Prevention of Cyber Bullying using Machine Learning Techniques

GMSN Gunawardana# and WMKS Ilmini

Department of Computer Science, General Sir John Kotelawala Defence University, Ratmalana, Sri Lanka

#35-cs-0017@kdu.ac.lk

Abstract- Cyber bullying has rapidly increased in the past few years with the growth of social media usage and the COVID-19 pandemic. This study uses a dataset of 65000 tweets , splitting them into training and testing sets. Data pre-processing was done using feature engineering methods such as vectorizing, and Bag of Words to prepare data to test machine learning models or classifiers to build a model. Five different classifiers were tested with dataset and Naïve Bayes Model and linear support vector classification model provided the best accuracy and prediction times in sequence. The Sentiment Analysis System was built using Naïve Bayes Model and it is deployed to the web interface using Flask to get user input and predict sentiment in the three key aspects of negative, positive and neutral. System tested with user inputs and gained accurate sentiment Scores (comment: "listen to my most beautiful friend singing with her beautiful voice" Scores: Compound- 0.97 Neutral – 0.166 Positive – 0.834 Negative – 0.0) with three key aspects. The aim of this research work is to utilize man-made consciousness at a specific level to pre-empt exploitation by recognizing the riskiest clients and accounts.

Keywords: cyber bullying, social networks, machine learning, sentiment analysis

I. INTRODUCTION

Cyberbullying is a type of tormenting that happens by means of web associated gadgets like cell phones, PCs, or tablets. Cyberbullying is the utilization of innovation to scare, annoy, undermine, torture, or embarrass an objective. As of this point many people face cyberbullying on daily basis and we look in to where it occurs, when the students were asked to indicate on which social media platforms they had experienced cyberbullying, the results showed Twitter 9%, YouTube 10%, WhatsApp 12%, Snap Chat 31%, Face Book 37%, Instagram 42%. And the study also revealed some surprising statistics about the number of people that are perpetrators of cyberbullying ("51 Critical Cyberbullying Statistics in 2020," n.d.). 69 percent of people report having done something abusive towards others online. 15 percent of people admit to having cyberbullies someone else online. These statistics are troubling as it indicates a general misunderstanding of definition of cyberbullying.

Over half of students who identify as being LGBTQ have experienced cyberbullying at some point (Affairs (ASPA), 2019). Girls are more likely to be a victim of cyberbullying than boys. Overall, around 36% of girls have reported being cyberbullied, as compared to 26% of boys ("Online Harassment 2017 | Pew Research Center," 2018). 83 percent of those who have been cyberbullied have also been bullied in person, and 69% of those who admitted to bullying online have also admitted to in-person bullying ("Nationwide teen bullying and cyberbullying study reveals significant issues impacting youth -- ScienceDaily," 2018). When we talk about the impacts of cyberbullying 64% of people who have been cyberbullied say it affects their ability to learn and feel safe at school ("Online Harassment 2017 | Pew Research Center," n.d.). Bullied students are twice as likely as other students to experience problems such as headaches and stomach aches (Gini and Pozzoli, 2013).

In this work we collect the group of words that has been used to bully and harass people to recognize such content and analyze those using linguistic analytics and sentiment analytics because with friends we aren't always polite. Also understanding the behaviors of the followers or people who are daily active on these people's account in case of twitter and such platforms analyze their behavior to prevent them from accessing those accounts and hashtags using machine learning and AI, if they are involved in bullying and harassing the victims in any kind of way. We can reduce the toxicity of these platforms and make many feel safe in their own space by doing that. The aim of this research work is to utilize man- made consciousness at a specific level to pre-empt exploitation by recognizing (and blocking, forbidding, or isolating) the most risky clients and accounts. We are going to achieve this by Understand the key characteristics of cyberbullying and people involved in cyberbullying, Identify risk factors and outcomes of cyberbullying, Clarify what measurement instruments will lead to consistent, Identify existing research gaps on cyberbullying and its prevention, Proposing suitable solution – Machine Learning System. In the second section of this paper we review previous works to prevent cyber bullying using machine learning and also other experiments that has been done to acknowledge the issue and focus on the future directions that this project can be taken and how to make the process efficient and increase accuracy of systems using machine language and sentiment analysis in a more depth level.

II. LITERATURE REVIEW

As for the existing networking systems or social media portals they sure do take necessary actions to reduce cyber bullying as it is increasing day by day and affecting younger generation a lot and most of the celebrities go through this a lot as their lives get high light attention all the time which general audience try to get involved without knowing many facts or anything about them personally. In America there are organizations that are constantly on watch of such cyberbullying to prevent it in real time in some platforms("Cyberbullying Organizations," 2018), but there is always a limit how much they can prevent at a time as Internet is a mass capacity. The facts and statistics of Cyberbullying for 2018-2020 shows that more parents than ever report that their children are getting bullied at school or online. Comparitech conducted a survey of over 1000 parents of children over the age of 5 and they found 47.7% of parents with children ages 6-10 reported their children were bullied, 56.4% of parents with children ages 11-13 reported their children were bullied, 59.9% of parents with children ages 14- 18 reported their children were bullied, 54.3% of parents with children ages 19 and older reported their children were bullied ("Cyberbullying Statistics and Facts for 2020," 2020).

