

Estudos (Inter) Multidisciplinares nas Ciências Exatas e Tecnologias

Henrique Ajuz Holzmann
João Dallamuta
Ricardo Vinicius Bubna Biscaia
(Organizadores)

Estudos (Inter) Multidisciplinares nas Ciências Exatas e Tecnologias

Henrique Ajuz Holzmann
João Dallamuta
Ricardo Vinicius Bubna Biscaia
(Organizadores)

2019 by Atena Editora
Copyright © Atena Editora
Copyright do Texto © 2019 Os Autores
Copyright da Edição © 2019 Atena Editora
Editora Chefe: Profª Drª Antonella Carvalho de Oliveira
Diagramação: Lorena Prestes
Edição de Arte: Lorena Prestes
Revisão: Os Autores



Todo o conteúdo deste livro está licenciado sob uma Licença de Atribuição Creative Commons. Atribuição 4.0 Internacional (CC BY 4.0).

O conteúdo dos artigos e seus dados em sua forma, correção e confiabilidade são de responsabilidade exclusiva dos autores. Permitido o download da obra e o compartilhamento desde que sejam atribuídos créditos aos autores, mas sem a possibilidade de alterá-la de nenhuma forma ou utilizá-la para fins comerciais.

Conselho Editorial

Ciências Humanas e Sociais Aplicadas

Profª Drª Adriana Demite Stephani – Universidade Federal do Tocantins
Prof. Dr. Álvaro Augusto de Borba Barreto – Universidade Federal de Pelotas
Prof. Dr. Alexandre Jose Schumacher – Instituto Federal de Educação, Ciência e Tecnologia de Mato Grosso
Prof. Dr. Antonio Carlos Frasson – Universidade Tecnológica Federal do Paraná
Prof. Dr. Antonio Gasparetto Júnior – Instituto Federal do Sudeste de Minas Gerais
Prof. Dr. Antonio Isidro-Filho – Universidade de Brasília
Prof. Dr. Constantino Ribeiro de Oliveira Junior – Universidade Estadual de Ponta Grossa
Profª Drª Cristina Gaio – Universidade de Lisboa
Prof. Dr. Deyvison de Lima Oliveira – Universidade Federal de Rondônia
Prof. Dr. Edvaldo Antunes de Faria – Universidade Estácio de Sá
Prof. Dr. Eloi Martins Senhora – Universidade Federal de Roraima
Prof. Dr. Fabiano Tadeu Grazioli – Universidade Regional Integrada do Alto Uruguai e das Missões
Prof. Dr. Gilmei Fleck – Universidade Estadual do Oeste do Paraná
Profª Drª Ivone Goulart Lopes – Istituto Internazionele delle Figlie de Maria Ausiliatrice
Prof. Dr. Julio Candido de Meirelles Junior – Universidade Federal Fluminense
Profª Drª Keyla Christina Almeida Portela – Instituto Federal de Educação, Ciência e Tecnologia de Mato Grosso
Profª Drª Lina Maria Gonçalves – Universidade Federal do Tocantins
Profª Drª Natiéli Piovesan – Instituto Federal do Rio Grande do Norte
Prof. Dr. Marcelo Pereira da Silva – Universidade Federal do Maranhão
Profª Drª Miranilde Oliveira Neves – Instituto de Educação, Ciência e Tecnologia do Pará
Profª Drª Paola Andressa Scortegagna – Universidade Estadual de Ponta Grossa
Profª Drª Rita de Cássia da Silva Oliveira – Universidade Estadual de Ponta Grossa
Profª Drª Sandra Regina Gardacho Pietrobon – Universidade Estadual do Centro-Oeste
Profª Drª Sheila Marta Carregosa Rocha – Universidade do Estado da Bahia
Prof. Dr. Rui Maia Diamantino – Universidade Salvador
Prof. Dr. Urandi João Rodrigues Junior – Universidade Federal do Oeste do Pará
Profª Drª Vanessa Bordin Viera – Universidade Federal de Campina Grande
Prof. Dr. Willian Douglas Guilherme – Universidade Federal do Tocantins

Ciências Agrárias e Multidisciplinar

Prof. Dr. Alexandre Igor Azevedo Pereira – Instituto Federal Goiano
Prof. Dr. Antonio Pasqualetto – Pontifícia Universidade Católica de Goiás
Profª Drª Daiane Garabeli Trojan – Universidade Norte do Paraná
Profª Drª Diocléa Almeida Seabra Silva – Universidade Federal Rural da Amazônia
Prof. Dr. Écio Souza Diniz – Universidade Federal de Viçosa
Prof. Dr. Fábio Steiner – Universidade Estadual de Mato Grosso do Sul
Profª Drª Girlene Santos de Souza – Universidade Federal do Recôncavo da Bahia
Prof. Dr. Jorge González Aguilera – Universidade Federal de Mato Grosso do Sul
Prof. Dr. Júlio César Ribeiro – Universidade Federal Rural do Rio de Janeiro
Profª Drª Raissa Rachel Salustriano da Silva Matos – Universidade Federal do Maranhão
Prof. Dr. Ronilson Freitas de Souza – Universidade do Estado do Pará
Prof. Dr. Valdemar Antonio Paffaro Junior – Universidade Federal de Alfenas

Ciências Biológicas e da Saúde

Prof. Dr. Benedito Rodrigues da Silva Neto – Universidade Federal de Goiás
Prof. Dr. Edson da Silva – Universidade Federal dos Vales do Jequitinhonha e Mucuri
Profª Drª Elane Schwinden Prudêncio – Universidade Federal de Santa Catarina
Prof. Dr. Gianfábio Pimentel Franco – Universidade Federal de Santa Maria
Prof. Dr. José Max Barbosa de Oliveira Junior – Universidade Federal do Oeste do Pará
Profª Drª Magnólia de Araújo Campos – Universidade Federal de Campina Grande
Profª Drª Natiéli Piovesan – Instituto Federal do Rio Grande do Norte
Profª Drª Vanessa Lima Gonçalves – Universidade Estadual de Ponta Grossa
Profª Drª Vanessa Bordin Viera – Universidade Federal de Campina Grande

Ciências Exatas e da Terra e Engenharias

Prof. Dr. Adélio Alcino Sampaio Castro Machado – Universidade do Porto
Prof. Dr. Alexandre Leite dos Santos Silva – Universidade Federal do Piauí
Profª Drª Carmen Lúcia Voigt – Universidade Norte do Paraná
Prof. Dr. Eloi Rufato Junior – Universidade Tecnológica Federal do Paraná
Prof. Dr. Fabrício Menezes Ramos – Instituto Federal do Pará
Prof. Dr. Juliano Carlo Rufino de Freitas – Universidade Federal de Campina Grande
Profª Drª Neiva Maria de Almeida – Universidade Federal da Paraíba
Profª Drª Natiéli Piovesan – Instituto Federal do Rio Grande do Norte
Prof. Dr. Takeshy Tachizawa – Faculdade de Campo Limpo Paulista

Dados Internacionais de Catalogação na Publicação (CIP) (eDOC BRASIL, Belo Horizonte/MG)	
E82	<p>Estudos (inter) multidisciplinares nas ciências exatas e tecnologias [recurso eletrônico] / Organizadores Henrique Ajuz Holzmann, João Dallamuta, Ricardo Vinicius Bubna Biscaia. – Ponta Grossa, PR: Atena Editora, 2019.</p> <p>Formato: PDF Requisitos de sistema: Adobe Acrobat Reader. Modo de acesso: World Wide Web. Inclui bibliografia ISBN 978-85-7247-798-7 DOI 10.22533/at.ed.987192611</p> <p>1. Ciências exatas – Pesquisa – Brasil. 2. Tecnologia. I. Holzmann, Henrique Ajuz. II. Dallamuta, João. III. Biscaia, Ricardo Vinicius Bubna.</p> <p style="text-align: right;">CDD 509</p>
Elaborado por Maurício Amormino Júnior – CRB6/2422	

Atena Editora
Ponta Grossa – Paraná - Brasil
www.atenaeditora.com.br
contato@atenaeditora.com.br

APRESENTAÇÃO

Neste livro são apresentados vários trabalhos, alguns com resultados práticos, outros com métodos de desenvolvimento para o ensino de tecnologias, bem como um enfoque em energias renovais.

