Scientific Research Publishing

Fake Profile Detection Using Machine Learning Techniques

Partha Chakraborty*, **Mahim Musharof Shazan, Mahamudul Nahid, Md. Kaysar Ahmed, Prince Chandra Talukder**

Department of Computer Science & Engineering, Comilla University, Cumilla, Bangladesh
Email: *partha.chak@cou.ac.bd, mm.shazan586@gmail.com, nahidmahamudul056@gmail.com, kaysarahmed@stud.cou.ac.bd, prince.cse.bd8@gmail.com

Abstract

Our lives are significantly impacted by social media platforms such as Facebook, Twitter, Instagram, LinkedIn, and others. People are actively participating in it the world over. However, it also has to deal with the issue of bogus profiles. False accounts are frequently created by humans, bots, or computers. They are used to disseminate rumors and engage in illicit activities like identity theft and phishing. So, in this project, the author'll talk about a detection model that uses a variety of machine learning techniques to distinguish between fake and real Twitter profiles based on attributes like follower and friend counts, status updates, and more. The author used the dataset of Twitter profiles, separating real accounts into TFP and E13 and false accounts into INT, TWT, and FSF. Here, the author discusses LSTM, XG Boost, Random Forest, and Neural Networks. The key characteristics are chosen to assess a social media profile's authenticity. Hyperparameters and the architecture are also covered. Finally, results are produced after training the models. The output is therefore 0 for genuine profiles and 1 for false profiles. When a phony profile is discovered, it can be disabled or destroyed so that cyber security problems can be prevented. Python and the necessary libraries, such as Sklearn, Numpy, and Pandas, are used for implementation. At the end of this study, the author will come to the conclusion that XG Boost is the best machine learning technique for finding fake profiles.

Keywords

Social Media, Fake Profiles, Random Forest, Neural Networks, XG Boost

1. Introduction

Social media plays a significant role in our lives today. Our lives nowadays rely

heavily on social media. Everyone uses social media, whether it be to share beautiful, expensive photos, follow celebrities, or talk with nearby and distant pals. It is a fantastic place for exchanging knowledge and interacting with others. However, everything has a drawback. Social media has a significant role in our lives, yet there have been times when it has become problematic.

There are 229 million daily active members of Twitter and 465.1 million monthly users. Furthermore, Facebook creates six new users per second, for a daily average of about 500,000 new users. Every day, a huge amount of information is posted on Twitter. On Twitter, one can access the most popular articles, the latest hashtag, news, and information on their most recent trip. Within the allotted 280 characters, people can reply, like, remark, exchange ideas, and express their viewpoints. There are often rumors, but there are also significant worries that are investigated. The various socioeconomic groupings get tense as a result of these rumors. Concerns around privacy, exploitation, cyberbullying, and false information have recently come to light. All of these activities involve the use of fake profiles. Humans, machines, and cybernetic beings may all create false accounts [1]. "Cyborg" accounts were once established by individuals but are now managed by machines.

False profiles are frequently made under fictitious identities, and they spread defamatory and abusive posts and images to influence society or advance anti-vaccine conspiracy theories, among other things. Phony personas are an issue on all social media platforms nowadays.

Most false profiles are made with spamming, phishing, and gaining more followers in mind. The fraudulent accounts are completely capable of committing online crimes. Fake accounts represent a serious risk, including identity theft and data breaches. When consumers access the URLs sent by these false accounts, all user information is sent to distant servers where it may be used against them. Furthermore, phony profiles purportedly created on behalf of businesses or individuals can damage their reputation and reduce the number of follows and likes they receive.

Social media propaganda is a challenge in addition to all of these. Conflicts arise as a result of false accounts spreading inaccurate and inappropriate information. The main objectives of this research project are given below:

➢ These fake profiles are also made to gain more followers.

➢ Phony profiles have been shown to hurt more people than other online crimes. Therefore, it's critical to spot a phony profile since the user is informed.

➢ The main goal of fake accounts on the site is to disseminate spam, information, and other false information.

➢ This study takes a look at the technical work that has been done and what is being done now to find fake profiles.

➢ To protect real users from people with bad intentions, it's important to find these fake identities.

Here the author discusses finding fraudulent Twitter identities in this exact

situation. Different machine learning models are used by the author. The dataset of Twitter profiles is used, with INT, TWT, and FSF being used for fake accounts and E13 and TFP being used for authentic accounts. Typical defenses against the creation of fake profiles include:

➢ When building social media accounts, techniques like user verification must be used.

➢ User behavior research must be used to find suspicious activity. It will be advantageous to use a bot detection system that uses real-time AI analysis.

➢ You must make use of automatic bot prevention technology. By making the LSTM, XG boost, random forest, and multi-layered neural network models, the author made a contribution to technology. These methods are some instances of machine learning with supervision.

Additionally, LSTM uses tweets to categorize data, and soon its output will be able to be paired with such a convolutional CNN architecture [2]. Several sections make up the document. Prior research, data preprocessing, technique, experimental results, model accuracy, conclusion, and forthcoming investigations are all presented in an organized manner.

2. Literature Review

This debate has long been ongoing: is it a boon or a bane? Additionally, all businesses sought to offer a platform that had fewer faults and offered a better user experience. This leads to daily updates and new developments. We looked at prior research addressing related issues because we found that there hasn't been much progress in finding fake people's identities on social media sites like Twitter.

