IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 4, April 2021

DOI 10.17148/IJARCCE.2021.10459

Sentiment Analysis of Twitter and Facebook Posts in NLP using Naïve Bayes Classifier

Prof. Farida Attar[1], Aisha Shaikh[2], Aiman Shaikh[3], Aquib Sayed[4]

Assistant Professor, M. H. Saboo Siddik College Of Engineering, Mumbai 8, India[1]

Student, M. H. Saboo Siddik College Of Engineering, Mumbai 8, India[2]

Student, M. H. Saboo Siddik College Of Engineering, Mumbai 8, India[3]

Student, M. H. Saboo Siddik College Of Engineering, Mumbai 8, India[4]

Abstract: The way in which people emote, express, share, opine thoughts have drastically changed in the past few decades, thanks to the Internet. Nowadays people can relentlessly express their thoughts, opinions on a social media platform, forums, product review sections, comment sections, online survey portals, etc., and the list is endless in the world of Internet. More importantly the major chunk of sentiments comes from social media platforms like Twitter and Facebook due to having a large number of daily active users under their belt.

Keywords: Sentiment analysis, Twitter, Facebook, Naïve Bayes Classifier, Social networks, classification and opinion.

I. INTRODUCTION

Nowadays, social networking services(SNS) are popularly used all around the globe, and the statistics for SNS users has been growing exponentially. People exhibit their happiness, frustration, inconveniences, amusement, excitement, thrill, and other multiple kind of emotions on a social platform like Twitter and Facebook and many other. But we our main focus is limited to these two platforms. It requires a deal of effort to summarize the actual emotion behind the textual information (unstructured) tweeted or posted by the users. There are numerous convoluted tendencies of the human mind which further complicates the process of sentiment analysis and poses a big challenge for natural language processing in general.

With smartphone tech made widely and easily accessible to masses and social media being the major contributor of the all the Internet consumption. Social media firms are capitalizing on the human tendencies by making the apps more addictive to increase their session times and ad revenues. Thus people are spending more time on their smartphones and social media than ever before. This generation which is witnessing an extremely interconnected virtual network where most of the interaction is virtual which includes exchanging their thoughts, ideas, emotions, sentiments, contemplations, beliefs, opinions, etc. hence the term “netizen” came into existence where people are heavily invested in the Internet culture and the online communities it offers.

This type of behavior shift provides the commercial and business sectors with a lot of free data to work with, which can be further used to improve their products and services and thus maintain and increase their overall demand and desire to use that product or service.

II. LITERATURE SURVEY

A. Sentiment analysis of twitter data

Social networks are the focal means to collect information about people’s views and thoughts concerning various topics as they devote hours on social media regularly and opine their thoughts and emotions. Sentiment analysis has numerous applications for diverse domains for instance in businesses to get responses, comments, criticisms for products through which businesses can track, learn and predict customers’ sentiment on social medias. Opinion and sentiment extraction have well explored with respect to this approach and various domains have been well researched. Rudimentary methods to mine text summarization are found to be extractive and abstractive in nature. In the extractive technique, words and word phrases are mined from the native text to effectuate a summary. In the abstractive method, the system attempts to assimilate an internal language representation and creates summary that is more alike to the summary done by human. The sentiment-aware systems nowadays have multiple applications from businesses to social sciences. Since Twitter is heavily relied on short tweets also known as micro-blogging with prominent emphasis on acronyms and abbreviations, traditional Natural Language Processing systems fail to extract the exact emotion behind the tweet,

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 4, April 2021

DOI 10.17148/IJARCCE.2021.10459

therefore many researchers advise using deep learning and machine learning techniques to determine the polarity of the text.

B. Machine Learning-Based Sentiment Analysis for Twitter Accounts

Accomplished Sentiment Analysis with the objective of arbitrating the accuracy of these analyzers, tweets were analyzed from various sentiments analyzers. The solitary determination of using variety of sentiment analyzers such as TextBlob, SentiWordNet, and WSD was to provide a brief comparison of their accuracies. By allocating a score from -1 to 1 based on the words used, where a negative score means a negative emotion and a positive score means a positive emotion and zero value is regarded as a neutral sentiment. It delivers a brief comparison of various approaches of sentiment analysis in the analysis of political views by implementing supervised machine-learning algorithms such as Naïve Bayes and Support Vector Machines(SVM).

