Low Level Design (LLD)
Money Laundering Prevention System
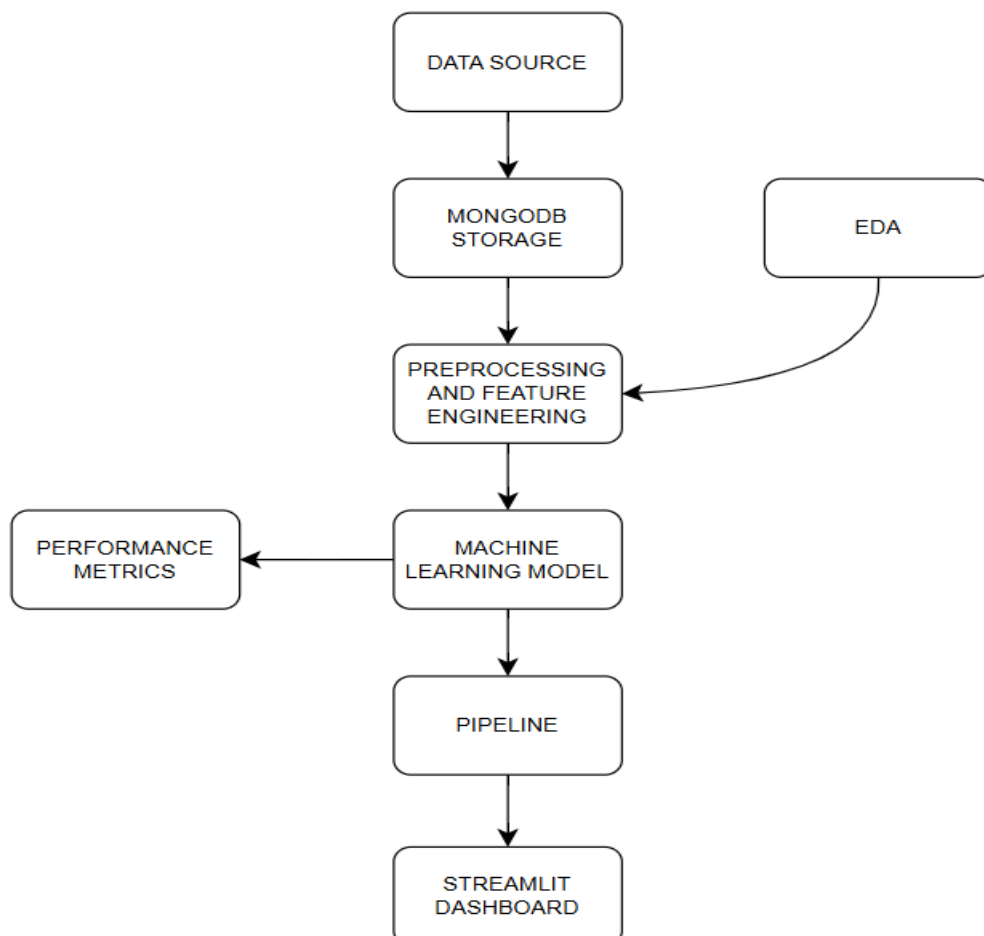
Written By: Mukul Palmia

Table of Contents

1. Problem Statement

The objective of this project is to develop a machine learning model capable of predicting whether a user is involved in money laundering. The model will achieve this by continuously analyzing their money transfer activities and other financial behaviours.

2. System Architecture

The system architecture consists of the following key components:

- Data Ingestion Module

- Data Validation Module

- Data Transformation Module

- Model Training Module

- Model Evaluation Module

- Model Deployment Module

3. Data Ingestion Module

The dataset is taken from google. The module will:

- Upload the data into a database using Python.

- Implement a function to collect/import data from the database and convert it into a DataFrame for further processing.

- Split the preprocessed data into training and testing sets and export them separately for further processing.

4. Data Validation Module

This module ensures data quality by:

- Validating the integrity and accuracy of the ingested data.

- Handling missing and erroneous values.

- Ensuring data consistency and adherence to predefined standards.

5. Data Transformation Module

The Data Transformation Module performs the following tasks:

- Feature engineering to create meaningful features from raw data.

- Data normalization to scale numerical features appropriately.

- Encoding categorical variables to make them suitable for machine learning models.

- Handling class imbalance using the SMOTE (Synthetic Minority Over-sampling Technique) method.

6. Model Training Module

This module is responsible for training the predictive model:

- Various machine learning algorithms such as Random Forest, Gradient Boosting, and Support Vector Machines (SVM) will be explored.

- Hyperparameter tuning will be conducted using GridSearchCV to achieve optimal performance.

7. Model Evaluation Module

The model's performance will be assessed using the following evaluation metrics:

- Accuracy Score

- F1 Score

- Precision Score

- Confusion Matrix

8. Model Deployment Module

This module focuses on deploying the trained model into a production environment:

- The model will be deployed via an API endpoint for real-time predictions.

- Users can submit new transaction data and receive instant predictions on potential money laundering activities.

9. User Interface

The project will feature a user interface where users can input transaction details. The model will analyze the input data and return predictions regarding potential money laundering risks.

10. Model Monitoring and Maintenance

To ensure continuous effectiveness, the deployed model will be:

- Regularly monitored for performance degradation.

- Retrained and updated as needed to maintain high accuracy and reliability.

11. Documentation

- The entire project, including source code, data, and model documentation, will be hosted on GitHub.

12. Conclusion

This document outlines the system architecture, key modules, and workflow for the money laundering prediction model. It serves as a foundational guide for the development, implementation, and maintenance of the project, ensuring a well-structured and systematic approach.