INSTITUTO FEDERAL DE SANTA CATARINA

MÁRIO ANDRÉ LEHMKUHL DE ABREU

Desempenho de técnicas de Aprendizado de Máquina em Reconhecimento de Voz

São José - SC

Dezembro/2019

RESUMO

Com os avanços tecnológico ocorrido nos últimos anos o mundo vem passando por transformações, sendo notável cada vez mais a interação dos seres humanos com aplicações computacionais. Assistentes pessoais ativados por voz, carros autônomos, reconhecimento facial, marcação automática em fotos de redes sociais entre outros já são uma realidade. Para a interação com essa aplicações, uma tecnologia que vem ganhando destaque na ultima década vem sendo usada, o Aprendizado de Maquina (Machine Learning). Esta tecnologia é uma área da inteligencia artificial (IA) que investiga como as máquinas podem aprender através da extração de padrões a partir de um conjunto de dados. A partir disso varias aplicações podem ser desenvolvidas em vários cenários diferentes. Uma que vem ganhando bastante presença no cotidiano da sociedade, é o reconhecimento de voz. Seu uso reflete em mais eficiência, na realização de ações, não sendo necessário utilizar as mãos. Além disso, podem dar maior acessibilidade às pessoas com limitações motoras. Desse modo esse trabalho tem como objetivo analisar e comparar o desempenho de várias técnicas de aprendizado de máquina úteis no reconhecimento de voz. Para servir como uma contribuição de informação, e verificar qual técnica apresenta o melhor resultado de uso, no desenvolvimento de sistema com reconhecimento de voz. Todas as análises serão realizadas em uma base de dados de vozes selecionadas. Os resultados obtidos serão avaliados em termos de eficiência, precisão (acurácia), sensibilidade e especificidade.

Palavras-chave: Reconhecimento de voz, Inteligencia artificial, Técnicas de aprendizado de máquina.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

SUMÁRIO

1 INTRODUÇÃO

Com os avanços tecnológico ocorrido nos últimos anos o mundo vem passando por transformações, sendo notável cada vez mais a interação dos seres humanos com aplicações computacionais. Assistentes pessoais ativados por voz, carros autônomos, reconhecimento facial, marcação automática em fotos de redes sociais entre outros já são uma realidade. Para a interação com essa aplicações, uma tecnologia que vem ganhando destaque na ultima década vem sendo usada, o Aprendizado de Maquina (Machine Learning)(DATA SCIENCE ACADEMY, 2019a).

O aprendizado de Máquina é uma área de pesquisa da Inteligência Computacional que estuda o desenvolvimento de métodos capazes de extrair conhecimento a partir de amostras de dados.(LORENA; CARVALHO, 2003) Em uma definição fundamental, representa o uso de algoritmos para extrair informações de dados brutos e representá-los por meio de algum tipo de modelo matemático. Com esse modelo conclusões são feitas a partir de outros conjuntos de dados. Para esse processo existem vários algoritmos disponíveis (DATA SCIENCE ACADEMY, 2019a).

No entanto aprender é uma ação muito elaborada e complexa. No Aprendizado de Máquina, fazer isso requer a extração de um padrão para que o mesmo seja primeiramente reconhecido. Isso visa detectar automaticamente regularidades dentro de um conjunto de dados, possibilitando que a máquina possa tomar decisões, como a seleção de dados em diferentes categorias ou classes. Essa classificação se baseia em conhecimentos preliminares ou dedutivos, e em informações estatísticas conseguidas a partir de referências. Para avaliar o resultados obtidos, termos de eficiência, precisão (acurácia), sensibilidade e especificidade são usados.(ARRIGONI, 2018).

A partir disso varias aplicações podem ser desenvolvidas em vários cenários diferentes, como na medicina e na saúde, transportes, agronegócio, comercio em geral, serviços públicos, cidades inteligentes e na indústria, com o conceito de Indústria 4.0, um termo recente que explica a aplicação das novas tecnologias nos principais processos industriais (BARBOSA, 2017). Segundo o Gartner (empresa que atua com consultorias e prospecções no mercado de TI), até 2020 todos os softwares corporativos terão alguma funcionalidade ligada ao Aprendizado de Máquina (DATA SCIENCE ACADEMY, 2019a).

Diante desses cenários, um que vem ganhando bastante presença no cotidiano da sociedade, é o reconhecimento de voz. Seu uso reflete em mais eficiência, na realização de ações, não sendo necessário utilizar as mãos (TRINDADE, 2018). Além disso, podem dar maior acessibilidade às pessoas com limitações motoras (ANDRADE et al., 2016). Computadores, smartphones e sistemas multimídias de carros, são alguns exemplos em que o reconhecimento de voz esta inserido. No caso dos computadores e smartphones, as aplicações do momento são os assistentes virtuais como, Siri (Apple), assistente Google, Cortana (Microsoft) e Amazon Echo com Alexa, tendo como objetivo atender seus usuários (TRINDADE, 2018).

Porém trabalhar com a voz não é uma tarefa fácil. O sinal de voz produzido pelo sistema vocal humano é muito complexo, resultando em uma difícil tarefa de caracterizá-lo. Inúmeros modelos matemáticos sofisticados existem que tentam simular a produção da voz humana, entretanto, sua capacidade de modelamento ainda é muito limitada.

Pelo fato de o sinal de voz carregar muita informação, não é possível sintetizar e analisá-lo diretamente. Desse modo, para desenvolver sistemas de reconhecimento de voz, é usado a técnica de sistemas modernos de Reconhecimento Automático de Voz (RAVs). Eles são responsáveis por produzir uma representação textual do sinal de voz. Para isso utilizam a combinação de algoritmos de extração

e reconhecimento de características. Com seu uso muitas aplicações podem ser desenvolvidas como, comandos de voz, diálogos interativos, transcrição de discursos gravado, ditado e pesquisa em documentos de áudio.

Os algoritmos de extração de características são usados para remover do sinal qualquer informação que não seja útil para o reconhecimento, ressaltando aspectos que contribuem para a identificação de diferenças. Dessa extração, vetores de características são gerados. Em seu uso tem-se disponível algumas técnicas para o modelamento acústico como, MFCC (Mel-Frequency Cepstral Coefficients), LPC (Linear Predictive Coding) e PLP (Perceptual Linear Prediction Coefficients).

No que diz respeito ao reconhecimento de características, etapa que classifica e reconhece a informação do sinal analisado, tem-se disponivel técnicas como HMM (Hidden Markov Models - Modelo oculto de Markov), modelamento linguístico, e técnicas de aprendizado de máquina como RNA (ou ANN, Artificial Neural Networks), SVM (Support Vector Machines - Máquina de vetores de suporte), Random F, Naive Bayes, arvore de decisão, KNN (K-Nearest Neighbors) e mais recentemente, DNN (Deep Neural Networks)(THIAGO, 2017).

1.1 Motivação

O uso de sistemas com reconhecimento de voz já é uma realidade em nossa sociedade. Seu objetivo tem por finalidade trazer mais eficiência na realização de ações, não sendo necessário utilizar as mãos. Além disso, por meio desse cenário permite dar maior acessibilidade à pessoas com limitações motoras. Computadores, smartphones e sistemas multimídias de carros, são alguns exemplos em que o reconhecimento de voz esta inserido, e seu uso é visto constantemente no cotidiano. No entanto o reconhecimento de voz ainda não ocorre de forma natural, havendo em alguns momentos a não realização de ações, devido a uma interpretação errada da fala, gerado por ruídos externos, entre outros fatores. A partir disso, este trabalho tem como objetivo apresentar um estudo sobre técnicas de reconhecimento e classificação de voz, para uso em sistemas de reconhecimento de voz, de modo a analisar qual técnica apresenta o melhor desempenho.

1.2 Objetivo geral

Estudar técnicas de aprendizado de máquina para reconhecimento e classificação de voz, com o propósito de uso em sistema com reconhecimento de voz.

1.3 Objetivos específicos

- Estudar e aplicar técnicas e algoritmos de aprendizado de máquina.
- Analisar o desempenho das técnicas a partir de um conjunto de dados pré-definido.
- Determinar, através dos resultados de testes, as técnicas que apresentem melhor desempenho de acordo com o objetivo geral do trabalho.

1.4 Organização do texto

O texto está organizado da seguinte forma: No Capítulo 2 é apresentado a fundamentação teórica onde serão discutidos conceitos e ferramentas que serão necessários para a implementação desse trabalho. No Capítulo 3 é apresentada a proposta, que descreve como vai ser executada a metodologia para o desenvolvimento do trabalho.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Característica da fala

Um fenômeno que existe desde o nascimento do ser humano é a sua voz, e se apresenta de várias maneiras, tais como choro, grito, risos e sons de fala. Ela é um dos meios de comunicação do indivíduo com o exterior e particularmente com seus semelhantes (GABANINI, 2003).

A fala é produzida pelo aparelho fonador, que é responsável pela produção da voz, mas também por funções vitais relacionadas a respiração e a alimentação. Os órgãos que compõem o aparelho fonador são a boca, os pulmões, a traqueia e a laringe, estes são controlados pelo sistema nervoso, que os adapta para moldar os sopros por eles emitidos por meio dos lábios, língua, dentes e palato gerando a fala.

Apesar de a maior parte das ondas sonoras saírem da boca, o som também é emitido das narinas, garganta e bochechas. Em seu funcionamento, o mecanismo de produção da fala, ocorre quando a pressão do ar contido nos pulmões segue pela traqueia até a laringe e chega a glote, que é o orifício entre as cordas vocais. Durante a respiração, na maiorias das vezes a glote esta aberta, permitindo a passagem de ar. Quando as cordas vocais são estiradas, a glote é temporariamente fechada, não permitindo a passagem de ar, modulando o ar em pulsos discretos. O trato vocal que começa na glote e termina nos lábios é um tubo acústico variável e não uniforme, que tem a função de agir como um tubo ressonante com a finalidade de filtrar o conjunto de pulsos produzidos.

