



Educação Matemática e suas Tecnologias 4

Atena Editora 2019

2019 by Atena Editora

Copyright © Atena Editora

Copyright do Texto © 2019 Os Autores

Copyright da Edição © 2019 Atena Editora

Editora Executiva: Profa Dra Antonella Carvalho de Oliveira

Diagramação: Natália Sandrini Edição de Arte: Lorena Prestes

Revisão: Os Autores

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

- Prof. Dr. Álvaro Augusto de Borba Barreto Universidade Federal de Pelotas
- Prof. Dr. Antonio Carlos Frasson Universidade Tecnológica Federal do Paraná
- 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. Gilmei Fleck Universidade Estadual do Oeste do Paraná
- Prof^a Dr^a Ivone Goulart Lopes Istituto Internazionele delle Figlie de Maria Ausiliatrice
- Profa Dra Juliane Sant'Ana Bento Universidade Federal do Rio Grande do Sul
- Prof. Dr. Julio Candido de Meirelles Junior Universidade Federal Fluminense
- Prof^a Dr^a Lina Maria Gonçalves Universidade Federal do Tocantins
- Profa Dra Natiéli Piovesan Instituto Federal do Rio Grande do Norte
- Profa Dra Paola Andressa Scortegagna Universidade Estadual de Ponta Grossa
- 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. Alan Mario Zuffo Universidade Federal de Mato Grosso do Sul
- Prof. Dr. Alexandre Igor Azevedo Pereira Instituto Federal Goiano
- Profa Dra Daiane Garabeli Trojan Universidade Norte do Paraná
- Prof. Dr. Darllan Collins da Cunha e Silva Universidade Estadual Paulista
- 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. 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. Gianfábio Pimentel Franco - Universidade Federal de Santa Maria

Prof. Dr. Benedito Rodrigues da Silva Neto - Universidade Federal de Goiás

Prof. a Dr. a Elane Schwinden Prudêncio – Universidade Federal de Santa Catarina

Prof. Dr. José Max Barbosa de Oliveira Junior - Universidade Federal do Oeste do Pará

Profa Dra Natiéli Piovesan – Instituto Federal do Rio Grande do Norte

Profa Dra Raissa Rachel Salustriano da Silva Matos – Universidade Federal do Maranhão

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. Eloi Rufato Junior - Universidade Tecnológica Federal do Paraná

Prof. Dr. Fabrício Menezes Ramos - Instituto Federal do Pará

Profa Dra Natiéli Piovesan - Instituto Federal do Rio Grande do Norte

Prof. Dr. Takeshy Tachizawa - Faculdade de Campo Limpo Paulista

Conselho Técnico Científico

Prof. Msc. Abrãao Carvalho Nogueira - Universidade Federal do Espírito Santo

Prof.^a Dr^a Andreza Lopes – Instituto de Pesquisa e Desenvolvimento Acadêmico

Prof. Msc. Carlos Antônio dos Santos - Universidade Federal Rural do Rio de Janeiro

Prof.ª Msc. Jaqueline Oliveira Rezende – Universidade Federal de Uberlândia

Prof. Msc. Leonardo Tullio - Universidade Estadual de Ponta Grossa

Prof. Dr. Welleson Feitosa Gazel - Universidade Paulista

Prof. Msc. André Flávio Gonçalves Silva – Universidade Federal do Maranhão

Prof.^a Msc. Renata Luciane Polsaque Young Blood - UniSecal

Prof. Msc. Daniel da Silva Miranda - Universidade Federal do Pará

Dados Internacionais de Catalogação na Publicação (CIP) (eDOC BRASIL, Belo Horizonte/MG)

E24 Educação matemática e suas tecnologias 4 [recurso eletrônico] /
Organizador Felipe Antonio Machado Fagundes Gonçalves. –
Ponta Grossa (PR): Atena Editora, 2019. – (Educação
Matemática e suas Tecnologias; v. 4)

Formato: PDF

Requisitos de sistema: Adobe Acrobat Reader

Modo de acesso: World Wide Web

Inclui bibliografia

ISBN 978-85-7247-350-7

DOI 10.22533/at.ed.507192405

 Matemática – Estudo e ensino – Inovações tecnológicas.
 Tecnologia educacional. I. Gonçalves, Felipe Antonio Machado Fagundes. II. Série.

CDD 510.7

Elaborado por Maurício Amormino Júnior - CRB6/2422

Atena Editora

Ponta Grossa – Paraná - Brasil

<u>www.atenaeditora.com.br</u>

contato@atenaeditora.com.br



APRESENTAÇÃO

Aobra "Educação Matemática e suas tecnologias" é composta por quatro volumes, que vêem contribuir de maneira muito significante para o Ensino da Matemática, nos mais variados níveis de Ensino. Sendo assim uma referência de grande relevância para a área da Educação Matemática. Permeados de tecnologia, os artigos que compõe estes volumes, apontam para o enriquecimento da Matemática como um todo, pois atinge de maneira muito eficaz, estudantes da área e professores que buscam conhecimento e aperfeiçoamento. Pois, no decorrer dos capítulos podemos observar a matemática aplicada a diversas situações, servindo com exemplo de práticas muito bem sucedidas para docentes da área. A relevância da disciplina de Matemática no Ensino Básico e Superior é inquestionável, pois oferece a todo cidadão a capacidade de analisar, interpretar e inferir na sua comunidade, utilizando-se da Matemática como ferramenta para a resolução de problemas do seu cotidiano. Sem dúvidas, professores e pesquisadores da Educação Matemática, encontrarão aqui uma gama de trabalhos concebidos no espaço escolar, vislumbrando possibilidades de ensino e aprendizagem para diversos conteúdos matemáticos. Que estes quatro volumes possam despertar no leitor a busca pelo conhecimento Matemático. E aos professores e pesquisadores da Educação Matemática, desejo que esta obra possa fomentar a busca por ações práticas para o Ensino e Aprendizagem de Matemática.

