

Sentiment Analysis of Posts and Comments of OSN

Abhishek Chaube¹, Vaidehi Dani², Trupti Dhapola³, Prof. Madhura Vyawahare⁴

^{1,2,3}Student, Dept. of Information technology Engineering, Pillai College of Engineering, Maharashtra, India

⁴Professor, Dept. of Computer Engineering, Pillai College of Engineering, Maharashtra, India

Abstract - Online social networking sites play an important role in the day to day life of everyone. Twitter is one of the most used platforms for social interaction. Twitter allows its users to have their own views, comments, and emotions majorly using texts and emoticons. There are many views, posts, and comments which may hurt the sentiments of other people. Monitoring such information over the discussion forums or groups is required but it becomes difficult as each individual has their own opinions and suggestions. Nowadays, depression-related posts have become a major concern. The proposed system aims to identify the sentiment behind the content posted on twitter to find depressed people. With sentiment analysis, we are targeting to find out the user's positive emotions (like happy, excited, surprised, satisfaction) and negative emotions (like depression, anxiety, stress, sad, angry). Then negative sentiments will further, be classified and the degree of depression is displayed. This study is significant to monitor the emotions of the users.

Keywords- Sentiment analysis, emotions, degree, comments.

1. INTRODUCTION

The project is based on Machine Learning and Data mining domain. Machine Learning is a field of computer science that gives the computer the ability to learn without being explicitly programmed. Machine Learnings' main focus is to provide algorithms that can be trained to perform a task. Data mining is the process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems. Data mining is an interdisciplinary subfield of computer science and statistics with an overall goal to extract information (with intelligent methods) from a data set and transform the information into a comprehensible structure for further use.

2. LITERATURE SURVEY

Sentiment Analysis

The author Neethu and Rajashree explained the need for sentiment analysis to match social media usage. Social media includes all the sites such as blogs, Facebook, Twitter, and other social networking sites. Sentiment score helps examine whether the tweet seems to be positive, Negative or Neutral. They have gathered reviews

on services and brands is done using opinion mining. The marketing-related sector is still at the forefront followed by finance, hospital, and tourism. Politics and Government are still a few of the bodies that are beginning to use sentiment analysis [1].

Sentiment analysis is done to detect the level of depression in paper [2]. The keyword-based approach is used to calculate the intensity of emotions. Keyword-based approach classifies text based on the presence of negative or positive polarity words such as happy, joyful, delighted, miserable, sad, terrified, and uninterested, happy, joyful, respectively. Weights are given to different words related to depression by evaluating the polarity of those words and the result is displayed in the metric form.

The author explains that Social Media has become very popular as a way of communication. In this research, authors have used a method for Automatic collection of data that can be used to train sentiment classifiers. They observe different emotions such as positive, negative, neutral, etc. They used different methods such as Tree Tagger, Syntactic Structures, N-gram and POS-tags amongst which POS-tags are able to identify sentiments strongly from emotional text [3].

The authors have built a model that analyzes sentiments on Twitter using Machine Learning techniques. They have applied Bigram, Unigram, Object-oriented features as an effective feature set for sentiment analysis. They have used a data set about 200000 tweets for training classifiers. They built a sentiment analysis model based on supervised learning such as Naive Bayes and Support Vector Machine for enhancing elective classification. Twitter APIs are used as a library tool to collect tweets from the internet for sentiment analysis and a system is built based on Naive Bayes (NB) and Support Vector Machine (SVM). Accuracy is not 100% as there are many grammar mistakes and repetitive words that disturb the feature dataset. This paper only works for English words and not any other languages [4].

Twitter sentiments are considered and analyzed using different Symbolic and Machine Learning techniques to identify sentiments from text. There are certain issues while dealing with identifying emotional keywords from tweets having multiple keywords. It is also difficult to handle misspellings and slang words. To deal with these issues, and efficient feature vector is created by doing

feature extraction in two steps after proper preprocessing. In the first step, twitter specific features are extracted and added to the feature vector. After that, these features are removed from tweets and again feature extraction is done as if it is done on the normal text. These features are also added to the feature vector. Classification accuracy of the feature vector is tested using different classifiers like Naive Bayes, SVM, Maximum Entropy and Ensemble classifiers [5].

How depression can be monitored by users using social media is explained in paper [6]. Sentiment analysis and affective computing methods are used to detect and monitor depression. This uses supervised and unsupervised learning methods and a lexicon-based approach is used. It detects depression and also provides a multimodal system to check facial expressions, images, and videos shared by the users [6].

