**Credit Card Fraud Detection Using Machine Learning**



Elizabeth Biju Arackal

20027845

Dublin Business School

This dissertation is submitted for the degree of

*Master of Science in Business Analytics*

*Supervisor name: Monika Smatralova*

January 2025

**Declaration**

I confirm that the content presented in this dissertation is entirely my own work, except where explicit reference is made to the contributions of others. This work has not been submitted, in full or in part, for assessment towards any degree or qualification at this institution or elsewhere. All collaborative or external inputs have been duly acknowledged within the text and in the Acknowledgements section.

Signed: Elizabeth Biju Arackal

Student Number: 20027845

Date: 06/05/2025

**Acknowledgements**

I would like to express my heartfelt thanks to my supervisor, Monika Smatralova, at Dublin Business School, for his invaluable support, insightful guidance, and encouragement throughout this project. His expertise and constructive feedback were instrumental in shaping the direction and outcome of this research. I am also grateful to Dublin Business School for the opportunity to undertake the Master of Science in Business Analytics program and for providing a learning environment enriched with academic resources and support. A special thank you goes to my family and friends, whose continuous motivation, patience, and belief in me have been a constant source of strength during this academic journey. Their encouragement has played a vital role in helping me stay focused and resilient. This project has been a rewarding and enriching experience, and I extend my sincere appreciation to everyone who contributed to its completion.

**ABSTRACT**

With the growth of digital payments, credit card fraud has become more sophisticated and harder to detect. This study applies advanced machine learning models to tackle the challenge and improve fraud detection accuracy in real-world scenarios. Credit card fraud remains a major concern for financial institutions due to its rising complexity and financial impact. This project explores the use of machine learning techniques CatBoost, XGBoost, and LightGBM to detect fraudulent transactions in a highly imbalanced dataset. After thorough preprocessing and feature engineering, models were trained and evaluated using precision, recall, F1-score, and confusion matrix. CatBoost emerged as the most effective model, offering a balanced performance in identifying fraud while minimizing false positives. The study also provides a detailed comparison of feature importance and addresses ethical, legal, and social considerations.

**CONTENTS**

ABSTRACT

LIST OF FIGURES

LIST OF TABLES

[CHAPTER 1 1](#_Toc197285884)

[INTRODUCTION 1](#_Toc197285885)

[1.1 BACKGROUND 1](#_Toc197285886)

[1.2 CHALLENGES IN DETECTING CREDIT CARD FRAUD 2](#_Toc197285887)

[1.3 AIM 3](#_Toc197285888)

[1.4 PROBLEM STATEMENT 3](#_Toc197285889)

[1.5 PROJECT OBJECTIVES 4](#_Toc197285890)

[1.6 RESEARCH QUESTIONS 4](#_Toc197285891)

[1.7 STRUCTURE OF THESIS 4](#_Toc197285892)

[CHAPTER 2 6](#_Toc197285893)

[RELATED WORKS 6](#_Toc197285894)

[2.1 INTRODUCTION 6](#_Toc197285895)

[2.1.1 Search Strategy 6](#_Toc197285896)

[2.1.2 Search Terms 6](#_Toc197285897)

[2.1.3 Search Eligibility Criteria 7](#_Toc197285898)

[2.1.4 Exclusion Criteria 7](#_Toc197285899)

[2.2 RELATED WORK 8](#_Toc197285900)

[2.3 LITERATURE GAPS 12](#_Toc197285901)

[CHAPTER 3 15](#_Toc197285902)

[ETHICAL CONSIDERATIONS 15](#_Toc197285903)

[3.1 ETHICAL CONSIDERATIONS 15](#_Toc197285904)

[3.2 SOCIAL CONSIDERATIONS 16](#_Toc197285905)

[3.3 LEGAL CONSIDERATIONS 17](#_Toc197285906)

[CHAPTER 4 19](#_Toc197285907)

[METHODOLOGY 19](#_Toc197285908)

[4.1 METHODOLOGY FRAMEWORK 19](#_Toc197285909)

[4.2 DATA COLLECTION 20](#_Toc197285910)

[4.3 DATA PREPROCESSING 21](#_Toc197285911)

[4.4 EDA 23](#_Toc197285912)

[4.5 FEATURE ENGINEERING 33](#_Toc197285913)

[4.6 MODELLING 35](#_Toc197285914)

[4.6.1 CatBoost - Categorical Boosting 35](#_Toc197285915)

[4.6.2 XGBoost - Extreme Gradient Boosting 36](#_Toc197285916)

[4.6.3 Light GBM (Light Gradient Boosting Machine) 37](#_Toc197285917)

[4.7 HYPER PARAMETER TUNING 37](#_Toc197285918)

[4.8 PERFORMANCE EVALUATION 38](#_Toc197285919)

[CHAPTER 5 39](#_Toc197285920)

[RESULT ANALYSIS AND DISCUSSIONS 39](#_Toc197285921)

[5.1 IMPLEMENTATION DETAILS 39](#_Toc197285922)

[5.2 RESULT ANALYSIS 40](#_Toc197285923)

[5.2.1 Cat Boost 40](#_Toc197285924)

[5.2.2 XGBoost 41](#_Toc197285925)

[5.2.3 LightGBM 42](#_Toc197285926)

[5.3 FEATURE IMPORTANCE 44](#_Toc197285927)

[5.4 CRITICAL ANALYSIS 45](#_Toc197285928)

[5.5 LIMITATIONS 46](#_Toc197285929)

[CHAPTER 6 48](#_Toc197285930)

[CONCLUSION 48](#_Toc197285931)

[6.1 CONCLUSION 48](#_Toc197285932)

[6.2 FUTURE WORKS 48](#_Toc197285933)

[REFERENCES 50](#_Toc197285934)

**LIST OF FIGURES**

[Figure 4. 1 Proposed Framework 19](#_Toc197285954)

[Figure 4. 2 Dataset Details 21](#_Toc197285955)

[Figure 4. 3 Null values 22](#_Toc197285956)

[Figure 4. 4 Unique values 23](#_Toc197285957)

[Figure 4. 5 Label Distribution 24](#_Toc197285958)

[Figure 4. 6 Category Distribution by is\_fraud' 25](#_Toc197285959)

[Figure 4. 7 Gender Distribution by is\_fraud' 26](#_Toc197285960)

[Figure 4. 8 Fraud Amt distribution 27](#_Toc197285961)

[Figure 4. 9 Fraud city pop distribution 28](#_Toc197285962)

[Figure 4. 10 Monthly Fraud Transactions 28](#_Toc197285963)

[Figure 4. 11 Seasonal Fraud Transactions 29](#_Toc197285964)

[Figure 4. 12 Average Hourly Fraud Counts 30](#_Toc197285965)

[Figure 4. 13 Average Daily Fraud Transactions 31](#_Toc197285966)

[Figure 4. 14 Scatter Plot of 'amt' 32](#_Toc197285967)

[Figure 4. 15 Outlier Analysis Using Scatter Plot 32](#_Toc197285968)

[Figure 4. 16 Confusion Matrix 38](#_Toc197285969)

[Figure 5. 1 CatBoost – Confusion Matrix 40](#_Toc197285970)

[Figure 5. 2 CatBoost – Classification Report 41](#_Toc197285971)

[Figure 5. 3 XGBoost– Confusion Matrix 42](#_Toc197285972)

[Figure 5. 4 XGBoost – Classification Report 42](#_Toc197285973)

[Figure 5. 5 LightGBM – Classification Report 43](#_Toc197285974)

[Figure 5. 6 LightGBM – Confusion Matrix 43](#_Toc197285975)

[Figure 5. 7 Top 10 Important Feature 44](#_Toc197285976)

**LIST OF TABLES**

[Table 1 Model Performance comparison 46](#_Toc197285993)

# CHAPTER 1

# INTRODUCTION

## 1.1 BACKGROUND

Fraud is not a new phenomenon; it has persisted throughout human history, continually adapting to changing economic and technological landscapes. In the modern financial ecosystem, driven by the proliferation of digital platforms and online transactions, fraud has become significantly more sophisticated. The motivations chiefly financial gain remains constant, but the methods have evolved, exploiting technological vulnerabilities and systemic loopholes.High-profile financial collapses, such as the Enron scandal in 2001 and the downfall of Lehman Brothers in 2008, highlight the catastrophic consequences that financial fraud can unleash not just for corporations, but for entire economies. These cases revealed how even heavily regulated firms with professional audits could manipulate earnings and conceal massive debts, raising serious concerns about transparency, accountability, and the robustness of regulatory frameworks (Lakis & Giriūnas, 2012)

Today, financial fraud spans a wide spectrum, from credit card and insurance fraud to complex money laundering operations. While certain monetary losses may be recovered, indirect costs such as chargebacks, operational disruptions, and long-term reputational damage are often irreparable and absorbed by institutions and consumers alike. As a result, the need for efficient, adaptive fraud detection systems has become more pressing than ever.Traditional rule-based fraud prevention systems, although useful in structured environments, struggle to keep pace with the dynamic and rapidly changing strategies employed by cybercriminals. These static systems often fail to identify novel or subtle forms of fraud. Consequently, financial institutions are increasingly shifting toward data-driven methods, particularly anomaly detection and machine learning (ML), which enable the identification of suspicious patterns in real-time and offer greater adaptability in high-volume, complex data environments (Sailusha et al., 2020).

