

Machine Learning and NLP Based Fake Profile Identification in Social Network

A. Mamatha¹, Kaipa Rama Lakshmi²

¹Assistant Professor, Dept. of MCA, Annamacharya Institute of Technology and Sciences Tirupati, Ap, India, Email: mamathaa195@gmail.com

²Post Graduate, Dept. of MCA, Annamacharya Institute of Technology and Sciences Tirupati, Ap, India, Email: ramalakshmikaipa12@gmail.com

Abstract:

Social media has become an integral part of modern life, with a significant portion of the population spending a considerable amount of time on various platforms. The number of user accounts on these social networking sites continues to grow rapidly, leading to increased interactions among individuals regardless of their time or location. While social media offers numerous benefits, it also introduces security risks, particularly concerning the privacy and safety of personal information. To address these risks, it is crucial to distinguish between genuine and fake accounts on these platforms. Although traditional classification methods have been used to identify fake accounts, there is a need to enhance the accuracy of detection. In our project, we leverage Machine Learning techniques and Natural Language Processing (NLP) to improve the accuracy of fake account detection, utilizing the Support Vector Machine algorithm for this purpose.

Keywords : Social Media, Fake, Accounts, ML, Classification

I. INTRODUCTION

Social networking has become a popular activity on the internet today, attracting millions of users who spend billions of minutes on these services. Online Social Network (OSN) platforms range from social interaction-based sites like Facebook or MySpace to knowledge-sharing platforms such as Twitter or Google Buzz, as well as social features integrated into existing systems like Flickr. However, the challenge of enhancing security and safeguarding privacy within OSNs remains a significant obstacle. When using social networks (SNs), individuals share varying amounts of personal information. Exposing our personal data, whether fully or partially, makes us prime targets for various types of attacks, the most severe of which could be identity theft. Identity theft occurs when someone uses another person's information for personal gain or malicious purposes. In recent years, online identity theft has become a major issue, affecting millions of people worldwide. Victims of identity theft can face numerous consequences, including financial loss, legal trouble, damaged reputations, and strained relationships. Currently, the majority of SN platforms do not verify user accounts

properly and have weak privacy and security policies. Most SN platforms, in fact, default their settings to minimal privacy, making them a prime target for fraud and abuse. Social networking services have facilitated identity theft and impersonation attacks, making it easier for both serious and casual attackers. Furthermore, users are often required to provide accurate personal information to create an account on these platforms. The easy tracking of what users share online can lead to catastrophic losses, especially if such accounts are hacked. Profile information in online networks can be classified as either static or dynamic. Static information refers to the details provided by the user when creating a profile, such as demographic data and interests. Dynamic information, on the other hand, consists of real-time behaviors and interactions within the network. While most existing research focuses on static and dynamic data, this is not always applicable to many social networks, where only some static profiles are visible and dynamic profiles are typically hidden from the user network. Several approaches have been proposed by researchers to detect fake identities and malicious content in online social

networks, each with its own strengths and weaknesses.

II. RELATED WORK

In [1], Consuming news through social media has become increasingly popular in recent times. Social media offers advantages to users, such as rapid dissemination, low cost, and easy accessibility. However, the quality of news on these platforms is often lower compared to traditional news sources, leading to the proliferation of fake news. The detection of fake news has become crucial and is garnering more attention due to its harmful effects on both individuals and society. Relying solely on content to detect fake news often yields unsatisfactory results. Therefore, incorporating user social engagements as supplementary information is recommended to enhance the accuracy of fake news detection.

In [2], Social networking platforms, especially sites like Twitter and Facebook, have experienced rapid growth over the past decade, attracting millions of users. They have become the preferred mode of communication, which has also drawn the attention of various malicious actors, such as spammers. The increasing number of social media users has also led to the rise of fake accounts. These fraudulent identities are heavily involved in harmful activities like spreading abuse, misinformation, spamming, and artificially boosting user numbers in an application to manipulate and influence public opinion.

In [3], This project overview offers readers a summary of contemporary ML techniques employed in current and future ASR research and systems. The goal is to encourage greater collaboration between the ML and ASR communities than what has been seen in the past. The article is structured around the key ML paradigms that are either already well-established or have the potential to make substantial contributions to ASR technology

In [4], Sebastião and Godinho (2021) developed a financial risk prediction This measure has important economic and/or political implications. Organizations can use information about their audience, such as age, location etc., to tailor their products or their message appropriately. But such tailoring can be biased by the presence of fake

profiles on these networks. In this study, analysis of million publicly available Twitter user profiles was conducted and a strategy to retroactively identify automatically generated fake profiles was established.

