REVEALING FAKE PROFILES THROUGH ARTIFICIAL BRAINPOWER

ABSTRACT
In this paper, we use machine learning, namely an artificial neural network to determine what are the chances that Facebook friend request is authentic or not. We also outline the classes and libraries involved. Furthermore, we discuss the sigmoid function and how the weights are determined and used. Finally, we consider the parameters of the social network page which are utmost important in the provided solution.

1. INTRODUCTION
In 2017 Facebook reached a total population of 2.46 billion users making it the most popular choice of social media. Social media networks make revenues from the data provided by users. The average user does not know that their rights are given up the moment they use the social media network's service. Social media companies have a lot to gain at the expense of the user. Every time a user shares a new location, new photos, likes, dislikes, and tag other users in content posted, Facebook makes revenue via advertisements and data. More specifically, the average American user generates about $26.76 per quarter. That number adds up quickly when millions of users are involved. In today's digital age, the ever-increasing dependency on computer technology has left the average citizen vulnerable to crimes such as data breaches and possible identity theft. These attacks can occur without notice and often without notification to the victims of a data breach. At this time, there is little incentive for social networks to improve their data security. These breaches often target social media networks such as Facebook and Twitter. They can also target banks and other financial institutions

LITERATURE SURVEY

1. In today's digital life, the dependency of computer technology is increasing day-by-day. So, the main target of hackers are social media networks such as Facebook, Instagram, and twitter. So, this is being a major cyber security problem these days. This intends to build an Artificial Intelligence solution to prevent the dangers of a bot in the form of fake profile on social media. The popularity of social networks has skyrocketed in recent years, with a glut of personal information being posted to sites like Facebook and LinkedIn by users. But this unchecked expansion has opened the door to problems such as false profiles and malicious actors. Fake accounts can be used for malicious purposes such as spreading spam or committing fraud, so it's important to have systems in place that can detect these fraudulent accounts. In this paper, we focus on social networks and analyze methods of detecting fake profiles, comparing them across various applications and measuring performance criteria.

2. There is a huge expansion in advancements nowadays. Mobiles are becoming shrewd. Innovation is related with online informal organizations which has turned into a section in each one\'s life in making new companions and keeping companions, their inclinations are known simpler. Be that as it may, this expansion in systems administration online makes numerous issues like faking their profiles, online pantomime having become increasingly more in present days. Clients are taken care of with more superfluous information during riding which are posted by counterfeit clients. Explores have seen that 20% to 40% profiles in internet based informal organizations like Facebook are phony profiles. This identification of phony profiles in internet based informal communities' results into arrangement utilizing systems.

3. These days, there is a noticeable increase in applied sciences. Mobile phones are becoming smarter. Technology is associated with online social networks, which have emerged as a part of everyone's life in terms of making new friends and retaining friends,. In this research, we employ computing device learning, specifically a synthetic neural community, to identify whether or not a Facebook buddy request is legitimate. We also outline the training and libraries that will be used. Finally, we reflect on consideration on the parameters of the social community web page which are utmost necessary in the furnished solution.

4. 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

5. The use of NLP (Natural Language Processing) techniques to identify "misleading" and "news" reports that originate from unreliable sources. only a count vector (Term Frequency Inverse Document Frequency) (word sizes contrasted to how (word sizes relative to how often they are used in other articles in your dataset) was generated using values from this function vector and those features with relative importance of 1.4 and that feature with relative importance 2.0 (or equal importance) (word) The linguistic models, however, ignore aspects such as word order and meaning. Two documents with completely different word counts can refer to the same thing. As a result, the data science community has put in place various measures to resolve this problem. Facebook is participating in a challenge on Kaggle to remove fabricated news stories from feeds on their social network using AI. fighting fake news is a straightforward job Can you separate "real" and "news" from "fakes?" Thus, the proposed study would have the false and the real news datasets as input, and use the Naive Bayesifier to construct a model that classifies articles by the words they contain. owing to the increased number of online information sources, it is difficult to know what is right and what is incorrect Therefore, the issue of "fake news" has gained further publicity. This research looks at historical and contemporary approaches for determining truth and falsity in text format, as well as how and why it occurs. This paper combines Nave Bayes Classifier, support vector machines, and semantic analysis to identify fake news, coming up with a system of three sections.

