IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

DETECTION OF CYBER BULLYING ON SOCIAL MEDIA USING MACHINE LEARNING ALGORITHMS

Dr. G. Satya Narayana
V. Susmitha
Computer science and engineering students
(professor of computer science and engineering)
J. Nagarani
Lingayas Institute of Management and
Dr. M. Chinnarao
P. Lavanya
Technology
(head of the department)

ABSTRACT: The rise of social media platforms has created new opportunities for individuals to connect and share information. However, these platforms have also given rise to cyberbullying, a pervasive and harmful form of online behavior that can have serious consequences. To combat cyberbullying, researchers have begun exploring the use of machine learning techniques to detect and prevent these incidents.

This project presents a comprehensive review of the current state-of-the-art in cyberbullying detection using machine learning, including an analysis of the various techniques and approaches that have been used, their effectiveness, and the challenges that remain. Through this review, we identify areas for future research and provide recommendations for improving the accuracy and effectiveness of cyberbullying detection using machine learning.

Keywords: *Cyberbullying, Machine Learning, Natural Language Processing, Text Classification, Online Safety.*

INTRODUCTION

Cyberbullying is the use of electronic communication to bully, threaten, or harass individuals or groups, commonly occurring on social media platforms, through text messages, or on online forums. This form of abuse encompasses various behaviors such as spreading rumors, sharing embarrassing content, or sending threatening messages, all of which can inflict severe psychological and emotional harm on victims.

Cyber bullying is a type of harassment that takes place online or through digital communication channels. It involves using electronic devices and online platforms to threaten, humiliate, or intimidate someone. The rise of social media and other online platforms has led to an increase in cyberbullying incidents, making it a significant societal problem that needs to be addressed.

One of the ways to tackle cyberbullying is to use machine learning techniques to detect and prevent it. Machine learning is a subfield of artificial intelligence that involves the use of algorithms and statistical models to enable computers to learn from data and make predictions or decisions without being explicitly programmed. By training machine learning models on large datasets of cyberbullying instances, it is possible to develop accurate and reliable cyberbullying detection systems.

The objective of this research is to develop a cyberbullying detection system using machine learning techniques. The system will be trained on large datasets of cyberbullying instances, including text, images, and videos. The system will use a combination of natural language processing (NLP) and computer vision techniques to analyze the content and context of digital communication channels, such as social media platforms, messaging apps, and online forums.

The proposed system will use supervised machine learning algorithms, such as logistic regression, decision trees, and support vector machines (SVM), to classify instances of cyberbullying accurately. The system will be evaluated using standard metrics such as precision, recall, and F1 score to measure its performance.

The potential benefits of this research are significant. A reliable and accurate cyberbullying detection system can help prevent cyberbullying incidents and protect vulnerable individuals. The system can also help parents, teachers, and other caregivers to monitor and control children's online activities and protect them from cyberbullying. Additionally, the system can help social media platforms and online forums to detect and remove cyberbullying content and improve their user experience.

In conclusion, developing a cyberbullying detection system using machine learning is a promising approach to addressing the issue of cyberbullying. The proposed system can help prevent cyberbullying incidents, protect vulnerable individuals, and improve the safety and security of online communities.

