International Journal of Engineering and Management Research e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-13, Issue-3 (June 2023)
https://ijemr.vandanapublications.com https://doi.org/10.31033/ijemr.13.3.4

Machine Learning Approaches for Fake User and Spammer Detection: A Comprehensive Review and Future Perspectives

Farheen Siddiqui[1] and Mohammad Suaib[2]
[1]Department of Computer Science & Engineering, Integral University, Lucknow, INDIA
[2]Associate Professor, Department of Computer Science & Engineering, Integral University, Lucknow, INDIA

[1]Corresponding Author: farheensiddiqui78687@gmail.com

Received: 02-05-2023 **Revised:** 14-05-2023 **Accepted:** 30-05-2023

ABSTRACT

The rise of digital platforms has given way to a surge in fraudulent activities, including the creation of fake user accounts and the prevalence of spammers. These malevolent actions present significant challenges to the security and integrity of these platforms, necessitating effective detection and prevention measures. This paper offers an extensive review of machine learning (ML) techniques currently employed for fake user and spammer detection. The paper explores a range of traditional ML algorithms such as decision trees, support vector machines, and logistic regression, as well as more complex deep learning models like convolutional neural networks (CNN) and recurrent neural networks (RNN). It also examines unsupervised and semi-supervised learning strategies that can be used when labeled data is scarce. Furthermore, we discuss the key challenges in detecting fake users and spammers, including the dynamic nature of spamming tactics, evolving deceptive strategies, data imbalance, and privacy issues. We propose potential solutions to these challenges like transfer learning, active learning, federated learning, and privacy-preserving ML techniques. The paper concludes with an exploration of emerging technologies such as explainable AI and reinforcement learning and their potential to enhance detection system performance and interpretability. It also provides insights into promising future research directions in this critical area.

***Keywords*—** Machine Learning, Spammer Detection, Fake User Detection, Fraud Detection, Deep Learning, Unsupervised Learning, Semi-Supervised Learning, Transfer Learning, Active Learning, Federated Learning, Explainable AI, Reinforcement Learning, Online Security

I. INTRODUCTION

The entrance of the digital age has brought with it a great deal of ease and opportunity; nevertheless, it has also led to an increase in fraudulent activities, most notably the manufacture of phoney user accounts and the growth of spammers. The digital age has brought with it a great deal of ease and opportunity; however, it has also led to an increase in fraudulent activities. These malicious activities pose significant problems not only to the safety and integrity of digital platforms but also to the user experience that these platforms offer. As a consequence of this, the development of trustworthy and effective methods for detecting behaviours of this kind and putting an end to them is an absolute necessity.

Techniques that are founded on machine learning (ML) provide a significant amount of potential in this sector. When it comes to their responsibility of identifying fraudulent users and spammers, they are incredibly effective due to their ability to learn from data, recognise patterns, and make predictions. This is because of their capacity to learn from data, recognise trends, and make predictions. Techniques that are considered to be more traditional in the realm of machine learning, such as decision trees, support vector machines, and logistic regression, have all seen significant application in this area. On the other hand, in order to account for the intricacies of online user behaviour as well as spamming methods, more complex models are required. It has been demonstrated that deep learning algorithms, such as convolutional neural networks (CNN) and recurrent neural networks (RNN), are effective in capturing complicated patterns and temporal connections, which enables improved detection abilities.

Nevertheless, employing ML for the goal of identifying fake users and spammers brings with it its own particular set of challenges. It may be tough for static models to continue to be effective over the course of time because of the dynamic and ever-changing nature of the spamming methods and misleading schemes. The problem of data imbalance, which arises when the number of real users substantially exceeds the number of fraudulent users and spammers, may inject bias into the learning process and influence how well the model works. This problem occurs when the number of real users greatly surpasses the number of fraudulent users and spammers. When working with sensitive user data, one of the most significant considerations becomes protecting the users' privacy.

There have been a number of different strategies proposed for consideration as possible answers to these issues. Transfer learning, active learning, and federated learning are some examples of strategies that can help handle the problems of data scarcity and model staleness while simultaneously respecting individuals' rights to privacy. These methods include transfer learning, active learning, and federated learning. It is possible that the deployment of AI in these systems that can be explained will have the potential to improve their interpretability and

transparency, which may, in turn, encourage users to have more faith in the systems.

Exciting new potential for the field's future study can be found in recent advancements in the field, such as algorithms for privacy-preserving machine learning and reinforcement learning. Learning from one's environment in order to form the most accurate possible judgements can be accomplished through the process of reinforcement learning, which could make the development of adaptive and secure detection systems conceivable. It's possible that the solution to the challenge of handling sensitive user data in an ethical and responsible manner lies in the application of machine learning techniques that safeguard the privacy of end users.

As a result of this, the ever-increasing complexity of fraudulent operations like the formation of phoney users and spamming necessitates continual study and innovation within the sector. This article's goal is to provide an overview of the current state of the art, the challenges that have been overcome, and the potential future directions in the field of false user and spammer identification utilising machine learning. It is our hope that it will serve as a comprehensive resource for academics, professionals, and policymakers who are interested in this vital topic.

II. LITERATURE REVIEW

A thorough awareness of the current techniques, difficulties, and possibilities in the field of fake user detection is required due to the fast changing digital world and the associated rise in harmful actions, such as the creation of false user accounts. This calls for a complete understanding of the current methods, challenges, and prospects. The increased interest and investment in this field has led to the production of a large body of work that covers a wide range of perspectives on this subject. Given its potential for pattern identification and predictive modelling, the use of machine learning methods, in particular, for the detection of fraudulent users has gained a substantial amount of interest in recent years. The purpose of this literature review is to provide a detailed perspective of the current environment and offer options for future study. The review will summarise and synthesise the results of 25 significant publications that have been published on this subject.

