**CHAPTER ONE**

**INTRODUCTION**

* 1. **BACKGROUND OF STUDY**

Credit card fraud is a significant problem for financial institutions and customers worldwide. Scammers are constantly finding new ways to exploit weaknesses in the payment processing system as more people use credit cards for financial transactions. Credit card theft can have serious consequences, including significant financial losses for individuals and businesses, decreased customer confidence, and damage to financial institutions' reputations.

Traditional fraud detection methods, such as rule-based algorithms and manual verification, are ineffective in addressing the complexity and scope of modern credit card fraud. Manual verification is time-consuming and impractical for real-time detection, while rule-based algorithms struggle to adapt to new fraud trends.

Fortunately, machine learning has emerged as a viable solution to overcome the challenges of credit card fraud detection. Machine learning algorithms can be trained using large datasets, enabling them to identify complex patterns and adjust to evolving fraud schemes. These algorithms have proven successful in various fields, such as image recognition, natural language processing, and recommendation systems, making them suitable for credit card fraud detection.

A cutting-edge machine learning algorithm will power the proposed Credit Card Fraud Detection System, providing a reliable and effective solution. This technology uses real-time transaction data to identify any fraudulent actions quickly, detecting irregularities with high accuracy through feature engineering, model creation, and data preparation. By reducing false positives, the system aims to improve fraud detection accuracy and enhance client confidence, ensuring financial stability and profitability for financial institutions. Furthermore, the system can promote the wider adoption of cashless payment systems, resulting in broader economic advantages. Through thorough analysis, careful experimentation, and evaluation, this research aims to develop credit card fraud detection methods that will become a crucial tool in securing financial transactions, protecting customers, and maintaining the integrity of the financial ecosystem.

**1.2 STATEMENT OF THE PROBLEM**

Due to the challenges that financial institutions encounter when dealing with credit card fraud by their clients.

**1. 3 AIM AND OBJECTIVES OF A CREDIT CARD DETECTION SYSTEM**

**AIM:**

The purpose of a Credit Card Fraud Detection System is to create a comprehensive and efficient solution that can swiftly and accurately detect and prevent fraudulent credit card transactions. The system aims to enhance security measures, protect cardholders, and minimize financial losses caused by credit card theft through the use of advanced machine learning algorithms and data processing methods.

**OBJECTIVES:**

* + - 1. Develop a system which can quickly analyze credit card transactions in real-time and identify any potential fraudulent activities.

1. One task is to develop machine learning models that can precisely distinguish between legitimate and fraudulent transactions.
2. Developing and Evaluating Models: You can train and improve machine learning models like logistic regression, decision trees, random forests, support vector machines, or deep learning models depending on the data and problem you are working with.
3. Achieving High Accuracy with Low False Positives.
4. Implement efficient data preprocessing methods to clean, transform, and ready the transaction data for model training.
5. Data Preprocessing and Imbalance Handling.
6. Feature Engineering and Selection: Find pertinent features in the transaction data that can offer helpful information for fraud detection.
7. To ensure a complete fraud detection solution, seamlessly integrate the system with current payment processing systems during deployment and integration.

**1.4 SIGNIFICANCE OF THE STUDY**

A credit card fraud detection system is incredibly important for safeguarding financial transactions

and protecting the interests of both cardholders and financial institutions. Its significance lies in several key areas:

1. It's important for financial institutions to follow a set of guidelines when it comes to preventing fraud and ensuring the safety of their clients. These regulations help to protect both the institution and its customers from potential harm.

2. Dealing with Evolving Theft Techniques: As time goes on, the tactics utilized to commit credit card theft are constantly evolving and becoming more intricate.

3. Credit card fraud often involves identity theft, where thieves steal personal information to carry out fraudulent transactions. This highlights the importance of preventing identity theft to protect individuals from financial harm.

4. A credit card fraud detection system protects finances by preventing fraudulent transactions.

5. When financial institutions provide reliable security measures, cardholders feel more comfortable using credit cards and making online transactions. Building trust and confidence in the safety of these transactions is important for encouraging their use..

6. The system effectively minimizes financial losses by detecting and preventing fraudulent transactions in real-time. This helps protect financial institutions, businesses, and customers from any adverse impact.

7. It's important for financial organizations to have robust fraud detection systems in place in order to maintain a positive reputation with customers.

8. Financial institutions have a crucial responsibility to protect their customers from fraud and ensure compliance with regulations. There are several rules in place that guide their operations and ensure that they are acting in the best interests of their clients. By following these guidelines, financial institutions can maintain the trust of their customers and help to create a safer and more secure financial environment for everyone.

**1.5 SCOPE OF THE STUDY**

The scope of this study is to create a credit card fraud detection system using a machine learning model. This involves analyzing various aspects of the system's development and evaluation. The study includes a comprehensive review of the methods and tools used in the financial industry to detect and prevent unauthorized credit card transactions.

**1.6 LIMITATIONS OF THE STUDY**

To improve the Credit Card Fraud Detection System, we need to understand its limitations. These include:

1. The system's accuracy and ability to work with real-world fraud events may be affected by limited access to transaction data or data that doesn't cover enough fraudulent events.
2. Explaining how a specific transaction was flagged as fraudulent can be challenging because it's hard to understand the decision-making process of the model.
3. Credit card fraud datasets often have a class imbalance issue where there are significantly more genuine transactions than fraudulent ones. This leads to unbalanced data.
4. Generalization: The system might perform well on the training and test datasets, but its performance in the real world on unused data from various sources may differ.
5. Latency in Real-Time Processing: Quick processing of high-speed transaction data is necessary for real-time credit card fraud detection.
6. When it comes to credit card fraud detection systems, financial institutions, particularly the smaller ones with limited resources, may have to bear substantial costs for implementation and maintenance.

**1.7 DEFINITION OF TERMS USED IN THE PROJECT**

Below are the definitions of important terminology used in the project description pertaining to the Credit Card Fraud Detection System:

1. **Precision:** A measure of correctly identified fraud among all positive predictions made by the model.

2 **F1-Score:** A single value is necessary to evaluate a model's overall performance, which balances both precision and recall metrics.

3. **Model development:** The process of creating and refining machine learning models involves using data and algorithms to generate predictions or judgments based on new and unexpected information.

4. **Machine Learning:** This is a type of artificial intelligence (AI) that allows computers to enhance their performance in a specific task without the need for explicit instruction.

5. **Data preprocessing:** In order to ensure accuracy and efficiency, the raw data is first cleaned, transformed, and prepared before it is fed into the machine learning models.

6. **Credit Card Fraud Detection System:** Our system uses advanced machine learning algorithms and data processing methods to swiftly and accurately detect and prevent fraudulent credit card transactions in real-time.

7**. Real-Time Processing**: A system's capacity to swiftly process and scrutinize data as it is received, leading to prompt responses and actions based on the most up-to-date information.

8. **Imbalanced Data:** In a dataset, there may be a significant difference in the number of instances between two classes, such as legitimate transactions and fraudulent transactions. This situation is commonly referred to as class imbalance.

