# SPOTTING FRAUDULENT TRANSACTIONS FROM CREDIT CARDS USING ML MODELS

**NAOMI JOBSON MITCHUAL**

# Table of Contents

# Introduction

Fraudulent transactions which refer to any unauthorised or deceitful activity conducted on an individual's or entity's account for financial gain or other benefits is a key aspect of financial crime. These transactions can take various forms, including credit card fraud, cheque fraud, identity theft, account takeover and many more.

Fraudulent transactions pose a significant threat to individuals, businesses, and financial institutions, leading to financial losses, reputational damage, and legal consequences. In addition to causing significant financial losses, financial fraud also erodes public faith in the financial system (Association of Certified Fraud Examiners, 2020).

Vigilance, security measures, and prompt reporting are crucial for detecting and preventing fraudulent activity. New technology advancements have also shown that machine learning techniques and AI tools can be applied successfully to detect fraudulent transactions in large amounts of transactional data. Such methods have the capacity to detect fraudulent transactions that human assessors may not be able to see and work on to meet real time deadlines.

Dhankhad  et al demonstrated that by  automatically mining large volumes of transactional data, financial institutions and other organisations can gain valuable intuitions into their operations, improve decision making process and increase efficiency. Financial institutions can also scrutinise past transactional data to find patterns related to fraudulent actions by utilising machine learning techniques like supervised learning, unsupervised learning and deep learning.

Furthermore, by monitoring transactions continuously in real-time, conducting reviews, performing routine due diligence using customers data and sending out alerts or red flags for suspected fraud will go a long way to integrate and improve fraud detection. When fraud cannot be prevented, it is desirable to detect it as rapidly as possible. (Dal Pozzolo et al., 2015). By implementing this proactive above-named strategy, potential losses and damages are minimized through timely action. Application of ML algorithms can be highly relevant in retail business where can be used to analyse customer purchase histories and predict future buying behaviour and optimise marketing strategies.

This project explores how real-time transaction monitoring systems and machine learning algorithms can be used to accurately identify fraudulent transactions and improve financial transaction, fraud detection and prevention capabilities. I will be looking for patterns and abnormalities using machine learning algorithms, leading to more accurate and adaptable fraud detection. I will use machine learning algorithms to detect fraudulent transactions by learning patterns from past fraudulent activities and flagging transactions that deviate significantly from these patterns.

# Aims and Objective

The aim of this project is to use knowledge gained from machine learning to build a system that will identify fraud within the financial industry specifically in retail transactions. It is anticipated that the results of this research will enhance the efficiency of analysing and pinpointing fraudulent activities.

Our three primary objectives include the following:

1. Reviewing existing literature on financial fraud detection to gain insights into various facets of the issue.

2. Addressing the challenges of detecting financial fraud using supervised machine learning methods on a readily accessible dataset.

3. Evaluating and contrasting various classification methodologies to determine the most effective approach for credit card fraud detection.

