IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

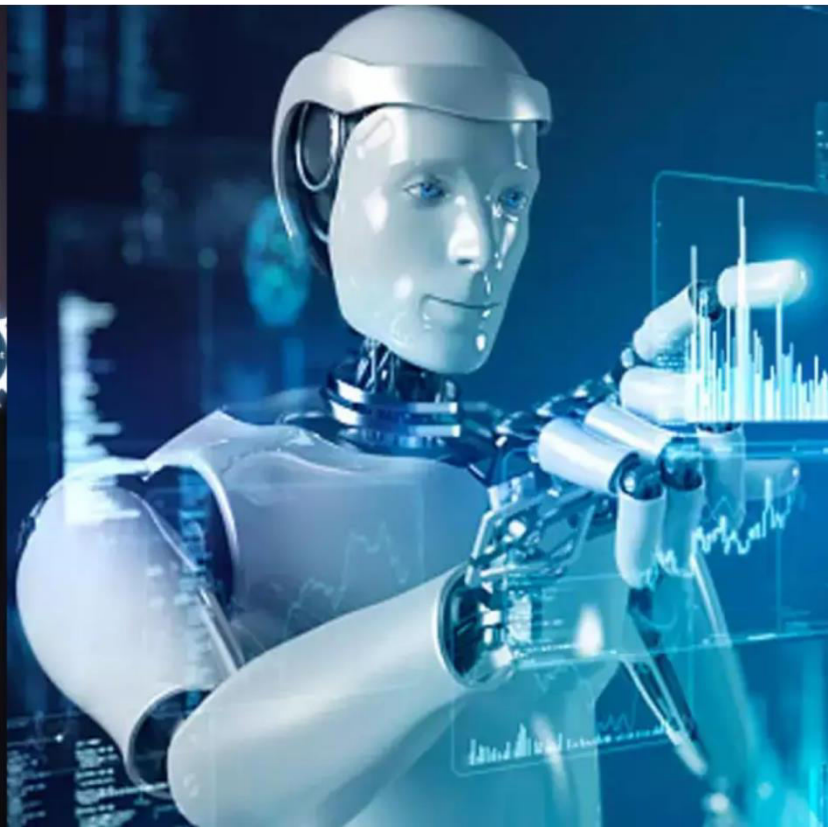
ISSN(O): 2319-8753
ISSN(P): 2347-6710



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 9.935

Volume 15, Issue 11, November 2026

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 11, November 2026

|DOI: 10.15680/IJIRSET.2026.1511021|

Fake Account Detection in Social Networks using Machine Learning

M.Jeevitha, N.Ashok Kumar, K.N.Sivakumar

ME(CSE) - 2nd Year Scholar, Department of Computer Science and Engineering, M.P.Nachimuthu M.Jaganathan Engineering College (An Autonomous Institution), Chennimalai, Erode, India

Associate Professor, Department of Computer Science and Engineering, M.P.Nachimuthu M.Jaganathan Engineering College (An Autonomous Institution), Chennimalai, Erode, India

Associate Professor/ Head, Department of Computer Science and Engineering, M.P.Nachimuthu M.Jaganathan Engineering College (An Autonomous Institution), Chennimalai, Erode, India

ABSTRACT: An online social network (OSN) is a virtual platform where users can create profiles, connect with other users, and share information, such as photos, videos, and text-based content. Examples of popular OSNs include Facebook, Twitter, Instagram, LinkedIn, and TikTok. OSNs have become an essential part of people's daily lives, enabling them to stay connected with friends, family, and colleagues, share experiences and ideas, and discover new content. However, OSNs can also present challenges, such as the risk of cyber bullying, the spread of misinformation, and the presence of fake accounts. Fake account detection in online social networks (OSNs) refers to the process of identifying and removing fraudulent accounts that are created with the intention of engaging in malicious activities, such as spreading false information, conducting scams, or engaging in cyber bullying. Traditionally we have different methods to classify fake and genuine profiles. In this project we are proposing Machine Learning and Natural Language Processing techniques to detect the fake accounts in Online Social Networks. We are using the Random Forest, Support Vector Machine and Naïve Bayes Algorithms