**CHAPTER TWO**

**LITERATURE REVIEW**

**2.1 ARTIFICIAL AND COMPUTATIONAL INTELLIGENCE**

Detecting credit card fraud is a difficult task, which is why artificial and computational intelligence were created. These technologies enable robots to perform tasks that humans excel at. Computational intelligence is a subset of AI that is used in dynamic contexts. To detect credit card fraud, there are several recommended methods, including rule induction approaches, decision trees, neural networks, Support Vector Machines, logistic regression, and meta-heuristics. Various methods can be used to identify credit card fraud, including a real-time framework that uses self-organizing maps for outlier analysis and a predictive algorithm for abnormal-looking transactions. However, there are some issues related to misclassifications and detecting fraud on cards with different limits. To address these problems, the MBO technique and the GASS algorithm can be used. The GASS algorithm combines genetic algorithms and scatter search to improve results.

**2.2 CREDIT CARD FRAUD TYPES**

1. **Application Fraud:** Fraudulent activities involving the use of stolen or falsified papers to create an account in someone else's name are known as application fraud. Criminals may resort to stealing or fabricating papers such as utility bills and bank records to establish a personal profile. The fraudster can then use the fraudulent or stolen documents to withdraw cash or obtain credit under the victim's name. Application fraud may also involve the use of a synthetic identity, which is a fictitious identity created by combining personal information from multiple sources. Once the fraudster has successfully established their identity and account, they have various options for exploiting the bank. One common tactic is to maximize their credit card spending by making large purchases with the new credit card and reselling the items for cash. (Gao, 2019)
2. **Account Take Over:** Account takeover is when someone tries to gain access to a customer's account, such as credit cards, email, banks, or SIM cards. This can lead to significant profits for fraudsters. To reduce risks, it is important to use risk-based authentication (RBA), as recommended by Forrester.
3. **Social Engineering Fraud:** Social engineering fraud is when a criminal pretends to be someone else in order to trick people into giving them money or information. Nowadays, fraudsters are using more sophisticated methods to deceive individuals and companies. One common tactic is to send fake emails that appear to be from a high-ranking employee, asking colleagues to transfer money to a fraudulent bank account. (Abhimanyu, 2016)

To obtain personal information, fraudsters may use various strategies, such as impersonating a bank or payment processor. The most commonly used social engineering method is telephone phishing, where the perpetrator tries to gain the victim's trust over the phone..

1. **Skimming**: Green plastic item used in an ATM slot to prevent fraudsters from putting a skimmer device Skimming is the stealing of personal information from an otherwise legitimate transaction. The thief can obtain a victim's card number by simple ways such as duplicating receipts or through more sophisticated methods such as employing a tiny electronic device (skimmer) to swipe and store hundreds of victims' card information. Taxis, restaurants, and bars are common places for skimming since the skimmer has the victim's payment card out of sight.
2. **Phishing**: One of the most common ways to steal personal information is through phishing. This is a type of cyber attack where the attacker pretends to be a trustworthy person, organization, or business in order to trick the victim into opening a message or carrying out a specific request. Often, the victim will receive an email or text message offering something they might want or need, with the hope that they will click on a link or download an attachment. During the COVID-19 pandemic, phishing attacks became more frequent as people spent more time online. In fact, researchers found that COVID-19 phishing attacks increased by 667% in the first few months of the pandemic. (Gamini , 2021)

**2.3 REVIEW OF RELATED WORKS**

A method that uses machine learning to detect credit card fraud was proposed by an ML Engineer using AdaBoost (Kuldeep, 2015). The technique involves using standard models, as well as hybrid models that utilize AdaBoost and majority voting methods. To evaluate the efficiency of the model, a publicly available data set was used, as well as a data set from a financial institution to analyze fraud. Noise was added to the data sample to measure the robustness of the algorithms. The experiments were conducted based on theoretical results, which showed that majority voting methods achieve good accuracy rates in detecting credit card fraud. Further evaluation of the hybrid model involved adding 10% and 30% noise to the sample data. Several voting methods achieved a score of 0.942 with 30% added noise, demonstrating that the voting method is more stable in the presence of noise. Therefore, it was concluded that the voting method is a reliable way to detect fraud in credit cards.

**(Zahra, 2016)** Proposed Deep autoencoder which is used to extract the best characteristics of the information from the credit card transaction. This will further add Softmax software to resolve the class label issues. An overcomplete autoencoder is used to map the data into a high dimensional space and a sparse model was used in a descriptive manner which provides benefits for the classification of a type of fraud. Deep learning is one of the most motivated and powerful techniques being employed for the detection of fraud in credit cards. These types of networks have a complex distribution of data which is very difficult to recognize. Deep autoencoder has been used in some stages to extract the best features of the data and for classification purposes. Also, higher accuracy and low variance are achieved within these networks. (Karthik, 2019)

Deep learning topologies for detecting fraud in online money transactions were proposed which became very active for some years to help facilitate security in Credit Cards. Abhimanyu (2016) This approach is derived from the artificial neural network with in-built time and memory components like long-term short-term memory and several other parameters. According to the efficiency of these components in fraud detection, almost 80 million online transactions through credit cards have been pre-labeled as fraudulent and legal. They have used a high-performance distributed cloud computing environment. The study proposed by the researchers provides an effective guide to the sensitivity analysis of the proposed parameters as per the performance of fraud detection. The researchers also proposed a framework for the parameter tuning of Deep Learning topologies for the detection of fraud. This enables the financial institution to decrease losses by avoiding fraudulent activities (Stempel, 2020).

**(Krishna, 2014)** Investigated several techniques that were used for detecting fraudulent transactions and provided a comparative study among them. Fraudulent transactions can be detected by utilizing either one of these or integrating any of these methods. The model can be trained more accurately by adding new features. Several data mining techniques are being used by banks and credit card companies for detecting fraud behaviors. The normal usage pattern of clients depending upon their past activities can be identified by applying any of these methods. Therefore, a comparative analysis is made here by studying different fraud detection techniques proposed over the years.

In their study, (Shiyang, 2018) employed two types of random forests to train normal and abnormal transaction behavior features. The researcher conducted a comparison of these two random forests based on their classifiers and performance in detecting credit card fraud. The data used in the study was sourced from an e-commerce company in China, and was used to evaluate the performance of these two types of random forest models. The study utilized a B2C dataset to identify and detect credit card fraud. The researcher's conclusion was that the proposed random forests produced good results on small datasets, but issues like imbalanced data made the model less effective than other datasets.

In 2018, John proposed a study to evaluate the performance of various algorithms when applied to highly skewed credit card fraud data. The dataset was created using 284,807 transactions made by European cardholders. A hybrid approach of under-sampling and oversampling was used on the skewed data. Three different techniques were applied to raw and preprocessed data in Python, and their performances were evaluated. Results showed that k-NN outperformed naïve Bayes and logistic regression approaches.

In 2015, Sharmistha conducted a study on the common crimes that occur in credit card applications. The study revealed that existing non-data mining methods are insufficient in preventing identity theft. To address this issue, a new data mining defense layer was proposed. The defense layer includes two algorithms: Communal Detection and Spike Detection, which can detect fraudulent activities in various applications.