# Methodology

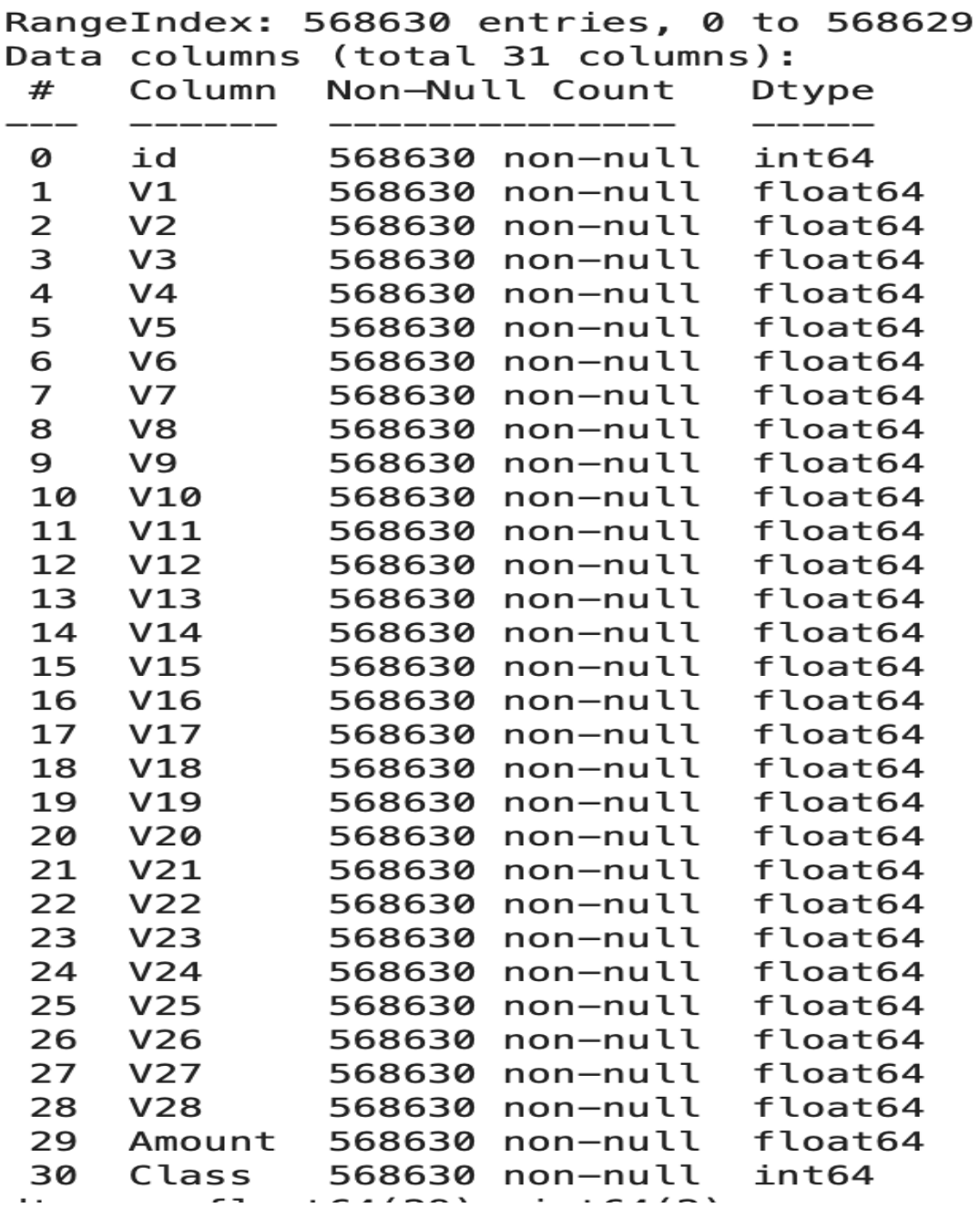
## Data Collection

The data for this study was sourced from Kaggle platform. It contains credit card transactions across Europe for the year 2023. The dataset spans several columns and contains 568,630 records. There are a total of 31 variables. However, many of the variables were anonymized due to the nature of their sensitivity i.e. some contained Personally Identifiable Information such as addresses, ID number, jobs, age etc. as shown in the anonymized data below in Table 1. These variables are marked by random names denoted by V1 - V27. Figure 1 above shows the first few rows of the dataset.

| | id | V1 | V2 | V3 | V4 | V5 | V6 | V7 | V8 | V9 | ... | V21 | V22 | V23 | V24 | V |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 0 | -0.260648 | -0.469648 | 2.496266 | -0.083724 | 0.129681 | 0.732898 | 0.519014 | -0.130006 | 0.727159 | ... | -0.110552 | 0.217606 | -0.134794 | 0.165959 | 0.126 |
| **1** | 1 | 0.985100 | -0.356045 | 0.558056 | -0.429654 | 0.277140 | 0.428605 | 0.406466 | -0.133118 | 0.347452 | ... | -0.194936 | -0.605761 | 0.079469 | -0.577395 | 0.190 |
| **2** | 2 | -0.260272 | -0.949385 | 1.728538 | -0.457986 | 0.074062 | 1.419481 | 0.743511 | -0.095576 | -0.261297 | ... | -0.005020 | 0.702906 | 0.945045 | -1.154666 | -0.605 |
| **3** | 3 | -0.152152 | -0.508959 | 1.746840 | -1.090178 | 0.249486 | 1.143312 | 0.518269 | -0.065130 | -0.205698 | ... | -0.146927 | -0.038212 | -0.214048 | -1.893131 | 1.003 |
| **4** | 4 | -0.206820 | -0.165280 | 1.527053 | -0.448293 | 0.106125 | 0.530549 | 0.658849 | -0.212660 | 1.049921 | ... | -0.106984 | 0.729727 | -0.161666 | 0.312561 | -0.414 |