Several methods classified profiles in accordance with account activity, the volume of requests that were responded to, the volume of communications that were delivered, and other characteristics. A graph-based system underlies the models. Others attempted to distinguish between cyborgs and robots using certain techniques. Below is a list of some earlier studies. Messages are deemed spam if certain words are present in them. This theory has been applied to identify phony social media profiles. Techniques for pattern matching were employed to find these words on social media. However, this criterion has a big disadvantage because new terms are constantly being created and used. On Twitter, it is also becoming common to utilize acronyms like lol, gbu, and gn.

In 2008, Sybil Guard [3] was created with the goal of reducing the tainting impact of social media assaults by Sybil. The occurrence of walk-random encounters was restricted, and each node's random walk was also Kleinberg's synthetic social network served as the dataset. At about the same time as the Sybil guard, a different strategy known as the Sybil restriction was also created. It operates under the same premise as the Sybil guard, except that the zone outside of Sybil is fast combining. To make it work, each node used an approach that included

many random variables. Additionally, ranking was determined by the frequency of walk intersection tails. In 2009, Sybil-infer was created. Using the supposition that randomized walks and the non-Sybil area are quickly combined, it makes use of methods like model-based sampling, greedy algorithms, and Bayesian networks. Threshold selection is a probability-based selection method. Using greedy search, Mislove's algorithm from 2010 chose profiles from the Facebook dataset based on metric adjusted conductivity. The Facebook Immune System, a new model that included random forest, SVM, and boosting approaches, was introduced in 2011. The feature loops were the selection technique, and they also employed the Facebook dataset.

Depending on how many of your friends may have tags or connection histories, Facebook uses an algorithm to identify bots. The aforementioned guidelines can be used to spot bot accounts, but they fall short when it comes to human-made false accounts. Bot detection employs unsupervised machine learning. In this technological method, information was compiled based on proximity rather than tagging. Co-attributes made it possible for grouping functions to distinguish the bots so well.

In 2012, a regression approach called the Sybil rank [3] [4] was created. Interaction, tagging, and wall postings are to order the profiles. False accounts are rated lower than true accounts, which are ones with a higher ranking. However, this approach was unreliable.

since occasionally an actual profile would receive a poor rating, even when it was excellent. The next type was the Sybil frame, which was created. It used a multistage level of categorization. It functioned in two stages, first with a content-based strategy and later with a structure-based one.

These methods have been used in some recent research on this subject. The authors of one of the earlier studies [5] developed a blacklist that is capable of telling the difference between phony features and fake accounts. Using dynamic CNN as a framework, a study [6] offers Deep Profile, a method that uses a supervised learning algorithm to detect fake accounts. The authors of a study [7] combined the SVM, RF, as well as Adaboost to detect the OSN fake account. A study [8] specifically used regression analysis and also the random forest classifiers technique to identify phony Instagram accounts. Different writers create diverse connected works [9]-[18].

3. Methodology

The author employed XG Boost, a random forest [19] method, and observable features from a profile-focused multi-layered neural network in this model. The model can easily read the extracted characteristics that were saved in a CSV file. Finally, whether a profile is genuine or not is finally determined by the training, testing, and analysis of the model. Because Google provides free GPU utilization, researchers chose Google Colab to build models. The 12-gigabyte (GB) Google Colab NVIDIA Tesla K80 GPU can run continuously for 12 hours. This technique

is quite good at identifying fake profiles. After being trained, this model's accuracy might be greater than in earlier comparable research. This design also emphasizes a visually pleasing framework. A representation of the system architecture is shown in **Figure 1** below.
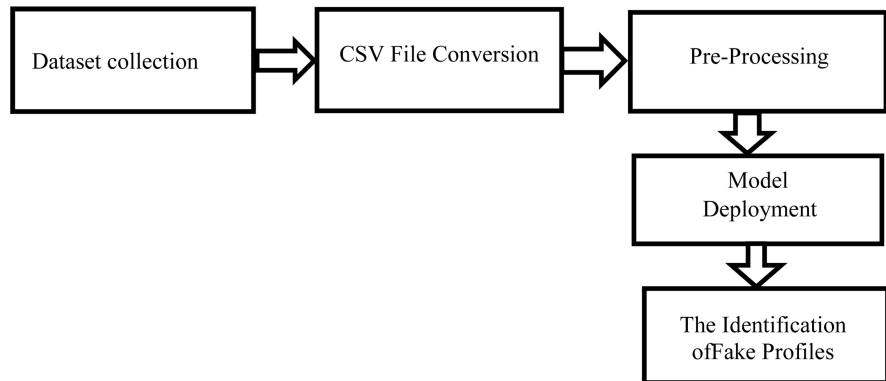

Figure 1. System architecture.

3.1. Dataset Collection

The MIB dataset was used by the author [20]. 3474 actual profiles and 3351 false profiles made up the data set. The data set utilized E13 and TFP for legitimate accounts and TWT, INT, and FSF for fraudulent ones. For machine extraction, the data is stored in CSV format.

