ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

Credit Card Fraud Detection Using Machine Learning

Hardik Manek[1], Sujai Jain[2], Nikhil Kataria[3], Chitra Bhole[4]

U.G. Student, Department of Computer Engineering, K.J. Somaiya Institute of Engineering and Information Technology, Mumbai, Maharashtra, India[1]

U.G. Student, Department of Computer Engineering, K.J. Somaiya Institute of Engineering and Information Technology, Mumbai, Maharashtra, India[2]

U.G. Student, Department of Computer Engineering, K.J. Somaiya Institute of Engineering and Information Technology, Mumbai, Maharashtra, India[3]

Assistant Professor, Department of Computer Engineering, K.J. Somaiya Institute of Engineering and Information Technology, Mumbai, Maharashtra, India[4]

ABSTRACT: The emergence of new technologies and the Internet is making transactions easier. But for online transactions, you do not have to appear in person somewhere in the transaction, so you are vulnerable to fraudulent attacks. You can impersonate a user and perform fraudulent transactions in a variety of ways. If the transaction is fraudulent, you can determine it by analyzing the previous transaction and comparing it to the current transaction. If the nature of the previous transaction and the current transaction vary considerably, the current transaction may be a fraudulent transaction. Banks and credit card companies use different methods to detect fraud, such as Genetic algorithms, Neural Networks and recent neighborhood algorithms. These methods analyze the client's past behavior and make decisions. In this paper, logistic regression and Autoencoder neural networks will be used to compare and the analysis will be done by implementing these algorithms.

KEYWORDS: Credit Card Fraud, Cashless Transactions, Genetic Algorithm, Neural Network, Logistic Regression.

I. INTRODUCTION

As described in our review paper [10] the payment card industry has seen an upward growth in recent years. Consumers can buy whatever they want just by sitting at home . The expansion is a big step forward for efficiency, with regard to accessibility and profitability, but also some drawbacks. Evolution is accompanied by greater vulnerability to threats. The problem of doing business through the Internet lies in the fact that neither the card nor the cardholder. You must be present at the point of sale. So it is impossible for the trader to check if the customer. It is the actual holder of the card or not. Financial institutions have attention focused on recent computational methods for managing the problem of credit card fraud. Credit card fraud detection is the process of identifying those transactions that are fraudulent in two types of legitimate (genuine) and fraudulent transactions. Credit card fraud detection is based on the analysis of a card's spending behaviour. The features are extracted and transformed from raw data as it is given to train model. This document compares different techniques.Features are extracted and transformed from raw data while giving it to train model [2]. Many techniques have been applied to credit card fraud detection,machine learning algorithms[4], artificial neural network, genetic algorithm, support vector machine, frequent item set mining, decision tree, migrating birds optimization algorithm. Bayesian performance and the neural network is evaluated in data on credit card fraud. Decision tree.We have demonstrated Autoencoder neural networks and logistic regression for fraud detection.

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

II. LITERATURE REVIEW

Ghosh and Reilly [2] used three-layer neural networks to detect fraud in 1994. The neural network was trained on examples of stolen card fraud, application fraud, counterfeit fraud, unreceived fraud issues (NRIs) and post-fraud orders.

An approach is proposed towards fraud detection in banking transactions in [4] using fuzzy clustering and neural network. In this approach, fraud detection is performed in three phases. The first step is to authenticate the initial user and verify card details. A blurry half - clustering algorithm is performed after completion of this operation to detect the behavior of normal user use based on past transactions.If a new transaction at this stage turns out to be uncertain, based on a neural network to determine whether it was actually a fraudulent transaction or whether the mechanism applies.

Kang Fu, Dawei Cheng, Yi Tu, and Liqing Zhang at [5] proposed a convolutional based neural network approach (CNN) to find fraudulent transactions. Convolutional neural network is a part of deep learning and is a type of advanced neural network composed of more than one hidden layer.In this document, finding more complex models of fraud and improving classification accuracy, a new feature of commercial entropy is proposed. In this document for the first time, CNN is used to detect fraud.used to detect frauds.

