ISSN No:-2456-2165

Credit Card Duplicity Spotting on Gaining the Knowledge on Machine Learning

T. Hemanth [1], K. Nikhil [2,*], K. Dheeraj [3], Srikanthyadav. M[4]
[1,2,3] UG student, Department of Infomation Technology,
[4] Department of Infomation Technology,
Vignan's Foundation for Science, Technology & Research, Guntur AP-522213, India

Abstract:- With the help of technologies like artificial intelligence (AI), machine learning, big data, blockchain, cloud computing, and IoT the technological revolution is speeding (IoT). There has been a dramatic increase in the number of cyber-attacks and criminal activities as a result of the widespread use of ever-improving internet technologies. Fraudulent use of credit cards is a major concern for the banking sector across the world. Credit card fraud is growing at an alarming rate and has become a major concern, especially as the amount of financial transactions utilising credit cards grows. Here, we've looked at some credit card fraud detection methods that can help protect against a variety of scams. The research problems were also discussed and analysed. For the purpose of detecting credit card fraud, we've deployed six widely-accepted machine learning approaches. A confusion matrix is created for each machine learning approach so that the algorithm's performance may be evaluated. Accuracy, precision, recall, specificity, misclassification and F1 score are used to evaluate its efficacy. Machine learning approaches can be useful in detecting credit card fraud, according to the results. For fraud detection, we propose utilising different machine learning algorithms, even though the findings demonstrate that each algorithm has a high degree of precision and recall.

Keywords:- *Credit Card Fraud, Fraud Detection by Machine Learning, Machine Learning Techniques.*

I. INTRODUCTION

Credit cards are increasingly being used to pay bills and conduct online purchases. As a result of this shift in payment methods, funds are now transmitted via digital means. With these cards, cashless systems have been worsened and cash credit has been relieved at the same time. At the same time, the usage of credit cards is on the rise, and this has led to an increase in credit card fraud. One of the most common types of credit card fraud is perpetrated by people who don't intend to pay back the money they've borrowed. A variety of credit card-related scams are described in detail in this section.

A. Application Fraud

The fraudster creates a phoney user account in order to get access to sensitive information like login and password, and then manipulates the application framework using this account. To make things worse, he steals customer service records.

B. Manual Credit Card Imprints

The magnetic strip on a credit card is used by fraudsters in this sort of scam. There will be falsified transactions in the future because of this knowledge.

C. Original Card Not Exist

When a card is used without the cardholder actually having it in their hands, the fraudster gets access to the account number and expiration date of the card.

D. Mail Non-Receipt Fraud

Every time a consumer applies for a new credit card, the process takes a while. Because of this, fraudsters utilise intercepted delivery to their advantage, changing the user's identity to their own and then making transactions, a practise known as Never Received Issue Fraud.

E. Counterfeit Card Fraud

All the properties of a genuine magnetic swipe card may be found in this replica card. Skimming is a method that may be used to accomplish this. This fictitious card may be used to make purchases because it is completely working.

F. Off track and Stolen Card Fraud

Due to some unknown circumstances, the card holder had their card stolen. If a fraudster manages to obtain the card, he or she may then use it to make purchases. Because internet transactions require a pin number, it's harder, but it can still be done.

G. False Merchant Sites

Phishing is a sort of fraud that resembles this type of scam. When a fraudster creates a false website, it has the appearance of being authentic. Customers will be enticed by the appealing designs and incentives, such as significant discounts, buy one, get one free, etc. Once a transaction has been completed, the cardholder's information is gathered and stored. This might be utilised in the future to conduct fraudulent transactions.

It is possible to identify credit card thefts using historical data by analysing the varied purchase habits of a certain customer. Banks and other credit card issuers can benefit from this data analysis by reducing the amount of money they lose to credit card fraud. When comparing historical data with current purchase habits, a statistical model is needed to identify fraudulent tendencies and alert banks to suspicious activity. This enables banks to quickly identify and eradicate CC frauds by rejecting suspicious transactions, hence reducing the risk of fraud.

