*International Journal of Interdisciplinary Innovative Research &Development (IJIIRD)*

**ISSN: 2456-236X**

Vol. 04 Issue 02 |2020

# Credit Card Fraud Detection Using Machine Learning

[1]*Aishwarya R Gowri*

*Department of MCA, computer science, Jain University, Jayanagar Bangalore, India*

**ABSTRACT**

***It is very essential for credit card companies to identify the fraudulent credit card transactions to ensure that their customers are not being charged for the items that they dint purchase. Data mining plays a major role in detecting fraudulent online credit card transactions and it becomes very challenging due to two major reasons such as the profiles of fraudulent and normal behaviours changes constantly and the credit card fraud data sets are highly skewed.Such problems can be addressed with machine learning and data science. And this technique is applied on pre-processed and raw data. This mechanism is implemented in python and the performance is evaluated based on the sensitivity, accuracy, Matthews correlation coefficient. It is easy to distinguish between fraudulent and genuine transition. In this process the main focus is on pre-processing and analysing the data sets and also deploying multiple anomaly detection algorithms such as forest isolation algorithm and local outlier factor.***

## 1. INTRODUCTION

Advanced technology has become the integral part of our life [1]. To satisfy the need of the society, almost in each work, we use the technology [2] [3]. In current era computer science is major subject [4]. It has many real life applications such as cloud computing [5], artificial intelligence [6], remote monitoring [7], Wireless sensor network [8, 9, 10], internet of things [11, 12, 13], Neural network [14, 15], FSPP [16, 17, 18], NSPP [19, 20, 21, 22, 23], TP [24, 25, 26], internet Security [27], uncertainty [28, 29, 30, 31, 32] and so on. Technology is the mode by which user can store, fetch, communicate and utilize the information [33]. So, all the organizations, industries and also every individual are using computer systems to preserve and share the information [34] . The internet security plays a major role in all computer related applications. The internet security appears in many real-life applications, e.g., home security, banking system, education sector, defence system, Railway, and so on. In this manuscript we discuss about the protection of authentication which is a part of internet security.

In credit card transaction the fraudulent transaction is unauthorized and unwanted usage of an account by someone apart from theowner of that account. Necessary prevention measures may be taken to prevent this abuse and also the behaviour of such fraudulent practices may be studied to reduce it and protect against similar occurrences within the future.Financial fraud may be a growing concern with far reaching consequences within thegovernment, corporate organizations, and finance industry. Credit card transactions have become a widespread mode of payment and the focus has been given to various computational methodologies in order to handle credit card fraud issue. Data mining technique is one of the notable and most popular methods used to solve credit card fraud detection problem. it's impossible to be sheer certain about verity intention and rightfulness behind an application or transaction. In reality, to find out possible evidences of fraud from the available data using mathematical algorithms is that the best effective option.In real time examples, a large stream of payment requests is quickly scanned by automatic tools that determine which transactions are authorize. Machine learning algorithms are employed to analyse all theauthorized transactions and report the suspicious ones. Thesereports are investigated by professionals who contact thecardholders to verify if the transaction was genuine orfraudulent.

Credit card fraud takes place when the physical card is stolen or some important data associated with the account is leaked. Credit card fraudster uses many ways to commit the fraud activities. Credit card fraud can be defined as when a person uses some other individual's credit card for their personal benefits without the knowledge of the card owner or the card issuer. Some of the other techniques used by the fraudster to commit credit card frauds are social engineering, where a fraudster will pretend to be someone genuine and ask number of questions and confuse the card holder resulting in giving out important information about the account. Also unauthorized use of stolen or lost card leads to credit card fraud. Some of the false merchant and cloning sites on the internet is another popular method used by many criminals with the skilled ability of hacking.

## 2. LITERATURE REVIEW

[1] E-commerce industry is growing rapidly and this leads to the increased usage of credit card payments for online purchases. In this paper investigation of the performance of logistic regression, random forest and decision tree for credit card fraud detection is carried out. The dataset for credit card fraud detection is gathered from kaggle and this dataset consists of over 2,84,808 credit card transaction data of a European bank. Fraud transactions are considered as positive class and the genuine transactions are considered as negative class.Dataset consists of imbalanced 0.172% of fraud transitions and the reaming transactions are genuine. Performance is evaluated based on accuracy, sensitivity, error rate and specificity.