The description of the various techniques are as follows:-
1. Credit Card Fraud Detection with a Neural Network by Sushmito Ghosh (IEEE 1994) [2] with the techniques, feed forward artificial Neural Network and data sets from Mallon bank 450000 transactions to train model.

2. Credit Card Fraud Detection using Convolutional Neural Network [4] by Tanmay kumar Behera (IEEE Computer Society, 2015) with the technique,Fuzzy Clustering and Neural Network using Synthetic data.

3. Credit Card Fraud Detection using Convolutional Neural Network by Kang Fu [5] (Springer, 2016 Convolutional Neural Network, Cost Based Sampling for imbalance data) with the technique Commercial Bank and data sets from 260 million transactions and 4000 fraud.

III. VARIOUS METHODOLOGIES FOR FRAUD DETECTION

1. ARTIFICIAL NEURAL NETWORKS

Artificial Neural Network(ANN) is one among the foremost powerful classifiers to search out hidden patterns among totally different attributes [3]. ANN works same as human brain. ANN consists of assorted layers throughout that initial layer is input layer and last layer is output layer. It should have form of hidden layer or no hidden layer. If Neural network embrace quiet one hidden layer, then it's deep learning every layer has completely different neurons, and each vegetative cell is connected with weighted edges. Output of every vegetative cell may even be a performance of its unit. This perform is named activation perform. Example of assorted activation functions used area unit sigmoid perform, step perform, function, linear perform etc.

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

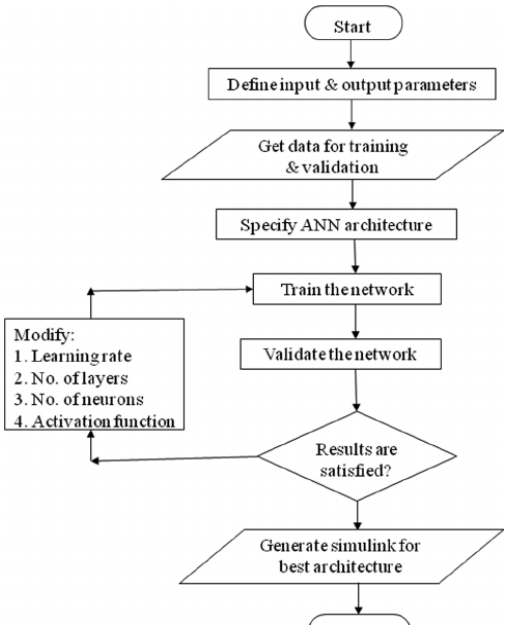

Fig 3.2 Artificial Neural Network

Most used perform is Sigmoid perform among all output layer has same type of neurons as classification label,each vegetative cell of output layer offers likelihood of being that category.In the Figure,four neurons square measure in input layer that creates five picks regarding to importance of input options. Neurons of second layer connected to output layers neurons. Neural network makes an alternative of weights on edges from data given there to for employment and regulate weights.

2. CONVOLUTIONAL NEURAL NETWORK

Convolutional Neural Network (CNN) is also a section of deep learning. Mapping of input into hidden layer represents one feature map each feature map represents one characteristic method of press neurons into feature map is termed convolution as shown in the Figure below.Subsampling reduces parameters of feature map fully connected layer is same as neural network .