The CD and SD algorithms can search through a large moving window, higher numbers of attributes, and various link types to generate results. However, since the system takes a significant amount of time to generate results, attackers do not have time to modify their behaviors in response to the algorithms. As a result, it is difficult to evaluate the true effectiveness of the algorithms.

To address this issue, future work can enhance the proposed algorithm to improve its adaptability and effectiveness.

In 2015, Dasgri Proposed a novel mechanism using which the payment of an invoice or bill is initiated. This approach is named as „NoCash‟ mobile application which is mainly used by the merchants through which the payment facility of clients can be eased. There is no need for NFC-Enabled Point of Sales (PoS) Machines in this approach and only mobile

phones are required. Minimizing the burden of clients for bringing cards when outside, by providing easy payment transferring mechanisms is the only aim for which this system is designed. The client‟s experience of shopping is improved when the NoCash application which includes many features is applied based on the increase in the number of NFC-based mobiles. To provide benefits to merchants, fraud activities are minimized using this proposed application. The application clients can be related to the expense history and minimize any unwanted costs using this proposed method.

In 2016, Delip Presented the evaluation of the performance of several sampling techniques on the classifier when they are applied to a credit card fraud data set with the class imbalance. The principal component analysis (PCA) is applied to real data as well as the variables time, amount, and class to achieve 28 principal components that are included within the data. There are ten thousand, fifteen thousand, and twenty thousand instances available respectively within the

three datasets. This approach applied five over-sampling and four under-sampling approaches. Further, on the data, few cost-sensitive and ensemble classifiers are applied.

In 2015, Luis proposed a method to improve credit card fraud detection performance by utilizing various signal processing techniques. The proposed method includes a modified version of the traditional iterative amplitude-adjusted Fourier transform (IAAFT) and the iterative surrogate signals on graph algorithms (ISSG). The main objective of this approach is to enhance detector training by generating surrogate samples from original fraud samples and reducing the estimate variance to improve detector performance. Since data streams constantly change and present various issues, it is crucial to provide a reliable augmentation of the scarce target population of frauds. In this experiment, real data was used to showcase the effectiveness of the proposed methods by measuring their performance using ROC curves and KPIs commonly used in the financial industry.

**2.4** **IMPLEMENTING REGULATIONS AND GOVERNANCE MEASURES TO COMBAT CREDIT CARD FRAUD**

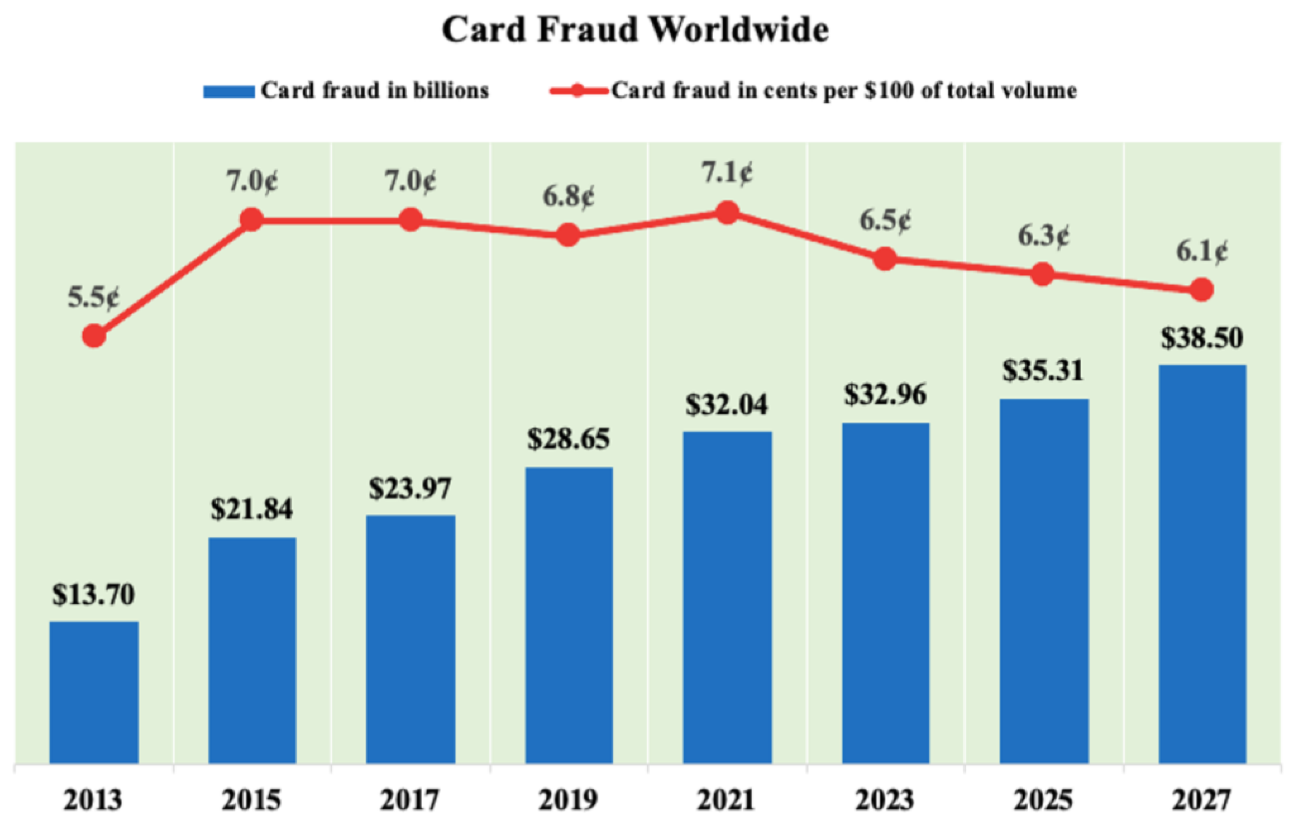
1. In September 2014, the US Department of Justice announced its plan to introduce stricter laws to combat foreign credit card trafficking. The current legislation is considered inadequate since it allows individuals in other countries to avoid punishment if they obtain and sell data outside of the United States without conducting illegal activities within the country. The Department of Justice is urging the United States Congress to revise the existing laws to make it unlawful for international criminals to hold, obtain, or sell a stolen credit card issued by a US bank, regardless of their location.

2. In the United Kingdom, credit cards are governed by the Consumer Credit Act 1974, which was amended in 2006. This act provides various protections and requirements, including the obligation for merchants or card issuers to refund any misuse of the card by the cardholder, unless the cardholder acted criminally. The regulation of banks in the UK is carried out by the Bank of England (BoE), the Prudential Regulation Authority (PRA) (a division of the BoE), and the Financial Conduct Authority (FCA), which oversees the day-to-day operations. There is no specific law or regulation that governs the credit card industry, but the Association for Payment Clearing Services (APACS) is the institution that all settlement members belong to. This organization operates under the Banking Consolidation Directive to monitor and regulate transactions. UK Finance is the association for the UK's banking and financial services sector, representing over 250 firms providing credit, banking, and payment-related services.

