Investigating Techniques for Detecting Fake Social Media Accounts

Perla Shiva Sindhu
MTECH. Integrated CSE with Spec. in Business Analytics
VIT Chennai
shivasindhu.perla2020@vitstudent.ac.in

Yashwanth Kumar A
MTECH. Integrated CSE with Spec. in Business Analytics
VIT Chennai
yashwanthkumara.2020@vitstudent.ac.in

Dr. Sujithra Kanmani R
Assistant Professor
School of Computer Science and Engineering
VIT Chennai
sujithrakanmani.r@vit.ac.in

ABSTRACT— The rise of fake social media accounts presents a serious challenge for maintaining trust in online platforms. To tackle this issue, researchers have turned to machine learning (ML) and deep learning (DL) models as promising solutions. Despite their effectiveness, these models face hurdles in real-life scenarios due to clever impersonation techniques and evolving tactics used by malicious actors. This paper offers an overview of the current landscape of detecting fake social media accounts using ML and DL models, discussing both their strengths and limitations. It delves into the practical challenges encountered, such as sophisticated spoofing methods and adaptive strategies employed by attackers. Additionally, the paper explores various detection strategies like anomaly detection, behavioural analysis, and content-based approaches. Finally, it suggests potential areas for future research aimed at improving the reliability and effectiveness of detection methods to combat the proliferation of fake social media accounts.

Keywords— Fake social media accounts, ML, DL, Detection challenges, Impersonation, Anomaly detection, Behavioural analysis, Content-based approaches, Research, Trust, Reliability.

I. INTRODUCTION

The prevalence of fake social media accounts has become a pressing concern, threatening the trustworthiness of online platforms. In response, researchers are turning to machine learning (ML) and deep learning (DL) techniques to detect these accounts. Despite their promise, these methods face challenges in real-world situations due to clever tactics used by impostors. This paper offers an overview of how ML and DL models are being used to tackle fake social media accounts, discussing both their strengths and the practical hurdles they encounter.

One major challenge is the sophisticated techniques employed by malicious actors, making it difficult for traditional detection methods to keep up. ML and DL models offer a promising solution by analysing patterns and behaviours to identify fake accounts. However, as attackers continuously evolve their strategies, researchers must adapt their detection methods accordingly. This paper delves into these challenges and explores the strategies researchers are using to overcome them. Various detection approaches are being explored, including anomaly detection, which flags unusual behaviour, and behavioural analysis, which examines patterns in user interactions. Content-based approaches analyse the substance of posts to identify inconsistencies or suspicious activity. Despite progress, there is still much to be done to improve the reliability and effectiveness of these methods. This paper concludes by highlighting the need for further research to enhance detection mechanisms and combat the proliferation of fake social media accounts.

Our research hypothesizes that by integrating machine learning algorithms with network analysis techniques, we can significantly enhance the detection accuracy of fake social media accounts. We posit that analysing user behaviour patterns, network structures, and content dissemination dynamics can reveal distinct markers indicative of fraudulent accounts. Furthermore, we anticipate that by employing sophisticated algorithms capable of processing vast amounts of data, we can uncover subtle nuances and anomalies characteristic of fake accounts that may elude traditional detection methods. Through empirical experimentation and evaluation, we aim to validate our hypothesis and demonstrate the efficacy of our proposed approach in detecting fake social media accounts across diverse

platforms and scenarios. Additionally, we seek to uncover insights into the underlying mechanisms driving the proliferation of fake accounts and contribute to the development of robust strategies for combating this pervasive threat in the digital realm.

II. OBJECTIVE

This study aims to explore, develop, and evaluate methodologies for detecting fake social media accounts, leveraging a combination of machine learning, network analysis, and data mining techniques. Firstly, we intend to conduct a comprehensive review of existing methodologies and tools employed in the detection of fake social media accounts. This review will provide insights into the state-of-the-art techniques and highlight gaps in current approaches.

