos *folds*, com o algoritmo *voting*, em uma acurácia superior àquela obtida quando utilizado o *dataset* proporcional.

Analisando as matrizes de confusão (Fig. D.24) observa-se que tanto o modelo treinado a partir do *dataset* proporcional quanto o modelo treinado a partir do *dataset* balanceado

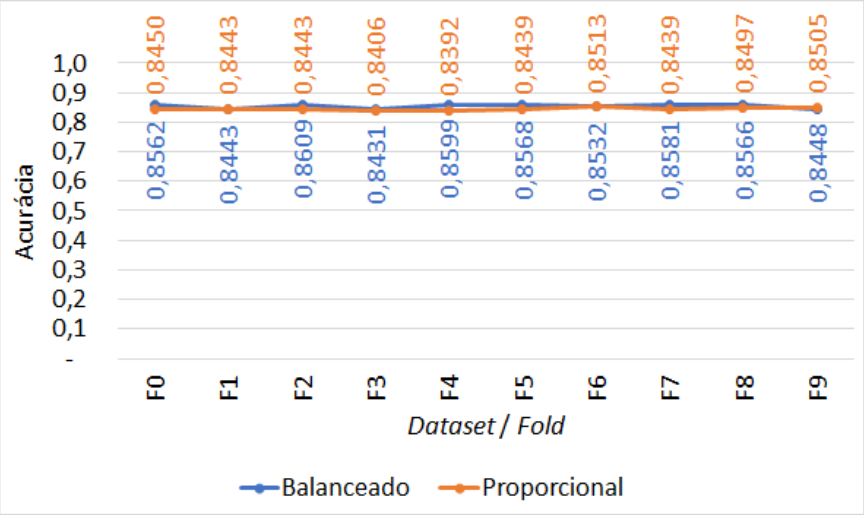
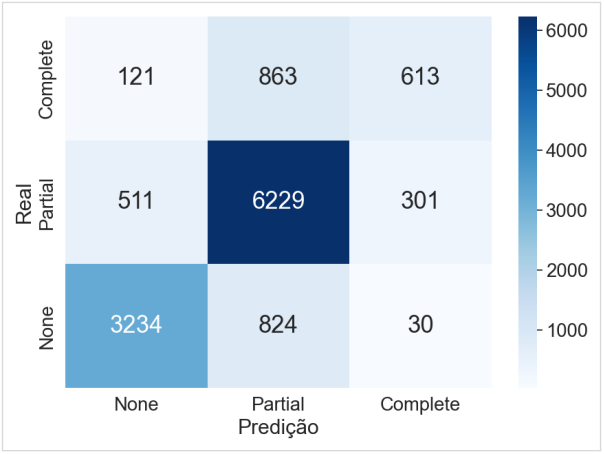
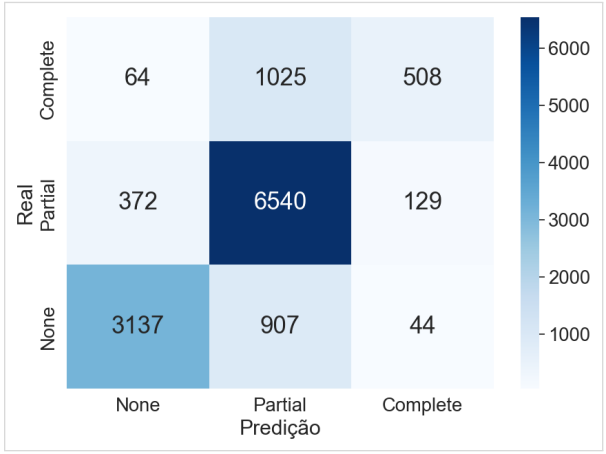


Figura D.23: Desempenho dos *folds* da validação cruzada para predição da métrica *integrity* do vetor CVSS a partir da descrição textual da vulnerabilidade utilizando o algoritmo *voting*

tiveram mais dificuldade na predição da classe *complete*. As predições incorretas dessa classe foram preditas, na maior parte dos casos, como *partial*.



