

Análise da qualidade do ar para classificação atmosférica

João Pedro Felix da Silva
UFPE
Recife, Brasil
jpfs2@cin.ufpe.br

Pedro Henrique A. G. Peixinho
UFPE
Recife, Brasil
phagp@cin.ufpe.br

José Victor de Luna da Silva
UFPE
Recife, Brasil
jvls@cin.ufpe.br

Victor Gabriel de Carvalho
UFPE
Recife, Brasil
vgc3@cin.ufpe.br

Abstract—No mundo moderno, estamos cada vez mais inseridos em uma sociedade em constante industrialização. Tal expansão técnico-industrial, apesar de benéfica em muitos casos, quando não é bem administrada, traz prejuízos para o meio ambiente, especialmente no que tange a nossa atmosfera. Diante disso, é fundamental o estudo das condições do ar que respiramos para que possamos evitar sua poluição e vivermos com mais qualidade. Para isso, estamos propondo um modelo que, dado um banco de testes, consegue classificar a qualidade de determinada amostra de ar.

Palavras chave—industrialização, qualidade, atmosfera, poluição, modelo.

I. INTRODUÇÃO

A qualidade do ar é conceituada como uma junção de fatores e características presentes na atmosfera de um determinado lugar que estabelecem o quão prejudicial pode ser frequentar naquele ambiente. Infelizmente, a qualidade do ar em nosso planeta enfrenta uma constante piora no decorrer das últimas décadas. A crescente expansão industrial e a maior utilização de combustíveis fósseis estão entre os maiores contribuintes para a escalada dos índices de poluição atmosférica. Consequentemente, tal fato traz consigo diversos prejuízos para a população e a natureza, acarretando em doenças, gastos e o famoso aquecimento global.

II. OBJETIVOS

Nosso trabalho tem como objetivo analisar os dados sobre a concentração de gases atmosféricos em um determinado local durante o ano e chegar a um modelo que possa determinar a qualidade do ar a partir da relação das informações obtidas nas experimentações. Tal modelo irá permitir que os locais sejam classificados de acordo com o nível da poluição atmosférica e tomar atitudes de acordo com esse nível. Isso irá contribuir para uma melhor intervenção nos ambientes que apresentem altos índices de poluição, permitindo uma ação mais rápida e eficaz para o controle da situação.

III. JUSTIFICATIVA

É inegável que a qualidade do ar é algo fundamental para a vida em nosso planeta, seja na respiração dos seres vivos, seja na dinâmica atmosférica das chuvas. Desse modo, torna-se fundamental analisar e entender a condição da atmosfera em que estamos inseridos, a fim de que possamos intervir no

aumento de poluição e no consequente decréscimo da qualidade do ar.

Sendo assim pelo estudo da qualidade do ar podemos ter uma noção básica de como a sociedade poderá agir ao longo dos próximos anos para minimizar os efeitos causados pela poluição proveniente dos automóveis, indústrias etc. que são uns dos principais motivos para a poluição do ar e atmosfera da terra, surtindo efeito no ecossistema e principalmente na sociedade.

IV. METODOLOGIA

Foi utilizado o método de pesquisa exploratória essencialmente em fontes primárias e secundárias para a obtenção dos dados e das informações necessárias para o progresso do projeto.

Para a realização do nosso projeto, será necessário o uso de um banco de dados (contendo a coleta dos índices atmosféricos em uma cidade italiana durante um ano) obtido na UCI Machine Learning Repository e o uso da linguagem de programação Python 3 (com as bibliotecas NumPy, Matplotlib, Scikit-Learn, SciPy, Pandas) na plataforma Colab para implementar a análise Bayesiana, famoso método de classificação de eventos.

Os dados atmosféricos analisados contêm a data e hora das medições (uma medição por hora durante um ano) onde foi registrada a concentração de CO (mg/m³), Hidrocarbonetos não metálicos (µg/m³), Benzeno (µg/m³), NOx (ppb), NO₂ (µg/m³), os dados também abrangem a umidade relativa do ar (%) a umidade absoluta e a temperatura ambiente do local (°C). Como mostrado na tabela abaixo.

