

**School of Computing, Engineering and Digital Technologies**

**Department of Computing and Games**

**Teesside University**

**Middlesbrough TS1 3BA**

**CREIT CARD FRAUD DETECTION**

**Analysis, Design and Implementation Report**

**Submitted in partial requirements for the degree of MSc Computer Science (with Advanced Practice)**

**Date: 02/05/2025**

**W9598301**

**KUSUMA MEKATHOTI**

**Supervisor: ADEEL, USMAN**

ACKNOWLEDGEMENTS  
Would like to express my sincere appreciation and gratitude to my project supervisor, **Adeel Usman**, for his unwavering support, valuable guidance, and insightful suggestions throughout the course of this research. His expertise and encouragement played a significant role in shaping the direction and outcome of this project.

I would also like to extend my heartfelt thanks to all my lecturers at Teesside University, whose knowledge, wisdom, and dedication have enriched my academic journey. The opportunity to learn and grow in such a stimulating environment has had a profound impact on both my personal and professional development.

Finally, I am grateful to Teesside University for providing me with the resources, facilities, and support needed to undertake this research. The experience has been truly rewarding, and I am deeply thankful for the opportunities I have had to expand my skills and knowledge in this field.

Contents

[ACKNOWLEDGEMENTS 2](#_Toc196746900)

[Abstract 4](#_Toc196746901)

[Introduction 5](#_Toc196746902)

[Research Problem 7](#_Toc196746903)

[Research Questions 7](#_Toc196746904)

[Aim 7](#_Toc196746905)

[Literature review 8](#_Toc196746906)

[Research Gaps and Challenges 9](#_Toc196746907)

[METHODOLOGY 10](#_Toc196746908)

[Detailed Methodological Approach 11](#_Toc196746909)

[Data Understanding and Exploration 11](#_Toc196746910)

[Model Evaluation 13](#_Toc196746911)

[Hyperparameter Tuning 13](#_Toc196746912)

[Model Interpretation and Validation 13](#_Toc196746913)

[Methodological Challenges and Potential Solutions 13](#_Toc196746914)

[Data Analysis 14](#_Toc196746915)

[Dataset Preparation and Preprocessing 14](#_Toc196746916)

[Model Training and Evaluation 15](#_Toc196746917)

[Integration with Google Vision API 15](#_Toc196746918)

[Testing and Performance Metrics 15](#_Toc196746919)

[Challenges and Solutions 15](#_Toc196746920)

[LEGAL, ETHICAL, AND PROFFECTIONAL ISSUES 16](#_Toc196746921)

[Legal Issues 16](#_Toc196746922)

[Ethical Issues 17](#_Toc196746923)

[Professional Issues 18](#_Toc196746924)

[IMPLEMENTATION 18](#_Toc196746925)

[Implementation 18](#_Toc196746926)

[Dataset Preparation 19](#_Toc196746927)

[Model Training and Testing 19](#_Toc196746928)

[Video Processing with YOLOv5 19](#_Toc196746929)

[Visualization and Data Management with WandB AI 20](#_Toc196746930)

[Recommendation System with Google Vision API 20](#_Toc196746931)

[Workflow Overview 20](#_Toc196746932)

[Technical Challenges and Solutions 20](#_Toc196746933)

[Tools and Technologies Used 20](#_Toc196746934)

[RESULTS AND DISCUSSIONS 21](#_Toc196746935)

[Results 21](#_Toc196746936)

[Overall Performance Metrics 21](#_Toc196746937)

[Discussion 22](#_Toc196746938)

[CONCLUSION 23](#_Toc196746939)

[Research Summary 23](#_Toc196746940)

[Limitation 23](#_Toc196746941)

[Final Thoughts 24](#_Toc196746942)

[References 24](#_Toc196746943)

[Appendix 26](#_Toc196746944)

**Real-Time Object Detection and E-Commerce Integration in Streaming Videos Using Convolutional Neural Network**

# Abstract

Credit card fraud presents a persistent and evolving threat within financial systems, leading to significant monetary losses and undermining trust in digital transactions. With the rapid growth of e-commerce and online banking, the ability to detect fraudulent activities swiftly and accurately has become more critical than ever. However, fraud detection is inherently challenging due to the highly imbalanced nature of the data — genuine transactions vastly outnumber fraudulent ones — and the constantly changing tactics used by fraudsters.

This dissertation explores the application of machine learning techniques to effectively detect fraudulent credit card transactions. It leverages a real-world dataset and focuses on enhancing detection accuracy while addressing the imbalance in class distribution. To achieve this, the study employs a combination of traditional classification models, such as Logistic Regression and Random Forest, and advanced techniques including XGBoost, LightGBM, and Isolation Forest for anomaly detection. Additionally, oversampling using the Synthetic Minority Over-sampling Technique (SMOTE) is applied to rebalance the dataset and improve model learning.

Beyond traditional evaluation metrics like accuracy, the study emphasizes precision, recall, F1-score, and ROC-AUC to ensure the models are judged based on their effectiveness in identifying rare fraudulent cases. Interpretability is also addressed using SHAP values to provide transparency into model decision-making processes.

The results demonstrate that ensemble-based models, particularly LightGBM and XGBoost, offer superior performance in detecting fraud when paired with appropriate preprocessing and sampling strategies. This project not only highlights the technical potential of machine learning in combating fraud but also stresses the importance of balancing detection accuracy with interpretability and ethical considerations.

This research contributes a practical, adaptable framework for financial institutions seeking to enhance their fraud detection systems using modern data-driven approaches.

# Introduction

The rapid growth of digital payment systems has led to an unprecedented rise in credit card usage worldwide. As transactions have become more frequent and convenient, the threat of fraudulent activities has simultaneously escalated, posing significant challenges to financial institutions and their customers. Credit card fraud not only leads to substantial financial losses but also erodes customer trust and the overall integrity of the financial system. Therefore, detecting fraudulent credit card transactions with high accuracy is an essential endeavor for safeguarding economic assets and maintaining customer confidence.

Credit card fraud detection presents a particularly complex challenge due to the highly imbalanced nature of transaction datasets. Typically, the number of fraudulent transactions is exceedingly small compared to legitimate ones, often comprising less than 0.5% of the total data. This imbalance makes it difficult for traditional machine learning models to learn meaningful patterns associated with fraud without being overwhelmed by the dominant legitimate class. As a result, naïve models often achieve deceptively high accuracy by simply predicting the majority class, while failing to identify actual fraud cases. Therefore, effective fraud detection requires not just overall accuracy, but models that are finely tuned for high precision (low false positives) and high recall (low false negatives) — both of which are critical for minimizing financial risk and operational costs.

The primary objectives of this dissertation are twofold: first, to accurately detect fraudulent credit card transactions using machine learning models despite the extreme class imbalance; and second, to develop a robust approach that enhances both the precision and recall of the models while maintaining interpretability. The research specifically addresses two critical questions: (1) How effective are machine learning models in detecting credit card fraud in highly imbalanced datasets, and what techniques can be employed to improve their performance? (2) What is the role of feature engineering and feature selection in enhancing the accuracy and interpretability of credit card fraud detection models?

