Computer Engineering Dept.

Group No :- 48
Date :-30/06/2020

Cyberbullying With Fake Account Detection

Mentor-Richard Joseph
Professor , CMPN

Jayesh Samtani	Sagar Sidhwa	Somesh Tiwari	Riya Wadhwani
V.E.S.I.T	V.E.S.I.T	V.E.S.I.T	V.E.S.I.T
2017.jayesh.samtani@ves.ac.in	2017.sagar.sidhwa@ves.ac.in	2017.somesh.tiwari@ves.ac.in	2017.riya.wadhwani@ves.ac.in

1. Abstract :

Enhancement in the Technology trend of using Social Networking is increasing day by day as of now there are more than 50 Crores active users are using different different social media platforms for the interaction which had affected their life just like a coin has two face in a similar way misuse of these platforms is going which cause the tremendous growth of the cyber crime and bullying for e.g Bullying Someone by sending the harmful messages , spreading of the harassment messages by using the fake accounts, using the abusive words on the social media etc.In this new era insulting a person by physically or emotionally is done by cyberbullying and by using fake accounts,So as a preventive measure to ensure the above things should not happen there is a need of detecting the cyber bullying and the fake accounts. In our study to stop this we'll use different Machine Learning algorithms for detecting the Cybercrime and fake accounts so as to report these issues to the system immediately and to stop the crimes to increase in future and develop a secure online environment.

2. Introduction :

Social networking sites have connected us to different parts of the world However, people are finding illegal and unethical ways to use these communities. We see that people, especially teens and young adults, are finding new ways to bully one another over the Internet. Close to 25% of parents in a study conducted by Symantec reported that, to their knowledge, their child has been involved in a cyberbullying incident.

Other than cyberbullying, Spreading False information is increasingly at a rapid pace. The number of users in social media is increasing exponentially. Instagram has recently gained immense popularity among social media users.

The major source of the fake news are the fake accounts.Business organizations that invest a huge Sum of money on social media influencers must know whether the following gained by that account is organic or not. Hence there is a huge need for the detection of these fake accounts.

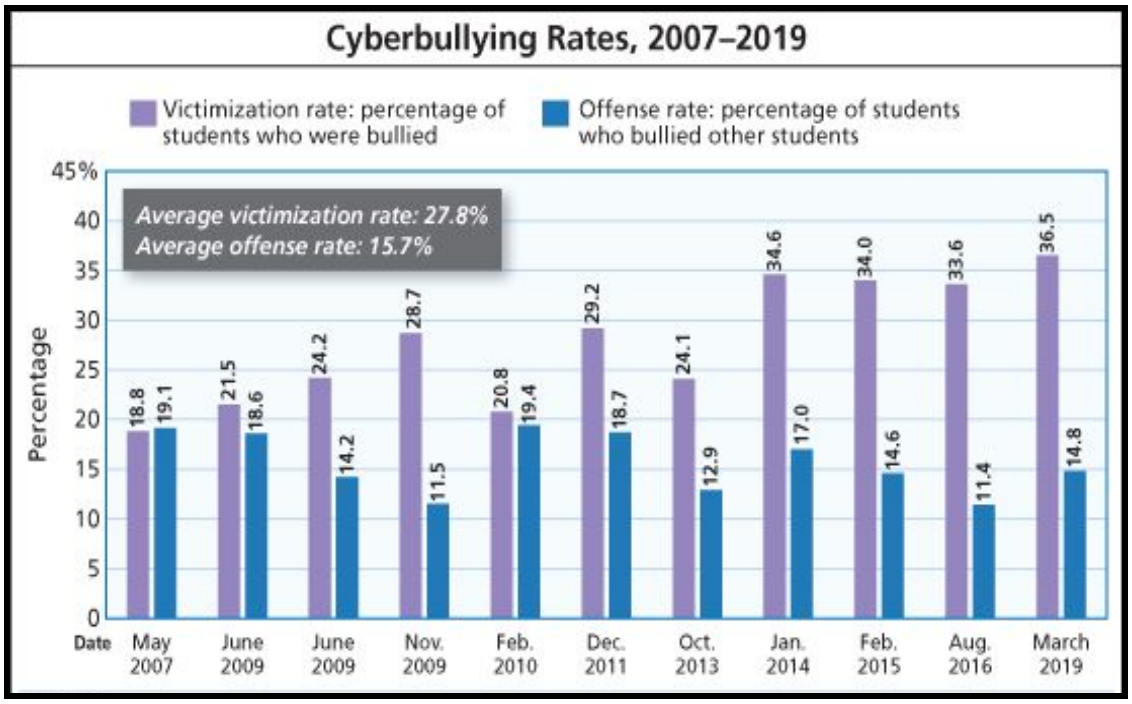

Fig 2.1-Graph showing the increase in the rate of CyberBullying in the recent years

