I.J. Mathematical Sciences and Computing, **2019, 4, 1-12**
Published Online November 2019 in MECS (http://www.mecs-press.net)
DOI: 10.5815/ijmsc.2019.04.01

Available online at http://www.mecs-press.net/ijmsc

Cyber Bullying Detection and Classification using Multinomial Naïve Bayes and Fuzzy Logic

Arnisha Akhter, Uzzal K. Acharjee, Md Masbaul A. Polash

Jagannath University, Dhaka-1100, Bangladesh

Received: 03 July 2019; Accepted: 11 August 2019; Published: 08 November 2019

Abstract
The advent of different social networking sites has enabled people to easily connect all over the world and share their interests. However, Social Networking Sites are providing opportunities for cyber bullying activities that poses significant threat to physical and mental health of the victims. Social media platforms like Facebook, Twitter, Instagram etc. are vulnerable to cyber bullying and incidents like these are very common now-a-days. A large number of victims may be saved from the impacts of cyber bullying if it can be detected and the criminals are identified. In this work, a machine learning based approach is proposed to detect cyber bullying activities from social network data. Multinomial Naïve Bayes classifier is used to classify the type of bullying. With training, the algorithm classifies cyber bullying as- Shaming, Sexual harassment and Racism. Experimental results show that the accuracy of the classifier for considered data set is 88.76%. Fuzzy rule sets are designed as well to specify the strength of different types of bullying.

Index Terms: Cyber Bullying, Multinomial naïve bayes classifier, Support vector machine, Fuzzy logic.

1. Introduction

The act of forcing, threatening or compulsion to abuse, threaten or violently dominate others is depicted as bullying. Cyber bullying is bullying by means of electronic media. This may come into existence in the form of posting rumors, threats, sexual remarks, a victim's personal information or pejorative labels. Young people are increasingly becoming susceptible to such harassment; a global mobile phone giant has found that in a new region-wide survey. The survey found that 49 percent of school students in Bangladesh have been victims of cyber bullying in a way or other [1]. In recent times, cyber bullying is growing at a remarkable magnitude in Bangladesh. Bullying or harassment can be identified by repeated behavior and intent to harm. The consequences

* Corresponding author.
E-mail address:

of cyber bullying become vital when the victim fails to cope with the emotional strain from abusive, threatening, humiliating and aggressive messages. The victims may have lower self-esteem, may grow suicidal tendency and a variety of emotional responses, including being scared, frustrated, angry and depressed. Sometimes cyber bullying comes out to be more harmful than traditional bullying.

The Cyber bullying Research Centre's research [2] showed that in 2013 about one in four teens had been the victim of cyber bullying and one in six teens was involved in the bullying. Their research also shows that cyber bullying is on the rise in every study. As both teens and adults continue to have an increased online presence; this number is expected to increase in near future.

The computational detection of cyber bullying is done based on the many classes of algorithms in the fields of machine learning. Natural language processing is another tool for social textual interaction analysis. The phenomena of social interaction analysis and sociolinguistics places an emphasis on specificity, uniqueness, the presence of the effect, ascription to the community and the personality of the individuals and their use of language; while statistical supervised and unsupervised methods emphasize generalization, abstraction and an exploitation of patterns in the data. The fields of social interaction analysis and sociolinguistics have a significant chasm and dissonance with the fields of machine learning and natural language processing. Another motivating factor behind this work is to investigate effectiveness of parameterization approach to exert the full power and weight of statistical machine learning and natural language processing involves the analysis of relevant parameters from the fields of sociology, psychiatry, and sociolinguistics, all three of which have been studying the phenomenon of bullying and meanness for decades [6].

