Cyberbullying Detection: A Machine Learning Approach using Social Media Data

Mohammad Zakariya
School Of Computer Application
Babu Banarasi Das University, Lucknow,India

Dr. Nupur Soni
Associate Professor
School Of Computer Application
Babu Banarasi Das University, Lucknow,India

Abstract- In recent years, the prevalence of cyberbullying on social media platforms has grown to such an extent that automated techniques to identify and stop it have become necessary. This study focuses on developing a machine learning-based model that uses text classification techniques to identify occurrences of cyberbullying in social media content. Using a dataset acquired from Kaggle that includes labeled social media data especially associated with different types of cyberbullying, the study uses a supervised learning approach. After preprocessing the dataset to eliminate noise and ensure uniformity, the Term Frequency-Inverse Document Frequency (TF-IDF) approach is used to extract features. Several machine learning algorithms, including Support Vector Machines (SVM), Decision Trees, Random Forests, and Naive Bayes, are trained and evaluated using standard classification metrics, such as accuracy, precision, recall, and F1-score. The experimental results show that the SVM model achieves the highest accuracy of 83%, outperforming the other algorithms in classifying both cyberbullying and non-cyberbullying content. The study's findings demonstrate the potential of machine learning techniques in combating cyberbullying on social media by automatically identifying harmful content, thus contributing to creating safer online spaces. The research also highlights challenges in handling imbalanced datasets and the need for further improvements in model performance.

Keywords- Cyberbullying Detection, Machine Learning, Social Media, Text Classification, Supervised Learning, Natural Language Processing (NLP), Term Frequency-Inverse Document Frequency (TF-IDF), Support Vector Machines (SVM), Decision Trees, Random Forests, Naive Bayes, Data Preprocessing, Feature Extraction, Model Evaluation, Imbalanced Dataset.

I. INTRODUCTION

The rise of social media sites such as Facebook, Twitter, and MySpace in the early 2000s marked the beginning of a new age associated with online communication.

On these sites, users may create profiles, submit updates, and communicate with individuals worldwide. Unlike earlier digital communication techniques, social media gave community building and user-generated

content a lot of weight. Over time, platforms added features like live streaming, multimedia sharing, and algorithm-driven content customization.Social media is a platform that enables users to engage with society and submit anything, including documents, videos, and images[1]. People access social media on their computers or smartphones. Facebook, Instagram, TikTok, Twitter, and other services are among the most popular.These days, social media is used for a variety of purposes, including education [2], business [3], and charitable causes[4]. Furthermore, social media is boosting the world economy by creating a lot of new job opportunities [5].Social media has grown at an exponential rate, with sites such as Instagram, TikTok, and Snapchat attracting millions of users in just a few years of launch. As the main platform for news, entertainment, education, and self-expression, social media is now intricately woven into everyday life. However, users are also at danger for additional threats like cyberbullying as a result of this widespread usage.Bullying has long been a social problem that has historically taken place in physical areas including communities, businesses, and schools. In the past, bullying frequently took the form of verbal abuse, physical threats, or social marginalization. Even if these actions were detrimental, they were usually limited to particular places and periods of time.

Abusive behavior became a more widespread problem as a result of the transition to online communication communication. With the rise of social media, cyberbullying—defined as the use of electronic communication to harass or cause harm to others—became a serious issue. Cyberbullying, in contrast to conventional bullying, is not restricted by time or location. The impact on victims is increased by the instantaneous and permanent dissemination of harmful posts, messages, or comments.Cyberbullying started to increase as social media platforms became more widely used. Because of the anonymity and reach offered by digital platforms, criminals can target people without worrying about the repercussions right away.

Social media platforms are become a big part of our daily life. We use social media for a number of purposes, including employment, education, entertainment, and personal growth. Social media is becoming increasingly significant in our daily lives due to the internet's and technology's quick advancements[6].The prevalence of cyberbullying among teenagers as a victim, offender, or bystander is very high[7]. According to Hinduja and Patchin, 36.5% of students have at some point in their lives been the victim of cyberbullying. This indicates that the most common form of online comments were those that were harmful or mean [8].

