RESEARCH PROPOSAL ★★★★★

MONEY LAUNDERING DETECTION OF SUSPICIOUS TRANSACTION USING MACHINE LEARNING ALGORITHM

Presented by:
Nur Adriana Batrisyia binti Mohd Subri
MCS241007

CONTENT OUTLINES

01

INTRODUCTION

★★★★★

PROBLEM BACKGROUND

Global Money Laundering (source: UNDOC)

2%-5% of world's GDP laundered annually

Equivalent to $800 billion - $2 trillion

Malaysia's 2023 AML Highlight (source: BNM Annual Report 2023)

317,435 Suspicious Transaction Reports (STRs)

31% increase from 2022

- Key Offences: Fraud, money laundering, tax evasion
- Over 100 individuals arrested and RM290 million recovered
- Disrupted 59,684 mule accounts

PROBLEM STATEMENT

Low effectiveness in money laundering prosecutions and convictions

Inadequately targeting high-risk offences, especially cross-border transactions

Limitations of rule-based techniques, ineffective for complex and hidden scheme

OBJECTIVES

01 To perform data preprocessing and exploratory data analysis

Handle noisy data and understand data distributions.

02 To implement machine learning algorithms

Learn patterns, identify anomaly transactions, and detect money laundering activities

03 To evaluate machine learning algorithms

Using metrics such as TPR, FPR, TNR, FNR, and AUC

PROJECT SCOPE

DATASET

- Synthetic Anti-Money Laundering Dataset (SAML-D)

ML ALGORITHM

- Support Vector Machines
- Decision Tree

PERFORMANCE METRICS

- True Positive Rate (TPR)
- False Positive Rate (FPR)
- True Negative Rate (TNR)
- False Negative Rate (FNR)
- Area Under the Curve (AUC)

02

LITERATURE REVIEW

RESEARCH GAPS

- Limitation of rule-based systems
- Limited real-world dataset on money laundering
- Imbalance transaction dataset
- Lack of research experimented on cross-border transaction dataset
- Computational limitation to analyze transaction dataset in real-time