As digital financial services expand, the financial sector faces escalating challenges in safeguarding transactions. Criminals continually devise new tactics to exploit system vulnerabilities, inflicting substantial losses on both organizations and individuals. Although prevention mechanisms are essential, post-transactional fraud detection has become equally critical to limiting damage, restoring trust in digital systems, and enhancing the responsiveness of financial institutions (Hilal et al., 2022). Modern ML-based systems significantly outperform traditional methods by offering superior accuracy, flexibility, and the ability to adapt to evolving fraud patterns. This has made machine learning a focal point for both researchers and industry practitioners seeking to strengthen fraud detection frameworks (Alarfaj et al., 2022).

## 1.2 CHALLENGES IN DETECTING CREDIT CARD FRAUD

Detecting fraud in credit card transactions presents significant challenges due to the dynamic and adversarial nature of fraudulent behaviour, as well as the characteristics of transaction data itself. As global credit card usage continues to surge, so does the volume of transactions, complicating efforts to flag fraudulent activity in real time. Fraudsters, constantly innovating, exploit system blind spots and modify their techniques to evade detection rendering static detection systems increasingly obsolete (Kulatilleke, 2022). A major challenge in this domain is the severe class imbalance in the data: fraudulent transactions typically account for less than 1% of the total transaction volume. This imbalance can bias machine learning models toward predicting the dominant (non-fraud) class, resulting in a high rate of false negatives, where actual fraud goes undetected (He & Garcia, 2009). This has a direct impact on the model’s utility in practice, as undetected fraud can lead to significant financial and reputational consequences.

Another issue is the limited availability of real-world, high-quality datasets. Owing to privacy regulations and the sensitive nature of financial data, institutions are often hesitant to share raw transaction records. When data is made publicly available, it is frequently anonymized or transformed (e.g., via Principal Component Analysis) to preserve confidentiality. While such preprocessing is essential for compliance, it may inadvertently affect the performance of ML models by obscuring meaningful patterns. Data quality itself is another critical factor. Errors in labelling such as incorrectly classifying a legitimate transaction as fraudulent or vice versa can mislead models during training, impairing their predictive accuracy. In addition, there is currently no standardized benchmark for evaluating the performance of fraud detection models. Variations in metrics, datasets, and evaluation protocols across studies make it difficult to assess and compare model efficacy objectively. Tackling credit card fraud is not solely about building more intelligent algorithms. It requires addressing foundational data issues, ensuring the availability of high-quality, representative datasets, and establishing robust and consistent evaluation frameworks.

## 1.3 AIM

The aim of this project is to develop and evaluate a comprehensive credit card fraud detection framework.

## 1.4 PROBLEM STATEMENT

Detecting credit card fraud is challenging due to highly imbalanced datasets and the evolving nature of fraudulent behaviour, requiring the development of robust machine learning models and feature analysis to ensure accurate, scalable, and real-world applicable fraud detection systems.

## 1.5 PROJECT OBJECTIVES

* Collect and preprocess a credit card transaction dataset for fraud detection.
* Perform exploratory data analysis (EDA) to understand data distribution and identify patterns.
* Develop and compare machine learning models and optimize model performance through hyperparameter tuning.
* Evaluate model performance using robust, task-specific metrics
* Identify and analyze the top 10 features contributing to fraud prediction.

## 1.6 RESEARCH QUESTIONS

* Which of these models provides the highest accuracy and reliability in identifying fraudulent transactions?
* What preprocessing and data balancing techniques are most effective in improving model performance on skewed credit card transaction data?
* How does hyperparameter tuning influence the performance of the selected models in fraud detection tasks?
* Which features have the greatest impact on predicting fraudulent behavior in credit card transactions?

## 1.7 STRUCTURE OF THESIS

The report is structured as follows:

Chapter 1 introduces the problem of credit card fraud, its significance, challenges in detection, and outlines the research aims, objectives, and questions.

Chapter 2 review the existing literature on fraud detection methods, highlighting key models, strategies, and research gaps in the domain.

Chapter 3 discusses the ethical, social, and legal implications of using machine learning for fraud detection, including privacy, fairness, and accountability.

Chapter 4 describes the dataset used, the preprocessing techniques, feature engineering steps, and the models implemented for fraud detection.

Chapter 5 presents the evaluation of CatBoost, XGBoost, and LightGBM models, compares their performance, and analyzes the top contributing features.

Chapter 6 summarizes key findings, acknowledges limitations, and proposes directions for future research and real-world implementation.

# CHAPTER 2

# RELATED WORKS

## 2.1 INTRODUCTION

In the chapter a complete review of existing works on credit card fraud detection using machine learning techniques. It highlights commonly used models, key challenges like data imbalance, and identifies gaps that this study aims to address.

### 2.1.1 Search Strategy

To ensure a comprehensive and relevant literature review, a systematic search strategy was employed to identify peer-reviewed studies, scholarly articles, and reputable reports related to credit card fraud detection using machine learning. The strategy involved a structured approach covering database selection, search techniques, keyword formulation, and the use of inclusion/exclusion criteria. The focus was specifically on scholarly works published from 2019 to 2024, to capture recent developments, algorithms, and evaluation strategies relevant to this rapidly evolving field.

### 2.1.2 Search Terms

The search strategy involved querying multiple academic databases known for their high-quality, peer-reviewed content. These included IEEE Xplore, ACM Digital Library, ScienceDirect, SpringerLink, Scopus, and Google Scholar. Advanced search options are used within each database to apply filters based on publication year, subject domain, and document type. Boolean operators (AND, OR), quotation marks for phrase searches, and truncation symbols are employed to refine and broaden the search results.

The following search terms and keyword combinations are used:

* "Credit card fraud detection" AND "machine learning"
* "Financial fraud" AND "AI models"
* "Credit card anomaly detection" OR "fraud classification"
* "XGBoost" OR "LightGBM" OR "CatBoost" AND "credit card fraud"
* "Imbalanced dataset" AND "fraud detection"

These search terms helped identify literature focused on model development, class imbalance solutions, model evaluation, and real-world applications of AI in fraud detection.

### 2.1.3 Search Eligibility Criteria

* Publications dated from 2019 to 2024 to ensure recency and relevance.
* Papers must be peer-reviewed, including journal articles, conference proceedings, or preprints from trusted repositories like arXiv.
* Research must focus on credit card fraud detection using machine learning, deep learning, or AI-based techniques.
* Studies must include a clear methodology, the use of realistic or benchmark datasets, and performance evaluation metrics (e.g., precision, recall, AUC).
* Only papers written in English are considered for consistency and accessibility.

### 2.1.4 Exclusion Criteria

* Papers published before 2019, unless highly cited and foundational.
* Studies focusing solely on rule-based systems without ML or AI components.
* Articles not directly addressing credit card or financial fraud detection (e.g., healthcare fraud, cyber intrusion).
* Research that lacks experimental results, performance metrics, or detailed model validation.
* Non-peer-reviewed sources, blog posts, news articles, and content from predatory or unverified journals.
* Duplicate publications or extended versions of already included studies.

## 2.2 RELATED WORK

Detecting credit card fraud is essential in protecting financial assets and upholding customer trust two key priorities for financial institutions. As fraudulent schemes become more advanced and harder to trace, there is growing pressure to develop detection systems that can effectively identify and prevent such activities before they result in significant financial losses. The complexity of modern digital transactions means that traditional systems often fall short, especially in handling card-not-present (CNP) transactions where existing deep learning models struggle to quantify uncertainty in predictions (Awosika et al., 2023; Habibpour et al., 2021).

The consequences of undetected fraud can be far-reaching. Beyond financial damages, institutions risk losing customer confidence in the security of digital payment systems. Moreover, in the growing space of Machine Learning as a Service (MLaaS), accountability becomes critical if services misrepresent their capabilities or deliver unreliable results, it can lead to major financial and reputational setbacks (Canim et al., 2019).

An effective fraud detection system must not only detect anomalies but also handle nonlinear relationships and high-cardinality categorical variables in a way that remains interpretable and practical for financial analysts and compliance teams (Raymaekers et al., 2021). At the same time, identity verification (IDV) plays a central role in supporting fraud prevention by ensuring that users interacting with the system are who they claim to be, thereby adding an additional layer of trust and security (Vaidya & Awasthi, 2025).

With the expansion of digital financial ecosystems, the tools and methods used to combat fraud must evolve accordingly. Machine learning, predictive modeling, and data analysis have proven to be highly effective in this space, helping to identify suspicious behavior based on complex, ever-changing patterns in transactional data. Various machine learning approaches ranging from supervised and unsupervised to hybrid techniques are now being used to uncover hidden anomalies and patterns, improving detection capabilities significantly (Shapira & Schuster, 2022).

