Arab American University

Faculty of Graduate Studies

Secure Internet Financial Transactions using Multifactor Authentication and Machine Learning

By

Alsharif Hasan Mohamad Abu Rbeian

Supervisor

Prof. Manuel Fernández-Veiga

This Thesis Was Submitted in Partial Fulfillment of the Requirements for

the Master's degree

in Cyber Security

January / 2024

I

Thesis Approval

Secure Internet Financial Transactions using Multifactor Authentication and Machine Learning

By

Alsharif Hasan Mohamad Abu rbeian

This thesis was defended successfully on 21/Feb/2024 and approved by:

Committee members	Signature
1. Prof. Manuel Fernández-Veiga (Supervisor)	……
2. Dr. Majdi Owda (Internal examiner)	……
3. Prof. Ana Fernández Vilas (External examiner)	……

Declaration

I certify that the thesis, "Secure Internet Financial Transactions using Multifactor Authentication and Machine Learning" is an entirely original work of mine, and hasn't been submitted for another degree or certification, and was completed for the Master's in Cyber Security at Arab American University - Palestine.

Student name: Alsharif Hasan Mohamad Aburbeian

Student ID: 202112885

Date: 15/March/2024

Acknowledgments

I deeply appreciate my godfather and thesis supervisor Prof. Manuel Fernández-Veiga**,** whose advice, constant support, and knowledge have been crucial throughout my academic endeavors. This research and my development as a scholar have been greatly influenced by your wisdom and support.

I pay a debt of appreciation to the institution I belong to; the Arab American University (AAUP) for creating an environment that encouraged growth and learning. My academic career has been greatly shaped by the direction of the faculty members and lecturers; Dr. Majdi Owda, Dr. Ahmad Hasasneh, Dr. Islam Amro9, and Dr. Huthaifa Ashqar. Your guidance and assistance have cleared my way and expanded my understanding, establishing the groundwork for the expert I am growing into. My godmother Dr. Amani Yousef Owda, your faith in my abilities and your dedicated support throughout my master's journey were the reasons for the success I achieved. Thank you for your commitment and for putting in me the values of knowledge and perseverance that will resonate throughout my life.

To the heartbeat of my existence, my mother and brother you have been my constant source of love and support as I have made my way through the wide world of life. You have been the silent cornerstones holding my dreams high with your kind heart and unwavering support. I'm grateful for your kind advice and never-ending support; without it, I couldn't have accomplished anything or overcome every challenge. Your grace and tenacity have continuously inspired me in the face of difficulties, and your wisdom continues to be the bedrock of who I am. My trip has been filled with crescendos and silences, but your love has been the song that has kept me going.

My beloved wife and kids, Abdullah, Osaid, and Hala who I adore, helped me get through the difficult times and long hours with your tolerance, empathy, and support. My academic success has been built upon the basis of your unfailing support.

To my friends and all those who helped me along the way, you have all contributed independently and significantly to my academic and personal development. Your assistance and efforts will always be valued components of this scholarly venture.

Abstract

The security of online financial transactions has emerged as a crucial concern in an era where financial services are becoming increasingly digital. The increasing use of digital platforms for banking, payments, and investment has given rise to a new wave of opportunities for both customers and cybercriminals. To address this problem, the present research offers a unique system that integrates machine learning (ML) with multifactor authentication (MFA). Using two levels of protection is the foundation of our system. The first layer uses two authentication factors, and the second layer is an embedded layer that asks for facial recognition from the user to successfully continue the purchase process if the ML model determines that the present transaction is fraudulent.

To select the best classifier for constructing the ML model, four supervised ML classifiers were put into practice. After testing many classifiers, including Random Forest (RF), Decision Trees (DT), Logistic Regression (LR), and Naïve Bayes (NB), the accuracy of each was 96.717%, 97.881%, 97.938%, and 92.354%, respectively. A front-end screen for an Android e-commerce application was created to demonstrate how the framework functions. You may configure our framework to operate on any digital e-commerce platform. A thorough analysis of the body of research on the subject and various methods for securing online transactions reveals that the integration of MFA and ML has great potential for providing the greatest level of security and a system that is easy to use. In future research, it could be useful to examine other authentication factors using a different dataset.

Table of content

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Description
MIM	Man-in-the-Middle
DoS	Denial of Service
SWIFT	Association for Worldwide Interbank Financial Telecommunication
IMF	International Monetary Fund
ROC	Receiver Operating Characteristic
ML	Machine Learning
FinTech	Financial Technology
PIN	Personal Identification Number
SIM	Subscriber Identity Module
OTP	One Time Password
SFA	Single Factor Authentication
2FA	Two Factor Authentication
MFA	Multi-Factor Authentication
EU	European Union
NIST	National Institute of Standards and Technology
AI	artificial intelligence
MDP	Markov Decision Process
GPS	Global Positioning System
SMS	Short Message Service
TIC	Transaction Identification Code
ID	Identification number
AUC	Area Under Curve
IP	Internet Protocol
RFID	Radio Frequency Identification
GSM	Global System for Mobile Communication
MTCNN	Multi-task Cascaded Convolutional Network

URL	Uniform Resource Locator
PCA	Principal Component Analysis
SMOTE	Synthetic Minority Over-sampling Technique
RF	Random Forest
DT	Decision Trees
LR	Logistics Regression
NB	Naïve Bayes
SVM	Support Vector Machines
KNN	K-Nearest Neighbors
GBM	Gradient Boosting Machines
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
TPR	True Positive Rate
FPR	False Positive Rate
LIME	Local Interpretable Model-Agnostic Explanations

Chapter 1

Introduction

1. Introduction

1.1. Overview

The majority of essential societal functions in the twenty-first century have grown reliant on digital infrastructure, especially the financial sector, whose operations are largely based on customer trust in the financial system as a whole. Nowadays, the majority of clients manage their accounts and make most point-of-sale payments using Internet banking, which has replaced banks as the main means of transferring money between institutions. Users store their personal and sensitive information on touch-screen devices such as smartphones and use them for online banking, E-commerce sites, and online bill payment processes [1].

Financial cybercrimes are trying to hack banking accounts, credit cards, or any other data associated with payment cards for financial purposes [2]. The hackers use a variety of techniques to target the security credentials, including Man-in-the-Middle attacks, secure socket layer attacks, impersonation attacks, password discovery attacks, session hijacking, eavesdropping, and Denial of Service (DoS) attacks. They were successful in accessing the database resources by using authentication credentials that were obtained illegally [3], [4]. Cybercriminals try to get access to bank employees' and consumers' confidential data and credentials to steal it and use it to access accounts and make illegal payment requests. Even the most inexperienced thieves can profit from the low risk and low expense of phishing attacks. Financial services can be turned disabled by

distributed denial of service attacks, which prohibit users from accessing their accounts and processing payments [5], [6].

A wealth of highly valuable personal data can be found in the reams of sensitive consumer data that banking institutions hold. The impacts of extensive data breaches, like the one that occurred in 2017 when Equifax stole the financial records of over 140 million individuals, destroy the confidence and trust that support the financial system. The cost of this lost trust means that, despite being hard to measure, the real cost of cyber robberies goes beyond just financial losses [7]. The banking industry is currently facing a significant cyber danger from hackers acting on state orders. With the support of state governments' resources, they possess the capacity to seriously damage the financial sector. North Korea maintains specialist teams dedicated to financial institution cyberattacks. Teams like the "Lazarus Group," whose famed "WannaCry" ransomware attacks led to losses in billions, attacks on the Association for Worldwide Interbank Financial Telecommunication (SWIFT) network have resulted in losses of over one billion dollars, with clear attempts to target national banking institutions [8], [9].

Securing these transactions from fraudulent actions becomes an issue in the light of rise of cybercrimes. Profit-driven cybercriminals find Canada's financial industry to be a lucrative target since it experiences millions of daily intrusion attempts and is more vulnerable to crimes made possible by technology, such as credit card fraud. The financial sector is the most vulnerable to cybercrime, with losses estimated to be three times higher than those of any other industry. According to a recent International Monetary Fund (IMF) assessment, banks might lose an average of US$97 billion a year, or nine percent of their total revenue, due to cybercrime [10].

1.2. Problem Statement and Motivation

Financial transactions through the internet become familiar in recent years, and over the preceding three years, the volume of phishing attacks aimed at stealing identities or accounts has risen to over triple [11]. Hackers gained access to internet users' accounts worldwide in the 3rd quarter of 2022, resulting in fifteen million privacy violations [12]. Because of this, it is more crucial than ever to protect them. Because of this, the study provides a framework to protect financial transactions over the Internet by merging machine learning (ML) techniques with multi-factor authentication (MFA). This adds extra security layers that can assist in shielding confidential data and stop fraud.

As a cyber security master's student, exploring the effectiveness and feasibility of these technologies in securing online transactions is both interesting and relevant to my field of study. Furthermore, it provides an opportunity to aid the creation of more advanced security solutions in the area of online transactions, which will have practical applications for businesses, individuals, and society as a whole.

1.3. Study Objectives

This study offers a framework to secure Internet financial transactions using MFA and ML. The goals of this research are: -

- Choose the suitable authentication factors to implement in the MFA model.
- Download a dataset of online credit card transactions to build the ML model.

- Data pre-processing phase consists of data cleaning, exploration, visualization, and feature selection.
- Investigating different supervised ML classifiers and choosing the best performer to build the ML model.
- Learning, evaluating, and creating the ML model.
- Test the result of the ML model using different metrics to improve the accuracy.
- Combine the ML model as a part of the MFA framework to gain the highest possible security and user-friendly model.
- Design an Android e-commerce application front-end screens.

1.4. Significance of the Study

The study's importance stems from its ability to:

- First, offering a model that can be implemented in the banking sector, E-commerce purchasing websites, and online payment systems.
- Second, using ML as part of MFA will achieve the highest possible security for securing financial transactions.
- Third, it shows the best way to use MFA in a convenient and user-friendly way.
- Fourth, provide a thorough examination of the most appropriate ML algorithms and training methods to use in combination with MFA for online transactions.
- Fifth, the possibility of modifying the ML algorithm to comply with the requirements of any electronic system and integrating this algorithm with MFA to provide secure, and user-friendly access to data.

1.5. Contribution

- Address a critical problem in the field of cyber security.
- Improve the safety of Internet financial transactions.
- Provide an evaluation of the feasibility and effectiveness of such a system.
- Build a robust understanding of the best practices for combining these technologies.
- Showing the best way to use MFA in a convenient and user-friendly way.

This will highlight the importance of using MFA and ML in Internet financial transaction security and help advance the state-of-the-art in this critical area.

1.6. Ethical Issues

In the current digital era, the widespread usage of internet-based financial transactions has fundamentally altered how people do banking and commerce. However, protecting these consumers from fraud is a significant issue that comes along with this ease of use. Financial institutions use a variety of methods, including ML, to analyze user behavior and prevent fraud. Financial institutions' usage of financial data as security solutions brought up moral issues that need to be addressed. Some of these issues are: -

- Data Privacy and Consent

In the world of online financial transactions, protecting users' privacy and getting their informed consent for data collection and processing are fundamental values. For users, to continue to have faith and confidence in the MFA and ML framework, there must be transparency about the usage

of personal data. Financial organizations should respect moral principles and safeguard users' privacy rights by highlighting the significance of informed consent and data transparency.

- Fairness and Bias

If ML algorithms are not properly controlled and minimized, they could worsen existing biased and discriminatory practices. Fairness in the training data used for ML models and the MFA system's decision-making processes must be ensured to mitigate this danger. Ensuring equal treatment of all users requires strategies for identifying and resolving prejudice, such as varied representation in training datasets and algorithmic fairness assessments.

- Safety and confidence

It is of the highest ethical importance to protect users' financial information against online attacks. It is essential to have policies in place to stop illegal access to or improper use of private data, building consumer confidence in financial institutions. Institutions can fulfill their ethical obligation to safeguard users' financial assets and data by giving security measures the highest priority and following best practices in cybersecurity.

- Social Impact

Using cutting-edge security measures in Internet banking can have a significant impact on society as a whole, especially for fraud-affected groups. It is critical to address adoption or access restrictions for vulnerable populations and consider the possible impact on financial services accessibility. Financial institutions can show their dedication to social justice and ethical responsibility by proactively addressing societal issues and fostering inclusivity.

In summary, maintaining the values of privacy, justice, security, autonomy, openness, responsibility, and societal effect requires addressing the ethical issues raised by MFA and ML technologies. Financial institutions can encourage confidence, improve user satisfaction, and promote the integrity of online financial transactions for all parties involved by placing a high priority on ethical behavior.

1.7. Thesis Arrangement

The following Figure 1.1 shows how the thesis is organized.

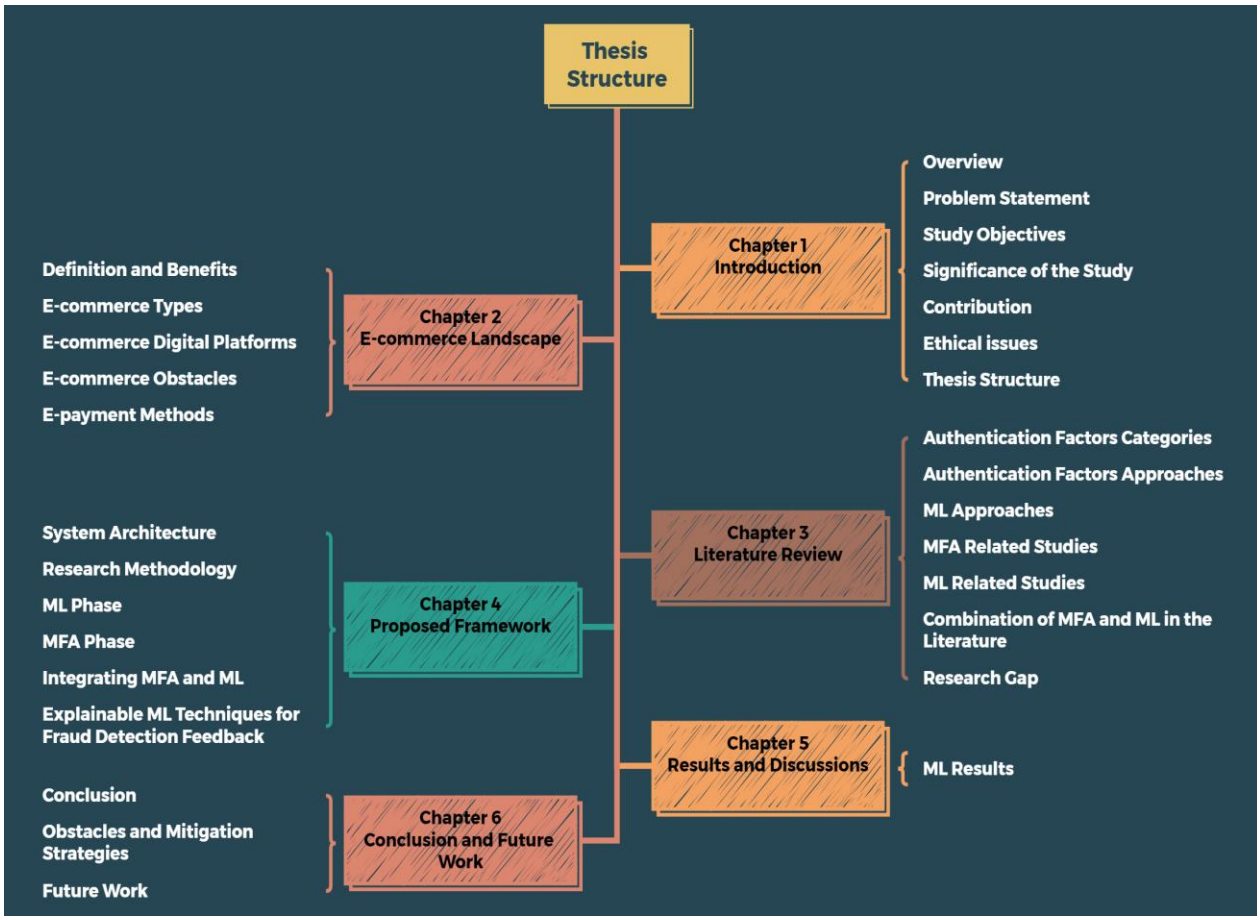

Figure 1.1. Thesis Structure

As shown in Figure 1.1, the thesis consists of six chapters which are: -

- **Chapter Two**: Explores the E-commerce concept landscape.
- **Chapter Three**: provide an overview of authentication factors categories and approaches, ML approaches. To address the research gap, this chapter will discuss ML and MFA related work.
- **Chapter Four**: shows the system architecture, the methodology used to perform the study, ML and MFA model building steps, illustrates the proposed framework which combines MFA and ML, and discusses the explainable ML techniques.
- **Chapter Five**: shows the ML model results.
- **Chapter Six**: Summarizes the overall thesis, discusses the obstacles that faced this study with the mitigation strategies, and suggests future work.

Chapter 2

E-commerce Landscape

2. E-commerce Landscape

2.1. Introduction

This chapter will discuss briefly the E-commerce concept from different perspectives such as definition, types, benefits, and challenges. An overview of the most common E-commerce platforms will be stated. Finally, this chapter will survey the common electronic payment techniques.

2.2. Definition and Benefits

Electronic commerce is a modern trading strategy that includes digital avenues for Internet shopping and the trading of products and services. In contrast to traditional trading, E-commerce runs online, made possible by any online platforms [13].

The potential of e-commerce to go beyond national boundaries and connect companies with a worldwide customer base is what makes it so inconvenient. This is in sharp contrast to the localized strategy of traditional trading, which requires clients to visit stores in person. The ease of use and accessibility of e-commerce is one of its main benefits. Clients can shop whenever and wherever they want, without being restricted by real store hours [14].

From a business owner's perspective, E-commerce also helps companies cut expenses by simplifying procedures and lowering overhead related to operating physical stores. E-commerce transactions' digital format also makes data-driven insights easier to obtain, enabling companies to customize customer experiences and sell products based on unique desires. Furthermore, e-commerce's worldwide scope creates new opportunities for companies to grow their market share and establish connections with a wide range of customers. E-commerce is a crucial tool for contemporary enterprises because of its convenience and capacity to execute payments safely over the Internet [15].

2.3. E-commerce Types

The literature has several categories for different forms of e-commerce, but the four primary categories are shown in Figure 2.1.

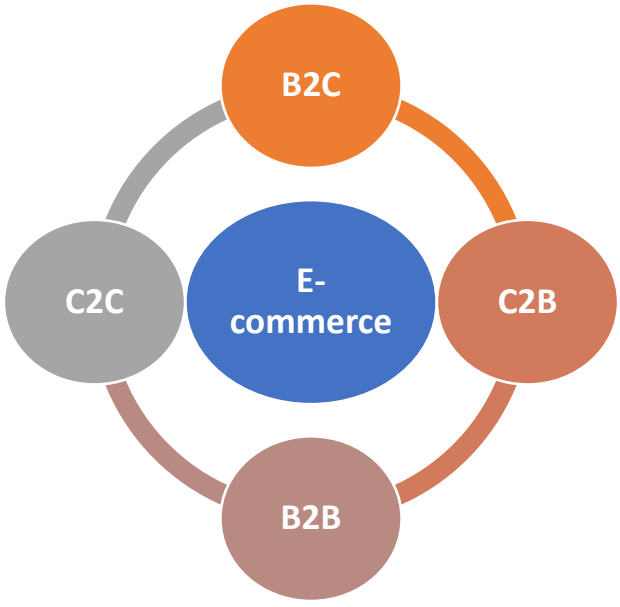

Figure 2.1. E-commerce Types

As shown in Figure 2.1, the main e-commerce types are B2C, C2B, B2B, and C2C which will be illustrated separately to highlight the differences between each type.

- **Business-to-Consumer (B2C)**

Electronic business partnerships between companies and their final clients; business-to-consumer e-commerce. This is the part of the website dedicated to online shopping, where traditional retail transactions usually occur. These types of partnerships can be ended, made easier, more complex, or sporadic. Ever since the Internet was created, this business model become more popular. There are now a lot of online stores and marketplaces that provide consumers with a huge landscape of things, including devices, technology, textbooks, and supplies. In contrast with conventional commerce's retail sales, buyers are usually more aware of the available educational material and it is widely acknowledged that buying something less expensive doesn't mean that the end-user experience is any less tailored or that manufacturing and shipping are any simpler [16].

- **Consumer-to-Business (C2B)**

In C2B, the trade of products is inverted. People also sell their goods or services to companies that specialize in particular categories. These events include venues where artists seek multiple logo concepts, to select and purchase just one successful submission. In this business sector, marketplaces that offer images, media, and design components for free are another well-liked format [17].

- **Business-to-Business (B2B)**

This kind of e-commerce is more concerned with the interchange of products and services between businesses than it is with individual customers. Companies using the Internet to purchase raw materials from vendors are one example [18].

- **Consumer-to-Consumer (C2C):**

Via online platforms, consumers can sell to other consumers directly in C2C e-commerce. A third party offers a digital platform that handles this type of exchange. These platforms make it easier to exchange goods and services. Examples include websites where people may purchase and sell secondhand goods, such as eBay [19].

2.4. E-commerce Channels

To enable smooth transactions and interesting consumer experiences, E-commerce utilizes a range of digital channels, some of which can be seen in Figure 2.2.

Figure 2.2. E-commerce Digital Channels

Figure 2.2 shows some E-commerce platforms which are [20], [21], [22], [23]: -

- **Websites**

websites, which act as companies' online shops. These websites are thoughtfully created to highlight goods or services, give comprehensive details, and have an easy-to-use interface for users. A carefully planned and easy-to-use website can serve as an online shop with many benefits for both customers and businesses [20].

- **Online Marketplace**

Digital marketplaces such as Amazon, eBay, and Etsy provide a venue for companies to promote and market their goods to a wide client group. These markets manage distribution and payment processing and offer an already-prepared client base [20].

- **Search Engines**

A major factor in bringing customers to e-commerce websites is search engine traffic. Companies use search engine optimization (SEO) tactics to make sure they show up widely in search results, which increases publicity and pulls in potential clients [21].

- **E-mail Marketing**

One popular digital avenue for connecting with current consumers and bringing in new ones is email marketing. Companies notify clients about new products, sales, and events by sending out e-newsletter emails with promotions, and customized offerings [21].

- **Social media platforms**

These channels are important to e-commerce because they give companies a direct line of communication with potential customers benefitting from the large number of users for each

platform. Social media networks such as TikTok, and Facebook allow businesses to display their items, target customers with ads, and even conduct payments [22].

- **Mobile Applications**

Another important digital channel is mobile applications, which give companies the ability to interact with clients on their smart devices and phones. To improve the overall shopping experience, these apps frequently include extra features like push notifications, customized suggestions, and easy payment alternatives [23].

2.5. E-commerce Obstacles

The E-commerce industry faces an enormous amount of challenges, some of these challenges are [24], [25], [26], [27], [28], [29]: -

- **Security Issues**

Preserving the confidentiality of consumer data and Internet transactions is a recurring problem. Cybersecurity concerns put customers and organizations at risk, especially, in light of the rise in fraud and data breaches [24].

- **Payment Gateway Problems**

Electronic payment gateways play a major role in e-commerce. Issues with these gateway's security, payment processing mistakes, and technical difficulties can affect user trust as well as experience and overall the E-commerce industry [25].

- **Multinational competition**

Due to the internet's extensive international scope, firms must contend with strong competition on an international scale. Adapting to varied consumer tastes and standing out in a crowded market become a huge challenge [26].

- **Management and Transportation**

Efficient and efficient shipment is vital for client satisfaction. Shipping expenses, delivery schedules, and guaranteeing that goods arrive to clients unaltered are among the difficulties [27].

- **Returning and Refunds**

It can be difficult to handle returns and refunds in a way that minimizes losses while satisfying customers. Achieving the ideal harmony between cost control and customer happiness is essential [28].

- **Various cultural practices and traditions**

Traditional negotiating is hampered by the fixed pricing schemes used by e-commerce platforms. Online price negotiating may be tough for people who are used to disputes face-to-face. For individuals who prefer the dynamic and engagement aspects of in-person negotiations, the lack of direct negotiation possibilities may prove to be a barrier [28].

- **Examination Challenges**

Another barrier to online shopping is still being unable to visually check things before making a purchase. While technology has made it easier to evaluate products virtually using pictures, descriptions, and evaluations from customers, the lack of a physical experience can raise questions. Without physical confirmation, customers could be reluctant to make a purchase, especially for commodities where physical characteristics like touch, smell, or size are important [29].

To get past these barriers, companies need to implement a well-thought-out plan, adapt constantly, and be dedicated to providing a safe and easy online purchasing experience.

2.6. E-payment Methods

In the world of technology, companies aim to give their customers a hassle-free buying experience by offering quick and simple checkout processes E-commerce payment platforms give customers a wide range of options, and grant them the freedom to select the payment method of their choice [30].

Electronic or E-commerce payment systems, allow transactions to be completed via an electronic method without requiring the presence of cash when conducting a payment. E-commerce employs E-payment methods to get paid for products. E-commerce payment systems have significantly transformed the process of conducting business online by simplifying it for both consumers and organizations [31]. In an E-payment system, a payment method connects an internet-based shop to the payment processing system of customer choice, see Figure 2.3.

