

**UFU –
UNIVERSIDADE
FEDERAL DE
UBERLÂNDIA**

**MODELO DE RASTREIO DE
PREDISPOSIÇÃO EM
ATAQUE CARDÍACO COM
CIÊNCIA DE DADOS
ATRELADA À LINGUAGEM
DE PROGRAMAÇÃO
PYTHON**

Monte Carmelo, MG

2023



Matheus Nunes Candido Senna

**MODELO DE RASTREIO DE
PREDISPOSIÇÃO EM
ATAQUE CARDÍACO COM
CIÊNCIA DE DADOS
ATRELADA À LINGUAGEM
DE PROGRAMAÇÃO
PYTHON**

Relatório apresentado à 7ª FEMIC - Feira
Mineira de Iniciação Científica.

Orientação do Prof. Matheus Nunes Candido
Senna

Monte Carmelo, MG

2023



RESUMO

Este projeto foi realizado a fim de promover melhor qualidade de vida à parte da população que está sujeita a sofrer futuramente de taquicardia e vir a óbito. Também tem a intenção de promover economia, rapidez e acurácia nos diagnósticos. A realização neste projeto foi de procurar, estudar e aplicar modelos válidos de aprendizagem de máquina e ciência de dados, a fim de promover confiabilidade e acurácia ao modelo criado. Dentre os procedimentos realizados há a coleta de dados públicos não-fictícios, uso de teoremas matemáticos funcionais dentro do modelo e validação do modelo através da curva ROC.

Obtendo 87% de precisão, o modelo preditivo se saiu bem e esta taxa de acerto é considerada muito boa para modelos preditivos de rastreio. O objetivo de acurácia deste modelo preditivo visava 75% e, merecidamente, alcançou 87%, portanto, o objetivo foi concluído com êxito. Portanto, o algoritmo tem validação boa, precisão alta e importância social muito grande.

Palavras-chave: ciência, dados, taquicardia, saúde, programação, python



SUMÁRIO

1 INTRODUÇÃO	5
2 JUSTIFICATIVA	6
3 OBJETIVO GERAL	7
4 METODOLOGIA	8
5 RESULTADOS OBTIDOS	9
6 CONCLUSÕES OU CONSIDERAÇÕES FINAIS	10
REFERÊNCIAS	11



1 INTRODUÇÃO

O projeto foi concebido a partir da linguagem de programação “Python”, que é a mais popular dentre as linguagens de programação que lidam com ciência de dados e aprendizagem de máquinas. A partir do banco de dados disponível na plataforma Kaggle, presente em: <https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset> acontecem as análises e os estudos.

No algoritmo de regressão logística presente no modelo, existe a fórmula de correlação de Pearson, que diz respeito ao grau de relacionamento entre duas variáveis.

A curva ROC, que é um método de validação de algoritmos foi aplicado ao algoritmo presente neste projeto e obteve resultado positivo. O objetivo de ROC é atestar de que um algoritmo obteve alta precisão através de sua eficácia e não por aleatoriedade. Isso quer dizer que, se outro banco de dados for aplicado ao mesmo modelo, a precisão continuará por volta de 87%. Isso não aconteceria caso o algoritmo possuísse curva ROC ruim. Para mais informações sobre isso, favor acessar: <https://www.leansaude.com.br/o-que-e-curva-roc-sensibilidade-e-especificidade/>

O modelo de aprendizagem de máquina funciona de acordo com a seguinte sucessão de etapas: Limpeza de dados, separação dos conjuntos de dados de treino e teste, processamento de acordo com o teorema do algoritmo, treino e teste.



2 JUSTIFICATIVA

Este projeto foi criado com a intenção de salvar vidas, rastreio precoce de vítimas, economia de laboratórios (devido ao não-desperdício de dinheiro com exames em pacientes cujos sofrem de outra condição), e aumento de expectativa de vida pois: “estima-se que 40 milhões de brasileiros tenham algum tipo de arritmia , que é responsável por mais de 320 mil mortes súbitas por ano “ de acordo com o hospital Leforte: leforte.com.br



3 OBJETIVOS

3.1 Objetivo geral

Rastrear a predisposição de pacientes x e y em ter taquicardia em algum período da vida e fazer com que a mesma não ocorra.

