IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⩩ Impact Factor 7.39 ⩩ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115173

Detection Of Cyberbullying On Social Media Using Machine Learning

Athira S[1], Joel Saji[2], Abin Biju[3], Shon Alex Chacko[4]

Assistant Professor, Department of Computer Science and Engineering, JCT College Of Engineering and Technology[1]

Computer Science And Engineering, JCT College Of Engineering And Technology, Coimbatore, India[2,3,4]

Abstract: Cyberbullying is a major problem encountered on internet that affects teenagers and also adults. It has lead to mishappenings like suicide and depression. Regulation of content on Social media platorms has become a growing need. The following study uses data from two different forms of cyberbullying, hate speech tweets from Twittter and comments based on personal attacks from Wikipedia forums to build a model based on detection of Cyberbullying in text data using Natural Language Processing and Machine learning. Three methods for Feature extraction and four classifiers are studied to outline the best approach. For Tweet data the model provides accuracies above 90% and for Wikipedia data it gives accuracies above 80%.

INTRODUCTION

Machine Learning is a system of computer algorithms that can learn from example through self-improvement without being explicitly coded by a programmer. Machine learning is a part of artificial Intelligence which combines data with statistical tools to predict an output which can be used to make actionable insights.

The breakthrough comes with the idea that a machine can singularly learn from the data (i.e., example) to produce accurate results. Machine learning is closely related to data mining and Bayesian predictive modeling. The machine receives data as input and uses an algorithm to formulate answers.A typical machine learning tasks are to provide a recommendation. For those who have a Netflix account, all recommendations of movies or series are based on the user's historical data. Tech companies are using unsupervised learning to improve the user experience with personalizing recommendation.

Machine learning is also used for a variety of tasks like fraud detection, predictive maintenance, portfolio optimization, automatize task and so on.Machine learning involves computers discovering how they can perform tasks without being explicitly programmed to do so. It involves computers learning from data provided so that they carry out certain tasks. For simple tasks assigned to computers, it is possible to program algorithms telling the machine how to execute all steps required to solve the problem at hand; on the computer's part, no learning is needed. For more advanced tasks, it can be challenging for a human to manually create the needed algorithms. In practice, it can turn out to be more effective to help the machine develop its own algorithm, rather than having human programmers specify every needed step. The discipline of machine learning employs various approaches to teach computers to accomplish tasks where no fully satisfactory algorithm is available. In cases where vast numbers of potential answers exist, one approach is to label some of the correct answers as valid. This can then be used as training data for the computer to improve the algorithm(s) it uses to determine correct answers. For example, to train a system for the task of digital character recognition, the MNIST dataset of handwritten digits has often been used.

PROBLEM DEFINITION

The current system working on cyberbullying on social media prediction works on a small dataset. The aim of our system is to work on a larger dataset to increase the efficiency of the overall system. The number of comments also affects the performance of the system, thus our aim is to detect the cyberbullying and to increase the efficiency of the prediction.

LITERATURE REVIEW

Following is some of the search which has been reviewed for the proposed system.

1) Towards the detection of cyberbullying based on social network mining techniques

AUTHORS: I. H. Ting, W. S. Liou, D. Liberona, S. L. Wang, and G. M. T. Bermudez

In recent years, users are widely intend to express and share their opinions over the Internet. However, due to the characters of social media, it appears negative use of social media. Cyberbullying is one of the abuse behavior in the Internet as well as a very serious social problem. Under this background and motivation, it can help to prevent the happen of cyberbullying if we can develop relevant techniques to discover cyberbullying in social media. Thus, in this paper we propose an approach based on social networks analysis and data mining for cyberbullying detection. In the approach, there are three main techniques for cyberbullying discovery will be studied, including keyword matching technique, opinion mining and social network analysis. In addition to the approach, we will also discuss the experimental design for the evaluation of the performance.

2) Supervised machine learning for the detection of troll profiles in twitter social network: Application to a real case of cyberbullying

AUTHORS: P. Galán-García, J. G. de la Puerta, C. L. Gómez, I. Santos, and
P. G. Bringas

The use of new technologies along with the popularity of social networks has given the power of anonymity to the users. The ability to create an alter-ego with no relation to the actual user, creates a situation in which no one can certify the match between a profile and a real person. This problem generates situations, repeated daily, in which users with fake accounts, or at least not related to their real identity, publish news, reviews or multimedia material trying to discredit or attack other people who may or may not be aware of the attack. These acts can have great impact on the affected victims' environment generating situations in which virtual attacks escalate into fatal consequences in real life. In this paper, we present a methodology to detect and associate fake profiles on Twitter social network which are employed for defamatory activities to a real profile within the same network by analysing the content of comments generated by both profiles. Accompanying this approach we also present a successful real life use case in which this methodology was applied to detect and stop a cyberbullying situation in a real elementary school.