Several attempts were taken to find accurate platforms to gather the information to conduct the process of prevention which has been successful to a great level. And there are different technologies used some in different and some in the same experiments to recognise what are the most accurate and efficient in preventing cyberbullying. In the previous reviews they have used gaming platforms to gather data automatically In their chat boxes on the issue which they have been successful to the expected level but could not achieve a solution. Multilingual systems are also in the discussion and a group has done a system review on Arabic language system to detect and prevent cyberbullying. From the review they have gathered all the data for them to come to the conclusion that it is possible preventing cyber bullying in different languages. As the issue getting severe day by day a group in IT field has come up with a causal theory and more effective empirical methods to investigate and mitigate this phenomenon, they leverage the control balance theory and their model examines the causes of cyberbullying from several novel angles. One major drawback to this method is that we did not observe people committing cyberbullying and there has been experiment in Automatic cyberbullying detection which is a task of growing interest, particularly in the Natural Language Processing and Machine Learning communities. In this work, they conduct an in-depth analysis of 22 studies on automatic cyberbullying detection, complemented by an experiment to validate current practices through the analysis of two datasets. They also complemented this approach with an extensive experiment to assess current practices, by using feature engineering. And the experiments to define what cyberbullying actually is also been conducted and their analysis proves its definition vary from one individual to another which sentiment analysis comes in handy in cyberbullying prevention.

Table 2-1. Feature Comparison and novelty of Machine Learning Models for Cyberbullying Detection and Sentiment Analysis

Paper (Research) Author	Technologies	Accuracy
A Multilingual System for Cyberbullying Detection: Arabic Content Detection using Machine Learning	Machine Learning (ML) Natural Language Processing (NLP) WEKA toolkit	58.7% - 61.2%
Systematic review on automatic cyberbullying detection.	Natural Language Processing Machine Learning.	67%
Expertand Machines against Bullies:A Hybrid Approach to Detect Cyberbullies.	Expert Systems Supervised Machine Learning.	68% - 72%
Cyberbullying Detection Using Sentiment Analysis In Social Media	Sentiment Analysis Naïve Bayes, Support Vector Machine and Neural Network	**SVM -** 89.39% **Naive Bayes** - 73.0328% **Convolutional Neural Network** - 48.6404%
Sentiment Analysis of Twitter Data	Naive Bayes, Support vector machine (SVM), and Bagging	**cross validation** Naïve Bayes - 56% SVM 41% Bagging 43%
Sentiment Classification using Machine Learning Techniques	Naïve Bayes Support Vector Machine (SVM)	**Naive Bayes** - 65.57% **SVM -** 45.71%
A Comparative Analysiof Machine Learning Classifiers for	Multinomial Naïve Bayes, Bernoulli Naïve Bayes and SVM	**BNB 70.75 % MNB 75.77 % SVM 74.09 %**
Twitter Sentiment	Unigrams and bigrams	
Sentiment Analysis on Product Reviews Using Machine Learning Techniques	Naïve Bayes Support Vector machine	**Naïve Bayes classifier** - 98.17% **SVM -** 93.54%
Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination	WEKA Tool, Sentiment analysis Lexico Based Approach, Machine Learning Approach, Naïve Bayes and SVM	**Decision Tree algorithm** - 62% **SVM - 66% Naive Bayes - 64%**
Machine Learning and Semantic Analysis of In-game Chat for Cyber Bullying, an automatic data collection	SQL Database Queries, AI-based sentiment Text analysis services.	55.7% - 59.6%
system.		
Prevent cyberbullying Using the Design	Better causal theory and more effective empirical methods Control balance theory	54.3% - 57.5%

Cyberbullying prevention attempts has been taking from a very long time but addressing severe and most danger issues has started in the near past. The researches and system reviews proves that the probability of cyberbullying prevention in getting increased as time pass by with technologies such as machine learning, Natural Language Processing, Sentiment Analysis, Linguistic Analysis and some more which makes internet much safer and personal space for the younger or adult generation to use without getting affected by the toxicity of individuals or groups.

III. METHODOLOGY

A. Approach

1) Modern Approach

Ten years prior, Computer Scientists would contact Data Centers and inquire as to whether they could give those huge number of instances of cyberbullying content – which they needed to search through to search for patterns and examples in how damage was dispensed. These days, they don't request information since it's generally accessible for them to web scratch.

From publicly available social media posts in the quantity they need (e.g., there are approximately 500 million tweets on Twitter every day). Alongside the diminished expense of modest equipment like (nearly) limitless extra room on hard drives and figuring multiprocessors that can crunch and mine information dangerously fast, this has permitted the field to make some astonishing mechanical forward leaps to decrease online abuse.

2) Why use Machine Learning

We take posts and use artificial intelligence within a machine learning framework – specifically deep learning to make determinations about them. She might first write multiple algorithms to do specific tasks. Together and collectively, those would form a neural network of layers, each with its own automated job to do.