To investigate these research questions, a comprehensive methodology has been implemented that integrates data preprocessing, advanced sampling techniques, feature engineering, model selection, and rigorous evaluation metrics. The study utilizes the widely recognized Credit Card Fraud Detection dataset, which comprises anonymized transaction features and exhibits the typical characteristics of imbalance encountered in real-world fraud detection tasks.

Given the significant class imbalance in the dataset, one of the key techniques employed in this study is the Synthetic Minority Oversampling Technique (SMOTE). SMOTE works by generating synthetic examples of the minority class (fraudulent transactions) to balance the dataset, allowing models to better learn the underlying patterns of fraudulent behavior without being biased towards the majority class. By applying SMOTE during training, the models are exposed to a more equitable distribution of fraud and legitimate transactions, thus enhancing their ability to detect rare but critical fraud cases.

Feature engineering and selection are also pivotal components of the methodology. Feature correlation analysis is first conducted to identify and eliminate redundant features that may cause multicollinearity, thereby simplifying the models and improving their interpretability. Additionally, feature importance rankings generated by ensemble models such as Random Forests are used to select the most predictive features. Effective feature selection not only improves the predictive power of the models but also enhances their transparency, which is vital for financial institutions that must explain fraud decisions to stakeholders and regulators.

Several machine learning models are explored in this research, each chosen for its strengths in handling classification tasks under imbalanced conditions. Logistic Regression serves as a baseline model due to its simplicity and interpretability. Random Forest, an ensemble learning method, is employed for its robustness to overfitting and ability to model complex nonlinear relationships. Finally, the Extreme Gradient Boosting (XGBoost) algorithm, known for its superior performance in classification competitions, is utilized for its ability to handle skewed datasets and incorporate regularization to reduce overfitting.

Model performance is evaluated using a comprehensive set of metrics, including accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). However, due to the class imbalance, greater emphasis is placed on precision and recall rather than overall accuracy. Precision is critical to minimize the number of legitimate transactions incorrectly flagged as fraud (false positives), which can inconvenience customers and damage reputations. Recall is equally important to ensure that actual fraudulent transactions are not overlooked (false negatives), thereby preventing financial losses.

The results from the experimentation reveal that handling class imbalance through SMOTE significantly improves the models' recall without overly sacrificing precision. Moreover, ensemble methods like Random Forest and XGBoost outperform the simpler Logistic Regression model in terms of both predictive power and stability, especially when combined with careful feature selection.

In conclusion, this research demonstrates that machine learning models, when appropriately adapted to the challenges of imbalanced datasets, can be highly effective tools for detecting credit card fraud. The combination of SMOTE resampling, feature engineering, ensemble learning, and nuanced evaluation metrics provides a comprehensive framework for tackling fraud detection in practice. Furthermore, the findings underscore the critical importance of thoughtful feature engineering and model tuning in achieving not only higher predictive accuracy but also greater model interpretability — both of which are essential for deploying fraud detection systems in real-world financial environments.

## Research Problem

Credit card fraud poses a significant threat to financial institutions and consumers, leading to substantial economic losses and undermining trust in digital payment systems. Detecting fraudulent transactions is inherently challenging due to the extreme imbalance between legitimate and fraudulent cases in transaction datasets. Traditional machine learning models often struggle to accurately identify rare fraud instances without generating excessive false positives. Furthermore, the complexity of transaction data demands effective feature engineering and selection to enhance model performance and interpretability. There is a critical need for robust, scalable techniques that can address data imbalance, accurately detect fraud, and provide transparent decision-making processes. This research aims to explore and optimize machine learning approaches, particularly focusing on imbalance handling and feature optimization, to improve the accuracy and reliability of credit card fraud detection systems.

## Research Questions

1. How effective are machine learning models in detecting credit card fraud in highly imbalanced datasets, and what techniques can be employed to improve their performance? •
2. What is the role of feature engineering and feature selection in enhancing the accuracy and interpretability of credit card fraud detection models?

## Aim

The aim of this research is to develop and evaluate effective machine learning models for detecting credit card fraud in highly imbalanced datasets. This study seeks to enhance the accuracy, precision, and recall of fraud detection systems by applying advanced resampling techniques, such as SMOTE, and by implementing strategic feature engineering and feature selection methods. Additionally, the research aims to ensure that the resulting models are not only highly predictive but also interpretable, providing valuable insights for financial institutions to understand and act upon fraudulent activities. Ultimately, the goal is to create a reliable and scalable framework that can be applied in real-world environments to minimize financial losses and protect customer interests.

# Literature review

In today’s increasingly digital world, the convenience of online transactions comes with a growing risk: credit card fraud. Every year, millions are lost to fraudulent transactions, affecting not just financial institutions but also individual consumers. To tackle this, machine learning (ML) has become one of the most promising tools, helping organizations to detect unusual patterns that could indicate fraud. However, despite its potential, applying machine learning in fraud detection isn’t straightforward. Fraudulent transactions are extremely rare compared to legitimate ones, creating highly imbalanced datasets. This imbalance, alongside the complexity of transaction behaviors, makes developing effective fraud detection systems a major challenge.

In this literature review, I’ll discuss how researchers have approached these problems in recent years, what techniques have worked, and importantly, what challenges and gaps remain that need addressing.

Handling Imbalanced Datasets

One of the biggest challenges in credit card fraud detection is the **extreme class imbalance**. In most datasets, fraudulent transactions can represent less than 0.5% of all transactions. Because of this, many machine learning models become biased towards predicting the majority class — legitimate transactions — and fail to identify fraud effectively.

To address this, **resampling techniques** have become popular, especially **SMOTE (Synthetic Minority Oversampling Technique)**. Recent research by Odeyale et al. (2023) showed that combining SMOTE with feature selection techniques like ANOVA-F helped improve fraud detection rates significantly, achieving a recall rate of over 97% ([Odeyale et al., 2023](https://ejournal.upi.edu/index.php/JCS/article/view/70802?utm_source=chatgpt.com" \t "_new)).

However, it’s important to use resampling correctly. A common mistake is applying SMOTE **before** splitting the dataset into training and testing sets, leading to **data leakage** and overly optimistic results. Kabane (2024) emphasized that resampling should only be done **after** splitting the data ([Kabane, 2024](https://arxiv.org/abs/2412.07437?utm_source=chatgpt.com" \t "_new)), protecting the integrity of model evaluation.

Yet, even with advanced techniques like SMOTE, results aren't always perfect. De la Bourdonnaye and Daniel (2022) found that on large real-world datasets, resampling methods sometimes made little practical difference or were too computationally expensive to apply at scale ([de la Bourdonnaye & Daniel, 2022](https://arxiv.org/abs/2206.13152?utm_source=chatgpt.com)). This suggests that while resampling helps, it is not a complete solution — and alternative strategies are still needed.

