Assignment 2: Credit Card Dataset Analysis

In this assignment, you are required to analyze the credit card dataset using different sampling techniques and evaluate the performance of machine learning classifiers. Follow the steps below carefully and document all findings in a detailed report.

Task 1: Dataset Preparation

1. Read the credit card dataset.

2. Print the ratio for each class in the original dataset.

Task 2: Sampling Techniques

Create three samples from the dataset, each containing 50,000 samples, by applying the following sampling techniques:

1. Random Sampling

2. Stratified Sampling

3. Clustered Sampling

For each of the three sample datasets, perform the following tasks:

Task 3: Sample Analysis

1. Print the ratio for each class in the sampled dataset.

2. Select the most important features using Wrapper Forward Feature Selection.

3. Create a new dataset using only the selected features.

4. Split the new dataset into training and testing sets using a test size of 0.33.

5. Train and test the following classifiers on the dataset:

- Decision Tree

- K-Nearest Neighbors (KNN) with k = 7

6. Evaluate each model's performance using the following metrics:

- Accuracy

- Precision for each class

- Recall for each class

- F-measure (weighted)

Task 4: Report Findings

Prepare a comprehensive report summarizing your findings. Include the following details for each sampling technique:

1. Class Ratios in the sampled dataset.

2. Selected Features for each sample using forward feature selection.

3. Evaluation Metrics for each classifier:
   - Accuracy
   - Precision
   - Recall
   - F-measure (weighted)

Sampling Technique	Class Ratios	Selected Features	Accuracy	Decision Tree			KNN		
				Precision	Recall	F-measure	Precision	Recall	F-measure

Task 5: Identify Best Sampling Technique

In the conclusion of your report, answer the following questions:

1. Which sampling technique produced the best results? Justify your answer using the evaluation metrics.

2. List all the features selected for each sampling technique.

3. Summarize the performance of each classifier for each sample.

Submission Requirements

- Submit the final report as a PDF or Word document.

- Include all source code used for analysis.

- Provide well-organized tables and charts (if applicable) to support your findings.

Grading Criteria

Your assignment will be graded based on the following:

- Correct implementation of all tasks.

- Accuracy and clarity of results.

- Depth of analysis in the report.

- Presentation and formatting of the report.