IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2320-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 5, May 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.512

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 10, Issue 5, May 2021||

DOI:10.15680/IJIRSET.2021.1005181

Offensive Text Detection Using Machine Learning Techniques

Mohit Pawar[1], Prathamesh Pansare[2], Ashitosh Gaikwad[3], Sushmita Ray[4]

Department of Computer Engineering, Bachelors of Engineering, All India Shri Shivaji Memorial Society's Institute of Information Technology, Pune University, Pune, Maharashtra, India [1,2,3,4]

ABSTRACT: Social media is a substantial information source. The examination of a large amount of user-generated content on social media can provide useful information for eliciting people's emotions and gaining a better knowledge of public mood and attitude. We provide a generic approach for analysing massive social data and predicting public emotions and mood in this paper.The framework is divided into two sections. To classify emotional content in user-generated social data, an ensemble classifier schema is first utilised, which blends a general domain-independent, knowledge-based tool with machine learning approaches.Then, based on the emotions identified, a graph-based approach is used to model the emotional level of a topic, and the topic's emotional graph is created, showing public feelings and mood on the issue. The case study's findings are quite positive. We presented a sentiment classification strategy employing the Support Vector Machine (SVM),Naive Bayes (NB) and Decision Tree(DT) algorithms in the suggested research. For categorization, the Twitter base streaming dataset was utilised. For testing and training, Machine Learning algorithms and Natural Launguage processing were employed for classification. The performance study demonstrates that the suggested system outperforms traditional classification techniques.

KEYWORDS: Machine Learning, NLP, Twitter data, SVM, NB, Decision Tree

I. INTRODUCTION

New means of communication, such as text messaging and microblogging, have evolved in the last decade and have grown widespread. While the breadth of information transmitted by tweets and texts is limitless, these brief communications are frequently used to communicate people's thoughts and feelings about what's going on in the globe. Tweets and messages are far shorter than documents, consisting of only a phrase or a headline. Additional feature of social media data, like Twitter messages, is that it provides rich structured information on the persons who took part in the conversation. People who have profiles on social media platforms (SMPs) but mask their identities for malevolent motives are on the rise. Sentiment Analysis is the computer process of finding and classifying opinions conveyed in a piece of text to determine if the writer has a favourable or negative attitude toward a given topic, product, etc. Sentiment analysis is the process of extracting a user's opinion from a written content. We uncover and evaluate the issue of the Bag-of-Words model, which disturbs word order and discards some semantic structure, as well as a familiar difficulty in the Polarity Shift Problem during Negation of Sentiment in the suggested study. The approaches used in traditional topic-based text classification are used in sentiment classification. However, there are a few drawbacks to this technique: 1) It is inaccurate to not handle a word or phrase. 2) Ignoring the polarity change issue. Natural language processing, machine learning, and other technologies will be used in the suggested system to handle this challenge. We suggest a basic yet effective model known as Naive Bayes, SVM, and DT.

II. LITERATURE SURVEY

Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection [2]. A method for detecting hate speech on Twitter. Our method is based on the automated collection of unigrams and patterns from the training set. These patterns and unigrams are then employed as features in a machine learning method, among other things. Our studies on a test set of 2010 tweets reveal that our method can determine if a tweet is offensive or not (binary classification) with an accuracy of 87.4 percent, and can determine whether a tweet is hostile, offensive, or clean with an accuracy of 78.4 percent (ternary classification).

|e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 10, Issue 5, May 2021||

DOI:10.15680/IJIRSET.2021.1005181

Research on text sentiment analysis based on CNNs and SVM[3]. A model combining Convolutional Neural Networks (CNNs) with SVM text sentiment analysis is suggested. The experimental findings suggest that the suggested technique successfully enhances text sentiment classification accuracy when compared to classic CNN, and that sentiment analysis based on CNNs and SVM is useful.

An approach AHTDT- Automatic Hate Text Detection Techniques in Social Media[4]. The automated hate text recognition approaches are highlighted, as well as their parametric comparison. Different hate text detection algorithms are used on social networking sites such as YouTube, Facebook, Whisper, Blogger, and others, and they rely on natural language processing, data mining, and machine learning domains. Detecting hate writing on social media ensures that children are safe online. Bag of Word (BOW) techniques have a few limitations, for example, a BOW model ignores the meaning of the word.

