**A** **PROJECT** **REPORT** **ON**

# Movie Recommendation

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR SEMINAR OF

# BACHELOR OF ENGINEERING (COMPUTER ENGINEERING)

**SUBMITTED** **BY**

Ms. Vaishnavi Khile

# DEPARTMENT OF COMPUTER ENGINEERING

**SHRI** **CHHATRAPATI** **SHIVAJI** **MAHARAJ** **COLLEGE** **OF** **ENGINEERING**

NEPTI, AHMEDNAGAR



**SAVITRIBAI** **PHULE** **PUNE** **UNIVERSITY** **2021-22**



**SHRI** **CHHATRAPATI** **SHIVAJI** **MAHARAJ** **COLLEGE** **OF** **ENGINEERING**

NEPTI, AHMEDNAGAR

# DEPARTMENT OF COMPUTER ENGINEERING

## CERTIFICATE

This is to certify **Our group** is a bonafide student of this institute and the work has been carried out by her under the supervision of **Prof.Shraddha Lande** and has presented and has successfully completed the DSBDA project on Tweet Application in the presence of the seminar evaluation panel for the partial fulfilment of the requirement of Savitribai Phule Pune University, for the award of the degree of Bachelor of Engineering (Computer Engineering).

***(*Prof.Shraddha Lande*)*** *(****Dr.*** ***H.*** ***B.*** ***Jadhav)***

Guide Seminar Coordinator

Department of Computer Engineering Department of Computer Engineering

#### (Prof. Yashanjali Mam) (Dr. M.P. Nagarkar)

Head Principal SCSMCOE

Place: Ahmednagar Date:

# ACKNOWLEDGEMENT

I take this opportunity to express my hearty thanks to all those who helped me in the completion of the Project of DSBDA on topic **Movie Recommendation**. I would especially like to express my sincere gratitude to my Guide **Prof.** **Prof. Shraddha Lande**, HOD of Computer Engineering Department **Prof.** **Jashanjali Mam** who extended their moral support, inspiring guidance and encouraging independence throughout this task. I would also thank our Principal **Dr.** **M.** **P.** **Nagarkar** for his great insight and motivation. Last but not least, we would like to thank my fellow colleagues for their valuable suggestions.

## NAME OF THE STUDENT

**ABSTRACT**

In this current era, social media plays a important role in data exchange, sharing their thoughts. Emotional Effect of a person maintains an important role on their day to day life. Sentiment Analysis is a procedureof analyzing the opinions and polarity of thoughts of the person. Twitter is a main platform on sharing the thought's, opinion and sentiments on different occasions. Twitter Sentimental Analysis is method of analyzing the emotions from tweets (message posted by user in twitter). Tweets are helpful in extracting the Sentimental values from the user. The data provide the Polarity indication like positive, negative or unbiassed values. It is focused on the person’s tweets and the hash tags for understanding the situations in each aspect of the criteria. The paper is to analyse the famous person’s id’s (@realdonaldtrump) or hash tags (#IPL2018) for understanding the mindset of people in each situation when the person has tweeted or has acted upon some incidents. The proposed system is to analyze the sentiment of the people using python, twitter API, Text Blob (Library for processing text). As the results it helps to analysis the post with a better accuracy

## TABLE OF CONTENTS

**CHAPTER** **TITLE** **PAGE** **NO.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr.** **No.** | | **Title** **of** **Chapter** | **Page** **No.** |
| **01** | | **Introduction** | 3 |
|  | 1.1 | Overview | 3 |
|  | 1.2 | Motivation | 5 |
| **02** | |  | 6 |
| **03** | | **Technology** | 8 |
|  | 3.1 |  | 8 |
|  | 3.2 |  | 9 |
|  |  | **Implementation** | 10 |
| **04** | | Implementation Details | 15 |
|  | 4.1 | Tools and Technologies Used | 15 |
|  | 4.2 |  | 15 |
|  |  |  |  |
|  | **4.3** | **Code** | 16 |
| **05** | | **Results** | 17 |
|  | 5.1 | Outcomes | 18 |
|  | 5.2 |  | 18 |
| **06** | | **Conclusions** | 19 |
|  | 6.1 | Conclusions | 19 |
|  | 6.2 |  | 19 |
|  | |  | 20 |
| **07** | | **References** | 21 |

### Chapter -1 Introduction

* 1. **Overview** **:**

We compare and use various different methods for sentiment analysis on tweets. The training dataset is estimated to be a csv file of type tweet\_id, sentiment, tweet

Where tweet id – unique identifying the tweet.

