

Análise de dados para previsão de defeito em sistemas de radiocomunicação naval*
Data analysis for defect prediction in naval radio communication systems

Alan Lopes Pombo

Universidade Federal Fluminense
alanpombo@id.uff.br

Vitor Hugo Ferreira

Universidade Federal Fluminense
vhferreira@id.uff.br

Marcio Zamboti Fortes

Universidade Federal Fluminense
mzamboti@id.uff.br

RESUMO ESTRUTURADO

O presente trabalho estudará um caso real. Cada tipo de embarcação deve atender aos requisitos mínimos de segurança marítima, entre muitos outros aspectos de segurança, o item de radiocomunicação pode levar ao risco de morte ou danos irreversíveis. É um dos principais itens da lista de verificação para que a embarcação seja liberada para navegar. A proposta é detectar um defeito antes que ele cause a interrupção da comunicação, com coleta de dados por meio da simulação de defeitos típicos, utilizando técnicas de classificação, que é uma subárea de Aprendizado de Máquina, atribuindo uma classe a cada nova observação com base nessas informações. A partir dos resultados da análise e do processamento, será feito um diagnóstico para tomar as medidas corretivas necessárias antes que o sistema seja interrompido. Hardware específico foi desenvolvido para permitir a obtenção dos dados tratados aqui.

Palavras-chave: Aprendizado de Máquina, Radio Comunicação, Análise de Dados

STRUCTURED ABSTRACT

This study focuses on a real case where each type of vessel must meet the minimum requirements of maritime safety. Among various safety aspects, radiocommunication is a critical component that, if compromised, can pose risks of death or irreversible damage. It stands as a key item on the checklist for a vessel to be cleared for sailing. The objective is to detect defects before they lead to communication failure. This is achieved through data collection involving the simulation of typical defects, utilizing classification techniques—a sub-area of Machine Learning. These techniques assign a class to each new observation based on the collected information. Following the analysis and processing of results, a diagnosis will be formulated to implement necessary corrective measures before any disruption occurs in the system. Specific hardware has been developed to facilitate the collection and treatment of data for this purpose.

Keywords: Machine Learning, Radio Communication, Data Analysis.

1. INTRODUÇÃO

Sistemas de comunicação estão presentes em muitas áreas, sendo de fundamental importância em diversos processos, são constituídos por vários equipamentos, a sua complexibilidade varia conforme sua tecnologia envolvida e conforme a quantidade de hardware necessário, eles são suscetíveis à degradação devido ao tempo de utilização e as intempéries da natureza [1].

Conhecendo a estrutura básica desses sistemas de comunicação é possível prever onde ocorrem essas falhas e em quais equipamentos, preferencialmente antes que estas falhas venham acarretar total paralização do sistema [1].

Devido ao histórico de manutenções e análises de defeitos em diversos equipamentos, e com o know-how neste segmento, é possível realizar uma manutenção mais objetiva e específica.

A manutenção preventiva seria uma boa prática, porém estamos falando de sistemas de comunicação para embarcações, que em vários casos não dispõem de mão de obra especializada qualificada para realização dessas manutenções preventivas. O custo financeiro torna-se elevado e por diversas vezes não identifica uma possível falha que venha a ocorrer [1].

A implementação de novas tecnologias e soluções tem o intuito de ir muito além dessas manutenções preventivas, com a utilização de hardwares embarcados desenvolvidos especificamente para essas soluções é possível reduzir essas falhas e adotar novas estratégias para planejar e desenvolver soluções [1], [2].

Esses sistemas computacionais embarcados são capazes de promover manutenções inteligentes, com o monitoramento contínuo de todo sistema, criando um log com todas as informações relevantes para o sistema de análise, assim permitindo maior precisão, rapidez e confiabilidade na detecção na origem das possíveis falhas as quais o sistema estaria mais suscetível [1], [2].