Subsequently, we seek to gain a deeper understanding of the characteristics and behaviours exhibited by fake social media accounts. This involves analysing patterns of activity, content dissemination strategies, and network interactions associated with fraudulent accounts. By identifying key features and indicators of fake accounts, we aim to inform the development of more effective detection algorithms.

III. RELATED WORK

[1] The paper "Fake Account Detection on Social Media using Random Forest Classifier" delves into the realm of identifying fake accounts on social media platforms through the utilization of machine learning algorithms, specifically random forest, logistic regression, and decision tree. Through a comparative analysis, it is revealed that the random forest classifier outperforms the other algorithms in terms of accuracy for detecting these fraudulent accounts. Emphasizing the detrimental impact of fake accounts, such as the dissemination of false information and malicious content, the study underscores the significance of effective detection methods.

[2] The paper "Using Machine Learning to Detect Fake Identities: Bots vs Humans" addresses the research gap concerning the detection of fake human identities on social media platforms (SMPs), contrasting with existing work focused primarily on identifying bot-generated accounts. It underscores the reliance of previous machine learning models on engineered features, such as the friend-to-followers ratio, derived from attributes like friend and follower counts available in account profiles. This study innovatively applies these engineered features to a dataset comprising fake human accounts, aiming to enhance the detection of such deceptive identities on SMPs.

[3] The paper "Fake news detection in social media based on sentiment analysis using classifier techniques" proposes a novel fake news detection method integrating sentiment analysis, multiple imputation, and feature extraction techniques. Achieving a remarkable accuracy of 99.8%, it outperforms existing approaches. Leveraging datasets ISOT and LIAR, it employs lexicon-based scoring for sentiment analysis and Multiple Imputation Chain Equation (MICE) for handling missing data. Term Frequency-Inverse Document Frequency (TF-IDF) aids in extracting effective features from text, subsequently classified using Naïve Bayes, passive-aggressive, and Deep Neural Network (DNN) classifiers. This research highlights the urgent need to combat fake news and mitigate its widespread impact.

[4] The Research paper "Detecting Fake Accounts in Online Social Networks at the Time of Registrations" introduces Ianus, a method for detecting Sybil accounts based on registration information, but its effectiveness may be compromised if attackers mimic benign user registration patterns. Evaluation using real-world WeChat datasets limits generalizability to other social networks, and lacking information on false positive rates and computational complexity are notable drawbacks. While Ianus shows promise in initial Sybil detection, further research is needed to address these limitations and validate its efficacy across diverse online social network platforms.

[5] The paper "Detecting Fake Accounts on Social Media" addresses the growing concern of fake accounts and bots on online social networks (OSNs), highlighting their negative impacts such as data theft and spreading false information. It introduces SVM-NN, a new algorithm designed to efficiently detect fake Twitter accounts and bots. SVM-NN utilizes feature selection and dimension reduction techniques and incorporates machine learning classifiers like Support Vector Machine (SVM) and Neural Network (NN). Results indicate SVM-NN's effectiveness, achieving a classification accuracy of around 98% on the training dataset while using fewer features compared to existing methods.

[6] The Research paper "Detection of Fake Accounts in Social Networks Based on One Class Classification" proposes a method for detecting fake social media accounts by utilizing users' similarities and network communications. It calculates similarity measures such as common neighbours and cosine coefficients from the social network graph's adjacency matrix. Principal Component Analysis (PCA) is applied to extract informative features and reduce data complexity. These features are then used to train a One Class Classification (OCC) algorithm, achieving a detection accuracy of 99.6% with a false negative rate of 0%. The study concludes that incorporating similarity measures and OCC algorithms enhances detection efficiency compared to multi-class algorithms. Experimental results

demonstrate the method's effectiveness in accurately identifying fake accounts on social networks.

