DETECTION OF CYBERBULLYING ON SOCIAL MEDIA USING MACHINE LEARNING AND DEEP LEARNING TECHNIQUE

Dr. Syed Asadullah Hussaini[1],,Summiya Sultana[2], Syeda Maleka Husna[3], Syeda FarahUnnisa[4]

[1]Associate Professor of CSE, ISL College of Engineering, Hyderabad, India.

[2,3,4,]B. E Student, Department of CSE, ISL College of Engineering, India.

Summiyasultana9@gmail.com, syedamaleka.227@gmail.com, farahsyeda54@gmail.com

Abstract: In the research of online social networks, detection of anonymous user behavior, detection of offensive data, etc. are traditional and important research works. This project is focused on the detection of offensive data, and bully statements in shared data of the social networks. For this system, using Machine Learning algorithms with Text Mining concepts predicted the offensive data to get more accurate results. This project proposed a system of "Cyber Bullying Detection (CBD) in Social Networking" for predicting bully data. The user must be blocked if the user uses offensive words for a threshold number of times, the model should be able to detect the words containing * marks. The model is to be built & comprise between ML (Support Vector Machine (SVM), Random Forest (RF), Naïve Bayes (NB), and Neural Network (NN)) and DL (ANN), CNN) This project is used two datasets, namely, 'Hate Speech and Offensive Language Dataset' and 'Harassment-Corpus Dataset'. Algorithms and calculated performance results for comparing the performance for both datasets. To demonstrate this system designed and developed a Python-based Django web application and showed the results.

1. INTRODUCTION

Many internet chat programs, such as WhatsApp, Messenger, and other social networking apps, include communication facilities for exchanging text messages, sharing media material, and sharing online data. Currently, several social networking sites have been established and are offering a wide range of services such as communication, multimedia services, and e-commerce on a large scale. Twitter, a popular social media platform, offers extensive micro-blogging capabilities. With over 700 million users, it generates a staggering 400 million micro-blogs daily. Research surveys indicate that a significant proportion, over 30%, of social media accounts across platforms such as Twitter, Facebook, and Sina are comprised of dummy, duplicate, or false profiles. [1]. However, the existing social media platforms do not prioritize services that include monitoring the anonymous activity of users. In the present system, social network sites must prioritize user microblogs and gather information about user activity, including whether they are anonymous or not. Some studies provide strategies for tracing attackers, such as using profile matching tools and network-based approaches. However, implementing such notions in real-time on a social network is not very feasible.

Extracting user information from microblogs is also not very feasible. Unidentified Users have the ability to effortlessly alter the public profile information. On social networks, users may exchange their communications via chat programs. Each social networking site has its own chat application, with Facebook being a prominent example. Another method involves the exchange of multimedia material, such as photographs or movies. These are the greatest examples for Facebook and Instagram. Chat programs are very beneficial for facilitating communication between users, allowing them to easily exchange information, opinions, views, and more. However, it may also lead to a security vulnerability in the user's safety, namely in the form of cyberbullying. Text-based material may pose a security hazard to users since individuals may utilize phony accounts to send cyberbullying messages. The references are [2] and [3]. Given these drawbacks, the identification of malevolent users is a prominent subject of research in the field of social media.

Existing System

Currently, numerous social networking sites include communications, multi-media, e-commerce, and more. Twitter, with 700 million users and 400 million microblogs every day, provides huge micro-blogging services. More than 30% of social media accounts like Twitter, Facebook, and Sina are false, according to study. However, modern social platforms do not monitor anonymous user activity. Current social media platforms must concentrate on user microblogs and collect user activity, whether anonymous or not. Few surveys include attacker tracking notions like profile matching and network-based approaches. Applying similar notions in real time to social networks is less practicable.

Proposed System

The suggested approach ‘Cyber Bullying Detection (CBD) in Social Networking’ uses a classification model to detect bullies and offensive utterances. Based on this need, the suggested architecture uses categorization analysis and user-side prediction. Figure 1 shows this architecture based on project requirements. This section describes architecture and project major module process. The CBD system uses Machine Learning algorithms to detect bullies and inappropriate utterances. Implement classification analysis to meet this need. Classification analysis involves developing and evaluating Deep Learning systems to compare performance. Based on these criteria, the suggested architecture has two flows: admin-handled categorization analysis and user-side prediction. The architecture (Figure 1) shows these two movements in two colors.

2. RELATED WORK

Currently, numerous social networking sites include communications, multi-media, e-commerce, and more. Twitter, with 700 million users and 400 million microblogs every day, provides huge micro-blogging services. Research shows that over 30% of social media accounts like Twitter, Facebook, and Sina are false [1]. However, modern social platforms do not monitor anonymous user activity. Current social media platforms must concentrate on user microblogs and collect user activity, whether anonymous or not.

Few surveys include attacker tracking notions like profile matching and network-based approaches. Applying similar notions in real time to social networks is less practicable. Also impractical is crawling user micro blogs for information. Anonymous Users may quickly edit public profiles.

Currently, numerous social networking sites include communications, multi-media, e-commerce, and more. Many anonymous user accounts are created quickly. Our goal should be monitoring anonymous users. Our suggested web-based program finds anonymous people based on user behavior. We compute user behavior based on chat sentences used with others. We categorize anonymous users using Machine Learning methods. We utilize Naïve Bayes method to classify people.

Problem Statement

Currently, numerous social networking sites include communications, multi-media, e-commerce, and more. Twitter, with 700 million users and 400 million micro-blogs every day, provides huge micro-blogging services. Research shows that over 30% of social media accounts like Twitter, Facebook, and Sina are false [1]. However, modern social platforms do not monitor anonymous user activity. Current social media platforms must concentrate on user microblogs and collect user activity, whether anonymous or not. Few surveys include attacker tracking notions like profile matching and network-based approaches. Applying similar notions in real time to social networks is less practicable. Also impractical is crawling user micro blogs for information. Anonymous Users may quickly edit public profiles.

