Detection of Cyberbullying on social media Using Machine Learning

Munnelli Shanmukanjali[1], Dharpalli Arun Praneeth Kumar[2], Pullareddy Varsheek Reddy[3], Dr.T.S.Mastan Rao[4]

[1,2,3]B.Tech Student, Department of Computer Science and Engineering, CMR Technical Campus, Medchal, Hyderabad, Telangana, India

[4]Associate Professor, Department of Computer Science and Engineering, CMR Technical Campus, Medchal, Hyderabad Telangana, India

Abstract- In the realm of online social networks, it is crucial to conduct research on the detection of anonymous user behavior and offensive content. This particular project focuses on detecting bully statements and offensive data in shared content of social networks. To achieve accurate results, the project proposes a system called "Cyber Bullying Detection (CBD) in Social Networking," which utilizes Machine Learning algorithms and Text Mining concepts. The project employs two datasets, namely the 'Hate Speech and Offensive Language Dataset' and 'Harassment-Corpus Dataset,' and utilizes three Machine Learning classifiers, including Support Vector Machine (SVM), Random Forest (RF), Naïve Bayes (NB), and Neural Network (NN) Algorithms to compare their performance on both datasets. The project also includes the design and development of a Python-based Django web application to demonstrate the system's results.

1.INTRODUCTION

Social networking apps, including chat applications such as WhatsApp and Messenger, provide a range of communication tools for text, media, and web data sharing. In recent years, social networking sites have expanded to offer extensive services such as multimedia and e-commerce. For instance, Twitter is a major platform for micro-blogging, with over 700 million users and 400 million micro-blogs produced daily. However, research has found that over 30% of accounts on social media services like Twitter, Facebook, and Sina are fake or duplicates. Despite this, social sites currently do not prioritize tracking the anonymous behavior of users.

Detecting malicious users has become a significant topic in the study of social media, as current social network sites do not prioritize tracking user behavior, especially that of anonymous users, and the practicality of implementing concepts like profile matching or network-based techniques in real-time is limited. Additionally, crawling user information from microblogs is impractical, and anonymous users can easily manipulate public profile information. Chat applications and multimedia data sharing (e.g., images, videos) are prevalent means of communication among social networking sites, with Facebook and Instagram serving as prime examples, but they also pose a security threat such as cyberbullying.

2. RELATED WORK

In the current trend many social networking sites are created and provide services of communications, multi-media services, e-commerce etc immensely. For example twitter social media provides major services of micro-blogging massively, it has more than 700 million users and 400 million micro-blogs produced per day. According to a research survey, more than 30% of dummy or duplicate or fake accounts are present in all social media services like twitter, facebook, sina etc [1]. But the current social sites do not focus on services like tracking the user behavior or anonymous behavior. In the current system, social network sites need to focus on the user microblogs and need to capture the user behavior whether he/she is an anonymous user or not.

Few surveys' providing concepts to tracking the attackers like using profile matching techniques and network based techniques etc. But in real-time, applying those concepts in social networks is less practical. Crawling the user information from the user micro blogs is also less practical. Anonymous Users can easily manipulate the public profile information.

In the current trend many social networking sites are created and provide services of communications, multi-media services, e-commerce etc immensely. Lot of anonymous user accounts are being created very rapidly. We need to focus on tracking the anonymous users. In our proposed system we are implementing a web based application which will find the anonymous users according to the user behavior. We calculate the user behavior according to the chat statements of the user which he/she does with others. By taking advantage of Machine Learning algorithms we classify the anonymous users. Here we are using Naïve Bayes algorithm to perform the classification of the users.

3.PROPOSED SYSTEM

The proposed system 'Cyberbullying Detection (CBD) in Social Networking' is to identify the bully or offensive statements using a classification model. Based on this requirement, the proposed architecture is designed based on two flows, namely, classification analysis and user side prediction. Based on the project requirement, this architecture is designed which is represented in figure 1. This section described the workflow of the architecture and project main modules. The main purpose of the CBD system is to identify the bully or offensive statements using Machine Learning algorithms. To achieve this requirement, we need to implement classification analysis. The classification analysis is the process of conducting training and testing processes for calculating the performance measures between various Machine Learning algorithms. Based on these requirements, the proposed architecture is designed with two flows, namely, classification analysis which is taken care of by the admin and user side prediction. In the architecture these two flows are represented in two different color formats.