03

RESEARCH METHODOLOGY

PROJECT LIFECYCLE

1 PROBLEM IDENTIFICATION

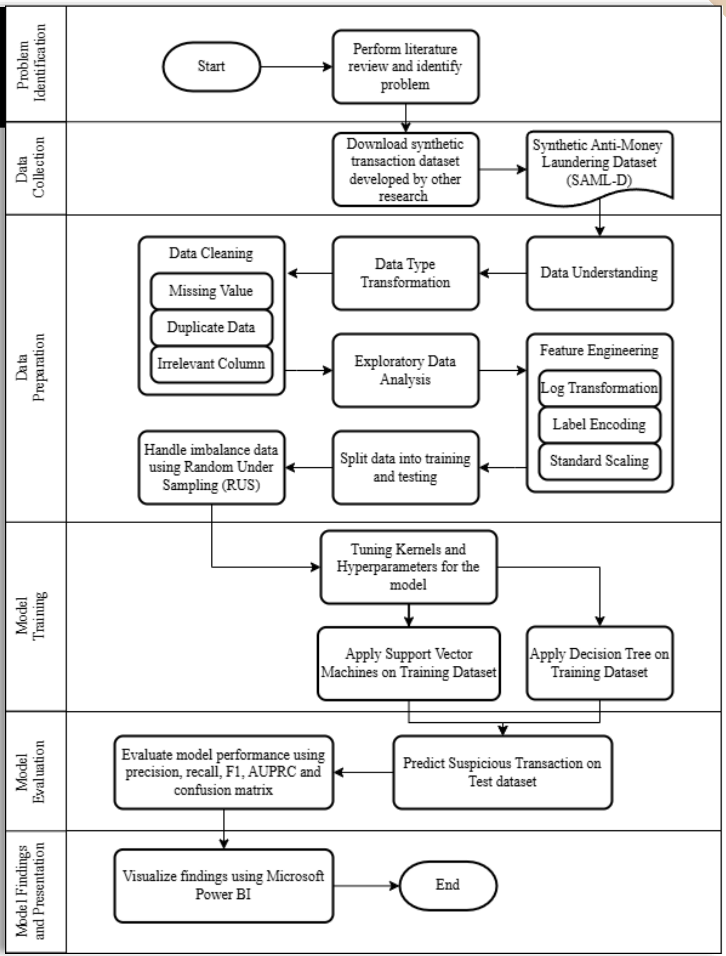
2 DATA COLLECTION

3 DATA PREPARATION

4 MODEL TRAINING

5 MODEL EVALUATION

6 MODEL FINDINGS AND PRESENTATION



1 PROBLEM IDENTIFICATION

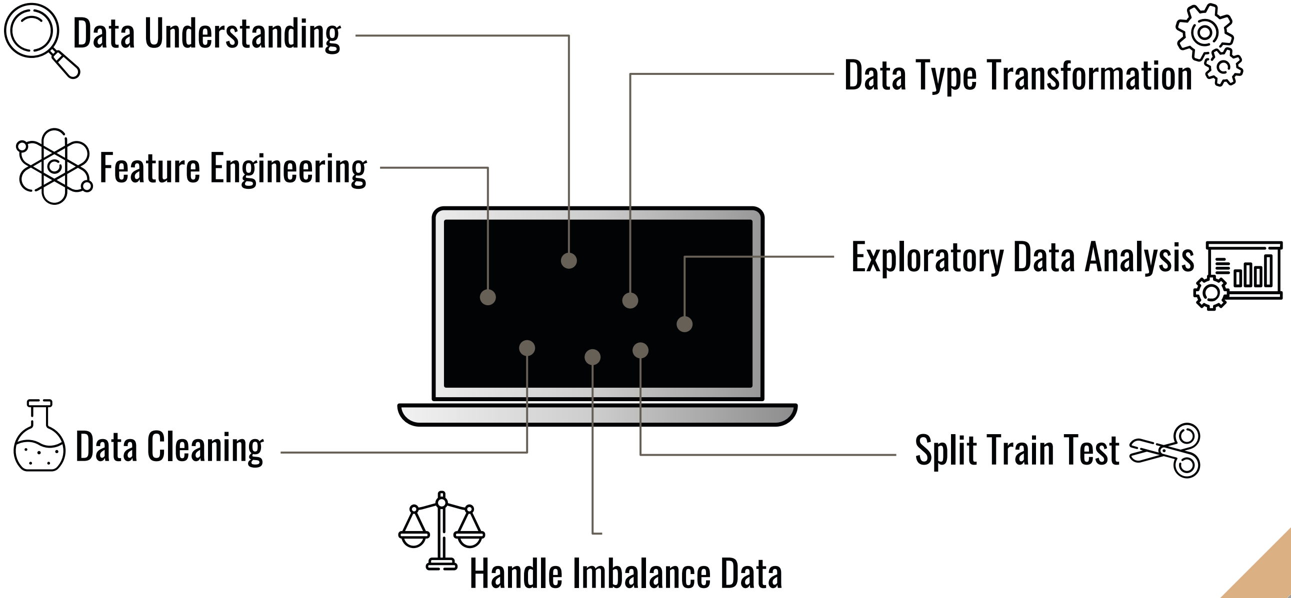
Understand the current challenges in money laundering detection through literature review

2 DATA COLLECTION

Download and extract SAML-D dataset in CSV format

Size	9,504,851 rows	
Attributes	12 Attributes 1) Time 2) Date 3) Sender_account 4) Receiver_account 5) Amount 6) Payment_currency	 7) Received_currency 8) Sender_bank_location 9) Receiver_bank_location 10) Payment_type 11) Is_laundering 12) Laundering_type

3 DATA PREPARATION



4 MODEL TRAINING

- Tuning Hyperparameters
- Apply Supervised Machine Learning on Training Dataset

5 MODEL EVALUATION

- Predict Suspicious Transaction on Test Dataset
- Evaluate Model Performance using Performance Metrics

