Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 06, June - 2024

Unmasking Deceit - A Fake Profile Detection using Artificial Neural Network

Mr.Anbazhagan.R, Mr. Jayakrishna.B,
Mr.Arun Kumar.S, Scholars
Dept of Artificial Intelligence and Data Science
Vel Tech High Tech Dr. Rangarajan
Dr. Sakunthala Engineering College
Tamilnadu,Chennai, Avadi

Mrs Geetha L, Assistant Professor
Dept of Artificial Intelligence and Data Science
Vel Tech High Tech Dr. Rangarajan
Dr. Sakunthala Engineering College
Tamilnadu,Chennai, Avadi

Abstract
In today's digital age, Online Social Networks (OSNs) have become a ubiquitous part of daily life, with individuals across all age groups dedicating significant time to engaging and sharing information worldwide. However, the expansive connectivity and extensive data exchange facilitated by these platforms also open avenues for misuse. Among the prevalent issues is the proliferation of fake accounts, utilized for nefarious purposes like targeting specific users or perpetrating cybercrimes. Consequently, there arises a pressing need for the development of robust systems capable of discerning genuine accounts from fraudulent ones. Researchers have pursued various methodologies, employing classification algorithms that leverage diverse account attributes including time-based, graph-based, content-based, and user-based features in their quest to combat this challenge on social networking sites. Machine learning algorithms have proven effective in identifying fake profiles, duplicates, spam, and bots on social networks, as demonstrated by extensive research in this area.

Keywords
Fake accounts, Instagram , machine learning, artificial neural network.

INTRODUCTION

In today's digital landscape, social networking sites have revolutionized how individuals interact, communicate, and forge connections with others.

Leveraging Web 2.0 technology, these platforms enable users to create profiles, connect with friends, share updates, and expand their networks. From Instagram to Facebook, Twitter to LinkedIn, these platforms have become ubiquitous, integral components of contemporary social life. However, alongside their widespread adoption comes the proliferation of fraudulent profiles, identity theft, and privacy breaches. . As the number of social network users grows, so does the prevalence of fake accounts, posing significant challenges for both users and platform providers. To address this issue, machine learning techniques offer promising solutions by analysing profile data and user behaviour to identify and mitigate the impact of fraudulent profiles across various online platforms.

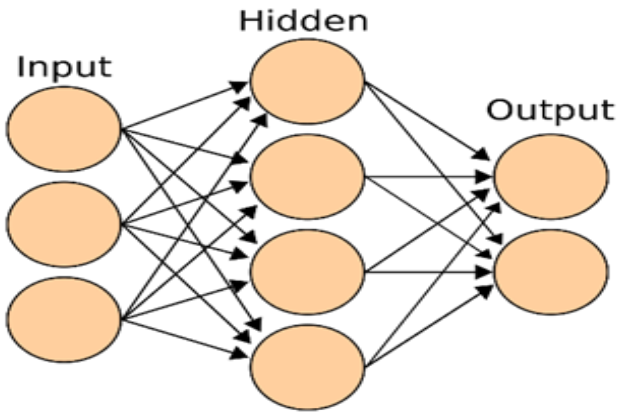
The identification of fake profiles through machine learning entails constructing a model capable of distinguishing between genuine and fraudulent profiles. This model is typically trained using a dataset comprising known fake and authentic profiles, gathered through methods such as manual identification by experts, social media platform crawling, or crowd-sourcing. Various profile features serve as classification criteria, including profile picture, name, location, age, gender, description, follower count, friend count, post frequency, engagement rate, activity time, and content. Machine learning algorithms like decision trees, random forests, support vector machines, and neural networks are employed to extract these features and classify profiles accordingly. Once trained, the model automates the classification process for new profiles. However, its effectiveness relies on factors such as dataset quality, feature selection, and algorithm choice, necessitating periodic updates to remain adept at identifying evolving techniques employed by fake profile creators.

