Detecting Credit Card Scams Using Data Science Techniques & Machine Learning

Kishore Kumar L[1], Tamilenban E[2], Mrs A.S.Hepsi Ajibah[3]
[1, 2, 3] Dept of Computer Science And Engineering
[1, 2, 3] GKM college of Engineering and Technology,Chennai, Tamilnadu, India.

Abstract- *Detecting credit card scams is a methodology to detect credit card frauds using machine learning and data science techniques by modelling the dataset. Applying machine learning to the credit card dataset and finding the parameters of the algorithm and calculating the performance metrics.*

Keywords- Credit Card Fraud Detection, Data Science Techniques, Machine Learning (3)

I. INTRODUCTION

A credit card is issued by a bank or financial services company that allows cardholders to borrow funds with which to pay for goods and services with merchants that accept cards for payment. Nowadays as everything is made cyber so there is a chance of misuse of cards and the account holder can lose the money. It is vital that credit card companies are able to identify fraudulent credit card transactions so that customers are not charged for items that they did not purchase. This type of problem can be solved through data science by applying machine learning techniques. It deals with the modelling of the dataset using machine learning with Credit Card Fraud Detection. In machine learning the main key is the data so modelling the past credit card transactions with the data of the ones that turned out to be a fraud. The built model is then used to recognize whether a new transaction is fraudulent or not. The objective is to classify whether the fraud had happened or not. The first step involves analysing and preprocessing data and then applying a machine-learning algorithm to the credit card dataset and finding the parameters of the algorithm and calculating their performance metrics.

II. LITERATURE REVIEW

A literature review is a body of text that aims to review the critical points of current knowledge on and/or methodological approaches to a particular topic. It is a secondary source and discusses published information in a particular subject area and sometimes information in a particular subject area within a certain time period. Its ultimate goal is to bring the reader up to date with current literature on a topic and forms the basis for another goal, such as future research that may be needed in the area and precedes a research proposal and maybe just a simple summary of sources. Usually, it has an organisational pattern and combines both summary and synthesis. A summary is a recap of important information about the source, but a synthesis is a reorganisation or reshuffling of information. It might give a new interpretation of old material or combine new with old interpretations or it might trace the intellectual progression of the field, including major debates. Depending on the situation, the literature review may evaluate the sources and advise the reader on the most pertinent or relevant of them

A. Credit Card Fraud Detection using Machine Learning: A Systematic Literature Review

Companies want to give more and more facilities to their customers. One of these facilities is the online mode of buying goods. Customers now can buy the required goods online but this is also an opportunity for criminals to do fraud. The criminals can theft the information of any cardholder and use it for online purchases until the cardholder contacts the bank to block the card. This paper shows the different algorithms of machine learning that are used for detecting this kind of transaction. The research shows the CCF is the major issue in the financial sector that is increasing with the passage of time. More and more companies are moving towards the online mode that allows the customers to make online transactions. This is an opportunity for criminals to theft the information or cards of other persons to make online transactions. The most popular techniques that are used to theft credit card information are phishing and Trojan. So a fraud detection system is needed to detect such activities.

B. A Research on Credit Card Fraudulent Detection System

Nowadays credit cards are more popular among private and public employees. By using the credit card, the users purchase the consumable durable products online, also transferring the amount from one account to another. The fraudster is detecting the details of the user transaction and doing the illegal activities with the card by phishing, Trojan virus, etc. The fraud may threaten the users with their sensitive information. In this paper, we have discussed various methods of detecting and controlling fraudulent activities.

This will be helpful to improve the security of card transactions in future. Credit card fraudulent activities which are faced by the people is one of the major issues. Due to these fraudulent activities, many credit card users are losing their money and their sensitive information. In this paper, we have discussed the different fraudulent detection and controlling techniques in credit cards and also it will be helpful to improve the security of the fraudsters in future to avoid illegal activities.

