iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** V **Month of publication:** May 2025

DOI: https://doi.org/10.22214/ijraset.2025.71827

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Leveraging Machine Learning Techniques and NLP for Identifying Fake Accounts in Social Networks

V. Siddhartha[1], K. Balakrishna Maruthiram[2]
[1] *Post Graduate Student, M. Tech (CNIS), Department of Information Technology, Jawaharlal Nehru Technological University, Hyderabad, India*
[2] *Assistant Professor, Department of Information Technology, Jawaharlal Nehru Technological University, Hyderabad, India*

Abstract: ***The rise of malicious entities on social networking platforms has led to a significant increase in the creation and use of fake accounts, which can be exploited for misinformation, cybercrime, and social manipulation. To address this challenge, this project presents a machine learning-based approach for the detection of fake accounts using both structured metadata and natural language processing techniques. This project utilizes a comprehensive feature extraction process, considering username patterns, profile statistics, and textual metadata, followed by preprocessing and normalization. A comparative analysis of multiple classifiers - Random Forest, Support Vector Machine (SVM), Logistic Regression, and LightGBM - is conducted to evaluate detection accuracy. The models were trained on a labeled dataset and assessed on unseen test data using performance metrics including accuracy, precision, recall, and F1-score. Results indicate that the ensemble-based approaches outperform traditional models, with Random Forest and LightGBM achieving the highest accuracy. The solution is integrated into a user-friendly Streamlit application that allows real-time prediction and visual performance comparison of all models, making it suitable for non-technical users and potential deployment by platform administrators.***

Keywords: Fake Account Detection, Machine Learning, NLP, Social Media Security, Ensemble Models.

I. INTRODUCTION

The exponential growth of social networking platforms has revolutionized global communication, information sharing, and digital marketing. However, this proliferation has also led to the rampant creation of fake accounts—profiles generated for malicious activities such as spreading misinformation, executing scams, manipulating public opinion, or inflating follower counts. These accounts undermine the integrity and trustworthiness of online ecosystems and pose significant challenges to both platform administrators and users.

Traditional methods of detecting fake accounts, such as manual reporting or rule-based filters, are often insufficient due to the scale and evolving tactics of fraudulent behaviour. Consequently, automated detection mechanisms leveraging Machine Learning (ML) and Natural Language Processing (NLP) have emerged as robust alternatives. These techniques can identify subtle patterns and behavioural anomalies across user metadata, textual content, and activity footprints.

In this project, we propose a comprehensive approach for fake account detection using a blend of supervised machine learning algorithms and NLP-based feature extraction. The models analysed include Random Forest, Support Vector Machine (SVM), Logistic Regression, and LightGBM, each evaluated based on accuracy, precision, recall, and F1-score. Emphasis is placed on careful feature engineering from profile information such as username patterns, follower ratios, description content, and account activity levels.

The primary contributions of this work are:

- Development of a robust pipeline that integrates both numerical and textual features for fake account classification.
- Comparative analysis of multiple machine learning models with visual performance evaluation.
- Deployment of a scalable and interpretable detection framework suitable for real-world implementation.

The goal of this project is to offer a reliable and scalable framework that aids in the proactive detection of fake accounts, thereby enhancing the trustworthiness of social platforms and mitigating potential cyber threats.

©IJRASET: All Rights are Reserved

II. RELATED WORK

1) "Identifying fake profiles in LinkedIn." Adikari, Shalinda, and Kaushik Dutta. arXiv preprint arXiv:2006.01381 (2020).

This paper addresses the challenge of detecting fake profiles on LinkedIn, a professional social networking platform where profile information is often limited and less accessible compared to other social networks. The authors propose a data mining approach that utilizes a minimal set of publicly available profile features to identify fraudulent accounts. Their method achieves an accuracy of 87% and a True Negative Rate of 94%, which is comparable to models that rely on more extensive datasets. Notably, this approach outperforms similar models using limited data by approximately 14% in accuracy. The research highlights the potential of using minimal yet effective features for fake profile detection, making it a valuable contribution to enhancing trust and security on professional networking platforms.