No trato vocal um dos componente principais que o compõem são os articuladores, que representam as partes moveis como maxilar, véu palatino, língua e lábios. Sendo o véu palatino uma peça fundamental, atuando na abertura e fechamento da cavidade nasal, controlando o som produzido pelas narinas

Nas cordas vocais o movimento oscilatório acontece eu uma frequência fundamental, denomina *pitch*, que varia durante a fala e esta relacionada com a magnitude das variações de pressão sobre a glote, da tensão nas cordas vocais e da massa das bordas vibrantes, resultando na entonação da pronuncia. As frequências produzidas, podem ser variadas, sendo para os homens de 80 a 150 Hz, para as mulheres de 150 a 250 HZ e nas crianças acima de 250 Hz.

A produção da onda sonora que é transmitida para fora do trato vocal pelos lábios e narinas, geram sons que são classificados como vocálicos e não-vocálicos. Os sons vocálicos são gerados quando a passagem do ar no trato vocal acontece continuamente e sem turbulência, o qual é representado pelas vogais. Já os sons não-vocálicos são gerados quando o trato vocal impõem resistência na passagem do ar, neste casso são representados por certas consoantes classificadas como explosivas e fricativas (VALIATI, 2000).

2.2 Sistemas de reconhecimento de voz

Uma área de pesquisa que vem sendo estudada nas últimas décadas, e tem sido a ambição de muitos pesquisadores é o desenvolvimento de sistemas para comunicação do homem com máquinas. Porém a ambição de equipar uma máquina com a característica de entender e falar naturalmente ainda esta um pouco distante (CIPRIANO, 2001). Apesar das várias pesquisas feitas, a diversidade de ambientes e locutores, ainda reflete um obstáculo no avanço do desenvolvimento de um sistema genérico e funcional (THIAGO, 2017). No entanto, muitos avanços já foram realizados desde então (MULATINHO, 2011).

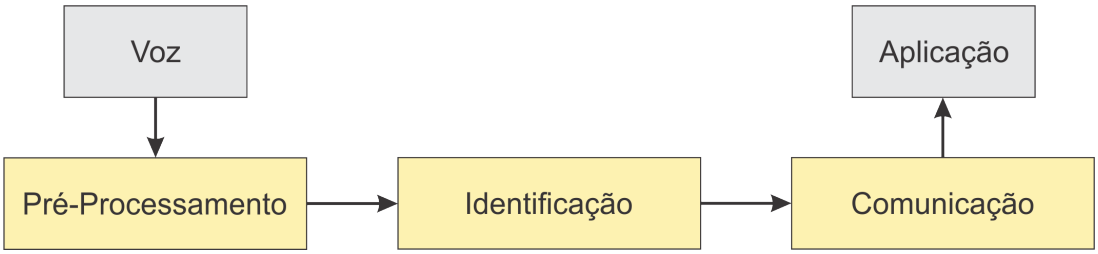
Em 1930, Homer Dudley, nos laboratórios Bell, propôs um dos primeiros modelos para análise e síntese de voz. Porém só em 1952, que Davis e Balashek, nos mesmos laboratórios Bell, desenvolveram um sistema para reconhecimento de dígitos isolados (números de zero a nove) para um único locutor (THIAGO, 2017). O sistema tinha como objetivo medir o espectro de ressonância nas regiões das vogais em cada número falado pelo locutor (MULATINHO, 2011).

No entanto só na década de 60 e 70 que as pesquisas na área realmente mostraram resultados. Em 1964 nos laboratórios da RCA (Radio Corporation of America), uma equipe liderada por Martin e seus colegas, desenvolveu um conjunto de técnicas de normalização para identificar o início e o fim da fala (THIAGO, 2017). Além de desenvolver o método, Martin fundou uma das primeiras empresas que criava e vendia soluções em reconhecimento automático de voz, a "Threshold Technology"(MULATINHO, 2011).Em 1975, tambem nos laboratórios Bell, Fumitada Itakura publicou pesquisas que obtiveram grandes progressos na área de identificação de características, como por exemplo, o LPC (Linear Predictive Coding).

A partir dos avanços obtidos desses estudos, sistemas de reconhecimento de voz, tiveram com o tempo uma mudança em sua abordagem, de uma máquina que responde a um pequeno numero de sons a um sistema sofisticado que pode responder a linguagem natural falada fluentemente (THIAGO, 2017).

Um sistema de reconhecimento automático de voz (RAV), tem como objetivo transformar um sinal de voz em uma sequencia de dados, servindo de informação para uma máquina tomar decisões. Em sua execução três tarefas fundamentais são realizadas como mostrado na figura 1.

Figura 1 – Etapas de um Sistema de reconhecimento automático de voz (RAV).



FONTE: PRÓPRIO AUTOR.

Neste sistema sua execução ocorre em três etapas: Pré-processamento, Identificação e Comunicação.

1. **Pré-processamento:** Ocorre a conversão A/D (Analógico/ digital), filtragem e extração dos parâmetros acústicos.
2. **Identificação:** É feito o reconhecimento da informação baseado em representações existentes dos padrões observados.
3. **Comunicação:** Os resultados são enviados para o ator externo que fará o uso do mesmo para tomar decisões.

Sobre os aspectos computacionais, a etapa de identificação, tem recebido grande interesse de pesquisa nas técnicas usadas. Em seu atual uso, se tem utilizado diferentes técnicas como modelos estatísticos e aprendizado de maquina .

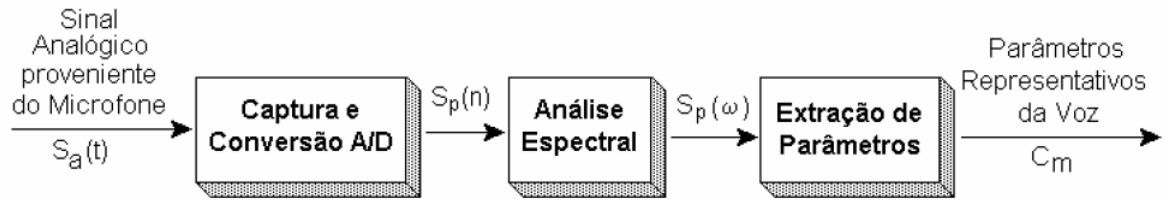
Um ponto a se analisar no uso do sistema RAV, é que alguns problemas podem aparecer no reconhecimento da voz, sendo eles:

- Na fala com palavras isoladas, no qual as palavras estão separadas por uma pausa, a tarefa é considerada simples. Em fala continua, em que os limites das palavras não são facilmente perceptíveis, ocorre confusão entre palavras e frases, o espaço de busca é maior e ocorre coarticulação, isto é, os fonemas e palavras produzidos são prejudicados pelos fonemas e palavras vizinhas.
- Se o vocabulário utilizado for grande, isso requer uma alta capacidade de armazenamento e alto custo computacional. Como solução, indica-se o uso de modelos de sub-palavras (sílabas, fonemas ou grupo de fonemas) no lugar do modelo de palavras.
- Se no reconhecimento estiver atribuído o reconhecimento do locutor, o dependente do locutor apresenta ótima precisão, porem necessita de treinamento para cada novo usuário, o que pode não ser desejável em determinadas aplicações. No independente do locutor a precisão é menor, devido os parâmetros característicos da voz dependerem do usuário (sexo, idade, sotaque, velocidade da fala, etc). No entanto não é necessário o treinamento para acada usuário, pois a voz do usuário é reconhecida sem treinamento.
- Ruído, distorções e tipo, direcionamento e posição do microfone, afetam o seu desempenho (CIPRIANO, 2001).

2.3 Etapas de extração de características e pré processamento

A primeira etapa de um sistema de reconhecimento automático de voz (RAV) é o pré-processamento. Nela é feita a conversão do sinal de analógico para digital, a análise espectral e a extração das características acústicas, como mostrado na figura 2.

Figura 2 – Etapas do Pré-Processamento.



FONTE: (CIPRIANO, 2001).

O objetivo da etapa de pré-processamento do sinal é estabelecer um conjunto de parâmetros que contenha informações uteis, para serem usadas na etapa de identificação.

2.3.1 Conversão A/D do sinal de voz

Os sinais analógicos são aqueles em que sua amplitude varia continuamente com o tempo, isto é, ela pode assumir qualquer valor pertencente a um intervalo contínuo de valores. Já um sinal digital apresenta amplitudes dentro de um conjunto de valores finito que varia de forma discreta com o tempo. Trabalhar com dados digitais é mais indicado pois o seu processamento é mais eficiente e confiável que os dados analógico (FERNANDES; PANAZIO, 2009).

A conversão analógico para digital ocorre da amostragem do sinal analógico $S_a(t)$, no período T segundos e da quantização das amostras para resultar no sinal digital $S(n) = S_a(n.T)$, onde $n = 0, 1, 2, ...$. Para que a informação do sinal se mantenha, o critério de *Nyquist* deve ser seguido, que diz que que a frequência de amostragem do sinal dever ser maior ou igual a duas vezes a sua frequência máxima

($f_s >= 2.f_{max}$). Para não ocorre a sobreposição do espectro de $S_a(t)$, impossibilitando a recuperação fiel do sinal original (efeito aliassing), o sinal de voz dever ser primeiramente passado em um filtro analógico passa-baixas com frequência de corte menor ou igual a metade da frequência de amostragem ($f_c <= f_s/2$), eliminando frequência com valores maiores que a metade da frequência de amostragem. Concluído esse processo o sinal é passado em um conversor analógico para digital.

