UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
PROJETO DE GRADUAÇÃO

FERNANDA GONÇALVES DA SILVA SOUZA

APLICAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA NA DETECÇÃO DE ANOMALIAS DO ESPECTRO DE RADIODIFUSÃO FM

VITÓRIA
2020

FERNANDA GONÇALVES DA SILVA SOUZA

APLICAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA NA DETECÇÃO DE ANOMALIAS DO ESPECTRO DE RADIODIFUSÃO FM

Parte manuscrita do Projeto de Graduação da aluna **Fernanda Gonçalves da Silva Souza**, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheira Eletricista.

Orientador: Prof. Dr. Patrick Marques Ciarelli

VITÓRIA
2020

FERNANDA GONÇALVES DA SILVA SOUZA

APLICAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA NA DETECÇÃO DE ANOMALIAS DO ESPECTRO DE RADIODIFUSÃO FM

Parte manuscrita do Projeto de Graduação da aluna **Fernanda Gonçalves da Silva Souza**, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheira Eletricista.

Aprovada em 11 de dezembro de 2020.

COMISSÃO EXAMINADORA:

_____

Prof. Dr. Patrick Marques Ciarelli
Universidade Federal do Espírito Santo
Orientador

_____

Prof. Dr. Jair Adriano Lima Silva
Universidade Federal do Espírito Santo
Examinador

_____

Prof. Dr. Evandro Ottoni Teatini Salles
Universidade Federal do Espírito Santo
Examinador

RESUMO

Dado que a Agência Nacional de Telecomunicações (ANATEL) tem, dentre suas atribuições, fiscalizar o espectro de frequências licenciadas, este trabalho mostra a possibilidade de automatizar o processo de fiscalização do espectro de radiodifusão FM da cidade de Vitória-ES a fim de otimizar o serviço, tornando-o mais rápido e eficaz. Para tal feito, amostras de frequência entre 87,4 a 108 MHz, contendo suas respectivas potências, foram obtidas por meio da leitura de um aparelho chamado *RFeye*, o qual pertence à citada Agência. Os dados obtidos da ANATEL passaram por um pré-processamento de sinal, e duas técnicas de aprendizado de máquina, chamadas k-vizinhos mais próximos (kNN, do inglês *k nearest neighborg*) e máquina de vetores de suporte (SVM, do inglês *support vector machine*), foram avaliadas para a identificação de portadoras consideradas “muito ruidosas”. Para validação das técnicas, foram utilizadas métricas para verificação do sucesso de cada algoritmo e análise de tempo de processamento. Os resultados experimentais indicaram que o algoritmo SVM com *kernel* polinomial é o mais apropriado para a tarefa. Portanto, uma *interface* foi criada usando este algoritmo, não só possibilitando o usuário fazer a análise automática de portadoras muito ruidosas, como também automatizando o processo de identificação de rádios licenciadas e detecção da variação de potência de cada portadora.

Palavras-chave: ANATEL. FM. Aprendizado de Máquinas. kNN. SVM.

ABSTRACT

Given that the National Telecommunications Agency (ANATEL) has among its duties to supervise the licensed frequency spectrum, this work shows the possibility of automating the process of supervising the FM broadcasting spectrum of Vitoria-ES in order to optimize the service, making it faster and more effective. For this purpose, frequency samples between 87.4 a 108 MHz , containing their respective powers, were obtained by reading a device called RFeye, which belongs to the Agency. The data obtained from ANATEL went through a signal pre-processing, and two Machine Learning techniques, called k nearest neighbors (kNN) and support vector machine (SVM), were evaluated for the identification of carriers considered "very noisy". For validation of the techniques, metrics were used to verify the success of each algorithm and processing time analysis. The experimental results indicated that the SVM algorithm with polynomial kernel is the most appropriate for the task. Therefore, an interface was created using this algorithm, not only enabling the user to perform automatic analysis of very noisy carriers, but also automating the process of identifying licensed radios and detecting the power variation of each carrier.

Keywords: ANATEL. FM. Machine Learning. kNN. SVM.

LISTA DE FIGURAS

Figura 1 – Largura de banda e frequência central de uma portadora.....19
Figura 2 – Estação RFeye 20
Figura 3 – Hierarquia de aprendizado 24
Figura 4 – Definição de classes kNN.....25
Figura 5 – Hiperplano alocado para a divisão de classes.....26
Figura 6 – Dados linearmente separados x dados não-linearmente separados.....27
Figura 7 – Mapeamento do espaço de entrada via função *kernel* 28
Figura 8 – Faixa de espectro que recebe o rótulo de normal 30
Figura 9 – Faixa de espectro que recebe o rótulo de anormal 30
Figura 10 – Matriz de confusão 31
Figura 11 – Limiar do filtro de mediana 34
Figura 12 – Picos detectados 34
Figura 13 – Limites para identificação de uma portadora.....35
Figura 14 – Comparação de potência de uma mesma portadora.....36
Figura 15 – Portadoras em que um dos limites divergem com o piso de ruído 37
Figura 16 – Portadora com classificação normal e seus respectivos limiares.....38
Figura 17 – Portadora com classificação anormal e seus respectivos limiares.....38
Figura 18 – Fluxograma para captura anomalias 39
Figura 19 – Funcionamento do *k-cross validation* para o problema 42
Figura 20 – Métricas referentes ao kNN em relação à quantidade de limiares 43
Figura 21 – Métricas referentes ao SVM, *kernel* linear, em relação à quantidade de limiares 44
Figura 22 –Métricas referentes ao SVM, *kernel* polinomial, em relação à quantidade de limiares 45
Figura 23 – Métricas referentes ao SVM, *kernel* RBF, em relação à quantidade de limiares 46
Figura 24 – Métricas referentes ao SVM, *kernel* sigmodal, em relação à quantidade de limiares 47
Figura 25 – *Interface* gráfica do *software* 50
Figura 26 – Sequência de botões.....51
Figura 27 – Aba aberta a partir do botão selecionar arquivos 52
Figura 28 – Exibição da *interface* após abrir um arquivo.....52

Figura 29 – Exibição do botão "Pré-visualização" do arquivo que foi selecionado ..54
Figura 30 – *Interface* após ativar o botão iniciar análise.....55
Figura 31 – Parte do CSV gerado.....55
Figura 32 – Janela referente aos dados ruidosos.....56
Figura 33 – Exibição do comando Ação de uma portadora ruidosa 57
Figura 34 – Exibição do comando Ação de uma portadora com variação de potência acima do permitido 58
Figura 35 – Janela referente às portadoras licenciadas e não licenciadas.....59
Figura 36 – Exibição do comando Ação de uma portadora identificada.....59

LISTA DE QUADROS

Quadro 1 – Funções de *Kernel* mais comuns.....28
Quadro 2 – Argumento *k* do kNN em relação ao número de limiares.....42
Quadro 3 – Tempos de construção, treinamento e classificação do kNN de acordo com os limiares.....44
Quadro 4 – Argumento d do SVM- polinomial em relação ao número de limiares ..45
Quadro 5 – Argumento σ do SVM- RBF em relação ao número de limiares.....46
Quadro 6 – Tempos de construção, treinamento e classificação do *kernel*: RBF de acordo com os limiares.....48
Quadro 7 – Tempos de construção, treinamento e classificação do *kernel*: polinomial de acordo com os limiares.....48

LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizado de Máquina
ANATEL	Agência Nacional de Telecomunicações
ES	Espírito Santo
FM	Frequência Modulada
IA	Inteligência Artificial
LGT	Lei Geral das Telecomunicações
MOSAICO	Sistema Integrado de Gestão e Controle do Espectro
kNN	*k Nearest Neighbors*
PDF	*Portable Document Format*
PMEC	Plano de Monitoração do Espectro utilizado por Aplicações Críticas
RBF	*Radial Basis Function*
RFeye	Sistema de Monitoramento de Frequência
ROC	*Receiver Operating Characteristic*
SVM	*Support Vector Machine*
TAE	Teoria do Aprendizado Estatístico
UO021	Unidade Operacional da ANATEL do Espírito Santo

LISTA DE SÍMBOLOS

$fi(t)$	Frequência instantânea (Hz)
f_0	Frequência da portadora não modulada (Hz)
K_F	Constante de modulação
$x(t)$	Sinal do modulador
A_m	Amplitude da portadora (V)
f_m	Frequência do sinal (Hz)
Δf	Desvio de frequência (Hz)
β	Índice de modulação angular
M	Número de raias significativas
fa	Frequência de amostragem (Hz)
k	Quantidade de vizinhos referente ao kNN
$\emptyset$	Representação das funções *kernel*
d	Grau da função polinomial
σ	Desvio padrão da RBF

SUMÁRIO

1 INTRODUÇÃO 12
1.1 Justificativa 13
1.2 Objetivos 14
1.3 Organização do Trabalho 15

2 FUNDAMENTAÇÃO TEÓRICA 16
2.1 Radiodifusão Sonora em Frequência Modulada 16
2.1.1 Conceitos da Modulação de Frequência 16
2.1.2 Taxa de Amostragem 20
2.1.3 Radiodifusão Sonora e os Critérios da ANATEL 21
2.2 Inteligência Artificial e Aprendizado de Máquina 23
2.2.1 kNN 24
2.2.2 SVM 26
2.3 Pré-processamento dos Dados 29
2.4 Métricas de Desempenho 30

3 METODOLOGIA E EXECUÇÃO DO PROJETO 33
3.1 Recurso Computacional Utilizado 33
3.2 Etapas de Desenvolvimento 33
3.2.1 Verificação de Rádio Não Licenciada 35
3.2.2 Verificação da Variação de Potência 36
3.2.3 Verificação de Portadoras com Ruídos Anormais 36
3.3 Fluxograma do Método Proposto 39

4 EXPERIMENTOS E RESULTADOS 41
4.1 kNN 41
4.2 SVM 44
4.3 *Interface* 49
4.3.1 Botão: Selecionar arquivo 51
4.3.2 Botão: Pré-visualização 53
4.3.3 Botão: Iniciar Análise 54
4.3.4 Botão: Portadoras Ruidosas 56
4.3.5 Botão: Portadoras Com Variação de Potência Maior ou Igual a 10% 57

4.3.6 Botão: Licenciada e Não Licenciada 58

5 CONCLUSÕES 60

REFERÊNCIAS BIBLIOGRÁFICAS 62

APÊNDICE A – LISTA DAS RÁDIOS DA GRANDE VITÓRIA.....66

1 INTRODUÇÃO

Na última década do século XX foi aprovado, no Brasil, o fim do monopólio estatal da operação de serviços de telecomunicações. Neste contexto, a Agência Nacional de Telecomunicações (ANATEL) foi criada pela Lei 9.472 de 16 de julho de 1997, conhecida por Lei Geral das Telecomunicações (LGT), a fim de regularizar tais serviços.

Uma das atribuições da Agência é fiscalizar as faixas de espectro de frequência licenciadas. Dentre essas, pode-se destacar a frequência modulada (FM), como fica claro na Resolução nº 67, de 12 de novembro de 1998, a qual se refere ao Regulamento Técnico para Emissoras de Radiodifusão Sonora em Frequência Modulada. De acordo com a Consulta Pública nº 6, de 08 de março de 2018, tem-se que

> As funções de monitoramento e fiscalização do espectro são atribuições típicas de Estado e visam a garantir a convivência harmônica e a correta utilização do recurso por seus usuários, combater o uso irregular e não autorizado do espectro, atuar na investigação de casos de interferências e prover suporte às atividades de planejamento do espectro. (AGÊNCIA NACIONAL DE TELECOMUNICAÇÕES, 2018a, p. 5).

Porém, até 2017, a fiscalização do espectro ocorria apenas a partir de denúncia. Em 2018 foi iniciado o Plano de Monitoração do Espectro utilizado por Aplicações Críticas (PMEC), visando a gestão do espectro de modo preventivo, no que diz respeito às radiofrequências consideradas críticas, abrangendo também a faixa de radiodifusão sonora em FM.

A gestão espectral se dá através de análise de dados. Atualmente, tal verificação é feita de forma visual, na qual o olho humano identifica possíveis problemas em portadoras identificadas. Neste panorama, a inteligência artificial (IA) pode se mostrar como uma solução para melhorar a precisão da análise de dados, além de ser um artifício capaz de otimizar o tempo de serviço de tal demanda.

“Técnicas de Inteligência Artificial, em particular de Aprendizado de Máquinas, têm sido utilizadas com sucesso em grande número de problemas reais” (FACELI et al., 2017). O aprendizado de máquina (AM) é uma área em desenvolvimento da ciência

da computação. Ela é focada em criar modelos analíticos automatizados capazes de realizar previsões a partir de banco de dados disponíveis (ANY CONSULTING, 2018).

Assim, o projeto aqui apresentado caracteriza-se como do tipo aplicado, tratando-se de um estudo explicativo, no qual se utiliza o método experimental para possibilitar a manipulação e o controle de variáveis. No caso deste projeto, as amostras espectrais serão manipuladas e as técnicas de aprendizado de máquina serão aplicadas. Como alguns algoritmos serão testados, a pesquisa é classificada como quantitativa, objetivando obter resultados satisfatórios e evitar contradições no processo de análise. Sob ponto de vista dos procedimentos técnicos a serem seguidos, tem-se que o projeto é uma pesquisa experimental.

1.1 Justificativa

De acordo com o Relatório Anual, disponibilizado para a comunidade, existem algumas realizações da ANATEL.

> No exercício de 2018, a Agência executou o primeiro piloto do Plano de Monitoração do Espectro utilizado por aplicações Críticas (PMEC). O plano teve como objetivo monitorar periodicamente as principais faixas de frequências utilizadas para comunicações de socorro e segurança, controle de tráfego aéreo e marítimo e faixas de uso proibido, conforme regulamentação e referências internacionais aplicáveis, e outras faixas de interesse, como as destinadas ao serviço de radiodifusão. O objetivo do Plano é identificar preventivamente possíveis interferências prejudiciais, bem como estações que estejam fazendo uso irregular ou não autorizado do espectro, com potencial de causar interferências prejudiciais em tais faixas (AGÊNCIA NACIONAL DE TELECOMUNICAÇÕES, 2018c, p. 61).

Com um plano de medida preventiva designado, ocorre uma coleta e análise de dados por parte dos agentes de fiscalização da ANATEL. A análise desses dados é feita de maneira manual nas Agências do Brasil, na qual o fiscal, após coletar os dados e processá-los em um programa para obter a imagem do espectro, tem que visualmente identificar se tem alguma portadora com comportamento anormal, o que torna uma tarefa que demanda muito tempo e por muitas vezes imprecisa. Este fato foi relatado pelos fiscais da ANATEL e constatado pela autora deste projeto, durante um estágio

realizado na unidade operacional da ANATEL no estado do Espírito Santo, também definida como UO021.

Dada a sequência de fatos e visto que a UO021, localizada em Vitória-ES, possui um histórico com dados coletados da faixa de frequência de FM, tendo comportamentos normais e anormais do espectro, este trabalho pretende aplicar técnicas de aprendizado de máquina como uma possibilidade de solução de detecção de anomalias na faixa de espectro definida. Isto é justificável, pois os comportamentos anormais, apesar de semelhantes, não são iguais, o que dificulta a aplicação de critérios simples para a detecção. Além disso, técnicas de aprendizado de máquina têm a habilidade de encontrar padrões nos dados, que nem sempre são fáceis de serem identificados por humanos, e conseguir generalizar o aprendizado, o que permite realizar tarefas de forma assertiva.

1.2 Objetivos

O objetivo geral do projeto é avaliar a aplicação de técnicas de aprendizado de máquina na detecção de anomalias das portadoras de emissoras de Radiodifusão Sonora, em FM, na faixa de frequência de 87,4 a 108 MHz da cidade de Vitória-ES.

Dentre os objetivos específicos, têm-se:

- Identificar na literatura técnicas de aprendizado de máquina úteis para a tarefa a ser desenvolvida;
- Averiguar e definir técnicas de pré-processamento de sinais a fim de otimizar os resultados esperados para o problema proposto;
- Definir e implementar em uma *interface* a técnica que tenha um desempenho satisfatório diante do contexto explorado;
- Desenvolver uma metodologia capaz de gerar dados de fácil interpretação para o usuário final.

1.3 Organização do Trabalho

Este trabalho possui a seguinte organização. Na seção 2 é apresentada a fundamentação teórica. Na seção 3 é descrita a metodologia desenvolvida. Os experimentos e resultados são detalhados na seção 4 e, por fim, as conclusões são delineadas na seção 5.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Radiodifusão Sonora em Frequência Modulada

A radiodifusão é a transmissão de ondas de radiofrequência moduladas que se propagam através do espaço (ROMAIS, 1994). Este tópico tem por objetivo abordar alguns conceitos da modulação FM, a fim de entender os dados coletados e tratados no projeto e mostrar como tal faixa é fiscalizada no Brasil.

2.1.1 Conceitos da Modulação de Frequência

Conhecida pela sigla FM (do inglês *frequency modulation*), a modulação de frequência é o sistema de modulação angular no qual a frequência instantânea é diretamente proporcional ao sinal modulador (CARVALHO, 2008, p. 128), conforme a equação (1).

$$fi(t) = f_0 + K_F.x(t) \tag{1}$$

Na qual f_0 é a frequência da portadora não modulada, K_F a constante de modulação de frequência e $x(t)$ é a parcela referente ao sinal modulador. Para as análises, Carvalho (2008) considera o sinal modulador como senoidal. Isto pode ser visto na equação (2), onde A_m é a amplitude da portadora e f_m a frequência do sinal, ambos em função do tempo (t).

$$x(t) = A_m(t).cos(2\pi.f_m(t)) \tag{2}$$

Logo, substituindo (2) em (1), encontra-se (3).

$$fi(t) = f_0 + K_F.A_m(t).\cos(2\pi.f_m(t)) \tag{3}$$

Como $x(t)$ é uma função de cosseno, logo, os valores máximos e mínimos da equação (2) são A_m e $-A_m$, respectivamente. Assim, tem-se que o desvio de frequência (Δf), neste caso, em relação à frequência não modulada f_0 é dado conforme a equação (4).

$$\Delta f = K_F . A_m \tag{4}$$

Com o desvio de frequência pode-se definir o índice de modulação angular (β) para FM, conforme exposto em Carvalho (2008), e apresentado na equação (5).

$$\beta = \frac{\Delta f}{f_m} \tag{5}$$

A compreensão do índice β, dado em radianos, serve para entender os tipos de faixas em FM. Carvalho (2008, p. 130) afirma que “o sinal modulado é uma função periódica no tempo, seu espectro é discreto e depende do índice de modulação angular para se distinguir em duas situações”, sendo esse sistema de faixa estreita (para β muito menor que 1 radiano) ou de faixa larga (para β maior que 1 radiano). Ainda sobre a divisão dos sistemas tem-se que

> A largura de faixa adequada para utilização no sistema de FM depende de sua aplicação. Com um grande desvio, o ruído será melhor suprimido como outras interferências, mas cuidados devem ser tomados para assegurar que o impulso de ruído em seus picos não torne excessivos. (RODRIGUES, 2004, p. 41)

No caso deste projeto de graduação, o objeto de estudo será um sistema de faixa larga. Para entender melhor o cálculo da largura de banda de frequência de cada faixa, tem-se que

> Como o sinal modulado em ângulo por sinal modulador senoidal de frequência f_m é um sinal periódico com frequência de repetição f_m, seu espectro é discreto, com componentes espectrais regularmente espaçadas de f_m. Há uma raia central (frequência f_o) e uma infinidade de raias laterais inferiores (frequências f_o - $n.f_m$) e superiores (frequências f_o + $n.f_m$), com n inteiro positivo, de 1 a + ∞. (CARVALHO, 2008, p. 132)

Assim, para um sinal modulador senoidal, o espectro do sinal modulado em ângulo tem um número teoricamente infinito de raias laterais. Porém, Carvalho (2008) demostra que as raias acima de certa ordem podem ser desprezadas por apresentarem magnitudes muito pequenas. O espectro significativo do sinal modulado é aquele que se estende de f_o - $M.f_m$ até f_o + $M.f_m$, sendo M o número de raias significativas.