Detection of cyber bullying and provision of subsequent preventive measures are the main courses of action to combat cyber bullying. The detection method should identify the presence of cyber bullying terms and classify cyber bullying activities in social networks such as Flaming, Harassment, Racism, and terrorism. Different types of bullying causes for different reasons and have diverse effect on the bullied person. So only identification of bully is not the major concern, detection of the types of bully is also a necessary objective in this field to mitigate the effects of cyber bullying. Therefore, classification of the bully type is as much important as identification to alleviate the risk of physical and mental damage of the bullied person. Again bullying is not always done with the same intention and same level of intensity. It differs in the varying altitude of strength of bullying, age of the person making bully and other things like social position and educational qualification of the person being bullied. Psychiatric conditions and sociolinguistic patterns are very much tough to be considered in identification of cyber bullying from analyzing just the social network data. Concealing of personal information in social network sites is common in people who are associated with cyber crime. Collection of required information is somewhat difficult for the cyber bully detection team and also difficult for the law enforcement agencies to find out the criminal.

In this paper, a system for identification and classification of cyber bully from Facebook comments is proposed. The system can be broadly divided into two parts: (1) identification and classification of cyber bully using multinomial naïve bayes classifier and (2) identification of bully strength using fuzzy rule sets. Multinomial naïve bayes classifier is used for multi class classification of text data. Cyber bullies can be classified with different types and each one is done with different intentions. Sexual harassment, racism and shaming have been taken in consideration for this work. A set of Fuzzy rules are defined for this system. These rules are applied on the identified bully comments to measure the strength of bully. The strength of bully signifies if there is any potential risk or not that is described in the methodology section.

The rest of the paper is organized as follows. Section 2 represents the related works and works using almost same technologies and different technologies used by other researchers of this field. The proposed methodology and work flow are described in section 3. Section 4 shows the output of the proposed system and performance comparison. Conclusions along with future enhancements possible are detailed in section 5.

2. Related Works

Consciousness about the effects and risks of cyber bullying is in rise that results in a number of works for

the detection and prevention of cyber bullying. An approach has been made for detecting cyber bullying in social networking sites using fuzzy logic and genetic algorithm [3]. Fuzzy rule sets are provided as knowledge of learning algorithm. For modeling adaptive and exploratory behaviors, the learning algorithm is made of a genetic algorithm. The output has classified bullying words present in the conversation.

Another approach aims to develop a set of tools for automatically detecting malicious entries and reporting them to PTA (Parent-Teacher Association) members [4]. Cyber bullying data collected from unofficial school Web sites are analyzed in two ways. Firstly, the entries are analyzed with a multifaceted affect analysis system with the aim of finding distinctive features for cyber bullying and apply them to a machine learning classifier. Secondly, a SVM based machine learning method is applied to train a classifier to detect cyber bullying. The system classifies cyber bullying entries with 88.2% of balanced F-score.

A system detects cyber bullying in YouTube video comments. Sexuality, culture, intelligence, and physical attributes are considered for classification of comments and determining what topic it is. The shortcoming of the system is that it performs less precise classification outcome and more false positives.

A number of natural language processing techniques are used to identify bullying traces [7]. Online and offline instances of bullying are traced. Sentiment analysis system is used to identify bullying and Latent Dirichlet Analysis to identify topics. The bullying instances are not accurately detected in this method.

Instagram images and associated comments are labeled using human labelers at the crowd-sourced CrowdFflower Website [8]. The labeled data analysis and correlations between different features are presented. Accuracy of classifiers for automatic detection of cyber bullying incidents are designed and evaluated using the labeled data.

Dinakar et al. 2011 [9] determined that it was possible to get better results by first labelling bullying into categories and then using a binary classifier for every category. They used sexuality, race/culture and intelligence as categories. A decision tree using JRip (an implementation of the propositional rule learning algorithm RIPPER) achieved the best accuracy while an SVM using the Sequential Minimal Optimization (SMO) algorithm for training was the most reliable method. The test data consisted of comments on Youtube videos.

In 2011 Reynolds et al. [10] used a data set from Formspring.me and rated the documents based on occurrences of bad words. They then used a machine learning tool to train a few different classifiers including decision trees and a support vector machine. A decision tree using the C4.5 algorithm performed best and reached a recall level of 78.5 percent.

To further improve classification, gender information can be used as shown by [5]. This study takes advantage of the fact that males and females use different types of vocabularies. More specifically there is a difference in which curse words are typically used. They then trained two separate SVM classifiers for the male and female test cases. The results showed an increase from 31 to 43 percent in precision, from 15 to 16 percent in recall and from 20 to 23 percent in F-measure.