Felipe Antonio Machado Fagundes Gonçalves

SUMÁRIO

CAPÍTULO 11
CONSTRUÇÕES MATEMÁTICAS COM GEOGEBRA: ALÉM DO DESENHO
Deire Lúcia de Oliveira
DOI 10.22533/at.ed.5071924051
CAPÍTULO 213
MATERIAL POTENCIALMENTE SIGNIFICATIVO COM O USO DA LOUSA DIGITAL PARA O ENSINO DE FUNÇÃO AFIM
José Roberto da Silva Maria Aparecida da Silva Rufino Celso Luiz Gonçalves Felipe
DOI 10.22533/at.ed.5071924052
CAPÍTULO 3
O DESENVOLVIMENTO DO PENSAMENTO PROPORCIONAL NAS ESCOLAS PAROQUIAIS LUTERANAS DO SÉCULO XX NO RIO GRANDE DO SUL
Malcus Cassiano Kuhn
DOI 10.22533/at.ed.5071924053
CAPÍTULO 443
O ENSINO DA MATEMÁTICA NAS SÉRIES INICIAIS DO ENSINO FUNDAMENTAL: UMA ANÁLISE DO PERFIL DOS PROFESSORES DA CIDADE DE CAJAZEIRAS-PB
Francisco Aureliano Vidal Waléria Quirino Patrício
DOI 10.22533/at.ed.5071924054
CAPÍTULO 553
FORMAÇÃO CONTINUADA DE PROFESSORES DE MATEMÁTICA PARA O USO DE SOFTWARES EM SALA DE AULA
Ailton Durigon Andrey de Aguiar Salvi Bruna Branco
Marcelo Maraschin de Souza
DOI 10.22533/at.ed.5071924055
CAPÍTULO 661
ESTATÍSTICA NA EDUCAÇÃO BÁSICA: O USO DE TECNOLOGIAS DIGITAIS EM PESQUISAS DE OPINIÃO
Felipe Júnio de Souza Oliveira
DOI 10.22533/at.ed.5071924056
CAPÍTULO 779
OS DESAFIOS DA MATEMÁTICA NA EDUCAÇÃO INCLUSIVA: UMA REVISÃO SISTEMÁTICA
Cíntia Moralles Camillo Liziany Muller
DOI 10 22533/at ed 5071924057

CAPÍTULO 887
UM OLHAR SOBRE A FACE OCULTA DOS REGISTROS DE REPRESENTAÇÃO SEMIÓTICA ENVOLVENDO SISTEMAS LINEARES
Wagner Gomes Barroso Abrantes Tula Maria Rocha Morais
Luiz Gonzaga Xavier de Barros DOI 10.22533/at.ed.5071924058
CAPÍTULO 997
UM MÉTODO PARA FACILITAR A RESOLUÇÃO DE DETERMINANTES
Fernando Cezar Gonçalves Manso
Diego Aguiar da Silva Flávia Aparecida Reitz Cardoso
DOI 10.22533/at.ed.5071924059
CAPÍTULO 10 111
UTILIZAÇÃO DE TÉCNICAS DE INTELIGÊNCIA COMPUTACIONAL PARA CARACTERIZAR PACIENTES CARDIOPATAS
Juliana Baroni Azzi Robson Mariano da Silva
DOI 10.22533/at.ed.50719240510
CAPÍTULO 11122
UMA PROPOSTA METODOLÓGICA PARA O ENSINO DE ÁLGEBRA NA EDUCAÇÃO BÁSICA: AS QUATRO DIMENSÕES DA ÁLGEBRA E O USO DO GEOGEBRA PARA ANÁLISE DOS SIGNIFICADOS DAS RELAÇÕES ALGÉBRICAS NAS PARÁBOLAS
Sarah Raphaele de Andrade Pereira Lúcia Cristina Silveira Monteiro
DOI 10.22533/at.ed.50719240511
CAPÍTULO 12132
SEQUÊNCIA DIDÁTICA ELETRÔNICA: UM EXPERIMENTO COM NÚMEROS DECIMAIS E O TEMA TRANSVERSAL TRABALHO E CONSUMO COM ESTUDANTES DO ENSINO FUNDAMENTAL
Rosana Pinheiro Fiuza Claudia Lisete Oliveira Groenwald
DOI 10.22533/at.ed.50719240512
CAPÍTULO 13145
CONTEÚDOS ALGÉBRICOS DA PROVA DE MATEMÁTICA DO "NOVO ENEM"
Alan Kardec Messias da Silva
Acelmo de Jesus Brito Luciana Bertholdi Machado
Marcio Urel Rodrigues
DOI 10.22533/at.ed.50719240513
CAPÍTULO 14157
EDUCAÇÃO MATEMÁTICA E CRIATIVIDADE: UMA ABORDAGEM A PARTIR DA PERSPECTIVA DE
SISTEMAS DE CRIATIVIDADE
Cleyton Hércules Gontijo DOI 10.22533/at.ed.50719240514
DOLTU.ZZ003/81.80.00/19Z4U014

