Vol. 7, Issue 2, pp. 107-118 | DECEMBER 2022
DOI: https://doi.org/10.53840/myjict7-2-163

USING NATURAL LANGUAGE PROCESSING TO DETECT OFFENSIVE TEXT AND CYBERBULLYING IN SOCIAL MEDIA: A REVIEW

[i*]Abdulkarim Faraj Alqahtani & Mohammad Ilyas
Department of Electrical Engineering and
Computer Science Florida Atlantic University, Boca Raton, FL, USA aalqahtani2021@fau.edu &

*(Corresponding author) e-mail: ilyas@fau.edu.my

Article history:
Submission date: 12 October 2022
Received in revised form: 14 November 2022
Acceptance date: 7 December 2022
Available online: 31 December 2022

Keywords:
Detection of offensive texts, Natural Language Processing, Cyberbullying, Sentiment Analysis, Machine Learning

Funding:
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing interest:
The author(s) have declared that no competing interests exist.

Cite as:
Abdulkarim Faraj Alqahtani, & Mohammad Ilyas. (2022). Using Natural Language Processing to Detect Offensive Text and Cyberbullying in Social Media: A Review. *Malaysian Journal of Information and Communication Technology (MyJICT), 7*(2), 107-118.
https://doi.org/10.53840/myjict7-2-163

Abstract

Social media is used extensively in various fields, including Natural Language Processing research. The exponential growth of user-generated social media content raises the potential for opinion mining to analyze consumer behavior. Sentiment Analysis concerns the computer treatment of people's views, ideas, and subjective experiences to recognize and extract sentiment from a text. The influence of religious conviction, halal awareness, halal certification, and food ingredients are all-inclusive in Muslim consumers' considerations while acquiring halal food. Those aspects assist in uncovering sentimental analysis on customer dependability toward halal certification during expressing issues related to halal food acquisition on social media. We aim to determine a correlation between the halal certification scheme and reliability prediction classification relating to customers' intention to acquire halal products. Identifying corresponding sentiment polarities in sentences is generated by utilizing the Malaya NLTK library to label them as positive, negative, or neutral. A sample of 895 tweets with the hashtag #sijilhalal was gauged and trained for an accurate prediction Machine Learning model using Random Forest, Logistic Regression, K-Nearest Neighbor, Decision Tree, and Naive Bayes. The Correlation Matrix for each model shows how features are related. As in its Confusion Matrix, we can observe an outcome overview of the model's accuracy, precision, recall, and F1-score. It successfully demonstrates the performance evaluation metric on model efficiency and applicability in choosing the best higher-accuracy model for predicting customer tendency toward halal food acquisition. In general, the Logistic Regression model performs the best in this study of predicting the occurrence of customer trustworthiness toward the acquisition of halal food. The findings show that consumers' belief in a food source and halal certification leads them to entirely acquire the food (tagged as loyal) or disregard the food (tagged as churn).

Introduction

Living with technology has become part of our lives and we completely depend on technology for many needs in our life. Thus, in recent times people communicate with each other through social media, whatever the environment that connects them. They may know others physically in work or school environments, so this communication may lead to negative effects when connecting with each other such as writing offensive text in social media or bullying (Husain & Uzuner, 2021). Currently, physical bullying has become cyberbullying with the prevalent use of technology. This phenomenon has increased in many countries around the world. It is risky for teens and considered harmful for them. Cyberbullying can be in many environments such as schools and workplaces via social media connections in multiple ways for example text messages, voice messages, and videos and image messages. Cyberbullying is widespread globally, especially in the age of teens, so the most common cyberbullying is in schools. Currently, most students use internet devices either for school requirements or for social media connections. Some of them use these devices to bully each other. Many studies emphasize that cyberbullying is victimization (Chun et al., 2020). This is a concerning issue as young people can harm each other via bullying each other. Some young people harm themselves when they are bullied. Young people spend most of their time in school and they can keep connected with others via social media, so it is necessary to create a safe environment for young people to prevent any harm that may potentially happen to them. In addition, cyberbullying can adversely impact the educational environment, and relationship between students. It is, therefore, essential to address these aspects for healthy and emotionally rewarding growth of students. This research paper provides some solutions that are suggested in many research publications to address this phenomenon in the school environment (Faucher et al., 2020).