Com o sinal convertido para digital, ele é inserido em um filtro de pré-enfase, com função de transferência, mostrada na equação 2.1, com o objetivo de equalizar o espectro de voz e aprimorar o desempenho da analise espectral, que acontece na próxima etapa (CIPRIANO, 2001).

$$H_{pre}(z) = 1 - a_{pre}.z^1 \tag{2.1}$$

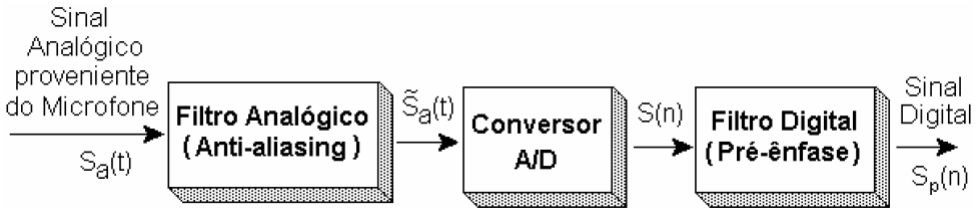
onde $0,9 \leq a_{pre} \leq 1$ é o coeficiente de pré-enfase.

A saída $S_p(n)$ do filtro de pré-enfase, tem relação com a entrada $S(n)$ por meio da equação 2.2 .

$$S_p(n) = S(n) - a_{pre}.S(n-1) \tag{2.2}$$

Na figura 3 é mostrado as etapas da conversão do sinal de voz analógico para digital.

Figura 3 – Etapas da conversão do sinal de voz analógico para digital.



FONTE: (CIPRIANO, 2001).

2.3.2 Análise espectral

A análise espectral tem como objetivo converter o sinal, que está no domínio do tempo, para o domínio da frequência. A conversão é de suma importância pois a audição e percepção humana possuem uma correlação maior com a representação espectral do sinal do que na temporal.

Em seu uso, a análise espectral é iniciada através do processo chamado janelamento, onde o sinal é dividido em quadros de duração fixa. Nesse processo ocorre uma superposição entre as janelas para aumentar a correlação entre as mais próximas e evitar grandes variações entre os parâmetros extraídos. Com o sinal de voz dividido em janelas, a analise espectral já é efetivamente aplicável.

Em sistemas RAV, a analise espectral é feita através do uso de dois principais métodos, que são os bancos de filtros FFT (Fast Fourier Transform) e a codificação preditiva linear ou LPC (Linear Predictive Coding). Com a conversão do sinal realizada para o domínio da frequência, o processo de extração dos parâmetros acústicos pode ser iniciado (THIAGO, 2017).

2.3.3 Extração das características

Esta etapa tem como principal objetivo, comprimir os dados de voz, eliminando qualquer informação que não contribua a analise dos dados e destacar os aspectos do sinal que sejam importante à detecção

das diferenças fonéticas. Na extração dos parâmetros acústicos do sinal de voz a seleção do método a ser usado é uma das etapas mais importantes do Pré-Processamento dos dados.

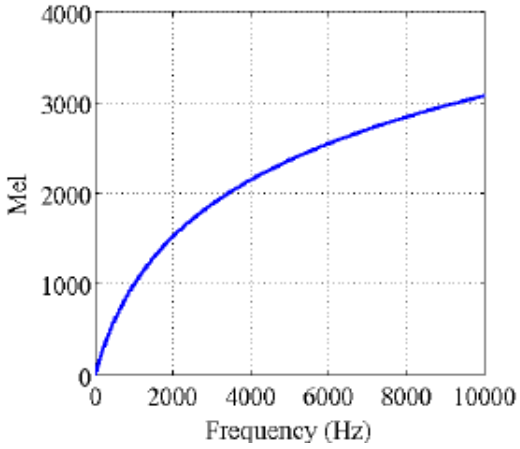
Nessa escolha, um dos métodos mais conhecidos para a extração de características em reconhecimento de voz é a obtenção dos Coeficientes Mel-Cepstrais (MFCC). Em um resumo historico, em 1974 este metodo foi mencionado por Bridle e Brown. E em 1980 desenvolvido com mais detalhes por Davis e Mermelstein. Para extrair um vetor de características contendo a informação linguística de um sinal de voz, a MFCC utiliza a escala Mel para analisar as diferentes frequências presentes no espectro.

A escala Mel foi experimentalmente desenvolvida na década de 1940 por Stevens e Volkmann para identificar como diferentes frequências eram interpretadas pelo aparelho auditivo humano, tendo como objetivo, descrever uma relação entre a frequência real e o que era interpretado. Com base nos resultados obtidos, foi concluído que a relação é linear de 0 a 1000Hz, e que para frequências superiores a 1000Hz, a relação pode ser definida de forma logarítmica. Com a definição que 1000 Hz equivale a 1000 mels, essa relação pode ser expressa matematicamente como:

$$F_{mel} = \frac{1000}{log(2)} \left[1 + \frac{F_{Hz}}{1000}\right] \tag{2.3}$$

onde F_{mel} é a frequência resultante na escala Mel medida em mels e F_{Hz} é a frequência medida em Hertz. Na figura 4 é mostrado a representação da equação 2.3 em uma plano cartesiano.

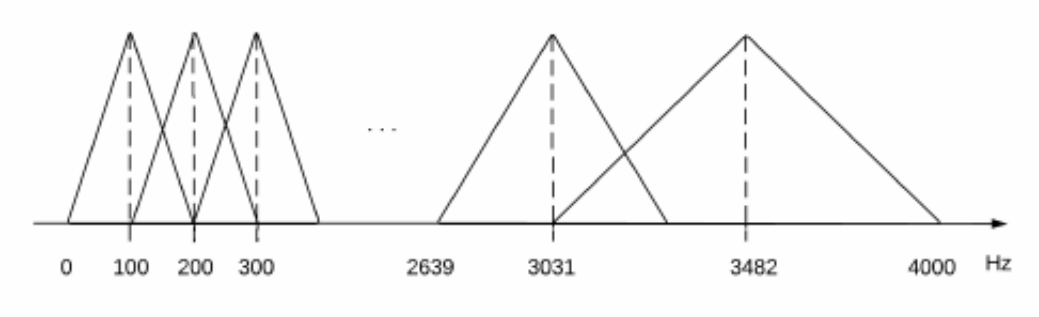
Figura 4 – Relação entre as escalas de Frequência e Mel.



FONTE: (COUVREUR et al., 2008).

No uso prático, a conversão do sinal para a escala Mel é feita usando filtros passa-banda, de resposta triangular, e com espaçamento e largura definidos por um intervalo constante de frequência Mel, como mostrado na figura 5. Na tabela 1, é mostrado um conjunto de frequências centrais de filtros que podem implementar a escala Mel.

Figura 5 – Banco de filtros na escala mel para uma frequência de amostragem de 8 kHz.



FONTE: (MARTINS; YNOGUTI, 2014).

Tabela 1 – Frequências centrais dos filtros na escala Mel

Índice dos Filtros	**Frequência central (Hz)**	**Índice dos Filtros**	**Frequência central (Hz)**
1	100	13	1516
2	200	14	1741
3	300	15	2000
4	400	16	2297
5	500	17	2639
6	600	18	3031
7	700	19	3482
8	800	20	4000
9	900	21	4595
10	1000	22	5278
11	1149	23	6063
12	1320	24	6964

FONTE: (MARTINS; YNOGUTI, 2014).

Tendo o sinal de voz convertido para a escala Mel, o "cepstrum"pode ser calculado. O cepstrum, também conhecido como o "espectro do espectro", é uma operação matemática que aplica a transformada discreta do cosseno (DCT - Discrete cosine transform) no logaritmo da energia, isto é, na saída de cada filtro. O processo de obtenção dos coeficientes MFCC, é mostrado matematicamente na equação 2.4:

$$Cn = \sum_{k=1}^{K} log(S_k) \cos[n \left(k - \frac{1}{2}\right) \frac{\pi}{K}], n = 1, ..., L \tag{2.4}$$

onde L é o número de coeficientes e S_k os coeficientes de potência da saída do k-ésimo filtro.

Geralmente, algumas das últimas amostras obtidas da DCT são descartadas, pois possuem pouca informação sobre o formato do trato vocal utilizado na produção da voz. Assim, estando concluído o vetor de coeficientes, que representa as características acústicas do sinal, o processo de reconhecimento pode ser iniciado utilizando técnicas de aprendizagem de máquina (THIAGO, 2017).

2.4 Aprendizagem de máquina

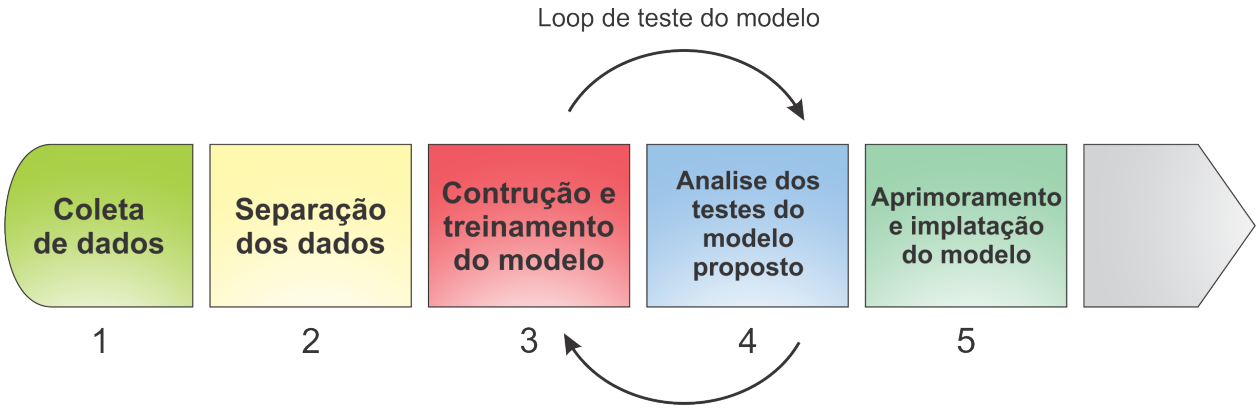
O aprendizado de Máquina (AM) é uma área de pesquisa da Inteligência Computacional que estuda o desenvolvimento de métodos capazes de extrair conhecimento a partir de amostras de dados.

