

UNIVERSIDADE FEDERAL DE SÃO CARLOS - UFSCar
DEPARTAMENTO DE ENGENHARIA FÍSICA

RUDIGER YUUIRO MUKUNO SHIBUYA

Previsão de MP_{10} através de redes neurais: estudos
de caso no Estado de São Paulo

São Carlos
21 de março de 2022

Rudiger Yuujiro Mukuno Shibuya

**Previsão de MP_{10} através de redes neurais: estudos de caso no
Estado de São Paulo**

Trabalho apresentado ao Curso de Engenharia Física da Universidade Federal de São Carlos como parte dos requisitos para a obtenção do Grau de Engenheiro Físico.

Universidade Federal de São Carlos

Orientador: Prof. Dra. Regina Maura de Miranda
Coorientador: Prof. Dr. Fabio Aparecido Ferri

São Carlos
21 de março de 2022

Dedico este trabalho à minha mãe por ter criado a pessoa que sou hoje em dia, mesmo após tantas adversidades

Agradecimentos

Agradeço imensamente a minha família por me apoiar e ajudar a tornar meu sonho realidade.

Agradeço aos meus amigos que me ajudaram nos tempos difíceis.

Agradeço meus professores por me ensinarem o que sei hoje.

Agradeço demais minha orientadora por aceitar me orientar mesmo em cima da hora.

Agradeço de coração a Pâmela Rosa por estar sempre me apoiando.

*“Eu tentei 99 vezes e falhei, mas na centésima tentativa eu
consegui, nunca desista de seus objetivos mesmo que esses
pareçam impossíveis, a próxima tentativa pode ser a vitoriosa.”*

Albert Einstein

Resumo

A poluição do ar é um problema sempre presente e cada vez mais preocupante para a saúde e bem estar da população, em particular, o material particulado MP_{10} tem se tornado alvo de estudos pelos problemas que sua exposição em grande quantidade ou tempo prologado pode gerar na saúde pública. Para monitorar a fim de controlar e entender mais sobre suas origens e comportamento, foi desenvolvido uma rede neural de Multicamadas de Perceptrons (MLP) com duas camadas de 40 neurônios e 4 parâmetros de análise para se prever a concentração do poluente material particulado (MP). Esta configuração foi usada a fim de aumentar a eficiência e precisão dos dados. Os quatro parâmetros escolhidos estão relacionados à meteorologia, por terem forte influência na dispersão dos poluentes, sendo estes: pressão atmosférica, velocidade do vento, umidade relativa do ar e temperatura ambiente. A rede neural foi treinada a partir dos dados de MP entre 1 de janeiro de 2017 até 1 de janeiro de 2022, onde o programa utilizava como base os dados dos dias anteriores para prever os valores dos dias seguintes.

Foram analisadas as regiões do Parque Dom Pedro II, Guarulhos, Santos e Jaú com dados das estações de qualidade do ar coletados pela da Companhia Ambiental do Estado de São Paulo (CETESB), disponíveis através da plataforma QUALAR. A partir dos resultados gerados, foi analisado um erro percentual médio entre os valores fornecidos e os valores previstos pela rede neural de, respectivamente, 30.93%, 27.77%, 25.12% e 24.69% para o MP_{10} , demonstrando resultados condizentes e próximos do valor real, com uma maior precisão para zonas menos populosas, por haver menos interações humanas afetando a concentração de materiais particulados. Foi realizado o mesmo procedimento para $MP_{2.5}$ nas regiões do Parque Dom Pedro II e Guarulhos, obtendo resultados diferentes e não muito conclusivos, porém, igualmente precisos em relação ao erro percentual médio de 30.66% e 32.94% respectivamente, mostrando que o $MP_{2.5}$ não possui tanta relação com os dados meteorológicos quanto o MP_{10} . No geral, os resultados obtidos foram satisfatórios pela precisão e pela informação a respeito das redes neurais e dos ciclos atuação das MP, mostrando que MLPs são confiáveis para prever os valores de MPs no geral para valores diários.

Palavras-chave: MP_{10} , $MP_{2.5}$, Previsão, Poluente, Redes Neurais, Meio Ambiente.

Abstract

The air pollution is an ever-present and increasingly worrying health problem and welfare of the population. One of the air pollutant in particular, the Particulate Matter (PM), is a target of studies because of the problems that the exposure in large amounts or prolonged time can generate in public health.

In order to control and understand more about the origins and behavior of the MP, a neural network of multilayer of perceptrons (MLP) was developed with two layers of 40 neurons and 4 analysis parameters to test its efficiency and accuracy for predicting PM. This configuration was used in order to increase the efficiency and accuracy of the data. The four input parameters chosen for the MLP are related to meteorology, as they have a strong influence on the dispersion of pollutants. The input chosen are: atmospheric pressure, wind speed, relative humidity and temperature environment. The neural network was trained with the PM data between January 1, 2017 until January 1, 2022, where the program used data from previous days as a basis to predict the values of the following days.

The regions of Parque Dom Pedro II, Guarulhos, Santos and Jaú were analyzed with data from air quality stations collected by the environmental company of the state of São Paulo (CETESB), available through the QUALAR platform. From the generated results, an average of percentage error was analyzed between the values provided and the values predicted by the neural network of each state, respectively, of 30.93%, 27.77%, 25.12% and 24.69% for the MP_{10} , showing results that are consistent and close to the real value, with greater precision for less populated areas, as there are fewer human interactions affecting the concentration of particulate materials. The same procedure was performed for $MP_{2.5}$ in the regions of Parque Dom Pedro II and Guarulhos, obtaining different and not very conclusive results, however, equally accurate in relation to the average percentage error of 30.66% and 32.94% respectively, showing that the $MP_{2.5}$ is not so co-related with meteorologic data than the MP_{10} . In general, the results obtained were satisfactory for the precision and for the information about the neural networks and the MP actuation cycles, showing that MLPs are reliable to predict the MP values in general for daily values.

Keywords: PM_{10} , $PM_{2.5}$, Prediction, Pollutant, Neural Network, Ambient.

Lista de abreviaturas e siglas

MP	Material Particulado
MP ₁₀	Material Particulado com diâmetro inferior a 10 micrômetros.
MP _{2,5}	Material Particulado com diâmetro inferior a 2,5 micrômetros.
MLP	Multicamadas de Perceptrons
CETESB	Companhia Ambiental do Estado de São Paulo
QUALAR	Qualidade do AR

Sumário

1	INTRODUÇÃO	9
2	REFERÊNCIA TEÓRICA	11
2.1	MP ₁₀	11
2.2	Redes Neurais	12
3	METODOLOGIA	15
3.1	Cetesb	15
3.2	Aquisição e processamento de dados	15
3.3	Códigos utilizados	17
4	APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS	18
5	CONCLUSÕES E CONSIDERAÇÕES FINAIS	24
	REFERÊNCIAS	25
	ANEXO A – TABELAS	27
	ANEXO B – CÓDIGOS	28
B.1	Código Utilizado para unir e tratar os dados brutos baixados direto do site:	28
B.2	Código da MLP utilizada:	28

1 Introdução

A poluição do ar é um assunto recorrente e cada vez mais importante ser estudado. Desde a revolução industrial, a poluição vem se agravando pela quantidade de poluentes emitidos na atmosfera, provindo principalmente da queima de combustíveis fósseis, emissão das atividades industriais e ações humanas. Juntamente com o avanço das indústrias, tecnologia e da rápida urbanização, a poluição avança a um ponto onde passou a atingir a população em grandes proporções, relacionando poluição atmosférica com problemas de saúde como Asma Bronquica, Doença pulmonar obstrutiva crônica (DPOC), Infecção respiratória aguda, câncer de pulmão, (ARBEX, 2012) chegando até mesmo a ter relação com problemas cognitivos e comportamentais (JOSEPH, 2016). Além de afetar condições meteorológicas e agravar as mudanças climáticas (PATERLINI, 2007).