5 rows × 31 columns

Table 1 – Anonymized Data (*First five rows)*



```
RangeIndex: 568630 entries, 0 to 568629
Data columns (total 31 columns):
 #    Column  Non-Null Count   Dtype
---   ------  --------------   -----
 0    id      568630 non-null  int64
 1    V1      568630 non-null  float64
 2    V2      568630 non-null  float64
 3    V3      568630 non-null  float64
 4    V4      568630 non-null  float64
 5    V5      568630 non-null  float64
 6    V6      568630 non-null  float64
 7    V7      568630 non-null  float64
 8    V8      568630 non-null  float64
 9    V9      568630 non-null  float64
 10   V10     568630 non-null  float64
 11   V11     568630 non-null  float64
 12   V12     568630 non-null  float64
 13   V13     568630 non-null  float64
 14   V14     568630 non-null  float64
 15   V15     568630 non-null  float64
 16   V16     568630 non-null  float64
 17   V17     568630 non-null  float64
 18   V18     568630 non-null  float64
 19   V19     568630 non-null  float64
 20   V20     568630 non-null  float64
 21   V21     568630 non-null  float64
 22   V22     568630 non-null  float64
 23   V23     568630 non-null  float64
 24   V24     568630 non-null  float64
 25   V25     568630 non-null  float64
 26   V26     568630 non-null  float64
 27   V27     568630 non-null  float64
 28   V28     568630 non-null  float64
 29   Amount  568630 non-null  float64
 30   Class   568630 non-null  int64
```

*Figure 1: Data Variables*

Each transaction is labelled as either fraudulent (1) or non-fraudulent (0), making it an ideal dataset for training and evaluating fraud detection systems. The column "*Class*" is the target column, and it contains the label as to whether the transaction is fraudulent or not (1 or 0).

## Exploratory Data Analysis (EDA)

In exploring the dataset through Exploratory Data Analysis (EDA), a notable observation emerges regarding the distribution of the target variable: it exhibits a balanced 50-50 ratio. This finding is visually depicted through both a bar plot and a pie chart, both of which vividly illustrate the equal proportion of occurrences between the two classes within the target variable. Such balance within the dataset is crucial for modelling tasks, as it ensures that neither class overwhelms the predictive capability of the model, thereby averting potential biases. This symmetrical distribution provides a solid foundation for subsequent analyses and model development, promising robust and equitable predictive outcomes. Figures 2 and 3 show plots depicting these distributions.

