Musik in bayern
ISSN: 0937-583x Volume 89, Issue 12 (Dec -2024)
https://musikinbayern.com DOI https://doi.org/10.15463/gfbm-mib-2024-*359*

CYBERBULLYING SCRUTINY FOR WOMEN SECURITY IN SOCIAL MEDIA

C. Sathish Kumar[1], K. Sindhuja[2], P. Sathya[3](Corresponding Author), R. Ganeshmurthi[4]

Department of Computer Science and Applications

SRM Institute of Science and Technology, Ramapuram, Chennai – 600089

Contact:sathishc@srmist.edu.in[1], sindhujk@srmist.edu.in[2], sathyap4@srmist.edu.in[3], ganeshmr@srmist.edu.in[4]

To Cite this Article

C. Sathish Kumar[1], K. Sindhuja[2], P. Sathya[3](Corresponding Author), R. Ganeshmurthi[4]" **CYBERBULLYING SCRUTINY FOR WOMEN SECURITY IN SOCIAL MEDIA"** *Musik In Bayern, Vol. 89, Issue 12, Dec 2024, pp73-83*

Article Info

Received: 14-10-2024 Revised: 12-11-2024 Accepted: 22-11-2024 Published: 07-12-2024

ABSTRACT

Social media platforms have become an essential part of our lives in the digital age, offering channels for networking, communication, and self-expression. But this ease of use has also brought forth a negative aspect: cyberbullying, a widespread menace that disproportionately impacts women. This abstract uses machine learning techniques—more particularly, the Random Forest algorithm—to investigate a novel strategy for enhancing women's security on social media. With its ensemble learning features, Random Forest shows to be a strong option for tackling the intricate and dynamic nature of cyberbullying. Machine learning has emerged as a potent tool in the identification and mitigation of online harassment. The program learns to identify trends, discern harmful content, and identify cases of cyberbullying against women by training on a variety of datasets and anticipating possible dangers. By taking a proactive and flexible stance, we hope to give women more confidence to use social networking sites without worrying about being harassed. The suggested approach relies on automatic content flagging and reporting, as well as real-time monitoring and response. Because Random Forest can handle big datasets and make judgments quickly, it's a great option for processing massive volumes of social media data in an efficient manner. Furthermore, the interpretability of the model makes it easier to pinpoint the essential elements that contribute to cyberbullying, leading to a more thorough comprehension of the phenomena. The study adds to the ongoing conversation about women's safety in digital environments by demonstrating the potential of machine learning. The results show that Random Forest algorithm with an accuracy of 95.86% performs better than other machine learning classifiers used for this research.

ISSN: 0937-583x Volume 89, Issue 12 (Dec -2024)
https://musikinbayern.com DOI https://doi.org/10.15463/gfbm-mib-2024-*359*
Keywords: *Machine Learning, Cyberbullying for women, Social Media, Accuracy, Logistic Regression, Random Forest, Support vector machine*

1. INTRODUCTION

The widespread expansion of social media platforms in recent years has completely changed how people connect, communicate, and exchange information throughout the world. These platforms present users with never-before-seen networking and self-expression options, but they also put users at risk of the sneaky menace known as cyberbullying, which disproportionately impacts women. Women's security and well-being have been negatively impacted by the rise in gender-based harassment in the digital sphere, which takes the form of explicit threats as well as disparaging remarks. Innovative approaches that use technology to stop cyberbullying are desperately needed as society struggles to provide a secure online environment for everyone. This study presents a ground-breaking strategy for combating the problem of cyberbullying against women on social media by anticipating possible dangers. By taking a proactive and flexible stance, we hope to give women more confidence to use social networking sites without worrying about being harassed. The suggested approach relies on automatic content flagging and reporting, as well as real-time monitoring and response. Because Random Forest can handle big datasets and make judgments quickly, it's a great option for processing massive volumes of social media data in an efficient manner. Furthermore, the interpretability of the model makes it easier to pinpoint the essential elements that contribute to cyberbullying, leading to a more thorough comprehension of the phenomena. This study adds to the continuing conversation about women's safety in digital environments by demonstrating the potential of machine learning and the Random Forest algorithm as essential instruments in building a more safe and inclusive online environment for women.

1.1 CYBERBULLYING

Cyberbullying is characterized as persistent, forceful, and purposeful harassment committed online by a person or group against a victim who is powerless to protect themselves. This kind of bullying encompasses hate speech, rumors, and abusive or sexual remarks.

