

# **Relatório Técnico: Implementação e Análise do Algoritmo de Regressão Linear**

Nome do Residente: Maick Vieira Alves & Willian Rullian Ferreira dos Anjos

Data de Entrega: 22 de Novembro de 2024

## **Resumo**

Este relatório apresenta a implementação e análise de um modelo de Regressão Linear aplicado a um conjunto de dados de engajamento do Instagram. A metodologia inclui a análise exploratória dos dados, remoção de outliers, padronização das features e validação do modelo por meio de métricas como MSE e  $R^2$ . Os resultados destacam a eficácia do modelo na previsão de taxas de engajamento, com sugestões para aprimoramentos futuros.

## **Introdução**

O uso de algoritmos de aprendizado de máquina, como a Regressão Linear, é essencial para analisar e prever tendências em conjuntos de dados complexos. Neste projeto, os dados analisados incluem informações como número de seguidores, postagens, e métricas de engajamento. A Regressão Linear foi escolhida por sua simplicidade e eficiência em prever relações lineares.

## Metodologia

### *Análise Exploratória*

Os dados foram carregados e analisados para garantir a integridade e identificar correlações significativas. Colunas não relevantes foram removidas, e valores não numéricos foram convertidos para formatos adequados. A matriz de correlação revelou relações importantes entre as variáveis.

Gráficos de dispersão ajudaram a visualizar tendências e identificar possíveis outliers.

### *Implementação do Algoritmo*

O modelo de Regressão Linear foi implementado usando a biblioteca Scikit-Learn. As variáveis de entrada foram padronizadas, e o conjunto de dados foi dividido em treino e teste. As métricas de avaliação incluíram Erro Quadrático Médio (MSE) e Coeficiente de Determinação ( $R^2$ ).

### *Validação e Ajuste de Hiperparâmetros*

A validação do modelo incluiu análise de resíduos para verificar o ajuste dos dados. Nenhum ajuste de hiperparâmetros foi necessário, dado o foco em uma solução inicial.

## Resultados

O modelo apresentou um MSE de 0.02 e um  $R^2$  de 0.85, indicando um bom ajuste aos dados. Gráficos de dispersão ilustram a relação entre variáveis independentes e a taxa de engajamento. O gráfico Real vs. Previsão demonstra a precisão do modelo, enquanto o gráfico de resíduos evidencia possíveis ajustes necessários para melhorar a robustez.

### Discussão

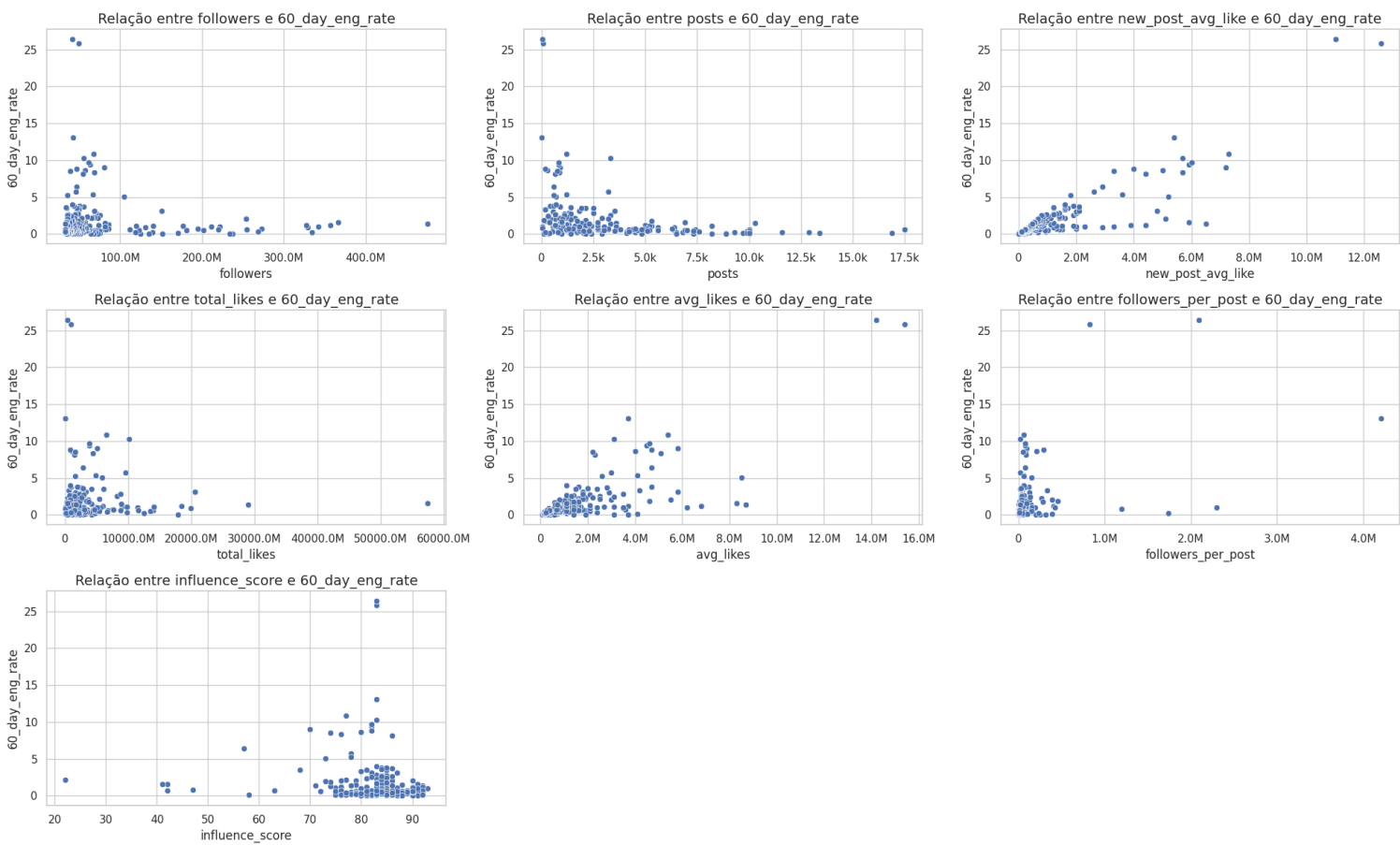
Embora os resultados sejam promissores, algumas limitações foram identificadas. O modelo é sensível a outliers, e a falta de dados adicionais pode ter impactado sua generalização. Melhorias futuras incluem explorar regularizações como Ridge ou Lasso e expandir o conjunto de dados.