Various machine learning approaches are employed in this work for the goal of identifying these scams, and they are used in this study as well. The accuracy of fraud detection is directly tied to the parameters of the evaluation. The values of these factors make it easier to determine whether or not the transaction is legitimate. The major goal of this study is to evaluate the effectiveness of several machine learning algorithms for detecting credit card fraud.

II. RELATED WORK

Various academics have utilised a variety of methods to identify credit card fraud. Some of the noteworthy credit card fraud detection work was reviewed in this section. [1] There are a number of artificial intelligence systems that use the physical unclonable functions [1] to facilitate and protect electronic transactions between distinct entities.

Various classifiers' performance is studied in this research [2] and a summary of the results is provided. [3] employed an ANN-GA hybrid model for credit card fraud detection in this article. ANNs are artificial neural networks, and GAs are genetic algorithms. To classify transactions, a neural network and a genetic algorithm have been deployed, respectively.

In this research, the author [4] describes how a neural network may be used to mine data to detect credit card fraud. Auto-associative architectures are employed in three levels to get the desired results in this project. They trained and tested the system on a set of synthetic data. With this solution, they were able to obtain extraordinarily high fraud detection rates.

The authors of an article [5] suggested a model for real-time fraud detection based on bidirectional neural networks. A model based on an Artificial Immune System was presented by the authors [6] of this research for the identification of credit card theft in the online environment. To ensure that they all performed at the same level, a logistic regression model and three different methods were utilised.

The authors of this study [7] describe a model for detecting credit card fraud based on the principles of a genetic algorithm. In this method, a two-phase synthesising algorithm was devised. Data generation is carried out in the first step before algorithms are applied to detect fraudulent transactions in the second phase.

The authors of this study [8] discuss a genetically programmed fuzzy system for detecting credit card fraud. Using this method, data from actual house insurance claims and credit card transactions may be utilised to generate categorization criteria.

In this study, researchers used a Hidden Markov Model (HMM) to identify credit card thefts with brief false alarms. It is possible to detect fraudulent credit card transactions using neural networks and support vector machines, according to the authors [10]. An RBF neural network with three layers of feed-forward is constructed.

Genetic algorithms (GA) were employed to choose support vectors in a Binary Support Vector System (BSVS) that was shown in this study [11]. A real negative rate was first determined by using the Self-Organizing Map (SOM) method, and then the data was trained using BSVS. An ANN (Artificial Neural Network) and decision tree hybrid was employed in this study [12].

Testing and implementation were carried out using a two-phased method. The dataset was generated in the first step using the classification outputs of decision trees and multilayer perceptrons. Feed this dataset into a multilayer perceptron for data categorization in the second phase. Phase 2: This model has a low percentage of false positives, which means it's dependable. A four-stage Bayesian network for fraud detection was created by the authors of this research [13]. As a result, they concluded that their suggested method would be excessively slow in comparison to other popular algorithms like K-nearest neighbour, neural networks, and regression.

The authors of this research [14] used a two-phase hybrid approach that incorporates neural networks and fuzzy clustering. The c-means clustering algorithm was proposed in phase one. In the second step, they feed suspicious transactions into the neural network to determine if it was fraudulent or not.

In study [15], the authors employed Bayesian networks and ANN, two machine learning approaches, to detect credit card fraud. ANNs were also discussed as a way to speed up Bayesian networks after only a minimal amount of training. Fuzzy logic and neural networks were used to create a system for fraud detection in this research [16]. They found that ANN was 33 percent more accurate than fuzzy logic in terms of precision. Fuzzy logic was utilised to assign a membership characteristic to each piece of data already in the system. The neural network was utilised for results validation.

III. METHODOLOGY

We took these procedures to deploy the machine learning algorithms outlined above. On Kaggle [17], a standard data set is used to evaluate the algorithms. Python has been used to provide a test environment in which various machine learning approaches may be evaluated.

