Project Report

"A Study of Credit Card Fraud Detection Using Statistical and Machine Learning Approaches."

A Project Report on,

“CREDIT CARD FRAUD DETECTION USING STATISTICAL AND MACHINE LEARNING TECHNIQUES”

An End-to-End Data Analytics Project Based on Real-World Credit Card Transaction Data (2019–2020)

Prepared By
Mr. Vaibhav Shrikant Kore
M.Sc. Statistics
Aspiring Data Analyst / Data Scientist

➢ Statistical Tools:

1) Exploratory Data Analysis (EDA)
Exploratory Data Analysis was performed to understand transaction behaviour and identify patterns related to fraudulent activities. The following statistical methods and visual tools were used:

- Descriptive Statistics (Mean, Median, Standard Deviation)
- Fraud vs Non-Fraud Count Analysis
- Distribution Analysis of Transaction Amount
- Box Plots for Outlier Detection
- Hour-wise, Day-wise and Category-wise Analysis
- Correlation Analysis using Heatmap

These techniques helped in understanding spending patterns, time-based behaviour, and differences between fraudulent and genuine transactions.

2) Time-Based Statistical Analysis
To analyse transaction frequency patterns, statistical analysis was performed on the inter-transaction time gap feature. The following methods were used:

• Descriptive Statistics (mean, median, standard deviation) to summarise transaction intervals.
• Box Plot Analysis to compare time gaps between fraudulent and genuine transactions.
• Mann–Whitney U Test to check whether the difference in time gaps is statistically significant.

This analysis helps in identifying unusually short transaction intervals that may indicate fraudulent behaviour.

3) Machine Learning Algorithms
To detect fraudulent credit card transactions, the following machine learning classification models were implemented and evaluated:

- Logistic Regression (with Class Weight Balancing)
- Logistic Regression using SMOTE
- Random Forest Classifier
- XGBoost Classifier

The models were evaluated using Confusion Matrix, Precision, Recall, F1-Score, and ROC-AUC to compare their fraud detection performance.

➢ Statistical Software:

1. **Python:** 
   Python was used as the primary tool for data preprocessing, feature engineering, exploratory data analysis, statistical analysis, and machine learning model development.

2. **MS-Excel:** 
   MS-Excel was used for basic data inspection and simple visualizations to quickly understand transaction summaries and fraud distribution.

3. **Jupyter Notebook:** 

   Jupyter Notebook was used as the development environment for executing Python code, visualizing results, and documenting the analysis step-by-step.

4. **Minitab:** 
   Minitab was used for conducting the normality test of the time-gap variable and for generating selected statistical graphs to support the time-based analysis.

Introduction

Credit card fraud is one of the major problems faced by the financial sector in today's digital world. Credit card fraud occurs when an unauthorized person uses someone's credit card details to make transactions without the cardholder's permission. Such fraudulent activities cause financial losses to both customers and banks and also reduce trust in digital payment systems.

With the rapid growth of online shopping, mobile banking, and digital payment platforms, the number of online transactions has increased significantly. As the use of credit cards has increased, opportunities for fraudulent activities have also grown. Many fraud cases occur because transactions are completed quickly and without physical verification, making it easier for fraudsters to misuse card information.

Fraud detection plays an important role in protecting customers and financial institutions. Early identification of fraudulent transactions helps banks prevent financial losses and protect customer accounts. Effective fraud detection systems also help in maintaining customer confidence in online payment systems and ensure the smooth functioning of digital financial services.

Traditional fraud detection methods are often not sufficient because fraud cases are very few compared to genuine transactions. This creates a problem known as class imbalance, where normal methods fail to correctly identify fraudulent activities. Therefore, advanced approaches are required to handle such challenges.

Statistical techniques help in understanding transaction behaviour by analysing patterns, variations, and differences between fraudulent and genuine transactions. Machine learning techniques further improve fraud detection by learning complex patterns from large datasets and making accurate predictions. By combining statistical analysis with machine learning approaches, it is possible to develop more effective and reliable credit card fraud detection systems.

Statement of the Problem

Credit card fraud detection is a challenging task because fraudulent transactions are very rare compared to genuine transactions. In most real-world credit card datasets, only a very small percentage of transactions are fraudulent, while the majority are normal. This situation is known as class imbalance, where the fraud class is much smaller than the non-fraud class.

Due to this imbalance, many traditional statistical and machine learning models fail to correctly identify fraudulent transactions. These models often perform well on genuine transactions but struggle to detect fraud cases, leading to low fraud detection rates. As a result, fraudulent activities may go unnoticed, causing financial losses to banks and customers.

Another major challenge in credit card fraud detection is the lack of proper analysis of transaction behaviour and time-based patterns. Many existing studies mainly focus on improving model accuracy without clearly analysing how transaction amount, time gaps, transaction frequency, and merchant behaviour contribute to fraud. Understanding these patterns is important for building effective and reliable fraud detection systems.

Therefore, there is a need for a study that focuses not only on detecting fraudulent transactions but also on analysing behavioural and time-based transaction patterns using statistical and machine learning approaches. Such an approach can help in improving fraud detection performance and provide better insights into the factors influencing credit card fraud.

Scope of the Study

The present study is limited to the analysis of credit card transaction data collected during the period 2019–2020. The research focuses on identifying fraudulent and non-fraudulent transactions using selected statistical techniques and machine learning approaches. The analysis is carried out in the context of credit card fraud detection by examining transaction amount, time-based patterns, customer behaviour, and merchant-related information. The study aims to evaluate the effectiveness of these methods in detecting fraudulent transactions. The scope of this research is restricted to the available dataset and does not include real-time fraud detection systems or direct implementation within banking environments.

Objectives of the Study

➢ To statistically examine time-based characteristics of credit card transactions, including transaction timing and inter-transaction gaps, in order to distinguish fraudulent behaviour from genuine transaction patterns.

➢ To analyse the association between merchant categories and transaction modes with the occurrence of fraudulent credit card transactions using exploratory and descriptive analysis.

➢ To construct and evaluate selected machine learning classification models for credit card fraud detection and assess their performance using appropriate evaluation metrics.

About Data

Description of Dataset:

a) Source of Dataset: The dataset used for this study has been obtained from the Kaggle website.

b) Name of the Dataset: Credit Card Transactions Fraud Detection Dataset (2019–2020)

c) Size of the Dataset

- **Number of observations (rows):** 1,048,575
- **Number of variables (columns):** 24

(Dataset Shape: 1048575 × 24)

d) Data Source: https://www.kaggle.com/datasets/kartik2112/fraud-detection

e) Data Format: The dataset is available in CSV (Comma-Separated Values) format. Each row represents an individual credit card transaction, and the header row contains the names of the variables.

Description of Variables:

Variable Name	Description
trans_date_trans_time	Date and time when the credit card transaction occurred
cc_num	Encrypted credit card number of the customer
merchant	Name of the merchant where the transaction was made
category	Category of the merchant (such as grocery, fuel, entertainment, etc.)
amt	Amount involved in the credit card transaction
gender	Gender of the cardholder
city	City in which the transaction took place
state	State where the transaction occurred
zip	ZIP code of the transaction location
lat	Latitude of the cardholder's location
long	Longitude of the cardholder's location

city_pop	Population of the city where the cardholder resides
job	Occupation of the cardholder
dob	Date of birth of the cardholder
unix_time	Transaction time represented in Unix timestamp format
merch_lat	Latitude of the merchant's location
merch_long	Longitude of the merchant's location
is_fraud	Fraud indicator variable (1 = Fraudulent, 0 = Genuine)
hour	Hour at which the transaction occurred
day	Day of the month on which the transaction was made
month	Month in which the transaction occurred
weekday	Day of the week of the transaction
time_gap	Time gap between consecutive transactions of the same card
age	Age of the cardholder calculated from date of birth

Dataset Relevance to the Study:

This dataset provides detailed transactional, temporal, geographical, and behavioural information required for analysing credit card fraud patterns. The presence of time-based variables, transaction amount, merchant details, and customer attributes makes the dataset suitable for applying statistical analysis and machine learning techniques to distinguish between fraudulent and genuine transactions.

“Exploratory Data Analysis”

Project Report on:

“A Study of Credit Card Fraud Detection Using Statistical and Machine Learning Approaches.”

➢Descriptive Statistics:

Statistic	cc_num	amt	zip	lat	long	city_pop	unix_time	merch_lat
Count	1,048,575	1,048,575	1,048,575	1,048,575	1,048,575	1,048,575	1,048,575	1,048,575
Mean	4.17E+17	70.28	48,801.59	38.53	-90.23	89,057.76	1.34E+09	38.53
Std Dev	1.31E+18	159.95	26,898.04	5.08	13.76	302,435.10	1.02E+07	5.11
Min	6.04E+10	1.00	1,257	20.03	-165.67	23	1.33E+09	19.03
25%	1.80E+14	9.64	26,237	34.62	-96.80	743	1.34E+09	34.73
50% (Median)	3.52E+15	47.45	48,174	39.35	-87.48	2,456	1.34E+09	39.36
75%	4.64E+15	83.05	72,042	41.94	-80.16	20,328	1.35E+09	41.96
Max	4.99E+18	28,948.90	99,783	66.69	-67.95	2,906,700	1.36E+09	67.51

Statistic	merch_long	is_fraud	hour	day	month	weekday	time_gap	age
Count	1,048,575	1,048,575	1,048,575	1,048,575	1,048,575	1,048,575	1,048,575	1,048,575
Mean	-90.23	0.0057	12.80	15.53	6.51	3.14	187.13	45.89
Std Dev	13.77	0.075	6.82	8.90	3.67	2.20	420.86	17.37
Min	-166.67	0	0	1	1	0	0	14
25%	-96.90	0	7	8	3	1	15	32
50% (Median)	-87.44	0	14	15	7	3	52	44
75%	-80.23	0	19	23	10	5	176	57
Max	-66.95	1	23	31	12	6	20,095	96

Interpretations:

a) The dataset contains 1,048,575 credit card transactions, indicating a large and reliable sample for fraud analysis.

b) The average transaction amount is around ₹70**,** while the maximum amount is very high, showing the presence of extreme values and outliers**,** which are important in fraud detection.

c) Most transactions are non-fraudulent**,** as indicated by the very low mean value of the is_fraud variable, confirming a highly imbalanced dataset.