CAPÍTULO 15164
LINGUAGEM, IMAGENS E OS CONTEXTOS VISUAIS E FIGURATIVOS NA CONSTRUÇÃO DO SABER MATEMÁTICO QUE NORTEIAM OS LIVROS DIDÁTICOS DE MATEMÁTICA
Alexandre Souza de Oliveira
DOI 10.22533/at.ed.50719240515
CAPÍTULO 16
LETRAMENTO ESTATÍSTICO NO ENSINO MÉDIO: ESTRUTURAS POSSÍVEIS NO LIVRO DIDÁTICO
Laura Cristina dos Santos Cileda de Queiroz e Silva Coutinho
DOI 10.22533/at.ed.50719240516
CAPÍTULO 17 184
UM ESTADO DA ARTE DE PESQUISAS ACADÊMICAS SOBRE MODELAGEM EM EDUCAÇÃO MATEMÁTICA (DE 1979 A 2015)
Maria Rosana Soares Sonia Barbosa Camargo Igliori
DOI 10.22533/at.ed.50719240517
CAPÍTULO 18195
SCRATCH: DO PRIMEIRO OLHAR À PROGRAMAÇÃO NO ENSINO MÉDIO
Taniele Loss Nesi
Renata Oliveira Balbino
Marco Aurélio Kalinke
DOI 10.22533/at.ed.50719240518
CAPÍTULO 19205
OBJETOS VIRTUAIS DE APRENDIZAGEM DISPONÍVEIS NO BANCO INTERNACIONAL DE OBJETOS EDUCACIONAIS PARA TRIGONOMETRIA EM TODOS OS NÍVEIS DE ENSINO
Erica Edmajan de Abreu
Mateus Rocha de Sousa Felícia Maria Fernandes de Oliveira
Edilson Leite da Silva
DOI 10.22533/at.ed.50719240519
CAPÍTULO 20216
MODOS DE RESOLUÇÃO DE PROBLEMAS MATEMÁTICOS REALIZADOS POR ALUNOS DO ENSINO FUNDAMENTAL
Milena Schneider Pudelco Tania Teresinha Bruns Zimer
DOI 10.22533/at.ed.50719240520
CAPÍTULO 21226
O PACTO NACIONAL PELA ALFABETIZAÇÃO NA IDADE CERTA (PNAIC): FORMAÇÃO E PRÁTICA
DOS PROFESSORES ALFABETIZADORES NO ENSINO DA MATEMÁTICA PARA ALUNOS SURDOS
Renata Aparecida de Souza Maria Elizabete Rambo Kochhann Nilce Maria da Silva

CAPÍTULO 22
INVESTIGANDO CONCEPÇÕES E EXPLORANDO POTENCIALIDADES NUMA OFICINA REALIZADA COM A CALCULADORA CIENTÍFICA NAS AULAS DE MATEMÁTICA DO ENSINO MÉDIO
José Edivam Braz Santana Kátia Maria de Medeiros
DOI 10.22533/at.ed.50719240522
CAPÍTULO 23248
O QUE REVELAM AS PESQUISAS REALIZADAS NA FORMAÇÃO DE PROFESSORES DE MATEMÁTICA NA EDUCAÇÃO À DISTÂNCIA
Francisco de Moura e Silva Junior
DOI 10.22533/at.ed.50719240523
CAPÍTULO 24259
NÚMEROS NEGATIVOS E IMPRENSA NO BRASIL: AS DISCUSSÕES NO PERIÓDICO <i>UNIÃO ACADÊMICA</i>
Wanderley Moura Rezende Bruno Alves Dassie
DOI 10.22533/at.ed.50719240524
SOBRE O ORGANIZADOR268

CAPÍTULO 10

UTILIZAÇÃO DE TÉCNICAS DE INTELIGÊNCIA COMPUTACIONAL PARA CARACTERIZAR PACIENTES CARDIOPATAS

Juliana Baroni Azzi

Universidade Federal Rural do Rio de Janeiro Seropédica – Rio de Janeiro

Robson Mariano da Silva

Universidade Federal Rural do Rio de Janeiro Seropédica – Rio de Janeiro

RESUMO: Com o aumento progressivo do número de óbitos recorrentes de Doenças Cardiovasculares no mundo, este assunto vem sendo cada vez mais abordado em estudos de diferentes áreas. Utilizando treze variáveis além do resultado de diagnose presentes na Heart Diseases Database, fez-se possível a caracterização de pacientes a partir de dois modelos. Em um modelo dito completo, no qual os pacientes foram classificados por Máquina de Vetor de Suporte, e foi considerado melhor devido sua estabilidade, obteve-se em sua melhor simulação, dentre 100 realizadas, uma acurácia de 92.1% acompanhada de 6.8% de falso negativo. Enquanto para o modelo fit, no qual as variáveis foram selecionadas Regressão Linear e posteriormente classificadas por SVM, a acurácia foi de 89.8% e o falso negativo de 11.1%.

PALAVRAS-CHAVE: Inteligência Computacional, Doenças Cardiovasculares, Regressão Linear Múltipla, Máquina de Vetor de Suporte **ABSTRACT:** With the progressive increase in the number of recurrent deaths of Cardiovascular Diseases in the world, this subject has been increasingly approached in studies of different areas. Using thirteen variables in addition to the diagnosis results present in the Heart Diseases Database, it became possible to characterize patients from two models. In a complete said model, in which the patients were classified by the Support Vector Machine, and was considered better due to their stability, the best simulation was obtained, among 100 performed, an accuracy of 92.1% accompanied by 6.8% of false negative. For the fit model, in which the variables were selected by Linear Regression and later classified by SVM, the accuracy was 89.8% and the false negative was 11.1%. **KEYWORDS:** Computational Intelligence, Cardiovascular Diseases. Multiple Linear Regression, Support Vector Machine

1 I INTRODUÇÃO

As Doenças Cardiovasculares (DCV) lideram o ranking dos fatores de maior incidência de mortalidade no mundo, alcançando um percentual de 31%. A Organização Mundial de Saúde (OMS) registrou no ano de 2016, cerca de 17,5 milhões de mortes por doenças que afetam o sistema circulatório humano, e

estatísticas apontam que esse número cegará a 23,9 milhões até 2030. No Brasil, a Sociedade Brasileira de Cardiologia (SBC) estimou que cerca de 381 mil pessoas foram a óbito por DCV no país em 2017, esta estimativa foi feita pelo Cardiômetro, um instrumento estatístico que calcula esses índices a partir de dados coletados na 10ª Classificação Internacional de Doenças (CID10), do sítio do DATASUS/MS, para o período de 2006 a 2016.

