ISSN: 2278-4632
Vol-14 Issue-02 Aug 2024

# SOCIAL SPAM DETECTION VIA CONVEX NON NEGATIVE MATRIX FACTORIZATION

**Dr.A.Nagaraju [1], Decharaju Vishnu Vardhan Rao (20S11A6751) [2], Mukkamalla Bharath Simha Reddy (20S11A6705) [3], Etikala Lokesh (20S11A6714) [4], Malladi Nithin Reddy (20S11A6722) [5],**
**PROFESSOR&HOD [1], UG STUDENTS [2,3,4,5,]**
**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING (DATA SCIENCE)**
**MALLA REDDY INSTITUTE OF TECHNOLOGY & SCIENCE,**
**Maisammaguda, Medchal (M), Hyderabad-500100, Telangana**

## ABSTRACT

***With the increasing popularity of social network platforms such as Twitter and Sina Weibo, a lot of malicious users, also known as social spammers, disseminate illegal information to normal users. Several approaches are proposed to detect spammers by training a classifier with optimization methods and mainly using content and social following information. Due to the development of spammers' strategies and the courtesy of some legitimate users, social following information becomes vulnerable to fake by spammers. Meanwhile, the possible social activities and behaviors vary significantly among different users, which leads to a large yet sparse feature space to be modeled by existing approaches. To address issues, in this paper, we propose a new approach named CNMFSD for spammer detection in social networks, which exploits both content information and users interaction relationships in an innovative manner. We have empirically validated the proposed method on a real-world Twitter dataset, and experimental results show that the proposed CNMFSD method improves the detection performance significantly compared with baselines.***

## Introduction:

Social networks, such as Twitter, Facebook, and Sina Weibo, are increasingly used to disseminatez and share information easily and quickly. However, it is a double-edged sword since the success of social networks also attracts more social spammers. They try to seize our privacy, send us unwanted information, publish malicious content and links, and promote commodity information, which thoroughly impacts social stability and organizational management models. According to a study by Nexgate, the number of social spammers grows so fast that one in two hundred social messages is spam. Meanwhile, to increase their influence and be undetected, spammers collude with each other to construct the criminal communities. Thus, social spammer detection is a challenging task for researchers. Successful social spammer detection presents its significance to improve the quality of user experience, and positively impact the overall value of the social systems going forward. In the past decade, researchers have tried different techniques to detect spammers, such as link analysis and content analysis. The methods of content-based detection of spammers mainly focus on analyzing and extracting users' features and then directly applying existing classification approaches such as support vector machines (SVM) to detect spammers. Recently, more advanced deep learning-based approaches have been proposed to detect social spammers only based on content. However, with the development of spamming strategies, these methods could not accurately detect spammers with new strategies, only relying on the extracted features. Another category of methods is proposed to detect spammers via social network analysis. These methods assume that spammers cannot establish an arbitrarily large number of social trust relations with legitimate users. The users, who have relatively low social influence or social status in social networks, will be determined as spammers. Unfortunately, only depending on network information, these methods are hard to distinguish between legitimate users and spammers. In this paper, we study the problem of social spammer detection with social interaction and content information. In essence, we investigate: how to model the social interaction information and content information properly; and how to seamlessly utilize both social interaction and content information for the problem we are studying. Our solutions to these two challenges result in a novel spammer detection framework name Convex-NMF based Supervised Spammer Detection with Social Interaction (CNMFSD). Based on statistical analysis, we observe that spammers and legitimate users have different characteristic distributions.

## LITERATURE SURVEY

**HUA SHEN [1]** is a Ph.D. candidate in computer science at Dalian University of Technology. She received a master's degree in Computer Applied Technology from Dalian Maritime University. She is currently an associate professor at

College of Mathematics and Information Science, Anshan Normal University, China. **Her research interests include data mining and machine learning.**

**BANGYU WANG [2]** is a graduate student in computer science at Dalian University of Technology. He received his scholar degree in software engineering from Dalian University of Technology. **His research interests include data mining, machine learning and information retrieval**.

**XINYUE LIU [3]** is an associate professor at School of Software, Dalian University, China, and got her Ph.D. from Dalian University of Technology, China. **Her research interests include data mining, machine learning and information retrieval**.

**XIANCHAO ZHANG [4]** is a full professor at Dalian University of Technology. He got his scholar and master degrees in mathematics from National University of Defense Technology, China, in 1994 and 1998, respectively. He got his Ph.D. in computer science from University of Science and Technology of China in 2000. From 2000 to 2003, he worked as a research and development manager in some international companies. He joined Dalian University of Technology in 2003. **His research interests include design and analysis of algorithms, machine learning, data mining, and information retrieval.**

## Existing System:

Many different methods have been proposed to combat social spammers firstly surveyed potential solutions and challenges in social spammer detection. They elaborated a classification of spammer detection techniques, including fake content, URL-based spam detection, detecting spam in trending topics, and fake user identification. In this paper, we only focus on the binary classification task, i.e., spammer or legitimate user identification. Many approaches employed machine learning methods to train a classifier to detect spammers. Some of them are jointly modeled user activities' information and the social following information to learn a classifier. Also, they proposed a hybrid technique that utilizes user-based, content-based, and graph-based characteristics for spammer profiles detection. They presented a policy for the detection of spammers on Twitter and used the popular techniques, i.e., Naive Bayes, clustering, and decision tree. An important line of research in spam detection relies on analyzing the tweet content, where suspicious use of hashtags or URLs is traced. The main objective is to study the semantics of short texts or messages in contrast with a set of Wikipedia text pages that are modeled and used as an aggregation of entities. Other directions adopted in detecting Twitter spammers focus on discovering traits or patterns that best describe the spammer's behavioral profile. In such works, the main contribution is to determine deceptive double characters for user profiles, which is done by analyzing nonverbal behavior variables as a function of time, such as follows and retweets.