(a) Modelo gerado a partir do *dataset* balanceado



(b) Modelo gerado a partir do *dataset* proporcional

Figura D.24: Matriz de confusão do teste de classificação da métrica *integrity* do vetor CVSS a partir da descrição textual da vulnerabilidade

D.2.7 Cálculo do *Base Score* a Partir dos Valores Preditos das Métricas do Vetor CVSS

Após realizar a predição de todas as seis métricas do vetor CVSS, versão 2, é possível calcular o *base score* aplicando a fórmula descrita na Subseção D.1. É importante salientar que o cálculo do *base score* a partir dos valores preditos das métricas do vetor CVSS retorna

o valor arredondado a uma casa decimal. Dessa forma, é possível obter exatamente o valor real do *base score*, diferente da abordagem do Apêndice C, onde o *base score* é truncado.

O desempenho do cálculo do *base score* a partir dos valores preditos das métricas do vetor CVSS é ilustrado na Figura D.25. A melhor média da acurácia das validações cruzadas foi de 57%, ou seja, 57% das predições foi exatamente igual ao *base score*, incluindo o decimal. Dentre os cinco melhores resultados, avaliando a média de acurácia dos *folds* das validações cruzadas, o algoritmo *Random Forest* aparece na duas primeiras posições, em primeiro lugar com treinamento realizado a partir do *dataset* balanceado e em segundo lugar com o treinamento realizado a partir do *dataset* proporcional. A terceira e a quinta posição foram ocupadas pelo algoritmo *Logistic Regression* treinados a partir dos *datasets* balanceado e proporcional, respectivamente. O algoritmo *Support Vector Machine*, treinado a partir do *dataset* balanceado, ficou na quarta posição. Em nenhum dos melhores cinco resultados foi aplicado algum dos algoritmos de redução de dimensionalidade.

Tendo em vista que o valor do *base score* obtido através do cálculo a partir dos valores preditos das métricas do vetor CVSS é arredondado à primeira casa decimal, conforme explicitado na Sub-subseção D.1.1, foi utilizada a métrica MSE para avaliar se as predições incorretas estão muito distantes do valor real (Fig. D.26). Entre os dez melhores MSE, o máximo medido foi 2,66. Isso significa que as predições incorretas não excessivamente discrepantes do valor real.

Também é possível constatar que, embora o modelo treinado a partir do *dataset* balanceado tenha apresentado melhor acurácia, o modelo treinado a partir do *dataset* proporcional apresentou um MSE mais baixo. Ao aplicar a métrica RMSE ao melhor resultado de MSE, modelo treinado com algoritmo *Random Forest* e *dataset* proporcional, o resultado é 1,4. Assim, em média, o erro na predição do *base score* é de 1,4 para mais ou para menos.

O gráfico plotado na Figura D.27 tem o objetivo de estimar a estabilidade e a capacidade de generalização do modelo. Nesse gráfico foi avaliada a amplitude no desempenho dos dez *folds*, bem como o desempenho da classificação do *dataset* de teste. O classificador que obteve o melhor resultado nos testes foi aquele treinado com o algoritmo *Random Forest* e o subconjunto, *fold*, F3 do *dataset* proporcional. Dentre os dez *folds* da validação cruzada, o menor desempenho aferido foi de 55% e o maior de 59%. Dessa forma, a amplitude foi de, aproximadamente, 4 pontos percentuais. Já o desempenho verificado do teste foi de 45%, o que representa uma diferença de 14 pontos percentuais para o *fold*