3. EXISTING SYSTEM

Malicious users create fake profiles to phish login information from unsuspecting users. A fake profile will send friend requests to many users with public profiles. These counterfeit profiles bait unsuspecting users with pictures of people that are considered attractive. Once the user accepts the request, the owner of the phony profile will spam friend requests to anyone this user is a friend. The fake profile's contents typically have links that lead to an external website where the damage happens. An unaware curious user clicking the bad link will damage their computer. The cost can be as simple as catching a virus to as bad as installing a rootkit turning the computer into a zombie. While Facebook has a rigorous screening to keep these fake accounts out, it only takes one fake profile to damage the computers of many.

3.1 LIMITATIONS OF SYSTEM

- Analyzing of account data becomes more critical for humans.
- Detecting of fake account take time

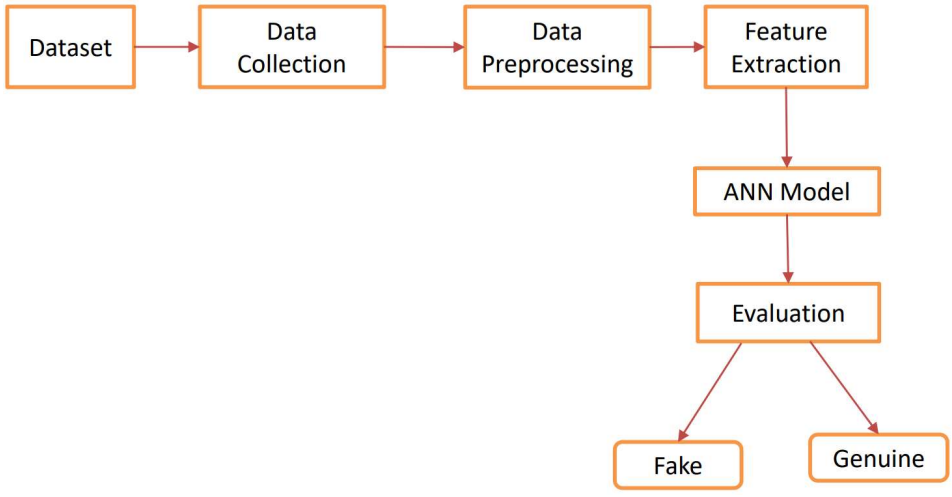
4. PROPOSED SYSTEM

In our solution, we use machine learning, namely an artificial neural network to determine what the chances that a friend request is authentic are or not. We utilize Microsoft Excel to store old and new fake data profiles. The algorithm then stores the data in a data frame. This collection of data will be divided into a training set and a testing set. We would need a data set from the social media sites to train our model. For the training set, the features that we use to determine a fake profile are Account age, Gender, User age, Link in the description, Number of messages sent out, Number of friend requests sent out, Entered location, Location by IP, Fake or Not. Each of these parameters is tested and assigned a value. For example, for the gender parameter if the profile can be determined to be a female or male a value of (1) is assigned to the training set for Gender. The same process is applied to other parameters. We also use the country of origin as a factor

4.1 ADVANTAGES OF PROPOSED SYSTEM :-

- Analyzing large amount of data becomes very easy.
- No human power requires.
- Using ANN algorithm easy to identify the status of account weather the account fake or genuine

5.SYSTEM ARCHITECTURE



6. IMPLEMENTATION

1. Admin Module: Admin will login to application by using username as 'admin' and password as 'admin' and then perform below actions.

a) Generate ANN Train Model: Admin will upload profile dataset to ANN algorithm to build train model. This train model can be used to predict fake or genuine account by taking new account test data.

b) View ANN Train Dataset: Using this module admin can view all dataset used to train ANN model.