[7] “A Review Article on Detection of Fake Profile on Social-Media” The widespread use of online social networking sites (OSNs) has revolutionized communication but also brought about challenges such as the proliferation of fraudulent profiles. These fake accounts are utilized for various malicious purposes, including spreading rumours and committing cybercrimes. Detecting fake profiles has become a critical concern, addressed through machine learning techniques. By analysing profile data and user behaviour, machine learning models can classify profiles as genuine or fake based on features like profile picture, name, activity, and engagement. Various algorithms like decision trees and neural networks are employed for classification, although model performance depends on dataset quality and feature selection. Continuous updates are necessary to combat evolving tactics used by fake profile creators.

[7] The paper “Fake Twitter Followers Detection using Machine Learning Approach” addresses the prevalence of fake user accounts on Twitter, often operated by automated bots for spamming and manipulating trends. Researchers propose a machine learning model leveraging publicly available Twitter user data, including activity patterns and profile information, to assess account authenticity. Utilizing algorithms such as logistic regression, long short-term memory, K-means, and random forest, the model evaluates account authenticity. Experimental findings demonstrate that the random forest algorithm yields the highest accuracy (0.7557), precision (0.7277), and F1 score (0.7943) compared to other algorithms tested.

[8] The paper “Optimized Feed Forward Neural Network for Fake and Clone Account Detection in Online Social Networks” addresses security challenges in Online Social Networks (OSNs), including fake accounts and clone attacks, which pose privacy threats to users. To combat these issues, the paper proposes a machine learning-based technique. It introduces an optimized feed forward neural network for detecting fake accounts, utilizing preprocessing techniques like removal of missing values and outlier data through K-means clustering. Additionally, principal component analysis and genetic algorithm are employed for feature extraction and selection, respectively, to optimize the neural network's objective function and reduce misclassification rates. Evaluation on a Twitter dataset demonstrates the proposed approach's effectiveness, comparing favourably with existing techniques in terms of accuracy, precision, and recall.

IV. METHODOLOGY

A. Existing Systems:

Existing systems for detecting fake social media accounts typically rely on traditional machine learning algorithms and heuristic-based approaches. These systems often utilize features such as account age, posting frequency, and engagement patterns to identify suspicious accounts. However, they may suffer from limited scalability and effectiveness in handling sophisticated fake account creation techniques. Some systems also incorporate manual verification processes or crowd-sourced labelling, which can be time-consuming and prone to human error. Overall, existing systems lack the robustness and efficiency required to combat the growing threat of fake accounts on social media platforms.

B. Proposed Systems:

In the proposed methodology, we aim to develop an advanced system for detecting fake social media accounts. The methodology involves data collection from various social media platforms, including attributes such as follower counts, posting behaviour, and engagement metrics. Next, we preprocess and clean the collected data to ensure its quality and consistency. Feature engineering is then performed to extract meaningful features from the data, which can help distinguish between genuine and fake accounts. Subsequently, we employ machine learning and deep learning algorithms for classification tasks, training the models on labelled data and optimizing their performance through techniques like hyperparameter tuning and cross-validation. The final step involves deploying the trained model to classify new social media accounts in real-time, continuously monitoring its performance and updating it to adapt to evolving strategies used by fake account creators.