[2] In this paper the author has used a case sensitive method which is based on Bays maximum risk and then it is presented using proposed cost measure. The dataset is based on the real life transaction data obtained from a European company and maintaining the confidentiality of the personal data. The accuracy of the algorithm used is 50%. The main significance of this paper is to reduce the cost. The result obtained was 23%.[3]This paper checks and investigates the performance of Random Forest, SVM, logistic regression and Decision tree on a highly skewed credit card fraud data. The dataset was gathered by a European cardholders consisting of about 2,84,786 transactions. The result obtained was 97.7% accuracy by Logistic regression, 97.5% by SVM and 98.6% precise accuracy obtained by Random Forest.[4]in this paper one of the best data mining algorithm called machine learning algorithm was introduced, which was used to recognize the credit card fraud. A half bread grouping framework with exception recognition was utilized in order to differentiate between misrepresentations of internet recreations. The framework obtained online calculations with factual data in order to distinguish various extraction types. This framework attained extreme location rate at 98% along with 0.1% fault rate. [5]This paper discusses about supervisor based classification using Bayesian network classifiers such as Naïve Bayes, K2, Tree Augmented Naïve Bayes (TAN, logistics and J48 classifiers. The datasets are pre-processed by using normalization and principal component analysis. Two datasets were used dummy dataset which represented the characteristics of the credit card data and newly generated dataset using data normalization and principal component analysis technique. All these classifiers achieved over 95% accuracy.

## 3. PROPOSED METHODOLOGY

The study discussed in this paper extends the understanding of the machine learning algorithms used to detect the anomalous activities.

The detailed architecture diagram with real life elements can be represented as:

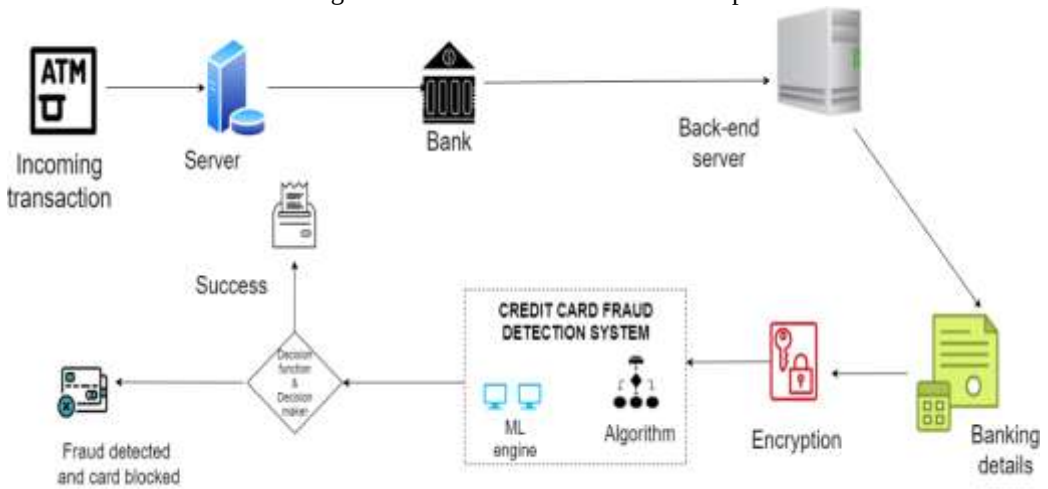


Figure 1: Blueprint of the proposed model

First, we gather the dataset from Kaggle, which is a data analysis website that provides datasets. The dataset consists of 31 columns and 8 rows among those 28 columns are named as v1-v28 in order to protect the sensitivity of the data. The remaining columns represent class, time and amount. Class is represented as 0 or 1 where 0 represents valid transaction and 1 represents fraudulent transaction. Amount represents the amount of money transferred and time manifest the time gap between fist transaction and the next transaction. This graph represents the time at with the transactions were done. Then a histogram is plotted for each column after checking the dataset.