According to some research studies, the students who are targeted for bullying behavior, develop a negative side effect such as poor academic performance and may develop mental health issues. In addition, cyberbullying impacts the social relationship between students in school, and may lead to being a victim. Moreover, bullying mostly happens in the presence of many students who are bystanders around the victim, so this makes this situation worse for the persons that are bullied (Salmivalli et al., 2021). There are some suggested solutions that help prevent or decrease the widespread existence of this phenomenon. The improvement of the knowledge of this phenomenon among students which can be affected by teachers is one of the solutions that help prevent this phenomenon. Also, some programs on the social side can improve the relationship between students, and prepare the new students to realize the risks and the impact of this phenomenon (D. W. Otter et al., 2021).

Technically, some tools can analyze the text data and classify as offensive or non-offensive, negative or positive, and bullying or non-bullying. Natural Language Processing (NLP) is one of the software tools that can be used for interpreting the human language by machine learning algorithms. NLP has many benefits that can help detect the textual data such as sentiment analysis, extract the related information, and categorize for analysis (Lauriola et al., 2021). However, the textual data is not easy to analyze and to extract information from it with high accuracy compared to numeric data. Textual data is complex and not clear when using models of NLP for analyzing and interpreting the human language.

According to some researchers, 32% of the contents of English language texts, are linguistically unclear. This is one of the major challenges facing the NLP to analyze the text data. In addition, this is also true for many other languages used in social media. One example of other language that have difficult to analyze by NLP is Arabic Language that is written from right to left, hinders the model from analyzing the text data. Also, Arabic language has some words that have different meaning which means that some Arabic countries use words for a positive meaning while other Arabic countries use the same words for negative meaning. The model will detect the words as negative text based on the training which may confuse and reduce the performance of the model's results. Thus, this is only one example of the little issues for Arabic language when analyzed by NLP, and there are some other challenges that NLP faces. This is one of the reasons that has hindered the development of NLP and reap the benefits of data

analysis (J. Wang et al., 2020). Nonetheless, it has been proven that NLP is very successful for many applications in recent times. Using NLP in managements of governments, provides many benefits such as citizens' participation during the response to disasters as they happen, and detection and analysis of crimes. Urban governances that use NLP to improve their management, the communication between citizen and mangers of city becomes easy for applications that support NLP (K. Mishev et al., 2020). NLP has many uses in a variety of fields. So when dealing with text data that needs Sentiment Analysis (SA), NLP can support that. SA can be used to identify the feelings of people about specific happenings or events. Also, SA can be used to reflect on a user's emotions about a situation that the user has experienced with reference to the user's comments regarding the situation. As a result, researchers resort to using SA in NLP which can help analyze and interpret a user's behavior based on the comments or tweets. By using the SA, one can detect the information that is relevant to other topics such as predictions for selling or buying products (Cai, M 2020). SA can also define and categorize the opinions of users of the platforms in social media and their expressions and predict their attitudes based on the analysis of their texts. Practically, sentiment analysis can identify the sentiments of users who sell a product whether they are satisfied by classifying their comments as positive, or unsatisfied by classifying their comments as negative. Using SA in NLP can be an optimal solution for detecting cyberbullying (Bharadwaj et al., 2020).

In this paper we have reviewed 11 research papers that identify solutions using machine learning algorithms with Natural Language Processing (NLP) that can detect and mitigate cyberbullying behavior. Also, we have proposed to test three machine learning algorithms which are named Support Vector Machine, Naive Bayes and Decision Tree for detection of cyberbullying text in assigned twitter's dataset.

Literature Review

NLP Detection

NLP is used in many fields to detect the textual data, so fake news is one of the issues that needs to be prevented or reduced by detecting it as real or fake news. Currently, fake news is one of the concerns that lack the real news which makes many people around the world worried in their life as there is an increasing use of social media platforms. In this paper, the authors build a model using machine learning algorithms with NLP to detect the text in the dataset that they used to classify as fake or real news. The dataset used in this model was provided by Kaggle.com which contains 6256 articles labeled as fake or real. After building the model and using multiple algorithms that were used to test the accuracy of detections. Random Forest method gave higher accuracy than Naive Bayes. To process the detection, some steps are executed when building the model starting with text processing which aims to clean the dataset such as tokenization, removing stop words and removing numbers, extracting features of TF & TFIDF, and stemming. In addition, the feature extracting step is used to extract the helpful data from the dataset, and the learning step is to learn the unsupervised algorithm to detect the fake or real news, and the final step is training the model to be able to classify the text data. The result that was achieved when building the model is an accuracy of 95.66% by using the Random Forest algorithm (T. D. Jayasiriwardene & G. U. Ganegoda, 2020).

