INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 1806-1810

FAKE PROFILE IDENTIFICATION IN SOCIAL NETWORK

Sonu Sharma[1], Pinnamcherla Aravind[2], Tirandas Charitha[3], Saggurthi Rani[4]

[1]Assistant Professor, Information Technology, ACE Engineering College, Telangana, India.

[2,3,4]Information Technology, ACE Engineering College, Telangana, India.

ABSTRACT

At present social network sites are part of the life for most of the people. Every day several people are creating their profiles on the social network platforms and they are interacting with others independent of the user's location and time. The social network sites not only providing advantages to the users and also provide security issues to the users as well their information. To analyse who are encouraging threats in social network we need to classify the social networks profiles of the users. From the classification, we can get the genuine profiles and fake profiles on the social networks. Traditionally, we have different classification methods for detecting the fake profiles on the social networks. But, we need to improve the accuracy rate of the fake profile detection in the social networks. In this paper we are proposing Machine learning and Natural language Processing (NLP) techniques to improve the accuracy rate of the fake profiles detection. We can use the Support Vector Machine (SVM) and Naïve Bayes algorithm.

Keywords: Network Sites, Social Network, investigation, Fake Profiles, Machine Learning, NLP, SVM.

1. INTRODUCTION

Social networking has become a widely used activity on the internet, attracting millions of users spending billions of minutes on platforms like Facebook, Twitter, and others. However, while these platforms offer various ways for people to interact and share information, ensuring security and privacy remains a significant challenge. Sharing personal information on social networks exposes individuals to various risks, including identity theft, where someone uses another person's information for personal gain. This has been a major concern, affecting millions worldwide, leading to financial loss, legal issues, damaged reputation, and strained relationships. Many social networking services lack stringent security measures, defaulting to minimal privacy settings, making them prime targets for fraud and abuse. The issue is compounded by the fact that most social networks do not verify user accounts thoroughly and have weak privacy policies. Moreover, users are required to provide accurate information when creating accounts, which, if compromised, can lead to severe consequences if misused or hacked. Online profiles in these networks contain static (provided by users) and dynamic (system-generated) information. While current research focuses on these aspects, most social networks display only a fraction of static profiles, with dynamic profiles often hidden from users, making it challenging to address. False profiles, those with inaccurate credentials, contribute to problems like privacy breaches, online bullying, and trolling on social networking sites. These profiles engage in malicious activities causing trouble for other users. Although platforms like Facebook have security measures like the Facebook Immune System (FIS) to protect against spam and phishing, they struggle to detect and mitigate the creation of fake profiles to a significant extent.

2. METHODOLOGY

2.1 Dataset

In our project, the database serves as the foundational repository for storing essential profile data attributes. These attributes encompass a range of crucial information, including usernames, follower counts, profile image URLs, and user locations. Each of these elements plays a pivotal role in shaping the user experience and functionality of our application. By meticulously organizing and maintaining this data within the database, we ensure seamless access and retrieval, enabling users to interact with the platform efficiently and effectively.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 1806-1810

