Research, Society and Development, v. 12, n. 4, e22512441277, 2023
(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v12i4.41277

Análise de Métricas de Desempenho sobre a Conjuntura de Intrusões em Redes IEEE 802.11 com Aprendizagem de Máquina no Hospital N.S.C.

Analysis of Performance Metrics on the Environment of Intrusions in IEEE 802.11 Networks with Machine Learning at Hospital N.S.C.

Análisis de Métricas de Rendimiento en Ambiente de Intrusiones en Redes IEEE 802.11 con Machine Learning en Hospital N.S.C.

Recebido: 30/03/2023 | Revisado: 10/04/2023 | Aceitado: 11/04/2023 | Publicado: 15/04/2023

Matheus Santos Andrade
ORCID: https://orcid.org/0000-0001-7274-9633
Instituto Federal de Educação, Ciência e Tecnologia de Sergipe, Brasil
E-mail: matheusbsi1992@gmail.com
Jonathas Carvalho de Freitas
ORCID: https://orcid.org/0009-0008-3268-6090
Universidade Tiradentes, Brasil
E-mail: jonathas200@gmail.com
Aldo César dos Santos Dultra
ORCID: https://orcid.org/0009-0009-9882-6962
Universidade Federal de Sergipe, Brasil
E-mail: aldocesar29@hotmail.com
Ubiratan Silva de Souza Junior
ORCID: https://orcid.org/0009-0006-0552-8295
Centro Universitário SENAC, Brasil
E-mail: ubiratan.souza@protonmail.com

Resumo

A segurança presente em redes IEEE 802.11 faz-se diariamente mais relevante. Porém, a segurança na rede IEEE 802.11 não tem acompanhado as ameaças com tanta significância. Por este motivo surge a proposta deprojetar um Sistema de Detecção de Intrusão-IDS baseada em aprendizagem de máquina que será capaz de possuir auto-aperfeiçoamento, visto que, irá criar um ambiente seguro, capaz de detectar todas as ameaças dissimuladas, *Deauthentication,* EAPOL-*Logoff* e *Beacon Flood,* em que foram lançadas em uma rede corporativa real. Com isto, correlacionado as métricas de desempenho, e entre uma delas, que preza pela qualidade da classificação, o *Matthews Correlation Coefficient.* A anomalia *Deauthentication* acima do classificador *Naive Bayes* foi obtido de (88,71%), já a valia de qualidade do classificador *Logistic Regression*(*Logistic*) equacionado a (88,69%), e não obstante, o *J48* apresentou um valor menor de (88,47%). Apesar disso, a identificação do ataque *Beacon Flood*, se deu por conta do algoritmo *Naive Bayes* exibindo a maior taxa de detecção (100,00%), seguido do *Logistic* (99,95%) e *J48* possuindo o menor valor (98,85%). Conseguinte, na detecção da anomalia EAPOL-*Logoff*, os classificadoresapresentaram similitude de (100,00%) e a demais, com a apresentação de uma detecção, em virtude de dados nãoanômalos (Normal), o *Naive Bayes* foi acometido de (89,92%), seguido do *Logistic* mantendo (89,89%), enquanto, o *J48* foi testado com uma taxa menor de (89,67%). Com as evidências do estudo proveem a possibilidade de que épossível desenvolver um sistema de detecção de intrusão baseado em redes *wireless*.

Palavras-chave: Ameaças; Qualidade; Evidências.

Abstract

The security present in IEEE 802.11 networks becomes more relevant every day. However, security on the IEEE 802.11 network has not kept pace with threats with as much significance. For this reason, the proposal arises to design an Intrusion Detection System-IDS based on machine learning that will be able to have self-improvement, since it will create a safe environment, capable of detecting all disguised threats, Deauthentication, EAPOL-Logoff and Beacon Flood, where they were launched on a real corporate network. With this, correlated the performance metrics, and among them, which values the quality of the classification, the Matthews Correlation Coefficient. The Deauthentication anomaly above the Naive Bayes classifier was obtained (88,71%), whereas the quality value of the Logistic Regression (Logistic) classifier was equated to (88,69%), and nevertheless, the J48 presented a lower value of (88,47%). Despite this, the identification of the Beacon Flood attack was due to the Naive Bayes algorithm showing the highest detection rate (100,00%), followed by Logistic (99,95%) and J48 having the lowest value

(98,85 %). As a result, in the detection of the EAPOL-Logoff anomaly, the classifications presented similarity of (100,00%) and the others, with the presentation of a detection, due to non-anomalous data (Normal), the Naive Bayes was affected by (89,92%), followed by Logistic maintaining (89,89%), while J48 was tested with a lower rate (89,67%). With the study evidences provide the possibility that it is possible to develop an intrusion detection system based on wireless networks.
Keywords: Threats; Quality; Evidences.

Resumen
La seguridad presente en las redes IEEE 802.11 cobra cada día más relevancia. Sin embargo, la seguridad en la red IEEE 802.11 no ha seguido el ritmo de las amenazas de tanta importancia. Por tal motivo surge la propuesta de diseñar un Sistema de Detección de Intrusos-IDS basado en *machine learning* que podrá tener autoperfeccionamiento, ya que creará un entorno seguro, capaz de detectar todas las amenazas encubiertas, *Deauthentication*, EAPOL-*Logoff* y *Beacon Flood*, donde se lanzaron en una red corporativa real. Con esto, correlacionó las métricas de desempeño, y entre ellas, la que valora la calidad de la clasificación, el Coeficiente de Correlación de *Matthews*. La anomalía *Deauthentication* arriba del clasificador *Naive Bayes* se obtuvo (88,71%), mientras que el valor de calidad del clasificador *Logistic Regression* (*Logistic*) se igualó a (88,69%), y sin embargo, el *J48* presentó un valor menor de (88,47%). A pesar de esto, la identificación del ataque *Beacon Flood* se debió a que el algoritmo *Naive Bayes* mostró la tasa de detección más alta (100,00%), seguido de *Logistic* (99,95%) y *J48* con el valor más bajo (98,85%). Como resultado, en la detección de la anomalía EAPOL-*Logoff*, las clasificaciones presentaron similitud de (100,00%) y las demás, con la presentación de una detección, por datos no anómalos (Normal), el *Naive Bayes* se vio afectado por (89,92%), seguido de *Logistic* de mantenimiento (89,89%), mientras que *J48* se probó con una tasa más baja (89,67%). Con el estudio se evidencia la posibilidad de que sea posible desarrollar un sistema de detección de intrusos basado en redes *wireless*.
Palabras clave: Amenazas; Calidad; Evidencias.

1. Introdução

As redes de computadores surgiram com a necessidade de interligar universidades ou centros acadêmicos. E posteriormente abraçadas pelas empresas de modo que trouxeram benefícios para indústria, comércio e domicílios (Arasaki & Della Flora, 2012). Em contrapartida, dispositivos com grande teor de mobilidade, cujas especificações seguem os métodos da família IEEE 802.11 (IEEE 802.11, 1999), como: *laptops*, celulares, *tablets* e dentre outros tornaram-se comuns e com o público diversificado sendo mais utilizados atualmente (Feng, 2012). No entanto, é corriqueira a presença de *hackers* nos meios eletrônicos, a vista que o número de usuários de *Internet* e a facilidade de aquisição de serviços nas redes de comunicação sem fio e produtos, faz-se da quantidade de processos financeiros despertarem interesse dos atacantes, para aplicarem golpes envolvendo ganhos monetários, espionagem industrial, extorsão, venda de informações e difamação da imagem do governo.

Embora os protocolos que visam a segurança das redes sem fio: WEP (*Wired Equivalent Privacy*) (IEEE 802.11, 1999) pretendia garantir um nível de segurança parecido com a rede cabeada (Morimoto, 2008) para proteger os quadros de dados que transportam dados e informações de controle através de seu cabeçalho, mas devida a sua diversidade de vulnerabilidades (Tews, 2007) e a não garantia da escalabilidade ao modelo, todavia (i.e., ficou ultrapassado), surgindo um novo padrão de segurança chamado de WPA (*Wi-Fi Protected Access*) (Wi-Fi Alliance, 2003) e WPA2 (*Wi-Fi Protected Access Version* 2) (IEEE 802.11i, 2004) auxiliando na proteção junto a confidencialidade e integridade a comunicação dos dados na rede. Apesar disso, não apresenta segurança aos quadros de controle que reserva o canal de comunicação na confirmação de dados na rede, e aos quadros de gerenciamento no reconhecimento da presença de uma rede sem fio, para iniciar a associação e desassociação de estações a algum AP (*Access Point*) (Linhares & Gonçalves, 2012).