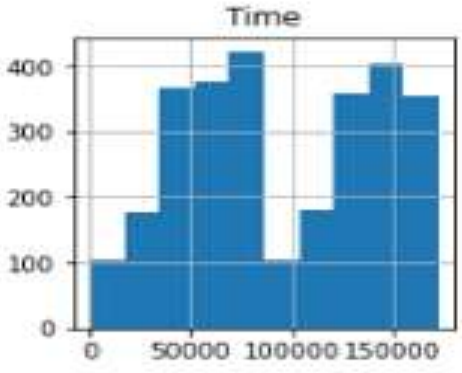


Figure 2: Time graph analysis

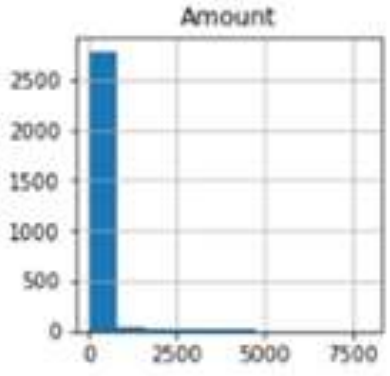


Figure 3: Time graph analysis

This histogram represents the amount of money transferred. This plotting is done in order to get the graphical representation of the data that can be used to ensure that there are no missing values in the dataset used. By describing the dataset, it will give useful information such as mean, median and count for each column. The data is processed from a set of algorithms and they are local outlier factor and isolation forest algorithm which is a part of sklearn. Sklearn is open source and free python library and this library is built using SciPy, Numpy and matplotlib.Jupyter Notebook was used to run the code and demonstrate this approach.

