

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

## Resumo

O projeto visa implementar um modelo preditivo para estimar a taxa de engajamento de influenciadores em diversas regiões do mundo. A previsão será realizada com base em variáveis independentes que estão fortemente correlacionadas com a variável dependente, ou em casos onde as variáveis não apresentam correlação entre si. O objetivo principal é criar um sistema eficiente de previsão, usando técnicas de regressão e análise de dados.

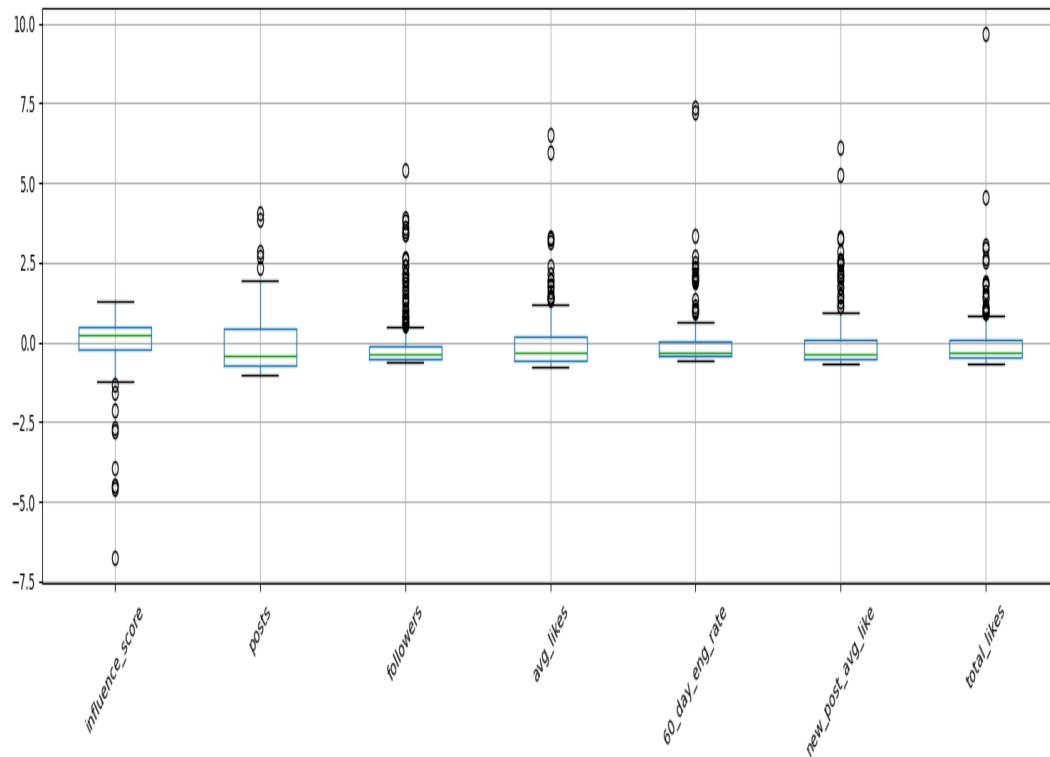
## Metodologia

O projeto utilizou três variações de modelos de regressão linear para prever a taxa de engajamento dos influenciadores: regressão linear simples, Ridge e Lasso. O processo de modelagem incluiu a análise e o pré-processamento dos dados, onde foram realizadas etapas de remoção de outliers para garantir a qualidade dos dados e minimizar a influência de valores extremos no modelo. Além disso, as variáveis independentes foram normalizadas para garantir que todas as features tivessem uma escala similar, o que favorece o desempenho dos modelos de regressão. A seleção dos modelos foi feita com base na eficiência de cada um para lidar com dados altamente correlacionados e na capacidade de regularização de Ridge e Lasso para evitar overfitting.

## Análise de Dados

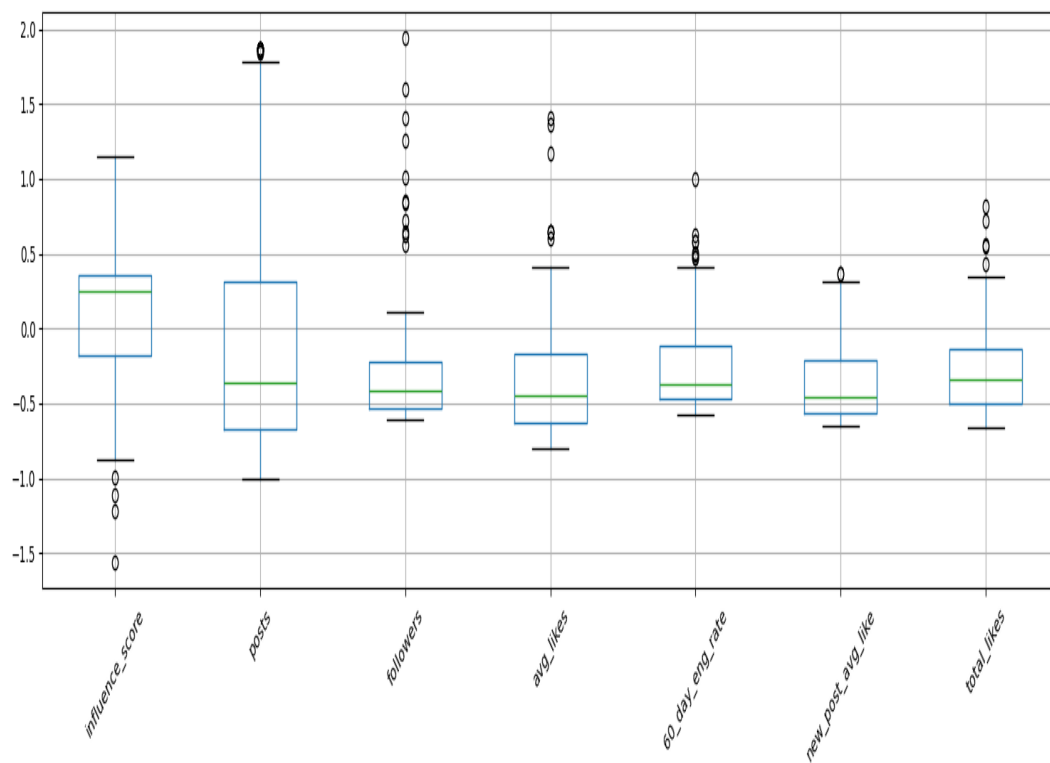
### *Distribuição das Features antes tratamento de Outliers*

