

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
CAMPUS DE GUARATINGUETÁ

MATHEUS CERQUEIRA DE JESUS

Técnicas de *Machine Learning* para estimar a correlação de queimadas em Cuiabá nas doenças respiratórias.

Guaratinguetá

2023

Matheus Cerqueira de Jesus

Técnicas de *Machine Learning* para estimar a correlação de queimadas em Cuiabá nas doenças respiratórias.

Trabalho de Graduação apresentado ao Conselho de Curso de Graduação em Engenharia Elétrica da Faculdade de Engenharia do Campus de Guaratinguetá, Universidade Estadual Paulista, como parte dos requisitos para obtenção do diploma de Graduação em Engenharia Elétrica .

Orientador: Profº Dr. Paloma Maria Silva Rocha Rizol

Coorientador: Mestranda Taynara de Oliveira Castellões

Guaratinguetá

2023

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
CAMPUS DE GUARATINGUETÁ

MATHEUS CERQUEIRA DE JESUS

ESTE TRABALHO DE GRADUAÇÃO FOI JULGADO ADEQUADO COMO PARTE DO
REQUISITO PARA A OBTENÇÃO DO DIPLOMA DE "**GRADUANDO EM ENGENHARIA**
ELÉTRICA "

APROVADO EM SUA FORMA FINAL PELO CONSELHO DE CURSO DE GRADUAÇÃO EM
ENGENHARIA ELÉTRICA

Profº Dr. NOME COMPLETO DO COORDENADOR
Coordenador

BANCA EXAMINADORA:

Profº Dr. Paloma Maria Silva Rocha Rizol
Orientador/UNESP-FEG

Profº Dr. Nome Completo do Membro Interno
UNESP-FEG

Profº Dr. Nome Completo do Membro Externo
Membro Externo

Janeiro , 2023

DADOS CURRICULARES

MATHEUS CERQUEIRA DE JESUS

NASCIMENTO	Data - Cidade / Sigla do Estado
FILIAÇÃO	Antonio Sergio de Jesus Maria do Socorro Cerqueira de Jesus
ANO INICIAL / ANO FINAL	Formação acadêmica ou Complementar (nome do curso e titulação) Instituição de Ensino
ANO INICIAL / ANO FINAL	Formação acadêmica ou Complementar (nome do curso e titulação) Instituição de Ensino
ANO INICIAL / ANO FINAL	Formação acadêmica ou Complementar (nome do curso e titulação) Instituição de Ensino

Sua dedicatoria deve ser digitada aqui.

AGRADECIMENTOS

Esse é um espaço reservado para os agradecimentos. Uma nota de rodapé¹ pode ser inserida desta forma para indicar alguma URL² que referencia o alvo de seu agradecimento.

¹ Essa é uma nota de rodapé

² <http:\www.url.com.br>

“Frase, citação, epígrafe.”
(Autor)

RESUMO

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

PALAVRAS-CHAVE: palavra-chave. palavra-chave. palavra-chave.

ABSTRACT

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

KEYWORDS: keyword. keyword. keyword.

LISTA DE ILUSTRAÇÕES

Figura 1	Estrutura das árvores de decisão	19
----------	--	----

LISTA DE TABELAS

Tabela 1 – Publicações no Scopus	16
Tabela 2 – Aplicações de IAs na medicina	18

LISTA DE ABREVIATURAS E SIGLAS

TCC	Trabalho de Conclusão de Curso
UNESP	Universidade Estadual Paulista
IA	Inteligência Artificial

LISTA DE SÍMBOLOS

α	Letra Grega Alfa
β	Letra grega Beta
γ	Letra grega Gama
e	Número de Euler
R\$	Unidade monetária Brasileira (Real)

SUMÁRIO

1	INTRODUÇÃO	15
1.1	objetivos	15
1.2	justificativas	16
1.3	delimitações da pesquisa	16
1.4	estrutura do trabalho	16
2	FUNDAMENTAÇÃO TEÓRICA	17
2.1	python e ciência de dados	17
2.2	Inteligência artificial no contexto da saúde	17
2.2.1	<i>Machine Learning</i>	18
2.2.1.1	<i>Decision Trees</i>	18
2.2.1.2	Random Forest	19
2.2.1.3	Cross Validation	19
2.3	Redes neurais recorrentes	19
2.4	previsão de séries temporais no contexto da saúde	19
2.4.1	Long short term memories	19
2.5	Métricas para avaliar os modelos de regressão	19
2.5.1	Raiz do erro quadrático médio	19
2.5.2	Erro médio absoluto	20
3	MATERIAIS E MÉTODOS	21
3.1	Descrição do problema	21
3.2	Obtenção dos dados	21
3.3	Modelagem do problema	21
4	ANÁLISE DOS RESULTADOS OBTIDOS	22
5	CONCLUSÃO	23
	REFERÊNCIAS	24
	APÊNDICE A – TÍTULO DO APÊNDICE A	25
	APÊNDICE B – TÍTULO DO APÊNDICE B	26
	ANEXO A – TÍTULO DO ANEXO A	27
	ANEXO B – TÍTULO DO ANEXO B	28

