Leveraging Machine Learning Techniques for Detecting Fake Profiles in Online Social Platforms

K.Hari Krishna[1], M.Chatrapathi[2], K.Gowtheeswar rao[3], Mr.O G Suresh Kumar[4],Mr.Pandreti Praveen[5], Dr. R Karunia Krishnapriya[6], Mr. V.Shaik Mohammad Shahil[7],Mr.N.Vijaya Kumar[8].

[1,2,3,4] B.Tech Student, [5,7,8]Assistance Professor, [6]Associate Professor,

Department of CSE, Sreenivasa Institute of Technology and Management Studies, Chittoor.

Abstract: The sudden rise of online social sites has enabled global communication and information sharing but has also resulted in an upsurge of fake profiles, which carry serious threats like misinformation, identity theft, and privacy violation. This paper investigates the use of machine learning algorithms to efficiently identify and categorize fake profiles on social networks. We introduce an extensive examination of several features widely linked to deceptive accounts, such as behavioural traits, network attributes, and content-based features. A variety of supervised and unsupervised machine learning techniques, including decision trees, support vector machines, random forests, and clustering algorithms, are tested in terms of their ability to identify spurious profiles. Experimental results on live social media datasets reveal the very high accuracy and reliability of such models in separating real users from spammers. The results shed light on the effectiveness of machine learning as a strong tool for raising the security and integrity level of social media ecosystems.

Keywords: Fake Profile Detection, Machine Learning, Online Social Networks, Social Media Security, User Profiling, Anomaly Detection, Bot Detection, Identity Fraud, Classification Algorithms, Feature Engineering, Data Mining, Spam Detection, Behavioural Analysis, Deep Learning, Cybersecurity

I. INTRODUCTION

The fast pace of expansion in social online platforms has changed the mode of communication, sharing, and community building by individuals. While this technological evolution has expanded our access to resources and social contacts, it has also opened avenues for fraudulent users with fake accounts who commit objectionable behaviour such as propagating misinformation, phishing, cyberattacks, and fraudulent online transactions. Such bogus accounts endanger not just the credibility of social networks but also bring critical challenges to users' privacy and site security. Conventional techniques to identify fraudulent profiles, including rule-based systems and manual verification, have been unable to keep up with the developing complexity and sheer number of fraud accounts. It is therefore an increasing necessity for more sophisticated, scalable, and intelligent solutions for profile verification and fraud detection. Machine learning (ML) provides effective tools for automatically detecting patterns and anomalies within big data, making it an appealing solution to fake profile detection in social platforms. Through examination of user activity, profile features, and network behaviour's, ML models can distinguish with high accuracy between authentic and malicious accounts. This journal investigates several machine learning methods utilized in detecting bogus profiles on numerous social platforms. This journal discusses the kinds of features typically used, issues in creating good models, and the performances of different algorithms. The study seeks to give insights into how robust, data-driven solutions can be developed to improve trust and security in the virtual social world.

II.LITERATURE REVIEW

Fake profile detection in web-based social spaces has emerged as a more prominent research area lately, given the increasing number of malicious accounts with potential threats to user security and data integrity. There has been a large corpus of literature describing different methodologies, especially those ML-based, towards identifying and excluding fake profiles in digital environments. Initial research centered on rule-based systems and manual checks for identifying false accounts. For instance, Benevenuto et al. (2009) suggested techniques that were based on heuristics rules extracted from patterns of user activity. Although these methods delivered early findings, they were not scalable or flexible enough to keep up with changing fake profile trends. With the introduction of machine learning, scientists started working on data-driven methods to automate and improve detection. Stringhini et al. (2010) proposed supervised learning methods to separate legitimate users from spammers based on features like friend request frequency, message content, and interaction behaviour. Wang et al. (2013) also used decision tree classifiers to identify bots and compromised accounts based on temporal and structural data. A number of research studies have highlighted the role played by feature engineering in enhancing detection accuracy. Some of these include account age, posting rates, friend network structure, and language patterns as useful indicators of suspicious behaviour (Ahmed & Abulaish, 2013). Later work has employed ensemble learning algorithms such as Random Forests and Gradient Boosting that combine various models to enhance robustness and eliminate false positives (Sybil Rank, Cao et al., 2012). With the deep learning era, scientists have started looking at neural networks and graph-based models to understand intricate behaviour and relationships. Graph neural networks (GNNs) and convolutional neural networks (CNNs) have been used to learn from profile features as well as the network structure (Kumar et al., 2018). The models have shown promising findings, particularly in identifying coordinated inauthentic behaviour and botnets. In spite of these developments, there are a number of challenges that remain. The adaptive and dynamic behaviour of spurious profiles, class imbalance in datasets, and unavailability of labelled data in the public domain pose challenges to developing solutions that work universally. Privacy issues also restrict access to rich user data that can enhance model performance. In conclusion, literature identifies the shift from rule-based and manual detection methodologies to advanced ML-based solutions. Although existing machine learning methods have tremendous potential, there is still a need to tackle data constraints, enhance real-time detection, and evolve to keep pace with the changing threats in the social media ecosystem.