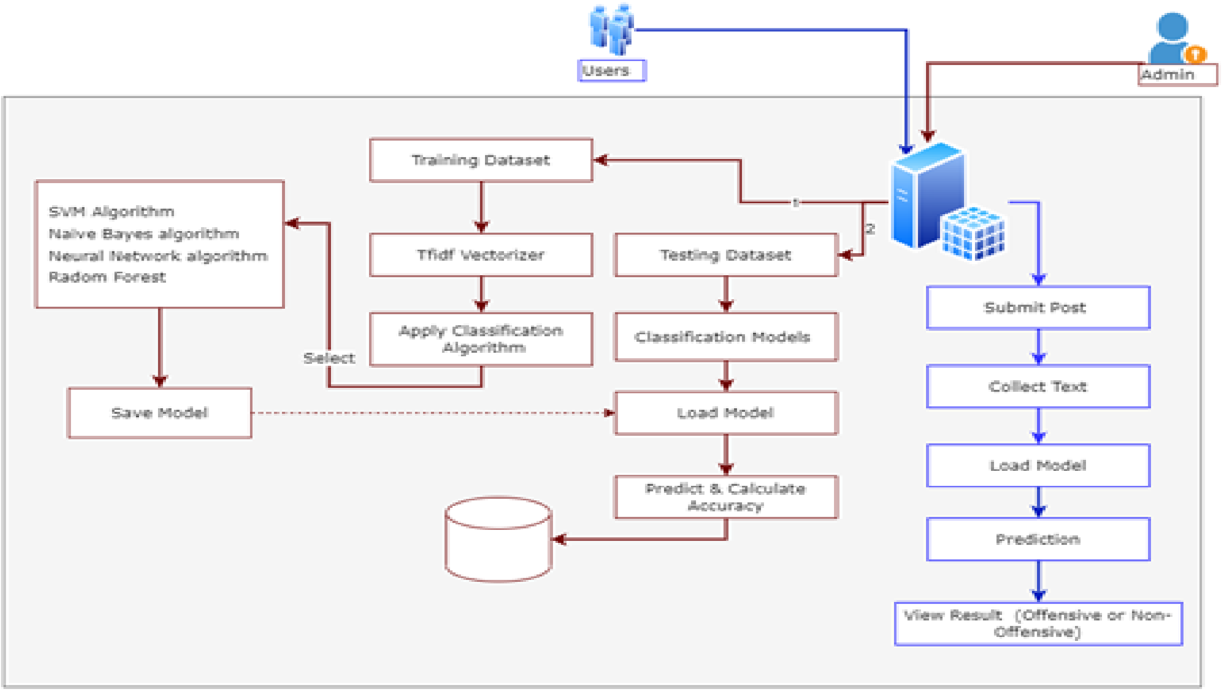


Fig:-1 CBD Architecture

Advantages:

We are depending on the chat history instead of the profile attributes or any other media attributes, so we can conclude with a minimum amount of the computation.

1. Due to machine learning classifications we can get accurate results

2. Cyberbullying detection process is automatic and time taken for detection is less and it works on the live environment.

3. The latest machine learning models are used for training models that are accurate.

Disadvantages:

1.Very less practical, we can't find the attacker using small tiny blogs
2.Based on the network based models we should effort heavy data for detecting
Modules:

Admin:
Admin is a main user of our application, admin will process main functionalities of the system. Admin can see the lists of users are available in the system. Admin can process the detection of anomaly users with Naïve Bayes algorithm. Admin will send a warning mail to the user. After that user if continue the same manner then admin will block the account of the anomaly user.

User:
In our system user is an end user of our application. We build a social application for users. They can make friends and share data among them. User also chat with his/her friends. When a user sends any bullying words to the others application will detect the user and send the warning mail to the user. User will get the notification of warning through mail. After getting the warning alert if user will continue the same manner then user will blocked in the system.

4.EXPERIMENTAL RESULTS

User's Account Creation:

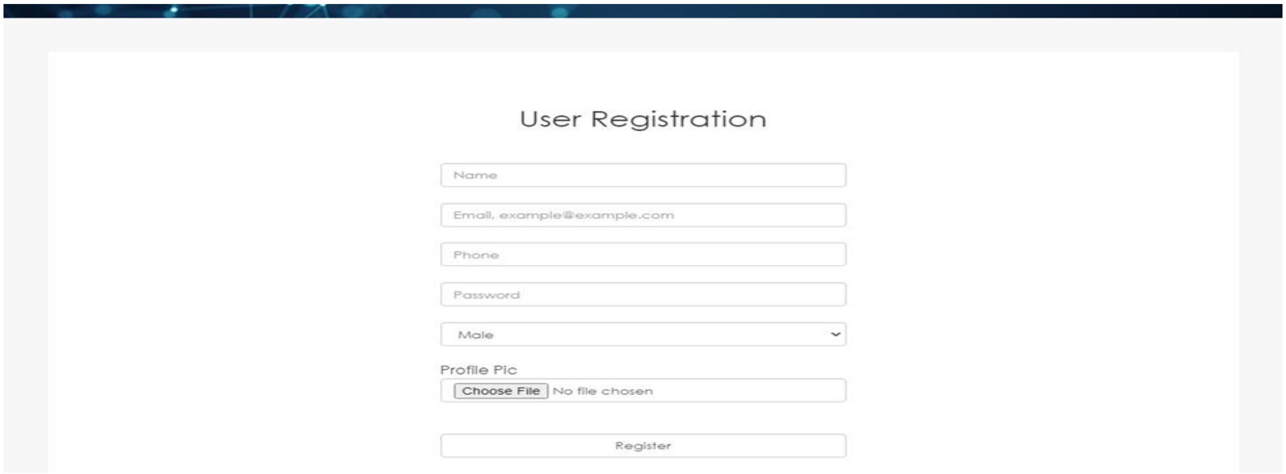


Figure 6.1 User Signup Page

Login page for user verification:

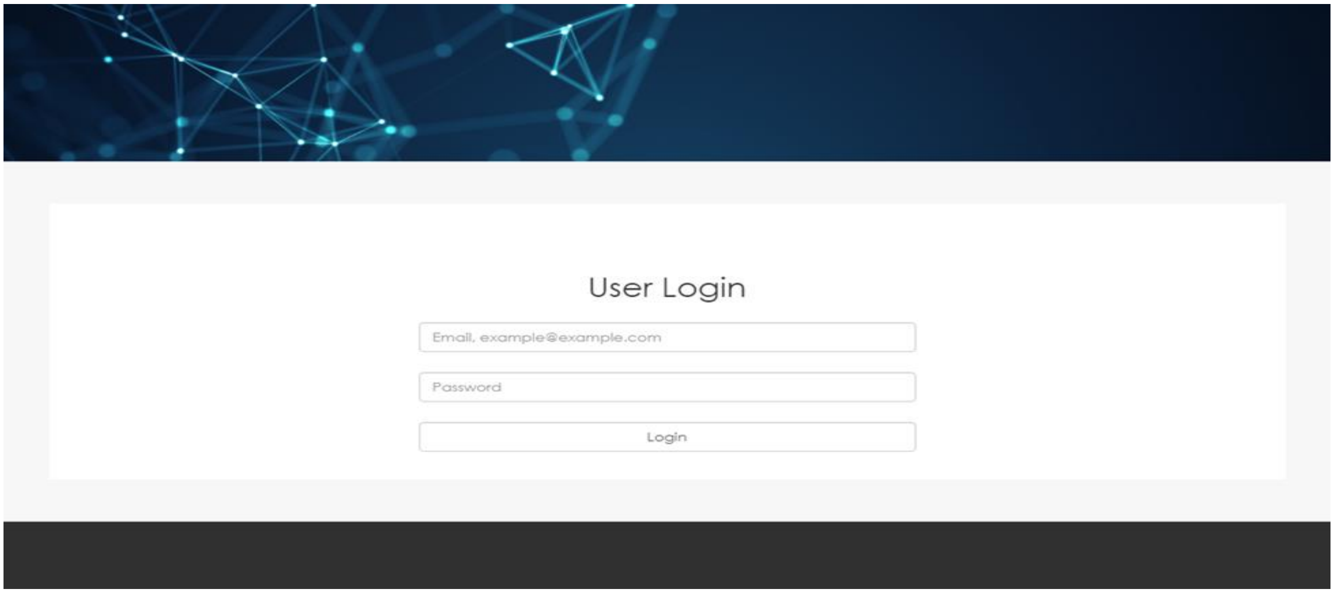
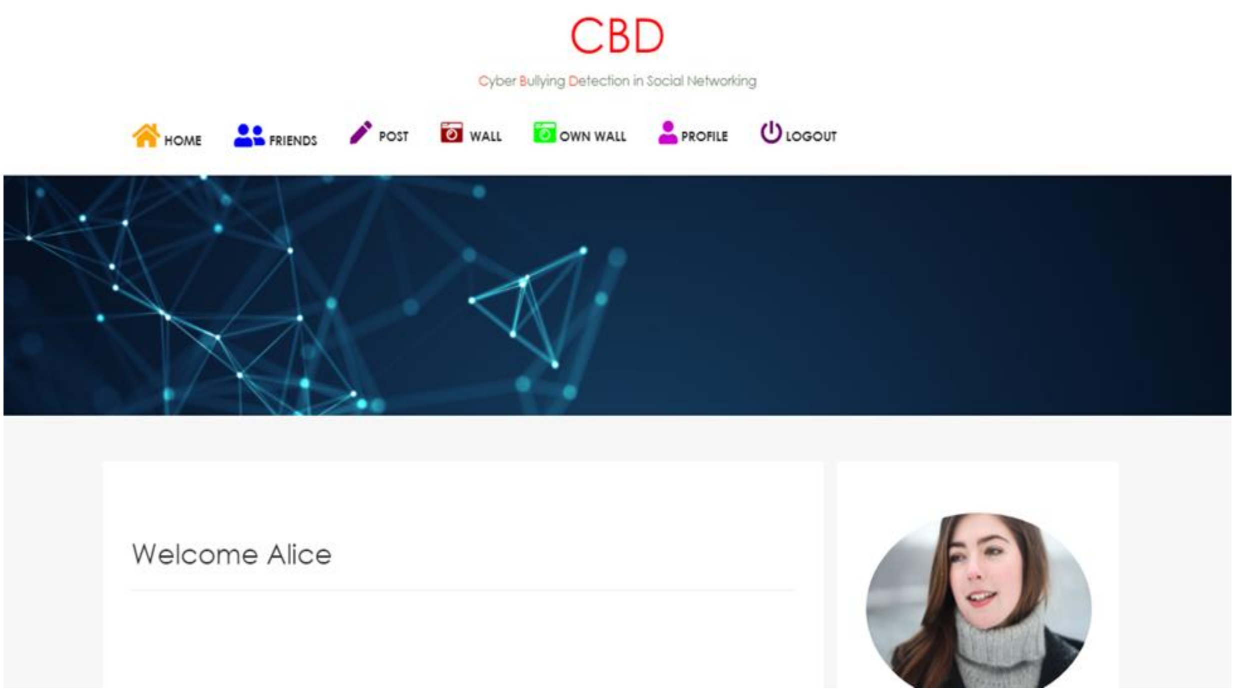