ANEXO C – TÍTULO DO ANEXO C	29
--	-----------

1 INTRODUÇÃO

Os últimos séculos foram marcados por períodos de intensos desenvolvimentos tecnológicos, sendo impulsionados pelas revoluções industriais. As indústrias são responsáveis por produzir a maior parte dos produtos essenciais para os seres humanos, bem como o desenvolvimento de novos produtos. Para a produção de carros cada vez mais potentes, escalas enormes de alimentos, dispositivos digitais, produtos farmacêuticos é necessária a produção de energia para o funcionamento das máquinas das indústrias, automóveis para o transporte de cargas e pessoas, e atender as populações das cidades e do campo.

No entanto, de acordo com a (AGENCY, 2023) a matriz energética mundial é majoritariamente composta por fontes de energia não renováveis e altamente poluentes, como os combustíveis fósseis, sendo eles, carvão mineral, petróleo e gás natural. Diante do consumo constante dessas matrizes energéticas os níveis de poluentes na atmosfera passaram a aumentar. E com isso a necessidade de realizar estudos para entender a influencia desses poluentes na saúde humana.

Segundo a (UNIDAS, 2021) uma estimativa realizada em 2021 aponta que a cada ano, a exposição à poluição atmosférica seja responsável por 7 milhões de mortes prematuras e milhões de anos de vida reduzidos de pessoas com uma vida saudável em todo o mundo. Outro agravante relacionado a poluição é o aumento temperatura média do planeta que segundo a (UNIDAS, 2022) em 2021 a temperatura média do planeta foi de 1,1°C acima da linha base pré-industrial. Diante dessa situação, (XU; HU; TONG, 2014) apontam que as altas e baixas temperaturas estão associadas com o aumento da incidência de casos de pneumonia e que as altas variações de temperatura diurnas e entre dias podem afetar o funcionamento do sistema respiratório.

Diante desse contexto, a demanda dos serviços de saúde estão cada vez maiores e situações em que os serviços operam acima da capacidade máxima estão cada vez mais frequentes. Conforme apresentado por (WHO, 2010) os sistemas e serviços de saúde necessitam de recursos essenciais para atuar, incluindo informações hospitalares, doenças e previsões. No entanto, de acordo com (SOYIRI; REIDPATH; SARRAN, 2012), os hospitais, serviços de saúde e fornecedores geralmente não estão adequadamente informados quando entram em uma situação de demanda acima do normal.

Este trabalho, reuniu dados e informações diárias relacionados a qualidade do ar, temperatura ambiente e internações ocorridas em hospitais de Cuiabá no estado do Mato Grosso entre 01/01/2013 e 31/12/2018 para treinar modelos de *Machine Learning* e redes neurais afim de prever o número de internações nos dias subsequentes a uma determinada data, utilizando como dados de entradas os valores de qualidade do ar e temperatura. Os resultados dos diferentes modelos foram comparados, observando principalmente o raiz do erro quadrático médio(RMSE).

1.1 OBJETIVOS

Este trabalho tem como objetivo utilizar técnicas de *Machine Learning* e redes neurais artificiais a fim de prever internações hospitalares de pacientes com doenças respiratórias causadas por poluições atmosféricas.

1.2 JUSTIFICATIVAS

Na plataforma Scopus foram realizadas três pesquisas a fim de verificar o contexto de trabalhos realizados na academia com temas semelhantes aos deste trabalho, conforme apresentado na Tabela 1. Nota-se que para a pesquisa com as palavras chave: Machine learning, respiratory diseases e air quality foram publicados somente 66 artigos, isso mostra ser um tema que ainda não foi muito pesquisado, apesar de estar em crescimento.

Ao adicionar Cuiabá nas palavras chave da busca, não foi retornado nenhum resultado, isso indica que o trabalho realizado é inédito.