Um compendio de temas e abordagens que constituem a base de conhecimento de profissionais que buscam estar atualizados e alinhados com as novas tecnologias .

A obra Estudos (Inter) Multidisciplinares nas Ciências Exatas e Tecnologias aborda os mais diversos assuntos sobre a aplicação de métodos e ferramentas nas diversas áreas das engenharias a fim de melhorar a relação ensino aprendizado, sendo por meio de levantamentos teórico-práticos de dados referentes aos cursos ou através de propostas de melhoria nestas relações.

Outro ponto de grande destaque, são as novas ferramentas utilizadas em um compendio relacionado ao ensino-aprendizagem, como ferramentas tecnológicas que facilitem o entendimento e executem um link entre aluno-professor-conteúdo.

Desta forma temas e abordagens que facilitam as relações entre ensino-aprendizado são apresentados, a fim de se levantar dados e propostas para novas discussões em relação ao ensino nas engenharias, de maneira atual e com a aplicação das tecnologias hoje disponíveis.

Boa leitura

Henrique Ajuz Holzmann

João Dallamuta

Ricardo Vinicius Bubna Biscaia

SUMÁRIO

CAPÍTULO 1	1
A TECNOLOGIA E AS MUDANÇAS NO MERCADO DE TRABALHO	
Eduardo Bruno de Almeida Donato Amanda Moura Camilo	
DOI 10.22533/at.ed.9871926111	
CAPÍTULO 2	9
CIBERCULTURA: ESPAÇO DE APRENDIZAGEM COLABORATIVA À LUZ DA TEORIA ATOR-REDE	
Diane Schlieck Martha Kaschny Borges	
DOI 10.22533/at.ed.9871926112	
CAPÍTULO 3	22
AUTORREGULAÇÃO DA APRENDIZAGEM: DESAFIOS E PERSPECTIVAS NA EDUCAÇÃO A DISTÂNCIA	
Milena Beatriz Silva Loubach Pollylian Assis Madeira Marcos Antônio Pereira Coelho Lucas Borcard Cancela	
DOI 10.22533/at.ed.9871926113	
CAPÍTULO 4	30
ANÁLISE COMPARATIVA ENTRE AS METODOLOGIAS DE ENSINO PEER INSTRUCTION E LECTURING	
Felipe Barbosa Araújo Ramos Antonio Alexandre Moura Costa Ademar França de Sousa Neto Luiz Antonio Pereira Silva Dalton Cézane Gomes Valadares Andressa Bezerra Ferreira	
DOI 10.22533/at.ed.9871926114	
CAPÍTULO 5	44
SISTEMA DE RECOMENDAÇÃO DE RECURSOS EDUCACIONAIS DIGITAIS, TENDO EM CONTA A DIVERSIDADE DOS ESTUDANTES	
Valentina Tabares Morales Néstor Darío Duque Méndez Yorely Bryjeth Ceballos Marta Rosecler Bez Silvana Vanesa Aciar	
DOI 10.22533/at.ed.9871926115	
CAPÍTULO 6	60
DIVULGAÇÃO CIENTÍFICA E ASTRONOMIA AMADORA NA ERA DA CONVERGÊNCIA DE MÍDIAS DIGITAIS: UMA ABORDAGEM DA EXPERIÊNCIA DO USUÁRIO	
Victor Alexandre Ferreira Luiz Agner	
DOI 10.22533/at.ed.9871926116	

CAPÍTULO 7	74
FÍSICA DA ALFACE: A PROMOÇÃO DA INTERDISCIPLINARIDADE ENTRE FÍSICA, INFORMÁTICA E FRUTICULTURA	
Lázaro Luis de Lima Sousa Sammya Kele Macena de Freitas Subênia Karine de Medeiros Neo	
DOI 10.22533/at.ed.9871926117	
CAPÍTULO 8	86
DOMÍNIOS DA TECNOLOGIA DA INFORMAÇÃO E COMUNICAÇÃO NA PROMOÇÃO E ATENÇÃO PRIMÁRIA À SAÚDE	
Diego Armando de Oliveira Meneses Adicinéia Aparecida de Oliveira	
DOI 10.22533/at.ed.9871926118	
CAPÍTULO 9	102
RASPBERRY PI COMO COMPUTADOR PARA USO ACADÊMICO NO IFRO <i>CAMPUS</i> PORTO VELHO ZONA NORTE	
Jhordano Malacarne Bravim Gabriel Augusto Fernandes Gonçalves Júlio Viana Filho Juliana Braz da Costa Ricardo Lopes Viera César	
DOI 10.22533/at.ed.9871926119	
CAPÍTULO 10	115
A PRIVACIDADE EM UM CENÁRIO <i>PANSENSITÍVEL</i> DE INTERNET DAS COISAS & CIDADES INTELIGENTES	
André Barbosa Ramiro Costa Maria Amália Oliveira de Arruda Câmara	
DOI 10.22533/at.ed.98719261110	
CAPÍTULO 11	129
ROBÔ AUTÔNOMO SEGUIDOR DE LINHA PARA AUXÍLIO RESIDENCIAL	
Gabriel Paiva Magalhães Wesley Miguel Dos Santos Peixoto	
DOI 10.22533/at.ed.98719261111	
CAPÍTULO 12	138
ALGORITMOS DE MACHINE LEARNING APLICADOS NA IDENTIFICAÇÃO DE GÊNERO POR MEIO DE FREQUÊNCIA DE VOZ	
Maicon Facco Daíse dos Santos Vargas Marcos Antônio de Azevedo de Campos Cleber Bisognin	
DOI 10.22533/at.ed.98719261112	
CAPÍTULO 13	151
O QUE PODEM OS ALGORITMOS?	
Gabrielle Granadeiro da Silveira	
DOI 10.22533/at.ed.98719261113	

CAPÍTULO 14 163

RECUPERAÇÃO DO ESTANHO PRESENTE EM PLACAS DE CIRCUITO IMPRESSO (PCI'S)
VISANDO À PRESERVAÇÃO DOS RECURSOS NATURAIS

Maria do Socorro Bezerra da Silva
Carlson Pereira Souza
André Luis Lopes Moriyama
Raffael Andrade Costa de Melo

DOI 10.22533/at.ed.98719261114

CAPÍTULO 15 175

INVESTIGAÇÃO DE MATERIAIS A BASE DE GRAFENO E HIDROTALCITA APLICADOS COMO
ADSORVENTES PARA REMOÇÃO BIFUNCIONAL DE MICROPOLUENTES EM ÁGUA

Eliane Kujat Fischer
Cintia Hisano
Rafael Aparecido Ciola Amoresi
Maria Aparecida Zaghete Bertochi
Rony Gonçalves Oliveira
Alberto Adriano Cavalheiro