2. OBJETIVO

Através do estudo de um caso real, este trabalho utilizará um banco de dados gerados em análises de defeitos reais obtidos em laboratório e em campo, com estes dados em mãos serão aplicadas técnicas de aprendizado de máquinas, para que estes dados sejam tratados e analisados de forma a possibilitar o diagnóstico precoce de possíveis defeitos.

O grande problema aqui é a falha no sistema de comunicação, estamos tratando de sistemas de radiocomunicação nas faixas de UHF/VHF com alta potência [1].

Para identificar as possíveis falhas foram criadas diversas situações de defeitos, e foi montada uma giga de testes e acionamentos para obtenção dos valores lidos através do hardware desenvolvido para esta solução. Estes dados foram tratados, filtrados e cuidadosamente classificados a fim de que com as técnicas de aprendizado de máquinas fossem possíveis categorizar cada tipo de problema apresentado e pudesse ser identificado antes que ele venha a causar a paralização de todo sistema e até mesmo a queima de forma geral.

Possibilitando que sejam tomadas as devidas providências de forma muito mais específica e pontual, o sistema tem sua implementação de forma muito simples e prática, sendo bem robusto e de fácil configuração, todos os dados são enviados para um servidor on-line onde são lidos e processados [1].

3. FUNDAMENTAÇÃO

Ao considerar a história do rádio, deve-se lembrar que algumas das primeiras aplicações foram colocadas em navios de guerra. A empresa Marconi está envolvida em comunicações marítimas há quase um século desde o trabalho pioneiro de seu fundador, *Guglielmo Marconi*. Para atender às crescentes necessidades dos usuários, a tecnologia evoluiu ao longo de gerações de Sistemas de Comunicações Integradas, para unidades de envio e recepção operadas digitalmente. O uso de receptores sem fio e antenas de anel *Bellini-Tosi* na área do cabeçalho foi desenvolvido pela *Marconi Company* antes da guerra e em 1912, o primeiro teste da válvula de transmissão entre a aeronave e o HMS *Calliope*, um cruzador leve. 75 navios foram posteriormente equipados com coletores de válvulas. Durante a década de 1960, a MOD UK, em colaboração com a indústria do Reino Unido, iniciou um programa de trabalho intitulado “Sistemas de Comunicação Integrados para Navios”. Os problemas das antenas de transmissão foram superados com o uso de antenas de banda larga HF integradas aos elementos estruturais do navio, como mastros e funis [2]. VHF/SSB - O transceptor/rádio VHF (*Very High Frequency* - Frequência Muito Alta- faixa de radiofrequências de 30 a 300 MHz) é um equipamento de radiocomunicação básico e obrigatório de acordo com os critérios da NORMAM (Normas da Autoridade Marítima [DPC) com alcance limitado a potência de transmissão, as condições de propagação de sinal, ao ganho e altura da antena [3], [4].

A Fig.1 ilustra um exemplo padrão de um radiocomunicador encontrado em navios e embarcações [5].

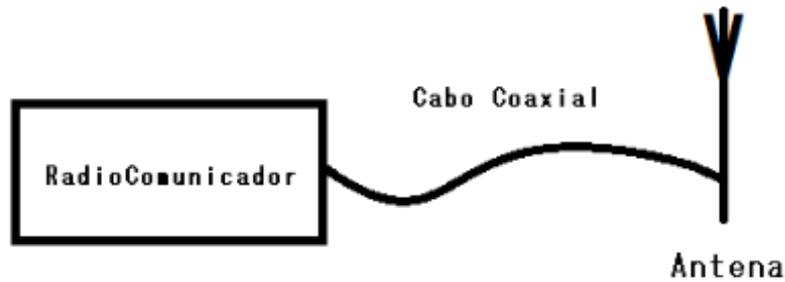
No esquema da Fig. 2 pode-se ver o esquema geral constituído pelo radiocomunicador, cabo coaxial e antena, e entre cada etapa existem as conexões, específicas para cada tipo de aplicação e que são de extrema importância para o perfeito funcionamento do sistema.

