A Project Report

On

Fake Social Media Profile Detection and Reporting

carried out as part of the course CSE CS3270

Submitted by

Sarthak Pundir (Registration No: 219310190)

Harshit Kalra (Registration No: 219301295)

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

In

Computer Science & Engineering

Department of Computer Science & Engineering,School of Computer Science and Engineering, Manipal University Jaipur,

March2024

Acknowledgement

This project would not have been completed without the help, support, comments, advice, cooperation, and coordination of various people. However, it is impossible to thank everyone individually; I am hereby making a humble effort to thank some of them.

I acknowledge and express my deepest sense of gratitude to my internal supervisor Dr. Aditya Sinha for his/her constant support, guidance, and continuous engagement. I highly appreciate his technical comments, suggestions, and criticism during the progress of this project Fake Social Media Profile Detection and Reporting.

I owe my profound gratitude to Dr. Neha Chaudhary, Head, Department of CSE, for her valuable guidance and for facilitating me during my work. I am also very grateful to all the faculty members and staff for their precious support and cooperation during the development of this project.

Finally, I extend my heartfelt appreciation to my classmates for their help and encouragement.

Student Signature

Student Name :
Registration No. :

MANIPAL UNIVERSITY
JAIPUR
(University under Section 2(f) of the UGC Act)

Department of Computer Science and Engineering
School of Computer Science and Engineering

Date: _____________

CERTIFICATE

This is to certify that the project entitled "**Fake Social Media Profile Detection And Reporting**" is a bonafide work carried out as Minor Project Midterm Assessment (Course Code: CS3270) in partial fulfillment for the award of the degree of Bachelor of Technology in Computer Science and Engineering, by **Sarthak Pundir and Harshit Kalra** bearing registration number **219310190 and 219301295**, during the academic semester VI of year 2023-2024.

Signature of the project guide:

Name of the project guide:

Place: Manipal University Jaipur, Jaipur

Contents

Abstract

Platforms such as Twitter, Facebook, Instagram, and LinkedIn are examples of social media, play a crucial role in our day-to-day existence. People worldwide are participating and are actively involved in it. However, it also encounters the issue of counterfeit. Fake profiles are typically created by either humans or bots. Cyborgs are designed to spread rumours, conduct phishing attacks, and breach data security. Stealing someone's personal information for personal gain is known as identity theft. Hence, in this article, we examine a detection model. Which distinguishes fake profiles from real profiles on Twitter, judging by the apparent factors such as the number of followers, friends, and status count by utilizing different machine learning techniques. The Twitter profile dataset includes TFP and E13 for real profiles, as well as INT, FSF and TWT for fake profiles. We are discussing Neural Networks, specifically Random Forest, XG Boost, and Long Short-Term Memory (LSTM). Key characteristics have been chosen to ascertain the legitimacy of a social media account. Additionally, the discussions about architecture and hyperparameters take place. Ultimately, the models are trained, and outcomes are achieved. The outcome we receive is a value of 0 for genuine profiles, while fake profiles receive a score of 1. Once a profile is identified as fake, it can be blocked or removed, preventing cybersecurity threats. Python3 is utilized for implementation along with all necessary libraries, similar to NumPy, Sklearn, and Pandas.

Keywords: Neural Network, Random Forest, XG Boost, social media, Fake profile

Fake Social Media Profile Detection

1.Introduction

Social media has become an essential component of our daily existence. From distributing appealing extravaganzas. Agent takes photos to track celebrities, and to socialize with friends near and far. Everyone is engaged in using social media. This is an excellent platform for sharing information and engage with individuals. However, there is a negative aspect to everything. With the rise of social media, it has become difficult finding a stable place in our lives, there are times when it has proven to pose an issue.

There are 330 million users who are active every month, and 145 million users who are active every day on the Twitter platform. Facebook gains approximately 500,001 new users per day and 6 additional Users at a rate of one per second. A huge amount of data is posted on Twitter daily. From popular trending subjects to the newest hashtags and updates to someone's latest information.

During your journey, you can find all the latest updates on Twitter. Individuals respond by either liking, commenting, or sharing their thoughts, expressing their viewpoints within the 280-character restriction. There are authentic topics that are talked about, but occasionally there are gossip. These rumours cause confusion, most of the time leading to disputes among various segments of the population. The worry about privacy, abuse.

In the recent past, cyberbullying [4] and spreading false information has been a growing concern. Fake profiles are responsible for carrying out these tasks. Artificial accounts may be created by people, machines, or a combination of both [2]. Cyborgs are profiles first made by people but then controlled by machines. False accounts are typically made using fake names and containing deceptive and harmful content. These profiles distribute posts and images in order to influence society which lead us to dealing with the issue of fraudulent profiles in the current times.

The main reasons for making fake profiles are to spam, phish, and obtain personal information. Gaining a larger number of followers. The fraudulent accounts have the capability to engage in cybercrime, criminal activities. Counterfeit accounts pose significant risks such as identity theft and fraud. Fake accounts are responsible for data breaches by sending different URLs to individuals. When users visit, their data is sent to remote servers where it could potentially be used against them. Additionally, the fraudulent profiles, seemingly generated on behalf of different organizations. For individuals, it can harm their image and result in a lower amount of likes and followers. In addition to these challenges, social media manipulation poses a hurdle. The counterfeit accounts contribute to the dissemination of false and inappropriate information, which lead to disputes.