DOI 10.22533/at.ed.98719261115

CAPÍTULO 16 188

A VOLTA MAIS RÁPIDA PARA OBTER REDUÇÃO DE CUSTO NA CONSTRUÇÃO CIVIL
CONSIDERANDO O CICLO DE VIDA DA EDIFICAÇÃO

Ivan Luiz Portugal Basile
Eduardo Ioshimoto
André Luiz Gonçalves Scabbia

DOI 10.22533/at.ed.98719261116

CAPÍTULO 17 204

DESENVOLVIMENTO E VIDA ÚTIL DE FARINHA A PARTIR DOS RESÍDUOS GERADOS NO
PROCESSAMENTO DE MÍNIMO DE CENOURA

Rosa Maria de Deus de Sousa
Celso Luiz Moretti
Cristina Maria Monteiro Machado
Leonora Mansur Mattos

DOI 10.22533/at.ed.98719261117

CAPÍTULO 18 217

VALIDAÇÃO DA TÉCNICA DE MODELAGEM COMPUTACIONAL PARAMÉTRICA BIDIMENSIONAL
SOB CAMPO DE VENTO UNIFORME

Marcelo Marques
Fernando Oliveira de Andrade
Elaine Patrícia Arantes
Isabela Arantes Ferreira
Tobias Bleninger
Alexandre Kolodynskie Guetter

DOI 10.22533/at.ed.98719261118

CAPÍTULO 19	228
ANÁLISE MULTITEMPORAL DA MALHA VIÁRIA DO ESTADO DA PARAÍBA COM A UTILIZAÇÃO DE TÉCNICAS E PROCEDIMENTOS DA CARTOGRAFIA DIGITAL E DO SENSORIAMENTO REMOTO	
Edmilson Roque da Silva Junior	
Emanoel Ferreira Cardoso	
Gilanildo Freires de Almeida	
Marcelo Laédson Morato Ferreira	
Renan Willer Pinto de Sousa	
DOI 10.22533/at.ed.98719261119	
CAPÍTULO 20	238
MAPEAMENTO LITO-ESTRUTURAL DA REGIÃO DE GURJÃO-PB	
Thayná Bel Pereira Guimarães	
Natanael Felipe Lorenzi de Oliveira	
DOI 10.22533/at.ed.98719261120	
CAPÍTULO 21	245
STUDY OF LIGHTNING BIFURCATION AND EFFECT ON RADIATION	
Fernando Júnio de Miranda	
DOI 10.22533/at.ed.98719261121	
CAPÍTULO 22	255
ESTIMATIVAS DO PODER EVAPORANTE DO AR PARA OS MUNICÍPIOS DE NOVO REPARTIMENTO E SANTANA DO ARAGUAIA NO ESTADO DO PARÁ	
Jocilene Teixeira do Nascimento	
Maria do Bom Conselho Lacerda Medeiros	
Valdeides Marques Lima	
Luane Laíse Oliveira Ribeiro	
Paulo Jorge de Oliveira Ponte de Souza	
Joaquim Alves de Lima Júnior	
Fabio Peixoto Duarte	
Helane Cristina Aguiar Santos	
Wellington Leal dos Santos	
Bianca Cavalcante da Silva	
DOI 10.22533/at.ed.98719261122	
CAPÍTULO 23	266
UTILIZAÇÃO DO MODELO TOPMODEL PARA ANÁLISE TEMPORAL DO SISTEMA CHUVA-VAZÃO NA BACIA DO RIO SÃO MIGUEL	
Ciro Couto Bento	
Cristiano Christofaro Matosinhos	
Welberth Pereira Dias	
Thiago Martins da Costa	
Hernando Baggio	
DOI 10.22533/at.ed.98719261123	
CAPÍTULO 24	279
STRENGTH PREDICTION OF ADHESIVELY-BONDED JOINTS WITH COHESIVE LAWS ESTIMATED BY THE DIRECT METHOD	
Ulisses Tiago Ferreira Carvalho	
Raul Duarte Salgueiral Gomes Campilho	
DOI 10.22533/at.ed.98719261124	

CAPÍTULO 25 292

PRINCIPAIS MATERIAIS E TÉCNICAS UTILIZADOS NA OXIDAÇÃO DE ÁLCOOL PARA USO EM CÉLULAS A COMBUSTÍVEL: UMA REVISÃO

Isaide de Araujo Rodrigues
Ziel Dos Santos Cardoso
Deracilde Santana da Silva Viégas
Vinicius Tribuzi Rodrigues Pinheiro Gomes

DOI 10.22533/at.ed.98719261125

CAPÍTULO 26 305

USO DE LISTAS DINÂMICAS EM APLICATIVO MÓVEL PARA INTERPOLAÇÃO DE DADOS DE TEMPERATURA DO AR, VISANDO O CONFORTO TÉRMICO

Arlson José de Oliveira Júnior
Silvia Regina Lucas de Souza
Guilherme dos Santos Sousa
William Duarte Bailo
Daniel de Oliveira

DOI 10.22533/at.ed.98719261126

SOBRE OS ORGANIZADORES..... 315

ÍNDICE REMISSIVO 316

ALGORITMOS DE MACHINE LEARNING APLICADOS NA IDENTIFICAÇÃO DE GÊNERO POR MEIO DE FREQUÊNCIA DE VOZ

Maicon Facco

Universidade Federal de Santa Maria
Santa Maria – Rio Grande do Sul

Daíse dos Santos Vargas

Universidade Federal de Santa Maria
Santa Maria – Rio Grande do Sul

Marcos Antônio de Azevedo de Campos

Universidade Federal de Santa Maria
Santa Maria – Rio Grande do Sul

Cleber Bisognin

Universidade Federal de Santa Maria
Santa Maria – Rio Grande do Sul

MACHINE LEARNING ALGORITHMS APPLIED IN GENDER IDENTIFICATION BY VOICE FREQUENCY

ABSTRACT: *Machine learning* techniques are widely applied in the most diverse areas. Through such techniques, it is possible to identify a person's gender through the frequency of his voice. Five supervised learning techniques were applied to 3000 voice frequency data aiming to select the model with the best performance in gender classification. The model selected was the Support Vector Machine (SVM), with an average accuracy of 97.71% of correct answers. **KEYWORDS:** Machine Learning, gender, voice.

RESUMO: Técnicas de machine learning são amplamente aplicadas nas mais diversas áreas. Por meio de tais técnicas, é possível identificar o gênero de uma pessoa por meio da frequência de sua voz. Foram aplicadas 5 técnicas de aprendizado supervisionado em 3000 dados de frequência de voz objetivando selecionar o modelo com melhor desempenho na classificação de gênero. O modelo selecionado foi o *Support Vector Machine* (SVM), com uma precisão média de 97,71% de acertos nas classificações.

PALAVRAS-CHAVE: Machine learning, gênero, voz.

1 | INTRODUÇÃO

Machine Learning (Aprendizado de Máquina) permite que computadores tomem decisões a partir da ajuda de algoritmos que reconhecem padrões e tornam-se capazes de fazer previsões. Um exemplo de previsão seria uma máquina poder classificar um objeto de acordo com informações armazenadas em sua memória, sejam imagens com cores, imagens de vários ângulos, formas, etc., ou seja, com base em padrões de dados pré-fixados. *Machine Learning* é uma área de grande importância ligada a Inteligência Artificial (IA) pois, segundo Batista (2003), a capacidade de

aprender é necessária para um comportamento inteligente. Mitchel (1997) afirma que a área de *Machine Learning* estuda métodos computacionais visando adquirir novos conhecimentos e habilidades, assim como novos meios de organizar o conhecimento existente.

