DEEP LEARNING APPROACHES FOR FRAUD DETECTION IN E – COMMERCE TRANSACTIONS

Video Link : https://youtu.be/Jj2223LHf6s

PREPARE BY : MOHAMED AZLAN AMEER OLI- MCS241050

SV : PROF. MADYA DR. MOHD. SHAHIZAN OTHMAN

PREPARED BY :

MOHAMED AZLAN AMEER OLI

MCS241050 Master Data Science Student

ASSOC. PROF. DR. MOHD SHAHIZAN BIN OTHMAN

Lecturer, Universiti Teknologi Malaysia (UTM)

TABLE OF CONTENTS

INTRODUCTION

Problem Statement

- Rise of cashless payments and online transactions introduces new fraud risks
- Traditional methods rely on manual checks or simple rule-based systems
- Struggle to process large volumes of financial data
- Class imbalance: Fraud cases are rare compared to legitimate transactions
- Leads to poor detection accuracy with conventional models
- Fraud tactics evolve rapidly, making static systems ineffective
- Highlights the need for intelligent, adaptive detection solutions (Nama & Obaid, 2024)

Proposed Deep Learning Approach

- Deep learning can uncover hidden patterns in large transactional datasets
- This research introduces an adaptive fraud detection system
- Combines Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) models
- Offers a more accurate and responsive solution for e-commerce fraud prevention

Research Aim

This project aims to identify fraud and non – fraud transactions in e – commerce using RNN and LSTM models and identify which is the best model to predict the fraudulent activities in e-commerce.

Research Objectives

UTM
UNIVERSITI TEKNOLOGI MALAYSIA

The objectives of this study are follows:

A. To investigate the deep learning – based approach for fraud transactions detection.

B. To implement the method used for fraud transactions detections based on deep learning method.

C. To predict the accuracy of the model used for fraud transaction detection

Research Scope

The scopes of this project are bound under the following constraints to accomplish this work:

A. The study utilizes the dataset of the synthetic dataset of Fraudulent activities in e-commerce.

B. The experiment related will be developed in Python programming.

C. The proposed model used Recurrent Neural Networks (RNNs) and Long Short – Term Memory (LSTM)

Expected Research Contribution

- Evaluate deep learning models for fraud detection
- Identify the most effective approach using RNN and LSTM
- Develop a real-time fraud detection system with a visual dashboard
- Help enhance security of e-commerce platforms

LITERATURE REVIEW

Literature Review

E - Commerce Transaction

Research that uses e - commerce transactions dataset to analyze the pattern of the transaction and help to detect fraud and non - fraud.

Deep Learning & Machine Learning Techniques

Research that studies various deep learning and machine learning techniques to detect fraud transactions.

Finding Best Model

Research that finds best models of Deep Learning that use to detect fraudulent activities at E - Commerce transactions.

Previous studies on results of Supervised Learning Method

Author / Year
Branco et al. (2020)
El Kafhali et al. (2024)
Benchaji et al. (2021)
Kumar & Swathi (2024)
Lin et al. (2021)
Nama & Al – Salam (2024)
Springer (2024)
Vanini et al.
Alarfaj et al. (2022)
Kodate et al.
Dantas et al.

Shows previous studies results of Unsupervised Learning Method

UTM UNIVERSITI TEKNOLOGI MALAYSIA

Author / Year	Unsupervised Lea
Li et al. (2025)	Contrastive Learning
Lu et al. (2021)	Graph Neural Networks (G Architecture
Ren et al. (2019)	Bipartite Graph + Clusterir
Kodate et al.	Community Detection in G
Kennedy et al. (Unsupervised Cleaning)	Interactive Cleaning + Clus Ensemble)

Shows the previous work of researcher in Deep Learning Methods

Author / Year	Research Title	Research Focus	
Branco et al. (2020)	Interleaved Sequence RNNs for Fraud Detection	Sequential modeling of transactions	Limite sequer
El Kafhali et al. (2024)	An Optimized Deep Learning Approach for Detecting Fraudulent Transactions	Deep Learning for fraud detection	Need f deep l
Benchaji et al. (2021)	Enhanced Credit Card Fraud Detection Based on Attention Mechanism and LSTM Deep Model	Attention – Enhanced LSTM for fraud detection	Low ex attenti detect
Kumar & Swathi (2024)	Fine – Tuned LSTM for Credit Card Fraud Detection and Classification	Fine – tuning LSTM for fraud classification	Lack o LSTM

Shows the previous work of researcher in Deep Learning Methods

MDPI Information (2024)	An Optimized Deep Learning Approach for Detecting Fraudulent Transactions	Deep learning model optimization for fraud	Need fc comput
Nama & Al – Salam (2024)	Financial Fraud Identification Using Deep Learning Techniques	Applying various DL models for fraud	Lack of methoc
Ren et al. (2019)	EnsemFDet: An Ensemble Approach to Fraud Detection Based on Bipartite Graph	Graph ensemble model for fraud	Sparse combin
[illegible]	An Intelligent Sequential Fraud [illegible]	Deep learning for sequential	Conver

