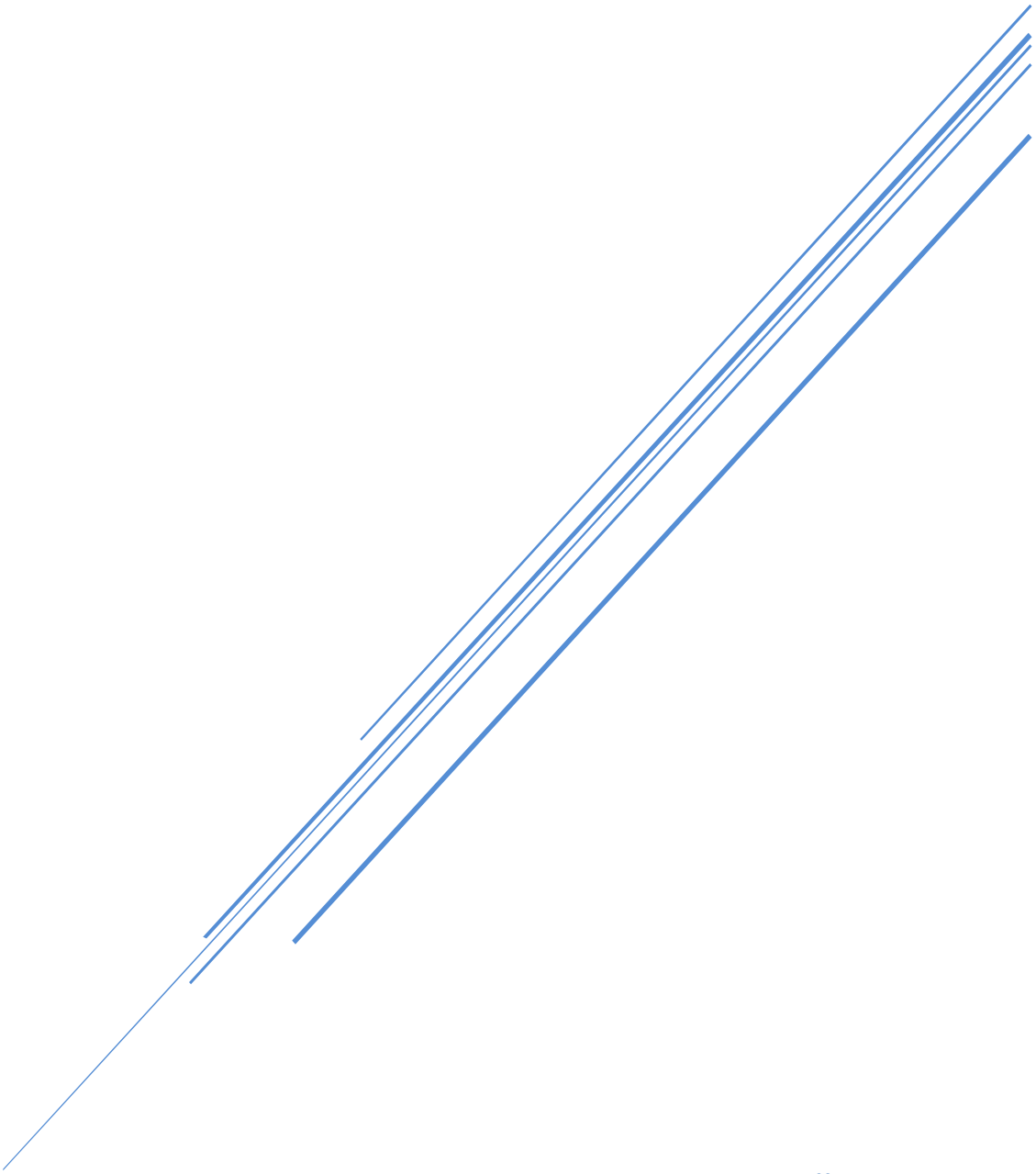
CREDIT CARD FRADULENT



Mustafa Tayyip BAYRAM

Furkan ÖCALAN

Table of Contents

List of Figures 1

ABSTRACT 2

1. Introduction 2

Purpose and Importance of the Project 4

Technology Areas Related to the Project 4

2. Requirements 6

Functional Requirements 7

Non-Functional Requirements 7

3. Detailed System Architecture and Design 8

Use-Case Model 8

Use Case Definitions 8

System Architecture Diagram 12

Input/Output Diagram 12

Entity Relationship Diagram (ERD) 13

Flow Chart 13

Project Risk Matrix 13

UML Class Diagram 14

4. Development 14

5. Platform Requirements 15

6. User Interface Design 15

Conclusion 19

References 19

List of Figures

Figure 1: Non-Cash Transactions Globally and Regionally – Historical Volumes, 2014-2019 2

Figure 2: Non-Cash Transactions Globally and Regionally – Forecast Growth Rates, 2019-2023F 3

Figure 3: Card Fraud Worlwide Volume 3

Figure 4: AI Schema 5

Figure 5: Use Case Diagram 8

Figure 6: System Architecture Diagram 12

Figure 7: Input/Output Classification Diagram 12

FIGURE 8: INPUT/OUTPUT VISUALIZATION DIAGRAM 12
FIGURE 9: ER DIAGRAM 13
FIGURE 10: FLOW CHART 13
FIGURE 11: CLASS DIAGRAM 14
FIGURE 12- SIGN-UP PAGE 16
FIGURE 13- SIGN-IN PAGE 16
FIGURE 14 - UPLOAD TRANSACTION FILE 17
FIGURE 15 - FILL THE TEXT AREA AND MAKE PREDICT 17
FIGURE 16 - ANALYZE PROCESS 18
FIGURE 17 - EXAMPLE OF MODEL RESULT 18
FIGURE 18 - EXAMPLE OF MODEL RESULT 19

ABSTRACT

We will explore credit card fraud detection using machine learning and deep learning in this project. Credit card fraud detection is critical for financial organizations in an era of digitization. If necessary, we want to use machine learning models and deep learning approaches. These methods and models will produce significant results when applied to financial data. Finally, new directions for improving both techniques and results are suggested.

1. Introduction

We are on our way to become a cashless society in today's world. According to the World Payments Report, overall non-cash transactions climbed by 10.1 percent in 2016 compared to 2015, reaching 482.6 billion! That's enormous! In addition, as indicated below, non-cash transactions are likely to rise steadily in the next years:

