Journal of Information Systems Engineering and Management
2025, 10(10s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A Hybrid Model for Detecting Fake Profiles in Online Social Networks: Enhancing User Trust

Rohini Bhosale[1],Vanita Mane[2]
[1]*Department of Computer Engineering,*
Ramrao AdikInstitute of Technology,
D Y Patil Deemed to be University, Nerul, India.
Department of Computer Science &Engineering (ICB),
Dwarkadas Jivanlal Sanghvi College of Engineering, Mumbai, India. rohini.bhosale@djsce.ac.in
[2]*Department of Computer Engineering,*
Ramrao Adik Institute of Technology,
D Y Patil Deemed to be University, Nerul, India.
vanita.mane@rait.ac.in

ARTICLE INFO

Received: 08 Nov 2024
Revised: 26 Dec 2024
Accepted: 10 Jan 2025

ABSTRACT

The multiplication of fake profiles in online social systems (OSNs) has developed as a basic challenge, debilitating client believe and security. This paper presents a novel cross breed show planned to distinguish fake profiles in OSNs by combining progressed machine learning methods to improve discovery precision and keep up client believe. Our approach coordinating both Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models, known for their quality in preparing successive information, into a crossover show that leverages the qualities of each method for more compelling location. Employing a dataset collected from Twitter, we conducted a comparative examination of different machine learning models, counting Naïve Bayes, Irregular Woodland, AdaBoost, and Support Vector Machines (SVM), to assess their execution in identifying fake profiles. In expansion, repetitive neural systems, such as LSTM and GRU, were tried separately some time recently being combined into the proposed crossover demonstrate. Each model's execution was assessed based on accuracy, review, F1 score, and exactness. Our findings illustrate that the cross breed LSTM-GRU show outflanks conventional machine learning calculations and person repetitive models, accomplishing prevalent discovery exactness and decreasing untrue positives. By saddling the complementary qualities of LSTM's capacity to capture long-term conditions and GRU's computational proficiency, the half breed demonstrate offers an progressed arrangement to fake profile location. This inquire about gives a comprehensive system for distinguishing fake profiles in OSNs, pointing to move forward the keenness of client intuitive and upgrade believe in online stages. The proposed show has critical suggestions for OSN security, especially in moderating the rising risk of false accounts.

Keywords: Fake Profile Detection, Online Social Networks (OSNs), Hybrid Model, Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Machine Learning Algorithms

I. Introduction

With the rise of innovation, there's presently a modern period of interfacing, talking, and working together through social systems online. These destinations are online places where individuals can meet unused individuals, share their encounters, and have talks with individuals all over the world. On the other hand, this move to computerized has moreover brought approximately numerous issues. The huge number of fake profiles is one of the foremost critical issues that must be unravelled in inventive and cautious ways. As individuals oversee the complicated web of social media, accounts that appear genuine are exceptionally critical for building believe and making genuine contacts. In any case, the troubling rise in fake accounts, which are frequently made with terrible eagerly, has harmed believe and made clients less secure. Fake accounts have impacts that go past online experiences. They can cause real-life issues like spreading wrong data, stalking, personality robbery, and changing how individuals feel

approximately things [1], [2]. This inquire about points to utilize cutting edge Machine Learning (ML) strategies to create finding fake profiles on social systems more precise. This will offer assistance unravel a major and developing issue. A parcel of consideration is paid to the complicated angles of profile coordinating, which is important for how clients interface with each other over diverse stages. It is evident what the objective is: to ensure clients from conceivable threats by building a solid framework that can discover and diminish the number of fake accounts. The creation of online social systems has changed how individuals see and interface with the world [3]. From the early days of Friendster and MySpace to the huge names of nowadays, like Facebook, Instagram, and Twitter, these systems are an vital portion of advanced life. Be that as it may, as these stages gotten to be more prevalent, so do the effects of bad individuals who need to require advantage of clients who do not know what's going on.

People make fake accounts for a lot of different reasons, from spreading false information for political reasons to committing cybercrimes like identity theft and financial scams. As these fake profiles get more complicated, they need more advanced defences [4]. This has led to study into more advanced machine learning methods that can successfully solve the problem. These days, finding fake accounts is very important in the digital world. Users depend on the reliability of profiles to help them choose who to connect with, what information to believe, and how to find their way around the huge amount of digital material available. If you interact with a fake page, bad things can happen, like falling for scams or phishing attempts or sharing false information without meaning to. The general security of online social networks is also at risk because of fake accounts [5]. These fake accounts hurt the platforms' reputation, which makes users less likely to trust them and use them. Having a reliable way to find and delete fake profiles is important for keeping digital communities healthy and thriving in a time when information spreads quickly.

Fake accounts have an impact on society as a entirety, not fair on the individuals who utilize them. Spreading wrong data, which is regularly done by fake accounts, can alter people's minds, alter the comes about of races, and part society separated. The reality that fake accounts can be utilized to trick or bug individuals online makes it indeed more clear that this issue ought to be settled in a careful way [6]. As individuals lose believe in online trades, it influences numerous regions, counting e-commerce and online tutoring. Individuals may be reluctant to purchase things online, share individual data, or connect online bunches, which can make it harder for individuals to associate in a great and valuable way on these destinations [7]. A test from Twitter is utilized in this think about to do a careful investigation of how to spot fake profiles. Twitter is extraordinary for examining changing patterns of client behavior since it is continuously changing and in genuine time. The huge number of users and speedy sharing of data on the stage make it conceivable to induce a part of data around how fake profiles carry on and how to spot them. The dataset features a assortment of user-generated fabric, like composing, pictures, and social joins [8]. The point of this ponder is to see into how difficult it is to spot fake profiles on social media locales, basically on Twitter, but the comes about may moreover be utilized on other destinations.

Natural Language Processing (NLP) is a key part of finding fake profiles. Most of what people do on social media is written down, like posts, comments, and profile descriptions [9]. Natural language processing (NLP) algorithms make it easier to read and understand written content, picking up on meanings, emotions, and small differences in language that help find fake accounts [10]. With the assistance of Characteristic Dialect Preparing (NLP), the proposed strategy can figure out the meaning of user-generated substance, discover discussion patterns, and spot oddities that may well be signs of trick. Not as it were does NLP offer assistance us get it dialect way better, it moreover makes a difference us get it how individuals act in advanced places. This app does more than its regular job; it's presently a key portion of finding fake profiles.

The most thoughts behind this ponder are Machine Learning (ML) and its subset, Profound Learning (DL). ML is based on the thought of learning from information, and it gives us a way to utilize information to discover patterns and exceptions that seem cruel fake profiles exist. Profound Learning's complex neural arrange structures make it less demanding for the show to discover complicated connections and profound representations in enormous datasets. This makes the disclosure handle indeed way better [11]. It is important to compare distinctive machine learning and profound learning models in arrange to discover perfect way the most perfect way to spot fake profiles. There are stars and cons to each strategy, such as Naïve Bayes, Irregular Timberland, AdaBoost, Back Vector Machines (SVM), Long Short-Term Memory (LSTM), Gated Repetitive Unit (GRU), and blended versions of these models. For illustration, a few may be way better at understanding information rapidly, whereas others may be way

better at being exact or managing with large amounts of information [12]. To urge the finest comes about at finding fake profiles, these components ought to be balanced when choosing the proper plan.