SYSTEM ANALYSIS

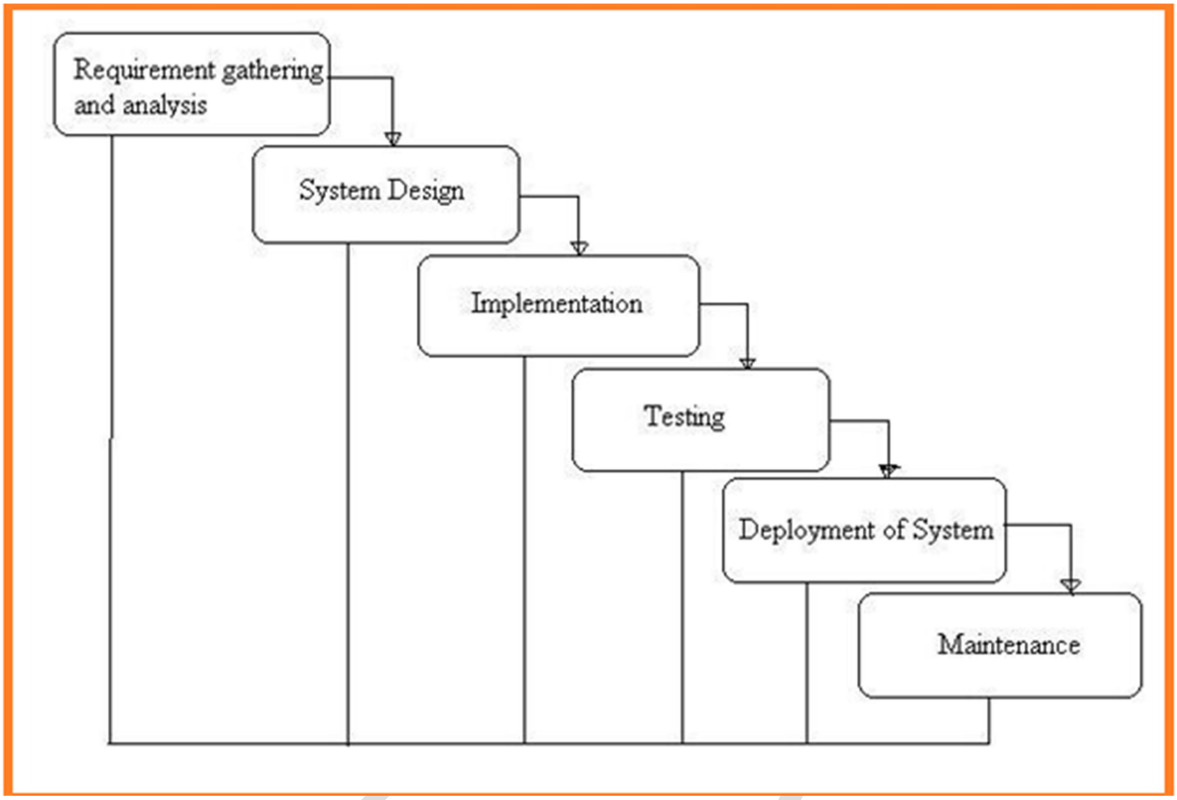


Fig:-1 Project SDLC

Requisites Accumulating and Analysis

This is the most important stage of any project because it's an academic leave for gathering requirements. We followed IEEE journals and gathered a large number of IEEE relegated papers before culling a paper titled "Individual web revisitation by setting and substance importance input." For the analysis stage, we gathered references from the paper and conducted a literature review of a few papers. In this stage, we gathered all the project requirements.

System Design

System design is divided into three categories: class diagrams, which provide information about various project classes and methods that must be used, GUI design, UML design, which facilitates the development of projects in an easy-to-understand manner with various actors and their utilizer cases through utilizer case diagrams, and flow of the project using sequence. The third and last step in the system design project is data base design, where we try to create a database based on the quantity of modules in our project.

IMPLEMENTATION

Architecture

In this architecture I describe the flow of the process of the project. This architecture will describe the online social user process of the chat procedure and detection of anonymous users, in this flow there is a manual procedure and another one is system procedure in the detection process. Let's describe the architecture components.

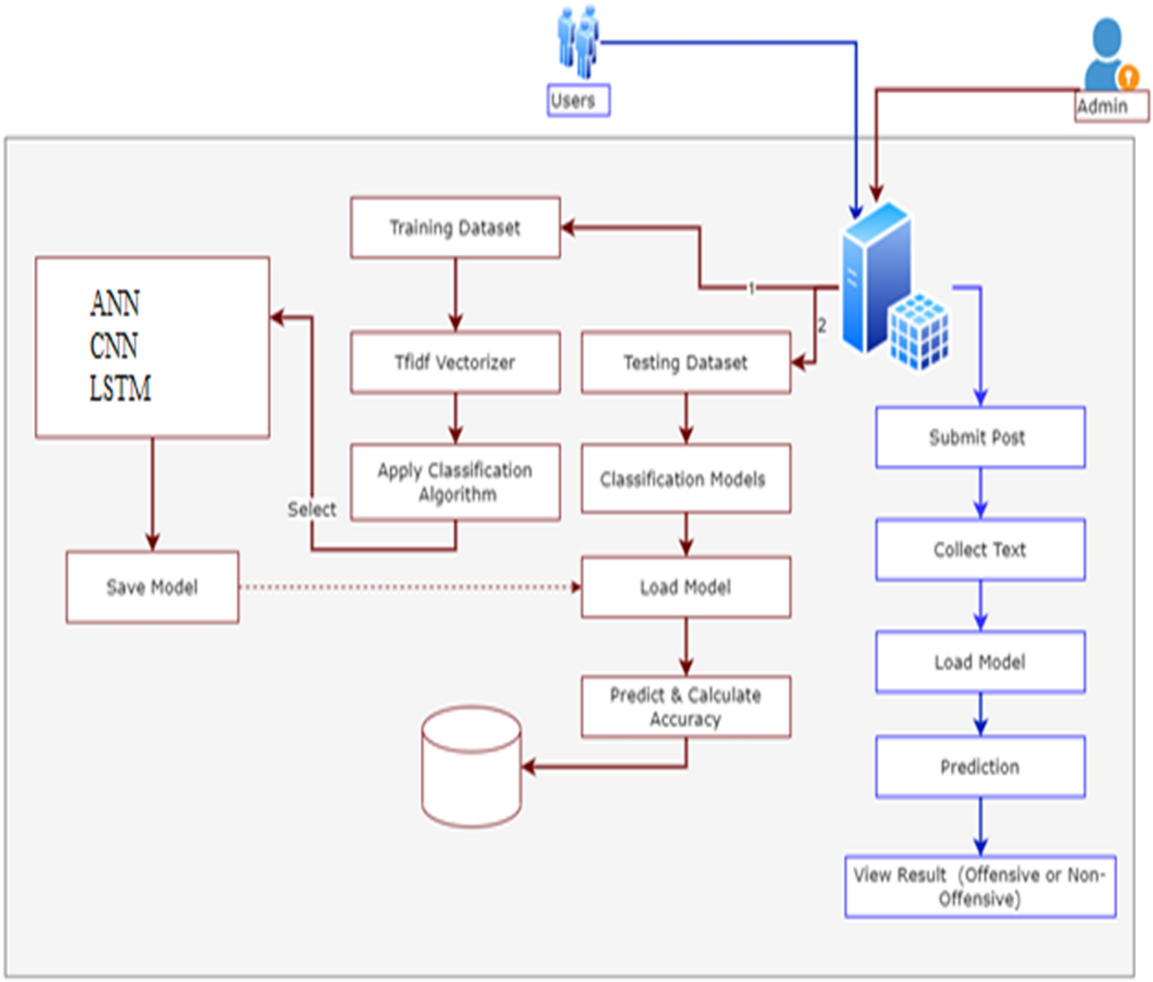
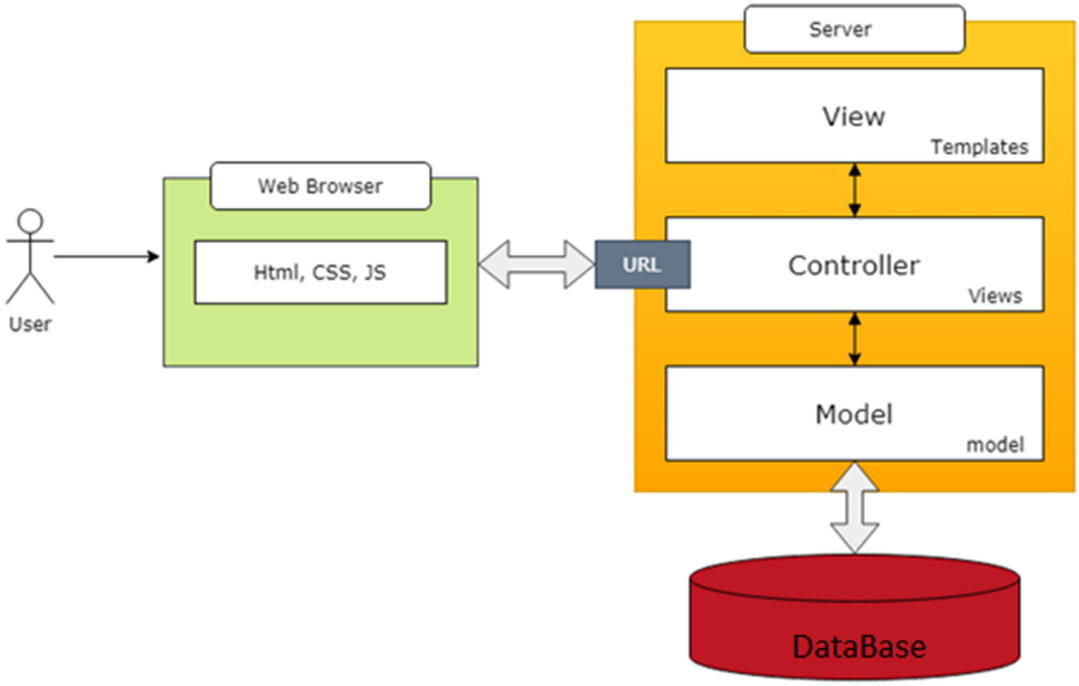