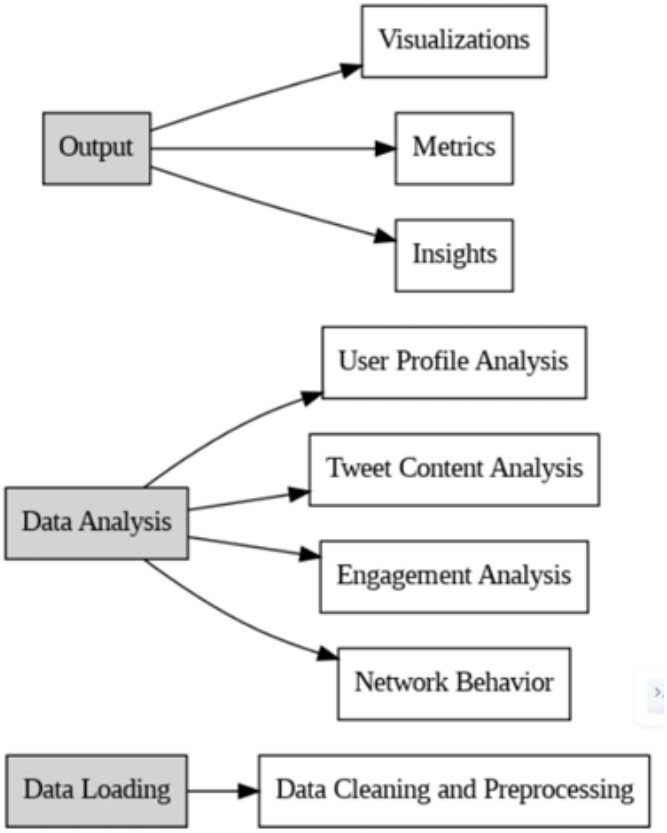

Fig 1.Architecture Diagram

C. Problem Definition:

The problem of detecting fake social media accounts is multifaceted, involving the identification of accounts created with malicious intent to deceive or manipulate users. Fake accounts can be used for various purposes, including spreading misinformation, engaging in fraudulent activities, or inflating follower counts. The primary challenge lies in distinguishing between genuine and fake accounts, as fake accounts often mimic the behaviour of real users to evade detection. Key issues include the development of effective feature extraction techniques, the selection of appropriate classification algorithms, and the integration of real-time monitoring capabilities. Additionally, the problem encompasses the need for continuous adaptation to emerging tactics employed by malicious actors, highlighting the dynamic nature of the fake account detection problem.

V. MAIN IDEA

The endeavor to counter the widespread proliferation of fake social media accounts hinges on harnessing the power of machine learning and deep learning models. This multifaceted approach delves into the intricacies of evolving impersonation techniques and the sophisticated strategies employed by malicious entities. By meticulously addressing these challenges, the primary goal is to cultivate robust detection mechanisms capable of discerning between genuine and fraudulent accounts.

The pursuit of this objective involves an extensive exploration of diverse methodologies, each offering unique insights and capabilities. Anomaly detection techniques scrutinize deviations from established behavioral norms, flagging activities that diverge significantly from authentic user patterns. This proactive approach allows for the early identification of suspicious accounts, minimizing their potential impact on the online ecosystem.

Furthermore, behavioral analysis plays a pivotal role in augmenting the credibility and integrity of online platforms. By analyzing nuanced behavioral cues and interaction patterns, machine learning algorithms can discern subtle indicators of authenticity or deceit. This granular understanding enables platforms to swiftly isolate and neutralize fraudulent accounts, safeguarding user trust and fostering a more secure digital environment.

VI. DATA PRE-PROCESSING

The provided code begins by loading two datasets: a training set and a test set, containing information about Twitter users. The training data consists of 5500 entries with 23 columns, where 'Choice' represents the target variable indicating whether a user account is genuine or fake. The data preprocessing step involves scaling the features using Standard Scaler and removing any duplicate entries from the dataset. After preprocessing, the features are split into the training and testing sets using an 80-20 ratio, with 4400 entries for training and 1100 for testing.

Scaling the features ensures that each feature contributes equally to the model training process, preventing any bias caused by features with larger scales. The train-test split allows for the evaluation of the machine learning model's performance on unseen data. By leveraging these preprocessing techniques and data splitting strategies, the subsequent machine learning model can be trained and evaluated effectively for the task of detecting fake Twitter, Instagram and accounts.

VII. PERFORMANCE ANALYSIS

A. RMSE:

RMSE quantifies how much the model's predictions deviate, on average, from the true values. A lower RMSE indicates better agreement between predicted and actual values, implying higher accuracy of the model's predictions.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$$

Each model is associated with its respective RMSE value, which serves as an evaluation metric indicating the average deviation of the predicted values from the actual values. The bar plot visualizes the RMSE values for both ML and DL models, enabling a comparison of their performance in terms of prediction accuracy.