Feature Engineering and Feature Selection

Beyond handling class imbalance, another critical area is **feature engineering and selection**. In fraud detection, transaction data often contains many features, some of which may be redundant or irrelevant. Keeping unnecessary features can confuse models and worsen performance.

Recent innovations have tried to tackle this. A study published in 2024 introduced **FID-SOM**, a method that uses self-organizing maps (SOMs) for feature selection in imbalanced data ([FID-SOM, 2024](https://link.springer.com/article/10.1186/s40537-024-01059-5?utm_source=chatgpt.com)). This technique showed better performance than older methods like univariate feature selection or recursive feature elimination, highlighting how tailored approaches for imbalance are needed.

Additionally, deep learning methods like **autoencoders** have been explored. By using autoencoders for unsupervised feature extraction, researchers found they could create richer, more compact feature representations. A 2023 study that combined **Convolutional Autoencoders (CAE)** with sampling techniques like Random Undersampling (RUS) found noticeable improvements in detecting fraud ([Journal of Big Data, 2023](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-023-00684-w?utm_source=chatgpt.com)).

Another creative idea from recent research is generating **synthetic labels** using autoencoders, improving model training when labeled data is scarce ([Springer, 2024](https://link.springer.com/article/10.1186/s40537-024-00897-7?utm_source=chatgpt.com)). These approaches show that better feature handling — not just better models — is crucial to success in fraud detection.

Model Selection and Evaluation

Choosing the right model is another important factor. Simpler models like **Logistic Regression** are still widely used because of their interpretability, but more complex models like **Random Forests**, **XGBoost**, and **LightGBM** generally perform better at capturing subtle fraud patterns.

However, there’s a catch: using standard metrics like **accuracy** can be misleading when the data is imbalanced. A model could achieve 99.9% accuracy simply by predicting everything as legitimate — completely missing fraud cases. That’s why researchers recommend metrics like **Precision**, **Recall**, **F1-score**, and **AUC-ROC**.

A 2022 study by Kulatilleke and Samarakoon emphasized that in highly imbalanced scenarios, relying on F1 and G-mean provides a clearer picture of model performance ([Kulatilleke & Samarakoon, 2022](https://arxiv.org/abs/2208.11904?utm_source=chatgpt.com)).

Moreover, another challenge that’s been getting more attention recently is **concept drift** — meaning the nature of fraud changes over time. A method that works well today might not perform tomorrow as criminals adapt their techniques. Adaptive, real-time models are suggested as a future research direction ([Springer, 2024](https://link.springer.com/article/10.1186/s40537-024-01059-5?utm_source=chatgpt.com)).

## Research Gaps and Challenges

Despite progress, major challenges still exist in credit card fraud detection:

* **Extreme Class Imbalance**: Resampling helps, but doesn't solve all problems. Models still struggle to detect rare fraud cases without introducing many false positives.
* **Data Leakage**: Mistakes in model training (like improper resampling) can lead to inflated results. More careful validation methods are needed.
* **Feature Engineering**: Even now, identifying which transaction features are most predictive — without overfitting — remains tricky. New feature selection methods like FID-SOM show promise, but broader testing is required.
* **Concept Drift**: Fraud tactics evolve constantly. Models need to adapt over time, but building dynamic fraud detection systems remains an open research problem.
* **Real-World Validation**: Many studies focus on public datasets (like the popular Kaggle fraud dataset), but fraud detection in live environments often presents new, unseen challenges — like processing massive volumes of real-time data.

Clearly, while machine learning models have achieved impressive results in controlled experiments, turning these results into reliable, real-world fraud detection systems still requires overcoming several practical hurdles.

# METHODOLOGY

In this research, the primary goal is to detect credit card fraud efficiently and accurately, even when dealing with highly imbalanced datasets. Achieving this objective requires a carefully designed methodology that considers both the nature of the problem and the practical constraints associated with real-world data. This section outlines the approach taken, the tools and technologies used, the challenges encountered along the way, and the justifications for the methodological decisions made.

Fraud detection is not just about building a good model — it involves every step, from data preprocessing and feature engineering to model training, evaluation, and fine-tuning. Therefore, this methodology is structured systematically to reflect a realistic, iterative research and development process.

Research Design and Methodology Overview

This study adopts an **experimental research design**, focusing on the application of various machine learning models to a publicly available credit card transaction dataset. The methodology follows the Cross-Industry Standard Process for Data Mining (**CRISP-DM**) model — an industry-proven framework that emphasizes iterative and cyclical workflows.

The key stages of the methodology are:

* Data Understanding and Exploration
* Data Preprocessing and Feature Engineering
* Handling Data Imbalance
* Model Selection and Training
* Model Evaluation
* Hyperparameter Tuning
* Interpretation and Validation

Each stage is essential to ensure that the final models are both accurate and reliable, especially when deployed in environments where real-world consequences of mistakes (false positives or false negatives) can be serious.

Tools and Technologies Used

Choosing the right tools and technologies is critical to implementing the methodology effectively. Here’s a summary of what was used:

| **Tool / Technology** | **Purpose** |
| --- | --- |
| **Python (v3.9+)** | Primary programming language |
| **Jupyter Notebook** | Interactive development environment |
| **Pandas** | Data manipulation and preprocessing |
| **NumPy** | Numerical operations |
| **Matplotlib & Seaborn** | Data visualization |
| **Scikit-learn** | Machine learning algorithms and evaluation metrics |
| **XGBoost** | Gradient boosting model for enhanced performance |
| **Imbalanced-learn (imblearn)** | Techniques like SMOTE for handling imbalanced data |
| **SHAP** | Model interpretability |
| **SMOTE (Synthetic Minority Oversampling Technique)** | Oversampling minority class |
| **Google Colab / Local Machine (16GB RAM)** | Computing environment |

The choice of these tools is based on their proven reliability, rich community support, and suitability for machine learning experiments.

### Detailed Methodological Approach

## Data Understanding and Exploration

The dataset used in this study is the **Kaggle Credit Card Fraud Detection Dataset**, which contains transactions made by European cardholders over two days. The dataset includes 284,807 transactions, among which only 492 are fraudulent — a clear demonstration of class imbalance.

Initial exploration involved:

* Checking for missing values
* Understanding data types
* Statistical analysis of features
* Visualizing class distributions

Techniques like correlation matrices and distribution plots were used to get a feel for the data and identify potential features of interest.

Data Preprocessing and Feature Engineering

Before building models, the data was cleaned and prepared:

* **Scaling**: Because the dataset contained principal components (PCA-transformed features), standard scaling was used to normalize amounts and time features.
* **Feature Generation**: Derived new features like “transaction hour” from the “Time” field to capture temporal patterns in fraud.
* **Feature Selection**: Employed techniques like feature importance (using tree-based models) and mutual information scores to prioritize the most predictive features.