According to a poll of 1,501 kids between the ages of 10 and 17 conducted in the USA, 12% of them acknowledged abusing someone online, 4% reported being the target of aggression, and 3% reported being both the victim and the bully [9]. In America, about 50% of teenagers report having been the victim of cyberbullying[10]. The victim of bullying experiences both mental and physical effects [11]. Because of the pain of cyberbullying, which is difficult to bear, the victims decide to commit self-destructive behaviors like suicide [12]. Therefore, it's critical to recognize and stop cyberbullying in order to safeguard adolescents.

Given the negative effects that cyberbullying has on its victims, appropriate measures to identify and subsequently stop it are desperately needed. Machine learning is one of the effective methods that uses data to create a model that classifies appropriate actions. Machine learning can help detect bullies' linguistic patterns and hence develop a model for detecting cyberbullying acts.In the digital age, the prevalence of cyberbullying is a serious social problem that necessitates creative solutions to safeguard people from harm while maintaining ethical standards. There is great potential for detecting and preventing cyberbullying through the use of machine learning techniques. By using sophisticated algorithms to efficiently evaluate social media data, this study seeks to close the gaps in the detection techniques now in use. The ultimate objective of these initiatives is to help build safer online environments where people can communicate without worrying about being harassed or abused.

II. RELATED WORK

Cyberbullying has become a major concern in the digital era, particularly among teenagers who are more engaged on social media platforms. Researchers are investigating a number of methods for detecting cyberbullying due to its prevalence, with a particular emphasis on machine learning and deep learning approaches. Key studies in the topic are summarized in this review of the literature, with an emphasis on their methods, conclusions, and limitations.

Enhanced Multi-label Classification Model

The improved multi-label classification model presented by Indumathi and Santhana Megala (2023) is designed to identify cyberbullying in text data from social media. By correctly recognizing different types of bullying in literature, this model marks a substantial improvement in managing the complexity of multi-label classification tasks. The study demonstrates significant gains in classification accuracy when compared to conventional models, highlighting the effectiveness of their methodology.

Advanced feature engineering approaches and the use of deep learning architectures, which allow the model to capture subtle language patterns linked to cyberbullying, are credited by the authors with this result. But they also draw attention to a number of difficulties. First, in real-time applications where speed and computing efficiency are crucial, the model's complexity presents a major obstacle to deployment. Furthermore, in order to operate at its best, the model depends largely on large, excellent training datasets, which aren't always easily accessible.

Despite these drawbacks, the study offers insightful information about the use of multi-label classification in the context of cyberbullying detection and identifies directions for further investigation, including improving the model for quicker inference and investigating semi-supervised learning techniques to lessen dependency on sizable datasets. The continued development of AI-driven solutions to address online abuse gains a solid knowledge from this work[13].

BERT-Based Detection Model

Guo, X., Anjum, U., and Zhan, J. (2022) utilized BERT, a state-of-the-art Bidirectional Encoder Representations from Transformers model, for cyberbullying detection. Their research highlights BERT's exceptional ability to understand contextual nuances in text, which makes it particularly effective in identifying subtle and complex instances of cyberbullying.

The findings demonstrate significant improvements in detection rates compared to traditional approaches, such as Support Vector Machines (SVMs), logistic regression, and earlier neural network models.

The authors conducted experiments across various datasets, showing the model's capability to adapt to diverse social media platforms like Twitter and Reddit. However, they also noted several challenges. The BERT-based model requires considerable computational resources, including powerful hardware and extended processing times, which may limit its accessibility for widespread application. Furthermore, the model's performance was found to be influenced by the quality and composition of the dataset used for training, raising concerns about generalizability across different contexts and languages.

To address these challenges, the study emphasizes the potential benefits of fine-tuning BERT for specific cyberbullying contexts, such as accommodating cultural differences, linguistic variations, and the dynamic nature of online behaviors. Guo et al. (2022) propose that future research should focus on optimizing computational efficiency and exploring transfer learning techniques to make these models more practical for deployment. This research underscores the critical role of advanced deep learning technologies like BERT in advancing the fight against cyberbullying while identifying areas that require further exploration and improvement [14].

