

**MONITORAMENTO DO OZÔNIO EM CARAPICUÍBA:
ANÁLISE DOS DADOS DO IEMA ENTRE 2015 E 2021**

**MONITORING OZONE IN CARAPICUÍBA:
ANALYSIS OF IEMA DATA FROM 2015 TO 2021**

**MONITOREO DEL OZONO EN CARAPICUÍBA:
ANÁLISIS DE LOS DATOS DEL IEMA ENTRE 2015 Y 2021**

Rita de Cássia Felix Carboni¹
Gustavo de Deus Conceição²
José Antonio Ramos Barreto Gomes³
Gustavo Cardoso Santos da Silva⁴

Artigo recebido em agosto de 2024
Artigo aceito em dezembro de 2024

DOI: 10.26853/Refas_ISSN-2359-182X_v11n03_06

RESUMO

O estudo investigou os níveis de ozônio (O₃) em Carapicuíba entre 2015 e 2021, utilizando dados do Instituto de Energia e Meio Ambiente (IEMA). A análise destacou a formação fotoquímica do ozônio, intensificada pela radiação solar e pela presença de precursores como NO_x e VOCs. Os resultados mostraram um aumento das concentrações de ozônio de 2015 a 2019, seguido por uma redução significativa durante a pandemia de COVID-19 devido às restrições nas atividades industriais e no tráfego. A concentração média horária de ozônio aumentava a partir das 10 horas, atingindo o pico entre 14 e 16 horas, refletindo a influência da radiação solar. A análise dos Índices de Qualidade do Ar (IQA_r) revelou que, na maioria das vezes, a qualidade do ar foi satisfatória, especialmente nos horários de pico. No entanto, em dias com altos níveis de emissão, a qualidade do ar pode ser classificada como moderada ou ruim, especialmente antes das intervenções da pandemia. A pesquisa sugere a importância de continuar monitorando e implementando estratégias de mitigação para reduzir as emissões de precursores de ozônio.

Palavras-chave: Ozônio (O₃); Radiação solar; Índices de Qualidade do Ar (IQA_r).

ABSTRACT

The study investigated ozone (O₃) levels in Carapicuíba between 2015 and 2021, using data from the Institute of Energy and Environment (IEMA). The analysis highlighted the photochemical formation of ozone, intensified by solar radiation and the presence of precursors such as NO_x and VOCs. Results

¹ Mestre em Educação com pós-graduação em Administração Geral e Bacharelado em Ciências da Computação. Professora na Fatec Carapicuíba. E-mail: rita.felix@fatec.sp.gov.br. Lattes: <http://lattes.cnpq.br/5879278240836544>. OrcId: <https://orcid.org/0009-0006-4662-4574>.

² Desenvolvedor Full Stack e Discente no curso Análise e Desenvolvimento de Sistemas na Fatec de Carapicuíba. E-mail: gustavodeusconceicao@gmail.com. OrcId: <https://orcid.org/0009-0004-2852-9050>.

³ Graduando em Análise e Desenvolvimento de Sistemas pela Fatec de Carapicuíba. E-mail: jose.gomes34@fatec.sp.gov.br. OrcId: <https://orcid.org/0009-0000-0587-7707>.

⁴ Graduando em Análise e Desenvolvimento de Sistemas pela Fatec de Carapicuíba. E-mail: gustavo.silva444@fatec.sp.gov.br. OrcId: gustavo.silva444@fatec.sp.gov.br.

showed an increase in ozone concentrations from 2015 to 2019, followed by a significant decrease during the COVID-19 pandemic due to restrictions on industrial activities and traffic. The hourly average concentration of ozone increased from 10 AM, peaking between 2 PM and 4 PM, reflecting the influence of solar radiation. The analysis of Air Quality Indices (IQA_r) revealed that most of the time, air quality was satisfactory, especially during peak hours. However, on days with high emission levels, air quality could be classified as moderate or poor, especially before the pandemic interventions. The research suggests the importance of continuing to monitor and implement mitigation strategies to reduce ozone precursor emissions.

Keywords: Ozone (O₃); Solar radiation; Air Quality Indices (IQA_r).

RESUMEN

El estudio investigó los niveles de ozono (O₃) en Carapicuíba entre 2015 y 2021, utilizando datos del Instituto de Energía y Medio Ambiente (IEMA). El análisis destacó la formación fotoquímica del ozono, intensificada por la radiación solar y la presencia de precursores como NO_x y VOCs. Los resultados mostraron un aumento en las concentraciones de ozono de 2015 a 2019, seguido de una disminución significativa durante la pandemia de COVID-19 debido a las restricciones en las actividades industriales y el tráfico. La concentración promedio horaria de ozono aumentaba a partir de las 10 de la mañana, alcanzando su punto máximo entre las 2 y las 4 de la tarde, reflejando la influencia de la radiación solar. El análisis de los Índices de Calidad del Aire (IQA_r) reveló que, la mayoría de las veces, la calidad del aire fue satisfactoria, especialmente durante las horas pico. Sin embargo, en días con altos niveles de emisión, la calidad del aire podía clasificarse como moderada o mala, especialmente antes de las intervenciones de la pandemia. La investigación sugiere la importancia de continuar monitoreando e implementando estrategias de mitigación para reducir las emisiones de precursores del ozono.

Palabras clave: Ozono (O₃); Radiación solar; Índices de Calidad del Aire (IQA_r).

1 INTRODUÇÃO

Desde que surgiram seus primeiros ancestrais na superfície da Terra, há aproximadamente um milhão de anos, na porção mais ao sul do continente africano, o homem tem atuado de forma transformadora e, muitas vezes, predatória sobre a natureza. A partir da descoberta do fogo, aproximadamente 800 mil anos antes de Cristo, o homem passou a contribuir de forma atuante, porém não consciente, para a deterioração da qualidade do ar e a sofrer as consequências desse ato (Braga et al., 2001).

Diante das catástrofes causadas pela poluição gerada pelo homem, torna-se essencial entender e contextualizar os poluentes atmosféricos e suas concentrações. Isso deve ser feito considerando as diferentes cidades, estados e países, bem como suas políticas ambientais. Os poluentes não apenas afetam as áreas próximas às suas fontes de emissão, mas também podem se dispersar por milhares de quilômetros, impactando regiões distantes (Leite et al., 2011).

Nesse contexto, torna-se crucial investigar a poluição atmosférica, com especial atenção à concentração de ozônio troposférico, cuja formação está diretamente relacionada às concentrações de óxidos de nitrogênio (NO_x) e compostos orgânicos voláteis (VOCs). Estudos sugerem que a presença desses precursores, emitidos principalmente por atividades industriais e veiculares, é determinante na formação de ozônio na troposfera (Zhou et al., 2020). Assim, a compreensão detalhada das concentrações desses poluentes e seus impactos é essencial para o desenvolvimento de políticas eficazes de controle da qualidade do ar.

Nos últimos anos, o ozônio (O_3) emergiu como um indicador crucial da qualidade do ar e da saúde ambiental em áreas urbanas, refletindo a concentração desse gás na atmosfera, resultado direto da emissão de poluentes. No município de Carapicuíba, situado no estado de São Paulo, Brasil, o rápido processo de industrialização e urbanização tem suscitado crescentes preocupações acerca da qualidade do ar e dos impactos da poluição atmosférica na saúde pública.

O ozônio é formado naturalmente na atmosfera por meio de reações químicas entre oxigênio (O_2) e radiação ultravioleta (UV) do Sol. Essas reações ocorrem principalmente na estratosfera, onde está localizada a camada de ozônio.

