International Journal of Scientific Research & Engineering Trends
Volume 7, Issue 1, Jan-Feb-2021, ISSN (Online): 2395-566X

Movie Stats: Sentiment Analysis of IMDB Reviews and Tweets of a Movie Using Naïve Bayes Classifier

Deep Rahul Shah
Department of Computer Engineering
MPSTME,NMIMS
Mumbai, India
deepshah3110@gmail.com

***Abstract*- Sentiment analysis, also referred to as opinion mining or emotion extraction is the classification of emotions within textual data. This technique has been widely used over the years to determine the sentiments, emotions within a particular textual data. I have designed a way where we as a user can view the real time sentiment analysis of movie reviews and tweets regarding a certain movie. I built a Sentiment Analysis model with an accuracy of 98.7 % and an F1 score of 0.99, which means the model is nearly 99% as accurate as a human in analyzing the Sentiment of the text. Many websites are providing you lengthy reviews, which makes it quite boring to read it, so the proposed system shows the real-time chart and sentiment analysis of the tweets and movie reviews in a user- friendly way. This is what makes MovieStats unique.**

***Keywords*- Movie Sentiment Analysis; Twitter Sentiment Analysis ; Naïve Bayes Supervised approach; Machine Learning**

I. INTRODUCTION

Sentiment analysis or emotion AI refers to the use of natu ral language processing, computational linguistics to system atically extract emotions, sentiments, opinions i.e. the subjective information in a piece of textual data. Opinion mining has found its use mainly within the market research allo wing a business to understand the sentiment regarding their products, services [1].

It not only allows the monitoring of opinions but also the likes and dislikes of people in general. Over the years Twitter has developed as one of the leading social media platforms allowing peopleto express their views, opinions, and sentiments concerning an event, incident in the form of tweets and IMDB is one of the most reputed and trusted websites for legitimate reviews about a certain movie. By monitoring these tweets and reviews, one can easily understand how they are being perceived and what improvements could be made. It has changed the outlook to a great extent, not only providing all the necessary information but also making it possible for people to be open about their views. Thus, IMDB and Twitter could be considered as one of the best ways of monitoring the opinions of people when it comes to movie reviews [2].

With the advances in Machine and Deep Learning, the ability to predict the sentiment within textual data has increased. It is because of these algorithms that it has become possible to predict the sentiments with great accuracy. Machine learning allows the learning of new tasks without being explicitly trained or programmed. Sentiment analysis models can be used to predict not only the sentiment but other subjective information within a piece of text.

Several Machine learning and deep learning algorithms could be used for predicting emotions, sentiments. In this project, I have used Naïve Bayes Classifier as it helps us achieve higher accuracy in Analysing the Sentiment of the movie. Multinomial Naive Bayes classification algorithm tends to be a baseline solution for sentiment analysis tasks. The basic idea of the Naive Bayes technique is to find the probabilities of classes assigned to texts by using the joint probabilities of words and classes.

II. IMPLEMENTATION

The target users for MovieStats are the people who are willing to spend around 100 to 150 minutes on a movie and wants to know what do people think about a certain movie, instead of reading lengthy reviews on IMDB or reading a million tweets about a movie on Twitter, they can go to the website homepage and enter the movie they are willing to see the detailed sentiment analysis of the movie.

MovieStats works for more than 15,000 + English movies and it also works with the Hindi and English movies released in and after 2020. As the user enters the movie name the Movie Reviews are fetched from IMDB website ,tweets are fetched using the Python Tweepy module and then the text is sent to our Sentiment Analysis model for the sentiments of the text to be

© 2021 IJSRET

analysed,and then the numbers of positive and negative reviews count are then analysed and visualized on the on the website page.

Fig 1. The home page where you can enter the movie name.

Users can also see the entire detailed reviews and tweets on the same dashboard page which were used for analyzing sentiments of the movie and the interactive charts make it easy for a user to understand the sentiment of the reviews and tweets of the movie.