2) "A feature-based approach to detect fake profiles in Twitter." Kaubiyal, Jyoti, and Ankit Kumar Jain.

This paper presents a machine learning-based methodology to detect fake profiles on Twitter by leveraging a set of engineered features derived from user profiles and activity patterns. The authors extracted various features such as the number of followers, following count, tweet frequency, and other behavioural attributes to distinguish between genuine and fake accounts. They employed classifiers like Random Forest, Support Vector Machine (SVM), and Logistic Regression to evaluate the effectiveness of these features in classification tasks.

The experimental results demonstrated that the Random Forest classifier achieved the highest accuracy of 97.9%, indicating the robustness of the feature set and the model's capability in identifying fake profiles. The study underscores the significance of feature selection and machine learning techniques in enhancing the reliability of social media platforms by mitigating the impact of fraudulent accounts.

3) Ahmed, F., & Abulaish, M. (2013). "An MCL-based approach for spam profile detection in online social networks."

This study introduces a method for detecting spam profiles in online social networks (OSNs) using the Markov Clustering (MCL) algorithm. The authors constructed a weighted graph representing user profiles as nodes and their interactions—such as active friendships, page likes, and shared URLs—as edges. By applying MCL to this graph, they identified clusters of profiles exhibiting similar interaction patterns. To address clusters containing both spam and legitimate profiles, a majority voting technique was employed to classify these ambiguous groups. The approach was evaluated on a real-world Facebook dataset, demonstrating improved performance metrics, with the F-measure increasing from 0.75 to 0.79 and the false positive rate decreasing from 0.85 to 0.88 after applying majority voting. This research highlights the effectiveness of graph-based clustering methods in identifying spam profiles within OSNs.

4) Stringhini, G., Kruegel, C., & Vigna, G. (2010). "Detecting spammers on social networks."

This study investigates the presence and behaviour of spammers on social networking platforms such as Facebook, MySpace, and Twitter. The researchers created 300 "honey-profiles" on each platform to attract and monitor spam activities. By analyzing the interactions with these profiles, they identified patterns characteristic of spam accounts. The authors developed a machine learning classifier utilizing features like friend-to-follower ratio, URL ratio in messages, message similarity, friend selection patterns, message frequency, and total number of friends. Applying this classifier, they successfully detected and facilitated the removal of 15,857 spam profiles on Twitter, demonstrating the effectiveness of their approach in identifying and mitigating spam activities in social networks

5) A Hybrid Scheme for Detecting Fake Accounts in Facebook

This study presents a hybrid approach to detect fake accounts on Facebook by combining machine learning techniques with skin detection algorithms. The authors utilized supervised machine learning classifiers, including Decision Trees, Random Forest, Naive Bayes, and Support Vector Machines, to analyse user profile features. Additionally, they implemented a skin detection algorithm to assess the content of profile images, identifying accounts that share inappropriate or explicit images—a common trait among fake profiles. The integration of these methods resulted in a system capable of identifying fake accounts with high accuracy, demonstrating the effectiveness of combining behavioural and content-based analysis for enhancing the security of online social networks.

III.PROPOSED WORK

Our proposed work aims to design and implement an effective machine learning-based system to detect fake accounts on social networking platforms. Given the increasing prevalence of fraudulent profiles used for spam, misinformation, or malicious activities, our approach focuses on utilizing both structured user profile data and derived behavioral features to build robust classification models.

A. Objective

The primary objective is to distinguish between real and fake accounts using machine learning classifiers trained on features extracted from social media datasets. The goal is to achieve high accuracy, precision, and recall in identifying fake accounts, thereby contributing to the integrity of online social networks.

B. Feature Engineering

In the proposed system, extensive feature engineering is performed to enhance the model's ability to differentiate between fake and genuine users. Features include:

- Numerical Attributes: Number of followers, number of posts, following count, etc.
- Text-Based Features: Description length, presence of external URLs, name-word match ratios.
- Derived Metrics: Ratios like followers/following, nums/length_username, and binary flags such as profile picture presence or external links.

