

# Análise de Modelo de Aprendizagem de Máquina envolvendo previsões meteorológicas na cidade do Recife

L. F. Azevedo, L. F. Guedes, L. M. Veloso, M. G. Souza, V. H. Mendonça

*#Centro de Informática - Universidade Federal de Pernambuco  
Recife, Brasil*

<sup>1</sup>[lfa5@cin.ufpe.br](mailto:lfa5@cin.ufpe.br)

<sup>2</sup>[lfgsm@cin.ufpe.br](mailto:lfgsm@cin.ufpe.br)

<sup>3</sup>[lmvs@cin.ufpe.br](mailto:lmvs@cin.ufpe.br)

<sup>4</sup>[mgs11@cin.ufpe.br](mailto:mgs11@cin.ufpe.br)

<sup>5</sup>[vhmq@cin.ufpe.br](mailto:vhmq@cin.ufpe.br)

**Resumo** — Na cidade do Recife o clima é bastante instável, e por vezes isso acaba por prejudicar algumas áreas ou grupos da cidade. Um modelo de Inteligência Artificial pode se mostrar uma ferramenta bastante útil na predição do tempo. Este trabalho tem como objetivo julgar o desempenho de certo modelo de IA na tarefa descrita acima.

**Palavras-Chave** — Análise de Dados, Estatística, Modelo de Aprendizagem, Previsões Meteorológicas e Inteligência Artificial

**Abstract** — In the city of Recife the climate is quite varied, and sometimes this ends up harming some areas or groups in the city. An Artificial Intelligence model can prove to be a very useful tool in weather prediction. This work aims to judge the performance of a certain AI model in the task described above.

**Keywords** — Data Analysis, Statistics, Learning Model, Weather Predictions and Artificial Intelligence.

## I. INTRODUÇÃO

A utilização de modelos de aprendizagem de máquina tornou-se uma prática extremamente estudada e pesquisada por diversos pesquisadores mundiais. Para que seja implementado um modelo adequado para o uso útil, tem-se a necessidade de, por meio de inferências estatísticas, realizar uma análise minuciosa dos resultados gerados por esse modelo.

A análise para a eficácia do modelo partiu de uma constatação de um conjunto de dados que retornam a probabilidade de acerto da previsão meteorológica desse algoritmo da cidade do Recife, localizado no nordeste brasileiro. Tem-se como objetivo, por meio dessa análise, concluir se o programa de aprendizagem de máquina está satisfatório para uma possível consideração de uso útil.

A metodologia utilizada, exemplificada com um determinado rigor no setor III, aborda quanto às ferramentas adotadas para a construção dos testes de hipóteses e para a possível validação do programa para a utilização adequada. Ao decorrer do setor citado, as explicações envolvendo o ambiente em que os dados foram extraídos e bibliotecas utilizadas para a avaliação por meio de linguagem de programação serão apresentadas de forma mais objetiva e adequada para o entendimento geral.

## II. OBJETIVOS

O principal resultado que almejamos com essa análise é descobrir a confiabilidade do modelo que estamos analisando, o que pode ser útil em uma cidade como Recife, que possui um clima altamente variado que frequentemente causa grandes impactos na vida de pessoas em situação de vulnerabilidade social, as quais vivem em encostas e morros. Nesta vista a predição do tempo na cidade é um importante item para manutenção do bem estar social da população.

Desta vista, nosso principal foco vai ser medir a acurácia do modelo de Inteligência Artificial para previsões meteorológicas na cidade do Recife com o auxílio de métodos estatísticos e de análise de dados.

### III. METODOLOGIA

#### A. Ambiente geográfico

O ambiente de estudo para a aplicação do modelo de aprendizagem é localizado em uma região caracterizada pelo clima Tropical Úmido (considerado como “Aw” ou “As” baseado na classificação climática de Köppen-Geiger). O clima da região é caracterizado por elevadas temperaturas, típicas dos aspectos tropicais, e uma alta precipitação atmosférica (valores elevados a 2000mm).

