Research Article

A novel Artificial Neural Networks and FCNN algorithms for identifications of Fake Profiles

[1]**P. Deepthi**, Asst. Prof, dept of CSE, Malla Reddy Institute of Technology, Kompally, Maisammaguda, Dulapally, Secunderabad, 500100

[2]**K. Bhavani**, Asst. Prof, dept of CSE, Malla Reddy Institute of Technology, Kompally, Maisammaguda, Dulapally, Secunderabad, 500100

[3]**B. Tirupathi Kumar**, Asst. Prof, dept of CSE, Malla Reddy Institute of Technology, Kompally, Maisammaguda, Dulapally, Secunderabad, 500100

ABSTRACT

Fake accounts in online networks are becoming more difficult to spot, and even the most fundamental tactics aren't always reliable. Consequently, I was prompted to create the suggested research using an Artificial Neural Networks (ANN) module from deep learning in order to identify if provided record peculiarities are from certified or counterfeit customers. According to previous customer records, this ANN will be computed based on that information from a given date. We'll create this ANN calculation using all of our previous customers' false profile data and verified record information, and we'll use it on all fresh test data to see whether the new record subtleties we get are from genuine or fraudulent clients. It is possible for malicious clients to hack into online interpersonal organisations, such as Facebook or Twitter, in order to steal or get access to a user's private information. Fake accounts on social media sites may be detected using an ANN model, which helps us protect our customers' data.

Key Words:

Artificial Neural Network, Deep Learning, Fake Profiles, Malignant Clients, Facebook, Twitter, Online Interpersonal Organization.

1. INTRODUCTION

In 2017 Facebook arrived at a complete populace of 2.46 billion clients settling on it the most well-known decision of online media [1]. Online social networks make incomes from the information given by clients. The normal client doesn't realize that their privileges are surrendered the second they utilize the web-based media networks administration. Online media organizations have a great deal to acquire to the detriment of the client. Each time a client shares another area, new photographs, likes, aversions, and label different clients in content posted. Facebook makes income by means of ads and information. All the more explicitly, the normal American client produces about $26.76 per quarter [2]. That number adds up rapidly when a great many clients are involved.

In todays advanced age, the always expanding reliance on PC innovation has left the normal resident powerless against violations, for example, information breaks and conceivable wholesale fraud. These assaults can happen without notice and frequently without notice to the survivors of an information break. As of now, there is minimal motivating force for informal communities to further develop their information security. These breaks regularly target web-based media organizations like Facebook and Twitter. They can likewise target banks and other monetary establishments.

In todays advanced age, the always expanding reliance on PC innovation has left the normal resident powerless against violations, for example, information breaks and conceivable wholesale fraud. These assaults can happen without notice and frequently without notice to the survivors of an information break. As of now, there is minimal motivating force for informal communities to further develop their information security. These breaks regularly target web-based media organizations like Facebook and Twitter. They can likewise target banks and other monetary establishments. There is by all accounts a newsworthy issue including online media networks getting hacked each day. As of late, Facebook had an information break which influenced around 50 million clients [3]. Facebook gives a bunch of unmistakably characterized arrangements that clarify how they manage the user's information [4].

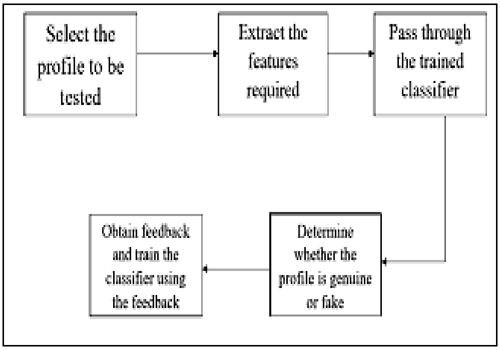

Select the profile to be tested
Extract the features required
Pass through the trained classifier
Determine whether the profile is genuine or fake
Obtain feedback and train the classifier using the feedback

Figure 1. Represent the Flow of Fake Profile Prediction

The approach does very little to forestall the consistent double-dealing of safety and security. Counterfeit profiles appear to fall through Facebook's worked in security highlights. A different risk of individual information being acquired for deceitful intentions is the presence of bots and phony profiles. Bots are programs that can accumulate data about the client without the client in any event, knowing. This interaction is known as web scratching. What is more awful, is that this activity is lawful. Bots can be covered up or come as a phony companion demand on an interpersonal organization site to access private data. The arrangement introduced in this paper expects to zero in on the risks of a bot as a phony profile on your web-based media. This arrangement would come as a calculation. The language that we decided to utilize is Python and by using python programming language and utilizing all the libraries we are going to show the performance of our proposed method in efficient way.

