IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, May 2022
Impact Factor: 6.252

Online Public Shamming using Machine Learning

Noopura Vaidya, Shivani Chandak, Shubham Dwivedy, Aatif Malik
Department of Computer Science Engineering
Sinhgad College of Engineering, Pune, Maharashtra, India

Abstract: *Social Media Platforms involve not millions but billions of users around the globe. Interactions on these easily available social media sites like Twitter have a huge impact on people. Nowadays, there is undesirable negative impact for daily life. These hugely used major platforms of communication have now become a great source of dispersing unwanted data and irrelevant information, Twitter being one of the most extravagant social media platforms in our times, the topmost popular microblogging services is now used as a weapon to share unethical, unreasonable amount of opinions, media. In this proposed work the dishonoring comments, tweets towards people are categorized into 9 types. The tweets are further classifying into one of these types or non- shaming tweets towards people. Observation says out of the multitude of taking an interested clients who posts remarks on a specific occasion, lions share is probably going to modify the person in question. Moreover, it is not the non-shaming devotee who checks the increment quicker but of shaming in twitter.*

Keywords: Negative, abusive, shamming, comments, tweets etc

I. INTRODUCTION

It is an Online Community characterized informally for the utilization of various sites of different genres that allows users to connect, discover Themselves their interests. These online made platform gives access to people across the globe to connect with people irrespective of their gender, age, religion. As everything comes with its own advantages and disadvantages, the children of this generation are introduced in a wrong way, before time to various levels of horrifying experiences here by losing their innocence meeting vulnerability. There are more issues which social network users are not aware of how they are attacked by hosted sites attackers. Today social media has become an integral part of our life, people utilize informal organizations, music, recordings, data, sharing pic- tures, etc. On a business level interpersonal organizations permits the user to communicate with different pages in the web. There is Online Web based shop- ping, promoting through advertisements for marketing. Social media platforms other than Twitter like Myspace, LinkedIn, and Facebook are also famous and connect various dots in the web world. The shaming which happens through this various social media platform have to be controlled as there is psychological dis- turbances, mental health problems happening because of these tweets. Here we have introduced offensive language detection, it is an activity of processing the natural languages and to figure out the shaming which is based on racism, related to religion, etc. The shaming detection of words are in the English Text Format for the comments, reviews on the movies, tweets, personal/political reviews, etc. The 9 types of classification are:

1. Abusive
2. Comparison
3. Passing Judgement
4. Religious
5. Sarcasm
6. Whataboutery
7. Vulgar
8. Spam
9. Non spam.

1.1 Literature Survey

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, May 2022
Impact Factor: 6.252

Dhamir Raniah Kiasat Desrul, AdeRo maDhony: In this paper, author presents an Indonesian abusive language detection system by accepting the problem using classifiers: Naives Bayes, SVM and KNN. They also perform feature process, similar information between words.

Guanjun Lin, Sun, Surya Nepal, JunZbang, Yamg Xiang, Senior Member, Houcinr Hassan: This paper explains how widely Cyberbullying happens and is granted a serious problem. Mostly its observed teenagers are victim of this type of crime like mail spam, facebook, twitter. Younger generation uses technology to learn but then they are harassed, threatened. They work on solving social and psychological problems of teenager's boys and girls by using innovative social network software. Reducing cyberbully involves two parts- First is robust technique for effective detection and other is reflective user interfaces.

Justin Cheng, Bernstien, Cristein Danescu, Niculesu, Mizil, Jure Leskove: Twitter trolling disturbs meaningful, motivational, emotional discussion in online communication by posting immature and provoking comments. A guessing model of trolling behaviour is designed which shows the mood of the user which will calculate and describe trolling behaviour and an individual history of trolling.

RajeshBasak, Sural, Senior Member, IEEE, Niloy Ganguly: As many of you know hate speech is a huge current problem. It is actually spreading, growing and particularly affects community such as a people of particular religion or people of particular colour or sudden race etc. This impacts our population highly. It is speech that threaten individuals base on natural language religion, ethnic origin, national origin, gender etc. This paper is also presenting the survey of hate speech. The online hate speech is also increasing our social media problems. The purpose is to implement a system that can detect and report hate to the constant authority using advance machine learning with natural language processing.