The studies that were chosen represent a wide variety of approaches and application domains, which is reflective of the breadth and depth of study that has been conducted in this topic. These papers present a wide variety of strategies for detecting and preventing fake user activity, ranging from more traditional machine learning techniques such as decision trees, support vector machines, and logistic regression to more advanced deep learning models such as convolutional neural networks and recurrent neural networks.

The analysis also takes into account the inherent difficulties that come with identifying bogus users. The papers that are now being reviewed often present a number of challenges, including, but not limited to, dynamic and ever-changing user behaviours, the tactics used by attackers, data imbalance, and the protection of personal privacy. Researchers have come up with a number of potential ways to address these issues, some of which include transfer learning, active learning, federated learning, and the incorporation of explainable AI.

The evaluation outlines possible future research areas, as indicated by the body of literature that was reviewed. Among them are the construction of dependable and generalizable models, the investigation of the possibilities presented by reinforcement learning, and the need of a moral and legal structure to govern the use of user data. This literature review aims to present a comprehensive picture of the current state of study in fake user identification using machine learning by providing a summary of these important issues and trends. This will pave the way for future studies and breakthroughs.

Table 1: Survey

Ref	Author Name	Title	Objective	Method Used	Description & Accuracy	Pros	Cons
[1]	Gomathy, C K.	The Spammer Detection and Fake User Identification on Social Networks	The focus here is on Twitter as a platform for analysing user activity in order to identify spam and fraudulent accounts.	This research makes use of a Machine Learning Technique, which is comprised of five different modules: an Admin Module, Data Collection Module, Train and Test Module, and Machine	The procedure begins with the creation of the system for online social networking and the collection of data through the use of the Tweepy Python package. After that, this information is put to use to	Pros: 1. A fuller comprehension of user behaviour might result from classifying users based not only on their own characteristics but also on the characteristics of the content	Cons: 1. Because there are no defined accuracy measures, it is difficult to determine how successful this strategy is. 2. If trending topics are used to collect data, this could limit the diversity of

				Learning Technique Module. Additionally, a Fake User Detection Module is included.	train and test a machine learning model that is based on the characteristics of both users and content. The final application of this model is to identify bogus users.	they interact with. 2. Because the model is trained on real-time data from Twitter, it may be more flexible to the many spamming strategies that are currently in use..	the training data, which could potentially cause the model to be biassed. 3. Due to the fact that the method was developed expressly to make advantage of Twitter's API and metadata, it may only be applicable to Twitter.
[2]	Ud Din, Ikram et al.	Spammer Detection and Fake User Identification on Social Networks	The purpose of this research is to examine and classify the several methods that have already been developed to identify spammers on Twitter, as well as to locate subject areas that call for additional investigation..	The research includes an in-depth analysis of the many methods that can be used to detect spam on Twitter. These methods can be broken down into the following subheadings: false content detection, URL-based spam detection, spam detection in hot topics, and fake user detection techniques.	The authors provided a comprehensive analysis and comparison of numerous methods for detecting spam based on a number of characteristics (for example, user, content, graph, structure, and time). They also brought attention to potential future research paths, such as the requirement for more advanced approaches for identifying bogus news and rumour sources. However, no particular accuracy measurements were supplied for any of the strategies that were evaluated.	Pros: 1. The paper presents an in-depth analysis of the various methods that are already in use and demonstrates how they can be of assistance to academics working in this area. 2. It proposes future study directions, which demonstrates that it is cognizant of the constantly changing nature of spamming techniques and the significance of ongoing research.	Cons: 1. It does not provide a new approach for detecting spam; rather, it discusses the several methods that have already been developed. 2. It does not provide any particular accuracy or effectiveness measurements for the approaches that are being reviewed. 3. The absence of an original and fresh contribution may reduce the effect of the study in comparison to other studies that propose novel methods.

[3]	Alaguvathana, P	Identification of spammer and fake accounts on social networks	The purpose of this investigation is to identify spam on Twitter by the utilisation of quantitative and qualitative data.	The research utilised a variety of approaches to data analysis, some of which included the examination of user names, the participants' connections to other users, follower numbers, image content, and account activities. In addition to this, it makes use of the Naive Bayes method and the Bayesian analysis on the Twitter dataset.	The research consisted of an examination of one thousand Twitter accounts, of which 500 were made up completely. Using a Twitter dataset consisting of 11,000 clients and 400,000 tweets, it presented a new model to differentiate between spam users and regular users of the platform. Nevertheless, there was no mention of a particular accuracy measure.	Pros: 1. The research presents a holistic strategy, which takes into account a variety of facets that are associated with a user's Twitter account. 2. It suggests an innovative strategy for differentiating between users who spam and those who do not spam. 3. It incorporates a substantial amount of data, which increases the reliability of the conclusions.	Cons: 1. The research does not provide a particular accuracy metric for the model that is being proposed, which makes it difficult to evaluate the usefulness of the model. 2. The fact that it relies on manual analysis and tangible evidence may prevent it from being scaled up. 3. Because this study focused primarily on Twitter, it is possible that the findings cannot be generalised to other social media sites.
[4]	Koggalahewa, Darshika	An unsupervised method for social network spammer detection based on user information interests	This research aims to address the constraints of data labelling and spam drifting through the utilisation of an unsupervised method for spammer identification.	This paper proposes a pure unsupervised technique that is built on the peer acceptance of a user inside a social network. This acceptance is estimated based on common shared interests between two users that span several related themes.	The primary contribution made by this research is the development of a method for the detection of spammers that is entirely unsupervised and is founded on the concept of users' acceptance by their peers. This method achieves an accuracy of 96.9% while without requiring labelled training datasets in any	Pros: 1. Because the unsupervised method does not rely on labelled datasets, it is more applicable and scalable than the supervised method. 2. The method circumvents the problems that are inherent in supervised methods, which include the labelling of data and the drifting of spam. 3. The	Cons: 1. Because the study does not compare its technique to any other unsupervised methods, determining how well it performs in comparison to other unsupervised approaches is challenging. 2. The emphasis placed on peer approval may not be sufficient to capture all forms of spamming behaviour. 3.