Para a determinação de *M*, Carlson (1986, apud CARVALHO, 2008, p. 327) cita que o Critério de Carson é o mais utilizado no caso senoidal discreto, como indicado na equação (6).

$$M = \beta + 1 \tag{6}$$

Em que *M* sempre será o número inteiro mais próximo do resultado da equação (6). Assim, a largura de banda (B) será dada por (7).

$$B = 2.M.f_m \tag{7}$$

Porém, o Critério de Carson pode ser estendido para os sinais não senoidais, como é o caso da transmissão de rádio. Com isso, a largura de banda passa a ser dada pela equação (8).

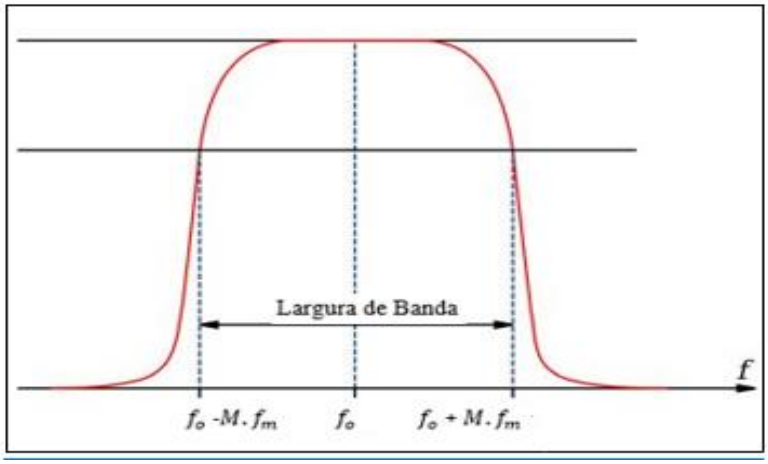
$$B = 2.(\beta + 1).f_m = 2.(\beta.f_m + f_m) \tag{8}$$

Substituindo a equação (5) em (8), chega-se na equação (9), que se refere a largura de banda.

$$B = 2(\Delta f + fm) \tag{9}$$

A Figura 1 ilustra alguns dos parâmetros abordados.

Figura 1 – Largura de banda e frequência central de uma portadora



Fonte: TESTAND & MEASUREMENT TIPS (2020).
Nota: Adaptado pela autora.

É importante salientar que o comportamento do sinal modulado ao ser transmitido pode sofrer algumas variações, na qual a envoltória deixa de ter um formato contínuo e passa conter ruídos. Han e Kamber (2000, apud FACELI et al., 2017, p. 34) afirmam que “ruído pode ser definido como uma variância ou erro aleatório no valor gerado ou medido para um atributo”.

Haykin (2011), explica que existem várias fontes potenciais de ruído em um sistema, sendo elas internas ou externas. As fontes internas incluem os ruídos que surgem devido às flutuações espontâneas de corrente ou tensão em circuitos elétricos. Já as fontes externas seriam referentes a ruídos atmosféricos, galácticos, entre outros.

“Em muitos casos, o ruído é produzido pelos próprios equipamentos ativos utilizados para implementar os sistemas de transmissão, tais como os amplificadores” UNIVERSIDADE DO MINHO (1999). Isto inclui não só os componentes utilizados nos repetidores e receptores, mas também a falta de técnicas de filtragem nos transmissores que podem ser grandes causadores de ruído. Logo, fica claro que alguns componentes da transmissão são os principais responsáveis por tais

acontecimentos. Na prática a autora do projeto ainda pôde identificar que uma rádio ao ter uma de suas antenas queimadas, propiciou um sinal ruidoso também.

Nesse contexto, a ANATEL é responsável por analisar os ruídos provenientes dos equipamentos da transmissão, não ruídos de fonte externas.

2.1.2 Taxa de Amostragem

Os dados analisados em tal projeto são amostras discretizadas no tempo. O aparelho responsável por coletar tais amostras é o *RFeye*, que é um sistema de monitoramento de espectro que opera na faixa de frequência de 10 MHz a 6 GHz, ilustrado na Figura 2.

Figura 2 – Estação RFeye

Fonte: AGC Telecom (2018).

O aparelho é programável, sendo possível ao operador inserir a faixa de frequência que quer monitorar, a quantidade de leitura por minuto que deseja captar e a taxa de amostragem que usará. Neste cenário, é relevante entender como na teoria é definida a taxa de amostragem.

Pelo Teorema da Amostragem de Nyquist/Shannon, mostrado em Carvalho (2008), tem-se que um sinal contínuo pode ser apropriadamente amostrado somente se ele não contiver componentes em frequência acima da metade da frequência de

amostragem. Logo, é possível verificar na equação (10), a frequência de amostragem fa deve ser maior que duas vezes da maior frequência do sinal amostrado.

$$fa > 2f_m \tag{10}$$

Assim. o Teorema de Nyquist prevê que que o sinal seja limitado em banda ou eu se utilize um filtro passa-baixas que limite a banda do sinal de entrada. Caso a frequência de amostragem não seja suficientemente alta, ocorre a sobreposição do espectro (*aliasing*), impossibilitando recuperar o sinal original e criando artefatos (OPPENHEIM, 1999).

2.1.3 Radiodifusão Sonora e os Critérios da ANATEL

No Brasil, a ANATEL é o órgão responsável por disciplinar as utilizações das faixas de frequência, deliberando resoluções de modo a padronizar o uso do espectro. A Resolução nº 596, de 6 de agosto de 2012, define ação de fiscalização como

> Atividade de acesso, obtenção e averiguação de dados e informações, por meio de procedimentos e técnicas aplicados por Agente de Fiscalização com a finalidade de reunir evidências para a apuração do cumprimento de obrigações e conformidades por parte da fiscalizada e verificar a forma de execução dos serviços de telecomunicações (AGÊNCIA NACIONAL DE TELECOMUNICAÇÕES, 2012)

Neste sentido, cabe ao agente de fiscalização zelar também pelo espectro de radiodifusão sonora em frequência modulada. A Resolução nº 67, de 12 de novembro de 1998, define que

> A faixa de radiodifusão sonora em frequência modulada estende-se de 87,4 a 108 MHz, e é dividida em 103 canais, cujas portadoras estão separadas de 200 kHz. Cada canal é identificado por sua frequência central, que é a frequência da portadora da estação de FM. [...] Percentagem de Modulação - É a relação entre o desvio de frequência e o desvio de frequência definido como 100% de modulação, expresso em percentagem. Para estações de radiodifusão sonora em frequência modulada, um desvio de frequência de ±75 kHz é definido como 100% de modulação. (AGÊNCIA NACIONAL DE TELECOMUNICAÇÕES, 1998).

Dessa forma, há um espaçamento de 50 kHz entre uma portadora e outra, fazendo com que não haja caso de intermodulação. Ainda sobre FM, o item 3.2.3 da Resolução

nº 67 da ANATEL aborda sobre tolerância de frequência, permitindo que a frequência central da emissão de uma emissora varie entre mais ou menos 2000 Hz de seu valor nominal.

Contudo, tais averiguações só podem ser feitas com o sinal não modulado e isolando o sistema do meio externo, já que as causas de ruído podem ser tanto internas quanto externas. Ou seja, não se pode notificar uma rádio sem fazer uma vistoria técnica nela.

A autora do trabalho constatou no estágio realizado na UO021 que uma prática feita pela Agência para averiguar tais parâmetros é identificar visualmente uma portadora muito ruidosa no espectro de frequência analisado, classificando o dado como anormal. Quando feita a verificação, o fiscal realiza uma vistoria técnica na estação da rádio referida, com foco nos parâmetros técnicos do transmissor, além de verificar outros possíveis componentes da estação responsável por tais anomalias.

Em tal vistoria, o fiscal operante desliga o sinal de modulação conectando o transmissor numa carga resistiva. Dado o processo, pode-se explorar o item 3.2.7 da Resolução nº 67 da ANATEL, o qual indica que

> Qualquer emissão presente em freqüências afastadas de 120 a 240 kHz, inclusive, da freqüência da portadora, deverá estar pelo menos 25 dB abaixo do nível da portadora sem modulação; as emissões em freqüências afastadas da freqüência da portadora de mais de 240 kHz até 600 kHz, inclusive, deverão estar pelo menos 35 dB abaixo do nível da portadora sem modulação. As emissões em freqüências afastadas de mais de 600 kHz da freqüência da portadora deverão estar (73+P) dB (P = potência de operação do transmissor, em dBk) abaixo do nível da portadora sem modulação, sendo 80 dB a maior atenuação exigida. (AGÊNCIA NACIONAL DE TELECOMUNICAÇÕES, 1998).

Em suma, analisar somente os dados gerados pela *RFeye* não são suficientes para atuar com a fiscalização. Mas tais dados servem de indícios de irregularidades, sendo de suma importância para fiscalizar uma rádio. Um desses indícios por parte da UO021 é a verificação de portadoras muito ruidosas.

Por último, tem-se uma verificação que pode ser feita sem ter que ir até à estação da transmissão. Ela está explicita no item 6.4 da Resolução nº 67 da ANATEL, o qual trata da potência de operação do transmissor. Neste, é deixado claro que tal potência

deve ser sempre mantida o mais próximo possível da autorizada e as eventuais variações devem ser restritas aos limites de mais ou menos 10% em condições normais.

2.2 Inteligência Artificial e Aprendizado de Máquina

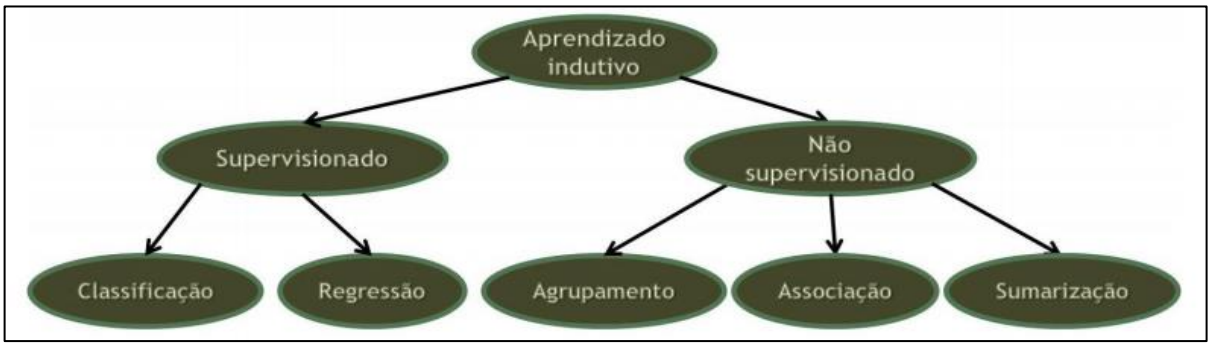
O termo Inteligência Artificial surgiu na década de 50, sendo um ramo da computação que tem por objetivo possibilitar que as máquinas simulem uma inteligência análoga a do ser humano.

Nas últimas décadas, com um crescente número de dados de diversos setores, tornou-se visível a necessidade da criação de ferramentas que automatizassem processos reduzindo a necessidade da intervenção humana. Para isso, essas ferramentas deveriam ser capazes de criar por conta própria uma hipótese ou uma função, a partir de experiências passadas, para resolver um problema. Ao processo de indução de hipóteses a partir de experiências dá-se o nome de aprendizado de máquinas (FACELI et al., 2017).

O AM é um ramo da inteligência artificial focado no raciocínio indutivo, que tem por objetivo desenvolver algoritmos que permitam o computador aprender uma determinada tarefa (BAIA, 2016).

As tarefas de aprendizado podem ser divididas em preditivas ou descritivas. A análise preditiva segue a linha de aprendizado supervisionado, já que tem a presença de um agente externo para indicar quais são os atributos de entrada e os de saída (rótulos). Já a descritiva busca o entendimento dos dados, sendo assim não supervisionado. A Figura 3 apresenta uma hierarquia de aprendizado de acordo com os tipos de tarefas (FACELI et al., 2017).

Figura 3 – Hierarquia de aprendizado



Fonte: Faceli, Lorena, Gama e Carvalho (2017).

Neste trabalho foram avaliadas duas técnicas de AM: o algoritmo dos vizinhos mais próximos (kNN, do inglês *k nearest neighborg*) e a máquina de vetores suporte (SVM, do inglês *support vector machine*). A primeira citada, kNN, foi testada não só pela sua simplicidade (FACELI et al., 2017), mas também porque obtém bons resultados em vários problemas de classificação, sendo muitas vezes usada como um classificador de referência. Já o SVM consegue lidar melhor com dados ruidosos (FACELI et al., 2017).

2.2.1 kNN

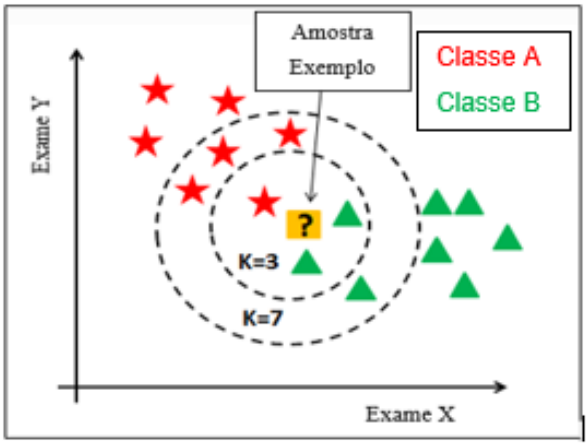
“O algoritmo dos vizinhos mais próximos, ou kNN, classifica um novo objeto com base nos exemplos do conjunto de treinamento que são próximos a ele” (FACELI et al., 2017, p. 59). Em tal algoritmo, o objeto a ser classificado é representado por um ponto no espaço definido pelos seus atributos.

Na fase de treinamento, o algoritmo memoriza os exemplos rotulados, ou seja, armazena as classes e os objetos que foram fornecidas à máquina. Ao surgir um dado em que a classe não é definida, há um cálculo das distâncias dele aos objetos armazenados. Para tal cálculo é usada, na maioria das vezes, a distância euclidiana. A classe mais frequente nos *k* exemplos mais próximos é recomendada para o dado de entrada.

Assim, o parâmetro mais importante para tal algoritmo é o valor de *k*. Este parâmetro é escolhido de acordo com a quantidade de “vizinhos” que o usuário quer usar para classificar os dados de entrada.

Para um melhor entendimento, pode-se imaginar duas classes, A e B, como é possível verificar na Figura 4. Nesta figura são ilustrados vários exemplos da classe A (estrela vermelha) e classe B (triângulo verde) que são representados por dois atributos cada: exame X e exame Y.

Figura 4 – Definição de classes kNN



Fonte: Luz (2019).
Nota: Adaptado pela autora.

Pode ser observado na Figura 4, que para *k* = 3 foram encontrados, em relação ao objeto com interrogação, dois vizinhos mais próximos da classe B, contra um da classe A, assim, a amostra é classificada como sendo da classe B. Já para *k* = 7, a amostra é classificada em A, por ter sido encontrados mais vizinhos da classe A do que da B. É importante notar que quando há duas categorias de rótulos é recomendado utilizar um *k* ímpar, para não correr o risco da máquina encontrar a mesma quantidade de vizinhos próximos para ambas as classes.

Como diferentes valores de *k* podem dar resultados diferentes para um mesmo conjunto de dados, é necessário definir o melhor *k*. Para essa definição, pode-se usar o *k-fold cross validation*, que é um procedimento usado para avaliar modelos de aprendizado de máquina em uma amostra de dados limitada.

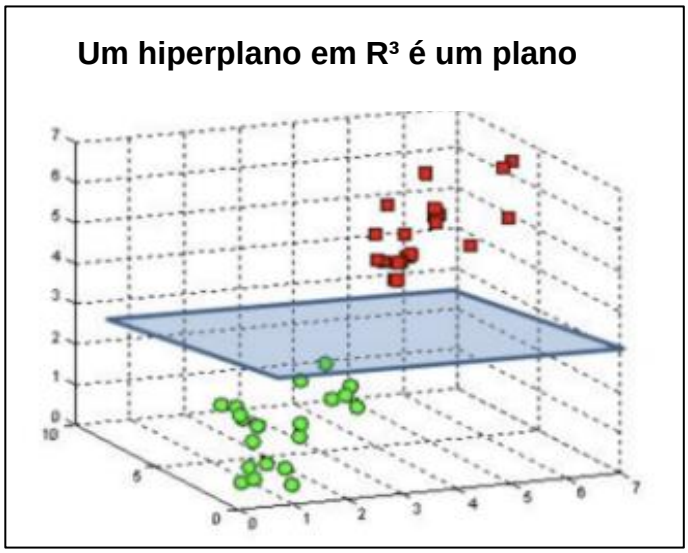
Apesar da simplicidade do kNN, ele pode trazer um esforço computacional, já que a distância terá que ser medida para cada dado. Outro fator relevante é que o objeto mais próximo pode ainda estar muito distante daqueles de classe definida, o que nem sempre traz uma veracidade para a análise dos fatos (FACELI et al., 2017).

2.2.2 SVM

As máquinas de vetores de suporte são fundamentadas na teoria de aprendizado estatístico, que consiste em estabelecer condições matemáticas para a escolha de um classificador. Tais condições consideram o desempenho do classificador no conjunto de treinamento e sua complexidade (FACELI et al., 2017).

O algoritmo, ao estabelecer tais condições matemáticas tem por objetivo encontrar um hiperplano no espaço *N*-dimensional, e então usá-lo como limite de decisão para a classificação dos objetos.

Figura 5 – Hiperplano alocado para a divisão de classes



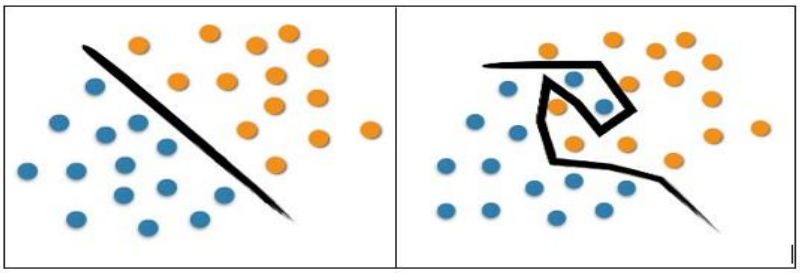
Fonte: *Towards Data Science* (2018).
Nota: Adaptado pela autora.

Na Figura 5 é possível observar um hiperplano bidimensional que separa as classes vermelhas e verdes. É importante observar que um hiperplano é alocado para que tenha distância máxima entre os dados de cada classe, o que é chamado de margem

máxima, que serve para que os dados futuros possam ser classificados com mais segurança (TOWARDS DATA SCIENCE, 2018).

Porém, o classificador de margem máxima é mais eficiente para os casos nos quais os dados analisados são linearmente separáveis, o que pode não ser adequado para dados ruidosos, como é o caso do espectro de frequência. Pode-se observar na Figura 6 a comparação de dados linearmente separados e dados não-linearmente separados.

