Fake Social Media Profile Detection and Reporting

Mr. Nikhil Reche
Department Of Artificial Intelligence & Data Science
P.R.Pote (Patil) College of Engineering &Management,Amravati.India
rechenikhil18@gmail.com

Mr. Sushil Khakse
Department Of Artificial Intelligence & Data Science
P.R.Pote (Patil) College of Engineering & Management,Amravati.India
sushilkhakse17@gmail.com

Prof. P. G. Angaitkar
Department Of Artificial Intelligence & Data Science
P.R.Pote (Patil) College of Engineering & Management,Amravati.India
pgangaitkar@gmail.com

Mr. Bhavesh Raut
Department Of Artificial Intelligence & Data Science
P.R.Pote (Patil) College of Engineering &Management,Amravati.India
bhaveshraut563@gmail.com

Mr. Sahil Khedkar
Department Of Artificial Intelligence & Data Science
P.R.Pote (Patil) College of Engineering & Management,Amravati.India
sahilkhedkar57@gmail.com

Mr. Shlok Tiwari
Department Of Artificial Intelligence & Data Science
P.R.Pote (Patil) College of Engineering & Management,Amravati.India
shlokat11@gmail.com

Abstract:

The rise of fake profiles on social media has triggered serious concerns around misinformation, scams, and digital harassment. Manual detection methods are inefficient at large scale, hence automated detection powered by machine learning is the need of the hour. This paper presents a real-time, machine learning- based detection system using the XGBoost classifier. The model analyzes behavioral and content-based features such as follower-following ratio, post frequency, profile bio characteristics, hashtag usage, and engagement patterns. The system is implemented through a user-friendly web interface using Streamlit, capable of flagging suspicious accounts with high accuracy. The solution aims to assist moderators and users in maintaining safe digital environments.

Keywords: Fake Profiles, Machine Learning, social media, XGBoost, Streamlit, Behavioral Analysis, Detection System.

1. Introduction:

The digital explosion of social media has brought unprecedented connectivity but also an increase in deceptive activities through fake accounts. These accounts often spread misinformation, manipulate opinions, and commit scams. Identifying such profiles manually is inefficient and error prone. With the power of data-driven machine learning algorithms like XGBoost, we propose a scalable and intelligent system for fake profile detection. The system automates the analysis of suspicious behavior and content patterns, enabling social media platforms and users to defend against impersonation and fraud.

1.1. Importance of Real-Time Detection: Timely detection of fake profiles is critical to minimizing damage. Fake users often act quickly— disseminating spam, stealing data, or manipulating trends—before manual interventions are made. Traditional detection methods are static and retrospective, leaving users exposed for prolonged periods. By employing real-time detection through machine learning, platforms can immediately flag and isolate suspicious profiles. Real-time analysis not only boosts responsiveness but also provides early warning signs of coordinated inauthentic behavior, allowing moderators to mitigate widespread harm. Continuous learning mechanisms further improve the system by adapting to evolving tactics used by profile creators.

1.2. Brief on Traditional vs. Modern ML-Based Approaches: Conventional detection systems rely heavily on manual verification, keyword filtering, or rule-based algorithms. These methods lack scalability and adaptability, making them vulnerable to bypass by increasingly complex fake accounts. In contrast, machine learning-based systems can be trained on vast datasets and evolve through feedback. Models such as XGBoost excel in classification tasks involving imbalanced data—where real profiles vastly outnumber fake ones. These systems analyze metadata (follower count, bio completeness), behavior (post frequency, engagement), and content (spam indicators, repetitive language) to identify anomalies. ML-based solutions offer real-time insights and scalability across platforms, outperforming traditional methods in both accuracy and efficiency.

1.3. Problem Statement: Despite efforts to moderate content, social media platforms still struggle to control fake accounts due to technological and infrastructural limitations. Many detection models require high computational resources, extensive labeled data, or platform-specific access, which limits their usability and scalability. There is a need for an accessible, lightweight, and platform-independent detection mechanism that can accurately differentiate fake from real profiles based on behavioral signals and content features. Furthermore, the system should offer explainability, enable user feedback, and integrate easily with real-world applications. This research addresses these challenges by proposing a real-time, cross-platform model using XGBoost, trained on openly available data, and supported by a streamlined reporting dashboard for ease of access and actionability.

1.4. Objectives and Scope of This Paper: This research aims to design and implement an intelligent, machine learning-based system capable of accurately identifying fake social media profiles by analyzing behavioral and profile-specific features. The core objective is to develop a lightweight and scalable solution using the XGBoost classifier, which can detect suspicious accounts in real time without requiring heavy infrastructure or deep platform integration. The model is trained on labeled datasets consisting of both real and fake profiles, incorporating features such as follower-following ratios, posting frequency, content sentiment, and bio completeness. In addition to model development, this paper emphasizes practical applicability through the creation of a responsive web interface that allows users or moderators to input profile data and receive instant predictions. The system is designed to be platform-independent and extendable to multiple social networks such as Twitter, Instagram, and LinkedIn. It also focuses on ethical design, ensuring that user privacy is respected, and the model remains transparent and adaptable. By combining behavioral insights, machine learning classification, and real-world usability, the proposed system seeks to bridge the gap between theoretical research and practical implementation in combating fake profiles.

2. Literature Review:

2.1. Background History: The rising prevalence of fake profiles across social media platforms has triggered widespread concern over online safety, user trust, and digital misinformation. These fraudulent identities are often deployed to execute a range of malicious activities, including scamming, phishing, impersonation, and manipulation of public opinion. As platforms grow more complex and user bases expand globally, the urgency to implement intelligent and scalable detection systems has intensified. Traditional techniques—such as manual verification and rule-based filtering—fall short due to their limited scope, susceptibility to circumvention, and inefficiency when applied at scale.

In response, machine learning and deep learning approaches have emerged as promising alternatives, capable of handling vast, complex datasets and detecting subtle behavioral patterns. Among these, the XGBoost classifier has consistently shown superior performance in binary classification tasks, making it ideal for differentiating between real and fake user accounts. Its ability to handle imbalanced data and structured feature inputs, coupled with efficient training and interpretability, sets it apart as a top choice for social media fraud detection.