Tabela 1 – Publicações no Scopus

Pesquisa	Período	Total de publicações
"Machine Learning"	2012 - 2022	445203
"Machine Learning"e "Respiratory Diseases"	2012 - 2022	2.264
"Machine Learning"e "Respiratory Diseases"e "Air Quality"	2012 - 2022	66

fonte: Scopus (2023)

1.3 DELIMITAÇÕES DA PESQUISA

Ao decorrer do trabalho foram verificadas algumas limitações responsáveis por restringir o alcance do estudo. Os dados utilizados para o treinamento dos modelos de *Machine Learning* e (RNA) possuem colunas com informações sobre a qualidade do ar, temperatura, umidade e internações por doenças respiratórias. Os dados de qualidade do ar são estimativas realizadas pelo SISAM para a cidade de Cuiabá-MT, assim considerou-se que todos os habitantes estão submetidos as mesmas condições.

Ao considerar que todas as pessoas estão influenciadas pelas mesmas condições, desconsidera-se as condições de onde essas pessoas vivem e com o que elas trabalham. Por exemplo, uma pessoa que trabalhe em uma marcenaria está sobre condições respiratórias piores do que outras pessoas e essa influência não foi levada em conta nos modelos por não possuir dados.

Os dados de internações representam apenas as internações ocorridas nos hospitais da rede pública (SUS). Dessa forma, todas as internações por doenças respiratórias ocorridas na rede particular de hospitais não foram contabilizadas. Ainda, as informações das internações por doenças respiratórias são diagnosticadas pelos médicos e inseridas no sistema SIH/SUS, portanto pode haver algum engano ao ser inserida uma informação errada no sistema.

1.4 ESTRUTURA DO TRABALHO

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo foi escrita uma descrição resumida das ferramentas, técnicas e conceitos utilizados no desenvolvimento do trabalho.

2.1 PYTHON E CIÊNCIA DE DADOS

25/01

Python é uma linguagem de programação de computadores multiparadigma e de código aberto (*open source*), que de acordo com (LUTZ, 2013) é otimizada para programar de forma produtiva, ler e entender os códigos com facilidade e qualidade de software.

Python é a linguagem mais utilizada do mundo segundo o rank da linguagens de (CARBONNELLE, 2023) ganhando de Java, Java Script e C++. Essa liderança explica a extensa comunidade que a linguagem possui, sendo isso uma vantagem para quem utiliza, uma vez que é possível encontrar milhares de exemplos de código para uma possível aplicação que alguém precise.

Uma ótima vantagem da linguagem é a incrível quantidade de bibliotecas disponibilizadas pela comunidade para as mais variadas aplicações. Isso torna o Python uma ótima ferramenta para trabalhar com inteligência artificial, *Machine Learning* e *deep learning*. Algumas das principais bibliotecas e *frameworks* utilizadas são Pandas e Numpy para a manipulação de dados, TensorFlow e Scikit Learning para a construção de modelos de *Machine Learning*, como os que serão abordados nos capítulos 2.2 e 2.4.

2.2 INTELIGÊNCIA ARTIFICIAL NO CONTEXTO DA SAÚDE

25/01

Inteligência artificial tem sido um assunto amplamente abordado pelas pessoas em vídeos, notícias e filmes. Muitos acreditam que ela pode ser a solução para todos os problemas existentes na sociedade. No entanto, é importante saber o que realmente é inteligência artificial e para o que ela realmente pode ser usada.

O professor (XIAO, 2022) define em seu livro inteligência artificial como sendo uma área de estudo dentro ciências da computação que possui o objetivo de fazer com que máquinas aprendam a interpretar e resolver problemas de maneira similar ao ser humano. Da mesma forma que um ser humano, aprende ao tentar resolver problemas e obter novas informações, uma inteligência artificial deve tomar uma ação a medida em que recebe novas informações e aprende a fim de melhorar sua performance.

Alguns exemplos de inteligência artificial presente no dia a dia de muitas pessoas são as assistentes virtuais como a Alexa da Amazon e a Cortana da Microsoft, outro exemplo são as ferramentas de anúncio na internet que aprendem com informações sobre os usuários e enviam anúncios de maior interesse. Mas como as IAs, estão relacionadas com a saúde das pessoas e os serviços de assistência médica?

De acordo com Trishan Panch:

A Inteligência Artificial e o aprendizado de máquina têm o potencial de ser o catalisador da transformação dos sistemas de saúde para melhorar a eficiência e a eficácia, criar margem para a cobertura universal de saúde e melhorar os resultados. (PANCH; SZOLOVITS; ATUN, 2018, p.1)

Nos sistemas de saúde existem processos que podem utilizar IAs. Dois exemplos são segundo (PANCH; SZOLOVITS; ATUN, 2018) o diagnóstico de doenças dos pacientes, realizando uma tarefa de classificação, se está ou não com a doença. O outro processo se dá durante o tratamento, envolvendo predição de uma melhora ou piora no quadro do paciente, monitorando os dados vitais. Algumas das aplicações existentes foram apresentadas na tabela 2.