In **Figure 2**, each indicator x-axis displays the characteristics that were utilized to recognize the fake profile. During the preprocessing, these were chosen. The number of entries for each feature that is present in the dataset is shown on the y-axis.
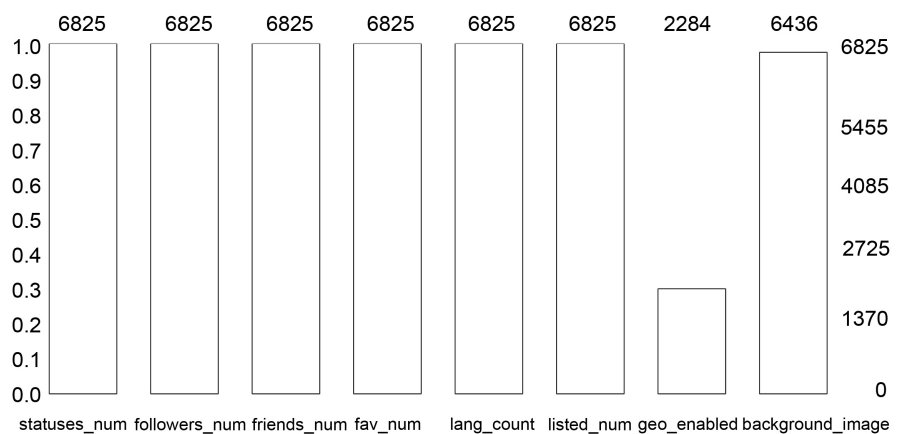

Figure 2. Dataset.

3.2. Model Development

In this section, the author presented the proposed solution to the challenge of detecting phony accounts by focusing on the features of such a situation. To begin with, a calculation was done to get the social network's graph's adjacency

matrix. After that, a calculation was made to determine the degrees to which nodes (social network users) are similarly based on their network friends. Following that, the similarity matrices for each of the stated metrics were constructed, including the similarity based on common friends, the similarity based on Jaccard, the similarity based on cosine, and any other relevant measures. At this point, several matrices were shown to show how similar the nodes were to each other.

All of the data was tagged as normal because the data in these circumstances is unbalanced and 98 - 99 percent of it relates to the exact majority class (normal users), making it difficult to understand the clarification of both the minority class (fake subscribers) and the overall accuracy of classifications. The SMOTE was used to get the statistics to reach equilibrium in order to tackle this problem.

3.3. CSV File Conversion

Due to the substantial dataset, the author chose to store it in Microsoft Excel. Then, for planned software to be able to make use of this data, the author exported these Excel files into CSV format. The following steps were taken at the time of the conversion of the Excel file to CSV:

➤ Launch the file that is being imported. Text Edit Notepad and other spreadsheet programs like Microsoft Excel and Google Sheets are both viable options for accomplishing this task. Alternatively, you might use Excel (Windows).
➤ Select File.
➤ Select the Save As option.
➤ If the author would like, the author can rename the file, and then choose the.csv extension (comma-delimited).
➤ Click the Save button.

After completing these steps, the author found a final dataset for our proposed system where the dataset was stored in a CSV file.

3.4. Proposed Methodology

The author used a variety of supervised algorithms, all with varying degrees of accuracy, to identify bogus Twitter profiles. Based solely on attributes that are visible, each model can identify a false profile. The same set of data is used to depict the accuracy and loss graphs for each supervised model. Additionally, comparison graphs of the accuracy of several models are displayed. The right optimization methods, loss functions, and logical operations are used to train the models. The list below includes a description of the models utilized.

3.4.1. Pre-Processing

Here the author adds one additional step of preprocessing before moving on to the models. Before the data set is delivered to a model, it is preprocessed. This approach seeks to determine whether a profile is genuine or fraudulent based on its look. All the specific details are now established. The categorical aspects have

been eliminated, leaving only the numerical data. Here the author chooses the following characteristics [21]:

friends	status count	fav num	geo enable	listed count	followers	lang count

A data set of accurate and inaccurate users is then combined, and each profile is given the additional label "isFake", a Boolean variable. The response related to profile X is then saved in the Y variable. Finally, zeros are used to replace any blank entries or NAN.