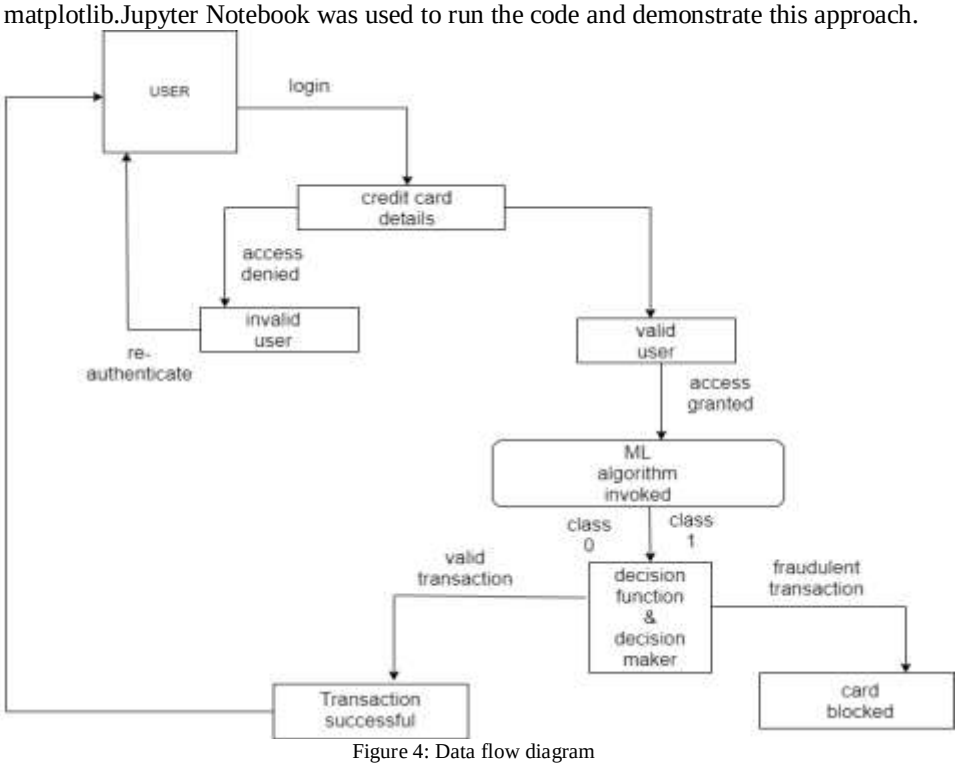


Figure 4: Data flow diagram

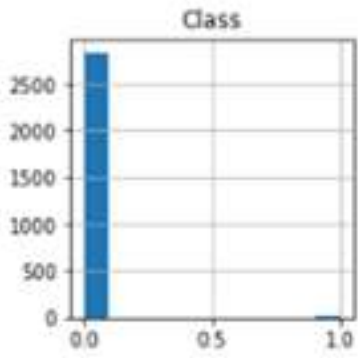


Figure 5: class value graph

In the above fig the user will enter the credit card information and login to the system then the users card details will be verified. If the details are valid then access permissions will be granted to the valid user If the user details are not valid then user should re-authenticate be entering the details. Once access is granted the machine learning algorithms will generate the class value of a particular transaction. If class is 0 then it is a valid transaction and the transaction will be successful and if the class value is 1 then the transaction is fraudulent and the credit card will be blocked.

### 3.1 Mathematical Formulation

#### 3.1.1 Isolation Forest Algorithm

It returns the anomaly score of each sample. It isolates the observation by randomly splitting the value between max and min value of the selected feature and it isolates the points having shorter path links or anomalies.

#### 3.1.2 Local Outlier Factor

It is unsupervised outlier detection method and it calculates the anomaly score of each sample. It will then measure the local deviation of the density of any given sample with its neighbours. Anomaly score depends on how isolated the object is with respect to the nearest neighbourhood.

According to this technique the density estimated of a point p is the number of p's neighbour divided by sum of distance to the point's neighbour.

Estimated destiny is:

$$f^{\wedge}(p)=k\sum x\in N(p)d(p,x) \ldots\ldots\ldots\ldots\ldots\ldots(1)$$

Here, N(p) is the set of neighbours of point p and x

k is the number of points in this set and d(p, x) is the number of distance between the points p and x

local outlier factor score

$$LOF(p)=1k\sum x\in N(p)f^{\wedge}(x)f^{\wedge}(p)\ldots\ldots.(2)$$

## 4. RESULT

The program will generate and detect number of false positive and then it will compare it with the actual values. This graph shows that the number of valid transactions are more when compared to the number of fraudulent transactions in a particular dataset. Class 0 represents the number of valid transactions and class 1 represents number of.

Number of fraud and valid transaction cases in the dataset

```
outlier fraction : len(Fraud)/float(len(Valid))
0.0017587055926837848
Fraud Cases: 5
Valid Transactions: 2843
```

Figure 6: Transactional analysis

## 5. CONCLUSION

Credit card fraud cases are increasing day by day and it is one of the major concerns in financial service sectors. This occurs when are no proper security measures are taken into consideration. In this paper an attempt is made to identify the number of fraudulent transactions in a particular dataset by using various machine learning algorithms such as local outlier factor and isolation forest method. Only a part of dataset was used in order to speed up the computational process. Future scope of improvement includes:

1. Large number of datasets can be stored using cloud storage and then fetch the datasets from the cloud storage repository.
2. User interface is not provided in this project.

## 6. ACKNOWLEDGEMENT

I would like to express my profound gratitude to professor Dr. MN Nachappa and project coordinators for their patient, encouragement and valuable assessments of this research work.