Figura 6 – Dados linearmente separados x dados não-linearmente separados



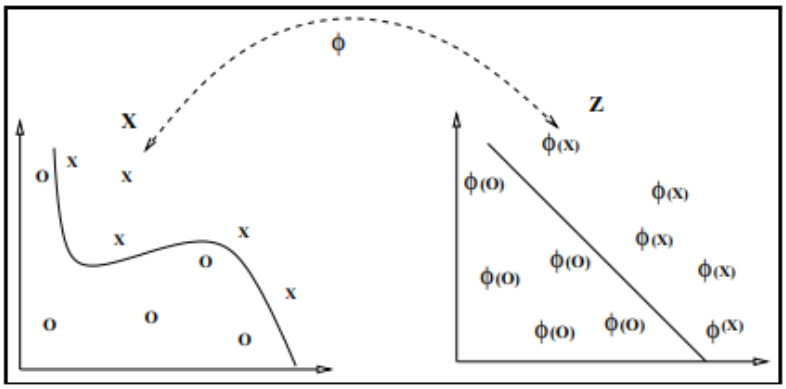
Fonte: Montenegro (2017).

Ao ocorrer classificações mais complexas, nas quais há a ocorrência de dados não-linearmente separáveis, usam-se equações de *kernel* para reorganizar tais objetos. Essas equações são artifícios matemáticos que servem para mapear os dados de um espaço *X* em um novo espaço Z (SANTOS, 2002), como pode ser visto na equação (11), onde ∅ representa as funções de *kernel*.

$$Z = \{\emptyset\ (x) | x\ \in X\} \tag{11}$$

Uma melhor representação é mostrada na Figura 7, a qual ilustra que a partir da mudança dimensional os dados não-linearmente separados no espaço X podem ser linearmente separáveis no espaço Z, onde o segundo costuma ter uma dimensão maior que o primeiro. É importante salientar que na figura os espaços X e Z se encontram representados em $\mathbb{R}^2$ para uma melhor visualização.

Figura 7 – Mapeamento do espaço de entrada via função *kernel*



Fonte: Santos (2002).

Logo, apesar da dimensão do espaço aumentar, a classificação do modelo se torna mais simples, já que na nova dimensão os dados podem ser linearmente separáveis. Mas é importante considerar que no caso de dimensões muito elevadas a computação pode ser bem custosa (FACELI et al., 2017).

Assim como o kNN depende da escolha do valor do *k* para obtenção de resultados satisfatórios, o SVM depende diretamente da escolha do *kernel* para obtenção de bons resultados. Na prática, os *kernels* mais utilizados são os polinomiais, as funções de base radial (RBF, do inglês *radial basis function*) e os sigmoidais. Uma boa escolha inicial é empregar o RBF (NATIONAL TAIWAN UNIVERSITY, 2003). Pode-se encontrar no Quadro 1 as funções dos kernels aqui apresentadas, onde *K* são matrizes semidefinidas positivas, em que cada elemento K_{ij} é definido por $K_{ij} = K\ (x_i, x_j)$, para todo $i, j = 1, \dots, n$, segundo Hebrich (2001, apud FACELI et al., 2017, p. 132). Sobre os parâmetros, *d* é o grau da função polinomial e σ é o desvio padrão da RBF.

Quadro 1 – Funções de *Kernel* mais comuns

Tipo de *kernel*	**Função**	**Parâmetros**
Polinomial	$\delta\ ((\mathrm{x_i}\ .\ x_j) + K)^d$	δ, K, d
RBF	$exp(-\sigma \mid\mid \mathrm{x_i} - x_j \mid\mid^2)$	σ
Sigmodal	$\tanh\ (\ \delta\ (\mathrm{x_i}\ .\ x_j) + K)$	δ, K

Fonte: Faceli, Lorena, Gama e Carvalho (2017).
Nota: Adaptado pela autora.

2.3 Pré-processamento dos Dados

O primeiro passo para a aplicação de uma técnica de AM é a preparação dos dados, para um uso posterior deles. Os dados devem ser representados por uma matriz de objetos X_n^d, em que *n* é o número de objetos e *d* o número de atributos (FACELI et al., 2017).

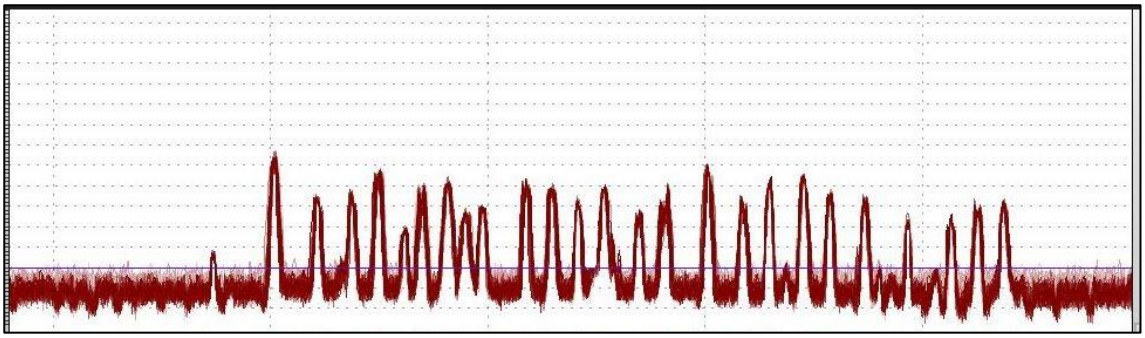
Entretanto, para se ter uma matriz de objetos, técnicas de pré-processamento de dados são frequentemente utilizadas, sendo útil não apenas para minimizar ou eliminar problemas como também para deixar os dados mais adequados para a situação (FACELI et al., 2017).

No caso específico deste trabalho, os dados coletados são amostras discretizadas de frequências em um intervalo entre 87,4 e 108 MHz, contendo informações de potência. Com a aplicação de técnicas de pré-processamento de sinais pode-se identificar as portadoras no sinal, as quais são os objetos usados na classificação. Antes de aplicar as técnicas de AM, é utilizado um extrator de características nas portadoras para representá-las na forma de uma matriz X_n^d, neste caso específico, *n* é o número de portadoras e *d* é o número de características para representar cada portadora.

As técnicas de AM analisarão o comportamento das portadoras a partir do espectro, e classificarão as mesmas como normais ou anormais. As portadoras anormais são aquelas que o fiscal identificou um comportamento ruidoso fora do comum.

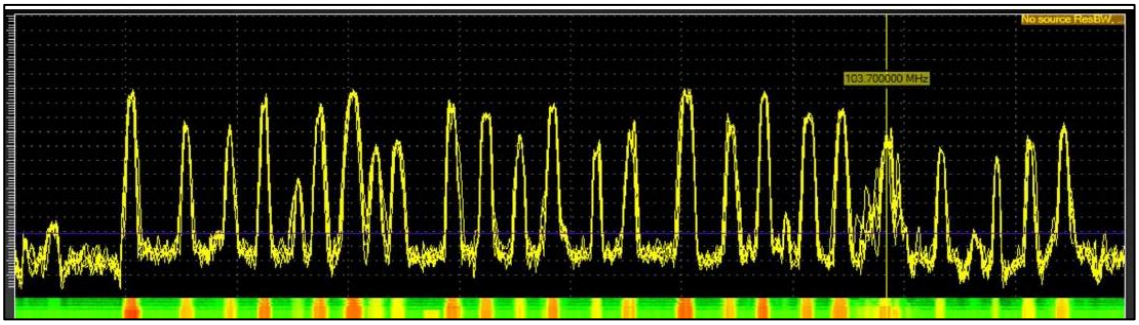
Na Figura 8 pode ser visto um espectro de frequência contendo portadoras com comportamentos normais. Por sua vez, na Figura 9, é possível verificar o comportamento anormal do espectro da portadora com frequência central de 103,7 MHz, imagem esta retirada de um relatório de fiscalização nº 61 (AGÊNCIA NACIONAL DE TELECOMUNICAÇÕES, 2019).

Figura 8 – Faixa de espectro que recebe o rótulo de normal



Fonte: ANATEL (2019).

Figura 9 – Faixa de espectro que recebe o rótulo de anormal



Fonte: ANATEL (2019).

É importante salientar que as portadoras anormais são aquelas que foram julgadas muito ruidosas por fiscais da ANATEL, e esta condição é definitiva para a abertura de processos de vistorias. A Figura 9 é uma tela de espectro que serviu para abrir um processo de vistoria à rádio 103,7 MHz.

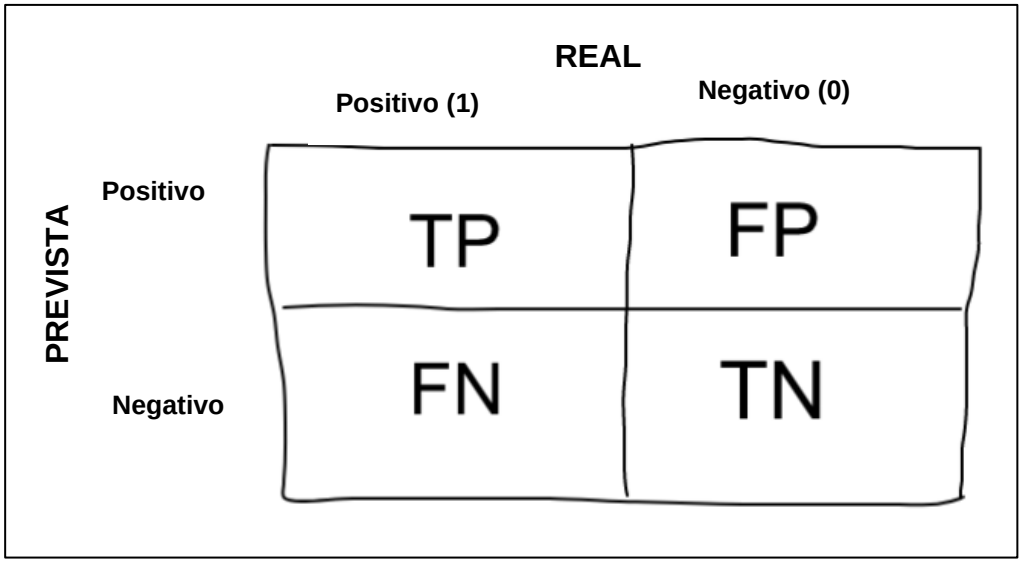
2.4 Métricas de Desempenho

Definidas as técnicas de aprendizado, tem-se que verificar o sucesso delas ao realizar a tarefa. Para isso, utiliza-se a matriz de confusão para indicar os erros e acertos dos modelos testados. Com ela é possível calcular métricas de avaliação como acurácia, precisão, *recall* e *F1-score* (SUNASRA, 2017).

Para entender melhor essas métricas, precisa-se analisar a matriz de confusão explanada na Figura 10. Ela é uma tabela, na qual as colunas exibem os números de casos reais (normal e anormal), anotados pelos especialistas, e as linhas os números

previstos pela máquina. Em positivo (1) tem-se a presença de anormalidade e em negativo (0) o dado é considerado normal.

Figura 10 – Matriz de confusão



Fonte: SUNASRA (2017).
Nota: Adaptado pela autora.

Nesse contexto, há os verdadeiros positivos (TP, do inglês *true positive*), que acontecem quando a classe real e a prevista são positivas. Os falsos positivos (FP, do inglês *false positive*), quando a classe real é negativa e a prevista é positiva. Os falsos negativos (FN, do inglês *false negative*), onde a classe real é positiva e a prevista é negativa. E, por último, verdadeiros negativos (TN, do inglês *true negative*) onde ambos, reais e previstos, são negativos (SUNASRA, 2017). Neste trabalho, as anomalias serão da classe positiva, pois estas são as que possuem o menor número de amostras, o que afeta as métricas, além de ser de interesse para o trabalho.

A partir da matriz de confusão podem ser obtidas diferentes métricas. Uma das mais comuns é a acurácia, que mede a porcentagem de classificações corretas, como pode ser vista na equação (12).

$$acurácia = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{12}$$

Já a precisão (P) mostra a porcentagem de dados anormais previstos que realmente aconteceram, como mostra a equação (13).

$$P = \frac{TP}{TP + FP} \times 100\% \tag{13}$$

Ao contrário da *precisão,* o *recall* (R) fornece informação sobre o desempenho baseado nos falsos negativos (SUNASRA, 2017). Ele indica a porcentagem de anormalidades que foram identificadas em relação à quantidade total. Tal métrica é possível ser verificada na equação (14).

$$R = \frac{TP}{TP + FN} \times 100\% \tag{14}$$

Por último, tem-se a *F1-score* (F1). Esta métrica é a média harmônica entre a precisão e o *recall,* como é possível verificar na equação (15).

$$F1 = \frac{2\ x\ P\ x\ R}{(R + P)} \times 100\% \tag{15}$$

Por se tratar de uma média harmônica entre *recall* e precisão, o *F1-score* é a principal métrica avaliada no trabalho. Embora a acurácia seja muito utilizada, ela não é recomendada para tarefas que envolvem bases de dados desbalanceadas, que é o caso deste trabalho, já que se tem muito mais portadoras normais (classe negativa) do que anormais (classe positiva). Logo, a avaliação das classes positivas, fica prejudicada neste sentido.

3 METODOLOGIA E EXECUÇÃO DO PROJETO

3.1 Recurso Computacional Utilizado

Para o desenvolvimento do trabalho e manipulação dos dados, foi necessário utilizar recursos computacionais. Para isso, o ambiente de desenvolvimento *python* foi utilizado. Neste sentido, o Anaconda, que é uma distribuição gratuita e de código aberto das linguagens de programação *python,* foi utilizada.

O computador da autora do projeto contempla com as configurações necessárias para a execução do programa, sendo um Aspire 5, com processador *Intel Core* i5 7200U, placa de vídeo NVIDIA *GeForce* 940MX, memória 8 GB DDR4, 1 TB de espaço em disco, *Windows 10,* e já com o Anaconda instalado.

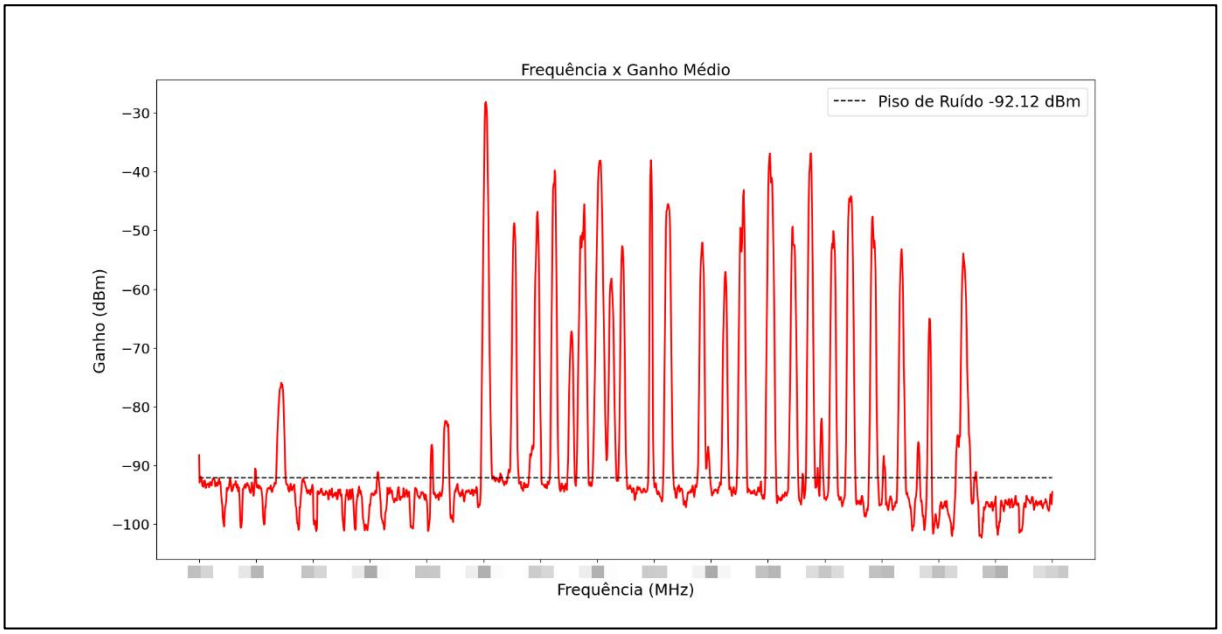
3.2 Etapas de Desenvolvimento

O desenvolvimento do projeto foi dado a partir das amostras discretas de ganho do espectro de FM da faixa de 80 a 110 MHz captadas pelo *RFeye*, a qual contempla a faixa de FM. Na faixa de frequência referida, a cada 20 kHz tem-se o ganho de cada frequência amostrada. O conjunto desses dados será chamado de vetor de amostras.

Na primeira etapa do trabalho é realizado o tratamento de sinal. Para tal feito, observou-se que um vetor de amostras é gerado a cada seis segundos. Com o intuito de otimizar o processamento e analisar um sinal mais suavizado, optou-se por fazer a média do sinal a cada dez amostras, obtendo assim um vetor de amostras por minuto. Com isso, o próximo passo realizado é a eliminação do piso de ruído do espectro.

Para tal, é calculada a mediana dos ganhos das frequências amostradas, definindo assim um piso de ruído. É importante salientar que isso só é efetivo porque nesse caso a quantidade de amostras das portadoras é menor que 50% do total de amostras do vetor, como é possível verificar na Figura 11. Neste exemplo, a mediana eliminou todos os ganhos abaixo de -93,12 dBm.

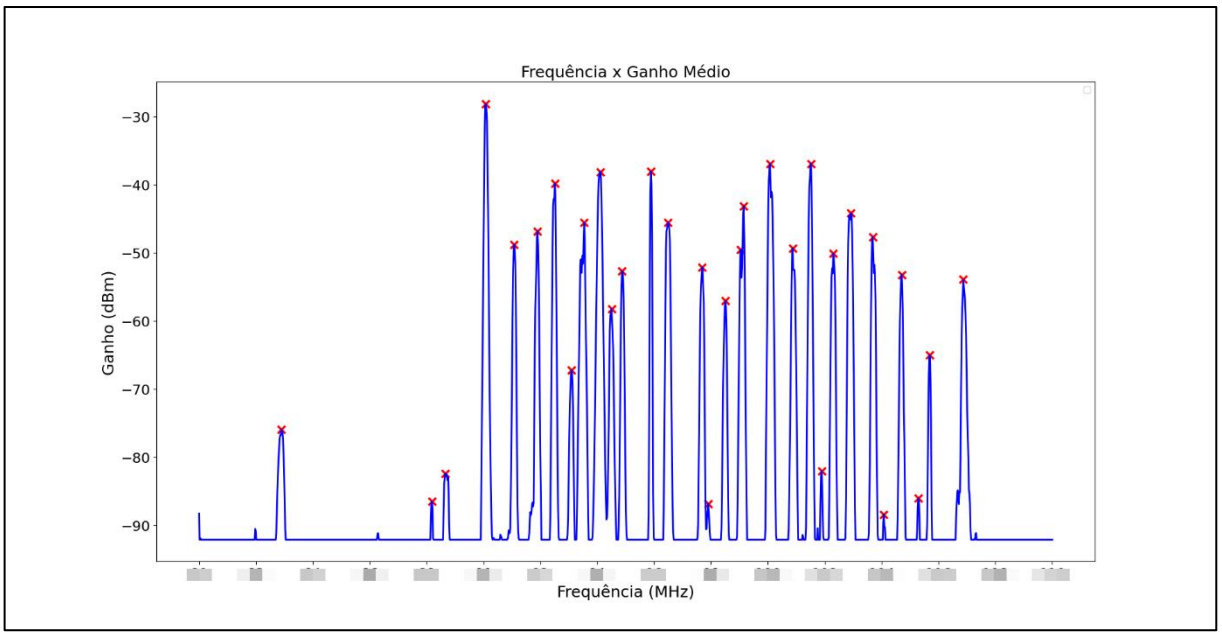
Figura 11 – Limiar do filtro de mediana



Fonte: Produzido pela própria autora.

Uma vez que o sinal foi suavizado, torna-se mais simples a manipulação dos dados. A Figura 12 mostra como, através do uso da função "*find_peaks*" (interna à biblioteca *"scipy_signal"*), foi possível realizar a detecção de cada um dos picos existentes na faixa de frequência.

