

# IDENTIFICATION OF FRAUD DETECTION FROM CREDIT CARD TRANSACTIONS USING ML

Dr. M. Radhika Mani,<sup>1</sup>, Komarathi Vyshnavi Devi<sup>2</sup>, Mavuri Divya<sup>3</sup>, Ganthakuri Sree Lakshmi Siddardha<sup>4</sup>, Karnasula Lakshmi Tulasi<sup>5</sup>, Dadi Satya<sup>6</sup>

PROFESSOR

DEPT OF COMPUTER SCIENCE AND ENGINEERING  
PRAGATI ENGINEERING COLLEGE(A), SURAMPALEM(EAST GODAVARI)A.P, INDIA

## **To Cite this Article**

Dr. M. Radhika Mani, , Komarathi Vyshnavi Devi , Mavuri Divya , Ganthakuri Sree Lakshmi Siddardha , Karnasula Lakshmi Tulasi , Dadi Satya, **IDENTIFICATION OF FRAUD DETECTION FROM CREDIT CARD TRANSACTIONS USING ML** ” *Journal of Science and Technology*, Vol. 08, Issue 04, -April 2023, pp25-31

## **Article Info**

Received: 20-02-2023

Revised: 14-03-2023

Accepted: 20-03-2023

Published: 12-04-2023

## **Abstract:**

This project is mainly focused on credit card fraud detection in real world. A phenomenal growth in the number of credit card transactions, has recently led to a considerable rise in fraudulent activities. In the modern era, the usage of the internet has increased a lot these days and becomes an essential part of the life. As the e-commerce has increased, the buying and selling the products over the internet becomes easier and more flexible. The usage of online shopping, online bill payment has increased a lot these days with the introduction of the modern technology like online banking, credit card payments. Due to the increase of the online payment and online shopping, the risk of credit card usage also increased as the credit card was used in many places as it becomes hard for the bank to distinguish the real transactions of the consumer versus the fraud transactions. Also, the credit card fraud transaction can also happen if the customer accidentally loses the credit card. So, it becomes complex for the banks to stop the fraud transactions at that point. This project represents about how to detect the fraud transactions and block the payments before processing by using the machine learning on a real-time basis.

## **I. Introduction**

There are various fraudulent activities detection techniques has implemented in credit card transactions have been kept in researcher minds to methods to develop models based on artificial intelligence, data mining, fuzzy logic and machine learning. Credit card fraud detection is significantly difficult, but also popular problem to solve. In our proposed system we built the credit card fraud detection using Machine learning. With the advancement of machine learning techniques. Machine learning has been identified as a successful measure for fraud detection. A large amount of data is transferred during online transaction processes, resulting in a binary result: genuine or fraudulent. Within the sample fraudulent datasets, features are constructed. These are data points namely the age and value of the customer account, as well as the origin of the credit card. There are hundreds of features and each contributes, to varying extents, towards the fraud probability. Note, the level in which each feature contributes to the fraud score is generated by the artificial intelligence of the machine which is driven by the training set, but is not determined by a fraud analyst. So, in regards to the card fraud, if the use of cards to commit fraud is proven to be high, the fraud weighting of a transaction that uses a credit card will be equally so. However, if this were to shrink, the contribution level would parallel. Simply make, these models self-learn without explicit programming such as with manual review. Credit card fraud detection using machine learning is done by deploying the classification and regression algorithms. We use supervised learning algorithm such as Random Forest algorithm to classify the fraud card transaction in online or by offline. Random forest is advanced version of Decision tree. Random forest has better efficiency and accuracy than the other machine learning algorithms. Random forest aims to reduce the previously mentioned correlation issue by picking only a subsample of the feature space at each split. Essentially, it aims to

make the trees de-correlated and prune the trees by fixing a stopping criterion for node splits, which I will be cover in more detail later

## II. LITERATURE SURVEY

### " The Use of Predictive Analytics Technology to Detect Credit Card Fraud in Canada"

**AUTHORS:** "Kosemani Temitayo Hafiz, Dr. Shaun Aghili, Dr. Pavol Zavarisky."

This research paper focuses on the creation of a scorecard from relevant evaluation criteria, features, and capabilities of predictive analytics vendor solutions currently being used to detect credit card fraud. The scorecard provides a side-by side comparison of five credit card predictive analytics vendor solutions adopted in Canada. From the ensuing research findings, a list of credit card fraud PAT vendor solution challenges, risks, and limitations was outlined.

