IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-5, Issue-8 E-ISSN: 2347-2693

Predicting and Detecting Hectoring on Social Media Using Machine Learning

Sakshi Gujral

Dept. of CSE, Indira Gandhi Delhi Technical University for Women, Delhi, India

**Corresponding Author:gujralsakshi8@gmail.com, Tel:91+7042618283*

Available online at: www.ijcseonline.org

Received: 20/Jul/2017,Revised: 30/Jul/2017, Accepted: 19/Aug/2017, Published: 30/Aug/2017

Abstract- The increase of use of Social networking sites in recent years has both pros and consequences. The idea is to have safer use of these Social media sites so as to obtain maximum benefits from them rather than having malicious effects from them. One of the misuses of these social networking sites like Twitter, Facebook, and Instagram is posting of absurd contents over these Social Medias. This content can be extremely harmful as it causes insult, depression, anxiety, peer pressure. This needs to be detected and reported for better use of social media. This measure will make an approach for better use of Internet yard. Hence, this research work aims at Predicting and detecting these harassing comments with the help of Machine learning Algorithms. The idea is to do Sentimental Analysis of the tweets obtained from Twitter, Pre-process them, apply Machine Learning algorithms along with Bag-Of-Words on the tweets and classify the tweets as positive and negative. Tweets classified as negative will help to detect the tweet as bullying or not. The proposed research work uses bag of words approach along with Laplace version of the Naive-Bayes classifier with Laplace function for Detection. For Prediction CART model with Bag-of-Words approach is used. The platform used here is R studio with various packages.

Keywords- Prediction, Detection, Hectoring, Bag-of-word ,Laplace, Confusion Matrix etc.

I. INTRODUCTION

Hectoring is a form of Bullying that make use of force, threat, or coercion to abuse, intimidate or aggressively dominate others. The behavior is often repeated and habitual. One essential prerequisite is the perception, by the bully or by others, of an imbalance of social or physical power, which distinguishes bullying from conflict. Behaviors used to assert such domination can include verbal harassment or threat, physical assault or coercion, and such acts may be directed repeatedly towards particular targets. Rationalizations of such behavior sometimes include differences of social class, race, religion, gender, sexual orientation, appearance, behavior, body language, personality, reputation, lineage, strength, size, or ability. If bullying is done by a group, it is called mobbing. Bullying ranges from one-on-one, individual bullying through to group bullying called mobbing, in which the bully may have one or more "lieutenants" who may seem to be willing to assist the primary bully in his or her bullying activities. Bullying in school and the workplace is also referred to as peer abuse

a) Individual bullying can be classified into four types. Collective bullying is known as mobbing, and can include any of the individual types of bullying.

b) Physical, verbal, and relational bullying are most prevalent in primary school and could also begin much earlier whilst continuing into later stages in individuals lives. It is stated that Cyber-bullying is more common in secondary school than in primary school.

Cyber-bullying[1]

Cyber bullying is the use of technology to harass, threaten, embarrass, or target another person. When an adult is involved, it may meet the definition of cyber-harassment or Cyber stalking, a crime that can have legal consequences and involve jail time. This includes email, instant messaging, social networking sites (such as Facebook), text messages, and cell phones.

Hectoring over various Social Media sites leading to adverse effects

Fig 1-Cyber-Bullying[1]

[1] https://www.worldpulse.com/sites/default/files/styles/post_cover_image/public/post/22246/65258/post_cover_image/6875666d4ff9736630e57ff55f2b52cd/cyber-bullying-finalcolor.png?itok=zT5k7_e_

II. LITERATURE REVIEW

The area of hectoring is relatively new field as use of social media has increased in last few years. It is still a research area to work upon. The absence of appropriate data sets for the hectoring makes it more challenging to deal with this problem. The literature survey indicates it as a text classification problem. So, various work related to sentimental analysis of twitter data is studied in this work.[2]

The literature Survey shows that text classification is dealt with supervised learning, unsupervised learning, tokenization and the most appropriate one is the hybridization of both machine learning algorithmsand Natural Language processing[3]. Literature helps to give us an idea of the work done for Sentimental Analysis as well as hectoring Detection. Since, Data obtained from the Social Networking sites is massive so some of the research work emphasize oh Map-Reduce framework[4].

