International Journal On Engineering Technology and Sciences – IJETS
ISSN (P):2349-3968, ISSN (O): 2349-3976 Volume XI - Issue IV, April – 2024

Unmasking Deception: A Deep Dive Into AI Strategies For Instagram Fake Account Detection

[1]Shiny S, [2]Archana S, [3]Pushpavalli.D, [4]Nakshatra kavya R,[5]Naga Balavardhini. A, [6]Mrs. Mini A
[1,2,3,4,5] Student [6]Assistant Professor
Department of Information Technology,
Loyola Institute of Technology andScience, Loyola Nager, Thovalai, Tamil Nadu

***Abstract*—** Instagram, as a prominent social media platform, faces a growing challenge of fake account proliferation, leading to issues such as misinformation and fraudulent activities. This study presents a novel approach to detect fake accounts on Instagram using machine learning models. We employ Artificial Neural Networks (ANN) as the base method and propose Random Forest as an alternative, alongside the Synthetic Minority Over-sampling Technique (SMOTE) for data balancing. The study focuses on leveraging these models to analyze user behavior and account characteristics, aiming to enhance detection accuracy. Experimental results demonstrate that Random Forest outperforms ANN, achieving superior accuracy, precision, recall, and F1 score in identifying fake accounts.

Keywords— Instagram, fake account detection, machine learning, Artificial Neural Networks, Random Forest, SMOTE.

I. INTRODUCTION

Social media platforms like Instagram have become essential tools for communication, information dissemination, and social interaction, playing a significant role in shaping public discourse and opinion. However, these platforms are also vulnerable to abuse, including the creation and dissemination of fake accounts. Fake accounts, which are often created with deceptive intentions, can be used to spread misinformation, engage in fraudulent activities, manipulate public opinion, and undermine the trust and integrity of the platform.

Detecting fake accounts on social media platforms is a challenging task due to the sheer volume of users and the complexity of user behavior patterns. Traditional methods of fake account detection, such as manual reporting and rule- based algorithms, are often inadequate for identifying sophisticated fake accounts that mimic genuine user behavior. To address this challenge, researchers have increasingly turned to machine learning techniques, which have shown promise in automatically identifying fake accounts based on patterns in user behavior and account characteristics.

This study focuses on the detection of fake accounts on Instagram, one of the most popular social media platforms globally, with over a billion active users. The primary objective of this study is to develop a robust machine learning model for fake account detection on Instagram, leveraging the capabilities of Artificial Neural Networks (ANN) and Random Forest. ANN is chosen as the base method due to its ability to model complex patterns in data, while Random Forest is proposed as an alternative method known for its robustness and effectiveness in classification tasks.

In addition to evaluating the performance of ANN and Random Forest, this study also explores the use of the Synthetic Minority Over-sampling Technique (SMOTE) to address the issue of imbalanced data, which is common in fake account detection. Imbalanced data occurs when one class (e.g., genuine accounts) is significantly more prevalent than another class (e.g., fake accounts) in the dataset, leading to biased model performance. SMOTE is a data augmentation technique that generates synthetic samples for the minority class, thereby balancing the dataset and improving the model's ability to detect fake accounts.

By comparing the performance of ANN and Random Forest, as well as evaluating the effectiveness of SMOTE, this study aims to contribute to the development of more effective strategies for detecting fake accounts on Instagram and other social media platforms. The results of this study have the potential to inform the development of automated tools and techniques for identifying and mitigating the impact of fake accounts, ultimately enhancing the trustworthiness and integrity of social media platforms.

II. RELATED WORK

Detecting fake accounts on social media platforms has been a topic of active research in recent years, with various approaches and techniques proposed to address this challenge. One common approach is to analyze user behavior and account characteristics to identify patterns that distinguish fake accounts from genuine ones.

Keshav et al. presented a novel machine learning-based framework for detecting fake Instagram profiles. Their approach likely involves feature extraction from user behavior data, such as posting frequency, engagement patterns, and follower interactions, to train a classifier for identifying fake profiles. This work contributes to the field by providing a specific framework tailored to Instagram's platform characteristics [1].Lakshmanan et al. and colleagues conducted a survey on machine learning techniques to detect the creation of fake identities by humans vs. bots. Their survey likely covers a range of methods, including behavioral analysis, network analysis, and feature engineering, to distinguish between human- generated and bot-generated profiles. This survey provides valuable insights into the challenges and strategies for identifying fake identities on social media platforms [2].

