**FRAUD DETECTION CASE STUDY**

*Steps to be followed*

**Notebook 1: Exploratory Data Analysis (EDA)**

**Objective**: Perform Exploratory Data Analysis on the provided dataset to gain insights and understand the data's characteristics.

**Steps recommended:**

1. **Introduction**:
   1. Provide an overview of the problem statement (fraud detection) and mention the dataset's source and context.
   2. Briefly describe the data dictionary to give context about the features.
2. **Loading Data**:
   1. Load the provided dataset into a pandas Data Frame.
   2. Perform basic checks like the number of rows and columns, data types, and missing values,etc.
3. **Data Exploration**:
   1. Explore the distribution of the target variable (fraud vs. non-fraud) to understand class imbalance.
   2. Explore individual feature distributions (numerical and categorical) using appropriate visualizations (histograms, box plots, etc.).
   3. Analyze correlations between features to identify potential relationships.
4. **Train-Test Split**:
   1. Split the data into training and testing sets, adhering to the mentioned train-test split ratio.
   2. Explain the reason for the chosen split ratio and its importance in model evaluation.
5. **Summary**:
   1. Summarize key findings from the EDA, including any insights about the data's characteristics and potential challenges in fraud detection.

**Notebook 2: Feature Engineering**

**Objective**: Transform and engineer features to improve the model's performance and predictive power.

**Steps recommended:**

1. **Feature Transformation**:
   1. Handle missing values (if any) using appropriate techniques like imputation or removal based on justification.
   2. Encode categorical variables using suitable encoding methods (label encoding, one-hot encoding) and explain the choice for each variable.
   3. Scale numerical features using techniques like Min-Max scaling or Standardization, etc.
2. **Feature Creation**:
   1. Create new features from existing ones if it adds value to the model (e.g., aggregations, interaction terms).
   2. Justify the creation of each new feature based on its relevance to fraud detection.
3. **Feature Selection**:
   1. Use techniques like correlation, feature importance from tree-based models, or other relevant methods to select the most important features.
   2. Explain the rationale behind selecting specific features for the model.
4. **Summary**:
   1. Summarize the feature engineering steps and the rationale behind each transformation and creation. Discuss the impact of feature engineering on the model's performance.

**Notebook 3: Model Development and Evaluation**

**Objective**: Develop and evaluate machine learning models for fraud detection using the processed data.

**Steps recommended:**

1. **Model Selection**:
   1. Choose appropriate machine learning algorithms (e.g., Logistic Regression, Random Forest, Gradient Boosting) based on the problem's nature.
   2. Justify the choice of models and explain why they are suitable for fraud detection.
2. **Model Training**:
   1. Train the selected models using the processed training data.
   2. Tune hyperparameters using techniques like Grid Search or Random Search to optimize model performance.
   3. Document the hyperparameters used and the reason behind their selection.
3. **Model Evaluation**:
   1. Evaluate models on the test dataset using relevant metrics (accuracy, precision, recall, F1-score, ROC AUC) as mentioned in the problem statement.
   2. Discuss the choice of evaluation metrics and their relevance to fraud detection.
   3. Provide detailed explanations of the results and compare the performance of different models.
4. **Conclusion**:
   1. Summarize the entire process, including EDA, feature engineering, and model development.
   2. Discuss challenges faced, lessons learned, and potential future improvements.

**Please note**: The above mentioned steps are to provide a guideline during solution development.