Figure 6.2 User Login Page

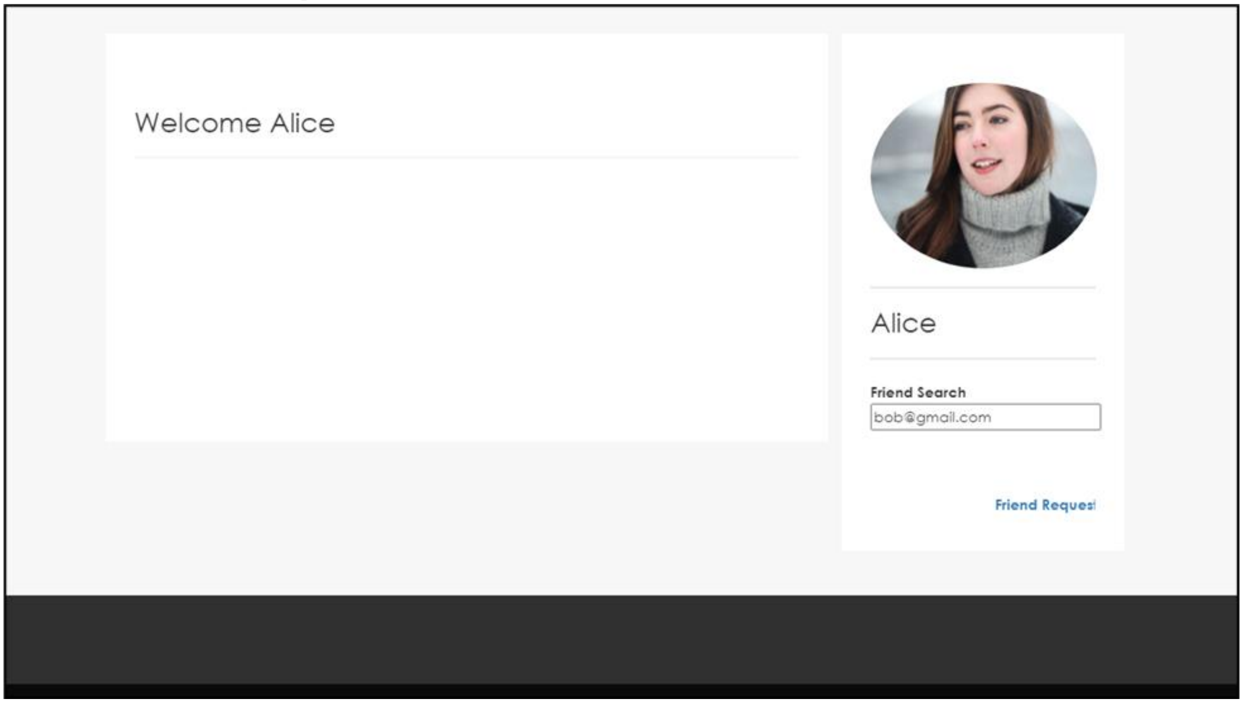
User Homepage



Welcome Alice

Figure 6.3 User Homepage

Search and Send Friend Requests

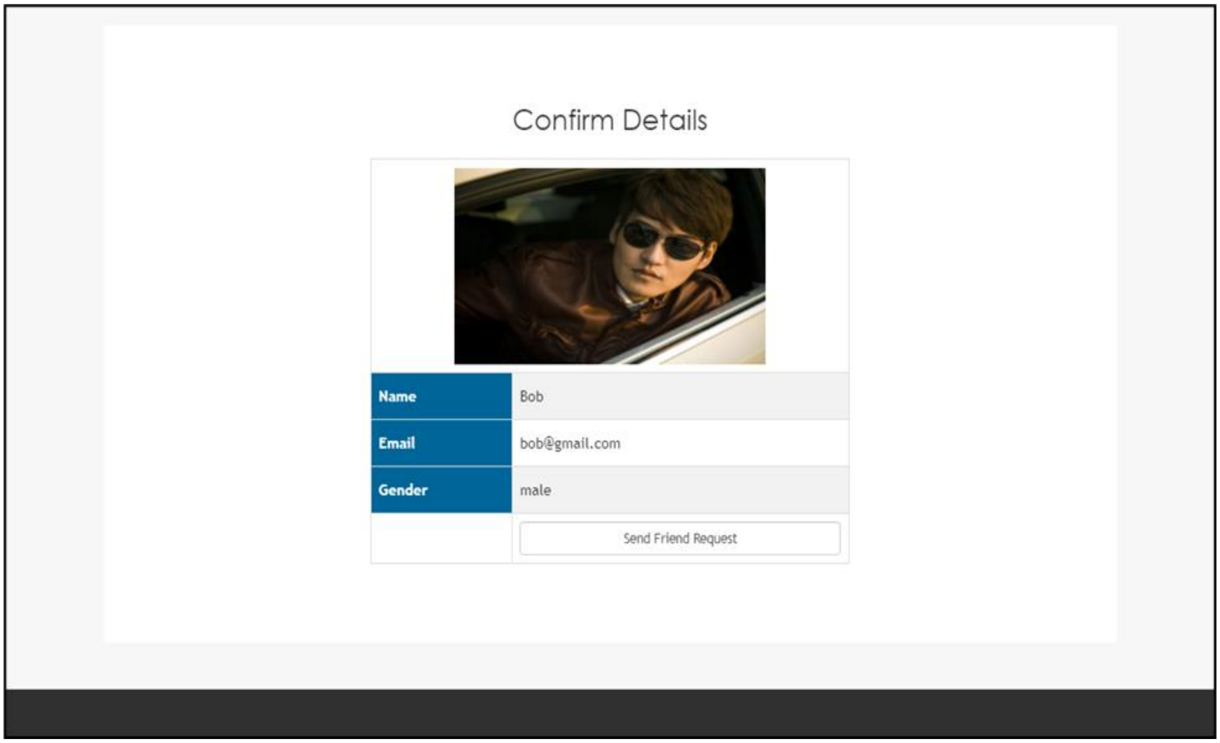


Welcome Alice

Alice

Friend Search: bob@gmail.com

