2020 International Conference for Emerging Technology (INCET)
Belgaum, India. Jun 5-7, 2020

Twitter Sentiments Analysis Using Machine Learninig Methods

Lokesh Mandloi
B.Tech Computer Science and Engineering
Meidcaps University
Indore, India
lokeshmandloi@outlook.com

Ruchi Patel
Assistant Professor
Dept. Computer Science and Engineering
Medicaps University
Indore, India
ruchipatel294@gmail.com

***Abstract*— Analysis of sentiments is the method of deciding whether the sentiments in the text is positive, negative or neutral. It is also known as material polarity or mining of opinions. The growth and advancement in social media platforms engaged a huge number of users. Social media platform like twitter where users can post their tweets in 280 characters. Because of the limited number of characters in tweets, it becomes easy for the sentiment analysis. On Twitter 550 millions of tweets are posted daily. Twitter also represents all age group people and also a fair representation of gender. Therefore, the sentiment analysis of twitter data becomes somewhat general sentiments of society. In this paper, we will compare various Machine Learning methods like the Naïve Bayes Classification method, Support Vector Machine Classification Method and Maximum Entropy Classification method. We will see how sentiments analysis is done by this classification algorithm and what is the accuracy and precision in these cases.**

Keywords*— *Sentiments Analysis, Opinion Mining, Twitter Sentiments Analysis.

I. INTRODUCTION

Twitter sentiments analysis make use of Natural Language Processing to evaluate a speaker's, writer's, or other person's mood and emotions through the piece of text. Through Sentiments Analysis, we can determine if a tweet of a user is positive negative or neutral. Social networking platforms such as Twitter, Facebook, Instagram, YouTube, etc. have been so popular for days now. They allow people to communicate, create networks, and share thoughts easily and promptly. Twitter has become an excellent medium for opinion creation and presentation. Twitter Sentiments Analysis can be used for real-time applications which can be a very helpful business. It can be used for people's sentiments on current political topics or trends. It can also be used for the review of movies using the trends on twitter [1].

As the audience on the media platforms grows continuously data from these sites can be used to analyses the sentiments of the people.

- Manufacturers or developers of the products of the can review their product by analyzing the sentiments of the people. That is how people reacting to their products.
- Marketing personal can see how people are reacting to their advertising campaign. They can analyze the sentiments related to this.
- Political parties can see how their political campaign is running and how people reacting to it. They can analysis which issue to be raised to not.
- Filmmakers can see how people are reacting to their newly released movie, by analyzing the sentiments of the people.

There are many reasons why we chose twitter data for sentiments analysis some of the reasons are given below:

- Twitter is used varied from regular people to actors, politician, businessmen, and various religious and social leader to post their opinion.
- The number of tweets on twitter daily is more than 500 million and that is enormous data for sentiments analysis.
- Twitter users range from daily users to celebrities, from business executives to political figures. So twitter reflects views of all groups.
- Twitter represents people of all age group and a high percentage of the business person is present on twitter. And people from many countries are present on twitter.
- Twitter application has more than 50 million downloads and used by many on web browser.

A popular use for this technology comes from its implementation in the social media field to explore how people feel about certain topics, specifically through the word-of-mouth of social media users in text posts or their tweets in the context of Twitter.

In this paper, we use different machine learning methods to analyze the sentiments of the people. Here we use machine learning methods like Naïve Bayes Classifier, Support Vector Machine method and Maximum Entropy method. Here we will compare these methods based on their accuracy and precision and see which method gives the best result. All the above methods are supervised learning methods. So, in all these cases we need to first train the data.

II. RELATED WORK

The word Sentiment has three layers of meaning:

- Opinion layer.
- Emotion layer.
- Idea colored by emotion layer.

978-1-7281-6221-8/20/$31.00 ©2020 IEEE

So from the meaning of Sentiments analysis we can say that it is the study of emotions of the user or people. If we go by the definition of Liu then, we say sentiment analysis is the field of research that analyzes the thoughts, feelings, perceptions, behaviors, and emotions of individuals towards things like such as goods, products, political parites, people, problems, incidents, issues, and their attributes. Farzindar distinguishes the study of feelings and the study of emotions to accentuate the slight difference. Examination of the emotions is more precisely categorized into minor details. Emotion is divided into six classes: rage, frustration, anxiety, happiness, and sorrow, excitement, most widely used in the literature [2]. There is presently no agreement about how many emotional groups should be included. Examination of the emotion is also called identification of moods. The distinction between the study of sentiments and the study of emotions goes over the range of this study. We can categorise Sense into various components like: holder, taget, dimension, and polarity. The growing part suits different tasks within a system. Holder denotes the person bearing the emotion.

