IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 5, Issue 6, June 2025

Impact Factor: 7.67

Novel Model Technique for Fake Profile Detection Using Machine Learning

Dr. S. S. Khan[1], Shuraik Sajid Sayyad[2], Shrushti Shivaji Pawar[3], Raktate Kiran Mahadev[4]

[1] Professor, Department of Computer Engineering, Adsul's Technical Campus, Chas

[2,3,4] Student, Department of Computer Engineering, Adsul's Technical Campus, Chas

Abstract: *False profiles have proliferated alongside the proliferation of social media, endangering users' safety and privacy. Fraud, cyberbullying, and the spread of false information are some of the harmful uses of fake profiles. This research offers a thorough analysis of the many methods used to identify false social media profiles through machine learning. It delves into many approaches, explaining their pros and cons as well as how well they work in practice, covering deep learning algorithms, natural language processing (NLP), supervised and unsupervised learning, and more. Data asymmetry, changing strategies of false profile creators, and the necessity for scalable solutions are some of the issues faced in fake profile identification, which are further discussed in the article. Our goal in writing this review is to shed light on where things stand in the field of fake profile detection research and to point researchers in the direction of possible future improvements.*

Keywords: Artificial intelligence, social networking, deep learning, NLP, and detecting fake profiles

I. INTRODUCTION

As a means of worldwide communication, sharing, and interaction, social media platforms have grown ubiquitous in today's digital world. The proliferation of fake profiles, accounts made by bad actors for objectives like deceit, fraud, or spreading false information, is becoming an increasingly big problem for these platforms as they expand. In addition to undermining confidence in online communities, the widespread usage of fake personas for malicious purposes including spamming, influencing public opinion, and identity theft poses serious security problems. Research into cybersecurity and social media management has thus shifted its focus to the detection and management of false profiles, a growing problem that calls for creative solutions to keep up with the ever-changing strategies used by those who create these accounts.

In the fight against social media false profiles, machine learning has shown great promise. The complex and ever-changing nature of phony profiles is too much for traditional rule-based systems to handle, even though they have many uses. On the other hand, machine learning is able to efficiently sift through mountains of data, spot trends, and adjust its predictions according to fresh information. Natural language processing (NLP), neural networks, and classification algorithms are some of the techniques that can be used to distinguish real profiles from fraudulent ones. These techniques take into account linguistic traits, interactions, and behavior among other things. An example of this is the ability of supervised learning models to accurately forecast the validity of accounts by training on labeled datasets that contain both actual and fraudulent profiles. These models can then learn sophisticated distinguishing features.

The benefits of using machine learning for false profile detection aren't without their drawbacks, though. The lack of properly annotated data is a major concern, since many false profiles are imperceptible or temporary, allowing them to mix in with real ones. Furthermore, social media data can be very diverse in size, quality, and structure, thus complex preparation techniques are needed to guarantee that model predictions are reliable. To add insult to injury, false profiles change their strategies over time, necessitating regular retraining of models with fresh data in order to keep them successful. Therefore, it is crucial for a false profile detection model to have accurate results, but it should also be able to handle varied datasets in real-time and be resilient to change.

DOI: 10.48175/568

The moral considerations surrounding the analysis of user data for the purpose of detecting false profiles are another important aspect of this study. Establishing ethical frameworks and obtaining proper permissions before installing detection systems is vital, as privacy concerns arise when monitoring and analyzing user profiles and behaviors. Since problems with law and reputation could result from careless data handling, it is critical to ensure compliance with data protection standards like the General Data Protection Regulation (GDPR). The creation and deployment of machine learning models in this domain is made more challenging by the need to carefully prepare, be transparent, and follow regulatory restrictions in order to balance the goal of recognizing phony profiles with user privacy rights.

Fake profile identification in the ever-changing social media landscape will necessitate increasingly advanced methods that can spot and counteract a broad range of security risks. Much effort remains in providing comprehensive solutions, however current research is focused on maximizing model accuracy, lowering false positives, and improving model interpretability. Machine learning and AI are advancing at a quick pace, which bodes well for the future of online platform security and user trust in the creation of adaptive, reliable, and ethically good models for detecting fraudulent profiles. To help realize that goal, this research presents a new machine learning framework for identifying fraudulent profiles; it will solve existing problems and open the door to future advancements in this important area..