Vários algoritmos são usados para criar classificadores para um conjunto de exemplos, sendo o termo "classificação"definido como o processo de atribuir a uma dada informação, o rótulo da classe a qual ela pertence. Desse modo, por meio de um conjunto de treinamento as técnicas de AM são usadas na indução de um classificador que consiga prever a classe de instancias quaisquer no contexto em que ele foi treinado (LORENA; CARVALHO, 2003).

Nos últimos anos o avanço dessa área em conjuntos com outras áreas tecnológicas, esta proporcionando que computadores tomem decisões sem estarem explicitamente programados. Permitindo que tarefas que por um bom tempo foram consideradas impossíveis de serem executadas por máquinas sejam automatizadas. Alguns exemplos disso são o reconhecimento de voz, a automação na área da robótica, carros autônomos e um melhor entendimento do genoma humano.

Na maiorias dos projetos que usam o aprendizado de máquina, o fluxo de trabalho comum que é executado é mostrado na figura 6. Nele pode-se observar que ao realizar todas as etapas, se espera como resultado um modelo satisfatório e eficaz.

Figura 6 – Processo de aprendizado de máquina.



FONTE: PRÓPRIO AUTOR.

- **Etapa 1:** O dados que serão usados para o desenvolvimento do projeto são identificados e coletados.

- **Etapa 2:** Os dados coletados passam por um processo de limpeza, através da seleção e adequação para serem utilizados. Além disso, é efetuada a divisão dos dados em duas partes, onde uma parte será usada para treinamento do modelo e a outra para os testes do modelo.

- **Etapa 3:** O modelo inicial é construído e um algoritmo ou método é selecionado. Esse é um modelo que parte de um conhecimento prévio dos dados, implicando na identificação da capacidade de assertividade do algoritmo escolhido. Dessa forma, é feito o treinamento desse modelo com os dados já separados da etapa anterior.

- **Etapa 4:** Após o treinamento do modelo, e os dados separados para testes terem sido aplicados ao mesmo, uma análise dos resultados será feita. Em casos em que os resultados não são satisfatórios, é necessário voltar para a etapa 3, como mostra o loop de teste de modelo na Figura x.

- **Etapa 5:** Representa que o modelo foi validado e aplicado, obtendo resultados adequados. Desse modo, o modelo já pode ser implantado, e até aprimorado(ARRIGONI, 2018).

2.5 Técnicas de aprendizagem de máquina

Ao se trabalhar com o aprendizado de máquina, três paradigmas podem ser escolhidos para fazer a classificação: aprendizagem supervisionada, não-supervisionada e por reforço (LORENA; CARVALHO, 2003).

No aprendizado supervisionado um conjunto de exemplos de treinamento é fornecido nos quais o rótulo da classe designada é conhecido, isto é , as entradas e a saídas são conhecidas. Cada exemplo é composto por um vetor de valores de características ou atributos e o rótulo da classe designada. O algoritmo então tem por objetivo criar um classificador que determine corretamente a classe de novos exemplos que ainda não possuam o rótulo da classe. Quando os rótulos das classes são discretos, esse problema é definido como classificação, já se forem contínuos é uma regressão.

No aprendizado não-supervisionado não se conhece a saída, os exemplos fornecidos são analisados e tenta-se observar se alguns deles podem ser agrupados de alguma forma. Ocorrendo os agrupamentos, é realizado uma analise para determinar o que cada agrupamento representa no contexto do problema que esta sendo analisado (MONARD; BARANAUSKAS, 2003).

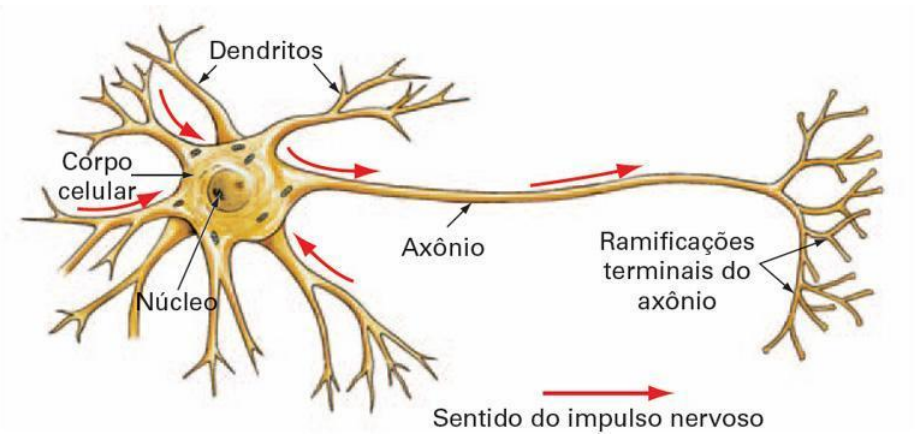
No aprendizado por reforço, o processo ocorre através de recompensas ou não ao classificador, dependendo do desempenho obtido na aproximação da função desejada (LORENA; CARVALHO, 2003).

Como os dados de entrada e saída são conhecidos, no desenvolvimento desse trabalho foram escolhidas para a classificação, técnicas de aprendizado supervisionado. Nas sub sessões seguintes são apresentados as técnicas escolhidas.

2.5.1 Redes Neurais Artificiais

A técnica de Redes Neurais Artificiais (RNA) do inglês *Artificial Neural Networks* (ANN) foram desenvolvidas a partir do neurônio biológico humano. O neurônio é uma célula do cérebro humano, especializado na transmissão de informações, pois possui propriedades de excitabilidade e condução de mensagens nervosas. Sua estrutura é constituída por 3 partes principais: o corpo celular ou a soma, do qual saem algumas ramificações chamadas de dendritos, e por uma outra ramificação, entretanto mais extensa, chamada de axônio. Nas extremidades dos axônios estão os nervos terminais, onde ocorre a transmissão das informações para outros neurônios, conhecida como sinapse. Na figura 7 é mostrado a estrutura do neurônio humano.

Figura 7 – Visão Simplificada do neurônio biológico humano.



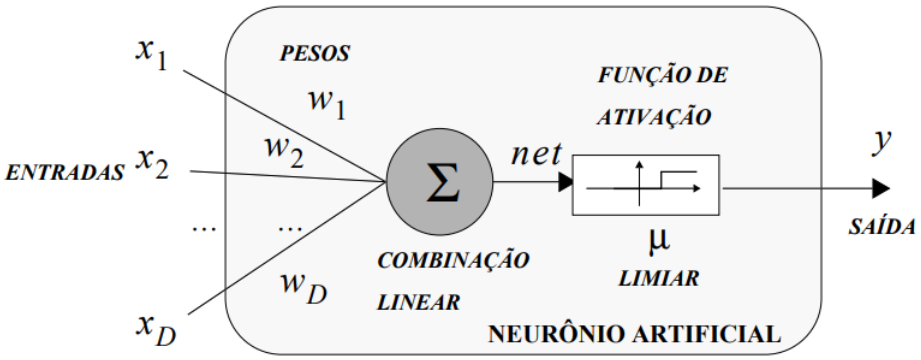
FONTE: (DATA SCIENCE ACADEMY, 2019b).

O cérebro é constituído por bilhões de neurônios, havendo entre eles centenas de bilhões de conexões, formando uma enorme rede de comunicação, chamada rede neural. Em cada neurônio há uma corpo celular, diversos dendritos e um axônio. Os dendritos recebem sinais elétricos de outros neurônios por meio das sinapses, que representa o processo de comunicação entre neurônios. O corpo celular faz o processamento dessa informação e envia para outro neurônio. Desse conjunto o corpo celular e os dendritos formam o conjunto de entrada do neurônio e o axônio a saída do fluxo de informação.

A partir da estrutura e funcionamento do neurônio biológico, vários pesquisadores tentaram simular este sistema em computador. Em 1943, os pesquisadores Warren McCulloch e Walter Pitts, apresentaram o modelo mais bem aceito, que de maneira simplificada implementa os componentes e o funcionamento de um neurônio biológico. Nesse modelo matemático, a rede neural artificial é um componente que calcula a soma ponderada de várias entradas, aplica uma função e passa o resultado adiante.

O funcionamento desse modelo relacionado com o neurônio biológico, é descrito da seguinte maneira: Os impulsos elétricos provenientes de outros neurônios são chamados de sinais de entrada, letra x. Dentre os vários estímulos recebidos, alguns excitarão mais e outros menos o neurônio receptor e essa medida é representada através dos pesos sinápticos, definido por wk_n. Onde k representa o índice do neurônio em questão e n se refere ao terminal de entrada da sinapse a qual o peso sináptico se refere. Quanto maior o valor do peso, mais excitatório é o estímulo. O corpo da célula é representado por dois módulos, o primeiro é o somatório dos estímulos (sinais de entrada) multiplicado pelo seu fator excitatório (pesos sinápticos), e posteriormente uma função de ativação, que definirá com base nas entradas e pesos sinápticos, qual será a saída do neurônio. O axônio representado pela saída yk é obtida pela aplicação da função de ativação. O estímulo pode ser excitatório ou inibitório, representado pelo peso sináptico positivo ou negativo respectivamente, como mostrado na figura 8.