Authors present improvement to the system by continuously detecting and extracting relevant information from textual data by using NLP. They use Twitter as a platform to extract the keywords that have retrieved relevant news with high performance. In the methodology section of their paper, they used a dataset that was collated by Twitter which contains more than 100,000 records in diverse fields of news such as education, health, politics, and sport. The dataset is considered a part of big data because it involves a large number of records and each record has many words which indicate the size of tweets. Building a model needs a pre-processing stage which helps the model to determine each kind of newspaper. The authors mentioned that they processed many steps to prepare the dataset. The first step removing URLs, punctuations, and emoticons that exist in a tweet also removes some words that do not need to be in the tweet such as stop words and frequency words. The lemmatization is used to transform

the word to the original form which helps the model to analyze the words. Also, it needs to be the spelling correction in the preprocessing stage, so using tokenization by NLTK which is one of Python's libraries the figure 1 shows the preprocessing stage for this paper. Therefore, the result shown in this model by using NLP tools is 67.6% accuracy (C. Sharma et al., 2021).

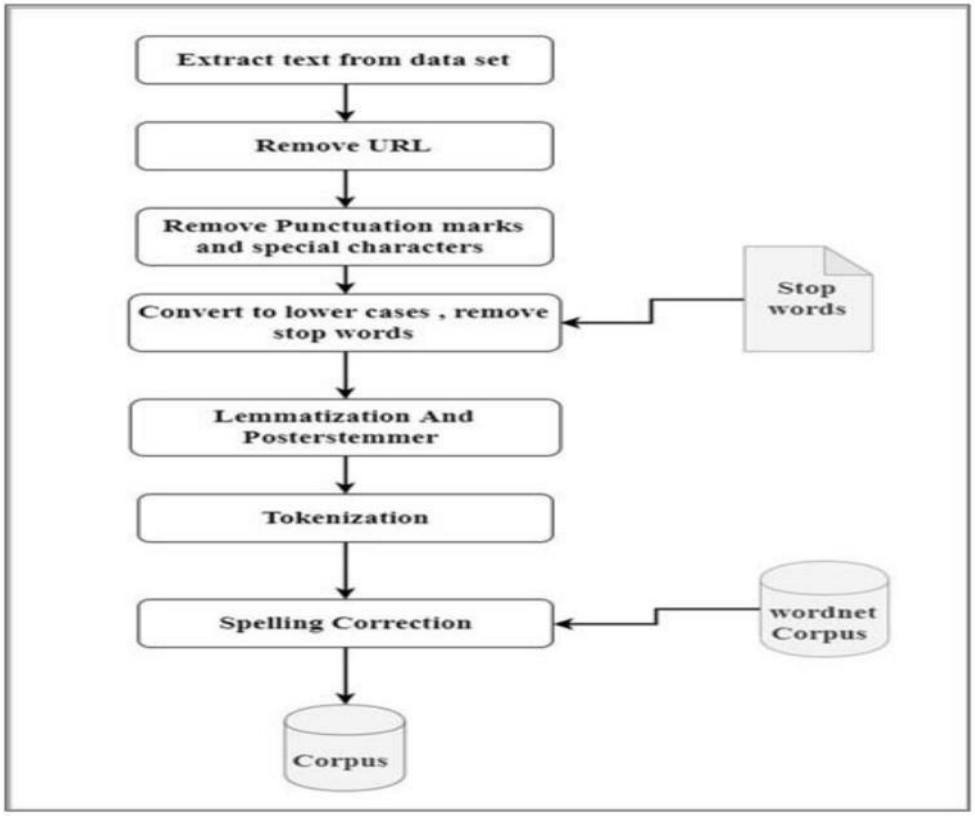

Figure 1: Data pre-processing (C. Sharma et al., 2021).

Cyberbullying Detection Using Machine Learning

As reported in (Jain et al., 2021), researchers developed a model by using Machine learning for the detection of cyberbullying on Twitter as the platform. Firstly, they discussed the risks of cyberbullying on social media and provided some solutions that help reduce this phenomenon. Also, they focus in their study on widespread presence of cyberbullying in schools. They emphasize that when students are bullied, they do not feel safe in school, and their ability for learning is decreased. They emphasize that by showing the results of analyzing the dataset that they chose. After they completed the preprocessing which appears in figure 2, they tested seven machine learning models to achieve the highest accuracy. The best result was found in their test which is 90% accuracy by the Support Vector Machine (SVM) model. As a solution, the authors in suggested cyberbullying detection that can be classified for each row of tweets that exist in their chosen dataset to detect the words that indicate the bullying regarding racism, sexism and hate speech. They use text mining to prove the performance and extract the result using multiple algorithms.