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-13, Issue-3 (June 2023)
https://doi.org/10.31033/ijemr.13.3.4

					way, shape, or form.	efficiency of the suggested strategy is demonstrated by the high level of accuracy, which is 96.9%.	It is not apparent whether or not the findings can be generalised to other types of social media platforms.
[5]	Gupta, Arushi	Improving Spam Detection in Online Social Networks	This study's purpose is to improve spam detection in online social networks, notably Twitter, by identifying spammers who submit content that is unwelcome, irrelevant, or spreads disinformation.	The research presents potential processes that are based on a variety of features that are present at both the tweet and user levels. These features include followers and followees, URLs, spam terms, replies, and hashtags. It uses the Naive Bayes, Clustering, and Decision Trees learning algorithms to make its determinations. An innovative integrated method that incorporates the benefits of these three algorithms has also been suggested as an alternative.	The results of spam identification were significantly improved by the proposed integrated strategy, which included the three algorithms. Spammers were identified with a detection rate of 99%, while the total accuracy was measured at 87.9%.	Pros: 1. This strategy is all-encompassing, as it takes into account a wide range of characteristics at both the tweet and user levels. 2. The integrated method successfully integrates the benefits of several different machine learning methods, which results in an improvement in accuracy. 3. The findings demonstrate a high level of accuracy in identifying spammers as well as non-spammers.	Cons: 1. Because different social media platforms offer different features and user behaviours, the strategy may not be directly applicable to those other social media platforms. 2. The paper does not examine any potential limitations or disadvantages of the integrated strategy, such as its cost in computational resources or its capacity to scale. 3. The research is primarily focused on Twitter, and there is a possibility that the findings cannot be generalised to other OSNs without additional examination.

[6]	Kondeti, Priyanka	Fake Account Detection using Machine Learning	The purpose of this study is to investigate how to combat the issue of fake accounts on social media platforms such as Twitter, which can disseminate misleading information and harmful content.	Several different machine learning methods, including Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), and K-Nearest Neighbours (KNN), are utilised in this work. In addition to this, it utilises two distinct normalisation methods, namely Z-Score and Min-Max, to further increase accuracy.	The strategy was successful in achieving a high accuracy and true positive rate, notably for Random Forest and KNN. As a result, the accuracy as a whole was improved to 98%. According to the findings of the study, the technique may also be utilised for other online social networking (OSN) websites, such as Facebook and LinkedIn.	Pros: 1. A comprehensive method to identify bogus accounts can be obtained through the application of a variety of machine learning techniques. 2. The implementation of normalisation strategies contributes to an increase in the model's degree of precision. 3. The method demonstrates a high degree of adaptability given its potential for use on various OSN platforms.	Cons: 1. Despite the fact that the system appears to have a high level of accuracy, the researchers who developed it believe there is potential for further development. 2. The research does not take into account either the scalability of using numerous machine learning methods or the computing cost of doing so. 3. The research was conducted mostly on Twitter; nevertheless, it may be applicable to other platforms; therefore, this hypothesis has to be confirmed by additional testing.
[7]	Rao, K. Sreenivasa	Detecting Fake Account on Social Media Using Machine Learning Algorithms	The purpose of the study is to identify malicious or fraudulent accounts on online social networks (OSNs) that have the potential to be exploited for the dissemination of false	In order to do detection, the research makes use of machine learning methods. These algorithms include Random Forest, Neural Networks, and Support	The accuracy of this research, which employs sophisticated machine learning algorithms to detect bogus accounts on social media networks, is 90%. In order to determine which	Pros: 1. The research makes use of machine learning methods, which are able to manage massive datasets as well as complicated relationships within the	Cons: 1. Despite having achieved a relatively high level of accuracy, there is still a margin of error of 10%, which indicates that there is room for progress. 2. The authors

			information and the development of political agendas.	Vector Machines (SVMs). The dataset is pre-processed by utilising a wide range of Python packages.	algorithm is most suited for the data set in question, a comparison model is utilised.	data. 2. The utilisation of Python modules during the preprocessing of data enables improved data management and consistency.	of the study do not specify which OSNs they looked at, which restricts how widely its conclusions may be applied. 3. The research does not address the scalability, computational cost, or time complexity of the machine learning techniques that were applied.
[8]	Khaled, Sarah	Detecting Fake Accounts on Social Media	The purpose of this study is to identify suspicious behaviours and fictitious accounts on Online Social Networks (OSNs), with a particular emphasis on Twitter.	For more effective identification of false accounts and bots, a brand-new method called SVM-NN has been presented. In addition to the Support Vector Machine (SVM) and the Neural Network (NN) algorithms, the method makes use of four different techniques for the selection of features and the reduction of dimensions.	The SVM-NN approach makes use of a smaller amount of features than other similar algorithms, but it is nevertheless able to accurately classify approximately 98% of the accounts in the training dataset.	Pros: 1. The SVM-NN method delivers an accuracy rate of 98%, which is rather high. 2. The method employs strategies for feature selection and dimension reduction, which can help in minimising the amount of time spent computing and enhancing the method's overall efficiency.	Cons: 1. The research study does not address the question of whether or not the newly developed SVM-NN algorithm would be just as successful when applied to various forms of data or social media sites. 2. It is not specified how the SVM-NN algorithm deals with the trade-off between precision and recall in the given example. 3. It is not discussed how well the SVM-NN method performs on data that it has not before seen.

