IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

CyberBully Detection Using Neural Network

Mohammed Tabrez[1], Atluri Prateek[2], Sanjay Raj Sharma[3],
Dr Sreedhar Bhukya[4], Dr. T. Vijaya Saradhi[5], B Vasundhara Devi[6]
Students, Department of Computer Science and Engineering
Professor, Department of Computer Science and Engineering[4,5]
Associate Professor, Department of Computer Science and Engineering[6]
Sreenidhi Institute of Science and Technology Hyderabad, Telangana, India

Abstract: *The advancements in technology, as well as the digitization of relationships, had a significant impact on the centennials' decision to maintain a social media account. Despite the entertainment provided by social media, cyberbullying has been identified as a real issue all over the world, with many centennials becoming victims. However, a few studies have been reported in detecting cyberbullying attempts on social media. As a result, a solution that employs appropriate data science techniques to detect cyberbullying attempts on social media would be ideal. The suspicious tweets dataset from Kaggle was used in this study to build three supervised learning predictive models, namely Naive Bayes, which were tuned using Random Grid Search and Keras tuner to indicate a suitable solution.*

Keywords: Cyberbullying

I. INTRODUCTION

Cyberbullying is the deliberate and persistent mistreatment or pestering of an individual using methods such as social media platforms, this behavior is intended to intimidate and degrade the target person according to mahlamgu owolawi in 2018 some instances of this behavior include disseminating fake information or humiliating photographs of the victim sending direct messages with abusive language or pretending to be the victim and sending unsolicited messages on their behalf these are but a few instances of the types of cyberbullying that may occur on social media according to a bark team study from 2017 internet abuse has caused a spike in cyberbullying which has resulted in some student suicides.

II. OBJECTIVES

The goal of this project is to perform proper data pre-processing into an appropriate format for data analytics processing by developing an appropriate text classification model to categorise cyberbullying intent and to assess the effectiveness of the predictive model using appropriate evaluation measures.

III. REVIEW OF RELATED LITERATURE

In order to comprehend earlier research, a number of journal papers about the detection of cyberbullying using data science are analysed. Additionally, the previously employed models are investigated and assessed as follows.Since the svm and nave bayes results in table 1 produce unsatisfactory results, this research improved the evaluation results for both svm and nave bayes machine learning algorithms.

Models	Research	Accuracy (%)	Precision (%)	Recall (%)
SVM	(Dalvi, Chavan, & Halbe, 2020)	52.7	71.0	71.0
Naïve Bayes		52.7	71.0	71.0
SVM	(Al-Ajlan & Ykhlef, 2018)	81.3	73.0	70.0
CNN		95.0	93.0	73.0
1D-CNN	(Ghosh, Chaki, & Kudeshia, 2021)	96.3	96.5	96.5
LSTM		94.1	94.8	94.3
BiLSTM		97.4	97.0	97.7

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

The table above displays some of the comparable past research on cyberbully detection using the data science approach. additionally In this study, LSTM is used to compare the outcomes of conventional machine learning versus deep learning.

IV. METHODS AND IMPLEMENTATION

4.1 Proposed Method

In addition to reducing incidents that may occur as a result of cyberbullying action, this project's goal is to identify cyberbullying intent within tweets from Twitter in order to assist parents, teenagers, and authorities with cyberbullying issues. The end result will not just be a predictive model, but will also provide output to decision-makers in the form of a bag of words and word cloud.

4.2 Methodology

Data Mining initiatives frequently follow the CRISP-DM (cross-industry standard process for data mining) methodology for organising planning and carrying out data mining initiatives it offers a planned and methodical procedure which consists of 6 phases.The first stage of the project is called business understanding where the business challenge is defined and the project goals for the cyberbullying tweet detection are determined, for instance the objective might be to immediately identify tweets that involve cyberbullying and alert the relevant authorities.The next phase is Data exploration and collection arewhich is called Data Understanding in order to do this it may be necessary to gather a sample of tweets from social media sites comprehend how the tweets are organised and determine the different types of cyberbullying that are frequently seen in the data. In the next phase the data preparation with removing or manipulating any noise, duplication, incomplete data by using pre-processing techniques as well as the creation of new variables.The modelling phase is where the modelling techniques are chosen based on the objective of the research which is to predict cyberbullying. Thus, support vector machine and Naïve Bayes machine learning algorithms are used in this project ,Lastly, the evaluation of the model is performed using the confusion matrix which will assess the accuracy, precision, recall, and F1-score of the model to perform critical evaluations of the model. The accuracy, precision, F1 score, and recall is determined based on the formula given below.