Another approach made by researchers from University of San Carlos [11] has harvested the cyber bullying related Facebook posts using a customized web scraper tool. Support Vector Machines (SVM) model is used for classification of these harvested data. For the considered data set, the study achieves the precision of 88% and the recall is 87%.One major problem is only 24 posts from a public page can be harvested using Facebook graph API. The limitation also expands in the field that some training data for the SVM (Support Vector Machine) model may have different interpretations. SVM cannot define thereat or curse differently which causes wrong classification.

VandanaNandakumar et al. has done another survey on Twitter data using Naive Bayes classifier algorithm and Support Vector Machine model [12]. The feature probabilities are calculated using Naïve Bayes Classifier Algorithm. A graph is plotted comparing among the two algorithms, Naïve Bayes Classifier Algorithm and Support Vector Machine. Comparison on the basis of precision factor is also done with the fact that the probabilities for each feature set are calculated independently from the twitter dataset and can evaluate the performance by predicting the output variable. The plotted graph shows that naïve bayes classifier shows better precision than support vector machine model. We can conclude that for text data classification Naive bayes

classifier shows better performance than the SVM model [12]. Again the previous work noted in [11] makes use of Facebook posts to identify and classify cyber bullying. In reality, people find it more embarrassing when there is bully in their posts done by others as comments. Facebook is the most used social networking site in Bangladesh and events of cyber bullying in Facebook posts is very common. People show offenses in comments of posts of celebrities, media persons, writers, politicians and what not. The posts may include image, sharing personal experience and thinking, motivational words or anything shared in Facebook. A system is modeled to identify and classify the bullying present in Facebook comments.

The process of detecting cyber bully activities begins with input dataset from social network. For this work Facebook is chosen. Input dataset is text comments posted in Facebook images and posts. Dataset is collected from the online data source kaggle.com (Dataset: Internet bulling for facebook comments). Figure 1 illustrates the block diagram of the proposed cyber bully detection and classification system.

Data pre-processing is done on input data to improve the quality of the research data. Subsequent analytical steps include removing stop words, extra characters and hyperlinks. Feature extraction is done after performing pre-processing on the input data. Features like Noun, Adjective and Pronoun from the text are obtained by feature extraction and frequency of words in the text are determined. The extracted features are then given to the Classification Algorithm.

Multinomial Naïve Bayes algorithm is used for the classification of comments for identifying bullying types. The classifier system is first trained with the training data set collected from Kaggle.com. A part (almost 20%) of the data set is considered as the test data set. Output of this classification algorithm indicates if the given data (comment) contains bullying words or not. This system is designed to detect the presence of three types of bullying- sexual harassment, racism and terrorism. This is a supervised approach because it uses tagged or labeled training data set.

Fuzzy rule sets are generated for identifying the strength of the bullying types. Along with comments, the gender and age of the commenter are also considered. Bad words (words signifying different bullying) are counted for each comment. Data dictionary for individual bullying types are built. Matching with the dictionary, number of bad words for each comment are counted.

Fuzzy rule sets are generated using three parameters- gender of the commenter, age of the commenter and the number of bad words contained in the comment. Conventionally, Fuzzy rule outputs are like – high, medium and low; indicating the potency of bully.

A. Collection of Data Set

This is the very first step in classification work. A data set is a collection of related, discrete items of related data that may be accessed individually or in combination or managed as a whole entity. The system is trained with labeled data sets for supervised classification. Data set containing comments posted in Facebook along with the gender and age of the commenter. An online data source, kaggle.com provides specific data sets for research purpose.

This is the very first step in classification work. A data set is a collection of related, discrete items of related data that may be accessed individually or in combination or managed as a whole entity. The system is trained with labeled data sets for supervised classification. Data set containing comments posted in Facebook along with the gender and age of the commenter. An online data source, kaggle.com provides specific data sets for research purpose.

B. Data Preprocessing

Text can come in a variety of forms from a list of individual words, to sentences to multiple paragraphs with special characters. Data pre-processing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. Data pre-processing is a proven method for resolving such issues.

Data pre-processing prepares raw data for further processing.

