ENGINEERING IN ADVANCED RESEARCHSCIENCE AND TECHNOLOGY

ISSN 2352-8648
Vol.01, Issue.03
April-2022
Pages: -680-685

# CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING ALGORITHM

**Dr. M. Chaitanya Kishore Reddy[1], Y. Srilatha[2], P. Deepak[3], M. Janaki[4]**

[1] chkishore.0007@gmail.com, *Professor & HoD-IT, NRI Institute of Technology, A.P, India-521212.*
[2] yadalasrilatha10@gmail.com, *UG scholar, B. tech, Dept.IT, NRI Institute of technology, A.P-521212.*
[3] deepakpunugupati@gmail.com, *UG scholar, B. tech, Dept.IT, NRI Institute of technology, A.P-521212.*
[4] marellajanaki393@gmail.com, *UG scholar, B. tech, Dept.IT, NRI Institute of technology, A.P-521212*

----------***----------

**Abstract** - *Credit card fraud is a field with perpetrators performing illegal actions that may affect other individuals or companies negatively. For instance, a criminal can steal credit card information from an account holder and then conduct fraudulent transactions. The activities are a potential contributory factor to how illegal organizations such as terrorists and drug traffickers support themselves financially. Within the machine learning area, there are several methods that possess the ability to detect credit card fraud transactions; supervised learning and unsupervised learning algorithms. This essay investigates the supervised approach (Random Forest) is evaluated on a real-life dataset of 284,807 transactions. Under those circumstances, the main purpose is to develop a "well-functioning" model with a reasonable capacity to categorize transactions as fraudulent or legit. As the data is heavily unbalanced, reducing the false-positive rate is also an important part when conducting research in the chosen area.*

**KEY WORDS:** Credit card, Fraud, Transaction, Machine learning, Feature extraction.

**1.INTRODUCTION:** Credit card is a small thin plastic or fiber card that contains information about the person such as picture or signature and person named on it to charge purchases and service to his linked account charges for which will be debited regularly. Now a day's card information is read by ATM's, swiping machines, store readers, bank and online transaction. Each card as a unique card number which is very important, its security is mainly relies on physical security of the card and also privacy of the credit card number. There is rapid increase in the credit card transaction which as led to substantial growth in fraudulent cases. Many datamining and statistical methods are used to detect fraud. Many fraud detection techniques are implemented using artificial intelligence, pattern matching. Detection of fraud using efficient and secure methods are very important. Credit card is a small thin plastic or fiber card that contains information about the person such as picture or signature and person named on it to charge purchases and service to his linked account charges for which will be debited regularly. Now a day's card information is read by ATM's, swiping machines, store readers, bank and online transaction. Each card as a unique card number which is very important, its security is mainly relies on physical security of the card and also privacy of the credit card number. There is rapid increase in the credit card transaction which as led to substantial growth in fraudulent cases. Many data mining and statistical methods are used to detect fraud. Many fraud detection techniques are implemented using artificial intelligence, pattern matching. Detection of fraud using efficient and secure methods are very important. Million and billions of people use the credit card for payment in both online and offline transaction, due to existence of widespread point of sale (POS). countless transaction occurred per minute everywhere in the planet. The reason behind fraud is negligence of user. when third person steal the most important information about credit card and user details easily fraud can be achieved. To detect what type of fraud,

occur during transaction, we need to face Several challenges. Fetching that among all the transactions is occurred and which one is real could be a task. Amongst the standard and very common ways of making payment globally and especially in North America, because of the presence of a far-reaching point of sale. A huge number of individuals around the globe use charge cards to buy products and services by getting credit for a time of half a month. Any helpful framework could be mishandled. The challenge is to recognize fraudulent credit card transactions so that the customers of credit card companies are not charged for items that they did not purchase. Main challenges involved in credit card fraud detection are:

- Enormous Data is processed every day and the model build must be fast enough to respond to the scam in time.
- Imbalanced Data i.e., most of the transactions (99.8%) are not fraudulent which makes it really hard for detecting the fraudulent ones
- Data availability as the data is mostly private.
- Misclassified Data can be another major issue, as not every fraudulent transaction is caught and reported.
- Adaptive techniques used against the model by the scammers.