2.User Module: Any user can use this application and enter test data of new account and call ANN algorithm. ANN algorithm will take new test data and applied train model to predict whether given test data contains fake or genuine details.

7. EXPECTED RESULTS

Online social networks such as Facebook or Twitter contains users details and some malicious users will hack social network database to steal or breach users information, To protect users data we are using ANN Algorithm.

To train ANN algorithm we are using below details from social networks

Account_Age, Gender, User_Age, Link_Desc, Status_Count, Friend_Count, Location, Location_IP, Status

All fake users main intention is to send friend request to normal users to hack their machine or to steal their data and never they will have many number of posts or have many following friends and their account age also will have less number of years. By analysing this features Facebook will mark whether user profile is fake or genuine. This Facebook profile data we downloaded from Facebook website and using this data to train ANN model. Below are some values from profile dataset.

Account_Age, Gender, User_Age, Link_Desc, Status_Count, Friend_Count, Location, Location_IP, Status

10,1,22,0,1073,237,0,0,0
10,0,33,0,127,152,0,0,0
10,1,46,0,1601,405,0,0,0
10,0,25,0,704,380,0,0,0
7,1,34,1,64,721,1,1,1
7,1,30,1,69,587,1,1,1
7,1,36,1,61,782,1,1,1
7,1,52,1,96,827,1,1,1

In above dataset all bold names are the dataset column names and all integer values are the dataset values. As ANN will not take string value so we convert gender values to 0 or 1, if male value is 1 and if female value is 0. In above dataset last column give us information of fake or genuine account if last column contains value 0 then account is genuine otherwise fake. All fake account will have less number of posts as their main intention is to send friend requests not posts, so by analysing this features Facebook mark that record with value 1 which means it's a fake account. We are using above dataset to train ANN model and this dataset saved inside code 'dataset' folder. After building train model we input test data with account details and ANN will give result as fake or genuine. Below are some values from test data

Account_Age, Gender, User_Age, Link_Desc, Status_Count, Friend_Count, Location, Location_IP

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

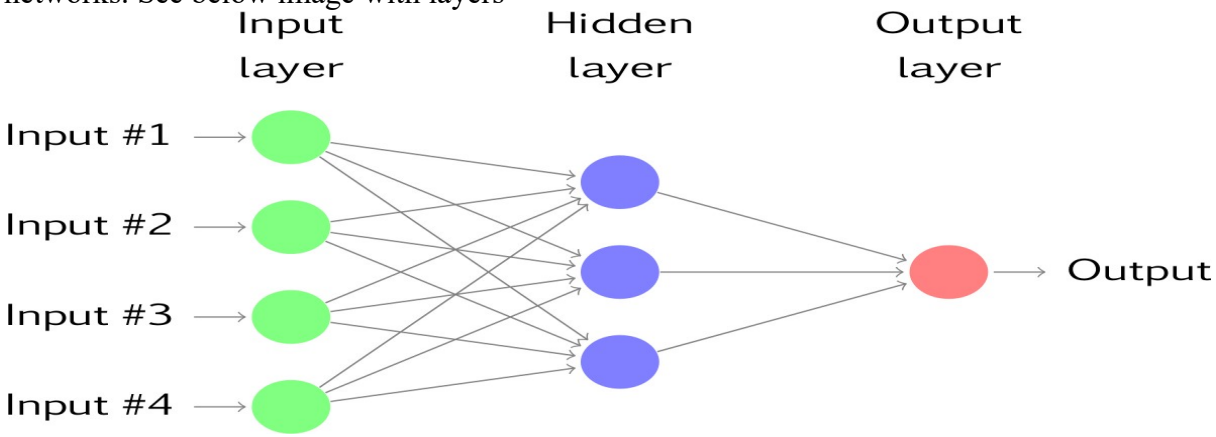
In above test data STATUS column and its value is there and ANN will predict status and give us result whether above test data is fake or genuine. In output we can see result of above test data.