Figure 6.4: Search Friends option

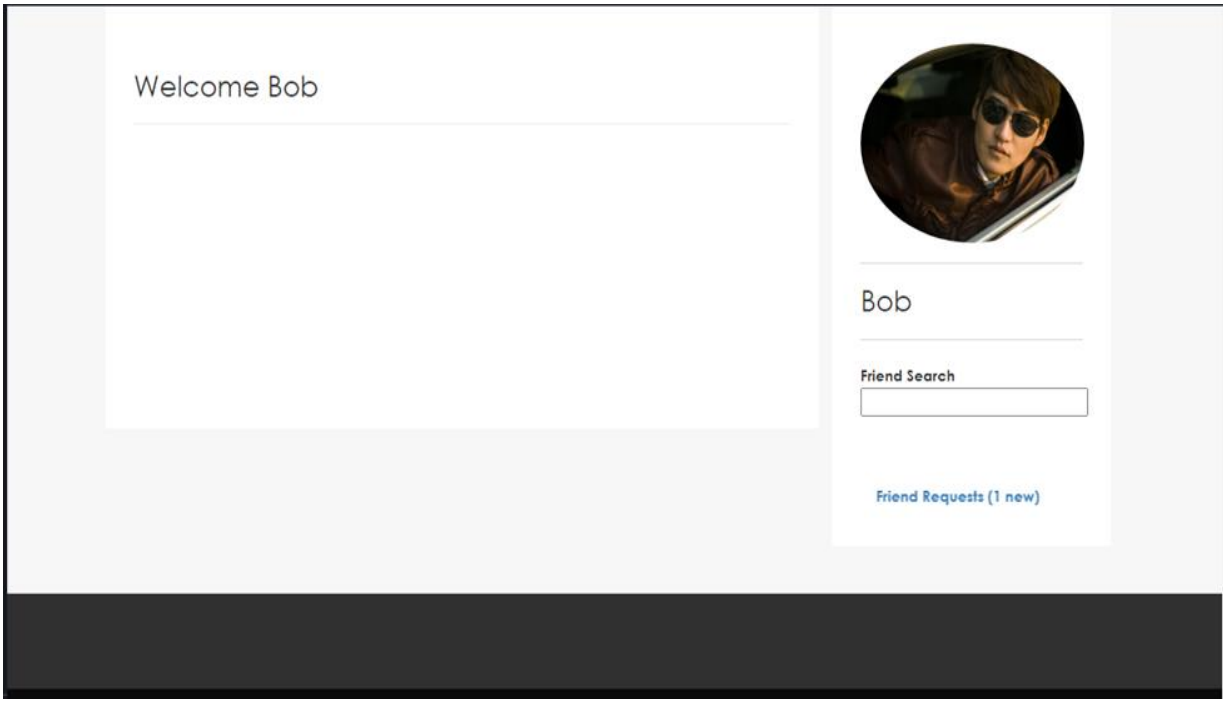


Confirm Details

Name	Bob
Email	bob@gmail.com
Gender	male

Figure 6.5: Send Friend request

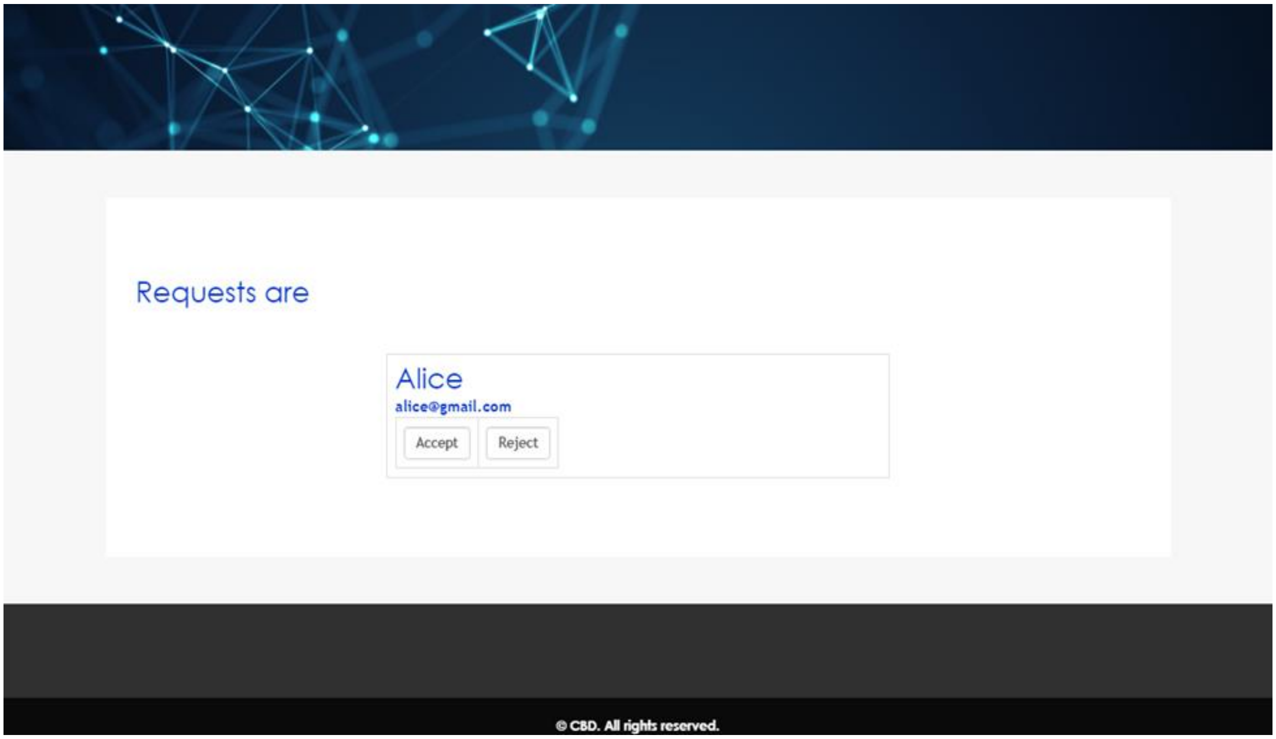
View Friend Requests and Decision on requests



