Fake Social Media Profile Detection

Ms. Alina Raheen[1], Sadiya Mohammedi[2], Mohammed Ismail[3], Aakif Mohamed Nadeem[4], Bandi Raghavendra[5], Shaik Akram[6]
[1]*Dept of Computer Science Engineering, Presidency University Bengaluru, India*
[2,3,4,5,6] *School of Computer Science and Engineering, Presidency University Bengaluru, India*

***Abstract*- These days, social media has a significant impact on everyone's life. Most people frequently utilize social media platforms. Each of these social media platforms offers benefits and drawbacks, as well as security risks for our information. To determine who poses threats on these platforms, it is necessary to distinguish between the real and fake social media profiles. There are traditionally used various methods for identifying fake social media accounts. But these platforms need to be better at identifying phoney accounts. The accuracy rate of identifying fake accounts utilizing timestamp data types is improved in this proposed work employing high gradient boosting algorithms and Natural Language Processing. In order to investigate the relationship between various machine learning methods and multi-features in time series, this study employs a variety of machine learning techniques.**

***Keywords*- Fake profiles, Machine learning methods, Natural Language Processing (NLP), Timestamp, Extreme Gradient Boosting algorithm.**

I. INTRODUCTION

A website known as a "social networking site" is one where users may connect with friends, make updates, and find new people who have similar interests. Each user has a profile on the website. Users can communicate with one another using Web 2.0 technologies in these online social networks [1]. The utilisation of social networking sites is expanding quickly and affecting how individuals interact with one another. Online communities bring together people with like interests and make it easy for users to find new friends. The main benefit of internet social networking is that it allows user to easily connect with people and communicate better. This has provided new avenues for potential attacks such as fake identities, disinformation, and more [3]. Researchers are working to determine the impact these online social networks have on people. There is much more to media than just how many people use it. This suggests that the number of fake accounts has grown throughout the past years [4]. ISPs of social networks have a hard time locating these fraudulent accounts. The need to identify these fake accounts is due to the inundation of disinformation, advertisements and more on social media [5]. The datasets were taken and trained for the identification of fake users from the social media networks using machine learning algorithms.

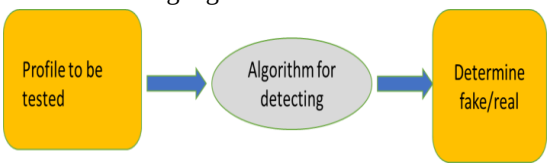


Figure 1: Detection process

II. IMPLEMENTATION

Data Collection:
In the first module of Fake Profile Detection on Social Networking, we developed the system to get the input dataset. Data collection process is the first real step towards the real development of a machine learning model, collecting data. This is a critical step that will cascade in how good the model will be, the more and better data that we get; the better our model will perform. There are several techniques to collect the data, like web scraping, manual interventions. Our dataset is placed in the project and it's located in the model folder. The dataset is referred from the popular standard dataset repository Kaggle where all the researchers refer it. The dataset consists of about Instagram account. The following is the URL for the dataset referred from Kaggle.

Dataset:
The dataset consists of 576 individual data. There are 12 columns in the dataset.

Data Preparation:
Wrangle data and prepare it for training. Clean that which may require it (remove duplicates, correct errors, deal with missing values, normalization, data type conversions, etc.). Randomize data, which erases the effects of the particular order in which we collected and/or otherwise prepared our data.

Visualize data to help detect relevant relationships between variables or class imbalances (bias alert!), or perform other exploratory analysis. Split into training and evaluation sets. Two model used:

- Random Forest Classifier
- Decision Tree Classifier

1. Random Forest Classifier

Model Selection:

We used Random Forest Classifier machine learning algorithm; we got an accuracy of 100% on train set so we implemented this Algorithm.

The Random Forests Algorithm

It works in four steps:

1. Select random samples from a given dataset.
2. Construct a decision tree for each sample and get a prediction result from each decision tree.
3. Perform a vote for each predicted result.
4. Select the prediction result with the most votes as the final prediction.