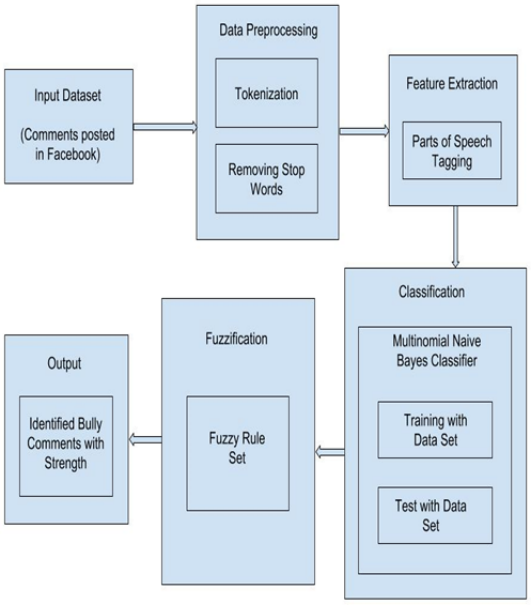

Fig. 1. Proposed cyber bully identification and classification system.

In this classification work, we have used NLTK, the Natural Language ToolKit that is one of the best-known and most-used NLP libraries in the Python ecosystem, useful for all sorts of tasks from removing stop words to tokenization, to part of speech tagging, and beyond.

Tokenization: Tokenization describes splitting paragraphs into sentences, or sentences into individual words. For the former Sentence Boundary Disambiguation (SBD) can be applied to create a list of individual sentences. This relies on a pre-trained, language specific algorithm like the Punkt Models from NLTK. This tokenizer divides a text into a list of sentences by using an unsupervised algorithm to build a model for abbreviation words, collocations and words that start sentences. It must be trained on a large collection of plain text in the target language before it can be used. The NLTK data package includes a pre-trained Punkt tokenizer for English. Sentences can be split into individual words and punctuation through a similar process. Most commonly this split across white spaces.

Removing stop words: A majority of the words in a given text are connecting parts of a sentence rather than showing subjects, objects or intent. Word like “the” or “and” can be removed by comparing text to a list of stop word.

C. Feature Extraction

Each individual word in a given sentence is tagged with appropriate parts of speech tags. Understanding parts of speech can make difference in determining the meaning of a sentence. Part of Speech (POS) often requires look at the proceeding and following words and combined with either a rule-based or stochastic method. It can then be combined with other processes for more feature engineering.

The NLTK toolkit provides PoS tagging attributes. Here, the parts of speeches must be defined like following:

- nounTags = [’NN’, ’NNP’, ’NNS’, ’NNPS’]
- verbTags = [’VB’, ’VBD’, ’VBG’, ’VBN’, ’VBP’, ’VBZ’]

- adjectiveTags = [’JJ’, ’JJR’, ’JJS’]
- adverbTags = [’RB’, ’RBS’, ’RBR’]

However, all PoS tags are not important for us to analyze. Verb and adverb tags do not have much significance in identification of bully characteristics. Only noun and adjective tags are taken in consideration for the next step for PoS tagging.

D. Classification using Multinomial Naïve Bayes Classifier

The considered dataset for running the proposed classification model is taken from kag- gle.com. A part of the whole data set is taken as the training data set; represented in Table 1. Traditionally 80% of the data set is considered as training data set. 800 comments from a dataset of 1000 have been taken as the training set. The data set contains individual comment, age of the commenter and the tag for bully type contained in respective comments. Age is not considered in the training phase; so truncated from training set.

TABLE 1. PART OF Training Data SET

Sentence	Bully Type	Age
fat rounded ball	Harassment	18
Wanna lick you	Harassment	19
you black ugly face	Racism	22
I love you beautiful	Others	33
that’s my nigga	Racism	62
You fat cow	Shaming	35

The test data set is also shown in Table 2 in brief. 200 comments from rest of the data set are considered as test data. The test data contains the comment and age of the commenter. Tags are removed as this data set is used for testing and the Multinomial Naive Bayes classifier classifies each comment with appropriate bully type.