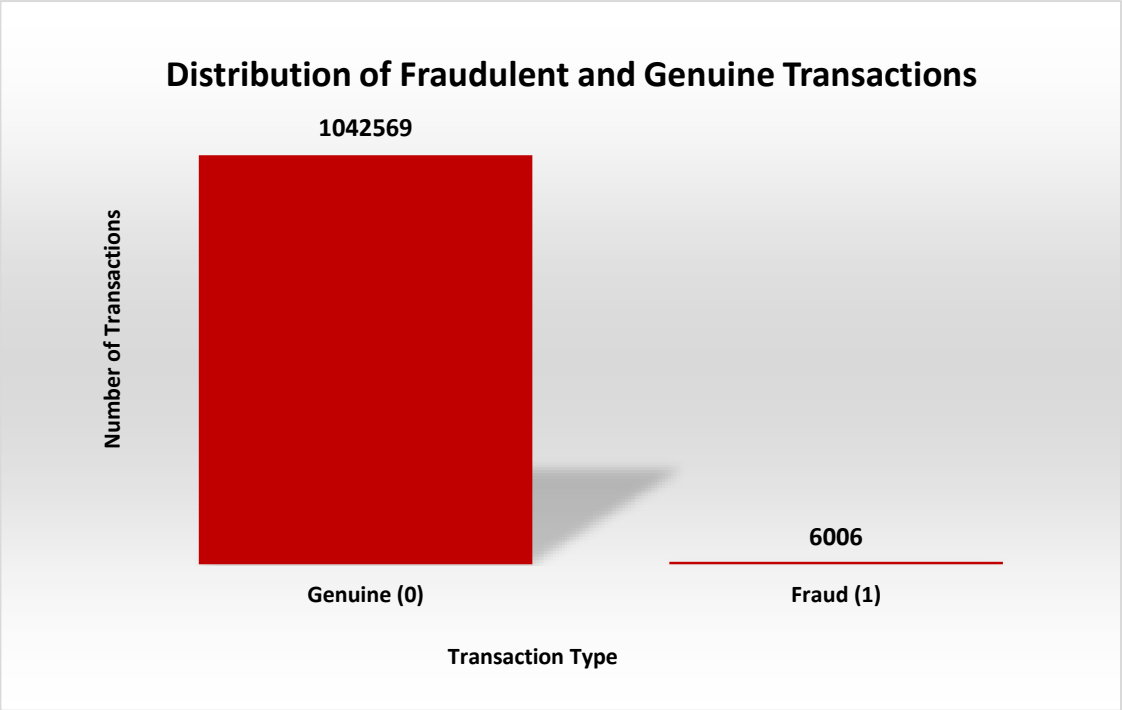
d) Transactions occur across all hours, days, and months**,** with an average transaction time around midday**,** suggesting continuous customer activity.

e) The average time gap between transactions is relatively high, but very small minimum values indicate rapid successive transactions**,** which can be suspicious.

f) The average customer age is about 46 years**,** with a wide age range, indicating fraud can occur across different age groups.

g) Geographic variables (latitude and longitude) show wide dispersion, reflecting transactions across multiple locations**,** which is useful for location-based fraud analysis.

➢Target Variable Distribution:



Interpretations:

The column chart clearly shows that the number of genuine transactions is extremely high compared to fraudulent transactions. Fraud cases are very few in number, indicating that fraudulent activities occur rarely in comparison to normal credit card usage.

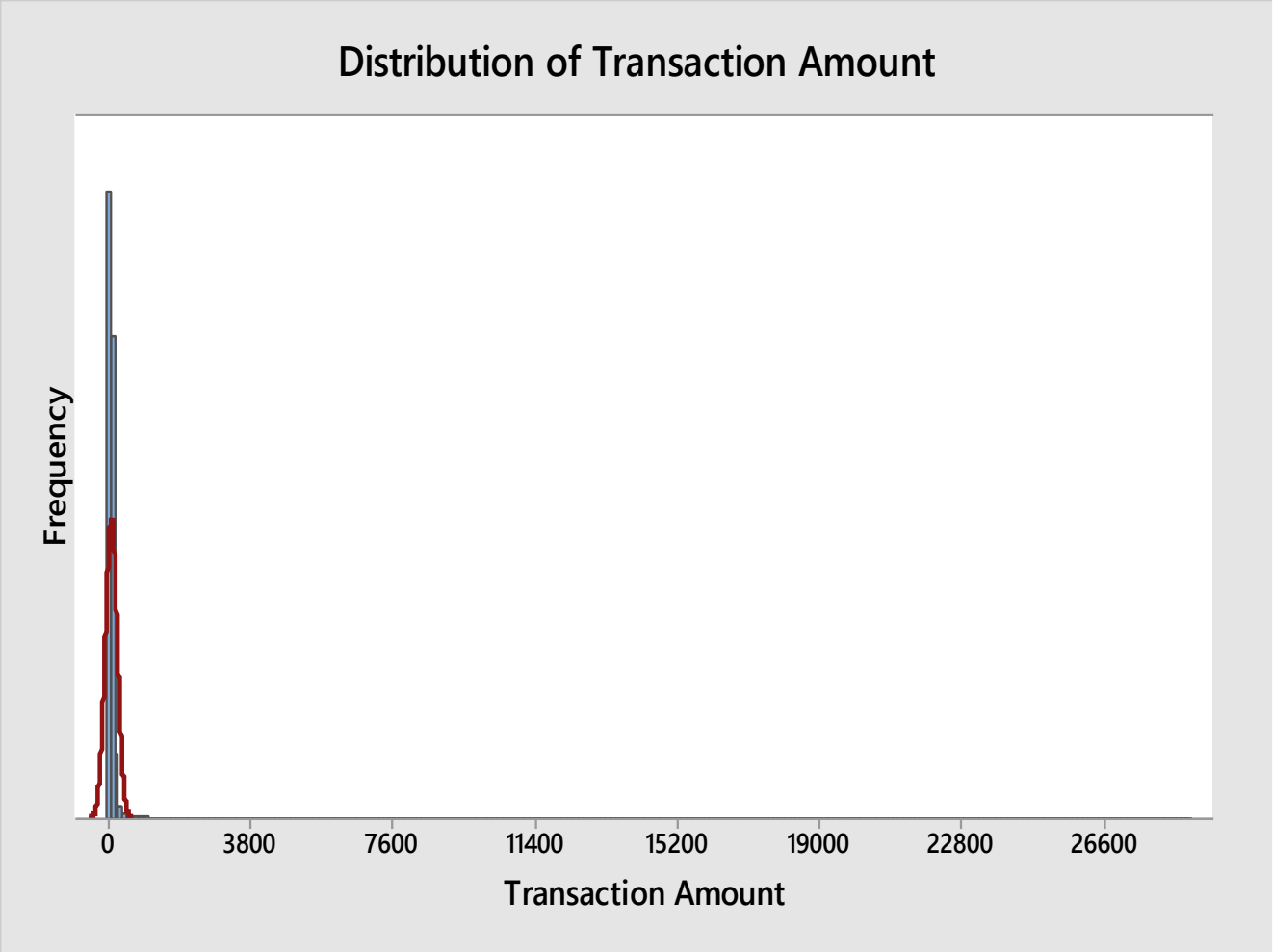
➢Descriptive Statistics of Transaction Amount by Fraud Status:

Fraud Status	Count	Mean Amount	Std. Dev.	Min	25%	Median (50%)	75%	Max
Non-Fraud (0)	1,042,569	67.63	153.70	1.00	9.60	47.22	82.47	28,948.90
Fraud (1)	6,006	530.57	391.33	1.18	241.58	391.17	901.95	1,371.81

Interpretations:

The table shows a clear difference in transaction amounts between fraudulent and genuine transactions. Fraudulent transactions have a much higher average and median amount compared to genuine transactions. This indicates that fraud cases generally involve larger transaction values than normal transactions.

➢**To study the distribution of transaction amounts using histogram:**



Interpretations:

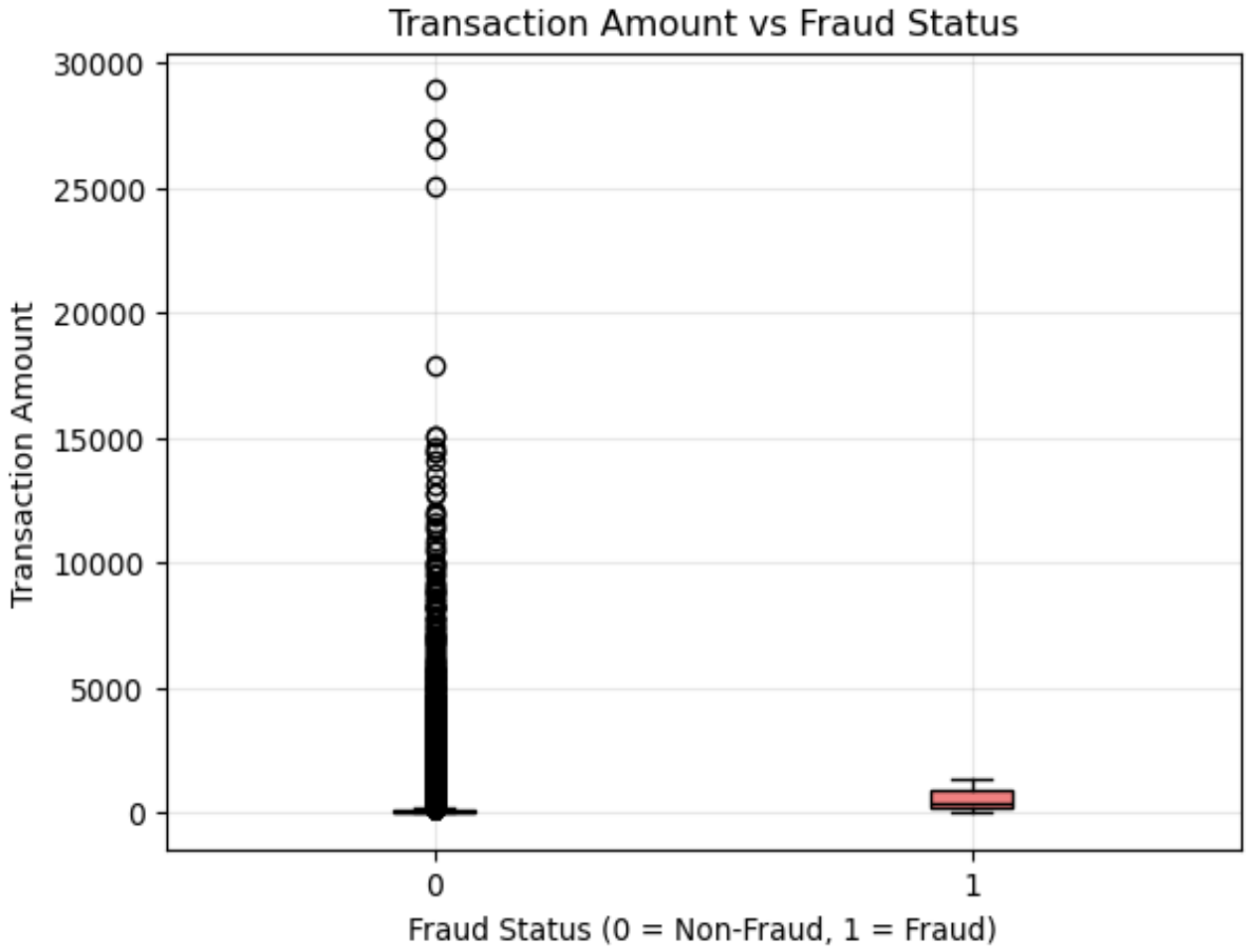
a) The histogram shows that most credit card transactions are of low value, indicating regular and routine spending behaviour.

b) The distribution is positively skewed, with a small number of transactions having very high amounts.

c) These high-value transactions occur less frequently and may represent unusual or suspicious activity.

d) Hence, transaction amount is an important variable for identifying potential fraud.