Multi-modal Cyberbullying Detection on Social Networks

Wang et al. (2020) proposed a multi-modal identification system that combines many forms of information, such as text, photos, and user interactions, to handle the complexities of modern cyberbullying. Their study highlighted how crucial it is to document the social context of cyberbullying occurrences, which frequently involve a variety of media and user interaction. The authors stated that by using hierarchical attention networks to process and interpret the multi-modal data, their model showed significant improvements in detection accuracy. Despite these developments, the system's capacity to detect in real time may be limited by the higher computational requirements of handling heterogeneous data, which would restrict its ability in hectic social media environments[15].

Feature Analysis for Cyberbullying Detection

Mahmud et al. (2022) conducted a thorough investigation of textual features to improve the effectiveness of machine learning models in identifying cyberbullying. The study methodically investigated how different linguistic and semantic factors contribute to higher categorization accuracy. The study found important characteristics, including word frequency, sentiment scores, and context-specific embeddings, that were highly predictive of cyberbullying material by examining textual patterns and linkages.

The results showed that models that included these crucial features performed better than those that used generic feature sets, underscoring the significance of careful feature selection. The study also demonstrated how linguistic cues and sophisticated natural language processing methods, such TF-IDF, can be combined to better capture the nuances of harmful online conduct.

Mahmud et al.'s contribution is that they offer researchers actionable insights on how to enhance machine learning models for detecting cyberbullying. The study laid the groundwork for creating more accurate and effective models for managing harmful online interactions by identifying and ranking features that improve detection accuracy[16].

Automatic Identification of Cyberbullying in Social Media Through the Use of an SVM-Activated Stacked Convolution LSTM Network

Buan and Ramachandra(2020) used a Support Vector Machine (SVM) activated stacked convolution Long Short-Term Memory (LSTM) network to create an automated cyberbullying detection model. By combining the advantages of LSTM layers for sequence modeling and convolutional layers for feature extraction, this novel method enabled the model to successfully detect patterns suggestive of cyberbullying in textual material. The authors noted increased effectiveness in identifying cyberbullying, especially on social media, where the rapid flow of information can make identification more difficult. However, the model's complexity can restrict its interpretability and accessibility, making it difficult to apply in the actual world without specialized resources and skills. This emphasizes the necessity of tools and user-friendly interfaces that can make it easier to apply such advanced models in real-world situations[17].

Automated Deep Learning Model Development

Jin et al. (2023) studied how AutoML libraries can help with the building of deep learning models for cyberbullying detection. AutoKeras, a tool for automating the designing and training of deep learning architectures, was used in their work. Researchers could quickly develop and assess neural network models by utilizing AutoKeras, which eliminated the need for in-depth understanding of deep learning methodologies.The methodology showed that AutoKeras may achieve similar performance to manually designed deep learning models while drastically reducing the complexity of model construction. The tool made it easier for researchers of different skill levels to choose hyperparameters, neural network layers, and optimization techniques.

The study demonstrated how AutoML frameworks might encourage approachable methods for complex machine learning tasks.

The dependence on automated pipelines, however, could restrict customization for certain activities that call for domain-specific modifications. The study of Jin et al. made a contribution by providing a workable way for scholars and practitioners to successfully implement deep learning techniques, especially in fields like cyberbullying detection where adoption may be hampered by technological obstacles[18].

Research Gaps

Even though cyberbullying detection techniques have advanced significantly, there are still a number of research gaps that need to be filled:

- **Limited Modalities in Detection**:

Text-based analysis for cyberbullying identification is the main focus of the majority of current research. On the other hand, cyberbullying can take many different forms on a variety of platforms, such as voice, video, and photos. Research on multi-modal detection methods that can combine and examine several data kinds is required in order to increase the precision and thoroughness of detection.

