Tampere University

Patricia Rodríguez Vaquero

LITERATURE REVIEW OF CREDIT CARD FRAUD DETECTION WITH MACHINE LEARNING

Faculty of Information Technology and Communication Sciences

Master of Science Thesis

November 2023

ABSTRACT

Patricia Rodríguez Vaquero: Literature Review of Credit Card Fraud Detection with Machine Learning Methods
Master of Science Thesis
Tampere University
Data Science
November 2023

This thesis presents a comprehensive examination of the field of credit card fraud detection, aiming to offer a thorough understanding of its evolution and nuances. Through a synthesis of various studies, methodologies, and technologies, this research strives to provide a holistic perspective on the subject, shedding light on both its strengths and limitations.

In the realm of credit card fraud detection, a range of methods and combinations have been explored to enhance effectiveness. This research reviews several noteworthy approaches, including Genetic Algorithms (GA) coupled with Random Forest (GA-RF), Decision Trees (GA-DT), and Artificial Neural Networks (GA-ANN). Additionally, the study delves into outlier score definitions, considering different levels of granularity, and their integration into a supervised framework. Moreover, it discusses the utilization of Artificial Neural Networks (ANNs) in federated learning and the incorporation of Generative Adversarial Networks (GANs) with Modified Focal Loss and Random Forest as the base machine learning algorithm. These methods, either independently or in combination, represent some of the most recent developments in credit card fraud detection, showcasing their potential to address the evolving landscape of digital financial threats.

The scope of this literature review encompasses a wide range of sources, including research articles, academic papers, and industry reports, spanning multiple disciplines such as computer science, data science, artificial intelligence, and cybersecurity. The review is organized to guide readers through the progression of credit card fraud detection, commencing with foundational concepts and advancing toward the most recent developments.

In today's digital financial landscape, the need for robust defense mechanisms against credit card fraud is undeniable. By critically assessing the existing literature, recognizing emerging trends, and evaluating the effectiveness of various detection methods, this thesis aims to contribute to the knowledge pool within the credit card fraud detection domain. The insights gleaned from this comprehensive review will not only benefit researchers and practitioners but also serve as a roadmap for the enhancement of more adaptive and resilient fraud detection systems.

As the ongoing battle between fraudsters and defenders in the financial realm continues to evolve, a deep understanding of the current landscape becomes an asset. This literature review aspires to equip readers with the insights needed to address the dynamic challenges associated with credit card fraud detection, fostering innovation and resilience in the pursuit of secure and trustworthy financial transactions.

Keywords: Fraud Detection, Machine Learning, Deep Learning, Financial Sector, Resampling Techniques, Overfitting.

Table of Contents

List Of Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
AP	Average Precision
AUC	Area Under the Curve
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DT	Decision Tree
ENN	Edited Nearest Neighbor
FN	False Negatives
FP	False Positives
FP-Growth	Frequent Pattern Growth
FPR	False Positive Rate
GA	Genetic Algorithms
GBM	Gradient Boosting Machines
GDPR	General Data Protection Regulation
IEEE	Institute of Electrical and Electronics Engineers
IF	Isolation Forests
IQR	Interquartile Range
KNN	K Nearest Neighbors
LASSO	Least Absolute Shrinkage and Selection Operator
LDA	Linear Discriminant Analysis
LOF	Local Outlier Factor
LOOCV	Leave-One-Out Cross Validation
LR	Logistic Regression
ML	Machine Learning
NB	Naive Bayes
PCA	Principal Component Analysis
POS	Point of Sale
PSD2	Payment Services Directive
PSPs	Payment gateways/payment service providers

RF	Random Forest
ROC	Receiver Operating Characteristics
SCA	Strong Customer Authentication
SMOTE	Synthetic Minority Oversampling Technique
SVM	Support Vector Machines
TG	Transaction Graph
TN	True Negatives
TP	True Positives
TPR	True Positive Rate
T-SNE	T-Distributed Stochastic Neighbor Embedding

1. Introduction

Credit card fraud presents a substantial challenge, impacting individuals, enterprises, and financial institutions on a global scale. As credit card usage continues to rise in both online and offline transactions, malevolent actors have become adept at identifying and exploiting weaknesses, leading to an uptick in fraudulent endeavors. In response to this issue, a range of approaches and strategies have been devised for the detection of credit card fraud.

The primary objective of this thesis is to carry out an exhaustive examination of the field of credit card fraud detection, as detailed in the existing literature. This study endeavor seeks to closely examine and evaluate the diverse range of approaches, algorithms, and technological innovations deployed in the detection of fraudulent transactions involving credit cards. By conducting a thorough review of previous scholarly works, academic publications, and industry reports, this thesis intends to provide a well-rounded understanding of the current state of credit card fraud detection methods.

In this thesis, I delve into the fundamental principles that underpin the domain of credit card fraud detection, and I examine various commonly employed data sources, such as transaction records, customer profiles, and historical data. My research entails the analysis of diverse fraud patterns and behaviors detectable through data examination, such as deviations in transaction amounts, temporal patterns, and geographic factors.

Furthermore, the literature review in this work assesses both the advantages and disadvantages associated with traditional rule-based and statistical techniques. It also assesses emerging technologies, such as machine learning, artificial intelligence, and data mining, and assesses their effectiveness within the domain of credit card fraud detection. The study highlights their significance in practical, real-world applications.

Moreover, this study will delve into the obstacles and constraints encountered within the domain of credit card fraud detection. These challenges encompass the escalating sophistication exhibited by fraudsters, the scarceness of annotated datasets, and the imperative for ongoing vigilance and adaptability in detection models. The paper will also delve into ethical concerns that arise with the utilization of personal data for fraud detection purposes and propose potential strategies to mitigate these ethical dilemmas.

Ultimately, this thorough examination of existing literature provides a foundation for future research initiatives within the field of credit card fraud detection. It not only identifies

shortcomings in the current state of knowledge but also reveals opportunities for further advancement in enhancing the accuracy and efficiency of fraud detection systems through the refinement of algorithms, methodologies, and strategies.

Through conducting an extensive examination of the existing literature, this thesis intends to offer valuable references to scholars, professionals, and decision-makers involved in the domain of credit card fraud detection. This endeavor seeks to enhance the existing reservoir of information by amalgamating and critically evaluating recent research and progress, ultimately fostering the creation of more proficient and resilient fraud detection systems.

This document demonstrates the need for a comprehensive analysis of the existing literature to acquire a deep understanding of credit card fraud detection methods. The introduction provides a clear overview of the thesis's objectives, scope, and potential contributions, setting the stage for an in-depth examination of the topic.

1.1. Background and Context

The global economy relies heavily on the banking sector, which is a crucial conduit for providing financial services to individuals, businesses, and governments. As pivotal financial institutions, banks gather deposits from customers and utilize these funds to extend loans, credit, and assorted financial services. The intricate role played by banks in the economy encompasses the facilitation of financial transactions, fostering economic growth, and adeptly managing various risks.

The banking sector is characterized by stringent regulations, overseen by governmental bodies and regulatory authorities tasked with establishing benchmarks for safety, soundness, consumer safeguarding, and the mitigation of financial malfeasance. Adherence to an array of regulations encompassing capital requisites, liquidity benchmarks, and transparency stipulations is an integral obligation for banks.

The emergence of online banking, mobile banking, and fintech enterprises has witnessed substantial transformation within the banking landscape. This dynamic shift has induced heightened competition, spurred innovation, and introduced novel hazards and complexities. Banks are thus compelled to harmonize their operational ethos with nascent technologies and business paradigms while upholding their core functions pertaining to risk management, customer service, and financial mediation.

The banking sector plays a pivotal role in fostering economic expansion and development. By furnishing businesses and individuals with credit, banks empower them to invest in, expand, and engender employment opportunities. These institutions also play a crucial role in financing essential infrastructure projects such as transportation networks and power generation facilities, which serve to elevate economic productivity and bolster competitive prowess.

However, the banking sector is not immune to such challenges. Banks grapple with the specter of credit losses, market volatility, and cyber vulnerability. Their mandate entails striking a judicious equilibrium between profit generation and the fulfillment of obligations towards clientele, investors, and society at large. Recent times have witnessed a call for augmented transparency, accountability, and ethical comportment within the realm of banking.

Credit card fraud, classified as a form of financial wrongdoing, encompasses the illicit utilization of another individual's credit card details to execute deceitful transactions or acquire cash advances. Within the domain of the banking sector, this kind of fraud results in substantial monetary losses for both financial organizations and cardholders.

To address this challenge, financial institutions such as banks and credit card companies utilize a range of strategies for detecting and preventing fraudulent activities. Chiefly, they employ advanced fraud detection algorithms and machine learning models that enable them to conduct immediate analyses of extensive transactional data. These models make use of a wide array of data elements, including the transaction histories of cardholders, their spending habits, geographic indicators, and various other variables, to identify potentially suspicious transactions.

Concomitantly, banks harness the expertise of dedicated teams of fraudulent analysts who undertake manual scrutiny of transactions and investigate suspicious activities. This human intervention is complemented by a toolkit of techniques, spanning the scrutiny of transaction logs, liaison with merchants, and analysis of closed-circuit television footage, aimed at unearthing and forestalling fraudulent transactions.

The banking sector's arsenal includes an array of security measures for averting credit card fraud. These measures span the spectrum from multifactor authentication to fortifying cardholder identity validation, encryption of sensitive data encompassing credit card numbers, and implementation of policies and procedures dedicated to fraud prevention.

However, despite these concerted endeavors, credit card fraud remains an enduring predicament in the banking domain. Adversarial actors consistently innovate tactics and

technologies to circumvent detection, thus necessitating sustained bank vigilance. Indeed, the banking sector continues to channel investments into research and development, channeling efforts to enhance fraud detection algorithms and devising novel security measures to fortify defenses against credit card fraud.

The consequences of banking sector fraud are pronounced, affecting both individuals and industry. For individuals, fraud begets financial losses, tarnished credit standing, and emotional distress. On a broader scale, fraud corrodes customer confidence, casts aspersions on individual banks, and begets financial losses through reimbursements and legal expenditures.

In the crusade against fraud, the banking sector deploys an array of countermeasures encompassing fraud detection and prevention software, comprehensive employee training regimens, and strategic liaisons with law enforcement agencies. Notwithstanding these concerted efforts, fraud perseveres as an unwavering peril to the banking domain, demanding an incessant commitment to vigilance and innovation to outpace the machinery of malefactors.

In the realm of credit card transactions, several key terms and concepts shape the landscape of payments, fraud prevention, and associated processes. Understanding these terms is essential for comprehending the intricate dynamics of credit card fraud detection and mitigation.

1.1.1. Key Terms and Concepts

The following are the main terms that are essential to understand what the thesis is about and what will be discussed in the following chapters.

The payment process is a multi-step procedure that facilitates communication between financial institutions, processing networks, and point of sale (POS) systems, enabling businesses to accept digital payments. It begins when the customer selects a product or service and proceeds to the checkout. During this phase, the customer chooses a payment method and provides the necessary payment information. This payment information is then transmitted to a payment gateway, which verifies its accuracy for approval.

Following this, the payment gateway proceeds to relay this data to the payment processor. The payment processor then dispatches the information to the customer's issuing bank via the appropriate card network, such as Visa or MasterCard. The issuing bank proceeds to perform validation checks on the payment particulars, evaluating the presence of adequate funds or

available credit to finalize the transaction. If all criteria are met, the issuing bank grants authorization for the transaction. The authorization status is relayed back through the same pathway (issuing bank to payment processor to payment gateway), and the customer is promptly informed of the success or failure of their payment.

In the event of a successful payment, the customer receives an order confirmation. In the end, the monetary transfer takes place, transferring funds from the customer's bank account to the merchant's bank account. This procedure is commonly referred to as the settlement process [PayPal, 2023]. This settlement stage may take several days to finalize.

It is important to note that there are two primary methods of payment: "card present," where physical cards are used for payment, and "card not present," where card details are utilized for payments, often over the phone or online. The term "card not present," in contrast, entails an elevated potential for fraudulent activities, as it does not require the physical presence of the card, making it more challenging for the seller to verify the identity of the purchaser.

Merchants are the people selling goods and services to cardholders. Examples include e-Commerce sites and physical stores. They are motivated to keep fraud low because it hurts their profitability. Not actively targeting them [PayPal, 2023].

Payment gateways/payment service providers (PSPs) are people selling payment-related services to merchants and/or acquiring banks. They provide access to front-end services such as checkouts and 3D Secure, access to payment methods (there are hundreds globally and plenty in Europe), ancillary services such as revenue optimization (retries, routing payments to different acquirers), and fraud detection [Ravelin, 2023].

3D Secure is a security protocol used in the context of online credit card transactions to prevent credit card fraud. It was developed by Visa as "Verified by Visa" and Mastercard as "Mastercard SecureCode," and is also known by other names depending on the card issuer, such as "American Express SafeKey" or "JCB J/Secure." The main objective of the 3D Secure system is to enhance the security and authentication process for online card transactions. This is achieved by introducing an additional authentication step during the online checkout process. In this step, the cardholder is required to verify their identity by entering a distinctive password, an SMS code, or a temporary PIN [Burns, 2023].

Nonetheless, the utilization of the 3D Secure protocol can occasionally result in increased instances of shopping cart abandonment. This is primarily due to an additional authentication step that can disrupt the checkout process, potentially dissuading customers from completing

their purchases. An enhanced iteration known as 3D Secure 2.0 has been introduced to tackle these challenges. This updated version adopts a more seamless approach to customer authentication, relying on a risk-based methodology. Furthermore, it extends its support beyond traditional browser-based payment methods to encompass a wide array of alternatives, such as wearables, in-app purchases, mobile payments, and digital wallets. An added benefit of this protocol is its commitment to enhancing risk evaluation by transmitting over 100 data points to issuing banks during the approval or rejection of transactions [EBANX, 2023].

Acquiring Banks (acquirer) serve the merchants, and they take on the risk that the merchant will remain solvent. If the merchant goes out of business, the acquiring bank is liable for the funds that the cardholders can ask for because they do not receive their goods/services. They performed this via a process known as chargeback. Therefore, acquiring banks are highly motivated to control their merchants' fraud levels. However, they earn money from processing chargebacks, so they are also motivated not to stop it completely. There are approximately 400 acquiring banks in Europe, and they are usually either a PSP/Acquirer or an Acquirer/Issuer combo.

With the European-wide regulation called PSD2, there is a component called strong customer authentication (SCA) that significantly impacts acquirers. The short story is that if their aggregate fraud is > 0.13%, then they must trigger SCA (think of 3D Secure or the like) on transactions above €100 [Fine, 2023].

This kills conversions and drives away cardholders, which in turn makes it harder to retain merchants or win new businesses. The long story is well worth reading, which you can do here.

Issuing Banks are the banks of cardholders who buy goods from merchants. As fraudsters target cardholders, most of the time and issuers are the ones who make the final decision to accept or decline a transaction. However, since they are often very large enterprises that run in an antiquated way, it is difficult to get access to them, and they take a long time to make decisions. They are likely to approach them via partnerships with payment experts, PSPs, and acquirers.

1.1.2. Parties Affected by Payment Fraud

Various parties susceptible to the impacts of payment fraud will be now examined. Initially, it is essential to address the effects on customers, the real cardholders. On average, customers

spend approximately two working days in the process of card cancellation and dealing with the aftermath [Ravelin Blog, 2023].

For the second group, namely the merchants or online sellers, fraud presents a significant financial burden. Merchants face the loss of their merchandise and the necessity to issue refunds in response to chargebacks. Furthermore, they are required to cover chargeback fees imposed by their payment providers. Notably, fraud ranks as the primary concern for 44% of finance professionals in this sector. The occurrence of false positives is particularly undesirable, as they obstruct legitimate customer transactions and fail to contribute to revenue growth. Additionally, the potential for reputational damage looms large [Finextra, 2019].

Finally, it is important to highlight the role of payment providers in Europe, who bear responsibility for addressing fraud issues within the framework of the Payment Services Directive (PSD2). If the aggregate fraud rate remains at or below 0.13%, payment providers have the flexibility to conduct risk assessments and bypass 3D Secure authentication for individual transactions [Fine, 2023].

It should be emphasized that people partake in deceptive actions for various reasons, encompassing factors like convenient opportunities, sociocultural pressures, and financial incentives [Fisher, 2015].

1.1.3. Types of Frauds

Now various forms of fraudulent activities that individuals engaged in fraudulent activities perpetrate within the mentioned categories will be explored.

A research study conducted by Ghosh and Reilly, titled "Credit Card Fraud Detection with a Neural-Network" [Ghosh & Reilly, 1994] offers insights into payment fraud and underscores the application of neural networks in fraud detection. Most online payment fraud originates from identity theft, accounting for approximately 71% of cases. In these instances, criminals acquire card details, assume the victim's identity, and make online purchases. The fraudster receives the purchased items, while the genuine cardholder can initiate a chargeback process, often incurring associated fees. Nonetheless, individuals engaging in fraudulent activities can employ a range of strategies. These may include the use of personal information (email accounts, user profiles, names, addresses, IP addresses, and personal devices), to establish a facade that appears to be that of genuine customers, rather than relying exclusively on pilfered credit card data.

Another form of fraud, known as friendly fraud, occurs when legitimate customers claim non-receipt of ordered goods, whether intentionally or due to a mistake, or reported damage upon delivery. Instead of seeking a refund directly from the seller, these customers opt to initiate a chargeback through their bank. It is worth noting that chargebacks are a frequent source of dispute within the e-commerce industry [Decorte, 2023].

In a recent research paper authored by R. Anderson and published in 2022 under the title "Challenges in Information Security from an Economic Standpoint" [Anderson, 2022], the concept of clean fraud is thoroughly examined. Clean fraud pertains to deceitful transactions that exhibit an outward appearance of legitimacy. This form of fraudulent activity has been causing mounting concerns for retailers, primarily due to its ability to bypass conventional security measures, such as the identification of previously blacklisted fraudulent accounts. Clean fraud entails the illicit utilization of pilfered credit card details to assume the identity of the legitimate cardholder.

