CHAPTER 1

INTRODUCTION

Introduction

According to Bank Negara Malaysia, money laundering is a method to transform ‘dirty’ illegal money into ‘clean’ legitimate appearance. The money may come from criminal activities such as drug trafficking and corruption, thus the offenders need to conceal its unlawful origin before they can luxuriously spend the money. In general, the money laundering process can be summarized into three steps which are placement, layering and integration as per Figure 1 below. Firstly, illegal money is placed into financial institution. Then, the money is transferred multiple times across many layers of accounts. Finally, after going through a cycle of complex transactions to disguise the origin, the money is then integrated back into the economy as lawful funds.

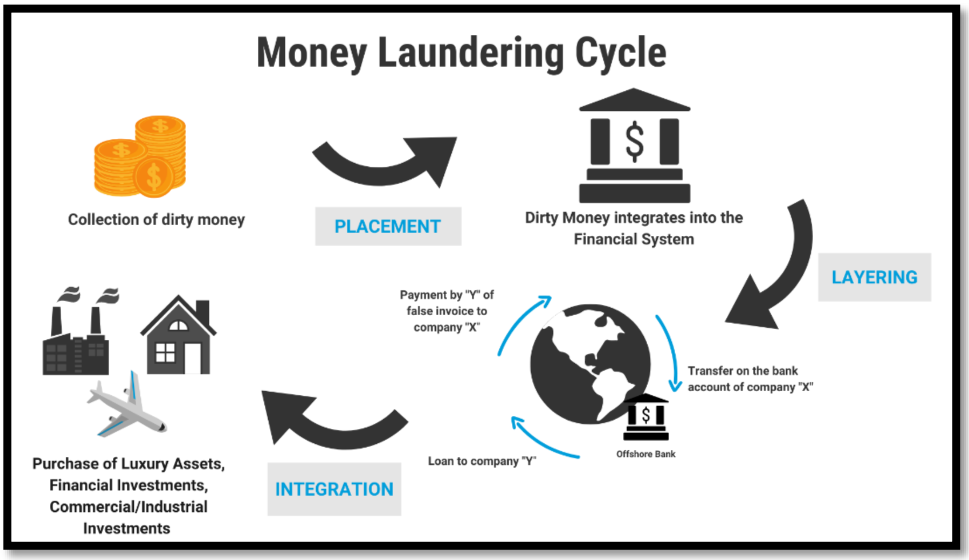

Figure 1: Money Laundering Cycle (United Nations Office on Drugs and Crime)

In Malaysia context, money laundering is a financial crime under the Anti-Money Laundering, Anti-Terrorism Financing and Proceeds of Unlawful Activities Act 2001 (AMLA). Based on

Paragraph 4(1) in AMLA, a money launderer is someone who clearly or subtly involves in an ambiguous transaction that is coming from unlawful proceeds of criminal activities which includes obtaining, receiving, transferring, converting, and concealing the true origin of illicit funds. (AMLA 2001). There is a long list of serious offence stated in the Second Schedule of AMLA comprising of more than 400 criminal offences under 49 different Acts such as Dangerous Drugs Act 1952 [Act 234], Financial Services Act 2013 [Act 758], and Kidnapping Act 1961 [Act 365]. These criminal activities are interrelated with money laundering as the criminal needs to 'clean' their money.

The complexity of money laundering activities encourages the use of machine learning algorithm as a promising approach to effectively combat this serious financial crime and maintaining integrity in the financial institutions. The development of predictive model plays a vital role as it can train on existing laundering scenarios in historical financial transactions to detect patterns, anomalies, and shady behaviours in large datasets. Machine learning algorithms which involve clustering and anomaly analysis contributes to flag suspicious transactions in real-time, thus improving efficiency and reducing reliance on traditional manual processes.

Problem Background

The United Nations Office on Drugs and Crime (UNDOC) estimated that around 2% to 5% of the world's total economic output is laundered globally every year. It brings to the significant amount of $800 billion - $2 trillion 'dirty' money (United Nations Office on Drugs and Crime). As of 2023, Malaysia received 317,435 Suspicious Transaction Report (STR) which was a 31% increase from last year where the reports are mainly on fraud, money laundering and tax offences. More than 100 individuals were arrested and RM290 million was recovered. Furthermore, 59,684 suspected mule accounts were identified and disrupted to hinder the process of disguising the origin of illicit funds vias multiple layering transactions (BNM Annual Report, 2023). In order to effectively combat money laundering activities, this project seeks to explore more advanced approach using machine learning algorithm to learn complex transaction patterns and enhance money laundering detection.

Problem Statement

Malaysia's Anti-Money Laundering and Counter Financing Terrorism (AML/CFT) regime is generally in compliance with Financial Action Task Force (FATF), however in terms of money laundering investigation and prosecution, it is still not showing a significant effectiveness even though the number of investigations is increasing. The total of money laundering prosecutions and convictions is still low, and Malaysia is not adequately targeting high-risk offences especially if it involves cross border transactions (Mutual Evaluation Report, 2015). This is because most financial institutions are using rule-based techniques to detect money laundering activities, but it is not powerful enough to identify the complex and hidden schemes used by criminals, especially in cross border transactions (Oztas et.al, 2023). Hence, there is a need to develop a machine learning approach to combat and stay ahead of sophisticated money laundering methods.

Project Objectives

This project aims to utilize supervised machine learning algorithm to efficiently detect money laundering activities as an effort to maintain financial integrity in Malaysia. The objectives of this project are:

1) To perform data preprocessing and exploratory data analysis (EDA) to handle noisy data and understand data distributions.
2) To implement machine learning algorithms to learn patterns, identify anomaly transactions and detect money laundering activities.
3) To evaluate model using metrics such as True Positive Rate (TPR), False Positive Rate (FPR), True Negative Rate (TNR), False Negative Rate (FNR) and Area Under the Curve (AUC).

Scope of the Project

Since it is difficult to obtain real financial transaction data due to privacy reasons and legal constraints, hence this project is using synthetic money laundering datasets called SAML-D developed by Berkan Oztas and five other researchers in their paper entitled 'Enhancing Anti-Money Laundering: Development of a Synthetic Transaction Monitoring Dataset'. This dataset includes 28 typologies of transactions which is the type of money flow after it is first placed into the financial institution, thus brings greater realism to the dataset. Plus, it also has geographic locations which are important to analyze cross border transactions.

This project will use machine learning algorithms such as Support Vector Machine (SVM) or Random Forest to detect money laundering activities. First, data preprocessing and exploratory data analysis are performed to have a better understanding of the data and discover patterns and indicators for money laundering activities. Next, machine learning algorithm is applied to detect suspicious transactions and flag as money laundering activities. The algorithm will then be evaluated by using metrics such as True Positive Rate (TPR), False Positive Rate (FPR), True Negative Rate (TNR), False Negative Rate (FNR) and Area Under the Curve (AUC).