Automatically, crafted by every one of these layers across the posts in the screen capture above – and a huge number of others, oppressive and not harmful – would be gathered, suitably gauged, and all in all used to acquire man-made reasoning in understanding what presents are undoubtedly on be poisonous. At that point, a calculation can perform sentiment analysis to make a determination of whether the next post is or is not toxic (sentiment polarity), what's more, therefore whether it ought to be hailed, hindered, or erased by a human mediator (whose decision-making is simplified through this system). This can then happen on every new post created by a user, automatically and on-the-fly("How Machine Learning Can Help Us Combat Online Abuse," 2017).

There are more layers to consider and evaluate (frequency of third-party reports on the post, use of emoticons, and how old the posting account is. I also know my example is not perfect, but hopefully you get the gist. As knowledge and technology in this area continues to develop, we will be increasingly able to identify what is abusive versus what is not.

negative, or impartial. It is a combination of natural language processing, text analysis and computational linguistics. In this process a sentence is considered positive if it has positive keywords and is considered negative if it has negative keyword. The comparison among the number of each type of contents decides the positivity and negativity of the whole content("(Tutorial) Simplifying Sentiment Analysis in Python- DataCamp," n.d.). This study tends to provide an algorithm that may help in analysis of words that may lead to crime detection especially in social sites. For our research, we are using machine learning sentiment analysis technique. The algorithm that we're using are Naive Bayes. For Naive Bayes and an initial training data set is required which need to be labelled with positive, negative and neutral sentiment accordingly. For that, we are using the pre-labeled data set from kaggle and e data we gathered and stored in csv file using Twitter API(Hassan, n.d.).

3) Technologies

Machine learning,
Natural Language Processing, Linguistic Analysis, Sentiment Analysis,
Google API Services

Algorithm – This is a list or set of rules that a PC will follow to achieve some undertaking or methodology by means of its computations.

Machine Learning – we use algorithms to get computer systems to go through content (images, text, whatever) and identify various trends and patterns across all of those data, based on what we have told them to look for. This can actually be done on unlabeled data as well, via what is called unsupervised learning.

Deep Learning – Basically, this is a subset of machine learning, yet, after we get the framework to distinguish patterns and examples across information by examining content, we request it to continually improve its likelihood from precisely grouping that content by persistently preparing itself on new information that it gets.

Natural language processing (NLP) – This includes utilizing machines to take human language in text or sound configuration – with the entirety of its nuances and subtleties including setting, manner of expression, idioms, and tone – and translating what is implied, in a perfect world with the precision that people have in understanding communicated words and expressions.

Sentiment analysis – This involves using NLP to identify and parse out emotions (affect) and other subjective notions within expressed words or phrases. Within this, there is sentiment polarity and a sentiment score.

B. Data Gathering

1) Twitter API / Kaggle

In this study data collected from two different sources, Data gathered from an online survey using google form to get an idea about knowledge about cyberbullying of general audience and its impact on them. Datasets were collected from Kaggle: Machine Learning Data Community Service and Twitter API for Machine Learning Models Training and Testing for sentiment Analysis. The dataset is sentiment140 dataset. It contains 65,000 tweets extracted using the twitter API. The tweets have been annotated (0 = negative, 1 = positive) and they can be used to detect sentiment. The SCV file contain 6 fields, which are. target: the polarity of the tweet (0 = negative, 1 = positive). ids: The id of the tweet date: the date of the tweet, flag: The query. If there is no query, then this value is NO_QUERY. user: the user that tweeted vi. text: the text of the tweet ("Sentiment140 dataset with 1.6 million tweets," 2020). Id, Target and Text Features were used for the research work.

