Project Design Phase I
Solution Architecture

Date	01 November 2023
Team ID	611609-1699536697
Project Name	Online Payments Fraud Detection Using ML
Maximum Marks	4 Marks

Solution Architecture:

Data Collection:

The initial phase involves gathering data from diverse sources such as user transactions, login activities, device details, IP addresses, and external providers offering geolocation and historical fraud data. This information can be collected through real-time streaming for immediate analysis and batch processing for historical data.

Data Preprocessing:

Following data collection, the next step is preprocessing. This involves cleaning and transforming the data, addressing issues like missing values, outliers, and inconsistencies. Categorical variables are converted into numerical representations, and numerical features are normalized or scaled to ensure uniformity.

Split Data:

To effectively evaluate models, the collected data is split into training and testing sets. This ensures that both sets contain a representative distribution of normal and fraudulent transactions, allowing for a thorough assessment of model performance.

Model Selection:

Choosing the appropriate model is crucial. Decision trees, random forests, gradient boosting, and neural networks are among the options. Selection is based on criteria such as scalability, interpretability, and accuracy.

Model Building:

The selected models are integrated into a real-time scoring engine and a rules engine. This combination of machine learning and rule-based systems enhances the overall fraud detection capabilities.

Model Training:

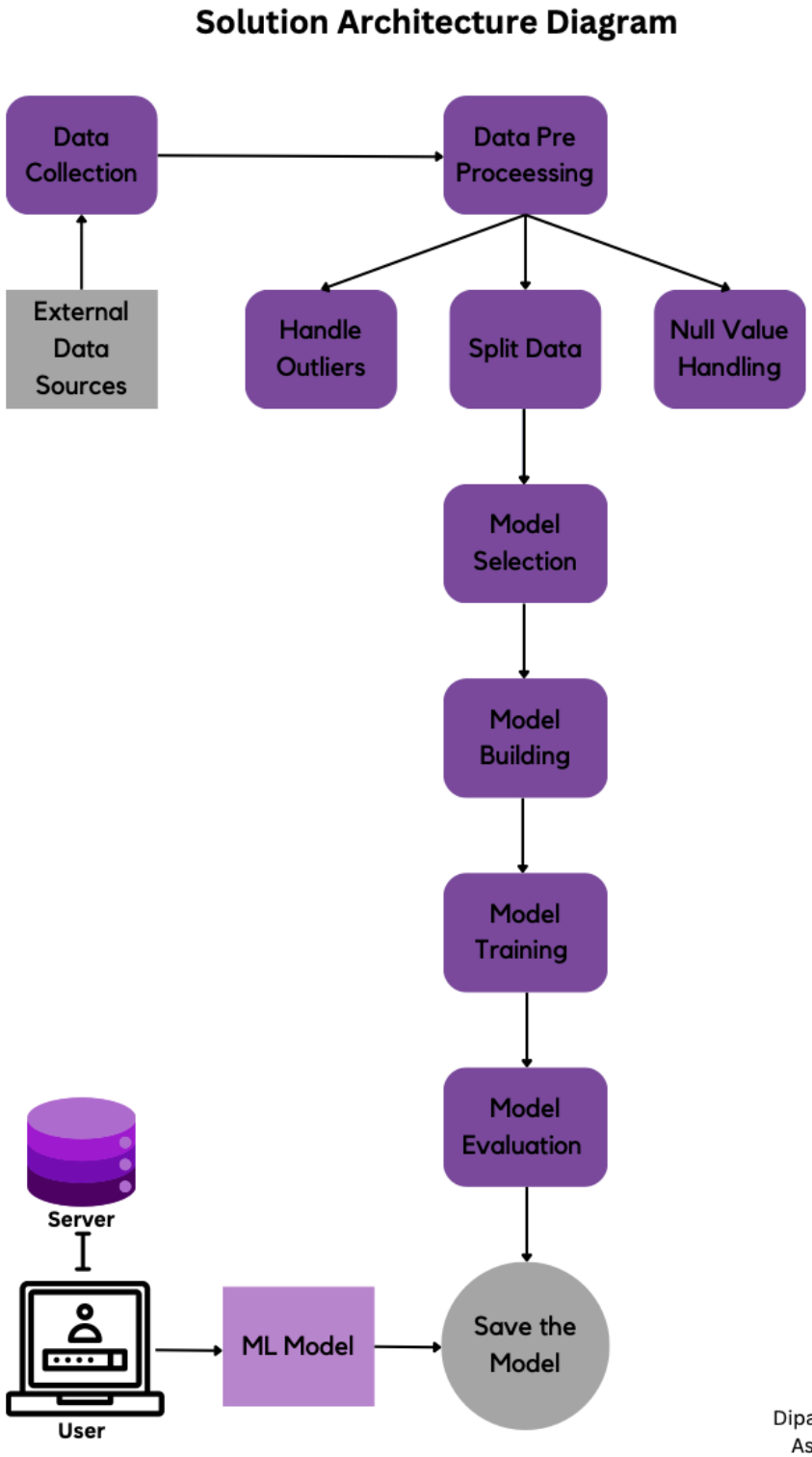
Using historical data, the chosen models undergo training. Ensemble methods, such as combining predictions from multiple models, are employed to improve accuracy and effectiveness.

Model Evaluation:

Once trained, models are evaluated using metrics such as precision, recall, F1 score, and ROC-AUC. Techniques like cross-validation and hyperparameter tuning contribute to optimizing model performance.

Save the Model:

The trained models are saved in a repository or cloud-based storage, with version control implemented for easy updates and rollbacks.



TEAM
Dipanshu Mandal
Ashlin Binshu
Arnab Ghosh