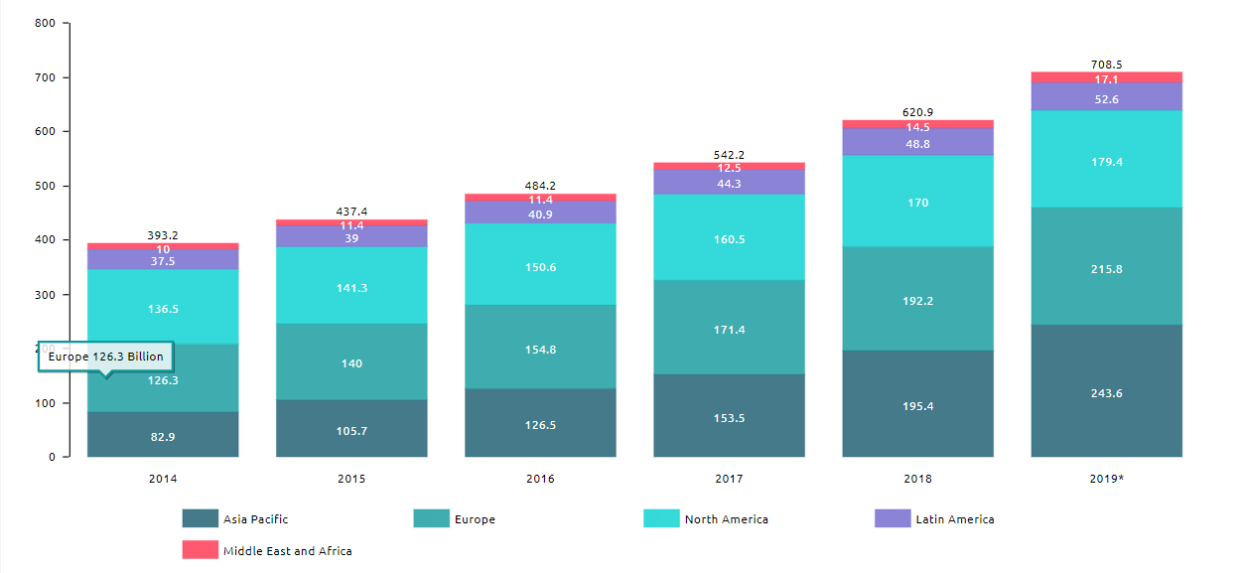


Figure 1: Non-Cash Transactions Globally and Regionally – Historical Volumes, 2014-2019

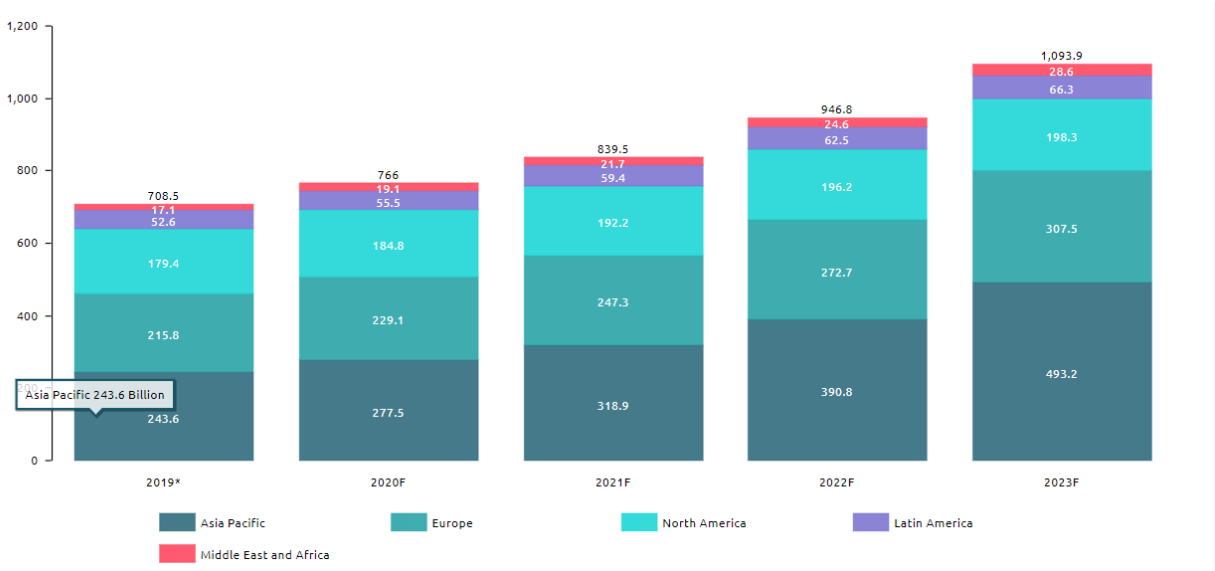


Figure 2: Non-Cash Transactions Globally and Regionally – Forecast Growth Rates, 2019-2023F

While this is good news, it also means that fraudulent transactions are becoming more common. Despite the implementation of EMV smart chips, credit card fraud continues to cost us a significant amount of money:

Figure 3: Card Fraud Worlwide Volume

This is becoming a severe problem since, in most cases, the victim of the fraud has no understanding what has transpired until the very end. So, in this project, we attempted to construct a Web App that would use Machine Learning to detect such types of frauds.

In other words, Fraud can be committed in a variety of ways and in a wide range of industries. To make a decision, the majority of detection systems combine a number of fraud detection datasets to create a connected picture of both legitimate and invalid payment data. IP address, geolocation, device identity, "BIN" data, global latitude/longitude, history transaction trends, and actual transaction information must all be considered while making this decision. In practice, this means merchants and issuers use analytically driven answers to detect fraud by using a set of business rules or analytical algorithms to internal and external data.

Purpose and Importance of the Project

The goal of this research is to use the transaction amount and PCA transforms to predict if a credit card transaction is fraudulent or not. Its goal is to hunt down credit card transaction data by looking for irregularities in the information. Credit card fraud detection is usually accomplished through the use of an algorithm that detects irregularities in transaction data and alerts the cardholder (as a precautionary measure) and the bank about any suspicious transactions.

Technology Areas Related to the Project

The project should be classified as a sub-branch of some fields. It is required to become familiar with the following areas in order to comprehend the general application.

- Artificial Intelligence (AI),
- Machine Learning (ML),
    - Logistic Regression
    - Principal Component Analysis (PCA)
    - K Nearest Neighbors (KNN)
- Deep Learning (DL),
    - One Dimensional CNN (1DCNN)
    - LSTM
- Web Development,
- Database Management Systems (DBMS)

a. Artificial Intelligence (AI):

It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable [1].
There are numerous, real-world applications of AI systems today. Some of the most common examples: speech recognition, customer service, computer vision, recommendation engines and automated stock trading.

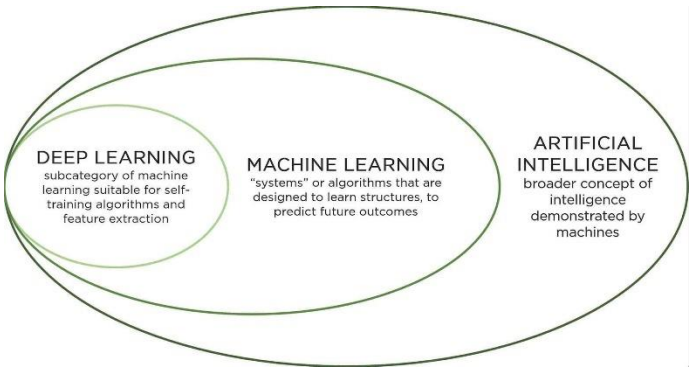


Figure 4: AI Schema

b. Machine Learning (ML):

Machine learning is an area of artificial intelligence (AI) and computer science that focuses on using data and algorithms to mimic the way people learn, with the goal of steadily improving accuracy [2].

