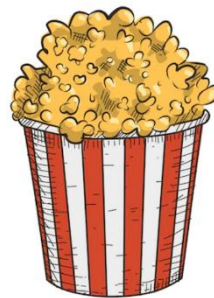


Section [I7]

IMDb Movies

Course Project Pattern Recognition



Taif Alharbi

Yara Alshehri

Ebtisam Alshehri

Introduction

IMDb Movies is the ultimate destination for movie lovers around the world. With a vast database of over 7 million titles, it is the go-to source for information on movies, TV shows, and celebrities. From classic films to the latest blockbusters, IMDb Movies has everything you need to stay up to date on your favorite flicks. Whether you're looking for reviews, ratings, trailers, or showtimes, IMDb Movies has got you covered.

Problem Description

The problem with IMDb Movies is that the reviews provided by users are often subjective and can vary greatly in quality, making it difficult for other users to determine the overall sentiment of a movie. This inconsistency in reviews can lead to confusion and frustration for those trying to make informed decisions about which movies to watch. To solve the problem, we can classify reviews into positive or negative.

Data Description

IMDb Movies is a dataset that contains information about movies reviews. The dataset consists of 50,000 movie reviews that are labeled as either positive or negative. The reviews were collected from IMDb website and are written in English language. Each review in the dataset is represented as a text string and has an associated label indicating whether it is positive or negative. The labels are binary with 1 indicating a positive review and 0 indicating a negative review. The dataset is suitable for sentiment analysis using Natural Language

Processing (NLP) techniques to determine whether a given movie review is positive or negative. NLP techniques such as tokenization, stop word removal, stemming, and sentiment analysis algorithms can be used to preprocess the data and extract meaningful features from text.

Method

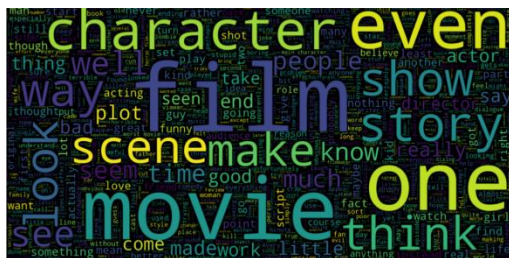
1-Support Vector Machine (SVM) is a popular classification algorithm used in Natural Language Processing (NLP) sentiment analysis for IMDb Movies dataset. SVM works by creating a hyperplane that separates the positive and negative reviews in the dataset. The SVM model is trained on a labeled dataset of movie reviews, where each review is labeled as positive or negative. Once trained, the model can predict the sentiment of new movie reviews by classifying them as either positive or negative based on text. The accuracy of the SVM model can be improved by tuning its hyperparameters.

2-Naïve Bayes classification algorithm is a popular machine learning technique used for sentiment analysis in Natural Language Processing (NLP). It is widely used to classify text data into positive or negative sentiments. In the case of IMDb Movies dataset, Naïve Bayes algorithm can be used to classify movie reviews as positive or negative based on the sentiment expressed in the text. The algorithm works by calculating the probability of a review belonging to a particular sentiment class based on the occurrence of words in the review. The model is trained on a labeled dataset of movie reviews with known sentiments and then used to predict the sentiment

Experiment and results



Positive Reviews



Negative Reviews

Discussion

Conclusion

The use of IMDb Movies sentiment analysis dataset provides insights into how people perceive and react to movies. This information can be useful for movie studios and filmmakers to understand audience reactions to their work and make adjustments to improve future projects.

References

1.9. Naive Bayes. (n.d.). Scikit-learn.

<https://scikitlearn.org/stable/md>

http://www.cs.cmu.edu/~jeff/naive_bayes.html

1.4. *Support Vector Machines.* (n.d.).
Scikit-learn.

<https://scikitlearn.org/stable/modules/svm.html>