Improving Accuracy of Twitter Fake Profile Detection Using Deep Learning

Chavan Akanksha[1], Kakde vaibhavi[2],Dhakane Vaibhav[3] ,Nalawade Prajkta[4]

[1-4] *Dept of Computer Engineering, SVPM'S College of Engineering, Malegaon (Pune),* India

---***---

ABSTRACT:
Outline the issue of fake accounts on popular social media platforms like Twitter, which spread false information, malicious content, and spam. Online social networks have grown rapidly, with billions of users worldwide. This growth has led to many fake ac- counts, causing problems like spam, fake news, and political manipulation. Fake accounts can also harm businesses financially and damage their reputation. Therefore, detecting these fraudulent accounts is crucial. Recently, researchers have been using neural network algorithms to identify fake accounts more effectively. Our system uses several types of neural networks, including feedforward and recurrent neural networks, as well as deep learning models, to address this issue. Specifically, we combine artificial neural networks (ANN) with principal component analysis (PCA) to create a reliable system for spotting fake accounts on social media. By collecting and processing data thoroughly, extracting important features, and training the ANN, we show that our method is better than traditional ones at detecting fake accounts. Our results highlight the potential for greater accuracy and efficiency in protecting the integrity of online social networks.

1. INTRODUCTION

Over the last two decades, there have been tremendous growths of social networking, bringing millions of users from all walks of life to this new form of online interaction. However, such rapid growth calls for a raising concern related to a new phenomenon: fake accounts that represent nobody at all. These bot accounts spread false information and manipulate web ratings by generating spam, amongst other abusive practices which are prohibited on sites such as Twitter. Examples of inappropriate use of social media include automated interaction, attempts at manipulating or misleading end-users, and worse behaviors such as posting malicious links, aggressive following and following/unfollowing others, multiple account creation, updating duplication, and reply and mention functions abuse. Authentic posts, however, adhere to the site's guidelines. This phenomenon has far- reaching effects since real-time tweets and messages are received, allowing information to swiftly reach a huge number of users. One of the biggest prob- lems with social media is spammers, who utilize their accounts for a variety of nefarious purposes, like spreading misinformation that can harm any company and have an im- pact on society as a whole.

Therefore, the research's goals will be focused on addressing the problem of identifying phony profile accounts on the Twitter online social network. This technology's primary goal is to stop the proliferation of different types of bogus news, advertisements, and followers. Furthermore, the rapid growth of social media platforms like Facebook, LinkedIn, Twitter, and Instagram has made it possible for over half of the world's population to be active internet users and engage in social media activities due to the enormous advancements in wireless communication technology. The proliferation of fake accounts has led to serious problems like the spread of fake news, political manipulation, hate speech, and spam activities that threaten the legitimacy and dependability of online social networks, even though the user base has grown exponentially. Machine learning techniques are becoming essential for identifying phony accounts in order to lessen these difficulties. But, cyber-thieves keep designing bots which can transcend such detection strategies. So, it is an ongoing cat-and-mouse game between the detection algorithms and malefactors. In this study, we approached a structured approach: the literature re-view, explanation of the proposed detection technique, and comparative analysis of results from various algorithms; thus, it gave a contribution towards the development of techniques to identify and fight the spread.

2. METHODOLOGY

This section describes the process used to identify phony Twitter profiles using a deep learning model. Data preprocessing, feature engineering, and model training and evaluation are the three main phases of our methodology. Preprocessing Data: The MIB dataset, which contains information on user profiles and activities, is first preprocessed. Natural Language Processing (NLP) methods such as language recognition, tokenization, stop word removal, and TF-IDF conversion are applied to textual features, especially user descriptions. Through this procedure, textual data is converted into a numerical representation that deep learning models can use. To make sure all features numerical and post-NLP— contribute equally during training, feature scaling is also used. Feature Engineering: In addition to the MIB dataset's raw characteristics, we investigate the development of new features that could improve the model's capacity to differentiate between authentic and fraudulent accounts. This could entail finding patterns suggestive of questionable activity (e.g., high favorite count with low status count) or computing ratios between current features (e.g., followers count / friends count). Model Training and Evaluation A deep learning model, namely a feed-forward neural network, serves as the primary classification engine. We test different network topologies, in- cluding the number of neurons and hidden layers, to obtain optimal performance. The model is trained on a portion of the preprocessed and maybe featureengineered MIB dataset. Evaluation parameters including as accuracy, precision, recall, and F1-score are used to assess the model's ability to identify fraudulent profiles on a separate hold- out test set. Software and Tools: During the data preprocessing and deep learning model implementation phases, libraries such as Tensor Flow may be utilized. However, for the sake of this research, we focus on a deep learning approach that is incorporated into a deep learning framework.