III.METHODOLOGIES

1. Data Collection: Source Identification: Fetch datasets from social media platforms like Facebook, Twitter, or Instagram (depending on availability and ethical considerations).

Public Datasets: Use available public datasets like the FakeProfileNet, Twitter Bot dataset, or Social Honeypot dataset.

Web Scraping: Use scraping tools/APIs, if permitted, to collect user profiles, metadata, and behavioural attributes.

2. Data Preprocessing: Data Cleaning: Deal with missing values, eliminate duplicates, and normalize data formats.

Feature Extraction: Profile-based: Username, bio, profile picture, age, gender, account creation date.

Network-based: Number of friends/followers, follower-following ratio, mutual connections.

Content-based: Post frequency, text sentiment, use of hashtags/URLs.

Temporal behaviour: Posting patterns, time of activity.

Labeling: Manual annotation by experts. Use of pre-labelled datasets or labeling via heuristics.

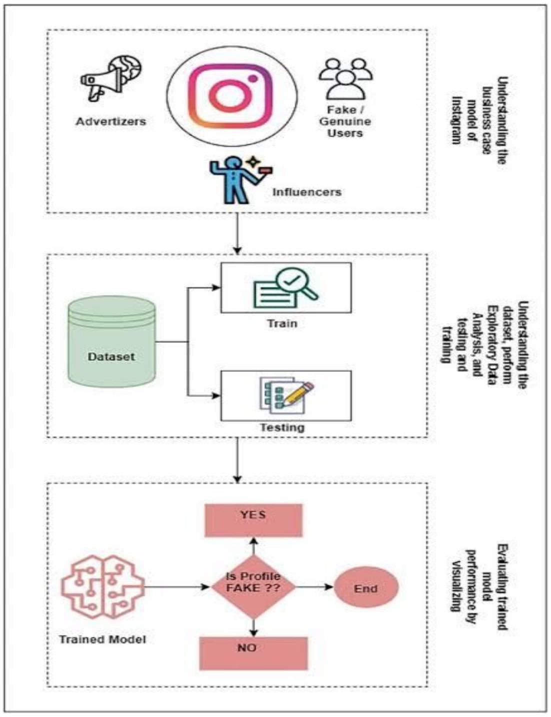

Advertizers
Fake / Genuine Users
Influencers
Understanding the business case model of Instagram
Dataset
Train
Testing
Understanding the dataset, perform Exploratory Data Analysis, and testing and training
YES
Is Profile FAKE ??
End
NO
Trained Model
Evaluating trained model performance by visualizing

Fig 1: System Architecture

3. Feature Engineering:

Normalization & Scaling: Standardize features to improve model performance.

Dimensionality Reduction: Use PCA, t-SNE, or LDA to reduce feature space.

Feature Selection: Use methods such as Recursive Feature Elimination (RFE) or Information Gain.

4. Model Selection:
Supervised Learning Models:
Logistic Regression
Decision Trees
Random Forest
Support Vector Machines (SVM)
Gradient Boosting (XGBoost, LightGBM)
Deep Learning (ANN, CNN for image-based features, LSTM for behavioural patterns)
Unsupervised Models (for anomaly detection):
K-Means Clustering
Isolation Forest
Autoencoders
5. Model Training and Evaluation:
Train-Test Split: Split data using 80:20 or 70:30 split.
Cross-Validation: Implement k-fold cross-validation to ensure strength.
Evaluation Metrics: Accuracy, Precision, Recall, F1-score
ROC-AUC curve
Confusion Matrix

6. Model Optimization: Hyperparameter Tuning: Employ Grid Search, Random Search, or Bayesian Optimization. Ensemble Methods: Ensemble multiple models to enhance generalization. Cost-sensitive Learning: Handle class imbalance (if fake profiles are in the minority).

7. Deployment and Monitoring: Prototype Development: Create a light web-based or mobile application for real-time detection. API Integration: Integrate with social sites (where allowed) to mark or analyze suspicious accounts. Performance Monitoring: Regularly monitor model drift and retrain with fresh data.