Porém, o estudo da poluição do ar é bastante complexo, já que a poluição atmosférica é composta por inúmeros tipos gases e partículas, desde os menos nocivos como gás carbônico, até poluentes mais problemáticos como gás metano e o material particulado (MP)

O material particulado é tema de inúmeros estudos e pesquisas, onde foi demonstrado sua concentração na atmosfera está diretamente relacionada à quantidade de pessoas internadas localmente (LUI, 2016). Pelo material particulado estar diretamente relacionado as condições meteorológicas, além da influência humana, é possível prever sua quantidade presente na atmosfera, através de análise de padrões e flutuações de diversos parâmetros. Uma boa forma de realizar essas previsões, está na utilização de redes neurais pela sua capacidade de encontrar padrões e inferências de diversos dados puros simultaneamente de forma coesa e confiável.

As redes neurais artificiais simulam as redes neurais biológicas do cérebro. Através de treinamento e persistência, as redes neurais artificiais são capazes de reconhecer padrões escondidos e correlacionar dados, os agrupando e classificando. Seu "aprendizado" ocorre extremamente rápido, com milhares de testes, treinos e validações por segundo. Onde, quanto mais dados e informações correlacionadas possuir, melhor será sua performance.

Mesmo com muitas pesquisas sobre materiais particulados e sua influência sobre a saúde, além de estudos de previsões através de redes neurais (MONTEIRO, 2007), pouco se sabe sobre efetividade dos dados de MP e comparação de resultados das redes neurais, além de melhorias e entendimento da macro-escala de emissão para o desenvolvimento de um programa mais preciso. Com isso em mente, o enfoque deste estudo foi o material particulado inalável MP_{10} para uma melhor compreensão do que está afetando a previsão das redes desenvolvidas e utilizadas atualmente, além disso, por ser um dos mais prejudiciais a saúde e por ser fortemente influenciado pelas condições meteorológicas, o que facilitaria no estudo de possíveis melhorias através da análise de diferentes regiões com diferentes

climas e condições.

2 Referencia teórica

Antes começar a realizar mudanças no clima e no meio ambiente para um mundo melhor, é preciso prever e simular o que pode acontecer para, então, poder analisar e atuar com maior efetividade. Com este intuito, um sistema de análise, como as redes neurais, seria uma boa abordagem para prever a quantidade de MP da atmosfera utilizando como base os parâmetros meteorológicos.

2.1 MP₁₀

O material particulado é um poluente classificado por seu tamanho, considerando o diâmetro aerodinâmico médio das partículas. As partículas inaláveis (MP₁₀) são partículas que possuem o diâmetro aerodinâmico médio de até 10 μm e as partículas inaláveis finas (MP_{2.5}) possuem o diâmetro aerodinâmico inferior a 2,5 μm (SOLCI, 2009).

As principais fontes de emissão de MP são, a ressuspensão de partículas no solo, emissões provenientes do sal marinho, queima de combustíveis fósseis, geração de energia como as usinas de carvão, processos industriais, usos de veículos, queimadas e construções. A composição química do material particulado é bastante variada, podendo conter desde materiais inofensivos como sódio e água até metais, nitratos e plásticos que podem ser extremamente tóxicos (LIMA, 2008).

A quantidade de MP presente na atmosfera tem variações sazonais, sendo que no verão é reduzida e no inverno a quantidade aumenta, devido às diferentes condições meteorológicas; além disso, existe uma variação diária também por conta da influência de parâmetros meteorológicos e das concentrações de precursores. Pode ocorrer dispersão e diluição ocasionadas por reações fotoquímicas (MAYER, 1999).

De acordo com Ramanathan (2007), esse poluente é capaz de afetar diretamente o clima do planeta através do espalhamento e absorção da radiação solar, alterando o balanço de radiação terra-atmosfera, e indiretamente, atuando como núcleos de condensação de nuvens. Um exemplo de atuação da MP no clima foi presenciado na cidade de São Paulo no dia 19 de agosto de 2019, quando o céu escureceu durante o dia por conta das cinzas emitidas pelas queimadas na Amazônia, atingindo medidas de concentração de MP_{2.5} acima do normal no dia ocorrido.(LANDULFO, 2021)

Além disso, sua concentração diária se correlaciona diretamente com a quantidade de pessoas internadas e mortas por doenças respiratórias e cardiovasculares diárias (THOL-GATE, 2002).

A inalação direta e contínua de MP₁₀ pode causar problemas graves de saúde por transpor as barreiras do sistema respiratório e se depositar nos alvéolos pulmonares, onde os mecanismos de expulsão destes poluentes não são eficientes, podendo causar danos que

só são perceptíveis em longo prazo de exposição (MOURA, 2016). Além disso, o particulado inalável pode se acumular nas vias respiratórias superiores, agravando problemas respiratórios como o da asma, bronquite e sinusite (SOLCI, 2009).

O perigo causado pela inalação de partículas depende não só da sua forma e tamanho, como também da composição química, do lugar no qual elas foram depositadas no sistema respiratório e da faixa etária da pessoa afetada, pois, pessoas idosas e crianças são mais suscetíveis a problemas causados pelos níveis elevados de poluentes no ar de forma geral (DOCKERY, 2012).

No estado de São Paulo as principais fontes de poluição atmosférica são as emissões veiculares, as industriais e a queima de biomassa com um maior agravante veicular e industrial no município de São Paulo, reduzindo sua influencia conforme é afastado da região metropolitana (ARAUJO, 2019). Na região costeira, a principal fonte vem do sal marinho e navios.(GIODA, 2017)

A previsão de material particulado, em especial o MP_{10} , é possível pela sua relação direta com dados meteorológicos, principalmente pela relação de fluxo de MP com a massa de ar fria, variando com grande influencia sob a pressão atmosférica regional (RIZZO, 2021). Com essa relação direta, é possível produzir uma rede neural para prever a quantidade de MP natural. Porém, utilizando apenas as condições meteorológicas de previsão não é considerado ações humanas, portanto, as previsões podem não ser tão precisas nas regiões com muita interação humana.

2.2 Redes Neurais

As redes neurais, também conhecidas como redes neurais artificiais, são sistemas de computação com nós interconectados que funcionam como os neurônios do cérebro humano, imitando a maneira como os neurônios biológicos enviam sinais uns para os outros (SIQUEIRA, 2018).

As redes neurais são compostas por camadas de nós, contendo uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída como mostra na figura 1. Cada nó ou neurônio artificial, também conhecido como Perceptron, conecta-se a outro por canais de comunicação com um peso e um limite associados. Cada nó faz apenas operações sobre seus dados locais, recebidas pelas comunicações a partir da soma ponderada dos sinais. Quando a soma ponderada, que chega dos outros nós, está acima do valor do limite especificado, esse nó será ativado enviando dados para a próxima camada da rede. Caso contrário, nenhum dado será transmitido para a próxima camada (ANDRE, 2020).

Existem diversos tipos de redes neurais, cada um com suas vantagens e desvantagens de acordo com sua utilização, elas são as Redes neurais *feedforward*, Redes neurais convolucionais, Redes neurais recorrente e Redes neurais autoencoders.

Nas Redes neurais *feedforward*, ou perceptrons multicamadas (MLPs) a informação é entregue de maneira antecipada de uma camada à seguinte seguindo sempre em frente.