3.4.2. Artificial Neural Network

Deep learning neural network systems behave similarly to the similar neuron networks largely seen in the human brain [22]. Each layer of the neural network contains neurons (nodes). The author made use of Keras' sequential. Three hidden layers, an output layer, and an input layer are all parts of the model's construction (**Figure 3**). Each has activation capabilities aside from the output layer. As a function for activating the output layer, sigmoid is used. The model was built using the Adam optimizer and the binary merge loss function. This model makes use of ANN with the aforementioned architecture. Lastly, the sigmoid function gives out a number between 0 and 1 that shows whether it thinks a given profile is fake or real based on its prediction.
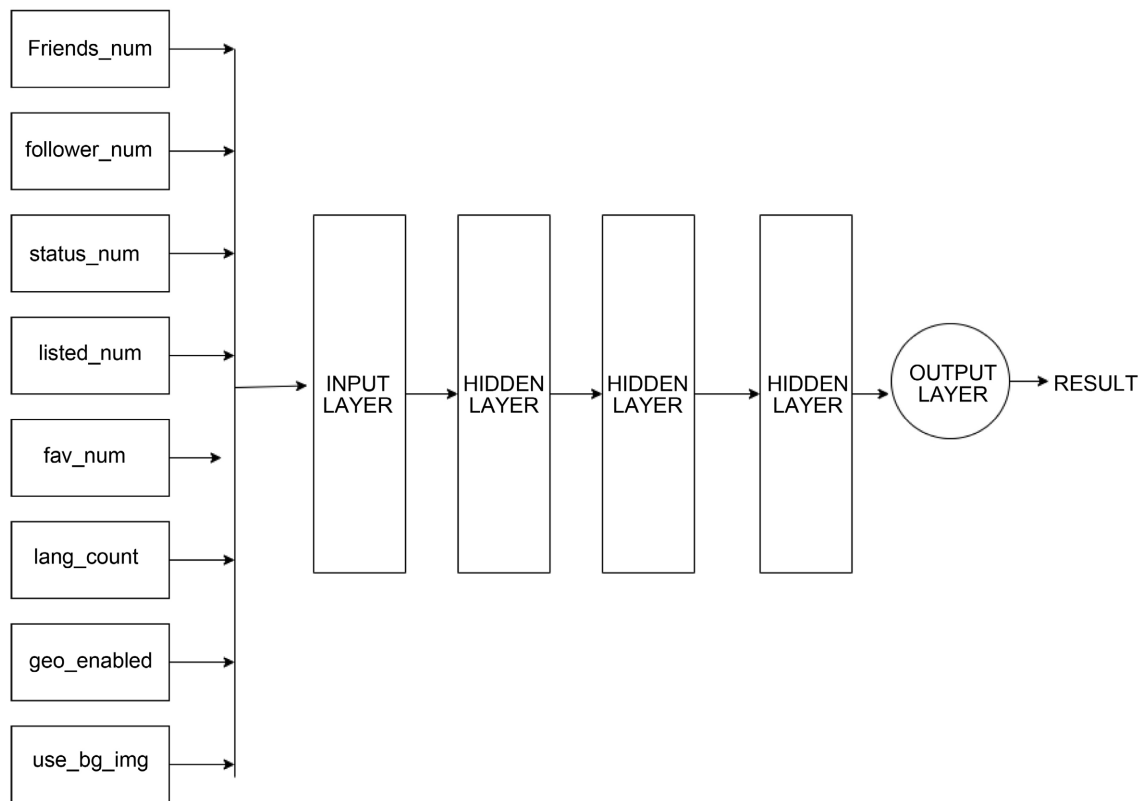
Figure 3. ANN architecture.

Hyper parameters

ReLUs (rectified linear units): The corrected activation functions are represented by the piecewise linear factor. Since ReLU is simple to train and produces superior results, it is often utilized as the primary learning algorithm in neural networks.

3.4.3. Random Forest

The ensemble learning approach known as random forest (or random-decision forest) is one example of this kind of method. Machine learning employs this technique because it is simple to apply both to classification and regression issues. Like in **Figure 4**, rather than depending on a single decision tree, the random forest uses predictions from each tree and predicts the result based on the votes of the majority of predictions. Random-forest, however, creates many more decision trees than the decision tree method does, and the final result seems to be the sum of nearly all of decision trees that have been created. For profile detection, the author employed the random forest method. The model takes in data and outputs relevant results. The bootstrap aggregating procedure is used to fit the trees (f_b) to the sample for the given set of $X = x_1, x_2, \cdots, x_n$ and $Y = y_1, y_2, \cdots, y_n$ answers. A random sample is selected at regular intervals (B times). After being trained, the following procedure is used to determine the results for a given sample(x):
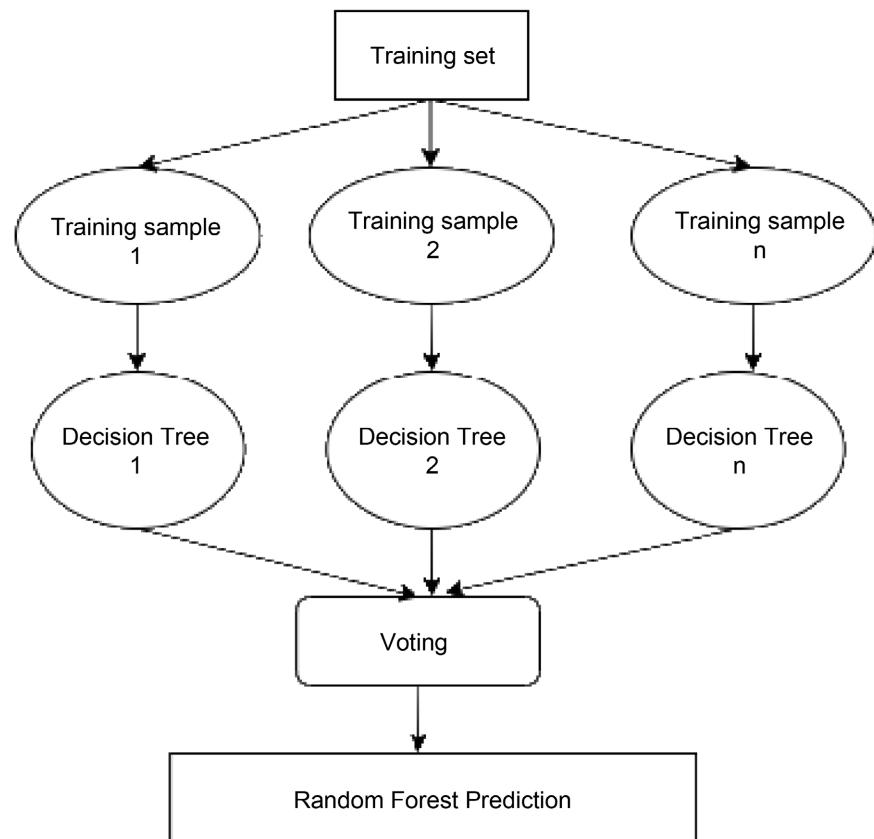