[9]	Singh, Naman	Detection of Fake Profile in Online Social Networks Using Machine Learning	The purpose of this study is to locate, recognise, and get rid of fraudulent accounts that have been created by users on various social networking sites.	The principles of machine learning are utilised, with the primary attention being placed on characteristics such as the number of friends and followers an account has on various social networking networks.	The research suggests leveraging publicly available characteristics from social media profiles, like the ratio of friends to followers, in order to identify phoney human accounts on social media. On the other hand, the correctness of this approach is not discussed in the paper.	Pros: 1. Since the approach makes use of data that is freely accessible to the public, it does not infringe on any user rights. 2. The concentration on phoney accounts that were created by humans is an approach to a less-explored area of spam detection.	Cons: 1. The research report does not go into extensive detail on the machine learning approaches that were applied. 2. Neither the efficiency nor the precision of the suggested method are discussed in any detail. 3. It is not clear how this system distinguishes between phoney and real profiles with comparable ratios of followers to friends.
[10]	Harish, K. and Kumar, R. Naveen	Fake Profile Detection Using Machine Learning	The purpose of this study is to identify and distinguish between false and real Twitter identities with the end goal of preventing cybersecurity risks.	The article makes use of machine learning techniques such as neural networks, LSTM, XG Boost, and random forest, and it focuses on features such as follower and friend counts, status updates, and other such things.	The authors suggest using certain characteristics as a means of determining whether or not a social media profile is genuine. In the study, the architecture and hyperparameters are discussed; however, the paper does not provide any detailed details on the correctness of the proposed models. The suggested approach indicates that,	Pros: 1. The strategy aims to identify phoney profiles on social media, which is an important step towards solving a critical problem in cybersecurity. 2. The research makes use of a wide range of cutting-edge machine learning algorithms to identify bogus profiles.	Cons: 1. The published article does not include any detailed details concerning the performance or accuracy of the proposed models. 2. It is not discussed whether or not the study was successful in actually having profiles deleted or disabled. 3. It is not addressed whether or whether the strategies that have been provided can

					in the event that a phoney profile is discovered, the profile can either be disabled or destroyed.		be applied to other social media platforms.
[11]	Ahmad Nazren Hakimi	Identifying Fake Account in Facebook using Machine Learning	The purpose of the research is to identify false accounts on Facebook and address the problem of illegal enterprises that provide services related to phoney accounts.	The process consists of gathering data, determining features, and using machine learning classifiers such as k-nearest neighbour, support vector machine, and neural network.	The purpose of the study is to identify critical traits that differentiate fake users from genuine Facebook users and to collect information on both real and fraudulent Facebook accounts. A classification precision of 82% can be attained using the K-nearest neighbour classifier.	focuses on spotting phoney accounts on Facebook, the social network that has the most users overall among internet platforms. Utilises the classifiers provided by machine learning to ensure precise detection.	The precision value is not extremely flawless, which indicates that there is potential for advancement. The report does not provide any information regarding the size of the dataset or its variety. Restricted to use solely on the Facebook platform.
[12]	Yasyn ELYUSUFI	Social Networks Fake Profiles Detection	The purpose of this research is to determine how to spot phoney profiles on social media platforms and evaluate the effectiveness of several classification methods, specifically Decision Tree (DT) and Naive Bayes (NB).	For the purpose of user profile classification, Decision Tree (DT) and Naive Bayes (NB) algorithms are among the methodologies that are utilised.	This research examines the effectiveness of two classification algorithms, one with an accuracy of 99% and the other with an accuracy of 78%, in the detection of fraudulent profiles on social media networks.	The report offers insights into the impact of utilising particular algorithms for the detection of false profiles.	The research does not disclose any detailed specifics regarding the dataset that was utilised or the accuracy that was reached by the classification methods.

[13]	M. Mamatha	Fake Profile Identification using Machine Learning Algorithms	The primary purpose of this research project is to determine whether or not Instagram profiles are authentic or fabricated. The study puts out the idea of training algorithms with historical data that includes both phoney and real user accounts. These trained models are then applied to the task of determining whether newly collected test data is authentic or not..	Machine learning-based methods	The dataset is pre-processed by using a number of different Python libraries, and then a comparison is carried out in order to choose a suitable algorithm. For the purpose of identifying false accounts, a number of different machine learning techniques, such as Random Forest, Neural Network, and Support Vector Machines, are utilised. The XGBoost algorithm has a good performance, with an accuracy of 96%.	- This article explains how to spot fraudulent profiles on Instagram and provides a strategy for doing so.	- With an accuracy of 96%, it is clear that the detection of bogus accounts is working effectively.
[14]	Yasyn Elyusufu	Social Networks Fake Profiles Detection Using Machine Learning Algorithms	The purpose of this study is to expose phoney profiles that are present on various social media platforms. In this study, many methods for identifying phoney profiles are discussed, with a primary emphasis placed on the examination of profile data and individual accounts. The	Random Forest, Decision Tree, and Naive Bayes are some examples of supervised machine learning algorithms.	The paper focuses on the role that fake identities play in advanced persistent threats and emphasises how important it is to identify bogus profiles as early as possible. The accuracy reached was 99.6%, showing that the detection of fraudulent profiles was carried out	- Excellent precision of 99.6%	- handling noisy data.

			development of fake profiles on social networks is regarded as a dangerous form of cybercrime; therefore, detection at an early stage is essential. In this study, we investigate a variety of techniques and strategies for detecting bogus profiles that have been offered in the existing research. The effectiveness of three supervised machine learning methods, namely Random Forest (RF), Decision Tree (DT-J48), and Naive Bayes (NB), is analysed and compared for the purpose of determining whether or not profiles are phoney or authentic.		quite successfully.		