A forma para verificação do funcionamento do sistema de comunicação é através do medidor de ROE ou VSWR, sempre encontramos a sigla VSWR (*Voltage Standing Wave Ratio*), ou em português, ROE (Relação de Ondas Estacionárias). Basicamente, a ROE é uma forma de se medir o casamento de uma antena com sua fonte de sinal. Exemplo: quando toda a energia gerada é transferida para a antena (modelo ideal), obtemos um ROE de 1:1 (um para um). Ou seja, um casamento perfeito. Na prática isto não ocorre, pois sempre existirão perdas [6].

Figura 1 – Exemplo Radio SSB [5]



Figura 2 – Esquema Geral (Fonte: Autoria Própria)



A medição e análise de ROE (SWR) são utilizadas para instalar e reparar linhas de transmissão de radiofrequência em uma antena. Se o transmissor estiver conectado à antena por um cabo, a impedância da antena e as características do cabo devem fazer o que é conhecido como casamento de impedância. Isso garante que mais energia seja transferida do cabo para a antena.

Se a impedância da antena e do cabo não for compatível, parte da tensão não será transferida. A energia que não é transmitida será então refletida como uma onda de volta ao transmissor. A interação entre a onda refletida e a onda fornecida pelo transmissor resulta em formas de onda estacionárias. A potência refletida de volta causa três problemas durante a transmissão de radiofrequência: A perda de potência na frequência do transmissor aumenta, a distorção é criada no transmissor devido à potência refletida e pode danificar os circuitos da máquina[6].

Podemos ver ilustrado na Fig.3 um modelo de um Wattímetro/Medidor de ROE (SWR) HF / VHF, geralmente utilizado para realizar a aferição dos valores de potência transmitida e refletidas.

Figura 3 – Wattímetro/Medidor de ROE (SWR) HF/VHF [7]



4. METODOLOGIA

Inicialmente montou-se todo setup e verificado o perfeito funcionamento dos equipamentos e dispositivos de medição. A partir desse ponto foram efetuadas 500 leituras do sistema em perfeito funcionamento. Dessa forma obtivemos um gráfico com os valores normais de funcionamento.

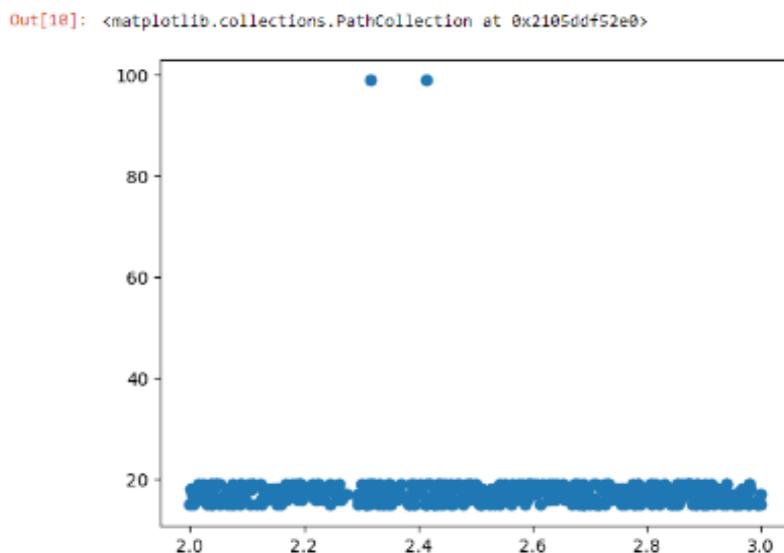
A interface utilizada foi Jupyter Notebook, utilizando a linguagem computacional Python. O ambiente Jupyter Notebook é integrado com o Anaconda Navigator. Para os

tratamentos, análises e modelagem foi usado a linguagem de computação Python. O ambiente de desenvolvimento será o Jupyter Notebook, no qual é integrado com o Anaconda Navigator, sendo uma plataforma de distribuição para linguagem Python e R, que é amplamente adotada, que funciona como um grande banco de vários repositórios de pacotes. Com a facilidade de ter todos de forma gráfica e não apenas por linha de comando.