2) Online Survey

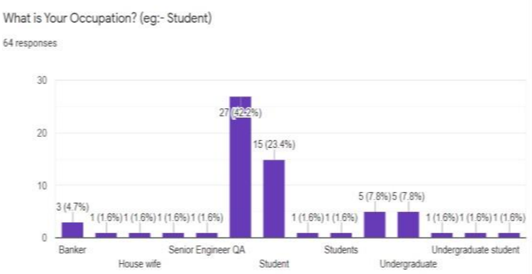

Figure 3.1: -: Occupations

Source: Author

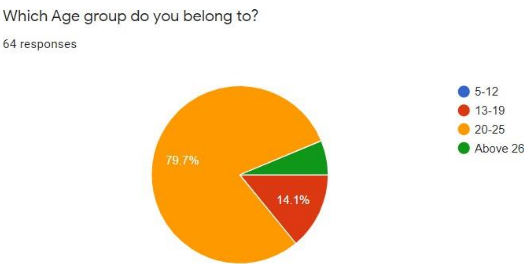

Figure 3.2: -: Age Groups

Source: Author

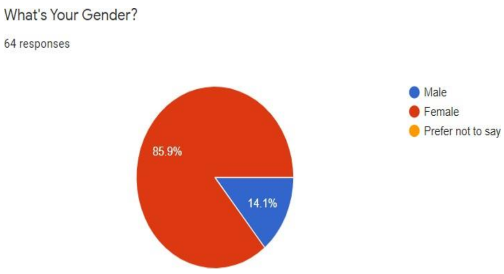

Figure 3.3: -: Gender

Source: Author

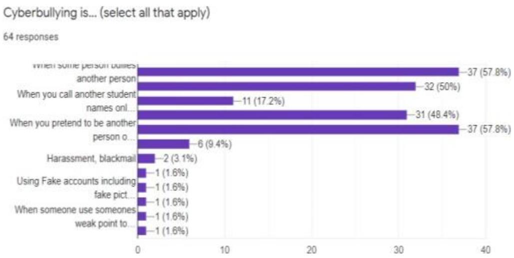

Figure 3.4: -: Cyberbullying Understanding

Source: Author

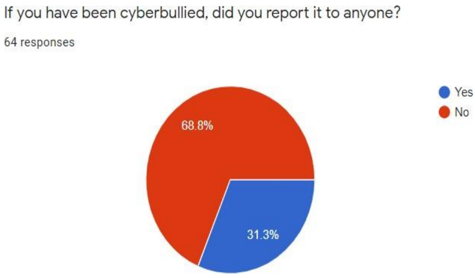

Figure 3.5: -: Cyberbullied Report

Source: Author

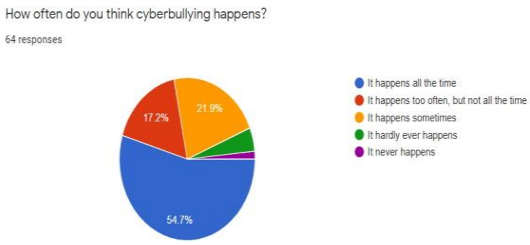

Figure 3.6: -: How often Cyberbullying Happen

Source: Author

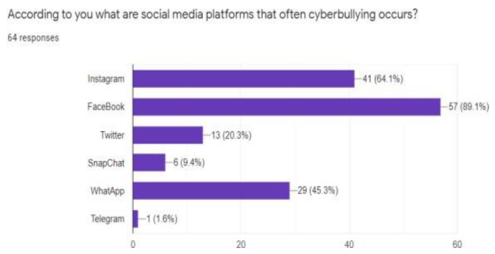

Figure 3.7: -: Often Cyberbullying happening Social Media Platforms

Source: Author

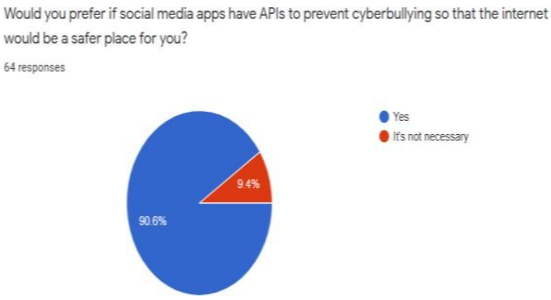

Figure 3.8: -: Preference for Cyberbullying detecting api

Source: Author

From the online survey we've conducted 64 responses were gathered. According to data gathered many people don't have a clear idea their root structure. The "root" for this situation may not be a genuine root word, yet a standard type of the first word. Impacts of stemming curved words stemming just possibly helped improved characterization exactness rather than utilizing better designed highlights and text enhancement approaches, for example, utilizing word implanting.

Lemmatization on the surface is very similar to stemming, where the objective is to eliminate affectations and guide a word to its root structure. The solitary contrast is that, lemmatization attempts to do it the appropriate way. It doesn't simply cleave things off, it really changes words to the genuine root. Stop words are a set of commonly used words in a language. The intuition behind using stop words is that, by removing low information words from text, we can focus on the important words instead (Kotsiantis et al., 2006)

whether they've been through cyberbullying or not. People are not still comfortable with sharing their bullying or bullied experiences in detail, and the case is even worse in countries like ours as the victim can get blamed for what they go through. So from these information only can gather a brief idea what generally people know about cyberbullying and the impact of the problem upon themselves.

C. Data Preprocessing

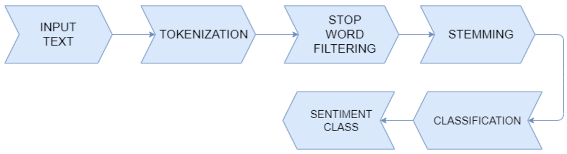

Figure 3.9 -: Data Pre-processing

Source: Author

To preprocess the text simply means to bring the text into a form that is predictable and analyzable for the task. A task here is a combination of approach and domain.