TABLE 2. PART OF TEST DATA SET

Sentence	Age
Looking like a gorgeous bitch	20
ugly black u	35
Too fat and ugly	45
Wearing glasses?? Ugly granny	28
Too much hot and curvy	33

The output is presented in Table 3 as a part of the whole test set. The output contains each comment tagged with the proper bully type classified by the Multinomial Naive Bayes classifier. The percentage of each type of bully is also included in the output. The age from test set remains here because the identified bully comments along with age are taken in later part for the implementation of fuzzy rule sets.

TABLE 3. OUTPUT FOR MULTINOMIAL NAÏVE BAYES CLASSIFICATION

Sentence	Output	Racism (%)	Harassment (%)	Shaming (%)
Looking like a gorgeous bitch	Harassment	12.40	85.20	3.10
ugly black u	Racism	97.56	1.32	1.12
Too fat and ugly	Shaming	1.56	1.34	97.20
Wearing glasses?? Ugly granny	Shaming	0.00	1.44	98.56
Too much hot and curvy	Harassment	0.64	99.36	0.00

E. Fuzzy Rule Set

Human beings make decisions based on rules. Although, we may not be aware of it, all the decisions we make are based on computer like if-then statements. Fuzzy machines, which always tend to mimic the behavior of man, work the same way. However, the decision and the means of choosing that decision are replaced by fuzzy sets and the rules are replaced by fuzzy rules. Fuzzy rules also operate using a series of if-then statements. For instance, if X then A, if Y then B; where A and B are all sets of X and Y. Fuzzy rules define patches, which is the key idea of this.

The Fuzzy rule sets presented in table 07 are applied on the output of Multinomial Naïve Bayes. This classifier classifies the input data set in different bullying types and fuzzy rule sets are applied only on comments that contain bullies. Where multinomial naïve bayes classifies comments for different bully types, Fuzzy rule set identifies the strength of that bully type. As bullying is proved to be a repeated work, we define the strength of the bully type to realize if there is any potential risk of further bullying. We can also grasp an idea about if there is any risk in real life that can occur in near future. Because study finds, most of the bullying cases occur for real life hostility [12].

The strength of bullies is defined as- high, medium and low. The fuzzy output 'High' can be termed as extreme level of bullying where there is potential risk of further bullying and possibility to come in existence in real life from virtual. Proper measurements should be taken to prevent further bullies and mitigate future risks. The fuzzy output 'Medium' means the bully is neither extreme nor mild. There is a little potential future risk but it can be repeated in virtual. Steps are required to stop further bullying and lessen stress of the bullied. The fuzzy output 'Low' indicates that the bully is somewhat mild and it can be discontinued. A notification to the stalker may stop this.

There are two parameters taken for building fuzzy rule sets- age of the commenter, number of bad words counted in the comment (counted by matching with bad word list for different bully types). Three levels are declared for these parameters- Low, Medium and High; shown in Table 4. The output is the 'Strength of Bully', also has three levels to define the robustness of the comment.

TABLE 4. PARAMETERS AND LEVELS FOR FUZZY RULE SETS

No	Parameter	Level
01	Age	Low, Medium, High
02	No_Bad_Words	Low, Medium, High
03	Strength_of_Bully	Low, Medium, High

The fuzzy rules for age are shown in Table 5. Ages from 18-65 are taken in consideration as there is an age restriction for using Facebook. Fuzzy provides partial membership, that's why we can take same value for different classes. Age is taken for 18-35 as low, 30-45 as medium and 40-65 as high. Figure 2 illustrates the three

level membership functions for age. The number of bad words is counted from each bullied comment. These comments are the output of the multinomial naïve bayes algorithm applied on the input data set.