Figura 12 – Picos detectados



Fonte: Produzido pela própria autora.

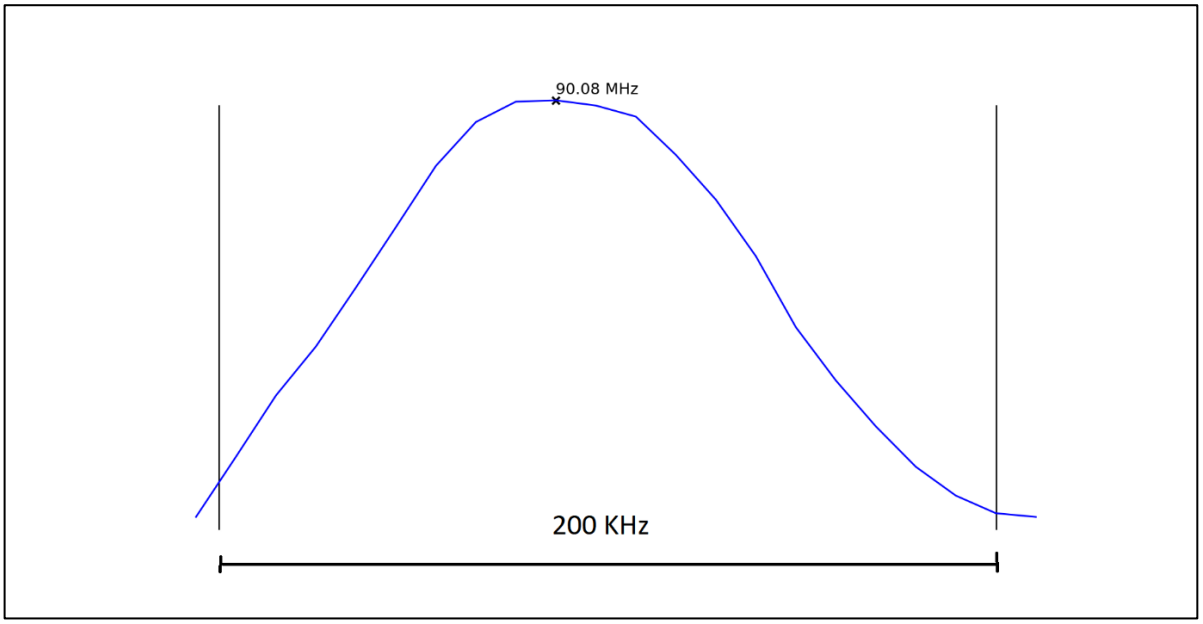
Com os picos de frequência achados, consequentemente podem ser identificadas as portadoras. Este é um dos principais objetos de análise deste trabalho.

Antes da aplicação do AM, duas verificações são feitas de modo a cumprir as normas da ANATEL. Estas averiguações são referentes a rádios licenciadas e a variação de potência do sinal.

3.2.1 Verificação de Rádio Não Licenciada

Inicialmente, foi levantado um banco de dados com as frequências centrais e nomes de cada rádio licenciada da Grande Vitória, no formato *comma-separated-values* (CSV), como pode ser visto no apêndice A. Então, pode-se comparar as frequências de pico encontradas na seção 3.2 com as do banco de dados. Se a frequência de pico se encontrar nos limites de um canal de FM, como visto na Figura 13, de uma frequência licenciada, o nome da rádio é associado a tal frequência, caso contrário, aponta-se uma rádio não licenciada.

Figura 13 – Limites para identificação de uma portadora



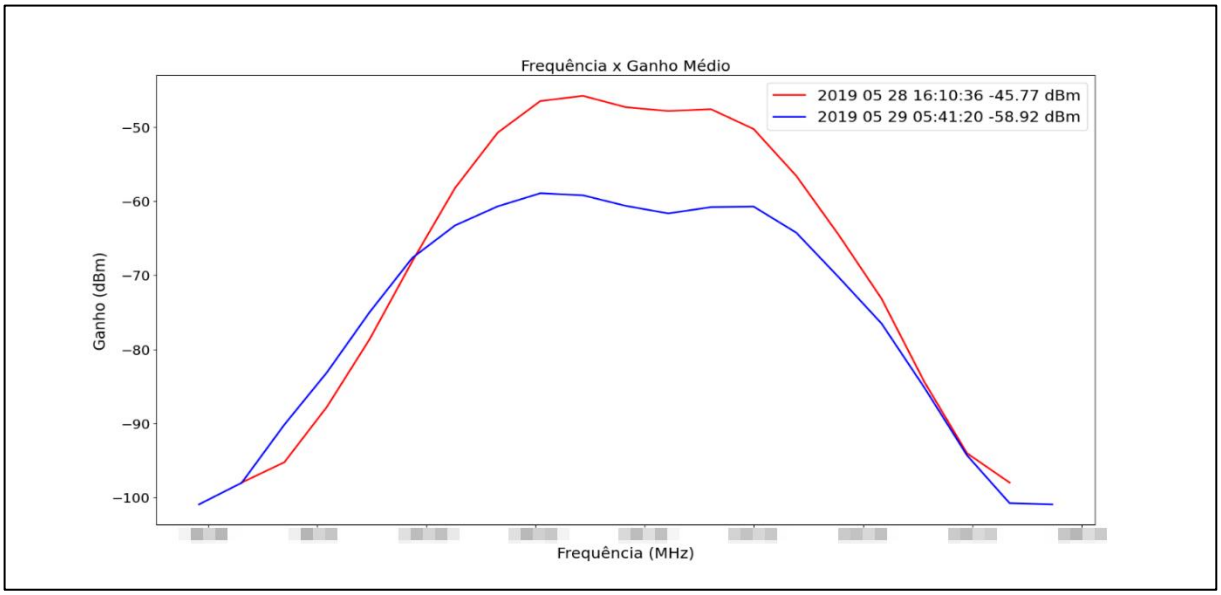
Fonte: Produzido pela própria autora.

3.2.2 Verificação da Variação de Potência

Antes da aplicação da AM, pode-se também analisar a potência das portadoras. Nesta perspectiva, os picos de potência das portadoras são armazenados e tais valores são comparados aos respectivos picos identificados em momentos posteriores. A cada leitura é verificado se tais potências variaram 10% da primeira potência gravada.

Por exemplo, tem-se que a primeira verificação da potência de uma determinada portadora, como mostra a Figura 14, foi feita no dia 28/05/2019, às 16:10:36, e foi identificada uma potência de -45,77 dBm. Em uma determinada leitura às 05:41:20 do dia 29/05/2019, a potência verificada foi -58,92 dBm. Ou seja, uma variação de cerca de 28% em relação à primeira. Logo, quando o feito acontecer, a frequência da portadora será gravada em memória para o relatório final.

Figura 14 – Comparação de potência de uma mesma portadora



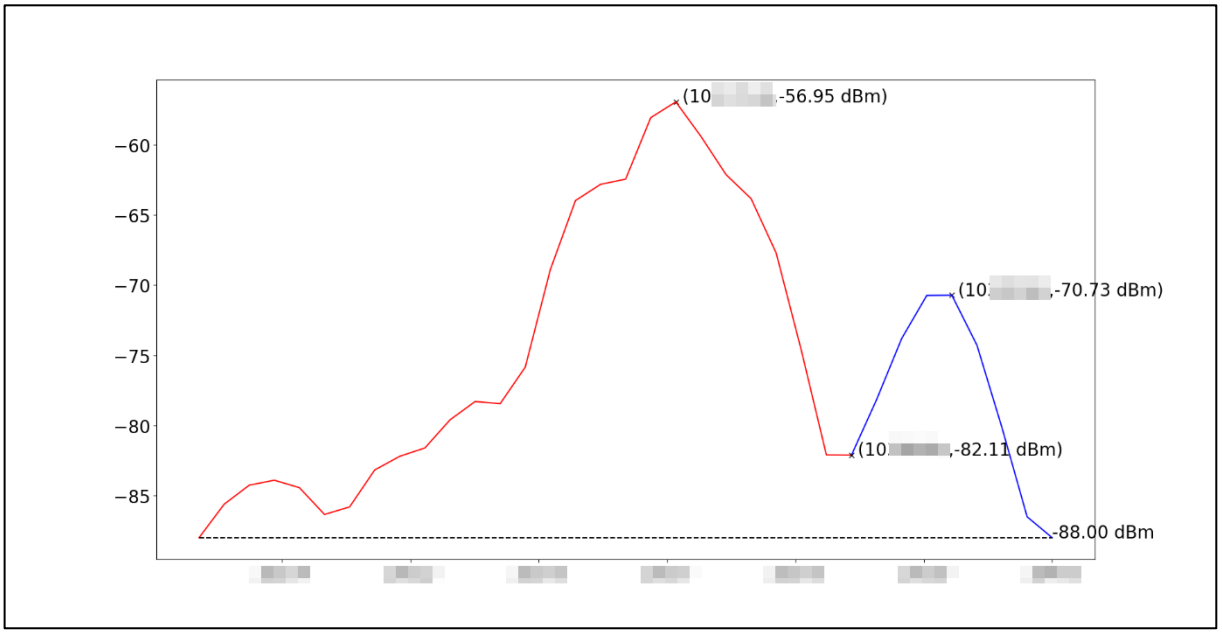
Fonte: Produzido pela própria autora.

3.2.3 Verificação de Portadoras com Ruídos Anormais

Para a aplicação do aprendizado de máquina na detecção de portadoras ruidosas, fez-se necessário analisar portadora a portadora individualmente. Tendo os picos já encontrados, os limites para cada portadora foram definidos como sendo os pontos de frequência que coincidem com o valor do piso de ruído. Porém, uma análise é feita anteriormente, na qual só é considerada uma portadora se o período verificado não

envolver dois picos, como observado no exemplo da Figura 15. Nesse caso, verificou-se que o valor do piso de ruído é de -88 dBm. Pode-se observar que entre as frequências de pico, as quais tem os ganhos de -55,85 dBm e -70,73 dBm não existe um ponto que coincida com o piso de ruído. Dessa forma, considerou-se o limite da portadora como o ponto mínimo entre os dois picos, que seria o ponto da figura onde está indicado o ganho de -81,11 dBm.

Figura 15 – Portadoras em que um dos limites divergem com o piso de ruído

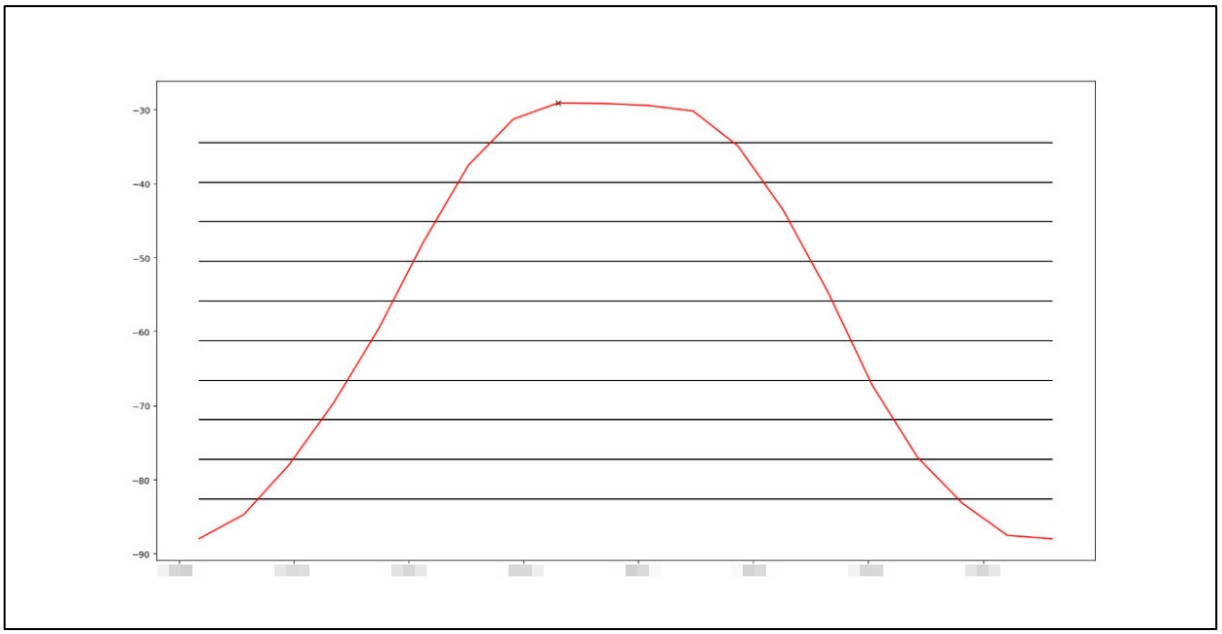


Fonte: Produzido pela própria autora.

Com as portadoras identificadas, estas podem ser rotuladas. A estratégia traçada para obter o vetor de características para representar as portadoras baseia-se na criação de limiares que cortam as mesmas, como indicado nas Figuras 16 e 17. O vetor de características é dado pela quantidade de vezes que cada limiar toca na portadora. No caso da Figura 16, o vetor de características de uma portadora normal é dado por (2,2,2,2,2,2,2,2,2,2). Já na Figura 17, tal vetor de uma portadora anormal é dado por (3,1,2,2,2,2,2,2,2,2).

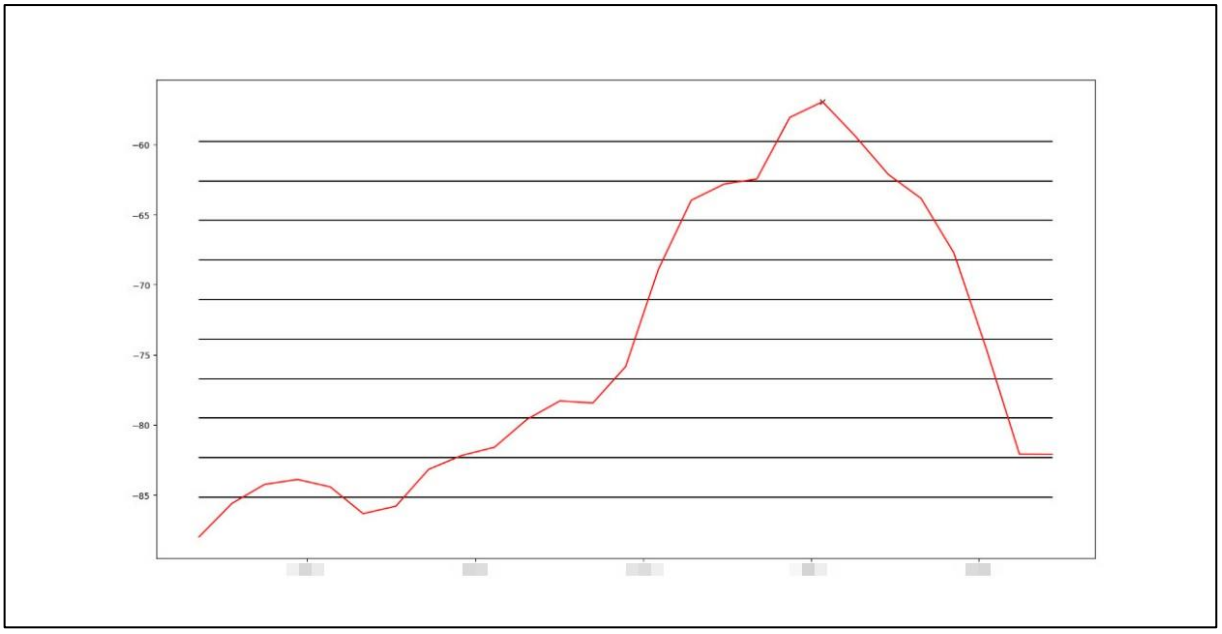
É importante salientar que para obter a caracterização das portadoras, a amplitude de cada uma delas é normalizada entre o pico e o piso de ruído. Logo depois, os limiares são distribuídos em espaçamentos iguais para cada portadora, de forma que todas as portadoras sejam representadas pelo mesmo número de limiares.

Figura 16 – Portadora com classificação normal e seus respectivos limiares



Fonte: Produzido pela própria autora.

Figura 17 – Portadora com classificação anormal e seus respectivos limiares



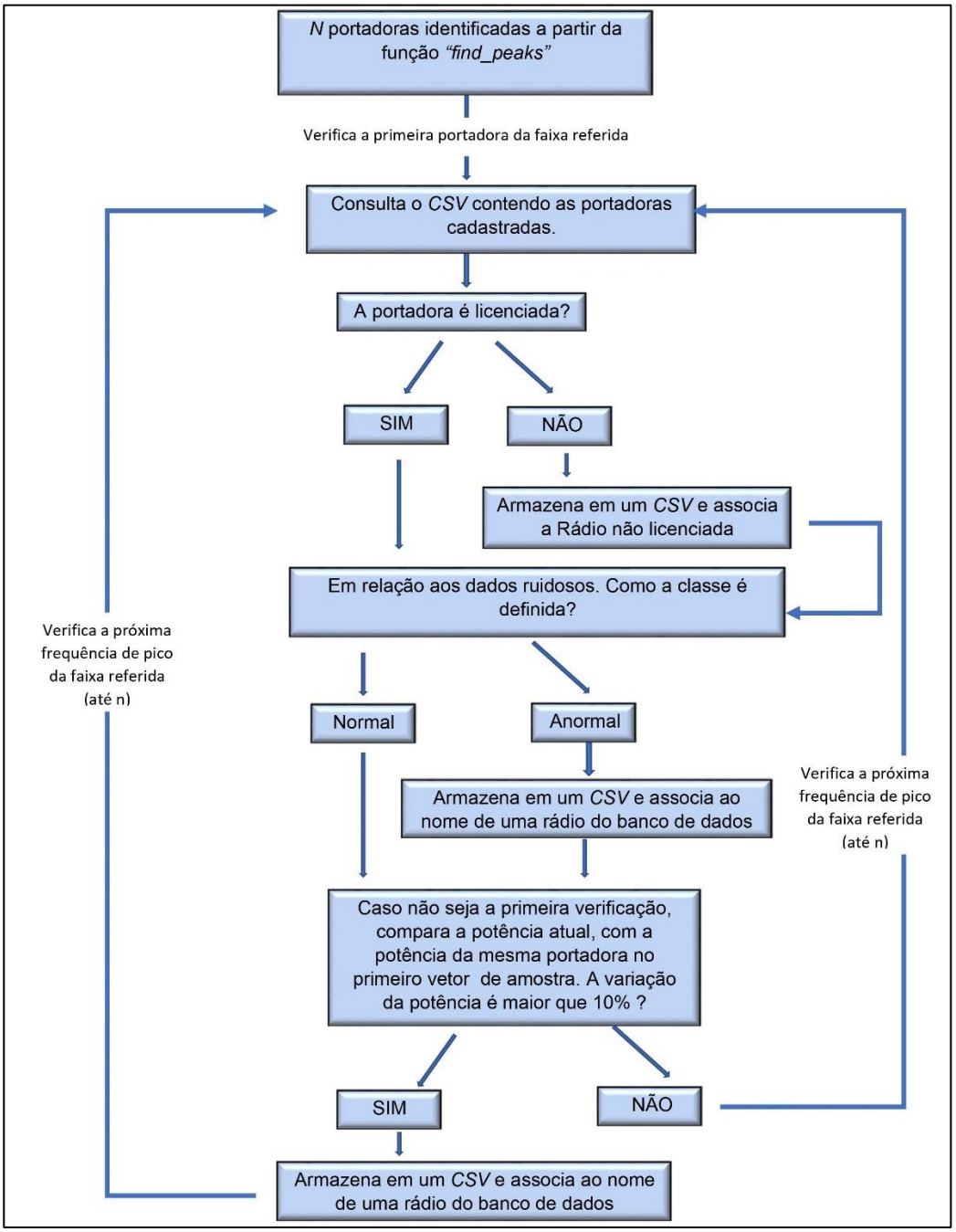
Fonte: Produzido pela própria autora.