A framework for sentiment analysis with opinion mining of hotel reviews[5]. Sentiment polarity is a system for extracting impartial assessments of hotel services from reviews that automatically builds a sentiment dataset for training and testing. To choose an appropriate machine learning method for the framework's classification component, a comparison study was developed using Naive Bayes multinomial, sequential minimum optimization, complement Naive Bayes and Composite hypercube on iterated random projections. In recent years, Twitter has grown in popularity as a microblogging social media platform, allowing millions of users to communicate their everyday thoughts and ideas via real-time status updates.

III. METHODOLOGY

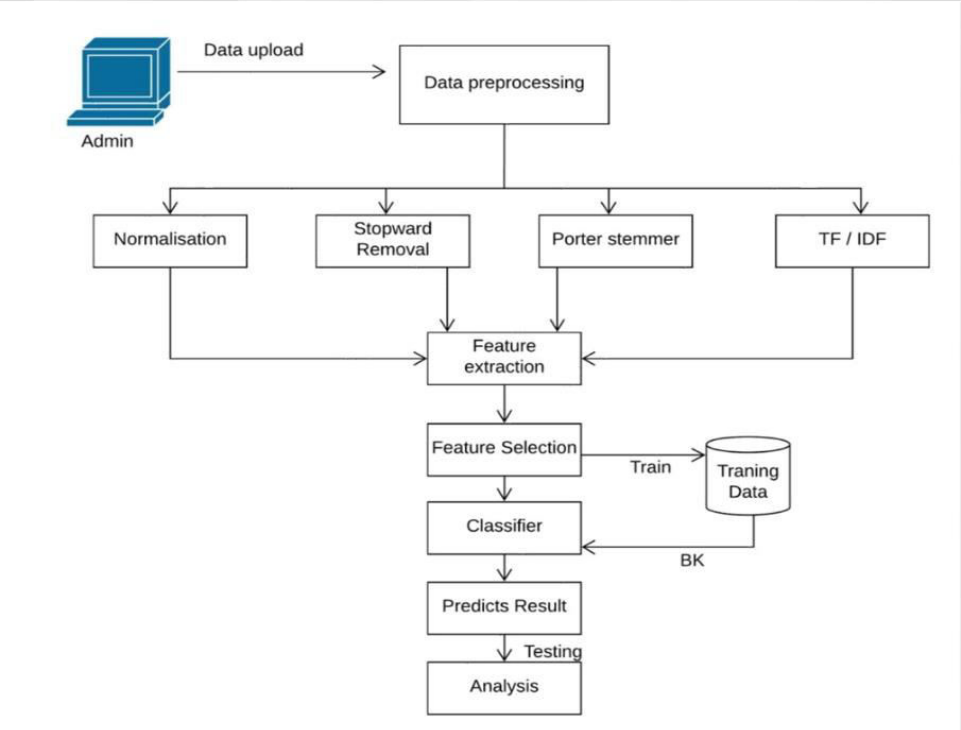

Figure 1: Proposed System Architecture

For this study project, the system gathered data from a variety of online sources, including Twitter API and some synthetic sources. The aforesaid architecture is depicted in Figure 1, which shows how the first system uses Twitter API to retrieve the data set based on input queries. Many times, internet data contains some incorrectly categorised instances, necessitating data pre-processing and normalisation. Tokenization, stop word removal, porter stemmer, and TF-IDF are all part of the natural language processing (NLP) stage of preprocessing. When all of the steps are

|e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 10, Issue 5, May 2021||

DOI:10.15680/IJIRSET.2021.1005181

complete, the system extracts the relevant characteristics from the whole data set. The extract feature was employed in the unigram approach. Once the features have been extracted, choose the best feature from the data. To choose the optimal feature, cosine similarity with TF-IDF and correlation with TF-IDF approaches were utilised. This is a hybrid technique for feature selection that may be used for both training and testing. Finally, the system uses a classification method to construct background knowledge in the train model, which includes both positive and negative levels. During the testing phase, the system runs a natural language processing (NLP) method on the full data set to extract and select the feature set. The accuracy was predicted using Artificial Neural Network (ANN) and Naive Bayes (NB) classifiers. Finally, the system's performance is evaluated using current classification techniques in the results section.