Welcome Bob

Bob

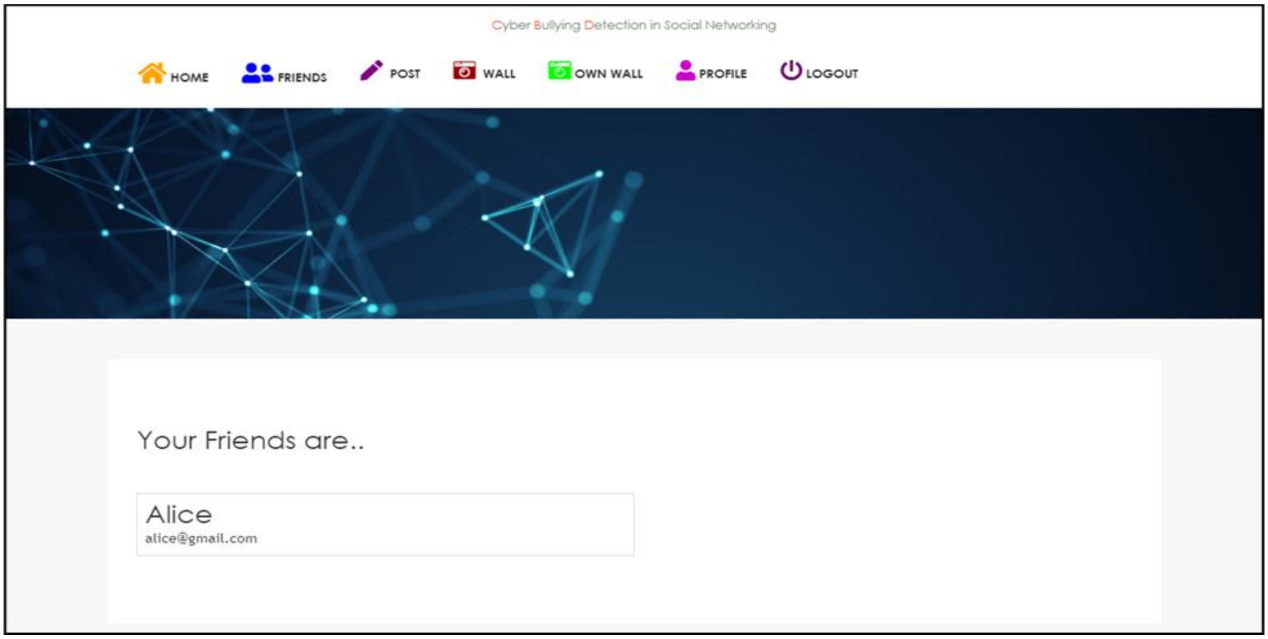
Figure 6.6: View Friend request



Requests are

Alice
alice@gmail.com

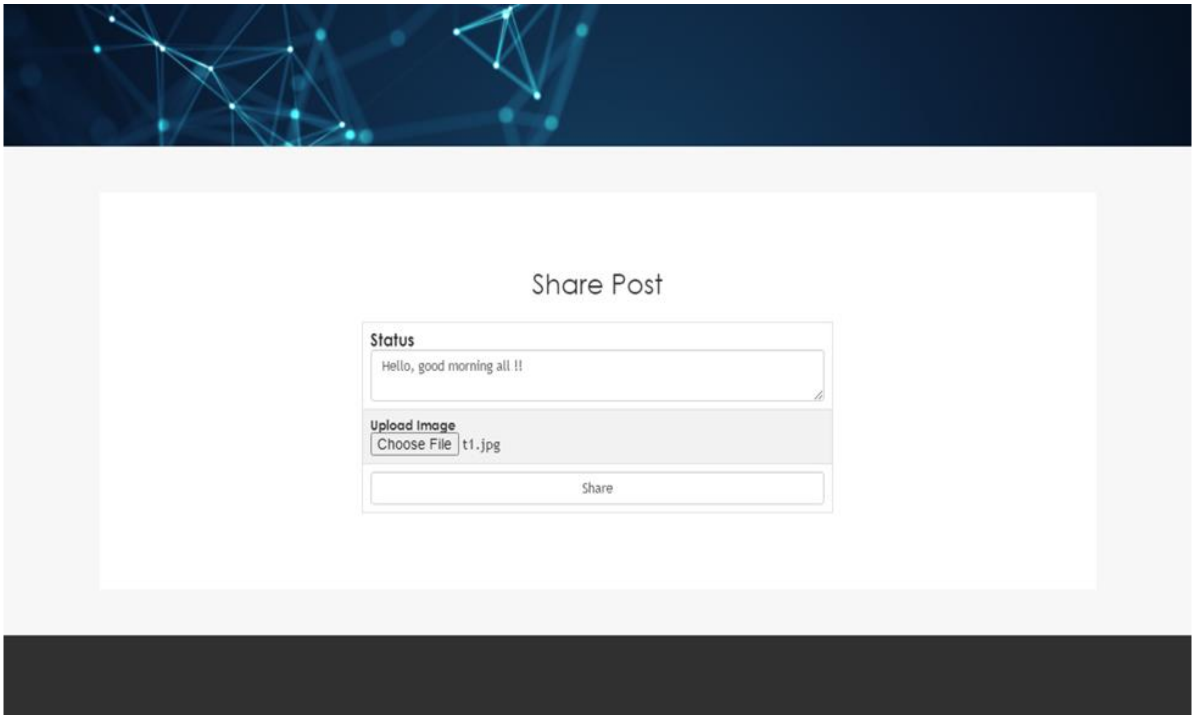
