USKUDAR UNIVERSITY ★ FACULTY OF ENGINEERING AND NATURAL SCIENCES

Credit Card Fraud and Machine Learning

B.Sc. Thesis by
Zeki Onur Yüksek

Department: Software Engineering

Thesis Supervisor: Dr.Öğr. Üyesi Tuğçe BALLI ALTUĞLU

26/06/2022

USKUDAR UNIVERSITY ★ FACULTY OF ENGINEERING AND NATURAL SCIENCES

Credit Card Fraud and Machine Learning

B.Sc. Thesis by
Zeki Onur Yüksek
(170209037)

Date of submission :	**26 June 2022**
Date of defence examination:	**26 June 2022**

Supervisor (Chairman) :	**Dr.Öğr. Üyesi Tuğçe BALLI ALTUĞLU**
Members of the Examining Committee :	**Doç.Dr. Türker Tekin ERGÜZEL**
	Dr.Öğr. Üyesi Kristin Surpuhi BENLİ
	Dr.Öğr. Üyesi Belaynesh CHEKOL

June 2022

FOREWORD

I would like to express my deep appreciation and thanks to my advisor Dr.Öğr. Üyesi Tuğçe BALLI ALTUĞLU. This work is supported by Uskudar University.

TABLE OF CONTENTS

Table of Contents

ABBREVIATIONS

FPSs   : Fraud Prevention systems
ML   : Machine Learning
POS   : Point of Sale
ATM   : Aoutomatic Teller Machine
DLNN   : Deep Learning And Neural Network

LIST OF FIGURES

LIST OF TABLES

Credit Card Fraud and Machine Learning

SUMMARY

Technology is developing at a pace that is hard to follow day by day. Along with the developing technology, many different new products take their place in people's lives and make their lives easier. Credit cards are at the forefront of these products that emerged with the rapid introduction of technology into our lives. Thanks to its worldwide prevalence and strong infrastructure, credit cards have become popular in a very short time.

It has entered their lives and has become a payment tool that they can use without fear. With the increase in the use of credit cards and, in parallel, the technology, which is difficult to follow, certain security vulnerabilities have been encountered. Due to the increase in the use of credit cards, a new challenge emerges for fraudsters. As a result of this situation, various institutions and organizations and people have suffered a great amount of material and moral damage. There are various algorithms and systems developed from the past to the present to detect credit card frauds. Within the scope of this thesis, firstly, researches on credit card and credit card frauds were conducted. Afterwards, it is aimed to find the most effective and efficient credit card fraud detection method that can be applied against such frauds by comparing the methods used to detect credit card frauds with the algorithm to be developed during the thesis. Analyzes were made using the Credit Card Fraud dataset obtained from the Kaggle database and Python libraries.

Kredi Kartı Dolandırıcılığı ve Makine Öğrenimi

ÖZET

Teknoloji her geçen gün takip edilmesi zor bir hızla gelişiyor. Gelişen teknoloji ile birlikte birçok farklı yeni ürün insanların hayatında yer almakta ve hayatlarını kolaylaştırmaktadır. Teknolojinin hayatımıza hızla girmesiyle ortaya çıkan bu ürünlerin başında kredi kartları gelmektedir. Dünya çapında yaygınlığı ve güçlü altyapısı sayesinde kredi kartları çok kısa sürede popüler hale geldi.

Hayatlarına girmiş ve korkmadan kullanabilecekleri bir ödeme aracı haline gelmiştir. Kredi kartı kullanımının artması ve buna paralel olarak takip edilmesi zor olan teknoloji ile birlikte bir takım güvenlik açıkları ile karşılaşılmıştır. Kredi kartı kullanımının artmasıyla birlikte dolandırıcılar için yeni bir zorluk ortaya çıkıyor. Bu durum sonucunda çeşitli kurum ve kuruluşlar ile kişiler büyük maddi ve manevi zarara uğramıştır. Kredi kartı dolandırıcılıklarını tespit etmek için geçmişten günümüze geliştirilmiş çeşitli algoritmalar ve sistemler bulunmaktadır. Bu tez kapsamında ilk olarak kredi kartı ve kredi kartı dolandırıcılıkları ile ilgili araştırmalar yapılmıştır. Daha sonra kredi kartı dolandırıcılıklarını tespit etmek için kullanılan yöntemler ile tez aşamasında geliştirilecek algoritma karşılaştırılarak bu tür dolandırıcılıklara karşı uygulanabilecek en etkili ve verimli kredi kartı dolandırıcılık tespit yönteminin bulunması amaçlanmaktadır. Kaggle veri tabanından ve Python kütüphanelerinden elde edilen Credit Card Fraud veri seti kullanılarak analizler yapılmıştır.