Apesar desses dados apontarem um crescimento no número de casos fatais de doenças cardiovasculares, Soares *et al.* (2015), identificou uma redução da mortalidade nos grupos de portadores de doenças do aparelho circulatório, doenças cerebrovasculares e doenças isquêmicas do coração, no Estado do Rio de Janeiro, em análise realizada no período de 1979 a 2010, e apontou que esta queda pode estar relacionada com a melhoria socioeconômica da população do estado, já que não pôde ser comprovada que se sucedeu a partir do desenvolvimento da tecnologia nos procedimentos necessários para o combate dessas anomalias, nem tampouco, por controle dos riscos dessas doenças.

ASBC, aponta a hipertensão arterial, colesterol, tabagismo, estresse, sedentarismo e diabetes, como os principais fatores para incidência das doenças cardiovasculares. Dentre esses índices, alguns foram confirmados por Mansur *et al.* (2016), em estudo que concluiu que ao menos 20% das mortes registradas no Brasil no período de um ano, de adultos acima dos 30 anos, deram se por DCV.

O diagnóstico dessas doenças, dá-se pela associação de ao menos dois exames de diagnose. Ferreira *et al.* (2016), exemplificou que a identificação de um Infarto Agudo do Miocárdio exige que a elevação plasmática dos marcadores de necrose miocárdica [MNM] seja obrigatória, porém associada a dor torácica (analisada por exame clínico), ou alterações no eletrocardiograma (segmento ST e onda T).

Dentre os exames mais utilizados para controle do funcionamento do sistema circulatório, podemos destacar o Eletrocardiograma (ECG), exame que permite estudar diversas propriedades da musculatura do coração, através de um Galvanômetro, medidor de diferença de potencial (ddp). Este permite uma análise da formação e condução de estímulo cardíaco, e assim o diagnóstico de problemas no ritmo cardíaco, problemas de condução cardíaca, sinais de insuficiência cardíaca dentre outras doenças. Como visto, apesar da importância desse exame, o mesmo não pode ser usado de forma isolada para diagnóstico de qualquer DCV.

De acordo com um modelo geométrico desenvolvido pelo *New York Obesity Research Center* (NYORC), Moraes (2016), afirmou ser possível caracterizar pacientes com cardiopatias através da medição dos perímetros braquial, da cintura, do quadril, da coxa e da panturrilha. Esta também é uma técnica que exige associação de outros exames para diagnose de doenças, principalmente por seus resultados apresentarem uma diferença considerada pequena entre pacientes cardiopatas e saudáveis.

Técnicas de Inteligência Computacional são comumente encontradas na literatura, sendo utilizadas com fins de categorização de doenças cardiovasculares.

A associação das técnicas Máquina de Vetor de Suporte (SVM), Redes Neurais MLP, Algoritmos Genéticos e Árvore de Decisão, foi feita por Tavares (2013), para classificar cardiopatias em crianças, utilizando uma base de dados não normalizada. O destaque deu-se a técnica de SVM com pesos, que obteve melhores resultados. Ubiratan (2014), obteve acurácia, especificidade e sensibilidade superiores a 92% no auxílio do diagnóstico de cardiopatia isquêmica, utilizando técnicas de Algoritmo Genético, Reconhecimento Baseado no Casos (RBC) e derivações da função de Distância Euclidiana. A Regressão Linear Simples e Múltipla, foi utilizada por Ishitani (2006), para investigar a associação entre indicadores de nível socioeconômico e mortalidade por DCV de adultos no Brasil, concluindo que existe uma relação inversa entre esses fatores, dando destaque a educação.

Para este trabalho serão apresentados dois modelos, o primeiro chamado de completo (no qual todas as variáveis do banco serão aplicadas à técnica de Máquina de Vetor de Suporte), e o segundo, chamado de modelo *fit* (que teve as variáveis aplicadas ao modelo SVM selecionadas por Regressão Linear Múltipla). Após a caracterização dos pacientes será realizada a comparação entre as técnicas, a fim de selecionar os melhores resultados e o melhor modelo.

2 I REVISÃO BIBLIOGRÁFICA

Doenças Cardiovasculares

As doenças cardiovasculares são originárias de todo o sistema circulatório humano, podendo afetar desde o músculo cardíaco até os vasos sanguíneos. Doenças coronárias, cerebrovasculares, arterial periférica, cardíaca reumática, cardiopatia congênita, trombose venosa profunda e embolia pulmonar, formam o grupo de DCV.

Oliveira *et al.* (2015), concluiu que um dos fatores preponderantes para DCV são as modificações bi psíquicas geradas pela condição hipoestrogênica, oriundas do ciclo de menopausa em mulheres. Já que nesta fase da vida a mulher apresenta um aumento no triglicerídeos e na lipoproteína de baixa intensidade no organismo, podendo então considerar que as doenças cardiovasculares também são consideradas o fator que mais leva mulheres acima de 50 anos a óbito.

Pode-se destacar a doença arterial coronariana, acidente vascular cerebral (AVC), acidente vascular encefálico (AVE), carditite reumática, cardiopatias congênitas em suas diversas formas, flebite, trombose e embolia pulmonar, como as DCV as com maior incidência na população. Sendo que estas, podem ser diagnosticadas através de exames como ecografia transesofágica, cintilografia, cateterismo, ECG, radiografia de tórax, ecocardiograma, ressonância magnética, entre outros.