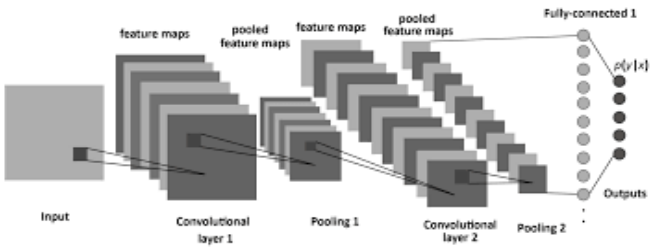

Fig 3.3 Convolutional Neural Network

CNN is with success applied in face recognition, character recognition,image classification,etc. Kang Fu, Dawei Cheng, Yi Tu, and Liqing Zhang[5] at projected a convolutional neural network primarily based approach to find fallacious transactions in mastercard. Input options square measure reworked into feature matrices so reborn into

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

pictures. For finding additional complicated fraud patterns and to enhance classification accuracy, a replacement feature commerce entropy is projected to alleviate the matter of the unbalanced dataset, they used price primarily based sampling technique to get completely different range of artificial frauds to coach the model. They applied CNN model as a result of it's appropriate for coaching giant size of information and CNN has mechanism to avoid overfitting.

IV. SYSTEM DESIGN

The data has been extracted from Kaggle.com. This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.[11]

There are total 30 columns out of which 2 features have been extracted using Principal Component Analysis (PCA). These features are considered for training the model. The system design follows two approach viz. Logistic Regression and Autoencoder neural network.

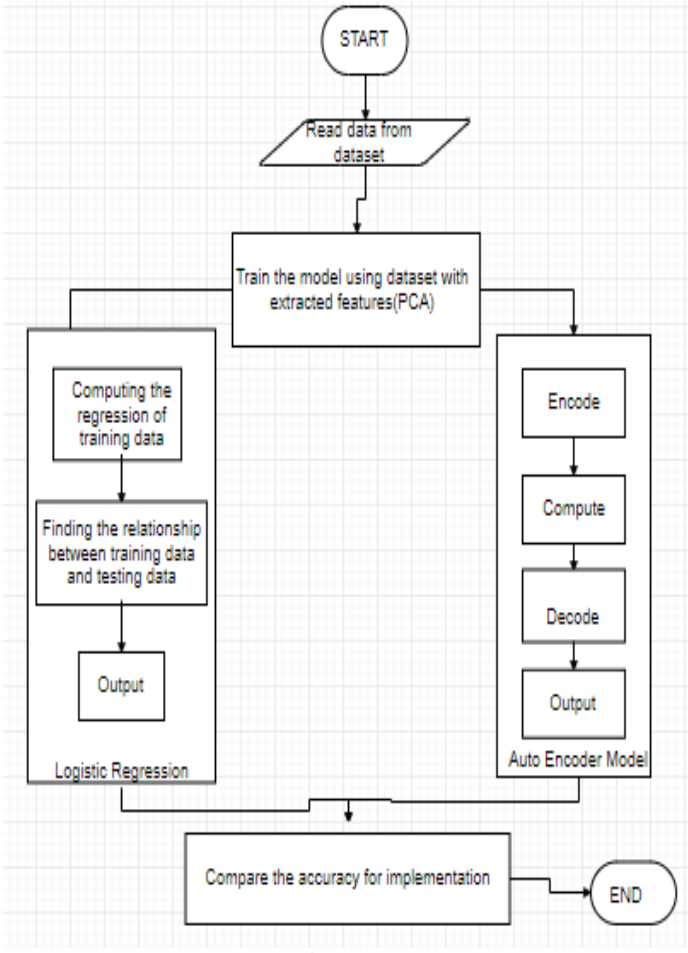

Fig 4.1 System Design

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

V. IMPLEMENTATION

1. LOGISTIC REGRESSION

In this study [3], classification models based on Artificial Neural Networks (ANN) and Logistic Regression (LR) are developed and applied on credit card fraud detection problem. This study is one of the first to output performance measure of Logistic Regression in credit card fraud detection with a real data set.

Logistic regression is that the most famed machine learning calculation following Linear regression. Logistic regression and Linear regression are comparative from multiple views. Still, the largest distinction is their use is Linear regression calculations are accustomed predict conjecture values, however Logistic regression is employed for classification assignments.