Além disso, é importante considerar o ozônio como um poluente secundário na troposfera. Na troposfera, abaixo da camada mais elevada, o ozônio é formado como resultado de reações fotoquímicas, principalmente com óxidos de nitrogênio (NO_x) e compostos orgânicos voláteis (VOCs). Essas reações ocorrem na presença de luz solar e são um fenômeno comum em áreas urbanas altamente industrializadas, como Carapicuíba. O ozônio troposférico, em grandes concentrações, pode causar graves complicações à saúde e ao meio ambiente.

Este artigo tem como objetivo principal investigar a presença e os efeitos dos poluentes atmosféricos em Carapicuíba, com especial atenção aos compostos de ozônio, até o ano de 2021. A análise do ozônio nesse contexto é crucial para avaliar a qualidade do ar e identificar possíveis riscos à saúde pública associados à poluição atmosférica. Partimos dos referenciais de que os níveis de ozônio em Carapicuíba estão intimamente ligados às atividades industriais e de transporte na região. A análise detalhada desses níveis pode fornecer insights cruciais para o desenvolvimento de estratégias de mitigação eficazes e políticas de controle da poluição.

2 REFERENCIAL TEÓRICO

Inicia-se o referencial teórico pela mineração de dados.

2.1 Mineração de dados

A Mineração de Dados (em inglês: Data Mining) é um processo não somente de identificação de padrões, mas também correlações e anomalias em determinado Conjunto de Dados (em inglês: Dataset), porém alguns especialistas podem ser mais profundos em relação a este tema.

Segundo Gillis; Stadman (2021), a mineração de dados é uma parte fundamental da análise de dados e uma das principais disciplinas da ciência de dados, que utiliza técnicas analíticas avançadas para encontrar informações úteis em conjuntos de dados. Em um nível mais granular, a mineração de dados é uma etapa do processo de descoberta de conhecimento em bancos de dados (KDD), uma metodologia de ciência de dados para coleta, processamento e análise de dados. A mineração de dados e o KDD às vezes são chamados de forma intercambiável, mas são mais comumente vistos como coisas distintas (Gillis; Stadman, 2021). Em um artigo publicado na Acta Paulista de Enfermagem, os autores realizaram uma revisão da literatura sobre a técnica de mineração de dados (Data Mining - DM). Eles observaram que o volume de dados armazenados é gigantesco e continua crescendo exponencialmente. O

processo de Descoberta do Conhecimento em Bases de Dados e DM inclui tarefas e métodos para extração de conhecimento útil, interessante e indispensável na tomada de decisões rápidas nas mais diversas áreas de conhecimento (Galvão et al, 2010).

A mineração de dados é uma etapa essencial no processo de descoberta de conhecimento em bancos de dados (KDD) que produz padrões ou modelos úteis a partir de dados (Fayyad; Piatetsky-Shapiro; Smyth, 1996). Amigo (2021), ressalta que a Mineração de Dados pode ser definida como um conjunto de métodos usados para extrair informações utilizáveis de grandes volumes brutos. conjuntos de dados. Deve-se notar que esta definição implica que a informação utilizável já está nos dados. No entanto, a complexidade dos dados e a natureza multivariada (multiway) dos dados significa que são incapazes de encontrar informações úteis sem ferramentas matemáticas poderosas. Basicamente, a mineração de dados visa separar o grão do feno ou encontrar padrões que já existem nos dados; no entanto, eles estão ocultos devido ao grande número de amostras e variáveis, ao ruído dos dados ou à dificuldade de vincular mais de duas variáveis ao mesmo tempo de forma univariada (uma variável por vez) (Amigo, 2021).

Com base nas referências teóricas apresentadas, podemos concluir que a mineração de dados é uma disciplina essencial na ciência de dados. Ela utiliza técnicas analíticas avançadas para descobrir tais elementos até mesmo sendo aplicados na questão da emissão de Ozônio, como por exemplo é feito pelo Serviço de Monitorização da Atmosfera Copérnico (sigla em inglês: CAMS - Copernicus Atmosphere Monitoring Service) no levantamento de dados (European Environment Agency, 2022). A mineração de dados é uma etapa crucial no processo de descoberta de conhecimento em bancos de dados (KDD), que envolve a coleta, processamento e análise de dados. Os especialistas concordam que o volume de dados armazenados está crescendo exponencialmente, tornando a mineração de dados ainda mais relevante. O processo de mineração de dados inclui tarefas e métodos para extrair conhecimento útil, interessante e indispensável para a tomada de decisões rápidas em várias áreas de conhecimento.

E em relação as tarefas, segundo Larose; Larose (2005), as tarefas mais comuns são:

- a) Descrição (Description) É a tarefa utilizada para descrever os padrões e tendências revelados pelos dados. A descrição geralmente oferece uma possível interpretação para os resultados obtidos. A tarefa de descrição é muito utilizada em conjunto com as técnicas de análise exploratória de dados, para comprovar a influência de certas variáveis no resultado obtido.
- b) Classificação (Classification) Uma das tarefas mais comuns, a Classificação, visa identificar a qual classe um determinado registro pertence. Nesta tarefa, o modelo analisa o conjunto de registros fornecidos, com cada registro já contendo a indicação à qual classe pertence, a fim de 'aprender' como classificar um novo registro (aprendizado supervisionado).
- c) Estimação (Estimation) ou Regressão (Regression) A estimação é similar à classificação, porém é usada quando o registro é identificado por um valor numérico e não um categórico. Assim, pode-se estimar o valor de uma determinada variável analisando-se os valores das demais.
- d) Predição (Prediction) A tarefa de predição é similar às tarefas de classificação e estimação, porém ela visa descobrir o valor futuro de um determinado atributo.
- e) Agrupamento (Clustering) A tarefa de agrupamento visa identificar e aproximar os registros similares. Um agrupamento (ou cluster) é uma coleção de registros similares entre si, porém diferentes dos outros registros nos demais agrupamentos. Esta tarefa difere da classificação pois não necessita que os registros sejam previamente categorizados (aprendizado não-supervisionado).

2.2 Processo KDD

Os acadêmicos Ralf Mikut e Markus Reischl classificam o KDD como o processo não trivial de identificação de padrões válidos, novos, potencialmente úteis e, em última análise, compreensíveis em dados por sua tamanha profundidade (Mikut; Reischl, 2011). KDD é a sigla para Descoberta de Cohecimento em Base (ou Banco) de Dados (do Inglês: Knowledge Discovery Database, as vezes com um hífen em Knowledge-Discovery), e ao ver a fala de Sankar e Pabitra, onde diz que o KDD utiliza dados coletados de bancos de dados internos e externos às organizações. Esses dados são analisados de diferentes ângulos e perspectivas para descobrir relacionamentos. O objetivo é aproveitar o ativo de dados para melhorar tomando uma decisão. KDD pode ser definido como o processo de identificação de produtos válidos, novos, potencialmente úteis e, em última análise, padrões compreensíveis nos dados.

Isso requer um banco de dados grande e bem integrado e um bom entendimento do negócio contexto dentro do qual o DM deve ser aplicado (Sankar; Pabitra, 2004), concluimos que o processo de Descoberta de Conhecimento em Base de Dados (KDD) é essencial para extrair informações valiosas e insights significativos a partir de grandes conjuntos de dados. Ele envolve a identificação de padrões válidos e úteis, que podem ser utilizados para melhorar a tomada de decisões em diversas áreas, desde negócios até pesquisa científica. Para realizar o KDD de forma eficaz, é fundamental ter acesso a bancos de dados internos e externos, permitindo uma análise abrangente e multifacetada dos dados.

