

Twitter sentimental analysis through machine learning And Comparative Analysis of Different Models

Tijil Mittal, Abhay Agarwal, Ujjwal Kansal

Mentor: dr. Mukesh Rawat Supervisor: Anjali Sharma

Department of Computer Science and Engineering,
Meerut Institute of Engineering and Technology

Abstract- In this paper, we are willing to adopt an approach that uses a Sentiment analyzer that incorporates Machine Learning. This paper also gives an examination of strategies of sentiment analysis in political sentiments by applying directed ML calculations. Hence we built up a program model for feeling and sentimental examination to deal with users and people sentiments. This paper provides details regarding the plan to analyze sentiments by extracting a tremendous number of tweets. Results classify clients' viewpoint through tweets into positive and negative, which is spoken to in a pie outline and html page

Index Terms-computationally, ML, positive, Sentiment analyzer.

I. INTRODUCTION

Online social media have gotten more consideration nowadays. Public opinions and feedback on a variety of subjects are given and spread effectively through various social media.[2]An online social platform, Twitter is an American microblogging and interpersonal interaction administration in which clients can post refreshes (tweets) to companions (adherents). It has become a gigantic dataset of the purported notions. on which clients post and connect with messages known as "tweets" without any than 140 characters. Tweets can communicate sentiments on various subjects, which can assist with coordinating promoting efforts in order to impart consumers' insights concerning brands and items, episodes of tormenting, occasions that produce instability, extremity forecast in political and sports discussions.

III. PROPOSED SYSTEM

The sentiments can be classified into three classes which are negative, positive and unbiased or neutral sentiments.

- I. Positive Sentiments: This highlights the inspirational disposition of the speaker about the content. Feelings with good assessments reflect joy, happiness, delight, grin and so forth. In the event of political audits, if the positive surveys/feelings about the legislators are more, it implies individuals are content with his work.
- II. Negative Sentiments: This alludes to the negative demeanor of the speaker about the content. Tweets with negative conclusions reflect bitterness, desire, scorn and so on. In the event of political audits, if the adverse surveys/assessments about the legislators are more, it implies individuals are not content with his work.

acknowledgment or dismissal of lawmakers, all in an electronic informal way. Sentiment Analysis can be characterized as a cycle that computerized mining of perspectives, sentiments, perspectives and feelings from text, speech, tweets and information base sources through Natural Language Processing (NLP).[6] Supposition investigation includes characterizing sentiments in text into classifications like "positive" or "negative" or "impartial".

II. LITERATURE SURVEY

Twitter, with an immense number of customers and a large number directives for consistently, has promptly transformed into a huge asset for relationship to invigilate their reputation and brands by removing and researching the inclination of the tweets by everybody about their things, organizations market and even about competitors included that, from the electronic media made ends with the mammoth advancement of the web, super volumes of speculation messages as tweets, overviews, web diaries or any discussion get-togethers and conversations are available for assessment, thus making the web the speediest, most including and viably open vehicle for supposition assessment. While there has been a decent measure of examination on how assessments are communicated in classes, for example, online audits and news stories, how slants are communicated given the casual language and message-length limitations of microblogging has been considerably less contemplated.[5] There were numerous examinations in conclusion, however nearly those zeroed in on a piece of writing or evaluates. A tweet is just restricted to 140 characters, so it is as various as a scrutinize.

- III. Nonpartisan/Neutral Sentiments: In this case no particular feeling is reflected about the content. It is neither shows positive nor negative attitude. Thus this class doesn't suggest anything in general, however it is significant to enhance the classification of positive and negative sentiments.

Advantages:

- 1 Irrelevant highlights are dispensed appropriately from the element vector for effective working of Machine Learning calculations and algorithms.[4]
2. ANN beats in supposition order than Naive Bayes classifiers

Sentiment Analysis of tweets using ML(Python)