A) Tratamento de Dados

O primeiro arquivo foi convertido em um *dataframe*. Os valores obtidos são referentes tensão que correspondem à potência em watts. Após uma análise dos dados de tensão, verificou-se que existe uma correlação percentual entre o valor de potência de rádio frequência transmitida e refletida e que esse valor em funcionamento normal está em torno de 80% de eficiência, ou seja, temos uma perda de sinal de até 20%. A Fig.4 ilustra uma plotagem dos dados analisados.

Figura 4— Dados lidos do sistema normal (Fonte: Autoria Própria)



Os dados são referentes a 500 leituras do valor de tensão que corresponde a taxa de eficiência do valor transmitido, entre as potências transmitidas e refletidas. Observa-se na Fig.4 que esses valores ficam abaixo dos 20%.

Existem dois pontos inseridos de forma manual para que seja possível a visualização da escala completa, fazendo uma comparação com os valores de defeitos essa percepção se torna mais perceptível.

Inicialmente foi implementado o modelo de árvore de decisão para que com isso fosse possível ter abrangência de outros modelos com a finalidade de aperfeiçoar o desempenho, como o método do *K Nearest Neighbor* (KNN).

B) Método de Aprendizagem

Os modelos preditivos têm propriedades diferentes. Alguns modelos podem lidar com dados complexos e outros têm dados lineares. Nesta seção, o método de classificação e outros algoritmos usados como classificador serão apresentados.

B.1) Classificadores

A classificação é um método de aprendizado supervisionado para rotular uma amostra com base em atributos. Aprendizado supervisionado significa que temos um rótulo de classificação ou número de regressões que o algoritmo aprenderá [8].

B.1.1) Árvore de Decisões

As árvores de decisão são uma maneira simples de acessar informações e são uma maneira eficiente de construir classificadores [9]. Pode-se usar o processo para construir uma Árvore de Decisão e ter uma série de consultas para prever a classe de destino. As vantagens deste modelo incluem suporte para dados não numéricos (em algumas implementações), pequena preparação de dados (sem necessidade de escala), suporte para gerenciamento de relacionamentos não lineares, importância de propriedades que são reveladas e fáceis de interpretar [8].

Uma árvore de decisão é um método de classificação que pode ser usado para apresentar os resultados de forma hierárquica. Em que, o atributo de entrada mais importante é apresentado na primeira posição, seguido pelos menos importantes. A principal vantagem do uso dessa abordagem é a tomada de decisão, permitindo que os usuários saibam quais fatores mais influenciam seu trabalho [9].

B.1.2) KNN

K Nearest Neighbor significa que *K* vizinho mais próximo é um algoritmo não paramétrico, cuja estrutura será determinada pela base de dados utilizada. O algoritmo possui uma variável chamada *K* que direciona o número de vizinhos.

Há um ponto, o algoritmo encontrará o ponto mais próximo dele, determinando sua classificação a priori. Para ver a que grupo pertence este ponto, é feita uma votação em que a maioria dirá a que parte o ponto pertence. Para isso, calcule a distância entre pontos adjacentes, em todas as classes possíveis, em função do número de pontos vizinhos, ou seja, do valor de *K*. Se o valor de *K* = 3, será contada a distância entre os 3 vizinhos mais próximos, método descritivo de classificação [10].

C) Métricas de Avaliação

No desenvolvimento de modelos de classificação, usando aprendizado de máquinas, deve-se existir a preocupação com a qualidade do classificador. As métricas são usadas para a escolha do modelo adequado para a resolução do problema analisado [11].