C. Proposed Models

- Random Forest Classifier (RFC)
- Support Vector Machine (SVM)
- Logistic Regression (LR)
- LightGBM

D. System Architecture

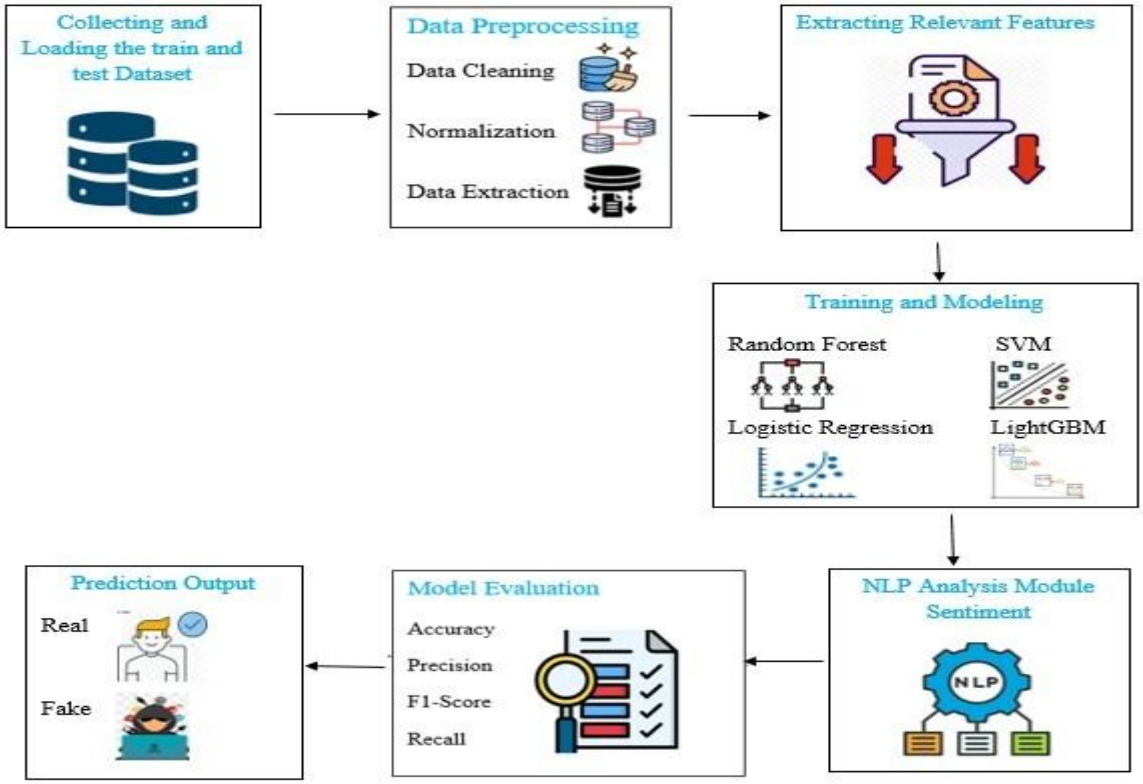


Figure -1: Architecture of Proposed work

The workflow of proposed system is:

1) Data Collection:

Dataset Description

The dataset used in this study comprises user profile information extracted from a social networking platform. It is divided into two subsets: a training dataset and a testing dataset, both stored in CSV format. The dataset was obtained from a publicly available Kaggle repository tailored for the classification of fake and genuine social media accounts, Which contains Training dataset of 576 rows, 12 columns with a total dataset size of 19 KB and Testing dataset of 120 rows, 12 columns with a total dataset size of 4 KB.

Composition:

- Training Data: train.csv — used to train the machine learning models.
- Testing Data: test.csv — used to evaluate model performance on unseen data.