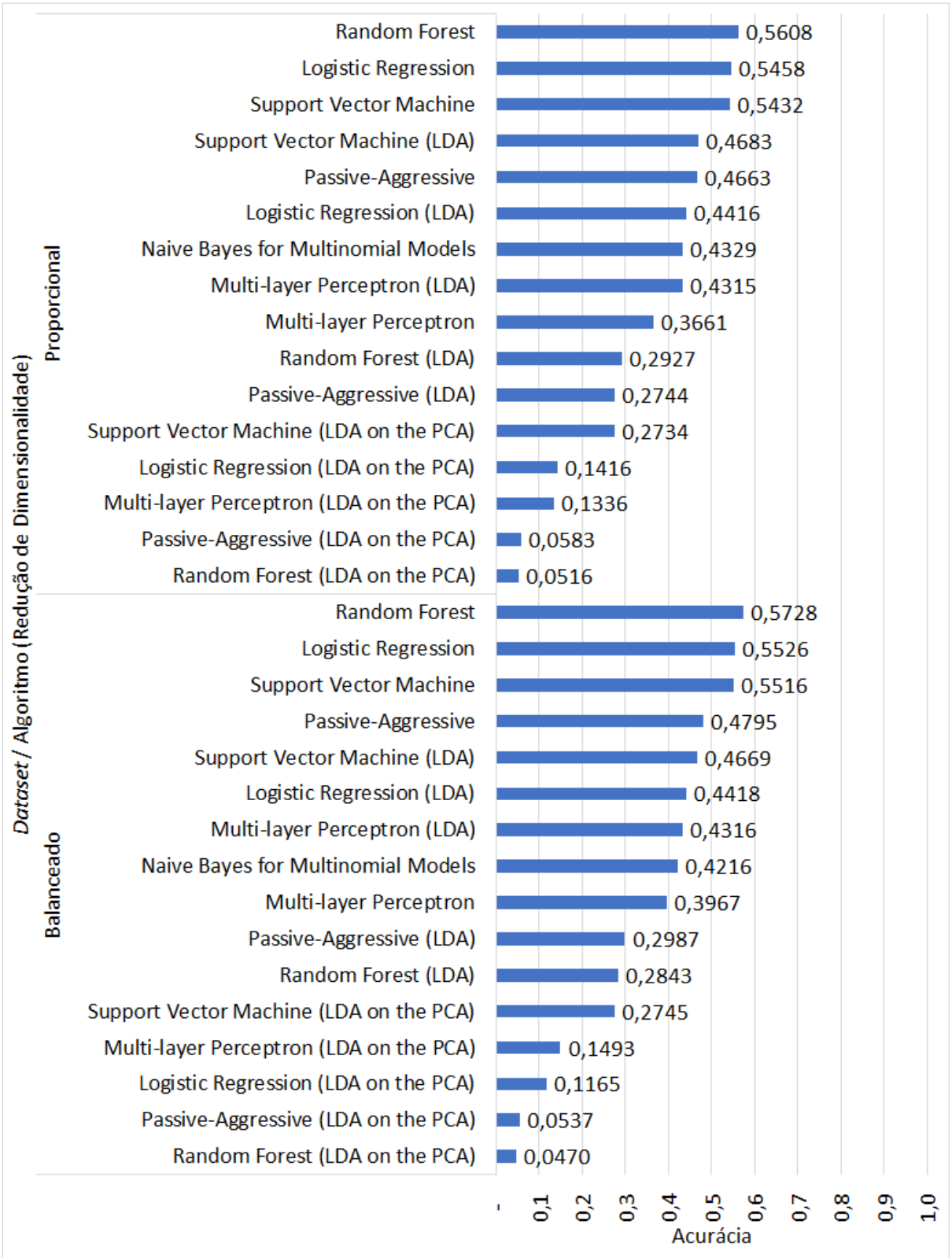


Figura D.25: Médias da acurácia na validação cruzada do *base score* calculado a partir dos valores preditos das métricas do vetor CVSS

