Section A: Theoretical Questions

These topics are focused on fundamental machine learning concepts and algorithms.

1. **Evaluation Metrics & Performance Measurement**
    - Precision, Recall, and F1 Score (Definition, Formula, Calculation)
    - Confusion Matrix (All derived metrics)
    - ROC Curve (Explanation and Interpretation)
    - Cohen's Kappa Value (Interpretation)
    - Bias and Variance Tradeoff
2. **Supervised Learning Algorithms**
    - Decision Tree (Working, Advantages, Disadvantages, Comparison with Random Forest)
    - Random Forest (Working, Advantages, Disadvantages)
    - Logistic Regression (Assumptions)
    - Naive Bayes (Assumptions)
    - K-Nearest Neighbors (Explanation with an example)
    - Stacking Classifier (Explanation and Advantages)
3. **Ensemble Learning Techniques**
    - Bagging vs. Boosting (Differences)
    - AdaBoost & Gradient Boosting (Differences and Advantages)
    - Ensemble Learning (Concept and Explanation)

Section B: Exploratory Data Analysis (EDA) and Data Preprocessing

This section focuses on handling datasets, checking for defects, and making data ready for modeling.

1. **Data Understanding**
    - Reading datasets and summarizing observations
    - Identifying numerical and categorical variables
    - Five-point summary for numerical variables
    - Descriptive statistics for categorical variables
2. **Handling Data Issues**
    - Identifying and fixing missing values
    - Detecting and handling outliers using Z-score and plots

- Checking and addressing class imbalance in the target variable
- Identifying textual data and converting it appropriately (Encoding)

3. **Feature Engineering & Data Transformation**
    - Encoding categorical features (One-hot encoding, Label encoding)
    - Correlation analysis and feature selection
    - Creating additional features if required
    - Removing unnecessary features
4. **Train-Test Split**
    - Splitting data into training and testing sets (70:30 split)

Section C: Model Building, Tuning & Evaluation

This section covers model selection, performance improvement, and business interpretation.

1. **Building a Base Model**
    - Choosing an initial model and explaining the choice
    - Calculating and interpreting Accuracy, Precision, Recall, F1 Score
    - Evaluating feature importance
    - Interpreting Cohen's Kappa Value
2. **Model Improvement Strategies**
    - Changes made to improve accuracy
    - Hyperparameter tuning (Using GridSearchCV)
    - Evaluating feature impact on model performance
    - Transforming features for better performance
3. **Final Model Evaluation**
    - Comparing performance before and after improvements
    - Evaluating F1 Score and ROC Score for the final model
    - Business interpretation of results (Which features impact predictions)
    - Discussing risks and limitations of the model