$$Accuracy = \frac{(True\ Positive + True\ Negative)}{(True\ Positive + True\ Negative + False\ Positive + False\ Negative)}$$

$$Precision = \frac{True\ Positive}{(True\ Positive + False\ Positive)}$$

$$Recall = \frac{True\ Positive}{(True\ Positive + True\ Negative)}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$= \frac{2(True\ Positive)}{2(True\ Positive) + False\ Positive + False\ Negative}$$

Eq. 1. Accuracy, Precision, Recall, and F1-score equation

afterwards deploying the models into use in a real-world so they can support business choices to do this it might be necessary to create a web application or api that can analyse tweets and categorise them as cyberbullying or non-cyberbullying.

A. Overview of the Process Model

The dataset acquired will be pre-processed from data that are dirty which has noise, duplication, and missing records by using data preprocessing techniques such as data cleaning. Variable selection will not be performed since the data consist of only two columns that contain the tweets and a Boolean variable indicating the cyberbullying intent. Within the pre-processing stage, data cleaning and data transformation are performed. After pre-processing has finished, the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

dataset will be divided into training and testing data sets with an 80:20 ratio. Afterwards,the model will be trained and tested with the preprocessed dataset through 3 algorithms which are support vector machine, Naïve Bayes, and longterm short memory. Once the model training has been completed, the model is evaluated to select the best model based on the evaluation results. Thus, after the best model has been identified, the best model will be deployed onto a web application

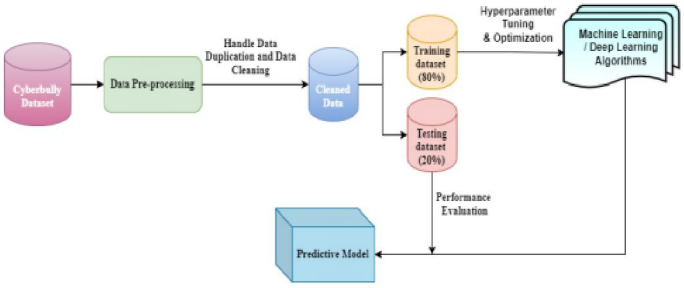


Figure 1. Process Flow Diagram

B. Data Pre-processing

Every data analytics project starts with EDA where the dataset is often examined to get insight into traits like data type and uncover trends to further investigate the kaggle dataset utilised for the cyberbully detection model contains suspicious tweets with ties to terrorism threats and cyberbullying the dataset which includes more than 60000 tweets was compiled from twitter the tweets message can be found enclosed in the initial column of the dataset which has a label indicating whether or not it is suspicious can be found in the second column.

C. Data Exploration (Pre-Cleaning)

Figure 2 depicts the variables that populate the label column the proportion of suspicious tweets vastly outnumbers the number of non-suspicious tweets the suspect tweets account for over 90 of the dataset with the remaining 10 occupied by the suspicious tweets the label column contains 50k records of non-suspicious records and 6k record of questionable records the data is not balanced in that given non-suspect data populates approximately 90 dataset while suspicious records populate the remainder so an oversampling approach known as synthetic minority over-sampling technique smote was used to perform balancing.

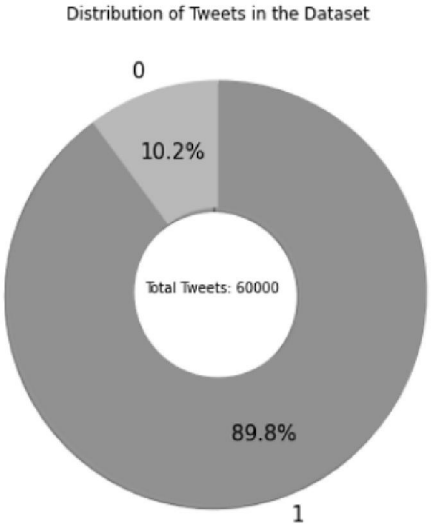


Figure 2. Pie Chart for label column

C. Data Cleaning

The first pre-processing step is to remove the duplication by using the "drop_duplicates" function. Once the duplicated data is removed, there is a total of 53,574 non-suspicious data and 6133 suspicious data. Thus, it has been found that all

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

the duplicated data are from non-suspicious data. Next, is to convert all the text into the lower case to make all the text data standardized and easier to be processed. Then, using the in-built function from the emoji library called "demojize" to convert all the possible emoji into text. Further data cleaning is performed where the message is tokenized into tokens by the NLTK library using the "word tokenize" function. Hence, each message within the data frame is converted into tokens. Secondly, tokens that contain punctuations are removed by using the string library to compare the tokens that match the punctuation library set. Then, removing the stop words using the NLTK stop words library is used to compare the tokens with the stop words set library. Lastly, lemmatization of each of the tokens using the wordnet and POS tagging is also included within the process to produce a more accurate result.