### " BLAST-SSAHA Hybridization for Credit Card Fraud Detection"

**AUTHORS:** "Amlan Kundu, Suvasini Panigrahi, Shamik Sural, Senior Member, IEEE, and Arun K. Majumdar"

This paper propose to use two-stage sequence alignment in which a profile Analyzer (PA) first determines the similarity of an incoming sequence of transactions on a given credit card with the genuine cardholder's past spending sequences. The unusual transactions traced by the profile analyzer are next passed on to a deviation analyzer (DA) for possible alignment with past fraudulent behavior. The final decision about the nature of a transaction is taken on the basis of the observations by these two analyzers. In order to achieve online response time for both PA and DA, we suggest a new approach for combining two sequence alignment algorithms BLAST and SSAHA.

### " Research on Credit Card Fraud Detection Model Based on Distance Sum"

**AUTHORS:** "Wen-Fang YU, Na Wang".

Along with increasing credit cards and growing trade volume in China, credit card fraud rises sharply. How to enhance the detection and prevention of credit card fraud becomes the focus of risk control of banks. It proposes a credit card fraud detection model using outlier detection based on distance sum according to the infrequency and unconventionality of fraud in credit card transaction data, applying outlier mining into credit card fraud detection. Experiments show that this model is feasible and ac.

### " Fraudulent Detection in Credit Card System Using SVM & Decision Tree."

**AUTHORS:** "Vijayshree B. Nipane, Poonam S. Kalinge, Dipali Vidhate, Kunal War, Bhagyashree P. Deshpande".

With growing advancement in the electronic commerce field, fraud is spreading all over the world, causing major financial losses. In current scenario, Major cause of financial losses is credit card fraud; it not only affects trades person but also individual clients. Decision tree, Genetic algorithm, Meta learning strategy, neural network, HMM are the presented methods used to detect credit card frauds. In contemplate system for fraudulent detection, artificial intelligence concept of Support Vector Machine (SVM) & decision tree is being used to solve the problem. Thus by implementation of this hybrid approach, financial losses can be reduced to greater extend.

## III.SYSTEM ANALYSIS

### 3.1. EXISTING SYSTEM

In existing System, research about a case study involving credit card fraud detection, where data normalization is applied before Cluster Analysis and with results obtained from the use of Cluster Analysis and Artificial Neural Networks on fraud detection has shown that by clustering attributes neuronal inputs can be minimized. And promising results can be obtained by using normalized data and data should be MLP trained. This research was based on unsupervised learning. Significance of this paper was to find new methods for fraud detection and to increase the accuracy of results. The data set for this paper is based on real life transactional data by a large European company and personal details in data is kept confidential. Accuracy of an algorithm is around 50%. Significance of this paper was to find an algorithm and to reduce the cost measure. The result obtained was by 23% and the algorithm they find was Bayes minimum risk.

#### **Drawbacks Of Existing System**

- In this paper a new collative comparison measure that reasonably represents the gains and losses due to fraud detection is proposed.
- A cost sensitive method which is based on Bayes minimum risk is presented using the proposed cost measure.

### **3.2. PROPOSED SYSTEM**

In proposed System, we are applying random forest algorithm for classification of the credit card dataset. Random Forest is an algorithm for classification and regression. Summarily, it is a collection of decision tree classifiers. Random forest has advantage over decision tree as it corrects the habit of over fitting to their training set. A subset of the training set is sampled randomly so that to train each individual tree and then a decision tree is built, each node then splits on a feature selected from a random subset of the full feature set. Even for large data sets with many features and data instances training is extremely fast in random forest and because each tree is trained independently of the others. The Random Forest algorithm has been found to provide a good estimate of the generalization error and to be resistant to over fitting.

#### **ADVANTAGES OF PROPOSED SYSTEM**

- Random forest ranks the importance of variables in a regression or classification problem in a natural way can be done by Random Forest.
- The 'amount' feature is the transaction amount. Feature 'class' is the target class for the binary classification and it takes value 1 for positive case (fraud) and 0 for negative case (not fraud).

## **IV.SYSTEM DESIGN**

### **4.1 SYSTEM ARCHITECTURE**

Below diagram depicts the whole system architecture of the proposed technique.