Mukul Anand, Dr. R. Eswari: In this paper the author uses Kaggle's toxic comment dataset for training the deep learning model and the data is categorized in harmful, deadly, gross, offensive, defame and abuse. On dataset various deep learning techniques get performed and that helps to analyse which deep learning techniques is better. In this paper the deep learning techniques like long short term memory cell and convolution neural network with or without the words GloVe, embeddings, GloVe. It is used for obtaining the vector representation for the words.

Alvaro Garcia-Recuero, Aneta Morawin and Gareth Tyson:In this research paper author uses the user's attributes and social graph metadata. The former includes the schema of account itself and latter includes the communicated data between sender and receiver. It uses the voting scheme for categorization of data. The sum of the vote decide that the message is acceptable or not. Attributes helps to identify the user account on OSN and graph-based schema used, the dymanics of scattered information across the network. The attributs uses the Jaccard index as a key feature for classifying the nature of twitter messages.

Justin Cheng, Michael Bernstein, Cristian Danescu Niculescu-Mizil Jure Leskovec: This study uses two primary trigger mechanism: the individual's mood and the surrounding context of discussion. This study shows that both negative mood and seeeing troll posts by others notably increases the chances of a user trolling and together doubles the chances. A sinister model of trolling behaviour shows that mood and discussion context together can explain trolling behaviour better than individuals' history of trolling. The result shows that ordinary people under right circumstances behave like this.

1.2 Functional Requirements

Registration and Authentication

The user should be able to create an account and login on the system using it. While logging in, the system should authenticate that it's the right user details.

Taking user's inputs

After logging in, users will have two options to choose one will be online part and other will be offline part. After that user have to upload the face image for mood detection. The user also plays the music file based on mood detection result.

Performance Requirements

The performance of the system lies in the way it is handled. Every user must be given proper guidance regarding how to use the system. The other factor which affects the performance is the absence of any of the suggested requirements.

Safety Requirements

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, May 2022
Impact Factor: 6.252

To ensure the safety of the system, perform regular monitoring of the system so as to trace the proper working of the system. An internal staff has to be trained to ensure the safety of the system. He has to be trained to handle extreme error cases.

Security Requirements

Secure Functional Requirements; this is a security related description that is integrated into each functional requirement. Typically, this also says what shall not happen. This requirement artifact can for example be derived from misuse cases. Only authorized doctor will be accessing this system.

II. METHODOLOGY

1. Lexicon-based Approach Lexicon-based methods make use of predefined list of words where each word is associated with a specific sentiment. The lexicon methods vary according to the context in which they were created and involve calculating orientation for a document from the semantic orientation of texts or phrases in the documents. Besides, also states that a lexicon sentiment is to detect word-carrying opinion in the corpus and then to predict opinion expressed in the text. has shown the lexicon methods which have a basic paradigm which are: i. Preprocess each tweet, post by remove punctuation ii. Initialize a total polarity score (s) equal 0 -> s=0 iii. Check if token is present in a dictionary, then If token is positive, s will be positive (+) If token is negative, s will be negative (-) iv. Look at the total polarity score of tweet post If s > threshold, tweet post as positive If s < threshold, tweet post as negative However, highlighted one advantage of leaning-based method, is that it has the ability to adapt and create trained models for specific purposes and contexts. In contrast, an availability of labeled data and hence the low applicability of the method of new data which is cause labeling data might be costly or even prohibitive for some tasks.

