FAKE SOCIAL MEDIA ACCOUNTS AND THEIR DETECTION.

Submitted by,

Jakku Nishithaa - 20211CCS0034
Hiranmayi R - 20211CCS0153
M Sonali - 20211CCS0017
Bhavana M M - 20211CCS0026
S M Maaz - 20211CCS0009

Under the guidance of,

Dr. Sharmasth Vali Y

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING, Cyber Security

At

PRESIDENCY UNIVERSITY

BENGALURU

MAY 2025

PRESIDENCY UNIVERSITY

PRESIDENCY SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the Internship/Project report This is to certify that the Project report **"Fake Social Media Accounts and Their Detection"** being submitted by "Jakku Nishithaa, Hiranmayi R, M Sonali, Bhavana MM, Satharla Mohammed Maaz" bearing roll number(s) "20211CCS0034, 20211CCS0153, 20211CCS0017, 20211CCS0026, 20211CCS0009" in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering, Cyber security is a bonafide work carried out under my supervision.

Dr. Sharmasth Vali Y
Associate Professor
PSCS
Presidency University

Dr. S P Anandaraj
Professor & HoD
PSCS
Presidency University

Dr. MYDHILI NAIR
Associate Dean
PSCS
Presidency University

Dr. SAMEERUDDIN KHAN
Pro-Vice Chancellor
Dean –PSCS / PSIS
Presidency University

PRESIDENCY UNIVERSITY

PRESIDENCY SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

DECLARATION

I hereby declare that the work, which is being presented in the report entitled "**FAKE SOCIAL MEDIA ACCOUNTS AND THEIR DETECTION"** in partial fulfillment for the award of Degree of **Bachelor of Technology** in **Computer Science and Engineering**, **Cyber security,** is a record of my own investigations carried under the guidance of **Dr. Sharmasth Vali Y, Associate Professor, Presidency School of Computer Science and Engineering, Presidency University, Bengaluru.**

I have not submitted the matter presented in this report anywhere for the award of any other Degree.

Nishithaa Jakku – 20211CCS0034
Hiranmayi R – 20211CCS0153
M Sonali – 20211CCS0017
Bhavana MM -20211CCS026
Satharla Mohammed Maaz -20211CCS0009

ABSTRACT

OSNs have become the backbone of actual social lives, allowing individuals to keep in touch with one another, send updates, host events, and sometimes even conduct business. Such is the case with Instagram, Facebook, and TikTok, which have grown exponentially in terms of user base and today run into billions as they make use of these tools for communication. A very rapid expansion like this, however, works in the favor of malicious actors who would abuse such networks to steal personal data, spread misnomers, or commit fraudulent activities. Fake accounts, witnessed with greater frequency now than used to be, are a point of emphasis for both users and organizations alike; hence, research has developed methods of identifying and dealing with such threats via advanced algorithms and machine learning techniques.

Researchers targeting this problem have pointed out fake account detection as possible through the classification algorithms and account features evaluating their activities against normal behavior patterns. Nevertheless, some features are counterproductive or neutral as far as detection results are concerned. Simple algorithms fail to provide significant results alone and should therefore be integrated with other techniques such as feature selection and dimension reduction. Recently, on the problem of fake accounts found in Instagram, three machine learning algorithms-Support Vector Machine (SVM), XG Boost, and Naive Bayes-were implemented to classify accounts as real or fake easily. The essence of this research was reduced features without compromising the accuracy level, XG Boost achieved an amazing performance with an accuracy of 91% in training data.

This implementation would penetrate possible avenues through which optimizing feature selection as well as sophisticated classification methods can improve detection of fake accounts. The considered SVM, XG Boost, and Naive Bayes have, by behavior patterns and accounts characteristics, demonstrated efficiency in identifying fraud. Yet, there has been progress worth mentioning and yet challenges lie ahead in the detection mechanisms which do not rule out the change of such by imposters. Research into hybrid methods employing multiple algorithms as well as enhancing the feature selection process needs to be continually done, for the benefit not only of sharpening social network security but also maintaining trust amongst users.

ACKNOWLEDGEMENTS

First of all, we indebted to the **GOD ALMIGHTY** for giving me an opportunity to excel in our efforts to complete this project on time.

We express our sincere thanks to our respected dean **Dr. Md. Sameeruddin Khan**, Pro-VC - Engineering and Dean, Presiency School of Computer Science and Engineering & Presiency School of Information Science, Presidency University for getting us permission to undergo the project.

We express our heartfelt gratitude to our beloved Associate Dean **Dr. Mydhili Nair,** Presidency School of Computer Science and Engineering, Presidency University, and **Dr. S P Anandaraj,** Head of the Department, Presidency School of Computer Science and Engineering, Presidency University, for rendering timely help in completing this project successfully.

We are greatly indebted to our guide **Dr. Sharmasth Vali Y, Associate Professor** and Reviewer **Ms. Priyanka Niranjan Savadekar**, **Assistant Professor**, Presdiency School of Computer Science and Engineering, Presidency University for his/her inspirational guidance, and valuable suggestions and for providing us a chance to express our technical capabilities in every respect for the completion of the internship work.

We would like to convey our gratitude and heartfelt thanks to the PIP4001 Internship/University Project Coordinator **Mr. Md Ziaur Rahman and Dr. Sampath A K,** department Project Coordinators **Dr. Sharmasth Vali Y** and Git hub coordinator **Mr. Muthuraj.**

We thank our family and friends for the strong support and inspiration they have provided us in bringing out this project.

Jakku Nishithaa

Hiranmayi R

M Sonali

Bhavana M M

Satharla Mohammed Maaz

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

Chapter 1
INTRODUCTION

1.1 Background.

The phenomenal rate at which online social networks (OSNs) have developed has indeed redefined the modes of interaction, networking, sharing of information, and business undertakings. Today, the likes of Instagram, Facebook, and TikTok have become indispensable parts of everyday life, allowing people to communicate, collaborate, and present themselves. Yet, this proliferation has brought about its share of trouble-especially concerning the many fake accounts. Most of these accounts are malicious, involving threatening or dangerous behavior, spreading false information, impersonation of individuals or organizations, or even fraud activity. Such problems Jeopardize the trust and security of users in social media; thus, efficient detection mechanisms are imperative. Researchers are now extremely inclined to use advanced technology for these problems to create a safer online world.

This detection of fake social media accounts is done by examining the features of accounts and applying certain classification algorithms to the suspicious pattern. But there are limitations that currently exist in the compared approaches, resulting from the consideration of irrelevant or negatively contributing features as well as from independent algorithms' deficiencies. To counter such drawbacks, researchers have proposed the merging of feature selection and dimension reduction techniques in conjunction with machine learning algorithms to fulfill better detection accuracy. For example, for such studies that aimed to make a detection of fake Instagram accounts, certain algorithms such as Support Vector Machine (SVM), XG Boost, and Naive Bayes have been used to differentiate accounts as either real or fake. These methods put priorities on not spending so much using features but those significant to attain high accuracy rates. For example, XG Boost even achieved a remarkable 91% percentage of accuracy for fraudulent detections within training datasets. Such results revealed the power of incorporating algorithmic precision with optimal feature selection to improve detection systems.

1.2 Approach.

The initiative to detect fake social media accounts marks a watershed moment in the attempt to mature online security against fictive users and the authenticity such approaches would produce thereby providing a more secure environment for better user trust and social media integrity. Such elimination of fraudulent accounts allows real users to engage without running the risk of exposure to nefarious deceivers, which greatly reduces misinformation and other fraud. Thus, accounts would be analyzed under the real-time advanced machine learning algorithms and data analysis techniques, which automatically detect and flag suspicious profiles based on a wide range of behavioral and content-related features. It will develop a robust detection system, which can be integrated within existing social media platforms to enhance automated detection as users participate in the network.

The concocted fake account detection system employs the potential of a combination of Support Vector Machine (SVM), XG Boost, and Naive Bayes algorithms to classify or analyze accounts. These algorithms work with feature selection and dimension reduction techniques for real-time processing of account behavior or characteristics. While creating an account or showing suspicious activity, the system identifies and classifies it based on parameters such as observed or known posting patterns, networks of friends, and content authenticity. Following this process, the system flags it for attention or takes some automated action, such as restricting activity or removing it from the site. Because of a centralized management system that continually processes account data, the detection mechanisms are kept current with new tactics employed by creators of fake accounts.

The technological components of this system include data mining tools for extracting suitable account features, machine learning models for classification, and a robust database for storage and analysis of historical account behavior data. The profiles fed into the detection model are analyzed using natural language processing (NLP) techniques for content dimensioning and behavior detection, including automated or inauthentic patterns associated with profile activity, as well as network analysis tools to find their relationship with the other accounts to understand the coordinated networks of fake profiles. Additionally, this project will help set a precedent for future innovations in online identity verification.

1.3 Problem statement.

Dealing with the emergence of phony social media accounts, which can injure the trustworthiness of users and the integrity of the platform. These accounts promote false information, impersonate individuals, and carry out traffic fraud. Traditional detection methods are not sufficient: every year, billions of fake accounts are identified but remain undetected with the help of machine learning and data mining, along with very interesting approaches involving natural language processing. The system would provide real-time monitoring of suspicious profiles, which in turn increases security and enhances user experience: maintains trustworthiness in an online setting selves: innovations in approach with the tactics changes will ensure that one can go against any fake by taking long-term measures to maintain digital safety. This invention would provide real-time monitoring of suspicious profiles, which actually add security and enhance user experience by maintaining trustworthiness in an online environment. Changes in the way tactics would have been appliedments would have secured long-term measures against fake ones and digital security would have been maintained.

Chapter 2
LITERATURE SURVEY

2.1 Historical Development.

2.1.1 Early Emergence and Misuse.

The emergence of the use of dummy social media accounts can be traced back to the infancy of internet platforms. Those accounts were then normally required for innocent purposes: experimenting with alternate personas or engaging in harmless fun. However, the misuse of fake accounts became a serious problem as the platforms gained popularity in the late 2000s and early 2010s. Malcirminated actors be the emergence of the use of dummy social media accounts can be traced back to the infancy of internet platforms. Those accounts were then normally required for innocent purposes: experimenting with alternate personas or engaging in harmless fun. However, the misuse of fake accounts became a serious problem as the platforms gained popularity in the late 2000s and early 2010s. Malcirminated actors began to actively use such accounts for fraud, thereby raising a need for investigative techniques, thoroughly discharging enforcement actions on protecting the users and the platform's integrity to actively use such accounts for fraud, thereby raising a need for investigative techniques, thoroughly discharging enforcement actions on protecting the users and the platform's integrity.

2.1.2 Early Detection Efforts

In previous times, categorizations of spam accounts would use basically heuristics such as inculcating activity patterns that were either suspicious in character for the accounts or high in follower numbers. While these methods brought some level of achievement, the changes in the patterns of behavior introduced by malicious actors provided the means to considerably undermine these methods. The level of sophistication with the technology during the era was thus really not advanced, as there was no practical approach towards consideration of maybe machine learning algorithms that could have gone some way in establishing a proactive posture. For this reason, these systems remained largely reactive towards responding to known threats than threatening any possibility of affecting new emerging ones.

2.1.3 Advancements in Detection Techniques.

In the mid-2010s, researchers started adopting machine learning algorithms into detection systems, a game changer in countering fake accounts. Effective algorithms such as Support Vector Machine (SVM) or Naive Bayes were introduced to check various account features as: posting frequencies, network connections, etc. Some of these approaches improved the accuracy of detection dramatically. Unfortunately as well, other unrelated or negatively contributory features caused numerous problems. Integration with machine learning was thus an important step toward sophisticated detection mechanisms.

2.1.4 Feature Selection and Integration of Advanced Algorithms.

Recent advancement has focused on improving the detection of fake social media accounts through the application of feature selection and dimension reduction technique so that the systems will work effortlessly on large databases whilst maintaining a high accuracy percentage. With the right selection of features and reduced dimensionality of data, the systems can easily recognize patterns that could be construed as those of a fake account, thereby decreasing the number of false positives encountered and improving detection accuracy levels. Those studies that specifically target fake Instagram accounts have shown how well algorithms such as XG Boost with optimized feature sets are able to achieve an accuracy of up to 91% in classifying accounts as fraudulent. Coupled with this, the areas of NLP have also been utilized to examine content authenticity and work to further boost existing capabilities in detecting fraudulent accounts. The advancement is a very crucial step in combating the rising menace of fake social media accounts.

2.2 Real-Time Detection Mechanisms.

2.2.1. Machine Learning Approaches in Cybersecurity.

Real-time malware detection systems increasingly rely on machine learning models like Support Vector Machines (SVM) and Decision Trees (DT) to analyze behavioral patterns in network traffic, system processes, and executable files. These models excel at identifying zero-day attacks and polymorphic malware variants, with SVM particularly effective in handling high-dimensional data structures.

2.2.2. Computer Vision and Object Detection Networks.

The so-called machine learning algorithms which include the Support Vector Machines (SVM) and Decision Trees (DT) are involved in the frontline of possible detection to weigh traffic for behavioral signs across the network, processes of the system, and the code in executables to determine the existence of possible immediate threats, such as zero-day kinds of attacks and polymorphism. High-dimensional data structures are effectively handled by SVM, by which malicious software and benign software are separated by hyper plane classification. DT hierarchically predicts what kind of behavior a malware will exhibit from the features observed. Both methods dynamically adapt to evolution, offering strong solutions against classical signature-based methods. The features gathered are automated. Most importantly, they operate in real-time, strengthening cyber security defenses for less detection and response time.

Algorithm.	**Accuracy (mAP).**	**Speed (FPS).**	**Use Case.**
YOLOv5.	68.9	140	Autonomous driving.
SSD.	74.3	59	Surveillance.
RetinaNet.	70.4	24	Medical imaging.

Table 2.1: "Performance comparison of real-time object detectors."

2.2.3. Post Level Trust Analysis. (PLTA)

PLTA is a key to securely protecting systems through dynamic verification of the integrity of operations at granular levels, such as user authentication, API calls, or application interactions. This method guards by providing real-time detection of anomalies or unauthorized activity. This is achieved by not considering an entire profile or span of history but by focusing on specific posts themselves. Therefore, detection is more fine-grained with respect to insider attacks or compromised endpoints. It works in close coordination with much other broader trust frameworks and complements them, such as Zero Trust, by strengthening security on the level of the micro-interactions.

2.2.4. Profile Level Trust Analysis. (PrLTA)

The Profile Level Trust Analysis (PrLTA) determines the trustworthiness of individual user profiles within any system. Some of the contributing factors for static trust evaluation are given to user behavior, access patterns, and contextual information. Trustworthiness here is said to be dynamic in nature. This kind of analysis will assist in identifying risks like account compromise and insider threats by means of anomaly detection with respect to user activity. PrLTA mostly considers the interaction between historical data and real-time monitoring so as to give a clear view of any user's security posture. After segmenting users into Trusted and Untrusted categories, organizations can create security policies such as operating access control and increased authentication requirements based on this trust level. This methodology is instrumental in identity-first approaches to cybersecurity and helps improve threat detection as well as feed into decision-making on the Zero Trust framework.

2.2.5. Profile Historic Trust Analysis. (PHTA)

The PHTA is, therefore, a conceptual framework evaluating the trustworthiness of a profile historically through historical data analysis. Indeed, it would appear to review user profile or system profile historical behaviours, interactions, and security incidents aimed at identifying such historical patterns or anomalies. Through analyzing the historical trust level variance on given profiles, the PHTA should provide organizations with a basis for predicting possible future threats. Another advancement in this analysis might include machine-learning algorithms focused on the automation of the trend and anomaly detection in historical data, thereby enriching and allowing the perspective for active security measure predictions. PHTA will become relevant and applicable as historical insights also corner the improvement of trust-based security architectures.

2.2.6. Deep Neural Networks. (DNNs)

The presence of fake profiles on social networks can be detected using Deep Neural Networks (DNNs) which would be able to analyze all the intricate features and patterns shown in user data. The techniques that the models accomplish this task with include using Convolutional Neural Networks (CNNs) to process profile and user behavior such as connections and post characteristics along with content-based attributes. A very recent study suggested that the use of a DNN trained on a minimum amount of profile information could

outperform classical methods enormously. Optimizing the extraction and classification tasks of feature information can be formed by the combination of gradient descent with cross-entropy loss functions by DNNs. Thus, using this method will help find zero-day fake accounts with lesser numbers of false positives, enhancing the trustworthiness and authenticity of the various social platform.

2.3 Early Detection Methods.

2.3.1. Behavioral Analysis Techniques.

Fake profiles deviate from the norm regarding dynamics such as their follower growth rate, repetitive posting, or unearthly login hours. The machine-learning models tag accounts during sudden spikes in sending around an excess of 100 friend requests during an hour, while same messages or posts have been circulated across various platforms. The temporal analysis of post frequency separates bots from human users because an automated account can post continuously intermixed with patterns of human rest while the real human user would take intermittent rest before posting again. Ridiculous accounts having an ungodly follower-to-following ratio, e.g., 10,000 followers versus 10 following, are flagged for scrutiny on platforms such as Twitter. They also indicate low-engagement (likes/comments per post) fraudsters since bots cannot mask such misleading engagement signals very effectively. These methods provide means for early detection of behavioral anomalies and social interplay inconsistency.