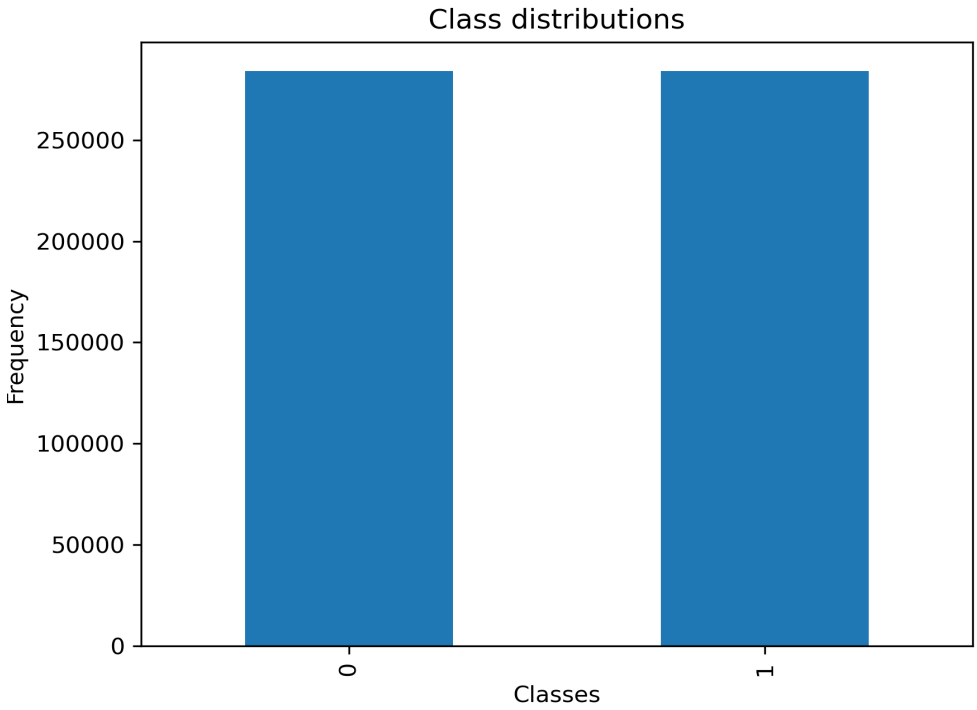



*Figure 2: Bar plot showing the class distribution.*

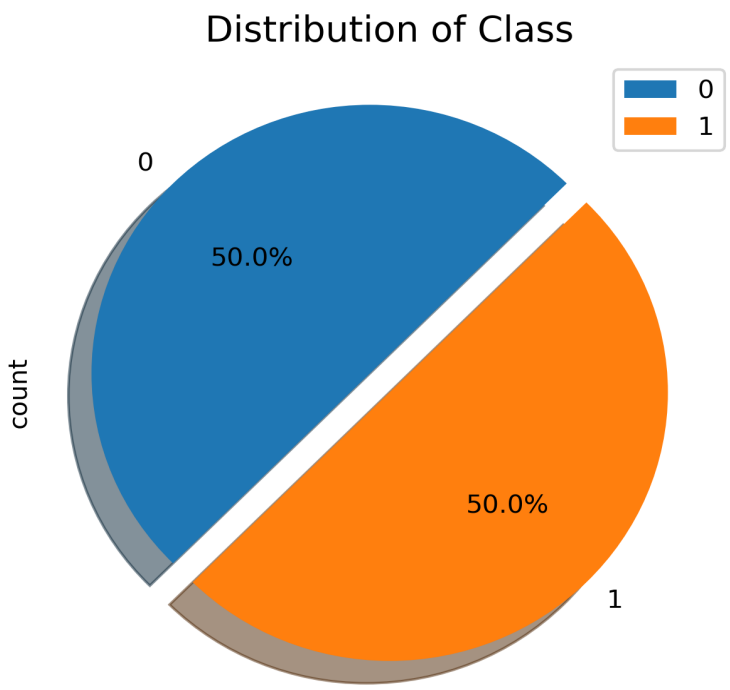



*Figure 3: Pie chart showing the class distribution.*

Visualising the amount involved in a transaction via a box plot reveals insightful details about the dataset's transactional characteristics. The whiskers extend from 0 to approximately 24000, indicating the overall range of transaction amounts. With a median value positioned at 12000, the box plot highlights the central tendency of the data. The interquartile range, delineated by the box, spans from 6000 to 18000, encapsulating the middle 50% of transaction values. This representation effectively communicates the spread and distribution of transaction amounts, providing a comprehensive overview for further analysis and model development. Figure 4 shows this plot.