A. *Dataset Description:*

Our study makes use of the "MIB" dataset, which was made freely available by Crescietal. (2015). 5,301 Twitter accounts in all, divided into actual and fraudulent accounts, make up this dataset

➢ *Actual Accounts:*

- The 469 accounts in the "Fake Project" dataset were gathered by human researchers at IIT-CNR in Pisa, Italy.
- Two sociologists from the University of Perugia in Italy confirmed that the 1481 authentic human stories in the "E13 (elezioni 2013)" dataset are accurate.

➢ *False Accounts:*

- The dataset known as "Fastfollowerz" comprises 1337 accounts. The dataset titled "Intertwitter" has 1169 accounts.
- In 2013, researchers bought 845 accounts from the market to create the "Twitter- technology" dataset.

B. *Data Preprocessing;*

Data Cleaning and Missing Value Imputation: If a significant percentage of values in the dataset are missing, we use suitable methods such as mean/median imputation or deletion to resolve the missing values. This guarantees that the model has all the data it needs to be trained.

➢ *User Description Text Preprocessing:*

- We apply Natural Language Processing (NLP) methods to user descriptions, which are textual elements that may contain useful information. This procedure entail Finding the prevailing language for each description is essential for the next steps in the language detection process. Tokenization is the process of breaking up text into discrete words or characters to create a sequence that may be processed further.
- Stop Word Removal: To lower dimensionality and concentrate on pertinent content, frequent words with little meaning (such as "the" and "a") are eliminated.
- Alternatives to Label Encoding for Textual Feature Representation: One-Hot Encoding: (if there aren't too many distinct descriptions): With this method, a new binary feature is produced for every distinct category. The associated feature value is set to 1 if a text sample falls into that category and to 0 otherwise. This lets the model discover connections between features while maintaining the categorical nature of the input.
- Word embeddings: these are useful when working with big datasets or when attempting to capture semantic links Words are

represented as vectors in a high-dimensional space using this potent approach. Semantic relationships are captured by placing words that have comparable meanings closer together. GloVe and Word2Vec are two well-liked word embedding techniques. To effectively represent textual data, you can use these pre-trained embeddings into your deep learning model.

Data Reduction After data preparation, we employ Principal Component Analysis (PCA) to decrease the dimensionality of the numerical features in the MIB dataset. High dimensionality can cause the "curse of dimensionality," which makes learning difficult because of sparse data and can also make training durations longer. The majority of the variance in the data is captured by principal components (PCs), a smaller set of features found using PCA. This allows us to reduce the amount of characteristics while preserving vital information. Our goal is to strike a compromise between computational efficiency and information retention by choosing a subset of informative PCs. Our deep learning model for false profile identification is then trained using this decreased dimensionality dataset, which could result in quicker training times and better model performance. Architecture Model We use a feed-forward neural network architecture with five consecutive layers in our deep learning model for Twitter false profile identification. By successfully classifying actual and fraudulent profiles, this architecture seeks to capture complex interactions between the preprocessed characteristics (numerical and post-NLP textual elements)

Figure 1: Architecture of Neural Networks 64 neurons make up the first layer, which is dense. Since dense layers are fully coupled, every neuron in the data is able to receive input from every feature. This layer applies a ReLU (Rectified Linear Unit) activation function and combines these inputs in a weighted linear fashion. By adding nonlinearity, ReLU enables the model to recognize increasingly intricate patterns. It promotes effective learning by letting only positive values through. A dropout layer is added after the initial dense layer. During training, dropout randomly deactivates a predetermined percentage of neurons (for example, 20%). By decreasing neuronal codependency and promoting the model to learn more resilient properties, this lessens overfitting. With 32 neurons, the third and fourth levels are similarly thick. These layers extract higher-level representations from the data and gradually improve the learned features. After the second and fourth dense layers, dropout layers are positioned carefully to increase feature robustness and avoid over-fitting. One neuron with a sigmoid activation function makes up the last layer. Every neuron in the preceding dense layer sends information to this cell. The model predicts that a profile is either true (closer to 0) or fraudulent (closer to 1), and the sigmoid function returns a number between 0 and 1. The model can learn increasingly complicated feature representations from the data thanks to this architecture's combination of dropout and ReLU layers. We can classify each profile as authentic or fake based on a predetermined threshold thanks to the likelihood score that the final output layer offers.