Few Research work make use of Recommendation engir that stores the pre-defined values in the Databases that hel in evaluating the input with pre-defined results. Researc works also make use of Lexical analysis[5] only and some if the work only considers Un-Supervised or supervise learning only[6]. In that scenario, accuracy is litt compromised. Content driven and network based approach also discussed in two research paper[7].

SVM classifier is used frequently in Literature worl However, Naïve Bayes is highly recommended for te classification. Text along with Image analysis is also done i one of the Research Work[8]. The data sets are generall obtained from Instagram.Pre-defined Data set available fo University but Twitter Data-Set for hectoring is not done in Literature. Various works are done for different Data-Set and compared.

III. PROPOSED WORK

After framing the research problem, a predictive prototype is defined for hybrid model comprising of the following iterative phases progressing sequentially in attaining the goal.

1. At stage 1, tweets are collected using twitter API and R Studio with the credentials. A balanced data set is obtained using bullying key-words like fat, ugly, freckles ,ginger etc. and non-bullying words like beautiful, love, care, support etc.
2. Data Enrichment acted as the second stage in which data filtration method is implemented to curtail the appropriate and inconsistent data from the source data, resulting in significant data cleaning as tweets are quite noisy. Data is labeled as positive or negative depending upon sentiment.
3. At third stage, tokenization is done using Bag-Of-Words Approach with unigram feature extraction.
4. Feature reduction is now done with the help of term-frequency using packages in R studio.
5. Data Set is subject to Prediction model with the help of CART, Decision Trees, Random Forest, Logistic Regression, accuracy is obtained and results are compared.
6. Now, Naive Bayes classifier is used with modified version of Laplace function to detect the tweets potentially as bullied or not.

Prediction Module-CART with BOW-

CART (Classification and Regression Trees) is a modern, c.1984, flavor of data mining that employs decision trees and can be used for a variety of business and scientific applications. Its advantages include quick insight into database patterns and significant relationships using simple tools such as graphs, charts and reports. It is highly useful for sentimental analysis.

Combining CART with bag of words is an effective approach and will increase the accuracy. As, it is One of the most used techniques to transforms text into independent variables is that called Bag of Words**.**

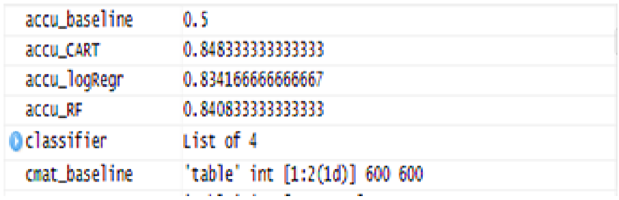
Laplace Naïve Bayes

A variation of the Naive Bayes algorithm known as Laplace Naive Bayes. In this method, the term frequencies are replaced by Boolean presence/absence features. The logic behind this being that for sentiment classification, word occurrence matters more than word frequency. Selection of this algorithm is done as it work well with the noisy data as well as easy to implement and literature surveys also reveals that this out-performs other classifiers when use with Bag-of-words unigram features.

IV. RESULTS

Proposed work is implemented with R-Studio with text mining and machine learning algorithm packages. Accuracies has been compared for various predictive algorithms. CART has been generated for the Data. Confusion matrix is built and various parameters are obtained and discussed for Detection module.