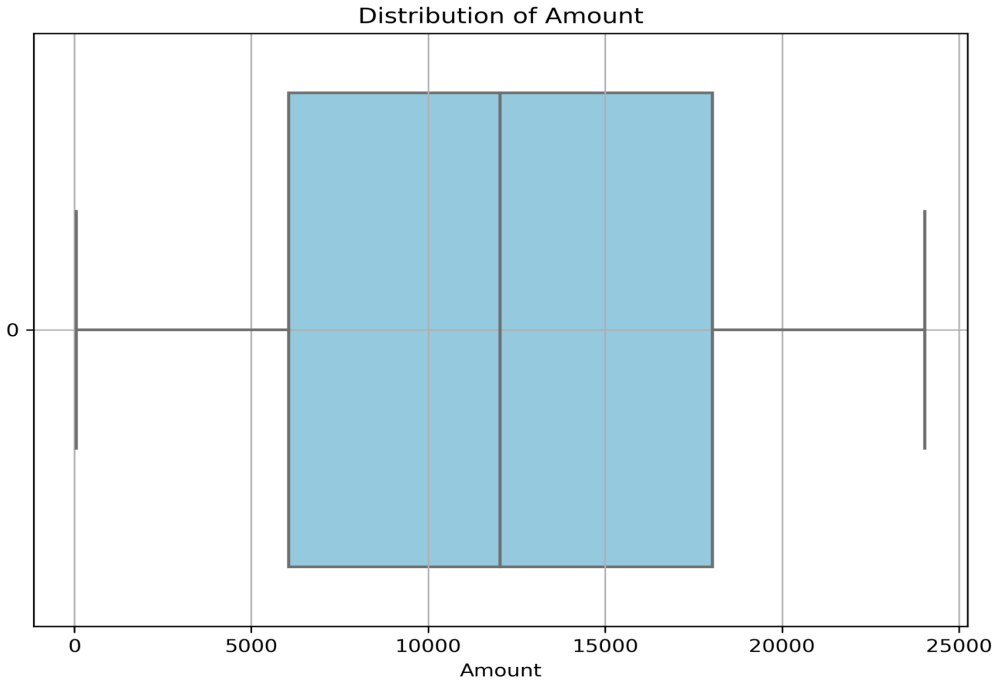



*Figure 4: Box plot showing the distribution of Amount involved in the transaction.*

## Data Preprocessing

Before utilising machine learning algorithms, several preprocessing steps were considered to clean and transform the data into a standardised format. This is done to make sure that the raw information is transformed into usable data. These steps include the following:

- Data Cleaning - Most machine learning algorithms struggle due to missing values and extreme outliers. It is important to ensure missing values and outliers are taken care of to ensure models converge gracefully. Luckily, this data did not contain any missing values. Also, the amount column does not have such extreme outliers.
- Normalisation/Standardisation - All numerical features were scaled to a common range to prevent certain features from dominating the model.
- Feature Selection - By selecting relevant and crucial features, which involves pinpointing the essential elements critical for fraud detection, thereby sifting through the data to detect and eliminate noise. This approach mitigated the risk of overfitting while concurrently improving the model's accuracy and interpretability.
- Handling Imbalanced Data - Using techniques like oversampling can be particularly useful for fraud detection as well as balancing the dataset. The “positive” class (frauds) is usually much smaller than the “negative” class (non-frauds), as fraudulent transactions are relatively rare compared to legitimate ones. However, identifying these rare fraudulent transactions is extremely important in the prevention of financial losses. Other algorithms that can handle class imbalances, such as random forests can also be employed. It’s important to note that though oversampling can improve the model performance, it can also lead to overfitting if not done carefully. Therefore, it’s a good idea to use techniques like cross-validation to ensure the model generalises well on unseen data. This data contained perfectly balanced classes with a 50-50 percentage ratio.
- Train-Test data set split - Splitting the data into training and testing sets. Typically, a large portion of the data is used for training and the rest for testing. In this study 75% of the data was used for training while 25% was used for testing. Stratified splitting on the target variable was used to ensure a class balance between the training and the testing sets.
- Feature selection - Feature selection entails identifying and choosing the most pertinent features that substantially contribute to fraud detection, taking into account their influence while diminishing computational complexity. The “Id” column was dropped as it is just used to uniquely identify a transaction and does not contribute to the model learning.

## Model Selection:

Fraud detection falls under supervised machine learning, and it is a binary classification problem to be specific as it entails assigning a transaction to either being fraudulent or not. After splitting the dataset into train and test sets, the following machine learning algorithms were fitted.

- Logistic Regression - Known for its simplicity and interpretability, it's ideal for binary classification tasks [3].
- Decision Trees - Able to capture complex decision boundaries and feature interactions.
- Random Forest - An ensemble technique that combines numerous decision trees, enhancing overall performance.
- Support vector Machines Classifier - a supervised machine learning algorithm for classification which works by finding the optimal hyperplane that best separates different classes.

# RESULTS

Logistic regression performs quite well with an overall accuracy of 95%. It shows a balanced performance in terms of precision and recall for both classes, indicating a good ability to classify instances of fraud (class 1) while minimising false positives (class 0). However, there's a slightly higher false negative rate (lower recall for class 1) compared to the false positive rate.

Classification Report - LogisticRegression

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.93 | 0.98 | 0.96 | 71079 |
| 1 | 0.98 | 0.93 | 0.95 | 71079 |
| accuracy | | | 0.95 | 142158 |
| macro avg | 0.96 | 0.95 | 0.95 | 142158 |
| weighted avg | 0.96 | 0.95 | 0.95 | 142158 |