There are a plethora of arrangement assignments done habitually by people as an example, transcription whether or not associate degree email could be a spam or not, characterizing whether or not a tumour is harmful or kind, ordering whether or not a website is pretend or not, and so on. These square measure run of the mill precedents wherever machine learning calculations will create our lives considerably less complicated a particularly basic and valuable calculation for characterization is that the logistical Regression calculation.

Sigmoid Function (Logistic Function) Calculated relapse calculation likewise utilizes a direct condition with free indicators to anticipate an esteem. The anticipated esteem can be any place between negative in term in ability to positive vastness. We require the yield of the calculation to be class variable, i.e 0-no, 1-yes. Along these lines, we are squashing the yield of the straight condition into a scope of [0,1]. To squash the anticipated an incentive somewhere in the range of 0 and 1, we utilize the sigmoid capacity.

$$z = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots.$$

$$h = g(z) = 1/(1 + e^{-z})$$

We take the output(z) of the straight condition and provide for the capacity g(x) which restores a squashed esteem h, the esteem h will lie in the scope of 0 to 1. To see how sigmoid capacity squashes the qualities inside the range, how about we imagine the chart of the sigmoid capacity.

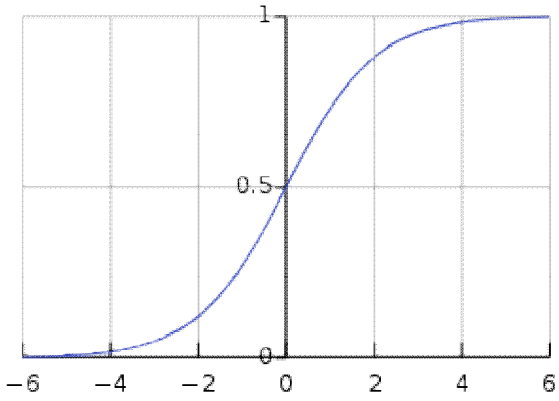

Fig 5.1 The Sigmoid Function

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

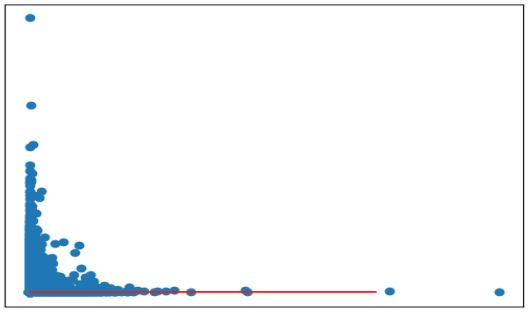

Fig 5.2 The outlier distinguishes the transactions into legitimate and non-legitimate. The clusters formed at bottom left corner depicts the non-fraudulent transactions and data points at the far end highlights the potential fraudulent transactions.

```
PS C:\Users\jansu\Desktop\fraud> python script.py
C:\Python\Python37-32\lib\site-packages\sklearn\externals\joblib\externals\cloudpickle\clo
udpickle.py:47: DeprecationWarning: the imp module is deprecated in favour of importlib; s
ee the module's documentation for alternative uses
  import imp
C:\Python\Python37-32\lib\site-packages\sklearn\linear_model\logistic.py:432: FutureWarnin
g: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this war
ning.
  FutureWarning)
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     56859
           1       0.58      0.18      0.28       103

   micro avg       1.00      1.00      1.00     56962
   macro avg       0.79      0.59      0.64     56962
weighted avg       1.00      1.00      1.00     56962
```

Fig 5.3 Precision achieved using Logistic Regression

2. AUTOENCODER NEURAL NETWORK

Autoencoders are a sort of Neural Network that is employed to find out information codings in an unsupervised manner. The most practicality of the encoder model is to search out an acceptable technique to encode the information such decoded data are equal to input. Autoencoder model consists of three main layers that are the input layer, output layer and one or additional hidden layer connecting them,although the quantity of nodes within the output layer are same as that of input layer.