2. Machine-learning-based Approach Machine learning methods often rely on supervised classification approaches where sentiment detection is framed as a binary which are positive and negative. This approach requires labeled data to train classifiers. This approach, it becomes apparent that aspects of the local context of a word need to be taken into account such as negative (e.g., Not beautiful) and intensification (e.g., Very beautiful). However, showed a basic paradigm for create a feature vector is:

1. Apply a part of speech tagger to each tweet post
2. Collect all the adjective for entire tweet posts
3. Make a popular word set composed of the top N adjectives
4. Navigate all of the tweets in the experimental set to create the following:

- Number of positive words
- Number of negative words
- Presence, absence or frequency of each word showed some examples of switch negation, negation simply to reverse the polarity of the lexicon: changing beautiful (+3) into not beautiful (-3). More examples: She is not terrific (6-5=1) but not terrible (-6+5=-1) either. In this case, the negation of a strongly negative or positive value reflects a mixed perspective which is correctly captured in the shifted value. However, has mentioned the limitation of machine-learning-based approach to be more suitable for Twitter than the lexical based method. Furthermore, stated that machine learning methods can generate a fixed number of the most regularly happening popular words which assigned an integer value on behalf of the frequency of the word in the Twitter.

F. Techniques of Sentiment Analysis The semantic concepts of entities extracted from tweets can be used to measure the overall correlation of a group of entities with a given sentiment polarity. Polarity refers to the most basic form, which is if a text or sentence is positive or negative. However, sentiment analysis has techniques in assigning polarity such as:

1. Natural Language Processing (NLP) NLP techniques are based on machine learning and especially statistical learning which uses a general learning algorithm combined with a large sample, a corpus, of data to learn the rules. Sentiment analysis has been handled as a Natural Language Processing denoted NLP, at many levels of granularity. Starting from being a document level classification task, it has been handled at the sentence level and more recently at the phrase level. NLP is a field in computer science which involves making computers derive meaning from human language and input as a way of interacting with the real world.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, May 2022
Impact Factor: 6.252

2. Case-Based Reasoning (CBR) Case-Based Reasoning (CBR) is one of the techniques available to implement sentiment analysis. CBR is known by recalling the past successfully solved problems and use the same solutions to solve the current closely related problems identified some of the advantages of using CBR that CBR does not require an explicit domain model and so elicitation becomes a task of gathering care histories and CBR system can learn by acquiring new knowledge as cases. This and the application of database techniques make the maintenance of large columns of information easier.

III. System Architecture

The algorithm used here is Random Forest. Random Forest is the most popular and powerful algorithm of machine learning.

3.1 Random Forest

- Step 1: Assume N as number of training samples and M as number of variables within the classifier.
- Step 2: The number m as input variables to decide the decision at each node of the tree; m should be much less than M.
- Step 3: Consider training set by picking n times with replacement from all N available training samples. Use the remaining of the cases to estimate the error of the tree, by forecasting their classes.
- Step 4: Randomly select m variables for each node on which to base the choice at that node. Evaluate the best split based on these m variables in the training set.
- Step 5: Each tree is fully grown and not pruned (as may be done in constructing a normal tree classifier). For forecasting, a new sample is pushed down the tree. It is assigned the label of the training sample in the terminal node it ends up in. This procedure is repeated over all trees in the ensemble, and the average vote of all trees is reported as random forest prediction. i.e. classifier having most votes.

3.2 Mathematical Model

The mathematical model for Shamming Detection
System is as-

S = {I, F, O}

Where, I = Set of inputs
The input consists of set of Words. It uses Twitter dataset.

F = Set of functions
F = {F1, F2,F3....FN}
F1: Tweets Extraction
F2: Tweets Preprocessing
F3: Feature Extraction
F4: Shamming Classification

Shamming Detection and Block Shamers

3.3 Dataset

A. Twitter

Twitter dataset is used for the classification purpose. In this social networking service users can freely communicate. They post and communicate with messages known as" tweets". Originally there was restriction of tweets characters that is 140, but from November 7, 2017, this limit was increased to 280 for all languages except Chinese, Japanese, and Korean. Registered users can post, like, and retweet tweets, but unregistered users can only read the messages. Users access Twitter through its website interface, through Short Message Service (SMS) or its mobile-device application software ("app"). Twitter, Inc. is based in San Francisco, California, and has more than 25 offices around the world.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, May 2022
Impact Factor: 6.252