6 MODEL FINDINGS & PRESENTATION

- Visualize findings using Microsoft Power BI

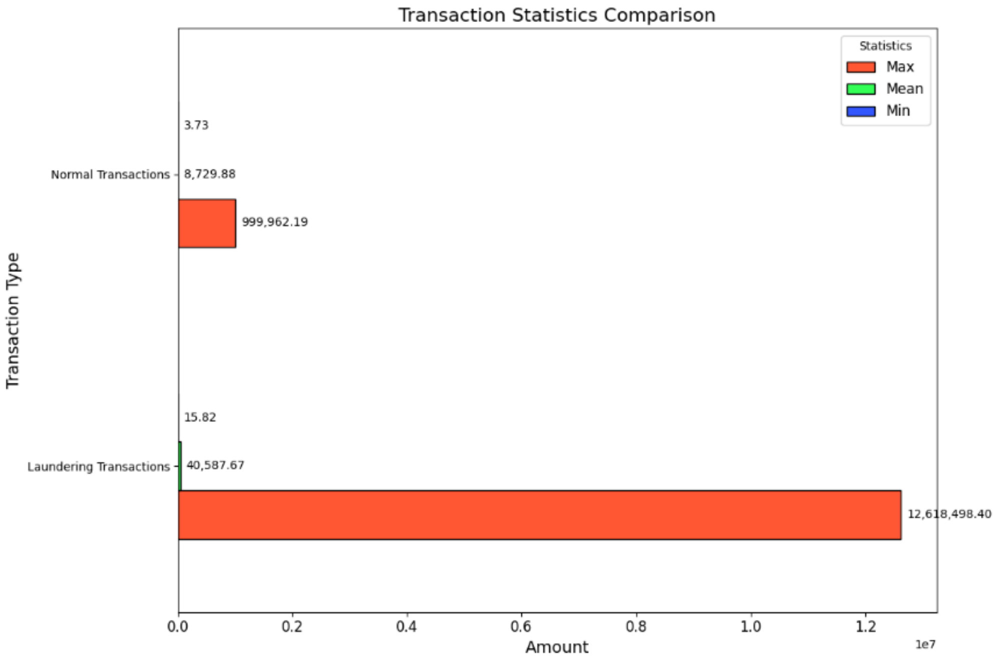
04

INITIAL RESULTS

★★★★★

EXPLORATORY DATA ANALYSIS

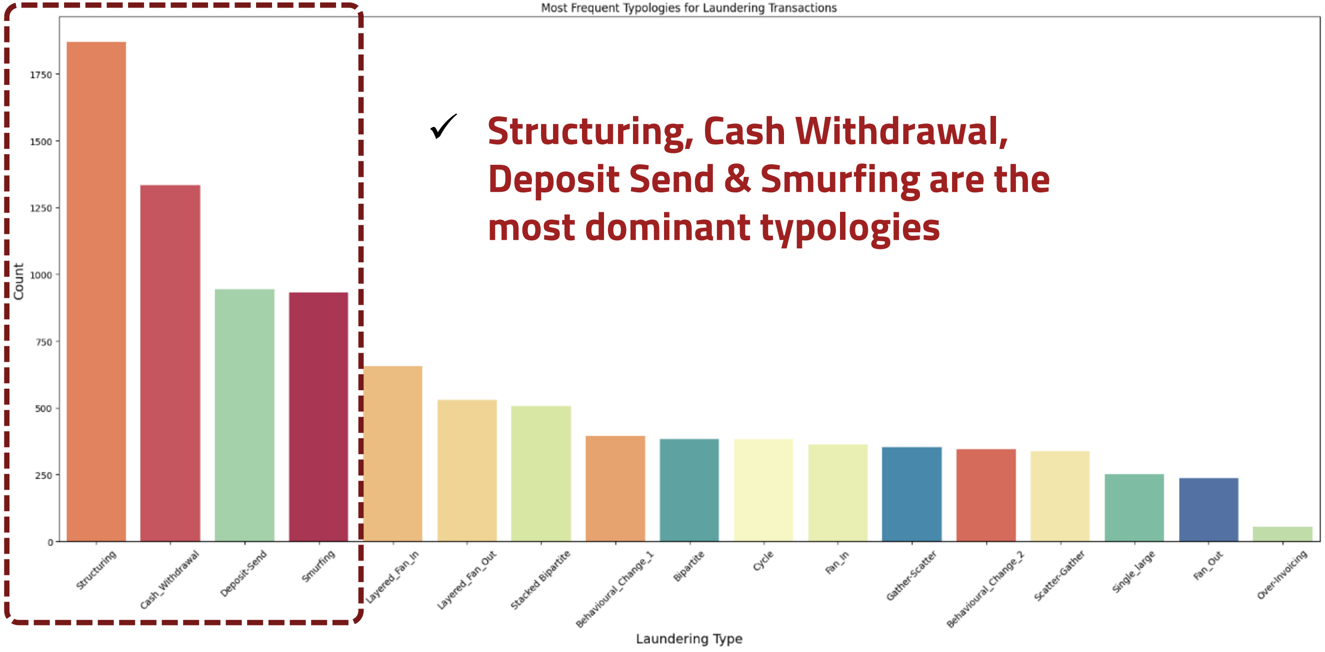
Identify Min, Max & Mean for Laundering and Normal Transactions