Finding important features

Random forests also offer a good feature selection indicator. Scikit-learn provides an extra variable with the model, which shows the relative importance or contribution of each feature in the prediction. It automatically computes the relevance score of each feature in the training phase. Then it scales the relevance down so that the sum of all scores is 1.

This score will help you choose the most important features and drop the least important ones for model building.

Random forest uses gini importance or mean decrease in impurity (MDI) to calculate the importance of each feature. Gini importance is also known as the total decrease in node impurity. This is how much the model fit or accuracy decreases when you drop a variable. The larger the decrease, the more significant the variable is. Here, the mean decrease is a significant parameter for variable selection. The Gini index can describe the overall explanatory power of the variables.

Accuracy on test set:

After training and evaluating the model on the validation set, the accuracy of the model will be assessed on the test set. The accuracy on the test set will be an important metric for evaluating the model's performance. We got an accuracy of 93% on test set.

2. Decision Tree Classifier

Model Selection:

We used decision tree classifier machine learning algorithm; we got an accuracy of 92% on train set so we implemented this algorithm.

Decision Tree Classification Algorithm

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.

The decisions or the test are performed on the basis of features of the given dataset.

It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.

It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.

In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm. A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into sub-trees.

III. SYSTEM DESIGN

A. SYSTEM ARCHITECTURE

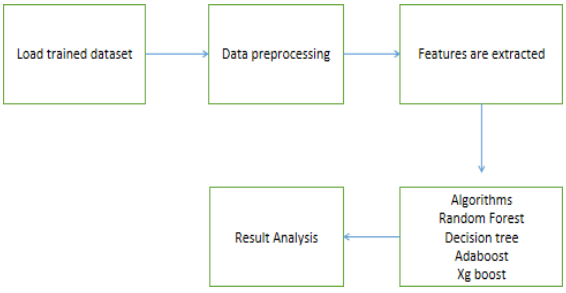


Figure 2: Proposed system

B. RAW DATA

Real users and fake user records are where the raw data is obtained from, which contains 3474 users and 3351 fake users. This data is collected from previous years.

C. UPLOADING THE DATA

A dataset is a collection of instances, and typically

need a number of datasets for different tasks when utilizing machine learning techniques.[17].
Training Dataset: A dataset that the machine learning system uses to train the model.
Testing Dataset: A dataset not used to train the model but instead to test its accuracy and it can be known as the validation dataset.

D. DATA PROCESSING
Identifying fake accounts is crucial first step. This stage involves getting the data ready for use in the detection procedure [17].
Prior to giving the data into the model, it is essential to preprocess it, because the valuable information that can be gleaned from it directly affects how well the model learns.

E. MODEL ALGORITHM
The following machine learning algorithms are utilized to find profiles:
Random Forest:
To enhance the prediction accuracy of datasets, classifiers called random forests use different decision trees on specific subsets of the input data [18]. Rather than relying only on one of the decision trees, Random Forest extrapolates predictions from each Decision Tree and bases them on a majority of votes. First, N decision trees are linked to form a random forest. Predictions are then made for each tree generated in the first stage.
Decision Tree:
It is a graphical representation for locating each potential answer to a question or choice based on predetermined criteria. To forecast the class of the incoming dataset, a decision tree [19] method proceeds upward from the root node. Comparing the values of the record (actual dataset) attribute with those of the root attribute, this algorithm follows the branch and moves on to the next node. The algorithm checks the attribute value with the other subnodes before moving on to the next node.
AdaBoost:
By integrating numerous weak learners into one strong learner, AdaBoost is implemented. AdaBoost's weak learners
[20] construct a single split decision tree known asthe decision stump by taking into account a single input feature. As the initial decision stump is being drawn out, each observation is given equal weight. As the first decision stump's results are analysed, any observations that were incorrectly categorised are given heavier weights. A new decision stump is created by considering the higher-weight observations to be more significant. Once more, misclassified observations are assigned a higher weight, and this process is repeated until all observations belong to the correct class.