Fig:-1 CBD Architecture

The suggested approach 'Cyber Bullying Detection (CBD) in Social Networking' uses a classification model to detect bullies and offensive utterances. Based on this need, the suggested architecture uses categorization analysis and user-side prediction. Figure 1 shows this architecture based on project requirements. This section describes architecture and project major module process. The CBD system uses Machine Learning algorithms to detect bullies and inappropriate utterances. Implement classification analysis to meet this need. Classification analysis involves training and evaluating Machine Learning algorithms to

compare performance. Based on these criteria, the suggested architecture has two flows: admin-handled categorization analysis and user-side prediction. The architecture (Figure 1) shows these two movements in two colors.



The basic levels of Testing:

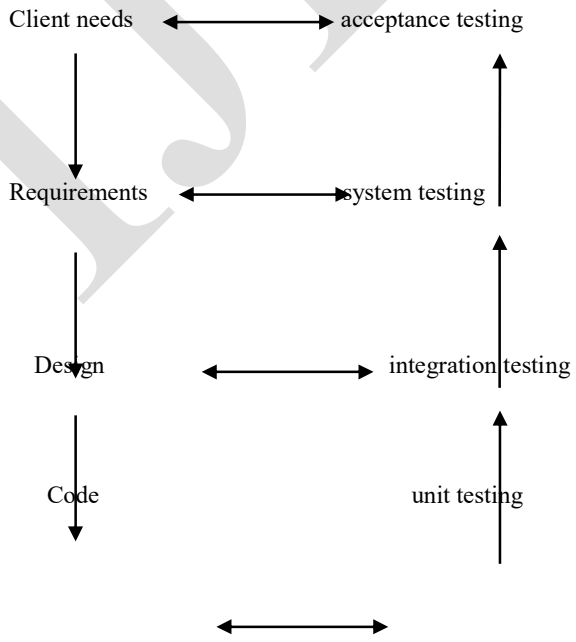


Figure: Levels of Testing

Manual testing on project application

TEST CASES

Test Case ID #1		**Test Case Description** - Validations in Registration Form		
S#	**Prerequisites**	**S#**	**Test Data Requirement**	
1	User should be Registered	1	Data should be valid	
Test Condition				
Entering data in registration form				
Step #	**Step Details**	**Expected Results**	**Actual Results**	**Pass/Fail/Not Executed/Suspended**
1	User gives First and Last Name	Pop showing email verification message	Enter valid email/password	Fail
2	Submitting the form without entering any details	Pop showing email verification message	Enter email /password	Fail
3	User enters invalid format of email id	Pop showing email verification message	Enter valid email id	Fail
4	User enters a phone number with < 10 digits	Pop showing email verification message	Enter valid phone number	Fail
5	Entering valid username and password	Pop showing email verification message	Pop showing email verification message	Pass

Table 1 Registration test case

Test Case ID #2		**Test Case Description** - Validations in Login Form		
S#	**Prerequisites**	**S#**	**Test Data Requirement**	
1	User should have an email id	1	Data should be valid	
Test Condition				
Entering data in login form				
Step #	**Step Details**	**Expected Results**	**Actual Results**	**Pass/Fail/Not Executed/Suspended**
1	User gives aemail	User logged in	Enter valid	Fail

	or password of <6 characters		email/password	
2	Submitting the form without entering any details	User logged in	Enter email /password	Fail
3	User enters wrong Email and (or) password	User logged in	Enter correct email /password	Fail