1.2 CYBERBULLYING ON SOCIAL MEDIA SITES

On social media platforms, the term "cyberbullying" describes the intentional use of these channels to harass, threaten, or injure someone. This detrimental conduct presents itself in a number of ways, such as threatening remarks, distributing misleading information, and disseminating disparaging material. Although social media promotes connectivity, it also serves as a breeding ground for this type of online harassment, which has an adverse effect on victims' mental, emotional, and occasionally even physical health. The fact that cyberbullying may rapidly reach a large audience and make it difficult for victims to escape its damaging repercussions makes it especially troubling. Cyberbullying occurrences are more common and severe because of the perpetrators' increased confidence due to the anonymity provided by internet platforms. To address this issue and establish safer digital spaces for people in general and for individuals in particular, a multifaceted strategy including awareness-raising, education, and technology solutions is needed for groups, such as women, who are disproportionately targeted.

2. LITERATURE SURVEY

Prajakta Ingle et al. [1]- After reviewing the literature on a number of machine learning algorithms, the study concluded that Light GBM was the most effective. A model was created to

identify bullying tweets in real time. For the classification, we took into account a number of user-specific and Twitter variables in addition to TF IDF embedding. A comprehensive report pertaining to tweets and analysis was presented. Future research can create a mechanism to categorize these tweets into different types of bullying.

Amgad Muneer et al. [2]- In an effort to investigate this matter, a global dataset consisting of 37,373 distinct tweets from Twitter was compiled. Additionally, seven machine learning classifiers—Light Gradient Boosting Machine (LGBM), Random Forest (RF), AdaBoost (ADB), Stochastic Gradient Descent (SGD), Naive Bayes (NB), and Support Vector Machine (SVM)—were employed. The performance measures of accuracy, precision, recall, and F1 score were employed to assess each of these methods in order to ascertain the classifiers' recognition rates when applied to the global dataset. One of these problems is cyberbullying, a serious worldwide problem that has an impact on both the victims and society as a whole. Numerous initiatives to stop, stop, or lessen cyberbullying have been presented in the literature; yet, these initiatives are realistic since they depend on the interactions between the victims.

Despoina Chatzakou et al.[3]- People of all demographics are affected by the concerning trends of cyberbullying and cyber aggression. Globally, more than half of the youth who use social media have experienced this kind of ongoing and/or coordinated online harassment. A wide range of emotions can be experienced by victims, and many of these can have detrimental effects including sadness, humiliation, and social isolation. These effects increase the likelihood of even more serious outcomes like suicide attempts. In this study, we make the initial tangible progress toward comprehending the traits of abusive conduct on Twitter, one of the biggest social media platforms available today. We compare users engaging in debates about seemingly regular issues, like the NBA, to those more likely to be hate-related, like the Gamer gate, by analyzing 1.2 million users and 2.1 million tweets controversy, or the gender pay inequality at the BBC station. We also explore specific manifestations of abusive behavior, i.e., cyberbullying and cyber aggression, in one of the hate-related communities (Gamer gate).

YuYi Liu et al. [4]- This research develops a comprehensive multi-dimensional feature set that considers cyberbullying factors that are based on individual, social network, episode, and language content. We created and developed cyberbullying detection models on the KNIME machine learning platform in order to evaluate the effectiveness of the suggested multi-dimensional feature set. Our cyberbullying detection models employed six distinct machine learning algorithms: Naïve Bayes, Decision Tree, Random Forest, Tree Ensemble, Logistic Regression, and Support Vector Machines. Our experimental findings show that all evaluated machine learning algorithms perform better when the suggested multi-dimensional feature set—that is, the set that is not restricted to language features—is applied.

Emre Cihan Ates et al. [5]- Actively identifying and regulating the items that comprise cyberbullying is the most fundamental method of defense against the harmful effects of cyberbullying. Given the numbers from today's social media and internet, it is hard to identify cyberbullying content using human labor alone. Detecting cyberbullying effectively is essential to creating a secure communication environment on social media. The goal of current research is to detect and eradicate cyberbullying with machine learning. There aren't many studies in Turkish for the detection of cyberbullying, despite the majority of studies being done on English texts. Studies on the Turkish language also employed restricted techniques and algorithms.

