ONLINE SHAMING DETECTION ON TWITTER USING MACHINE LEARNING

Vaishnavi Supekar[1], Shivani titre[2], Anagha Dalvi[3], Pratiksha Kamble[4], Namrata Salunke [5],

Nutan Deshmukh [6]

MKSSS's Cummins College of Engineering for Women, Pune

Abstract - **In this paper, pro- posed mechanize the errand of open disgracing detection in Twitter. Disgracing tweets are sorted into six kinds: oppressive, correlation, condemning, strict, jokes on close to home issues, and whataboutery, and each tweet is characterized into one of these sorts or as no disgracing. It is evident that in all the actions of an interested buyer who submits a comment in a particular case, a large part of them will likely kill the person in question. Curiously, it is likewise the disgracing whose supporter checks increment quicker than that of the regarded in Twitter. At long last, in light of arrangement and grouping of disgracing tweets.**

Keywords— **oppressive, correlation, condemning, whataboutery.**

- INTRODUCTION

It will be an online communication network defined as using dedicated websites that allow users to interact with other users or find people with similar social media platforms that allow people around the world to stay in touch no matter what their age. Children in particular are being introduced to a horrible world of bad experiences and abuse. Users of social networking sites may be unaware of a number of potential dangers to attackers. Today the Internet has become a part of people's daily lives, people use social media to share photos, music, videos, etc. Social networks like Facebook LinkedIn, Myspace, and twitter have become very popular lately. The discovery of obscene language is the task of analyzing the natural language that is responsible for determining whether there is discrimination (e.g., religious affiliation, racism, freedom, etc.) in a given document and organizing the file document accordingly. The text that will be split to get offensive wording is in English text format that can be published via tweets, comments on social networks, movie reviews, political reviews, and comments.

- MOTIVATION

  Nowadays, social networking sites involve billions of users worldwide.
  User communication with these social networking sites, such as twitter, has significant and unpleasant consequences at times in everyday life.
  Trollers are trying to disrupt meaningful messages to online communities by sending unwanted comments.

- OBJECTIVES OF PROJECT
  To classify and automatic sorting of disgracing tweets. To provide information on shameful and embarrassing incidents.

- DATABASE REQUIREMENTS
  MySQL Database

  Software requirements (platform choice)

  We need software requirements such as The Operating System of windows 7 & above .The application server we are using is Apache tomcat 7 & above .An application that is written in Java .HTML, JDK 1.8 & JSP used for the front end designed for UI design. The Java script is used by the user to connect to a web page. Java Server Pages is a collection of software-assisted software. developers to dynamically create pages based on Html, XML etc. We use Mysql because the data is used for general purposes, e-commerce and login request.

  Hardware requirements

  Hardware requirements we need are Pentium Processor-III at 1.1 GHz and RAM is 256 MB (min). Hard disk Must be 20 GB with Floppy drive 1.44 MB. Keyboard is pf Standard Windows Keyboard and Mouse with two or three buttons. and the Monitor of SVGA
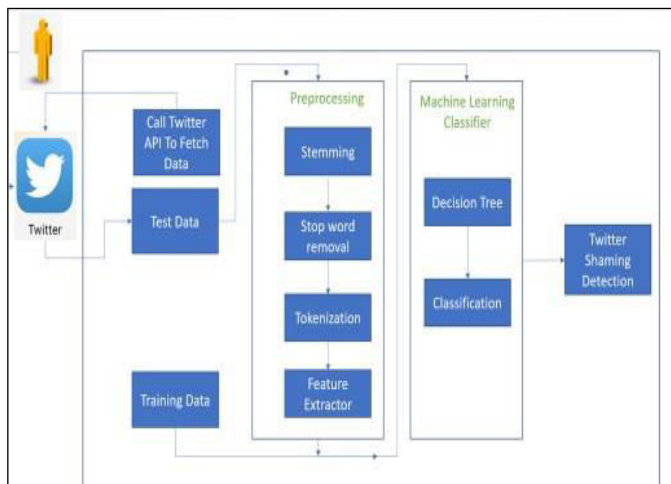
- *SYSTEM ARCHITECTURE*

Figure 1: System architecture

User Classes and Characteristics

- *User*
  - Registration
  - Login
  - Send tweets
  - View tweets
  - Logout

- *Admin Module*
  - Login
  - View Users
  - Authorize Users
  - Find disgracing tweets
  - Block disgracing tweets