The major Contribution of paper is given as:

- This study utilizes Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models to boost detection performance, introducing an innovative hybrid approach that combines the strengths of both architectures. LSTM and GRU, as recurrent neural networks, are adept at capturing sequential dependencies in data, making them ideal for tasks involving temporal patterns.
- The proposed hybrid model merges LSTM's ability to retain long-term dependencies with GRU's efficiency in handling shorter sequences, addressing limitations of each individual architecture. This combination enhances the model's capability to detect complex patterns associated with fake profiles, resulting in improved accuracy.

II. Related Work

Individuals are paying a parcel of consideration to finding fake profiles on online social systems (OSNs). This can be since fake accounts are getting to be more common and harmed client believe. A study of diverse thoughts has been put forward, from straightforward machine learning strategies to complex profound learning models. These strategies are implied to meet the developing require for solid following frameworks that can keep social stage users' trades secure. Dealing with bad actors' continually changing techniques may be an assignment that drives the improvement of these sorts of models. A parcel of prior considers utilized rule-based strategies and straightforward statistical models to undertake to discover fake profiles. These strategies generally utilized set up characteristics, just like the number of companions, the number of posts, and the age of the account, to tell the distinction between genuine and fake profiles. These strategies worked some of the time, but they weren't exceptionally great since they couldn't keep up with unused dangers and the information from social systems was difficult to get it. For case, [13] focused the utilize of highlights that are carefully chosen by hand, indicating out that inflexible strategies are constrained in a advanced world that changes rapidly. Additionally, [14] talked almost how difficult it is to keep spotting precision tall when as it were utilizing rule-based frameworks, particularly since fake accounts are getting more intelligent all the time. Since of these issues, machine learning (ML) strategies have come up as a more adaptable way to discover fake profiles. ML models might naturally learn patterns that point to fake profiles without having to utilize set rules by utilizing enormous datasets. For this, a study of individuals has utilized calculations like Naïve Bayes, Irregular Woodland, and Bolster Vector Machines (SVM). For occurrence, [15] appeared that Arbitrary Woodland models were superior than rule-based strategies since they seem adjust to modern information more rapidly. But the fact that ML-based acknowledgment frameworks still had to depend on include engineering, which means that people had to choose out critical highlights by hand to prepare the demonstrate [16].

Individuals like to utilize Repetitive Neural Systems (RNNs) and their varieties, like Long Short-Term Memory (LSTM) and Gated Repetitive Unit (GRU), to see at successive information, just like the text-based trades that happens a lot study on social systems. Since they can learn high-level representations from crude information on their claim, these models have appeared guarantee in making acknowledgment more exact. For example, [17] showed that LSTM is good at finding trends over time, especially when the way users behave over time is a key part of finding fake accounts. In the same way, [18] looked into the benefits of GRU, focusing on how well it works with computers and how well it compares to LSTM in many situations. As a way to improve recognition even more, hybrid models that use more than one method have become more popular. When ML and DL methods are used together, they make recognition systems more reliable. For instance, [19] looked into how Random Forest and LSTM models could work together and found that this kind of combined method could be more accurate and flexible than using just one model. Comparable to this, [20] recommended a blended demonstrate that employments both SVM and profound learning to discover fake profiles more precisely by utilizing the finest parts of both ML and DL. A developing collection of consider has too looked at how to utilize Common Dialect Handling (NLP) to see at the content that individuals post on social systems. NLP lets you spot fake profiles by looking at designs of composed discussion, which are frequently exceptionally distinctive between genuine and fake accounts. The consider [21] talked almost how vital it is to utilize disposition examination and content classification to discover little changes in dialect that can be signs of unscrupulous behavior. At last, [22] appeared that utilizing both characteristic dialect preparing (NLP) and profound learning models like LSTM and GRU together can offer

assistance discover fake profiles utilizing text-based characteristics. This strategy has worked particularly well on destinations like Twitter, where brief, real-time discussions grant us a parcel of data around how individuals carry on.

A part of work has been made in finding fake destinations, but there are still issues. One of the most issues is that awful individuals are continuously changing the ways they do things. As frameworks for finding fake accounts get way better, so do the strategies utilized to create and run them. The research [23] said that unfriendly assaults are getting more astute. In these assaults, fake profiles are made to trap identifying models. This appears how imperative it is to keep inquiring about versatile checking frameworks that can keep up with these changing dangers. Moreover, [24] brought up the issue of scale. As social systems proceed to develop, so does the amount of information that must be taken care of. This implies that real-time checking frameworks must be both exact and quick. Analysts are working difficult to discover fake profiles on social systems like Facebook and Twitter. Unused improvements in machine learning, profound learning, and blended models are making observing frameworks more grounded and more exact. Indeed in spite of the fact that victory has been made, issues like changing unfriendly techniques and issues with scaling up mean that more development is required. Using NLP methods together with progressed ML and DL models may be a cheerful way to progress the distinguishing proof of fake accounts and, within the conclusion, re-establish client believe in online social stages.

Table 1: Summary of related Work in fake profile detection

Approach	Key Finding	Algorithm Used	Limitation	Advantages	Scope
Rule-based detection	Effective for early detection of fake profiles	Manual feature selection	Requires constant updates, static nature	Simple and easy to implement	Suitable for small datasets, outdated for evolving threats
Feature-based classification	Identified key user behavior metrics for detection	Naïve Bayes	Limited to predefined features	Fast and computationally efficient	Applicable to real-time social networks monitoring
Random Forest	Outperforms traditional rule-based systems	Random Forest	Requires heavy feature engineering	High accuracy with structured data	Useful in complex decision-making systems
Sentiment analysis in text data	Captures linguistic differences to detect fake accounts	NLP, Sentiment Analysis	Difficult to generalize across platforms	Strong performance with text-based data	Expands to text-dominant platforms like Twitter
Social graph analysis	Reveals suspicious connections and anomalies in user networks	Graph-based algorithms	Ineffective without rich social graph information	Identifies network behavior anomalies	Broad applicability across large social networks
Temporal behavior analysis	Effective in identifying inconsistent behavior over time	LSTM, Time-Series Analysis	High computational costs	Strong for long-term user behavior monitoring	Suitable for platforms where activity tracking is key
Hybrid LSTM-GRU	Combines strengths of both LSTM and GRU for improved detection	LSTM, GRU	Computationally intensive	High accuracy for sequential data	Ideal for complex behavior prediction