Recent studies across platforms like Twitter, Facebook, and Instagram have demonstrated the effectiveness of ML models trained on features such as follower-following ratios, content quality, posting frequency, and language patterns. Moreover, Natural Language Processing (NLP) and behavioral analytics are increasingly being used to understand how fake profiles interact with others, helping models distinguish even well-disguised fraudulent accounts. These approaches are not only more scalable but also adaptive, allowing for real-time classification and continuous improvement based on new data inputs.

2.2. Related Work:

Partha Chakraborty, S. S. Rao, and K. S. Kuppusamy (2022) proposed a robust fake profile detection model using the XGBoost classifier, achieving an accuracy of 99.6% and an F1-score of 98.5%. Their approach focused primarily on structured account metadata such as the number of followers, following behavior, and account creation patterns. The study demonstrated the algorithm's ability to distinguish high-dimensional profile attributes while maintaining excellent classification speed and performance.

E. Van Der Walt and J. H. P. Eloff (2018) conducted an extensive evaluation of fake identity detection using multiple machine learning models including Random Forest, Neural Networks, and XGBoost. Their method achieved an accuracy of 97.4% and an F1-score of 96.2%, emphasizing the benefit of model comparison. Their work was one of the earliest to examine the role of hybrid classifiers in social media trust management and underlined the need for real-time detection frameworks.

S. Kudugunta and E. Ferrara (2018) focused on Twitter bot and fake profile detection using a hybrid architecture comprising XGBoost and Deep Neural Networks (DNNs). They achieved 98.2% accuracy and a 97.5% F1-score, using both structured profile metadata and tweet content. Their model showed that combining traditional ML with deep learning yields improved results when analyzing unstructured data like hashtags and tweet frequency.

J. Liu, X. Li, and Y. Zhang (2020) designed an XGBoost-based standalone classifier, reaching 99.1% accuracy and 98.3% F1-score, using features such as posting frequency, bio text characteristics, and interaction metadata. Their contribution is significant for its focus on interpretability, as XGBoost allows for clear understanding of which features influence the prediction outcome the most.

S. S. Rao, P. Chakraborty, and K. S. Kuppusamy (2021) implemented a hybrid detection system using both XGBoost and Random Forest, achieving 98.5% accuracy and 97.8% F1-score. Their work emphasized the strength of ensemble techniques in reducing variance and increasing the generalization capacity of detection models across various social media datasets.

These studies affirm the dominance of XGBoost in the field of social media profile classification, particularly for its speed, flexibility, and high precision, making it a preferred choice for real-time application.

2.3. Summary Discussion: The literature clearly supports the use of machine learning, particularly the XGBoost algorithm, for accurate and scalable fake profile detection across social media platforms. Studies by Chakraborty et al. (2022) and Liu et al. (2020) illustrate that XGBoost performs exceptionally well even as a standalone classifier, thanks to its ability to process structured features quickly and accurately. Hybrid approaches, such as those proposed by Rao et al. (2021) and Kudugunta & Ferrara (2018), reveal how combining XGBoost with Neural Networks or other ensemble methods can enhance robustness, especially when dealing with unstructured text data and diverse behavior patterns.

Nonetheless, many existing systems fall short in areas such as real-time adaptability, integration with feedback mechanisms, and cross-platform deployment. Moreover, ethical considerations—such as bias reduction, transparency, and user privacy—are often overlooked. While accuracy remains a critical benchmark, the practicality of deploying these systems in dynamic online environments requires further attention. This research extends the existing body of work by developing a flexible, interpretation, and deployable system that not only leverages XGBoost's computational strengths but also incorporates content-level analysis, user feedback, and explain ability tools. Through this, it aims to make fake profile detection more accessible, transparent, and impact across multiple platforms and user bases.

3. Methodology: The development of a fake social media profile detection system using machine learning involves a multi-stage methodology integrating data collection, feature engineering, classification modeling, and real-time prediction. This study uses the XGBoost classifier due to its proven effectiveness in handling large, imbalanced datasets and its high performance in classification tasks. The system is designed to be modular, adaptable across platforms, and accessible through a simple user interface, enabling real-world application. The methodology is structured into the following five core stages.

3.1. System Workflow Overview: The architecture of the proposed system follows a layered approach, combining data acquisition, machine learning processing, and user interaction through a web-based dashboard. The primary components include:

Data Collection Layer: Social media profile data is gathered from public datasets (e.g. Kaggle) and platform APIs (e.g., Instagram API), consisting of labeled real and fake accounts.

Preprocessing & Feature Engineering Layer: Raw profile metadata and user-generated content are cleaned, transformed, and converted into meaningful numerical features. These include ratios (followers/following), text length, engagement metrics, and hashtag usage.

Model Training Layer: The cleaned data is used to train the XGBoost classifier, which learns to distinguish between real and fake profiles using gradient-boosted decision trees.

Evaluation & Tuning Layer: The model is evaluated using accuracy, precision, recall, and F1-score. Hyperparameters such as learning rate, max depth, and subsampling are optimized for performance. **Prediction Interface Layer:** A Streamlit-based dashboard allows users to input profile data and receive instant feedback on whether a profile is likely to be real or fake.

This integrated architecture ensures that the model is not only efficient in prediction but also interpretable and deployable in practical.