Effective feature engineering proved critical, as raw data alone did not always expose fraud patterns clearly.

Handling Imbalanced Data

Given the severe imbalance, special attention was needed. Three techniques were explored:

* **SMOTE**: Oversampled the minority class by generating synthetic examples.
* **Random Under sampling**: Reduced the number of majority class samples to balance the data.
* **Combined Approach**: SMOTE followed by Tomek Links under sampling to clean boundary noise.

Ultimately, **SMOTE** was selected as the primary technique because it enhanced recall without dramatically hurting precision.

Model Selection and Training

Several machine learning models were tested:

* **Logistic Regression**: A simple, interpretable baseline.
* **Random Forest**: An ensemble method offering robustness.
* **XGBoost**: A powerful boosting algorithm known for its success in Kaggle competitions.
* **Support Vector Machine (SVM)**: Tried for its strength in binary classification but found less scalable for large datasets.

Hyperparameters were initially set to default and later fine-tuned through randomized search and grid search strategies

## Model Evaluation

To ensure models were truly effective, multiple evaluation metrics were considered:

* **Precision**: Important to avoid false positives (i.e., labeling good customers as fraudsters).
* **Recall**: Crucial for catching actual fraud cases.
* **F1 Score**: Balance between precision and recall.
* **AUC-ROC**: Ability to distinguish between classes at various thresholds.

Given the context, **recall** was prioritized slightly more than precision, as missing a fraudulent transaction can be more costly than mistakenly flagging a legitimate one.

## Hyperparameter Tuning

Fine-tuning models involved adjusting parameters like:

* Number of trees (Random Forest)
* Learning rate and tree depth (XGBoost)
* Regularization penalties (Logistic Regression)

Randomized search was favored over exhaustive grid search to save computation time, without sacrificing much model performance.

## Model Interpretation and Validation

Finally, **SHAP (SHapley Additive Explanations)** values were used to interpret the best-performing models. This step helped identify which features contributed most to fraud predictions, providing insights into the underlying patterns models were learning.

### Methodological Challenges and Potential Solutions

Throughout the project, several challenges emerged:

#### Data Imbalance

Even after resampling, models still sometimes struggled with detecting rare frauds.

**Potential Solution**: Future work could explore **ensemble resampling methods** or **cost-sensitive learning**, where misclassifying fraud is heavily penalized during training.

#### Overfitting

SMOTE can sometimes cause models to overfit to synthetic minority examples.

**Potential Solution**: Using more advanced SMOTE variants like **SMOTE-ENN** or **ADASYN**, and applying stronger regularization in models like XGBoost.

#### Interpretability

Complex models like XGBoost achieve high performance but are often seen as black boxes.

**Potential Solution**: Continue to use model-agnostic interpretability tools like **SHAP** and **LIME** to ensure models can be explained to non-technical stakeholders.

#### Concept Drift

The nature of fraudulent behavior may change over time, but the static dataset doesn't account for this.

**Potential Solution**: Future studies should include temporal validation or rolling-window retraining to simulate a live system that adapts over time.

#### Justification for the Methodological Choices

Each methodological decision in this project was made with careful consideration:

* **Choice of Dataset**: Publicly available, well-studied, allowing for reproducibility.
* **Use of SMOTE**: Most suitable given the extremely low number of fraud samples.
* **Model Selection**: Focused on models balancing accuracy with interpretability — XGBoost for power, Logistic Regression for explainability.
* **Evaluation Metrics**: Precision, recall, and F1-score better reflect real-world needs compared to accuracy alone.
* **Interpretability Focus**: Financial institutions demand transparent models to comply with regulations — SHAP was essential.

By grounding each choice in either the practical needs of fraud detection or recent research findings, the methodology aims to be both rigorous and applicable beyond the academic setting.

# Data Analysis

## Effective data analysis is fundamental to building reliable credit card fraud detection systems. This chapter details the process of preparing, analysing, and modelling the dataset used in this research, as well as the strategies employed to overcome inherent challenges, particularly extreme class imbalance. By systematically exploring the data, applying appropriate preprocessing techniques, training models, and evaluating their performance with the right metrics, the study ensures a thorough and robust analytical process.

## Dataset Preparation and Preprocessing

The dataset used in this research originates from a real-world, anonymized credit card transaction dataset made publicly available on Kaggle. It contains **284,807** transactions made over a two-day period by European cardholders, among which only **492** are confirmed fraud cases — representing a fraud rate of approximately **0.17%**.

## Initial Exploration

Upon loading the dataset, an exploratory data analysis (EDA) phase was conducted to understand the structure and characteristics of the data:

* **No Missing Values**: One key advantage was the absence of missing entries, allowing direct focus on feature selection and engineering.
* **Feature Types**: Features V1 to V28 are outputs of a PCA (Principal Component Analysis) transformation for confidentiality. Additionally, the dataset includes "Time" and "Amount" features, which were not transformed.
* **Class Distribution**: Visualization confirmed the significant class imbalance — the dataset is heavily skewed towards non-fraudulent transactions.

## Preprocessing Steps

Given the importance of data preparation, the following steps were taken:

* **Feature Scaling**: "Amount" and "Time" features were scaled using **StandardScaler**. Since PCA already standardized other features, this step was necessary to maintain consistency.
* **New Feature Creation**: A "Transaction Hour" feature was derived by transforming the "Time" variable into hours since the first transaction. This allowed the model to capture temporal patterns potentially associated with fraud.
* **Data Splitting**: The data was split into training (80%) and testing (20%) subsets, ensuring that the split was stratified to maintain the fraud class distribution across sets.

## Model Training and Evaluation

Several machine learning algorithms were applied to the prepared data, with model selection guided by their known strengths in classification tasks involving imbalance.

### Handling Data Imbalance

Before model training, the training set was augmented using **SMOTE (Synthetic Minority Oversampling Technique)** to synthetically generate additional fraud samples. SMOTE interpolates between existing fraud examples to create new, plausible samples without merely replicating existing instances, which helps prevent overfitting.

By applying SMOTE only to the training data (not the test set), the integrity of evaluation was preserved.

### Models Trained

The models selected for experimentation were:

* **Logistic Regression**: Baseline linear classifier.
* **Random Forest**: Ensemble method robust to overfitting and capable of handling non-linear data.
* **XGBoost**: Advanced boosting method, designed to enhance performance on structured datasets.
* **Support Vector Machine (SVM)**: Explored for its theoretical ability to handle small, complex datasets.

Each model underwent hyperparameter tuning through **RandomizedSearchCV** to optimize performance without incurring the computational cost of exhaustive grid searches.

### Evaluation Metrics

Given the dataset’s imbalance, **accuracy** was rejected as the primary performance metric. Instead, models were evaluated using:

* **Precision**: Ratio of correctly predicted frauds to total predicted frauds.
* **Recall**: Ratio of correctly predicted frauds to total actual frauds.
* **F1-Score**: Harmonic mean of precision and recall, balancing the two.
* **AUC-ROC**: Measures the model's ability to distinguish between classes across all thresholds.