KEYWORDS: OSN, Cyber Bullying, RF, SVM, Fake Accounts

I. INTRODUCTION

Artificial Neural Networks (ANNs) can be used to identify fake profiles on social media platforms. ANNs are a type of machine learning algorithm that are inspired by the structure and function of the human brain. ANNs can learn to identify patterns in data, making them well-suited for tasks like identifying fake profiles. To use ANNs for identifying fake profiles, a dataset of profiles is needed that contains both real and fake profiles. This dataset is used to train the ANN on features that are indicative of fake profiles, such as the number of friends, the frequency of posts, and the types of content shared. Once the ANN is trained, it can be used to classify new profiles as either real or fake. The ANN takes in the features of a profile as input and outputs a binary classification (real or fake) One of the key advantages of using ANNs for identifying fake profiles is that they can learn to identify patterns that are not immediately obvious to humans. For example, ANNs can learn to identify subtle differences in the language used by fake profiles that are not easily detected by humans. However, it's important to note that ANNs are not perfect and may still make errors in identifying fake profiles. It's also important to ensure that the training dataset is representative of the population of profiles on the social media platform in question, to avoid biases in the classification of profiles. Additionally, ethical considerations should be taken into account when developing such systems, such as ensuring user privacy and preventing discrimination.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 11, November 2026