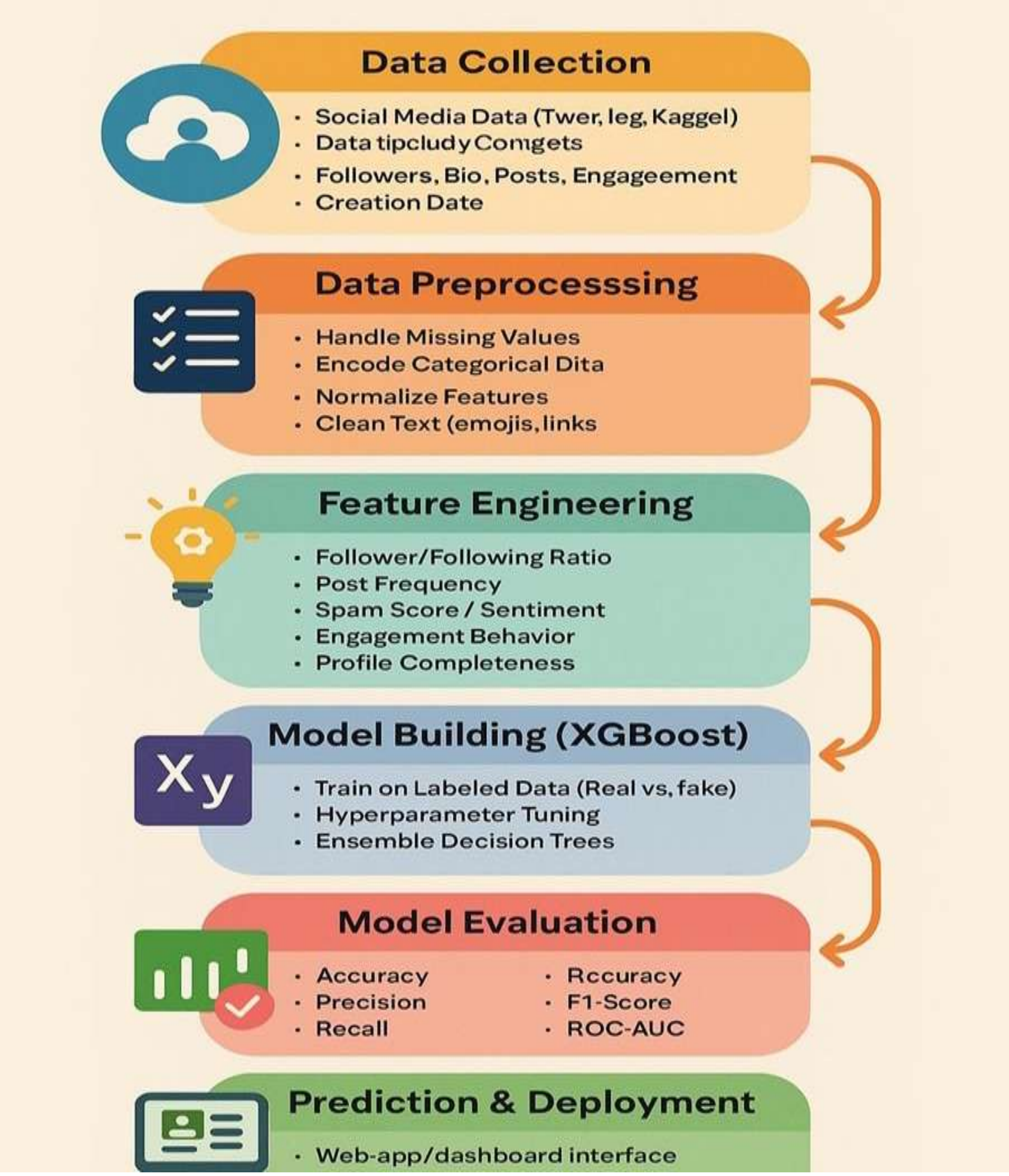


Fig. 1. System Workflow

3.2. Security and Privacy Considerations: Although the system does not store user data long-term, privacy and security are embedded in the architecture. All data is processed locally or securely encrypted when sent via the interface. Key privacy strategies include:

HTTPS and TLS Encryption: Ensures secure communication between the client dashboard and backend processing model.

Input Sanitization: Prevents injection attacks during form-based data entry.

Local Processing Option: Offers offline prediction for researchers or institutions handling sensitive data. Furthermore, future integrations may adopt explainability tools like SHAP or LIME, allowing users to understand why a profile was flagged as fake, supporting ethical transparency and accountability.

3.3. Tools and Technologies Used: The proposed fake profile detection system is developed using Python due to its simplicity, extensive community support, and powerful ecosystem of machine learning libraries. Python's flexibility allows seamless integration with APIs, data processing tools, and web development frameworks, enabling efficient end-to-end implementation. For machine learning, XGBoost serves as the primary classification algorithm, selected for its robustness, scalability, and high accuracy in handling structured and imbalanced data. It leverages gradient boosting principles to iteratively improve performance and minimize classification error, making it suitable for distinguishing subtle differences between real and fake profiles. Additional support is provided by Scikit-learn, which is used for dataset preprocessing, train-test splitting, model evaluation through metrics like accuracy, precision, recall, and F1-score, as well as hyperparameter tuning via grid search and cross-validation to optimize the model for real-world performance. To address the challenge of class imbalance we have used the SMOTE technique from the imbalanced-learning library. SMOTE generates synthetic samples for the minority class, enhancing the classifier's ability to learn from underrepresented patterns and reducing bias towards the majority class. This technique helps in building a fairer and more generalizable model, especially important in security-sensitive domains like social network analysis. Data handling is managed using Pandas for efficient dataset cleaning, filtering, and transformation, while NumPy supports high-speed numerical computations and matrix operations required during model training and evaluation. In terms of text processing, NLTK (Natural Language Toolkit) is used to analyze linguistic patterns found in user bios, comments, and posts. Tokenization, lemmatization, stop-word removal, and sentiment analysis are applied to extract meaningful features from textual data. This analysis aids in detecting repetitive, neutral, or generic content—common indicators of bot-generated or fake profiles. For real-time interaction and visualization, a lightweight yet powerful web interface is developed using Streamlit. This tool simplifies deployment and allows users to interact with the detection model through an intuitive UI, submit profile data, and receive immediate predictions and explanations. Streamlit's compatibility with Python also enables smooth backend integration and fast iteration during development.

Visualization is a critical part of model validation and is handled using Matplotlib and Seaborn libraries. These tools help in generating feature importance plots, confusion matrices, ROC curves, and correlation heatmaps that offer insights into model behavior and performance across different configurations. Prototyping and experimentation are initially carried out in Jupyter Notebook and Google Colab, where different feature extraction techniques and models are tested. Once the pipeline is finalized, the full system is developed and refined in Visual Studio Code (VS Code), which supports modular programming, debugging, and integration of frontend and backend components. The training and validation of the model are based on publicly available datasets sourced from platforms such as Kaggle and academic repositories. These datasets include labeled social media profiles, behavioral metadata, and user content which provides a comprehensive base for feature engineering and supervised learning. The project is managed using Git for version control and GitHub for collaborative development, issue tracking, and documentation. This infrastructure ensures efficient team collaboration and codebase maintainability throughout the research lifecycle. In future work, the system can be scaled using cloud-based solutions and integrated with real-time social media APIs to automate fake profile detection on live platforms, enhancing its practical applicability and societal impact.