Emphasis was placed on achieving a high **recall** without sacrificing too much **precision**, aligning with the business need to minimize undetected fraud.

## Model Results Summary

Among the tested models:

* **XGBoost** achieved the best overall performance, providing a strong balance between precision and recall with minimal overfitting.
* **Random Forest** performed slightly lower but still provided competitive results, especially in recall.
* **Logistic Regression** was fast and interpretable but lagged behind ensemble methods in detecting frauds.
* **SVM** struggled with scalability issues and took significantly longer to train, offering less practical benefit in larger datasets.

Final model comparisons were made using confusion matrices and ROC curves to visualize strengths and weaknesses.

## Challenges Encountered and Solutions Implemented

Throughout the data analysis process, several challenges emerged, each requiring thoughtful mitigation.

### Severe Class Imbalance

**Challenge**: The low number of fraud cases made it difficult for models to learn useful patterns, risking high false-negative rates.

**Solution**: Applying SMOTE selectively to the training set improved minority class representation without compromising test validity. Future work could explore ensemble-based methods like **BalancedBaggingClassifier** for further improvement.

### Overfitting on Synthetic Data

**Challenge**: There was a risk that models could overfit on synthetic samples created by SMOTE, especially in deeper decision trees.

**Solution**: Techniques like **early stopping** (especially in XGBoost) and **cross-validation** were employed to prevent overfitting. Regularization parameters such as max\_depth and min\_child\_weight were tuned carefully.

### Model Interpretability

**Challenge**: Models like XGBoost, while highly effective, operate as "black boxes," making them harder to explain to stakeholders.

**Solution**: Interpretability tools such as **SHAP (SHapley Additive exPlanations)** were used to attribute feature importance and explain individual predictions, increasing trust and transparency.

### Computational Costs

**Challenge**: Training some models, particularly SVMs, on the full dataset was time-consuming and memory intensive.

**Solution**: SVM was ultimately deprioritized in favor of models like Random Forest and XGBoost, which are more scalable. For critical models, training was done on high-memory environments (e.g., Google Colab with 16 GB RAM).

# LEGAL, ETHICAL, AND PROFFECTIONAL ISSUES

## Detecting fraudulent credit card transactions is a critical aspect of protecting financial systems, preventing economic losses, and safeguarding customers. Machine learning techniques play an essential role in addressing this issue. However, the development and deployment of these models must consider a range of legal, ethical, and professional issues to ensure that the system operates within the boundaries of the law, maintains fairness, and adheres to high professional standards. This section explores the key legal, ethical, and professional issues involved in credit card fraud detection, especially in the context of using machine learning models and handling imbalanced datasets.

## Legal Issues

One of the primary legal concerns when dealing with credit card fraud detection is data privacy and protection. Financial institutions and organizations are required by law to protect sensitive customer data, including credit card details, transaction history, and personal information. Different countries have various legal frameworks to ensure data protection:

* **General Data Protection Regulation (GDPR)**: In the European Union, GDPR mandates strict rules regarding the collection, storage, and processing of personal data. Organizations need to obtain explicit consent from individuals before collecting their data, and they are required to provide mechanisms for individuals to access, correct, or delete their data.
* **California Consumer Privacy Act (CCPA)**: In the U.S., California’s CCPA provides similar protections, allowing consumers to opt out of the sale of their data, access their data, and request its deletion.
* **Payment Card Industry Data Security Standard (PCI DSS)**: This set of standards governs the protection of cardholder data. Organizations that process, store, or transmit credit card information are required to follow these standards to prevent data breaches and fraud.

When using machine learning models for fraud detection, organizations must ensure that they comply with these regulations. This means:

* **Data anonymization and encryption**: It is crucial that personal data used for training the fraud detection models is anonymized or encrypted. Direct access to personally identifiable information (PII) should be minimized.
* **Limited data usage**: Data used for training and testing models should be limited to what is necessary for the task, reducing the risk of over-collection or misuse.
* **Transparency and accountability**: Organizations should be transparent with customers regarding how their data is being used and the steps taken to protect it.

## Regulatory Compliance

Regulatory compliance is another legal concern for financial institutions. Different financial regulatory bodies govern credit card fraud detection practices, including the **Financial Conduct Authority (FCA)** in the UK and **Federal Financial Institutions Examination Council (FFIEC)** in the U.S. These bodies enforce laws and guidelines to ensure that financial institutions protect consumers and minimize fraud risk.

Machine learning-based fraud detection systems must be designed to meet these regulatory requirements. For example, if an institution is using an automated system to flag fraudulent transactions, it must ensure that the system provides accurate, reliable, and transparent results. This is critical because financial institutions may be held accountable for failing to prevent fraudulent activities, especially if their systems were not compliant with regulatory standards.

## Liability for False Positives and Negatives

A significant challenge in fraud detection is the handling of false positives (legitimate transactions flagged as fraudulent) and false negatives (fraudulent transactions that are not detected). From a legal perspective:

* **False positives**: When a legitimate transaction is incorrectly flagged as fraudulent, it may lead to an inconvenience for the customer, such as their card being temporarily blocked or their account frozen. In extreme cases, customers may take legal action for damages incurred due to the wrongful blocking of their transactions or the negative impact on their credit score.
* **False negatives**: Conversely, failing to detect a fraudulent transaction could result in significant financial losses for both the customer and the financial institution. If a model fails to identify fraudulent activity, institutions may be liable for the resulting losses, especially if it is found that the system did not meet the required standards for fraud prevention.

Organizations must strike a balance between minimizing false positives and maximizing fraud detection to limit legal risks.

## Ethical Issues

## Bias and Fairness in Machine Learning Models

Bias is a significant ethical issue in credit card fraud detection. Machine learning models are only as good as the data they are trained on. If the data used to train fraud detection systems contains biases—such as certain demographic groups being overrepresented or underrepresented—then the resulting model can be unfair and discriminatory. For example, if the training data predominantly consists of transactions from high-income individuals, the model may have a higher rate of false negatives for transactions from low-income individuals.

Key ethical concerns include:

* **Discrimination against minority groups**: If fraud detection models unfairly target specific demographic groups (e.g., based on age, gender, race, or location), they may disproportionately flag individuals from those groups as potential fraudsters. This can lead to unjust financial scrutiny or customer discrimination, harming the affected individuals.
* **Data-driven inequality**: Bias in machine learning can exacerbate existing inequalities. If fraud detection systems are not inclusive in their training data, they may not perform well for customers from diverse backgrounds, leading to unequal protection against fraud for certain groups.

To mitigate bias, organizations must ensure that the data used to train fraud detection systems is representative of the full range of customers and that efforts are made to detect and correct for any inherent biases. Additionally, regular audits of the models' performance across different demographic groups should be conducted.

