PROJECT REPORT

ON

**TWITTER SENTIMENT ANALYSIS**

2021-2022

Under the Guidance of

**Mr. Akshay Rajput**, Asst Professor

Deptt. Of Computer Science & Application



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING GRAPHIC ERA DEEMED TO BE UNIVERSITY, DEHRADUN

Submitted by:

Aniket

2017462

**CERTIFICATE**

Certified that ANIKET(2017462) has developed a mini project on “ TWITTER SENTIMENT ANALYSIS” for the CSE 4th Semester Mini Project Lab in Graphic Era Deemed to be University, Dehradun. The project carried out by students is their work to the best of my knowledge.

ANIKET

2017462

**ACKNOWLEDGMENT**

Hereby I’m submitting the project report on **TWITTER SENTIMENT ANALYSIS**, as per the scheme of Graphic Era University, Dehradun.

In this connection, I would like to express my deep sense of gratitude to our beloved institution Graphic Era University and also like to express our sincere gratitude and indebtedness to **Prof. (Dr). Kamal Ghansala,** founder of GEU, Dehradun.

I would like to thank particularly **Mr. AKSHAY RAJPUT** sir my mentor for his patience, support, and encouragement throughout the completion of this project and for having faith in me

Finally, I would like to thank all the faculty members of the Department of Computer Science and Application, friends, and my parents for their constant encouragement, support, and help throughout the project's conduction.

ANIKET(2017462)

MINI PROJECT

**TWITTER SENTIMENT ANALYSIS**

**Introduction**

millions of people are using social network sites to express their emotions, and opinions and disclose their daily lives. However, people write anything such as social activities or any comments on products. The online communities provide an interactive forum where consumers inform and influence others. Moreover, social media provides an opportunity for businesses that gives a platform to connect with their customers such as social media to advertise or speak directly to customers for connecting with customers’ perspectives of products and services. In contrast, consumers have all the power when it comes to what consumers want to see and how consumers respond. With this, the company’s success & failure is publicly shared and end up with word of mouth. However, the social network can



change the behavior and decision-making of consumers, for example, mentions that 87% of internet users are influenced in their purchases and decision by customer reviews. So that, if organizations can catch up faster on what their customers think, it would be more beneficial to organize to react on time and come up with an excellent strategy to compete with their competitors.

**TOOLS USED:-**

**LANGUAGE:**

● Python

**SOCIAL MEDIA:**

● Twitter

**LIBRARIES:**

● Numpy

● Pandas

● Seaborn

**PLATFORM FOR DEVELOPMENT:**

● Google colaboratory

**Technology and approach:**

**Opinion Mining:-** Opinion mining refers to the broad area of natural language processing, text mining, and computational linguistics, which involves the computational study of sentiments, opinions, and emotions expressed in a text. Although, a view or attitude based on emotion instead of reason is often colloquially referred to as a sentiment. Hence, lending to an equivalent for opinion mining or sentiment analysis. stated that opinion mining has much application

domains including accounting, law, research, entertainment, education, technology, politics, and marketing. In earlier days many social media have given web users avenues for opening up to express and share their thoughts and opinions

**Twitter Sentiment:-**

Analysis The sentiment can be found in the comments or tweets to provide useful indicators for many different purposes. Also stated that sentiment can be categorized into two groups, which are negative and positive words. Sentiment analysis is a natural language processing technique to quantify an expressed opinion or sentiment within a selection of tweets.

**Lexicon-based Approach:-**

i. Preprocess each tweet, and post by removing punctuation

ii. Initialize a total polarity score (s) equal 0 -> s=0

iii. Check if the token is present in a dictionary, then If the token is positive, s will be positive (+) If the token is negative, s will be negative (-)

iv. Look at the total polarity score of tweet post If s > threshold, tweet post as positive If s < threshold, tweet post as negative

**Machine-learning-based Approach:-**

i. Apply a part of speech tagger to each tweet post

ii. Collect all the adjectives for entire tweet posts

iii. Make a popular word set composed of the top N adjectives

iv. Navigate all of the tweets in the experimental set to create the following:

• Number of positive words • Number of negative words

• presence, absence, or frequency of each word



**Techniques of Sentiment Analysis:-** The semantic concepts of entities extracted from tweets can be used to measure the overall correlation of a group of entities with a given sentiment polarity. Polarity refers to the most basic form, which is if a text or sentence is positive or negative. However, sentiment analysis has techniques for assigning polarity such as:

1. **Natural Language Processing (NLP)**:-

NLP techniques are based on machine learning and especially statistical learning which uses a general learning algorithm combined with a large sample, a corpus, of data to learn the rules. Sentiment analysis has been handled as a Natural Language Processing denoted NLP, at many levels of granularity. Starting from being a document-level classification task, it has been handled at the sentence level and more recently at the phrase level. NLP is a field in computer science that involves making computers derive meaning from human language and input as a way of interacting with the real world.

