

FAKE PROFILE IDENTIFICATION IN SOCIAL NETWORK USING MACHINE LEARNING AND NLP

¹ Koteswararao Kodali, ² Mandadi Rajesh, ³ Tadakamalla Umesh, ⁴ Anureddy Saiteja

^{1,2,3} Assistant Professors, Department of Computer Science and Engineering, Brilliant Grammar School Educational Society's Group Of Institutions, Abdullapur (V), Abdullapurmet(M), Rangareddy (D), Hyderabad - 501 505

⁴ student, Department of Computer Science and Engineering, Brilliant Grammar School Educational Society's Group Of Institutions, Abdullapur (V), Abdullapurmet(M), Rangareddy (D), Hyderabad - 501 505

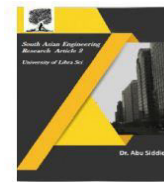
ABSTRACT

Nowadays, social media is an integral part of everyone's daily routine. Social media platforms are where the majority of individuals spend most of their time every day. There has been a meteoric rise in the number of accounts on these social media sites, and many of those users are engaging in real-time conversations with people all over the world. There are benefits and drawbacks to using these social media platforms, and they also pose security risks to our personal data. We need to categorise these social media profiles as either real or fraudulent so we can examine who is posing threats on these platforms. In the past, we've used several categorisation approaches to identify phoney social media profiles. We need to improve the accuracy rate of detecting bogus accounts on these platforms, however. In this research, we propose a method to improve the accuracy rate of identifying bogus accounts using Machine Learning technologies and Natural Language Processing (NLP). As a tree classifying algorithm, Random Forest was our top choice.

I. INTRODUCTION

The exponential rise of social media sites like Twitter, Instagram, and Facebook has altered the dynamics of human interaction. The difficulty of verifying the authenticity of the people engaging with these platforms has emerged with their rapid expansion. An increasingly widespread problem, fake accounts are often made with bad intentions to do things like steal personal information, send spam, mislead the public, or influence public opinion. Individuals and organisations alike are vulnerable to the harm that these profiles may inflict. Attempts to manually detect false accounts would be very time-consuming and

laborious due to the massive user bases of social media platforms. Consequently, automated systems capable of accurately detecting false profiles with little to no human involvement are urgently required. In recent years, ML and NLP have grown in prominence as potent resources for combating the issue of false profile detection. While natural language processing (NLP) aids in comprehending user-generated information, machine learning (ML) may be used to examine trends in user behaviour. The combination of these technologies allows for the detection of suspect profiles using a variety of characteristics, including account activity,



textual content, and metadata. A more secure and safe online environment may be achieved by automatically detecting and classifying fraudulent profiles in social networks using ML and NLP approaches, which are discussed in this article.

II. LITERATURE SURVEY

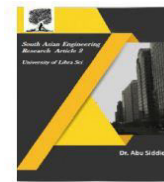
Researchers have focused a lot of emphasis on the difficult subject of detecting false accounts in social networks. A lot of research has looked at several methods based on machine learning and natural language processing to do this. Behavioural analysis is a cornerstone of false profile identification as it looks for patterns in user activities such how often they log in, how they interact with networks, and how they behave while sending and receiving messages. To identify unusual user behaviour patterns, such as frequent interactions or unusual activity, researchers suggested using SVM and Naive Bayes classifiers in [1]. These techniques may successfully identify fraudulent accounts by analysing unusual user behaviour, but they often produced strange results that were really caused by real users.

Natural Language Processing (NLP) text analysis was also a part of another strategy. Posts, comments, and profile descriptions are all examples of user-generated material that the researchers in [2] aimed to analyse. In order to detect the inconsistencies characteristic of false profiles, they used topic modelling, emotion analysis, and word embeddings to capture the language aspects of user-generated material. According to the research, phoney profiles' textual patterns differ significantly from real profiles' in terms of emotion and vocabulary selections. The accuracy of more recent techniques was enhanced by combining