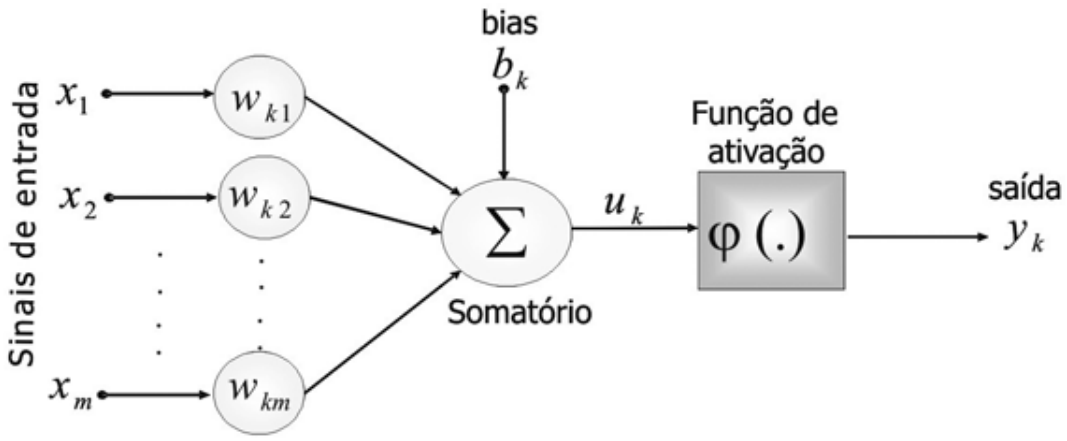
Figura 8 – Visão do Neurônio de McCulloch e Pitts.



FONTE: (RAUBER, 2005).

O modelo também possui incluída ao somatório da função de ativação uma polarização ou bias de entrada, com o objetivo de aumentar o grau de liberdade desta função e a capacidade de aproximação da rede. Seu valor é ajustado da mesma forma que os pesos sinápticos. Com o bias um neurônio pode apresentar saída não nula, ainda que todas as suas entradas sejam nulas. Por exemplo, caso não houvesse o bias e todas as entradas fossem nulas, então o valor da função de ativação seria nulo. Na figura 9 é mostrado essa descrição.

Figura 9 – Visão Simplificada do Neurônio Matemático.



FONTE: (VERONEZ et al., 2009).

- **Sinais de entrada (X1, X2, . . . , Xn):** São os dados externos normalmente normalizados que alimentam o modelo preditivo.
- **Pesos sinápticos (W1, W2, . . . , Wn):** São valores usados para ponderar os sinais de entrada da rede. Esses valores são aprendidos durante o treinamento.
- **Combinador linear ($\sum$):** Soma todos os sinais de entrada que foram ponderados pelos respectivos pesos sinápticos, produzindo um potencial de ativação.
- **Limiar de ativação (θ):** Determina qual será o valor escolhido para que o resultado produzido pelo combinador linear possa gerar um valor de disparo de ativação.
- **Potencial de ativação (u):** É o resultado obtido pela diferença do valor produzido entre o combinador linear e o limiar de ativação. Caso o valor seja positivo (u 0), então o neurônio produz um potencial excitatório. Caso contrário, o potencial será inibitório.
- **Função de ativação (g):** Limita a saída de um neurônio em um intervalo de valores.
- **Sinal de saída (y):** É o valor de saída. Este ainda pode ser usado como entrada de outros neurônios que estão sequencialmente interligados (DATA SCIENCE ACADEMY, 2019b) .

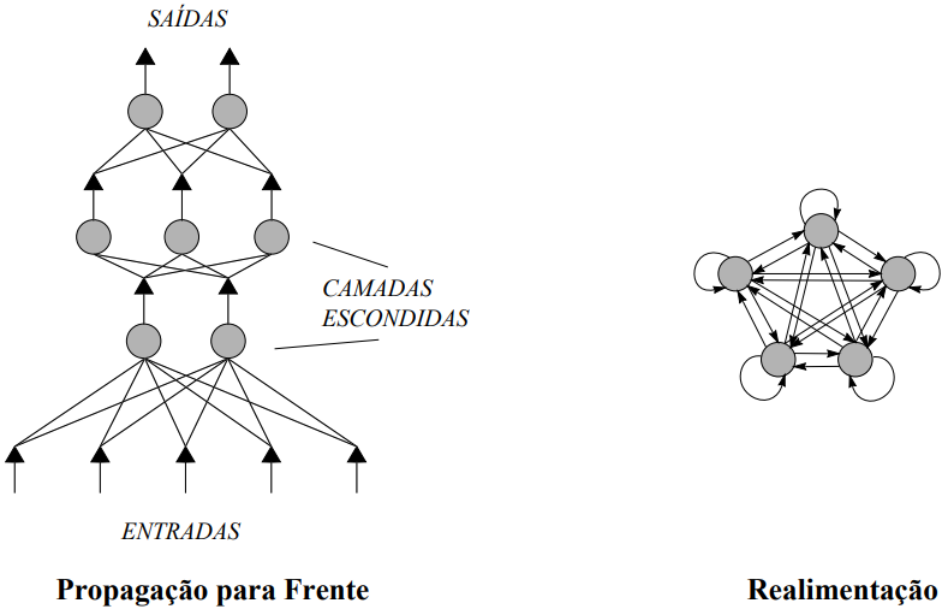
O potencial do calculo baseado em redes neurais, esta na criação de conjuntos de neurônios interligados entre si. O processamento local de elementos em paralelo, estabelece a inteligencia da rede. Um elemento da rede recebe um estimulo nas suas entradas, processa esse sinal e origina um novo sinal de saída, que é recebido por outros elementos.

Nesse sentido a fundamentação da topologia dos neurônios pode ser feita em relação ao método de propagação da informação recebia. A diferença é definida em propagação para frente (*feedforward*) e redes realimentadas (*recurrent*). Em redes de propagação para frente o fluxo de informação é unidirecional. Os neurônios que recebem simultaneamente a informação são agrupamentos em camadas. As camadas que não estão conectadas às entradas e nem às saídas da rede são denominadas camadas escondidas ou intermediarias. Exemplos desse tipo de rede são o *perceptron*, o *perceptron* multi-camada e o o *ADALINE*.

As redes realimentadas têm conexões entre os neurônios sem restrições. O seu comportamento desempenha um papel fundamental dinâmico. Em certas situações os valores de ativação da rede passam por

um processo de relaxação até apresentar um estado estável. Como exemplo se tem a rede auto-associativa. Na figura 10 é mostrada essas topologias (RAUBER, 2005).

Figura 10 – Principais topologias de redes neurais artificiais.



FONTE: (RAUBER, 2005).

Nos últimos anos os modelos baseados em redes neurais artificiais são os mais destacados, pois resolveram problemas de inteligencia artificial (IA), em que outras técnicas utilizados tinham poucos resultados.(DATA SCIENCE ACADEMY, 2019b).

2.5.2 Máquinas de vetores de suporte (SVM)

A técnica de máquina de vetores de suporte, em inglês *Support Vector Machine* (SVM) é um método de aprendizagem para problemas de reconhecimento de padrões. Foi introduzida por Vapnik (1995) através da teoria estatística de aprendizagem, no qual utiliza do fundamento de separação ótima entre classes, onde se as classes são separáveis, então o resultado é obtido de modo a separar o máximo as classes. (NASCIMENTO et al., 2009).

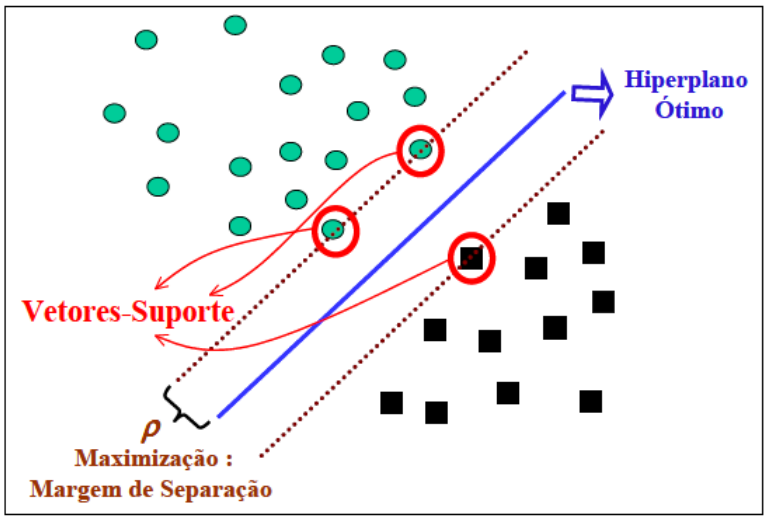
Sua proposta esta em resolver problemas de classificação e análise de regressão. No entanto é principalmente usado em problemas de classificação de duas classes, isto é, atuando como um classificador binário, onde é aplicado em tarefas de classificação de dados lineares e não lineares (Ahmad et al., 2018).

O funcionamento básico de uma SVM ocorre através do fornecimento de duas classes e um conjunto de pontos que pertencem a essas classes. Com essas informações um hiperplano é determinado separando os pontos de forma a colocar o maior numero de pontos da mesma classe do mesmo lado, enquanto atribui o valor mais alto a distancia de cada classe a esse hiperplano. A distância de uma classe a um hiperplano é a menor distância entre ele e os pontos dessa classe, no qual é denominado margem de separação. Já a construção do hiperplano é feita por um subconjunto dos pontos das duas classes, chamados de vetor de suporte (GEVERT et al., 2010).

Sendo os dados de treinamento separáveis, o hiperplano ótimo no espaço de características é o que apresenta a máxima margem de separação p, como mostrado na figura 11. Em dados de treinamento

onde as amostras das diversas classes não são separáveis (superposição), uma generalização dos dados deve ser realizada (ZUBEN; ATTUX,).