Entretanto, com o surgimento da emenda IEEE 802.11w (IEEE 802.11w, 2009), que inclui proteção aos quadros de gerenciamento, que só foi ratificada em 2009, após uma década do surgimento da emenda IEEE 802.11, o que permitiu uma gama de desenvolvimento de ataques, visando a interoperabilidade da rede, bem como a prática de captura de informações sensíveis sendo conduzidos nesses quadros.

Apesar da natureza em particular das redes sem fio, junto às emendas (IEEE 802.11i e IEEE 802.11w) resolvendo partes das vulnerabilidades encontradas neste percalço nas redes IEEE 802.11, os incidentes nas redes sem fio que são (*e.g.* causados pela realização de ataques de negação de serviço, perda de informações pela solicitação de transmissão falsa que estão sendo armazenadas no AP, que seriam enviadas as estações vinculadas e não estariam prontas para receber, causando a rejeição de informações, além disso o bloqueio do uso do canal de comunicação por um período de tempo estipulado, e dentre outros).

Com isto, há uma maneira de inibir tais "mazelas" e é através de um Sistema de Detecção de Intrusão (do Inglês *Intrusion Detection System* - IDS) que proporciona a coibição de eventos buscando identificar, diagnosticar e tratar anomalias para manter uma rede operando (Barford et al., 2002), porém com o cenário heterogêneo das redes sem fio vem a tornar complexa a sua justa avaliação, e sendo assim, o objetivo deste artigo é de apresentar uma abordagem na construção de um conjunto de dados que representa uma rede sem fio bem como a avaliação através de aprendizagem de máquina que surge como a necessidade de aprimorar o desempenho de alguma atividade através de experiência ou na descoberta entre similitudes de dados homogêneos (Mitchell, 1997), para aumentar a segurança em IDS.

2. Metodologia

Esta seção descreve um resumo sobre o que foi adotado para o desenvolvimento do conjunto de dados baseado no IEEE 802.11 e também a uma visão geral sobre a realização dos experimentos que foram adotados no desenvolvimento do trabalho. O procedimento principal inclui, a criação do conjunto de dados, pré-processamento, normalização e classificação.

2.1 Geração do Conjunto de Dados

O conjunto de dados gerado é de um cenário real contido no Hospital Nossa Senhora da Conceição, localizado na cidade de Lagarto-SE, região centro-sul do Estado e sobre as coordenadas geográficas: Latitude: -10.912929561173492, Longitude: -37.673240474073125. A via a isso, a coleta de dados aconteceu entre os dias 11/01/2023 até 16/01/2023 intercalados/variando de dias e entre o intervalo de $\pm 1hr$ para a geração do *dataset*. Incluído de uma rede *wireless* aos componentes como, HTTP/HTTPS, SMTP, POP3, IMAP e SSH. Representando a uma rede corporativa com diversos usuários autenticados, além da contenção de criptografia WPA2 habilitada para uma rede segura.

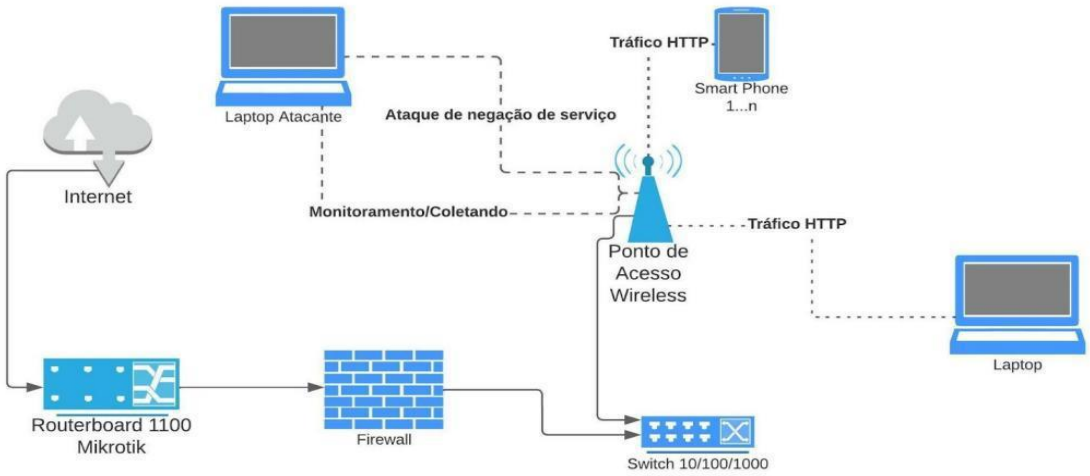
O cenário apontado na (Figura 1) se delimita ao ponto de existirem vários dispositivos enviando/recebendo dados à infraestrutura da rede, entretanto, é da responsabilidade da estação agressora (*Laptop* Atacante) através do Linux (Kali Linux) a captação e/ou monitoramento de dados de rádio na transmissão do quadro MAC correspondente ao IEEE 802.11, além da geração de ataques simultâneos e categóricos com as técnicas de negação de serviço:

1) *Deauthentication*: Este tipo de ataque afeta os quadros de gerenciamento, com o envio simultâneo de quadros irreais que vem a forçar o dispositivo conectado a ser desautenticado da rede (Ahmad & Tadakamadla, 2011). Detalhes sobre a utilização deste ataque é através da ferramenta *aireplay-ng* no pacote *Aircrack-ng* (Aircrack-ng, 2022);

2) EAPOL-*Logoff*: Este ataque prejudica os quadros de gerenciamento e controle, com uma inundação de pacotes EAPOL forjados e os envia para o AP deletar o estado de autenticação de um usuário autenticado e associado (Ahmad & Tadakamadla, 2011). Para este tipo de ataque a ferramenta *Mdk3* (Mdk3, 2022) foi utilizado;

3) *Beacon Flood*: Ataque que causa danos aos quadros de gerenciamento, e isto através da emissão de gama de pacotes com vários SSIDS falsos ao espectro da frequência da rede, assim trazendo desordem ao usuário que tentar se conectar ao AP (Ahmad & Tadakamadla, 2011). Para o uso deste tipo de ataque foi utilizado a ferramenta *Mdk3* (Mdk3, 2022).

Figura 1 - Topologia da rede *Wireless* (WPA2) contida de um segmento no Hospital Nossa Senhora da Conceição (H.N.S.C).



Fonte: Autores (2023).

A Figura 1, demonstra as várias formas de interconectar, em redes de computadores, sendo relacionados por componentes de uma rede de comunicação de dados. Entretanto, em via a trazer intercomunicação através de ondas de rádio frequência, para esta topologia, os quadros trocados entre os nós (*Smart Phone*, *Laptop* atacante ou não e o ponto de acesso *wireless*), são transmitidos através da atmosfera.

2.2 Pré-Processamento e Normalização

Na fase de pré-processamento que teve o auxílio da ferramenta *Wireshark* (Wireshark, 2022), sobre a necessidade da estação agressora mantendo apenas de atributos significantes ao quadro MAC (*Protocol Version*, *Type*, *To DS*, *From DS*, *More Fragment*, *Retry*, *Power Management*, *More Data*, *WEP*, *Order*, *Duration*, *Transmitter address*, *Destination address*, *Source address*, *Receiver address*, *BSS Id* e *Sequence number*). Isto realiza, uma devida organização para a coleta de dados proposto a este estudo. Apontando, a identificação de ataques que impactam o funcionamento da rede sem fio IEEE 802.11, a ser utilizado como referência, em diferentes abordagens de diversos ecossistemas *wireless*, e com a necessidade de um atributo *Info* na fácil identificação do tipo de ataque e/ou dito como normal, além da representatividade na sua quantidade ao (Quadro 1).

Quadro 1 - Valores da amostragem de dados do tipo Normal, *Deauthentication*, EAPOL-*Logoff* e *Beacon Flood*.

Info	Quantidade de amostras
Normal	9134
Deauthentication	5094
EAPOL-*Logoff*	1428
Beacon Flood	1047
Total	16703

Fonte: Autores (2023).