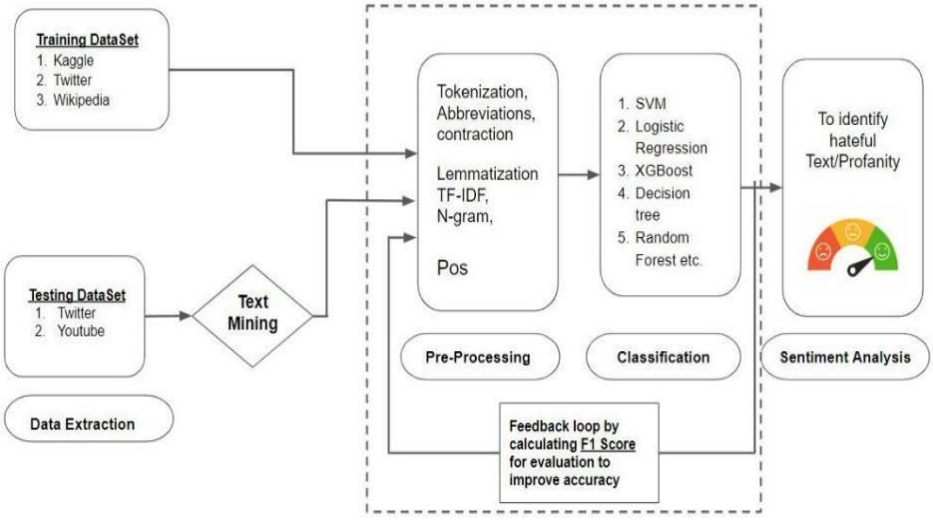

Figure 2: Functional Block Diagram of method followed (Jain et al., 2021).

In addition, machine learning helps predict the behavior of users as most people communicate via social media. The author in (Mahat, 2021) emphasize that social media environment involves a misuse of technologies which can lead to aggression by writing social media comments. Cyberbullying is one of the bad behaviors that needs to be detected and prevented to avoid potential harm to victims. Based on their paper, they created a model that is used to detect the bullying words in the dataset which is large data. As well, they tried many machine learning algorithms for analyzing the dataset, and they used equations that were used to evaluate the accuracy of the model. As they predicate to determine the cyberbullying in social media as follows:

Accuracy = (TP+TN)/(TP+FP+TN+FN)
where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives

After they evaluate many models, the model that gives high accuracy to predict the cyberbullying was Support Vector Machine (SVM).

Offensive Text Detection

Authors in (Hani et al., 2019) divided their research paper into three main parts which are related to cyberbullying. These parts start by discussing cyberbullying as a Cyber-Crime, surveying the cyberbullying issue, and providing research that helps solve this problem. In particular, they developed a model that can classify tweets collected from Twitter as a dataset to be offensive or non-offensive. In addition, the highest accuracy that resulted in their model is 92.9% when using Stochastic Gradient Descent (SGD) Classifier and 92.7% when using Bagging Classifier which gives better results in the F1 Score Test.

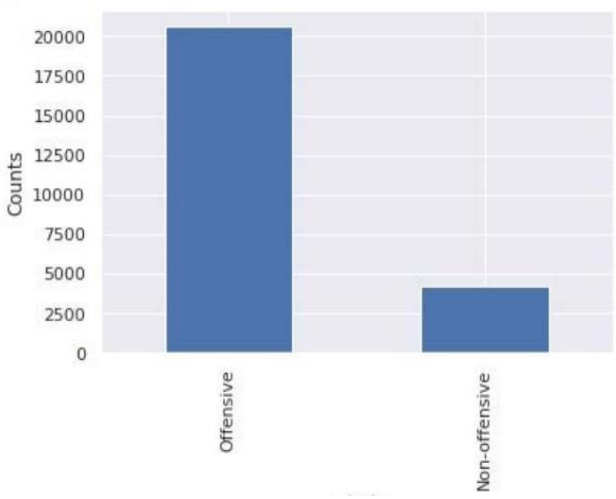

Figure 3: Non-offensive and Offensive Classes Mapping of Dataset Portion-II (Hani et al., 2019).

Based on (Al-Garadi et al., 2019), it was suggested that in their case study, automatic detection of offensive text on social media. They limited the kinds of offensive texts that need to be detected such as fake news, hate speech, etc. as in general disinformation. Thus, they developed a model by using NLP techniques to classify Hierarchical multitask learning which is used to retrieve the relevant information from the document. Often, the information in the documents has other relevant information. In their case study, they show the best solutions that detect the offensive text with a huge amount of data, and this approach can be replicated in any kind of dataset that has disinformation.