Deep learning-based classification	Automatic detection without heavy feature engineering	LSTM, CNN	Requires large datasets for training	High-level feature extraction capabilities	Applicable to text, image, and multi-modal data
Random Forest and LSTM Hybrid	Enhanced detection accuracy through hybrid approaches	Random Forest, LSTM	Complex implementation	Combines the strengths of multiple models	Suitable for real-time detection and scalable systems
SVM with deep learning	Improved precision in detecting fake accounts	SVM, Deep Learning Hybrid	Not effective for imbalanced datasets	High precision and recall rates	Best suited for medium to large datasets
Adversarial detection model	Detects fake accounts designed to evade standard algorithms	GANs, Adversarial Networks	Requires continuous retraining	Adapts to evolving threats	Critical for future OSNs with evolving threats
NLP-based hybrid	Strong at detecting fake profiles based on communication style	NLP, Hybrid ML Models	Limited to text-based platforms	Leverages both content and user interaction	Effective for platforms with text-heavy interactions
Anomaly detection with ML	Identifies outliers in user behavior for detection	Isolation Forest, K-Means	Limited to specific types of anomalies	Effective for detecting unusual activity	Scalable for large datasets
Deep graph neural networks	Extracts complex patterns in social networks to detect fakes	Graph Neural Networks (GNNs)	Requires comprehensive graph information	Captures relational dynamics	Suitable for deep analysis in network-based platforms

III. Methodology

A organized and careful approach is required to come up with a great way to spot fake accounts on social media. For this ponder, the "twibot-20" dataset, which has information from Twitter accounts, is carefully arranged ahead of time as portion of the strategy. The essential JSON file is carefully looked over to induce valuable information out of it, like account insights, tweet content, and social organize points of interest. Amid this handle, steps are taken to urge freed of copy records, fill in any lost values, and settle any issues, which secures the dataset's security. Once the information has been cleaned up, the center changes to include extraction and choice, where key variables are positioned by how likely they are to appear extortion. These components incorporate the date the account was made, how frequently tweets are sent, the number of clients compared to the number of individuals who are taking after, engagement measures such as offers and likes, and patterns in dialect utilize. The choice handle is exceptionally imperative since it finds the foremost vital characteristics that offer assistance tell the distinction between genuine and fake accounts.

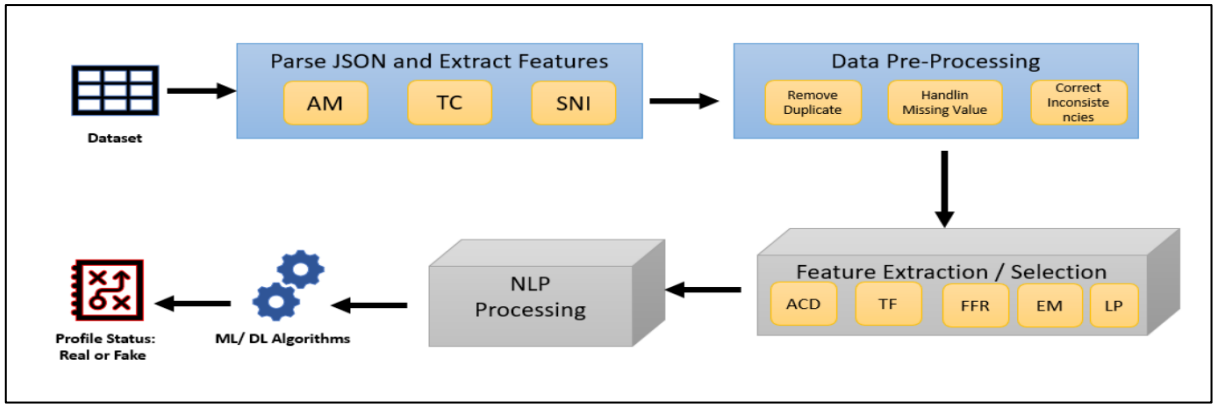

Figure 1: Overview of proposed system model

Building on this base, the way employments Common Dialect Handling (NLP) methods that are particular to Twitter to see at content. To form the tweet substance less demanding to studied, strategies like tokenization, stopword expulsion, stemming, lemmatization, and content normalization are utilized to clean it up. After the content has been handled, it is utilized with progressed methods such as the bag-of-words demonstrate, TF-IDF calculations, and word embeddings to drag out valuable data from it. There are both account-level features and prepared printed information in this carefully cleaned dataset, which makes it a solid establishment for examination and show building. With these characteristics in put, distinctive classification strategies can be utilized to discover fake profiles. A few of the calculations that are utilized are Naïve Bayes, Arbitrary Timberland, AdaBoost, Back Vector Machines (SVM), Long Short-Term Memory (LSTM), Gated Repetitive Unit (GRU), and a demonstrate that combines LSTM and GRU. Each program has its claim benefits that make it less demanding to spot fake names on social systems, which are exceptionally complicated places to be. Figure-1 appears the recommended method, which incorporates these steps and makes beyond any doubt that there's a organized and adaptable way to discover fake accounts.

A. Data preprocessing

The primary step is to utilize the advertised JSON record to stack the "twibot-20" dataset. Let D stand for the dataset, which has n tests. This JSON file is perused to induce the critical characteristics required to discover fake accounts. These incorporate account subtle elements (AM), tweet content (TC), and data around the social organize (SNI). Here are some examples of the extraction handle: Within the to begin with step, the dataset is stacked. Key highlights like client data, tweet action, and social contacts are taken from the JSON structure. These parts are exceptionally vital for finding trends and oddities that are connected to fake profiles. Account data incorporates things just like the date the account was created, the number of clients, and the number of individuals who are observing the account. These things assist you figure out in case the account is genuine. Tweet substance investigation incorporates looking at the real messages, such as how regularly they are posted, what dialect is utilized, and how clients interact with each other. Data almost a social organize incorporates the joins and discussions between clients, which appears how the account acts within the greater organize.

This information sorting prepare makes sure that all the imperative information is legitimately assembled and sorted so that it can be analyzed assist. In this way, it sets the organize for afterward steps like feature engineering and classification, which make it conceivable to precisely recognize fake accounts. The information that has been prepared is appeared as $D = \{ M , \mathbf{TL} , \mathbf{SNA} \}$ D={AM,TC,SNI}. This information is utilized to train machine learning models that discover fake profiles.

- **Normalization (Min-Max Scaling):**

$$X' = (X - X\_min) / (X\_max - X\_min)$$

Where:

- X' is the normalized value,
- X is the original feature value,

- X_min and X_max are the minimum and maximum values of the feature, respectively.

2. **Term Frequency-Inverse Document Frequency (TF-IDF):**

$$TF - IDF(t,d) = TF(t,d) * \log\left(\frac{N}{DF(t)}\right)$$

B. Feature Selection and Extraction

Within the location of fake social media profiles, successful highlight determination and extraction are significant steps to upgrade the exactness of machine learning models. The method starts with distinguishing key highlights that separate veritable accounts from false ones. Imperative highlights for social media account investigation incorporate account creation date, tweet recurrence, follower-to-following proportion, engagement measurements (likes, retweets, answers), and profile movement. These highlights offer assistance in evaluating anomalous behavior designs ordinary of fake accounts.

For text-based highlights, Common Dialect Handling (NLP) strategies are connected to extricate profitable data from tweet substance. This incorporates tokenization, stopword evacuation, stemming, and lemmatization to normalize the content. After preprocessing, strategies like Term Frequency-Inverse Record Recurrence (TF-IDF) and word embeddings are utilized to capture important representations of content information. By selecting the foremost important highlights and extricating designs from both account-level metadata and tweet substance, the dataset gets to be more reasonable and enlightening for show preparing. Appropriate include determination diminishes clamor and computational complexity, driving to more productive and exact classification of fake profiles. This step lays the establishment for executing machine learning calculations to identify false accounts on social systems.