Literature Review

UNIVERSITI TEKNOLOGI MALAYSIA

Research Gaps

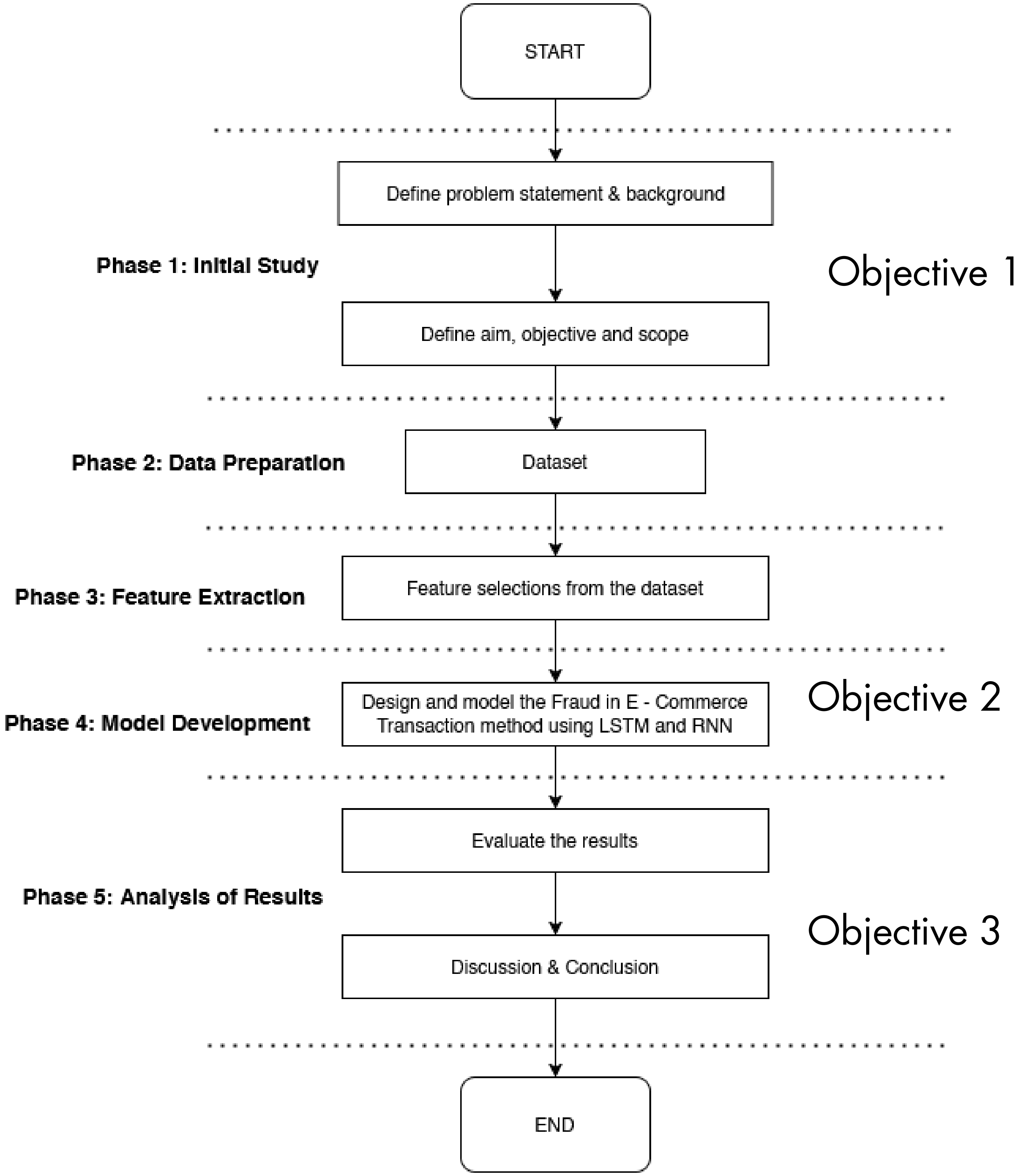
- **Data Limitation in Current Models:** While models like LSTM and RNN perform well in fraud detection, they require large labeled datasets, which are often unavailable in real-world scenarios—highlighting the need for semi-supervised or unsupervised methods.
- **Narrow Focus on Credit Card Fraud**: Most studies focus on credit card fraud, not the broader and more complex e-commerce fraud landscape that includes multiple platforms and payment types.
- **Need for Advanced, Real-Time E-Commerce Solutions**: Future research should develop deep learning models that handle limited data, work in real-time, are more interpretable, and are tailored to the unique characteristics of e-commerce fraud.

Solutions

- Adopt Semi-Supervised and Unsupervised Learning Techniques
- Design E-Commerce-Specific Fraud Detection Models
- Integrate Real-Time and Multi-Modal Data Processing

RESEARCH METHODOLOGY

Research Frame Work



1. Phase 1: Initial Study – Conducts background research and explores existing issues in e-commerce fraud detection.
2. Phase 2: Conceptual Design and Development – Outlines the structure of the proposed deep learning model and selects appropriate methods and tools.
3. Phase 3: Model Development – Builds the deep learning models (LSTM and RNN) for fraud detection based on the defined concepts.
4. Phase 4: Implementation – Applies the developed models to the selected dataset to detect fraudulent transactions.
5. Phase 5: Analysis of Results – Compares the performance of LSTM and RNN models, aiming to