Regressão Linear

A relação existente entre uma variável dependente e uma ou várias variável(is) independente(s), pode ser encontrada de forma estatística, através de uma análise

de regressão. Esta, é representada por um modelo matemático representado por uma equação que associa as variáveis, em um gráfico, chamado de diagrama de dispersão. Quando a relação existente é de uma variável dependente pra uma independente, o modelo é de Regressão Linear Simples, caso exista mais de uma variável independente, é um modelo de Regressão Linear Múltiplo.

Esta técnica, portanto, define a influência de uma variável X (investigativa), sobre um valor esperado de uma variável Y (resposta), objetivando analisar e identificar alterações em E[Y], ou seja, no valor esperado de Y, a fim de verificar se o mesmo sofre alterações devido as condições de interação com a variável Y, enquanto a variável X informa sobre o comportamento de Y.

A variável resposta desse modelo, em geral se comporta de forma linear, quadrática, cúbica, exponencial ou logarítmica. Essa identificação é de suma importância para o modelo, para que haja uma explicação do mesmo. A aproximação dos pontos no diagrama de dispersão é feita pelo Método de Mínimos Quadrados (MMQ), método que realiza a soma dos quadrados da distância entre os pontos do modelo e os pontos do diagrama, gerando uma relação entre X e Y, sempre buscando o menor erro. A utilização dessa técnica é fundamental, já que nem todos os pontos do modelo se ajustarão perfeitamente aos pontos do diagrama.

Máquina de Vetor de Suporte

Com base na teoria do Aprendizado de Máquina (AM), do inglês *Machine Learning*, pode-se dizer que esta técnica estuda o desenvolvimento de algoritmos capazes de aprender com os próprios erros, para realizar previsões sobre dados, ou seja, suas decisões serão tomadas a partir de experiências anteriores acumuladas, adquirindo automaticamente novos conhecimentos.

O AM é dividido em dois principais padrões, a partir da forma com que o algoritmo se relaciona com o meio, podendo ser supervisionada ou não supervisionada. No aprendizado supervisionado, o algoritmo recebe um conjunto de treino, responsável por definir o que o algoritmo irá buscar, em seguida, o conjunto de teste é mapeado dentro de cada categoria desejada, produzindo padrões de saída corretos para cada nova entrada. Já o aprendizado não supervisionado, agrupa as entradas de acordo com medidas de qualidade, não existindo classes pré-definidas para os atributos, portanto, este padrão visa o estabelecimento de classes.

Características como precisão e velocidade de classificação dos dados, robustez e escalabilidade do sistema e a clareza de resposta do modelo são essenciais para que o processo se cumpra obtendo os melhores resultados. O SVM se destaca devido a características como essas, além ser uma teoria bem definida.

O Support Vector Machine (SVM) foi criado com base na Teoria do Aprendizado Estatístico, que propõe uma maximização da capacidade de generalização, buscando uma classificação eficiente do conjunto de treino, e minimização do risco estrutural do sistema, que representa a probabilidade de classificação errônea de padrões ainda

não conhecidos pela máquina.

Na literatura, a Máquina de Vetor de Suporte, encontra se correlacionada a problemas de classificação e regressão. Os vetores de suporte são necessários para a definição de um hiperplano ótimo, ou seja, uma função capaz de separar as classes. As funções desse hiperplano ótimo são definidas com base na teoria do aprendizado estatístico, desenvolvido por Vipnik, e chamado de dimensão Vipnik-Chervonenkis ou dimensão VC. A importância dessa dimensão se encontra no fato de que se a mesma for definida corretamente, o aprendizado se torna confiável.

Essa técnica se SVM vem se destacando dentre outras técnicas de Inteligência Computacional, quando se trata de reconhecimento de padrão, devido aos seus resultados superiores, quando comparado com as Redes Neurais, por exemplo.

3 I MATERIAIS E MÉTODOS

Para aplicação das técnicas escolhidas para este trabalho, foram utilizados os dados extraídos da *Heart Disease Database*, base de dados de domínio público, que foi subdividida em quatro subconjuntos e utilizaremos o de Claverland, obtidos por Robert Detrano no *Cleverland Clinic Foundation*. Trezentos e três pacientes foram incluídos nos estudos, dos quais, 164 foram classificados como saudáveis e 139 doentes. Para cada um, foram expostos 76 atributos, dos quais apenas 13 foram utilizados, considerando o número irrisório de dados faltantes, e então permitindo uma melhor análise. As variáveis aplicadas ao modelos foram: idade, gênero, tipo de dor no peito, pressão arterial em repouso, colesterol sérico, concentração de açúcar no sangue em jejum, resultados eletrocardiográficos em repouso, ritmo cardíaco máximo alcançado, angina induzida por exercício, depressão da onda ST induzida pelo exercício em relação ao repouso, inclinação do pico de segmento ST durante o exercício, número de grandes vasos coloridos por fluoroscopia e talassenia, além da décima quarta coluna que apresenta o diagnóstico de cada paciente.

Foram realizadas cem simulações para cada um dos dois modelos simulados. Em ambos os modelos foi utilizada a função *kernel* C-svc, com um valor de C unitário. Dos 303 pacientes, foram considerados 297 para as simulações, tendo em vista que os outros 6 possuíam dados faltantes, que impossibilitavam um melhor desempenho dos modelos. Desses 297, 70% formava o conjunto de treino, e 30% o conjunto de teste. O primeiro modelo simulado foi chamado de completo, onde todas as 13 variáveis do banco foram consideradas em uma rotina de Máquina de Vetor de Suporte. O segundo modelo, chamado *fit*, as 13 variáveis foram analisadas por Regressão Linear Múltipla, onde um Modelo Linear Generalizado (MGL) selecionou apenas 6 variáveis que apresentaram correlação, sendo elas: gênero, pressão arterial em repouso, tipo de dor no peito, angina induzida por exercício, número de grandes vasos coloridos por fluoroscopia e talassenia, em seguida essas foram classificadas pelo SVM.