C. NLP Processing on Tweeter Dataset

A number of Common Dialect Preparing (NLP) strategies are utilized to clean and get ready the content information in a Twitter collection so that it can be analyzed.

- Tokenization is the primary step. This can be the method of breaking the first text into little pieces called tokens. This strategy helps separate each word so that it can be examined assist. "Usually a great day!" would be turned into ["This", "can", be "a", "extraordinary", "day", "!"].
- After that, stopword expulsion is done. Stopwords are ordinary words like "could," "be," a and "the" that do not include anything useful to the think about. Able to center on more vital terms by getting freed of these words. The stopwords "is" and "a" would be taken out in this case, taking off ["This", "incredible", "day", "!"].
- At that point, stemming is utilized to induce words back to their most fundamental frame. For occurrence, "running" turns into "run." In this case, the composing is made easier, which makes a difference computers discover patterns. Within the same way, lemmatization makes this indeed superior by changing words into their lexicon frame depending on their setting. "Way better" seem gotten to be "wagered" through stemming, but "great" would remain the same through lemmatization.
- At last, normalization makes beyond any doubt that the content is reliable by doing things like changing all the words to lowercase. After this final step, the dataset is more organized and uniform, and it is now ready for highlight extraction and encourage handling to discover fake profiles.

Table 2: Processing of textual content through various stages

Original Text	Tokenized Text	Stopwords Removed	Stemmed Text	Lemmatized Text	Normalized Text
I love learning new things!	["I", "love", "learning", "new", "things", "!"]	["love", "learning", "new", "things"]	["love", "learn", "new", "thing"]	["love", "learn", "new", "thing"]	["love", "learn", "new", "thing"]
This is an amazing	["This", "is", "an", "amazing",	["amazing", "project"]	["amaz", "project"]	["amazing", "project"]	["amazing", "project"]

project.	"project", "."]				
Can’t wait to see more updates!	["Can’t", "wait", "to", "see", "more", "updates", "!"]	["wait", "see", "more", "updates"]	["wait", "see", "more", "updat"]	["wait", "see", "more", "update"]	["wait", "see", "more", "update"]
The results are very promising.	["The", "results", "are", "very", "promising", "."]	["results", "promising"]	["result", "promis"]	["result", "promising"]	["result", "promising"]
You should definitely try this.	["You", "should", "definitely", "try", "this", "."]	["definitely", "try"]	["definit", "tri"]	["definitely", "try"]	["definitely", "try"]

D. Machine Learning Methods

1. Naive Bayes

A random machine learning method called Naïve Bayes is often used to sort text into groups, like finding fake social network accounts. Naïve Bayes can sort Twitter accounts into groups based on things like the text of tweets, engagement measures, and how users act. It works by thinking that each feature is separate from every other feature, which makes the process easier. Even though it's very basic, Naïve Bayes is very good at finding spam or fake profiles, especially when it's paired with text data feature extraction methods like TF-IDF.

Prior Probability Calculation:

$$P(C) = \frac{Number\ of\ samples\ in\ class\ C}{Total\ number\ of\ samples}$$

- Where P(C) is the prior probability of class C (e.g., fake or genuine profile).

Likelihood Calculation:

$$P(x_i|C) = \frac{Number\ of\ times\ feature\ x_i appears\ in\ class\ C}{Total\ number\ of\ features\ in\ class\ C}$$

- Where P(x_i|C) is the likelihood of feature x_i given class C.

Posterior Probability Calculation (Bayes' Theorem):

$$P(C|X) = \frac{P(C) * product(P(x_i|C) for\ i = 1\ to\ n)}{P(X)}$$

- Where P(C|X) is the posterior probability of class C given feature set X, and P(X) is the evidence.

Classification:

$$C_{pred} = argmax_C P(C|X)$$

- Where C_pred is the predicted class that maximizes the posterior probability.

2. Random Forest

Random Forest could be a sort of outfit learning that builds numerous choice trees and after that joins the estimates they make. Each tree within the forest makes a figure, and the lesson with the foremost votes is chosen as the result. This is often accomplished through a larger part voting framework. Choice trees with N levels are used to discover the anticipated lesson C.

$$C = \boldsymbol{arg}Ci\boldsymbol{max} \sum_{j=1}^{N} I(yj - ci)$$

This strategy picks the class that has the foremost votes over all choice trees as the conclusion result. This makes the framework more stable and accurate than a single choice tree. Multiple trees lower the chance of overfitting and boost generalization, which is why random forest could be a common choice for employments that ought to classify things.

3. AdaBoost

The title AdaBoost comes from the phrase "Adaptive Boosting." It could be a sort of outfit learning that combines powerless models into a more grounded, more accurate demonstrate. As appeared in h (x) h t (x), the most thought behind AdaBoost is to combine the comes about of a few frail learners, which might not do well on their claim, to create a solid predictor. Most of the time, these powerless learners are basic models, such as choice stumps (moreover called single-level choice trees). The ultimate prediction in AdaBoost is found by including up the figures from all the frail categories and figuring out their weights. Typically how to create the conclusion show:

$$F(x) = sign \sum_{t=1}^{T} \left(T\alpha tht(x)\right)$$

AdaBoost works by iteratively altering the weights of misclassified illustrations. After each frail learner makes its forecasts, the calculation increments the weight of erroneously classified occasions, driving the another powerless learner to center more on the harder-to-classify illustrations. Over time, the demonstrate gets to be more exact because it centers on minimizing blunders, making AdaBoost viable for classification errands where information may be troublesome to classify with a single demonstrate.

4. Support vector machine

Support Vector Machines (SVMs) are a strong guided learning method that can be used for both sorting and predicting. The main idea behind SVMs is to find a hyperplane that best separates data points that belong to different groups. In a binary classification problem, the decision function tells us which side of the hyperplane a given data point is on, which tells us what class it is likely to belong to.

$$f(x) = sign \sum_{i=1}^{N} (\alpha iyiK(xi, x) + b)$$

SVM works by locating the best hyperplane that makes the difference between the two classes as big as possible. This makes sure that the model for classifying works well with new data. With the help of support vectors and kernel functions, SVM can deal with complicated data distributions, which makes it useful for many classification jobs.

E. Deep Learning Model

1. LSTM

LSTM is a type of recurrent neural network (RNN) that is meant to find long-term relationships in linear data. This makes it perfect for jobs like finding fake social media accounts. LSTM is used in this study to look for trends in how people behave and what they share over time. LSTM is great at dealing with the timing parts of social media interactions because it keeps important information from earlier in the process and gets rid of useless data.

