Sentimental Analysis on Movie Reviews

By

JAHAN GAGAN - 18BCE357
HIREN GOHIL - 18BCE359

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
Ahmedabad 382481

Sentimental Analysis on Movie Reviews

Mini Project – III

Submitted in partial fulfillment of the requirements

For the degree of

Bachelor of Technology in Information Technology

By

JAHAN GAGAN - 18BCE357
HIREN GOHIL - 18BCE359

Guided By
Prof. Kruti Lavingia
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
Ahmedabad 382481

CERTIFICATE

This is to certify that the Mini Project -III entitled "**Sentiment Analysis on Movie Reviews**" submitted by **Jahan Gagan** (**18BCE357**) and **Hiren Gohil** (**18BCE359**), towards the partial fulfillment of the requirements for the degree of Bachelor of Technology in Information Technology/Computer Engineering of Nirma University is the record of work carried out by him/her under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination.

Kruti Lavingia
Assistant Professor
Department of Computer Science & Engg.,
Institute of Technology,
Nirma University,
Ahmedabad

Dr. Madhuri Bhavsar (HOD)
Dept. of Computer Science & Engineering,
Institute of Technology,
Nirma University,
Ahmedabad

ACKNOWLEDGEMENT

The success of any work depends on the efforts made by the individual, but it cannot be achieved without the support of other people who are getting help. The perfect session of our mini project termination was an excellent experience, which provides us with understanding an invitation to learn about the various sockets concepts and its benefits. Without support and cooperation of people on this opportunity, we should make it difficult to complete this project. For our leading guide and honest efforts in the project, we will review our project guidelines "**Professor Kruti Lavingia**". Thank you very much for victory. They took a keen interest in making difficulties easier. Apart from that, they are also the source of inspiration for us.

ABSTRACT/ Outline

We have designed an aspect-oriented approach to analyze text i.e., Movie reviews in our case and assign a sentiment tag to each movie aspect. For that the movie reviews are first scraped and then analyzed with the Natural Language Toolkit (NLTK) package of python. After getting a sentiment tag we have compared it with the website ratings, the result obtained showed our analyzed text produces accurate and targeted sentiment tag. At last the concept is implemented as Android Mobile Application.

CONTENTS

Table of Contents

Table of Figures

CHAPTER - 1 : INTRODUCTION

1.1 General

Project Name	Sentiment Analysis on Movie Reviews
Project Type	Mobile Application
Front-End	Android Java
Back-End	Python Flask
Developed For	Users that are enthusiastic for movies verdict
IDE Used	Google Colab
Scraping Website	imdb.com
Documentation Tool	MS Office
Submitted to	Department of Computer Science and Engineering
Developed By	Jahan Gagan (18BCE357) Hiren Gohil (18BCE359)
Guided By	Professor Kruti Lavingia

Table 1 Introduction

1.2 What is Sentiment Analysis?

Sentiment analysis is a text analysis method that can determine the polarity (for example, positive or negative tags) in a text, whether it is an entire document, paragraphs, sentences or sentences.

Understanding the emotions of customers is critical to the company because customers can express their thoughts and feelings more openly than ever. By automatically analyzing customer feedback, brand owners can focus on listening to customers and tailoring products and services to meet their needs.

Sentiment analysis applications:

Social media monitoring
Brand monitoring (**To which our project belongs to**)
Voice of customer (VoC)
Customer service Market research

Brand Monitoring:

1. Analyze news articles, blog posts, forum discussions, and other text on the Internet over a period of time to see the sentiment of a fixed audience
2. Automatically classify all the urgency of mentioning your trademark online.
3. Automatically remind designated team members about online raises of their work areas.
4. Automate any or all of these processes.
5. Get a better perception of the brand's online image by collecting various thrilling insights and analysis

1.2.1 What is the NLTK Package?

The NLTK Module module is a large suite of tools designed to help you use the full natural language processing method (NLP). NLTK can help your machine help you identify paraphrasing sentences, word distribution, speech parts of these words, highlight key points, and understand the overall content of the text. In this series, we will explore the field of opinion or spirit analysis.

We have used following concepts of NLTK in our project

1. Tokenizing words
2. NLTK stop words
3. NLTK – stemming
4. Tokenizing sentences

1.3 Objective

Objective is to build a model that is used to classify the movies based on user reviews.

The second objective is to come up with pertinent points that could reveal the interesting facts about the movie.

CHAPTER - 2 : LITERATURE SURVEY

2.1 General

Literature Survey covers various websites and papers mentioned in Appendix - A that introduced the NLP, NLTK package and its various concepts. Along with that the study of the Flask framework was carried out. Flask is a python based framework used to create web applications but we have used it to create API for our android application.API returns sentiment analysis in the form of JSON and parsing that JSON into android finally we have implemented our mentioned objectives. Also, hands on to the python programming, specially the concept of regular expressions in python which is mainly used for cleaning the data.