I. METHODOLOGY TO IDENTIFY THE FAKE ACCOUNTS:

- Identifying fake accounts on social media has become a tedious issue for various sites such as Instagram, Facebook, Twitter, etc. It is generally done using several machine learning algorithms. Few methods which were previously used are not giving results up to the mark.
- Neural Network : The most common way a computer works is to assign commands or algorithms to a computer and create an output based result. But what if you do not know the problem-solving algorithm? Will your computer still be able to provide solutions? If we use standard techniques, then the computer will not be able to solve the problem unless you give specific instructions. Here comes the concept of Neural Networks. We can still solve such a problem by training the network as our system will learn by itself and will provide a solution that is close to some accuracy. The name Neural Networks was coined in 1943 but was not used at that time due to a lack of technology. Neural Networks learn by example. Neural Networks are based on organic matter i.e. brain cells and the way information is processed within the brain.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 06, June - 2024

- ANN :



- ANN is found in a group of connected units or nodes called artificial neurons, which form a model of neurons in the bloodstream freely. Each connection, such as synapses in the blood, can transmit signal to other neurons. The artificial neuron receives signal and processes it and can display neurons connected to it. The "signal" on the connection is a real number, and the output of each neuron is calculated by another function that does not match the input line. Connections are called genes. Neurons and edges usually have a corrective weight as the study progresses. Neurons can reduce the signal sent only if the combined signal exceeds that limit. Typically, neurons are divided into layers. Different layers can make different modifications to inputs. Symbols range from the first layer (input layer), to the last layer (output layer), probably after multiple layers.

➢ The active domain for the next project was Public Discovery. Public discovery is the key to understanding the structure of complex networks, and ultimately useful information is extracted from them. In this project, we have come up with a framework for which we can obtain a false profile using machine learning algorithms to ensure human health.

1. Separation begins in selecting the profile that needs to be separated.

2. Once the profile is selected, the useful features are removed for the purpose of partitioning.

3. Features are assigned to a well-constructed section.

4. The separator is trained regularly as new information can be added to the separator.Separate and determine whether the profile is true or false.

5. The result of the phase algorithm is then verified and the answer is sent back to the separator.

6. As the number of training details increases the detector becomes more accurate in predicting fake profiles.

BENEFITS:

- Social networking sites have improved our lives, but there are still many problems with social networking. Privacy issues, online exploitation, possible harassment, harassment, etc. This is mainly done with false profiles.
- For this project, we have come up with a framework for which we can obtain a false profile using machine learning algorithms to protect public health.

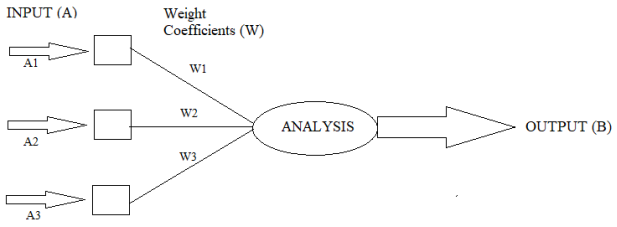
DETECTION STRATEGY:

➢ The data which we collected is used to derive the values for the parameters such as artificial activity and spam behavior. We considered the following terms for deciding if an account is fake or not:

1. Is profile picture present or not: A profile picture is the image which is displayed for an account in all its interactions across the platform. This thing plays a vital role in detection of fake account.
2. Num /Length of username: A username is a unique name assigned to each user. It can include characters as well as numbers. A valid username can be something which is close to user's actual name, and which includes minimum numerical values.
3. How many full name words are present: Name of account holder can include just first name or first name and surname, or it might include name of a brand/something else. Taking into consideration how many full name words are present helps in identifying if an account is fake or not.
4. Num/length of full name: This includes the length of account holder's name. If account name is absent, the chances of account being fake increases.
5. If account name and username are equal: Having account name and username same or similar gives more chances of account being real. If it doesn't match, it is more probable that the account is fake.
6. Length of account description: An account description or bio is a small summary found underneath the username. Number of words and characters present in an account's description can be considered while detecting fake/real account. Fake accounts generally have no bio/less bio.
7. Is any external URL present in bio: Fake accounts generally make use of external URL in their bio/description for promotional activities.
8. If the account is private or not: An Instagram account can be private or public. Posts shared by public account are visible to anyone on the platform, whereas the posts shared by private account are only visible to their respective followers.
9. Total number of posts of that account: An Instagram post is something which a user shares on the platform.
10. Total number of followers of an Instagram account: An Instagram Follower is a user who clicks on the follow button and follows your account. These users are able to give engagement on any of your posts.
11. Total number of following by an Instagram account: Following refers to all the users you follow on the platform. The data collected is used to get the values for

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 06, June - 2024

the above-mentioned parameters. Using these parameters different decision tress are formed. And hence after applying gradient boosting algorithm, we are able to determine if the account is fake or real.



MODULE DESCRIPTION

- Execute the code once all the libraries are installed The module has several features that make it perfect for Fake account detection. It has large data among them by using various modules it ensure quick detection with accurate with accurate results . Additionally, the module's fast processing speed ensures quick detection times.

