International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 07 | July 2022 www.irjet.net p-ISSN: 2395-0072

CREDIT CARD FRAUD DETECTION USING ARTIFICIAL NEURAL NETWORK (ANN) ALGORITHM

Sara Sangeetha E.G[1], Thamarai Selvi.M[2], Sirija.M[3] , Reena.R[4]

[1,2] *Department of Computer Science and Engineering, Prince Shri Venkateshwara Padmavathy Engineering College, Ponmar, Chennai*
[3] *Assistant Professor, Department of Computer Science and Engineering, Prince Shri Venkateshwara Padmavathy Engineering College, Ponmar, Chennai*
[4] *Associate Professor, Department of Computer Science and Engineering, Prince Shri Venkateshwara Padmavathy Engineering College, Ponmar, Chennai*

Abstract - *Due to the rise and rapid growth of E-commerce, use of credit card for online purchases has dramatically increased and it caused an explosion in the credit card fraud. In our project we mainly focused on finding out whether a transaction is fraud or genuine for that we were using a feature totally based only on time i.e.) a particular period of time the occurred transaction data. And extracted feature is given as input for further implementation using Artificial Neural Network (Deep Learning) and Support Vector Machine (Supervised Learning), accuracy of the process is individually collected and compared with one and another. while comparing ANN provides the better accuracy of above 95% while SVM provides only 93% and at last using confusion matrix total number of transactions, number of genuine transactions and number of fraud transactions will be displayed as output based on the data given as input.*

Key Words: Feature Extraction, Fraud Detection, Online Payment, Credit Card, Deep Learning, Artificial Neural Network(ANN),Machine Learning, Support Vector Machine(SVM).

1. INTRODUCTION

The fraud in credit card transaction occurs when the stealer uses the other person card without authorization of the respective person by stealing the necessary information like PIN, password and other credentials with or without the physical card. Using fraud detection module involving machine learning and deep learning, we can find out whether the upcoming transaction is fraud and legitimate.

Machine Learning is the treading and most used technology because of its various applications and less time consumption, more accurate in result. Machine learning is a technology that deals with the algorithm, which provides the computer, a capability to study and advance through experience without being explicitly programmed. Machine learning has application in multiple fields. Example: medical, diagnosis, regression etc.

Machine learning involves the combination of algorithm and statically models which allow computer to perform the task without hard coding then a model is built through a training data and then it is tested on the trained model.

Deep learning is a part of machine learning techniques that makes use of neural networks. Some of methods that come under deep learning are artificial neural network, Convolution neural network, auto encoders, recurrent neural networks, restricted Boltzmann machine etc.

Deep learning makes uses of neural networks, which resembles the human brain in processing the data and making the decision. Here we used both Deep Learning and Machine Learning techniques but Deep Learning Algorithm outperformed based on accuracy. For implementation process we were used “PYTHON” programming language since it is simple and easy to read, learn and write. And also we used some of the python libraries and packages called NUMPY, Pandas, SCIKITLEARN, KERAS, MySQL and TKINTER for data analysis, data manipulation, use of linear algebraic operation, storage purpose, used for Graphical User Interface (GUI).

2. TYPES OF CREDIT CARD FRAUDS

There are different kinds of frauds that are seen on e-commerce sites. Offline theft and robbery occur near ATMs ; while online theft can occur over the internet and mobile phones.

1 Application fraud: Customer's credentials are stolen by the fraudster, then he creates a fake account and the transactions takes place.

2 Electronic or manual card imprints: The fraudster will skim the information that is present on the card and uses the credentials and fraud transaction takes place

3 Card not present: This is a type where actual physical card is not present during transaction

4 Counterfeit card fraud: All the data from magnetic strip will be copied by the fraudster where the real card looks like original card and the same card can be used for fraud

5 Lost/stolen card: This happens due to loosing of the card by the cardholder or by stealing the card from the cardholder.

6 Card id theft: This happens when the id of the cardholder is stolen and fraud takes place.

7 Mail non-received card fraud: While issuing the credit card there will be process of sending a mail to the recipient, fraud can occur here by defrauding the mail or phishing.

8 Account Takeover: The fraudster takes the complete control of the account holder and makes a fraud.

9 Fake fraud in website: A malicious code will be introduced by the fraudster which does their work in the website

10 Merchant collision: The details of the cardholder are shared by the third party or the fraudster by merchants without cardholder authorization.

3. PROBLEM STATEMENT

Now -a-days, most of them are using credit cards for buying the goods which are so much in need but can't afford at the moment. In order to meet the needs, credit cards are used and the fraud associated with it is also increasing. So, there is a need to create and implement a model that's fit well and predicts at higher accuracy.

4. OBJECTIVES

- The main objective of the research is to find a fraudulent transaction in credit card transactions.
- Comparison between the supervised learning and deep learning and deep learning algorithm outperformed based on accuracy.

