International Journal for Research in Engineering and Emerging Trends (IJREET), Volume 6, Issue 2, October, 2022
ISSN: 2545-4523 (Online)

ANN based Fake User Profile Detection

Dipanshu Sharma[1], Er. Ravinder Singh Madan[2]

1, M.Tech Scholar, IEC University Solan, Himachal Pradesh
2, HOD – CSE, IEC University Solan, Himachal Pradesh

***Abstract*— It's general known that social networking sites like Facebook, Twitter, and Histogram enjoy a significant amount of traffic because to their widespread use. The users of these social networking websites as well as users of other forms of media are in continual communication with one another and their contemporaries. They use these social networks to communicate information about themselves that is open to the public as well as information that is kept private. The allure of social networking websites has resulted in the engagement of a sizeable population of internet users. The notoriety contributes to the challenges that are experienced by the websites since it encourages the development of false accounts. The individuals who are in possession of bogus accounts steal the personal information of other users and then disseminate it throughout other social media sites. In our strategy, we suggest using machine learning methods such as neural networks and support vector machines (SVM) to identify bogus accounts on social media sites such as Twitter and Facebook. These methods are included in our plan. The approach that is being suggested displays the results of a simulation that was executed by using a variety of data mining tools for the strategy. We categorize the data in this to discover the fraudulent data accounts on social media websites by making use of a simple solution for the short user interaction with a data mining tool, false accounts from accessible data using the machine learning approaches that were described before.**

Keywords—ANN; SVM; machine learning

I. INTRODUCTION

Because of our ever-increasing dependence on computer technology in this day and age, the typical person is now at a greater risk of being a victim of crimes such as the unauthorized exposure of personal information and even identity theft. It is possible that these assaults may take place with no prior warning at all, and the individuals whose personal information has been stolen will often be unaware that anything has happened to them. On the other hand, it is not impossible for them to. There aren't a lot of compelling reasons for social networks to beef up the level of data security that they provide for their users at the present, so it's unlikely that this will happen anytime soon. This is due to the method in which technological advancements have occurred. The great majority of the time, security flaws and vulnerabilities of this kind will have an effect on social media networks like Facebook and Twitter. [7] They are also able to target monetary establishments such as banks and other organizations operating in the monetary sector. This is one of the abilities that they have at their disposal.

The process of determining whether or not a friend request made via Facebook is legitimate by using the principles of machine learning and, more particularly, an artificial neural network to the analysis. In addition to this, could you perhaps provide a synopsis of the several classes and libraries that are relevant to the topic that is the focus of the discussion that is now taking place? In addition to this, we will talk about the sigmoid function, the method for assigning weights, as well as the way in which the model makes use of the information that is gathered from the weights. In conclusion, it is essential to take into account the needs of the social networking page, seeing as how these criteria are of the biggest significance in regard to the solution that has been described.

The unauthorized collecting of personally identifiable information is already fraught with risk; however, the use of bots and phoney accounts increases the stakes of this risk to a whole new level. Bots are computer programs that can gather information about a user even if the user is unaware that information is being collected about them. This is because bots can collect information even if the user is logged out of their account. This is due to the fact that bots are able to gather data even if the user is not currently signed into their account. This is because bots have the capability of gathering information even while the user is not logged into their account, which is the reason for this behavior. If a person visits a website that saves a record of the information that they submit, then there is a possibility that something like this may occur to that person. The activity that is being carried out in this instance is known as web scraping, which is a term that appropriately depicts the strategy that is being used.

Because this behaviour is not only tolerated by the law, but also actively encouraged by it, the predicament is a great deal more dangerous than it would have been in any other circumstance. The administrators of a social networking site have the option of either keeping bots hidden or enabling them to appear as false friend requests so that bots can access private information on the site. Both of these options allow bots to access information that is normally protected from public view.