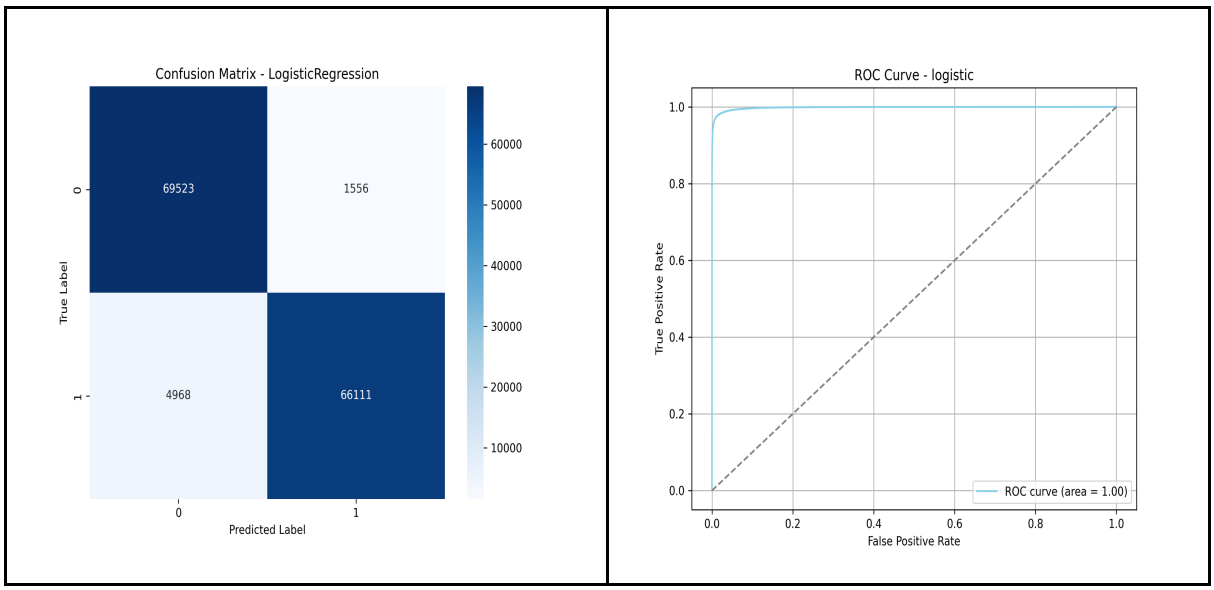



*Figure 5: Classification report, Confusion matrix and ROC curve for Logistic regression model*

Decision trees achieve exceptional performance with a perfect accuracy score. It demonstrates excellent precision, recall, and F1-score for both classes, indicating that it effectively captures the underlying patterns in the data. However, a perfect score might suggest overfitting, so it's essential to assess its performance on unseen data.

Classification Report - Decision Tree

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 1.00 | 0.99 | 1.00 | 71079 |
| 1 | 0.99 | 1.00 | 1.00 | 71079 |
| accuracy | | | 1.00 | 142158 |
| macro avg | 1.00 | 1.00 | 1.00 | 142158 |
| weighted avg | 1.00 | 1.00 | 1.00 | 142158 |