4. System Implementation: The implementation of the fake profile detection system was carried out in a phased manner, starting from setting up the development environment to deploy a working web-based prediction interface. Each component plays a critical role in ensuring the model performs accurately and is user-friendly for real-world use.

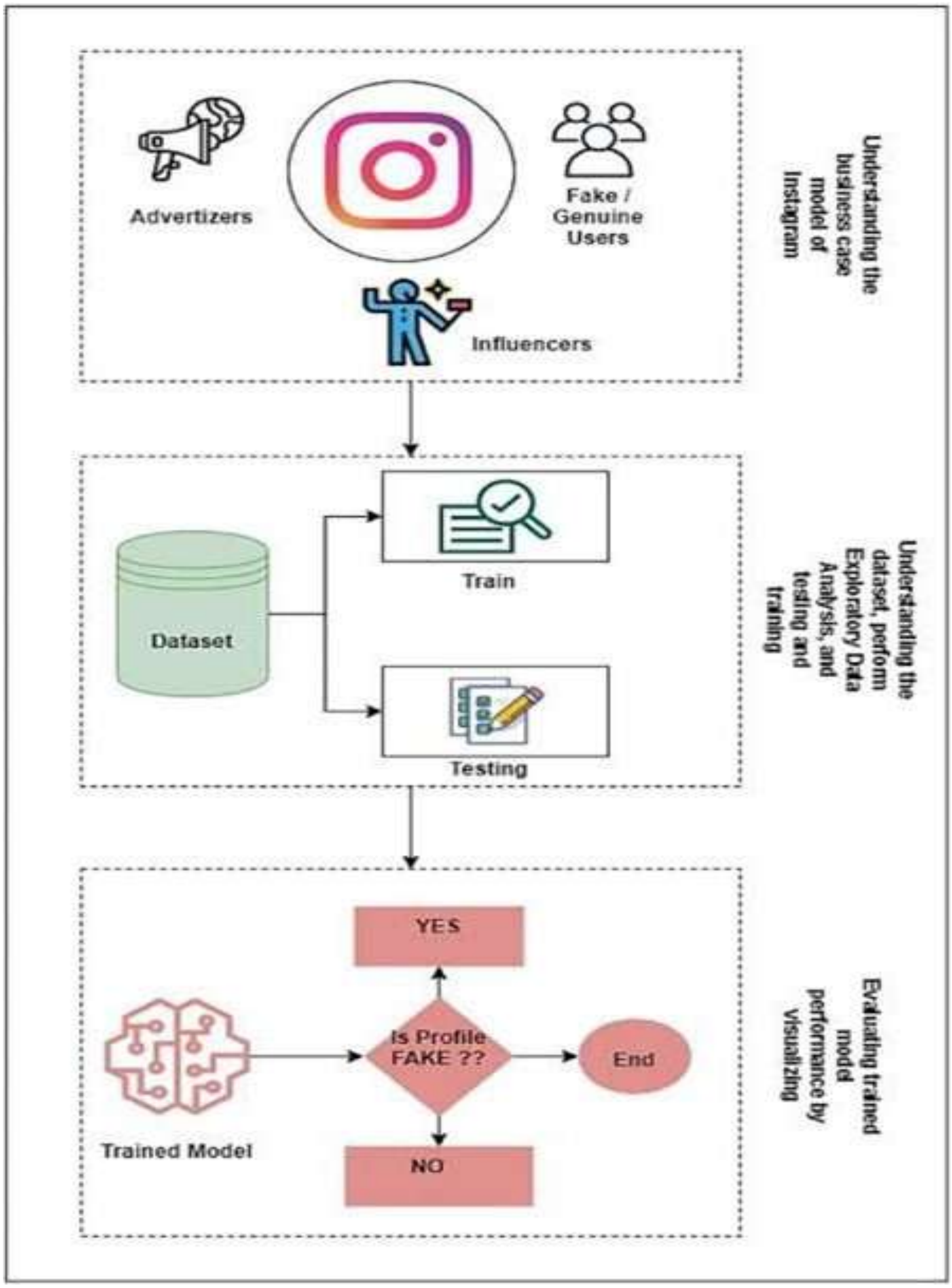

Advertizers
Fake / Genuine Users
Influencers
Understanding the business case model of Instagram
Dataset
Train
Testing
Understanding the dataset, perform Exploratory Data Analysis, and testing and training
YES
Is Profile FAKE ??
End
NO
Trained Model
Evaluating trained model performance by visualizing

Fig. 2. System Workflow

4.1. Setting up the Environment: The implementation phase began by selecting a suitable development environment to build and test the fake social media profile detection system. The team utilized Python as the core programming language due to its extensive support for machine learning, data analysis, and web application development. Initial experimentation and model training were conducted using platforms like Jupyter Notebook and Google Colab, which offered an interactive environment for data preprocessing, feature engineering, and visualization. For more advanced development and integration of the model with the user interface, Visual Studio Code was used as the main IDE. To support the system's functionality, a variety of Python libraries were installed and configured. Key machine learning libraries such as XGBoost and Scikit-learn were essential for model building and evaluation. Pandas and NumPy were used for data manipulation and numerical processing, while Matplotlib and Seaborn were implemented for visualizing data trends, performance metrics, and model comparisons. The imbalanced-learn library, specifically SMOTE, was included to address class imbalance in the dataset by generating synthetic samples of the minority class (fake profiles). For natural language preprocessing tasks such as cleaning bio text and identifying spam indicators, NLTK (Natural

Language Toolkit) was used. The data used for training the model was sourced from publicly available datasets on Kaggle, including collections like "Fake Instagram Profiles" and "Bot vs Human Twitter Accounts." These datasets contained useful metadata such as the number of followers and followings, the total number of posts, profile bios, hashtag patterns, and engagement behavior. Although this phase primarily relied on static datasets, the environment was also configured to support live data fetching via Twitter API and Instagram Graph API in future development, enabling real-time predictions and system updates.