D. Data Transformation

After the data cleaning has been completed data is divided into training and test datasets with a ratio of 8020 is executed for training purposes next a count vectorizer is done before training the model it is a basic tool for tokenizing a set of text documents and creating a vocabulary of known words as well as encoding new documents with that vocabulary the result is an encoded vector with the full vocabularies length and an integer count of how many times each word appears in the document a sparse matrix will be generated which contains a lot of zeros by using the transform function brownlee 2020 after the count vectorizer has been executed the vectorizer is saved as every time it is executed again as it would be not unique and resulting in a mismatch in the model later.

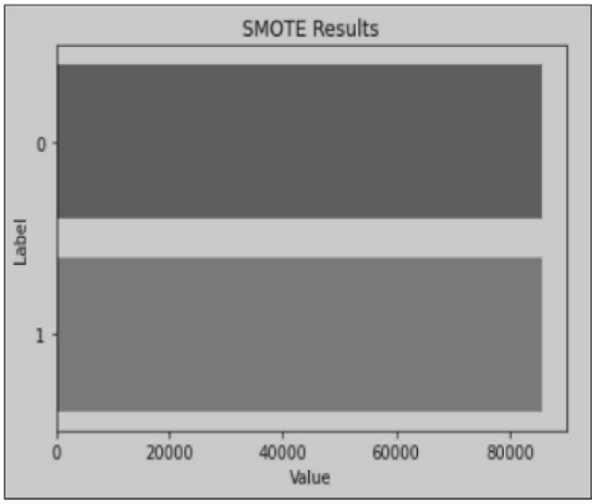


Figure 3. SMOTE Result

smote employs knn algorithm for producing data points by selecting closest adjacent point from minor class this method repeats until minor and major classes are equalized smote mechanism is employed to vectorized training dataset to prepare it for ml models this balances suspicious data 0 with non-suspicious data 1 resulting in a balanced data division which eliminates overfitting figure 3 depicts how smote contributes to this equilibrium.

Data Exploration (Post-Cleaning): An illustration of words is called a word cloud that gives the basic idea of the dataset in this case the word cloud represents the tweet messages the larger word size means that the word is more significant than the other words as seen in figure 4 most of the words are related to now work today and many more.

Figure 4. Word Cloud

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

V. UML DIAGRAMS

Class Diagram: Class diagrams are useful for representing the static structure of a system, and for understanding the relationships between the classes and their attributes and operations. They are also useful for identifying the interfaces and dependencies between the classes, and for specifying the attributes and operations of the classes in more detail.

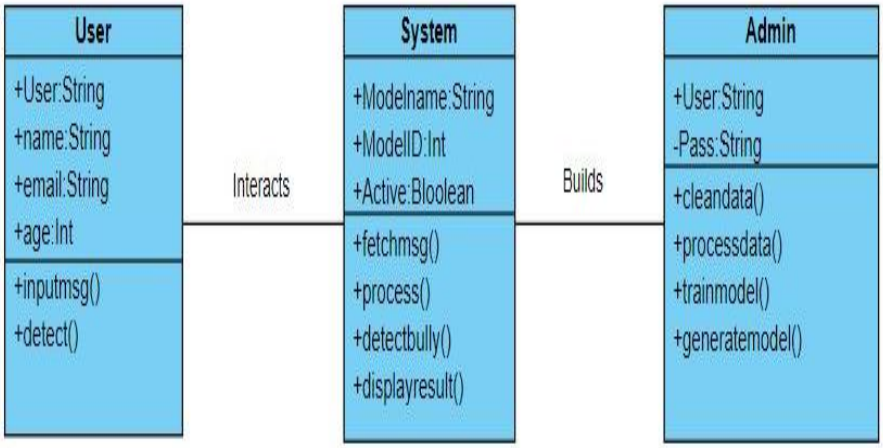


Figure 5 Class Diagram

UseCase Diagram: In this example, the actor interacts with the system to initiate the process of identifying cyberbullies by sending a tweet. The administrator then runs a number of operations on the data, such as preprocessing, training a machine learning model, testing the model, analysing the results, and possibly improving the model within the context of a more general detection use case.