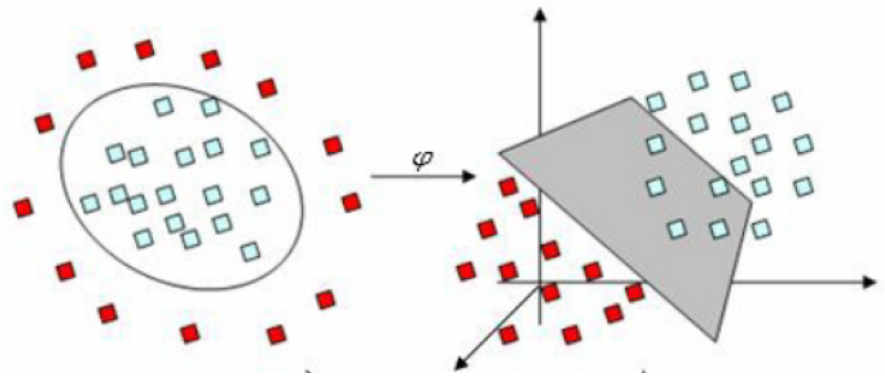
Figura 11 – Dados separados linearmente através do hiperplano que permite maior maximização da margem.



FONTE: (ZUBEN; ATTUX,).

Quando não é possível separar os dados linearmente, é utilizado um classificador não linear que usa funções do kernel para estimar as margens. O objetivo dessas funções é maximizar as margens entre os hiperplanos, sendo as funções do kernel mais empregadas a linear, polinomial, base radial e sigmóide (Ahmad et al., 2018). No funcionamento dessas funções os dados de entrada são mapeados em um espaço de dimensão maior, onde os dados são separados por meio de um hiperplano, tornando-os linearmente separáveis, como mostrado na figura 12 (GEVERT et al., 2010).

Figura 12 – Dados não separáveis linearmente projetados de um plano bidimensional para um plano tridimensional.



FONTE: (GEVERT et al., 2010).

Exemplos de aplicações onde o SVM pode ser encontrado estão em vários contextos como na categorização de textos, na análise de imagens e em Bioinformática (LORENA; CARVALHO, 2007).

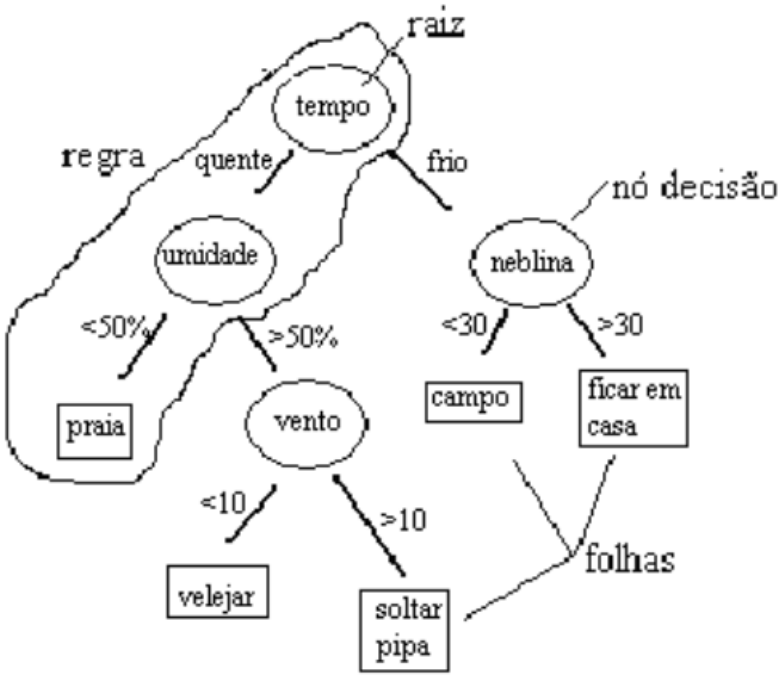
2.5.3 Árvores de decisão

A técnica de árvore de decisão, em inglês *Decision Tree* (DT), é um método usado em problemas de classificação e regressão, que particiona recursivamente um conjunto de treinamento, até que cada subconjunto resultante apresente casos de uma única classe. Para isso, na construção da arvore o algoritmo escolhido examina e compara a distribuição de classes.

A construção da arvore de decisão é fundamentada no modelo Top-Down, isto é, a sua estrutura inicia do nó raiz e vai até as suas folhas. Apesar dos algoritmos desse grupo conterem diferenças importantes em sua lógica de execução, todos utilizam da técnica de dividir para conquistar. Que diz que o problema deve ser dividido sucessivamente em vário problemas menores, até uma solução para cada um dos problemas mais simples seja encontrada. Nesse fundamento, os algoritmos tentam descobrir maneiras de dividir sucessivamente o conjunto em vários subconjuntos, até que cada um apresente apenas uma classe ou que uma das classes seja a maioria, não sendo necessário novas divisões (CASTANHEIRA, 2008).

Na figura 14 é mostrado um exemplo de uma árvore de decisão. Em sua estrutura os seguintes termos são utilizados:

Figura 13 – Estrutura de uma árvore de decisão.



FONTE: (CASTANHEIRA, 2008).

- **Nó:** São todos os elementos que aparecem na árvore;
- **Nó Raiz:** É o primeiro nó do topo da árvore;
- **Nó pai e nó filho:** São os nós logo abaixo do nó raiz que se dividem em sub nós. Um nó dividido em sub nós é chamado nó pai de sub nós, enquanto os sub nós são filhos do nó pai;
- **Nó de Decisão:** Quando um sub nó se divide em outros sub nós.
- **Folhas ou Terminal:** São os nós que não têm filhos, os últimos elementos da árvore;

Em sua interpretação cada nó de decisão realiza um teste para algum atributo, cada ramo descendente representa um valor desse atributo, o conjunto de ramificações são diferentes, as folhas definem uma classe, e cada percurso da arvore iniciando do nó raiz até a folha, constitui uma regra de classificação. Analisando essa estrutura, é possível extrair regras do tipo "se-então"para melhor compreensão dos resultados.

Para definir as partições da árvore, a regra utilizada é o quão útil o atributo é para a classificação. A partir dessa regra, um determinado ganho de informação é aplicado a cada atributo. Dessa forma, o atributo escolhido como atributo teste para o nó corrente, é o que contem o maior ganho de informação. Por meio dessa aplicação, um novo processo de partição é iniciado. Quando a árvore é usada para classificação, as regras de partição mais conhecidas são a entropia e o índice de Gini

A entropia é o calculo do ganho de informação que mede a impureza dos dados. Em um conjunto de dados, mede a falta de homogeneidade dos dados de entrada em relação a sua classificação. Já o índice Gini, desenvolvido por Conrado Gini em 1912, calcula o grau de heterogeneidade dos dados, sendo usado para calcular a impureza de um nó.

Um problema encontrado no uso de arvores de decisão, é o *Overfitting*, que significa o ajuste exagerado dos dados de treinamento. Na execução do algoritmo de partição recursiva, o mesmo estende a sua profundidade até o ponto de classificar corretamente os elementos do conjunto de treinamento. Nesse processo, quando o conjunto de treinamento não possui ruído, o número de erros no treinamento pode ser zero. Porem, quando o conjunto possui ruído, ou quando o conjunto de treinamento não é representativo, este algoritmo pode produzir o *overfitting*.

Segundo Breiman (1998), uma maneira de impedir o problema do *overfitting* e melhorar a classificação é fazer a poda da árvore. A podagem da arvore tende a ser feita em duas situações. Para parar o crescimento da árvore mais cedo, conhecido como pré-podagem ou poda descendente. Ou com a árvore já completa, conhecido como pós-podagem ou poda ascendente.

Para determinar o numero ideal de nós, utiliza-se uma representação gráfica que mostra o percentual de erro no conjunto de treinamento, versus o numero de nós da arvore. Quando o erro no conjunto de teste começar a crescer, o numero de nós nesse ponto é considerado ideal(SILVA, 2005).

Algoritmos que usam a ideia das arvores de decisão são bastante aplicados por retornarem modelos preditivos de fácil interpretação, estabilidade e alta precisão. O método *Top-Down Induction of Decision Tree* (TDIDT) é um dos mais utilizados na aplicação de arvore de decisão e é referencia para outros algoritmos como o *Classification and Regression Tree* (CART) e o C5.0.

Em sua execução o TDIDT faz uma busca recursiva por atributos que melhor dividem o conjunto de observações em sub-conjuntos. O CART é caracterizado por ter a capacidade de assimilar relações entre dados, mesmo não havendo visíveis relações. Em seu uso faz a construção de arvores binarias, onde cada nó interno possui como saída dois nós filhos. Já no C5.0 o algoritmo transforma árvores treinadas em conjuntos de regras if-then, e avalia qual ordem essas regras devem ser aplicadas (ARRIGONI, 2018)

No seu uso prático, as árvores de decisão foram aplicadas em vários domínios da ciência e engenharia como pesquisas de produtos farmacêuticos, saúde pública, biologia celular, consumo de energia elétrica, estudos de transporte, entre outros(Rivera-Lopez; Canul-Reich, 2018).

2.5.4 Floresta aleatória - Random Forest

A técnica floresta aleatória, do inglês Random Forest (RF), é um método desenvolvido por Breiman (2001), para resolver problemas de classificação e regressão de métodos de aprendizagem em

arvores, por meio do bootstrap dos dados de treinamento, aumentando o uso do algoritmo conhecido como CART (classification e Regression Trees) proposto por Breiman et al. (1984).

