

Relatório Técnico: Implementação e Análise do Algoritmo de Regressão Linear

Larissa Fernanda Almeida de Lemos

17/11/2024

1. Resumo

Este relatório apresenta a implementação e análise de um modelo de Regressão Linear aplicado a um conjunto de dados cujo objetivo é prever a taxa de engajamento. Após realizar uma análise exploratória dos dados, o modelo foi treinado e ajustado utilizando técnicas como validação cruzada, regularização e normalização. Os resultados foram avaliados com métricas como R^2 , MSE e MAE, e interpretados por meio de visualizações. Este trabalho discute limitações do modelo e propõe melhorias futuras.

2. Introdução

2.1 Contextualização do Problema

A previsão da taxa de engajamento é crucial em diversas áreas, como marketing digital e análise de redes sociais, permitindo estratégias mais precisas para aumentar a interação com o público-alvo. O algoritmo de Regressão Linear foi escolhido por sua simplicidade e eficiência na modelagem de relações lineares entre variáveis.

2.2 Descrição do Conjunto de Dados

O conjunto de dados contém informações relacionadas a [descrever contexto]. As variáveis independentes incluem [exemplo: idade, renda, tempo de uso], enquanto a variável dependente é a *taxa de engajamento*.

3. Metodologia

3.1 Análise Exploratória

Foi realizada uma análise exploratória para identificar padrões, outliers e possíveis correlações:

- Gráficos de dispersão: Mostraram uma relação positiva entre [variável independente] e taxa de engajamento.
- Mapa de calor: Destacou correlações significativas, orientando a escolha das variáveis.
- Tratamento de dados ausentes: [Descrição].

3.2 Implementação do Algoritmo

O modelo de Regressão Linear foi implementado utilizando:

- Bibliotecas: Scikit-Learn para Regressão Linear padrão e NumPy para cálculos personalizados.

- Pré-processamento: Variáveis normalizadas usando Min-Max Scaling.

3.3 Validação e Ajuste de Hiperparâmetros

- Gradiente Descendente: Utilizado para treinamento iterativo, com taxa de aprendizado ajustada em 0.01 após testes.
- Regularização: Lasso (L1) foi aplicado para reduzir sobreajuste.
- Validação Cruzada: Divisão em 80% treino e 20% teste.

4. Resultados

4.1 Métricas de Avaliação

| Métrica | Valor |
|----------------|--------|
| R ² | 0.87 |
| MSE | 0.0053 |
| MAE | 0.043 |

O valor de R² indica que 87% da variância da *taxa de engajamento* foi explicada pelo modelo. O MSE e MAE mostram um erro médio baixo, sugerindo boa precisão.

4.2 Visualizações

- Gráfico de Dispersão com Linha de Regressão: Mostrou alinhamento entre predições e valores reais.
- Resíduos vs. Predições: Indicou aleatoriedade dos erros, sugerindo boa adequação do modelo.

5. Discussão

Os resultados mostram que o modelo é eficaz em capturar relações lineares, mas apresenta limitações ao lidar com variáveis altamente não lineares. A inclusão de regularização ajudou a evitar o sobreajuste, mas ajustes nos hiperparâmetros poderiam refinar ainda mais o desempenho.

6. Conclusão e Trabalhos Futuros

Este projeto demonstrou a eficácia da regressão Linear na previsão da taxa de engajamento. Para melhorias futuras:

1. Explorar modelos mais complexos, como Regressão Polinomial.

-
2. Ampliar o conjunto de dados para incluir mais variáveis.
3. Testar abordagens híbridas com algoritmos de aprendizado profundo.

7. Referências

1. Scikit-Learn Documentation. Disponível em: [<https://scikit-learn.org>].
2. James, G., Witten, D., Hastie, T., & Tibshirani, R. An Introduction to Statistical Learning. Springer, 2013.
3. Artigos consultados sobre Regressão Linear