Forget Gate: The forget gate decides which information from the previous cell state should be discarded.

$$f_t = \sigma\left(W_f * \left[h_{\{t-1\}}, x_t\right] + b_f\right)$$

Input Gate: The input gate controls which values from the input will update the cell state.

$$i_t = \sigma\left(W_i * \left[h_{\{t-1\}}, x_t\right] + b_i\right) C_{tilde_t} = \tanh\left(W_C * \left[h_{\{t-1\}}, x_t\right] + b_C\right)$$

Cell State Update: The new cell state is calculated as:

$$C_t = f_t * C_{\{t-1\}} + i_t * C_{tilde_t}$$

Output Gate: The output gate controls the output based on the cell state and the current input.

$$o_t = \sigma(W_o * [h_{\{t-1\}}, x_t] + b_o) h_t = o_t * \tanh(C_t)$$

2. GRU

Another type of recurrent neural network (RNN) is the gated recurrent unit (GRU). This network is often used for jobs that depend on an order, like finding fake social network accounts. GRU works like LSTM, but its structure is easier. This makes it faster to compute while still being good at detecting how data changes over time. When it comes to finding fake profiles, GRU helps look at trends like how often people tweet, what language they use, and how engaged they are with the content. GRU controls the flow of information by using two gates, reset and update, to decide what parts of the old information to keep and which to throw away.

Update Gate:

$$z_t = \sigma(W_z * [h_{\{t-1\}}, x_t] + b_z)$$

Reset Gate:

$$r_t = \sigma(W_r * [h_{\{t-1\}}, x_t] + b_r)$$

Current Memory Content:

$$h_{tilde_t} = \tanh(W_h * [r_t * h_{\{t-1\}}, x_t] + b_h)$$

Final Hidden State:

$$h_t = (1 - z_t) * h_{\{t-1\}} + z_t * h_{tilde_t}$$

3. Hybrid LSTM+GRU

This model uses the best features of the Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) designs to make sequential data analysis faster. For example, it can be used to find fake social network accounts. LSTM is great at finding long-term relationships in data, while GRU uses less computing power and does a better job with shorter patterns. The combined method uses both models together, taking advantage of LSTM's ability to keep important information over longer sequences and GRU's speed at handling shorter relationships. This combo makes the model more accurate and flexible overall, which makes it very useful for hard jobs that need to be done quickly.

LSTM Forget Gate: The forget gate of LSTM decides what information to discard from the previous cell state.

$$f_{t_{LSTM}} = \sigma(W_{f_{LSTM}} * [h_{\{t-1\}_{LSTM}}, x_t] + b_{f_{LSTM}})$$

LSTM Input Gate: The input gate of LSTM decides what new information to store in the cell state.

$$i_{t_{LSTM}} = \sigma(W_{i_{LSTM}} * [h_{\{t-1\}LSTM}, x_t] + b_{i_{LSTM}}) C_{tilde_{t_{LSTM}}} = \tanh(W_{C_{LSTM}} * [h\{t-1\}_{LSTM}, x_t] + b_{C_{LSTM}})$$

LSTM Cell State Update: The cell state is updated using the forget gate and the new input.

$$C_{t_{LSTM}} = f_{t_{LSTM}} * C_{\{t-1\}_{LSTM}} + i_{t_{LSTM}} * C_{tilde_{t_{LSTM}}}$$

GRU Update Gate: The update gate in GRU decides how much of the previous information needs to be passed to the next state.

$$z_{t_{GRU}} = \sigma(W_{z_{GRU}} * [h_{\{t-1\}_{GRU}}, x_t] + b_{z_{GRU}})$$

GRU Reset Gate: The reset gate in GRU determines how much of the previous hidden state to forget.

$$r_{t_{GRU}} = \sigma(W_{r_{GRU}} * [h_{\{t-1\}_{GRU}}, x_t] + b_{r_{GRU}})$$

Final Hybrid Hidden State: The final hidden state combines both LSTM and GRU hidden states.

$$h_t = (1 - z_{t_{GRU}}) * h_{\{t-1\}GRU} + z_{t_{GRU}} * \tanh(W_{h_{GRU}} * [r_{t_{GRU}} * h\{t-1\}_{GRU}, x_t] + b_{h_{GRU}})$$

IV. RESULT AND DISCUSSION

When different machine learning models are used to find fake accounts on social networks, the results shown in Table 3 demonstrate a thorough evaluation. Critical performance measures, such as Accuracy, Precision, Recall, and F1 Score, all given in numbers, are used to judge each model. These measures are necessary to find out how well the models work at finding fake profiles and reducing the number of wrong labels. First, let's look at how well each model did and figure out what the results mean. With a precision of 89.6%, a recall of 87.69%, and an F1 score of 89.23%, the Naïve Bayes (NB) model gets a score of 88.61%. Even though this model works, it's not perfect for this job because it assumes that features are independent. When looking for fake profiles, things like activity numbers, account age, and tweet content are all connected in complicated ways that make it harder for NB to pick up on these subtleties. Naïve Bayes still does pretty well, especially since it has a good F1 score, which means it has a good mix between accuracy and memory. Random Forest (RF) does a lot better than Naïve Bayes. It has an F1 score of 90.9%, an accuracy of 93.45%, a precision of 94.25%, a recall of 91.2%, and a recall of 94.25%. RF is an ensemble method that builds several decision trees and then adds up all of their estimates, which makes them more reliable and useful in more situations. Random Forest is very good at telling the difference between real and fake profiles, as shown by its better accuracy and precision. This is likely because it can record complex feature interactions. But recall is a little lower than accuracy, which means that even though it is good at finding fake profiles, it might miss some.

Table 3: Performance Comparison of Machine Learning Models and DL model with proposed model for Fake Profile Detection

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1 Score (%)**
Naïve Bayes (NB)	88.61	89.6	87.69	89.23
Random Forest (RF)	93.45	94.25	91.2	90.9
AdaBoost	94.4	92.2	93.8	95.61
Support Vector Machine (SVM)	92.14	93.47	90.8	90.2
GRU	95.25	95.66	94.25	95.78
LSTM	93.6	94.15	94.55	93.75
Hybrid LSTM+GRU	**98.89**	**98.9**	**98.63**	**99.14**

AdaBoost is another ensemble method that does a great job. Its accuracy is 94.4%, its precision is 92.2%, its recall is 93.8%, and its F1 score is 95.61%, which is very high. AdaBoost's main strength is that it can make weak classifiers better by focusing on the examples that other classifiers got wrong, comparision illustrate in figure 2.