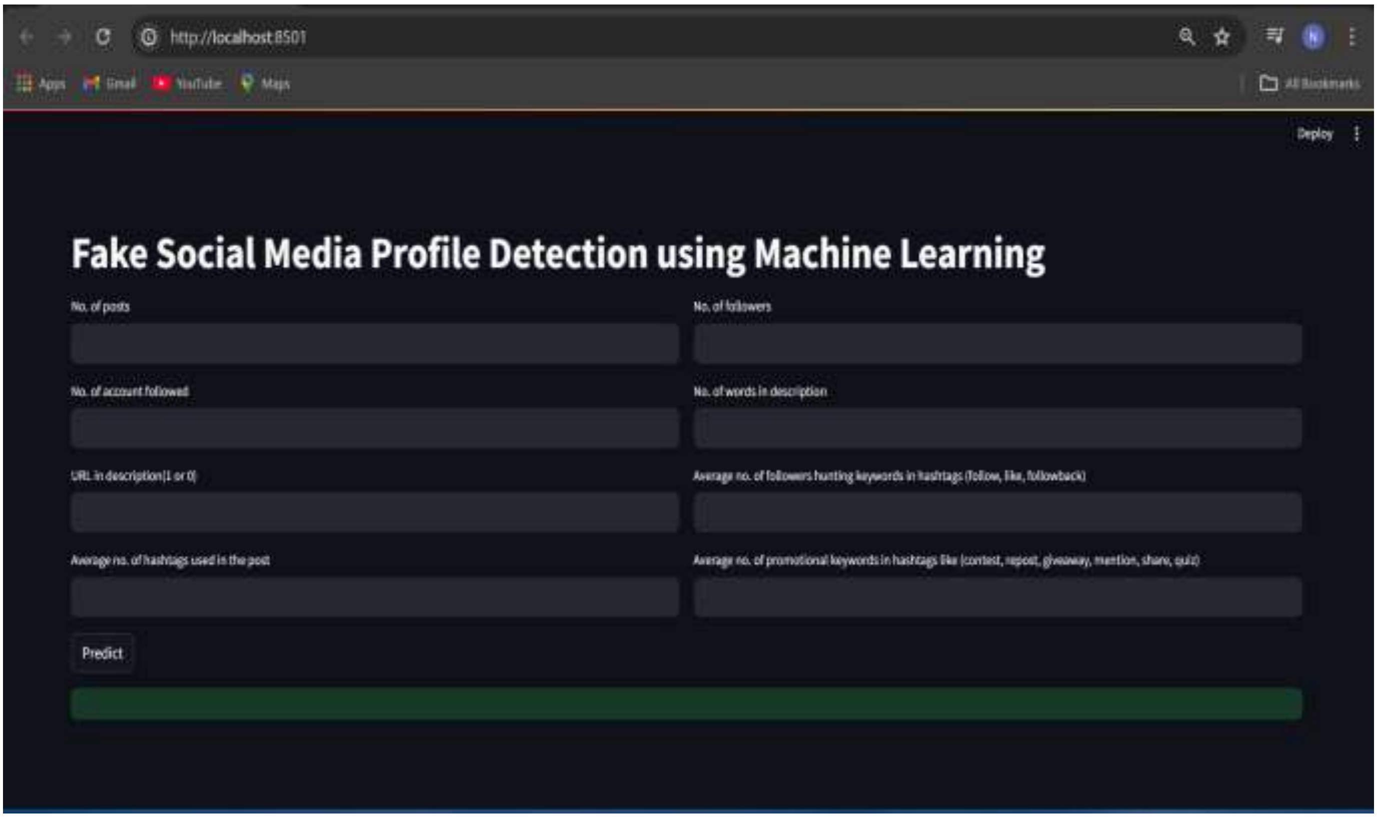
4.2. Implementation Details: After setting up the environment and preparing the datasets, the implementation began with a structured data preprocessing phase. The raw profile data collected from Kaggle, and other sources was first cleaned to ensure consistency and accuracy. Missing values were handled either by imputation or removal, and categorical fields such as “URL present” or “has profile picture” were encoded into binary format (1 or 0). Numerical data like follower count, number of posts, and hashtag counts were normalized to bring all features onto a similar scale. Text-based attributes such as bios and captions were cleaned by removing unwanted characters, emojis, hyperlinks, and repeated keywords using regular expressions and NLP tools like NLTK.

Following preprocessing, feature engineering was conducted to extract meaningful indicators of fake behavior. Several behavioral and content-based features were derived, such as the follower-to-following ratio, posting frequency, bio completeness, average number of hashtags per post, and the presence of promotional or suspicious keywords like “giveaway,” “follow back,” or “contest.” These features are commonly associated with fake profiles and significantly contribute to classification accuracy. The dataset was highly imbalanced, with real profiles far outnumbering fake ones. To solve this, SMOTE (Synthetic Minority Oversampling Technique) was applied to generate synthetic samples of fake profiles, creating a balanced dataset for training the model. The core model used in the system is the XGBoost classifier, chosen for its high performance in binary classification tasks and its efficiency in handling large, structured datasets. The model was trained on 80% of the dataset while the remaining 20% was used for validation. Hyperparameters such as learning rate, maximum tree depth, and number of estimators were tuned using grid search to optimize the model's performance. Once trained, the model was evaluated using key performance metrics including accuracy, precision, recall, and F1-score. A confusion matrix was also generated to visually assess the number of true positives, false positives, and misclassifications.

To make the system accessible to non-technical users, a web-based interface was developed using Streamlit, a lightweight Python framework for interactive applications. The dashboard allows users to manually input profile details such as the number of posts, followers, hashtags used, and bio description length. Upon submission, the data is passed to the backend XGBoost model, and the interface instantly returns whether the profile is likely to be real or fake. The goal of this implementation was not just to achieve high accuracy, but also to ensure the system is simple, responsive, and ready for deployment in real-world environments such as social media moderation tools or third-party verification platforms.

4.3. System Execution Details: Once the model was trained and integrated into the system, the execution phase involved testing the workflow of the detection system and evaluating how effectively it responds to real-world profile data. The core functionality was built into a web-based interface using Streamlit, allowing users or moderators to enter relevant profile attributes through a simple input form. These input fields included the number of posts, the number of followers and followings, the presence of a URL in the bio, the average number of hashtags used per post, and the frequency of suspicious keywords such as “giveaway,” “contest,” “likeforlike,” or “follow back.” Users could also input the number of words in the description or bio, which is another critical feature in differentiating between real

and fake accounts. When the user enters the profile details, the data is processed in real time by the backend model. The XGBoost classifier, trained on the labeled dataset with behavioral and content-based features, evaluates the inputs and predicts whether the profile is real or fake. The model then outputs the classification result along with a confidence score, providing a transparent and user-friendly prediction. This interaction is seamless and takes only a few seconds to return the result, making it suitable for use in live moderation environments or research tools. To provide visual feedback and support decision-making, the interface can display summary statistics and relevant prediction indicators. In future extensions, the system may include additional features such as user feedback submission, result explanations using tools like SHAP (Shapley Additive Explanations), and report generation for suspicious profiles. The combination of a reliable machine learning model with an intuitive frontend ensures that users can easily interact with the system, understand its outputs, and apply it for practical applications like flagging, reporting, or auditing suspicious social media accounts.