**2. RANDOM FOREST:**

A random forest is a supervised machine learning algorithm that is constructed from decision tree algorithms. This algorithm is applied in various industries such as banking and e-commerce to predict behavior and outcomes. A random forest is a machine learning technique that's used to solve regression and classification problems. It utilizes ensemble learning, which is a technique that combines many classifiers to provide solutions to complex problems. A random forest algorithm consists of many decision trees. The 'forest' generated by the random forest algorithm is trained through bagging or bootstrap aggregating. Bagging is an ensemble meta-algorithm that improves the accuracy of machine learning algorithms.

- It's more accurate than the decision tree algorithm.
- It provides an effective way of handling missing data.
- It can produce a reasonable prediction without hyper- parameter tuning.
- It solves the issue of overfitting in decision trees.

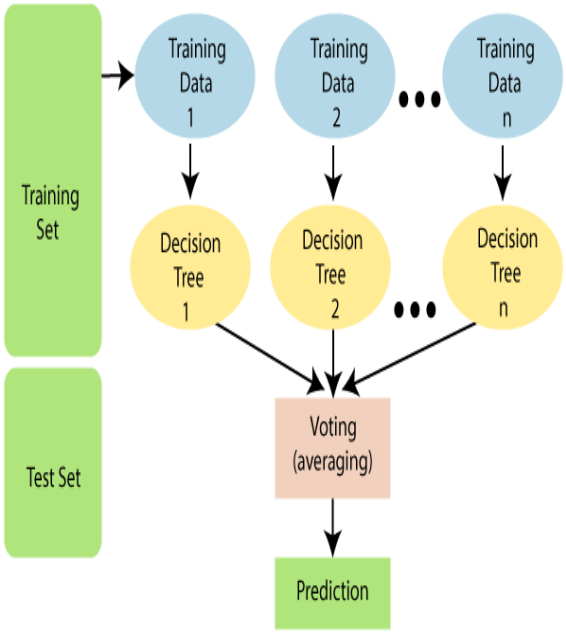


**Fig. 1.** Random Forest

**3. Existing System**: In existing System, 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.

**3.1. Disadvantages of Existed System**:

1. In this paper a new collative comparison measure that reasonably represents the gains and losses due to fraud detection is proposed.

2. A cost sensitive method which is based on Bayes minimum risk is presented using the proposed cost measure.

**4.Proposed System**: In proposed System, we are applying Random Forest algorithm. 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. The proposed system has the following advantages compared to other approaches that currently existing system exists for:

1. Random Forest ranks the importance of variables in a regression or classification problem in a natural way can be done by Random Forest.

2. 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)

**5. System Architecture:** Every software has a model. Before the implementation of the software, architecture is drawn within the sort of any model or any diagram. So, it become quite easier to know the flow of the software and it also help within the easier implementation of the software. The client interacts with the developer team so that client can make them understand that what kind of software he wants. According to his idea the developer team makes diagram in order that it can become easy for the both of the parties. If any changes are to be made, then the client could also ask the developer to make the required change. The explanation of the above architecture is as follows: System architecture of proposed method consists of mainly two tasks. Here the architecture diagram of credit card fraud detection is shown below.