Table 2 Login test case

Results screen

User's Account Creation

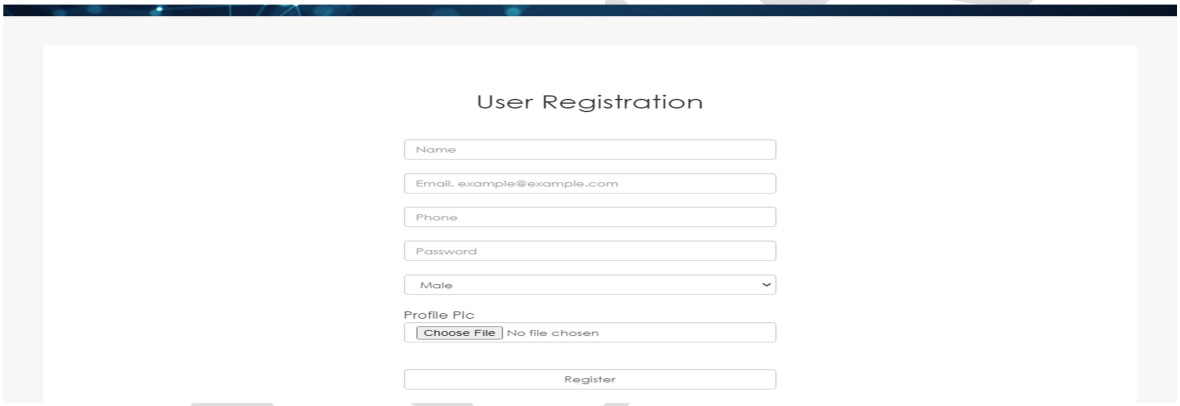


Figure 1 User Signup Page

or user verification

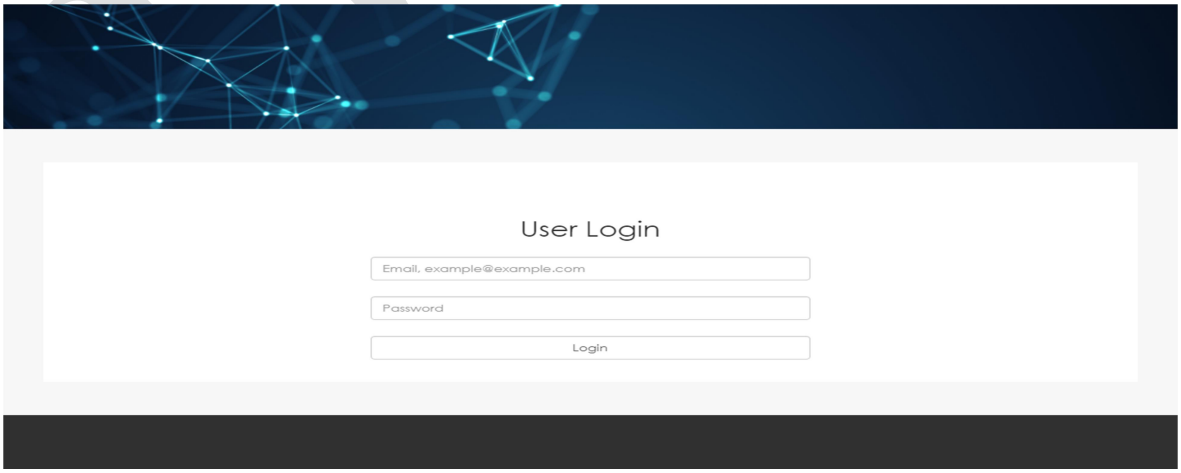


Figure 2 User Login Page

User Homepage

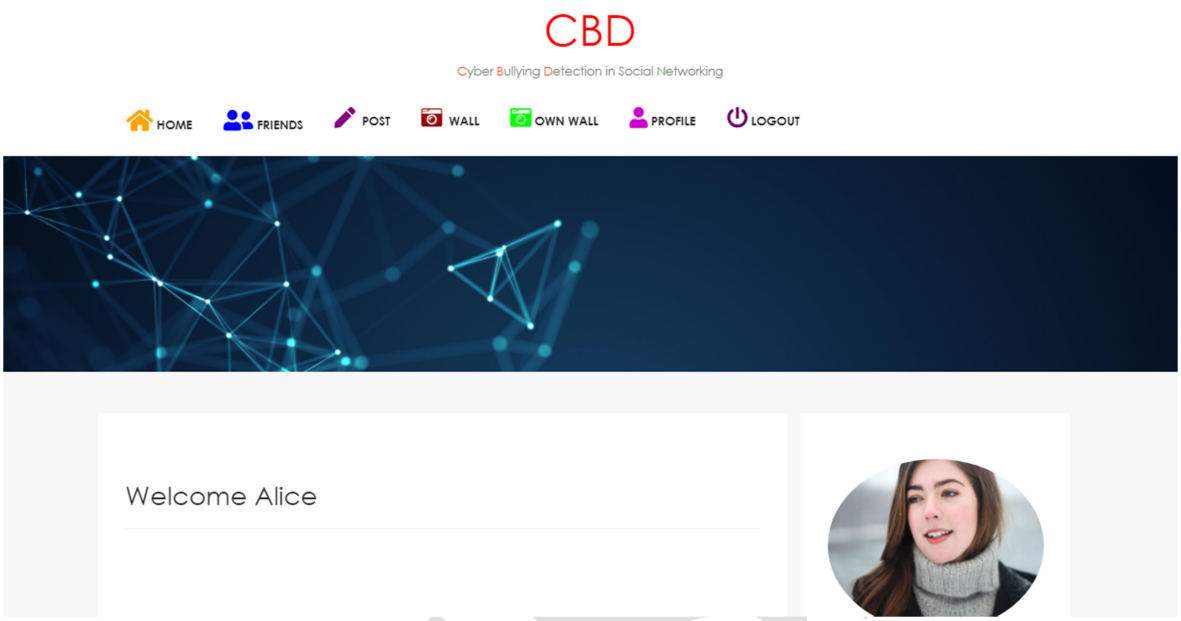


Figure 3 User Homepage

Search and Send Friend Requests

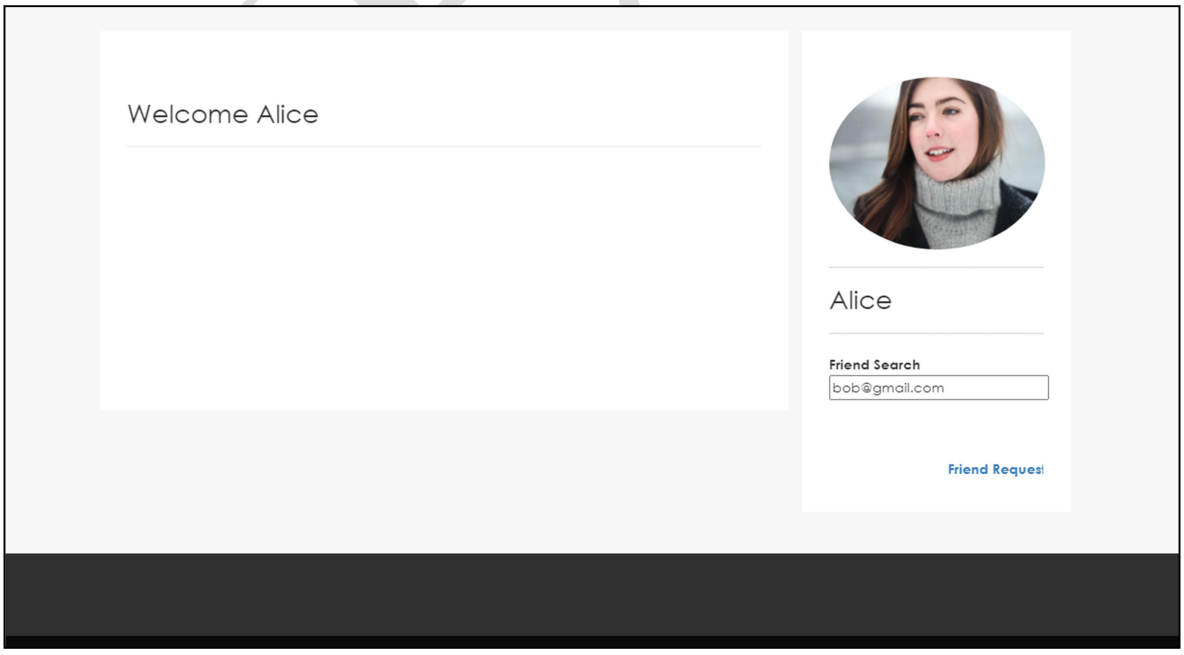


Figure 4: Search Friends option

