



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 9, September 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.165

 9940 572 462

 6381 907 438

 ijircce@gmail.com

 www.ijircce.com

Application Analysis of Big Data in Digital Media

Alfiya Javeed, Dr. Sujata SR

PG Student, Dept. of CSE, Sri Siddhartha Institute of Technology, Tumakuru, Karnataka, India

Associate Professor, Dept. of CSE, Sri Siddhartha Institute of Technology, Tumakuru, Karnataka, India

ABSTRACT: With the rapid growth of social networks and micro blogging websites, communication between people from different cultural and psychological backgrounds has become more direct, resulting in more and more ‘cyber’ conflicts between these people. Consequently, hate speech is used more and more, to the point where it has become a serious problem invading these open spaces and difficult to detect fake accounts which spread hate speech against group of people or individual. Hate speech refers to the use of aggressive, violent or offensive language, targeting a specific group of people sharing a common property, whether this property is their gender (i.e., sexism), their ethnic group or race (i.e., racism) or their believes and religion.

KEYWORDS: Big Data; Digital Media, Hate Speech, Fake Account.

I. INTRODUCTION

Expanding new demand and predicting future development in Digital Media can be solely depend on Big Data. Big Data supports Digital Media efficiently in the industry which helps to strength social and economic growth. Digital Media has become primary medium for individuals to communicate and interact with individuals using the Internet. It contains an interactive social network where users participate, connect, reshape, and exchange their ideas and thoughts with other users. One of the most popular Digital Media application is Twitter one of the popular website started in 2006, where users interact with other users via post messages of around 180 characters, known as “tweet”.The spammers are attracted to this popularity spreads hateful URLs within tweets, sending unsolicited message to other users, spreading rumours, tweeting abusive words and trolling the user. The users having no idea about the ill intentions of fake account holder shares their problems and their weakness, which the fake account holder use it for his/her benefit.

II. METHODOLOGY

The collection method and tools are very different for abundant set of data. The processing and storage of enormous data is managed to control the software and hardware costs and error rate in the best ways. User Acceptance of a system is the key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keeping in touch with the prospective system users at the time of developing and making changes wherever required. User friendly interface is designed which helps the user to easily understand and work on the system who is new to it. The user’s fake or real account can be known by using machine learning algorithms based on the amount of time sends friend requests, Tracking the user time, Average clicks per session.

Neural Network, Support Vector Machine and newly developed algorithm, SVM-NN are the machine learning algorithms used in distinguishing the real and fake accounts.

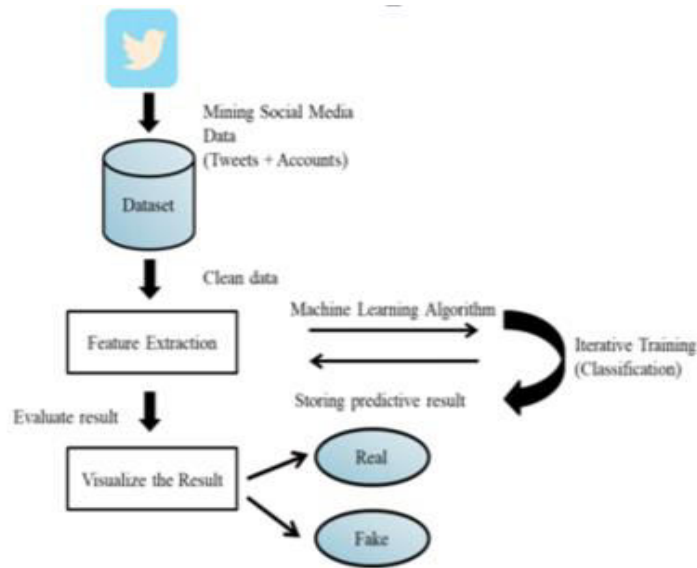


Figure 1 Twitter Fake Account Detection

2.1 Neural Networks

Neural Network Algorithms are used for training and predicting the results of models with reference to previously trained models. A base algorithm Feed-forward back propagation algorithm is applied as the initial step. The Neural Network uses three classifiers to calculate the highest accuracy of finding out fake accounts.

