

Previsão da Temperatura Máxima Utilizando Regressão Linear

Leonardo Henrique de Oliveira Matos, 10389516¹, Victor Junqueira Colombaro, 10395711¹

¹Ciência da Computação
Faculdade de Computação e Informática
Universidade Presbiteriana Mackenzie
São Paulo – SP – Brasil

Resumo. *Este projeto emprega a regressão linear para estimar a temperatura máxima diária em Seattle, utilizando dados climáticos como precipitação, velocidade do vento e condições meteorológicas. O modelo demonstrou boa capacidade de prever a temperatura, considerando variáveis numéricas e categóricas. A metodologia envolveu análise exploratória de dados, preparação do dataset, normalização de variáveis e avaliação do modelo através de análises gráficas de resíduos e outras métricas relevantes.*

1. Introdução

1.1. Contextualização

A previsão de condições climáticas é uma aplicação clássica de técnicas de machine learning, especialmente quando se trata de variáveis contínuas, como a temperatura. Com o avanço de algoritmos de regressão e a disponibilidade de dados históricos, é possível prever a temperatura de dias futuros com base em padrões observados no clima.

1.2. Justificativa

A previsão precisa da temperatura é essencial para diversos setores, incluindo planejamento diário, agrícola e industrial. Com a ajuda de *machine learning*, particularmente através de regressão linear, podemos desenvolver um modelo eficiente que prevê a temperatura máxima diária com base em variáveis climáticas. Para garantir que o modelo possa lidar com variáveis categóricas, como condições de tempo (ensolarado, chuvoso, nublado), estas foram transformadas em variáveis *dummies*, seguindo a descrição por GÉRON (2019) [2]. Isso permite que o modelo processe informações categóricas de maneira adequada.

Além disso, utilizamos o *R-squared* (R^2) como uma métrica fundamental para avaliar o desempenho do modelo. O valor de R^2 varia de 0 a 1, onde valores mais próximos de 1 indicam que o modelo explica bem a variabilidade dos dados.

1.3. Objetivo

O objetivo deste projeto é construir um modelo de regressão capaz de prever a temperatura máxima com base em variáveis climáticas, como precipitação, velocidade do vento, e condições meteorológicas (clima).

1.4. Opção do projeto

Optou-se pela utilização de um framework de aprendizado de máquina, especificamente o scikit-learn, para implementar o modelo de regressão e realizar a análise exploratória dos dados com Python.

2. Descrição do Problema

Prever a temperatura máxima diária é um problema complexo, uma vez que a temperatura é influenciada por múltiplos fatores climáticos que interagem de forma não linear. Neste estudo, busca-se modelar essa relação utilizando a regressão linear, apesar das limitações impostas pela natureza não linear dos fenômenos climáticos.

3. Descrição do Dataset

O dataset utilizado contém dados climáticos da cidade de Seattle. As variáveis incluem:

- **date:** Data da observação.
- **precipitation:** Quantidade de precipitação em milímetros.
- **temp_max:** Temperatura máxima do dia (alvo de previsão).
- **temp_min:** Temperatura mínima do dia.
- **wind:** Velocidade do vento.
- **weather:** Condições meteorológicas categorizadas (ensolarado, chuvoso, nublado, etc.).

A preparação dos dados incluiu:

- Conversão de variáveis categóricas (como o tempo) em variáveis dummies.
- Normalização de variáveis contínuas (precipitação e vento) usando *StandardScaler*.

4. Metodologia e Resultados Esperados

Gráficos de correlação foram utilizados para visualizar as relações entre as variáveis contínuas, como precipitação, temperatura mínima e velocidade do vento. Posteriormente, o dataset foi preparado, o que incluiu a normalização das variáveis contínuas (precipitação e vento) com a função *StandardScaler*, além da conversão de variáveis categóricas (como as condições climáticas) em variáveis dummies.

Após a preparação dos dados, foi implementado um modelo de **Regressão Linear** para prever a temperatura máxima diária com base nas variáveis preditoras.

Como parte da análise final, foi gerado um gráfico de dispersão comparando os valores reais e previstos da temperatura máxima, a fim de avaliar visualmente a precisão do modelo. Além disso, foi conduzida uma **análise de resíduos** para identificar a distribuição dos erros e verificar a presença de possíveis padrões não capturados pelo modelo.

Espera-se que o modelo de regressão linear forneça previsões precisas da temperatura máxima, especialmente em dias com condições climáticas estáveis. O modelo deve apresentar menor erro em dias com baixa variabilidade climática e demonstrar bom desempenho com os dados climáticos disponíveis.

5. Bibliografia

References

- [1] Scikit-learn: Machine Learning in Python, *scikit-learn.org*. Disponível em: <https://scikit-learn.org>. Acesso em: 29 set. 2024.

- [2] GÉRON, Aurélien. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2 ed. Sebastopol: O'Reilly, 2019.
- [3] Weather Dataset. *Kaggle.com*. Disponível em: <https://www.kaggle.com>. Acesso em: 20 set. 2024.