Recent work has explored the use of user-dependent neural sequence models, which incorporate user-specific behavior patterns into prediction models. These models have shown improved accuracy by tailoring detection to individual behavior, rather than applying a one-size-fits-all approach (Boyd et al., 2020). Improvements in optimization techniques, such as those involving weak error analysis in stochastic gradient descent (SGD), have also contributed to the robustness and efficiency of model training processes (Jentzen et al., 2021).

Integrating advanced algorithms with robust modeling and data analysis has led to systems that not only detect fraud more accurately but also offer transparency and adaptability. This has been a major shift in the field of fraud analytics, extending its application beyond credit card transactions to areas like insurance claims and digital identity verification. One of the main challenges continues to be data imbalance the relatively low proportion of fraudulent transactions compare d to legitimate ones. This, along with the limited availability of public datasets, has pushed researchers to develop novel machine learning solutions capable of learning meaningful representations despite data limitations (Bockel-Rickermann et al., 2023; Kulatilleke, 2022; Psychoula et al., 2021).

Supervised learning methods have become particularly prominent in credit card fraud detection, leveraging labelled transaction data to distinguish between legitimate and fraudulent activity. These models have shown considerable success in identifying patterns within historical data and applying them to new, unseen data—making them an essential component of modern fraud detection systems (West et al., 2015).

Decision trees, known for their interpretability and flexibility, are commonly used to segment data based on rules that isolate fraudulent transactions. However, their tendency to overfit, especially in high-dimensional data, has led to the use of ensemble techniques like Random Forests and Gradient Boosted Decision Trees (GBDT). A notable approach is WOTBoost, which enhances classification performance by combining weighted oversampling with boosting techniques to better capture minority (fraudulent) classes (Zhang et al., 2019). Other hybrid methods, such as SWOE-SB, integrate shrinkage estimation for categorical features and spline binning for continuous variables, achieving a balance between model interpretability and predictive power (West et al., 2015).

SVMs have proven particularly effective for detecting fraud when the boundary between legitimate and fraudulent transactions is complex or nonlinear. By using kernel functions, SVMs transform input features into higher-dimensional spaces where distinctions become more apparent. Innovative methods, including quantum anomaly detection via quantum kernel PCA and one-class SVMs, have further expanded their capabilities (N. Liu & Rebentrost, 2018). Temporal Knowledge Distillation (TKD) also offers an edge by using previously trained models to guide current classifications in time-sensitive scenarios (Kurshan & Shen, 2023). Still, the scalability of SVMs remains a concern, especially with large transaction datasets.

Logistic regression continues to be a popular choice due to its simplicity and suitability for binary classification tasks. While basic forms of logistic regression face challenges with imbalanced datasets, enhanced versions such as secure and online-efficient logistic regression have improved their suitability for real-time fraud detection (J. Liu et al., 2023). The adoption of GPU-accelerated logistic regression has also improved training efficiency for large-scale data environments, reinforcing its relevance in the fraud detection space.

Statistical and regression-based techniques remain foundational in predictive modeling for fraud detection. Classical methods like Naive Bayes, Decision Trees, K-Neare st Neighbors, and SVMs offer valuable baseline comparisons and are effective in identifying patterns despite class imbalance (Bockel-Rickermann et al., 2023). Advanced sampling methods, such as KDE, help address imbalance by estimating minority class densities and generating synthetic samples to improve classifier performance (Fajardo et al., 2018).

Parallel logistic regression models have also shown promise in reducing computation time, an essential feature for real-time detection systems (Li et al., 2018). Ensemble methods such as DeepBalance, which use deep belief networks trained on balanced bootstrapped samples, have demonstrated strong performance by incorporating random feature selection to improve model robustness (Lunghi et al., 2023).

The effectiveness of fraud detection models relies not only on the algorithms themselves but also on the evaluation criteria used to judge them. Metrics such as precision, recall, F1-score, and AUROC are commonly used to assess performance, while ROC curves help compare models across different thresholds (Kyriienko & Magnusson, 2022). Newer techniques like eLDA classifiers have shown improved control over false negatives and false positives, enhancing overall reliability (Cheng et al., 2024).

## 2.3 LITERATURE GAPS

Detecting fraudulent transactions in credit card data presents a complex set of challenges due to the high dimensionality, dynamic patterns, and variability of user behavior embedded in transactional data streams. One of the primary issues is the complexity and unpredictability of fraud, which often emerges in forms that are not previously annotated or clearly defined in training data (Shapira & Schuster, 2022). The irregular timing of transactions and distinct, user-specific spending behaviors make it difficult for conventional models to generalize effectively (Boyd et al., 2020). Moreover, the adaptive strategies used by fraudsters require equally adaptive detection methods that can evolve over time to remain effective.

A major technical limitation lies in the use of traditional linear models, such as logistic regression, which often fail to handle nonlinear relationships and high-cardinality categorical variables both of which are common in real-world financial datasets. In addition, identity verification (IDV) systems critical in preventing fraudulent access—face challenges in balancing robust security requirements with the need for a smooth user experience, while simultaneously keeping up with global regulations and emerging threats. On the data quality front, typical issues such as missing values, invalid entries, unknown categories, and wide datasets with too many features can severely hinder the performance and reliability of machine learning models.

To overcome these challenges, recent research emphasizes the importance of adopting advanced machine learning techniques capable of managing complex, high-dimensional data and dynamically responding to new fraud tactics. Robust preprocessing frameworks are also essential to address data inconsistencies and improve model accuracy. Techniques such as ensemble learning, anomaly detection, and real-time classification have shown promise in analysing large-scale transaction data and identifying fraud patterns with greater precision. Furthermore, the use of scalable identity verification methods and explainable AI tools is critical to ensuring both security and transparency in fraud detection systems.

Despite notable progress in this field, several research gaps remain unaddressed. First, there is a significant lack of real-time fraud detection models. Many current systems are tested offline in static environments and are not optimized for immediate action limiting their practical use in stopping fraud before financial damage occurs. Second, the availability of authentic transaction data remains limited. Due to privacy and confidentiality concerns, most datasets used in research are either anonymized or synthetically generated, which may not fully represent the complexity and behaviour of real-world fraud.

Another persistent issue is the over-reliance on accuracy as a performance metric. In the context of highly imbalanced datasets where fraudulent transactions are rare accuracy can be misleading. There is a growing need for studies to adopt more appropriate evaluation metrics, such as precision, recall, F1-score, and the AUC-PR, to reflect true model performance.

Finally, there is underutilization of hybrid models that combine the strengths of supervised and unsupervised learning. While gradient boosting methods like XGBoost and unsupervised models like autoencoders are popular independently, fewer studies have explored how these can be integrated into a unified pipeline for improved performance. Exploring such combinations could lead to systems that are both robust in detection and flexible in adapting to new fraud strategies.

Addressing these gaps requires not only technical innovation but also collaborative efforts across data providers, regulatory bodies, and researchers to create frameworks that are practical, transparent, and secure. The financial sector must continue to invest in adaptive technologies and interdisciplinary research to stay ahead of evolving fraud tactics and protect users in an increasingly digital world.

# CHAPTER 3

# ETHICAL CONSIDERATIONS

## 3.1 ETHICAL CONSIDERATIONS

The use of machine learning models for fraud detection must be guided by ethical principles to ensure the fair and responsible treatment of individuals and their data. Key ethical aspects include:

* **Data Privacy and Confidentiality:**

Credit card transaction data involves highly sensitive personal and financial information, making data privacy a fundamental ethical concern. It is essential to ensure that all user data is anonymized and encrypted to prevent unauthorized access or the potential identification of individuals. Access to such data should be strictly limited to authorized personnel directly involved in model development and analysis.

* **Bias and Fairness:**

Machine learning models are only as unbiased as the data they are trained on. When historical datasets contain imbalances or discriminatory patterns, these can be unintentionally learned and reinforced by the model. This poses a significant risk, as certain demographics might be disproportionately flagged for fraud due to skewed training data. Misclassifications can particularly affect underrepresented or vulnerable groups, potentially leading to unfair treatment or service denial. To ensure fairness, it is important to regularly audit models for biased behaviour across variables such as age, gender, geography, and financial behavior. Employing fairness-aware algorithms and using balanced, representative datasets can help mitigate such risks and promote equitable outcomes.

* **Transparency and Explainability:**

Transparency in decision-making is essential when deploying AI systems in critical domains like finance. Users and stakeholders must be able to understand the rationale behind why a transaction has been flagged as fraudulent.

* **Accountability:**

Despite the advantages of automation, human oversight remains a cornerstone of ethical AI deployment. Fraud detection systems should not operate in isolation, particularly when a transaction’s classification could have serious consequences for the user. Establishing clear accountability structures is also necessary to determine who is responsible for errors, misclassifications, or user grievances. Maintaining a balance between automation and human intervention ensures not only better accuracy but also safeguards user rights and ethical integrity.