Cyberbullying in Social Media

Social media is an attractive environment for cyberbullying in current times. Thus, the authors in show the results of the accuracy of the model that they developed to detect and prevent cyberbullying in social media. The model was built by using unsupervised machine learning algorithms using Neural Network (NN) and Support Vector Machine (SMV). NN achieved 92.8% accuracy while SMV achieved 90.3% accuracy. Based on their statistical findings that include the total words that include bullying and non- bullying (Mangaonkar et al., 2022).

Authors in (Chandra et al., 2019) emphasized in their case study the adverse impact of cyberbullying in social network environments, including harm to the victims as mental health issues. Thus, they highly recommended finding solutions for this phenomenon. The goal of the technical part in this research paper is to provide a real time model that can detect the words related to bullying. The detection model can detect and try to stop the offensive words before data access to the central server in an effort to reduce the time needed to analyze the data. This system uses distributed detection which is divided into many nodes called detection nodes that can utilize Machine Learning algorithms. Thus, each node is considered a server that can detect any tweets that come from clients such as smart devices.

Table 1: Statistics of The Dataset (Chandra et al., 2019).

Total number of Conversations	1608
Number of cyberbullying	804
Number of non-Cyberbullying	804
Number of distinct words	5628
Number of token	48843
Maximum Conversation size	773 Characters
Minimum Conversation size	59 Characters

Cyberbullying Using Deep Learning

Deep learning is one of the models that detect the negative text on some social media platforms. Also, deep learning detects and gives better performance than some other models. The authors in (M. T. Ahmed, et al., 2021) develop a model by using deep learning to detect cyberbullying. They used three datasets on different platforms in the social network, so each dataset contains 3000 examples that are used for detection on three different platforms which are Twitter, Formspring, and Wikipedia. In the data processing stage, they removed stop words, punctuations, and numbers from the dataset. After they processed these steps, they executed the model and get accuracy for each dataset with the highest accuracy being 79.1%.

Table 2: Accuracy results on different datasets (M. T. Ahmed, et al., 2021).

Wikipedia	Twitter	Formspring
75.5%	79.1%	72%

Deep learning and machine learning algorithms are not suitable for all human languages. This means the model can detect and give accuracy for a dataset that contains English language text, but it does not work with other datasets that have another language such as French or Arabic, etc. As each language needs a separate model that can be developed by using machine learning algorithms, the authors in develop a model by using deep learning to detect the cyberbullying for Bengali and Romanized Bangali languages. They used a dataset that contains two languages which are mentioned above, and 12000 records that are mixed between Bangali and Romanized Bangali languages. They built a model that can detect the words related to bullying in social media and processed the preprocessing stage for the dataset to have the dataset cleaned and prepared for classification. They used multiple algorithms to test the dataset and to achieve higher accuracy. The highest accuracy that they achieved is 84% which was achieved by the naive Bayes algorithm (L. Cheng, et al., 2021).

Methodology

In this section, we will describe and demonstrate the implementation of our work in this paper and the process by which the experimental work was conducted. Thus, we have reviewed 11 numbers of prior research papers that have been conducted in detection cyberbullying by NLP. We have summarized each paper based on the dataset that have been used in this paper such as the source, the platforms of social media, and the number of records in the dataset. Also, we referenced the methods that have been used for detection for each paper, so some papers used machine learning

and other used deep learning, and show the accuracies that have achieved for each paper.

Implementation and data collection

This section focuses on the experimental data collection and preparation, model selection and construction, performance evaluation metrics, and results. The various models are compared based on various performance. We obtained a labeled dataset of tweets from a reliable source, kaggle.com. This dataset consists of 47000 tweets that have been labeled according to the type of cyberbullying they contain. The categories include age, ethnicity, gender, religion, other types of cyberbullying, and tweets that do not contain cyberbullying. The data has been evenly distributed such that there are approximately 8000 tweets in each category. In our experiment, there are 39751 tweets that have been processed, the category of other types of cyberbullying not used because it impacts on the accuracy of models and it hasn't categorized to types cyberbullying. Also, there are 250 rows have been found as duplicated. Figure 4 visualized the dataset (J. Wang, et al., 2020).

