Cyberbullying tweet classification using Machine learning

Mohamad Suhail Polur
700745813
dept.Computer Science
University of Central Missouri
mxp58130@ucmo.edu

Adithya Venkatapuram
700742057
dept.Computer Science
University of Central Missouri
axv2057@ucmo.edu

Swathi Jonnalagadda
700744259
dept.Computer Science
University of Central Missouri
sxj42590@ucmo.edu

Maher Sarosh Abdul
700736941
dept.Computer Science
University of Central Missouri
mxa69410@ucmo.edu

***Abstract*—According to a recent survey, the number of people accessing the internet is growing tremendously. From 2000 to 2009 the number went from 394 million to 1.8 billion in the next 10 years 44.3 percent of the people are going to use the internet.With the evolution of internet usage of social media usage also increased drastically. During the pandemic the usage of social media increased more. Major challenge in social media posts is anonymity of the tweets or messages sometimes these messages contain different offensive content also. These kinds of messages or tweets are referred to as cyberbullying. During Covid-19 people encountered various bullying tweets about the pandemic.The statistics of cyberbullying are almost 37 percent of students are falling prey to cyber bullying and 87 percent of these commiting suicides,academic performance decline and depressed . With the evolution of Artificial Intelligence and NLP(Natural Language Processing) dealing with data i.e. huge corpus became easy. NLP allows us to understand the polarities of the corpus and linguistic contexts. These are not the exhaustive list of tasks that NLP can do. The main objective of this project are: Converting the text into vectorizer using count vectorizer Classifying the type of cyber bullying from the given tweet using SVM and Multinomial Naive bayes i.e. multi class classification**

[1] ***Index Terms*—SVM(Support Vector Machine), COVID-19, NLP(Natural Language processing), Bag of words, Cyberbullying tweets and Multinomial Naive Bayes.**

I. Introduction

The goal of the project is to find or filter the abusive messages from the tweets. The challenge with dealing with abusive content is to find or collect the relevant vocabulary of the topic.After collecting the vocabulary we classify the text messages into abusive or normal messages.In this paper we are using n-gram and bi-gram analysis.We use TFIDF vectorizer for feature extraction.After the feature extraction we apply classification algorithms KNN(K-Nearest Neighboring), MNB(Multinomial Naive Bayes) and SVM(Support Vector Machine) algorithms to classify the posts or tweets . One term that is popular nowadays is opinion spamming, which means intentionally writing a review about the product or item in order to damage the reputation of the certain organization. These kinds of activities impact the organizations or businesses in a negative way. There are multiple ways we can find whether the reviews of a certain product are genuine or not using the following steps collecting the dataset, preprocessing the dataset, opinion unique words, extraction of the vocabulary, labeling of the unique words, polarity determination of the words and in the final step result comparison [2]. These tweets can be identified using the following hints too much of emotion, grammar of the sentences. In this paper cyber threat sentences are classified using machine learning techniques.Cyber threat is one of the major issues nowadays mainly on minors to protect people from these threats we use threat miner concept to identify the related words from the sentences. For this task we use a bespoke library to create a dictionary out of the sentences . With the advent of cyber bullying and hate speech analysis. Offensive language detection also plays a crucial role in identifying the offensiveness in the given sentences. In this paper we are using GA(Genetic Algorithm), SVM(Support Vector Machine) on the ArCybc corpus and achieved 87.8 percent of the F1 score and 88.2 percent of accuracy.

In the experimental results we have observed that SVM outperformed the Multinomial Naive Bayes algorithm. Following figures show the classification reports of SVM and Multinomial Naive Bayes. After the experimental analysis we conclude that precision, recall and F1-score are high for following categories: age,gender, ethnicity, not cyber bullying types in case of SVM. When we consider Multinomial Naive Bayes, these metrics are uniform.

Main features of the project are: Implementing Multinomial Naive Bayes classifier for multi classification of type cyber bullying. This algorithm combines multiple algorithms and works based on probabilistic learning. Main theme behind the multinomial naive bayes is Bayes theorem In this project

[1]https://github.com/MohamadSuhail/Cyber-bullying-tweet-prediction

we explore different text processing libraries and for model creations we use scikit-learn library. For data visualization we use seaborn and matplotlib.The whole project is implemented in the Python programming language in PyCharm IDE.

For model evaluations we use various metrics of accuracy to generalize the model performance. To understand the models in depth using classification reports we publish recall, F1 score and precision metrics.

To measure the separability of the model decision boundary of each model is visualized using scikit-learn library to understand the area under the curve we used ROC-AUC curves. In this project we explore different text processing libraries and for model creations we use scikit-learn library. For data visualization we use seaborn and matplotlib.The whole project is implemented in the Python programming language in PyCharm IDE.