Depois de cada portadora ser representada por um vetor de características, elas podem ser usadas como entrada para os algoritmos de AM.

3.3 Fluxograma do Método Proposto

A Figura 18 ilustra o processo de identificação de uma rádio, captura da variação de potência e a classificação entre normal e anormal para cada portadora. Os dados são armazenados para uma posterior aferição. Cabe lembrar que tal fluxograma representa o processo para cada portadora.

Figura 18 – Fluxograma para captura anomalias



Fonte: Própria Autora.

Os resultados coletados nesta etapa são armazenados tanto em memória quanto em um arquivo CSV, permitindo que estes sejam exibidos em uma *interface* posteriormente.

4 EXPERIMENTOS E RESULTADOS

Todos os experimentos foram realizados com os dados extraídos de uma *RFeye* localizada em Vitória-ES, na Enseada do Suá.

Os dados utilizados nos experimentos são compostos por 577 vetores de amostras (espectros). Desses 577 vetores de amostras, 16.944 portadoras foram identificadas, sendo 1.152 consideradas anormais e 15.792 normais.

Para a representação das portadoras na forma de vetores de características foram testados quatro limiares diferentes, sendo esses 5, 10, 15 e 20.

As técnicas de AM avaliadas foram kNN e SVM. No caso do kNN foi utilizada a biblioteca do python *"sklearn.neighbors*" com função *"KNeighborsClassifier"*. E para o SVM a biblioteca *"sklearn.svm"* com a função "*svm.SVC(kernel='*X*')*", onde 'X' é um parâmetro referente as funções de *kernel*, sendo eles linear, polinomial, RBF e sigmoidal.

4.1 kNN

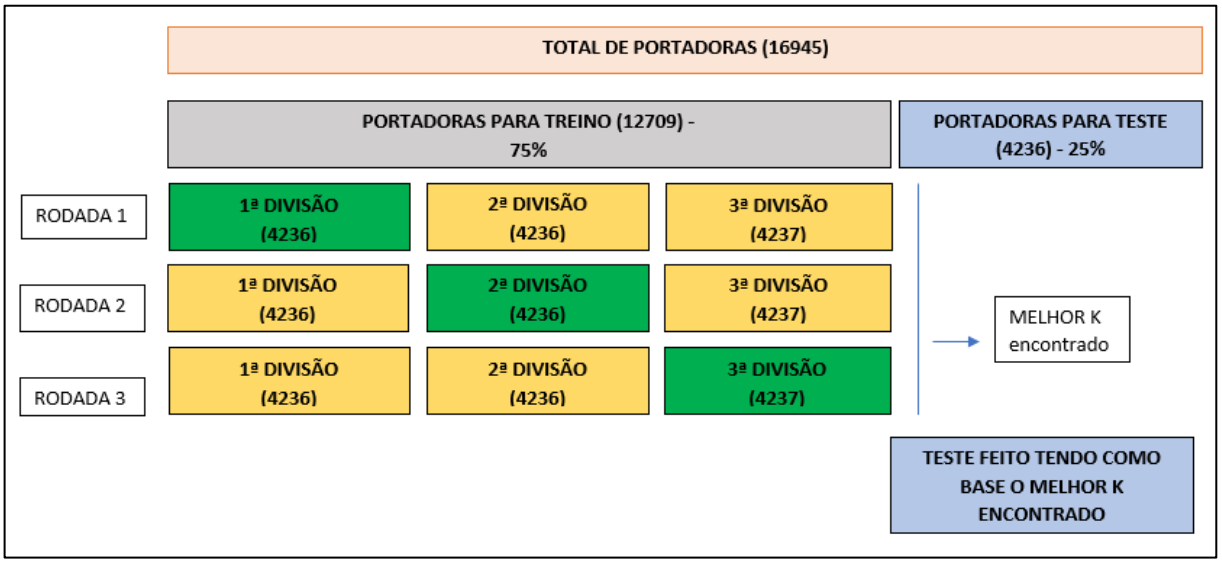
Para a aplicação do kNN*,* é necessário definir o valor de *k*, o qual pode ser obtido através da técnica de *k-fold cross validation*. Para isso, foi feita uma divisão estratificada das portadoras, utilizando 75% delas para treino e 25% para o teste.

Logo, foi utilizado *k-fold cross validation* sobre os 75% dos dados. Na técnica, o usuário define *folds* (*cv*), em que os dados serão divididos, para depois realizar o procedimento. Foi usada nesta etapa a função *"GridSearchCV"* do python.

A Figura 19 mostra o funcionamento do *k-fold cross validation*, onde *cv* = 3 e o *k* escolhido para o kNN pode ser um número ímpar entre 1 e 11 (definido pela autora do projeto). Logo, das 12.709 portadoras para o treinamento (75% das portadoras totais), tem-se uma divisão em 3 partes. Realizada a divisão, a cada rodada uma dessas partes é usada para validação, indicado pela cédula verde, e as outras duas

são usadas para o treinamento, indicadas pela cor amarela. Dessa forma, a cada rodada os valores de *k* são avaliados. Após fazer 3 rodadas, o *k* que retornar a melhor média de *recall, F1-score*, acurácia e precisão é selecionado. Depois de encontrado o melhor valor de *k*, um teste é feito com as 4.236 portadoras (retângulo indicado pela cor azul), usando os 75% de portadoras para treino (retângulo indicado pela cor cinza). Logo, métricas são geradas para a avaliação do algoritmo.

Figura 19 – Funcionamento do *k-cross validation* para o problema



Fonte: Produzido pela própria autora.

Assim, para cada quantidade de limiares, foi identificado um melhor valor de *k* para o kNN, como pode ser visto no Quadro 2.

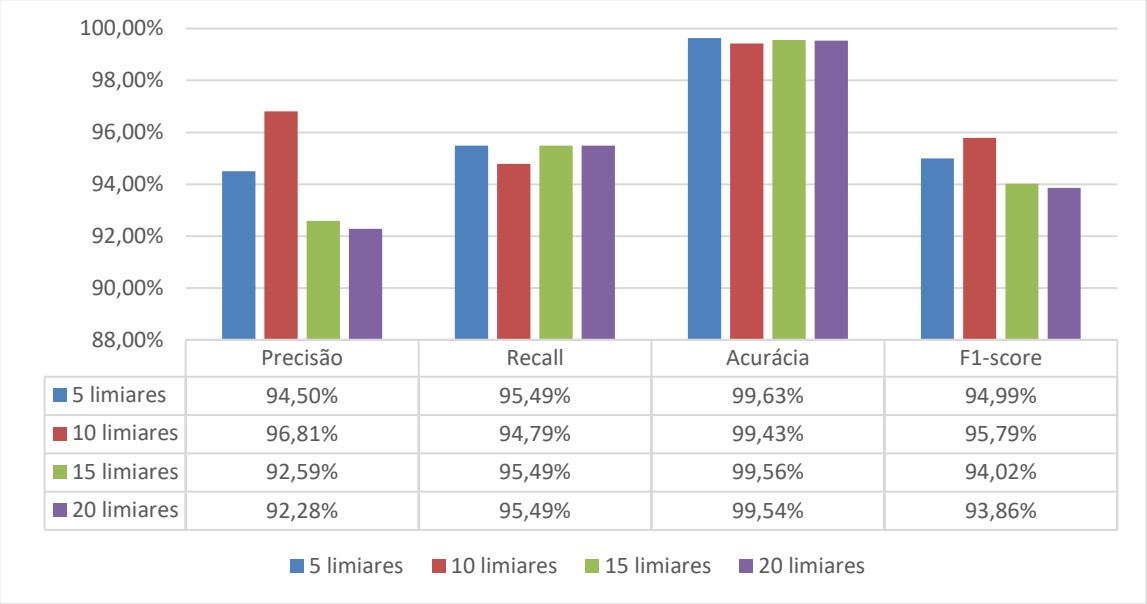
Quadro 2 – Argumento *k* do kNN em relação ao número de limiares

Quantidade de limiares	***K* obtido**
5	3
10	1
15	7
20	1

Fonte: Produzido pela própria autora.

Com o *k* definido pôde-se calcular as métricas de acurácia, precisão, *recall* e *F1-score* para cada quantidade de limiares, considerando a classe positiva como as portadoras anormais. Tais resultados podem ser vistos na Figura 20.

Figura 20 – Métricas referentes ao kNN em relação à quantidade de limiares



Fonte: Produzido pela própria autora.

Pode ser observado na Figura 20 que as métricas alcançaram taxas acima dos 92%, com destaque para os vetores de características de 10 limiares, que se comportaram melhores em dois quesitos, *F1-score* e precisão. Sendo *F1-score* é a métrica que melhor serve para análise, considera-se que 10 limiares é cenário mais adequado.

Depois da etapa de teste, pode-se fazer uma análise referente ao tempo, na qual foi verificado quanto tempo gastou-se para criar um vetor de características, o treinamento do kNN e, por final, tempo de classificação de cada portadora.

Logo, pode-se ver no Quadro 3 que o tempo de construção de um vetor de características por portadora e o de classificação se deu na ordem de milissegundos, já o de treinamento se deu na ordem de microssegundos, mostrando que o treinamento do algoritmo é cerca de 1.000 vezes mais rápido, que a construção e a classificação do vetor de características. Nota-se, ainda, com o aumento no número de limiares, os tempos aumentaram, o que é compreensível por se tratar de vetores de características maiores, ou seja, com mais dados. Contudo, considerando o número de portadoras por minuto e o tempo de processamento de cada uma, pode-se observar que o sistema consegue funcionar *online*.

Quadro 3 – Tempos de construção, treinamento e classificação do kNN de acordo com os limiares

Quantidade de Limiares	Tempo de construção do vetor de característica (ms)	Tempo do treinamento (µs)	Tempo de classificação (ms)
5	1,0722	0,1777	0,7753
10	1,1254	0,1652	0,8852
15	1,1389	0,1947	1,9621
20	1,1552	0,2003	2,0785

Fonte: Produzido pela própria autora.

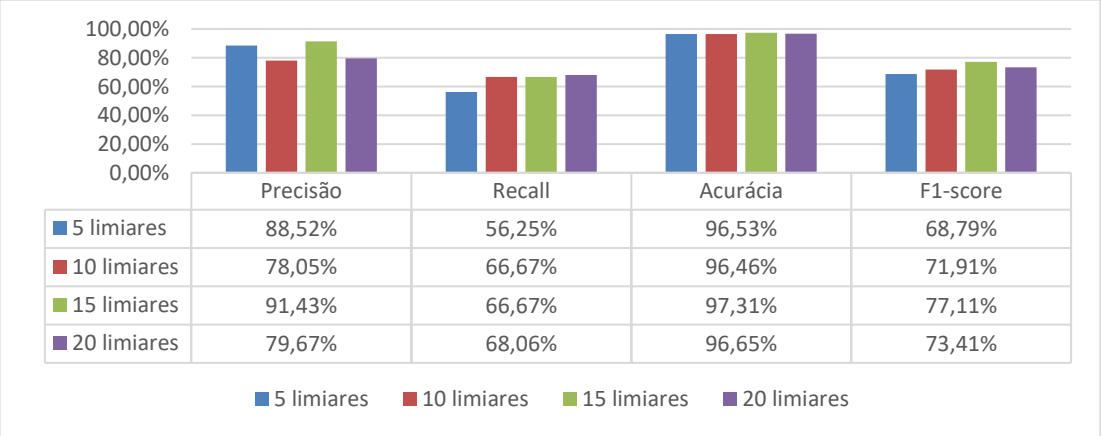
4.2 SVM

Se tratando do SVM, foram utilizadas as mesmas 16.945 portadoras para obtenção das métricas. Para a prática foi necessário o uso da biblioteca "*sklearn.svm*" do python para a implementação do algoritmo. Ela já vem configurada com valores pré-definidos para os parâmetros padrões. Porém, no caso dos kernels polinomial e para o RBF, houve a tentativa de se obter os melhores valores de *d* e σ respectivamente.

Para os *kernels* de RBF e polinomial foi utilizada a técnica de *k-fold cross validation* com a mesma configuração da Figura 19. Já para os *kernels* linear e sigmodal, não se otimizou nenhum parâmetro, logo, aplicou-se 75% das portadoras para treino e o restante para teste.

A primeira análise a ser feita foi em relação às métricas, referente ao *kernel* linear. A Figura 21 apresenta as métricas considerando a classe anormal como classe positiva.

Figura 21 – Métricas referentes ao SVM, *kernel* linear, em relação à quantidade de limiares



Fonte: Própria Autora.

Pode-se perceber na Figura 21 que apenas a acurácia tem resultados maiores que 90%. Em contrapartida, as métricas de *recall* e *F1-score* não chegam a 80%, o que não traz um resultado de classificação tão confiável.

Sobre o *kernel* polinomial, foi utilizado o *k-fold cross validation* para identificar o melhor valor de *d* entre os números inteiros 2, 3 e 4 (graus definidos arbitrariamente pela autora). Assim, pode ser encontrado o melhor *d* para cada quantidade de limiares, como pode ser visto no Quadro 4.

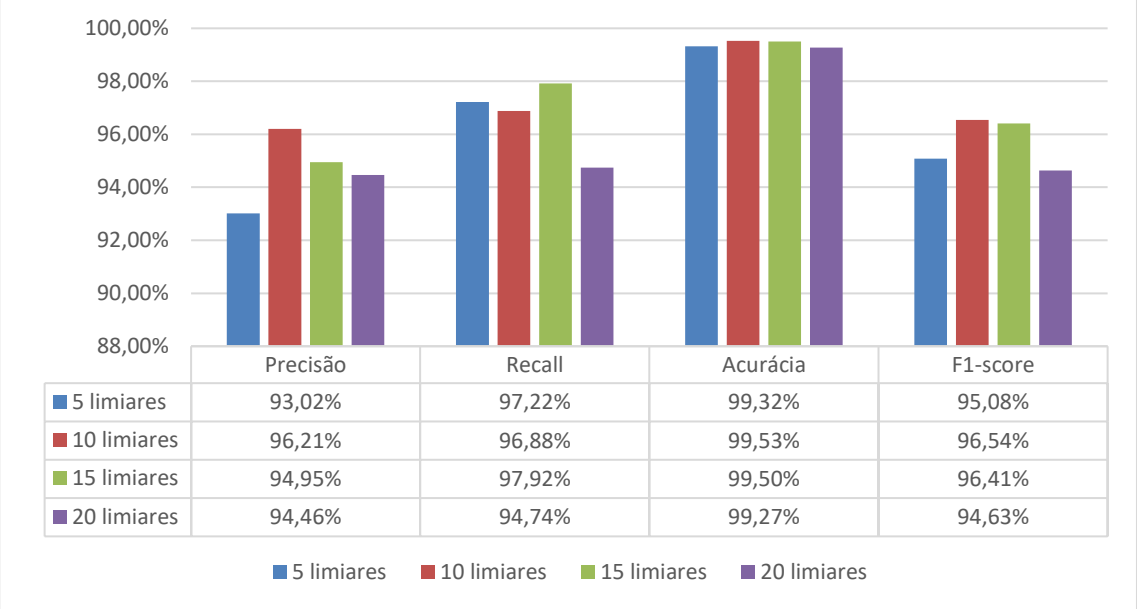
Quadro 4 – Argumento d do SVM- polinomial em relação ao número de limiares

Quantidade de limiares	$\boldsymbol{d}$ **obtido**
5	3
10	4
15	3
20	2

Fonte: Produzido pela própria autora.

Os resultados das métricas deram acima de 90%, como podem serem vistos na Figura 22. Sobre a eficiência do algoritmo, os vetores de características de 10 limiares mostraram-se com o melhor desempenho se tratando de *F1-score,* precisão *e* acurácia.

Figura 22 –Métricas referentes ao SVM, *kernel* polinomial, em relação à quantidade de limiares



Fonte: Produzido pela própria autora.

Sobre o *kernel* de *RBF*, a validação através do *k-fold cross validation* se deu procurando o melhor σ entre 0,1 a 10, com o passo de 0,1 (valores definidos arbitrariamente pela autora). Após a aplicação, o melhor σ para cada limiar foi encontrado, descrito no Quadro 5.

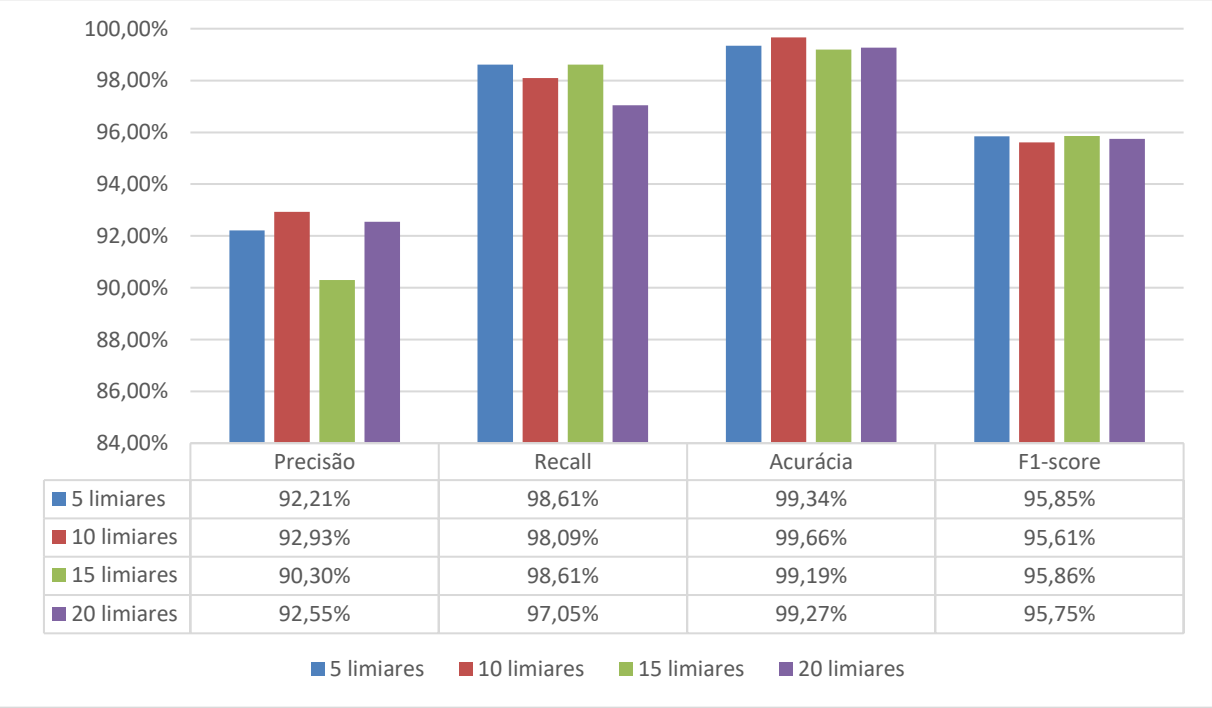
Quadro 5 – Argumento σ do SVM- RBF em relação ao número de limiares

Quantidade de limiares	**σ obtido**
5	1
10	0,3
15	0,1
20	0,2

Fonte: Produzido pela própria autora.

Logo, as métricas acurácia, precisão, *recall* e *F1-score* podem ser calculadas para cada quantidade de limiares, considerando como classes positivas as portadoras anormais. Tais resultados podem ser vistos na Figura 23.