## Disadvantages of Existing System:

□ The system is not implemented Convex-NMF based Supervised Spammer Detection with Social Interaction (CNMFSD).
□ The system is not implemented any ml classifier for test and train the datasets.

## Proposed System:

□ The system proposes a three-stage optimization model that conducts feature extraction and classifier learning simultaneously. First, we use Convex Non-negative Matrix Factorization (CNMF) and Non-negative Matrix Factorization (NMF) to induce latent feature from content information, then train an SVM classifier and finally, refine latent features using social interaction information as the input representations of the classifier. Through iteratively learning among content information, social interaction regularization, and classification model, the proposed method can train an accurate classifier.
□ The system proposes a novel method to induce latent features and a novel social interaction regularization term. Using CNMF, we get the latent content matrix of spammers and legitimate users, respectively, and then obtain the user feature latent matrix by NMF according to the latent content matrix. The latent feature refine process is guided by the social interaction relationship matrix and the label information.

 The proposed system evaluates our method on a large-scale real-world social network data set from Twitter, one of the largest social networks in the world. The experimental results show that the proposed framework can identify more spammers compared with baseline approaches.We conduct experiments to demonstrate the significance of using CNMF to induce latent features for spammers and legitimate users, respectively, and validate the effectiveness of the new social interaction regularization term

## Advantages of Proposed System

 The proposed system refines the latent features with predicted label information and social interaction information with the help of svm classifier.
 The proposed system implemented UNLABELED USER CONTENT FACTORIZATION.

### SYSTEM DESIGN

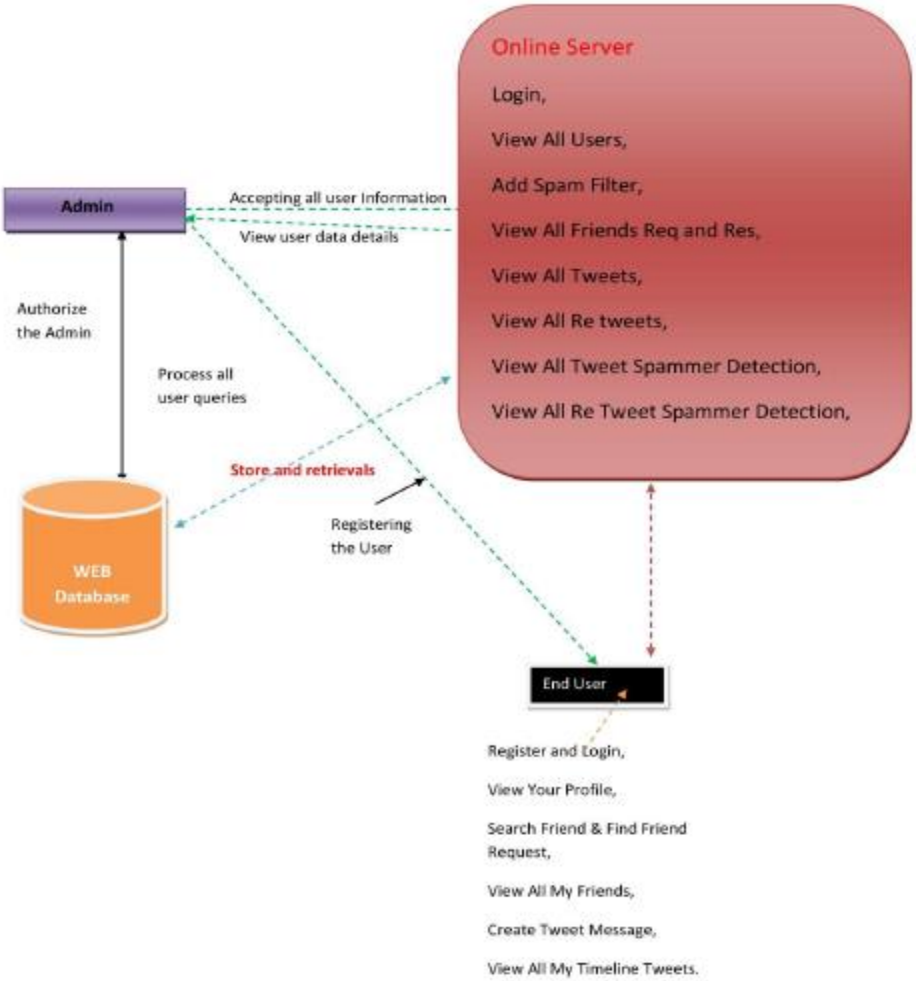


*Fig-1 Block Diagram*

## Flow Chart 1: End User

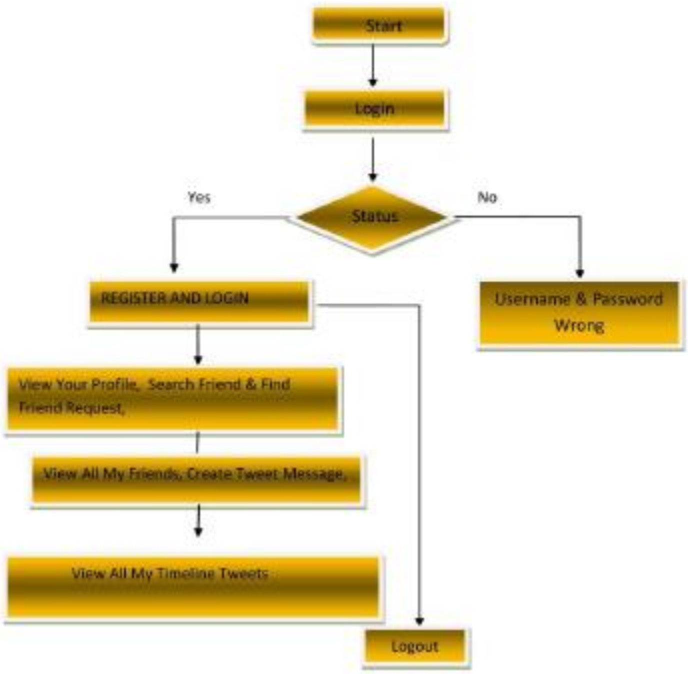

Start
Login
Yes
Status
No
REGISTER AND LOGIN
Username & Password Wrong
View Your Profile, Search Friend & Find Friend Request,
View All My Friends, Create Tweet Message,
View All My Timeline Tweets
Logout