TABLE I. CONCENTRAÇÃO DOS GASES (10/03/2004)

Hora	CO (mg/m ³)	Hidrocarbonetos não metálicos (µg/m ³)	Benzeno (µg/m ³)	NOx (ppb)
18:00	2,6	150	11,9	166
19:00	2	112	9,4	103
20:00	2,2	88	9,0	131
21:00	2,2	80	9,2	172
22:00	1,6	51	6,5	131

Hora	NO ₂ (µg/m ³)	Humidade relativa (%)	Humidade absoluta	Temperatura (°C)
18:00	113	48,9	0,7578	13,6
19:00	92	47,7	0,7255	13,3
20:00	114	54,0	0,7502	11,9
21:00	122	60,0	0,7867	11,0
22:00	116	59,6	0,7888	11,2

Além disso, para cada gás supracitado, foi feita, também, uma medição dos mesmos através de sensores de oxidação. Como exemplificado na tabela abaixo.

TABLE II. MEDIDAS REGISTRADAS PELOS SENSORES DE OXIDAÇÃO (10/03/2004)

Hora	Sensor de resposta para CO	Sensor de resposta para HCNM	Sensor de resposta para NO _x	Sensor de resposta para NO ₂	Sensor de resposta para O ₃
18:00	1360	1046	1056	1692	1268
19:00	1292	955	1174	1559	972
20:00	1402	939	1140	1555	1074
21:00	1376	948	1092	1584	1203
22:00	1272	836	1205	1490	1110

Primeiramente, os dados do banco serão baixados e formatados usando a biblioteca Pandas, com o objetivo de ser preparado para a posterior análise.

Posteriormente, os dados serão submetidos à análise Bayesiana, onde a relação entre os dados formará um modelo que funcionará para prever futuros casos-teste quanto à qualidade do ar. Tal modelo, conhecido como Classificador Naive Bayes é utilizado em machine learning e se baseia no Teorema de Bayes,

$$P(B_i|A) = \frac{P(A|B_i) P(B_i)}{\sum_j P(A|B_j) P(B_j)}$$

que classifica a probabilidade da ocorrência de um evento B_i dado que um evento A tenha ocorrido. Podemos utilizar essa propriedade para classificar um novo evento (como as características do ar de algum lugar) comparando-as com padrões já vistos nos casos de treinamento. Contudo, o Classificador Naive Bayes assume a independência entre os eventos analisados, o que pode prejudicar seu desempenho em alguns casos, pois nem sempre existe tal autonomia. Essa característica foi a responsável pelo nome “naive” (ingênuo em inglês) que o classificador recebe. Mas, para fins acadêmicos, os resultados obtidos são suficientemente satisfatórios.

A biblioteca NumPy será essencial para a realização dos cálculos científicos que envolvem o Classificador de Bayes, e a Matplotlib ajudará na visualização das etapas e dos resultados do modelo devido à capacidade de exibir visualizações do projeto, como tabelas e gráficos.

V. CRONOGRAMA DE ATIVIDADES

Para melhor fluxo do trabalho e maior organização da equipe, o trabalho foi dividido entre as 6 semanas e foram criados 8 tópicos (etapas) centrais para a dividir o projeto.

Lista de atividade	Semana 1 (07/11-13/11)	Semana 2 (14/11-20/11)	Semana 3 (21/11-27/11)
Escolha do tema e do banco de dados	X		
Elaboração da Proposta de projeto	X		
Tratamento do banco de dados		X	X
Implementação do Classificador Bayesiano			X
Treinamento do algoritmo			
Teste com novos casos			
Formulação do relatório			
Apresentação e entrega do projeto			
Lista de atividade	Semana 4 (28/11-04/12)	Semana 5 (05/12-11/12)	Semana 6 (12/12-16/12)
Escolha do tema e do banco de dados			
Elaboração da Proposta de projeto			
Tratamento do banco de dados			
Implementação do Classificador Bayesiano	X		
Treinamento do algoritmo	X		
Teste com novos casos (experimentos)		X	
Formulação do relatório		X	X
Apresentação e entrega do projeto			X

REFERENCIAS

- [1] P. L. MEYER, “Probabilidade - Aplicações à estatística”, 2ª ed, Washington.
- [2] Vito, Saverio. (2016). Air Quality. UCI Machine Learning Repository.
- [3] Min. do Meio Ambiente, “Qualidade do Ar”, disponível em: <https://antigo.mma.gov.br/cidades-sustentaveis/qualidade-do-ar.html>.
- [4] P. C. T. Gomes, “Classificação com Naive Bayes”, disponível em: <https://www.datageeks.com.br/naive-bayes/>.