Step 1: Transact with a credit card by loading and reading its transaction data into the system's memory
Step 2: Normalize each value in the dataset to a certain format.
Step 3: Fill in missing values to cleanse the data.
Step 4: Separate the entire dataset into two sections: Test dataset and Train dataset
Step 5: Apply several machine learning techniques and develop models for each methodology.
Step 6: Create the confusion matrix for each strategy
Step 7: utilising the confusion matrix, calculated the values of each technique's assessment parameters
Step 8: Comparing and contrasting the results of the various methods of evaluation

A. Data Set Description

In order to obtain this data, you may go to: https://mlg-ulb/creditcardfraud/home. Due to difficulties with confidentiality, this dataset does not include any background information on the 28 features out of the 30 features. Thus, principal component analysis generates the values of these attributes. Principal components such as 'Amount' and 'Time' have not been turned into features. When fraud occurs, the class is set to 1 and when it does not occur, the class is set to 0.

B. Performance Evaluation Parameters

A confusion matrix is a common tool for assessing the effectiveness of a classification statistical model. The target class number N is used to create a confusion matrix. The N x N matrix provides us with a full view of our model's performance and the kinds of errors it is committing.

According to actual and expected positive and negative class values (see fig. 1), there are four parameters shown. TP and TN refer to the values when the model accurately predicts the positive or negative class. For example, when a model predictions something wrongly in a positive or negative category, it is known as a False Positive (FP).

Any machine learning technique's performance is evaluated using the confusion matrix parameters. Based on assessing the many sorts of values, such as true positives, false negative and the like in the single confusion matrix, these parameters have been developed. Confidence matrix parameters are used to calculate accuracy, recall, and precision, among other things.

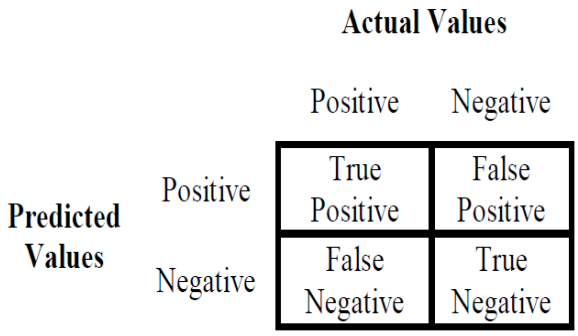

Fig. 1. Confusion Matrix

IV. RESULTS AND DISCUSSION

The outcomes of the machine learning approaches outlined above are examined and demonstrated in this part. Dataset anomalies and how to use them for training and validation of datasets must be understood for performance analysis. Overall, there are 29,415 normal instances and 592 bogus ones in the dataset, with a total of 294807. A parameter value for the confusion matrix of each approach is calculated in table 1 using normal and fraudulent scenarios as examples. The estimated values of the evaluation parameters for each approach are shown in Table 1. The amount of precision achieved by each approach is almost the same.

Table.1. Performance analysis of ML approaches

Technique	Accuracy	Precision	Recall	Specificity	Fl Score
Decision Tree	0.9991	0.9995	0.9995	0.7484	0.9995
Isolation Forest	0.9995	0.9996	0.9999	0.7707	0.9997
K-NN	0.9995	0.9996	0.9999	0.8846	0.9997
Logistic Regression	0.9992	0.9994	0.9998	0.6730	0.9996
Random Forest	0.9993	0.9996	0.9997	0.7799	0.9997
Support Vector Machine	0.9995	0.9996	0.9999	0.7692	0.9997

V. CONCLUSION

For the purpose of detecting credit card fraud, we used six different machine learning approaches, which we tested in this article. A close examination of the various methods reveals that not all of the procedures yielded the same results in all circumstances. The results of fraud detection algorithms are also dependent on the type of dataset that's being analysed. Techniques that yield excellent accuracy but require a lot of time and money to learn are available. When used to huge datasets, certain strategies produce subpar outcomes when applied to small data sets only. Some strategies perform better with pre-processed and sampled data, whereas others perform better with raw, unprocessed data. Another thing to keep in mind when dealing with anomaly identification is that the outlier class of modelling might be useless or even harmful. Focus must be placed on the structure and distribution of typical data. Detecting credit card fraud while it is still in transit requires a system that can detect fraudulent activity and identify it precisely, with as few classification errors as possible.