Subbaraju Pericherla et al. [6]- Social media sites like Facebook and Twitter offer a fantastic forum for the public to express their thoughts, feelings, and opinions via text message, picture, or video. The user-friendly Graphical User Interface (GUI) that allows users to share content from

their cellphones and other electric devices with just a few clicks or taps has piqued the attention of the general public in using these networks. However, some individuals engage in cyberbullying behaviors such as abusive remarks, trolling, and hostile remarks. In light of this, the research focuses on analyzing different techniques and strategies for using machine learning to identify instances of cyberbullying in social media posts. The majority of methods currently in use do not take into account sarcastic text and only take into account a few number of content attributes in order to identify instances of cyberbullying.

Aaminah Ali et al. [7] - Cyberbullying is a type of bullying when people are insulted or harmed via the use of technology. Sarcasm is one area of this threat that has to be addressed despite the numerous answers and tactics that scholars have suggested. The purpose of this study is to draw attention to earlier studies and suggest a method for identifying cyberbullying and the sarcastic component of it. The outcomes demonstrated that the SVM classifier outperformed the other classifiers. The purpose of this specific study was to investigate machine learning-based cyberbullying detection. The phrase "cyberbullying" is broad and has several meanings.

Mohammed Ali Al-Garadi et al. [8] - Online human networks, rich human behavior-related data, and user-generated content have all undergone revolution thanks to these social technologies. However, a new kind of violence and aggressiveness that only happens online has been brought about by the misuse of social technologies, such as social media (SM) platforms. This research highlights a new method of displaying hostile behavior on social media websites. Additionally described are the reasons behind the development of prediction models to combat aggressive behavior in SM. We conduct a thorough analysis of cyberbullying prediction models and pinpoint the primary problems associated with their development on social media.

Manuel F. López et al. [9] - In this paper, we investigate various methods that consider the latency in cyberbullying detection on social media platforms. We employ two distinct early detection models, namely threshold and dual, in a supervised learning approach. While the latter involves two machine learning models, the former uses a more straightforward strategy. To the best of our knowledge, this is the first study looking into cyberbullying early detection. Specifically addressing this issue, we suggest two feature groups and two early detection techniques.

3. PROPOSED SYSTEM

Cyberbullying is a severe problem that many people experience on social media, especially women. Given that social networking sites are among the websites that teenagers use the most, a suggested system for women's security in social media utilizing machine learning might be built in order to address this issue. As a result, bullying may target those who are already at risk. This work explains the machine learning method for identifying cyberbullying. To find the offensive or abusive comment, they investigate the two feature detection hypotheses. This will have negative effects on them because of the statement. Cyberbullying is a major problem on social networking sites, per a new survey. In their piece, they concentrated on bullies, victims, and onlookers on social media. They increased the viability of automated cyberbullying by employing a number of binary classifications. In a similar vein, they use the linear support vector machine to alter the feature set. The suggested approach analyzes crimes committed against women on social media. In order to examine the data and find patterns and trends that can aid in comprehending the issue and creating workable solutions, the system will employ machine learning algorithms. The objective of the first strategy is restricted to the social media network. The second method is

ISSN: 0937-583x Volume 89, Issue 12 (Dec -2024)
https://musikinbayern.com DOI https://doi.org/10.15463/gfbm-mib-2024-*359*
address-focused (similar to hashtags), and only topic-based analysis may be performed. The third technique is a data characteristic that is previously predefined.

4. METHODOLOGY

Real-time cyberbullying monitoring and detection is possible with the machine learning system, which can facilitate the prompt identification and handling of abusive and harassing situations. The technology can be tailored to recognize particular forms of cyberbullying to which women are especially susceptible, so facilitating the provision of focused protection and assistance. Easily scaled to monitor many social media channels at once, and trained on copious volumes of data to offer complete defense and assistance. It can lower the expenses related to cyberbullying by automating the identification and monitoring of manual observation and involvement. Machine learning has the potential to offer a more effective, efficient, and adaptable solution to the significant problems of cyberbullying and women's safety on social media. Real-time identification of cases of abuse and harassment can be beneficial to lessen the effects of cyberbullying and encourage a more secure online space for females.