2.2 SVM-NN

The SVM-NN is the proposed algorithm used in analyzing and maintaining the accounts which implements C-support and kernel functions. The SVM-NN is used to identify the account as real or fake. The input data(Tweets and Accounts) is used for twitter dataset with some features. The training dataset contains data pre-processing which includes two steps: Machine learning technique and Feature Extraction are used in processing the data which is used for selecting number of features.

III. LITERATURE SURVEY

In paper [1], The scourge of cyber bullying has assumed alarming proportions with an ever-increasing number of adolescents admitting to having dealt with it either as a victim or as a bystander. Anonymity and the lack of meaningful supervision in the electronic medium are two factors that have exacerbated this social menace. Comments or posts involving sensitive topics that are personal to an individual are more likely to be internalized by a victim, often resulting in tragic outcomes.

In paper [2], It is unclear how close to or removed from their online personas users actually are in real life. This means that scholars should be careful to take data for what it truly represents (e.g., traces of behaviour), but not infer too much about possible attitudes, emotions, or motivations of those whose behaviour created the data although some seem happy to make such inferences. Some of the data obtain from user may be faked or unreal.

In paper [3], The concept of Social Media is top of the agenda for many business executives today. Decision makers, as well as consultants, try to identify ways in which firms can make profitable use of applications such as Wikipedia, YouTube, Facebook, Second Life, and Twitter.

In paper [4], Cyber Bullying, which often has a deeply negative impact on the victim, has grown as a serious issue among adolescents. To understand the phenomenon of cyber bullying, experts in social science have focused on personality, social relationships and psychological factors involving both the bully and the victim. Recently computer science researchers have also come up with automated methods to identify cyber bullying messages by identifying bullying-related keywords in cyber conversations.

In paper [5], compromising social network accounts has become a profitable course of action for cybercriminals. By hijacking control of a popular media or business account, attackers can distribute their malicious messages or disseminate fake information to a large user base. The impacts of these incidents range from a tarnished reputation to multi-billion dollar monetary losses on financial markets.

IV. EXISTING SYSTEM

The role of big data technologies in various industries is very obvious. The internet users even they are smart and sometimes not aware of these fake accounts they accept the requests while online purchasing or any spam message and regret after misguided and suffered in the hand of fraudster. The fake account holders are really good in disguising their personality and behaviour they may appear to be genuine and nice person but they have malicious thoughts. They have hostile intentions create issues in an individual life by hate-speech, morphing or sharing the pictures of user without their consent, spreading fake news, blackmailing them by their weakness by sharing their photo's or personal information. The users accept the requests of these fake accounts and suffer in the process. Thus, detecting fake accounts in digital Media is necessary in order to avoid the scams.

V. PROBLEM STATEMENT

Generally, people are using Twitter to get updates from personal and professional connections, as well as to see the current news of the world. Due to no restriction on what to post, people feel free to post any post on Twitter; as a result, it is effortless to post negative comments, hateful messages on Twitter.

Social networks in this context present a better and more rich, yet less reliable and full of noise, source of information. Short text message contains lot of un-constructed worlds and cannot process all short text messages this may slow down the process.

Most recently, Facebook and Twitter received a large number of Hate Speech (HS) related posts on Delhi's (India capital) Shaheen bag protest against National Register of Citizens (NRC), Citizenship Amendment Act (CAA) and National Population Register (NPR), which began on 11 December, 2019.2 Again during the COVID-19 coronavirus epidemic, HS-related tweets were started trending on Twitter (with the hashtag #hatespeech, #HateSpeech) and Facebook.

VI. OBJECTIVE

- Implementation of module which identifies hate speech on OSN.
- Implementation of filter module which identifies special characters.
- Identify user activity on OSN.
- Auto report of abnormal behavior of user to administrator.
- Identify social network activates using sentiment analysis statistics
- Implementation of machine learning module to filter bad words from posted messages.