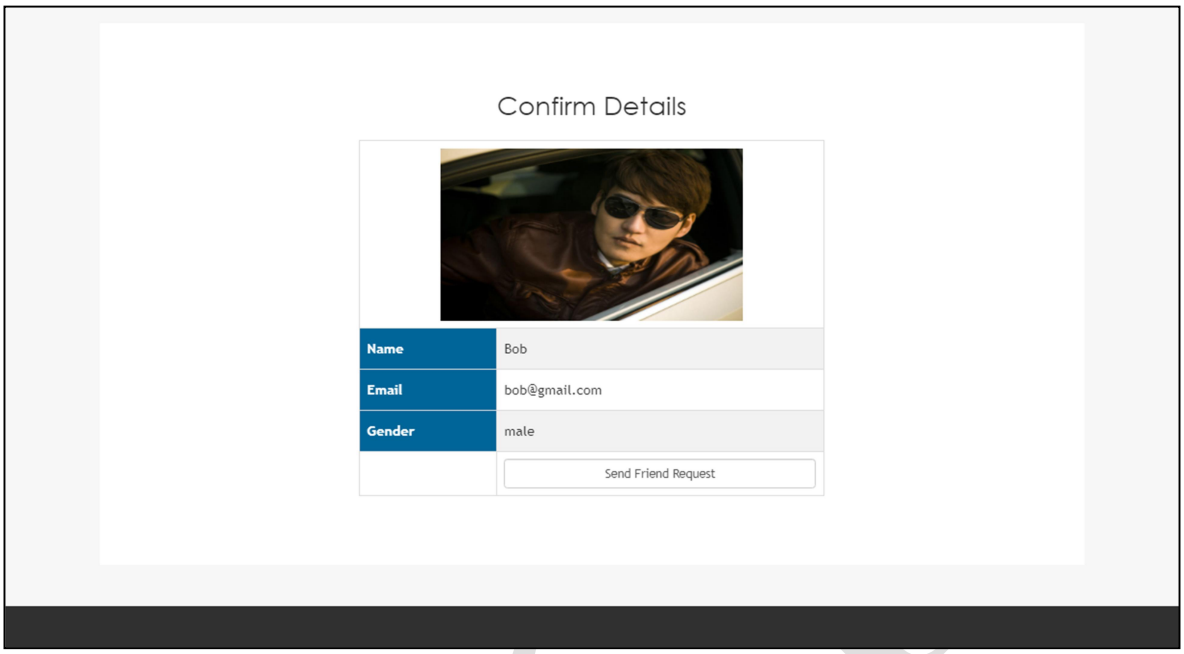


Figure 5: Send Friend request

View Friend Requests and Decision on requests

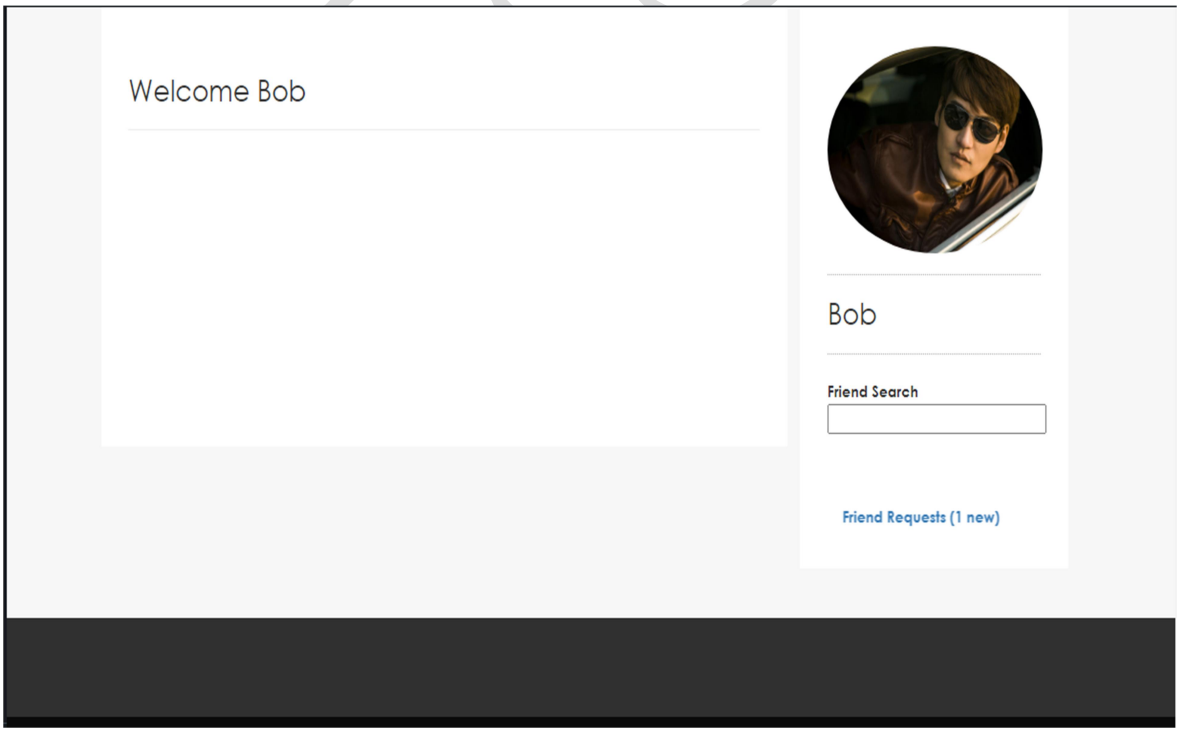


Figure 6: View Friend request

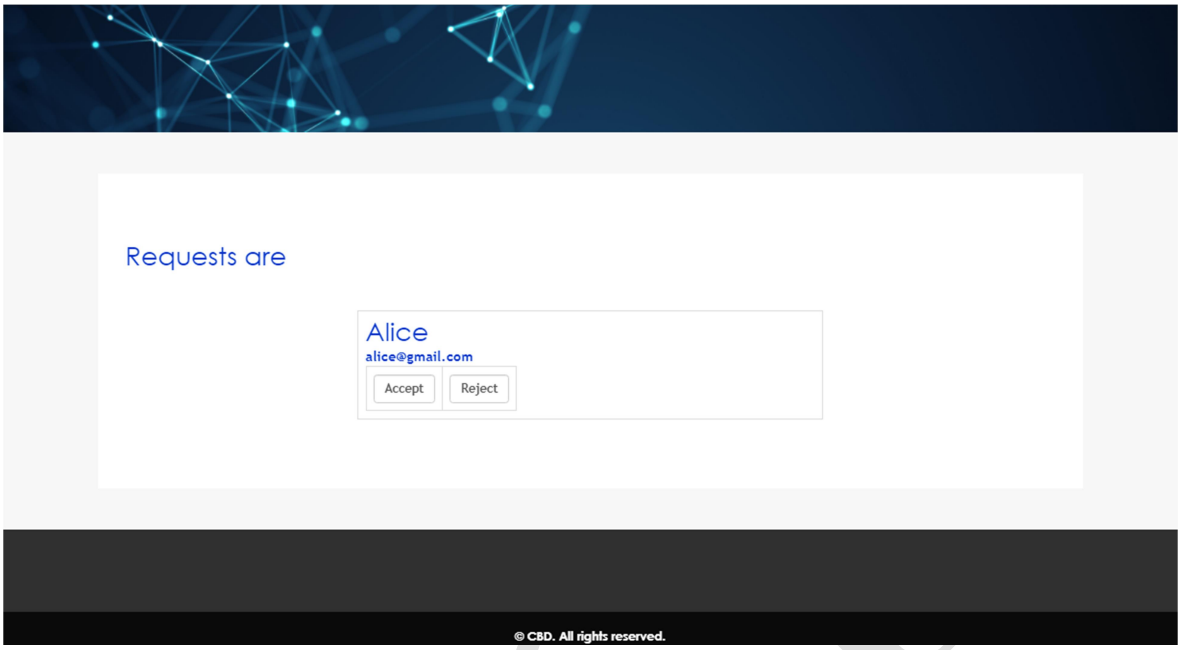


Figure 7: Decision on friend request

View Friends

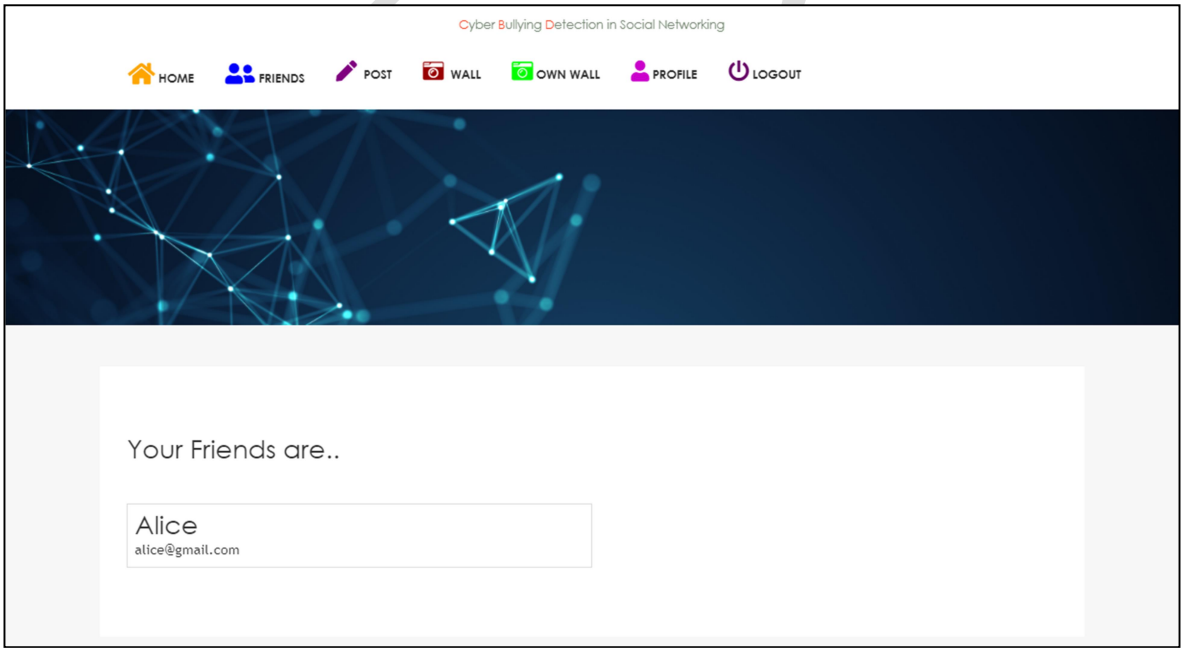


Figure 8: Decision on friend request

Post Share

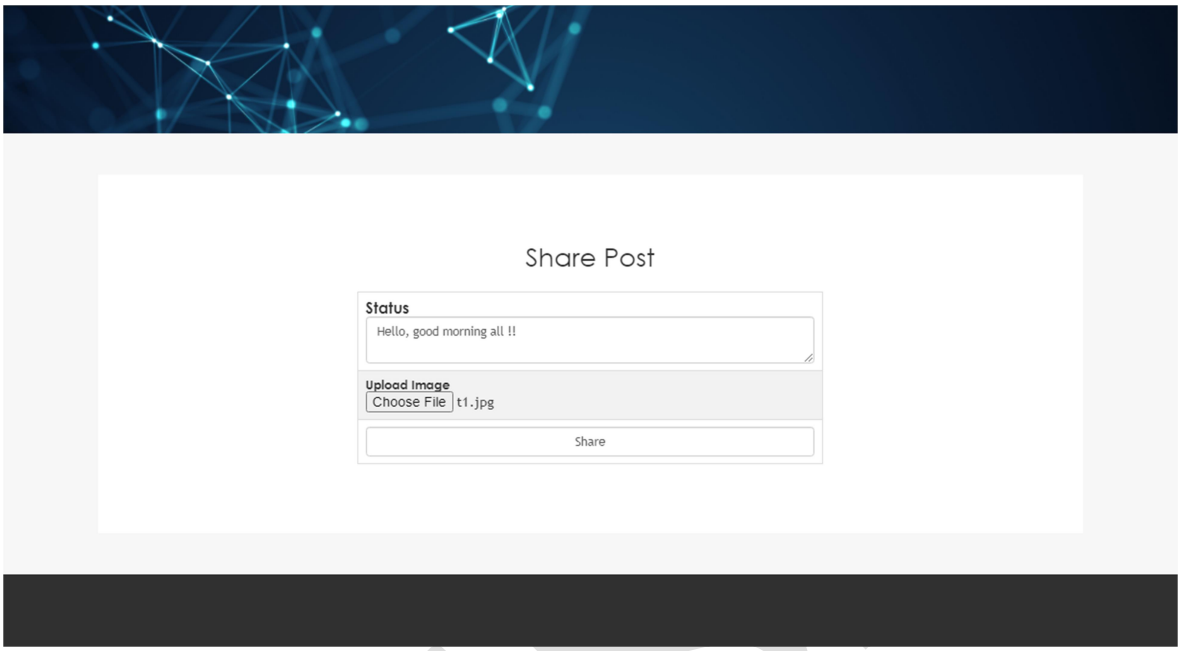


Figure 9: Post Share

Own Wall