Além disso, é necessário um profundo entendimento do contexto do negócio ou do problema em questão para garantir que os resultados obtidos sejam relevantes e aplicáveis. Sumamente, o KDD é um processo complexo e não trivial, mas pode proporcionar insights valiosos que podem impulsionar o progresso e o sucesso em diversos campos.

2.3 Random forest

Random Forest, ou Floresta Aleatória, (Breiman, 2001) é um algoritmo estatístico ou de aprendizado de máquina para previsão.

Os pesquisadores Wang, Jinyang et al. descrevem a aplicação do modelo de Random Forest como uma abordagem não trivial para prever a concentração de ozônio na troposfera, que é um problema complexo e multifacetado (Wang et al., 2024). O Random Forest é um algoritmo de aprendizado de máquina que utiliza uma coleção de árvores de decisão para fazer previsões ou classificações.

No contexto da previsão de ozônio, o Random Forest é usado para analisar dados coletados de instrumentos de monitoramento de ozônio, como o OMI (Ozone Monitoring Instrument), e identificar padrões válidos, novos e potencialmente úteis nos dados. Os dados são analisados de diferentes ângulos e perspectivas para descobrir relacionamentos entre as variáveis e a concentração de ozônio. O objetivo é aproveitar o ativo de dados para melhorar a previsão da concentração de ozônio e, assim, auxiliar na tomada de decisões relacionadas à gestão da qualidade do ar.

A aplicação do Random Forest na previsão de ozônio requer um grande e bem integrado banco de dados e um bom entendimento do contexto ambiental dentro do qual o modelo deve ser aplicado. Outro bom exemplo é o de Pan, Qilong; Harrou, Fouzi; Sun, Ying, que também destacam a importância do Random Forest na previsão da poluição por ozônio. Eles avaliaram o desempenho preditivo de dezenove modelos de aprendizado de máquina para a previsão da poluição por ozônio. Especificamente, eles avaliaram como a incorporação de características

usando Random Forest afeta a previsão da concentração de ozônio e investigaram o uso de medidas com atraso temporal para melhorar a precisão da previsão (Pan; Harrou; Sun, 2023).

Um experimento parecido, porém, feito por Domínguez-García, Rodrigo; Arellano-Vázquez, Magali, focou na avaliação de diferentes modelos, como Random Forest (RF), Support Vector Regression (SVR) e Gradient Boosting (GB), para prever a concentração de ozônio 24 horas adiantado, usando dados do Sistema de Monitoramento Atmosférico da Cidade do México e variáveis meteorológicas que influenciam o fenômeno de dispersão e formação de ozônio (Domínguez-García; Arellano-Vázquez, 2024).

Concluimos que a aplicação do Random Forest na previsão de ozônio é essencial para extrair informações valiosas e insights significativos a partir de grandes conjuntos de dados. Ele envolve a identificação de padrões válidos e úteis, que podem ser utilizados para melhorar a tomada de decisões em diversas áreas, desde a gestão da qualidade do ar até a pesquisa científica. Para realizar a previsão de ozônio de forma eficaz, é fundamental ter acesso a bancos de dados internos e externos, permitindo uma análise abrangente e multifacetada dos dados.

Além disso, é necessário um profundo entendimento do contexto ambiental para garantir que os resultados obtidos sejam relevantes e aplicáveis. Em suma, a aplicação do Random Forest na previsão de ozônio é um processo complexo e não trivial, mas pode proporcionar insights valiosos que podem impulsionar o progresso e o sucesso em diversos campos.

2.4 Métricas de avaliação

As métricas de avaliação são fundamentais para medir o desempenho dos modelos de Machine Learning. Neste contexto, quatro métricas populares são frequentemente utilizadas:

- 1) **Acurácia:** A acurácia é uma métrica que mede a proporção de predições corretas feitas pelo modelo em relação ao total de predições. Em termos técnicos, a acurácia é a quantidade de acertos do nosso modelo dividido pelo total da amostra. Por exemplo, em um algoritmo que classifica imagens de animais como gatos e não gatos, a acurácia seria a soma de tudo que o algoritmo acertou (sejam eles verdadeiros positivos ou verdadeiros negativos) dividido pelo total de amostras.
- 2) **Precisão:** A precisão mede a quantidade de vezes que o modelo acerta em relação ao total de vezes que ele tenta acertar. Em outras palavras, de todos os dados classificados como positivos, quantos são realmente positivos. A precisão é o número de exemplos que seu modelo previu como positivos e acertou dividido pelo número total de exemplos que ele previu como positivos.
- 3) **Recall:** O recall mede a quantidade de vezes que o modelo acerta em relação ao total de vezes que ele deveria ter acertado. Ou seja, qual a porcentagem de dados classificados como positivos comparado com a quantidade real de positivos que existem em nossa amostra.
- 4) **F1-Score:** O F1-Score é uma métrica que combina a precisão e o recall de maneira equilibrada. Ele é especialmente útil quando temos um conjunto de dados desbalanceado, ou seja, quando uma classe é muito mais frequente do que as outras. O F1-Score é calculado a partir da média harmônica entre a precisão e o recall, fornecendo uma medida balanceada do desempenho do modelo.

De acordo com Mateus Pádua, as métricas de avaliação, incluindo Acurácia, Precisão, Recall e F1-score, são cruciais para avaliar o desempenho de modelos de Machine Learning

(Padua, 2020). Filho (2023) destaca a importância da precisão, recall e F1-score em Machine Learning, ressaltando sua aplicação e relevância na avaliação de modelos.

Em resumo, a acurácia responde à pergunta: qual a proporção de gatos e não gatos que foram corretamente classificados? A precisão responde: qual a proporção dos dados classificados como gatos eram realmente gatos? O recall responde: entre todas as amostras que realmente eram de gatos, qual a proporção classificada como gatos? E o F1-Score é uma maneira de observar em um único número a precisão e o recall.

2.5 Poluentes

De acordo com a Companhia Ambiental do Estado de São Paulo (CETESB, 2013), poluente é definido como qualquer substância presente no ar cuja concentração possa torná-lo impróprio, nocivo ou ofensivo à saúde humana, causando inconvenientes ao bem-estar público, danos a materiais, fauna e flora, ou prejudicando a segurança, uso e gozo da propriedade, bem como as atividades normais da comunidade.

Para simplificar a diversidade de poluentes encontrados na atmosfera, eles são comumente classificados em duas categorias: poluentes primários e secundários. Os poluentes primários são emitidos diretamente pelas fontes de emissão, enquanto os poluentes secundários são formados na atmosfera por meio de reações químicas entre poluentes primários e componentes naturais da atmosfera. Exemplos de poluentes primários incluem dióxido de enxofre (SO₂), óxidos de nitrogênio (NO_x) e material particulado (PM), enquanto ozônio troposférico (O₃) e material particulado orgânico (MPO) são exemplos de poluentes secundários.

2.6 Ozônio (O₃)

“Oxidantes fotoquímicos” é a denominação que se dá à mistura de poluentes secundários formados por reações entre os óxidos de nitrogênio e compostos orgânicos voláteis, na presença de luz solar, sendo estes últimos liberados na queima incompleta e evaporação de combustíveis e solventes. O principal produto dessa reação é o ozônio, por isso mesmo utilizado como parâmetro indicador da presença de oxidantes fotoquímicos na atmosfera. Tais poluentes formam a chamada névoa fotoquímica ou “smog fotoquímico”, que possui esse nome porque causa na atmosfera diminuição da visibilidade. (CETESB, 2013).

O ozônio troposférico, também conhecido como ozônio "mau", é de fato um problema sério de qualidade do ar. Diferente da camada de ozônio estratosférico, que é benéfica ao bloquear os raios UV nocivos do sol, o ozônio troposférico é formado através de reações químicas entre poluentes atmosféricos, como óxidos de nitrogênio (NO_x) e compostos orgânicos voláteis (COVs), na presença de luz solar. Ele se concentra em níveis mais baixos da atmosfera, onde as pessoas respiram.