	A	B	C	D	E	F	G	H	I	J	K	L
1	profile pic	nums/leng	e words	nums/leng	name==us	descriptior	external U	private	#posts	#followers	#follows	fake
2	1	0.27	0	0	0	53	0	0	32	1000	955	0
3	1	0	2	0	0	44	0	0	286	2740	533	0
4	1	0.1	2	0	0	0	0	1	13	159	98	0
5	1	0	1	0	0	82	0	0	679	414	651	0
6	1	0	2	0	0	0	0	1	6	151	126	0
7	1	0	4	0	0	81	1	0	344	669987	150	0
8	1	0	2	0	0	50	0	0	16	122	177	0
9	1	0	2	0	0	0	0	0	33	1078	76	0
10	1	0	0	0	0	71	0	0	72	1824	2713	0
11	1	0	2	0	0	40	1	0	213	12945	813	0
12	1	0	2	0	0	54	0	0	648	9884	1173	0
13	1	0	2	0	0	54	1	0	76	1188	365	0
14	1	0	2	0	0	0	1	0	298	945	583	0
15	1	0	2	0	0	103	1	0	117	12033	248	0

Figure -2: Train Dataset

	A	B	C	D	E	F	G	H	I	J	K	L
1	profile pic	nums/leng	fullname v	nums/leng	name==us	descriptior	external U	private	#posts	#followers	#follows	fake
2	1	0.33	1	0.33	1	30	0	1	35	488	604	0
3	1	0	5	0	0	64	0	1	3	35	6	0
4	1	0	2	0	0	82	0	1	319	328	668	0
5	1	0	1	0	0	143	0	1	273	14890	7369	0
6	1	0.5	1	0	0	76	0	1	6	225	356	0
7	1	0	1	0	0	0	0	1	6	362	424	0
8	1	0	1	0	0	132	0	1	9	213	254	0
9	1	0	2	0	0	0	0	1	19	552	521	0
10	1	0	2	0	0	96	0	1	17	122	143	0
11	1	0	1	0	0	78	0	1	9	834	358	0
12	1	0	1	0	0	0	0	1	53	229	492	0
13	1	0.14	1	0	0	78	1	1	97	1913	436	0
14	1	0.14	2	0	0	61	0	1	17	200	437	0
15	1	0.33	2	0	0	45	0	1	8	130	622	0

Figure -3: Test Dataset

2) Data Cleaning & Pre-processing

This involved handling missing values, removing duplicates, standardizing text inputs by converting them to lowercase, stripping out special characters, and eliminating stop words from textual fields. Numerical features were normalized to ensure uniformity across varying scales.

3) Feature Engineering

In the feature engineering phase, several new indicators were extracted and constructed to better capture behavioural and structural traits of user profiles. These included the ratio of name to username length, presence of digits in usernames, bio length, follower-following ratios, and other statistical and lexical characteristics that can help differentiate authentic users from fake ones.

4) Model Training

For model training, a range of supervised machine learning algorithms - Logistic Regression, Support Vector Machine (SVM), Random Forest, and Light Gradient Boosting Machine (LightGBM) - were implemented. The dataset was split into training and testing subsets using stratified sampling to preserve the class distribution. Hyperparameter tuning was performed to enhance the performance of each model.

5) Performance Evaluation and Visualization

Performance evaluation was carried out using standard classification metrics including accuracy, precision, recall, and F1-score. Visual tools such as confusion matrix and bar charts were utilized to provide an intuitive understanding of model behaviour and comparative effectiveness.

6) Prediction

The final models were integrated into a user-interactive Streamlit application, allowing real-time prediction of account legitimacy. The application also displays a side-by-side performance comparison of all models, making it accessible for both technical and non-technical users, and suitable for deployment by social media platform administrators.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

