Pune Institute of Computer Technology
Dhankawadi, Pune

A SEMINAR REPORT
ON

COMPARISON OF MACHINE LEARNING MODELS FOR CREDIT CARD FRAUD DETECTION

SUBMITTED BY

Vishwajeet Ekal
Roll No. 31216
Class TE2

Under the guidance of
Prof. Y.A. Handge

DEPARTMENT OF COMPUTER ENGINEERING
Academic Year 2019-20

DEPARTMENT OF COMPUTER ENGINEERING
Pune Institute of Computer Technology
Dhankawadi, Pune-43

CERTIFICATE

This is to certify that the Seminar report entitled

"COMPARISON OF MACHINE LEARNING MODELS IN CREDIT CARD FRAUD DETECTION"

Submitted by
Vishwajeet Ekal Roll No. 31216

has satisfactorily completed a seminar report under the guidance of Prof. Y.A. Handge towards the partial fulfillment of third year Computer Engineering Semester II, Academic Year 2019-20 of Savitribai Phule Pune University.

Prof. Y.A. Handge
Internal Guide

Prof. M.S.Takalikar
Head
Department of Computer Engineering

Place:
Date:

ACKNOWLEDGEMENT

I sincerely thank our Seminar Coordinator Prof. B.D.Zope and Head of Department Prof. M.S.Takalikar for their support.

I also sincerely convey my gratitude to my guide Prof. Y.A. Handge, Department of Computer Engineering for her constant support, providing all the help, motivation and encouragement from beginning till end to make this seminar a grand success.

Contents

List of Tables

List of Figures

Abstract

In Computer Engineering Machine learning is being used in a wide range of application domains to discover patterns in large datasets. Classification which is one of the most important aspects of machine learning has been widely used for credit card fraud detection.

With the development of modern technology credit cards became popular mode for cashless transactions due to ease of use and convenience. But along with that credit card fraud is also increasing considerably.Every year credit card fraud costs consumers and the financial companies billions of dollars. It is important that credit card companies are able to recognize fraudulent credit card transactions so that customers are not charged for items that they did not purchase. Supervised classification is used for this purpose.

Classification techniques widely used for fraud detection are K-Nearest Neighbor algorithms (KNN), Support Vector Machines (SVM), Naïve Bayes (NB). Above mentioned methods can either be used alone or in collaboration using meta-learning techniques to build classifiers. Here we compare the mentioned methods.

Keywords

Machine Learning, Fraud Detection, Credit Card Fraud, KNN, Naive Bayes.

1 INTRODUCTION

People use credit cards for the procurement of products and services in day-to-day life. The transactions may be offline or online. In recent years E-commerce has changed everything with rapid growth of internet. Along with this rapid increase of e-commerce, using credit card has become a convenient and necessary part of our life. In our day-to-day life credit cards assists online as well as card swiping procurement. Following are some advantages of using credit cards:

- Ease of purchase
- Keep good customer history
- Protection of Purchases

Well this also leads to increasing online transactions using credit cards. Credit cards play an important role in today's economy. Also at the same time the frauds involving credit cards are increasing. The problem of fraud is a serious issue that threaten credit card transaction. With the growth of modern technology credit card fraud is also increasing. Data from Crime Complaint Center show that there has been a significant rise in credit card frauds in last decade.

Fraud is wrongful or criminal deception intended to result in a financial or personal gain. Credit card fraud on other hand is the illegal use of credit card or its information without the knowledge of its owner. There are mainly two groups:

- Application Fraud-

  Here criminals use false identities to apply for new cards from bank or issuing company.

- Behavioral Fraud-

  It has four types stolen/lost-card, mail-theft, counterfeit card and card holder not present fraud.

Fraud detection systems are highly complex and hard to build. Also 100% accuracy is hard to achieve. There are certain difficulties of credit card fraud detection:

- Overlapping Data
- Lack of Adaptability
- No Standard evaluation criteria
- Cost of fraud detection
- Imbalanced data

A good fraud detection system must address all these difficulties.

Classification of credit card fraud detection techniques:

- Fraud Analysis(misuse detection)

  This is a supervised classification task(transaction level).

- User Behavior Analysis(anomaly detection)

  This involves the use of unsupervised methodologies.

It's important to understand the differences between these two methods. The fraud analysis is detection of known fraud tricks, while user behavior analysis detects novel frauds.

2 MOTIVATION

Credit cards are popular mode of cashless transactions both online and offline. In recent years the use of credit card has increased. Along with that the credit card fraud is also growing. During this age of global communication improvement, economic fraud is drastically increasing. There is billions of dollars of loss due fraudulent transaction every year.

Due to this credit card is considered as a ”nice target for fraud”. In very short time attackers can get lots of money. These fraudulent transactions are hard to separate from genuine ones. Thus having an efficient method for fraud detection is necessary.

There are many ways to identify fraudulent transactions. The main motive of this paper is to compare some of the models used for fraud detection. This project will compare SVM, KNN, and Naive Bayes models.

3 A SURVEY ON PAPERS

3.1 Credit Card Fraud Identification Using Machine Learning Approaches

Papaer presents the survey of present and past research on credit card fraud detection. An analysis of the different techniques of fraud detection which are secure and consumer friendly is given.

The dataset contains the transactions created by European card holders. Local Outlier factor, Isolation Forest and Support Vector machine are the techniques used for data processing. While comparing the three models mentioned above it was observed that Isolation Forest performed far better than the other two.

3.2 Analysis on Credit Card Fraud Identification Techniques based on KNN and Outlier Detection

Fraud detection is highly complex, and no system can guarantee 100% satisfaction result rate. This paper revolves around the use of KNN algorithm for fraud detection. A brief literature survey of past research is given. In this paper a comparison of various techniques is also given (Logistic Regression, Decision Tree, ANN, SVM, Hidden Markov Model).

It also states properties of good fraud detection model. This paper also defines the structure and working of KNN model. It explains outlier detection technique. Also the two main types of outlier detection are discussed.

4 PROBLEM DEFINITION AND SCOPE

4.1 Problem Definition

To implement and compare various classification algorithms to classify fraudulent and genuine credit card transactions from given dataset.

4.2 Scope

To reduce the loss due to fraudulent transactions it's necessary to have a good fraud detection model. Machine learning can be used to build such models. Here the scope is limited to identifying fraudulent transactions from the dataset by using various methods mentioned above. We will compare their performances.

The future scope may be defined where the above mentioned models are used in collaboration with ensemble or meta-learning techniques to build classifiers.

5 DIFFERENT MACHINE LEARNING ALGORITHMS

5.1 Support Vector Machine (SVM)

It is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular co-ordinate. Then we find out ideal hyperplane that separates the two classes.

It can work with both linear and non-linear scenarios. It has high prediction accuracy and performance rate but is limited to two classified classes only.

5.2 Naive Bayes

Naive Bayes is one of the powerful and easy-to-train machine learning algorithms that is used for classification. It is an extension of the Bayes theorem wherein each feature assumes independence. These are multipurpose classifiers. It is used for a variety of tasks such as spam filtering and other areas of text classification.

It's an easy and quick way to predict class of the data. It can also be used for multi-class prediction. When the assumptions of independence is valid, Naive Bayes is much more capable than other algorithms. Furthermore, you will require less training data.

5.3 KNN

A simple yet important classification algorithm. These algorithms does not make any assumptions about how the data is distributed. It stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). It finds several applications in data mining, pattern recognition and intrusion detection

Its a non-parametric algorithm. Mostly used in real-life scenarios.

6 Implementation

6.1 Data and Preprocessing

The Kaggle dataset for credit card fraud detection contains transactions made by credit cards in September 2013 by European cardholders. This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class(frauds) account for).172% of all transactions. The dataset does not have any missing or null values.

It contains only numerical input variables which are result of PCA(Principal Component Analysis) transformation. Unfortunately, due to confidentiality issues, we cannot get the original features and more background information about the data. Features V1, V2, ... V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'.

Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'Amount' is the transaction Amount, this feature can be used for example-dependant cost-senstive learning. Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise.

