Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 5, May 2024

Multi Relations and Context-Aware Fake User Profile Detection Using Ensemble and LSTM Models

Dr. R. Nagalakshmi M.E.,Ph.D,
Computer Science and Engineering
SRM Institute of Science and Technology
Chennai, India

Abirami M
Computer Science And Engineering
SRM Institute of Science and Technology
Chennai, India

Clament Clive
Computer Science and Engineering
SRM Institute of Science and Technology
Chennai, India

Sabapathy MS
Computer Science and Engineering
SRM Institute of Science and Technology
Chennai, India

***Abstract*— Online Social Networks (OSNs) serve as crucial platforms for internet users, facilitating various daily activities such as content sharing, news reading, messaging, product reviews, and event discussions. However, these networks also attract different types of spammers, including online fraudsters, advertising campaigners, catfishes, and social bots. These malicious actors exploit the trust network by creating fake profiles to disseminate content and perpetrate scams, posing significant harm to users and service providers alike.**

From the perspective of OSN service providers, fake profiles tarnish the network's reputation and result in bandwidth loss. Detecting these malicious users requires substantial manpower and sophisticated automated methods. With the widespread integration of OSNs into daily social life, issues like fake profiles and online impersonation have escalated, yet viable solutions remain elusive. In this paper, a framework is proposed for the automatic identification of fake profiles, offering efficiency and scalability. This framework employs classification techniques such as Ensemble Learning and Long Short-Term Memory (LSTM) to classify profiles into fake or genuine classes. By automating the detection process, this method becomes feasible for OSNs with millions of profiles, eliminating the need for manual examination.

***Keywords*— SMP (Social Media Platform), OSN (Online Social Networks), ML (Machine Learning), DL (Deep Learning), LSTM (Long Short-Term Memory), RMSE (Root Mean Square Error), MSE (Mean Square Error).**

I. INTRODUCTION

Online Social Network Analysis (OSNA) stands at the forefront of emerging research domains. An online social network (OSN) refers to the interconnected nodes (individuals, organizations, nations, or web pages) worldwide linked by various relationships such as interactions, distances, hyperlinks, etc. Since its inception, OSNs have revolutionized how people communicate, express themselves, and interact with the outside world. For instance, when considering purchasing a new product, individuals often rely on Google reviews rather than solely relying on friends' advice. Presently, there exists a plethora of social networking sites including Facebook, Twitter, Google+, Flickr, LinkedIn, etc., with nearly every individual being a member of at least one OSN. These networks are experiencing exponential growth both in terms of user numbers and interconnections across different geographical locations. Despite their widespread popularity, OSNs pose significant security and privacy challenges such as spam, scams, phishing, clickjacking, harassment, stalking, defamation, identity theft, and third-party personal information disclosure. The concentration of users' personal, professional, social, and political data in one location attracts social spammers (cybercriminals) who can cause harm to both users and service providers. These cybercriminals employ various tactics including identity theft attacks, creation of fake profiles, and automated crawling across multiple social networking sites. Motivations behind creating fake profiles range from advertising, defamation, and social engineering to data collection for research or specialized marketing. OSNs generate vast amounts of user-generated content, making them lucrative targets for spammers. The primary objective of these cybercriminals is to exploit users by exposing them to unwanted information such as pornography or by deceiving them into divulging personal, professional, political, or financial information. Various methods, including the creation of fake profiles, are employed by adversaries to hack users' data, compromising their security and privacy. Consequently, users' personal, professional, and financial information is no longer secure in the online social networking environment.

In recent years, online social networks have witnessed exponential growth, offering individuals myriad opportunities to express themselves publicly, communicate

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 5, May 2024

with friends, and share information globally. A recent survey estimated that a significant percentage of adult internet users engage with online social network sites such as Twitter, LinkedIn, Google+, and Facebook. As of October, Facebook boasts over one billion monthly active users, with each user having an average of hundreds of friends and uploading billions of pictures. Additionally, American internet users spend billions of minutes on Facebook monthly, making it the leading web brand in the United States.