*Fig-2 Flow Chart of End User*

### Hardware Requirements:

- Processor: i3 and above
- RAM: 4 GB
- Space on Hard Disk: 20 GB

### Software Requirements:

- Eclipse IDE
- JDK 1.8
- SQL YOG
- MYSQL
- TOMCAT

### Operating Systems Supported:

- Windows 7
- Windows 10
- Windows 11

### Technologies and Languages used to Develop:

- ✓ JAVA
- ✓ J2EE (JSP, Servlet)
- ✓ CSS
- ✓ HTML
- ✓ JAVA SCRIPT
- ✓ MySQL

### Debugger and Emulator:

- ✓ Eclipse

## INPUT AND OUPUT DESIGN

### INPUT DESIGN:

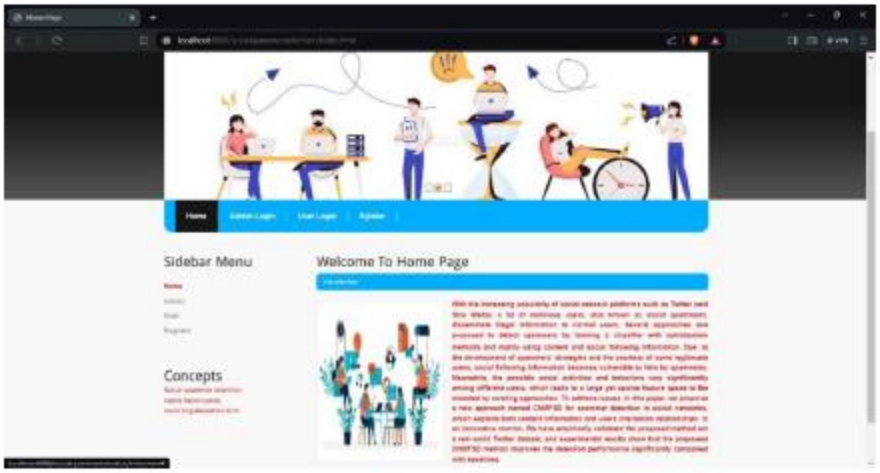
Input Design plays a vital role in the life cycle of software development, it requires very careful attention of developers. The input design is to feed data to the application as accurate as possible. So inputs are supposed to be designed effectively so that the errors occurring while feeding are minimized. According to Software Engineering Concepts, the input forms or screens are designed to provide to have a validation control over the input limit, range and other related validations. This system has input screens in almost all the modules. Error messages are developed to alert the user whenever he commits some mistakes and guides him in the right way so that invalid entries are not made. Let us see deeply about this under module design.

Input design is the process of converting the user created input into a computer-based format. The goal of the input design is to make the data entry logical and free from errors. The error is in the input are controlled by the input design. The application has been developed in user-friendly manner. The forms have been designed in such a way during the processing the cursor is placed in the position where must be entered. The user is also provided with in an option to select an appropriate input from various alternatives related to the field in certain cases. Validations are required for each data entered. Whenever a user enters an erroneous data, error message is displayed and the user can move on to the subsequent pages after completing all the entries in the current page.

### OUTPUT DESIGN:

The Output from the computer is required to mainly create an efficient method of communication within the company primarily among the project leader and his team members, in other words, the administrator and the clients. The output of VPN is the system which allows the project leader to manage his clients in terms of creating new clients and assigning new projects to them, maintaining a record of the project validity and providing folder level access to each client on the user side depending on the projects allotted to him. After completion of a project, a new project may be assigned to the client. User authentication procedures are maintained at the initial stages itself. A new user may be created by the administrator himself or a user can himself register as a new user but the task of assigning projects and validating a new user rests with the administrator only. The application starts running when it is executed for the first time. The server has to be started and then the internet explorer in used as the browser. The project will run on the local area network so the server machine will serve as the administrator while the other connected systems can act as the clients. The developed system is highly user friendly and can be easily understood by anyone using it even for the first time.

## RESULTS



Sidebar Menu

Home
Admin
User
Register

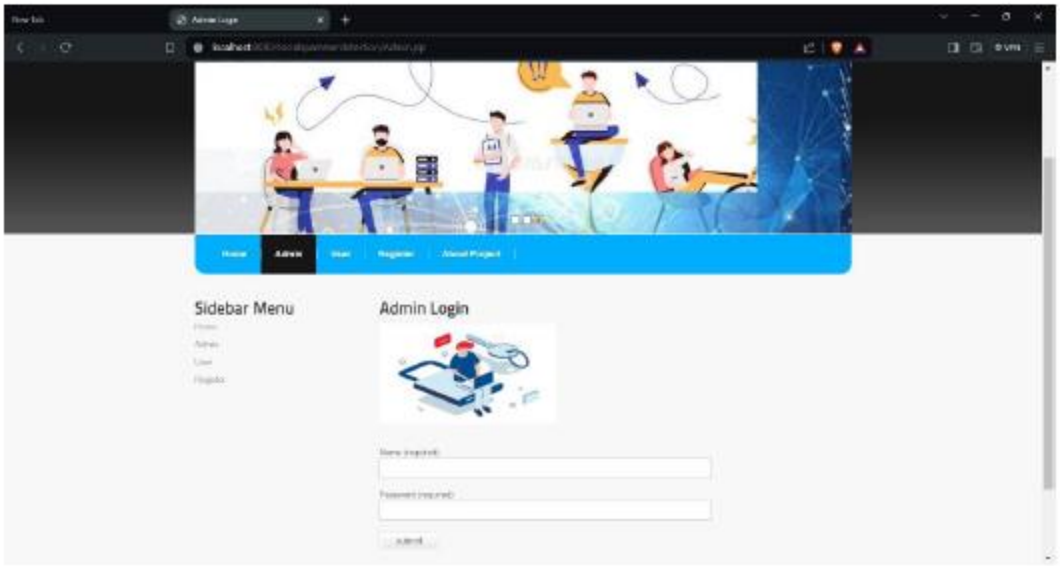
Concepts

[illegible]