Para o auxílio do entendimento existem quatro possíveis classes, em classificações binárias [11]:

1. Verdadeiro positivo (VP): O método classificou como positivo, e a classe realmente era positiva.
2. Verdadeiro negativo (VN): O método classificou como negativo, e a classe realmente era negativa.
3. Falso positivo (FP): O método classificou como positivo, mas a classe era negativa.
4. Falso negativo (FN): O método classificou como negativo, mas a classe era positiva.

C.1) Matriz Confusão

Uma maneira simples de representar dados, esta apresentada na Fig. 5, onde há uma indicação do número de eventos que o sistema possui em quatro categorias [11].

Figura 5– Matriz Confusão Genérica [11]

Matriz de confusão		Classe predita	
		Positiva	Negativa
Classe original	Positiva	VP	FN
	Negativa	FP	VN

C.2) Acurácia

A acurácia é uma métrica considerada simples e importante. Ele irá verificar a porcentagem de acertos, calcular a relação entre o número de acertos e o número digitado [11].

$$Acurácia = \frac{\text{Total de acertos}}{\text{Total de Itens}} \qquad Acurácia = \frac{VP+VN}{VP+FN+VN+FP} \quad (1)$$

C.3) Sensibilidade

Também conhecida como revocação ou *recall*, essa métrica avalia a capacidade do método em obter resultados considerados positivos [11].

$$Sensibilidade = \frac{VP}{VP+FN} \quad (2)$$

C.4) Precisão

A precisão é uma métrica que avalia o número de verdadeiros positivos sobre a soma de todos os valores positivos [11].

$$precisão = \frac{VP}{VP+FP} \quad (3)$$

C.5) F-Score

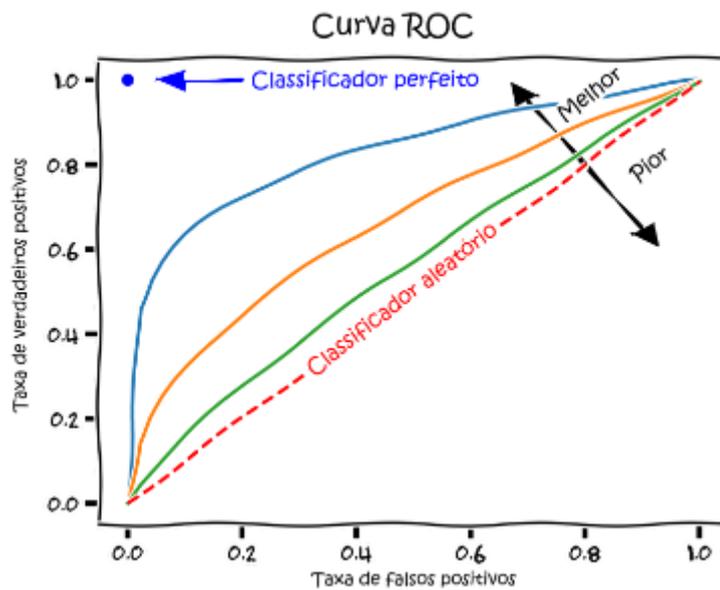
Uma métrica de uso de média harmônica calculada com base na precisão e revocação [11].

$$f1 = 2 * \frac{\text{precisão} * \text{revocação}}{\text{precisão} + \text{revocação}} \quad (4)$$

C.6) Curva ROC

A curva ROC (*Receive Operating Characteristic*) é um gráfico que permite testar um classificador binário. Exemplo de trabalho na Fig. 6 [11].

Figura 6– Exemplo de funcionamento curva ROC [11]



D) Aplicabilidade

A precisão pode ser usada onde os falsos positivos são considerados mais perigosos do que os falsos negativos. Por exemplo, se você classificar uma ação como um bom investimento, o modelo precisa ser preciso, mesmo que acabe classificando um bom investimento como ruim (Falso Negativo) nesse processo. Ou seja, o modelo deve ser preciso em sua classificação, pois consideramos um bom investimento quando na verdade não é, podendo-se perder grandes somas de dinheiro.