|DOI: 10.15680/IJIRSET.2026.1511021|

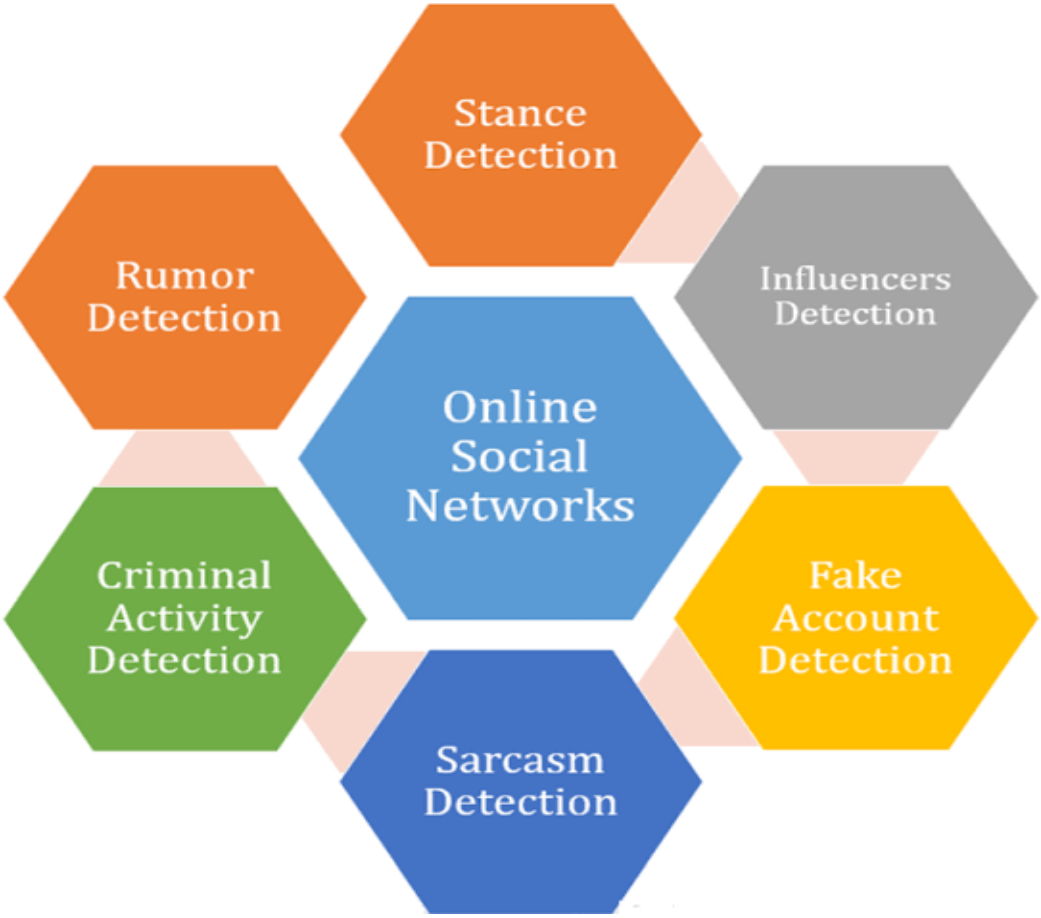


Fig 1: Fake account detection in social network

In recent years, the widespread use of online social networks has created an environment in which individuals can easily create and maintain multiple profiles. However, some of these profiles may be created with the intention of deceiving others or spreading false information, known as "fake profiles." Detecting fake profiles is a crucial task to maintain the authenticity and trustworthiness of social networks. Machine learning (ML) algorithms and natural language processing (NLP) techniques have emerged as effective tools for detecting fake profiles in online social networks. In this project, we aim to explore the use of ML algorithms and NLP techniques for detecting fake profiles in online social networks. Our focus will be on identifying the key features that distinguish fake profiles from real ones and building models that can accurately classify profiles as either real or fake. The outcome of this paper can help social networks to improve their security and user trust by eliminating fake profiles. ML algorithms and NLP techniques is multifaceted. Firstly, the presence of fake profiles can negatively impact the user experience of social networks by spreading false information, promoting malicious activities, and compromising user privacy. Therefore, it is essential to develop robust methods to detect fake profiles to maintain the authenticity and credibility of social networks.

II. LITERATURE SURVEY AND RELATED WORK

In 2018, Yeh-Cheng chen and Shystunfelix Wu [1] have presented Fake Buster: A Robust fake Account detection by Activity Analysis. They proposed an innovative method to detect fake account in OSNs(Online Social Networks). It is develop for accurately detecting fake account among social network users, based on various activity collection and analysis. In this research they have use Random forest, along with C$.5 and Adaptive Boosting, with decision stump as a second classifier that created behind it to focus on the instance in the training data , in case the accuracy of the first

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 11, November 2026

|DOI: 10.15680/IJIRSET.2026.1511021|

classifier is less effective. After finish training, a cluster of features for each testing account will input into models and output a prediction with rank score indicating the likelihood of being fake account.

In 2019, Faiza Masood, Ghana Ammad, Ahmad Almogren, Assad Abbas, Hasan Ali Khathak, Ikram Uddin, MohsenGuizani, and MansourZuair [2] have presented in their work Spammer detection and fake user identification on social network. A review of techniques used for detecting spammers on Twitter. Spammers can be identified based on: (i) fake content, (ii) URL based spam detection, (iii) detecting spam in trending topics, and (iv) fake user identification. The proposed taxonomy of spammer detection on twitter is categorized into four main classes, namely,(i) fake content, (ii) URL based spam detection, (iii) detecting spam in trending topics, and (iv) fake user identification. The first category (fake content) includes various techniques, such as regression prediction model, malware alerting system, and Lfun scheme approach. In the second category (URL based spam detection), the spammer is identified in URL through different machine learning algorithms. The third category (spam in trending topics) is identified through Naïve Bayes classifier and language model divergence. The last category (fake user identification) is based on detecting fake usersthrough hybrid techniques.

Farhan, Muhammad Ibrohim, Indra Budi [3] have presented in their work Malicious Account Detection on Twitter based on Tweet Account features using Machine Learning. In this research, build a malicious account detection that can distinguish genuine accounts from malicious accounts using only tweet features of the accounts. Also managed to build a multiclass classification for the two types of malicious accounts,fake followers and spam bots using only tweet features. Lastly, found the best combination of algorithms, features, and data transformation scenario that suits best of our problem.

III. METHODOLOGY

Data pre-processing is an important step for the creation of a machine learning model. Initially, data may not be clean or in the required format for the model which can cause misleading outcomes. In pre-processing of data, we transform data into our required format. It is used to deal with noises, duplicates, and missing values of the dataset. Data pre-processing has the activities like importing datasets, splitting datasets, attribute scaling, etc. Pre-processing of data is required for improving the accuracy of the model. For achieving better results from the applied model in Machine Learning, the format of data in a proper manner. Kaggle provides you pre-processed dataset. But how this data is pre-processed is discussed below. Natural Language Processing (NLP) pre-processing techniques are used to transform raw text data into a format that can be easily understood and analysed by machine learning algorithms. Here are some NLP pre-processing techniques used in project.