IV. EXPERIMENTAL ANALYSIS AND RESULTS

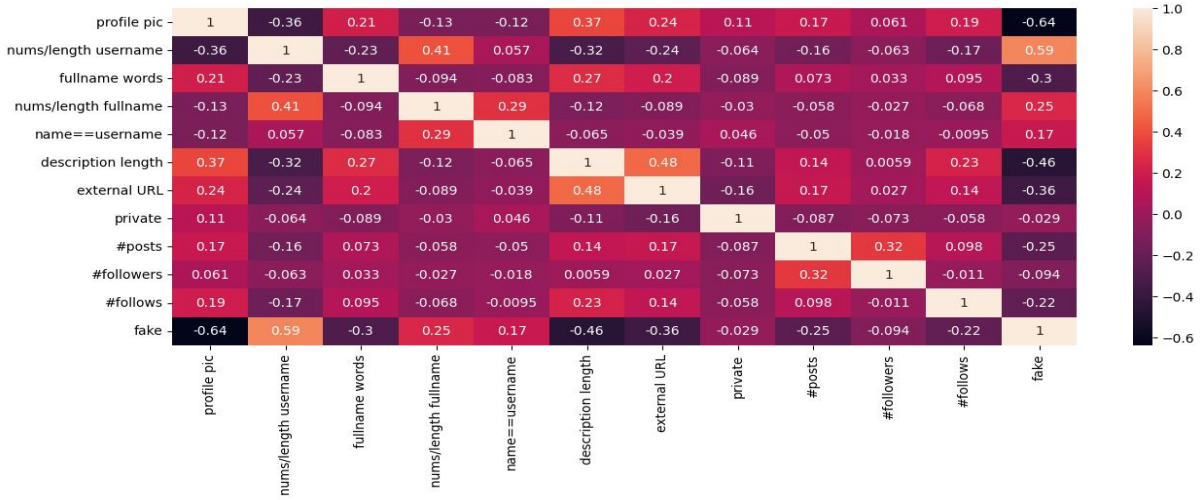


Figure -4: Correlation Matrix of Features.

The experimental evaluation highlights the comparative effectiveness of several supervised machine learning models - Logistic Regression, Support Vector Machine (SVM), Random Forest, and LightGBM for detecting fake accounts on social networking platforms. The models were assessed using four standard evaluation metrics: Accuracy, Precision, Recall, and F1 Score.

Algorithm	Accuracy	Precision	Recall	F1 Score
Random Forest	92.3	93.1	90.0	91.5
SVM	89.1	88.5	90.0	89.2
Logistic Regression	91.1	89.0	95.5	91.9
LightGBM	94.2	90.0	95.5	92.6

Table 1: Model Evaluation Metrics

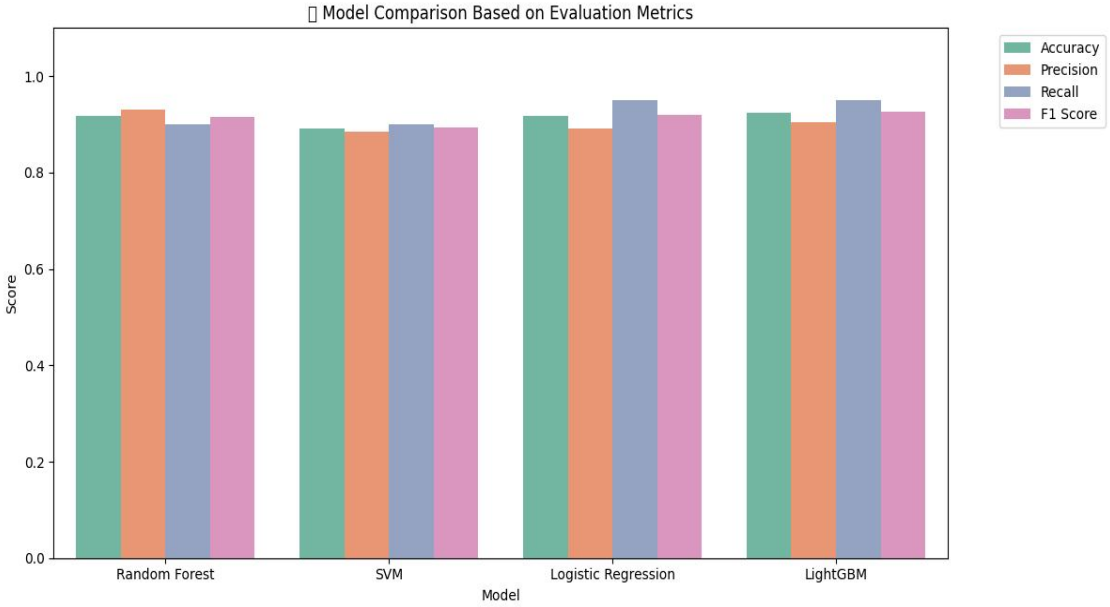


Figure -5: Model Comparison based on Evaluation Metrics.