- ✓ Maximum amount in laundering transactions is significantly higher than normal transactions
- ✓ Both transactions have extremely small minimum amount
- ✓ **Laundering transactions often involve extreme values**

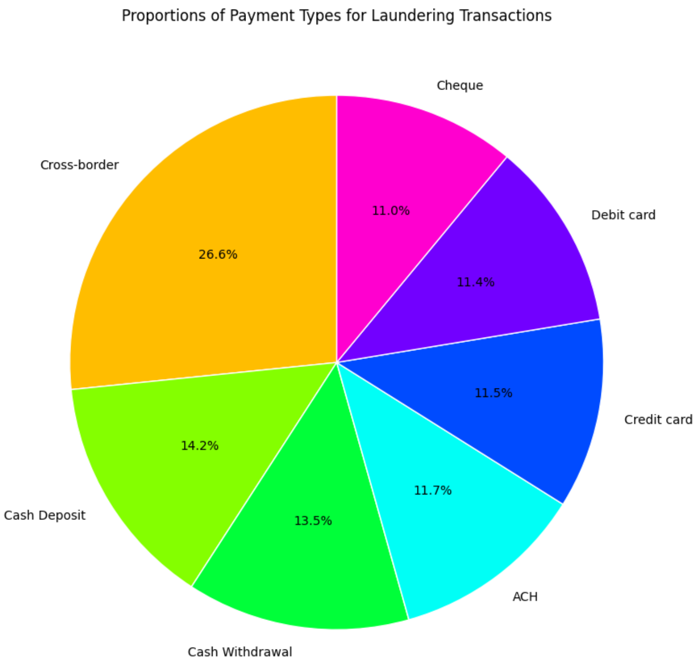
EXPLORATORY DATA ANALYSIS

Identify Most Frequent Typologies for Laundering Transactions



EXPLORATORY DATA ANALYSIS

Identify Most Frequent Payment Types for Laundering Transactions

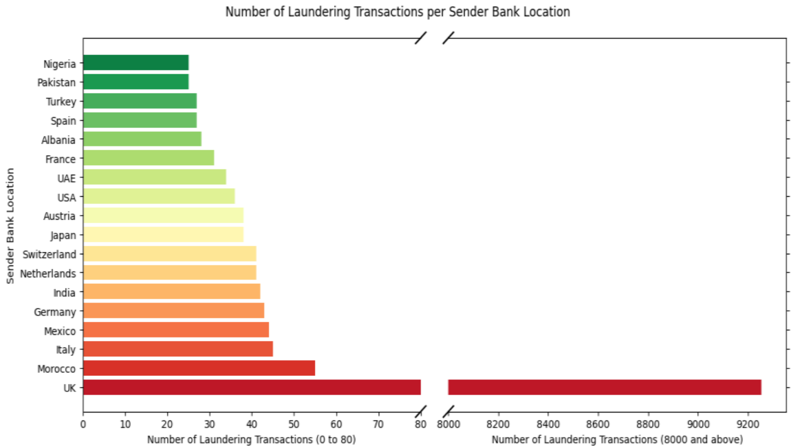


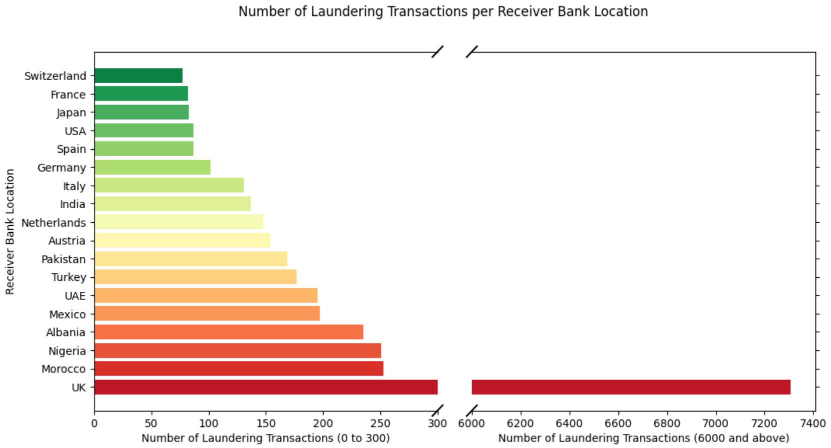