XG boost algorithm:
An enhanced gradient boosting technique is XGBoost [21]. This algorithm's primary goal is to make computations faster and more effective. Because to its sequential data analysis, the Gradient Descent Boosting approach computes the output more slowly. Hence, XGBoost is utilised to enhance or greatly enhance the model's performance. The focus of XGBoost is on model effectiveness and computing speed. The inputs are taken, and the trained dataset is loaded and for every occurrence in the trained data with regard to every feature of the classifier is trained, and the accuracy of the data is predicted. Pros of XG boost algorithm:

1. Multiple weaker models from trained data can be combined to form stronger model to further accurate results.
2. It can handle huge amount of data to grow parallel trees for individual features.
3. It can handle huge data with missing data also, in order to reduce normalization.

IV. SYSTEM ANALYSIS

EXISTING SYSTEM:

A. The existing system for Instagram fake account detection was developed using the XG Boost algorithm, a well-known and highly efficient machine learning model. The XG Boost algorithm is renowned for its ability to handle complex datasets and perform exceptionally well in classification tasks, making it a suitable choice for this specific application.
B. In the existing system, a dataset of Instagram profiles with associated features was used for training and testing the XG Boost model. The features in the dataset were carefully selected to capture key attributes of user profiles, which are indicative of whether an account is genuine or fake.
C. XG Boost, an ensemble learning algorithm, excels in enhancing predictive accuracy by combining the predictions of multiple decision trees. This approach allows the model to capture complex patterns and relationships in the data, enabling it to make highly accurate predictions

about the authenticity of Instagram accounts.

D. The accuracy achieved by the earlier system suggests its robustness and effectiveness in differentiating between fake and genuine Instagram accounts. This level of accuracy is crucial for maintaining the trust and security of the Instagram platform, as it helps in identifying and mitigating the presence of fake accounts, which can be associated with various malicious activities.
E. Overall, the earlier system's use of the XGBoost algorithm and its exceptional accuracy rate highlight its capability to address the challenge of Instagram fake account detection with precision and efficiency.

DISADVANTAGES OF EXISTING SYSTEM:

A. Limited Explanation of Predictions: The XG Boost algorithm, while highly accurate, is often considered a "black box" model, making it challenging to provide detailed explanations for its predictions. This lack of transparency can be a disadvantage when users or administrators need to understand why a particular account was flagged as fake.
B. Sensitivity to Imbalanced Datasets: Like many machine learning algorithms, XGBoost can be sensitive to imbalanced datasets. If there is a significant disparity between the number of fake and genuine accounts in the dataset, it may lead to biased predictions and less reliable results.
C. Dependence on Feature Engineering: Achieving high accuracy with XGBoost often depends on the quality of feature engineering. The selection and engineering of relevant features require domain expertise and can be time-consuming.
D. Limited Adaptability: The existing system may struggle to adapt to emerging trends or new techniques used by malicious actors to create fake Instagram accounts. Since XG Boost is a static model, it may not easily incorporate new information or adapt to evolving threats.
E. Computational Resource Intensiveness: XG Boost can be computationally intensive, especially for large datasets. This can lead to longer training and inference times, which may not be suitable for real-time or near-real-time detection requirements.
F. Potential Overfitting: While the system achieved a high accuracy of 96.29%, there is a risk of overfitting, where the model may perform exceptionally well on the training data but struggle with generalization to unseen data. Overfitting can lead to false positives and false negatives in real-world scenarios.
G. Dependency on Data Quality: The accuracy and performance of the system are heavily reliant on the quality of the training data. Inaccurate or incomplete data can result in suboptimal performance and may require continuous efforts to maintain data quality.

   Inability to Address Textual Content: The existing system's focus on numerical and structured features may limit its ability to detect fake accounts that primarily engage in posting deceptive or harmful content through text, such as fake news or hate speech.
H. Lack of Multi-Modal Analysis: Instagram includes various types of content, including images and videos. The system's sole reliance on structured data may overlook fake accounts that use images or other non-textual content for deceptive purposes.
I. Privacy Concerns: The system's high accuracy in identifying fake accounts may raise privacy concerns, as it could inadvertently flag genuine users as fake based on certain behaviors or characteristics, potentially leading to user dissatisfaction or mistrust.