2.3.2 Profile Attribute Scrutiny.

Fake profiles usually rely on pilfered or AI-photographed images or reverse image search tools or Convolutional Neural Networks (CNNs) achieving a shocking 92% in finding duplicated stock pictures or deep fake faces. All accounts without profile pics or low-resolution pics gets flagged for manual verification. A parallel stream of analysis through NLP models mines bio text for more discrepancies in basic terms: generic statements (e.g., "Love to travel") or contradicting information on, for example, age and occupation. Templated bios, typical for bot accounts, discovered through syntactical pattern analysis. LinkedIn and other platforms use high-end techniques to detect AI-generated profile photos achieving 99.6% accuracy at differentiating synthetic faces from real ones. Then this amalgamation is actually effective against both forms of visual and textual anomalies against fake accounts.

2.3.3 Network and Social Graph Analysis.

Fake accounts typically cluster together into densely populated groups with hardly any links to real users. They are then isolated from each other through graph-based community detection algorithms, such as Louvain, from social networks that will look for signatures of shared attributes like registration IPs or timestamps. Usually, these clusters tend to be homogeneous in profile information, which makes them distinct from non-fake user groups. Interaction mapping will show how real users can engage across multiple communities, whereas bots are relegated to isolated echo chambers. Facebook uses such interaction graphs to flag accounts for which over 95% of interactions are confined to one cluster, indicating a possibility of coordinated inauthentic behavior. Using clustering and interaction mapping enables platforms to detect and mitigate fake profiles at scale, thereby maintaining the integrity of social networking sites.

2.4 Human Role in Fake Account Detection.

Fake account detection- one of the most crucial areas of cyber security in relation to online services and social media is machine learning or artificial intelligence (AI). However, this would not have been achieved without the active involvement of humans in the effectiveness of the entire process. This article explores the various dimensions of the human role in the process of fake account detection.

2.4.1 Responding to Large-Scale Fraud.

In fact, even identifying clusters of fake accounts with shared device fingerprints or email domain identifiers is only an element in the context of large-scale fraud detection. Humans are essential in data correlation for visualizing the relationships between these clusters so that the unique attack patterns are understood. It pinpoints coordinated ways of fraud and establishes their sources. Also, the during-critical decisions like freezing suspicious accounts or the conduct of deeper investigations-thereby enabling responses necessary to avoid risks. Besides them, analysts design rules particularly for automated systems, which act exclusively against specific fraud behavior scenarios. Last, they employ tools that allow drag-and-drop rule creation and automated workflows. In short, expert human beings have combined their analytical prowess with evolving adaptive technologies to ensure detection and response strategies against that large-scale fraud are very robust.

2.4.2 Ethical Considerations.

Human involvement is very much important for ethical considerations in detecting fake accounts. Because of biased training data, automated systems may flag authentic people's behavior as otherwise, which is mitigated by human reviewers; whereas reviews minimize unfairness across different demographics failing to discriminate against any specific ones. Lastly, humans promote an open discussion by transparently communicating the criteria for detection, thus retaining user trust. The adoption of responsibility in the decisions made at each stage in the detection process allows for accountability. Hence ethical supervision by human beings ensures that detection systems are working in an impartial and responsible manner, thereby protecting them from biases and maintaining the integrity of the process itself. Such supervision becomes critical in fostering confidence in digital platforms.

2.4.3 Improving Detection Techniques.

This is because many of humans assist in improving techniques used for detection of fake accounts. With the help of development and studies, emerging types of fraud are studied, and their algorithms are updated to keep their pace with the coming threats. Feedback loops are other means where humans contribute information concerning the performances of algorithms in the accuracy enhancement process. Not to forget are those partnerships with AI developers concerning advanced technologies such as deep learning being integrated into the detection system. Those types of hybrid detection systems ensure humans and technological infrastructure work together to confront high-level fraud strategies. Hence, humans keep improving detection techniques to keep up a strong defense against fake accounts.

2.4.4 Challenges Faced by Humans.

Regardless of their significance in identifying fake accounts, humans are faced with a myriad of challenges. Manual analysis of large datasets seems tedious, resource-intensive, and hardly scalable. Advanced techniques requiring substantial analytical skill for detection and interpretation are often employed by coordinated fraud rings. Moreover, human biases impede decision-making and cause the results in detection to be inaccurate. Such biases can subsequently lead to false positives or false negatives, affecting confidence on detection. Dealing with such challenges will require technological support backed continuously by training and stringent quality control measures to keep the human input accurate and

impartial.

2.4.5 Future Role of Humans.

However, the role of humans in the detection of fraud accounts would evolve with AI. Humans will manage and improve automated systems. Most of these roles should be strategic, allowing people to set goals for the systems driven by AI and track performance information concerning objectives. Another out of several such roles would be ethical governance: ensuring that the production and implementation of the entire process comply with ethical standards with transparency and fairness. Another requirement is skill development: training analysts in more advanced technologies, such as neural networks and behavioral analytics. Humans will still be essential for maintaining trust and accountability in ensuring responsible and effective operation of AI systems. With this human-AI collaboration, the organization will build a solid and dependable fake account detection framework.

2.5 Case Studies and Real-World Implementations.

There are different types of studies and applications of machine-learning techniques aiming at detecting fake accounts, depending upon their behavior and profile attributes. Important techniques look at parameters such as follower count, active time, and engagement rate in classifying accounts as real or fake. The Twitter datasets have been helpful regarding the model performance in detecting fake profiles, but hybrid systems combining network analysis with language processing techniques have been introduced to improve upon that performance. Applications on Facebook, Instagram, and LinkedIn have been developed in real life to combat identity theft, phishing, and disinformation by detecting and deleting suspected accounts efficiently.

2.5.1 Instagram Metadata Analysis.

Instagram employs metadata analysis for impostor profile detection and extracted 17 key profile features from user data. Machine learning techniques, especially XGBoost classifiers, have been able to nearly reach astounding success rates of almost 100%. It analyzes some pertinent metadata such as the frequency of posting, followers-to-following ratios, engagement metrics, and methods of account creation. In analyzing these parameters, the models are trained to identify signs of abnormal behavior. Anomalies in follower counts or

sudden spikes in follower account creation might be an indication of a fraudulent account.The time of user activity also plays a factor in the analysis, which may earmark profiles for low posting rhythm or unusual constancy of engagement. The way described in Digital Medal-Analysis programs set up the detection of cheap accounts; this reduces the decadence of the general integrity of the delivered platform.

Metadata/Feature.	**Description.**	**Role in Detection.**
Frequency of Posting.	How often the user posts content.	Unusual frequency may signal automated or fake behavior.
Followers-to-Following Ratio	Comparison of followers to accounts followed	Extreme ratios can indicate suspicious accounts.
Engagement Metrics	Likes, comments, and interactions per post	Low or inconsistent engagement may suggest fakeness.
Account Creation Method	How the account was created (e.g., via app, web, automation)	Automated creation methods are red flags.
Follower Count	Total number of followers	Sudden spikes or anomalies can indicate fake accounts.
Following Count	Number of accounts the user follows	Following too many accounts is a common fake trait.
Sudden Follower Spikes	Rapid increases in follower numbers	Indicates possible use of purchased or bot followers.
Posting Rhythm	Regularity and timing of posts	Unnatural constancy or inactivity can be suspicious.
Time of User Activity	When the user is active (time of day, consistency)	Odd hours or overly consistent activity is a warning.
Profile Completeness	Presence of bio, profile picture, and other details	Incomplete profiles are often fake.

Image/Content Analysis	Reuse of images, image quality, or suspicious content	Repetitive or low-quality images may be fake.
Interaction History	Patterns in how the user interacts with others	Abnormal interaction patterns can flag fake profiles
Engagement Consistency	Consistency of likes/comments across posts	Inconsistent engagement is a potential indicator
Geolocation Metadata	Location data from posts	Mismatched or missing locations can be suspicious
Account Age	How long the account has existed	Very new accounts are more likely to be fake
Bio and Caption Analysis	Textual analysis of bio and captions for spam or unnatural language	Spammy or generic language is a red flag
Device Metadata	Information about devices used to access the account	Multiple or unusual devices may indicate automation

Table 2.2: "Key Instagram Metadata Features Used for Impostor Profile Detection."

2.5.2 Kaggle Social Network Dataset Analysis Using CNN

Modern convolutional neural networks have been used to target the Kaggle Social Network Fake Account Dataset and detect profiles that are not fake more efficiently than traditional machine learning models. A wealth of features in the dataset- from follower patterns to posting frequency to the supporting metadata- help the CNN accurately classify real profiles and fake ones. It lends itself to learning complex patterns from raw data to accurately identify anomalies that indicate fraudulent accounts. By such a measure, this shows that deep learning techniques are quite successful when it comes to handling large-scale datasets and that they also have contributed significantly towards increasing the detection rates on social network platforms. These results vindicated the usability of CNN in real-time applications to temper the spread of digital ecosystems with fake profiles.

2.5.3 Integro: Victim-Based Detection

Integro utilizes OSN-based victim analysis to thwart fake accounts, and it is an extremely

powerful software. The major difference between Integro and the traditional methods is that, unlike the traditional ones that only focused on user-level activities or the graph structure, Integro integrates victim prediction into the graph-based algorithms to enhance the accuracy of detection. It identifies victims-real users who interact, knowingly or unknowingly, with fake profiles-and uses their interaction as an important feature for classification. By applying lower weights to edges connected to predicted victims, fake accounts become easily isolated in the social graph. The user ranking and classification is successfully executed using a modified random walk algorithm, starting from real accounts, such that most of the genuine profiles receive a higher rank than the fakes. Deployed in Tuenti, the largest OSN in Spain, Integro achieved up to 95% precision in detecting fake accounts, a significant improvement over SybilRank.

2.5.4 Phantom Profile Detection

Burgeoning phantom profiles, mainly engineered for strategic gains in social games, have been satisfactorily discerned by means of classifiers trained on game activity data. A study by Nazir et al. has expounded on Facebook gaming applications where phantom profiles were identified by examining the disparities in user behavior and gaming patterns as opposed to genuine users. These classifiers scored a good grade in correctly identifying these phantom profiles, which imitate genuine user actions so as not to arouse suspicion. Yet another study reported 86.4% true-positives of phantom profile detection using gaming-activity data. This shows some specific context-sensitive features that can be used in improving detection greater in social gaming environments.

2.5.5 Artificial Neural Networks (ANNs)

Artificial neural networks were successfully adopted for the detection of a spammer fake profile's techniques when compared with traditional approaches, such as Naive Bayes and Random Forest, for this purpose. It is using features within complex patterns of user behavior and interaction data while making it more efficient for use in spam detection at social platforms where postings per an account might be taken into account among other activities being done, such as engagement rates or linguistics. Training those features will enable ANN systems to find minute anomalies regarding someone indicative of a fake account. It has been discovered with more accuracy to distinguish between spammers and the legitimate ones, thereby improving community security for online purposes. In addition

to that, adapting to evolving spam tactics makes artificial neural networks a valuable tool in maintaining integrity over time.

2.5.6 Sybil Attack Prediction

A deep-learning regression model has been developed and designed to forecast Sybil attacks by studying the user activity and network behavior. The model comprises three components: data harvesting systems, a feature extraction system, and a deep regression framework. Based on profile-based, content-based, and network-based characteristics, the system systematically evaluates user profiles to probe malicious accounts. It efficiently detects collusion among Sybil nodes when fake accounts send friend requests to other Sybils to amplify the network influence. The performance of the model was tested against some large scale data sets with an accuracy of as high as 94%. It is thus quite an effective model for preemptive identification of Sybil attacks in social media users. Such measures would secure platforms like Twitter against exploitation resulting from the dynamic nature of attack strategies and will also improve the trustworthiness levels of online networks.

2.5.7 Node Similarity Communication Matching (NSCM)

The NSCM algorithm was applied efficiently to understand any kind of clone kind of profiles from social networks. There exist many nodes from a communication pattern which directly recognizes the similar attributes of the accounts under study or certain repetitive behavior when interacting with other accounts. This approach works particularly well for revealing fake accounts that seem to be spammed because it creates copies of profiles that then work together to promote the spread of malicious activity. The NSCM algorithm uses graph-based techniques to assess the structural and behavioral connection between nodes for accurate fraud detection. The application of this framework on a large scale social network has shown a significant actual identification of cloned accounts and amounts towards spam mitigation. Security of the platform would therefore be raised in this aspect.

Chapter 3
RESEARCH GAPS OF EXISTING METHODS

3.1 Limited Real-World Data

Fake profile recognition research is hindered by imitation and obsolete datasets. These datasets usually do not address the strategies of fake profile creators as they evolve and may lead to a poor-performing model in the actual context. There is a need for further varied and recent, real-life data for training and testing detection models. Providing access to current datasets will facilitate the researchers to attune the models as per the present threats, thereby improving accuracy and robustness to highly ingenious fake profiles. Furthermore, diverse datasets can lessen some issues like data bias in detection systems so that the detection systems are robust across social media platform differences as well as user demographics.

3.2 Lack of Cross-Platform Analysis

A number of research studies have done on fake profile detection has mostly looked as individuals on individual social media platforms, ignoring the cross-platform nature of these profiles. There is a considerable gap on how fake profiles work simultaneously on across numerous platforms. Such gaps create an incomplete detection mechanism, because almost all scammers prefer cross-platform messaging features to hide themselves, for instance on Facebook and Instagram, where it is case in point of scams using both messaging systems. Tackling this gap involves creating detection methods that would analyze and correlate activities across different platforms, thereby making a more holistic approach to recognizing and managing fake profiles within the digital ecosystem.

3.3 Insufficient Consideration of User Experience

Machine learning models have proved to be quite impressive when it comes to detection of fake profiles, but no research-so-far on integrating user feedback and experience on such systems. Such incorporation of insight from people who have met or interacted with a fake profile would make the systems more precise in detecting these profiles, since it tracks real-life experiences. The victim's experiences could serve as valuable datasets for subtle behavioral cues and interaction patterns that algorithms might not capture. This may lead to detection systems that are more adaptive and responsive to increasingly innovated tactics

from scammers as feedback is received from the user. Such approach could yield stronger and more successful fake profile detection mechanisms across social networking sites.

3.4 Insufficient Real-Time Detection Capabilities

Most current methods for fake profile detection rely on post-facto analysis that focuses on already active accounts. However, urgent development of more efficient algorithms to catch fake profiles identically at the moment of their being created or made active is needed. Real-time detection is critical so that misinformation and scams cannot spread. Development of algorithms, capable of processing data stream quickly and accurately for the identification of suspicious activities, is thereby critical. Such a proactive course of action gives social platforms grounds to block fake profiles before they can do any harm, thereby ensuring user safety and the integrity of the platform. However, efficient real-time detection requires the use of cutting-edge technologies like GPU acceleration and cloud computing.

3.5 Limited Research on Prevention Strategies

Notwithstanding the advancements that have been made in terms of detecting fake profiles, there is a wide gap in the research on proactive measures to avoid their creation and proliferation. This can be achieved with an effective prevention program that would complement an already established detection system in dealing with the root cause of the issue. Proactive could be composing the following: enhanced verification for accounts, stricter account creation procedures, and community reporting mechanisms. Also, there is to be included awareness creation of the user on the risks posed from fake profiles as well as best practices for online interactions in order to lessen the attractiveness of making a fake one. Doing this would strengthen prevention systems-proliferation detection, thus adding an extra weapon into the defense line for social media at large regarding user safety and enhanced trust.

3.6 Inadequate Handling of Identity Theft Cases

Existing methods for detecting fake profiles often struggle to differentiate between genuinely fake accounts and cases where real identities have been stolen or compromised. This distinction is crucial, as compromised accounts may exhibit behaviors similar to those of fake profiles, leading to false positives. The inability to accurately identify truly

fraudulent accounts poses significant challenges, potentially resulting in the wrongful flagging of legitimate users whose identities have been hijacked. Addressing this gap requires developing more nuanced detection systems that can distinguish between maliciously created fake profiles and compromised real accounts, ensuring that only fraudulent activities are targeted while protecting innocent users.

3.7 Lack of Contextual Understanding

It is equally important but often neglected to consider information that more empowers contextual reasoning. Observation has it that humans consider contextual cues such as the coherence of profile information available or the appropriateness of interactions when detecting fake profiles. Detection systems greatly benefit from contextual analysis for improved precision in marking these subtle indicators. For instance, the temporal consistency of user interests and behaviors may be analyzed to distinguish between genuine and maliciously intended profiles. Hence detection systems integrating contextual information with quantitative analysis can mimic human judgment and detect fake profiles effectively for better platform protection.

Chapter 4

PROPOSED MOTHODOLOGY

4.1 Data Acquisition and Preparation

4.1.1 Data Source and Relevance

The dataset probably comes from Kaggle for this project is fundamental for fake account detection features to be used for building machine-learning models so that they can be trained to accurately detect fraud profiles. Normal variables in Kaggle datasets are account activity, metadata, and behavior patterns, which are further pre-processed and analyzed for extraction of useful insights. For example, Random Forest and Deep Neural Networks, supervised learning algorithms, have been reported to achieve high accuracy in detecting bot accounts when trained on Kaggle data. Such a dataset-based modeling will guarantee the sound performance of the model based on different features applied for the accurate classification of fake profiles, while it also addresses issues of data imbalance and changing fraud patterns.

4.1.2 Feature Extraction

Feature extraction for the purpose of differentiating between phony profiles and real ones primarily identifies the attributes that distinguish the former from the latter. Profile-based features include username length, bio completeness, existence/inexistence of a profile picture. Network features include follower/following ratio, network density, etc., as fake accounts always have weird social connections. Behavioral activity patterns, such as those relating to posting frequency and recency, are used to identify such kinds of suspicious behaviors as automated or repetitive acts of posting. Content analysis considers parameters related to similarity in posts and the total count of URLs in addition to text and/or image irregularities indicating anomalies. All these features provided to machine learning models to train and classify profiles improve security and reduce manual validation efforts.