Quando inalado em concentrações elevadas, o ozônio troposférico pode causar uma série de problemas de saúde, como irritação respiratória, tosse, falta de ar, piora de doenças respiratórias pré-existentes (como asma e bronquite) e até mesmo danos aos pulmões a longo prazo.

Além de prejuízos à saúde, o ozônio pode causar danos à vegetação. É sempre bom ressaltar que o ozônio encontrado na faixa de ar próxima do solo, onde respiramos, chamado de “mau ozônio”, é tóxico. Entretanto, na estratosfera (cerca de 25 km de altitude) o ozônio tem

a importante função de proteger a Terra, como um filtro, dos raios ultravioletas emitidos pelo Sol. (CETESB, 2013).

Portanto, é importante monitorar e controlar os níveis de ozônio troposférico para proteger a saúde pública e melhorar a qualidade do ar. Isso geralmente envolve a redução das emissões de poluentes que contribuem para sua formação, como os mencionados NO_x e COVs, e medidas para limitar a exposição humana em períodos de alta concentração.

O próximo tópico, a ser retratado, é sobre os índices de qualidade do ar (IQAr).

2.7 Índices de qualidade do ar (IQAr)

Segundo a FEPAM (2018), O Índice de Qualidade do Ar é uma ferramenta matemática utilizada para transformar as concentrações medidas dos diversos poluentes em um único valor adimensional que possibilita a comparação com os limites legais de concentração para os diversos poluentes (Padrões de Qualidade do Ar - PQAr).

Os IQA geralmente levam em consideração uma variedade de poluentes atmosféricos, como dióxido de enxofre (SO₂), óxidos de nitrogênio (NO_x), material particulado (PM), ozônio troposférico (O₃) e monóxido de carbono (CO). Cada poluente tem seus próprios padrões de qualidade do ar definidos por agências reguladoras com base em pesquisas científicas sobre seus efeitos na saúde humana e no meio ambiente.

Os IQA podem ser apresentados em uma escala numérica ou em categorias, como "bom", "moderado", "prejudicial para grupos sensíveis" e "prejudicial para todos". Essas categorias geralmente refletem os níveis de poluentes presentes no ar e os possíveis impactos na saúde pública.

Os IQA são frequentemente utilizados por agências de proteção ambiental, autoridades de saúde pública e organizações não governamentais para alertar o público sobre a qualidade do ar e fornece orientações sobre atividades ao ar livre. E neste trabalho para analisar os índices do O₃ presente na cidade de Carapicuíba que serão explorados nos próximos capítulos.

3 MÉTODO

Os dados utilizados neste trabalho são referentes à ferramenta de qualidade do ar da IEMA (2021). Os dados escolhidos são referentes ao estado de São Paulo, organizados em uma planilha que, ao agruparmos os dados, possui aproximadamente 1 gigabyte de informações brutas referentes aos anos de 2015 a 2021 para a maioria dos municípios do estado. No entanto, neste trabalho, serão avaliados apenas os dados referentes ao município de Carapicuíba. A seguir, estão os tópicos referentes à extração de conhecimento (KDD - Knowledge Discovery in Databases).

3.1 KDD

Para o processo de KDD e desenvolvimento do projeto será utilizado Python e as seguintes tecnologias: Visual Studio Code; Excel; Pandas; Matplotlib; Scikit-learn; e Numpy.

3.1.1 Seleção

Para realizar a leitura da planilha, utilizou-se a biblioteca Pandas, e iniciamos a seleção dos dados conforme demonstrado na Figura 1.

Figure 1 - Processo de Seleção dos dados

```
import pandas as pd

df = pd.read_csv("./SP_poluicao_dados.csv").drop(columns=['Unnamed: 0'])

print("Quantidade de linhas no DataFrame original:", len(df))

display(df.head())
```

[4] ✓ 23.1s Python

Quantidade de linhas no DataFrame original: 10980515

	ID	Data	Hora	Estacao	Codigo	Poluente	Valor	Unidade	Tipo
0	0	2015-01-01	01:00	Americana - Vila Santa Maria	SP01	MP10	65.0	ug/m3	automatica
1	1	2015-01-01	02:00	Americana - Vila Santa Maria	SP01	MP10	98.0	ug/m3	automatica
2	2	2015-01-01	03:00	Americana - Vila Santa Maria	SP01	MP10	79.0	ug/m3	automatica
3	3	2015-01-01	04:00	Americana - Vila Santa Maria	SP01	MP10	53.0	ug/m3	automatica
4	4	2015-01-01	05:00	Americana - Vila Santa Maria	SP01	MP10	35.0	ug/m3	automatica

Fonte: Os autores

Com a visualização das primeiras linhas dos dados, observou-se a primeira medição referente a Americana – Vila Santa Maria. A partir dos filtros necessários ao projeto, criou-se um dataset contendo apenas os dados do poluente 03 (ozônio) e do município de Carapicuíba, como demonstrado na Figura 2.

Figure 2 - Seleção dos dados de Carapicuíba e do poluente 03

```
estacoes = ['Carapicuíba']
poluente = ['03']
df_filtrado = df[df['Estacao'].isin(estacoes) & df['Poluente'].isin(poluente)]
print("\nQuantidade de linhas no DataFrame filtrado:", len(df_filtrado))
display(df_filtrado.describe())
```

[12] Python

Quantidade de linhas no DataFrame filtrado: 56340

	ID	Valor
count	5.634000e+04	56340.000000
mean	4.974299e+06	35.698953
std	3.226230e+06	30.591168
min	1.801890e+05	0.000000
25%	1.576492e+06	13.000000
50%	4.983404e+06	30.000000
75%	8.174583e+06	50.000000
max	9.679523e+06	298.000000

Fonte: Os autores

Para validar as datas de início e término da coleta de dados referentes ao poluente, utilizamos funções que demonstram as primeiras e últimas amostras do *dataset*, como ilustrado nas Figuras 3 e 4, e por fim salvando o *dataset*.

Figure 3 - Demonstração das primeiras amostras de dados

```
display(df_filtrado.head())
```

[8] ✓ 0.0s Python

ID	Data	Hora	Estacao	Codigo	Poluente	Valor	Unidade	Tipo
180189	2015-01-01	01:00	Carapicuiba	SP12	O3	29.0	ug/m3	automatica
180190	2015-01-01	02:00	Carapicuiba	SP12	O3	20.0	ug/m3	automatica
180191	2015-01-01	03:00	Carapicuiba	SP12	O3	15.0	ug/m3	automatica
180192	2015-01-01	04:00	Carapicuiba	SP12	O3	13.0	ug/m3	automatica
180193	2015-01-01	05:00	Carapicuiba	SP12	O3	13.0	ug/m3	automatica

Fonte: Os autores

Figure 4 - Demonstração das últimas amostras de dados.

```
display(df_filtrado.tail())
df_filtrado.to_csv("arquivo_filtrado.csv", index=False)
```

[9] Python

ID	Data	Hora	Estacao	Codigo	Poluente	Valor	Unidade	Tipo
9679519	2021-12-31	20:00	Carapicuiba	SP12	O3	16.0	ug/m3	automatica
9679520	2021-12-31	21:00	Carapicuiba	SP12	O3	9.0	ug/m3	automatica
9679521	2021-12-31	22:00	Carapicuiba	SP12	O3	5.0	ug/m3	automatica
9679522	2021-12-31	23:00	Carapicuiba	SP12	O3	2.0	ug/m3	automatica
9679523	2021-12-31	23:55	Carapicuiba	SP12	O3	1.0	ug/m3	automatica

Fonte: Os autores

3.1.2 Pré-processamento

Com o intuito de preservar a qualidade dos dados, realizou-se uma análise de dados nulos na tabela, constatando que não há indícios de dados nulos, conforme demonstrado na Figura 5.

