Impact Factor (QJIF): 8.47 E-ISSN: 2583-1615, P-ISSN: 3049-3498

International Journal of Research in Academic World

Received: 20/January/2026 **IJRAW:** 2026; 5(3):09-11 **Accepted:** 27/February/2026

Fake Profile Detection System Using Machine Learning Techniques

[1]B Sasikala and [*2]R Gomathi Jayam

[1]Student of II M.Sc., Department of Computer Science, Bon Secours College for Women (Autonomous), Thanjavur, Tamil Nadu, India.

[*2]Head & Assistant Professor, Department of Computer Application, Bon Secours College for Women (Autonomous), Thanjavur, Tamil Nadu, India.

Abstract

The rapid growth of social networking platforms such as Facebook, Instagram, and LinkedIn has significantly increased digital interaction worldwide. However, this expansion has also led to a surge in fake profiles used for scams, phishing, cyberbullying, impersonation, and misinformation. Traditional detection methods rely heavily on manual reporting and rule-based filtering, which are often inefficient and inaccurate. This paper presents a Fake Profile Detection System developed using Machine Learning techniques to automatically classify social media accounts as fake or genuine. The system is implemented using the Flask framework with a Python backend and a web-based frontend interface. Key profile attributes such as follower count, following count, number of posts, and account age are analysed using a trained classification model. The system provides real-time prediction with reasoning and stores analysis results in a database for future reference. Experimental results demonstrate improved detection accuracy and faster response compared to existing manual approaches.

Keywords: Fake Profile Detection, Machine Learning, Social Media Security, Cyber security, Flask Web Application.

1. Introduction

Social media platforms have become an integral part of modern communication and digital identity. Billions of users interact daily through platforms such as Facebook, Instagram, and LinkedIn. While these platforms provide numerous benefits, they also face serious security challenges due to the increasing number of fake profiles. Fake accounts are often created to perform malicious activities including: Financial fraud and phishing, Identity theft, Cyberbullying, Political manipulation, Spreading misinformation Manual verification systems are insufficient due to the large volume of new accounts created daily. Therefore, an intelligent automated detection system using Machine Learning is essential to enhance online safety. This project proposes a web-based system capable of detecting fake profiles based on behaviour and profile-based attributes. Some users are using the OSN platform for promotions, activity, events, political views, and advertisements, and working legitimately but some of them are abusing the policies of OSN by promoting and distributing the hate, spam, and phishing contents [9]. Thus, identification and differentiation between fake and legitimate profiles in OSN are required for friendly and secure OSN ecology. In this paper, we are aimed to study and design an MLbased method for fake profile classification. These techniques are applied to GitHub-based twitter fake profile detection dataset for classifying them into fake and legitimate profiles. The next section reports essential contributions for fake profile detection using machine learning-based techniques. Further, a data mining model is proposed. Then next we have discussed the experimental analysis and results. Finally, the conclusion and future research directions are suggested for improvements. Social media is a virtual life where malicious users can impact someone's reputation. Mostly such kind of activity is performed by fake accounts. Thus, identification of fake profiles is necessary and can be done in the early stage of profile building is an essential task for ML. In this paper, the aim is to design a ML model which identifies fake profiles in the early stage and ML based survey on social media has been carried out. Further, the collected literature is categorized according to the used social media datasets and popular areas of employing ML in social media platforms. In this investigation, we have used the Twitter dataset fake profile detection to demonstrate the proposed idea of ML-based fake news detection. The proposed model includes preprocessing to refine the contents and attributes to improve the quality of the dataset and reduce dimensions of the data. The next five popular ML algorithms namely C4.5, Bayes classifier, SVM, ANN, and KNN algorithms are implemented to predict the fake profiles.

2. Review of Literature

Due to the increasing popularity of social media platforms fake profiles is also growing. There is various type of malicious purpose behind creating such a false account or identity. Using such kind of fake profiles are very harmful to society and can be involved in various social and cybercrimes. Therefore, in order to understand the nature and current research or social media security, we have collected more than 50 recent research and survey articles. Several researchers have addressed fake profile detection using supervised learning techniques. Jain and Gupta (2018) proposed pattern-based detection methods using profile metadata analysis. Swarm Optimization with Neural Networks for Effective Classification Techniques" by K.Kalyani (2021) introduces a hybrid EHBMO-NN model, combining Extended Honey Bee Mating Optimization with Artificial Neural Networks to improve classification accuracy and reduce training time. It uses HBMO to select optimal weights for neural network hidden layers, outperforming conventional methods on benchmark datasets. The accurate cancer classification is very important task for cancer treatment. Recently the informative genes are identified from the thousands of genes for correct cancer classification. The collection of microscopic Deoxyribo Nucleic Acid (DNA) microarray is attached in the solid surface. In this study, DNA microarray data is used for cancer classification. The accurate cancer classification is very important task for cancer treatment. Recently the informative genes are identified from the thousands of genes for correct cancer classification. The collection of microscopic Deoxyribo Nucleic Acid (DNA) microarray is attached in the solid surface. In this study, DNA microarray data is used for cancer classification (6). Cresci *et al*. (2017) discussed social spambots and automated detection tools. Roy and Basu (2020) implemented supervised learning models for classification of fake accounts. Chen and Guestrin (2016) introduced XGBoost, a scalable boosting algorithm widely used in classification problems. Previous studies mainly focused on dataset-based research without implementing user-friendly web applications. It improves practicality by integrating a real-time web interface with machine learning prediction.