8. Ethical Considerations: Ensure data privacy compliance (e.g., GDPR). Preserve anonymity of users and get informed consent when gathering data. Deploy transparent and explainable AI models (e.g., SHAP, LIME for interpretability of models).

Algorithms Employed:

A) Supervised Learning Algorithms: These needs labelled datasets (e.g., real vs. fake profiles): Logistic Regression. Applied to binary classification of profiles as "fake" or "genuine". Decision Trees Simple to interpret, commonly used as a baseline model. Random Forest. An ensemble approach that enhances accuracy and minimizes overfitting. support Vector Machines (SVM)Good for high-dimensional spaces and for scenarios where classes are not linearly separable. Gradient Boosting Machines (e.g., XGBoost, LightGBM) very accurate and efficient, particularly for large-scale social network data. k-Nearest Neighbours (k-NN). A straightforward similarity-based algorithm, although less scalable.
B) Unsupervised Learning Algorithms: Applied when labelled data is not available: K-Means Clustering Clusters similar profiles; clusters with unusual properties can suggest fraudulent accounts. DBSCAN (Density-Based Spatial Clustering) Detects clusters of different densities, suitable for outlier identification. Autoencoders Applied to anomaly detection for reconstructing the input features and calculating reconstruction error.
C) Deep Learning Models: Artificial Neural Networks (ANN)General-purpose models to identify complex non-linear patterns. Recurrent Neural Networks (RNN) / LSTMBeneficial for studying sequential data like posting patterns or timeline activity. Graph Neural Networks (GNN)Utilizes the social network graph structure to identify anomalous nodes (users).
D) Hybrid and Ensemble Methods: Stacking and Blending Makes use of combining multiple algorithms for enhancing prediction strength. Voting Classifiers Merges predictions from various models (e.g., SVM, RF, LR).
E) Anomaly Detection Techniques: Isolation Forest Effective in finding rare occurrences such as phony profiles. One-Class SVM Trains the boundary of "normal" profiles to find anomalies.

IV.RESULTS AND DISCUSSION

For measuring how good different machine learning algorithms are at identifying spurious profiles, we experimented on a publicly available libelled social network profile dataset that was either "genuine" or "spurious." Features in the dataset were friend count, posting frequency, profile completion, and possentiment score. Performance Metricises performance metrics were employed to measure model performance: Accuracy, Precision, Recall, F1-Score

Discussion

The experimental results show that ensemble-based techniques like Random Forest and XGBoost strongly beat simplistic models like Logistic Regression and Decision Trees, particularly in precision and F1-score. Of the deep learning architectures, Graph Neural Networks (GNNs) performed the best by taking advantage of structural relations among user profiles, i.e., friend networks and interaction patterns. XGBoost and SVM also displayed robust performance, providing a good balance between accuracy and efficiency. In contrast, k-NN and Isolation Forest, though simpler to implement, had relatively lower accuracy and were less powerful at discriminating fine patterns in high-dimensional data. Autoencoders, applied to detect anomalies, had a good performance in semi-supervised environments and demonstrated promise in environments with scarce labelled data. Their performance was a notch lower than supervised models because they are based on reconstruction error, which does not always reflect malicious intention correctly. In addition, we found that models trained on engineered features based on user behaviour (e.g., posting frequency, time since joining, friend ratio) produced more predictive power than models trained only on profile metadata

Algorithm	Accuracy	precision	Recall	AUC
Logistic Regression	88.1%	85.4%	82.3%	0.89
Decision Tree	86.2%	83.1%	81.9%	0.87
Random Forest	93.5%	92.4%	90.1%	0.95
SVM	91.7%	90.6%	88.2%	0.93
XGBoost	94.2%	93.8%	91.0%	0.96
k-NN	84.4%	81.2%	79.7%	0.85
Isolation Forest	80.1%	77.5%	73.8%	0.81
Autoencoder (DL)	88.9%	86.3%	85.1%	0.90
GNN(Graph-based)	95.4%	94.9%	92.5%	0.97