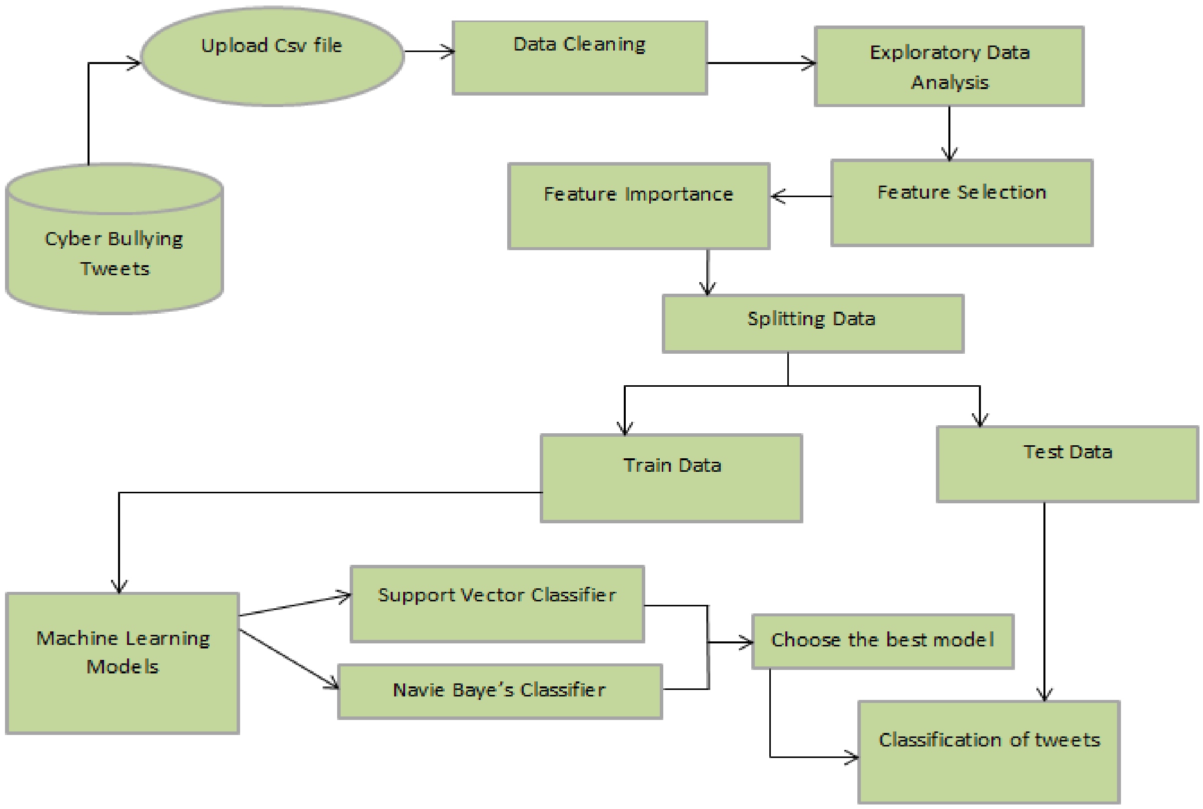
PROBLEM STATEMENT:

The problem statement for "Cyberbullying Detection Using Machine Learning" is to develop an automated system that can identify cyberbullying behavior in social media platforms, emails, and other digital communication channels. The system must be able to analyze text and multimedia data and classify them as either cyber bullying or not. The system needs to be accurate in detecting cyberbullying behavior and not misclassify harmless communication. It should also be able to identify patterns and trends in cyberbullying behavior to help develop preventive measures. The development of such a system is important because cyberbullying can have serious consequences for individuals, including depression, anxiety, and even suicide. It can also have a negative impact on society as a whole, as it can lead to social isolation and disengagement. To develop a system that can accurately detect cyberbullying, a large dataset of labeled examples will be needed, along with various machine learning algorithms, such as natural language processing, computer vision, and deep learning. The system will also require ongoing training and updates to stay current with evolving cyberbullying tactics and behaviors.

DESIGN:

SYSTEM ARCHITECTURE:

System architecture is a conceptual model that describes the structure and behaviour of multiple components and subsystems.



USE CASE DIAGRAM:

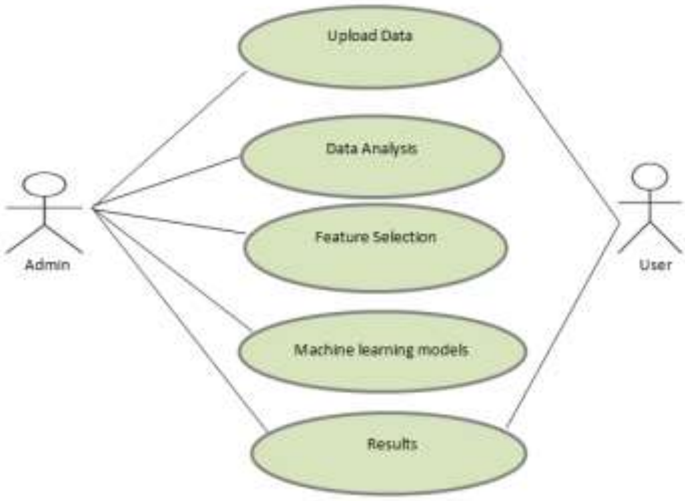
- A use case diagram is used to represent the dynamic behaviour of a system. It encapsulates the system's functionality by incorporating use cases, actors, and their relationships.