Figura 23 – Métricas referentes ao SVM, *kernel* RBF, em relação à quantidade de limiares



	Precisão	Recall	Acurácia	F1-score
5 limiares	92,21%	98,61%	99,34%	95,85%
10 limiares	92,93%	98,09%	99,66%	95,61%
15 limiares	90,30%	98,61%	99,19%	95,86%
20 limiares	92,55%	97,05%	99,27%	95,75%

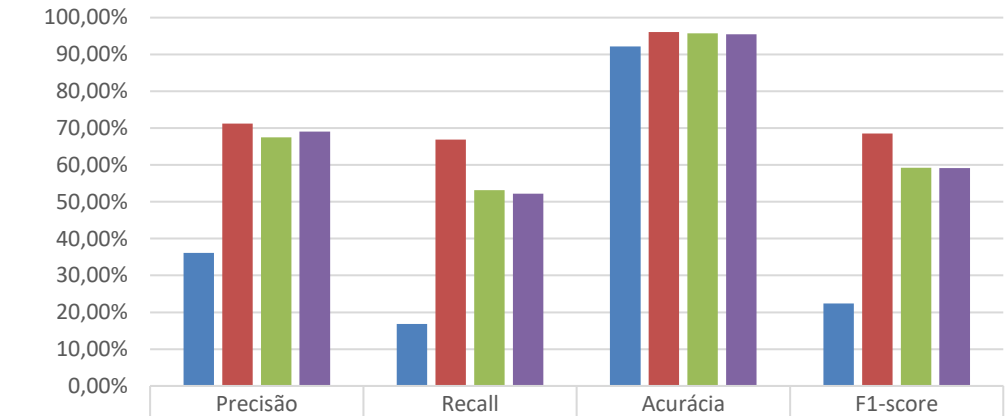
Fonte: Produzido pela própria autora.

Pode-se observar que as métricas em relação ao RBF são, na maioria dos casos, acima de 92%. Além disso, o vetor de características com 5 limiares e 15 limiares

tiveram desempenhos muito parecidos. Como os últimos mostraram melhores desempenhos, fez-se necessário considerar o tempo de processamento como elemento decisivo.

Por último, sobre o *kernel* sigmodal, mostrado na Figura 24, pode-se observar resultados muito inferiores aos *kernels* analisados anteriormente. O *F1-score*, métrica mais relevante para análise, não chegou a 70% em nenhum dos casos.

Figura 24 – Métricas referentes ao SVM, *kernel* sigmodal, em relação à quantidade de limiares



	Precisão	Recall	Acurácia	F1-score
5 limiares	36,12%	16,82%	92,21%	22,37%
10 limiares	71,23%	66,89%	96,11%	68,52%
15 limiares	67,49%	53,19%	95,74%	59,21%
20 limiares	69,10%	52,23%	95,45%	59,12%

Fonte: Produzido pela própria autora.

Em se tratando do tempo gasto para a criação de um vetor de características, o treinamento do SVM e o tempo de classificação de cada portadora, apenas os casos dos *kernels* RBF e polinomiais foram avaliados, visto que os *kernels linear e* sigmodal mostraram resultados bem inferiores aos demais. Os tempos em função do número de limiares são apresentados nos Quadros 6 e 7.

Quadro 6 – Tempos de construção, treinamento e classificação do kernel: RBF de acordo com os limiares

Quantidade de Limiares	Tempo de construção do vetor de característica (ms)	Tempo do treinamento (µs)	Tempo de classificação (µs)
5	0,851	13,9512	8,6116
10	0,895	12,7732	6,9997
15	0,897	39,1301	21,8271
20	0,967	45,0789	20,9372

Fonte: Produzido pela própria autora.

Quadro 7 – Tempos de construção, treinamento e classificação do *kernel*: polinomial de acordo com os limiares

Quantidade de Limiares	Tempo de construção do vetor de característica (ms)	Tempo do treinamento (ms)	Tempo de classificação (µs)
5	0,9505	16,82	3,0446
10	0,9018	0,9758	4,7037
15	0,9708	0,3148	4,7061
20	1,0487	0,1944	7,0332

Fonte: Produzido pela própria autora.

Com relação ao Quadro 6, tem-se que o tempo de construção do vetor de características ficou por volta de um milissegundo por portadora, aumentando de acordo com a quantidade de limiares. Considerando o tempo de treinamento e de classificação, observa-se que todos ocorreram na ordem de dezenas de microssegundos. Nota-se que a variação do valor de *F1-score* para o kernel RBF, quando analisando todas as quantidades de limiares, foi inferior a 0,05%. Portanto, a escolha de 5 limiares por portadora é a mais viável a ser usada com este kernel, já que tem também os menores tempos.

Sobre o Quadro 7 (*kernel* polinomial), observa-se que o tempo de construção de vetores tem pouca variação, flutuando em torno de um milissegundo por portadora. O tempo de treinamento, nesse caso, também se deu na ordem de milissegundos, mostrando que em relação ao *kernel* RBF, o polinomial exige um maior custo computacional nesse quesito. Por último, se tratando do tempo de classificação, tem-se que este varia entre 3 e 7 µs, aumentando de acordo com a quantidade de limiares.

É importante destacar que os tempos de classificação dos kernels polinomial e RBF foram cerca de 100 a 1.000 vezes menores do que os obtidos pelo kNN, tornando o algoritmo de SVM mais eficiente para a aplicação.

Considerando os tempos de classificação e os resultados obtidos nas métricas de desempenho, o algoritmo que melhor contempla a identificação de portadoras anormais é o SVM com kernel polinomial de quarta ordem, utilizando 10 limiares. Com essa escolha, foram alcançados 96,21% de precisão, 96,88% de *recall,* 99,53% de acurácia e 96,54% de *F1-score*.

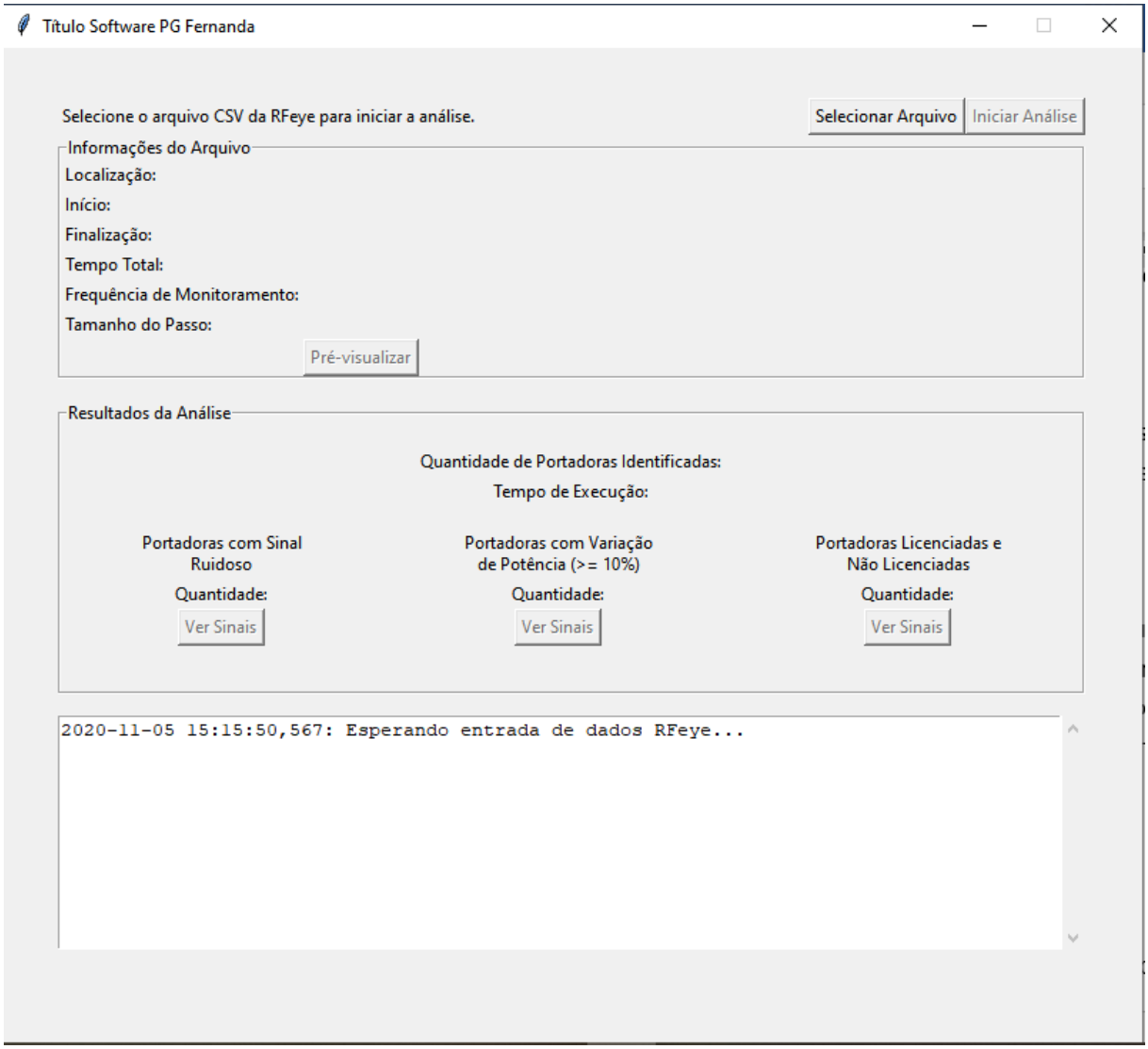
4.3 Interface

Com a metodologia de análise de rádio clandestina, variação de potência e com o parâmetro de aprendizado de máquina definido, pode-se criar uma *interface* para o usuário identificar as anomalias que o espectro de frequência pode apresentar. É importante salientar que os nomes, eixos de frequência e as frequências centrais das rádios identificadas serão *hachurados* com o intuito de preservar a identidade de cada entidade.

Para confecção da *interface* se utilizou a biblioteca "*tinker",* do python. Com ela, foi possível empregar os chamados "*widgets"*, que nesse contexto foram os *frames* (áreas retangulares), *buttons* (botões), *labels (*exibidores de texto), *canvas* (caixas de desenho) e o *scrolledtext* (caixa de texto com barra de rolagem).

A *interface* criada conta com seis botões e dois quadros. Os botões da parte superior contam com as opções de carregar arquivo e iniciar a análise. O botão central é referente a uma pré-visualização dos dados. Por fim, os três botões inferiores indicam as análises feitas na metodologia. O primeiro quadro exibe informações do arquivo carregado, e o segundo os resultados da análise, como é possível verificar na Figura 25.

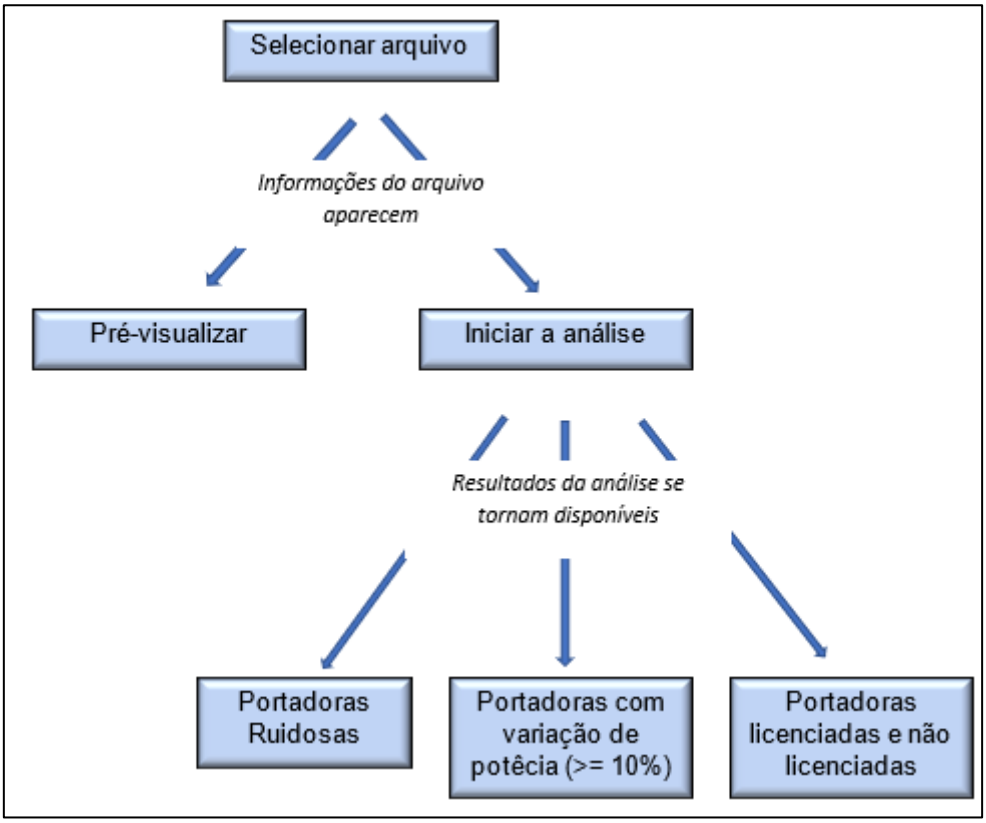
Figura 25 – *Interface* gráfica do *software*



Fonte: Produzido pela própria autora.

A sequência de funcionamento de botões pode ser observada no fluxograma da Figura 26, onde é mostrada a relação entre os botões.

Figura 26 – Sequência de botões



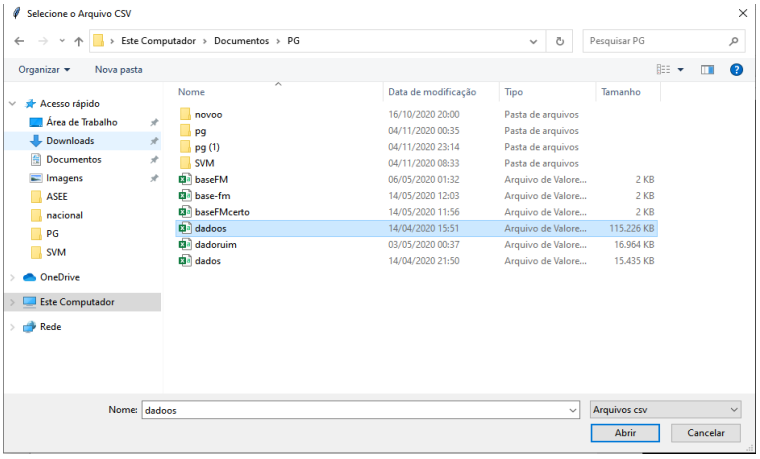
Fonte: Produzido pela própria autora.

É importante ressaltar que para o uso da *interface*, foi necessário reservar uma pasta de documentos no computador contendo um arquivo CSV com as rádios licenciadas e outro de treinamento do SVM polinomial de quarta ordem, conforme definido na seção 4.2.

4.3.1 Botão: Selecionar Arquivo

Ao clicar o botão “Selecionar o Arquivo”, é aberto para o usuário uma janela para a busca do arquivo CSV desejado, o qual se refere aos vetores de amostras que devem ser analisados, como mostrado na Figura 27.

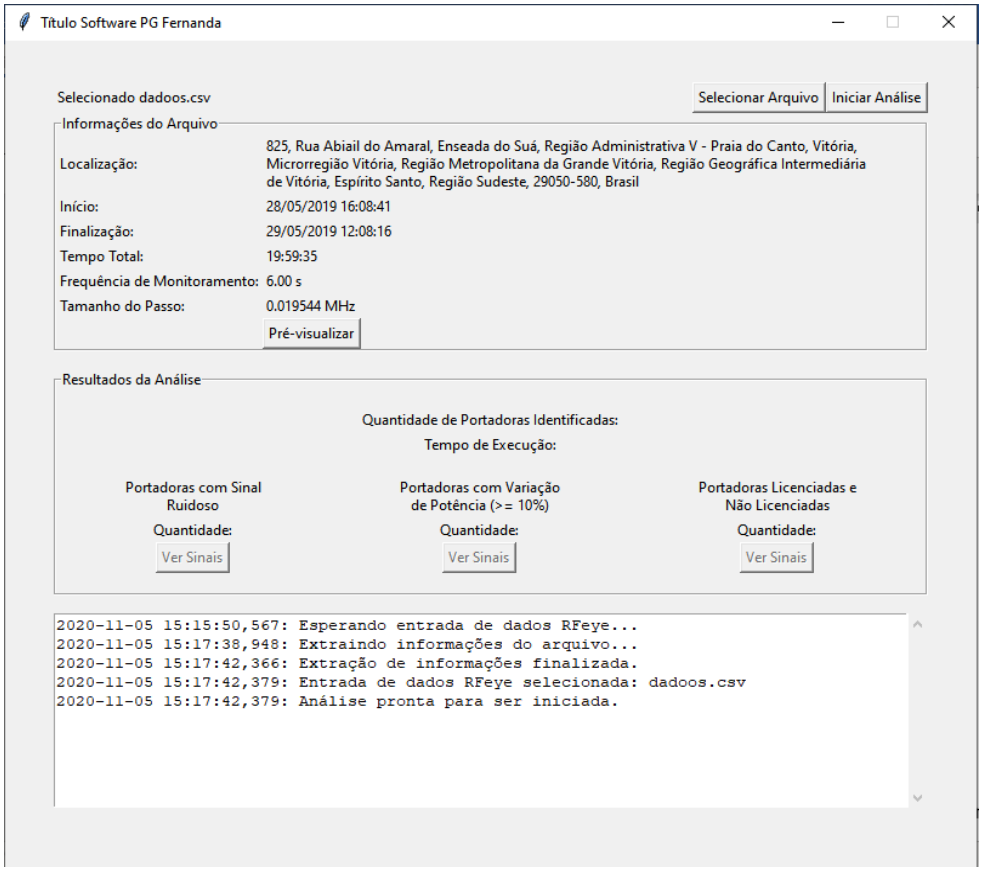
Figura 27 – Aba aberta a partir do botão selecionar arquivos



Fonte: Produzido pela própria autora.

Ao clicar no botão "*abrir*", a *interface* assumirá as atualizações presentes na Figura 28. Nesse caso, pode-se perceber que além do quadro de informações do arquivo é exibido um *log* que indica quando o botão "*iniciar a análise"* será disponibilizado.

Figura 28 – Exibição da *interface* após abrir um arquivo



Fonte: Produzido pela própria autora.

No quadro de informações do arquivo, verifica-se que o campo de localização está preenchido. Essa refere-se ao lugar onde os dados da *RFeye* foram coletados. Isso só é possível, pois nos arquivos que contêm os vetores de amostras também têm especificadas as coordenadas de longitude e latitude. Assim, com a biblioteca *"geopy",* do *python,* as coordenadas geográficas são transformadas em endereço.