These fraudulent accounts are made with the purpose of gaining more followers as well. Who would not desire this? How to stay trendy on social media? In order to gain a large number of followers, individuals must strive to increase their social media presence. In general, it has been noted that fake profiles lead to greater consequences, causing more damage than a different cyber-crime. Therefore, it is crucial to identify an artificial account.

In this particular situation, we are discussing the identification of fraudulent accounts on Twitter utilize different machine learning algorithms. The data from Twitter accounts E13 and TFP is used for authentic accounts, while INT, TWT, and FSF are utilized for counterfeit accounts. When it comes to dealing with the creation of fake profiles, typical defences include:

(A) Incorporating methods like user verification is essential when setting up accounts on social media platforms.

(B) User behaviour analysis is necessary in order to identify unusual activities. The bot detection solution, which utilizes real-time AI analysis, will be advantageous.

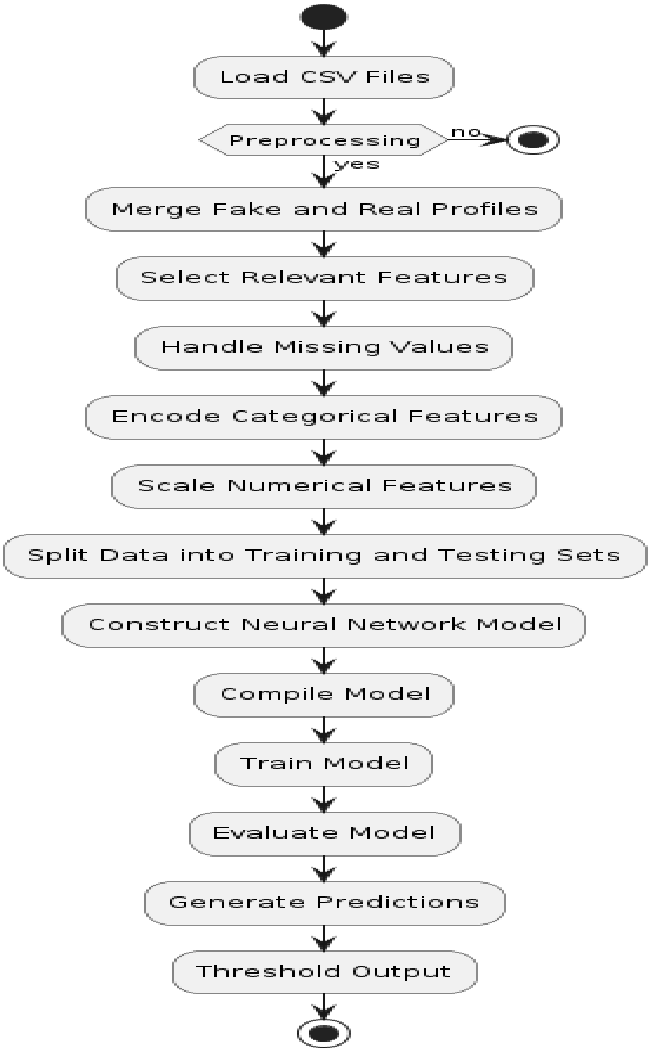
(C) It is essential to utilize an automated tool for protecting against bots.

Our technical contribution involves the creation of a mul ti-layer neural-network model, random-forest model, a model of XG-boost, and a model of LSTM. The aforementioned models are models of supervised machine learning.

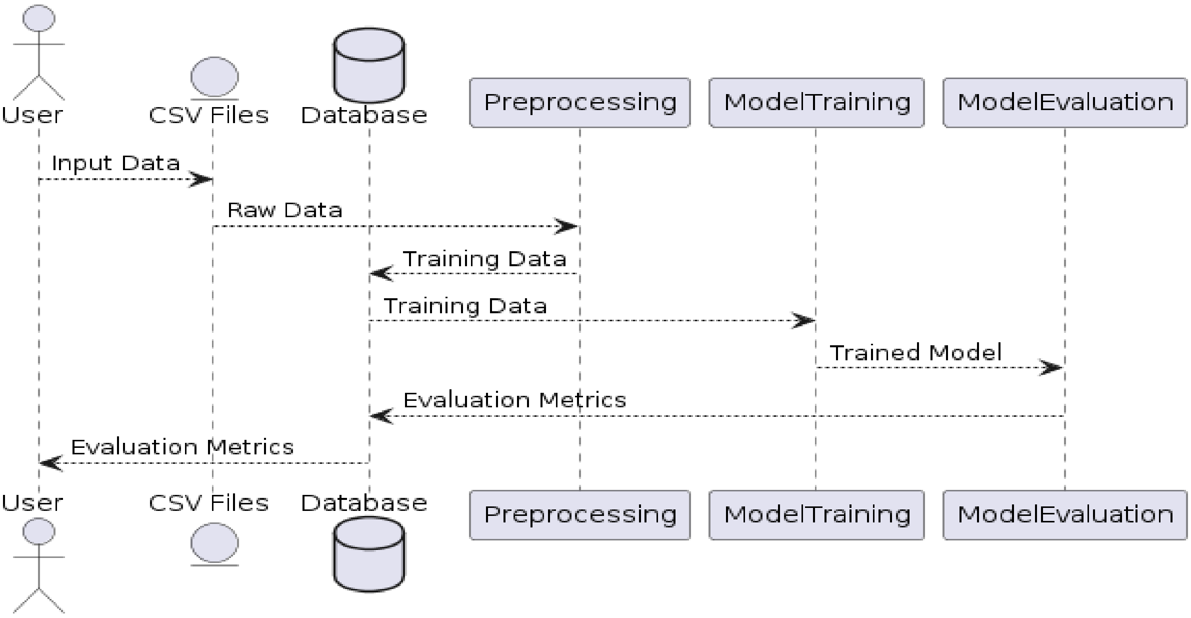
In addition, the LSTM makes classifications using tweets, and the outcome can be integrated with a convolutional neural network in the upcoming future [6]. The paper is divided into segments. The previous studies, data preparation, approach, findings from experiments, precision of the models, summary, and future work are described in order.