ANN algorithms Details

To demonstrate how to build aANN neural network based image classifier, we shall build a 6 layer neural network that will identify and separate one image from other. This network that we shall build is a very small network that we can run on a CPU as well. Traditional neural networks that are very good at doing image classification have many more parameters and take a lot of time if trained on normal CPU. However, our objective is to show how to build a real-world convolutional neural network using TENSORFLOW.

Neural Networks are essentially mathematical models to solve an optimization problem. They are made of neurons, the basic computation unit of neural networks. A neuron takes an input (say x), do some computation on it (say: multiply it with a variable w and adds another variable b) to produce a value (say; z= wx+b). This value is passed to a non-linear function called activation function (f) to produce the final output(activation) of a neuron. There are many kinds of activation functions. One of the popular activation function is Sigmoid. The neuron which uses sigmoid function as an activation function will be called sigmoid neuron. Depending on the activation functions, neurons are named and there are many kinds of them like RELU, TanH.

If you stack neurons in a single line, it's called a layer; which is the next building block of neural networks. See below image with layers

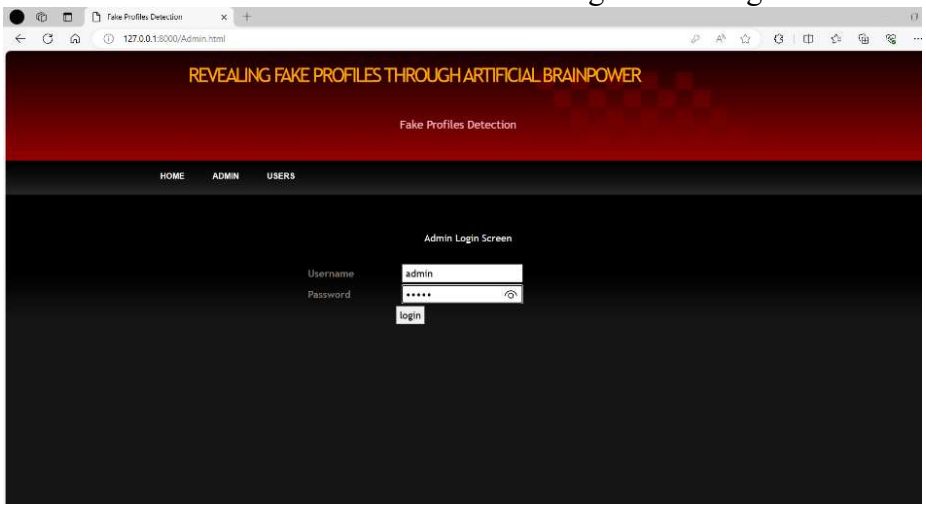


To predict image class multiple layers operate on each other to get best match layer and this process continues till no more improvement left.