2.2 Architecture

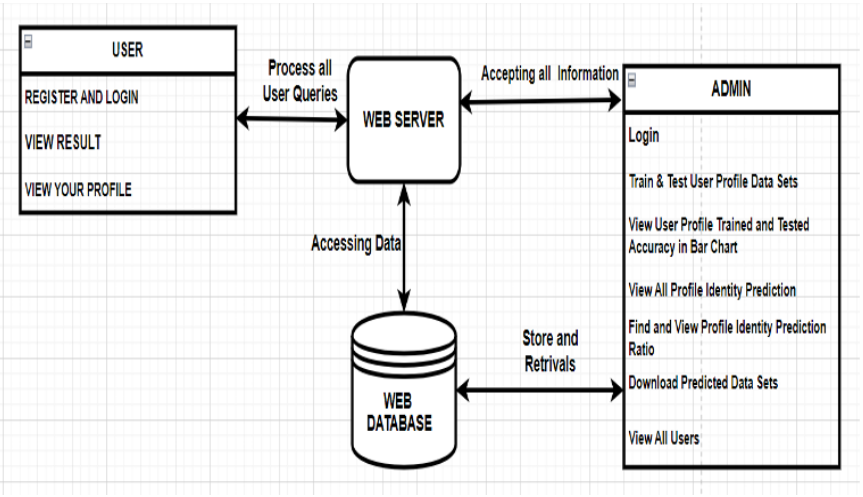

Figure 1: Architecture

3. LITERATURE SURVEY

[1] Michael Fire et al. (2012). "Strangers intrusiondetection-detecting spammers and fake profiles in social networks based on topology anomalies." Human Journal 1(1): 26-39.Günther, F. and S. Fritsch (2010). "neuralnet: Training of neural networks." The R Journal 2(1): 30-38

Fake and Clone profiles are creating dangerous security problems to social network users. Cloning of user profiles is one serious threat, where already existing user's details are stolen to create duplicate profiles and then it is misused for damaging the identity of original profile owner. They can even launch threats like phishing, stalking, spamming etc. Fake profile is the creation of profile in the name of a person or a company which does not really exist in social media, to carry out malicious activities. In this paper, a detection method has been proposed which can detect Fake and Clone profiles in Twitter. Fake profiles are detected based on number of abuse reports, number of comments per day and number of rejected friend requests, a person who are using fake account. For Profile Cloning detection two Machine Learning algorithms are used. One using Random forest Classification algorithm for classifying the data and Support Vector Machine algorithm. This project has worked with other ML algorithms, those training and testing results are included in this paper.

[2] Dr. S. Kannan, Vairaprakash Gurusamy, "Preprocessing Techniques for Text Mining", 05 March 2015.

Preprocessing is an important task and critical step in Text mining, Natural Language Processing (NLP) and information retrieval (IR). In the area of Text Mining, data preprocessing used for extracting interesting and non-trivial and knowledge from unstructured text data. Information Retrieval (IR) is essentially a matter of deciding which documents in a collection should be retrieved to satisfy a user's need for information. The user's need for information is represented by a query or profile, and contains one or more search terms, plus some additional information such as weight of the words. Hence, the retrieval decision is made by comparing the terms of the query with the index terms (important words or phrases) appearing in the document itself. The decision may be binary (retrieve/reject), or it may involve estimating the degree of relevance that the document has to query. Unfortunately, the words that appear in documents and in queries often have many structural variants. So before the information retrieval from the documents, the data preprocessing techniques are applied on the target data set to reduce the size of the data set which will increase the effectiveness of IR System The objective of this study is to analyze the issues of preprocessing methods such as Tokenization, Stop word removal and Stemming for the text documents.

[3] Shalinda Adikari and Kaushik Dutta, Identifying Fake Profiles in LinkedIn, PACIS 2014 Proceedings, AISeL

As organizations increasingly rely on professionally oriented networks such as LinkedIn (the largest such social network) for building business connections, there is increasing value in having one's profile noticed within the network. As this value increases, so does the temptation to misuse the network for unethical purposes. Fake profiles have an adverse effect on the trustworthiness of the network as a whole, and can represent significant costs in time and effort in building a connection based on fake information. Unfortunately, fake profiles are difficult to identify. Approaches have been proposed for some social networks; however, these generally rely on data that are not publicly available for LinkedIn profiles. In this research, we identify the minimal set of profile data necessary for identifying fake profiles in LinkedIn, and propose an appropriate data mining approach for fake profile identification. We demonstrate that, even

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 1806-1810

with limited profile data, our approach can identify fake profiles with 87% accuracy and 94% True Negative Rate, which is comparable to the results obtained based on larger data sets and more expansive profile information. Further, when compared to approaches using similar amounts and types of data, our method provides an improvement of approximately 14% accuracy.

[4] Z. Halim, M. Gul, N. ul Hassan, R. Baig, S. Rehman, and F. Naz,"Malicious users' circle detection in social network based on spatiotemporal co-occurrence," in Computer Networks and Information Technology (ICCNIT),2011 International Conference on, July, pp. 35–390.