Figura 5 - Validação e exclusão de dados nulos da amostra de dados.

```
display(df_filtrado.isnull().sum())
df_filtrado = df_filtrado.dropna()
```

[17] ✓ 0.0s Python

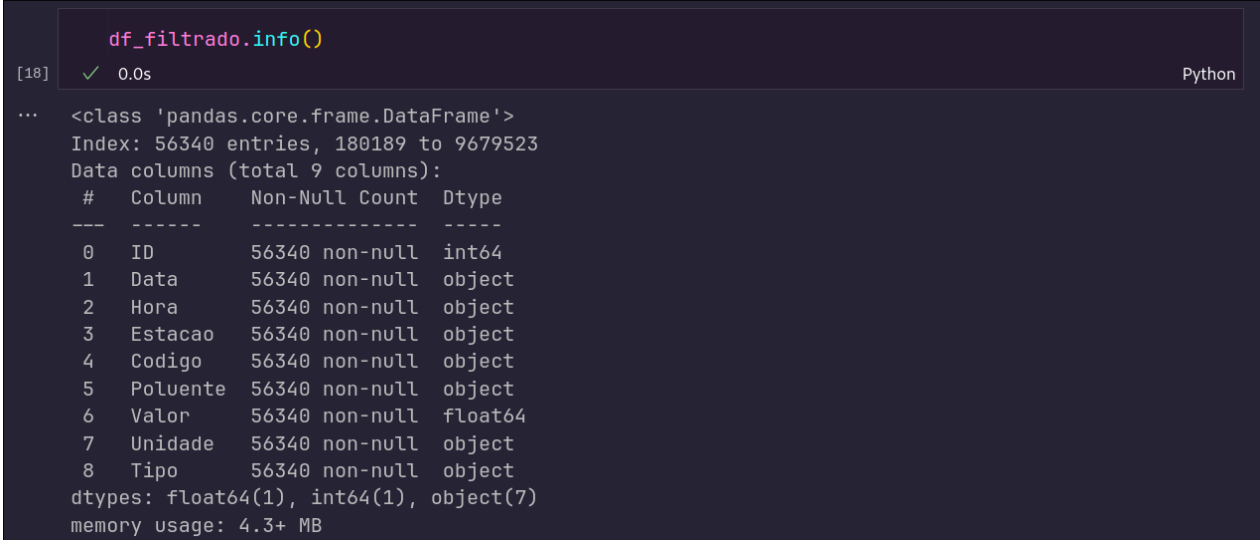
ID	0
Data	0
Hora	0
Estacao	0
Codigo	0
Poluente	0
Valor	0
Unidade	0
Tipo	0
dtype:	int64

Fonte: Os autores

3.1.3 Transformação

O projeto conta com uma planilha contendo nove colunas. Não foi necessária nenhuma transformação adicional, pois os tipos identificados pela própria ferramenta foram suficientes para o restante do projeto, como demonstrado na Figura 6.

Figura 6 - Demonstração do tipo de cada coluna



```

df_filtrado.info()
[18] ✓ 0.0s Python
...
<class 'pandas.core.frame.DataFrame'>
Index: 56340 entries, 180189 to 9679523
Data columns (total 9 columns):
#   Column      Non-Null Count  Dtype
---  -
0   ID           56340 non-null    int64
1   Data         56340 non-null    object
2   Hora         56340 non-null    object
3   Estacao      56340 non-null    object
4   Codigo       56340 non-null    object
5   Poluente     56340 non-null    object
6   Valor        56340 non-null    float64
7   Unidade      56340 non-null    object
8   Tipo         56340 non-null    object
dtypes: float64(1), int64(1), object(7)
memory usage: 4.3+ MB

```

Fonte: Os autores

3.1.4 Data Mining

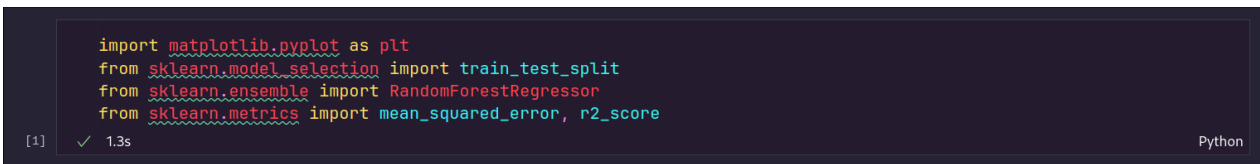
Na etapa de Data Mining, aplicaram-se métodos estatísticos e algoritmos de aprendizado de máquina para extrair padrões significativos dos dados. Neste projeto, utilizou-se a técnica de Random Forest para determinar as características mais relevantes.

3.1.5 Random Forest

A técnica de Random Forest foi utilizada para realizar previsões e classificações baseadas nos dados transformados. Esse método é conhecido por sua robustez e precisão em tarefas de classificação e regressão, sendo adequado para lidar com grandes volumes de dados complexos. No projeto, o modelo de Random Forest foi aplicado para prever os níveis de ozônio (O₃) em Carapicuíba.

Para iniciar o desenvolvimento, foi necessário importar as bibliotecas necessárias, como ilustrado na Figura 7.

Figura 7 - Importando as bibliotecas necessárias



```

import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, r2_score
[1] ✓ 1.3s Python

```

Fonte: Os autores

Seguiu-se o processo de identificar a fonte de dados, referenciando o dataset e extraindo as informações necessárias para o desenvolvimento do algoritmo, ver Figura 8.

Figura 8 - Preparando os dados

```
data = pd.read_csv("./arquivo_filtrado.csv")

data['DataHora'] = pd.to_datetime(data['Data'] + ' ' + data['Hora'])
data = data.drop(columns=['Data', 'Hora'])
```

[9] ✓ 0.1s Python

Fonte: Os autores

Para realizar a previsão dos níveis de ozônio, os dados foram divididos em variáveis independentes (X) e variável dependente (y). A variável dependente selecionada foi a coluna 'Valor', conforme demonstrado na Figura 9, juntamente as colunas irrelevantes para o modelo, como 'ID', 'Estacao', 'Codigo', 'Poluente', 'Valor', 'Unidade', e 'Tipo', foram removidas do dataset, mantendo apenas as colunas relevantes para o desenvolvimento do modelo.

Figura 9 - Definindo os eixos

```
y = data['Valor']
X = data.drop(columns=['ID', 'Estacao', 'Codigo', 'Poluente', 'Valor', 'Unidade', 'Tipo'])
```

[10] ✓ 0.0s Python

Fonte: Os autores

Para identificar as características mais relevantes, extraímos informações de data e hora usando o Pandas, criando colunas no DataFrame com os valores de hora, dia, mês e dia da semana, derivados da coluna original. Em seguida, removemos a coluna original.

Figura 10 - Extração de características de data e hora

```
X['hour'] = data['DataHora'].dt.hour
X['day'] = data['DataHora'].dt.day
X['month'] = data['DataHora'].dt.month
X['dayofweek'] = data['DataHora'].dt.dayofweek

X = X.drop(columns=['DataHora'])
```

[11] ✓ 0.0s Python

Fonte: Os autores

A figura 11 ilustra o processo de divisão dos dados em conjuntos de treinamento e teste usando a função *train_test_split* da biblioteca Scikit-learn, com 20% dos dados reservados para teste e uma semente aleatória definida para reprodutibilidade. Após a divisão, o modelo de Random Forest é treinado com o conjunto de treinamento utilizando 100 árvores (*n_estimators=100*).

