



# COMPARATIVE ANALYSIS OF ALGORITHMS USED IN CREDIT CARD FRAUD DETECTION SYSTEM

Subhadra S Menon<sup>1</sup>; Dr. Anjana S Chandran<sup>2</sup>

<sup>1</sup>Scholar, SCMS, Cochin, Kerala, India, [subhadrasmenon@gmail.com](mailto:subhadrasmenon@gmail.com)

<sup>2</sup>Assistant Professor, SCMS, Cochin, Kerala, India, [anjana@scmsgroup.org](mailto:anjana@scmsgroup.org)

---

**Abstract:** Machine learning is an application of Artificial Intelligence which is the scientific study of algorithms and that computer use in order to perform a specific given task without making use of explicitly provided instructions and completely relying on patterns and inference instead. This kind of learning is best suited for making predictions and decisions without programming them explicitly. Using machine learning one can analyze different kinds of data and draw conclusions from them. In this paper The Credit Card Fraud Detection is taken care of. This includes considering previous credit card transactions by keeping in mind the ones that usually turns out to be fraud. This model is later on used to identify if a new transaction is fraudulent or not. The paper aims to detect hundred percent of the fraudulent transactions that are occurring along with bringing down the incorrect fraud classifications. For finding out the fraudulent transactions different algorithms are made used of. A comparative study between Random forest, SVM, Naïve Bayes, LOF is performed.

**Keywords:** SVM- Support Vector Machine, LOF-Local Outlier Factor

---

## 1. Introduction

Machine Learning is a widely used as an application of Artificial Intelligence using which machine automatically learns and tries to improve from experience without being programmed explicitly. The system that is taken here is Credit Card Fraud Detection System where many fraudulent transactions occurs and it is quite difficult for detecting how and when such situations comes. In order to predict the number of such transactions occurring, different algorithms are used. A comparison between four different algorithms namely Random Forest, Support Vector Machine, Local Outlier Factor and Naïve Baye's algorithm is performed and is in terms of accuracy, precision and different other parameters.

Credit card fraud is divided into two types: Offline fraud and On-line fraud.

**Offline fraud** is done by making use of a stolen physical card at call center or any other place.

**On-line fraud** is done through internet, phone, shopping, web, or without card holder's knowledge

## 2. Literature Review

Vldamir Zaslavsky, Anna sthrizak in "Credit Card Fraud Detection Using Self-Organizing maps" Uses self-organizing map (SOM) algorithm to analyze how the transactions occur which helps us in finding any suspicious transactions when credit card is being used.. The SOM is an unsupervised neural network which allows us to greatly solved problems that includes recovery of missing data and retrieval of patterns. [2]



*Aditya Oza* in “*Fraud Detection using Machine Learning*” have used the data set on logistic regression which is a technique to find a linear decision boundary followed by support vector machine that creates a classification hyper plane and class weight based approach to sample different weights belonging to fraud and non fraud classes.[5]

*Samaneh Sorounejad, Zahra Zojaji, Reza Ebrahimi Atani, Amir Hassan Monadjemi* In “*A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective*” highlights the fact that credit card has become an unavoidable part of everyone’s life. The paper highlights the key differences between fraud analysis as well as user behaviour analysis approaches. Has mentioned about supervised, unsupervised as well as hybrid techniques have been used for detection. The data sets have been taken, did all the pre processing and given to the algorithm forgetting the results. These data sets itself can be categorical based or numerical based for each algorithm. [7]

*Heta Naik, Prashasti Kanikar* in “*Credit card Fraud Detection based on Machine Learning Algorithms*” takes different machine learning algorithms have been compared in order to find which is the best one. Naive Bayes which comes after Bayes theorem, logistic regression, J48 and Ada Boost. After implementing all these algorithm it was concluded that highest accuracy was given by logistic regression and Ada boost accuracy of hundred percent. [10]

*Linda Delamaire, Hussein Abdou, John Pointon* in “*Credit card fraud and detection techniques: review*” highlights the fact that fraud is one of the most difficult issues in the credit card industry. As Detection techniques, author has discussed decision tree technique, genetic algorithm and other algorithms, clustering technique, neural network and so on. The author made an argument that one of the ethical problem that is faced is to predict those who are fraudulent and those who are genuine customers [12]