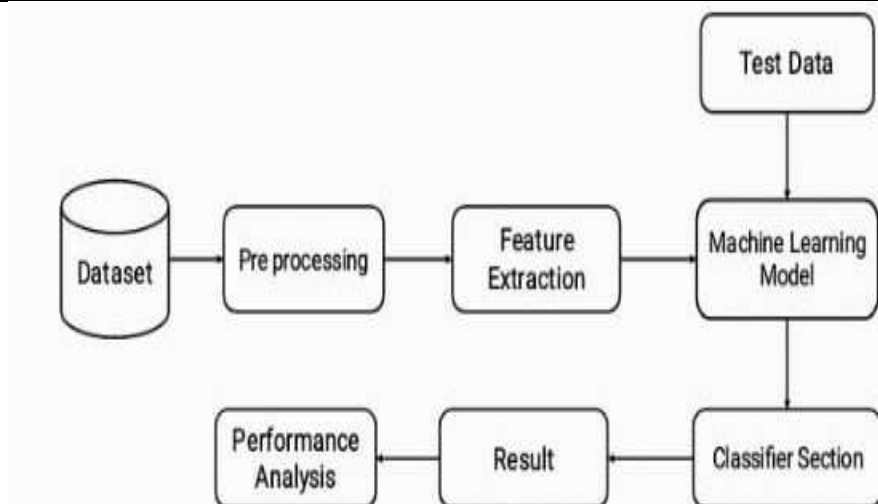


Fig. 4.1.1 System Architecture

## V.SYSTEM IMPLEMENTATION

### 5.1. MODULES

- DATA COLLECTION
- DATA PRE-PROCESSING
- FEATURE EXTRACTION
- EVALUATION MODEL

Module description:

#### 5.1.1 Data Collection

Data used in this paper is a set of product reviews collected from credit card transactions records. This step is concerned with selecting the subset of all available data that you will be working with. ML problems start with data preferably, lots of data (examples or observations) for which you already know the target answer. Data for which you already know the target answer is called labelled data.

#### 5.1.2 Data Pre-processing

Organize your selected data by formatting, cleaning, and sampling from it. Three common data pre-processing steps are:

**Formatting:** The data you have selected may not be in a format that is suitable for you to work with. The data may be in a relational database and you would like it in a flat file, or the data may be in a proprietary file format and you would like it in a relational database or a text file.

**Cleaning:** Cleaning data is the removal or fixing of missing data. There may be data instances that are incomplete and do not carry the data you believe you need to address the problem. These instances may need to be removed. Additionally, there may be sensitive information in some of the attributes and these attributes may need to be removed from the data entirely.

**Sampling:** There may be far more selected data available than you need to work with. More data can result in much longer running times for algorithms and larger computational and memory requirements. You can take a smaller

representative sample of the selected data that may be much faster for exploring and prototyping solutions before considering the whole dataset.

### 5.1.3 Feature Extraction

Next thing is to do Feature extraction is an attribute reduction process. Unlike feature selection, which ranks the existing attributes according to their predictive significance, feature extraction actually transforms the attributes. The transformed attributes, or features, are linear combinations of the original attributes. Finally, our models are trained using Classifier algorithm. We use classify module on Natural Language Toolkit library on Python. We use the labelled dataset gathered. The rest of our labelled data will be used to evaluate the models. Some machine learning algorithms were used to classify pre-processed data. The chosen classifiers were Random Forest. These algorithms are very popular in text classification tasks.

### 5.1.4 Evaluation Model

Model Evaluation is an integral part of the model development process. It helps to find the best model that represents our data and how well the chosen model will work in the future. Evaluating model performance with the data used for training is not acceptable in data science because it can easily generate overoptimistic and over fitted models. There are two methods of evaluating models in data science, Hold-Out and Cross-Validation. To avoid over fitting, both methods use a test set (not seen by the model) to evaluate model performance. Performance of each classification model is estimated base on its averaged. The result will be in the visualized form. Representation of classified data in the form of graphs. Accuracy is defined as the percentage of correct predictions for the test data. It can be calculated easily by dividing the number of correct predictions by the number of total predictions.