Figura 11 - Divisão dos dados. conjuntos de treinamento, teste e treinamento modelo Random Forest

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

rf = RandomForestRegressor(n_estimators=100, random_state=42)
rf.fit(X_train, y_train)
```

[12] ✓ 8.0s Python

Fonte: Os autores

A Figura 12 apresenta o processo de avaliação do modelo Random Forest treinado. O modelo foi utilizado para fazer previsões no conjunto de teste, e a precisão das previsões foi avaliada usando o erro quadrático médio (Mean Squared Error - MSE) e o coeficiente de determinação (R-squared - R²). Os resultados obtidos foram um MSE de 409.0317 e um R² de 0.5601, que foram exibidos no console.

O MSE de 409.0317 indica a média dos quadrados dos erros, ou seja, a diferença média ao quadrado entre os valores previstos e os valores reais. Um MSE menor indica um modelo mais preciso. Já o R² de 0.5601 significa que aproximadamente 56% da variabilidade nos níveis de ozônio pode ser explicada pelas variáveis independentes do modelo. Este valor de R², embora moderado, sugere que o modelo possui uma capacidade razoável de explicar as variações nos níveis de ozônio, mas também indica que há espaço para melhorias.

Figura 12 - Avaliação do modelo Random Forest

```

y_pred = rf.predict(X_test)

mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

display(f'Mean Squared Error: {mse}')
display(f'R-squared: {r2}')

```

[13] ✓ 0.3s Python

```

... 'Mean Squared Error: 409.0317727195436'
... 'R-squared: 0.5601823269412356'

```

Fonte: Os autores

A Figura 13 ilustra o uso da biblioteca Pandas para criar um DataFrame a partir de um objeto, indexando-o pelas colunas `X_train.columns`. Em seguida, o DataFrame é ordenado pela coluna 'importance' em ordem decrescente. Utilizando a biblioteca Matplotlib, é gerado um gráfico de barras com o DataFrame 'feature_importances_'. O eixo x é rotulado como 'Features', e o eixo y como 'Importância'.

Figura 13 - Visualização da Importância das Características

```

feature_importances = pd.DataFrame(rf.feature_importances_,
                                  index = X_train.columns,
                                  columns=['importance']).sort_values('importance',
                                                                      ascending=False)

print(feature_importances)

plt.figure(figsize=(10, 6))
plt.bar(feature_importances.index, feature_importances['importance'])
plt.xlabel('Features')
plt.ylabel('Importância')