Figure 4. Random forest architecture.

$$f = \frac{1}{B} \sum_{b=1}^{B} f_b(x') \tag{1}$$

3.4.4. Extreme Gradient Boost

Another ensemble learning technique for regression is XG Boost. Subsampling of different parameters of Stochastic Gradient boosting this algorithm is implemented.

The disadvantage of random forest is that it works best when all the inputs are present, or when there are no missing values. The author employs a gradient boosting approach to get around this.

In accordance with the boosting procedure, $F_0(x)$ is initialized initially

$$f_0(x) = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, \gamma) \tag{2}$$

The gradient of the loss function is then calculated iteratively after that

$$\gamma_{im} = -\alpha \left[\frac{\delta L(y_i, F(x_i))}{F(x_i)} \right] \tag{3}$$

Finally, it defines the boosted model $F_m(X)$.

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \tag{4}$$

The learning rate is α.

The multiplicative factor is γ_n.

3.4.5. Long Short-Term Memory

The author of this work developed an LSTM-based framework for assessing a profile's credibility using tweets. While training an LSTM on this website and tweets, the author handled the filter that is used to remove the identifier strings from each tweet.

➢ Lowercase letters are used to write on all tokens.

➢ Tweets are no longer allowed to use stop words.

The author then used an embedding layer to build vector representations of the incoming words from these blockchain-enabled tweets. The output is then made by moving the single 32-dimensional vector output of the LSTM forward within layers that are triggered by sigmoid functions.

4. Experimental Results and Discussion

The outcomes of each model's training and testing are as follows. For the LSTM neural network, the ROC curves of stochastic forest, XG boost, and other approaches are given, along with model accuracy comparisons, loss versus eras graphs, and model comparisons.

4.1. Neural Network

For the trained neural network, the model accuracy graph and models' loss graphs are as follows (**Figure 5** and **Figure 6**).
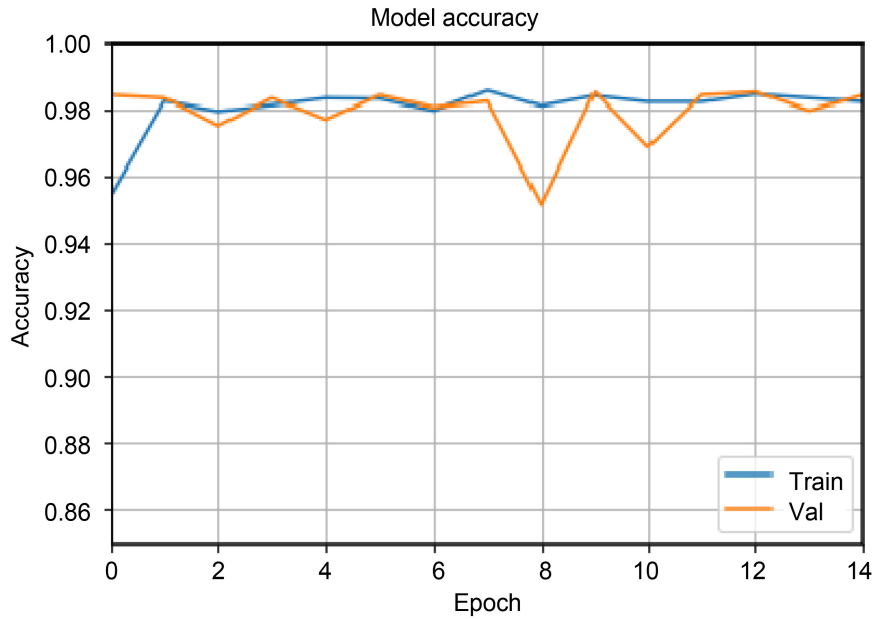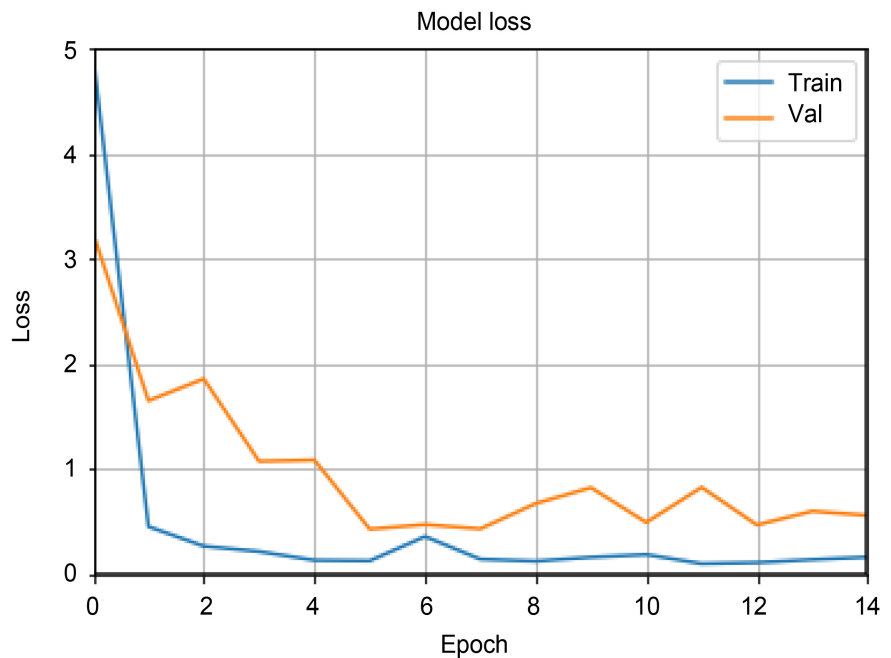
Figure 5. Model accuracy.