## 3.2 SOCIAL CONSIDERATIONS

Machine learning-based fraud detection systems impact not only technical operations but also broader societal dynamics. Social considerations include:

* **Impact on Customers:**

The implementation of fraud detection systems can have a direct impact on customers, particularly when false positives occur. Incorrectly flagging legitimate transactions may lead to blocked payments, frozen accounts, or embarrassing situations for users. Such experiences can be frustrating and inconvenient, especially during urgent transactions. To mitigate this, systems must be designed with a high true-positive rate while minimizing false positives. Additionally, it is essential to provide easily accessible and responsive customer support channels, allowing affected users to appeal or resolve issues promptly and effectively.

* **Public Trust and Acceptance:**

Public trust plays a vital role in the success of fraud detection systems. If customers perceive these systems as overly intrusive or unfair, it can erode their confidence in digital banking platforms. Trust may be compromised when users feel they are being constantly monitored or when fraud alerts are delivered in a vague or impersonal manner. Maintaining this trust requires transparent and empathetic communication, especially when fraud is suspected. Alerts and notifications should be designed to inform users clearly and respectfully, offering guidance on the next steps rather than inducing panic or confusion.

* **Digital Divide and Accessibility**

Not all users possess the same level of digital literacy or familiarity with fraud detection mechanisms. This digital divide can result in confusion or inaction when users encounter fraud alerts or blocked transactions. It is crucial to design fraud detection interfaces that are accessible to all, including support for multiple languages and consideration for users with disabilities.

## 3.3 LEGAL CONSIDERATIONS

The legal implications of developing and deploying machine learning models in financial environments are significant. Key areas include:

* **Data Protection and Privacy Laws:**

Credit card transaction data is governed by strict data protection regulations designed to safeguard users’ personal and financial information. Key legal frameworks such as the General Data Protection Regulation (GDPR) in the European Union and the California Consumer Privacy Act (CCPA) in the United States impose stringent requirements on how personal data is collected, stored, and processed. These include obtaining informed consent from users, allowing them the right to access, modify, or delete their personal data, and restricting data usage strictly to its original intended purpose in this case, fraud detection. Non-compliance with these regulations can result in severe financial penalties and reputational damage, making it imperative for fraud detection systems to be designed with data privacy and legal compliance as core principles.

* **Liability and Accountability:**

In automated fraud detection systems, determining legal accountability can be complex. If a system incorrectly flags a legitimate transaction or blocks access to a user's account, the question arises—who is responsible? Is it the software developer, the financial institution deploying the model, or the third-party service provider? Clear legal guidelines are needed to define responsibility for automated decisions and to ensure that users have the right to contest such decisions. This includes establishing transparent procedures for reviewing misclassifications and implementing appeal mechanisms so that affected individuals can seek redress or clarification.

* **Compliance with Financial Sector Regulations:**

Financial institutions are heavily regulated by government and industry bodies, such as the Reserve Bank of India (RBI), the U.S. Securities and Exchange Commission (SEC), and equivalent authorities in other regions. Fraud detection models must comply with sector-specific anti-fraud mandates and demonstrate adherence to regulatory expectations. This includes ensuring that models are auditable, interpretable, and do not violate ethical or financial standards. Regulators may also require periodic reporting on model performance and evidence of fairness, accuracy, and robustness in detecting fraud without discriminating against specific user groups.

# CHAPTER 4

# METHODOLOGY

## 4.1 METHODOLOGY FRAMEWORK

The methodology adopted for this study follows a structured pipeline designed to ensure effective fraud detection and robust model performance. It begins with data collection, followed by exploratory data analysis (EDA) to uncover patterns, anomalies, and trends within the dataset. The next phase involves data preprocessing, which includes handling missing values, normalizing numerical features, and encoding categorical variables to make the data suitable for machine learning algorithms.

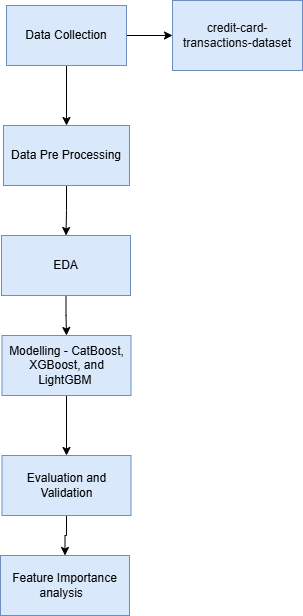


Figure 4. Proposed Framework

The processed data is then used to train and evaluate three advanced models CatBoost, XGBoost, and LightGBM chosen for their ability to handle complex, imbalanced datasets effectively. Hyperparameter tuning is conducted to optimize each model's performance, followed by a comprehensive evaluation using metrics such as precision, recall, F1-score, and AUC. Additionally, a feature importance analysis is performed to identify the top 10 most influential predictors contributing to fraud detection. This end-to-end methodology is visually summarized in Figure 4.1, which presents the proposed model development framework.,

## 4.2 DATA COLLECTION

Credit card fraud datasets play a crucial role in building and testing models that aim to detect fraudulent activity. For this study, the dataset used is titled "Credit Card Transactions Dataset" (*Credit Card Transactions Dataset*, n.d.). It is a simulated dataset designed to closely resemble real-world credit card transactions, containing both legitimate and fraudulent records. The dataset includes features such as transaction amount, time, customer ID, and transaction type all anonymized to protect privacy.

One of the most important fields is the is\_fraud label, which identifies whether a given transaction is fraudulent (1) or not (0). The data is simulated and anonymized, there are limitations when it comes to directly applying the results to live financial systems, though it remains a strong starting point for model development and evaluation. Figure 4.2 displays the structure of the credit card transactions dataset, which contains 1,296,675 entries and 23 columns, including transaction details, customer information, merchant data, and a binary is\_fraud label indicating fraudulent activity.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 4. Dataset Details

## 4.3 DATA PREPROCESSING

Data preprocessing is a vital step in preparing raw data for machine learning tasks, as it ensures the dataset is clean, consistent, and structured appropriately for analysis. This process starts with data cleaning, which includes handling missing values, correcting data inconsistencies, and removing duplicates or outliers that may distort model training. In this dataset, it was confirmed that there are no duplicate rows, indicating a high level of data integrity and reducing the risk of bias in model outcomes. Following this, data transformation techniques such as feature scaling, normalization, and encoding of categorical variables are applied. These transformations help balance feature contributions, standardize inputs, and improve overall model performance. Figure 4.3 illustrates the missing value distribution across all features in the dataset. As shown, most of the columns are complete, with only one feature merch\_zipcode having missing values (195,973 entries). This insight is critical for designing appropriate imputation strategies or deciding whether to drop the feature based on its importance and contribution to the model.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 4. Null values

Figure 4.4 presents the number of unique values for each feature in the credit card transactions dataset, which helps in understanding the diversity and cardinality of the data. Features like trans\_num, trans\_date\_trans\_time, and unix\_time have a high number of unique entries, indicating that each transaction is uniquely identified and timestamped. Categorical features such as merchant (693), category (14), job (494), and state (51) show moderate variability, which is important for model learning without causing high dimensionality. The is\_fraud column has only 2 unique values (0 for legitimate and 1 for fraud), confirming it as a binary classification target. Notably, merch\_zipcode has 28,336 unique values, indicating a very high cardinality, which may require encoding techniques or dimensionality reduction to prevent model overfitting.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 4. Unique values

## 4.4 EDA

Exploratory Data Analysis plays a vital role in understanding the structure, trends, and patterns within a dataset before building any machine learning models. It helps in identifying relationships between features, uncovering anomalies, and detecting potential data quality issues such as outliers or missing values. EDA also provides insights into feature distributions, correlations, and class imbalances factors that significantly affect model performance and interpretation. In fraud detection, understanding the distribution of fraudulent versus legitimate transactions is crucial due to the typically extreme imbalance in real-world datasets. Figure 4.5 shows the distribution of the target variable is\_fraud, which indicates whether a transaction is legitimate (0) or fraudulent (1). From the bar chart, the dataset is highly imbalanced, with 1,289,169 legitimate transactions and only 7,506 fraudulent ones. This imbalance is typical in fraud detection scenarios and poses a challenge for machine learning models, which may become biased towards predicting the majority class.

A graph with a bar

AI-generated content may be incorrect.

Figure 4. Label Distribution

Figure 4.6 illustrates the distribution of credit card transactions across different purchase categories, segmented by the is\_fraud label, which indicates whether the transaction was fraudulent (1) or legitimate (0). The fig shows that both fraudulent and non-fraudulent occur in categories like gas\_transport, grocery\_pos, shopping\_pos, and home. Interestingly, fraudulent transactions are notably higher in categories such as grocery\_pos and shopping\_net, which suggests that these segments might be more vulnerable to fraud. Categories like entertainment, travel, and personal\_care show comparatively fewer fraud cases, which may reflect lower transaction volume or better fraud prevention in these areas.

A graph with numbers and a number of data

AI-generated content may be incorrect.

Figure 4. Category Distribution by is\_fraud'

Figure 4.7 illustrates the distribution of credit card transactions by gender, categorized by the is\_fraud variable. The chart shows the total number of legitimate (0) and fraudulent (1) transactions for each gender Female (F) and Male (M). From the chart, it is observed that females account for a slightly higher number of total transactions (706,128) compare d to males (583,041). However, when it comes to fraud, the counts are quite similar: 3,735 fraudulent transactions for females and 3,771 for males. This indicates that fraudulent activity is relatively balanced between genders, suggesting no significant gender-based bias in fraud occurrence.

The visualization highlights that while transaction volumes differ slightly by gender, the likelihood of fraud is not heavily skewed toward either group. This insight is useful for ensuring fairness in fraud detection models and indicates that gender may not be a strong standalone predictor of fraud.

A bar graph with numbers and text

AI-generated content may be incorrect.

Figure 4. Gender Distribution by is\_fraud'

Figure 4.8 compares the distribution of transaction amounts (amt) between fraudulent and non-fraudulent transactions. It provides valuable insight into spending behavior patterns associated with fraud. The histogram for fraudulent transactions shows a multimodal distribution, indicating that fraudulent amounts tend to cluster around certain values particularly near low amounts (under $100), around $300, and again in the higher range near $900–$1000. This suggests that fraudsters often target both small and moderately high-value transactions, possibly to evade detection or exploit transaction limits. The histogram for non-fraudulent transactions reveals a heavily right-skewed distribution, with most transaction amounts falling below $100. There are a few high-value legitimate transactions extending up to around $30,000, but they are extremely rare. This stark contrast highlights that while legitimate transactions are mostly of low value, fraudulent ones span a broader and more varied amount range. This insight is crucial for building models that can identify suspicious transaction amounts more effectively.

A graph of a function

AI-generated content may be incorrect.

Figure 4. Fraud Amt distribution

Figure 4.9 compares the distribution of city population (city\_pop) for fraudulent and non-fraudulent transactions, helping to assess whether fraud is more prevalent in areas with specific population sizes. The distribution of fraudulent transactions shows a sharp peak in cities with very low populations, suggesting that a significant proportion of fraud incidents occur in smaller or less populated areas. While there are some fraud cases in mid- to high-population cities, their frequency diminishes considerably as the population increases. The non-fraudulent transactions also exhibit a strong skew toward low-population cities, but the overall volume is much higher. The shape of the distribution mirrors that of fraudulent transactions but at a much larger scale, reflecting the fact that most of all transactions fraudulent or not tend to occur in smaller cities, possibly due to the demographic distribution in the dataset. The figure indicates that fraud is not exclusively concentrated in metropolitan areas and may occur more frequently in smaller cities, which could have implications for fraud detection strategies based on geographic and demographic features.

A comparison of a graph

AI-generated content may be incorrect.

Figure 4. Fraud city pop distribution

Figure 4.10 illustrates the trend of monthly fraudulent transactions from January 2019 to June 2020. The bar chart shows fluctuations in fraud activity over time, helping to identify any seasonal or cyclical patterns. The number of fraudulent transactions was relatively high in the early months of 2019, peaking in December 2019 with 592 cases, which represents the highest monthly fraud count during the observed period. After that, there was a noticeable decline in early 2020, reaching a low in April 2020 with 302 cases, before rising again in May 2020.

A graph of blue bars

AI-generated content may be incorrect.

Figure 4. Monthly Fraud Transactions

Figure 4.11 displays the number of fraudulent transactions categorized by season across the years 2019 and 2020, offering a clearer view of potential seasonal trends in credit card fraud activity. In Winter 2019 recorded the highest number of fraudulent transactions (1,615 cases), suggesting a possible spike in fraud during the holiday and year-end shopping season. Summer 2020 shows the lowest fraud count (334 cases), possibly reflecting reduced economic activity or tighter security measures during that period, which may coincide with pandemic-related slowdowns. Other seasons like Spring and Autumn in both years show consistent levels of fraud (around 1,200–1,300 cases), while Summer 2019 also reflects a relatively lower volume compare d to other seasons. The analysis provides important insights for financial institutions, indicating that fraud tends to increase during colder months, particularly around the holiday season, which can guide the timing of intensified monitoring and fraud prevention strategies.

A graph of blue bars with numbers

AI-generated content may be incorrect.

Figure 4. Seasonal Fraud Transactions

Figure 4.12 illustrates the average number of fraudulent transactions occurring at each hour of the day, highlighting distinct peaks during specific time windows. The fig shows two prominent spikes in fraud activity: one between 10 PM and midnight (22:00–23:00), where fraud counts reach their highest average (80.5 and 79.3), and another smaller but consistent cluster from midnight to 3 AM (with averages around 25–27). In contrast, fraud counts are significantly lower throughout the daytime and early evening hours, particularly between 4 AM and 8 PM, where the average fraud rate remains under 5 transactions per hour. This pattern suggests that fraudsters may be more active during late-night hours, possibly exploiting times when users and fraud monitoring systems are less vigilant.

A graph of a number of blue squares

AI-generated content may be incorrect.

Figure 4. Average Hourly Fraud Counts

Figure 4.13 displays the average number of fraudulent transactions occurring on each day of the week, revealing clear variations in fraud activity across weekdays and weekends. The fig shows that fraud incidents peak on weekends, with Saturday (175.3) and Sunday (173.7) having the highest average fraud counts. This trend suggests that fraudsters may be more active when users are potentially less attentive to their accounts or when financial institution oversight might be reduced. Monday also shows a high fraud count (168.9), possibly reflecting delayed processing or reporting from the weekend. Conversely, Wednesday (122.7) and Tuesday (133.6) record the lowest averages, indicating relatively lower fraud activity during midweek. These insights can help guide scheduling for fraud monitoring teams and the allocation of fraud detection resources across the week.

A graph of blue rectangular bars

AI-generated content may be incorrect.

Figure 4. Average Daily Fraud Transactions

Figure 4.14 presents a scatter plot of transaction amounts (amt) plotted against the index of transactions, providing a visual overview of how transaction values are distributed across the entire dataset. The plot reveals that most transactions cluster below $5,000, forming a dense horizontal band. There are numerous outliers scattered across the index with transaction amounts exceeding $10,000 and even reaching up to nearly $30,000. These high-value transactions, although relatively rare, are critical to monitor closely since they could indicate significant fraud attempts or high-risk activity. The scatter plot effectively highlights the presence of extreme values and helps in identifying the need for outlier detection techniques or logarithmic scaling during preprocessing to normalize skewed distributions and improve model stability.

A red dotted graph with numbers

AI-generated content may be incorrect.

Figure 4. Scatter Plot of 'amt'

Figure 4.15 shows an outlier analysis of transaction amounts (amt) using a scatter plot. The chart distinguishes between normal transaction values (in solid red) and outliers (in light pink), based on a defined threshold of $2,700, marked by a dashed horizontal line. Transactions that exceed this threshold are flagged as outliers and visually stand apart from the dense band of normal values below the line. Detecting and handling these outliers correctly can enhance both the robustness and accuracy of machine learning predictions.

A screen shot of a graph

AI-generated content may be incorrect.

Figure 4. Outlier Analysis Using Scatter Plot

## 4.5 FEATURE ENGINEERING

In feature engineering, specific data cleaning and dimensionality reduction steps are applied to enhance model performance and interpretability. For transaction records with an amount (amt) greater than $2,700 are removed from the dataset. This threshold was selected based on outlier analysis, and the decision resulted in the loss of only 430 records, which constitutes just 0.03% of the total dataset—a negligible impact considering the dataset's large size. Removing these high-value outliers helps to stabilize the model and reduce noise without sacrificing meaningful data. A set of columns was identified and removed for being either redundant, high-cardinality, or irrelevant for model training. These include personal identifiers (first, last, street, etc.), high-cardinality location details (city, state, zip, merch\_zipcode), transaction-specific IDs (trans\_num, unix\_time), and an unnamed index column. Dropping these fields not only reduces data dimensionality and risk of overfitting but also ensures privacy protection and regulatory compliance by eliminating personally identifiable information. The streamlined feature set focuses the model on the most relevant and generalizable patterns for fraud detection.

In the feature engineering phase, new variables are derived from existing timestamp and location features to enrich the dataset with more informative attributes for fraud detection. First, the trans\_date\_trans\_time column, which stores the original timestamp of each transaction, was converted into a datetime format. Multiple temporal features are extracted—including the year, month, day, weekday, hour, minute, and second—to capture detailed behavioral patterns in transaction timing. Additionally, a trans\_season variable was created to group transactions into seasons (Winter, Spring, Summer, Autumn) based on the month. These temporal components help identify time-related fraud trends, such as increased activity during weekends, holidays, or late-night hours. After extraction, the original timestamp column was dropped to avoid redundancy.

The date of birth (dob) field was transformed into a datetime format to compute each cardholder’s age at the time of the transaction. This was done by subtracting the year of birth from the transaction year, providing a card\_holder\_age feature, which may influence purchasing behavior or risk profiles. Both dob and the intermediate birth\_year columns are then removed to maintain a clean dataset and ensure data privacy.

Geographic information was leveraged to compute the distance between the customer’s location and the merchant’s location using the geopy library. This calculation uses the latitude and longitude values of both parties and applies the geodesic formula, which accurately accounts for the Earth's curvature. The resulting distance feature can be a strong fraud indicator, as unusually long distances between a cardholder and the point of transaction may signify suspicious activity. These engineered features collectively enhance the model's ability to detect complex fraud patterns by incorporating behavioral, temporal, and spatial context.