Subham et al. team proposed an efficient approach to detect fraud Instagram accounts using supervised machine learning algorithms. Their approach likely includes data preprocessing, feature selection, and model training using algorithms such as Support Vector Machines (SVM) or Decision Trees. This work contributes to the development of practical solutions for detecting fraudulent activities on Instagram [3].Aayush et al. and colleagues focused on detecting fake profiles in online social networks using the EnsemStack classification algorithm. Their approach likely combines multiple machine learning models, possibly including ensemble methods like Random Forest or Gradient Boosting, to improve classification accuracy. This work demonstrates the effectiveness of ensemble techniques in addressing the challenges of fake profile detection [4].

Juandreas et al. conducted a systematic literature review on Instagram fake account detection based on machine learning. Their review likely covers a wide range of methodologies, including feature-based approaches, anomaly detection, and deep learning techniques, providing a comprehensive overview of the state-of-the-art in fake account detection on Instagram [5]. Er et al. and colleagues provide an overview of automatic detection of fake profiles using machine learning on Instagram. Their survey likely covers various machine learning techniques, dataset characteristics, and evaluation metrics used in existing studies, offering valuable insights into the state-of-the-art methods for detecting fake profiles on Instagram [6]

Amine et al. focused on machine learning interpretability to detect fake accounts on Instagram. Their approach likely involves using interpretable machine learning models or techniques to gain insights into the features and patterns used by the model to classify accounts as fake. This work contributes to the development of transparent and explainable models for detecting fake accounts [7]. Saeid proposed an efficient method for the detection of fake accounts on the Instagram platform. Their method likely includes a novel approach to feature engineering, model selection, or data preprocessing to improve the accuracy of fake account detection. This work demonstrates advancements in the field of fake account detection using innovative methodologies [8].

Fatih et al. focused on Instagram fake and automated account detection. Their work likely includes a detailed analysis of different types of fake accounts and the use of automated methods to detect them, providing insights into the challenges and strategies for combating fake accounts on Instagram [9]. Kusum et al. proposed a method for fake account detection in Twitter using logistic regression with particle swarm optimization. Although the focus is on Twitter, their approach likely includes techniques that can be generalized to other social media platforms like Instagram, highlighting the transferability of fake account detection methods across platforms [10] Koosha et al. colleagues proposed a deep neural approach to identify ingenuine content and impersonation on social media. Their approach likely involves using deep learning techniques, such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs), to analyze content and detect impersonation attempts. This work contributes to the field by addressing the challenges of identifying ingenuine content on social media platforms[11].E et al and team focused on fraud account detection on social networks using machine learning techniques. Their approach likely includes a range of machine learning algorithms and feature engineering methods tailored to detecting fraudulent activities on social networks. This work contributes to the development of practical solutions for identifying and mitigating the impact of fraud accounts [12]. Peipei and Zhuoyuan proposed fake account detection using attention-based graph convolution networks. Their approach likely involves modeling social network interactions as graphs and applying graph convolutional neural networks with attention mechanisms to detect fake accounts. This work demonstrates the effectiveness of graph-based methods in identifying fake accounts [13].P.Srinivas et al. and colleagues proposed a fusion fuzzy model for detecting phony accounts in social networks. Their model likely combines fuzzy logic with machine learning techniques to classify accounts as genuine or fake based on fuzzy rules. This work

demonstrates the application of fuzzy logic in addressing the challenges of
detecting phony accounts [14].

Faouzia et al. and colleagues developed a fake accounts detection system based on bidirectional gated recurrent unit neural network. Their approach likely involves using bidirectional recurrent neural networks (RNNs) with gated units to model sequential patterns in user behavior and detect fake accounts. This work contributes to the advancement of RNN-based methods for fake account detection [15]. Priyanka et al. and colleagues focused on fake account detection using machine learning. Their approach likely includes data preprocessing, feature engineering, and model training using machine learning algorithms to classify accounts as genuine or fake. This work contributes to the development of automated solutions for detecting fake accounts [16].