3.2 Objetivos específicos

- Rastreio
- Economia de laboratórios
- Rapidez , salvando tempo de espera



4 METODOLOGIA

Enfrentado pelo criador deste projeto foram as ambiguidades de vários conteúdos divergentes entre si encontrados na internet e as falhas na taxa de precisão e validação.

A metodologia consiste em em aprendizado reverso; que significa que se obtém o entendimento de um assunto através da análise de seus resultados. Isso, inconscientemente causa o entendimento de forma rápida e prática, pois, ao invés de se fazer perguntas e obter respostas, obtemos as causas através dos resultados. Exemplo: “Um carro esportivo Ferrari consome menos que 50% por cento da frota de rodovia nas estradas brasileiras” – Através disso, pensa-se: “Os carros esportivos Ferrari são mais leves e possuem apenas 2 acentos, apesar de seu motor de alto performance.”.

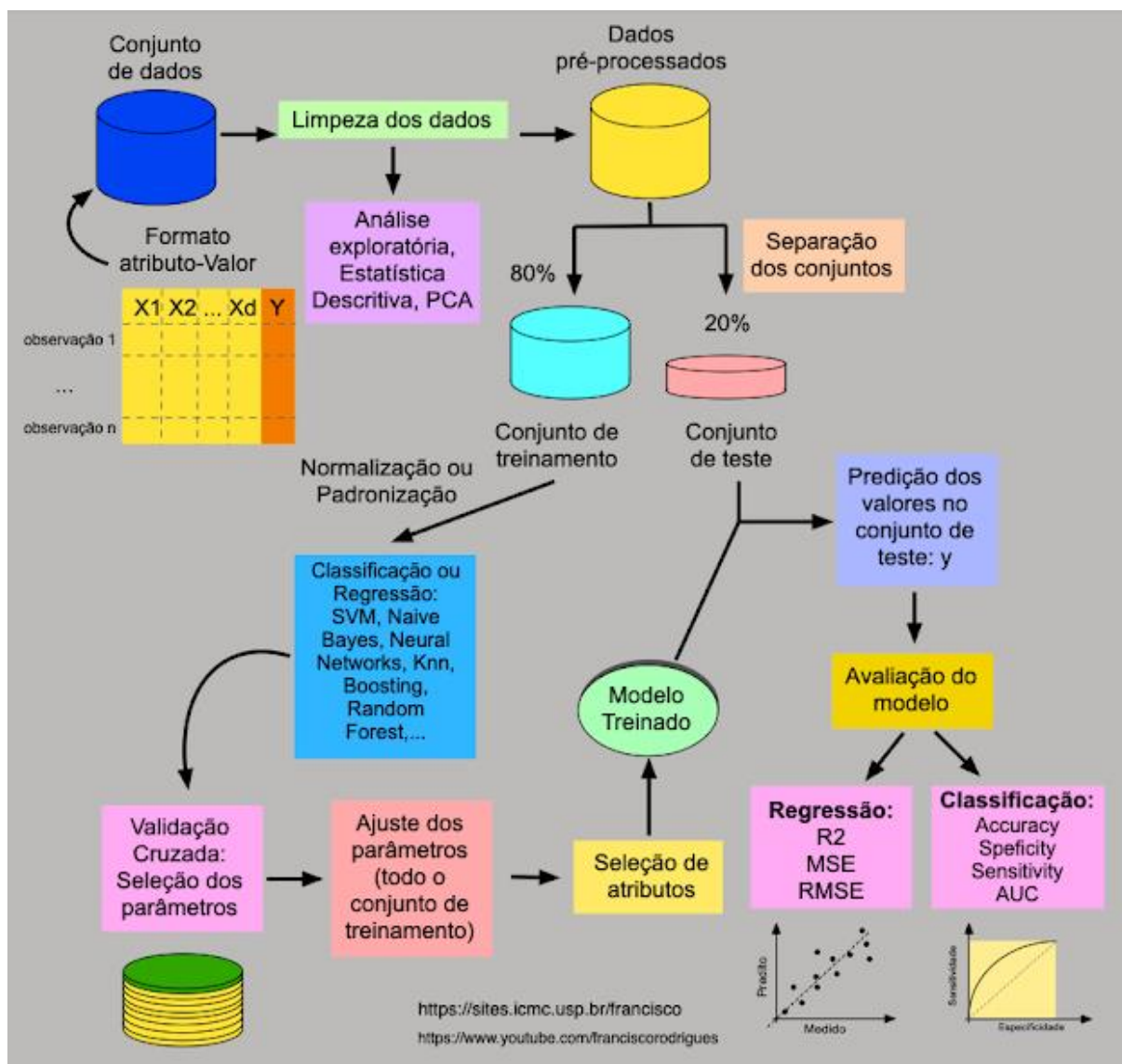
Deste modo, olhando para os resultados do estudo de casos do modelo preditivo, podemos, a partir dos resultados, olhar as características de um paciente e relacionar com sua predisposição à taquicardia.