2. LITERATURE SURVEY

The most important phase in the software development process is the literature review. You must first determine how long it will take to construct the new application or model as well as how much it will cost. To begin creating the application, we need to ensure that all of these components are in place and have been approved.

1) Counterfeit Profile Identification in Online Social Networks

Many new inventions are being created every day. The mobile phone is becoming better and better. Keeping in touch with old friends and creating new ones has become easier thanks to online informal groups, which have become a part of everyone's daily routine. However, the proliferation of online systems management has led to a rise in difficulties like as online pantomime and fabricating one's own profile. While riding, customers are bombarded with a lot of useless information that is submitted by fake customers. In informal groups like Facebook, researchers have shown that up to 40% of the profiles are fake. The presence of fake accounts in web-based informal groups leads to the use of systems via arrangement.

2) Utilization of Artificial Neural Networks to Identify Fake Profiles

Artificial intelligence (AI), especially a fake neural network, is used in this article to determine whether or not Facebook buddy demand is legitimate. The classes and libraries are also laid up in this manner. In addition, we discuss the sigmoid capacity and how it is still up in the air and being exploited. Finally, we'll look at the interpersonal organisation page's limits, which are of the utmost importance in the given solution. Bots and fraudulent profiles pose a variety of threats to the integrity of an individual's personal information. Bots are software programmes that can gather information about a user without the user's knowledge. Web scratching is the name given to this repetitive motion. That this is legal just makes things worse. It is possible to conceal bots or to disguise them as a bogus buddy request on a social networking site in order to get access to sensitive information. The system proposed in this research aims to eliminate the dangers of having a bot pose as a real person on your social media accounts. To put it another way, this layout would have been calculated. Python is the language we've chosen to work with. In order to determine if a current online companion demand from a customer is legitimate or a fake companion demand asking for data, the calculator would have the ability to do so. In order to create our model and afterwards verify whether the profiles are fake or real, we would need the aid of online media companies to make this computation work. The computation might potentially be included into the user's web browser as a standard feature.

3) Recognition of phony cash note utilizing convolutional neural networks (2016).

Records in internet-based web-based media provide a wealth of information, including names, sexual orientation, coworkers, enthusiasts, proclivities, and geographic coordinates. This information is both public and private in half of the time. As private information is unavailable, we must rely on public data to identify phoney profiles for relational connection. Regardless, if our recommended strategy is adopted by the associations of relational cooperation, they will be able to exploit the private information of their customers to avoid security difficulties due to improper treatment. To distinguish between real and bogus profiles, a profile's qualities are taken into account. We used the following methods to identify fake profiles: After determining the attributes of a candidate, it is necessary to choose a group of profiles that are either false or real for the order calculation's instructional reason. 1337 fake customers and 1481 real customers were used to create our dataset, which includes numerous attributes such as the number of friends, how many fans rely on the service and what languages are spoken. In the second step, the teaching dataset is arranged in accordance with a set of rules. It benefits from the practise dataset and is expected to provide appropriate tastefulness scores for the testing dataset. There are no test marks left for the informed classifier to use as an assurance. 6.

3. EXISTING SYSTEM

Fake accounts are used by criminals to steal login information from unwary users. Using a false profile, you can reach out to a large number of people with public profiles. False profiles entice naive viewers by displaying photos of persons who are seen as appealing. If the user accepts the friend request, the owner of the fake profile will begin sending friend invitations to everyone the person has added to their social network.

LIMITATIONS OF EXISTING SYSTEM

The following are some of the major drawbacks of the current setup. They are as follows:

1) In most cases, the text of the bogus profile contains links leading to an external website where the harm is done.

2) If an inquisitive person clicks on the faulty link, their computer will be harmed. Installing a rootkit and turning your machine into a zombie may cost you as little as acquiring a virus.

3) Even though Facebook has a stringent screening process, it only takes one false profile to infect the computers of a large number of its users.

4) Sometimes the fake profile not only grabs the users information but also create damage for the personal computers.

