Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 16, Issue. 1: 2025
ISSN : **1906-9685**

FAKE PROFILE IDENTIFICATION IN SOCIAL NETWORK USING NLP

Mr.B.AMARNATH REDDY[1], K. SINDHU[2]
#1 Assistant Professor #2 M.C.A Scholar
Department of Master of Computer Applications,
Qis College of Engineering and Technology

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 analyze, 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.

INTRODUCTION
Social networking has end up a well-known recreation within the web at present, attracting hundreds of thousands of users, spending billions of minutes on such services. Online Social network (OSN) services variety from social interactions-based platforms similar to Face book or MySpace, to understanding dissemination-centric platforms reminiscent of twitter or Google Buzz, to Social interaction characteristic brought to present systems such as Flicker. The opposite hand, enhancing security concerns and protecting the OSN privateness still signify a most important bottleneck and viewed mission. When making use of Social network's (SN's), one of a kind men and women share one-of-a-kind quantities of their private understanding. Having our individual know-how entirely or in part uncovered to the general public, makes us excellent targets for unique types of assaults, the worst of which could be identification theft. Identity theft happens when any individual uses character's expertise for a private attain or purpose. During the earlier years, online identification theft has been a primary problem considering it affected millions of people's worldwide. Victims of identification theft may suffer unique types of penalties; for illustration, they would lose time/cash, get dispatched to reformatory, get their public image ruined, or have their relationships with associates and
loved ones damaged. At present, the vast majority of SN's does no longer verifies ordinary users‟ debts and has very susceptible privateness and safety policies. In fact, most SN's applications default their settings to minimal privateness; and consequently, SN's became a best platform for fraud and abuse. Social Networking

offerings have facilitated identity theft and Impersonation attacks for serious as good as naïve attackers. To make things worse, users are required to furnish correct understanding to set up an account in Social Networking web sites. Easy monitoring of what customers share on-line would lead to catastrophic losses, let alone, if such bills had been hacked. Profile information in online networks will also be static or dynamic. The details which can be supplied with the aid of the person on the time of profile creation is known as static knowledge, the place as the small print that are recounted with the aid of the system within the network is called dynamic knowledge. Static knowledge includes demographic elements of a person and his/her interests and dynamic knowledge includes person runtime habits and locality in the network. The vast majority of current research depends on static and dynamic data. However this isn't relevant to lots of the social networks, where handiest some of static profiles are seen and dynamic profiles usually are not obvious to the person network. More than a few procedures have been proposed by one of a kind researcher to realize the fake identities and malicious content material in online social networks. Each process had its own deserves and demerits.

LITERATURE SURVEY

1. **Author**: S. Kumar & N. Shah (2018)

Title: False Information on Web and Social Media
Merits:

- Combines user behavior and content analysis.
- Uses supervised learning for detection.

Demerits:

- Limited focus on multilingual content.
- High dependence on labeled data.

2. **Author**: S. Cresci, R. Di Pietro, M. Conti, M. Tesconi (2015)

Title: Fame for Sale: Efficient Detection of Fake Twitter Followers
Merits:

- High accuracy in detecting fake Twitter profiles.
- Utilizes graph-based features and account metadata.

Demerits:

- Twitter-specific; may not generalize to other platforms.
- Ineffective against sophisticated bots.

3. **Author**: P. Kaur, M. Kaur, H. S. Gill (2020)

Title: Detection of Fake Profiles on Online Social Networks using Machine Learning
Merits:

- Compares multiple ML classifiers (SVM, RF, etc.).
- High precision and recall for SVM.

Demerits:

- Small dataset used.
- Lack of real-time performance evaluation.

4. **Author**: B. Kudugunta & A. Ferrara (2018)

Title: Deep Neural Networks for Bot Detection
Merits:

- Uses LSTM networks for behavior analysis.
- Learns complex temporal patterns.

Demerits:

- Requires large datasets and training time.
- May overfit on unseen bot behaviors.

5. **Author**: T. Al-Qurishi et al. (2017)

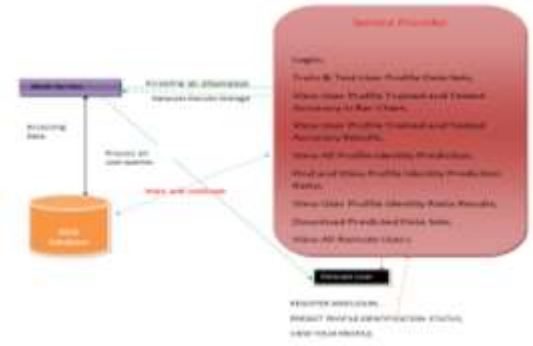
Title: Sybil Defense Techniques in OSNs
Merits:

- Good overview of identity deception.
- Highlights hybrid detection approaches.