➢To compare transaction amounts for fraudulent and non-fraudulent transactions using box plot:



Interpretations:

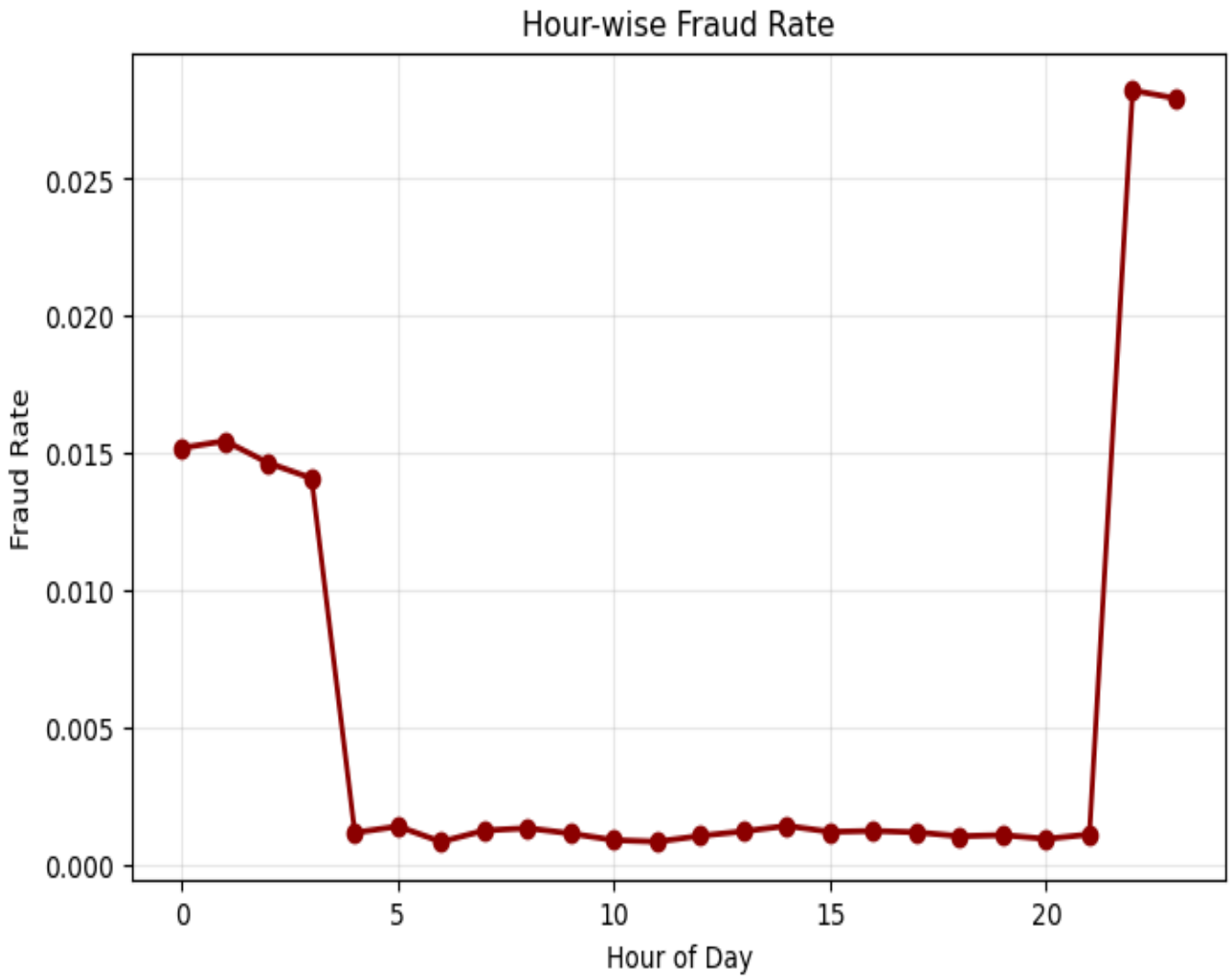
a) The box plot clearly shows a difference between fraudulent and non-fraudulent transactions.

b) Non-fraud transactions are mostly concentrated at lower transaction amounts with fewer extreme values.

c) Fraudulent transactions tend to involve higher amounts and show several high-value outliers.

d) This indicates that unusually large transaction amounts are more likely to be associated with fraud.

➢To study the hour-wise pattern of fraudulent transactions:



Interpretations:

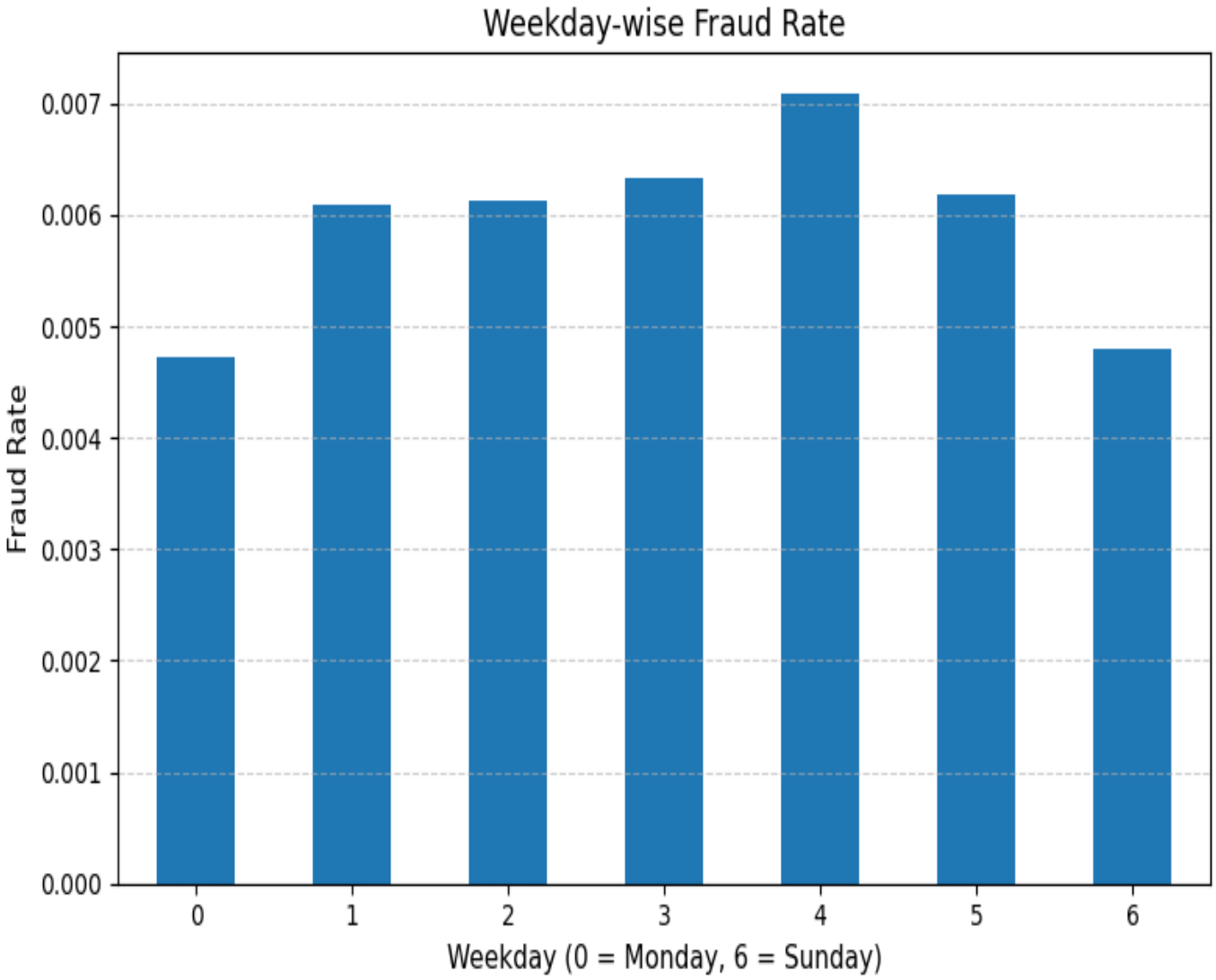
a) The fraud rate is not uniform throughout the day and varies across different hours.

b) Higher fraud activity is observed during late night and early morning hours, when transaction monitoring is generally lower.

c) During daytime and evening hours, the fraud rate remains comparatively low and stable.

d) This pattern indicates that transaction time plays an important role in identifying suspicious activities.

