Credit Card Fraud Detection Using Artificial Neural Networks

A.Peter[1]*, K. Manoj[2], P.Kumar[3]

Abstract

People are becoming more used to shopping and paying for things online, which has led to an increase in online payment fraud. Credit cards, debit cards, and net banking are the three primary methods of online payment. Nowadays, more people are using credit cards as a simple method of payment, which leads to an increase in credit card fraud. In this study, we used the Neuralnet R Package to create a neural network-based system to identify credit card fraud. The many parameters that were taken into account during training and testing the neural network are presented in this research. We have also given our experiment's findings and conclusions.

Keywords: Credit Card, Scatter plot, Artificial Neural Network, Confusion Matrix.

DOI Number: 10.14704/Nq.2022.20.17.Nq88098 **Neuroquantology 2022; 20(17): 772-776**

I.INTRODUCTION

When someone uses another person's credit card without that person's permission, credit card fraud occurs. This can happen with or without the physical card when the necessary information, such as the PIN, password, and other credentials, are stolen. We can determine whether or not the upcoming transaction is fraudulent using a fraud detection module that uses machine learning and deep learning. Nowadays, the majority of them use credit cards to buy items they desperately need but cannot currently afford. Credit cards are used to fulfil the needs, and the fraud that goes along with it is growing. A model that fits well and predicts with greater accuracy must be developed and put into use as a result.

Machine learning is the most prevalent and widely utilised technology because of its many uses, quick turnaround times, and reliable outcomes. Machine learning is a field of technology that deals with the algorithms that give computers the ability to learn from experience and develop without being explicitly programmed. The use of machine learning is widespread. Combining algorithms and statistical models in machine learning enables computers to carry out tasks without the need for hard coding. A model is built using training data, and it is then tested using the trained model.

TYPES OF CREDIT CARD FRAUDS

On e-commerce websites, frauds of all stripes can be seen. In a number of methods, including those used by Kang et al. (2016), Ghosh and Reilly (1994), Sahin and Duman (2011), and Nielsen (2017), the researchers were taught with actual data by banks. While online theft can happen via the internet and mobile phones, offline theft and robberies happen close to ATMs.

- **Application fraud:** The fraudster steals the customer's login information before creating a phoney account and doing the transactions.
- **Manual or electronic card imprints:** The fraudster will read the card's information, utilise the credentials, and conduct the fraudulent transaction.
- **Card not present:** This is a transaction where the actual physical card is not present.
- **Fraud involving counterfeit cards:** The fraudster copies all of the magnetic strip data from cards that seem identical to the real thing in order to commit fraud.

***Corresponding Author: -** A.Peter
Address:-[1*,2]Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli, India-627 012.
[3] CITE, Manonmaniam Sundaranar University, Tirunelveli, India-627 012.
Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

- **Lost or stolen card:** This can happen if the cardholder loses it or if someone steals it from them. Card id theft: This happens when the id of the cardholder stolen and fraud takes place.
- **Letter non-received card fraud:** When a credit card issued, a mail will sent to the recipient. Fraud here might happen through mail fraud or phishing.
- **Account Takeover:** In this method, the fraudster seizes total authority over the account holder in order to commit fraud.
- **Website fake fraud:** The fraudster will insert harmful code that performs their actions in the website.
- **Merchant collision:** Without the cardholder's consent, the merchants share the cardholder's information with a third party or a fraudster.

Credit card use has become a more common method of payment because to advancements in e-commerce and communication technology, and transaction fraud is also on the rise, according to Taha and Malebary (2020). They employed the optimised light gradient boosting machine, which combines Bayesian-based hyper-parameter optimization with light gradient boosting machine parameter tuning (LightGBM). They employed two sets of real world, publicly available datasets, including both fraudulent and non-fraudulent transactions, in this method. Their proposed approach outperform-med other techniques in terms of accuracy when compared. Accuracy is 98.40%, area under the receiver operating characteristics curve (AUC) is 92.88%, precision is 97.34%, and the F1-score is 56.95% for the proposed system.