Ambos os modelos foram implementados no *software* livre chamado de R. Este que vem sendo vastamente utilizado para aplicações de modelos lineares e não lineares, testes estatísticos clássicos, análise de séries temporais, classificação, agrupamento e técnicas gráficas altamente expansíveis. O próprio *software* gerou em meio a simulação um sumário dos dados para cada um dos conjuntos (treino e teste), nas Tabelas 1 e 2 serão apresentados esses valores estatísticos das variáveis contínuas.

	Idade	Pressão arterial	Colesterol	Ritmo cardíaco	Depressão onda ST
Valor mínimo	29	94	126	71	0
1° Quartil	48	120	210.8	131.8	0
Mediana	55.5	130	239	154.5	0.75
Média	54.58	132.3	244.6	149.2	1.058
3° Quartil	61.25	140	271.8	167.2	1.65
Valor máximo	76	192	564	202	6.2

Tabela 1: Sumário de valores do conjunto de treino

	Idade	Pressão arterial	Colesterol	Ritmo cardíaco	Depressão onda ST
Valor mínimo	35	100	149	97	0
1° Quartil	47	120	214	138	0
Mediana	56	130	254	151	0.8
Média	54.46	130.3	253.8	150.5	1.051
3° Quartil	60	140	288	163	1.6
Valor máximo	77	200	409	194	4.4

Tabela 2: Sumário de valores do conjunto de teste

Os resultados das 100 simulações de cada um dos modelos, forneceram valores de erro de treino, erro de validação cruzada, número do vetor suporte, acurácia, sensibilidade, especificidade e valor falso negativo. A partir desses resultados, é possível construir a Matriz Confusão do modelo, como pode ser visto através da formação de cada um desses índices nas fórmulas seguintes.

A acurácia pode ser calculada conforme a Eq. 1, e esta garante o grau de confiabilidade do modelo.

$$Acurácia = \frac{VP + VN}{N} = \frac{(Verdadeiro\ positivo + Verdadeiro\ negativo)}{Total\ lote} \tag{1}$$

A sensibilidade apresenta a capacidade do sistema de reconhecimento dos pacientes doentes, enquanto a especificidade a capacidade de reconhecimento dos saudáveis. Estas são calculadas pelas Eq. 2 e 3, respectivamente.

Sensibilidade =
$$\frac{VP}{VP+FN} = \frac{N\'umero\ de\ resultados\ de\ testes\ verdadeiros\ positivos}{Todos\ os\ doentes\ af\ etados}$$
 (2)

onde VP = Verdadeiro Positivo e FN = Falso Negativo.

Especificidade =
$$\frac{VN}{VN+FP} = \frac{N\'umero\ de\ resultado\ de\ teste\ verdadeiros\ negativos}{Todos\ os\ doentes\ n\~ao\ afetados}$$
 (3)

onde VN = Verdadeiro Negativo e FP = Falso Positivo.

4 I RESULTADOS E DISCUSSÕES

Em busca de melhores resultados uma mesma base foi aplicada a dois distintos modelos (completo e *fit*), que nos permitiram analisar cada um de forma individual e por fim compará-los, a fim de escolher o melhor. Foram realizadas 100 simulações para cada um dos modelos, ambas com um conjunto de treino formado por 70% dos dados, e um conjunto de teste com os 30% restantes. Com o *software* R foram aplicados SVM e Regressão Linear para selecionar as variáveis do modelo *fit* (gênero, pressão arterial em repouso, tipo de dor no peito, angina induzida por exercício, número de grandes vasos coloridos por fluoroscopia e talassenia), e então, foi obtido como resposta do sistema o erro de treino, erro de validação cruzada, número de vetores de suporte, acurácia, sensibilidade, especificidade e falso negativo.

A aplicação da técnica de SVM foi escolhida dentre as de Aprendizado de Máquina devido aos bons resultados já encontrados anteriormente na literatura em aplicações semelhantes a realizada neste trabalho. Além de querer confrontá-la com resultados já obtidos em modelos bastante parecidos aos aqui desenvolvidos.

O erro de treino analisa sempre os mesmos valores presentes no conjunto de treino da base, portanto é considerado mais simples, e apresentará um valor inferior ao do erro de validação cruzada que é calculado toda vez que uma nova informação é introduzida no modelo, sendo assim, este garante uma maior robustez ao sistema.

Com base nos resultados do sistema foi construída a Tabela 3.

	Erro de treino	Erro de loocv	N° de vetores suporte	Acurácia	Sensibilidade	Especificidade	Falso Negativo
Valor mín.	7.2%	12.1%	101	75.2%	59%	71.4%	2.5%
Mediana	10.5%	17.3%	114	83.1%	78.3%	87.2%	21.6%
Média	10.5%	17.3%	114.4	82.8%	77.8%	87.2%	22.1%
Valor máx.	13.9%	22.1%	128	92.1%	97.5%	97.9%	40.9%

Tabela 3: Estatística dos resultados do modelo completo

Diante da diferença, já vista, entre os erros de treino e erro de validação cruzada, e observando os valores mínimos e máximos do modelo completo, 7.2% e 13.9% para erro de treino e 12.1% e 22.1% para o de *loocv*, podemos perceber que até mesmo suas variações são diferentes, sendo a de treino inferior a de validação cruzada. A

complexidade do sistema também não apresentou uma alteração brusca entre os valores de mínimo e máximo (101 e 128), confirmando isso pela comparação entre a média e mediana desse índice.

Este modelo completo, em sua melhor simulação, obteve uma acurácia de 92.1%, garantindo que em sua simulação mais assertiva, existe pouco mais de 92% de chance de o modelo acertar no diagnóstico de um paciente. E em sua pior simulação, esse valor foi de 75.2%, o que não é um valor consideravelmente baixo.