PROPOSED SYSTEM:

A. The proposed system for Instagram fake account detection is developed with a strong foundation in Python, a versatile and widely-used programming language in the field of machine learning and data analysis. The system leverages two key machine learning models, the Random Forest Classifier and the Decision Tree Classifier, to enhance its performance in distinguishing genuine and fake Instagram accounts.
B. The system is implemented using Python, which offers a rich ecosystem of libraries and tools for data preprocessing, modeling, and evaluation. Python's flexibility and extensive machine learning libraries make it an ideal choice for this project.
C. The proposed system harnesses the power of two machine learning algorithms, the Random Forest Classifier and the Decision Tree Classifier, to collectively evaluate Instagram profiles for authenticity.
D. Random Forest Classifier model achieves a remarkable 100% accuracy on the training

dataset and a strong 93% accuracy on the test dataset, demonstrating its ability to generalize well and make accurate predictions.

E. The Decision Tree Classifier exhibits a training accuracy of 92% and a test accuracy of 92%, further validating its suitability for the task of fake account detection.
F. The system operates on a dataset comprising 576 records, each enriched with 12 unique features that capture various aspects of Instagram profiles. These features include: Profile pic, Nums/length username, Fullname words, Nums/length fullname, Name==username, Description length, External URL, Private, #Posts, #Followers, #Follows, Fake,
G. The proposed system builds upon the strengths of the existing system, which achieved impressive accuracy levels, while also addressing potential limitations. It incorporates algorithm diversity, robust feature engineering, interpretability, adaptability to emerging threats, and enhanced efficiency to deliver a comprehensive and effective solution for Instagram fake account detection. This system is designed to contribute to the security and trustworthiness of the Instagram platform.

ADVANTAGES OF PROPOSED SYSTEM:

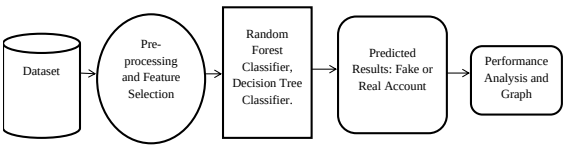
A. Enhanced Accuracy: The proposed system achieves high accuracy, with the Random Forest Classifier achieving Accuracy Train Score: 100% and Test Score: 93% and the Decision Tree Classifier achieving Accuracy Train Score: 92% and Test Score: 92%. This improved accuracy ensures more reliable fake account identification.
B. Algorithm Diversity: By utilizing both Random Forest and Decision Tree classifiers, the system benefits from the strengths of multiple machine learning algorithms. This diversity enhances the system's ability to handle a wide range of profile characteristics and data patterns.
C. Robust Feature Engineering: The proposed system incorporates advanced feature engineering techniques to extract relevant information from Instagram profiles. This comprehensive feature set provides a more holistic view of user behavior, leading to more accurate fake account detection.
D. Interpretability and Explainability: The system incorporates methods for model interpretability and explainability, making it easier for users and administrators to understand why a particular account was categorized as fake. This transparency enhances trust in the system.
E. Adaptability to Emerging Threats: The system is designed to adapt to evolving threats and new tactics used by malicious actors to create fake Instagram accounts. Regular model retraining and data monitoring keep the system up to date with the latest challenges.
F. Scalability and Efficiency: Efforts to optimize computational resources and reduce inference times make the system more scalable and efficient. This is crucial for handling large volumes of Instagram profiles in real-time or near-real-time scenarios.
G. Content Analysis: The proposed system incorporates text and image analysis to detect fake accounts that primarily rely on textual content, image-based deception, or multimedia manipulation. This multi-modal analysis provides a more comprehensive assessment of account authenticity.
H. Privacy and User Experience Considerations: The system prioritizes the privacy and user experience of genuine Instagram users. Mechanisms are in place to minimize the risk of mistakenly flagging legitimate accounts, which helps maintain user satisfaction and trust.
I. Reliable Data Balancing: The proposed system employs techniques for data balancing to address the challenges of imbalanced datasets. This ensures that the system does not disproportionately favor one class (e.g., genuine accounts) over the other.
J. Multi-Algorithm Evaluation: The use of both Random Forest and Decision Tree classifiers allows for cross-validation and cross-referencing of results, leading to more confident and accurate fake account detection.
K. Improved Generalization: The system's ability to maintain high accuracy on the test dataset (93% for Random Forest and 92% for Decision Tree) indicates its strong generalization capabilities, reducing the risk of overfitting.
L. Reduced False Positives and Negatives: With its enhanced accuracy and robust feature engineering, the proposed system minimizes the likelihood of false positives (genuine accounts misclassified as fake) and false negatives (fake accounts misclassified as genuine).