**3.** In Australia, the concept of identity encompasses both the identification of living or deceased individuals and the identification of corporate entities.

The following covers:

1. Identity fabrication is the process of creating a false identity.
2. Identity manipulation is the process of changing one's own identity.
3. Identity theft is defined as the theft or assumption of a pre-existing identity (or a major portion thereof), with or without authorization, and whether the subject is alive or deceased.
4. Identity crime is a broad phrase that refers to activities/crimes in which a perpetrator utilizes a manufactured identity, a manipulated identity, or a stolen/assumed identity to commit a crime(s).



**Fig 2.1:** A chart the statistics of credit card fraud worldwide.

**Source**: <https://www.ag.gov.au/RightsAndProtections/IdentitySecurity/Documents/Identity-Crime-and-Misuse-in-Australia-2013-14.pdf>

**2.5 TYPES OF CREDIT CARD ATTACK**

Credit card attacks, also known as credit card fraud, encompass various methods used by malicious actors to gain unauthorized access to credit card information or conduct fraudulent transactions. Some common types of credit card attacks include:

1. Card Skimming: In card skimming attacks, criminals place a small device, known as a skimmer, on point-of-sale (POS) machines, ATMs, or gas pumps to capture the magnetic stripe data from credit cards as they are swiped. This information is then used to create cloned cards or make unauthorized transactions.

2. Phishing: Phishing attacks involve sending fraudulent emails, messages, or websites that impersonate legitimate entities, such as banks or online retailers, to trick users into revealing their credit card information or login credentials. These emails often contain malicious links or attachments that lead to fake websites designed to steal sensitive data.

3. Carding: Carding is a type of credit card fraud where attackers use stolen credit card information to make small, inconspicuous transactions to test the validity of the card before attempting larger purchases.

4. Card Not Present (CNP) Fraud: In CNP fraud, criminals use stolen credit card information to make online or phone transactions where a physical card is not required. Since the card is not present during the transaction, it becomes more challenging to verify its legitimacy, making CNP fraud prevalent in e-commerce and remote payment scenarios.

5. Carding Bots: Cybercriminals use automated bots to perform large-scale testing of stolen credit card information across multiple websites to identify valid cards and make unauthorized purchases.

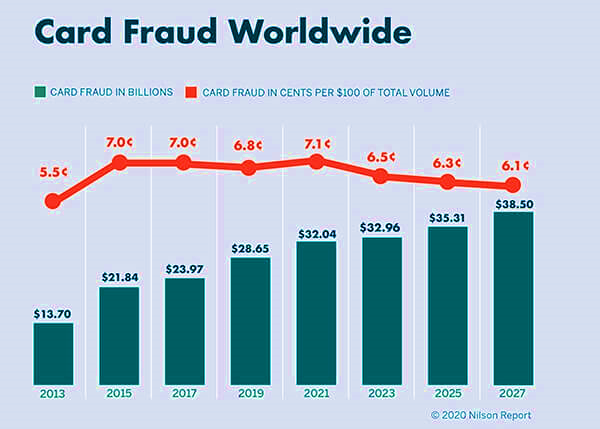
6. Account Takeover: In an account takeover, attackers gain unauthorized access to a user's online account associated with a credit card, either by guessing weak passwords or using stolen credentials obtained through data breaches or phishing attacks. Once inside the account, they can make unauthorized transactions or change account information.

7. Card Not Received (CNR) Fraud: CNR fraud occurs when criminals intercept or steal newly issued credit cards before the legitimate cardholder receives them. The attackers then use the stolen cards for fraudulent transactions.

8. Lost or Stolen Card Fraud: This type of fraud occurs when a legitimate credit card is physically lost or stolen, and the thief uses it to make unauthorized purchases before the cardholder can report the incident.

9. Card Present (CP) Fraud: CP fraud happens when an attacker uses a stolen credit card to make purchases in person, typically by presenting a counterfeit or cloned card, or by providing the card's details manually at a physical point of sale.

10. Man-in-the-Middle (MITM) Attacks: In MITM attacks, criminals intercept communication between the cardholder and a legitimate merchant during an online transaction, enabling them to steal credit card information and perform unauthorized transactions.



**Fig 2.1** Famous Credit Card Fraud Statistics

**Source**:https://www.google.com/url?sa=i&url=https%3A%2F%2Fsdk.finance%2Fcredit-card-fraud-detection-big-players-

**2.6 DISPARITIES AND ETHICAL DILEMMAS IN CREDIT CARD FRAUD**

**Generation Differences**

Millennials are the most common victims of all types of fraud, including credit and debit card fraud, digital wallet and payment fraud, banking, and tax fraud. They are followed by the GenXers and subsequently the GenZers.

Millennials spend the most time of any generation trying to recover money lost due to fraudulent charges, challenging fraudulent charges, and reviewing accounts for fraudulent or odd behavior. GenZers experienced fraud most often through digital payment apps such as PayPal, Venmo, and Square. The other generations experienced most of their issues through credit card fraud. Baby Boomers were found to have the lowest instances of fraudulent charges, and also spent the least amount of time trying to recover money due to fraudulent charges or to dispute these charges.

**2.7 FRAUDULENT CHARGEBACK**

Fraudulent fraud occurs when a client uses their credit card to make an online purchase and then requests a chargeback from the issuing bank after getting the purchased goods or services. When a chargeback is accepted, the financial transaction is canceled and the customer is reimbursed for the money spent. When a chargeback occurs, depending on the payment method used, the seller may be held accountable.

On the Internet, friendly fraud is ubiquitous, affecting both actual goods purchases and digital transactions. To prevent digital transaction fraud, prepaid cards have been presented as a possible solution to ensure customer payment.

In 2003, an Internet merchant sued MasterCard for having credit card laws and fees that made Internet businesses particularly vulnerable to friendly fraud. When a fraudulent transaction occurs, such as friendly fraud, Internet sellers may face a significant amount of the losses. A new sort of friendly fraud has been reported in Europe in recent years, involving bank transfers rather than credit card payments. One method of combating friendly fraud is to include a feature in the product that checks in with the merchant's database. If a chargeback is filed, the merchant can direct the product to suspend service. This method applies to digital subscription services as well as any other online product that requires upgrades or logins.. Because the merchant is normally paid a fee for incurring a chargeback, this is not a comprehensive solution.

**2.7.1 Merchants' costs**

Merchants must pay a chargeback fee regardless of the outcome of the dispute, which normally runs from $20 to $100. According to a LexisNexis survey from 2016, chargeback fraud costs merchants $2.40 for every $1 lost. This is due to product losses, banking fees and penalties, and administrative expenses. According to a 2018 Aite Group report on chargeback charges, U.S. CNP fraud losses in 2017 were $4 billion, with a projected increase to $6.4 billion by 2020.

**2.7.2 Electronic Transaction**

Electronic commerce: Friendly fraud thrives in the digital products market, where fraudsters have an advantage. Pornography and gambling sites are frequently targeted. It is difficult for the merchant to prove that the consumer received the items or services bought. Again, the use of card security codes can demonstrate that the cardholder (or, in the case of the three-digit security codes written on the backs of US credit cards, someone with physical possession of the card or at least knowledge of the number and the code) was present, but entering a security code at purchase does not prove that delivery was made, especially for online or over-the-phone purchases where shipping occurs after the contract is finalized.

