Volume No:4, Issue No:10 (March-2019) ISSN No : 2454-423X (Online)

International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

Credit Card Fraud Detection System Using Machine Learning Techniques

J. Devi Sushma
Department of Computer Science & Engineering,
NS Raju Institute of Technology,
Visakhapatnam, Andhra Pradesh-531173, India.

M. Hemanthkumar
Department of Computer Science & Engineering,
NS Raju Institute of Technology,
Visakhapatnam, Andhra Pradesh-531173, India.

P.H.V.Raviteja
Department of Computer Science & Engineering,
NS Raju Institute of Technology,
Visakhapatnam, Andhra Pradesh-531173, India.

P.Vinay
Department of Computer Science & Engineering,
NS Raju Institute of Technology,
Visakhapatnam, Andhra Pradesh-531173, India.

P.V.S.Prabhakar
Department of Computer Science & Engineering,
NS Raju Institute of Technology,
Visakhapatnam, Andhra Pradesh-531173, India.

Abstract

In this research, a technique for `Credit Card Fraud Detection' is developed. As fraudsters are increasing day by day. And fallacious transactions are done by the credit card and there are various types of fraud. So to solve this problem combination of technique is used like Genetic Algorithm, Behavior Based Technique and Hidden Markov Model. By this transaction is tested individually and whatever suits the best is further proceeded. And the foremost goal is to detect fraud by filtering the above techniques to get better result. Fraud is one of the major ethical issues in the credit card industry. The main aims are, firstly, to identify the different types of credit card fraud, and, secondly, to review alternative techniques that have been used in fraud detection. The sub-aim is to present, compare and analyze recently published findings in credit card fraud detection. This article defines common terms in credit card fraud and highlights key statistics and figures in this field. Depending on the type of fraud faced by banks or credit card companies, various measures can be adopted and implemented. The proposals made in this paper are likely to have beneficial attributes in terms of cost savings and time efficiency. The significance of the application of the techniques reviewed here is in the minimization of credit card fraud. Yet there are still ethical issues when genuine credit card customers are misclassified as fraudulent. By using logistic regression model, the number of fraudulent and normal transactions based on the dataset get the classification report and accuracy.

Keywords:
Credit card fraud, Fraud detection technique, Data mining and Logistic Regression method

1. Introduction

Credit card is a small plastic card issued by a bank, building society, etc., allowing the holder to purchase goods or services on credit [1]. It allows you to borrow money from a bank to make purchases, whether you're buying a burger or a round-trip ticket to London [2]. As long as you pay back the money you borrow within the "grace period" of 25-30 days; you don't have to pay extra [3]. If you don't pay it back in that time period, you'll have to pay interest i.e., "a percentage of the money you owe the bank" on top of what you borrowed.

Cite this article as: J.Devi Sushma, M.Hemanthkumar, P.H.V.Raviteja, P.Vinay & P.V.S.Prabhakar, "Credit Card Fraud Detection System Using Machine Learning Techniques", International Journal of Research in Advanced Computer Science Engineering, Volume 4, Issue 10, 2019, Page 26-30.

By using credit cards, the purchases are made easy [4]. It can also be useful in times of emergencies like, healthcare. Some credit cards offer additional benefits, such as discounts from particular stores or companies, bonuses such as free airline miles or travel discounts, and special insurances[5-8]. It also provides additional features incase if the card is damaged or lost. In these cases, the card can be blocked and the transactions are stopped. The biggest disadvantage of credit cards is that they encourage people to spend money that they don't have. Like cash, sometimes credit cards can be stolen [9-10]. They may be physically stolen (if you lose your wallet) or someone may steal your credit card number (from a receipt, over the phone, or from a Web site) and use your card to rack up debts [11]. There are many situations in which the transactions made using the credit card to be fraud. In this project, we predict and classify the credit card transactions as fraud and not fraud transactions.

2. Objective of the project

There is a substantial rise in some technologies like “machine learning, artificial intelligence, deep learning” and other relevant fields of information technology. These technologies help us automate the credit card fraud classification process. Automation saved a huge amount of time and work in detecting the fraudulent transactions present in the dataset.

Module1:

Part1: Reading the dataset from the file “creditcard.csv” which has 2,84,807 transactions.

Part2: In order to get all variables in an equivalent range, we subtract the mean and divide by the standard deviation such that the distribution of the values is normalized.

Module2:

Here we perform normalization. In order to get all variables in an equivalent range, we subtract the mean and divide by the standard deviation such that the distribution of the values is normalized.

Module3:

Plotting of dataset is done in this module.

Module4:

In this module, we first define some models like the “Logistic Regression, Gaussian Naïve-Bayes and Decision Tree Classifier”, and then loop through a training and testing set. First, we train the model by the training set and then validate the results with the testing set.

3. Limitations of the project

Credit card must be used by the authorized users. The people cannot use the credit cards without the permission of the owner of the credit card. The permissions may be violated in case it is stolen.

3.1 Existing system

In the past, this was done by employees, who checked all transactions manually. Earlier, fraudulent transactions were detected using the outlier detection in which we find the critical anomalies present in the data. In data mining, anomaly detection (also outlier detection) is the identification of items, events or observations which do not conform to an expected pattern or other items in a dataset. There are few more methods like decision tree, cluster techniques and neural networks etc

3.2 Disadvantages of existing system

The main disadvantage of the existing technique is that it is difficult to find the correct number of clusters we want We could run different algorithm first to see the distance between any number of clusters (how discriminate are they to each other). We also tend to lose information from heavy outliers since those have to be put in a cluster as well. Statistical fraud detection problem is a very difficult problem in that there are very few examples of fraud.

The great majority of transactions are legitimate, which makes the fraud detection difficult.

3.3 Proposed system

The performance of fraud detection in credit card transactions is greatly affected by the sampling approach on dataset, selection of variables and detection technique(s) used. The Credit Card Fraud Detection Problem includes modeling past credit card transactions with the knowledge of the ones that turned out to be fraud. Hence we are using the technique of machine learning for fraud detection. In this we take the real bank dataset and split the dataset into training set and testing set and then apply the Logistic Regression method.

4. System Analysis

A Software Requirements Specification (SRS) – a requirements specification for a software system – is a complete description of the behavior of a system to be developed. It includes a set of use cases that describe all the interactions the users will have with the software. In addition to use cases, the SRS also contains non-functional requirements. Non-functional requirements are requirements which impose constraints on the design or implementation (such as performance engineering requirements, quality standards, or design constraints).

4.1 Software requirement specification

It is a complete description of the system that is to be developed. It includes a set of use cases that describes all the interactions the users will have with the software. Use cases that describe all the interactions the users will have with the software. Use cases are also known as functional requirements. In addition to the use cases, the SRS also contains non-functional requirements. In addition to use cases, there are also functional (or supplementary) requirements. Non-functional requirements are requirements which impose constraints on the design or implementation (such as performance engineering supplementary) requirements. Non-functional requirements, quality standards, or design constraints).

a) **Operating system:** An operating system (OS) is software that manages computer hardware and software resources and provides common services for computer programs. Operating system used is windows 10 with 64-bit operating system. Operating System: Windows 10 with 32-bit

b) **Programming Language:** A programming language is a formal language that
specifies a set of instructions that can be used to produce various kinds of output. Programming languages generally consist of instructions for a computer. Programming languages can be used to create programs that implement specific algorithms.

c) **Hard Drives:** Hard drives store all of your computer information once your system is turned off. Hard drive of 4GB is minimum required.

d) **Processor:** A processor is the logic circuitry that responds to and processes the basic instructions that drive a computer. Processor being used is Intel® Core i5 @ 2.20 GHz.

RAM	:	4G
Processor	:	Intel ® Core i5
Speed	:	320

5. Implementation and Results

5.1 Introduction

Credit card transactions are given to various supervised learning models in order to find the precision with which it can classify the fraudulent and non-fraudulent transactions.

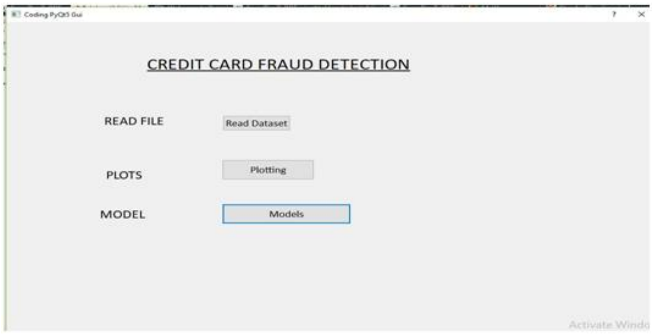
5.2 Explanation of Key Functions

5.2.1 Python

Python is a general-purpose programming language that is becoming more and more popular for doing data science. Companies worldwide are using Python to harvest insights from their data and get a competitive edge. Unlike any other Python tutorial, this course focuses on Python specifically for data science.

Advantages of Python

- Easy and fast to learn
- Simple to get support
- Python's syntax is clear and readable
- Fast to code, i.e. it is easier and faster to code a problem in Python than in C, C++ or Java, just to mention a few other languages
- Python is portable, i.e. it runs on multiple platforms and systems, like e.g. Unix, Linux and Microsoft Windows
- Object oriented


Coding PyQt5 Gui

CREDIT CARD FRAUD DETECTION

READ FILE Read Dataset

PLOTS Plotting

MODEL Models

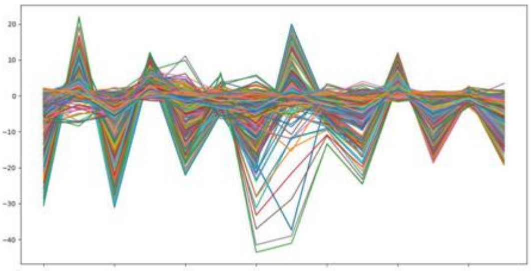


Fig. 1 Plot of features V1-V14 of fraud cases

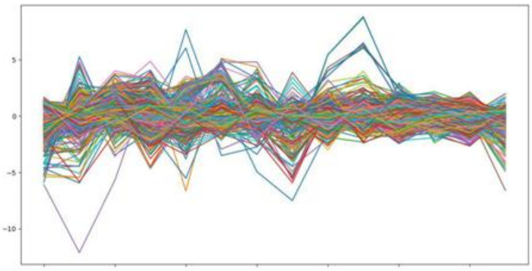


Fig. 2 Plot of features V1-V14 of normal cases

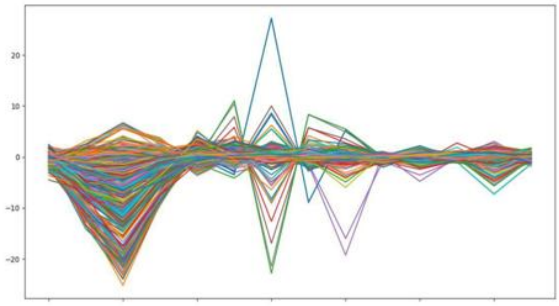


Fig. 3 Plot of features V15-V28 of fraud cases

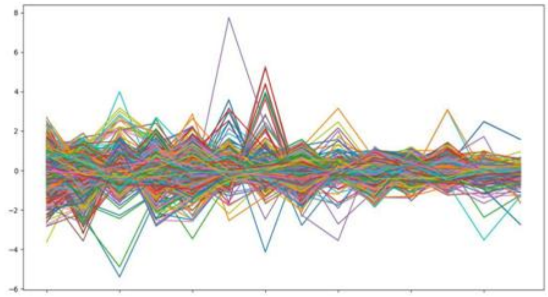


Fig. 4 Plot of features V15-V28 of normal cases

Table -1 The software requirements specification

Test No	Test case	Excepted output	Actual output	Result
1	Running the program after importing all the required libraries	The shape of the dataset, plots and the classification report with accuracy has to be printed	The shape of the dataset, plots and the classification report with accuracy are printed	Passed
2	Running the PY file without improving the Logistic Regression library	A name error has to be displayed indicating Logistic Regression is not defined	A name error is displayed indicating logistic regression is not defined	Passed
3	Running the file without creating a ui but loading the ui in the code	File not found error has to be displayed	File not found error has to be displayed	Passed
4	Running the PY file by changing the name of the attribute	Attribute error has to be displayed	Attribute error has to be displayed	Passed

6. Conclusion

- Machine learning has been recognized as a successful measure for fraud detection. A great deal of data is transferred during online transaction processes, resulting in a binary result, genuine or fraudulent.

- In this process we estimate the precision, recall, f1score and support function for the dataset on all the models. In the case of Logistic Regression the 0 classes (transactions without fraud) are predicted with 100% precision and recall and the 1 classes (transactions which are fraud) are predicted with 88% precision. This means that 12% of the transactions which are fraudulent remain undetected by the system. But, 88% is still quite good.

- In the case of Decision Tree Classifier the the 0 classes (transactions without fraud) are predicted with 100% precision and recall and the 1 classes (transactions which are fraud) are predicted with 82% precision. This means that 18% of the transactions which are fraudulent remain undetected by the system.

- In the case of Gaussian Naive Bayes the 0 classes (transactions without fraud) are predicted with 100% precision and 98% recall and the 1 classes (transactions which are fraud) are predicted with 6% precision. This means that 94% of the transactions which are fraudulent remain undetected by the system.

- According to these results it is clear that the precision with which the Logistic Regression model detects the transactions is much more than the decision tree classifier and the gaussian naive bayes.

7. References

Resources that are used for making this analysis are shown below which provided huge information for the whole process:

[1] Introduction to Data Mining: Pang-Ning Tan & Michael Steinbach, Vipin Kumar, Pearson.

[2] Credit Card Fraud Detection:A case study by Ayushi Agarwal , Shiv Kumar , Amit Kumar Mishra(IEEE paper)

[3] Credit card fraud detection using Hidden Markov Model by Divya Iyer , Arti Mohanpurkar , Sneha Janardhan, Dhanashree Rathod , Amruta Sardeshmukh (IEEE paper)

[4]https://www.youtube.com/watch?v=Z5WKQr4H4Xk&t=26 6s

[5]https://pandas.pydata.org/pandasdocs/stable/api.html#id5

[6]https://www.youtube.com/watch?time_continue=420&amp;v=XWo3nY06RgQ

[7]https://www.quora.com/What-is-logistic-regression/answer/Vinay-Kumar-R- 12? srid=3Okwy

[8] https://youtu.be/953DXVpN5_Q

[9] https://www.kaggle.com/agpickersgill/credit-card-fraud-detection/data

[10] https://en.wikipedia.org/wiki/Credit_card_fraud

[11] https://docs.python.org/2/tutorial.