Because of the ever-increasing dependence on computer technology in this day and age of the digital revolution, the average person is now more susceptible to crimes such as data breaches and probable identity theft. This is because of the fact that the digital revolution is in full swing. It is possible that these assaults will take place without any prior warning, and the individuals whose private information has been hacked will often never learn what has taken place to them. At this point in time, there is not a great lot of reason for social networks to increase the amount of data protection that they offer for their users. This is something that is something that should be addressed in the future. The majority of the time, social media networks like Facebook and Twitter are the ones to be affected by these sorts of security flaws and vulnerabilities. [Citation needed] They are also able to target monetary institutions like banks and other financial organisations. This is one of their capabilities.

It seems as if there is a noteworthy event concerning the hacking of social media networks virtually on a daily basis. These occurrences are often reported in the media. A data breach recently occurred at Facebook, and it is estimated that the privacy of around 50 million users' sensitive information was exposed as a result of the incident. Facebook supplies its users with a set of terms and conditions that clarify what the company does with the user's data. These terms and conditions have been thoroughly discussed. You may read more about these policies on Facebook's official website. Even while the rule isn't particularly successful, it does make an attempt to stop people from repeatedly compromising the privacy and safety of other people by making it illegal for them to do so. It would seem that the security processes that are incorporated into Facebook may be easily circumvented by using fake accounts since they are so straightforward.

Participating in various online social networks is one of the most frequent things that people do when they are browsing the internet. Users are able to interact with their friends and family, express their ideas and views, and engage in discussions about current events thanks to these networks. Online communities that fulfil this definition include Facebook, Twitter, and LinkedIn, just to name a few examples. Twitter has quickly risen to become one of the most popular social networks that can be accessed online. The messages that users send to Twitter are referred to as "tweets," which is also the name of the platform. Typically, these communications are limited to a maximum of 140 characters. A good example of a microblogging service is Twitter, which first appeared on the scene in 2006. Twitter has around 330 million active users, who together publish approximately 500 million tweets each and every single day. Twitter is often attacked by spammers who use the network for nefarious purposes as a result of its enormous popularity. These goals could include, but are not limited to, the propagation of malicious URLs inside tweets, the fabrication of misleading information, and the transmission of unwanted messages to other users. According to the regulation that Twitter has established about the protection of its users' privacy, an account is regarded as spam if it has a large number of individuals following it but a relatively small number of people really following them back. It is true that the popularity and reputation of an account are negatively correlated with the number of individuals who follow that account.

The great majority of the approaches that have been reported in the relevant literature have produced spammer detection systems that are wholly dependent on account-related features. This is the case for the vast majority of these approaches. These criteria include the number of people who follow you and who you follow, the amount of tweets you post, and a wide range of additional considerations. These features have been used to calculate a variety of scores, such as the fifo score, which represents the ratio of the number of an account's followings to the number of its followers, and the reputation score, which is computed as the ratio of the number of an account's followers to the sum of its followers and followings. Both of these scores are based on the ratio of the number of an account's followers to the total number of the account's followers and followings. Both of these scores are determined by the proportion of an account's total number of followers and followings to the number of the account's followers alone. The features of the account that are being analysed serve as the basis for determining both of these ratings. On the other hand, spammers on Twitter may be able to sidestep these measures to detect spam by purchasing followers through a third-party marketplace. Because of this, the dependability of certain capabilities is decreased in comparison to what it would be in the absence of this factor. This gives us the push we need to build new skills that are more strong so that we can identify spammers on Twitter, and it does so in a timely manner.