Results Of Models



accu_baseline	0.5
accu_CART	0.848333333333333
accu_logRegr	0.834166666666667
accu_RF	0.840833333333333
classifier	List of 4
cmat_baseline	'table' int [1:2(1d)] 600 600

Fig 2-Accuracies of Prediction Models

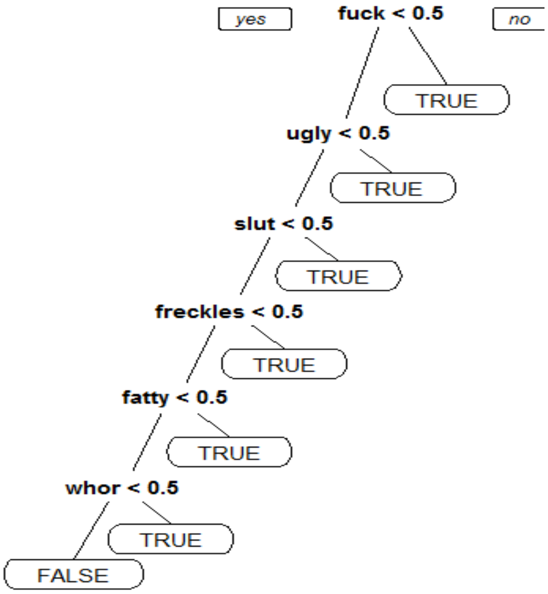


Fig 3-CART Generation

Confusion Matrices & Statistics

In the field of machine learning and specifically the problem of statistical classification, a confusion matrix, also known as an error matrix, is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one (in unsupervised learning it is usually called a matching matrix).

Sensitivity (TPR) = -also called the true positive rate, the recall, or probability of detection in some fields) measures the proportion of positives that are correctly identified as such (i.e. the percentage of sick people who are correctly identified as having the condition).

TPR=TP/P

Specificity (TNR)-also called the true negative rate measures the proportion of negatives that are correctly identified as such (i.e., the percentage of healthy people who are correctly identified as not having the condition).

TNR =TN/N

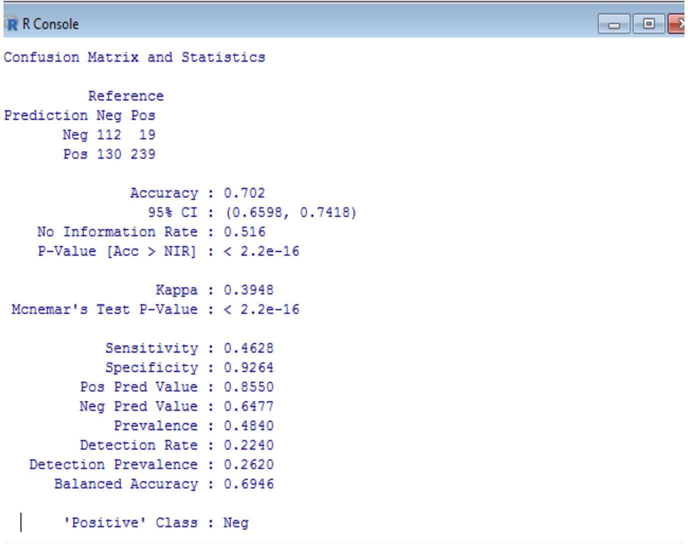
Precision (PPV))- Precision is the fraction of the documents retrieved that are relevant to the user's information need.
PPV=TP/(TN+FN

Accuracy=Given a set of data points from a series of measurements, the set can be said to be *precise* if the values are close to the *average value* of the quantity being measured, while the set can be said to be *accurate* if the values are close to the *true value* of the quantity being measured.

Accuracy= (TP+TN)/ (TP+FP+FN+TN)

Kappa Measure

The kappa statistic is a measure of how closely the instances classified by the *machine learning classifier* matched the data labeled as *ground truth*, controlling for the accuracy of a random classifier as measured by the expected accuracyKappa= (observed accuracy-expected accuracy)/ (1-expected accuracy)



```
Confusion Matrix and Statistics

          Reference
Prediction Neg Pos
       Neg 112  19
       Pos 130 239

               Accuracy : 0.702
                 95% CI : (0.6598, 0.7418)
    No Information Rate : 0.516
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.3948
 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.4628
            Specificity : 0.9264
         Pos Pred Value : 0.8550
         Neg Pred Value : 0.6477
             Prevalence : 0.4840
         Detection Rate : 0.2240
   Detection Prevalence : 0.2620
      Balanced Accuracy : 0.6946

       'Positive' Class : Neg
```