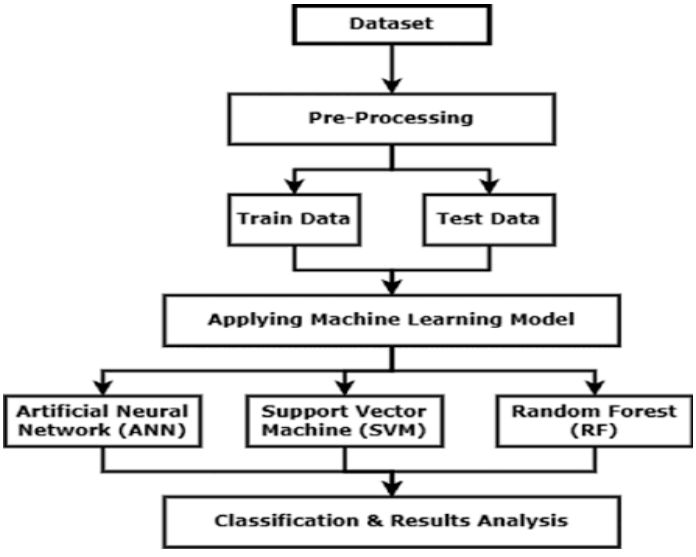


Fig 2: Fake User Account Detection in Online Social Media Networks Using Machine Learning and Neural Network Techniques

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 11, November 2026

|DOI: 10.15680/IJIRSET.2026.1511021|

The methodology for fake account detection in Online Social Networks (OSNs) using Random Forest, Naive Bayes, SVM, and NLP techniques involves the following steps: First, collect a large dataset of user accounts in the OSN, including both genuine and fake accounts. Next, extract a set of features from each account, including user-generated content and linguistic patterns, using NLP techniques. Then, use Random Forest, Naive Bayes, and SVM algorithms to select the most important features and train the classifiers on the labeled dataset. After that, deploy the trained classifiers to detect fake accounts in the OSN using the selected features and linguistic patterns. Finally, monitor the performance of the system over time and update the classifiers as needed to adapt to new types of fake accounts and evolving attack strategies. By integrating NLP techniques with machine learning algorithms, the proposed methodology can provide a more accurate and robust solution for fake account detection in OSNs.

IV. RESULT ANALYSIS

Naive Bayes is a probabilistic machine learning algorithm that is commonly used for classification tasks. The algorithm is based on Bayes' theorem, which describes the probability of a hypothesis given evidence. In the context of classification, Naive Bayes works by calculating the probability that a data point belongs to a particular class based on its features. The algorithm assumes that each feature is independent of the others, which is why it is called "naive." This simplifies the calculations and makes the algorithm faster and more efficient. There are several types of Naive Bayes classifiers, including: Gaussian Naive Bayes: This classifier assumes that the features are normally distributed. Multinomial Naive Bayes: This classifier is commonly used for text classification tasks and assumes that the features are counts of occurrences of words. Bernoulli Naive Bayes: This classifier is also commonly used for text classification tasks and assumes that the features are binary (e.g., whether a word is present or not).

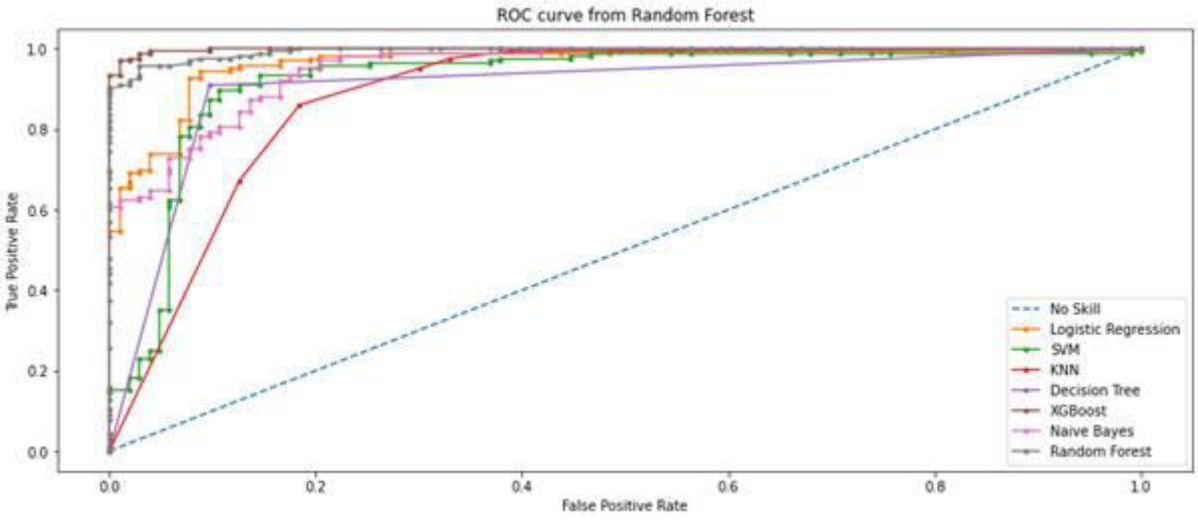


Fig 3: Improvement of a Machine Learning Model Using a Sentiment Analysis Algorithm

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of over fitting to their training set. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance. Random forests are frequently used as "blackbox" models in businesses, as they generate reasonable predictions across a wide range of data while requiring little configuration.