Users now have the capacity to interact with people located in different parts of the world as a direct result of the proliferation of online social networks (OSNs) such as Facebook and Twitter. Open Social Networks (OSNs) provide users with a platform on which they can establish and exchange personal profiles, as well as text, photo, audio, and video content; discover and make friends; and so on. Users can also use this platform to do things like shop for clothes, listen to music, and watch movies. Websites for social networking are often employed when it comes to marketing a company's goods or services over the internet in the capacity of online advertising and promotion. OSNs are used as a platform by government organisations to facilitate the delivery of government services to residents in an effective manner as well as to educate and enlighten individuals about a variety of different scenarios. This is done to facilitate the delivery of government services to residents in an effective manner. This is done in order to improve the efficiency with which governmental services may be delivered to the public. As a direct consequence of the widespread use of social networking sites, an enormous quantity of information has been disseminated all over the globe. This information may be accessible on social networking websites in the form of reviews, such as those that can be found on Amazon and Yelp; posts, messages, and comments on Facebook; advertising that have been liked on Facebook; and tweets on Twitter. Users place a significant amount of faith in the information that is made accessible on OSNs, and as a result, their perspectives are either directly or indirectly influenced by this material. [Here's a good example:] Facebook is regarded to be one of the biggest open social networks (OSNs) in the whole world since it has 1.44 billion monthly active members and has the potential to be a rich source of information. Additionally, Facebook has the potential to be a rich source of advertising revenue.

II. PROPOSED WORK

In order to demonstrate how to build an ANN neural network based image classifier, we are going to build a 6 layer neural network that can recognise and differentiate one photo from others. This will allow us to demonstrate how to build an ANN neural network based image classifier. This will be a very simple network that is also capable of being run on a centralised processing unit that we will build. Conventional neural networks need a substantial amount of time and contain a significant number of training parameters when compared to other forms of neural networks. This is because traditional neural networks are trained on a typical central processing unit (CPU). Nevertheless, the purpose of this is to show how to build a convolutional neural network using TENSORFLOW that is useful in the real world.

In order to differentiate real accounts from bogus ones, the suggested study makes use of techniques such as support vector machines (SVM) and neural networks (NN). The feature set that aids in the detection of bogus accounts will soon be made available on Facebook, Twitter, and Twitter. The same feature set is available on Twitter as well. The proposed approach ought to be able to realise the greater recall and f-measure values that are necessary to recognise phoney accounts on social media sites such as Facebook, Twitter, or Twitter. There are a few different approaches to machine learning that are examples of procedures that provide accurate results. Some of these approaches include neural networks and support vector machines. When it comes to the categorization of data, neural networks and support vector machines provide superior outcomes.

Freeman et al. (2015) conducted a study with the intention of establishing the efficacy of various different approaches to the detection of false accounts in online social networks. According to the conclusions of the research, the phrase "fake accounts" refers to the use of online social platforms by malicious persons in order to spread abuse and spam across the whole system and engage in fraudulent activity. It's possible for a single individual to generate hundreds of fake accounts in order to broaden the scope of their activities and interact with a massive number of real people. This is feasible as a result of the capability of the platform to handle numerous logins for each individual user. It is essential, not only for the safety of genuine users but also for the maintenance of the trustworthiness of all platforms and networks, to identify fraudulent accounts and to take appropriate action against them in order to ensure that the system continues to maintain its reputation for reliability.

On the other hand, the fabricated testimonies of many individuals provide the impression that the statements are real at first look. It is hard to identify fraudulent accounts based on those characteristics since, for example, they have believable profiles and names that seem real (Freeman et al., 2015). This research has offered a description of the scalable technique that may be utilised to find the cluster of fraudulent accounts that were established by the same user. These accounts were created by the same individual. A supervised machine learning pipeline that is able to differentiate between excellent and poor software is the most critical strategy to implement. Attacks will be avoided as a result of this measure.

The numerous machine learning algorithms may simulate a problem with background information or knowledge for the process of model preparation, which helps in choosing the optimal method that will take the

provided input data and produce the best result. This aids in the machine learning community's efforts to advance artificial intelligence (AI).

1. Support Vector Machine (SVM): Support-vector machines (also known as support-vector networks) are supervised learning models with corresponding learning algorithms that examine data used for regression analysis and classification. Support-vector machines (SVMs) are also known as support-vector networks. The method, when applied to the labelled training data that is supplied, produces an ideal hyper plane that is capable of classifying new samples (supervised learning).