Logistic Regression:

It is often used for classification and predictive analytics. Logistic regression estimates the probability of an event occurring, such as voted or didn't vote, based on a given dataset of independent variables. Since the outcome is a probability, the dependent variable is bounded between 0 and 1. In logistic regression, a logit transformation is applied on the odds—that is, the probability of success divided by the probability of failure. This is also commonly known as the log odds, or the natural logarithm of odds, and this logistic function is represented by the following formulas [3]:

$$\text{Logit}(pi) = 1/(1+ \exp(-pi))$$

$$\ln(pi/(1-pi)) = Beta\_0 + Beta\_1*X\_1 + \ldots + B\_k*K\_k$$

Principal Componen Analysis (PCA):

PCA is a method for lowering the dimensionality of such datasets, boosting interpretability while minimizing information loss. It accomplishes this by generating new uncorrelated variables that optimize variance in a sequential manner [4].

K Nearest Neighbors (KNN) (If needed):

The k-nearest neighbors method, often known as KNN or k-NN, is a non-parametric, supervised learning classifier that makes classifications or predictions about the grouping of individual data points based on closeness. It may be used for both regression and classification applications, however it is most commonly employed as a classification technique, based on the idea that comparable points can be discovered close together [5].

c. Deep Learning (DL) (If needed):

Deep learning is a subset of machine learning, which is essentially a neural network with three or more layers. These neural networks attempt to simulate the behavior of the human brain albeit far from matching its ability allowing it to "learn" from large amounts of data. While a neural network with a single layer can still make approximate predictions, additional hidden layers can help to optimize and refine for accuracy.
Deep learning algorithms are incredibly complex, and there are different types of neural networks to address specific problems or datasets [6]. For example,

One Dimensional CNN(1DCNN)

A convolutional neural network, or CNN, is a deep learning neural network sketched for processing structured arrays of data such as portrayals. similar to ANN,CNN has the same hidden layer structure in addition to special convolution layers with a different number of channels in each layer. The word convolution is linked with the idea of moving filters that capture the key information from the data. CNN is widely used in image processing as it automatically performs the feature reduction which makes it less prone to overfitting and thus training CNN does not require heavy data pre-processing. [7]

Long short-term memory (LSTMs)

Long Short-Term Memory is a kind of recurrent neural network. In RNN output from the last step is fed as input in the current step. LSTM was designed by Hochreiter & Schmidhuber. It tackled the problem of long-term dependencies of RNN in which the RNN cannot predict the word stored in the long-term memory but can give more accurate predictions from the recent information. As the gap length increases RNN does not give an efficient performance. LSTM can by default retain the information for a long period of time. It is used for processing, predicting, and classifying on the basis of time-series data. [8]

Finally, deep learning has many kinds of applications in healthcare, customer service, financial services, law enforcement, and many other fields.

d. Web Development:

Web development, also known as website development, refers to the tasks associated with creating, building, and maintaining websites and web applications that run online on a browser [9].

e. Database Management System:

A database management system (DBMS) is a software package designed to define, manipulate, retrieve, and manage data in a database. A DBMS generally manipulates the data itself, the data format, field names, record structure and file structure. It also defines rules to validate and manipulate this data.
Database management systems are set up on specific data handling concepts, as the practice of administrating a database evolves. The earliest databases only handled individual single pieces of specially formatted data. Today's more evolved systems can handle different kinds of less formatted data and tie them together in more elaborate ways. [10]

2. Requirements

Functional Requirements

These are the requirements that the end user expresses as essential features that the system should provide. As part of the contract, all these functionalities must be included into the system.

Non-Functional Requirements

According to the project contract, these are the quality restrictions that the system must meet. The importance of these aspects, as well as the amount to which they are implemented, vary by project.

Functional Requirements
1. The user must be authorized when they log in.
1.1 User must be able to sign up if user is not registered.
2. User must be fill the transactions to the text area or upload as a file in required formats for new analyze.
2.1 The system gives an error message if the file does not match the format.
3. The system must display and save analysis results.
3.1 The system shall be able to tag outcome as positive, negative in the matter of fradulent.
3.2 The system must be show the results.
4. User must be able to examine previous analyze results.
5. User can edit user's information.
Non-Functional Requirements
1. Fradulent detection accuracy must be greater than 85%.
2. Fradulent detection recall must be greater than 80%.
3. Web systems loading time must be less than 5 seconds.
4. Web system must have responsive design.

3. Detailed System Architecture and Design

Use-Case Model

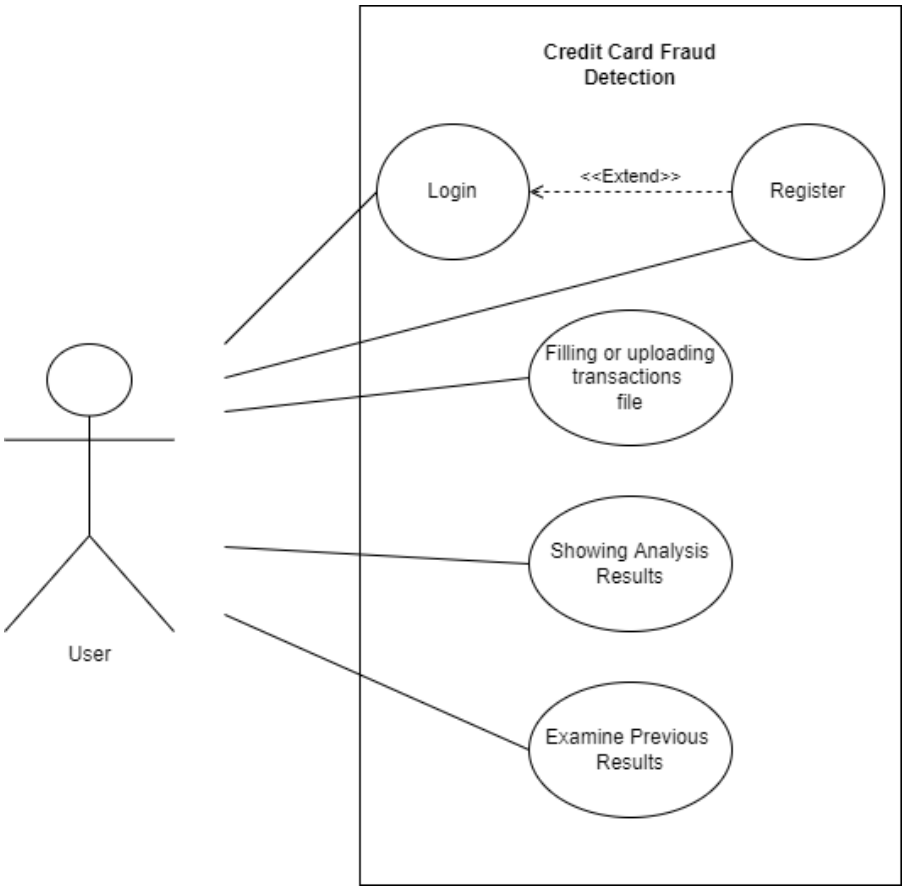


Figure 5: Use Case Diagram

Use Case Definitions

Use Case Name	Login
Description	Login case describes how the user logs into a system. If the user exists, then directs to main page and have an access. Otherwise, user doesn't exist, or wrong password would be some problems.
Actors	User
Pre-Condition	The device must be connected to the internet. The user must exist in the database.
Post-Condition	The user will be logged in

Main Scenario	Serial No	Steps
	1	Enter username and password
	2	Login information will be validated
	3	User will be redirected to the main menu
Extension	2a	Invalid username Error message will be shown
	2b	Invalid password Error message will be shown

Use Case Name	Register	
Description	If no one else has registered with the same email address, register use case will add the user's information to the database.	
Actors	User	
Pre-Condition	The device must be connected to the internet. The user must not exist in the database.	
Post-Condition	The user will be registered in	
Main Scenario	**Serial No**	**Steps**
	1	Enter username, name, email, password.
	2	User information will be added to the database.
	3	User will be redirected to the login page.
Extension	2a	If email or username already exists. Error message will be shown.