A retirada pode ser usada nos casos em que os falsos negativos são considerados mais perigosos do que os falsos positivos. Por exemplo, o modelo ainda deve detectar todos os pacientes doentes, mesmo que classifique algumas pessoas saudáveis como doentes (falso estado) no sistema. Ou seja, o modelo deve ser altamente recuperável, pois classificar um paciente doente como saudável pode ser um desastre.

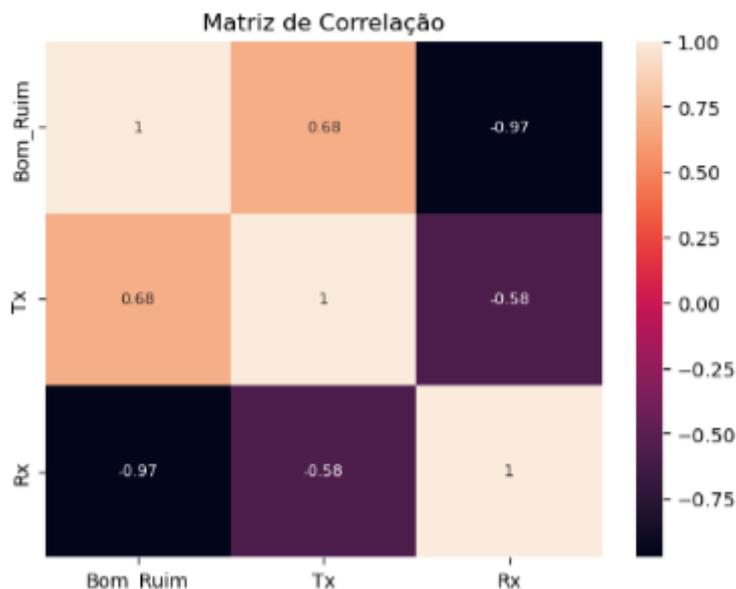
O F1-Score é simplesmente uma maneira de observar apenas uma métrica em vez de duas (precisão e recall) em uma determinada situação. É uma interpretação harmonizada dos dois valores, mais próxima dos valores menores do que uma simples interpretação aritmética. Ou seja, se você tiver um F1 Score baixo, isso é uma indicação de baixa precisão ou recuperação.

5. RESULTADOS

Neste item, serão apresentados os resultados dos modelos de aprendizado de máquina para classificação. Modelos com saída binária, onde uma regra condicional é utilizada para nomear a variável da saída de áudio. Quando a tensão está abaixo de 1 volts, é atribuído um "1", indicando uma operação normal, ou seja, um valor de potência refletida quase zero, caso contrário, é atribuído um "0", identificando assim quando existe um defeito, neste caso sem antena desconectada, gerando assim um valor de potência refletida muito elevada.

Após captura e tratamento dos dados foi construída uma matriz de correlação, com objetivo de verificar possíveis relações, apresentada na Fig.7.

Figura 7– Matriz de Correlação (Fonte: Autoria Própria)



Pode-se utilizar uma matriz de confusão com os valores encontrados para calcular as métricas analíticas (Fig.8).

Figura 8– Matriz Confusão Padrão (Fonte: Autoria Própria)

		Predicted	
		True	False
Sample	True	True Positive (TP)	False Negative (FN)
	False	False Positive (FP)	True Negative (TN)

True Positive: valores previstos pelo modelo como um bom ajuste e positivos para a amostra (previsão de acertos).