For model evaluations we use various metrics of accuracy to generalize the model performance. To understand the models in depth using classification reports we publish recall, F1 score and precision metrics.

To measure the separability of the model decision boundary of each model is visualized using scikit-learn library to understand the area under the curve we used ROC-AUC curves.

In this paper we are proposing a classification model to analyze sentiments into cyber bullied messages from normal messages using multiple machine learning algorithms. Applying multiple algorithms a comparative analysis of the algorithms and selecting the best algorithm considering the different performance measures.

With the advent of the internet people are accessing different kinds of information to process and read. One of the popular languages is Chinese. There are multiple screening languages in the existing literature but the recent studies shows that there is huge improvement in the classification accuracy if add the WAE topic modelling method before we classify the text. In this experiment we have implemented a WAE model with an SVM classifier and the results show that the accuracy of the model improved drastically. The classification of texts using distance measurement is a new method which effectively classifies the text. In this paper we are proposing various distance measurement parameters like Euclidean distance and cosine similarity metrics. The two methods achieved similar accuracy on the sample 1000 documents. This experiment is conducted on the Bangla dataset.

In this paper we are proposing a model that can classify the text in a novel approach. In the first stage the raw text is preprocessed using text preprocessing methods like BeautifulSoup and regular expression. In the second step we select the features of statistical methods like mutual information. In the third step we extract features which are useful to the classification and discard the unnecessary words using LDA topic modelling method. In the last stage we use ensemble classifiers to classify the text.

Nowadays text classification has a huge area to research. In this paper we are proposing a text classification model to classify the text in 3 stages: Attribute value discretization, text set information systems and determine text vectors. This research proved that the accuracy parameters are improved drastically with precision, recall and F1 score. In this paper we are proposing a n-gram BiLSTM method to classify the texts. The implementation is broken into two methods one is one-vs-one method and the second one one-vs- rest method. In the experimental analysis results show that N-gram BiLSTM outperformed the traditional LSTM classifiers. Traditional text classifiers follow the below assumptions: one is there should be enough training data to achieve good results and the other one is we must train the data with a good classification model. In this paper we are proposing a model which classifies the comment review text classifier using a transfer learning model TrAdaBoost classifier which outperformed the tradi- tional algorithms. In this study, a sentiment analysis pipeline is implemented and integrated into the continuing open source cross-media analysis system. The chat room cleaner, NLP, and sentiment analyzer are included in the pipeline. Prior to the integration, we contrast lexicon-based and machine learning methodologies, two main kinds of sentiment analysis methods. Our major goal is to determine the best technique for extracting feelings from forum discussion threads. We employ Stanford coreNLP library's Recursive Neural Tensor Network (RNTN) model and Apache-Hadoop framework's lexicon-based sentiment prediction algorithm in our research. The issue of dynamic polarity shift in review analysis is researched in this work. First, the Apriori algorithm is used to extend the sentimentally ambiguous terms based on context and create the triples of (sentiment object, sentiment word, and sentiment polarity) that make up the sentiment ambiguous lexicon. Then, utilise the condition random field model (CRFs) to fine-grainedly analyse sentiment orientation based on the sentiment ambiguous lexicon to extract emotional aspects from comments. Experimental results using product corpora from the computer and mobile phone industries demonstrate the viability of the suggested approach and assist boost sentiment analysis's precision.

II. Motivation

Cyberbullying is the act of harassing people through an electronic medium that is also known as online bullying. The forms of cyberbullying includes hate speech, sexual remarks, spreading a victims personal information on social media platforms. Not all negative comments and remarks are attributed to cyberbullying. Main victims in the process are teenagers with these harassments teengers are committing suicides, low grades in academics and falling into depression.

In the year of 2010 several high profile cases were filed so from this year there were so many awareness camps being conducted in the USA. The seriousness of the cyberbullying crime depends on the region and demographics of the region. There is an increasing trend in cyber bullying during the Covid-19 since the quarantine of the teenagers and surge of internet usage. It is published in the Journal of Social Psychology that there is an increase in pro-cyberbullying but research also suggests that teenagers between the 4-12 grades in Canada show less cyberbullying due to close monitoring of parents in the quarantine time. Statisticians point out that sophisticated computer techniques could solve the problem of identifying the cyberbullying texts and we can prevent the potential threats on teenagers.

