

Relatório Técnico - Predição de Valor de Venda no Varejo com Random Forest

1. Objetivo do Projeto

Desenvolver um sistema de predição de valor de vendas para uma rede varejista com base em dados históricos, utilizando o algoritmo Random Forest. A previsão de vendas é uma ferramenta essencial para planejamento de estoque, campanhas de marketing e tomada de decisões estratégicas.

2. Conjunto de Dados

O dataset utilizado inclui:

- ID do pedido
- Data do pedido
- Segmento do cliente
- Localização (país, estado, cidade)
- Categoria e subcategoria do produto
- Valor da venda

3. Metodologia

3.1 Análise Exploratória de Dados (EDA)

Foram realizadas as seguintes atividades:

- Verificação de valores nulos e tipos de dados
- Análise descritiva e distribuição do valor de venda
- Análise temporal das vendas por mês
- Exploração de frequências de categorias e localizações

3.2 Engenharia de Features

- Extração de ano e mês da data do pedido
- Conversão de variáveis categóricas em indicadores via OneHotEncoder
- Separação dos dados em treino (80%) e teste (20%)

3.3 Algoritmo Escolhido

O algoritmo selecionado foi **Random Forest Regressor**, por sua robustez, capacidade de lidar com dados heterogêneos e boa performance em tarefas de regressão com dados estruturados.

4. Avaliação do Modelo

As métricas obtidas no conjunto de teste foram:

- **MSE (Erro Quadrático Médio):** baixo
- **MAE (Erro Absoluto Médio):** consistente com a dispersão

- **R² (Coeficiente de Determinação):** bom ajuste dos dados
- **Validação cruzada:** R² médio estável entre os folds

5. Importância das Variáveis

As features mais relevantes identificadas foram:

- Subcategoria do produto
- Estado e cidade
- Mês do pedido (indicando possível sazonalidade)
- Segmento de cliente

6. Conclusões e Recomendações

- O modelo se mostrou eficaz para prever o valor de venda com base nos atributos fornecidos.
- Identificou-se padrões temporais que podem ser explorados em campanhas sazonais.
- A granularidade geográfica (estado/cidade) e o tipo de produto são fortemente correlacionados com o volume de vendas.

7. Limitações e Trabalhos Futuros

- O modelo não considera variáveis externas (feriados, preços, promoções).
- Futuras versões podem explorar modelos como XGBoost, LightGBM e redes neurais.
- Inclusão de features econômicas (inflação, PIB) pode melhorar previsão.

8. Referências

- Scikit-learn documentation
- Artigos sobre previsão de demanda no varejo

Autor: [Seu Nome Aqui]

Data: [Inserir data do projeto]