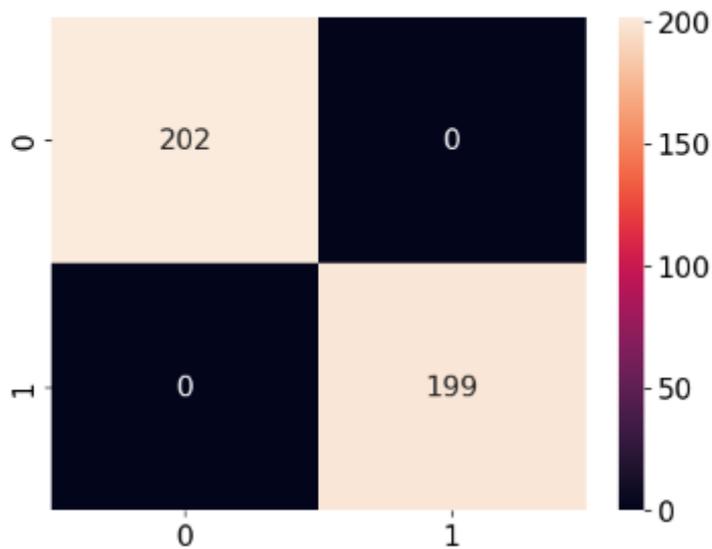
True Negative: os valores são previstos como negativos no modelo e negativos na amostra (previsão de acertos).

False Positive: os valores previstos são positivos no modelo e negativos na amostra (erro de previsão).

False Negative: os valores são previstos como negativos no modelo e positivos na amostra (erro de previsão).

A Fig. 9 é a matriz confusão referente ao modelo de Árvore de Decisão, esses valores foram usados para os cálculos das métricas de avaliação apresentados no item 4.

Figura 9– Matriz Confusão modelos Árvore de Decisão (Fonte: Autoria Própria)



A Fig.10 ilustra a matriz confusão referente ao modelo de KNN, esses valores foram usados para os cálculos das métricas de avaliação apresentados no item 4.

Para apresentar os resultados das métricas, os valores calculados foram dispostos na Tabela 1. Na Tabela 1 são apresentados os valores referentes aos 2 modelos utilizados com intuito de comparação e escolha do melhor modelo para o estudo, porém como o sistema é linearmente separável, qualquer método utilizado atinge 100% de eficácia.

Figura 10 – Matriz Confusão modelo KNN (Fonte: Autoria Própria)

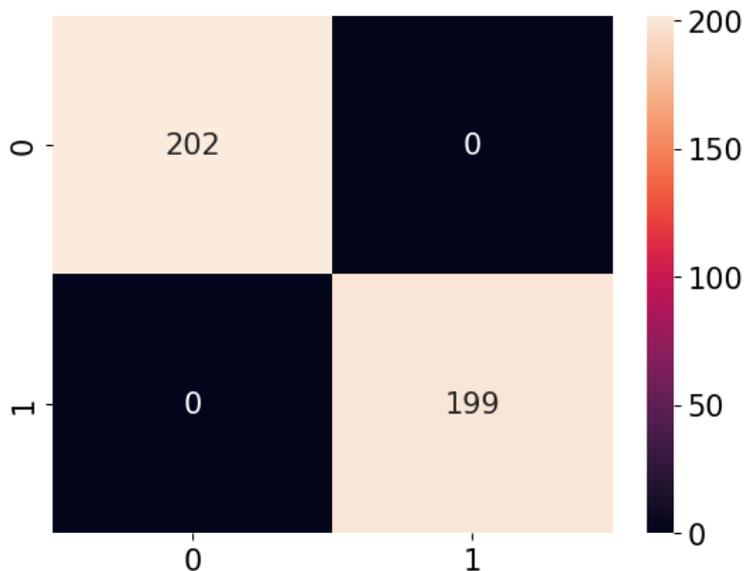


Tabela 1- Métricas de Avaliação – Comparação (Fonte: Autoria Própria)

	Árvore de Decisão	KNN
Acurácia	100%	100%
Sensibilidade	100%	100%
Precisão	100%	100%
F-score	100%	100%

6. CONCLUSÃO

Após a análise dos resultados através das métricas de avaliação dos modelos, foi constatado que o problema teve os resultados similares, devido aos valores medidos durante as simulações feitas em laboratório.