Figure 6.7: Decision on friend request



Your Friends are..

Alice
alice@gmail.com

Figure 6.8: Decision on friends

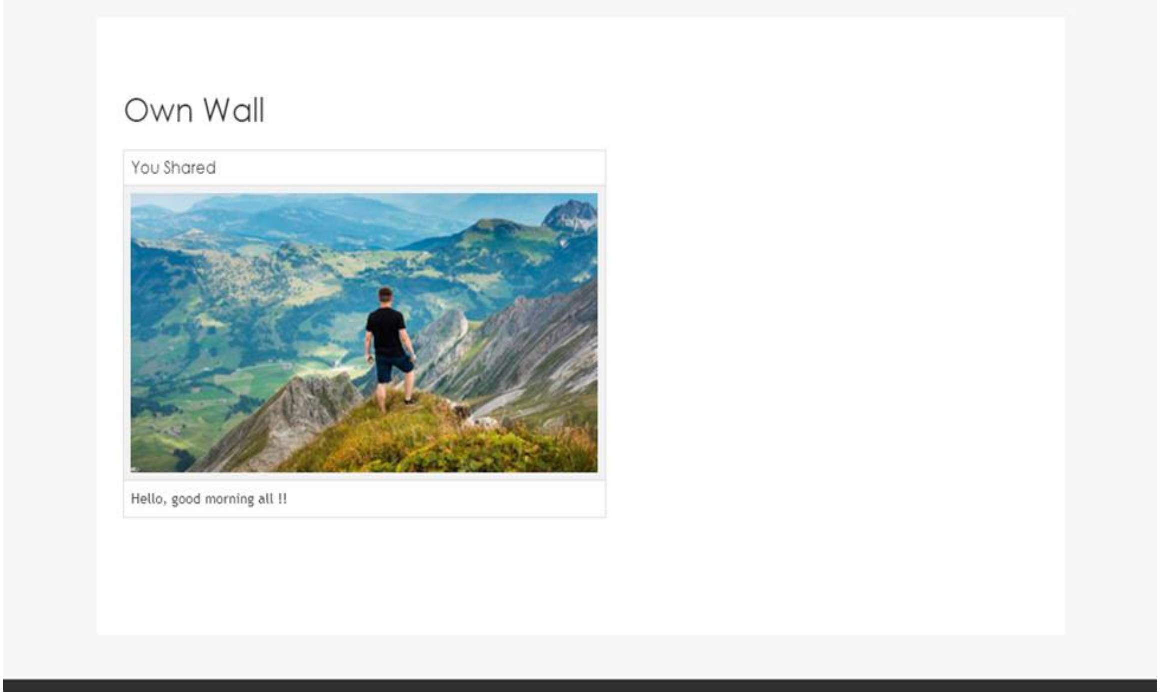


Share Post

Status: Hello, good morning all !!