- *User Interfaces*
  - Home page
  - User and Admin Login Page
  - Data Collection Page,
  - Preprocessing Page, Classification Page.
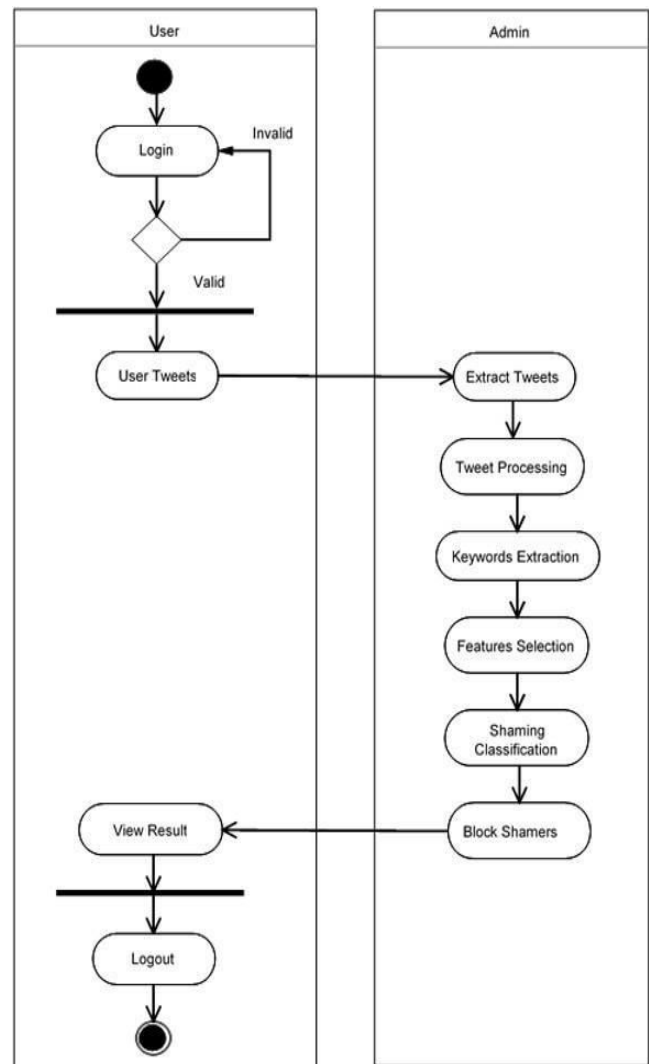 - Result Page.

- *ACTIVITY DIAGRAM:*

Figure 2: Activity diagram

- *TECHNOLOGY*
  - Java: Java programming language (technology). Originally developed by James's gosling in sun microsystems, they are now part of Oracle Corporation. It was released in 1995 as part of the solar system. Java platform. Language has improved a lot of its syntax from c and c ++ Java application is often integrated into a byte code (section file) that can be done on any virtual machine (JVM). Java is currently one among the foremost widely used programming languages. It has an estimated 10 million users.
  - better than most web-based languages out there - certainly higher than powerful scripts like Python, PHP, Perl, and Ruby.Static typing will help when the app gets large (you'll catch errors at compile time instead of run time).
  - Modern IDEs like intelligent and eclipse are great for analyzing your code, and allow you to view the call

management category, type the management category, and use methods and categories.

MySQL: - MySQL is a fast, easy-to-use. MySQL is being developed, marketed, and supported by MySQL, a Swedish company.

- MySQL is released under an open-source license. So, you have nothing to pay for it.
- MySQL is a powerful program.
- It handles a huge amount of data and functionality of the most expensive and powerful database packages.
- MySQL uses a standard form for popular SQL data languages.
- MySQL works on multiple operating systems and has multiple languages including PHP, PERL, C, C ++, Java, etc.
- MySQL is very friendly to PHP, the language most interested in web development.
- MySQL is customized. The open-source GPL license allows system developers to modify MySQL software in their specific areas.