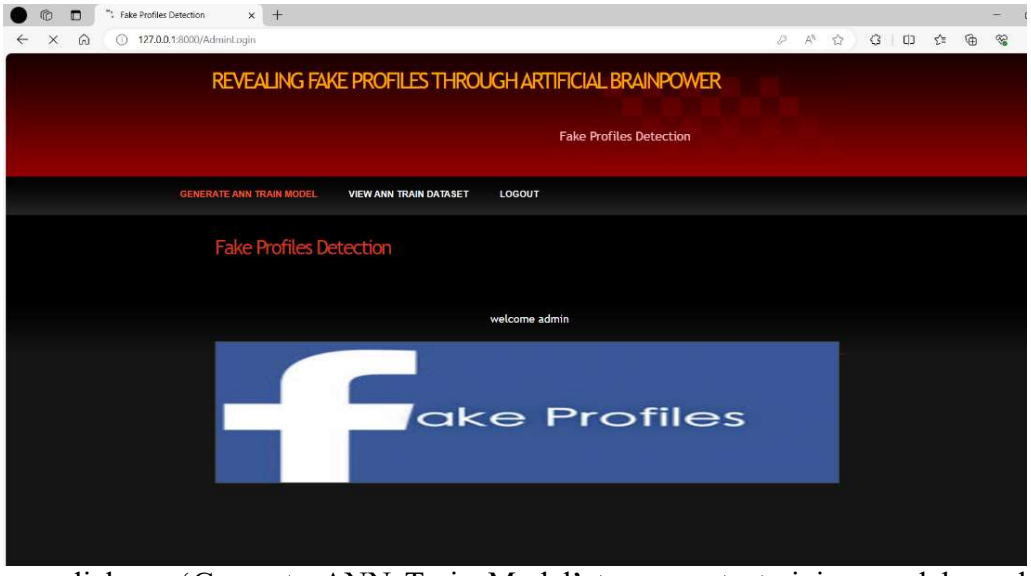
Screen shots

Deploy this application on DJANGO server and then run in browser enter URL as 'http://localhost:8000/index.html' to get below screen

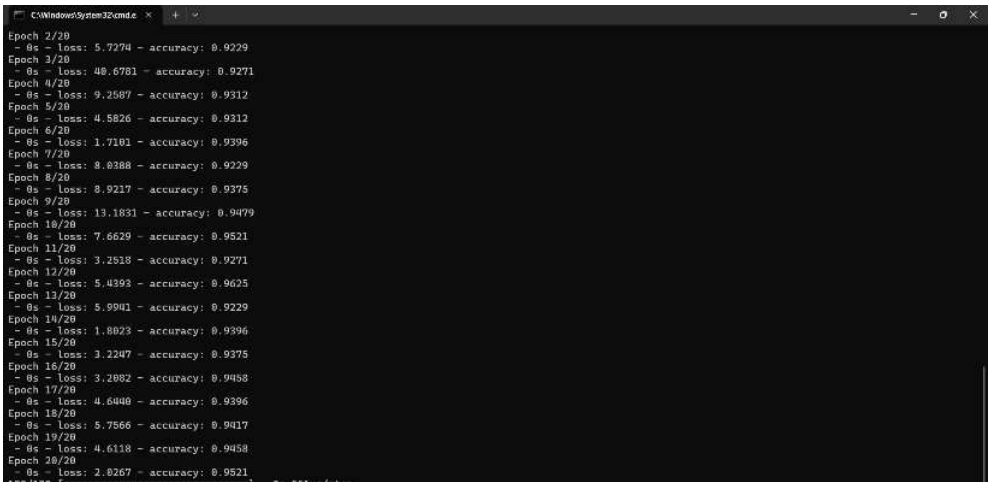
In above screen click on 'ADMIN' link to get below login screen



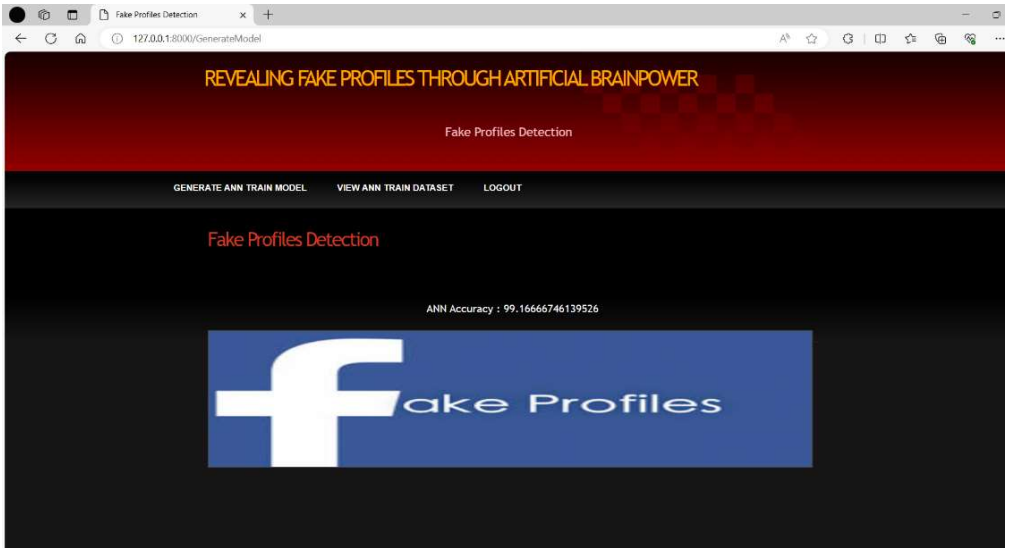
In above screen enter admin and admin as username and password to login as admin. After login will get below screen



