CS432/532: Final Project Report

Project Title: Movie analysis based on indexes

Team Member(s): Aravind Reddy Putnala, Madhan Thangavel

Abstract – When we sit to relax to watch a movie, we always spend some time in finding the appropriate movie worth spending that time. We have abundant data available on the internet in form of blogs, list by some websites. The issue is those lists are subjective to few alike. So here we find a scope to analyze the information of standard reviews and present to user.

Index Terms—IMDB ratings, Genre and Director of movie.

I. PROBLEM

Most of us don't want to spend much time on analyzing the available on internet or surf through all the prominent review providers before watching a movie. That time is worth watch movie.

In this we are creating a stand-alone application which would like to analyze the data corresponding to in over the top (OTT) platforms like NETFLIX, HULU, and AMAZON PRIME. We will be working with huge chunk of data, where we will get the opportunity to implement our learned skills. And also, there exists plethora of data related to this and it is interesting to organize the existing data into the form more convenient for the user end. Which would easy them in searching on various genre.

The analyzed data can be utilized by the OTT platforms, movie production companies and other relevant agencies. The data organization based on different indexes will give a broad view of the direction to proceed for example if they have to suggest the latest to people based on their age group then this analyzed data can be utilized to for suggestions

II. SOFTWARE DESIGN AND IMPLEMENTATION

A. Data-set

We are using the mongodb data set for the analysis of our work. The basic data-set contains the id which will be unique for every, the title of movie, year of the movie released, the age group which are allowed to watch.

Along with it the review of widely accepted ratings like the IMDB rating, Rotten Tomatoes ratings. The dataset also contains the availability on various platforms like Netflix, Hulu, Prime Video and Disney+. The name of the Director, Genres which includes more than one, the name of the country produced which would be significant to region, the languages in which the movie in available on the respective platforms and Runtime.

Data set is in the form of csv file obtained from online portfolio and converted it into json format to support the stand alone application which is run using NodeJS.

B. Software Design and NoSQL-Database and Tools Used

We use mongo Database as the NoSQL-Database associated with mongo compass NodeJS and studio 3t in considering the ease and number of functions it offers for data manipulation.

Our data set consists of Id, Title of the movie, Year released, Age group, IMDb, Netflix, Hulu, Prime, Disney+, Directors, Genre, Country, Language and run time of the movie. The data-set is prepared by extending the details of the, converting the fields into required, relevant formats, dropping

the useless fields. Each data has fields like, ID, movie title, released year, age group who can watch and the rating of other applications like IMDB, Rotten Tomatoes, Netflix, Amazon Prime, and Disney, name of the directors, movie Genre, language of the movie, Country released and run time of the movie are assigned a polarity by running through library using a programming language script.

The final data-set is then inserted into the mongo database. We have written queries to retrieve data based on our interest of search as specified above. The analysis based on the movies and queries retrieved is discussed in the following sections

C. Supported Queries

The application supports the following queries:

- Search for a movie: retrieves the information of movie that includes movie name, artist, and genre. If there are more than one movie with the same name, the query is made more specific by asking the user to enter the artist's name.
- Search for movies by a director: retrieves the information of all movies by a particular director.
- Search for movies by a genre: retrieves the information of all movies of a particular genre.
- Get the hot charts for a year: retrieves the information of all movies present in the hot charts for a particular year.
- Get all the top movies: retrieves the information of all top-ten movies in every year between 1958 - 2019
- Get all the top genres: retrieves the top-ten genres for each week in every year between 1958-2019
- Get all the top language: retrieve the information based on the user's language preferences.
- Get the movies on age and IMDb: retrieve the information based on selected age group and movie rating with respect to Netflix, Hulu, Prime Video and Disney+.
- Get the movies based on rotten tomatoes: movies retrieved between two different percentages in rotten tomatoes.
- Get the movies based on language and profile: Movies retrieved based on user's choice language and different online platforms ratings.
- Get the movies based on year: retrieve the movies based on released between two years.
- Get the movies on aggregation of rating and year: retrieve the information based on the rating and aggregation of year.
- Get the movie based on aggregation and year: retrieve the movie information based on grouping the year and directors.

III. SENTIMENT ANALYSIS

The sentiment analysis is performed to understand how the movies and music preference of people has changed over the years.

The article "Sentiment analysis of all billboard Hot 100 movies over time (1958–2019)" on medium.com provided us with a basic idea on how to perform a sentiment analysis [1].

A. Movie Sentiment Analysis

The average polarity is calculated for each year and a line graph is plotted against the time. From the plot it can be observed that the lyrics of movies have become increasingly negative over the years. Further, movies were most positive in the year 1979 and least positive in the year 2019.

B. Genre Sentiment Analysis

The opinion conveyed by the user towards a movie can be understood by doing Sentiment Analysis on the movie review. In the current work we focus on Genre Specific Aspect Based Sentiment Analysis of Movie Reviews. Using the aforementioned dataset and considering movie genres like action, comedy, crime, drama and

horror, we develop a fine grained unsupervised analysis model using lexicons that are context specific to each genre under consideration has emerged as the most popular music genre showing the change in the music preference of people.

The Rotten Tomatoes movie review dataset is a corpus of movie reviews used for sentiment analysis, originally collected by Pang and Lee [1]. In their work on sentiment tree banks, Socher et al. [2] used Amazon's Mechanical Turk to create fine-grained labels for all parsed phrases in the corpus. This competition presents a chance to benchmark your sentiment-analysis ideas on the Rotten Tomatoes dataset. You are asked to label phrases on a scale of five values: negative, somewhat negative, neutral, somewhat positive, and positive. Obstacles like sentence negation, sarcasm, terseness, language ambiguity, and many others make this task very challenging.

IV. PROJECT OUTCOME

The source code is on github: https://github.com/Madhan-GitHub/Movies_Analysis_Based_On_Indexes/

V. ACKNOWLEDGMENT

We thank our Professor, Kyoung-Don Kang for all his assistance for this project. We offer our sincere appreciation for all the learning opportunities.

REFERENCES

[1] Salim Zubair, "Sentiment Analysis of all Billboard Hot 100 Movies Over Time (1958–2019)," https://towardsdatascience.com/sentiment-analysisof-all-billboard-hot-100-movies-over-time-1958-2019-3329439e7c1a

[2] MongoDB – Database Query Insertion https://docs.mongodb.com/manual/tutorial/insert-documents/

[3] MongoDB – To gather and merge data https://rstudio-pubs-static.s3.amazonaws.com/384729_2d449865dfc4457dbd99b2198c666667.html

[4] Mongo Compass – Performing application GUI https://www.bmc.com/blogs/mongodb-compass/

[5] Studio 3t – https://studio3t.com/features/ To Import and export features

[6]https://ieeexplore.ieee.org/document/7275981 Sentiment Analysis – IEEE Paper

[7] https://www.w3schools.com/nodejs/ - NodeJS – Tutorial and concepts of programming