A ideia principal seria de que uma máquina deveria aprender com seus erros. Se uma máquina pudesse “observar” determinada ocorrência, deveria ser capaz de armazenar aquela informação determinando um desfecho para ela. Quando observada uma situação que não se assemelha com a mesma ocorrência, a máquina deveria identificar que esta não faz parte das ocorrências similares a que foi observada anteriormente.

É possível, por meio de generalizações e especificações, criar algoritmos que induzam uma hipótese cujo conceito é consistente com todos os exemplos. Muitos algoritmos de *Machine Learning* utilizam operações de generalização e especificação para criar hipóteses a partir de exemplos. Segundo interpretação da definição de Bratko (1990), uma hipótese somente pode ser útil se puder ser utilizada no reconhecimento correto de exemplos além dos utilizados na indução da hipótese.

Além de tomar decisões, pode fazer uso de algoritmos para fazer previsões nas mais diversas áreas, como por exemplo, na bolsa de valores, ou na previsão da ocorrência de uma determinada doença baseado nos históricos de saúde de pacientes que já sofreram da mesma, ou até mesmo realizar a previsão do gênero de uma pessoa baseado na frequência da voz do indivíduo, que é o foco deste trabalho.

2 | REVISÃO DA LITERATURA

2.1 Sistemas de aprendizado

A partir de dados fornecidos por um processo externo é que o aprendizado é efetuado. Sendo assim, no contexto de *Machine Learning*, a máquina programada toma decisões baseada em experiências acumuladas contidas em casos resolvidos com sucesso (WEISS & KULIKOWSKI, 1991). Conforme Monard & Baranauskas (2003), existem dois principais tipos de aprendizado de máquina: aprendizado supervisionado e aprendizado não supervisionado. Para a realização neste trabalho foi usada a aprendizagem supervisionada.

2.1.1 Aprendizado supervisionado

No aprendizado supervisionado é informado ao sistema computacional (sistema de aprendizado) um conjunto de dados $\mathbf{D} = (D_1, D_2, \dots, D_n)$, onde cada dado $D_i \in \mathbf{D}$, para $i = 1, \dots, n$, possui uma classificação associada. Essa classificação define qual a classe que o dado pertence. Ou seja,

$$D = \left((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \right),$$

onde $D_i = (x_i, y_i)$, no qual x_i é um vetor de valores que representam os atributos do dado D_i e y_i é o valor da classe desse dado. O objetivo deste tipo de aprendizado é induzir um mapeamento geral dos vetores x para os valores y . A partir disso, o objetivo é construir um modelo $y = g(x)$ que permite prever valores y para dados previamente não observados. Mas na realidade, o aprendizado supervisionado não é capaz de construir com exatidão tal função g , mas sim uma função aproximada de g , pois o número de dados normalmente não é suficiente para caracterizar esta função.

Em outras palavras, de posse de um banco de dados rotulados, já se sabe qual é a saída correta e que esta deve ser semelhante ao banco, demonstrando uma relação entre a entrada e a saída. Enquadram-se aqui os problemas de regressão e classificação. Em problemas de regressão, tenta-se prever os resultados em uma saída contínua, ou seja, há um mapeamento para alguma função contínua. Já nos problemas de classificação, a previsão é focada na previsão de resultados em saídas discretas, ou seja, há um mapeamento de variáveis de entrada em categorias distintas.

2.2 Machine learning para dados de classificação

Um grande problema no *Machine Learning* (aprendizado de máquina) é referente à classificação. Segundo Izbicki, trata-se de um problema muito parecido com o de fazer previsões em regressão.

Dada uma amostra $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n) \sim (X, Y)$, o objetivo do Machine Learning é elaborar uma função $g(x)$ que possa ser utilizada para fazer previsão de novas observações $(X_{m+1}, Y_{m+1}), (X_{m+2}, Y_{m+2}), \dots, (X_n, Y_n)$, ou seja, queremos elaborar uma função tal que $g(x_{m+1}) \approx y_{m+1}, g(x_{m+2}) \approx y_{m+2}, \dots, g(x_n) \approx y_n$. Aqui o problema é que a variável resposta Y é qualitativa, tornando-se assim um problema de classificação, ao contrário dos problemas de regressão, nos quais a variável resposta é quantitativa.

O estudo de regressão para variável resposta quantitativa baseia-se em uma função $R(g(x)) = E[(Y - g(X))^2]$, denominada função de risco. Quando se trata de variáveis qualitativas, esta função de risco já não faz mais sentido, visto que agora a variável resposta é qualitativa, ou seja, não se aplica a um problema de classificação.

Para esse tipo de problema, sugere-se como função de risco

$$R(g(x)) = E[I(Y \neq g(X))] = P(Y \neq g(X)),$$

ou seja, a probabilidade de erro em uma nova observação (X, Y) , onde $I(Y \neq g(X))$ é a função indicadora, também conhecida como função de perda 0-1.

A melhor função classificadora $g(x)$ segundo a função de risco $R(g)$ é dada por

$$g(\mathbf{x}) = \arg \max_{d \in \mathcal{C}} P(Y = d|\mathbf{x}),$$

onde \mathcal{C} é um conjunto de valores de classificação e d um valor qualquer de \mathcal{C} . Em outras palavras, deve-se classificar x como sendo da classe com maior probabilidade a posteriori. A função classificadora $g(x)$ é também conhecida como classificador de Bayes.

Neste trabalho faz-se uso de validação cruzada para estimar o risco dos métodos de classificação. O banco de dados deve ser dividido aleatoriamente em três partes: treinamento, validação e teste. Preferencialmente a parte do banco de dados direcionada ao treinamento e validação deve ser de $\frac{2}{3}$ ou $\frac{70\%}{100\%}$, e a parte do banco de dados direcionada ao teste de $\frac{1}{3}$ ou $\frac{30\%}{100\%}$, embora possa ser dividido em outros tamanhos.

A divisão do banco de dados deve ser feita aleatoriamente, conforme verifica-se abaixo, por exemplo, em uma amostra de tamanho n :

$$\begin{array}{c} \text{Treinamento e validação (70\%)} \\ \hline \begin{array}{ccc} \text{Treinamento (70\%)} & \text{Validação (30\%)} & \text{Teste (30\%)} \\ \hline (\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_s, Y_s), & (\mathbf{X}_{s+1}, Y_{s+1}), \dots, (\mathbf{X}_m, Y_m), & (\mathbf{X}_{m+1}, Y_{m+1}), \dots, (\mathbf{X}_n, Y_n) \end{array} \end{array}$$

Para fazer a avaliação de modelos, ao se tomar uma decisão sobre qual apresenta o melhor desempenho, é necessário avaliar acertos e/ou erros em exemplos fora do conjunto de treinamento, ou seja, no conjunto de testes. A importância de se avaliar um modelo é verificar se ele é bom o suficiente para ser útil e se seu desempenho é mais satisfatório que o de outros modelos.

Usa-se o conjunto de treinamento para estimar a função classificadora $g(x)$ e usa-se o conjunto de validação para estimar $R(g(x))$ através da fórmula

$$\hat{R}(g(\mathbf{x})) := \frac{1}{m-s} \sum_{i=s+1}^m I(Y_i \neq g(\mathbf{X}_i)),$$

ou seja, avaliar a proporção de erros no conjunto de validação. Em outras palavras, utiliza-se o conjunto de treinamento para fazer a aprendizagem do algoritmo e o conjunto de validação para estimar os parâmetros do modelo. Depois do algoritmo ser treinado, avalia-se sua generalização sobre o conjunto de testes.