C. An Efficient Techniques for Fraudulent detection in Credit Card Dataset: A Comprehensive study

Nowadays, credit card transactions are one of the most famous modes of financial transaction. Increasing trends of financial transactions through credit cards also invite fraud activities that involve the loss of billions of dollars globally. It has also been observed that fraudulent transactions have increased by 35% since 2018. A huge amount of transaction data is available to analyse the fraud detection activities that require analysis of behaviour/abnormalities in the transaction dataset to detect and ignore the undesirable action of the suspected person. The proposed paper lists a compressive summary of various techniques for the classification of fraud transactions from the various datasets to alert the user to such transactions. In the last decades, online transactions are growing rapidly and are the most common tool for financial transactions.

The increasing growth of online transactions also increases threats. Therefore, in keeping in mind the security issue, nature, and anomaly in the credit card transaction, the proposed work represents the summary of various strategies applied to identify the abnormal transaction in the dataset of credit card transaction datasets.This dataset contains a mix of normal and fraud transactions; this proposed work classifies and summarises the various classification methods to classify the transactions using various Machine Learning-based classifiers. The efficiency of the method depends on the dataset and classifier used.

The proposed summary will be beneficial to the banker, credit card user, and researcher to analyse to prevent credit card fraud. The future scope of this credit card fraud detection is to explore the things in each and every association and bank to live a safe and happy life. The data must be balanced in each place and we are getting the best results.

D. A Review On Credit Card Fraud Detection Using Machine Learning

In recent years credit card fraud has become one of the growing problems. A large financial loss has greatly affected individual people using a credit card and also the merchants and banks. Machine learning is considered one of the most successful techniques to identify fraud. This paper reviews different fraud detection techniques using machine learning and compares them using performance measures like accuracy, precision and specificity. The paper also proposes an FDS which uses a supervised Random Forest algorithm. With this proposed system the accuracy of detecting fraud in credit cards is increased. Further, the proposed system uses the learning to rank approach to rank the alert and also effectively addresses the problem of concept drift in fraud detection.

This paper has reviewed various machine learning algorithms to detect fraud in credit card transactions. The performances of all these techniques are examined based on accuracy, precision and specificity metrics. We have selected the supervised learning technique, Random Forest, to classify the alert as fraudulent or authorised. This classifier will be trained using feedback and a delayed supervised sample. Next, it will aggregate each probability to detect alerts. Further, we proposed a learning to rank approach where alerts will be ranked based on priority. The suggested method will be able to solve the class imbalance and concept drift problem. Future work will include applying semi-supervised learning methods for the classification of alerts in FDS.

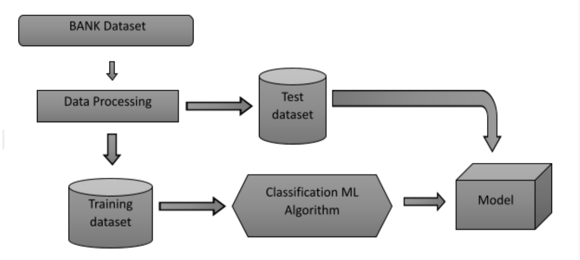
E. Credit Card Fraud Detection and Prevention using Machine Learning

This research focused mainly on detecting credit card fraud in the real world. We must collect the credit card data sets initially for the qualified data set. Then provide queries on the user's credit card to test the data set. After random forest algorithm classification method using the already evaluated data set and providing the current data set[1]. Finally, the accuracy of the results data is optimised.Then the processing of a number of attributes will be implemented, so that affecting fraud detection can be found in viewing the representation of the graphical model. The technique's efficiency is measured based on accuracy, flexibility, and specificity, precision. The results obtained with the use of the Random Forest Algorithm have proved much more effective.

III. RESEARCH METHODOLOGY

The proposed model is to build a classification model to classify whether it's fraud or not. The dataset of previous credit card cases is collected and it is used to make the machine learn about the problem. The first step involves the analysis of data where each and every column is analysed and the necessary measurements are taken for missing values and other forms of data. Outliers and other values which do not have much impact are dealt with.