PROPOSED SYSTEM

SVM is basically used to plot a hyperplane that creates a boundry between data points in number of features (N)-dimensional space. To optimize the margin value hinge function is one of best loss function for this. Linear SVM is used in the following case which is optimum for linearly seperable data. In case of 0 misclassification, i.e. the class of data point is accurately predicted by our model, we only have to change the gradient from the regularisation arguments. A random forest consists of many individual decision trees which individually predict a class forgiven query points and the class with maximum votes is the final result. Decision Tree is a building block for random forest which provides a predicition by decision rules learned from feature vectors. An ensemble of these uncorrelated trees provide a more accurate decision for classification or regression.

ALGORITHMS

Support Vector Machine

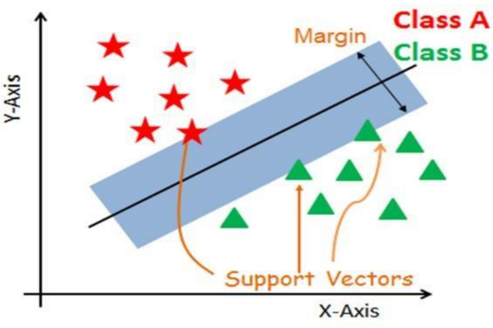
SVM is basically used to plot a hyperplane that creates a boundry between data points in number of features (N)-dimensional space.

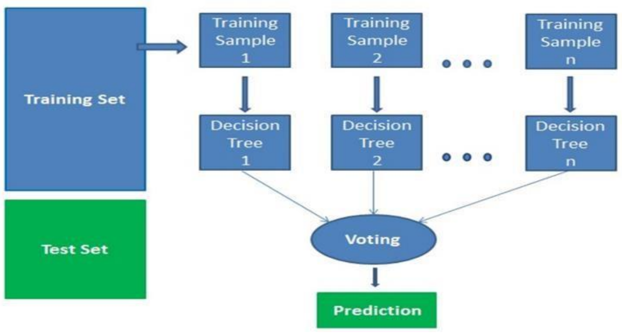
To optimize the margin value hinge function is one of best loss function for this. Linear SVM is used in the following case which is optimum for linearly seperable data.

Random Forest Classifier

A random forest consists of many individual decision trees which individually predict a class forgiven query points and the class with maximum votes is the final result.

Decision Tree is a building block for random forest which provides a predicition by decision rules learned from feature vectors.An ensemble of these uncorrelated trees provide a more accurate decision for classification or regression.





EXPECTED RESULT

The goal of our project is to predict whether the comment is offensive or non offensive and it is personal or non personal attack. We mainly looking forward for the celebrities who cannot look all the comments in there post by our invention we can check many comments at the same time so that we provide more time and energy. We implement two algorithms to work efficient and more accurate than the existing system. Our proposed system algorithms are 1. Support Vector

DOI: 10.17148/IJARCCE.2022.115173

Machine 2. Random Forest Classifier these two algorithms are more efficient and more accurate than the other systems. **Linear Kernel** A linear kernel can be used as normal dot product any two given observations. The product between two vectors is the sum of the multiplication of each pair of input values.

$K(x, xi) = sum(x * xi)$

Polynomial Kernel A polynomial kernel is a more generalized form of the linear kernel. The polynomial kernel can distinguish curved or nonlinear input space.

$K(x,xi) = 1 + sum(x * xi)^d$

Where d is the degree of the polynomial. d=1 is similar to the linear transformation. The degree needs to be manually specified in the learning algorithm.

Radial Basis Function Kernel The Radial basis function kernel is a popular kernel function commonly used in support vector machine classification. RBF can map an input space in infinite dimensional space.

$K(x,xi) = exp(-gamma * sum((x – xi^2))$

CONCLUSION

Cyber bullying across internet is dangerous and leads to mishappenings like suicides, depression etc and therefore there is a need to control its spread. Therefore cyber bullying detection is vital on social media platforms. With avaibility of more data and better classified user information for various other forms of cyber attacks Cyberbullying detection can be used on social media websites to ban users trying to take part in such activity In this paper we proposed an architecture for detection of cyber bullying to combat the situation. We discussed the architecture for two types of data: Hate speech Data on Twitter and Personal attacks on Wikipedia.

FUTURE SCOPE

The scope of the project is reduce the cyberbulliying on the social media using SVM and Random forest classifier algorithms and it also provide more efficient than the existing algorithms

REFERENCES

[1] I. H. Ting, W. S. Liou, D. Liberona, S. L. Wang, and G. M. T. Bermudez, "Towards the detection of cyberbullying based on social network mining techniques," in Proceedings of 4th International Conference on Behavioral, Economic, and Socio Cultural Computing, BESC 2017, 2017, vol. 2018-January, doi: 10.1109/BESC.2017.8256403.

[2] P. Galán-García, J. G. de la Puerta, C. L. Gómez, I. Santos, and P. G. Bringas, "Supervised machine learning for the detection of troll profiles in twitter social network: Application to a real case of cyberbullying," 2014, doi: 10.1007/978-3319-01854-6_43.

[3] A. Mangaonkar, A. Hayrapetian, and R. Raje, "Collaborative detection of cyberbullying behavior in Twitter data," 2015, doi: 10.1109/EIT.2015.7293405.

[4] R. Zhao, A. Zhou, and K. Mao, "Automatic detection of cyberbullying on social networks based on bullying features," 2016, doi:
10.1145/2833312.2849567.