

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

Alunos: Kevin Borges e Brenda Trindade

Data de Entrega: 17/11/2024

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

Kevin Borges e Brenda Trindade

Data de Entrega: 17/11/2024

## Resumo

Este relatório apresenta a implementação e análise de um modelo de Regressão Linear para prever a taxa de engajamento de influenciadores do Instagram com base em dados históricos. Utilizamos um conjunto de dados que inclui métricas como número de seguidores, média de curtidas e pontuação de influência. O modelo foi ajustado e avaliado com métricas de desempenho, como MSE, MAE e  $R^2$ , além de regularizações Lasso e Ridge. Concluimos que o modelo fornece previsões satisfatórias, com possibilidade de refinamentos futuros.

## Introdução

Com o crescimento das redes sociais, especialmente o Instagram, entender métricas de engajamento tornou-se essencial para criadores de conteúdo e marcas. Este projeto visa construir um modelo preditivo capaz de estimar a taxa de engajamento dos influenciadores nos últimos 60 dias. A escolha pela Regressão Linear se justifica por sua simplicidade, interpretabilidade e eficiência em cenários com variáveis correlacionadas.

Conjunto de Dados: O dataset contém informações como o número de seguidores, posts realizados, curtidas médias por postagem e uma métrica chamada `influence_score`.

## Metodologia

## 1. Análise Exploratória

Uma matriz de correlação foi gerada para identificar as variáveis mais relevantes. Observou-se forte correlação positiva entre a taxa de engajamento (60\_day\_eng\_rate) e variáveis como média de curtidas e influence\_score.

## 2. Implementação do Algoritmo

Pré-processamento:

- Valores ausentes foram tratados removendo registros com a variável dependente nula e preenchendo demais valores com a média.
- Valores abreviados (e.g., "K", "M") foram convertidos em números reais.

Divisão dos Dados:

O conjunto foi dividido em 80% para treino e 20% para teste. As variáveis foram normalizadas utilizando StandardScaler para evitar viés devido a escalas distintas.

Treinamento:

O modelo de Regressão Linear foi ajustado com as variáveis mais correlacionadas.

## 3. Validação e Ajuste de Hiperparâmetros

Foram aplicados métodos de regularização Lasso e Ridge para reduzir possíveis problemas de sobreajuste. A validação cruzada (k=5) foi realizada para medir a consistência do modelo.

## Resultados

### 1. Métricas de Avaliação

Os resultados obtidos com o modelo de Regressão Linear são apresentados abaixo:

- MSE: 0.00002134

- MAE: 0.0037

- $R^2$ : 0.9458

Para os modelos regularizados:

- Lasso:  $R^2 = 0.9402$

- Ridge:  $R^2 = 0.9441$

A validação cruzada revelou  $R^2$  médio de 0.9362, indicando boa generalização.

## 2. Visualizações

Gráficos como matriz de correlação e dispersão entre valores reais e previstos foram gerados para auxiliar na interpretação dos resultados.

## Discussão

Os resultados mostram que o modelo linear capturou bem a relação entre as variáveis independentes e a taxa de engajamento. No entanto, limitações como a ausência de dados categóricos e possíveis vieses nas métricas de entrada podem ter influenciado a precisão. A inclusão de mais dados e o uso de algoritmos mais complexos podem melhorar o desempenho.

## Conclusão e Trabalhos Futuros

O projeto demonstrou a viabilidade do uso de Regressão Linear para prever taxas de engajamento. No futuro, planeja-se:

- Explorar métodos não lineares, como redes neurais.
- Incorporar variáveis categóricas, como região e nicho do influenciador.
- Refinar a coleta e pré-processamento de dados para maior consistência.

## Referências

1. Scikit-learn Documentation. Disponível em: <https://scikit-learn.org>
2. Pandas Documentation. Disponível em: <https://pandas.pydata.org>
3. Seaborn Documentation. Disponível em: <https://seaborn.pydata.org>
4. Artigos sobre Métricas de Engajamento no Instagram.
5. Tutoriais Python para Ciência de Dados.