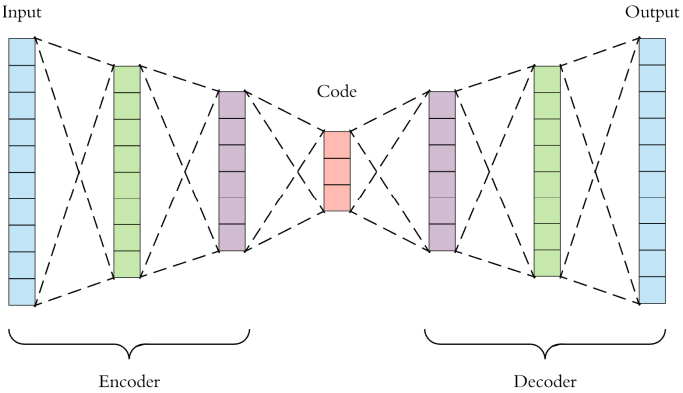

Fig 5.4 Autoencoder Model

Figure Structure of an Autoencoder Model with 3 connected hidden layers which encodes input x and fives output x(Z) f W,b(x) = x The Reconstruction error is minimized by using traditional squared error given by: $L(x, x) = ||xx||$.

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

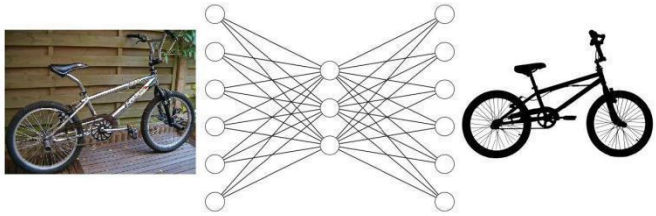

Fig 5.5 Demonstration of Autoencoder Model

A Simple demonstration of Autoencoder model is given in above Figure in which first part that is input layer and hidden layer encodes a bicycle and subsequently, the 2nd part i.e. outer layer and hidden layer decodes to achieve similar bicycle as the output.

Mohamad Zamini, and Gholamali Montazer [8] have implemented an autoencoder with three hidden layer and a k-means clustering has been used and tested on 284807 transactions from European banks. Based on the results, the accuracy of this method was 98.9%, as well as 81% TPR which outperforms in comparison with others.

The below implementation is of our autoencoder model:

```
x = Sys.time()
trctrl <- trainControl(method = "repeatedcv", number = 5, repeats = 2)
model <- train(as.factor(Class) ~ .,
        data=training,
        method="dnn",
        preProcess = c("center", "scale"),
        tuneGrid=expand.grid(layer1=3,layer2=2,layer3=0,hidden_dropout=0,visible_drop out = 0))
y = Sys.time()
time = y-x
print(time)
## Time difference of 7.257171 mins
model
## Stacked AutoEncoder Deep Neural Network
## 199366 samples
## 30 predictor
##   2 classes: 'Valid', 'Fraudulent'
## Pre-processing: centered (30), scaled (30)
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 199366, 199366, 199366, 199366, 199366, 199366, ...
## Resampling results:
##   Accuracy   Kappa
##   0.9982553  0
predicts <- predict(model,newdata = testing)
confusionMatrix(predicts,testing$Class)
## Confusion Matrix and Statistics
```

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

```
##          Reference
## Prediction   Valid Fraudulent
##   Valid   85294     147
##   Fraudulent 0       0
##              Accuracy : 0.9983
##                95% CI : (0.998, 0.9985)
## No Information Rate : 0.9983
## P-Value [Acc > NIR] : 0.5219
##                 Kappa : 0
##  Mcnemar's Test P-Value : <2e-16
##            Sensitivity : 1.0000
##            Specificity : 0.0000
##         Pos Pred Value : 0.9983
##         Neg Pred Value : NaN
##             Prevalence : 0.9983
##         Detection Rate : 0.9983
## Detection Prevalence : 1.0000
##       Balanced Accuracy : 0.5000
```

The accuracy of the model is 99.83%. However, in order to maximize accuracy, the model classifies all the points as Valid since the overall number of available Fraudulent transactions is significantly lower.