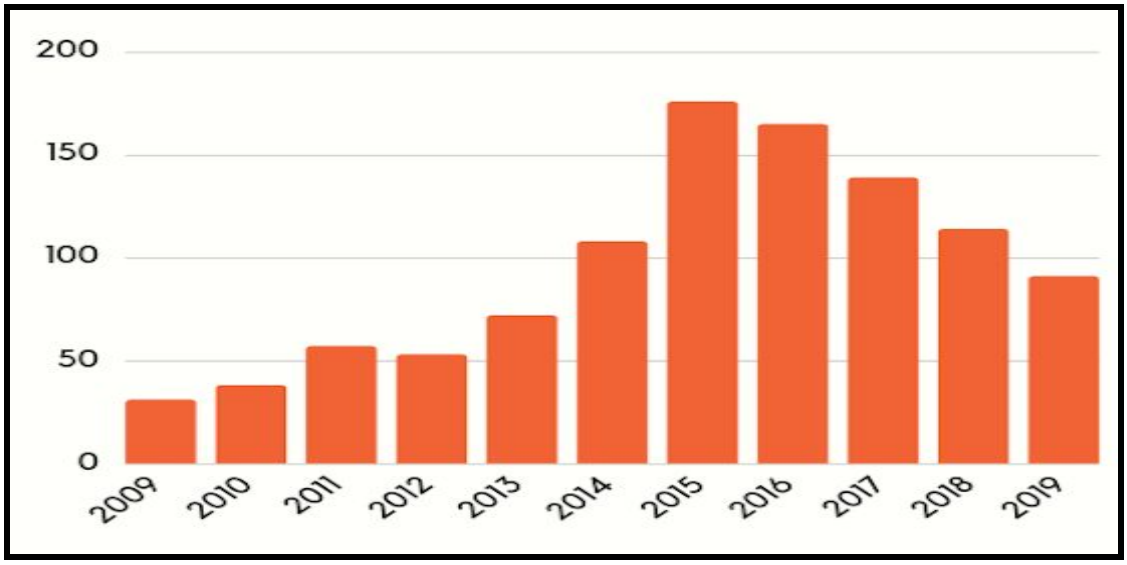

Fig 2.2-Graph showing the increase in the number of fake accounts

3. Problem Statement :

Nowadays, cybercrime is one of the common issues everyone is facing and it is impacting the people, in which some are long period of sadness, anger, irritability, loss of interest in activities, being restless, anxious and worried, even in some cases they go into depression and take steps to scarify their life.It is unfortunate that there are no special Anti-Cyber Bullying Laws in India yet.There are some common types of cyberbullying that is Flaming, Harassment, Denigration,Impersonation,Trickery. So to detect cyberbullying we have to make some software that will detect it and then report it to www.cybercrime.gov.in. Similarly, we will detect fake accounts.

4. Proposed Solution :

The proposed approach contains three main steps namely Preprocessing, features extraction and classification step.

In the preprocessing step we clean the data by removing the noise and unnecessary text.
The preprocessing step is done in the following: -
Tokenization: In this part we take the text as sentences or whole paragraphs and then output the entered text as separated words in a list. -
Lowering text: This takes the list of words that got out of the tokenization and then lower all the letters Like: 'THIS IS AWESOME' is going to be 'this is awesome'. -
Stop words and encoding cleaning: This is an essential part of the preprocessing where we clean the text from those stop words and encoding characters like \n or \t which do not provide meaningful information to the classifiers.

The second step of the proposed Model is the features extraction step. In this step the textual data is transformed into a suitable format applicable to feed into machine learning algorithms.
The last step in the proposed approach is the classification step where the extracted features are fed into a classification algorithm to train, and test the classifier and hence use it in the prediction phase. We will use classifiers, namely, SVM (Support Vector Machine), Naive Bayes,Random Forest,Decision Tree , Logistic Regression and Neural Network.
The neural network will contain three layers: Input, hidden, output layer. The output layer is a Boolean output.

Accuracy of different algorithms will be Compared to get the best possible result.
For the fake profile detection this paper proposes the detection process starts with the selection of the profile that needs to be tested. After selection of the profile the suitable attributes ie., features are selected on which the classification algorithm is being implemented ,the attributes extracted are passed to the trained classifier .