**CHAPTER THREE**

**RESEARCH METHODOLOGY**

This research focuses on the Credit Card Fraud Detection System, a collection of methods and tactics designed to detect fraud in a dataset obtained. This chapter also discusses the data gathering process, model selection, and methodology employed. This is performed by ensuring that the transaction is genuine and that you are working with the proper cardholder.

**3.1 METHODOLOGY OF EXISTING SYSTEM**

In compared to more modern and advanced fraud detection systems such as a Machine Learning Model System, prior ways of detecting credit card theft were typically less successful and had a number of disadvantages. One or more of the major disadvantages is:

1. **Manual By Hand:** In old fraud detection systems, analysts had to manually assess transactions to check for any fraud tendencies. This approach was not scalable for managing big transaction volumes, and it was time-consuming and error-prone.
2. **Inadequate real-time data analysis:** Traditional approaches typically lacked the power to analyse massive amounts of data. Because they were ill-equipped to manage the massive data streams generated by Internet transactions, it was impossible to detect fraudulent tendencies quickly and properly.
3. **Excessive false-positive rates:** Due to the limitations of rule-based systems and less complex algorithms, older fraud detection approaches produced more false positives. This resulted in increased operational expenditures, as well as unnecessary annoyance for real customers whose transactions were incorrectly flagged as fraudulent.
4. **Inability to recognize new fraud trends:** As fraudsters' strategies continued to change, more traditional techniques found it difficult to recognize new fraud patterns. They frequently lacked the capacity to take fresh data into account while modifying their detection methods.
5. **Reactive strategy:** The traditional methods were largely reactive in nature, emphasizing fraud detection after it had already happened. Due to this reactive strategy, it was possible for businesses and customers to sustain financial losses as a result of fraudulent transactions that went unnoticed until they were reported by customers.
6. **Poor consumer profiling:** Traditional methodologies' use of restricted client profiles hampered the capacity to properly spot unusual behavior. As a result, they may miss fraud patterns that differ only slightly from a customer's typical transactional behavior.
7. **Inadequate real-time monitoring:** The identification of fraud is primarily reliant on timeliness. Older systems lacked real-time monitoring capabilities, making it more difficult to detect and respond to fraudulent behavior.

**3.2 AREA AND POPULATION OF THE STUDY**

Credit card fraud detection systems can be applied to a wide range of companies and areas where credit card transactions are common. These systems are commonly used in the following areas:

1. Financial organizations: Banks, credit card issuers, and other financial institutions use fraud detection systems to protect their customers' accounts and halt unauthorized credit card transactions.
2. Point of Sale (POS) Systems: Retail establishments and companies that accept credit cards in person frequently include fraud detection features in their POS systems.
3. Tourism and Hospitality: To stop fraudulent transactions and safeguard consumer data, airlines, hotels, and other travel-related firms use fraud detection systems.
4. Telecommunications: To identify illegal usage of or fraudulent actions involving mobile payments, telecommunication businesses that provide mobile services and subscriptions utilize fraud detection technologies.
5. Online shopping: Online merchants and e-commerce platforms employ fraud detection systems to identify and prevent fraudulent conduct that takes place during online purchases.
6. Payment Gateways: Companies that process credit card transactions on behalf of retailers use fraud detection software to keep their payment networks safe from fraudulent transactions.

**3.3 OVERVIEW PROPOSED METHODOLOGY**

The proposed approach strives at developing a system that recognizes financial fraud.

**BENEFITS OF THE PROPOSED SYSTEM**

Machine learning (ML) approaches for detecting credit card fraud offer multiple significant benefits over human or traditional rule-based systems. The following are some of the primary advantages:

1. Behavioral Analysis: Machine Learning allows for the creation of detailed customer profiles, enabling the system to identify unusual behavior and detect fraudulent activities based on deviations from normal spending patterns.
2. Improved Accuracy: Machine learning algorithms can analyze massive amounts of transaction data and continually learn from patterns and irregularities. This allows them to detect fraudulent behavior with more precision than static rule-based systems.
3. Flexibility in the Face of New Fraud Patterns: Machine learning algorithms can adapt to new and evolving fraud patterns without the need for manual upgrades. When fraudsters change their approach, the system may constantly learn and improve its detection techniques. False Positives: Because ML algorithms are better at distinguishing between legitimate and fraudulent transactions, there are fewer false positives. Customers will have a better user experience since legitimate transactions are less likely to be recorded incorrectly.
4. Analysis of Large-Scale Data: ML algorithms can efficiently handle and evaluate large amounts of transaction data. This scalability is critical in dealing with the huge number of credit card transactions that occur every day.
5. Increased Fraud Prevention: ML-powered systems may detect fraud early in the transaction process, allowing fraudulent transactions to be avoided, decreasing the effect of fraud on both enterprises and customers.
6. Costs-Effectiveness: While initial development and installation may necessitate an investment, ML-based solutions can result in long-term cost benefits. Businesses can reduce possible financial losses and operational expenditures connected with fraud investigations by avoiding fraud and lowering false positives.
7. Continual Learning: ML systems may learn from fresh data indefinitely, allowing them to keep up with the newest fraud trends and improve their detection skills over time.

**3.3 METHOD OF DATA COLLECTION**

Credit card fraud detection systems use a variety of data collecting methods to detect and prevent fraudulent transactions. To make appropriate judgments, these systems often use a combination of past transaction data and extra contextual information.

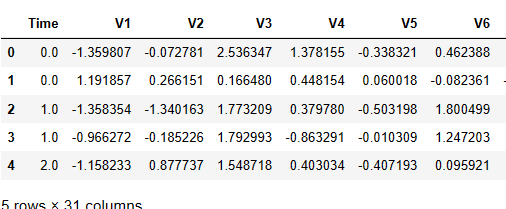
The dataset was gathered from https://kaggle.com, an open data warehouse that covers credit card transactions done by European cardholders in September 2013.

This dataset contains 492 frauds out of 284,807 transactions that happened over the course of two days. The dataset is very uneven, with positive transactions accounting for 0.172% of all transactions. It only has numerical input variables that are the outcome of a PCA transformation. We are unable to give the original features and further background information about the data owing to confidentiality concerns. The main components derived with PCA are features V1, V2,... V28; the only features that have not been changed with PCA are 'Time' and 'Amount'. Feature 'Time' is the number of seconds that have passed between each transaction and the first transaction in the dataset. The feature 'Amount' represents the transaction Amount; this feature may be utilized for cost-sensitive learning, for example. Feature 'Class' is the answer variable, with a value of 1 indicating fraud and 0 otherwise.