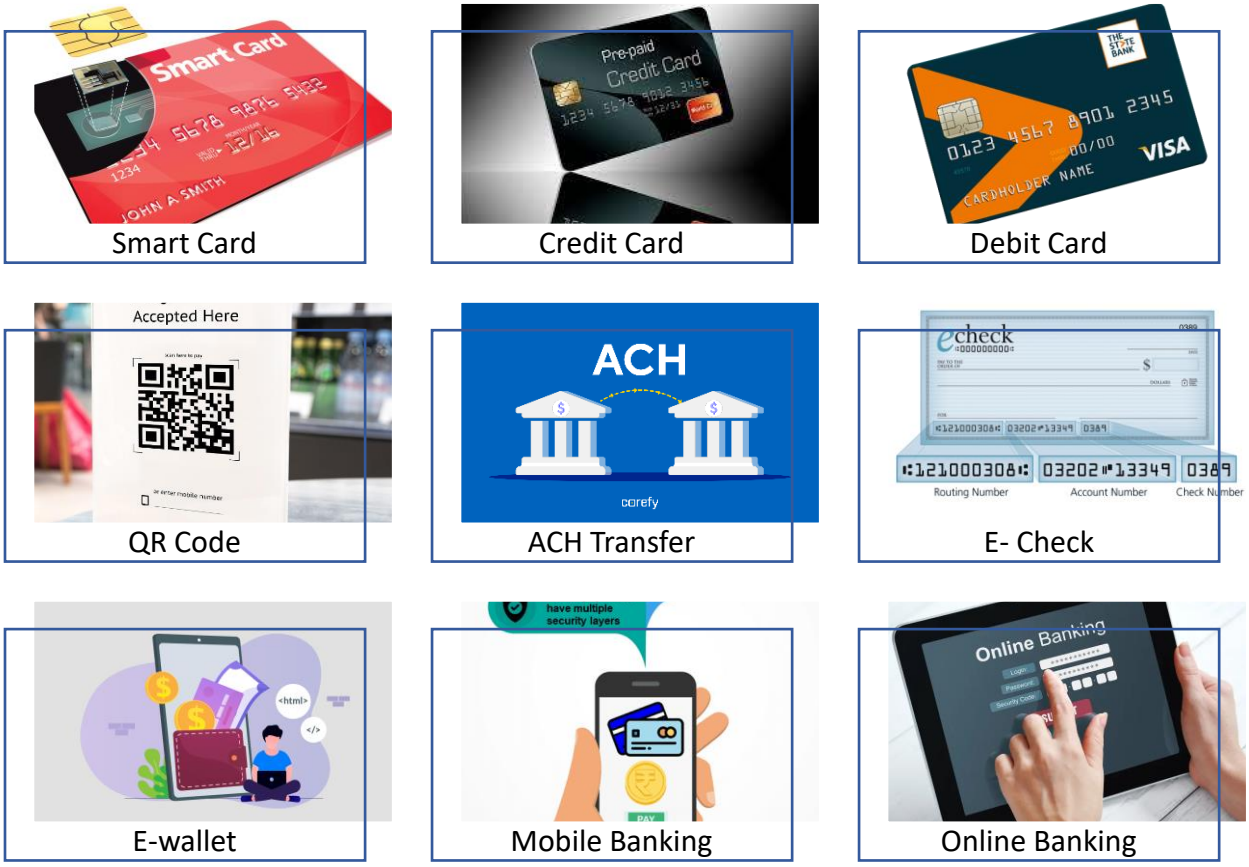

Figure 2.3. E-payment Channels [194], [195], [196], [197]

Consumers have several safe e-commerce payment choices to choose from which work with banks to clear business cash. Many payment methods exist as Figure 2.3 shows which are: -

- **Credit Cards**

Typically provided by a bank, a credit card is a card that enables its owner to make purchases of goods or services or make cash withdrawals on credit. By using the card, you acquire debt that you must repay later. Among the methods of payment that are most often used worldwide are credit cards. Credit cards enable users to accrue a continuous debt load, which is subject to interest charges. A charge card only delays the customer's payment until a later time; in contrast, a credit

card typically entails a third-party business that pays back the seller when the buyer repays their purchase. [32].

- **Debit Card**

A debit card is an identity card connected to the user's bank account. What separates the debit card from the credit card is that the debit card takes money directly out of funds from the user at the moment of the transaction [32].

- **Smart Card**

Credit cards, debit cards, and smart cards have somewhat similar appearances. On the other hand, smart cards have an integrated microprocessor chip. Account balance and a customer’s personal data can both be stored on it. Smart cards provide for faster processing at reduced rates [32].

- **E-Wallet**

One way to define the E-Wallet is as an account that allows users to perform payments without the need to provide E-payment method details, allowing them to securely store multiple payment method details. Every day, the number of users is growing. people who use e-wallets can speed up the checkout process by avoiding having to enter their card information every time [32].

- **E-Check**

An electronic check, or e-check, is a payment made electronically from your checking account. An e-check functions similarly to a traditional check, except rather than removing a sheet of paper from your checkbook, you submit payment details. Via an e-check approval form. This gets your payment ready for electronic processing [33].

- **ACH Transfers**

A bank-to-bank electronic money transfer is called an Automated Clearing House (ACH) transfer. Money is taken out of one account and put into another. These transactions are typically less expensive than bank transfers and take advantage of the ACH network. Similar to how regular bills or subscription payments are made, ACH debit transfers involve taking money out of an account. Money can be transferred (pushed) to other accounts via ACH credit transactions. This kind of transfer can take the form of an agreement between a company and its workers or customers [33].

- **Online banking**

Direct bank account payments are a handy way for clients to make transactions. The customer must sign up with their bank account without needing to acquire any type of card to use this payment method. When making a purchase, the client simply must give their PIN and net banking ID and the payment will be processed from the user's bank account directly [34].

- **Transactional QR Code**

The popularity of using QR codes to make payments has increased dramatically. It is a square grid of rectangles or squares aligned to form a pixel pattern code. The data contained in every part is the code. This data may include transactional information, retailer information, and other information. Using a smartphone or tablet to scan the QR code is necessary to make payments [34].

- **Mobile Payment**

Shoppers may use mobile devices to make purchases quickly and conveniently with mobile payments. All that is required is to install a payment application. After that, to make purchases, the user must link his bank account to the installed application. The app gets a payment request when

a customer decides to use it to purchase from an internet retailer, and it has to authorize it to complete the transaction successfully [35].

In summary, e-commerce payment methods have become essential for various reasons, including ease of use, safety, accessibility from anywhere in the world, user experience, and flexibility in response to emerging technology. Businesses engaged in e-commerce that prioritize payment options and do it thoroughly and user-centric will be more likely to succeed in the competitive and ever-changing digital marketplace [36].

2.7. Summary

This chapter covered the exploration of the many aspects of e-commerce briefly, revealing the subtleties that form the online market. It began by providing a precise description and outlining the many advantages of e-commerce, demonstrating how it exceeds traditional trading and provides numerous benefits for users and enterprises.

A variety of interconnections within this continually changing environment were revealed by the unfolding of the different types of e-commerce kinds, which included B2B, B2C, C2C, and C2B models. A thorough exploration of digital platforms revealed how essential they are to enabling Internet payments and creating outstanding user interaction. The chapter overcame the obstacles that customers and businesses face, including safety issues, restricted bargaining opportunities, trust-related concerns, the difficulties of virtual examination of products, and the unique characteristics of E-payment systems.

Finally, the information that has been synthesized from the chapter's various sections sheds light on the complex and ever-changing environment of e-commerce, giving readers a general understanding of its definition, complex nature, and the challenges that may affect the ongoing expansion.

The next chapter will explore the authentication factors categories and approaches, ML approaches, and discuss the MFA and ML related work to address the research gap.

Chapter 3

Literature Review

3. Literature Review

3.1. Introduction

FinTech is defined as an emerging financial technology company that improves and mechanizes financial services. The shortcomings of stringent laws governing conventional banking systems are largely responsible for the growth of financial innovations such as mobile wallets, banking, payment gateways, etc. The increased acceptance of financial services is a result of the technological advancements that offer quick, affordable, and user-friendly financial services [37]. Generally speaking, financial technologies like online transactions also called wireless transactions; allow wireless-based retailers to handle and enable payments powered by E-commerce platforms. Internet financial transactions can be made using a credit card, smart card, or mobile device [38], [39], [40], [41], [42].

This chapter will offer: -

- Exploring authentication factors categories.
- Discuss the authentication factors approaches.
- Survey the ML different approaches.
- Review the MFA related work.
- Review the ML related studies for securing financial transactions.

- Discuss the combination of MFA and ML in the literature.
- Identify the research gap in the literature.
- Summarize the main points and goals of this chapter in brief.

3.2. Authentication Factors Categories

Authentication factors are the crucial components that ascertain the validity of an individual's identification during the sign-in procedure, acting as the core components of today's security frameworks. The authentication techniques can be one of three categories; something you know, something you have, and something you are [43]. See Figure 3.1

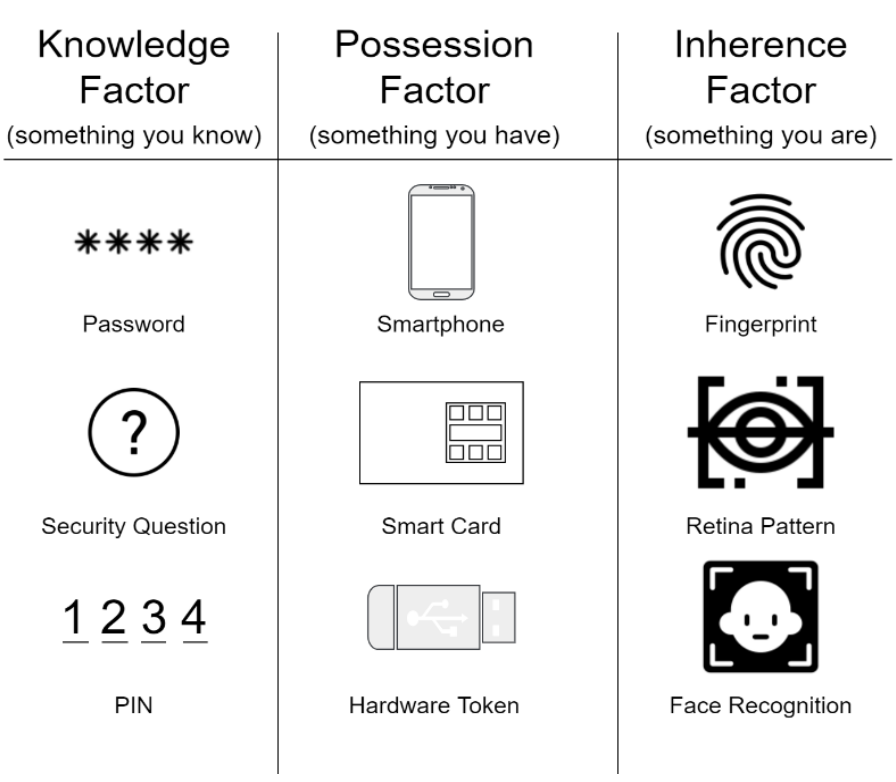

Figure 3.1. Authentication Factors Categories [198]

There are different factors examples to authenticate users as shown in Figure 3.1, which will be discussed in different sections.

3.2.1. Something You Know

This category includes factors that are knowledge-based also called "memorized factors." The conventional username and password combination is the most used example. Users must provide information, like a password or personal identification number (PIN), that should only be known to them. There are some problems with passwords. A violent person may easily rob, crack, shoulder surf, or predict the password. Another method of this type of authentication is a security question. Users can set security questions in certain systems. Security questions ask for responses to things like car type and favorite hoppy that are easy for someone else to know from a discussion through clever manipulation (social engineering) [44].

3.2.2. Something You Have

This category includes elements of a particular physical form item that an individual owns. To satisfy the Possession Factor, a user must show that they have a tangible object, such as a SIM card, smartphone, smart card, or hardware OTP token. The Possession Factor was much simpler to use with the development of contemporary technologies. Many authentication techniques are considerably more simple and more secure than the traditional user name and password. Unlike the Knowledge Factor, which is easy to hack, the Possession Factor verifies if a person possesses hardware or not. An intruder can gain access to this hardware remotely, carry out a successful swapping assault, or steal the hardware. All of these tasks are still far more challenging than executing a straightforward brute-force attack [45].

Using this kind of authentication factor carries the danger of allowing hostile actors to access your device. When all communication takes place over a network in the modern world, the hostile actor does not even need to take the device. To persuade users to let them remotely access your device, they may employ social engineering techniques. Occasionally, the malevolent party need not even have any kind of access to the user's phone. Certain authentication techniques are exposed to Man-in-the-Middle (MITM) assault, wherein cybercriminals can get a user's identity by intercepting the exchange between the user and the security system. An attacker's task of breaking into the device is considerably more difficult when using a robust authentication mechanism based on the possession factor [43].

3.2.3. Something You Are

These components come from the individual biological or behavioral characteristics of the user. They include actions like typing speed and movement, biometric characteristics like fingerprints, facial and voice recognition, retina pattern scan, and facial recognition. Because biometric identification is accurate and convenient, it has become more and more popular. Nevertheless, its implementation may need additional resources and give rise to privacy and data security problems [46].

3.3. Authentication Factors Approaches

The amount of information users must provide to verify their identity often varies depending on how sensitive the data and digital resources are. For instance, to access their online accounts, customers of online retail sites frequently just need to submit one reliable credential like a

password. Although anyone may see what users bought on a specific website, the user's private information is safe. However, since financial institutions deal with far more valuable data, such as account balances and payments, they frequently demand that users give a minimum of two pieces of authenticated information to gain access to their online accounts.

3.3.1. Single-Factor Authentication

Users can verify themselves with SFA by providing just one piece of confirmed data. This data might be anything from a biometric (like a fingerprint) to a knowledge factor (like a password). Because single-factor authentication (SFA) is a simple and easy method to use, it has been widely utilized to secure communication between two entities, such as the use of a PIN code [47]. Keep in mind that SFA isn't always less safe than MFA or 2FA. SFA does not relate to the method of authentication employed, but rather to the number of factors required. Due to its great exploratory potential, SFA is the least reliable authentication method [48], [49]. The account is instantly compromised when the password is shared. Additionally, a dictionary attack [50], a rainbow table [51], or social engineering techniques [52] can be used to get illicit access. After it was shown that, due to several security flaws, single-factor authentication is inappropriate for offering adequate protection [53]. But fingerprints are another kind of SFA, and as they are hard to forge, they are considered one of the safest techniques available. See Figure 3.2 which illustrates the working principle of SFA.

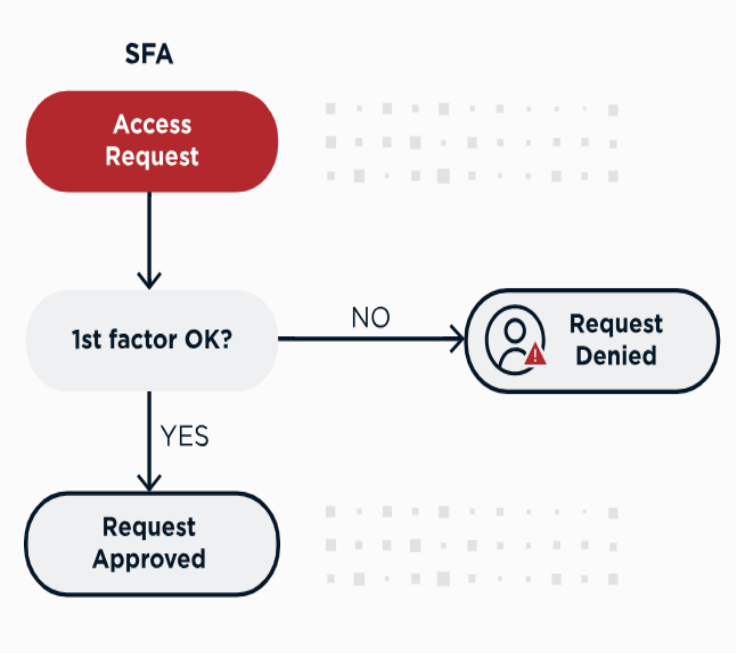

Figure 3.2. Single-Factor Authentication Workflow [199]

As shown in Figure 3.2, users of SFA must authenticate with one piece of verifiable information. The system will verify the user's identity in three steps: -

- Users input the necessary data, which may include their fingerprints, a PIN, or a password.
- The online facility verifies that the data it has received matches the authentication data it has on record.
- Users get access if the information they gave for authentication matches the data in the system's database. Otherwise, the access will be rejected [54].

3.3.2. Two-Factor Authentication

Two-factor authentication (2FA) was created to provide additional protection for sensitive data in which a user must provide two identification data to confirm identity [55], [56], [57]. Since primary login credentials and passwords are frequently lost or stolen, 2FA may be used to help make sure

that private data is protected. Based on the European Union Regulation (EU) 2015/1502 [58], a robust authentication system necessitates the utilization of a minimum of two elements related to separated categories. Similar requirements are addressed in NIST papers [58], [59], which also provide a high-security level authentication method based on two authentication elements. For instance, for users who want to access their systems, sign-on procedures may ask for their usernames and passwords, which are pieces of information about what users know, in addition to a fingerprint which are pieces of information about what users are [60]. Alternatively, sign-in procedures may ask users to provide their passwords and usernames (something users know) as well as information proving they own stored in their smartphones (something users have) [61]. Users that utilize 2FA must authenticate with two pieces of verified information, see Figure 3.3.

Figure 3.3 shows the steps to verify user identity using 2FA which are: -

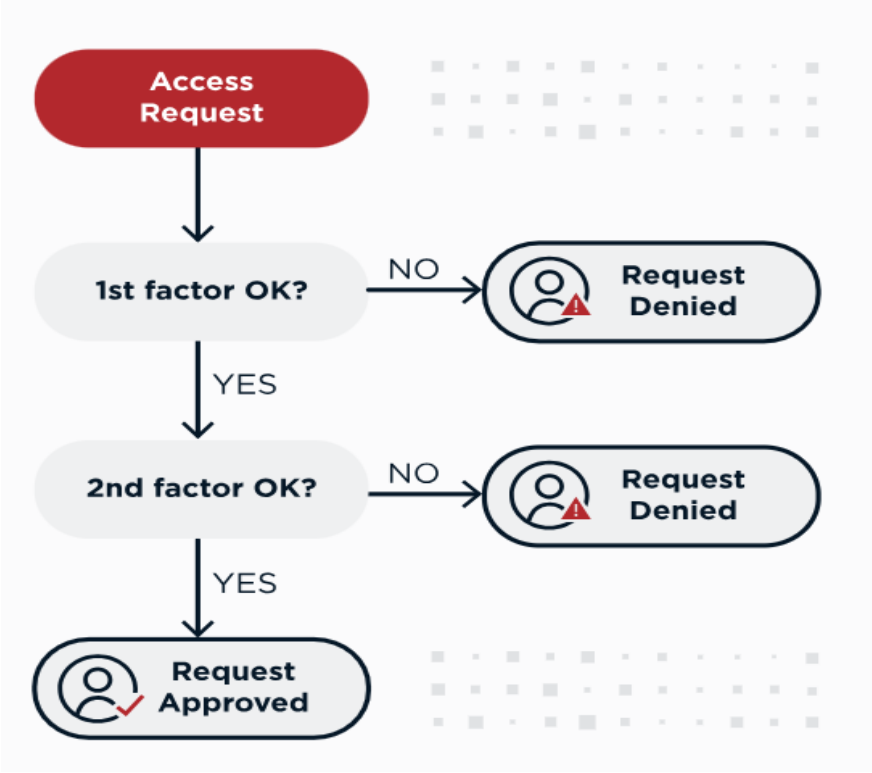

Figure 3.3. Two-Factor Authentication Workflow [199]

- Users enter the first needed piece of data.
- The online resource verifies the data by comparing it with the authentication data it has on file.
- If the authentication data they gave corresponds to the data in the system, users are prompted to submit the second identification data, which may be a face recognition or a one-time password (OTP).
- Users successfully log in, if the authentication information meets the information in the system. Otherwise, users are not allowed to access the system [55], [56], [57].

3.3.3. Multi-Factor Authentication

Verification mechanism at which users must deliver multiple authentication credentials (more than two) when requesting access to an online system [62]. MFA was recommended to significantly raise security levels and add effective protection against account theft. MFA greatly increases the difficulty for to expose information systems, even if passwords or personal identification numbers (PINs) are compromised by using a layered architecture. Similar to 2FA, distinct categories must include the bits of verifiable information that are sought. In addition to requiring users to enter their usernames and passwords (something users know), sign-in procedures may additionally ask for a smartphone or hardware token (something users have), and biometrics, like a fingerprint or retinal scan (something users are). Typically, MFA requires more than two certifications to authenticate, see Figure 3.4.

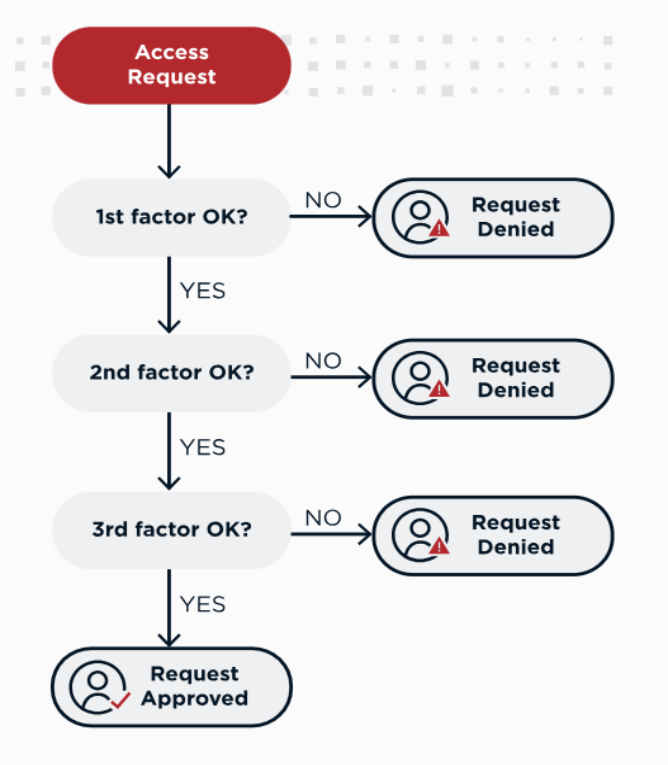

Figure 3.4. Multi-Factor Authentication Workflow [199]

As shown in Figure 3.4, the steps to verify user identity using MFA are: -

- Users enter the first needed piece of data.
- The online resource verifies the data by comparing it with the authentication data it has on file.
- If the authentication data they gave corresponds to the data in the system, users are prompted to submit the second identification data.
- If the user passes the second factor successfully, users will be requested to submit the third identification data.
- Users are given access if the submitted data meets the data in the system. Otherwise, users are not allowed to access the system [63], [64].

3.4. Machine Learning Approaches

The study of computer algorithms that learn on their own based on practice and a specific data set is known as ML [65]. Without the requirement for programming, ML algorithms use training data to build a model that produces predictions or judgments [66]. Numerous systems utilize ML including speech recognition, computer vision [67], email screening [68], and medicine [69]. ML systems fall into four main types based on the decision provided: -

3.4.1. Supervised Learning

This approach relies on supervision. This implies that the "labeled" data is used to make the algorithm learn from it, and then the algorithm will predict the output [70]. Mapping the input data to the output data is the main aim of the supervised ML [71]. Systems that use the supervised ML approach include fraud detection, risk assessment, spam filtering, and more. The main tasks of supervised ML are regression and classification. Classification algorithms are employed to manage classification issues that require categorical results, like "Yes" or "No", “Male” or “Female”, “cat” or “dog” etc. The groups of data that have been identified in the set of data are estimated by the classification algorithms. Applications that use this approach include email filtering [72] and spam detection [73]. The Random Forest, Decision Tree, Logistic Regression, and Support Vector Machine algorithms are a few well-known classification techniques [74], [75]. Regression is utilized to tackle issues where there is a linear connection between the input and output features. These algorithms are used in the prediction of continuous output variables, including market movements [76] and weather forecasts [77]. Several well-liked regression techniques include Decision Tree, Lasso, Multivariate, and Simple Linear regression [78].

3.4.2. Unsupervised Learning

This approach doesn't require supervision, as the name implies. It denotes that unsupervised ML uses an unlabeled dataset to train the system, which then makes output predictions on its own without human oversight [79]. Unsupervised learning involves training models using unlabeled and unclassified data, and letting the model make decisions on its own without any external oversight. Separating data into groups based on patterns, and differences is the main objective of this approach.

Two further categories of this approach are Clustering and Association. If we want to identify the innate groups within a dataset, we employ the clustering methodology. It is a method of clustering where similar items are put together into one group. Grouping customers according to their shopping behavior is an illustration of the clustering algorithm in action [80]. Principal component analysis, mean-shift method, K-Means clustering, and independent component analysis are a few of the well-liked clustering techniques.

Association algorithms can be employed to find unique correlations between elements in a large dataset [81] and this is the main goal of the association algorithms. This approach is used in many applications such as online usage analytics [82], market basket analysis [83], and others. “Eclat” and “FP-growth” algorithms are examples of “association rule” learning methods [84]. Famous applications of Unsupervised Learning; Singular Value Decomposition [85], Network Inspection [86], Suggestions Systems [87], and Identification of Anomalies [88].