## Transparency and Explainability

Machine learning models, particularly deep learning models, can sometimes operate as "black boxes," meaning it is difficult to understand how they arrive at a decision. In the context of credit card fraud detection, a lack of explainability could raise ethical concerns regarding transparency and accountability.

* **Lack of explainability**: Customers who have their transactions flagged as fraudulent should have the right to understand why their transaction was flagged. If a model’s decision-making process is opaque, customers may not trust the system or may feel that they have been treated unfairly.
* **Model interpretability**: From an ethical standpoint, fraud detection systems should be interpretable so that customers can request clarifications if they believe a mistake was made. This is important for maintaining customer trust and ensuring that their rights are respected.

Ethical guidelines in AI and machine learning emphasize the importance of explainability. Researchers and practitioners must work on techniques to make fraud detection models more transparent, such as using simpler, interpretable models or integrating post hoc interpretability tools.

## Informed Consent

Informed consent is another key ethical issue. Financial institutions must ensure that customers are fully informed about how their data is being used. This involves explaining the purposes for which their data will be used, including for fraud detection, and obtaining explicit consent.

The ethical principle of informed consent extends beyond just collecting customer data. It also involves ensuring that customers have the option to opt-out of being monitored or to access their data. This can help address concerns about data surveillance and the unauthorized use of personal information.

## Professional Issues

## Accountability and Responsibility

Professionals involved in the development and deployment of machine learning models for fraud detection must be accountable for their actions. This includes:

* **Data scientists**: Data scientists must ensure that the models they develop are thoroughly tested, validated, and meet the necessary legal and ethical standards.
* **Developers**: Developers are responsible for implementing the fraud detection system in a way that ensures transparency, privacy, and fairness.
* **Management**: Organizational leaders must ensure that adequate resources are allocated to data protection and compliance efforts, and that ethical considerations are prioritized in decision-making.

A clear structure of accountability should be established within the organization to ensure that each stakeholder is aware of their responsibilities and that any issues are promptly addressed.

## Professional Competence

Another critical professional issue is ensuring that those working on fraud detection systems possess the necessary knowledge and skills. Given the complexity of machine learning algorithms and the regulatory environment surrounding financial systems, it is essential that professionals involved in fraud detection are competent in:

* **Machine learning**: Professionals should have expertise in machine learning, particularly in the areas of model selection, evaluation, and dealing with imbalanced datasets.
* **Financial regulations**: Understanding the legal requirements and compliance standards related to financial transactions and data protection is crucial for building models that meet industry standards.
* **Ethical AI**: Developers must be well-versed in the ethical implications of AI and machine learning and be committed to ensuring fairness, transparency, and privacy in their work.

Professional organizations, such as the **Association for Computing Machinery (ACM)** and **IEEE**, provide guidelines and best practices for AI and machine learning ethics. Practicing professionals should adhere to these standards to ensure that their work is aligned with industry norms and ethical expectations.

# IMPLEMENTATION

## Implementation

The implementation of a credit card fraud detection system using machine learning involves several stages, including data collection, preprocessing, model development, training, evaluation, and deployment. This section provides an in-depth explanation of each phase, highlighting the methodologies and techniques typically used to build an effective fraud detection system. The goal is to create a model capable of accurately identifying fraudulent transactions while handling the challenges associated with imbalanced datasets.

## Data Collection and Understanding

The first step in implementing a machine learning-based fraud detection system is acquiring the relevant data. Financial institutions and payment processors typically have vast amounts of transactional data, which can be used to detect patterns indicative of fraudulent behavior.

### Source of Data

For machine learning models, the most commonly used data comes from historical credit card transactions. These transactions typically contain the following features:

* **Transaction Amount**: The value of the transaction.
* **Merchant Information**: The type and name of the merchant involved in the transaction.
* **Transaction Timestamp**: The date and time the transaction occurred.
* **Location Data**: The location where the transaction was made, often extracted from the cardholder’s geolocation.
* **Cardholder Information**: Demographic information about the customer, such as their spending habits, geographical location, and account status.
* **Transaction ID and Account ID**: Unique identifiers for both the transaction and the cardholder.

To build an effective fraud detection model, it is essential that the dataset includes both fraudulent and legitimate transactions. The dataset will typically be highly imbalanced, with fraudulent transactions representing a very small percentage of the overall data, making it challenging to train accurate models. Specialized techniques for handling imbalanced datasets will be explored later in the implementation.

### Exploratory Data Analysis (EDA)

Before jumping into model building, a thorough exploratory data analysis (EDA) is necessary to understand the data and identify any underlying patterns or potential issues. During EDA, the following steps are typically performed:

* **Data distribution**: Analyze the distribution of various features (such as transaction amounts and time) to identify potential anomalies or outliers that may need to be handled.
* **Class imbalance check**: Given that fraudulent transactions are rare, this is an important step. Visualizing the ratio of fraudulent to legitimate transactions helps in determining the severity of the imbalance.
* **Correlation analysis**: Assessing correlations between different features can reveal relationships that may help improve model performance.
* **Missing data analysis**: Check for missing or incomplete records, which may require imputation or removal, depending on their impact on model performance.

### Data Preprocessing

Data preprocessing is a crucial step in preparing the data for machine learning models. In the context of credit card fraud detection, the following preprocessing steps are common:

* **Handling Missing Values**: Missing values in transactional data need to be either removed or imputed using techniques like mean/mode imputation, regression imputation, or more sophisticated methods like k-Nearest Neighbors (k-NN) imputation.
* **Normalization/Standardization**: Since machine learning models like decision trees or neural networks are sensitive to scale, it is often necessary to normalize or standardize numerical features (e.g., transaction amounts, merchant categories) to ensure they contribute equally to the model’s performance.
* **Categorical Encoding**: Categorical features such as merchant name, customer ID, or transaction type need to be encoded into a numerical format. Methods like **One-Hot Encoding** or **Label Encoding** can be used to handle these categorical variables.
* **Feature Engineering**: Feature engineering is the process of creating new features from existing ones to improve model performance. For example:
  + **Rolling Aggregates**: Generate rolling statistics for each customer, such as the average transaction amount over the past 30 days, to detect unusual behavior.
  + **Time-based Features**: Features such as the hour of the day, day of the week, or time since the last transaction can provide valuable insights into potential fraud patterns.

### Handling Imbalanced Datasets

Since fraudulent transactions are rare, an imbalance between legitimate and fraudulent transactions is inevitable. This imbalance poses a significant challenge for machine learning algorithms, which tend to perform poorly when class distribution is skewed.

Several techniques can be employed to address the class imbalance:

* **Resampling**:
  + **Oversampling** the minority class (fraudulent transactions) by replicating existing instances or generating synthetic samples using techniques such as **SMOTE (Synthetic Minority Over-sampling Technique)**.
  + **Undersampling** the majority class (legitimate transactions) to balance the dataset, though this can lead to loss of valuable information.
* **Class Weights**: Most machine learning algorithms allow the specification of class weights, which assigns a higher weight to the minority class. This forces the model to pay more attention to fraudulent transactions during training.
* **Anomaly Detection**: Given that fraud is rare, anomaly detection models can be useful. These models learn the patterns of legitimate transactions and identify outliers (fraudulent transactions) that deviate significantly from normal behavior.