Phase 1: Initial Study

UNIVERSITI TEKNOLOGI MALAYSIA

Background & Problem Statement

- Rise of online and cashless transactions increases complexity in fraud detection.
- Traditional rule-based systems struggle with large-scale and imbalanced data.
- Fraudulent transactions are rare, making them harder to detect accurately.
- Need for real-time, adaptive models due to evolving fraud techniques.
(Nama & Obaid, 2024)

Deep Learning in Fraud Detection

- RNN and LSTM are effective in capturing transaction patterns.
- Real-world deployment faces issues with imbalanced & growing datasets.
- Bayesian optimization helps improve model accuracy and efficiency.(Branco et al., 2020; Lin et al., 2021; El Kafhali et al., 2024)

Phase 2: Data Preparation

UTM UNIVERSITI TEKNOLOGI MALAYSIA

Dataset Summary

- Total Transactions: 1,472,952
- Features: 16
- Non-Fraudulent: ~95%
- Fraudulent: ~5%

Data Preparation Steps:

Removed missing values, duplicate rows, and inconsistencies
Irrelevant data excluded to improve model accuracy
Text Preprocessing: Converted to lowercase, removed irrelevant terms
Date Formatting: Standardized transaction timestamps

Phase 3: Feature Extraction

UTM UNIVERSITI TEKNOLOGI MALAYSIA

Feature Extraction & Preprocessing

Extracted key features: Transaction Date, Payment Method, Product Category, Customer Location, Device Used

Standardized text: lowercased, removed extra spaces

Reformatted date for temporal analysis

Removed missing values and duplicate entries

Encoded categorical variables for model input

Used heatmaps and bar plots to identify predictive features

Final output: Cleaned and scaled dataset ready for RNN and LSTM model training

	Transaction ID	Customer ID	Transaction Amount	Transaction Date	Payment Method	Product Category	Quantity	Customer Age	Customer Location	Device Used	IP Address	Shipping Address	Billing Address	Is Fraudulent	Account Age Days	Transaction Hour
0	15d2e414-8735-46fc-9e02-80b472b2580f	d1b87f62-51b2-493b-ad6a-77e0fe13e785	58.09	2024-02-20 05:58:41	bank transfer	electronics	1	17	Amandaborough	tablet	212.195.49.198	Unit 8934 Box 0058\nDPO AA 05437	Unit 8934 Box 0058\nDPO AA 05437	0	30	5
1	0bfee1a0-6d5e-40da-a446-d04e73b1b177	37de64d5-e901-4a56-9ea0-af0c24c069cf	389.96	2024-02-25 08:09:45	debit card	electronics	2	40	East Timothy	desktop	208.106.249.121	634 May Keys\nPort Cherylview, NV 75063	634 May Keys\nPort Cherylview, NV 75063	0	72	8
2	e588eef4-b754-468e-9d90-d0e0abfc1af0	1bac88d6-4b22-409a-a06b-425119c57225	134.19	2024-03-18 03:42:55	PayPal	home & garden	2	22	Davismouth	tablet	76.63.88.212	16282 Dana Falls Suite 790\nRothhaven, IL 15564	16282 Dana Falls Suite 790\nRothhaven, IL 15564	0	63	3
3	4de46e52-60c3-49d9-be39-636681009789	2357c76e-9253-4ceb-b44e-ef4b71cb7d4d	226.17	2024-03-16 20:41:31	bank transfer	clothing	5	31	Lynnberg	desktop	207.208.171.73	828 Strong Loaf Apt. 646\nNew Joshua, UT 84798	828 Strong Loaf Apt. 646\nNew Joshua, UT 84798	0	124	20
4	074a76de-fe2d-443e-a00c-f044cdb68e21	45071bc5-9588-43ea-8093-023caec8ea1c	121.53	2024-01-15 05:08:17	bank transfer	clothing	2	51	South Nicole	tablet	190.172.14.169	29799 Jason Hills Apt. 439\nWest Richardtown, ...	29799 Jason Hills Apt. 439\nWest Richardtown, ...	0	158	5

Phase 4: Model Development

UTM
UNIVERSITI TEKNOLOGI MALAYSIA

Model Development & Evaluation

- Models: LSTM and RNN used to detect fraudulent transactions in e-commerce.
- Data Split: 80% training, 20% testing; input reshaped to 3D (samples, time steps,
features).

LSTM Architecture:

- Input layer
- LSTM layer with 64 units
- Dense layer with sigmoid activation (binary output)

Training Setup:

- Loss: Binary cross-entropy
- Optimizer: Adam
- Batch size: 64 | Multiple epochs
- Validation set used to prevent overfitting

Evaluation Metrics:

- Accuracy, Precision, Recall, F1-Score, Confusion Matrix
- RNN: Trained and evaluated using the same approach for result comparison

Phase 5: Analysis of Results

UTM
UNIVERSITI TEKNOLOGI MALAYSIA

In this phase, the output of the fraudulent transaction detection is analyzed. The model between LSTM and RNN comparison in terms of accuracy and prediction has been validated. The output is based supervised learning on fraud and non – fraud transactions. The performance measure of the data has been discussed in this phase.

Exploratory Data Analysis (EDA)

- EDA helps visualize data to uncover patterns and anomalies.
- Begins with understanding the data and identifying potential issues.
- Checks for missing values and inconsistencies, which are handled by removal or imputation.
- Analyzes data distribution, averages, and overall structure.
- Applies SMOTE for handling class imbalance and improving dataset quality.
- Uses visualizations (graphs, charts) to highlight trends and insights.
- Detects and addresses outliers to ensure reliable analysis.
- Final findings summarized through visual and statistical outputs.