3.4.3. Semi-Supervised Learning

This approach uses both labeled and unlabeled data during the training stage [89]. Semi-supervised learning employs mostly unlabeled data, but it may also work with some labeled data [90]. This

approach was suggested to bypass the issues of “supervised, unsupervised learning” [91]. The objective of this approach is to make efficient use of all data “labeled and un-labeled”. As opposed to the exclusive use of labeled data in supervised learning, data that share the same characteristics are grouped using an unsupervised learning approach, which subsequently helps to label the unlabeled ones and finally gains a labeled dataset. This is because obtaining labeled observations is comparatively more expensive than obtaining unlabeled ones [92].

3.4.4. Reinforcement Learning

Through a feedback-based approach, reinforcement learning enables an artificial intelligence (AI) “agent” to automatically investigate its environment by striking and trailing, acting, picking up lessons from past mistakes, and enhancing its performance. Maximizing rewards is the aim of a reinforcement learning agent, as they are rewarded for good actions and punished for negative ones. Unlike supervised learning, which uses labeled data, the reinforcement approach depends on the agent learning experiences [93]. Information theory [94], game theory [95], [96], operation research [97], and multi-agent systems [98] are examples of fields that use this type of ML. The “Markov Decision Process” (MDP) is employed to formalize an issue involving reinforcement learning. The agent in MDP engages with the surroundings and takes actions continuously; the surroundings react to every action and create a new state. There are two basic categories for reinforcement learning: Positive Reinforcement Learning, which focuses on adding elements to make the preferred pattern of behavior more likely to happen. The processes of positive and negative reinforcement learning are opposed. It increases the possibility that a preferred activity will happen again by avoiding the undesirable event [99].

3.5. Multi-factor Authentication Related Studies

Using a combination of the user's global positioning system (GPS), one-time password (OTP), and personal identification number (PIN), the study [100] suggested an MFA architecture to protect transactions via the Internet. By a predefined separation distance between the current transactional device and the user's smartphone, their framework seeks to accept or reject the transaction based on that distance. The transaction is accepted if the distance is less than or equal to the predefined distance. The transaction is rejected in any other case.

In [101], they put out another architecture for secure wireless payment systems; it makes use of an MFA mechanism depending on “short message service SMS”, “transaction identification code TIC”, and username-password combinations. TICs are distinct transaction identification numbers that clients of banks or other financial institutions get. This code offers more secure transaction authentication than OTP. TIC codes are stored as secret codes on mobile phones and are only used once before being encrypted and decrypted. With mobile phones, the user may initiate safe online transactions quickly by selecting a TIC from a stored list, eliminating the need to memorize and input a complex TIC code for each payment.

The study of [102] paired multi-layer authentication methods with MFA authentication based on risk assessment principles to offer layered MFA architecture. Five levels make up the established model, and every level has one or more aspects of authentication, such as possession, knowledge, or biometric-based criteria. To satisfy layering requirements, the model was enhanced by adding control information components in the last two levels.

To protect the MPESA mobile money application, the study [103] suggested an MFA architecture based on PIN, employing device-specific abbreviation for identification (ID), and voice

recognition. When a user requests access to an application to complete a transaction, the system utilizes the credentials they have stored in a database to authenticate their identity.

To provide an MFA schema, a password, OTP, and fingerprint were utilized in the research of [104], [105] to protect electronic payment systems. After entering the system using the password, the user is prompted to provide a fingerprint for verification when they reach the transfer page. The system sends an OTP after the transaction details are complete to ensure a successful completion of the operation.

An MFA framework that included biometric verification, OTP, and username-password authentication was utilized in [106], [107] to protect mobile banking apps. The user first logs in using his pre-registered credentials. Next, an OTP is given to his phone, and finally, fingerprint verification completes the transaction.

Another research with a similar design was put out by [108], who used a facial recognition schema, an OTP, and a PIN. First, the bank obtains personal data from the individual, including their phone number, PIN, and facial photo. Second, the user needs to provide his face photo and PIN to log in. Following authentication of the PIN and face photo, the user is shown a menu to choose a service, such as making a deposit, paying bills, or checking their balance. To complete the procedure, the system creates an OTP and transmits it to the user's mobile device for validation.

Adding additional levels of authentication is done by [109]. Four factors were deployed to secure the grid environment which are password, user ID, biometrics, and the user's current location which can be obtained through the GPS. The addition of the fourth component (user location) improves the security standards necessary for large distributed systems such as Banking Grid settings.

All mentioned studies in this section implemented an MFA approach without addressing the importance of whether the system is user-friendly or not, which affects the usability and attitude toward using such a system. The user's negative feelings towards MFA were mentioned and proved by many studies in the literature [110], [111], [112], [113]. Because of that, this study aims to implement a user-friendly MFA system.

3.6. Machine Learning Related Studies

The ML model used in this study is related to the classification of in-progress purchase transactions to avoid financial fraud. So, this section will discuss credit card fraud detection using ML.

Classifying credit card activities is usually a binary problem [114]. In this case, a credit card transaction can be classified as either legitimate (positive class) or fraudulent (negative class). The main characteristic that sets credit card transaction data apart is an uncommon occurrence. Often, the features of fraudulent and authorized transactions are similar. Fraudsters are always coming up with new ways to mimic the behavior of real credit card holders. As a result, the characteristics of honest and dishonest behavior are always evolving. This intrinsic characteristic causes a distribution that has a large bias towards the negative class (legal transactions) since fewer truly fraudulent incidents are discovered in credit card transaction data. For instance, the dataset that was analyzed by [115], [116], and [117] includes fraud cases in all transactions that account for 20%, 0.025%, and 0.172%, respectively.

There are several ways to sample the highly unbalanced data. The experimental results shown in [118] demonstrate that an algorithm with the greatest TP rate and the smallest FP rate is created by consciously splitting learning data in a fifty percent to fifty percent ratio between fraud and

legitimate data, using a random sample methodology similar to that used in [115]. Stratified sampling is used in the study of [117] to under-sample non-fraud transactions to obtain the best ratio to gain the best performance. An experiment was conducted by examining 50:50, 10:90, and 1:99 fraud variations in real cases. As a result, the 10:90 distribution performs better when it comes to performance comparisons with the 1:99 distribution.

To maintain important patterns in the data, stratified sampling was utilized in [119]. During this research, stratified sampling is combined with under- and oversampling of the positive cases. Classifying credit card data as legitimate or fraudulent is a common data mining classification problem that forms the basis of fraud detection [120], [121].

In the actual world, credit card fraud is typically detected in two different methods. First, the costly and unreliable manual fraud detection method of employing data mining. Second, is the application of rule-based or expert systems, which are employed to retain and alter fraud awareness to meaningfully understand the data and prevent fraud incidents. These expert systems can be divided into three groups.: supervised, unsupervised, and semi-supervised. In unsupervised fraud detection, transactions from outliers are recognized as possible cases of fraudulent activity. In contrast, supervised fraud detection techniques use data of legal and illegal transactions to classify new data as illegal or legal [116]. A semi-supervised technique, however, combines both supervised and unsupervised approaches.

Many algorithms were examined to resolve the current problem. For instance, quick research by [122] claims that their classifier can identify transactions that are fraudulent or not with an accuracy of 90% and 98.6%, respectively.

In the study by [123], two distinct types of random forest algorithms with different foundational classifiers are used to train the behavioral characteristics of legal and illegal payments. The best algorithm achieves 96.77% accuracy, precision, recall, and F1 score of 89.46%, 95.27%, and 96.1%, respectively.

The authors examine several methods for fraud detection in the study [124]. The findings reveal that the precision, recall, F1 score, and accuracy of the random forest classifier for class 1 were all 99.7%, meaning that 31%, 89%, and 46% of the transactions were fraudulent.

A different study [125] looks into several classification techniques for severely unbalanced datasets, including “naïve Bayes, logistic regression, random forest, and decision trees”. According to their results, the RF classifier performs the best, with 96.77% accuracy, 100% precision, 91.11% recall, and 95.43% F1 score.

A study was conducted by [126] who examines many supervised ML algorithms. The researchers assess their algorithm's performance using average precision and “area under the curve AUC”. According to findings, the best algorithms gain an average precision of 84.83% and an AUC of 91.48%.

The effectiveness of “decision trees, random forests, logistic regression, k-nearest neighbor, support vector machines, and random forests” is examined in another comparative study [127]. With an accuracy of 88%, the random forest algorithm was the most effective one after the five were put into practice.

The unbalanced dataset, in which most of the payment data are legal, is the primary issue of credit card fraud detection. Because of this problem, supervised learning systems can predict payments that are not fraudulent.

All mentioned studies in this section even the implemented algorithm results were not good nor they don’t evaluate their results using different metrics. Because of that, this study builds an ML model for financial fraud detection that deals with the fact of an unbalanced dataset. This study examined different ML algorithms and the better performer was chosen. To improve the model’s performance, many indicators such as accuracy, precision, recall, F1-score, and ROC curve were employed to assess each algorithm’s performance.

3.7. The Combination of MFA and ML in the Literature

One type of MFA that adjusts to the risk profile of the users is called risk-based authentication. To determine the user's degree of risk, the study [128] determines the authentication techniques that may affect user confidentiality by designing a risk engine that integrates with the system. This engine looks at the user's historical login logs and deploys machine-learning techniques to create an appropriate pattern and risk level for authentication factors for every user.

The study of [129] also utilizes risk-based authentication and MFA to establish safe and easy authentication methods. They developed two separate libraries, one for backend servers and one for Android applications. The server-side library of the study included an ML risk engine. The choice of authentication elements was informed by the risk levels that this ML engine determined using user-specific information including Internet Protocol (IP) addresses, device types, and access times.

The study by [130], for instance, suggested a 2FA system where the user first logs in with his username and password before employing neural networks to recognize faces.

In [131], they use two ML classifiers to examine user actions as an authentication framework. Following user login, they apply K-nearest Neighbor and random forest to evaluate player actions when playing a particular game with two fingers. They then use the information that they collect to verify the user's authenticity as a continuing authentication schema.

For the attendance system, another 2FA strategy that utilizes Radio Frequency Identification (RFID), IOT, and ML was carried out by [132]. An RFID tag, an RFID reader, a microcontroller, and a “global system for Mobile communication (GSM)” module was used in the first stage of authentication. A camera equipped with the “Multi-task Cascaded Convolutional Network (MTCNN)” model (using ML) was used for the second authentication. If both are acceptable, attendance will be awarded to the students. Parents will be informed about their child's attendance if it does not work.

3.8. Research Gap

The research gap is the shortage of knowledge on the potential of merging ML techniques with MFA to raise the safety of Internet financial transactions. The utilization of MFA schema without addressing the fact of users' negative feelings toward MFA systems is not the best way to secure financial transactions. To enhance security, ML techniques have been extensively employed in isolation, but integrating their applications with MFA has not received much attention.

Some studies talk about this possible combination, for example in [128], [129], ML was used for ranking the authentication factors denoting which one may be vulnerable. In [131], ML was used for continuous checking for user authenticity by evaluating user actions when using the system, so

the detection of fraud access will be after the user logs in to the system and this is not sufficient way too. In [130], [132] They used ML to enhance the face recognition quality of the users.

Because of the promised power of ML techniques in financial fraud detection and prevention [133], [134], [135], [136], this research uses an ML model as an embedded layer of security in the MFA framework to classify the current process into fraud or legitimate and asking for additional authentication factors in the scene of fraud. In this way, legitimate users will interact with a 2FA system without being annoyed by many factors to complete a purchase. Therefore, this research will bridge the gap in the literature by offering a framework that combines MFA with ML to gain the highest possible security along with a user-friendly system.

3.9. Summary

This chapter surveyed the different authentication factors types and approaches, offered an overview of the different ML approaches, and discussed the differences between them. Moreover, to address the research gap, ML, and MFA related studies were discussed.

The chapter's primary role was to identify and draw attention to significant research gaps in the body of current literature. By highlighting the shortcomings of the previous work in the field of financial transaction security, the author was inspired to present a novel framework that not only raises the security level but also provides a user-friendly system. This will guide future research endeavors in this dynamic field and a link between the theoretical and practical components of the possible combination of ML and MFA.

As the dissertation progresses, the proposed framework, approaches, and solutions meant to strengthen digital security while also improving user experience will be discussed in the next chapter.

Chapter 4

The Proposed Framework

4. The proposed Framework

4.1. Introduction

The previous chapter illustrated the authentication factors types and techniques, explored the ML algorithms categories, and stated the research gap by discussing the related studies in the literature. This chapter aims to: -

- Illustrates the general system architecture.
- Shows the study approach that was employed.
- Discuss the dataset, data preprocessing, and the justification for ML classifier choice.
- Presents the authentication factors that are utilized to build the MFA model.
- Illustrates the proposed MFA framework working principle.

4.2. System Architecture

This study suggests a framework for safe online transactions. This framework can be used with any e-commerce platform, such as a website where customers can make purchases using laptops or an application where customers can use tablets or mobile devices. Figure 4.1 displays the components of the system. These elements consist of:

- An e-commerce platform (website or application).
- Authentication factors for the user's identity verification.
- ML model to classify purchase processes.

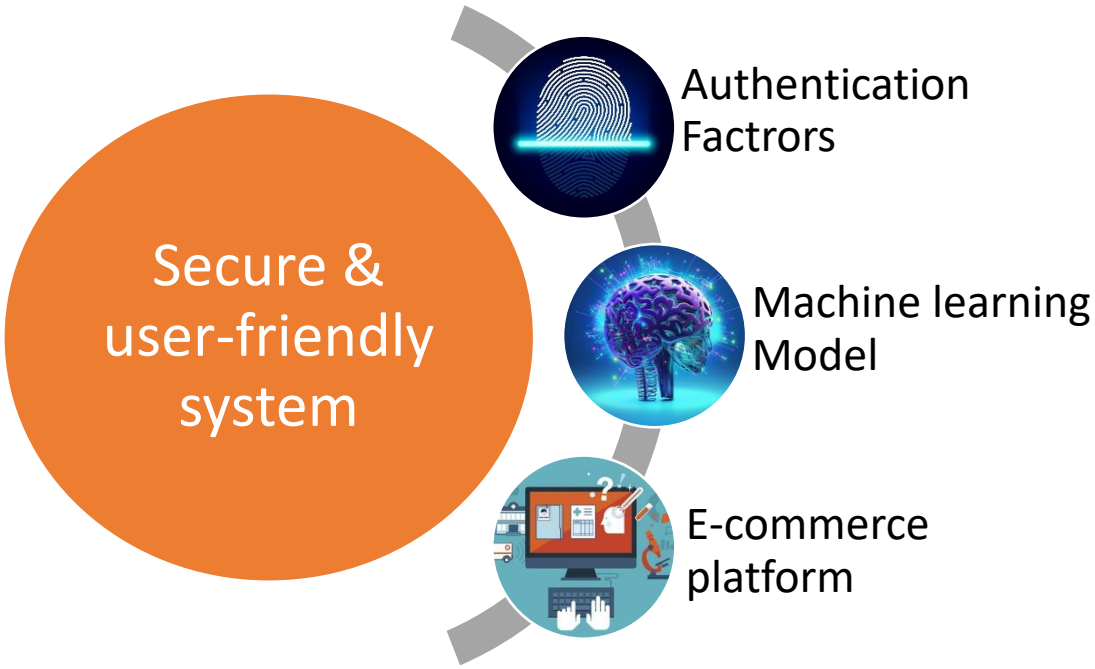

Figure 4.1. System Architecture

Figure 4.1 shows the components of our proposed framework, to gain a secure and user-friendly system. An E-commerce application or website must exist, choose the suitable authentication factors, train the ML model, and finally combine ML as an embedded part of the MFA system. Further illustration of the system workflow will be provided in the next sections.

4.3. Research Methodology

The study's methodology depends on integrating ML with MFA to build a system that is both highly secure and easy to use. Refer to Figure 4.2 for an overview of the research methodology.

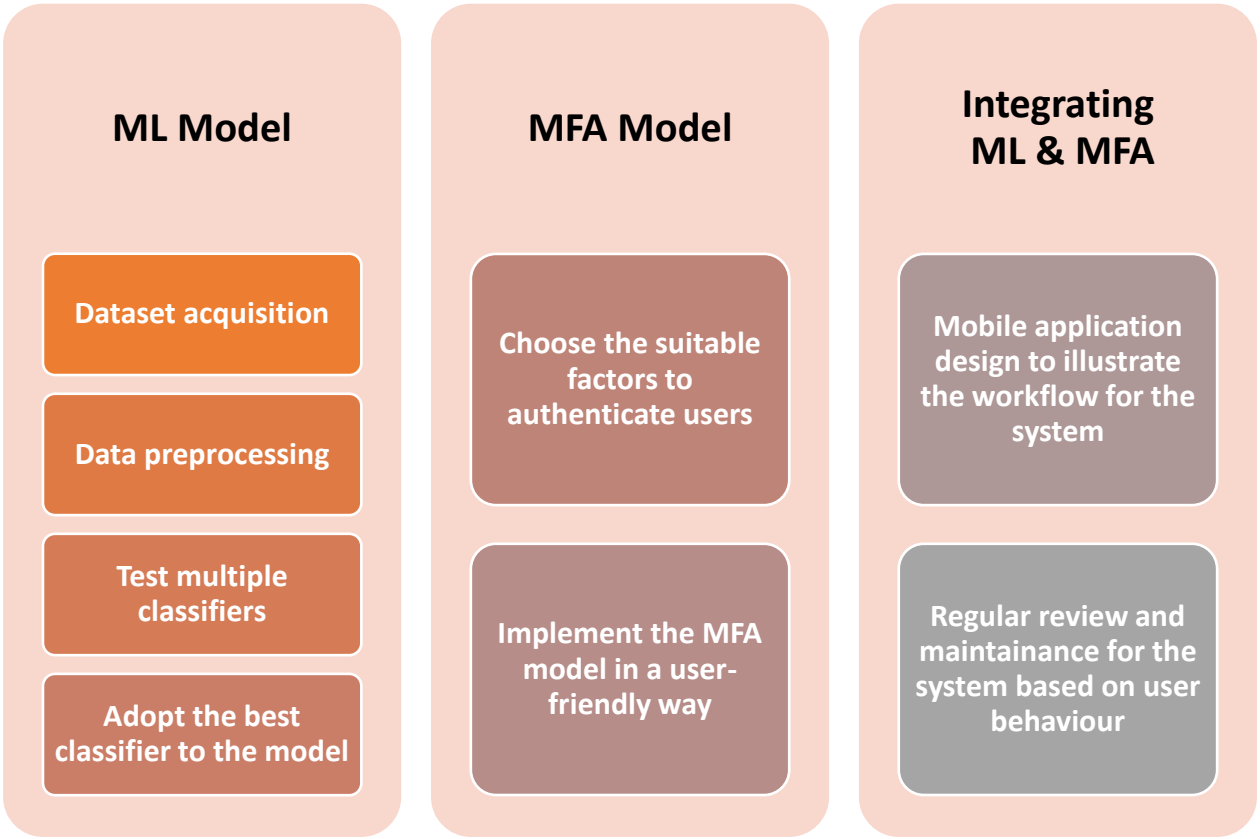

Figure 4.2. Methodology

The methodology is divided into three primary categories, as Figure 4.2 illustrates:

1. To begin with, the ML portion involves downloading an open-source credit card fraud dataset, cleaning the dataset, testing various classifiers, and finally adopting the best performer into the ML model.

2. In the MFA section, the appropriate elements for user authentication were selected and the model architecture was established to produce an MFA implementation that is easy for users to comply with.

3. application screens for e-commerce applications were designed to make it easy to understand the proposed framework.

4.4. Machine Learning Phase

This section will illustrate the journey to build the ML model. Figure 4.3 shows the steps in general terms. The dataset, data preprocessing, and a justification for the used ML algorithms will be discussed in different sections.

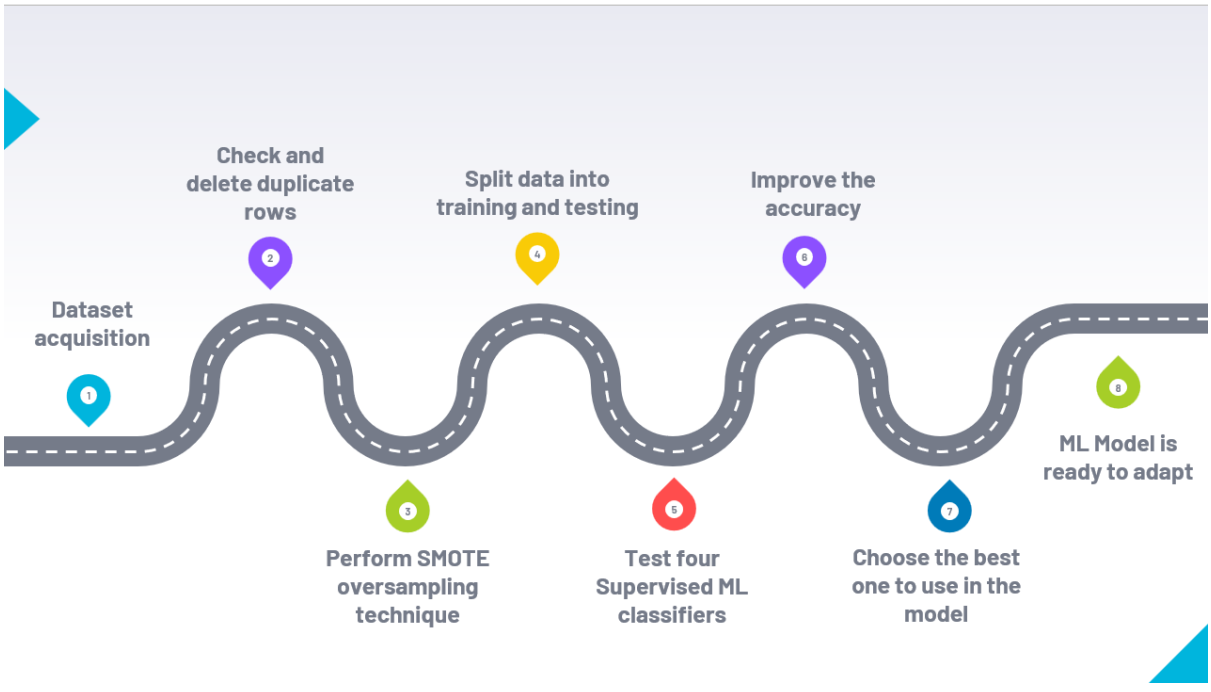

Figure 4.3. Roadmap for Machine Learning Model

The roadmap as seen in Figure 4.3 starts with the dataset acquisition, after performing the dataset cleaning, different classifiers were tested and the best one was chosen to build the model.

4.4.1. Dataset

The Kaggle website is where the dataset was located using the URL [https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud]. It includes September 2013 credit card purchases made by cardholders across Europe. The 31 characteristics in the dataset are "V1–V28, Time, Amount, and Class". The features from "V1- V28" are numeric values and have been treated with principal component analysis (PCA). The dataset owner clarified in the data card that they cannot provide any further explanation or metadata about these variables (V1-V28) because of the user's confidentiality issues.

Principal component analysis, or PCA, is an established approach to analyzing big datasets with many dimensions or characteristics per observation, improving data interpretability while retaining as much information as possible, and facilitating multidimensional data visualization. Technically speaking, PCA is a statistical method for reducing the dimensions of a dataset. To achieve this, the data are converted linearly into new arrangement systems where most of the data variation is explained by dimensions that are less than the original ones [137]. See Table 4.1 which shows a screenshot of the dataset.

Table 4.1 Dataset sample

Time	V1	V2	V3	V4	V5	V6	V7
0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599

0.0	1.191857	0.266151	0.16648	0.448154	0.060018	-0.082361	-0.078803
1.00	-1.358354	-1.340163	1.773209	0.37978	-0.503198	1.800499	0.791461
1.00	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609
2.00	-1.158233	0.877737	1.548718	0.403034	0.407193	0.095921	0.592941

V8	**V9**	**...**	**V26**	**V27**	**V28**	**Amount**	**Class**
0.098698	0.363787	...	**-0.189115**	0.133558	-0.021053	149.62	0
0.085102	-0.255425	...	**0.125895**	-0.008983	0.014724	2.69	0
0.247676	-1.514654	...	**-0.139097**	-0.055353	-0.059752	378.66	1
0.377436	-1.387024	...	**-0.221929**	0.062723	0.061458	123.5	0
-0.270533	0.817739	...	**0.502292**	0.219422	0.215153	69.99	0

Table 4.1 displays a sampling of the dataset. The features that have not been modified by PCA are the amount and time features, as seen in the table. The seconds that elapsed between transactions are represented by the “Time” feature. The feature “Amount” is a representation of the transaction amount. The feature "Class," is set to 1 in fraud situations and 0 in legal transactions.

4.4.2. Data Preprocessing

There are 285,299 transactions in the dataset. With 492 frauds out of 284,807 transactions, or 0.172% of all transactions in the positive class, the sample is extremely skewed. Since the imbalanced dataset may cause no preferred actions, it is inappropriate to analyze it directly in the model. The classifier will be skewed only to identify the negative (legitimate) class, while positive samples (the fraudulent class) are very likely to be false [138].

A preprocessing method called the “Synthetic Minority Over-sampling technique (SMOTE)”, is used to rectify a class imbalance in a dataset. In the real world, when training a model on a dataset that contains few samples of a certain class (e.g., manufacturing failures, fraudulent transactions, diagnosis of rare diseases), it leads to insignificant performance [139]. It is not always practical to collect more data because of this problem. The majority class should be under-sampled as one method of resolving this problem. But in the process, we will lose a lot of data that would have helped us train the model [140]. Another solution is to oversample the minority class. Stated differently, we replicate observations of the minority class. The criticism of the oversampling method is the overfitting problem since the model keeps learning from the same examples [141]. SMOTE is useful in this situation.