## Model Selection and Training

After preprocessing the data, the next step is selecting appropriate machine learning models and training them on the dataset.

### Model Choice

Several machine learning algorithms can be used for credit card fraud detection. The choice of model depends on factors such as interpretability, performance, and ability to handle imbalanced datasets. Some common models include:

* **Logistic Regression**: A simple yet powerful algorithm that can be used as a baseline for fraud detection. It outputs probabilities, which can be useful for identifying high-risk transactions.
* **Random Forest**: An ensemble method that creates multiple decision trees and aggregates their predictions. It is robust to overfitting and can handle both binary classification and imbalanced datasets effectively.
* **Gradient Boosting Machines (GBM)**: Algorithms like **XGBoost** and **LightGBM** are widely used for fraud detection due to their ability to handle large datasets and produce highly accurate models.
* **Neural Networks**: Deep learning models, particularly feedforward neural networks, can be used to model complex relationships between features. However, they often require larger datasets and more computational power.
* **Anomaly Detection Models**: Algorithms like **Isolation Forest** and **One-Class SVM** are particularly effective when dealing with imbalanced data, as they focus on detecting outliers.

### Training the Model

Once a model has been selected, the next step is to train it using the prepared dataset. During training, the model learns the relationship between the input features and the target variable (fraud or not fraud). Key steps in this process include:

* **Cross-Validation**: To avoid overfitting and ensure the model generalizes well, k-fold cross-validation is used. The dataset is divided into k subsets, and the model is trained on k-1 subsets, with the remaining subset used for validation. This process is repeated for each subset.
* **Hyperparameter Tuning**: Hyperparameters such as the learning rate, number of trees (for Random Forest), or the number of layers (for neural networks) can significantly impact model performance. Techniques like **Grid Search** or **Random Search** can be used to find the best combination of hyperparameters.

### Model Evaluation

Evaluating the model's performance is crucial to determine how well it is detecting fraudulent transactions. The following metrics are commonly used in fraud detection:

* **Accuracy**: The proportion of correct predictions (both fraud and legitimate transactions) to the total number of predictions. However, accuracy is often misleading in imbalanced datasets.
* **Precision**: The ratio of true positives (correctly predicted fraudulent transactions) to the total predicted positives (both true positives and false positives). High precision ensures that fraudulent transactions are detected with minimal false alarms.
* **Recall**: The ratio of true positives to the total actual positives (true positives and false negatives). High recall ensures that most fraudulent transactions are identified.
* **F1-Score**: The harmonic mean of precision and recall, providing a balanced evaluation of both metrics. This is particularly important in fraud detection, where both false positives and false negatives need to be minimized.
* **ROC-AUC**: The area under the Receiver Operating Characteristic curve. AUC measures the ability of the model to distinguish between fraudulent and legitimate transactions, with a higher value indicating better model performance.

### Model Interpretation

Given that fraud detection systems have a significant impact on customers, it is important to be able to interpret the model's decisions. **SHAP (Shapley Additive Explanations)** or **LIME (Local Interpretable Model-Agnostic Explanations)** are popular tools for explaining model predictions. These techniques provide insight into which features (e.g., transaction amount, merchant type, etc.) are driving the model’s decisions.

## Model Deployment

Once the model has been trained and evaluated, it is ready for deployment. The deployment process involves integrating the fraud detection system into the existing infrastructure of the financial institution or payment processor.

### Real-Time Processing

For fraud detection to be effective, the model must operate in real-time. This requires efficient processing of incoming transaction data, where the model predicts whether each transaction is fraudulent or legitimate within milliseconds. To achieve this:

* **Streamlining the Model**: Ensure that the model can process new data points efficiently, without introducing latency.
* **Batch Processing**: For systems with limited real-time processing capabilities, batch processing may be used to evaluate transactions periodically.

### Continuous Monitoring and Updates

Fraud patterns evolve over time, so it is essential to regularly update the fraud detection model to maintain its effectiveness. This can be done through **model retraining** using new transaction data, and by periodically reviewing the system’s performance to detect any degradation.

# RESULTS AND DISCUSSIONS

## Results

The Logistic Regression model was trained and evaluated to detect fraudulent credit card transactions in a highly imbalanced dataset. The classification report, confusion matrix, and ROC curve were used to assess the model's performance.

The classification report showed that for legitimate transactions (class 0), the model achieved a precision of 1.00, a recall of 0.98, and an F1-score of 0.99. For fraudulent transactions (class 1), the model achieved a precision of 0.09, a recall of 0.95, and an F1-score of 0.17. The overall accuracy across all transactions was 98%, calculated on a total of 18,239 transactions.

The macro-averaged precision and recall were 0.55 and 0.97, respectively, with a macro-averaged F1-score of 0.58. The weighted average F1-score was 0.99, heavily influenced by the large proportion of legitimate transactions.

The confusion matrix indicated that the model correctly classified 17,794 legitimate transactions and incorrectly classified 401 legitimate transactions as fraud. For the fraudulent transactions, the model correctly identified 42 cases and misclassified 2 as legitimate.

The Receiver Operating Characteristic (ROC) curve showed an AUC (Area Under Curve) of 0.9846, suggesting excellent model performance in distinguishing between the two classes. The ROC curve itself showed a sharp rise towards the top-left corner of the plot, indicating a high true positive rate with a low false positive rate across a range of thresholds.

Visual inspection of the confusion matrix confirmed that false negatives (missed fraudulent transactions) were extremely rare, while false positives (legitimate transactions flagged as fraud) were more common.

Overall, the Logistic Regression model demonstrated strong recall for the minority fraud class while maintaining a high overall accuracy, albeit at the cost of lower precision for fraudulent transaction predictions.

A screenshot of a computer

AI-generated content may be incorrect.

## Overall Performance Metrics

| **Metric** | **Value** |
| --- | --- |
| **Accuracy** | 98% |
| **Precision (Fraudulent Class 1)** | 0.09 |
| **Recall (Fraudulent Class 1)** | 0.95 |
| **F1-Score (Fraudulent Class 1)** | 0.17 |
| **Macro Avg Precision** | 0.55 |
| **Macro Avg Recall** | 0.97 |
| **Macro Avg F1-Score** | 0.58 |
| **Weighted Avg F1-Score** | 0.99 |
| **True Negatives (TN)** | 17,794 |
| **False Positives (FP)** | 401 |
| **True Positives (TP)** | 42 |
| **False Negatives (FN)** | 2 |
| **ROC-AUC Score** | 0.9846 |

**Brief Notes:**

* **High recall (0.95)**: most frauds detected.
* **Low precision (0.09)**: many false alarms.
* **Excellent ROC-AUC (0.9846)**: strong overall separability.