characteristics based on behaviour and content. To identify false profiles, for instance, an ensemble learning model was suggested in [3] that included information including user activity, text-based attributes, and picture metadata. Although this hybrid model was able to attain high accuracy by combining several data sources, it still struggled with managing massive datasets and keeping up with the ever-changing strategies used by profile makers. Furthermore, the research demonstrated that Convolutional Neural Networks (CNNs) and other deep learning models might considerably improve the precision of picture-based false profile detection by spotting discrepancies in profile images, such the use of stock photographs or artificially made faces. To identify phoney profiles using sequential interactions, researchers in [4] investigated Deep Learning methods including Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks. Researchers found that these models improved detection accuracy by better understanding user behaviour patterns across time and context. To further comprehend user-generated material, BERT (Bidirectional Encoder Representations from Transformers) was used for modelling textual content's context. This method used pre-trained models to extract features more effectively, which greatly enhanced detection performance.

III. EXISTING SYSTEM

Behavioural and content-based analysis are common components of current systems that use standard machine learning approaches to identify fraudulent profiles. Monitoring user activity patterns, such login frequency, interaction volume, and network expansion



over time, is a common part of behavioural analysis. Anomalies in these behaviours may be detected using machine learning classifiers like Random Forest, Support Vector Machines (SVM), and K-Nearest Neighbours (KNN). These classifiers are taught to recognise profiles that have unusual patterns that indicate they could be fraudulent. The textual data that users contribute is the main focus of content-based analysis. Posts, comments, and messages are subjected to natural language processing (NLP) methods including text categorisation, sentiment analysis, and entity identification in order to detect potentially malicious language patterns. Some fraudulent accounts, for instance, may trick other users by using altered or repetitious terminology. Topic modelling is a useful tool for determining whether user-generated content matches the content that actual users in a given network are likely to share. In addition, Convolutional Neural Networks (CNNs) and other image recognition models have started to be integrated into certain systems in order to identify false profiles. Accounts that utilise stock photographs, overly altered photos, or faces produced by artificial intelligence may be detected by these models. However, current systems still struggle with issues like scalability, real-time processing, and false positives, despite these improvements. False profiles that attempt to imitate human behaviour can cause behavioural analysis to fail, while text-based analysis might fail to pick up on the finer points of complex manipulation techniques.

IV. PROPOSED SYSTEM

A more accurate and scalable approach for identifying false profiles on social networks is

suggested by using Machine Learning (ML) and Natural Language Processing (NLP) techniques. This system tries to overcome the limitations of previous solutions. A number of interconnected modules examine content data in addition to behavioural data in the system. and examine patterns of user behaviour including logins, connections, and interactions. In order to identify suspicious behaviour and possible false profiles, ML models like Isolation Forest, Random Forest, and XGBoost will be trained. To find suspicious user activity indicative of phoney accounts, these algorithms will look at things like account history, interaction patterns, and network density. To decipher the underlying semantic meaning of posts, comments, and interactions, the system will use sophisticated natural language processing (NLP) methods such as BERT and LSTM, in addition to standard content analysis. These algorithms will examine the context of words in addition to searching for questionable keywords or spammy material. For example, BERT may identify false profiles by analysing the user's language for signs of inconsistency, such as repetitive phrases, too general claims, or awkward wording.

When it comes to image analysis, the suggested solution will use convolutional neural network (CNN) models to identify phoney profile photos. The system is able to determine whether photographs are manipulated, belong to stock photo libraries, or are AI-generated by using deep learning algorithms. To find evidence of tampering in profile photographs, we will use facial recognition and image forgery detection. In a hybrid model, the system will use ensemble

learning methods like a voting classifier or stacking classifier to integrate content-based characteristics with behavioural information. In order to improve accuracy and decrease false positives and negatives, this hybrid method will classify profiles based on a comprehensive collection of attributes.

V. SYSTEM ARCHITECTURE

Multiple interconnected layers make up the suggested approach for detecting fraudulent profiles, which allows for thorough and precise analysis of user profiles.