Steps of Exploratory Data Analysis (EDA)

Data Collection

	Transaction ID	Customer ID	Transaction Amount	Transaction Date	Payment Method	Product Category	Quantity	Customer Age	Customer Location	Device Used	IP Address	Shipping Address	Billing Address	Is Fraudulent	Account Age Days	Transaction Hour
0	15d2e414-8735-46fc-9e02-80b472b2580f	d1b87f62-51b2-493b-ad6a-77e0fe13e785	58.09	2024-02-20 05:58:41	bank transfer	electronics	1	17	Amandaborough	tablet	212.195.49.198	Unit 8934 Box 0058\nDPO AA 05437	Unit 8934 Box 0058\nDPO AA 05437	0	30	5
1	0bfee1a0-6d5e-40da-a446-d04e73b1b177	37de64d5-e901-4a56-9ea0-af0c24c069cf	389.96	2024-02-25 08:09:45	debit card	electronics	2	40	East Timothy	desktop	208.106.249.121	634 May Keys\nPort Cherylview, NV 75063	634 May Keys\nPort Cherylview, NV 75063	0	72	8
2	e588eef4-b754-468e-9d90-d0e0abfc1af0	1bac88d6-4b22-409a-a06b-425119c57225	134.19	2024-03-18 03:42:55	PayPal	home & garden	2	22	Davismouth	tablet	76.63.88.212	16282 Dana Falls Suite 790\nRothhaven, IL 15564	16282 Dana Falls Suite 790\nRothhaven, IL 15564	0	63	3
3	4de46e52-60c3-49d9-be39-636681009789	2357c76e-9253-4ceb-b44e-ef4b71cb7d4d	226.17	2024-03-16 20:41:31	bank transfer	clothing	5	31	Lynnberg	desktop	207.208.171.73	828 Strong Loaf Apt. 646\nNew Joshua, UT 84798	828 Strong Loaf Apt. 646\nNew Joshua, UT 84798	0	124	20
4	074a76de-fe2d-443e-a00c-f044cdb68e21	45071bc5-9588-43ea-8093-023caec8ea1c	121.53	2024-01-15 05:08:17	bank transfer	clothing	2	51	South Nicole	tablet	190.172.14.169	29799 Jason Hills Apt. 439\nWest Richardtown, ...	29799 Jason Hills Apt. 439\nWest Richardtown, ...	0	158	5

Figure 4.1: Fraudulent E-Commerce Transactions Dataset

Import and Inspect Dataset

```
Both fraudulent and non-fraudulent transactions found.
Number of fraudulent transactions: 73838
Number of non-fraudulent transactions: 1399114
```