A target defines the chosen person as the source of the emotion. A target determines the individual chosen as the origin of the sentiment. Polarity can be described in both positive and negative aspects, or can be represented in three positive, negative and neutral aspects. Feature defines the particular aspect or attribute of the objective to which the emotion is expressed. Take the following example1: Samsung Phone cost lower than ten thousand, does not gives a good performance. Aspect is also an essential part of sentiment analysis.

- Different levels of analysis are given below:

Analysis of sentiment can be classified according finer details of the text. Past research focuses primarily on the various levels given below levels:

- Document-level sentiments analysis:

In this level analysis is done to assess whether the sentiment conveyed in an entire text is positive, negative or neutral. Take an instance, provide product feedback, the system will be able to assess the overall polarity of sentiment. Sentiments of the level of the document imply that a fragment of text communicates feelings about a particular target. Although this is usually appropriate for products analysis, film reviews, hotel reviews, etc., it does not extend to circumstances where several targets are evaluated in a document [6].

- Sentence level sentiments analysis:

In this level we analyzes if the views expressed in a sentence are positive, negative or neutral. Sentence level sentiments analysis can be done in two ways. The first way to classify the sentiments in three different tags that are positive, negative or neutral. And the second method is to find the subjective of the sentence to differentiate between the text which has already been classified and which haven't been classified and then mark them with tags such as positive, negative or neutral. The problems in determining the sentence level is that each sentence is related to semantically and syntactically to some other part of the text. This role, therefore, requires contextual knowledge, local as well as global.

- Aspect level sentiments analysis:

Evaluation of sentiments at the aspect-level is when the aspects are extracted from the text using various mechanisms and then the sentiment is analyzed for each subject. This can also be defined as a study of emotions at the feature level. You may decide the feelings of more than one individuals present in one sentence. These can be of three types which are given in brief below:

- Extracting aspects of target,
- Determining aspects-wise polarity,
- Summarizing the overall analysis.

III. SENTIMENTS ANALYSIS

Sentiments Analysis is a task which mainly focuses on textual data and we expect there to be a huge amount of text data. This data is processed and analyses for sentiments that are expressed. Our training set also required a lot of text data from twitter to analyses.

A. Data Gathering

Social Media Site like twitter uses a source of text. Twitter API is used to gather text data. So, for preparing the test set we need to perform the following task:

Register the twitter application for getting our credentials. First of all, make a twitter account and through it register to the twitter developer account. On twitter developer account citing the valid reason for our academic research creates an app. When we complete the creation of the app it will provide us with some keys like consumer key and access token key and their secret key. Using these keys in our program we can access data for our project work.

B. Pre-Processing

After collecting the textual data from twitter, the next step is pre-processing. That is implemented in python. There are several steps involved in pre-processing they are as follows:

Conversion of upper-case letter to lower case latter example: TOPPER to topper

- Tokenization: -

  Tokenization is done with the help of installing the NLP (Natural Language Processing) package. It means the removal of Hash Tags, and the conversion of text to token. And removing Numbers like(1, 2, 3, 4...) URL and Targets (@) etc.

- Removal of non- English words

  Twitter supports many languages but since our project deals with English so, we remove non-English words.

- Emotion Replacement

  It is very important for evaluating the Sentiments of the users. Hence, we substitute the emotional terms with their polarity by contrasting them in the dictionary of emotions.

- Removal of stop words

  For sentiments analysis we need to remove the stop words like (a, an, the… etc.) Which does not play any significant role in sentiments analysis.

C. *Feature Extracting*

The selection of the words which are useful from the tweet of the user is called feature extraction. In feature extraction we extract the aspect from, pre-processed twitter data set.

- There are three types of features mainly unigram, bigram, and n-gram feature.
- POS i.e. Parts of speech tags like adjective, adverb, verb, and noun is a good indicator of subjectivity and sentiments analysis.
- One of the difficult things to interpret is negation which usually changes the polarity of the content

D. *Feature Selection*

For sentiment analysis, appropriate feature selection techniques are used which have a significant role to play for defining relevant attributes and increasing classification (machine learning) accuracy. They are categorized into four major categories namely: -

- Natural Language Processing (NLP)
- Statistical based
- Clustering Based
- Hybrid based

E. *Classifying Methods*

There are several Machine Learning Classifying method we will apply here are:

- Naïve Bayes Classifier
- Support Vector Machine
- Maximum Entropy Method

IV. Different Machine Learning Methods for Sentiments Analysis

There are different machine learning method to analyses the sentiment of the tweets of the users but here we will use three methods which are as follows: -

A. *Naïve Bayes Classifier*

Naïve Bayes is the supervised machine Learning algorithm which uses Bayes theorem for classification problems. It is mostly used in text classification which includes a dataset for high-dimensional training. It is one of the simplest and most powerful classification algorithms that helps to create fast machine learning models that can make predictions quickly. It is mostly used for text analysis, sentiments analysis and classifying articles. It make use of following Bayes theorem [5]:-

$$Prob(A/B) = (Prob(B/A)*Prob(A))/Prob(B)$$

Where:

- Prob(A|B) is know as (posterior probability) of hypothesis A which occurs when some condition is already provided.
- Prob(B|A) it is also called likelihood probability it is the probability of evidence E when we presumes that given hypothesis is true.
- Prob(A) it is the prior known probability of A and does not include any condition.
- Prob(B) it is the prior known probability of A and does not include any condition.

B. *Support Vector Machine (SVM)*

SVM is a supervised machine learning method which is used for both classifications as well as a regression problem. Classification and regression both are a subcategory of supervised machine learning. Classification is something that can be defined as predicting a label whereas Regression is about predicting quantity. So the main task of the Support Vector Machine classifier is to perform classification. That is it classifies the data in different classes by drawing a hyperplane them, which differentiate between different classes which we plot in n-dimensional space [4].

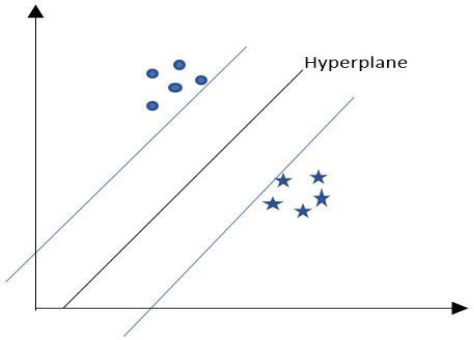

Fig. 1. Represent Support Vector Machine

The hyperplane that is drawn by the SVM is drawn with the help of a mathematical function called Kernels. The data point which is closest to the hyperplane is called the support vector and the method is called the support vector machine.

C. *Maximum Entropy Method*

Maximum Entropy is also a Supervised Machine Learning method. It is used very well in many applications of natural language processing. Sometimes it gives better performance than Naïve Bayes Classification in text classification. Maximum Entropy Classifier is the probabilistic classifier. It is distinguished from the Naïve Bayes classifier because Naïve Bayes considers that the events are independent from each other while the Maximum Entropy approach does not presume that the events are independent. It selects the data which best fits the training data and has the maximum entropy among them. It can be used for a different problem like text classification, sentiment analysis, language detection and image classification. Mostly we use It when we do not have much information about the prior distribution and it is very risky to make any prior

assumption. It required more time as compared to the Naïve Bayes method to train the data set. [8, 9].

V. Result And Discussion

The accuracy and precision of the different machine learning from above are follows:

A. Formula for Calculation of Accuracy and Precision

So, based on our study we have divided the tweets from the users on the basis of the sentiments that is positive, negative or neutral. Here we have created a contingency matrix for reference to our below tabular data.

TABLE I. Contingency Matrix

Actual Value				
Predicted Value		Positive Tweets	Negative Tweets	Neutral Tweets
	Positive Tweets	True Negative	False Positive	False Positive
	Negative Tweets	False Negative	True Negative	False Negative
	Neutral Tweets	False Neutral	False Neutral	True Neutral

Here, True Positive are the positive tweets which are actually classified positive. Where else, False Positive are the tweets which are positive but, classified as negative or neutral. Similarly True Negative are the negative tweets which are classified as negative where else False Negative are the Negative Tweets which are classified as positive or neutral tweets. True Negative are the tweets which are actually found neutral and classified as neutral where else False Negative are the tweets which are neutral tweets but, classified as positive or negative.

1) Precision Calculation

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

2) AccuracyCalculation

$$\text{Accuracy} = \frac{\text{Number of currect Predicted data}}{\text{Total Number of data}}$$

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative} + \text{True Neutral}}{\text{Total}} * 100$$

B. Accuracy and Precision Calculation in Naïve Bayes

Based on our study we found following data According to the contingency matrix table form above Table 1.