Com o intuito de calcular o desempenho médio de um modelo é necessário repetir o processo de treinamento k vezes para se estimar distintos parâmetros através de taxa de acerto médio e variância, e isto pode ser feito através de métodos de validação cruzada, sendo usado neste trabalho o método *k-fold cross validation*, conforme apresentado na seção a seguir.

2.3 Método de validação cruzada k-fold

O método de validação cruzada *k-fold* consiste em particionar o conjunto de dados de treinamento e validação de tamanho m em k subconjuntos mutuamente exclusivos (ou seja, cada subconjunto terá um tamanho de m/k , e posteriormente, o uso de $k - 1$ subconjuntos para a estimação dos parâmetros do modelo (dados de treinamento), fazendo-se o cálculo da acurácia do modelo, sendo o subconjunto restante (dados de validação ou de teste) empregados na validação do modelo. O algoritmo é treinado k vezes, cada vez com um conjunto de dados diferente sendo deixado de fora para a realização da validação. Após, estima-se o desempenho por meio do erro médio ou taxa de acerto média sobre os k conjuntos de validação.

2.4 Medidas de performance

Por vezes, a função de risco $R(g(\mathbf{x})) = E[I(Y \neq g(\mathbf{X}))] = P(Y \neq g(\mathbf{X}))$ não traz toda a informação sobre a razoabilidade de g . Para evitar este tipo de problema, é comum gerar juntamente ao modelo escolhido na fase de treinamento e validação as chamadas matrizes de confusão.

De acordo com (PRIDA, 2015), uma matriz de confusão é uma forma de representação da qualidade obtida em um modelo, sendo expressa por meio da correlação de informações dos dados de referência com os dados classificados.

Suponha a matriz de confusão abaixo 2×2 , onde V indica verdadeiro, F indica falso, P indica positivo (classificação concordante) e N indica negativo (classificação discordante).

	Valor Verdadeiro	
Valor Predito	$Y=0$	$Y=1$
$Y=0$	VN	FN
$Y=1$	FP	VP

Baseado na matriz acima, são definidas algumas medidas relacionadas aos modelos usados para Machine Learning, a saber:

- Sensibilidade: $S = VP / (VP + FN)$;
- Especificidade: $E = VN / (VN + FP)$;
- Valor Preditivo Positivo: $VPP = VP / (VP + FP)$;
- Valor Preditivo Negativo: $VPN = VN / (VN + FN)$.

Uma outra medida comum nos modelos de *Machine Learning* é a denominada Acurácia ou Precisão. Tal medida é simplesmente a proporção de dados que foram classificados corretamente em toda a matriz de confusão. É muito útil na seleção de melhores modelos de classificação para determinadas variáveis quando estes estão

sendo comparados.

Também há um outro coeficiente denominado coeficiente Kappa, e é uma das variáveis que podem ser quantificadas após construir a matriz de confusão, sendo um índice que retrata o grau de concordância dos dados, gerando, assim, um aspecto de confiabilidade e precisão dos dados classificados (PERROCA e GAIDZINSKI, 2003). O valor do coeficiente Kappa varia no intervalo [0, 1], sendo que quanto mais próximo de 1, melhor a qualidade das classificações dos dados.

2.5 ALGUNS MÉTODOS DE CLASSIFICAÇÃO

Alguns métodos de classificação são baseados nos chamados classificadores plug-in. Tais classificadores consistem em estimar a quantidade $P(Y = c|\mathbf{x})$ e tomar como função classificadora $g(\mathbf{x}) = \arg \max_{c \in \mathcal{C}} \hat{P}(Y = c|\mathbf{x})$. Esta função classificadora conecta o estimador da probabilidade condicional na fórmula do classificador $g(\mathbf{x})$ ótimo. ou seja, o conceito consiste na criação de classificadores, o que é equivalente a estimar o valor de $P(Y = c|\mathbf{x})$.

2.5.1 ANÁLISE DISCRIMINANTE LINEAR (LDA)

Assume-se na análise discriminante linear que cada distribuição condicional $\mathbf{X}|Y = c$ segue uma distribuição normal multivariada. As distribuições, embora possuam médias distintas, têm a mesma matriz de variância e covariância.

Assim,

$$f(\mathbf{x}|Y = c) = \frac{1}{\sqrt{(2\pi)^d} |\Sigma^{-1}|} e^{-(\mathbf{x}-\mu_c)' \Sigma^{-1} (\mathbf{x}-\mu_c)}$$

A estimação dos parâmetros é feita através do método de máxima verossimilhança:

$$\hat{\mu}_c = \frac{1}{|\mathcal{C}_c|} \sum_{k \in \mathcal{C}_c} X_k; \hat{\Sigma} = \frac{1}{n} \sum_{c \in \mathcal{C}} \sum_{k \in \mathcal{C}_c} (\mathbf{x}_k - \hat{\mu}_c) (\mathbf{x}_k - \hat{\mu}_c)',$$

onde $\mathcal{C}_c = \{j = 1, \dots, n : Y_j = c\}$.

Constrói-se então o classificador plug-in através das estimativas de $\hat{f}(\mathbf{x}|Y = c)$ por meio de

$$g(\mathbf{x}) = \arg \max_{c \in \mathcal{C}} \hat{P}(Y = c|\mathbf{x}) = \arg \max_{c \in \mathcal{C}} \hat{f}(\mathbf{x}|Y = c) \hat{P}(Y = c).$$

No caso de classificações binárias (casos de “sim” ou “não”, “0” ou “1”), utiliza-se como regra de classificação a função $g(\mathbf{x})=1$, que é uma regra de decisão em um

hiperplano de \mathbb{R}^d . Isso vem do fato de que, após a realização de alguns cálculos,

$$\frac{\hat{P}(Y = 1|\mathbf{x})}{\hat{P}(Y = 0|\mathbf{x})} \geq L \Leftrightarrow \frac{\hat{f}(\mathbf{x}|Y = 1)\hat{P}(Y = 1)}{\hat{f}(\mathbf{x}|Y = 0)\hat{P}(Y = 0)} \geq L \geq \alpha\mathbf{x}' \geq L''.$$

onde L, L' e L'' são constantes.

2.5.2 SUPPORT VECTOR MACHINES(SVM)

Proposto por Cortes and Vapnik (1995), o support vector machine é um método de classificação que se diferencia de outros métodos principalmente porque em momento algum estimam-se as probabilidades $\hat{P}(Y = c|\mathbf{x})$.

Considere uma função linear $f(\mathbf{x}) := \beta_0 + \beta_1x_1 + \dots + \beta_dx_d$ e seja Y assumindo valores em um conjunto $\mathcal{C} = \{-1, 1\}$.

O classificador $g(\mathbf{x})$ dado pelo support vector machine é dado por

$$\begin{cases} g(\mathbf{x}) = -1, & \text{se } f(\mathbf{x}) < 0, \\ g(\mathbf{x}) = 1, & \text{se } f(\mathbf{x}) \geq 0. \end{cases}$$

Supõe-se que existe um hiperplano que separa perfeitamente as observações do conjunto de treinamento segundo sua classificação (classe à qual pertencem), ou seja, descreve-se a existência de uma função $f(\mathbf{x})$ linear tal que $f(\mathbf{x}_i) < 0$ se, e somente se, $y_i = -1$. Logo, para todo $i = 1, \dots, n$, $y_i(\beta_0 + \beta_1x_{i1} + \dots + \beta_dx_{id}) = y_if(\mathbf{x}_i) > 0$.