Use Case Name	Filling or uploading transactions file
Description	The user fill the text area or uploads the transactions file to be analyzed. Initial step to start the analysis.

Actors	User	
Pre-Condition	The device must be connected to the internet. The user must be logged in. File type will be suitable for standards.	
Post-Condition	The file will be sent for analysis.	
Main Scenario	**Serial No**	**Steps**
	1	The user clicks upload button and select the transcation file in requested formats or fill the text area .
	2	After file is approved by the user, analysis process begins.
Extension	2a	Uploading invalid file. Error message will be shown
	2b	If the file is not approved by the user, process will not begin for the file.

Use Case Name	Showing Analysis Results	
Description	Defines showing analysis results by basic visualization techniques that tells the user there is a fraud or not. User can examine easily what system performed from the file.	
Actors	User	
Pre-Condition	The device must be connected to the internet. The user must be logged in. The user must be uploaded reviews file.	
Post-Condition	The user will be examined results	
Main Scenario	**Serial No**	**Steps**
	1	User opens the main menu
	2	User uploads the transaction file or fill the text area
	3	Result visualizations-data are shown to the user and saved.

Extension	2a	Uploading invalid file Error message will be shown
	2b	Uploading file in different format Error message will be shown

Use Case Name	Examine Previous Results from Dashboard	
Description	Defines checking previous results from user's records if user has already done some analysis. That presents a comparison between previous results and new one.	
Actors	User	
Pre-Condition	The device must be connected to the internet. The user must be logged in. The user must be done analysis before.	
Post-Condition	The user will be checked results	
Main Scenario	**Serial No**	**Steps**
	1	User opens the dashboard
	2	User checks the results and make comparison
Extension	2a	Hasn't done analysis yet Show message about not existing analysis

System Architecture Diagram

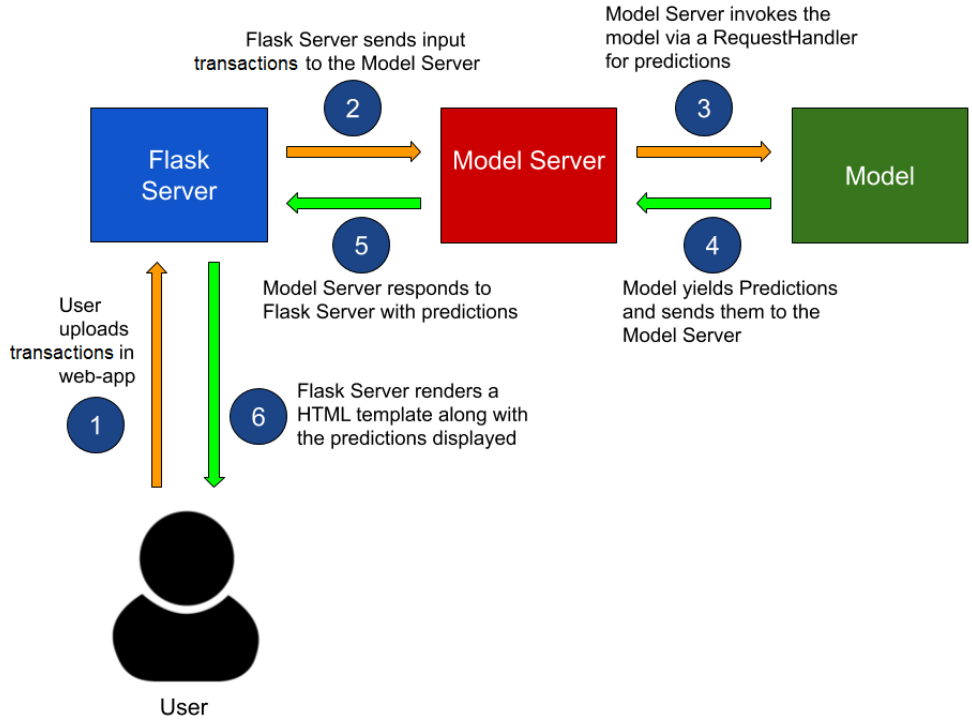


Figure 6: System Architecture Diagram

Input/Output Diagram

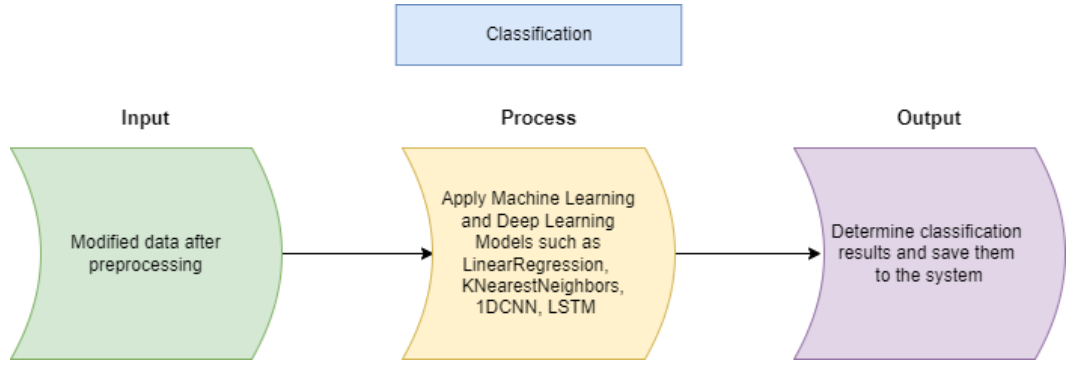


Figure 7: Input/Output Classification Diagram

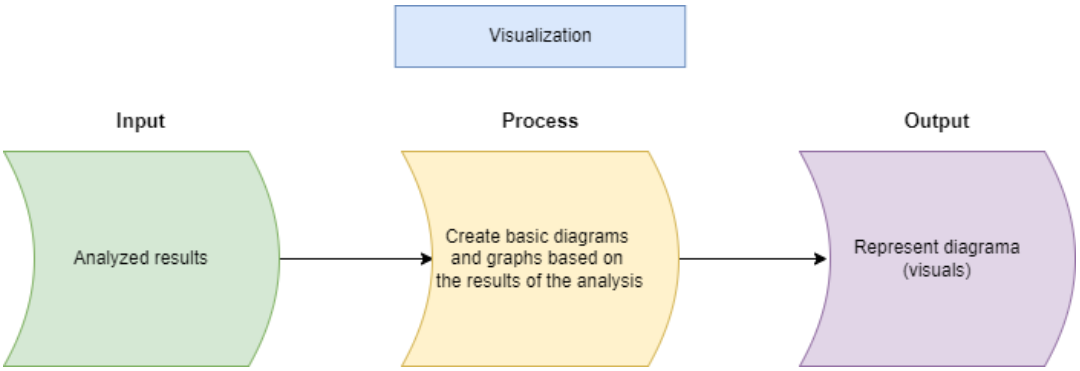


Figure 8: Input/Output Visualization Diagram

Entity Relationship Diagram (ERD)

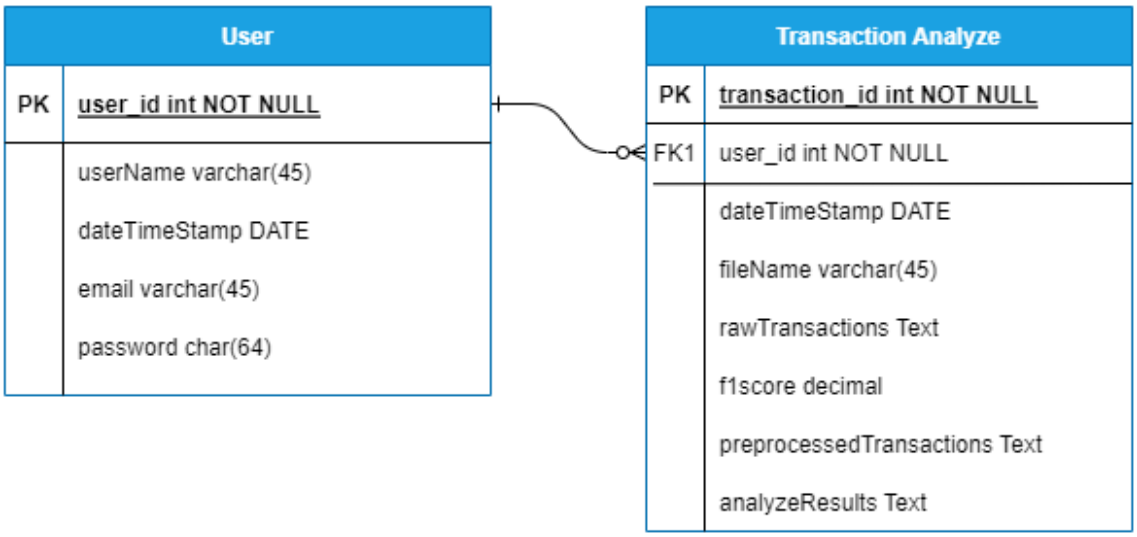


Figure 9: ER Diagram

Flow Chart

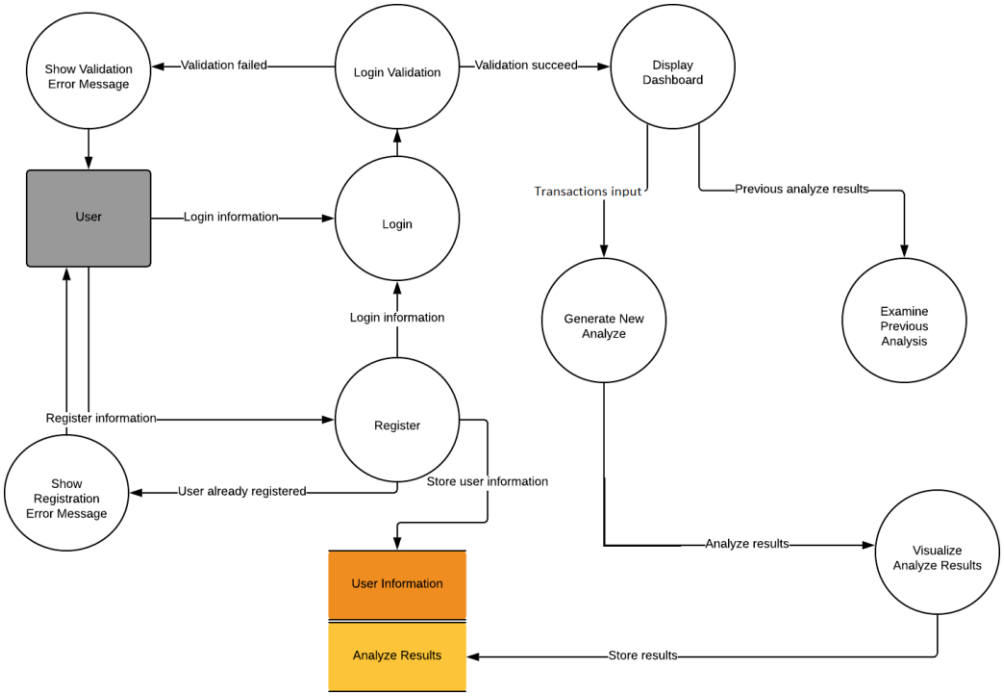


Figure 10: Flow Chart

Project Risk Matrix

Risk	Importance Level	Probability
Changing Requirements	High	Low

Lack of Coding	High	Low
Delayed Deadline of Project	High	Low
Project is not Completed and not Delivered	Very High	Very Low
Software Response Time	Low	Medium
Unable to access desired F1-Score	Medium	Medium

UML Class Diagram

The UML Class diagram is a graphical notation used to construct and visualize object-oriented systems. A class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's:

- classes,
- their attributes,
- operations (or methods),
- and the relationships among objects. [11]

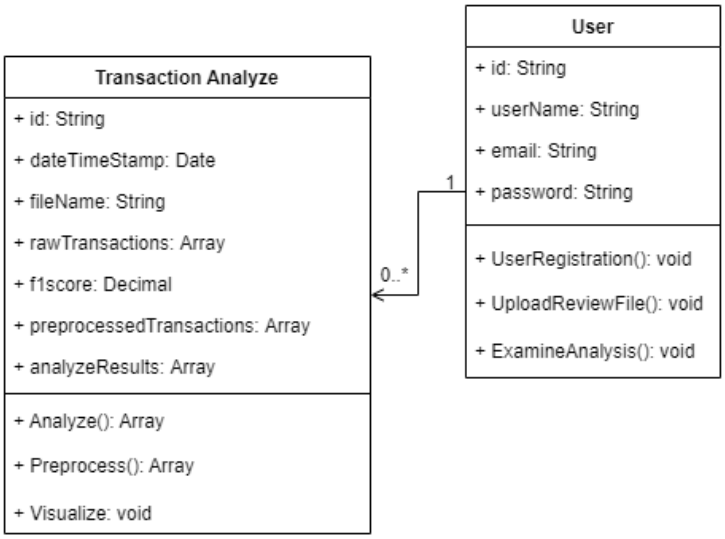


Figure 11: Class Diagram

4. Development

1. Web Application

Web application will be developed using Flask framework which is a framework of Python. The main purpose is to develop a web application that can provide ease of use and accessibility, providing a result of analyzes using a variety of visualization techniques. It is going to be more suitable with our model because we are going to develop it with Python. Besides, with SQLAlchemy ORM, we can

handle our database works more easily. We are going to use HTML, CSS and Bootstrap inside of the project.

2. Database Management System: MySQL

MySQL has been chosen because it is the world's most popular open-source database is that it provides comprehensive support for every application development need. Besides, it has simplicity.

Why MySQL?