2 Design Description

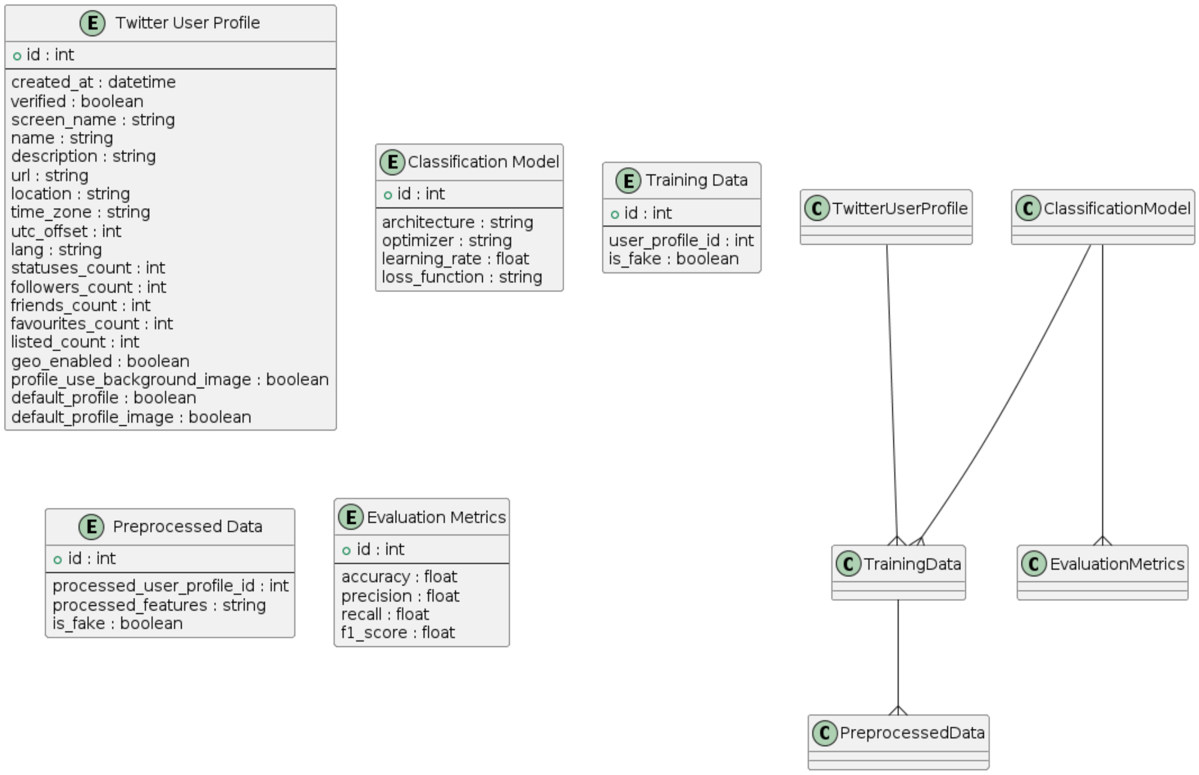
2.1 Flow Chart



2.2 Data Flow Diagrams (DFDs)



2.3 Entity Relationship Diagram (E-R Diagram)



3 Review of prior research

Social-media: A blessing or curse, this query has always endured. All-companies have focused on creating a pla tform that has minimal errors and offers a superior user experience. Every day, new advancements and revisions are made. Not enough efforts have been made to detect fake human-identities on social media platforms like Twitter, so we decided to review previous research on similar issues.