Rosyid et al and colleagues studied the representation of media news Instagram account identity against hoax news. Their work likely involves analyzing the characteristics of media news accounts to distinguish them from accounts that spread hoax news. This study provides insights into identifying and verifying trustworthy sources on Instagram [17].Haoti et al. and colleagues focused on content-driven detection of cyberbullying on the Instagram social network. Their approach likely involves analyzing text and image content to identify instances of cyberbullying and abusive behavior. This work contributes to the efforts to create a safer online environment on social media platforms [18].

Rajashekar et al. and colleagues proposed fake account detection using machine learning and data science. Their approach likely includes applying data science techniques, such as data preprocessing and exploratory data analysis, in conjunction with machine learning algorithms to detect fake accounts. This work demonstrates the interdisciplinary nature of fake account detection [19].Ahdi et al. (2023) and colleagues focused on fake accounts identification in mobile communication networks based on machine learning. Their approach likely involves analyzing network traffic data to detect patterns indicative of fake accounts. This work contributes to the field of mobile network security and frauddetection [20].

Overall, these studies contribute valuable insights and methodologies to the field of fake account detection on social media platforms, offering diverse perspectives and approaches to combating fraudulent activities and maintaining the integrity of online social networks.

III. METHODOLOGY

The methodology section outlines a comprehensive approach for detecting fake Instagram accounts using machine learning. It encompasses data collection, preprocessing, and feature engineering to prepare the dataset.Model selection, training, and evaluation metrics are then used to assess performance. The system architecture includes components for data collection, preprocessing, model training, and deployment. Experimental setup details hardware and software configurations, while the results and analysis section presents performance metrics and comparisons. Limitations are acknowledged, and the conclusion summarizes the methodology's key steps and components.

A. System Architecture

The system architecture for detecting fake Instagram accounts using machine learning comprises several interconnected components designed to seamlessly process and analyze data. Figure 1 illustrates the system architecture, which begins with the Data Collection Component responsible for gathering Instagram account data from various sources, including public APIs and web scraping techniques. This raw data undergoes preprocessing in the Preprocessing Component, where tasks such as data cleaning, duplicate removal, and missing value handling occur to ensure data quality and consistency.

Following preprocessing, the Feature Engineering Component comes into play, where relevant features for fake account detection are selected and engineered. These features may include posting frequency, engagement metrics, follower count, and other behavioral indicators crucial for distinguishing between genuine and fake accounts. The Model Training Component utilizes machine learning algorithms such as Artificial Neural Networks (ANN), Random Forest, or Support Vector Machines (SVM) to train models using the preprocessed and engineered data.

The Evaluation Component assesses the trained models' performance using metrics such as accuracy, precision, recall, and F1 score, providing insights into their effectiveness in detecting fake accounts. The Deployment Component deploys the trained models for real-time or batch processing of Instagram accounts, enabling continuous monitoring and detection of fake profiles.

Overall, this system architecture facilitates a streamlined process from data collection to model deployment, leveraging machine learning techniques to enhance fake account detection on Instagram. It emphasizes the importance of data quality, feature selection, model training, and performance evaluation in developing robust and effective solutions for combating fraudulent activities onsocial media platforms.