## Discussion

The results highlight several important insights into the performance of Logistic Regression when applied to credit card fraud detection in an imbalanced dataset.

Firstly, the model’s outstanding recall of 0.95 for fraudulent transactions suggests that it is highly effective at identifying fraud. This is a critical requirement in fraud detection systems, where missing fraudulent activities (false negatives) can lead to significant financial and reputational damages. Even at the expense of precision, maximizing recall is often prioritized in practical fraud detection scenarios, where it is preferable to investigate false alarms than to overlook actual frauds.

However, the extremely low precision (0.09) for fraudulent transactions indicates that the majority of transactions flagged as fraud were actually legitimate. While a high recall minimizes the risk of missed fraud, a low precision can create operational inefficiencies, such as increased manual reviews by fraud investigation teams and potential customer dissatisfaction when legitimate transactions are blocked or challenged unnecessarily. In a real-world application, this trade-off would need to be carefully managed depending on the institution's tolerance for false positives and resource availability.

The confusion matrix provided further evidence of the model’s behavior. It demonstrated that while false negatives were nearly eliminated (only 2 missed fraudulent transactions), false positives were relatively high (401 legitimate transactions incorrectly flagged). Given the extreme class imbalance, this result is not unexpected. Logistic Regression models typically produce a probability output, and in this case, the threshold for classifying a transaction as fraudulent was likely tuned to favor sensitivity over specificity.

The high ROC-AUC score of 0.9846 is particularly encouraging, indicating that the model has an excellent ability to separate fraudulent from legitimate transactions overall. The ROC curve’s steep ascent toward the top-left corner supports this, showing that even across various thresholds, the model consistently maintains a strong true positive rate while keeping the false positive rate reasonably low.

The macro-averaged metrics also provide useful information. A macro-averaged F1-score of 0.58 reflects that when treating each class equally, the model’s balanced performance is moderate. This underlines the difficulty in achieving high performance across both classes simultaneously in highly imbalanced datasets. The weighted average metrics, skewed by the dominance of legitimate transactions, were significantly higher, demonstrating how accuracy and weighted metrics can be misleading if interpreted without considering class distribution.

The use of resampling techniques to balance the training data likely contributed to the model’s ability to achieve high recall for fraudulent transactions. Without balancing the dataset, the model might have been biased towards predicting only legitimate transactions, given their overwhelming majority in the data.

An important observation is the potential for threshold adjustment to fine-tune the balance between precision and recall. Logistic Regression outputs probabilities, and by adjusting the threshold above the default 0.5, it may be possible to achieve a better balance between identifying frauds and minimizing false alarms. However, any threshold adjustment would need to be informed by the specific operational needs and cost structures associated with fraud investigation and customer relations.

It is also worth noting that the Logistic Regression model's inherent simplicity brings advantages beyond performance metrics. Its transparency and interpretability make it particularly suitable for domains like finance, where explainability of decisions is crucial for compliance and customer communication. Although feature importance was not directly explored in the presented results, Logistic Regression coefficients offer a straightforward way to understand which features most influence the model’s decision-making process.

Despite its simplicity, the model’s performance demonstrates that traditional machine learning models like Logistic Regression remain strong contenders for fraud detection tasks when properly tuned and evaluated. While more complex models, such as ensemble methods or deep learning, might achieve marginally better precision-recall trade-offs, Logistic Regression offers a strong baseline with easier interpretability and faster deployment times.

In summary, the results show that the Logistic Regression model is capable of effectively detecting fraudulent credit card transactions in a highly imbalanced dataset by maximizing recall, even though it sacrifices precision. The high ROC-AUC and strong recall metrics are particularly promising, suggesting that with additional threshold tuning or the inclusion of ensemble techniques, further performance improvements could be realized.

# CONCLUSION

## Research Summary

This research focused on detecting fraudulent credit card transactions using machine learning techniques, particularly logistic regression, on a highly imbalanced dataset. The model was evaluated using key performance metrics, including accuracy, precision, recall, F1-score, and ROC-AUC. Results demonstrated that while the model achieved a high overall accuracy of 98%, precision for fraudulent transactions was relatively low (0.09), highlighting the challenge of false positives. However, the recall for fraudulent cases was extremely high (0.95), meaning that most actual fraud cases were successfully identified. The ROC-AUC score of 0.9846 confirmed that the model performed exceptionally well in distinguishing fraudulent from legitimate transactions. Despite the simplicity of logistic regression, it proved highly effective, especially when combined with appropriate resampling and evaluation techniques. The study emphasized the importance of considering recall over precision in fraud detection systems, where the cost of missing a fraudulent transaction is significantly higher than falsely flagging a legitimate one. The findings suggest that even basic machine learning models, if properly tuned, can achieve competitive results in sensitive applications like fraud detection.

## Limitation

While the model achieved promising results, several limitations were identified. Firstly, the dataset was highly imbalanced, which influenced overall performance metrics like accuracy and weighted F1-scores, possibly making them appear overly optimistic. Although resampling and careful evaluation were used, more sophisticated techniques such as SMOTE or ensemble methods could further improve precision. Secondly, only logistic regression was explored; advanced models like random forests, XGBoost, or neural networks were not tested, which may outperform logistic regression, especially in complex fraud patterns. Thirdly, feature engineering was limited; the study did not deeply explore creating new variables or feature selection methods that might boost model interpretability and performance. Additionally, the model evaluation was based on historical data and may not fully capture evolving fraud strategies used in real-world scenarios. Model generalizability to new, unseen types of fraud remains uncertain. Finally, operational costs associated with high false positive rates were not quantitatively assessed, which is crucial for deploying fraud detection systems in a live environment. Future work should address these gaps by testing more algorithms, applying advanced balancing techniques, and considering real-world operational impacts.

# Final Thoughts

This research demonstrated the effectiveness of logistic regression in detecting fraudulent credit card transactions in a highly imbalanced dataset. While the model achieved a high overall accuracy, the performance metrics highlighted the importance of balancing precision and recall, especially in fraud detection systems. A recall of 0.95 showcased the model’s ability to identify most fraudulent transactions, which is crucial given the potentially high costs associated with missing fraud cases. However, the trade-off came with a low precision (0.09), leading to many false positives that would require manual investigation.

Despite these challenges, the findings underscore the strength of simple models like logistic regression when used in combination with appropriate data handling and evaluation techniques. The high ROC-AUC score (0.9846) also indicates that logistic regression can effectively distinguish between fraudulent and legitimate transactions, making it a valuable tool in fraud detection.

Nevertheless, the study also revealed key limitations, including the need for better precision, more sophisticated resampling techniques, and the potential of exploring more advanced models. Future work should focus on improving model performance with feature engineering, advanced algorithms, and deeper analysis of false positive impacts in a real-world operational environment. Ultimately, this research paves the way for further exploration into machine learning models for financial fraud detection.

# References

# Appendix