*VENKATA RATNAM GANJI, SIVA NAGA PRASAD MANNEM, Venkata Ratnam Ganji* in “*Credit card fraud detection using anti-k nearest neighbor algorithm*” discusses that banks have always used fraud warning system for several years. He tries to bring in a new algorithm which is data stream outlier detection algorithm based on k nearest neighbour (SODRNN). The main advantage of this algorithm is that it needs just one pass of scan where the others require the database can many times. [14]

*Krishna Kumar Tripathi, Mahesh A. Pavaskar* in “*Survey on Credit Card Fraud Detection*” Methods has compared different fraud detection techniques which include effusion approach using Dempster-Shafer theory and Bayesian learning, meta learning strategy, artificial immune system, neural network and so on [16]

*Suman and Nutan* in “*Review Paper on Credit Card Fraud Detection*” tell that there is increase in fraud that result in the loss of dollars worldwide. Author has presented around 7-8 detection method namely Gass algorithm, Bayesian algorithm, neural network, logistic regression and so on. [20]

*Navanshu Khare and Saad Yunus Sait* in “*Credit Card Fraud Detection Using Machine Learning Models and Collating Machine*” in the paper discusses about finance fraud being one of the growing problem even after many techniques have been discovered. Data mining and machine learning one of the two techniques used for solving credit card fraud detection. Initially the data set will be taken from a source and then cleaning and validation will be performed to remove redundancy. It is then randomly partitioned into k equal sized sub samples. From these k sub samples models will be created for decision tree, random forest, logistic regression and later on a comparison is made among sensitivity, specificity, accuracy and precision. It is found that random forest shows the most precise and high accuracy with the data set provided which is equal into 98.6%. [23]

*S. Benson Edwin Raj, A. Annie Portia* in “*Analysis on Credit Card Fraud Detection Methods*” the rise in e-commerce has increased the number of credit card frauds. Fuzzy Darwinian detection uses genetic programming to develop fuzzy logic rules for classifying suspicious and non suspicious. This system can attain



good accuracy and produces a low false alarm. When this was compared with several other algorithm like fusion approach using Dempster-Shafer theory and Bayesian learning, BLAST-SSAHA, Hidden Markov model, with 100% accuracy Fuzzy Darwinian system was concluded to give the best results in detecting fraudulent transactions.[25]

*Sahayasakila.V, D. Kavya Monisha, Aishwarya, Sikkakolli VenkatavisalakshiseshasaiYasaswi* in “Credit Card Fraud Detection System using Smote Technique and Whale Optimization Algorithm” mainly discusses about whale optimization algorithm and smote (synthetic minority oversampling technique) algorithm. Smote is used to solve class imbalance and from real instances it shrinks out the minority instances. Whale optimization is having more efficiency and it works in a pattern of searching a prey, encircling the prey, bubble net scratch around behaviors of whale. The author concludes by telling that Whale swarm optimization algorithm is better to obtain the ideal value by using BP network to find the errors in the transactions data set. [27]

*Dushyant Singh, Saubhagya Vardhan, Dr. (Mrs.) Neha Agrawal* in “Credit Card Fraud Detection Analysis” analyses the credit card transaction datasets by making histograms and co-relation matrix of the variables. Certain Pattern is made out from the analysis to design and implement a new fraud detection algorithm. Local Outlier Factor Machine Learning technique is used to implement the algorithm and the results shows the precision of the algorithm is very high but the accuracy remains low. [29]

*Ibtissam Benchaji, Samira Douzi, Bouabid El Ouahidi* in “Using Genetic Algorithm to Improve Classification of Imbalanced Datasets for credit card fraud detection” has mentioned that developing a fraudulent detection system is quite a difficult challenge. It is difficult to find the minority class objects that are present in skewed data sets. The two algorithms discussed here are SMOTE and Ensemble learning. [31]

### 3. Credit Card Fraud Detection Overview

Machine Learning is a widely used as an application of Artificial Intelligence using which machine automatically learns and tries to improve from experience without being programmed explicitly. Main focus is always on the development of new computer programs that is ready to access data and use it to learn for them. This is seen to be used in a number of systems to detect whether anything wrong is happening or not.