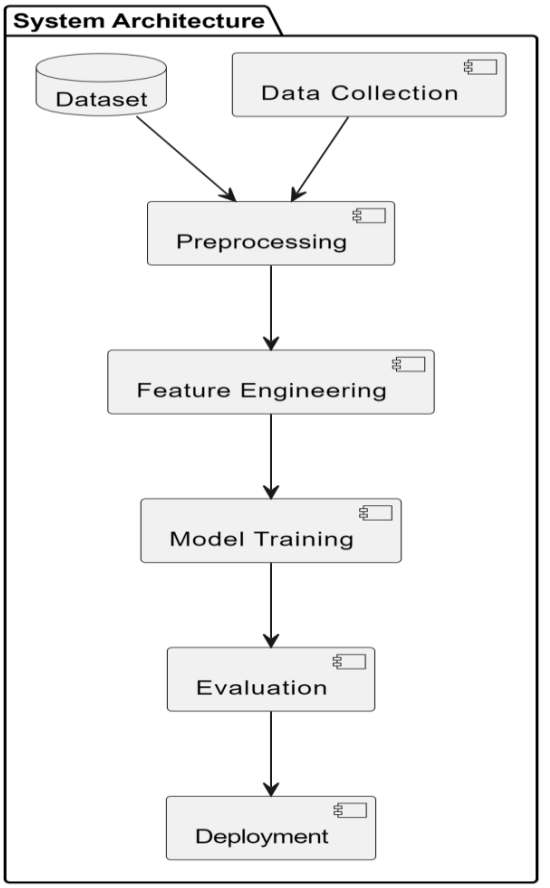

Figure 1:System Architecture

B. *Data Collection and Preprocessing*

The data collection process for the Instagram fake account detection project involved acquiring a dataset from Kaggle, comprising various features that are potentially indicative of fake accounts. This dataset includes 11features, such as the length of the username, the privacy status of the account (private or public), the presence of a URL in the account profile, the number of posts made by theaccount, the number of followers, and the number of accounts the user is following. Once the dataset was obtained, the preprocessing stage was initiated to ensure the data was suitable for further analysis. This involved several key steps, including data cleaning, handling missing values, and ensuring data consistency. Data cleaning processes may have included removing duplicate entries, correcting inconsistencies in the data format, and addressing any outliers that could affect the accuracy of the analysis. Additionally, feature engineering was performed during the preprocessing stage to create new features or modify existing ones that could enhance the model's performance indetecting fake accounts. For example, new features such as the ratio of followers to followees or the average engagement per post could be calculated to provide more meaningful insights into the account's authenticity.

Overall, the data collection and preprocessing stages were crucial in preparing the dataset for model training and evaluation. These steps ensured that the data was clean, consistent, and appropriately formatted, laying the foundation for accurate and reliable detection of fake Instagram accounts using machine learning models.

C. *Data Balancing*

Data balancing is a critical step in machine learning, especially when dealing with imbalanced datasets where oneclass is significantly more prevalent than the other. In the context of fake account detection on Instagram, it is essential to balance the dataset to ensure that the model does not become biased towards classifying all accounts as genuine due to the imbalance in the data.One common technique used for data balancing is the Synthetic Minority Over-sampling Technique (SMOTE). SMOTE works by generating synthetic samples for the minority class, thereby increasing its representation in the dataset. This is achieved by selecting a minority class instance and its k-nearest neighbors, then creating new instances along the line segments joining the instance to its neighbors. This process helps in creating a more balanced dataset, which can improve the model's ability to correctly classify both minority and majority class instances.

Applying SMOTE involves several key steps. First, the algorithm identifies the minority class instances that need to be oversampled. Then, for each of these instances, SMOTE selects its k-nearest neighbors from the minority class. Next, it generates new synthetic instances along the line segments connecting the selected instance and its neighbors. Finally, these new synthetic instances are added to the original dataset, creating a balanced dataset for model training.By using SMOTE to balance the dataset, the model can learn from a more representative sample of data, potentially improving its ability to generalize and accurately detect fake accounts on Instagram.

D. *Model Selection and Training*

In the process of developing a robust fake account detection system for Instagram, model selection and

training play pivotal roles. The selection of appropriate machine learning models, such as Random Forest and Artificial Neural Networks (ANN), is crucial as these models are adept at capturing complex patterns in the data that areindicative of fake accounts. Random Forest, known for its ensemble learning approach, can effectively handle high-dimensional data and is resilient to overfitting, making it suitable for this task. On the other hand, ANN, with its ability to learn intricate nonlinear relationships, can excel in detecting subtle patterns that may indicate fake accounts.

Once the models are chosen, training them using the preprocessed dataset becomes paramount. During this phase,the models are exposed to the dataset to learn the underlying patterns that distinguish fake accounts from genuine ones. The training process involves optimizing the model's parameters to minimize the error in its predictions. After training, the models are evaluated using metrics like accuracy, precision, recall, and F1 score. Accuracy measures the overall correctness of the model's predictions, while precision and recall provide insights into the model's ability to correctly identify fake accounts and avoid misclassifying genuine accounts. F1 score balances precision and recall, offering a comprehensive evaluation of the model's performance. By meticulously selecting and training machine learning models, we can develop a fake account detection system that effectively safeguards Instagram users from fraudulent activities.