O SVM, na presença de muitos hiperplanos, busca por aquele que fica mais distante de todos os pontos observados. Se $f(\mathbf{x})$ separa perfeitamente os dados, tem-se que $|f(\mathbf{x})| = y_if(\mathbf{x}_i)$, dado que $|f(\mathbf{x})|$ é a distância do ponto \mathbf{x} à reta dada por $f(\mathbf{x})$. Ou seja, o SVM busca $f(\mathbf{x})$ tal que $y_if(\mathbf{x}_i)$ seja grande para todo \mathbf{x}_i , pois se $|f(\mathbf{x})|$ é muito alto, \mathbf{x} está muito longe do plano separador. Assim, caso haja um hiperplano que separa perfeitamente os dados, o SVM busca o hiperplano com coeficientes $\boldsymbol{\beta}$ tal que

$$(\boldsymbol{\beta}, M) = \arg \max_{\boldsymbol{\beta}, M} M$$

sujeito às restrições:

(i) $\sum_{i=1}^d \beta_i^2 = 1$ e

(ii) para todo $i = 1, \dots, n$, $y_if_{\boldsymbol{\beta}}(\mathbf{x}_i) \geq M(1 - \epsilon_i)$, em que $\epsilon_i > 0$ e $\sum_{i=1}^n \epsilon_i \leq C$

Enfatiza-se aqui que pode ser maior que um, permitindo assim que $y_if_{\boldsymbol{\beta}}(\mathbf{x}_i)$ seja negativo, o que significa que a i -ésima amostra pode cair no lado errado do hiperplano, ou seja, pode ser mudada sua classificação. C é conhecido como um tuning parameter, ou seja, quanto maior seu valor, maior a chance de cair no lado errado do hiperplano

as amostras têm, e é um limitante para as observações, além de sempre ser maior que ϵ_i .

2.5.3 MÉTODO DOS k VIZINHOS MAIS PRÓXIMOS (k NN)

Um dos métodos mais populares no Machine Learning (BENEDETTI, 1977), o método dos k vizinhos mais próximos. Este método tem como base estimar a função de classificação $g(x)$ para uma dada configuração das covariáveis Y dos k -vizinhos mais próximos de x .

Define-se como classificador a função

$$g(x) = \text{moda}_{i \in \mathcal{N}_x} y_i,$$

em que \mathcal{N}_x é o conjunto das k -observações mais próximas de x , isto é,

$$\mathcal{N}_x = \{i \in \{1, \dots, n\}: d(x_i, x) \leq d_x^k\},$$

e d_x^k é a distância do k -ésimo vizinho mais próximo de x até x . Ou seja, busca-se entre as observações mais próximas do vetor de covariáveis x a classe mais frequente que foi observada.

2.5.4 ÁRVORES DE CLASSIFICAÇÃO E REGRESSÃO (CART)

Uma árvore cria uma partição do espaço das covariáveis em regiões distintas e disjuntas R_1, \dots, R_j . Sendo assim, a predição para a variável resposta Y de uma observação com covariáveis x que estão em R_k é dada por

$$g(x) = \text{moda}\{y_i: x_i \in R_k\},$$

ou seja, para realizar a classificação de x , observa-se a qual região pertence a observação x e, então, calcula-se a moda amostral do conjunto de treinamento que pertencem à essa região.

A criação de uma árvore de classificação é feita por meio de duas etapas. Na primeira, busca-se criar uma árvore de tal forma que leve a partições puras, partições onde os valores de Y nos conjuntos de treinamento em cada uma das folhas sejam homogêneos, construindo uma árvore completa e complexa; a segunda etapa é a poda da árvore, evitando assim um sobreajuste, que é quando um modelo estatístico se ajusta muito bem ao conjunto de dados, porém é ineficaz para previsão de novos resultados.

Na busca da melhor partição, busca-se minimizar o índice de Gini, dado por

$$\mathcal{P}(T) = \sum_R \sum_{c \in \mathcal{C}} \widehat{p}_{R,c} (1 - \widehat{p}_{R,c})$$

onde R é uma das regiões induzidas pela árvore e $\widehat{p}_{R,c}$ é a proporção de observações classificadas como sendo da categoria c entre as que caem na região R . A escolha do índice de Gini se dá por este ser mais sensível a mudanças nas proporções de cada categoria nos nós.

Obtém-se uma árvore pura quando o índice de Gini é mínimo, e isso ocorre quando todas as proporções são 0 ou 1. Sendo assim, a ocorrência de uma árvore pura significa que cada folha contém somente observações de uma única classe.

2.5.5 FLORESTAS ALEATÓRIAS (RF)

Esta abordagem consiste em criar árvores distintas e combinar seus resultados para melhorar as predições dadas pelas árvores individualmente. A ideia principal na criação de florestas aleatórias é modificar o método de criação das árvores de classificação para que elas se tornem diferentes umas das outras. Sendo assim, em cada passo é permitido que seja escolhida somente uma dentre $m < d$ covariáveis para ser utilizada em cada um dos nós da árvore, ao invés de decidir qual das d covariáveis será utilizada em cada um dos nós, onde as m covariáveis são escolhidas aleatoriamente dentre as covariáveis originais, e é sorteado um novo subconjunto de covariáveis a cada nó criado. O valor de m , segundo (IZBICKI) pode ser escolhido por meio do método de validação cruzada, mas geralmente é feito por meio da fórmula $m \approx \sqrt{d}$. O valor de m é mantido constante durante o processo de crescimento das árvores e cada árvore é “cultivada” em suas maiores extensões possíveis, não havendo poda em nenhum momento.

Para criar as B árvores distintas, utilizam-se B amostras bootstrap da amostra original. Para cada uma das amostras criadas, cria-se uma árvore utilizando as técnicas usadas em árvores aleatórias.

3 | METODOLOGIA

Os dados para este artigo foram coletados de milhares de amostras de vozes, cada uma sendo classificada em masculina ou feminina. A base de dados para este trabalho foi obtida no Kaggle (<https://www.kaggle.com>), plataforma online de competições de Data Science públicas, privadas e acadêmicas, onde as competições são patrocinadas por empresas e oferecem prêmios financeiros pela melhor solução. Para não utilizar o mesmo banco de dados da plataforma, foi selecionada de forma aleatória uma subamostra das 3168 amostras de vozes, falas e expressões masculinas e femininas, resultando em uma base de dados com 3000 amostras de vozes, composta por 1109 homens e 992 mulheres.

As amostras foram processadas usando análise acústica e utilizando 20 variáveis, sendo elas: frequência média (meanfreq), desvio padrão da frequência (sdfreq), frequência mediana (medianfreq), primeiro quantil (q25), terceiro quantil (q75), intervalo interquantílico (IQR), assimetria (assim), curtose (kurt), entropia espectral (sp.ent), planicidade espectral (sfm) , moda frequencial (modafreq), centroide de frequência (centroide), média da frequência fundamental medida no sinal acústico (meanfun), frequência fundamental mínima no sinal acústico medida por meio do sinal acústico (minfun), frequência fundamental máxima no sinal acústico medida por meio do sinal acústico (maxfun), média da frequência dominante medida no sinal acústico (meandom), mínimo da frequência dominante medida no sinal acústico (mindom), máximo da frequência dominante medida no sinal acústico (maxdom), faixa de frequência dominante medida no sinal acústico (dfrange), e índice de modulação (modindx).

Por meio da aprendizagem supervisionada, aplicou-se a técnica de *Machine Learning* para realizar a previsão do gênero de cada pessoa da amostra, e dividiu-se a base de dados em duas partes, sendo 70% para treinamento e validação e 30% para a fase de teste, selecionados aleatoriamente. Foram testados 5 diferentes métodos de classificação, aplicados através do método de validação cruzada *k-fold*, em 10 blocos, e utilizou-se a medida de precisão Acurácia, visto que, conforme a literatura, é a melhor medida quando a variável de interesse é de classificação, como no caso da variável gênero. Para este estudo foi utilizado exclusivamente o software R Core Team (2018). utilizando o comando `set.seed(1234)` como semente para a aleatorização dos dados.