- MySQL is a Relational Database Management System or RDBMS which means that it stores and presents data in tabular form, organized in rows and columns.
- MySQL is more secure as it consists of a solid data security layer to protect sensitive data from intruders and passwords in MySQL are encrypted.
- MySQL is available for free to download and use from the official site of MySQL.
- MySQL is compatible with most of the operating systems, including Windows, Linux, NetWare, Novell, Solaris and other variations of UNIX.
- MySQL provides the facility to run the clients and the server on the same computer or on different computers, via internet or local network.
- MySQL has a unique storage engine architecture which makes it faster, cheaper and more reliable.
- MySQL gives developers higher productivity by using views, Triggers and Stored procedures

Object Relational Model (ORM)

The most significant reason to use an ORM is to be able to have a sophisticated, object-oriented business model while still being able to store it and run efficient queries against a relational database efficiently.

- SQLAlchemy: It provides a full suite of well-known enterprise-level persistence patterns, designed for efficient and high-performing database access, adapted into a simple and Pythonic domain language. [11]

3. Security Preventions

User Passwords

User passwords wWill be stored as hashed. Flask libraries will be use for adding this security layer.

5. Platform Requirements

A browser for the client(s) and user which support Javascript.

A internet for the client(s) and user.

Internet Explorer	Microsoft Edge	Mozilla Firefox	Chrome	Opera	Safari
8 +	Latest	Latest	22 +	17 +	12 +