Welcome To Home Page

With the increasing popularity of social network platforms such as Twitter and Sina Weibo, a lot of malicious users, also known as social spammers, disseminate illegal information to normal users. Several approaches are proposed to detect spammers by training a classifier with optimization methods and mainly using content and social following information. Due to the development of spammers' strategies and the practice of some legitimate users, social following information becomes vulnerable to fake by spammers. Meanwhile, the possible social activities and behaviors vary significantly among different users, which leads to a large yet sparse feature space to be modeled by existing approaches. To address issues, in this paper, we propose a new approach named CMSF-SD for spammer detection in social networks, which exploits both content information and users' interaction relationships in an innovative manner. We have empirically validated the proposed method on a real-world Twitter dataset, and experimental results show that the proposed CMSF-SD method improves the detection performance significantly compared with baselines.

*Fig-3: Output Of Home Page*



Sidebar Menu

Home
Admin
User
Register

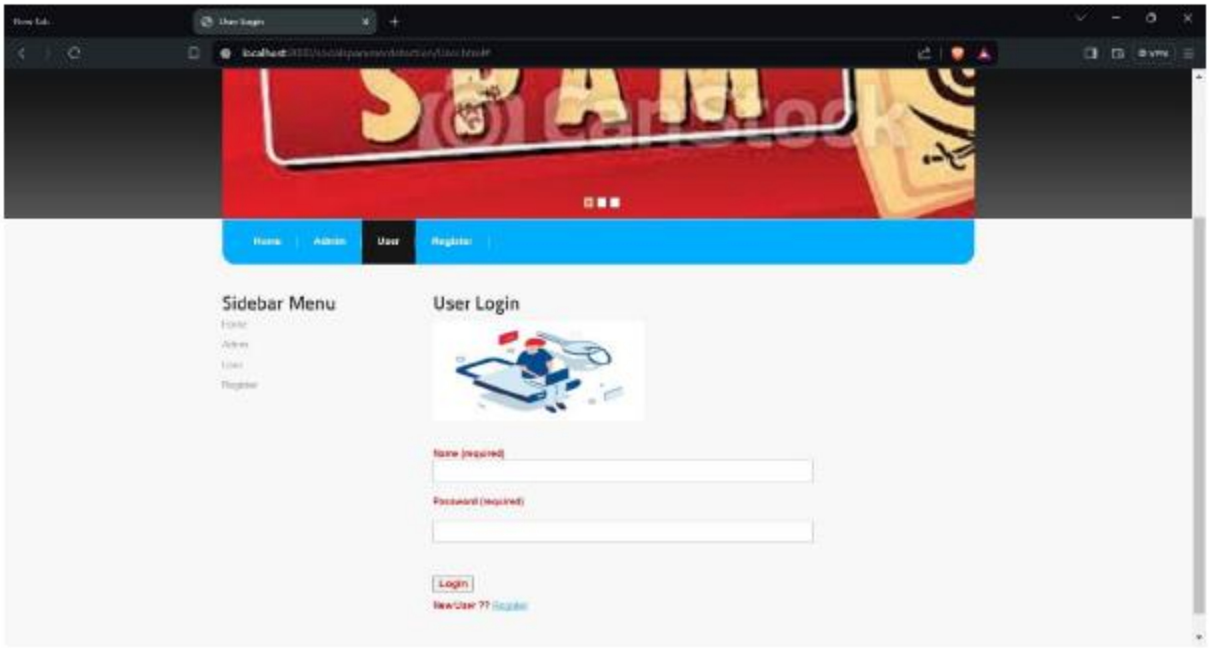
Admin Login

Name (required)

Password (required)

submit

*Fig-4: Admin Login Page*

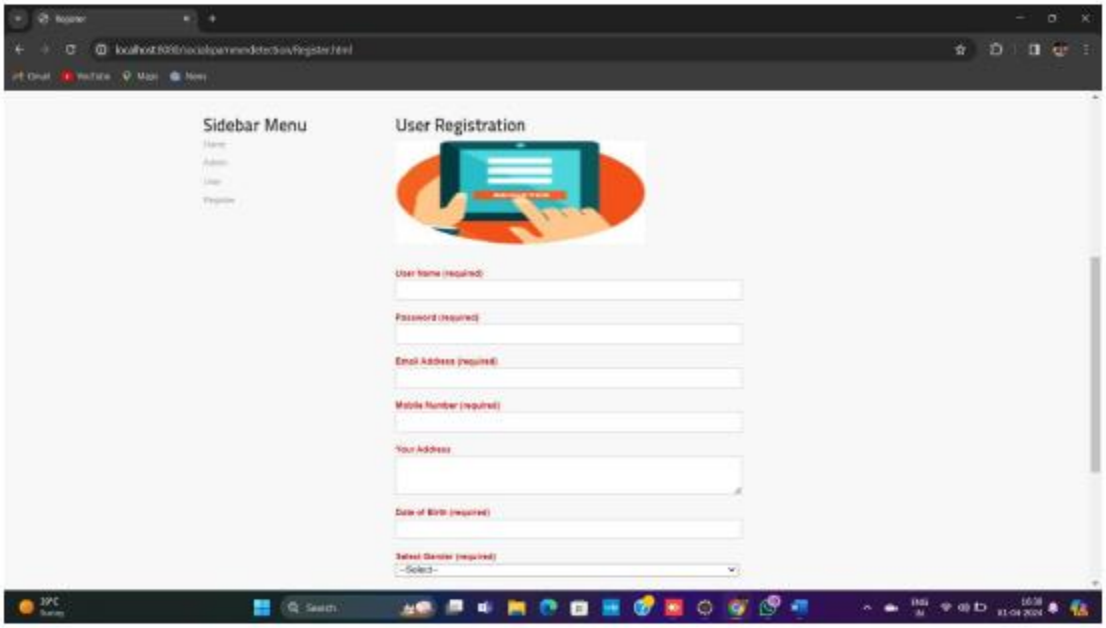


User Login

Name (required)