C. Data Reduction:

Following data preprocessing, we use Principal Component Analysis (PCA) to reduce dimensionality in the MIB dataset's numerical characteristics. The "curse of dimensionality," which causes data to become sparse and learning to become challenging, can result from high dimensionality, which can also lengthen training times. Principal components (PCs), a smaller group of features identified by PCA, are responsible for capturing the most important variation in the data. This enables us to keep important information while reducing the number of features. Our strike a compromise between computing efficiency and information retention by choosing a subset of in- formative PCs. Our deep learning model for false profile identification is then trained using this reduced-dimensionality dataset, which could result in quicker training times and better model performance.

D. Algorithms:

- *NLP:*

A subfield of artificial intelligence (AI) called natural language processing (NLP) deals with the use of natural language in computer-human interaction. It includes a broad range of activities designed to make it possible for machines to comprehend, interpret, and produce data in human language.

- Breaking down text into smaller parts, such as words, phrases, or sentences, is known as tokenization. For many NLP projects, this is the initial step.
- Part-of-Speech (POS) Tagging: giving each word in a phrase a grammatical tag (such as noun, verb, or adjectival).
- Named Entity Recognition (NER): NER is the process of identifying and categorizing entities that are mentioned in text, including names of people, groups, locations, dates, etc.4. Parsing: examining a sentence's grammatical structure to determine its syntactic linkages.
- Sentiment analysis: Identifying the sentiment—whether neutral, negative, or positive—expressed in a document.
- Machine Translation: Automatically translating text between languages.
- Writing Generation: Producing writing that resembles that of a human being in response to prompts or contexts.
- Question Answering: The process of automatically determining responses to queries in natural language.
- Text Summarization: distilling lengthy texts into concise synopses while keeping the most crucial details.
- Topic Modeling: Determining the primary subjects covered in a group of documents.

➢ *PCA:*

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more subtopics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named "Heading 1," "Heading 2," "Heading 3," and "Heading 4" are prescribed.

- Standardization: Set the variables to have a standard deviation of one and a mean of zero.
- Covariance Matrix: Determine the relationship between each variable and all other variables.
- Eigende composition: Determine the covariance matrix's eigenvalues (the magnitude of variance) and eigenvectors (the directions of maximum variance).
- Principal Component Selection: To preserve the greatest amount of variance, select the top eigenvectors according to their matching eigenvalues.
- Projection: To create a new feature space, project the original data onto the chosen principal components.

➢ *Feed Forward Neural Network :*

A feedforward neural network, in which node connections do not produce cycles, is the most basic type of artificial neural network. This structure consists of an input layer, an output layer, and one or more hidden layers. There is just one direction in which information moves from input to output. Each layer's nodes use weighted connections to digest information from the previous layer before sending it to the next layer, where activation functions are often applied.

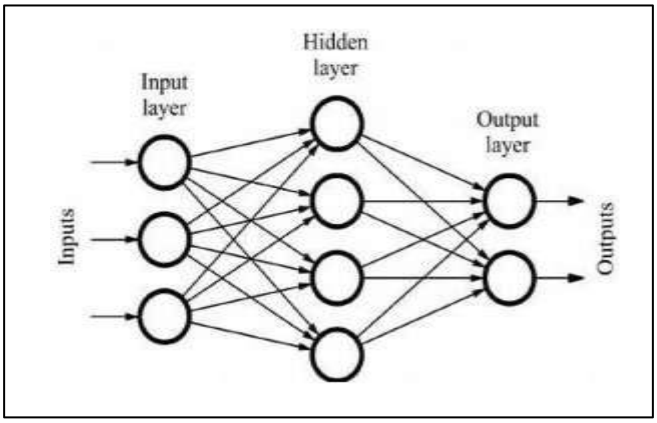


Fig 1: Feedforward Neural Network

- Input Layer: Gets the first features or data. A feature (such as the pixel value in image recognition) is represented by each node. It serves as the neural net- work's data entry point.
- Hidden Layers: In between input and output are intermediary layers. Nodes use activation functions and weighted connections to carry out

intricate transformations. They allow the network to use input data to learn abstract features.