Com valores elevados para sensibilidade e especificidade, em suas melhores simulações destes pontos de vista, 97.5% dos diagnósticos positivos foram para pacientes realmente doentes, enquanto 97.9% foram de resultados negativos para pacientes saudáveis.

Falso negativo é um fator de suma importância quando se trata de resultados na área da biomedicina, principalmente em diagnósticos médicos. Este índice representa a porcentagem de diagnósticos assertivos, ou seja, menores valores representam menos erros de resultados de exames. Para este caso obteve-se um valor bastante baixo, 2.5%, o que garante uma baixa probabilidade de erro.

Para escolha de melhor e pior simulações, foi escolhido o valor de acurácia, por esse ser o real valor de avalição de confiabilidade do modelo, portanto, na Tabela 4 serão apresentadas essas simulações com seus valores.

Simulações	Erro de treino	Erro de loovc	N° de vetores suporte	Acurá- cia	Sensibi- lidade	Especifi- cidade	Falso Negativo
Pior	7.6%	13.6%	101	75.2%	64.8%	82.6%	35.1%
Melhor	12.9%	21%	123	92.1%	93.1%	91.1%	6.8%

Tabela 4: Pior e melhor simulações do modelo completo

Para o valor mais alto de acurácia, obteve-se um valor de 6.8% de falsos negativos, e valores de sensibilidade e especificidade acima de 91%, mostrando que este poderia ser um modelo confiável, para auxiliar em diagnósticos de doenças cardiovasculares.

Para o modelo *fit*, em que as variáveis foram escolhidas por Regressão Linear, foi feita a análise estatística apresentada na Tabela 5.

	Erro de treino	Erro de loocv	N° de vetores suporte	Acurá- cia	Sensibi- lidade	Especifi- cidade	Falso Negativo
Valor mín.	12%	13.8%	92	69.6%	59.5%	69.3%	10.8%
Mediana	15.8%	19.8%	114	80.8%	76.8%	84.6%	23.1%
Média	15.6%	19.9%	112.99	80.2%	76.5%	83.6%	23.4%
Valor máx.	19.7%	26.6%	130	89.8%	89.1%	95.4%	40.4%

Tabela 5: Estatística dos resultados do modelo fit

Tanto o erro de treino quanto o de validação cruzada, em seus menores valores, apresentaram um percentual acima do encontrado para o modelo completo, afirmando então, que o modelo *fit* apresentou mais erros em seus resultados quando comparado ao modelo completo. Já quando falamos do número de vetores de suporte, este modelo (*fit*), mostrou-se menos complexo em sua melhor simulação, apresentando a necessidade de 92 vetores de suporte, enquanto o modelo completo precisou de 101.

Com uma acurácia de 89.8%, o modelo *fit* também pode ser considerado confiável, já que este é um valor alto, porém, sua pior simulação ficou abaixo de 70%, o que pode mostrar uma certa instabilidade do modelo. Essa instabilidade é confirmada pela variação entre os valores mínimos e máximos de sensibilidade e especificidade. Apesar de apresentarem valores máximos elevados, os valores mínimos baixo, não transferem a confiança desejada por quem busque bons resultados. Da mesma forma, podemos avaliar o valor de falso negativo, que obteve 10.8% de diagnósticos errados em sua melhor simulação, contudo apresentou 40.4% de exames negativos para doença, de pacientes saudáveis.

Na Tabela 6, serão mostradas as simulações consideradas pior e melhor, também tomando como parâmetro, como no modelo anterior, os valores da acurácia.

Simulações	Erro de treino	Erro de loocv	N° de vetores suporte	Acurá- cia	Sensibi- lidade	Especifi- cidade	Falso Negativo
Pior	14.9%	19.3%	107	69.6%	68.1%	71.1%	31.8%
Melhor	19.7%	23.5%	119	89.8%	88.8%	90.9%	11,1%

Tabela 6: Pior e melhor simulações do modelo fit

Para a melhor simulação do modelo chamado de *fit*, em que a acurácia ficou próxima de 90%, o sistema mostrou-se pouco menos complexo por ter a necessidade de 119 vetores de suporte, porém, apresentou valores de erros um pouco acima do que foi visto para o modelo completo. Com sensibilidade e especificidade consideradas boas para esta melhor simulação, o valor de falso negativo, também representou uma quantidade de erros considerada baixa, mas não tanto quanto no modelo anterior.

A pior simulação confirma o que foi dito anteriormente, com relação a confiabilidade do modelo, devido a grande diferença de valor para os índices relacionados a matriz confusão, quando comparados a melhor simulação.

5 I CONCLUSÕES

Diante dos resultados obtidos no modelo completo, que utilizou todas as variáveis do banco de dados (idade, gênero, tipo de dor no peito, pressão arterial em repouso, colesterol sérico, concentração de açúcar no sangue em jejum, resultados eletrocardiográficos em repouso, ritmo cardíaco máximo alcançado, angina induzida

por exercício, depressão da onda ST induzida pelo exercício em relação ao repouso, inclinação do pico de segmento ST durante o exercício, número de grandes vasos coloridos por fluoroscopia e talassenia), e nos obtidos no modelo *fit*, que selecionou as variáveis (gênero, pressão arterial em repouso, tipo de dor no peito, angina induzida por exercício, número de grandes vasos coloridos por fluoroscopia e talassenia) por Regressão Linear, pode-se dizer que além de satisfatórios, foram conclusivos para que fosse escolhido um modelo que se destacasse diante do outro apresentado.