III. Objectives

The main objectives of the project are:

- Classifying the text into one of 5 categories of cyber-bullying i.e. age, gender, other cyberbullying, religion, ethnicity and normal(not cyberbullying) categories.
- To vectorize the text inputs using the Bag of Words model that is a count vectorizer.
- Implementing various machine learning models SVM(Support Vector Machine) and Multinomial Naive Bayes classifiers on the preprocessed text data.
- Conducting a comparative analysis with prediction accuracies using precision, recall and F1 score metrics.
- Selecting the best model to conclude the whole process

IV. Related work

In this paper we are proposing deep learning techniques to compare the performance of the algorithms on text classification tasks. To classify the text we have implemented the attention based mechanism to classify the text. In this paper we have tested the mechanism on 3 benchmark datasets; those are news dataset, reviews datasets.After applying the classification algorithms we have implemented the applications using high accuracy algorithm [7]. The most important step in the text based classification or any other classification task in machine learning begins with the cleaning process. The main principle of machine learning is that good data produces good results. CNN is coined as the best classification method. How this works with preprocessing works with CNN. After the experimental analysis we have found that there is a weak effect on CNN after the preprocessing method. All these experiments are conducted on Turkish data of different datasets [10]. With the increase in use of the internet and the production of huge amounts of data there is a need for improvement in the accuracy of the text classification models. With deep learning models like BERT(Bidirectional Encodes Representation from Transformers) are used to improve the accuracy of the classifiers. Texts are converted into vectors and the vectors are given as input to the BERT model. This model is tested on the task of MRPC and it gained 93.25 pecentage F1 score on the data. This model can be used in the future evaluation [14]. The applications of text classification are widely spread to different areas like optimising the search engines, news classification , information extraction and preprocessing of electronic emails. The text classification is extended to multiclass classification also. In this paper we have implemented the multiple machine learning models to classify the 5 categories of the news including politics, economy, sports, health and technology. To classify the text we have implemented Support Vector Machine, K-Nearest Neighbour, Decision tree, Multinomial Naive Bayes and Bernoulli Naive Bayes algorithms. After the experimental analysis we have concluded that Naive Bayes algorithm performs the better accuracy on the Turkish dataset with 90 percent accuracy. In the future we can implement the algorithm in web based platforms [6]. In the recent studies a fusion model that a fusion decision layer proves to be the best model to classify the texts and the sentences. In the fusion model each layer consists of the following algorithms KNN, BP Net, SVM as the feature layers. These layers are constructed based on the DS theory. For the comparative analysis we have constructed a baseline model to compare the results with the fusion model. In the end we have compared the results with the fusion model. It is proven that the precision of the fusion model is high compared to the other models [17] . In this paper we have proposed a short text classification model to perform the classification of short text. Main methods used in this paper are referenced from the previous papers collected; those are feature extraction in combination with the traditional algorithms [5].

In the text mining converting the text into vectors is an important segment. To do this task we have numerous methods. One of the effective methods is the count vectorizer which converts the text or frequency of the words into vectors. In this model we have implemented the countvectorizer on the dataset of UCI Reuter news dataset the performance of the Support Vector Machine. The performance is further segregated into linear kernel and polynomial kernel. SVM with linear kernel achieves better accuracy than the polynomial kernel [1] . The performance of the classification of text documents is further improved with the feature selection methods. With the improved accuracy these models can be applied in Parts of Speech tagging, news classification, spam filtering etcThe main features of the project are extracting the features as single word and multiple words. The context of the single words and 2 words differ in many situations. For the feature selection we have used the Chi-Square statistical tool to select the features and Naive Bayes and SVM algorithms to classify the texts [15].

There are several tasks we can do with text based corpus one is finding the semantics of the text using deep learning. In this method we are combining the ALBERTNET with CNN to find the similarity of the texts. The similarity of the texts is generated based on the multi label concept.It copies the labels from the target to source text [9]. Across the world we have many different languages. One of the difficult languages is Chinese. In this paper we are proposing a machine learning model to classify a representation method for the Chinese language model. The pipeline of the model includes text representation, feature selection and classification. In the conclusion part we discuss the advantages and disadvantages of the model in depth [2].

With the tremendous increase of the internet there is a need for advanced text classification methods like topic modelling. This paper talks about the differences between the traditional machine learning algorithms and the proposed algorithms and how they differ in different aspects in this task. Research shows the traditional algorithms can not handle the tasks well and there is an improvement needed for SVM by 20 percent and LDA performs better than the SVM algorithm [9].

Text classification tasks are moving towards the non english language analysis. In this paper we are proposing a model which classifies the Tibetian text. Main steps involved in the Tibetian text classification are: Preprocessing the text like removal of stop words, constructing TF-IDF vectorizer and Word2 vector method.Traditional classification methods uses RNN-LSTM method to classify the texts but in this paper we are proposing a Bidirectional LSTM method to classify the text [16].