Due to the user-friendly nature of Facebook, users tend to disclose extensive personal details, including date of birth, personal pictures, workplace, email address, high school name, relationship status, and even phone numbers. Identity deception on big data platforms like social media is a growing concern due to the continuous expansion and evolution of these platforms. Social media has become a preferred means of communication, making it a prime target for spammers and scammers. Traditional cyber threats such as spamming now manifest in different forms on social media platforms, posing new challenges to users' security and privacy. In the contemporary era, online social networks have become integral to everyone's social life, revolutionizing the way people interact and communicate. These platforms have significant implications for various domains including science, education, grassroots organizing, employment, and business. Researchers are actively studying the impact of online social networks on individuals and society as a whole. Teachers leverage these platforms to facilitate learning, creating a friendly environment for students through online classroom pages, homework assignments, and discussions. Employers utilize social networking sites for talent recruitment and background checks, streamlining the hiring process. While most OSNs are free, some charge membership fees or rely on advertising for revenue generation. Governments can leverage these platforms to gauge public opinion swiftly and efficiently.

Popular social networking sites such as Facebook, Twitter, LinkedIn, Google+, MySpace, and others have witnessed exponential growth in both size and influence over the last decade. However, along with the benefits of enhanced communication and information sharing, OSNs also present numerous challenges including privacy issues, cyberbullying, social engineering, and online impersonation. Fake accounts, which do not represent real individuals or organizations, are a prevalent issue on platforms like Facebook. Despite efforts to combat them, the detection of fake accounts remains a challenging task for both social networking platforms and security researchers.

In conclusion, online social networks have revolutionized how people interact and communicate, offering unprecedented opportunities for connection and collaboration. However, with these opportunities come significant security and privacy challenges that must be addressed to ensure a safe and secure online environment for users.

II. LITERATURE REVIEW

[1] Aditya and Sachin Nandan provided a novel approach to identifying fake profiles on social media platforms using heterogeneous social media analysis coupled with deep learning techniques. In contrast to existing methods that primarily rely on singular types of data, such as textual content or network structure, our approach integrates diverse sources of information, including textual, visual, and behavioral attributes. By harnessing the power of deep learning models, we aim to enhance the accuracy and efficiency of fake-profile detection in heterogeneous social media environments.

[2]. Awais Vasin and Zain Khalid proposed an innovative approach to recognizing fake profiles on social media platforms through the application of big data analytics. The proliferation of fake profiles has become a significant concern, posing threats to online security, integrity, and user trust. Leveraging the vast amounts of data generated within social media ecosystems, our methodology harnesses big data analytics techniques to identify patterns, anomalies, and trends indicative of fraudulent behavior. Through the integration of machine learning algorithms and data mining methodologies, we aim to enhance the efficiency and accuracy of fake profile recognition, thereby safeguarding users and enhancing the credibility of online platforms.

[3] Eswara Venkata Sai Raja developed a method for detecting fake profiles on online social networks by leveraging logistic regression and the gradient descent algorithm. With the proliferation of social media platforms, the presence of fake profiles has become a significant concern, leading to issues such as misinformation, identity theft, and fraud. Our approach aims to address this challenge by analyzing various features associated with user profiles and interactions, training logistic regression models using gradient descent optimization to classify profiles as genuine or fake. Through experimentation and evaluation, we demonstrate the efficacy and reliability of our method in detecting fake profiles within online social networks.

[4] Susmitha Saha and Samuel Mathew have done a systematic review that explores the landscape of social media bot detection techniques using deep learning methods. With the pervasive presence of bots in social media ecosystems, understanding and effectively identifying automated accounts have become crucial for maintaining the authenticity and integrity of online interactions. Through a comprehensive examination of existing literature, this paper synthesizes key methodologies, datasets, challenges, and advancements in social media bot detection leveraging deep learning approaches. By analyzing and categorizing various techniques, we provide insights into the state-of-the-art methodologies and highlight avenues for future research and development in this rapidly evolving domain.