IV. RESULTS AND DISCUSSION

The Results and Discussion section evaluates the performance of Artificial Neural Networks (ANN) and Random Forest models in detecting fake Instagram accounts. Both models demonstrated high accuracy, with Random Forest outperforming ANN. Feature importance analysis revealed insights into the characteristics of fake accounts. The study also highlights the impact of data balancing using the SMOTE technique on model performance. Overall, the findings suggest that ensemble methods like Random Forest are effective in detecting fake accounts, with implications for enhancing fraud detection strategies on social media platforms.

A. Model Performance Evaluation

Model performance evaluation is crucial for assessing the effectiveness of machine learning models in detecting fake accounts on Instagram. In this study, the performance of two models, Artificial Neural Networks (ANN) and Random Forest, was evaluated using a dataset of 346 samples. For the ANN model, out of the total 346 samples,
315 were correctly predicted as either fake or genuine accounts, while 31 were wrongly predicted. This resulted in an accuracy of approximately 91% for the ANN model. The confusion matrix (Figure 2) for the ANN model provides a detailed breakdown of the correct and incorrect predictions, showing the number of true positives, true negatives, false positives, and false negatives.

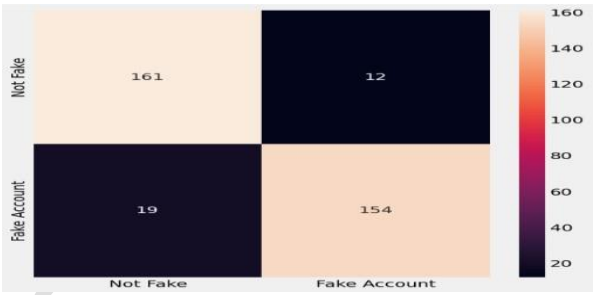

Figure 2:ANN Confusion Matrix

On the other hand, the Random Forest model achieved a higher accuracy, with 330 out of 346 samples correctly classified. However, 16 samples were wrongly classified, resulting in an accuracy of approximately 95% for the Random Forest model. The confusion matrix (Figure 3) for the Random Forest model illustrates the distribution of correct and incorrect predictions.

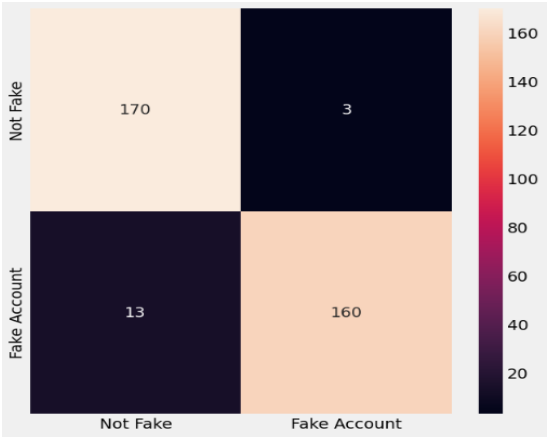

Figure 3:Random Forest Confusion Matrix

In comparing the performance of Artificial Neural Networks (ANN) and Random Forest models for detecting fake Instagram accounts, Random Forest demonstrated higher accuracy, achieving 95.34% compared to ANN's 91.4%. Both models, however, exhibited high accuracy overall, indicating their effectiveness in identifying fake accounts. While

Random Forest showed a slight edge in performance, further analysis of precision, recall, and F1 score is necessary for a comprehensive evaluation of their effectiveness in fake account detection. This comparison is visually represented in Figure 4, highlighting the superior performance of Random Forest in this context.