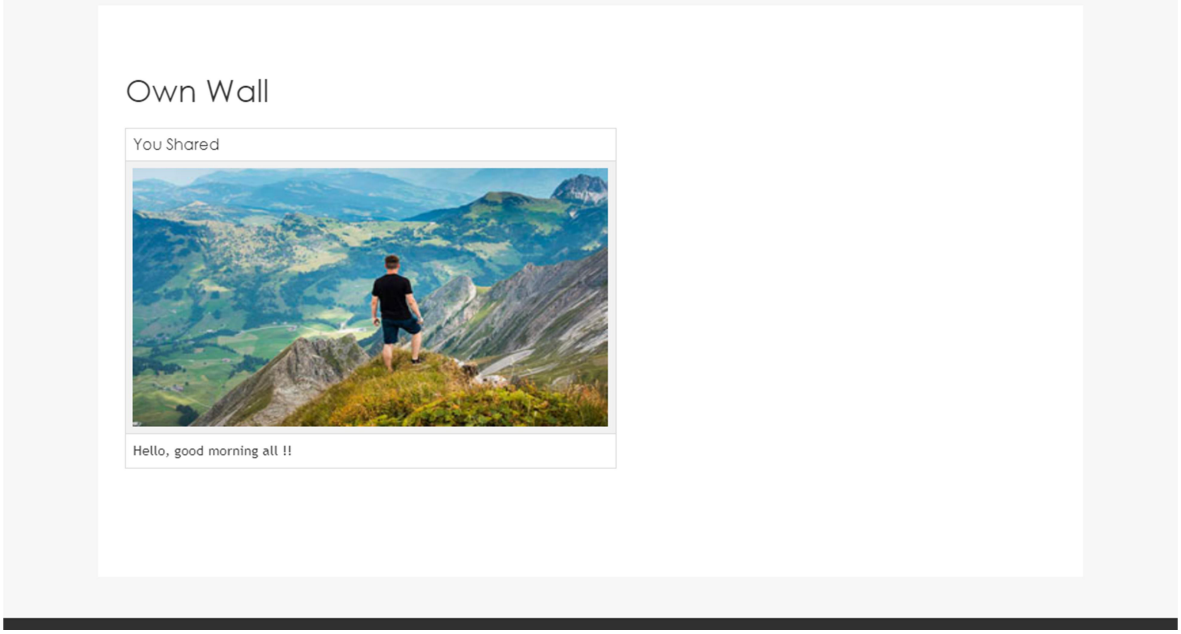


Figure 10: View Own Wall

View Wall

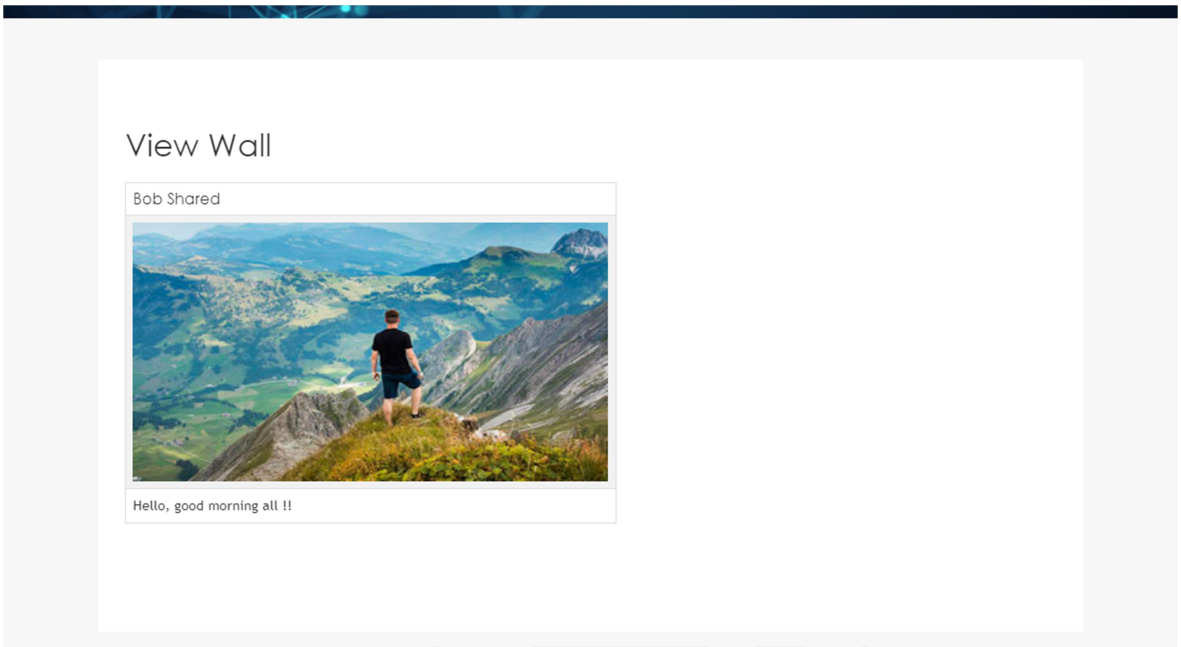


Figure 11: View Wall

CONCLUSION

Currently, several social networking sites have been established and are offering a wide range of services such as communication, multimedia services, and e-commerce on a large scale. A large number of anonymous user accounts are being created at a quick pace. We must prioritize the surveillance of anonymous users. Our research involves analyzing user behavior based on their chat interactions with others. By using the benefits of Deep Learning algorithms, we are able to categorize anonymous users.

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. Manils, “How unique and traceable are usernames?” in Proc. 11th Int. Conf. Privacy Enhancing Technol., 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.