Purpose

- The main purpose of a use case diagram is to portray the dynamic aspect of a system. It accumulates the system's requirement, which includes both internal as well as external influences.

Components

- USE CASE
- ACTOR
- ASSOCIATION OR COMMUNICATION LINK IN BOUNDARY OF SYSTEM



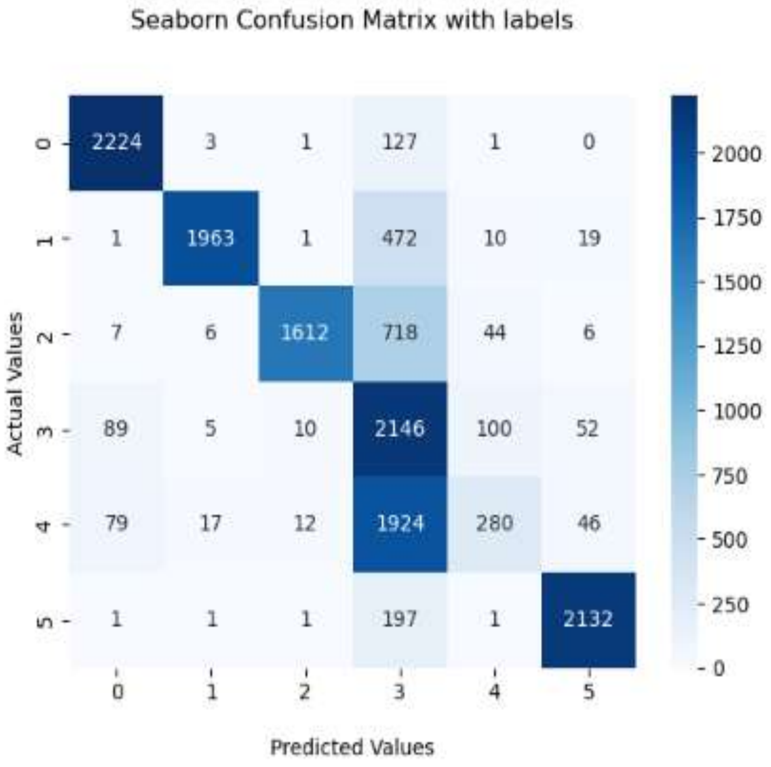
MODEL INFORMATION:

1. SVC:

SVC stands for Support Vector Machine Classifier, which is a type of machine learning model used for classification tasks. The SVC model works by finding the optimal hyperplane (i.e., a line or a plane) that separates the data into different classes in a way that maximizes the margin (i.e., the distance between the hyper plane and the closest data points of each class). This hyper plane is determined by identifying a set of support vectors, which are the data points closest to the hyper plane.

In practical terms, the SVC model works by taking a set of input data and output labels, and using this to learn the optimal hyper plane that can be used to classify new, unseen data. The model is trained by finding the set of support vectors that maximizes the margin between the classes. Once trained, the SVC model can be used to predict the class of new data based on its input features.

SVC models can be used for a wide range of classification tasks, from image recognition to fraud detection. They are particularly useful when the data is not linearly separable, as they can use a technique called kernel trick to transform the data into a higher dimensional space where it is more easily separable.

