

RELATÓRIO DE AVALIAÇÃO DO MODELO BASELINE DE MACHINE LEARNING

1. Introdução

Este relatório apresenta o desenvolvimento e a avaliação de um modelo baseline de Regressão Linear. O objetivo principal é criar uma primeira referência de desempenho, permitindo comparações com modelos mais avançados nas próximas etapas.

2. Preparação dos Dados

Os dados foram divididos em três conjuntos:

- 60% para treino
- 20% para validação
- 20% para teste

Essa divisão ajuda a garantir que o modelo seja avaliado de forma justa e não sofra influência excessiva de um único conjunto.

3. Treinamento do Modelo

Um modelo de Regressão Linear foi utilizado como baseline. Esse tipo de modelo é simples, interpretável e adequado para iniciar o processo de modelagem. Ele busca encontrar uma relação linear entre as variáveis independentes e o alvo.

4. Métricas de Avaliação

As seguintes métricas foram calculadas:

- MSE (Mean Squared Error): mede o erro quadrático médio.
- RMSE: raiz do erro quadrático médio, interpretável na unidade do alvo.
- MAE (Mean Absolute Error): mede o erro médio absoluto.
- R^2 : indica quanto da variação do alvo é explicada pelo modelo.

Interpretação:

- MSE e RMSE maiores indicam erros mais elevados.
- MAE mostra o erro médio direto.
- R^2 varia entre 0 e 1; quanto mais próximo de 1, melhor o modelo.

5. Análise de Overfitting

O desempenho foi comparado entre treino e validação. Caso o modelo apresente métricas muito melhores no treino do que na validação, isso indica overfitting. No baseline, espera-se que o modelo seja simples o suficiente para evitar esse problema.

6. Análise de Resíduos

Os resíduos foram observados para avaliar se seguem um comportamento aleatório. A presença de padrões pode indicar problemas como:

- Heterocedasticidade
- Linearidade inadequada
- Outliers influentes

No baseline, ainda é comum encontrar padrões, reforçando a necessidade de modelos futuros mais robustos.

7. Feature Importance

As três variáveis mais importantes foram identificadas com base na magnitude dos coeficientes da Regressão Linear. Valores mais altos indicam maior impacto no resultado previsto.

8. Conclusões

O modelo baseline fornece uma referência inicial de desempenho. Ele é simples e rápido de treinar, mas possui limitações:

- Pode não capturar relações não-lineares
- Pode ser sensível a outliers
- Pode sofrer com multicolinearidade

Ainda assim, cumpre o papel de ponto de partida para melhorias.

9. Próximos Passos (Etapa 4)

- Testar modelos mais complexos (Árvore, Random Forest, XGBoost)
- Realizar tuning de hiperparâmetros
- Criar gráficos mais avançados de avaliação
- Reforçar storytelling focado no impacto dos resultados

RELATÓRIO AMPLIADO – FIM