Fig 4-Detection Module Results

V. CONCLUSION

Prediction of tweets for hectored contents is quite important for the Social Networking Host as it can help to mitigate the adverse effects of depression, anxiet, suicides, murders. Moreover, it helps in maintaining the image of the host of Social Network in the market.This will help in better business.CART model does sophisticated job for the same.It has outperform other models when combined with Bag-of-words approach.This type of work is absent in the literature as there is absence of sophisticated Data Set.Moreover,hectoring is a burning topic that has been triggeered with burgeon use of Social network in recent times.CART gives 84.8% Accuracy that is quite appealing.

Dection of hectoring tweets is also studied in this work.Hectoring is an offensive crime that can led to suicide,murder,depression especially by school students.Hence,it's timely detection is important whenerver a person reports anything about it.This approach uses both bag-of –words approach with Uni-Gram feature as single hectored word can also cause adversity with Laplace Naïve Bayes gives fair results of 70.2%.

VI. FUTURE SCOPE

The tweets can be collected with more hectoring keywords so as to minimise maximum adversities.Since,Social Networking generates Data in TB'S and this Data is increasing at faster

Pace so Big data comes into picture for Data Anlysis.Data can be collected from multiple Social Networks like Facebook,Instagram,Twitte,You-Tube,Meet-Up etc.Along with contents like tweets,Facebook status,posts,Image analysis,videoes analysis can also be done,As Image also forms great source of bullying.In that case R studio can be used wit Spark framework to deal big Data.

REFERENCES

[1]. Bello-Orgaz, Gema, Jason J. Jung, and David Camacho. "Social big data: Recent achievements and new challenges." *Information Fusion* 28 (2016): 45-59.

[2]. Frommholz, Ingo, Haider M. Al-Khateeb et all"On textual analysis and machine learning for cyberstalking detection." *Datenbank-Spektrum* 16, no. 2 (2016),pp-127-135..

[3]. Al-garadi, Mohammed Ali, Kasturi Dewi Varathan, and Sri Devi Ravana. "Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network." *Computers in Human Behavior* 63 (2016), pp. 433-443.

[4]. Chen, Ying, Yilu Zhou et all "Detecting offensive language in social media to protect adolescent online safety." In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, pp. 71-80. IEEE, 2012..

[5]. Parker, Shawniece L., and Yen-Hung Hu. "Content Mining Techniques for Detecting Cyberbullying in Social Media." *Virginia Journal of Science* 67, no. 3 (2016),pp 1-8.

[6]. Sheela, L. Jaba. "A Review of Sentiment Analysis in Twitter Data Using Hadoop." *International Journal of Database Theory and Application* 9, no. 1 (2016): 77-86.

[7]. Zhao, Rui, Anna Zhou, and Kezhi Mao. "Automatic detection of cyberbullying on social networks based on bullying features." In *Proceedings of the 17th international conference on distributed computing and networking*, p. 43. ACM, 2016.

[8]. Ismail, Mohamed Maher Ben, and Ouiem Bchir. "Insult detection in social network comments using possibilistic based fusion approach." In *Computer and Information Science*, pp. 15-25. Springer International Publishing, 2015.

[9]. Nagar, Himanshu, Chetna Dabas, and J. P. Gupta. "*Navie Bayes and K-Means Hybrid Analysis for Extracting Extremist Tweets*", *ACM Conference,* pp 27-32.

Author Profile

Sakshi Gujral was born in Delhi,1991.She completed her Bachelor in Technology from Guru Gobind Singh Indraprastha University, Delhi in Computer Science and Engineering. Passion for Computer Science enthusiast her to pursue Master in this field.Sakshi Has just defended her Masters in Technology from Indira Gandhi Delhi Technical University in Mobile & Pervasive Computing. Her research areas include IOT, Data Analytics ,Data Mining, Big-Data.