International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 08 | Aug 2022 www.irjet.net p-ISSN: 2395-0072

Emotion Recognition By Textual Tweets Using Machine Learning

Vijaykumar G[1]

Dept of MCA, Vidya Vikas Institute of Engineering And Technology, Karnataka, India

---

Abstract - Opinion mining has become difficult due of the abundance of user-generated content on social media. Twitter is used to gather opinions about products, trends, and politics as a microblogging site.

Sentiment analysis is a technique used to examine people's attitudes, feelings, and opinions toward anything. It may be applied to tweets to examine how the public feels about news, legislation, social movements, and political figures. Natural language processing and machine learning are both regarded as having a category called sentiment analysis. It is used to separate, identify, or represent views from various information structures, such as news, audits, and articles, and it classifies them as positive, neutral, or negative. From tweets in several Indian languages, election results are tough to forecast.

To get tweets in Hindi, we used the Twitter Archiver programme. We used data (text) mining to examine 48,276 tweets that mentioned five national political parties in India over the course of a period of time. Both supervised and unsupervised methods were applied.

***Key Words*: Naive Bayes, SVM, Decision Tree, Long Short-Term Memory, and NRC Lexicon Emotion.**

1. INTRODUCTION

In today's environment, text or opinion mining is useful for gauging public opinion of recently released goods, such as movies, songs, books, and other media. It also distinguished between recommendations and opinions that were good, negative, and neutral. The general people are now accustomed to posting their feelings about the political leader on social media. In order to learn about the political leaders' opinions and engage the public through TV programmes, YouTube, etc., many reporters have conducted interviews with them.

The effort of using surveys and polls to research people's opinions is very time- consuming and expensive. Sentiment analysis is a type of data mining technique that employs NLP to determine the prevailing opinion.

It is the practice of categorizing viewpoints into three groups, such as "positive," "negative," and "neutral." This data quantifies public reactions to certain people, organizations, and political discourses, showing the environmental orientation of the data. Consequently, based on social media tweets, our goal is to examine how online users feel about each political party, its leaders, and its actions.

1.1 Objectives

The project's primary goal is to predict India's five national political parties. To do this, we combined supervised and unsupervised methods. We built our classifier using Naive Bayes, a decision tree NRC Lexicon, and the SVM method, and we categorized the test data as positive, negative, neutral, and eight other categories of emotions.

➢ Building an approach that will be used to predict election outcomes using emotions.

➢ Creating a framework that is simple to use.

1.2 Scope

Opinion mining becomes challenging since there is so much customer content on social media. Twitter is used to gather opinions about customers, trending products, and political opinion as a microblogging site. Sentiment analysis is a method for examining people's attitudes, feelings, and opinions toward various topics. It may be applied to tweets to examine how the public feels about various topics, including news, policy, social movements, and individuals.

2. Existing System

Corporations are motivated by sentiment analysis to identify consumer preferences for brands, products, and services. Additionally, it is crucial in evaluating data on businesses and sectors to keep them in mind when conducting entity reviews. By extracting a large number of tweets with the aid of prototypes, Sarlan et al.

He built a sentiment analysis, and the results categorized customers' thoughts expressed in tweets into negative and positive categories. They separated their research into two parts. The primary section is depend on a literature review and uses current methodologies and methods for sentiment analysis. The second section describes the operations and requirements of the application before it is developed.

Disadvantages

➢ Extraction keyword is improper.

➢ POS tagging is not incorporated for calculation of tweet weights.

➢ Parallel processing time is high.

➢ Lexicon based content parsing is traditional method and leads to less accuracy.

3. Proposed System

In this research, different techniques have been used for methodology in Machine Learning (ML) for its objectives. Various approaches and procedures were used to assess versatile experiments. The Twitter dataset used in this experiment was scrapped from the

Kaggle repository and subjected to many classifier applications. The dataset is first pre- processed by deleting unnecessary data. The data was then divided into a training set and a testing set. The test set component is 20%, whereas the training set was given an allocation of 80%. The training set is then used to apply feature engineering approaches. On the training set and test set, various machine learning classifiers are trained.

Advantages of Proposed Methodology:

➢ Stochastic processing of tweets provides hight rate of classification of keywords from tweet.

➢ POS tagging implementation for calculation of tweet weights.

➢ Parallel tweet processing for improvisation of accuracy.

4. System Design

Machine learning systems design is the process of defining the software architecture, infrastructure, algorithms, and data for a machine learning system to satisfy particular requirements. By outlining the intricacies of how the programme should be created, the software design will be used to assist in the development of software for web apps.

It is customary to begin with context-level records before moving on to the flow diagram, which shows how the system interacts with external suppliers who serve as data assets and hyperlinks. Simple record flow models are used to represent the machine's interactions with the outside world using context diagrams.