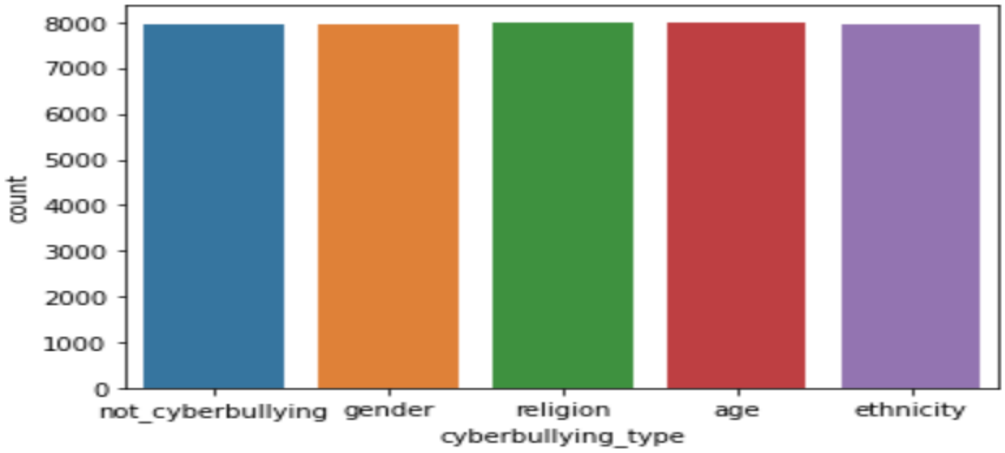

Figure 4: shows labelled distribution

Data preprocessing

Text preprocessing is an important step in the process of building a model. It involves breaking down the text into smaller pieces called tokens, converting all words to lowercase, removing unnecessary words called stop words, reducing words to their base form through stemming, and reducing words to their base form through lemmatization. After clean the text, the dataset need to be split to taring set and testing set to perform the model. In order to input it into a machine learning model, it need converting text data into numerical data which can be doing by feature extraction which called Vectoring the text. These techniques are commonly used to simplify and reduce the complexity of the text, which helps to improve the model's performance. Figure 5 show the experimental work.

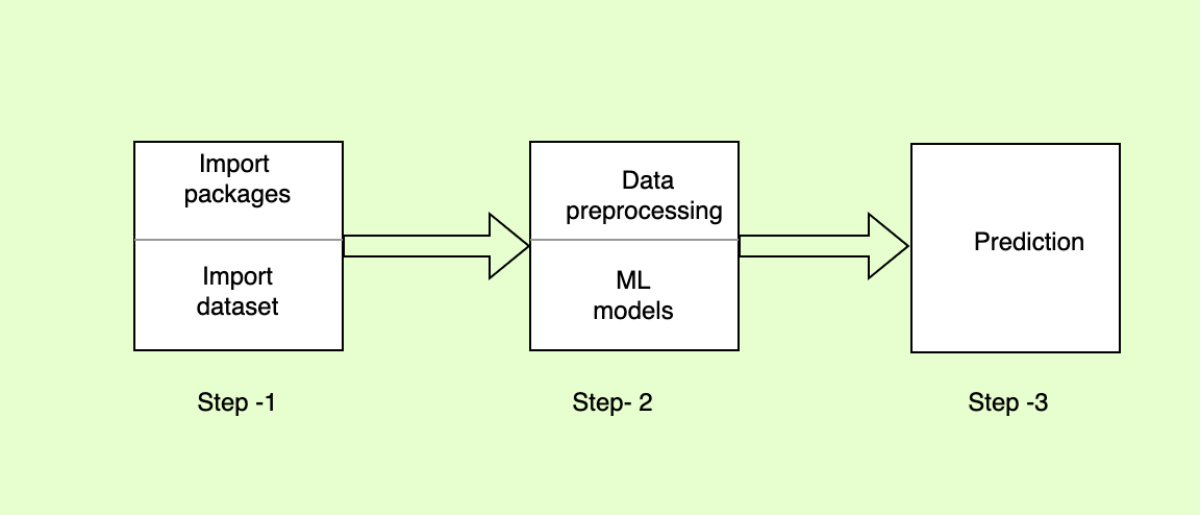

Figure 5: shows the experimental work

The first step involves importing necessary packages and accessing the data sources. The second step involves selecting and preparing the data for machine learning models. The final step involves using the trained model to make predictions and deploy it for use in detecting and classifying the desired categories with a satisfactory level of accuracy.