➢**To study the weekday-wise pattern of fraudulent transactions:**



Interpretations:

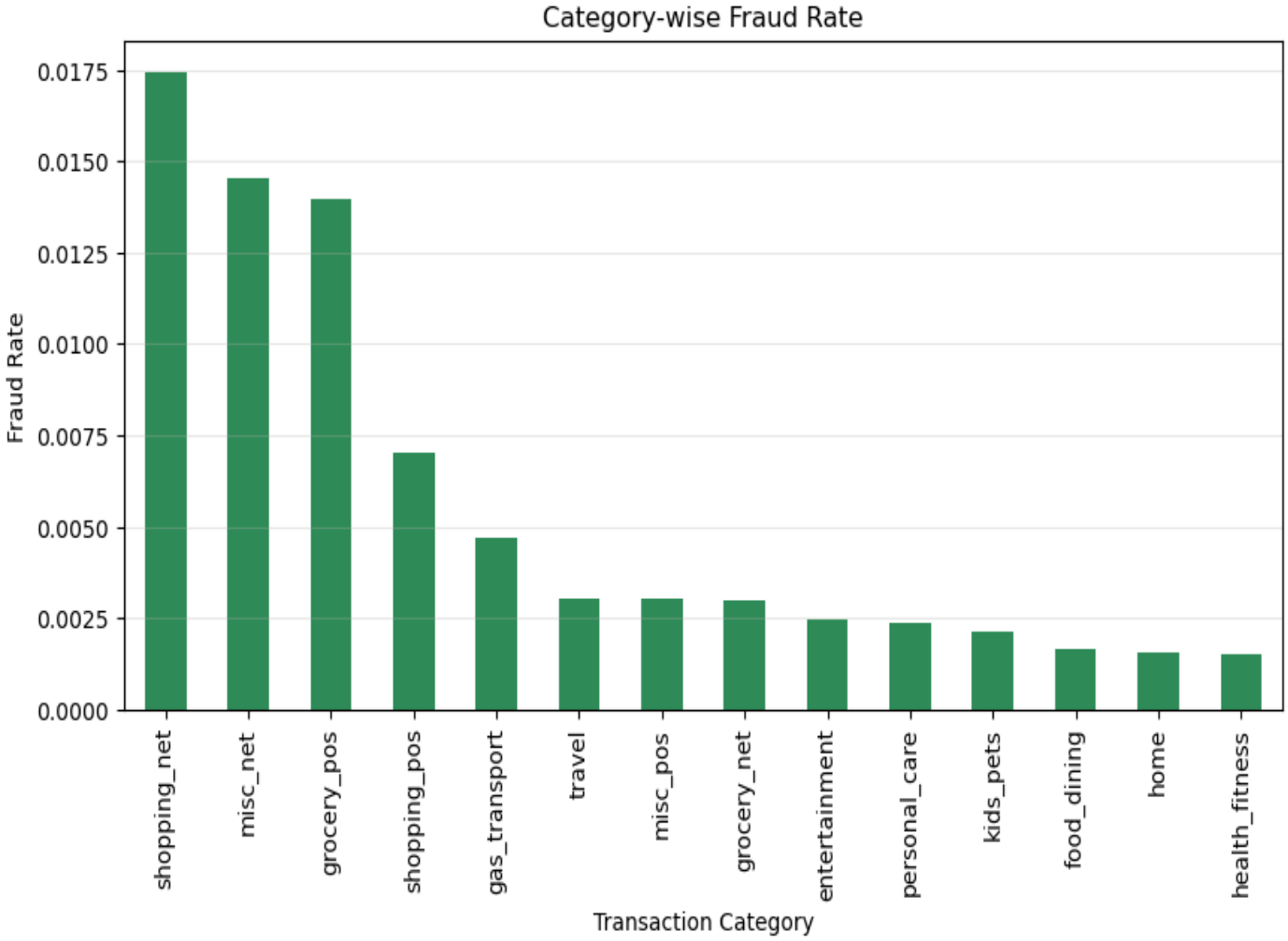
a) Fraudulent transactions occur across all days of the week, but the rate is not exactly the same.

b) A slightly higher fraud rate is observed during weekends, especially on Sunday.

c) Weekdays show a more stable and lower fraud pattern compared to weekends.

d) This suggests that fraudsters may take advantage of reduced banking activity during weekends.

➢To study the category-wise distribution of fraudulent transactions:



Interpretations:

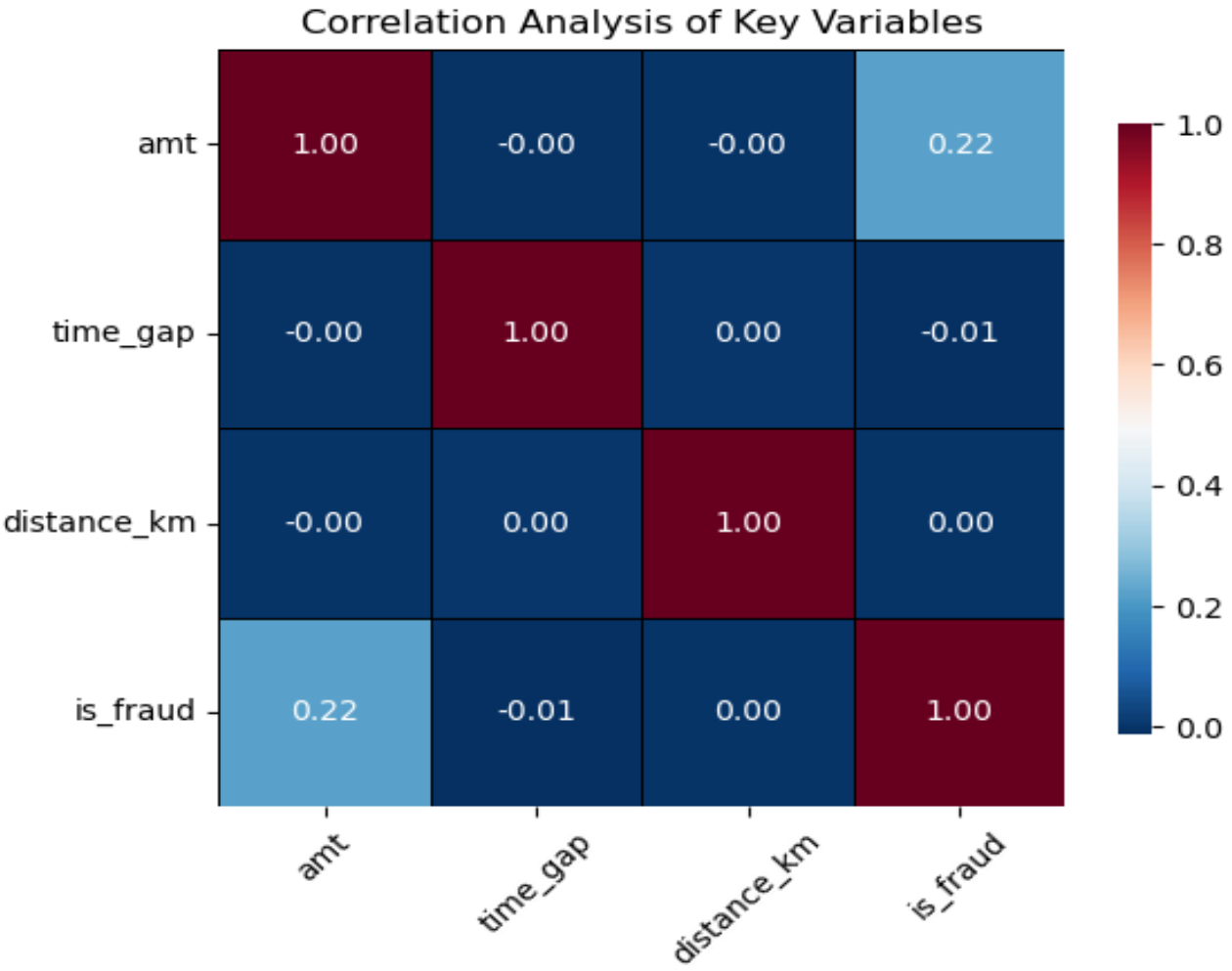
a) Fraud rate differs significantly across different transaction categories.

b) Certain categories show a higher concentration of fraudulent transactions, indicating higher risk.

c) Categories related to online and non-essential purchases tend to have higher fraud rates.

d) This analysis highlights that merchant category is an important factor in fraud detection.

➢To study the relationship between transaction variables and fraud using correlation heatmap:



Interpretations:

a) The heatmap shows that transaction amount has a positive association with fraud, indicating that higher-value transactions are more likely to be fraudulent.

b) A weak negative relationship is observed between time gap and fraud, suggesting that fraudulent transactions often occur within shorter time intervals.

c) Transaction distance shows negligible correlation with fraud, indicating that distance alone is not a strong factor in fraud detection.

d) Overall, the results indicate that fraud is influenced by a combination of factors rather than a single variable.

“Time Based Statistical Analysis”

Project Report on:

“A Study of Credit Card Fraud Detection Using Statistical and Machine Learning Approaches.”

➢Descriptive Statistics of Time Gap by Fraud Status:

Fraud Status	Count	Mean Amount	Std. Deviation	Min	25%	Median (50%)	75%	Max
Non-Fraud (0)	1,042,569	187.48	421.41	0.0	15.0	52.0	176.0	20095.0
Fraud (1)	6,006	126.90	305.12	0.0	10.0	29.0	86.0	8388.0

Interpretations:

a) The average time gap between transactions is lower for fraudulent transactions compared to genuine ones. This indicates that fraud transactions tend to occur in quicker succession.

b) The median (50%) time gap for fraud cases is much smaller than for non-fraud cases, showing that in typical fraud behaviour, transactions happen within short intervals.

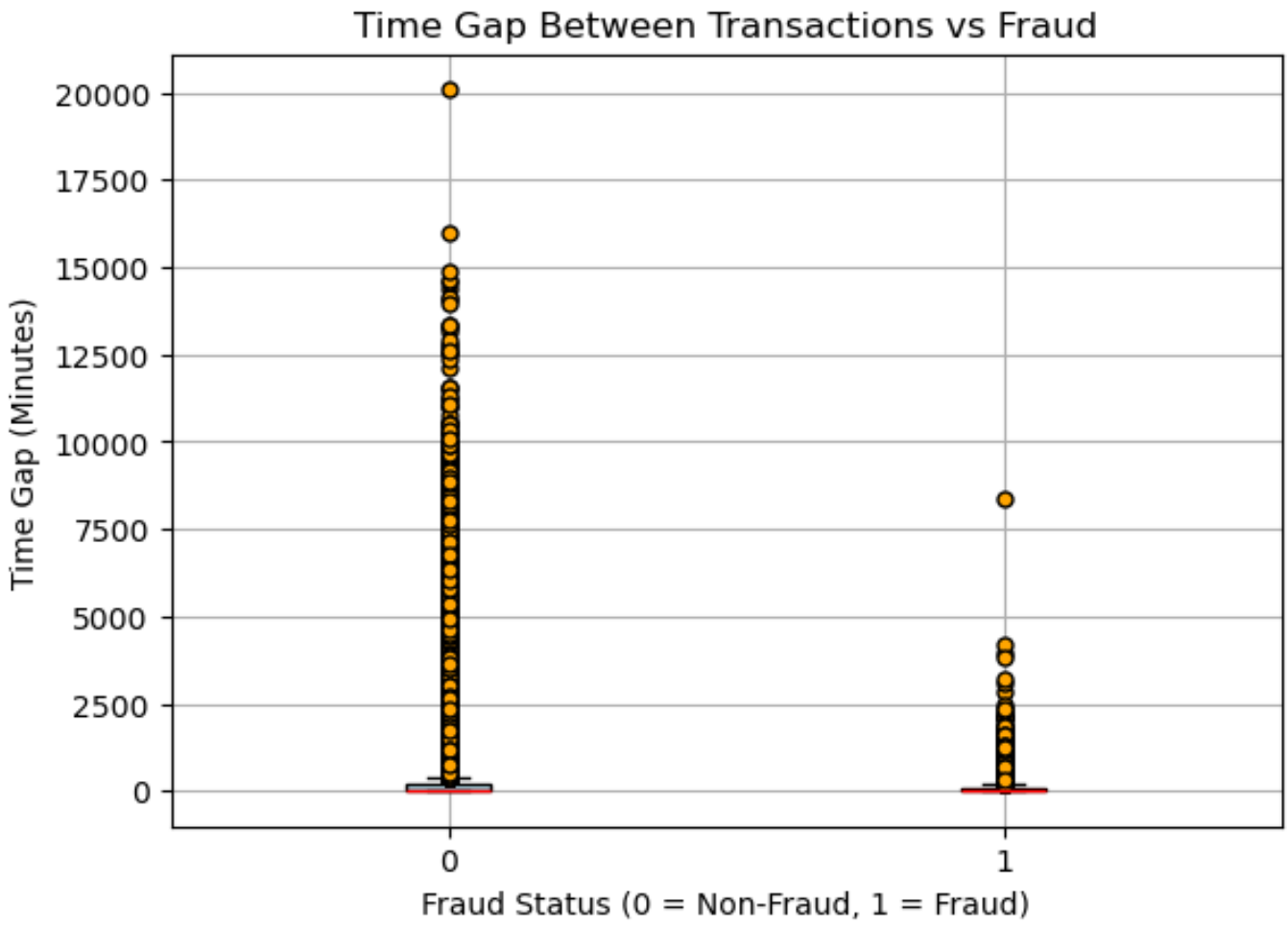
c) The spread (standard deviation**)** of time gaps is higher for non-fraud transactions, meaning genuine customers have more varied and irregular spending intervals.