Entretanto, a coleta de amostras peculiares exige a normalização que foi realizada através do *Java*[1] (Java, 2022) a fim de evitar ruídos sobre os dados reais, em específico com os rótulos, de Normal com o valor 0 (zero), e os demais ataques,

[1] A linguagem de programação *Java* é orientada a objetos, com o intuito de ser executada em qualquer plataforma ou até mesmo dispositivos.

Deauthentication valor correspondente a 1(um), *Beacon Flood* ao seu valor 2 (dois) e EAPOL-*Logoff* respectivo a 3 (três). Facilitando o todo, com o balanceamento propriamente dito na extração dos campos para um conjunto de dados normalizado, em obtenção de algoritmos de aprendizagem de máquina e as avaliações por métricas de desempenho.

2.3 Algoritmos de Aprendizagem de Máquina

Para introduzir efeito a este projeto foram relacionados algoritmos de aprendizagem de máquina. Com a categorização do conjunto de dados, para uma nova observação ao qual é pertencente, e é sobre o domínio da ferramenta *Waikato Environment for Knowledge Analysis* (*Weka*) (Witten et al., 2016) e os seus princípios funcionais que são uma característica desejável para IDS (Scarfone & Mell, 2007). Com a casualização da interface de programação de aplicação (do Inglês *Application Programming Interface* - API) em *Java* que o estudo das associações de anomalias ao IEEE 802.11 foram submetidos.

1) *Logistic Regression*: O modelo baseado em regressão logística tem objetivo de criar relação de dependência direta entre a variável de classe e as características, buscando trazer valores compreendidos entre 0 e 1, valores estes que representam a probabilidade de retornar o valor 1 para a expressão linear:

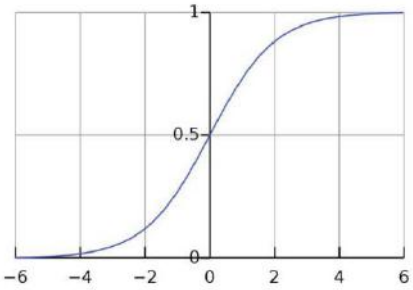
$$\theta x = \theta_0 + \theta_{1x1} + \theta_{2x2} + \ldots + \theta_{nxn} \quad \text{(Eq.1)}$$

Os valores de X e θ são vetores que alimentam a função hipótese. Tal função tem funcionalidade de determinar o valor de θ para que retorne o y esperado baseado no valor de entrada de x. A função hipótese é dada pela sigmoide:

$$h = g(z) = \frac{1}{1+e^{-z}} \quad \text{(Eq.2)}$$

Os valores gerados pela função são representados no gráfico abaixo (ver Figura 2). Nota-se que a função gera 2 assintotas, 1 tendendo a 0 para valores negativos de z e outra tendendo a 1 para valores positivos.

Figura 2 - Gráfico sigmoide para a função g(z)(Eq.2)



Fonte: Autores (2023).

Para que os valores apresentados na (Figura 2) sejam aproximados do modelo real, se faz necessário usar outra equação para tal, função esta denominada de função de perda logarítmica. Apresentada a seguir:

$$\text{Custo}\,(h\theta(x), y) = -\log(h\theta(x)) \text{ se } y = 1 \text{ ou } -\log(1 - h\theta(x)) \text{ se } y = 0 \quad \text{(Eq.3)}$$

Este modelo é mais adequado para métodos de classificação binária, podendo ser utilizado em funções multiclasses com maior mão de obra. Em *Weka*, a regressão logística é a implementação do algoritmo *Logistic e* pode ser encontrado por (Cessie & Houwelingen, 1995).

2) *J48*: Este tipo de algoritmo representa a forma de árvores binárias de decisão, mas com grande estabilidade entre tempo de precisão e cálculo, com menos esforço de treinamento sobre o algoritmo para classificadores não lineares. Pela *Weka*, a árvore de decisão é a implementação do algoritmo *J48*. Detalhes deste algoritmo pode ser encontrado em (Quinlan, 2014). Eventualmente a uso de entropia sobre o grau de incerteza dos elementos aleatórios e de ganho de informações comumente são as mais utilizadas neste tipo de algoritmo (Ravipati & Abualkibash, 2019). Apesar disso, a entropia irá calcular a homogeneidade das amostras detalhadas, da forma que, os dados analisados completamente como homogênea serão respectivos a zero, senão a entropia com a perspectiva a 1 (Ravipati & Abualkibash, 2019) na emissão da fórmula:

$$\mathrm{E(S)} = \sum_{x=1}^{x} -P_I \, log_2 \, P_x \qquad \text{(Eq.4)}$$

Já o ganho de informações vem com a obstinação de construir uma árvore de decisão, para estimar a informação sobre cada atributo retornando o maior ganho sobre o atributo independente. Isto induz, o ganho de informação (T,X) que aplica o recurso acerca do atributo X; porém, a Entropia(T) de encontro ao completo conjunto de dados, e a Entropia(T, X) com a sua aplicabilidade ao recurso sofre uma devida desvantagem, em relação ao ajuste do modelo na tratativa da divisão de dados de treinamento forte, e reduzir consideravelmente a precisão do teste (Ravipati & Abualkibash, 2019).

$$\text{Ganho de Informação (T, X)} = \text{Entropia(T)} - \text{Entropia (T, X)} \qquad \text{(Eq.5)}$$

3) *Naive Bayes*: Algoritmo altamente escalonável, exigindo um alto número de variáveis lineares (preditores) em um problema de aprendizagem. Calculando a probabilidade condicional de cada atributo seguido de uma aplicação contida do teorema de *Bayes* (ver equação 6), para determinar a probabilidade relativa sobre as características dos atributos, no sentido da previsão do resultado (Aggarwal, 2014). Maiores detalhes da implementação do algoritmo *Naive Bayes*, em *Weka*, pode ser encontrado por (John & Langley, 1995).

$$\mathrm{P(A|B)} = \frac{\mathrm{P(B|A)} \times \mathrm{P(A)}}{\mathrm{P(B)}} \qquad \text{(Eq.6)}$$

2.4 Métricas de Desempenho

Para esta devida situação, se dar no envolvimento de métricas, em *micro average* (Abracadabra, 2018), a serem usadas como medidas de desempenho específicas ao estudo. As abordagens adotam a viabilidade sobre a amostragem, e para isto foram adequados ao estudo:

1) Verdadeiro Positivo indica que o conjunto de dados classificados como ataque pelo modelo de classificação.

2) Verdadeiro Negativo prevê uma resposta do tipo não (normal) e está condizente com o dado observado na conexão.

3) Falso Positivo relaciona o número de instâncias classificadas como normais, mas sendo identificadas como anomalias pelo classificador.

4) Falso Negativo prevê uma conexão anômala como não, mas esta deveria ser sim.

5) Taxa de verdadeiro positivo (TPR, que é o tipo de sensibilidade) e a probabilidade de um teste real seja positivo. A (equação 7) define o tipo da medida.

$$TPR = \frac{\text{VP}}{\text{VP+FN}} \quad \text{(Eq.7)}$$

6) Taxa de falso positivo. Sendo a (equação 8) na determinância, de que o FP são os números de falsos positivos e o VN é o número de verdadeiros negativos, com a sua probabilidade a ser disparada quando o valor da intrusão for verdadeiro, mas ele é determinado como negativo.

$$FPR = \frac{\text{FP}}{\text{FP+VN}} \quad \text{(Eq.8)}$$

7) Acurácia analisa a capacidade de classificar corretamente um objeto de dados como normal ou anômalo. É definida com a (equação 9).

$$Acurácia = \frac{\text{VP+VN}}{\text{FP+VP+VN+FN}} \quad \text{(Eq.9)}$$

8) Precisão avalia a quantidade de classificações positivas que estão condizentes ao conjunto de dados. A (equação 10) define o tipo de medida.

$$Precisão = \frac{\text{VP}}{\text{VP+FP}} \quad \text{(Eq.10)}$$

9) *Recall* é a relação do classificador que pôde reconhecer os números de ataques positivos. A (equação 11) define a estrutura da medida.

$$Recall = \frac{\text{VP}}{\text{VP+FN}} \quad \text{(Eq.11)}$$

10) *F-Measure* considerada como uma precisão do classificador, além de definir uma média harmônica entre a medida de precisão e *recall*. Para o seu uso a (equação 12) é definida.

$$F - Measure = 2 \times \frac{\text{Precisão} \times \text{Recall}}{\text{Precisão+Recall}} \quad \text{(Eq.12)}$$