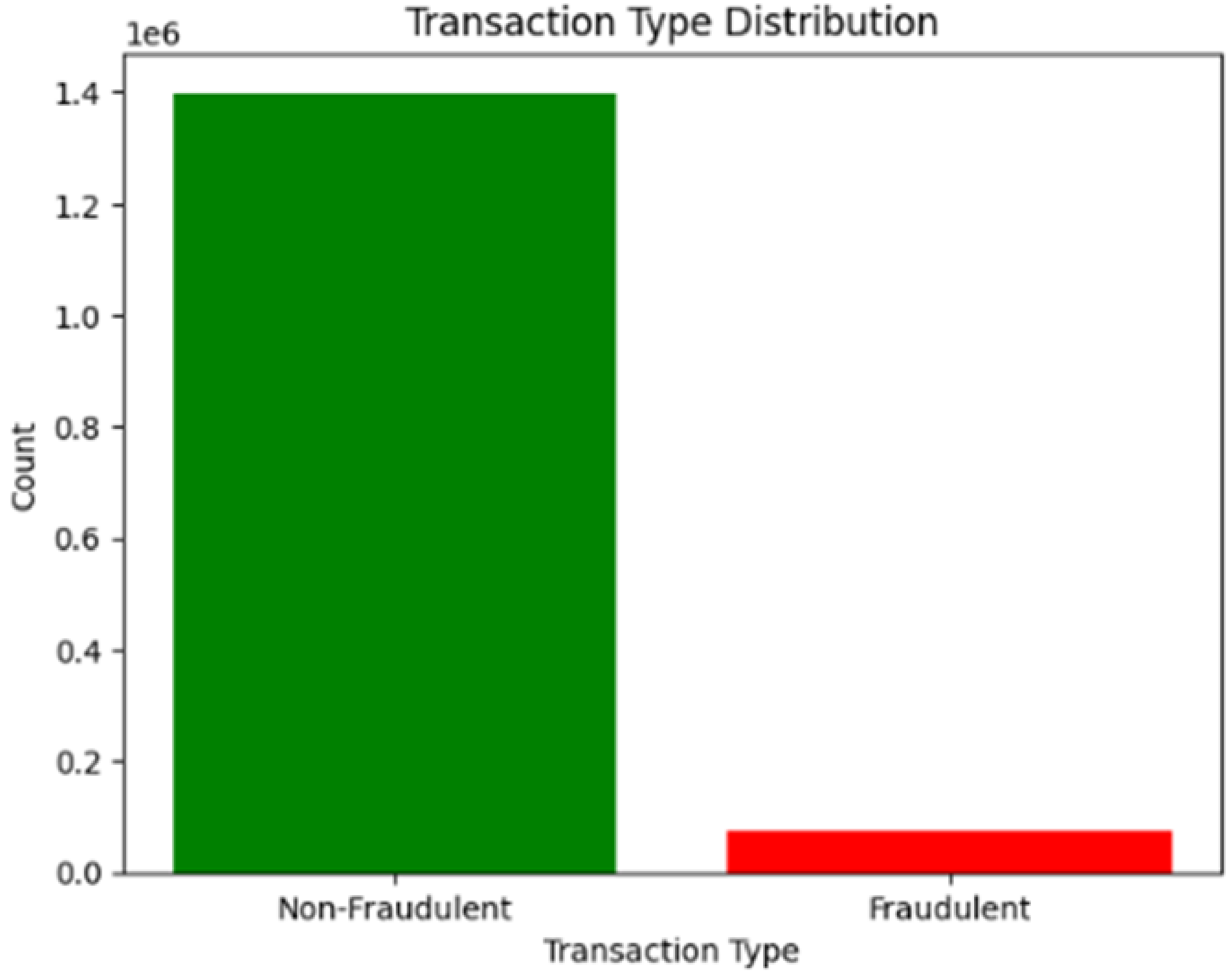


Figure 4.2: Transactions Type Distribution

Demographic and Distribution Data

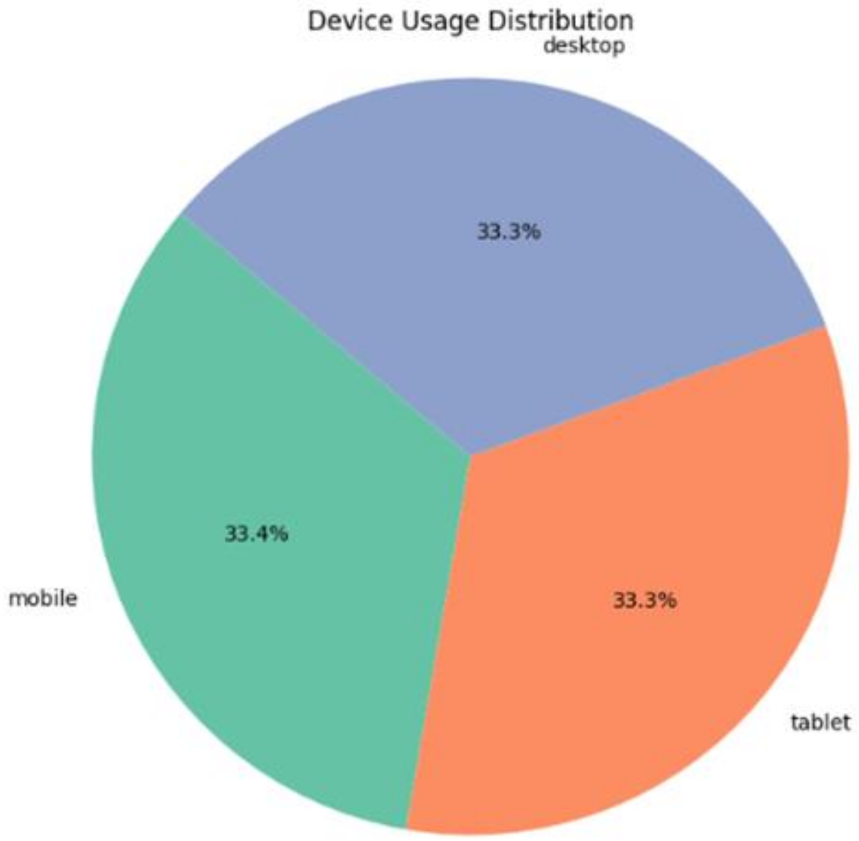


Figure 4.4: Device Usage Distribution

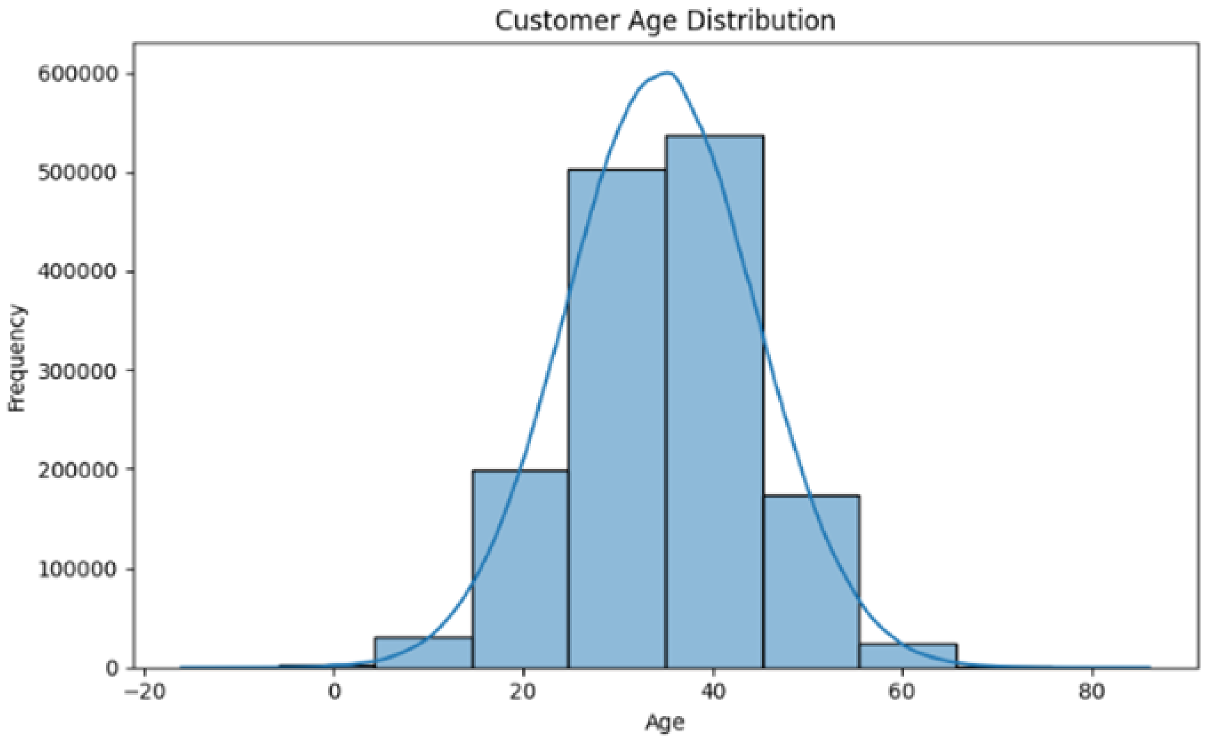


Figure 4.3: Customer Age Distribution

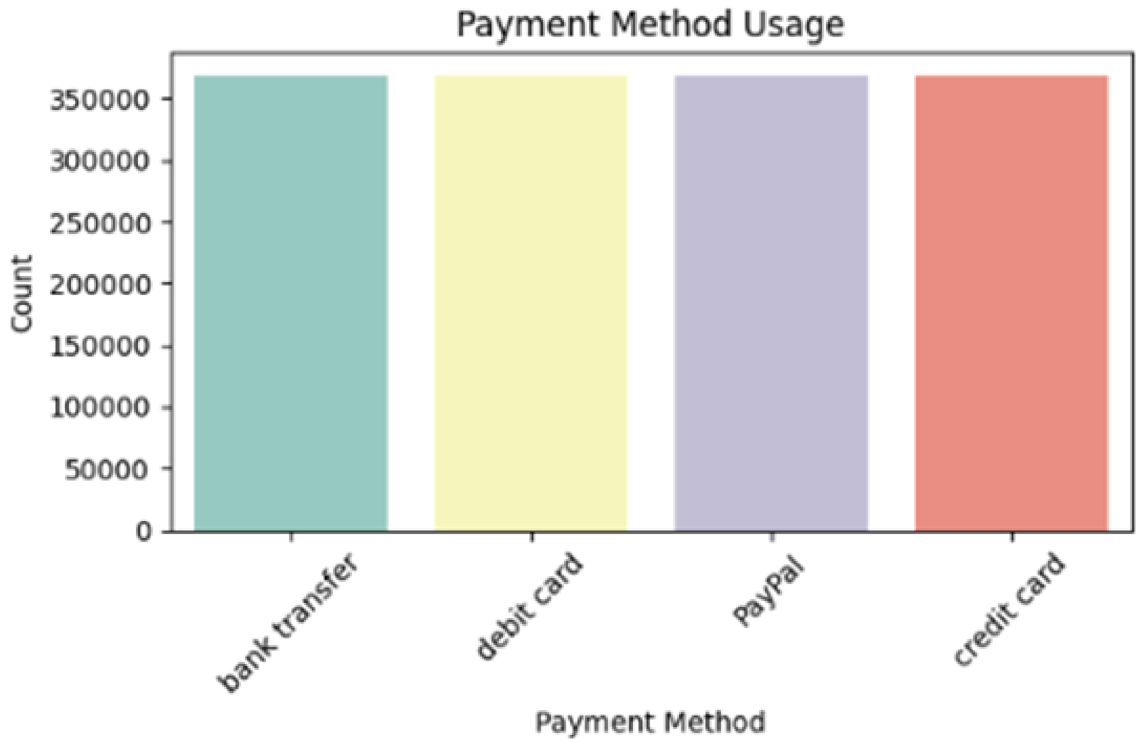


Figure 4.5: Payment Method Usage

Data Cleaning

```python
# Step 1: Drop rows with missing target or duplicate entries
df = df.drop_duplicates()
df = df.dropna(subset=['Is Fraudulent'])  # Replace with your actual target column name

# Convert 'Transaction Date' to datetime format
df['Transaction Date'] = pd.to_datetime(df['Transaction Date'])

# Standardize categorical text (lowercase)
text_columns = ['Payment Method', 'Product Category', 'Customer Location', 'Device Used']
for col in text_columns:
    df[col] = df[col].str.lower().str.strip()

df = df.reset_index(drop=True)
df.head()
rows,columns = df.shape
print(f"The dataset contains {rows} rows and {columns} columns.")
```

```
The dataset contains 1472952 rows and 16 columns.
```

Figure 4.6: Data Cleaning Code

Using SMOTE Model for Balancing Data

UTM UNIVERSITI TEKNOLOGI MALAYSIA

```
After SMOTE:
Number of fraudulent transactions: 932742
Number of non-fraudulent transactions: 1399114
```