According to studies by Makki et al. (2019), credit card theft results in significant financial losses. The majority of researchers have been working on this to offer a cutting-edge technique to eliminate this loss, but the majority of the approaches that are currently available are expensive, time-consuming, and labor-incentive tasks. After conducting numerous experimental experiments, the authors concluded that the uneven classification of the dataset is the primary cause of the incorrect results. These imbalanced datasets employed in these imbalance classifications led to the model's erroneous predictions and subsequent financial loss. As a result, they discovered that the best algorithms in terms of accuracy, AUCPR, and sensitivity are the LR, C5.0 decision tree method, SVM, and ANN. They have used the balanced dataset in order to train these models.

Several authors, including Jiang et al. (2018), proposed a novel, multi-stage process. The process begins with the collection of cardholder transactions, followed by the aggregated transactions based on behavioural patterns, classification of the dataset, training of the model, and testing of the model. In the event that any anomalous behaviour occurs, the system is given input via a feedback mechanism.

The ratio of credit card fraud to regular transactions is a little bit appropriate, thus Sohony et.al (2018) presented an ensemble learning strategy. They discovered that neural networks and Random Forest work best together to deliver a higher level of accuracy for identifying fraud incidents. They also conducted experiments using significant real-world credit card transactions. Neural networks and Random Forest are combined in ensemble learning. Credit card fraud has gradually increased over the past few years. Machine-learning algorithms are used in a variety of ways to find and stop fraudulent transactions. Two novel data-driven methodologies were introduced, using the best anomaly strategy for credit card transaction fraud. The two methods are the T^2 control chart and choosing the kernel parameters.

According to Sadgali et.al. (2018) research, people today prefer digital and paperless transactions, hence financial transactions like internet, credit card, and mobile ones are becoming more and more common. Millions of transactions were made, and every single one of them was the victim of fraud. Numerous academics have examined, created, and developed the concept for applying machine learning to detect fraud. To determine which model is optimal for fraud detection in card transactions, they presented a comparison of the completely machine-learning methodology. To assess the precision of fraud detection,

Prusti and Rath (2019) created an application using machine learning techniques such as Decision Tree (DT), K-Nearest Algorithm (KNN), Extreme Learning Machine (ELM), Multilayer Perceptron (MLP), and Support Vector Machine (SVM). Through the fusion of the DT, SVM, and KNN approaches, they created a model. For effective data interchange across numerous heterogeneous systems, they employed two web-based protocols called Simple Object Access Protocol (SOAP) and Representational State Transfer (REST). Based on an accuracy parameter, the results of five machine-learning algorithms were compared. SVM outperformed other algorithms by 81.63 percent, while their proposed hybrid system had a higher accuracy rate of 82.58%. Unsupervised credit card fraud detection method employing autoencoder-based clustering was proposed by Zamini and Montazar (2018). When evaluated on a European dataset and compared to other systems, they performed well using three hidden layers and k means for clustering.

The artificial neural network is used by the proposed method to detect fraud in credit card transactions. Based on prediction, performance is evaluated and accuracy is calculated. The dataset used in the experiment consist of 12 and so on and last attribute give the outcome of the transaction in either 0 or 1.

II. METHODOLOGY:

Scatter plots display the relationship between two continuous variables by placing one variable on the x-axis and another on the y-axis. The input variable is on the x-axis of a scatter plot for regression, while the response variable is on the y-axis. You may generally assess if there is a linear link between several variables by using scatterplot matrices. This is especially useful for identifying particular characteristics that might correlate with the credit card dataset.

Ciaburro & Venkateswaran (2017), ANN is a term used to describe a collection of nonlinear statistical modelling techniques that are based on and inspired by the structure of the human brain. ANNs are particularly adapted to the problem of detecting credit card fraud since they may be used to simulate any complex transactional pattern. A neuron is the fundamental building block of a neural network. It accepts several inputs, adds them up, applies a transfer function and then generates the result as either a model prediction or as input to other neurons. A neural network is a structure made up of numerous such neurons that are systematically connected. Feed-forward neural networks, commonly referred to as multilayer perceptron's, are the most widely used neural networks.