The SMOTE algorithm can be summarized as follows at a high level [142]: -

- Determine how a sample differs from its closest neighbor.
- Take an arbitrary value between 0 and 1 and multiply the difference by it.
- Include this variation in the sample to generate a new simulated example in the space of features.
- Proceed to the next closest neighbor until the user-specified number is reached.

“SMOTE” is the method most commonly used in the literature [143], [144], [145], [146], [147], [148], [149], [150]. Before performing the oversampling technique, a total of 1081 duplicate rows were found and deleted from the dataset. After cleaning the dataset, the SMOTE oversampling technique was implemented to get a balanced dataset. Figure 4.4 displays the dataset's distribution before and after oversampling. A further illustration of SMOTE and PCA techniques is provided in the Appendix part.

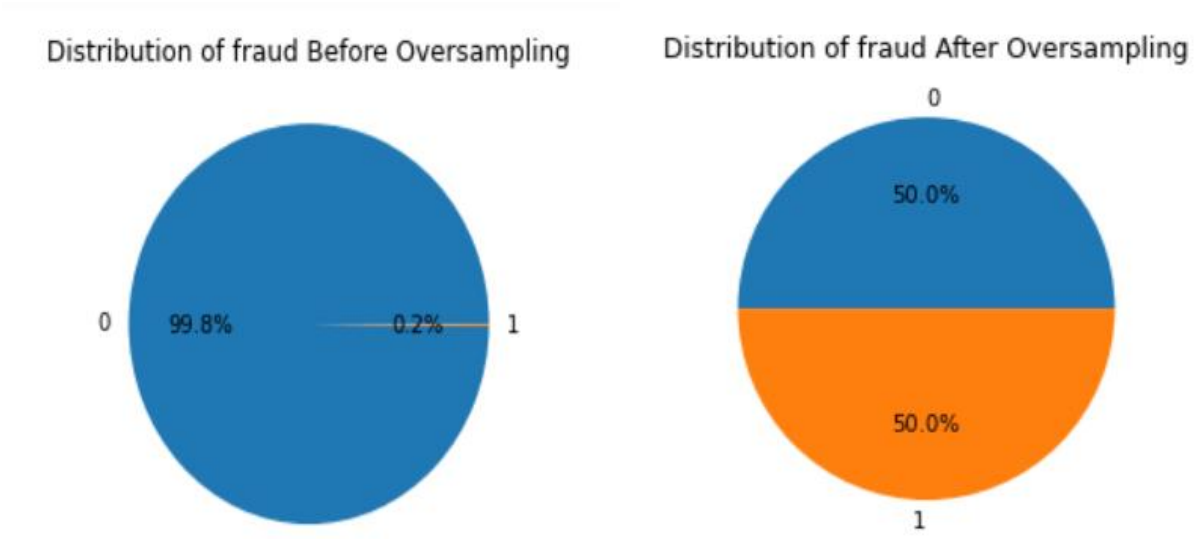

Figure 4.4. Dataset Distribution (not-fraud is '0', fraud is '1')

Figure 4.4 makes it obvious that 283253 transactions with a 50:50 distribution were counted for each class following the SMOTE oversampling approach. Ultimately, the dataset was separated into training and testing subsets (20% of transactions with a count of 113302 and 80% of transactions with a count of 453204). Finally, the dataset becomes ready for testing various supervised ML classifiers.

4.4.3. The Choice of Machine Learning Classifiers

A range of "ML" techniques are essential for locating fraudulent payments in the field of credit card fraud detection. The frequently employed techniques are: -

1. **Random Forest (RF)**

RF is an ensemble learning technique that blends several decision trees. A random subset of the features and data are chosen to create each tree. A new transaction is input, it is routed through

each tree, and the individual trees' majority vote determines the outcome. RF is resistant to overfitting and efficient in identifying intricate patterns in transaction data [133].

2. **Decision Trees (DT)**

Decision trees sequentially create a tree-like structure of decisions by creating subgroups of the dataset depending on attributes. The algorithm chooses the feature at each node that best divides the instances into separated classes. The transaction is classified as legal or fraudulent based on how it moves through the tree [143].

3. **Logistics Regression (LR)**

A statistical approach called logistic regression calculates the probability that a payment is fraudulent. A logistic function, shaped like an S, is employed to associate input features with a probability score. The transaction is categorized as fraudulent if the probability score is higher than a certain point; otherwise, it is categorized as legal [144].

4. **Naïve Bayes (NB)**

Based on the Bayes theorem, It is assumed that characteristics, given the class, are conditionally independent. It determines how likely it is that a transaction's characteristics will appear in both authentic and fraudulent transactions. The transaction is assigned to the class with the highest chance [145].

5. **Support Vector Machines (SVM)**

This algorithm aims to locate the hyperplane that divides transactions into the most distinct classes. To optimize the margin between the classes, the hyperplane is selected. After that, transactions are categorized according to which side of the hyperplane they are on [146].

6. **Neural Networks (Deep Learning)**

Layers of networked nodes, or neurons, process transaction data in neural networks. Through the process of learning hierarchical representations from the data, deep learning models; such as CNNs and deep neural networks can identify intricate patterns[147].

7. **K-Nearest Neighbors (KNN)**

By calculating the degree of similarity between a transaction and its k nearest neighbors in feature space this algorithm categorizes transactions. The transaction is categorized based on the vast majority class [148].

8. **XGBoost and Gradient Boosting Machines (GBM)**

These ensemble techniques enhance classification accuracy by combining several decision trees. They create trees iteratively, concentrating on fixing mistakes from the earlier trees to produce a very accurate model [149].

9. **K-Means Clustering**

K-Means uses feature similarity to divide transactions into clusters. Transactions that do not belong to any cluster or that are part of a small, separate cluster can be used to identify outliers [150].

10. **Isolation Forest**

Isolation Forest effectively isolates anomalies by using random partitioning. Because anomalies require fewer divisions to be isolated from the bulk of transactions, they can be isolated in fewer steps [151].

The mentioned algorithms, each with distinct operating principles, provide a variety of methods for spotting and stopping credit card fraud by examining transaction data and looking for odd

trends or anomalies. Several criteria, including the size and complexity of the dataset and the trade-off between computing efficiency and model truthfulness, influence the choice of algorithm [152]. To determine the best strategy for a particular credit card fraud detection issue, researchers frequently test out various algorithms and ensemble approaches [153], [154], [155], [156].

Tackling the problem of fraud detection, which often involves binary classification (0 for transactions that are not fraudulent and 1 for fraudulent transactions) four supervised "ML" methods were selected which are: Naïve Bayes (NB), Logistic Regression (LR), Decision Trees (DT), and Random Forest (RF).

Interestingly, Random Forest (RF) was chosen due to its ensemble-based approach, which is a favored technique for detecting fraud because of its capability to manage complex, high-dimensional data without overfitting [133]. Because of their interpretability, Decision Trees (DT) help understand the mental processes that result in fraudulent behavior [143]. Because logistic regression (LR) provides modeling simplicity and efficiency and completely aligns with binary classification jobs, its incorporation was made possible [144]. Naïve Bayes (NB) has proven effective in managing categorical data, which is commonly encountered in fraud detection scenarios, despite its simplicity [145].

This study investigates many classifiers to determine which model works best for fraud detection. To successfully identify legitimate fraudulent transactions, MFA systems in the real world must strike a balance between security and usability, which is why finding the classifier with the best accuracy was the goal. These classifiers are also possibly more straightforward and practical to integrate with an MFA system.

4.5. Multi-Factor Authentication Phase

In many systems, authentication factors are essential for confirming users' identities. Each component has advantages and disadvantages that may affect which use cases they are appropriate for. An outline of the benefits and drawbacks of popular authentication factors is provided below [157], [158], [159].

4.5.1. Knowledge-Based Authentication (Something You Know)

Advantages:

- Relatively Simple to Reset: When a password is lost, it can be reset or altered.
- Low Cost: Implementing it usually does not cost much
- Widespread: The majority of users are accustomed to and frequently utilize knowledge-based authentication.

Drawbacks:

- Phishing: Phishing attacks can fool users into disclosing their passwords
- Password Guessing Vulnerability: Brute-force attacks and password guessing can target weak or widely used passwords.
- Reusing Passwords: People are used to using identical passwords across various accounts, this will make it simpler to be hacked [160], [161], [162].

4.5.2. Possession-Based Authentication (Something You Have)

Advantages:

- Phishing-resistant: Possession-based factors are more resilient to phishing attempts than knowledge-based factors.
- Enhanced Security: An extra layer of resistance is offered by possession-based considerations, such as a tangible token or a mobile device.
- Two-Factor Authentication (2FA): To improve safety, it is frequently used in conjunction with two-factor authentication.

Drawbacks:

- Loss Risk: Physical tokens are susceptible to theft or loss, which could jeopardize security.
- Additional Cost: Putting real tokens into use and distributing them can be expensive.
- Inconvenience: Carrying and using actual tokens may be inconvenient for users [163], [164].

4.5.3. Biometric Authentication (Something You Are)

Advantages:

- Sharing: It is challenging to copy or distribute biometric data.
- Convenience: There is no need for users to carry physical tokens or memorize passwords.
- High Security: Because biometrics are personal to each person, they provide a high level of security.

Drawbacks:

- Complexity and Cost: Putting biometric authentication systems into place can be costly, and they could need specific technology.

- Privacy Issues: Privacy concerns arise from the use and storage of biometric data.
- False Positives and Negatives: Biometric systems can result in false positives, which authenticate the incorrect individual, or false negatives, which prevent the authorized user from accessing the system [165], [166].

4.5.4. Behavior-Based Authentication (Something You Do)

Advantages:

- Hard to duplicate: It is challenging to duplicate user activity, such as typing patterns.
- Ongoing Authentication: Behavior-based authentication can watch users continuously and adjust to their actions.

Drawbacks:

- Privacy Issues: Privacy concerns may arise from the ongoing observation of user activity.
- False Positives: If a user's behavior drastically changes, behavior-based systems may generate false positives [167].

The particular security needs and system usability elements must be considered when selecting authentication factors and how to combine them. In the real world, a lot of systems employ MFA to minimize the drawbacks of each element while utilizing its advantages.

During the MFA stage of this study, the selection of authentication factors to strengthen the safety of Internet financial payments was carefully considered. Username-password, fingerprint, one-time password (OTP), and face recognition were used to create a robust security architecture.

These standards were selected due to their unique benefits and ability to provide a multi-tiered security solution. Username-password combinations are mostly used in online accounts and provide a basic level of security. A further degree of dynamic security is additionally provided by using OTPs, which guarantee that a time-sensitive code is needed to access the system [168]. By verifying the user's identity using particular facial features, face recognition, the third authentication element, uses biometric technology to increase security and reinforce the system's resistance to unauthorized access [169].

This new approach not only increases system security but also facilitates a more user-friendly MFA implementation. By making use of modern biometric authentication technology, the study attempts to find an acceptable balance between enhanced security and user comfort.

4.6. Combining Multi-Factor Authentication with Machine Learning

After building the ML model and determining the authentication factors to use in the final MFA model, both models will be combined in the final MFA framework. This section will discuss the workflow of the proposed MFA framework and the application screens that have been designed to make the idea easy to understand.

4.6.1. The Proposed Framework

This section shall talk about the system's operation. To authenticate users, the suggested system uses four factors: username and password, one-time password, ML classification, and face recognition. Refer to Figure 4.5 which shows the functionality of the proposed system.

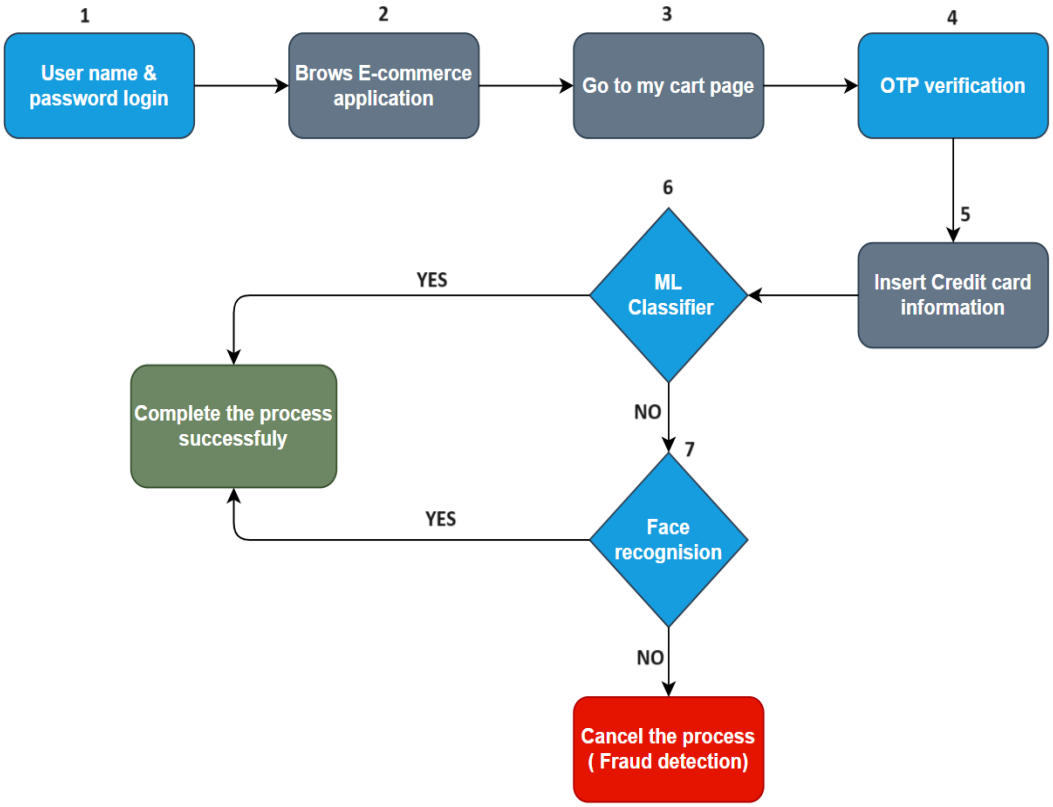

Figure 4.5. Workflow of the Proposed System

As seen in Figure 4.5, the user will first login to the application using the username and password. After exploring the various products and selecting what to buy, the user will proceed to the cart page to complete the transaction. The user must then complete the OTP verification before being forwarded to the credit card information page. Currently, the ML model will assess and categorize this payment as either legal or illegitimate. To successfully finish the purchasing process, the user would be prompted for facial recognition identification if the classification has been determined to be fraudulent; if not, the procedure would be terminated.

In this way, the user will interact with two factors to verify the identity as a first layer of security to complete his transaction. The ML model will work as an embedded layer of security without bothering the user and evaluate the current transaction, asking the user for face recognition if the classification is fraud and asking for no more authentication factors if the classification is legitimate. In the end, by utilizing “ML” capabilities and integrating them with the “MFA”, the proposed model will offer both a high level of security and a user-friendly system.

4.6.2. Application Screens Design

Screens for Android e-commerce applications were created to make the concept simple to understand. The application interfaces' layout made it possible to properly deploy the ML and MFA solutions. To ensure user-friendliness and ease of use, a user-centered approach was adopted during the design phase. Creating a user interface that is both visually beautiful and flexible, with a logical and clear screen flow during the purchasing process, is part of the design approach.

An HP laptop with an “Intel(R) Core (TM) i5-10210U CPU @ 1.60GHz 2.11GHz” served as the development environment for this investigation. The processing power needed for the phases of application design and development was supplied by this hardware setup. The "Adalo" website was the key piece of software used to customize the application displays.

Adalo is an empowering no-code platform that lets people and companies create web and mobile applications without requiring a deep understanding of coding. Adalo makes it simple for users to visually build and customize app components with its drag-and-drop functionality and user-friendly interface, similar to putting together a puzzle. The main features of Adalo are [170], [171], [172]:

- **No-Code Creation**

Coding complexity is less important while using Adalo. Without having to learn a lot of programming, users browse a visual environment, choosing and organizing elements. This method democratizes the process of creating apps, enabling a large number of users to realize their ideas [178].

- **Design Tools**

A toolbox for creating app screens and user interfaces is provided by the platform. Users can design and personalize every element, including buttons, input fields, and navigation parts, to create a unique and visually appealing user experience [178].

- **API Integrations**

Through API integrations, Adalo provides access to external features and services. This increases the app's potential by making it possible to seamlessly integrate identity confirmation services, MFA systems, or even ML capabilities, which improves the app's security and user experience [179].

- **Data handling**

Its ability to import and export data makes working with outside sources easier. To ensure an efficient and effective system, this feature is essential for integrating ML models, MFA data, and other crucial components into the app [179].

- **Cooperation and compatibility**

The platform promotes cooperation, enabling smooth connections with various services, design tools, and platforms. By allowing users to work together with outside tools or integrate other

services, this collaborative capability improves the design process and improves the functionality and appearance of the app [180].

Adalo's methodology is centered on facilitating the method of developing applications by offering a link between the accessibility of user-friendly design tools and the complex nature of coding procedures. Without requiring a deep understanding of coding, a broad range of users may construct visually beautiful, technically sound, and useful apps thanks to its feature-rich set and collaborative possibilities [180].

A clear visual depiction of the safety measures needed to successfully conduct a purchase through the application can be found in Figure 4.6, Figure 4.7, and Figure 4.8. This graphic aid makes it easier to comprehend the security procedures that are part of the procurement process.

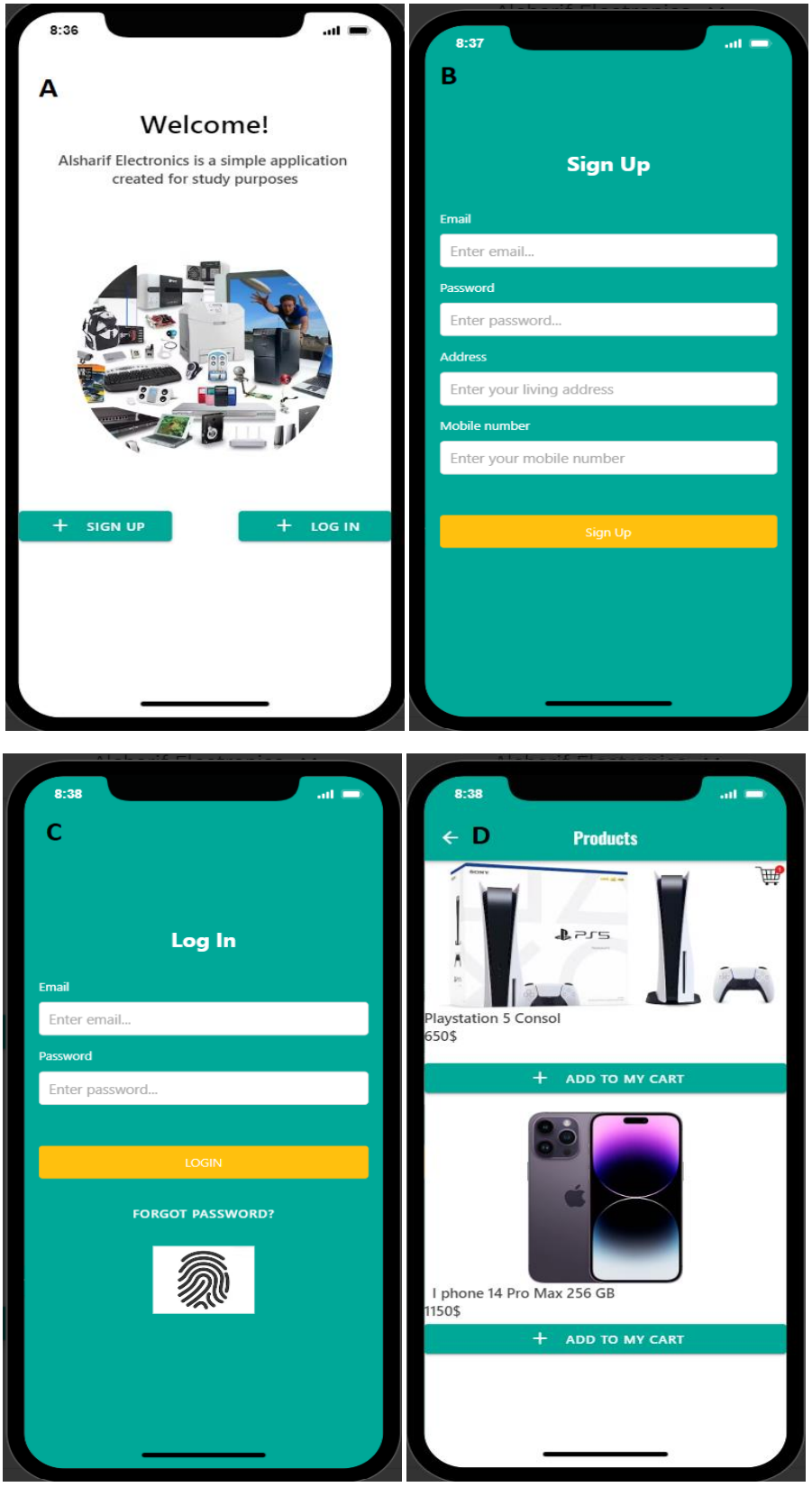

Figure 4.6. Application Screens (A-D)

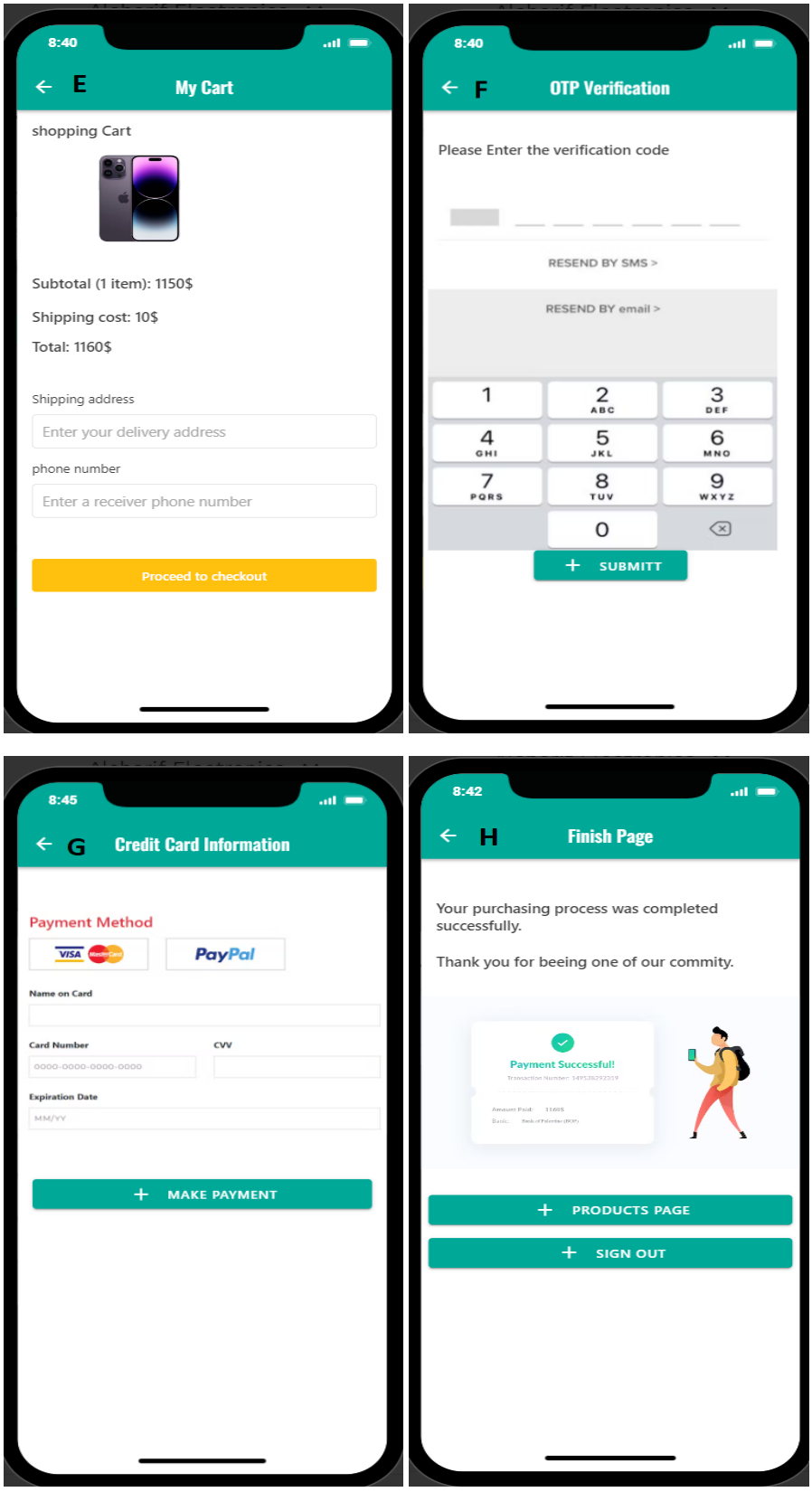

Figure 4.7. Application Screens (E-H)

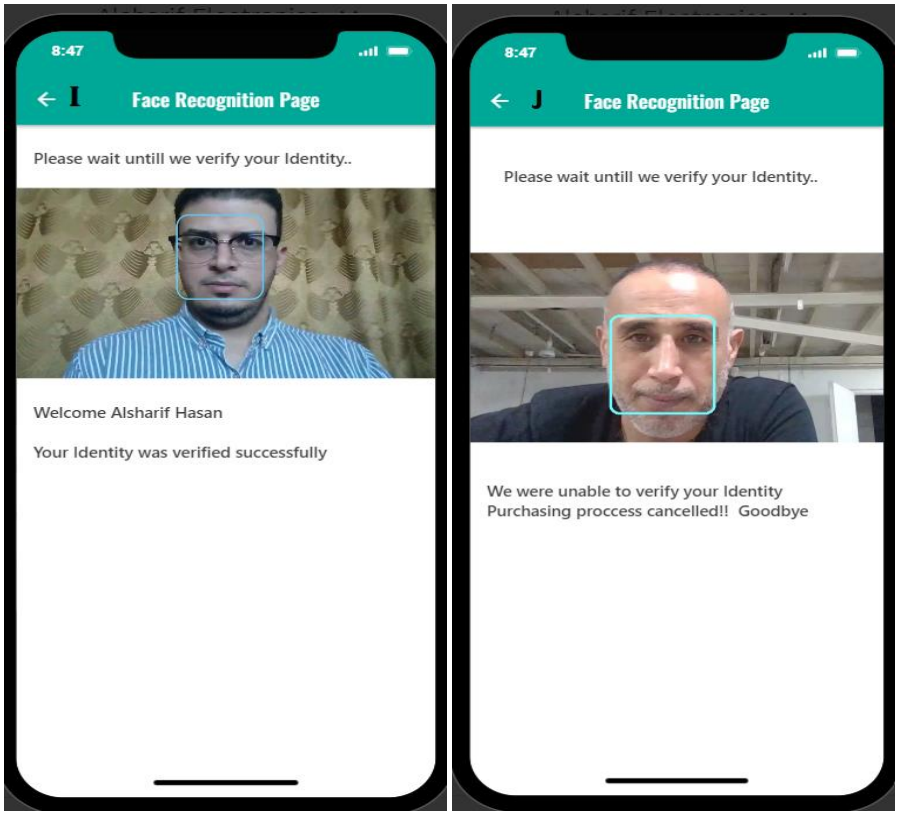

Figure 4.8. Application Screens (I-J)

Figure 4.6, (A) shows the application welcome page, the user must select between going to the sign-up or login screen. The user will create an account by entering some details in the sign-up screen Figure 4.6, (B), including their address, mobile number, password, and email address.