2.2 Scope

With the lethal growth of multi-channel data such as social or mobile data, companies need powerful techniques to assess and evaluate customer confidence. Yet, companies are happy to analyze customer behavior. But in today's highly competitive environment, this type of consumer analysis is outdated.

Now, companies need to analyze and understand customer attitudes, preferences and feelings in the field of sentiment analysis. Without NLP, business owners will have to make serious efforts to analyze beliefs.

So belongs to the analysis of movie reviews, it can be possible that users use the app and based on it further demand for other interesting facts or can be demanded by film industry producers to analyze the negative reviews i.e., what made our film wrong like cast, budget, shooting locations etc.

CHAPTER - 3 : PROJECT DESCRIPTION

3.1 Steps Carried Out

For our project, we have used a review collection of the Bollywood movies.
The collection contains a variable number of reviews for each movie. So,
For our implementation we have taken maximum 25 reviews from the collection for each movie.

To achieve this, we had done a web scraping from imdb website and collected reviews from imdb and wrote it into the excel sheet.

- **Approach**
  1. Created an excel sheet which contains the name of 1000's of bollywood movies
  2. Fetched movie name from sheet
  3. Search the movie by appending "imdb reviews" to it and make a google search. Through a search visit to the website and scrapped 25 reviews(if available) from it.
  4. After scrapping the reviews, it is written to the excel sheet.

- **Steps**
  1. **Cleaning :** It includes
     a. Conversion of reviews to the lowercase
     b. Removing URLs
     c. Removing stop words
     d. Removal of !, #, @, multiple space, digits, dot(.) and comma(,) colon(:), semicolon(;), apostrophe(')
     (Inshort accepting only strings with lowercase)

  2. Sentiment : Now, the actual development of the project started. From reviews in an excel sheet, using a sentiment library, sentiment analysis is done.

  Based on sentiment value (rounded to 3 decimal places), rank is assigned

between 1 to 10 because imdb review rating was from out of 10 so it make easy to check the model's accuracy by comparing both the data's.

Now based on rating of user's review, outlier points (say min and max) are calculated. Based on outlier values, sentiment value of scraped reviews are compared and if found below min or above max then, that value of review is replaced with the average of all sentiment values of reviews of that a particular movie.

Q. Why average value?

Actually first we have tried with the frequent value of review sentiment but it doesn't fit correctly with imdb rating.

Then we have tried with standard deviation but it was not suitable for negative reviews.

Then finally we found average value to be more appropriate.

Q. How concluded appropriately?

By doing mean square error on imdb reviews and sentiment values we found the answer was 1.902 for 1890 movies which is very obvious and minimum compared to other methods.

3. Deployment

It was a quite difficult task for us, because, for that we needed a hands on practice on python programming and to create API in it we had to learn a python framework called **Flask Framework.** However we did it and created a JSON object which contains our result. Then comes a step for integrating flask to android, which was again a quiet time oriented task. Atlast we did it too and deployed our project in android as a front-end and flask as a back-end.

