SENTIMENT ANALYSIS OF IMDB MOVIE REVIEWS

A UVA Data Science Case Study by Alexa Owen

[1]

Reviews, most of us don't write them, but most of us read them. Whether it is for a product, book, movie, or something else we often read reviews before wasting money and time. However, it is sometimes difficult to tell if a review is meant to be positive or negative. This is where you come in.

Prompt: IMDB has come to the UVA School of Data Science and asked students to analyze their data for movies that were released in 2021. They are confused why some reviews are bad, but the same reviewer gives a high rating, or the opposite. IMDB wants to make sure its users are getting the most out of the reviews others provide.

Sounds easy enough.

The data they have provided you contains the following [2]:

1. An index column
2. The written movie review
3. A numerical rating
4. The author of the review
5. The title of the review

Deliverable: IMDB has asked you to perform sentiment analysis on their movie reviews from 2021. The type of sentiment analysis completed is up to the analyst and the only part of the data set IMDB has requested to be included in the analysis was the written movie review. IMDB would like you to present a recap of your project and your results to them in a presentation format. Furthermore, they would like you to upload the materials used in your project to a GitHub site for easy access.

[1] "File:IMDB logo 2016.SVG," Wikimedia Commons. [Online]. Available: https://commons.wikimedia.org/wiki/File:IMDB_Logo_2016.svg. [Accessed: 28-Apr-2023].
[2] D. Deshpande, "IMDB movie reviews 2021," Kaggle, 31-May-2021. [Online]. Available: https://www.kaggle.com/datasets/darshan1504/imdb-movie-reviews-2021?select=IMDB_Movies_2021.db. [Accessed: 28-Apr-2023].