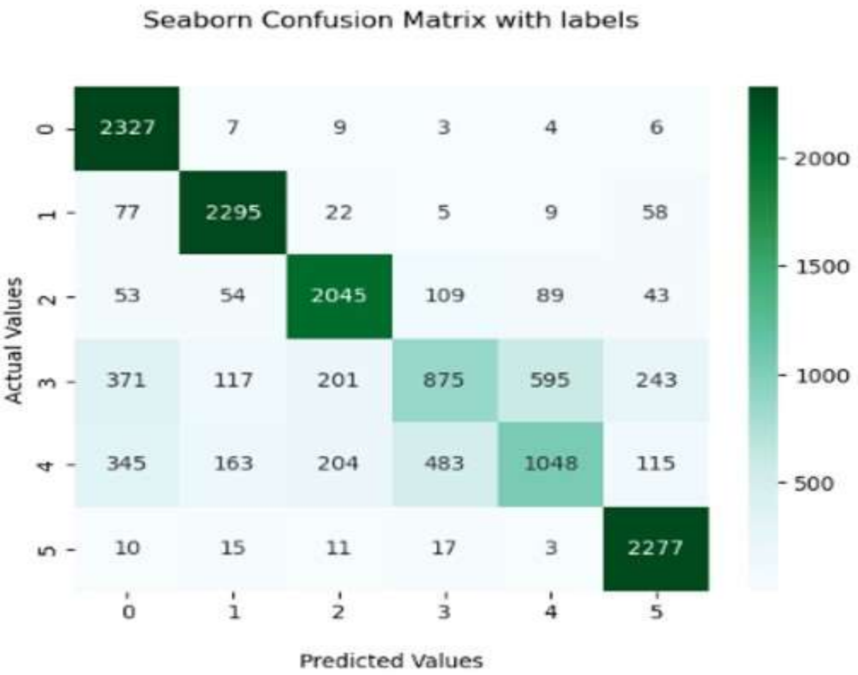


2. Navie Bayes (Nb)

NB stands for Naive Bayes, which is a type of probabilistic machine learning model used for classification tasks. The NB model is based on Bayes' theorem, which states that the probability of a hypothesis (i.e., the output class) is proportional to the prior probability of the hypothesis and the likelihood of the evidence (i.e., the input features) given the hypothesis.

The NB model works by assuming that the input features are conditionally independent given the output class. This is known as the "naive" assumption, as it is often not true in practice, but simplifies the model and makes it computationally efficient. Based on this assumption, the NB model estimates the probability of each output class given the input features using the Bayes' theorem and the prior probabilities of the output classes.

There are several variants of the NB model, including the Gaussian NB, Multinomial NB, and Bernoulli NB. Gaussian NB assumes that the input features follow a Gaussian distribution, while Multinomial NB and Bernoulli NB are used for discrete input features such as word counts in text classification tasks.



EVALUATION METRICS

The performance of the proposed architecture is evaluated based on several statistical measures in addition to our new metric, defined as the corona score.

Accuracy

Accuracy is a metric that quantifies the competency of the method in defining the correct predicted cases

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

- True Positive: True Positive is equal to the number of correct predicted positive cases.
- False Positive: False Positive is equal to the number of incorrect predicted positive cases.
- True Negative: True Negative is equal to the number of correct predicted negative cases.
- False Negative: False Negative is equal to the number of incorrect predicted negative cases

Recall

Recall is the sensitivity of the method

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

Precision

Precision is the ratio of the unnecessary positive case to the total number of positives.

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

Specificity

Specificity is the ratio of correct predicted negatives over negative observations.

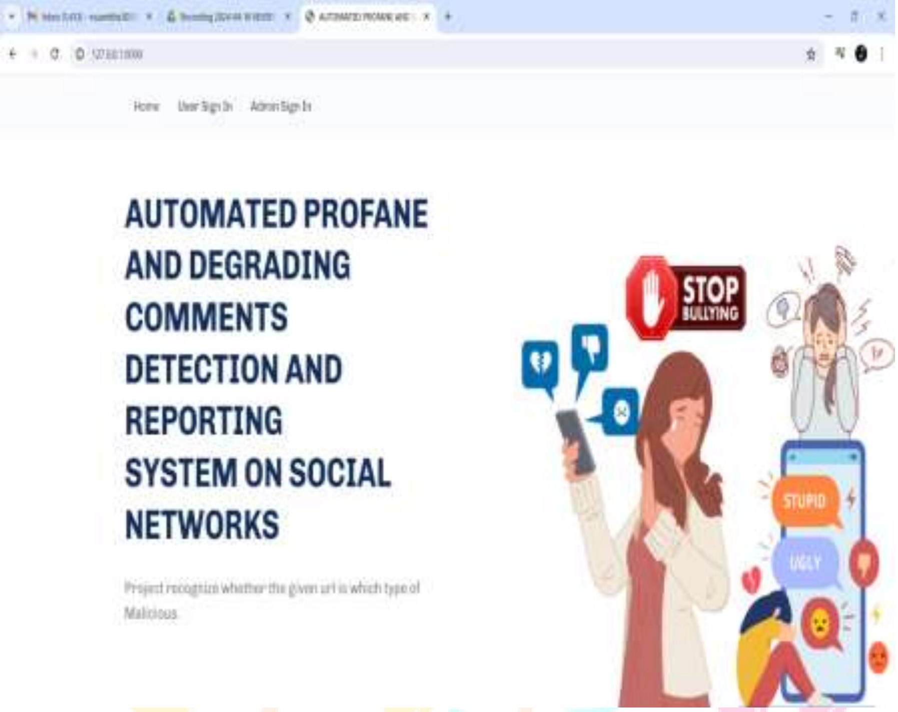
$$Specificity = \frac{TN}{TN+FP}.$$

F1-Score

F1-Score is the measure of the quality of detection.