Demerits:

- Theoretical; lacks implementation details.
- Focuses more on Sybil attacks than generic fake profiles.

System Architecture:



The image shows a **system architecture** for a web-based user identity prediction and profiling system involving three main components: the **Service Provider**, **Web Server**, and **Web Database**, with interactions from a **Remote User**.

1. Remote User

- **Actions:**
  - Register and log in to the system.
  - Predict their profile identification status.
  - View their own profile.
- **Purpose:** A user who accesses the system remotely to check their identity prediction and view their profile information.

2. Web Server

- **Roles:**
  - Acts as the **central processing unit**.
  - Processes all user queries.
  - **Accepts, processes, and stores information** such as:
    - Datasets,
    - Prediction results,
    - Accuracy outcomes.
  - Sends processed data to the web database.
- **Function:** Manages data flow between users, the service provider, and the web database.

3. Web Database

- **Roles:**
  - Stores and retrieves data.
  - Handles all user-related and dataset-related data for analysis and prediction.
- **Data Stored:**
  - User profiles,
  - Prediction results,
  - Accuracy metrics.

4. Service Provider

- **Actions Available:**
  - Login to the system.
  - Train and test user profile data sets.
  - View trained and tested accuracy in bar chart format.
  - View individual and overall prediction accuracy.
  - Find/view profile identity prediction ratio.
  - View/download predicted data sets.
  - Monitor all remote users.

- **Purpose:** The service provider acts as an **admin or system operator**, managing the models, visualizing data, and overseeing all remote user activities.

Implementation:

CONCLUSION

In this paper, we proposed machine learning algorithms along with natural language processing techniques. By using these techniques, we can easily detect the fake profiles from the social network sites. In this paper we took the Face book

Data set to identify the fake profiles. The NLP pre-processing techniques are used to analyze the dataset and machine learning algorithm such as SVM and Naïve Bayes are used to classify the profiles. These learning algorithms are improved the detection accuracy rate in this paper.

FUTURE ENHANCEMENT

In the future, the Fake Profile Identification system can be significantly enhanced by incorporating advanced deep learning models such as transformers (e.g., BERT, Roberta) to better understand and analyse complex linguistic patterns in user-generated content. Real-time detection mechanisms can be implemented to flag suspicious profiles instantly, improving platform security. Additionally, multilingual and code-mixed language support can be introduced to handle diverse language usage, especially in regions like India. The system can also be extended to work across multiple social media platforms, identifying identity duplication and coordinated fake profile networks. Adaptive learning models that continuously evolve with new patterns of fake behaviour will improve long-term effectiveness. Furthermore, the integration of user interaction analysis, such as likes, shares, and comments, can offer deeper insights into suspicious activity. The inclusion of explainable AI (XAI) will help users and administrators understand the reasons behind classification decisions. Moreover, incorporating voice or image verification using multimodal AI can enhance the robustness of profile authentication. Feedback loops involving user reports on false positives or negatives will enable continuous improvement. Lastly, collaboration with law enforcement through API-based reporting tools can help take real-world action against harmful fake profiles.

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). "Neural net: Training of neural networks." The R Journal 2(1): 30-38

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

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

[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.

[5] Liu Y, Gummadi K, Krishnamurthy B, Mis love A," Analysing Facebook privacy settings: User

expectations vs. reality", in: Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference, ACM, pp.61–70.

[6] 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.

[7] Stein T, Chen E, Mangla K," Facebook immune

system. In: Proceedings of the 4th workshop on social network systems", ACM 2011, pp

[8] Saeed Abu-Nimeh, T. M. Chen, and O. Alzubi,

"Maliciousand SpamPosts in OnlineSocialNetworks,"Computer,vol.44,no.9, IEEE2011,pp.23–

28.

[9] 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

[10] Kuzmenko, P. and K. Musiał (2006). Social capital in online social networks. Knowledge-Based Intelligent

Information and Engineering Systems, Springe

Authors:

Mr. B. Amarnath Reddy is an Assistant Professor in the Department of Master of Computer Applications at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. He earned his M.Tech from Vellore Institute of Technology(VIT), Vellore. His research interests include Machine Learning, Programming Languages. He is committed to advancing research and fostering innovation while mentoring students to excel in both academic and professional pursuits.

Ms. **K SINDHU** has received her MCA (Masters of Computer Applications) from QIS college of Engineering and Technology Vengamukkapalem(V), Ongole, Prakasam dist., Andhra Pradesh-