Paper [7] aims to identify the opinion mining and sentiment analysis components for extracting both English and Malay words on Facebook. Information, in terms of texts, are extracted and clustered into emotions. This work begins with transforming unstructured information into meaningful lexicons after extracting Facebook's contents. All of the meaningful lexicons are stored in a database after manual identifications are carried out. With sentiment analysis, emotions are classified into happy (positive), unhappy (negative) and emotionless [7].

User activities and interactions in the tourism domain are analyzed. In particular, the emotions of the users regarding their forthcoming trips are studied with the objective to characterize interdependencies between them. Social network analysis is applied to examine interactions between the users. To capture their emotions, text mining techniques, and sentiment analysis is applied to construct a measure, which is based on free-text comments in a travel forum [8].

3. PROPOSED SYSTEM

The system architecture is given in Figure 1. Each block is described in this Section.

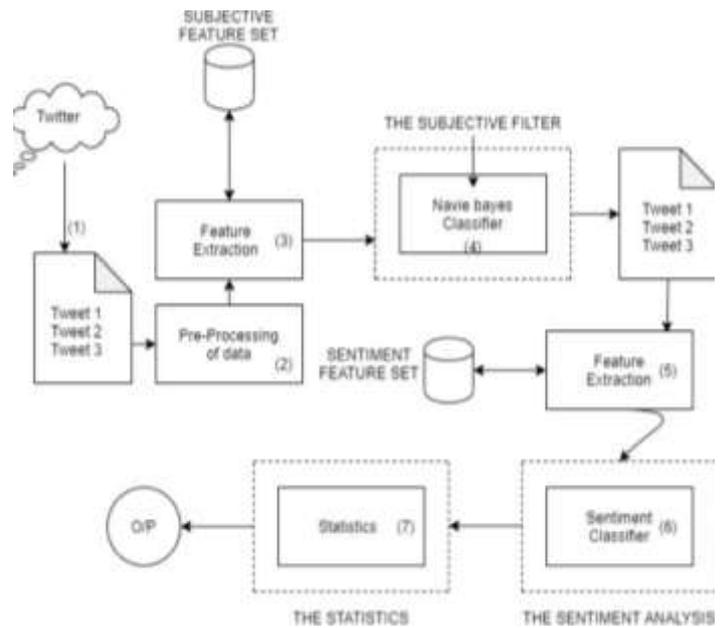


Fig- 3.1: Proposed System Architecture

Data collector: Personalization is a concern about adapting to interests, comments, emoticons, and preferences of users on social networking sites to collect data scrapped by such websites.

Noise filter: The filtered out sentences can be further narrowed down to only some specific words that can be triggering or words that are required to further carry out testing.

Smart filter: The words that are filtered out are then compared with words that are stored and that commend different emotions.

Result viewer: The data that was mined from different social networking sites after being filtered, compared and then sorted according to a specific emotions is then measured down and calculated according to every emotion which is then calculated all together to get a score which will then specify the percentage of each emotion and the ranking of the same.

4. IMPLEMENTATION

4.1 Software

Operating System	Windows XP Professional With Service pack 2
Programming Language	Python
Database	Oracle 9

Table- 4.1: Software details

4.2 Hardware

Processor	2 GHz Intel
HDD	180 GB
RAM	2 GB

Table- 4.2: Hardware details

4.3 Inputs Details

Standard datasets are taken from Tweepy. Tweepy is used to get live data from twitter. The input data will be tweets extracted from the user. Firstly it takes the consumer key, consumer secret, access key and access secret from twitter developer available easy for each user. These keys will help the API for authentication. Then the user is asked to enter the keyword and number of tweets to search about.