$$\text{F1 Score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Results:

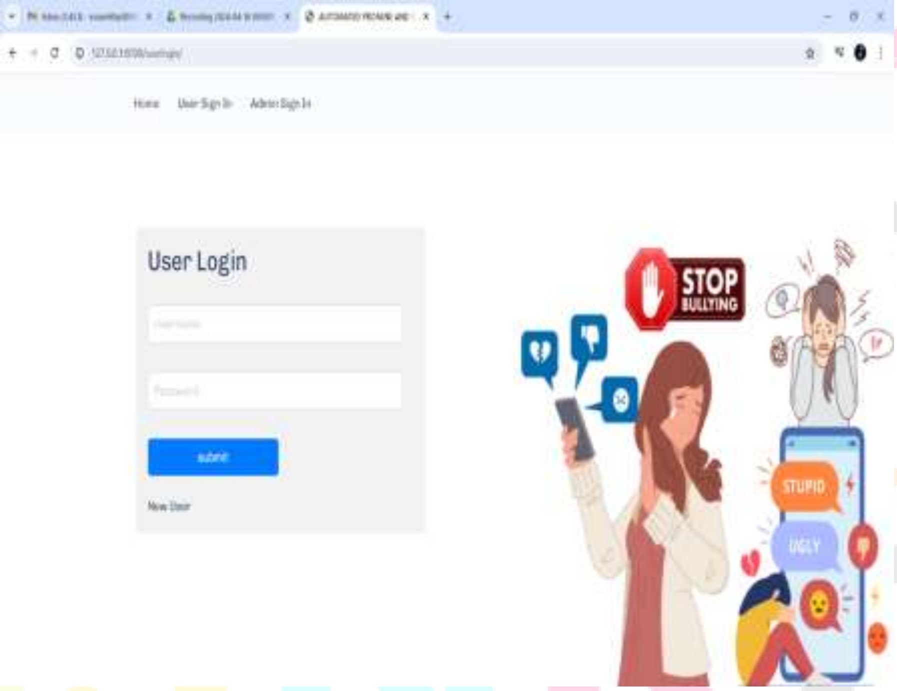


Home User Sign In Admin Sign In

AUTOMATED PROFANE AND DEGRADING COMMENTS DETECTION AND REPORTING SYSTEM ON SOCIAL NETWORKS

Project recognize whether the given url is which type of Malicious

Figure 1: home page



Home User Sign In Admin Sign In

User Login

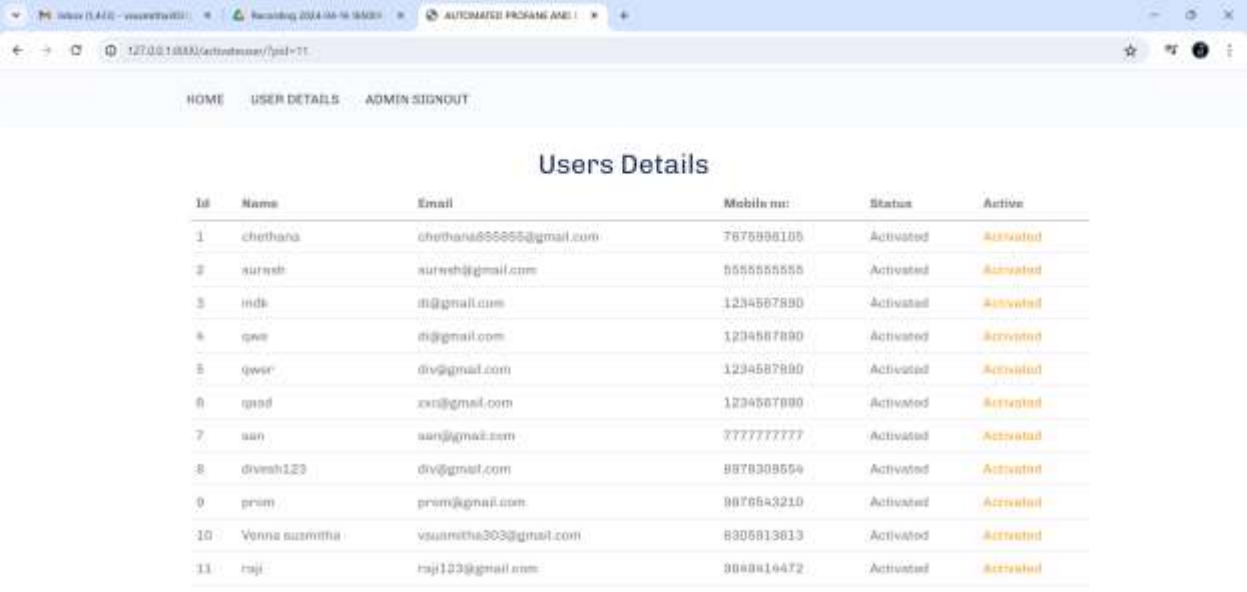
Username

Password

submit

New User

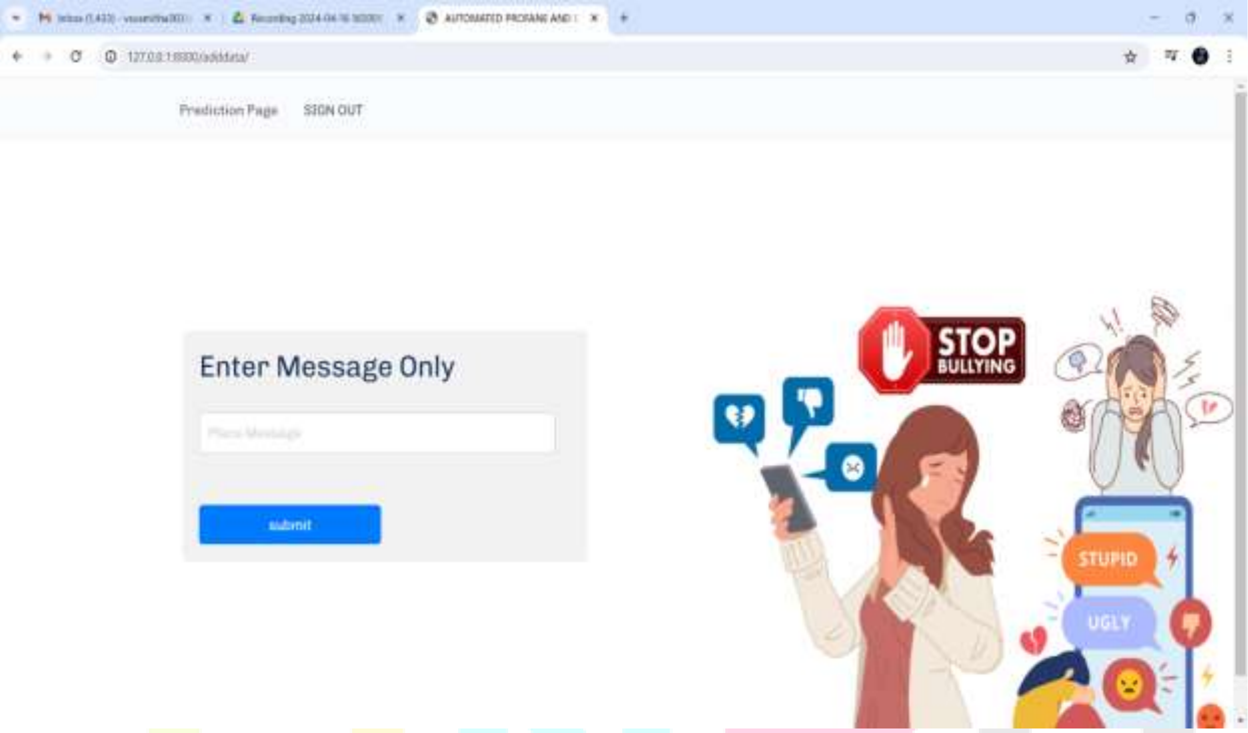
Figure 2:user login



Users Details

Id	Name	Email	Mobile no:	Status	Active
1	chethana	chethana655655@gmail.com	7675998105	Activated	Activated
2	suresh	suresh@gmail.com	5555555555	Activated	Activated
3	indk	di@gmail.com	1234567890	Activated	Activated
4	qwe	di@gmail.com	1234567890	Activated	Activated
5	qwer	div@gmail.com	1234567890	Activated	Activated
6	qasd	zxc@gmail.com	1234567890	Activated	Activated
7	san	san@gmail.com	7777777777	Activated	Activated
8	divesh123	div@gmail.com	8978309554	Activated	Activated
9	prem	prem@gmail.com	9876543210	Activated	Activated
10	Venna susmitha	vsusmitha303@gmail.com	8305813813	Activated	Activated
11	raji	raji123@gmail.com	9848414472	Activated	Activated