Affiliate fraud represents a form of deceptive activity wherein malicious individuals exploit traffic and registrations to deceive a merchant into believing they are garnering genuine consumer interest that is, in fact, non-existent. Numerous enterprises participate in or oversee affiliate marketing initiatives, which generate revenue by distributing links and content. Regrettably, affiliate fraud can take the guise of relatively straightforward actions, such as repeatedly refreshing a webpage or inundating users with spam emails and pop-up advertisements, thereby fabricating an illusion of substantial traffic [Dutta et al., 2010].

Triangulation fraud is a deceptive practice employed by cybercriminals who create counterfeit or imitation websites to lure unsuspecting buyers with enticing offers on inexpensive products. These fraudulent websites can sometimes surface in advertisements or be delivered to users via phishing attempts in emails, tricking them into visiting the site. This illicit scheme involves acquiring and subsequently utilizing credit card information that has been illicitly obtained, a practice commonly referred to as triangulation fraud. The term "triangulation" is derived from the three-step process it involves: enticing buyers, pilfering their personal details, and incorporating these details into a broader fraudulent operation. Furthermore, it is worth noting that established businesses can also suffer reputational damage due to such fraudulent activities [Cheliatsidou, 2021].

Overpayments, sometimes referred to as a payout scam, represent a form of stolen-card fraud. In this scheme, the perpetrator poses as someone in need of third-party services linked to a purchase. Subsequently, they propose paying the seller not only for the merchandise's price but

also an additional amount meant for the fictitious third party, often accompanied by a supplementary convenience fee (resembling a tip) to facilitate the arrangement. The essence of the deception lies in the non-existence of the promised third-party service, ultimately leading to the fraudster pocketing extra funds while leaving the seller entangled in a dispute [Stripe, 2023].

Alternative refunds involve a scheme where an individual with fraudulent intentions purposely overpays a transaction amount. Subsequently, they reach out to the company, asserting that they made an unintentional error in entering the payment. The fraudster then seeks a partial reimbursement to correct this alleged mistake but asserts that they have deactivated the associated card. Consequently, they request a refund to be processed through an alternative means, such as a check or wire transfer, outside the conventional card network channels.

As an illustration, consider a scenario where an individual with malicious intent contributes $500 to a charitable organization and subsequently reaches out to the organization, claiming that the intended donation was only $50. The person then requests a reimbursement of $450, employing an alternative payment method. Consequently, no funds are reimbursed to the initial credit card. If the genuine cardholder challenges the fraudulent transaction, the charity not only bears responsibility for the contested sum but also incurs a loss equivalent to the amount transferred through the alternate payment channel.

Marketplace fraud refers to situations where sellers fail to fulfill their obligation of delivering purchased goods after customers have made payments. One common fraudulent practice involves testing the validity of a payment card, either single or multiple cards, on one website before utilizing them for illicit transactions on another platform. Acquiring card information through illicit means, such as purchasing it on the dark web, represents the swiftest and most accessible method for obtaining a substantial quantity of card details. Additional tactics utilized by fraudsters include creating counterfeit websites and intercepting payment pages. Surprisingly, a significant portion of consumers, less than 25%, possess an understanding of the techniques employed by fraudsters, and merely 20% are aware that it is the retailers who bear the financial burden of such fraudulent activities [Stripe, 2023].

1.1.4. How Fraudsters Operate Online

In this chapter the methodology used by fraudsters to commit fraud will be introduced, and how they can be recognized [Ravelin, 2023].

Scammers employ sophisticated privacy tools like Anti-Detect and Kameleo to evade detection through browser IDs. These software solutions empower them to generate numerous virtual machine instances within browser windows. While this tactic does enhance their anonymity, it is worth noting that manipulating their apparent location remains a key red flag for potentially fraudulent activity. Fraudsters have the capability to identify the registration location of pilfered card information and can manipulate it to appear as if they are situated in that specific location.

Online phone numbers can be acquired using card information without requiring access to a physical phone. Consequently, individuals may reach out to the customer's mobile service provider to arrange call redirection to their designated number for the purpose of verifying purchases, if necessary.

Rather than making conspicuous large orders that could expose their intentions, they adopt the guise of authentic customers by gradually building, modifying, and removing items from their shopping carts before ultimately placing a substantial order.

To enhance their credibility and establish new accounts, they may incorporate additional customer data such as device IDs, driver's licenses, or blend this information with other customer particulars. This deceptive practice is frequently employed in cases of bank fraud.

The obligation falls on the seller to cover chargeback expenses incurred with their chosen payment provider. These charges have the potential to reach up to $50. Moreover, card schemes impose restrictions on the volume of chargebacks an online seller can experience before facing more substantial penalties. To mitigate the likelihood of chargebacks, merchants find it beneficial to allocate resources toward enhancing fraud detection and prevention measures. Payment providers that include fraud detection as an integral component of their services not only enhance the security for online sellers but also decrease the potential fee exposure [Ravelin, 2023].

1.1.5. Three pillars of fraud protection

The management of fraud protection relies on three primary guidelines outlined in this section. Payments meeting specific fraudulent criteria are subject to either blocking or further scrutiny, particularly in cases involving high-value orders that exhibit a higher likelihood of fraudulent activity. However, relying solely on rule-based approaches carries inherent risks, as there is a potential to unintentionally impede legitimate customer transactions. For instance, activating a

rule that blocks all transactions exceeding $500 would invariably affect numerous genuine customers. Nevertheless, rule-based methods can be employed with a 90% accuracy rate in situations where there is no requirement for ambiguity in decision-making, such as consistently flagging payments originating from exceedingly high-risk countries or regions. Combining rule-based approaches with machine learning proves to be a beneficial strategy.

Utilizing machine learning algorithms to categorize transactions into low, medium, or high risk levels is a common practice. These models can operate in real-time, leveraging historical data and concurrently incorporating new information. Furthermore, the synergy between machine learning models and graph networks is evident. For instance, one can instruct the machine learning model to identify and flag expansive networks for further scrutiny. It can also be programmed to block transactions originating from networks that exhibit rapid growth, thereby preventing potential fraudsters from exploiting multiple accounts for fraudulent purchases. In the research paper authored by G. Liu, J. Tang, Y. Tian, and J. Wang, titled "Graph Neural Network for Detecting Credit Card Fraud," the authors explore the utilization of Graph Neural Networks (GNNs) in addressing the issue of transaction fraud detection. They introduce a novel concept called the Transaction Graph (TG), which comprises weighted multiple graphs to comprehensively represent transaction data for enhanced fraud detection [Liu et al., 2021].

1.1.6. Identifying Potential Fraud

The utilization of potentially false information, such as fabricated phone numbers or email addresses comprising random characters, can raise suspicion within fraud detection systems. Moreover, discrepancies in customer details during multiple transactions, like employing the same email address with a different name or variations in payment information (such as using the same card but altering the shipping address or using multiple cards linked to a common shipping address), can serve as potential red flags for fraudulent activities.

Furthermore, specific patterns warrant vigilance, such as repeated transactions with the same card originating from an identical IP address or consistent use of the same customer’s name and email address. In instances where each failed attempt is associated with a distinct credit card, any successful payment should be treated as carrying significantly higher fraud risk.

In addition, the presence of suspicious or scripted communication, with messages sent to multiple sellers containing common phrases, can serve as another indicator of potential fraud.

These messages can often be cross-referenced using search engines to identify if particular sentences or content have been widely reused.

Lastly, unusually large orders, particularly those comprising numerous identical high-value items or expensive merchandise, should prompt heightened scrutiny, as they may suggest fraudulent intentions. These elements collectively contribute to a comprehensive approach to detect and prevent credit card fraud in online transactions.

It is worth noting that such behavior can be assessed for consistency with typical customer actions [Diadiushkin et al., 2019]. Despite seemingly low prosecution rates, public concern about payment fraud is significant, as demonstrated by the "Consumer attitudes to payment fraud survey" conducted by the Ravelin company [ECB, 2022].

1.2. Importance of Machine Learning

Machine learning holds a crucial position within the field of credit card fraud detection. With the increasing frequency of online financial transactions, the problem of credit card fraud has emerged as a significant worry for both individuals and businesses. Traditional rule-based systems designed for fraud detection often possess limitations when it comes to adapting to evolving patterns. Conversely, machine learning offers an efficacious solution by harnessing algorithms capable of learning from historical data to spot fraudulent activities in real-time.

One of the pivotal advantages of machine learning in credit card fraud detection pertains to its capacity to analyze copious volumes of data swiftly and accurately. Machine learning algorithms can efficiently process extensive transaction data encompassing cardholder details, merchant information, transaction amounts, geographic locations, and timestamps. This thorough examination enables the identification of patterns and irregularities associated with fraudulent activities. With the continuous integration of new data, these algorithms can adapt and enhance their detection capabilities progressively.

Another critical facet of machine learning in credit card fraud detection lies in its ability to uncover hitherto unseen or unfamiliar fraud patterns. Traditional rule-based systems lean heavily on pre-established rules and thresholds that may fail to encapsulate emerging fraud tactics. In contrast, machine learning algorithms possess the adeptness to pinpoint nuanced patterns and anomalies that may suggest fraudulent behavior, even if such behaviors haven't been explicitly defined within the rule set. The flexibility of machine learning models enables

them to proactively detect emerging forms of fraudulent behavior, outsmarting fraudsters in the process.

Furthermore, machine learning contributes to the reduction of false positives, which entail legitimate transactions being erroneously flagged as fraudulent. By scrutinizing historical data and assimilating insights from past decisions, machine learning algorithms can fine-tune their detection models, thereby minimizing false positives while maintaining a high level of accuracy. This not only enhances the customer experience by reducing unwarranted transaction declines but also conserves valuable time and resources for businesses by allowing them to focus on authentic fraud cases.

2. Traditional Approaches

Old school/traditional fraud detection relies solely on rules to block fraudulent payments. But using them also results in a lot of false positives as you block genuine customers if, for example, you define a threshold for when to block [Ravelin Insights, 2023].

Furthermore, prices can change over time; therefore, threshold values need to be updated.

Additionally, rules are most of the time dependent on yes/no answers and are therefore less flexible for adjustment or limit judgment on the risk scale. Using rules implies that the library must continue to expand as fraud evolves. Making the system slower and heavier maintenance. Therefore, they are inefficient and difficult to scale.

Fraudsters also impersonate genuine customers, making them harder to detect [Ravelin Insights, 2023].

2.1. Traditional Approach

Before the advent of machine learning algorithms, the field of credit card fraud detection predominantly depended on rule-based systems, manual verification processes, and various authentication methods [Nguyen, 2020].

Rule-based systems have been used to identify potentially fraudulent transactions based on predefined rules. These rules could be as simple as flagging transactions above a certain amount

or as complex as identifying unusual transaction patterns. For example, if a card is used in two different countries within a short time frame, it is flagged as suspicious. These systems use a set of logic-based rules to categorize data into non-suspicious or suspicious activities. As an illustration, consider a scenario where a transaction surpasses a specific predefined limit, leading to its identification as a potential candidate for fraudulent activity. Another rule could be that if a series of transactions occur in quick succession, this could also be flagged as suspicious.

Manual verification is another common method for detecting credit card fraud. This method involves human analysis and the verification of transactions. For instance, if a transaction seems out of the ordinary for a particular customer's typical spending habits, the credit card company might contact the customer directly to verify the transaction. Similarly, if a transaction is made in a location far from the customer's normal location, it might be flagged for manual verification.

Authentication measures have also been implemented to prevent credit card fraud. These measures include requiring additional information during the transaction process, such as a PIN or zip code. Some systems have also used challenge questions to verify the identity of the cardholder. This method is based on the principle of multifactor authentication, which involves verifying the cardholder's identity using multiple pieces of evidence or factors. These factors can include something the user knows (such as a password), something the user has (such as a physical card), and something the user is (such as a fingerprint) [Nguyen, 2020].

These approaches have historically proven their effectiveness and continue to find relevance in contemporary settings, frequently alongside more recent machine learning techniques. Nevertheless, they do come with certain constraints, such as reliance on human intervention and the challenge of staying abreast of constantly evolving fraudulent schemes. Consequently, machine learning algorithms have gained prominence as crucial instruments in combating credit card fraud, primarily due to their capacity to autonomously acquire and adjust to emerging fraudulent patterns [Kulatilleke, 2022].

2.2. Limitations

The older methods for credit card fraud detection listed above, although effective in many cases, also have several limitations listed below [Mohari et al., 2021].

The first limitation is rule-based systems that require constant updating to keep up with evolving fraud patterns. Because these rules are manually defined, they may not effectively adapt to new

types of fraud or subtle variations in fraud patterns (limited adaptability). These systems can generate a significant number of false positives, thereby leading to unnecessary verification procedures and potential customer dissatisfaction (false positives). Rule-based systems can become inefficient as the number of rules increases, as every transaction must be checked against each rule (inefficiency).

The second limitation is manual verification methods that are labor-intensive and time-consuming. Additionally, they depend on the availability and judgment of human operators, which can lead to inconsistencies. These methods can be seen as intrusive or inconvenient by customers, particularly if they are frequently contacted to verify transactions.

The last limitations are authentication measures in the form of user experience and data security. User experience measures while effective in preventing unauthorized access, multifactor authentication can negatively impact the user experience by adding additional steps to the transaction process. Data security methods depend on secure storage and transmission of sensitive information, such as passwords and PINs. If this information is compromised, the effectiveness of the authentication measures is significantly reduced [Mohari et al., 2021].

In addition, all these methods struggle with the inherent imbalance in credit card fraud detection data, where legitimate transactions vastly outnumber fraudulent transactions. This can lead to models that are very accurate overall but perform poorly in detecting relatively rare instances of fraud [Sulaiman et al., 2022].

As a result of these limitations, there has been a significant shift toward the adoption of machine learning and artificial intelligence methodologies in the realm of credit card fraud detection. These approaches possess the ability to autonomously acquire knowledge and adjust to emerging fraud patterns, efficiently process substantial datasets, and effectively address the imbalances present in fraud detection datasets [Kropelnytsky & Vidjikant, 2023].

3. Machine Learning

Machine Learning, often abbreviated as ML, constitutes a specialized domain within the realm of artificial intelligence (AI). Its primary objective lies in the development of computer programs adept at independently acquiring knowledge from data sources. This fundamental idea is grounded in the principle that systems have the ability to independently recognize patterns and

make decisions with limited human involvement. Machine Learning equips computers with the capacity to improve their performance by accumulating knowledge and experience over time [GeeksforGeeks, 2023].

Tom M. Mitchell formally defined Machine Learning as follows: "A computer program is considered to acquire knowledge from experience E concerning a particular category of tasks T, as gauged by performance measure P, if its performance in tasks within T, assessed by P, demonstrates improvement with the accumulation of experience E" [Wikipedia, 2023].

Machine learning algorithms possess the capacity to go beyond pre-defined algorithms and conduct contextualized inference, enabling them to analyze and interpret data while considering its contextual nuances and relationships. This capability empowers them to make more precise and context-aware predictions or decisions [MIT, 2021]. They achieve this by initially identifying patterns within data through algorithms and subsequently applying these patterns to forthcoming data. The process begins with the provision of top-notch data to the machines, followed by their training using a variety of machine learning models and diverse algorithms. The selection of algorithms depends on the unique features of the available data and the specific automation task at hand [GeeksforGeeks, 2023].

Machine Learning finds extensive application across various domains, encompassing recommendation systems, identification of malware threats, automation of business processes, filtering out spam, predictive maintenance scheduling, and the detection of fraudulent activities, to name a few [Nigam, 2023].

Furthermore, machine learning algorithms can be categorized into four main types [GeeksforGeeks, 2022].

One of these categories, known as **Supervised learning**, involves the analysis of labeled data, enabling the algorithm to understand how to correlate input data with specific output labels. This type of algorithm is frequently employed in tasks involving classification and prediction [Corbo, 2023].

Two distinct problem-solving applications are addressed through the utilization of supervised learning [IBM, 2023]:

- **Classification:** In this context, supervised learning employs algorithms to accurately categorize test data into predefined classes or categories. It involves the recognition of specific entities within a dataset and the subsequent inference of appropriate labels or definitions for these entities. Notable classification algorithms encompass linear

classifiers, support vector machines (SVM), decision trees, k-nearest neighbors, and random forests.

- **Regression:** Supervised learning plays a pivotal role in exploring the connections between dependent and independent variables, offering significant utility in generating predictions, such as forecasting the sales revenue of a specific business. Widely adopted regression techniques encompass linear, logistic, and polynomial regression algorithms [IBM, 2023].

Some commonly employed supervised learning algorithms, as discussed in a prior work [Alam, 2021], include:

- **Linear classifiers** rely on a linear predictor function, which involves the combination of a feature vector with a specific set of weights. Well-known instances of linear classifiers include Logistic Regression and the Naive Bayes classifier.
- **Support Vector Machines (SVM)** have a dual role, as they are versatile tools suitable for both classification and regression assignments. Their primary aim involves determining the most effective line or decision boundary for segregating the dataset into distinct categories.
- **Decision tree algorithms** ascertain straightforward decision rules from data features to make predictions about the value of a target variable.
- **K-nearest neighbors (k-NN)** is a classification algorithm that assigns class membership to an object. This technique categorizes an object by considering the majority decision from its closest neighbors, leading to the assignment of the object to the class that is most frequently found among its k-nearest neighbors.
- **Neural networks**, drawing loose inspiration from the human brain, comprise a set of algorithms designed for the purpose of identifying patterns. These algorithms analyze sensory information by means of machine perception, which involves assigning labels or grouping raw input data [Alam, 2021].

Unsupervised learning as a branch of machine learning involves the process of pattern recognition within unannotated datasets through techniques such as clustering and similarity identification. The fundamental goal of unsupervised learning revolves around the discovery of the intrinsic patterns within data, grouping data points according to their similarities, and ultimately expressing the dataset in a more compact and organized fashion [GeeksforGeeks, 2022].

There are three primary categories of unsupervised learning algorithms [Wu et al., 2021]:

Clustering algorithms have an important role in the task of categorizing comparable instances according to their intrinsic characteristics. They find extensive applications in scenarios where the goal is to unveil the underlying structure of a dataset or swiftly detect unknown patterns within a vast dataset. The most popular clustering algorithms will be introduced below.