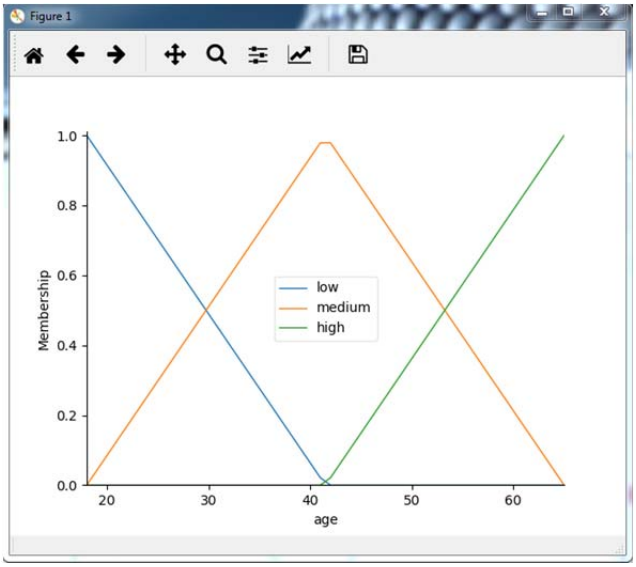

Fig. 2. Three level membership function for Age.

TABLE 5. FUZZY RULES FOR AGE

No	Rules
01	IF a>=18 AND a<=35 THEN Age= Low
02	IF a>=30 AND a<=45 THEN Age = Medium
03	IF a>=40 (AND a<=65) THEN Age = High

There are data dictionaries for each type of bullies. These dictionaries are collected from kaggle.com. The comments found out for different bully types are then matched with the data dictionaries to count the number of bad words. Partial levels for number of bad words are defined as: 1-3 as low, 2-4 as medium and 4-6 as high; depicted in Table 6. Figure 3 illustrates the three level membership function for number of bad words.

TABLE 6. FUZZY RULES FOR NUMBER OF BAD WORDS

No	Rules
01	IF n>=1 AND n<=3 THEN Bad_Words_Count = Low
02	IF n>=2 AND n<=5 THEN Bad_Words_Count = Medium
03	IF n>=4 AND n<=7 THEN Bad_Words_Count = High

Fuzzy rule sets for defining the strength of bullied comments are represented in Table 7. Two input parameters- age and number of bad words; each having three levels- as high, medium or low, produces nine set of rules. The output parameter is strength of the comments. This parameter is also categorized as high, medium or low as per defined by the rule sets.

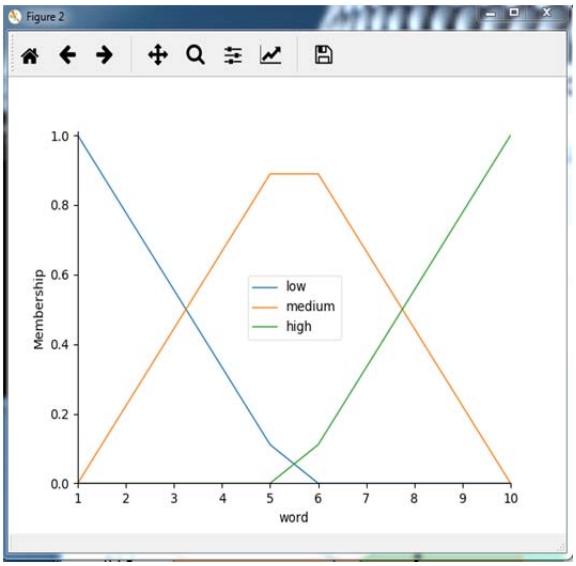

Fig. 3. Three level membership function for Number of Bad Words.

TABLE 7. FUZZY RULES FOR IDENTIFYING STRENGTH OF BULLY COMMENTS.

No	Fuzzy Rules
01	IF Age =Low AND Bad_Words_Count= Low THEN S = Low
02	IF Age =Low AND Bad_Words_Count= Low THEN S = Low
03	IF Age =Low AND Bad_Words_Count = High THEN S =High
04	IF Age =Medium AND Bad_Words_Count = Low THEN S = Medium
05	IF Age =Medium AND Bad_Words_Count = Medium THEN S = Medium
06	IF Age = Medium AND Bad_Words_Count = High THEN S = High
07	IF Age =High AND Bad_Words_Count = Low THEN S = Medium
08	Age =High AND Bad_Words_Count = Medium THEN S = High
09	IF Age =High AND Bad_Words_Count = High THEN S = High

The output for multinomial naïve bayes classification shown in Table 3 identifies and classifies the bully comments. The comments tagged with different bullies are given in the fuzzy system for identification of strength of the bully. Table 8 shows the output for fuzzy rule sets for the considered test data.