VI. CONCLUSION

In this paper, we have executed and compared Logistic Regression and Autoencoder Neural Network for fraud detection in banking transactions. Results supported our implementation, Logistic Regression model helped to attain accuracy of 58%. Since then, the quintessential goal of increasing accuracy was satisfied through the Autoencoder Neural Network, with accuracy of 99.83%. Hence by analyzing the result, it is observed that Autoencoder Neural Network has clearly outperformed the Logistic Regression model. The reason behind the superior performance of Autoencoder Neural Network is it has the capability of self-learning through backpropagation. Autoencoder Neural Network uses the sigmoid function as one of the activation function in the hidden layer which means that Logistic Regression is a subset of Autoencoder Neural Network. There are multiple layers in Autoencoder Neural Network in our implementation there are 4 layers which perform the distinct function of encoding, computing, and decoding. While Logistic Regression is considered as a single layer neural network which alone performs the complete function of binary classification, therefore, resulting in reduced accuracy. Consequently, even though the neural network is computationally more expensive it has the advantage of learning more complex and non-linear functions.

REFERENCES

[1] Alejandro Correa Bahnsen, Djamila Aouada, Aleksandar Stojanovic, Bjurn Ottersten, Feature engineering strategies for credit card fraud detection, 0957-4174/ 2016 Elsevier.
[2] Ghosh, S., Reilly, D.L.: Credit card fraud detection with a neural- network.In: Proceedings of the Twenty-Seventh Hawaii International Conference on System Sciences,1994,vol. 3,IEEE(1994)
[3] “Detecting credit card fraud by ANN and logistic regression”,Y. Sahin , E. Duman, 2011 International Symposium on Innovations in Intelligent Systems and Applications
[4] Tanmay Kumar Behera, Suvasini Panigrahi, Credit Card Fraud Detection: A Hybrid Approach Using Fuzzy Clustering and Neural Network, IEEE Computer Society, 2015
[5] Kang Fu, Dawei Cheng, Yi Tu, and Liqing Zhang, Credit Card Fraud Detection Using Convolutional Neural Networks, Springer International Publishing AG 2016.
[6] Krishna Modi, Bhavesh Oza, Outlier Analysis Approaches in Data Mining, IJIRT vol 3 issue 7.

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

[7] Sam Maes, Karl Tuyls, Bram Vanschoenwinkel, Bernard Manderick, Credit card fraud detection using bayesian and neural network- , International Naiso Congress on Neuro Fuzzy Technology, 2002.
[8] "Credit Card Fraud Detection using autoencoder based clustering",Mohamad Zamini ,Gholamali 2018 9th International Symposium on Telecommunications (IST)
[9] Abhinav Srivastava, Amlan Kundu, Shamik Sural, Senior Member, IEEE , and Arun K. Majumdar, Senior Member, IEEE , Credit Card Fraud Detection Using Hidden Markov Model , IEEE transactions on dependable and secure computing, vol. 5, no. 1, january- march 2008.
[10] Abhinav Shrivastava, Amlan Kundu, Shamik Sural,Senior member IEEE and Arun k Majumdaar, senior member IEEE"Credit card Fraud detection using Hidden Markov Model",IEEE transactions on depend- able and secure computing,vol 5,no 1,january-march-2008
[11] Michael Nielsen(2017,March15), Deep learning available,https://neuralnetworksanddeeple arning.com/chap6.html
[12] Hardik Manek, Sujai Jain, Nikhil Kataria, Chitra Bhole. "Review on Various Methods for Fraud Transaction Detection in Credit Cards," International Journal for Innovative Engineering and Management Research, Vol 7, No. 12, 2018.