Fake Social Media Profile Detection using Machine Learning

No. of posts

No. of followers

No. of account followed

No. of words in description

URL in description(1 or 0)

Average no. of followers hunting keywords in hashtags (follow, like, followback)

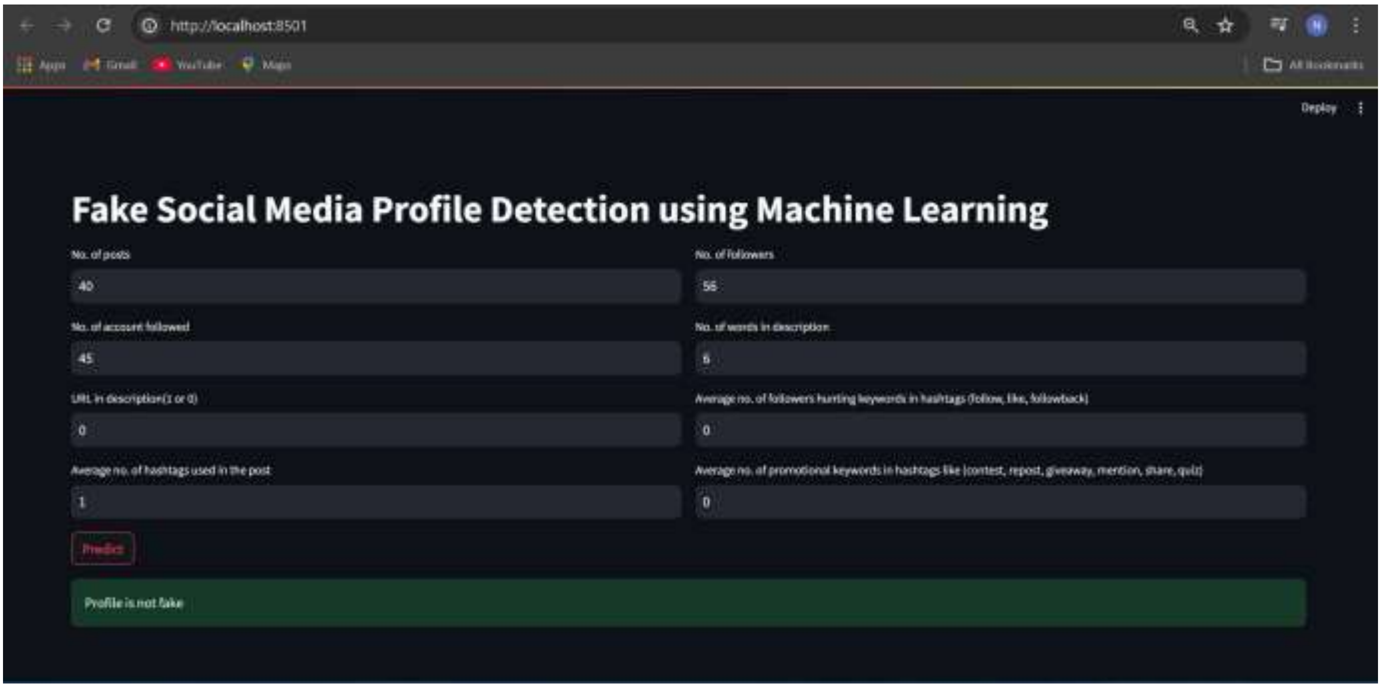
Average no. of hashtags used in the post

Average no. of promotional keywords in hashtags like (contest, repost, giveaway, mention, share, quiz)

Predict

Fig 3. Homepage

4.4 Result Analysis:



Fake Social Media Profile Detection using Machine Learning

No. of posts
40

No. of followers
56

No. of account followed
45

No. of words in description
6

URL in description(1 or 0)
0

Average no. of followers hunting keywords in hashtags (follow, like, followback)
0

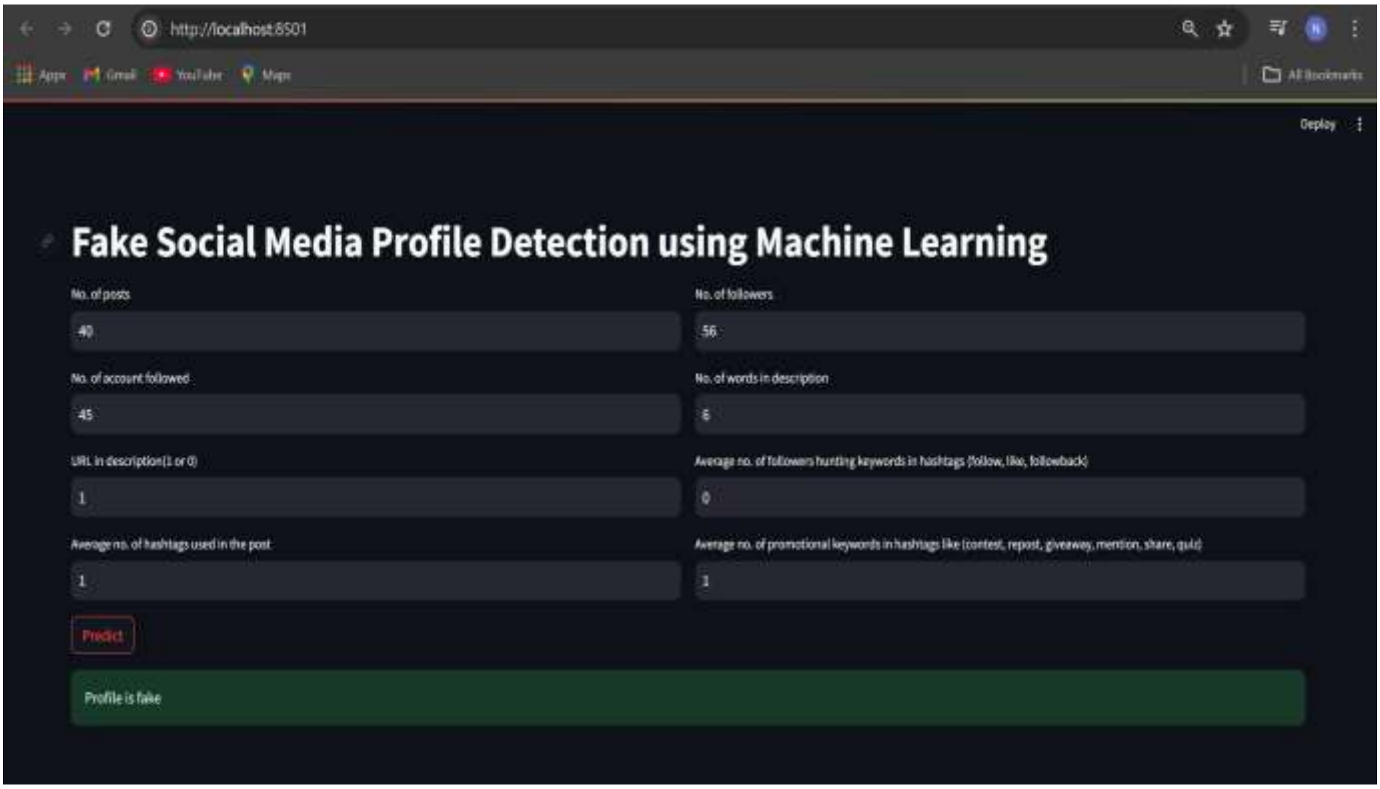
Average no. of hashtags used in the post
1