4. PROPOSED METHODOLOGY

In our proposed work, we use AI with CNN model in order to find out the fake profiles who try to create malware attempt on genuine user records. In general we try to use MS-Excel to store old and new phony information profiles. Initially we try to calculate the information in starting point and then try to find out the outline information about all users at that point. Now we try to do assortment of information based on user information partitioned into a preparation set and a testing set. For partition of information we require some information collected from online social media to prepare our current model. In our current model we try to assume the fake profiles are gathered from a online social network such as facebook or twitter and we try to concentrate on a phony profile like Account age, Gender, User age, Link in the portrayal, Number of messages conveyed, Number of companion demands conveyed, Entered area, Location by IP, Fake or Not. Each and every field has some importance and we try to figure out information from any filed. For instance, for the sexual orientation boundary if the profile not really set in stone to be a female or male a worth of (1) is relegated to the preparation set for Gender. A similar cycle is applied to different boundaries. We likewise utilize the nation of beginning as a factor From the above figure 2, we can clearly identify there are three layers present in neural network such as: Input Layer, Hidden Layer and Output Layer.

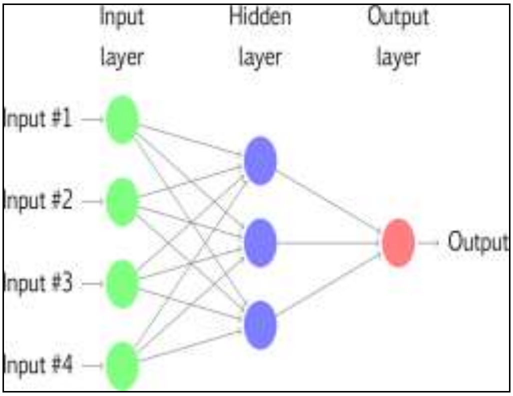


Figure 2. Represent the Architecture flow of CNN

Initially the data or input is collected from input layer and the input will be undergo several pre-processing stages and then enter into hidden layer. In this current model, the input is nothing but several clients profiles are collected and examined to identify the fake profiles and normal profiles. Once the profile is loaded as input now we need to enter into hidden layer. In the hidden layer we can see all the mathematical processing and problem formulation steps. At this stage we try to apply calculations for identifying the fake profiles based on some features. In this stage we will take some features and match the features with previous genuine profiles to identify the similarity of fake and normal profiles.

Now the hidden layer will process the input and then processed information is send as outcome for the end user. This outcome can be displayed on output layer and this will give clear idea about the output.

In order to train the current ANN model, we use the following details or features which are collected from real time social network such as facebook.

1) Account_Age,

2) Gender,

3) User_Age,

4) Link_Desc,

5) Status_Count,

6) Friend_Count,
7) Location,
8) Location_IP, and
9) Status

Normally for every fake user, they will have some intention like send friend requests to normal users to gain illegal access of genuine profiles. So they try to hack the genuine users profile data and collect as much of information they can.Based on some continuous observations of fake profiles, we shortlisted some attributes which are used to find out whether profile is genuine or fake. To train an ANN model, we collected this Facebook profile data from the site and used it to train the model. Some of the profile dataset's values are shown below.

The above dataset contains nearly 9 columns and each and every column name has its own importance and for easy reference they kept in bold. The corresponding values are separated by using comma between each and every value.Here the features what we consider for fake profile identification are holding numerical values, because ANN model will not accept strings as matching parameters.In some cases if there is any string value present in input dataset,those string values are mapped with numerical values by replacing with either zero or one.For example if we take gender as one feature,the gender filed contains two possibilities either male or female. So if the male value is 1 and the female value is 0, we change the gender values to 0 or 1. If the last column in the dataset contains a value of 0, the account is genuine; otherwise, it is false. Fake accounts will always have fewer posts since their primary goal is to send friend requests rather than create content, hence Facebook marks that record as a fake account by studying this characteristic. Our ANN model is trained using the aforementioned dataset, which is stored in the "dataset" subdirectory of the code. Inputing test data and account information after creating a train model will allow an artificial neural network (ANN) to determine if the data is fraudulent or real. The numbers shown below were obtained via the use of test data.

An ANN algorithm will be able to tell us if the test data above is genuine or fake based on the STATUS column and its value. We can see the results of the aforementioned tests in the output.