Figure 6. Model loss.

The accuracy and loss graphs shown above represent the results of 15 epochs of operation. Beginning at 0.97 and having reached its optimum, which is 0.98, the accuracy fluctuates initially during the course. The loss graph likewise has a local minimum of less than 0.5 before beginning at 1 for the test dataset and 4 for the validation data. It uses the binary cross-entropy function to calculate the loss. The machine first assigns random weights to each feature before giving each feature a particular weight.

4.2. Random Forest and Other Approaches

Several model's accuracy, such as decision trees, xgboost, random forests, and ada boosts, is shown in the comparison plot below (**Figure 7**). The XG boost, which is equal to 0.996, produces the highest level of precision. Additionally, decision trees and random forests both have an accuracy of about 0.99. The author now gets an ADA boost, at last.

The accuracy comparison (**Figure 7**) and ROC curve graphics (**Figure 8** and **Figure 9**) are shown below.
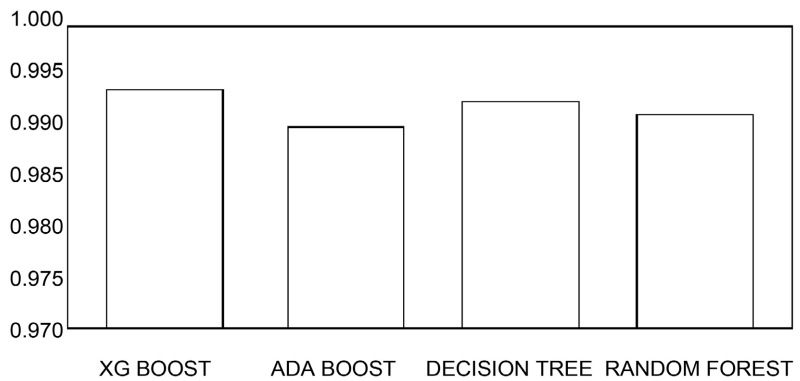
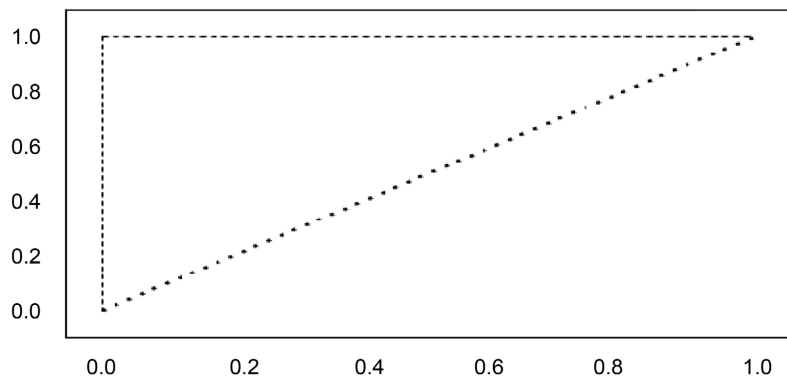Figure 7. Different model's accuracy.
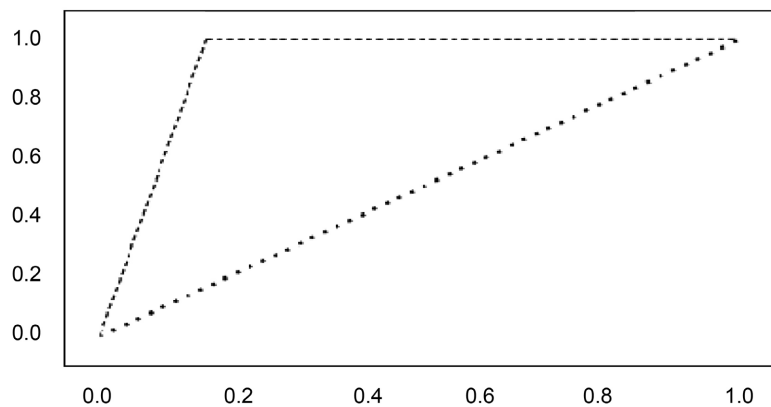
Figure 8. XG boost ROC curve.