6. User Interface Design

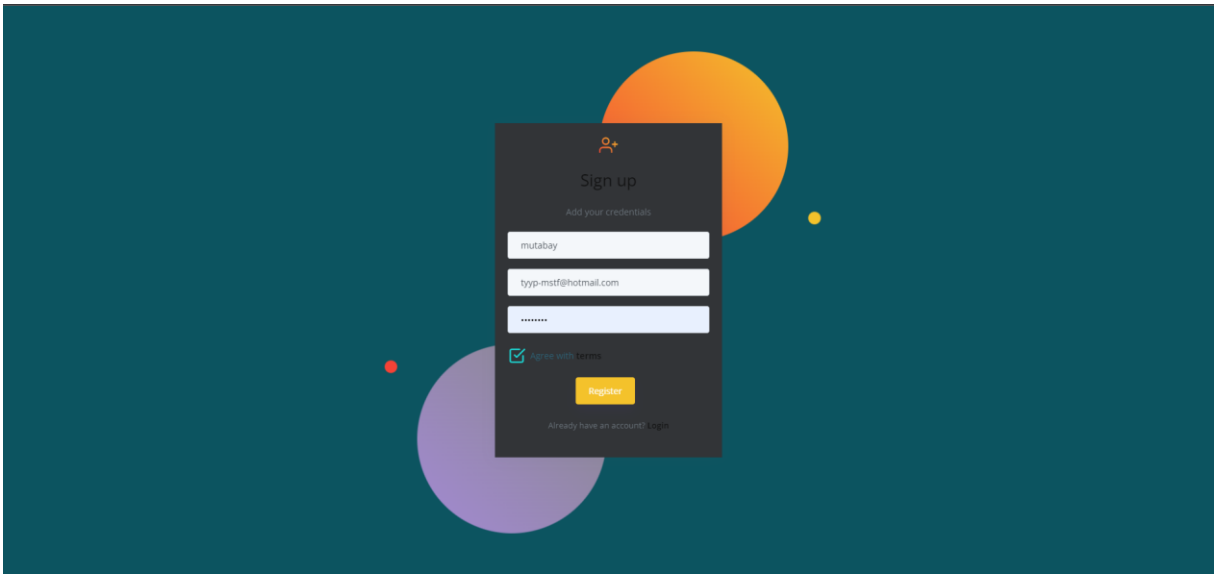


Figure 12- Sign-up Page

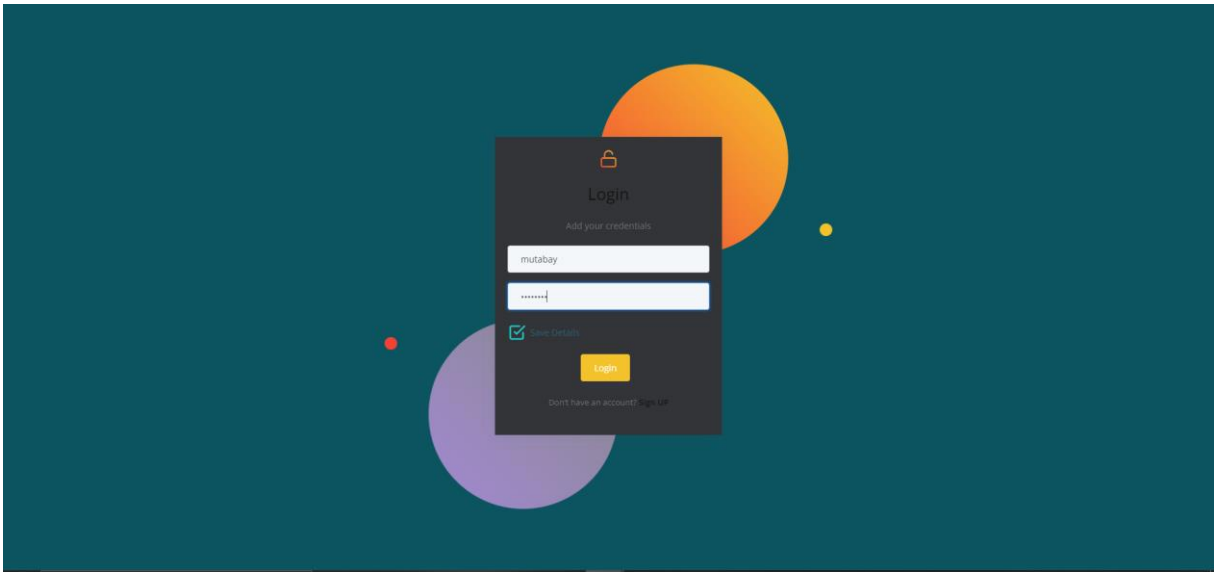


Figure 13- Sign-in Page

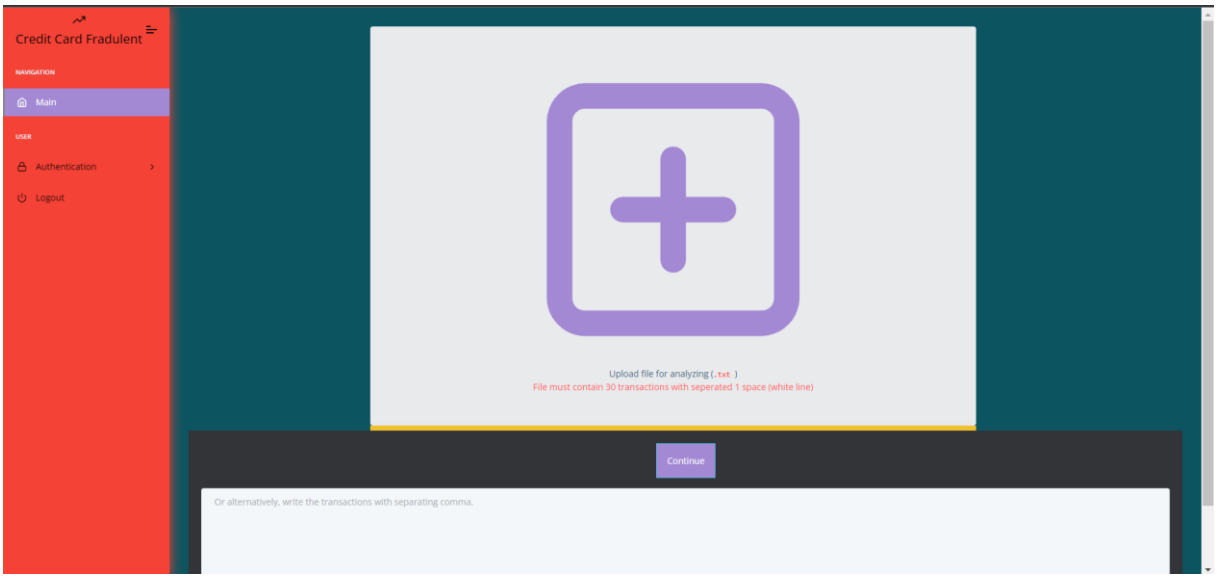


Figure 14 - Upload Transaction File

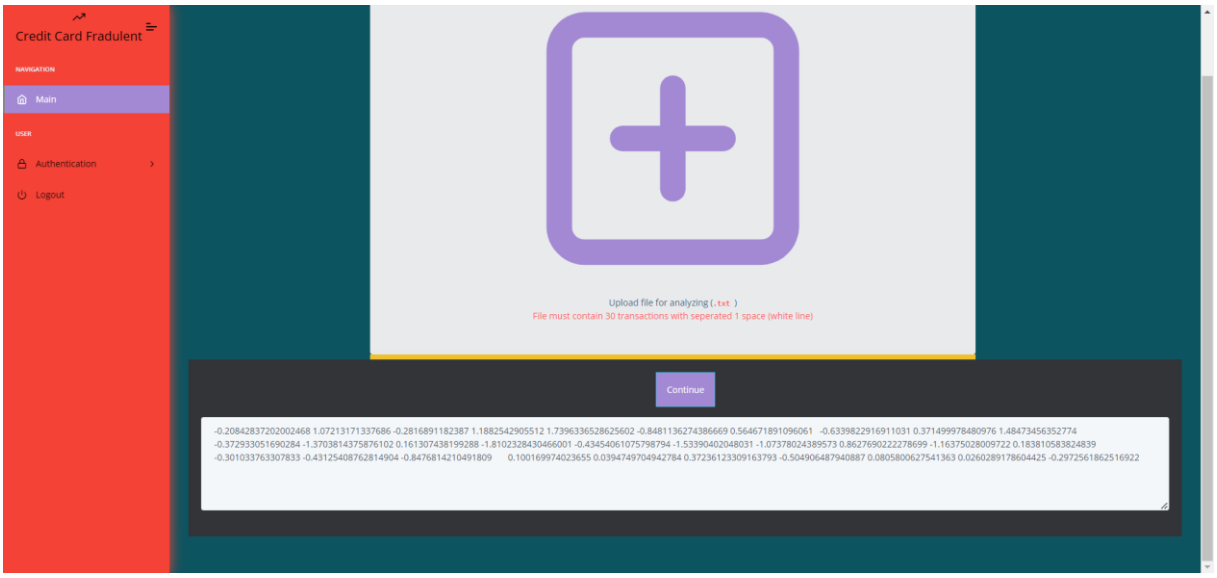


Figure 15 - Fill the text area and make predict

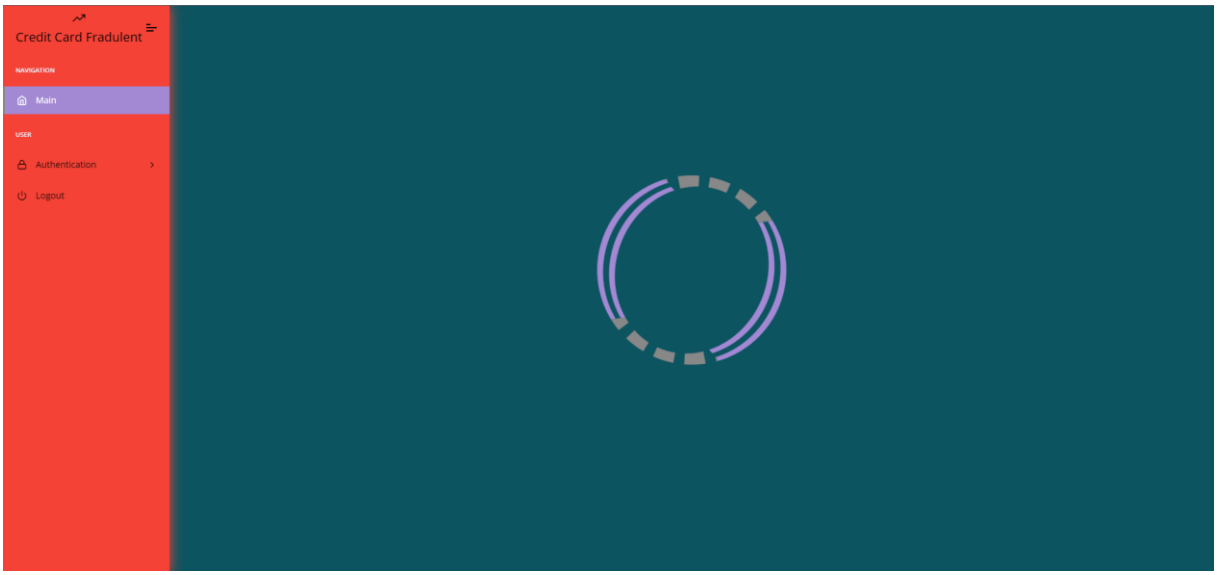


Figure 16 - Analyze Process

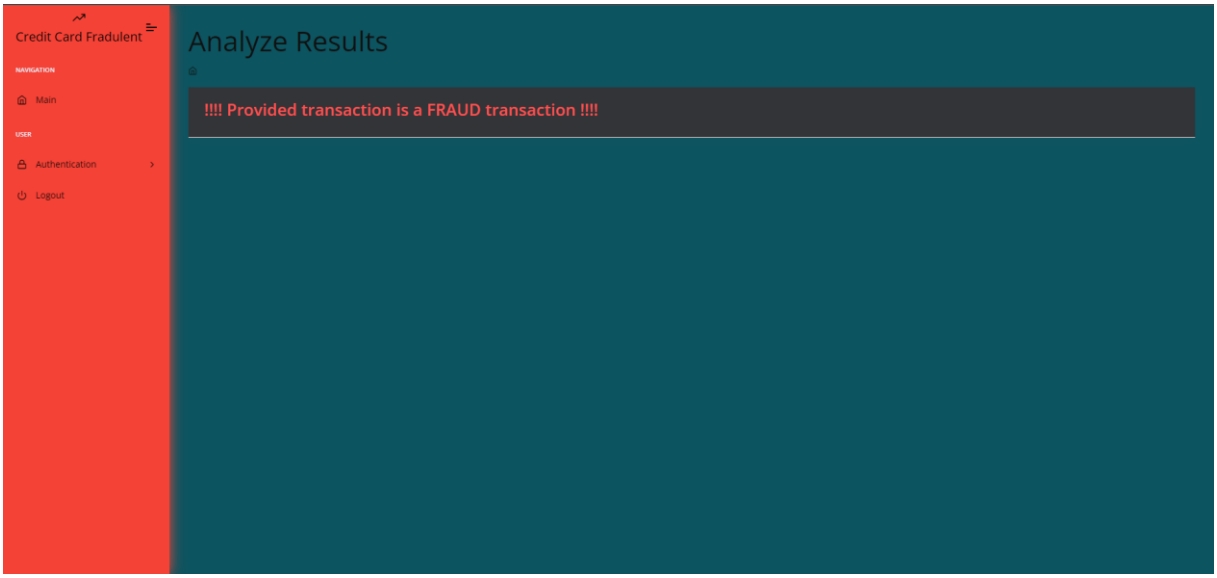


Figure 17 - Example of Model Result

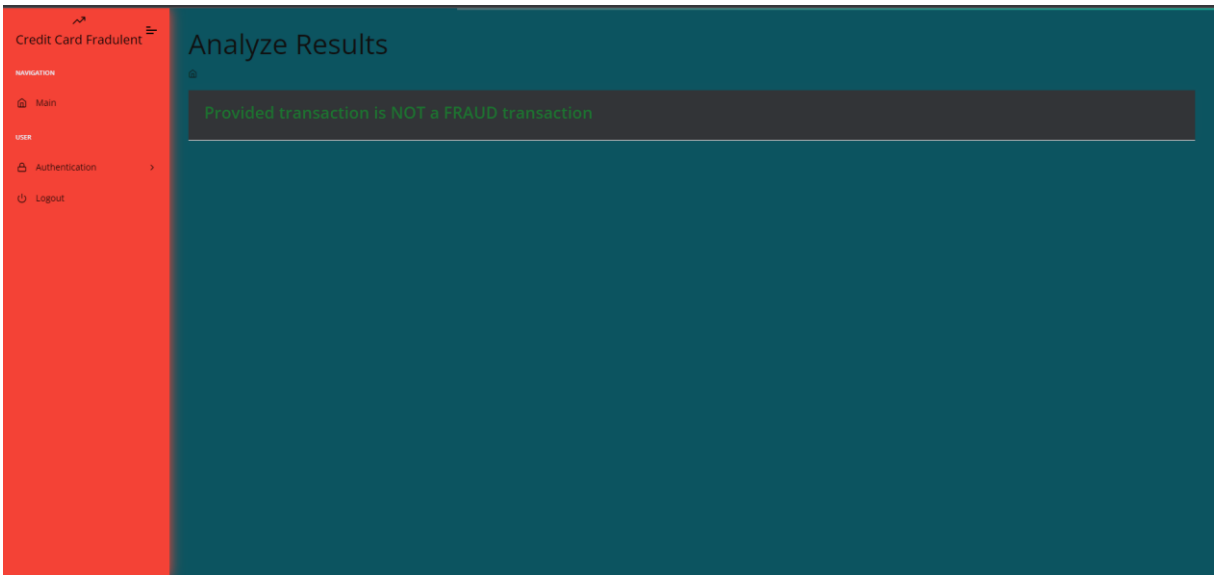


Figure 18 - Example of Model Result

Conclusion

Credit cards are a convenient way to pay for things, but they, like all other monetary payment methods, are vulnerable to hacking and other types of fraud. To address this issue, a solution is required to recognize transaction patterns and identify those that are fraudulent, making it much easier to spot such transactions in the future.

Machine Learning (ML) and Deep Learning (DL) are excellent tools for this task since they will assist us in identifying patterns in the data. If given adequate data, machine learning and deep learning can assist in creating excellent results. Furthermore, as technology progresses, ML and DL will improve, making it easier for a human to forecast if a transaction is fraudulent or not much more accurately.

References

[1 J. McCarthy, «WHAT IS ARTIFICIAL INTELLIGENCE?,» Computer Science Department, [Çevrimiçi].
] Available: https://homes.di.unimi.it/borghese/Teaching/AdvancedIntelligentSystems/Old/IntelligentSystems_2008_2009/Old/IntelligentSystems_2005_2006/Documents/Symbolic/04_McCarthy_whatisai.pdf. [Erişildi: 26 12 2021].

[2 IBM, What is Machine Learning?, 2020.

]

[3 IBM, What is logistic regression?.
]

[4 I. T. J. a. J. Cadima, Principal component analysis: a review and recent developments, 2016.
]

[5 IBM, What is the k-nearest neighbors algorithm?.
]

[6 IBM, «What is Deep Learning?,» [Çevrimiçi]. Available: https://www.ibm.com/cloud/learn/deep-
] learning. [Erişildi: 26 12 2021].