- ✓ Cross-border transactions has the largest proportions followed by Cash Deposit and Cash Withdrawal
- ✓ ACH, Credit Card, Debit Card, and Cheque have relatively similar proportion

- ✓ **Cross-border transactions is the most preferred payment method by launderers**

EXPLORATORY DATA ANALYSIS

Identify the High-Risk Bank Location

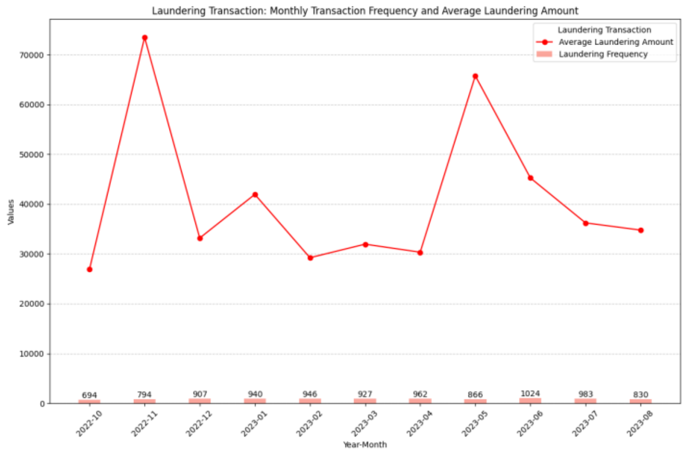


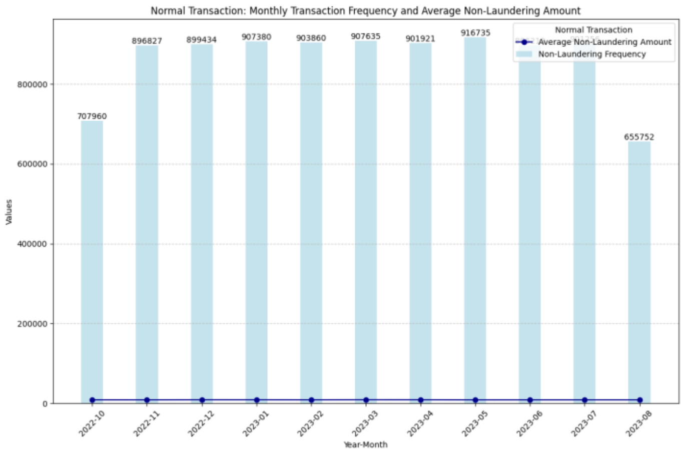


✓ **UK and Morocco are the most high-risk bank locations as it seems to be a central hub for both sending and receiving illicit money from laundering transactions**