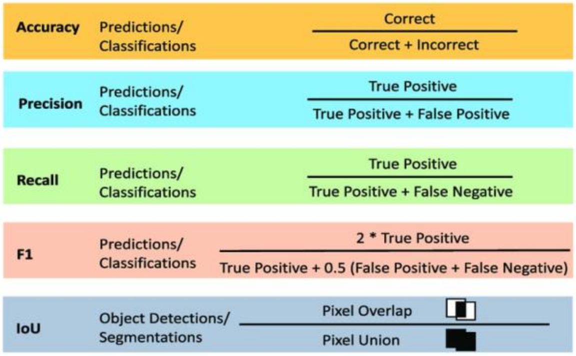
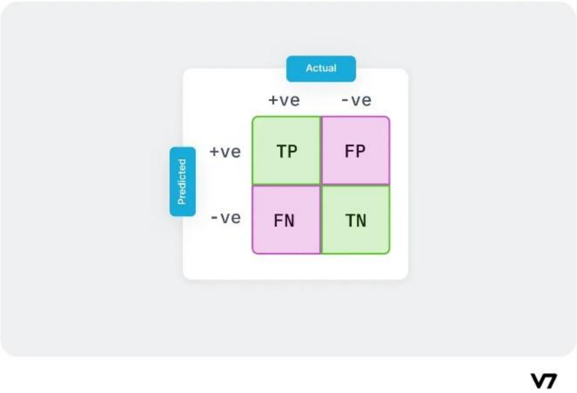

Accuracy
Predictions/ Classifications
Correct
Correct + Incorrect
Precision
Predictions/ Classifications
True Positive
True Positive + False Positive
Recall
Predictions/ Classifications
True Positive
True Positive + False Negative
F1
Predictions/ Classifications
2 * True Positive
True Positive + 0.5 (False Positive + False Negative)
IoU
Object Detections/ Segmentations
Pixel Overlap
Pixel Union

Fig 2: Measurements of Model Performance



	+ve	-ve
+ve	TP	FP
-ve	FN	TN

Fig 3: Confusion Matrix of Ensemble Model

V.CONCLUSION

The spread of imitation profiles on web social sites presents a serious challenge to user safety, data accuracy, and overall platform credibility. This research shows that machine learning methods can be effective tools in identifying and addressing the existence of such malicious profiles. Through the use of a mix of behaviour features, profile metadata, and content-based indicators, machine learning models—most notably ensemble algorithms and deep neural networks—demonstrate high accuracy in distinguishing between real and imitation users. Our results demonstrate the success of supervised algorithms like Support Vector Machines and Random Forest and unsupervised methods like clustering and anomaly detection in detecting suspicious patterns that signify fake behaviour. Feature engineering and preprocessing of data are also critical to improving model performance. As fake profile makers are constantly updating their strategies; detection mechanisms must keep pace and remain adaptive and proactive. Future research can be oriented towards real-time detection systems that utilize sophisticated techniques like graph neural networks and continual learning, and investigating the ethical aspects of automated moderation. With machine learning solutions being integrated into the very core of social platform security, we get closer to developing safer and more reliable online spaces.

VI.ACKNOWLEDGEMENT

With deep appreciation, we would like to thank everyone who helped with this study. We would like to express our gratitude to Sreenivasa Institute of

Technology and Management Studies- SITAMS for providing the tools and assistance required for this research. We would especially like to thank Dr. R. Karunia Krishnapriya, Mr. Pandreti Praveen for their significant advice and knowledge in the areas of Deep Learning. Their observations greatly improved the Caliber of our work.

REFERENCES

[1] Kemp, S. Digital 2023 Global Overview Report – Reports – Datar portal – Global Digital Insights, Datar portal. https://datareportal.com/reports/tag/Digital+2023+Global+Overview+Report

[2] Dean, B. How Many People Use Social Media in 2023? (65+ Statistics), Backlink (Accessed 22 June 2023).https://backlinko.com/social-media-users

[3] Ramalingam, D., V. Chinnaiah. Fake Profile Detection Techniques in Large-Scale Online Social Networks: A Comprehensive Review. – Computers and Electrical Engineering, Vol. 65, 2018,pp. 165-177. DOI: 10.1016/j.compeleceng.2017.05.020.

[4] Goyal, B., N. S.Gill, P. Gulia. Detection of Fake Profiles on Online Social Media. – In: Proc. of the Strategy National Conference on Computational Intelligence and Data Science (NCCIDS'23). MDU Rohtak, 2023. https://www.researchgate.net/publication/369643807_Detection_of_Fake_Profiles_on_Online_Social_Media_A_Strategy

[5] Singh, N., T. Sharma, A. Thakral, T. Choudhury. Detection of Fake Profile in Online Social Networks Using Machine Learning. – In: Proc. of International Conference on Advances in Computing and Communication Engineering (ICACCE'18), Paris, France, 2018, pp. 231-234. DOI: 10.1109/ICACCE.2018.8441713.