Traditional classification method has two main drawbacks, one is the curse of dimensionality and poor performance towards classification. To overcome these 2 main challenges we are implementing the fasText method. In this method we have observed that the dimensionality of the corpus or vocabulary is reduced to produce satisfactory results in terms of all the accuracy parameters like precision, recall and F1 score [3].

Due to digitalization of documents in all the government offices most of the documents now are preserved in digital format so there is a necessity to recognise text and segment them.In this paper we are proposing a digital text classification for this task we have implemented a Logical Text Classification Strategy method(LTCS) the pipeline of this method follows: image processing, segmentation of the image, feature extraction , classification of documents and image post processing [13]

With the advances of text classification we are concentrating on complex classification tasks such as patent classification. In the existing methods this task is achieved through LSHTC(Large Scale Hierarchical Text Classification) method which is a computationally expensive one. In our paper we are proposing a comparative study of two methods that are hierarchical learning methods SVM and K-NN [11]. In the experimental analysis we have concluded that K-NN outperforms the SVM model [17]. With the increase of production of data in the media industry everyday by huge amounts there is a growing demand for automatic classification of text [8]. In this paper we are proposing a novel approach to classify the text that is Latent Dirichlet Allocation (LDA) and to get the features from the corpus we are using topic modelling and for multiclass classification we have deployed multi class softmax regression.These models are implemented on the real time database of news which achieved good results [4] [12] [18].

V. Data Description

The dataset for the project is collected from the open source repository Kaggle. The dataset contains 47000 tweets from 6 categories:age,gender, ethnicity, religion, other types of cyberbullying and not cyber bullying. Roughly each category of tweets contains 8000 tweets. The dataset contains two feature columns one is text body and the other one is category or the type of cyberbullying it is targeting towards.

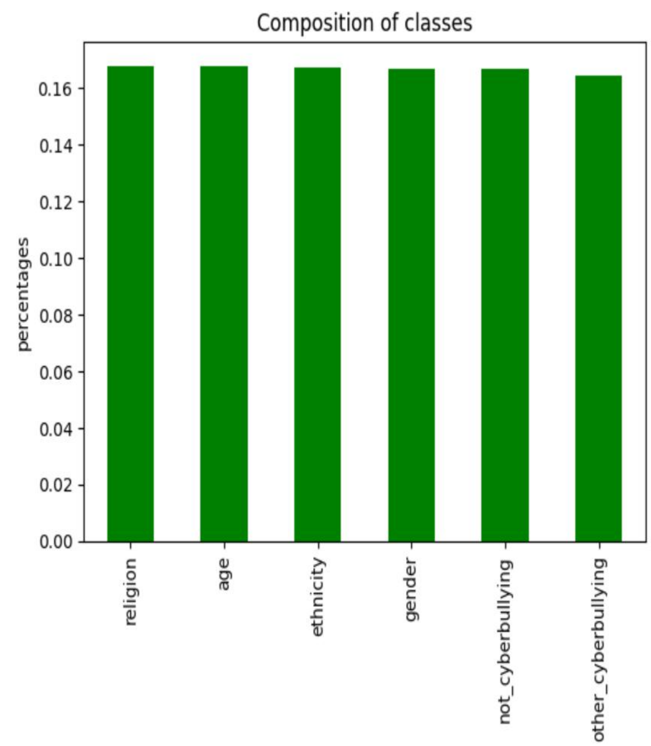

Figure 1. Composition of cyberbullying classes

VI. Proposed Framework

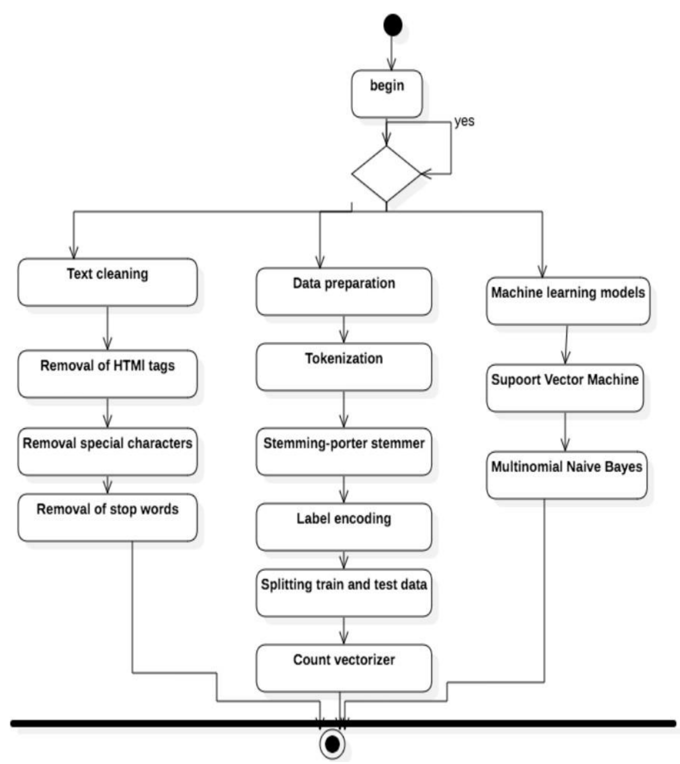

Figure 2. Activity diagram

Project implementation is broken down into following steps:

Data collection: Data is collected from the Kaggle repository.