2. Neural Networks, Also Known As The term "neural network" now refers to an artificial neural network that is constructed of artificial neurons or nodes. This kind of network is currently in use. In the context of artificial neurons, a neural network (NN) is a connected collection of actual or made-up neurons that processes information using a mathematical model. Artificial neurons might be genuine or made-up neurons.

The following procedures are included in the methodology:

1. Acquiring the relevant data set The first step in recognising a false profile on either Facebook or Twitter is to get the relevant data set for either of the two platforms, depending on which one is available. A survey methodology is used to collect the data set from Twitter or Facebook in preparation for the task that is planned. Using the survey method, we collect the data sets that users of Facebook or Twitter have posted to those platforms. In order to accomplish this goal, we create a Google form. Because the Google form has a wide array of questions, we are better able to differentiate between accounts that have no data and accounts that have been fraudulently created.

2. Filtration of the dataset: After that, we go on to the next step, which involves filtering the data set that we have gathered using a randomised filtration approach. Through the use of randomization, the positions of the profiles within the dataset are shifted about in an unpredictable manner. The wrong values in the dataset are also removed using the filtering approach, and the removed value is replaced with an average of the upper and lower column values in the dataset.

3. Clustering of the dataset: This step takes place after the filtering procedure has been applied to the data set. In order to partition the dataset into distinct groups, the K-media clustering method is used. In order to identify a collection of fraudulent accounts, the clustering methodology is used. K-mediod is a better clustering approach than K-mean since it is less sensitive to the presence of outliers in the data.

4. Feature selection: At this point in the process, the feature set will be put through the feature selection process. In the process of feature selection, principal component analysis is one of the methods used. In addition, principal component analysis is used to aggregate the characteristics of the linked factors and provide a value to the combined set. If a feature selection approach is used, it is possible to achieve a higher level of accuracy while using a less number of features. Because by using the feature selection strategy, we have the ability to omit the qualities that have the lowest weight, or even zero weight.

5. Classification of the dataset The classification method is utilised once the feature selection and clustering processes have been completed. In order to categorise the data, SVM and neural networks are used. Two methods of machine learning that can handle a wide range of data sets efficiently are the support vector machine (SVM) and neural networks. Because it needs less time to be trained, SVM runs faster while it's actually being used.

6. A comparison of the findings reveals that the neural network and SVM provide two separate outputs when using either the current method or the strategy that was recommended. The results of both processes, in addition to those obtained via the use of other approaches, are compared. In it, a comparison is made between the qualities and length of time required by the hybrid approach that has been offered and the method that has already been developed. The outputs of neural networks and support vector machines (SVM) are compared in the hybrid method that was recommended, and the outcomes that have a higher level of accuracy are taken into account.

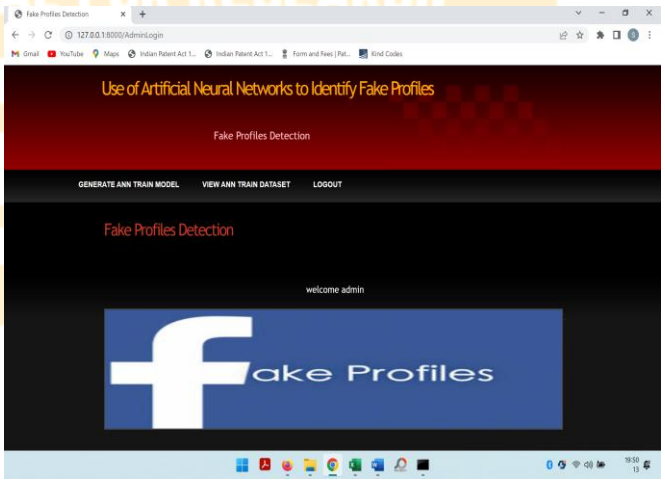

Fig 1. User Interface

To access the screen below, go to the "ADMIN" link on the previous screen.