[5] Padmaveni and Aravindhar proposes an innovative approach to identifying fake profiles in online social networks using finite automata. With the proliferation of fake profiles posing significant challenges to the integrity and trustworthiness of online platforms, the need for effective detection methods is paramount. Leveraging the concept of finite automata, the proposed method aims to analyze patterns and behaviors associated with fake profiles, enabling

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 5, May 2024

efficient and accurate identification. Through a detailed examination of the methodology, experimental results, and implications, this paper sheds light on the potential of finite automata in enhancing the security and reliability of online social networks.

[6] Shruti Joshi, Himanshi Gupta and Nagariya, Neha Dhanotiya presents a comprehensive overview and survey of techniques for identifying fake profiles in online social networks. With the exponential growth of social media platforms, the presence of fake profiles has become a pervasive issue, threatening the authenticity and trustworthiness of online interactions. Through a systematic examination of existing literature and methodologies, this paper synthesizes key approaches, challenges, and advancements in fake profile detection. By categorizing and analyzing various techniques, datasets, evaluation metrics, and emerging trends, the paper provides insights into the state-of-the-art practices and directions for future research in the domain of fake profile identification.

[7] Srijan Kumar, Xikun Zhang, and Jure Leskovec reveals that the pervasive presence of online social networks has revolutionized the way individuals connect, communicate, and share information in virtual environments. However, alongside genuine users, these platforms host a multitude of fake profiles, created with malicious intent to deceive, manipulate, or exploit unsuspecting users. Detecting and mitigating the presence of fake profiles is crucial for preserving the authenticity and trustworthiness of online interactions. In this paper, a machine learning approach is proposed as a solution to identify and differentiate fake profiles from genuine ones within online social networks.

[8] Rodrigo, Raquel and Ronaldo provided a comprehensive survey of botnet detection techniques in computer networks. With the proliferation of botnets posing significant threats to network security, detecting and mitigating their presence has become a critical task. Through a systematic examination of existing literature, this survey categorizes and analyzes various botnet detection methods, including signature-based, anomaly-based, and behavior-based approaches. Additionally, the paper discusses key challenges, datasets, evaluation metrics, and emerging trends in botnet detection, offering insights into the state-of-the-art practices and directions for future research in the field.

[9] Yuanshun Yao and Bimal Viswanath examines automated crowd turfing attacks and defense mechanisms in online review systems. Crowd turfing involves the manipulation of online reviews through the use of fake accounts or automated bots, influencing public perception and consumer decisions. Through an in-depth analysis, the paper identifies various automated crowd turfing techniques, including account automation, content generation, and review injection. Additionally, the paper explores defense strategies, such as anomaly detection, reputation scoring, and behavioral analysis, aimed at mitigating the impact of crowd turfing attacks on online review platforms.

[10] Sarah Bugrara and Sen-ching S. Cheung presents a dynamic approach to detecting fake profiles on LinkedIn, a professional networking platform. With the increasing prevalence of fake profiles posing threats such as identity theft and fraudulent activities, the need for effective detection methods is paramount. Through a detailed analysis, the paper proposes a dynamic approach that leverages temporal and behavioral features to identify anomalies associated with fake profiles. Experimental results demonstrate the effectiveness of the proposed method in distinguishing between genuine and fake profiles on LinkedIn, offering insights into its potential applications in enhancing platform security and user trust.

[11] Mohammad Ali Bashiri, Mahdi Jalili and Mohammad Hossein Moattar proposes a framework for detecting fake profiles in social networks using similarity measures. With the proliferation of fake profiles undermining the authenticity and trustworthiness of online interactions, effective detection methods are essential for maintaining the integrity of social networking platforms. Through the utilization of similarity measures, the proposed framework aims to identify suspicious patterns and discrepancies in profile attributes and behaviors, enabling the differentiation between genuine and fake profiles. Experimental results demonstrate the efficacy of the framework in detecting fake profiles, offering insights into its potential applications in enhancing platform security and user trust.

[12] Al-Garadi and Mohammed Shuib addresses the issue of online social network deception by focusing on the detection and mitigation of fake profiles. With the increasing prevalence of fake profiles across various social media platforms, the integrity and credibility of online interactions are compromised, leading to risks such as identity theft, misinformation, and fraud. Through a detailed analysis, the paper explores techniques for detecting fake profiles, including machine learning algorithms, behavioral analysis, and social network analysis. Additionally, the paper discusses strategies for mitigating the impact of fake profiles, emphasizing the importance of collaboration between platform administrators, researchers, and users to combat online social network deception effectively.