VII. DESIGN OF THE SYSTEM

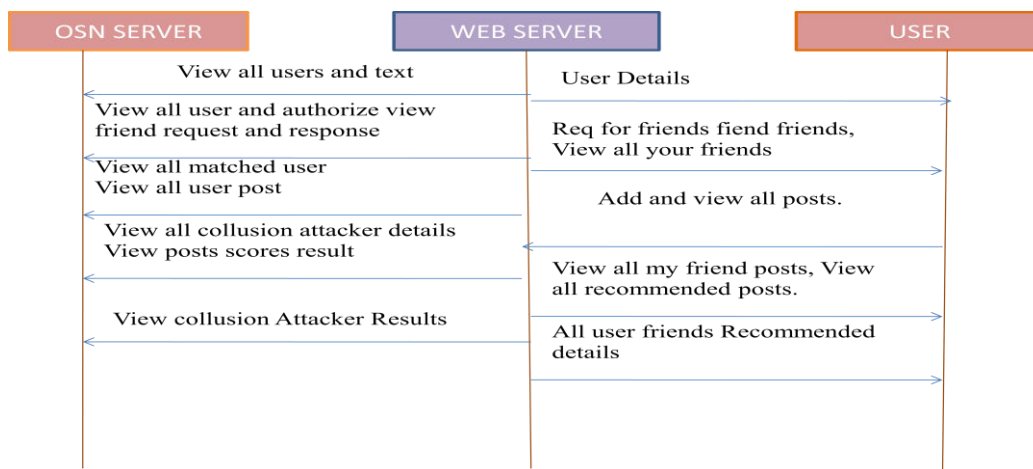


Figure 2 Design of System

The web browser acts a median between OSN server and user, it exchanges the information between this modules. The web browser collects and stores the data of users which will be analysed by the OSN server first then by the user. The OSN server after viewing and authorize the data the web browser sends the user details to user. The server view all

users and authorize, view Friend Request and Response then a request for friend, find friends, view all the friend is send to user .After accepting the friend request the user view all the post of the user, the server view all matched users, view all user posts. The server then checks for any collusion Attacker Details, view posts score result, the user checks friend posts and recommended posts. The user checks all the friends recommended details and cross verify for any attacks and threat by them.

VIII. FLOW CHART

The user register its details such as User_Id and password then login to a Digital Media website. User_id is unique each user have their specific unique identity. A single user can have more than one account. If the user name and password is correct then it will view users profile, if not it will not view the profile it will give a message a username and password is incorrect. Once user is login to website the user will view his profile were they can send friend request, find new friends, view other friends and sometimes unknown user profile. The user can add post, view all posts by them, view their friend’s posts, view the recommended posts, other users lifestyle and their day-to-day activities. The users can even view their friends user posts and can send friend request to them if their mutual friends. The user can logout once done with the interaction with other users.

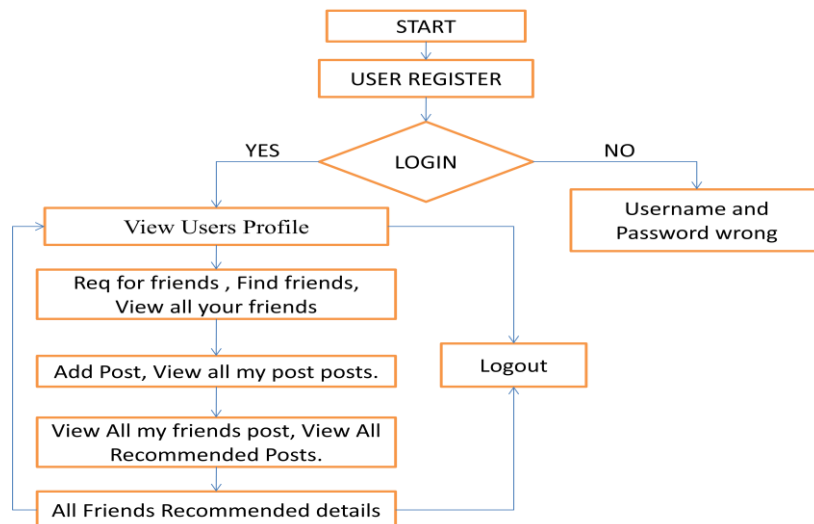


Fig 3 User Registration in a Digital Media.

