

Beatriz Ribeiro Santana

Milane Souza Andrade

Relatório Técnico: Implementação e Análise do Algoritmo de Regressão Linear

Resumo

Este relatório apresenta o desenvolvimento de um modelo de Regressão Linear aplicado à previsão da taxa de engajamento de influenciadores no Instagram. O trabalho incluiu etapas de análise exploratória de dados, configuração do modelo, ajustes de hiperparâmetros e avaliação dos resultados. As principais métricas de avaliação, como MSE, MAE e R^2 , indicaram um bom desempenho do modelo, com visualizações complementares que validam a consistência das previsões realizadas.

Introdução

A previsão de métricas relacionadas ao engajamento de influenciadores digitais tem ganhado destaque no contexto atual de marketing digital. Este projeto utiliza Regressão Linear para modelar a relação entre características como número de seguidores, média de curtidas e taxa de engajamento, com o objetivo de entender como essas variáveis se correlacionam.

O conjunto de dados utilizado foi obtido do **Kaggle**, contendo informações sobre influenciadores do Instagram, como o número de seguidores, média de curtidas por postagem e outras métricas relacionadas. O uso da Regressão Linear justifica-se pela simplicidade e interpretabilidade do algoritmo, sendo ideal para um primeiro experimento com os dados disponíveis.

Metodologia

Análise Exploratória de Dados:

Inicialmente, os dados foram analisados quanto a sua completude e distribuição. As variáveis com valores ausentes foram tratadas utilizando imputação média para variáveis numéricas. Foram geradas visualizações gráficas para identificar relações entre as variáveis e possíveis outliers.

Implementação do Algoritmo:

O modelo base foi construído utilizando a Regressão Linear simples, e versões regularizadas, como Lasso (L1) e Ridge (L2), foram testadas para evitar problemas de multicolinearidade e melhorar a generalização. O conjunto de dados foi dividido em treinamento (80%) e teste (20%), e as variáveis foram normalizadas para facilitar a convergência do modelo.

Validação e Ajuste de Hiperparâmetros:

Para garantir o melhor ajuste do modelo, foi utilizada a validação cruzada com $k=5$. A escolha dos hiperparâmetros, como o valor de alpha para as regularizações, foi feita com base nos resultados da validação.

Resultados

Métricas de Avaliação

O desempenho do modelo no conjunto de teste foi avaliado com as métricas:

- **Mean Squared Error (MSE):** 8.798348747772804e-05
- **Mean Absolute Error (MAE):** 0.006222270049585442
- **Coeficiente de Determinação (R^2):** 0.8611000127940588

Impacto das Variáveis no Modelo

Os coeficientes estimados pelo modelo mostraram que as variáveis mais impactantes na previsão da taxa de engajamento foram:

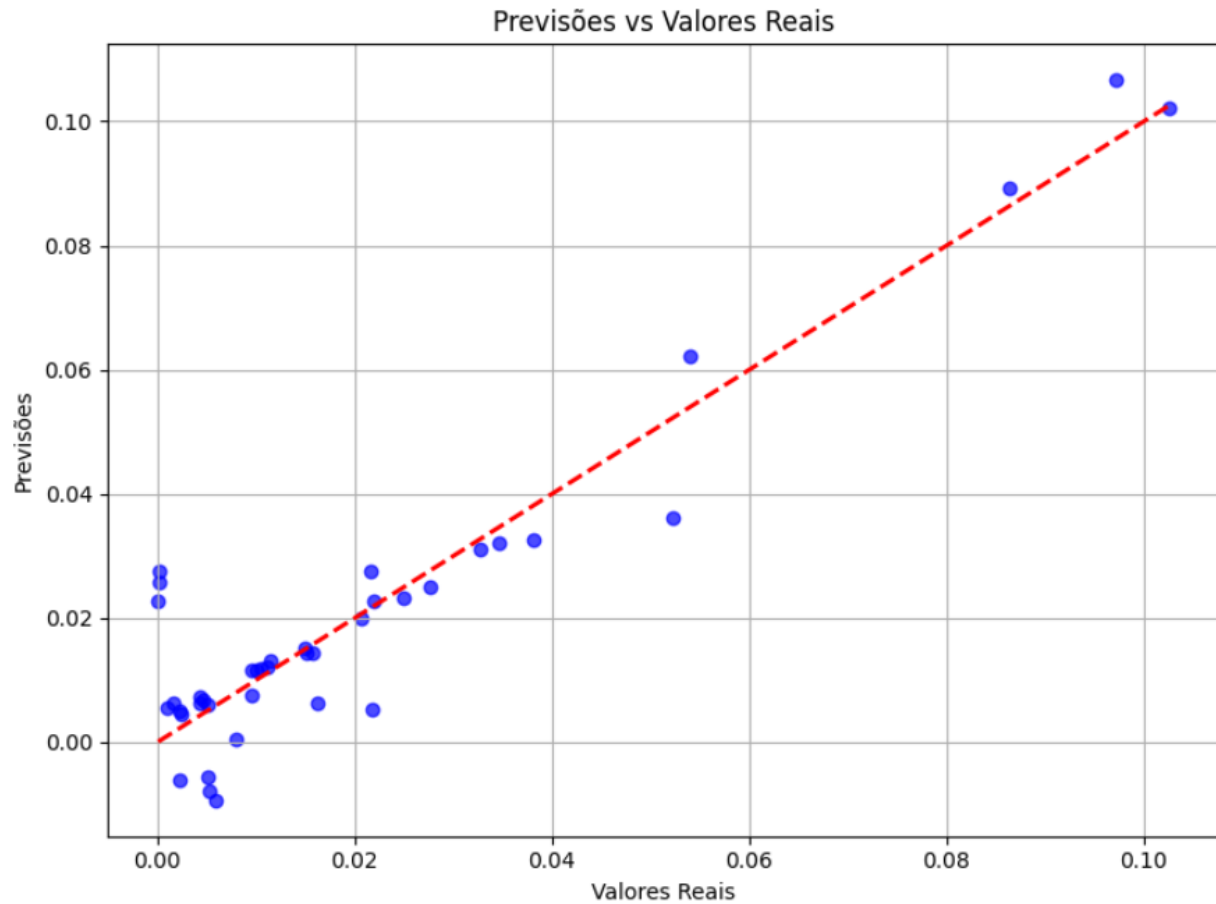
1. Média de Curtidas por Postagem
2. Número de Seguidores

