Data Processing Requirements for Money Laundering Prevention

1. Introduction

The document delineates the comprehensive data processing strategy for our anti-money laundering project. It

elaborates on the sequential phases of data ingestion, validation, transformation, and storage, highlighting the pivotal role

these processes play in fostering accurate and reliable analytical predictions.

2. Data Ingestion

Our system is engineered to accommodate data from multifaceted sources, including CSV files, databases,

and real-time API integrations. It ensures data integrity, consistency, and format through meticulous validation processes,

equipped with robust error handling and logging mechanisms to monitor and resolve any data ingestion discrepancies.

3. Data Validation

The system rigorously applies data validation checks to ascertain the accuracy and integrity of the

input data against predefined standards. These checks include, but are not limited to, assessments of missing values, data

type consistency, range validations, and the identification of outliers.

4. Data Transformation

Transformation processes are tailored to optimize data for model training and prediction,

incorporating feature engineering, normalization, encoding of categorical variables, and strategies for handling missing

values. Additionally, data balancing techniques are employed to mitigate over-fitting and bias in machine learning models.

5. Data Storage

MongoDB is utilized as the primary data storage solution for this project, offering scalable, secure, and efficient data management capabilities to support the system's requirements.

6. Conclusion

This document encapsulates the essential methodologies and procedures for effective data processing within the anti-money laundering project framework. It serves as a guiding document for the development team, ensuring adherence to best practices in data handling, privacy, security, and monitoring to achieve reliable system performance.