6.2 Software

Python version used: 3.7.4 Jupyter Notebook version used: 6.0.1

S.No	Software/Package	Version
1	Jupyter	143.5 MB
2	pandas	0.25.1
3	scikit-learn	0.0.21.3
4	seaborn	0.9.0
5	matplotlib	3.1.1

Table 1: Software Used

6.3 Results

	Algorithm	Average Accuracy
	KNN	0.9995
	SVM	0.9994
	NB	0.9784

Table 2: Results

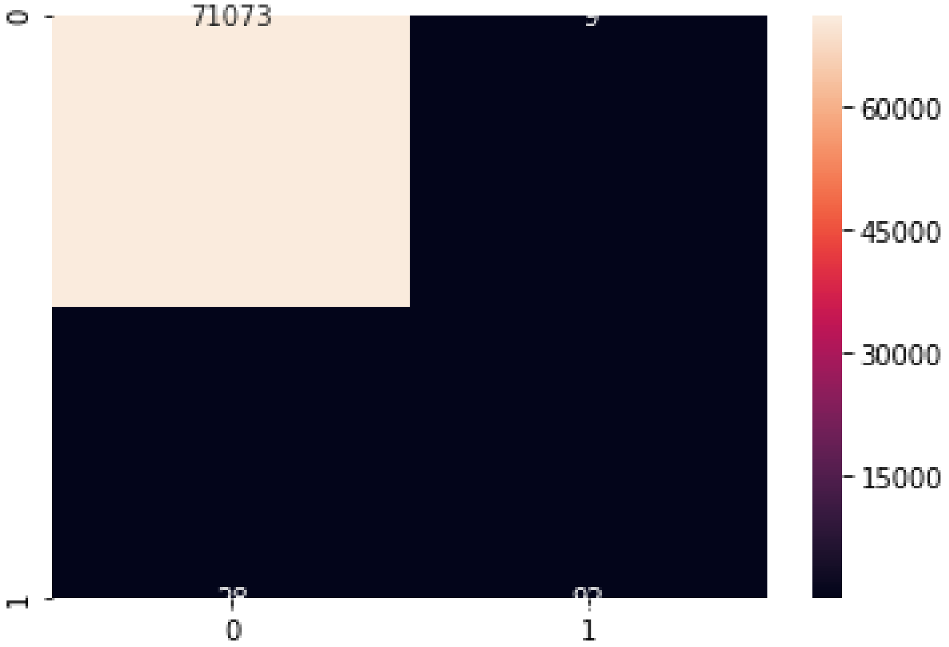

Figure 1: KNN

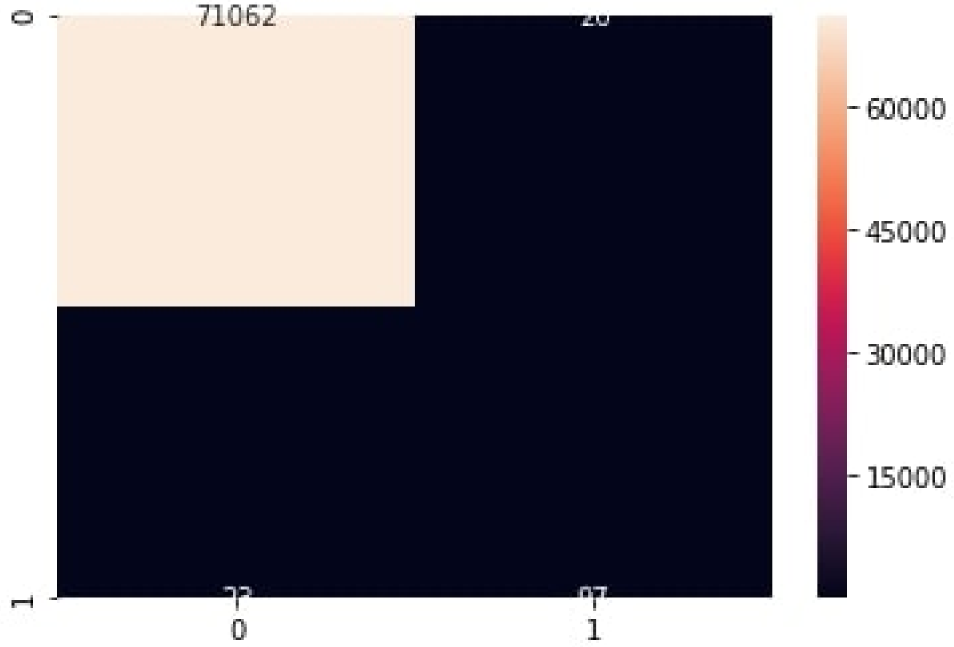

Figure 2: SVM

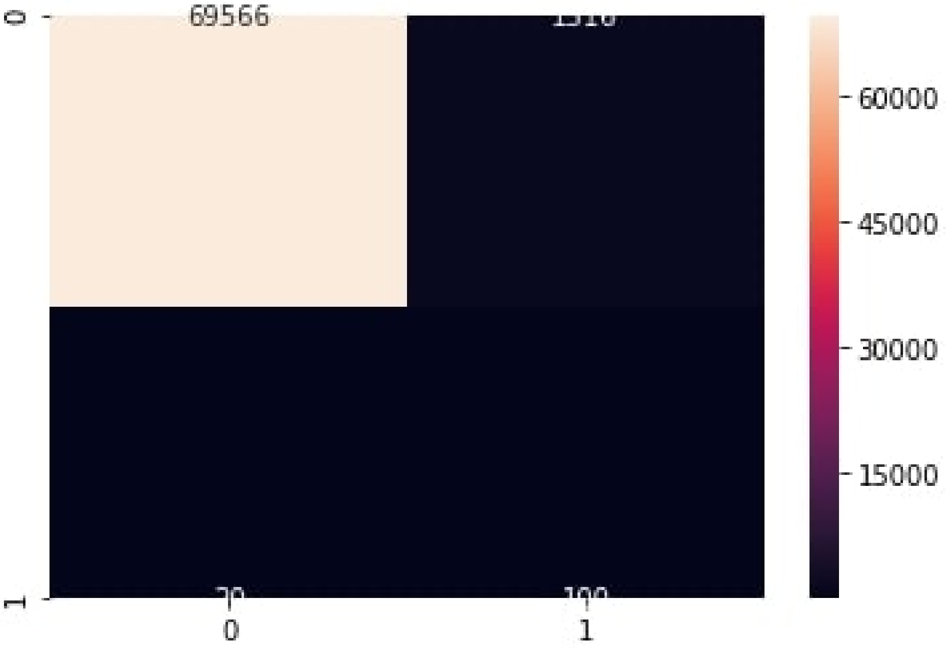

Figure 3: Naive Bayes

7 CONCLUSION

With rise of e-commerce number of credit card users increased. Along with which the number of credit card frauds also increased. This results in loss of billions of dollars every year. That is why it's important for both consumers as well as financial companies to have system to identify or detect such fraudulent transactions. In this study different classification algorithms such as KNN, Naive Bayes, SVM were used to build a basic fraud detection model. It was observed that KNN has the highest average accuracy of 99.95% followed by SVM with an accuracy of 99.94%.

References

[1] P. Kumar and F. Iqbal, ”Credit Card Fraud Identification Using Machine Learning Approaches,” 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), CHENNAI, India, 2019, pp. 1-4.

[2] N. Malini and M. Pushpa, ”Analysis on credit card fraud identification techniques based on KNN and outlier detection,” 2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), Chennai, 2017, pp. 255-258. Available.

[3] SamanehSorournejad, Zahra Zojaji, Reza Ebrahimi Atani, Amir Hassan Monadjemi, ”A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective”, 2016

[4] ”Credit Card Fraud Detection”, https://www.kaggle.com/mlg-ulb/creditcardfraud

[5] ”Machine Learning A-Z”, https://www.superdatascience.com/courses/machine-learning

[6] ”Detecting Credit Card Fraud Using Machine Learning”, https://towardsdatascience.com/detecting-credit-card-fraud-using-machine-learning-a3d83423d3b8

attach your review and visit log here......

attach plagiarism report here.....