Password (required)

*Fig-5: User Login Page*



User Registration

User Name (required)

Password (required)

Email Address (required)

Mobile Number (required)

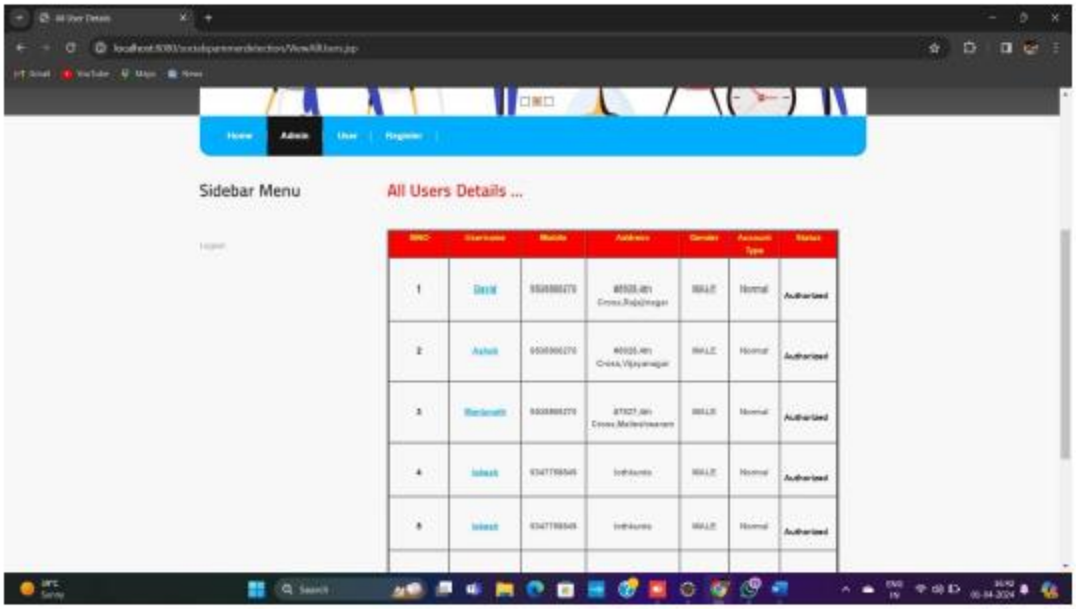
Your Address

Date of Birth (required)

Select Gender (required)

--Select--

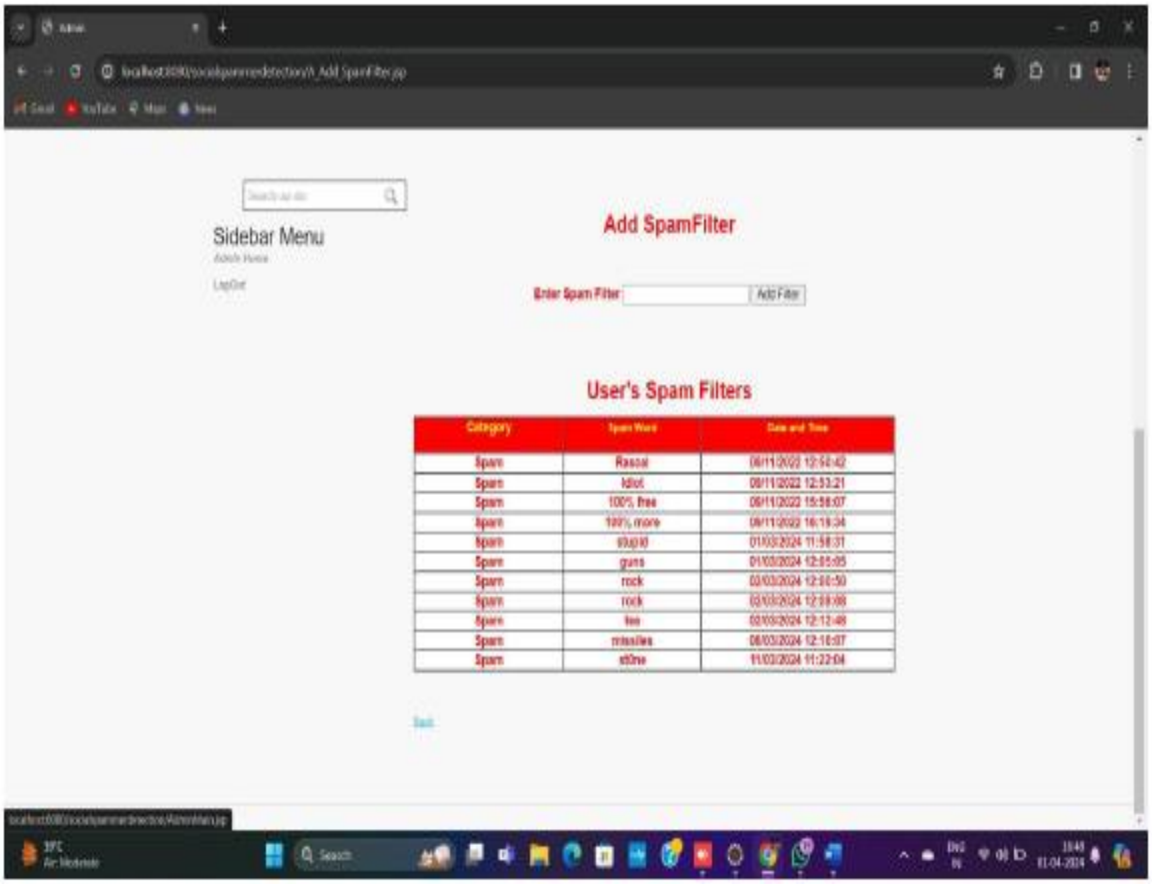
*Fig-6: New User Register Page*



All Users Details ...