Average no. of promotional keywords in hashtags like (contest, repost, giveaway, mention, share, quiz)
0

Predict

Profile is not fake

Fig 4. Real Profile Prediction



Fake Social Media Profile Detection using Machine Learning

No. of posts: 40

No. of followers: 56

No. of account followed: 45

No. of words in description: 6

URL in description(1 or 0): 1

Average no. of followers hunting keywords in hashtags (follow, like, followback): 0

Average no. of hashtags used in the post: 1

Average no. of promotional keywords in hashtags like (contest, repost, giveaway, mention, share, quiz): 1

Predict

Profile is fake

Fig 5. Fake Profile Prediction

To spot fake social media profiles, we used an XGBoost classification model trained on behavioral and profile-related features. These features included things like the number of posts, how many accounts the profile follows and is followed by, whether there’s a URL in the bio, hashtag usage, and the presence of promotional or follower-hunting keywords.

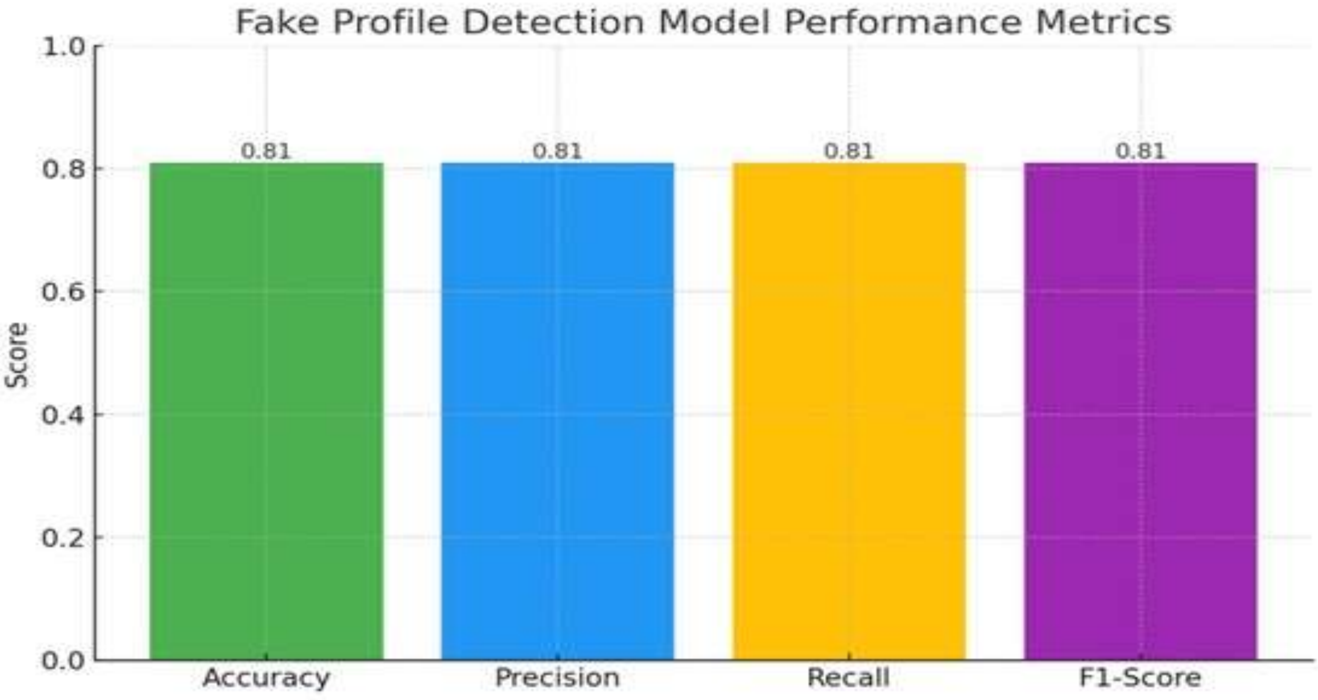


Fig 6. Performance Analysis Graph

Performance Metrics:

1. **Accuracy-score:** 0.8102 or 81.02%
2. **F1-score:** 81%
3. **Precision:**
   - Precision for Class 0 (Not Fake) = 0.86
   - Precision for Class 1 (Fake) = 0.75
   - Precision (Weighted) = 81%
4. **Recall:**
   - Recall for Class 0 (Not Fake) = 0.78
   - Recall for Class 1 (Fake) = 0.84
   - Recall (Weighted) = 81%

5. Opportunities and Challenges:

5.1. Opportunities: The proposed system offers several promising opportunities in the field of fake profile detection and digital safety. One of the major strengths is the high classification accuracy achieved through the XGBoost model, which effectively learns patterns from structured and imbalanced data. This accuracy is enhanced by the model's ability to highlight feature importance, allowing analysts to understand which features such as follower-to-following ratio, URL presence, or hashtag frequency are most indicative of fake behavior. The system is also well-suited for real-world deployment due to its lightweight design and fast inference capabilities, making it ideal for real-time use in social media moderation tools or external verification platforms. Additionally, the modularity of the system makes it scalable across various platforms beyond Twitter or Instagram. With minor adjustments to input features, it can be adapted to detect fake accounts on Facebook, LinkedIn, Reddit, and emerging platforms. By automating the detection process, the system significantly reduces the burden on manual moderators and allows for quicker, data-driven decision-making. The use of explainability tools like SHAP or LIME can further enhance trust by allowing users and administrators to understand how each prediction was made. Beyond detection, the system can provide valuable behavioral insights for researchers and analysts, uncovering trends in fraudulent activity and contributing to a safer and more transparent digital ecosystem.