[15]	Ahmad Homsi	Detecting Twitter Fake Accounts using Machine Learning and Data Reduction Techniques	Through the application of machine learning algorithms and various methods of data reduction, the purpose of this work is to identify phoney Twitter accounts. On the internet, it is not uncommon to find fake identities, which can then be exploited for a variety of nefarious objectives. The primary objective of this work is the development and improvement of machine learning algorithms that can detect bogus accounts in an effective and efficient manner. The MIB Twitter Dataset is used to train four different machine learning algorithms (J48, Random Forest, Naive Bayes, and KNN), as well as two different data reduction techniques (PCA and Correlation). The findings present a comprehensive analysis of	J48, Random Forest, Naive Bayes, and KNN are some examples of machine learning algorithms. PCA and correlation are two examples of data reduction techniques.	The purpose of this research is to illustrate how successful it is to combine the Correlation algorithm with the Random Forest algorithm in order to detect phoney accounts on Twitter. The paper achieves an accuracy of 98.6%.	- High accuracy of 98.6%	Data management is low

			how well various algorithms function, with a combination of correlation and random forest obtaining an accuracy of 98.6%.				
[16]	Fatema A. Sarhan	Fake Accounts Detection in Online Social Networks using Hybrid Machine Learning Models	This study's goals are to (1) identify phoney accounts in online social networks (OSNs) and (2) evaluate and evaluate the results of various machine learning approaches. The use of social networks online has skyrocketed in popularity across the globe; nevertheless, these platforms are also susceptible to the development of fraudulent and malevolent accounts, which can result in the dissemination	Hybrid machine learning methods	Through the utilisation of hybrid machine learning models, the research was able to identify bogus accounts in online social networks with an accuracy of 92.14%.	- High accuracy of 92.14%	Hybrid algorithm complexity is high

			of misleading or damaging information. This problem is going to be tackled head-on by the research project by utilising hybrid machine learning models for the purpose of fraudulent account detection.				
[17]	Amna Kadhim Ali	Fake Accounts Detection on Social Media using Stack Ensemble System	Finding bogus accounts on the social media site Twitter will be the focus of this research project. This study intends to present a method for detecting fake accounts utilising a stack ensemble system with preprocessing techniques and supervised machine learning algorithms. Given the extensive use of social media and the accompanying hazards of damaging information through fake accounts, the purpose of this study is to propose a mechanism for detecting fake accounts..	The Spearman correlation coefficient, the chi-square test, and the ensemble system (also known as the stack approach).	When it came to identifying bogus accounts on Twitter, the stack ensemble system obtained a detection accuracy of 99%..	- Utilises a stack ensemble system, which mixes numerous machine learning methods in order to improve detection performance. Stack ensembles are also known as "ensembles of ensembles." - Incorporates preprocessing approaches to extract useful features, which has the potential to improve the false account identification accuracy. - Obtains a high accuracy of 99%, which indicates outstanding ability in identifying bogus accounts.	- Because there is so little information supplied regarding the particular implementation of the stack ensemble system, it is difficult to evaluate its robustness and determine whether or not it has any potential limits. - The extent to which the proposed mechanism is applicable to other social media sites is not addressed in this discussion. - There is no consideration given to the possibility of bias or limitations associated with the dataset that was used.

[18]	Voitovych Olesia	Detection of Fake Accounts in Social Media	This study's purpose is to identify false accounts on social media platforms, with a particular emphasis on Facebook as its primary target. In times of war, fake identities are frequently utilised for the purpose of conducting cyberattacks and manipulating information. The research analyses a wide range of metrics and characteristics that are associated with fake accounts on Facebook, and it ranks those characteristics according to the degree to which they contribute to the fakeness of the accounts..	Supported vector machine (SVM), decision-making system	By employing a specifically crafted dataset consisting of actual and fake account attributes, the decision-making system that was constructed, which was based on SVM, was able to reach an accuracy of 97% when detecting fraudulent accounts.	- Gives a methodical examination of measurements and qualities that contribute to the detection of fraudulent accounts, which enables a thorough understanding of the issue. - Makes use of a decision-making system that is predicated on SVM, a potent machine learning algorithm. - Accomplishes a high level of accuracy in the detection of bogus accounts, with a score of 97%, indicating effective performance.	Because there is so little information available regarding the particulars of how the decision-making system is implemented, it is impossible to evaluate both its robustness and any potential limits it may have. The extent to which the established approach can be applied to other types of social media platforms is not addressed in this discussion. The dataset that was utilised is not specified in depth, which may potentially limit the reproducibility of the results as well as their application.
[19]	Dr Vijay Tiwari	Analysis and Detection of Fake Profile over Social Network	The purpose of this study is to investigate and identify phoney accounts that can be found on social networks, with a particular emphasis on social engineering as the primary source of	ML models literature review	ML models	- This article presents an overview of approaches and techniques for detecting fraudulent profiles in existing social networks. It also offers insights into the current state of	- Because the exact methods and strategies that are employed for phoney profile detection are not disclosed in detail, comprehension of the suggested solution is limited. Because the