11) Taxa de Alarme Falso apenas calcula o número de predições incorretas do algoritmo classificador pelo número de verdadeiro positivo. A (equação 13) defini a característica da sua utilização.

$$Taxa\ de\ Alarme\ Falso = \frac{\text{FP + FN}}{\text{VP}} \quad \text{(Eq.13)}$$

12) MCC ou *Matthews Correlation Coefficient* é o coeficiente que mede a qualidade da classificação (Liu et al., 2014). A (equação 14) está delimitada para o uso.

$$MCC = \frac{(\text{VP}\times\text{VN}) * (\text{FP}\times\text{FN})}{\sqrt{(\text{VP}+\text{FP})*(\text{VP}+\text{FN})*(\text{VN}+\text{FP})*(\text{VN}+\text{FN})}} \quad \text{(Eq.14)}$$

13) ROC ou *Receiver Operator Characteristic Curve*, sendo uma medida da área de precisão do modelo que indica a compensação entre a taxa de verdadeiros positivos $TPR = \frac{\text{VP}}{\text{VP+FN}}$ e a taxa de falso positivo $FPR = \frac{\text{FP}}{\text{FP+VN}}$. Isto indica ao modelo propriamente dito e o conjunto de teste, a analisar os dados corretos e a compensação de dados incorretos.

14) Tempo que representa a construção do modelo de classificação até os resultados.

2.5 Classificação

O que determina o deslumbramento dos dados é ditado com as características do (Pseudocódigo 1). Apresentando de uma forma simples, as partes envolvidas de instâncias de dados relacionados aos algoritmos (*Naive Bayes*, *J48* e *Logistic*), em uma classificação supervisionada com base em padrões e associações dos dados rotulados a este estudo ao padrão de 70% em cima da base de treinamento e 30% sobre a base de teste. Além disso, a validação cruzada que divide em partes o conjunto de dados na garantia de que os mesmos estejam em uma forma aleatória de i=10 ou $E= \frac{1}{10}\sum_{i=10}^{i} E_i$ i.e., isolando o grupo de dados em via para treinamento para estimar os modelos, enquanto outra parte faz a relação ao teste, validando cada um dos modelos. E, os dados pré avaliados então identificados, em uma instância da classe respectiva e preditiva, são previamente analisados os seus resultados através de métricas de desempenho.

Pseudocódigo 1: Modelo de Processamento

1. Leitura do *Dataset*
2. Divisão da instância de dados em treinamento de 70%
3. Divisão da instância de dados em teste de 30%
4. Percorra cada modelo indicado (*Naive Bayes*, *J48* e *Logistic*)
5. Modelo (*Naive Bayes*, *J48* e *Logistic*) classifique os dados de treino
6. Avalie o modelo com os dados de teste
7. Realize a validação cruzada $E= \frac{1}{10}\sum_{i=10}^{i} E_i$
8. Identifique a classe de atributo (*Info*)
9. Percorra as classes de atributos $\sum_{i=4}^{i} E_i$ validas (*Info*)
10. Analise os dados através das métricas de desempenho
11. Apresente os valores das métricas de desempenho de cada classificador
12. Apresente o tempo em *ms* e o tempo total de cada modelo classificador

Fonte: Autores (2023).

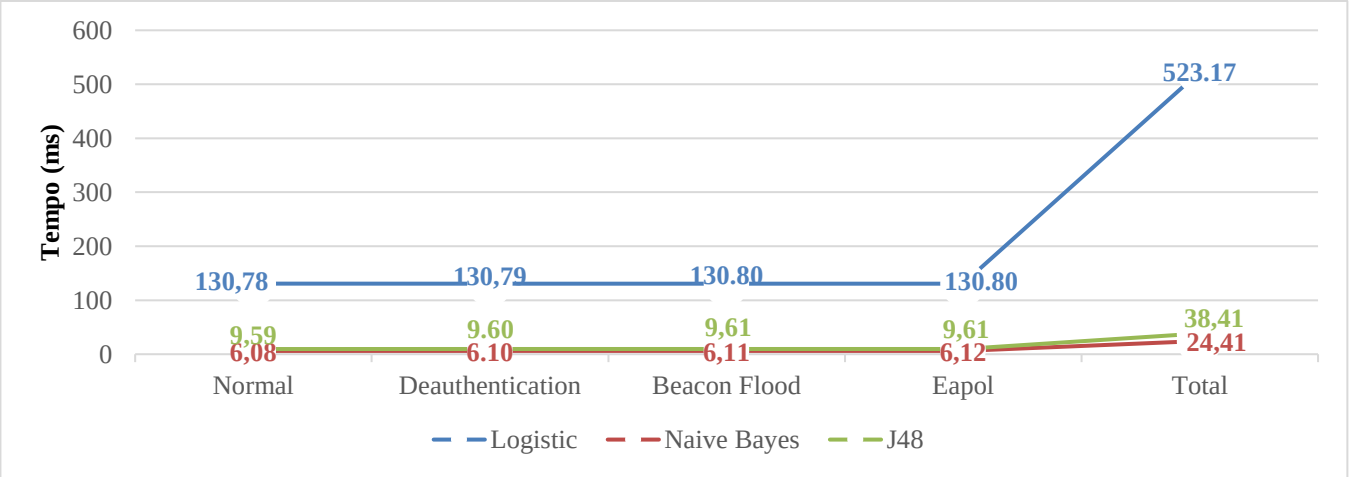
3. Resultados e Discussão

Nesta seção é basicamente, o desempenho dos modelos de classificação de aprendizagem de máquina que são avaliados sobre o cenário da rede *Wireless* ao contexto do IEEE802.11, em detecção das anomalias descritas a este estudo. Como descrito, avalia as devidas métricas de desempenho e a fluidez (ou rapidez). E para os experimentos utilizamos 1 (um) processador Intel(R) Core(TM) i7-3770 CPU @ 3.40GHz, 3401 *MHz*, 4 Núcleo(s), 8 Processador(es) Lógico(s), além de 8 (oito) GB's de memória *ram* a 1333 *MHz* e o sistema operacional *Windows* 10. Já o ambiente de desenvolvimento de *software* foi utilizado, o *IntelliJ* IDEA 2022.3.1 (*Community Edition*), API *Apache Spark* e API *Weka* em *Java*.

3.1 Tempo da Análise das Métricas de Desempenho

Com o objetivo de avaliar as métricas de desempenho e até, a relação de cálculo de tempo total computacional dos classificadores propostos ao estudo, obtivermos a velocidade de processamento, em milissegundos (*ms*) de 16.703 instâncias do conjunto de dados (ver Figura 3). Onde os tipos de dados anômalos, *Deauthentication*, *Beacon Flood*, EAPOL-*Logoff* e o dado não anômalo Normal, sendo identificados os seus tempos computacionais únicos, por cada um dos classificadores propostos ao estudo e o somatório total de cada um deles.

Figura 3 - Consumo computacional dos respectivos algoritmos, *Logistic*, *Naive Bayes* e *J48.* Significado por valores totais de cada uma das equivalências conforme (Normal, *Deauthentication*, *Beacon Flood* e EAPOL-*Logoff*), em *ms*.



Fonte: Autores (2023).

Os resultados analisados, na (Figura 3) indica que, na detecção do dado não anômalo (Normal), o classificador *Naive Bayes* obteve um custo de (6,08 *ms*), já o tipo de ataque *Deauthentication*, o seu valor foi respectivo a (6,10 *ms*), não obstante, para a anomalia *Beacon Flood* o valor apresentado é de apenas (6,11 *ms*), já o conseguinte ataque EAPOL-*Logoff* o valor foi de seu exato (6,12 *ms*) com um total de apenas (24,41 *ms*) sendo o melhor classificador, em custo computacional na constatação dos ataques supracitados. Contudo, o *J48* no tocante ao tipo de dado Normal foi obtido um custo computacional de aproximadamente (9,59 *ms*), em respectivo ao dado anômalo *Deauthentication* a importância de (9,60 ms), seguindo os outros tipos de ataques, *Beacon Flood* e EAPOL-*Logoff* com a similitude de (9,61 *ms*), e já se tratando do tempo total, o *J48* teve um desempenho não igual ao seu antecessor, mas de apenas (38,41 *ms*) acumulados. Sendo uma gama (γ) de dados, o *Logistic* e o seu respectivo e determinado valor de (130,78 *ms*) na detecção de dados normais, bem aproximado ao tipo de ataque *Deauthentication* com aproximadamente (130,79 *ms*), todavia, a anomalia *Beacon Flood* e EAPOL-*Logoff* foram obtido de valores similares (130,80 *ms*), com a consternação dos valores atingidos sobre o algoritmo *Logistic*, é de se admirar que o seu valor total é o maior de todos, com o seus exatos (523,17 *ms*).