- **K-Means Clustering** represents an iterative technique employed to divide a collection of n data sets into k separate and non-intersecting clusters. In this process, every individual data point gets assigned to the cluster that possesses the closest mean to it.
- **Hierarchical clustering** is characterized by a sequence of gradual partitions. It begins by grouping all items into a single cluster and proceeds incrementally until each cluster consists of just one individual entity.
- **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)** is a density-based clustering algorithm designed to identify clusters of varying shapes and sizes within extensive datasets. It excels at handling noisy data and outliers, making it a versatile tool for cluster discovery.

Dimensionality Reduction Algorithms serve the purpose of decreasing the number of random variables in consideration by identifying a set of primary variables. These methods come into play when dealing with a high count of input variables or dimensions, which can potentially result in overfitting and an elevated computational burden. Most popular dimensionality reduction algorithms are listed up next.

- **Principal Component Analysis (PCA)** is a method employed to accentuate variances and establish prominent patterns within a dataset. It proves especially valuable for simplifying exploration and visualization tasks.
- **Autoencoders** belong to the realm of artificial neural networks, designed to acquire efficient representations of input data. They find utility in tasks such as feature extraction, the development of generative data models, dimensionality reduction, and even data compression.

Association Rule Learning Algorithms play a crucial role in uncovering meaningful connections between variables within extensive databases. Their primary application lies in conducting market basket analyses. The most popular association rule-learning algorithms are as follows.

- **Apriori** stands as a timeless algorithm employed in data mining to unearth association rules. It is tailored to function efficiently within databases containing numerous transactions, such as the products purchased by customers at a store.

- **FP-Growth (Frequent Pattern Growth)** is an algorithm for extracting frequent patterns in transaction data. Compared to Apriori, it has the advantage of not requiring explicit generation of candidate sets, which makes it faster in practice.

These algorithms find wide-ranging utility across various domains, including medical imaging, anomaly detection, and recommendation systems. The selection of a specific algorithm hinges on the unique requirements of the task at hand [Wu et al., 2021].

In the realm of **reinforcement learning**, the algorithm engages in actions that yield the greatest rewards, making it a valuable tool frequently applied in scenarios involving AI systems designed for gaming and robotic navigation. Through a process of iterative trial and error, the reinforcement learning algorithm strives to optimize its performance and attain the highest possible reward [GeeksforGeeks, 2022].

Hierarchical clustering is characterized by a sequence of gradual partitions. It begins by grouping all items into a single cluster and proceeds incrementally until each cluster consists of just one individual entity. An illustrative use case for this approach can be found in the realm of anomaly detection [GeeksforGeeks, 2022].

Ensemble methods represent a class of meta-algorithms designed to enhance predictive modeling by integrating multiple machine learning techniques. These methods serve various purposes, such as reducing variance (bagging), mitigating bias (boosting), or enhancing prediction accuracy (stacking) [Alam, 2021].

Ensemble techniques can be categorized into two distinct groups:

- **Sequential Ensemble Methods:** Within this specific class, the process of sequentially creating base learners is evident, as demonstrated by the utilization of algorithms such as AdaBoost. The underlying rationale behind sequential methods is to capitalize on the interdependence among base learners. By assigning higher weights to previously misclassified examples, sequential ensembles aim to improve overall predictive performance.
- **Parallel Ensemble Methods:** Contrarily, parallel ensemble methods create base learners in parallel, as seen in Random Forest, for instance. These methods leverage the independence among base learners to significantly reduce errors by means of averaging.

In general, most ensemble techniques utilize a single foundational learning algorithm to generate consistent base learners, thus maintaining consistency throughout the ensemble.

However, it is worth noting that there exist approaches that incorporate heterogeneous learners, involving different types of base learning algorithms [Alam, 2021].

One important consideration to bear in mind is choosing the right machine learning algorithm, which depends on various factors. These factors encompass the problem you intend to tackle, the attributes of your dataset, and the computational capabilities available to you.

It is worth noting that various algorithms are tailored to specific task categories, and there is no universal approach that fits every scenario [Brownlee, 2020].

3.1. Models Used in Fraud Detection

Fraud detection represents a crucial application of Machine Learning, employing various models to achieve its objectives. Some of the most used models for fraud detection are as follows.

Decision trees stand out as a prominent supervised learning algorithm, primarily utilized in classification scenarios. This algorithm is adaptable to both categorical and continuous input-output variables. The fundamental principle involves partitioning the dataset into two or more homogeneous subsets, guided by the most significant attributes or independent variables, with the aim of creating distinct groups [Databricks, 2019].

On the contrary, Random Forest stands out as a versatile machine learning method that can handle regression and classification tasks with ease. Moreover, it incorporates dimensional reduction methods, handles missing values, identifies outliers, and undertakes essential data exploration steps effectively. This strategy belongs to the domain of ensemble learning, wherein a group of less powerful models collaboratively work together to create a resilient model [Team, 2023].

Logistic regression is a statistical method that employs a logistic function to model a binary outcome variable in its simplest form. In the domain of regression analysis, logistic regression, also known as logit regression, serves the purpose of estimating the parameters within a logistic model. It essentially constitutes a variant of binomial regression [Di Stefano, 2022].

Support Vector Machines (SVMs) have emerged as highly popular supervised learning techniques, particularly valuable for tackling classification and regression tasks. Although SVMs find their main utility in the realm of machine learning classification, they are designed to create an ideal line or decision boundary that effectively separates an n-dimensional space into

separate categories. This property enables the easy assignment of new data points to their respective classes in future scenarios [Pykes, 2021].

K-Nearest Neighbors (KNN) emerges as one of the simplest algorithms employed in Machine Learning for addressing both regression and classification problems. The core principle behind KNN algorithms lies in utilizing data to classify new data points, relying on measures of similarity, such as distance functions. Classification is achieved through a majority vote mechanism involving neighboring data points [Team, 2023].

Neural networks encompass a set of algorithms that draw inspiration from the organization of the human brain. Their fundamental objective revolves around the recognition of patterns. They analyze sensory information using machine perception, which involves tasks such as labeling or clustering raw input data [Skillfloor, 2023].

The selection of these models is contingent upon the distinct attributes of the data and the inherent complexities of the fraud detection issue. It is important to recognize that the effectiveness of these models may vary, sometimes requiring adjustments or enhancements to match the particular requirements of the given task [Team, 2023].

3.2. Advantages

Machine learning methods offer several advantages over traditional rule-based systems, manual verification, and authentication measures for detecting credit card fraud.

Machine learning models possess the ability to acquire knowledge from datasets and dynamically adjust to emerging fraud strategies without the need for manual rule modifications. This inherent adaptability endows them with remarkable flexibility, enabling them to effectively counter the ever-evolving tactics employed by fraudulent actors. These models can autonomously refine and fine-tune their algorithms by leveraging fresh data inputs, enabling them to discern intricate patterns and correlations within the data that might elude human investigators or rule-based systems [Opus Consulting, 2021].

These techniques exhibit a high level of effectiveness in handling extensive datasets, surpassing the scalability of rule-based systems and manual verification approaches. They possess the capability to swiftly process and scrutinize substantial volumes of transactional data, facilitating increased efficiency and faster identification of fraudulent activities [AltexSoft, 2020].

These models excel in discerning intricate data patterns and relationships that might elude rule-based systems. Their remarkable ability to pinpoint subtle anomalies and fraudulent activities with enhanced precision contributes to a reduction in both false positives and false negatives [AltexSoft, 2020].

Detecting credit card fraud frequently entails managing imbalanced data, wherein fraudulent occurrences are infrequent when contrasted with legitimate transactions. To address this imbalance and enhance the identification of fraudulent transactions, machine learning algorithms can be tailored. Methods such as under-sampling, over-sampling, and the utilization of the synthetic minority oversampling technique (SMOTE) can be applied to address data distribution imbalances, ultimately enhancing the model's performance [Satpathy, 2023].

Various machine learning approaches can be employed in the realm of credit card fraud detection, each possessing distinct advantages. For instance, Decision Trees and Random Forests exhibit proficiency in handling both categorical and numerical data types, rendering them capable of generating easily interpretable models. Neural Networks, on the other hand, excel at capturing intricate nonlinear relationships within the data. Support Vector Machines prove effective in managing high-dimensional datasets. This diversity in techniques facilitates the choice of the most appropriate model, aligning with the unique characteristics of the data and the specific problem under consideration [Varun Kumar K S et al., 2020].

Machine learning models possess the ability to incorporate feature selection methods with the objective of identifying the most relevant attributes in each dataset. This incorporation serves a dual role by both improving the model's effectiveness and alleviating its computational demands. Genetic Algorithms (GA) are commonly employed in the realm of credit card fraud detection for the purpose of feature selection [Ileberi et al., 2022].

These algorithms exhibit a unique capacity for detecting intricate patterns and anomalies, often beyond the scope of human analysts. Leveraging historical fraud patterns, they acquire knowledge and subsequently apply it to uncover potentially fraudulent activities in new transactions [Softjourn, 2022].

An alternative approach involves utilizing unsupervised machine learning algorithms to spot anomalies or outliers within the data. Given the rarity and distinct characteristics of fraudulent transactions in comparison to most typical transactions, they can be classified as anomalies. Consequently, machine learning proves exceptionally effective in the domain of fraud detection [Bajaj, 2021].

3.3. Challenges

Although machine learning methods offer significant advantages for credit card fraud detection, they also encounter their own set of challenges [Kulatilleke, 2022].

Effective model training in machine learning necessitates substantial volumes of data. However, obtaining these datasets can be difficult because of privacy and confidentiality concerns. Financial institutions are often reluctant to share transaction data due to regulatory and privacy concerns [Kulatilleke, 2022]. Some of the most used credit card fraud detection datasets include the following.

Credit Card Fraud Detection Dataset is available on Kaggle and contains credit card transactions, with a mix of genuine and fraudulent transactions. It is a common dataset used for developing and testing fraud detection algorithms [Kaggle, 2018].

IEEE-CIS Fraud Detection Dataset was provided as part of a Kaggle competition organized by the Institute of Electrical and Electronics Engineers (IEEE). It includes various features related to credit card transactions, and the task is to predict whether a transaction is fraudulent [Kaggle, 2019].

In addition to real datasets, there are synthetic datasets generated to simulate credit card transactions. These datasets can be useful for testing algorithms and models in controlled environments [Borgne, 2022].

These datasets provide a starting point for the development of credit card fraud detection models. However, it is crucial to understand that detecting credit card fraud is a continuous and evolving challenge. As such, there is a constant influx of new datasets being generated and utilized for research and development purposes.

Therefore, exploring other sources, such as research papers, academic institutions, and industry collaborations, is recommended for access to the latest and most comprehensive datasets in this field.

The dataset containing fraudulent transactions exhibits a substantial class imbalance, where the count of legitimate transactions far surpasses that of fraudulent ones. This imbalance poses a potential risk of introducing bias into predictive models, causing them to prioritize the majority class and consequently generating a significant number of false negatives. To address this issue, researchers employ a range of methods, including oversampling, undersampling, and the

generation of synthetic data. However, it is important to note that each of these techniques presents its own unique set of challenges [Kulatilleke, 2022].

Machine learning models, particularly intricate ones like neural networks, often pose challenges in terms of their interpretability. This inherent complexity can create difficulties in comprehending the rationale behind the identification of specific transactions as fraudulent. This issue can have repercussions in terms of regulatory adherence and effective customer communication [AltexSoft, 2020].

The efficiency of fraud detection systems hinges on their ability to promptly identify fraudulent transactions. Considering the potentially vast quantities of transaction data, reaching into the millions of transactions per day, it becomes imperative for classification times to be as low as tens of milliseconds. This demand aligns closely with the need for parallelization and scalability within fraud detection systems [De Jesus, 2019].

Fraudsters constantly changed their tactics to evade detection. This means that fraud detection models must be updated regularly to keep up with these evolving patterns. However, training and deploying new models is time-consuming [De Jesus, 2019].

Machine learning models may encounter the issue of overfitting, characterized by their proficient performance on training data but subpar performance on unfamiliar data. This predicament becomes especially pertinent in the context of fraud detection, given the evolving nature of fraud patterns over time. To address this challenge, strategies like cross-validation and regularization are employed to alleviate overfitting tendencies. However, it should be noted that implementing these techniques can introduce intricacies into the model training procedure [Xue, 2019].

Transactional data typically contain a plethora of categorical attributes, including customer identifiers, terminal references, and card types. Machine learning algorithms struggle to handle categorical features effectively, necessitating their conversion into numerical representations. Various techniques are commonly employed to achieve this transformation, such as feature aggregation, graph-based conversion, and the utilization of deep learning methodologies like feature embeddings [Lucas, 2020].

The last two challenges can be seen as part of the broader issue of class overlap. When dealing solely with raw transaction data, differentiating between genuine and fraudulent transactions becomes an extremely arduous task. This predicament is typically addressed through the

application of feature engineering methods, which enrich the raw payment data with contextual information [Lucas, 2020].

Conventional classification metrics, like the mean misclassification error or AUC ROC, are not well-suited for detection problems due to the presence of class imbalance and the intricate cost structure inherent to fraud detection. An effective fraud detection system must excel at identifying fraudulent transactions while minimizing false positives. Consequently, a multitude of metrics must often be considered to comprehensively evaluate the performance of fraud detection systems. Surprisingly, there remains no consensus within the field regarding the ideal set of performance measures to adopt [Lucas, 2020].

Despite encountering these obstacles, machine learning remains a promising strategy for detecting credit card fraud due to its capacity to acquire and adjust to intricate data patterns. A range of strategies are currently under investigation and being put into practice to surmount these challenges. These methods encompass the utilization of synthetic datasets for model training, the application of advanced deep learning techniques, and the creation of hybrid machine learning architectures. Additionally, tactics such as augmenting the minority class, diminishing the majority class, and employing cost-sensitive learning can aid in handling imbalanced data. Researchers are actively engaged in addressing these issues to enhance the efficacy and efficiency of machine learning-based systems designed for fraud detection [Kulatilleke, 2022].

4. Data Preprocessing Techniques

Data preprocessing encompasses the tasks of refining, converting, and structuring unprocessed data into a suitable format for training machine learning models. Effective data preprocessing holds significant importance as it profoundly influences the model's performance and precision by determining the quality of input data.

The precise preprocessing procedures needed are contingent upon both the nature of the data and the machine learning algorithm being employed. To illustrate, certain algorithms may exhibit sensitivity to the magnitude of input variables, whereas others might possess the capability to handle categorical variables without necessitating encoding [Maharana, 2022].

Data preprocessing is also known as Feature Engineering (this involves deriving new variables from available data. This could be as simple as creating interaction terms between variables, or as complex as applying domain-specific calculations).

Data preprocessing involves multiple techniques as described below.

4.1. Data Cleaning

This process entails the recognition and rectification of inaccuracies within the data, encompassing actions like addressing missing information, eliminating redundant entries, and rectifying discrepancies.

4.1.1. Handling missing values

Preparing data for machine learning models involves a crucial process since the majority of algorithms are unable to process incomplete data directly. The absence of data can arise from a range of factors such as human errors, equipment malfunctions, or individuals choosing not to respond to specific questions.

There are several solutions to deal with missing values whether to remove instances with missing data, fill in missing values with zeros or the mean, or use advanced imputation techniques [Emmanuel, 2021].

Several methods for handling missing data will be introduced next.

One approach to address missing data is deletion, which is a straightforward method involving the removal of instances or variables with missing values. Nevertheless, it is generally discouraged due to the potential loss of valuable information.

Another strategy is imputation, which entails substituting missing data with estimated values. Common techniques for imputation include mean, median, and mode imputation, as well as regression imputation and K-nearest neighbor imputation. It is important to bear in mind that imputation can introduce bias and alter the underlying data distribution.

Prediction models can also be employed to handle missing values by utilizing machine learning algorithms. For instance, one can train a model like Linear Regression or Decision Tree on the available data and then utilize this model to predict the missing values [Ogunbiyi, 2022].

Furthermore, ensemble methods such as bagging and boosting can be employed to address missing data. This approach involves creating multiple models and amalgamating their predictions. However, additional research is necessary to explore the effectiveness of these methods further.

Some algorithms such as Naive Bayes and certain tree-based models can handle missing values directly. They treat missing values as a unique category or use statistical methods to handle them [Brownlee, 2020b].

4.1.2. Outlier Detection

Identifying anomalies is of paramount importance during the initial data preparation stage in the realm of machine learning.

Outliers represent data points that deviate significantly from the overall dataset, and their presence is often attributable to inconsistent data entry, erroneous observations, or genuinely extreme data points. These exceptional data points have the potential to disrupt the data's distribution, which could impact the effectiveness of machine learning models. There are various methods available for the detection of these outliers [Boukerche, 2020].

One set of methods revolves around statistical approaches, which rely on the inherent statistical properties of the data. For example, data values that deviate beyond 1.5 times the interquartile range (IQR) from either the lower or upper quartile are commonly identified as outliers.

Similarly, data points exceeding three standard deviations from the mean may also be considered outliers. While these methods are relatively straightforward to grasp and implement, they may not be optimal when dealing with multidimensional data.

Another category of methods, known as distance-based techniques, identifies outliers based on their proximity to other data points. As an example, the Local Outlier Factor (LOF) algorithm assesses the deviation in local density for a specific data point when compared to its nearby data points. It detects instances that exhibit notably lower density when compared to their neighboring data points, indicating the possibility of them being outliers.

Clustering methods, on the other hand, aim to group similar data points into clusters and identify points that fall outside of these clusters as outliers. One example is the K-means algorithm, which labels data points not belonging to any cluster as outliers.

Anomaly detection systems, such as the Isolation Forest algorithm, are specifically tailored for outlier detection. The Isolation Forest accomplishes this by isolating observations through a process of random feature selection and subsequent random selection of split values within the chosen feature's range. The underlying rationale is that isolating anomalies is relatively straightforward because only a minimal number of conditions are needed to separate these exceptional cases from the normative observations.

4.2. Data Transformation

This process encompasses modifications to the scale, nature, or arrangement of the dataset's variables or attributes. Instances of such adjustments involve the standardization or normalization of numerical variables, aimed at achieving uniform scales. This practice can enhance the effectiveness of algorithms sensitive to scale variations, like gradient descent-based techniques. Additionally, it involves the encoding of categorical variables.