- NUMPY AND PANDAS:
  Numpy and pandas are two essential tools for data science. Numpy is a powerful numerical library and pandas is a data analysis library. Our dataset contains some various accounts around the world. We will use pandas to load the data and NumPy for fake account detection..

- TENSORFLOW_HUB/TENSORFLOW:
  TensorFlow is a powerful open-source software library for data analysis and machine learning. We will be using the TensorFlow Hub library to load pre-trained models for fake account detection.

- MATPLOT LIB:
  Matplotlib is a powerful Python library used for creating static, animated, and interactive visualizations in Python. It is widely used for data visualization tasks, such as plotting graphs, histograms, scatter plots, and more. We will be using the matplot library to create plots for fake account detection

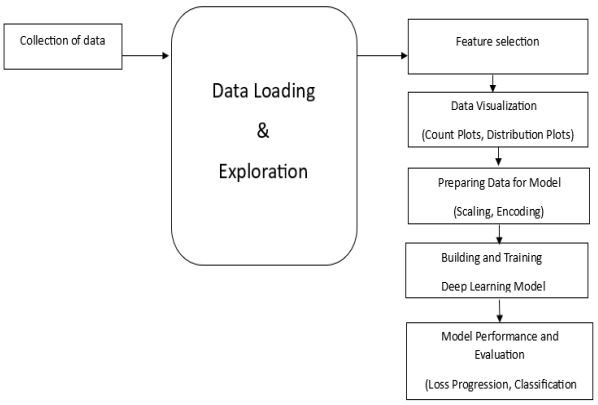
- SEABORN :
  Seaborn is a Python data visualization library based on Matplotlib that provides a high-level interface for drawing attractive and informative statistical graphics. We will be using Seaborn library is to primarily focused on creating visually appealing plots for data exploration and analysis, so it can also be useful for fake account detection tasks.

- KERAS:
  Keras is a high-level neural networks API written in Python that runs on top of TensorFlow, Theano, or Microsoft Cognitive Toolkit. It is designed to enable fast experimentation with deep neural networks and provides a user-friendly interface for building and training neural network models. And we are using keras to make the implementation of neural network easily
  Keras simplifies the process of building, training, and deploying deep learning models, making it a valuable tool for fake account detection tasks that involve analyzing large amounts of data and learning complex patterns.

- SKLEARN:
  Scikit-learn (sklearn) library using for machine learning tasks, including classification, regression, clustering, and dimensionality reduction, it provides a rich set of tools and algorithms that can be used to build and evaluate models for this purpose. We are using sklearn library so it can be useful in fake account detection and we are using it for to implement machine leraning models and statistical modelling and hence it provides a comprehensive toolkit for building and evaluating machine learning models, making it a valuable resource for fake account detection tasks.

WORK FLOW

Fake account detection on Instagram using a neural network. The process begins with the importation of essential libraries, including Pandas, Matplotlib, Seaborn, TensorFlow, and scikit-learn. Subsequently, training and testing datasets are loaded and subjected to exploratory data analysis (EDA) to understand the dataset's characteristics. Data preprocessing steps involve checking for missing values, standardizing numerical features, and converting labels into a categorical format. The neural network model is constructed with multiple hidden layers, employing ReLU activation functions and dropout layers to enhance generalization. Following model compilation, the data is split into training and testing sets, and feature scaling is applied. The model is then trained and evaluated, with performance metrics visualized, including loss progression during training and a confusion matrix. The entire workflow provides a structured approach to building and assessing a neural network for fake account detection, emphasizing the importance of data exploration, preprocessing, and rigorous evaluation.



Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 06, June - 2024

EXISTING WORK

Existing fake account detection systems employ a multifaceted approach involving several techniques:

1. Machine Learning: Both supervised and unsupervised learning techniques are used to analyze user behavior and profile characteristics, helping to distinguish between genuine and fake accounts.

2. Behavioral Analysis: This involves examining posting frequency, engagement patterns, and content v sharing behaviors to identify abnormal or bot-like activities. Patterns such as excessive posting or lack of interaction can be indicative of fake accounts.

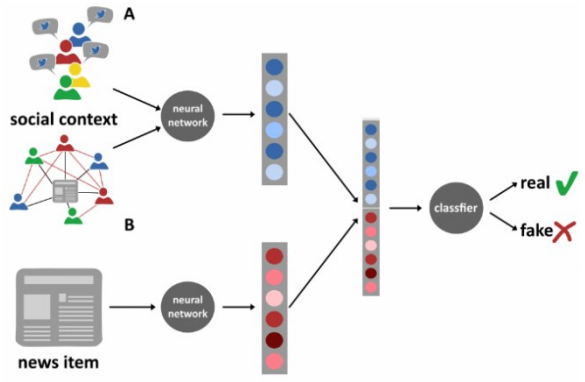
3. Profile Information Scrutiny: Analyzing the completeness, consistency, and uniqueness of profile details helps flag suspicious accounts. Incomplete profiles, mismatched information, or generic details often point to fake accounts.