In preparation for machine learning model training, categorical features within the dataset are transformed into numerical values using Label Encoding. This is an essential preprocessing step, as most machine learning algorithms require input features to be in numerical format. Label encoding assigns a unique integer to each category in a feature column, preserving the categorical nature of the data while making it machine-readable.

In preparation for machine learning model training, categorical features within the dataset are transformed into numerical values using Label Encoding. This is an essential preprocessing step, as most machine learning algorithms require input features to be in numerical format. Label encoding assigns a unique integer to each category in a feature column, preserving the categorical nature of the data while making it machine-readable.

Since that fraud detection is a highly imbalanced classification problem—with fraudulent transactions forming a very small fraction of the total—the model must be guided to avoid bias toward the majority class (non-fraud). To handle this, class weights are computed using compute\_class\_weight from sklearn.utils. This function calculates weights inversely proportional to class frequencies, ensuring that the minority class (fraudulent transactions) receives more attention during training. The computed weights are stored in a dictionary (class\_weights\_dict) and later passed to the machine learning model during training. This approach helps reduce the likelihood of false negatives (i.e., undetected fraud) and improves overall model performance, especially in detecting rare but critical fraud cases.

## 4.6 MODELLING

To detect fraudulent credit card transactions effectively, three high-performance gradient boosting algorithms were selected for model development and evaluation: CatBoost, XGBoost, and LightGBM. These algorithms are widely used in tabular data problems due to their superior ability to handle non-linear relationships, categorical variables, and class imbalance, while delivering fast, accurate predictions. Each model was trained using the pre-processed dataset and evaluated using metrics suited for imbalanced classification tasks.

### 4.6.1 CatBoost - Categorical Boosting

CatBoost is a gradient boosting algorithm developed by Yandex specifically to handle categorical data more effectively than traditional boosting algorithms. Unlike other models that require explicit one-hot encoding or label encoding of categorical variables, CatBoost natively supports categorical features through a technique known as ordered target statistics. This allows it to convert categories into numerical values while avoiding target leakage. One of the distinguishing features of CatBoost is its use of ordered boosting, which processes data in a way that prevents overfitting, even on small datasets. CatBoost builds symmetric trees, which means that at each level of the decision tree, the same split is used across all branches. This structure leads to faster inference times and a more efficient model representation. Additionally, CatBoost automatically handles missing values and requires minimal hyperparameter tuning to achieve competitive performance. These properties make it particularly suitable for fraud detection tasks, where categorical fields such as merchant category, customer job title, or location are critical and where overfitting on small subsets of fraud cases is a significant risk. In this project, CatBoost served as a highly interpretable and accurate model that handled high-cardinality categorical variables with ease.

### 4.6.2 XGBoost - Extreme Gradient Boosting

XGBoost is one of the most widely used machine learning algorithms for structured data tasks, including fraud detection. It implements a highly optimized form of gradient boosting decision trees (GBDT) with performance enhancements like parallelized tree construction, cache-aware access, and regularization. One of the key advantages of XGBoost is its use of second-order Taylor approximation, which leverages both first and second derivatives of the loss function for more precise gradient updates. This improves convergence speed and accuracy, especially in noisy datasets. XGBoost also supports L1 (Lasso) and L2 (Ridge) regularization, which helps prevent overfitting particularly important when dealing with imbalanced data, as is typical in credit card fraud detection. The algorithm allows users to define custom loss functions and evaluation metrics, making it flexible for scenarios where standard accuracy may be insufficient. In this study, XGBoost was fine-tuned with hyperparameters such as learning rate, maximum depth, and number of estimators. Class weights were applied to ensure that the model paid adequate attention to the minority fraud class, helping to improve recall and reduce false negatives. XGBoost’s robustness and strong generalization ability made it an essential part of the model comparison.

### 4.6.3 Light GBM (Light Gradient Boosting Machine)

LightGBM is a highly efficient gradient boosting framework developed by Microsoft that is designed for performance and scalability. Unlike traditional boosting algorithms that grow decision trees level-wise, LightGBM uses a leaf-wise tree growth strategy, which splits the leaf with the largest loss reduction. This approach allows LightGBM to produce deeper, more complex trees that often result in better accuracy compared to level-wise methods. Additionally, it uses histogram-based learning, where continuous features are bucketed into discrete bins to speed up computation and reduce memory usage. Another major advantage of LightGBM is its ability to handle large datasets with high-dimensional and sparse features efficiently. It also supports native handling of categorical variables by finding the best splits without requiring explicit encoding. These capabilities make it particularly well-suited for fraud detection, where real-time performance and the ability to learn from massive transactional datasets are crucial. In this project, LightGBM was configured to optimize detection of fraudulent transactions by tuning its depth, learning rate, and boosting iterations. Its ability to deliver fast training with high predictive power made it a strong contender among the three models evaluated.

## 4.7 HYPER PARAMETER TUNING

To further enhance model performance and ensure optimal parameter configurations, Optuna, a state-of-the-art automatic hyperparameter optimization framework, was employed for tuning the CatBoost, XGBoost, and LightGBM models. Optuna uses a define-by-run approach and leverages Bayesian optimization to efficiently search the hyperparameter space by learning from previous trials. Unlike manual tuning or grid search, Optuna dynamically selects promising hyperparameter combinations using a tree-structured Parzen estimator (TPE), significantly reducing computation time while improving model accuracy.

## 4.8 PERFORMANCE EVALUATION

Evaluating the performance of fraud detection models requires more than just accuracy especially in highly imbalanced datasets, where the number of non-fraudulent transactions vastly outnumbers fraudulent ones. In such scenarios, metrics like precision, recall, F1-score, and the confusion matrix offer a clearer and more meaningful assessment of a model's effectiveness. These metrics provide insights into how well the model identifies fraudulent activity while minimizing false alarms. Precision measures the proportion of transactions predicted as fraudulent that were fraudulent. Recall measures the proportion of actual fraud cases that the model correctly identified. The F1-score is the harmonic mean of precision and recall. It provides a balanced measure that considers both false positives and false negatives.



Figure 4. Confusion Matrix

# CHAPTER 5

# RESULT ANALYSIS AND DISCUSSIONS

## 5.1 IMPLEMENTATION DETAILS

The implementation was carried out using Google Colab, a cloud-based Python environment that provides free GPU access, making it ideal for rapid prototyping and model training. The libraries and tools are:

**Data Manipulation and Handling**

1. pandas: For loading, cleaning, transforming, and manipulating structured tabular data and is used for DataFrame operations, datetime conversions, and feature engineering.
2. numpy: For efficient numerical operations and handling arrays.Used in mathematical computations like calculating class weights and distance metrics.

**Visualization**

1. matplotlib.pyplot: For creating static visualizations such as histograms, bar plots, and scatter plots. Used to visualize transaction distributions, outliers, and feature relationships.
2. seaborn: Built on top of matplotlib for more attractive and informative statistical plots.

**Model Development and Training**

1. CatBoostClassifier (from catboost): For training CatBoost gradient boosting models.
2. XGBClassifier (from xgboost): For implementing the XGBoost algorithm with customizable hyperparameters.
3. LGBMClassifier (from lightgbm): For training LightGBM models with histogram-based and leaf-wise growth.

## 5.2 RESULT ANALYSIS

### 5.2.1 Cat Boost

The CatBoost model was implemented using the CatBoostClassifier from the CatBoost library with several optimized hyperparameters. Key parameters included a tree depth of 10, a learning rate of 0.2, and 2000 estimators. Regularization was applied using l2\_leaf\_reg=8, and the Bernoulli bootstrap method was used for stochastic sampling.

A graph with numbers and a blue square

AI-generated content may be incorrect.

Figure 5. CatBoost – Confusion Matrix

Figure 5.1 (Confusion Matrix), the CatBoost model achieved excellent classification performance, correctly identifying 1,244 out of 1,382 actual fraud cases (True Positives) while only misclassifying 138 frauds as non-fraud (False Negatives). Simultaneously, it misclassified only 95 non-fraudulent transactions as fraud (False Positives), while accurately classifying 257,858 legitimate transactions (True Negatives). This demonstrates the model’s ability to maintain a strong balance between minimizing both false positives and false negatives. The classification report in Figure 5.2 further supports this, showing a precision of 0.93, recall of 0.90, and an F1-score of 0.91 for the fraud class (label 1). The overall accuracy is nearly perfect at 1.00, with macro and weighted averages of precision, recall, and F1-score also exceeding 0.95. These metrics underscore the model's robustness and effectiveness in detecting fraudulent transactions with both high specificity and sensitivity.

A screenshot of a computer screen

AI-generated content may be incorrect.