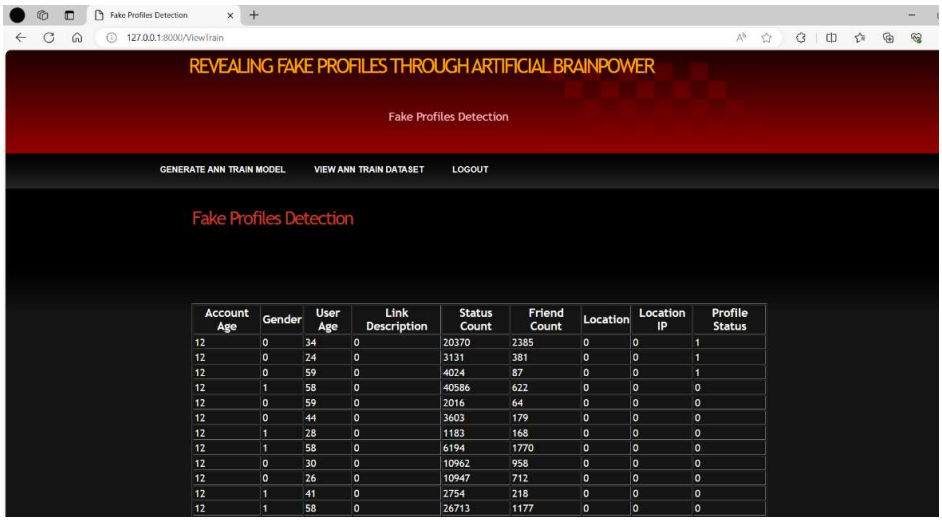
In above screen click on 'Generate ANN Train Model' to generate training model on dataset. After clicking on that link you can see server console to check ANN processing details with accuracy



In above black console we can see all ANN details.

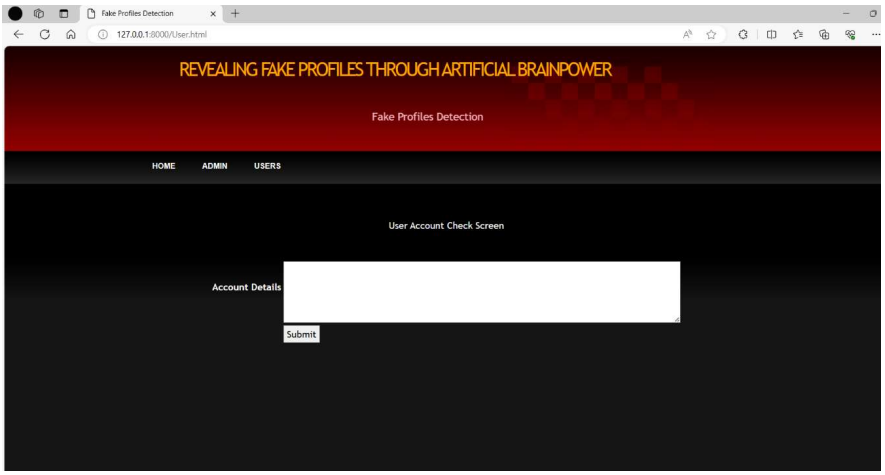


In above screen we can see ANN got 98% accuracy to train all Facebook profile. Now click on 'View Ann Train Dataset' link to view all dataset details