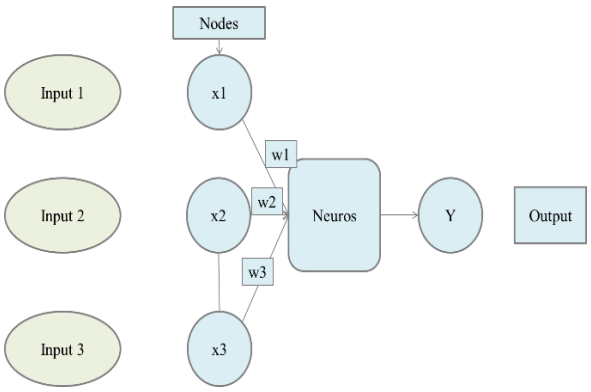

Figure 1: Architecture of Artificial neural network

The biological model for ANN is the human brain. In the human brain, neurons are connected in the same way as nodes are connected in an artificial neural network. The construction of an ANN with input, output, and hidden layers is shown in Fig. 1. The inputs are x_1, x_2,..., x_n and the output is y. The weights for each of the inputs, w_1...w_n. This neural network employs three hidden layers.

Confusion Matrix:

A confusion matrix is a table that is used to describe how well a classification system performs. A confusion matrix depicts and summarises a classification algorithm's performance. Four fundamental properties are comprised of the confusion matrix and used to specify the classifier's measuring metrics. These four numbers are:

- **TP (True Positive):** TP represents the number of frauds who have properly classified to have malignant nodes, meaning they have the fraud.
- **TN (True Negative):** TN represents the number of correctly classified credit card users who are non-fraud.
- **FP (False Positive):** FP represents the number of misclassified customers with the fraud but actually, they are non-fraud. FP is

also known as a *Type I error*.

- **FN (False Negative):** FN represents the number of customers misclassified as non-fraud but actually, they are fraud. FN known as *Type II error*.

Accuracy, precision, recall, and F1 score are performance metrics for algorithms that are determined using the TP, TN, FP, and FN shown in figure 2 below. The ratio of patients who were correctly classified (TP+TN) to all patients (TP+TN+FP+FN) is a measure of an algorithm's accuracy.

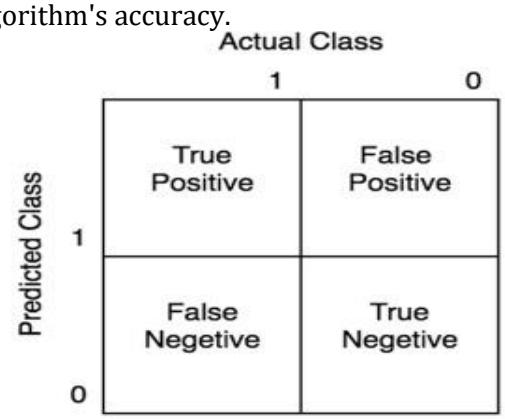

Figure 2: Confusion matrix plot

III.RESULT AND DISCUSSION:
DATA DESCRIPTION

The German Credit card data collected from the **ics.unic** website and the detailed data description explained the table1. There are 26 variables have been collected and importance variables only consider for this analysis. The complete analysis was made by R studio version 4.0.5 with the help of Neuralnet package.