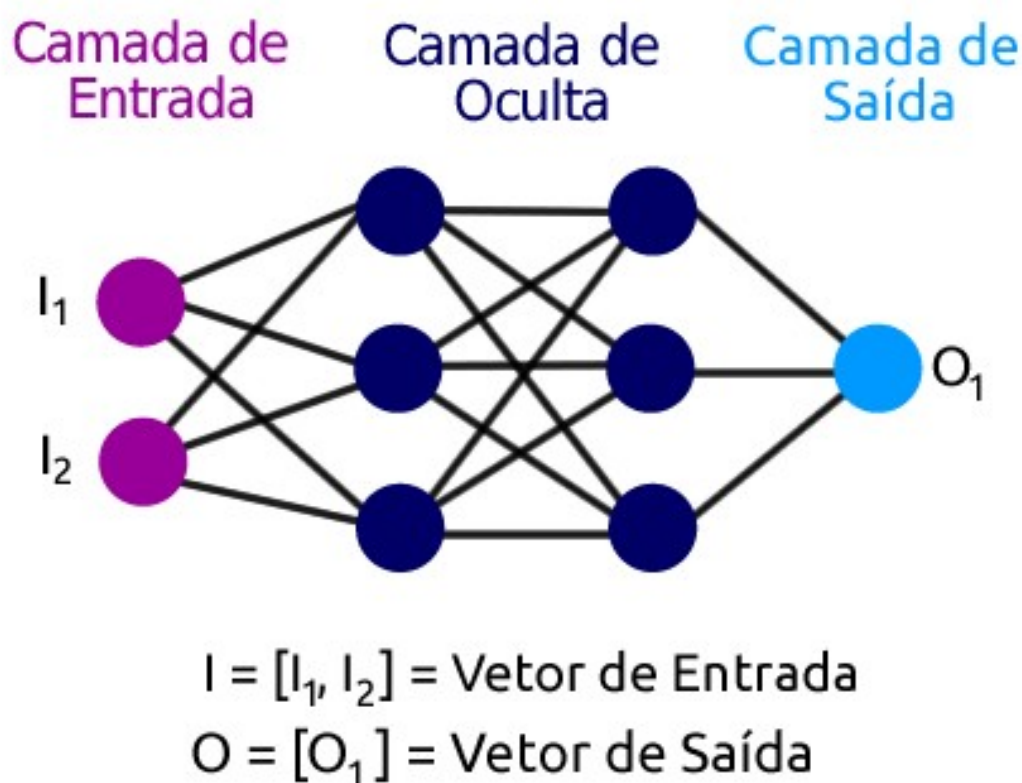


Figura 1 – Ilustração de funcionamento da Rede Neural

Não há loops de feedback (IBM, 2020).

As Redes neurais convolucionais (RNCs) contêm cinco tipos de camadas: de entradas, de convolução, de agrupamento, as completamente conectadas e as de saída. Cada camada tem um propósito específico e sua utilização se encontra mais na área de classificação de imagens, detecção de objetos e processamento de linguagem natural (IBM, 2020).

As Redes neurais recorrente (RNRs) usam informações sequenciais, diferentemente das redes neurais tradicionais, as entradas de uma rede neural recorrente não são independentes umas das outras e os resultados para cada elemento dependem da computação dos elementos precedentes (GOMEZ,).

As Redes neurais autoencoders procuram modelar as entradas por si só e, portanto, o método é considerado não supervisionado. Conforme as camadas são adicionadas, outras abstrações são formuladas em camadas mais altas. Essas abstrações podem, então, ser usadas por classificadores lineares ou não lineares (GOMEZ,).

Usando algoritmos, as redes neurais podem reconhecer padrões escondidos e correlações em dados brutos, agrupá-los e classificá-los. Com isso podem aprender e modelar relações entre entradas e saídas de dados que são não-lineares e complexos; realizar generalizações e inferências; revelar relacionamentos, padrões e previsões ocultas e modelar dados altamente voláteis (como dados de séries temporais financeiras) e variâncias neces-

sárias para prever eventos raros (como detecção de fraudes). Além disso tudo, ela já está presente em nosso dia-a-dia pois uma das redes neurais mais conhecidas é o algoritmo de procura do Google (LUI, 2016).

Neste trabalho, foi programado, no ambiente MatLab, uma Rede neural do tipo *feedforward* de Multicamadas de perceptron (MLP) composta por várias camadas de neurônios artificiais para fazer a previsão de MP_{10} . A eficiência da conexão entre duas camadas de neurônios é modelada por uma matriz W de peso e tamanho $n \times m$, onde m é o número de neurônios na camada anterior e n o número de neurônios da próxima camada. A ativação dos neurônios de cada camada é uma matriz dada pela equação:

$$A = f(W.X + B) \quad (1)$$

Onde f é uma função limiar, X é a matriz coluna com a saída dos m neurônios da camada anterior e B é uma matriz coluna com n entradas. Neste trabalho a função do limiar f utilizado foi a função sigmóide.

$$f(x) = 1/(1 + e^{-x}) \quad (2)$$

A camada de entrada da rede contém os valores médios da velocidade do vento, umidade relativa, temperatura, pressão e concentração de MP_{10} no dia e a camada de saída contém um único neurônio que representa a concentração de MP_{10} no dia seguinte. Foram adotadas duas camadas escondidas com 40 neurônios cada e o treinamento foi feito usando o método Backpropagation (PAULINO, 2020).

3 Metodologia

Atualmente, a utilização de MLPs para estudos de MP é bem reconhecida por seus valores precisos de previsão de dados diários, com erro médio de valores absolutos chegando a 9,2158Lui (2016). Para este trabalho, foi implementado o valor da pressão atmosférica às informações de entradas normalmente utilizadas de velocidade do vento, temperatura e umidade relativa do ar e pressão no intervalo de 5 anos para análise e teste da rede neural, o intervalo de 5 anos foi escolhida pela maior continuidade de dados disponíveis em um intervalo de tempo. Após serem testados diversos valores de camadas, nós e relações de treino, teste e validação para melhorar a performance, foram encontrados o valores de melhor desempenho quando utilizado uma rede de duas camadas de 40 neurônios com as relações de treino, teste e validação de 70%, 15% e 15% respectivamente.

3.1 Cetesb

”A Cetesb – Companhia Ambiental do Estado de São Paulo é a agência do Governo do Estado responsável pelo controle, fiscalização, monitoramento e licenciamento de atividades geradoras de poluição, com a preocupação fundamental de preservar e recuperar a qualidade das águas, do ar e do solo.”(CETESB, <https://cetesb.sp.gov.br/>, site)

A Cetesb possui diversas estações de qualidade do ar espalhadas pelo estado de São Paulo. Cada estação mede parâmetros meteorológicos e poluentes, com estações que medem de forma automática ou manual. As medidas das estações automáticas são armazenadas de hora em hora em nuvem e podem ser acessadas através do site QUALAR para todos os públicos. A figura 2 mostra a localização das estações de qualidade do ar distribuídas pelo estado de São Paulo, onde é perceptível uma maior concentração de estações próxima da região metropolitana.

Os dados foram adquiridos através do site da CETESB (<https://cetesb.sp.gov.br/>), que fornece dados de velocidade do vento, temperatura, umidade relativa, pressão atmosférica, MP_{10} , $MP_{2.5}$, nível de ozônio ou gás carbônico no ar, etc. Que são armazenados separadamente no site QUALAR (<https://qualar.cetesb.sp.gov.br/>).

3.2 Aquisição e processamento de dados

Existem dois tipos de redes de monitoramento obtidos pela Cetesb, os convencionais, que coletam os dados manualmente, e os automáticos, que coletam dados em intervalos de minutos, sendo fornecida a média horária. Para maior precisão e estabilidade de dados, foram escolhidos apenas os automáticos com dados no intervalo entre 01 de janeiro de 2017 e 1 de janeiro de 2022 por 5 anos ser um bom tempo para treinamento da rede neural e pelo período possuir dados mais recentes e completos.