Figure 3: user details

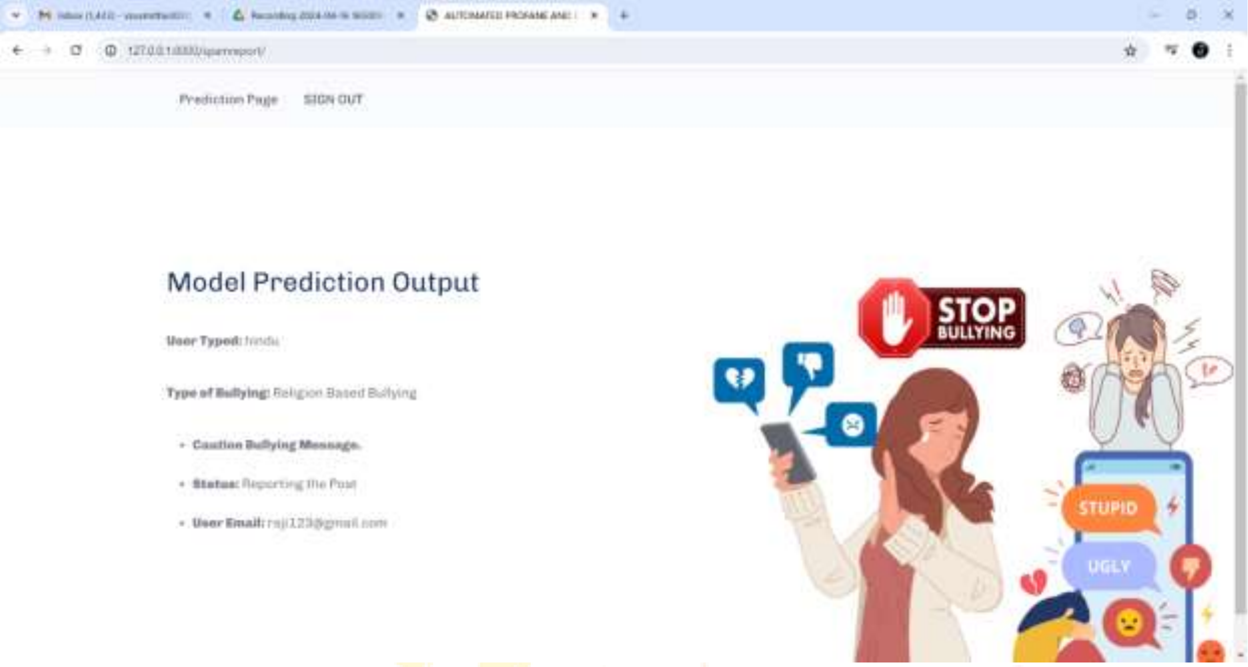


Enter Message Only

Place Message

submit

Figure 4: prediction page



Prediction Page SIGN OUT

Model Prediction Output

User Typed: hindu

Type of Bullying: Religion Based Bullying

- **Caution Bullying Message.**
- **Status:** Reporting the Post
- **User Email:** raji123@gmail.com

Figure 5: model prediction output

CONCLUSION

In conclusion, cyber bullying detection using machine learning is an important area of research that has gained significant attention in recent years. The use of machine learning models such as Support Vector Machine (SVM) and Naive Bayes (NB) have shown promising results in detecting and classifying instances of cyberbullying in online social media platforms with accuracies of 72 % and 75 %. SVM models are effective in dealing with high dimensional data and can handle non-linearly separable data by using kernel methods. On the other hand, NB models are relatively simpler to implement, computationally efficient, and can handle high-dimensional data with many input features.