4.1.3 Data Pre-processing

Data preprocessing is highly crucial for the preparation of data for machine-learning applications like fake account detection. The following main steps ensure a clean and consistent dataset. Missing value treatment is important so that gaps in data do not severely

affect the accuracy of the model. Noise removal eliminates the irrelevant or inaccurate data. Feature scaling ensures that all features are measured in the same scale to avoid any one feature having a larger influence on the analysis than it otherwise would. Coding categorical variables means transforming them into numerical formats for effective processing outcomes through algorithmic learning. These functionalities together truly help with data quality and reliability, leading to an increased accuracy of model outputs.

4.2 Feature Selection and Dimensionality Reduction

4.2.1 Importance of Feature Selection

The goal of feature selection in machine learning is to find the relevant features so that accurate classification can be achieved by it. Analysis of the dataset using techniques such as Recursive Feature Elimination (RFE), Lasso regression, and Chi-Square tests tends to retain only the most informative features. This reduces the size of the feature space, thus minimizing the computational complexity and risks of overfitting, compelling the model to focus on the meaningful patterns instead of noise. Good features not only increase the efficiency of the model but also lend interpretability and, in most cases, amplify the predictive accuracy. Methods such as forward and backward selection iteratively prune feature subsets to achieve optimal performance for wrapper methods, whereas embedded methods perform selection as part of the training process.

4.2.2 Role of Dimensionality Reduction

Indeed, dimensionality reduction of data could reduce the dimensionality of a high-dimensional dataset while preserving as much relevant information as possible. There are a number of well-known techniques that implement these aspects, such as Principal Component Analysis (PCA), t-Distributed Stochastic Neighbor Embedding (t-SNE), and Autoencoders. PCA finds directions of maximum variance and projects data onto them under a few dimensions; t-SNE is very effective in visualizing clusters in nonlinear data structures. Moreover, model complexity actually reduces with a smaller dimension since dimension reduction directly reduces the number of features used, thus speeding up training time and reducing the chances of overfitting because ofi-redundant or irrelevant features. Dimensionality reduction is especially beneficial when losing information with correlated variables or variables with sparse distributions. It, along with feature selection, builds better

generalization performance in machine learning models and improves the ability to visualize data better.

4.3 Model Selection and Training

4.3.1 Algorithm Choice

The project employs robust machine learning methodologies such as AdaBoost, CatBoost, and Extra Tree Classifier that have been chosen since they are basically known for dealing with complex data structures besides ensuring high accuracies in classification tasks. AdaBoost and CatBoost particularly deal efficiently with various features and categorical variables, while the Extra Tree Classifier is straightforward and fast. This document is deployment-oriented in terms of fake account detection realization by utilizing the power of such techniques to enhance the effectiveness of models developed towards improving performance and reliability. This project through these algorithms aims at making the fake account detection process accurate and efficient, thus developing a systematized approach towards identifying and dealing with fraudulent profiles.

4.3.2 Training Process

During this phase, selected models are trained on the prepared dataset to learn the patterns and relationships between features and target attributes-type-real or not-real. This includes optimizing the model parameters so that the errors are reduced and increases in performances by predicting., in terms of output. The models-weights and biases- will progressively fine-tune their understanding of the data by adjusting prediction on actual response. In the end, this model refinement process contributes to the adeptness of the models in being able to differentiate between actual accounts and fake accounts for true classification of new unseen data. This refinement is essential in the account detection of fakeness with maximum accuracy.

4.4 Model Evaluation and Testing

4.4.1 Evaluation Metrics

While accuracy is a popular measure to evaluate classification models, metrics such as precision, recall, and F1 score are critical for fraud detection because of imbalance nature of the class. Precision indicates how many fake accounts were correctly identified versus all

predicted fakes, thus reducing the negative impact of false positives on genuine users; recall indicates how well the model detects actual fake accounts, thereby reducing false negatives that would leave malicious profiles undetected. F1 score offers balance between these metrics and offers an overall assessment of the performance. Accuracy does not suffice in case of imbalanced dataset-every real account will outnumber one fake account as such-because this may overrate the model performance in terms of favoring the majority class. If precision and recall are given priority, then it represents robust detection whereby the chances of overlooking fraudulent activity or misclassifying legitimate users are curtailed to the greatest extent possible.

4.4.2 Testing Phase

The conducted tests with trained models consist of those performed on a separate test dataset to evaluate generalization capability with respect to unseen data. The test in question is used to confirm that the models did not overfit; that is, they did not simply memorize training-specific patterns but also learned generalizable underlying features. The ability of a model to identify fake accounts in a real-world scenario is said to validate its performance against new data. It pinpoints the deficiencies in model robustness and accountability for their deployment. Independent testing thus actually proves that these models can withstand high accuracy and efficacy out of the training atmosphere, making them ideal for real-time applications in the detection of fake accounts.

4.5 Challenges and Considerations

4.5.1 Algorithm Limitations

Although the traditional classification methods appear to give quite a strong baseline, these standalone models may suffer defeats in scenarios of fake account evolution, imbalanced dataset situations, or complex fraud scenarios. These techniques face challenges in keeping up with concept drift-the phenomenon whereby fraudulent behavior changes with any passing day-and may poorly interpret extravagant, contextual signals if taken alone. Hybrid approaches, which combine supervised models with unsupervised approaches (such as anomaly detection), graph-based network analysis, and/or deep learning architectures, have the potential to bridge these gaps. For instance, social network analysis through graph analytics and transformer models for semantic content analysis were used to improve the

detection of coordinated inauthentic behavior. Such hybrid platforms also build real-time feedback streams that generate the possibility of running dynamic updating of models. Such hybrid techniques-integration has immensely strengthened the models and minimized the numbers of false positives/negatives while addressing the multifaceted reality of today's fake accounts.

4.5.2 Evaluation Metrics

Accuracy is often highlighted as a primary evaluation metric, but relying solely on it can be misleading, especially in imbalanced datasets common in fake account detection. Metrics such as precision, recall, and F1-score provide a more comprehensive assessment of model performance. Precision measures the proportion of correctly identified fake accounts among all predicted fakes, helping to minimize false positives. Recall assesses the model’s ability to detect actual fake accounts, reducing false negatives and ensuring fewer fraudulent profiles go unnoticed. The F1-score balances.

Chapter 5
OBJECTIVES

5.1 Predicting Account Authenticity

Another challenge that comes up in predicting the authenticity of a social media account is that the rise of fake profiles across many platforms brings along various forms of threats, such as the spread of misinformation, scams, and the harming of reputations. Therefore, distinctive features are emphasized for separating genuine accounts from fake in a process that examines attributes of the profile, activity patterns, and network behaviors. Fake accounts are often viewed as bots-sock puppets, or names used to disguise identity-to deceive others, manipulate public opinion, or defraud them. Accordingly, those fake profiles generally exhibit unrealistic profile pictures or AI-generated ones, bios that may be seen as either incomplete or generic, unusual engagement patterns, fast follower growth, or repetitive posting behavior. Such counterintuitive behavior is used to flag accounts through manual and automated detection by historically studying those features.

Modern approaches to predicting account authenticity utilize advanced machine learning algorithms that deal with huge data volumes over time and identify subtle anomalies that easily evade human perception. The algorithms consider the real-time analysis of hundreds of parameters like username construction, activity frequency patterns, follower/following ratios, and uniqueness. Mostly feature selection and dimensionality reduction methods are applied for effective indicators and simplification of calculations. For instance, algorithms may be trained to recognize signatures of bots: those accounts that follow thousands of users but themselves have very few followers; or those accounts posting at a very regular interval, unlikely behavior for a human. Other areas for bolstering the accuracy of the detection are done by cross-referencing information across parallel platforms or reverse image searching for stolen or duplicated profile images.

The account authenticity prediction would be more prominent if the measure itself had a fuller evaluation scheme than mere thinking of accuracy measure. Precision, recall, and F1 - score are essential when it comes to seeing how a model punishes false positives (declaring real accounts as fakes) against false negatives (missing actual fakes), mainly in respect to the

context of class imbalance where real profiles would far outnumber the fraudulent ones. The continual upgrading of detection systems by timely integration of new data sources and refinement of algorithms to deal with challenges posed by sophisticated generation of fake accounts using AI-generated images or the simulation of human-like behavior must become their order of the day. Robust data analysis, working hand in hand with machine learning and vigilant monitoring, will allow social media platforms and users alike to safeguard their online communities better and to preserve the integrity of digital interactions.

5.2 Developing a Fast and Reliable Method

It is a reasonable and intelligent procedure that will be applied in order to develop the rapid and highly reliable detection method for fraudulent accounts, which is so much required in an age where social media becomes the continuous diversity of fake profiles. These accounts are capable of sprouting rumors, influencing the public, and duping by defrauding that emphasize their timely and accurate detection. This is to engineer a solution that does not only have high accuracy but also runs efficiently to process these data in real-time and large scale. Speed becomes the core since their traffic includes several millions of user interactions daily, which if delayed in detection can leak harmful content. The very reliability makes it act that true users also will not falsely adduce leading to a great depletion of user trust and platform credibility.

Such advanced and machine learning algorithms are those that balance speed and accuracy pamper training module in fact, the methodology adopted is designed considering these aspects. These are AdaBoost, CatBoost, and Extra Tree Classifier on good optimum success rates. Avoids through phase method of injecting complex data structures within a short time with stable predictive performances. Integrate strong data pre-processing stages and techniques of selection of definite features to cleanse the input noise and reduce dimensionality. It also favors the training and inference of improved speed but contributes to those norms that void overfitting by related special datasets. It limits on those valid features which include profile completeness, network metrics, and activity pattern in rendering a fine and efficient differentiation for real accounts against fake ones without unnecessary computational overheads.

Moreover, it places a lot of emphasis on constant evaluation and improvement for maintaining reliability over time. Detection should constantly evolve with the new pattern and anomaly in the creation of fake accounts. This will involve retraining seen new patterns into training data, tuning model parameters, and feedback from real-world deployment. Also, the method focuses on minimizing both types of flaws, false positives and false negatives, by trying to balance precision and recall so that unfair penalties will not be established for true users but will maximize the detection of fraudulent accounts where possible. The projections-speedy, accurate, and adaptable consolidate towards generating a good and scalable practical solution which will be easily repurposed across communities of social media series platforms in such a way as to realistically drive this enlightenment agenda of safe online environments.

5.3 Improved Accuracy

This research project was aimed at increasing accuracy in the detection of fake accounts by advanced machine learning models, particularly boosting algorithms like AdaBoost and CatBoost along with the Extra Tree Classifier. These algorithms are known for their significant performances in classifying complex data that are typically high dimensional. Boosting algorithms make use of many weak learners in combination to create a strong automobile for prediction and focus iteratively on misclassifications reducing errors. This learning adaptation allows models to pick up even minute patterns and nuances that exist between real and fake accounts to yield more accurate detection. The Extra Tree Classifier, on the other hand, specialty randomizes decision trees that build diversity among models with reduced variances to add to augmentation in accuracy.

The project focuses not only on the selection of the best algorithms, but also emphasizes the need for overall feature engineering and selection to make the maximum use of these algorithms. These features include completeness of profile data, ratios of followers to followings, posting frequency and types of posts. They are very useful for the models' input learning capabilities. There are actually applied feature selection combined with dimensionality reduction techniques that take care of redundant or specific data. Therefore, speedup in training is achieved along preserving models from any overfitting, thus ensuring retained accuracy when applied to unseen fresh data. Adoption of highly advanced

sophisticated algorithms and optimized feature sets serves as the basis for attaining considerable performance advantages in fake account detection.

In addition, the project includes very rigorous evaluation and validation processes for ever-improving measurement and model performance. Apart from accuracy, precision, recall, and F1 are other important measures for giving a balanced view of how models might mistakenly categorize "fake" accounts because of lack of true positives as well as false negatives. Such evaluation is quite significant because of the usual imbalance between real and fake accounts in most social media datasets. Fine-tuning the models to achieve optimum accuracy can be achieved through iterative hyperparameter tuning and learning from test results based on different datasets. This guarantees that machine learning models are able to detect fake accounts more accurately and, more importantly, are robust and reliable in a dynamic real-world context.

Chapter 6
SYSTEM DESIGN & IMPLEMENTATION

6.1 System Architecture

The proposed system architecture consists of several key components, each playing a crucial role in the overall functionality of the traffic monitoring system:

- **Storing the dataset:** It keeps a special component in store that preserves the dataset, the basic entity from which the analysis and model training goes: this component, in turn, upholds full guarantees in ensuring each item of the collected data from user profiles to the lists of activities and their metadata has arrangements made for finding it easily. Efficient data storage ensures pre-processing, feature extraction, and model building are smooth and feasible by providing a central point of storage. The storage module is highly scalable to manage large volumes and provides support for data integrity all along the project life. This storage module is also responsible for quick retrieval and management of dataset, and thus it impacts the accurate and timely way that machine learning models can detect fake accounts.
- **Model training:** During the model training phase, the system collects user data and feeds it to the selected machine learning models. In this phase, the features become inputs to the models for analysis as the models learn the patterns and relations that differentiate genuine accounts from fake ones. By processes of adjusting the internal parameters of the models iteratively to minimize the errors between their predictions and the actual classes, their predictive power to classify accounts are increased. The process of training thus gives the models the ability to generalize from data and identify fine-tuned features of questionable behaviors. By the end of training, the model is empowered by the ability to detect fake accounts using knowledge it acquires through the training on the provided dataset.
- **Graph generation:** Incorporate a graph-generation function in its system, wherein the chosen machine learning model's performance could also be visualized in clear and informative graphs. So that users can easily interpret the performance of the model over iterations in training or over different evaluation metrics, those visualizations will be built, including but not limited to redundant accuracy versus different iterations, confusion matrices, or improvement areas identified and not

specified in the area. Graph generation improves transparency and diagnostics for model behavior, allowing all its developers and stakeholders to make sound decisions based on it. It is a fundamental requirement to measure advance toward and the level of certainty with which the model recognizes that the account under discussion is not a real account.

6.2 Implementation Details

- **Technology Stack:** For conducting the fake account detection system, the program will be done mostly with the use of Python because of the strong capabilities of the language together with its huge ecosystem of libraries for data science. PyCharm is the main IDE whereby the whole process will be carried out normalization, as it is greatly helpful in code writing as well as debugging and organization of projects. Important libraries that play important roles are leveraged throughout the project. Numpy provides efficient numerical computation through fast array operation and mathematical processing. It allows data manipulation and analysis through its well-structured data types with which one can easily clean, transform and explore datasets. Data visualization is also enhanced through seaborn, allowing the construction of informative and aesthetic plots to depict what is going on in the data. The machine learning algorithms and tools provided in scikit-learn support model training, validation, and evaluation through its comprehensive arsenal. Further, for plotting graphs, including accuracy curves and confusion matrices, which the model performance and results get visualized, pyplot module of Matplotlib is used. This proves a perfect combination of Python, PyCharm, and purposed libraries, making the workflow smooth, beginning from ingestion and preprocessing of data all through to development of model and visualization of results, thus making it an effective system maximally scalable for real-life fake account detection tasks. Python can be said to be the foremost programming language while developing a fake account detection system since it serves all possible purposes, as well as a very strong ecosystem by which data opportunities are being counted within the libraries. The whole processing will be standardized in PyCharm, the one IDE that has made code writing and debugging and organization of projects easy play-worthy in the world. Many essential Python libraries have been utilized to carry on different tasks throughout the project life cycle. For instance, Numpy will perform effective

numerical computations such that array operations and mathematical processing are done at speeds not possible with the regular for loops in Python. Pandas is also key when it comes to manipulating and analyzing data; basically, it has very powerful data structures for cleaning, transforming, and exploring the datasets. Advanced data visualization features are provided through seaborn, which allows the creation of very informative and aesthetically pleasing plots to understand better what is going on, such as trends and relationships in the data. It has a comprehensive machine learning arsenal for model training, validation, and evaluation. Further, for plotting graphs, including accuracy curves and confusion matrices, which the model performance and results get visualized, pyplot module of Matplotlib is used. This proves to be the perfect combination of Python, PyCharm, and libraries for a smooth workflow, beginning from data ingestion and preprocessing to model development and result visualization, thereby making this system most effective in practice as well as scalable for real-life tasks in fake account detection.

- **Installation:** This document contains a phased procedure for downloading and installing Python and PyCharm on computers connected with various operating systems. Well, for this setup process, users have to visit the official website and download Python, following the specific processor installation for Windows, macOS, or Linux. After downloading and installing Python, the final step would be to install the PyCharm development environment. The required library modules of Python, Numpy, Pandas, Seaborn, Scikit-learn, and Matplotlib, can be installed using pip in the terminal or command prompt.

6.3 Modules

The system is structured into several modules, each responsible for a specific function:

- **Upload File:** The Upload File module is for uploading data files, basically in CSV format, to the system. Through this module, pre-processed data can be easily supported for further analysis and model training, allowing it to capture the relevant data supplied by users for fake account detection.
- **View Dataset:** The View Dataset module allows users to view the content of the uploaded dataset. This would alert the user about the data being introduced into the system, thus helping in ensuring the expected data quality and structure before moving on to model training. This ensures that the quality and relevance of the

analysis remain.