4. Network Analysis: Studying connections and interactions between accounts helps detect patterns associated with fake profiles. Fake accounts often exhibit unusual connection patterns, such as being part of dense clusters with other fake accounts.

5. Image Analysis: Using reverse image search and assessing image characteristics helps identify instances of reused or AI-generated profile pictures. This step is crucial for detecting accounts that use stolen or artificially generated images.

PROPOSED WORK PRE-REQUISITES / TECHNOLOGIES USED

The project requires intermediate knowledge of Python programming and familiarity with libraries such as NumPy, Pandas, Matplotlib, Seaborn, Scikitlearn, and TensorFlow. Concepts from statistics, probability, and artificial neural networks (ANN) are also essential to understand and implement the machine learning and deep learning models



METHODOLOGY

The methodology consists of several key steps to ensure a comprehensive approach to fake account detection:

1.Data Loading and Exploration: The initial step involves importing the necessary libraries and loading the training and testing datasets from CSV files. Basic information about the data is displayed, and checks for null values are conducted. Data distributions and correlations are visualized using various graphs such as histograms, count plots, heatmaps, and distribution plots. This step helps understand the data structure and identify any anomalies or patterns.

2. Data Preprocessing: This step addresses issues related to null and missing values. Input features are standardized to ensure consistent scale across all variables, and label encoding is performed to convert categorical labels into numerical form. Normalizing the data ensures that the features are within a specific range, which is crucial for the effective training of the deep learning model.

3.Model Building and Training: A deep learning model is constructed using TensorFlow/Keras. The model comprises multiple dense layers with ReLU activation functions and dropout layers to prevent overfitting. The model is compiled with the Adam optimizer and categorical crossentropy loss. Training the model involves fitting it to the training data for a specified number of epochs with a validation split. Monitoring the training progress helps in fine-tuning the model parameters.

4.Performance Evaluation: The performance of the model is evaluated by visualizing the loss progression during training and validation. Predictions are made on the test set, and metrics such as the classification report and confusion matrix are generated to assess model performance. Thesmetrics provide insights into the accuracy, precision, recall, and overall effectiveness of the model in detecting fake accounts.

5.Model Refinement (Optional): An additional model with a different architecture may be defined and trained for comparison. This step allows for experimentation with different configurations to achieve the best possible performance.

6.Conclusion and Recommendations: The findings are summarized, and suggestions for improvement are provided. These include adding comments for better code readability, exploring different evaluation metrics, and considering additional features or techniques to enhance the model's accuracy and robustness.

ADVANTAGES OF PROPOSED WORK

The proposed work offers several advantages, including the utilization of advanced deep learning architectures such as transformers to improve detection accuracy and efficiency. Incorporating behavioral biometrics, such as typing patterns and device interaction behavior, provides enhanced authentication and detection capabilities. Blockchain integration helps secure user identities and transactions using decentralized and tamper-resistant technology. NLP integration empowers AI with language understanding, enabling tasks like sentiment analysis, speech recognition, chatbot interactions, and information extraction.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 06, June - 2024

CHALLENGES AND FUTURE DIRECTIONS

While the proposed system shows promise, there are several challenges that need to be addressed to ensure its effectiveness and reliability. One significant challenge is the availability and quality of labeled data. High-quality labeled datasets are crucial for training accurate models, but obtaining such data can be difficult due to privacy concerns and the dynamic nature of social media. Additionally, fake account detection models must be continuously updated to adapt to evolving tactics used by malicious actors. This requires ongoing monitoring and maintenance, which can be resource-intensive. Another challenge is ensuring the model's robustness and generalizability across different social media platforms. While this project focuses on Instagram, fake account detection techniques need to be adaptable to other platforms with varying user behaviors and account characteristics. This requires a flexible approach to feature extraction and model training.

FUTURE DIRECTIONS

To address these challenges and further improve fake account detection systems, future work could explore the following directions: Data Augmentation and Synthetic Data: To mitigate the issue of limited labeled data, techniques such as data augmentation and the generation of synthetic data can be employed. These methods can expand the training dataset and improve model robustness.