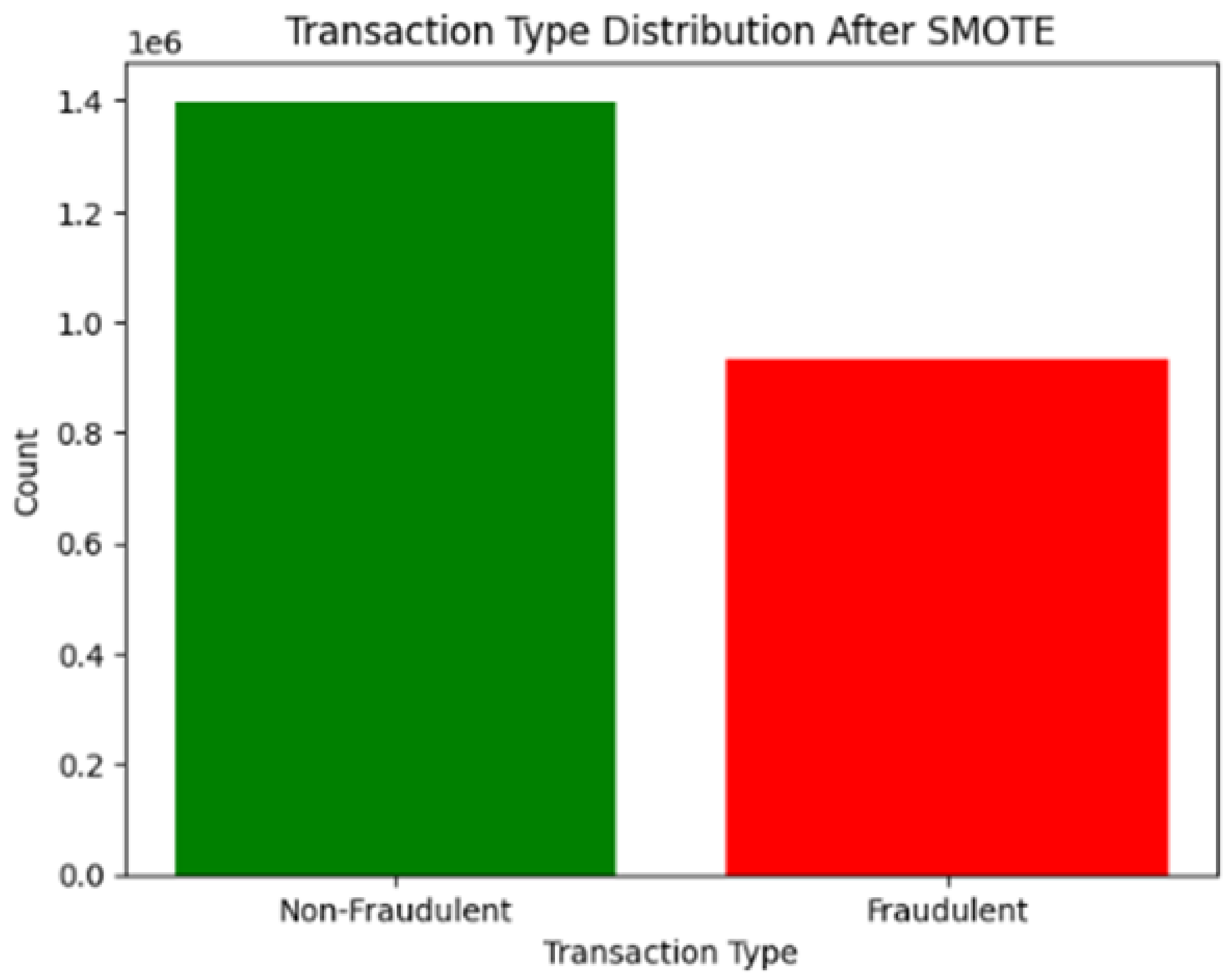


Figure 4.7: Transaction Type Distribution After SMOTE

Feature Extraction

UTM UNIVERSITI TEKNOLOGI MALAYSIA

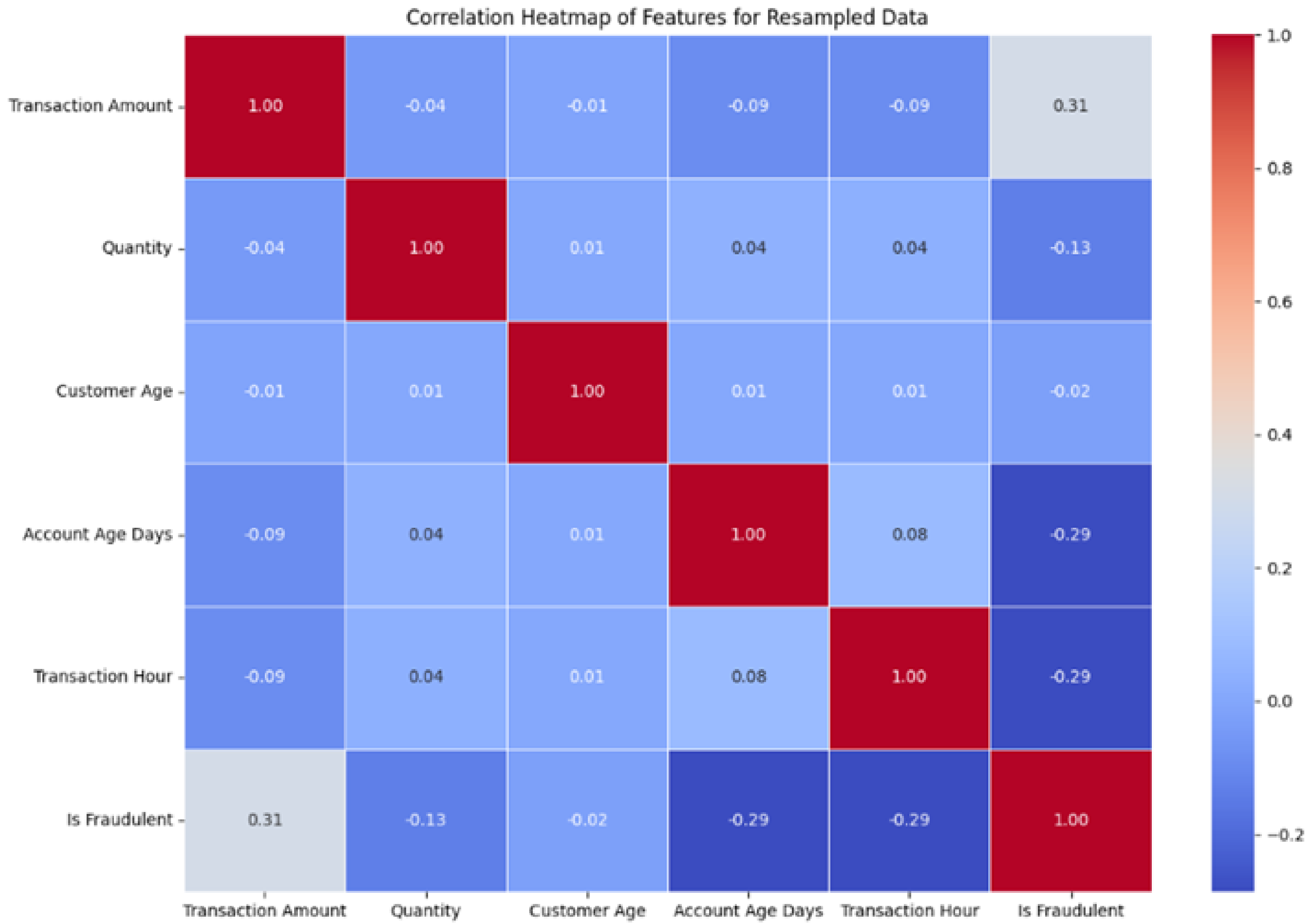


Figure 4.8: Correlation Heatmap of Features for Resampled Data

Data Modeling

UTM UNIVERSITI TEKNOLOGI MALAYSIA

```
# Build the model
model = Sequential()
model.add(LSTM(128, return_sequences=True, input_shape=input_shape))
model.add(Dropout(0.4))
model.add(LSTM(64))
model.add(Dense(1, activation='sigmoid'))
```

Figure 4.9: LSTM Modeling

```
model = Sequential([
    SimpleRNN(128, return_sequences=True, input_shape=input_shape),
    Dropout(0.4),
    SimpleRNN(64),
    Dense(1, activation='sigmoid')
])
```

Figure 4.10: RNN Modeling

INITIAL FINDING AND RESULTS

Initial Results

UTM UNIVERSITI TEKNOLOGI MALAYSIA

```
Test Loss: 0.4989
Test Accuracy: 0.7574
ROC-AUC Score (from sklearn): 0.8316
Test AUC (from model.evaluate): 0.8316

Classification Report:
              precision    recall  f1-score   support

           0       0.84      0.75      0.79    278751
           1       0.66      0.76      0.71    175171

    accuracy                           0.76    453922
   macro avg       0.75      0.76      0.75    453922
weighted avg       0.77      0.76      0.76    453922
```

Figure 4.11: Initial Results of LSTM