Figura 2 – Localização das estações (CETESB, <https://cetesb.sp.gov.br/>)

Devido às diferenças entre os poluentes monitorados nas estações de qualidade do ar, onde alguns possuem mais informações disponíveis que outros, foram escolhidos 4 estações de regiões diferentes com maior variedade de medições e diferenças de clima com a finalidade de comparar as regiões e as diferenças de desempenho da rede neural. As regiões escolhidas foram, o município Jaú, por ser uma cidade mais afastada da grande metrópole. A região de Pimentas no município de Guarulhos, uma cidade periférica da grande São Paulo. O bairro Parque Dom Pedro II da grande São Paulo, localizado no centro da capital paulista e, por ultimo, o centro do município de Santos, uma região costeira para o oceano Atlântico.

Como o site QUALAR só possui dados por hora e só permite baixar arquivos contendo valores de um ano inteiro por vez. Foram baixados 5 arquivos por estação, onde foi, posteriormente, juntados em um único arquivo e extraído a média de cada dia das informações de velocidade, temperatura e umidade relativa do ar, além da pressão atmosférica, MP_{10} e $MP_{2.5}$. Depois de unir os arquivos, os dados foram tratados e os valores inexistentes foram preenchidos com a média geral de cada coluna, tudo isso através do uso da linguagem de programação Python com a biblioteca Pandas.

Com os dados prontos, foi utilizado o processamento da rede neural MLP na linguagem de programação MATLAB, utilizando como entrada os valores de pressão, velocidade, temperatura e umidade relativa do ar, além os valores de MP_{10} para treinamento e comparação com a previsão para todos as estações, repetindo os mesmos procedimentos para

o $MP_{2.5}$ das regiões de Guarulhos e Parque Dom Pedro II por serem os únicos a possuir as medições de $MP_{2.5}$.

3.3 Códigos utilizados

Para este estudo, foram utilizados dois códigos, a da rede neural MLP com 2 camadas de 40 nós B.2 que possui como divisão de parâmetros de treino a 70%, teste e valores absolutos de 15% cada, feita na linguagem de programação MatLab. E o código de junção e tratamento de dados na linguagem python B.1 dos arquivos adquiridos através do site QUALAR. Os códigos utilizados estão em apêndice.

4 Apresentação e discussão dos resultados

Com base nas tabelas de média geral 1 e 2, é possível perceber uma boa concordância entre os valores das medições e das previsões com as médias idênticas ou muito próximas de medição e previsão. Foi estudado, também, as médias anuais mostradas nas tabelas no anexo A, mostrando uma grande relação, em cada ano, com variações maiores de valores absolutos e desvio padrão nas cidades grandes e menores variações nas cidades pequenas.

A partir da MLP utilizada, foram gerados dois gráficos por região, um com os valores comparativos de MP_{10} medidos e MP_{10} Previstos e o outro gráfico com o erro numérico absoluto da comparação de cada dia, além de gerar os dados do erro médio e do erro médio percentual da previsão em relação à medida efetuada. Para as regiões de Guarulhos e Parque Dom Pedro II, foram gerados os mesmos gráficos para $MP_{2,5}$ a fim de comparar os resultados.

Os resultados obtidos foram representados pelos gráficos mostrados nas figuras fig.3, fig. 4, fig.5, fig.6, fig.7 e fig.8, que representam os valores respectivos da MP_{10} e $MP_{2,5}$ do Parque dom pedro II, MP_{10} e $MP_{2,5}$ da região de Guarulhos, MP_{10} da região de Santos e MP_{10} da região de Jaú. Onde o gráfico azul e verde representa respectivamente os valores medidos pela CETESB e previstos pela rede neural e o gráfico vermelho representa o valor do erro absoluto entre o valor medido e o valor previsto.

Com base nos gráficos e nos valores médios de cada cidade é possível preceber uma

Tabela 1 – Médias gerais e Desvio Padrão de cada estação retirados a partir dos dados diários e respectiva quantidade de medidas de MP_{10} em $\mu g/m^3$

Estações de MP_{10}	Média geral Medição ($\mu g/m^3$)	Média geral Previsão ($\mu g/m^3$)	Quantidade de Medidas
Pq. d. pedro II	28.12 ± 15.32	26.78 ± 17.92	1829
Guarulhos	31.15 ± 16.72	30.48 ± 19.23	1830
Santos	18.80 ± 07.02	19.07 ± 06.93	1840
Jaú	25.02 ± 15.07	24.87 ± 15.48	1840

Tabela 2 – Médias gerais e Desvio Padrão de cada estação retirados a partir dos dados diários e respectiva quantidade de medidas de $MP_{2,5}$ em $\mu g/m^3$

Estações de $MP_{2,5}$	Média geral Medição ($\mu g/m^3$)	Média geral Previsão ($\mu g/m^3$)	Quantidade de Medidas
Pq. dom pedro II	17.02 ± 8.82	16.43 ± 10.72	1830
Guarulhos	18.60 ± 10.62	17.19 ± 12.69	1830

precisão muito grande de 69.07%, 72.23%, 74.88% e 75.31% para o MP_{10} das regiões do Parque Dom Pedro II, Guarulhos, Santos e Jaú respectivamente. Alguns gráficos possuem uma linha reta que seriam dias que a CETESB não obteve os dados e eles foram tratados de forma a minimizar o erro do programa ao atribuir a média geral das medições. Porém, mesmo assim, esses dados faltando pode ter influenciado negativamente na previsão dos dados por gerar dados falsos para a rede neural treinar.

A partir dos gráficos das regiões, é possível analisar, menos na cidade de Santos, os ciclos de alta e baixa da concentração de MP com o passar dos dias, o que condiz com estudos de ciclos de poluentes ao longo do tempo, variando diariamente e anualmente como comentado no referencial teórico. As concentrações são mais altas nos meses de inverno devido às condições meteorológicas, inverno seco, menos entradas de frente e atmosfera mais estável. A ausência de ciclos no município de Santos deve-se a localização da cidade, por ser uma cidade costeira, com variações sazonais menos intensas, que não influenciam a concentrações dos poluentes como nas outras regiões.

Um ponto a ser levantado está na possibilidade do erro percentual das regiões serem diretamente proporcionais a sua quantidade de habitantes, onde Jaú com 24,69%, Santos com 25,12%, Guarulhos com 27,77% e São Paulo com 30,93% possuem suas respectivas populações de 151.881, 433.991, 1.404.694 e 12.396.372 (IBGE, 2020). Onde os locais com maiores populações possuem mais interferência humana com uma maior quantidade de veículos, construções, indústrias e pessoas que influenciariam na concentração de MP_{10} da região. Pela interação humana ser um parâmetro fora do escopo de análise do processamento da rede neural, a quantidade de MP_{10} prevista teria um erro maior conforme as influências humanas aumentassem. Além disso, a estação de qualidade do ar da cidade de Jaú não fornecia os valores de pressão atmosférica, diminuindo a precisão dos resultados, conforme analisado com as outras regiões, a ausência da pressão reduzia a precisão e, mesmo assim, o erro percentual médio de Jaú foi 6,24% menor do que o da região do Parque Dom Pedro II.

Não houve mudança significativa entre MP_{10} e $MP_{2.5}$ na região do Parque Dom Pedro II com erros percentuais semelhantes mesmo com a diferença de erros absolutos, demonstrando que a quantidade de MP_{10} detectado foi maior que a quantidade de $MP_{2.5}$ de modo geral. Já para a região de Guarulhos, o percentual muda bastante podendo ser relacionado aos valores absolutos mais baixos detectados para $MP_{2.5}$ e, por isso, gerando um erro maior para pequenas variações causadas pelas atividades humanas, outro possível motivo seria pelo $MP_{2.5}$ não ser tão dependente das condições meteorológicas, tornando a análise com maior erro. O erro absoluto é menor que o do MP_{10} pelas menores concentrações, porém, quando comparado o erro percentual é fácil notar que não há muita diferença, provando que a rede neural possui uma precisão estável.