- **Contextual Understanding**:

The context in which cyberbullying takes place is frequently not understood by current models. Sarcasm, cultural jokes, and social dynamics are just a few examples of the the nuances of language that can have a big impact on how messages are understood. In order to better comprehend interactions on social media platforms, future study should try to include contextual elements and social aspects.

- **Psychological and Emotional Factors**:

Although some research have started to include psychological elements like personality traits and emotional states, comprehensive models that make good use of these components are still lacking. Improved identification and intervention techniques may result from a better understanding of the psychological effects of cyberbullying on victims as well as the motivations of offenders.

- **Motivations of Perpetrators**:

Comprehending the motivations behind cyberbullying is equally crucial. Peer pressure, cultural dynamics, or personal concerns are some of the reasons why someone can harass others online. By including these factors into detection models, researchers can develop more advanced techniques that address the root causes of bullying behavior in addition to identifying it.This could result in preventive actions.

III. METHODOLOGY

This section outlines the systematic approach adopted for developing the automated cyberbullying detection model. The methodology involves several stages, including data collection, preprocessing, feature extraction, model selection and training, and model evaluation. Below is a detailed description of each stage.

A. Data Collection

The dataset utilised in this study was obtained from Kaggle, a platform that offers publicly available datasets. It is made up of text data that was originally collected from social media platforms like Twitter and Facebook and labelled with cyberbullying detection-related categories. This dataset was appropriate for the goals of this study because it guaranteed diversity and relevance.

The pre-labeled nature of this dataset, which simplifies the supervised learning process, and its applicability to actual cyberbullying detection cases led to its selection. The dataset's structure and size allowed it to be used for training and assessing a variety of machine learning models, giving this work a strong basis.

B. Natural Language Processing (NLP)

As part of the NLP pipeline, several important procedures were conducted to prepare the textual data for machine learning analysis. Among these are feature extraction and data preprocessing, which are explained below:

1) Data Preprocessing

Before applying machine learning models, the textual data was preprocessed to assure consistency and quality. The actions listed below were taken:

- **Character Removal:** To reduce noise, special characters, punctuation, and digits were eliminated.
- **Lowercasing**: To maintain uniformity, the text was transformed to lowercase.
- **Tokenization**: In order to analyze individual word patterns, sentences were divided into words, or tokens.
- **Stop-Word Removal**: Words that are often used but have little contextual meaning, such as "the" and "is," were not included.
- **Lemmatization**: In order to unify variations, words were reduced to their base or root forms (e.g., "running" and "run").

By following these procedures, the text data was guaranteed to be homogeneous, clean, and appropriate for feature extraction and analysis.

2) Feature Extraction

Term Frequency-Inverse Document Frequency (TF-IDF), one of the Natural Language Processing (NLP) approaches utilized in this study, was utilized to convert textual data into numerical vectors appropriate for machine learning models. By giving each phrase a weight that represents its significance within a document in relation to its occurrence throughout the full dataset, TF-IDF works.

While the inverse document frequency (IDF) component lessens the weight of commonly used terms that occur often across several documents, the term frequency (TF) component counts the number of times a word appears in a particular document. By highlighting important textual elements and reducing the influence of high-frequency, less informative words, this strategy improves the representation of important and distinguishing phrases. The textual data was successfully converted into a format that preserved significant linguistic patterns by utilizing TF-IDF, which made it easier for the model to find relevant features for the identification of cyberbullying.

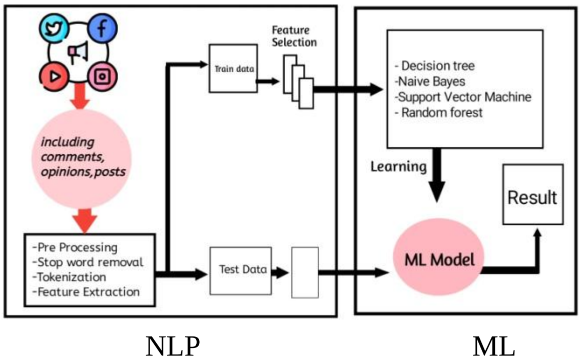

Fig.1. Proposed framework for cyberbullying detection

C. Model Selection and Training

For this study, four machine learning algorithms were chosen: Naive Bayes, Random Forests, Decision Trees, and Support Vector Machines (SVM). To categorize textual data as either cyberbullying or non-cyberbullying, each system was trained using the TF-IDF characteristics. A full discussion of these algorithms and their use cases is provided below.