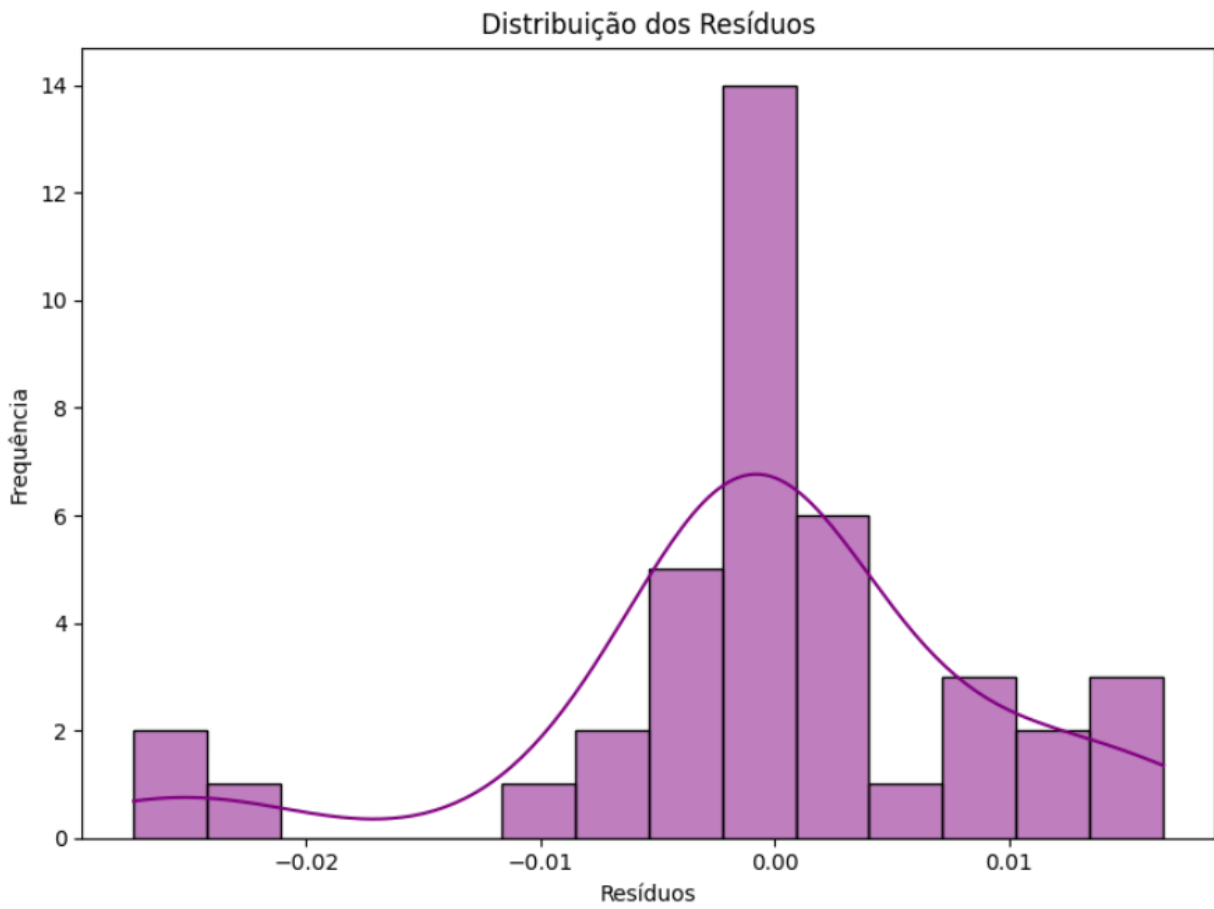
A análise visual dos coeficientes foi feita por meio de gráficos de barras.

Visualizações

Foram gerados os seguintes gráficos para ilustrar os resultados:

Previsões vs Valores Reais: Verifica a proximidade entre os valores previstos e os reais.





Erro Absoluto vs Valores Reais: Identifica regiões onde os erros foram maiores.

Conclusão e Trabalhos Futuros

O projeto demonstrou a aplicabilidade da Regressão Linear para problemas de previsão de engajamento em redes sociais. O uso de regularização e validação cruzada foi essencial para a construção de um modelo robusto.

Como trabalhos futuros, sugere-se explorar algoritmos mais complexos, como redes neurais ou árvores de decisão, e incluir variáveis contextuais para aprimorar a precisão das previsões.

Referências

1. **Scikit-learn Documentation.** Disponível em: <https://scikit-learn.org>
2. **Dataset Kaggle:** Top Instagram Influencers Data.
3. **ReportLab Documentation.** Disponível em: <https://www.reportlab.com>