

# Credit Risk Assessment using Tree-based Models

## Task 1: Understanding Decision Tree Fundamentals

Objective: Develop a solid understanding of decision tree components and structure.

Instructions:

1. Draw a basic decision tree structure for credit risk assessment using: income level, credit score, employment duration, existing debt
2. Label all components (root node, internal nodes, leaf nodes, branches)
3. Explain the splitting criteria at each node
4. Document why each feature was chosen for different levels

Expected Outcome: A well-labeled decision tree diagram with written explanations for each component and decision point.

## Task 2: Data Preparation and Initial Analysis

Objective: Prepare and analyze credit risk dataset for decision tree modeling.

Instructions:

1. Load the credit risk dataset and perform: display first 10 rows, calculate basic statistics, identify missing values and outliers
2. Create visualizations showing: distribution of risk categories, relationship between income and default risk, correlation matrix
3. Write a summary of key findings

Expected Outcome: Comprehensive data analysis report with visualizations and insights.

## Task 3: Implementation of Gini Index and Entropy

Objective: Understand and implement different splitting criteria.

Instructions:

1. Implement two decision trees using Gini index and Entropy as splitting criteria
2. Calculate and compare: node impurity at each split, information gain for each feature
3. Document which criterion performs better for credit risk assessment and why

Expected Outcome: Comparative analysis of Gini index vs. Entropy with supporting calculations.

#### **Task 4: Model Development and Tuning**

Objective: Build and optimize a decision tree model for credit risk prediction.

Instructions:

1. Split data into training (70%) and testing (30%) sets
2. Create initial decision tree model with default parameters
3. Perform hyperparameter tuning: test different `max_depth` values, adjust `min_samples_split`, modify `min_samples_leaf`
4. Document the impact of each parameter on model performance

Expected Outcome: Optimized decision tree model with documented tuning process.

#### **Task 5: Model Evaluation and Performance Analysis**

Objective: Evaluate model performance and handle class imbalance.

Instructions:

1. Calculate and analyze: confusion matrix, precision/recall/F1-score, ROC curve and AUC score
2. Implement techniques to handle class imbalance: use class weights, apply SMOTE, adjust decision threshold
3. Compare performance metrics before and after balancing

Expected Outcome: Comprehensive model evaluation report with balanced class handling.

#### **Task 6: Pruning and Model Improvement**

Objective: Implement pruning techniques to prevent overfitting.

Instructions:

1. Apply pre-pruning techniques: set maximum depth, define minimum samples per leaf, establish minimum impurity decrease
2. Implement post-pruning: use cost complexity pruning, create pruning path visualization
3. Compare model performance before and after pruning

Expected Outcome: Analysis of pruning impact on model performance with visualizations.