The system that is taken here is Credit Card Fraud Detection System where a number of fraudulent transactions occur and it is quite difficult to detect how and when these situations arise. For the prediction of the number of transactions occurring, different algorithms are used. The simplest way to detect these kinds of fraudulent transactions is to monitor the behavior of the card holder and then establish individual customer profiles. This helps us to observe the pattern and conclude which is standard operating procedure for cardholders and which are those ones that require scrutiny.

As the E-Commerce sites increased there has been a successive increase in the use of credit cards for online shopping. As this increased the number of frauds related to credit cards also increased

As a first step one has to collect different data which is transactions in this case. It needs to be pre processed before it is taken for analysis. In this stage the data is closely observed and the pattern of occurrence is figured out. If there is a huge deviation from the actual pattern then surely something wrong has happened and we need to detect them fast and solve it as soon as possible. For doing this there are several algorithms use which falls under either supervised or unsupervised methods.

A classifier can be used here which is an algorithm that is used for implementing classification process especially in concrete implementation. It in turn refers to a mathematical function implemented by an algorithm maps input data into a suitable category. It is thus an instance of supervised learning in which training set of properly identified observations are available. [4]

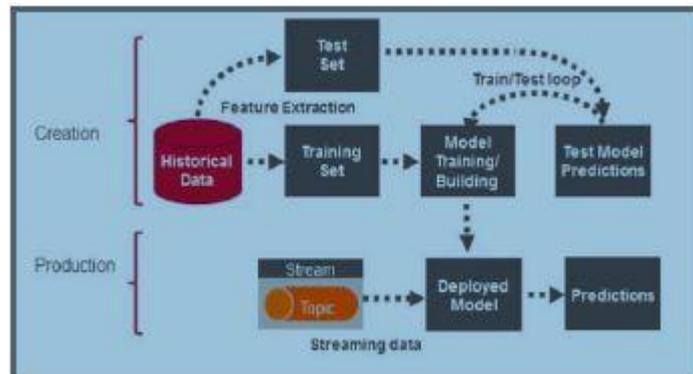


Fig1: Classifier Steps

### 3.1

#### [a] Advantages of Credit Card Fraud Detection

1. Lost as well as stolen card feature will be easier for stopping fraudulent transactions
2. Credit card validation detects errors as a sequence of numbers and thus easily detects valid and invalid Numbers.

#### [b] Disadvantages of Credit Card Fraud Detection

1. **Imbalanced data:** The credit card fraud detection data always has imbalanced nature that means very small percentages of the credit card transactions are found to be fraudulent. So it makes their detection imprecise and difficult.
2. **No usage of standard metrics:** A standard evaluation criterion that is used for the assessment and comparison of the results for fraud detection is not proper.

### 3.2 Algorithms used for Detection

For any algorithm the basic steps are follows:

- Step 1: Reading the dataset.
- Step 2: Random Sampling is done on the data set so that it is balanced.
- Step 3: Dividing the dataset into two sets namely Train dataset and Test dataset.
- Step 4: Feature selection is then applied for the proposed models.
- Step 5: Accuracy and performance metrics will be calculated to know the efficiency of various algorithms.
- Step6: Finally retrieving the best algorithm based on efficiency for the given dataset.

#### 3.2.1 Local Outlier Factor Algorithm

**Local Outlier factor (LOF)** is an algorithm which is mainly used for anomaly detection. It finds the data points that are anomalous by calculating the local deviation of a given point with respect to its neighboring data points. It is based purely on the concept of local density, where locality can be defined by the closest neighbors, whose distance will be used to figure the density. Making such a comparison between the object's local densities to the neighbor's local densities, one can easily identify those regions of similar density, and those points that have comparatively lower density when compared to their neighbors. Those will be considered as outliers.

One can use either supervised learning or unsupervised learning method for this process. Unsupervised methods does not need any prior knowledge of which transactions are fraudulent and which are non fraudulent, but instead they will detect changes in behavior or any sort of unusual transactions. In supervised methods, the

models will be trained to distinguish between fraudulent and non-fraudulent behavior to add new observations to classes. Supervised methods also require exact identification of fraudulent transactions in the databases and can be used for detecting frauds of a type which have been already occurred. An advantage of making use of unsupervised methods method is because those undiscovered frauds which were not detected before may be detected. They are also trained to differentiate between previously found fraud and those legitimate transactions.