ANN ALGORITHM

To show how an ANN neural network-based picture classifier may be built, we aim to develop a 6-layer neural network to detect and differentiate each profile from one another. There is no need to worry about memory since we will be building this model on a tiny dataset. Instead of using GPU, we may store and access the network directly in CPU memory. It takes a lot of time to train traditional neural networks that are effective at picture classification on a typical CPU.

Using TENSORFLOW, we want to develop a CNN in this project.

Mathematical modelling is used to tackle an optimization issue using CNN models in general. It is the fundamental unit of computation in neural networks, which is why they are formed of neurons.

Take an input (say x), do some calculation on it (say: multiplying it by w and adding another variable B) to get the output (say, the value b).

$z= wx + b$

The ultimate output (activation) of a neuron is generated by passing this value via a non-linear function termed the activation function (f).

Activation functions come in a wide variety of shapes and sizes. Sigmoid is a well-known activation function. As the name suggests, sigmoid neurons are those that employ the sigmoid function as an activation function. Neurons are called based on their activation functions, and there are a variety of them, such as RELU and TanH. This is the next building element of neural networks, and it is termed a layer when neurons are stacked in one line. Figure 2 shows all of this in great detail.

5. IMPLEMENTATION STAGE

Implementation is a stage where the theoretical design is converted into programmatically manner. Here we try to divide the application into number of modules and then coded for deployement.Here we used python as programming language with tensorflow to show the performance of our current application. The proposed application is mainly divided into 2 modules. They are as follows:

1) Admin Module

2) User Module

5.1 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.

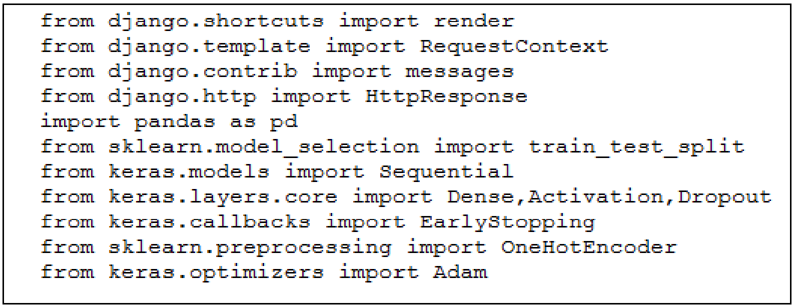
5.2 USER MODULE

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. Here the user module is one which is mainly used to show the similarity between normal profile and fake profile.

6. EXPERIMENTAL RESULTS

Using Python as our programming language, we are attempting to demonstrate the performance of our suggested new application. In order to get the required result, we first need to import all of the relevant libraries and then load the input dataset.

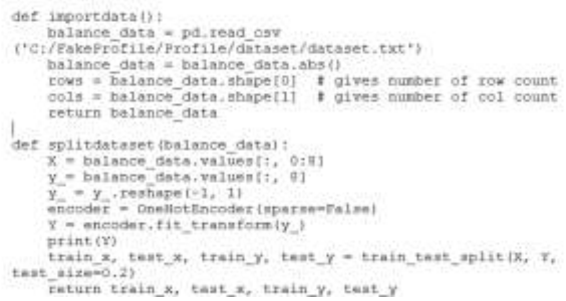
IMPORT LIBRARIES



```
from django.shortcuts import render
from django.template import RequestContext
from django.contrib import messages
from django.http import HttpResponse
import pandas as pd
from sklearn.model_selection import train_test_split
from keras.models import Sequential
from keras.layers.core import Dense,Activation,Dropout
from keras.callbacks import EarlyStopping
from sklearn.preprocessing import OneHotEncoder
from keras.optimizers import Adam
```

Several libraries and packages are plainly visible in the above window in support of the present purpose. As a result, we make an effort to load and import all relevant libraries into our programme.

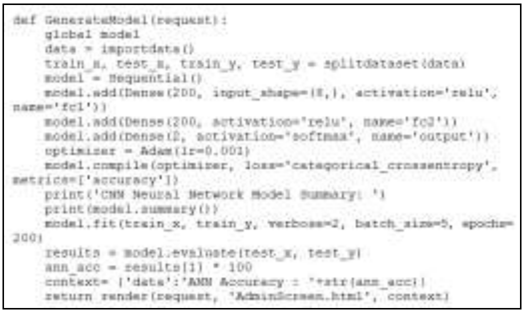
DATA SET PRE-PROCESSING



```
def importdata():
    balance_data = pd.read_csv
('C:/FakeProfile/Profile/dataset/dataset.txt')
    balance_data = balance_data.abs()
    rows = balance_data.shape[0]  # gives number of row count
    cols = balance_data.shape[1]  # gives number of col count
    return balance_data
|
def splitdataset(balance_data):
    X = balance_data.values[:, 0:8]
    y_= balance_data.values[:, 8]
    y_ = y_.reshape(-1, 1)
    encoder = OneHotEncoder(sparse=False)
    Y = encoder.fit_transform(y_)
    print(Y)
    train_x, test_x, train_y, test_y = train_test_split(X, Y,
test_size=0.2)
    return train_x, test_x, train_y, test_y
```