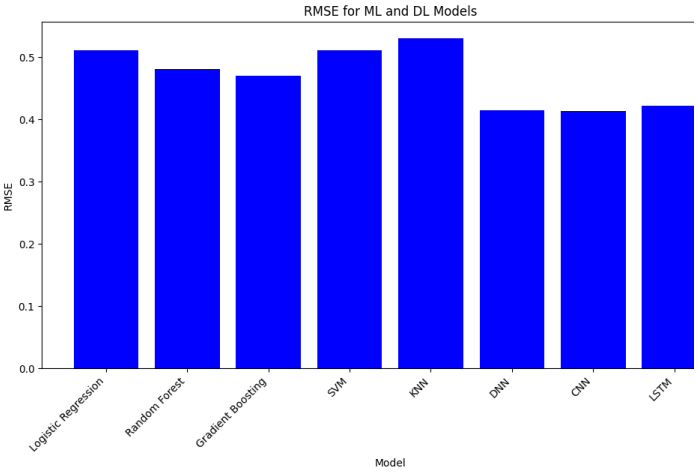

Fig 2. RMSE for ML and DL models on Twitter

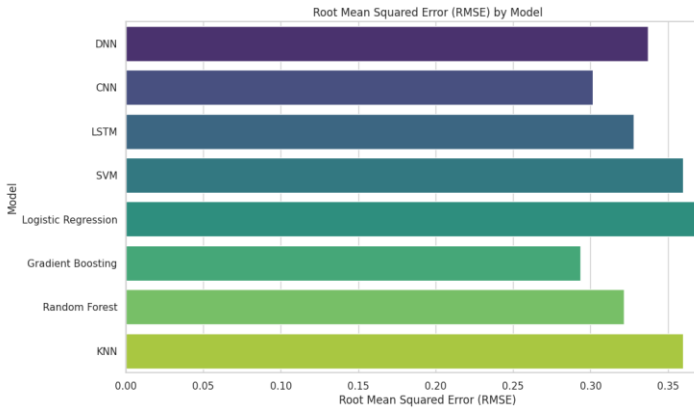

Fig 3. RMSE for ML and DL models on Instagram

B. MAE:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|$$

MAE represents the average absolute deviation of the model's predictions from the true values. Like RMSE, a lower MAE indicates better agreement between predicted and actual values, suggesting higher accuracy of the model's predictions.