Musik in bayern
ISSN: 0937-583x Volume 89, Issue 12 (Dec -2024)
https://musikinbayern.com DOI https://doi.org/10.15463/gfbm-mib-2024-*359*

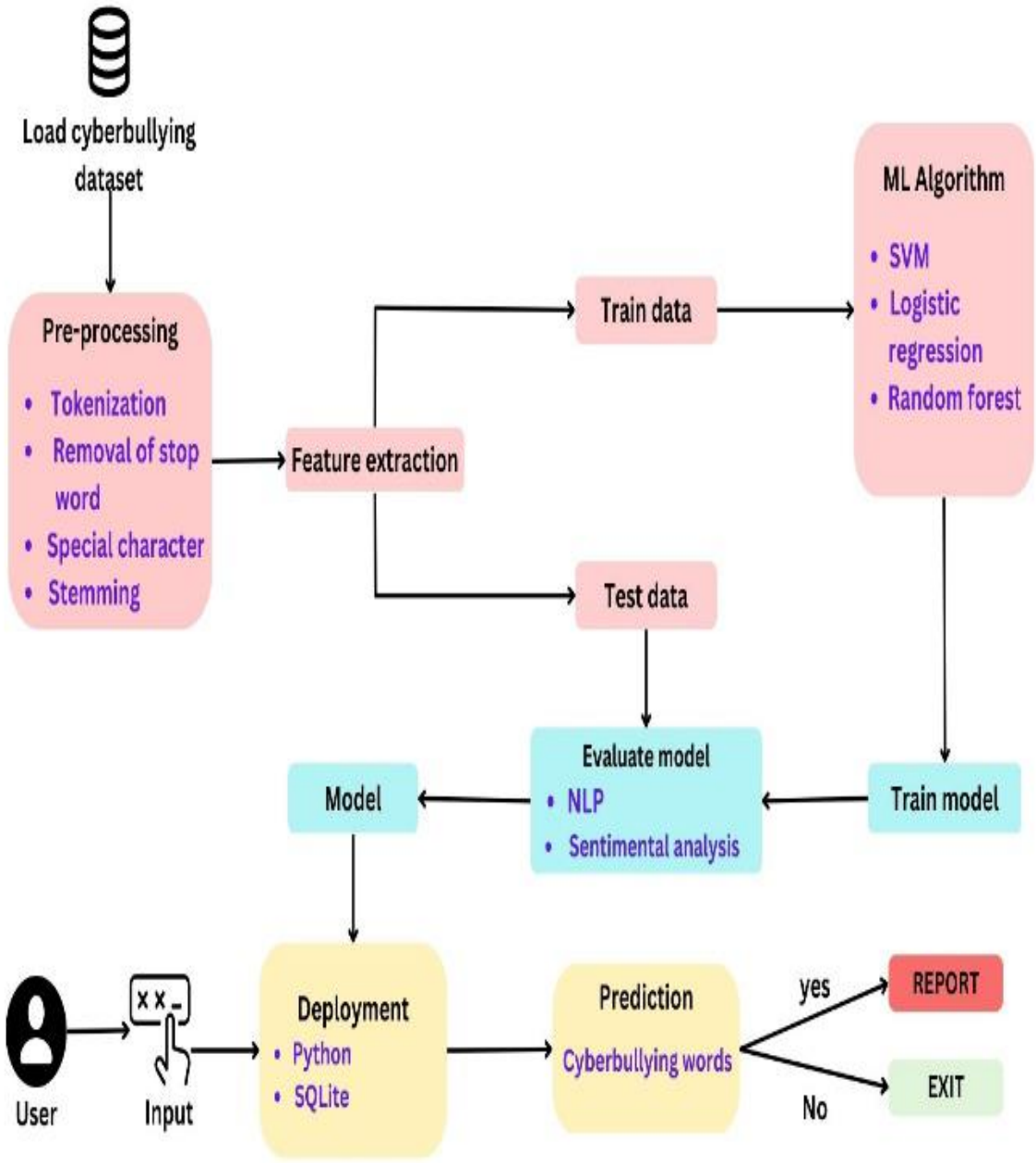

Figure 1: General Architecture for monitoring and detecting cyberbullying

4.1 Data set collection

Assembling a dataset to gather information about the various forms of cyberbullying that target women on social media. This data could be gathered from various sources such as social media platforms, online forums, and surveys. The data set for this research is taken from github **[10]**. This dataset contains 500 samples exclusively meant for research in cyberbullying. This dataset integrates multiple modalities, including text, image features and video features content,

along with user metadata and other metrics. This may increase the precision with which incidents of cyberbullying are detected.