A metodologia do modelo consiste na a seguinte sucessão de etapas: Limpeza de dados, separação dos conjuntos de dados de treino e teste, processamento de acordo com o teorema do algoritmo, treino e teste.



5 RESULTADOS OBTIDOS

O modelo preditivo em questão funciona de acordo com o fluxograma a seguir



Fonte: <https://sites.icmc.usp.br/francisco>



6 CONCLUSÕES OU CONSIDERAÇÕES FINAIS

Conclui-se que este projeto é um pouco complexo e que dificuldades foram enfrentadas desde o seu primórdio. A grande ajuda nisto foi a plataforma Kaggle, onde é possível encontrar dados não-fictícios com finalidade de estudo e análises. Assegura-se de que é possível ampliar a a acertividade deste algoritmo e que ainda há muito o que ser trabalhado neste projeto.



REFERÊNCIAS

As referências bibliográficas de seu artigo devem ser apresentadas em ordem alfabética e de acordo com a norma da ABNT – NBR 6023.

Siga os exemplos abaixo, sendo que as palavras em *itálico* são destaques meramente ilustrativos e não devem ser incluídos em seu manuscrito final:

Livros:

SCHWERTL, Simone Leal. Matemática básica. 2. ed. Blumenau: Edifurb, 2010. 113 p, il.

Capítulos de Livros:

BRANDT, Paulo Roberto. Geração de energia para o desenvolvimento regional no médio vale do Itajaí. In: Desenvolvimento e meio ambiente em Santa Catarina : a questão ambiental em escala local/regional, Joinville : Ed. UNIVILLE, 2006. p.[119]-125.

Periódicos:

CIÊNCIA E OPINIÃO. Curitiba: Centro Universitário Positivo. 2003.

Artigos de periódicos:

TOZZI, M.; OTA, J. Vertedouro em degraus. Revista da Vinci, Curitiba, v.1, n.1, p. 9-28, 2004.

Monografias, dissertações e teses:

PERES, Adriano; BARBI, Ivo; UNIVERSIDADE FEDERAL DE SANTA CATARINA, Centro Tecnológico. Uma nova família de inversores com comutação suave empregando a técnica de grampeamento ativo, 2000. 162p, il. Tese (Doutorado).



Publicações periódicas consideradas em parte (suplementos, fascículos, números especiais:

ARC DESIGN. Mestres da Arquitetura: Oscar Niemeyer. São Paulo: Quadrifoglio, n. 35, mar. - abril, 2004.

Artigos de jornais:

MOREIRA, T. Debate sobre software livre chega ao celular. Valor Econômico, São Paulo, 04 out. 2004. p. B4.

Trabalhos em eventos

LODER, L.L. A formação de identidades e a construção de saberes em um curso de engenharia elétrica. Anais: XXXVIII – Congresso Brasileiro de Ensino de Engenharia. Fortaleza: UFC, 2010.

Internet:

MOURA, G. C. de M. **Citação de referências e documentos eletrônicos.** Disponível em: <<http://www.elogica.com.br/users/gmoura/refere.html>> Acesso em: 09 out. 1996.



APÊNDICE 1 OU ANEXO 1

É possível, com a ciência de dados, criar uma ponte para a interdisciplinaridade no novo ensino médio. Devido ao fato de o aprendizado reverso existir na ciência de dados, se pode aprender várias coisas da mesma maneira e o entendimento seria instantâneo. Desta maneira a sociedade brasileira caminhará em direção ao mundo da tecnologia, e fará também com que a cultura seja maior num país de ensino abaixo da média dos países do mundo afora. A ciência de dados com a metodologia de aprendizado reverso pode ser empregado em todos os campos do conhecimento e do trabalho. Isso geraria um mundo sem desperdício de fonte, de trabalho, de energia e muito mais. É real e clara esta evidência percebida.