

Solution Approach:

1. **Understanding and Exploring the Data:**

- Begin by gaining a clear understanding of the dataset and exploring the various features.

2. **Data Cleaning:**

- Handle missing values and treat outliers to ensure the dataset is clean and reliable.

3. **Exploratory Data Analysis (EDA):**

- Conduct univariate analysis to examine individual features.
- Perform bivariate analysis to explore relationships between variables.

4. **Data Preparation for Modeling:**

- Check for skewness in the data and address it to ensure fair analysis.
- Manage the data imbalance, especially considering that only 0.172% of the records are fraudulent transactions.

5. **Data Splitting:**

- Split the dataset into training and testing sets.
- Apply scaling techniques (such as normalization) to standardize the data.

6. **Model Building:**

- Train the model using various algorithms such as Logistic Regression, SVM, Decision Trees, Random Forest, and XGBoost.
- Fine-tune the hyperparameters using Grid Search with Cross Validation to find the optimal configuration.

7. **Model Evaluation:**

- Given the significant class imbalance, accuracy may not be the best metric for this problem.
- Focus on finding a balance between Precision and Recall instead of purely relying on accuracy.
- Evaluate the model based on ROC curve performance, aiming for a high True Positive Rate (TPR) and a low False Positive Rate (FPR) to reduce misclassifications.