3.2 Análise de Métricas de Performance sobre a Conjuntura de Intrusões em Redes IEEE 802.11 com Aprendizagem de Máquina no Hospital N.S.C.

A análise de métricas de *performance* para a combinação de ataques restritos ao estudo é analisada no Quadro 2, Quadro 3, Quadro 4, Quadro 5, apresentando alguns dos resultados obtidos, tanto na fase de treinamento, como que diz respeito à fase de teste e a comparação aos classificadores de aprendizagem de máquina propostos. Todos os resultados atingidos foram formadas de uma validação cruzada de $E = \frac{1}{10}\sum_{i=10}^{i} E_i$ ou de 10 vezes. Isto com as apresentações ditadas sobre

desempenho do estudo, que pode ser avaliado utilizando algumas e/ou várias métricas da matriz de confusão: acurácia, precisão e *recall* (Tarca et al., 2007), bem como as taxas de números verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos de cada classe. Entretanto, a uma complementação dos dados em amostragem com *F-measure*, a taxa de alarme falso, taxa de verdadeiro positivo, taxa de falso positivo, ROC e MCC.

3.3 Análise de Métricas de Performance Sobre o Então Tipo de Dado não Anômalo - Normal

Durante a fase de detecção respectivos ao projeto, o tipo de dado não anômalo (Normal) ao classificador *Naive Bayes*, *Logistic* e *J48*, foram bem analisados. Identificando, no Quadro 2, as taxas de verdadeiros positivos entre as marcas de (8.249,0) ao *Naive Bayes*, a indicação do *Logistic* apresentando um número (8.246,0), além de *J48* e seus (8.226,0) números. Haja visto que, os números de similitudes, em verdadeiros negativos de número (7.569,0), falsos positivos de número (0,0) e a taxa de falso positivo com (0,00%) sendo bem viável para um sistema de detecção de intrusão, além de uma precisão de exatamente (100,00%) entre todos os classificadores. Relacionado ao desempenho, em identificar que estes valores são coexistentes ao tipo de dado, como normal, além de induzir a uma acurácia de (94,70%) respectivo ao *Naive Bayes*, bem aproximado o *Logistic* (94,68%) e o classificador *J48* (94,56%). Apesar de, identificar os falsos negativos com (885,0) números sobre o *Naive Bayes* e *Logistic* com (888,0) números e logo acima (908,0) números ao *J48*.

Quadro 2 - Métricas de desempenho em avaliação dos classificadores (*Naive Bayes*, *Logistic* e *J48*) pelo tipo de dado não anômalo (Normal).

Classificador	VP	VN	FP	FN	A %	P %	R %	F-M %	TAF %	TVP %	TFP %	RC %	MCC %	T (*ms*)
Naive Bayes	8.249,0	7.569,0	0,0	885,0	94,70	100,00	90,31	94,91	10,73	90,31	0,00	96,70	89,92	6,08
Logistic	8.246,0	7.569,0	0,0	888,0	94,68	100,00	90,28	94,89	10,77	90,28	0,00	96,59	89,89	130,78
J48	8.226,0	7.569,0	0,0	908,0	94,56	100,00	90,06	94,77	11,04	90,06	0,00	96,47	89,67	9,59
Total de amostras: 16.703,0														

Legenda: VP = Verdadeiro positivo; VN = Verdadeiro negativo; FP = Falso positivo; FP = Falso negativo; A = Acurácia; P = Precisão; R = Recall; F-M = F-*measure*; TAF = Taxa de alarme falso; TVP = Taxa de verdadeiro positivo; TFP = Taxa de falso positivo; RC = ROC ou *Receiver Operator Characteristic Curve*; MCC = *Matthews Correlation Coefficient*. Linhas com dados em seu volume total de (VP, VN, FP e FN) se diferenciam das devidas percentagens (%) entre os tratamentos. Fonte: Autores (2023).

Conforme, citado ao estudo no Quadro 2, existem poucas divergências de valores entre os classificadores apresentados através das métricas de desempenho. E isto, é respectivo com o *recall* do *Naive Bayes* (90,31%), já aproximado o *Logistic* (90,28%) e o classificador *J48* apresenta (90,06%). Apesar disso, o *F-measure* equivale a um valor um tanto quanto parecido do *Naive Bayes* e *Logistic*, alcançando (94,91%) e (94,89%), enquanto o *J48* com a taxa de (94,77%). A maior taxa de alarme falso foi apresentado por *J48* (11,04%), o *Logistic* com (10,77%) e o *Naive Bayes* com a menor taxa de (10,73%).

Respectivamente, o *Naive Bayes* em detecção de dados não anômalos em taxa de verdadeiro positivo foi de (90,31%), seguido do *Logistic* com uma taxa de (90,28%) e o *J48* e os seus respectivos (90,06%). Já a área da curva ROC ao algoritmo *Naive Bayes* (96,70%), um pouco abaixo o *Logistic* apresentando (96,59%) e *J48* com a taxa de (96,47%). Já se tratando de qualidade da classificação conforme o MCC, o *J48* obteve o menor valor (89,67%), seguido do *Logistic* demonstrando uma taxa de (89,89%) e a proporção do *Naive Bayes* sendo a maior entre todos os classificadores (89,92%). Entrentanto, a indicação da apuração de tempo computacional é muito relevante para a detecção do tipo não anômalo, sendo o *Naive Bayes* que representa (6,08 *ms*), seguido do *J48*, e seu perculiar tempo de (9,59 *ms*) e o por último o algoritmo *Logistic* denotando (130,78 *ms*).

3.4 Análise de Métricas de Performance Sobre o Então Tipo de Dado Anômalo - *Deauthentication*

Quanto a busca pela detecção de dados anômalos ao estudo, que é do tipo *Deauthentication.* Existem dados sobre as métricas de desempenho correspondentes ao Quadro 3. Ao qual os classificadores *Naive Bayes* e *Logistic* demonstram que arquivaram a mesma métrica de verdadeiro positivo, com seu exato número (5.094,0), enquanto a diferença do *J48* e seus (5.085,0) números de verdadeiros positivos. A diferença entre os resultados é que o *Naive Bayes* apresentou (10.724,0) verdadeiros negativos e falsos positivos (885,0), e todavia, o classificador *Logistic* apresentou (10.722,0) verdadeiros negativos e (887,0) falsos positivos, contudo, o *J48* obteve um valor menor de verdadeiros negativos, com exato (10.715,0) números e um valor maior de falsos positivos (894,0) números. Apesar disso vale salientar, que o *Naive Bayes* e *Logistic* obtiveram uma taxa de falsos negativos meramente igual a (0,0), e isto vem a identificar que o *J48* obteve um número (9,0) de falsos negativos praticamente maior. No entanto, pelo leve aumento de números respectivos ao verdadeiro positivo, verdadeiro negativo, falso positivo e falso negativo, o *J48* apresentou um valor um pouco abaixo da acurácia com apenas (94,59%), enquanto, os algoritmos *Naive Bayes* e *Logistic* adquiriram uma taxa de (94,70%) e (94,69%).

Quadro 3 - Métricas de desempenho em avaliação dos classificadores (*Naive Bayes*, *Logistic* e *J48*) pelo tipo de dado anômalo (*Deauthentication*).

Classificador	VP	VN	FP	FN	A %	P %	R %	F-M %	TAF %	TVP %	TFP %	RC %	MCC %	T (*ms)*
Naive Bayes	5.094,0	10.724,0	885,0	0,0	94,70	85,20	100,00	92,01	17,37	100,00	7,62	96,12	88,71	6,10
Logistic	5.094,0	10.722,0	887,0	0,0	94,69	85,17	100,00	91,99	17,41	100,00	7,64	96,03	88,69	130,79
J48	5.085,0	10.715,0	894,0	9,0	94,59	85,05	99,82	91,85	17,76	99,82	7,70	95,99	88,47	9,60
Total de amostras: 16.703,0														

Legenda: VP = Verdadeiro positivo; VN = Verdadeiro negativo; FP = Falso positivo; FP = Falso negativo; A = Acurácia; P = Precisão; R = Recall; F-M = F-*Measure*; TAF = Taxa de alarme falso; TVP = Taxa de verdadeiro positivo; TFP = Taxa de falso positivo; RC = ROC ou *Receiver Operator Characteristic Curve*; MCC = *Matthews Correlation Coefficient*. Linhas com dados em seu volume total de (VP, VN, FP e FN) se diferenciam das devidas percentagens (%) entre os tratamentos. Fonte: Autores (2023).