5. EXISTING SYSTEM

The existing systems are carried out by considering machine learning algorithms like Support Vector Machine, Naïve Bayes, k-Nearest Neighbor and so on and some of them used random dataset. Very few have used artificial neural network for credit card fraud detection.

DISADVANTAGES:

- It produces lot of tables with relatively small number of columns.
- Data restrictions.
- Requires huge processing time.

6. PROPOSED SYSTEM

In the Proposed system we use the Artificial Neural Network to find the fraud in the credit card transactions. Performance is measured and accuracy is calculated based on prediction. And also classification algorithm such as Support vector machine is used to build a credit card fraud detection model. We compared both algorithms and made a decision that artificial neural networks predicts well than support vector machine and gives the outcome of the transaction in either 0 or 1.

ADVANTAGES:

- Provides high accuracy.
- Corrects the problem of overfitting.
- Added layer of safety

7. ARCHITECTURE DIAGRAM

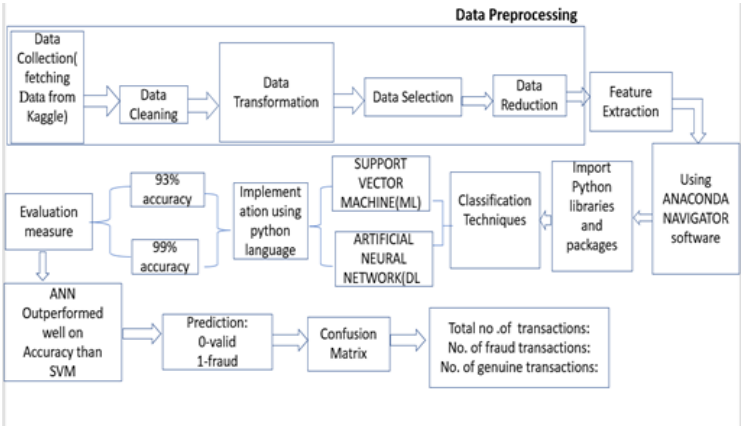


Fig -1

8. MODULES

8.1. DATA PREPROCESSING:

This section explains about the implementation of the algorithm used for proposed system. In this paper, the implementation starts from the collection of data (Data Collection). Then data pre-processing is carried out that includes data cleaning (Filling any missing values in the transaction by using mean, median, standard deviation techniques) and normalizing the data. Dataset is split into two data set as train data and test data and model is trained and tested to measure the accuracy. Finally, system predicts whether transaction is fraud or non-fraud.

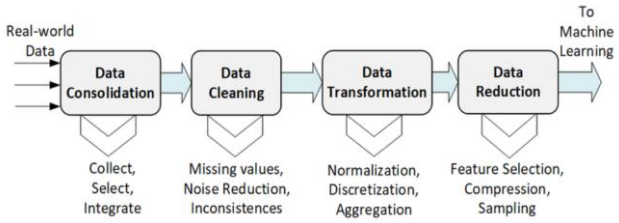


Fig -2

8.2. FEATURE EXTRACTION

- It is the method of reducing the input variable to our model by using only relevant data and getting rid of noise in data.
- Here we use features based only on time i.e.) particular period of time occurred transactional data.

8.3. IMPLEMENTATION:

i) SOFTWARE:

Anaconda Navigator is a desktop graphical user interface (GUI) included in Anaconda® distribution that allows you to launch applications and easily manage CONDA packages, environments and channels without using command-line commands. Navigator can search for packages on Anaconda Cloud or in a local Anaconda Repository.

ii) PACKAGES AND LIBRARIES USED:

Some of the python library and packages used in proposed system are as follows:

1. **NUMPY**

NUMPY is a python library. Abbreviation of NUMPY is numerical python library. NUMPY package is used for multidimensional arrays and linear algebraic operations.

2. **Pandas**

Pandas is a python library. Pandas is used for data analysis and data manipulation tool. It is used to read the dataset and load the dataset. It is fast, flexible when working with data.

3. **SCIKITLEARN**

A python package which is suitable for statistical model and machine learning models. A best suited python package for machine learning modeling.

4. **KERAS**

KERAS is advanced stage of neural network application programming interface (API). It is able of run on top of tensor flow. KERAS is mainly used while implementing deep learning algorithms such as CNN, RNN because its user friendly, modularity, and easy to extensibility. It runs on both CPU and GPU. In the experiment of finding the fraud or non-fraud credit card transaction we had used KERAS along with backend running tensor flow. This KERAS along with tenor flow backend makes excellent choice for training neural network architecture.

5. **MySQL**

MySQL is database which is used for storage purpose. In the experiment of fraud identification in card transaction we had used MySQL for storing the user details namely user name, password, email-id and phone number. While entering into application, user needs to register by providing the credential. These credentials are stored in database. Thereafter, user needs to login by giving username and password. The application will validate the login and registered information than user is moved to next window.