- **Support Vector Machines (SVM):**

SVM is a supervised learning technique that determines the best hyperplane for classifying data points. It is useful for text categorization problems and performs well in high-dimensional areas.SVM is best suited for datasets that are well-structured and require high precision and recall. It works especially well with textual data because of its capacity to manage high-dimensional feature spaces.

- **Decision Trees:**

A decision tree is a rule-based model that creates a tree-like structure by dividing the data into subsets according to feature values. A class label is represented by each leaf node, and a characteristic is represented by each internal node.

Decision trees work well with datasets where interpretability is important because of their tree form, which makes it simple to see and comprehend how decisions are made.

- **Random Forests:**

Random Forest is an ensemble learning technique that builds numerous decision trees during training and then aggregates their results to increase classification accuracy. By averaging several models, it reduces overfitting.

Random Forests are good for complicated, high-variance datasets because they improve robustness and generalization over a single decision tree.

- **Naive Bayes:**

Naive Bayes is a probabilistic classifier that relies on Bayes' Theorem. It is computationally efficient since it presumes feature independence.When it comes to text classification jobs where features (like words) are independent, Naive Bayes works well. It is renowned for being quick and easy to use, and it performs well with small to medium-sized datasets.

D. Model Evaluation

A crucial first step in assessing machine learning models' effectiveness, robustness, and suitability in detecting cyberbullying in social media content is their evaluation. The evaluation metrics and methods used to assess the trained models' performance are covered in this section.

Evaluation Metrics

A number of common categorization criteria were used in order to thoroughly assess the models' performance. These metrics ensure a balanced evaluation by measuring various aspects of the models' performance:

- **Accuracy**:

Accuracy measures the model's overall correctness by comparing the number of properly predicted occurrences to the total number of instances. It is computed as follows:

$$\text{Accuracy} = \frac{\text{TP} \; + \; \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}}$$

Where:

- TP: True Positives (correctly identified cyberbullying instances)
- TN: True Negatives (correctly identified non-cyberbullying instances)
- FP: False Positives (non-cyberbullying instances misclassified as cyberbullying)
- FN: False Negatives (cyberbullying instances misclassified as non-cyberbullying)

- **Precision**:

Precision provides a measure of relevance by calculating the percentage of accurately predicted positive cases among all instances classed as positive. Reducing false alarms is especially crucial when it comes to cyberbullying detection. The precision formula is as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

- **Recall:**

Recall, also known as sensitivity, assesses the model's ability to recognize genuine positive situations by computing the proportion of correctly identified positives among all true positives. It is very important for identifying every case of cyberbullying. The recall formula is:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- **F1-Score**:

The F1-score, which is the harmonic mean of these two metrics, can help balance the trade-off between precision and recall.

When the dataset is imbalanced, as is frequently the case with cyberbullying detection, it is extremely helpful. The equation is:

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

These metrics were generated for each algorithm to establish the best-performing model for identifying cyberbullying. Comparative analysis assisted in figuring out the algorithm that achieved the highest accuracy and reliability in differentiating between cyberbullying and non-cyberbullying content.

IV. RESULT

A. Data Description

The dataset was sourced from Kaggle and included 47,692 tweets about cyberbullying and its associated categories. To enhance model performance and guarantee text consistency, the dataset was preprocessed.

B. Preprocessing and Feature Extraction

The preprocessing steps included removing irrelevant characters, converting the text to lowercase, tokenization, stop word removal, and lemmatization. The preprocessed text data was then used to extract features using the Term Frequency-Inverse Document Frequency (TF-IDF) approach. The vocabulary size was set to 5000, limiting the feature space to the most common and informative terms.