| SINO | Username | Mobile | Address | Gender | Account Type | Status |
|---|---|---|---|---|---|---|
| 1 | David | [illegible] | #8928,4th Cross,Rajajinagar | MALE | Normal | Authorized |
| 2 | Ashok | [illegible] | #8928,4th Cross,Vijayanagar | MALE | Normal | Authorized |
| 3 | Manjunath | [illegible] | #7927,4th Cross,Malleshwaram | MALE | Normal | Authorized |
| 4 | [illegible] | [illegible] | [illegible] | MALE | Normal | Authorized |
| 5 | [illegible] | [illegible] | [illegible] | MALE | Normal | Authorized |

*Fig-7: View and Authorize User from Admin*



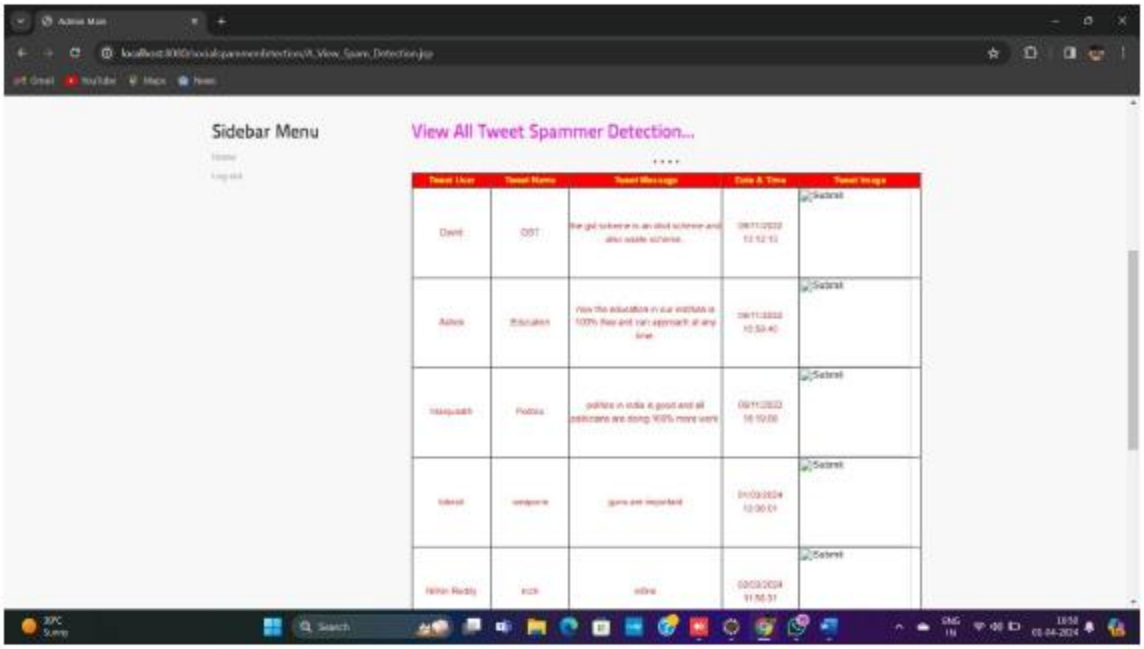
Add SpamFilter

Enter Spam Filter

User's Spam Filters

| Category | Spam Word | Date and Time |
|---|---|---|
| Spam | Rascal | 08/11/2022 12:50:42 |
| Spam | Idiot | 08/11/2022 12:53:21 |
| Spam | 100% free | 08/11/2022 15:58:07 |
| Spam | 100% more | 08/11/2022 16:19:34 |
| Spam | stupid | 01/03/2024 11:58:31 |
| Spam | guns | 01/03/2024 12:05:05 |
| Spam | rock | 02/03/2024 12:00:50 |
| Spam | rock | 02/03/2024 12:09:08 |
| Spam | lee | 02/03/2024 12:12:48 |
| Spam | missiles | 08/03/2024 12:10:07 |
| Spam | st0ne | 11/03/2024 11:22:04 |