4.2.1. Feature Scaling

Feature scaling plays a vital role in the data preprocessing phase, aiming to normalize the span of independent variables or data attributes. This method plays a pivotal role in ensuring that all attributes contribute equally to a model, preventing any single feature from exerting undue influence due to variations in their scales. It becomes imperative to employ feature scaling when working with datasets that encompass attributes exhibiting dissimilar scales, units of measurement, or orders of magnitude [Bhandari, 2023]. We will now introduce the most widely recognized feature scaling techniques.

One prevalent approach is normalization, also referred to as Min-Max scaling. In this technique, values are adjusted and rescaled to fall within the range of 0 to 1. The transformation is achieved through the formula X' = (X - Xmin) / (Xmax - Xmin), where X signifies the original value, X'

represents the normalized value, and Xmax and Xmin denote the maximum and minimum values of the respective feature.

Another scaling method is standardization, which alters features to have a mean of zero and a standard deviation of one. The standardization formula is expressed as $X' = (X - \mu) / \sigma$, where X corresponds to the original value, X' stands for the standardized value, μ represents the mean of the feature values, and σ signifies the standard deviation of the feature value.

4.2.2. Encoding Categorical Variables

Categorical variables refer to variables with a limited and typically fixed set of potential values. These variables encompass characteristics such as color (e.g., red, blue, green), size (e.g., small, medium, large), and geographic designations (e.g., city, suburban, rural) [Brownlee, 2020c]. Various techniques exist for encoding categorical variables.

One method is Label Encoding, which involves converting each distinct category into a numerical value. For instance, in the case of a color feature with categories red, green, and blue, they could be encoded as 0, 1, and 2, respectively. It is important to note that this approach can inadvertently imply an ordinal relationship between the numerical values, which may mislead the model.

On the other hand, One-Hot Encoding transforms each category value into a new column and assigns binary values of 1 or 0. This creates a binary vector representation for each integer value, where all values are zero except for one marked as 1. Unlike Label Encoding, One-Hot Encoding does not assume any inherent order among categories, making it a suitable choice for non-ordinal variables.

Ordinal Encoding is best for categorical variables with natural rank ordering. Categories are assigned an integer based on their ordering values. For example, for a feature like size with categories small, medium, and large, we could encode these as 0, 1, and 2 respectively.

4.3. Feature Selection

This process entails the determination of the most pertinent input parameters for the given task. In some cases, certain variables might prove superfluous or unrelated and can be omitted to decrease the data's dimensionality. This practice serves to diminish the quantity of input variables, leading to enhanced computational efficiency and the potential for improved model performance. Diverse approaches to feature selection are at one's disposal, encompassing filter-based, wrapper-based, and embedded methods [Brownlee, 2020d].

4.3.1. Filter-based feature selection

This approach employs statistical techniques to assess the connection between each input factor and the target variable. It gauges the correlation or interdependence among input factors and subsequently identifies the most pertinent features using these assessments. Filter-based techniques prove to be computationally efficient and can be highly successful in the process of feature selection. The specific statistical measure employed hinges on the data types associated with both the input and output variables.

4.3.2. Wrapper-based feature selection

This approach assesses various feature subsets through the process of training and assessing a machine learning model for each subset individually. It employs a search strategy, which could include techniques like sequential forward selection or best-first search, to identify the most effective subset of features that enhances model performance to its fullest potential. While wrapper-based methods may demand significant computational resources, they offer the advantage of yielding more informative and valuable insights.

4.3.3. Embedded feature selection

This approach integrates feature selection seamlessly into the machine learning model training procedure. It identifies relevant features by assessing their significance or by examining the coefficients generated by the model. An illustration of such an integrated technique is the Least Absolute Shrinkage and Selection Operator (LASSO), which employs L1 regularization to promote sparsity among feature coefficients, thus achieving the effective selection of a subset of features. Embedded methodologies exhibit efficiency as they concurrently execute feature selection within the model training phase.

4.4. Data Reduction

4.4.1. Dimensionality Reduction

Dimensionality reduction stands as a pivotal concept within the field of machine learning. It entails the process of decreasing the quantity of features or variables within a dataset, all the while preserving vital information and discernible patterns. This process serves several purposes, such as simplifying data, preventing overfitting, decreasing computational complexity, and enhancing the performance of machine learning models [Kumar, 2023].

The challenge commonly referred to as the "curse of dimensionality" emerges when handling data with a high number of dimensions, leading to significant difficulties and constraints. When the quantity of features expands, it results in data sparsity, causing a rapid exponential expansion in the volume of the data space. Consequently, the performance of machine learning models may suffer. Dimensionality reduction methods offer a solution to these issues by diminishing the feature count and streamlining data representations [Vadapalli, 2020].

A variety of dimensionality reduction techniques are at our disposal, each with its unique approach and underlying assumptions. Some commonly used techniques include the following [Brownlee, 2020e].

Principal Component Analysis (PCA) constitutes a linear technique that aims to reduce the dimensionality of data by projecting it onto a lower-dimensional space, with the primary objective of maximizing the preservation of data variance. It identifies the principal components,

representing the directions along which the data exhibits the most variation, to create a condensed data representation.

Linear Discriminant Analysis (LDA) represents a supervised method for reducing dimensionality, with the primary goal of discovering a lower-dimensional subspace that effectively enhances the distinction between different data classes or categories. This technique finds frequent application in tasks related to classification.

t-Distributed Stochastic Neighbor Embedding, often abbreviated as t-SNE, proves to be a valuable tool for the visualization of complex, high-dimensional data when it is necessary to project it into a lower-dimensional space. This approach preserves local data structure and interrelationships between data points, proving effective for tasks such as exploratory data analysis and clustering.

4.4.2. Sampling

Sampling methods have a crucial importance in the field of credit card fraud detection, especially when dealing with datasets that exhibit significant class imbalances. In such scenarios, where the number of legitimate transactions far outweighs that of fraudulent ones, these methodologies assume a critical role in mitigating the class imbalance issue and enhancing the effectiveness of fraud detection models.

Credit card fraud detection inherently faces the challenge of dealing with a substantial class imbalance due to the infrequent occurrence of fraudulent transactions compared to legitimate ones. This skewed distribution can result in the development of biased models that exhibit subpar performance in fraud detection. Sampling techniques aim to redress this imbalance by harmonizing the distribution of both fraudulent and legitimate transactions in the dataset.

When employing sampling methods in the context of fraud detection model evaluation, it becomes essential to evaluate its performance using a well-balanced dataset. Placing exclusive reliance on accuracy as the sole metric can lead to misinterpretation, especially when dealing with imbalanced class distributions. Instead, it is recommended to take into account a range of metrics, including precision, recall, F1-score, and the area under the ROC curve (AUC-ROC), in order to conduct a thorough assessment of the model's effectiveness. Employing these techniques allows for the adjustment of class distribution to achieve balance before initiating the training process for the fraud detection model.

Various sampling methods are commonly utilized in credit card fraud detection [Alamri & Ykhlef, 2022].

The Random Undersampling technique involves the selection of a random subset from the majority class (representing legitimate transactions) to equalize its size with that of the minority class (comprising fraudulent transactions). This methodology effectively tackles the problem of class imbalance. Conversely, the Random Oversampling method replicates instances from the minority class to augment its size, achieving a balanced distribution. While this approach prevents information loss associated with undersampling, it may also introduce the risk of overfitting.

The Synthetic Minority Over Sampling Technique, often referred to as SMOTE, stands out as a significant approach for creating synthetic samples within the minority class through interpolation between pre-existing instances. By mimicking authentic fraud patterns, SMOTE helps counteract the scarcity of fraudulent transactions, contributing to a more robust fraud detection model.

4.5. Data Splitting

Data splitting is a fundamental practice in machine learning, which involves partitioning a given dataset into three distinct subsets for various purposes: training, hyperparameter tuning, and model evaluation. This division is essential to ensure that the model undergoes different phases of assessment, preventing it from becoming overly specialized on the training data and enabling it to generalize effectively to new, unseen data [Gillis, 2022].

The process of data splitting serves several crucial purposes, including accurately gauging the model's performance, facilitating the selection of the most suitable model, and confirming its ability to generalize effectively. Furthermore, it plays a vital role in mitigating the risk of overfitting, ensuring a realistic assessment of the model's capabilities.

The allocation of data among these three sets: training, validation, and testing varies depending on the dataset size and the specific problem at hand. While there is no universally optimal split ratio, a commonly adopted distribution is 70% for training, 20% for validation, and 10% for testing. Nevertheless, this ratio can be adjusted to align with the data's availability and the requirements of the problem.

The training set, which comprises the majority of the data, plays a crucial role in the process of fine-tuning the model's parameters. During this phase, the model acquires insights into patterns and relationships within the data to make accurate predictions.

The validation set, also known as the dev set or cross-validation set, plays a critical role in fine-tuning the model's parameters and assessing its performance. This set is invaluable for selecting the most effective model and preventing overfitting. It is essential for hyperparameter tuning and model selection.

The testing dataset plays a crucial role in assessing the model's overall performance, serving as the ultimate benchmark. It serves as an impartial yardstick for gauging the model's capacity to generalize to data it has never encountered before. To preserve the integrity of this assessment, it is essential to abstain from employing the testing dataset for any kind of model training or parameter tuning [Gillis, 2022].

In Figure 1, we present a graphical representation of how data is usually split.

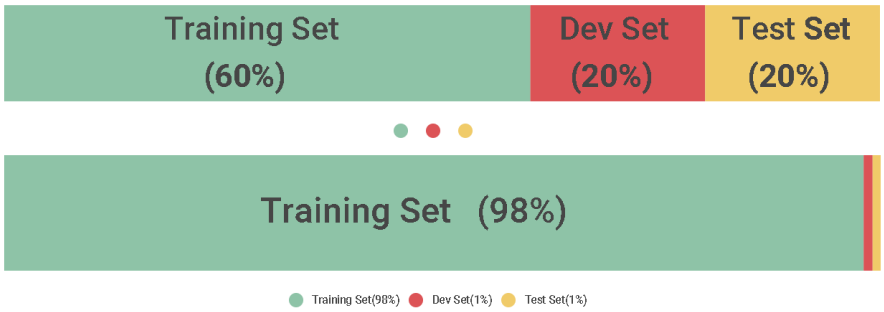

Figure 1 Data splitting technique, [Kumar, 2021]

4.6. Hyperparameter Tuning

Hyperparameter tuning, alternatively referred to as hyperparameter optimization, involves the task of selecting the most suitable values for a machine learning algorithm's hyperparameters. These hyperparameters serve as the settings that govern the algorithm's learning process and its behavior. It is important to note that, unlike model parameters, which are acquired through data-driven learning, hyperparameters must be specified manually by the practitioner.

The primary objective of hyperparameter tuning lies in identifying the ideal combination of hyperparameter values that can yield the highest level of model performance and generalization capability. This undertaking is of paramount importance because different hyperparameter configurations can have a significant impact on how a machine learning algorithm performs and generalizes when applied to a specific dataset [Brownlee, 2020f].

Numerous methods and approaches are at one's disposal when it comes to hyperparameter tuning.

4.6.1. Grid Search

Grid search is a method that revolves around the creation of a matrix of hyperparameter values, wherein a methodical exploration is conducted to identify the most favorable configuration. This comprehensive process entails an exhaustive examination of all potential combinations, assessing the model's effectiveness either through cross-validation or an independent validation dataset. It is worth noting that grid search can pose significant computational demands, especially when confronted with an abundance of hyperparameters or a broad spectrum of feasible values [Brownlee, 2021].

4.6.2. Random Search

Random search entails the process of sampling hyperparameter values in a random manner from predefined probability distributions. This allows for a more efficient exploration of the hyperparameter space than a grid search. Using random sampling, it is possible to cover a wide range of values and potentially discover better combinations [Brownlee, 2021].

4.6.3. Bayesian Optimization

Bayesian optimization represents a sophisticated approach, employing probabilistic models to capture the interplay between hyperparameters and model performance. This methodology harnesses this insight to conduct a strategic exploration for the most suitable hyperparameter

settings. Bayesian optimization stands out as a valuable tool, especially when dealing with extensive and intricate search spaces, as it enhances the efficiency of the search procedure [David, 2020].

4.6.4. Genetic Algorithms

Genetic algorithms draw their inspiration from the mechanisms of natural evolution and employ a population-centric strategy to explore and discover optimal hyperparameter configurations. This involves creating an initial population of hyperparameter sets, evaluating their performance, and then evolving the population through selection, crossover, and mutation operations. Genetic algorithms can be effective in finding good hyperparameter combinations, especially when there are interactions or dependencies between hyperparameters [David, 2020].

4.6.5. Automated Hyperparameter Tuning

Many machine learning frameworks and libraries provide built-in tools and algorithms for automated hyperparameter tuning. These tools often combine various techniques, such as grid search, random search, and Bayesian optimization, to automatically search for the best hyperparameter values. Examples include scikit-learn's GridSearchCV and RandomizedSearchCV classes, as well as libraries such as Optuna and Hyperopt [David, 2020].

It is crucial to emphasize that the process of hyperparameter tuning necessitates the utilization of either a distinct validation set or cross-validation. This approach is essential for obtaining an impartial assessment of the model's performance. Additionally, it is crucial to consider computational resources and time constraints when selecting a hyperparameter tuning strategy [Brownlee, 2021].

4.7. Cross Validation

Cross-validation stands as a pivotal method within the realm of machine learning, playing a crucial role in assessing a model's performance by making use of data that has not been seen before. In this method, the dataset is segmented into multiple partitions or folds. During this procedure, one of these folds is designated as the validation set, and the model is trained on the remaining folds. This cyclic process is reiterated multiple times, with each iteration using a different fold as the validation set. Ultimately, the outcomes from each validation iteration are averaged to yield a more reliable and consistent evaluation of the model's overall performance [GeeksforGeeks, 2023b].

The accompanying diagram (Figure 2) illustrates a cross-validation technique combined with the data split method.

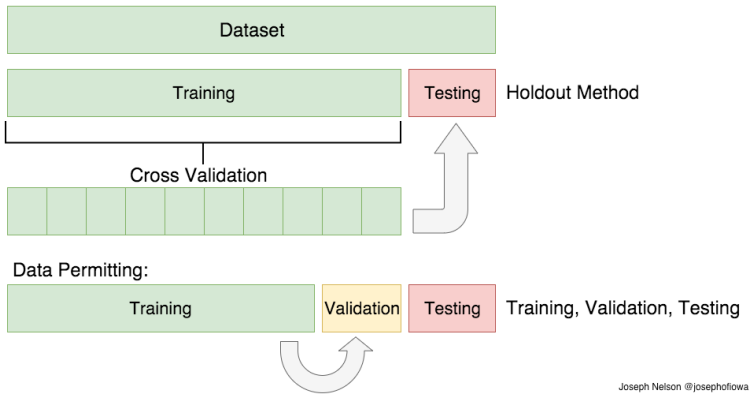

Figure 2 Train/Test Split and Cross Validation, [Bronshtein, 2023]

It helps estimate the performance of a machine learning model by providing a more reliable measure of its generalization ability. It allows for the evaluation of different models and hyperparameter settings, aiding in the selection of the best model for deployment.

Cross-validation plays a pivotal role in addressing the issue of overfitting, which occurs when a model becomes too tailored to the training data and performs poorly when exposed to new and unseen data. Through the process of evaluating the model on multiple validation sets, cross-validation provides a more precise assessment of the model's ability to generalize its learnings beyond the training data. In essence, it measures the model's ability to perform capably on novel, unseen data instances.

Various cross-validation techniques can be employed, including k-fold cross-validation, leave-one-out cross-validation, and stratified cross-validation. The choice of which technique to employ hinges upon factors such as the dataset's size and characteristics, as well as the demands of the modeling task.

Figure 3 provides a visual depiction of the general concept of Cross-validation.

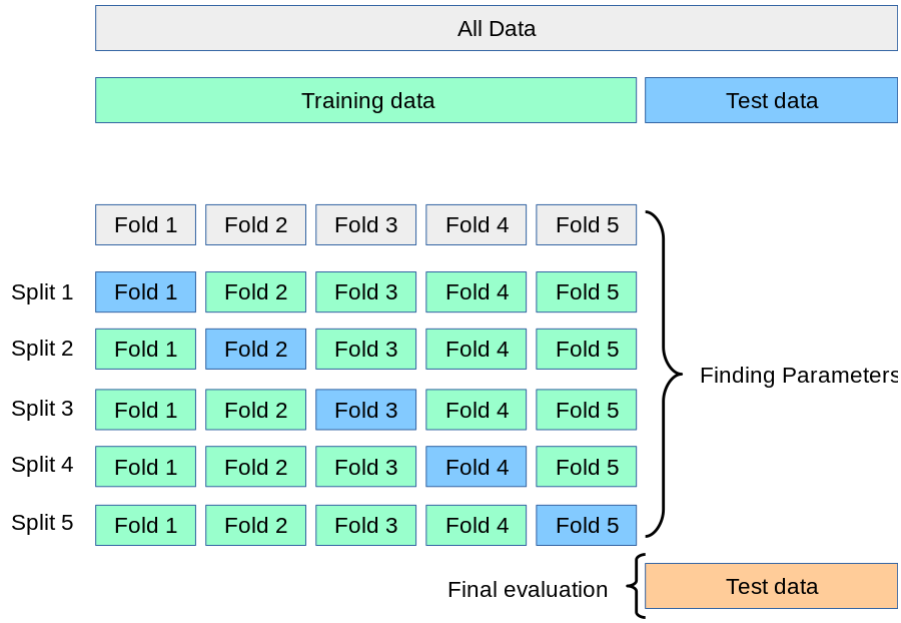

Figure 3 Cross-validation, [Scikit-learn, 2023b]

4.7.1. K-Fold Cross Validation

In this method, the dataset gets divided into k sections or segments. The model undergoes training using k-1 sections, leaving one section for evaluation as the test set. This sequence is reiterated k times, with a unique partition designated as the validation set in each iteration. The outcomes from these repeated iterations are aggregated to gauge the model's performance. For a more comprehensive grasp of this cross-validation method, please refer to Figure 4, which provides a visual representation of this technique.