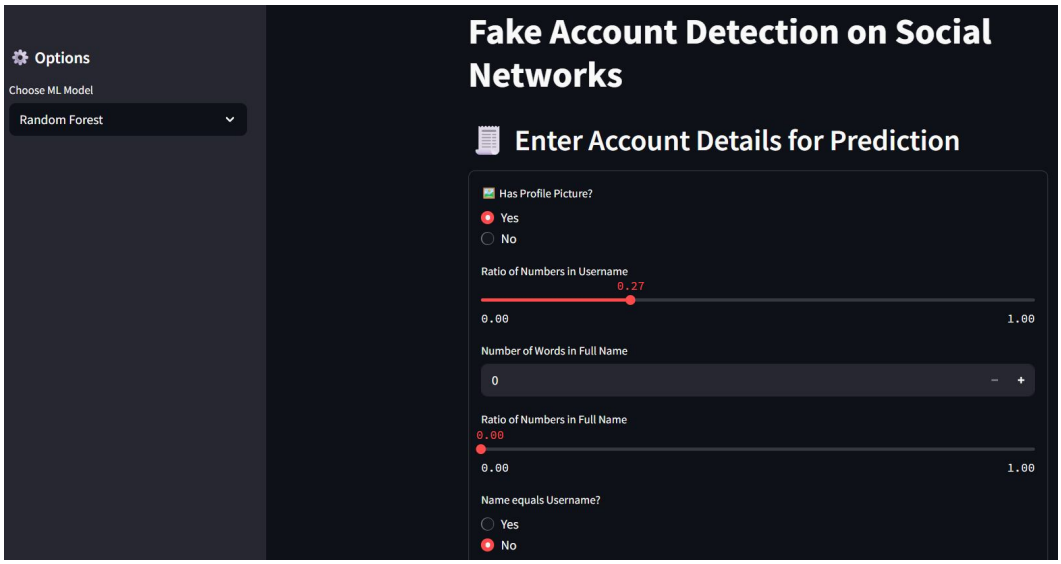


Figure -6: Output Screen-1.