d) Fraud transactions show shorter lower quartile (25%) values**,** suggesting that many fraud events happen almost immediately after a previous transaction.

e) Although some fraud cases have large gaps, the overall pattern shows fraud is associated with rapid, closely spaced transactions**,** which is a strong behavioural fraud indicator.

These results confirm that time gap between transactions is an important feature for identifying suspicious activity**.**

➢To compare time gaps between consecutive transactions for fraudulent and genuine cases:



Interpretations:

a) Fraudulent transactions tend to happen in quick succession, showing shorter time gaps between transactions.

b) Non-fraudulent transactions are more spread out over time, indicating regular customer behaviour.

c) This suggests that time gap is a key indicator for identifying suspicious or fraudulent activity.

d) Transactions with unusually short gaps can be flagged for further investigation.

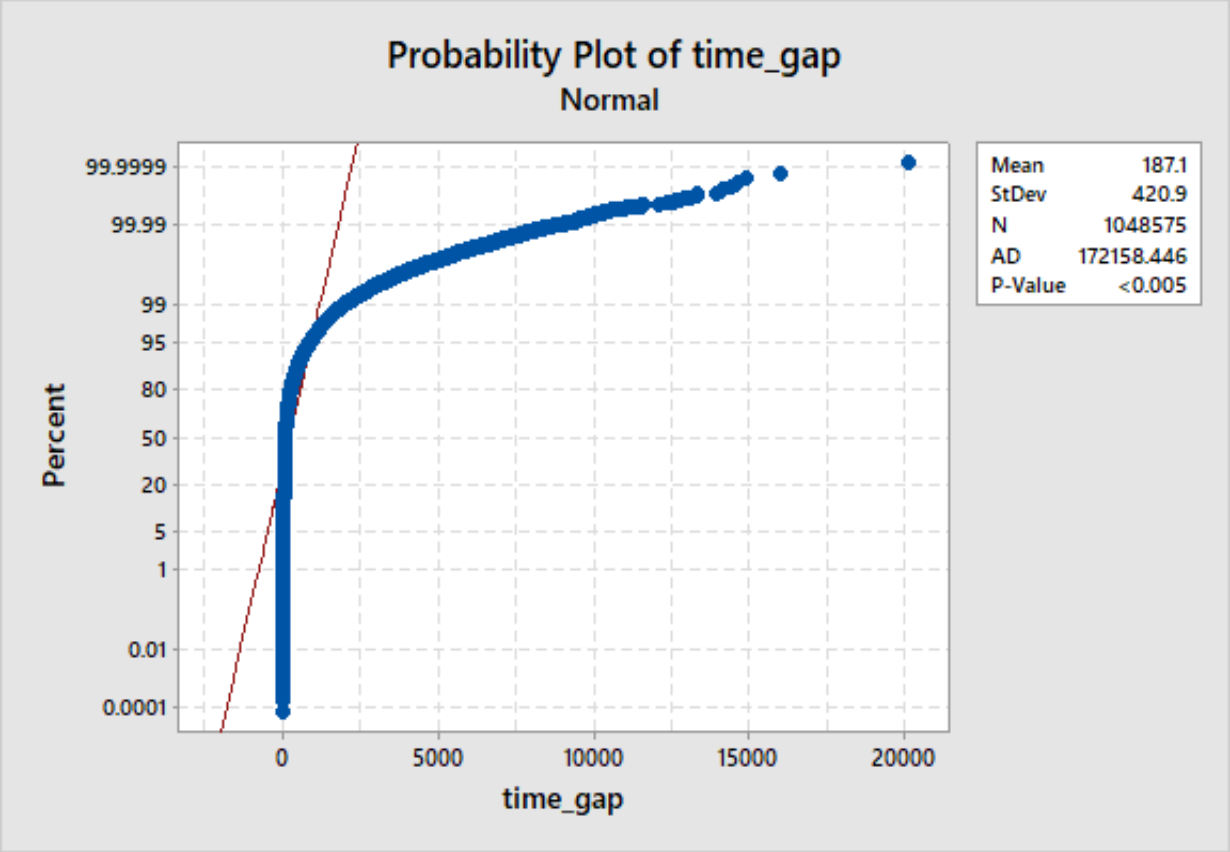
➢To check whether the time gap variable follows a normal distribution:

Hypothesis: -

Ho: - The time gap data is normally distributed.

v/s

H1: - The time gap data is not normally distributed.



Interpretations:

Since the P-value < level of significance (0.05),
we reject the null hypothesis and conclude that the time gap data is not normally distributed.

Since the data is not normally distributed, a non-parametric test is more appropriate for further statistical analysis.

➢Mann–Whitney U Test: - To test whether the time gap between transactions differs for fraudulent and non-fraudulent transactions:

Hypothesis: -

Ho: - There is no significant difference in the time gap between fraudulent and genuine transactions

v/s

H1: - There is a significant difference in the time gap between fraudulent and genuine transactions.

Level of Significance**: -**

The test was conducted at a 5% level of significance ($\alpha = 0.05$).

Output: -

```
Mann-Whitney U Statistic: 2597440318.0
p-value: 4.172204050961408e-115
```

Decision Rule

- If **$p < 0.05$** → Reject H_0
- If **$p \geq 0.05$** → Fail to reject H_0

Interpretations:

Since the p-value < level of significance (0.05)**,**
we reject the null hypothesis and conclude that there is a statistically significant difference in the time gap between fraudulent and genuine transactions**.**

This indicates that fraudulent transactions tend to occur with different transaction timing patterns compared to normal transactions. In particular, fraud transactions are more likely to happen with shorter time gaps**,** suggesting rapid or suspicious transaction behaviour.

Therefore, time gap between transactions is an important indicator for identifying potential credit card fraud.

“Machine Learning Algorithms”

Project Report on:

“A Study of Credit Card Fraud Detection Using Statistical and Machine Learning Approaches.”

Data Mining Classifiers

In this section, modelling is performed to predict fraudulent credit card transactions using Python software. The dataset is divided into training and testing data so that the model is trained on the training dataset and its performance is evaluated on the testing dataset. Here, the target variable is is_fraud. Different classification algorithms are applied, and their performance is measured.

➢Logistic Regression:

Logistic Regression is a supervised machine learning algorithm used for classification problems. It is mainly used when the target variable is binary. In this study, Logistic Regression is used to classify credit card transactions as fraudulent or genuine**.** Since the response variable is categorical (fraud or non-fraud), Logistic Regression estimates the probability of a transaction being fraudulent using a logistic function. This model is simple, interpretable, and helps in understanding how different transaction features influence the likelihood of fraud.

➢Random Forest Classifier:

Random Forest is an ensemble machine learning algorithm used for classification and regression tasks. It works by building multiple decision trees during training and combining their outputs to make the final prediction. In this study, Random Forest is used to classify credit card transactions as fraudulent or genuine. This model helps in improving prediction accuracy and reducing overfitting by averaging the results of many trees. It is capable of handling large datasets and capturing complex patterns in transaction behaviour, making it effective for fraud detection.

➢XGBoost Classifier:

XGBoost (Extreme Gradient Boosting) is an advanced ensemble machine learning algorithm based on gradient boosting decision trees. It builds models sequentially, where each new tree corrects the errors made by the previous ones. In this study, XGBoost is used to classify credit card transactions as fraudulent or genuine with high accuracy. The algorithm includes regularization techniques that help prevent overfitting and improve model generalization. It is efficient, handles large datasets well, and is especially powerful in detecting complex and non-linear fraud patterns in transaction data, making it highly suitable for credit card fraud detection tasks.

• Confusion Matrix

A Confusion Matrix is a performance evaluation tool used for classification models. It compares the actual values with the predicted values and summarizes the results in a tabular form.

	Predicted Negative	**Predicted Positive**
Actual Negative	True Negative (TN)	False Positive (FP)
Actual Positive	False Negative (FN)	True Positive (TP)

- **True Positive (TP):** Fraud transaction correctly predicted as fraud
- **True Negative (TN):** Genuine transaction correctly predicted as genuine
- **False Positive (FP):** Genuine transaction wrongly predicted as fraud
- **False Negative (FN):** Fraud transaction wrongly predicted as genuine

The confusion matrix helps in understanding the types of errors made by the model.

• Measures of Performance Evaluation

To evaluate the effectiveness of the classification models, the following performance metrics are used:

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

Accuracy represents the overall percentage of correctly classified transactions.

ii) Recall (Sensitivity or True Positive Rate): - TP/(TP+FN)

Recall measures the proportion of actual fraudulent transactions that are correctly identified.
It is very important in fraud detection because missing a fraud case (FN) is costly.

iii) Precision: - TP/(TP+FP)

Precision indicates the proportion of transactions predicted as fraud that are actually fraudulent.
High precision means fewer genuine transactions are incorrectly flagged as fraud.

iv) F1-Score: - 2TP/(2TP+FP+FN)

The F1-score is the harmonic mean of Precision and Recall.
It provides a balance between detecting fraud correctly and avoiding false alarms.