[13] Long Wang, Bo Li, and Feng Hui Ren presents an approach for detecting fake profiles in social networks at the time of account creation. With the increasing prevalence of fake profiles across various social media platforms, the need for proactive detection methods is essential to prevent the proliferation of fraudulent accounts. Through a detailed analysis, the paper proposes techniques for assessing the authenticity of user profiles based on attributes such as registration information, activity patterns, and behavioral cues. Experimental results demonstrate the effectiveness of the proposed approach in identifying fake profiles during the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 5, May 2024

account creation process, offering insights into its potential applications in enhancing platform security and user trust.

[14] Shahim Ahmad proposed an article that delves into the strategies and methodologies employed in detecting fake profiles across various social media platforms. It examines a spectrum of techniques, including linguistic analysis, network-based approaches, and machine learning algorithms, offering an overview of the current landscape of fake profile detection in online social networks. The paper synthesizes findings from diverse studies to provide insights into the challenges, advancements, and future directions in the field of fake profile detection on social media platforms. [15] Vineetha Venugopal and Saravanan introduces a novel approach for detecting and verifying cloned profiles within online social networks. It employs MapReduce-based clustering and classification techniques to analyze similarities and patterns among profiles. By leveraging scalable data processing and classification algorithms, the method aims to differentiate authentic accounts from cloned ones, offering a robust framework for enhanced profile authentication in online social networks.

III. PROBLEM STATEMENT

Impersonation occurs when users create social media accounts mirroring legitimate ones, often with malicious intent. These imposters mimic popular individuals, brands, or companies, prevalent across major platforms like Instagram. While some impersonators may be harmless, others engage in deceitful practices such as boosting account popularity, disseminating untrustworthy content, or perpetrating brand abuse. Consequently, numerous lawsuits, including in the United States, have addressed criminal impersonation as a fraudulent act involving assuming false identities or misrepresenting affiliations. However, detecting such activities manually proves slow and cumbersome. Thus, there's a pressing need for automated detection techniques to assist social media companies in combating this issue effectively.

IV. SYSTEM DESIGN

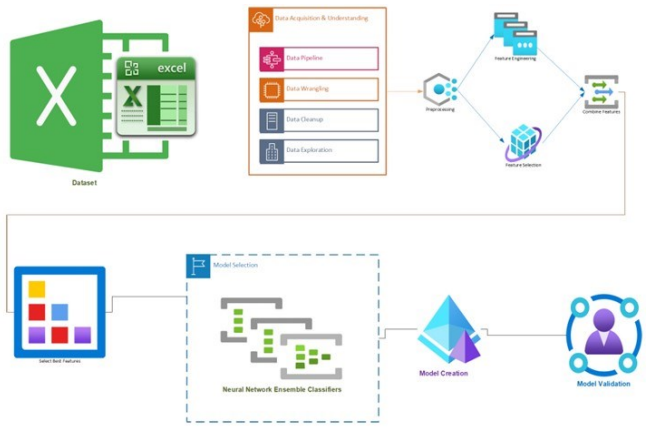

Fig 1. architecture diagram

A. Dataset Preprocessing

Initially, data analysis involves two key steps: data acquisition and data integration. During data acquisition, information is gathered from the real world, often containing incompleteness, noise, or redundancy. To enhance data quality, a process known as data cleansing is employed, aimed at eliminating redundancy and ensuring consistency. This includes addressing missing values and smoothing data through techniques like clustering, binning, and regression. Subsequently, data from diverse sources such as files and databases is integrated into a unified source. Following integration, preprocessing steps are implemented to reduce and transform the data without altering its fundamental characteristics. Preprocessing involves removing inconsistent and noisy data, a critical stage in preparing data for further analysis.