For this we make use of any algorithm named as SODRNN, which stands for Stream Outlier Detection based on Reverse k-Nearest Neighbors. This algorithm consists of two processes: Stream Manager and Query Manager, along with that there is an entire window that is to be allocated in memory. Stream Manager is responsible for receiving the incoming stream data objects and it in turn updates the window. When new objects come they only need to update knnlist and rknnlist of  $n$  those influenced objects in that current window. When a new object is inserted, there should be single pass of scan to the current window to figure out all objects whose  $k$  nearest neighbors will get influenced. Kknnlists of these influenced objects will then update their rknnlists in a single stretch. When user demands some query of the top  $n$  outliers, the Query Manager will scan the current window and return those  $n$  objects whose  $RNNk(p)$  is small as the outliers of the given query.

### 3.2.2 Support Vector Machine

SVM is one of the most popular machine learning algorithm used for regression .SVM algorithm is a supervised machine learning algorithm that is applied for doing anomaly detection in the one-class setting Such techniques use one class learning techniques for SVM and they learn a region which consist of the training data instances The basic idea of this classification algorithm is the construction of a hyper plane in the form of a decision plane which marks the distance between the positive and negative mode maximum . The advantage of SVM mainly comes from two important properties that they possess - kernel representation and margin optimization. Kernels are mainly used to learn complex regions that are present. A kernel function defines the dot product of projections of the different data points in a high dimensional feature space. In SVMs, the classification function defines a hyper-plane separating the different classes of data. The basic technique is to find the smallest hyper sphere that is present in the kernel space which contains all training instances, and then tries to determine on which side of hyper sphere a test instance may lay. Algorithm finds a special kind of linear model, followed by a maximum margin hyper plane, and then it classifies all these training instances correctly by dividing them into correct classes through a hyper plane.

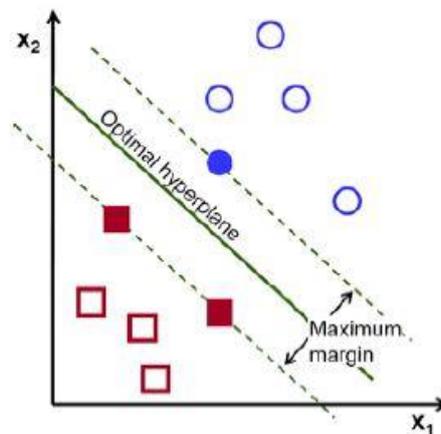


Fig2: SVM Model Graph

In SVM algorithm, plotting of graph is done when each data item is considered as a point in  $n$ -dimensional space where  $n$  is considered as the number of features, where the value of feature is equivalent to the value of a particular coordinate. Then, classification will be performed by locating the hyper-plane that separates these two classes properly.



### 3.2.3 Random Forest

Random Forest is an algorithm which is used for regression and classification. In short it is a set of decision tree classifiers. Random forest has priority over decision tree because it does not do over fitting to its training set. A subset of these training set is considered and is sampled randomly to train every individual tree and later on the decision tree will be built, and each node then splitting on a feature that is selected from a random subset of the complete feature set. Even for bigger data sets with several features and data instances training can be extremely fast in random forest since each tree is trained independently of the remaining ones. The Random Forest algorithm is found to provide a very good estimate of the generalization error.

Bagging Algorithm is used for creating random samples. Data set DS1 is given for n rows and m columns and another data set DS2 is created for sampling n cases at random that includes replacement from the original data. From dataset DS1, 1/3rd of rows are left out and this is called as Out of Bag samples. A new dataset DS2 is then trained to this model and Out of Bag samples will be used for determining unbiased estimate of the occurring error. Out of these m columns,  $M \ll m$  columns is then selected at each node in the data set. The M columns are in turn selected at random. Usually, by default, choice of M, is  $m/3$  for regression tree and M will be  $\sqrt{m}$  for classification tree. There is no pruning in random forest i.e., each tree is grown completely. The Accuracy of random tree is always considered better when compared with the remaining set of algorithms.