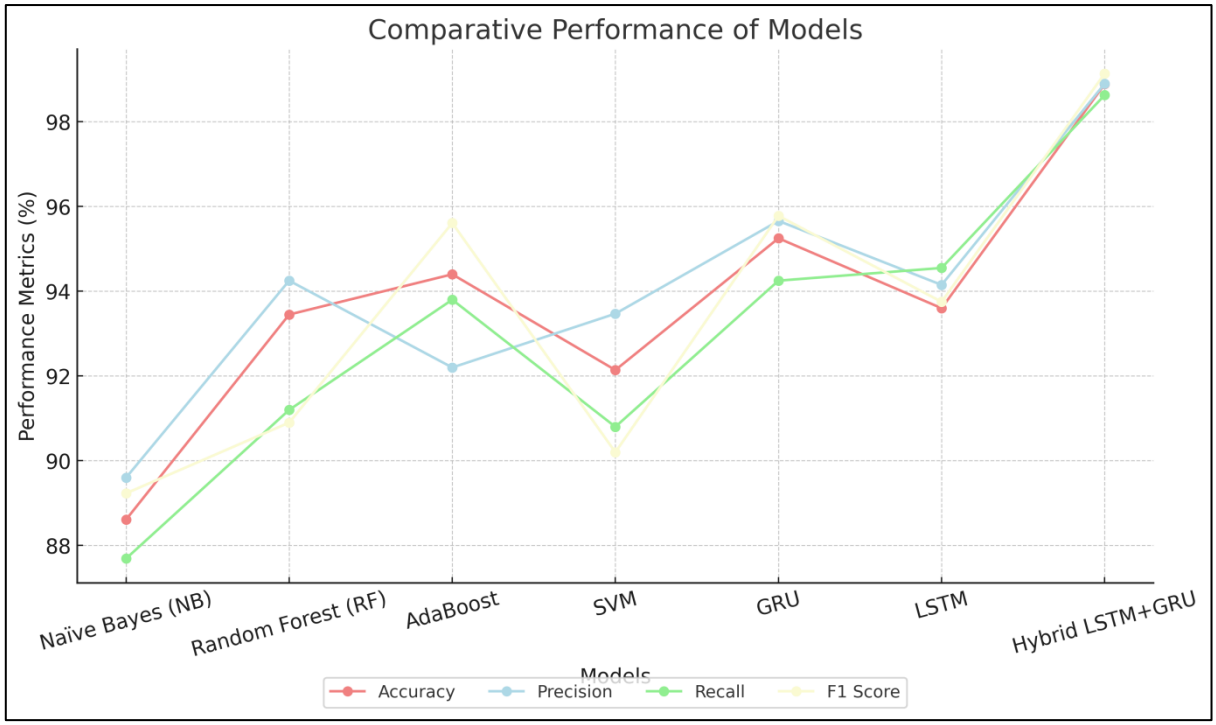

Figure 2: Comparative analysis of model performance

This repeated process makes the model more accurate and reliable, as shown by its high results. It's important to note that the F1 score, which is a harmonic mean of accuracy and recall, shows that AdaBoost does a great job of finding fake profiles and reducing the number of false hits. Support Vector Machine (SVM) does a good job; it gets 92.14% of the time right, 93.47% of the time right, 90.8% of the time right, and has an F1 score of 90.2%. SVM works well, especially when it comes to accuracy, where it does better than Naïve Bayes and is close to AdaBoost. SVM works well for binary classification jobs since it tries to find the best hyperplane that makes the difference between classes as big as possible. But SVM may not work as well as AdaBoost and Random Forest because it is easily confused by noise and outliers. This could be why it has lower recall and F1 score.

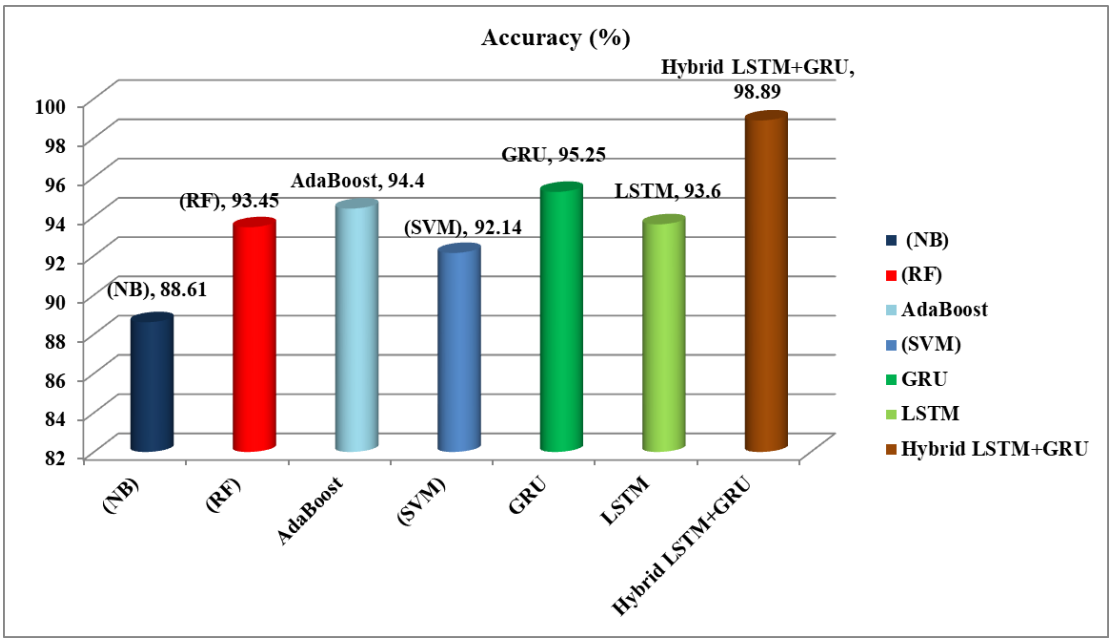

Figure 3: Accuracy comparison of ML and DL model with proposed method

With an F1 score of 95.78%, the GRU model, a type of recurrent neural network (RNN), does a great job. Its accuracy is 95.25%, its precision is 95.66%, its recall is 94.25%, and its accuracy is 95.25%. GRU's best feature is

that it can handle sequential data, which makes it perfect for looking at how users behave over time. When it comes to social networks, GRU can effectively find trends in how users interact, what they share, and how often they post, which leads to high memory and accuracy, as comparison shown in figure 3. The F1 score shows that GRU does a good job of finding fake profiles while also avoiding false hits. One more type of RNN is LSTM, which has an F1 score of 93.75%, an accuracy of 93.6%, a precision of 94.15%, a recall of 94.55%, and a recall of 94.55%. The structure of LSTM is made to keep long-term relationships in sequential data. This is very important for finding fake profiles where behavior patterns might not be obvious at first glance. Even though LSTM doesn't do as well as GRU in this case, its high recall means that it is very good at finding fake accounts, though it might sometimes get real accounts wrong. All of the tests show that the Hybrid LSTM+GRU model works the best. It has an F1 score of 99.14%, an accuracy score of 98.89%, a precision score of 98.9%, and a memory score of 98.63%. This combination model takes advantage of both LSTM and GRU's skills by mixing them. LSTM can record long-term relationships, and GRU is good at speeding up computations. The model that was made is very good at both accuracy and memory, which means it can find fake accounts on social networks very accurately and reliably.