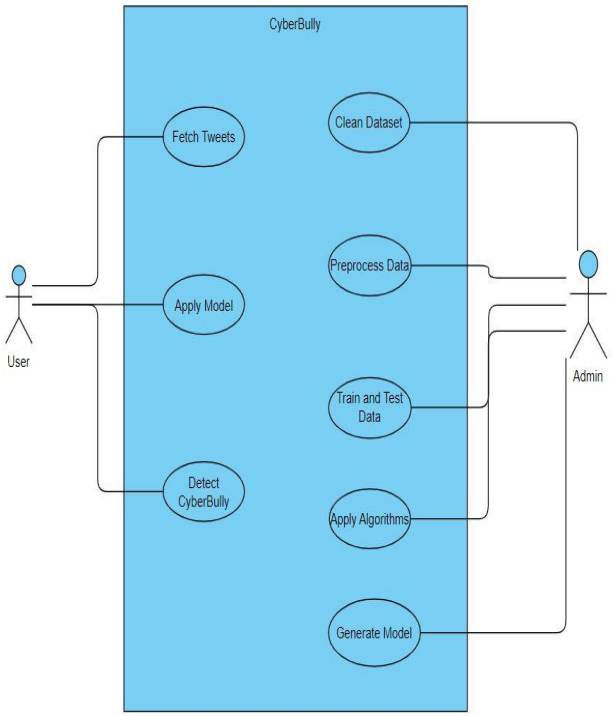
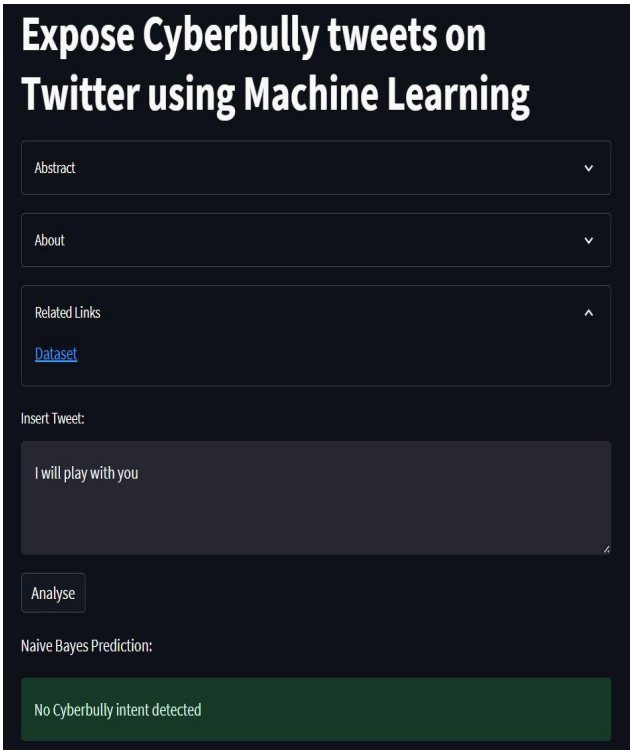


Figure 6. UseCase Diagram

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

VI. OUTPUT

Figure 7. Positive Cyber Intent



VII. CONCLUSION

Giving a solution using technological advancements such as a predictive model is beneficial for the social media domain. There are four models produced good results with Naïve Bayes being chosen as the best model with 88% accuracy. Proper data preprocessing and optimisation methods are used to build a more effective model with higher accuracy than in the past. Several improvements can be made where the layers in the LSTM algorithm can be improved as many options are available and more unique choices and combinations for the traditional algorithm like the random forest could be applied as well

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

REFERENCES

[1]. Brownlee, J. (2020, 28 June). How to Encode Text Data for Machine Learning with scikit-learn. Retrieved from Machine Learning Mastery:https://machinelearningmastery.com/prepare-text-data-machine-learning-scikit-learn/ .

[2]. Gandhi, R. (2018, May 6). Naive Bayes Classifier.Retrieved June 4, 2021,fromTowards DataScience:https://towardsdatascience.com/naive -bayes-classifier-81d512f50a7cR. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[3]. Ghosh, S., Chaki, A., & Kudeshia, A. (2021).Cyberbully Detection Using 1D-CNN and LSTM. Proceedings of International Conference on Communication, Circuits, and Systems. Bhubaneswar: KIIT University. doi:10.1007/978-981-33-4866-0_37

[4]. Smart Vision. (2021). What is the CRISP-DM methodology? Retrieved May 30, 2021, from Smart Vision: https://www.sveurope.com/crisp-dm-methodology/#one.