### 3.2.4 Naïve Bayes

Naive Bayes is a classification technique based on Bayes' Theorem assuming the independence among predictors. This is a very simple as well as powerful algorithm. In simple terms, a Naive Bayes classifier tries to assume the presence of a particular feature within a class is being unrelated to the presence of any other feature within the same class. This model is particularly useful for very large data sets and also easy to build. Along with the feature of simplicity, Naive Bayes is known to outperform even highly sophisticated classification. Bayes theorem is all about finding probability of event occurring given probability of another event that has been already occurred.

$$P(A/B) = (P(B/A) P(A)) / P(B)$$

Where, P(A) – Priority of A and

P(B) – Priority of B and

P(A/B) – Posteriori priority of B

Naïve Bayes algorithm quite fast. This algorithm reacquires less number of training data.

## 4. Conclusion

At its most basic level, machine learning refers to any type of computer program that can “learn” by itself without having to be explicitly programmed by a human. Using different machine learning algorithm on can thus figure out how different systems work and based upon that they can perform operations on the. Here we have taken Credit Card Fraud Detection. Different algorithms have been compared as it was found that Random Forest is the most suitable algorithm in terms of efficiency and precision and is the most preferred algorithm.



## References

- [1] Analysis of Credit Card Fraud Detection Techniques: based on Certain Design Criteria Masoumeh Zareapoor, Seeja.K.R, M.Afshar.Alam *International Journal of Computer Applications (0975 – 8887) Volume 52– No.3, August 2012.*
- [2] Credit Card Fraud Detection Using Self-Organizing maps, Vldamir Zaslavsky, Anna sthrizak.
- [3] INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY- DETECTION AND PREDICTION OF CREDIT CARD FRAUD TRANSACTIONS USING MACHINE LEARNING Kaithekuzhical Leena Kurien\*1 & Dr. Ajeet Chikkamannur2 ISSN: 2277-9655, Impact Factor: 5.164.
- [4] International Journal of Pure and Applied Mathematics Volume 118 No. 20 2018, 825-838 -Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models 1Navanshu Khare and 2Saad Yunus Sait.
- [5] Fraud Detection using Machine Learning Aditya Oza.
- [6] Data Mining-based Fraud Detection Research CLIFTON PHUA1\*, VINCENT LEE1, KATE SMITH1 & ROSS GAYLER2.
- [7] A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective.SamanehSorournejad1, Zahra Zojaji2, Reza Ebrahimi Atani3, Amir Hassan Monadjemi4.
- [8] A review of Fraud Detection Techniques: Credit Card Khyati Chaudhary, Jyoti Yadav Bhawna Mallick *International Journal of Computer Applications (0975 – 8887) Volume 45– No.1, May 2012.*
- [9] DETECTING CREDIT CARD FRAUD USING MACHINE LEARNING by Lukas Frei.
- [10] Credit card Fraud Detection based on Machine Learning Algorithms, Heta Naik, Prashasti Kanikar *International Journal of Computer Applications (0975 – 8887) Volume 182 – No. 44, March 2019.*
- [11] International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 1, January 2016 Copyright to IJARCCCE DOI 10.17148/IJARCCCE.2016.5109 39, Credit Card Fraud Detection Ishu Trivedi1, Monika2, Mrigya Mridushi3.
- [12] Banks and Bank Systems, Volume 4, Issue 2, 2009 Linda Delamaire (UK), Hussein Abdou (UK), John Pointon (UK) Credit card fraud and detection techniques: review.
- [13] CREDIT CARD FRAUD DETECTION BASED ON BEHAVIOR MINING NIMISHA PHILIP1, SHERLY K.K2.
- [14] Credit card fraud detection using anti-k nearest neighbour algorithm VENKATA RATNAM GANJI, SIVA NAGA PRASAD MANNEM, Venkata Ratnam Ganji et al. / International Journal on Computer Science and Engineering (IJCSSE).
- [15] International Journal of Computer Science and Mobile Computing A Monthly Journal of Computer Science and Information Technology ISSN 2320–088X IJCSMC, Vol. 3, Issue. 4, April 2014, pg.780 A Revived Survey of Various Credit Card Fraud Detection Techniques Neha Sethi1, Anju Gera2.
- [16] Distributed Data Mining in Credit Card Fraud Detection Philip K. Chan, Florida Institute of Technology Wei Fan, Andreas L. Prodromidis, and Salvatore J. Stolfo, ColumbiaUniversity.
- [17] International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 11, November 2012) Survey on Credit Card Fraud Detection Methods Krishna Kumar Tripathi1, Mahesh A. Pavaskar2.
- [18] International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-5S2, January 2019 A Comparative Analysis of Various Credit Card Fraud Detection Techniques Yashvi Jain, NamrataTiwari, ShripriyaDubey, Sarika Jain.
- [19] International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 24 (2018) pp. 16819-16824 © Research India Publications. <http://www.ripublication.com> Machine Learning for Credit Card Fraud Detection System Lakshmi S V S S1, Selvani Deepthi Kavila2.
- [20] International Journal of Computer Trends and Technology (IJCTT) – volume 4 Issue 7–July 2013 ISSN: 2231-2803 <http://www.ijcttjournal.org> Page 2206 Review Paper on Credit Card Fraud Detection 1Suman 2Nutan.
- [21] Machine Learning 2: 195 198, 1987 © 1987 Kluwer Academic- Publishers, Boston Manufactured in The Netherlands EDITORIAL Research Papers in Machine Learning.