De modo geral, os valores obtidos foram precisos, semelhantes aos estudos relacionados. Usando como base o estudo Lui (2016) que utilizou apenas os parâmetros de umidade,

temperatura e velocidade do vento com uma rede neural MLP de 15 neurônios, atingiu um erro absoluto médio de 7,13 na cidade de São Carlos. Quando comparados aos resultados obtidos, 6.82, 7.25, 3.89 e 5.27 pode-se dizer que os valores obtidos foram melhores. Além disso, São Carlos, a cidade em que o estudo de Lui (2016) foi baseado, tem uma população mais próxima de Jaú e Santos do que São Paulo e Guarulhos, 254.484 pessoas, com um valor de erro mais próximo de São Paulo e Guarulhos. Porém, para uma análise comparativa melhor, seria necessária a utilização de dados iguais.

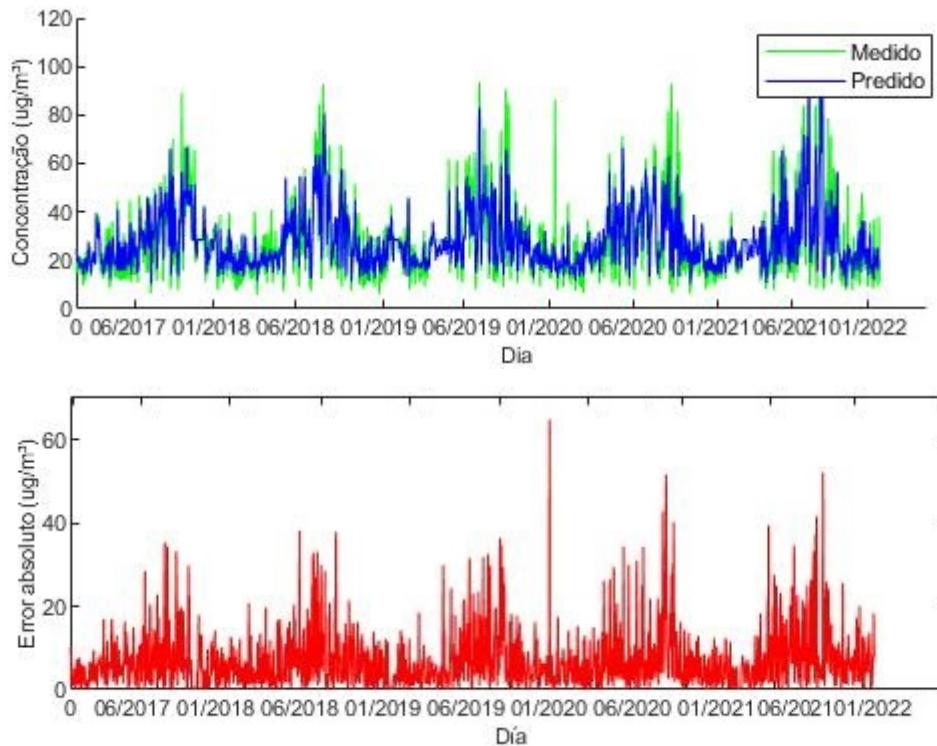


Figura 3 – Gráficos de comparação entre dados de MP_{10} medidos e previstos pelo programa e Erro absoluto entre os dois para a região Parque Dom Pedro II. Com valores médios de erro absoluto de 6.8192 e erro percentual de 30.93%

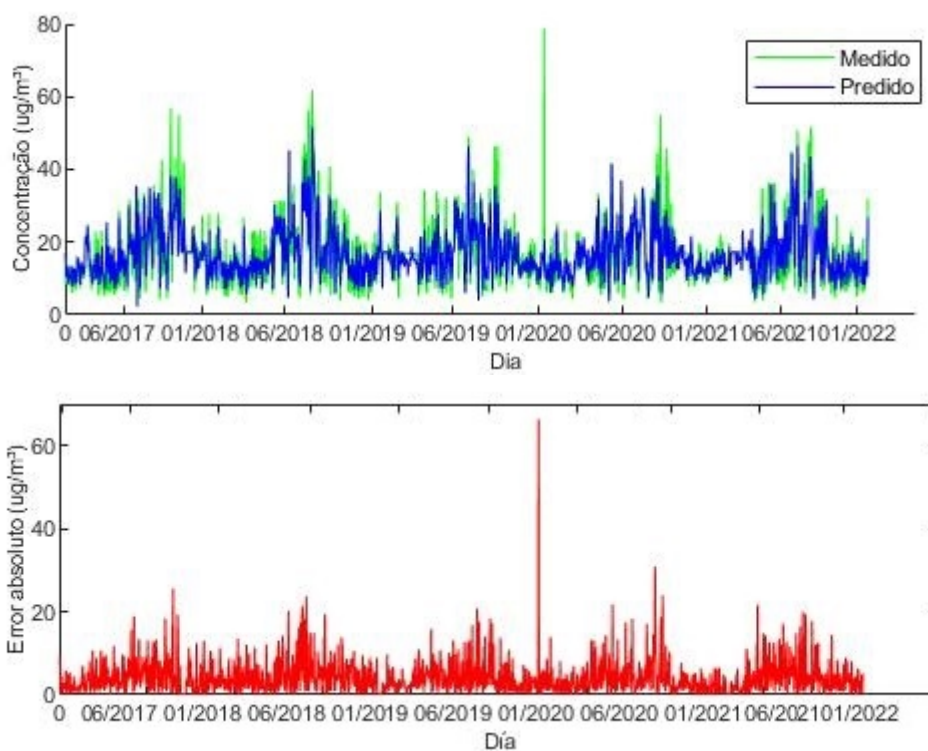


Figura 4 – Gráficos de comparação entre dados de $MP_{2.5}$ medidos e previstos pelo programa e Erro absoluto entre os dois para a região Parque Dom Pedro II. Com valores médios de erro absoluto de 4.1097 e erro percentual de 30.66%

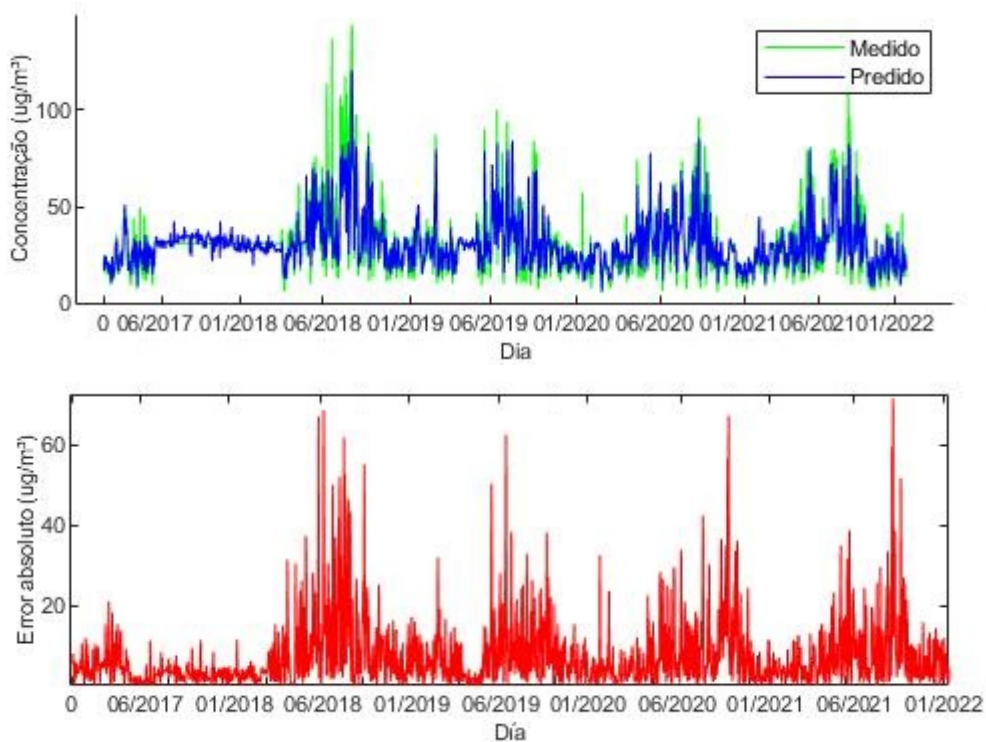


Figura 5 – Gráficos de comparação entre dados de MP_{10} medidos e previstos pelo programa e Erro absoluto entre os dois para a região de Guarulhos. Com valores médios de erro absoluto de 7.2535 e erro percentual de 27.77%