ADVANTAGES

- To create awareness by providing proper guidance and knowledge.
- The user can connect globally with other user without any issues.

IX. MODULES FOR IMPLEMENTATION

- **Users:** Social Network users can create an account, send and receive message to other users. The users by creating digital media accounts can improve their knowledge, create awareness, can earn income, Influence other lifestyles, Endorse their brand, learn new concepts.
- **OSN:** OSN Server is one of the important module which helps to visualize, analyse, response to users by their behaviour and activities on Digital Media. Server can view all users and Authorize, view friend Request and Response, View All Matched Users, View All User Post Posts, View All Posts Recommended Details, View All Friend Recommended Details, View All Collusion Attacker Details.
- **Naïve Bayes Classifier:** Naïve Bayes classifiers are a collection of classification algorithm based on Bayes Theorem. It is simple and most effective classification algorithm which works well with large datasets using machine learning algorithm that can make quick prediction as it doesn't require as much training data. Naïve Bayes Algorithms are mostly used in sentiment analysis, spam filtering, recommendation systems, to check account activities and predict account is not fake or a fake account.

X. RESULTS

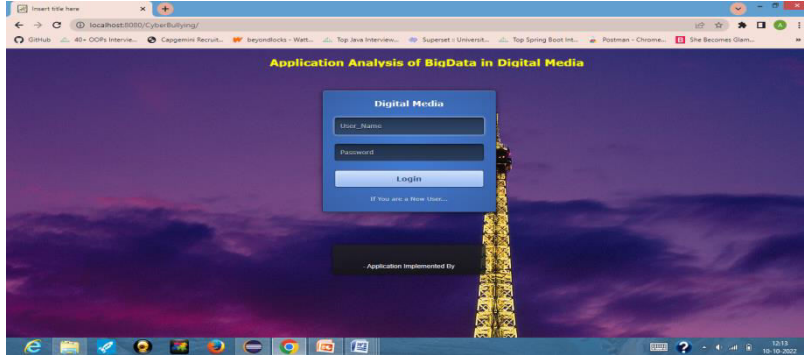


Fig 4 Login to a Website

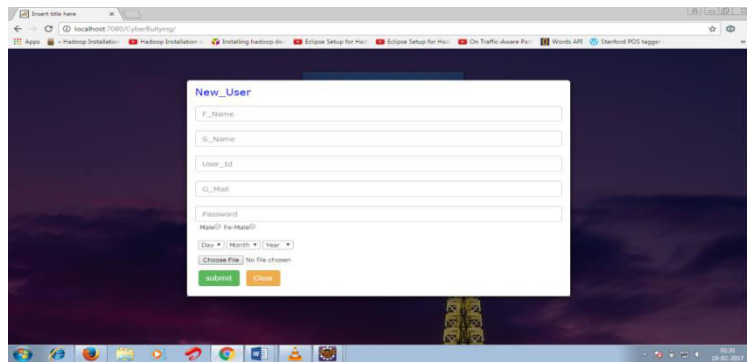


Fig 5 Entering the User data

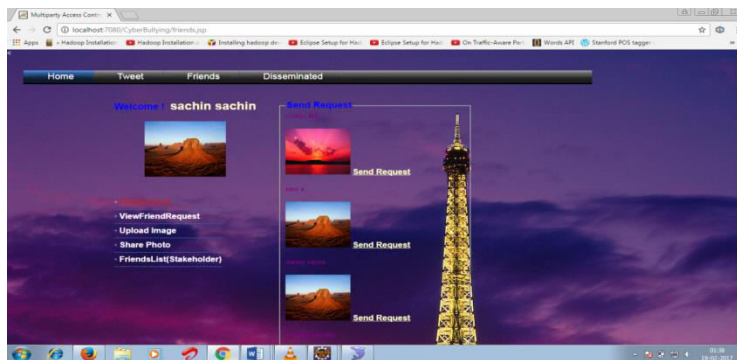


Fig 7 Sending Friend Request to other users.