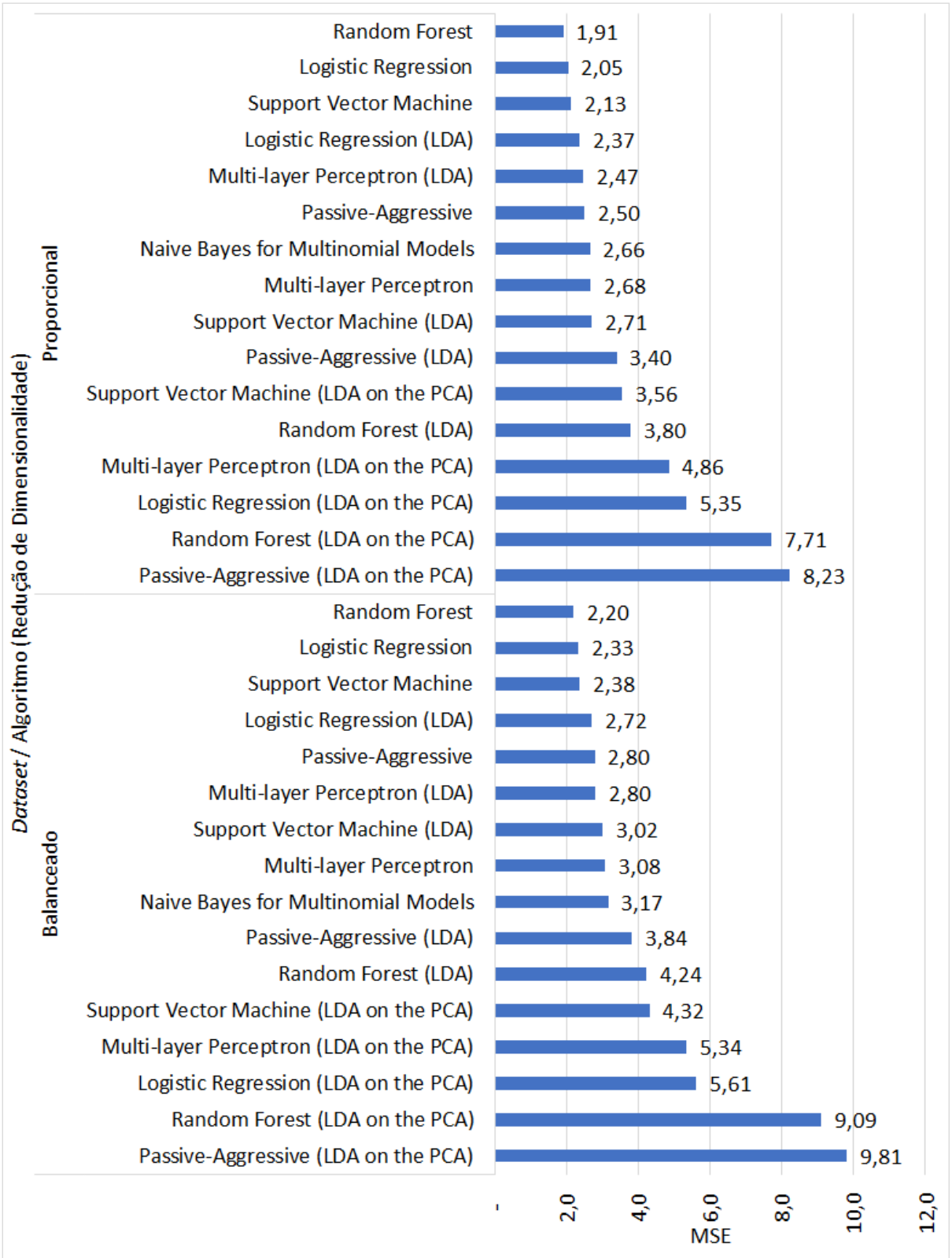


Figura D.26: Médias do MSE na validação cruzada do *base score* calculado a partir dos valores preditos das métricas do vetor CVSS

com melhor desempenho na validação, utilizado para classificar o *dataset* de teste.