4 | RESULTADOS E DISCUSSÃO

As medidas descritivas dos modelos utilizados no banco de dados podem ser verificadas por meio da Tabela 1.

	Acurácia					
	Mínimo	1° Quartil	Mediana	Média	3° Quartil	Máximo
LDA	0,948	0,957	0,962	0,964	0,969	0,986
CART	0,943	0,957	0,964	0,962	0,970	0,976
KNN	0,681	0,712	0,722	0,727	0,746	0,781
SVM	0,962	0,968	0,979	0,977	0,985	0,995
RF	0,957	0,973	0,979	0,976	0,981	0,991
	Coeficiente Kappa					
	Mínimo	1° Quartil	Mediana	Média	3° Quartil	Máximo
LDA	0,895	0,914	0,923	0,927	0,938	0,971
CART	0,885	0,914	0,929	0,924	0,940	0,952
KNN	0,355	0,423	0,440	0,450	0,488	0,558
SVM	0,924	0,936	0,957	0,954	0,969	0,990
RF	0,914	0,945	0,957	0,952	0,962	0,981

Tabela 1 – Medidas descritivas dos modelos de Machine Learning utilizados.

Fonte: Os Autores.

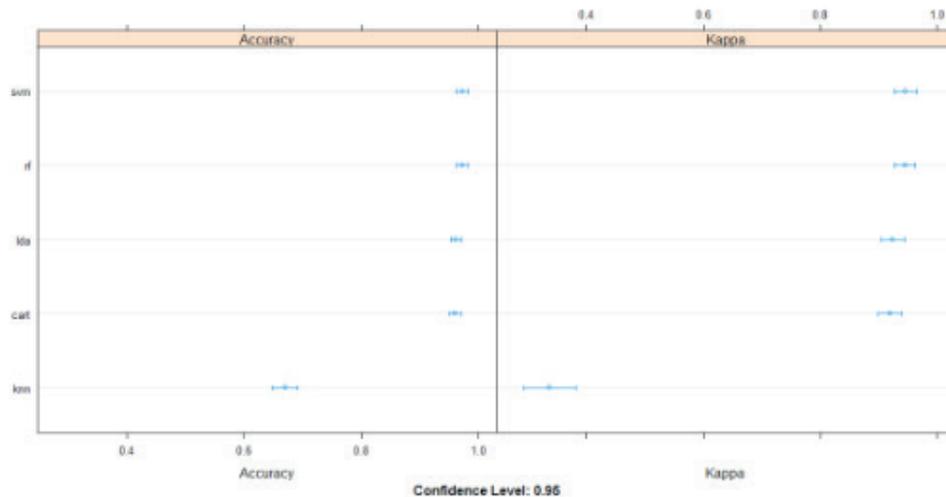


Figura 1 - Comparação entre os intervalos de confiança dos algoritmos de Machine Learning para os parâmetros Acurácia e Coeficiente Kappa

Fonte: Os Autores.

Com base na Tabela 1, observa-se que o modelo que obteve melhor acurácia e melhor coeficiente Kappa foi o SVM, seguido dos modelos RF, LDA e CART, respectivamente. O modelo com os piores índices foi o modelo KNN. A partir da Figura 1 é possível visualizar o ordenamento dos modelos, por meio dos respectivos intervalos de confiança, de acordo com a acurácia e o coeficiente Kappa. Posteriormente, testou-se os índices dos modelos utilizando os 30% da base de dados sorteados para validação. Tal validação ocorreu utilizando-se a matriz de confusão. Como resultado tem-se que o modelo SVM apresentou os melhores índices, quando comparado aos outros 4 modelos, tanto para especificidade, quanto sensibilidade.

O resultado final para validação deste modelo está na Tabela 2, a qual apresenta a matriz de confusão com seus respectivos valores. Com base nessa matriz tem-se que a acurácia desse modelo é de 0,9855, podendo variar entre 0,9754 e 0,9923 num intervalo de confiança de 95%. O resultado do coeficiente Kappa é de 0,9710. A especificidade desse modelo é de 99,16%, e sua sensibilidade de 97,88%.

Valor predito	Valor verdadeiro	
	Feminino	Masculino
Feminino	415	4
Masculino	9	471

Tabela 2 – Matriz de confusão do modelo SVM.

Fonte: Os Autores.

A partir da Tabela 2 verifica-se também que 415 mulheres e 471 homens tiveram sua classificação correta, enquanto 13 pessoas tiveram sua classificação trocada. Isso mostra que, de fato, o modelo SVM teve desempenho superior.

5 | CONCLUSÃO

A classificação de uma frequência de voz é complexa, e encontrar um modelo que realize tal distinção é uma tarefa que exige tempo e paciência. No entanto, ao fazer o uso de técnicas de *Machine Learning*, este trabalho é simplificado e realizado em um curto espaço de tempo. Dentre os cinco modelos aplicados, o que obteve o melhor desempenho foi o *Support Vector Machines*, com 97,71% de precisão média nos dados referentes a treinamento e validação, e precisão de 98,45% nos dados referentes à base de teste. O método dos k vizinhos mais próximos teve o pior desempenho (72,68% de precisão) e não sendo recomendado para a classificação de gêneros baseada na frequência de voz, utilizando as variáveis propostas.

A escolha do modelo SVM está diretamente relacionada ao banco de dados específico. Em outras palavras, este modelo não deve ser considerado como um algoritmo definitivo para o tratamento de frequências de voz e classificação de gênero, pois caso haja uma classificação que pareça incorreta, esta deve ser verificada. Um banco de dados com 3000 informações é uma amostra muito pequena comparada com o número de pessoas que habitam o planeta, o que implica que este modelo serve somente para fins de pesquisa, e não para classificações de gênero definitivas baseado nas frequências de voz.

REFERÊNCIAS

- BATISTA, Gustavo Enrique de Almeida Prado et al. **Pré-processamento de dados em aprendizado de máquina supervisionado**. 2003. Tese de Doutorado. Universidade de São Paulo.
- BENEDETTI, Jacqueline K. On the nonparametric estimation of regression functions. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 39, n. 2, p. 248-253, 1977.
- BRATKO, Ivan. **Prolog programming for artificial intelligence**. Pearson education, 2001.
- CORTES, Corinna; VAPNIK, Vladimir. Support-vector networks. **Machine learning**, v. 20, n. 3, p. 273-297, 1995.
- IZBICKI, Rafael; DOS SANTOS, Tiago Mendonça. **Machine Learning sob a ótica estatística**, 2017.
- MITCHELL, Tom M. et al. Machine learning. 1997. **Burr Ridge, IL: McGraw Hill**, v. 45, n. 37, p. 870-877, 1997.
- MONARD, Maria Carolina; BARANAUSKAS, José Augusto. Conceitos sobre aprendizado de máquina. **Sistemas inteligentes-Fundamentos e aplicações**, v. 1, n. 1, p. 32, 2003.
- PERROCA, Márcia Galan; GAIDZINSKI, Raquel Rapone. Avaliando a confiabilidade interavaliadores de um instrumento para classificação de pacientes: coeficiente Kappa. **Revista da Escola de Enfermagem da USP**, v. 37, n. 1, p. 72-80, 2003.
- PRINA, Bruno Zucuni; TRENTIN, Romario. GMC: Geração de Matriz de Confusão a partir de uma classificação digital de imagem do ArcGIS®. **Anais do XVII Simpósio Brasileiro de Sensoriamento Remoto-SBSR**, p. 137, 2015.