The methodology and performance of every model is as follows:

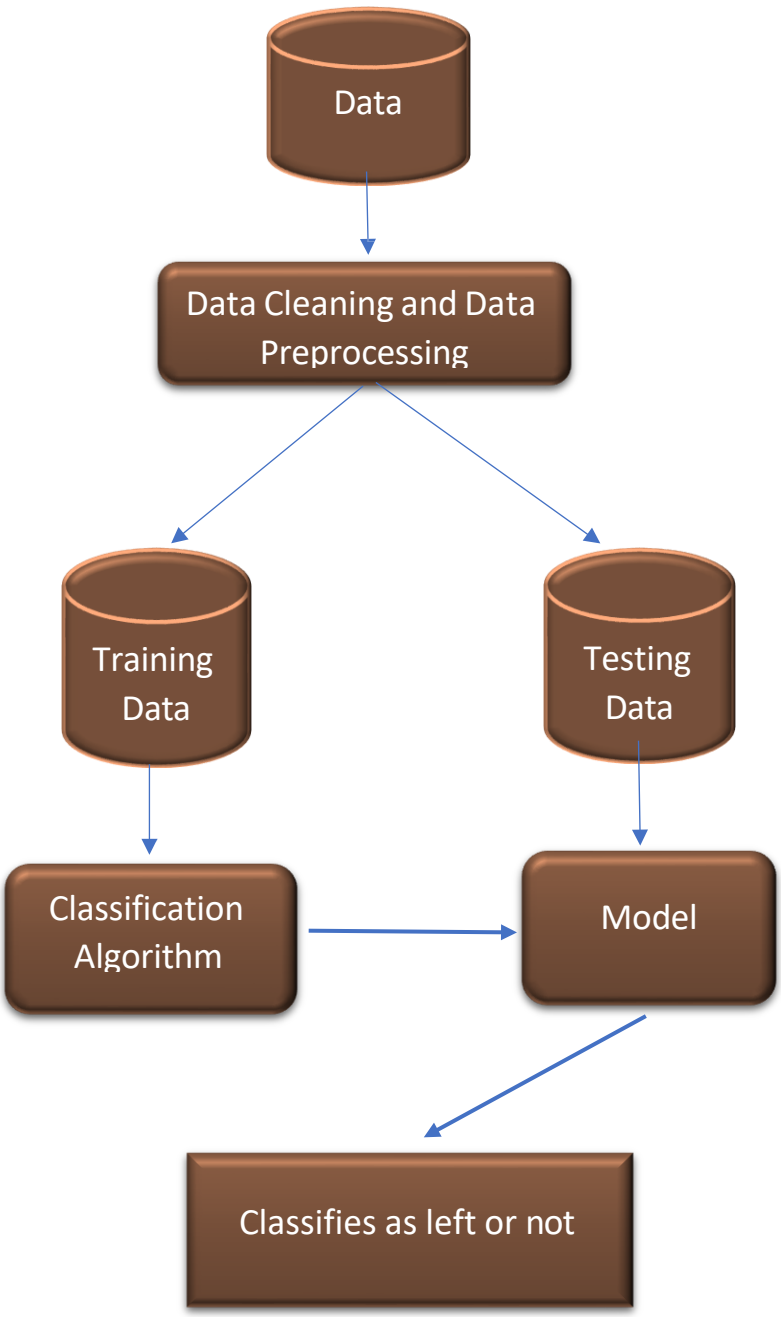


Fig: Workflow Diagram

➢Logistic Regression (Class Weight = Balanced):

• Confusion Matrix:

n = 209,715	**Predicted: Genuine (0)**	**Predicted: Fraud (1)**
Actual: 0	197,544	10,970
Actual: 1	293	908

• Classification Report:

Class	**Precision**	**Recall**	**F1-score**	**Support**	**Accuracy**
Genuine (0)	1	0.95	0.97	208,514	**0.95**
Fraud (1)	0.08	0.76	0.14	1,201	
Macro Avg	0.54	0.85	0.56	209,715	
Weighted Avg	0.99	0.95	0.97	209,715	

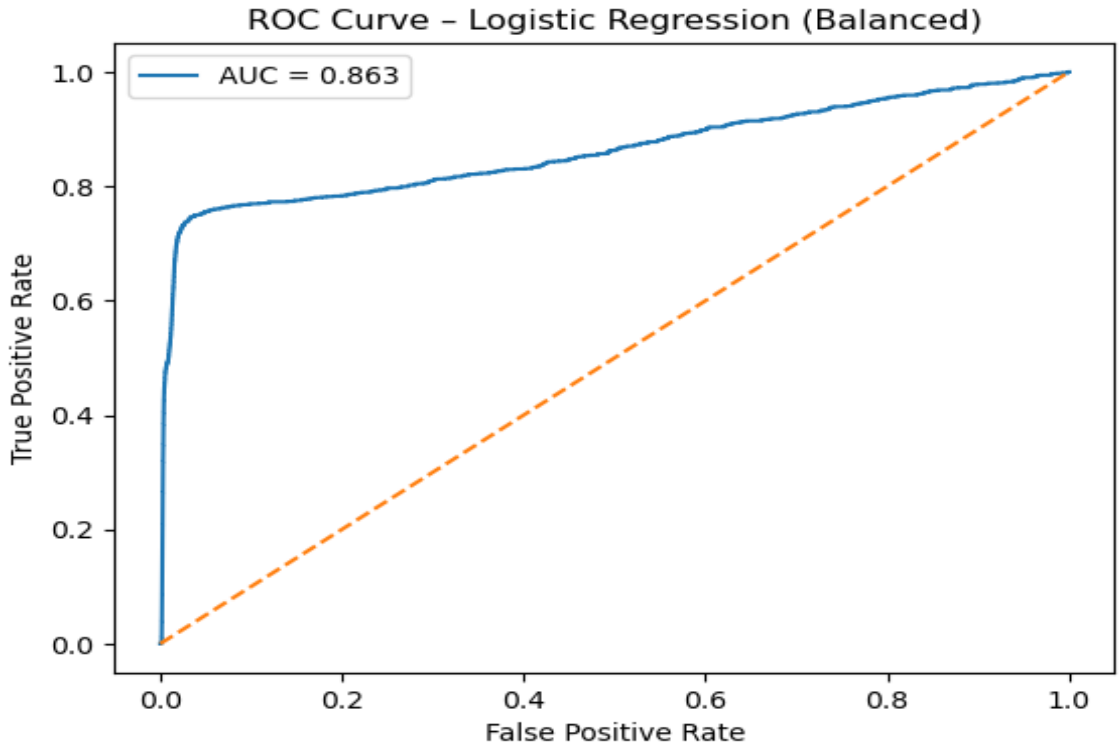
• Receiver Operating Characteristic (ROC) Curve & Cross-Validation Performance:

ROC AUC = 0.852

Stratified 5-Fold Cross-Validation ROC AUC Scores: 0.863, 0.862, 0.856, 0.868, 0.861

Mean CV ROC AUC = 0.862

• Receiver Operating Characteristic Curve:



Interpretations:

1.The Logistic Regression model correctly classifies most transactions, showing good overall performance on the test dataset, with an overall accuracy of 94%**.**

2.The model achieves high recall for fraudulent transactions, meaning it detects most fraud cases. This is very important in fraud detection, where catching fraud matters more than avoiding a few false alarms.

3.Some genuine transactions are predicted as fraud, but this trade-off is acceptable because missing a fraud case is more costly than investigating an extra alert.

4.The ROC AUC value of 0.852 (85.2%) shows that the model has good ability to distinguish between fraudulent and genuine transactions.

5.The ROC curve AUC of 0.863 (86.3%) and the mean cross-validation AUC of 0.862 (86.2%) indicate that the model performs consistently across different data splits and generalizes well.

➢Logistic Regression (SMOTE):

• Confusion Matrix:

n = 209,715	**Predicted: 0**	**Predicted: 1**
Actual: 0	197,241	11,273
Actual: 1	294	907

• Classification Report:

Class	Precision	Recall	F1-score	Support	Accuracy
Genuine (0)	1	0.95	0.97	208,514	0.94
Fraud (1)	0.07	0.76	0.14	1,201	
Macro Avg	0.54	0.85	0.55	209,715	
Weighted Avg	0.99	0.94	0.97	209,715	

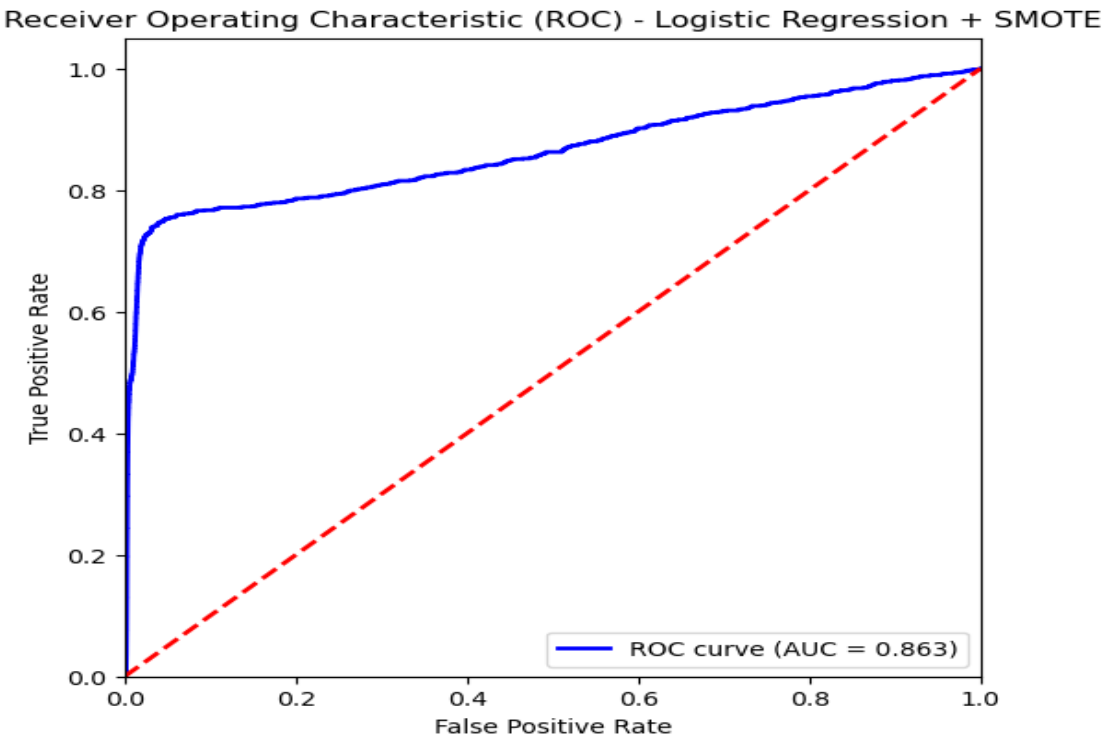
• Receiver Operating Characteristic (ROC) Curve & Cross-Validation Performance:

ROC AUC = 0.851

Stratified 5-Fold Cross-Validation ROC AUC Scores: 0.870, 0.868, 0.868, 0.869, 0.868

Mean CV ROC AUC = 0.869

• Receiver Operating Characteristic Curve:



Interpretations:

1.The Logistic Regression model trained using SMOTE correctly classifies most transactions and shows improved performance in learning fraud patterns from the imbalanced dataset, with an overall accuracy of about 94%**.**

2.The model achieves high recall for fraudulent transactions, meaning it detects most fraud cases. This improvement is mainly due to SMOTE, which balances the training data by increasing fraud samples.