B. Feature Selection

Feature subsets in datasets are typically determined through filters, with the selection depending on the data size. Various learning algorithms such as random forest (RF) and Relief-F are then applied to these subsets to assess the data. RF aggregates the output of individual decision trees, each trained on a random subset of features, to produce the final output. This process involves filtering less optimal feature subsets to obtain the final result. In contrast, Relief-F computes the feature score for each feature and ranks them accordingly, filtering features based on their rank.

Choosing the appropriate subset based on consistency criteria can be challenging. To address this, methods like cross-validate filter (CRV), Ensembler filter (EF), and partitioning filter (PF) are employed as needed. CRV divides features into subsets and evaluates the performance of each subset. Poor-performing features are filtered out, leaving behind the best feature. PF, on the other hand, divides the dataset into chunks and selects the partition where the model performs the best.

C. Train the Model for fake User Profile Detection

Ensemble learning is a versatile machine learning approach aimed at enhancing predictive performance by aggregating predictions from multiple models. By combining various basic models, ensemble methods can produce a single optimized prediction model. Bagging, stacking, and boosting represent the fundamental types of Ensemble learning techniques, each offering unique advantages for predictive modeling projects. Additionally, training multiple instances of the same ensemble algorithm on diverse datasets is another effective ensemble strategy.

Data sampling from the training set can be achieved through bagging, which involves random sampling with replacement, or pasting, which entails sampling without replacement. Boosting, on the other hand, differs from traditional ensemble methods by sequentially training machine learning models, with each subsequent model addressing the weaknesses of its predecessor.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 5, May 2024

LSTM (Long Short-Term Memory) blocks are integral components of recurrent neural networks, featuring three or four gates to regulate information flow in and out of memory. These gates utilize the logistic function to compute values between 0 and 1, controlling the flow of information by selectively allowing or denying it. Specifically, input gates regulate new information entering memory, forget gates manage information retention, and output gates govern the utilization of memory content for output computation. LSTM blocks contain weights (W and U) that dictate gate operation, facilitating memory maintenance based on input values and previous time step outputs. Training LSTM blocks typically involves Backpropagation through time to optimize weights and minimize loss, allowing the LSTM model to learn and adapt over time.

V. RESULT

S.no	Algorithms	Accuracy
1.	Logistic regression	90%
2.	K-Nearest neighbour	85%
3.	Decision Tree	90%
4.	Random forest	92%
5.	Support Vector machine	52%
6.	XGBoost	92%
7.	Gradient Boosting Classifier	91%
8.	RF classifier	91%
9.	Extensible trees classifier	91%
10.	Adaboost classifier	90%
11.	SVC	79%

Fig 3: Table

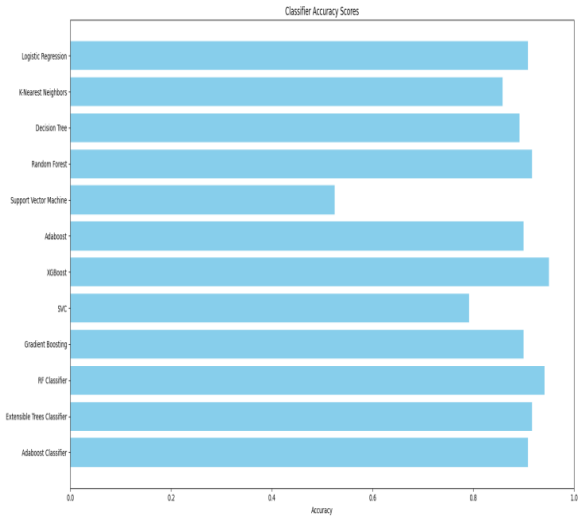

Fig 4: Accuracy table

VI. FUTURE ENHANCEMENT AND CONCLUSION

In forthcoming research endeavors, there will be a deliberate focus on enhancing the existing feature set utilized in this paper's investigation by incorporating insights from the realm of social sciences, particularly psychology. The primary objective will revolve around augmenting the corpus with novel features derived from analogous attributes utilized in this study, sourced from Social Media Platforms (SMPs). The overarching aspiration is that these newly engineered features will yield superior outcomes in the identification of identity deception prevalent on SMPs.