Text cleaning: First step in Machine learning is cleaning the raw data. If we process the raw data as it is we may encounter high training time and low accuracy scores. From the collected data we remove the parts which do not contribute to the context of the sentence. In the first step of preprocessing we remove the stop words from the corpus. Stop words are the list words which have less importance in the linguist context or meaning of the sentence. For example a, is, the , are so on.

Data preparation: In the data preparation step we encode the categorical target columns. We convert the preprocessed text to vector using CountVectorizer. Count vectorizer takes the rows as documents or sentences and columns as unique words. It converts the sentences to vectors using documents and unique words.

Machine learning algorithms: After converting the text into vectors we apply SVM(Support Vector Machine) and Multinomial Naive Bayes classifiers to to classify the polarities of the reviews.

SVM, Multinomial Naive Bayes and Count Vectorizer are implemented using scikit-learn

A. Methods

Support Vector Machine can be used for both regression and classification tasks. In the case of classification, Support Vector Machine is an algorithm which separates the two classes of vectors using a decision boundary. In general SVM can be used to classify the two classes i.e. binary classification and more than two classes which is multi class classification. In our paper we are using SVM for a multi class classification problem that is cyberbullying type of category prediction. There are 5 classes in this problem set. In case of text classification also we need to convert the text into vectors. Here, the vectors are the numbers which can be represented in the coordinate space. So SVM decides the best fitting line for the two classes. In case of non-linear data SVM classifies the data using kernel tricks to make a hyperplane to separate the classes.

Pseudo code:

- Start
- Input the dataset
- Classify the dataset
- Apply the SVM machine learning with kernel functions like linear, polynomial, radial and sigmoid functions
- Display the decision boundary
- If the obtained accuracy is not good we need to tune the hyper parameters
- End

Multinomial Naive Bayes follows a probabilistic approach to classify the things.The main advantages of using the Multinomial Naive Bayes classifier is low complexity and we can avoid the repeated training cycles to obtain good accuracy.

Pseudo Code:

- Split the sample S into set of n-terms
- For each class compute the vector W of n features in the document C where wi is the frequency of the word in the document
- Evaluate the full probability p(Ck) of the class Ck the class Compute the posterior probability pr(Ck—W)
- Determine the class Cs of the sample and then evaluate the algorithms complexity

Logistic regression is used when the dependent or input variables are categorical. There are many types of logistic regression, binary, multi class regression. In our project we have implemented a multi class logistic regression model to predict the type of the cyber bullying. Logistic regression has various applications in various fields like predicting the email as spam or ham, whether the tumour is malignant or benign. Unlike linear regression it uses the sigmoid function to classify the classes.For binary classes the output will be 0 or 1 and the hypothesis it will follow is Z= WX+B where B is the bias.

The hyper parameters of the logistic regression solver, penalty and C values.Logistic regression handles the multiclass classification problems also this can be done using one vs all or multinomial logistic regression. This model is trained using a training dataset and tested using the test set.To evaluate the model we used various metrics to measure the performance of the model. We have constructed the classification report and confusion matrix to evaluate the test performance of the model.

Count vectorizer is a tool which converts the text into numbers or forms the vectors using the frequency of the words. This vector is formed using the rows as the text document and columns as the vocabulary of the corpus. With this we form a matrix of rows vs columns and each value in the matrix represents the count of the word in the document.This way of representation is called a sparse matrix. If the word does not appear in the document it will be represented with 0. This method is implemented using sklearn Countvectorizer object.In the first step we initialise the object and transform the training and testing sets with the count vectorizer.