Different Classifier algorithms such as Gradient Booster, random forest Decision trees ,Support Vector Machine and Neural Networks such as RNN and CNN can be used.
The model generated by the learning algorithm should both fit the input data correctly and correctly predict the class labels of the learning algorithm is to build the model with good generality capability.

Data set of both fake and genuine profiles with various attributes like number of friends ,followers, status count. Dataset is divided into training and testing data. Classification algorithms are trained using training dataset and testing data set is used to determine the efficiency of the algorithm .From the dataset used 80% of both (real and fake) are used to prepare a training data set and 20% of both profiles are used to prepare a testing dataset.

5. Methodology / Block Diagram :

In this Project our aim is to detect Cyberbullying along with Fake account Detection with respect to the marginal number of attributes the proposed methodology consists of different steps.

First step is the preprocessing of the datasets and find the minimum set of attributes from two datasets i.e Cyberbullying and the second one is the fake account so in this step we will separate number of the attributes that will actually used to identify these bullying, contains abusive words etc and save them the dataset for further processing in the second dataset we will preprocess and extract the attributes e.g name, contact, email, etc and will be stored for further processing and to identify that weather the account is fake or not.

The Second step is to create different Machine Learning Models like Support Vector Machine Classifier, Random Forest algorithm,Naïve Bayes, Logistic Regression,K-mean clustering,ADT and BFT tree and Neural Networks etc and applying the following algorithms on the datasets of Cyberbullying and the fake account by splitting the datasets in to training and test approximately in the ratio of 80:20 and to find the algorithm which best suits or fit for our system to achieve the highest accuracy.

Third step is to test the messages that are extracted from the chats or the tweets or the blog which is posted which can cause bullying or use of abusive words and then sending the words to different Machine Learning models that are created for testing or checking to give the input to these models and running these models to test that which one of them gives the best solution and fits best in to our system.

In the Fourth step is to find whether the account that causes bullying to someone is Fake or Not so the attributes of the account is given as input to the models and the model that will give us the best accuracy is used as the best fit to our system and the results that we will get from the

third step and fourth step will be combined and will be send and reported to the system that if the cyberbullying is done by the person or not is yes then we can report this to the cyber branch and this will help them to take an action by them.

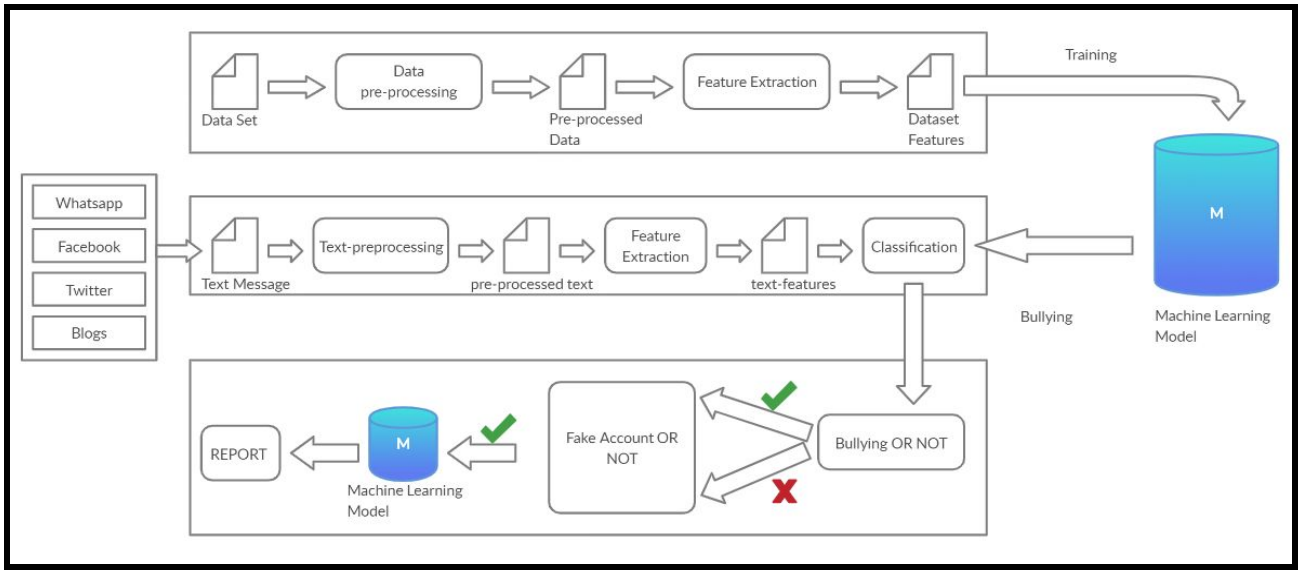

Fig 5-Block Diagram

6. Hardware , Software and tools Requirements :

- Intel Pentium Processor
- RAM>4GB
- Django
- Machine Learning Algorithms
- Anaconda