Then preprocessed data is used to build the classification model where data will be split into two parts one is for training and the remaining data for testing purposes. Machine learning algorithms are applied to the training data where the model learns the pattern from the data and the model will deal with test data or new data and classify whether it's fraud or not. The algorithms are compared and the performance metric of the algorithms are calculated.



The architecture of Proposed model

Exploration data analysis of variable identification:

1. Loading the given dataset
2. Import required libraries packages
3. Analyse the general properties
4. Find duplicate and missing values
5. Checking unique and count values

Data Wrangling:

In this section of the report will load the data, check for cleanliness, and then trim and clean the given dataset for analysis. Make sure that the document steps carefully and justify cleaning decisions.

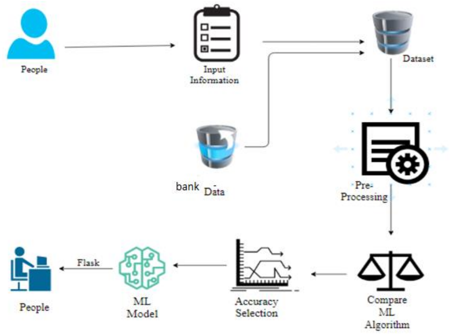
Data collection:

The data set collected for predicting given data is split into the Training set and a Test set. Generally, 7:3 ratios are applied to split the Training set and Test set. The Data Model which was created using Random Forest, logistic, Decision tree algorithms and Support vector classifier (SVC) are applied to the Training set and based on the test result accuracy, Test set prediction is done.

Preprocessing:

The data which was collected might contain missing values that may lead to inconsistency. To gain better results data needs to be preprocessed so as to improve the efficiency of the algorithm. The outliers have to be removed and also variable conversion needs to be done.

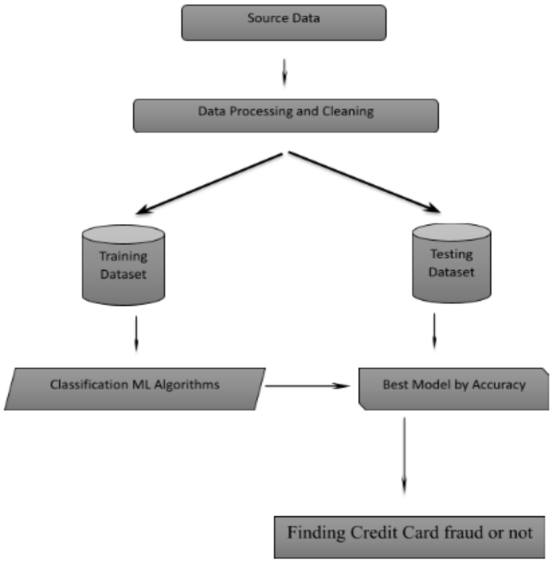
IV. SYSTEM ARCHITECTURE



Building the classification model:

The prediction of credit card fraud, A high accuracy prediction model is effective because of the following reasons: It provides better results in classification problems.

- It is strong in preprocessing outliers, irrelevant variables, and a mix of continuous, categorical and discrete variables.
- It produces out of bag estimate error which has proven to be unbiased in many tests and it is relatively easy to tune with.


Source Data
Data Processing and Cleaning
Training Dataset
Testing Dataset
Classification ML Algorithms
Best Model by Accuracy
Finding Credit Card fraud or not

Workflow diagram

Construction of a Predictive Model:

Machine learning needs data gathering and has a lot of past data. Data gathering has sufficient historical data and raw data. Before data pre-processing, raw data can't be used directly. It's used to pre-process then, what kind of algorithm with the model. Training and testing this model working and predicting correctly with minimum errors. Tuned model involved by tuned time to time with improving the accuracy.

V. CONCLUSION

The analytical process started with data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. The best accuracy on the public test set is a higher accuracy score. This application can help to find the Prediction of credit card fraud or not.