Upon creating an account, the user has two options for logging in Figure 4.6, (C): either use their email address (user name) and password to access the application, or use their fingerprint to access it. This is the first authentication factor. The products available in the e-commerce application are displayed in Figure 4.6, (D) where users can look over and add any item to their cart.

Once browsing is complete, the user will go to the My Cart page Figure 4.7, (E) to edit the products he has selected and the total amount of the transaction. To get in touch with the delivery firm, the user needs to input the shipping address and mobile number. One Time Password (OTP) will be delivered to the user's mobile device after completing this step and selecting Proceed to Check Out. The user must input the sent number into the screen shown in Figure 4.7, (F) (second authentication factor). After successful OTP verification, the user will be taken to the credit card details page Figure 4.7, (G).

While the user fills in the credit card information, the ML model will evaluate the purchasing process. Based on the training and model covered in Chapter 3, the ML model will determine whether the purchase process is legitimate or fraudulent. If it is determined to be legitimate the purchase process will be finished successfully as shown in Figure 4.7, (H).

If the ML model classifies the current process as fraud, the user will be prompted to provide their face recognition, which is the third authentication factor. If face recognition is successful Figure 4.8, (I), the purchase process will be finished successfully Figure 4.8, (H), if the user fails in face recognition **Error! Reference source not found.**,(J) the transaction is considered fraudulent and the purchasing process is canceled.

4.7. Explainable ML Techniques for Fraud Detection Feedback

The term "explainable ML" is defined as the ability of ML models to offer easy-to-comprehend justifications for their interpretations or predictions. Explainable ML algorithms provide many advantages to financial organizations, clients, and regulatory agencies when it comes to detecting fraud in online financial transactions.

Explainable ML is an effective technique for informing parties involved in online financial transactions about fraud. This approach provides stakeholders with more clarity, accountability, and transparency, allowing them to make better decisions and successfully stop fraudulent activity. Consider a bank that employs ML techniques to identify fraudulent transactions within its online banking platform. The organization wants to provide information about how the ML model makes decisions and which features are most important in spotting fraudulent behavior with its fraud detection team and other relevant stakeholders. Some of the explainable ML techniques are: -

4.7.1. Feature Importance Analysis

- The ML model can shed light on the characteristics or factors that matter most in detecting fraud. As important indicators of possible fraud, the model might note unusual transaction amounts, transaction frequency, geographic regions, or times of day.
- Based on these crucial characteristics, the fraud detection team can utilize this data to prioritize their investigations and concentrate on transactions that have the highest chance of being fraudulent.

4.7.2. Local Interpretable Model-Agnostic Explanations (LIME)

- The LIME approach offers local interpretations for each prediction the ML model makes. For example, LIME can produce an explanation indicating which indicators were most important in predicting a given transaction that the model flagged as possibly fraudulent.
- Stakeholders can confirm the predictions made by the model and take suitable action, like accepting or rejecting the transaction, by looking over these clarifications to understand the reasoning behind the model's choice for each transaction.

4.7.3. Shapley Guidelines

- Shapely values consider all potential feature combinations to assign each feature's contribution to the model's output. Shapley values, for instance, can be used to measure how various features (like transaction amount, account balance, or transaction history) affect the model's conclusion when a transaction is reported as fraudulent.
- Stakeholders can increase the accuracy of the ML model over time by refining their fraud detection techniques and gaining a thorough grasp of how different aspects influence the chance of fraud by using Shapley values.

4.7.4. Visualization tools

- User-friendly displays of the ML model's decision-making process can be achieved through interactive visualization tools. Stakeholders, for example, can easily see patterns and anomalies by examining a dashboard that displays the feature ratio between fraudulent and non-fraudulent transactions.
- With the use of these visualization instruments, stakeholders may make responsible choices about risk management and fraud prevention tactics by gaining practical insights from the ML model's predictions.

Through the use of explainable ML techniques like feature importance analysis, LIME, Shapley values, and visualization tools, the financial institution can provide stakeholders with comprehensive feedback regarding fraud detection judgments. As a result, stakeholders can confirm the correctness of the model, understand the reasons driving its predictions, and take proactive steps to successfully avoid and minimize fraudulent activity.

4.8. Summary

This chapter aimed to dig deeply into the creative solution to enhance the security of online financial transactions. The system architecture, which gave a summary of the components supporting the framework was revealed at the beginning of the chapter. The study process was examined in detail, and the methodical strategy used to look into and create the framework was clarified. After that, the chapter broke down the ML stage, covering important elements including dataset selection, data preprocessing, and the reasoning behind the machine learning classifier selection.

Subsequently, the phase of MFA was implemented, showing the comprehensive strategy for user verification. The study came to a critical point when investigating the intersection of MFA and ML. The chapter broke down the proposed system's workflow, outlining the steps involved in user interactions and security checks.

It also looked at the application design, which serves as the foundation for an intuitive user interface and guarantees that the framework blends in smoothly with the user's experience. The mobile application design demonstrated the usefulness of combining MFA processes with ML. It was carefully built through the combination of hardware (HP laptop) and the adaptable "Adalo" software. The application panels showed an easy-to-use interface that led users through a safe transaction procedure, from choosing a product and creating an account to implementing multi-layered authentication during transactions, which included facial recognition and OTP verification.

Additionally, this chapter discussed the importance of using explainable ML techniques which helps humans to understand the ML decisions and try to adapt the model's inputs to enhance results.

Finally, this chapter formed the foundation of the research and stated the groundwork for the in-depth examination. Findings will come in the next chapter.

Chapter 5

Results and Discussions

5. Results and Discussions

5.1. Introduction

The previous chapter showed the system components, methodology used to secure financial transactions, discussed the ML model-building steps, determined the factors to authenticate users, and finally illustrated how the MFA framework will communicate with the users in a friendly and safe manner.

This chapter will show: -

- ML model results.
- Experimental environment.
- Metrics and measures used to improve the efficiency of the applied ML model.

5.2. Machine Learning Results

This study tested many classifiers on a well-selected dataset that was covered in the dataset section, such as “Naïve Bayes (NB), Logistic Regression (LR), Random Forest (RF), and Decision Trees (DT)”. The findings show how well these classifiers perform in comparison when it comes to detecting fraudulent transactions. Each model's efficacy was assessed using assessment measures such as ROC-AUC, F1-score, accuracy, precision, and recall.

5.2.1. Experiment

The same device that was used to design the application screens (HP laptop) was utilized to maintain the results. The experiment's code was written in Python, and an Anaconda Jupyter Notebook V3 was used to conduct it.

The experiment aimed to construct a trustworthy detection model with precise classification and identification capabilities for fraudulent transactions. Data for testing and training were taken out of the dataset. To overcome the imbalanced dataset issue and avoid bias when implementing multiple classifiers, the SMOTE oversampling technique was employed. Finally, to find the optimal settings, a grid search, and standard scaler were performed to achieve the best precision feasible.

- **Anaconda**

Comparable to a scientist's toolbox, Anaconda comes with many pre-installed libraries, packages, and tools that are essential to our ML. We can coordinate different libraries using Anaconda such that they function as a unit. We can concentrate on the main findings of our research rather than tinkering with software setups because of this capacity to manage dependencies and package versions. Anaconda is a powerful tool that offers libraries for statistical analysis, ML, data processing, and visualization. Our research environment is complete as it guarantees that we have all the resources we require at our disposal. In addition, the “Conda package manager” provided by Anaconda makes library installation simple and flexible, allowing us to keep up with the rapidly changing field of data science. Our study can be replicated and our code is protected from the unexpected effects of library updates [173].

- **Jupyter**

In terms of data exploration and code creation, Jupyter is a vital component of our research technique. We can experiment and iterate with ease thanks to Jupiter's collaborative and interactive environment. Jupyter Notebook serves as a canvas that unites code, information, justifications, and graphics. Our research workflow is improved by Jupiter's interactive features, which allow us to run code cells, view findings, and make quick changes to the code.

Another distinguishing feature of Jupyter is collaboration. Sharing Jupyter Notebooks allows users to share their research with collaborators and coworkers, making it a shared resource. This collaborative tool facilitates the exchange of ideas and best practices, peer review, and information sharing. Because Jupyter is dynamic and interactive, it guarantees that this study is a collaborative effort in which insights from all viewpoints enhance the final product [174].

- **Python**

The Python programming language is the foundation of our investigation. We made a deliberate decision to use Python as the principal language for our ML model. Python is the perfect language for our research needs because of its reputation for elegance, readability, and versatility. The simple syntax of the language makes documentation, readability, and code maintenance easier. Another strength of Python is its adaptability. The language has a large ecosystem of ML and data science-specific libraries and packages. Python gives us the tools we need to create our model easily, with libraries like "NumPy" for mathematical procedures, "Pandas" for modification of data, "Scikit-Learn" for ML, and "Matplotlib and Seaborn" for visualization.

Additionally, our code can run on several operating systems because of Python's cross-platform compatibility, which enhances the usability and scalability of our research. The language is a great option for research and development because of its vibrant and sizable community, which offers a multitude of resources and assistance [175].

To offer transparency and reproducibility, the ML model used in this study has its source code available on "GitHub" for the general public to view. Readers and researchers who are interested in looking into the specific implementation can access the source code repository at https://github.com/Alsharif-hasan/Credit-Card-Fraud-Detectio-Model/tree/main.

- **Essential Libraries Used in the Research**

The collection of libraries that have been painstakingly incorporated to solve certain problems is a crucial component of the study process.

1. Matplotlib and Seaborn: With the help of data visualization packages like Matplotlib and Seaborn, researchers can visualize data and make intricate patterns easier to understand. Pie charts that show the distribution of fraud incidents before and after oversampling can be created more easily with Matplotlib. Seaborn enhances the visuals' readability and beauty by working in tandem with Matplotlib. These libraries are really helpful in understandably providing the study results [176].

2. SMOTE and Imblearn: Handling class imbalances in the dataset is a common ML difficulty. The 'imblearn' library offers SMOTE as a solution. SMOTE is an essential method to enhance the accuracy of our classifiers since it rebalances the data. Making certain that our model is trained on a 50:50 dataset, this oversampling strategy helps to avoid biases [177].

3. Pandas: Pandas are an essential tool for data manipulation that helps with loading, cleaning, and analyzing datasets. By efficiently managing duplicate rows, extracting essential features, and getting the data ready for modeling, it supports data preparation. Pandas make data preparation more efficient and less prone to errors by streamlining the process [177].

4. Scikit-Learn: It provides a large array of tools for reducing dimensionality, and selecting models for clustering, regression, and classification. Scikit-Learn is essential to the implementation of ML model code, including RF, NB, DT, and LR. It makes the ML model creation, training, and evaluation simple with its intuitive interfaces, which makes it an essential component of our research [178].

5. Standard Scaler: The feature data is normalized or standardized using "StandardScaler", an element of the "Scikit-Learn" preprocessing package. The features of zero mean and one standard deviation are guaranteed by standardization. For many ML methods, this preprocessing stage is crucial in ensuring that every attribute contributes similarly to the model's performance. "StandardScaler" is utilized in this research code to enhance model accuracy by scaling the input data before feeding it into ML classifiers [179].

6. Confusion Matrix and Classification Report: These are crucial components of research implementation. An evaluation of the effectiveness of ML classifiers was performed with the help of these tools, which are a component of the "Scikit-Learn" metrics module. TP, TN, FP, and FN predictions are broken down in confusion matrices, and complete statistics including precision,

recall, F1-score, and support are provided in classification reports for each class. These resources are essential for assessing the accuracy and instance classification accuracy of the model [180].

7. Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC): To figure out the efficacy of the implemented ML model, ROC curves and AUC values are computed and visualized in the code. ROC curves and AUC are important metrics for evaluating binary classification models, even though they are not stand-alone libraries. While the AUC measures the classifier's efficiency to distinguish between categories, the ROC curve illustrates the relationship between the TP rate and the FP rate. These metrics are used to investigate the ML model efficiency [180].

8. ROC Curve and AUC Plotting: To measure the model's performance in classification, the code generates ROC curves, and AUC values are used. This demonstrates how to depict these important metrics using tools like Matplotlib, even though it isn't specifically credited to any one library. Researchers can effectively communicate the study findings by using Matplotlib to create graphical representations of model performance that are easy to understand and instructive [181].

In conclusion, combining these tools and libraries creates a rich ecosystem that makes it possible to apply, assess, and visualize ML models. Their crucial role in evaluating the safety of online financial transactions allows them to conduct comprehensive and significant research.

5.2.2. Metrics and Results

Because the dataset is skewed, the model's dependability cannot be demonstrated by testing it and only proving its accuracy. The confusion matrix, precision, recall, F1 score, and ROC curve examinations were implemented to assess the results.

- **Confusion Matrix**

Confusion matrix is a measure that shows information about correctly and incorrectly identified classes. An output matrix (2 by 2) measuring that displays the values of TP, TN, FP, and FN is the confusion matrix [182]. See Figure 5.1.

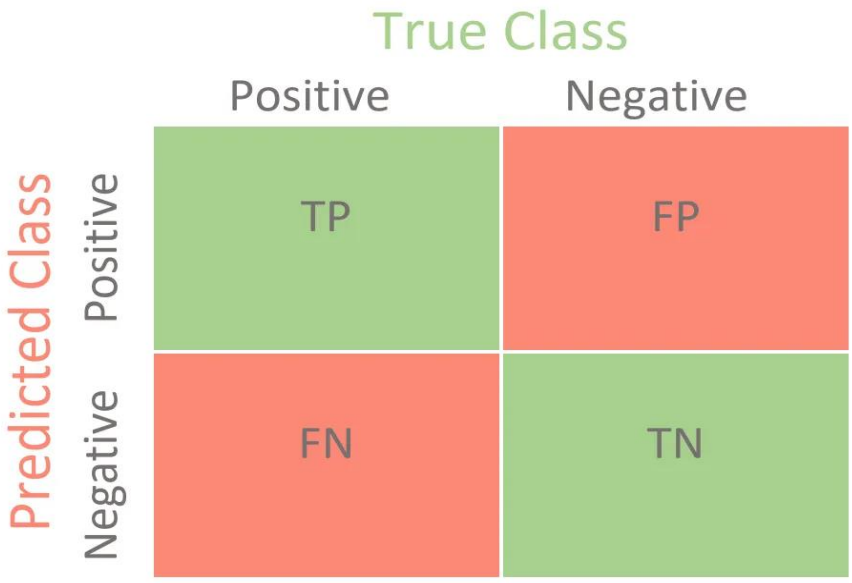

Figure 5.1. Confusion Matrix [182]

Figure 5.1 represents the values of the confusion matrix, True Positive (TP) indicates that the positive prediction made by your model is correct. A True Negative (TN) indicates that your prediction of a negative was accurate. False Positive (FP) denotes a false positive prediction. It is incorrect to predict a negative result when you make a False Negative (FN). The confusion matrix

results of the implemented algorithms are shown in Figure 5.3, Figure 5.2, Figure 5.4, and Figure 5.5.

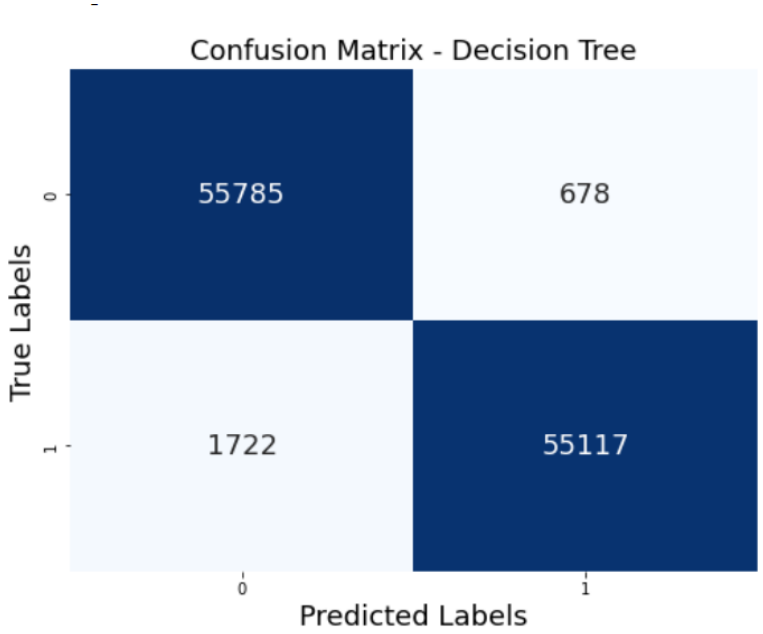

Figure 5.2. Decision Trees Results

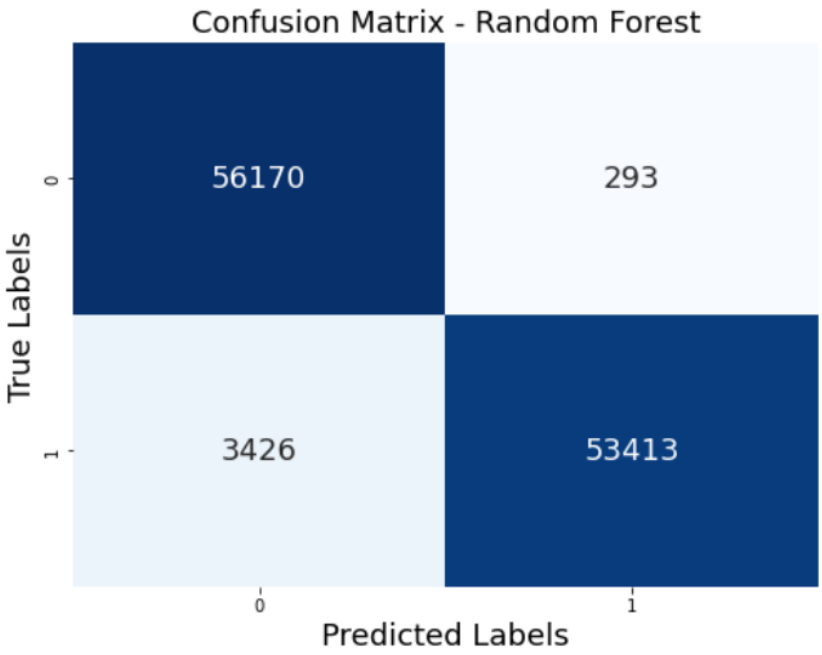

Figure 5.3. Random Forest Results

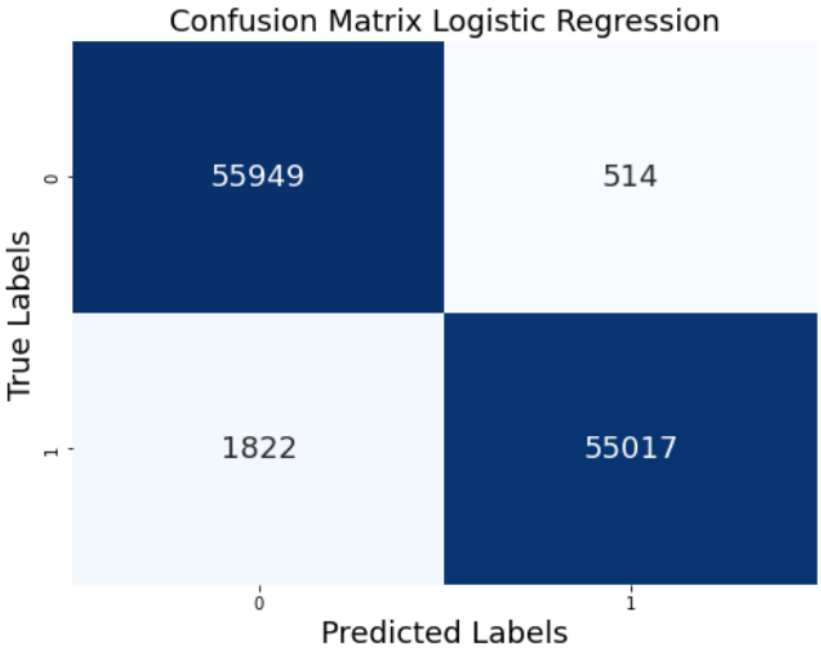

Figure 5.4. Logistic Regression Results

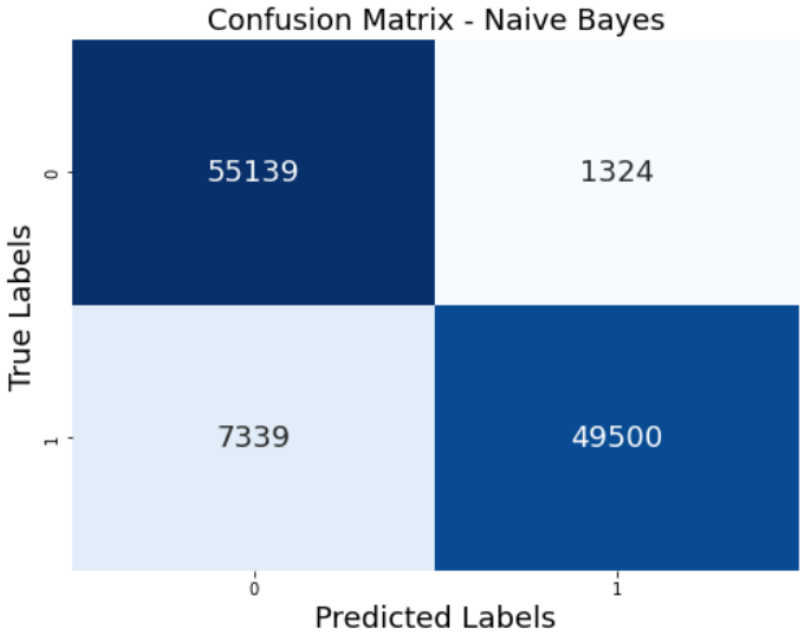

Figure 5.5. Naïve Bayes Results

Figure 5.2, yielded the Decision Tree classifier results: TP=55785, TN=55117, FP=678, and FN= 1722. Figure 5.3 presents the Random Forest classifier results, the accurate estimates (TP=56170; TN=53413), and the wrong estimates (FP=1564; FN=3426). As shown in Figure 5.4, the Logistic Regression results are TP = 55949, TN = 55017, FP = 514, and FN = 1822. In contrast, the findings of the Naïve Bayes classifier shown in Figure 5.5 are TP = 55139, TN = 49500, FP = 1324, and FN = 7339.

- **Classification Report**

The percentage of accurately predicted outcomes is called accuracy. Calculating the number of TPs and TNs, and then dividing the result by the overall predictions, is the way to determine the classifier's overall accuracy (Equation (1)) [183].

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \quad (1)$$

But accuracy is not necessarily the best indicator of a classifier's effectiveness; especially when there is an imbalance in the classes. To have a deeper grasp of the classifier's performance, various measures such as accuracy, recall, and F1-score should be used to evaluate its performance. The number of outputs that are correctly identified serves as a measure of precision (Equation (2)) [183].

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \quad (2)$$

The proportion of True Positives that the model successfully detected is known as recall (Equation (3)) [183].

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \quad (3)$$

On the other hand, the F1 score can be described as the balanced mean of precision and recall (Equation (4)) [184].