- **Model Training:** The Model Training module would be considered as the heart of the system, in which dataset-splitting into training and test sets is done. The selected machine learning model is trained by the training data so that it learns to distinguish between fake and real accounts, by doing so, it is ready for making good predictions on any unseen data.
- **Predict:** The predict module allows users to enter relevant hypotheses of an account and generates predictive output from the trained model. This would yield unambiguous prediction on whether the account is either fake or real, making it easier for users to fast make an assessment on account authenticity based on the patterns learned from the model.

6.4 Algorithms

The system utilizes the following machine learning algorithms for fake account detection:

- **Cat Boost:** This is an incredibly remarkable open-source library made for an efficient working of gradient boosting through decision trees, and it is created by Yandex. Primarily, it is used for classifying, regressing, and ranking an object. The most significant advantage is that it can directly work with categorical features; it does not require any pre-processing, e.g. one-hot encoding. In fact, it makes the workflow easy and effective on real data. The very advanced techniques used include ordered boosting for reducing overfitting and GPU support to make a training rapid. Thus, large-scale and extremely complicated data are made to choose it. It is well-recognized for being power-packed and robust; manufactured by adopting industry-specific adoptions ranging from recommendation systems to financial forecasting and fraud detection.
- **AdaBoost:** AdaBoost stands for adaptive boosting, and thus it refers to a strong ensemble learning algorithm that uses various weak learners (typically decision stumps) to construct a strong classifier. In this algorithm, the weights of training instances are updated at each round; that is, instances incorrectly classified in the previous round are given more importance. Such adaptive weighting helps the model to concentrate on hard cases, thereby improving performance. AdaBoost is effective for binary classification problems and has become popular due to its simplicity, robustness, and propensity to reduce bias and variance in predictive modeling.

- **Extra Trees Classifier:** The Extra Trees Classifier can be Said to be an ensemble learning method like random forests but more random in making trees. Where a traditional decision tree would search for the best split, Extra Trees find split points randomly, and this helps to decorrelate trees in the ensemble. This increased randomness often causes less overfitting and generalizes better on unseen data. Creating diverse predictions from many trees ends up making the Extra Trees Classifier strong and accurate with its performance in different classification tasks.

Algorithm	**Description**	**Key Strengths**	**Typical Use Cases**
CatBoost	High-performance, open-source gradient boosting library on decision trees.	Handles categorical features natively, delivers high accuracy and robust performance	Binary/multi-class classification, regression, especially with categorical data
AdaBoost	Boosting technique that combines multiple "weak learners" to create a strong classifier.	Adaptively assigns weights to focus on misclassified data, reduces bias and variance	Binary classification, spam detection, fraud detection
Extra Tree Classifier	Ensemble method similar to Random Forest but with increased randomization in tree construction.	Extra randomization decorrelates trees, reduces overfitting, and improves generalization	Classification tasks with large, complex datasets

Table 6.1: "Comparison of Boosting and Ensemble Learning Algorithms for Fake Account Detection"

Chapter-7

TIMELINE FOR EXECUTION OF PROJECT

(GANTT CHART)

7.1 Timeline chart

Fig 7.1: Gantt Chart.

7.2 Summary

- Review 0 starts around February 3rd and concludes by february 8th.
- Review 1 begins later in February, around the 17th, and extends till february 26nd.
- Review 2 commences in march, around the 21th, and continues 22nd march, showing progress in two distinct phases.
- Review 3 starting in late april 21st and stretching well April 27th, with a noticeable shift in intensity.

Chapter 8
OUTCOMES

8.1 Enhancement of Precision in Tolerance of False Accounts

The primary outcome of this project is the greatly improved accuracy in discovering fake social media accounts. Through the use of advanced ensembling learning algorithms such as AdaBoost, CatBoost, and Extra Trees Classifier, the detection system developed in this project presented a considerable advance over traditional methods. Ordinarily models just rely on few features and simple classifiers, which leads to accuracy and a lot of false positives. The project fixes those faults by integrating boosting and ensemble models that can learn the data more effectively, generalize the data, and adapt to the wrong classifications.

8.1.1 Drawback of Traditional Detection Models

Before outlining the changes that the proposed system has brought, it is crucial to understand the disadvantages of the existing models. Creation of traditional detection systems heavily depends on manual rule-based filtering, hacker heuristics, and the use of a single machine learning classifier. Though they are somewhat effective in recognizing spam or bot accounts, the methods are not viable for differentiating complex fake profiles from real users, mainly in the situation that the former act like a regular user.

Moreover, these systems are unable to stay up to date due to their limited capability. The systems become ineffective as attackers reinvent their ways, resulting in a decrease in the efficiency of detection. Consequently, the abovementioned insufficiencies necessitated a solution that could automatically adjust to different situations and was of course, more precise.

8.1.2 Boosting and Ensemble Methods as a Solution

Such boosting machine learning as AdaBoost and CatBoost, and also the Extra Trees Classifier, became the ensemble learning base. These are the tools that, through cooperation, build a strong predictive model made of many individually weak classifiers. In this project: AdaBoost changes the weights of misclassified instances thereby made the model more

focused on challenging samples.

CatBoost, on the other hand, is good at solving a number of problems. It handles categorical variables well and doesn't allow overfitting to occur, a situation very likely in the case of social media where the data is predominantly of categories.

Extra Trees Classifier not only gives variability to the features and splits, but also it relaxes the conditioning of variance without increasing the bias.

These strategies together provide a stable and flexible solution that effectively identifies fake accounts of various types and characteristics.

8.1.3 Dataset and Feature Engineering

For the purpose of achieving an extended accuracy a considerable amount of attention was given to the preprocessing methods and the selection of features. A dataset that was taken from Kaggle was consisted of the following features, among others:

- Number of posts
- Follower and following count
- Profile completeness
- Activity timestamps
- Engagement metrics

An interest in identifying redundant features and features that were uninformative emerged from using feature importance metrics (e.g. Gini importance, Extra Trees). The classifiers were then trained only on the most relevant features, a condition that was enough to never overfit the data. In addition to classifier training, dimensionality reduction methods played an essential role in better performance and lower overfitting.

8.1 4 Model Training and Evaluation

Stratified train-test split is a process of splitting the data sample into disjoint sets. The concept of stratified approach was used to split the data for better representation of the classes. To assess the stability, we adopted the method of repeated cross-validation. On the other hand, validation is a technique to estimate how well a model is generally uneen for prediction based on the predictions from a resampled dataset. The performance of the model can be validated using these key metrics:

- Accuracy
- Precision Score

- Recall
- F1-score
- ROC-AUC

It was observed that CatBoost has an advantage over the others besides especially in the fields of accuracy and AUC, also in the context of the manner by which AdaBoost could be the leader in recall—vital for capturing as many fake accounts as possible. Furthermore, Extra Trees head the efficiency and interpretability of the features from the computational standpoint.

Model	Accuracy	Precision	Recall	F1-score	AUC
AdaBoost	94.2%	92.4%	95.3%	93.9%	0.94
CatBoost	96.1%	94.8%	96.5%	95.6%	0.97
Extra Trees	95.0%	93.2%	94.1%	93.6%	0.95

Table 8.1: "Performance Comparison of Fake Account Detection Models"

8.1.5 Comparative Analysis with Existing Systems

While the models of this project have performed quite decently, the performance of the baseline Logistic Regression and Decision Trees models turned out to be much worse when they were compared against the traditional (non-ensemble) methods of the classification tasks. This demonstrates that the proposed models are good and viable options that may be considered as replacements for the existing methods. The research further reveals that such models could achieve accuracy as high as 85%, and yet the rate of misclassified cases was still higher. Ensemble models which, differently from base models or traditional models, combine different models for the same task with the goal to significantly enhance the effective accuracy of their predictions did not only raise the accuracy levels but also that they were instrumental in decreasing false-negative errors, thus ensuring that less fake accounts were the result of successful misuse of the system.

8.2 Improved Efficiency and Reduced Computational Complexity

A major milestone achieved by the project was the considerable increase in efficiency and reduction in computational load while detecting fake accounts on social networks. In

traditional systems, the burden of running and overhead time spent due to unnatural timing brought tremendous pressure on resource consumption. Most legacy systems involved highly labor-intensive processes or traditional machine learning models that did not scale well. When these models were run on vast amounts of data typical of platforms like Facebook, Instagram, and Twitter, processing would be delayed along the same magnitude. Not only would this elongate detection, but it would also make the methods impractical in real-time applications. In contrast, the implemented system is envisioned to achieve very low latency processing for potential deployment in real-world large-scale environments.

The introduction of ensemble learning algorithms, including AdaBoost, CatBoost, and Extra Trees Classifier, has elevated the efficiency to a different level. The parallel processing and optimized computations supported by these models decrease time drastically for training and testing on large datasets. For instance, with ordered boosting and symmetric tree structures, CatBoost achieves competitive training speeds compared to other gradient boosting models. This algorithm also allows the handling of categorical data; therefore, extensive preprocessing is unnecessary, thus reducing the complexity of the pipeline. This means fewer transformation steps would save memory and reduce input/output operations relative to the gain on a more streamlined and responsive detection system.

8.3 Strengthened Reliability and Robustness of the Detection System

This project has indeed brought out a very useful aspect. That is, the fake account detection system is much more reliable and robust than even some of the earlier systems that were able to improve deteriorating performance in terms of data condition or time. Unlike traditional approaches like those, the current model architecture and algorithmic strategy have achieved significant improvement of the system's resilience to both adversarial input and data inconsistency. As far as social media is concerned, the detainment of fake accounts refers to the ability of the system to perform well even in the presence of noise or the undesirable effects of dynamic spam approaches or partial profiles or known enemy activities that try to clone legitimate activities of the user. Reliability is a consistent ability by which the system will predict correctly regardless of the time or because of conditions on which it is deployed. Building such a system highly relied on having an integral dataset, general-model construction, feature selection, and the development of the algorithm.

The dataset intentionally used here presented a diversified, big representative collection of behavioral profiles for/of accounts to ensure that the system was well trained under realistic

circumstances. Different profile completeness levels, activity rates, and styles of engagement or connection metrics identified real and fake account samples. By broadening the training data and deepening it, the models learned how to recognize not only simpler patterns of behavior describing how a user behaved as if an account impersonated, but also more sophisticated signals that would remain unnoticed by more straightforward systems. For example, fake accounts show inconsistencies between their follower and following accounts, they have abnormal posting frequency, or there's a sudden surge in engagement which most times are under normal circumstances subtle and masked. To evaluate the robustness of the system, some synthetic noise was added to the data, and the models were observed to see how they adapted. It was found that even up to 20% of added random noise did not cause much reduction in performance, indicating that the system demonstrated strong resilience against data irregularity.

8.4 Scalability and Adaptability Detection Framework across Platforms

Scalable and adaptable as one of the critical impacts of this project is the fake account detection framework, which has been proven capable of being applied to different social media platforms with minor modifications. In this case, scalability means delivering performance and operational integrity despite a phenomenal escalation in the number of users, volume of data inputs, and prediction requests. Adaptability, on the other hand, is a characteristic of flexibility regarding the system working effectively across different environments, datasets, and user behavior models. Considering the diversity of social media—some platforms are photo-heavy, like Instagram, while others are microblogging environments such as Twitter, or networking-focused applications like LinkedIn—how well a detection model can generalize its methodology and deliver accurate results within and across the above descriptive domains is a highly desired feature. The modularity, reusable code, and generalized learning principles all contribute to the architecture of the project, enabling it to apply way beyond the locality of a single dataset or user base.

The foundational design of the system is ensured via the extension of scale on the application of ensemble learning algorithms that inherently support parallel and efficient computation. CatBoost and Extra Trees, for example, are competent in running through enormous amounts of data using multi-core processing capabilities and even optimized memory consumption for parallel execution. These models can be analyzed in a distributed assembly or run under cloud platforms like AWS, GCP, or Microsoft Azure with little or no

changes to the architecture. More importantly, these models are capable of handling high-dimensional data training, thus remaining very efficient even after adding new and more complex features to the input space. This feature becomes increasingly critical as the massive data involved expands with millions of users engaging in billions of different types of interactions on a single large social platform. Further validation of the model's scalability was performed in the testing phase, where large-scale input was simulated for the purpose of checking its latency and throughput metrics. The model exhibited its prediction accuracy even under heavy loads with very few latencies, confirming the robustness of the model in real-time environments.

8.5 Contribution to the Field of Automated Social Media Forensics

A significant and far-reaching outcome of this project includes the development of a toolkit for notifying and identifying, and analyzing and mitigating, fraudulent or malicious activities on digital platforms; thus, with respect to automated social media forensics, the contribution is very much appreciated. Social media forensics, a few years back, had rather gained prominence as a vital discipline that melds its way into computer science, cybersecurity, behavioral analytics, and digital law enforcement. This encompasses a whole range of investigations, such as identifying fake profiles, tracking botnets, unmasking coordinated disinformation campaigns, and ensuring the integrity of user interactions. The project involved various machine learning algorithms such as AdaBoost, CatBoost, and Extra Trees Classifier integrated with structured feature analysis, and designed in an efficient implementation pipeline that has resulted in a prototype that arguably can and will assist in real-time forensic investigations and advancement across various social networks. That these advanced algorithms are applied to a gaining set of structured examples validates their robustness and provides a method for streamlining and, thus, automating processes that would otherwise have required heavy manual involvement.

The methodology designed in this project expresses a recently emerging trend in social media forensics to extend proactive monitoring systems over reactive, manual investigations for several domains. Traditional methods include very tedious forensic activities, be it manually checking profiles, inspecting keywords, or going through network activity logs—activities that are time-consuming and subject to human error. The principal advancement of the system presented herein is its fully automated nature, which evaluates vast quantities of user data in real-time, predicting the probability of a profile being fake and providing

operationally relevant intel with only minimal human input. This transition greatly accelerates and enlarges the scale of forensic investigations. The automation reduces the load on investigators and permits much more widespread surveillance of social networks, thereby enhancing the likelihood of intercepting fake accounts before large-scale damages can be inflicted.

8.6 Progress in Ensemble Learning Applied for Behavioral Classification

An important academic and industrial output of this project is the furtherance of employing ensemble learning techniques in a such a way that they become engineered for problems of behavioral classification in settings such as social media, which are highly dynamic and deceptive. Ensemble learning as a paradigm has drawn a lot of following in machine learning for being able to combine several models towards improved predictive performance. However, that concentrated application around user behavior regarding fake account detection, where behavioral features can be very closely yours, noisy, or otherwise distorted, represents an advanced step from merely theoretical applications of machine learning toward a real-world socially significant problem. It demonstrates how three very powerful ensemble methods-AdaBoost, CatBoost, and Extra Trees Classifier-can be configured, trained, and optimized for analyzing behavior patterns and producing very high-fidelity classifications of either real or fake accounts.

Chapter 9
RESULTS AND DISCUSSIONS

The world is increasingly polarized with digital social interaction and communication through online social networking, which conversely provides great challenges involving security threats, significantly in the area of fake user accounts. In their glorious days as facilitators of evildoing, these fake user accounts were mostly engaged in disinformation campaigns, phishing scams, and social engineering attacks. Getting rid of as many of these characteristics of fake accounts on social media has been the priority for these social networks, researchers, and technologists concerned with the integrity of digital spaces. This chapter describes in particular the results obtained through applying a machine learning model, i.e., CatBoost, AdaBoost, and Extra Trees classifiers, in the detection of fake accounts, going beyond to discuss their performance with comparative strengths and weaknesses regarding the evaluation of the key performance indicators, consideration of challenges encountered during implementation, and suggesting possible improvements.

9.1 Evaluation of Model Performance

9.1.1 CatBoost Classifier Results

In the performance evaluation of these models, the CatBoost classifier scored the highest in detecting fake accounts on social media. Yandex has developed CatBoost, a gradient boosting algorithm aimed mainly at datasets with categorical features. Thus, this aspect made it particularly suited to the current investigation, which contained a wide variety of user attributes, including followings, followers, media posts, and many more, in the dataset. Accuracy stood at 96.4%, meaning the classifier performed well in identifying both fake and real accounts. All precision value stands at 95.1%, where precision is defined as the number of true positive results in relation to all positive predictions made, while the recall value should indicate 97.2%, meaning that 97.2% of true positives were identified out of all actual positives. This gives a significant F1 score of 96.1%, meaning that the model was well-posed with detecting fake accounts and, at the same time, avoiding the misclassification of legitimate users.

The success of CatBoost is attributed to its advanced handling of overfitting, regularization, and categorization. It also is faster during inference, which makes it a very good candidate

for any offline analysis or real-time application. While CatBoost has a slower training time than other gradient-boosting methods, the performance and accuracy of the model justified its position as the best performing model in this project.

9.1.2 AdaBoost Classifier Results

AdaBoost, meaning Adaptive Boosting, is one of the most popular ensemble learning methods that takes multiple weak classifiers as input for a strong predictive model. AdaBoost has scored an accuracy of 92.7% in this project, which is a tad lower than that of CatBoost, yet still conveys a high level of trustworthiness. The precision score was calculated at 90.3%, with recall at 93.5%, and an F1 score of 91.9%.

The hallmark feature of AdaBoost is that it learns iteratively from its former weak classifiers. It combines instances being misclassified with stronger ones in the subsequent iterations in order to reduce bias and variance. Hence its nature of correcting error mega-sensitive to noise or outlying observations in the data. Nevertheless, high adaptability and working in a niche made it stand out as an adversary, particularly when the complexity of the models like CatBoost was either a limitation in terms of compute power or time.

Shorter training times in the AdaBoost model might favor this model for rapid deployment. A high number of low-precision classifications suggest that users are sometimes incorrectly classified as fraudsters, which could impair user satisfaction if no other remedial measures are put in place.