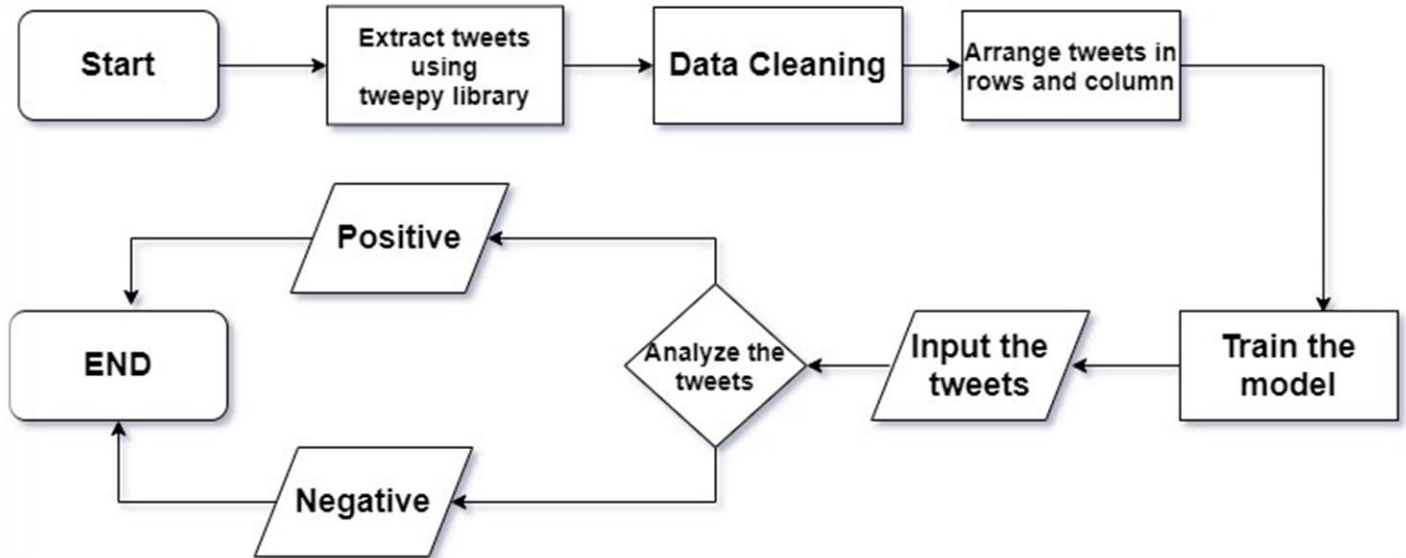


Fig 1: Workflow for the proposed system of sentimental analysis by machine learning

IV. Machine Learning Methods:

It joins AI calculations to find the notion via preparing on a known dataset. This way to deal with notion arrangement is administered and permits powerful content order. AI order requires two distinct arrangements of archives, specifically for preparing and testing. A preparation set is utilized A way to deal with estimation investigation utilizing Artificial Neural Network with near examination by a programmed classifier to learn and separate credits of records, and a test set is utilized to check the exhibition of the programmed classifier. There are many AI procedures embraced to arrange the surveys. AI strategies like NB, ANN ME, and SVM have accomplished better exhibitions in text arrangement.

A. Naive Bayes

Naive Bayesian Classifier is basically a probabilistic classifier which is used in Classification. It is based on principle i.e. every pair of features being classified is independent of each other or we can say it predicts on the basis of likelihood of an article.

B. Support Vector Machines (SVM)

Support Vector Machine is a powerful classifier that works both on linearly and non-linearly separable data. It basically finds an optimal hyperplane that best separates our data so that distance from nearest points in space to itself (also

called margin) is maximized. While these nearest points are called Support Vectors.

For non-linearly separable data, it uses something called 'Kernel Trick'.

C. XGBoost

XGBoost is basically an implementation of Gradient Boosted decision trees where X stands for 'eXtreme'. In this algo, decision trees are created and arranged in a sequential way. Weights play an important role in XGBoost. Weights are assigned to all the independent variables which are then fed into the decision tree which finally predicts results. These individual classifiers then ensembled together to give a strong and more precise model which has more accuracy than the last ones. It can work on all the applications like regression, classification, ranking, and user-defined prediction problems.

D. Logistic regression

Logistic regression is actually a classification algorithm not a regression one. It is a supervised learning algorithm. It can be used for both binary (0-1) as well as multi-class classification problems. This is an extension of linear regression as it models the data using Sigmoid Function.

$$g(z) = \frac{1}{1+e^{-z}}$$

Here, the output from the linear regression is put on as input to the sigmoid function so that we can have a number in the range of [0,1] and that number can be compared with the

threshold value which can help us in the classification purpose.

E. Artificial Neural Networks

An Artificial Neural Network (ANN) learning is robust to errors in the training data and has been successfully applied for learning real-valued, discrete-valued, and vector-valued functions containing problems such as text classification, information extraction, semantic parsing, question answering, language generation, machine translation, and speech and character recognition. ANNs are built out of a densely interconnected set of simple units, where each unit i.e. a neuron takes a number of real-valued inputs and produces a single real-valued output.

```

Import tweepy library
Enter the consumer key and access token
Run loop to extract the data set
for tweet in tweepy till 4000 tweets
csvWriter = csv.writer(csvFile) //store it in a csv file
  
```

V. Statistical analysis:

A. Precision:

Precision is fact of how accurate our model is out of the positive predicted ones. It also tells how many of them are actual positive.

$$\begin{aligned}
 \text{Precision} &= \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \\
 &= \frac{\text{True Positive}}{\text{Total Predicted Positive}}
 \end{aligned}$$

B. Recall:

Recall calculates how many of the Actual Positives our model captures through labeling it as Positive (True Positive).