Table 5.1 shows the outcomes for every classifier.

$$\text{F1 score} = 2 \times \frac{(\text{precision} \times \text{recall})}{(\text{precision} + \text{recall})} \quad (4)$$

Table 5.1. Classification Report Results (all classifiers)

Classifier	Accuracy	Class	Precision	Recall	F1-score	support
		0	0.94	0.99	0.97	56463

Random Forest	96.717	1	0.99	0.94	0.97	56463
Decision Tree	**97.881**	0	0.97	0.99	0.98	56463
		1	0.99	0.97	0.98	56463
Logistic Regression	**97.938**	0	0.97	0.99	0.98	56463
		1	0.99	0.97	0.98	56463
Naïve Bayes	92.354	0	0.88	0.98	0.93	56463
		1	0.97	0.87	0.92	56463

Decision Tree and Logistic Regression achieved nearly equal levels of accuracy, as Table 5.1 illustrates. Decision Tree achieves an accuracy of 97.881%. With precision, recall, and F1 score of 97%, 99%, and 98% for class 0 (legal transaction) and 99%, 97%, and 98% for class 1 (fraud transaction) respectively. Logistic Regression gains an accuracy of 97.938%, with precision, recall, and F1 score of 97%, 99%, and 98% respectively for class 0 (legitimate transaction), precision, recall, and F1 score of 99%, 97%, and 98% respectively for class 1 (fraud transaction).

- The ROC Curve

A graphical representation that indicates the degree to which a binary classifier algorithm can identify issues depending on its ability to differentiate levels. Plotting the “true positive rate TPR” and recall, sometimes called sensitivity relative to the “false positive rate FPR”, also referred to as

the likelihood of false alerts, is one way to do this [185]. Refer to Figure 5.6, which displays the ROC curve findings for the classifiers that were put into use.

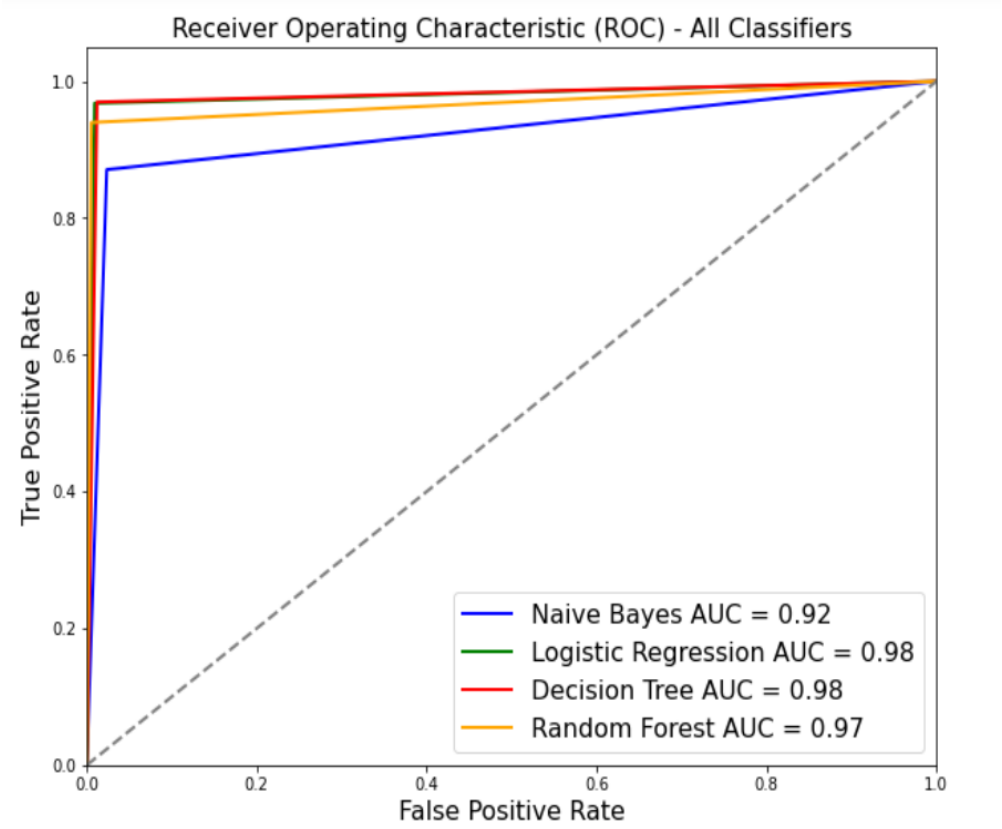

Figure 5.6. The ROC Curve Results (all classifiers)

AUC values range from 0 to 1, where 0.5 denotes a classifier that is no more successful than a wild guess and 1 denotes perfect performance. Figure 5.6 shows that the random forest classifier's AUC was 0.97. It shows that the classifier successfully classifies 97% of positive cases as such and 97% of negative cases as such. It also shows that the classifier has a low false positive rate, or the percentage of cases in which it incorrectly labels negative cases as positive. Decision Tree, Logistic Regression, and Naïve Bayes have AUCs of 0.98, 0.98, and 0.92, in that order. Lastly, the

ML section results do not conflict with previous research in the field [125], [136], [154], [186], [187], [188].

5.3. Summary

The integration of ML and MFA to protect Internet financial transactions was discussed and put into practice in this chapter. The investigation made an experimental setup that included an HP laptop, and Python language was used to write the code in the Anaconda Jupyter platform. This potent combination made it easier to design and assess ML models, guaranteeing a reliable and effective procedure.

The accuracy of the implemented classifiers; Random Forest (RF), Decision Trees (DT), Logistic Regression (LR), and Naïve Bayes (NB) were 96.717%, 97.881%, 97.938%, and 92.354% respectively. Other metrics and measures were utilized to evaluate the ML model results which are; recall, precision, F1-score, confusion matrix, and ROC curve.

This way provided a thorough assessment of the models' discriminating ability between both classes (1 is fraud, 0 is legitimate), and provided significant insight into the predictive power, accuracy, and error rates of the models.

Chapter 6

Conclusion and Future Work

6. Conclusion and Future Work

6.1. Conclusion

The pressing need to improve Internet financial transaction security justifies the study's significance. The dangers of internet-based transactions have grown more noticeable in an era where financial services are becoming more and more digitally engaged. Users and financial institutions are in danger from cyber threats such as identity theft, data breaches, and fraudulent operations. The existing dependence on single- or dual-factor authentication techniques is showing itself to be inadequate in the face of the constantly changing cyber threat environment and the implementation of MFA faces the truth of user annoyance and therefore affects the usage of such systems.

Because of that, this study offered a novel framework to safeguard Internet financial transactions and overcome the previous work in the literature by adding more layers of security, while offering a user-friendly system. Taking advantage of ML ability and making it work as an embedded and additional layer of security within an MFA framework, was the strength and distinction of this study. The proposed framework can be deployed in any e-commerce website, application, or platform where the user creates an account and needs to perform a payment process.

During the thesis journey, every chapter in this study explored a different aspect of enhancing the security of Internet financial transactions. Chapter One established the foundation by outlining the importance of safe online financial transactions. It discussed the rise of cyber dangers and the necessity of protecting financial data to explain why the digital banking industry needs stronger security measures. The chapter also provided an overview of the problem statements, study objectives, significance of the study, contribution to the body of knowledge, ethical issues, and thesis general organization.

The second chapter offered a thorough analysis of the e-commerce environment, explaining its importance and various facets in the digital sphere. The chapter started with an introduction that places e-commerce's evolution in perspective. From there, it defined the term's scope and outlined the advantages it provided for both consumers and enterprises. By examining different e-commerce models and digital platforms, the chapter sheds light on the range of channels that are available for conducting online transactions. The chapter additionally looked at some of the obstacles to e-commerce adoption and illuminated the difficulties that companies face in this ever-changing landscape. Finally, it explored e-payment techniques and examined the variety of choices for enabling safe and practical online transactions.

An in-depth review of the state of the art of ML and authentication factors about financial security was provided in Chapter Three. It carefully covered the entire range of authentication factors, including possession and knowledge-based, biometric, and behavior-based authentication. The chapter also covered the field of ML by highlighting the working principle of the different ML

algorithm categories and discussed the related work. Finally, the chapter discussed other studies that tried to combine ML and MFA and identified gaps that the current work seeks to fill.

Chapter Four, which served as the study's core, provided an in-depth analysis of the techniques developed to support Internet financial transactions. It walked through the chosen research approach while introducing the architecture of the suggested framework. This includes building the ML model by acquiring a dataset, preprocessing the data, testing many ML classifiers, and choosing the best performer. Notably, it also highlighted the use of authentication factors by classifying various authentication techniques and how they work together. The climax of this chapter was the integration of ML and MFA into an application architecture, which offered a comprehensive and realistic example of this integration.

The study results were shown in Chapter Five. This chapter detailed the results of ML, and carefully analyzed measures including ROC curves, confusion matrices, and classification reports. It highlighted the pressing need to utilize different metrics to improve ML results, especially when working with unbalanced datasets.

6.2. Obstacles and Mitigation Strategies

6.2.1. Obstacles

Authentication factors: Potential user resistance or discomfort with the chosen methods presents one challenge when choosing the MFA for security purposes. Sometimes people feel that MFA is

too complicated and annoying to use, which can cause resistance or lower user acceptance of such systems. The challenge is to choose a secure factor to authenticate users along with taking into consideration the ease of use.

Data availability: One major obstacle is the absence of necessary data. Organizations may choose to hide financial transaction data according to privacy and security considerations, and the needed datasets may not be publicly accessible due to the sensitivity of this data. One of the main challenges is getting access to representative and comprehensive databases.

Data quality: The study may be greatly impacted by the quality of the accessible data. Data that is missing, incorrect, unbalanced, or inconsistent can make ML models and authentication systems less effective and possibly produce biased or incorrect results. In particular, most of the available datasets are transformed using the PCA transformation technique to satisfy the user's privacy policy.

Technical limitations: Technical barriers, such as compatibility issues or limited storage capacity, and processing speed may restrict the ability to handle, process, and store large volumes of data efficiently. These limitations might impact the implementation of the proposed MFA and ML framework.

6.2.2. Mitigation Strategies

To address the challenges mentioned in Section 6.2.1, the research implemented the following strategies: -

Suitable MFA implementation: a user-centric strategy was used to gain adaptable, secure, and user-friendly system implementation. The adaptive implementation of the MFA system led to interaction with only two factors when put into practice. A third factor is required if the ML algorithm classifies the transaction as fraud. This preserves strong security standards while simultaneously improving usability.

Data cleaning and preprocessing: Using techniques to remove errors and deal with unbalanced datasets that could affect the models' accuracy in cleaning and preparing data. This was done successfully and discussed in Section 4.4 (ML phase).

Replication: Conducting the ML analysis at different times to confirm and guarantee the reliability and consistency of the results while reducing the influence of anomalies or errors.

Algorithm and analysis suitability: Using the right statistical techniques and ML algorithms to analyze the data while taking hardware constraints into account. Identifying and evaluating the best algorithms for the particular use case of safe financial transactions was done carefully. The implemented ML algorithms were simple and accurate to overcome the hardware limitations and facilitate the integration of ML and MFA into one model.

6.3. Future Work

Future work on the integration of improved biometric recognition systems is a promising option, in addition to increasing the range of authentication elements and improving the dataset. Working

on a dataset of customer purchases with full knowledge about these data will enhance the ML model's accuracy. While some technologies, like fingerprint scanning and facial recognition, are currently widely used, there are many more opportunities due to the ongoing advancements in biometric authentication. This could entail using behavioral biometrics like keystroke dynamics or gait analysis, as well as voice recognition and iris scanning. The MFA system may be strengthened even further by these cutting-edge biometric authentication techniques, increasing its resilience and usability.

Moreover, the framework of the study might be modified and applied in other fields outside of e-commerce, like online banking, healthcare, or government services, where user authentication and safe transactions are critical. Adapting the framework to these diverse domains' unique security and usability needs could produce beneficial outcomes and advance online security in a variety of industries.

Last but not least, keeping up with the most recent advancements in cyber security and ML is essential as technology keeps on developing. To keep the framework at the top of security technology, future research might concentrate on combining cutting-edge ML algorithms and cyber security measures. Also, keeping up with evolving cyber threats and vulnerabilities can help with proactive security measures and preemptive defense mechanisms that keep possible cyber-attacks away.

References

Abdollahi, G., & Leimstoll, U. (2011). A Classification for Business Model Types in E-commerce. *AMCIS 2011 Proceedings - All Submissions*. https://aisel.aisnet.org/amcis2011_submissions/88

Abdulghani, A. Q., Ucan, O. N., & Alheeti, K. M. A. (2021). Credit Card Fraud Detection Using XGBoost Algorithm. *Proceedings - International Conference on Developments in ESystems Engineering, DeSE*, *2021-December*, 487–492. https://doi.org/10.1109/DESE54285.2021.9719580

Abidin, T. F., Yusuf, B., & Umran, M. (2010). Singular Value Decomposition for dimensionality reduction in unsupervised text learning problems. *ICETC 2010 - 2010 2nd International Conference on Education Technology and Computer*, *4*. https://doi.org/10.1109/ICETC.2010.5529649

Aburbeian, A. H. M., & Ashqar, H. I. (2023). Credit Card Fraud Detection Using Enhanced Random Forest Classifier for Imbalanced Data. *Lecture Notes in Networks and Systems*, *700 LNNS*, 605–616. https://doi.org/10.1007/978-3-031-33743-7_48/COVER

Aburbeian, A. M., & Ashqar, H. I. (2023). Credit Card Fraud Detection Using Enhanced Random Forest Classifier for Imbalanced Data. *Lecture Notes in Networks and Systems. In the Proceedings of the 2023 International Conference on Advances in Computing Research (ACR'23). Springer, Cham*, 605–616. https://doi.org/10.1007/978-3-031-33743-7_48/COVER

Adalo: Design & Build Custom Apps • No Code Required. (n.d.). Retrieved November 27, 2023, from https://www.adalo.com/?via=thuong&gclid=CjwKCAiAmZGrBhAnEiwAo9qHiVGvZPs8aoAFs0TtX4eiOIC0CMzKvF8j-23UK6dktVjmcqFFU8CjkhoCza4QAvD_BwE

Adjabi, I., Ouahabi, A., Benzaoui, A., & Taleb-Ahmed, A. (2020). Past, Present, and Future of Face Recognition: A Review. *Electronics 2020, Vol. 9, Page 1188*, *9*(8), 1188. https://doi.org/10.3390/ELECTRONICS9081188

Ah Kioon, M. C., Wang, Z. S., & Deb Das, S. (2013). Security Analysis of MD5 Algorithm in Password Storage. *Applied Mechanics and Materials*, *347–350*, 2706–2711. https://doi.org/10.4028/WWW.SCIENTIFIC.NET/AMM.347-350.2706

Ahmed, N., Amin, R., Aldabbas, H., Koundal, D., Alouffi, B., & Shah, T. (2022). Machine Learning Techniques for Spam Detection in Email and IoT Platforms: Analysis and Research Challenges. *Security and Communication Networks*, *2022*. https://doi.org/10.1155/2022/1862888

Akanksha, E., Jyoti, Sharma, N., & Gulati, K. (2021). Review on Reinforcement Learning, Research Evolution and Scope of Application. *Proceedings - 5th International Conference on Computing*

Methodologies and Communication, ICCMC 2021, 1416–1423. https://doi.org/10.1109/ICCMC51019.2021.9418283

Alenzi, H. Z., & Aljehane, N. O. (2020). Fraud Detection in Credit Cards using Logistic Regression. *International Journal of Advanced Computer Science and Applications*, *11*(12), 540–551. https://doi.org/10.14569/IJACSA.2020.0111265

Alhakami, H., & Alhrbi, S. (2020). Knowledge based Authentication Techniques and Challenges. *International Journal of Advanced Computer Science and Applications*, *2*, 727–732. https://doi.org/10.14569/IJACSA.2020.0110291

Ali, G., Dida, M. A., & Sam, A. E. (2021). A Secure and Efficient Multi-Factor Authentication Algorithm for Mobile Money Applications. *Future Internet 2021, Vol. 13, Page 299*, *13*(12), 299. https://doi.org/10.3390/FI13120299

Alqahtani, A. A. S., El-Awadi, Z., & Min, M. (2021). A Survey on User Authentication Factors. *2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON 2021*, 323–328. https://doi.org/10.1109/IEMCON53756.2021.9623159

Alzoubi, H. M., Alshurideh, M. T., Kurdi, B. Al, Alhyasat, K. M. K., & Ghazal, T. M. (2022). The effect of e-payment and online shopping on sales growth: Evidence from banking industry. *International Journal of Data and Network Science*, *6*(4), 1369–1380. https://doi.org/10.5267/J.IJDNS.2022.5.014

Aouedi, O., Piamrat, K., Hamma, S., & Perera, J. K. M. (2022). Network traffic analysis using machine learning: an unsupervised approach to understand and slice your network. *Annales Des Telecommunications/Annals of Telecommunications*, *77*(5–6), 297–309. https://doi.org/10.1007/S12243-021-00889-1/METRICS

Archetti, C., & Bertazzi, L. (2021). Recent challenges in Routing and Inventory Routing: E-commerce and last-mile delivery. *Networks*, *77*(2), 255–268. https://doi.org/10.1002/NET.21995

Arnold, D., Blackmon, B., Gibson, B., Moncivais, A. G., Powell, G. B., Skeen, M., Thorson, M. K., & Wade, N. B. (2022). The Emotional Impact of Multi-Factor Authentication for University Students. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3491101.3516809

Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017). Credit card fraud detection using machine learning techniques: A comparative analysis. *Proceedings of the IEEE International Conference on Computing, Networking and Informatics, ICCNI 2017*, *2017-January*, 1–9. https://doi.org/10.1109/ICCNI.2017.8123782

Back, A. (2023). *4 Payment Methods Most Widely Accepted in the US in 2023*. Pay.Com. https://pay.com/blog/4-methods-of-payments-accepted-in-us

Baesens, B., Höppner, S., & Verdonck, T. (2021). Data engineering for fraud detection. *Decision Support Systems*, *150*, 113492. https://doi.org/10.1016/J.DSS.2021.113492

Bansal, A., Khosla, T., & Saini, V. K. (2023). Security Challenges and various methods for Increasing Security in E-Commerce Applications. *International Journal for Research in Applied Science and Engineering Technology*, *11*. https://doi.org/10.22214/ijraset.2023.48475

BasuMallick, C. (2023). *Principal Component Analysis Working and Applications | Spiceworks - Spiceworks*. Spicework. https://www.spiceworks.com/tech/big-data/articles/what-is-principal-component-analysis/

BERR, J. (2017). *"WannaCry" ransomware attack losses could reach $4 billion - CBS News*. CBC News. https://www.cbsnews.com/news/wannacry-ransomware-attacks-wannacry-virus-losses/

Bingi, P., Mir, A., & Khamalah, J. (2000). The Challenges Facing Global E-Commerce. *Information Systems Management*, *17*(4), 22–30. https://doi.org/10.1201/1078/43193.17.4.20000901/31249.5

Bin Sulaiman, R., Schetinin, V., & Sant, P. (2022). Review of Machine Learning Approach on Credit Card Fraud Detection. *Human-Centric Intelligent Systems 2022 2:1*, *2*(1), 55–68. https://doi.org/10.1007/S44230-022-00004-0

Bolton, R. J., & Hand, D. J. (2001). Unsupervised Profiling Methods for Fraud Detection. *Credit Scoring and Credit Control VII*, 235–255.

Bonneau, J., Herley, C., Van Oorschot, P. C., & Stajano, F. (2015). Passwords and the evolution of imperfect authentication. *Communications of the ACM*, *58*(7), 78–87. https://doi.org/10.1145/2699390

Boonkrong, S. (2021). Methods and Threats of Authentication. *Authentication and Access Control*, 45–70. https://doi.org/10.1007/978-1-4842-6570-3_3

Boute, R. N., Gijsbrechts, J., van Jaarsveld, W., & Vanvuchelen, N. (2022). Deep reinforcement learning for inventory control: A roadmap. *European Journal of Operational Research*, *298*(2), 401–412. https://doi.org/10.1016/J.EJOR.2021.07.016

Bouveret, A. (2018). Cyber risk for the financial sector: A framework for quantitative assessment. *International Monetary Fund*. https://books.google.com/books?hl=ar&lr=&id=n7QZEAAAQBAJ&oi=fnd&pg=PA3&dq=Bouveret,+Antoine.+2018.+%E2%80%9CCyber+Risk+for+the+Financial+Sector:+A+Framework+f

or+Quantitative+Assessment.%E2%80%9D+IMF+Working+Paper,+June+22.&ots=47D52Uzv-3&sig=ljA0IxUGzKpUl6seu1IN245VtW8

Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Analytical Methods*, *6*(9), 2812–2831. https://doi.org/10.1039/C3AY41907J

Brown, P. N., Borowski, H. P., & Marden, J. R. (2019). Security against impersonation attacks in distributed systems. *IEEE Transactions on Control of Network Systems*, *6*(1), 440–450. https://doi.org/10.1109/TCNS.2018.2838519

Bui, V. H., Hussain, A., & Su, W. (2022). A Dynamic Internal Trading Price Strategy for Networked Microgrids: A Deep Reinforcement Learning Based Game-Theoretic Approach. *IEEE Transactions on Smart Grid*. https://doi.org/10.1109/TSG.2022.3168856

Burkart, N., & Huber, M. F. (2021). A Survey on the Explainability of Supervised Machine Learning. *Journal of Artificial Intelligence Research*, *70*, 245–317. https://doi.org/10.1613/JAIR.1.12228

Burr, W., Dodson, D., & Polk, ; W Timothy. (2004). Archived NIST Technical Series Publication Archived Publication Series/Number: Title: Publication Date(s): Withdrawal Date: Superseding Publication(s) Electronic Authentication Guideline Electronic Authentication Guideline. *NIST Special Publication 800*. https://doi.org/10.6028/NIST.SP.800-63v1.0.1

Carcillo, F., Le Borgne, Y. A., Caelen, O., Kessaci, Y., Oblé, F., & Bontempi, G. (2021). Combining unsupervised and supervised learning in credit card fraud detection. *Information Sciences*, *557*, 317–331. https://doi.org/10.1016/J.INS.2019.05.042

Celebi, M. E., & Aydin, K. (2016). Unsupervised learning algorithms. *Cham: Springe*, 1–558. https://doi.org/10.1007/978-3-319-24211-8/COVER

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/JAIR.953

Cheng, L., Liu, F., & Yao, D. D. (2017). Enterprise data breach: causes, challenges, prevention, and future directions. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *7*(5), e1211. https://doi.org/10.1002/WIDM.1211

Chetalam, L. J. (2018). ENHANCING SECURITY OF MPESA TRANSACTIONS BY USE OF VOICE BIOMETRICS. *Diss. United States International University-Africa*.

Chevers, D. (2019). The impact of cybercrime on e-banking: A proposed model. *CONF-IRM 2019 Proceedings*. https://aisel.aisnet.org/confirm2019/11

Christina, V., Karpagavalli, S., Suganya, G., & Phil, # M. (2010). Email Spam Filtering using Supervised Machine Learning Techniques. *IJCSE) International Journal on Computer Science and Engineering*, *02*(09), 3126–3129.