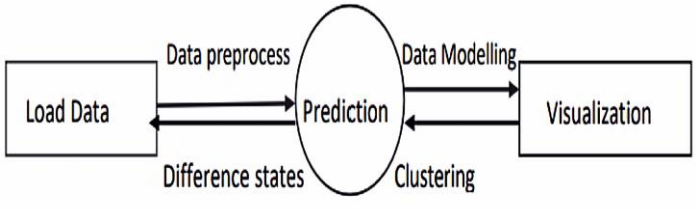

Fig 4.1Context diagram

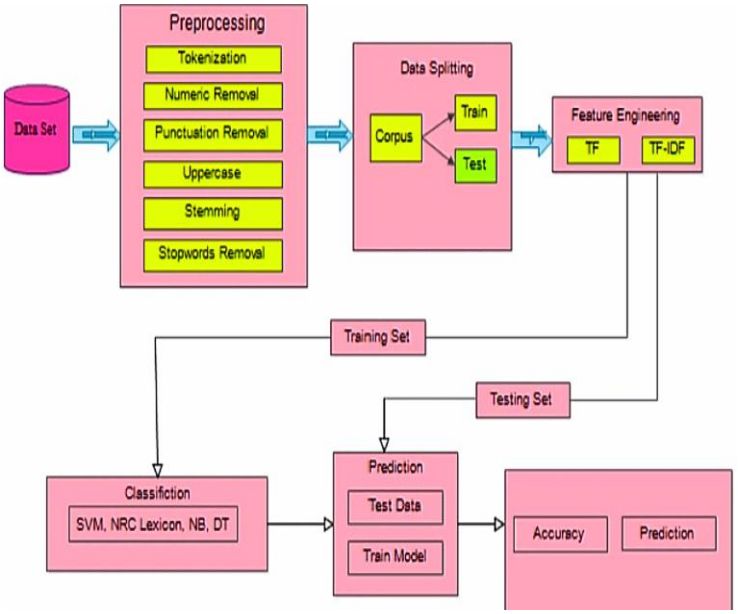

Fig: 4.2 Architecture Design

5. Detailed Design

By describing the specifications of how the programme should be created, the software design will be used to assist in the development of software for web apps. The software design specifications, which are literary and graphical documentation of the software design for the project, comprise use case models, sequence diagrams, and other supporting requirement data.

The design document is for a fundamental system that will serve as solid evidence for the development of a system that offers a fundamental level of capacity to demonstrate that it can be successfully employed on a broad scale for production purposes.

5.1 Use Case Diagram

A use case diagram is made up of a conversation relationship between the actor and the use case as well as an actor graph, a hard list of use cases, and a device boundary. Each use case is a specific piece of functionality that a machine offers to its users, and the use case diagram explains how a device interacts with outside actors. An actor is represented as a stick figure with the call of the actor beneath the parent, and a use case is represented as an ellipse holding the use case's call.

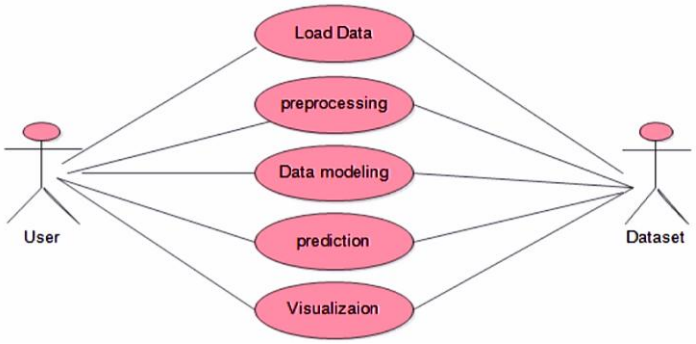

Fig: 5.1 Use Case Diagram

5.2 ACTIVITY DIAGRAM

An activity diagram uses graphics to represent the flow of control or the order of occurrences within a system. Activity diagrams are often used in business process modelling. In a use case graphic, they can also describe the steps. Sequential and concurrent activities can both be modelled. An activity diagram will always have a start (also known as an initial state) and an end (a final state).

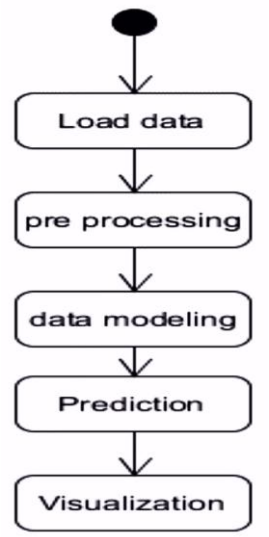

Fig: 5.2 Activity Diagram

6. Implementation

The programming language employed for this project is Python. Python has dynamic typing and garbage collection. The programming paradigms that are supported include procedural, object-oriented, and functional programming, to name just a few. Python is frequently referred to as a "batteries included" language because of its extensive standard library. Machine learning methods are used in this project.

Sentimental analysis extracts and identifies subjective information from a text or sentence by using context-based text mining. Here, the key idea is to use machine learning methods like LSTM to extract the sentiment from the text (Long short-term memory). This text categorization approach examines the incoming text to identify whether the emphasized emotion is positive or negative, as well as the likelihood that the positive or negative claims are true.

6.1 ALGORITHMS IMPLEMENTATION:

6.1.1 NRC Lexicon:

Eight fundamental emotions (angry, joy, sadness, fear, trust, surprise, anticipation and disgust) and two sentiments are listed in the NRC Emotion Lexicon (negative and positive).