Certain techniques ca tegorized profiles according to the account's-activity, including the quantity of responses to requests, sent messages, and other factors. The models utilize a system that is based on graphs. Certain techniques also focused on distinguishing between bots and cyborgs. Below are mentioned some previous studies. If specific words are present in a message, it will be categorized as spam. This idea has been employed to identify fraudulent accounts on social networking sites. Pattern matching techniques were utilized for identifying those words on social media. However, the main disadvantage of this ru le is that as time goes on, there is a constant growth and adoption of new words. Additionally, the popularity of acronym such as lol, gbu, and gn is increasing on Twitter. Sybil-Guard [13] was created in 2008 to reduce the negative impact of Sybil attacks on social media networks. The random walk was limited by each node and relied on random walk interactions happening. Kleinberg's artificial social network dataset was utilized.

Another method known as the Sybil limit was created at the same time as the Sybil guard approach. Similar to Sybil-guard, it also operates under the assumption that the non-Sybil area is highly blended. The method involved each node performing several random walks. The ranking depended on how often walk intersections had tails. Sybil-infer was created in the year 2009. It utilized techniques such as greedy-algorithm, Bay esian inference, and Monte-Carlo sampling assuming that the ‘non Sybil’ region and random walks are fast mixing. The selection method is based on probability thresholds. Mislove's 201- algorithm used greedy search on the Facebook dataset to choose profiles based on normalized conductance metric.

A new Facebook immune system model was introduced in 2011, utilizing random forest, SVM, and boosting methods. The Facebook dataset was also utilized, with the selection technique being the feature loop. Facebook uses an algorithm to identify bots by analyzing the amount of friends they have, which may be linked to tagging or relationship history. The above regulations can detect bot accounts but are unsuccessful in detecting fake-accounts created by humans. Unsupervised machine learning was employed to identify bots. Instead of assigning labels, data was gathered according to proximity in this approach. The bots were identified through their exceptional grouping functions thanks to shared characteristics. Sybil rank [1][13] was created in 2012 and relies on a system that is graph-based. The rankings of the profiles were determined by their interactions, tags, and wall posts. The profiles designated as real have a high rank, while the fake ones have a lower rank. However, this approach was not consistently accurate due to situations where a genuine profile received a low ranking.

Then, a different model was created known as the Sybil frame. It employed a classification system with multiple stages. It operated in a two-fold process, first utilizing a content-focused strategy and then implementing a structure based technique. Filtering is also considered one of the previous methods. A recent harmful action is identified, leading to the account being included in the blacklist. When it comes to fake accounts created by humans, they have a tendency to adjust in order to avoid being blacklisted. Studies were also conducted to identify fraudulent accounts using criteria such as engagement-rate and ‘artificial activity’. An engagement rate refers to the proportion of audience interaction with a post. The formula for calculating the engagement rate is multiplying the total number of interactions by 100 and then dividing by the total number of followers. These engagements may take the form of likes, shares, or comments. Artificial behavior relies on the quantity of ‘shares, likes, and comments’ generated by a specific account. Lack of information and the verification status of email are also seen as artificial activity. In our study, we utilized a multi-layered neural network, the random forest [9] method, and XG Boost to analyze the observable traits of a profile. The collected characteristics are saved in a CSV file for easy interpretation by the model. After completing training, testing, and evaluating, the model is able to determine whether a profile is legi timate or not. We utilized Google Col ab to train our mo dels due to the availability of free GPU provided by Google. The NVIDIA Tesla K80 GPU with 12GB in Google Colab can be utilized for a maximum of 12 consecutive hours. All the models were programmed in Python3.

4.Methodology

We integrated numerous supervised methods to identify fake Twitter accounts. Techniques, all aiming for the same outcome but varying in levels of precision. Every model is able to identify fake account developed solely using visual characteristics.

Each of these supervised models is trained on the identical dataset, and their accuracy is compared accordingly. Charts showing profitability and loss are created. Additionally, there is a graph comparing the accuracy of. There are signs of various models being suggested. The optimization is applied appropriately when training the models. Different components of a neural network include layers, functions that measure the loss, and functions that introduce non-linearity. The aforementioned models that were utilized.

4.1 Pre-processing

Before moving forward with the models, we add an additional step known as pre-processing. The dataset undergoes pre-processing prior to being inputted into a model. Our model is designed to focus on Identifying whether a profile is fake or real by analysing its visible attributes. From this point forward, all specific details are established. The only type of information is numerical data and the non-numeric attributes are eliminated. The listed characteristics are selected [10]:

friends	Followers	Status_count	Listed_count	Geo_enabled	Fav_count	Lang_num

Then, the dataset containing fake and real users is combined into a single dataset. An extra tag for every profile called "isFake" in the form of a Boolean variable. It is at that moment. The Y variable stores the response related to profile X. In conclusion, the empty cells or "Not a Number" values are replaced with zeros.

4.2 Artificial Neural Network

Neural networks [8] in the human brain, mimicking their ability to learn patterns and make decisions based on data. Complex network within a human brain consists of interconnected nodes that process input data contains neurons (nodes). We employed the Keras sequential function. The design of the model features activation function is present in an input layer, three hidden layers, and an output layer. Apply 'ReLU activation function' to all layers except for the final output layer. The activation function used is Sigmoid, the final layer of output. The model was built with optimizer: Adam and loss function: binary_crossentropy. In our model, we utilize an artificial neural network with the architecture described above. Sigmoid function ultimately yields an output ranging from 0 to 1, depending on the forecasting a specific profile, distinguished as either counterfeit or authentic.