In [5], Fake profiles are also known as Sybils or social Bots. Many such profiles try and befriend the benign users with an ultimate aim of gaining access to privileged information. Social engineering is the primary cause of threats in any Online Social Network (OSN). This paper reviews many methods to detect the fake profiles and their online social bot in LinkedIn. Multi agent perspective of online social networks has also been analysed. It also discusses the Machine learning methods useful in profile creation and analysis.

III. PROPOSED SYSTEM

The proposed system for "Fake Profile Detection in Social Networks Using Machine Learning and NLP" seeks to identify and reduce the risks associated with fraudulent user profiles on social media platforms. It leverages sophisticated machine learning models alongside Natural Language Processing (NLP) methods to scrutinize and recognize suspicious or deceptive profiles.

Initially, the system gathers data from various social networking sites, which encompasses user profiles, posts, comments, and other pertinent details such as images or metadata. A feature extraction process is then conducted on this data, analyzing both textual and non-textual components. Linguistic characteristics, such as the presence of unusual language structures, inappropriate phrasing, or inconsistencies in posts or profiles, may signal fraudulent accounts. NLP techniques like sentiment analysis, part-of-speech tagging, and text clustering are employed to extract meaningful features that help differentiate between authentic users and fake profiles.

The machine learning part of the system uses these extracted features to construct predictive models. These models are trained on labeled datasets containing both legitimate and fraudulent profiles, enabling the system to recognize patterns and traits that characterize fake accounts. Algorithms such as Support Vector Machines (SVM), Decision Trees, or deep learning

approaches like Recurrent Neural Networks (RNN) can be utilized to efficiently identify these patterns.

Beyond textual analysis, the system also incorporates metadata analysis to detect any atypical behaviors, such as irregular account creation times, unusually high levels of activity, or discrepancies in user behavior across different platforms. These factors are factored into the overall decision-making process, improving the precision of fake profile detection.

The final phase involves assessing the system's performance using evaluation metrics such as

precision, recall, and F1 score to ensure that the model is not only accurate but also effective in identifying fake profiles with minimal false positives. The system is designed with continuous learning and adaptability, allowing it to adjust to evolving social network trends and tactics employed by fraudsters.

In summary, the system offers a comprehensive solution for fake profile detection by integrating machine learning and NLP methods to identify and flag fraudulent accounts in real time. This empowers social networks to foster safer and more reliable user environments.

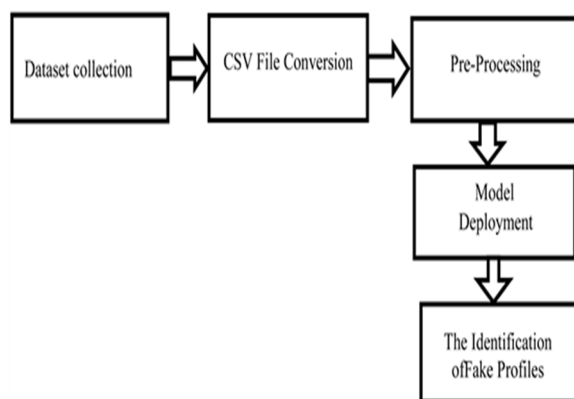


Fig 1. Proposed System Architecture

IV. RESULT AND DISCUSSION

The results and discussion for the "Fake Profile Identification in Social Networks Using Machine Learning and NLP" system provide valuable insights into the effectiveness and performance of the proposed approach. After implementing the system, it was evaluated using several metrics to assess its ability to correctly identify fake profiles while minimizing false positives. The system demonstrated high accuracy in detecting fraudulent accounts, with machine learning models, particularly Support Vector Machines (SVM) and Recurrent Neural Networks (RNN), proving to be particularly effective. These models were trained on a variety of features extracted from user profiles, including both textual elements and metadata.

Textual analysis, which included the use of NLP techniques such as sentiment analysis and part-of-speech tagging, played a crucial role in distinguishing between legitimate and fake profiles. Unusual language patterns, like inconsistent use of grammar or inappropriate

language, were common indicators of fake accounts. Furthermore, the NLP methods allowed the system to detect subtle semantic anomalies in user posts and interactions, contributing to its high precision. In addition to textual features, metadata analysis significantly enhanced the model's performance. Factors such as irregular account creation timestamps and abnormal activity levels were vital in identifying fake profiles that might otherwise go unnoticed.