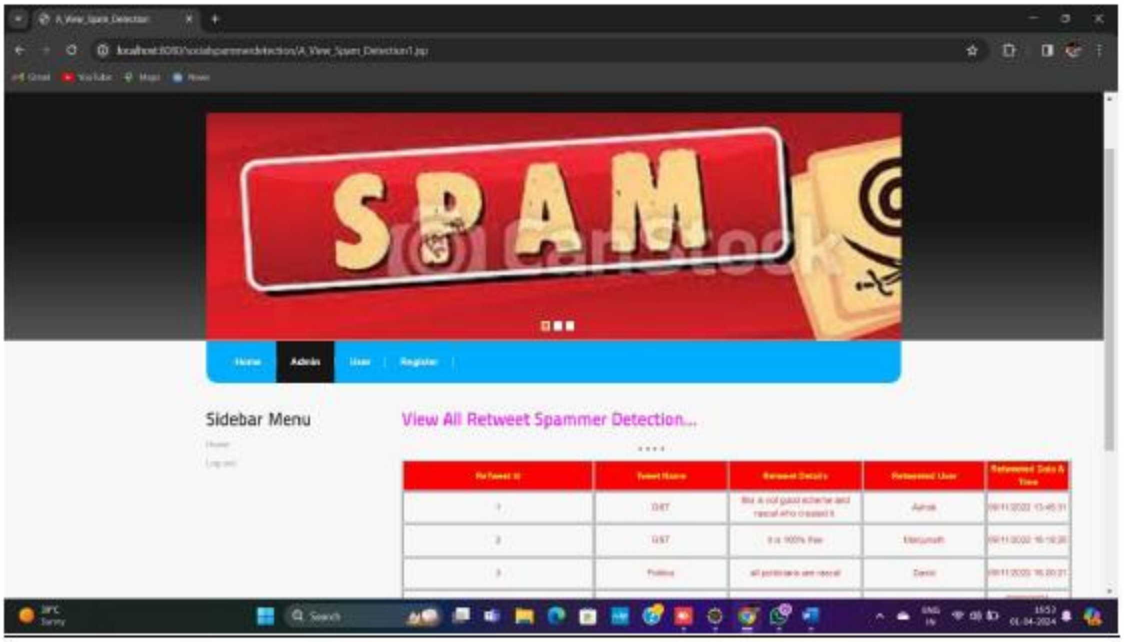
*Fig-8: Adding Spam Filters*


Admin Main
localhost:8080/socialspammerdetection/A_View_Spam_Detection.jsp
Sidebar Menu
Home
Log out

View All Tweet Spammer Detection...

| Tweet User | Tweet Name | Tweet Message | Date & Time | Tweet Image |
|---|---|---|---|---|
| David | GST | the gst scheme is an idiot scheme and also waste scheme.. | 08/11/2022 12:12:13 | Submit |
| Admin | Education | now the education in our institute is 100% free and can approach at any time | 08/11/2022 10:59:40 | Submit |
| Manjunath | Politics | politics in india is good and all politicians are doing 100% more work | 09/11/2022 16:10:00 | Submit |
| [illegible] | [illegible] | guns are important | 01/03/2024 12:00:01 | Submit |
| Nithin Reddy | [illegible] | [illegible] | 02/03/2024 11:56:31 | Submit |

*Fig-9: View Of All Tweet Spammer Users from Admin*



View All Retweet Spammer Detection...

| ReTweet Id | Tweet Name | Retweet Details | Retweeted User | Retweeted Date & Time |
|---|---|---|---|---|
| 1 | GST | this is not good scheme and rascal who created it | Admin | 09/11/2022 13:45:31 |
| 2 | GST | it is 100% free | Manjunath | 09/11/2022 16:19:20 |
| 3 | Politics | all politicians are rascal | David | 09/11/2022 16:20:21 |

*Fig-10: View Of All ReTweet Spammer Users from Admin*

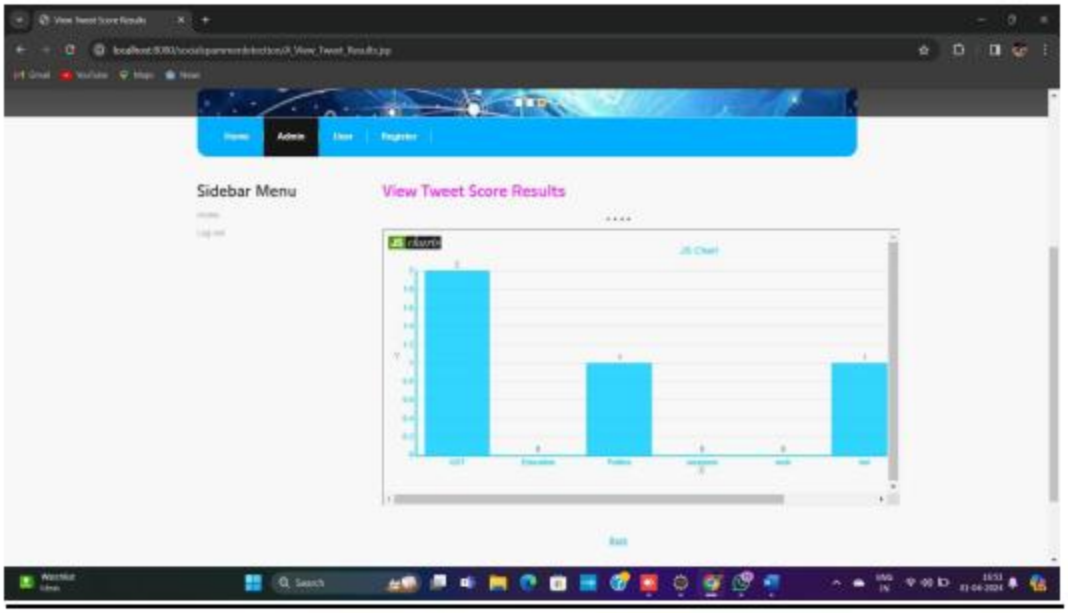


*Fig-11: View Of Tweet Score results By Using Spam Filters*

## Conclusion:

In this paper, we propose a new framework by taking advantage of content and social interaction information for social spammer detection. Different from existing methods that utilize users' the following information, the proposed method CNMFSD integrates users' interaction information based on the trained classification model. In addition, we introduce a new strategy to induce latent features using CNMF in spammers and legitimate users space for improving the performance of detecting spammers. Experimental results on a real dataset show that CNMFSD obtains better detection performance compared with existing methods. In this work, we employ Convex-NMF to learn latent user features for legitimate users and spammers, respectively. Such a fine-grained learning strategy makes the proposed model obtain accurate latent user representations, which further helps the model to achieve better performance. Besides, introducing social interaction into this task can also improve prediction performance. Although the proposed model outperforms baselines, it also has some disadvantages. First, in the classifier training stage, we do not consider the social interaction graph, which is trained solely based on the outputs from CNMF. Second, we use tf-idf to extract the user content matrices. However, spammer always posts some normal tweets to imitate the behavior of legitimate users. Thus, it is essential to distinguish the importance of tweets when we extract the user content matrix.