- *TEST CASES AND TEST RESULTS*

Test Case_ID	Description	Test case I/P	Actual Result	Expected result	Test case criteria (P/F)
101	Enter the case insensitive Username click on Submit button.	Username	Error comes	Error Should come	P
102	Enter the case sensitive Username click on Submit button.	Username	Accept	Accept Username	P
201	Enter the case insensitive Password click on Submit button.	Password	Error comes	Error Should come	P
202	Enter the case sensitive Password click on Submit button	Password	Accept	Accept	P
301	Enter the case insensitive Mobile Number click on Submit button	Mobile Number	Error comes	Error Should come	P
302	Enter the case sensitive Mobile Number click on Submit button.	Mobile Number	Accept	Accept	P

Table 1: Test cases and test results

- *Test Cases:*

Testing of project problem statement using generated test data (using mathematical models, GUI, Function testing principles, if any) selection and appropriate use of testing tools, testing of UML diagram's reliability. We use Unit Testing using JUnit. Module-ID: -

- Modules to be tested: -Registration

- Enter the case insensitive Username click on Submit button.
  Expected: It should display error.
- Enter the case sensitive Username click on Submit button.
  Expected: It should accept.
- Enter the case insensitive Password click on Submit button.
  Expected: It should display error.
- Enter the case sensitive Password click on Submit button.
  Expected: It should accept.
- Enter the case insensitive Mobile Number click on Submit button.
  Expected: It should display error.
- Enter the case sensitive Mobile Number click on Submit button.
  Expected: It should accept.
- Enter the wrong address and click on Submit button.
  Expected: It should display error.
- Enter the correct address and click on Submit button.
  Expected: It should accept.

- *ALGORITHM:*

Random Forest

- Assume N as range of training samples and M as range of variables in the classifier.
- The amount m as input variables to determine the choice at each node of the tree; m ought to be a lot of lower than M.
- Think about training set by selecting n times with replacement from all N avail- ready training samples. Use the remaining of the cases to estimate the error of the tree, by prediction their classes.
- In random way choose m variables for every node on which to base the selection at that node. valuate the most effective split supported these m variables within the training set.
- Each tree is full-grown and not cropped (as could also be tired constructing a standard tree classifier).

Sentiment Analysis Algorithm:

Input: Text File (comment or review) T, The sentiment lexicon Output: Smt = {P, Ng and} and strength S where P: Positive, Ng: Negative, N: Neutral • Initialization: SumPos = SumNeg =0, where, SumPos: accumulates the polarity of positive tokens tismt in T, SumNeg:accumulates the polarity of negative tokens ti-smt in T,

- Begin
- for every ti in TF
- do
- Search for ti in lex
- if ti in P-list
- then

- SPos ← SPos + ti-st
- else if ti in p-list
- then
- SNeg ←SNeg + ti-st
- End if
- End for
- if SPos > |SNeg| then
- st = Pos
- S = SPOS / (SPos + SNeg)
- else if SPos < |SNeg|
- then
- St= Neg
- S = SPos / (SPos + SNeg)
- end if
- end

- *FUTURE SCOPE*

  The projected System are often enforced as a twitter feature to seek out shaming content on twitter. we've worked with solely the terribly simplest models; we are able to improve those models by adding further data like closeness of the word with a negation word. We can decide to model human confidence in our system. as an instance if we have five human labellers labelling every tweet, we can plot the tweet within the 2-dimensional sound judgment / sound judgement and positivism / negativity plane whereas differentiating between tweets during which all 5 labels agree, only four agree, only three agree or no majority vote is reached. we have a tendency to might develop our custom price perform for developing with optimized category boundaries specified highest weightage is given to those tweets during which all five labels agree and because the variety of agreements begin decreasing, therefore do the weights assigned. during this approach the consequences of human confidence are often visualized in sentiment analysis.

- *CONCLUSION*

  Disgracing tweet detection has lead to identify disgracing contents. Disgracing words can be extracted from social media. Applications allows the user to get disrespect full words with the data and their overall polarity in percentage is calculated using classification by using machine learning. possible solution for responding the risk of online Public Disgracing in twitter by classifying the disgracing comments in six types, choosing appropriate features and designing a set of classifiers to detect it.

- *REFRENCES*

1. https://scholar.google.com/citations?user=V6fBDKgAAAAJ&hl=en

2. https://dl.acm.org/doi/10.1145/2998181.2998213

3. https://ieeexplore.ieee.org/document/7937783

4. https://inpressco.com/wp-content/uploads/2021/02/Paper208997-1000.pdf

5. https://Statistical_Twitter_Spam_Detection_Demystified_Per.pdf