

Nanodegree Engenheiro de Machine Learning

Proposta de projeto final

Histórico do assunto ¶

Na era onde o lucro se torna o fator mais importante de um produto, nas grandes obras cinematográficas cujos orçamentos são milionários, o lucro ou prejuízo de um filme na maioria das vezes determinam se a franquia vai ter uma continuação ou não. Quais são os fatores mais importantes para o sucesso de um filme? Se o filme é de ação, se o ator X é o protagonista ou se o diretor Y está dirigindo o filme. Pensando nisso como podemos utilizar a ciência de dados para prever um possível sucesso ou fracasso financeiro de um filme antes do mesmo ser produzido.

A inspiração deste projeto é pessoal, pois sou um aficionado por filmes e presto muito atenção qual é o diretor, atores que farão parte do elenco do filme e se a franquia está dando lucros para as produtoras, pois assim teremos mais filmes da franquia.

Descrição do problema

Será que é possível prever o sucesso de um filme se deve apenas do orçamento gasto, o gênero ou o elenco? As críticas de por parte dos seus espectadores são um fator importante para este sucesso? Após a identificação desses fatores é possível prever com base em novas características de filme qual será a receita do mesmo?

Conjuntos de dados e entradas

Para o este projeto será utilizada uma base de dados disponível do site Kaggle com 45 Mil filmes possuindo as informações de orçamento, elenco, receita, gênero do filme e rating dos filmes pelos usuários do site TMBD e GroupLens estando disponível este conjunto de dados em <https://www.kaggle.com/rounakbanik/the-movies-dataset> (<https://www.kaggle.com/rounakbanik/the-movies-dataset>).

Descrição da solução

Para entender os fatores de sucesso de um filme, será realizada uma análise de quais são as variáveis mais importantes ao sucesso do filme e ao final do projeto teremos as respostas se o elenco influencia mais o sucesso, o orçamento é mais importante ou o gênero é crucial para isto. Ao final do projeto poderemos prever com base nos filmes já lançados se um novo filme que será produzido terá o resultado esperado ou não.

Modelo de referência (benchmark)

Será relacionados os resultados encontrados com notícias sobre o resultado financeiro de filmes que não constam na base de dados, como por exemplo, se os filmes com o ator estão sendo bem recebida por público e consequentemente uma boa receita. Posteriormente será introduzido dados de filmes recentes no modelo e analisado se o resultado encontrado é condizente com a realidade, como por exemplo se o novo filme está com a receita conforme o modelo previa.

Métricas de avaliação

Para avaliarmos se o modelo está adequado será comparado o resultado obtido da correlação dos dados com notícias relacionadas existentes que comprovem os dados apresentados e se esta o modelo está prevendo com uma boa precisão a bilheteria de novos filmes com os resultados divulgados destes filmes.

Design do projeto

Inicialmente será realizada uma análise exploratória dos dados para entender quais são tipos de dados disponíveis e um tratamento de dados para obter novos dados, por exemplo, o elenco do filme está todo relacionado em campo da base de dados, devendo ser extraído o nome dos primeiros atores do elenco deste campo. Após o tratamento destes dados será realizado um estudo de quais são os atributos mais correlacionados com a receita do filme e posteriormente a geração de um modelo que utilize estes atributos

In []: