

A Study on Credit Card Fraud Detection using Machine Learning

Ajayi Kemi Patience, Dr. Lakshmi J. V. N

Department of Masters of Computer Applications,
JAIN (Deemed-to-be University), Bengaluru, Karnataka, India

ABSTRACT

Due to the high level of growth in each number of transactions done using credit card has led to high rise in fraudulent activities. Fraud is one of the major issues related to credit card business, since each individual do more of offline or online purchase of product via internet there is need to developed a secured approach of detecting if the credit card been used is a fraudulent transaction or not. Pattern involves in the fraud detection has to be re-analyze to change from reactive approach to a proactive approach. In this paper, our objectives are to detect at least 95% of fraudulent activities using machine learning to deployed anomaly detection system such as logistic regression, k-nearest neighbor and support vector machine algorithm.

KEYWORDS: Credit Card, Fraud Detection, Machine Learning, Logistic Regression, K-Nearest Neighbor, Support Vector Machine Algorithm

How to cite this paper: Ajayi Kemi Patience | Dr. Lakshmi J. V. N "A Study on Credit Card Fraud Detection using Machine Learning" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-4 | Issue-3, April 2020, pp.801-804, URL: www.ijtsrd.com/papers/ijtsrd30688.pdf



IJTSRD30688

Copyright © 2020 by author(s) and International Journal of Trend in Scientific Research and Development Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) (<http://creativecommons.org/licenses/by/4.0>)



1. INTRODUCTION

In recent years banking sector as be paramount among services offered to customer by its bank which involves in credit card facility that is offered. However, credit card issues as increase drastically in the aspect of security threat such as, theft of credit card used in fraudulent transaction or using method of hacking people credit card. Early detection of credit card improves and enable protection of each transaction done, database size constantly changing which are vital information needs to develop effectively algorithm to help analyze each activity involves, the data stream analysis can be done.

Fraud is an unauthorized use of account that belong to someone that is not the owner of the account. Fraud detection involves in various method used to prevent activities of unauthorized usage of credit card by using detection methodology. A fraud detection system (FDS) should effectively and efficiently detect fraudulent used of

credit card transaction. The theory in credit card detection is either fraudulent transaction or legitimate transaction, our propose system model will predict fraud activities and reduce false positive and false negative hypothesis.

However, the common use of online shopping and ecommerce by entering the Card verification value (CVV), customer credit card passwords and vital information are at risk. The traditional method of detecting fraudulent behavior is replace with online fraud detection software using machine leaning algorithm. Method used are fraudulent pattern and prediction of the transaction. machine learning has successfully rate of fraud detecting using supervised algorithm, unsupervised algorithm and reinforcement. The existing system uses cluster analysis and artificial neural networks on fraud detection in this paper is to proposed credit card fraud detection system using supervised algorithm.

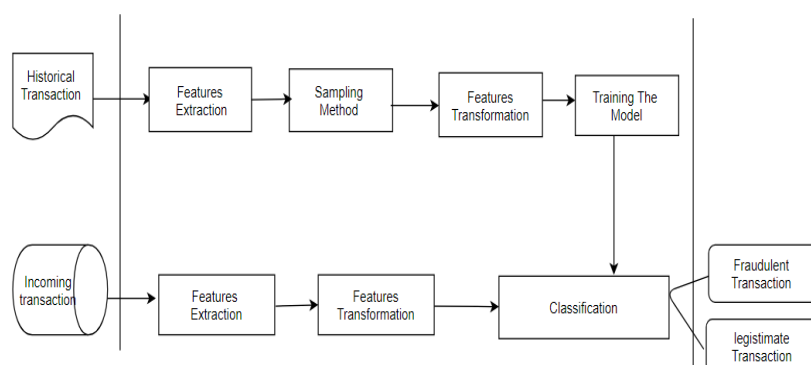


Fig. 1: Fraud Detection Process

2. RELATED WORK

1. "Machine Learning Approaches for Credit Card Fraud Detection". "S. Venkata Suryanarayana, G.N. Balaji G. Venkateswara Rao".

This paper proposes four algorithms in which are Naïve Bayes, Decision tree, K-Nearest Neighbour and support vector machine algorithm. The model was trained and tested using Waikato environment for knowledge analysis (WEKA). It was developed in java by university of Waikato in New Zealand, the approach ensures that all data that was represented once as attest data and multiple training data to produce accurate output.

2. "Detection and Prediction of Credit Card Transaction Using Machine Learning". "Kaithekuzhical Leena Kurian and Dr. Ajeet Chikkamannur".

In this study, the classification model on logistic regression and random forest are developed and applied to fraud detection. Its techniques suggest a well-suited system model that prove that the predicted fraudulent transaction is genuine. Its main objective is to detect and predict various credit card transaction used in an unauthorised way. It also states the limitation of random forest is due to multiple tree in forest which enable the algorithm to be slow.

3. "Credit Card Fraud Detection Using Machine Learning and Data Science". "Aditya Saini, Swarna Deep Sarkar, Shadab Ahmed, S P Maniraj".

This research paper focus on analysis and pre-processing of each data set by deploying anomaly detection algorithm such as local outlier factor and isolation forest algorithm. The paper suggests the approach of Jupyter notebook platform to make a program in python, this program can be executed in cloud using google collab platform which support all python notebook files.

4. "Machine Learning Based Credit Card Analysis Modelling, Detection, And Deployment". "Shivakumar Goel And Hitesh Patil".

The study represents how the combination of different clustering and machine learning algorithm can be used to developed a very large scale to detect the fraudulent transaction and use to ensure that credibility of the payment system. This paper analyses the factor mapping use to connect different factor that mainly associated with credit card. After then pattern matching model using clustering algorithm and lastly prediction model using artificial neural network (ANN).

5. "Credit Card Fraud Detection Using Random Forest" "Devi Meenakshi. B, Janani. B, Gayathri .S , Mrs Indira N"

There is a phenomenal growth in the number of credit card transaction, the paper applies the use of random forest algorithm for classification of credit card dataset. The algorithm estimates the general error and to be resistant to over fitting, the propose system rank the important of variable in a regression or classification problem. Random forest is an advanced version of decision tree, the study explains the tree de-correlated and prune the tree by fixing a stopping criterion for node split.

3. METHODOLOGY

Fraud detection is a classification task use to predict fraud or legit transaction. this approach uses machine learning algorithm to detect anomaly transaction which are logistic regression, k-nearest neighbor, support vector machine algorithm. The performance of the algorithm is compare based on specificity, accuracy and precision. The diagram below represents the system architecture.

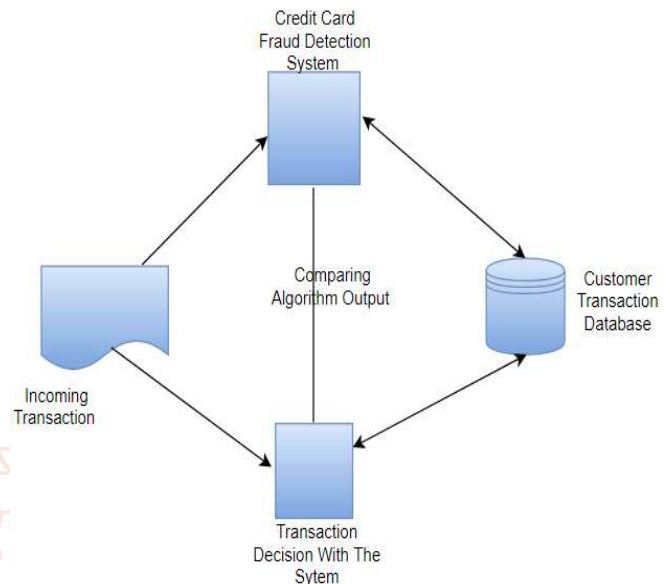


Fig. 2: System Architecture

3.1. CLASSIFICATION TECHNIQUES

1. Logistic regression

Logistic regression is a supervised classification algorithm based on the decision that brought into the pictures. The aspect of regression is relying on classification of problems. It used dichotomous (binary) variables such as female/ male or true/ false, there are two values labeled "0" and "1". It also estimates the relationship between dependent variables to one or more independent variables using sigmoid function also called logistic function. In classification based of logistic regression can only take discrete values for a set of elements(or features) X & Y, it deals with the argument using the threshold which are, high recall/ low precision this reduce the total number of false negative without reducing the total number of false positive while low recall/high precision this reduce the total number of false positive without reducing number of false positive. Logistic regression categories are based on binomial, ordinal and multinomial.

Application of logistic regression

logistic regression helps in medical field, it predicts the probability of risk of patient in developing a disease. In marketing logistic regression help the customer purchase property and it will predict if the value of the property will rise or not.

2. K-Nearest Neighbor (KNN)

K-nearest neighbor is a machine learning algorithm that is used by regression and classification problem also the output depends if (KNN) is used for either regression or classification. Some examples are lazy learning, instance base learning. Each function is locally and all calculation is delay until each of the function is evaluated.

KNN performance is influence using three factors:

- Number of neighbor use to arrange each new sample.
- Distance metric used to discover the nearest neighbor.
- Distance act use to extract a classified k-nearest neighbor.

Some of the dataset is generated for training set use by the algorithm but no exact training step is followed.

Application of K-Nearest Neighbor

K-nearest neighbor is used in terms of finance it includes currency exchange rate, stock market planning investment strategies. It's also used for medical prediction for analysis of micro array gene (KNN) is mostly popular algorithm that can easily categories text or text mining

3. Support vector machine

Support vector machine it analyzes each data used for regression and classification. It effectively uses a non-linear classification called kernel trick.

4. EXPERIMENTS ANALYSIS AND COMPARISON

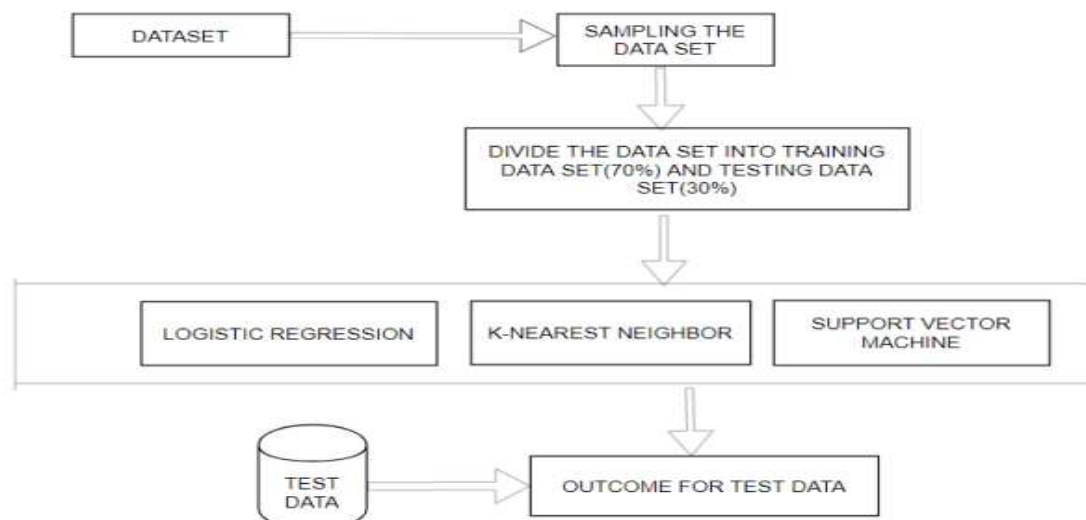


Fig. 3: System Analysis

Dataset

Firstly, our dataset was obtained from Kaggle in order to import the dataset and convert the data into data frames format. Our dataset has 12 attributes features placed in the dataset that was obtained and do random sampling of each data.

Training and testing data

Testing is done on each dataset that was provided after training process occur. Giving 70% data training and remaining 30% testing dataset, the outcome of test will then compare the algorithm with the accuracy and performance of each transaction.

Applying algorithm

We demonstrate the performance of each algorithm mention above. The following classification of algorithm use to train and test dataset, the metrics uses confusion matrix, precision and recall score.

A. logistic regression

The performance of logistic regression is described as below:

Cross validation mean score of logistic regression: 97.2%

The concept of SVM involve are:

- **Margin** it is the gap along two lines on the short distance point of dissimilar classes.
- **Support vector data** point that are short distance to the hyperplane is called support vector.
- **Hyperplane** it is a decision space which is along a set of objects having dissimilar classes.

Outline by separate hyperplane if given a training data the output from the algorithm is optimal hyperplane.

Application of Support vector machine

Support vector machine can recognize handwritten characters and help in hypertext categories to reduce the use of hand label training. (SVM) algorithm can classified images and also used by scientist and biologist.

Model accuracy 97%

Confusion matrix

[273 5]
[4 42]

Table 1: Classification report for logistic regression:

	precision	recall	F1score	support
0	1.00	1.00	1.00	273
1	1.00	1.00	1.00	42
Accuracy			1.00	324
Macro	1.00	1.00	1.00	324
Weighted avg	1.00	1.00	1.00	324

B. K- nearest neighbor

The performance of K- nearest neighbor is describe as below:

Cross validation mean score of K-Nearest Neighbor: 96.5%

Model accuracy 97%

Confusion matrix

[272 5]
[5 42]

Table 2: Classification report for K- nearest neighbor

	precision	recall	F1score	support
0	0.84	1.00	0.95	272
1	1.00	0.84	0.89	42
Accuracy			0.94	324
Macro	1.00	1.00	0.94	324
Weighted avg	1.00	1.00	0.94	324

C. Support vector machine

The performance of Support vector machine is described as below:

Cross validation mean score of Support vector machine: 98%

Model accuracy 98%

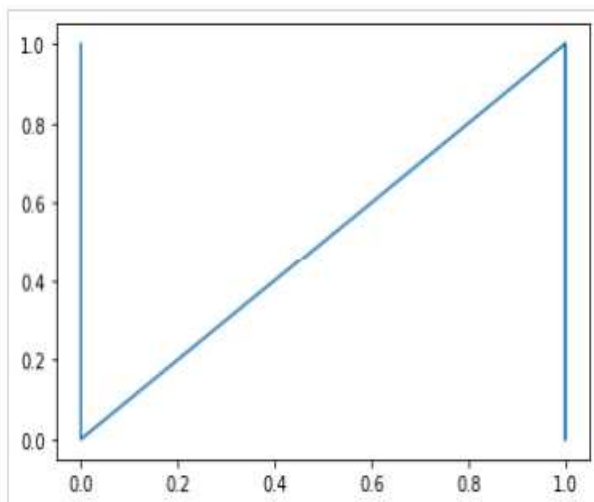
Confusion matrix

[276 7]

[1 40]

Table 3: Classification report for Support vector machine:

	precision	recall	F1score	support
0	1.00	1.00	0.95	276
1	0.89	0.84	0.89	40
Accuracy			0.90	324
Macro	0.90	0.92	0.90	324
Weighted avg	0.90	0.92	0.90	324

**Fig. 4: ROC Curve for all the three models.**

5. FUTURE SCOPE

This project enables us to analyze using logistic regression, k-nearest neighbor, support vector machine algorithm for credit card fraud detection technique. In the further

enhancement of the project we can use neural network with combination of Naïve-Bayes may help to detect anomaly behavior in fraud detection.

6. CONCLUSION

Credit card monitoring is an essential task for merchant bank which involve the customer and its bank, effective need for better improvement by using the combination of one or more algorithm. This paper has explained in detail the involvement of machine learning in use for fraud detection alongside with its algorithm. Analysis and comparison of the three algorithm and also enumerate the various application of each algorithm.

REFERENCE

- [1] "Machine learning approaches for credit card fraud detection". "S. Venkata Suryanarayana, G. N. Balaji G. Venkateswara Rao". International journal of engineering and technology (IJET) 7(2) PP. 917-920 (2018).
- [2] "Detection and prediction of credit card transaction using machine learning". "Kaithekuzhical Leena Kurian and Dr. Ajeet Chikkamannur". International journal of engineering science and research technology (IJESRT): 8(3) ISSN: 277-9655 (2019).
- [3] "Credit card fraud detection using machine learning and data science". "Aditya Saini, Swarna Deep Sarkar, Shadab Ahmed, S P Maniraj". International journal of engineering research and technology (IJERT) vol:8 issue 09 ISSN: 2278-0181(2019).
- [4] "Machine learning based credit card analysis modelling, detection, and deployment". "Shivkumar Goel and Hitesh Patil" International journal of advanced research (IJAR) ISSN: 2320-5407 (2017).
- [5] "Credit card fraud detection using Random forest" "Devi Meenakshi. B, Janani. B, Gayathri. S, Mrs Indira.N" International research journal of engineering and technology (IRJET) VOL. 06 issue: 03 ISSN 2395-0072. (2019)
- [6] "Detecting Credit Card Fraud by Decision Trees and Support Vector Machines". Hong Kong, China: The International Multi Conference of Engineers and Computer Scientists. Sahin, Y., & Duman, E. (2011).
- [7] "A Comparative Study of Credit Card Fraud Detection Using Machine Learning for United Kingdom Dataset". "V. Sandeep, T. Supriya, P. Hiranya, Dr.M.SUJATHA" International Journal of Computer Science and Information Security (IJCSIS), Vol. 17, No. 9, September (2019).
- [8] "Credit card fraud detection using anti-K-Nearest Neighbour algorithm" by Venkata Ganji,Siva Naga Prasad Mannem International Journal on Computer Science and Engineering(IJCSE), Vol.4 No. 06 June(2015).