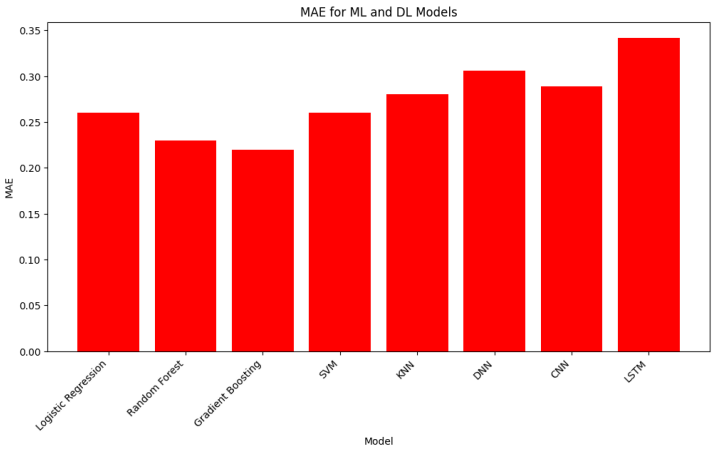

Fig 4.MAE for ML and DL models on Twitter

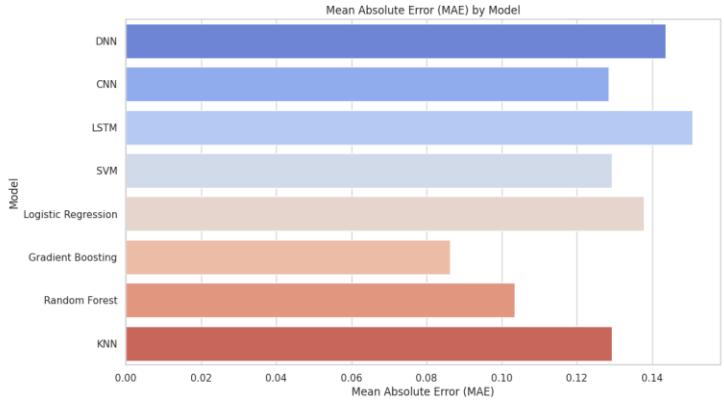

Fig 5. MAE for ML and DL models on Instagram

C. Model Metrics

i) Precision:

Precision measures the accuracy of positive predictions made by the model.

Precision = TP / (TP + FP)

ii) Recall:

Recall, also known as sensitivity or true positive rate, measures the ability of the model to correctly identify all positive instances in the dataset.

Recall = TP / (TP + FN)

iii) F1 Score:

F1 Score is the harmonic mean of Precision and Recall, providing a balance between the two metrics.

F1 Score = 2 * (Precision * Recall) / (Precision + Recall)

“F1 Score ranges from 0 to 1”

Fig 6. Model Precision, Recall and F1 score on Twitter

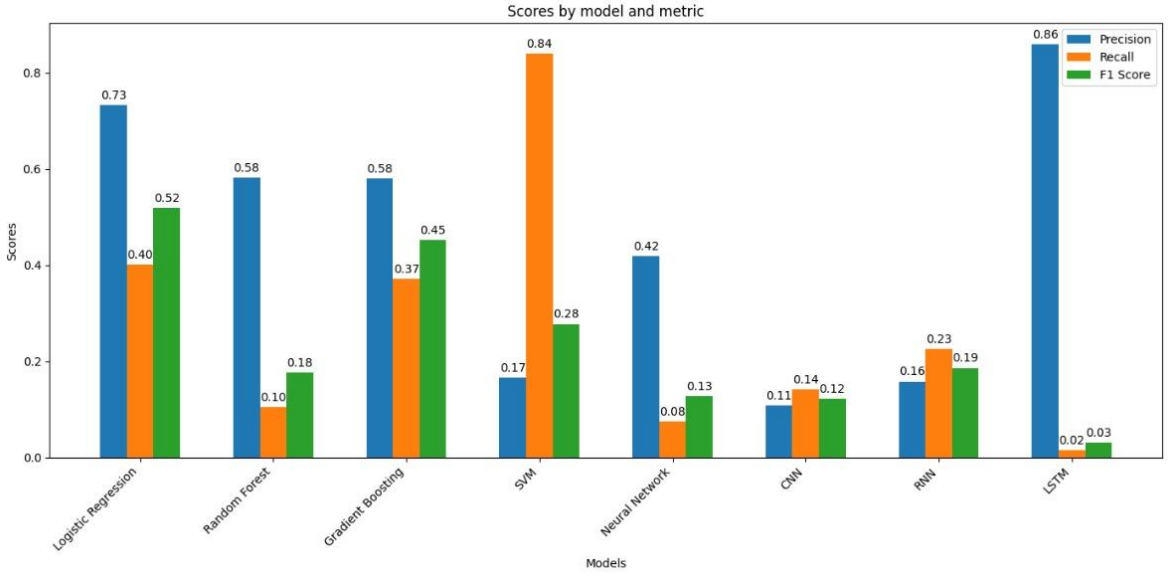

Fig 7. Model Precision, Recall and F1 score on Instagram

D. Model Test Accuracy

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

Refers to the proportion of correctly classified instances by a machine learning model on unseen data, providing insight into its predictive performance. Higher accuracy values indicate better model performance in accurately predicting outcomes.

