MONEY LAUNDERING DETECTION OF SUSPICIOUS TRANSACTION USING MACHINE LEARNING ALGORITHM

NUR ADRIANA BATRISYIA BINTI MOHD SUBRI

UNIVERSITI TEKNOLOGI MALAYSIA

CHAPTER 5

CONCLUSION

5.1   Summary

This research proposal aims to achieve the first object of this project which is to perform data preprocessing and Exploratory Data Analysis (EDA) to handle noisy data and understand data distributions. To achieve this objective, it has been successfully passed halfway through this project lifecycle where three out of six phases has been completed. Starting with Phase 1, this project has identified the major challenges that hinders the effort to combat money laundering crimes. Moving on to Phase 2, synthetic transaction dataset that was developed by other researchers has been downloaded and extracted to train the supervised machine learning models as real transaction dataset is difficult to obtain.

The major achievements from this research proposal are highly contributed from the activities in Phase 3 which is data preparation. Data preparation was carried out by performing seven key activities to ensure that the training data used for model development is usable and of high quality to ensure high accuracy of model prediction. The achievements from Phase 3 to this research proposal are:

    i.   Understand the data shape, column names, data types, and various categories of payment types, transaction typologies, countries of bank location, and currencies involved.

    ii.   Transform the data types for certain columns such as sender accounts into string format to make it easier to process for analytics and visualizations.

    iii.   Removing irrelevant columns and validating that there are no missing columns or duplicated data exist in the dataset to provide a cleaned dataset.

    iv.   Performing exploratory data analysis to discover the patterns and indicators for money laundering activities such as the most frequent payment types and high-risk countries.

    v.   Transforming features in raw data to improve the accuracy and efficiency of machine learning models using feature engineering techniques such as log transformation, label encoding, and standard scaling.

    vi.   Splitting the dataset into training and testing set with ratio of 70:30.

    vii.   Handling imbalance data in training dataset using Random Under Sampling to avoid overfitting or underfitting during model training.

5.2   Future Work

Future works for this project is to continue the project lifecycle from Phase 4 until Phase 6 which are to:

    i.   Perform hyperparameter tuning to optimize the Support Vector Machines and Decision Tree for better performance.

    ii.   Train the dataset using Support Vector Machines and Decision Tree to learn the patterns and features in the dataset.

    iii.   Predict laundering or normal transactions based on testing dataset using Support Vector Machines and Decision Tree.

    iv.   Evaluate the machine learning algorithm accuracy using selected performance metrics.

    v.   Visualize findings and insights using Power BI.