Tabela 2 – Aplicações de IAs na medicina

Diagnóstico	Prognóstico e predição
Análise de imagem: Mamografia	Hospitalização por doença cardíaca
Análise de sinais: Monitoramento intraparto	Risco de acidente cardiovascular
Análise de imagem: identificação da retinopatia diabética	Predição de resultados em câncer colorretal

fonte: Adaptado de (PANCH; SZOLOVITS; ATUN, 2018)

Diante disso, é notável que as IAs podem contribuir para melhorar os serviços de saúde, impactando a vida de milhares de pessoas. A seguir serão abordadas algumas técnicas de aprendizado de máquinas utilizadas neste trabalho.

2.2.1 *Machine Learning*

26/01

2.2.1.1 *Decision Trees*

26/01

Árvores de decisão (*Decision Trees*) são algoritmos de *machine learning* muito versáteis, que podem ser utilizados tanto para resolver problemas de classificação quanto de regressão. Como o próprio nome diz as *Decision Trees* possuem uma estrutura hierárquica em árvore, contendo o nó raiz (*root node*), as ramificações, nós de decisão (*decision nodes*) e as folhas (*leaf nodes*). Uma estrutura simples desse algoritmo está apresentada na Figura 1.

O modelo de *Decision Trees* utilizado nesse trabalho realizou uma tarefa de regressão, ou seja, uma *Regression Tree*, nessa variação todos os nós da árvore possuem valores numéricos. Imaginando um conjunto de pontos (x, y), O primeiro passo para montar a árvore é obter o nó raiz, ele é encontrado ao calcular iterativamente para o valor médio entre dois x adjacentes o erro quadrático médio (MSE) com a média dos valores de y a esquerda do ponto médio e a média dos valores de y a direita do ponto médio. O valor médio que obtiver o menor MSE será o nó raiz da árvore.

Para finalizar todo o processo de construção o processo é repetido para o lado esquerdo do nó raiz e para o lado direito até que não sobrem mais valores possíveis para as folhas.

Figura 1 – Estrutura das árvores de decisão



fonte: Produção do próprio autor.

2.2.1.2 Random Forest

26/01

2.2.1.3 Cross Validation

26/01

2.3 REDES NEURAIIS RECORRENTES

27/01

2.4 PREVISÃO DE SÉRIES TEMPORAIS NO CONTEXTO DA SAÚDE

27/01

2.4.1 Long short term memories

27/01

2.5 MÉTRICAS PARA AVALIAR OS MODELOS DE REGRESSÃO

27/01

2.5.1 Raiz do erro quadrático médio

27/01

2.5.2 Erro médio absoluto

27/01

3 MATERIAIS E MÉTODOS

3.1 DESCRIÇÃO DO PROBLEMA

28/01

Escolha da cidade de Cuiabá Numero de hanitantes queimadas na cidade
explicar os principais poluentes utilizados nos modelos

3.2 OBTENÇÃO DOS DADOS

28/01

Quantidade de dados filtragem dos dados tratamento dos dados nulos

3.3 MODELAGEM DO PROBLEMA

28/01

temperatura, umidade, pm25 temperatura, umidade, pm25, oxonio

4 ANÁLISE DOS RESULTADOS OBTIDOS

28/01

1. uma coisa

- outra coisa

2. tres coisas

5 CONCLUSÃO

28/01

Deve ser fundamentada no texto, contendo deduções lógicas e correspondentes aos objetivos propostos.

Responder os objetivos

REFERÊNCIAS

AGENCY, I. E. **Energy Statistics Data Browser**. 2023. Disponível em: <<https://www.iea.org/data-and-statistics/data-tools/energy-statistics-data-browser?country=WORLD&fuel=Energy%20supply&indicator=TESbySource>>.

CARBONNELLE, P. **PYPL PopularitY of Programming Language**. 2023. Disponível em: <<https://pypl.github.io/PYPL.html>>.

LUTZ, M. **Learning Python, 5th Edition**. Sebastopol, Califórnia - EUA: O'REILLY, 2013.

PANCH, T.; SZOLOVITS, P.; ATUN, R. Artificial intelligence, machine learning and health systems. **Journal of Global Health**, v. 8, n. 8, p. 1 – 8, 2018.

SOYIRI, I. N.; REIDPATH, D. D.; SARRAN, C. Forecasting peak asthma admissions in london: an application of quantile regression models. **International Journal of Biometeorology**, v. 1, n. 2, p. 1 – 11, 2012.