Many people today use social networking sites as a part of their everyday lives. They create their own profiles on the social network platforms every day, and they interact with others regardless of their location and time. In addition to providing users with advantages, social networking sites also present security concerns to them and their information to them. We need to classify the social network profiles of the users to figure out who is encouraging threats on social networks. From the classification, we can figure out which profiles are genuine and which are fake. As far as detecting fake profiles on social networks is concerned, we currently have different classification methods. However, we must improve the accuracy of detecting fake profiles in social networks. We propose the use of a machine learning algorithm and Natural Language Processing (NLP) technique in this paper so as to increase the detection rate of fake profiles. This can be achieved using Support Vector Machines (SVM) and Naïve Bayes algorithms.

[5] Liu Y, Gummadi K, Krishnamurthy B, Mislove A," Analyzing Facebook privacy settings: User expectations vs. reality", in: Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference,ACM,pp.61–70.

Human Activity Recognition is a field concerned with the recognition of physical human activities based on the interpretation of pre-trained dataset for action recognition with time series classification. It is very difficult and challenging work to identify the one's action through the CCTV cameras for the security purpose and other purpose as well. Traditionally, hand-crafted features are relied upon to develop the machine learning models for activity recognition. However, that is a challenging task and requires a high degree of domain expertiseand feature engineering. With the development in deep beural networks, it is much easier as models can automatically learn features from raw sensor data, yielding improved classificationresults. In this project, I present a novel approach for human activity recognition using ensemble learning of multiple convolutional neural network (CNN) models. CNN model (ZDConv Model) are trained on the publicly available dataset in which the Kinetics 400dataset is used and ensembles of the model are created. The captured video was processed through the trained CNN model for the activity recognition.