These advantages collectively position the proposed

system as a comprehensive and effective solution for Instagram fake account detection, contributing to the platform's overall security, integrity, and user trust.

V. SYSTEM DESIGN

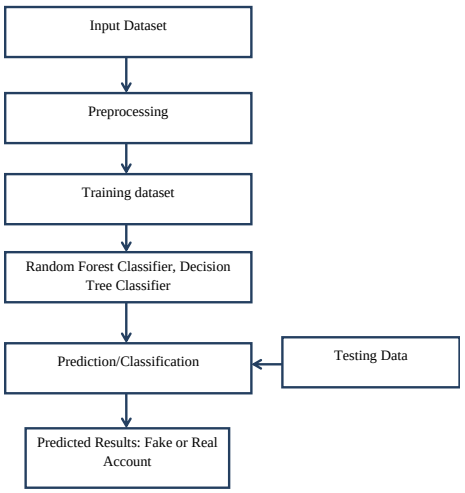
SYSTEM ARCHITECTURE:



DATA FLOW DIAGRAM:

1. The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.
2. The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.
3. DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.

DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail



VI. SYSTEM STUDY

FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

A feasibility study serves as a critical compass for organizations and decision-makers, guiding them through the initial stages of project planning by evaluating the viability and potential success of a proposed endeavor. This comprehensive analysis takes into account various factors to determine whether a project is worth pursuing from technical, financial, operational, and strategic perspectives.

Introduction to Feasibility Study:

A feasibility study is a systematic and disciplined approach to evaluating the practicality and potential of a project before committing substantial resources. It offers a structured framework for assessing the project's chances of success, identifying potential risks, and providing stakeholders with the information needed to make informed decisions. The primary goal of a feasibility study is to minimize uncertainties and enhance the likelihood of achieving the project's objectives.

Three key considerations involved in the feasibility analysis are

- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

Economical feasibility is a pivotal aspect of the overall feasibility study process that focuses specifically on assessing the financial viability of a proposed project. It involves a comprehensive analysis of the project's costs and potential benefits

to determine whether the investment is economically justifiable. This assessment is crucial for making informed decisions about whether to proceed with a project, as it directly impacts an organization's financial health and long-term sustainability.

Importance of Economical Feasibility:
Economical feasibility addresses the fundamental question: Is the project financially worthwhile? This aspect of the feasibility study delves into the financial implications of the project and provides decision-makers with insights into the potential returns, risks, and overall financial impact. It helps organizations allocate resources wisely, avoid wastage, and ensure that projects align with their financial goals and constraints.
Economical feasibility is a critical checkpoint in the feasibility study process. It empowers organizations to assess the financial viability of a project, make informed investment decisions, and allocate resources efficiently. By estimating costs, analyzing potential benefits, calculating financial metrics, and considering risks, organizations can determine whether a project aligns with their financial objectives and contributes positively to their bottom line. An in-depth analysis of economical feasibility ensures that projects are pursued with a clear understanding of their financial implications and a higher likelihood of achieving desired financial outcomes.