Crowdsourcing was used to hand annotate the images.

6.1.2 SVM Algorithm:

It is believed that the Support vector machine (SVM) effectively does sentiment analysis. SVM characterizes preference, restricts and employs procedures for evaluation, and evaluates records that are obtained inside the index region. All magnitude's vector arrangements include significant information. To accomplish this, the information (represented as a vector) has been sorted by kind. The border is then stratagem-classified in two training sessions.

6.1.3 Decision Tree:

The supervised ML approach known as the decision tree algorithm is frequently employed in applications involving regression and classification. The main issue, known as attribute selection, is choosing the root node of a tree at each level. The most popular techniques for attribute selection are knowledge gain and the Gini index. In this study, the Gini index is used to calculate the likelihood of the root node by squaring the attribute value total and subtracting one.

6.2 Dataset Table

Sl no	Attributes	Attribute type	Description
01	created_at	Integer	Date and time of tweet creation.
02	tweet_id	Integer	Unique ID of the tweet.
03	tweet	String	Full tweet text.
04	likes	Integer	Number of likes.
05	retweet_count	Integer	Number of retweets.
06	user_id	Integer	User ID of tweet creator.
07	user_name	String	Username of tweet creator.
08	user_description	String	Description of self by tweet creator.
09	user_followers_count	Integer	Followers count on tweet creator.
10	user_location	String	Location given on tweet creator's profile.

Table:6.2 Dataset Table for Election Result Prediction

7. TESTING

This chapter gives the various test cases performed to check for the effective execution of the venture. Testing is a procedure of cross verification of the designed system model under active state and various inputs. This process is carried out in various ways. The main objective of software development life cycle is to produce a product with no errors or very few errors. In

the processes of achieving hassle free software we plan testing and test cases. Software testing is done for the success of the application. The testing is done mainly to check whether the product meet the requirement of the user properly. It is used to check the bugs and errors in the system or to find out the defects of the system.

7.1 Test Case Scenario on Emotion:

Test Case Number	Testing Scenario	Expected result	Result
01	No dataset is provided	Alert "provided an input file"	Pass
02	Dataset with no data	No –databased visualize	Pass
03	Testing Machine learning algorithm	The output predicts types of emotion in a particular party.	Pass
04	Clustering on dataset	The output should cluster successfully based on different parties.	Pass
05	Testing of accuracy	The output predicts the emotions in a Particular party	Pass
06	Integration testing	The output predicts the emotions in a Particular party	Pass
07	Dataflow testing	The output predicts the emotions in a Particular party	Pass
08	User interface testing	The output predicts the emotions in a Particular party.	Pass
09	Event based user interface testing	The output predicts the emotions in a Particular party	Pass
10	Classification Module	Should be classify successfully based on emotions tweet by user	Pass
11	Extraction Module	It will Extract the data from the dataset.	Pass

Table 7.1: Test Case Scenario on Emotion

7.2 SYSTEM TESTING:

A framework's modules are merged for execution after each one has undergone thorough testing. At that point, it is necessary to undertake top-down testing, which begins with upper-level modules and moves down to lower-level ones, to determine whether the entire framework is operating as intended.

TC ID	Test Case Description	Inputs	Expected Output	Actual Output	Status
01	Executing the Application	Running the application with required inputs	Application should run Without any interrupts	Runs successfully without any interrupts	Pass
02	Connection String	Clicking on every button in the application	Every form data should be displayed	Displays form data	Pass

Table 7.2 System Testing

CONCLUSIONS

Managing client interactions, human computer interface, information retrieval, more natural text-to-speech systems, and sociological and literary analysis are just a few of the practical uses for emotion detection and creation. There are, however, very few resources with restricted coverage for emotions, and those are only available in English. In this study, we demonstrate how a large term-emotion association vocabulary may be produced swiftly and affordably by harnessing the strength and collective knowledge of the masses. There are entries for more than 10,000 word-sense pairs in this lexicon, called EmoLex. Each entry lists a word- sense pair's associations with the eight fundamental emotions.

REFERENCES

1. Jyoti Ramteke, Samarth Shah, "Election result prediction using Twitter sentiment analysis"

2. Febby Apri Wenando; Regiolina Hayami, "Tweet Sentiment Analysis for 2019 Indonesia Presidential Election Results using Various Classification Algorithms"

3. Rincy Jose; Varghese S Chooralil, "Prediction of election result by enhanced sentiment analysis on Twitter data using Word Sense Disambiguation"

4. Dimosthenis Beleveslis; Christos Tjortjis, "A Hybrid Method for Sentiment Analysis of Election Related Tweets"

5. Haji Binali; Chen Wu; Vidyasagar Potdar, "Computational approaches for emotion detection in text"

6. Umar Rashid; Muhammad Waseem Iqbal, "Emotion Detection of Contextual Text using Deep learning"

7. Shadi Shaheen Wassim El-Hajj," Emotion Recognition from Text Based on Automatically Generated Rules"