Contudo, com o alto número de falsos positivos (894,0), *J48* apresentou a pior precisão com (85,05%). *Naive Bayes* e *Logistic* obtiveram uma margem um pouco maior (85,20%) e (85,17%). De forma, a ter uma completude de ataque correspondente a *Deauthentication*, o *recall* de *J48* foi de (99,82%) e, *Naive Bayes* e *Logistic* alcançaram (100,00%). Com a média harmônica entre a precisão e *recall* foi então obtido, o *F-measure* do classificador *J48* (91,85%), logo, seguido *Naive Bayes* e *Logistic* a valores de (92,01%) e (91,99%). Além disso, na taxa de alarme falso o algoritmo *J48* resultou em um valor basicamente maior (17,76%), e respectivamente o *Naive Bayes* e *Logistic* obtiveram valores com uma taxa menor apresentada (17,37%) e (17,41%). No entanto, o Quadro 3 ajuda a salientar os valores designados sobre a taxa de verdadeiro positivo ao algoritmo *Naive Bayes* e *Logistic* (100,00%) e logo abaixo, o *J48* e seus (99,82%). Além disso, a taxa de falso positivo, em *Naive Bayes* é de apenas (7,62%), *Logistic* (7,64%), mas sendo resultante para *J48* (7,70%). No entanto, a área da curva ROC foi de um parecer, em *Logistic* aos seus (96,03%), já ao *Naive Bayes* uma leve alta de (96,12%) e o *J48* simplesmente com a taxa de (95,99%). Mesmo assim, o coeficiente de qualidade MCC sobre o *Naive Bayes* é respectivo a (88,71%), com o *Logistic* a um custo de (88,69%) e posteriormente a árvore de decisão (*J48*) obteve uma taxa de apenas (88,47%). Todavia, vale salientar, que o tempo computacional do *Logistic* foi o maior entre todos os classificdores, acompanhado por (130,79 *ms*), na detecção da anomalia *Deauthentication*, seguido do menor valor apresentado pelo *Naive Bayes* (6,10 *ms*) e, logo após, o *J48* e o seu valor computacional de (9,60 *ms*).

3.5 Análise de Métricas de Performance Sobre o Então Tipo de Dado Anômalo - *Beacon Flood*

Mediante a classificação seguinte ao *dataset* identificado a este estudo, no Hospital Nossa Senhora da Conceição (H.N.S.C.). O tipo de dado anômalo *Beacon Flood* pode ser analisado através de medidas de desempenho caracterizados aos classificadores interpostos ao Quadro 4. Todavia, há um número de verdadeiros positivos (1.047,0), entre, *Naive Bayes*, *Logistic* e *J48*. Entretanto, não existe similitude de dados aos números de verdadeiros negativos, com o *Naive Bayes* (15.656,0), logo em seguida, o *Logistic* e seus (15.655,0) números, e seguido do *J48* (15.633,0) números. Além disso, existe uma discrepância de falsos positivos, o *Naive Bayes* apresenta um número (0.0), em seguida um pouco acima ou bastante irrelevante, o *Logistic* e seu número (1,0), logo o *J48* relevantes (23,0) números. Com o desenvolvimento sobre os números de falsos negativos onde houve praticamente (0.0) ou nulo sobre os classificadores apresentados.

Quadro 4 - Métricas de desempenho em avaliação dos classificadores (*Naive Bayes*, *Logistic* e *J48*) pelo tipo de dado anômalo (*Beacon Flood*).

Classificador	VP	VN	FP	FN	A %	P %	R %	F-M %	TAF %	TVP %	TFP %	RC %	MCC %	T (*ms)*
Naive Bayes	1.047,0	15.656,0	0,0	0,0	100,00	100,00	100,00	100,00	0,00	100,00	0,00	100,00	100,00	6,11
Logistic	1.047,0	15.655,0	1,0	0,0	99,99	99,90	100,00	99,95	0,10	100,00	0,01	100,00	99,95	130,80
J48	1.047,0	15.633,0	23,0	0,0	99,86	97,85	100,00	98,91	2,20	100,00	0,15	99,95	98,85	9,61
Total de amostras: 16.703,0														

Legenda: VP = Verdadeiro positivo; VN = Verdadeiro negativo; FP = Falso positivo; FP = Falso negativo; A = Acurácia; P = Precisão; R = Recall; F-M = F-*Measure*; TAF = Taxa de alarme falso; TVP = Taxa de verdadeiro positivo; TFP = Taxa de falso positivo; RC = ROC ou *Receiver Operator Characteristic Curve*; MCC = *Matthews Correlation Coefficient*. Linhas com dados em seu volume total de (VP, VN, FP e FN) se diferenciam das devidas percentagens (%) entre os tratamentos. Fonte: Autores (2023).

Com a construção de taxas a serem analisadas pela acurácia propriamente dita, houve um desempenho muito bom entre todos os algoritmos, sendo o *Naive Bayes* apresentando (100,00%) de assertividade, o *Logistic* com a sua taxa de (99,90%), e pouco abaixo o *J48*, com os seus (99,86%). Já a precisão do *Logistic* apresentou (99,90%) e em seguida, *Naive Bayes* (100,00%), enquanto levemente abaixo o *J48* (97,85%). Em seguida foram obtidos os valores do *recall*, a uma paridade entre todos os algoritmos, com uma taxa de (100,00%). Da mesma forma, o Quadro 4 apresenta valores, em *F-measure* de (100,00%) ao *Naive Bayes*, no *Logistic* é visualizado a taxa de (99,95%), com a completeza do algoritmo *J48* busca seu valor a uma taxa de (98,91%). Apesar disso, a taxa de alarme falso combinado ao *J48* foi o maior (2,20%), no entanto, a taxa do *Naive Bayes* com o valor de (0,00%) e *Logistic* (0,10%). Entretanto, os algoritmos supracitados obtiveram um valor de similaridade de taxa de verdadeiro positivo (100,00%). Já a taxa de falso positivo para o *Naive Bayes* foi de exatamente (0,00%) e *Logistic* de (0,01%), e o *J48* apresentou (0,15%). Em argumentação da taxa de falso positivo, o classificador *Naive Bayes* obteve a menor taxa com seu exato (0,00%), acompanhado do classificador *Logistic* e a taxa de (0,01%), e concomitante o *J48* a uma taxa de (0,15%). Para a obtenção da área da curva ROC, os algoritmos citados ao estudo atingiram um valor de (100,00%) entre o *Naive Bayes* e *Logistic*, logo, o *J48* obteve a taxa de (99,95%). O MCC a ser apresentando a qualidade do *Naive Bayes* com o maior valor (100,00%), o *Logistic* com (99,95%) e em prossecução com o valor abaixo dos outros algoritmos, o *J48* (99,85%). E através, de um poder computacional, foi preciso analisar no Quadro 4, que o tempo computacional do algoritmo *Naive Bayes*, foi o mais momentoso, com (6,11 *ms*), em seguida com um valor um pouco acima, o classificador *J48* exibindo um taxa de (9,61 *ms*), em sequência, o maior valor entre todos os classificadores apresentado pelo *Logistic* e o exato (130,80 *ms*).

3.6 Análise de Métricas de Performance Sobre o Então Tipo de Dado Anômalo - EAPOL-*Logoff*

Finalmente, para a detecção do ataque EAPOL-*Logoff*, em um Quadro 5. Salienta, que os números interpostos a cada um dos algoritmos correlacionados ao estudo foram muito bem apresentados, com uma valia de 100% sobre as taxas respectivas diante das métricas (Acurácia, Precisão, *Recall*, *F-measure*, Taxa de Alarme Falso, Taxa de Verdadeiro Positivo, Taxa de Falso Positivo, ROC e MCC), além ter números agradáveis de verdadeiros positivos (1.428,0) entre, *Naive Bayes*, *Logistic* e *J48.* E seguidos de números de verdadeiros negativos (15.275,0) entre todos os classificadores designados, e com um ótimo valor de falsos negativos (0.0), em seguida atraídos a um número (0,0) de falsos positivos.

