Solution Approach

1. Data Understanding and Exploration

- Gain insights into the dataset and identify key features.

2. Data Cleaning

- Address missing values.
- Handle outliers.

3. Exploratory Data Analysis (EDA)

- Conduct univariate analysis.
- Perform bivariate analysis.

4. Data Preparation for Modeling

- Assess and correct data skewness to ensure accurate analysis.
- Address data imbalance, noting that only 0.172% of transactions are fraudulent.

5. Data Splitting

- Divide the dataset into training and testing sets.
- Normalize the data.

6. Model Building

- Train various models such as Logistic Regression, SVM, Decision Tree, Random Forest, and XGBoost.
- Optimize hyperparameters using Grid Search Cross-Validation to find the best parameters.

7. Model Evaluation

- Given the data imbalance, accuracy is not the ideal metric.
- Focus on balancing Precision and Recall.
- Aim for a high ROC score with a high True Positive Rate (TPR) and low False Positive Rate (FPR) to minimize misclassifications.