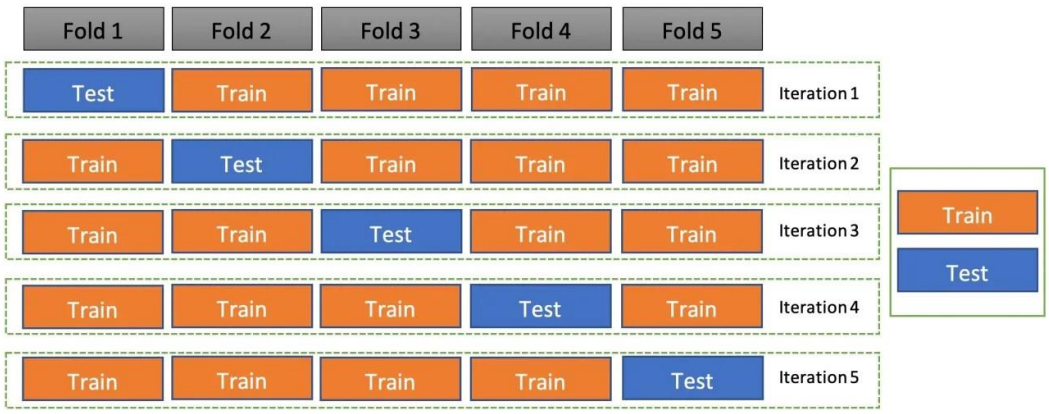

Figure 4 K-fold Cross-Validation, [Ranjan, 2021]

4.7.2. Leave-One-Out Cross Validation (LOOCV)

In this approach, the training of the model involves utilizing the entire dataset apart from one specific instance, which is reserved for testing purposes. This iterative procedure is applied to every individual instance within the dataset. The Leave-One-Out Cross-Validation (LOOCV) technique offers the advantage of leveraging all available data points for training; nevertheless, it may entail significant computational costs, particularly when dealing with extensive datasets. For a visual representation of the LOOCV methodology, please consult Figure 5.

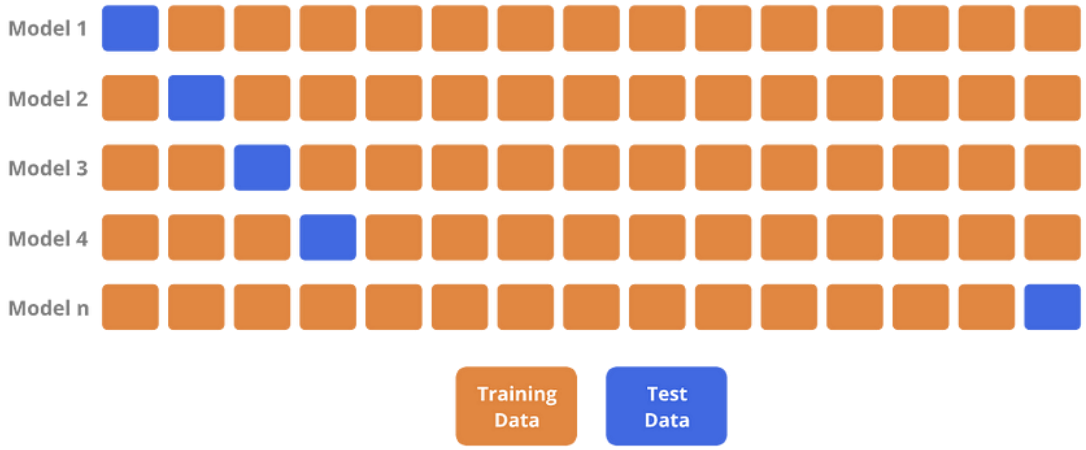

Figure 5 Leave-One-Out Cross-Validation, [Chaturvedi, 2023]

4.7.3. Stratified Cross Validation

Stratified Cross Validation is a method employed to ensure that each fold maintains a balanced representation of target classes, with approximately equal proportions as the complete dataset. This strategy is commonly utilized in classification assignments, particularly when confronted with imbalanced datasets, wherein one category is less frequently represented in comparison to the others. The core concept behind Stratified Cross-Validation is to guarantee that each fold remains a faithful reflection of the entire dataset by preserving the class distribution. This practice aids in obtaining more accurate estimates of both bias and variance [Dekanovsky, 2021].

4.7.4. Monte Carlo Cross Validation

Monte Carlo Cross-Validation, sometimes known as Monte Carlo resampling, stands as a valuable method for evaluating the performance of machine learning models. It proves particularly beneficial in situations with constrained data availability, where the objective is to maximize its utility [Cerqueira, 2022].

Unlike the traditional k-fold cross-validation method, where the data is divided into k distinct, non-overlapping folds for repeated model training and testing, Monte Carlo Cross-Validation involves random data partitioning into training and testing sets, performed multiple times. Each of these partitions is termed an iteration, and the model's performance is subsequently averaged across all iterations.

One of the notable advantages of Monte Carlo Cross-Validation is its capacity to provide a more robust evaluation of the model's performance, especially in scenarios featuring limited data or imbalanced datasets. Nevertheless, It is important to note that this approach can impose higher computational demands due to its typically larger number of iterations [Hu, 2023].

5. Performance Evaluation Metrics

In the forthcoming chapter, we will illustrate why the mean misclassification error is an inadequate performance measure because of the challenging characteristics of the fraud detection task, marked by its cost sensitivity and skewed class distribution. To exemplify this point, consider a transaction dataset containing only 0.1% of fraudulent transactions. Even a basic dummy model that labels all transactions as legitimate achieved an impressive accuracy of 0.99. This phenomenon is widely acknowledged within the realm of fraud detection, prompting the incorporation of alternative evaluation measures. These alternative performance metrics encompass indicators like recall, specificity, precision, the F1 score, AUC (Area Under the Curve) of the ROC (Receiver Operating Characteristics) curve (also referred to as AUROC - Area Under the Receiver Operating Characteristics), and Average Precision (AP) [Tharwat, 2020]. In the subsequent sections, we will undertake a comprehensive analysis of these metrics, elucidating their respective strengths and weaknesses. While these metrics play a crucial role in assessing the efficacy of fraud detection systems, it is important to note that there exists no unanimous consensus regarding the preferred evaluation metric.

Metrics such as recall, specificity, precision, and F1 score, often referred to as threshold-based metrics, possess notable limitations owing to their reliance on a decision threshold. Determining this threshold in practical scenarios is challenging and heavily influenced by business-specific constraints. These metrics are frequently supplemented by AUC ROC and, more recently, Average Precision (AP). AUC ROC and AP aim to provide a holistic assessment of performance across all conceivable decision thresholds, earning them the designation of threshold-free metrics. Presently, the AUC ROC is the prevailing metric for appraising fraud detection accuracy [Pozzolo, 2015]. However, recent research has underscored its shortcomings in the context of highly imbalanced problems, such as fraud detection [Muschelli, 2019], advocating the use of the precision-recall curve and AP metric as preferable alternatives [Saito & Rehmsmeier, 2015].

5.1. Threshold-based metrics

Threshold-based metrics are a group of performance measures that depend on the choice of a discrimination threshold in binary classification problems. The most common threshold-based metrics are Recall (Sensitivity), Specificity, Precision, and F1 Score [Kanika, 2021].

5.1.1. Confusion Matrix

A confusion matrix serves as a tool for summarizing the performance of a classification algorithm. It offers a detailed perspective on the classifier's accuracy, presenting both correct and incorrect predictions categorized by class. In essence, it not only sheds light on the classifier's mistakes but also elucidates the nature of these errors [Brownlee, 2016].

A confusion matrix consists of four elements for binary classification problems [GeeksforGeeks, 2023c].

- True Positives (TP): These correspond to instances where the model accurately identifies the positive class—i.e., when the actual class is positive, and the model predicts it as positive.

- True Negatives (TN): These instances arise when the model correctly predicts the negative class—meaning the actual class is negative, and the model correctly predicts it as negative.
- False Positives (FP) or Type I Error: These scenarios arise when the model erroneously predicts the positive class, specifically when the actual class is negative, but the model predicts it as positive.
- False Negatives (FN) or Type II Error: These cases occur when the model incorrectly predicts the negative class, that is, when the actual class is positive, but the model predicts it as negative.

The confusion matrix size is nXn for multiclass classification problems, where n is the number of classes [GeeksforGeeks, 2023c].

Using these four values, several performance metrics were calculated.

Accuracy: (TP + TN) / (TP + TN + FP + FN)

Precision: TP / (TP + FP)

Recall (or Sensitivity): TP / (TP + FN)

F1 Score: 2 * (Precision * Recall) / (Precision + Recall)

Specificity: TN/(TN+FP)

The confusion matrix offers a more comprehensive view of the model's performance than accuracy alone, as it can reveal whether the model is biased towards a particular class or whether it is difficult to distinguish between classes [Brownlee, 2016].

For a better understanding of the process, consult Figure 6, where a confusion matrix is represented.

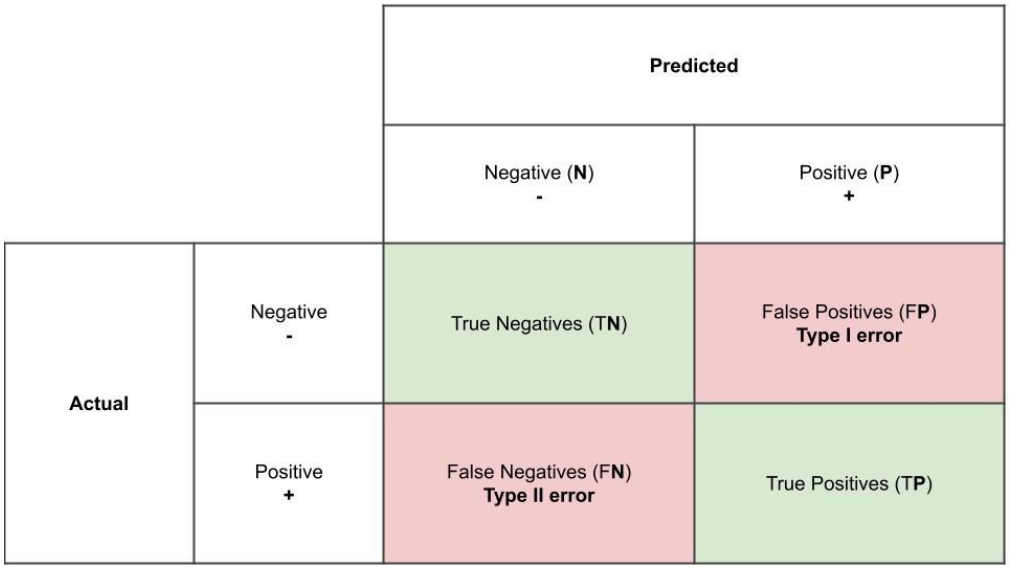

Figure 6 Confusion Matrix, [nbshare, 2023]

5.1.2. Mean Misclassification Error

The mean misclassification error, commonly known as the misclassification rate, serves as a crucial metric in the field of machine learning. It plays a pivotal role in quantifying the frequency with which a model incorrectly categorizes instances. This metric is computed by dividing the number of erroneous predictions by the total count of predictions [Zach, 2022].

To express it mathematically:

Misclassification Rate = Number of Incorrect Predictions / Total Number of Predictions

The misclassification rate possesses a range spanning from 0 to 1. When it reaches 0, it signifies that the model has made no misclassifications, while a value of 1 indicates that all predictions are erroneous. Therefore, it is highly desirable to achieve a lower misclassification rate as it corresponds to a higher accuracy level.

In the context of a binary classification problem, you can calculate the misclassification rate using the confusion matrix as follows:

Misclassification Rate = (False Positives + False Negatives) / (True Positives + True Negatives + False Positives + False Negatives)

It is important to keep in mind that while the misclassification rate provides a broad measure of model performance, it does not distinguish between various types of errors, such as false positives and false negatives. Consequently, it may not be the most suitable metric for datasets with imbalances or applications where different types of errors have varying associated costs [Brownlee, 2020b].

5.1.3. Cost Matrix and Weighted Loss

A cost matrix, alternatively referred to as a confusion matrix or error matrix, serves as a structured representation used for assessing the effectiveness of an algorithm, often one employed in supervised learning. In this matrix, each row corresponds to the instances assigned to a predicted class, while each column pertains to the instances belonging to an actual class [Jain, 2018].

In the context of binary classification, a cost matrix typically takes the form of a 2 × 2 table, providing a means to quantify the expenses or penalties associated with correctly identifying true positives, incorrectly labeling false positives, accurately identifying true negatives, and erroneously classifying false negatives [Brownlee, 2020b].

	Actual Positive $y_i = 1$	Actual Negative $y_i = 0$
Predicted Positive $c_i = 1$	C_{TP_i}	C_{FP_i}
Predicted Negative $c_i = 0$	C_{FN_i}	C_{TN_i}

Figure 7 Cost-Sensitive Cost-Matrix, [Bahnsen, 2016]

To delve further into the concept introduced in Figure 7 concerning the cost-sensitive cost-matrix, the abbreviations used are listed below.

CTN: True Negatives - denoting instances that are negative and correctly classified as such.

CFN: False Negatives – indicating instances that are positive but erroneously classified as negative.

CFP: False Positives - representing instances that are negative but inaccurately categorized as positive.

CTP: True Positives - signifying instances that are positive and accurately classified as positive.

Weighted loss, conversely, refers to the practice of assigning distinct weights or costs to various types of classification errors. For instance, in a medical diagnostic scenario, the cost of a false negative (failing to detect a disease) might outweigh that of a false positive (diagnosing a healthy patient as having the disease). Consequently, a higher cost could be ascribed to false negatives compared to false positives.

The overall cost of misclassification can be calculated as the sum of the costs associated with false negatives and false positives, each multiplied by their respective costs:

Total Cost = C(0,1) * False Negatives + C(1,0) * False Positives

The objective is to minimize the total cost in the context of cost-sensitive learning.

In machine learning, cost-sensitive learning techniques are employed to address imbalanced classification challenges, where the costs associated with different types of misclassification

errors vary. The cost matrix is instrumental in defining these costs and adapting the learning algorithm accordingly [Brownlee, 2020b].

5.2. Threshold-free metrics

These evaluation metrics analyze the classifier's effectiveness across a spectrum of thresholds, offering a more holistic assessment of the model's performance. Among the commonly used threshold-independent metrics are the Area Under the Receiver Operating Characteristic (AUC ROC) curve and the Average Precision (AP) [Leung, 2023].

5.2.1. Receiver Operating Characteristic (ROC)

The Receiver Operating Characteristic (ROC) curve is a widely employed visual aid in the domains of machine learning and statistics for evaluating the effectiveness of binary classification models. To create this curve, one maps the True Positive Rate (TPR), synonymous with sensitivity or recall, against the False Positive Rate (FPR), recognized as 1-specificity, while modifying the threshold parameters [Google, 2022].

In the ROC curve representation:

The Y-axis corresponds to the True Positive Rate (TPR), denoting the fraction of actual positive instances, such as individuals with a specific condition, correctly identified as positive by the model.

The X-axis represents the False Positive Rate (FPR), indicating the proportion of true negative instances, like healthy individuals, mistakenly classified as positive by the model [Google, 2022].

The ROC curve illustrates a trade-off relationship between sensitivity (TPR) and specificity (1 - FPR). As the model's discrimination threshold decreases to classify more instances as positive, both the count of true positives and false positives increases, impacting TPR and FPR [Pozzolo, 2015].

In a perfect scenario, an impeccable classifier would display an ROC curve that intersects the upper-left corner of the graph, indicating a sensitivity of 100% (meaning it has no false negatives) and a specificity of 100% (meaning it has no false positives). The Area under the ROC curve

(AUC), also referred to as the Area Under the Receiver Operating Characteristics (AUROC), quantifies the model's ability to differentiate between positive and negative classes. An AUC score of 1.0 represents flawless classification, while an AUC of 0.5 suggests a model that lacks classification capability, like random guessing [Glen, 2020].

To provide a clearer picture of the process, Figure 8 illustrates the AUC ROC curve concept and makes it clearer.

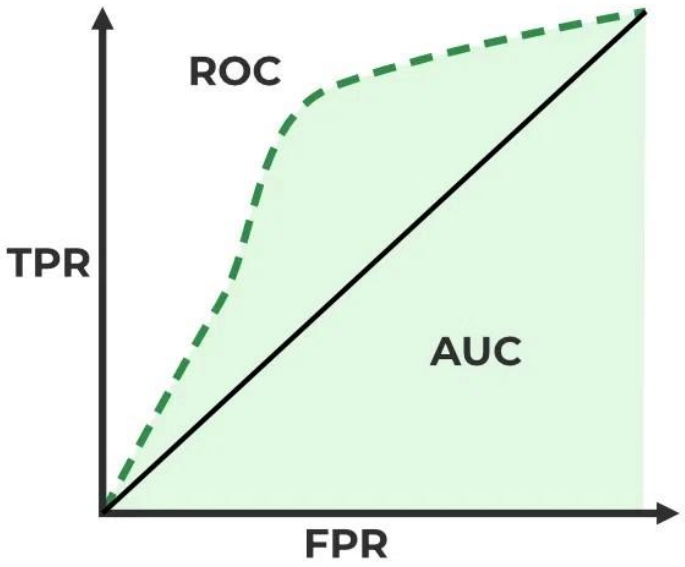

Figure 8 AUC ROC Curve, [GeeksforGeeks, 2023b]

5.2.2. Precision-Recall Curve

The Precision-Recall Curve functions as a visual depiction illustrating how well a binary classifier system performs when its discrimination threshold is adjusted. This curve is constructed using two key parameters, namely Precision and Recall.

Precision serves as an indicator of the relevance of results and is computed as the quotient of true positives (TP) and the sum of true positives and false positives (FP). It can be interpreted as the probability that a positive prediction made by the model is indeed correct [The scikit-yb developers, 2019].

The concept of Recall, sometimes referred to as sensitivity or the true positive rate, assesses the classifier's comprehensiveness and its capacity to correctly detect all positive instances. This measurement is determined as the proportion of true positives to the total number of true positives and false negatives (FN). It provides insight into the likelihood of the model accurately recognizing a positive instance within the class [The scikit-yb developers, 2019].