- Weighted Connections: Weighted connections between nodes in neighboring layers. These establish the strength of the connection and are modified based on error during training. They stand for the significance of input features in predic- tion- making.
- Activation functions introduce non-linearity. applied to the weighted aggregate of the inputs, allowing the network to identify complex patterns. They facilitate the discovery of complex relationships within the data.

3. ARCHITECTURE OF SYSTEM

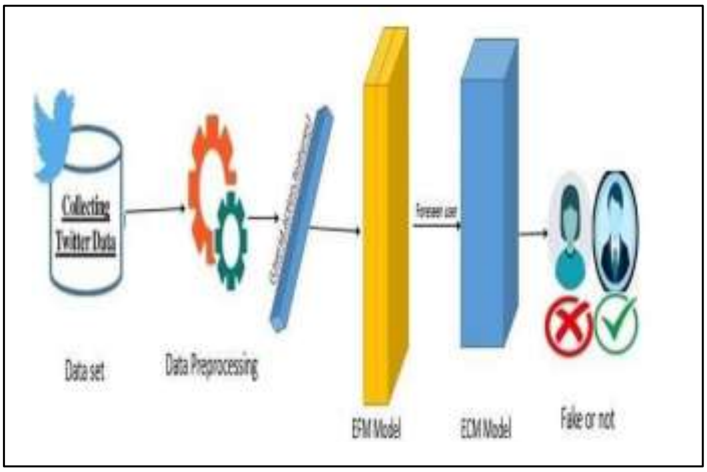


Fig 2: Architecture of System

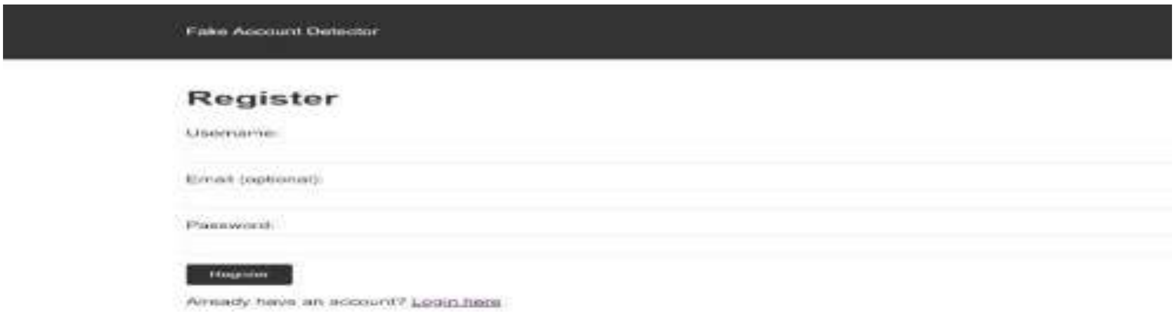
The JODIE model is optimized for early detection of malicious Twitter users by dynamically analyzing temporal user interactions. Key operations include updating and projecting user embedding, creating a trajectory to model interactions over time. The embedding foreseeing model (EFM) uses JODIE to predict future interactions, while the embedding classification model (ECM) classifies users as fake or legitimate. This layered approach to data processing in social networks helps proactively identify malicious behavior patterns before they escalate

4 .LITERATURE SURVEY

- Detecting Twitter Fake Accounts using Machine Learning and Data Reduction Techniques. In order to detect phony social media accounts, previous research used a variety of datasets and machine learning techniques. For more precise detection, recent research favors deep learning models like neural networks. This underscores the transition to neural network-based methods and under- scores the importance of diverse datasets.
- Twitter Fake Account Detection. Rise in fake accounts on Twitter poses risks such as spreading fake news and spam. To differentiate between authentic and fraudulent users, feature-based detection techniques track user activity. Different attribute sets and detection techniques were investigated in a number of studies. In order to efficiently manage numerical attributes, some studies concentrate on discretization approaches.
- Detecting Fake Accounts on Social Media. Introduces SVM-NN, a novel algorithm for fake Twitter account detection. Techniques for dimension reduction and feature selection were used in preprocessing. SVM-NN achieves good classification accuracy, outperforming other classifiers.
- Detection of Fake Profile in Online Social Networks Using Machine Learning. suggests using SVM-NN to detect phony Twitter accounts. Machine learning methods and feature selection strategies are used, with SVM-NN demonstrating the best results. It has been observed that correlation-based feature selection methods are more successful than PCA.
- In Using Machine Learning to Detect Fake Identities: Humans versus Bots The author trained and assessed the machine learning models using a corpus of social media accounts. In order to improve the detection of fraudulent ac- counts, engineering elements were added to the corpus, which was made up of social network attributes. Metrics like accuracy, F1 score, and precision-recall area under curve (PR-AUC) were used to assess each model's efficacy.