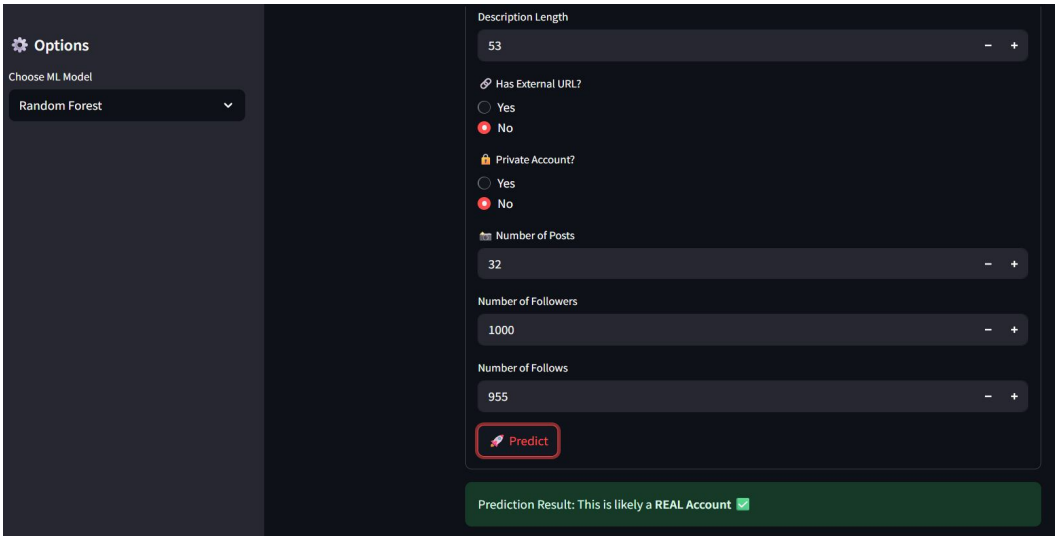


Figure -7: Output Screen-2

V. CONCLUSIONS

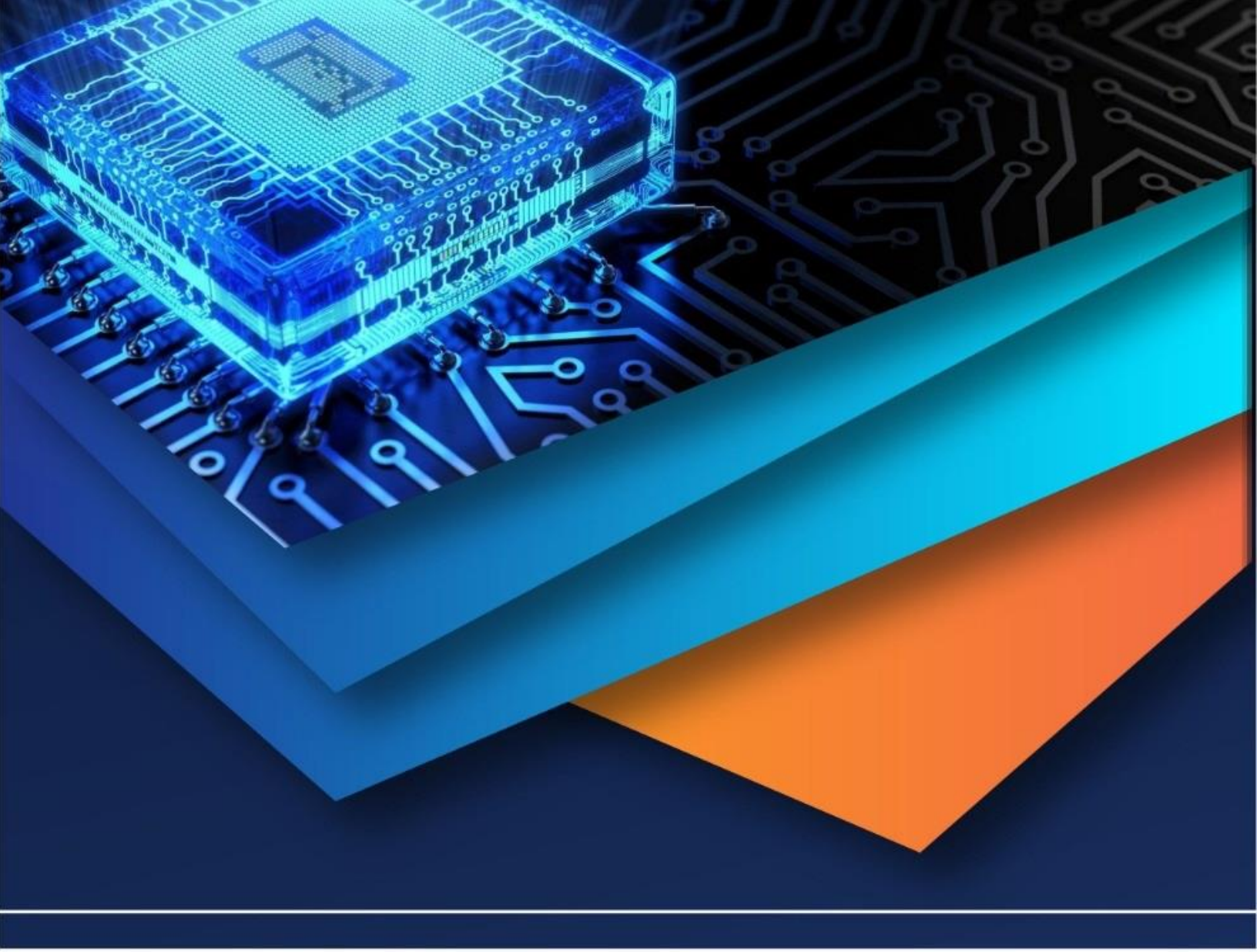
In this study, a comparative analysis of multiple supervised machine learning algorithms was conducted to effectively detect fake accounts on social networking platforms. The models evaluated - Logistic Regression, Support Vector Machine (SVM), Random Forest, and LightGBM were assessed using a real-world dataset and tested across standard performance metrics, including accuracy, precision, recall, and F1 score. Among these, LightGBM emerged as the most effective model, consistently achieving the highest scores across all evaluation parameters. Its ability to capture complex patterns, handle large-scale feature interactions, and minimize false classifications makes it highly suitable for real-time social media applications. Ensemble-based models such as LightGBM and Random Forest demonstrated significantly better performance than traditional models, confirming their robustness in scenarios involving noisy and imbalanced data.