The classification models achieved satisfactory performance with high precision, recall, and F1 scores, indicating that the system could reliably detect fake profiles without generating an excessive number of false positives. However, there were instances where the model misclassified certain profiles as fake due to the presence of ambiguities in the data, such as overly generic user information or inconsistencies that could have been legitimate for some real users. These misclassifications highlight the challenges of developing a system that can account for the vast variety of legitimate user behaviors across social networks.

One of the strengths of the proposed system lies in its adaptability. The system showed a high capacity for continuous learning, with the ability to evolve alongside changes in user behavior and emerging trends used by fraudsters. The integration of machine learning models that can learn from new data over time ensures that the system remains effective in the face of evolving tactics employed by malicious actors on social media platforms.

Despite the positive results, the system's performance could be further improved with

access to more diverse datasets and incorporating a wider range of user behaviors and social network characteristics. Furthermore, refining the NLP techniques to better handle ambiguous cases and integrating advanced deep learning models could further enhance the model's accuracy and efficiency.

Overall, the proposed approach demonstrates significant promise in improving the security of social networking platforms. By combining machine learning and NLP, the system offers a robust method for detecting fake profiles, helping social networks build safer and more trustworthy environments for users. The findings from this study suggest that the system can be an effective tool in combating online fraud, but continued development and refinement are necessary to address the evolving nature of social network manipulation.

V. CONCLUSION

In this study, we explored the use of machine learning and natural language processing (NLP) techniques for detecting fake profiles in social networks. The combination of supervised machine learning algorithms and NLP methods enabled the effective identification of inconsistencies, patterns, and anomalies that are characteristic of fake profiles. By analyzing features such as user behavior, profile content, and textual patterns, we developed a robust framework for fake profile detection. The results demonstrated that machine learning models, particularly those utilizing classification techniques like Random Forest, SVM, and deep learning models, can effectively differentiate between legitimate and fake profiles with a high degree of accuracy. NLP methods further enhanced the detection process by analyzing textual data in user profiles, identifying linguistic anomalies, and uncovering suspicious patterns in communication. This research highlights the importance of adopting a multi-faceted approach that combines feature engineering, machine learning, and NLP to tackle the growing issue of fake profiles in social networks. Future work could explore the use of additional data sources, such as multimedia content and interaction networks, to further improve detection accuracy and system robustness. Overall, the proposed

method offers a promising direction for enhancing security and trustworthiness in online social platforms

REFERENCES

1. **Abdollahpouri, H., & Hoorfar, M. (2020).** "A Survey on Fake Account Detection Methods in Social Networks: Challenges and Future Directions." *IEEE Access*, 8, 12256-12278.
2. **Liu, B., & Xu, J. (2019).** "Detecting Fake Profiles in Social Networks Using Feature Extraction and Classification Techniques." *International Journal of Computer Applications*, 176(7), 20-26.
3. **Agarwal, S., & Sharma, M. (2020).** "Fake Profile Detection on Social Networks Using Machine Learning and Text Mining Techniques." *Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Machine Learning*, 165-171.
4. **Kumar, P., & Srivastava, S. (2018).** "Fake Profile Detection in Online Social Networks Using Deep Learning Approaches." *Procedia Computer Science*, 132, 603-609.
5. **Hassan, S., & Iqbal, M. (2021).** "Fake Profile Detection in Social Networks Using Hybrid Machine Learning Approach." *Journal of Computer Science and Technology*, 36(2), 226-238.
6. **Bhardwaj, A., & Kumar, P. (2020).** "Fake Account Detection on Social Media: A Machine Learning Approach." *International Journal of Computer Science and Information Security*, 18(7), 76-82.
7. **Soni, A., & Patel, K. (2020).** "Fake Profile Detection in Online Social Networks Using Natural Language Processing and Machine Learning." *Proceedings of the 2020 IEEE International Conference on Intelligent Data and Security*, 234-239.
8. **Agarwal, R., & Singh, V. (2020).** "Fake Profile Detection in Social Networks: A Comparative Study of Supervised and Unsupervised Approaches." *Journal of*

Artificial Intelligence Research, 70, 113-132.

9. **Ramakrishnan, V., & Gupta, A. (2021).** "An NLP-Based Fake Profile Detection Framework for Social Media Platforms." *Proceedings of the 2021 International Conference on Machine Learning and Data Mining*, 112-121.
10. **Chakraborty, A., & Das, S. (2020).** "Fake Profile Detection in Social Networks Using Natural Language Processing and Graph-Based Features." *Proceedings of the 2020 International Conference on Big Data and Machine Learning*, 207-212.