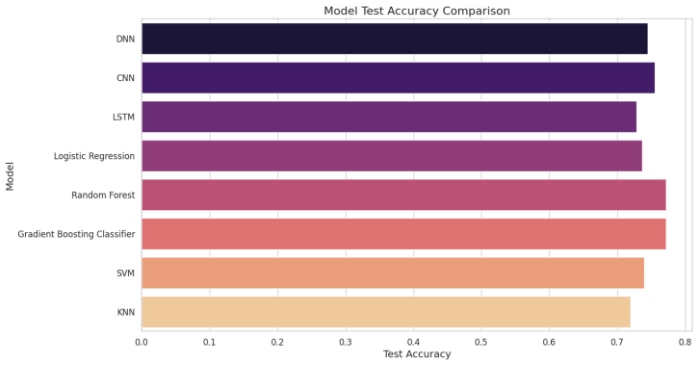

Fig 8. Model Test Accuracy on Twitter

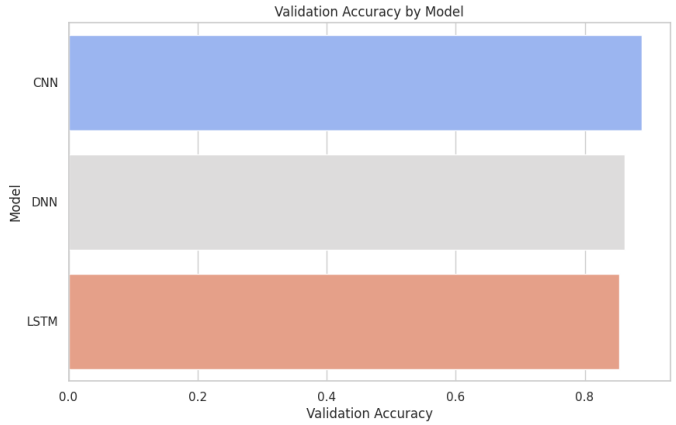

Fig 9. Validation Accuracy on Instagram

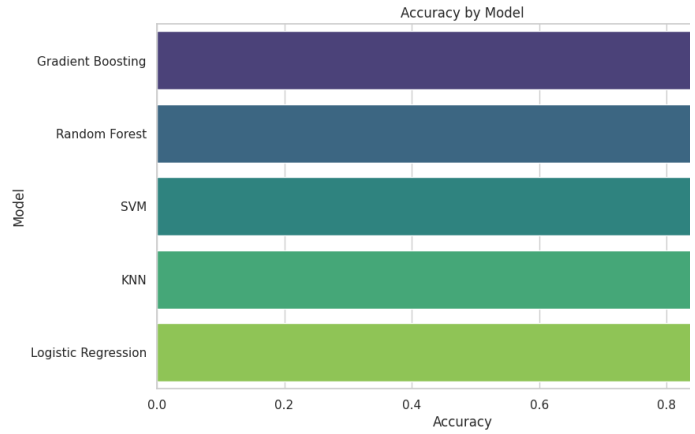

Fig 10. Model Test Accuracy score on Instagram

IX. CONCLUSION

The discussion on fake social media account detection underscores the importance of leveraging machine learning (ML) and deep learning (DL) models to address the proliferation of fraudulent profiles. Despite real-life challenges such as evolving impersonation techniques and adaptive behaviours by malicious actors, these models offer promising solutions. By analysing features like activity patterns and profile information, ML algorithms like logistic regression and random forest, alongside DL models like DNN and CNN, can effectively differentiate between genuine and fake accounts. However, model performance varies, with some, like Random Forest and Gradient Boosting Classifier, consistently exhibiting higher accuracy. Overall, the integration of these models into detection mechanisms holds immense potential for curbing the spread of fake accounts and enhancing the credibility of online platforms.