4. RESULTS

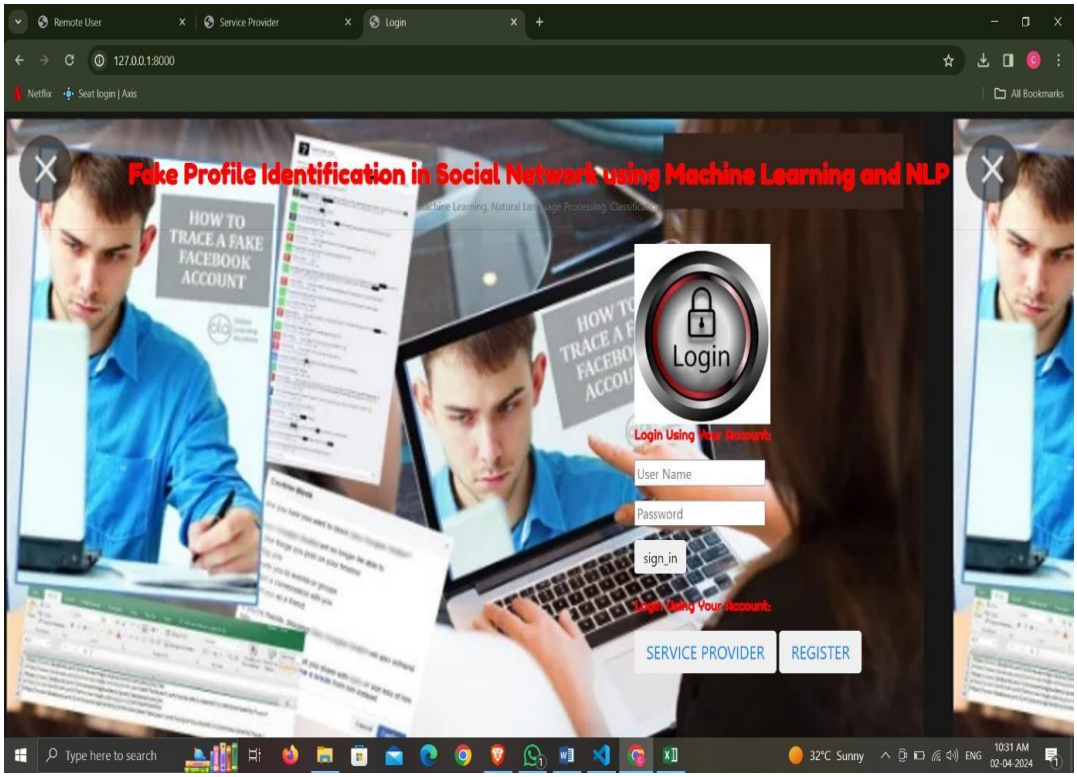

Figure 2: Login Screen

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 1806-1810

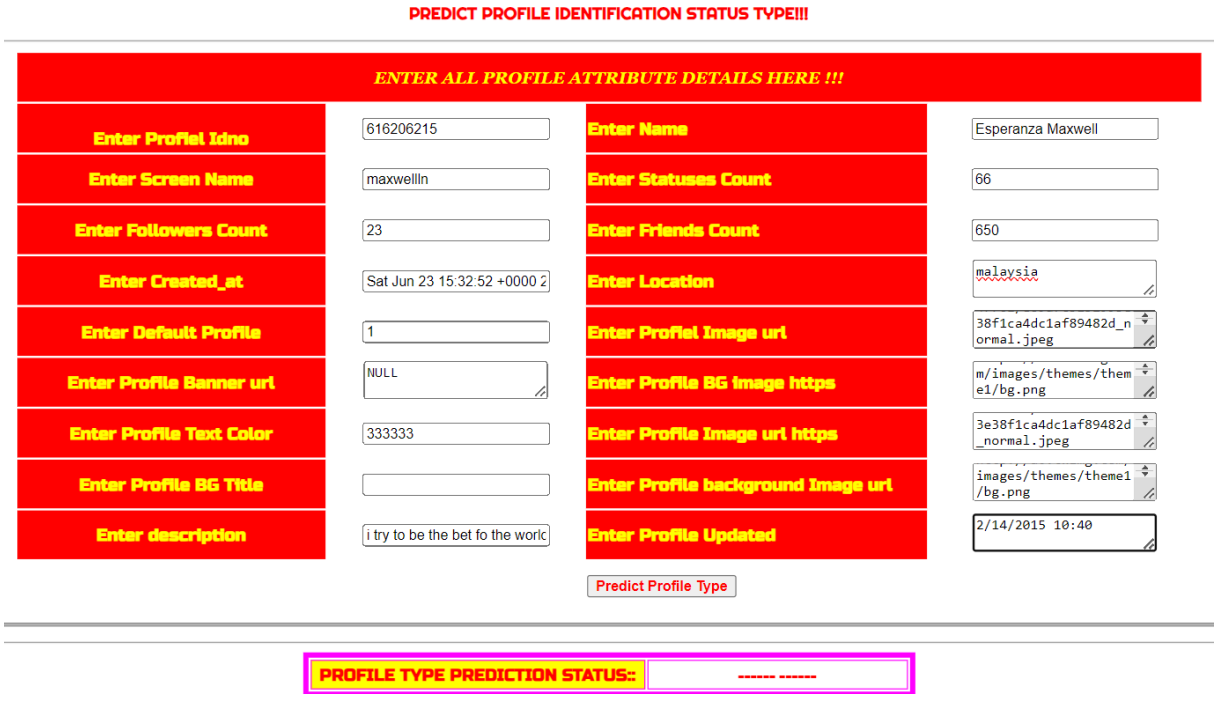

Figure 3: Data Entry Screen

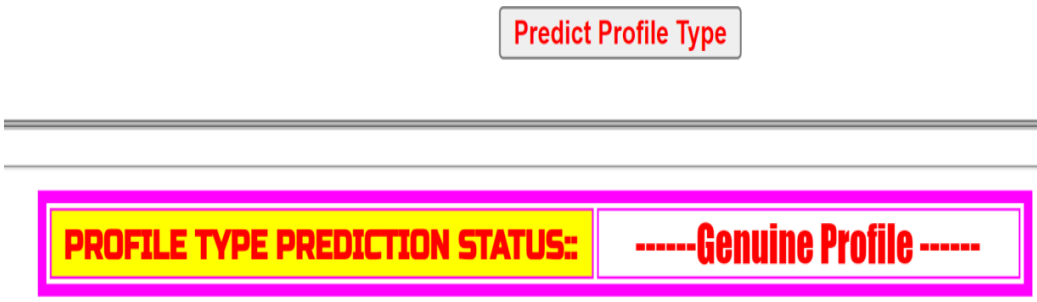

Figure 4: Profile Type Prediction Status

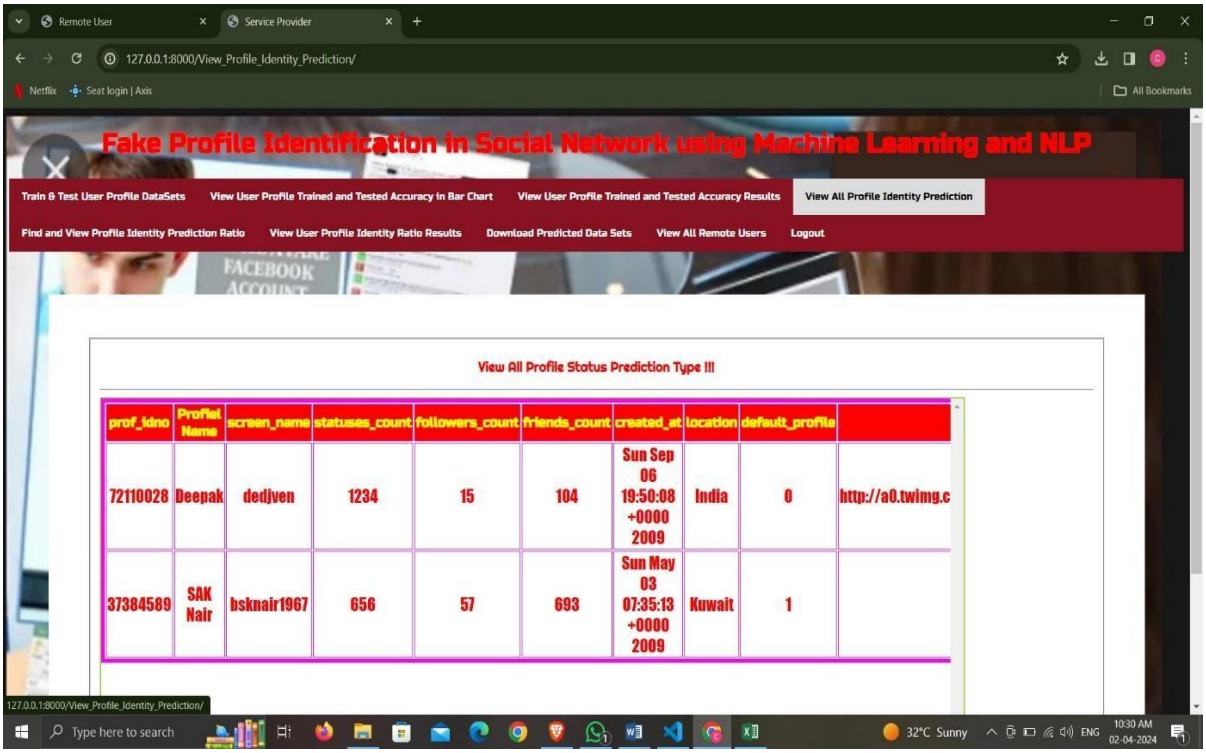

Figure 5: View all profile Identity Prediction

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 1806-1810

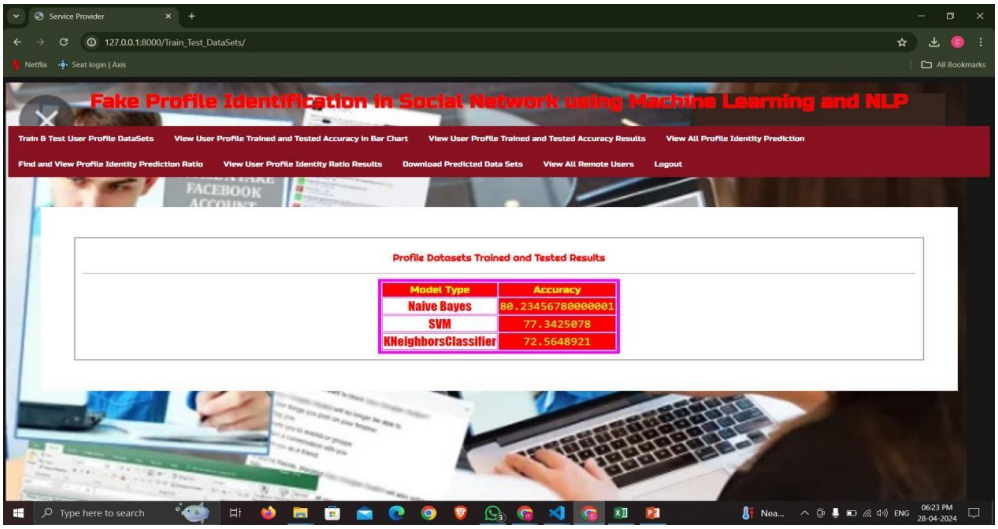

Figure 6: Profile Datasets Trained and Tested Results

5. CONCLUSION

In this paper, we have presented a novel approach leveraging machine learning algorithms in conjunction with natural language processing techniques to address the pervasive issue of fake profiles on social network sites. Our methodology offers a robust framework for identifying and flagging fraudulent accounts, particularly within the Facebook dataset utilized for experimentation. Through meticulous NLP pre-processing, we meticulously analyze the dataset, extracting meaningful insights that inform the subsequent classification process. Specifically, we deploy well-established machine learning algorithms such as Support Vector Machines (SVM) and Naïve Bayes, harnessing their capabilities to accurately classify