Sentiment - IS either 1 (positive) and 0 (negative)

The tweet – the tweet enclosed in “”.

Similarly, the test dataset is a csv file of type tweet\_id, tweet, please note that csv headers are not expected and should be removed from the training and test datasets.

# Motivation:

In the past years, the young generation people are moving towards the social media like Google Plus, WhatsApp, Facebook, Twitter, etc. The social media is also revolving with those people to get them involved by making current trending insights concepts that is trending within a second. In the recent years, the people are exposing their social related issues through several social media by comments, reviews, posts, hashtags, emoji’s, etc. which was followed by many people and those tweets become popular soon. Moreover, the social media is also bringing tremendous opportunity platform for businesses to connect with the consumers so easily. People rest on mostly user produced content like, comments, over online for making the decision. Example: if anyone has to buy a product or make a decision, they initiallysearch its reviews online, converse about it on social media. The content that is displayed for that product is mainly taken into the point as well as the discussion in the social media is also noticed and these made the way to make our business a success. To automate our analysis based on the reviews or comments in the social media by the people, for a sentimental analysis. Sentimental Analysis (SA) is introduced to the world to tell us the information is correct or wrong in each scenario using the social media tags. Thus, we can know about how world or people are reacting to every aspect currently going in the world.

**Chapter** **–** **3**  **Implemention**

**Sentiment Analytics**

* **Business –** In marketing field companies use it to develop their strategies, to understand customers, feeling towards products or brand, how people respond to their campaigns or product launches and why consumres don’t buy some product.
* **Politics –** In political field, it is used to keep track of political view, to detect consistency and inconsistency between statements and actions at the government level.
* **Public Action –** Sentiment analysis also is used to monitor any analysis social phenomena, for the spotting of potentially dangerous situations and determining the general mood of the blogosphere.

**Step 1: Install and Import Libraries**

Before analysis, you need to install **textblob** and **tweepy** libraries using ***!pip install*** command on your Jupyter Notebook.

Tweepy supports both OAuth 1a (application-user) and OAuth 2 (application-only) authentication. Authentication is handled by the tweepy.AuthHandler class.

OAuth 2 is a method of authentication where an application makes API requests without the user context. Use this method if you just need read-only access to public information.

You first register our client application and acquire a consumer key and secret. Then you create an AppAuthHandler instance, passing in our consumer key and secret.

Before the authentication, you need to have **Twitter Developer Account.**If you don’t have, you can apply by using this [**link**](https://developer.twitter.com/). Getting Twitter developer account usually takes a day or two, or sometimes more, for your application to be reviewed by **Twitter.**

**Step 2: Authentication for Twitter API**

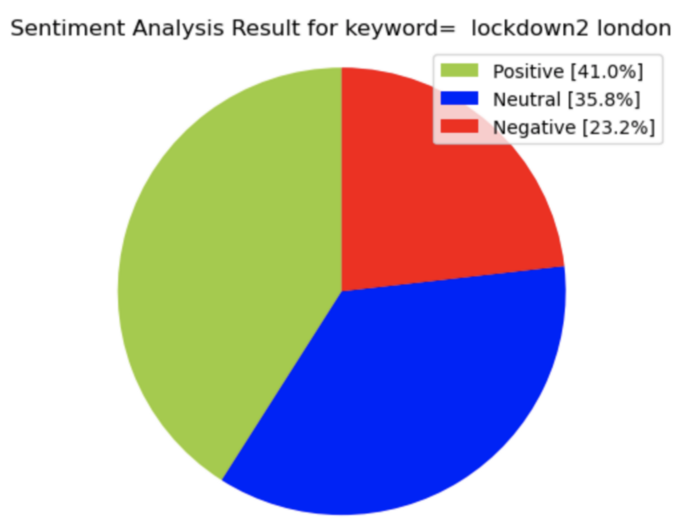
After your authentication, you need to use tweepy to get text and use Textblob to calculate **positive**, **negative**, **neutral**, **polarity** and **compound** parameters from the text.