Cohen, S. (2021). The evolution of machine learning: past, present, and future. *Artificial Intelligence and Deep Learning in Pathology*, 1–12. https://doi.org/10.1016/B978-0-323-67538-3.00001-4

Cottam, J. A., Heller, N. C., Ebsch, C. L., Deshmukh, R., MacKey, P., & Chin, G. (2020). Evaluation of Alignment: Precision, Recall, Weighting and Limitations. *Proceedings - 2020 IEEE International Conference on Big Data, Big Data 2020*, 2513–2519. https://doi.org/10.1109/BIGDATA50022.2020.9378064

Crawford, M., Khoshgoftaar, T. M., Prusa, J. D., Richter, A. N., & Al Najada, H. (2015). Survey of review spam detection using machine learning techniques. *Journal of Big Data*, *2*(1), 1–24. https://doi.org/10.1186/S40537-015-0029-9/TABLES/9

Crisci, C., Ghattas, B., & Perera, G. (2012). A review of supervised machine learning algorithms and their applications to ecological data. *Ecological Modelling*, *240*, 113–122. https://doi.org/10.1016/J.ECOLMODEL.2012.03.001

Dasgupta, D., Roy, A., & Nag, A. (2016). Toward the design of adaptive selection strategies for multi-factor authentication. *Computers & Security*, *63*, 85–116. https://doi.org/10.1016/J.COSE.2016.09.004

Dasgupta, D., Roy, A., & Nag, A. (2017). Multi-Factor Authentication. *Advances in User Authentication*, 185–233. https://doi.org/10.1007/978-3-319-58808-7_5

De Cristofaro, E., Du, H., Freudiger, J., & Norcie, G. (2013). A Comparative Usability Study of Two-Factor Authentication. *ArXiv Preprint ArXiv:1309.5344*. https://doi.org/10.14722/usec.2014.23025

de Luna, I. R., Liébana-Cabanillas, F., Sánchez-Fernández, J., & Muñoz-Leiva, F. (2019). Mobile payment is not all the same: The adoption of mobile payment systems depending on the technology applied. *Technological Forecasting and Social Change*, *146*, 931–944. https://doi.org/10.1016/J.TECHFORE.2018.09.018

Deridder, Z., Siddiqui, N., Reither, T., Dave, R., Pelto, B., Vanamala, M., & Seliya, N. (2022). Continuous User Authentication Using Machine Learning and Multi-finger Mobile Touch Dynamics with a Novel Dataset. *2022 9th International Conference on Soft Computing and Machine Intelligence, ISCMI 2022*, 42–46. https://doi.org/10.1109/ISCMI56532.2022.10068450

Doycheva, K., Horn, G., Koch, C., Schumann, A., & König, M. (2017). Assessment and weighting of meteorological ensemble forecast members based on supervised machine learning with

application to runoff simulations and flood warning. *Advanced Engineering Informatics*, *33*, 427–439. https://doi.org/10.1016/J.AEI.2016.11.001

Duman, E., Buyukkaya, A., & Elikucuk, I. (2013). A novel and successful credit card fraud detection system implemented in a Turkish bank. *Proceedings - IEEE 13th International Conference on Data Mining Workshops, ICDMW 2013*, 162–171. https://doi.org/10.1109/ICDMW.2013.168

Dumortier, J. (2016). Regulation (EU) No 910/2014 on Electronic Identification and Trust Services for Electronic Transactions in the Internal Market (eIDAS Regulation). *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.2855484

Dutson, J., Allen, D., Eggett, D., & Seamons, K. (2019). Don't Punish all of us: Measuring User Attitudes about Two-Factor Authentication. *Proceedings - 4th IEEE European Symposium on Security and Privacy Workshops, EUROS and PW 2019*, 119–128. https://doi.org/10.1109/EUROSPW.2019.00020

Eid Alanzi, T., & Naif Alatawi, M. (2022). A Secure Two-factor Authentication Framework Based on Deep Learning. *Journal of Research in Science and Engineering (JRSE)*, *4*(6). https://doi.org/10.53469/jrse.2022.04(06).24

El Naqa, I., & Murphy, M. J. (2015). What Is Machine Learning? *Machine Learning in Radiation Oncology*, 3–11. https://doi.org/10.1007/978-3-319-18305-3_1

Elreedy, D., & Atiya, A. F. (2019). A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance. *Information Sciences*, *505*, 32–64. https://doi.org/10.1016/J.INS.2019.07.070

Fernández, A., García, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. *Journal of Artificial Intelligence Research*, *61*, 863–905. https://doi.org/10.1613/JAIR.1.11192

Finezza, T. (2022). *Effects of Mobile Banking on Digital Economy: Boon or a Curse? - Finezza Blog*. Finezza. https://finezza.in/blog/effects-mobile-banking-digital-economy/

Fourati, A., Ben Ayed, H. K., Kamoun, F., & Benzekri, A. (2002). A SET based approach to secure the payment in mobile commerce. *Proceedings - Conference on Local Computer Networks, LCN*, *2002-January*, 136–137. https://doi.org/10.1109/LCN.2002.1181777

Gaikwad, J. R., Deshmane, A. B., Somavanshi, H. V, Patil, S. V, & Badgujar, R. A. (2014). Credit Card Fraud Detection using Decision Tree Induction Algorithm. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, *6*, 2278–3075.

Gentleman, R., & Carey, V. J. (2008). Unsupervised Machine Learning. *Bioconductor Case Studies*, 137–157. https://doi.org/10.1007/978-0-387-77240-0_10

Gianey, H. K., & Choudhary, R. (2018). Comprehensive Review On Supervised Machine Learning Algorithms. *Proceedings - 2017 International Conference on Machine Learning and Data Science, MLDS 2017*, *2018-January*, 38–43. https://doi.org/10.1109/MLDS.2017.11

Groß, S., Lein, S., & Steinbrecher, S. (2005). A multilateral secure payment system for wireless LAN hotspots. *Lecture Notes in Computer Science*, *3592*, 80–89. https://doi.org/10.1007/11537878_9/COVER

Guma, A. (2022). Development of a secure multi-factor authentication algorithm for mobile money applications. *PhD Thesis. NM-AIST*. http://dspace.nm-aist.ac.tz/handle/20.500.12479/1782

Gunson, N., Marshall, D., Morton, H., & Jack, M. (2011a). User perceptions of security and usability of single-factor and two-factor authentication in automated telephone banking. *Computers & Security*, *30*(4), 208–220. https://doi.org/10.1016/J.COSE.2010.12.001

Gunson, N., Marshall, D., Morton, H., & Jack, M. (2011b). User perceptions of security and usability of single-factor and two-factor authentication in automated telephone banking. *Computers & Security*, *30*(4), 208–220. https://doi.org/10.1016/J.COSE.2010.12.001

Gunson, N., Marshall, D., Morton, H., & Jack, M. (2011c). User perceptions of security and usability of single-factor and two-factor authentication in automated telephone banking. *Computers & Security*, *30*(4), 208–220. https://doi.org/10.1016/J.COSE.2010.12.001

Guo, M. H., Xu, T. X., Liu, J. J., Liu, Z. N., Jiang, P. T., Mu, T. J., Zhang, S. H., Martin, R. R., Cheng, M. M., & Hu, S. M. (2022). Attention mechanisms in computer vision: A survey. *Computational Visual Media*, *8*(3), 331–368. https://doi.org/10.1007/S41095-022-0271-Y/METRICS

Gupta, A. (2014). E-COMMERCE : ROLE OF E-COMMERCE IN TODAY'S BUSINESS. *International Journal of Computing and Corporate Research*, *4*(1), 1–8.

Han, Y., Duan, L., & Zhang, R. (2019). Jamming-Assisted Eavesdropping over Parallel Fading Channels. *IEEE Transactions on Information Forensics and Security*, *14*(9), 2486–2499. https://doi.org/10.1109/TIFS.2019.2901821

Hao, J., & Ho, T. K. (2019). Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language. *Journal of Educational and Behavioral Statistics*, *44*(3), 348–361. https://doi.org/10.3102/1076998619832248/SUPPL_FILE/DS_10.3102_1076998619832248.ZIP

Hassan, M. A., & Shukur, Z. (2021). A Secure Multi Factor User Authentication Framework for Electronic Payment System. *2021 3rd International Cyber Resilience Conference, CRC 2021*. https://doi.org/10.1109/CRC50527.2021.9392564

Haug, C. J., & Drazen, J. M. (2023). Artificial Intelligence and Machine Learning in Clinical Medicine, 2023. *New England Journal of Medicine*, *388*(13), 1201–1208. https://doi.org/10.1056/NEJMRA2302038/SUPPL_FILE/NEJMRA2302038_DISCLOSURES.PDF

Heartfield, R., & Loukas, G. (2015). A Taxonomy of Attacks and a Survey of Defence Mechanisms for Semantic Social Engineering Attacks. *ACM Computing Surveys (CSUR)*, *48*(3). https://doi.org/10.1145/2835375

Hoo, Z. H., Candlish, J., & Teare, D. (2017). What is an ROC curve? *Emergency Medicine Journal*, *34*(6), 357–359. https://doi.org/10.1136/EMERMED-2017-206735

Hruschka, H. (2021). Comparing unsupervised probabilistic machine learning methods for market basket analysis. *Review of Managerial Science*, *15*(2), 497–527. https://doi.org/10.1007/S11846-019-00349-0/TABLES/17

Huang, H., Zavareh, A. A., & Mustafa, M. B. (2023). Sentiment Analysis in E-Commerce Platforms: A Review of Current Techniques and Future Directions. *IEEE Access*, *11*, 90367–90382. https://doi.org/10.1109/ACCESS.2023.3307308

Huang, Y., Chai, Y., Liu, Y., & Shen, J. (2019). Architecture of next-generation e-commerce platform. *Tsinghua Science and Technology*, *24*(1), 18–29. https://doi.org/10.26599/TST.2018.9010067

Huang, Z., & Chen, K. (2002). Electronic payment in mobile environment. *Proceedings - International Workshop on Database and Expert Systems Applications, DEXA*, *2002-January*, 413–417. https://doi.org/10.1109/DEXA.2002.1045930

Hu, Q., Du, B., Markantonakis, K., & Hancke, G. P. (2020). A session hijacking attack against a device-assisted physical-layer key agreement. *IEEE Transactions on Industrial Informatics*, *16*(1), 691–702. https://doi.org/10.1109/TII.2019.2923662

Huseynov, F., & Özkan Yıldırım, S. (2019). Online Consumer Typologies and Their Shopping Behaviors in B2C E-Commerce Platforms. *SAGE Open*, *9*(2). https://doi.org/10.1177/2158244019854639/ASSET/IMAGES/LARGE/10.1177_2158244019854639-FIG3.JPEG

Ingale, M., Cordeiro, R., Thentu, S., Park, Y., & Karimian, N. (2020). ECG Biometric Authentication: A Comparative Analysis. *IEEE Access*, *8*, 117853–117866. https://doi.org/10.1109/ACCESS.2020.3004464

Itoo, F., Meenakshi, & Singh, S. (2021a). Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection. *International Journal of Information Technology (Singapore)*, *13*(4), 1503–1511. https://doi.org/10.1007/S41870-020-00430-Y/METRICS

Itoo, F., Meenakshi, & Singh, S. (2021b). Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection. *International Journal of Information Technology (Singapore)*, *13*(4), 1503–1511. https://doi.org/10.1007/S41870-020-00430-Y/METRICS

Jain, A., Purwar, A., & Yadav, D. (2021). Credit Card Fraud Detection Using K-Means and Fuzzy C-Means. *Handbook of Research on Innovations and Applications of AI, IoT, and Cognitive Technologies. IGI Global, 2021*, 216–240. https://doi.org/10.4018/978-1-7998-6870-5.CH016

Jain, S. (2019). A comparative analysis of various credit card fraud detection techniques. *International Journal of Recent Technology and Engineering*, 2277–3878. https://www.researchgate.net/publication/332264296

Jain, V., Malviya, B., & Arya, S. (2021). An Overview of Electronic Commerce (e-Commerce). *Journal of Contemporary Issues in Business and Government* , *27*(3), 665–670. https://doi.org/10.47750/cibg.2021.27.03.090

Jaspher, G., Kathrine, W., & Kirubakaran, E. (2011). Four-Factor based Privacy Preserving Biometric Authentication and Authorization Scheme for Enhancing Grid Security. *International Journal of Computer Applications*, *30*(5), 975–8887.

Jia, J., & Wang, W. (2020). Review of reinforcement learning research. *Proceedings - 2020 35th Youth Academic Annual Conference of Chinese Association of Automation, YAC 2020*, 186–191. https://doi.org/10.1109/YAC51587.2020.9337653

Jiang, T., Gradus, J. L., & Rosellini, A. J. (2020). Supervised Machine Learning: A Brief Primer. *Behavior Therapy*, *51*(5), 675–687. https://doi.org/10.1016/J.BETH.2020.05.002

Joshi, P., & Dumbre, G. M. (2017). Basic Concept of E-Commerce. *RESEARCH JOURNAL OF MULTIDISCIPLINARY STUDIES*, *3*, 2454–8499. www.irjms.in

Kadiyala, P., Shanmukhasai, K. V., Budem, S. S., & Maddikunta, P. K. R. (2021). Anomaly Detection Using Unsupervised Machine Learning Algorithms. *Signals and Communication Technology*, 113–125. https://doi.org/10.1007/978-981-16-6186-0_6/COVER

Kariapper, R. (2021). Attendance system using RFID, IoT and Machine learning: A two factor verification approach. *Systematic Reviews in Pharmacy*, *12*(3), 314–321. https://scholar.google.com/scholar?hl=ar&as_sdt=0%2C5&q=+Attendance+system+using+RFID

%2C+IoT+and+Machine+learning%3A+A+two+factor+verification+approach%E2%80%8F&btnG=

Katsini, C., Belk, M., Fidas, C., Avouris, N., & Samaras, G. (2016). Security and usability in knowledge-based user authentication: A review. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3003733.3003764

Kaur, S., Kaur, G., & Shabaz, M. (2022). A Secure Two-Factor Authentication Framework in Cloud Computing. *Security and Communication Networks*, *2022*. https://doi.org/10.1155/2022/7540891

Kemp, L. (2021). *Most Used Payment Methods In Australia*. Expert Easy. https://www.experteasy.com.au/blog/most-used-payment-methods-in-australia/

Kennedy, E., & Millard, C. (2016). Data security and multi-factor authentication: Analysis of requirements under EU law and in selected EU Member States. *Computer Law & Security Review*, *32*(1), 91–110. https://doi.org/10.1016/J.CLSR.2015.12.004

Khattri, V., & Singh, D. K. (2019). Implementation of an Additional Factor for Secure Authentication in Online Transactions. *Journal of Organizational Computing and Electronic Commerce*, *29*(4), 258–273. https://doi.org/10.1080/10919392.2019.1633123

Kherif, F., & Latypova, A. (2020). Principal component analysis. *Machine Learning: Methods and Applications to Brain Disorders*, 209–225. https://doi.org/10.1016/B978-0-12-815739-8.00012-2

Kiran, S., Guru, J., Kumar, R., Kumar, N., Katariya, D., & Sharma, M. (2018). Credit card fraud detection using Naïve Bayes model based and KNN classifier. *International Journal of Advance Research*. www.IJARIIT.com

Konoth, R. K., van der Veen, V., & Bos, H. (2017). How anywhere computing just killed your phone-based two-factor authentication. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9603 LNCS*, 405–421. https://doi.org/10.1007/978-3-662-54970-4_24/COVER

Kou, Y., Lu, C. T., Sirwongwattana, S., & Huang, Y. P. (2004). Survey of fraud detection techniques. *Conference Proceeding - IEEE International Conference on Networking, Sensing and Control*, *2*, 749–754. https://doi.org/10.1109/ICNSC.2004.1297040

Krol, K., Philippou, E., De Cristofaro, E., & Sasse, M. A. (2015). "`They brought in the horrible key ring thing!" Analysing the Usability of Two-Factor Authentication in UK Online Banking. *ArXiv Preprint ArXiv:1501.04434*. https://doi.org/10.14722/usec.2015.23001

Kulatilleke, G. K., & Mary, Q. (2022a). Challenges and Complexities in Machine Learning based Credit Card Fraud Detection. *ArXiv Preprint ArXiv:2208.10943*. https://arxiv.org/abs/2208.10943v1

Kulatilleke, G. K., & Mary, Q. (2022b). Challenges and Complexities in Machine Learning based Credit Card Fraud Detection. *ArXiv Preprint ArXiv:2208.10943*. https://arxiv.org/abs/2208.10943v1

Kumar, M. S., Soundarya, V., Kavitha, S., Keerthika, E. S., & Aswini, E. (2019a). Credit Card Fraud Detection Using Random Forest Algorithm. *2019 Proceedings of the 3rd International Conference on Computing and Communications Technologies, ICCCT 2019*, 149–153. https://doi.org/10.1109/ICCCT2.2019.8824930

Kumar, M. S., Soundarya, V., Kavitha, S., Keerthika, E. S., & Aswini, E. (2019b). Credit Card Fraud Detection Using Random Forest Algorithm. *2019 Proceedings of the 3rd International Conference on Computing and Communications Technologies, ICCCT 2019*, 149–153. https://doi.org/10.1109/ICCCT2.2019.8824930

Kungpisdan, S., Srinivasan, B., & Le, P. D. (2004). A secure account-based mobile payment protocol. *International Conference on Information Technology: Coding Computing, ITCC*, *1*, 35–39. https://doi.org/10.1109/ITCC.2004.1286422

Laudon, K. C., & Traver, C. G. (2020). E-commerce 2019 : business, technology, society. *Pearson*, 840. https://thuvienso.hoasen.edu.vn/handle/123456789/12556

Leuprecht, C. (2019). Mitigating cyber risk across the financial sector. *JSTOR*. https://www.jstor.org/stable/pdf/resrep26129.15.pdf

Li, J., Ji, L., Zhang, C., & Li, H. (2022). Optimal couple-group tracking control for the heterogeneous multi-agent systems with cooperative-competitive interactions via reinforcement learning method. *Information Sciences*, *610*, 401–424. https://doi.org/10.1016/J.INS.2022.07.181

Lindholm, A., Wahlström, N., Lindsten, F., & Schön, T. B. (2019). Supervised Machine Learning Lecture notes for the Statistical Machine Learning course. *Department of Information Technology, Uppsala University: Uppsala, Sweden, 112*.

Lin, Y. B., Chang, M. F., & Rao, H. C. H. (2000). Mobile prepaid phone services. *IEEE Personal Communications*, *7*(3), 6–14. https://doi.org/10.1109/98.847918

Liu, C., Chan, Y., Hasnain, S., Kazmi, A., & Fu, H. (2015). Financial Fraud Detection Model: Based on Random Forest. *International Journal of Economics and Finance*, *7*(7). https://doi.org/10.5539/ijef.v7n7p178

Liu, X. Y., Wu, J., & Zhou, Z. H. (2009). Exploratory undersampling for class-imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, *39*(2), 539–550. https://doi.org/10.1109/TSMCB.2008.2007853

Li, Y. F., & Liang, D. M. (2019). Safe semi-supervised learning: a brief introduction. *Frontiers of Computer Science*, *13*(4), 669–676. https://doi.org/10.1007/S11704-019-8452-2/METRICS

Lopez, C., Tucker, S., Salameh, T., & Tucker, C. (2018). An unsupervised machine learning method for discovering patient clusters based on genetic signatures. *Journal of Biomedical Informatics*, *85*, 30–39. https://doi.org/10.1016/J.JBI.2018.07.004

Lutkevich, B. (2021). *What is smart card? | Definition from TechTarget*. TechTarget. https://www.techtarget.com/searchsecurity/definition/smart-card

MacIej, B., Imed, E. F., & Kurkowski, M. (2019). Multifactor Authentication Protocol in a Mobile Environment. *IEEE Access*, *7*, 157185–157199. https://doi.org/10.1109/ACCESS.2019.2948922

Maćkiewicz, A., & Ratajczak, W. (1993). Principal components analysis (PCA). *Computers & Geosciences*, *19*(3), 303–342. https://doi.org/10.1016/0098-3004(93)90090-R

Marjuni, A., Azman, M. N. A. A., Mustofa, H. A., & Sukadari, S. (2022). Development of the Android-Based Mobile Application "Mywheel Alignment" for Wheel Alignment Topics in Automotive Technology Courses at Vocational Colleges. *Asian Journal of Vocational Education And Humanities*, *3*(2), 17–25. https://doi.org/10.53797/AJVAH.V3I2.3.2022

Mbona, R. M., & Yusheng, K. (2019). Financial statement analysis: Principal component analysis (PCA) approach case study on China telecoms industry. *Asian Journal of Accounting Research*, *4*(2), 233–245. https://doi.org/10.1108/AJAR-05-2019-0037/FULL/PDF

Misbahuddin, M., … B. B.-, Trusted, A. &, & 2017, undefined. (2017). Design of a risk based authentication system using machine learning techniques. *In 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, 1–6. https://ieeexplore.ieee.org/abstract/document/8397628/

Mohajon, J. (2020). *Confusion Matrix for Your Multi-Class Machine Learning Model | by Joydwip Mohajon | Towards Data Science*. Meduim. https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826

Mohammed, B., Hasan, S., & Mohsin Abdulazeez, A. (2021). A Review of Principal Component Analysis Algorithm for Dimensionality Reduction. *Journal of Soft Computing and Data Mining*, *2*(1), 20–30. https://doi.org/10.30880/jscdm.2021.02.01.003

Mohammed, M. M., & Elsadig, M. (2013). A multi-layer of multi factors authentication model for online banking services. *Proceedings - 2013 International Conference on Computer, Electrical and Electronics Engineering: "Research Makes a Difference", ICCEEE 2013*, 220–224. https://doi.org/10.1109/ICCEEE.2013.6633936

Mohammed, R., Rawashdeh, J., & Abdullah, M. (2020). Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results. *2020 11th International Conference on Information and Communication Systems, ICICS 2020*, 243–248. https://doi.org/10.1109/ICICS49469.2020.239556

Mohapatra, S., & Sanjay Mohapatra. (2013). E-Commerce Strategy. *Springer US*, 155–171. https://doi.org/10.1007/978-1-4614-4142-7

Montiel, J., Halford, M., Alan, alaneu, Saulo Martiello Mastelini mastelini, F., Bolmier, G., Vaysse, R., Zouitine, A., Murilo Gomes, H., Read, J., Bifet, A., Martiello Mastelini, S., Sourty, R., & Abdessalem, T. (2021). River : machine learning for streaming data in Python. *The Journal of Machine Learning Research*, *22*, 1–8. https://doi.org/10.5555/3546258.3546368

Muduroglu, E. (2023). *The Most Popular eCommerce Payment Methods In 2023 | exactly*. Exactly. https://exactly.com/blog/ecommerce-payment-methods

Nel, J. D., & Badenhorst, A. (2020). A conceptual framework for reverse logistics challenges in e-commerce. *International Journal of Business Performance Management*, *21*(1–2), 114–131. https://doi.org/10.1504/IJBPM.2020.106119

Ng, A. Y., & Jordan, M. I. (2001). On Discriminative vs. Generative Classifiers: A comparison of logistic regression and naive Bayes. *Advances in Neural Information Processing Systems*, *14*.

Nguyen, G., Dlugolinsky, S., Bobák, M., Tran, V., López García, Á., Heredia, I., Malík, P., & Hluchý, L. (2019). Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey. *Artificial Intelligence Review*, *52*(1), 77–124. https://doi.org/10.1007/S10462-018-09679-Z/TABLES/5

Niveditha, G., Abarna, K., & Akshaya, G. V. (2019). Credit Card Fraud Detection Using Random Forest Algorithm. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology © 2019 IJSRCSEIT |*, *5*(2), 2456–3307. https://doi.org/10.32628/CSEIT195261

Nowé, A., Vrancx, P., & De Hauwere, Y. M. (2012). Game theory and multi-agent reinforcement learning. *Adaptation, Learning, and Optimization*, *12*, 441–470. https://doi.org/10.1007/978-3-642-27645-3_14/COVER

Ometov, A., Bezzateev, S., Mäkitalo, N., Andreev, S., Mikkonen, T., & Koucheryavy, Y. (2018). Multi-Factor Authentication: A Survey. *Cryptography 2018, Vol. 2, Page 1*, *2*(1), 1. https://doi.org/10.3390/CRYPTOGRAPHY2010001

O'sullivan, A., & Sheffrin, S. M. (2003). *Economics: Principles in Action*.

Palekar, V., Kharade, S., Zade, H., Ali, S., Kamble, K., & Ambatkar, S. (2020). Credit Card Fraud Detection Using Isolation Forest. *International Research Journal of Engineering and Technology*. www.irjet.net

Paliouras, G., Papatheodorou, C., Karkaletsis, V., & Spyropoulos, C. D. (2002). Discovering user communities on the Internet using unsupervised machine learning techniques. *Interacting with Computers*, *14*(6), 761–791. https://doi.org/10.1016/S0953-5438(02)00015-2

Paul, H., & Nikolaev, A. (2021). Fake review detection on online E-commerce platforms: a systematic literature review. *Data Mining and Knowledge Discovery*, *35*(5), 1830–1881. https://doi.org/10.1007/S10618-021-00772-6/TABLES/3

Petrosyan, A. (2023). *Data records breached worldwide 2023 | Statista*. Statista. https://www.statista.com/statistics/1307426/number-of-data-breaches-worldwide/

Petsas, T., Tsirantonakis, G., Athanasopoulos, E., & Ioannidis, S. (2015). Two-factor authentication: Is the world ready? quantifying 2FA adoption. *Proceedings of the 8th European Workshop on System Security, EuroSec 2015*. https://doi.org/10.1145/2751323.2751327

Ping, T. (2022). *Single-factor, Two-factor, and Multi-factor Authentication | Ping Identity*. Ping Identity. https://www.pingidentity.com/en/resources/identity-fundamentals/authentication/single-factor-two-factor-multi-factor-authentication.html

Pinkston, D. A. (2018). North Korean Cyber Threats. *North Korean Cyber Threats*, 89–119.