Raw data contains several non contextual literals which do not contribute to the context of the corpus; we use regular expressions to remove these literals and punctuation marks. In this step we remove the stopwords from the corpus stop words are the words which don’t help in finding the context of the text. Some of the stopwords in the English language are: is, are, contractions like won’t, can’t. Word tokenization is also part of preprocessing of the text; each sentence is split into words which are technically called as tokens. For this task we have implemented the NLTK tokenizer to tokenize the input text. After tokenization we have implemented the stemming of the preprocessed text to stem the text we used Porter stemming method. Based on multiple conditions or rules we stem the words to produce less computationally expensive vectors to process

Tools and technologies used in this project to implement are: Python programming language to execute the source code To execute the code we chose google colaboratory for its interactiveness. Basic Python libraries used to implement the code are: Numerical Python i.e. numpy and Pandas library which analyses the data. Exploratory data analysis is conducted to understand the underlying structure of the data.To visualise the data and its representations we have used seaborn and matplotlib visualisation libraries. Main project flow of the project is implemented using scikit-learn library. Models like SVM, Multinomial Naive Bayes and Logistic Regression are initialised and trained using scikit-learn library and trained. The data is split into training and testing sets. Test performance is evaluated using a test dataset. For the performance evaluation also the scikit-learn library is used.

Scikit-learn is a free software machine learning library which offers ready to use machine learning models and methods. This mainly developed in the Python programming language. This library extensively uses the other python libraries Pandas and numpy. These libraries are the basic building blocks for scientific computing and data analysis.

To measure the performance of the model we have implemented the classification report and confusion matrix. These two explore the true/false positives and true/false negatives. From classification report we can calculate the precision , recall and F1 score of the models.

- Precision = TruePositives / (TruePositives + FalsePositives)
- recall = TruePositives / (TruePositives + FalseNegatives)
- F1 score = 2*(precision*recall)/(precision+recall)

Precision identifies when False Positive is high. For example, positive or negative reviews detection. a false positive means that an review that is positive comment or review has been identified as positive (predicted positive review).

Recall identifies when False negative is high. For example, positive or negative reviews detection. a false negative means that an review that is negative comment or review has been identified as positive (predicted negative review).

F1 score is the harmoinic mean of the precision and recall equally focuses on the false positives and false negatives

VII. Results Analysis

A. SVM

SVM classifier is used to classify the multi class categories of the cyberbullying corpus. In the first stage data is split into training and testing. After the splitting of whole corpus model is trained using training set and test set or the unseen data is used to evaluate the performance of the data on the cyberbullying classification task. To evaluate the performance of the model various parameters are calcualted to compare the accuracy of the test samples. In the first level of evaluation classification matrix is used which displays the each class i.e. 6 categories precision, recall and F1 score values. Not all the machine learning models are 100 percent accurate so classification report gives the summary of the accuacy of the models but the it will not any information about the quantity of the misclassified samples of the test data compared to the ground truth. To display the misclassified samples of each category we have constructed the confustion matrix. Both classification report and confusion matrix are constructed using scikit-learn library.

B. Multinomial Naive Bayes

Multinomial Naive Bayes classifier works based on probabilistic learning. This can be used for multi class classification tasks. Our problem type is multi class classification which classifies the cyber bullying categories into one of the predicted categories based on the training accuracy and quality of the data.

C. Logistic regression

Logistic regression is the basic classifier which can be used as baseline model to compare the performance of the models. In this project we have considered the Logistic regression model as the baseline model. This model is imported using scikit-learn library and evaluated using scikit-learn evaluation methods classification report and confusion matrix. The performance of the test samples evaluated using above methods and compared with the other classification models in the project to get the better view of working of the classification task.

To test the performance of the classification task in general 3 different classifiers are used. These 3 algorithms differ in their working and functioning. So the comparative analysis gives the better understanding of the cyberbullying task.

	precision	recall	f1-score	support
0	0.98	0.97	0.98	1589
1	0.99	0.98	0.99	1539
2	0.91	0.84	0.87	1573
3	0.69	0.47	0.56	1601
4	0.59	0.84	0.69	1592
5	0.95	0.94	0.95	1638
accuracy			0.84	9532
macro avg	0.85	0.84	0.84	9532
weighted avg	0.85	0.84	0.84	9532

Figure 3. SVM classification report

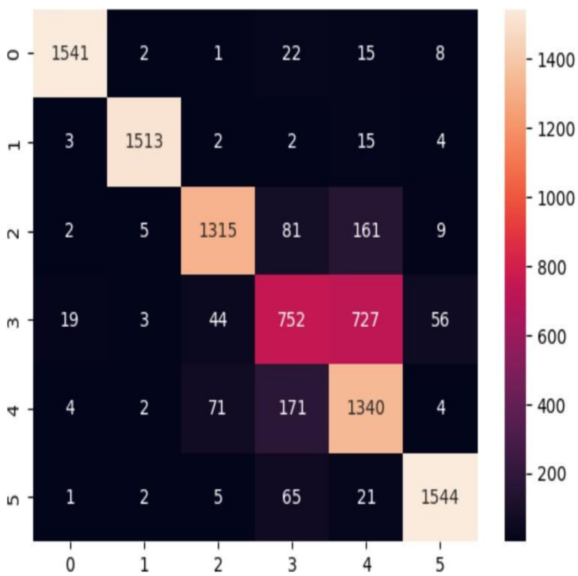

Figure 4. SVM confusion matrix

VIII. Results summary

Conducting comparative analysis of the algorithms we have found the following observations:

- Support Vector Machine achieved 84 percent accuracy on the test dataset
- It scored highest precision for gender category and lowest precision for other types of cyber bullying category
- It scored highest recall for gender category and least for religion category
- It scored highest F1 score for gender category and least for religion category
- The resultant confusion matrix is displayed for the SVM algorithm
- The overall classification accuracy of the multinomial niave bayes is 77 percent

	precision	recall	f1-score	support
0	0.71	0.99	0.83	1529
1	0.85	0.94	0.89	1644
2	0.85	0.83	0.84	1627
3	0.67	0.39	0.50	1558
4	0.66	0.51	0.58	1620
5	0.84	0.98	0.90	1554
accuracy			0.77	9532
macro avg	0.76	0.77	0.76	9532
weighted avg	0.76	0.77	0.76	9532

Figure 5. Multinomial Naive Bayes classification report

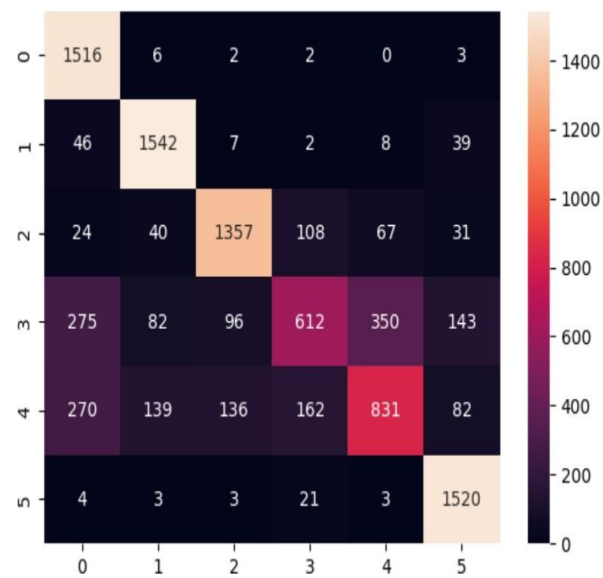

Figure 6. Confusion matrix of multinomial Naive Bayes

- It scored highest precision for gender and ethnicity categories and least for other cyber bullying category
- It scored highest recall for age category and least for religion category
- It scored highest F1 score gender category and least religion category
- the resultant confusion matrix is displayed
- Logistic regression model achieved 83 percent accuracy on the test dataset
- It scored highest precision for gender category and least for religion category

	precision	recall	f1-score	support
0	0.97	0.98	0.97	1529
1	0.99	0.99	0.99	1644
2	0.89	0.85	0.87	1627
3	0.58	0.55	0.57	1558
4	0.60	0.66	0.63	1620
5	0.96	0.95	0.95	1554
accuracy			0.83	9532
macro avg	0.83	0.83	0.83	9532
weighted avg	0.83	0.83	0.83	9532

Figure 7. Logistic regression classification report

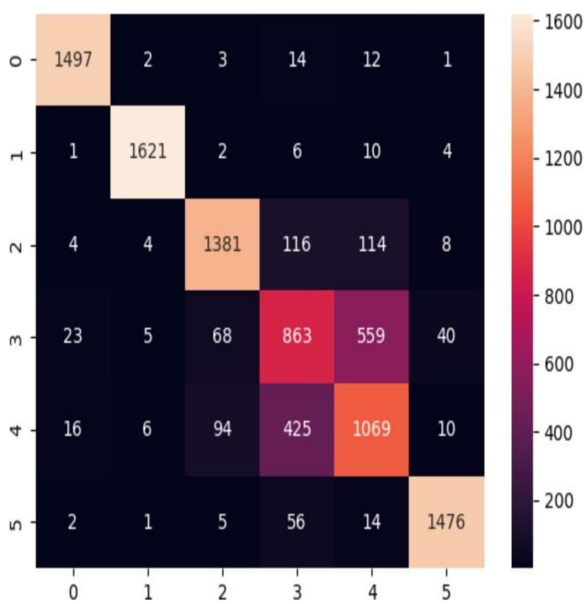

Figure 8. Confusion matrix of logistic regression

- It scored highest recall for gender category and least for reliogion category
- It scored highest F1 score for gender category and least for religion category
- resultant confusion matrix is displayed

To conclude the project Logistic regression outperformed the Multinomial Naive Bayes and Support Vector Machine algorithms. Highest scores achieved in the following categories overall age,gender and least scores are achieved religion and other cyber bullying categories. In the future score we would like to implement the algorithms with feature extraction methods to further improve the accuracy of the models.After improving the model accuracy these are deployed into web application.