Todos os modelos conseguiram prever, de forma satisfatória, a partir da potência refletida em relação a potência transmitida. Sendo evidente a identificação do defeito. Com isso indicando a manutenção e/ou troca do dispositivo com defeito no sistema de comunicação.

Foram feitas duas classificações, uma sendo operação normal e outro sistema sem antena, o sistema está preparado para receber outros dados referentes a outras classes de defeitos. E outros sinais, como tensão de alimentação, corrente consumida pelo sistema, no entanto essas opções ainda não foram implementadas pelo hardware que foi desenvolvido, sendo essa uma proposta de melhoria do projeto proposto, a leitura desses dados, assim como tratamento dos dados e nova avaliação das métricas, para melhor eficácia nas previsões dos defeitos.

REFERÊNCIAS

[1] J. L. Bosa, “Sistema embarcado para manutenção inteligente de atuadores elétricos,” Trabalho de Conclusão de Curso, Universidade Estadual do Rio Grande do Sul, 2009.

[2] R.J. Eassom, R.L.J. Awcock and B.M. Sosin, “HF transmitters and receivers for naval”, Proceedings of the 1995 International Conference on 100 Years of Radio, pp.62-68, London, UK, 1995. doi: 10.1049/cp:19950791.

[3] T.Thomas, Tudo que você precisa saber sobre a diferença entre VHF e UHF, 2019. Disponível em: <https://firstsourcewireless.com/pt-br/blogs/blog/all-you-need-to-know-about-the-difference-between-vhf-and-uhf> > Acesso em 4 de novembro de 2022.

[4] Diretoria de Portos e Costas, NORMAN – NORMAS DE AUTORIDADE MARITIMA, 2022.

[5] ICOM, “IC-M330GE Rádio Fixo de VHF Marítimo c/ DSC Classe D e GPS”, 2002. Disponível em: <https://www.nautiradar.pt/pt/produtos/icom/ic-m330ge-radio-fixo-de-vhf-maritimo-c-dsc-classe-d-e-gps> > Acesso em 4 de novembro de 2022.

[6] Relação de Ondas Estacionárias, 2019. Disponível em: https://www.wikifox.org/pt/wiki/Rela%C3%A7%C3%A3o_de_ondas_estacion%C3%A1rias > Acesso em 4 de novembro de 2022.

[7] Radiohaus, CN-501H2 - Wattímetro/Medidor de ROE (SWR) HF / VHF Daiwa , Disponível em: <https://www.radiohaus.com.br/produto/5504/cn-501h2-wattimetro-medidor-de-roe-swr-hf-vhf-daiwa?nomodal> > Acesso em 4 de novembro de 2022.

[8] M. Harrison, *Machine Learning - Guia de Referência Rápida*, São Paulo, 2020.

[9] P. G. Crepaldi, R. N. P. Avila, J. P. M. D. O. Paulo, R. Rodrigues and R. L. Martins, “Um Estudo Sobre a Árvore de Decisão e sua Importância na Habilidade de Aprendizado,” *Rev. Saber*, vol. 6, pp. 1–14, 2008.

[10] F. Luz, “Algoritmo KNN para Classificação”, 2019. Disponível em: <https://inferir.com.br/artigos/algoritmo-knn-para-classificacao/> > Acesso em 05 de julho de 2023.

[11] D. Mariano, “Métricas de avaliação em Machine Learning: acurácia, sensibilidade, precisão, especificidade e F-score”, In: book: *BIOINFO - Revista Brasileira de Bioinformática e Biologia Computacional*, 2021. doi: 10.51780/978-6-599-275326-15.