## Future Enhancement:

In future work, we will directly use raw tweets as the model input to learn user representations by distinguishing the importance of each tweet via deep learning techniques. After that, we plan to use graph neural networks to model social interactions among users.

## BIBLIOGRAPHY

***1. Aliaksandr Barushka and Petr Hajek. Spam detection on social networks using cost-sensitive feature selection and ensemble-based regularized deep neural networks. Neural Computing and Applications, 32(9):4239–4257, 2020.***
***2. Qiang Fu, Bo Feng, Dong Guo, and Qiang Li. Combating the evolving spammers in online social networks. Computers & Security, 72:60–73, 2018.***
***3. Zhijie Zhang, Rui Hou, and Jin Yang. Detection of social network spam based on improved extreme learning machine. IEEE Access, 8:112003–112014, 2020.***

*4. Nexgate2013. 2013 state of social media spam. http://nexgate.com/wpcontent/ uploads/2013/09/Nexgate-2013-State-of Social-Media-Spam-Research-Report.pdf.*
*5. Dehai Liu, Benjin Mei, Jinchuan Chen, Zhiwu Lu, and Xiaoyong Du.Community based spammer detection in social networks. In International Conference onWeb-Age Information Management, pages 554–558 Springer, 2015.*
*6. Faiza Masood, Ahmad Almogren, Assad Abbas, Hasan Ali Khattak, Ikram Ud Din, Mohsen Guizani, and Mansour Zuair. Spammer detection and fake user identification on social networks. IEEE Access, 7:68140– 68152, 2019.*
*7. Sanjeev Rao, Anil Kumar Verma, and Tarunpreet Bhatia. A review on social spam detection: Challenges, open issues, and future directions. Expert Systems with Applications, 186:115742, 2021.*
*8. Chao Chen, Jun Zhang, Yi Xie, Yang Xiang, Wanlei Zhou, Mohammad Mehedi Hassan, Abdulhameed AlElaiwi, and Majed Alrubaian. performance evaluation of machine learning-based streaming spam tweets detection. IEEE Transactions on Computational social systems, 2(3):65– 76, 2015.*
*9. Xianghan Zheng, Zhipeng Zeng, Zheyi Chen, Yuanlong Yu, and Chunming Rong. Detecting spammers on social networks. Neurocomputing, 159:27–34, 2015.*
*10. Chao Yang, Robert Harkreader, and Guofei Gu. Empirical evaluation and new design for fighting evolving twitter spammers. IEEE Transactions on Information Forensics and Security, 8(8):1280–1293, 2013.*
*11. Zi Chu, Indra Widjaja, and Haining Wang. Detecting social spam campaigns on twitter. In International Conference on Applied Cryptography and Network Security, pages 455–472. Springer, 2012.*
*12. Mohd Fazil, Amit Kumar Sah, and Muhammad Abulaish. Deepsbd: A deep neural network model with attention mechanism for socialbot detection. IEEE Transactions on Information Forensics and Security, 16:4211–4223, 2021.*
*13. Zulfikar Alom, Barbara Carminati, and Elena Ferrari. A deep learning model for twitter spam detection. Online Social Networks and Media, 18:100079, 2020.*
*14. Xinbo Ban, Chao Chen, Shigang Liu, Yu Wang, and Jun Zhang. Deeplearn features for twitter spam detection. In 2018 International Symposium on Security and Privacy in Social Networks and Big Data (SocialSec), pages 208–212. IEEE, 2018.*
*15. Saptarshi Ghosh, Bimal Viswanath, Farshad Kooti, Naveen Kumar Sharma, Gautam Korlam, Fabricio Benevenuto, Niloy Ganguly, and Krishna Phani Gummadi. Understanding and combating link farming in the twitter social network. In Proceedings of the 21st international conference on World Wide Web, pages 61–70, 2012.*
*16. Yin Zhu, Xiao Wang, Erheng Zhong, Nathan Liu, He Li, and Qiang Yang. Discovering spammers in social networks. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 26, pages 171–177, 2012.*
*17. Xia Hu, Jiliang Tang, Yanchao Zhang, and Huan Liu. Social spammer detection in microblogging. In Twenty-third international joint conference on artificial intelligence. Citeseer, 2013.*
*18. David M Beskow and Kathleen M Carley. Bot conversations are different: leveraging network metrics for bot detection in twitter. In 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis an Mining (ASONAM), pages 825–832. IEEE, 2018.*
*19. Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He. Twitterrank: findin topic-sensitive influential twitterers. In Proceedings of the third ACM international conference on Web search and data mining, pages 261–270, 2010.*
*20. Christian Thurau, Kristian Kersting, Mirwaes Wahabzada, and Christian Bauckhage. Convex non-negative matrix factorization for massive datasets. Knowledge and information systems, 29(2):457–478, 2011.*
*21. Chris HQ Ding, Tao Li, and Michael I Jordan. Convex and seminonnegative matrix factorizations. IEEE transactions on pattern analysis and machine intelligence, 32(1):45–55, 2008.*
*22. Paul Heymann, Georgia Koutrika, and Hector Garcia-Molina. Fighting spam on social web sites: A survey of approaches and future challenges. IEEE Internet Computing, 11(6):36–45, 2007.*
*23. Malik Mateen, Muhammad Azhar Iqbal, Muhammad Aleem, and Muhammad Arshad Islam. A hybrid approach for spam detection for twitter. In 2017 14th International Bhurban Conference on Applied Sciences and Technology (IBCAST), pages 466–471. IEEE, 2017.*
*24. Arushi Gupta and Rishabh Kaushal. Improving spam detection in online social networks. In 2015 International conference on cognitive computing and information processing (CCIP), pages 1–6. IEEE, 2015.*
*25. Sangho Lee and Jong Kim. Warningbird: A near real-time detection system for suspicious urls in twitter stream. IEEE transactions on dependable and secure computing, 10(3):183–195, 2013.*