Task = approach + domain

Lowercasing all text data, although commonly overlooked, is one of the simplest and most effective form of text preprocessing. Stemming is the way toward decreasing intonation in words to

D. Building Machine Learning Model for Sentiment Analysis

$$P(y|x_1, \ldots, x_n) = \frac{P(y)P(x_1, \ldots, x_n|1)}{P(x_1, \ldots, x_n)}$$

1) Naïve Bayes Classifier

Naive Bayes classifier is based on the Bayes' Theorem and is a supervised learning approach. Specifically, a supervised learning algorithm takes a known set of input data and known responses to the data, and trains a model to generate reasonable predictions for the response to new data. Naive Bayes classifier is used for sentiment analysis purposes due to its high accuracy. Although it is a simple theorem, it performs almost as well as many other complicated approaches. It is essentially a set of supervised learning algorithms based on the application of Bayes' theorem with the "naive" assumption of independence between every pair of features [3]. Given a class variable and a dependent feature vector through, Bayes' theorem states the following relationship:

For all i, this relationship is further simplified to:

$$P(y \mid x_1, \ldots, x_n) = \frac{P(y) \prod_{i=1}^{n} P(x_i \mid y)}{P(x_1, \ldots, x_n)}$$

This means that - the probability that classification y is correct, given the features , , and so onx1x2equals the probability of y times the product of each x feature given y, divided by the probability of the features.

2) Experimental Setup

We implemented the python nltk package for Naive Bayes classification. The training sets need to be labelled in order to recognize the category a corpus is classified upon. For our case, we are trying to detect bullying and hence we need to find out if a particular tweet is positive or negative or is opinionated/neutral. The negative tweets are regarded as cyberbullying related tweets. The labelled tweets were then stored in CSV file. If it is positive or neutral, then there is no harm done and we leave it at that. However, if a tweet is negative, we can successfully identify cyberbullying. Now the question remained, how accurate was the detection of this negative tweet. We collected a large data set as mentioned before for training the classifier in order to increase the accuracy.

After the bigram features were extracted and added to the feature vector, we trained the Naive Bayes classifier using the built in package function with the 1 million tweets that were collected and annotated. Then we moved on to testing the polarity of the test data. In order to determine how much precise and accurate our classifier was we also found out some metrics like precision, accuracy, recall and f-score.

D. Deploying the Trained Model for Prediction

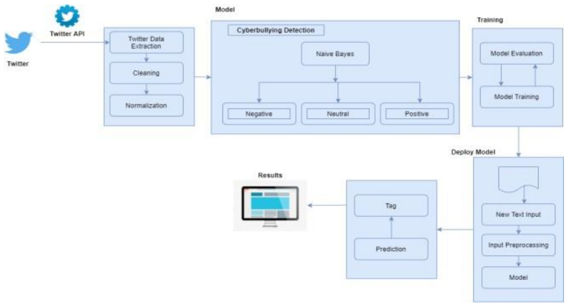

Figure 3-10. Deploy the developed model with Flask

Source: Author

The data is a collection of tweets tagged as 1 as positive or 0 as negative that was collected from twitter API and Kaggle. First, the dataset is used to build a prediction model that will accurately classify which texts are spam. Naive Bayes classifiers are a popular statistical technique of sentiment predicting. They typically use bag of words features to identify negative tweets. Not only Naive Bayes classifier is easy to implement but also provides very good result. After training the model, it is desirable to have a way to persist the model for future use without having to retrain. And we can load and use saved model. Models are persisted in a certain format specific to the language in development. And the model will be served in a micro-service that expose endpoints to receive requests from client.

Having prepared the code for classifying Sentiment of tweets we will develop a web application that consists of a simple web page with a form field that lets us enter a message. After submitting the message to the web application, it will render it and gives us a result of negative or positive. The app.py file contains the main code that will be executed by the Python interpreter to run the Flask web application, it included the ML code for classifying text sentiment.

E) Results

A comparison study was done in order to conclude what the best Algorithm is to use in Sentiment Analysis Model. Algorithm, Accuracy: Test, Precision: Test, Recall: Test, F1 Score: Test, Prediction Time, Accuracy: Train, Precision: Train, Recall: Train, F1 Score: Train, Training

Time Features were used to analyze the results of the pipeline.

Algorithm	Accuracy: Test	Precision: Test	Recall: Test	F1 Score: Test	Prediction Time	Accuracy: Train	Precision: Train	Recall: Train	F1 Score: Train	Training Time
LogisticRegression	0.893372	0.901663	0.922279	0.920579	0.002992	0.94456	0.990241	0.977282	0.983719	0.895054
DecisionTreeClassifier	0.923214	0.954808	0.927041	0.94072	0.023008	0.998845	0.999943	0.9983	0.999121	3.687034
LinearSVC	0.916732	0.946599	0.92551	0.935936	0.006946	0.997019	0.998298	0.997167	0.997733	0.435515
MultinomialNB	0.926344	0.964089	0.940306	0.942721	0	0.978726	0.956763	0.959039	0.9579	0.003988
KNeighborsClassifier	0.857606	0.895161	0.887245	0.891186	45.926606	0.897727	0.927596	0.915982	0.921753	0.015959

Classification Summary of Algorithm was done by using Accuracy: Test, Precision: Test, F1 Score: Test, Recall: Train Features. The gained results are shown in the Figure 3.4.