EXPLORATORY DATA ANALYSIS

Identify Monthly Transaction Frequency and Average Laundering Amount by Transaction Type



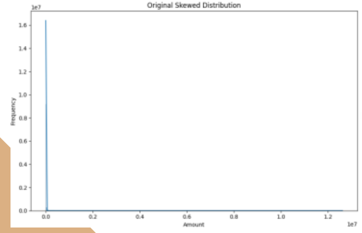


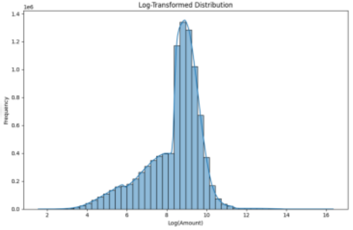
- ✓ **The sharp contrast in frequency and amount emphasize that laundering transactions occurrence are rare but usually involve larger amounts of money.**

FEATURE ENGINEERING

Log Transformation

Reduced the highly skewed dataset





Label Encoding

Transform categorical features into numerical variables

Payment Type
Cash Deposit
Cross-border
Cheque
ACH
Cash Deposit

Payment Type
1
5
3
0
1

Standard Scaling

Scale the numerical features using standard normal distribution

Amount
1459.15
6019.64
14328.44
11895.00
115.25

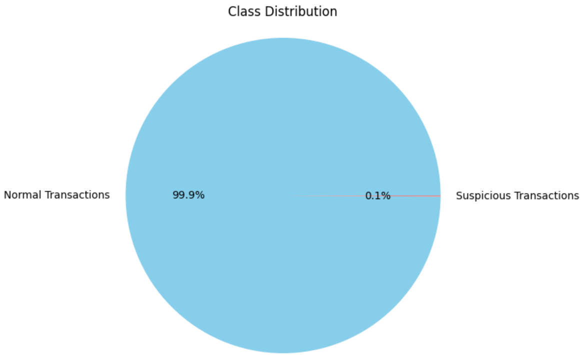
Amount
-0.756957
0.253092
0.871335
0.738639
-2.561195

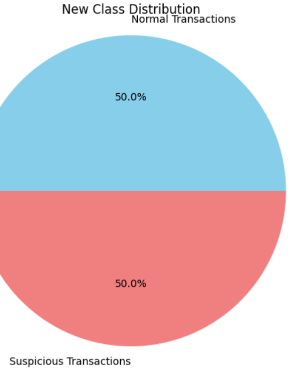
SPLIT TRAIN-TEST DATASET

70% Training Set	30% Testing Set
x-train: (6653396,11)	x-test: (2851456,11)
y-train: (6653396,1)	y-test: (2851456,1)

Handling Class Imbalance

Random Under Sampling





Reducing the majority samples (normal transactions) to match the size of minority class (laundering transactions)

	Tranining Set before RUS	Training Set after RUS
Normal Transactions	6,646,428	6,968
Laundering Transactions	6,968	6,968

05

CONCLUSION

★★★★★

ACHIEVEMENTS

- Achieved the first objective of this project; to perform data preprocessing and Exploratory Data Analysis
- Passed halfway through the project lifecycle; completed Phase 1 until Phase 3
- Dataset has been cleaned and features have been transformed to prepare for model training
- Training dataset has been balanced to avoid overfitting or underfitting during model training.

FUTURE WORK

- Perform hyperparameter tuning to optimize the Support Vector Machines and Decision Tree
- Train the dataset using Support Vector Machines and Decision Tree
- Predict laundering and normal transactions using testing dataset and evaluate model performance
- Visualize findings using Power BI

THANK YOU!