**3.4 METHOD OF DATA PRE - PROCESSING**

Creating a data-driven model that identifies fraudulent transactions is the first step in designing a credit card fraud detection system using machine learning. The following is a system design overview for a machine learning-based credit card fraud detection system:

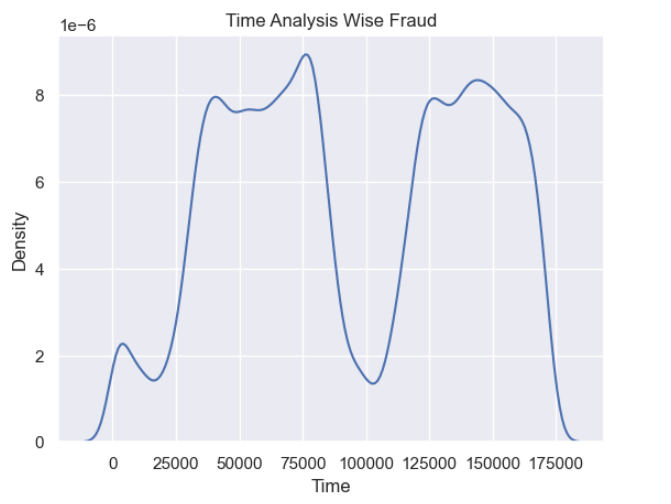
1. Data Collection: Gather historical transaction data, including both valid and fraudulent transactions, as well as pertinent information such as transaction amount, location, timestamp, cardholder information, and so on.

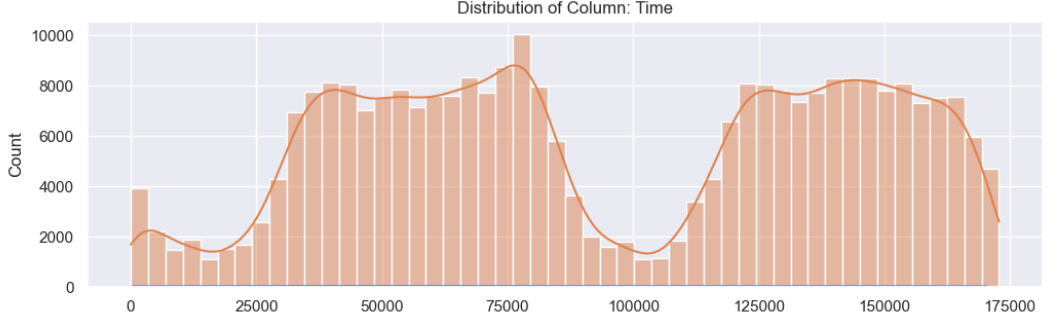


**Fig 3.1:** Showing Data for credit card fraud extracted from Kaggle

**Source:** Extracted from the project

2. Data Preprocessing: Clean up the data, deal with missing values, and eliminate duplicates or unnecessary characteristics. Normalize or scale the characteristics to ensure that machine learning algorithms can efficiently handle them.



****

**Fig 3.2** Time-Wise Fraud Analysis

**3.5 FEATURE EXTRACTION**

Feature Extraction or Feature Engineering Create new relevant features that can give more information for fraud detection. After carefully reviewing the complete dataset, the following characteristics were carefully picked for the project: Time, Amount, and Class. This dataset's Time is measured in seconds from the first transaction logged, meaning that it contains two days' worth of transactions. Because the features were created using PCA, the physical significance of each individual feature is unimportant. PCA does not modify 'Time' or 'Amount'. The response variable is the 'Class' feature, which has a value of 1 if the transaction is fraudulent and 0 if it is not. The information is classified as label and target data.

**3.6 MODEL FORMATION OF THE MODEL**

Model Selection: the best machine learning algorithms for fraud detection based on the

characteristics of the data and the problem at hand include Support Vector Classifier, Random Forest, and K-Fold Model for correcting the data biases.

**Random Forest Classifier:** The Random Forest Classifier is a commonly used machine learning algorithm that performs both classification and regression tasks. It utilizes an ensemble learning method, which combines many decision trees to enhance accuracy and prevent the overfitting of the data. In this case, 100 decision trees were employed to assess accuracy and make predictions.

**Support Vector Classifier (SVC):** The Support Vector Classifier (SVC) was used to detect data fraud. Credit card fraud detection frequently includes maintaining complicated and imbalanced information, making SVC in such cases less efficient and effective. Other machine learning techniques, such as Random Forest and K-fold Classifier, were utilized inaddition**.**

**3.7 METHOD OF MODEL TRAINING AND EVALUATION PROCESS**

**Model Education**

For model training and assessment, the data was divided into training and testing sets. Using the identified fraud indicators, train the specified machine learning models using historical data. 70% for the training and 20% for the exam size.

**Model Assessment:**

Evaluate the efficacy of trained models using evaluation measures like as accuracy, precision,

recall, and F1-score to properly analyze, validate our predictions, and optimize hyperparameters for the best outcomes.

**CHAPTER FOUR**

**RESULT AND DISCUSSION**

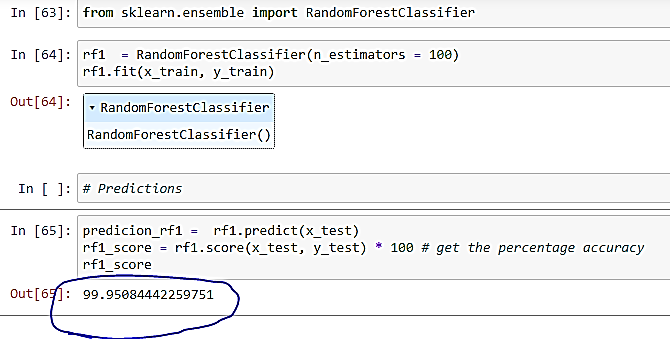
**4.1 INTRODUCTION**

This chapter digs into the outputs of our model, paying special emphasis to the Harmony Mean Value (also known as the F1 score), as well as our data precision and forecast accuracy. We also give infographics and presentations to help you understand our findings.

**4.2 RESULTS OF DATA IDENTIFICATION AND COLLECTION**

To begin, determine where to collect the required data sources for credit card transactions and fraud labels. Then, specify the data properties and features that are required for credit card fraud detection, such as Amount, Time, and Class.

The next stage was to collect data after identifying the data sources. Collect the indicated data from the source (Kaggle) and combine and merge the datasets to create a full dataset for analysis. Cleaning was performed to check for missing values, eliminate duplicates, and resolve any data quality concerns that may have affected the model's performance, and our data was completely clean. Oversampling, undersampling, and employing the Synthetic Minority Over-sampling Technique (SMOTE) were used to balance the quantity of fraudulent and genuine transactions. To analyze the model's performance, the data was separated into training, validation, and test sets. At the end of the process, we discovered that after using the Random Forest Classifier of n\_estimator of 100 Decision Trees, we obtained an accuracy of 99.95, which is extremely unusual. We then used a validation process using one of the Machine Learning Algorithms called K-Fold Classifier to correct Data Biasness and Imbalances.



**Figure 4.1** Displays a predictive model of a Random Forest Classifier with an accuracy score of 99.95.

**Source**: The source of the model is derived from the source code.

**4.3 DISCUSSION OF THE FEATURE SELECTION RESULT**

When it comes to detecting credit card fraud, depending just on "Amount" and "Class" as attributes may not be sufficient. Let's look at the benefits and downsides of employing these two features.