- Identify relevant sources: Determine which sources provide information on crime against women and children in India. This may include government agencies such as the National Crime Records Bureau, police departments, and non-governmental organizations that track violence against women and children.
- Determine data fields: Decide which fields are relevant to the dataset, such as the type of crime, location, age of victim, and demographic information.
- Gather data: Collect data from the identified sources. This may involve accessing online databases, filing requests for information, or manually compiling data from reports.
- Clean and process data: Once the data is collected, it may need to be cleaned and processed to ensure accuracy and consistency. This may involve removing duplicate entries, standardizing formatting, and addressing missing data.
- Analyse data: After cleaning and processing, the data can be analysed to identify trends and patterns in crime against women and children. This analysis can help identify areas of high incidence and inform policy and intervention strategies.
- Ensure privacy and confidentiality: It's important to ensure that the dataset is collected and used in a way that protects the privacy and confidentiality of victims and individuals mentioned in the data.

Overall, collecting a dataset of crime against women and children in India is a complex process that requires careful consideration of sources, data fields, and privacy concerns. However, by collecting and analysing this data, we can better understand the scope and nature of these crimes and develop effective interventions to prevent them.

4.2 Data preprocessing

Data preprocessing is an important step in using machine learning algorithms to analyze crime against women and develop women safety solutions.

- Data cleaning: Raw data may contain errors, inconsistencies, and missing values that need to be addressed before analysis. Data cleaning involves identifying and correcting or removing these issues to ensure the accuracy and completeness of the dataset.
- Data integration: In some cases, data may be collected from multiple sources, each with its own format and structure. Data integration involves combining these sources into a single, cohesive dataset that can be used for analysis.
- Data transformation: Machine learning algorithms may require data to be transformed into a specific format or range before it can be used effectively. Data transformation can involve scaling data, converting categorical data into numerical data, or applying other mathematical functions.
- Data reduction: Large datasets may be difficult or time-consuming to analyse, and may contain irrelevant or redundant information. Data reduction involves removing or consolidating features in the dataset to simplify analysis and improve algorithm performance.
- Data sampling: In some cases, the dataset may be too large to analyse comprehensively. Data sampling involves selecting a representative subset of the dataset to analyse, which can reduce computational complexity and improve algorithm performance.

4.3 NLP processing

The machine learning model could not be constructed using the text data. For this reason, we must convert to a vector format. In this approach, the vectorization operations are carried out via the NLP toolbox.

ISSN: 0937-583x Volume 89, Issue 12 (Dec -2024)
https://musikinbayern.com DOI https://doi.org/10.15463/gfbm-mib-2024-*359*

4.4 Model Selection

To create effective machine learning models, the program compares the algorithms for using Random Forest and Logistic Regression based models. The model will be chosen based on its improved cross-validation performance.

4.5 Model Deployment

To determine the risk factor of a certain location, machine learning algorithms like SVM and logistic regression will be utilized. Model development is the process of training the machine learning models on the train data and testing on the testing data.

4.6 MACHINE LEARNING ALGORITHMS

Numerous models and algorithms that can be used for a variety of use cases fuel predictive analytics systems. To get the most out of a predictive analytics system and use data to make wise decisions, you must decide whether predictive modeling approaches are appropriate for your business. Within the statistical domain, machine learning is characterized as an artificial intelligence application in which preexisting data is processed or aided in the processing of statistical data via algorithms. Even though machine learning uses automation techniques, human supervision is still necessary. High levels of generalization are required in machine learning to create systems that function well on data examples that have not previously been observed. Within computer science, machine learning is a relatively young field that offers a compilation of methods for analyzing data. While many of these techniques lack the foundation of well-established statistical methods (e.g., principal component analysis and logistic regression), others do.

The majority of statistical methods operate under the paradigm of selecting a specific probabilistic model from a group of related models that best fits the observed data. Similar to this, the majority of machine learning approaches are made to identify models that solve certain optimization problems by finding the models that best suit the data; the only difference is that these models are no longer limited to probabilistic ones.

Thus, machine learning techniques have an advantage over statistical techniques in that the latter do not require underlying probabilistic models, but the former do. The classical statistical techniques are typically overly strict, even if some machine learning techniques incorporate probabilistic models.

A wider class of more adaptable alternative analytical techniques that are more appropriate for contemporary data sources may be made available via machine learning. Statistical agencies must investigate the potential applications of machine learning techniques in order to ascertain whether these methods can better serve their needs in the future than more conventional ones.