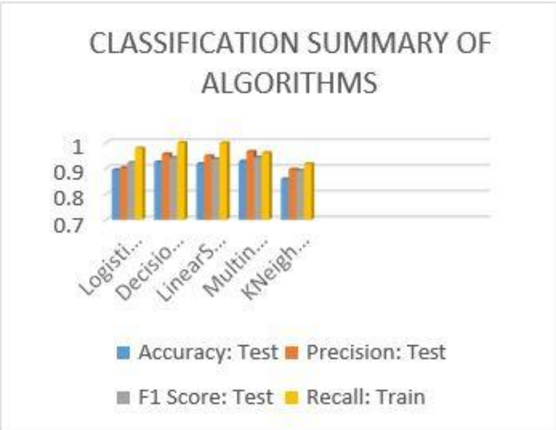

Figure 3-11. Classification Summary of Algorithms

Source: Author

As shown in the figure 3.4
Best Accuracy: 0.926 – MultinomialNB
Best F1 Score: 0.943 – MultinomialNB
Best Precision: 0.964 – MultinomialNB
Best Recall: 0.940 – MultinomialNB

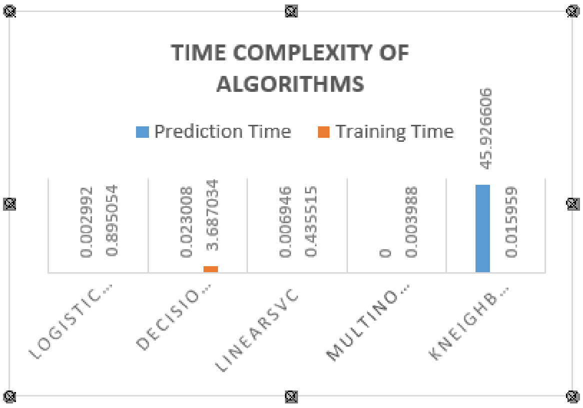

Figure 3-12. Time Complexity of Algorithms

In the figure 3.5 is shown the Time Complexity of Algorithms. According to the results gained

Best training time is 0.004 - MultinomialNB
Best prediction time is 0.0. - MultinomialNB
Worst Training time 3.6870 – Decision Tree Classifier Worst Prediction Time 45.92 – Kneighbors Classifier
Worst Training time 3.6870 – Decision Tree Classifier Worst Prediction Time 45.92 – Kneighbors Classifier
Sentiment Analysis Model was build Using Naïve Bayes Classifier. As out prediction Targets we used Negative, Positive and Neutral states to analyze the tweets Sentiment.

I am happy -> 2.15
I am very bad -> -1.29
this movie should have been great. -> 2.14 great -> 2.14
great great -> 4.28
great great great -> 6.41
great great great great -> 8.55 bad bad bad bad -> -5.18

the above-mentioned results shows prediction Score got from the model. Here 2.15 is positive sentiment and -1.29 is a negative sentiment. And to get more accuracy the words arrays are used to train the system for predictions.
User interfaces are shown below with the user inputs and system outputs.

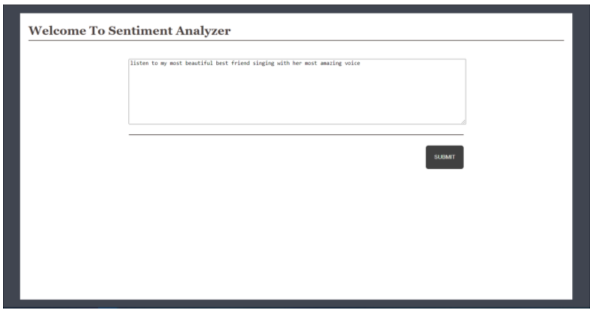

Figure 3-13. Insert Text

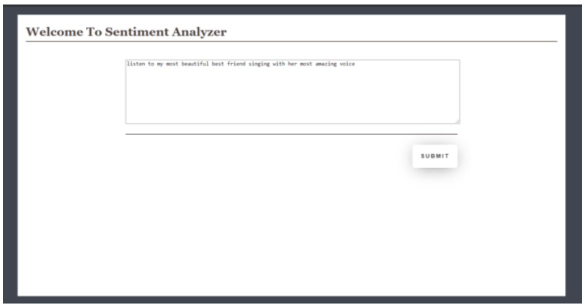

Figure 3-14. Submit the text for sentiment analyze

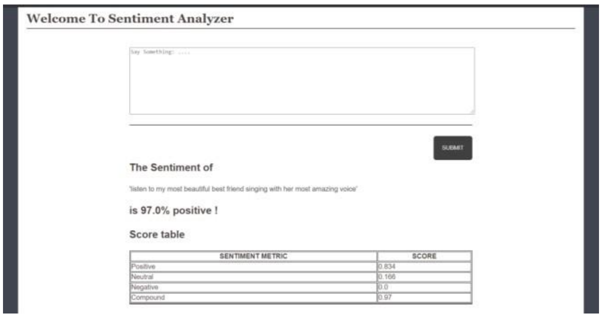

Figure 3-15. Get the results with Sentiment Matric Scores (Positive)

IV. DISCUSSION AND CONCLUSION

Considering the researches that has been done the highest accuracy they have gained are among 70% to 95% percent which claims they are successful at achieving their goal(Birjali et al., 2017). Some researches were there which has good F1 Scores as well. In our research work we did a comparison study to conclude which algorithm the best was considering measurement features mentioned in the results section to analyze sentiment of the tweets with our sentiment analysis model. As the results we gain from the evaluation Naïve Bayes Classifier gives the best accuracy 0.926, Precision 0.964, F1 Score 0.943 of the test dataset and Best recall of the train dataset. The Training time and Prediction Time was also given with Naïve Bayes Model. While Linear SVC, Decision Tree Logistic Regression, Kneighbor Classifiers gives next best results in sequence.

