

Relatório Análise de Dados Cinematográficos - PProductions

Autor(a): Isabela Araujo Costa

1. INTRODUÇÃO

O relatório tem como propósito apresentar o projeto desenvolvido para o desafio de Ciência de Dados, onde uma breve situação fictícia é demonstrada, apresentar dados revelantes para um estúdio da Hollywood para uma tomada de decisão. Todavia, o projeto busca mostrar reais habilidades de análise detalhada. Dessa forma, considerando o máximo de fatores possíveis, a partir de um conjunto de dados cinematográficos, fornecido para a realização do desafio, foram aplicados limpeza de dados, transformações e desenvolvimento para padronizar variáveis como faturamento, duração e ano de lançamento. Dessa forma, possibilitando uma análise mais confiável. A plataforma foi desenvolvida em Streamlit, que integra análises exploratórios, modelagem preditiva e visualizações dinâmicas

2. ANÁLISE

O projeto teve como objetivo realizar análise detalhadas. Deste modo, durante a construção do código em Python, foi imprescindível aplicar filtros: interativos de ano, gênero, certificação e diretor. Assim, permite uma visão segmentada da base, o dashboard exibe métricas de alto nível como nota média do IMDB, bilheteria total, diretor mais frequente e o filme mais popular, além de gráficos sobre distribuição de notas, relação entre duração e avaliação e evolução das médias por ano, permitindo que analistas e executivos explorem de forma intuitiva informações mais aprofundadas sobre os filmes para identificar padrões de sucesso comercial e crítico. A análise textual da coluna Overview utilizou nuvens de palavras e frequências de termos por gênero, revelando padrões narrativos e temáticos que diferenciam produções de drama, ação e comédia, oferecendo subsídios criativos para roteiristas e produtores.

3. MODELAGEM

Na etapa de modelagem, foi desenvolvido um modelo de regressão baseado em Random Forest para prever notas do IMDB. O modelo considerou variáveis como duração, metaspore, número de votos, faturamento, elenco, diretor e gênero, alcançando bons resultados em métricas como MAE e R^2 . A análise de importância das variáveis mostrou que metaspore, votos e elenco são fatores determinantes no desempenho crítico.

4. CONCLUSÃO

Em suma, a utilização de Machine Learning pode ajudar a empresa a se adaptar de acordo como a situação atual do mercado, compreendendo suas tendências e prever mudanças futuras. Essas aplicações, podem ajudar empresas que possuem o mesmo intuito da "PProductions", extrair informações essenciais para realizar planejamentos e poder gerenciar suas operações sem riscos, fornecendo vantagens para um futuro mais estável. O estudo de caso de The Shawshank Redemption demonstra robustez

do modelo. A previsão da nota aproximou-se bastante da avaliação real, confirmando que atributos como elenco estrelado, direção reconhecida e narrativa dramática estão fortemente associados ao sucesso crítico. Além disso, o dashboard inclui seções que respondem diretamente às perguntas apresentadas no documento: recomendação de filmes para público desconhecido, identificação dos fatores ligados à bilheteria, exploração de insights do Overview e previsão de notas. Assim, o projeto entrega não apenas uma análise estática. Porém, visa aprofundar em assuntos fundamentados em evidências, capaz de orientar decisões de investimento milionárias no setor cinematográfico.