## 7. REFERENCES

[1] M. BM and H. Mohapatra, "Human centric software engineering," International Journal of Innovations & Advancement in Computer Science (IJIACS), vol. 4, no. 7, pp. 86-95, 2015.
[2] H. Mohapatra, C Programming: Practice, Vols. ISBN: 1726820874, 9781726820875, Kindle, 2018.
[3] H. Mohapatra and A. Rath, Advancing generation Z employability through new forms of learning: quality assurance and recognition of alternative credentials, ResearchGate, 2020.
[4] H. Mohapatra and A. Rath, Fundamentals of software engineering: Designed to provide an insight into the software engineering concepts, BPB, 2020.
[5] V. Ande and H. Mohapatra, "SSO mechanism in distributed environment," International Journal of Innovations & Advancement in Computer Science, vol. 4, no. 6, pp. 133-136, 2015.
[6] H. Mohapatra, "Ground level survey on sambalpur in the perspective of smart water," EasyChair, vol. 1918, p. 6, 2019.
[7] H. Mohapatra, S. Panda, A. Rath, S. Edalatpanah and R. Kumar, "A tutorial on powershell pipeline and its loopholes," International Journal of Emerging Trends in Engineering Research, vol. 8, no. 4, 2020.
[8] H. Mohapatra and A. Rath, "Fault tolerance in WSN through PE-LEACH protocol," IET Wireless Sensor Systems, vol. 9, no. 6, pp. 358-365, 2019.
[9] H. Mohapatra, S. Debnath and A. Rath, "Energy management in wireless sensor network through EB-LEACH," International Journal of Research and Analytical Reviews (IJRAR), pp. 56-61, 2019.
[10] H. Mohapatra and A. Rath, "Fault-tolerant mechanism for wireless sensor network," IET Wireless Sensor Systems, vol. 10, no. 1, pp. 23-30, 2020.
[11] H. Mohapatra and A. Rath, "Detection and avoidance of water loss through municipality taps in india by using smart tap and ict," IET Wireless Sensor Systems, vol. 9, no. 6, pp. 447-457, 2019.
[12] M. Panda, P. Pradhan, H. Mohapatra and N. Barpanda, "Fault tolerant routing in heterogeneous environment," International Journal of Scientific & Technology Research, vol. 8, pp. 1009-1013, 2019.
[13] D. Swain, G. Ramkrishna, H. Mahapatra, P. Patra and P. Dhandrao, "A novel sorting technique to sort elements in ascending order," International Journal of Engineering and Advanced Technology, vol. 3, pp. 212-126, 2013.
[14] H. Mohapatra, "HCR using neural network," 2009.
[15] V. Nirgude, H. Mahapatra and S. Shivarkar, "Face recognition system using principal component analysis & linear discriminant analysis method simultaneously with 3d morphable model and neural network BPNN method," Global Journal of Advanced Engineering Technologies and Sciences, vol. 4, p. 1, 2017.
[16] R. Kumar, S. Edalatpanah, S. Jha, S. Gayen and R. Singh, "Shortest path problems using fuzzy weighted arc length," International Journal of Innovative Technology and Exploring Engineering, vol. 8, pp. 724-731, 2019.
[17] R. Kumar, S. Jha and R. Singh, "A different approach for solving the shortest path problem under mixed fuzzy environment," International Journal of fuzzy system Applications, vol. 9, no. 2, pp. 132-161, 2020.
[18] R. Kumar, S. Jha and R. Singh, "Shortest path problem in network with type-2 triangular fuzzy arc length," Journal of Applied Research on Industrial Engineering, vol. 4, pp. 1-7, 2017.
[19] S. Broumi, A. Dey, M. Talea, A. Bakali, F. Smarandache, D. Nagarajan, M. Lathamaheswari and R. Kumar, "Shortest path problem using Bellman algorithm under neutrosophic environment," Complex & Intelligent Systems, vol. 5, pp. 409--416, 2019.
[20] R. Kumar, S. Edalatpanah, S. Jha, S. Broumi, R. Singh and A. Dey, "A multi objective programming approach to solve integer valued neutrosophic shortest path problems," Neutrosophic Sets and Systems, vol. 24, pp. 134-149, 2019.
[21] R. Kumar, A. Dey, F. Smarandache and S. Broumi, "A study of neutrosophic shortest path problem," in Neutrosophic Graph Theory and Algorithms, F. Smarandache and S. Broumi, Eds., IGI-Global, 2019, pp. 144-175.
[22] R. Kumar, S. Edalatpanah, S. Jha and R. Singh, "A novel approach to solve gaussian valued neutrosophic shortest path problems," International Journal of Engineering and Advanced Technology, vol. 8, pp. 347-353, 2019.
[23] R. Kumar, S. Edaltpanah, S. Jha, S. Broumi and A. Dey, "Neutrosophic shortest path problem," Neutrosophic Sets and Systems, vol. 23, pp. 5-15, 2018.
[24] R. Kumar, S. Edalatpanah, S. Jha and R. Singh, "A Pythagorean fuzzy approach to the transportation problem," Complex and Intelligent System, vol. 5, pp. 255-263, 2019.
[25] J. Pratihar, R. Kumar, A. Dey and S. Broumi, "Transportation problem in neutrosophic environment," in Neutrosophic Graph Theory and Algorithms, F. Smarandache and S. Broumi, Eds., IGI-Global, 2019, pp. 176-208.