REFERENCES

[1] Aleskerov, E., Freisleben, B., & Rao, B. (1997). Card watch: A neural network-based database mining system for credit card fraud detection. Paper presented at the Proceedings of the IEEE/IAFE 1997 computational intelligence for financial engineering (CIFEr).

[2] Bahnsen, A. C., Aouada, D., Stojanovic, A., & Ottersten, B. J. E. S. w. A. (2016). Feature engineering strategies for credit card fraud detection. 51, 134-142.

[3] Bhatla, T. P., Prabhu, V., & Dua, A. J. C. b. r. (2003). Understanding credit card frauds. 1(6), 1-15. Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. J. D. s. s. (2011). Data mining for credit card fraud: A comparative study. 50(3), 602-613.

[4] Bolton, R. J., & Hand, D. J. J. S. s. (2002). Statistical fraud detection: A review. 17(3), 235-255.

[5] Brause, R., Langsdorf, T., & Hepp, M. (1999). Neural data mining for credit card fraud detection. Paper presented at the Proceedings 11th International Conference on Tools with Artificial Intelligence.Chan, P. K., Fan, W., Prodromidis,A. L., Stolfo, S. J. J. I. I. S., & Applications, T. (1999). Distributed data mining in credit card fraud detection. 14(6), 67-74.

[6] Dal Pozzolo, A., Caelen, O., Le Borgne, Y.-A., Waterschoot, S., & Bontempi, G. J. E. s. w. a. (2014). Learned lessons in credit card fraud detection from a practitioner perspective. 41(10), 4915-4928.

[7] Delamaire, L., Abdou, H., Pointon, J. J. B., & systems, B. (2009). Credit card fraud and detection techniques: a review. 4(2), 57-68. Dhanapal, R., & Gayathri, P. J. I. J. o. C. S. I. (2012). Credit card fraud detection using a decision tree for tracing Email and IP. 9(5), 406.

[8] Dheepa, V., & Dhanapal, R. J. I. J. o. S. c. (2012). Behaviour-based credit card fraud detection using support vector machines. 2(07), 2012.Donepudi, P. K. (2014). Technology Growth in Shipping Industry: An Overview.American Journal of Trade and Policy,1(3), 137-142. https://doi.org/10.18034/ajtp.v1i3.503

[9] Donepudi, P. K. (2014a). Voice Search Technology: An Overview.Engineering International,2(2), 91-102. https://doi.org/10.18034/ei.v2i2.502Donepudi, P. K. (2015). Crossing Point of Artificial Intelligence in Cybersecurity.American Journal of Trade and Policy,2(3), 121-128. https://doi.org/10.18034/ajtp.v2i3.493

[10] Duman, E., & Ozcelik, M. H. J. E. S. w. A. (2011). Detecting credit card fraud by genetic algorithm and scatter search. 38(10), 13057-13063. Excell, D. J. C. F., & Security. (2012).

[11] Bayesian inference–the future of online fraud protection. 2012(2), 8-11. Foster, D. P., & Stine, R. A. J. J. o. t. A. S. A. (2004). Variable selection in data mining: Building a predictive model for bankruptcy. 99(466), 303-313.

[12] Gaikwad, J. R., Deshmane, A. B., Somavanshi, H. V., Patil, S. V., Badgujar, R. A. J. I. J. o. I. T., & Engineering, E. (2014). Credit Card Fraud Detection using Decision Tree Induction Algorithm. 4(6).

[13] Ghosh, S., & Reilly, D. L. (1994). Credit card fraud detection with a neural network. Paper presented at the System Sciences, 1994. Proceedings of the Twenty-Seventh Hawaii International Conference on.Juszczak, P., Adams, N. M., Hand, D. J., Whitrow, C., Weston, D. J. J. C. S., & Analysis, D. (2008). Off-the-peg and bespoke classifiers for fraud detection. 52(9), 4521-4532