

Relatório do Teste Técnico: Estágio em IA

Objetivo

Este relatório documenta a execução do teste técnico para previsão de compra de casas em um site imobiliário, conforme as especificações do desafio. O objetivo foi construir um modelo de classificação para prever se um usuário realizará a compra de uma casa (variável Compra = 1 ou 0) com base nas seguintes informações:

Idade

Renda Anual (em \$)

Gênero

Tempo no Site (min)

Anúncio Clicado

Estatísticas Descritivas

Métrica, Idade, Renda Anual (em \$), Tempo no Site (min), Compra (0 ou 1)

Média, 38.51, 58,253.97, 17.35, 0.33

Desvio Padrão, 12.68, 25,612.06, 7.72, 0.47

Mínimo, 18.00, 30,000.00, -1.00, 0.00

Máximo, 59.00, 100,000.00, 29.85, 1.00

Nota: Foi identificado um valor inválido na coluna Tempo no Site (min) com valor negativo.

Etapas Realizadas

1. Análise Exploratória dos Dados

Foram identificados valores ausentes em colunas como Idade, Renda Anual (em \$) e Anúncio Clicado.

Algumas colunas necessitavam padronização e tratamento para uso em modelos de classificação.

A variável-alvo Compra estava desbalanceada, com uma proporção maior de exemplos da classe 0 (não comprou).

2. Tratamento e Preparação dos Dados

Valores ausentes foram tratados da seguinte forma:

Idade e Renda Anual (em \$): Substituídos pela média da coluna.

Anúncio Clicado: Substituído pelo valor mais frequente (Sim ou Não).

Variáveis categóricas foram codificadas:

Gênero: Feminino = 0, Masculino = 1.

Anúncio Clicado: Sim = 1, Não = 0.

Foi realizada a normalização das variáveis numéricas para escalas comparáveis.

O dataset foi dividido em conjuntos de treino (80%) e teste (20%).

3. Construção do Modelo de Classificação

Modelo Utilizado: Regressão Logística

Métricas de Avaliação:

Precisão, Recall, F1-Score e Acurácia.

Matriz de confusão para avaliar erros de classificação.

4. Avaliação do Modelo

Métrica, Classe 0 (Não Comprou), Classe 1 (Comprou)

Precisão, 63%, 60%

Recall, 92%, 19%

F1-Score, 75%, 29%

Acurácia Geral, → 62%,

O modelo mostrou bom desempenho para a classe 0, mas dificuldade em identificar compradores (classe 1).

A limitação foi atribuída ao desbalanceamento de classes.

Sugestões de Melhoria

Balanceamento de Classes:

Utilizar técnicas como SMOTE ou undersampling para equilibrar as proporções de classes.

Testar Outros Modelos:

Implementar Random Forest ou Gradient Boosting para comparar os resultados.

Validação Cruzada:

Usar k-fold cross-validation para uma avaliação mais robusta.

Exploração de Hiperparâmetros:

Ajustar hiperparâmetros do modelo para melhorar o desempenho.

Feature Engineering:

Criar novas variáveis baseadas nas relações existentes entre as colunas.

Conclusão

O desafio foi concluído. O modelo implementado apresentou acurácia geral de 62%, com bom desempenho para identificar não-compradores (classe 0), mas dificuldade em classificar compradores (classe 1).