plt.show()

```

[14] ✓ 0.2s Python

```

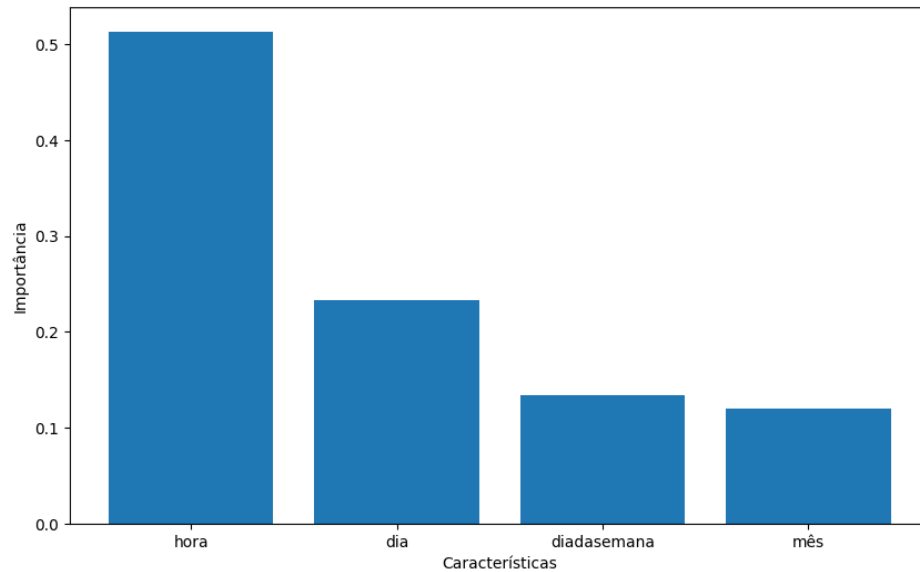
...
importance
hour      0.513274
day       0.233150
dayofweek 0.134110
month     0.119466

```

Fonte: Os autores

O resultado mostra um gráfico de barras com quatro barras representando diferentes características: dia, hora, dia da semana e mês, cada uma com seus respectivos escores de importância listados abaixo em formato de tabela. Como ilustrado na Figura 14, conforme é possível determinar a importância de cada característica e sua determinada importância, sendo a com maior pontuação a hora, que determinada a característica mais relevante para determinar o nível de ozônio.

Figura 14 – Visualização gráfica da Importância das Características



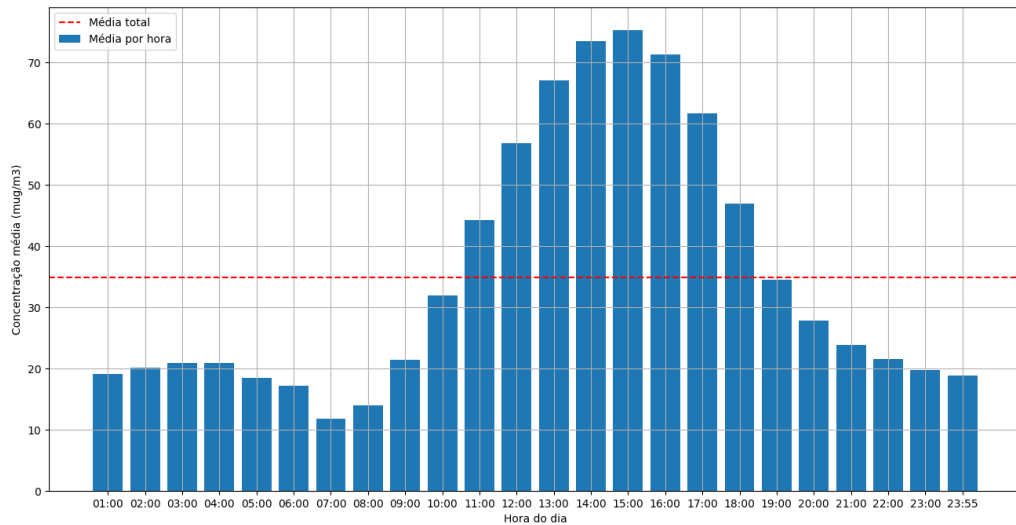
Fonte: Os autores

4 RESULTADOS E DISCUSSÃO

Dividem-se os resultados da pesquisa em duas subseções: análise descritiva dos dados e, em seguida, mostram-se os correlação entre os municípios de São Paulo.

4.1 Análise descritiva

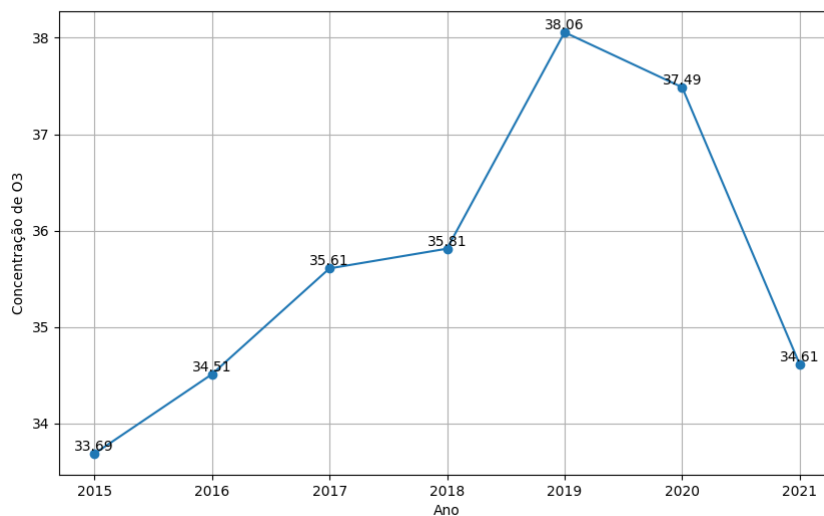
Os resultados da análise dos dados de ozônio (O_3) em Carapicuíba foram organizados e visualizados utilizando a biblioteca Matplotlib, conforme apresentado nas figuras demonstradas no decorrer deste tópico. Este estudo utilizou os dados fornecidos pelo Instituto de Meio Ambiente (IEMA) até o ano de 2021. As figuras 15 e 16 ilustram, respectivamente, a concentração média horária de ozônio ao longo de um dia típico e a evolução anual da concentração de ozônio no município.

Figura 15 - Concentração média por hora de O₃ em Carapicuíba

Fonte: Os autores

A concentração média horária de ozônio em Carapicuíba aumenta a partir das 10:00 horas, atinge o pico entre 14:00 e 16:00 horas, e diminui gradualmente após esse período. Esse padrão reflete a formação fotoquímica do ozônio, intensificada pela radiação solar e pela presença de precursores como óxidos de nitrogênio (NO_x) e compostos orgânicos voláteis (VOCs), reforçando os resultados referentes ao random forest.

A linha tracejada vermelha na Figura 15 representa a média diária de concentração de ozônio, mostrando que durante as horas de maior insolação, os níveis de ozônio superam essa média, destacando a influência da radiação solar na formação do poluente. Estudos como o de Zhou et al. (2020) corroboram a correlação entre a intensidade solar e os níveis de ozônio troposférico.

Figura 16 - Evolução da concentração do poluente O₃ em Carapicuíba

Fonte: Os autores

De 2015 a 2019, a concentração anual de ozônio em Carapicuíba aumentou devido ao crescimento urbano e industrial, que elevou as emissões de precursores como óxidos de nitrogênio (NOx) e compostos orgânicos voláteis (VOCs). A partir de 2020, observou-se uma leve queda, seguida por uma diminuição mais acentuada em 2021, atribuída principalmente às restrições da pandemia de COVID-19, que reduziram significativamente as atividades industriais e o tráfego veicular. Outros fatores, como mudanças nas políticas ambientais, variações meteorológicas e alterações nas atividades industriais e de transporte, também influenciaram essa variação. Estudos como o de Sicard et al. (2020) documentaram o impacto das restrições da pandemia na qualidade do ar, corroborando essa tendência de redução dos níveis de ozônio.

Os resultados foram interpretados à luz dos Índices de Qualidade do Ar (IQA_r) estabelecidos por órgãos reguladores. Conforme a tabela de classificação do IQA_r, em diversos momentos do dia, os níveis de ozônio em Carapicuíba os limites são considerados saudáveis, como será mostrado na figura 17 e 18, sendo avaliado durante o período com maiores picos de emissão de O₃, às 15 horas.

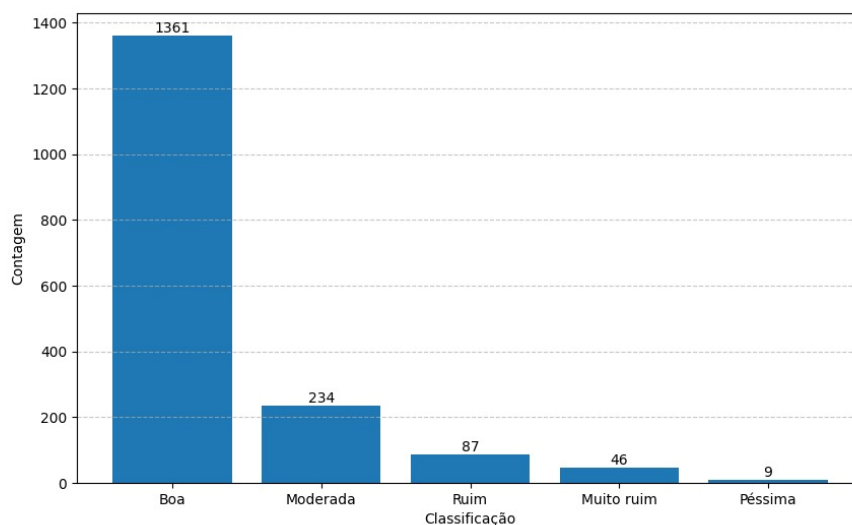
Tabela 1 - Estrutura do índice brasileiro de qualidade do ar e efeitos à saúde

Classe	O ₃ (µg/m ³) 8h
Boa	0 - 100
Moderada	100 - 130
Ruim	130 - 160
Muito ruim	160 - 200
Péssima	200 - 800

Fonte: Adaptado do IEMA (2021)

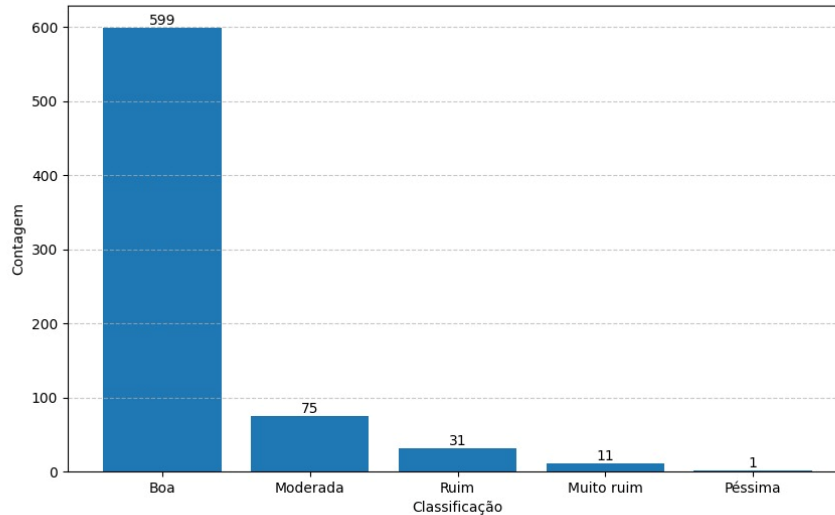
A Tabela 1 apresenta os níveis de classificação do IQA_r, que foram usados como referência para avaliar os dados de ozônio.

Figura 17 - Avaliação dos índices de poluição do ar às 15 horas, de acordo com a frequência de identificação dos níveis de qualidade, antes da pandemia



Fonte: Os autores

Figura 18 - Avaliação dos índices de poluição do ar às 15 horas, de acordo com a frequência de identificação dos níveis de qualidade, durante a pandemia



Fonte: Os autores

A predominância da categoria "Boa" indica que, na maioria das vezes, a qualidade do ar em Carapicuíba às 15 horas era satisfatória. Esta informação é crucial para a saúde pública, pois um ar de boa qualidade reduz os riscos de doenças respiratórias e cardiovasculares.

Na Figura 17, observa-se a presença de 234 registros na categoria "Moderada", sugerindo que, em algumas ocasiões, a qualidade do ar não era ideal, mas ainda aceitável para a maioria da população. No entanto, as 87 ocorrências de "Ruim" e as 46 de "Muito Ruim" apontam para períodos em que a poluição atingiu níveis preocupantes, exigindo atenção e medidas de mitigação. As 9 ocorrências de "Péssima" indicam momentos críticos de poluição extrema, que podem ter sérios impactos na saúde pública, especialmente para grupos vulneráveis, como crianças, idosos e pessoas com doenças crônicas.

Na Figura 18, os dados mostram uma melhora substancial em relação aos registros anteriores. As categorias "Moderada", "Ruim", "Muito Ruim" e "Péssima" apresentam uma redução significativa nos índices de O₃ troposférico em Carapicuíba. Especificamente, os novos valores são: "Boa" com 599 registros, "Moderada" com 75, "Ruim" com 31, "Muito Ruim" com 11, e "Péssima" com apenas 1 ocorrência. Esses números evidenciam a eficácia das políticas de redução de emissão de poluentes atmosféricos implementadas na região. A queda nos registros das categorias de qualidade de ar mais preocupantes demonstra um avanço crucial na melhoria da qualidade do ar e, conseqüentemente, na proteção da saúde pública.