[26] J. Pratihar, S. E. R. Kumar and A. Dey, "Modified Vogel's Approximation Method algorithm for transportation problem under uncertain environment," Complex & Intelligent Systems (Communicated).
[27] J. Sakhnini, H. Karimipour, A. Dehghantanha, R. Parizi and G. Srivastava, "Security aspects of Internet of Things aided smart grids: A bibliometric survey," Internet of Things, pp. 100-111, 2019.
[28] S. Gayen, F. Smarandache, S. Jha and R. Kumar, "Interval-valued neutrosophic subgroup based on interval-valued triple t-norm," in Neutrosophic Sets in Decision Analysis and Operations Research, M. Abdel-Basset and F. Smarandache, Eds., IGI-Global, 2019, p. 300.
[29] S. Gayen, F. Smarandache, S. Jha, M. Singh, S. Broumi and R. Kumar, "Introduction to plithogenic subgroup," in Neutrosophic Graph Theory and Algoritm, F. Smarandache and S. Broumi, Eds., IGI-Global, 2020, pp. 209-233.
[30] S. Gayen, S. Jha, M. Singh and R. Kumar, "On a generalized notion of anti-fuzzy subgroup and some characterizations," International Journal of Engineering and Advanced Technology, vol. 8, pp. 385-390, 2019.
[31] S. Gayen, F. Smarandache, S. Jha, M. K. Singh, S. Broumi and R. Kumar, "Introduction to plithogenic hypersoft subgroup," Neutrosophic Sets and Systems, vol. 33, p. Accepted, 2020.
[32] S. Gayen, S. Jha and M. Singh, "On direct product of a fuzzy subgroup with an anti-fuzzy subgroup," International Journal of Recent Technology and Engineering, vol. 8, pp. 1105-1111, 2019.
[33] Behura and H. Mohapatra, "IoT Based Smart City with Vehicular Safety Monitoring," EasyChair, vol. 1535, 2019.
[34] P. H, M. H and R. A.K, "WSN-Based Water Channelization: An Approach of Smart Water," Smart Cities—Opportunities and Challenges. Lecture Notes in Civil Engineering, vol. 58, pp. 157-166, 2020.
[35] S. k. Lakshmi S V, "Machine Learning For Credit Card Fraud Detection System," International Journal of Applied Engineering Research, vol. Volume 13, 2018.
[36] G. B. D. S. Jain R., " A hybrid approach for credit card," International Journal of Computer Applications 139(10), 2016.
[37] S. Y. S. Navanshu Khare, "Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models," International Journal of Pure and Applied Mathematics, vol. Volume 118, 2018.
[38] J. K. K. S. B. H. H. Y. Ramyashree. K, "A Hybrid Method for Credit Card Fraud Detection Using Machine Learning Algorithm," International Journal of Recent Technology and Engineering (IJRTE), Vols. Volume-7, no. Issue-6S4, April 2019.
[39] D. N.Sivakumar, ""Fraud Detection in Credit Card Transaction: Classification, Risks and Prevention Techniques"," International Journal of Computer Science and Information, vol. Volume (2), pp. pp.1379-1386, 2015.
[40] S. S. D. S. S P Maniraj, "Credit Card Fraud Detection using Machine," International Journal of Engineering Research & Technology (IJERT), vol. Vol. 8 , no. Issue 09, September-2019.