C. Model Training and Evaluation

The preprocessed data was used to train and assess four classification models: Support Vector Machine (SVM), Decision Tree, Random Forest, and Naive Bayes. The dataset was divided into training and testing sets in an 80/20 ratio to ensure a reliable evaluation of model generalization.

D. Model Performance Comparison

The performance of each model was evaluated using common measures such as precision, recall, F1-score, and accuracy.

- **SVM Results:**

The Support Vector Machine (SVM) model obtained an overall accuracy of 83%. The detailed performance metrics for each class are shown in Table 1.

Table 1. Classification Report for SVM

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Age	0.96	0.97	0.97	1603
Ethnicity	0.99	0.97	0.98	1603
Gender	0.90	0.83	0.86	1531
Not Cyberbullying	0.61	0.52	0.56	1624
Other Cyberbullying	0.59	0.73	0.65	1612
Religion	0.96	0.94	0.95	1566

- **Decision Tree Result:**

Overall accuracy for the Decision Tree model was 79%. While it performed well in the "Age" and "Ethnicity" categories, it did not perform well in "Not Cyberbullying" and "Other Cyberbullying." The detailed performance metrics for each class are shown in Table 2.

Table 2. Classification Report for Decision Tree

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Age	0.97	0.96	0.97	1603
Ethnicity	0.99	0.96	0.97	1603
Gender	0.84	0.83	0.83	1531
Not Cyberbullying	0.49	0.49	0.49	1624
Other Cyberbullying	0.52	0.55	0.54	1612
Religion	0.93	0.93	0.93	1566

- **Random Forest Result:**
  The overall accuracy of the Random Forest model was 82%. It performed well in the majority of classifications, especially "Age"
  and "Ethnicity." The detailed performance metrics for each class are shown in Table 3.

Table 3. Classification Report for Random Forest

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Age	0.98	0.97	0.98	1603
Ethnicity	0.98	0.98	0.98	1603
Gender	0.88	0.84	0.86	1531
Not Cyberbullying	0.57	0.50	0.53	1624
Other Cyberbullying	0.56	0.67	0.61	1612
Religion	0.95	0.96	0.95	1566

- **Naive Bayes Result:**
  The overall accuracy of the Naive Bayes model was 77%. It did well on "Age" and "Religion," but it did poorly on "Not Cyberbullying" and "Other Cyberbullying." The detailed performance metrics for each class are shown in Table 4.

Table 4. Classification Report for Naive Bayes

Class	Precision	Recall	F1-Score	Support
Age	0.79	0.96	0.87	1603
Ethnicity	0.87	0.91	0.89	1603
Gender	0.78	0.83	0.80	1531
Not Cyberbullying	0.67	0.43	0.52	1624
Other Cyberbullying	0.63	0.57	0.60	1612
Religion	0.83	0.96	0.89	1566

To better understand the differences in performance, Figure 1 displays the overall accuracy for each model. Figure 2 displays the F1-scores of all models for each category (e.g., Age, Gender, Ethnicity). Figure 3 displays the variation in accuracy, precision, recall, and F1-score across models.