TRANSFER LEARNING

Leveraging transfer learning techniques can help apply knowledge gained from detecting fake accounts on one platform to other platforms. This can enhance the model's adaptability and reduce the need for extensive retraining.

REAL-TIME DETECTION SYSTEMS

Developing real-time detection systems can help identify and mitigate the impact of fake accounts more quickly. This involves optimizing the model for low-latency inference and integrating it with social media platforms for continuous monitoring.

USER EDUCATION AND REPORTING MECHANISMS:

Educating users about the signs of fake accounts and improving reporting mechanisms can complement automated detection systems. User reports can provide valuable data for model training and refinement.

INTERDISCIPLINARY APPROACHES

Combining insights from fields such as cybersecurity, psychology, and sociology can enhance the understanding of fake account behaviors and inform the development of more effective detection techniques.

CONCLUSION

The proposed system provides an end-to-end pipeline for detecting fake Instagram accounts using deep learning. The system includes comprehensive steps such as data loading, exploration, preprocessing, model building, training, and evaluation. Suggestions for improvement include adding comments for better code readability, exploring different evaluation metrics, incorporating advanced deep learning architectures, and investigating behavioral biometrics and blockchain integration for enhanced security. The integration of cutting-edge technologies such as deep learning, behavioral biometrics, blockchain, and NLP reflects a forward-looking approach to enhance xi accuracy, security, and user experience. By addressing the challenges posed by fake accounts on social media platforms, the proposed system aims to contribute to a safer and more trustworthy social media environment. The project underscores the importance of leveraging advanced technologies to maintain the integrity of online communities and ensure a reliable digital ecosystem.

Spark MLlib, deep on it. The agent will receive a warning SMS automatically if the detected value exceeds the threshold.Our plan to develop a high- frequency, high-mobility, low- power water monitoring system makes it special.

Kartik Maheshwari & Adrija Chakraborty, [6] were reported that, the proposed system was successfully implemented to determine the turbidity, TDS, flow rate and the level of water for a given sample. The data obtained from the sensors are uploaded to the Thing Speak dashboard for online monitoring purpose. Besides, an SMS alert is sent to the user whenever the turbidity and TDS values have crossed the threshold limit defined for good quality water. As per the obtained results, the proposed system can be considered as suitable water quality monitoring system. Our system has made water quality testing more economical, convenient, and reliable with timely feedback.

REFERENCES:

[1] F. C. D. Da Silva, A. C. B. Garcia, and S. W. M. Siqueira, “A Systematic Literature Mapping on Profile Trustworthiness in Fake News Spread,” in 2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design, CSCWD 2022, 2022, pp. 275– 279. doi: 10.1109/CSCWD54268.2022.9776232.

[2] R. Singh Boparai and R. Bhatia, “Detectionof Fake Profiles in Online Social Networks-A Survey.” [Online]. Available: https://ssrn.com/abstract=4159087

[3] O. Kadam and N. Surse, “Detection of Fake Social Network Account,” vol. 6, pp. 2456–0774, 2021, doi: 10.51319/2456-0774.2021.4.0013.

[4] Sarah Khaled, Neamat El-Tazi, and Hoda MO Mokhtar. "Detecting fake accounts on social media." In 2018 IEEE international conference on big data (big data), pp. 3672-3681. IEEE, 2018.

[5] Kumud Patel, Sudhanshu Agrahari, and Saijshree Srivastava. "Survey on fake profile detection on social sites by using machine learning algorithm." In 2020 8th international conference on reliability, infocom technologies and optimization (trends and future directions)(ICRITO), pp. 1236-1240. IEEE, 2020

[6] Preethi Harris, J. Gojal, R. Chitra, and S. Anithra. "Fake Instagram Profile Identification and Classification using Machine Learning."

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 06, June - 2024

In 2021 2nd Global Conference for Advancement in Technology (GCAT), pp. 1-5. IEEE, 2021

[7] Adikari, S., Dutta, K., 2014. Identifying Fake Profiles in Linkedin, in: PACIS 2014 Proceedings. Presented at the Pacific Asia Conference on Information Systems.

[8]. Maclean, J. Melton, J. Melton, A. R. Simon, and M. Chisholm, Data Mining : Concepts and Techniques. 1999. [9]. S. Kiruthiga, “Detecting Cloning Attack in Social Networks Using Classification and Clustering Techniques,” IEEE, 2014. [10]. Harini .N., Smruthi.M.,"A Hybrid Scheme for Detecting FakeAccounts in Facebook" ISSN: 2277- 3878, (IJRTE)International Journal of Recent Technology and Engineering (2019) , Issue-5S3, Volume.