UNIDAS, O. das N. **Novas diretrizes da OMS sobre qualidade do ar reduzem valores seguros para poluição**. 2021. Disponível em: <<https://brasil.un.org/pt-br/145721-novas-diretrizes-da-oms-sobre-qualidade-do-ar-reduzem-valores-seguros-para-poluicao>>.

UNIDAS, O. das N. **Temperatura média global tem 50exceder 1,5°C até 2026**. 2022. Disponível em: <<https://brasil.un.org/pt-br/181236-temperatura-media-global-tem-50-de-chance-de-exceder-15degc-ate-2026#:~:text=De%20acordo%20com%20os%20novos,%2C%20saltou%20para%20quase%2050%25.>>>

WHO. **Health service delivery**. Genève, Switzerland: World Health Organization, 2010.

XIAO, P. **Artificial Intelligence Programming with Python: From Zero to Hero**. Hoboken, Nova Jersey - EUA: Wiley, 2022.

XU, Z.; HU, W.; TONG, S. Temperature variability and childhood pneumonia: An ecological study. **Environmental Health**, v. 1, n. 8, p. 1 – 8, 2014.

APÊNDICE A – TÍTULO DO APÊNDICE A

Quisque facilisis auctor sapien. Pellentesque gravida hendrerit lectus. Mauris rutrum sodales sapien. Fusce hendrerit sem vel lorem. Integer pellentesque massa vel augue. Integer elit tortor, feugiat quis, sagittis et, ornare non, lacus. Vestibulum posuere pellentesque eros. Quisque venenatis ipsum dictum nulla. Aliquam quis quam non metus eleifend interdum. Nam eget sapien ac mauris malesuada adipiscing. Etiam eleifend neque sed quam. Nulla facilisi. Proin a ligula. Sed id dui eu nibh egestas tincidunt. Suspendisse arcu.

APÊNDICE B – TÍTULO DO APÊNDICE B

Maecenas dui. Aliquam volutpat auctor lorem. Cras placerat est vitae lectus. Curabitur massa lectus, rutrum euismod, dignissim ut, dapibus a, odio. Ut eros erat, vulputate ut, interdum non, porta eu, erat. Cras fermentum, felis in porta congue, velit leo facilisis odio, vitae consectetur lorem quam vitae orci. Sed ultrices, pede eu placerat auctor, ante ligula rutrum tellus, vel posuere nibh lacus nec nibh. Maecenas laoreet dolor at enim. Donec molestie dolor nec metus. Vestibulum libero. Sed quis erat. Sed tristique. Duis pede leo, fermentum quis, consectetur eget, vulputate sit amet, erat.

ANEXO A – TÍTULO DO ANEXO A

Sed mattis, erat sit amet gravida malesuada, elit augue egestas diam, tempus scelerisque nunc nisl vitae libero. Sed consequat feugiat massa. Nunc porta, eros in eleifend varius, erat leo rutrum dui, non convallis lectus orci ut nibh. Sed lorem massa, nonummy quis, egestas id, condimentum at, nisl. Maecenas at nibh. Aliquam et augue at nunc pellentesque ullamcorper. Duis nisl nibh, laoreet suscipit, convallis ut, rutrum id, enim. Phasellus odio. Nulla nulla elit, molestie non, scelerisque at, vestibulum eu, nulla. Ut odio nisl, facilisis id, mollis et, scelerisque nec, enim. Aenean sem leo, pellentesque sit amet, scelerisque sit amet, vehicula pellentesque, sapien.

ANEXO B – TÍTULO DO ANEXO B

Sed consequat tellus et tortor. Ut tempor laoreet quam. Nullam id wisi a libero tristique semper. Nullam nisl massa, rutrum ut, egestas semper, mollis id, leo. Nulla ac massa eu risus blandit mattis. Mauris ut nunc. In hac habitasse platea dictumst. Aliquam eget tortor. Quisque dapibus pede in erat. Nunc enim. In dui nulla, commodo at, consectetur nec, malesuada nec, elit. Aliquam ornare tellus eu urna. Sed nec metus. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas.

ANEXO C – TÍTULO DO ANEXO C

Phasellus id magna. Duis malesuada interdum arcu. Integer metus. Morbi pulvinar pellentesque mi. Suspendisse sed est eu magna molestie egestas. Quisque mi lorem, pulvinar eget, egestas quis, luctus at, ante. Proin auctor vehicula purus. Fusce ac nisl aliquam ante hendrerit pellentesque. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Morbi wisi. Etiam arcu mauris, facilisis sed, eleifend non, nonummy ut, pede. Cras ut lacus tempor metus mollis placerat. Vivamus eu tortor vel metus interdum malesuada.