

Análise Preditiva com Árvore de Decisão em Machine Learning

Arthur Romani e Renan Manera - CC5N

19 de maio de 2025

Introdução

Este trabalho tem como objetivo aplicar técnicas de Machine Learning com foco na estrutura de **Árvore de Decisão**, um modelo baseado na estrutura de árvore binária amplamente utilizado em classificação e regressão. A análise se baseia nos dados históricos das ações PETR3 e PETR4 da Petrobras entre 2015 e 2016.

Contexto e Fundamentação

As **árvores de decisão** são estruturas hierárquicas binárias que permitem classificar ou prever valores com base em condições sequenciais. Cada nó interno representa uma condição (divisão), cada ramo representa o resultado da condição, e cada folha representa uma predição ou classe final.

Aspectos da Árvore de Decisão

Nesta aplicação, observamos:

1. **Estrutura da árvore binária:** composta por nós com até dois filhos, representando decisões binárias baseadas nas variáveis dos dados.
2. **Profundidade dos nós:** cada nó tem uma profundidade que indica sua posição na hierarquia. As folhas estão nos níveis mais profundos e são os nós terminais.
3. **Nós modificados:** foram utilizados dados de PETR3 e PETR4, e as variáveis de entrada alteraram a estrutura da árvore gerada dinamicamente.
4. **Folhas alcançadas:** para cada amostra de teste, a árvore chega a uma folha específica, que representa a predição para aquele conjunto de características.
5. **Regras de divisão:** a cada nó, o algoritmo utiliza um critério estatístico (ex: redução de variância) para definir a melhor divisão. Isso se relaciona diretamente à estatística e à correlação entre as variáveis.
6. **Caminhos compartilhados:** amostras com características semelhantes tendem a seguir o mesmo caminho pela árvore até uma folha comum, representando similaridade nos dados.

Modelo Testado

Utilizamos o `RandomForestRegressor`, que combina várias árvores de decisão. Apesar de sua sofisticação, o modelo apresentou desempenho inferior:

- **RMSE:** 1,4830
- **R²:** -1,9215

Isso sugere dificuldades na generalização dos dados de série temporal sem tratamento específico para dependência sequencial.

Visualização da Predição

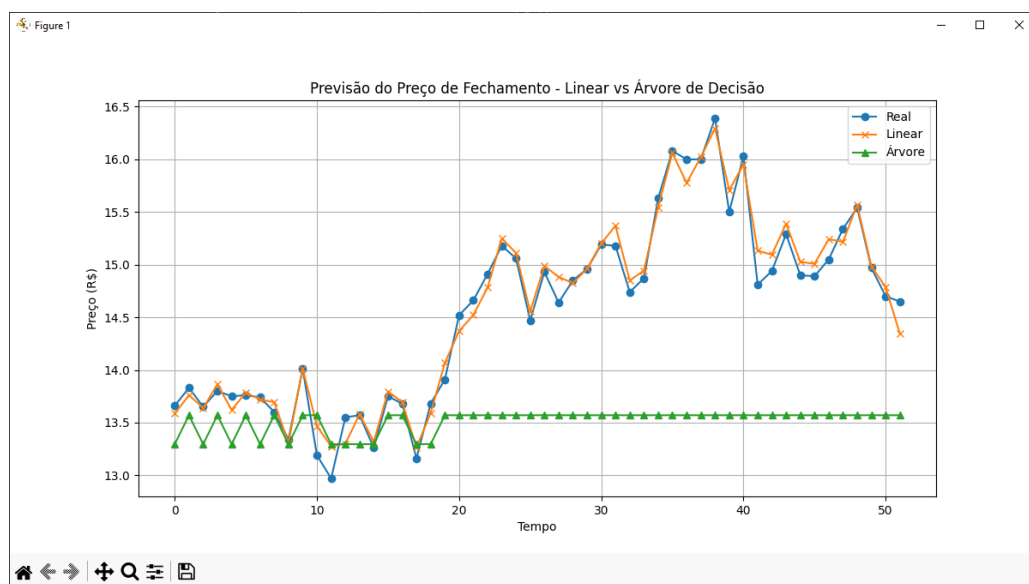


Figura 1: Resultado da previsão com modelo Random Forest.

Conclusão

As árvores de decisão são ferramentas fundamentais em Machine Learning, principalmente por sua interpretabilidade e relação direta com a estrutura de árvores binárias estudadas em Estrutura de Dados. Apesar dos resultados limitados do modelo Random Forest neste caso, a análise da estrutura da árvore permite compreender melhor a relação entre as variáveis e a previsão.