TABLE 8. OUTPUT FOR FUZZY RULE SETS

Sentence	Output	Strength of Bully
Looking like a gorgeous bitch	Harassment	Medium
ugly black u	Racism	Medium
Too fat and ugly	Shaming	Low
Wearing glasses?? Ugly granny	Shaming	Low
Too much hot and curvy	Harassment	Medium

3. Accuracy and Performance Analysis

This section demonstrates the accuracy measure of the proposed classification system and comparative analysis of the performance relative to existing works using another model. The accuracy [15] is calculated using equation (1).

$$Accuracy = (Number\ of\ correctly\ classified\ comments/Number\ of\ comments\ in\ the\ dataset) * 100\% \quad (1)$$

Thus the error rate is calculated as (2).

$$Error\ Rate = (100 - AccuracyRate)\% \quad (2)$$

Table 9 represents percentage of accuracy for prediction of correct classes. The proposed Multinomial Naive Bayes classification model shows an accuracy rate of 88.89%. The considered data set is used for the implementation of the existing SVM model of classification for Facebook posts [11]. The SVM model gives an accuracy rate of 76.38%. Comparison shows that the proposed model yields a far better accuracy for prediction for the considered data set.

TABLE 9. COMPARISON OF ACCURACY AND ERROR RATE BETWEEN SVM MODEL AND PROPOSED CLASSIFICATION SYSTEM

Performance Factor	SVM Model (%)	Proposed Model (%)
Accuracy	76.38	88.89
Error Rate	23.62	11.11

The time is calculated for each run of both models and taken in consideration in fragments of data sets. The time complexity graph in Figure 4 shows that the proposed Multinomial Nave Bayes Classifier is having lower run time complexity than the SVM classification model. Run time complexity is calculated as the absolute difference between the time before the algorithm starts and time at which the algorithm finishes running. It is calculated in milliseconds.

4. Conclusion

This work is conducted for detection and classification of cyber bullying from comments in Facebook posts. Bullies are classified in three categories- Shaming, harassment and racism. A Multinomial Naive Bayes classifier is used for the classification of bully comments. Percentage for each bully type is also calculated. The accuracy is 88.76% for the classification model using considered data set. Authors find it satisfactory but more accuracy can be gained in further implementations. The strength of individual bully comment is identified by applying fuzzy rule sets designed for this work. The data set considered for the classification includes comments, age who commented and tags for individual comment. The identification and classification model

does not take into account the gender parameter. Gender specific classification has been done previously [5]. The authors have shown an analysis on gender specific patterns of cyber bulling in social networks. Again multinomial naive bayes classification works by calculating the conditional probability of frequency of words. Though multinomial naive bayes classification works best for text classification of multi class problems, the accuracy for prediction can still be improved through further research. Consequences of cyber bullying events can be mitigated if the level of negative aspects can be assumed and take proper measures against that.

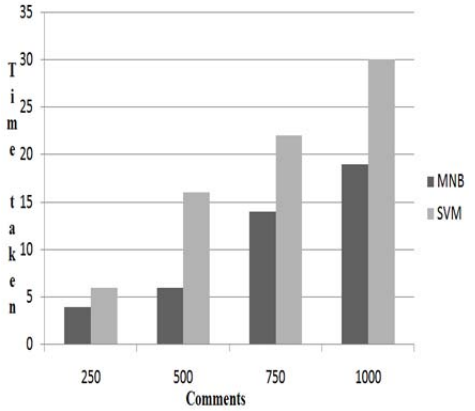

Fig. 4. Time Comparison between Proposed Multinomial Naïve Bayes (MNB) and Support Vector Machine (SVM) model.