INTRODUCTION
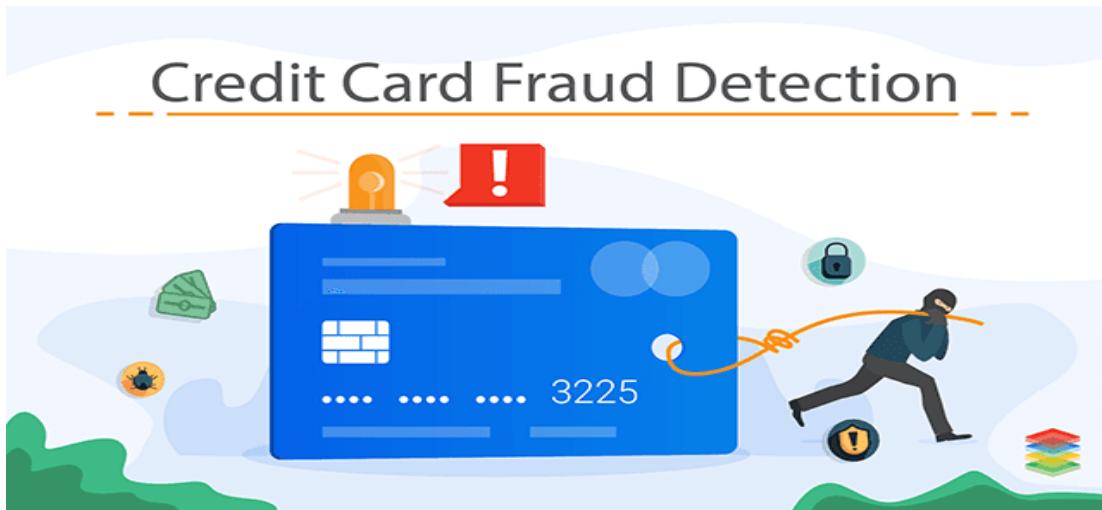
Figure 1 Credit Card Fraud Detection

Payment can be made by credit cards, POST devices used at shopping points, which banks give to people they serve. You can also withdraw cash from ATMs. Credit cards also make people's lives easier in terms of paying their expenses in installments. In this way, people reduce the amount of monthly spending by dividing them by a certain number of months instead of paying in one go.

Credit card fraud is a very broad term for theft and fraud committed using this card or using this card at the time of payment. The purpose may be to purchase goods without paying or to transfer funds from an account without authorization. Credit card fraud is also an addition to identity theft because credit cards are personal cards with personal information on them. Unauthorized 'fraud' in credit card transactions and unwanted use of an account by someone other than the owner of that account. Necessary preventive measures can be taken. To stop this abuse and the behavior of such scammers, practices can be worked on to minimize and protect against it.

All countries and companies around the world incur huge financial losses in credit card fraud. For this reason, certain analyzes are made by using the data obtained from credit card transactions in the study, and as a result of this analysis, it is aimed to prevent credit card fraud.

Scope of the Thesis

An algorithm will be developed to detect credit card fraud and the algorithm will be tested on the data set to be used in the application.The methods used today will be applied to the data set to be used in the application.

Background

Many people experience credit card fraud. According to the [1]Security.org Team's survey, 58% of people in America have been scammed at least once. In 81% of the scenarios where the fraudsters are successful, the system does not give a warning. It emphasizes the importance of the system in credit card fraud.
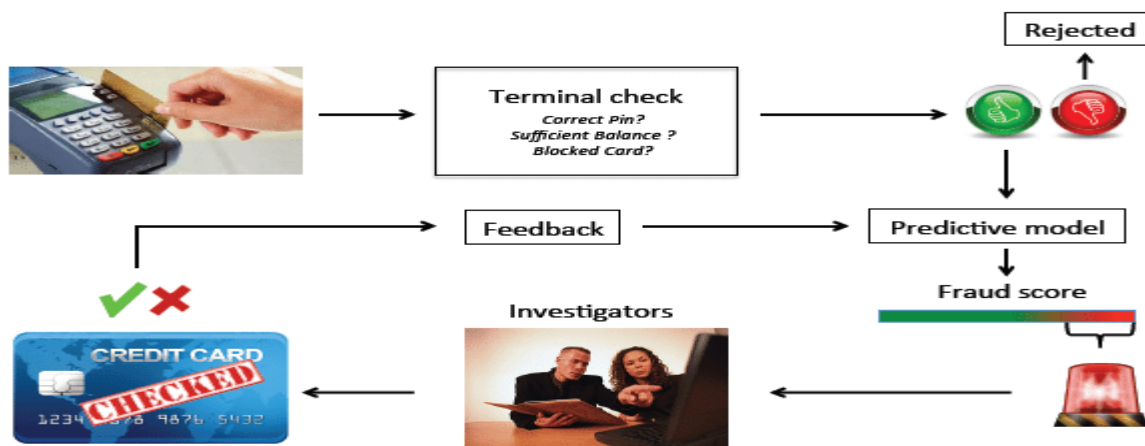

Figure 2 Credit Card Fraud Detection Way

LITERATURE REVIEW

When the studies in the literature are examined, it is seen that the credit card fraud detection algorithms are mostly

It is understood that it is based on data mining. For example, in credit card fraud, the questionable. Transactions can be identified by dividing them into two classes, legitimate (legal) and fraudulent transactions.

Credit card transaction dataset was studied in Credit Card Fraud Transaction Detection thesis prepared by K. Soylu in 2018. The model was trained with the training set for 70% of the data set, and the success of the model was tested with the corresponding test set for 30%. In the light of these data, the fraud detection rate of the deep learning model is 78.7%. By working on

the same dataset with the Random forest model, the accuracy error rate was determined as 0.00008.[3]

Results, artificial neural networks and logistic regression approaches are more efficient than the decision tree in solving the related problem.

proved to be performing well. Decision trees are used to detect credit card fraud another study in which the decision tree approach is investigated in the solution of the researched problem is DVM emphasizes that it performs better than the machine learning has given all methods very high accuracy. Random Forest machine learning technique with multiple decision trees was the technique that gave the best results. It is one of the techniques tested in our project.

METHODS OF LITERATURE SURVEY

The data set and methods used to help detect Credit Card Fraud are explained in this section.
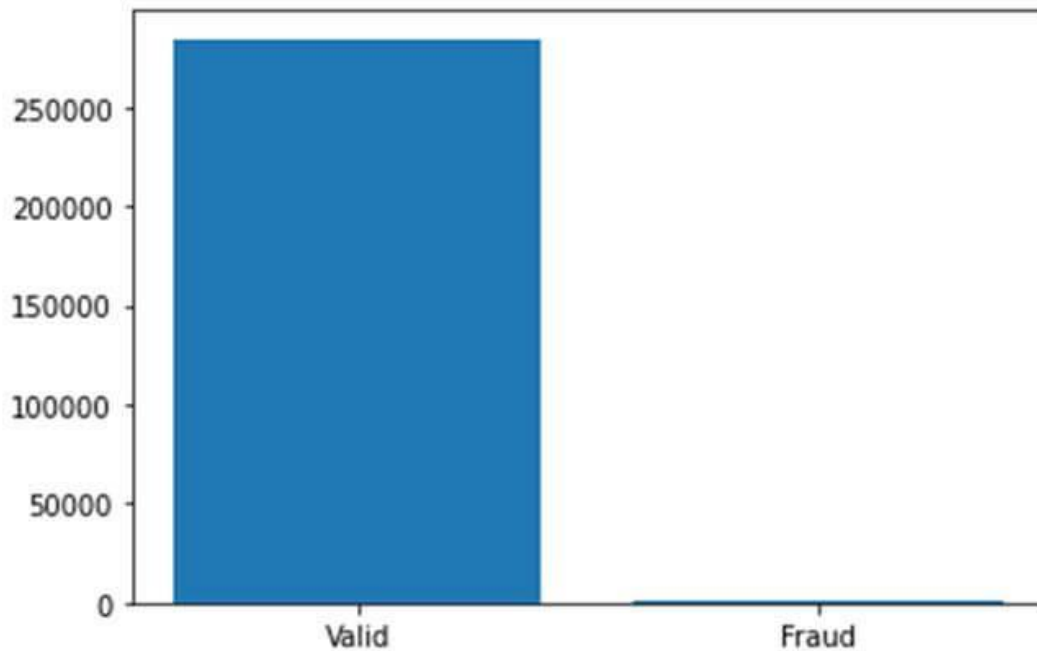
Data Set

The data set used in the study was obtained from the Kaggle database. The features and explanations of the dataset are given below.[2]The datasets include transactions made by European cardholders with credit cards in September 2013.

The relevant dataset presents fraudulent transactions, of which 492 out of 284,807 transactions occurred within two days. The positive class (fraud) accounts for 0.172% of all transactions. It contains only the numerical input variables that are the result of the PCA transformation. Due to privacy issues, no further information about the original features and data could be gathered.

• V1, V2,…..,V28 are the main components obtained with PCA.

• Properties that are not converted with PCA are 'Time' and 'Quantity'.

• The 'Time' property includes the seconds that elapse between each operation and the first operation in the dataset.

• The 'Amount' property is the transaction amount, this property is used for cost sensitive learning.

• The property 'Class' is the response variable and takes the value 1 in case of fraud and 0 otherwise.

There is a fraud rate of 0.172% in the transactions in the dataset. This imbalance between classes makes it very difficult to classify data. In order to solve the unbalanced data problem, under-sampling is generally performed. That is, all class instances in the data set are reduced to the smallest number of class instances.
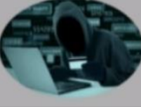

Table 1 Data Set

Python

Data scientists, who are responsible for processing and organizing data with scientific methods, algorithms, and other related techniques, daily review large amounts of data sets and deduce what is important. Python excels at scalability. MATLAB is much faster than languages like R and Stata. Scalability is the main reason why Youtube chose to migrate their processes to Python. Also, cloud software Titan has written more than 4 million Python codes for the scalability of its applications. Python's data science libraries are very popular among data scientists. Many data science libraries such as Numpy, Scipy, StatsModels and sci-kit-learn continue to be added, while existing ones continue to be developed.

Therefore, for data scientists, Python becomes a solid programming language that fulfills most needs and helps solve seemingly unsolvable problems.

Figure 3 Confusion Matrix

Deep Learning And Neural Network

Deep Learning is a machine learning method. It allows us to train artificial intelligence to predict outputs with a given dataset. Both supervised and unsupervised learning can be used to train artificial intelligence. Artificial neural networks are information processing technology developed by examining the learning ability of the human brain. It is an artificial intelligence method created by simulating biological neurons. Computer learning and It is a system that can produce comments and results on the subject of incomplete information or knowledge. Artificial neural networks emerge as a result of the systematic combination of a specified number of artificial nerve cells.
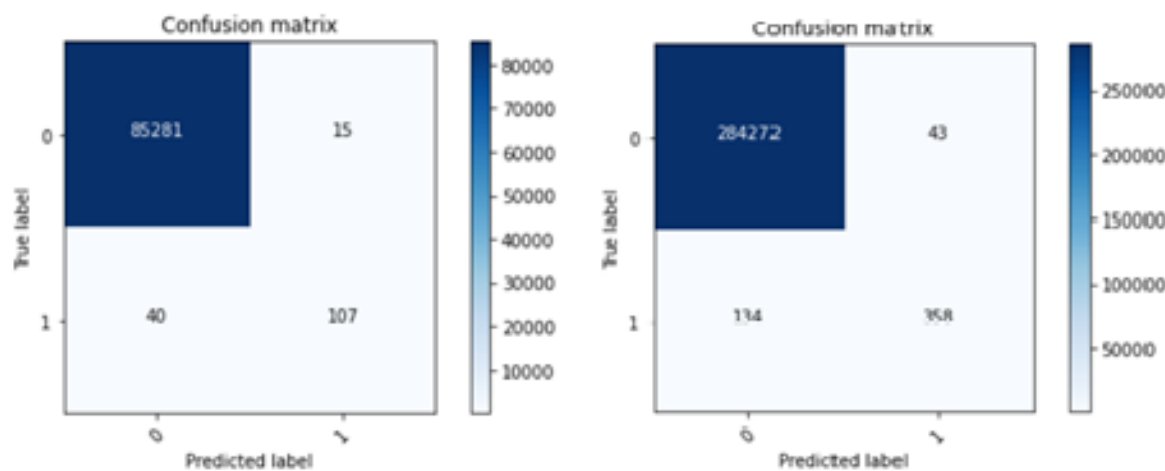

Table 2 DLNN

Decision Trees

Tree-based learning algorithms are among the most used supervised learning algorithms. In general, they can be adapted to the solution of all the problems considered.

Methods such as decision trees, random forest, gradient boosting are widely used in all kinds of data science problems. Therefore, it is very important for data analysts to learn and use these algorithms. The decision tree algorithm is one of the data mining classification algorithms. It is easy to understand and interpret. The tree structures used can be visualized
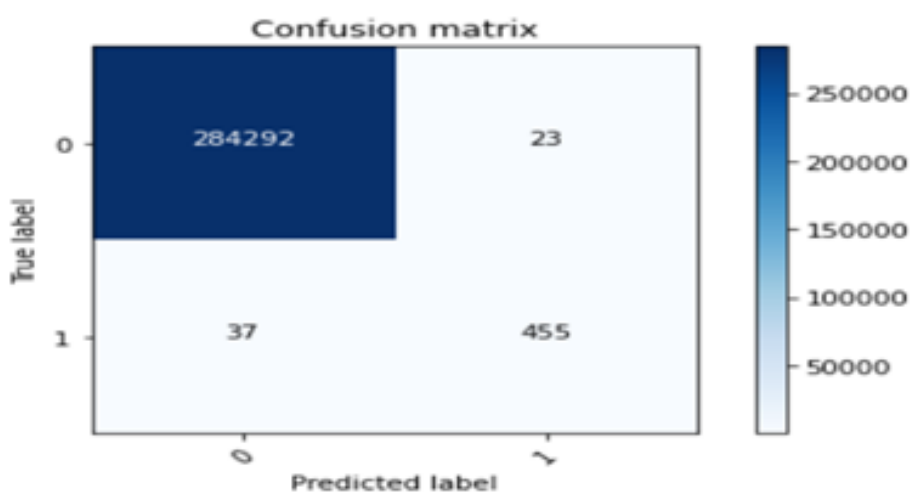
Table 3 Decision Trees

Random Forest

Random Forest algorithm; It enables you to generate various models and create classifications by training each decision tree on a different observation sample over multiple decision trees. Ease of use and flexibility; It has accelerated its adoption and widespread use as it addresses both classification and regression problems. The most liked point about the algorithm is; It provides the opportunity to rediscover your dataset and to explore it more deeply by creating various models on your dataset.
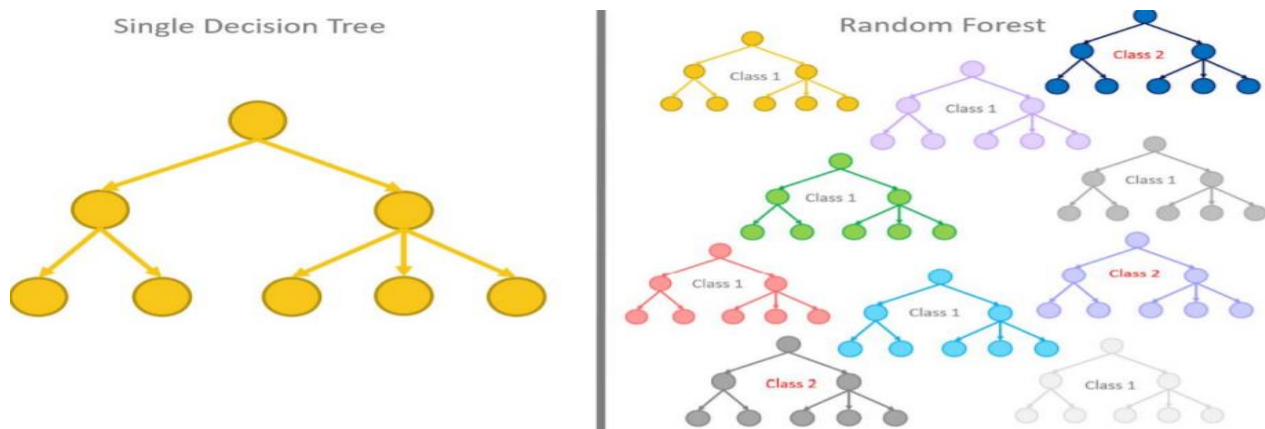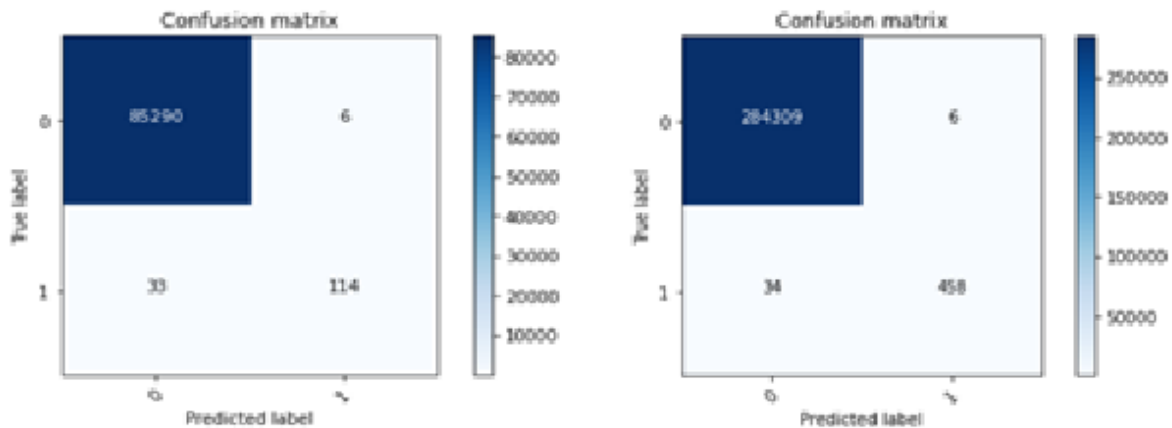

Figure 4 Random Forest

It was the method I got the best results in my random forest project. I was able to detect the most margin of error and the most forgery with this method

Table 4 Random Forest

CONCLUSION AND FUTURE DIRECTIONS

In this study, machine learning techniques such as Deep Learning and Neural Network, Decision Tree and Random forest were used to detect fraud in the credit card system. The accuracy for Neural Network, Decision tree and random forest classifier is 99.0, 99.93 and 99.95, respectively. Comparing all three methods, he found that the random forest classifier was better than the neural Network and the decision tree.The size of the data set used in this study accuracy rates obtained in line with Availability and reliability are very high.To improve this work, in Data Miningcan be tested with other algorithms used. Thus,the resulting rates are comparable and from the Artificial Neural Network.the achievability of a better result can be checked.

Table 5 Precison, Recall, F1

	Precison (%)	Recall (%)	F1 (%)
DNN	85.3	78.9	82.0
Decision Tree	82.7	74.8	78.6
Random Forest	95	77.5	85.3

(The F1 score combines Recall and Precision into one metric as a weighted average of the two.)

REFERENCES

Goy, G., Gezer, C., & Gungor, V. C. Makine Öğrenmesi Yöntemleri ile Kredi Kartı Sahteciliği Tespiti.

Goy, G., Gezer, C., & Gungor, V. C. (2019, September). Credit card fraud detection with machine learning methods. In *2019 4th International Conference on Computer Science and Engineering (UBMK)* (pp. 350-354). IEEE.

KESKENLER, M. F., Deniz, D. A. L., & AYDİN, T. (2021). Yapay zeka destekli ÇOKS yöntemi ile kredi kartı sahtekarlığının tespiti. *El-Cezeri*, *8*(2), 1007-1023.

Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017, October). Credit card fraud detection using machine learning techniques: A comparative analysis. In *2017 international conference on computing networking and informatics (ICCNI)* (pp. 1-9). IEEE.

Varmedja, D., Karanovic, M., Sladojevic, S., Arsenovic, M., & Anderla, A. (2019, March). Credit card fraud detection-machine learning methods. In *2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH)* (pp. 1-5). IEEE.

Yee, O. S., Sagadevan, S., & Malim, N. H. A. H. (2018). Credit card fraud detection using machine learning as data mining technique. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, *10*(1-4), 23-27.

Sailusha, R., Gnaneswar, V., Ramesh, R., & Rao, G. R. (2020, May). Credit card fraud detection using machine learning. In *2020 4th international conference on intelligent computing and control systems (ICICCS)* (pp. 1264-1270). IEEE.

Yee, O. S., Sagadevan, S., & Malim, N. H. A. H. (2018). Credit card fraud detection using machine learning as data mining technique. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, *10*(1-4), 23-27.

Beigi, S., Amin Naseri, M. (2020). Credit Card Fraud Detection using Data mining and Statistical Methods. *Journal of AI and Data Mining*, 8(2), 149-160. doi: 10.22044/jadm.2019.7506.1894

[1] https://www.security.org/digital-safety/credit-card-fraud-report/

[2] https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

[3] SOYLU K."Kredi Kartı Sahte İşlem Tespiti"