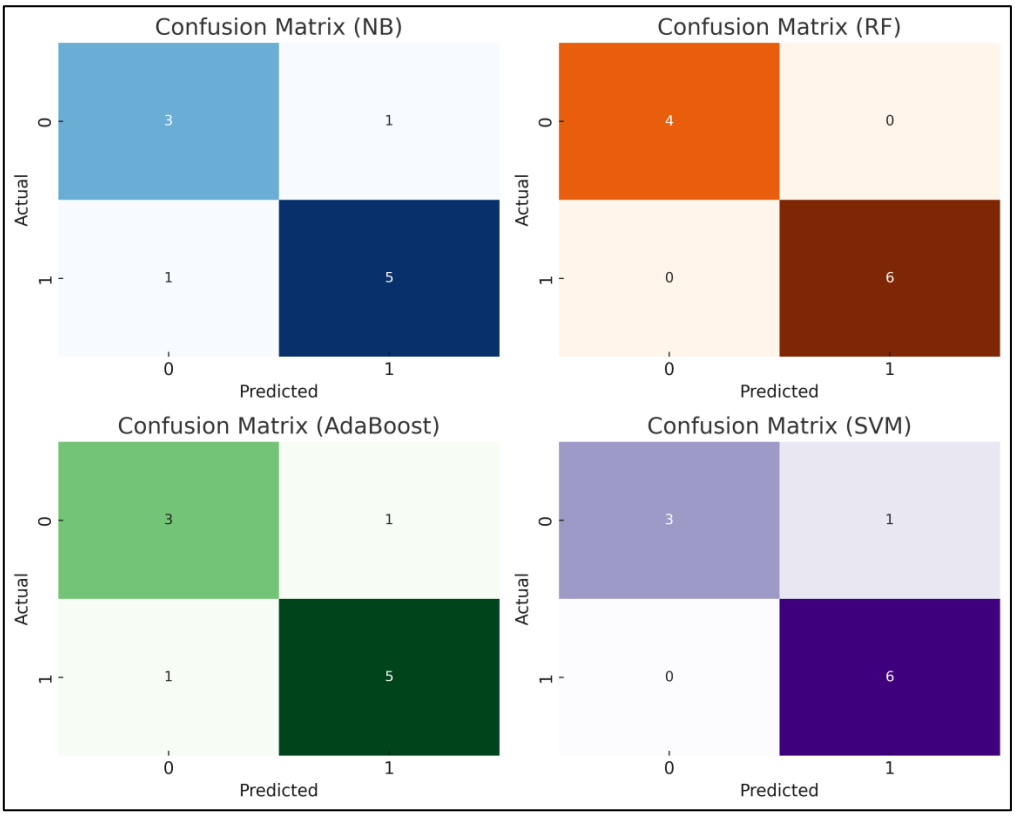

Figure 4: Confusion matrix ML Models

The almost perfect F1 score shows that this model does a good job of balancing accuracy and memory, keeping false positives and false negatives to a minimum. The Hybrid LSTM+GRU model definitely does better than all the others, showing higher F1 score, accuracy, precision, and memory. The results show that using both LSTM and GRU together can pick up on both long-term and short-term relationships in user behavior. This makes it a very good way to find fake profiles. Naïve Bayes and SVM are two traditional machine learning models that do a good job, but they are not as good as AdaBoost, GRU, and LSTM, which are more advanced ensemble and recurrent models. Overall, using advanced RNNs, especially the Hybrid LSTM+GRU, makes it much easier to spot fake profiles, confusion matrix for ML and DL model shown in figure 4 and 5 respectively. This makes them perfect for dealing with the complicated and changing nature of social media data.

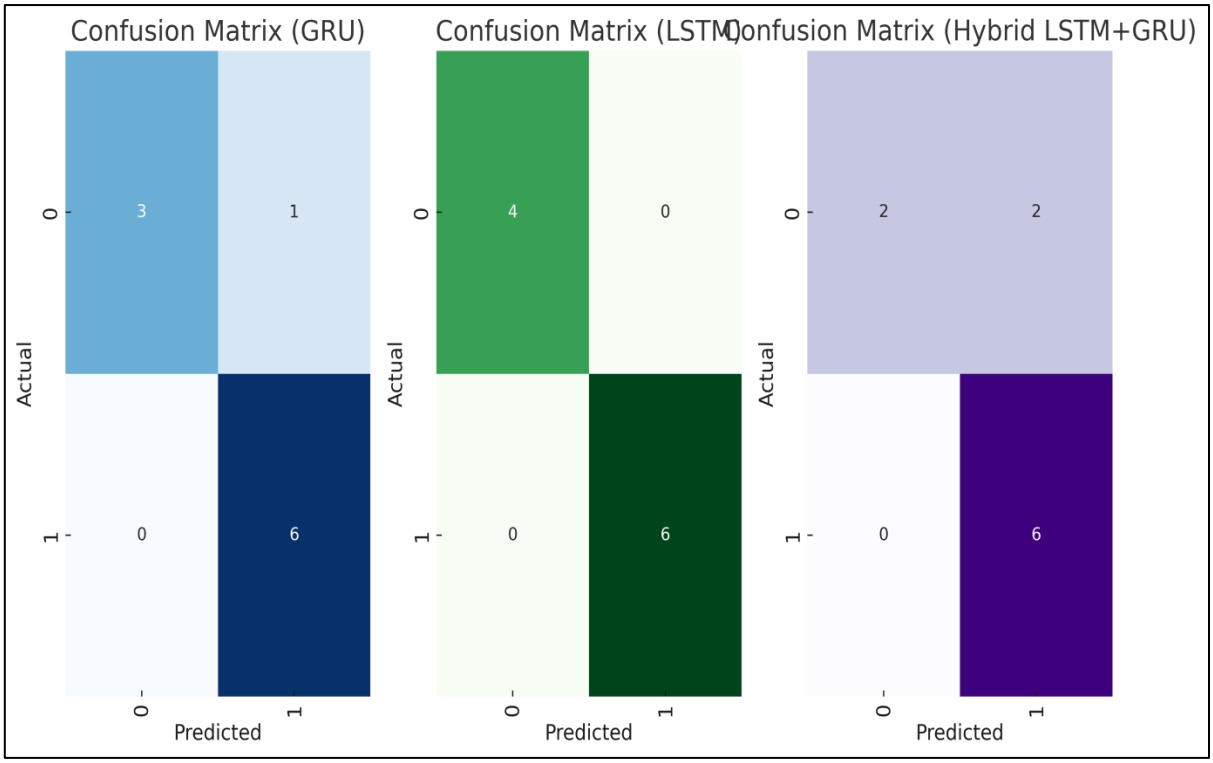

Figure 5: Confusion matrix for DL model with Proposed Method

V. CONCLUSION

In this consider appeared a blended show that employments Long Short-Term Memory (LSTM) and Gated Repetitive Unit (GRU) structures to way better discover fake profiles on social systems, with the most objective of boosting client believe. Having a parcel of fake profiles is exceptionally unsafe since it can lead to character burglary, money related tricks, and the spread of untrue data. Finding these profiles rapidly and accurately is exceptionally imperative for keeping social stages legitimate. The blended demonstrate is awesome at catching both long-term and short-term behavioral patterns, which are vital for finding fake accounts since it employments the benefits of both LSTM and GRU. The comparison of diverse machine learning models, such as Naïve Bayes, Arbitrary Woodland, AdaBoost, Bolster Vector Machines (SVM), GRU, and LSTM, appears that the Crossover LSTM+GRU demonstrate continuously does way better than the ancient ways of doing things. The hybrid show may be an exceptionally great way to discover fake profiles since it has tall precision, accuracy, review, and F1 score. This combined strategy viably gets around the issues that single models have, like not being able to handle time connections or the trouble of finding the proper adjust between quick computing and precise expectation. Characteristic Dialect Handling (NLP) strategies like tokenization, word embeddings, and disposition examination are too utilized to make strides the set of highlights. This lets fake profiles be found based on substance and client behavior. The truth that the blended demonstrate can learn from a wide run of inputs appears how well it can spot diverse sorts of tricks. The recommended joint LSTM+GRU demonstrate may be a huge step forward within the battle against fake profiles on social systems, and it gives a solid way to form computerized spaces more secure and more reliable for clients.