References

[1] The daily star: https://www.thedailystar.net/bytes/49-bangladeshi-school-pupils-face-cyberbullying-287209.
[2] Kids health: https://kidshealth.org/en/parents/cyberbullying.html.
[3] J. B.Sri Nandhini, “Online social network bullying detection using intelligence techniques", Advanced Computing Technologies and Applications, Procedia Computer Science, vol. 1215, pp. 485{492, 2015.
[4] Hosseinmardi, Homa, A. Mattson, Sabrina, Ra_q, R. Ibn, Han, Richard, L. Qin, Mishra, and Shivakant, “Detection of cyberbullying incidents on the instagram social network." 2015.
[5] R. O. M. Dadvar, F. d. Jong and D. Trieschnigg, “Improved cyberbullying detection using gender information", Twelfth Dutch-Belgian Information Retrieval Workshop, pp. 23-25, 2012.
[6] Charles E. Notar , Sharon Padgett, Jessica Roden, “Cyberbullying: A Review of the Literature" Universal Journal of Educational Research 1(1): 1-9, 2013.
[7] I. McGhee, J. Bayzick, A. Kontostathis, L. Edwards, A. Mcbride, , and E. Jakubowski, “Learning to identify internet sexual predation," International Journal on Electronic Commerce 2011, vol. 15, pp. 103-122, 2011.
[8] H. Hosseinmardi, S. A. Mattson, R. IbnRa_q, R. Han, Q. Lv, and S. Mishra, “Detection of cyberbullying incidents on the instagram social net-work," 2015.
[9] K. Dinakar, R. Reichart, and H. Lieberman, “Modeling the detection of textual cyberbullying," 2011.
[10]K. Reynolds, A. Kontostathis, and L. Edwards, “Using machine learning to detect cyberbullying," Machine Learning and Applications and Workshops (ICMLA), vol. 2, pp. 241-244, 2011.
[11]K. D. Gorro, M. J. G. Sabellano, K. Gorro, C.Maderazo, and K. Capao, “Classification of cyberbullying in facebook using selenium and svm," 3rd International Conference on Computer and Communication Systems, pp. 183-186, 2018.
[12]V. Nandakumar, B. C. Kovoor, and S. M. U, “Cyberbulling revelation in twitter data using naive bayes classifier algorithm," International Journal of Advanced Research in Computer Science, vol. 2, pp. 511{513,

2018.
[13]Kaggle: https://www.kaggle.com/arnisha/facebook-comments-bullies.
[14]Scikit learn: https://www.scikit-learn.org/stable/modules/model evaluation.html.
[15]https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234.

Authors' Profiles

Arnisha Akhter received her B.Sc. and M.Sc. degree in Computer Science and Engineering from Jagannath University, Bangladesh, respectively in the year 2016 and 2018. She has been working as a Lecturer in the department of Computer Science and Engineering, Daffodil International University, Bangladesh, since 2017. She has published several good research articles in high ranked conference proceedings. Her research interest includes Artificial Intelligence, Data Mining, Wireless Ad Hoc and sensor Networks etc. She has been recognized as a IEEE young professional member since 2016.

Dr. Uzzal K. Acharjee obtained his Ph.D degree in Applied Physics, Electronics & Communication Engineering from University of Dhaka, Bangladesh in 2014. He received the B.Sc. and M.Sc. degree in Applied Physics, Electronics and Computer Science from the University of Dhaka, Bangladesh, in the year 1999 and 2000, respectively. He is serving as Professor in the Department of Computer Science and Engineering, Jagannath University, Bangladesh. His research interests include the area of Artificial Intelligence, Neural Networks, Deep Learning, Data Mining, Wireless Networks, Social Networks etc.

Md Masbaul A. Polash received the B.Sc. degree and M.Sc. degree in Computer Science and Engineering from University of Dhaka, Bangladesh. He received the Ph.D. degree in Computer Science at the Institute for Integrated and Intelligent Systems, Griffith University, Australia, in 2017. Since 2011, he has been a faculty member in the department of Computer Science and Engineering, Jagannath University, Bangladesh. He has published several research articles in top quality journals and conference proceddings. His research interest includes combinatorial optimization, operations research, artificial intelligence etc. Mr. Polash has been awarded a number of research awards from different organizations and conferences.

How to cite this paper: Arnisha Akhter, Uzzal K. Acharjee, Md Masbaul A. Polash," Cyber Bullying Detection and Classification using Multinomial Naïve Bayes and Fuzzy Logic", International Journal of Mathematical Sciences and Computing(IJMSC), Vol.5, No.4, pp.1-12, 2019. DOI: 10.5815/ijmsc.2019.04.01