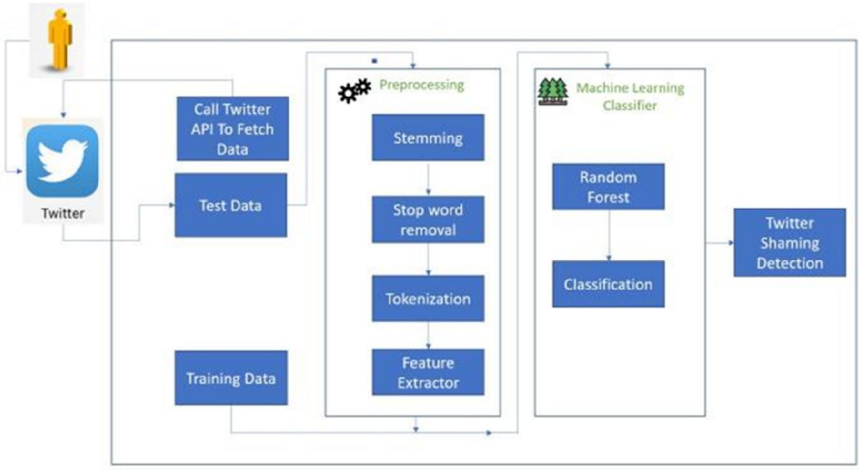

Chart 1: System Architecture

3.4 Data Flow Diagram

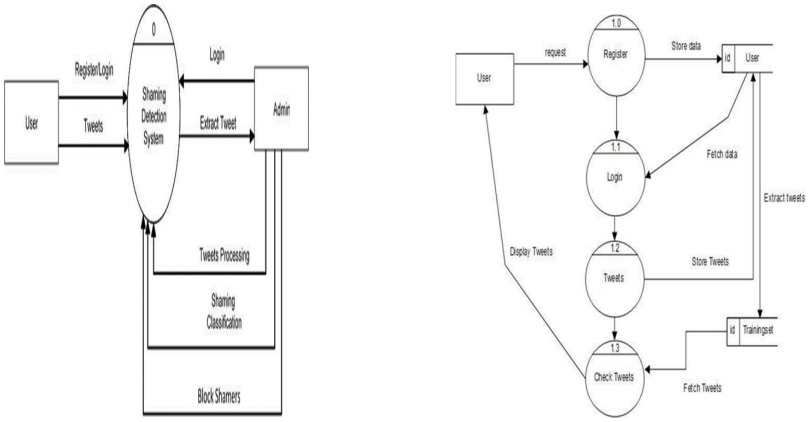

Figure 1: Data Flow Diagram

IV. CONCLUSION

Public detection has led to identify Shaming contents. Shaming words can be mined from social media. Shaming detection has become quite popular with its application. This system allows users to find offensive word count with the data and their overall polarity in percentage is calculated using classification by machine learning. But add some points incumbent on everyone to consider both contexts and consequences.

REFERENCES

[1]. Rajesh Basak, Shamik Sural, Senior Member, IEEE, niloy Ganguly, and Soumya K. Ghosh, Member, IEEE, "Online Public Shaming on Twitter: Detection, Analysis and Mitigation", IEEE Transaction on Computational Social System, Vol. 6, No. 2, APR 2019.

[2]. Guntur Budi Herwanto , Annisa Maulida Ningtyas , Kurniawan Eka Nu- grahaz , I Nyoman Prayana Trisna" Hate Speech and Abusive Language Classification using fastText" ISRITI 2019.

[3]. Chaya Libeskind, Shmuel Liebeskind" Identifying Abusive Comments in Hebrew Facebook" 2018 ICSEE.

[4]. Mukul Anand, Dr.R.Eswan" Classification of Abusive Comments in social media using Deep Learning" ICCMC 2019.

[5]. Dhamir Raniah Kiasati Desrul , Ade Romadhony" Abusive Language De- tection on Indonesian Online News Comments" ISRITI 2019