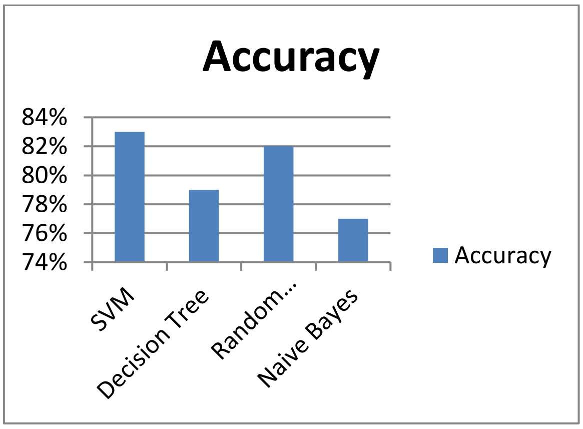

Fig.2. Accuracy Comparison Across Models

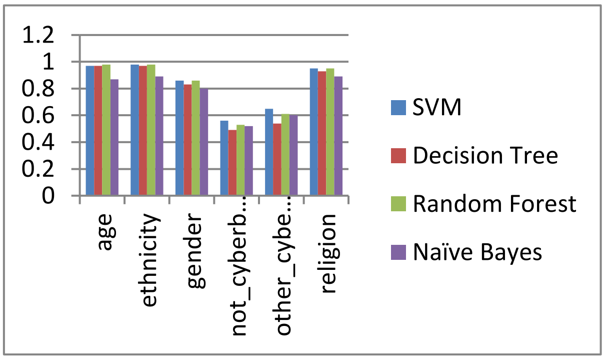

Fig.3. Compare the F1-scores of all models for each category (e.g., Age, Gender, Ethnicity,Other_cyberbullying,not_cyberbullying,religion).

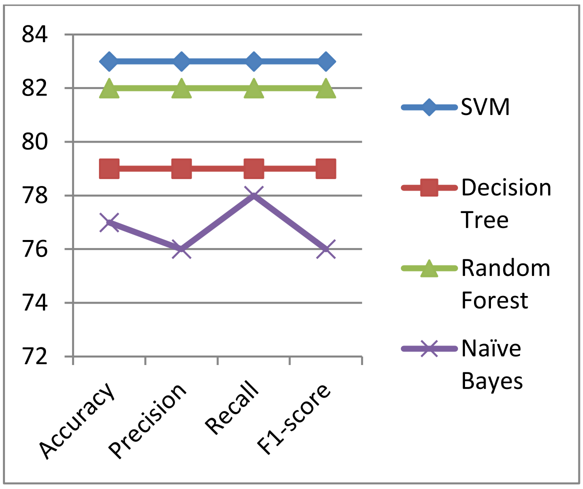

Fig.4. Show the variation in accuracy, precision, recall, and F1-score across models.

V. CONCLUSION

This study emphasizes the importance of machine learning in addressing the growing issue of cyberbullying in the digital age. Through the analysis of publicly accessible social media data from sites such as Facebook and Twitter, the study created and assessed machine learning models to efficiently identify cyberbullying.

The suitability of four models—Support Vector Machine (SVM), Random Forest, Decision Tree, and Naive Bayes—for recognizing different forms of cyberbullying was evaluated. With accuracies of 83% and 82%, respectively, SVM and Random Forest were the best performers. SVM performed best in well-defined categories, displaying strong and balanced performance, whereas Random Forest excelled in nuanced categories with complicated linguistic patterns.

Due to overfitting, the Decision Tree model suffered with ambiguous categories but excelled in precise ones like age and race. Despite being computationally efficient, Naive Bayes performed the worst overall and had trouble capturing contextual dependencies.

Despite the limitations of individual models, the study highlights how machine learning techniques can improve the identification of cyberbullying, providing insightful information and advancing ongoing research on online harassment prevention.

VI. FUTURE SCOPE

While the results of this study are promising, there are several avenues for future research to further improve the detection of cyberbullying on social media. First, addressing the issue of imbalanced datasets is critical because it affects the accuracy and reliability of classification models. Techniques such as oversampling, undersampling, or using more advanced models like deep learning could potentially enhance the detection process.

Incorporating multimodal data, such as photographs, videos, and user behavior patterns, in addition to textual data, could result in more effective and comprehensive cyberbullying detection systems. To increase classification accuracy and generalization, future research might potentially investigate the use of transfer learning or pre-trained language models like BERT or GPT, which have demonstrated good performance in tasks involving natural language processing.

Real-time cyberbullying detection is another area that needs work, as this would call for more scalable systems and better processing speeds. Last but not least, it might be investigated to include the detection model into social media platforms for real-time content moderation, guaranteeing prompt action to stop the spread of damaging content. These developments would help create more efficient, scalable, and successful strategies to stop cyberbullying on social media.