Model Selection

Three machine learning classifiers (as assigned below) were selected and tested on prepared data using a pipeline method, in which each of the following classifiers was applied one after the other:

- Support Vector Machine
- Naive Bayes
- Decision Tree

Performance Metrics

Several performance metrics were calculated and analyzed. These include accuracy, precision, recall, and F1 score. These metrics are commonly used to evaluate the performance of a classifier in a classification task. The results of these metrics were observed for both the training and testing datasets.

Accuracy- (TN + TP)/(TN+TP+FN+FP) It is calculated as the number of true positive (TP) and true negative (TN) predictions divided by the total number of predictions made (TP + TN + false positive (FP) + false negative (FN)).

Precision-(TP)/(TP+FP) It is calculated as the number of true positive predictions divided by the total number of positive predictions made (TP + FP).
Recall-(TP)/(TP+FN) It is calculated as the number of true positive predictions divided by the total number of actual positive instances in the data (TP + FN).

F1 Score- 2*((precision*recall)/(precision+recall)) It is calculated as the product of precision and recall divided by the sum of precision and recall (V. Jain, et al., 2021).

Results and Discussion

For our review, some papers suggest using deep learning for large dataset, and other papers suggest using machine learning for adequately sized dataset. Also, machine learning takes short time

for detection while deep learning need more time. Some papers discussed the challenges that faced to detect other languages such as Arabic and French. The results of our experiment achieved by three models are shown in Table 3.

Table 3: Summary of results using three models.

Algorithm	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Support Vector Machine	0.93	0.93	0.93	0.93
Decision Tree	0.91	0.92	0.91	0.91
Naive Bayes	0.84	0.85	0.84	0.83

Challenges That Faced to Detect Cyberbullying

While cyberbullying is a concern in recent times, it is not easy to build models that detect and prevent this issue. Many challenges are faced by researchers and developers when building or improving models for the detection of cyberbullying in social media. While social media environment involves the data that help describe human behavior, the data has massively increased which leads to the use of big data analysis. To analyze big data, it requires high performance for tools and consumes a long time to execute the preprocessing steps. Moreover, the newest data is not labeled, so this step is one of the preprocessing stages and it is difficult and impacts the accuracy of results. In addition, the language change is considered a challenge, meaning each generation has different idioms. When the meaning of words changes the model will detect based on its training dataset (Al-Garadi et al., 2019).

Conclusions

In conclusion, this paper discusses in the first section the risks of cyberbullying in many environments especially in the school. Also, the paper reviews some of the technical solutions that are mentioned in some studies which help realize the damages caused by cyberbullying. NLP is one of these technical solutions and presents the benefits of applying NLP in many applications. In the second section, this paper reviews many of the prior research work that discussed the issue and provides some technical solutions using some tools. Thus, this section is divided into three parts discussing tools that are used for detecting cyberbullying on many platforms in social media. NLP is used in many examples to detect, predict and analyze text data, and provide high accuracy of the models that are used to detect a cyberbullying or offensive text. Using deep learning is also an optimal tool to detect the positive or negative text. However, it needs a high-performance computer to analyze big data. Moreover, in this paper, we proposed a machine learning method for detecting cyberbullying. We tested our model using three different classifiers: Support Vector Machine, Naive Bayes, and Decision Tree. To extract features, we used TF-IDF and sentiment analysis. Our results showed that the Support Vector Machine classifier had an accuracy of 0.93%, while the Decision Tree classifier had an accuracy of 0.91% and the Naive Bayes classifier had an accuracy of 0.84%. In addition, this paper proposed to review 11 papers that have been examined for detection cyberbullying by summarizing their mothed for detection and their results.

References

Chun, J., Lee, J., Kim, J., & Lee, S. (2020). An international systematic review of cyberbullying measurements. *Computers in human behavior, 113, 106485.*

López-Meneses, E., Vázquez-Cano, E., González-Zamar, M. D., & Abad-Segura, E. (2020). Socioeconomic effects in cyberbullying: Global research trends in the educational context. *International journal of environmental research and public health, 17(12), 4369.*

Faucher, C., Cassidy, W., & Jackson, M. (2020). Awareness, policy, privacy, and more: Post-secondary students voice their solutions to cyberbullying. *European Journal of Investigation in Health, Psychology and Education, 10(3), 795-815.*

Salmivalli, C., Laninga-Wijnen, L., Malamut, S. T., & Garandeau, C. F. (2021). Bullying prevention in adolescence: solutions and new challenges from the past decade. *Journal of research on adolescence, 31(4), 1023-1046.*

D. W. Otter, J. R. Medina and J. K. Kalita, (2021). A Survey of the Usages of Deep Learning for Natural Language Processing," in *IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 2, pp. 604-624, Feb. 2021, doi: 10.1109/TNNLS.2020.2979670*.