			security risks. This paper provides an overview of a variety of methodologies and machine learning techniques that can be helpful in identifying and analysing bogus profiles. It draws attention to the growing frequency of phoney profiles and the possibilities for information manipulation and covert agendas that these profiles can facilitate.			research in this area. - The importance of methods that use machine learning in identifying and analysing social bots is emphasised here. - Educates people about the growing risk posed by phoney profiles, as well as the possibility for information theft and manipulation that these profiles can provide.	accuracy and performance parameters of the detection methods are not discussed, determining how well they work can be difficult. - The study does not contain any new or original research; rather, it focuses on reviewing already established methodologies . As a result, it is possible that it does not offer any fresh perspectives or innovative contributions to the subject.
[20]	Er. Pranay Meshram	Automatic Detection of Fake Profile Using Machine Learning on Instagram	Using techniques from machine learning, the purpose of this paper is to identify phoney profiles on the photo-sharing platform Instagram. The article acknowledges the increase in cybercrimes, particularly those targeting women, and relates the growth to the influence of social media platforms such as Facebook, Instagram, and Twitter. Specifically, the research	Logistic Regression, Random Forest	Accuracy: 95.5%	This article addresses the problem of phoney profiles on Instagram, which can have detrimental effects, both for individuals and for society as a whole. Detects bogus profiles automatically by employing methods based on machine learning in order to streamline the procedure. - Identifies bogus profiles with a high	Because the particular features and datasets that were utilised for the training of the machine learning models are not given, our knowledge of the methodology is restricted. The research is limited to Instagram, therefore it is possible that false profiles on other social media platforms would go unnoticed. - There is no consideration given to any

			focuses on the rise in cybercrimes targeting women. Using logistic regression and the random forest algorithm, the purpose of this work is to determine whether or not Instagram accounts are phoney or genuine.			degree of precision, achieving a rate of 95.5% accuracy.	of the possible drawbacks or difficulties associated with the approach of machine learning, which leaves room for additional research and development. - The study does not provide any insights on the potential impact of detecting and eliminating false profiles on Instagram users and the general user experience of the platform.
[21]	Edward Kwadwo Boahen	Detection of Compromised Online Social Network Account with an Enhanced Knn	By addressing the constraints of the various machine learning algorithms that are currently in use, the purpose of this research is to identify accounts on online social networks (OSNs) that have been compromised. In this study, the difficulties associated with discovering hacked accounts are discussed, and an improved approach to machine learning known as Word Embedding	Word Embedding, KNN	Accuracy: 99.98% for recall, 99.97% for precision	- Addresses the principal risk posed by hijacked accounts on online social networks. - Presents an innovative strategy that combines word embedding and KNN to enhance accuracy and feature selection. - High levels of accuracy are achieved, with recall rates of 99.98% and precision rates of 99.97%. - Includes a comparison and analysis with other	Because the particular specifics of the word embedding technique and the dataset that was used are not published, the understanding of the strategy is restricted. The limitations and difficulties of the currently available machine learning approaches are noted in passing but are not explored in depth here. It is mentioned that there is an optimisation procedure for

			and KNN (WE-KNN) is proposed as a solution. For the purpose of determining how well the suggested model performs, it is assessed with the help of benchmark datasets and compared to several other methods.			methods that are comparable, proving that the findings have been improved.	the purpose of further improving accuracy, but no further explanation is provided. There is no mention made of the possible impact that recognising and mitigating compromised accounts could have on OSN users and on the network's security.
[22]	Putra Wanda	DeepProfile: Finding fake profile in online social network using dynamic CNN	Using a dynamic CNN technique, the purpose of this article is to search for and identify false accounts that are present in online social networks. In this paper, a deep neural network approach known as DeepProfile is proposed. DeepProfile makes use of a dynamic CNN and a unique pooling layer in order to achieve increased performance. When compared to traditional learning algorithms, the studies show encouraging results, with improved accuracy and a lesser loss in	Dynamic CNN with WalkPool pooling layer	Accuracy: AUC = 0.9547	- Enables enhanced feature extraction and representation through the application of deep learning algorithms. - This paper presents a dynamic CNN architecture that has an innovative pooling layer. When compared to traditional learning algorithms, it achieves a higher level of accuracy while incurring a lower level of loss. - Illustrates the performance by making use of evaluation criteria such as the area under the curve (AUC)	Because the particular specifics of the dynamic CNN architecture and the implementation of the pooling layer are not revealed, the understanding of the approach is restricted. The conversation on potential changes and new directions for the future may be more in-depth and laser-focused. - It is not emphasised how important it is to identify fraudulent accounts on online social networks or the potential applications of the DeepProfile algorithm. - Only a limited

			the detection of bogus profiles.			and the ROC curve.	comparison is made with other methods that are considered state of the art in false profile detection. - The dataset that was used and how well it represented real-world circumstances are neither stated nor explained.
[23]	Faouzia Benabbou	Fake accounts detection system based on bidirectional gated recurrent unit neural network	By analysing the content of users' tweets, the purpose of this research is to identify false accounts that have been created on social networking sites. The system that has been suggested use the bidirectional gated recurrent unit (BiGRU) model in conjunction with the global vectors (GloVe) word embedding technique in order to determine whether or not Twitter user accounts are authentic. Evaluation and comparison of the accuracy and precision of the technique with baseline	Bidirectional Gated Recurrent Unit (BiGRU)	Accuracy: 99.44% Precision: 99.25%	- Makes use of a model known as a bidirectional gated recurrent unit (BiGRU), which has been demonstrated to be useful in sequential data analysis. - Captures semantic and syntactic context by utilising the global vectors (GloVe) word embedding technique. - Delivers superior results in terms of accuracy and precision compared to baseline models such as LSTM and CNN. - Pays attention to the content of the tweets that users post because this	Because the particular implementation details of the BiGRU model and the GloVe word embedding technique are not revealed, one's comprehension of the strategy is hindered. Neither the dataset that was utilised nor the degree to which it was indicative of actual-life situations were provided. Accuracy and precision are the only two measures that are considered for use in the review process; other metrics, such as recall and F1 score, are not even mentioned. -