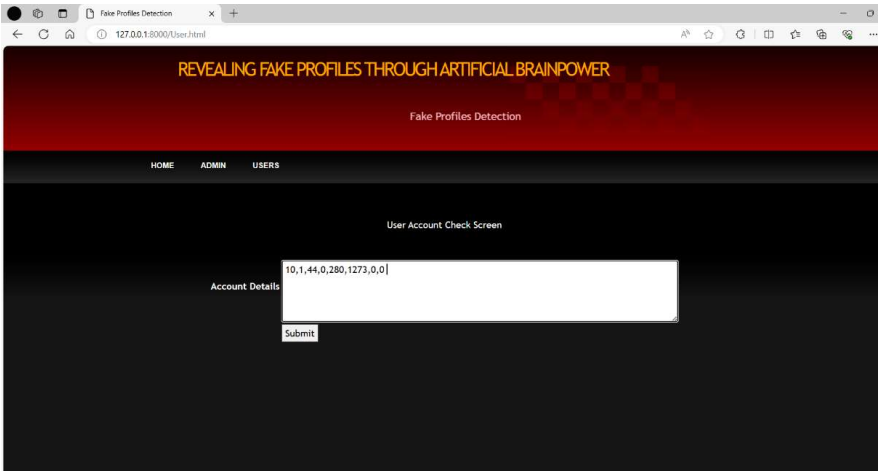
Account Age	Gender	User Age	Link Description	Status Count	Friend Count	Location	Location IP	Profile Status
12	0	34	0	20370	2385	0	0	1
12	0	24	0	3131	381	0	0	1
12	0	59	0	4024	87	0	0	1
12	1	58	0	40586	622	0	0	0
12	0	59	0	2016	64	0	0	0
12	0	44	0	3603	179	0	0	0
12	1	28	0	1183	168	0	0	0
12	1	58	0	6194	1770	0	0	0
12	0	30	0	10962	958	0	0	0
12	0	26	0	10947	712	0	0	0
12	1	41	0	2754	218	0	0	0
12	1	58	0	26713	1177	0	0	0

In above screen we can see all train data and scroll down to view all records. Now ANN train model is ready and you can logout and click on 'User' link to get below screen.

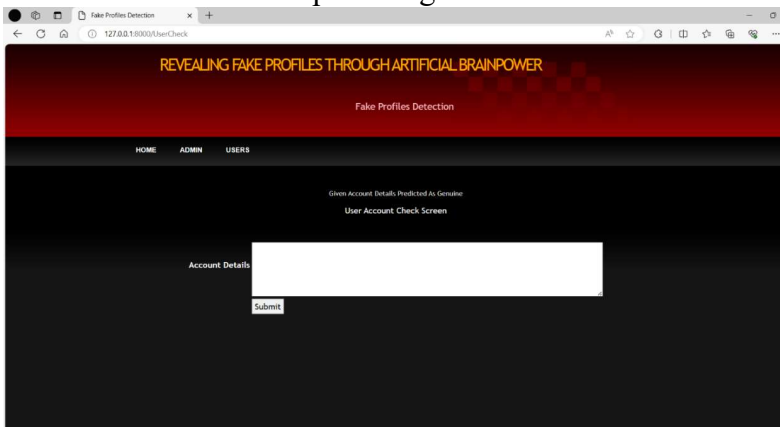


In above screen enter some test account details to get prediction/identification from ANN. You can use below records to check

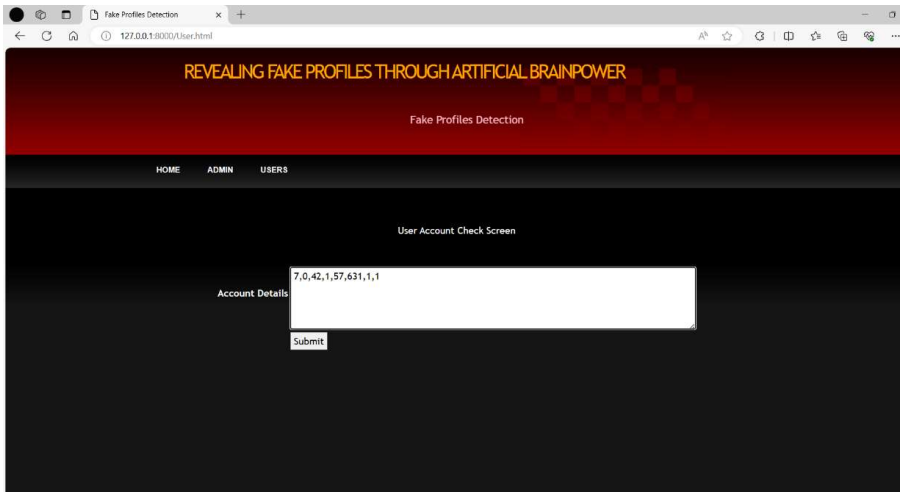
10,1,44,0,280,1273,0, 0
10,0,54,0,5237,241,0,0
7,0,42,1,57,631,1,1
7,1,56,1,66,623,1,1