A diferença na acurácia de cinco pontos percentuais entre a segunda abordagem (Apêndice C), predição do *base score*, e a terceira abordagem, cálculo do *base score* a partir das métricas do vetor CVSS, pode, à primeira vista, levar ao entendimento de que a segunda abordagem é superior à terceira. No entanto, é importante ressaltar que na segunda abordagem o *base score* é truncado, o que diminui a margem de erro do modelo. O mesmo fato influencia o resultado do RMSE, que registrou, na segunda abordagem, o valor de 1,4, contra 2,66 da terceira abordagem. Na segunda abordagem a diferença entre o melhor e o pior desempenho é de dois pontos percentuais, contra os quatro da terceira abordagem. No entanto, o classificador que alcançou o melhor resultado em ambas abordagens foi treinado usando o mesmo algoritmo (*Random Forest*) e o mesmo tipo de *dataset* (balanceado). No teste, a segunda abordagem apresentou um desempenho superior em seis pontos percentuais, se comparado ao da terceira abordagem. O truncamento do *base score* faz diferença na medição do desempenho do classificador. Contudo, não necessariamente implica no desempenho da obtenção da severidade a partir de seu valor.

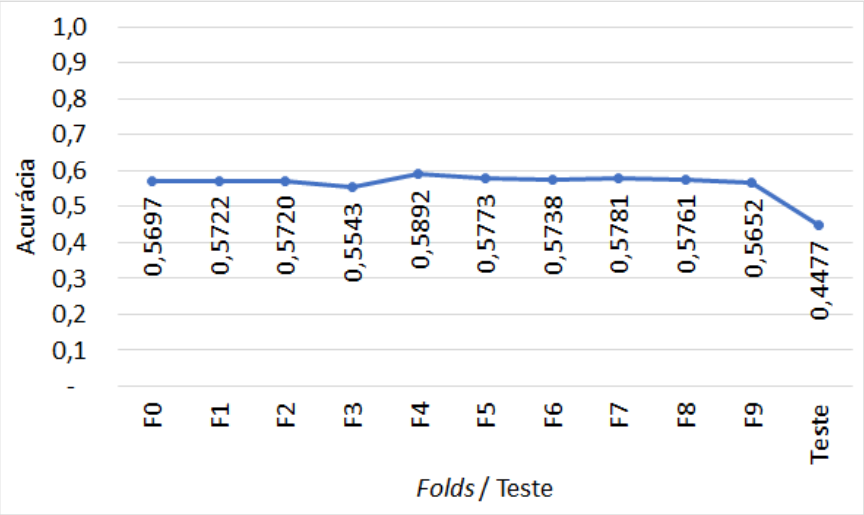


Figura D.27: Desempenho do teste e de cada *fold* do classificador que apresentou a melhor média dos *folds* para o *base score* calculado a partir dos valores preditos das métricas do vetor CVSS (algoritmo *Random Forest*, *dataset* balanceado)

Ao considerar as dez melhores médias dos resultados do cálculo do *base score* a partir dos valores preditos das métricas do vetor CVSS, constata-se que a composição do *dataset* teve pouca influência no resultado final. Isso porque cinco em dez desses melhores resultados utilizaram o *dataset* balanceado para treinamento do modelo. Nos demais 50%, os modelos foram treinados a partir do *dataset* proporcional. Além disso, os dois melhores resultados foram alcançados em modelos treinados pelo algoritmo *Random Forest*, cada

um com um tipo de *dataset*, e a diferença na acurácia desses dois modelos foi de 1 ponto percentual. Dos dez melhores resultados, considerando a acurácia, somente 20% utilizou algoritmos de redução de dimensionalidade. Entre os dez piores resultados, todos os modelos usaram um dos dois algoritmos de redução de dimensionalidade abordados no presente estudo.

Consoante Figura D.28, o algoritmo *voting* registrou a acurácia (média dos dez *folds*), de 56% quando treinado a partir do *dataset* balanceado, e de 55%, quando treinado a partir do *dataset* proporcional. Sendo assim, o algoritmo *voting* não superou os resultados obtidos pelos modelos construídos a partir dos algoritmos de classificação de forma individual. A utilização do *dataset* balanceado implicou, em 80% dos casos, com o algoritmo *voting*, em uma acurácia superior àquela obtida quando utilizado o *dataset* proporcional.