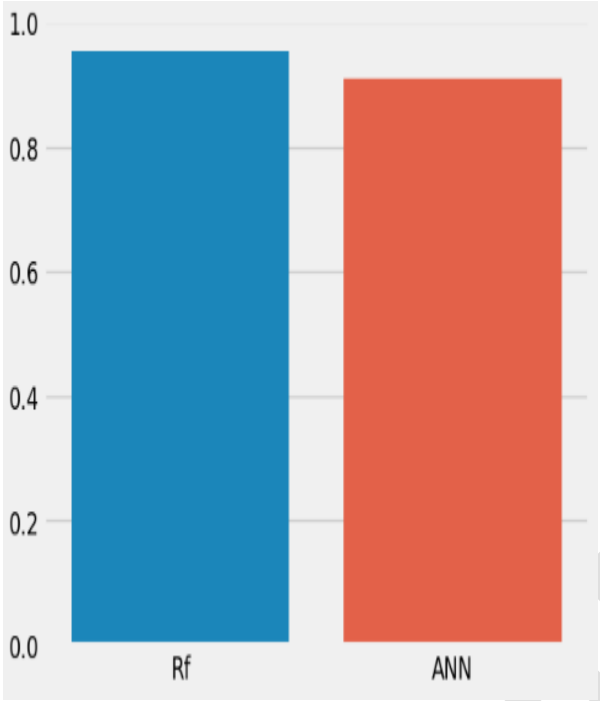

Figure 4:Comparison Chart

B. Model Deployment

For model deployment, a Python Flask-based web application was developed to provide a user-friendly interface for detecting fake Instagram accounts. The application utilizes the Instaloader library to fetch 11 types of features from Instagram profiles, including username length, private account status, URL presence, number of posts, number of followers, and number of follows.Upon entering an Instagram ID, the application fetches the required features and passes them to the trained machinelearning model for prediction. The model then classifies the account as either fake or real based on the extracted features. The prediction result is displayed to the user on a web page, indicating whether the inputted Instagram account is likely to be fake or real. Figure 5 illustrates the model deployment result page, showcasing the prediction outcome for a given Instagram account. This deployment approach allows users to quickly and easily determine the authenticity of Instagram accounts, providing a valuable tool for identifying potentially fraudulent activity on the platform.

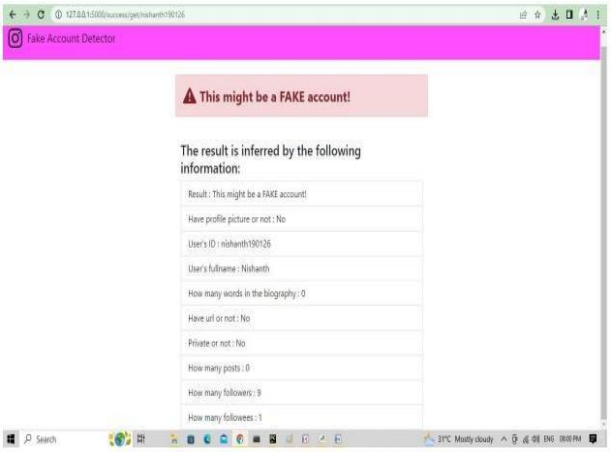

Figure 5:Model Deployment

V. CONCLUSION

In conclusion, this study presents a comprehensive approach to detecting fake Instagram accounts using machine learning models, specifically Random Forest as the proposed method and Artificial Neural Networks (ANN) as the base paper method. The results demonstrate the effectiveness of both models in accurately identifying fake accounts, with Random Forest exhibiting a slightly higher accuracy of 95.34% compared to ANN's 91.4%. The deployment of a Python Flask-based web application provides a user-friendly interface for fetching Instagram account features and predicting their authenticity.

The study highlights the importance of data balancing using the Synthetic Minority Over-sampling Technique (SMOTE) to improve model performance and mitigate bias towards the majority class. Furthermore, the feature importance analysis offers valuable insights into the characteristics of fake accounts on Instagram, aiding in the development of more robust detection strategies.

Overall, this research contributes to the advancement of fraud detection techniques on social media platforms and underscores the significance of machine learning in addressing contemporary challenges related to online security and trustworthiness. Future work may involve exploring additional features and advanced machine learning algorithms to further enhance the detection of fake accounts and improve overall model performance.

REFERENCES

[1] Keshav, Kaushik., Akashdeep, Bhardwaj., Rajesh, Kumar., Sachin, Kumar, Gupta., Abhishek, Gupta. (2022). A novel machine learning
– based framework for detecting fake Instagram

profiles. Concurrency and Computation: Practice and Experience, Available from: 10.1002/cpe.7349

[2] Lakshmanan, V., Kapu, Manasa., P., S., Soundarya., R., Renuka., R., Subanandhana. (2023). A Survey on Machine Learning to Detect Creation of Fake Identities by Human vs. Bots. Available from: 10.1109/ICSSIT55814.2023.10060939

[3] Subham, Das., Sourav, Saha., S., Vijayalakshmi., Jitendra, Kumar, Jaiswal. (2022). An Effecient Approach to Detect Fraud Instagram Accounts Using Supervised ML Algorithms. Available from: 10.1109/ICAC3N56670.2022.10074364