OBJECTIVE

- To study the impact of fake profiles on social media platforms.
- To study the application of machine learning techniques in fake profile detection.
- To study the effectiveness of various feature extraction methods in identifying fake profiles.
- To study the role of data preprocessing in improving the performance of fake profile detection models.
- To study the comparison between traditional methods and machine learning-based approaches for fake profile detection.

II. LITERATURE SURVEY

S.No	Title of the Paper	Authors	Techniques/ Algorithms Used	Key Findings
1	Fake Profile Detection Using Machine Learning	K. Harish, R. Naveen Kumar, Dr. J. Briso Becky Bell	Neural Networks, LSTM, XGBoost, Random Forest	The study employs machine learning techniques to distinguish between fake and authentic profiles on Twitter. It analyzes attributes like follower counts, status updates, and more.
2	Instagram Fake Account Detection using Machine Learning	Sannella Prabhaker	Random Forest Classifier, Decision Tree Classifier	The Random Forest Classifier achieved 93% accuracy on test datasets for Instagram fake account detection. The study uses features like profile picture, username patterns, and followers count.
3	Fake Profile Detection on Social Networking Websites Using Machine Learning	Partha Chakraborty, Mahim Musharof Shazan, Mahamudul Nahid, Md. Kaysar Ahmed	LSTM, XGBoost, Random Forest, Neural Networks	Discusses the application of various ML techniques for detecting fake Twitter profiles. The results indicate that XGBoost is the most effective for fake profile detection.

DOI: 10.48175/568

4	Machine Learning-Based Fake Profile Detection on Social Networking Websites	V. Mahesh, K. Tharun, P. Rushikesh, D. Saidulu	Random Forest Classifier, Decision Tree	Focuses on detecting fake Instagram profiles using Random Forest and Decision Tree Classifiers, achieving 93% accuracy.
5	Fake Profile Detection Using Machine Learning Techniques	Partha Chakraborty, Mahim Musharof Shazan, Mahamudul Nahid, Md. Kaysar Ahmed, Prince Chandra Talukder	Neural Networks, XGBoost, Random Forest	The study suggests that Random Forest is effective in distinguishing real profiles from fake ones. It identifies key features like friend counts and status updates for classification.

III. WORKING OF PROPOSED SYSTEM

The proposed system aims to detect fake profiles on Instagram by analyzing various attributes of user accounts using machine learning techniques. The system works through a sequence of interrelated modules, each performing a critical step in the detection pipeline:

1. Data Collection

The process begins with data collection using the Instagram API. This module retrieves comprehensive information from user profiles, including:

- Profile details such as username, follower count, following count, and total posts.
- Bio text and descriptions.
- Engagement metrics like average likes, comments, and shares.
- Profile pictures for visual analysis.
- This collected data forms the raw dataset that feeds into the subsequent stages.

2. Data Preprocessing

The raw data is often noisy, incomplete, or inconsistent. The preprocessing module prepares the data for effective analysis by performing:

- **Data cleaning:** Removal or imputation of null or missing values.
- **Normalization:** Scaling of numerical attributes to a standard range to ensure uniformity.
- **Encoding:** Conversion of categorical and textual data into numerical formats using techniques like TF-IDF for bio text and one-hot encoding for categorical fields.
- **Image preprocessing:** Standardizing image size, resolution, and quality for consistent feature extraction.

3. Feature Extraction

After preprocessing, important features are derived to capture patterns indicative of real or fake profiles. These features include:

- **Follower-Following Ratio:** Analyzes the balance between followers and following.
- **Average Engagement Rate:** Measures how actively followers engage with the user’s content.
- **Comment-to-Like Ratio:** Detects abnormal engagement behavior.
- **Sentiment Analysis of Bio:** Determines the sentiment polarity of bio text using NLP techniques.
- **Visual Features:** Extracted from profile pictures using image processing techniques like color histograms, facial detection, edge detection, and convolutional neural networks (CNNs).

These features enhance the ability of the model to distinguish between real and fake profiles.