Upload Image: t1.jpg

Figure 6.9: Post Share

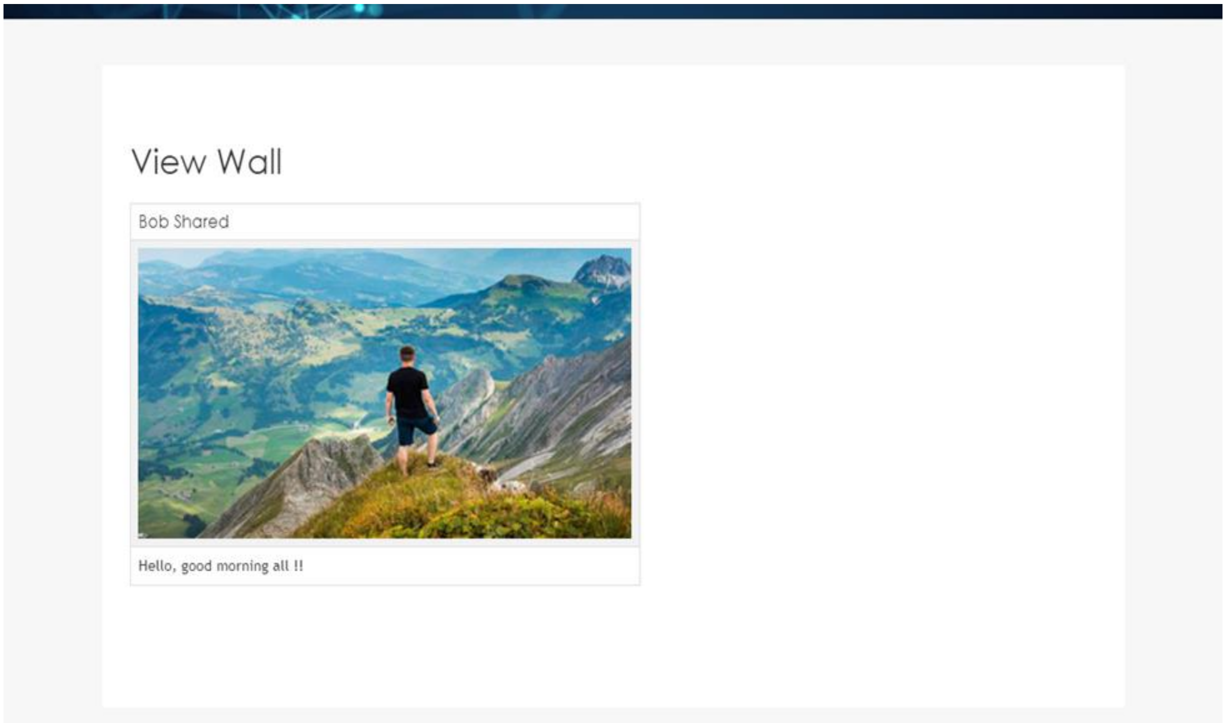


Own Wall

You Shared

Hello, good morning all !!

Figure 6.10: View Own Wall



View Wall

Bob Shared

Hello, good morning all !!

Figure 6.11: View Wall

5.CONCLUSION

In current trend many social networking sites are created and provide services of communications, multi-media services, e-commerce etc immensely. Lot of anonymous user accounts are being created very rapidly. We need to focus on tracking the anonymous users. In our project we have calculated the user behavior according to the chat statements of the user which he/she does with others. By taking advantage of Machine Learning algorithms we classify the anonymous users. Here we are using Naïve Bayes algorithm to perform the classification of the users.

6.ACKNOWLEDEMENTS

We express our gratitude to CMR Technical Campus for their assistance in our project titled "Detection of cyberbullying on social media using machine learning". We extend our sincere thanks to the Chairman, Director, Deans, Head of the Department, Department of Computer Science and Engineering, our guide, and the teaching and non-teaching faculty members for their invaluable suggestions and guidance throughout our work.

REFERENCES

[1] T. Iofciu, P. Fankhauser, F. Abel, and K. Bischoff, “Identifying users across social tagging systems,” in Proc. 5th Int. AAAI Conf. Weblogs Social Media, 2011, pp. 522–525.

[2] Identifying Users Across Social Tagging Systems Tereza Iofciu, Péter Fankhauser, +1 author Kerstin Bischoff, prieto et al. 2009

[3] D. Perito, C. Castelluccia, M. A. Kaafar, and P. Manila, “How unique and traceable are usernames?” in Proc. 11th Int. Conf. Privacy Enhancing Technology., 2011, pp. 1–17.

[4] J. Liu, F. Zhang, X. Song, Y. I. Song, C. Y. Lin, and H. W. Hon, “What’s in a name?: An unsupervised approach to link users across communities,” in Proc. 6th ACM Int. Conf. Web Search Data Mining, 2013, pp. 495–504.