5.2. Challenges: Despite its advantages, the system also presents certain challenges that must be addressed to ensure sustained performance and reliability. One of the primary challenges is obtaining high-quality, labeled datasets. Social media data is often restricted due to privacy policies and platform limitations, which can result in incomplete or inconsistent training samples. Moreover, fake profiles continuously evolve to avoid detection, making it necessary to frequently retrain the model with updated data to maintain its effectiveness. Another issue is the risk of false positives, where legitimate users might be incorrectly flagged as fake due to unique posting styles or limited activity. This could lead to user dissatisfaction or trust issues if not handled carefully. Integrating the model into a real-time environment also introduces technical complexity, especially when dealing with live data streams and concurrent user requests. While the current model handles structured data efficiently, it does not yet include advanced analysis of images or deep natural language processing, which could allow some sophisticated fake profiles to bypass detection. In addition, although user data is not stored, the system must still comply with global data protection regulations such as the GDPR and CCPA. Ensuring transparency while maintaining the model's complexity is another challenge, as interpreting decision logic in ensemble models like XGBoost is not always straightforward. Overcoming these challenges will be crucial for deploying the system at scale while preserving accuracy, fairness, and user trust.

6. Conclusion: The proliferation of fake social media profiles presents a significant challenge to the safety, integrity, and authenticity of online platforms. These accounts are often used to manipulate conversations, spread misinformation, conduct scams, or engage in identity theft. The research presented in this project introduces a machine learning-based system using the XGBoost algorithm for accurately detecting such fake profiles. By leveraging structured behavioral data such as follower-following ratios, posting frequency, hashtag patterns, and bio completeness, the system identifies suspicious accounts with a high degree of precision.

The model was trained on publicly available datasets and evaluated using key metrics like accuracy, precision, recall, and F1-score, all of which showed strong performance. The inclusion of SMOTE addressed class imbalance issues, ensuring the model could generalize well across both real and fake profiles. Additionally, a user-friendly web interface was developed using Streamlit to enable real-time predictions, making the system accessible for practical applications such as moderation, research, or social network audits. While the system performs effectively, it also emphasizes the importance of transparency, ethical AI usage, and data privacy, which are critical in any large-scale deployment.

Overall, the project demonstrates that with the right combination of feature engineering, model tuning, and usability considerations, machine learning can be a powerful tool in detecting and mitigating the spread of fake social media accounts. It reduces manual moderation efforts, enhances platform security, and contributes to a more trustworthy digital environment.

7. References:

- Partha Chakraborty, S. S. Rao, and K. S. Kuppusamy, “Fake Profile Detection Using Machine Learning Techniques,” Journal of Computer and Communications, vol. 10, pp. 74–87, 2022. [Online]. Available: https://www.scirp.org/pdf/jcc_2022102614142569.pdf
- E. Van Der Walt and J. H. P. Eloff, “Using Machine Learning to Detect Fake Identities: Bots vs Humans,” IEEE Access, vol. 6, pp. 6540–6549, 2018. [Online]. Available: https://repository.up.ac.za/handle/2263/63916
- S. Kudugunta and E. Ferrara, “Deep Neural Networks for Bot Detection,” Information Sciences, vol. 467, pp. 312–322, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025518304685
- J. Liu, X. Li, and Y. Zhang, “XGBoost-Based Fake Profile Detection on Social Media,” Proceedings of [Conference/Journal], 2020. [Online]. Available: https://www.researchgate.net/publication/353589259
- S. S. Rao, P. Chakraborty, and K. S. Kuppusamy, “A Hybrid Approach for Fake Profile Detection on Social Media using XGBoost and Random Forest,” International Journal of Engineering and Management Research, vol. 13, no. 3, pp. 265–270, June 2023. [Online]. Available: https://ijemr.vandanapublications.com/index.php/ijemr/article/download/1298/1131
- Sharma and A. Gupta, “Profile Imposter Detection on Instagram Using XGBoost and SVM,” Proceedings of the International Conference on Computational Intelligence and Emerging Technologies, pp. 24–29, 2024. [Online]. Available: : https://www.atlantis- press.com/article/126001987.pdf
- S. Kumar and R. Patel, “Fake Social Media Accounts and Their Detection,” International Journal for Multidisciplinary Research, vol. 7, no. 2, pp. 1–10, April 2025. [Online]. Available: https://www.ijfmr.com/papers/2025/2/40569.pdf
- M. A.Rahman and S. A. Smith, “Unmasking Fake Social Network Accounts with Explainable AI,” International Journal of Advanced Computer Science and Applications, vol. 15, no. 3, pp. 125–1352024. [Online]. Available: https://thesai.org/Downloads/Volume15No3/Paper_125-Unmasking_Fake_Social_Network_Accounts.pdf

- M. Shazan, M. Nahid, M. Ahmed, and P. Talukder, “Fake Profile Detection on Social Networking Sites Using XGBoost,” International Journal of Progressive Research in Engineering Management and Science, vol. 5, no. 4, pp. 498–501, April 2025. [Online]. Available: https://www.ijprems.com/uploadedfiles/paper//issue_4_april_2025/39554/final/fin_ijprems1744186361.pdf
- M. Shazan, M. Nahid, M. Ahmed, and P. Talukder, “Fake Profile Detection Using Machine Learning Techniques,” Journal of Computer and Communications, vol. 10, pp. 74–87, 2022. [Online]. Available: https://www.scirp.org/pdf/jcc_2022102614142569.pdf