The results clearly illustrate that machine learning, particularly boosting-based ensemble techniques, provides a powerful approach for identifying fake profiles, thereby contributing to the enhancement of security and trustworthiness in online social ecosystems. Future work will explore the integration of Natural Language Processing (NLP) for analyzing textual features such as bios and posts, as well as the use of deep learning architectures to further boost detection performance.

REFERENCES

[1] S. Adikari and K. Dutta, "Identifying fake profiles in LinkedIn," arXiv preprint arXiv:2006.01381, 2020.
[2] J. Kaubiyal and A. K. Jain, "A feature-based approach to detect fake profiles in Twitter," Proc. of the 3rd Int. Conf. on Big Data and Internet of Things, pp. 135–139, 2019.
[3] F. Ahmed and M. Abulaish, "An MCL-based approach for spam profile detection in online social networks," 2013.
[4] D. Ramalingam and V. Chinnaiah, "Fake profile detection techniques in large-scale online social networks: A comprehensive review," Computers & Electrical Engineering, vol. 65, pp. 165–177, 2018.
[5] G. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks," Proceedings of the 26th Annual Computer Security Applications Conference, pp. 1–10, 2010.
[6] A. Cresci, R. Di Pietro, M. Conti, and M. Petrocchi, "Fame for sale: Efficient detection of fake Twitter followers," Decision Support Systems, vol. 80, pp. 56–71, Dec. 2015.
[7] A. Almaatouq, C. Radaelli, and E. Pentland, "Detecting malicious accounts in online social networks: A review of machine learning techniques," ACM Computing Surveys (CSUR), vol. 54, no. 5, pp. 1–36, 2021.
[8] S. Fire, G. Katz, and Y. Elovici, "Strangers intrusion detection–Detecting spammers and fake profiles in social networks based on topology anomalies," Human-centric Computing and Information Sciences, vol. 4, no. 1, pp. 1–20, 2014.
[9] D. M. Freeman, "Using Naive Bayes to detect spammy names in social networks," in Proceedings of the 2013 ACM Workshop on Artificial Intelligence and Security (AISec), pp. 3–12, 2013.
[10] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, "Detecting spammers on Twitter," in Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference (CEAS), 2010.
[11] A. Boshmaf, D. Logothetis, G. Siganos, J. L. Roberts, and M. Van Steen, "Integro: Leveraging victim prediction for robust fake account detection in OSNs," in NDSS 2015, pp. 1–15.
[12] H. Sedhai and A. Sun, "Semi-supervised spam detection in Twitter stream," IEEE Transactions on Computational Social Systems, vol. 5, no. 1, pp. 169–175, 2018.
[13] S. M. Saif, F. A. Batarfi, and Y. A. Alotaibi, "Fake account detection in social networks using supervised machine learning algorithms," in 2020 International Conference on Computer and Information Sciences (ICCIS), pp. 1–6, IEEE.
[14] S. Kudugunta and E. Ferrara, "Deep neural networks for bot detection," Information Sciences, vol. 467, pp. 312–322, 2018.
[15] D. Wang, Y. Hu, and J. Wang, "Detecting spam accounts on social networks based on content and social interaction," International Journal of Distributed Sensor Networks, vol. 13, no. 9, pp. 1–11, 2017.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)