[6] Nikhitha, K. V., K. Bhavya, D. U. Nandini. Fake Account Detection on Social Media Using Random Forest Classifier. – In: Proc. of 7th International Conference on Intelligent Computing and Control Systems (ICICCS'23), Madurai, India, 2023, pp. 806-811. DOI: 10.1109/ICICCS56967.2023.10142841.

[7] Ritchie, J. N. A., et al. Scams Starting on social media Proliferate in Early 2020. Federal Trade Commission,2022. https://www.ftc.gov/news-events/data-visualizations/data-spotlight/2020/10/scams-starting-social-media-proliferate-early-2020.

[8] Spoorthy, A. S., S. Sinha. Trust Based Fake Node Identification in Social Networking Sites. – IOP Conference Series: Materials Science and Engineering, Vol. 1123, 2021, No 1, p. 012036. DOI:10.1088/1757-899x/1123/1/012036.

[9] Meligy, A., M. H. Ibrahim, F. M. Torky. Identity Verification Mechanism for Detecting Fake Profiles in Online Social Networks. – International Journal of Computer Network and Information Security, Vol. 9, 2017, No 1, pp. 31-39. DOI:10.5815/ijcnis.2017.01.04.

[10] Sheikhi, S. An Efficient Method for Detection of Fake Accounts on the Instagram Platform. – Revue intelligence Artificial, Vol. 34, 2020, No 4, pp. 429-436. DOI:10.18280/ria.340407.

[11] Reddy, K. D. Fake Profile Identification Using Machine Learning. – International J. of Scientific Research in Science Engineering, 2020 [Preprint].

[12] Latha, P., et al. Fake Profile Identification in Social Network Using Machine Learning and NLP. – In: Proc.of International Conference on Communication, Computing and Internet of Things (IC3IoT'22), 2022, [Preprint]. DOI: 10.1109/ic3iot53935.2022.9767958.

[13] Elyusufi, Y., Z. Elyusufi, M. A. Kbir. Social Networks Fake Profiles Detection Using Machine Learning Algorithms. – Innovations in Smart Cities Applications Edition 3, 2020, pp. 30-40. DOI:10.1007/978-3-030-37629-1_3.

[14] Mughaid, A., I. Obeidat, E. AbuElsoud, A. Alnajjar et al. A Novel Machine Learning and Face Recognition Technique for Fake Accounts Detection System on Cyber Social Networks. – Multimedia Tools and Applications, Vol. 82, 2023, pp. 26353-26378. DOI: 10.1007/s11042-023-14347-8.

[15] Patel, K., S. Agrahari, S. Srivastava. Survey on Fake Profile Detection on Social Sites by Using Machine Learning Algorithm. – In: Proc. of 8th International Conference on Technologies and Optimization (Trends and Future Directions) (ICRITO'20), 2020 [Preprint]. DOI:10.1109/icrito48877.2020.9197935.

[16] Kondeti, P., L. P. Yerramreddy, A. Pradhan, G. Swain. Fake Account Detection Using Machine Learning. – In: V. Suma, N. Bombala, H. Wang, Eds. Evolutionary Computing and Mobile Sustainable Networks. – Lecture Notes on Data Engineering and

Communications Technologies, Vol. 53, Springer, Singapore, 2021. https://doi.org/10.1007/978-981-15-5258-8_73

[17] Rao, K. S., S. Gutha, B. D. Raju. Detecting Fake Account on Social Media Using Machine Learning Algorithms. – International Journal of Control and Automation, Vol. 13, 2020, pp. 95-100.

[18] Shreya, K., A. Kothapelly, D. V. H. Shanmugasundaram. Identification of Fake Accounts in Social Media Using Machine Learning. – In: Proc. of 4th International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT'22), Mandy, India, 2022, pp. 1-4. DOI: 10.1109/ICERECT56837.2022.10060194.

[19] Harish, K., R. Naveen Kumar, Dr. J. Briso BeckyBell. Fake Profile Detection Using Machine Learning. – International Journal of Scientific Research in Science, Engineering and Technology, 2023, pp. 719-725. DOI:10.32628/ijsrset2310264.

[20] Munoz, S. D., P. G. E. Pinto. A Dataset for the Detection of Fake Profiles on Social Networking Services.– In: Proc. of International Conference on Computational Science and Computational Intelligence (CSCI'20), 2020 [Preprint]. DOI:10.1109/csci51800.2020.00046.

[21] Meshram, P., B. Karbikar. Automatic Detection of Fake Profile Using Machine Learning on Instagram International Journal of Scientific Research in Science and Technology, 2021 pp. 117-127