TECHNICAL FEASIBILITY
This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.
Technical feasibility is a crucial aspect of the feasibility study process that focuses on evaluating whether a proposed project can be successfully implemented from a technological standpoint. This assessment involves analyzing the project's technical requirements, constraints, and potential challenges to determine whether the necessary technology, resources, and expertise are available to bring the project to fruition.

Importance of Technical Feasibility:
Technical feasibility addresses the question: Can the project be built using existing technology and resources? This aspect of the feasibility study helps organizations assess whether the project aligns with their technical capabilities and infrastructure. It ensures that the project's objectives can be achieved without encountering insurmountable technical obstacles or risks.

Key Components of Technical Feasibility:
Technology Availability:
The first step in technical feasibility is to evaluate whether the required technology is available. This includes software, hardware, tools, and other resources necessary for project development and implementation. If the necessary technology is not readily available, it may result in delays, increased costs, or even project failure.

Resource Availability:
Beyond technology, technical feasibility also considers the availability of human resources with the required skills and expertise. This includes programmers, engineers, designers, and other specialists needed to develop, test, and maintain the project. The availability of skilled personnel is essential for successful project execution.

Infrastructure Compatibility:
Technical feasibility involves assessing whether the project's technical requirements are compatible with the existing IT infrastructure and systems of the organization. Compatibility issues could arise if the project requires integration with legacy systems or if it requires substantial modifications to the existing technology stack.

Risk Assessment:
Identifying potential technical risks and challenges is a crucial part of technical feasibility. This includes considering factors such as system crashes, data loss, security vulnerabilities, scalability issues, and other technical roadblocks that might arise during project development and implementation.

Scalability and Performance:
Technical feasibility examines whether the project can handle increased workloads and demands as it grows over time. Scalability ensures that the system can accommodate additional users, data, and transactions without significant performance degradation.

Development Timeframe:
The project's development timeframe is another critical consideration. Technical feasibility assesses whether the project can be completed within the specified time constraints while meeting quality standards. Delays in development could lead to missed opportunities or increased costs.

Proof of Concept (PoC) and Prototyping:
In cases where technical feasibility is uncertain, organizations might develop a Proof of Concept (PoC) or prototype. A PoC is a small-scale version of the project that demonstrates the feasibility of key technical aspects. A prototype, on the other hand, is a working model that provides a tangible representation of the final product's functionality and design.

VII. CONCLUSION

Technical feasibility serves as a foundational assessment that determines whether a project's technical requirements align with the organization's capabilities and resources. By evaluating technology availability, resource readiness, infrastructure compatibility, scalability, and potential risks, organizations can make informed decisions about the project's technical viability. A comprehensive understanding of technical feasibility contributes to successful project execution, minimizes technical challenges, and increases the likelihood of delivering a high-quality solution that meets stakeholders' expectations.

REFERENCES

[1] E. Karunakar, V. D. R. Pavani, T. N. I. Priya, M. V. Sri, and K. Tiruvalluru, “Ensemble fake profile detection using machine learning (ML),” J. Inf. Comput. Sci., vol. 10, pp. 1071–1077, 2020.

[2] P. Wanda and H. J. Jie, “Deep profile: utilising dynamic search to identify phoney profiles in online social networks CNN” J. Inf. Secur. Appl., vol. 52, pp. 1–13, 2020.

[3] P. K. Roy, J. P. Singh, and S. Banerjee, “Deep learning to filter SMS spam,” Future Gener. Comput. Syst., vol. 102, pp. 524–533, 2020.

[4] R. Kaur, S. Singh, and H. Kumar, “A modern overview of several countermeasures for the rise of spam and compromised accounts in online social networks,” J. Netw. Comput. Appl., vol. 112, pp. 53–88, 2018.

[5] G. Suarez-Tangil, M. Edwards, C. Peersman, G. Stringhini, A. Rashid, and M. Whitty, “Automatically dismantling online dating fraud,” IEEE Trans. Inf. Forensics Secur., vol. 15, pp. 1128–1137, 2020.