Overall, both SVM and NB models can be used in cyber bullying detection with varying degrees of effectiveness, depending on the specific use case and the nature of the data. It is important to note that these models are not a complete solution to the problem of cyberbullying detection, and must be used in conjunction with other methods and techniques for effective detection and prevention. With further research and development, machine learning models such as SVM and NB have the potential to greatly enhance our ability to detect and prevent cyberbullying, and promote a safer and more inclusive online environment.

REFERENCES

1. Rafiq, R. I., Hosseinmardi, H., Han, R., Qin, L. Q., Mishra, S., & Mattson, S. A. (2015). Careful what you share in six seconds. In Advances in Social Networks

Analysis and Mining. https://doi.org/10.1145/2808797.2809381

2. Felix, E. D., Sharkey, J. D., Green, J. C., Furlong, M. J., & Tanigawa, D. (2011). Getting precise and pragmatic about the assessment of bullying: The development of the California Bullying Victimization Scale. Aggressive Behavior, 37(3), 234–247. https://doi.org/10.1002/ab.20389

3. Zhao, R., & Mao, K. (2017). Cyberbullying Detection Based on Semantic-Enhanced Marginalized Denoising Auto-Encoder. IEEE Transactions on Affective Computing, 8(3), 328–339. https://doi.org/10.1109/taffc.2016.2531682

4. Srinath, A., Johnson, H., Dagher, G. G., & Long, M. (2021). Bully Net: Unmasking Cyber bullies on Social Networks. IEEE Transactions on Computational Social Systems, 8(2), 332–344. https://doi.org/10.1109/tcss.2021.3049232

5. Pitsilis, G., Ramampiaro, H., & Langseth, H. (2018). Effective hate-speech detection in Twitter data using recurrent neural networks. Applied Intelligence, 48(12), 4730–4742. https://doi.org/10.1007/s10489-018-1242-y

6. Tripathy, J. K., Chakkaravarthy, S. S., Satapathy, S. C., Sahoo, M., & Vaidehi, V. (2020). ALBERT-based fine-tuning model for cyberbullying analysis. Multimedia Systems, 28(6), 1941–1949. https://doi.org/10.1007/s00530-020-00690-5

7. Tokunaga, R. S. (2010). Following you home from school: A critical review and synthesis of research on cyberbullying victimization. Computers in Human Behavior, 26(3), 277–287. https://doi.org/10.1016/j.chb.2009.11.014

8. Chen, J., Yan, S., & Wong, K. (2020). Verbal aggression detection on Twitter comments: convolutional neural network for short-text sentiment analysis. Neural Computing and Applications, 32(15), 10809–10818. https://doi.org/10.1007/s00521-018-3442-0

9. Squicciarini, A., Rajtmajer, S. M., Liu, Y. W., & Griffin, C. A. (2015). Identification and characterization of cyberbullying dynamics in an online social network. In Advances in Social Networks Analysis and Mining. https://doi.org/10.1145/2808797.2809398

10. Al-Hassan, A., & Al-Dossari, H. (2021). Detection of hate speech in Arabic tweets using deep learning. Multimedia Systems, 28(6), 1963–1974. https://doi.org/10.1007/s00530-020-00742-w

11. Smith, P., Mahdavi, J., Carvalho, M. B., Fisher, S., Russell, S., & Tippett, N. (2008). Cyberbullying: its nature and impact in secondary school pupils. Journal of Child Psychology and Psychiatry, 49(4), 376–385. https://doi.org/10.1111/j.1469-

7610.2007.01846.x

12. Chatzakou, D., Leontiadis, I., Blackburn, J., De Cristofaro, E., Stringhini, G., Vakali, A., & Kourtellis, N. (2019). Detecting Cyberbullying and Cyberaggression in Social Media. ACM Transactions on the Web, 13(3), 1–51. https://doi.org/10.1145/3343484

13. Jazayeri, M. (2007). Some Trends in Web Application Development. In International Conference on Software Engineering. https://doi.org/10.1109/fose.2007.26

14. Hani, J., Nashaat, M., Ahmed, M., Emad, Z., Amer, E., & Mohammed, A. (2019). Social Media Cyberbullying Detection using Machine Learning. International Journal of Advanced Computer Science and Applications, 10(5). https://doi.org/10.14569/ijacsa.2019.0100587

15. Cheng, L. P., Li, J., Silva, Y. N., Hall, D. A., & Liu, H. (2019). XBully. In Web Search and Data Mining. https://doi.org/10.1145/3289600.3291037