This system was build and trained and tested using Naïve Bayes Model considering the results and the system very accurately predicted the sentiment of the inputs put in by user. The build system was deployed to a Web Interface for user input and predict the sentiment of the input using the Machine Learning model and result is given in a table in three aspect scores which are negative, positive and neutral. The Naïve Bayes Model trains data fastest compared to other classifiers which makes it easier to use large data sets to increase Accuracy of the system.

V. FUTURE WORKS

In our work we build a Sentiment analysis model using Naïve Bayes classifier to predict the sentiment of user input text real time. To extend the research work further the classifiers such as Bagging Classifier, stochastic gradient descent Classifier, Random Forest Classifier and Ada Boost Classifier. Other Feature Engineering Methods also can be used to get the best out of datasets. In our research work cyberbullying prediction will be done using twitter Chabot, but cyberbullying comments deletion can't be done. As a solution for that the user input comments can be translated into decent set of words using Natural Language Process techniques and libraries. As Image Processing is a rising field, can be used to analyze cyberbullying using Screenshots or images detected by the system. If these goals are achieved in the future Internet will be a safe space for every generation, every gender and basically human race.

REFERENCES

51 Critical Cyberbullying Statistics in 2020 [WWW Document], n.d. . BroadbandSearch.net. URL https://www.broadbandsearch.net/blog/cyber-bullying-statistics (accessed 1.22.21).

Affairs (ASPA), A.S. for P., 2019. Facts About Bullying [WWW Document]. StopBullying.gov. URL https://www.stopbullying.gov/resources/facts (accessed 1.22.21).

Al-garadi, M.A., Varathan, K.D., Ravana, S.D., 2016.

Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network. Computers in Human Behavior 63, 433–443.

https://doi.org/10.1016/j.chb.2016.05.051 Birjali, M., Beni-Hssane, A., Erritali, M., 2017. Machine

Learning and Semantic Sentiment Analysis based Algorithms for Suicide Sentiment Prediction in Social Networks. Procedia Computer Science 113, 65–72. https://doi.org/10.1016/j.procs.2017.08.290

Brewer, R., Kieliszewski, C.A., 2014. Exploring Interaction Design for Advanced Analytics and Simulation, in: Nah, F.F.-H. (Ed.), HCI in Business, Lecture Notes in Computer Science. Springer

International Publishing, Cham, pp. 3–14. https://doi.org/10.1007/978-3-319-07293-7_1

Camacho, S., Hassanein, K., Head, M., 2014. Understanding the Factors That Influence the Perceived Severity of Cyber-bullying, in: Nah, F.F.-H. (Ed.), HCI in Business, Lecture Notes in Computer Science.

Springer International Publishing, Cham, pp. 133– 144. https://doi.org/10.1007/978-3-319-07293-7_13

Cleaning & Preprocessing Text Data for Sentiment Analysis | by Muriel Kosaka | Towards Data Science [WWW Document], n.d. URL https://towardsdatascience.com/cleaning-preprocessing-text-data-for-sentiment-analysis-382a41f150d6 (accessed 4.14.21).

Cyberbullying Prevention and Intervention Efforts: Current Knowledge and Future Directions - Dorothy L. Espelage, Jun Sung Hong, 2017 [WWW Document], n.d. URL https://journals.sagepub.com/doi/10.1177/0706743 7 16684793 (accessed 1.29.21).

Cyberbullying Statistics and Facts for 2020, n.d. .

Comparitech. URL https://www.comparitech.com/internet-providers/cyberbullying-statistics/ (accessed 1.23.21).

Cyberbullying Survey: Key Findings - Panda Security [WWW Document], n.d. URL https://www.pandasecurity.com/en/mediacenter/fa m ily-safety/cyberbullying-survey/ (accessed 11.14.20).

Dadvar, M., de Jong, F., 2012. Cyberbullying detection: a step toward a safer internet yard, in: Proceedings of the 21st International Conference on World Wide Web, WWW '12 Companion. Association for Computing Machinery, New York, NY, USA, pp. 121–126. https://doi.org/10.1145/2187980.2187995

Dadvar, M., Trieschnigg, D., de Jong, F., 2014. Experts and Machines against Bullies: A Hybrid Approach to Detect Cyberbullies, in: Sokolova, M., van Beek, P. (Eds.), Advances in Artificial Intelligence, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 275–281. https://doi.org/10.1007/978-3-319-06483-3_25

Frontiers | Technological Resources to Prevent Cyberbullying During Adolescence: The Cyberprogram 2.0 Program and the Cooperative Cybereduca 2.0 Videogame | Psychology [WWWDocument], n.d. URL https://www.frontiersin.org/articles/10.3389/fpsyg.2 018.00745/full (accessed 11.14.20).