5 CONSIDERAÇÕES FINAIS

A análise detalhada dos dados sugere que a formação de ozônio em Carapicuíba é fortemente influenciada por fatores ambientais e antropogênicos. A correlação observada entre a intensidade solar e a concentração de ozônio destaca a necessidade de estratégias de mitigação que considerem os padrões diurnos e sazonais da poluição. Medidas como a redução de

emissões de NOx e VOCs, especialmente em áreas urbanas e industriais, são essenciais para controlar os níveis de ozônio.

Além disso, a diminuição das concentrações de ozônio em 2020 e 2021 ressalta o impacto das intervenções humanas na qualidade do ar, reflexo das medidas de isolamento da pandemia. Este período oferece uma oportunidade única para avaliar a eficácia das políticas de restrição de emissões e considerar a implementação de práticas sustentáveis a longo prazo.

Em consideração final, os resultados deste estudo fornecem uma visão abrangente da dinâmica do ozônio em Carapicuíba e destacam a importância da monitorização contínua e da implementação de políticas ambientais rigorosas. A abordagem metodológica empregada, combinando a análise de dados com visualizações gráficas detalhadas, proporciona uma base sólida para futuras pesquisas e ações direcionadas à melhoria da qualidade do ar na região.

6 REFERÊNCIAS

AMIGO, E. J. **Data Mining, Machine Learning, Deep Learning**, 2021.

BREIMAN, L. **Machine Learning**. 2001.

BRAGA, A. et al. **Poluição atmosférica e saúde humana**. Revista USP, São Paulo, n.5, p.58-71, set./nov. 2001.

CETESB. **Qualidade do Ar: Poluentes**. 2 dez. de 2013. Disponível em: <https://cetesb.sp.gov.br/ar/poluentes/>. Acesso em: 18 maio 2024.

DOMÍNGUEZ-GARCÍA, R.; ARELLANO-VÁZQUEZ, M. **Evaluation of Machine Learning Models for Ozone Concentration Forecasting in the Metropolitan Valley of Mexico**. Disponível em: <<https://www.mdpi.com/2076-3417/14/4/1408>>. Acesso em: 31 maio 2024.

EUROPEAN ENVIRONMENT AGENCY. **Current state of the ozone layer**. Disponível em: <<https://www.eea.europa.eu/en/topics/in-depth/climate-change-mitigation-reducing-emissions/current-state-of-the-ozone-layer>>. Acesso em: 17 maio 2024.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. **From Data Mining to Knowledge Discovery in Databases**. AI Magazine, v. 17, n. 3, p. 37-54, 1996.

FEPAM. IQAr - **Índice de Qualidade do Ar**. 1 nov. 2018. Disponível em: <https://fepam.rs.gov.br/iqar-indice-de-qualidade-do-ar>. Acesso em: 18 maio 2024.

FILHO, Mario. **Precisão, Recall e F1 Score Em Machine Learning**. Mario Filho | Machine Learning, 2023. Disponível em: <https://mariofilho.com/precisao-recall-e-f1-score-em-machine-learning/>. Acesso em: 31 maio 2024.

GALVÃO, N. D. et al. **A violência no cotidiano dos profissionais de saúde: uma revisão integrativa**. Acta Paulista de Enfermagem, São Paulo, v. 23, n. 3, p. 403-409, 2010. Disponível em: <<https://www.scielo.br/j/ape/a/Lzj9vW6Fp4QVdXyNKhmtfvv/>>. Acesso em: 17 maio 2024.

GILLIS, A. S.; STADMAN, C. **Data mining**. Disponível em: <<https://www.techtarget.com/searchbusinessanalytics/definition/data-mining>>. Acesso em: 17 maio 2024.

INSTITUTO DE ENERGIA E MEIO AMBIENTE (IEMA). **Plataforma da Qualidade do Ar**. set. 2021. Disponível em: <https://energiaeambiente.org.br/produto/plataforma-da-qualidade-do-ar>. Acesso em: 18 maio 2024.

LAROSE, D. T.; LAROSE, C. D. **Discovering Knowledge in Data: An Introduction to Data Mining**. John Wiley and Sons, Inc, 2005.

LEITE, R. C. M. et al. **Utilização de regressão logística simples na verificação da qualidade do ar atmosférico de Uberlândia**. Engenharia Sanitária e Ambiental, v.16, n.1, jan./mar. 2011.

LIU, Y.; ZHOU, Y.; LU, J. **Exploring the relationship between air pollution and meteorological conditions in China under environmental governance**. Sci Rep 10, 14518 (2020). DOI: <https://doi.org/10.1038/s41598-020-71338-7>.

MIKUT, R.; REISCHL, M. **Data Mining Techniques and Applications**. WIREs Data Mining and Knowledge Discovery, v. 4, n. 1, p. 7-13, 2014. Disponível em: <https://www.wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.24>. Acesso em: 20 maio 2024.

PADUA, Mateus. **Machine Learning – Métricas de avaliação: Acurácia, Precisão e Recall, F1-score**. Medium, 2020. Disponível em: <https://medium.com/@mateuspdua/machine-learning-m%C3%A9tricas-de-avalia%C3%A7%C3%A3o-acur%C3%A1cia-precis%C3%A3o-e-recall-d44c72307959>. Acesso em: 31 maio 2024.

PAN, Q.; HARROU, F.; SUN, YING. **A comparison of machine learning methods for ozone pollution prediction**. Disponível em: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-023-00748-x>. Acesso em: 31 maio 2024.

RUSSO, P. R. **A qualidade do ar no município do Rio de Janeiro: análise espaço-temporal de partículas em suspensão na atmosfera**. Revista de Ciências Humanas, v.10, n.1, p.78-93, jan./jun. 2010.

SANKAR, K. P.; PABITRA, B. S. **Data Mining and Warehouse**. New Delhi: PHI Learning Pvt. Ltd, 2004.

SICARD, Pierre; DE MARCO, Alessandra; AGATHOKLEOUS, Evgenios; FENG, Zhaozhong; XU, Xiaobin; PAOLETTI, Elena; RODRIGUEZ, José; CALATAYUD, Vicent. **Amplified ozone pollution in cities during the COVID-19 lockdown**. 15 set. 2020. Disponível em: <https://www.sciencedirect.com/science/article/pii/S004896972033059X>. Acesso em: 18 maio 2024.

WANG, J.; et al. **Caracterization of tropospheric ozone pollution, Random Forest and Trend Prediction**. Disponível em: <https://enveurope.springeropen.com/articles/10.1186/s12302-024-00863-3>. Acesso em: 31 maio 2024.