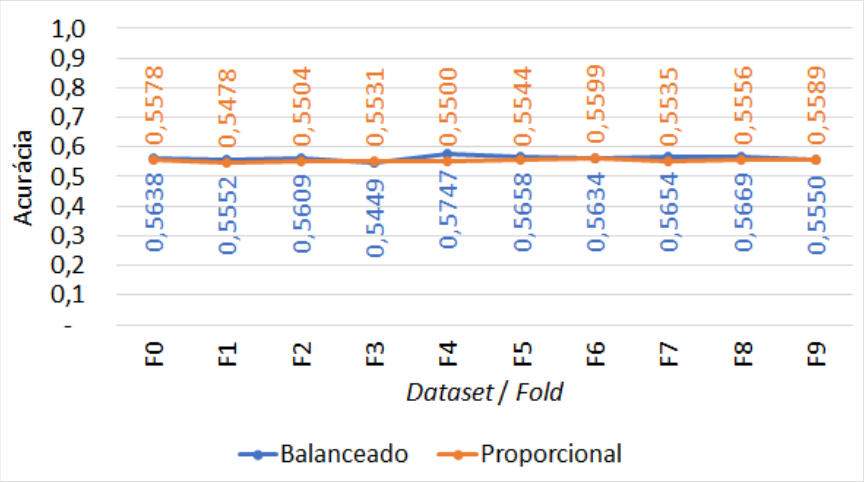
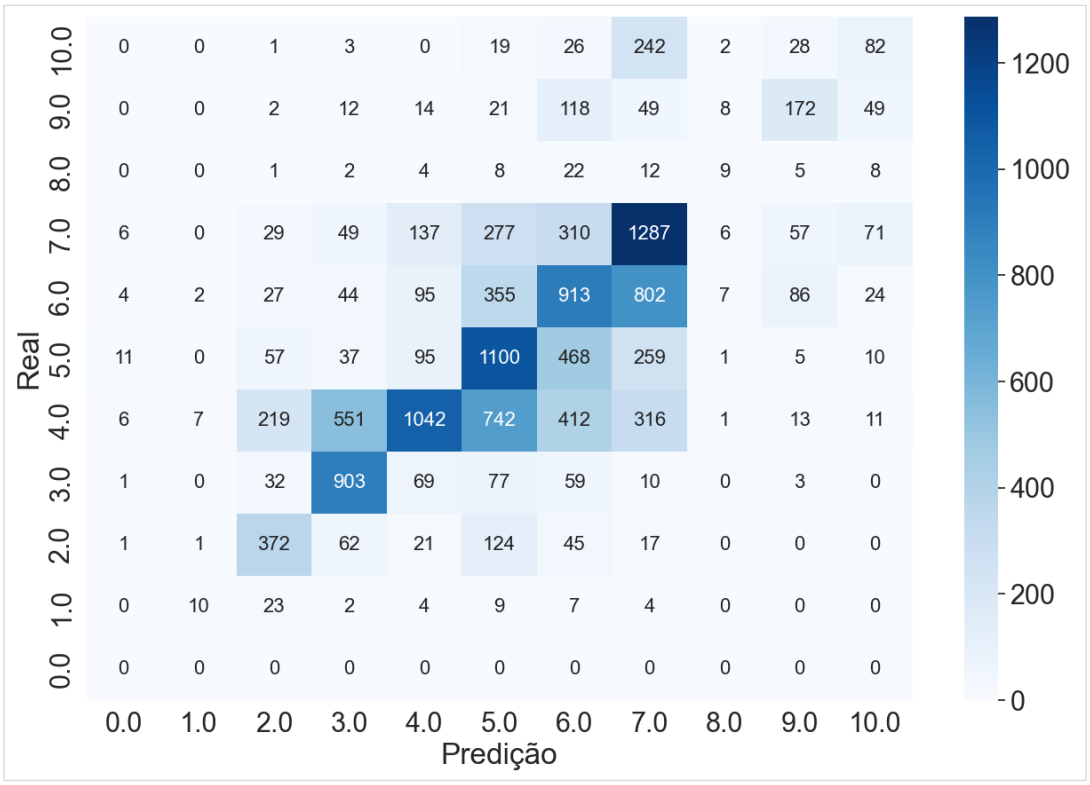
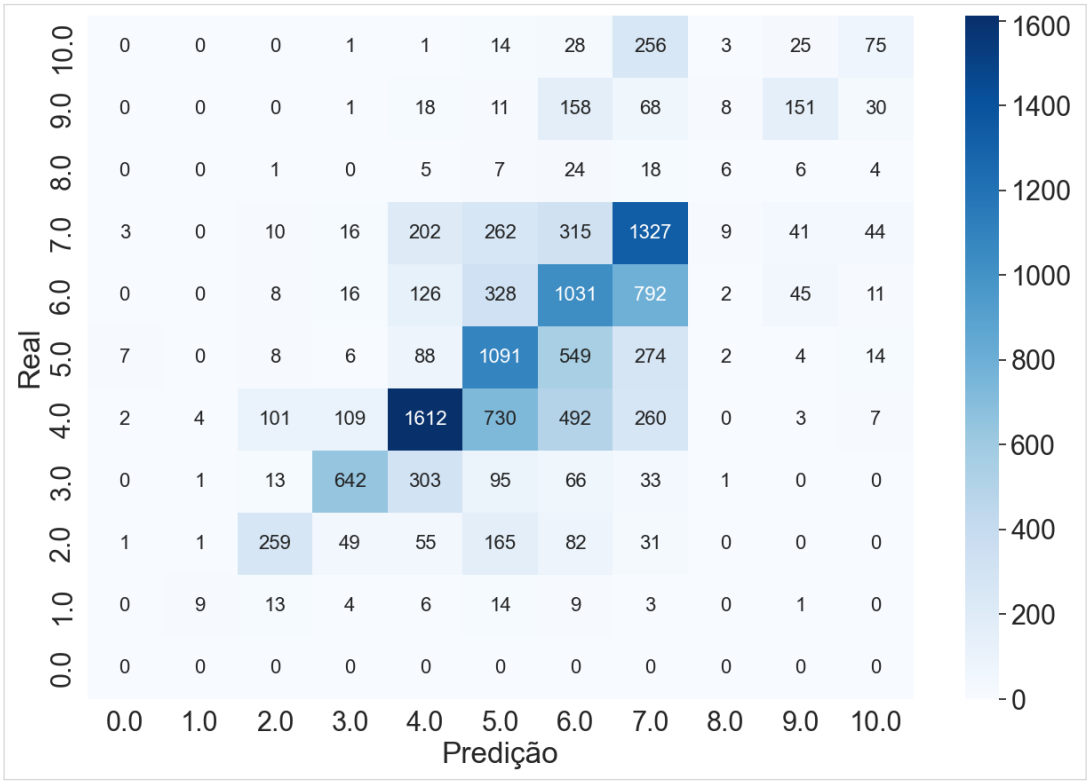


Figura D.28: Desempenho dos *folds* da validação cruzada para o *base score* calculado a partir dos valores preditos das métricas do vetor CVSS

Analisando as matrizes de confusão do resultado do teste (Fig. D.29), observa-se, no teste realizado a partir do modelo treinado com o *dataset* proporcional, uma concentração mais próxima à diagonal, o que justifica um MSE de 2,22 e um RMSE de 1,5. No teste realizado a partir do modelo treinado com o *dataset* balanceado a concentração também não fica distante da diagonal, corroborando o MSE de 2,38 e um RMSE também de 1,5. Cabe destacar que, embora toda a análise do cálculo do *base score* a partir dos valores preditos das métricas do vetor CVSS seja realizado com valores arrendondados à primeira casa deciaml, foi necessário truncá-los especificamente na matriz de confusão, a fim de viabilizar a visualização das informações.



(a) Modelo gerado a partir do *dataset* balanceado



(b) Modelo gerado a partir do *dataset* proporcional

Figura D.29: Matriz de confusão do teste de classificação do *base score* calculado a partir dos valores preditos das métricas do vetor CVSS