[7 H. T. M. A. a. A. B. Thanh Thi Nguyen, Deep Learning Methods for Credit Card Fraud Detection,
] 2020.

[8 geeksforgeeks, Deep Learning | Introduction to Long Short Term Memory, 2021.
]

[9 Brainstation, «What Is Web Development?,» [Çevrimiçi]. Available: https://brainstation.io/career-
] guides/what-is-web-development. [Erişildi: 28 12 2021].

[1 techopedia, «Database Management System (DBMS),» [Çevrimiçi]. Available:
0] https://www.techopedia.com/definition/24361/database-management-systems-dbms. [Erişildi:
28 12 2021].

[1 «SQLAlchemy,» [Çevrimiçi]. Available: https://www.sqlalchemy.org/. [Erişildi: 28 12 2021].
1]

[1 «Sequalize,» [Çevrimiçi]. Available: https://sequelize.org/. [Erişildi: 28 12 2021].
2]

[1 IBM, "CRISP-DM Help Overview," [Online]. Available: https://www.ibm.com/docs/en/spss-
3] modeler/SaaS?topic=dm-crisp-help-overview. [Accessed 26 12 2021].

[1 IBM, «CRISP-DM Help Overview,» [Çevrimiçi]. Available: https://www.ibm.com/docs/en/spss-
4] modeler/SaaS?topic=dm-crisp-help-overview.

[1 IBM, «Recurrent Neural Networks,» [Çevrimiçi]. Available:
5] https://www.ibm.com/cloud/learn/recurrent-neural-networks. [Erişildi: 26 12 2021].

[1 IBM, «What is an application programming interface (API)?,» [Çevrimiçi]. Available:
6] https://www.ibm.com/topics/api. [Erişildi: 3 1 2022].

[1 V. Paradigm, «UML Class Diagram Tutorial,» [Çevrimiçi]. Available: https://www.visual-
paradigm.com/guide/uml-unified-modeling-language/uml-class-diagram-tutorial/. [Erişildi: 3 1

7] 2022].

[1 T. Rascia, «Node.js, Express.js, and PostgreSQL: CRUD REST API example,» [Çevrimiçi]. Available:
8] https://blog.logrocket.com/nodejs-expressjs-postgresql-crud-rest-api-example/#whatisarestfulapi. [Erişildi: 2 1 2022].

[1 Ş. E. ŞEKER, «Yönetim Bilişim Sistemleri Ansiklopedisi Cilt 3, Sayı 3 Eylül 2016,» [Çevrimiçi].
9] Available: https://ybsansiklopedi.com/wp-content/uploads/2016/09/aralik20163-3kapak.pdf#page=19. [Erişildi: 18 18 2021].

[2 Wikipedia, «Long short-term memory,» [Çevrimiçi]. Available:
0] https://en.wikipedia.org/wiki/Long_short-term_memory. [Erişildi: 26 12 2021].