Figure 5. CatBoost – Classification Report

### 5.2.2 XGBoost

The XGBoost model was implemented using the XGBClassifier with carefully tuned hyperparameters to address the class imbalance and enhance generalization. The model utilized a maximum tree depth of 7, a learning rate of 0.2, and 2000 boosting rounds. Regularization parameters such as reg\_lambda=1 (L2 regularization) and reg\_alpha=3 (L1 regularization) were used to prevent overfitting. To better handle class imbalance, the scale\_pos\_weight parameter was set based on the ratio of class weights. Figure 5.3 (Confusion Matrix), the XGBoost model accurately identified 1,246 fraudulent cases and 257,802 legitimate transactions. However, it misclassified 136 fraud cases as non-fraud (False Negatives) and 151 non-fraud cases as fraudulent (False Positives). While the number of misclassifications is slightly higher than that of the CatBoost model, the results still indicate strong performance in handling imbalanced data. The classification report shown in Figure 5.4 reinforces this conclusion. The model achieved a precision of 0.89, recall of 0.90, and F1-score of 0.90 for the fraud class, indicating a well-balanced ability to minimize both false alarms and missed detections. The macro-averaged and weighted metrics both registered 0.95 and above, with an overall accuracy of 1.00, demonstrating that the model is highly capable in real-world fraud detection scenarios.

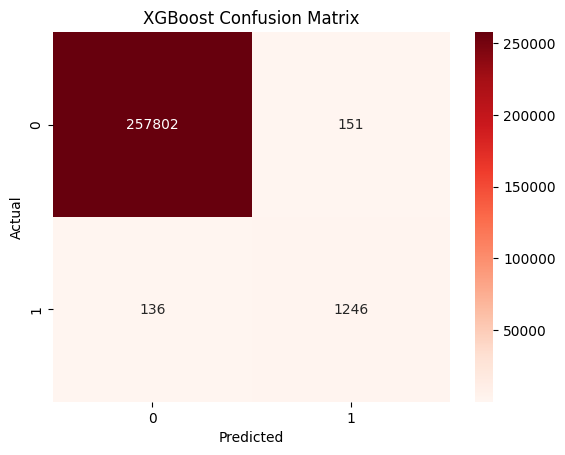


Figure 5. XGBoost– Confusion Matrix

A screenshot of a computer screen

AI-generated content may be incorrect.

Figure 5. XGBoost – Classification Report

### 5.2.3 LightGBM

The LightGBM model is a highly efficient gradient boosting framework that is well-suited for large-scale and high-dimensional data. It uses a leaf-wise tree growth strategy, which allows for faster training and better accuracy by focusing on leaf nodes with the highest gain. In this implementation, the model is configured with a maximum depth of 8, 64 leaves per tree, and a learning rate of 0.03. It is trained with 2000 estimators, using regularization parameters reg\_lambda=3 and reg\_alpha=1 to prevent overfitting. The scale\_pos\_weight is adjusted to handle the significant class imbalance between fraudulent and non-fraudulent transactions. Figure 6.6 displays the confusion matrix for the LightGBM model, which shows excellent predictive performance. It correctly classified 257,738 legitimate transactions and 1,272 fraudulent ones, while misclassifying 110 frauds as legitimate (false negatives) and 215 legitimate ones as frauds (false positives). This balance indicates that the model is highly accurate while maintaining a good trade-off between precision and recall. Figure 5.5 provides the classification report for the LightGBM model. The model achieved a precision of 0.86 and recall of 0.92 for the fraudulent class, resulting in an F1-score of 0.89. The overall accuracy is perfect due to the large class imbalance, but the macro average F1-score of 0.94 gives a more balanced view of performance. These results affirm the model’s capability to effectively detect fraudulent transactions while keeping misclassification rates low.

A screenshot of a computer screen

AI-generated content may be incorrect.

Figure 5. LightGBM – Classification Report

A green and white graph

AI-generated content may be incorrect.

Figure 5. LightGBM – Confusion Matrix

## 5.3 FEATURE IMPORTANCE

As part of this study, an important step involved identifying which features contributed most to detecting fraudulent credit card transactions. Three different machine learning models CatBoost, XGBoost, and LightGBM each of generates a feature importance score as part of its output. These scores help us understand which variables the models rely on most when making predictions.

Collected the importance values from each of the three models and calculated the average importance for every feature. The top 10 most important features were then visualized using a bar chart, as shown in Figure 5.7. The transaction amount (amt) is the most influential feature. This is expected, as unusually high or low amounts often trigger suspicion. The second most important feature is the cardholder’s age (card\_holder\_age), suggesting that certain age groups may exhibit different transaction behaviors that help in identifying fraud.

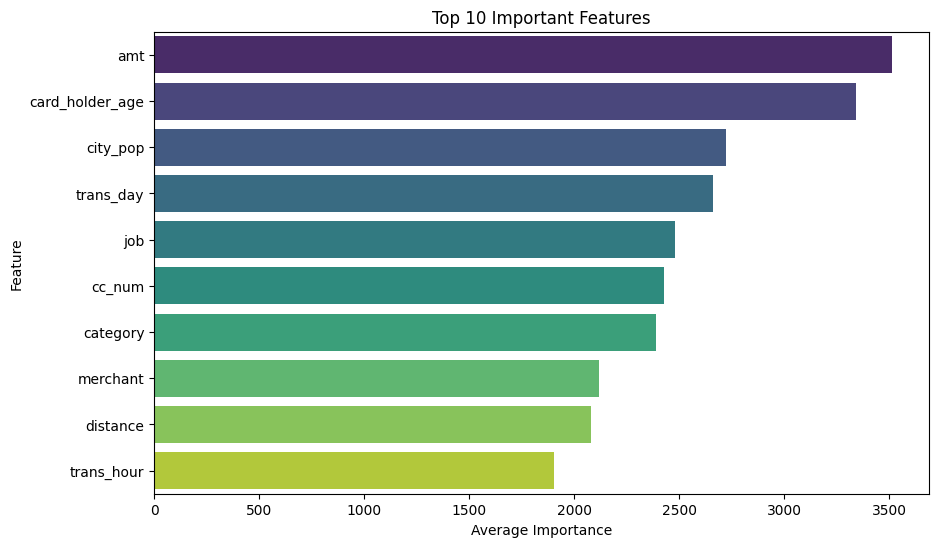


Figure 5. Top 10 Important Feature

## 5.4 CRITICAL ANALYSIS

In this study, we implemented and evaluated three advanced gradient boosting algorithms CatBoost, XGBoost, and LightGBM for credit card fraud detection. Each model was trained and tested using the same dataset, preprocessing steps, and feature set, enabling a fair and consistent comparison across multiple performance metrics.

CatBoost emerged as the most balanced and accurate model in this comparison. It handled categorical features efficiently without requiring extensive manual encoding, which helped preserve the quality of the data. The model achieved high precision (0.93), recall (0.90), and F1-score (0.91) for detecting fraudulent transactions, along with the lowest number of false negatives among the three models (138). These results suggest that CatBoost is highly reliable for identifying fraud without over flagging real transactions.

XGBoost also delivered strong performance, with an F1-score of 0.90 for the fraud class. However, it recorded slightly more false positives (151) and showed marginally lower precision (0.89) for fraud detection compared to CatBoost. LightGBM performed well in terms of recall (0.92), which means it was highly sensitive in catching most fraud cases. However, it had the lowest precision (0.86), indicating a higher number of false positives. This can be problematic in real-world applications where too many false alerts may lead to customer dissatisfaction and loss of trust.

Considering the sensitivity of fraud detection systems and the cost of false alarms, CatBoost is selected as the best-performing model in this study. It offers the strongest balance between accuracy, robustness, and interpretability, making it well-suited for production deployment in financial systems.

Table 1: Model Performance comparison

Table Model Performance comparison

|  |  |  |  |
| --- | --- | --- | --- |
| **Metric** | **CatBoost** | **XGBoost** | **LightGBM** |
| **Precision (Fraud)** | 0.93 | 0.89 | 0.86 |
| **Recall (Fraud)** | 0.90 | 0.90 | 0.92 |
| **F1-score (Fraud)** | 0.91 | 0.90 | 0.89 |
| **Accuracy** | 1.00 | 1.00 | 1.00 |
| **False Positives** | 95 | 151 | 215 |
| **False Negatives** | 138 | 136 | 110 |

## 5.5 LIMITATIONS

The proposed methodology for identifying credit card fraud with modern gradient boosting algorithms, although certain limitations must be recognized. The dataset employed, while realistic and comprehensive, remains simulated and anonymous. This constrains the applicability of the findings to real-world contexts where transaction patterns and user behaviours may differ markedly and are frequently affected by intricate, developing fraud tactics.

Secondly, despite employing strategies such as class weighting and hyperparameter optimization, the substantial class imbalance in the dataset continues to pose a considerable barrier. Fraudulent transactions represent a minimal fraction of the data; however, they can nonetheless induce model bias towards the majority class. This disparity may hinder the model's capacity to identify infrequent yet critical fraud incidents across varying operational contexts. Fraud detection systems necessitate ongoing model retraining and adaption due to the evolving nature of fraud behaviours. This static modeling technique presumes the underlying data distribution is stable, which may not be accurate in a dynamic financial ecosystem. Even though feature importance scores partially addressed explainability, more sophisticated interpretability methods, such as SHAP values or LIME, were not thoroughly examined, which is crucial for real-world applications where transparency and justification of automated decisions are imperative. These constraints underscore the necessity for more research utilizing more varied, real-time, and annotated datasets, as well as the incorporation of contextual fraud information and ongoing model assessment in operational environments.