REFERENCES

[1] C. Fuchs, *Social media: A critical introduction*. Sage, 2017.

[2] N. Selwyn, “Social media in higher education,” *The Europa world of learning*, vol. 1, no. 3, pp. 1–10, 2012.

[3] H. Karjaluoto, P. Ulkuniemi, H. Kein¨anen, and O. Kuivalainen, “Antecedents of social media b2b use in industrial marketing context:customers’ view,” *Journal of Business & Industrial Marketing*, 2015.

[4] W. Akram and R. Kumar, “A study on positive and negative effects of social media on society,” *International Journal of Computer Sciences and Engineering*, vol. 5, no. 10, pp. 351–354, 2017.

[5] D. Tapscott *et al.*, *The digital economy*. McGraw-Hill Education,, 2015.

[6] K. Maity, S. Saha, and P. Bhattacharyya, “Emoji, Sentiment, and Emotion Aided Cyberbullying Detection in Hinglish,” IEEE Trans. Comput. Soc. Syst., pp. 1–10, 2022, doi: 10.1109/TCSS.2022.3183046.

[7] A. Ioannou *et al.*, “From risk factors to detection and intervention: A metareview and practical proposal for research on cyberbullying,” in *2017 IST-Africa Week Conference (IST-Africa)*, Windhoek, May 2017, pp. 1–8, doi:10.23919/ISTAFRICA.2017.8102355.

[8] J. W. Patchin, “2019 Cyberbullying Data,” *Cyberbullying Research Center*, Jul. 09, 2019. https://cyberbullying.org/2019-cyberbullyingdata (accessed Oct. 11, 2019).

[9] A. A. Mazari, “Cyber-bullying taxonomies: Definition, forms, consequences and mitigation strategies,” in *2013 5th International Conference on Computer Science and Information Technology*, Amman, Jordan, Mar. 2013, pp. 126–133, doi:10.1109/CSIT.2013.6588770.

[10] S. Bastiaensens, H. Vandebosch, K. Poels, K. Van Cleemput, A. Desmet, and I. De Bourdeaudhuij, “Cyberbullying on social network sites. An experimental study into bystanders’ behavioural intentions to help the victim or reinforce the bully,” *Computers in Human Behavior*, vol. 31,pp. 259–271, 2014.

[11] D. L. Hoff and S. N. Mitchell, “Cyberbullying: Causes, effects, and remedies,” *Journal of Educational Administration*, 2009.

[12] S. Hinduja and J. W. Patchin, “Bullying, cyberbullying, and suicide,” *Archives of suicide research*, vol. 14, no. 3, pp. 206–221, 2010.

[13] V. Indumathi and S. Santhana Megala. Enhanced Multi-label Classification Model for Bully Text Using Supervised
Learning Techniques, pages 763–778. Springer, pakistan, 2023.

[14] Xiaoyu Guo, Usman Anjum, and Jusin Zhan. Cyberbully detection using BERT with augmented texts. In 2022 IEEE
International Conference on Big Data (Big Data), pages 1246–1253. IEEE, 12 2022.

[15] K. Wang, Q. Xiong, C. Wu, M. Gao, and Y. Yu, “Multi-modal cyberbullying detection on social networks,” in 2020 International Joint
Conference on Neural Networks (IJCNN). IEEE, 2020, pp. 1–8.

[16] Mahmud MI, Mamun M, Abdelgawad A. A Deep Analysis of Textual Features Based Cyberbullying Detection Using Machine Learning.In2022 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT) 2022 Dec 18 (pp. 166-170). IEEE.

[17] T. A. Buan and R. Ramachandra, “Automated cyberbullying detection in
social media using an svm activated stacked convolution lstm network,”
in Proceedings of the 2020 the 4th International Conference on Compute
and Data Analysis, 2020, pp. 170–174.

[18] Jin H, Chollet F, Song Q, Hu X. AutoKeras: An AutoML Library for Deep Learning. Journal of Machine Learning Research. 2023;24(6):1-6.