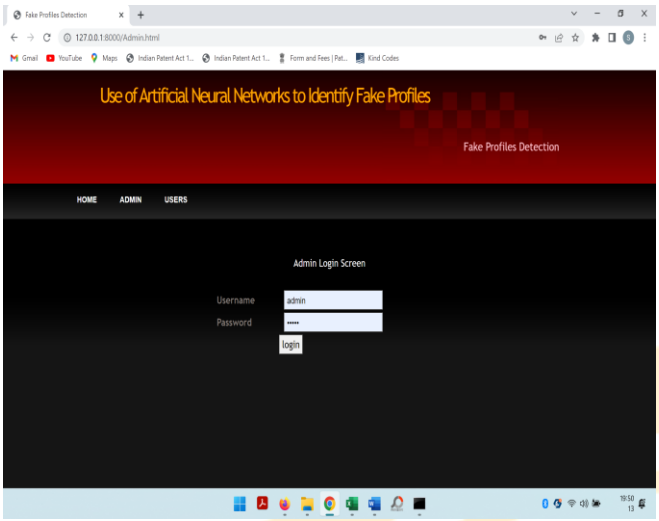

Fig 2. Login

To log in as admin, on the screen above, enter admin and admin for both the username and the password. After logging in, you'll see the screen below.

To construct a training model based on the dataset, on the page that you're now on, click the button labelled "Generate ANN Train Model." When you follow that link, you will be sent to the server console, where you may inspect the ANN processing details with precision.

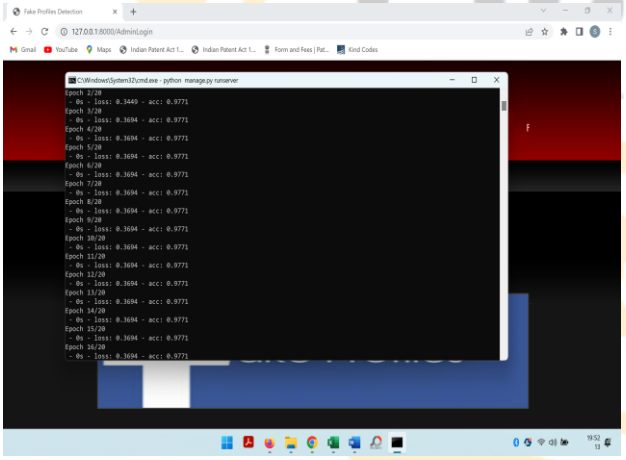

Fig. 3 ANN Processing

All ANN information is shown in the black console that sits above this page.

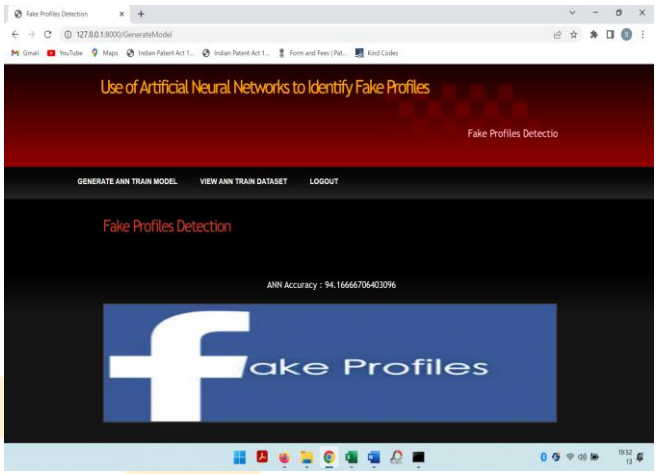

Fig. 4 Accuracy

We can see that ANN achieved a level of accuracy of 94% while training all of the Facebook profile data using the ANN method on the screen that was just shown to us.