# CHAPTER 6

# CONCLUSION

## 6.1 CONCLUSION

In this project the main aim is to create and assess a resilient credit card fraud detection system utilizing sophisticated machine learning models—CatBoost, XGBoost, and LightGBM—on a realistic simulated dataset. By employing thorough preprocessing, developing characteristics such transaction time variables and geographical distance, and meticulously addressing class imbalance, the models were trained to accurately differentiate between legal and fraudulent transactions. The assessment indicated that all three models exhibited outstanding performance, with CatBoost marginally surpassing the others in precision, recall, and F1-score for identifying fraudulent transactions. The examination of feature importance corroborated that transaction amount, cardholder age, and city population significantly influenced predictions, mirroring actual fraud tendencies. The work provides useful insights into the implementation of intelligent fraud detection systems by integrating interpretability, robust classification metrics, and scalable model frameworks. It emphasizes the necessity of ongoing surveillance, access to authentic transaction data, and flexible modeling to sustain performance in dynamic financial settings.

## 6.2 FUTURE WORKS

To proposed method has successfully demonstrated the use of advanced machine learning models for credit card fraud detection, there are several practical directions in which the work can be further developed. One of the key areas for improvement is bringing the model into a real-time setting. In actual banking and payment systems, detecting fraud instantly is critical. The current models work in a batch setting, but future versions can be adapted to process live transaction data as it flows in, enabling timely alerts and quicker responses. Another important step forward would be the inclusion of richer and more detailed features. The current dataset, although realistic, lacks some of the behavioural and contextual information that could make fraud detection even more accurate. Adding data like IP addresses, device types, login history, or previous user transaction patterns could provide deeper insights and help detect more subtle types of fraud. Making the model explainable is another valuable direction. In industries like banking, it’s not enough to say that a transaction is suspicious you also need to explain why. Integrating tools like SHAP or LIME would make it easier for analysts and auditors to understand the model’s reasoning and take informed actions. Another real-world challenge is that fraud patterns constantly change. To keep the system effective, it’s important to monitor and update the model regularly. This might involve retraining with newer data or adding components that can detect when the model’s performance begins to drop. Approaches like federated learning allow institutions to collaborate on model improvement without sharing raw customer data.

# REFERENCES

Alarfaj, F. K., Malik, I., Khan, H. U., Almusallam, N., Ramzan, M., & Ahmed, M. (2022). Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms. *IEEE Access*, *10*, 39700–39715. https://doi.org/10.1109/ACCESS.2022.3166891

Awosika, T., Shukla, R. M., & Pranggono, B. (2023). *Transpare ncy and Privacy: The Role of Explainable AI and Federated Learning in Financial Fraud Detection* (arXiv:2312.13334). arXiv. https://doi.org/10.48550/arXiv.2312.13334

Bockel-Rickermann, C., Verdonck, T., & Verbeke, W. (2023). Fraud analytics: A decade of research: Organizing challenges and solutions in the field. *Expert Systems with Applications*, *232*, 120605. https://doi.org/10.1016/j.eswa.2023.120605

Boyd, A., Bamler, R., Mandt, S., & Smyth, P. (2020). *User-Dependent Neural Sequence Models for Continuous-Time Event Data* (arXiv:2011.03231). arXiv. https://doi.org/10.48550/arXiv.2011.03231

Canim, M., Kundu, A., & Payne, J. (2019). *Uncheatable Machine Learning Inference* (arXiv:1908.03270). arXiv. https://doi.org/10.48550/arXiv.1908.03270

Cheng, Y., Wang, C.-H., Potluru, V. K., Balch, T., & Cheng, G. (2024). *Downstream Task-Oriented Generative Model Selections on Synthetic Data Training for Fraud Detection Models* (arXiv:2401.00974). arXiv. https://doi.org/10.48550/arXiv.2401.00974

*Credit Card Transactions Dataset*. (n.d.). Retrieved May 4, 2025, from https://www.kaggle.com/datasets/priyamchoksi/credit-card-transactions-dataset

Fajardo, V. A., Findlay, D., Houmanfar, R., Jaiswal, C., Liang, J., & Xie, H. (2018). *VOS: A Method for Variational Oversampling of Imbalanced Data* (arXiv:1809.02596). arXiv. https://doi.org/10.48550/arXiv.1809.02596

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., & Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, *77*, 354–377. https://doi.org/10.1016/j.patcog.2017.10.013

Habibpour, M., Gharoun, H., Mehdipour, M., Tajally, A., Asgharnezhad, H., Shamsi, A., Khosravi, A., Shafie-Khah, M., Nahavandi, S., & Catalao, J. P. S. (2021). *Uncertainty-Aware Credit Card Fraud Detection Using Deep Learning* (arXiv:2107.13508). arXiv. https://doi.org/10.48550/arXiv.2107.13508

He, H., & Garcia, E. A. (2009). Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239

Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances. *Expert Systems with Applications*, *193*, 116429. https://doi.org/10.1016/j.eswa.2021.116429

Jentzen, A., Kuckuck, B., Neufeld, A., & Wurstemberger, P. von. (2021). Strong error analysis for stochastic gradient descent optimization algorithms. *IMA Journal of Numerical Analysis*, *41*(1), 455–492. https://doi.org/10.1093/imanum/drz055

Kulatilleke, G. K. (2022). *Challenges and Complexities in Machine Learning based Credit Card Fraud Detection* (arXiv:2208.10943). arXiv. https://doi.org/10.48550/arXiv.2208.10943

Kurshan, E., & Shen, H. (2023). *Temporal Knowledge Distillation for Time Sensitive Financial Services Applications* (SSRN Scholarly Paper 4941375). Social Science Research Network. https://doi.org/10.2139/ssrn.4941375

Kyriienko, O., & Magnusson, E. B. (2022). *Unsupervised quantum machine learning for fraud detection* (arXiv:2208.01203). arXiv. https://doi.org/10.48550/arXiv.2208.01203

Lakis, V., & Giriūnas, L. (2012). The concept of internal control system: Theoretical aspect /. *Ekonomika*, *91*(2), 142–152. https://doi.org/10.15388/Ekon.2012.0.890

Li, J., Liu, Y., Jia, Y., Ren, Y., & Nanduri, J. (2018). *Predictive Modeling with Delayed Information: A Case Study in E-commerce Transaction Fraud Control* (arXiv:1811.06109). arXiv. https://doi.org/10.48550/arXiv.1811.06109

Liu, J., Cui, J., & Chen, C. (2023). Online Efficient Secure Logistic Regression based on Function Secret Sharing. *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 1597–1606. https://doi.org/10.1145/3583780.3614998

Liu, N., & Rebentrost, P. (2018). Quantum machine learning for quantum anomaly detection. *Physical Review A*, *97*(4), 042315. https://doi.org/10.1103/PhysRevA.97.042315

Lunghi, D., Simitsis, A., Caelen, O., & Bontempi, G. (2023). Adversarial Learning in Real-World Fraud Detection: Challenges and Perspectives. *Proceedings of the Second ACM Data Economy Workshop*, 27–33. https://doi.org/10.1145/3600046.3600051.

Psychoula, I., Gutmann, A., Mainali, P., Lee, S. H., Dunphy, P., & Petitcolas, F. A. P. (2021). *Explainable Machine Learning for Fraud Detection* (arXiv:2105.06314). arXiv. https://doi.org/10.48550/arXiv.2105.06314

Raymaekers, J., Verbeke, W., & Verdonck, T. (2021). *Weight-of-evidence 2.0 with shrinkage and spline-binning* (arXiv:2101.01494). arXiv. https://doi.org/10.48550/arXiv.2101.01494

Sailusha, R., Gnaneswar, V., Ramesh, R., & Rao, G. R. (2020). Credit Card Fraud Detection Using Machine Learning. *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 1264–1270. https://doi.org/10.1109/ICICCS48265.2020.9121114

Shapira, G., & Schuster, A. (2022). *Unsupervised Frequent Pattern Mining for CEP* (arXiv:2207.14017). arXiv. https://doi.org/10.48550/arXiv.2207.14017

Vaidya, A., & Awasthi, A. (2025). *Zero-to-One IDV: A Conceptual Model for AI-Powered Identity Verification* (arXiv:2503.08734). arXiv. https://doi.org/10.48550/arXiv.2503.08734

West, J., Bhattacharya, M., & Islam, R. (2015). *Intelligent Financial Fraud Detection Practices: An Investigation* (arXiv:1510.07165). arXiv. https://doi.org/10.48550/arXiv.1510.07165

Zhang, W., Ramezani, R., & Naeim, A. (2019). WOTBoost: Weighted Oversampling Technique in Boosting for imbalanced learning. *2019 IEEE International Conference on Big Data (Big Data)*, 2523–2531. https://doi.org/10.1109/BigData47090.2019.9006091