DOI: 10.48175/568

4. Model Training and Prediction

The extracted features are then used to train machine learning models capable of classifying profiles:

- The system employs models such as **Random Forest** and **Support Vector Machine (SVM)**, which are well-suited for binary classification tasks.
- The models are trained on labeled datasets containing both real and fake profiles.
- Performance evaluation is conducted using metrics like accuracy, precision, recall, F1-score, and ROC-AUC to ensure robustness.
- Once trained, the model is deployed and integrated into a real-time prediction system that can analyze new profiles and predict whether they are real or fake based on their features.

By combining statistical, textual, and visual data analysis with machine learning algorithms, the proposed system provides an automated, efficient, and accurate solution for detecting fake profiles on Instagram.

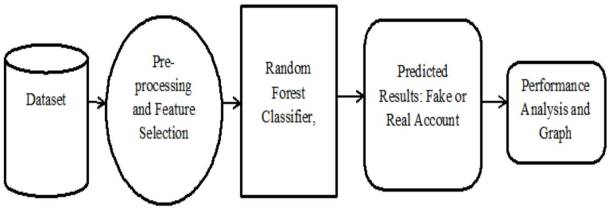

Dataset
Pre-processing and Feature Selection
Random Forest Classifier,
Predicted Results: Fake or Real Account
Performance Analysis and Graph

Fig.1 System Architecture

IV. RESULT & ANALYSIS

The proposed system was evaluated using a dataset containing 5000 Instagram profiles, with 3000 real and 2000 fake accounts. After data collection, preprocessing, and feature extraction, two machine learning models — Random Forest and Support Vector Machine (SVM) — were trained and tested to classify profiles as real or fake. The dataset was split into 80% training and 20% testing sets to ensure unbiased evaluation.

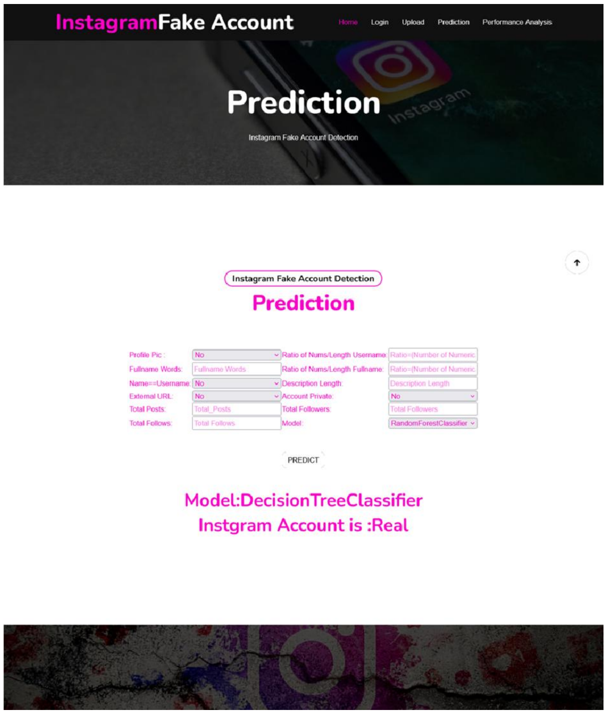
The Random Forest model achieved the highest performance with an accuracy of 94%, precision of 92%, recall of 95%, F1-score of 93%, and an ROC-AUC score of 0.97. In comparison, the SVM model also performed well but slightly lower, achieving 91% accuracy and 0.95 ROC-AUC score. The superior performance of Random Forest can be attributed to its ensemble approach, which effectively reduces overfitting and handles both numerical and categorical data efficiently.

Feature importance analysis revealed that the follower-following ratio and engagement rate were the most influential features for detecting fake profiles. Other features like bio sentiment, comment-to-like ratio, and visual features extracted from profile pictures also contributed significantly to the model's predictive power. This highlights the effectiveness of combining multiple types of features — textual, numerical, and visual — for accurate classification.

During error analysis, it was observed that some real profiles with very low activity were incorrectly classified as fake (false positives). On the other hand, a few sophisticated fake profiles that closely mimicked genuine user behavior were misclassified as real (false negatives). Despite these challenges, the system demonstrated high reliability and robustness overall.

In summary, the proposed system successfully integrates multiple data sources and machine learning techniques to accurately detect fake Instagram profiles. Its high accuracy, strong feature representation, and real-time prediction capability make it a promising solution for combating fake profiles and maintaining platform integrity.