Quadro 5 - Métricas de desempenho em avaliação dos classificadores (*Naive Bayes*, *Logistic* e *J48*) pelo tipo de dado anômalo (EAPOL-*Logoff*).

Classificador	VP	VN	FP	FN	A %	P %	R %	F-M %	TAF %	TVP %	TFP %	RC %	MCC%	T (*ms)*
Naive Bayes	1.428,0	15.275,0	0,0	0,0	100,00	100,00	100,00	100,00	0,00	100,00	0,000	100,00	100,00	6,12
Logistic	1.428,0	15.275,0	0,0	0,0	100,00	100,00	100,00	100,00	0,00	100,00	0,000	100,00	100,00	130,80
J48	1.428,0	15.275,0	0,0	0,0	100,00	100,00	100,00	100,00	0,00	100,00	0,000	100,00	100,00	9,61
Total de amostras: 16.703,0														

Legenda: VP = Verdadeiro positivo; VN = Verdadeiro negativo; FP = Falso positivo; FP = Falso negativo; A = Acurácia; P = Precisão; R = Recall; F-M = F-*Measure*; TAF = Taxa de alarme falso; TVP = Taxa de verdadeiro positivo; TFP = Taxa de falso positivo; RC = ROC ou *Receiver Operator Characteristic Curve*; MCC = *Matthews Correlation Coefficient*. Linhas com dados em seu volume total de (VP, VN, FP e FN) se diferenciam das devidas percentagens (%) entre os tratamentos. Fonte: Autores (2023).

Como não houve perda, mas ganho de informações ao que indica o Quadro 5, as taxas de dados da acurácia, em todos os classificadores apresentados ao estudo foi de uma relação de (100,00%). Logo, em seguida os mesmos (*Naive Bayes*, *Logistic* e *J48*) acrescentaram mais detalhes aos seus, com uma precisão de (100,00%) e um *recall* absurdo de (100,00%). O sistema apresenta uma ideia muito relacionada, para o teste da métrica *F-measure*, com um detalhe apresentando (100,00%) de todos os classificadores, e com as predições propriamente ditas ao estudo foram bem relevantes fazendo com que a taxa de alarme falso não ultrapassasse a gama de (0,00%) entre todos os algoritmos. Contudo, a taxa de verdadeiro positivo foi bastante elevado, com uma taxa de (100,00%) entre todos os classificadores. Dito isso, a taxa de falso positivo foi de seus relevantes ao *Naive Bayes*, *Logistic* e *J48*, de apenas (0,00%), já a área da curva ROC apresentou uma significância de (100,00%) entre o *Naive Bayes*, *Logistic* e *J48*, e observa-se que todos os classificadores foram bem qualificados a taxa de (100,00%) sobre o MCC. Contudo, friza que o custo computacional apresentado pelo Quadro 5, vem a salientar, que o *Naive Bayes* identifica um valor menor do que os outros classificadores, com o acurado aproveitamento de (6,12 *ms*), em consequência da detecção do ataque EAPOL-*Logoff*, o classificador *J48* apresenta um valor um pouco acima (9,61 *ms*), e no que diz respeito ao classificador *Logistic*, exibe o maior valor (130,80 *ms*).

3.7 Discussão

Todavia, um IDS para analisar, processar e classificar as informações em intrusão ou normal é primordial para tomada de decisão que venha a ocorrer sobre a rede *wireless*. Esta primazia buscou uma grande eficiência, em razão da obtenção de resultados sobre um grande número de verdadeiro positivo. Entretanto, o número de falso positivo foi satisfatório, sendo pequeno ou até zero, e este estudo expôs. Apesar disso, para relacionar a estrutura do projeto, o modelo de Aminanto et al. (2022) utilizou do conjunto de dados *Aegean* Wi-Fi *Intrusion Dataset* 2 (AWID2) que tem sido corroborado em diversos estudos e da rede neural convolucional (CNN) na classificação de ataques a rede Wi-Fi com a perspectiva sobre a métrica de avaliação F1-Score com pontuação de (99,73%) em detecção de anomalia.

Assim, na base de dados WSN-DS, Quincozes & Kazienko (2020) abordaram o classificador de *J48* (Árvore de Decisão), além de *Naive Bayes*, *REP Tree*, *Random Tree* e *Random Forest*, em processamento na melhor categorização de "buracos cinzentos" e "buracos negros" (i.e., ataques com similitude a DoS em redes de sensores sem fio), e enquanto ao uso de *Random Tree* (Árvore Aleatória) vem a categorizar melhor a detecção de "inundações", mas avaliando os dados por acurácia. Constatando durante a detecção de "buraco negro", que o *J48* alcançou (97,88%) de acurácia, enquanto *REP Tree* (97,89%), sendo os algoritmos mais precisos. Entretanto, o *Naive Bayes* (97,47%), *Random Forest* e *Random Tree* apresentaram uma taxa de (97,71%) e (97,72%). No entanto, quanto a detecção de "buraco cinza", o *J48* e *REP Tree* arquivaram a mesma acurácia (98,11%). Apesar disso, o *Naive Bayes* apresentou a menor acurácia (97,50%), e entre o *Random Forest* e *Random Tree* obtiveram (98,06%) e (98,07%). Posteriormente, os autores detectaram ataques de "inundação", obtendo o *Random Forest* com um valor de (99,13%) de acurácia, com um pouco menor acurácia (99,11%) ao *Random Tree*, e os demais algoritmos apresentaram variações de dados.

E em via a isto, na abordagem de Qin et al. (2018) vêm a utilizar o conjunto de dados *Aegean* Wi-Fi *Intrusion Dataset* (AWID) com a seleção de 18 atributos úteis em vez de 154, para a uma *performance* na melhoria da precisão de detecção de anomalias através de máquina de vetor de suporte (SVM) com a aproximação de 89,18%, 87,34% e 99,88%, em ataques de "inundação", ataques de "injeção" e dados normais. Outros resultados promissores também foram obtidos, quando Patil & Agarkhed (2020) utilizaram o paradigma da função *Radial Bias*, para a melhor taxa de detecção de anomalias em WSN (Redes de Sensor sem Fio) e ao encontro da baixa quantidade de falsos positivos, com base no conhecimento de técnicas de árvore de decisão na acurácia de 98,00% sobre os ataques *Sleep Deprivation* e *Sinkhole*.

Assim os benefícios de aprendizagem de máquina por meio do ambiente proposto, em detecção de intrusão em redes sem fio, são percebidos através do aumento de verdadeiros positivos e uma baixa gama de falsos positivos. Mas com base nos resultados logo visto acima, o algoritmo mais conciso, irá depender do tipo de ataque. Pois, em vista ao tipo de ataque EAPOL-*Logoff*, os classificadores *Naive Bayes*, *Logistic* e *J48* obtiveram ótimos resultados, entre todas as métricas de avaliação. E, em nossas descobertas revelam que a detecção de *Beacon Flood*, foram obtidos um ótimo desempenho em métricas de avaliação, principalmente entre o *Naive Bayes* e o *Logistic*, indicando maiores resultados precisos, em vez do *J48*. E em geral, por se tratar de um tempo em (*ms*), o algoritmo *Naive Bayes* demonstrou ser mais rápido. Já se tratando do tipo de ataque *Deauthentication* e o seu envio simultâneo de quadros irreais. Sucedeu valores designados de falsos positivos sobre uma média (888,6) números em via aos algoritmos *Naive Bayes*, *Logistic* e *J48*, mas acrescentam que as taxas designadas são formidáveis. Contudo, não apresentando um desempenho tanto quanto aos demais ataques, o EAPOL-*Logoff*, *Beacon Flood*, e o conseguinte tipo não anômalo (Normal), que tal representação dos algoritmos proporcionam resultados significativos.

Portanto, o uso de aprendizagem de máquina demonstra ser uma tecnologia profunda e com várias descobertas existentes sobre o quadro MAC 802.11, em via a ataques correspectivos nas redes *wireless*.