4.6.1 LOGISTIC REGRESSION

Logistic regression is a statistical method that is used for building machine learning models where the dependent variable is dichotomous: i.e. binary. Logistic regression is used to describe data and the relationship between one dependent variable and one or more independent variables. The independent variables can be nominal, ordinal, or of interval type.

4.6.2 K-NEAREST NEIGHBOUR OR KNN

K-nearest neighbors (KNN) is a machine learning algorithm used for classification and regression tasks. In the training phase, it memorizes all available data points. When predicting a new point, KNN calculates distances to find the k-nearest neighbors and assigns a label based on majority voting for classification or averaging for regression. The choice of 'k' influences model sensitivity,

ISSN: 0937-583x Volume 89, Issue 12 (Dec -2024)
https://musikinbayern.com DOI https://doi.org/10.15463/gfbm-mib-2024-*359*
and it is a non-parametric algorithm, making no assumptions about data distribution. Implementation typically involves libraries like scikit-learn in Python, where training, prediction, and evaluation are straightforward processes.

4.6.3 RANDOM FOREST

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

5. RESULTS AND DISCUSSION

The following evaluation metrics are used to identify the best performing machine learning algorithm for analyzing the instances of cyberbullying.

5.1 *ACCURACY*

Accuracy in machine learning is a metric used to assess the performance of a classification model. It measures the ratio of correctly predicted instances to the total number of instances in the dataset.

5.2 *PRECISION*

Precision is a metric used in machine learning to evaluate the performance of a classification model, particularly in scenarios where the focus is on minimizing false positives. It is the ratio of correctly predicted positive observations to the total predicted positives.

5.3 *RECALL*

Recall, also known as sensitivity or true positive rate, is a metric used in machine learning to assess the ability of a classification model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives.

5.4 *F1-SCORE*

The F1-score is a metric in machine learning that combines precision and recall into a single value. It is especially useful when there is an uneven class distribution, and you want to balance the trade-off between false positives and false negatives.

Based on the results, the below given table 1 identifies the best performing algorithm. The results show that Random Forest algorithm gives a precision value of 82.90%, recall value of 83.43% f1-score value of 81.7% and an accuracy of 95.86%.

Machine Learning algorithm	**Precision**	**Recall**	**F1-Score**	**Accuracy**
Logistic Regression	81.45%	81.58%	81.7%	87.42%
K-Nearest Neighbor	68.6%	65.3%	67.5%	69.2%
Random Forest	**82.90%**	**83.43%**	**81.7%**	**95.86%**

Musik in bayern
ISSN: 0937-583x Volume 89, Issue 12 (Dec -2024)
https://musikinbayern.com DOI https://doi.org/10.15463/gfbm-mib-2024-*359*

Table 1. Comparison of classifier models based on Precision, Recall, F1-score and Accuracy

Figure 2 gives a comparison analysis of different classifier models based on precision, recall, f1-score and accuracy. The results show that the logistic regression model produces an accuracy of 87.42%, the K-Nearest Neighbor model gives an accuracy of 69.2% and the random forest model gives an accuracy of 95.86%. The results show that random forest performs better than all other algorithmic models.