3.Some genuine transactions are predicted as fraud, but this trade-off is acceptable because missing a fraud case is more costly than investigating a few extra alerts.

4.The ROC AUC value of 0.851 (85.1%) shows that the model has good ability to distinguish between fraudulent and genuine transactions.

5.The ROC curve AUC of 0.863 (86.3%) and the mean cross-validation AUC of 0.869 (86.9%) indicate that the model performs consistently across different data splits and generalizes well.

➢ Random Forest Classifier:

• Confusion Matrix:

n = 209,715	**Predicted: 0**	**Predicted: 1**
Actual: 0	208,489	25
Actual: 1	305	896

• Classification Report:

Class	Precision	Recall	F1-score	Support	Accuracy
Genuine (0)	1	1	1	208,514	1
Fraud (1)	0.97	0.75	0.84	1,201	
Macro Avg	0.99	0.87	0.92	209,715	
Weighted Avg	1	1	1	209,715	

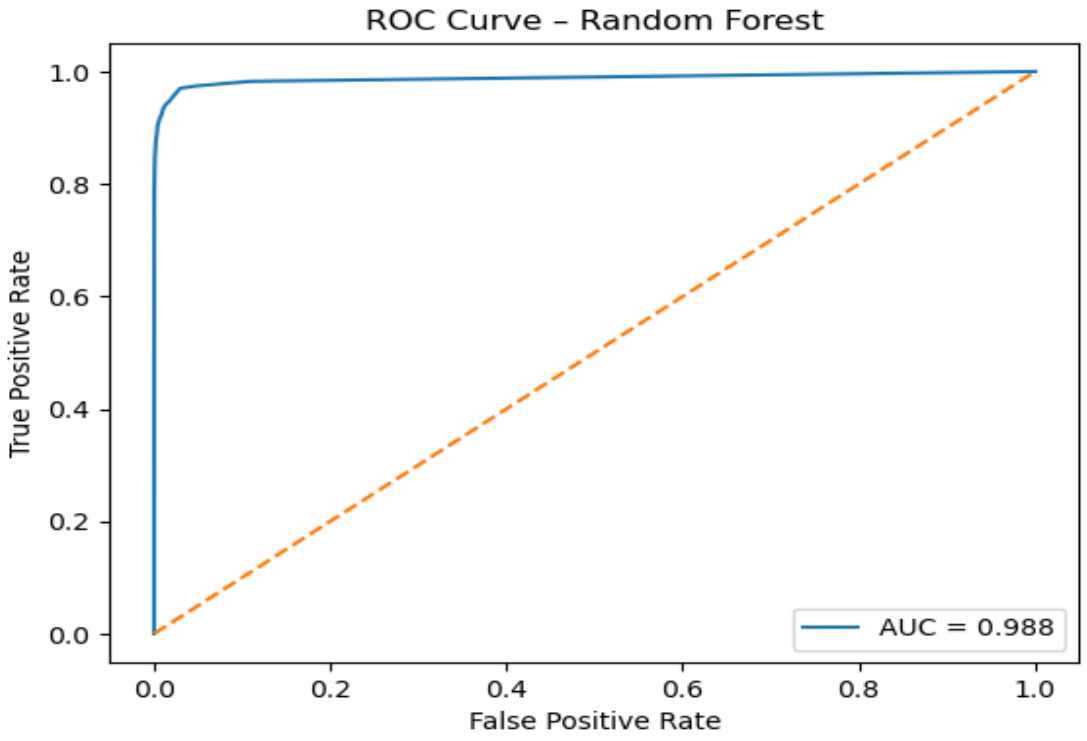
• Receiver Operating Characteristic (ROC) Curve & Cross-Validation Performance:

ROC AUC = 0.873

Stratified 5-Fold Cross-Validation ROC AUC Scores: 0.993, 0.993, 0.993, 0.991, 0.992

Mean CV ROC AUC = 0.992

• Receiver Operating Characteristic Curve:



Interpretations:

1.The Random Forest model correctly classifies almost all genuine transactions and detects a large proportion of fraudulent transactions, with an overall accuracy of about 99%**.**

2.The model achieves high precision **(**0.97**)** for fraud detection, meaning most transactions predicted as fraud are actually fraudulent.

3.The recall for fraud cases is 0.75 (75%)**,** indicating that the model successfully identifies most fraudulent transactions, although a small number of fraud cases remain undetected.

4.The ROC AUC value of 0.873 (87.3%) shows that the model has strong ability to distinguish between fraudulent and genuine transactions.

5.The ROC curve AUC of 0.988 (98.8%) and the mean cross-validation AUC of 0.992 (99.2%) indicate that the model performs very consistently across different data splits and demonstrates excellent generalization capability.

➢ XGBoost Classifier:

• Confusion Matrix:

n = 209,715	**Predicted: 0**	**Predicted: 1**
Actual: 0	208,436	78
Actual: 1	221	980

• Classification Report:

Class	Precision	Recall	F1-score	Support	Accuracy
Genuine (0)	1	1	1	208,514	0.999 (~1.00)
Fraud (1)	0.93	0.82	0.87	1,201	
Macro Avg	0.96	0.91	0.93	209,715	
Weighted Avg	1	1	1	209,715	

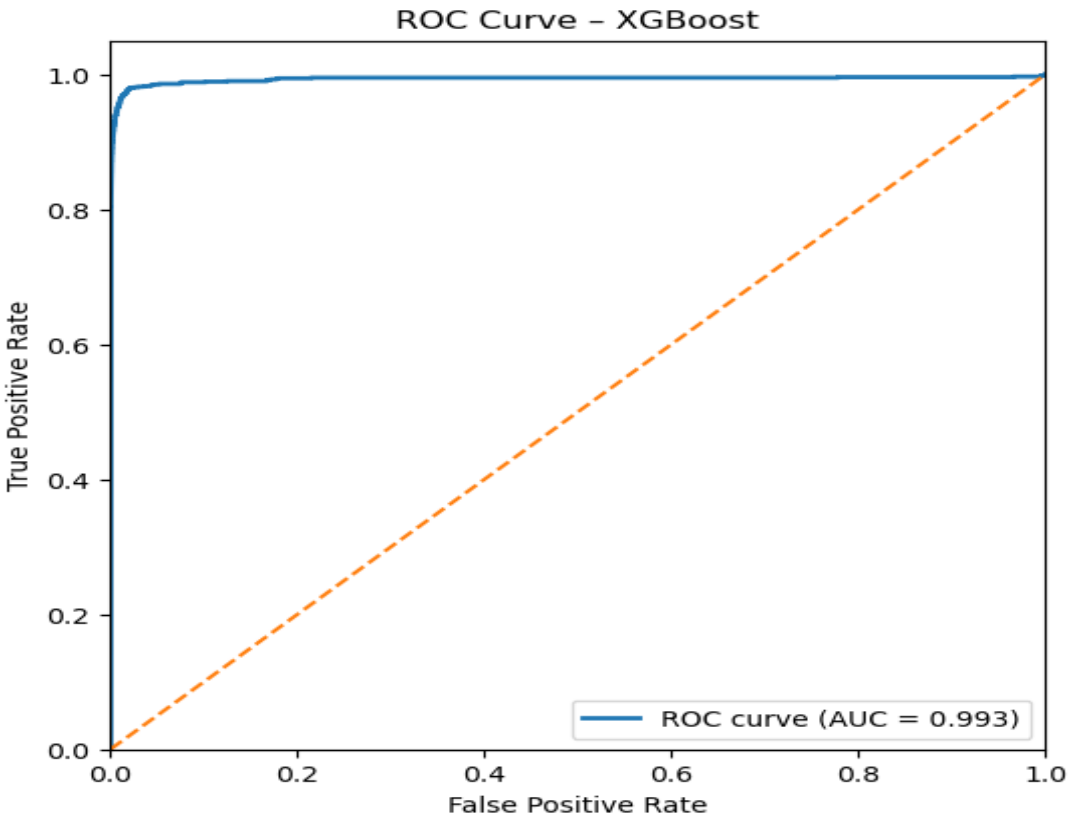
• Receiver Operating Characteristic (ROC) Curve & Cross-Validation Performance:

ROC AUC = 0.908

Stratified 5-Fold Cross-Validation ROC AUC Scores: 0.984, 0.983, 0.989, 0.993, 0.978

Mean CV ROC AUC = 0.985

• Receiver Operating Characteristic Curve:



Interpretations:

1.The XGBoost model correctly classifies almost all transactions and delivers outstanding fraud detection performance, with an overall accuracy of about 99%**.**

2.It achieves very high precision and strong recall for fraudulent transactions, meaning most fraud cases are detected while keeping false alarms low.