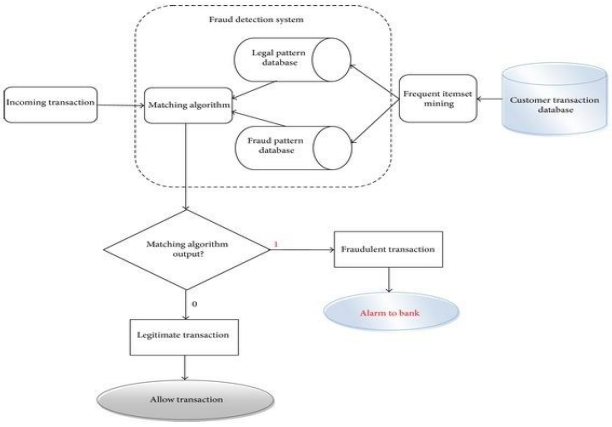


**Fig. 2.** System Architecture

The steps followed in architecture diagram are

Step 1: Initially load the dataset named as "credit card fraud data set" which is downloaded from kaggle website into jupyter notebook.

Step 2: Data Cleaning Process can do in Step2. Check the Missing values in the dataset. Detect the missing values if are present and remove or fill the missing values with mean, median and most repeated value. Convert the data into desired form if they are in undesired form.

Step 3: In Step3 verify the data again and check the missing values.

Step 4: Select the two variables X, Y and copy the attributes in the dataset into the variables which are helpful to fraud detection

Step 5: Split the dataset into training data and testing data.

Step 6: Fit the model to the training data. The model we selected for fraud detection status is "Random Forest" which is supervised learning algorithm tested on labeled data.

Step 7: Test the fraud or not by using the tested data and model. In this way we enter new data for checking fraud status by using the previous data.

Step 8: Check the accuracy of credit card fraud detection of new data and print.

**6. Flow Chart:** A flowchart is a picture of the separate steps of a process in sequential order. It is a generic tool that can be adapted for a wide variety of purposes, and can be used to describe various processes, such as a manufacturing process, an administrative or service process, or a project plan.

Flowchart Basic Procedure:

1. Define the process to be diagrammed. Write its title at the top of the worksurface.
2. Discuss and decide on the boundaries of your process: Where or when does the process start? Where or when does it end? Discuss and decide on the level of detail to be included in the diagram.
3. Brainstorm the activities that take place. Write each on a card or sticky note.

4. Arrange the activities in proper sequence.
5. When all activities are included and everyone agrees that the sequence is correct, draw arrows to show the flow of the process.

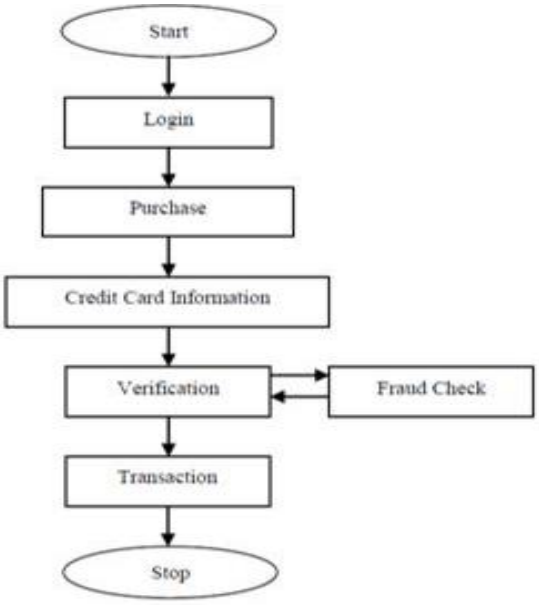


**Fig. 3.** Flow Chart

**7.Conclusion**: Hence, we have acquired the result of an accurate value of credit card fraud detection i.e.,0.995611109160493 (99.56%) using a random forest algorithm with new enhancements. In comparison to existing modules, this proposed module is applicable for the larger dataset and provides more accurate results. The Random Forest algorithm will provide better performance with many training data, but speed during testing and application will still suffer. Usage of more pre-processing techniques would also assist.

**8. Future Scope:** Our future work will try to represent this into a software application and provide a solution for credit card fraud using the new technologies like Machine Learning, Artificial Intelligence and Deep Learning.