## VI. RESULTS



Fig.6.1 Random Forest accuracy

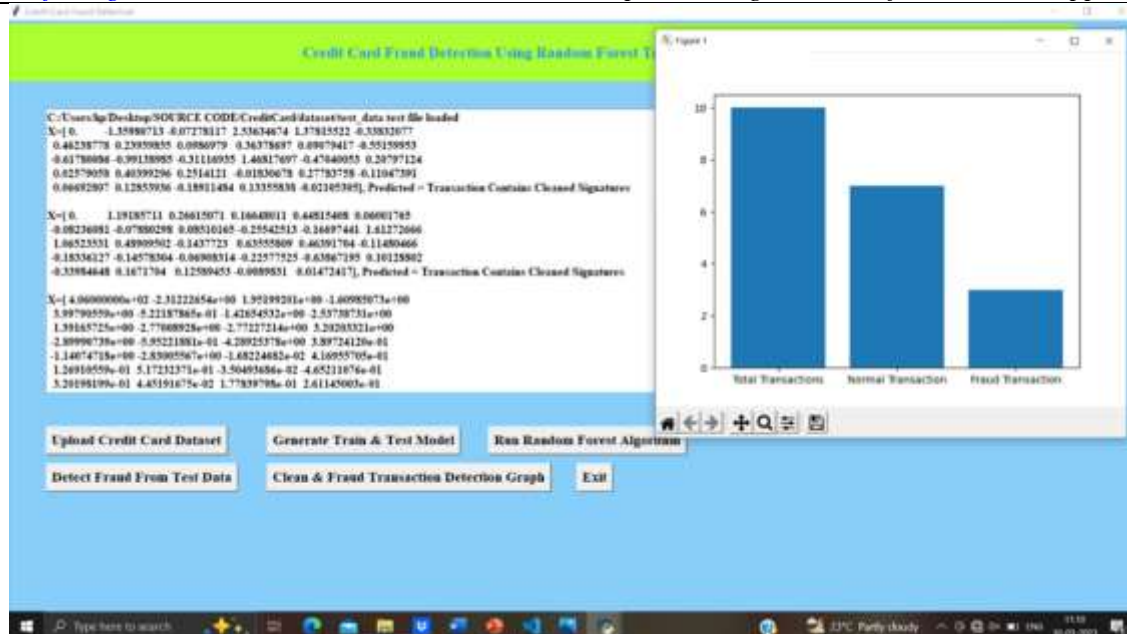


Fig.6.2 Fraud Transaction Detection Graph

## VII. CONCLUSION AND FUTURE WORK

The Random Forest algorithm will perform better with a larger number of training data, but speed during testing and application will suffer. Application of more pre-processing techniques would also help. The SVM algorithm still suffers from the imbalanced dataset problem and requires more preprocessing to give better results at the results shown by SVM is great but it could have been better if more preprocessing have been done on the data. In future we plan to enhance the existing algorithm and train it with another credit card dataset having a greater number of features.

## REFERENCES :

- [1] Sudhamathy G: Credit Risk Analysis and Prediction Modelling of Bank Loans Using R, vol. 8, no-5, pp. 1954-1966.
- [2] LI Changjian, HU Peng: Credit Risk Assessment for ural Credit Cooperatives based on Improved Neural Network, International Conference on Smart Grid and Electrical Automation vol. 60, no. - 3, pp 227-230, 2017.
- [3] Wei Sun, Chen-Guang Yang, Jian-Xun Qi: Credit Risk Assessment in Commercial Banks Based On Support Vector Machines, vol.6, pp 2430-2433, 2006.
- [4] Amlan Kundu, Suvasini Panigrahi, Shamik Sural, Senior Member, IEEE, "BLAST-SSAHA Hybridization for Credit Card Fraud Detection", vol. 6, no. 4 pp. 309-315, 2009.
- [5] Y. Sahin and E. Duman, "Detecting Credit Card Fraud by Decision Trees and Support Vector Machines, Proceedings of International Multi Conference of Engineers and Computer Scientists, vol. I, 2011.
- [6] Sitaram patel, Sunita Gond , "Supervised Machine (SVM) Learning for Credit Card Fraud Detection, International of engineering trends and technology, vol. 8, no. -3, pp. 137- 140, 2014.

- 
- [7] Snehal Patil, Harshada Somavanshi, Jyoti Gaikwad, Amruta Deshmane, Rinku Badgajar," Credit Card Fraud Detection Using Decision Tree Induction Algorithm, International Journal of Computer Science and Mobile Computing, Vol.4 Issue.4, April- 2015, pg. 92-95
- [8] Dahee Choi and Kyungho Lee, "Machine Learning based Approach to Financial Fraud Detection Process in Mobile Payment System", vol. 5, no. - 4, December 2017, pp. 12-24.