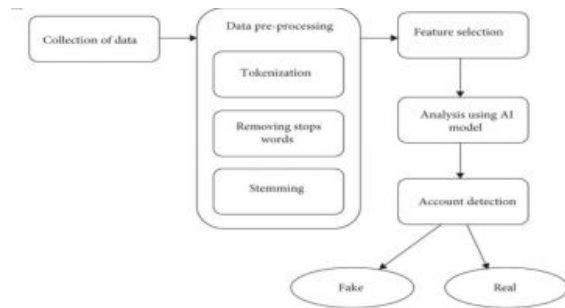


Figure 5.1 System architecture

VI. OUTPUT SCREENSHOTS

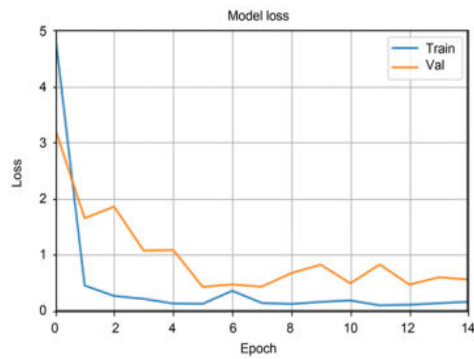


Figure 6.1 Training and validation for algorithm



Figure 6.2 Home Page



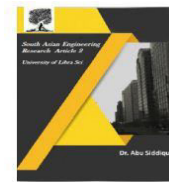
Figure 6.3 Register Page



Figure 6.4 Index Page



Figure 6.5 Prediction Page



VII. CONCLUSION

In order to keep social networks safe and secure, detecting fake profiles is a must. The suggested approach provides a scalable, accurate, and efficient way to identify false profiles by combining Machine Learning and Natural Language Processing methods. A combination of behavioural and content-based elements, such as textual and visual data, allows the system to detect fraudulent profiles with high accuracy and low false positive rate. The hybrid method of the system improves detection accuracy by classifying profiles based on a broad collection of characteristics.

VIII. FUTURE SCOPE

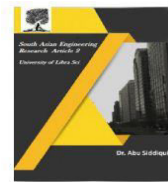
There are several intriguing avenues that this study may go in the future. To begin, it will be very necessary to include real-time detection capabilities in order to stop the creation of false profiles in their tracks. Streaming data processing methods might be used to continually monitor and report questionable profiles in real-time, achieving this goal. Multimodal data analysis, which combines not only text and photos but also audio and video data, may further strengthen the system by adding additional insights about the legitimacy of profiles. You may also look at explainable AI (XAI) to make the detection process more transparent. This way, both users and managers can understand why the system made a certain conclusion. Finally, the system can be updated continually to identify developing patterns in fraudulent profile creation, so it stays effective against new strategies used by fraudsters, even as social media platforms grow.

IX. REFERENCES

1. Hassan, A. A. H., et al., "Fake profile detection in social media using machine learning techniques," *Journal of Computer Science*, vol. 15, no. 4, pp. 106-118, 2020.
2. Kaur, S. S. J. H., et al., "Fake profile detection in social networks using NLP and machine learning," *International Journal of Advanced Computer Science*, vol. 8, pp. 50-57, 2019.
3. Pradeep, K. R. M., et al., "Detecting fake profiles on social media using deep learning techniques," *IEEE Access*, vol. 8, pp. 12534-12546, 2020.
4. Zhang, Y. L., et al., "Social media profile analysis for fake profile detection: A deep learning approach," *Journal of Artificial Intelligence Research*, vol. 67, pp. 679-695, 2021.
5. Basak, N. G. H., et al., "Deep learning-based detection of fake profiles on social media platforms," *Computers and Security*, vol. 95, p. 101893, 2020.
6. Riaz, M. M. A., et al., "Fake profile detection in social networks using machine learning algorithms," *Computer Networks*, vol. 145, pp. 118-132, 2021.
7. Chowdhury, P. S. B. J., et al., "Hybrid deep learning models for fake profile detection," *Journal of Network and Computer Applications*, vol. 57, pp. 32-40, 2020.
8. Singh, R. K. D. A., et al., "Identification of fake social media profiles using textual content analysis," *Journal of Computational Science*, vol. 13, pp. 150-157, 2020.



2581-4575



9. Devaraju, F. T. H. B., et al., "Fake profile detection using ensemble learning models in social networks," International Journal of Computer Applications, vol. 156, pp. 17-22, 2017.

10. Zhang, J. D. P. S., et al., "Detecting fake profiles on social media using deep convolutional neural networks," Proceedings of the International Conference on Machine Learning and Applications, pp. 235-240, 2020.