V1	Status of existing checking account	1:0 DM, 2:200 DM, 3:200/Salary Assignment for at least 1 year, A14: no checking account
V3	Credit History	0: no Credits taken/all credits paid back duly, 1:all credits at this bank paid back duly, 2: Existing credits paid back duly till now, 3: Delay in paying off in the past, 4:critical account/other credits existing (not at this bank)
V6	Saving Account/bonds	1: <100DM, 2: 100-500DM, 3:500-1000DM, 4:>1000DM, 5: unknown/no savings account
V7	Present employment since	1: unemployment, 2: <1 year, 3=1-4 years, 4: 4-7 years, 5=>7 years
V9	Personal status and sex	1: Male(divorced/separated), 2: female(divorced/separated), 3: male(single), 4: male(married/widowed), 5: female(single)
V12	Other Debtors/guarantors	1: none, 2: co-applicant, 3:quarantor
V14	Other instalment plans	1: bank, 2: stores, 3:none
V15	Housing	1: rent, 2: own, 3: for free
V16	Number of existing credits at this bank	number of cards
V19	Telephone	1: none, 2: Yes, registered under the customer's name
V20	Foreign worker	1: yes, 2: No
TARGET	Fraud transaction	0-No, 1-Yes
Source: https://archive.ics.uci.edu/ml/machine-learning-databases/statlog/german/		

Table1: data description for German Credit Card Fraud

SCATTER PLOT

Model the data split in terms of training and testing data 80/20 before executing the artificial neural network. The correlation between the variables and scatter plot explained in Figure 3 along with a full histogram. In contrast to the other variables, the correlation between the two variables for housing and credit history is extremely positive. When compared to other variables, the Foreign Workers (V20) and Other Debtors (V12) have a very low correlation.

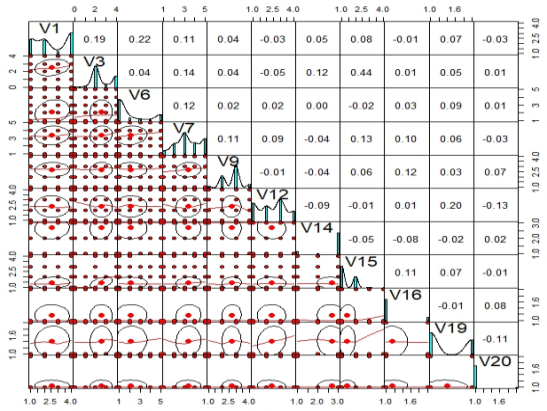

Figure 3: Scatter plot Matrix

ARTIFICIAL NEURAL NETWORK

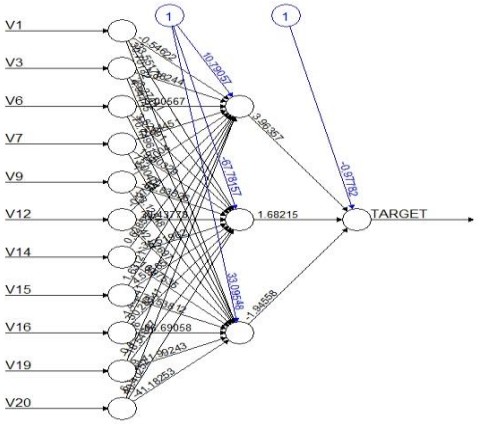

Figure 4: Artificial Neural Network for Credit Card Fraud

The figure 4 represent the left-most nodes (i.e. input nodes) are the raw data variables for German Credit Card data. The arrows in black (and associated numbers) are the weights which say that how much that variable contributes to the next node. The blue lines are the bias weights. The middle nodes (i.e. anything between the input and output nodes) are the hidden nodes. This is where the image analogy helps. Each of these nodes constitutes a component that the network is learning to recognize.

CONFUSION MATRIX

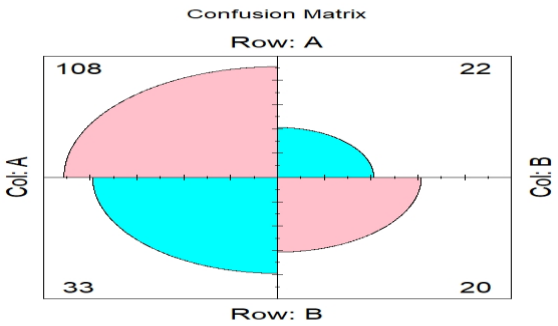

Figure 5: Confusion Matrix

The figure 5, represent the confusion matrix for predicted class for fraud and non-fraud. The left upper and right lower values are the correctly predicted and others are wrongly classified that means true positive, true negative, false positive and false negative.