$$\begin{aligned}
 \text{Recall} &= \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \\
 &= \frac{\text{True Positive}}{\text{Total Actual Positive}}
 \end{aligned}$$

C. F1 Score:

F1 is a function of Precision and Recall

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = \frac{2 \cdot \text{TP}}{\text{TP} + \text{FP} + \text{FN}}$$

TP = number of true positives

FP = number of false positives

FN = number of false negatives

VI. SAMPLE DATA SET

```

[b'Fair, election, !, \xf0\x9f\x97\xb3, find, ...,
https://t.co/rEwHUwnwrc, ']
[@vincecable, :, On, @BBCNewsnight, dismiss, claim,
post, No, Confidence, motion, PM, free, defer, election,
Oct, 31, ., Cross, par\xe2\x80\xa6, ']
['', b'@KarenLWms, It, appears, work, contract, w, /, feds,
,, case, ,, fired,\xe2\x80\xa6, https://t.co/7PnsLbFGnJ, ', '']
[@AltRightCanada, :, Google, Manipulating, Search,
Engine, Results, To, Get, Justin, Trudeau, Re, -, Elected,
https://t.co/ecnvxXTnGR, cdnpoli, cbcnl, qan\xe2\x80\xa6,
']
b'Google, Manipulating, Search, Engine, Results, To, Get,
Justin, Trudeau, Re, -, Elected, https://t.co/ecnvxXTnGR,
cdnpoli, cbcnl\xe2\x80\xa6, https://t.co/RXXD5zydKb, ']
['', b''RT, @XposeBlackCrime, :, JudicialWatch, -, Why,
Left, ACTIVELY, Fighting, Election, Integrity, ],
TomFitton, 's, Issue, Update, -, https://t.co/\xe2\x80\xa6, ', '']
[@Silvsports, :, @AOC, Talk, straw, man, argument, .,
Boys, boys, saying, acts, natural, happen, ., , The,
above\xe2\x80\xa6, ']
  
```

TABLE I. DATA SET USED IN TESTING ALGORITHMS

VII. STATISTICAL DATA

Table II. TABLE SHOWING THE STATISTICAL DATA IN TERMS OF NUMERICAL VALUES OF VARIOUS MACHINE LEARNING ALGORITHMS

| MODELS | DATA SET | APPROACH | ACCURACY (Approx) |
|---------------------|-----------------|------------|-------------------|
| XGboost | Election tweets | Supervised | 82 |
| Naive Bayes | Election tweets | Supervised | 80 |
| Logistic Regression | Election tweets | Supervised | 85 |
| SVM | Election tweets | Supervised | 90 |
| ANN | Election tweets | Supervised | 96.27 |

VIII. Conclusion:

Through the investigation of results, the point of this exploration is accomplished in distinguishing the most appropriate calculation for opinion examination on twitter information regarding the chose dataset. ANN model outcomes in having the most elevated exactness of 96.2% among the chose calculations for the conclusion investigation of Twitter information regarding the chose dataset.

Catchphrases: Machine Learning, Sentiment Analysis, Twitter information, Deep Learning, Naïve Bayes, Twitter Sentiment Analysis

IX. REFERENCES

- [1] IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 18, Issue 2, Ver. V (Mar-Apr. 2016), PP 64-69
- [2] Artem Oppermann Feb 25,2019. Sentiment Analysis with Deep Learning of Netflix Reviews.[<https://towardsdatascience.com/sentiment-analysis-with-deep-learning-62d4d0166ef6>] Accessed November 23,2020.
- [3] Sergio Verahonda Oct 9 .An easy tutorial about Sentiment Analysis with Deep Learning and Keras.[<https://towardsdatascience.com/an-easy-tutorial-about-sentiment-analysis-with-deep-learning-and-keras-2bf52b9cba91>] Accessed November 23,2020.
- [4] International Journal of Computer Science Trends and Technology(IJCST)-Volume 4 Issue 3 ,May-June 2016.
- [5] International Journal of Electrical and Computer Engineering (IJECE) Vol. 6, No. 1, February 2016, pp. 357~366 ISSN: 2088-8708, DOI: 10.11591/ijece.v6i1.8982 .
- [6] September 29, 2020. Machine Learning (ML) for Natural Language Processing (NLP) [<https://www.lexalytics.com/lexablog/machine-learning-natural-language-processing>] Accessed November 23,2020.
- [7] G. Vinodhini and RM. Chandrasekaran, "Sentiment Analysis and Opinion Mining: A survey", International Journal of Advanced Research in Computer Science and Software Engineering, (pp.281 – 291) -2012.
- [8] Ankush Chavan Apr 24. Automatic Tweet Liker using Tweepy and Django — Part 1 [<https://ankush-chavan.medium.com/automatic-tweet-liker-using-tweepy-and-django-215a89838e23>] Accessed November 23,2020.