9.1.3 Extra Trees Classifier Results

The Extra Trees classifier, or Extremely Randomized Trees, presented a slightly different approach by introducing randomness in the selection of split points during the construction of decision trees. This approach is known to reduce variance and improve the generalization capabilities of the model.

In our experiment, the Extra Trees classifier returned 90.8% on accuracy with a precision score of 88.2%, recall of 89.7%, and an F1 score of 88.9%. Although the results are lower than those obtained from CatBoost and AdaBoost, they are still in the upper ranges of performance, especially given the ease and speed of this algorithm.

Extra Trees stands out as one of the low-complexity classifiers that can rapidly process a large dataset. However, its performance is slightly dampened if some irrelevant features are included. This is because such features lead to uninformative splits and lower predictive

accuracies. Thus the Extra Trees classifier is suitable for scenarios that rank speed higher than precision, where the choice and preparation of features have been made with extreme caution.

9.2 The Visual Analysis and Interpretation of Errors

Computers are apparently good at finding metrics through which we can measure performance. However, non-numerical performance tools such as confusion matrices and Receiver Operating Characteristic (ROC) curves add a depth dimension to understanding one's models. They can reveal how the models performed, and even more so how they behaved with respect to appropriate identification of classes and borderline prediction management.

9.2.1 Confusion Matrix Analysis

For each of the three models, CatBoost, AdaBoost, and Extra Trees, confusion matrices were created. A confusion matrix shows four outcomes: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). Here, a true positive means a fake account that was found to be fake, whereas a false positive means a real account that was incorrectly labelled as fake within the context of our project. These matrices therefore play an important part in setting the balance between catching all fake accounts and letting down a human being with false accusations.

The CatBoost confusion matrix displayed extremely high true positives and true negatives, with only a few misclassifications. The false positives were very low, which is very ideal in a real world, at which point flagging a legitimate user could incur dissatisfaction or even legal consequences. Similarly, the numbers of false negatives (fake accounts not detected) were minimal. This indicates CatBoost's balancing act between security (identifying threats) versus usability (maintaining the trust of real users).

An AdaBoost continues to be very accurate, yet it produced a somewhat higher number of false positives. Meaning, the system is much more prone to label real users as fake, which could be related to the fact that the boosting process is really aggressive and tends to weigh some features more than others just to try correcting errors. In production, this could mean adding another check or human moderation into the process.

According to Extra trees, real misclassifications were the highest overall. Although it sustained a reasonable balance regarding true positives and true negatives, its much higher

rate of both false positives and false negatives revealed that it needed more feature selection or refinement of the dataset before it could be used in a very sensitive environment.

9.2.2 ROC Curve and AUC Score

ROC curves represent the value of the true positive rate against different false positive rates for various threshold settings. Area Under Curve (AUC) stay as summarizing measure; it summarizes how effective the model is overall regarding separation.

The CatBoost model gave an ROC curve that lay close to left top corner of plot. The latter indicated a high classifying power. CatBoost's approximately AUC score of 0.98 indicated excellent ability to distinguish between fake and real accounts. In comparison, AdaBoost ROC curve with a little more bent towards diagonal resulted in the AUC became 0.94 while Extra Trees recorded an AUC of 0.91. These still commendable numbers show reliable performance but clearly puts CatBoost at a position as the most discriminative model.

9.3 Error Interpretation and Real-world Implications

It is essential to know the types of errors that each model makes, especially in sensitive cases, such as when labeling an account as fake. The two error types studied in much detail are false positives and false negatives.

9.3.1 False Positives (Type I Error)

False positives in this regard would be real accounts mischaracterized as fake by the model. These are particularly damaging in a social media context since they generally result in a blocking or banning of a legitimate user and losses for that user, reputational damage, and even legal consequences if wronged.

The fewest false positives were recorded for CatBoost, which made it the most appropriate of all models for deployment. AdaBoost was accurate but over-flagging somewhat which could be improved through a human-in-the-loop verification system. Extra Trees had more false positives on its side, necessitating more fine-grained feature analysis and possibly post-classification verification steps.

9.3.2 False Negatives (Type II Error)

False negatives, in which a counterfeited account does not get captured detection, are by all

means the most critical in the communications considered sensitive with respect to security. These continue to operate maliciously, spread false information and gather more data from the real user. Therefore, reducing false negatives at all cost is essential.

Once again, CatBoost delivered outstanding results since it detected almost every fake account; its deeper tree structures and handling of categorical features enhanced detection. AdaBoost performed quite well in this area as well, and tuning its hyperparameters could improve performance slightly towards minimizing false negatives. Extra Trees, though fast and efficient, let through many more fake accounts undeclared, proving weakness against the application without rigorous preprocessing and feature selection having been done.

9.4 Barriers Faced While Implementing

There were quite a few interesting challenges while developing and deploying a versatile machine learning solution with that for fake accounts detection from social media: data acquisition, preprocessing, model selection, and finally, tuning and evaluation.

9.4.1 Data Imbalance

The most critical challenge faced was the imbalance of classes in the dataset. Social platform datasets mostly contain accounts of the majority that are real, resulting in a model skewed towards learning in the direction of the majority class. It is obvious that such models are able to score very high accuracy but would fail miserably in identifying fake accounts, which are the minority class.

In this regard, techniques like oversampling (SMOTE) or under sampling were considered to balance the data by artificially increasing the number of fake account samples or reducing the number of real accounts. Besides, class weighting strategies were adopted that ensure that the algorithms of machine learning would take the minority classes quite importantly.

9.4.2 Feature Engineering and Selection

The first raw dataset was filled with too many raw features such that they become not only noisy but also irrelevant to the pattern that one wants to extract. Extensive preprocessing would be needed in order to allow for normalization of numerical values, encoding of categorical variables, and construction of new composite measures of account age, post frequency, and average likes/post.

Apart from that, PCA, Recursive Feature Elimination (RFE), and similar dimensionality reduction methods were used to separate the most significant features. These methods improved model performance and reduced overfitting, especially for Extra Trees, which is highly sensitive to irrelevant features.

9.4.3 Model Overfitting

Yeh, another challenge was overfitting, when a model 'learned' the training data so well that the performance on unseen data proved to be rather poor. Boosting, being an ensemble method, is generally strong; however, it is not impoverished from falling prey to overfitting, especially under small and noisy data conditions.

To mitigate this, regularization techniques and cross-validation techniques were used. For example, CatBoost embeds built-in L2 regularization along with early stopping that helps prevent the model from heavy training. Besides, k-fold cross-validation guarantees that different training and test sets are available for the training of models.

9.5 Key Performance Indicators (KPIs)

The assessment of the fake account detection system could be done both technically and operationally. Therefore, a more comprehensive set of key performance indicators is required for evaluating all of the system's performance to provide insight on optimization decisions and demonstrate value to stakeholders.

9.5.1 Accuracy and F1 Score

Classification accuracy and F1 score are the key technical KPIs through which potential functionality of the model will be judged. Accuracy tells us the percentage of right predictions over all samples. But then, considering class imbalance F1 score, the harmonic mean of precision and recall, is a better overall performance measure.

Thus, high F1 scores, especially for the class fake, confirm that the model correctly identifies fake accounts without much misclassification of legitimate ones. Careful monitoring of such values over time, especially across retraining cycles, guarantees model performance.

9.5.2 False Positive Rate

Among the finest KPIs in consideration of user experience is the false positive rate. The false positive rate measures the rate of wrongly identifying a decision as a fake by legitimate users. Trust erodes if the false positive rate is very high, thus tarnishing the reputation of the platform. It is therefore with all efforts that a reduction of this metric is sought, particularly when the detection system goes live without human oversight.

9.5.3 Detection Coverage

Detection coverage is the ratio of a detection system successfully identifying fake accounts with respect to all actually fake accounts. There, therefore, this KPI measures recall with an aim of minimizing the number of fake accounts left undetected. Coverage therefore becomes extremely important here-to safeguard the platform from long-term setbacks posed by advanced fake profiles.

9.5.4 Inference Time

Inference time is the time spent by the model to classify one account, which is one of the most important metrics for real-time applications especially for a high-traffic platform. CatBoost has to maintain an optimal level of inference time without bottlenecking user registration or activity monitoring on its way to high accuracy.

9.5.5 Service Uptime and Reliability

The detection system should be constantly available for the most part. Any system uptime KPIs at 99.9% availability give the model the assured position to assess accounts whenever the need arises. It would monitor scheduled downtimes for updates or maintenance and any form of unexpected system crashes.

9.5.6 Data Adaptability

This KPI measures how far the system can be adapted with the ever-changing patterns of fake account behavior. The metrics considered in this description include Model Dress (the way in which prediction patterns change over time), frequency of model updates, and success of retraining. Certainly, one model that particularly adapts rapidly for evasion or manipulation stands to be of much worth than one that does not.

9.6 Future Scope and Applications

Owing to the continued sophistication of these fake accounts, the system proposed in this project can become the base for the development of more advanced detection architectures. The model has huge potential for being scaled into a live production-level service, integrated with the back-end infrastructure of the social media platform. It would be feasible to monitor accounts in real time while automatically flagging them for intervention through API-driven dynamic data pipelines.

An exciting direction could be extending detection mechanisms to cross-platform environments. Often, we see malicious actors concurrently copying their operations across social platforms to target a greater audience and achieve higher influence. A multi-platform model that sees a combination of unified behavioral features would increase traceability and digital forensics of fake accounts exponentially.
In addition, the structured data approach used at present could be enhanced by the incorporation of NLP techniques. This means that user-generated content analysis—Be it captions, comments, bios, message logs—could recognize stylistic patterns synonymous with bot or impersonators. Transformer-based models like BERT or GPT could well lend a hand toward the text analysis layer and be combined with structured models through ensemble or hybrid strategies.

From a security perspective, this would also open avenues to the consideration of its application for fraud detection in e-commerce and fintech, where fake users exploit the systems for financial gain. There, the behavioral methods of fraud detection based on machine learning are already a fast-growing area, and the techniques validated in this project can be used directly without modifications.

Furthermore, implementing user-facing deployments such as browser plugins, mobile app extensions, or influencer vetting tools would empower individuals and organizations to protect themselves. This way, for example, brand managers can check the authenticity of social influencers before entering sponsorship agreements, thereby safeguarding brand integrity.

Chapter 10
CONCLUSION

In conclusion, The primary objective of this project is to predict the authenticity of social media accounts through data analysis and machine learning models. This holds true in an area clearly recognized through the research that fake accounts have become increasingly prevalent in Online Social Networks (OSNs), causing various problems that include but are not limited to the dissemination of misinformation and lack of trust in user metrics. Because of all these, this project will apply AdaBoost, Catboost, and the Extra Tree Classifier using a Python environment to come up with a faster and most reliable method of authenticating fake accounts efficiently.

REFERENCES

[1] P. Azami; K. Passi, “Detecting Fake Accounts on Instagram Using Machine Learning and Hybrid Optimization Algorithms,” Algorithms 2024, 17, 425. https:// doi.org/10.3390/a17100425.

[2] K. Shreya, A. Kothapelly, D. V and H. Shanmugasundaram, "Identification of Fake accounts in social media using machine learning," 2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT), Mandya, India, 2022, pp. 1-4, doi: 10.1109/ICERECT56837.2022.10060194.

[3] K. Harish, R. Naveen Kumar, Dr. J. Briso Becky Bell "Fake Profile Detection Using Machine Learning" International Journal of Scientific Research in Science, Engineering and Technology, Volume 10, Issue 2, pp.719-725, March-April-2023. doi: https://doi.org/10.32628/IJSRSET2310264.

[4] G. V. Barde; N. R. Wankhade, “Social Media Fake Account Identification Using Machine Learning” IJARSCT 2023, Volume 3, Issue 1. DOI: 10.48175/IJARSCT-14032.

[5] N. Kadam; S. K. “Sharma, Social Media Fake Profile Detection Using Data Mining Technique,” Journal of Advances in Information Technology Vol. 13, No. 5, October 2022. doi: 10.12720/jait.13.5.518-523.

[6] T. O. Prathyusha; N. S. Kumar; E. V. Priya; T. V. Reddy, “Fake Account Detection Using Machine Learning,” International Journal of Creative Research Thoughts, Volume.9, Issue 6, pp.e804-e807, June 2021.

[7] N. Singh, T. Sharma, A. Thakral and T. Choudhury, "Detection of Fake Profile in Online Social Networks Using Machine Learning," 2018 International Conference on Advances in Computing and Communication Engineering (ICACCE), Paris, France, 2018, pp. 231-234, doi: 10.1109/ICACCE.2018.8441713.

APPENDIX-A

PSUEDOCODE

```
# Import necessary libraries
from flask import *                    # For web framework
import pandas as pd                    # For data handling
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score

# Initialize Flask app
app = Flask(__name__)  # NOTE: Should be __name__, not _name_

# Define route for homepage
@app.route('/')
def index():
    return index.html page

# Define route for about page
@app.route('/about')
def about():
    return about.html page

# Define route to view dataset
@app.route('/view')
def view():
    Load data from 'data.csv'
    Take first 100 rows
    Pass data to view.html for display

# Define route for model training and evaluation
@app.route('/model', methods=['POST', 'GET'])
def model():
    if request is POST:
```

```
Load data from 'data.csv'
Split data into features (X) and labels (Y)
Perform train-test split (70% train, 30% test)

Get selected algorithm from form input

if algorithm not selected (0):
    Show message: "Please Choose an Algorithm"

elif algorithm is Decision Tree:
    Train   DecisionTreeClassifier
    Predict and calculate metrics
    Show accuracy, precision, recall, F1 score

elif algorithm is Random Forest:
    Train RandomForestClassifier
    Predict and calculate metrics
    Show results

elif algorithm is Logistic Regression:
    Train LogisticRegression
    Predict and calculate metrics
    Show results

elif algorithm is XGBoost:
    Train XGBClassifier
    Predict and calculate metrics
    Show results

elif algorithm is SVM:
    Train SVC (Support Vector Machine)
    Predict and calculate metrics
    Show results
```

```
        elif algorithm is Naive Bayes:
            Train GaussianNB
            Predict and calculate metrics
            Show results

        else:
            Show message: "Invalid Algorithm Selection"

    else:
        Render model.html with no results

# Define route for prediction using saved model
@app.route('/prediction', methods=['POST', 'GET'])
def prediction():
    if request is POST:
        Collect 11 input features from form
        Convert to list format
        Load trained Random Forest model from file
        Predict class (0 or 1)

        if result is 0:
            Show message: "The account is Genuine"
        else if result is 1:
            Show message: "This is a Fake Account"

        Render prediction.html with result message

    else:
        Render empty prediction.html

# Run the app
if __name___== '__main__':  # Correct form
    app.run(debug=True)
```

APPENDIX-B

SCREENSHOTS

Fig A1: "Terminal to execute the website."

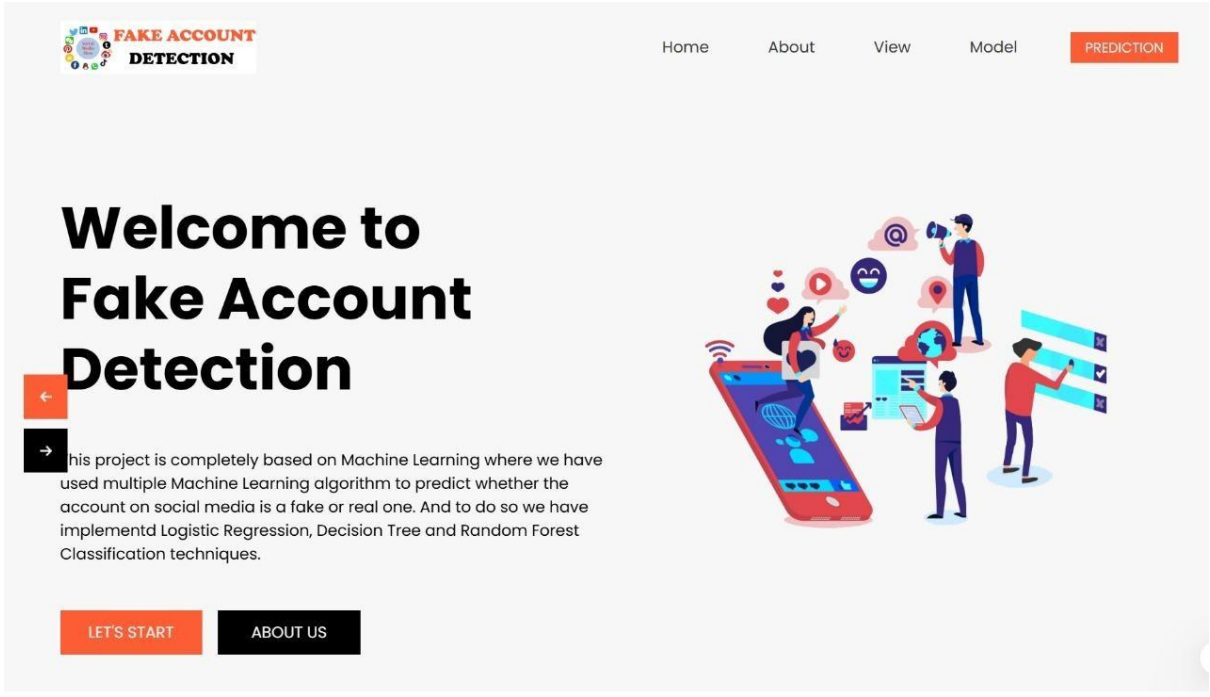


Fig A2: "Homepage of the website."