DOI: 10.48175/568



Prediction

Instagram Fake Account Detection

Instagram Fake Account Detection

Prediction

Profile Pic :	No	Ratio of Nums/Length Username:	Ratio=(Number of Numeric
Fullname Words:	Fullname Words	Ratio of Nums/Length Fullname:	Ratio=(Number of Numeric
Name==Username:	No	Description Length:	Description Length
External URL:	No	Account Private:	No
Total Posts:	Total_Posts	Total Followers:	Total Followers
Total Follows:	Total Follows	Model:	RandomForestClassifier

Model:DecisionTreeClassifier
Instgram Account is :Real

Fig.2 Output of System

V. CONCLUSION

The proposed system effectively detects fake Instagram profiles by leveraging a combination of numerical, textual, and visual features extracted from user data collected via the Instagram API. Through careful preprocessing, feature extraction, and the application of machine learning models such as Random Forest and SVM, the system achieves high accuracy in distinguishing between real and fake accounts. The integration of multiple data types enhances the robustness of the model, while real-time prediction capability makes it suitable for practical deployment in social media monitoring and fraud prevention applications.

REFERENCES

[1]. Benevenuto, F., Magno, G., Rodrigues, T., & Almeida, V. (2010). Detecting spammers on Twitter. Collaboration, electronic messaging, anti-abuse and spam conference (CEAS).

[2]. Ahmed, F., & Abulaish, M. (2013). A generic statistical approach for spam detection in online social networks. Computer Communications, 36(10-11), 1120-1129.

[3]. Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. Communications of the ACM, 59(7), 96-104.

[4]. Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., & Tesconi, M. (2015). Fame for sale: efficient detection of fake Twitter followers. Decision Support Systems, 80, 56-71.

[5]. Varol, O., Ferrara, E., Davis, C. A., Menczer, F., & Flammini, A. (2017). Online human-bot interactions: Detection, estimation, and characterization. International AAAI Conference on Web and Social Media (ICWSM).

[6]. Kumar, S., West, R., & Leskovec, J. (2016). Disinformation on the web: Impact, characteristics, and detection of Wikipedia hoaxes. Proceedings of the 25th international conference on World Wide Web.

[7]. Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2012). Detecting automation of Twitter accounts: Are you a human, bot, or cyborg? IEEE Transactions on Dependable and Secure Computing, 9(6), 811-824.

[8]. Instagram Graph API Documentation. Meta Platforms Inc. Retrieved from: https://developers.facebook.com/docs/instagram-api

[9]. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273-297.

[10]. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.

[11]. Zhou, X., & Kapoor, G. (2011). Detecting spammers on social networks. Proceedings of the 2011 International Conference on Web Intelligence, Mining and Semantics.

[12]. Almaatouq, A., Radaelli, L., Pentland, A., & Shmueli, E. (2016). Are you your friends' friend? Poor perception of friendship ties limits the ability to promote behavioral change. PLOS ONE, 11(3), e0151588.

[13]. Lee, K., Eoff, B. D., & Caverlee, J. (2011). Seven months with the devils: A long-term study of content polluters on Twitter. Fifth International AAAI Conference on Weblogs and Social Media.

[14]. Subrahmanian, V. S., Azaria, A., Durst, S., Kagan, V., Galstyan, A., Lerman, K., & Menczer, F. (2016). The DARPA Twitter Bot Challenge. Computer, 49(6), 38-46.

[15]. Wang, A. H. (2010). Detecting spam bots in online social networking sites: A machine learning approach. Data and Applications Security and Privacy XXIV, 335-342.

[16]. Saberi, M., Qazvinian, V., & Radev, D. (2011). Detecting fake websites using generative models. Computational Linguistics and Intelligent Text Processing, 285-296.

[17]. Zhang, C., & Paxson, V. (2011). Detecting and analyzing automated activity on Twitter. Proceedings of the 12th international conference on Passive and Active Measurement.

[18]. Sharma, A., & Kaushik, A. (2013). An approach for detection of fake profiles in online social networks. International Journal of Engineering Research & Technology (IJERT), 2(12).

[19]. Aggarwal, C. C. (2015). Data Mining: The Textbook. Springer

DOI: 10.48175/568