

Proposta para projeto prático

Gabriel D’Luca Souza Viana¹, Vitor Antônio de Lima Silva²

*Universidade Federal de Pernambuco – Centro de Informática
Caixa Postal 7.851 – 50.732-970 – Recife, PE – Brasil
gdsu@cin.ufpe.br, vals@cin.ufpe.br*

Abstract

Um dos mais tradicionais websites para agregação de críticas, o Metacritic sintetiza avaliações de publicações reputadas sobre um determinado filme, álbum, jogo eletrônico ou série de televisão. Desde 1999, o site é responsável por realizar uma agregação de notas e converter para uma escala de 0 à 100 pontos, reunindo a nota atribuída por cada crítico em uma única avaliação ponderada global, denominada Metascore. Os pesos diferem de acordo com o crítico e o tipo de produto, sendo classificados conforme o detalhamento, articulação e concisão de escrita. Dessa forma, algumas publicações possuem maior influência no escore global do que outras. Nossa proposta para o projeto final da disciplina é conduzir análises e estudos no Metacritic, de forma a verificar o cenário atual de resenhas de filmes a partir das avaliações do The New York Times.

1. Coleta

Pretendemos realizar a coleta das pouco mais de 12,000 críticas feitas pelo The New York Times, a partir da página da publicação no Metacritic. Coletaremos: nome do filme, metascore, nome do crítico, nota do crítico, texto
5 da resenha e link com os detalhes. Em seguida, partiremos para a coleta dos dados de cada filme avaliado através da página de detalhes do filme. Desta, extraíremos: nome do filme, índice de aclamação, escore dos usuários, gêneros,

¹Estudante de graduação do curso bacharelado em Sistemas de Informação

²Estudante de graduação do curso bacharelado em Sistemas de Informação

diretor, idiomas e data de lançamento.

2. Pré-processamento

10 Converteremos os elementos coletados de cada página do site para seu tipo de atributo correspondente. Trataremos dados ausentes, já que existem filmes avaliados pelo The New York Times que não possuem Metascore ou User score. Por fim, também será necessário utilizar técnicas de discretização para agrupar a data de lançamento do filme por quinquênio e classificar avaliações como
15 positivas, mistas ou negativas. Isso feito, realizaremos a integração dos dados.

3. Análise exploratória

Conduziremos análises de estatísticas descritivas para obter maior conhecimento sobre os dados que estaremos trabalhando. A página da publicação do The New York Times no Metacritic também dispõe de algumas informações e
20 parte de nossa análise consiste em verificar se estas são condizentes. Iremos elaborar visualizações e utilizaremos modelos de regressão para analisar algumas perguntas que gostaríamos de responder, entre elas:

- Quão forte é a influência no Metascore por parte da nota atribuída pelos críticos do The New York Times?
- 25 • Quais gêneros costumam ser atribuídos a maiores e menores índices de aclamação?
- O quanto o Metascore influencia no score global dos usuários?
- Existe alguma associação entre as demais características (idiomas, classificação, tempo de duração, ...) do filme e o Metascore?
- 30 • Dentre os críticos do The New York Times, quais atribuem um maior número de notas positivas, mistas e negativas em suas avaliações?

Por fim, também tentaremos utilizar técnicas de classificação de texto com o Scikit-learn para prever o Metascore do filme através da sinopse da resenha.