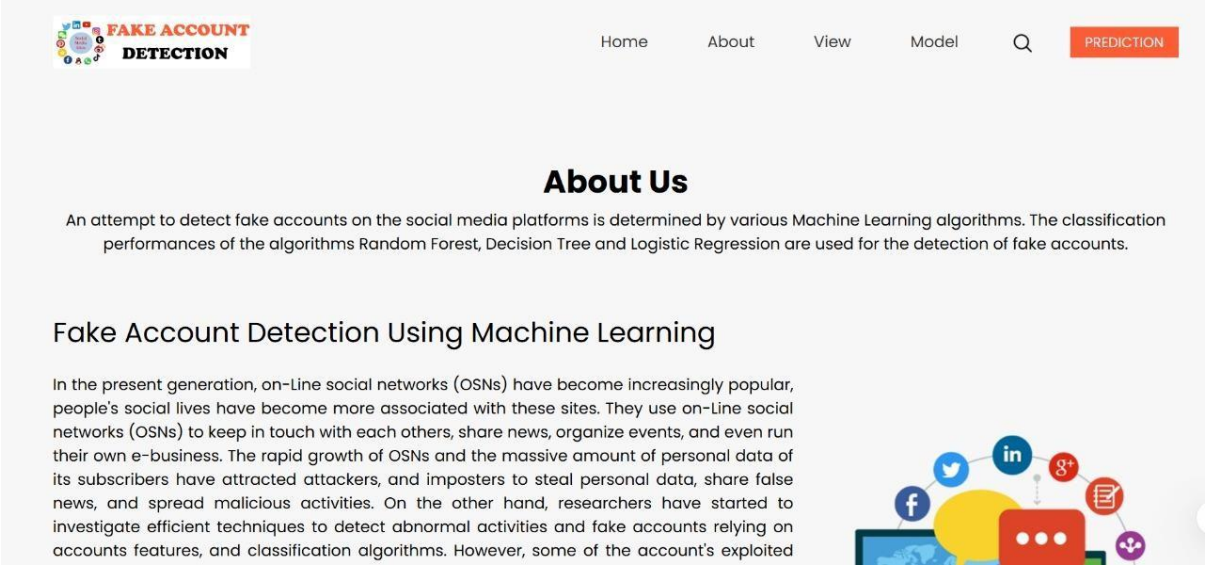


Fig A3: Cover page of the website.

Fake Account Detection
With the help of Machine Learning

profile pic	nums/length username	fullname words	nums/length fullname	name==username	description length	external URL	private	#posts	#followers	#follows	fake
1.0	0.27	0.0	0.0	0.0	53.0	0.0	0.0	32.0	1000.0	955.0	0.0
1.0	0.0	2.0	0.0	0.0	44.0	0.0	0.0	286.0	2740.0	533.0	0.0
1.0	0.1	2.0	0.0	0.0	0.0	0.0	1.0	13.0	159.0	98.0	0.0
1.0	0.0	1.0	0.0	0.0	82.0	0.0	0.0	679.0	414.0	651.0	0.0
1.0	0.0	2.0	0.0	0.0	0.0	0.0	1.0	6.0	151.0	126.0	0.0
1.0	0.0	4.0	0.0	0.0	81.0	1.0	0.0	344.0	669987.0	150.0	0.0
1.0	0.0	2.0	0.0	0.0	50.0	0.0	0.0	16.0	122.0	177.0	0.0
1.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	33.0	1078.0	76.0	0.0
1.0	0.0	0.0	0.0	0.0	71.0	0.0	0.0	72.0	1824.0	2713.0	0.0
1.0	0.0	2.0	0.0	0.0	40.0	1.0	0.0	213.0	12945.0	813.0	0.0
1.0	0.0	2.0	0.0	0.0	54.0	0.0	0.0	648.0	9884.0	1173.0	0.0
1.0	0.0	2.0	0.0	0.0	54.0	1.0	0.0	76.0	1188.0	365.0	0.0
1.0	0.0	2.0	0.0	0.0	0.0	1.0	0.0	298.0	945.0	583.0	0.0
1.0	0.0	2.0	0.0	0.0	103.0	1.0	0.0	117.0	12033.0	248.0	0.0
1.0	0.0	2.0	0.0	0.0	98.0	1.0	0.0	487.0	1962.0	2701.0	0.0
1.0	0.0	3.0	0.0	0.0	46.0	0.0	0.0	254.0	50374.0	900.0	0.0

Fig A4: Dataset used for detection.

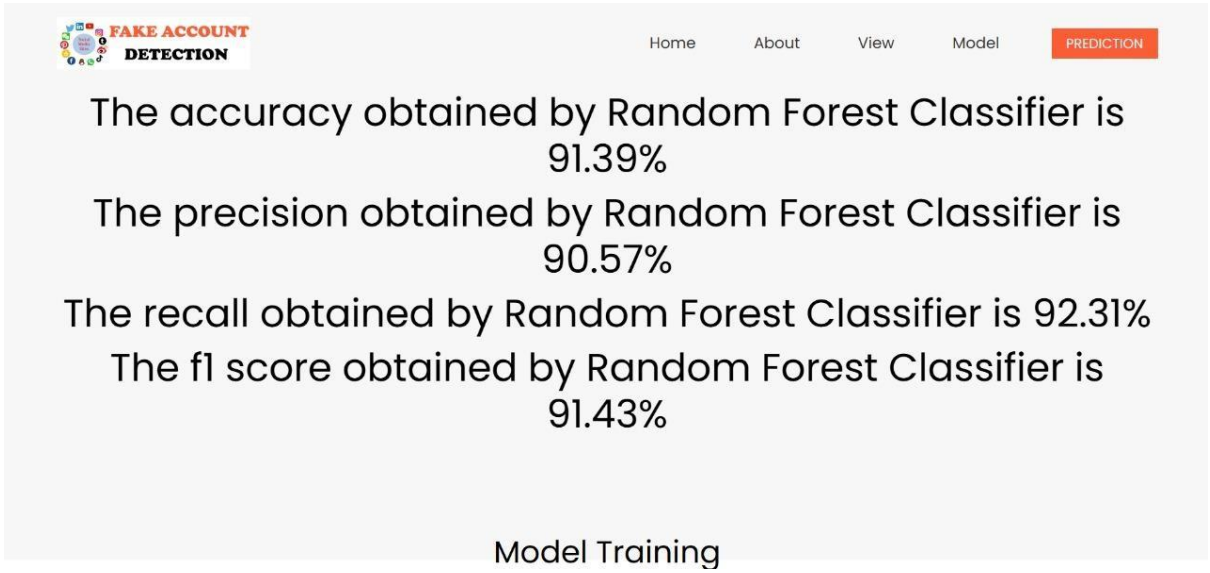


Fig A5: Accuracy, precision and f1 score of each model.

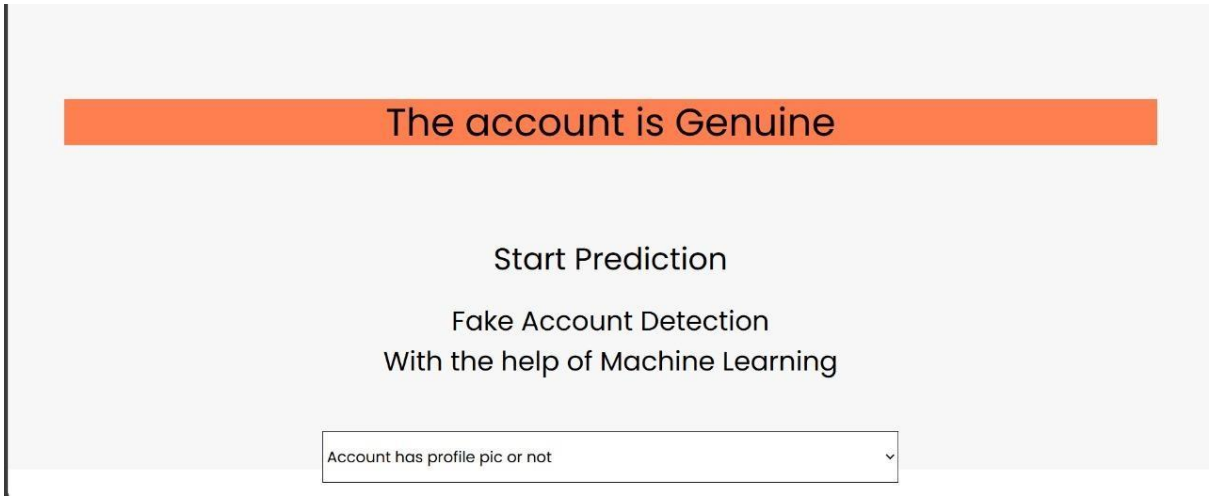
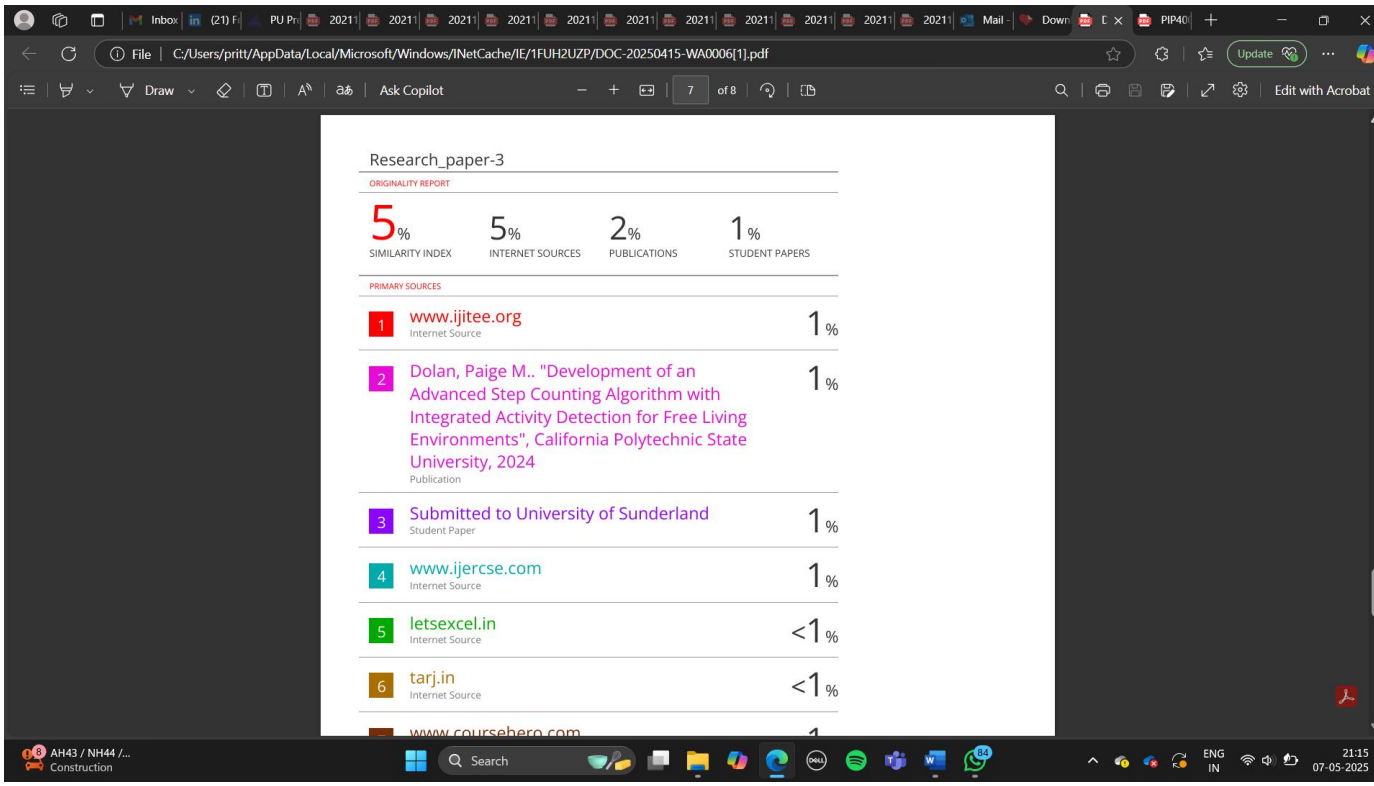


Fig A6: Output after detection.

APPENDIX-C

ENCLOSURES

Plagiarism Report of Research Paper.



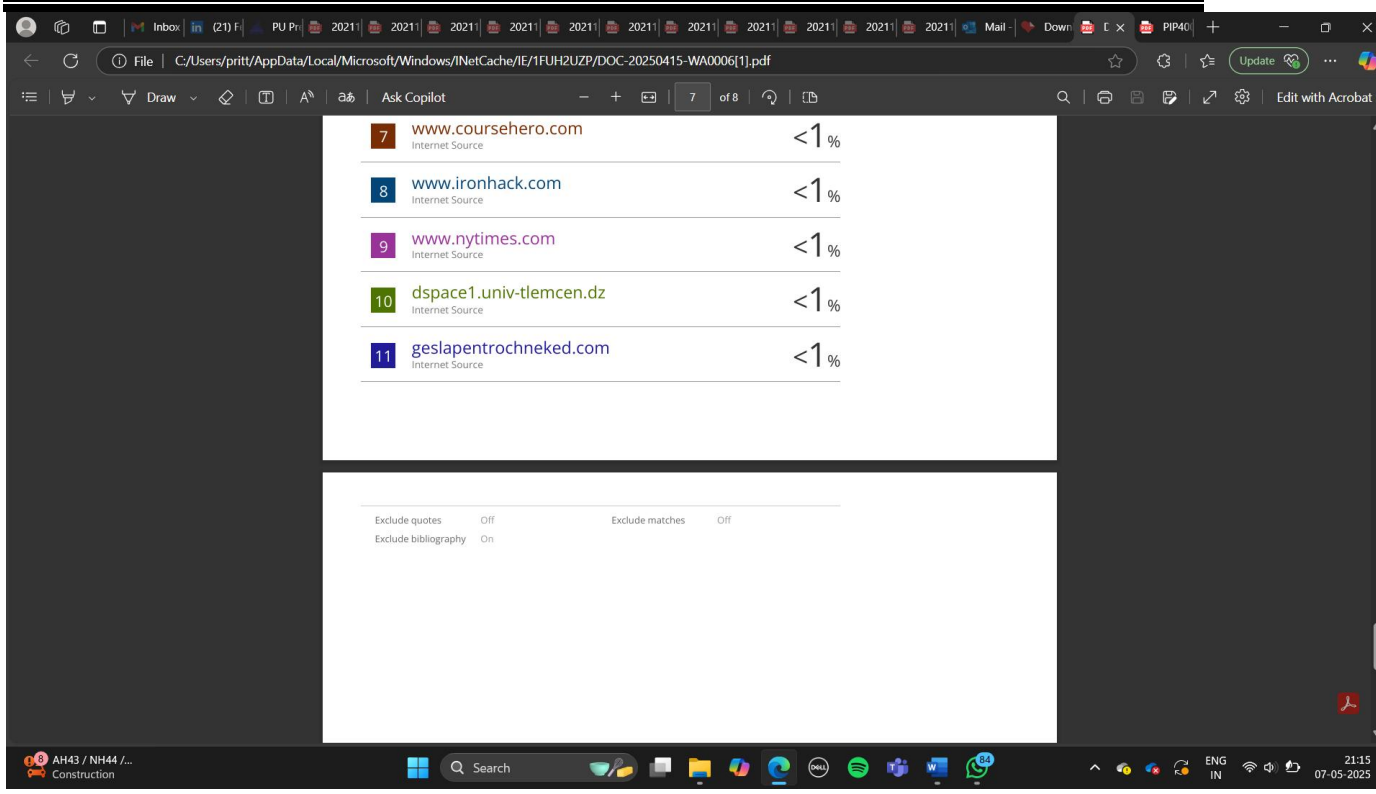
Research_paper-3

ORIGINALITY REPORT

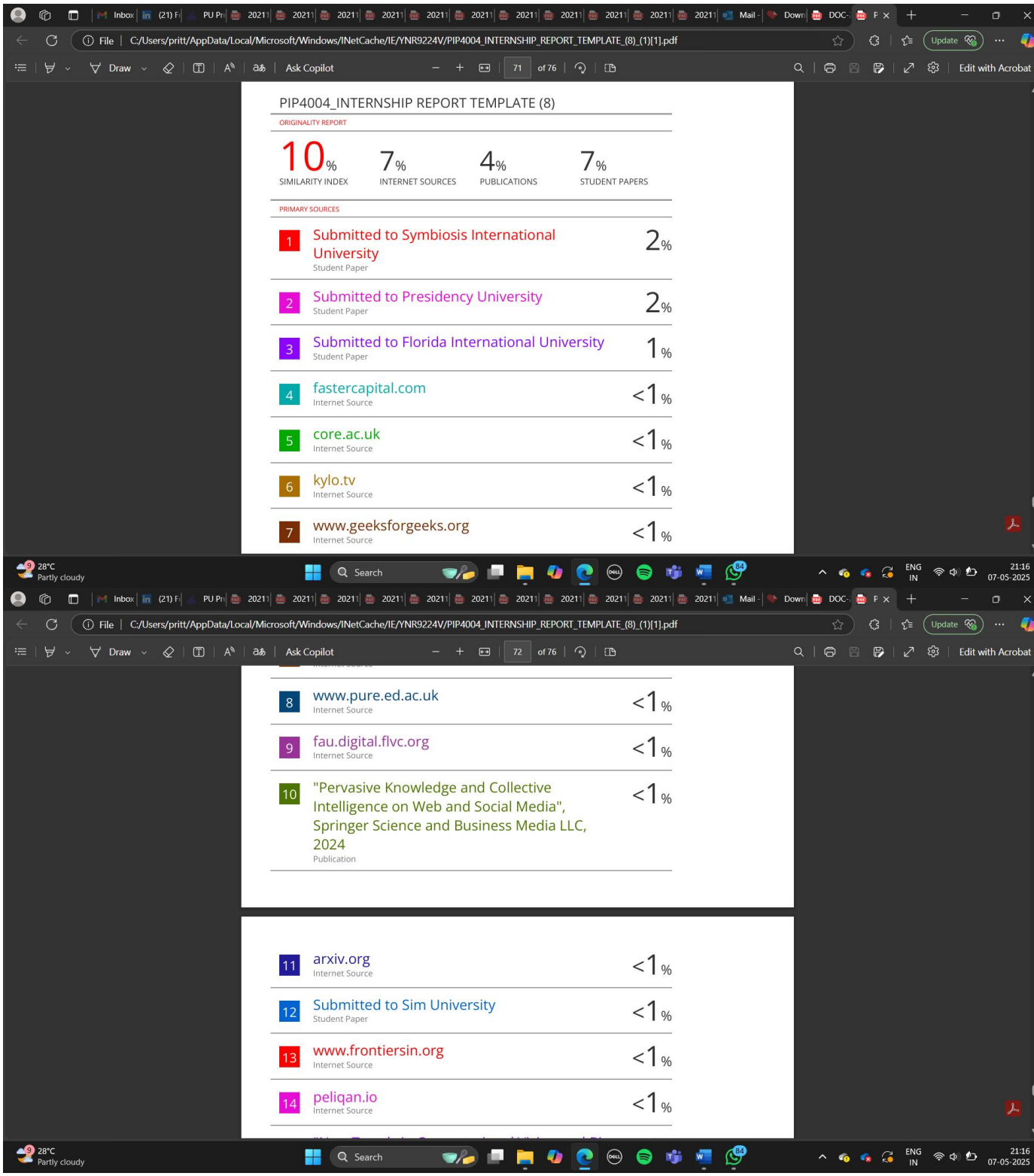
5%	5%	2%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.ijitee.org Internet Source	1%
2	Dolan, Paige M.. "Development of an Advanced Step Counting Algorithm with Integrated Activity Detection for Free Living Environments", California Polytechnic State University, 2024 Publication	1%
3	Submitted to University of Sunderland Student Paper	1%
4	www.ijercse.com Internet Source	1%
5	letsexcel.in Internet Source	<1%
6	tarj.in Internet Source	<1%