The Precision-Recall curve provides a clear visual representation of the balance between Precision and Recall as threshold values change. When the area under this curve is substantial, it indicates a strong performance in both precision and recall. High precision corresponds to a minimal false positive rate, while high recall indicates a minimal false negative rate [Brownlee, 2018].

Average precision (AP) serves as a concise metric, offering a consolidated evaluation by calculating the weighted average of precision values obtained at various thresholds. This calculation considers the rise in recall relative to the previous threshold as the weighting criterion. In essence, AP delivers a solitary numeric indicator that encapsulates the overall performance of the Precision-Recall curve [Scikit-learn, 2023].

To make the concept more tangible, Figure 9 illustrates the precision-recall curve.

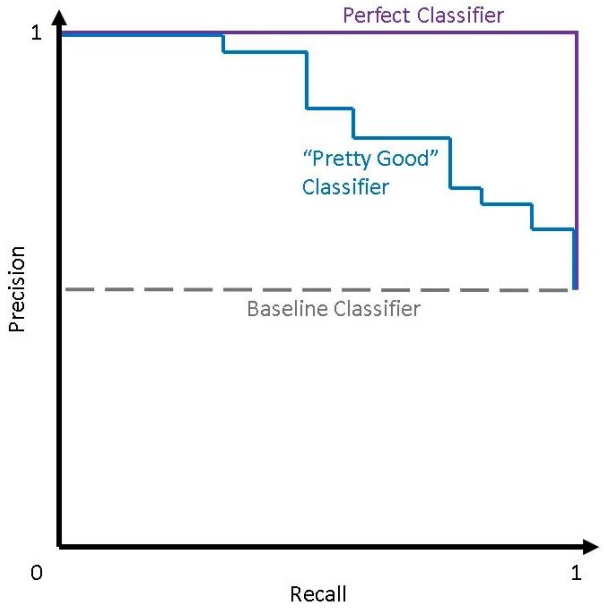

Figure 9 Precision-Recall Curve, [Steen, 2021]

6. Review of Existing Literature

In this chapter, a thorough and in-depth exploration of the vast landscape of credit card fraud detection is presented, with a primary emphasis on the pivotal role played by machine learning techniques. This encompasses not only a comprehensive survey of the existing literature, but also a meticulous dissection of the multifaceted facets that constitute the realm of credit card fraud detection.

One cannot begin to grasp the significance of machine learning in this context without acknowledging the profound challenges that beset the domain of credit card fraud detection. These challenges, as illuminated by our extensive literature review, are multifarious and often involve sophisticated and rapidly evolving tactics employed by fraudulent entities. Recognizing these intricacies is paramount as it underscores the urgency of employing advanced and adaptive methodologies to tackle the ever-shifting landscape of financial fraud.

Amidst this intricate backdrop, this review elucidates the pivotal importance of harnessing machine learning algorithms as a potent tool to bolster the efficacy and precision of fraud detection mechanisms. This study provides a compelling rationale for integrating machine learning into credit card fraud detection systems. It emphasizes the ability of these algorithms to analyze vast datasets, discern intricate patterns, and adapt in real time, thereby staying one step ahead of fraudulent activities. Furthermore, exploration underscores how machine learning approaches are indispensable for addressing the inherent imbalance in fraudulent detection datasets, where legitimate transactions vastly outnumber fraudulent transactions.

Delving deeper into this chapter, the focus moves towards a detailed exposition of Machine and Deep Learning credit card fraud detection algorithms. These algorithmic gems have emerged as the vanguard for contemporary fraud detection, boasting remarkable capabilities in identifying and mitigating fraudulent transactions. This chapter discusses the intricate architecture and function of these algorithms, elucidating their strengths and limitations in the context of credit card fraud detection.

6.1. Ileberi et al. - A machine learning based credit card fraud detection using the GA algorithm for feature selection

Ileberi et al. (2022) conducted a study in which they utilized the Genetic Algorithm (GA) as a machine learning approach for feature selection in the realm of credit card fraud detection. The GA algorithm, drawing inspiration from natural selection and genetics, serves as a robust optimization method designed to identify and retain the most pertinent attributes within a dataset, while eliminating those that are unimportant or duplicative.

Credit card fraud detection presents several challenges. First, traditional rule-based systems are limited in their ability to adapt and learn from new patterns emerging in fraudulent activities. Second, credit card transactions generate vast amounts of data every day, making it difficult for human analysts to manually review each transaction for potential fraud indicators. Therefore, an automated approach that can accurately detect fraudulent transactions is necessary.

The importance of feature selection cannot be overstated when it comes to enhancing the effectiveness and efficiency of credit card fraud detection systems. By selecting relevant features while eliminating irrelevant ones from the dataset, we can reduce computational complexity and improve model performance.

Prior investigations have delved into a range of machine learning methods aimed at detecting credit card fraud. These methods encompass supervised classification approaches like logistic regression, decision trees, support vector machines (SVM), neural networks, and ensemble strategies such as random forests.

Although these methods have demonstrated potential in effectively identifying fraudulent transactions, they frequently encounter challenges such as elevated rates of false positives or computational inefficiency, primarily attributed to the extensive array of attributes found within credit card transaction datasets. In this investigation, we aim to address these limitations by integrating the Genetic Algorithm (GA) for feature selection specifically in the context of credit card fraud detection.

The approach used for feature selection in credit card fraud detection utilizes a genetic algorithm (GA), which is a methodology that is guided by the principles of population-driven optimization and is inspired by the mechanisms of natural evolution. The process commences by establishing an initial population comprising potential solutions, each of which represents distinct feature subsets extracted from the dataset. These solutions undergo reproduction,

mutation, and crossover operations to generate new offspring populations that inherit favorable characteristics from their parents.

Candidate solutions undergo assessment using appropriate evaluation metrics like accuracy, precision, recall, and the F1-score. Those solutions demonstrating superior performance are chosen as "parents" for the subsequent generation, while those exhibiting subpar performance are removed from consideration. This iterative procedure persists until convergence is achieved or a predefined termination condition is satisfied.

In the study authored by Ileberi and colleagues, the primary focus revolves around the application of machine learning techniques in the realm of credit card fraud detection. Specifically, their approach integrates the Genetic Algorithm (GA) for the purpose of feature selection. Genetic algorithms, commonly referred to as GAs, represent a widely employed method within the field of machine learning for feature selection. The essence of GAs lies in their ability to simulate natural selection and evolutionary processes. This is accomplished by iteratively selecting and combining features based on their fitness, ultimately working towards enhancing the overall performance of the model.

The significance of feature selection in the development of effective credit card fraud detection models cannot be overstated. It entails the crucial task of singling out the most pertinent attributes or features from the available dataset, attributes that play a substantial role in distinguishing fraudulent transactions from legitimate ones.

In their research, the scholars undertook a comparative analysis, pitting their GA-based approach against several state-of-the-art methods such as Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), Artificial Neural Network (ANN), and Naive Bayes (NB) for fraud detection. These experiments were conducted using an authentic credit card transaction dataset sourced from an undisclosed financial institution. The research utilized a dataset comprising credit card transactions carried out by European cardholders during a two-day period in September 2013. Notably, this dataset exhibited a significant class imbalance issue, with a mere 0.172% of the transactions classified as fraudulent. To address this class imbalance challenge, the Synthetic Minority Oversampling Technique (SMOTE) was employed.

Ileberi et al.'s paper aims to enhance existing fraud detection systems by improving their accuracy and reducing false positive rates.

This study makes a significant contribution by incorporating a Genetic Algorithm as a tool for feature selection. Genetic Algorithms are heuristic search methods inspired by natural selection.

In this specific application, we employ the GA algorithm to identify the most pertinent and meaningful features within credit card transaction data. The importance of feature selection lies in its ability to decrease the dataset's dimensionality, potentially enhancing the efficacy of machine learning models and diminishing computational overhead.

In the preprocessing phase, several techniques were employed to enhance data quality and mitigate potential sources of bias or noise that might impact model performance. This included the application of normalization and the removal of outliers.

Furthermore, to bolster the reliability of the study's findings, cross-validation methodologies were seamlessly integrated. This approach involved the division of the dataset into multiple subsets, serving the dual purpose of training and testing. The overarching aim was to prevent overfitting issues while guaranteeing that each subset encompassed a balanced representation of both fraudulent and legitimate transactions.

The assessment of model performance encompassed the utilization of a range of performance metrics, such as accuracy, precision-recall curves, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC). These metrics in combination enabled a thorough evaluation of the performance of various models across diverse evaluation criteria.

Furthermore, to gauge the effectiveness of the proposed approach, direct comparisons were drawn between the results obtained from the GA algorithm and those achieved by traditional machine learning algorithms.

The experimental results showcased the outstanding performance of the GA-based models. For instance, GA-RF achieved an exceptional overall accuracy of 99.98%, while other classifiers, such as GA-DT, also demonstrated remarkable accuracy, reaching 99.92%. These results clearly surpassed the performance of existing methods.

Notably, GA-DT exhibited an impressive AUC of 1, accompanied by a perfect accuracy of 100%. Similarly, the GA-ANN model achieved a high AUC of 0.94 and a flawless accuracy rate of 100%. These findings underscore the effectiveness of the GA algorithm in enhancing model performance, as evidenced by both accuracy and AUC metrics.

The findings of this study are significant, as they demonstrate the potential of machine learning algorithms, specifically employing genetic algorithms for feature selection, in enhancing credit card fraud detection systems. By leveraging big data and machine learning techniques such as the GA algorithm, financial institutions can better protect their customers from fraudulent activities while minimizing false positives and preserving a seamless user experience.

However, it is important to note that this study has certain limitations. The experiments were conducted on a specific dataset with its own characteristics, which might limit generalizability to other datasets or scenarios. Additionally, further research is needed to explore additional evaluation metrics and compare against state-of-the-art techniques in credit card fraud detection.

In conclusion, credit card fraud detection using machine learning algorithms is crucial in today's digital society. This study has shown that employing the GA algorithm for feature selection enhances model performance by selecting relevant features while discarding irrelevant ones. By accurately identifying fraudulent activities, financial institutions can protect their customers' assets and reduce monetary losses due to fraudulent transactions. Future research should focus on validating these findings on different datasets and exploring advanced evaluation metrics for more comprehensive analysis.

6.2. Carcillo et al. - Combining unsupervised and supervised learning in credit card fraud detection

Credit card fraud is a prevalent issue in the financial industry, posing substantial challenges for both individuals and organizations. The rapid advancement of technology has made it easier for criminals to engage in fraudulent activities, necessitating the development of robust fraud detection techniques. In their paper titled "Combining Unsupervised and Supervised Learning Approaches for Credit Card Fraud Detection," Carcillo, Le Borgne, Caelen, Kessaci, Oblé, and Bontempi (2021) propose a novel method that combines unsupervised and supervised learning algorithms to enhance credit card fraud detection accuracy.

Unsupervised learning constitutes a subset of machine learning, focusing on the training of models using unlabeled datasets to reveal underlying patterns and deviations within the data. This methodology does not depend on pre-established categories or tags; instead, it endeavors to recognize atypical behaviors or transactions through statistical analysis. In the context of credit card fraud detection, unsupervised learning techniques prove valuable in identifying outliers or irregular patterns that might signal potential fraudulent actions.

Several unsupervised learning algorithms are commonly employed in fraud detection tasks. One such algorithm is clustering, which groups similar data points together based on their

characteristics. Anomalies or outliers can then be identified as instances that do not belong to any cluster or fall into sparsely populated clusters.

Another popular technique is outlier detection using density estimation methods such as Local Outlier Factor (LOF) or Isolation Forests (IF). These methods calculate the density of each data point compared to its neighbors and flag those with low densities as potential anomalous instances.

Supervised learning entails training models using labeled data where each instance belongs to a specific class or category. In credit card fraud detection, this typically involves creating a classifier model trained on historical transaction records tagged as either genuine or fraudulent.

Numerous supervised learning techniques have found application in the realm of fraud detection. Logistic regression, Support Vector Machines (SVM), and Random Forests are frequently utilized for their proficiency in managing imbalanced datasets and achieving precise instance classification.

While there is potential in the application of unsupervised learning methods for the identification of credit card fraud, these approaches are confronted by various constraints and shortcomings. One significant challenge is the high number of false positives generated during anomaly detection. Anomalies detected by unsupervised algorithms may not always correspond to actual fraudulent transactions, leading to unnecessary investigations and inconvenience for customers.

Additionally, unsupervised learning approaches struggle with adapting to evolving fraud patterns. As criminals continuously devise new tactics, these algorithms may fail to detect emerging fraudulent activities that do not match previously identified patterns.

The authors begin by acknowledging the importance of credit card fraud detection, emphasizing the need for effective and adaptive methods in an evolving landscape of fraudulent activities. Traditional rule-based systems often fall short in capturing complex patterns of fraud, making machine learning an attractive alternative.

Carcillo et al. (2021) proposed a novel approach that merges elements of unsupervised and supervised learning methodologies. This inventive strategy harnesses the advantages of both approaches to improve the precision and effectiveness of credit card fraud detection systems. The unsupervised component, notably Isolation Forest and Local Outlier Factor (LOF) algorithms, is responsible for identifying potentially fraudulent transactions, while the supervised component, utilizing Gradient Boosting Machines (GBM), further refines the predictions.

The incorporation of unsupervised learning techniques is a notable feature of this research. Isolation Forest and LOF are well-suited for anomaly detection in credit card transactions. Isolation Forest isolates anomalies efficiently by creating binary partitions in the data, while LOF measures the local density deviation of a data point compared to its neighbors. These algorithms excel in identifying outliers, which are often indicative of fraudulent transactions.

The supervised learning component, employing GBM, refines the initial unsupervised results. GBM is a powerful ensemble learning technique known for its ability to handle imbalanced datasets and provide accurate predictions. By integrating supervised learning into the process, the authors aim to reduce false positives and enhance the overall detection accuracy.

Their method utilizes an Autoencoder neural network for anomaly detection, which reconstructs input data while minimizing reconstruction error. Instances with high reconstruction errors are flagged as potential outliers or anomalies. These instances are subsequently passed on to a classifier trained on labeled data for classification into genuine or fraudulent transactions.

To evaluate the effectiveness of their proposed method, Carcillo et al. conducted experiments using publicly available credit card transaction datasets from real-world scenarios. They compared the performance of their combined approach against traditional unsupervised and supervised methods.

The authors utilized evaluation metrics such as precision, recall, F1 score, receiver operating characteristic (ROC) curve analysis, and area under the curve (AUC) values to assess the accuracy of their model predictions across different thresholds.

Carcillo et al.'s experiments demonstrated that their combined approach outperformed both unsupervised and supervised methods individually, achieving higher precision, recall, and F1 scores. The Autoencoder-based anomaly detection step effectively reduced false positives compared to standalone unsupervised algorithms.

The authors observed that the combined approach was particularly effective in detecting previously unseen fraudulent patterns by leveraging the generalization capabilities of the supervised classifier model. This adaptability is crucial in mitigating emerging fraud threats.

In comparison to existing methods for credit card fraud detection, Carcillo et al.'s approach offers several advantages. By combining unsupervised and supervised techniques, their method addresses the limitations faced by each individual method, resulting in improved accuracy, and reduced false positives.

Furthermore, their two-step process allows for a more focused investigation of potential anomalies flagged by the unsupervised algorithm, saving time and resources in manual review processes.

In conclusion, Carcillo et al. (2021) make a significant contribution to the field of credit card fraud detection with their innovative hybrid approach. By seamlessly integrating unsupervised and supervised learning techniques, they show promise in enhancing the accuracy and efficiency of fraud detection systems. Their research highlights the significance of capitalizing on the strengths of various machine learning paradigms to tackle complex real-world problems effectively. Future research may explore further refinements and extensions of this hybrid methodology while continuing to adapt to the evolving landscape of credit card fraud.

6.3. Ghosh et al. – Comparative analysis of applications of machine learning in credit card fraud detection

The realm of financial security places significant emphasis on the identification and prevention of credit card fraud, constituting a pivotal domain for research and practical utilization. In light of ongoing technological advancements, credit card fraud detection has evolved into a more intricate field. This evolution has underscored the importance of devising and applying advanced methodologies for fraud detection. Notably, machine learning techniques have risen to prominence as valuable instruments for bolstering the precision and efficacy of detection systems. These techniques have demonstrated their effectiveness in scrutinizing substantial datasets and discerning discernible trends indicative of fraudulent activities.

Prior research has demonstrated the importance of machine learning algorithms in addressing credit card fraud (Strelcenia & Prakoonwit, 2022). These algorithms leverage historical transaction data to train models that can detect anomalies or patterns indicative of fraudulent activity. By utilizing these algorithms, financial institutions can improve their ability to detect suspicious transactions accurately.

Ghosh et al. (2023) open their paper by emphasizing the importance of credit card fraud detection, framing it as a critical challenge in the financial industry. They acknowledge the detrimental impact of fraudulent activities on both financial institutions and consumers, underlining the need for effective and adaptive detection methods.

The authors delve into the role of machine learning as a powerful approach to addressing credit card fraud. They recognize the limitations of rule-based systems and the potential of machine learning algorithms to capture intricate patterns of fraud. Ghosh et al. provide a valuable context by explaining why machine learning is well-suited for this task.

A central focus of Ghosh et al.'s (2023) research is the comparative analysis of various machine learning algorithms and techniques in the context of credit card fraud detection. They systematically review and evaluate the performance of different methods, including supervised, unsupervised, and hybrid approaches. This comprehensive analysis sheds light on the strengths and weaknesses of each technique.

In their research, Ghosh and colleagues introduced a novel approach aimed at enhancing the precision of credit card fraud detection while safeguarding data privacy. Their innovative strategy involves the utilization of a neural network (ANN) within the context of federated learning. This framework offers a viable resolution to the challenge of maintaining data confidentiality, particularly in sectors like banking and finance, where stringent regulations like the General Data Protection Regulation (GDPR) restrict the sharing of data on centralized servers.

Ghosh et al. introduced a novel hybrid model that employs a two-tiered training process. Initially, real-time data is trained locally on edge devices. Subsequently, the resulting model is disseminated centrally to be utilized by various banks and research centers. This strategy significantly improves the precision of identifying fraudulent transactions while safeguarding the confidentiality of user data.