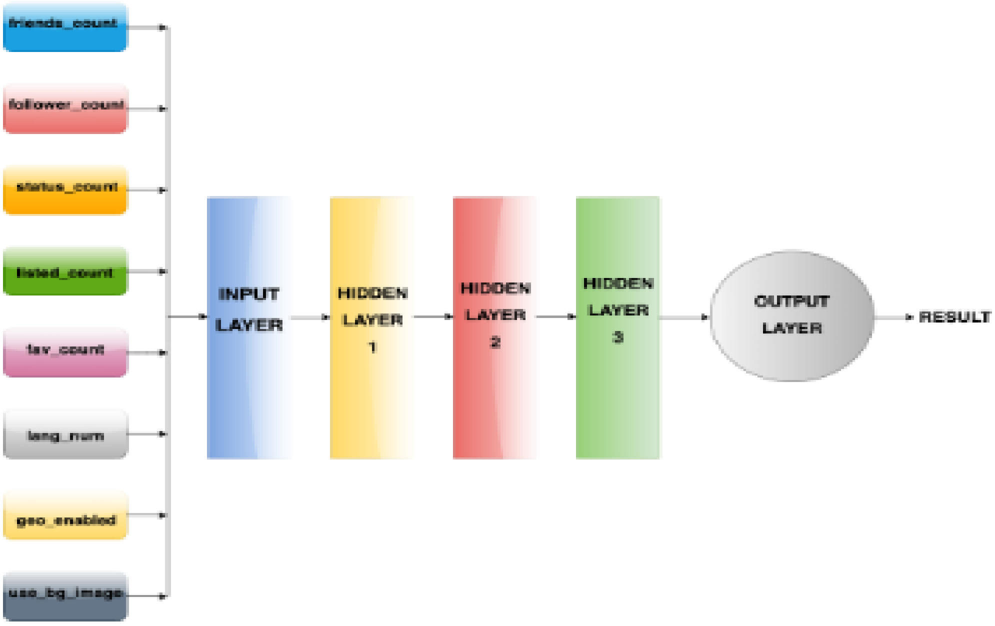


Fig. 1. ANN Architecture

Hyper parameters

– Rectified Linear Units (ReLU): Rectified-Linear activation function is a function that is linear in different segments. ReLU (Fig. 2) is commonly used as the activation function in neural networks due to its ease of training and superior results.

$$R(z) = \max(0, z) \tag{1}$$

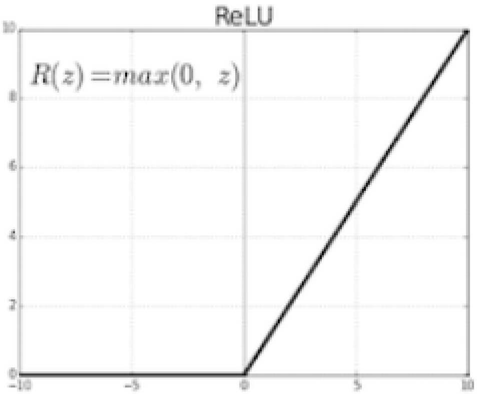


Fig. 2. Rectified Liner Units.

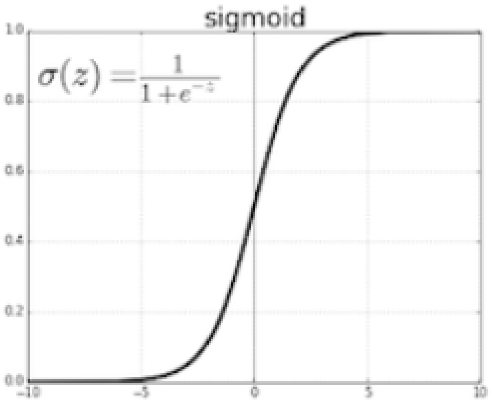


Fig. 3. Sigmoid Function.

– Sigmoid Function: The logistic function is also referred to as the Sigmoid Function. The sigmoid function (Fig. 3) is utilized when values within the range of 0.0 to 1.0 are needed. It is an activation function that is not linear and can be differentiated to find the slope at any two points.

$$\sigma(z) = \frac{1}{1+e^{-z}} \qquad (2)$$

- If z is very large then e^{z} is close to zero and

  $$\sigma(z) = \frac{1}{1+0} \approx 1 \qquad (3)$$

- If z is very small then e^{z} is large then

  $$\sigma(z) = \frac{q}{1+large\ number} \approx 0 \qquad (4)$$

4.3 Random Forest

Random forest, or random-decision-forest, is a technique that belong to the group of ensemble learning techniques. This technique is utilized in machine learning is highly effective in solving regression problems because of its simplicity and efficiency. Random forest, in contrast to the decision tree approach, produces several decision trees. The trees combine to form the final output, which is the culmination of their collective decisions. In the same manner, we utilized the random forest [9] method for profile detection. Information is provided to the model, resulting in outputs being generated. During the time when the training is happening, the bootstrap aggregating algorithm is used for the provided set of X = x_1, sample randomly B times responses x_2 to x_n and Y = y_1 to y_n, is chosen and matched the trees(fb) with the sample. Once the training is complete, the predictions can be made. The formula specified below is used to calculate a specific sample (x').