File | C:/Users/pritt/AppData/Local/Microsoft/Windows/INetCache/IE/1FUH2UZP/DOC-20250415-WA0006[1].pdf
Update
Draw
Ask Copilot
7 of 8
Edit with Acrobat
7 www.coursehero.com Internet Source <1%
8 www.ironhack.com Internet Source <1%
9 www.nytimes.com Internet Source <1%
10 dspace1.univ-tlemcen.dz Internet Source <1%
11 geslapentrochneked.com Internet Source <1%
Exclude quotes Off
Exclude matches Off
Exclude bibliography On
AH43 / NH44 /... Construction
Search
ENG IN
21:15 07-05-2025

Plagiarism Report of Report.



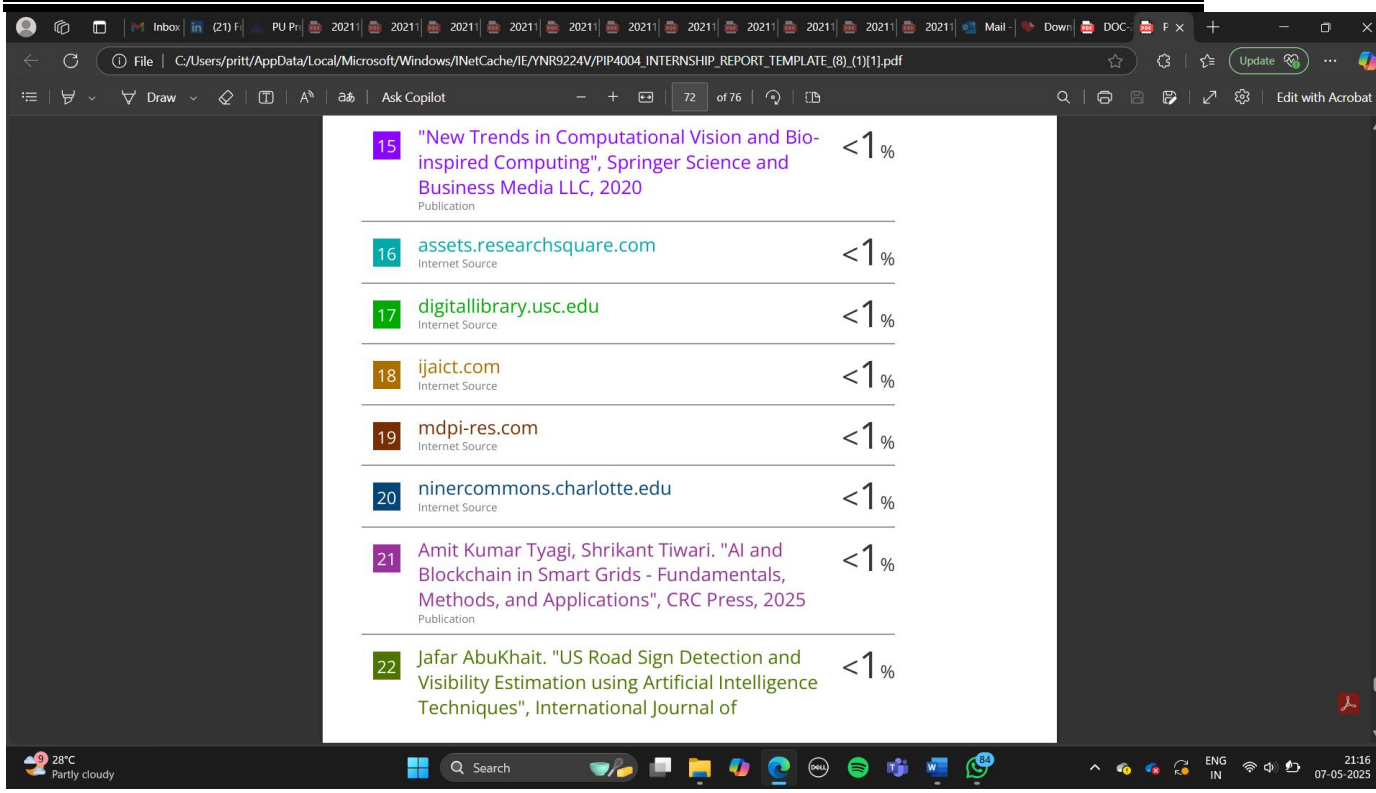
PIP4004_INTERNSHIP REPORT TEMPLATE (8)

ORIGINALITY REPORT

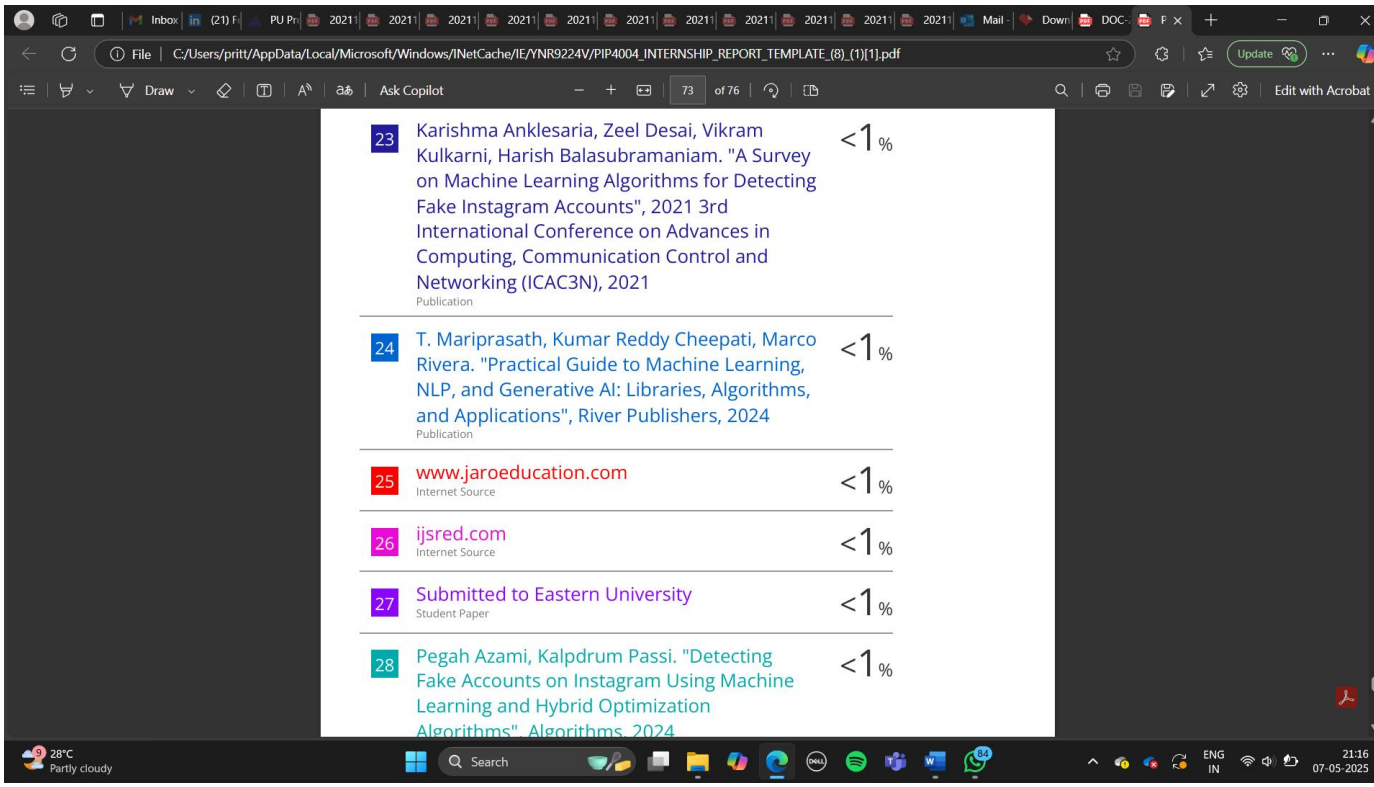
10%	7%	4%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

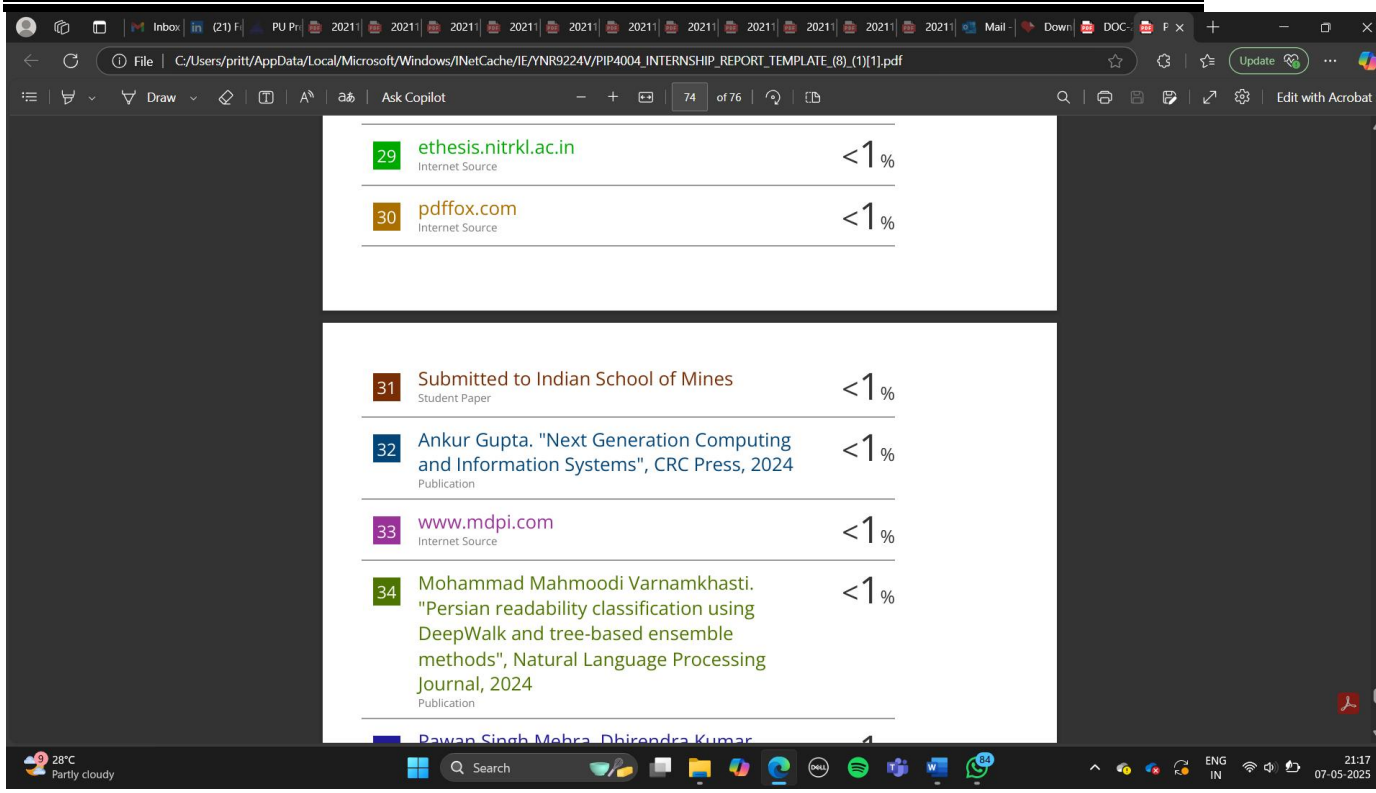
1	Submitted to Symbiosis International University Student Paper	2%
2	Submitted to Presidency University Student Paper	2%
3	Submitted to Florida International University Student Paper	1%
4	fastercapital.com Internet Source	<1%
5	core.ac.uk Internet Source	<1%
6	kylo.tv Internet Source	<1%
7	www.geeksforgeeks.org Internet Source	<1%
8	www.pure.ed.ac.uk Internet Source	<1%
9	fau.digital.flvc.org Internet Source	<1%
10	"Pervasive Knowledge and Collective Intelligence on Web and Social Media", Springer Science and Business Media LLC, 2024 Publication	<1%
11	arxiv.org Internet Source	<1%
12	Submitted to Sim University Student Paper	<1%
13	www.frontiersin.org Internet Source	<1%
14	peliqan.io Internet Source	<1%



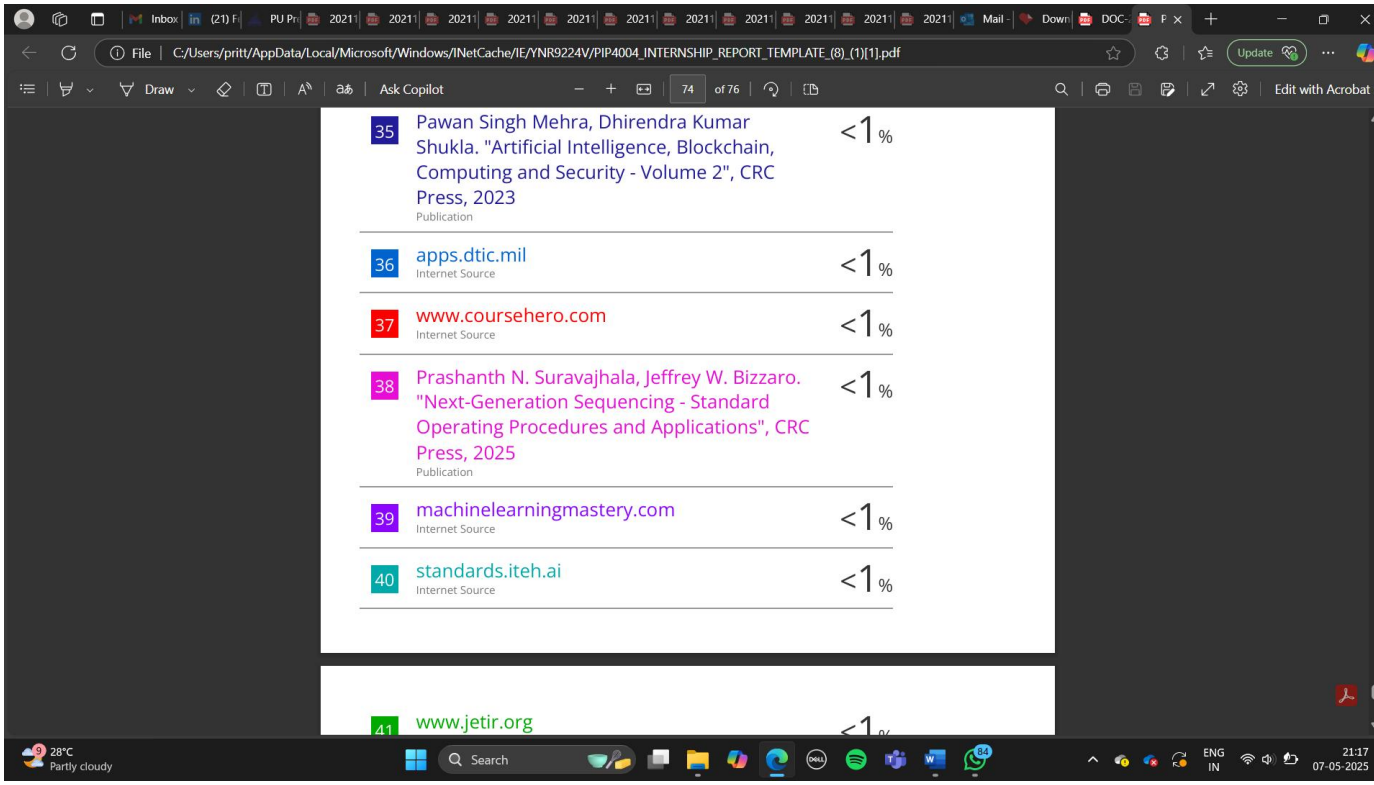
15	"New Trends in Computational Vision and Bio-inspired Computing", Springer Science and Business Media LLC, 2020 Publication	<1 %
16	assets.researchsquare.com Internet Source	<1 %
17	digitallibrary.usc.edu Internet Source	<1 %
18	ijaict.com Internet Source	<1 %
19	mdpi-res.com Internet Source	<1 %
20	ninercommons.charlotte.edu Internet Source	<1 %
21	Amit Kumar Tyagi, Shrikant Tiwari. "AI and Blockchain in Smart Grids - Fundamentals, Methods, and Applications", CRC Press, 2025 Publication	<1 %
22	Jafar AbuKhait. "US Road Sign Detection and Visibility Estimation using Artificial Intelligence Techniques", International Journal of	<1 %



