

# Artigo aí de Machine Learning

YVENS REBOUÇAS

Universidade de Fortaleza

yvensre@gmail.com

## I. INTRODUÇÃO

A indústria cinematográfica é um dos setores de entretenimento mais populares e lucrativos do mundo. Segundo dados da *Motion Picture Association of America* (MPAA), a receita de bilheteria nos Estados Unidos aproxima-se da cifra de US\$ 9.5 bilhões no ano de 2006 [?]. No entanto, a produção de um curta ou longa metragem requer valores altos de investimento inicial. Um filme pertencente a uma companhia integrante da MPAA, por exemplo, custa em média US\$ 100.3 milhões, incluindo os custos de produção, comercialização e *marketing* [?]. Contudo, como toda atividade de empresa, a produção de filmes é uma atividade de risco, tendo em vista a relativa imprevisibilidade da preferência do público.

Este trabalho procura analisar os perfis de diferentes filmes, suas características e avaliações de usuários, para identificar possíveis padrões e informações que possivelmente caracterizam se um filme será um sucesso ou um fracasso. Tais informações minimizariam o risco na produção de filmes e auxiliariam diretores e produtores nesse processo, guiando-os por parâmetros que aumentariam significativamente as chances de sucesso no empreendimento.

A maioria das pesquisas nessa área busca analisar as críticas e *reviews*, classificando-as como positivas ou negativas, para compreender a opinião do público quanto ao filme [?, ?, ?]. Outra abordagem comum são os sistemas de recomendação que tentam, através das críticas e *reviews* prévias de um determinado usuário, identificar possíveis filmes recomendáveis [?]. Trabalhos com o foco na recepção do público aos filmes são raros e escassos.

Partindo dos dados coletados na *Internet Movies Database* (IMDb), nós coletamos diferentes variáveis (gênero, duração, faixa etária, etc.) e, a partir destas e de algoritmos de *machine learning*, classificá-los como "Bom", "Regular" ou "Ruim". As classificações foram definidas a partir das avaliações médias dos usuários (*rating* no site). Utilizando árvores de decisão conseguimos uma média de 66% de acerto, indicando possíveis padrões e variáveis a serem mais profundamente estudadas.

## II. ESTADO DA ARTE

Conceitos Básicos.

- Árvore de decisão.
- Outros algoritmos.

Falar dos trabalhos relacionados.

## III. SOLUÇÃO PROPOSTA

Solução conceitual

- Instanciação.
- Falar sobre a evolução do modelo.

## IV. RESULTADOS DA AVALIAÇÃO

### I. Prova de Conceito

Metodologia.

- Resultados.
- Avaliação.
- Limitações.

### II. Discussão

Discussão dos resultados.

Gráficos.

## V. CONCLUSÃO

Conclusão geral do trabalho.

## REFERENCES