5. Ijteba Sultana, Dr. Mohd Abdul Bari ,Dr. Sanjay," *Routing Performance Analysis of Infrastructure-less Wireless Networks with Intermediate Bottleneck Nodes*", International Journal of Intelligent Systems and Applications in Engineering, ISSN no: 2147-6799 IJISAE,Vol 12 issue 3, 2024, Nov 2023

6. Md. Zainlabuddin, "*Wearable sensor-based edge computing framework for cardiac arrhythmia detection and acute stroke prediction*", Journal of Sensor, Volume2023.
7. Md. Zainlabuddin, "*Security Enhancement in Data Propagation for Wireless Network*", Journal of Sensor, ISSN: 2237-0722 Vol. 11 No. 4 (2021).
8. Dr MD Zainlabuddin, "*CLUSTER BASED MOBILITY MANAGEMENT ALGORITHMS FOR WIRELESS MESH NETWORKS*", Journal of Research Administration, ISSN:1539-1590 | E-ISSN:2573-7104 , Vol. 5 No. 2, (2023)
9. Vaishnavi Lakadaram, " Content Management of Website Using Full Stack Technologies", Industrial Engineering Journal, ISSN: 0970-2555 Volume 15 Issue 11 October 2022
10. Dr. Mohammed Abdul Bari,Arul Raj Natraj Rajgopal, Dr.P. Swetha ," *Analysing AWSDevOps CI/CD Serverless Pipeline Lambda Function's Throughput in Relation to Other Solution*", International Journal of Intelligent Systems and Applications in Engineering , JISAE, ISSN:2147-6799, Nov 2023, 12(4s), 519–526
11. Ijteba Sultana, Mohd Abdul Bari and Sanjay," *Impact of Intermediate per Nodes on the QoS Provision in Wireless Infrastructure less Networks*", Journal of Physics: Conference Series, Conf. Ser. 1998 012029 , CONSILIO Aug 2021
12. M.A.Bari, Sunjay Kalkal, Shahanawaj Ahamad," *A Comparative Study and Performance Analysis of Routing Algorithms*", in 3rd International Conference ICCIDM, Springer - 978-981-10-3874-7_3 Dec (2016)
13. Mohammed Rahmat Ali,: BIOMETRIC: AN e-AUTHENTICATION SYSTEM TRENDS AND FUTURE APLLICATION", International Journal of Scientific Research in Engineering (IJSRE), Volume1, Issue 7, July 2017
14. Mohammed Rahmat Ali,: BYOD.... A systematic approach for analyzing and visualizing the type of data and information breaches with cyber security", NEUROQUANTOLOGY, Volume20, Issue 15, November 2022
15. Mohammed Rahmat Ali, Computer Forensics -An Introduction of New Face to the Digital World, International Journal on Recent and Innovation Trends in Computing and Communication, ISSN: 2321-8169-453 – 456, Volume: 5 Issue: 7
16. Mohammed Rahmat Ali, Digital Forensics and Artificial Intelligence ...A Study, International Journal of Innovative Science and Research Technology, ISSN:2456-2165, Volume: 5 Issue:12.
17. Mohammed Rahmat Ali, Usage of Technology in Small and Medium Scale Business, International Journal of Advanced Research in Science & Technology (IJARST), ISSN:2581-9429, Volume: 7 Issue:1, July 2020.
18. Mohammed Rahmat Ali, Internet of Things (IOT) Basics - An Introduction to the New Digital World, International Journal on Recent and Innovation Trends in Computing and Communication, ISSN: 2321-8169-32-36, Volume: 5 Issue: 10

19. Mohammed Rahmat Ali, Internet of things (IOT) and information retrieval: an introduction, International Journal of Engineering and Innovative Technology (IJEIT), ISSN: 2277-3754, Volume: 7 Issue: 4, October 2017.

20. Mohammed Rahmat Ali, How Internet of Things (IOT) Will Affect the Future - A Study, International Journal on Future Revolution in Computer Science & Communication Engineering, ISSN: 2454-424874 – 77, Volume: 3 Issue: 10, October 2017.
21. Mohammed Rahmat Ali, ECO Friendly Advancements in computer Science Engineering and Technology, International Journal on Scientific Research in Engineering(IJSRE), Volume: 1 Issue: 1, January 2017
22. Ijteba Sultana, Dr. Mohd Abdul Bari ,Dr. Sanjay, "*Routing Quality of Service for Multipath Manets, International Journal of Intelligent Systems and Applications in Engineering"*, JISAE, ISSN:2147-6799, 2024, 12(5s), 08–16;
23. Mr. Pathan Ahmed Khan, Dr. M.A Bari,: Impact Of Emergence With Robotics At Educational Institution And Emerging Challenges", International Journal of Multidisciplinary Engineering in Current Research(IJMEC), ISSN: 2456-4265, Volume 6, Issue 12, December 2021,Page 43-46
24. Shahanawaj Ahamad, Mohammed Abdul Bari, Big Data Processing Model for Smart City Design: A Systematic Review ", VOL 2021: ISSUE 08 IS SN : 0011-9342 ;Design Engineering (Toronto) Elsevier SCI Oct : 021
25. Syed Shehriyar Ali, Mohammed Sarfaraz Shaikh, Syed Safi Uddin, Dr. Mohammed Abdul Bari, "Saas Product Comparison and Reviews Using Nlp", Journal of Engineering Science (JES), ISSN NO:0377-9254, Vol 13, Issue 05, MAY/2022
26. Mohammed Abdul Bari, Shahanawaj Ahamad, Mohammed Rahmat Ali," Smartphone Security and Protection Practices", International Journal of Engineering and Applied Computer Science (IJEACS) ; ISBN: 9798799755577 Volume: 03, Issue: 01, December 2021 (International Journal,U K) Pages 1-6
27. .A.Bari& Shahanawaj Ahamad, "Managing Knowledge in Development of Agile Software", in International Journal of Advanced Computer Science & Applications (IJACSA), ISSN: 2156-5570, Vol: 2, No: 4, pp: 72-76, New York, U.S.A., April 2011
28. Imreena Ali (Ph.D), Naila Fathima, Prof. P.V.Sudha ,"Deep Learning for Large-Scale Traffic-Sign Detection and Recognition", Journal of Chemical Health Risks, ISSN:2251-6727/ JCHR (2023) 13(3), 1238-1253
29. Imreena, Mohammed Ahmed Hussain, Mohammed Waseem Akram" An Automatic Advisor for Refactoring Software Clones Based on Machine Learning", Mathematical Statistician and Engineering ApplicationsVol. 72 No. 1 (2023)
30. Mrs Imreena Ali Rubeena,Qudsiya Fatima Fatimunisa "Pay as You Decrypt Using FEPOD Scheme and Blockchain", Mathematical Statistician and Engineering Applications: https://doi.org/10.17762/msea.v72i1.2369 Vol. 72 No. 1 (2023)
31. Imreena Ali , Vishnuvardhan, B.Sudhakar," Proficient Caching Intended For Virtual Machines In Cloud Computing", International Journal Of Reviews On Recent Electronics And Computer Science , ISSN 2321-5461,IJRRECS/October 2013/Volume-1/Issue-6/1481-1486

32. Heena Yasmin, A Systematic Approach for Authentic and Integrity of Dissemination Data in Networks by Using Secure DiDrip, INTERNATIONAL JOURNAL OF PROFESSIONAL ENGINEERING STUDIES, Volume VI /Issue 5 / SEP 2016

33. Heena Yasmin, Cyber-Attack Detection in a Network, Mathematical Statistician and Engineering Applications, ISSN:2094-0343, Vol.72 No.1(2023)
34. Heena Yasmin, Emerging Continuous Integration Continuous Delivery (CI/CD) For Small Teams, Mathematical Statistician and Engineering Applications, ISSN:2094-0343, Vol.72 No.1(2023)