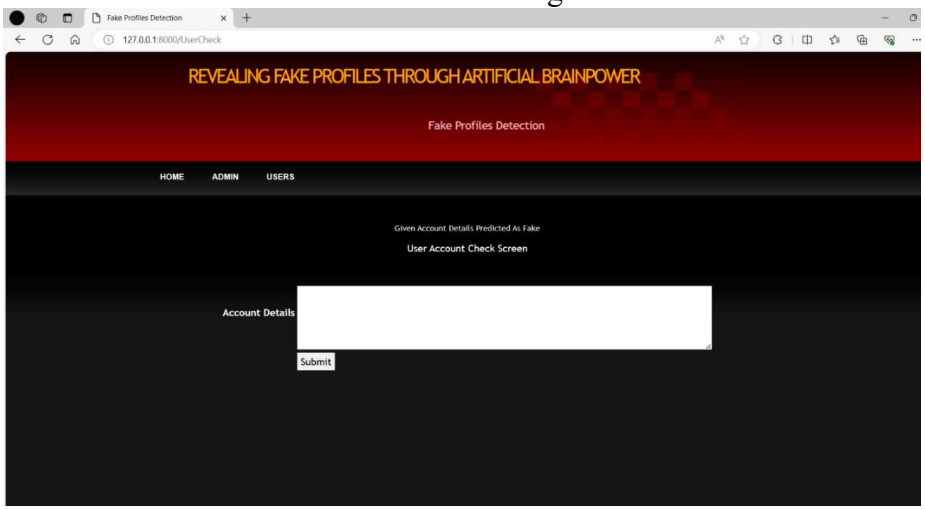
For above input will get below result



In above screen we can see the result predicted as genuine account



For above account details we got below result



In above screen we got result as fake for given account data

CONCLUSION

we use machine learning, namely an artificial neural network to determine what are the chances that a friend request is authentic are or not. Each equation at each neuron (node) is put through a Sigmoid function. We use a training data set by Facebook or other social networks. This would allow the presented deep learning algorithm to learn the patterns of bot behavior by back propagation, minimizing the final cost function and adjusting each neuron's weight and bias.

SCOPE FOR FUTUREWORK

Each input neuron would be a different, previously chosen feature of each profile converted into a numerical value (e.g., gender as a binary number, female 0 and male 1) and if needed, divided by an arbitrary number (e.g., age is always divided by 100) to minimize one feature having more influence on the result than the other. The neurons represent nodes. Each node would be responsible for exactly one decision-making process

Future advancements in identifying fake profiles through artificial neural networks (ANNs) will focus on multimodal learning and advanced feature extraction for enhanced accuracy. Real-time detection capabilities and robust defense mechanisms against adversarial attacks will ensure efficient and reliable perform

BIBLIOGRAPHY

Code snippets for any errors http://stackoverflow.com/
Android Development Guide https://www.udemy.com/android
Xml and Layout Guide https://www.androidhive.com/
Connecting to Firebase Docs https://firebase.google.com
Software Testing http://en.wikipedia.org/wiki/Software_testing
Manual Testing http://en.wikipedia.org/wiki/Manual_testing
Performance Testing http://en.wikipedia.org/wiki/Software_performance_testing