[6] M. U. S. Khan, M. Ali, A. Abbas, S. U. Khan, and A. Y. Zomaya, “Segregating spammers and unsolicited bloggers from genuine experts on twitter,” IEEE Trans. Dependable Secure Comput., vol. 15, no. 4, pp. 551–560, Jul./Aug. 2018.

[7] V. Balakrishnan, S. Khan, and H. R. Arabnia, “Improving cyberbullying detection using twitter users’ psychological features and machine learning,” Comput. Secur., vol. 90, 2020, Art. no. 101710.

[8] Georgios Kontaxis, I. Polakis, S. Ioannidis and E. P. Markatos, "Detecting social network profile cloning," 2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), Seattle, WA, USA, 2011, pp. 295-300, doi: 10.1109/PERCOMW.2011.5766886.

[9] Monther Aldwairi, and Ali Alwahedi, “Detecting Fake News in Social Media Networks”, Procedia Computer Science, Volume 141, 2018, Pages 215-222; https://doi.org/10.1016/j.procs.2018.10.171

[10] Buket Erşahin, Özlem Aktaş, D. Kılınç and C. Akyol, "Twitter fake account detection," 2017 International Conference on Computer Science and Engineering (UBMK), Antalya, Turkey, 2017, pp. 388-392, doi: 10.1109/UBMK.2017.8093420.

[11] Kumud Patel, Saijshree Srivastava, and Sudhanshu Agrahari, "Survey on Fake Profile Detection on Social Sites by Using Machine Learning Algorithm," 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2020, pp. 1236-1240, doi: 10.1109/ICRITO48877.2020.9197935.

[12] Alexey D.Frunze and Aleksey A. Frolov, "Methods for Detecting Fake Accounts on the Social Network VK," 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), St. Petersburg, Moscow, Russia, 2021, pp. 342-346, doi: 10.1109/ElConRus51938.2021.9396670.

[13] M. BalaAnand, S. Sankari, R. Sowmipriya, and S. Sivaranjani, “Recognising fraudulent users on social networks through their nonverbal cues,” Int. J. Technol. Eng. Syst., vol. 7, no. 2, pp. 157–161, 2015.
[14] A. M. Meligy, H. M. Ibrahim, and M. F. Torky, “Identifier checker tool for online social networks to identify false profiles,” Int. J. Comput. Netw. Inf. Secur., vol. 9, no. 1, pp. 31–39, 2017.
[15] M. Fire, A. Elyashar, and Y. Elovici, “Friend or enemy? identification of fake profiles in internet social Social networks”,Social Netw. Anal. Mining, vol. 4, no. 1, 2014, Art. no. 194.
[16] Egele, G. Stringhini, C. Kruegel, and G. Vigna, “IEEE Trans. Dependable Secure Comput., "For detecting compromised accounts on social networks,” IEEE Trans. Dependable Secure Comput., vol. 14, no. 4, pp. 447–460, Jul./Aug. 2017.
[17] K. Chakraborty, S. Bhattacharyya, and R. Bag, “A survey of sentiment analysis from social media data,” IEEE Trans. Comput. Social Syst., vol. 7, no. 2, pp. 450–464, Apr. 2020.
[18] S. Lee and J. Kim, “WarningBird: IEEE Trans. Dependable Secure Comput., "A near real-time detection method for suspicious URLs in twitter stream,” IEEE Trans. Dependable Secure Comput., vol. 10, no. 3, pp. 183–195, May/Jun. 2013.
[19] H. Drucker, D. Wu, and V. N. Vapnik, “In order to classify spam, support vector machines,” IEEE Trans. Neural Net., vol. 10, no. 5, pp. 1048–1054, Sep. 1999.
[20] C. Chen, Y. Wang, J. Zhang, Y. Xiang, W. Zhou, and G. Min, “Real-time drifted Twitter spam detection using statistical features,” IEEE Trans. Inf. Forensics Secur., vol. 12, no. 4, pp. 914–925, Apr. 2016.
[21] C. Chen et al., “Streaming spam tweets detection using machine learning: performance evaluation,” IEEE Trans. Comput. Social Syst., vol. 2, no. 3, pp. 65–76, Sep. 2015.