In fig. 5 screen enter some test account details to get prediction/identification from ANN. You can use below records to check

10, 1, 44, 0, 280, 1273, 0, 0

10, 0, 54, 0, 5237, 241, 0, 0

7, 0, 42, 1, 57, 631, 1, 1

7, 1, 56, 1, 66, 623, 1, 1

Fig. 5 Output for fake user detection

Table 1: Comparison with different techniques

	Naïve Bayes [4]	NLP [9]	Forest Regression [5]	ANN (proposed)

Accuracy	89%	92%	93%	94%

III. CONCLUSION

According to the findings of the study, it has been established that the number of people using social media platforms is growing on a daily basis, and that practically everyone possesses a mobile device that is capable of using social media platforms. However, nobody is aware of any controversy that may occur, such as spam being sent to their accounts and a great deal more.

In this essay, the author discusses all of the variables that are related with the identification of fake accounts. He adds that identifying false accounts on the internet is one way in which the online social network helps safeguard OSN creators and their users from a variety of hazardous behaviours. The primary and the majority of the detection systems make an effort to assess the genuine account user, taking into consideration factors such as honesty, and they start by analysing the structure of graph level and user level operation. These defence mechanisms are not effective against the hostile effort in which phoney accounts try to conceal their activity by mimicking the actions of actual users. In order to persuade and promote public opinions, actions such as spamming, artificial inflation, distributing disinformation, and spreading abuse to a large number of users on a platform are solely included in the employment of these fake and fraudulent identities. In this study, a feature-based approach was used in order to analyse the false profiles that were deployed on the various apps.

REFERENCES

[1] M. Egele, G. Stringhini, G. Stringhini, and G. Vigna, “Towards Detecting Compromised Accounts on Social Networks,” IEEE, vol. 5971, no. c, 2015.

[2] D. M. Freeman and T. Hwa, “Detecting Clusters of Fake Accounts in Online Social Networks Categories and Subject Descriptors,” AISec, 2015.

[3] K. B. Kansara, “Security against sybil attack in social network,” ICICES, no. Icices, 2016.

[4] A. M. Meligy, “Identity Verification Mechanism for Detecting Fake Profiles in Online Social Networks,” IJCNIS, no. January, pp. 31–39, 2017.

[5] Ashraf Khalil, Hassan Hajjdiab, and Nabeel Al-Qirim , Detecting Fake Followers in Twitter: A Machine Learning Approach, International Journal of Machine Learning and Computing, Vol. 7, No. 6, December 2017.

[6] S. Khaled, N. El-Tazi and H. M. O. Mokhtar, "Detecting Fake Accounts on Social Media," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 3672-3681.

[7] S. Khaled, N. El-Tazi and H. M. O. Mokhtar, "Detecting Fake Accounts on Social Media," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 3672 3681.

[8] N. Singh, T. Sharma, A. Thakral and T. Choudhury, "Detection of Fake Profile in Online Social Networks Using Machine Learning," 2018 International Conference on Advances in Computing and Communication Engineering (ICACCE), Paris, 2018, pp. 231-234.

[9] Xiao, C., Freeman, D.M. and Hwa, T., 2015, October. Detecting clusters of fake accounts in online social networks. In Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security (pp. 91-101).

[10] Boshmaf, Y., Logothetis, D., Siganos, G., Lería, J., Lorenzo, J., Ripeanu, M. and Beznosov, K., 2015, February. Integro: leveraging victim prediction for robust fake account detection in osns. In NDSS (Vol. 15, pp. 8-11).

[11] Y. Boshmaf, D. Logothetis, G. Siganos, J. Lería, J. Lorenzo, M. Ripeanu, K. Beznosov, H. Halawa, "Íntegro: Leveraging victim prediction for robust fake account detection in large scale osns", Computers & Security, vol. 61, pp. 142-168, 2016.

[12] Kaubiyal, J. and Jain, A.K., 2019, August. A feature based approach to detect fake profiles in Twitter. In Proceedings of the 3rd international conference on big data and internet of things (pp. 135-139).