5.RESULT

1. Rasistration



Register

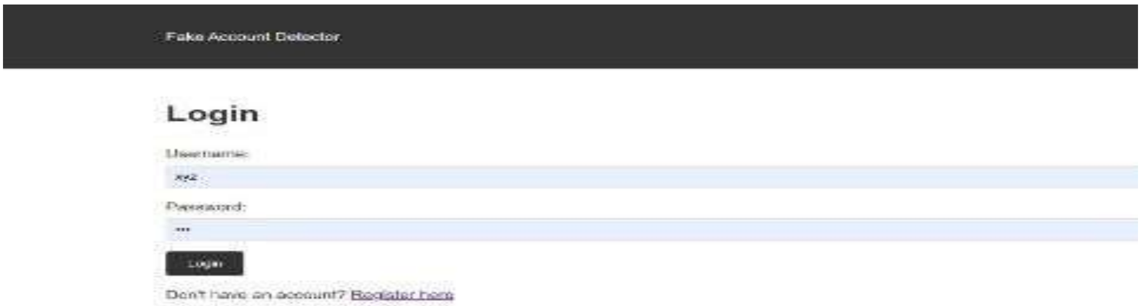
Username:

Email (optional):

Password:

Already have an account? Login here

2. Login



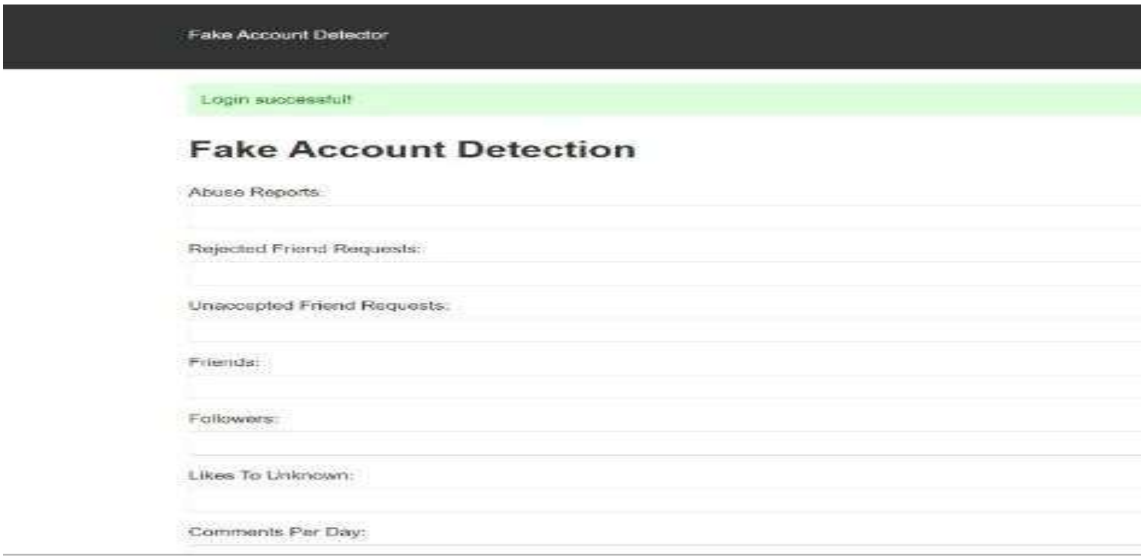
Login

Username: xyz

Password: •••

Don't have an account? Register here

3. Fake account detection



Login successful!