TABLE II. Predicted Value for Naïve Bayes

Actual Value					
Predicted Value		Positive	Negative	Neutral	Total
	Positive	612	32	46	690
	Negative	13	123	17	153
	Neutral	14	18	125	157
	Total	639	173	188	1000

Bases on the Above Table 2. We calculate Accuracy and Precision for Naïve Bayes.

$$\text{Accuracy} = \frac{612 + 123 + 125}{1000} * 100$$

Accuracy = 86%

$$\text{Precision} = \frac{612}{690} * 100$$

Precision = 88.695652 %

C. Accuracy and Precision is Support Vector Machine

Based on our study we found following data According to the contingency matrix table form above Table 1.

TABLE III. Predicted Value for Support Vector Machine

Actual Value					
Predicted Value		Positive	Negative	Neutral	Total
	Positive	516	41	123	680
	Negative	18	114	29	161
	Neutral	21	22	116	159
	Total	555	177	268	1000

Bases on the Above Table 3 we calculate Accuracy and Precision for Support Vector Machine

$$\text{Accuracy} = \frac{516 + 114 + 116}{1000} * 100$$

Accuracy = 74.6%

$$\text{Precision} = \frac{516}{680} * 100$$

Precision = 75.88235235%

D. Accuracy and Precision in Maximum Entropy Method

Based on our study we found following data According to the contingency matrix table form above Table 1.

TABLE IV. PREDICTED VALUE FOR MAXIMUM ENTROPY METHOD

Actual Value					
		Positive	Negative	Neutral	Total
Predicted Value	Positive	564	37	70	671
	Negative	17	108	16	141
	Neutral	18	16	154	188
	Total	599	161	240	1000

Bases on the Above Table 4 we calculate Accuracy and Precision for Naïve Bayes.

$$\text{Accuracy} = \frac{564 + 108 + 154}{1000} * 100$$

Accuracy = 82.6%

$$\text{Precision} = \frac{564}{671} * 100$$

Precision = 84.0536512%

E. Comparison of Accuracy and Precision in Machine Learning Methods

TABLE V. COMPARISON OF ACCURACY AND PRECISION OF DIFFERENT MACHINE LEARNING METHOD

Name of the Method	Accuracy	Precision
Naïve Bayes Classifier	86	88.695652
Support Vector Machine	74.6	75.88235235
Maximum Entropy Method	82.6	84.0536512

VI. CONCLUSION

The different machine learning technique of data analysis of twitter are discussed like Naïve Bayes, SVM and Maximum Entropy Method. The analysis of twitter data is being done in various aspects to mine the sentiments. This study defines the concept of opinion in sentiment analysis of Twitter. Sentiment analysis deals with opinion classified into positive, negative and neutral. The study shows that the machine learning method such as Naïve Bayes has the highest accuracy and can be consider as the baseline learning methods as well as in some cases Maximum Entropy methods are very effective. More work in future is needed to improve the performance measures.

REFERENCES

[1] Evolutionary Machine Learning Techniques: Algorithms and Applications (Algorithms for Intelligent Systems) by Seyedali Mirjalili, Hossam Faris, Ibrahim Alijarah.

[2] Sentiment Analysis Using Support Vector Machine 1Aamera Z. H. Khan, 2Dr. Mohammad Atique, 3Dr. V. M. Thakare in International Journal of Advanced Research in Computer Science and Software Engineering(http://ijarcsse.com/Before_August_2017/docs/papers/Special_Issue/ITSD2015/25.)

[3] Twitter Sentiment Analysis Using Support Vector Machine and K-NN Classifiers International Journal of Scientific and Research Publications(http://www.ijsrp.org/research-paper-1018.php?rp=P827862)

[4] SENTIMENT ANALYSIS OF TWEETS USING SUPPORT VECTOR MACHINE Suman Rani1 , Jaswinder Singh Suman Rani et al, International Journal of Computer Science and Mobile Applications, Vol.5 Issue. 10, October- 2017, pg. 83-91

[5] Sentiment Analysis using Maximum Entropy Algorithm in Big Data Durgesh Patel , Sakshi Saxena , Toran Verma, International Journal of Innovative Research in Science, Engineering and Technology(http://www.ijirset.com/upload/2016/may/246_49_Sentiment.pdf)

[6] International Research Journal of Engineering and Technology (IRJET) Volume: 03 Issue: 08 | Aug-2016 www.irjet.net, Sentiment Analysis Using SVM and Maximum Entropy Snehal L. Rathod, Sachin N.Deshmukh

[7] International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-8 June, 2019, Sentiment Analysis Using Naïve Bayes Classifier Sentiment Analysis Using Naïve Bayes Classifier