1. **Amount**: The "Amount" refers to the transaction value and is extremely important in detecting fraudulent activity. Typically, fraudulent transactions involve far larger or smaller sums than real ones. The model may identify patterns of fraudulent transactions linked with certain transaction values by utilizing "Amount" as a feature.

2. **Class**: The label "Class" is used to distinguish a fraudulent (labeled as 1) transaction from a valid (labeled as 0) transaction. This label is crucial for supervised learning algorithms as it helps train the model to distinguish between the two types of transactions by providing them with ground truth labels. Considering the utilization of "Amount" and "Class" as features, the following potential outcomes can be expected:

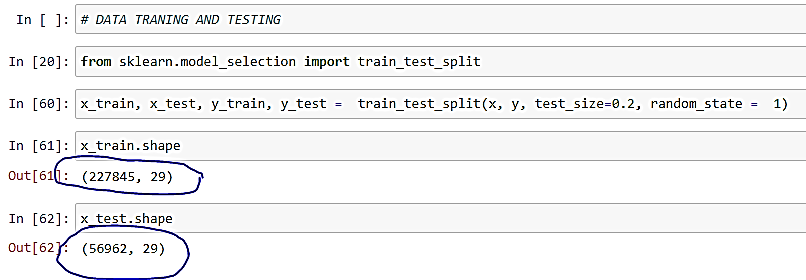
**1. Basic Baseline Model:** Using only the "Amount" and "Class" characteristics, a simple baseline model for credit card fraud detection may be developed. This algorithm would forecast based just on the transaction amount, making it simple to spot any possible outliers.

**2. Limited Discriminatory Power:** While "Amount" can be useful, it may have limits in properly discriminating between fraudulent and lawful transactions. Many genuine transactions can have varied quantities, while some fraudulent transactions may have sums that are comparable to lawful transactions.

**4. Potential Overfitting:** There is a risk of overfitting if we only rely on "Amount" and "Class" for detecting fraud, particularly if the patterns of fraud don't solely depend on transaction amounts. This can cause our model to not perform well when it encounters new and unseen data.

**4.4 RESULT FROM THE FEATURE SELECTION PROCESS**

From the feature Selection, below are the results:



**Fig 4.2** Showing the result from the feature selection

**Source**: The source of the model is derived from the source code

**4.5 RESULTS OF MODEL FORMULATION**

When formulating a model for credit card fraud detection using Random Forest and Support Vector Classifier (SVC), the results can differ depending on various factors such as the data quality, feature engineering, hyperparameter tuning, and evaluation metrics. In this overview, I will provide general information on these two algorithms.

**1. Random Forest:**

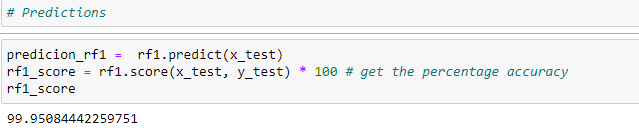
**Advantages:**

1. The Random Forest model is reliable and less likely to overfit as it combines multiple decision trees in an ensemble approach.
2. It can handle high-dimensional data and is relatively less sensitive to feature scaling.
3. You can easily access feature importance analysis, which gives you a better understanding of the features that have the most impact on the model's predictions.
4. The Random Forest algorithm is capable of handling both linear and nonlinear relationships that exist between the features and the target variable.

**Potential Results:**

1. Random Forest is effective for credit card fraud detection with diverse and well-prepared datasets.
2. It has the ability to handle class imbalances with great effectiveness and can seamlessly work with both categorical and continuous features.

The Below shows the result of the Model Formulation:



**Fig 4.3** Showing the Result Formation of Random Forest Classifier Model

Source: Extracted from the Source Code Used in the project

**2. Support Vector Classifier (SVC)**

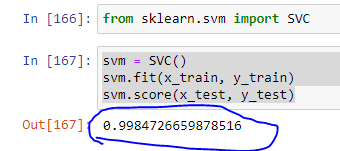
**Advantages:**

* 1. SVC is capable of effectively managing high-dimensional data, even when it involves nonlinear relationships that have been transformed by kernel functions.
  2. It is effective in finding the optimal hyperplane that separates classes in feature space, making it suitable for tasks with clear boundaries.

**Potential Results**

1. SVC can provide good performance when the data is linearly separable or close to linearly separable.
2. It should be noted that while Support Vector Classification (SVC) is generally a reliable method for fraud detection, it may not be as effective when dealing with larger datasets that are heavily imbalanced, such as those commonly encountered in credit card fraud detection.

The Below shows the result of the SVC Model for the Credit Card Fraud Detection System



**Fig 4.2** Showing the Result Formation of Support Vector Classifier Model

Source: Extracted from the Source Code Used in the project.

**Comparison of the two Models:**

1. In general, Random Forest requires less hyperparameter tuning and is easier to use out-of-the-box when compared to SVC.
2. When dealing with nonlinear data, achieving optimal performance with SVC may require more careful selection of hyperparameters such as kernel type and C-parameter.
3. Random Forest is a highly versatile tool for detecting credit card fraud as it can handle various data types and distributions effectively.
4. SVC might struggle with datasets that have a high degree of class overlap or significant class imbalance.

**4.6 PERFORMANCE EVALUATION OF MODEL**

The performance of a model for credit card fraud detection must be evaluated in order to determine how effectively the model is functioning and to discover areas for improvement. There are a variety of assessment criteria that are routinely used for classification tasks such as credit card fraud detection. Here are some of the most important assessment metrics:

**1**. **Accuracy**: Our test prediction had an accuracy of more than 90%, indicating that we properly identified both true positives and true negatives out of the total cases in the dataset.

**2**. **Precision**: The fraction of real positive predictions (fraudulent transactions) in comparison to the total number of anticipated positive occurrences. It assesses the model's capacity to prevent false positives, which is critical in fraud detection to decrease false alarms.

**3. F1 Score:** A statistic that balances accuracy and recall is the harmonic mean of the two. This is particularly useful when dealing with unequal class distribution. We attained an F1 Score of 99.95 and Over 90% utilizing the SV Classifier in the instance of the Random Forest Classifier.

**4.7 DISCUSSION OF MODEL FORMULATION**

Creating a successful credit card fraud detection system necessitates rigorous model construction. This entails selecting an appropriate machine learning algorithm, designing the structure of the model, and tweaking its parameters to achieve optimal performance. Let's go over the key components of model development for credit card fraud detection.

The machine learning algorithm used is critical and can have a substantial influence on the model's performance. Random Forest, Support Vector Classifier (SVC), and K-fold Classifier are commonly used methods for detecting bias in data.

Random Forest is employed due of its resilience, capacity to handle high-dimensional data, and effective management of unbalanced datasets, whereas SVC can be strong when the data is well-separated but may not be the best choice for highly imbalanced datasets.

**4.8 DISCUSSION OF RESULTS ON VALIDATION AND EVALUATION PERFORMANCE ON MODEL FORMULATION**

To avoid **overfitting,** the model's performance should be evaluated using a different test dataset that the model has not seen during training.