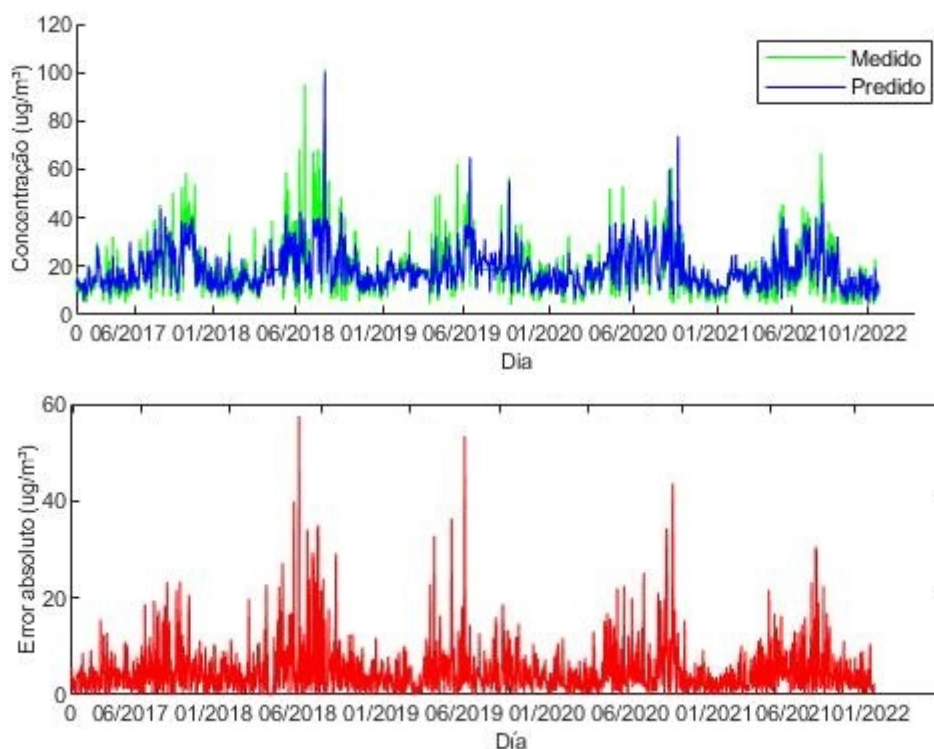


Figura 6 – Gráficos de comparação entre dados de $MP_{2.5}$ medidos e previstos pelo programa e Erro absoluto entre os dois para a região de Guarulhos. Com valores médios de erro absoluto de 4.9438 e erro percentual de 32.94%

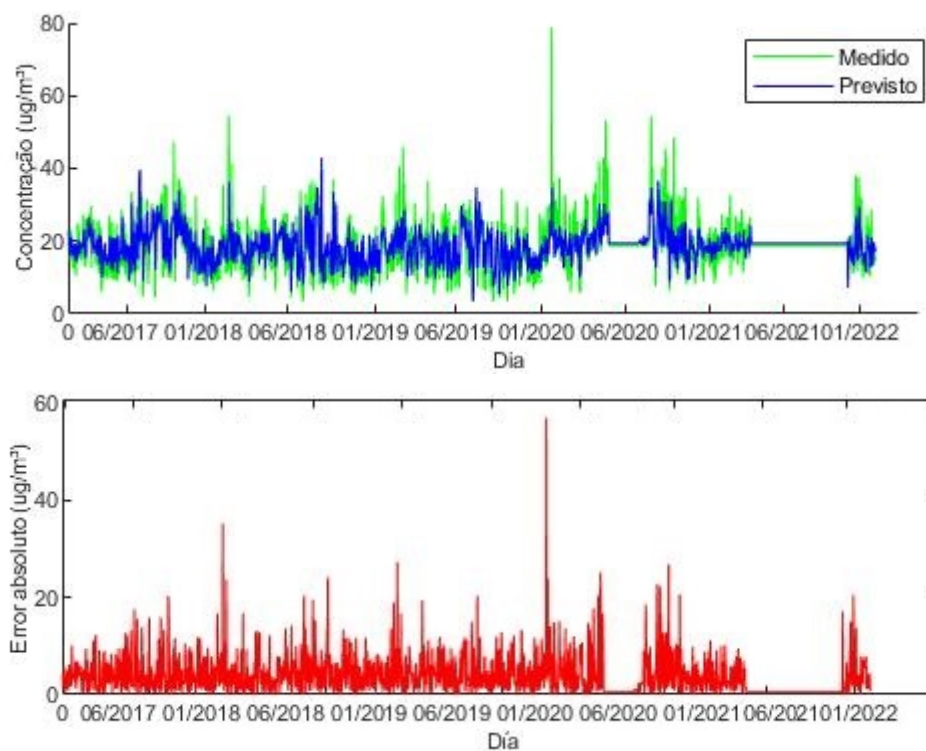


Figura 7 – Gráficos de comparação entre dados de MP_{10} medidos e previstos pelo programa e Erro absoluto entre os dois para a região de Santos. Com valores médios de erro absoluto de 3.8918 e erro percentual de 25.12%

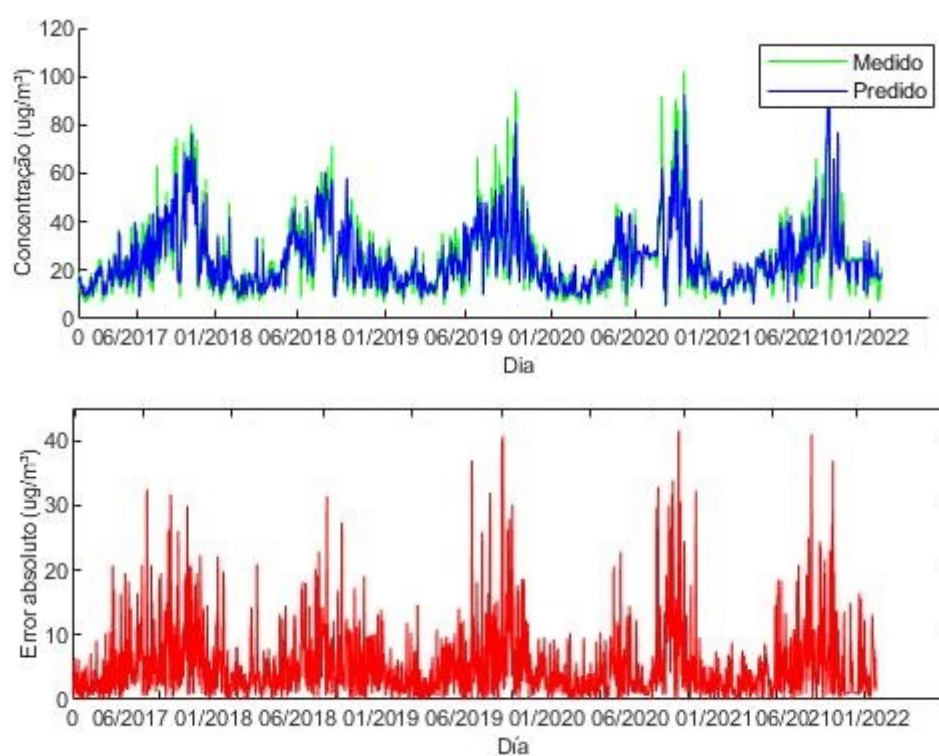


Figura 8 – Gráficos de comparação entre dados de MP_{10} medidos e previstos pelo programa e Erro absoluto entre os dois para a região de Jaú. Com valores médios de erro absoluto de 5.2736 e erro percentual de 24.69%