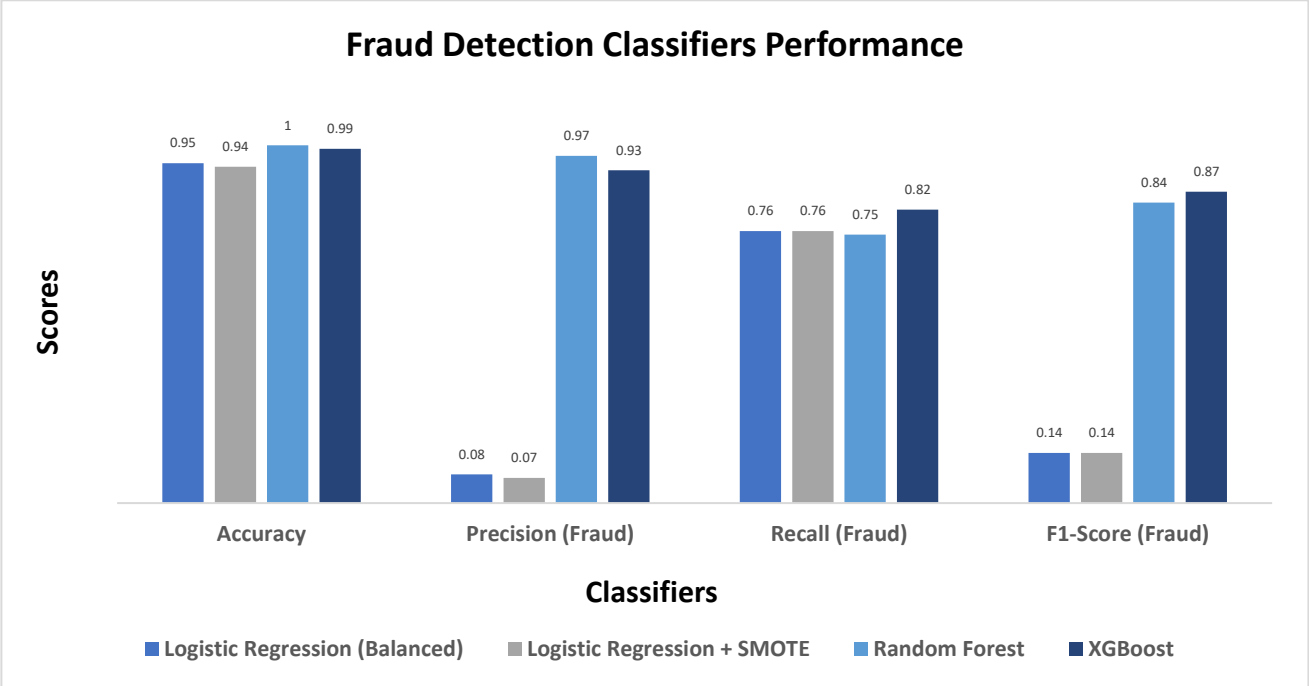
3.The ROC AUC value of 0.91 (91%) shows excellent ability to distinguish between fraudulent and genuine transactions.

4.The ROC curve AUC of 0.993 (99.3%) and the mean cross-validation AUC of 0.985 (98.5%) confirm that the model is highly stable, performs consistently across different data splits, and generalizes very well to unseen data.

5.Compared to the other models, XGBoost provides the best balance between fraud detection and false positive control, making it the most suitable model for real-world credit card fraud detection.

➢**Performance Metrics for Fraudulent Transactions (Class = 1):**

Classifier	Accuracy	Precision (Fraud)	Recall (Fraud)	F1-Score (Fraud)	Support (Fraud N)
Logistic Regression (Balanced)	95%	0.08	0.76	0.14	1,201
Logistic Regression + SMOTE	94%	0.07	0.76	0.14	1,201
Random Forest	100%	0.97	0.75	0.84	1,201
XGBoost	99.90%	0.93	0.82	0.87	1,201



Interpretations:

From the above table, we can observe that **XGBoost** achieves the highest F1-score (0.87) and strong recall (0.82) for detecting fraudulent transactions, indicating it successfully identifies the majority of fraud cases while keeping false positives relatively low.

Random Forest also performs well, with the highest precision (0.97), meaning most transactions it flags as fraud are truly fraudulent, although its recall (0.75) is slightly lower.

Both **Logistic Regression models** (Balanced and SMOTE) show moderate recall (0.76) but very low precision (0.07–0.08), implying that while they catch many fraud cases, a large number of genuine transactions are incorrectly flagged as fraud.

Overall, for practical fraud detection, **XGBoost and Random Forest** provide the best balance between catching fraud and minimizing false alerts, making them the most reliable choices for

Major Findings

a) The dataset shows a high-class imbalance, where fraudulent transactions form a very small proportion compared to genuine transactions. This confirms that credit card fraud detection is a rare-event problem and requires specialized analytical techniques.

b) A clear difference is observed in transaction amount patterns between fraudulent and genuine transactions. Fraudulent transactions have significantly higher average and median amounts, indicating that high-value transactions are more likely to be associated with fraud.

c) The distribution of transaction amounts is positively skewed, with most transactions being of small value and a few extremely high-value transactions. These extreme values often correspond to suspicious or fraudulent behaviour.

d) Time of transaction plays an important role in fraud occurrence. A relatively higher fraud rate is observed during late-night and early-morning hours, suggesting that fraudsters may take advantage of reduced monitoring during these periods.

e) Fraudulent activity varies across the days of the week, with slightly higher fraud rates observed during weekends. This indicates that temporal patterns are useful indicators in fraud detection.

f) The merchant category is strongly associated with fraud occurrence. Certain transaction categories, especially those related to online and non-essential purchases, show higher fraud rates compared to routine spending categories.

g) The time gap between consecutive transactions is significantly different for fraudulent and genuine transactions. Fraudulent transactions tend to occur within shorter time intervals, indicating rapid successive usage of the card.

h) The Mann–Whitney U test confirms that the difference in time-gap behaviour between fraudulent and non-fraudulent transactions is statistically significant, proving that transaction timing is a strong behavioural indicator of fraud.

i) Correlation analysis shows that fraud is influenced by multiple variables together rather than a single factor. Transaction amount and time-based variables contribute more strongly to fraud detection than geographic distance alone.

j) Among all machine learning models tested, Logistic Regression shows high recall but very low precision, meaning it detects many fraud cases but also produces a large number of false alarms.

k) The Random Forest classifier achieves very high precision and overall accuracy, showing strong capability in correctly identifying genuine transactions while still detecting most fraud cases.

l) The XGBoost classifier delivers the best overall performance, achieving the highest F1-score and ROC-AUC value among all models. It provides the best balance between detecting fraudulent transactions and minimizing false positives.

m) The strong ROC-AUC performance of XGBoost indicates excellent discrimination ability, meaning the model effectively distinguishes between fraudulent and genuine transactions even in an imbalanced dataset.

Recommendations

Based on the statistical analysis and machine learning results, the following recommendations are suggested to improve real-world credit card fraud detection systems:

• Financial institutions should give special attention to high-value transactions, as fraudulent transactions tend to involve significantly larger amounts compared to genuine transactions.

• Transactions occurring during late-night and early-morning hours should be monitored more carefully, since fraud activity is relatively higher during these time periods.

• Banks should incorporate time-gap monitoring systems that flag cards making multiple transactions within very short intervals, as rapid successive transactions are a strong indicator of fraudulent behaviour.

• Merchant categories with higher observed fraud rates should be placed under risk-based monitoring, where transactions from such categories undergo additional verification or scrutiny.

• Fraud detection systems should not rely on a single variable. Instead, institutions should use multi-factor behavioural analysis**,** combining transaction amount, time patterns, merchant category, and customer behaviour for better detection.

• Since fraud datasets are highly imbalanced, organizations should implement class-imbalance handling techniques such as resampling or cost-sensitive learning while training detection models.

• Among the models studied, XGBoost is recommended for practical deployment because it provides the best balance between identifying fraudulent transactions and minimizing false alerts.

• Even with a strong model, human review mechanisms should remain in place for high-risk flagged transactions to reduce customer inconvenience due to false positives.

• Fraud detection models should be updated regularly using new transaction data so that the system can adapt to changing fraud patterns and emerging techniques used by fraudsters.

• Organizations should also invest in real-time fraud detection infrastructure**,** where suspicious transactions are flagged instantly before financial loss occurs.

CONCLUSION

This study focused on detecting fraudulent credit card transactions using a combination of statistical analysis and machine learning techniques. The analysis revealed that fraudulent behaviour differs significantly from genuine transaction patterns in terms of transaction amount, timing, merchant category, and transaction frequency.

Statistical exploration showed that fraudulent transactions generally involve higher amounts and occur within shorter time intervals. Time-based behavioural analysis, supported by the Mann–Whitney U test, confirmed that transaction timing patterns of fraud cases are statistically different from genuine ones. These findings highlight the importance of behavioural and temporal features in fraud detection.

Several machine learning models were developed and evaluated to classify transactions as fraudulent or genuine. Logistic Regression demonstrated good fraud detection capability but produced a high number of false alarms. Random Forest showed strong performance with high precision. However, XGBoost emerged as the best-performing model, achieving the highest F1-score and ROC-AUC value. It provided the most effective balance between detecting fraud cases and minimizing incorrect fraud alerts.

Overall, the study proves that combining statistical insights with advanced machine learning algorithms significantly improves fraud detection performance. The results suggest that intelligent, data-driven fraud detection systems can help financial institutions reduce financial losses, protect customers, and strengthen trust in digital payment systems.

This project demonstrates that modern analytical techniques can play a vital role in addressing real-world financial fraud problems and provides a strong foundation for further research and real-time fraud detection system development.

Limitations of the Study:

• This study is based on a secondary dataset from Kaggle representing transactions from 2019–2020. Fraud patterns may change over time, so the findings may not fully reflect current real-world fraud behaviour.

• The dataset does not include certain practical banking features such as device details, transaction channel, or customer spending history, which could further improve fraud detection accuracy.

• The models were developed and tested on historical (offline) data and were not implemented in a real-time banking system. Hence, actual real-world performance may vary.

• Credit card fraud behaviour can differ across countries, banks, and customer groups. Therefore, the results of this study cannot be generalized to all financial environments**.**

• Only selected statistical and machine learning techniques were used. Other advanced methods may provide different or improved results.

References:

1.Lucas, Yvan & Jurgovsky, Johannes. (2020). *Credit card fraud detection using machine learning: A survey*. 10.48550/arXiv.2010.06479.

2.Waspada, I., Handayani, A. D., & Suryanegara, M. (2022). *Performance Analysis of Isolation Forest Algorithm in Fraud Detection of Credit Card Transactions*. *Procedia Computer Science, 197*, 346–353.

3.Ghanem, M., Elkaffas, S. M., & Madbouly, M. (2022). *Machine Learning Technique for Credit Card Fraud Detection*. *Arab Academy for Science, Technology, and Maritime Transport*.

4.Abu Rbeian, Alsharif Hasan & Ashqar, Huthaifa. (2023). *Credit Card Fraud Detection Using Enhanced Random Forest Classifier for Imbalanced Data*. 10.48550/arXiv.2303.06514.

5.Gostkowski, M., Krasnodębski, A., & Niedziółka, A. (2024). *Credit Card Fraud Detection Using Machine Learning Techniques*. *European Research Studies Journal, 27(2)*, 571–585.

6.Wang, Y. (2025, March 27). *A Data Balancing and Ensemble Learning Approach for Credit Card Fraud Detection* [Preprint]. *arXiv*.