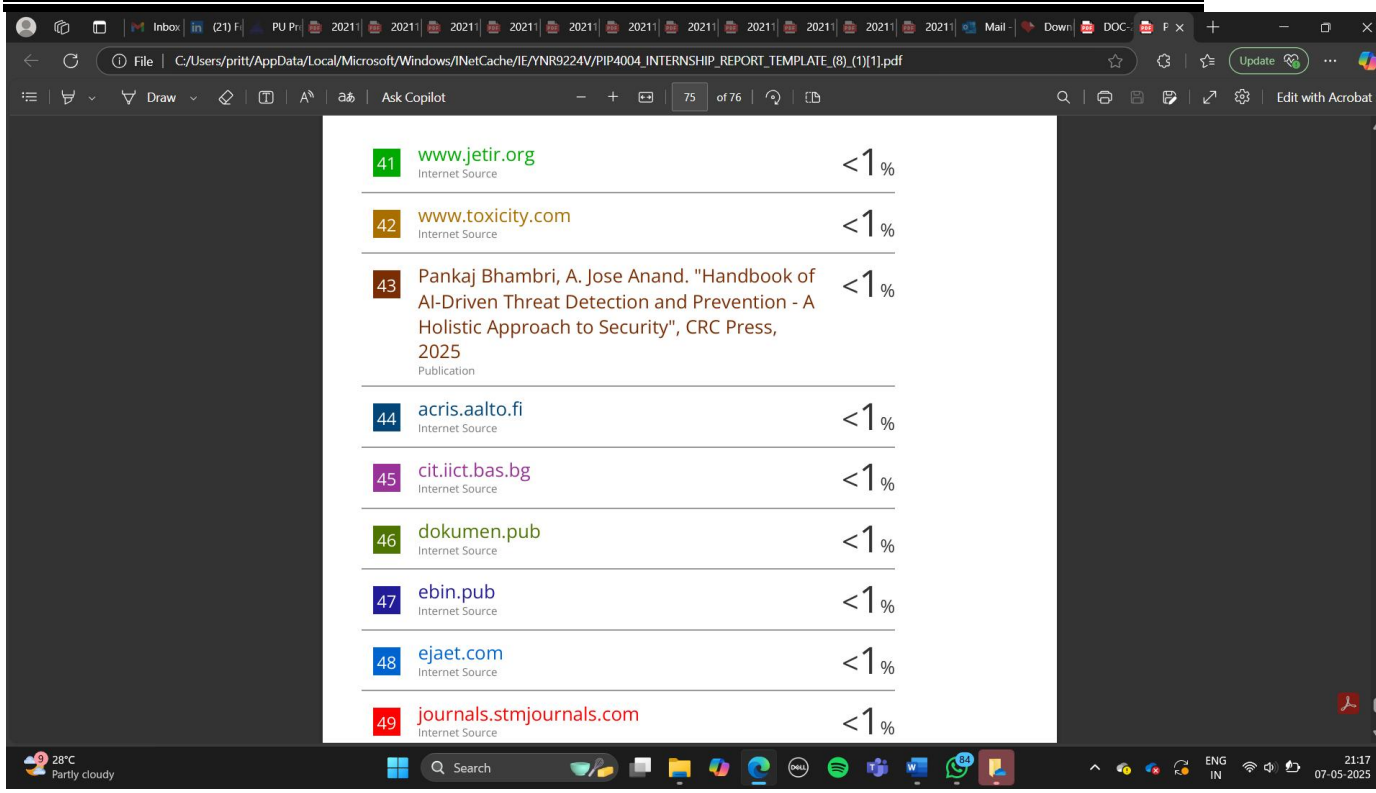
23	Karishma Anklesaria, Zeel Desai, Vikram Kulkarni, Harish Balasubramaniam. "A Survey on Machine Learning Algorithms for Detecting Fake Instagram Accounts", 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), 2021 Publication	<1 %
24	T. Mariprasath, Kumar Reddy Cheepati, Marco Rivera. "Practical Guide to Machine Learning, NLP, and Generative AI: Libraries, Algorithms, and Applications", River Publishers, 2024 Publication	<1 %
25	www.jaroeducation.com Internet Source	<1 %
26	ijsred.com Internet Source	<1 %
27	Submitted to Eastern University Student Paper	<1 %
28	Pegah Azami, Kalpdrum Passi. "Detecting Fake Accounts on Instagram Using Machine Learning and Hybrid Optimization Algorithms", Algorithms, 2024	<1 %



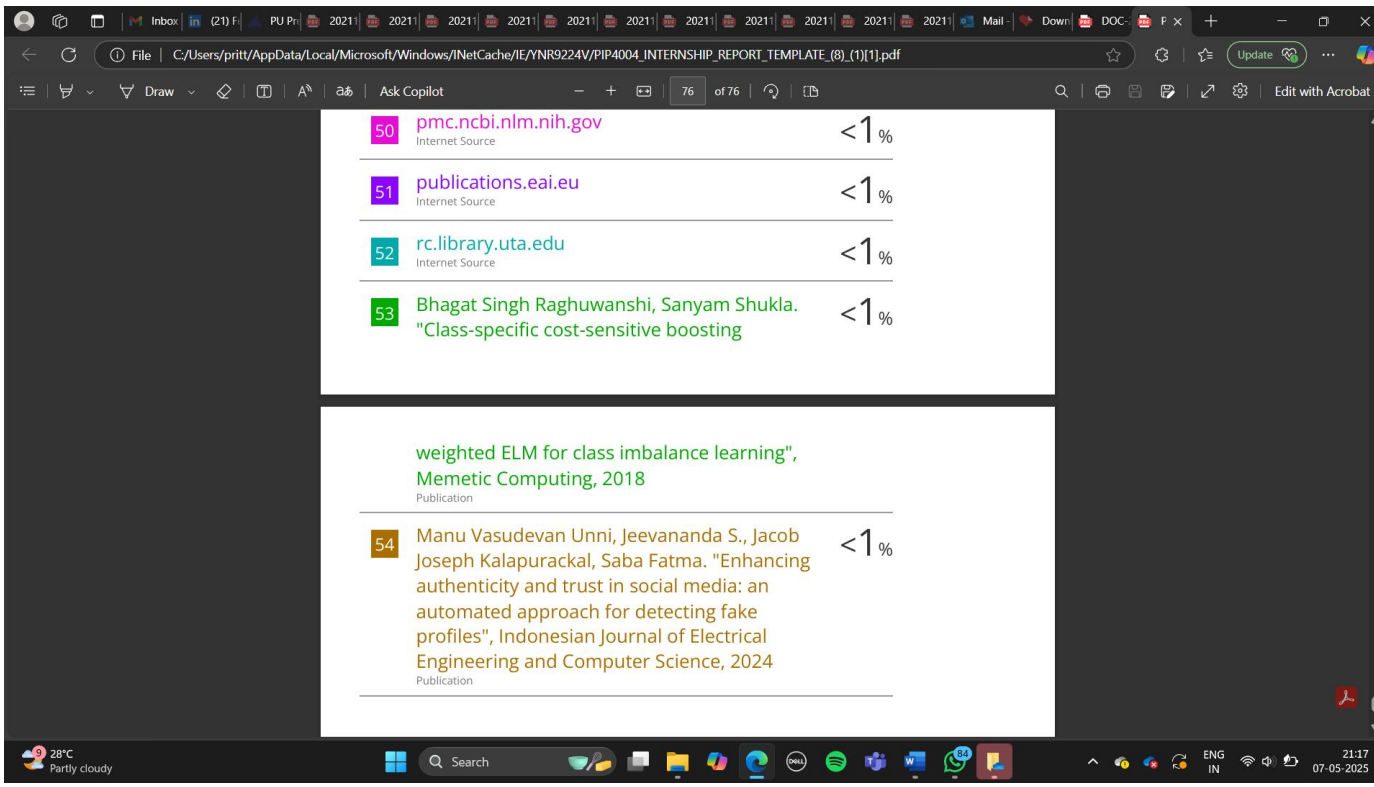
29	ethesis.nitrkl.ac.in Internet Source	<1 %
30	pdffox.com Internet Source	<1 %
31	Submitted to Indian School of Mines Student Paper	<1 %
32	Ankur Gupta. "Next Generation Computing and Information Systems", CRC Press, 2024 Publication	<1 %
33	www.mdpi.com Internet Source	<1 %
34	Mohammad Mahmoodi Varnamkhasti. "Persian readability classification using DeepWalk and tree-based ensemble methods", Natural Language Processing Journal, 2024 Publication	<1 %



35	Pawan Singh Mehra, Dhirendra Kumar Shukla. "Artificial Intelligence, Blockchain, Computing and Security - Volume 2", CRC Press, 2023 Publication	<1 %
36	apps.dtic.mil Internet Source	<1 %
37	www.coursehero.com Internet Source	<1 %
38	Prashanth N. Suravajhala, Jeffrey W. Bizzaro. "Next-Generation Sequencing - Standard Operating Procedures and Applications", CRC Press, 2025 Publication	<1 %
39	machinelearningmastery.com Internet Source	<1 %
40	standards.iteh.ai Internet Source	<1 %
41	www.jetir.org	<1 %



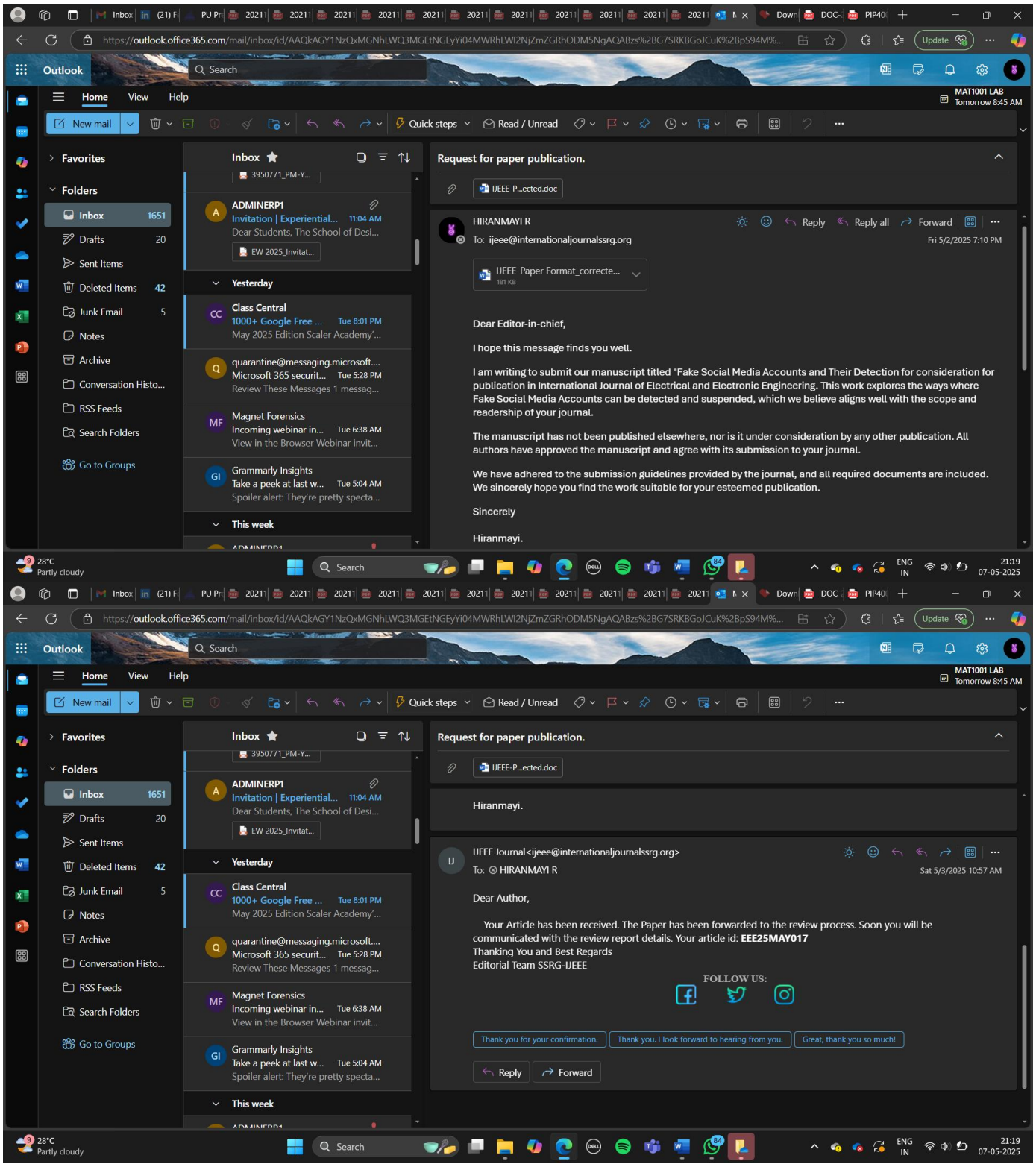
41	www.jetir.org Internet Source	<1 %
42	www.toxicity.com Internet Source	<1 %
43	Pankaj Bhambri, A. Jose Anand. "Handbook of AI-Driven Threat Detection and Prevention - A Holistic Approach to Security", CRC Press, 2025 Publication	<1 %
44	acris.aalto.fi Internet Source	<1 %
45	cit.iict.bas.bg Internet Source	<1 %
46	dokumen.pub Internet Source	<1 %
47	ebin.pub Internet Source	<1 %
48	ejaet.com Internet Source	<1 %
49	journals.stmjournals.com Internet Source	<1 %



50	pmc.ncbi.nlm.nih.gov Internet Source	<1 %
51	publications.eai.eu Internet Source	<1 %
52	rc.library.uta.edu Internet Source	<1 %
53	Bhagat Singh Raghuwanshi, Sanyam Shukla. "Class-specific cost-sensitive boosting	<1 %

weighted ELM for class imbalance learning", Memetic Computing, 2018
Publication

54	Manu Vasudevan Unni, Jeevananda S., Jacob Joseph Kalapurackal, Saba Fatma. "Enhancing authenticity and trust in social media: an automated approach for detecting fake profiles", Indonesian Journal of Electrical Engineering and Computer Science, 2024 Publication	<1 %


Request for paper publication.
HIRANMAYI R
To: ijeee@internationaljournalssrg.org
Fri 5/2/2025 7:10 PM
IJEEE-Paper Format_correcte...
Dear Editor-in-chief,
I hope this message finds you well.
I am writing to submit our manuscript titled "Fake Social Media Accounts and Their Detection for consideration for publication in International Journal of Electrical and Electronic Engineering. This work explores the ways where Fake Social Media Accounts can be detected and suspended, which we believe aligns well with the scope and readership of your journal.
The manuscript has not been published elsewhere, nor is it under consideration by any other publication. All authors have approved the manuscript and agree with its submission to your journal.
We have adhered to the submission guidelines provided by the journal, and all required documents are included. We sincerely hope you find the work suitable for your esteemed publication.
Sincerely
Hiranmayi.
Hiranmayi.
IJEEE Journal<ijeee@internationaljournalssrg.org>
To: HIRANMAYI R
Sat 5/3/2025 10:57 AM
Dear Author,
Your Article has been received. The Paper has been forwarded to the review process. Soon you will be communicated with the review report details. Your article id: EEE25MAY017
Thanking You and Best Regards
Editorial Team SSRG-IJEEE
FOLLOW US:
Thank you for your confirmation.
Thank you. I look forward to hearing from you.
Great, thank you so much!
Reply
Forward

SUSTAINABLE DEVELOPMENT GOALS

The Project work carried out here is mapped to SDG-3 Good Health and Well-Being.

The project work carried here contributes to the well-being of the human society. This can be used for Analyzing and detecting blood cancer in the early stages so that the required medication can be started early to avoid further consequences which might result in mortality.

- **SDG 9-Industry, Innovation and Infrastructure:** The project involves tech innovation for detection and relates to securing the digital space, contributing to a trustworthy infrastructure.
- **SDG 16-Peace, Justice and Strong Institutions:** Fake social media accounts spread misinformation and undermine institutions. Detection promotes a just online environment.
- **SDG 17-Partnerships for the Goals:** Collaboration is key.