**Step 3: Getting Tweets With Keyword or Hashtag**

The scenario in this post like that, the user should type keyword or hashtag (**lockdown2 london**) and type how many tweets (**2500**) that want to get and analyse.

You could get 2500 tweets and;

* **1025 (41.0%)** of tweets include positive sentiment
* **580 (23.2%)** of tweets include negative sentiment
* **895 (35.8%)** of tweets include neutral sentiment



**Step 4: Cleaning Tweets to Analyse Sentiment**

When you have a look tweet list you can see some duplicated tweets, so you need to drop duplicates records using **drop\_duplicates** function.

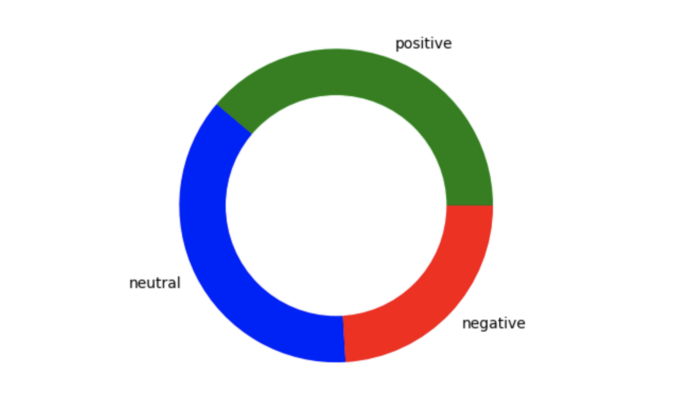
Image by the author

Our new data frame has **1281** unique tweets.

Firstly, I create new data frame (tw\_list) and a new feature(text), then clean

**Step 5: Sentiment Analyse**

Now, I can use cleaned text to calculate polarity, subjectivity, sentiment, negative, positive, neutral and compound parameters again sentiment. For this one, create 3 new data frame



Now you can prepare to create worcloud using 1281 tweets, So you can realize that which words most used in these tweets. To create a worcloud, firstly let’s define a function below, so you can use wordcloud again for all tweets, positive tweets, negative tweets etc.

Applying count vectorizer provides the capability to preprocess your text data prior to generating the vector representation making it a highly flexible feature representation module for text. After count vectorizer, it is possible to analyze the words with two or three or whatever you want.

Applying stemmer is also provides the root of words. So you can eliminate words that come from the same root, such as ;