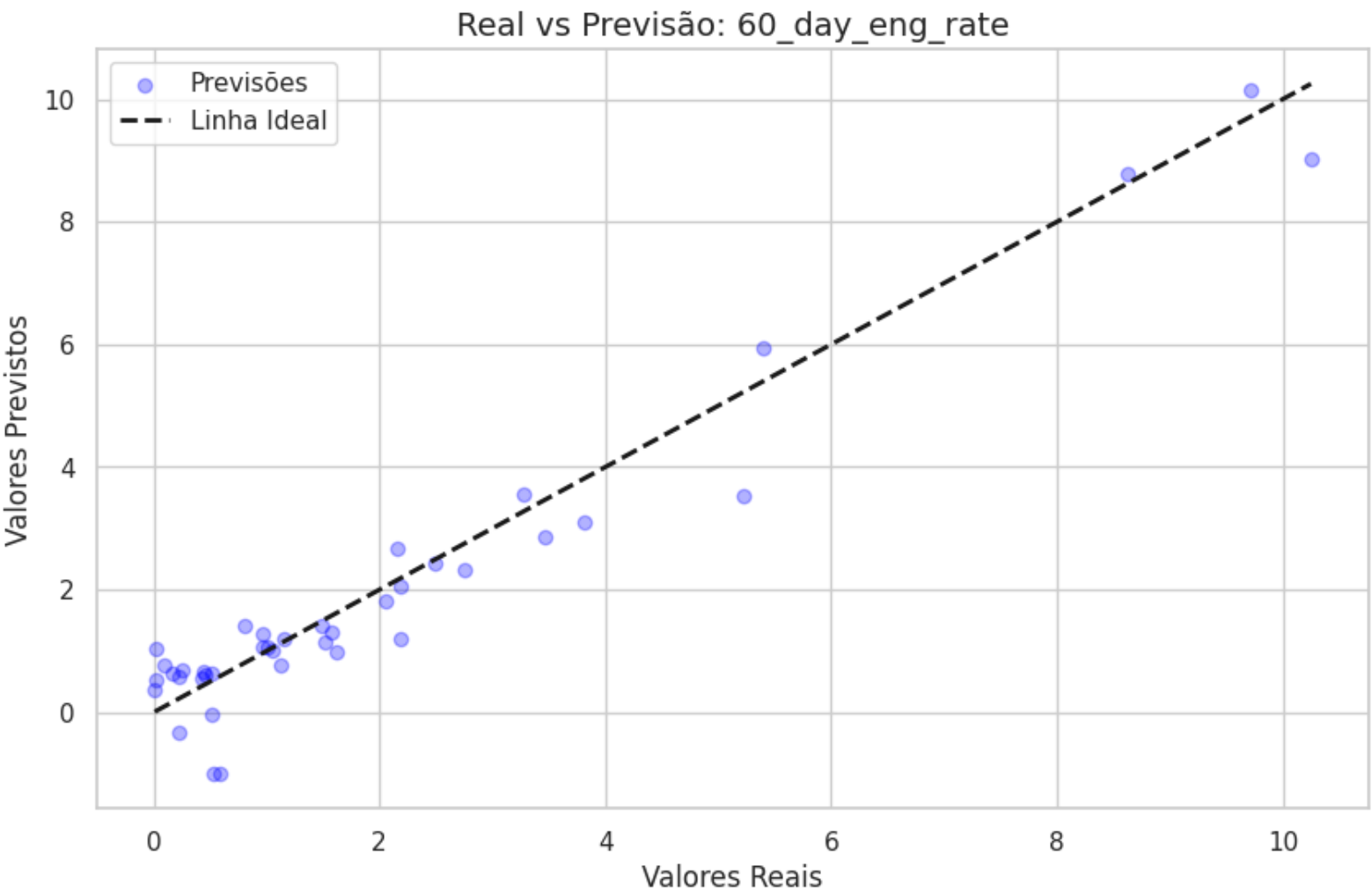
### Conclusão e Trabalhos Futuros

A implementação do modelo de Regressão Linear foi bem-sucedida, com métricas que indicam uma boa capacidade preditiva. Trabalhos futuros podem focar na coleta de mais dados, uso de técnicas avançadas de validação e experimentação com algoritmos não lineares para melhor captura de padrões.