5 Conclusões e considerações finais

Foi observado uma relação direta entre a quantidade populacional da região com seu erro percentual, onde Jaú com população de 151.881 teve um erro percentual de 24,69%, Santos com 433.991 e erro de 25,12%, Guarulhos com 1.404.694 e erro 27,77% e São Paulo com 12.396.372 e erro 30,93%. O erro aumentado pode ter relação com as maiores concentrações de veículos e partículas ressuspendidas diariamente pela própria população. Além disso, o método de MLP para previsão de MP_{10} demonstrou ser muito confiável, com um erro percentual médio extremamente baixo e com dados coerentes por toda a análise, o $MP_{2,5}$ demonstrou possuir uma menor relação com as condições meteorológicas do que o MP_{10} , mas não deixou de ser confiável. Esse MLP pode ser utilizado para alertar a população de altas de MP, preparação para uma eventual alta na entrada de hospitais, além de possibilitar análises de ações que podem ou não afetar a quantidade de MP no ar com uma análise comparatória de previsão e medição em curto tempo.

Levando em conta a discrepância perceptível de 20% entre os erros absolutos nos dados entre cidades com grande e baixa população, seria de grande interesse, para melhorar a precisão da rede neural, se tivesse um parâmetro previsível e confiável de atuações humanas como volume de carros, previsão de fluxo de pessoas ou quantidade de construções em andamento no dia. Além disso, outros processos que ajudariam a melhorar a acurácia das redes neurais seria uma melhor forma de tratar os dados ausentes sem gerar problemas para a rede neural. Mesmo com resultados muito bons, é possível melhorar a precisão caso sejam testados outros tipos de redes neurais ou com mais parâmetros diferentes de entrada. Porém, a MLP por si só já trás um resultado preciso e confiável para valores diários.

Para um estudo futuro, a inclusão dos dias e das estações à entrada da MLP para a previsão de um período maior de análise afim de treinar os ciclos sazonais, ajudando a máquina a prever tais ciclos. Um possível estudo futuro seria a utilização de redes neurais para prever o fluxo de carros ou de pessoas utilizando variáveis de tempo como dias úteis, feriados, eventos anuais e condições meteorológicas. Pois, com uma previsão de fluxo minimamente confiável, será possível realizar diversas pesquisas em todos os campos antropológicos sem a necessidade de sempre se estudar no passado.

Referências

- ANDRE. *Redes neurais artificiais*. 2020. Disponível em: <<https://sites.icmc.usp.br/andre/research/neural/>>. Acesso em: 26 de março de 2022.
- ARAUJO, J. M. de. Estudo da poluição atmosférica no estado de são paulo associada ao material particulado a partir de satélite. *Repositório institucional*, Unifesp, p. 12–26, 2019. Disponível em: <<https://repositorio.unifesp.br/handle/11600/55936>>.
- ARBEX, e. a. M. A. A poluição do ar e o sistema respiratório. *SciELO*, Núcleo de Estudos em Epidemiologia Ambiental - USP, p. 644–652, 2012.
- CETESB. *Histórico*. <https://cetesb.sp.gov.br/>. Disponível em: <<https://cetesb.sp.gov.br/historico/#:~:text=Criada%20em%2024%20de%20julho,da%20Polui%C3%A7%C3%A3o%20das%20%C3%81guas%20e>>. Acesso em: 28 de março de 2022.
- DOCKERY, C. A. P. I. . D. W. Health effects of fine particulate air pollution: Lines that connect. *Journal of the Air Waste Management Association*, p. 715–740, 2012. Disponível em: <<https://doi.org/10.1080/10473289.2006.10464485>>.
- GIODA, G. M. . M. F. C. Q. . A. Estudo da qualidade do ar no entorno da pontifícia universidade católica do rio de janeiro. *Dignidade Re-vista*, PUC Rio, p. 14, 2017. Disponível em: <<http://periodicos.puc-rio.br/index.php/dignidaderevista/article/view/403>>.
- GOMEZ, I. *Redes Neurais: O que são e qual sua importância?* Disponível em: <https://www.sas.com/pt_br/insights/analytics/neural-networks.html#:~:text=Redes%20neurais%20s%C3%A3o%20sistemas%20de,tempo%20%E2%80%93%20aprender%20e%20melhorar%20continuamente.> Acesso em: 26 de março de 2022.
- IBGE. *Instituto Brasileiro de Geografia e Estatística*. 2020. Disponível em: <<https://www.ibge.gov.br/cidades-e-estados/sp.html>>. Acesso em: 30 de março de 2022.
- IBM. *Redes neurais: O que são redes neurais?* 2020. Disponível em: <<https://www.ibm.com/br-pt/cloud/learn/neural-networks>>. Acesso em: 26 de março de 2022.
- JOSEPH, G. W. E. . S. V. Air pollution and human behavior. *American psychological association*, 750 First Street NE, Washington, DC 20002-4242, p. 23–32, 2016.
- LANDULFO, G. A. M. . I. S. A. . A. C. . F. J. S. L. . A. C. Y. . A. A. G. . J. J. da S. . E. Influence of a biomass-burning event in pm2.5 concentration and air quality: A case study in the metropolitan area of são paulo. *Sensors*, Universidade Federal de São Paulo, Diadema, SP, Brazil, p. 21 – 425, 2021. Disponível em: <<https://www.mdpi.com/1424-8220/21/2/425/pdf>>.
- LIMA, A. C. C. de. Concentração de material particulado inalável mp10 em espírito santo do pinhal - sp. *CC*, Espírito Santo do Pinhal, p. 133–144, 2008.
- LUI, E. S. *Previsão de material particulado a curto e médio prazos com o uso de redes neurais artificiais*. Dissertação (Doutorado) — Escola de Engenharia de São Carlos, USP, 2016.