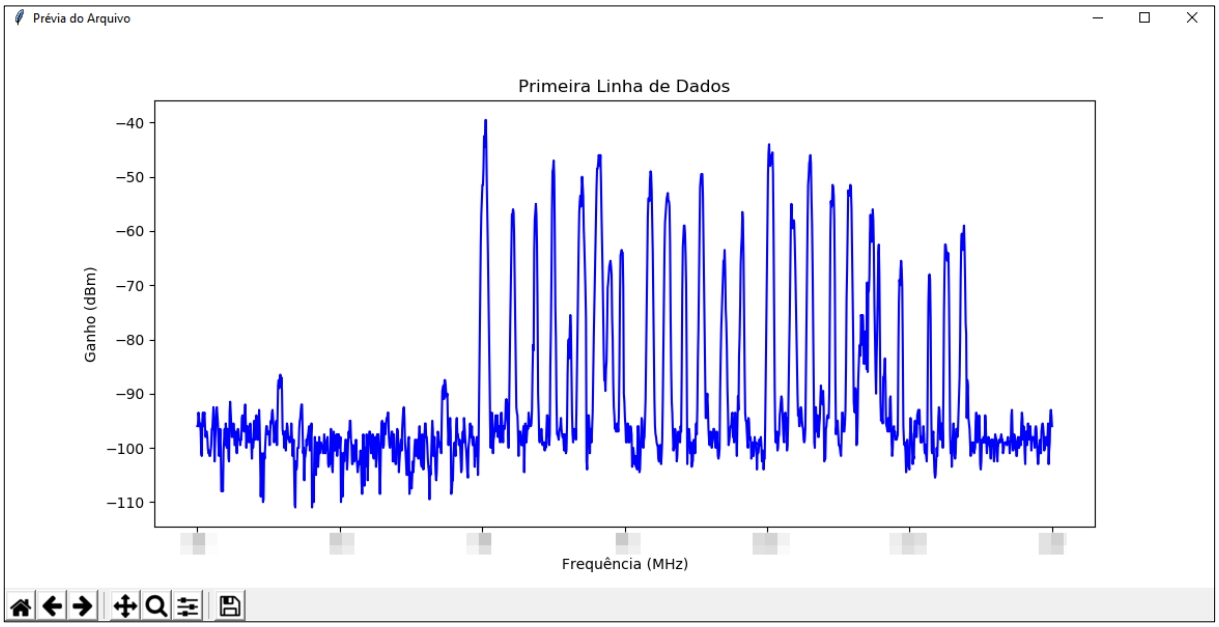
Se tratando da hora/data de início e fim da leitura, tem-se que as informações foram capturadas verificando o primeiro e último vetor de amostras do arquivo CSV selecionado, já que antes dos dados de cada vetor encontram-se as informações de data e hora. Assim, com a diferença de hora/data dos vetores de amostras, é possível obter também o período total em que foi feita a leitura. Por último, é observada a diferença no tempo de leitura entre o primeiro e segundo vetor de amostras, obtendo assim a frequência de monitoramento.

O quadro também exibe o tamanho do passo. O passo seria a taxa de amostragem do espectro de frequência.

4.3.2 Botão: Pré-visualização

O botão de "pré-visualização" serve apenas para exibir o primeiro vetor de amostras do arquivo em forma de gráfico, fazendo com que o usuário possa ter conhecimento do arquivo que ele quer analisar. A Figura 29 mostra a exibição do primeiro vetor de amostras, sem nenhum tratamento do arquivo selecionado.

Figura 29 – Exibição do botão "Pré-visualização" do arquivo que foi selecionado

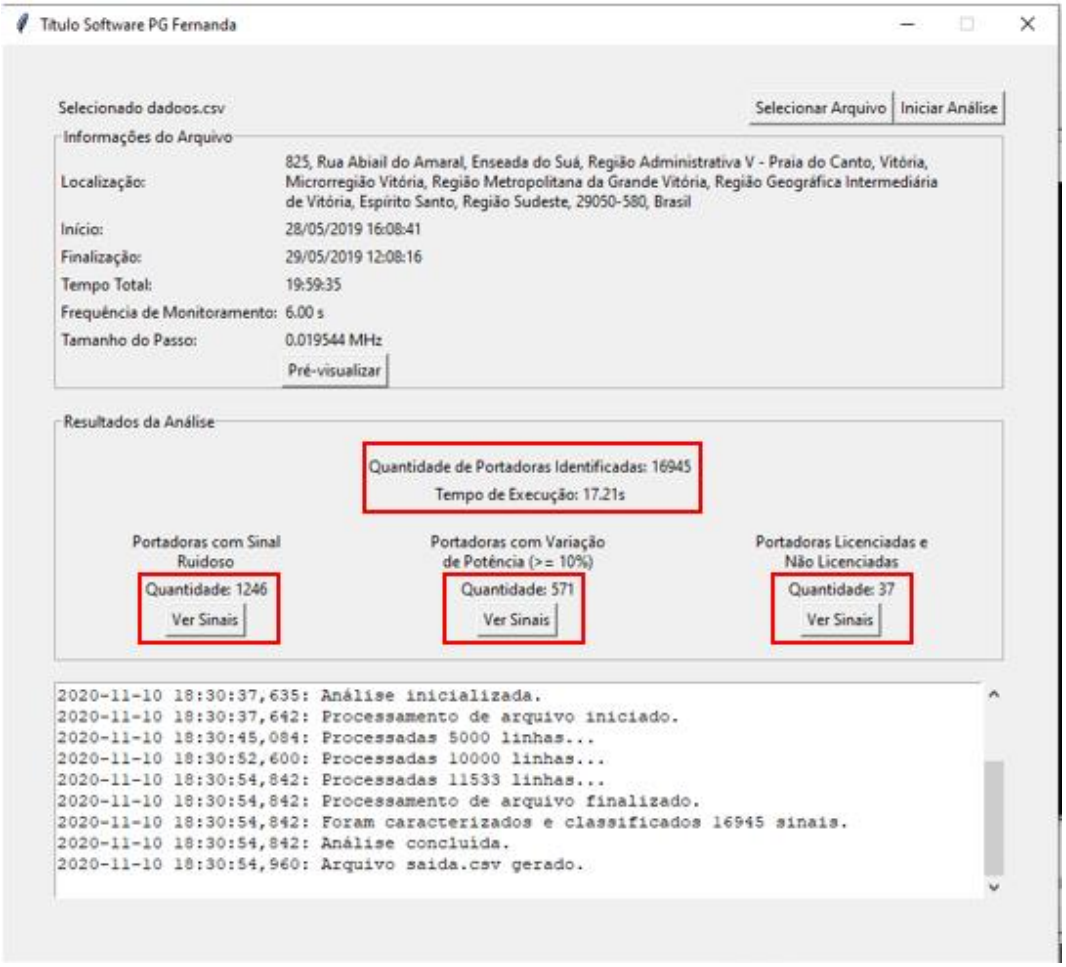


Fonte: Produzido pela própria autora.

4.3.3 Botão: Iniciar Análise

Ao clicar no botão “Iniciar Análise” toda metodologia da Figura 18 é carregada, sendo possível verificar na Figura 30 o momento em que ela começa a atuar, indicada como “Análise inicializada”, no *log*. A partir deste momento, o usuário acompanha o processamento do arquivo até a avaliação ser concluída. Após o fim do processo, os itens destacados são exibidos. No final, é gerado um arquivo chamado “saída.csv” com o formato mostrado na Figura 31.

Figura 30 – *Interface* após ativar o botão iniciar análise



Fonte: Produzido pela própria autora.

Figura 31 – Parte do CSV gerado

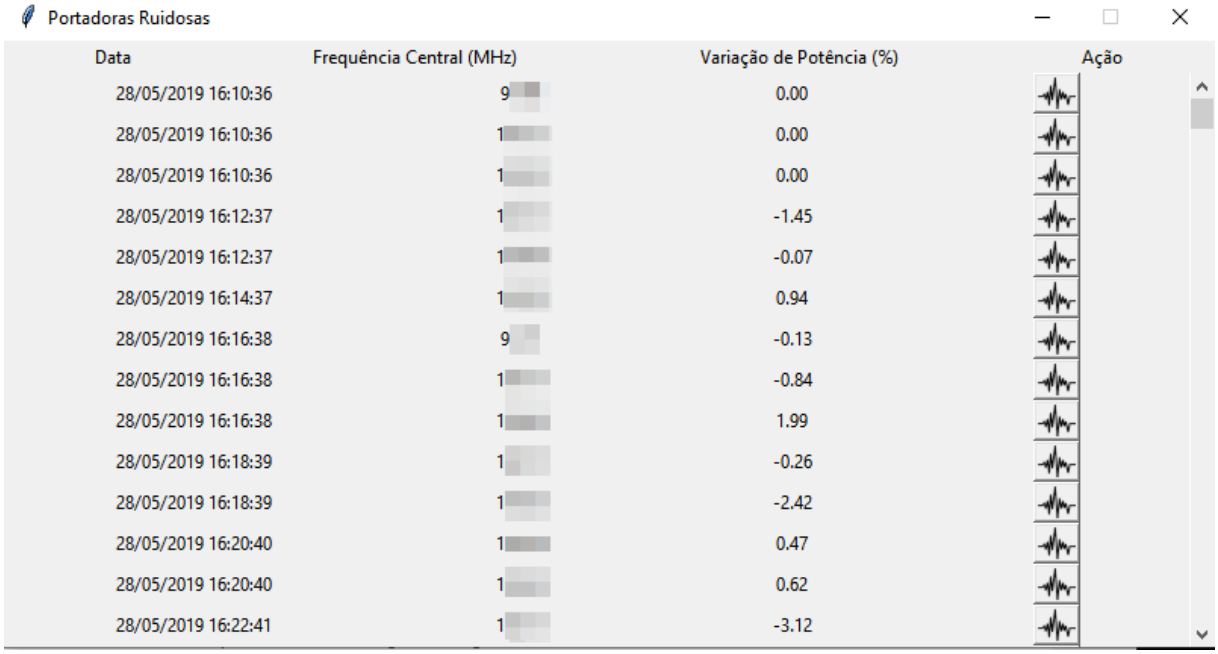
Data	Razão Social da Rádio	Freq. Cen	Variação c	Anomalia
28/05/2019 16:10	RÁDIO NÃO LICENCIADA	8	0.00	Não Licenciada
28/05/2019 16:10	[illegible]	9	0.00	Ruidosa
28/05/2019 16:10	[illegible]	1	0.00	Ruidosa
28/05/2019 16:10	RÁDIO NÃO LICENCIADA	1	0.00	Ruidosa
28/05/2019 16:10	RÁDIO NÃO LICENCIADA	1	0.00	Não Licenciada
28/05/2019 16:12	RÁDIO NÃO LICENCIADA	8	-0.54	Não Licenciada
28/05/2019 16:12	[illegible]	1	-1.45	Ruidosa
28/05/2019 16:12	RÁDIO NÃO LICENCIADA	1	-0.07	Ruidosa
28/05/2019 16:12	RÁDIO NÃO LICENCIADA	1	-0.07	Não Licenciada
28/05/2019 16:14	RÁDIO NÃO LICENCIADA	8	-0.31	Não Licenciada
28/05/2019 16:14	RÁDIO NÃO LICENCIADA	1	0.00	Não Licenciada

Fonte: Produzido pela própria autora.

4.3.4 Botão: Portadoras Ruidosas

O botão “portadoras ruidosas” é referente à metodologia da seção 3.2.3. Ao clicar neste botão, a janela referente à Figura 32 é aberta, listando todos os dados que foram guardados em memória. Nela, são exibidas a data e a hora que uma determinada anomalia ocorreu, além da frequência central e a variação de potência. A janela ainda possui mais um botão, disposto na coluna de ação.

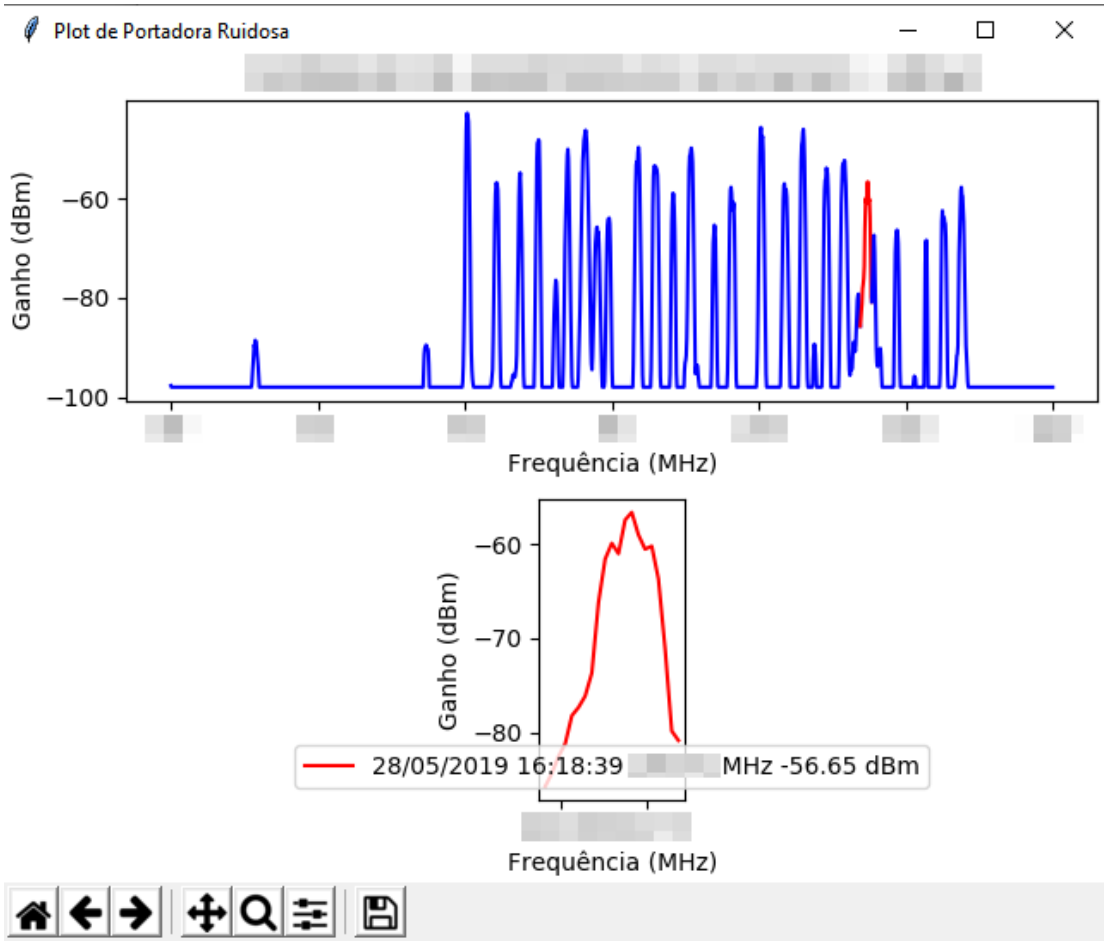
Figura 32 – Janela referente aos dados ruidosos



Fonte: Produzido pela própria autora.

Os botões referentes à coluna “*Ação*” têm a utilidade de exibir o comportamento da portadora anômala, de acordo com a data/hora em que ela foi identificada. Exibe-se assim a própria portadora e onde ela se localiza no espectro de frequência. Já que a frequência central da portadora é conhecida, é possível associar tal frequência ao banco de dados das rádios da Grande Vitória, como pode ser visto na Figura 33.

Figura 33 – Exibição do comando Ação de uma portadora ruidosa

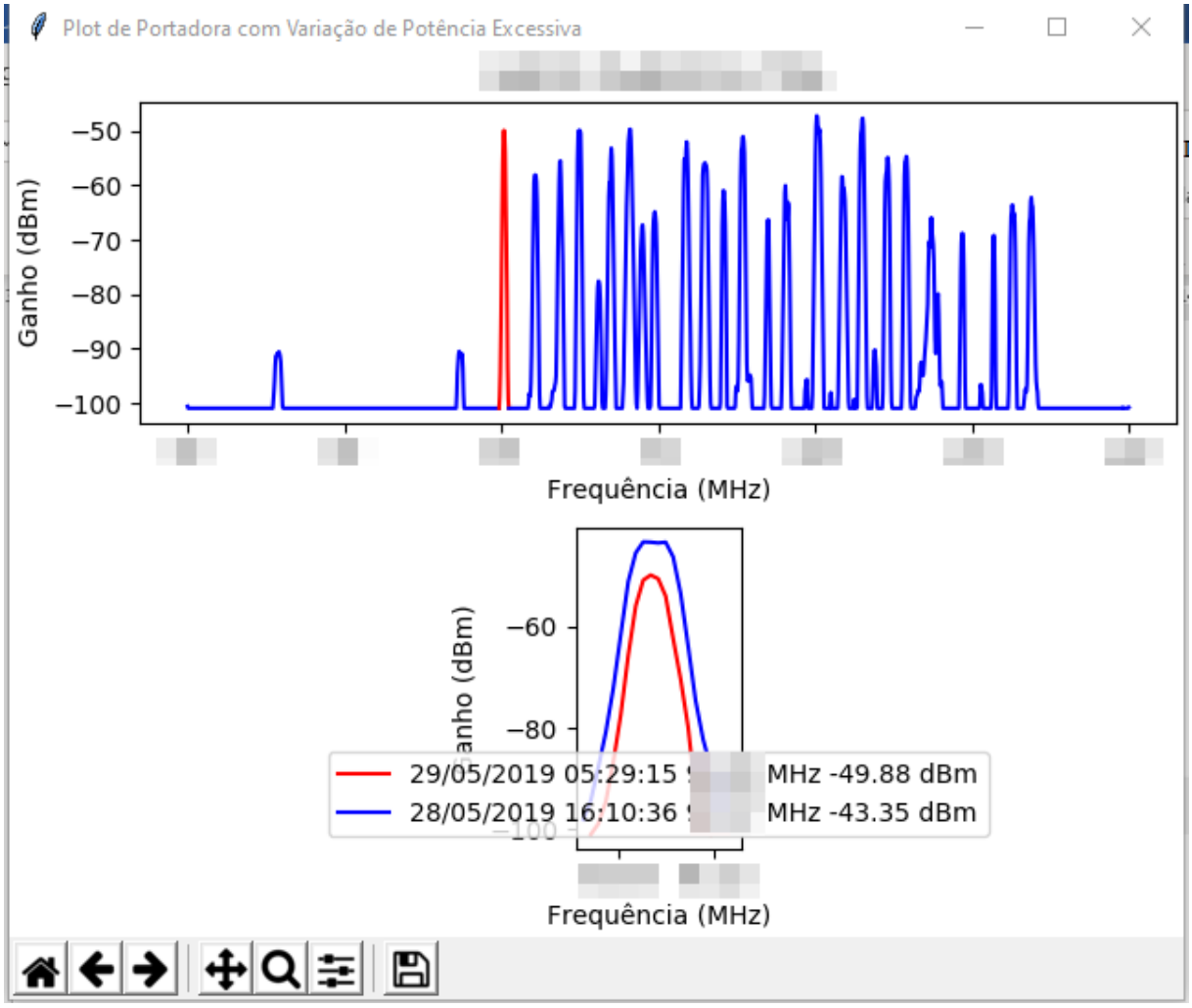


Fonte: Produzido pela própria autora.

4.3.5 Botão: Portadoras Com Variação de Potência Maior ou Igual a 10%

O botão de “portadoras com variação de potência” está relacionado à metodologia da seção 3.2.2. Ao clicar neste botão, uma janela semelhante à Figura 32 é aberta. Porém, os botões referentes a coluna de “*Ação*” exibem o comportamento das portadoras com uma variação de potência maior que 10%, apresentando a variação entre a potência da portadora de referência e a da anômala identificada. Tal exibição pode ser vista na Figura 34.

Figura 34 – Exibição do comando Ação de uma portadora com variação de potência acima do permitido



Fonte: Produzido pela própria autora.

4.3.6 Botão: Licenciada e Não Licenciada

O botão “licenciada e não licenciada” está associado à metodologia da seção 3.2.1. Ao clicar neste botão, uma janela semelhante a Figura 35 é aberta. Ela mostra o primeiro momento que tal rádio foi verificada e as frequências centrais de canais cadastrados mais próximos que se associam ao nome da emissora. O botão *“Ação”,* assim como nas demais verificações apresenta a portadora individual e onde ela se localiza no espectro de frequência, como mostrado na Figura 36.