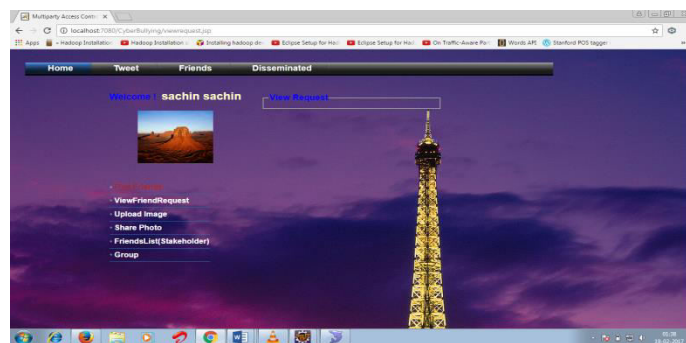


Fig 8 Accepting or Rejecting others friend request.

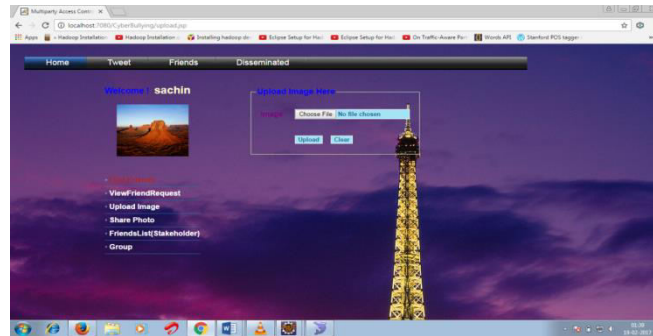


Fig 9 Uploading a Post or Image in your profile

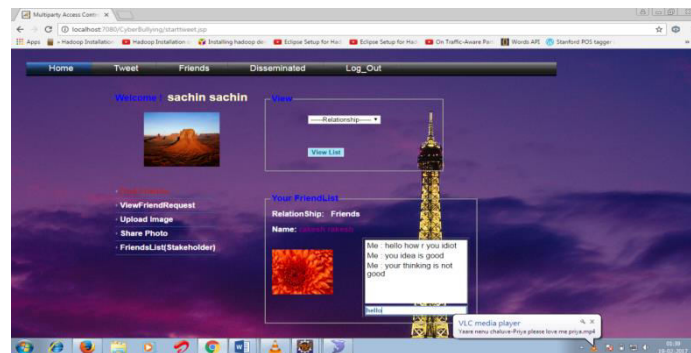


Fig 11 Abusive Comments in Post.

XI. CONCLUSION

The role of big data technologies in various industries is very obvious. The Digital Media have a huge user base, using Big data it will improve the application quality and effect of Digital Media products. We proposed a new strategy for detecting hate speech on Twitter. Our proposed method classifies tweets into hateful, offensive, and clean categories by automatically detecting hate speech patterns and the most common unigrams, as well as emotive and semantic aspects. For the binary classification of tweets into offensive and non-offensive, our suggested method achieves an accuracy of 87.4 The ternary classification of tweets into hateful, offensive, and clean had an accuracy of 78.4 percent. We will strive to construct a richer dictionary of hate speech patterns in the future, which may be used in conjunction with a unigram dictionary to detect hostile and offensive online messages. We'll conduct a quantitative investigation of the prevalence of hate speech across different genders and analyse it.