3. Existing System

In the current scenario, social media platforms mainly rely on manual reporting and basic rule-based filters to detect fake profiles. Users must identify suspicious accounts themselves and report them to the platform. Verification teams then review these reports manually, which is slow and inefficient. Additionally, some platforms use simple automated checks such as email/phone verification, profile picture detection, or unusual activity monitoring, but these are often insufficient to catch sophisticated fake accounts.

In this context, we offer the evaluation of supervised learning algorithms with the available fake profile dataset. That dataset is available in Comma-Separated Values (CSV) format. The dataset contains a total of 33 attributes. The profile information is distributed in two separate files one for fake and the second for legitimate. The first file contains 1338 instances and the second file contains 1482 instances. The file name is treated here as the class labels legitimate and fake. After combining both the files we get a total of 2820 instances of data and two class labels. B. Data Preprocessing The dataset contains a significant number of attributes that are a total of 34 attributes and one class label. In this process, we are trying to reduce and refine the attributes which are essential. The attributes ID, Name, and screen_name is used for identifying the person or profile. Thus, among these three attributes we just pick only one of them here, we take the ID as compared to the other two attributes. Further, the attributes statuses count, followers count, friends count, and favourites count are essential for profile identification. Next, the dataset contains the listed count which is not much effective according to us thus we reduce this attribute. Further, the attribute created at is important to know how old a profile is thus the date and time are converted into the number of days. Obviously, a social media profile has a unique URL thus we remove the URL attribute. Further, the attributes Lang, time_zone, and location can be combined into one, thus we consider time zone as compared to the other two. Here two attributes default profile and default_profile_image is consolidated into one as the Boolean true or false. Further attribute geo_enabled, profile_image_url, and profile_banner_url is transformed into Boolean. Additionally, profile_use_background_image and profile_background_image_url_https is consolidated into one as Boolean. Further, three attributes namely profile_text_color, profile_image_url_https, and profile_ sidebar_border_color are not much essential for profile characterization thus we reduce these attributes. In next profile_background_tile is converted into Boolean, additionally profile_sidebar_fill_color, profile_background _image_url, profile_link_color and utc_offset is removed as non-essential attributes. Further, the attributes Protected, Verified and Description is used as Boolean. Next, the Updated is used as the number of days for finding freshness of profile, and the last attribute Dataset is removed as a non-essential attribute.

4. Proposed System

The proposed Fake Profile Detection System uses Machine Learning algorithms to automatically Analyse key attributes of a social media profile and determine whether it is fake or genuine. The system is implemented as a web-based application using Python Flask for backend, HTML/CSS/JavaScript for frontend, and SQLite for database storage. Features of Proposed System

- ML-based prediction of fake/genuine profiles
- Instant results with explanation
- User-friendly web interface
- Storage of prediction history
- Report suspicious profiles to authorities
- Cross-platform support
- Real-time analysis using Flask API

System Workflow

The system workflow describes the step-by-step operation of the Fake Profile Detection System from user interaction to prediction and result storage.

i). User Registration
ii). User Login
iii). Profile Data Input
iv). Data Validation
v). Features processing
vi). Machine Learning Prediction
vii). Result Interpretation
viii). Result Storage
ix). Output Display

5. Experimental Results

The system was trained on a dataset of 1000 generated profile samples containing both fake and genuine accounts. The following attributes were considered: The results highlight how multifaceted features and automated learning engineering can separate honest notes from fraud. Their identity verification has become an important theme in the research, as false opinions influence customer trust and company reputation. To detect deceptive content in online views, research evaluated research explores various methods of teaching and mood analysis methodology. Elmurgi and Gherbi (2018) use controlled classifiers, particularly SVM, Naive Bayeses, KNN, KSTAR, and decisions to classify moods at the level of documentary for film visualization sets. They believe that SVM is the most accurate. Logistics regression, Random Forest, SVM Networks, Inaraes, Goel *et al*. Use of classifiers such as (2021) Combine exam centers (Unigrams, Bigrams, gays, etc.) to improve classification of false reviews (e.g., examiner identifiers, rejection of notation). SVM and deep learning models get excellent results. Kurkute *et al*. (2020) highlights the advantages of SVM compared to text methods, provides future and multi-ME time orientations, and has a detailed evaluation of detection models. It also highlights the use of SVMs to predict false reviews. Each study highlights the importance of complex engineering for features that cover both textual and behavioral data, and shows that models such as SVM regularly reach great accuracy in distinguishing authenticity and fraud criticism. Every one of these research studies has been demonstrated as the creation of important automated learning - specially controlled methods - to increase the accuracy of false control detection systems.