$$\hat{f} = \frac{1}{B} \sum_{b=1}^{B} f_b\,(x') \qquad (5)$$

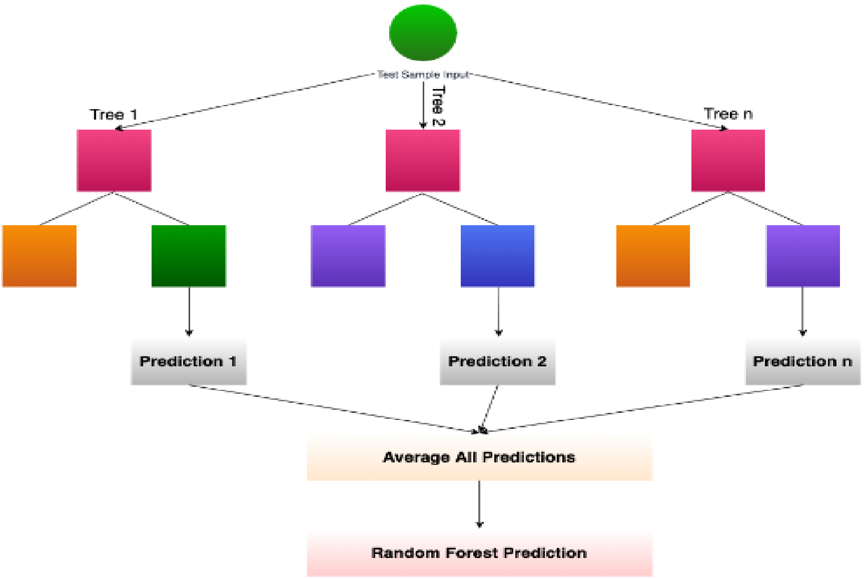


Fig. 4. Random Forest Architecture

4.4 Extreme Gradient Boost

XG Boost is an additional ensemble learning approach employed for regression tasks. It enhances the stochastic gradient boosting technique. Random forest performs best when it has access to all inputs, no value is absent. To surmount this issue, we employ a gradient boosting algorithm. According to the boosting algorithm, the first step is to initialize F_0(x).

$$F_0(x) = argmin_{\gamma \sum_{i=1}^{n} L(y_i,\gamma)} \tag{6}$$

Next, the gradient of the loss function is calculated iteratively.

$$r_{im} = -\alpha[\frac{\partial L(y_i,F(x_i))}{(F(x_i))}] \tag{7}$$

The model Fm(x) that has been enhanced is ultimately specified.

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \tag{8}$$

α is the learning rate

γ_m is the multiplicative f actor

5 Experimental Results

5.1 Dataset

The dataset that we utilized was obtained from MIB [17]. There were a total of 6825 profiles in the data set, with 3474 being real and 3351 being fake. The chosen data included 'TFP and E13' for real accounts and 'INT, TWT and FSF' for fraudulent accounts. The information is saved in a CSV file format that the machine can easily read. All the markers on the x-axis represent the characteristics utilized in identifying the counterfeit profile. These were chosen in the initial processing stage. The number of entries that each feature in the dataset has is represented on the y-axis.

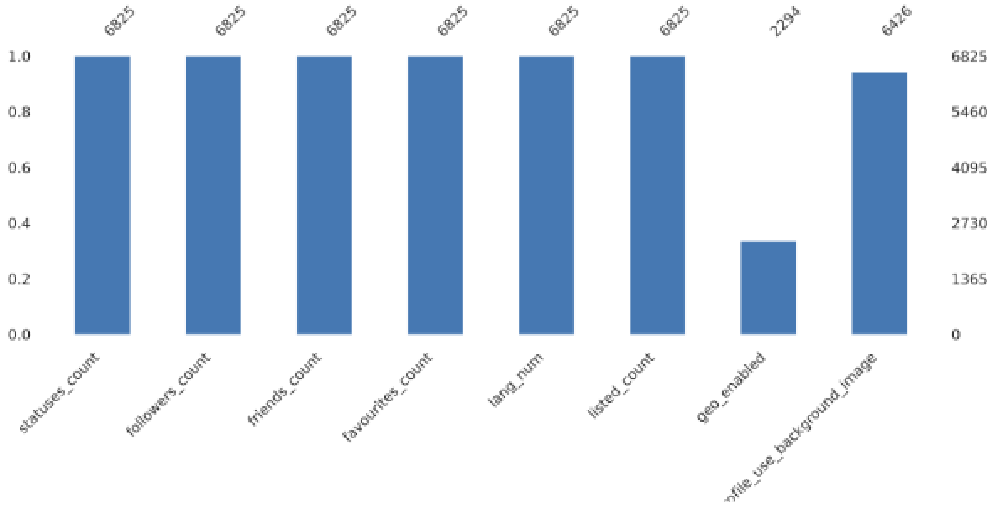
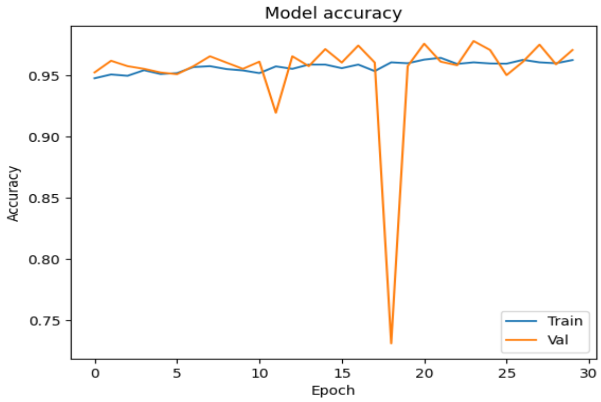