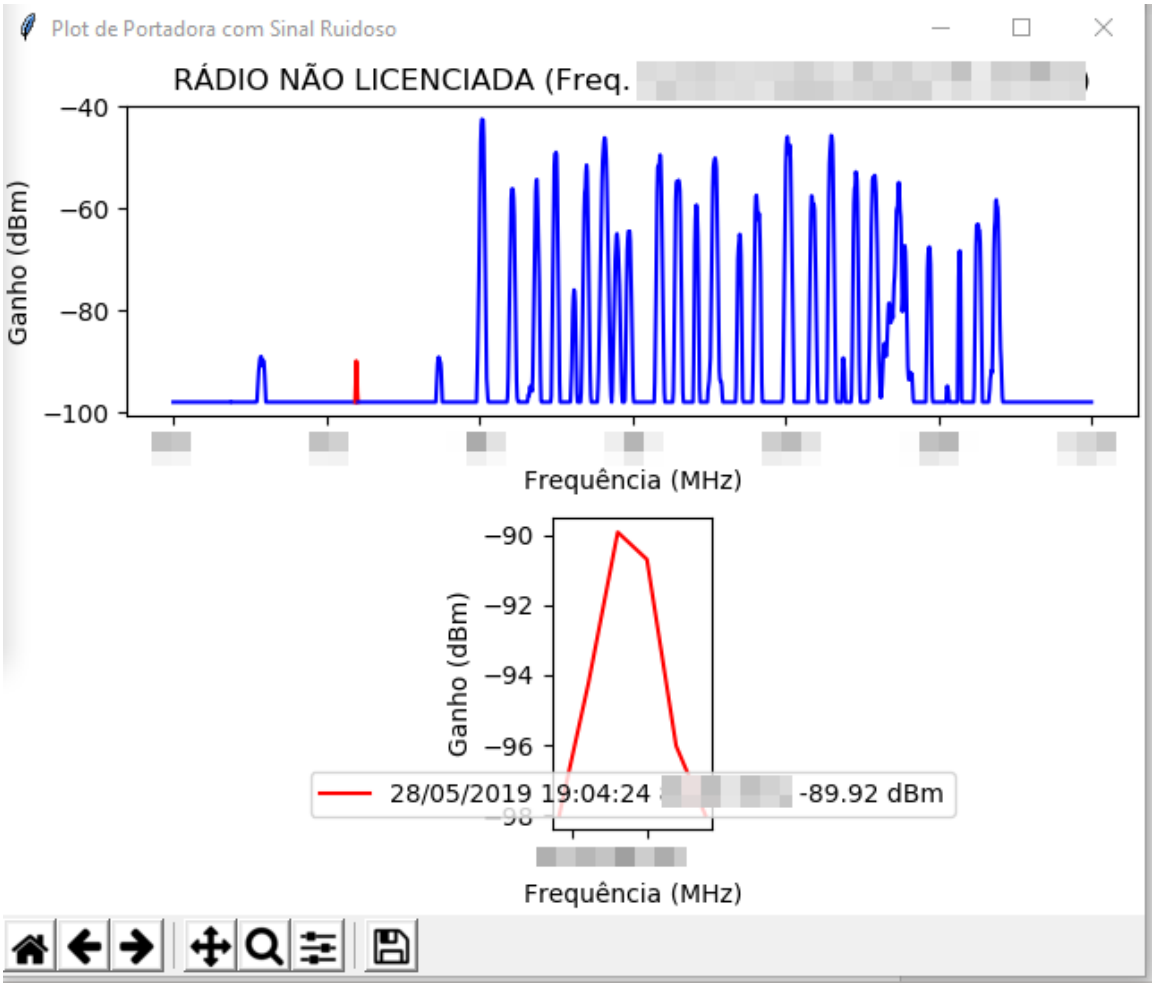
Figura 35 – Janela referente às portadoras licenciadas e não licenciadas

Portadoras Licenciadas e Não Licenciadas

Data	Freq. Central Cadastrada (MHz)	Portadora	Ação
28/05/2019 16:10:36	82.90	RÁDIO NÃO LICENCIADA	
28/05/2019 16:10:36	88.70	RADIO MARATAIZES FM LTDA	
28/05/2019 16:10:36	90.10	RADIO PLANETA LTDA	
28/05/2019 16:10:36	91.10	FUNDACAO ROMULO NEVES BALESTRERO	
28/05/2019 16:10:36	91.90	SM COMUNICACOES LTDA	
28/05/2019 16:10:36	92.50	A GAZETA DO ESPIRITO SANTO RADIO E TV LTDA	
28/05/2019 16:10:36	93.10	RADIO GAETA LTDA	
28/05/2019 16:10:36	93.50	METROPOLE FM LTDA	
28/05/2019 16:10:36	94.10	FUNDACAO EDUCATIVA MESTRE ALVARO	
28/05/2019 16:10:36	94.50	SISTEMA ALAN KARDEC DE RADIODIFUSAO LTDA	
28/05/2019 16:10:36	94.90	EMPRESA ESPIRITOSANTENSE DE RADIODIFUSAO LTDA	
28/05/2019 16:10:36	95.90	FUNDACAO ROBERTO RABELLO DE COMUNICACAO SOCIAL	

Fonte: Produzido pela própria autora.

Figura 36 – Exibição do comando Ação de uma portadora identificada



Fonte: Produzido pela própria autora.

5 CONCLUSÕES

Neste projeto foi apresentado uma maneira de automatizar um processo que é feito de forma manual na ANATEL. Assim, o objetivo geral que era avaliar técnicas de AM na detecção de anomalias das portadoras de FM, na faixa de 87,4 a 108 MHz da cidade de Vitória-ES foi cumprido.

Técnicas de pré-processamento foram aplicadas, anomalias foram detectadas e as métricas de avaliação foram verificadas. Além disso, duas verificações adicionais foram feitas, que se refere a identificação das rádios licenciadas e não licenciadas, além de uma análise da variação de potência.

Apesar do kNN apresentar bons resultados referentes às métricas de avaliação, ele teve um tempo de classificação cerca de 1000 vezes maior que o SVM, logo, se mostrou um algoritmo com alto custo computacional. Além disso, foi observado que o SVM com kernel polinomial de quarta ordem apresentou resultados melhores.

Cumprindo um objetivo específico, uma *interface* foi feita de modo a simplificar o contato dos usuários com os códigos desenvolvidos, a partir dela também um CSV foi gerado, com o objetivo de as análises serem guardadas ao longo do tempo. A *interface* ainda fornece o tempo de leitura e processamento dos dados, no qual foi visto que para quase um dia de dados coletados o tempo de execução foi cerca de 17 segundos.

A metodologia para achar a portadora ruidosa mostrou-se eficaz no sentido de identificar portadoras em mais de uma posição, apresentando robustez no método, mostrando assim que o algoritmo não estava viciado.

Destaca-se que com a automatização da tarefa é possível obter não só respostas rápidas, como também resultados que não seriam possíveis de serem alcançados por simples inspeção visual.

Para estudos futuros, com relação ao AM, pode-se treinar a máquina com mais dados, para aumentar a confiabilidade das análises geradas, além de utilizar a curva ROC

como métrica de desempenho. Outro ponto sugerido, seria o funcionamento do programa em tempo real, onde pode-se estudar a possibilidade de acoplar a *interface* diretamente aos dados coletados.

Ainda para trabalhos futuros pode-se considerar a largura de banda como aspecto importante na classificação de uma portadora, visto que também pode ser um fator relevante para classificar um dado anômalo. Por último, dado os resultados positivos dos experimentos, abre-se margem para as técnicas de AM serem aplicadas em outras faixas do espectro, por exemplo, nas faixas disponíveis para o uso do 5G, usando-as para verificar as faixas livres.

Depois de todo o estudo feito, e verificando através do estágio na ANATEL que várias categorias de atividades são realizadas pela equipe de fiscalização da UO021, a análise automatizada de alguns parâmetros do espectro de FM pode otimizar o tempo do fiscal de Vitória-ES, além de trazer uma maior confiabilidade nas análises. É importante ressaltar que apesar da automatização, o fiscal ainda é responsável por verificar as análises, sendo assim agente fundamental para inspecionar o espectro de frequência.

REFERÊNCIAS BIBLIOGRÁFICAS

AGC TELECOM. **RFeye node**. 2018. Disponível em: http://agc.com.br/produto/rfeye-node/. Acesso em: 21 out. 2019.

AGÊNCIA NACIONAL DE TELECOMUNICAÇÕES. Conselho Diretor. Consulta Pública nº 6, de 08 de março de 2018. Proposta de Reavaliação do Modelo de Gestão do Espectro. **Diário Oficial da União**: seção 01, Brasília, DF, ano 2018, n. 6, p. 12, 08 mar. 2018a.. Disponível em: https://sei.anatel.gov.br/sei/modulos/pesquisa/md_pesq_documento_consulta_externa.php?eEP-wqk1skrd8hSlk5Z3rN4EVg9uLJqrLYJw_9INcO7RYMO8DYCc72PYTkQyGS4LWeI-1eeTR-vrhKsIYBm6_u1PHtBpsA7xkPho2ugnFFPkwKpuRCiNpjc66nzBWZ9I. Acesso em: 25 set. 2019.

AGÊNCIA NACIONAL DE TELECOMUNICAÇÕES. Conselho Diretor. Resolução nº 67, de 12 de novembro de 1998. Aprova o Regulamento Técnico para Emissoras de Radiodifusão Sonora em Frequência Modulada. **Diário Oficial da União**, seção 01, Brasília, DF, ano 1998, n. 208-E, p. 28, 12 nov. 1997. Disponível em: https://www.anatel.gov.br/legislacao/resolucoes/13-1998/168-resolucao-67. Acesso em: 25 set. 2019.

AGÊNCIA NACIONAL DE TELECOMUNICAÇÕES. Conselho Diretor. Resolução nº 596, de 12 de novembro de 2012. Aprova o Regulamento de Fiscalização. **Diário Oficial da União**, seção 01, Brasília, DF, ano 2012, n. 154, p. 52, 12 de nov. 2012. Disponível em: https://www.anatel.gov.br/legislacao/resolucoes/2012/308-resolucao-596. Acesso em: 25 set. 2020.

AGÊNCIA NACIONAL DE TELECOMUNICAÇÕES. **Plano de Ação e Fiscalização**. Brasília, DF: ANATEL, 2018b. Disponível em: https://sei.anatel.gov.br/sei/modulos/pesquisa/md_pesq_documento_consulta_externa.php?eEP-wqk1skrd8hSlk5Z3rN4EVg9uLJqrLYJw_9INcO5CGGQnWfvk2CBdxKuEIYaL9NofDnAFLx4y1UpfqWUY-k_X3GNhy_ZW6umxu-4CAq-Xz6V7MS5e6Pmvi6fIHWi8. Acesso em: 28 maio 2019.

AGÊNCIA NACIONAL DE TELECOMUNICAÇÕES. **Relatório Anual 2018**. Brasília, DF: ANATEL, 2018c. Disponível em: https://sei.anatel.gov.br/sei/modulos/pesquisa/md_pesq_documento_consulta_externa.php?eEP-wqk1skrd8hSlk5Z3rN4EVg9uLJqrLYJw_9INcO5qCMSr1EMcIpgzizi6JB11ZfIGUkLQFuv-nQ2rlXU4epBnwl_RtrJMm4TFmg79Z9fViw1qMyG2njP4-8SlBgOO . Acesso em: 10 set. 2019.

AGÊNCIA NACIONAL DE TELECOMUNICAÇÕES. **Relatório De Fiscalização Nº 61**. Vitória: ANATEL, 2019. Disponível em: https://sei.anatel.gov.br/sei/modulos/pesquisa/md_pesq_documento_consulta_externa.php?eEP-wqk1skrd8hSlk5Z3rN4EVg9uLJqrLYJw_9INcO7dBWekWfSTL_OIJui-

GaSh-wxoLPvgDXbE-YpUacwu_Ws9WxhEILGzPFFcnO1xIwPVEM-eS5PIUUcqG7nnb5CL. Acesso em: 19 dez. 2020.

ANY CONSULTING. **O que é machine learning e por que será a tecnologia mais importante de 2018**. 2018. Disponível em: https://www.anyconsulting.com.br/o-que-e-machine-learning/. Acesso em: 15 out. 2019.

BAIA, C. **Aprendizado de Máquinas.** 2016. Disponível em: http://carlosbaia.com/2016/04/22/machine-learning/. Acesso em: 20 out. 2019.

BRASIL. **Lei nº 9.472, de 16 de julho de 1997.** Dispõe sobre a organização dos serviços de telecomunicações, a criação e funcionamento de um órgão regulador e outros aspectos institucionais. Brasília, DF: Congresso Nacional, [1997]. Disponível em: http://www.planalto.gov.br/ccivil_03/LEIS/L9472.htm. Acesso em: 25 set. 2019.

CARLSON, A. B. **Communication Systems:** An Introduction to Signals and Noise in Electrical Communication. 1. ed. Nova York: McGraw-Hill, 1986 apud CARVALHO, M. R. **Comunicações Analógicas e Digitais**. 1. ed. Vitória: [*s.n*], 2008.

CARVALHO, M. R. **Comunicações Analógicas e Digitais**. 1. ed. Vitória: [*s.n*], 2008.

FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. L. F. **Inteligência Artificial**: Uma Abordagem de Aprendizado de Máquina. 1. ed. Rio de Janeiro: LTC, 2017.

HAN, J.; KAMBER, M. **Data Mining:** Concepts and Techniques. 1. ed. Massachusetts: Morgan Kaufmann, 2000 apud FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. L. F. **Inteligência Artificial**: Uma Abordagem de Aprendizado de Máquina. 1. ed. Rio de Janeiro: LTC, 2017.

HAYKIN, S.; MOHER, M. **Sistemas de Comunicação**. 5. ed. Porto Alegre: Bookman, 2011.

HEBRICH, R. **Learning Kernel Classifiers:** Theory and Algorithms. 1 ed. Cambridge: MIT Press, 2001 apud FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. L. F. **Inteligência Artificial**: Uma Abordagem de Aprendizado de Máquina. 1. ed. Rio de Janeiro: LTC, 2017.

LUZ, F. Algoritmo KNN para Classificação. **Inferir**, Brasília, DF, fev. 2019. Seção Artigos. Disponível em: https://inferir.com.br/artigos/algoritimo-knn-para-classificacao/. Acesso em: 20 set. 2020.

NATIONAL TAIWAN UNIVERSITY. **A Practical Guide to Support Vector Classification.** 2003. Disponível em: https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf. Acesso em: 29 nov. 2020.

OPPENHEIM, A. V; SCHAFER, R. W; BUCK, J. R. **Discrete-Time Signal Processing**. 2. ed. Nova Jersey: Prentice Hall, 1999.

QSL. **Modulação angular**: FM e PM. 2007. Disponível em: https://www.qsl.net/py4zbz/teoria/fmpm.htm. Acesso em: 30 set. 2019.

RODRIGUES, W. **Modulação em Frequência**. 2004. Disponível em: http://www.jdbte.com.br/wjrteleco/unit%2004.pdf. Acesso em 30 set. 2019. (Apostila)

ROMAIS, C. **O Que é** – Rádio Em Ondas Curtas. 1. ed. Brasília: Brasiliense, 1994.

REIS, E. E.; COSTA, E. P. **Atividades de Fiscalização:** Radiointerferência (Radiodifusão). Brasília, DF, 2004. (Apostila)

SANTOS, E. M. **Teoria e Aplicação de Support Vector Machines à Aprendizagem e Reconhecimento de Objetos Baseado na Aparência.** 2002. Dissertação (Mestrado em Informática) – Programa de Pós-Graduação em Informática, Universidade Federal da Paraíba, Campina Grande, 2002. Disponível em: http://docs.computacao.ufcg.edu.br/posgraduacao/dissertacoes/2002/Dissertacao_EulandaMirandadosSantos.pdf. Acesso em: 20 set. 2020.

SAS. **Inteligência Artificial:** O que é? Qual importância?. 2019. Disponível em: https://www.sas.com/pt_br/insights/analytics/inteligencia-artificial.html. Acesso em: 20 out. 2019.

SUNASRA, M. Performance Metrics for Classification problems in Machine Learning. **Medium**, Mumbai, nov. 2017. Seção MohammedS. Disponível em: https://medium.com/@MohammedS/performance-metrics-for-classification-problems-in-machine-learning-part-i-b085d432082b. Acesso em: 20 set. 2020.

TELECOM & ITSolutions. **Instrumentação e Medidas em Telecomunicações**. São Paulo, 2000. (Apostila)

TESTAND & MEASUREMENT. **Are you sure you know what bandwidth means?**. 2020. Disponível em: https://www.testandmeasurementtips.com/are-you-sure-you-know-what-bandwidth-means-faq/. Acesso em: 28 nov. 2020.

TOWARDS DATA SCIENCE. **Support Vector Machine – Introduction to Machine Learning Algorithms**. 2018. Disponível em: https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47. Acesso em: 15 out. 2020.

UNIVERSIDADE DE BRASILIA. **Tutorial de SVM.** 2017. Disponível em: https://lamfo-unb.github.io/2017/07/13/svm/. Acesso em: 20 set. 2020.

UNIVERSIDADE DO MINHO. **Alguns aspectos que condicionam o desempenho dos sistemas de transmissão.** 1999. Disponível em: http://www3.dsi.uminho.pt/adriano/Teaching/Comum/FactDegrad.html. Acesso em: 19 dez. 2020.

APÊNDICE A – LISTA DAS RÁDIOS DA GRANDE VITÓRIA

88,7, RADIO MARATAIZES FM LTDA, Itapemirim
90,1, RADIO PLANETA LTDA, Serra
90,5, RADIO VITORIA S/A, Vitória
91,1, FUNDACAO ROMULO NEVES BALESTRERO, Vitória
91,7, SOCIEDADE ALFREDENSE DE RADIODIFUSAO LTDA, Alfredo Chaves
91,9, SM COMUNICACOES LTDA, Vila Velha
92,5, A GAZETA DO ESPIRITO SANTO RADIO E TV LTDA, Vitória
93,1, RADIO GAETA LTDA, Guarapari
93,5, METROPOLE FM LTDA, Vila Velha
94,1, FUNDACAO EDUCATIVA MESTRE ALVARO, Viana
94,5, SISTEMA ALAN KARDEC DE RADIODIFUSAO LTDA, Santa Leopoldina
94,9, EMPRESA ESPIRITOSANTENSE DE RADIODIFUSAO LTDA, Guarapari
95,9, FUNDACAO ROBERTO RABELLO DE COMUNICACAO SOCIAL, Vila Velha
96,5, MEAIPE EMPRESA DE COMUNICACOES LETDA, Guarapari
97,1, A GAZETA DO ESPIRITO SANTO RADIO E TV LTDA, Serra
97,7, FUND. UNIVERSIT. DE PESQ. ECONOM. E SOCIAIS DE V. VELHA, Serra
98,5, ASSOCIACAO DE MORADORES PRAIA DO CANTO, Vitória
99,1, NASSAU EDITORA RADIO E TV LTDA, Vitória
99,5, NASSAU EDITORA RADIO E TV LTDA, Vitória
100,1, RADIO ASTRAL, Aracruz
100,3, CAMARA DOS DEPUTADOS, Guarapari
100,9, FUNDACAO NOVO MILENIO, Guarapari
101,5, FUNDACAO NOSSA SENHORA DA PENHA DO ESPIRITO SANTO, Vitória
101,9, SISTEMA ARACRUZ DE RADIODIFUSAO LTDA, Aracruz
102,3, RADIO FM 102 LTDA, Cariacica
102,7, SUPER RADIO DM LTDA, Ibiraçu
102,9, FUNDACAO EDUCATIVA E CULTURAL DONA DADA, Vila Velha
103,7, FUNDACAO DE ASSISTENCIA E EDUCACAO - FAESA, Cariacica
104,7, FUNDACAO CECILIANO ABEL ALMEIDA, Vitória
105,1, RADIO CAPIXABA LTDA, Vitória
105,3, CRISTO REI COMUNICACOES LTDA, Santa Teresa
105,7, SUPER RADIO DM LTDA, Domingos Martins
106,3, SISTEMA NORTE DE RADIO LTDA, Serra
106,9, FUNDACAO EDUCATIVA E CULTURAL PEDRO TRES, Vila Velha

107,3, FUNDACAO CULTURAL SANTA MARIA DE DEUS, Ibiraçu
107,9, PREFEITURA MUNICIPAL DE CARIACICA, Cariacica