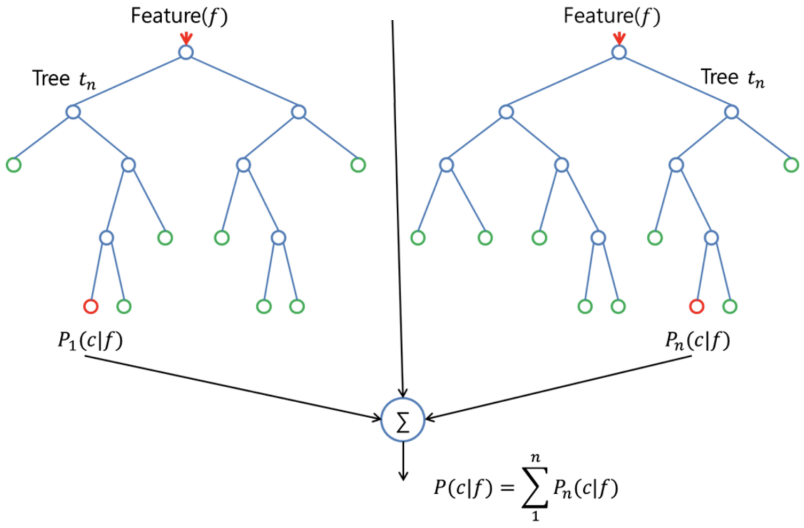
Na grande maioria, os métodos de decisão em arvore possuem viés baixo, mas alta variância, resultando na produção de um sobre ajuste (overfitting). Utilizando RF o problema da variância é tratado usando um numero n de árvores de modo que ao produzir n observações independentes f_1, f_2, ..., f_n, uma para cada nó final de árvores, a variância diminui através da média dessa n observações. Isso porque a variância individual de cada árvore é maior do que a variância da média delas (MOTA, 2019).

O termo floresta é utilizado, pois a técnica cria uma combinação (*ensemble*) de arvores de decisão, que na maiorias das vezes são treinadas com o método de *bagging*. Este método possui como fundamento a ideia que a combinação dos modelos de aprendizado aumenta o resultado geral. Em uma definição resumida, o algoritmo de florestas aleatórias gera várias árvores de decisão e as combina para conseguir uma predição com maior acurácia e mais estabilidade (MEDIUM, 2018a).

A precisão do classificador de RF é muito boa, possui forte capacidade de generalização, tem alta velocidade computacional e os parâmetros configuráveis são muito poucos. Em particular, o RF é apropriado para calcular a função não linear de variáveis e pode refletir a interação entre variáveis (Liu et al., 2019).

O processo de treinamento do algoritmo RF é baseado no método de bagging. No seu funcionamento, Considerando um conjunto de dados D, com d tuplas. A cada interação i, onde (i = 1,2,...,k) um conjunto de treinamento D_i de d tuplas é amostrado com elementos do conjunto de dados D. Isto é, do conjunto de dados iniciais, k amostras são formadas com instâncias e atributos escolhidos de forma aleatória. Assim, cada conjunto D_i passa pelo processo de treinamento de uma árvore de decisão. Através da média do resultado de todas as árvores o resultado final é obtido (CONTI et al., 2019). Na figura 14 é mostrado uma estrutura genérica do funcionamento desse algoritmo com duas árvores de decisão.

Figura 14 – Estrutura de uma floresta aleatória com 2 árvores de decisão.



FONTE: (MEDIUM, 2018a).

2.5.5 Naive Bayes

A técnica Naive Bayes é um modelo que foi proposto pelo reverendo Thomas Bayes no século XVIII(TORRES, 2005). Ela é um algoritmo classificador probabilístico que calcula um conjunto de probabilidades, contando as combinações e a frequência de valores em um determinado conjunto de dados.

O algoritmo utiliza o teorema de Bayes e considera que todos os atributos são independentes, fornecido o valor da variável de classe. Essa suposição de independência condicional dificilmente é usada em aplicações do mundo real, o que a caracteriza como ingênua, mas o algoritmo demostra aprender rapidamente e ter um bom desempenho em vários problemas de classificação supervisionada.

Para o cálculo das probabilidades é utilizado o teorema de Bayes. Nesse calculo a probabilidade de um documento d com o vetor x = $<x_1, ..., x_n>$ pertencer à hipótese h é mostrado na equação 2.5:

$$P(h_1|x_i) = \frac{P(x_i|h_1)P(h_1)}{P(x_i|h_1)P(h_1) + P(x_i|h_2)P(h_2)} \tag{2.5}$$

Onde, $P(h_1|x_i)$ é uma probabilidade posterior, e $P(h_1)$ é a probabilidade anterior associada à hipótese h_1. Para m hipóteses diferentes, $P(x_i)$ é descrito pela equação 2.6:

$$P(x_i) = \sum_{j=1}^{n} P(x_i|h_j)P(h_j) \tag{2.6}$$

Dessa forma a equação 2.5 pode ser modificada, como mostrado na equação 2.7 abaixo (PATIL; SHEREKAR et al., 2013).

$$P(h_1|x_i) = \frac{P(x_i|h_1)P(h_1)}{P(x_i)} \tag{2.7}$$

2.5.6 K-Vizinhos mais próximos

A técnica K-Vizinhos mais próximos, no inglês K-Nearest Neighbors (K-NN) é um método classificador onde o aprendizado é fundamentado na similaridade que um dado (vetor) tem de outro. O treinamento utiliza um conjunto de dados composto por vetores n-dimensionais e cada elemento desse conjunto defini um ponto no espaço n-dimensional.

No seu funcionamento, para descobrir a classe de um elemento que não pertença ao conjunto de dados de treinamento, o KNN busca K elementos do conjunto que estejam mais próximos do elemento desconhecido, isto é, que tenham a menor distância. Estes K elementos são chamados de K-vizinhos mais próximos, e verificando a qual classes esses K vizinhos pertencem, a classe com maior numero de elementos será atribuída à classe do elemento desconhecido .

Para o calculo da distância entre dois pontos, as métricas mais comuns são a distância Euclidiana, Manhattan e Minkowski, sendo a primeira a mais utilizada. A baixo é descrito este calculo.

Seja $X = (x_1, x_2, ..., x_n)$ e $Y = (y_1, y_2, ..., y_n)$ dois pontos no $\Re^n$

- A distância Euclidiana entre X e Y é descrita pela equação 2.8:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_n - y_n)^2} \tag{2.8}$$

Se cada variável for considerada ter certa importância, um peso relativo pode ser incluso, desse modo a distância Euclidiana ponderada pode ser descrita como mostrado na equação 2.9. Nas distâncias Manhattan e Minkowski, os pesos também podem ser aplicados (SILVA, 2005).

$$d(x, y) = \sqrt{w_1(x_1 - y_1)^2 + w_2(x_2 - y_2)^2 + ... + w_n(x_n - y_n)^2} \tag{2.9}$$

Na utilização do algoritmo KNN, a sua execução ocorre por meio das seguintes etapas:

1. Recebe um dado não classificado;

2. Mede a distância (Euclidiana, Manhattan, Minkowski ou Ponderada) do dado desconhecido com todos os outros dados que já estão classificados;

3. Obtém-se as K menores distâncias;

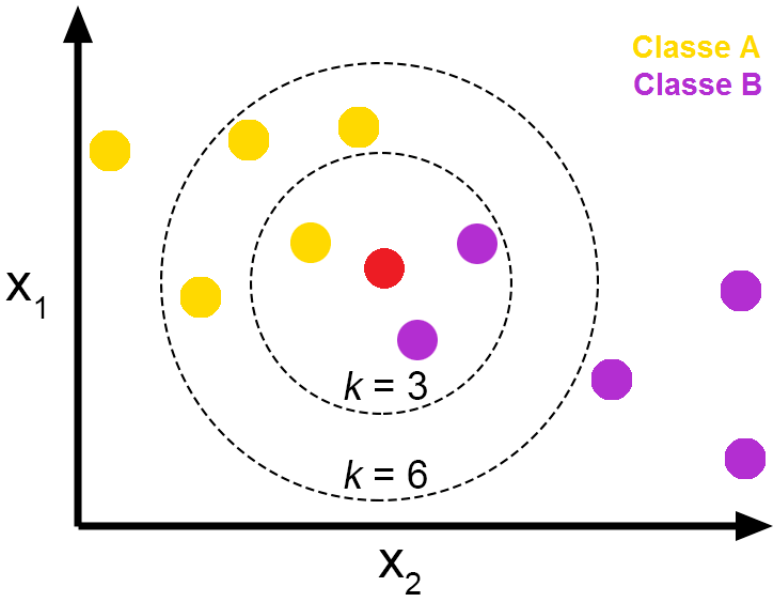
4. Verifica-se a classe de cada um dos dados que tiveram a menor distância e se conta a quantidade de cada classe que aparece;

5. Se escolhe como resultado a classe que mais apareceu dentre os dados que tiveram as menores distâncias;

6. Classifica-se o dado desconhecido com a classe escolhida do resultado da classificação que mais apareceu.

Na figura 15 é mostrado esse processo. Olhando a figura tem-se um dado não classificado representado em vermelho, e todos os os outros dados já classificados, definidos em amarelo e roxo, que respectivamente representam a classe A e a classe B. O calculo da distancia do dado desconhecido com todos os outros é feito para verificar quais estão mais próximos, isto é, tem as menores distancias. Com o resultado do calculo, se escolhe três ou seis dos dados mais próximos e verifica qual é a classe que mais aparece. Nesse exemplo é usado K igual a três. Os dados mais próximos do dado desconhecido são os que estão dentro do primeiro circulo (olhando de dentro para fora). Nele há três dados já classificados, e observando a classe que mais predomina, se conclui que o dado roxo (Classe B), é o mais predominante, pois possui dois dados em relação ao amarelo que possui só um. Desse modo, o dado desconhecido passa a ser classificado como sendo o roxo da classe B. Caso o valor de K seja igual a 6, o dado desconhecido passa ser classificado como sendo o amarelo da classe A (MEDIUM, 2018b).