SIMON, Herbert A. Por que as máquinas aprendem? In: **Aprendizado de máquina**. Morgan Kaufmann, 1983. p. 25-37.

WEISS, Sholom M.; KULIKOWSKI, Casimir A. **Computer systems that learn: classification and prediction methods from statistics, neural nets, machine learning, and expert systems**. Morgan Kaufmann Publishers Inc., 1991.

SOBRE OS ORGANIZADORES

Henrique Ajuz Holzmann: Professor da Universidade Tecnológica Federal do Paraná (UTFPR). Graduação em Tecnologia em Fabricação Mecânica e Engenharia Mecânica pela Universidade Tecnológica Federal do Paraná. Mestre em Engenharia de Produção pela Universidade Tecnológica Federal do Paraná. Doutorando em Engenharia e Ciência dos Materiais pela Universidade Estadual de Ponta Grossa. Trabalha com os temas: Revestimentos resistentes a corrosão, Soldagem e Caracterização de revestimentos soldados.

João Dallamuta: Professor da Universidade Tecnológica Federal do Paraná (UTFPR). Engenheiro de Telecomunicações pela UFPR. Especialista em Inteligência de Mercado pela FAE Business School. Mestre em Engenharia pela UEL. Trabalha com os temas: Inteligência de Mercado, Sistemas Eletrônicos e Gestão Institucional.

Ricardo Vinicius Bubna Biscaia: Professor da Universidade Tecnológica Federal do Paraná (UTFPR). Graduado em Engenharia Mecânica pela Universidade Tecnológica Federal do Paraná. Mestre em Engenharia Mecânica pela Universidade Tecnológica Federal do Paraná. Doutorando em Engenharia de Produção pela UTFPR. Trabalha com os temas: análise microestrutural e de microdureza de ferramentas de usinagem, modelo de referência e processo de desenvolvimento de produto e gestão da manutenção.

ÍNDICE REMISSIVO

A

Alface 74, 75, 77, 78, 79, 80, 81, 82, 83, 84, 85

Algoritmos 45, 46, 57, 119, 137, 138, 139, 148, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 306, 311

Aprendizagem colaborativa 9, 12, 13, 14, 15, 18, 20

Arduino 85, 136

Arquitetura de informação 60, 63

Astrofotografia 60, 61, 63, 64, 65, 66, 68, 69, 70, 72, 73

Ativismo de dados 151, 159

Autorregulação 22, 23, 25, 26, 27, 28, 29

B

Benchmark 102, 103, 109, 110

C

Cibercultura 9, 10, 11, 12, 14, 17, 18, 20

Ciclo de vida da edificação 188, 189, 190, 191, 192, 196, 197, 198, 199, 201

Cidades inteligentes 115, 117, 118, 121, 122, 123, 124, 125

Comparação de metodologias de ensino 30

Comunicação digital 60

Construção civil 188, 189, 196, 201, 202

D

Dados 2, 6, 7, 32, 35, 36, 37, 38, 39, 43, 45, 46, 48, 49, 50, 63, 67, 71, 74, 77, 79, 80, 81, 86, 90, 91, 92, 93, 97, 103, 109, 112, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 151, 152, 153, 156, 158, 159, 160, 161, 179, 192, 198, 199, 200, 202, 207, 219, 221, 228, 229, 230, 231, 232, 233, 236, 239, 241, 244, 255, 257, 258, 259, 260, 261, 262, 263, 264, 266, 268, 270, 271, 272, 274, 275, 277, 295, 296, 299, 305, 306, 307, 311, 312, 313

Divulgação científica 60, 62, 65, 66, 71, 72

Domínios 18, 86, 93, 94, 95, 98, 118

E

EaD 22, 23, 24, 25, 27, 28

Eletrodeposição 163, 164, 165, 166, 169

Engenharia do valor 188, 189, 192, 193, 196, 197, 202

Ensino 7, 9, 11, 12, 13, 14, 15, 16, 17, 19, 20, 22, 23, 24, 25, 26, 27, 29, 30, 31, 32, 33, 34, 35, 36, 37, 39, 40, 41, 42, 43, 65, 74, 75, 76, 80, 84, 85, 103

Estanho 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 295

F

Física 31, 33, 45, 65, 73, 74, 75, 76, 77, 83, 84, 85, 103, 205, 216, 217, 218, 219, 225, 226, 268, 292

G

Gênero 138, 139, 147, 149

I

IFRO 102, 103, 104, 105, 106, 108, 110, 112, 113

Interdisciplinaridade 24, 65, 74, 75, 84, 85

Internet das coisas 6, 113, 115, 117

L

Lecturing 30, 31, 32, 33, 34, 35, 36, 37, 39, 40, 41, 42, 43

M

Machine Learning 129, 138, 139, 140, 142, 145, 147, 148, 149, 150, 153

Mudanças 1, 2, 5, 6, 7, 8, 14, 27, 91, 133, 146, 160, 190, 243, 267, 272

N

Necessidades educativas especiais 44

O

OpenCV 129, 130, 133, 137

P

Peer Instruction 30, 31, 32, 33, 34, 35, 36, 37, 39, 40, 41, 42, 43

Placas de circuito impresso 163, 164, 165, 166, 174

Privacidade 96, 115, 117, 118, 119, 120, 121, 122, 124, 125, 126, 128, 160

Profissões 1, 2, 3, 4, 5, 6, 7

Proteção de dados 115, 117, 120, 121, 122, 124, 125, 126, 160

R

Raspberry Pi 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 131

Recursos educacionais digitais 44, 45, 47, 49, 52, 53

Redução de custo 188, 189, 192, 196, 198, 200, 201, 202

Remoção 96, 163, 165, 166, 168, 169, 170, 171, 173, 175, 177, 182, 184, 185, 191, 205, 309, 310

Repositórios 44, 45, 46, 47, 48, 49, 50, 52, 57

Robô autônomo 129, 130, 137

S

Saúde eletrônica 86

Seguidor de linha 129, 130, 131, 137

Sistema de recomendação 44, 52, 53, 54, 56

Sistemas 6, 12, 44, 45, 46, 65, 70, 72, 73, 86, 87, 88, 89, 90, 91, 92, 93, 95, 98, 100, 103, 107, 109, 130, 139, 149, 152, 154, 155, 157, 165, 190, 193, 217, 226, 230, 236, 277, 302, 314, 315

T

Tecnologia de Informação e Comunicação 86

Tecnologias digitais 9, 10, 13, 15, 16, 17, 18

Telessaúde 86, 93, 99, 100

Teoria Ator-Rede 9, 10, 11, 20

Teste de Friedman 30

Teste de Wilcoxon 30

U

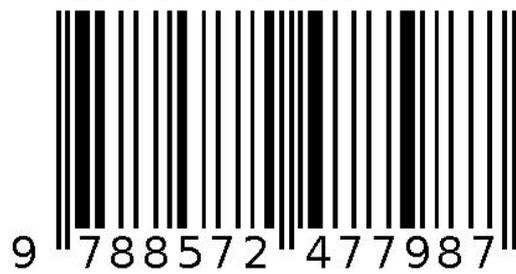
UX 60, 62

V

Vigilância 115, 118, 120, 153

Voz 4, 129, 138, 139, 149, 160

Agência Brasileira do ISBN
ISBN 978-85-7247-798-7



9 788572 477987