The researchers employ various machine learning algorithms such as logistic regression, decision trees, random forests (RF), support vector machines (SVM), naive Bayes classifiers, k-nearest neighbors (k-NN), neural networks (ANN), gradient boosting machines (GBM), and XGBoost. These algorithms are trained on the dataset to develop models capable of identifying fraudulent transactions.

Random Forest, as utilized in credit card fraud detection, has proven effective in predicting class of regression problems and performing well on limited datasets. However, its limitations become apparent in real-time scenarios where the algorithm's performance is slower, and it takes a longer time to make predictions. Furthermore, it lacks the capability of training large volumes of data effectively.

In contrast, the utilization of Artificial Neural Networks (ANNs) exhibits significant potential, primarily attributed to their capacity to handle extensive datasets and their decentralized memory architecture. ANNs have demonstrated effective performance when integrated with diverse functions and algorithms.

The application of the Support Vector Machine (SVM) method involves the classification of consumer behavior into two categories: fraudulent or non-fraudulent transactions. SVM demonstrates high accuracy when applied to datasets with limited features. Nevertheless, its effectiveness diminishes in real-time scenarios, particularly when handling extensive datasets.

They provide insights into how these techniques handle credit card transaction data and their respective strengths in mitigating false positives and negatives.

The paper also explores the application of unsupervised learning techniques like clustering and anomaly detection in credit card fraud detection. The authors discuss the advantages of algorithms like k-means clustering and Isolation Forest in identifying irregularities within transaction data.

Ghosh et al. (2023) recognize the potential of hybrid approaches that combine both supervised and unsupervised techniques to leverage the benefits of both paradigms. They investigate the performance of these hybrid models and their ability to enhance detection accuracy.

To assess the efficacy of different machine learning methodologies, the researchers utilize a diverse set of performance measures, encompassing metrics like accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC). These evaluative tools facilitate a thorough examination of model performance and facilitate valuable comparisons.

By comparing these metrics across multiple algorithms, Ghosh et al. can identify which techniques offer the highest level of fraud detection accuracy and efficiency. Their framework provides valuable knowledge into the strengths and weaknesses of various machine learning approaches for credit card fraud detection.

The study conducted by Ghosh et al. offers noteworthy insights into the effectiveness of diverse machine learning algorithms when applied to the realm of credit card fraud detection. Their research outcomes reveal that specific algorithms consistently exhibit superior performance in metrics such as accuracy, precision, recall, and F1 score when compared to their counterparts.

For instance, neural networks demonstrate superior performance compared to other algorithms such as logistic regression or naive Bayes classifiers. This finding highlights the potential benefits of utilizing deep learning techniques when detecting fraudulent activities in credit card transactions.

The results presented by Ghosh et al. provide valuable insights into applications of machine learning in credit card fraud detection (Ghosh et al., 2022). The superiority of neural networks suggests that more complex models with deeper layers can effectively capture intricate patterns indicative of fraudulent behavior.

These findings align with previous research conducted in this area (Strelcenia & Prakoonwit, 2022), emphasizing the importance of employing advanced machine learning techniques for accurate fraud detection.

While Ghosh et al.'s study offers valuable contributions to understanding applications of machine learning in credit card fraud detection (Ghosh et al., 2022), some limitations and challenges should be acknowledged. The researchers may have encountered limitations in terms of the dataset used, potentially causing bias or incomplete representation of fraudulent transactions. Additionally, the chosen machine learning algorithms may not encompass all possible approaches, limiting the scope of their analysis.

Based on the gaps identified by Ghosh et al.'s study (Ghosh et al., 2022), future research directions could focus on exploring more advanced deep learning architectures for credit card fraud detection. Additionally, incorporating ensemble methods that combine multiple machine learning algorithms could enhance overall model performance and further improve fraud detection accuracy.

It is also recommended to address potential privacy concerns associated with utilizing personal transaction data for fraud detection purposes. Ensuring proper data anonymization techniques and complying with privacy regulations will be crucial in maintaining consumer trust while effectively detecting fraudulent activities.

The research conducted by Ghosh et al. (2023) offers a valuable contribution to the field of credit card fraud detection. Their comparative analysis of machine learning applications provides insights into the strengths and weaknesses of different techniques, facilitating informed decisions for financial institutions and researchers. By highlighting the potential of various approaches and emphasizing the significance of machine learning in combating credit card fraud, this paper guides future efforts in enhancing detection systems. As the landscape of fraud

continues to evolve, ongoing research in this area remains crucial, and Ghosh et al.'s work provides a solid foundation for further exploration and refinement of machine learning-based fraud detection methodologies.

In conclusion, credit card fraud detection is a pressing issue that necessitates effective solutions. Ghosh et al.'s comparative analysis of applications of machine learning in credit card fraud provides a comprehensive understanding of the effectiveness and limitations of different machine learning techniques. Their proposed hybrid solution, which combines the ANN method with a federated learning framework, shows promise in enhancing the accuracy of credit card fraud detection while ensuring data privacy. Their findings highlight the superiority of neural networks and emphasize the importance of employing advanced machine learning techniques to accurately detect fraudulent transactions. This research contributes to ongoing efforts aimed at improving credit card fraud prevention through innovative machine learning approaches.

6.4. Mondal et al. - Handling Imbalanced Data for Credit Card Fraud Detection

Identifying instances of credit card fraud stands as a vital responsibility within the contemporary digital landscape, characterized by the prevalence of online financial transactions. However, this task presents significant challenges due to the presence of imbalanced data. The problem of imbalanced data in fraud detection is a prevalent issue, as fraudulent transactions are typically much less frequent than genuine ones. Frequently, this disparity tends to result in a skewed focus of the predictive model on the more prevalent class (legitimate transactions), leading to inadequate detection of the less common class (fraudulent transactions).

The study carried out by Mondal et al. (2021) addresses the challenge of fraud detection with a focus on managing imbalanced data.

Imbalanced data is a common issue faced by researchers and practitioners when dealing with credit card fraud detection. The occurrence of imbalanced data can be attributed to several factors, such as the rarity of fraudulent activities compared to legitimate transactions and underreporting due to fear or inconvenience among victims. The implications of imbalanced data in credit card fraud detection are severe since traditional classification algorithms tend to favor the majority class, leading to inadequate identification and prediction of fraudulent transactions.

Several techniques have been proposed over the years to address the imbalance issue in credit card fraud detection. These techniques include oversampling, undersampling, and hybrid methods that combine both approaches.

Oversampling involves replicating minority instances or generating synthetic ones using various algorithms like Synthetic Minority Over-sampling Technique (SMOTE). This technique helps balance out the distribution between minority and majority classes but may lead to overfitting or redundancy if not applied carefully.

Undersampling, on the other hand, aims at reducing samples from the majority class by random selection or clustering methods like Tomek links or Edited Nearest Neighbor rule. While undersampling can effectively address imbalance issues, it runs the risk of losing important information present in majority class examples.

Hybrid methods combine oversampling and undersampling techniques to create a balanced dataset. These methods aim to leverage the strengths of both approaches while minimizing their limitations. One example is the SMOTEENN algorithm, which first applies SMOTE to oversample the minority class and then utilizes ENN (Edited Nearest Neighbor) to remove noisy instances.

These techniques can be specifically applied to credit card fraud detection by balancing the dataset before training machine learning models. This ensures that fraudulent transactions receive adequate representation during model training, leading to improved performance in detecting credit card fraud.

When evaluating the performance of fraud detection models, several metrics are commonly used, including accuracy, precision, recall, and F1-score. Accuracy evaluates the general accuracy of a model's predictions, whereas precision assesses the fraction of accurately identified fraudulent transactions among all transactions predicted as fraudulent. Recall, alternatively referred to as sensitivity or true positive rate (TPR), gauges the model's ability to detect real fraudulent transactions within the entire set of actual fraudulent transactions within the dataset. The F1-score amalgamates precision and recall into a unified metric that achieves a balance between these two performance aspects.

These evaluation metrics play a crucial role when dealing with imbalanced datasets since traditional accuracy measures may not provide an accurate picture due to skewed class distributions. Instead, focusing on specific metrics like recall or F1-score can give better insights into how well a model performs in identifying rare events such as credit card fraud.

Mondal et al.'s paper titled "Handling Imbalanced Data for Credit Card Fraud Detection" (2021) proposes an approach that tackles imbalanced data using modified focal loss within an XGBoost framework (Trisanto et al., 2021). The authors assert that the utilization of focal loss offers a solution to the challenges arising from class imbalance in machine learning. This method involves assigning greater importance, in terms of weight, to minority class samples that are misclassified during the training phase. The authors acknowledge the prevailing scenario in which a vast majority of transactions are legitimate, with a significantly smaller proportion being fraudulent. This imbalance presents a formidable obstacle for conventional machine learning algorithms, which typically exhibit a bias toward the majority class and encounter difficulties in identifying infrequent instances of fraud.

The authors delve into the consequences of imbalanced data, emphasizing the potential for high false negative rates in fraud detection systems. False negatives are particularly problematic in this context, as they allow fraudulent activities to go undetected, leading to substantial financial losses.

Mondal et al. (2021) review various techniques designed to address class imbalance, with a focus on resampling methods. They discuss oversampling of the minority class (fraudulent transactions) and undersampling of the majority class (legitimate transactions). The authors explore the benefits and drawbacks of these techniques in mitigating class imbalance.

While the paper provides valuable insights into handling imbalanced data, it would be beneficial to see a comparison with other over-sampling techniques. Furthermore, the paper could also benefit from exploring the effect of varying the ratio of synthetic samples generated by SMOTE to further optimize the model performance.

One notable component of their research involves investigating techniques for generating synthetic data, including the utilization of the Synthetic Minority Over-sampling Technique (SMOTE). The primary objective of these methods is to produce artificial instances of the underrepresented class in order to achieve dataset balance. The authors offer valuable insights into the impact of these approaches on enhancing model performance.

Mondal et al. reported that utilizing both SMOTE and various machine learning algorithms led to a substantial enhancement in the effectiveness of fraud detection. Notably, the Random Forest algorithm exhibited superior performance across precision, recall, and F1-score metrics.

This work contributes to the field by demonstrating a successful approach to handling imbalanced data in credit card fraud detection. By combining SMOTE with machine learning

algorithms, Mondal et al. have presented a model that can improve the detection of fraudulent transactions, thereby enhancing the security of credit card transactions.

In their experimental setup, Mondal et al. used a publicly available credit card fraud dataset and compared the performance of their proposed approach with other techniques discussed earlier. They highlight the significance of various metrics such as precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC) for evaluating the efficacy of credit card fraud detection models. Their findings demonstrate that their methodology outperformed conventional XGBoost models that did not incorporate focal loss, showcasing superior performance in terms of recall, precision, and F1-score.

While Mondal et al.'s approach shows promising results, it is important to consider its limitations. One potential drawback is that the modified focal loss technique may lead to longer training times due to higher computational requirements. Additionally, the effectiveness of this approach may vary depending on the specific characteristics of different datasets or types of fraud patterns.

When comparing Mondal et al.'s approach with other relevant research papers addressing imbalanced data in credit card fraud detection (Trisanto et al., 2021), it becomes evident that there are multiple strategies available for handling imbalance issues in this domain. Each approach has its strengths and weaknesses, which need to be carefully considered based on the specific context and dataset characteristics.

The central contribution of the paper lies in its exploration of methods to handle class imbalance in credit card fraud detection. By systematically reviewing and evaluating these techniques, Mondal et al. offer valuable guidance for practitioners and researchers in the field. Their work serves as a practical resource for improving the accuracy of fraud detection systems.

Mondal et al.'s (2021) research provides a significant contribution to the field of credit card fraud detection by addressing the critical issue of imbalanced data. Their comprehensive review and evaluation of sampling techniques and synthetic data generation methods offer insights into the strategies available to improve detection accuracy in the face of class imbalance. As credit card fraud continues to evolve, handling imbalanced data remains a key challenge, and Mondal et al.'s work equips researchers and practitioners with valuable tools to address this challenge effectively. Their findings have implications for the development of more robust and accurate fraud detection systems, ultimately benefiting both financial institutions and consumers.

In conclusion, imbalanced data poses significant challenges in credit card fraud detection. Several strategies have emerged to tackle this problem, encompassing oversampling, undersampling, and combinations of both. The assessment of model performance on imbalanced datasets heavily relies on evaluation metrics like accuracy, precision, recall, and the F1-score. Mondal et al.'s paper provides a valuable contribution to the field of credit card fraud detection by addressing the significant issue of imbalanced data. The proposed model, which combines SMOTE utilizing modified focal loss within an XGBoost framework shows promise in handling imbalance data for credit card fraud detection but also has limitations that need further investigation.

7. Conclusions

In conclusion, this thesis has undertaken a comprehensive exploration of the vast and dynamic field of credit card fraud detection through an extensive literature review. By delving into the key themes, methodologies, and technologies that have shaped the landscape of credit card fraud detection, this study has provided valuable insights into the evolution of this critical domain.

This thesis began by establishing the significance of credit card fraud detection in the current digital age, emphasizing the increasing need for robust and adaptive detection systems. It then navigated through a rich body of literature, highlighting the evolution of fraud detection techniques from traditional rule-based systems to more advanced machine learning and artificial intelligence-based approaches. Throughout this journey, the thesis showcased the strengths and limitations of various methodologies, including GA-RF, GA-DT, GA-ANN, and the use of Generative Adversarial Networks (GANs) with Modified Focal Loss and Random Forest as the base machine learning algorithm.

While these methods have shown promise in improving credit card fraud detection, they are not without limitations. For instance, GA-ANN, while offering a data-driven and adaptive approach, can be computationally expensive and require extensive parameter tuning. GANs with Modified Focal Loss and Random Forest integration, on the other hand, may face challenges in interpretability and require substantial computational resources.

Additionally, the research identified the definition of various outlier scores with consideration of different levels of granularity and their integration into supervised approaches. These outlier scores have provided new perspectives on anomaly detection, yet their effectiveness in real-world scenarios and scalability may require further investigation.

Furthermore, exploring Artificial Neural Networks (ANNs) in the context of federated learning is a promising avenue for improving privacy in credit card fraud detection. However, it is important to address the challenges related to model aggregation, communication efficiency, and security in federated learning settings.

This literature review also underscores the importance of collaboration between academia and industry to stay ahead in the cat-and-mouse game of fraud detection. It highlights the necessity of adapting to new fraud schemes and techniques, as fraudsters continually evolve their tactics.

In summary, this thesis offers a comprehensive overview of the credit card fraud detection landscape, from its historical roots to its current state and future prospects. It is hoped that this research will serve as a valuable resource for scholars, practitioners, and policymakers in the field of financial security, guiding them towards more effective and innovative approaches to combatting credit card fraud in an increasingly digital world. Because the ever-evolving nature of fraud presents new challenges and opportunities, this literature review lays the foundation for future research and advancements in the domain of credit card fraud detection, considering the latest developments and their associated limitations.

Bibliography

[Alam, 2021] Alam, B. (2021). Supervised Learning Algorithms explained [Beginners Guide] | GoLinuxCloud. *GoLinuxCloud*. https://www.golinuxcloud.com/supervised-learning-algorithms/

[Alamri & Ykhlef, 2022] Alamri, M., & Ykhlef, M. (2022). Survey of Credit card anomaly and Fraud Detection using sampling techniques. *Electronics*, *11*(23), 4003. https://doi.org/10.3390/electronics11234003

[AltexSoft, 2020] *AltexSoft*. (2020, February 27). Fraud Detection: How machine learning systems help reveal scams in fintech, healthcare, and eCommerce. *AltexSoft*. https://www.altexsoft.com/whitepapers/fraud-detection-how-machine-learning-systems-help-reveal-scams-in-fintech-healthcare-and-ecommerce/

[Anderson, 2022] Anderson, R.. (2002). Why Information Security is Hard-An Economic Perspective. 358 - 365. 10.1109/ACSAC.2001.991552.