The primary contribution of this paper lies in its demonstration that the engineered features previously utilized for identifying fake accounts generated by bots do not exhibit similar efficacy in detecting fake accounts. Our contributions through this work are manifold. Firstly, we curated datasets specifically tailored for the detection of both fake and automated accounts, providing a foundational resource for future research in this domain. Secondly, we proposed novel derived features optimized for the classification of both fake and automated accounts, enriching the feature space and enhancing detection capabilities.

Moreover, we introduced a sophisticated cost-sensitive feature reduction technique leveraging genetic algorithms. This method facilitates the selection of the most relevant features for automated account classification, effectively optimizing performance while mitigating computational overhead. Additionally, we addressed the inherent imbalances within the fake account dataset by employing an algorithmic approach, ensuring more equitable representation and fostering more accurate model training.

Furthermore, we conducted comprehensive evaluations across the collected datasets, exploring various pattern recognition methodologies. Notably, ensemble and neural network-based approaches emerged as particularly promising, achieving the highest F-score for automated account detection. This thorough examination underscores the significance of employing diverse methodologies to effectively tackle the nuanced challenges posed by fake and automated account detection in online environments.

REFERENCES

[1] Aditya and Sachin nandan, “Heterogenous Social Media Analysis for Efficient Deep Learning Fake-Profile Identification” in the proceeding Of Institute Of Electrical And Electronics Engineers ,2023

[2] Awais Vasin and Zain Khalid “Fake profile recognition using big data analytics in social media platforms” in the proceeding of International Journal of Computer Applications in Technology, 2022

[3] Eswara Venkata Sai Raja “Fake Profile Detection Using Logistic Regression and Gradient Descent Algorithm on Online Social Networks” in the proceedings of EAI Endorsed Transactions on Scalable Information Systems, 2023

[4] Susmitha Saha and Samuel Mathew, “Social media bot detection with deep learning methods: a systematic review” in the proceedings of Big data and society, 2023

[5] Padmaveni and Aravindhar “Finite Automata for Fake Profile Identification in Online Social Networks” in the proceedings of International Conference Intelligent Computing and Control Systems, 2023

[6] Shruti Joshi, Himanshi Gupta and Nagariya, Neha Dhanotiya, “Identifying Fake Profile in Online Social Network: An Overview and Survey” in the proceedings of Conference on cybersecurity, 2020

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 5, May 2024

[7] Srijan Kumar, Xikun Zhang, and Jure Leskovec, "Fake Profile Detection in Online Social Networks: A Machine Learning Approach" in the proceedings of Conference on Knowledge Discovery and Data Mining (KDD), 2016
[8] Rodrigo, Raquel and Ronaldo, "Botnet detection: A survey" in the proceedings of Security engineering and application, 2022
[9] Yuanshun Yao and Bimal Viswanath, "Automated crowdturfing attacks and defenses in online review systems" in the proceedings of conference on computer and communications, 2023
[10] Sarah Bugrara and Sen-ching S. Cheung, "Detecting Fake Profiles on LinkedIn: A Dynamic Approach" int he proceedings of information science, 2023
[11] Mohammad Ali Bashiri and team, "A Review Article On Detection Of Fake Profile On Social-media" in the proceedings of International Journal of Innovative Research in Computer Science & Technology, 2023
[12] Al-garadi and Mohammed Shuib, "Online Social Network Deception: Detection and Mitigation of Online Social Network Fake Profiles" in the proceedings of International Conference on Cybercrime, Security and Digital Forensics (CyberForensics), May 2012
[13] Long Wang, Bo Li, and Fenghui Ren "Detecting Fake Profiles in Social Networks at Account Creation Time" in the processing of International Conference on Advances in Social Networks Analysis and Mining, 2023
[14] Shahim ahmad, "A Review Article On Detection Of Fake Profile On Social-media" in the proceedings of International Journal of Innovative Research in Computer Science & Technology, 2023
[15] Vineetha Venugopal and Saravanan A., "Detection And Verification Of Cloned Profiles In Online Social Networks Using Map reduce Based Clustering And Classification" in the proceedings of International Journal of intelligent Systems and Applications, 2020