4. Screen Shots

Figure 1 Splash Screen

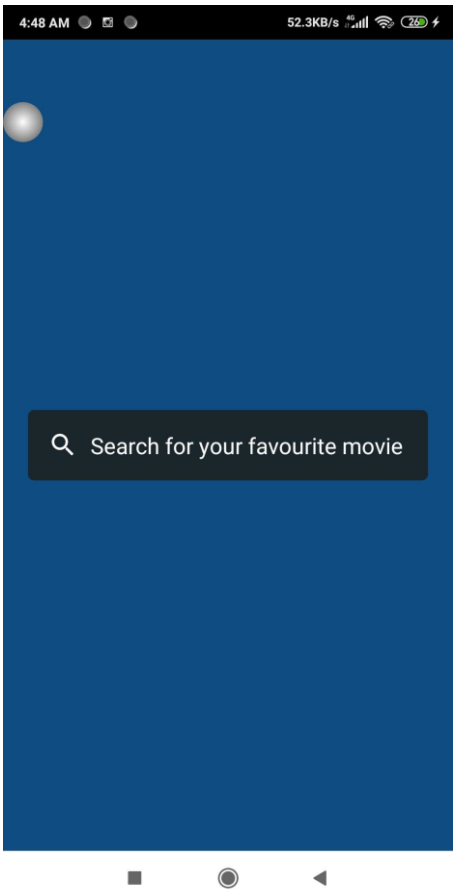

Figure 2 Movie name as a parameter

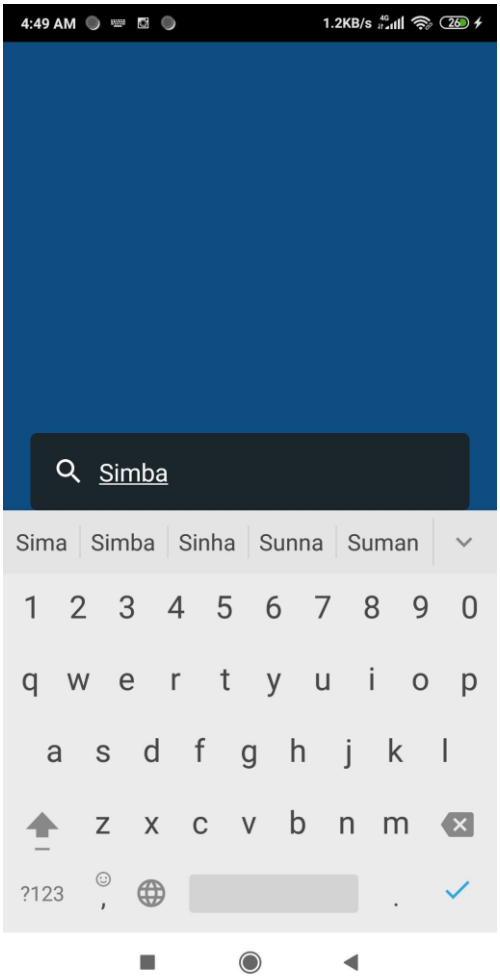

Figure 3 Movie as a parameter

Figure 4 Processing

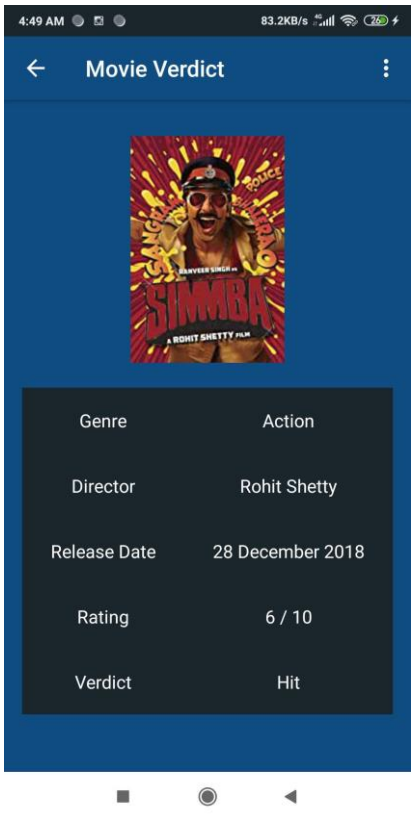

Figure 5 Output

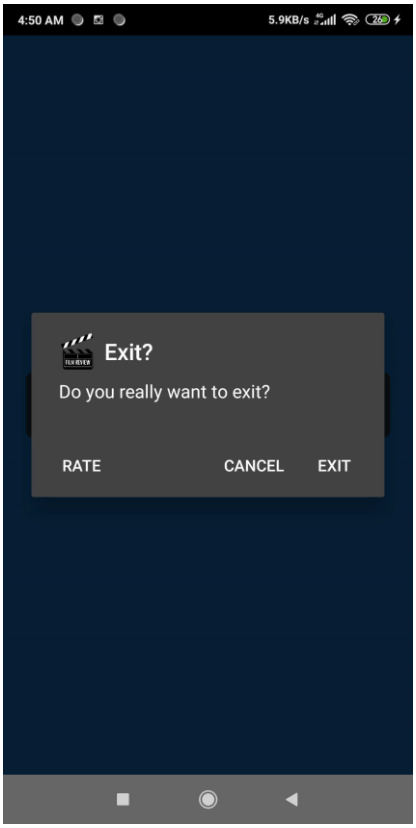

Figure 6 Exit Confirmation

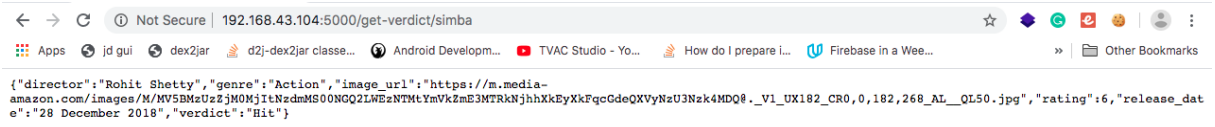

Figure 7 API Response

Video

CHAPTER - 4 : Conclusion

4.1 Conclusion

After performing all the mentioned approaches and step we had successfully developed an android application that takes simply movie name as an argument and scraps banner of movie, release date and reviews from imdb website and after performing nlp operations it gives rating and final verdict based on user reviews and by doing so, we successfully achieved our objectives.

Appendix – A List of Useful Websites

https://www.analyticsvidhya.com/blog/2017/01/ultimate-guide-to-understand-implement-natural-language-processing-codes-in-python/

https://www.udemy.com/course/nlp-natural-language-processing-with-python/

https://www.youtube.com/watch?v=vGjQ1B3FNnU

https://www.imdb.com/

https://www.edureka.co/blog/web-scraping-with-python/

https://realpython.com/python-web-scraping-practical-introduction/