V. CONCLUSION AND FUTURE SCOPE

In our Paper, we proposed machine learning algorithms along with natural language processing techniques to classify the fake and genuine profiles. By using these techniques, we can easily detect the fake profiles from the social network sites. In our project we took the Face book Data set to identify the fake profiles. The NLP pre-processing techniques are

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 11, November 2026

|DOI: 10.15680/IJIRSET.2026.1511021|

used to analyse the dataset and machine learning algorithms such as Random Forest, SVM and Naïve Bayes are used to classify the profiles. These learning algorithms are improved the detection accuracy rate in our project. The accuracy of present proposed system is 73.9%. The future scope of fake account detection in OSNs using machine learning algorithms such as random forest, SVM, Naive Bayes, and NLP is quite vast. Some of the potential areas of development are: Integration with deep learning: Deep learning techniques such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) can be used to improve the performance of fake account detection algorithms. Sentiment analysis: Sentiment analysis can be used to analyse the sentiment of user-generated content and metadata to detect fake accounts. Graph-based techniques: Graph-based techniques can be used to analyse the network structure of OSNs to detect fake accounts. For example, centrality measures can be used to identify accounts that are highly connected to other fake accounts. Real-time detection: Real-time detection of fake accounts can be achieved using streaming algorithms that process data in real-time and adapt to changing data distributions. Multi-modal data analysis: Multi-modal data analysis can be used to analyse different types of data such as text, images, and videos to detect fake accounts. Transfer learning: Transfer learning can be used to transfer knowledge learned from one OSN to another OSN, improving the accuracy of fake account detection algorithms in new OSNs. Overall, the future scope of fake account detection in OSNs using machine learning algorithms such as random forest, SVM, Naive Bayes, and NLP is promising and offers many opportunities for further research and development.

REFERENCES

[1] Michael Fire et al. (2012). "Strangers intrusion detection-detecting spammers and fake profiles in social networks based on topology anomalies." Human Journal 1(1): 26-39.Günther, F. and S.
[2] Fritsch (2010). "neuralnet: Training of neural networks." The R Journal 2(1): 30-38
[3] Dr. S. Kannan, Vairaprakash Gurusamy, “Pre-processing Techniques for Text Mining”, 05
[4] Shalinda Adikari and Kaushik Dutta, Identifying Fake Profiles in LinkedIn, PACIS 2014 Proceedings, AISeL
[5] Z. Halim, M. Gul, N. ul Hassan, R. Baig, S. Rehman, and F. Naz,“Malicious users’ circle detection in social network based on spatiotemporal co-occurrence,” in Computer Networks and Information Technology (ICCNIT),2011 International Conference on, July, pp. 35–390.
[6] Liu Y, Gummadi K, Krishnamurthy B, Mislove A,” Analyzing Facebook privacy settings: User expectations vs. reality”, in: Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference,ACM,pp.61–70.
[7] Mahmood S, Desmedt Y,” Poster: preliminary analysis of google?’s privacy. In: Proceedings of the 18th ACM conference on computer and communications security”,ACM 2011, pp.809–812.
[8] Stein T, Chen E, Mangla K,” Facebook immune system. In: Proceedings of the 4th workshop on social network systems”, ACM 2011. [9] Saeed Abu-Nimeh, T. M. Chen, and O. Alzubi, “Malicious and Spam Posts in Online Social Networks,” Computer, vol.44, no.9, IEEE2011.
[10] J. Jiang, C. Wilson, X. Wang, P. Huang, W. Sha, Y. Dai, B. Zhao, Understanding latent interactions in online social networks, in: Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement, ACM, 2010, pp. 369–382.
[11] Kazienko, P. and K. Musiał (2006). Social capital in online social networks. Knowledge- Based Intelligent Information and Engineering Systems, Springer



INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details