2. **Case-Based Reasoning (CBR):**-

Case-Based Reasoning (CBR) is one of the techniques available to implement sentiment analysis. CBR is known for recalling the past successfully solved problems and using the same solutions to solve the current closely related problems. identified some of the advantages of using CBR that CBR does not require an explicit domain model and so

elicitation becomes a task of gathering care histories and the CBR system can learn by acquiring new knowledge as cases. This and the application of database techniques make the maintenance of large columns of information easier.

3. **Artificial Neural Network (ANN)**:-

mentioned that Artificial Neural Network (ANN) or known as a neural network is a mathematical technique that interconnects a group of artificial neurons. It will process information using the connections approach to computation. ANN is used in finding the relationship between input and output or to find patterns in data.

4. **Support Vector Machine(SVM)**:-

Support Vector Machine to detect the sentiments of tweets. together with the stated SVM can extract and analyze to obtain upto70%-81.3% of accuracy on the test set. collected training data from three different Twitter sentiment detection websites which 

collected training data from three different Twitter sentiment detection websites which mainly use some pre-built sentiment lexicons to label each tweet as positive or

negative. Using SVM trained from these noisy labeled data, they obtained 81.3% sentiment classification accuracy

**Application Programming Interface(API) :-**

Alchemy API performs better than the others in terms of the quality and the quantity of the extracted entities. As time passed the PythonTwitter Application Programming Interface (API) is created by collected tweets. Python can automatically calculate the frequency of messages being retweeted every 100 seconds, sorted the top 200 messages based on their there-tweeting frequency, and stored them in the designated database. As the Python Twitter API only included Twitter messages for the most recent six days, collected the data needed to be stored in a different database.

**RESULT AND DISCUSSION**

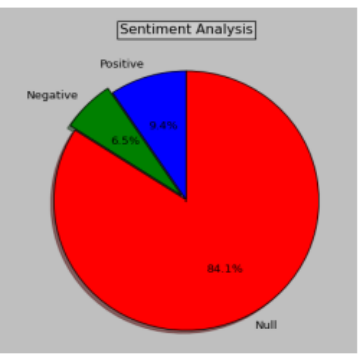
**Twitter Retrieved**:-

To associate with Twitter API, developers need to agree on the terms and conditions of the development Twitter platform which has been provided to get authorization to access data. The output from this process will be saved in a JSON file. The reason is, that JSON (JavaScript Object Notation) is a lightweight data-interchange format that is easy for humans to write and read. Moreover, stated that JSON is simple for machines to generate and parse. JSON is a text format that is language-independent but uses a convention that is known to programmers of the C-family of languages, including Python and many others. However, output size depends on the time for retrieving tweets from Twitter.

**Information Presented:-**

The result will be shown in a pie chart which is representing a percentage of positive, negative, and null sentiment hashtags. For null hashtag is representing the hashtags that were assigned zero value. However, this program can list the top ten positive and negative hashtags.





**CONCLUSION AND RECOMMENDATION**:-

Twitter sentiment analysis is developed to analyze customers' perspectives on the critical to success in the marketplace. The program is using a machine-based learning approach which is more accurate for analyzing a sentiment; together natural language processing techniques will be used. As a result, the program will be categorized sentiment into positive and negative, which is represented in a pie chart and Html page Although, the program has been planned to be developed as a web application, due to the limitations of Django which can only work on Linux server or LAMP. Thus, it cannot be realized. Therefore, further enhancement of this element is recommended in a future study.

6



**REFERENCES**

● G. Kalia, “A Research Paper on Social Madia: An Innovative Educational Tool”, Vol.1, pp. 43-50, Chitkara University, 2013.

● http://ieeexplore.ieee.org/xpls/abs\_all.jsp?arnumber=5428313

**Thank you**