Fig. 5. Data set

5.2 Graphs and Charts

After completing ‘training and testing’ on all of the models, the subsequent results were achieved. Graphs showing the accuracy and loss of the model over the epochs are displayed for the neural network. comparison of mo del accuracy and ‘ROC curve’ for random and LSTM networks, XG boost and alternative techniques.

Neural Net work: The ‘model-accuracy-graph’ and ‘model-loss-graph’ for the trained neural network are as follows:



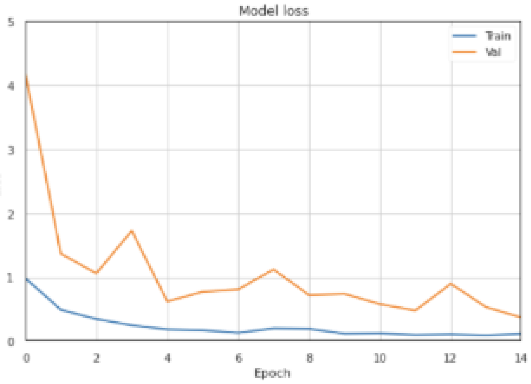


Fig. 6. Model Accuracy

Fig. 7. Model loss

After completing 15 iterations, the graphs showing accuracy and loss are shown above. At first, the accuracy fluctuates as it progresses from 0.97 and eventually peaks at 0.98. Likewise, the testing data's loss graph starts at 1, while the validation data's loss graph starts at 4 and eventually reaches a mini-mum_point below 0.5. The ‘Binary cross-entropy’ function is utilized for determining the loss. At first, each feature is assigned random weights before the machine eventually assigns a unique weight to each feature.

Random Forest as well as alternative techniques: In the upcoming comparison chart, we will-ass ess the effectiveness of different models like random forest, xg boost, and ada boost. and tree that is utilized in the decision-making process. XG boost achieves its peak accuracy level at 0.996. Moreover, we possess a decision tree and a random forest that both have nearly the same accuracy of 0.99. At last, ADA boost is now accessible. The accuracy comparison histogram and the ROC curves are shown below.

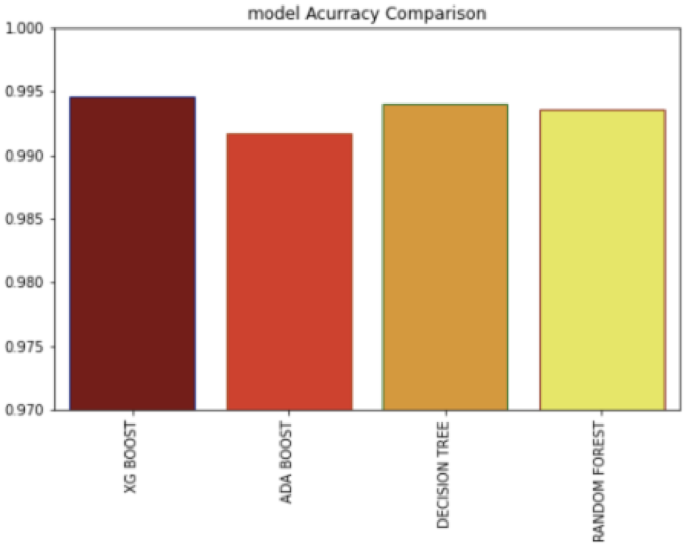


Fig. 8. Accuracy of Different Models

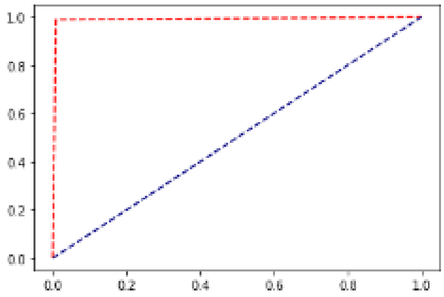


Fig. 9. ROC curve XG Boost

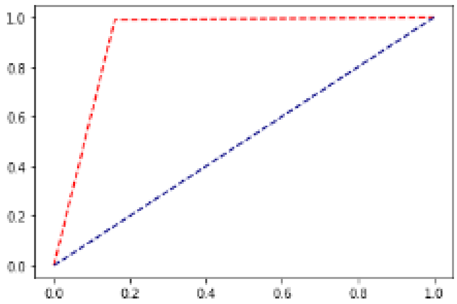


Fig. 10. ROC cruve Random Forest

6 Conclusion

In-this model, we incorporated the Neural-Network, Random-Forest, and XGBoost. Enhance machine-learning techniques for training ‘our system’ to identify fake-Twitter profiles, documents created from data that can be seen. Once we trained, validated, and tested our models on. After analysing the MIB data set, we ultimately deduce that the highest level of accuracy is achieved by XG Boost method reached a success rate of 99.46%, with ANN and random forest closely behind. More progress can be made by merging images of side views with the cat-egorical. Use both categorical and numerical data and apply it with a CNN. In addition, incorporating additional items as well. We have combine parameters from various models, and build a live model to

attain improved outcomes.