REFERENCES

[1]. Fragkos, G., Minwalla, C., Plusquellic, J., & Tsiropoulou, E. E. (2020). Artificially Intelligent Electronic Money. IEEE Consumer Electronics Magazine, 10(4), 81-89.
[2]. Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017, October). Credit card fraud detection using machine learning techniques: A comparative analysis. In 2017 International Conference on Computing Networking and Informatics (ICCNI) (pp. 1-9). IEEE.
[3]. Patidar, R., & Sharma, L. (2011). Credit card fraud detection using neural network. International Journal of Soft Computing and Engineering (IJSCE), 1(32-38).
[4]. G Aleskerov, E., Freisleben, B., & Rao, B. (1997, March). Cardwatch: A neural network based da-tabase mining system for credit card fraud detection. In Proceedings of the IEEE/IAFE 1997 computational intelligence for financial engineering (CIFEr) (pp. 220-226). IEEE.
[5]. Krenker, A., Volk, M., Sedlar, U., Bešter, J., & Kos, A. (2009). Bidirectional Artificial Neural Networks for Mobile-Phone Fraud Detection. Etri Journal, 31(1), 92-94.
[6]. Brabazon, A., Cahill, J., Keenan, P., & Walsh, D. (2010, July). Identifying online credit card fraud using artificial immune systems. In IEEE Congress on Evolutionary Computation (pp. 1-7). IEEE.
[7]. RamaKalyani, K., & UmaDevi, D. (2012). Fraud detection of credit card payment system by genetic algorithm. International Journal of Scientific & Engineering Research, 3(7), 1-6.
[8]. Bentley, P. J., Kim, J. W., Jung, G. H., & Choi, J. U. (2000). Fuzzy darwinian detection of credit card fraud. In Proceedings of the Korea Information Processing Society Conference (pp. 277-280). Korea Information Processing Society.
[9]. Bhusari, V., & Patil, S. (2011). Application of hidden markov model in credit card fraud detec-tion. International Journal of Distributed and Parallel Systems, 2(6), 203.
[10]. Ghosh, S., & Reilly, D. L. (1994, January). Credit card fraud detection with a neural-network. In System Sciences, 1994. Proceedings of the Twenty-Seventh Hawaii International Conference on (Vol. 3, pp. 621-630). IEEE.
[11]. Chen, R. C., Chen, T. S., & Lin, C. C. (2006). A new binary support vector system for increasing detection rate of credit card fraud. International Journal of Pattern Recognition and Artificial Intelligence, 20(02), 227-239.
[12]. Keskar, V. (2020). Comparing different models for credit card fraud detection. Journal of Critical Reviews, 7(2), 981-986.
[13]. Ezawa, K. J., & Norton, S. W. (1996). Constructing Bayesian networks to predict uncollectible telecommunications accounts. IEEE Expert, 11(5), 45-51.
[14]. Behera, T. K., & Panigrahi, S. (2015, May). Credit card fraud detection: a hybrid approach using fuzzy clustering & neural network. In 2015 Second International Conference on Advances in Computing and Communication Engineering (pp. 494-499). IEEE.
[15]. Maes, S., Tuyls, K., Vanschoenwinkel, B., & Manderick, B. (2002, January). Credit card fraud detection using Bayesian and neural networks. In Proceedings of the 1st international naiso congress on neuro fuzzy technologies (pp. 261-270).
[16]. Razooqi, T., Khurana, P., Raahemifar, K., & Abhari, A. (2016, April). Credit card fraud detection using fuzzy logic and neural network. In Proceedings of the 19th Communications & Networking Symposium (pp. 1-5).
[17]. Kaggle: Your Machine Learning and Data Science Community. Kaggle.com. (2021). Retrieved 8 May 2021, from https://www.kaggle.com/.