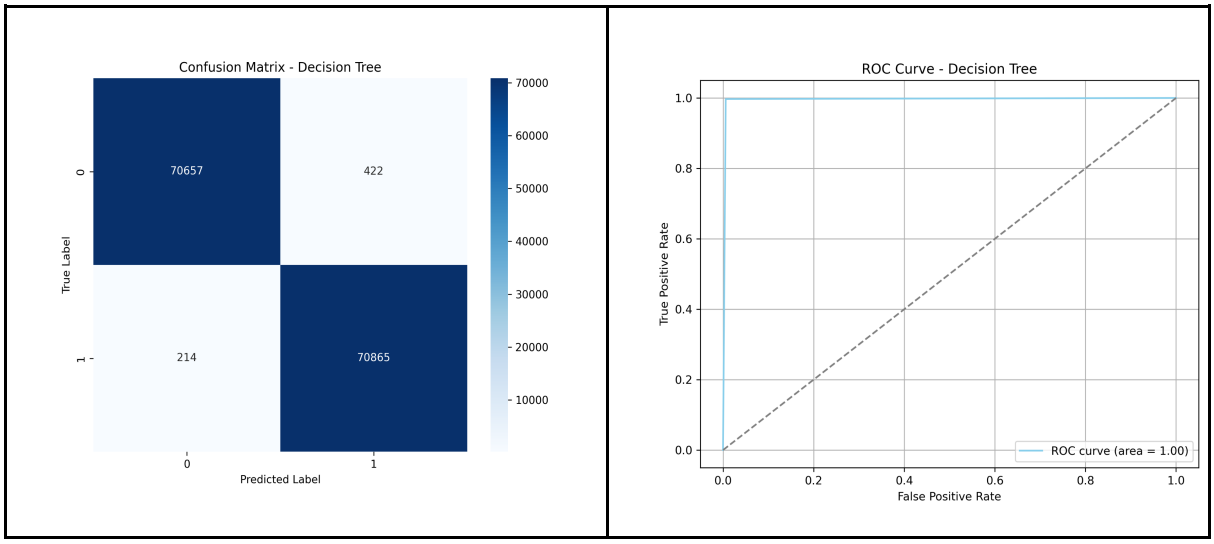



*Figure 6: Classification report, Confusion matrix and ROC curve for Decision tree model*

SVC also demonstrates strong performance with an accuracy of 98%. It maintains high precision and recall for both classes, indicating its effectiveness in correctly identifying instances of fraud while minimising misclassifications. The balanced precision and recall scores suggest a robust performance across different metrics.

Classification Report - SVC

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.97 | 0.99 | 0.98 | 71079 |
| 1 | 0.99 | 0.97 | 0.98 | 71079 |
| accuracy | | | 0.98 | 142158 |
| macro avg | 0.98 | 0.98 | 0.98 | 142158 |
| weighted avg | 0.98 | 0.98 | 0.98 | 142158 |

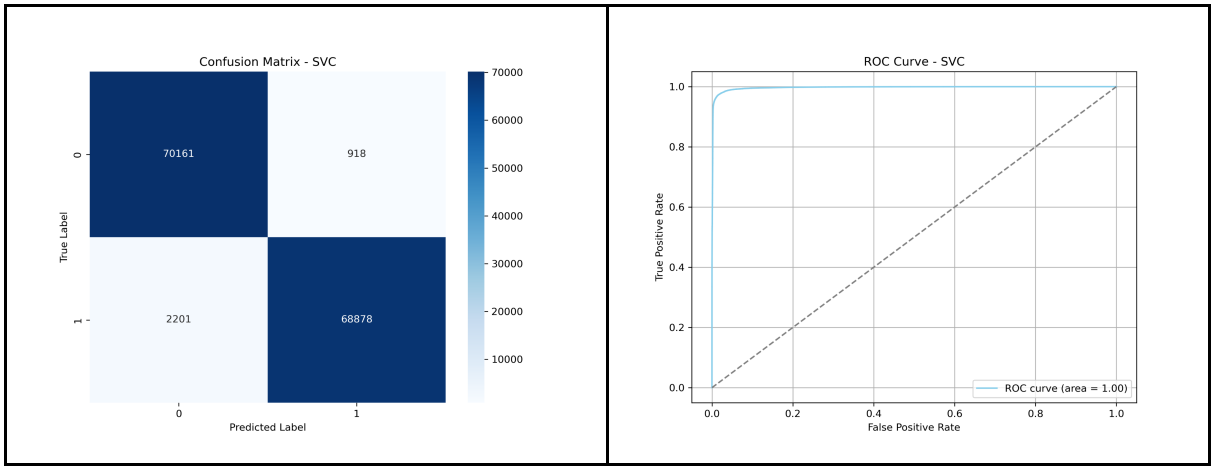



*Figure 7: Classification report, Confusion matrix and ROC curve for Support Vector Machine model*

Lastly, Random Forest achieves a perfect accuracy score, similar to decision trees. It demonstrates excellent precision, recall, and F1-score for both classes, suggesting a strong ability to classify instances accurately. However, like decision trees, the perfect score might indicate overfitting, requiring validation on unseen data.

Classification Report - RandomForest

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 1.00 | 1.00 | 1.00 | 71079 |
| 1 | 1.00 | 1.00 | 1.00 | 71079 |
| accuracy | | | 1.00 | 142158 |
| macro avg | 1.00 | 1.00 | 1.00 | 142158 |
| weighted avg | 1.00 | 1.00 | 1.00 | 142158 |