7 Acknowledgment

We offer our heartfelt thanks to Dr. Aditya Sinha for his support, valuable feedback, and patience throughout this project. His fervent support motivated us to stay on track with our progress.

8 References

1. Gergo Hajdu, Yaclaudes Minoso, Rafael Lopez, Miguel Acosta, Abdelrahman Elleithy: Use of Artificial Neural Networks to Identify Fake Profiles.

2. Est´ee Van Der Wal: Using Machine Learning to Detect Fake Identities: Bots vs Humans.

3. Aleksei Romanov, Alexander Semenov, Oleksiy Mazhelis and Jari Veijalainen: Detection of Fake Profiles in Social Media.

4. Yasyn ELYUSUFI, Zakaria ELYUSUFI, M'hamed Ait KBIR:Social Networks Fake Profiles Detection Based on Account Setting and Activity

5. Stefano Cresci, Roberto Di Pietro, Marinella Petrocchi, Angelo Spognardi, Maurizio Tesconi: Fame for sale: Efficient detection of fake Twitter followers.

6. Sneha Kudugunta, Emilio Ferrara:Deep Neural Networks for Bot Detection

7. Rohit Raturi: Machine Learning Implementation for Identifying Fake Accounts in Social Network August International Journal of Pure and Applied Mathematics (2018)

8. M. Likitha, K. Rahul, A. Prudhvi Sai, A.Mallikarjuna Reddy: Design and Development of Artificial Neural Networks to Identify Fake Profiles.

9. https://www.irjet.net/archives/V6/i12/IRJET-V6I12189.pdf

10. Yasyn Elyusufi, Zakaria Elyusufi, A¨ıt Kbir M'hamed: SocialNetworksFakeProfilesDetectionUsingMachineLearningAlgorithms https://www.researchgate.net/publication/339012245/

11. Naman Singh, Tushar Sharma, Abha Thakral, Tanupriya Choudhury: Detection of Fake Profile in Online Social Networks Using Machine Learning in International Conference on Advances in Computing and Communication Engineering.(2018)

12. Jari Veijalainen, Aleksei Romanov, and Alexander Semenov:Revealing Fake Profiles in Social Networks by Longitudinal Data Analysis.

13. Devakunchari Ramalingam , Valliyammai Chinnaiah:Fake profile detection techniques in large-scale online social networks: A comprehensive review.

14. Kharaji MY, Rizi FS: An IAC approach for detecting profile cloning in online social networks. (2014)[1]

15. Yu H, Gibbons PB, Kaminsky M, Xiao F. Sybillimit: A near-optimal social network defense against sybil attacks. In: IEEE symposium in security and privacy, (2008)

16. Fire M, Goldschmidt R, Elovici Y: Online social networks: threats and solutions. IEEE Commun Surv Tut9 (2014)

17. Sarah Khaled, Hoda M. O. Mokhtar, Neamat El-Tazi

18. Mulamba D, Ray I, Ray I. SybilRadar: A graph-structure based framework for sybil detection in on-line social networks. In: Proceedings of IFIP in- ternational information security and privacy conference.(2016)

19. Conti M, Poovendran R, Secchiero M.: Fakebook- Detecting fake profiles in online social networks. In: Proceedings of the international conference on advances in social networks analysis and mining, (2012)

20. Wang G, Jiang W, Wu J, Xiong Z: Fine-grained feature-based social influence evaluation in online social networks. IEEE Trans Parallel Distrib Syst (2014)

21. Silvia Mitter, Claudia Wagner, and Markus Strohmaier: A categorization scheme for socialbot attacks in online social networks. In Proc. of the 3rd ACM Web Science Conference (2013).

22. Norah Abokhodair, Daisy Yoo, and David W McDonald. Dissecting a social botnet: Growth, content and influence in Twitter. In Proc. of the 18th ACM Conf. on Computer Supported Cooperative Work Social Computing (2015).

9 Course Outcome

Title: "Detection of Fake Twitter User Profiles using Machine Learning"

Patent Pending

This course outcome document outlines the intellectual property rights associated with the project titled "Detection of Fake Twitter User Profiles using Machine Learning." The copyright protects the original work created by the author, prohibiting unauthorized reproduction, distribution, or transmission of the content without prior written permission.

Additionally, the project is currently pending patent approval, which seeks to protect the novel methods and technologies developed as part of the project. The patent application aims to safeguard the project's innovative aspects from unauthorized use or replication by others.