Algorithms Used

1. : Stop word Removal Approach
Input: Stop words list L[], String Data D for remove the stop words. Output: Verified data D with removal all stop words.
Step 1: Initialize the data string S[].
Step 2: initialize
a=0,k=0 **Step 3:** for
each(read a to L)
If(a.equals(L[i]))
Then Remove
S[k] End for
Step 4: add S to D.
Step 5: End Procedure

Stemming
Algorithm
Input : Word w
Output : w with removing past participles as well.
Step 1: Initialize w
Step 2: Intialize all steps of Porter stemmer
Step 3: for each (Char ch from w)
If(ch.count==w.length()) &&
(ch.equals(e))
Remove chfrom(w)
Step
4:if(ch.endswith(ed
)) Remove 'ed'
from(w)
Step 5: k=w.length()
If(k (char) to k-3
.equals(tion)) Replace
w with te.
Step 6: end procedure

TF-IDF
Input : Each word from vector as Term T, All vectors V[i…n] Output : TF-IDF weight for each T
Step 1 : Vector = {c1, c2, c3….cn}
Step 2 : Aspects available in each comment
Step 3 : D = {cmt1, cmt2, cmt3,
cmtn} and comments available in
each document
Calculate the Tf score as

|e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 10, Issue 5, May 2021||

DOI:10.15680/IJIRSET.2021.1005181

Step 4 :tf (t,d) = (t,d)
t=specific term
d= specific document in a term is to be found.

Step 5 :idf = t □ sum(d)
Step 6: Return tf *idf

Classification algorithm (SVM)

Input : Training Rules Tr[], Test Instances Ts[],
Threshold T. Output : Weight w=0.0
Step 1 : Read each test instance from (TsInstnace from Ts)

Step 3 : Read each train instance from (TrInstnace from Tr)
Step 6 :if (w >= T)
Step 7 : Forward feed layer to input layer for feedback FeedLayer[] □ {Tsf,w}
Step 8 :optimized feed layer weight, Cweigt□FeedLayer[0]
Step 9 : Return Cweight

IV. RESULTS AND DISCUSSIONS

We generate matrices for accuracy in the suggested system performance evaluation. The system is built using the Java architectural framework. The data comprises roughly 70,000 user comments, some of which are good and some of which are critical. Finally, the algorithm categorizes all of the comments as favorable, negative, or neutral. When it comes to aspect categorization, negative handling is also useful.

The accuracy of the system with multiple machine learning algorithms assisting with various Natural language Processing (NLP) aspects is shown in Figure 2.