			models such as LSTM and CNN.			information can be quite helpful in identifying bogus accounts.	This article does not provide a comparison to other methods that are considered to be state-of-the-art in the identification of bogus accounts. - There is no discussion of the possible restrictions and difficulties associated with the approach that has been offered.
[24]	Vernika Singh	Preventing Fake Accounts on Social Media Using Face Recognition Based on Convolutional Neural Network	Utilising face recognition techniques that are underpinned by convolutional neural network (CNN) algorithms, the purpose of this article is to find a way to put a stop to the production of false accounts on various social media sites. Face recognition, prediction, classification, and clustering are some of the methods that will be utilised in this study with the end goal of reducing the number of fraudulent profiles and making them	Convolutional Neural Network (CNN)	Good precision	Face recognition methods built on convolutional neural networks (CNN) are utilised in this system, which enables for improved detection of fraudulent profiles. - Places an emphasis on the prevention of the establishment of phoney profiles, which is one way to help make the online world safer. - Improves the accuracy of false profile identification by utilising a variety of methods including	Because the correctness of the approach that is being recommended is not disclosed, determining whether or not it is effective might be challenging. Because the particular CNN architectures and methodologies that are utilised in face recognition are not defined, our comprehension of the technology is restricted. It is not disclosed whose dataset was utilised for training and testing the model, which raises issues about how

			easier to see. The correctness of the approach that has been proposed is not specified and will be determined by subsequent experimentation on its implementation..			prediction, classification, and clustering. - Addresses the specific issue of phoney accounts that are prevalent across several social media networks.	representative the hypothetical situations are of the real world. The comparison with other bogus profile eradication methods, such as existing systems or methodologies, is not provided. - There is no discussion of the possible restrictions and difficulties associated with the approach that has been offered.
[25]	Putra Wanda	RunMax: fake profile classification using novel nonlinear activation in CNN	The purpose of this research is to address the problem of false accounts in online social networks (OSN) by presenting a convolutional neural network (CNN) method named RunFake as a potential solution to the problem. In order to achieve more precision when classifying phoney accounts, the suggested approach integrates a recently developed activation function with the name	Convolutional Neural Network (CNN)	Accuracy: Precision = 94.00, Recall = 93.21, F1-Score = 93.42, AUC = 0.9547	- Improves the accuracy of the false profile classification process by adding a new activation function to the last layer of the CNN model. This function is called RunMax. - Makes use of user profile data as characteristics, which might provide helpful information for determining which accounts are authentic and which are false. - Indicates that	The specifics of the RunMax activation function, such as its formulation and features, are not revealed, which makes it more difficult to comprehend the benefits that it offers in comparison to more conventional activation functions. - It is not disclosed whose dataset was utilised for training and testing the model, which raises issues about how representative

			RunMax into the last layer of the CNN model. It has been reported that the proposed method has an accuracy of 94.00 for Precision, 93.21 for Recall, and 93.42 for F1-Score, with an area under the curve (AUC) score of 0.9547 when using user profile data as features.			the proposed method is effective by reporting promising accuracy metrics such as precision, recall, F1-score, and area under the curve (AUC). - Tackles the particular problem of phoney profiles that are prevalent in online social networks (OSN).	the hypothetical situations are of the real world. - Because there is no offered comparison with other methods or approaches for fake profile classification, it is difficult to evaluate the originality and competitiveness of the suggested RunFake methodology. - There is no discussion of the possible restrictions and difficulties associated with the approach that has been offered.

III. CONCLUSION

To summarise, the proliferation of digital platforms has been accompanied by an increase in fraudulent behaviour. This activity includes the establishment of phoney user accounts as well as spamming. This is particularly relevant to the former case. Comprehensive detection and prevention methods are essential because the hazards that these attacks represent to the honesty and safety of digital platforms necessitate their implementation. In this regard, the ability of machine learning algorithms to learn from data, spot patterns, and make accurate predictions offers a significant amount of promise.

Traditional machine learning techniques, such as decision trees, support vector machines, and logistic regression, have found broad use in a variety of contexts. On the other hand, more complicated deep learning models like convolutional and recurrent neural networks have shown increasing success in dealing with the intricacies of online user behaviour and spamming methods. Examples of these models include neural networks.

Despite this, challenges continue to persist. The dynamic and ever-evolving nature of spamming activities, deceptive schemes, data imbalance, and privacy problems, among other things, all create significant challenges that must be overcome. As a direct response to this issue, a variety of distinct educational methods, such as federated learning, active learning, and transfer learning, have been developed. These techniques are intended to protect the privacy of users while also resolving the issues of insufficient data and outmoded models. Both the interpretability and the transparency of these detection systems stand to benefit from the application of explainable artificial intelligence (AI), which also has the potential to improve both.

When looking to the future, the exciting new advances in machine learning, such as algorithms that protect users' privacy and reinforcement learning, give tremendous prospects for future research. These tactics could be helpful in stimulating the development of detection systems that are flexible and robust, and that manage user data in an ethical and responsible manner.

In the end, detecting and blocking spammers and false users is a complex undertaking that is constantly evolving and requires for ongoing study and invention. This is because spammers and false users can take many forms. This evaluation's goals are to promote and direct similar endeavours by providing a comprehensive picture of the existing scene, highlighting the challenges that are currently being met, and pointing towards potential future chances. In

the rapidly evolving digital environment in which we live, it is necessary that we make a coordinated effort as a community to battle spammers and phoney users. This will ensure that the integrity of the system, the trust of its users, and its security will be maintained. In order to be successful, this fight requires the involvement of researchers, industry professionals, and politicians.

REFERENCES

[1] Gomathy, C K. (2021). The spammer detection and fake user identification on social networks. *International Journal for Research in Applied Science and Engineering Technology*, 9. DOI: 10.22214/ijraset.2021.38760.

[2] Ud Din, Ikram, & Masood, Faiza & Ammad, Ghana & Almogren, Ahmad & Abbas, Assad & Khattak, Hasan Ali & Guizani, Mohsen. (2019). Spammer detection and fake user identification on social networks. *IEEE Access, 7*, 1-14. DOI: 10.1109/ACCESS.2019.2918196.