**Cross-validation** techniques may be used to provide a more credible assessment of the model's performance, especially when data is restricted; hence, **the K-Fold Classifier** was used to fix this. This resulted in an accuracy of more than 90%, which is an excellent forecast.

**CHAPTER FIVE**

**SUMMARY, CONCLUSION, AND RECOMMENDATION**

**5.1 SUMMARY**

Credit card fraud is an expensive and ongoing issue for both financial institutions and individuals throughout the world. Because credit cards are increasingly being used for financial transactions, fraudsters are continually devising new methods to attack flaws in the payment processing system. Credit card fraud has serious effects, resulting in considerable financial losses for both people and businesses, a loss of client trust, and harm to financial institutions' reputations.

**5. 2 CONCLUSION**

This is a Machine Learning-based Project focused at generating models for financial institutions and customers worldwide, as credit card theft is a persistent and costly issue. Because more individuals are utilizing credit cards for financial transactions, fraudsters are always devising new ways to exploit flaws in the payment processing system. Credit card fraud has substantial consequences, including significant financial losses for both people and businesses, a loss of client trust, and damage to the reputations of financial institutions.

**5.3 RECOMMENDATION**

This project is suggested for every financial organization that conducts any financial transaction using a credit card. This study has created a model to aid in the detection of credit card fraud during the influx and outflow of transactions.

**REFERENCE**

Al Rubaie Al Rubaie (2021) EMH. Improvement in credit card fraud detection using ensemble classification technique and user data. International Journal of Nonlinear Analysis and Applications. 2021;12(2):1255–1265. doi: 10.22075/IJNAA.2021.5228.

Al Smadi & Min Al Smadi B, Min M (2020). A critical review of credit card fraud detection techniques. 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON); Piscataway: IEEE; 2020. pp. 732–736.

Al-Faqeh Al-Faqeh AWK, Zerguine A, Al-Bulayhi MA, Al-Sleem AH, Al-Rabiah AS (2021). Credit card fraud detection via integrated account and transaction submodules. Arabian Journal for Science and Engineering. 2021;46(10):10023–10031. doi: 10.1007/s13369-021-05856-5.

AbhimanyuSiciliano, Robert (2016). ["What Is Account Takeover Fraud?"](https://web.archive.org/web/20170912103303/https:/www.thebalance.com/what-is-account-takeover-fraud-1947476). the balance. Archived from [the original](https://www.thebalance.com/what-is-account-takeover-fraud-1947476).

Dastgir Krebs, Brian (2014). ["Adobe hacked: customer data, source code compromised"](http://www.smh.com.au/it-pro/security-it/adobe-hacked-customer-data-source-code-compromised-20131004-hv1wl.html). The Sydney Morning Herald. The Sydney Morning Herald Newspaper.

Dilip V. (2016). The Principal Component Analysis, “Evaluation of Performance of Several Sampling on Classifier”

Gamini Gamini P, Yerramsetti ST, Darapu GD, Pentakoti VK, Raju VP ((2021). Detection of credit card fraudulent transactions using boosting algorithms. Journal of Emerging Technologies and Innovative Research (JETIR) 2021;8(2):2021.

Gao Gao J, Zhou Z, Ai J, Xia B, Coggeshall S (2019). Predicting credit card transaction fraud using machine learning algorithms. Journal of Intelligent Learning Systems and Applications. 2019;11(3):33–63. doi: 10.4236/jilsa.2019.113003.

Gupta, Lohani & Manchanda Gupta A, Lohani MC, Manchanda M (2021). Financial fraud detection using naive Bayes algorithm in highly imbalance data set. Journal of Discrete Mathematical Sciences and Cryptography. 2021;24(5):1559–1572. doi: 10.1080/09720529.2021.1969733.

Irby, LaToya, (2019). ["The Balance 9 Ways to Keep Credit Card Fraud From Happening to You"](https://web.archive.org/web/20201130235346/https:/www.thebalance.com/ways-avoid-credit-card-fraud-960797).

Karthik, Mishra & Reddy Karthik VSS, Mishra A, Reddy US (2022). Credit card fraud detection by modelling behaviour pattern using hybrid ensemble model. Arabian Journal for Science and Engineering. 2022;47(2):1987–1997. doi: 10.1007/s13369-021-06147-9.

Karthik Karthik R, Navinkumar R, Rammkumar U, Mothilal KC (  
2019). International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 2019. Supervised machine learning algorithms for credit card fraudulent transaction detection; pp. 2456–3307.

Kikerpill, Kristjan, and Andra Siibak (2021). "MAZEPHISHING: THE COVID-19 PANDEMIC AS CREDIBLE SOCIAL CONTEXT FOR SOCIAL ENGINEERING ATTACKS."

KrishnaTucker, Eric, (2014). Techniques for detecting fraudulent transaction. ["Prosecutors target credit card thieves overseas"](http://hosted.ap.org/dynamic/stories/U/US_CREDIT_CARD_THEFT?SITE=AP&SECTION=HOME&TEMPLATE=DEFAULT&CTIME=2014-09-12-17-56-33)

KuldeepWoolston, Sarah (2017). ["Machine Learning Methods for Credit Card Fraud Detection"](https://www.proquest.com/docview/1954696965). Proquest. [ProQuest](https://en.wikipedia.org/wiki/ProQuest_(identifier)) [1954696965](https://search.proquest.com/docview/1954696965).

Le Borgne, Yann-Aël; Bontempi, Gianluca (2021). "Machine Learning for Credit Card Fraud Detection - Practical Handbook". Neural Networks in the book titled "Business Applications of Neural Networks". World Scientific. ISBN 9789810240899.

Luis Little, Allan (2015). ["Overseas credit card scam exposed"](http://news.bbc.co.uk/2/hi/uk_news/7953401.stm). bbc.co.uk.com.

Mayer Brown (2023) ["Hong Kong Monetary Authority Issued Two Circulars to Authorized Institutions Regarding Payment Cards"](https://www.mayerbrown.com/en/perspectives-events/publications/2023/04/hong-kong-monetary-authority-issued-two-circulars-to-authorized-institutions-regarding-payment-cards).

McCurry, Justin (2016). "100 thieves steal $13m in three hours from cash machines across Japan". The Guardian.

Sharmistha Adsit, Dennis (2015). ["Error-proofing strategies for managing call center fraud"](https://web.archive.org/web/20110615000035/http:/www.isixsigma.com/index.php?option=com_k2&view=item&id=1854&Itemid=1&Itemid=1). isixsigma.com.

Stempel, Jonathan (2020). "Home Depot reaches $17.5 million settlement over 2014 data breach". Reuters.

William Westhoven (2016). ["Theft ring rigged Florham Park ATM, attorney general says"](http://www.dailyrecord.com/story/news/2016/11/17/theft-ring-rigged-florham-park-atm-attorney-general-says/94020118/). [Daily Record (Morristown)](https://en.wikipedia.org/wiki/Daily_Record_(Morristown))

Zetter, Kim (2010). ["TJX Hacker Gets 20 Years in Prison"](https://www.wired.com/2010/03/tjx-sentencing/). WIRED. Wired Magazine.