Percebe-se uma vasta quantidade de outliers em métricas como followers, avg\_likes, new\_post\_avg\_like e na variável dependente 60\_day\_eng\_rate. Esses outliers devem ser tratados da melhor forma possível para que o modelo seja menos enviesado pelo overfitting.

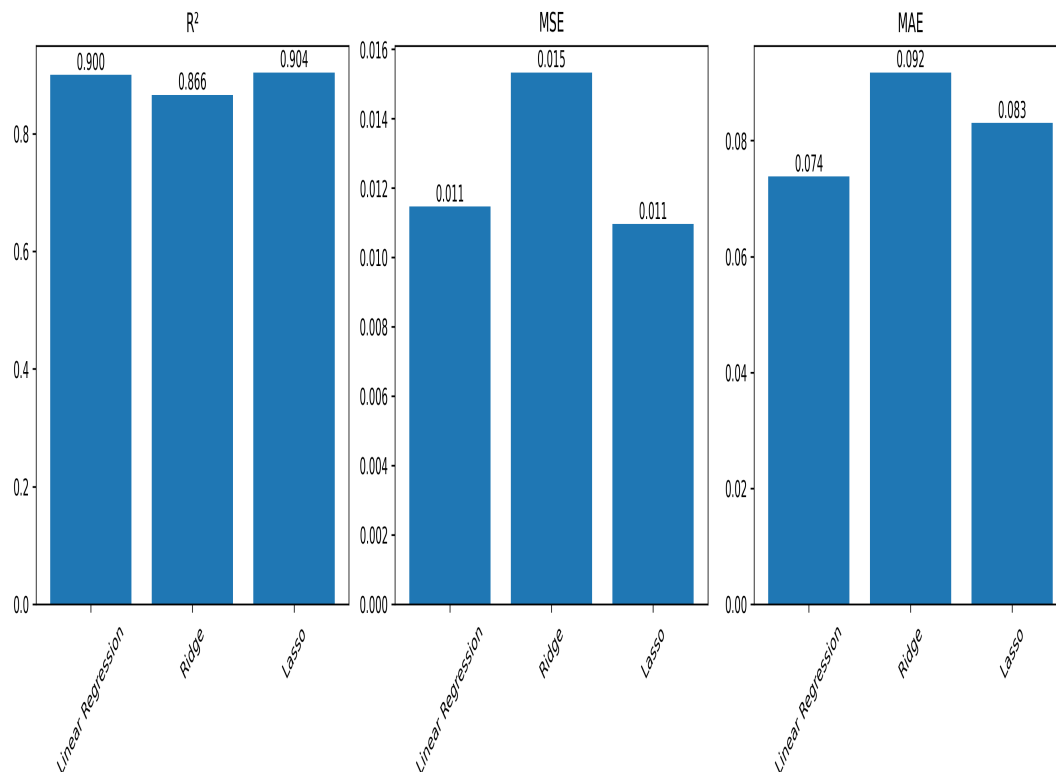


### ***Distribuição das Features após tratamento de Outliers***

Aqui já temos uma redução considerável na quantidade de outliers em todas as variáveis independentes citadas anteriormente, principalmente na new\_post\_avg\_like que apenas sobrou um outlier. Dessa forma, nosso modelo não será tão afetado pelos outliers no data set.



## Resultados



RESULTADOS DOS MODELOS ===== Linear Regression: ----- R²: 0.8996 MSE: 0.0115 MAE: 0.0738 CV Score (média): 0.8326 CV Score (desvio): 0.1446 Coeficientes: influence\_score: 0.0072 posts: -0.0061 followers: -0.2722 avg\_likes: -0.0483 new\_post\_avg\_like: 1.2016 total\_likes: -0.0582  
===== Ridge: ----- R²: 0.8657 MSE: 0.0153 MAE: 0.0918 CV Score (média): 0.8187 CV Score (desvio): 0.0829 Coeficientes: influence\_score: -0.0038 posts: -0.0236 followers: -0.2445 avg\_likes: 0.1458 new\_post\_avg\_like: 0.8073 total\_likes: -0.0303  
===== Lasso: ----- R²: 0.9040 MSE: 0.0110 MAE: 0.0831 CV Score (média): 0.8556 CV Score (desvio): 0.0726 Coeficientes: influence\_score: -0.0000 posts: -0.0310 followers: -0.1999 avg\_likes: 0.0000 new\_post\_avg\_like: 0.8960 total\_likes: -0.0000  
=====

## Conclusão

Com base nos resultados obtidos, o modelo de Regressão Linear demonstrou melhor performance para a predição de taxas de engajamento, apresentando um equilíbrio adequado entre complexidade e acurácia.