Subhadra S Menon *et al*, International Journal of Computer Science and Mobile Applications,  
Vol.8 Issue. 3, March- 2020, pg. 15-23

**ISSN: 2321-8363**

**Impact Factor: 5.515**

- [22] International Journal of Computer Science and Mobile Computing A Monthly Journal of Computer Science and Information Technology ISSN 2320-088X IMPACT FACTOR: 6.199 *IJCSMC*, Vol. 8, Issue. 8, August 2019, pg.43 – 49 Credit Card Fraud Detection Techniques: A Review Sonal Mehndiratta, Mr. Kamal Gupta.
- [23] International Journal of Pure and Applied Mathematics Volume 118 No. 20 2018, 82 5-838, ISSN: 1314-3395 (on-line version) Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models Navanshu Khare and Saad Yunus Sait.
- [24] International Conference on Computer, Communication and Electrical Technology – ICCET2011, 18th & 19th March, 2011 978-1-4244-9394-4/11/\$26.00 ©2011 IEEE 152 Analysis on Credit Card Fraud Detection Methods S. Benson Edwin Raj, A. Annie Portia.
- [25] Periodicals of Engineering and Natural Sciences ISSN 2303-4521 Vol.6, No.2, December 2018, pp.394-400 Using Isolation Forest in anomaly detection: the case of credit card transactions Soumaya Ounacer, Hicham Ait El Bour, Younes Oubrahim, Mohamed Yassine Ghomari and Mohamed Azzouazi.
- [26] 18th International Symposium INFOTEH-JAHORINA, 20-22 March 2019 978-1-5386-7073-6/19/\$31.00 ©2019 IEEE Credit Card Fraud Detection - Machine Learning methods Dejan Varmedja, Mirjana Karanovic, Srdjan Sladojevic, Marko Arsenovic, Andras Anderla.
- [27] International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249-8958, Volume-8 Issue-5, June 2019 190 *Published By: Blue Eyes Intelligence Engineering & Sciences Publication Retrieval Number D6468048419/19©BEIESP*, Credit Card Fraud Detection System using Smote Technique and Whale Optimization Algorithm Sahayasakila.V, D. Kavya Monisha, Aishwarya, Sikhakolli VenkatavisalakshiseshsaiYasaswi.
- [28] International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 24 (2018) pp. 16819-16824 © Research India Publications. <http://www.ripublication.com> Machine Learning for Credit Card Fraud Detection System Lakshmi S V S S, Selvani Deepthi Kavila.
- [29] International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 05 Issue: 11 | Nov 2018 [www.irjet.net](http://www.irjet.net) p-ISSN: 2395-0072 © 2018, IRJET | Impact Factor value: 7.211 | ISO 9001:2008 Certified Journal | Page 1600, Credit Card Fraud Detection Analysis Dushyant Singh, Saubhagya Vardhan, Dr. (Mrs.) Neha Agrawal.
- [30] Real-time Credit Card Fraud Detection Using Machine Learning Anuruddha Thennakoon<sup>1</sup>, Chee Bhagyani<sup>2</sup>, Sasitha Premadasa<sup>3</sup>, Shalitha Mihiranga<sup>4</sup>, Nuwan Kuruwitaarachchi<sup>5</sup>.
- [31] Using Genetic Algorithm to Improve Classification of Imbalanced Datasets for credit card fraud detection Ibtissam Benchaji, Samira Douzi, Bouabid El Ouahidi.