In the above window we can clearly see the dataset is loaded and it is pre-processed and converted into test and train folders. Once the pre-processing step is completed next we can able to test the model on sample user profiles and check how efficiently the current model is able to identify fake profile or not.

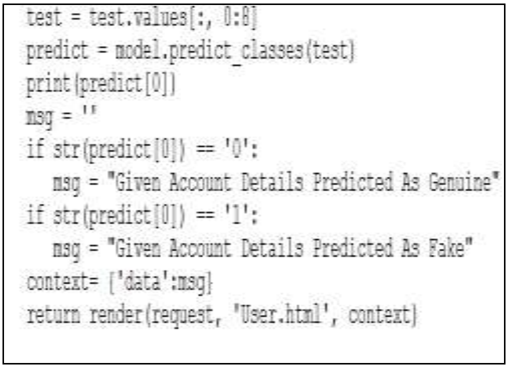
APPLY ANN MODEL



```
def GenerateModel(request):
    global model
    data = importdata()
    train_x, test_x, train_y, test_y = splitdataset(data)
    model = Sequential()
    model.add(Dense(200, input_shape=(8,), activation='relu',
name='fc1'))
    model.add(Dense(200, activation='relu', name='fc2'))
    model.add(Dense(2, activation='softmax', name='output'))
    optimizer = Adam(lr=0.001)
    model.compile(optimizer, loss='categorical_crossentropy',
metrics=['accuracy'])
    print('CNN Neural Network Model Summary: ')
    print(model.summary())
    model.fit(train_x, train_y, verbose=2, batch_size=5, epochs=
200)
    results = model.evaluate(test_x, test_y)
    ann_acc = results[1] * 100
    context= {'data':'ANN Accuracy : '+str(ann_acc)}
    return render(request, 'AdminScreen.html', context)
```

From the above window we can clearly see the ANN model is defined for the proposed module. Here we try to apply ANN model to check the efficiency of our current application in order to check whether user profile is genuine or fake.

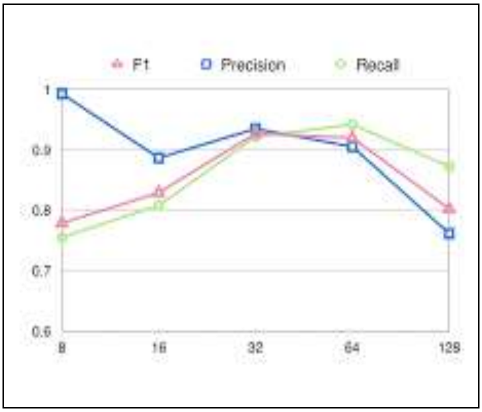
PREDICT THE OUTCOME



```
test = test.values[:, 0:8]
predict = model.predict_classes(test)
print(predict[0])
msg = ''
if str(predict[0]) == '0':
   msg = "Given Account Details Predicted As Genuine"
if str(predict[0]) == '1':
   msg = "Given Account Details Predicted As Fake"
context= {'data':msg}
return render(request, 'User.html', context)
```

From the above window we can clearly test condition is verified on 9 attributes which are applied to check whether given profile is genuine or fake. Here we can able to see two cases for prediction, one is fake profile prediction and other one is normal profile.

GROWTH GRAPH



From the above window we can clearly identify the following growth factors such as Precision, F1 Score, Recall for the given dataset by using ANN model.

7. CONCLUSION

The proposed research is the first time a new method of determining the most accurate analysis of predicted fake profiles in social media has been developed. In general till now there is no accurate method which can able to find out the fake profiles very accurately and efficiently from online social networks and hence a lot of valuable users are lost their sensitive information illegally. Hence this motivated me to do this proposed work by taking

Artificial Neural Network (ANN) model and then try to find out the fake profiles very accurately and efficiently. By conducting various experiments on our proposed model by taking some sample dataset, our comparative results clearly state that our proposed ANN model achieves high level of accuracy in order to predict the fake profiles.