Initial Results

```
Test Loss: 0.5001
Test Accuracy: 0.7577
ROC-AUC Score (from sklearn): 0.8284
Test AUC (from model.evaluate): 0.8283

Classification Report:
              precision    recall  f1-score   support

   Non-Fraud       0.83      0.76      0.79    278751
       Fraud       0.66      0.76      0.71    175171

    accuracy                           0.76    453922
   macro avg       0.75      0.76      0.75    453922
weighted avg       0.77      0.76      0.76    453922
```

Figure 4.13: Initial Results of RNN

Future Work

UTM
UNIVERSITI TEKNOLOGI MALAYSIA

Enhanced the Data Balancing Method

Although SMOTE was used to address class inconsistency, the minority class detection may be further improved by combining SMOTE with methods like cost – sensitive learning or ensemble under – sampling.

Model Optimization and Tuning Parameter

Future work should focus on optimizing hyperparameters such as Bayesian Optimization to enhance the performance of RNN and LSTM models. Furthermore, reducing overfitting may also be achieved by implementing the batch normalization or dropout layers.

Advanced Architectures

Using advanced neural networks such as Bidirectional LSTM and GRU (Gated Recurrent Units) may identify deeper sequential patterns. Also, time series analysis may better for model fraud patterns.

- THANK YOU -