Data Processing Requirements (DPR) Document

Credit Card Default Prediction

Table of Contents

1. Introduction

The Data Processing Specifications (DPS) document delineates the project's data processing pipeline and associated prerequisites. It articulates the sequential processes encompassing data intake, validation, transformation, and storage, all aimed at guaranteeing the dependability and precision of predictions.

2. Data Ingestion

- The system should support data ingestion from various sources, such as CSV files, databases, or real-time API integration.
- The data ingestion process should include mechanisms for validating the data integrity, consistency, and format.
- Error handling and logging should be implemented to track any issues or anomalies during the data ingestion phase.

3. Data Validation

- System perform data validation checks to ensure the correctness and quality of input data which were pre-defined requirements.
- Common validation checks include checking for missing values, data type consistency, range validation, and outlier detection.

4. Data Transformation

- Data transformation step prepare the data for model training and prediction.
- This involve feature engineering, normalization, encoding categorical variables, and handling missing values.
- Performed data up-sampling on target feature to balance the data and ensure that the data doesn't create over-fi ing or bias to ML model.

5. Conclusion

The Data Processing Requirements (DPR) document delineates the essential factors and procedural steps involved in data processing for the project. It encompasses data ingestion, validation, transformation, storage, privacy, security, and monitoring aspects. This document serves as a crucial reference for the development team, ensuring the efficient and reliable processing of data within the system.