- MAYER, H. Air pollution in cities. *ScienceDirect*, Universidade de Freiburg, Alemanha, p. 5–11, 1999. Disponível em: <[https://doi.org/10.1016/S1352-2310\(99\)00144-2](https://doi.org/10.1016/S1352-2310(99)00144-2)>.
- MONTEIRO, M. A. C. S. Desenvolvimento de um sistema de avaliação e previsão da qualidade do ar para Portugal. *ProQuest*, Universidade de Aveiro, Portugal, p. 32–36, 2007. Disponível em: <<https://www.proquest.com/openview/2c2d2b58d2f8f51052e59752b6e811b7/1?pq-origsite=gscholar&cbl=2026366&diss=y>>.
- MOURA, I. F. S. *Avaliação de MP10 na região metropolitana de Belo Horizonte*. Dissertação (Mestrado) — Tecnologia nuclear - Universidade Federal de Minas Gerais, 2016.
- PATERLINI, W. C. Fontes e composição das partículas atmosféricas na área urbana e rural da região central do estado de São Paulo. *Repositório Institucional Unesp*, UNESP - Universidade Estadual Paulista, p. 1–77, 2007.
- PAULINO, F. O. Uso de red backpropagation para prognóstico de la concentración diaria de mp10. p. 1–3, 2020.
- QUALAR. *Qualidade do Ar de SP*. <https://qualar.cetesb.sp.gov.br/>. Disponível em: <<https://qualar.cetesb.sp.gov.br/qualar/home.do>>. Acesso em: 30 de março de 2022.
- RAMANATHAN, C. E. C. . G. C. R. . M. V. R. . D. K. . V. Capturing vertical profiles of aerosols and black carbon over the Indian Ocean using autonomous unmanned aerial vehicles. *Nature*, Scripps Institution of Oceanography, La Jolla, California, p. 448–575, 2007. Disponível em: <<https://hal.archives-ouvertes.fr/hal-00303026>>.
- RIZZO, M. C. Q. D. O. . A. D. . L. V. Air pollution persistent exceedance events in the Brazilian metropolis of São Paulo and associated surface weather patterns. *ORIGINAL PAPER*, Universidade Federal de São Paulo, Diadema, SP, Brazil, p. 1–4, 2021. Disponível em: <<https://doi.org/10.1007/s13762-021-03778-1>>.
- SIQUEIRA, L. N. A. L. . J. T. B. . L. V. da Silva Manoel H. de N. Marinho Thiago Antonini Alves Yara de S. T. . H. V. Redes neurais feedforward aplicadas na avaliação do impacto da poluição atmosférica e variáveis climáticas na saúde humana. In: ISBN (INEGI/FEUP). *Conference on Theoretical and Experimental Mechanics and Materials*. 11. ed. Porto/Portugal: National Congress on Experimental Mechanics, 2018. p. 275–284.
- SOLCI, A. de M. F. . M. C. Caracterização do mp10 e mp2,5 e distribuição por tamanho de cloreto, nitrato e sulfato em atmosfera urbana e rural de Londrina. *SciELO Brasil*, Universidade Estadual de Londrina, PR, Brasil, p. 23–32, 2009. Disponível em: <<https://doi.org/10.1590/S0100-40422009000700013>>.
- THOLGATE, P. B. B. . P. S. Air pollution and health. *The Lancet*, Universidade de Utrecht, Finlândia, p. 1233–1242, 2002. Disponível em: <[https://doi.org/10.1016/S0140-6736\(02\)11274-8](https://doi.org/10.1016/S0140-6736(02)11274-8)>.

ANEXO A – Tabelas

Tabela 3 – Médias anuais a partir dos dados diários de MP₁₀ medidas de cada estação em $\mu\text{g}/\text{m}^3$

Ano	Pq dom pedro II	Guarulhos - Pimentas	Santos - Centro	Jaú
2017	26.60 ± 18.6	22.74 ± 14.90	18.67 ± 14.55	26.04 ± 22.27
2018	27.80 ± 20.83	39.27 ± 41.46	17.24 ± 14.6	24.6 ± 17.75
2019	29.83 ± 23.04	31.15 ± 26.80	17.39 ± 16.8	25.21 ± 17.75
2020	27.24 ± 21.69	28.23 ± 24.15	22.38 ± 15.24	24.30 ± 22.97
2021	29.55 ± 25.47	29.89 ± 25.28	19.66 ± 12.57	25.20 ± 20.94

Tabela 4 – Médias anuais a partir dos dados diários de MP₁₀ previstas de cada estação em $\mu\text{g}/\text{m}^3$

Ano	Pq dom pedro II	Guarulhos - Pimentas	Santos - Centro	Jaú
2017	24.94 ± 18.12	22.11 ± 15.11	18.07 ± 14.20	25.72 ± 21.50
2018	28.16 ± 23.23	38.72 ± 42.82	17.81 ± 15.87	24.79 ± 17.61
2019	30.03 ± 25.12	31.91 ± 26.66	17.67 ± 16.66	25.4 ± 16.25
2020	26.55 ± 22.97	28.98 ± 24.69	22.42 ± 15.91	24.32 ± 23.27
2021	29.12 ± 26.27	28.50 ± 25.02	19.53 ± 12.12	25.98 ± 21.52

Tabela 5 – Médias anuais a partir dos dados diários de MP_{2.5} medidas de cada estação em $\mu\text{g}/\text{m}^3$

Ano	Pq dom pedro II	Guarulhos - Pimentas
2017	17.36 ± 17.62	18.16 ± 17.92
2018	19.23 ± 12.47	20.75 ± 28.78
2019	16.82 ± 15.24	19.48 ± 16.51
2020	15.23 ± 14.09	17.51 ± 19.18
2021	17.78 ± 16.18	18.01 ± 13.45

Tabela 6 – Médias anuais a partir dos dados diários de MP_{2.5} previstos de cada estação em $\mu\text{g}/\text{m}^3$

Ano	Pq dom pedro II	Guarulhos - Pimentas
2017	16.67 ± 13.29	17.94 ± 16.07
2018	17.62 ± 14.74	21.26 ± 29.35
2019	17.43 ± 13.57	20.1 ± 17.88
2020	16.58 ± 13.3	18.24 ± 16.11
2021	16.75 ± 14.81	16 ± 15.11

ANEXO B – Códigos

B.1 Código Utilizado para unir e tratar os dados brutos baixados direto do site:

```
import pandas as pd
import numpy as np
ano=['2017-2018','2018-2019','2019-2020','2020-2021','2021-2022']
nome='Santos'
doc='.csv'

df=[0,0,0,0,0]

for i in range(0,len(ano)):
    arquivo = str(nome + ano[i] + doc)
    df[i]= pd.read_csv(arquivo, encoding='latin-1')

df_final = pd.concat(df, ignore_index=True)
for i in list(df_final.columns):
    try:
        media=df_final[i].mean()
        df_final[i].fillna(value=media,inplace=True)
    except:
        continue
df_final.to_csv(f'Raw/{nome}.csv', index=False, encoding='latin-1')

arquivo1 = 'Santos'
arquivo2 = 'SantosMP10'
tipo='.csv'

df1=pd.read_csv(arquivo1+tipo, encoding='latin-1')
df2=pd.read_csv(arquivo2+tipo, encoding='latin-1')

df_final = pd.merge(df1,df2)
df_final = df_final.groupby(np.arange(len(df_final))//24).mean()
df_final.to_excel(arquivo1+' Final.xlsx', encoding='latin-1', index=False)
```

B.2 Código da MLP utilizada:

```
clc
[numbers,strings]=xlsread('Guarulhos Final.xlsx','Sheet1');
inputs=numbers(2:1829,1:4);
outputs=numbers(2:1829,6);
persist = numbers(2:1829,4);
persist = persist';
```

```

inputs=inputs';
outputs=outputs';
[mapedinputs,mapinp]=mapminmax(inputs,-1,1);
[mapedoutputs,mapout]=mapminmax(outputs,-1,1);
net=newff(mapedinputs,mapedoutputs,[40 40]);
net.layers1.transferFcn='tansig';
net.layers2.transferFcn='tansig';
net=init(net);
net.trainFcn='trainscg';
net.divideFcn='divideint';
net.divideParam.trainRatio=0.7;
net.divideParam.testRatio=0.15;
net.divideParam.valRatio=0.15;
net.trainParam.epochs=2000;
net.trainParam.max_fail = 150;

[net,tr]=train(net,mapedinputs,mapedoutputs);
Sout=sim(net,mapedinputs);
[DeMapedResponse,PS]=mapminmax('reverse',Sout,mapout);
Error=abs(outputs-DeMapedResponse);
MediaErroAbs=mean(Error);
ErroPerc = Error./outputs;
MediaErroPerc=mean(ErroPerc);
ErrorPersist = abs(persist-outputs);
MediaErroPersist = mean(ErrorPersist);
ErroPercPers = ErrorPersist./outputs;
MediaErroPercPers=mean(ErroPercPers);
disp('Erro Absoluto Médio');
MediaErroAbs
disp('Erro percentual Médio');
MediaErroPerc
figure(1)
plot(Error,'r')

xlabel('Día')
ylabel('Error absoluto (ug/m³)')
filename='erro_absoluto_mlpmp';
figure(2)
hold on

```

```
plot(outputs,'g')
plot(DeMappedResponse,'b')
xlabel('Dia')
ylabel('Concentração (ug/m3)')
legend('Medido','Predido')
Error;
```