X. FUTURE WORK

Future research in the realm of detecting fake social media accounts could focus on advancing machine learning and deep learning models to better counter increasingly sophisticated impersonation techniques and evolving strategies employed by malicious entities. This entails exploring innovative feature engineering methods, data augmentation techniques, and interdisciplinary collaborations to bolster model performance and generalization capabilities. Additionally, integrating real-time monitoring systems, anomaly detection algorithms, and user behaviour analysis approaches can contribute to swift and proactive detection and mitigation of fake account activities. Furthermore, the exploration of blockchain technology and ethical considerations regarding privacy and algorithmic biases are essential facets to address for ensuring responsible and effective deployment of fake account detection mechanisms in social media platforms.

XI. REFERENCES

[1]K. V. Nikhitha, K. Bhavya and D. U. Nandini, "Fake Account Detection on Social Media using Random Forest Classifier," 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2023, pp. 806-811, doi: 10.1109/ICICCS56967.2023.10142841

[2] E. Van Der Walt and J. Eloff, "Using Machine Learning to Detect Fake Identities: Bots vs Humans," in IEEE Access, vol. 6, pp. 6540-6549, 2018, doi: 10.1109/ACCESS.2018.2796018.

[3] Balshetwar, S.V., RS, A. & R, D.J. Fake news detection in social media based on sentiment analysis using classifier techniques. Multimed Tools Appl 82, 35781–35811 (2023). https://doi.org/10.1007/s11042-023-14883-3

[4] S. Khaled, N. El-Tazi and H. M. O. Mokhtar, "Detecting Fake Accounts on Social Media," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 3672-3681, doi: 10.1109/BigData.2018.8621913.

[5] He, P., Zhang, X., Lin, C. et al. Towards understanding bogus traffic service in online social networks. Front Inform Technol Electron Eng 25, 415–431 (2024). https://doi.org/10.1631/FITEE.2300068

[6] Saxena, A., Saxena, P., Reddy, H. (2022). Fake News Detection Techniques for Social Media. In: Biswas, A., Patgiri, R., Biswas, B. (eds) Principles of Social Networking. Smart Innovation, Systems and Technologies, vol 246. Springer, Singapore. https://doi.org/10.1007/978-981-16-3398-0_15

[7] M. Zeshan Shabbir, I. Naseer, S. Akhter, M. Abubakar, G. F. Issa and M. Hassaan Mehmood, "Fake Twitter Followers Detection using Machine Learning Approach," 2023 International Conference on Business Analytics for Technology and Security (ICBATS), Dubai,

United Arab Emirates, 2023, pp. 1-7, doi: 10.1109/ICBATS57792.2023.10111260.

[8] K. Mohanapriya, N. Sangavi, A. Kanimozhi, V. R. Kiruthika and P. Dhivya, "Optimized Feed Forward Neural Network for Fake and Clone Account Detection in Online Social Networks," 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2023, pp. 476-481, doi: 10.1109/ICSCDS56580.2023.10104616.

[9] S. R. Ramya, R. Priyanka, S. S. Priya, M. Srinivashini and A. Yasodha, "SVM Based Fake Account Sign-In Detection," 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2023, pp. 509-514, doi: 10.1109/ICOEI56765.2023.10125850.

[10] Feng, Y. (2022). Misreporting and Fake News Detection Techniques on the Social Media Platform. Highlights in Science, Engineering and Technology, 12, 142-152. https://doi.org/10.54097/hset.v12i.1417

[11] A. Bhattacharya, R. Bathla, A. Rana and G. Arora, "Application of Machine Learning Techniques in Detecting Fake Profiles on Social Media," 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2021, pp. 1-8, doi: 10.1109/ICRITO51393.2021.9596373

[12] D. Punkamol and R. Marukatat, "Detection of Account Cloning in Online Social Networks," 2020 8th International Electrical Engineering Congress (iEECON), Chiang Mai, Thailand, 2020, pp. 1-4, doi: 10.1109/iEECON48109.2020.229558.