REFERENCES

[1] Sai Pooja, G., Rajarajeswari, P., Yamini Radha, V ., Na vya Krishna.G., Naga Sri Ram.B., Recognition of phony cash note utilizing convolutional neural networks(2016). Global Journal of Innovative Technology and Exploring Engineering, 58-63,8(5).

[2] Mohammed Ali Al-Garadi,Mohammad Rashid Hussain, Henry Friday Nweke, Ihsanali, Ghulammujtaba1, Harunachiro Ma, Hasanalikhattak, Andabdullahgani &quot;Predicti-Ngcyber Bullying On Social Networks.

[3] Yadongzhou,Daewookkim,Junjiezhang,(Member,Ieee),Lili Liu1, Huanjin3, &quot;(IEEE)ProGuard: Detecting Malicious Accounts in SocialNetwork-Based Online Promotions&quot;.

[4] Mauro Conti University of Padua, Radha Poovendran University of Washington, Marco Secchiero University of Padua,&quot;FakeBook: Detecting Fake Profiles in, ACM/IEEE International Conference on Advances in Social Networks Analysis and Mining.

[5] ni .N., Smruthi.M.,&quot;A Hybrid Scheme for Detecting FakeAccounts in Facebook&quot; ISSN: 2277-3878, (IJRTE)International Journal of Recent Technology and Engineering (2019) , Issue-5S3, Volume-7.

[6] NarsimhaGugulothu,JayadevGyani, Srinivas Rao Pulluri &quot;A Comprehensive Model for Detecting Fake Profiles in Online Social Networks(2016)&quot;.

[7] Dr.Narsimha.G, Dr.JayadevGyani, P. Srinivas Rao ,&quot;Fake Profiles Identification in Online Social Networks Using Machine Learning and NLP(2018)&quot;, International Journal of Applied Engineering Research ISSN 0973-4562, Number 6, Volume 13.

[8] Reddy, A. V. N., &amp;Phanikrishna, C. Form following based information extraction and article acknowledgment utilizing profound learning neural

networks(2016). Paper introduced at the Proceedings on second International Conference on Next Generation Computing Technologies in 2016, NGCT 2016, 352-354. doi:10.1109/NGCT.2016.7877440.

[9] V. Rama Krishna,&amp; K.Kanaka Durga. Programmed location of ill-conceived sites with common clustering.(2016) International Journal of Electrical and Computer Engineering, 6(3), 995-1001. doi:10.11591/ijece.v6i3.9878

[10] D.Rajeswara Rao &amp; V.Pellakuri. Preparing and improvement of counterfeit neural organization models: Single layer feedforward and multi-facet feedforward neural network(2016). Diary of Theoretical and Applied Information Technology, 150-156,84(2).

[11] Challa, N.,. Pasupuleti, S. K, &amp; Chandra, J. V. A pragmatic way to deal with E-mail spam channels to shield information from cutting edge tenacious threat.(2016) Paper introduced at the Proceedings of IEEE International Conference on Circuit, Power and Computing Technologies, ICCPCT 2016, doi:10.1109/ICCPCT.2016.7530239.

[12] D.Rajeswara Rao , &amp; P.Vidyullatha. AI procedures on multidimensional bend fitting information dependent on r_ square and chi_square methods(2016). Global Journal of Electrical and Computer Engineering, . doi:10.11591/ijece.v6i3.91556(3), 974-979.

[13] K.Anand., &amp;J.Kumar,Anomaly location in internet based informal community: A review. Paper introduced at the Proceedings of the(2017) International Conference on Inventive Communication and Computational Technologies,ICICCT 2017, 456-459. doi:10.1109/ICICCT.2017.7975239

[14] Pradeepini, G., Patil, S. T. ,&amp; Bangare, S. L (2017). Mind cancer order utilizing blended technique approach. Paper introduced at the International Conference on Information Communication and Embedded Systems, ICICES, doi:10.1109/ICICES.2017.8070748

[15] Jaya Lakshmi, R., &amp;Subba Rao, G. V. A re-helpful calculation to further develop picture recuperation in compacted detecting. Diary of Theoretical and Applied Information Technology (2017), 95(20), 5443-5453.