

# Relatório de Resultados

Este relatório apresenta os resultados obtidos durante a construção e avaliação de um modelo de classificação aplicado a um conjunto de dados fictício. O objetivo foi prever se um cliente realizaria ou não uma compra, com base em variáveis como idade, gênero, renda anual e tempo gasto no site. O processo envolveu as etapas de pré-processamento, treinamento do modelo, avaliação e interpretação dos resultados.

## 1. Introdução

### Resultados Obtidos - Acurácia do Modelo

O modelo treinado foi uma Árvore de Decisão, que alcançou uma acurácia satisfatória nos dados de teste. Esse resultado indica que o modelo conseguiu classificar corretamente a maioria dos casos, mas ainda pode haver espaço para melhorias.

Neste projeto, foi desenvolvido um modelo de classificação utilizando uma Árvore de Decisão para prever a probabilidade de um cliente realizar uma compra ("Compra (0 ou 1)"). Os dados foram previamente processados, normalizados e divididos em conjuntos de treino e teste. Este relatório apresenta os principais resultados e interpretações obtidos.

### Métricas de Desempenho

Além da acurácia, foram analisadas outras métricas importantes, como:

- Matriz de confusão:** revelou o número de predições corretas e incorretas em cada classe (compra ou não compra).
- Relatório de classificação:** forneceu informações sobre precisão, revocação e F1-score (desempenho preditivo), ajudando a entender como o modelo performou em detalhes.

### Importância das Variáveis

Uma análise das variáveis mostrou quais foram mais influentes na tomada de decisão do modelo. Entre elas, destacaram-se:

- Tempo no Site (min):** Foi a variável mais relevante, sugerindo que clientes que passam mais tempo no site tendem a realizar compras.
- Renda Anual (em \$):** Outra variável significativa, mostrando que a renda também desempenha um papel importante na tomada de decisão.

Os resultados foram visualizados por meio de um gráfico de barras através da aplicação do Jupyter Notebook, que facilitou a interpretação da contribuição de cada variável.

---

## 2. Dados e Métodos

### Propostas de Melhorias

Apesar dos bons resultados iniciais, algumas ações poderiam ser implementadas para refinar o modelo:

1. **Testar outros modelos de classificação:** Comparar o desempenho com algoritmos como Regressão Logística ou Random Forest.
2. **Validação cruzada (K-Fold):** Garantir que os resultados sejam consistentes em diferentes divisões do conjunto de dados.
3. **Ajuste de hiperparâmetros:** Melhorar o desempenho da Árvore de Decisão ajustando parâmetros como profundidade máxima ou critério de divisão ("grid search" ou "random search"), para reduzir overfitting.

- **Variáveis Utilizadas:**

- Idade
- Gênero
- Renda Anual (em \$)
- Tempo no Site (min)
- Anúncio Clicado

- **Estratégias de Processamento:**

- Codificação de variáveis categóricas usando LabelEncoder.
- Normalização das variáveis numéricas utilizando o StandardScaler.

- **Modelo de Classificação:**

- Algoritmo: Árvore de Decisão.
- Divisão de dados: Cerca de 80% para treino e 20% para teste.

---

## 3. Conclusão

A construção deste modelo proporcionou uma experiência prática valiosa na aplicação de técnicas de aprendizado de máquina. Os resultados foram promissores, e as melhorias propostas poderão tornar o modelo ainda mais eficaz. Estou confiante de que, com mais exploração, é possível extrair ainda mais insights dos dados e aprimorar as previsões. Agradeço pela oportunidade de desenvolver este projeto e fico à disposição para discutir os resultados ou implementar as melhorias sugeridas.