REFERENCES

[1] P. Wanda, “RunMax: fake profile classification using novel nonlinear activation in CNN,” Soc. Netw. Anal. Min., vol. 12, no. 1, pp. 1–11, 2022, doi: 10.1007/s13278-022-00983-9.

[2] M. Vyawahare and S. Govilkar, “Fake profile recognition using profanity and gender identification on online social networks,” Soc. Netw. Anal. Min., vol. 12, no. 1, pp. 1–13, 2022, doi: 10.1007/s13278-022-00997-3.

[3] T. Sudhakar, B. C. Gogineni, and J. Vijaya, “Fake Profile Identification Using Machine Learning,” Proc. 2022 IEEE Int. Women Eng. Conf. Electr. Comput. Eng. WIECON-ECE 2022, pp. 47–52, 2022, doi: 10.1109/WIECON-ECE57977.2022.10150753.

[4] D. Sharma and E. R. S. Madan, “ANN based Fake User Profile Detection,” Int. J. Res. Eng. Emerg. Trends, vol. 6, no. 2, pp. 460–465, 2022.

[5] B. O. Saracoglu, “Initialization of profile and social network analyses robot and platform with a concise systematic review,” Mach. Learn. with Appl., vol. 7, no. January, p. 100249, 2022, doi: 10.1016/j.mlwa.2022.100249.

[6] A. M. Meligy, H. M. Ibrahim, and M. F. Torky, “Identity Verification Mechanism for Detecting Fake Profiles in Online Social Networks,” Int. J. Comput. Netw. Inf. Secur., vol. 9, no. 1, pp. 31–39, 2017, doi: 10.5815/ijcnis.2017.01.04

[7] P. Hajek, A. Barushka, and M. Munk, “Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining,” Neural Comput. Appl., vol. 32, no. 23, pp. 17259–17274, 2020, doi: 10.1007/s00521-020-04757-2.

[8] S. Gupta, B. Verma, P. Gupta, L. Goel, A. K. Yadav, and D. Yadav, “Identification of Fake News Using Deep Neural Network-Based Hybrid Model,” SN Comput. Sci., vol. 4, no. 5, p. 679, 2023, doi: 10.1007/s42979-023-02117-0.

[9] Chekuri, “Fake profile detection in using machine learning,” J. Eng. Sci., vol. 13, no. 08, pp. 751–754, 2022

[10] M. Conti, R. Poovendran, and M. Secchiero, “FakeBook: Detecting fake profiles in on-line social networks,” Proc. 2012 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Mining, ASONAM 2012, pp. 1071– 1078, 2012, doi: 10.1109/ASONAM.2012.185.

[11] K. K. Bharti and S. Pandey, “Fake account detection in twitter using logistic regression with particle swarm optimization,” Soft Comput., vol. 25, no. 16, pp. 11333–11345, 2021, doi: 10.1007/s00500-021-05930- y.

[12] S. R. Sahoo and B. B. Gupta, Real-time detection of fake account in twitter using machine-learning approach, vol. 1086. Springer Singapore, 2021.

[13] A. Mitra, A. Kundu, M. Chattopadhyay, and A. Banerjee, “An Approach to Detect Fake Profiles in Social Networks Using Cellular Automata-Based PageRank Validation Model Involving Energy Transfer,” SN Comput. Sci., vol. 3, no. 6, p. 423, 2022, doi: 10.1007/s42979-022-01315-6.

[14] A. Mewada and R. K. Dewang, “CIPF: Identifying fake profiles on social media using a CNN-based communal influence propagation framework,” Multimed. Tools Appl., 2023, doi: 10.1007/s11042-023-16685-z.

[15] Durga, P.; Sudhakar, D.T. The use of supervised machine learning classifiers for the detection of fake Instagram accounts. J. Pharm. Negat. Results 2023, 14, 267–279.

[16] Prakash, O.; Kumar, R. Fake Account Detection in Social Networks with Supervised Machine Learning. In International Conference on IoT, Intelligent Computing and Security. Lecture Notes in Electrical Engineering; Agrawal, R., Mitra, P., Pal, A., Sharma Gaur, M., Eds.; Springer: Singapore, 2023; Volume 982, pp. 287–295.

[17] Kanagavalli, N.; Sankaralingam, B.P. Social Networks Fake Account and Fake News Identification with Reliable Deep Learning. Intell. Autom. Soft Comput. 2022, 33, 191–205.

[18] Bhattacharyya, A.; Kulkarni, A. Machine Learning-Based Detection and Categorization of Malicious Accounts on Social Media. In Social Computing and Social Media. HCII 2024. Lecture Notes in Computer Science; Coman, A., Vasilache, S., Eds.; Springer: Cham, Switzerland, 2024; Volume 14703, pp. 328–337.

[19] Goyal, B.; Gill, N.S.; Gulia, P.; Prakash, O.; Priyadarshini, I.; Sharma, R.; Obaid, A.J.; Yadav, K. Detection of Fake Accounts on Social Media Using Multimodal Data With Deep Learning. IEEE Trans. Comput. Soc. Syst. 2023, 1–12.

[20] M. Bende, M. Khandelwal, D. Borgaonkar and P. Khobragade, "VISMA: A Machine Learning Approach to Image Manipulation," 2023 6th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2023, pp. 1-5, doi: 10.1109/ISCON57294.2023.10112168.

[21] Wani, M.A.; Agarwal, N.; Jabin, S.; Hussain, S.Z. Analyzing real and fake users in Facebook network based on emotions. In Proceedings of the 2019 11th International Conference on Communication Systems & Networks (COMSNETS), Bengaluru, India, 7–11 January 2019; pp. 110–117.

[22] Gupta, A.; Kaushal, R. Towards detecting fake user accounts in facebook. In Proceedings of the 2017 ISEA Asia Security and Privacy (ISEASP), Surat, India, 29 January–1 February 2017; pp. 1–6.

[23] Boshmaf, Y.; Logothetis, D.; Siganos, G.; Lería, J.; Lorenzo, J.; Ripeanu, M.; Beznosov, K. Integro: Leveraging victim prediction for robust fake account detection in OSNs. In Proceedings of the Network and

Distributed System Security Symposium 2015 (NDSS'15), San Diego, CA, USA, 8–11 February 2015; pp. 1–15.

[24] Hamed, S.K.; Ab Aziz, M.J.; Yaakub, M.R. Fake News Detection Model on Social Media by Leveraging Sentiment Analysis of News Content and Emotion Analysis of Users' Comments. Sensors 2023, 23, 1748. https://doi.org/10.3390/s23041748