Fake Account Detection

Abuse Reports:

Rejected Friend Requests:

Unaccepted Friend Requests:

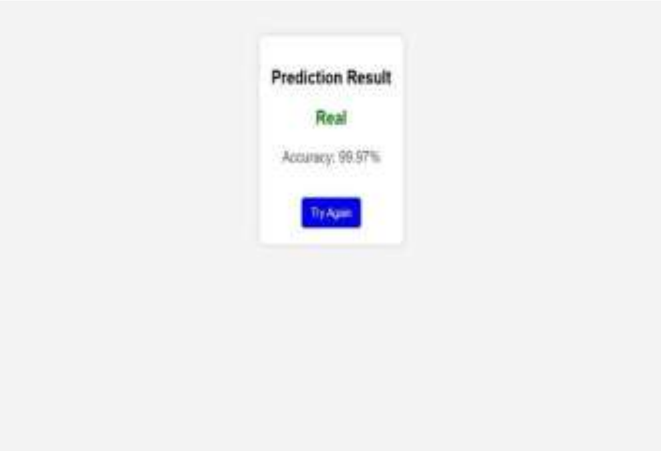
Friends:

Followers:

Likes To Unknown:

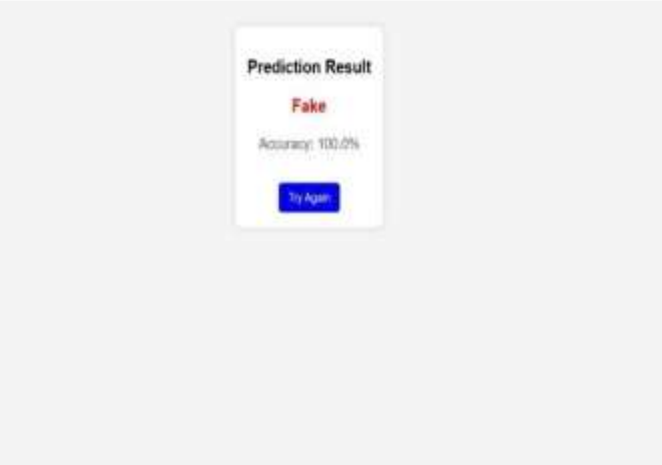
Comments Per Day:

4. Result


Prediction Result

Real

Accuracy: 99.97%


Prediction Result

Fake

Accuracy: 100.0%

6. CONCLUSION

In conclusion, our deep learning-based experiment on Twitter Fake profile identification offers promising results and insights for improving online security. Through the creation and assessment of our deep learning model, we were able to determine how well it could accurately identify Fake profiles. By examining user activity, account in- formation, and tweet content, a deep learning model offers a practical way to identify Fake Twitter profiles. Its excellent accuracy, scalability, and automation make it a use- ful tool for spotting fraudulent accounts.

FUTURE DIRECTIONS

This study creates opportunities for more investigation. The effectiveness of various deep learning architectures and optimization strategies for detecting phony profiles can be examined. Furthermore, experimenting with different text preparation techniques or adding additional user-related elements (such as network activity) might improve model performance. All things considered, this study shows how deep learning can be used to identify fake Twitter profiles. We can help create a more dependable and trust- worthy online environment by consistently improving methods and investigating novel approaches. Finally, additional study is required to ascertain whether the idea can be applied to different social media platforms.

REFERENCES

1. Sarangam Kodati, Kumbala Pradeep Reddy, Sreenivas Mekala, PL Srinivasa Murthy, and P Chandra Sekhar Reddy, Detection of Fake Profiles on Twitter Using Hybrid SVM Algorithm, "2021.

2. Buket Er¸sahin, Ozlem Akta¸s, Deniz Kılınc¸, and Ceyhun Akyol,¨ ”Twitter Fake Account Detection,” 2017.

3. Mohammad Abu Snober, ”Detecting Twitter Fake Accounts using Machine Learning and Data Reduction Techniques,” ResearchGate, 2021.

4. Ruben Sanchez-corcuera , Arkaitz Zubiaga ,”Early detection and prevention of malicious user behaviour on Twitter using Deep learning technique”,2024.

5. Louzar Oumaima, Ramdi Mariam, Baida Ouafae, Lyhyaoui Abdelouahid,"Fake Account Detection in Twitter using Long Short Term Memory and Convolutional Neural Network",2024.

6. Faisal S. Alsubaei,"Article Detection of Inappropriate Tweets Linked to Fake Accountson Twitter",2023.

7. K. Harish, R. Naveen Kumar, Dr. J. Briso Becky Bell ,"Fake Profile Detection Using Machine Learning ",2023.

8. Unreliable Users Detec- tion in Social Media Deep Learning Techniques for Automatic Detec- tion, 2020.