7. Proposed Evaluation Measures:

Effectiveness of the proposed method is accomplished by machine learning methods. Here cross-validation technique, and the basic metrics and evaluation of the classifier performance will be used.

Cross-validation is a technique in which we train our model using the subset of the data-set and then evaluate using the complementary subset of the data-set.

It is a technique used to evaluate predictive models. In this technique, the original samples are divided into two categories: training set for model training and test set for evaluation. The original sample is randomly divided into k subsamples with equal size. One of these subsamples is considered as evaluative data in order to test the model, and the rest of them, k-1 subsamples, are considered as training data. The cross-validation process is repeated k times for k subsamples, each time for one of them as evaluative data. The first advantage of this method is

that all samples are used for both training and validation process, and the second one is that each sample is used for validation just once.

A confusion matrix is a summary of prediction results on a classification problem.The number of correct and incorrect predictions are summarized with count values and broken down by each class. This is the key to the confusion matrix.The confusion matrix shows the ways in which your classification model is confused when it makes predictions.
It gives you insight not only into the errors being made by your classifier but more importantly the types of errors that are being made.It is this breakdown that overcomes the limitation of using classification accuracy alone.
The evaluation is based on a confusion matrix and associated metrics. The variables TP, FP, TN, and FN in the confusion matrix refer to the following:

true positive (TP): number of fake nodes that are identified as fake nodes,
false positive (FP): number of normal nodes that are identified as fake nodes,
true negative (TN): number of normal nodes that are identified as normal nodes,
false negative (FN): number of fake nodes that are identified as normal nodes.

To evaluate the classifier, accuracy and area under curve (AUC) are used. The AUC is performance metrics for binary classifiers; the closer this AUC is to one, the more favorable the final performance of the classification will be. By comparing the ROC curves with the AUC, it captures the extent to which the curve is up in the northwest corner. The metrics which are introduced below are used to calculate the ROC.

True negative rate (TNR) =TN/ (TN+FP).
False positive rate (FPR) =FP/ (FP+TN).
True positive rate (TPR) = TP/ (TP+FN).
False negative rate (FNR) =FN/ (FN+TP).
There is another measure which is used to evaluate the performance:
Accuracy= (TP+TN)/ (TP+FP+TN+FN).

8. Conclusion:

In this paper, we proposed an approach to detect cyberbullying using machine learning techniques. We will evaluated our model on two classifiers SVM and Neural Network and we used TFIDF and sentiment analysis algorithms for features extraction
By using machine learning algorithms to its full extent, we have eliminated the need for manual prediction of a fake account, which needs a lot of human resources and is also a time-consuming process.

9. References :

1) S. Gurajala, J. S. White, B. Hudson, and J. N. Matthews, “Fake Twitter accounts: Profile characteristics obtained using an activity-based pattern detection approach,” in Proceedings of the 2015 International Conference on Social Media & Society (SMSociety'15), Toronto, Ontario, Canada, 2015.View at: Publisher Site | Google Scholar.

2) N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: synthetic minority over-sampling technique,” Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002.View at: Google Scholar

3)I. Jolliffe, Principal Component Analysis, 2002.View at: MathSciNet.

4)S. Sperandei, “Understanding logistic regression analysis,” Biochemia Medica, vol. 24, no. 1, pp. 12–18, 2014.View at: Publisher Site | Google Scholar.

5)R. Kohavi, “A study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection,” in Proceedings of the in 14th international joint conference on Artificial intelligence, pp. 20–25, 1995.View at: Google Scholar.

6)https://machinelearningmastery.com/confusion-matrix-machine-learning/

7)http://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/

8)J. W. Patchin and S. Hinduja, “Bullies Move Beyond the Schoolyard; a Preliminary Look at Cyberbullying,” Youth Violence and Juvenile.

9)N. E. Willard, Cyberbullying and Cyberthreats: Responding to the
Challenge of Online Social Aggression, Threats, and Distress.

10) J. C. Platt, “Fast Training of Support Vector Machines using Sequential Minimal Optimization.