6. **TKINTER**

TKINTER is python library which is used for Graphical User interface (GUI). It can be used on both Unix and Windows platform. We can create it by importing TKINTER module then GUI is created and one or more widgets are added finally, called in loop.

iii) PROGRAMMING LANGUAGE USED:

We have used python as programming language. Python is beginner's language, which provides various applications. In recent years, python had set the new trend because it is easy to use, interpreted, object-oriented, high-level, scripting language. It provides rich packages and libraries that used in machine learning.

iv)CLASSIFICATION TECHNIQUES:

The following algorithms are used in implementation:

1. Support Vector Machine
2. Artificial Neural Network

Support Vector Machine(SVM):

Support Vector Machine (SVM) is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well its best suited for classification.

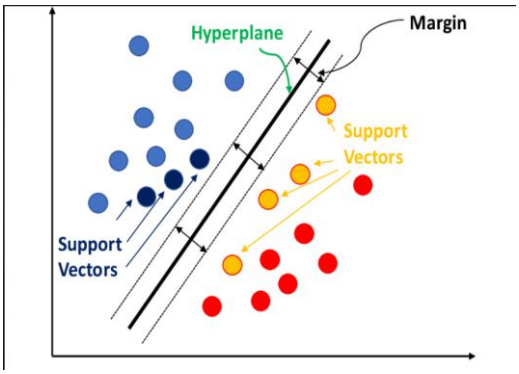


Fig -3

Pseudocode:

• Importing the necessary packages

Example: import pandas as pd

• def SVM

Step 1: Start

Step 2: Reading the dataset. pd.read.csv (file name) # reads the dataset file

Step 3: Data cleaning and data preprocessing

• Resampling the data as normal and fraud class i.e. normal = 0 and fraud =1

• Under sampling of data is done

• Data is scaled (if any null value then eliminated) and normalized.

• Dataset is split into two sets as train data and test data using split()

Step 4: Training the data using the SVM algorithm

• SVM classifier is called as classifier .predict () # which predicts whether the transaction is fraud or not.

Step 5: Calculating the fraud transactions and valid transactions, then calculating the recall, precision and accuracy

Step 6: STOP

Artificial Neural Network (ANN):

Artificial Neural Network ANN is an efficient computing system whose central theme is borrowed from the analogy of biological neural networks. ANNs are also named as "artificial neural systems," or "parallel distributed processing systems," or "connectionist systems." ANN acquires a large collection of units that are interconnected in some pattern to allow communication between the units. These units, also referred to as nodes or neurons, are simple processors which operate in parallel.

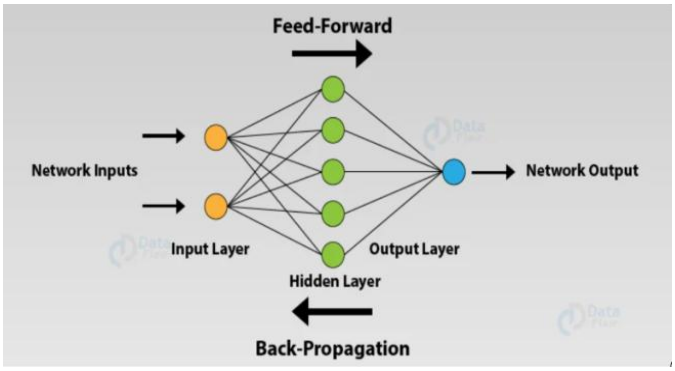


Fig -4

Pseudocode:

This algorithm has two parts namely, Training part and testing part.

Training part:

Def ANN:

Step 1: Start

Step2: Loading and observing the dataset

• pd.read.csv(.csv) # Reads the dataset

• Resampling of data

• Standard Scaler() #scaling and normalization of data

Step 3: Data pre-processing

• Train-test-split() #Splitting of data

Step 4: Training the model

• Dense() #Adding data to activation function

Step 5Analyzing the model

• Prediction of fraud is made and this trained data is stored. It can be used to test (training the model takes longer time so it is stored)

Step 6: Stop

Testing part:

Def ANN

It is carried out similar way only difference is that the stored trained model is used to test the data and classify it.

EVALUATION MEASURE:

The end result is evaluated based on the confusion matrix and precision, recall and accuracy is calculated. It contains two classes: actual class and predicted class. The confusion matrix depends on these features:

True Positive: if both the values are positive that is 1.

True Negative: if both values are negative that is 0.

False Positive: this is the case where true class is 0 and non-true class is 1.

False Negative: It is the case when actual class is 1 and non-true class is 0.