profiles as genuine or fake. Furthermore, our experimental results underscore the efficacy of our proposed methodology, showcasing a notable improvement in the detection accuracy rate. By combining advanced NLP techniques with powerful learning algorithms, we achieve a refined approach to discerning authentic profiles from fraudulent ones, thereby enhancing the integrity of social network platforms. This research not only contributes to the ongoing efforts in combating online deception but also underscores the potential of interdisciplinary approaches in tackling complex challenges within the digital realm. As social networks continue to evolve, our work lays a foundation for further advancements in the realm of profile authentication and online security.

6. REFERENCES

[1] Michael Fire et al. (2012). "Strangers intrusion detection-detecting spammers and fake profiles in social networks based on topology anomalies." Human Journal 1(1): 26-39. Günther, F. and S. Fritsch (2010). "neuralnet: Training of neural networks." The R Journal 2(1): 30-38 Dr. S. Kannan, Vairaprakash Gurusamy, "Preprocessing Techniques for Text Mining", 05 March 2015.

[2] Shalinda Adikari and Kaushik Dutta, Identifying Fake Profiles in LinkedIn, PACIS 2014 Proceedings, AISeL Z. Halim, M. Gul, N. ul Hassan, R. Baig, S. Rehman, and F. Naz,"Malicious users' circle detection in social network based on spatiotemporal co-occurrence," in Computer Networks and Information Technology (ICCNIT),2011 International Conference on, July, pp. 35–390.

[3] Liu Y, Gummadi K, Krishnamurthy B, Mislove A," Analyzing Facebook privacy settings: User expectations vs. reality", in: Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference, ACM, pp.61–70.

[4] Mahmood S, Desmedt Y," Poster: preliminary analysis of google?'s privacy. In: Proceedings of the 18th ACM conference on computer and communications security", ACM 2011, pp.809–812.

[5] Stein T, Chen E, Mangla K," Facebook immune system. In: Proceedings of the 4th workshop on social network systems", ACM 2011, pp

[6] Saeed Abu-Nimeh, T. M. Chen, and O. Alzubi, "Malicious and Spam Posts in Online Social Networks," Computer, vol.44, no.9, IEEE2011, pp.23–28.

[7] J. Jiang, C. Wilson, X. Wang, P. Huang, W. Sha, Y.Dai, B. Zhao, Understanding latent interactions in online social networks, in: Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement, ACM, 2010, pp. 369–382

[8] Kazienko, P. and K. Musiał (2006). Social capital in online social networks. Knowledge-Based Intelligent Information and Engineering Systems, Springer