## 9. References

[1] Gupta, Shalini, and R. Johari. "A New Framework for Credit Card Transactions Involving Mutual Authentication between Cardholder and Merchant." International Conference on Communication Systems and Network Technologies IEEE, 2011:22-26

[2] A Neural Network Ensemble with Feature Engineering for Improved Credit Card Fraud Detection Ebenezer Esenogho; Ibomoiye Domor Mienye; Theo G. Swart; Kehinde Aruleba; George Obaido IEEE Access Year: 2022 | Volume: 10 | Journal Article | Publisher: IEEE

[3] Supervised Machine Learning Algorithms for Credit Card Fraud Detection: A Comparison Samidha Khatri; Aishwarya Arora; Arun Prakash Agrawal 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence) Year: 2020 | Conference Paper | Publisher: IEEE

[4] Comprehensive Analysis for Fraud Detection of Credit Card through Machine Learning Parth Roy; Prateek Rao; Jay Gajre; Kanchan Katake; Arvind Jagtap; Yogesh Gajmal 2021 International Conference on Emerging Smart Computing and Informatics (ESCI) Year: 2021 | Conference Paper | Publisher: IEEE

[5] Fraud Detection in Credit Card Data using Unsupervised Machine Learning Based Scheme Arun Kumar Rai; Rajendra Kumar Dwivedi 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC) Year: 2020 | Conference Paper | Publisher: IEEE

[6] Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study Sahil Dhankhad; Emad Mohamme; Behrouz Far 2018 IEEE International Conference on Information Reuse and Integration (IRI) Year: 2018 | Conference Paper | Publisher: IEEE

[7] Real-time Credit Card Fraud Detection Using Machine Learning Anuruddha Thennakoon; Chee Bhagyani; Sasitha Premadasa; Shalitha Mihiranga; Nuwan Kuruwitaarachchi 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)Year: 2019 | Conference Paper | Publisher: IEEE

[8] Credit Card Fraud Detection using Machine Learning Anjali Singh Rathore; Ankit Kumar; Depanshi Tomar; Vasudha Goyal; Kaamya Sarda; Dinesh Vij 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART) Year: 2021 | Conference Paper | Publisher: IEEE

[9] Credit Card Fraud Detection with Machine Learning Methods Gokhan Goy; Cengiz Gezer; Vehbi Cagri Gungor 2019 4th International Conference on Computer Science

and Engineering (UBMK) Year: 2019 | Conference Paper | Publisher: IEEE

**Author's Profile:**

Dr. M. Chaitanya Kishore Reddy received Ph.D. in computer science and Engineering from Jawaharlal Nehru Technological University, Kakinada. He is currently working as a HOD, Department of Information Technology at NRI Institute of Technology. He has published papers at international conference on Research Advancement in Computer Science and communication (ICRACS), International Conference on Recent Trends in Engineering, Science and Technology (ICRTEST), International Conference on Communications, Signal Processing Computing and Information Technologies (ICCSPCIT), International Journal of Engineering and Computer Science (IJECS), International Journal and Magazine of Engineering and Technology, management and Research(IJMETMR) in Mobile Ad-hoc Networks and Wireless sensor networks.

Srilatha. Y is currently studying B. Tech in the stream of Information Technology in NRI Institute of Technology. She completed a Mini project on "Pedestrian detection and car detection in videos" and completed Certification on cyber security course in Udemy and had done python internship in VN technologies.

Deepak. P is currently studying B. Tech in the stream of Information Technology in NRI Institute of Technology. He had done a Mini Project called "Disha: The Safe" and completed NPTEL Certification courses on "Programming in Java" and "The Joy of Computing using Python".

Janaki. M is currently studying B. Tech in the stream of Information Technology in NRI Institute of Technology. She did a mini project on Novel Design f Line Follower Robot with Obstacle Apphrension which helps the industries to make the work easy and automated.