Gini, G., Pozzoli, T., 2013. Bullied Children and Psychosomatic Problems: A Meta-analysis. Pediatrics 132, 720–729. https://doi.org/10.1542/peds.2013-0614

Haidar, B., Chamoun, M., Serhrouchni, A., 2017.

Multilingual cyberbullying detection system: Detecting cyberbullying in Arabic content, in: 2017 1st Cyber Security in Networking Conference (CSNet). Presented at the 2017 1st Cyber Security in Networking Conference (CSNet), pp. 1–8. https://doi.org/10.1109/CSNET.2017.8242005

Hassan, A., n.d. Cyberbullying detection using sentiment analysis in social media.

Hinduja, S., Patchin, J., 2017. Cultivating youth resilience to prevent bullying and cyberbullying victimization. Child abuse & neglect 73, 51–62. https://doi.org/10.1016/j.chiabu.2017.09.010

How to Get, Use, & Benefit From Twitter's API [WWW Document], n.d. URL https://blog.hubspot.com/website/how-to-use-twitter-api (accessed 4.14.21).

How Twitter Is Fighting Harassment & Cyberbullying [WWW Document], n.d. URL https://blog.hubspot.com/marketing/twitter-harassment-cyberbullying (accessed 11.14.20).

Ismail, H., Harous, S., Belkhouche, B., 2016. A Comparative Analysis of Machine Learning Classifiers for Twitter Sentiment Analysis.

Jagdale, R., Shirsat, V., Deshmukh, S., 2019. Sentiment Analysis on Product Reviews Using Machine Learning Techniques: Proceeding of CISC 2017. pp. 639–647. https://doi.org/10.1007/978-981-13-0617-

4_61

Kolchyna, O., Souza, T., Treleaven, P., Aste, T., 2015. Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination.

Kotsiantis, S., Kanellopoulos, D., Pintelas, P., 2006. Data Preprocessing for Supervised Learning.

International Journal of Computer Science 1, 111– 117.

Kurka, D.B., Godoy, A., Von Zuben, F.J., 2016. Online Social Network Analysis: A Survey of Research Applications in Computer Science. arXiv:1504.05655 [physics].

Myers, C.-A., Cowie, H., 2019. Cyberbullying across the Lifespan of Education: Issues and Interventions from School to University. International Journal of Environmental Research and Public Health 16, 1217. https://doi.org/10.3390/ijerph16071217

Nationwide teen bullying and cyberbullying study reveals significant issues impacting youth -- ScienceDaily [WWW Document], n.d. URL https://www.sciencedaily.com/releases/2017/02/17 0 221102036.htm (accessed 1.22.21).

(PDF) Using IT Design to Prevent Cyberbullying [WWW Document], n.d. URL https://www.researchgate.net/publication/3176393 5 2_Using_IT_Design_to_Prevent_Cyberbullying (accessed 1.29.21).

Ptaszynski, M., Dybala, P., Matsuba, T., Masui, F., Rzepka, R., Araki, K., 2010. Machine Learning and Affect Analysis Against Cyber-Bullying.

Python, R., n.d. Reading and Writing CSV Files in Python – Real Python [WWW Document]. URL https://realpython.com/python-csv/ (accessed 4.14.21).

Rosa, H., Pereira, N., Ribeiro, R., Ferreira, P.C., Carvalho, J.P., Oliveira, S., Coheur, L., Paulino, P., Veiga Simão, A.M., Trancoso, I., 2019. Automatic cyberbullying detection: A systematic review.

Computers in Human Behavior 93, 333–345. https://doi.org/10.1016/j.chb.2018.12.021

Sentiment140 dataset with 1.6 million tweets [WWW Document], n.d. URL https://kaggle.com/kazanova/sentiment140 (accessed 4.11.21).

Talpur, B.A., O'Sullivan, D., 2020. Cyberbullying severity detection: A machine learning approach. PLoS One

15. https://doi.org/10.1371/journal.pone.0240924 The Priority for Cyberbullying Prevention, 2018. .

Cyberbullying Research Center. URL https://cyberbullying.org/priority-cyberbullying-prevention (accessed 11.14.20).

The-Annual-Bullying-Survey-2017-1.pdf, n.d. (Tutorial) Simplifying Sentiment Analysis in Python -

DataCamp [WWW Document], n.d. URL https://www.datacamp.com/community/tutorials/simplifying-sentiment-analysis-python (accessed 4.14.21).

Xu, J.-M., Jun, K.-S., Zhu, X., Bellmore, A., 2012. Learning from Bullying Traces in Social Media, in: Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Presented at the NAACL-HLT 2012, Association for Computational Linguistics, Montréal, Canada, pp. 656–666.

Zapf, D., Einarsen, S., 2003. Bullying and emotional abuse in the workplace. International perspectives in research and practice. Individual antecedents of bullying: Victims and perpetrators 165–184.

Zhao, R., Zhou, A., Mao, K., 2016. Automatic Detection of Cyberbullying on Social Networks based on Bullying Features. https://doi.org/10.1145/2833312.2849567