Algorithm 1: SVM-NN

Result: feature subsets classification accuracy

- 1 Identify list of reduced features using PCA, Correlation, Regression, SVM;
- 2 Set feature subsets to s;
- 3 Split your data into testing and training using 8 cross validation;
- 4 Set the training identifying labels to rLabel Set the testing identifying labels to sLabel
- 5 for each s do
- 6 Use SVM classification algorithm to Train the model using the training set, and the identifying labels rLabel.
- 7 Predict the output using the SVM trained model, and set the output decision-values to decisionV
- 8 Train NN model using decisionV, and the identifying labels rLabel .
- 9 Predict the testing set output using the SVM trained model, and set the output



decision-values to testingDecisionV
10 Test NN using the testingDecisionV, and NN
trained model, set the output to nnPredicted
11 Calculate NN prediction for each s accuracy
using the sLable, and nnPredicted
12 end
13 calculate the average accuracy for each fold
compared with the other subsets. As in spearman's rank-
order Correlation best pattern was (1 0 0 0 0 0 1 0 0 0
1 1 0 1 1 0), Multiple linear Regression best pattern was
(0011110111001111), and Wrapper-SVM best pattern was
(0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1). The detailed accuracy
results of this experiment were explained in Figure

REFERENCES

- [1] Karthik Dinakar, Roi Reichart, Henry Lieberman “Modeling the Detection of Textual Cyberbullying” MIT Media Lab, computer Science & Artificial Intelligence Laboratory Massachusetts Institute of Technology Cambridge, MA 02139 USA.
- [2] Mazur (2017), Kearon & Harrison (2019) Orgad (2019) and Murthy (2018) “To be wary of data collected from social media to a certain degree, because they may be the result of active design efforts by users”
- [3] “Users of the world, unite! The challenges and opportunities of Social Media” ESCP Europe, 79 Avenue de la Republique, F-75011 Paris, France.
- [4] “Cyber Bullying Detection Using Social and Textual Analysis” Department of Applied Computer Science. The University of Winnipeg, Winnipeg, MB, Canada, The Media Lab Massachusetts Institute of Technology Cambridge, MA, USA, Department of Computer Science University at Albany – SUNY Albany, NY, USA.
- [5] (2018) Human or ‘bot’? doubts over italian comic beppe grillo’s twitter followers.
- [6] (2018) How concerned are you that there are fake accounts and bots on social media platforms that are used to try to sell you things or influence you?
- [7] P. Patel, K. Kannoopatti, B. Shanmugam, S. Azam, and K. C. Yeo, “A theoretical review of social media usage by cyber-criminals,” in Computer Communication and Informatics (ICCCI), 2017 International Conference on. IEEE, 2017, pp. 1–6.
- [8] R. Kaur and S. Singh, “A survey of data mining and social network analysis based anomaly detection techniques,” Egyptian informatics journal, vol. 17, no. 2, pp. 199–216, 2016.
- [9] Y. Boshmaf, D. Logothetis, G. Siganos, J. Ler’ia, J. Lorenzo, M. Ripeanu, K. Beznosov, and H. Halawa, “Integro: Leveraging victim prediction for robust fake account detection in large scale osns,” Computers & Security, vol. 61, pp. 142–168, 2016.
- [10] Y. Boshmaf, M. Ripeanu, K. Beznosov, and E. Santos-Neto, “Thwarting fake osn accounts by predicting their victims,” in Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security. ACM, 2015, pp. 81–89.
- [11] B. Viswanath, M. A. Bashir, M. Crovella, S. Guha, K. P. Gummadi, B. Krishnamurthy, and A. Mislove, “Towards detecting anomalous user behavior in online social networks.” in USENIX Security Symposium, 2014, pp. 223–238.
- [12] A. S. M. Salih and A. Abraham, “Novel ensemble decision support and health care monitoring system,” Journal of Network and Innovative Computing, vol. 2, no. 2014, pp. 041–051, 2014.
- [13] G. Wang, T. Konolige, C. Wilson, X. Wang, H. Zheng, and B. Y. Zhao, “You are how you click: Clickstream analysis for sybil detection.” in USENIX Security Symposium, vol. 9, 2013, pp. 1–008.
- [14] L. Alvisi, A. Clement, A. Epasto, S. Lattanzi, and A. Panconesi, “Sok: The evolution of sybil defense via social networks,” in Security and Privacy (SP), 2013 IEEE Symposium on.



INNO  **SPACE**
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi[®]
cross **ref**

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 **9940 572 462**  **6381 907 438**  **ijircce@gmail.com**



www.ijircce.com

Scan to save the contact details