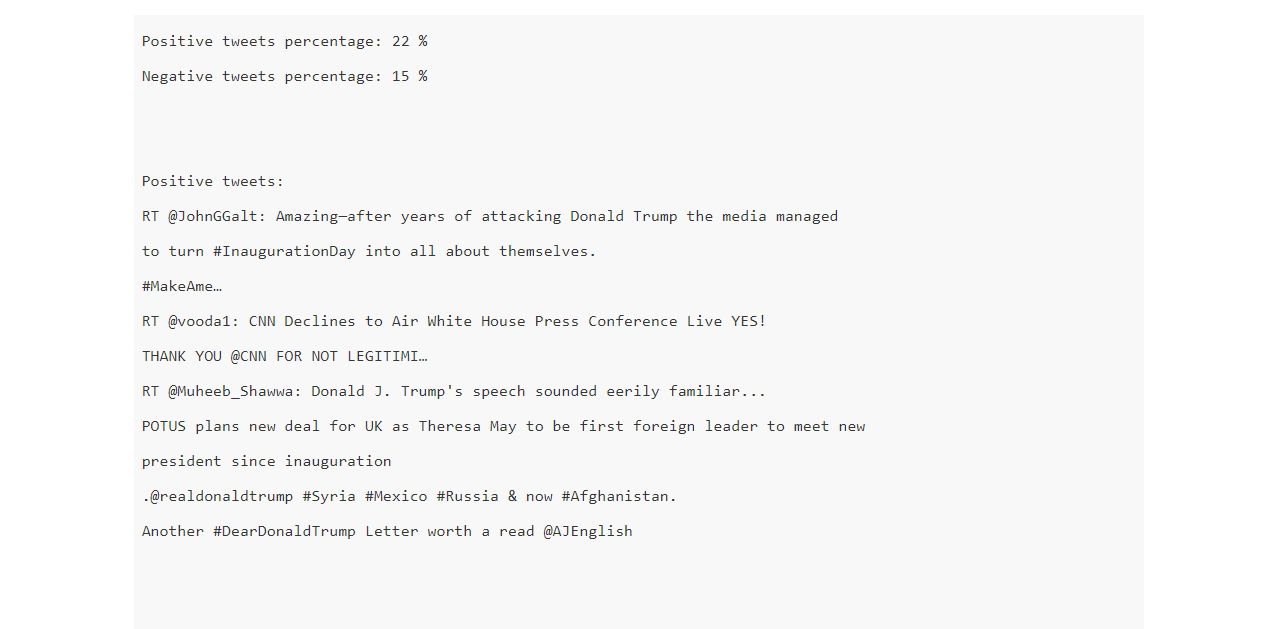
* connect
* connection
* connected
* connections
* connects

comes from **“connect”.**

### Outcomes:

**Output / Screenshots:**

Output 1 :



### Output 2 :

### 

### Chapter -6 Conclusion

* 1. **Conclusion:**

Twitter sentiment analysis comes under the category of text and opinion mining. It focuses on analyzing the sentiments of the tweets and feeding the data to a machine learning model to train it and then check its accuracy, so that we can use this model for future use according to the results. It comprises of steps like data collection, text preprocessing, sentiment detection, sentiment classification, training and testing the model. This research topic has evolved during the last decade with models reaching the efficiency of almost 85%-90%. But it still lacks the dimension of diversity in the data. Along with this it has a lot of application issues with the slang used and the short forms of words. Many analyzers don’t perform well when the number of classes are increased. Also, it’s still not tested that how accurate the model will be for topics other than the one in consideration. Hence sentiment analysis has a very bright scope of development in future.

### Chapter References

1. Jansen,B.J.; Zhang,M.; Sobel,K.; and Chowdury,A. (2009), “Twitterpower: Tweets as electronic word of mouth”, Journal of the American Society for Information Science and Technology 60(11):2169–2188.
2. 2. Pak, A., and Paroubek, P (2010), “Twitter as a corpus for sentiment analysis and opinion mining”. In Proc. of LREC.
3. 3. Pang, B., and Lee, L. (2008), ”Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval” 2(1-2):1– 135.
4. 4. Wilson, T. Wiebe, J.; and Hoffmann, (P. 2009),”Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. Computational Li nguistics”, 35(3):399–433.
5. 5. M Hu and B Liu. (2004),”Mining and summarizing customer reviews. KDD”.
6. 6. L. Barbosa, J. Feng. “Robust Sentiment Detection on Twitterfrom Biased and Noisy Data”. COLING 2010: Poster Volume,pp. 36-44.
7. 7. J. Kamps, M. Marx, R. J. Mokken, and M. De Rijke, “Using wordnet to measure semantic orientations of adjectives,” 2004.
8. 8. Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. Technical report, Stanford.
9. 9. David Zimbra, M. Ghiassi and Sean Lee, “Brand-Related Twitter Sentiment Analysis using Feature Engineering and the Dynamic Architecture for Artificial Neural Networks”, IEEE 1530-1605, 2016.
10. 10. Varsha Sahayak, Vijaya Shete and Apashabi Pathan, “Sentiment Analysis on Twitter Data”, (IJIRAE) ISSN: 2349-2163, January 2015.