

RELATÓRIO TÉCNICO: IMPLEMENTAÇÃO E ANÁLISE DO ALGORITMO DE REGRESSÃO LINEAR

JOSÉ EDUARDO MENDES DA SILVA

TAÍS FARIAS RIBEIRO DE SOUZA

17 DE NOVEMBRO DE 2024

RESUMO

Este relatório expõe o desenvolvimento de um modelo preditor baseado no algoritmo de Regressão Linear para prever a taxa de engajamento de influenciadores no Instagram.

INTRODUÇÃO

Com o crescimento do marketing digital, a previsão da taxa de engajamento de influenciadores é de extrema importância para que as campanhas sejam eficazes. Com isso, este projeto utiliza a Regressão Linear devido à sua simplicidade e interpretabilidade para modelar essa relação. Os dados utilizados foram obtidos através do Top Instagram Influencers Data, que possui informações como número de seguidores, curtidas, comentários e outras métricas.

METODOLOGIA

A primeira etapa do projeto envolveu uma análise exploratória abrangente dos dados. Foram realizadas inspeções visuais e estatísticas para identificar potenciais relações entre variáveis e a taxa de engajamento. Utilizando gráficos de dispersão e histogramas, buscamos evidenciar padrões e correlações. A análise inicial revelou que algumas variáveis, como o número de seguidores e o volume de postagens, tinham uma correlação intuitiva com a taxa de engajamento, embora correlações espúrias também fossem evidentes.

Após a exploração dos dados, a Regressão Linear foi implementada utilizando a biblioteca Scikit-Learn em Python. Diversas configurações foram testadas, como o tratamento de dados ausentes, normalização e padronização de variáveis para assegurar a estabilidade e eficiência do modelo. A função de custo foi minimizada utilizando o método de gradiente descendente, com ajustes na taxa de aprendizado e número de épocas para garantir uma convergência adequada. Além disso, técnicas de regularização, como Lasso (L1) e Ridge (L2), foram aplicadas para evitar overfitting e melhorar a generalização do modelo.

Durante a otimização, priorizou-se a utilização de validação cruzada para avaliar a eficácia do modelo em dados não vistos. Esta técnica permitiu a aferição da robustez do modelo, minimizando os riscos de ajuste excessivo. Além disso, métodos de

seleção de variáveis foram empregados para isolar as mais significativas, com base em análises de correlação e importância dos coeficientes.

RESULTADO

As métricas de avaliação escolhidas para este projeto incluem o coeficiente de determinação (R^2), o Erro Quadrático Médio (MSE) e o Erro Absoluto Médio (MAE).

Os coeficientes obtidos foram interpretados para compreender o impacto de cada variável independente. Por exemplo, um coeficiente positivo associado ao número de seguidores indica que um aumento no número de seguidores tende a elevar a taxa de engajamento, enquanto outras variáveis, como a frequência de postagens, exibiram comportamentos mais complexos. A normalização e técnicas de regularização contribuíram para um modelo mais equilibrado e menos sujeito a variações extremas.

DISCUSSÃO

Os resultados sugerem que o modelo de Regressão Linear tem potencial para prever a taxa de engajamento dos influenciadores com um grau razoável de precisão. No entanto, algumas limitações foram identificadas. O modelo é linear por natureza, o que pode não capturar relações não lineares entre as variáveis. A escolha da taxa de aprendizado e do número de épocas influenciou fortemente a convergência, exigindo ajustes cuidadosos.

CONCLUSÃO E TRABALHOS FUTUROS

Este projeto demonstrou a aplicabilidade da Regressão Linear para prever a taxa de engajamento de influenciadores no Instagram. As principais descobertas incluem a importância da seleção de variáveis e da otimização cuidadosa de parâmetros. Futuros aprimoramentos podem envolver a exploração de modelos não lineares, a inclusão de mais variáveis explicativas, ou o uso de técnicas avançadas de engenharia de características.

REFERÊNCIAS

JHA, Suraj. Top Instagram Influencers Data Cleaned. Disponível em: <https://www.kaggle.com/datasets/surajjha101/top-instagram-influencers-data-cleaned>. Acesso em: 17 nov. 2024.