OVERALL STATISTICS:

Accuracy	0.70
Kappa	0.22
Sensitivity	0.77
Specificity	0.48
Prevalence	0.77
Balance Accuracy	0.63
Mcnemar's Test P value	0.18

Table 2: overall Statistics

The table 2 represent the overall measures for the collected dataset. The ANN model overall model accuracy is 70% and Sensitivity and Prevalence are 0.77. The balance accuracy is 0.63. In machine learning methods more than 70 percentage of accuracy is the acceptable fit. Furthermore, the Artificial Neural Network model was well fitted with the help of three hidden layers and prediction also fitted.

IV. CONCLUSION

The field of employing neural networks to identify credit card fraud is very broad. Our research focuses on the neural model's fundamental application. Results were obtain-ned using a neural network with a single hidden layer and real, albeit sparse, data. By adding more transactions to the training sets and modifying the neural network's design, the results produced can be further optimised. The model correctly fitted and categorised data with a 70% accuracy rate.

REFERENCES:

A.A. Taha, S.J. Malebary, An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine, IEEE Access 8 (2020) 25579–25587.

S. Makki, Z. Assaghir, Y. Taher, R. Haque, M. Hacid, H. Zeineddine, An experimental study with imbalanced classification approaches for credit card fraud detection, IEEE Access 7 (2019) 93010–93022.

C. Jiang, J. Song, G. Liu, L. Zheng, W. Luan, Credit card fraud detection: a novel approach using aggregation strategy and feedback mechanism, IEEE Internet Things J. 5 (5) (Oct. 2018) 3637–3647.

Michael Nielsen (2017, March 15), Deep learning available, https://neuralnetworksanddeeplearning.com/chap6.html.

Sohony, R. Pratap, U. Nambiar, Ensemble learning for credit card fraud detection, in: Proceedings of the ACM India Joint International Conference on Data Science and Management of Data Association for Computing Machinery, New York, NY, USA, 2018, pp. 289–294,

P.H. Tran, K.P. Tran, T.T. Huong, C. Heuchenne, P. HienTran, T.M.H. Le, Real time data-driven approaches for credit card fraud detection, in: Proceedings of the 2018 International Conference on E-Business and Application. Association for Computing Machinery, New York, NY, USA, 2018, pp. 6–9,

Sadgali, N. Sael, F. Benabbou, Fraud detection in credit card transaction using neural networks, in: Proceedings of the 4th International Conference on Smart City Applications (SCA '19). Association for Computing Machinery, New York, NY, USA, 2019, pp. 1–4,

D. Prusti, S.K. Rath, Web service based credit card fraud detection by applying ma- chine learning techniques, in: Proceedings of the TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), Kochi, India, 2019, pp. 492–497.

M. Zamini, G. Montazer, Credit card fraud detection using autoencoders based clustering, in: Proceedings of the 9th International Symposium on Telecommunications (IST), Tehran, Iran, 2018, pp. 486–491,

Kang Fu, Dawei Cheng, Yi Tu, and Liqing Zhang, Credit Card Fraud Detection Using Convolutional Neural Networks, Springer International Publishing AG 2016.

Ghosh and Reilly, "Credit card fraud detection with a neural-network," *1994* Proceedings of the Twenty-Seventh Hawaii International Conference on System Sciences, 1994, pp. 621-630,

Y. Sahin and E. Duman, Detecting Credit Card Fraud by Decision Trees and Support Vector Machines, IMECS vol 1, 2011.

Ciaburro, G., & Venkateswaran, B. (2017). Neural Networks with R: Smart models using CNN, RNN, deep learning, and artificial intelligence principles. Packt Publishing Ltd.

Miller, J. D. (2017). Statistics for Data Science: Leverage the power of statistics for Data Analysis, Classification, Regression, Machine Learning, and Neural Networks. Packt Publishing Ltd.