

# IMDb - Relatório Final das Análises

## Resultados e Escolha do Modelo

### Random Forest

Root Mean Squared Error (RMSE): 0.000254

$R^2$  (Coeficiente de Determinação): 0.999999

#### Interpretação dos Resultados

O  $R^2$  Score de 0.999999 indica que o modelo é capaz de explicar 99.9999% da variabilidade dos dados de resposta em torno da média. Este é um indicador extremamente alto, sugerindo que o modelo captura quase todas as variações nos dados.

O RMSE de 0.000254 é muito baixo, indicando que as previsões do modelo estão extremamente próximas dos valores reais. Um RMSE baixo é desejável, pois indica que o modelo comete poucos erros ao fazer previsões.

#### Considerações sobre o Overfitting

Os resultados excepcionalmente altos, tanto do  $R^2$  Score quanto do RMSE, sugerem que o modelo pode estar super ajustado aos dados de treinamento. Overfitting ocorre quando o modelo aprende não apenas os padrões subjacentes nos dados, mas também o ruído. Isso resulta em desempenho impecável nos dados de treinamento, mas pode levar a um desempenho inferior em novos dados não vistos.

Indicadores de Overfitting:

- Alto  $R^2$  Score nos dados de treinamento.
- Baixo RMSE nos dados de treinamento.

# IMDb - Relatório Final das Análises

## Regressão Linear

Root Mean Squared Error (RMSE): 0.113616

$R^2$  (Coeficiente de Determinação): 0.241217

### Interpretação dos Resultados

O  $R^2$  Score de 0.241217 indica que o modelo de Regressão Linear é capaz de explicar apenas 24.1217% da variabilidade dos dados de resposta em torno da média. Embora este valor não seja alto, ele ainda fornece uma visão útil sobre as relações entre as variáveis.

O RMSE de 0.113616 é moderado, indicando que as previsões do modelo têm uma margem de erro razoável. Embora maior do que o RMSE do modelo Random Forest, ainda está dentro de um limite aceitável para muitas aplicações.

## Conclusão

Após avaliar os dois modelos, com base nas métricas RMSE e  $R^2$ , constatamos que o modelo de Regressão Linear apresentou um desempenho mais robusto e consistente. Este modelo não apenas se aproximou significativamente dos resultados do site oficial do IMDb, mas também mostrou menos tendência a overfitting, tornando suas previsões mais confiáveis para novos dados.

Em contrapartida, o modelo Random Forest Regressor, apesar de suas métricas excepcionalmente altas, apresentou indícios de super ajuste aos dados de treinamento, o que pode prejudicar seu desempenho em dados não vistos. Além disso, enfrentou problemas na etapa de predição, que impactaram negativamente sua eficácia.

Portanto, a Regressão Linear se mostrou a escolha mais adequada para este contexto.