REFERENCES

[1] Sakib Al Hasan, Md Gulzar Hussain, Joy Protim, Md Mostafizur Rahman, Nawshad Fahim, Mehrab Zaman Chowdhury, and Ahmed Iqbal Pritom. Classification of multi-labeled text articles with reuters dataset using svm. In *2022 International Conference on Science and Technology (ICOSTECH)*, pages 01–05. IEEE, 2022.

[2] Shihong Chen, Qichao Kuang, Xiaohui Yu, Suifeng Li, and Ruoyao Ding. A multi-label classification algorithm for non-standard text. In *2022 International Conference on Asian Language Processing (IALP)*, pages 206–211. IEEE, 2022.

[3] Xiaobo Chen and Huimin Zhang. Rail surface defects detection based on faster r-cnn. In *2020 International Conference on Artificial Intelligence and Electromechanical Automation (AIEA)*, pages 819–822. IEEE, 2020.

[4] Ankita Dhar, NiladriSekhar Dash, and Kaushik Roy. Classification of text documents through distance measurement: An experiment with multi-domain bangla text documents. In *2017 3rd International Conference on Advances in Computing, Communication & Automation (ICACCA)(Fall)*, pages 1–6. IEEE, 2017.

[5] Zhou Faguo, Zhang Fan, Yang Bingru, and Yu Xingang. Research on short text classification algorithm based on statistics and rules. In *2010 Third international symposium on electronic commerce and security*, pages 3–7. IEEE, 2010.

[6] Fatih G ürcan. Multi-class classification of turkish texts with machine learning algorithms. In *2018 2nd international symposium on multidisciplinary studies and innovative technologies (ismsit)*, pages 1–5. IEEE, 2018.

[7] Gang Jin. Application optimization of nlp system under deep learning technology in text semantics and text classification. In *2022 International Conference on Education, Network and Information Technology (ICENIT)*, pages 279–283. IEEE, 2022.

[8] Zhenzhong Li, Wenqian Shang, and Menghan Yan. News text classification model based on topic model. In *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, pages 1–5. IEEE, 2016.

[9] Yajing Ma, Yonghong Li, Xiaolong Wu, and Xiang Zhang. Chinese text classification review. In *2018 9th International Conference on Information Technology in Medicine and Education (ITME)*, pages 737–739. IEEE, 2018.

[10] Mehmet Umut Salur and İlhan Aydin. The impact of preprocessing on classification performance in convolutional neural networks for turkish text. In *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, pages 1–4. IEEE, 2018.

[11] M Tamilselvi, G Ramkumar, G Anitha, P Nirmala, and S Ramesh. A novel text recognition scheme using classification assisted digital image processing strategy. In *2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, pages 1–6. IEEE, 2022.

[12] Lin Xun, Li Zhishu, Zhou Yong, and Xue Yuan. Text classification algorithm study based on rough set theory. In *2010 International Forum on Information Technology and Applications*, volume 1, pages 117–120. IEEE, 2010.

[13] Tengjun Yao, Zhengang Zhai, and Bingtao Gao. Text classification model based on fasttext. In *2020 IEEE International Conference on Artificial Intelligence and Information Systems (ICAIIS)*, pages 154–157. IEEE, 2020.

[14] Bihui Yu, Chen Deng, and Liping Bu. Policy text classification algorithm based on bert. In *2022 11th International Conference of Information and Communication Technology (ICTech))*, pages 488–491. IEEE, 2022.

[15] Yujia Zhai, Wei Song, Xianjun Liu, Lizhen Liu, and Xinlei Zhao. A chi-square statistics based feature selection method in text classification. In *2018 IEEE 9th International conference on software engineering and service science (ICSESS)*, pages 160–163. IEEE, 2018.

[16] Bo Zhang. News text classification algorithm based on machine learning technology. In *2022 International Conference on Education, Network and Information Technology (ICENIT)*, pages 182–186. IEEE, 2022.

[17] Jian Zhang, Hai Zhao, and Bao-Liang Lu. A comparative study on two large-scale hierarchical text classification tasks' solutions. In *2010 International Conference on Machine Learning and Cybernetics*, volume 6, pages 3275–3280. IEEE, 2010.

[18] Yunxiang Zhang and Zhuyi Rao. n-bilstm: Bilstm with n-gram features for text classification. In *2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*, pages 1056–1059. IEEE, 2020.