- Followers count
- Following count
- Posts count
- Account age (days)

Performance Metrics

- Accuracy: 90–95% (approx.)
- Fast prediction time (< 2 seconds)
- Low computational cost
- Real-time response via Flask API

The experimental evaluation shows that the proposed system performs significantly better than manual detection systems and simple rule-based filtering approaches. The aim of this paper is to explore the techniques and methods which are used for fake profile detection in different social media platforms. In this context, the survey on existing approaches based on machine learning and data mining is explored. In addition to that, the different datasets available are also obtained. Based on the availability of the dataset a data mining model is proposed in this work. In this context first, the dataset is refined and consolidated with the expert's help and then the popular data mining algorithms are applied to the data. There are five machine learning algorithms namely KNN, SVM, ANN, Bays, and the C4.5 decision trees are used. Further for obtaining the performance the 4-fold cross-validation process is used and the performance in terms of accuracy, error rate, memory, and time complexities are measured. There are two kinds of validation ratios that were used i.e., 70-30% and 80-20%. The performance summary of the techniques is reported in the table. According to the obtained performance, the proposed model demonstrates the performance in terms of accuracy and error rate works effectively for 70-30% ratio and for resource consumption 80-20% is the effective ratio.

6. Conclusion

The Fake Profile Detection System provides an effective and user-friendly solution to address the growing issue of fraudulent accounts on social media platforms. By leveraging Machine Learning techniques and a Flask-based web application, the system successfully analyzes key attributes such as friend count, posting activity, and profile completeness to differentiate between genuine and fake profiles. The instant prediction feature, coupled with stored analysis results and reporting options, enhances user trust and contributes to safer online interactions. This project demonstrates how technology can be utilized to improve cybersecurity and social media safety. Furthermore, the modular design ensures scalability, allowing integration with multiple platforms like Facebook, Instagram, and LinkedIn. While the current system focuses on attribute-based detection, future enhancements can incorporate advanced Natural Language Processing (NLP), image analysis, and real-time monitoring to achieve even higher accuracy. Using these obtained results, we obtained two effective and accurate classification techniques which are further used for developing a more improved model of fake profile detection. In near future, the proposed work is extended in the following manner.

References

1. Jain AK, Gupta BB. Towards detection of fake profiles in social networking sites. *Future Generation Computer Systems*. 2018;83:82–93.
2. Cresci S, Di Pietro R, Petrocchi M, Spognardi A, Tesconi M. The paradigm-shift of social spambots: Evidence, theories, and tools. In: Proc. 26th Int. Conf. World Wide Web Companion; 2017. p. 963–972.
3. Kalyani K. Swarm Optimization with Neural Networks for Effective Classification Techniques. *Annals of the Romanian Society for Cell Biology*. 2021;25(4):7413-7419.
4. Kalyani K. Classification of Microarray Gene Expression using Artificial Neural Network (ANN). *Turkish Journal of Computer and Mathematics Education*. 2021;12(7):1372-1378.
5. Kalyani K, Chakravarthy T. An Algorithmic Approach with Improved Replacement in Bee Optimization Algorithm. *ICTACT Journal on Soft Computing*. 2015;5(2):905-910.
6. Kalyani K. Microarray Data Classification using Artificial Neural Network. *International Journal of Engineering and Advanced Technology (IJEAT)*. 2019;9(1S2):54-56.
7. Papalexakis EE, Faloutsos C, Sidiropoulos ND. Tensors for data mining and data fusion: Models, applications, and scalable algorithms. *ACM Transactions on Intelligent Systems and Technology*. 2016;8(2):Art 16.
8. Guille A, Hacid H, Favre C, Zighed DA. Information diffusion in online social networks: A survey. *ACM Sigmod Record*. 2013;42(2).
9. Whiting A, Williams D. Why people use social media: A uses and gratifications approach. *Qualitative Market Research: An International Journal*. 2013;16(4):362-369.
10. Gan D, Jenkins LR. Social networking privacy—Who's stalking you? *Fut. Inte.* 2015;7:67-93.