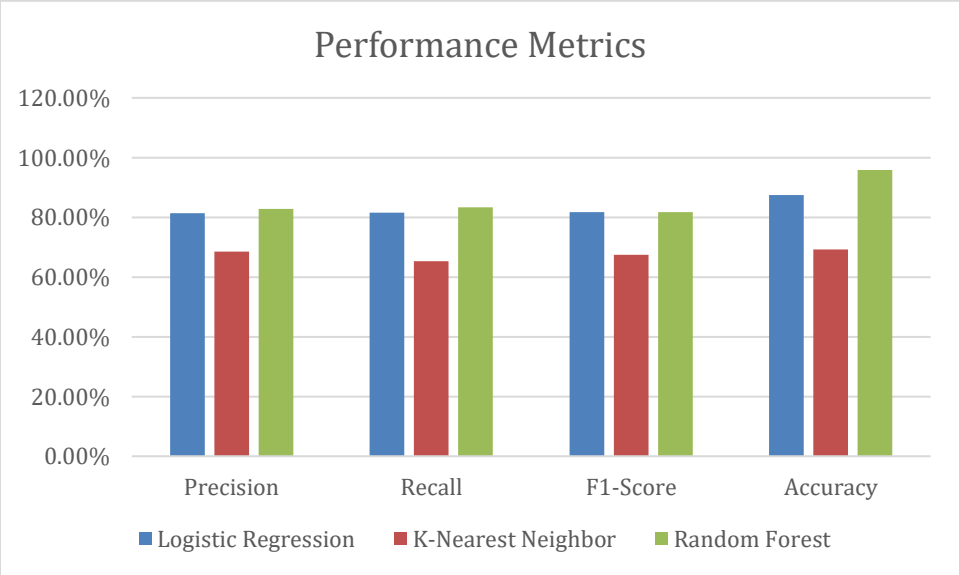

Figure 2. Comparison of classifier models based on Precision, Recall, F1-score and Accuracy

6. CONCLUSION

Conclusively, the analysis of women's safety is a comprehensive and intricate endeavor that necessitates meticulous evaluation of numerous aspects, such as data collection, preprocessing, analysis, and interpretation. We can comprehend the nature and extent of this problem by identifying patterns and trends in the crime data that we examine using machine learning algorithms. We may also use this research to create efficient plans of action and interventions to stop and address crimes against women. It is crucial to remember that studying crime against women does not, by itself, provide a remedy. It is imperative that we utilize the knowledge gleaned from data analysis to guide decisions and initiatives. To better defend women's rights and safety, this may entail enacting focused measures, raising public awareness, and fortifying legal frameworks. We collected a variety of data sets to identify texts that involved cyberbullying. This social media sentiment collection technology makes it possible to collect quantitative data globally, including the frequency of people's emotional feelings around harassment, providing a strong basis for improving and empowering security. The likelihood that these evil operations will have an impact on society would surely be reduced by this technique of analyzing popular emotional sentiments and providing advice to reduce insecure situations that develop in society. The proposed system with random forest algorithm gives an accuracy of 95.86% which is better than the other classifiers used such as logistic regression and k-nearest neighbor. This proposed system would include a wide range of criteria for identifying offensive language, together with proactive measures by the government and public education to protect women from sexual

harassment and other types of abuse. Additionally, this gadget uses it to alert the woman to potentially hazardous situations. Additionally, the system outperforms other systems in terms of efficiency. It also effectively provides better safety to ladies.

REFERENCES

[1] M. E. Hossain, M. W. Rahman, M. T. Islam and M. S. Hossain, "Manifesting a mobile application on safety which ascertains women salus in bangladesh", *International Journal of Electrical and Computer Engineering*, vol. 9, no. 5, pp. 4355, 2019.
[2] L. Mubassara, M. I. I. Towhid, S. Sultana, A. I. Anik, M. Salwa, M. M. H. Khan, et al., "Cyber child abuse in bangladesh: A rural population-based study", *World*, vol. 8, no. 1, 2021.
[3]. M. M. Hossain, M. Asadullah, A. Rahaman, M. S. Miah, M. Z. Hasan, T. Paul, et al., "Prediction on domestic violence in bangladesh during the covid-19 outbreak using machine learning methods", 2021.
[4] J. Thavil, V. Durdhawale and P. Elake, "Study on smart security technology for women based on iot", *International Research Journal of Engineering and Technology (IRJET)*, vol. 4, no. 02, 2017.
[5] W. Akram, M. Jain and C. S. Hemalatha, "Design of a smart safety device for women using iot", *Procedia Computer Science*, vol. 165, pp. 656-662, 2019.
[6] Sumathy, P. D. Shiva, P. Mugundhan, R. Rakesh and S. S. Prasath, "Virtual friendly device for woman security", *J. Phys. Conf. Ser.*, vol. 1362, no. 1, 2019.
[7] C. M. Akhila and J. Raju, "B with U- IoT based Woman Security System", vol. 8, no. 08, pp. 507-509, 2019.
[8] Dhruvil Parikh, Pallavi Kapoor, Shital Karnani and Sudhir Kadam, "IoT based Wearable Safety Device for Woman", *Int. J. Eng. Res.*, vol. V9, no. 05, pp. 1086-1089, 2020.
[9] Priyanka Das, Asit Kumar Das, "Crime analysis against women from online newspaper reports and an approach to apply it in dynamic environment", (ICBDAC) 2017 (IEEE Xplore :October 2017)
[10] GitHub - kmkrphd/Cyberbullying-Dataset retrieved on 10/11/24.