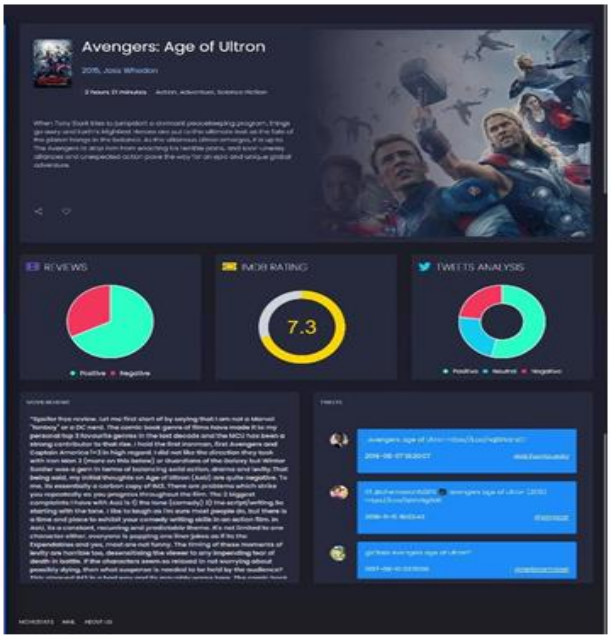

Fig 2. Dashboard of MovieStats.

User can also share the sentiment analysis of the movie over mail to a friend. The tweets are being analyzed using the TextBlob library with an accuracy of around 60% - 70% and the Movie reviews are being analyzed by the sentiment analysis model that I built with an accuracy around 98.7%. Both texts were analyzed differently to create a comparative analysis between the TextBlob library and my Model.

User's can also view the detailed list of Reviews and Tweets with the time and date they were posted. The IMDB Reviews are taken based on the one which was highly voted to the one with less votes.

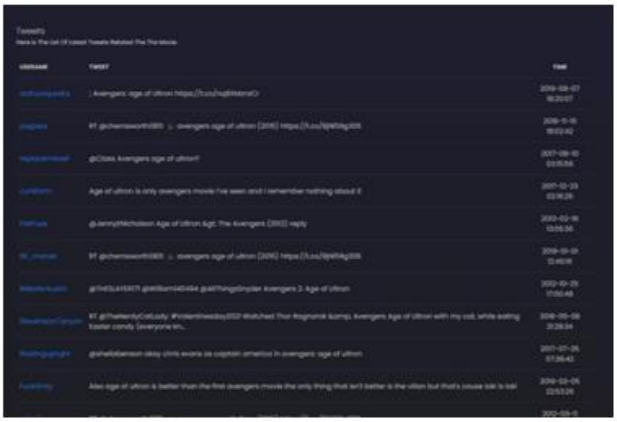

Fig 3. Detailed View of Tweets.

Fig 4. Detailed View of Movie Reviews.

	precision	recall	f1-score	support
0	0.99	0.98	0.99	580
1	0.98	1.00	0.99	804
accuracy			0.99	1384
macro avg	0.99	0.99	0.99	1384
weighted avg	0.99	0.99	0.99	1384

Fig 5. Classification Report of the Sentiment Analysis model.

III. INFERENCE

Following are the benefit and features of MovieStats:-

- HigherAccuracy Sentiment analysis with an accuracy ≈ 99% and f1-score 0.99

© 2021 IJSRET

- One Stop station for visualizing and real-time analy zing the Sentiments of the IMDB Reviews and Twe ets of the Certain Movie with an feature of accessing the profile of person tweeting and sharing the review.
- User Friendly visualizing and analyzing of data.
- This features are not available on most of the web sites related to movies.

IV.CONCLUSION

In an ideal scenario if an user wants to know how well the movie did ,instead of reading lengthy boring reviews and going through a million tweets about,MovieStats provides you high accuracy Sentiment Analysis of both in an user friendly way making it easier for the user to analyze.

Ass uming we have more users we can use the data anony mously to recognise pattern and train Machine Learning algorithms to predict what kind of movie user would like to see. To cover the Bollywood and regional cinema we are able to analyze sentiment of the Regional and Bolly wood movie but we are not accurately able to identify the sentiments of the tweets as they are posted in regional languages and our model cannot analyze the sentiments of any other language except

English, but I am even planning to add a dictionary for as many regional languages as possible. In the world where OTT platforms are booming and demand for web series are increasing am also working to add a section where user can see the Analysis of a certain season or an episode of the series.

REFERENCES

[1] B.Pang,L.Lee,andS.Vaithyanathan,sentiment classific ation using machine learning techniques, Proceedings of the ACL-02 conference on Empirical methods in natural language processing, vol.10, 2002, pp.79-86.
[2] Edosomwan,Simeon,etal."The history of social media and its impact on business."Journal of Applied Mana gement and entrepreneurship 16.3(2011):79-91

© 2021 IJSRET