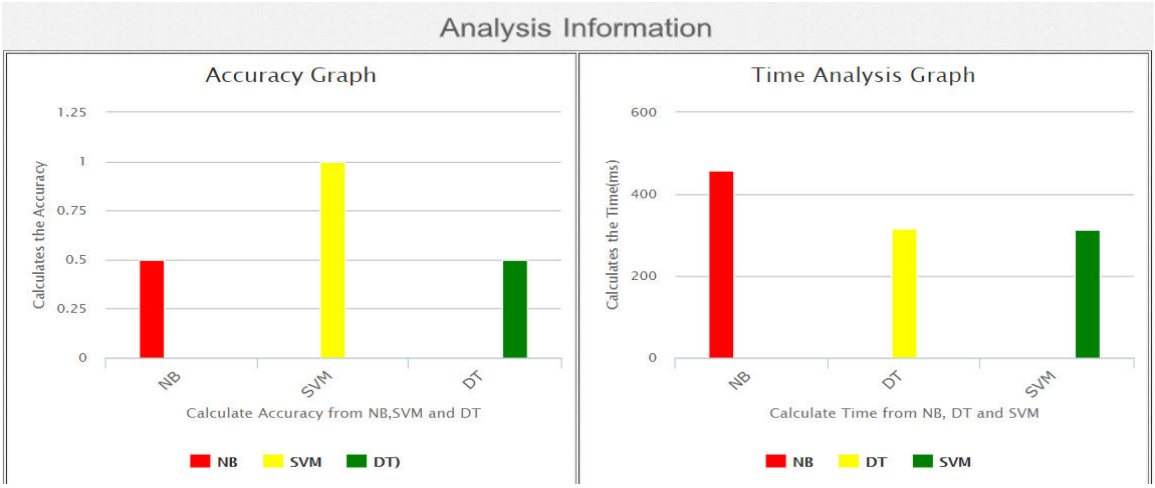

Figure 2 : System accuracy with various machine learning algorithms

Finally, based on the above graphs, we can conclude that the SVM algorithm delivers greater classification accuracy than other supervised machine learning algorithms.

|e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 10, Issue 5, May 2021||

DOI:10.15680/IJIRSET.2021.1005181

V. CONCLUSION

The preprocessing module in the proposed study includes sub modules for tokenization, stop word removal, and stemming. To avoid ambiguity, aspects with the same meaning but different names are handled using synonyms. The sheer magnitude of actual data has a number of implications. The sheer volume of data implies that the methods used to process it be sufficiently rapid, as slower models may not be able to handle it in real time. On the other hand, it enables us to filter the data and focus only on the portions that satisfy a certain standard of quality. It was e.g. observed that not all of the posts bearing negative sentiment contain a reasonable complaint, expressing the cause of the negative sentiment. In most circumstances, the sentiment class will not be the final outcome of a practical application. The product should contain useful information that can be acted upon. As a result, we should limit our social media posts to those that are directly tied to a product, service, or feature of the latter. The postings should then be combined in order to show the results in a clear and understandable manner. One of the goals of future research should be to focus on this.

REFERENCES

[1] Watanabe, Hajime, MondherBouazizi, and TomoakiOhtsuki. "Hate speech on twitter: A pragmatic approach to collect hateful and offensive expressions and perform hate speech detection." IEEE Access 6 (2018): 13825-13835.
[2] Gaydhani, Aditya, et al. "Detecting hate speech and offensive language on twitter using machine learning: An n-gram and tfidf based approach." arXiv preprint arXiv:1809.08651 (2018).
[3] Wiegand, Michael, et al. "A survey on the role of negation in sentiment analysis." Proceedings of the workshop on negation and speculation in natural language processing. 2010.
[4] Al-Hassan, Areej, and Hmood Al-Dossari. "Detection of hate speech in social networks: asurvey on multilingual corpus." Computer Science & Information Technology (CS & IT) 9.2 (2019): 83.
[5] Badjatiya, Pinkesh, et al. "Deep learning for hate speech detection in tweets." Proceedings of the 26th International Conference on World Wide Web Companion.International World Wide Web Conferences Steering Committee, 2017.
[6] Del Vigna12, Fabio, et al. "Hate me, hate me not: Hate speech detection on facebook." (2017).
[7] Pratiwi, Nur Indah, Indra Budi, and IkaAlfina. "Hate Speech Detection on Indonesian Instagram Comments using FastText Approach." 2018 International Conference on Advanced Computer Science and Information Systems (ICACSIS). IEEE, 2018.APA
[8] Şahi, Havvanur, YaseminKılıç, and Rahime Belen Sağlam. "Automated Detection of Hate Speech towards Woman on Twitter."2018 3rd International Conference on Computer Science and Engineering (UBMK).IEEE, 2018. [9] Saksesi, Arum Sucia, Muhammad Nasrun, and CasiSetianingsih."Analysis Text of Hate Speech Detection Using Recurrent Neural Network." 2018 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC).IEEE, 2018.
[10] Watanabe, Hajime, MondherBouazizi, and TomoakiOhtsuki. "Hate speech on twitter: A pragmatic approach to collect hateful and offensive expressions and perform hate speech detection." IEEE Access 6 (2018): 13825-13835.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.512

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details