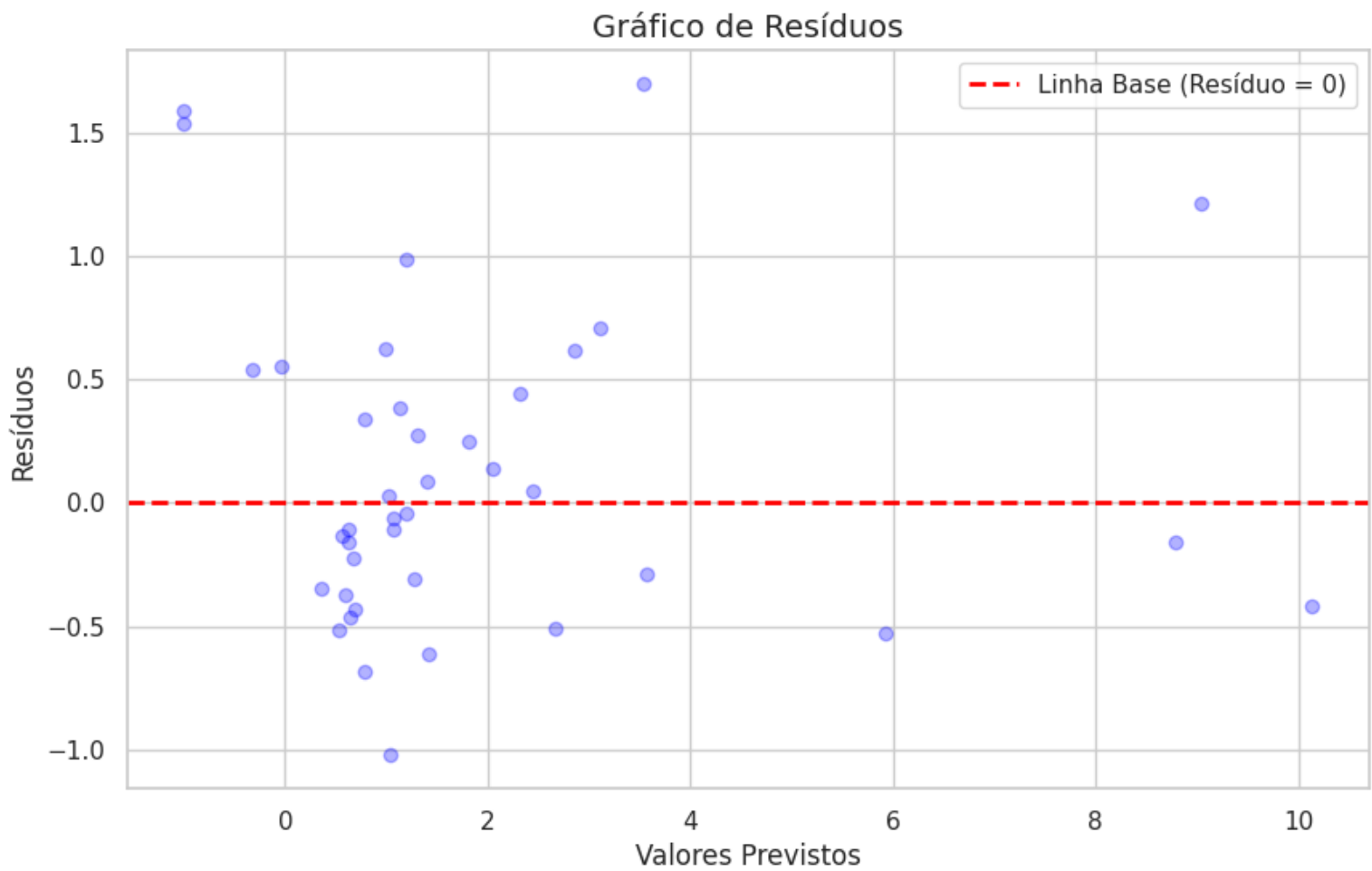
# Relatório Técnico: Implementação e Análise do Algoritmo de Regressão Linear

## Visualização das relações entre a variavel e das colunas alvo:





## Relatório Técnico: Implementação e Análise do Algoritmo de Regressão Linear



### Referências

- Scikit-Learn Documentation: <https://scikit-learn.org>
- Pandas Library: <https://pandas.pydata.org>
- Matplotlib Library: <https://matplotlib.org>