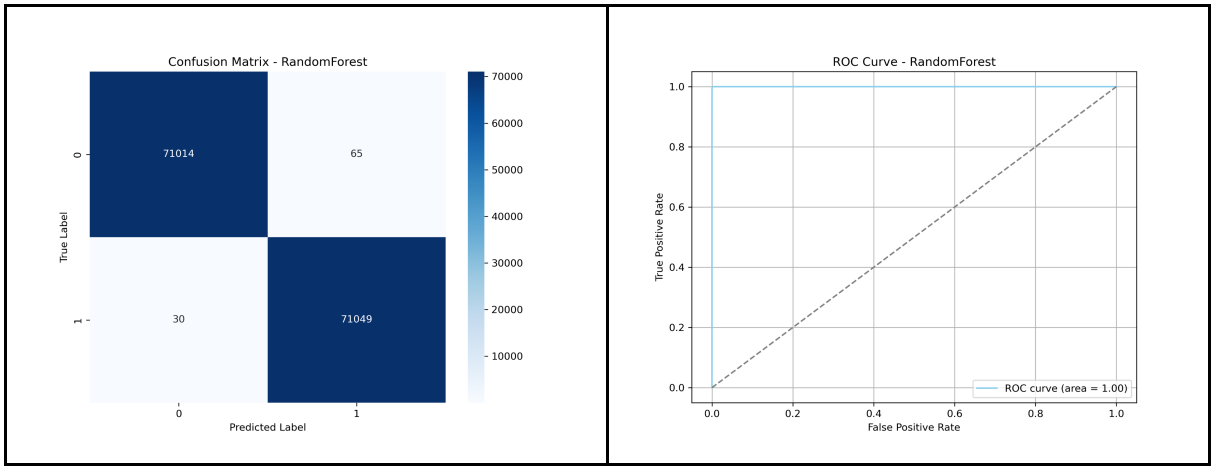



*Figure 8: Classification report, Confusion matrix and ROC curve for Random Forest model*

Overall, all models perform exceptionally well, with decision trees and random forest showing perfect accuracy scores, albeit with potential overfitting concerns. Logistic regression and SVC offer slightly lower but still robust performances, with a good balance between precision and recall for fraud detection. Decision trees and Random Forest classifiers showed perfect results.

# Ethical Considerations

In this study, the columns containing personally identifiable information (PII) data were anonymized for ethical consideration. By obfuscating sensitive details such as names, addresses, and account numbers, the privacy and confidentiality of individuals within the dataset were rigorously upheld. Anonymization serves as a crucial safeguard against the potential misuse or unauthorised access to personal data, aligning with ethical principles of data protection and privacy preservation.

By prioritising ethical considerations, this study underscores the commitment to responsible data handling practices and the protection of individuals' privacy in the pursuit of knowledge and insight.

# Limitations of the Study

One of the primary limitations of this study arises from the anonymization of data columns, which hindered the ability to discern interactions and correlations between various features within the dataset. The obscured nature of the data impeded comprehensive feature engineering efforts, as crucial insights into the nature and significance of each column were obscured.

Consequently, the study faced challenges in identifying which features most effectively predict the legitimacy or fraudulent nature of transactions. Without clear visibility into the underlying data structure, it became increasingly difficult to ascertain the predictive power of individual features and their collective impact on the target variable.

This limitation underscores the importance of balancing data privacy and analytical insights, highlighting the trade-offs inherent in anonymization practices within sensitive domains such as fraud detection.

Future research endeavours may benefit from alternative approaches to data anonymization that preserve privacy while facilitating robust analysis and modelling capabilities.

# Project Access

The project can be accessed on the GitHub repository through the link below.

https://github.com/naomimitchual/credit-card-fraud-detection

# References


1. D. Pozzolo, O. Caelen, R. A. Johnson and G. Bontempi, "Calibrating Probability with Undersampling for Unbalanced Classification," 2015 IEEE Symposium Series on Computational Intelligence, Cape Town, South Africa, 2015, pp. 159-166, doi: 10.1109/SSCI.2015.33.

2. Raghavan, P., & Gayar, N. E. (2020). Fraud Detection using Machine Learning and Deep Learning. In 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE) (pp. 334-339). IEEE. https://doi.org/10.1109/ICCIKE47802.2019.9004231

3. Mijwil, M. M., & Salem, I. E. (2020). Credit card fraud detection in payment using machine learning classifiers. *Asian Journal of Computer and Information Systems (ISSN: 2321–5658)*, *8*(4).

4. Dhankhad, S., Mohammed, E., & Far, B. (2018, July). Supervised machine learning algorithms for credit card fraudulent transaction detection: a comparative study. In *2018 IEEE international conference on information reuse and integration (IRI)* (pp. 122-125). IEEE.