Figure 9. Random forest ROC curve.

4.3. Discussion

Fake accounts on Twitter have the ability to change concepts like influence and popularity, which could have an effect on the economy, the political system, and society. They are dangerous for social media networks. This work uses a variety of algorithms to recognize false profiles, as the authors claimed in the introduction, so that makes sure that users won't be alarmed or damaged by malicious people. The authors of one of the previous research developed a blacklist that effectively distinguishes the fake features from the fake accounts. Various algorithms for machine learning are compared in this study to show which ones produce the best results (XGBoost 99.6%), even though those results are higher (94.9%) than those of the previous spam word list-based method (91.1%). In a study that used dynamic CNN, DeepProfile was introduced as a method that employs a supervised learning algorithm to foresee phony accounts. Another intriguing technique to determine sybil features based mostly on registration time was employed by another study [23]. Because they had similar IP phone numbers and addresses to the sybil, many legitimate individuals were incorrectly labeled as false positives, according to the investigation's authors. In different-sized towns, they had rates of false positives of 7%, 3%, and 21%. The study's authors had a 95% accuracy rate, which was a stunning result. In a study [24] that used feature extraction using phony profiles, the SVM-NN classification system had the highest performance of 98.3% in predicting sybil profiles.

5. Conclusion

The author used the CNN Model, Random Forests, and XG Boost supervised learning methodologies in this architecture to teach the system how to recognize fraudulent Twitter accounts depending on the information that is readily available. The main limitations of this project are that it works only on visible data and has no real-time application. By running a CNN on the numerical and categorical data as well as the profile photos, more tasks can be done. Also, adding more parameters, combining multiple models, and making a model that works in real time could lead to better results. The regions in the model and data may be given various degrees of prominence depending on their size or their particular significance in the recognition process. For instance, using this strategy would make it easier to pinpoint regions where extremely complex problems must be located, such as those that occasionally arise and the latter. Despite their complexity, these hybrid models ought to yield superior outcomes. However, occasionally combining these approaches may not have a significant impact on the outcome. The model will then be prepared for more social media sites like LinkedIn, Snapchat, WeChat, QQ, etc.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Van Der Walt, E. and Eloff, J. (2018) Using Machine Learning to Detect Fake Identities: Bots vs Humans. *IEEE Access*, **6**, 6540-6549. https://doi.org/10.1109/ACCESS.2018.2796018

[2] Kudugunta, S. and Ferrara, E. (2018) Deep Neural Networks for Bot Detection. *Information Sciences*, **467**, 312-322. https://doi.org/10.1016/j.ins.2018.08.019

[3] Ramalingam, D. and Chinnaiah, V. (2018) Fake Profile Detection Techniques in Large-Scale Online Social Networks: A Comprehensive Review. *Computers & Electrical Engineering*, **65**, 165-177. https://doi.org/10.1016/j.compeleceng.2017.05.020

[4] Hajdu, G., Minoso, Y., Lopez, R., Acosta, M. and Elleithy, A. (2019) Use of Artificial Neural Networks to Identify Fake Profiles. 2019 *IEEE Long Island Systems*, *Applications and Technology Conference* (*LISAT*), Farmingdale, 3 May 2019, 1-4. https://doi.org/10.1109/LISAT.2019.8817330

[5] Swe, M.M. and Myo, N.N. (2018) Fake Accounts Detection on Twitter Using Blacklist. 2018 *IEEE/ACIS* 17*th International Conference on Computer and Information Science* (*ICIS*), Singapore, 6-8 June 2018, 562-566. https://doi.org/10.1109/ICIS.2018.8466499

[6] Wanda, P. and Jie, H.J. (2020) DeepProfile: Finding Fake Profile in Online Social Network Using Dynamic CNN. *Journal of Information Security and Applications*, **52**, Article ID: 102465. https://doi.org/10.1016/j.jisa.2020.102465

[7] Kodati, S., Reddy, K.P., Mekala, S., Murthy, P.S. and Reddy, P.C.S. (2021) Detection of Fake Profiles on Twitter Using Hybrid SVM Algorithm. *E3S Web of Conferences*, **309**, Article No. 01046. https://doi.org/10.1051/e3sconf/202130901046

[8] Meshram, E.P., Bhambulkar, R., Pokale, P., Kharbikar, K. and Awachat, A. (2021) Automatic Detection of Fake Profile Using Machine Learning on Instagram. *International Journal of Scientific Research in Science and Technology*, **8**, 117-127. https://doi.org/10.32628/IJSRST218330

[9] Chakraborty, P., Muzammel, C.S., Khatun, M., Islam, S.F. and Rahman, S. (2020) Automatic Student Attendance System Using Face Recognition. *International Journal of Engineering and Advanced Technology* (*IJEAT*), **9**, 93-99. https://doi.org/10.35940/ijeat.B4207.029320

[10] Sayeed, S., Sultana, F., Chakraborty, P. and Yousuf, M.A. (2021) Assessment of Eyeball Movement and Head Movement Detection Based on Reading. In: Bhattacharyya, S., Mršić, L., Brkljačić, M., Kureethara, J.V. and Koeppen, M., Eds., *Recent Trends in Signal and Image Processing*, Springer, Singapore, 95-103. https://doi.org/10.1007/978-981-33-6966-5_10