Puh, M., & Brkić, L. (2019). Detecting credit card fraud using selected machine learning algorithms. *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2019 - Proceedings*, 1250–1255. https://doi.org/10.23919/MIPRO.2019.8757212

Rai, A. K., & Dwivedi, R. K. (2020). Fraud Detection in Credit Card Data Using Machine Learning Techniques. *Communications in Computer and Information Science*, *1241 CCIS*, 369–382. https://doi.org/10.1007/978-981-15-6318-8_31/COVER

Raschka, S. (2015). *Python Machine Learning*. Packt publishing ltd. https://books.google.ps/books?hl=ar&lr=&id=GOVOCwAAQBAJ&oi=fnd&pg=PP1&dq=pytho

n&ots=NdgEN9XRZG&sig=s7nEXwjBsC_xq5rdbK8QnSM-h6o&redir_esc=y#v=onepage&q=python&f=false

Raschka, S., Patterson, J., & Nolet, C. (2020). Machine Learning in Python: Main Developments and Technology Trends in Data Science, Machine Learning, and Artificial Intelligence. *Information 2020, Vol. 11, Page 193*, *11*(4), 193. https://doi.org/10.3390/INFO11040193

Ratnam, V., Ratnam Ganji, V., & Naga Prasad Mannem, S. (2012). Credit card fraud detection using anti-k nearest neighbor algorithm. *International Journal on Computer Science and Engineering*, *4*(6), 1035–1039. https://www.researchgate.net/publication/236962626

RB, A., & KR, S. K. (2021). Credit card fraud detection using artificial neural network. *Global Transitions Proceedings*, *2*(1), 35–41. https://doi.org/10.1016/J.GLTP.2021.01.006

Reades, J. (2020). Teaching on Jupyter. *REGION*, *7*(1), 21–34. https://doi.org/10.18335/REGION.V7I1.282

Reddy, P., Reddy, E., Viswanath, P., & Reddy, B. E. (2018). Semi-supervised learning: a brief review. *International Journal of Engineering &Technology*, *7*(1), 81–85. https://doi.org/10.14419/ijet.v7i1.8.9977

Rodrigues, A. R. L. (2023). Enhanced Multi-Factor Authentication for Mobile Applications. *University of Coimbra, Portugal, MS Thesis*. https://estudogeral.uc.pt/handle/10316/107821

Rublon, T. (2021). *What Are the Three Authentication Factors? - Rublon*. Rublon. https://rublon.com/blog/what-are-the-three-authentication-factors/

Ruppel, C., Underwood-Queen, & Harrington. (2003). e-Commerce: The Roles of Trust, Security, and Type of e-Commerce Involvement. *E-Service Journal*, *2*(2), 25. https://doi.org/10.2979/ESJ.2003.2.2.25

Sadeghi, A.-R., & Schneider, M. (2003). Electronic Payment Systems. *Digital Rights Management. Lecture Notes in Computer Science, . Springer, Berlin, Heidelberg.*, 113–137. https://doi.org/10.1007/10941270_9

Saha, P. (2021). Prediction And Detection Model Of Systemic Lupus Disease By Using Machine-Learning And Artificial Intelligence Along With Jupyter Anaconda Navigator Simulation. *Webology*, *18*(6). http://www.webology.org

Sailusha, R., Gnaneswar, V., Ramesh, R., & Ramakoteswara Rao, G. (2020). Credit Card Fraud Detection Using Machine Learning. *Proceedings of the International Conference on Intelligent Computing and Control Systems, ICICCS 2020*, 1264–1270. https://doi.org/10.1109/ICICCS48265.2020.9121114

Sanyal, S., Tiwari, A., & Sanyal, S. (2010). A multifactor secure authentication system for wireless payment. *Advanced Information and Knowledge Processing*, *53*, 341–369. https://doi.org/10.1007/978-1-84996-074-8_13/COVER

Sarkar, A., & Singh, B. K. (2020). A review on performance,security and various biometric template protection schemes for biometric authentication systems. *Multimedia Tools and Applications*, *79*(37–38), 27721–27776. https://doi.org/10.1007/S11042-020-09197-7/METRICS

Sayed, S. A., Abdel-Hamid, Y., & Hefny, H. A. (2023). Artificial intelligence-based traffic flow prediction: a comprehensive review. *Journal of Electrical Systems and Information Technology 2023 10:1*, *10*(1), 1–42. https://doi.org/10.1186/S43067-023-00081-6

Scaria, B. A., & Karman Megalingam, R. (2019). Enhanced E-Commerce Application Security Using Three-Factor Authentication. *Proceedings of the 2nd International Conference on Intelligent Computing and Control Systems, ICICCS 2018*, 1588–1591. https://doi.org/10.1109/ICCONS.2018.8662831

Schmidhuber, J. (2015). On Learning to Think: Algorithmic Information Theory for Novel Combinations of Reinforcement Learning Controllers and Recurrent Neural World Models. *ArXiv Preprint ArXiv:1511.09249* . https://arxiv.org/abs/1511.09249v1

Schneier, B. (2005). Two-factor authentication. *Communications of the ACM*, *48*(4), 136. https://doi.org/10.1145/1053291.1053327

Sensuse, D. I., Sipahutar, R. J., Jamra, R. K., Suryono, R. R., & Kautsarina. (2020). Challenges and recommended solutions for change management in indonesian e-commerce. *2020 International Conference on Information Technology Systems and Innovation, ICITSI 2020 - Proceedings*, 250–255. https://doi.org/10.1109/ICITSI50517.2020.9264950

Shah, S. U., Hadi, F. E., & Minhas, A. A. (2009). New factor of authentication: Something you process. *Proceedings - 2009 International Conference on Future Computer and Communication, ICFCC 2009*, 102–106. https://doi.org/10.1109/ICFCC.2009.79

Sharif, M. H. U., Mohammed, M. A., Sharif, M. H. U., & Mohammed, M. A. (2022). A literature review of financial losses statistics for cyber security and future trend. *Cyber Security and Future Trend*, *16*(2), 138–156. https://doi.org/10.30574/WJARR.2022.15.1.0573

Shrestha, R. B., Razavi, M., & Prasad, P. W. C. (2020). An unsupervised machine learning technique for recommendation systems. *CITISIA 2020 - IEEE Conference on Innovative Technologies in Intelligent Systems and Industrial Applications, Proceedings*. https://doi.org/10.1109/CITISIA50690.2020.9371817

Shuhaiber, A., Alhosani, S., Albadi, F., & Almarri, Q. (2022). “Sidekick” Application: A Smart Mobile Application for Generation Z. *Proceedings - 2022 9th International Conference on Wireless Networks and Mobile Communications, WINCOM 2022*. https://doi.org/10.1109/WINCOM55661.2022.9966450

Sint, K., & Kyaw, S. (2019). Analysis on the Strength and Weakness of Current Authentication Systems to Overcome Their Limitations. *International Journal of Scientific Engineering and Technology Research (IJSETR)*, 463–468. www.ijsetr.com

Sivathanu, B. (2019). Adoption of digital payment systems in the era of demonetization in India: An empirical study. *Journal of Science and Technology Policy Management*, *10*(1), 143–171. https://doi.org/10.1108/JSTPM-07-2017-0033/FULL/XML

Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation. *AAAI Workshop - Technical Report*, *WS-06-06*, 24–29. https://doi.org/10.1007/11941439_114/COVER

Song, Z., Sun, Y., Wan, J., Huang, L., & Zhu, J. (2019). Smart e-commerce systems: current status and research challenges. *Electronic Markets*, *29*(2), 221–238. https://doi.org/10.1007/S12525-017-0272-3/METRICS

Sourabh, & Arora, B. (2022). A Review of Credit Card Fraud Detection Techniques. *Lecture Notes in Electrical Engineering*, *832*, 485–496. https://doi.org/10.1007/978-981-16-8248-3_40/COVER

Stancin, I., & Jovic, A. (2019). An overview and comparison of free Python libraries for data mining and big data analysis. *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2019 - Proceedings*, 977–982. https://doi.org/10.23919/MIPRO.2019.8757088

Stylios, I., Kokolakis, S., Thanou, O., & Chatzis, S. (2022). Key factors driving the adoption of behavioral biometrics and continuous authentication technology: an empirical research. *Information and Computer Security*, *30*(4), 562–582. https://doi.org/10.1108/ICS-08-2021-0124/FULL/XML

Tanouz, D., Subramanian, R. R., Eswar, D., Reddy, G. V. P., Kumar, A. R., & Praneeth, C. H. V. N. M. (2021). Credit card fraud detection using machine learning. *Proceedings - 5th International Conference on Intelligent Computing and Control Systems, ICICCS 2021*, 967–972. https://doi.org/10.1109/ICICCS51141.2021.9432308

Tan, Z., Jamdagni, A., He, X., Nanda, P., Liu, R. P., & Hu, J. (2015). Detection of Denial-of-Service Attacks Based on Computer Vision Techniques. *IEEE Transactions on Computers*, *64*(9), 2519–2533. https://doi.org/10.1109/TC.2014.2375218

Telo, J. (2019). ANALYZING THE EFFECTIVENESS OF BEHAVIORAL BIOMETRICS IN AUTHENTICATION: A COMPREHENSIVE REVIEW. *Tensorgate Journal of Sustainable Technology and Infrastructure for Developing Countries*, *2*(1), 19–36. https://research.tensorgate.org/index.php/tjstidc/article/view/13

Tiwari, P., Mehta, S., Sakhuja, N., Kumar, J., & Singh, A. K. (2021). Credit Card Fraud Detection using Machine Learning: A Study. *ArXiv Preprint ArXiv:2108.10005*. https://arxiv.org/abs/2108.10005v1

Usama, M., Qadir, J., Raza, A., Arif, H., Yau, K. L. A., Elkhatib, Y., Hussain, A., & Al-Fuqaha, A. (2019). Unsupervised Machine Learning for Networking: Techniques, Applications and Research Challenges. *IEEE Access*, *7*, 65579–65615. https://doi.org/10.1109/ACCESS.2019.2916648

Usmani, M., Adil, S. H., Raza, K., & Ali, S. S. A. (2016). Stock market prediction using machine learning techniques. *2016 3rd International Conference on Computer and Information Sciences, ICCOINS 2016 - Proceedings*, 322–327. https://doi.org/10.1109/ICCOINS.2016.7783235

van Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. *Machine Learning*, *109*(2), 373–440. https://doi.org/10.1007/S10994-019-05855-6/FIGURES/5

Velásquez, I., Caro, A., & Rodríguez, A. (2018). Authentication schemes and methods: A systematic literature review. *Information and Software Technology*, *94*, 30–37. https://doi.org/10.1016/J.INFSOF.2017.09.012

Velazquez, P. V., Bobek, V., Vide, R. K., & Horvat, T. (2022). Lessons from Remarkable FinTech Companies for the Financial Inclusion in Peru. *Journal of Risk and Financial Management 2022, Vol. 15, Page 62*, *15*(2), 62. https://doi.org/10.3390/JRFM15020062

Wang, C., Wang, Y., Chen, Y., Liu, H., & Liu, J. (2020). User authentication on mobile devices: Approaches, threats and trends. *Computer Networks*, *170*, 107118. https://doi.org/10.1016/J.COMNET.2020.107118

Wang, D., Gu, Q., Cheng, H., & Wang, P. (2016). The request for better measurement: A comparative evaluation of two-factor authentication schemes. *ASIA CCS 2016 - Proceedings of the 11th ACM Asia Conference on Computer and Communications Security*, 475–486. https://doi.org/10.1145/2897845.2897916

Wang, D., He, D., Wang, P., & Chu, C. H. (2015). Anonymous Two-Factor Authentication in Distributed Systems: Certain Goals Are Beyond Attainment. *IEEE Transactions on Dependable and Secure Computing*, *12*(4), 428–442. https://doi.org/10.1109/TDSC.2014.2355850

Wang, D., & Wang, P. (2015). Offline dictionary attack on password authentication schemes using smart cards. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial*

Intelligence and Lecture Notes in Bioinformatics), *7807*, 221–237. https://doi.org/10.1007/978-3-319-27659-5_16/COVER

Wang, Z., Liu, K., Li, J., Zhu, Y., & Zhang, Y. (2019). Various Frameworks and Libraries of Machine Learning and Deep Learning: A Survey. *Archives of Computational Methods in Engineering*, 1–24. https://doi.org/10.1007/S11831-018-09312-W/METRICS

Wasfi, H., & Stone, R. (2023). Usability and Security of Knowledge-based Authentication Systems: A State-of-the-Art Review. *IJACSA) International Journal of Advanced Computer Science and Applications*, *14*(5). www.ijacsa.thesai.org

Xuan, S., Liu, G., Li, Z., Zheng, L., Wang, S., & Jiang, C. (2018a). Random forest for credit card fraud detection. *ICNSC 2018 - 15th IEEE International Conference on Networking, Sensing and Control*, 1–6. https://doi.org/10.1109/ICNSC.2018.8361343

Xuan, S., Liu, G., Li, Z., Zheng, L., Wang, S., & Jiang, C. (2018b). Random forest for credit card fraud detection. *ICNSC 2018 - 15th IEEE International Conference on Networking, Sensing and Control*, 1–6. https://doi.org/10.1109/ICNSC.2018.8361343

Yang, X., Song, Z., King, I., & Xu, Z. (2023). A Survey on Deep Semi-Supervised Learning. *IEEE Transactions on Knowledge and Data Engineering*, *35*(9), 8934–8954. https://doi.org/10.1109/TKDE.2022.3220219

Zadeh, M. J., & Barati, H. (2019). Security improvement in mobile baking using hybrid authentication. *ACM International Conference Proceeding Series*, 198–201. https://doi.org/10.1145/3369114.3369151

Zhang, D., Bhandari, B., & Black, D. (2020). Credit Card Fraud Detection Using Weighted Support Vector Machine. *Applied Mathematics*, *11*(12), 1275–1291. https://doi.org/10.4236/AM.2020.1112087

Zhao, S., & Hu, W. (2018). Improvement on OTP authentication and a possession-based authentication framework. *International Journal of Multimedia Intelligence and Security*, *3*(2), 187. https://doi.org/10.1504/IJMIS.2018.096406

Zviran, M., & Erlich, Z. (2006). Identification and Authentication: Technology and Implementation Issues. *Communications of the Association for Information Systems*, *17*(1), 4. https://doi.org/10.17705/1CAIS.01704

Appendix

7. Appendix

7.1. Appendix A. SMOTE

The category distribution of the skewed dataset is equalized by applying the SMOTE oversampling approach. The less frequent instances are chosen [139]. Next, at each point along the line connecting the instances, a new instance is created. In other words, the method uses K Nearest Neighbors to select a random neighbor and select a random instance from the minority class. In the feature space, the synthetic instance is built between two instances. Using SMOTE has a disadvantage in that it does not take the majority class into account when generating synthetic instances. When the classes significantly overlap with one another, this may lead to problems [189]. This study performed the SMOTE technique to balance the dataset, see Figure 7.1 which illustrates how this technique works.

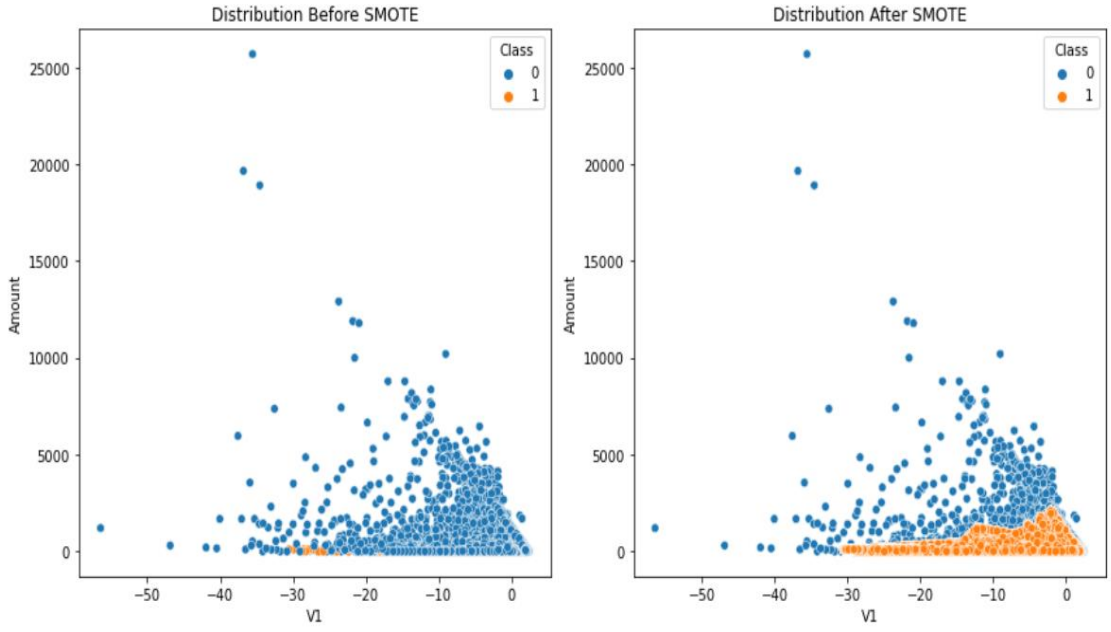

Figure 7.1. The Distribution of Data before and after Performing SMOTE

A scatter plot was used to visualize the data distribution for two features in the dataset, as seen in Figure 7.1, the fraud data “class 1” was very little before performing the SMOTE technique.

The Python code that was used to do this is provided in Figure 7.2.

```
import matplotlib.pyplot as plt
import seaborn as sns
from imblearn.over_sampling import SMOTE
import pandas as pd
# Load the dataset
df = pd.read_csv(r'C:\Users\Msys\Desktop\New folder\creditcard.csv')
# Check for duplicates and remove them
duplicate_rows = df.duplicated()
num_duplicates = duplicate_rows.sum()
print("Number of duplicate rows is:", num_duplicates)
df = df.drop_duplicates()
# Visualize the distribution before SMOTE using a scatter plot
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.title('Distribution Before SMOTE')
sns.scatterplot(x='V1', y='Amount', hue='Class', data=df)
# Select the features and target
X = df.drop('Class', axis=1)
y = df['Class']
# Create the SMOTE oversampler
oversampler = SMOTE()
# Oversample the data
X_over, y_over = oversampler.fit_resample(X, y)
# Convert the oversampled data back to a DataFrame for visualization
df_over = pd.concat([pd.DataFrame(X_over, columns=X.columns), pd.DataFrame({'Class': y_over})], axis=1)
# Visualize the distribution after SMOTE using a scatter plot
plt.subplot(1, 2, 2)
plt.title('Distribution After SMOTE')
sns.scatterplot(x='V1', y='Amount', hue='Class', data=df_over)
plt.tight_layout()
plt.show()
```

Figure 7.2. Python Code to Perform the SMOTE Oversampling

7.2. Appendix B. PCA

A statistical method called “principal component analysis (PCA)” aids the dimensionality reduction of big, complicated datasets by identifying the main elements with the most information and eliminating noise or less significant data while keeping all the essential features [190]. With this method, a big set of information is reduced in size while preserving almost all of the original data.

As reducing the size of an image removes certain features you could see in the larger version, lowering the data's variable count will certainly shrink its accuracy. The reduction of dimensions aims to simplify things by harming some degree of accuracy, Anyway, Shorter datasets are easier to interpret. This facilitates the examination of data points by ML algorithms. [191]. See Figure 7.3.

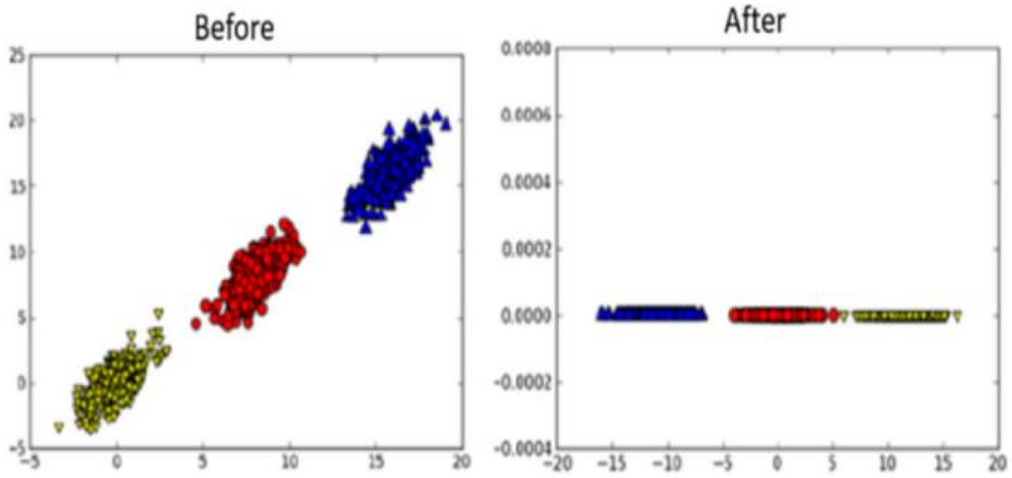

Figure 7.3. Data before and after Performing the PCA [200]

As seen in Figure 7.3, the main goal for PCA is to reduce the dimensionality by reducing the input data while preserving its main characteristics “primary component PC”. These primary components have the qualities listed below [192]:

- The linear arrangement of the initial attributes must match the primary feature.
- These components have a vertical nature. This suggests that there is no correlation between the two variables.
- Each component's importance decreases from 1 to n. This suggests that the PC with a value of "n" is the least significant and the PC with the value "1" is the most important.

The PCA can performed by applying the steps in Figure 7.4.

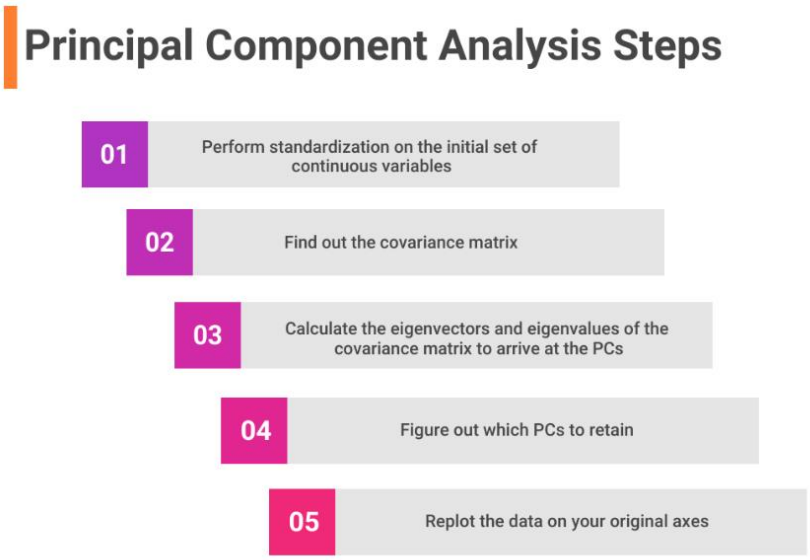

Figure 7.4. Steps to Perform the PCA Technique [201]

As shown in Figure 7.4, which illustrates how PCA works. First, before using PCA, standardization of continuous variables is essential to remove biases and variances. Subtracts each variable with the mean and divide the result by the standard deviation. The covariance matrix is then calculated to determine any redundancies and comprehend the connections between the variables. Subsequently, the principal components (PCs) of the covariance matrix are then found by computing the eigenvectors and eigenvalues of the matrix. The greatest variation in the data set is captured by these additional variables, called PCs. By classifying the eigenvectors according to their eigenvalues, the importance of every PC is determined. A feature vector is used to determine which components should be kept or discarded. Lastly, the transposition of the feature vector is multiplied by the transposition of the initial dataset to reposition the data onto the axes of the major components. Dimension reduction and a more readable data representation in a smaller feature space are made possible by this method.

PCA is utilized in many applications such as healthcare and biology, recognizing faces, financial sector, and Image compression.

However, the related application to this study is the financial sector. In a complex financial situation, PCA minimizes the number of dimensions. Assume for the moment that 200 products are included in the investment banker's portfolio. The problem is made much more complicated by the fact that a 200-by-200 correlation matrix is required to properly assess these stocks. However, PCA can help identify the 20 primary components that best characterize the volatility in the stock. This would provide a simplified solution while also providing information on the movements of all 200 stocks [193].

7.3. Appendix C. Publication in Peer-Reviewed Journal

This thesis has been published as a peer-reviewed paper in the “AI” Journal, underscoring the significance of its contributions to the field of Internet financial transaction security. The paper, titled [Secure Internet Financial Transactions: A Framework Integrating Multi-Factor Authentication and Machine Learning], digs further into the complexities of enhancing security measures in online financial transactions, a topic of increasing importance in today's digital landscape. By undergoing rigorous peer review, this publication validates the relevance and impact of the research presented in this thesis. The dissemination of this work in a respected academic journal not only advances scholarly discourse but also contributes to practical solutions for addressing the evolving challenges of cyber threats in financial services.

Interested readers can access the published paper online via its DOI:

[https://doi.org/10.3390/ai5010

الملخص

لقد برز أمن المعاملات المالية عبر الإنترنت باعتباره مصدر قلق بالغ في عصر أصبحت فيه الخدمات المالية رقمية بشكل متزايد. أدى الاستخدام المتزايد لمنصات الخدمات المصرفية والمدفوعات الرقمية إلى ظهور موجة جديدة من الفرص لكل من المستخدمين ومجرمي الإنترنت. من أجل معالجة هذه المشكلة، يقدم البحث الحالي نظامًا فريدًا يدمج التعلم الآلي (ML) مع المصادقة متعددة العوامل (MFA). يعتمد النظام المقترح على استخدام مستويين من الحماية بحيث تستخدم الطبقة الأولى عاملين اثنين للمصادقة، والطبقة الثانية عبارة عن طبقة مضمنة تطلب التعرف على وجه المستخدم لمواصلة عملية الشراء بنجاح إذا قرر نموذج التعلم الآلي أن المعاملة الحالية احتيالية.

لبناء نموذج التعلم الآلي، تم وضع أربعة مصنفات للتعلم الآلي موضع التنفيذ وهي: الغابات العشوائية(RF) ، وأشجار القرار(DT) ، والانحدار اللوجستي (LR) والمصنف البايزي (NB) ، بلغت دقة كل منها 96.717%، و97.881%، و97.938%، و92.354%، على التوالي. كما تم إنشاء شاشات لتطبيق تجارة إلكترونية لنظام أندرويد لتوضيح مبدأ عمل النظام.

إن إطار العمل المقترح في هذه الدراسة يمكن تطويره للعمل على أي منصة تجارة إلكترونية رقمية. بعد الدراسة والتحليل لمجموعة الأبحاث حول هذا الموضوع والأساليب المختلفة لتأمين المعاملات عبر الإنترنت تبين أن تكامل المصادقة متعددة العوامل وتعلم الآلة بإمكانه توفير أعلى مستوى من الحماية بالاضافة الى توفير أنظمة سهلة الاستخدام.

في الأبحاث المستقبلية، قد يكون من المفيد فحص عوامل المصادقة الأخرى باستخدام مجموعة بيانات مختلفة.