O local analisado corresponde ao espaço referente a cidade do Recife, localizado no nordeste brasileiro no estado de Pernambuco. A localização da cidade sofre influências de fenômenos geográficos específicos que acarretam instabilidades para uma análise precisa das informações.

#### B. Sobre a coleta de dados

Os dados obtidos foram extraídos de um intervalo de 1000 dias, uma amostra ideal para a análise da eficácia de um modelo de aprendizagem de máquina, iniciando-se a partir da data de 01/01/2019 até 27/09/2021 na cidade do Recife. Considera-se que, durante esse período, fenômenos como El Niño e La Niña impactaram nas condições meteorológicas da região analisada.

#### C. Tecnologias adotadas para a análise de dados

A fim de se chegar em um objetivo satisfatório, certas tecnologias foram utilizadas para que fosse realizado a análise dos dados extraídos das previsões meteorológicas. A adoção da linguagem de programação “Python” tornou-se crucial para o estudo, pois essa tecnologia possui diversos meios para a geração de resultados agradáveis, por meio da geração de gráficos, cálculos envolvendo as medidas centrais e outras questões. Entre as principais bibliotecas utilizadas pode-se citar Numpy, Matplotlib, Scipy, Seaborn e Statistics (biblioteca nativa da linguagem Python). Os gráficos gerados ao decorrer do seguinte relatório escrito foram gerados por meio dos seguintes arquivos de códigos que podem ser acessados por esse [link](#).

### IV. SOBRE A BASE DE DADOS

A base de dados apresenta informações sobre a porcentagem de acertos em previsões meteorológicas feitas por um modelo de aprendizado de máquina. Cada linha da base representa um dia de previsão da IA.

TABELA I  
MEDIDAS CENTRAIS DE DISPERSÃO

Medidas	Valores
Curtose	0.006006568742133567
Moda	38.26
Máximo	97.32
Mínimo	2.53
Média	50.46723
Mediana	51.07
1° Quartil	40.385000000000005
3° Quartil	60.4975
Variância	224.70795798508507
Desvio padrão	14.990262105283051

A tabela I mostra as medidas centrais de dispersão, o que nos apresenta que a maioria dos dados estão no centro, entre 40% e 70%, e estão bem próximos da média e mediana. A variância e o desvio padrão nos informa que há alguns outliers nos dados.

### V. TESTE DE HIPÓTESE I

Nossa hipótese nesta análise é de que vão haver diversos outliers nos resultados por conta do clima variado e instável da cidade do Recife.

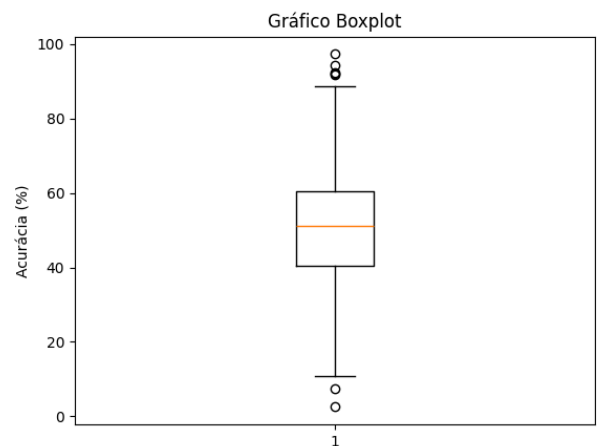


Fig. 1. Gráfico boxplot - Acurácia

A Fig. 1. é um gráfico um boxplot, que reforça a ideia de que grande parte dos dados estão próximos

à média e à mediana, havendo alguns outliers, tanto na parte superior, quanto na inferior, o que é normal, pois, como foi citado anteriormente, o local da coleta de dados foi feita em uma região instável.