• Precision defined as follows:

Precision = true positive / Actual result

Precision = true positive/(true positive + false positive)

• Recall defined as follows:

Recall = true positive / predicted result

Recall = true positive/(true positive + false negative)

• Accuracy defined as:

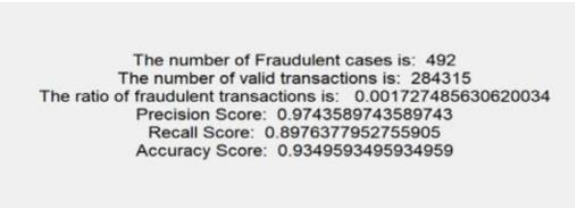
Accuracy = (true positive + true negative)/ total

RESULT:

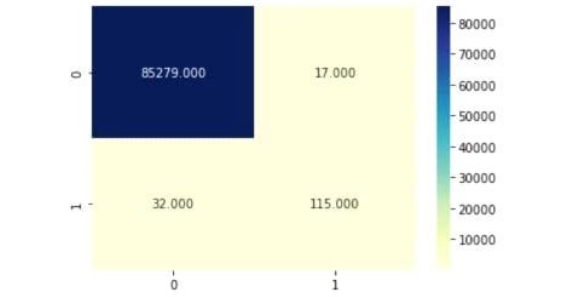
Thus the accuracy of our model is increased by using Artificial Neural Network (ANN) algorithm. This algorithm is best suited to give higher accuracy when compared to all other Machine Learning algorithms.

9. OUTPUT

FOR SVM,



FOR ANN,



Total no. of. transactions: 284807

No. of. genuine transaction: 284315

No. of. fraud transaction:492

The ratio of fraudulent transactions:0.00173047500

Accuracy:99%

Precision score:0.8115942028985508

Recall score:0.9992860737567735

10. OUTPUT

In this research, we have proposed a method to detect the fraud in credit card transactions based on deep learning. We first compare it with machine learning algorithm such as Support vector machine and finally we have used the Artificial Neural Network, which would fit fine to model for detecting a fraud in credit card transactions. In our model, by using an artificial neural network (ANN) which gives accuracy about above 95% is best suited for credit card fraud detection. In this research work, data pre-processing, normalization and under-sampling carried out to overcome the problems faced by using an imbalanced dataset.

11. FUTURE ENHANCEMENT

This model can further be improved with the addition of more algorithms into it. However, the output of these algorithms needs to be in the same format as the others. Once the condition is satisfied, the modules are easy to add as done in the code. This provides a great degree of modularity and versatility to the project.

REFERENCES

[1] Andrew. Y. Ng, Michael. I. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes", Advances in neural information processing systems, vol. 2, pp. 841-848, 2002.

[2] A. Shen, R. Tong, Y. Deng, "Application of classification models on credit card fraud detection", Service Systems

and Service Management 2007 International Conference, pp. 1-4, 2007.

[3] A. C. Bahnsen , A. Stojanovic , D. Aouada , B. Ottersten , "Cost sensitive credit card fraud detection using Bayes minimum risk", Machine Learning and Applications (ICMLA). 2013 12th International Conference, vol. 1, pp. 333-338, 2013.

[4] B. Meena, I. S .L. Sarwani , S. V. S. S. Lakshmi," Web Service mining and its techniques in Web Mining" IJAEGT, Volume 2,Issue 1 , Page No.385-389.

[5] F. N. Ogwueleka , "Data Mining Application in Credit Card Fraud Detection System", Journal of Engineering Science and Technology, vol. 6, no. 3, pp. 311-322, 2011.

[6] G. Singh, R. Gupta, A. Rastogi, M. D. S. Chandel, A. Riyaz, "A Machine Learning Approach for Detection of Fraud based on SVM", International Journal of Scientific Engineering and Technology, vol. 1, no. 3, pp. 194-198, 2012, ISSNISSN: 2277-1581.

[7] K. Chaudhary, B. Mallick, "Credit Card Fraud: The study of its impact and detection techniques", International Journal of Computer Science and Network (IJCSN), vol. 1, no. 4, pp. 31-35, 2012, ISSNISSN: 2277-5420.

[8] M. J. Islam, Q. M. J. Wu, M. Ahmadi, M. A. Sid Ahmed, "Investigating the Performance of Naive-Bayes Classifiers and K-Nearest Neighbor Classifiers", IEEE International Conference on Convergence Information Technology, pp. 1541-1546, 2007.

[9] R. Wheeler, S. Aitken, "Multiple algorithms for fraud detection" in Knowledge-Based Systems, Elsevier, vol. 13, no. 2, pp. 93-99, 2000.

[10] S. Patil, H. Somavanshi, J. Gaikwad, A. Deshmane, R. Badgujar, "Credit Card Fraud Detection Using Decision Tree Induction Algorithm", International Journal of Computer Science and Mobile Computing (IJCSMC), vol. 4, no. 4, pp. 92-95, 2015, ISSNISSN: 2320-088X.