**Solution Approach**

1. **Problem Statement**

Credit card fraud is a significant challenge in the financial sector, involving unauthorized use of credit card details for purchases or withdrawals. The goal of this project is to develop a classification model that accurately predicts fraudulent transactions to protect customers and minimize financial losses. The dataset is highly imbalanced, with fraudulent transactions making up only 0.172% of the total, which introduces challenges in model development and evaluation.

1. **Data Sources**

The dataset contains credit card transactions made by European cardholders in September 2013, provided for research purposes. The data includes:

* Time: Time elapsed since the first transaction in seconds.
* Amount: Transaction amount.
* V1 to V28: Features derived using Principal Component Analysis (PCA) to ensure anonymity.
* Class: Target variable indicating fraud (1) or legitimate (0).

1. Data Processing
   1. Data understanding and exploring
   2. Data Cleaning
      1. Handling missing value.
      2. Outlier treatment.
   * Exploratory data analysis
     1. Univariate analysis.
     2. Bivariate analysis.
2. Prepare the data for modelling
   1. Check the skewness of the data and mitigate it for fair analysis.
   2. Handling data imbalance as we see only 0.17% records are the fraud transaction.
3. Split the data into train and test
   1. Scale the data (normalization)
4. Model building
   1. Train the model with various algorithm such as logistic regression, SVM, Decision Tree, Random Forest, XGBoost etc.
   2. Tune the hyperparameter with Grid Search Cross Validation and find the optimal values of the hyperparameters.
5. Model evaluation
   1. As we see the data is heavily imbalanced, Accuracy may not be the correct measure for this particular case.
   2. We have to look for a balance between Precision and Recall over Accuracy.
   3. We also have to find out the good ROC score with high TPR and low FPR in order to get the lower number of misclassifications.
6. Project Outcomes
   1. Expected Results:
      1. A classification model capable of accurately detecting fraudulent transactions.
      2. Minimized false negatives (fraudulent transactions classified as legitimate).
      3. Insights into features contributing most to classification decisions.
7. Conclusion

This project demonstrates how to handle imbalanced datasets, preprocess data, and select evaluation metrics for fraud detection. The results can inform real-world fraud detection systems, helping financial institutions mitigate risks effectively.