Lauriola, I., Lavelli, A., & Aiolli, F. (2022). An introduction to deep learning in natural language processing: models, techniques, and tools. *Neurocomputing, 470, 443-456.*

Cai, M. (2021). Natural language processing for urban research: A systematic review. *Heliyon, 7(3), e06322.*

K. Mishev, A. Gjorgjevikj, I. Vodenska, L. T. Chitkushev and D. Trajanov, (2020). "Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers," in *IEEE Access, vol. 8, pp. 131662- 131682.*

Bharadwaj, Pranav and Shao, Zongru, (2019). Fake News Detection with Semantic Features and Text Mining *International Journal on Natural Language Computing (IJNLC) Vol.8, No.3, June 2019, Available at SSRN: https://ssrn.com/abstract=3425828*

T. D. Jayasiriwardene and G. U. Ganegoda, (2020). Keyword extraction from Tweets using NLP tools for collecting relevant news, *2020 International Research Conference on Smart Computing and Systems Engineering (SCSE), 2020, pp. 129-135, doi: 10.1109/SCSE49731.2020.9313024.*

C. Sharma, R. Ramakrishnan, A. Pendse, P. Chimurkar and K. T. Talele, (2021). Cyber-Bullying Detection Via Text Mining and Machine Learning, *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2021, pp. 1-6, doi: 10.1109/ICCCNT51525.2021.9579625.*

Jain, V., Saxena, A. K., Senthil, A., Jain, A., & Jain, A. (2021, December). Cyber-Bullying Detection in Social Media Platform using Machine Learning. *In 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART) (pp. 401-405). IEEE.*

Hani, J., Nashaat, M., Ahmed, M., Emad, Z., Amer, E., & Mohammed, A. (2019). Social media cyberbullying detection using machine learning. *Int. J. Adv. Comput. Sci. Appl, 10(5), 703-707.*

Mangaonkar, A., Pawar, R., Chowdhury, N. S., & Raje, R. R. (2022). Enhancing collaborative detection of cyberbullying behavior in Twitter data. Cluster Computing, 1-15.

Chandra, S., & Das, B. (2022). An approach framework of transfer learning, adversarial training and hierarchical multi-task learning-a case study of disinformation detection with offensive text. *In Journal of Physics: Conference Series (Vol. 2161, No. 1, p. 012049). IOP Publishing.*

Al-Garadi, M. A., Hussain, M. R., Khan, N., Murtaza, G., Nweke, H. F., Ali, I., ... & Gani, A. (2019). Predicting cyberbullying on social media in the big data era using machine learning algorithms: review of literature and open challenges. *IEEE Access, 7, 70701-70718.*

M. Mahat, (2021). Detecting Cyberbullying Across Multiple Social Media Platforms Using Deep Learning, *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2021, pp. 299-301, doi: 10.1109/ICACITE51222.2021.9404736.*

M. T. Ahmed, M. Rahman, S. Nur, A. Islam and D. Das, (2021). Deployment of Machine Learning and Deep Learning Algorithms in Detecting Cyberbullying in Bangla and Romanized Bangla text: A Comparative Study, *2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), 2021, pp. 1-10, doi: 10.1109/ICAECT49130.2021.9392608.*

L. Cheng, Y. N. Silva, D. Hall and H. Liu, (2020). Session-Based Cyberbullying Detection: Problems and Challenges, *in IEEE Internet Computing, vol. 25, no. 2, pp. 66-72, 1 March-April 2021, doi:*

10.1109/MIC.2020.3032930.

Ali, S., AL ADWAN, M. N., QAMAR, A., & HABES, M. (2021). Gender discrepancies concerning social media usage and its influences on students academic performance. *Utopía y Praxis Latinoamericana, 26(1), 321-333.*

Husain, F., & Uzuner, O. (2021). A survey of offensive language detection for the arabic language. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP), 20(1), 1-44*.

J. Wang, K. Fu, C.T. Lu, (2020). SOSNet: A Graph Convolutional Network Approach to Fine-Grained Cyberbullying Detection, *Proceedings of the 2020 IEEE International Conference on Big Data (IEEE BigData 2020), December 10-13, 2020.*

V. Jain, A. K. Saxena, A. Senthil, A. Jain and A. Jain, (2021). Cyber-Bullying Detection in Social Media Platform using Machine Learning, *2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART), 2021, pp. 401-405, doi: 10.1109/SMART52563.2021.9676194.*