[11] Chakraborty, P., Yousuf, M.A. and Rahman, S. (2021) Predicting Level of Visual Focus of Human's Attention Using Machine Learning Approaches. In: Shamim Kaiser, M., Bandyopadhyay, A., Mahmud, M. and Raym K., Eds., *Proceedings of International Conference on Trends in Computational and Cognitive Engineering*, Springer, Singapore, 683-694. https://doi.org/10.1007/978-981-33-4673-4_56

[12] Muzammel, C.S., Chakraborty, P., Akram, M.N., Ahammad, K. and Mohibullah, M. (2020) Zero-Shot Learning to Detect Object Instances from Unknown Image Sources. *International Journal of Innovative Technology and Exploring Engineering* (*IJITEE*), **9**, 988-991. https://doi.org/10.35940/ijitee.C8893.029420

[13] Sultana, M., Ahmed, T., Chakraborty, P., Khatun, M., Hasan, M.R. and Uddin, M.S. (2020) Object Detection Using Template and Hog Feature Matching. I*nternational Journal of Advanced Computer Science and Applications*, **11**, 233-238. https://doi.org/10.14569/IJACSA.2020.0110730

[14] Faruque, M.A., Rahman, S., Chakraborty, P., Choudhury, T., Um, J.S. and Singh, T.P. (2021) Ascertaining Polarity of Public Opinions on Bangladesh Cricket Using Machine Learning Techniques. *Spatial Information Research*, **30**, 1-8. https://doi.org/10.1007/s41324-021-00403-8

[15] Sarker, A., Chakraborty, P., Sha, S.S., Khatun, M., Hasan, M.R. and Banerjee, K. (2020) Improvised Technique for Analyzing Data and Detecting Terrorist Attack Using Machine Learning Approach Based on Twitter Data. *Journal of Computer and Communications*, **8**, 50-62. https://doi.org/10.4236/jcc.2020.87005

[16] Ahammad, K., Shawon, J.A.B., Chakraborty, P., Islam, M.J. and Islam, S. (2021) Recognizing Bengali Sign Language Gestures for Digits in Real Time using Convolutional Neural Network. *International Journal of Computer Science and Information Security* (*IJCSIS*), **19**, 11-19.

[17] Sultana, M., Chakraborty, P. and Choudhury, T. (2022) Bengali Abstractive News Summarization Using Seq2Seq Learning with Attention. In: Tavares, J.M.R.S., Dutta, P., Dutta, S. and Samanta, D., Eds., *Cyber Intelligence and Information Retrieval*, Springer, Singapore, 279-289. https://doi.org/10.1007/978-981-16-4284-5_24

[18] Ahmed, M., hakraborty, P. and Choudhury, T. (2022) Bangla Document Categorization Using Deep RNN Model with Attention Mechanism. In: Tavares, J.M.R.S., Dutta, P., Dutta, S. and Samanta, D., Eds., *Cyber Intelligence and Information Retrieval*, Springer, Singapore, 137-147. https://doi.org/10.1007/978-981-16-4284-5_13

[19] Reddy, S.D.P. (2019) Fake Profile Identification Using Machine Learning. *International Research Journal of Engineering and Technology* (*IRJET*), **6**, 1145-1150.

[20] Khaled, S., El-Tazi, N. and Mokhtar, H.M. (2018) Detecting Fake Accounts on Social Media. 2018 *IEEE International Conference on Big Data* (*Big Data*), Seattle, 10-13 December 2018, 3672-3681. https://doi.org/10.1109/BigData.2018.8621913

[21] Elyusufi, Y. and Elyusufi, Z. (2019) Social Networks Fake Profiles Detection Using Machine Learning Algorithms. In: Ahmed, M.B., Boudhir, A.A., Santos, D., El Aroussi, M. and Karas, İ.R., Eds., *Innovations in Smart Cities Applications Edition* 3, Springer, Cham, 30-40. https://doi.org/10.1007/978-3-030-37629-1_3

[22] Joshi, U.D., Singh, A.P., Pahuja, T.R., Naval, S. and Singal, G. (2021) Fake Social Media Profile Detection. In: Srinivas, M., Sucharitha, G., Matta, A. and Chatterjee, P., Eds., *Machine Learning Algorithms and Applications*, Scrivener Publishing LLC, Beverly, MA, 193-209. https://doi.org/10.1002/9781119769262.ch11

[23] Yuan, D., Miao, Y., Gong, N. Z., Yang, Z., Li, Q., Song, D., Wang, D. and Liang, X. (2019) Detecting Fake Accounts in Online Social Networks at the Time of Registrations. *Proceedings of the* 2019 *ACM SIGSAC Conference on Computer and Communications Security*, London, 11-15 November 2019, 1423-1438. https://doi.org/10.1145/3319535.3363198

[24] Roy, P.K. and Chahar, S. (2020) Fake Profile Detection on Social Networking Websites: A Comprehensive Review. *IEEE Transactions on Artificial Intelligence*, **1**, 271-285. https://doi.org/10.1109/TAI.2021.3064901