C. Sentiment Analysis of Facebook Statuses using Naïve Bayes classifier for language learning

Facebook enables individuals Sentiment Analysis of Facebook have their own records to express sentiments and pass on feelings by means of writings and in addition emoji. We give enhanced characterization methods for slant investigation of Facebook statuses and characterizing the content as either positive or negative or impartial. The creators gave a decent overview of different methods created on the web sentiment analysis. Among the machine learning strategies, they thought about three directed methodologies. The first approach is using the “Naive Bayes” method and the other approach using the “SVM”. The information utilized is irreproducible Further our work can be improved by following changes of individuals' estimation on explicit subject and the time reliance of our information can be investigated to examine their patterns it might additionally give intriguing results, if we think about the transient highlights on this examination and not to concentrate exclusively on past posts or exchanges.

III.EXISTING SYSTEM

A. Sentiment analysis of twitter data

There are primarily three distinct classifications levels i.e. Document-Level, Sentence-Level, Aspect-Level. Apart from being challenging, Twitter also functions on the principle of aspect level classification which directly aims for the opinion or the entity itself. The proposed system implements aspect level sentiment analysis on Twitter Data or based on tweets into three main categories [1]:

- Positive
- Negative
- Neutral

B. A novel classification approach based on Naive Bayes for Twitter sentiment analysis

This system Illustrates the operation flow of the proposed approach which adopts two methods based on MNB. The first one targets to calculate the weights more minute and precisely based on the training set divided into positive and negative, while the seconds one is built such a way that it seeks to modify the weights using the mean of weight differences for automatic feature selection. With the existing MNB scheme, the occurrence of each word in the training document set, the value of D is computed. [4].

C. Machine Learning-Based Sentiment Analysis for Twitter Accounts

In this system, we are not aiming to propose any sentiment-prediction technique, but analyzing the notable sentiment techniques in election domains. Due to this, the dataset is primarily based off of users’ tweets about political bias and opinions across Pakistan. These type of analyzers are also being utilized in domains such as health, disease and various other branches of social sciences. This paper aims to solidify the sentiment analysis of the Urdu tweets by first translating the Urdu tweets into English. To further add on to that, our outcomes from analyzers are thus affirmed from statistical machine learning classifiers like Naïve Bayes and SVM in Weka. To put that into perspective, this system also has demonstrated the results of both: sentiment analyzers and their machine-learning capabilities using a tabular format to offer a crystal clear estimate about the accuracy percentages obtained [2].

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 4, April 2021

DOI 10.17148/IJARCCE.2021.10459

IV. PROPOSED SYSTEM

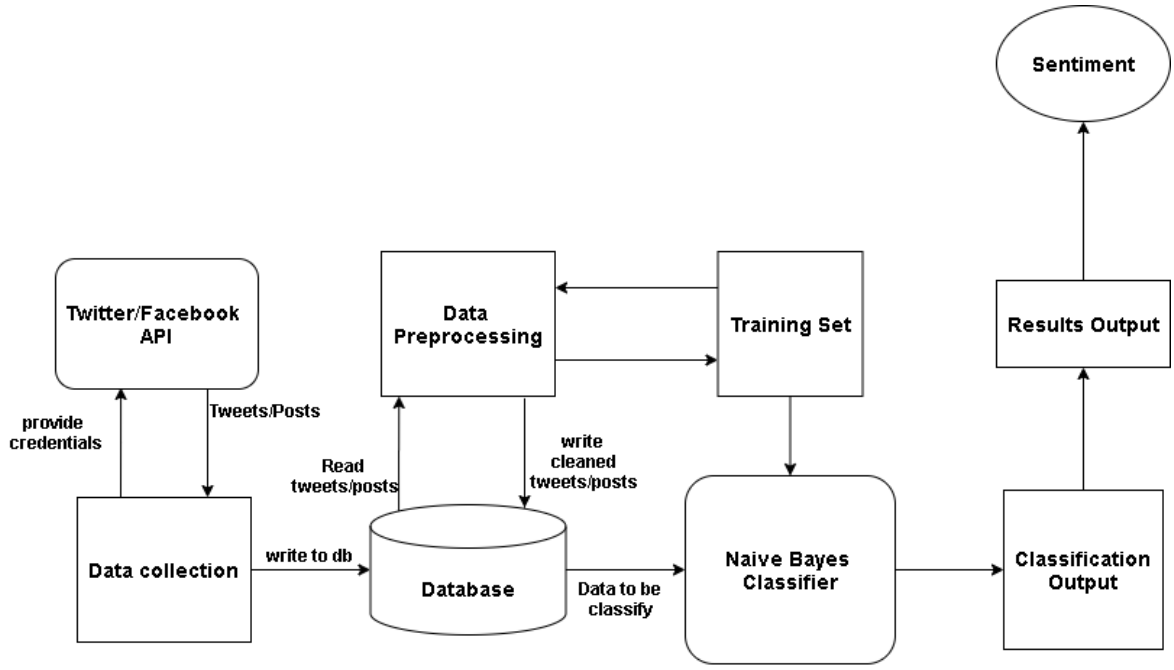

Fig. 1 System Block Diagram

Our System has the following steps:

1. Data collection using Twitter/Facebook API using developer accounts because publicly large datasets of the social media are not available.
2. Next up is Data pre-processing which involves simplifying and cleaning the data by performing spell correction, removing stop words, Unicode, URLs, punctuation handling, stemming, etc. to remove noise from the data.
3. Applying Naive Bayes Classifier on these tweets/posts in order to categorize them. Prior to this, the classification model is thoroughly trained using the training data sets.
4. Classified tweets/posts may belong to any of these three categories (positive, negative & neutral)
5. The results of the sentiments are demonstrated using data visualization.

V. ALGORITHM

Naïve Bayes classifier is a probabilistic, supervised machine learning algorithm which is applied to classify the tweets/posts based on the extracted feature set from the words. The algorithm works on the following theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Using the Bayesian terminologies, we can rewrite the equation as:

$$Posterior = \frac{prior\ x\ likelihood}{evidence}$$

Implementation code for training the classifier algorithm:

```
# Training the Naive Bayes classifier
ts = nltk.classify.apply_features(extract_features,tweets)
classifir = nltk.NaiveBayesClassifier.train(ts)
# Training the Naive Bayes classifier
ts = nltk.classify.apply_features(extract_features,tweets)
classifir = nltk.NaiveBayesClassifier.train(ts)
neg_cnt = 0
```

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 4, April 2021

DOI 10.17148/IJARCCE.2021.10459

```
pos_cnt = 0
for obj in test_neg:
    result = classifier.classify(extract_features(obj.split()))
    if(result == 'Negative'):
        neg_cnt = neg_cnt + 1
for obj in test_pos:
    result =  classifier.classify(extract_features(obj.split()))
    if(result == 'Positive'):
        pos_cnt = pos_cnt + 1
```

VI. FUTURE SCOPE

Currently our system only deals with classifying tweets into positive, negative and neutral based on the entered user handle or the hashtag.

i. Furthermore, the system can incorporate features such as actual emotion detection like joy, anger, disgust, sorrow, happiness, frustration and many more as the system learns and improves.
ii. This system can also be made capable of predicting things like sarcasm, irony, pun, exaggeration, modern and trending acronyms
iii. Negative words like "not" have a prominent effect of inverting the polarity. Our current classifier is not capable of handling this as expected.

VII. CONCLUSION

In this paper, the approach mainly focuses on to derive the superficial sentiment behind the tweet/post in consideration the amount of negative and positive words and thus classify into positive, neutral and negative.

The workflow of the proposed system also emphasizes on the subjectivity and polarity of the tweet which in turn results into better accuracy of the sentiment prediction.

ACKNOWLEDGMENT

We express our sincere gratitude to our Principal **Dr. Ganesh Kame**, M. H. Saboo Siddik College of Engineering, Mumbai for providing the facilities to carry out the project work. We thank our beloved Head of the I.T. Department, **Er. Sanam Kazi** who gave us full support and valuable suggestions. We are greatly thankful to our guide **Er. Farida Attar** for her kind support and guidance involved in successful completion of the implementation and research study.

REFERENCES

[1]. Kiruthika M., Sanjana Woonna, Priyanka Giri, “Sentiment Analysis of Twitter Data”, International Journal of Innovations in Engineering and Technology(IJIET), 2016.

[2]. Ali Hasan, Sana Moin, Ahmad Karim, Shahaboddin Shamshirband, “Machine Learning-Based Sentiment Analysis for Twitter Accounts”, Multidisciplinary Digital Publishing Institute, 2018.

[3]. Christos Troussas, Maria Virvou, Kurt Junshean Espinosa, Kevin Llaguno, Jaime Caro, “Sentiment Analysis of Facebook Statuses using Naïve Bayes classifier for language learning”, International Conference on Information, Intelligence, Systems and Applications (IISA), 2013.

[4]. Junseok Song, Kyung Tae Kim, Byungjun Lee, Sangyoung Kim, Hee Yong Youn, “A novel classification approach based on Naïve Bayes for Twitter sentiment analysis”, KSII Transactions on Internet and Information Systems, 2017.