Em aplicações de Inteligência Computacional, em geral, duas variáveis de resposta do sistema são fundamentais para que se diga que um modelo é confiável, e em ambas o modelo completo, obteve destaque diante do modelo *fit*, com valores de acurácia de 92.1% e falso negativo de 6.8% em uma mesma simulação, comparados a 89.8% de acurácia e 11.1% de falso negativo no modelo *fit*. A importância desses valores, associados aos valores de sensibilidade (93.1% para a melhor simulação do modelo completo e 88.8% para o *fit*), que representam a capacidade do modelo em acertar diagnósticos de pacientes doentes, e em geral é muito utilizado em modelos médicos, dá-nos a segurança de se ter um resultado de exame com diagnóstico correto, sendo assim, podemos considerar que os valores obtidos neste trabalho foram satisfatórios e superiores aos de Bhatia *et al.*(2008), que aplicou a mesma base de dados a um modelo SVM, e obteve uma acurácia de 72,55% em sua melhor simulação, e ao de Ho & Chou (2001), que apresentou um percentual de erro de 81% em suas respostas para diagnósticos de tais doenças, utilizando também um modelo de Máquina de Vetor de Suporte.

REFERÊNCIAS

Bhatia, S., *et al.* **SVM based decision support system for heart disease classification with integer-coded genetic algorithm to select critical features**. *World Congress on Engineering and Computer Science*. San Francisco: USA, 2008.

D.W. Aha. *Heart Disease Databases*. Disponível em: www.ics.uci.edu/pub/machine-learning-databases/heart-disease.names - Acessado em dezembro/2017

Ferreira, A.R.P.A., Silva, M.V., Maciel, J. **Eletrocardiograma no infarto agudo do miocárdio**: O que esperar?. *International Journal of Cardiovascular Sciences.*, v.3, n.29, p.198-209, 2016.

Ho, C.S., Chou, J.S. **Fuzzy ARTRON**: A general-purpose classifier empowered by fuzzy ART and error back-propagation learning. *Journal of Information Science and Engineering*, v.13, n.17, p.683-695, 2001.

Hoffmann, R. Análise de Regressão: Uma introdução à Econometria. 4 ed. Piracicaba: Hucitec, 2014.

Ishitani, L.H., *et al.* **Desigualdade social e mortalidade precoce por doenças cardiovasculares no Brasil**. Rev Saúde Pública, v.40, n.4, p.1-8, 2006.

Lorena, A.C., Carvalho, A.C.P.L.F. **Introdução às máquinas de vetor suporte**. Relatório técnico. São Paulo: Universidade de São Paulo, 2003.

Mansur, A.D.P., Favarato, D. Mortalidade por doenças cardiovasculares no Brasil e na região metropolitana de São Paulo. São Paulo: Instituto do Coração (InCor) – HCFMUSP, 2012.

Moraes, V.C.S., *et al.* **Identificação do risco de cardiopatia através do estudo combinado de circunferências corporais**. Acta Biomédica Brasiliensia., v.7, n.1, p.31-39, 2016.

Oliveira, A.S. **Fatores de Risco Cardiovascular em Mulheres Pós-Menopausa**. Canoas: UNILASALLE, 2015.

Organização Pan-Americana de Saude & Organização Mundial de Saúde Brasil. **Doenças Cardiovasculares**. 2017. Disponível em: www.paho.org/bra/index.php?option=com_content&view= article&id=5253%3Adoencas-cardiovasculares&catid=845%3Anoticias&Itemid=839 – Acessado em maio/2017

Passos, U.R.C. Computação evolutiva e aprendizado de máquina aplicados ao apoio do diagnóstico da cardiopatia isquêmica. Dissertação. Campos dos Goytacazes: Universidade Cândido Mendes, 2014.

The R Foundation. The R Project for Statistical Computing. 2017. Disponível em: https://www.r-project.org/ - Acessado em outubro/2017

Soares, G.P., Klein, C.H., Silva, N.A.S., Oliveira, G.M.M. **Evolução da Mortalidade por Doenças do Aparelho Circulatório nos Municípios do Estado do Rio de Janeiro, de 1979 a 2010**. Arq. Bras. Cardiol., v.104, n.5, p.356-365, 2015.

Sociedade Brasileira de Cardiologia. **Cardiômetro**: Mortes por doenças cardiovasculares no Brasil. 2016. Disponível em: http://www.cardiometro.com.br/ - Acessado em agosto/2018

Stitson, M.O., Weston, J.A.E., Gammerman, A., Vovk, V. & Vapnik, V. *Theory of support vector machines*. Relatório técnico. Londres: *University of London*, 1996.

Tavares, T.R. **Utilização de técnicas de inteligência artificial para classificação de crianças cardiopatas em base de dados desbalanceada**. Dissertação. Recife: Universidade Federal de Pernambuco, 2013.

Thermo Scientific. Interpretação dos resultados dos testes. 2012. Disponível em: www.phadia.com/pt-BR/Diagnostico-de-auto-imunidade/Saber-mais/Avalia cao-dos-Resultados-dos-Testes/#Sens Spec – Acessado em novembro/2017

WHO: *World Health Organization*. **Cardiovascular diseases**. 2017. Disponível em: http://www.who.int/news-room/cardiovascular-diseases-(cvds) – Acessado em junho/2018

SOBRE O ORGANIZADOR

FELIPE ANTONIO MACHADO FAGUNDES GONÇALVES Mestre em Ensino de Ciência e Tecnologia pela Universidade Tecnológica Federal do Paraná(UTFPR) em 2018. Licenciado em Matemática pela Universidade Estadual de Ponta Grossa (UEPG), em 2015 e especialista em Metodologia para o Ensino de Matemática pela Faculdade Educacional da Lapa (FAEL) em 2018. Atua como professor no Ensino Básico e Superior. Trabalha com temáticas relacionadas ao Ensino desenvolvendo pesquisas nas áreas da Matemática, Estatística e Interdisciplinaridade.

Agência Brasileira do ISBN ISBN 978-85-7247-350-7

9 788572 473507