[3] Paravel, Alaguvathana & Rangasamy, Suganya & Ilampiray, P & Sriraam, Natarajan & Premkumar, M. (2021). Identification of spammer and fake accounts on social networks. *Journal of Physics: Conference Series, 1916*. 012095. DOI: 10.1088/1742-6596/1916/1/012095.

[4] Koggalahewa, Darshika & Xu, Yue & Foo, Ernest. (2022). An unsupervised method for social network spammer detection based on user information interests. *Journal of Big Data, 9*. DOI: 10.1186/s40537-021-00552-5.

[5] Gupta, Arushi & Kaushal, Rishabh. (2015). Improving spam detection in Online Social Networks. *Proceedings - 2015 International Conference on Cognitive Computing and Information Processing*, DOI: 10.1109/CCIP.2015.7100738.

[6] Kondeti, Priyanka & Yerramreddy, Lakshmi Pranathi & Pradhan, & Swain, Gandharba. (2021). *Fake account detection using machine learning*. DOI: 10.1007/978-981-15-5258-8_73.

[7] Sreenivasa Rao, Kuncham & Gutha, Sreeram & Raju, Bandrapalli. (2020). Detecting fake account on social media using machine learning algorithms. *International Journal of Control and Automation, 13*, 95-100.

[8] Khaled, Sarah & El-Tazi, Neamat & Mokhtar, Hoda. (2018). *Detecting fake accounts on social media*, 3672-3681. DOI: 10.1109/BigData.2018.8621913.

[9] Singh, Naman & Sharma, Tushar & Thakral, Abha & Choudhury, Tanupriya. (2018). *Detection of fake profile in online social networks using machine learning*. DOI: 10.1109/ICACCE.2018.8441713.

[10] Harish, K. & Kumar, R. & Bell J, Briso Becky. (2023). Fake profile detection using machine learning. *International Journal of Scientific Research in Science, Engineering and Technology*, 719-725. DOI: 10.32628/IJSRSET2310264.

[11] Hakimi, Ahmad & Ramli, Suzaimah & Wook, Muslihah & Zainudin, Norulzahrah & Hasbullah, Nor & Wahab, Norshahriah & Matrazali, Noorafiza. (2019). *Identifying fake account in facebook using machine learning*. DOI: 10.1007/978-3-030-34032-2_39.

[12] Elyusufi, Yasyn & Elyusufi, Zakaria & M'hamed, Aït Kbir. (2019). *Social networks fake profiles detection based on account setting and activity*. 1-5. DOI: 10.1145/3368756.3369015.

[13] Mallampeta, Mamatha & Datta, M & Ansari, Umme & Shaik, Subhani. (2021). fake profile identification using machine learning algorithms, *11*, 60-65. DOI: 10.9790/9622-1107036065.

[14] Elyusufi, Yasyn & Elyusufi, Zakaria & M'hamed, Aït Kbir. (2020). *Social networks fake profiles detection using machine learning algorithms*. DOI: 10.1007/978-3-030-37629-1_3.

[15] Homsi, Ahmad & Al-Nemri, Joyce & Naimat, Nisma & Kareem, Hamzeh & Al-Fayoumi, Mustafa & Abu Snober, Mohammad. (2021). *Detecting twitter fake accounts using machine learning and data reduction techniques*. 88-95. DOI: 10.5220/0010604300880095.

[16] Sarhan, Fatema & Mattar, Ebrahim. (2023). Fake accounts detection in online social networks using hybrid machine learning models. *International Journal of Simulation: Systems, Science & Technology, 24*. DOI: 10.5013/IJSSST.a.24.02.02.

[17] Kadhim, Amna & Abdullah, Abdulhussein. (2022). Fake accounts detection on social media using stack ensemble system. *International Journal of Electrical and Computer Engineering, 12*, 3013-3022. DOI: 10.11591/ijece.v12i3.pp3013-3022.

[18] Kupershtein, Leonid & Voitovych, Olesya & Vitalii, Holovenko. (2022). Detection of fake accounts in social media. *Cybersecurity Education Science Technique, 2*. DOI: 10.28925/2663-4023.2022.18.8698.

[19] Tiwari, Vijay. (2017). *Analysis and detection of fake profile over social network*. DOI: 10.1109/CCAA.2017.8229795.

[20] Meshram, Pranay & Bhambulkar, Rutika & Pokale, Puja & Kharbikar, Komal & Awachat, Anushree. (2021). Automatic detection of fake profile using machine learning on instagram. *International Journal of Scientific Research in Science and Technology*, 117-127. DOI: 10.32628/IJSRST218330.

[21] Boahen, Edward & Wang, Changda & Elvire, Bouya-Moko. (2020). Detection of compromised online social network account with an enhanced knn. *Applied Artificial Intelligence, 34*, 1-15. DOI: 10.1080/08839514.2020.1782002.

[22] Wanda, Putra & Jie, Huang. (2020). DeepProfile: Finding fake profile in online social network using dynamic CNN. *Journal of Information Security and Applications*, *52*, 102465. DOI: 10.1016/j.jisa.2020.102465.

[23] Benabbou, Faouzia & Boukhouima, Hanane & Sael, Nawal. (2022). Fake accounts detection system based on bidirectional gated recurrent unit neural network. *International Journal of Electrical and Computer Engineering (IJECE)*, *12*, 3129. DOI: 10.11591/ijece.v12i3.pp3129-3137.

[24] Singh, Vernika & Shanmugam, Dr Raju & Awasthi, Saatvik. (2021). *Preventing fake accounts on social media using face recognition based on convolutional neural network*. DOI: 10.1007/978-981-15-9509-7_4.

[25] Wanda, Putra. (2022). RunMax: fake profile classification using novel nonlinear activation in CNN. *Social Network Analysis and Mining*, *12*. DOI: 10.1007/s13278-022-00983-9.