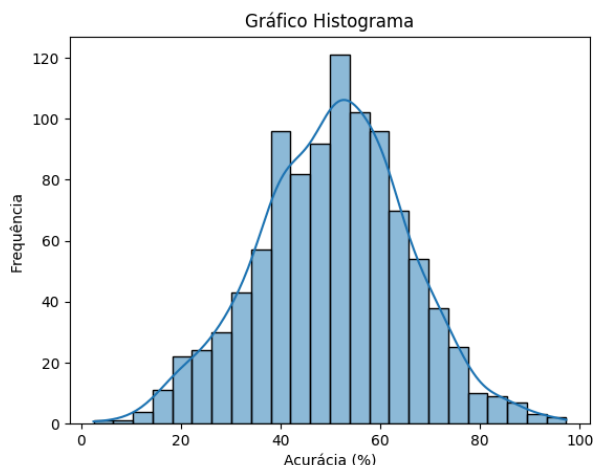


Fig. 2 gráfico histograma- frequência x acurácia

Fig. 2. levanta fortes indícios de que a distribuição dos dados coletados segue a distribuição normal, uma vez que a curvatura formada lembra um sino, característica típica desse tipo de distribuição, o que poderá ser confirmado com o teste de hipótese 2.

## VI. TESTE DE HIPÓTESE II

Conforme a figura 2 sugere, vamos verificar a distribuição que os dados seguem.

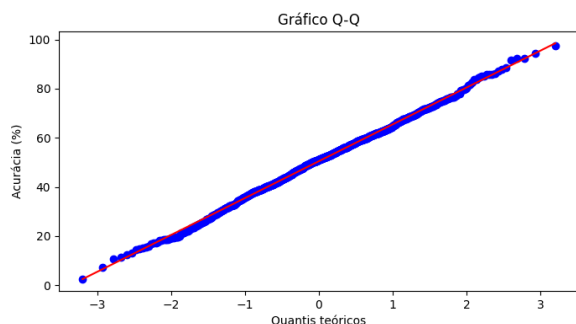


Fig. 3 Gráfico Q-Q

A Fig. 3. é um gráfico um Q-Q, o qual pode-se observar que reforça a ideia de que os dados

analisados se aproximam de uma distribuição normal, uma vez que os quantis do conjunto se apresentam em cima da reta de 45 graus.

## VII. CONCLUSÕES

O modelo de IA teve um desempenho razoável na predição do clima da cidade do Recife, uma vez que representou com fidelidade as nuances que ocorrem (outliers), que é algo recorrente do clima do local, apesar de ter grande grau de dificuldade de previsão. Além disso, a porcentagem de acurácia se mostrou satisfatória, principalmente quando se analisa as medidas de dispersão que foram mostradas aqui, o que destaca sua provável confiança para ser utilizado nos objetivos para o qual foi criado.

## REFERÊNCIAS

- [1] Arnfield, A. John. "Köppen climate classification." In: Encyclopædia Britannica, Encyclopædia Britannica, inc., 2024. Disponível em: <https://www.britannica.com/science/Koppen-climate-classification>. Acesso em: [19/03/2024].
- [2] Weather Spark. Clima característico em Recife, Brasil durante o ano. Disponível em: <https://pt.weatherspark.com/y/31432/Clima-caracter%C3%ADstico-em-Recife-Brasil-durante-o-ano>. [Acesso em: 19/03/2024].
- [3] Declan V. "Python Data Analysis with Pandas and Matplotlib" In: Our Coding Club, Our Coding Club, inc., 2024. Disponível em: <https://ourcodingclub.github.io/tutorials/pandas-python-intro/#:~:text=matplotlib%20is%20a%20Python%20package,%2C%20maps%2C%20and%20other%20visualisations>. Acesso em: [18/03/2024].
- [4] Aayush B. "Performance Metrics in Machine Learning" In: neptune.ai, Disponível em: <https://neptune.ai/blog/performance-metrics-in-machine-learning-complete-guide> Acesso em: [19/03/2024].