[Bahnsen, 2016] Bahnsen, A. (2016). Introduction to Example-Dependent Cost-Sensitive Classification — costcla documentation. http://albahnsen.github.io/CostSensitiveClassification/Intro.html

[Bajaj, 2021] Bajaj, V. (2021, December 15). Unsupervised Learning For Anomaly Detection - Towards Data Science. *Medium*. https://towardsdatascience.com/unsupervised-learning-for-anomaly-detection-44c55a96b8c1

[Bhandari, 2023] Bhandari, A. (2023). Feature Engineering: Scaling, Normalization, and Standardization (Updated 2023). *Analytics Vidhya*. https://www.analyticsvidhya.com/blog/2020/04/feature-scaling-machine-learning-normalization-standardization/

[Borgne, 2022] Borgne, Y. L. (2022, January 6). Machine Learning for Credit card fraud Detection - towards data science. Medium. https://towardsdatascience.com/machine-learning-for-credit-card-fraud-detection-a-jupyter-book-for-reproducible-research-8ca5edad7b5d

[Boukerche, 2020] Boukerche, A., Zheng, L., & Alfandi, O. (2020). Outlier detection. *ACM Computing Surveys*, *53*(3), 1–37. https://doi.org/10.1145/3381028

[Bronshtein, 2023] Bronshtein, A. (2023). Train/Test split and cross validation in Python - towards data science. *Medium*. https://towardsdatascience.com/train-test-split-and-cross-validation-in-python-80b61beca4b6

[Brownlee, 2016] Brownlee, J. (2016). What is a Confusion Matrix in Machine Learning. *MachineLearningMastery.com*. https://machinelearningmastery.com/confusion-matrix-machine-learning/

[Brownlee, 2018] Brownlee, J. (2018). How to use ROC curves and Precision-Recall curves for classification in Python. *MachineLearningMastery.com*. https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-classification-in-python/

[Brownlee, 2020] Brownlee, J. (2020). A tour of machine learning algorithms. *MachineLearningMastery.com*. https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/

[Brownlee, 2020b] Brownlee, J. (2020). Cost-Sensitive learning for imbalanced classification. *MachineLearningMastery.com*. https://machinelearningmastery.com/cost-sensitive-learning-for-imbalanced-classification/

[Brownlee, 2020b] Brownlee, J. (2020). How to Handle Missing Data with Python. *MachineLearningMastery.com*. https://machinelearningmastery.com/handle-missing-data-python/

[Brownlee, 2020c] Brownlee, J. (2020). Ordinal and One-Hot encodings for categorical data. *MachineLearningMastery.com*. https://machinelearningmastery.com/one-hot-encoding-for-categorical-data/

[Brownlee, 2020d] Brownlee, J. (2019). How to Choose a Feature Selection Method for Machine Learning. *MachineLearningMastery.com*. https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/

[Brownlee, 2020e] Brownlee, J. (2020). 6 dimensionality reduction Algorithms with Python. *MachineLearningMastery.com*. https://machinelearningmastery.com/dimensionality-reduction-algorithms-with-python/

[Brownlee, 2020f] Brownlee, J. (2020). Tune hyperparameters for classification machine learning algorithms. *MachineLearningMastery.com*. https://machinelearningmastery.com/hyperparameters-for-classification-machine-learning-algorithms/

[Brownlee, 2021] Brownlee, J. (2021). How to Manually optimize Machine Learning model hyperparameters. *MachineLearningMastery.com*. https://machinelearningmastery.com/manually-optimize-hyperparameters/

[Burns, 2023] Burns, C. (2023). What is 3-D Secure? Fraud Prevention Solution Explained. *Chargebacks911*. https://chargebacks911.com/3-d-secure/

[Carcillo et al., 2021] Carcillo, F., Le Borgne, Y., Caelen, O., Kessaci, Y., Oblé, F. and Bontempi, G. (2021). Combining unsupervised and supervised learning in credit card fraud detection. Information Sciences. https://doi.org/10.1016/j.ins.2019.05.042

[Cerqueira, 2022] Cerqueira, V. (2022). Monte Carlo Cross-Validation for time series - towards data science. *Medium*. https://towardsdatascience.com/monte-carlo-cross-validation-for-time-series-ed01c41e2995

[Chaturvedi, 2023] Chaturvedi, N. (2023). Leave-One-Out Cross-Validation - DataDrivenInvestor. *Medium*. https://medium.datadriveninvestor.com/leave-one-out-cross-validation-32fa248c1739

[Cheliatsidou, 2021] Cheliatsidou, A., Sariannidis, N., Garefalakis, A., Azibi, J., & Kagias, P. (2021). The international fraud triangle. *Journal of Money Laundering Control*, *26*(1), 106–132. https://doi.org/10.1108/jmlc-09-2021-0103

[Corbo, 2023] Corbo, A. (2023). What is supervised learning? *Built In*. https://builtin.com/machine-learning/supervised-learning

[Databricks, 2019] *Detecting Financial Fraud at Scale with Decision Trees and MLflow on Databricks*. (2019, May 2). Databricks. https://www.databricks.com/blog/2019/05/02/detecting-financial-fraud-at-scale-with-decision-trees-and-mlflow-on-databricks.html

[David, 2020] David, D. (2020). Hyperparameter Optimization Techniques to Improve Your Machine Learning Model's Performance. *freeCodeCamp.org*. https://www.freecodecamp.org/news/hyperparameter-optimization-techniques-machine-learning/

[De Jesus, 2019] De Jesus, A. (2019). Machine Learning for Credit Card Fraud – 7 Applications for Detection and Prevention. *Emerj Artificial Intelligence Research*. https://emerj.com/ai-sector-overviews/machine-learning-for-credit-card-fraud/

[Decorte, 2023] Decorte, D. (2023). Friendly Fraud cost: What do you lose to chargeback abuse? *Chargebacks911*. https://chargebacks911.com/friendly-fraud-costs/

[Dekanovsky, 2021] Dekanovsky, V. (2021). Complete guide to Python's cross-validation with examples. *Medium*. https://towardsdatascience.com/complete-guide-to-pythons-cross-validation-with-examples-a9676b5cac12

[Di Stefano, 2022] Di Stefano, A. (2022, August 30). Machine learning for fraud detection: fighting crime with algorithms. *itransition*. https://www.itransition.com/machine-learning/fraud-detection

[Diadiushkin et al., 2019] Diadiushkin, A., Sandkuhl, K., & Maiatin, A. (2019b). Fraud Detection in Payments Transactions: Overview of existing approaches and usage for instant payments. *Complex Systems Informatics and Modeling Quarterly*, *20*, 72–88. https://doi.org/10.7250/csimq.2019-20.04

[Dutta et al., 2010] Dutta, Amitava; Roy, Rahul; and Seetharaman, Priya, "Drivers of Knowledge Contribution in Open Fora: Findings from Wikipedians" (2010). AMCIS 2010 Proceedings. 441. https://aisel.aisnet.org/amcis2010/441

[EBANX, 2023] EBANX. (2023). What is 3D Secure (3DS)? | Payments Explained. https://www.ebanx.com/en/resources/payments-explained/3d-secure/

[ECB, 2022] European Central Bank. (2022, December 20). *Study on the payment attitudes of consumers in the euro area (SPACE) – 2022*. https://www.ecb.europa.eu/stats/ecb_surveys/space/html/ecb.spacereport202212~783ffdf46e.en.html

[Emmanuel, 2021] Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., & Tabona, O. (2021). A survey on missing data in machine learning. *Journal of Big Data*, *8*(1). https://doi.org/10.1186/s40537-021-00516-9

[Finextra, 2019] Finextra. (2019). TD Bank: Treasury professionals anticipate benefits from blockchain. Finextra Research. https://www.finextra.com/pressarticle/77223/td-bank-treasury-professionals-anticipatebenefits-from-blockchain/wholesale

[Fisher, 2015] Fisher, K. (2015). The Psychology of fraud: What motivates fraudsters to commit crime? *Social Science Research Network*. https://doi.org/10.2139/ssrn.2596825

[GeeksforGeeks, 2022] GeeksforGeeks. (2022). Machine learning algorithms Types of machine learning algorithms. *GeeksforGeeks*. https://www.geeksforgeeks.org/machine-learning-algorithms/

[GeeksforGeeks, 2023] GeeksforGeeks. (2023). What is Machine Learning. *GeeksforGeeks*. https://www.geeksforgeeks.org/what-is-machine-learning/

[GeeksforGeeks, 2023b] GeeksforGeeks. (2023). Cross validation in machine learning. *GeeksforGeeks*. https://www.geeksforgeeks.org/cross-validation-machine-learning/

[GeeksforGeeks, 2023b] GeeksforGeeks. (2023). Guide to AUC ROC Curve in Machine Learning. *GeeksforGeeks*. https://www.geeksforgeeks.org/auc-roc-curve/

[GeeksforGeeks, 2023c] GeeksforGeeks. (2023). Confusion Matrix in machine learning. *GeeksforGeeks*. https://www.geeksforgeeks.org/confusion-matrix-machine-learning/

[Ghosh & Reilly, 1994] S. Ghosh, D.L. Reilly. (1994). Credit card fraud detection with a neural network, in: Proceedings of the Annual International Conference on System Science.

[Ghosh et al., 2023] Ghosh, S. Bilgaiyan, S. Gourisaria, K. and harma, A. (2023). Comparative Analysis of Applications of Machine Learning in Credit Card Fraud Detection. https://doi.org/10.1109/ISCON57294.2023.10112099

[Gillis, 2022] Gillis, A. S. (2022). data splitting. *Enterprise AI*. https://www.techtarget.com/searchenterpriseai/definition/data-splitting

[Glen, 2020] Glen, S. (2020). *Receiver Operating Characteristic (ROC) curve: Definition, example - Statistics How to*. Statistics How To. https://www.statisticshowto.com/receiver-operating-characteristic-roc-curve/

[Google, 2022] Google for Developers. (2022). Classification: ROC Curve and AUC. https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc

[Hu, 2023] Hu, W. (2023). Monte Carlo Simulation with Python - Wendy Hu - Medium. *Medium*. https://medium.com/@whystudying/monte-carlo-simulation-with-python-13e09731d500

[IBM, 2023] *IBM. (2023). What is Supervised Learning? IBM*. https://www.ibm.com/topics/supervised-learning

[Ileberi et al., 2022] Ileberi, E., Sun, Y., & Wang, Z. (2022). A machine learning based credit card fraud detection using the GA algorithm for feature selection. Journal of Big Data, 9(1). https://doi.org/10.1186/s40537-022-00573-8

[Jain, 2018] Jain, V. (2018). Confusion & Cost Matrix helps in calculating the accuracy, cost and various other measurable factors in classification problem. *Medium*. https://medium.com/@inivikrant/confusion-cost-matrix-helps-in-calculating-the-accuracy-cost-and-various-other-measurable-a725fb6b54e1

[Kaggle, 2018] *Credit card fraud Detection*. (2018, March 23). Kaggle. https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

[Kaggle, 2019] IEEE-CIS Fraud Detection | Kaggle. https://www.kaggle.com/c/ieee-fraud-detection

[Kanika, 2021] Kanika, J. Singla and Nikita. (2021). "Comparing ROC Curve based Thresholding Methods in Online Transactions Fraud Detection System using Deep Learning," 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2021, pp. 9-12, doi: 10.1109/ICCCIS51004.2021.9397167.

[Kropelnytsky & Vidjikant, 2023] Kropelnytsky, Y., & Vidjikant, S. (2023, March 29). An overview of detecting and preventing credit card fraud by using new technology. Softjourn Inc. https://softjourn.com/insights/detecting-and-preventing-credit-card-fraud

[Kulatilleke, 2022] Kulatilleke, G. K. (2022, August 20). Challenges and Complexities in Machine Learning based Credit Card Fraud Detection. arXiv.org. https://arxiv.org/abs/2208.10943

[Kumar, 2021] Kumar, S. (2021). Data splitting technique to fit any Machine Learning Model. *Medium*. https://towardsdatascience.com/data-splitting-technique-to-fit-any-machine-learning-model-c0d7f3f1c790

[Kumar, 2023] Kumar, N. (2023). Dimensionality reduction technique. *Spark by {Examples}*. https://sparkbyexamples.com/machine-learning/dimensionality-reduction-technique/

[Leung, 2023] Leung, K. (2023). Financial Fraud Detection with AutoXGB - Towards Data Science. *Medium*. https://towardsdatascience.com/autoxgb-for-financial-fraud-detection-f88f30d4734a

[Liu et al., 2021] G. Liu, J. Tang, Y. Tian and J. Wang, "Graph Neural Network for Credit Card Fraud Detection," 2021 International Conference on Cyber-Physical Social Intelligence (ICCSI), Beijing, China, 2021, pp. 1-6, doi: 10.1109/ICCSI53130.2021.9736204.

[Lucas, 2020] *Lucas, Y. (2020, October 13). Credit card fraud detection using machine learning: A survey. arXiv.org. https://arxiv.org/abs/2010.06479*

[Maharana, 2022] Maharana, K., Mondal, S., & Nemade, B. I. (2022). A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*, *3*(1), 91–99. https://doi.org/10.1016/j.gltp.2022.04.020

[MIT, 2021] *MIT Sloan*. (2021, April 21). *Machine learning, explained.* MIT Sloan. https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained

[Mohari et al., 2021] Mohari, A., Dowerah, J., Das, K., Koucher, F., & Bora, D. J. (2021). Credit Card Fraud Detection Techniques: A review. In Springer eBooks (pp. 157–166). https://doi.org/10.1007/978-981-16-1048-6_12

[Mondal et al., 2021] Mondal I., Haque, M., Hassan, A., Shatabda, S. (2021). Handling Imbalanced Data for Credit Card Fraud Detection. 24th International Conference on

Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2021, pp. 1-6, https://10.1109/ICCIT54785.2021.9689866

[Muschelli, 2019] Muschelli, J. (2019). ROC and AUC with a Binary Predictor: a Potentially Misleading Metric. *Journal of Classification*, *37*(3), 696–708. https://doi.org/10.1007/s00357-019-09345-1

[nbshare, 2023] nbshare. (2023). Learn and Code Confusion Matrix with Python. https://www.nbshare.io/notebook/626706996/Learn-And-Code-Confusion-Matrix-With-Python/

[Nguyen, 2020] Nguyen, T. T. (2020, December 7). Deep learning methods for credit card fraud Detection. arXiv.org. https://arxiv.org/abs/2012.03754v1

[Nigam, 2023] Nigam, S. (2023, August 11). What is Machine Learning? Definition, Features, Types | Leverage Edu. *Leverage Edu*. https://leverageedu.com/discover/general-knowledge/science-and-technology-what-is-machine-learning/

[Ogunbiyi, 2022] Ogunbiyi, I. A. (2022). How to handle missing data in a dataset. *freeCodeCamp.org*. https://www.freecodecamp.org/news/how-to-handle-missing-data-in-a-dataset/

[Opus Consulting, 2021] Opus, Inspiring Payment Innovation. (2021). Rule-Based vs. Machine Learning: Effective Fraud Prevention Models. *Opus | Inspiring Payment Innovation*. https://www.opusconsulting.com/rule-based-vs-machine-learning-effective-fraud-prevention-models/

[PayPal, 2023] PayPal. (2023). Online payment processing guide for small business owners. https://www.paypal.com/us/brc/article/how-online-payments-processing-works

[Pozzolo, 2015] Pozzolo, A. D. (2015). *Adaptive machine learning for credit card fraud detection*. https://www.semanticscholar.org/paper/Adaptive-Machine-Learning-for-Credit-Card-Fraud-Pozzolo-Bontempi/bcfbf068dff507b9ef11240e69f96d24f5d89fc1

[Pykes, 2021] Pykes, K. (2021, December 16). Using Machine Learning To Detect Fraud - Towards Data Science. *Medium*. https://towardsdatascience.com/using-machine-learning-to-detect-fraud-f204910389cf

[Ranjan, 2021] Ranjan, G. K. (2021). Introduction to k-fold Cross-Validation in Python. *SQLRelease*. https://sqlrelease.com/introduction-to-k-fold-cross-validation-in-python

[Ravelin Blog, 2023] Ravelin. (2023). Why is payment fraud so persistent? Ravelin Blog. https://www.ravelin.com/blog/why-is-payment-fraud-so-persistent

[Ravelin Insights, 2023] Ravelin. (2023). Machine learning for fraud detection. Ravelin. https://www.ravelin.com/insights/machine-learning-for-fraud-detection

[Ravelin, 2023] Ravelin. (2023). Insights Online payment fraud. Ravelin. https://www.ravelin.com/insights/online-payment-fraud

[Saito & Rehmsmeier, 2015] Saito, T., & Rehmsmeier, M. (2015). The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLOS ONE*, *10*(3), e0118432. https://doi.org/10.1371/journal.pone.0118432

[Satpathy, 2023] Satpathy, S. (2023). SMOTE for Imbalanced Classification with Python. *Analytics Vidhya*. https://www.analyticsvidhya.com/blog/2020/10/overcoming-class-imbalance-using-smote-techniques/

[Scikit-learn, 2023] Scikit-learn. (2023). *Precision-Recall*. https://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html

[Scikit-learn, 2023b] Scikit-learn. (2023). 3.1. Cross-validation: evaluating estimator performance. (2023). https://scikit-learn.org/stable/modules/cross_validation.html

[Skillfloor, 2023] Skillfloor. (2023, August 18). Fraud Detection with Machine Learning: Identifying Anomalies. *Medium*. https://medium.com/@skillfloor_29561/fraud-detection-with-machine-learning-identifying-anomalies-a42952aa08f9

[Softjourn, 2022] Softjourn. (2022, September 27). Pattern recognition. *Softjourn Inc*. https://softjourn.com/insights/pattern-recognition

[Steen, 2021] Steen, D. (2021). Precision-Recall Curves - Doug Steen - medium. *Medium*. https://medium.com/@douglaspsteen/precision-recall-curves-d32e5b290248

[Stripe, 2023] Stripe. (2023). Common types of online fraud. Stripe Documentation. https://stripe.com/docs/disputes/prevention/fraud-types

[Sulaiman et al., 2022] Sulaiman, R. B., Schetinin, V., & Sant, P. (2022). Review of Machine Learning Approach on Credit Card Fraud Detection. Human-centric Intelligent Systems, 2(1–2), 55–68. https://doi.org/10.1007/s44230-022-00004-0

[Team, 2023] Team, A. (2023). Machine Learning In Fraud Detection: An In-Depth Analysis – Avenga. *Avenga*. https://www.avenga.com/magazine/fraud-detection-machine-learning/

[Tharwat, 2020] Tharwat, A. (2020). Classification assessment methods. *Applied Computing and Informatics*, *17*(1), 168–192. https://doi.org/10.1016/j.aci.2018.08.003

[The scikit-yb developers, 2019] Precision-Recall Curves — Yellowbrick v1.5 documentation. (2019). https://www.scikit-yb.org/en/latest/api/classifier/prcurve.html

[Vadapalli, 2020] Vadapalli, P. (2020). Top 10 Dimensionality Reduction Techniques For Machine Learning. *upGrad blog*. https://www.upgrad.com/blog/top-dimensionality-reduction-techniques-for-machine-learning/

[Varun Kumar K S et al., 2020] Varun Kumar K S, Vijaya Kumar V G , Vijayshankar A , Pratibha K, 2020, Credit Card Fraud Detection using Machine Learning Algorithms, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 07 (July 2020). https://doi.org/10.17577/ijertv9is070649

[Wikipedia, 2023] Wikipedia contributors. (2023). Machine learning. *Wikipedia*. https://en.wikipedia.org/wiki/Machine_learning

[Wu et al., 2021] Wu, X., Liu, X., & Zhou, Y. (2021). Review of unsupervised learning techniques. In *Springer eBooks* (pp. 576–590). https://doi.org/10.1007/978-981-16-6324-6_59

[Xue, 2019] Xue, Y. (2019). An Overview of Overfitting and its Solutions. *Journal of Physics*, *1168*, 022022. https://doi.org/10.1088/1742-6596/1168/2/022022

[Zach, 2022] Zach. (2022). Misclassification rate in Machine Learning: definition & example. *Statology*. https://www.statology.org/misclassification-rate/

[Fine, 2023] Fine, R. (2023, October 19). PSD2, SCA, and 3DS2: Understanding The Basics Of European Regulations. https://www.spreedly.com/blog/psd2-sca-3ds2-basics