

Projeto de Rotatividade de Usuários

Resultados da Regressão Logística



Preparado para: Equipe de Liderança doWaze

OVERVIEW

A equipe de dados do Waze está desenvolvendo um projeto que visa aumentar o crescimento geral da empresa, evitando a perda mensal de usuários. Como parte do esforço para aprimorar a retenção, o Waze deseja entender melhor o comportamento dos usuários. Modelos de regressão logística binomial geralmente oferecem flexibilidade e poder preditivo, que podem ser usados para orientar decisões de negócios mais amplas. A equipe construiu um modelo de regressão logística. **Este relatório apresenta detalhes e insights da Etapa 5, os quais impactam o desenvolvimento futuro do projeto como um todo.**

PROJECT STATUS

Etapa 5 - Modelagem de Regressão



Objetivo: Aplicar dados do usuário para construir e analisar um modelo de regressão logística binomial.



Métodos:

- Criada características de interesse para os stakeholders e cenário de negócios
- Avaliada as características quanto à multicolinearidade
- Construído o modelo de regressão
- Avaliado o desempenho do modelo

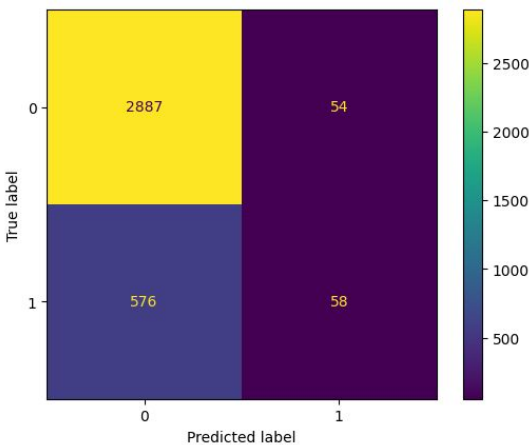


Impacto: Os resultados do modelo podem revelar relações significativas entre variáveis e prever resultados binários, proporcionando orientação para decisões em setores como marketing e desenvolvimento de produtos.

NEXT STEPS

- ➔ Recomendamos usar as principais descobertas do modelo para orientar futuras explorações no projeto.
- ➔ Apesar de não ser adequado para decisões de negócios cruciais, o modelo destaca a necessidade de mais dados relacionados ao cancelamento de usuários e sugere aprimorar o perfil de usuário para impulsionar o crescimento global e evitar cancelamentos mensais no aplicativo.

KEY INSIGHTS



Legenda: 1 = cancelado e 0 = retido

- A eficácia de um modelo de regressão logística binomial é determinada pelos escores de acurácia, precisão e recall; em particular, o recall é essencial, pois mostra o número de usuários cancelados.
- O modelo possui precisão mediana (53% das previsões positivas estão corretas), mas um recall muito baixo, identificando apenas 9% dos usuários cancelados. Isso significa que o modelo faz muitas previsões de falsos negativos e falha em capturar usuários que cancelarão.
- A variável "activity_days" foi a característica mais importante no modelo, apresentando uma correlação negativa com o cancelamento do usuário.
- Na EDA anterior, a taxa de cancelamento do usuário aumentou à medida que os valores em "km_per_driving_day" aumentaram. No modelo, a distância dirigida por dia foi a segunda variável menos importante.