Figura 15 – Classificação de um dado desconhecido utilizando k-NN.



FONTE: (MEDIUM, 2018b).

O classificador KNN possui o numero de K-vizinhos como único parâmetro livre, o qual é controlado pelo usuário para obter uma melhor classificação. Esta classificação pose ser exaustiva computacionalmente se utilizado um conjunto com muito dados. Entretanto, em aplicações especificas seu uso é recomendado (SILVA, 2005).

3 PROPOSTA

Para este trabalho se propõe a implementação de algoritmos de técnicas de aprendizado de máquina para analisar os seus desempenhos no reconhecimento de voz. O aprendizado que será usado é o supervisionado, e as técnicas escolhidas são: Redes Neurais Artificiais, Máquinas de vetores de suporte (SVN), Árvores de decisão, Random Forest, Naive Bayes e K-Nearest Neighbors (KNN). Para o reconhecimento da voz, serão usados uma base de dados processada com MFCC (Mel-Frequency Cepstral Coefficients, composta por palavras isoladas, sendo estas palavras dígitos de zero a nove, pronunciados em inglês. A base de dados utilizada será da *Texas Instruments*, e os algoritmos serão desenvolvidos na linguagem R. Os resultados obtidos serão avaliados em termos de eficiência, precisão (acurácia), sensibilidade e especificidade. De modo a verificar se os resultados são satisfatórios, os mesmos serão comparados com resultados de um trabalho relacionado a esse tema. O trabalho relacionado ainda vai ser definido.

Para o desenvolvimento desse trabalho, se definiu um cronograma de atividades como mostrado na tabela 2

Tabela 2 – Cronograma das atividades previstas

Atividades	Mês					
	Fev	Mar	Abr	Mai	Jun	Jul
Desenvolvimento dos algoritmos das técnicas de aprendizado de máquina	√	√	√	√		
Teste do reconhecimento		√	√	√		
Avaliação dos resultados			√	√	√	
Escrita do TCC				√	√	√
Defesa do TCC						√

REFERÊNCIAS

Ahmad, I. et al. Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection. *IEEE Access*, v. 6, p. 33789–33795, 2018. ISSN 2169-3536. Citado 2 vezes nas páginas 25 e 26.

ANDRADE, C. Araujo de et al. Avaliação de um módulo de reconhecimento de fala na plataforma arduino. *Revista Principia - Divulgação Científica e Tecnológica do IFPB*, v. 1, p. 19, 06 2016. Citado na página 11.

ARRIGONI, T. R. Reconhecimento de silhueta de automóveis para carros autônomos utilizando aprendizado de máquina. 2018. Citado 3 vezes nas páginas 11, 21 e 28.

BARBOSA, J. P. *Os três pilares da inovação – Machine Learning, Big Data e IoT*. 2017. Disponível em: <http://igti.com.br/blog/os-tres-pilares-da-inovacao-machine-learning-big-data-e-iot>. Acesso em: 19 Outubro. 2019. Citado na página 11.

CASTANHEIRA, L. G. Aplicação de técnicas de mineração de dados em problemas de classificação de padrões. *Belo Horizonte: UFMG*, 2008. Citado na página 27.

CIPRIANO, J. L. G. Desenvolvimento de arquitetura para sistemas de reconhecimento automático de voz baseados em modelos ocultos de markov. 2001. Citado 3 vezes nas páginas 15, 17 e 18.

CONTI, J. P. J. et al. *Redes neurais recorrentes e expoente de Lyapunov aplicados a séries temporais financeiras*. Dissertação (Mestrado) — Universidade Tecnológica Federal do Paraná, 2019. Citado na página 29.

COUVREUR, L. et al. Audio thumbnailing. *QPSR of the numediart research program*, v. 1, p. 67–85, 01 2008. Citado na página 19.

DATA SCIENCE ACADEMY. *Capítulo 1 – Deep Learning e a Tempestade Perfeita*. 2019. Disponível em: <http://deeplearningbook.com.br/deep-learning-a-tempestade-perfeita>. Acesso em: 17 Outubro. 2019. Citado na página 11.

DATA SCIENCE ACADEMY. *Capítulo 4 – O Neurônio, Biológico e Matemático*. 2019. Disponível em: <http://deeplearningbook.com.br/o-neuronio-biologico-e-matematico>. Acesso em: 23 novembro. 2019. Citado 3 vezes nas páginas 22, 24 e 25.

FERNANDES, T. G.; PANAZIO, A. N. Do analógico ao digital: amostragem, quantização e codificação. *II Simpósio de Iniciação Científica da Universidade Federal do ABC-SIC-UFABC*, 2009. Citado na página 17.

GABANINI, A. P. N. *A Voz Humana*. 2003. Disponível em: <http://www.profala.com/arttf57.htm>. Acesso em: 03 Novembro. 2019. Citado na página 15.

GEVERT, V. G. et al. Modelos de regressão logística, redes neurais e support vector machine (svm s) na análise de crédito a pessoas jurídicas. *RECEN-Revista Ciências Exatas e Naturais*, v. 12, n. 2, p. 269–293, 2010. Citado 2 vezes nas páginas 25 e 26.

Liu, S. et al. Random forest-based track initiation method. *The Journal of Engineering*, v. 2019, n. 19, p. 6175–6179, 2019. ISSN 2051-3305. Citado na página 29.

LORENA, A. C.; CARVALHO, A. C. de. Introduçaoas máquinas de vetores suporte. *Relatório Técnico do Instituto de Ciências Matemáticas e de Computaçao (USP/Sao Carlos)*, v. 192, 2003. Citado 3 vezes nas páginas 11, 21 e 22.

LORENA, A. C.; CARVALHO, A. C. de. Uma introdução às support vector machines. *Revista de Informática Teórica e Aplicada*, v. 14, n. 2, p. 43–67, 2007. Citado na página 26.

MARTINS, R.; YNOGUTI, C. Normalização do locutor em sistemas de reconhecimento de fala para usuários crianças. In: . [S.l.: s.n.], 2014. Citado na página 20.

MEDIUM. *Aprendendo em uma Floresta Aleatória*. 2018. Disponível em: <https://medium.com/machina-sapiens/o-algoritmo-da-floresta-aleatória-3545f6babdf8>. Acesso em: 02 Dezembro. 2019. Citado na página 29.

MEDIUM. *KNN (K-Nearest Neighbors) 1*. 2018. Disponível em: <https://medium.com/brasil-ai/knn-k-nearest-neighbors-1-e140c82e9c4e>. Acesso em: 01 Dezembro. 2019. Citado 2 vezes nas páginas 31 e 32.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. *Sistemas inteligentes-Fundamentos e aplicações*, v. 1, n. 1, p. 32, 2003. Citado na página 22.

MOTA, A. A. L. *Previsão de prêmio e a ocorrência de sinistros no mercado de seguro agrícola brasileiro*. Tese (Doutorado) — Universidade de São Paulo, 2019. Citado na página 29.

MULATINHO, G. M. Reconhecimento de voz para aplicações em automação implementado em fpga. Universidade Estadual Paulista (UNESP), 2011. Citado 2 vezes nas páginas 15 e 16.

NASCIMENTO, R. F. F. et al. O algoritmo support vector machines (svm): avaliação da separação ótima de classes em imagens ccd-cbers-2. *Simpósio Brasileiro de Sensoriamento Remoto*, v. 14, p. 2079–2086, 2009. Citado na página 25.

PATIL, T. R.; SHEREKAR, S. et al. Performance analysis of naive bayes and j48 classification algorithm for data classification. *International journal of computer science and applications*, v. 6, n. 2, p. 256–261, 2013. Citado na página 30.

RAUBER, T. W. Redes neurais artificiais. *Universidade Federal do Espírito Santo*, 2005. Citado 2 vezes nas páginas 23 e 25.

Rivera-Lopez, R.; Canul-Reich, J. Construction of near-optimal axis-parallel decision trees using a differential-evolution-based approach. *IEEE Access*, v. 6, p. 5548–5563, 2018. ISSN 2169-3536. Citado na página 28.

SILVA, L. M. Uma aplicação de árvores de decisão, redes neurais e knn para a identificação de modelos arma não-sazonais e sazonais. *Rio de Janeiro. 145p. Tese de Doutorado-Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro*, 2005. Citado 3 vezes nas páginas 28, 31 e 32.

THIAGO, E. R. de S. Reconhecimento de voz utilizando extração de coeficientes mel-cepstrais e redes neurais artificiais. 2017. Citado 5 vezes nas páginas 12, 15, 16, 18 e 20.

TORRES, J. Utilização do corte percentual na categorização de documentos da web com o algoritmo naive bayes. In: . [S.l.: s.n.], 2005. Citado na página 30.

TRINDADE, R. *Siri, Alexa e Google: Veja como usar as assistentes de voz no celular*. 2018. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2018/05/31/siri-alexa-e-google-veja-como-usar-as-assistentes-de-voz-no-celular.htm>. Acesso em: 19 Outubro. 2019. Citado na página 11.

VALIATI, J. F. Reconhecimento de voz para comandos de direcionamento por meio de redes neurais. 2000. Citado na página 15.

VERONEZ, M. et al. Estimativa de alturas geoidais para o estado de são paulo baseada em redes neurais artificiais. *Revista Brasileira de Geofísica*, v. 27, p. 583–593, 12 2009. Citado na página 24.

ZUBEN, F. J. V.; ATTUX, R. R. Máquinas de vetores-suporte. Citado na página 26.