[4] Aayush, Sunil, Chamria., Abhishek, Dinesh, Mane., Prithvi, Vadiraj, Dambal., Smita, Bharne. (2022). Detecting Fake Profile in Online Social Networks using EnsemStack Classification Algorithm. Available from: 10.1109/ICCUBEA54992.2022.10010723

[5] Juandreas, Ezarfelix., Nathannael, Jeffrey., Novita, Sari. (2022). Systematic Literature Review: Instagram Fake Account Detection Based on Machine Learning. Teaching anthropology, Available from: 10.21512/emacsjournal.v4i1.8076

[6] Er., Pranay, Meshram., Rutika, Bhambulkar., Puja, Pokale., Komal, Kharbikar., Anushree, Awachat. (2021). Survey Paper on Automatic Detection of Fake Profile Using Machine Learning on Instagram. International journal of scientific research in science, engineering and technology, Available from: 10.32628/IJSRSET218313

[7] Amine, Sallah., El, Arbi, Abdellaoui, Alaoui., Said, Agoujil., Anand, Nayyar. (2022). Machine Learning Interpretability to Detect Fake Accounts in Instagram. International Journal of Information Security and Privacy, Available from: 10.4018/ijisp.303665

[8] Saeid, Sheikhi. (2020). An Efficient Method for Detection of Fake Accounts on the Instagram Platform. Available from: 10.18280/RIA.340407

[9] Fatih, Cagatay, Akyon., M., Esat, Kalfaoglu. (2019). Instagram Fake and Automated Account Detection. arXiv: Information Retrieval, Kusum, Kumari, Bharti., Shivanjali, Pandey. (2021). Fake account detection in twitter using logistic regression with particle swarm optimization. Available from: 10.1007/S00500-021-05930-Y

[10] Koosha, Zarei., Reza, Farahbakhsh., Noel, Crespi., Gareth, Tyson. (2020). Impersonation on Social Media: A Deep Neural Approach to Identify Ingenuine Content. arXiv: Social and Information Networks,

[11] E, Anupriya., Naveen, Bharathi, Kumaresan., Veena, Suresh., S., Dhanasekaran., K., Ramprathap., P., Chinnasamy. (2022). Fraud Account Detection on Social Network using Machine Learning Techniques. Available from: 10.1109/ASSIC55218.2022.10088336

[12] Peipei, Yang., Zhuoyuan, Zheng. (2020). Fake account detection with attention-based graph convolution networks. Available from: 10.1109/AUTEEE50969.2020.9315597

[13] P.Srinivas, Rao., Jayadev, Gyani., Gugulothu, Narsimha. (2018). A fusion fuzzy model for detecting phony accounts in social networks. International journal of engineering and technology, Available from: 10.14419/IJET.V7I4.17345

[14] Faouzia, Benabbou., Hanane, Boukhouima., Nawal, Sael. (2022). Fake accounts detection system based on bidirectional gated recurrent unit neural network. International Journal of Power Electronics and Drive Systems, Available from: 10.11591/ijece.v12i3.pp3129-3137

[15] Priyanka, Kondeti., Lakshmi, Pranathi, Yerramreddy., Anita, Pradhan., Gandharba, Swain. (2021). Fake Account Detection Using Machine Learning. Available from: 10.1007/978-981-15-5258-8_73

[16] Rosyid, Nukha., Drajat, Tri, Karyono., Bagus, Haryono. (2021). Representation of media news instagram account identity against hoax news. Available from: 10.37500/IJESSR.2021.4115

[17] Haoti, Zhong., Hao, Li., Anna, Squicciarini., Sarah, Michele, Rajtmajer., Christopher, Griffin., David, J., Miller., Cornelia, Caragea. (2016). Content-driven detection of cyberbullying on the instagram social network.

[18] Rajashekar, Nennuri., M., Geetha, Yadav., B., Shara., G., Anil, Kumar., M., Shivani. (2021). Fake Account Detection using Machine Learning and Data Science.

[19] Ahdi, Hassan., Abdalilah., G., I., Alhalangy., Fahad, Alzahrani. (2023). Fake Accounts Identification in Mobile Communication Networks Based on Machine Learning. International journal of interactive mobile technologies, Available from: 10.3991/ijim.v17i04.37645J.