4. Conclusão

Neste estudo propusemos o lançamento de anomalias em uma rede corporativa contida de WPA2, na fomentação do quadro MAC 802.11 e os seus atributos (*Protocol Version*, *Type*, *To DS*, *From DS*, *More Fragment*, *Retry*, *Power Management*, *More Data*, *WEP*, *Order*, *Duration*, *Transmitter Address*, *Destination Address*, *Source Address*, *Receiver Address*, *BSS Id* e *Sequence Number*). Além disso, apresentado por uma técnica de aprendizagem supervisionada, em sub-amostragem aleatória da classe majoritária aplicada ao uso de filtragem a uma determinada seleção de percentagem (30%), para sobreamostragem aleatória da classe minoritária, além da condição de uma validação cruzada e os classificadores prevendo um rótulo de classe categórico, a uma instância atribuída da característica (i.e., *Info*) de informação ao tipo de ataque correspondente ao estudo.

Com um conjunto de dados real. Tanto quanto, sabemos que o pré-processamento de dados foi bastante importante, assim a introduzir melhor desempenho, em relação aos tipos de ataques identificados ao estudo. Apesar disso, as métricas contidas atingindo valores acima da média respectivamente, o que é comparável e observado em estudos existentes. E entretanto, operamos com melhores detalhes, sobre os algoritmos citados (*Naive Bayes*, *Logistic* e *J48*).

Portanto, como trabalho futuro pretendemos melhorar o conjunto de dados e o modelo de processamento, afim de aplicar métodos (i.e., algoritmos para classificar em tempo real), além de reconhecer outros tipos de ataques conforme semi-automática ou automática aplicado em aprendizagem de máquina.

Referências

Abracadabra (2018). Micro- and Macro-average of Precision, Recall and F-Score. Website Tomaxent. https://tomaxent.com/2018/04/27/Micro-and-Macro-average-of-Precision-Recall-and-F-Score/.

Aggarwal C. C. (2014). *Data Classification*: Algorithms and Applications. Chapman & Hall/CRC.

Ahmad, M. S., & Tadakamadla, S. (2011). Short Paper: Security Evaluation of IEEE 802.11w Specification. In *Proceedings of the Fourth ACM Conference on Wireless Network Security*.Association for Computing Machinery, 53–58. http://dx.doi.org/10.1145/1998412.1998424.

Aircrack-ng. AIRCRACK-NG(2022). http://www.aircrack-ng.org/doku.php.

Aminanto, M. E., Wicaksono, R. S. H., Aminanto, A. E., Tanuwidjaja, H. C., Yola, L., & Kim, K. (2022). Multi-Class Intrusion Detection Using Two-Channel Color Mapping in IEEE 802.11 Wireless Network. *IEEE Access*, 10, 36791–36801. https://doi.org/10.1109/ACCESS.2022.3164104.

Arasaki, A. M. & Della Flora, J. C. L. (2012). *Teste de intrusão em redes sem fio padrão 802.11*. 63p. Monografia - Curso de Pós-Graduação em Redes de Computadores e Segurança de Dados. Centro Universitário Filadélfia de Londrina - UniFil, Londrina.

Barford, P., Kline, J., Plonka, D., & Ron, A. (2002). A Signal Analysis of Network Traffic Anomalies. In *Proceedings of the 2nd ACM SIGCOMM Workshop* on Internet Measurment. Association for Computing Machinery, 71–82. https://doi.org/10.1145/637201.637210.

Cessie, S. L., & Houwelingen, J. C. V. (1992). Ridge Estimators in Logistic Regression. *Journal of the Royal Statistical Society. Series C* (*Applied Statistics*), 41(1), 191–201.http://dx.doi.org/10.2307/2347628.

Feng, P. (2012). Wireless LAN security issues and solutions. In *2012 IEEE Symposium on Robotics and Applications* (*ISRA*), 921–924. https://doi.org/10.1109/ISRA.2012.6219343.

IEEE Standard for Information Technology- Telecommunications and Information Exchange Between Systems- Local and Metropolitan Area Networks-Specific Requirements- Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. (2003). *ANSI/IEEE Std 802.11, 1999 Edition (R2003)*, i-513. https://doi.org/10.1109/IEEESTD.2003.95617.

IEEE Standard for Information technology - Telecommunications and information exchange between systems - Local and metropolitan area networks - Specific requirements. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 4: Protected Management Frames. (2009). *IEEE Std 802.11w-2009 (Amendment to IEEE Std 802.11-2007 as amended by IEEE Std 802.11k-2008, IEEE Std 802.11r-2008, and IEEE Std 802.11y-2008)*, 1–111.https://doi.org/10.1109/IEEESTD.2009.5278657.

IEEE Standard for information technology-Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements-Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Amendment 6: Medium Access Control (MAC) Security Enhancements. (2004). *IEEE Std 802.11i-2004*, 1–190. https://doi.org/10.1109/IEEESTD.2004.94585.

Java (2022). Java. https://www.java.com.

John, G. H., & Langley, P. (1995). Estimating Continuous Distributions in Bayesian Classifiers. *In Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann Publishers Inc. 338–345. https://dl.acm.org/doi/10.5555/2074158.2074196.

Linhares, A.G., & Gonçalves, P. A. da S. (2012). Uma análise dos mecanismos de segurança de redes IEEE 802.11: WEP, WPA, WPA2 e IEEE 802.11 w. 1-10. https://www.cin.ufpe.br/~pasg/gpublications/LiGo06.pdf.

Liu, Y., Cheng, J., Yan, C., Wu, X., & Chen, F. (2015b). Research on the Matthews Correlation Coefficients Metrics of Personalized Recommendation Algorithm Evaluation. *International Journal of Hybrid Information Technology*, 8(1), 163–172. https://gvpress.com/journals/IJHIT/vol8_no1/14.pdf.

Mdk3. Penetration Testing Tools. (2022). https://en.kali.tools/?p=34.

Mitchell, T. (1997). *Machine Learning (Mcgraw-Hill International Edit)*. McGraw-Hill Education (ISE Editions).

Morimoto, C. E. (2008). *Redes, Guia Prático*. Sul Editores.

Patil, B., & Agarkhed, J. (2020). An Exploratory Machine Learning Technique for Investigating Intrusion in Wireless Sensor Networks. In *2020 IEEE Bangalore Humanitarian Technology Conference* (*B-HTC*), 1–6. https://doi.org/10.1109/B-HTC50970.2020.9297969.

Qin, Y., Li, B., Yang, M., & Yan, Z. (2018). Attack Detection for Wireless Enterprise Network: a Machine Learning Approach. In *2018 IEEE International Conference on Signal Processing, Communications and Computing* (*ICSPCC*), 1–6. https://doi.org/10.1109/ICSPCC.2018.8567797.

Quincozes, S. E., & Kazienko, J. F. (2020). Machine Learning Methods Assessment for Denial of Service Detection in Wireless Sensor Networks. In *2020 IEEE 6th World Forum on Internet of Things* (*WF-IoT*), 1–6. https://doi.org/10.1109/WF-IoT48130.2020.9221146.

Quinlan, J. (1995). MDL and Categorical Theories (Continued). In A. Prieditis & S. Russell (Eds.), *Machine Learning Proceedings 1995*. Morgan Kaufmann. 464–470.

Ravipati, R. D., & Abualkibash, M. (2019). Intrusion Detection System Classification Using Different Machine Learning Algorithms on KDD-99 and NSL-KDD Datasets - A Review Paper. *SSRN Electronic Journal*, 11(3),1-16. http://dx.doi.org/10.2139/ssrn.3428211.

Scarfone, K. A., & Mell, P. M. (2007). *Guide to Intrusion Detection and Prevention Systems (IDPS)*. National Institute of Standards and Technology, 800-94.

Tarca, A. L., Carey, V. J., Chen, X.-w., Romero, R., & Drăghici, S. (2007). Machine Learning and Its Applications to Biology. *PLoS Computational Biology*, *3*(6), e116.

Tews, E. (2007). Attacks on the WEP Protocol. *Cryptology ePrint Archive*, 471, 1-125. https://eprint.iacr.org/2007/471.pdf.

Wi-Fi Alliance (2003). Wi-Fi Protected Access: Strong, Standards-based, Interoperable Security for Today's Wi-Fi Networks. https://www.cs.kau.se/cs/education/courses/dvad02/p1/Papers%20Wireless/Wi-Fi%20Protected%20Access%20-%20Whitepaper.pdf.

Wireshark (2022).The world's most popular network protocol analyzer. https://www.wireshark.org/.

Witten, I. H., Frank, E., A, H. M., & Pal, C. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier Science & Technology Books.