**MINI PROJECT – I**

**(2020-21)**

**Tweet sentiment analysis by ML**

**SYNOPSIS**



**Institute of Engineering & Technology**

**Team Members**

Deepak

(171500092)

Kumar Ashutosh Sharma

(171500169)

**Supervised By**

**Mr. VAIBHAV DIWAN**

Technical Trainer

***Department of Computer Engineering & Applications***

***About The Project:-***

This project addresses the problem of sentiment analysis in twitter; that is classifying tweets according to the sentiment expressed in them: positive, negative or neutral. Twitter is an online micro-blogging and social-networking platform which allows users to write short status updates of maximum length 140 characters. It is a rapidly expanding service with over 200 million registered users - out of which 100 million are active users and half of them log on twitter on a daily basis - generating nearly 250 million tweets per day. Due to this large amount of usage we hope to achieve a reflection of public sentiment by analysing the sentiments expressed in the tweets. Analysing the public sentiment is important for many applications such as firms trying to find out the response of their products in the market, predicting political elections and predicting socioeconomic phenomena like stock exchange. The aim of this project is to develop a functional classifier for accurate and automatic sentiment classification of an unknown tweet stream.

***Problem Statement:-***

1. A major benefit of social media is that we can see the good or bad things people say about the particular brand or personality
2. The bigger your company gets difficult it becomes to keep a handle on how everyone feel about your band for large companies with thousands of daily mentions on social ,media and news sites, it’s extremely difficult to do this manually
3. To combat this problem, sentimental analysis software are necessary. The soft wares can be used to evaluate the people’s sentiment about particular or personality.

***Motivation:-***

We have chosen to work with twitter since we feel it is a better approximation of public sentiment as opposed to conventional internet articles and web blogs. The reason is that the amount of relevant data is much larger for twitter, as compared to traditional blogging sites. Moreover the response on twitter is more prompt and also more general (since the number of users who tweet is substantially more than those who write web blogs on a daily basis). Sentiment analysis of public is highly critical in macro-scale socioeconomic phenomena like predicting the stock market rate of a particular firm. This could be done by analysing overall public sentiment towards that firm with respect to time and using economics tools for finding the correlation between public sentiment and the firm’s stock market value. Firms can also estimate how well their product is responding in the market, which areas of the market is it having a favourable response and in which a negative response (since twitter allows us to download stream of geo-tagged tweets for particular locations. If firms can get this information they can analyze the reasons behind geographically differentiated response, and so they can market their product in a more optimized manner by looking for appropriate solutions like creating suitable market segments. Predicting the results of popular political elections and polls is also an emerging application to sentiment analysis. One such study was conducted by Tumasjan et al. in Germany for predicting the outcome of federal elections in which concluded that twitter is a good reflection of offline sentiment

***Description:-***

Sentiment Analysis is a technique widely used in text mining. Twitter Sentiment Analysis, therefore means, using advanced text mining techniques to analyze the sentiment of the text (here, tweet) in the form of positive, negative and neutral. It is also known as Opinion Mining, is primarily for analyzing conversations, opinions, and sharing of views (all in the form of tweets) for deciding business strategy, political analysis, and also for assessing public actions.

Enginuity, Revealed Context, Steamcrab, MeaningCloud, and SocialMention are some of the well-known tools used for the analysis of Twitter sentiment. R and Python are widely used for [sentiment analysis](https://www.digitalvidya.com/blog/sentiment-analysis/) dataset twitter.

As a part of [Natural Language Processing](https://www.digitalvidya.com/blog/natural-language-processing-with-python), algorithms like SVM, Naive Bayes is used in predicting the polarity of the sentence. sentiment analysis of Twitter data may also depend upon sentence level and document level

***TECHMOLOGY USED:-***

1. ***Python :-***

Python is a high-level, general-purpose and a very popular programming language. Python programming language (latest Python 3) is being used in web development, Machine Learning applications, along with all cutting edge technology in Software Industry. Python Programming Language is very well suited for Beginners, also for experienced programmers with other programming languages like C++ and Java.

1. ***Machine Learning:-***

**Machine learning** (**ML**) is the study of computer [algorithms](https://en.wikipedia.org/wiki/Algorithm" \o "Algorithm) that improve automatically through experience and by the use of data.It is seen as a part of [artificial intelligence](https://en.wikipedia.org/wiki/Artificial_intelligence" \o "Artificial intelligence). Machine learning algorithms build a model based on sample data, known as "[training data](https://en.wikipedia.org/wiki/Training_data" \o "Training data)", in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as [email filtering](https://en.wikipedia.org/wiki/Email_filtering" \o "Email filtering) and [computer vision](https://en.wikipedia.org/wiki/Computer_vision" \o "Computer vision), where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks

1. ***Data Set:-***

A data set (or dataset) is a collection of [data](https://en.wikipedia.org/wiki/Data" \o "Data). In the case of tabular data, a data set corresponds to one or more [database tables](https://en.wikipedia.org/wiki/Table_(database)" \o "Table (database)), where every [column](https://en.wikipedia.org/wiki/Column_(database)" \o "Column (database)) of a table represents a particular variable, and each [row](https://en.wikipedia.org/wiki/Row_(database)" \o "Row (database)) corresponds to a given record of the data set in question. The data set lists values for each of the variables, such as height and weight of an object, for each member of the data set. Each value is known as a datum. Data sets can also consist of a collection of documents or files.

***Requirements and Specifications:-***

1. ***Hardware Requirements:***
2. Processor : Intel dual Core ,i5
3. RAM : 4GB
4. Hard disk : 500 GB
5. ***Software Requirements: -***
6. Windows 7,8,10
7. Google Colab,Project Jupyter

***Applications: -***

1. Twitter sentiment analysis allows you to keep track of what’s being said about your product or service on social media, and can help you detect angry customers or negative mentions before they turn into a major crisis.
2. At the same time, Twitter sentiment analysis can provide interesting insights. What do customers love about your brand?  What aspects get the most negative mentions?

***Advantages: -***

1. You can develop a more insightful, data-based marketing strategy
2. Understood your customers
3. Take a look at brand perception
4. Measure your marketing campaign
5. Find industry leaders and influencers

***Disadvantage:-***

**Sentiment analysis** of short texts such as single sentences and **Twitter** messages is challenging because of the limited contextual information that they normally contain. Effectively solving this task requires strategies that combine the small text content with prior knowledge and use more than just bag-of-words.

***Future Scope :-***

1. Data Pre - Processing using more parameter to get best sentimental
2. Updating Dictionary for new synonym and antonyms of already existing words.
3. Context Sentimental Analysis may be implemented in future for accuracy purposes

***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.

**THANK YOU**