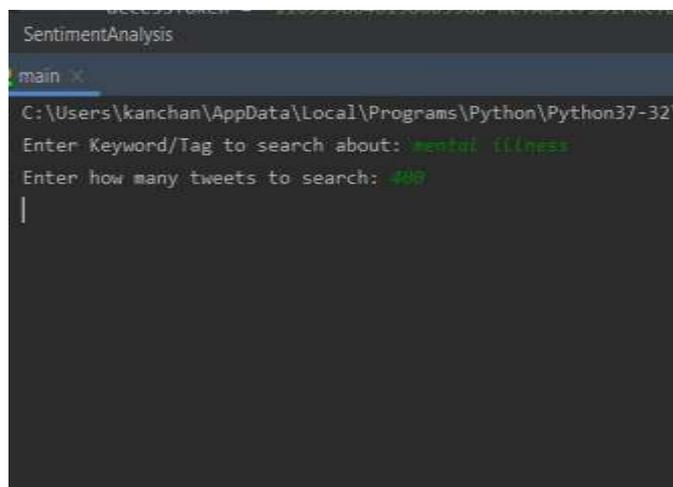


Fig -4.1: Input data

4.4 Output Details

In order to evaluate the proposed system, experiments were conducted on tweets related to the keyword entered by the user. The data was collected from hundreds of users who use twitter to express their views and opinions. Firstly, it shows the authentication of the twitter developer account. It gives the date and time and other basic information about when the account was created. Then it displays the tweets based on the search keyword. It will show the number of tweets the user wants to see.

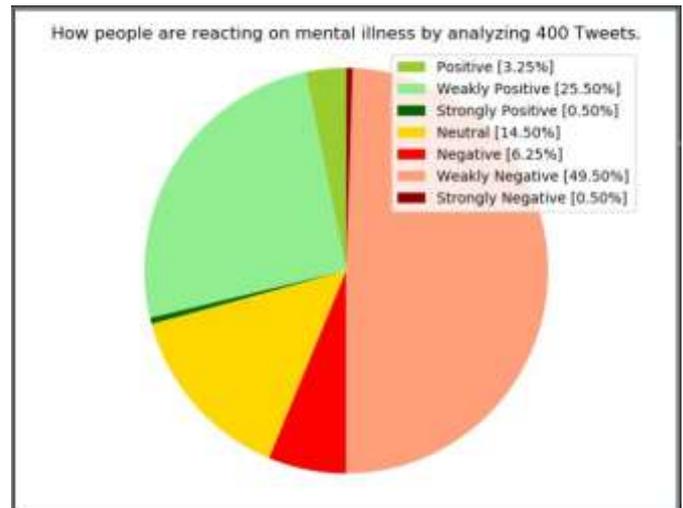


Fig -4.2: Output Data

5. CONCLUSIONS

The application helps up find sentiments related to a specific subject, argument, problem, the issue using keywords associated to topic of interest. The data that is the sentiment of social media users is collected and analyzed. This data can be used for various purposes such as consumer requirement analysis, impact caused on users due to certain issues and so on. The analyzed data then can be used to improve user experiences in the future.

6. ACKNOWLEDGEMENT

We would first like to thank our Project Coordinator and our guide, **Prof. Madhura Vyawahare**, who keeps encouraging us to do such projects which helps improve practical knowledge. We also thank her for solving our queries and helped us through this project.

We also thank our project coordinator, **Prof. Gayatri Hegde** for her guidance on the planning and implementation of our project.

We give special gratitude to our Principal, **Dr. Sandeep Joshi**, who always encouraged and motivated us to do innovative things that will improve our knowledge.

We also thank our H.O.D of Information Technology Department, **Dr. Satishkumar Varma**, for this opportunity of practically implementing the concepts we study, through Project. This would not have been possible without the opportunity. We are thankful to all who provided us an opportunity to complete this project.

REFERENCES

[1] Neethu M S, Rajasree R, "Sentiment Analysis in Twitter using Machine Learning Techniques"(2013)

[2] Shahid Shayya, Noor Ismavati Jaffar, "Sentiment Analysis of Big Data: Methods, Applications, and Open Challenges"(2016)

[3] Young sub-Han," Sentiment Analysis on Social Media Using Morphological Sentences Pattern Model", (2017)

[4] Alexander Pak, Patrick Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining"(2017)

[5] Neethu MS And Rajasree R, "Sentiment Analysis in Twitter using Machine Learning Techniques"(2016)

[6] Chiara Zucco, Barbara Calabrese, Mario Cannataro, "Sentiment Analysis and Affective computing for depression monitoring." (2017)

[7] N. Azmina, Nasiroh Omar, "Sentiment Analysis: Determining people's emotions in Facebook" (2017)

[8] Julia Neidhardt, Hannes Werthner," Predicting happiness: user interactions and sentiment analysis in an online travel form"(2018)

[9] S. Alami, O. Elbeqqali "Detecting Suspicious Profiles Using Text Analysis Within Social Media" Journal Of Theoretical And Applied Information Technology,(2017)