**Assignment Code: DA-AG-014**

# Ensemble Learning | **Assignment**

**Instructions:** Carefully read each question. Use Google Docs, Microsoft Word, or a similar tool to create a document where you type out each question along with its answer. Save the document as a PDF, and then upload it to the LMS. Please do not zip or archive the files before uploading them. Each question carries 20 marks.

**Total Marks**: 200

**Question 1:**  What is Ensemble Learning in machine learning? Explain the key idea behind it.

**Answer:**

**Question 2:** What is the difference between Bagging and Boosting?

**Answer:**

**Question 3:** What is bootstrap sampling and what role does it play in Bagging methods like Random Forest?

**Answer:**

**Question 4:** What are Out-of-Bag (OOB) samples and how is OOB score used to evaluate ensemble models?

**Answer:**

**Question 5:** Compare feature importance analysis in a single Decision Tree vs. a Random Forest.

**Answer:**

**Question 6:** Write a Python program to:

- Load the Breast Cancer dataset using
  `sklearn.datasets.load_breast_cancer()`
- Train a Random Forest Classifier
- Print the top 5 most important features based on feature importance scores.

(*Include your Python code and output in the code box below.*)

**Answer:**

**Question 7**: Write a Python program to:

- Train a Bagging Classifier using Decision Trees on the Iris dataset
- Evaluate its accuracy and compare with a single Decision Tree

(*Include your Python code and output in the code box below.*)

**Answer:**

**Question 8**: Write a Python program to:

- Train a Random Forest Classifier
- Tune hyperparameters `max_depth` and `n_estimators` using GridSearchCV
- Print the best parameters and final accuracy

(*Include your Python code and output in the code box below.*)

**Answer:**

```



```

**Question 9**: Write a Python program to:

- Train a Bagging Regressor and a Random Forest Regressor on the California Housing dataset
- Compare their Mean Squared Errors (MSE)

(*Include your Python code and output in the code box below.*)

**Answer:**

```



```

**Question 10:** You are working as a data scientist at a financial institution to predict loan default. You have access to customer demographic and transaction history data.

You decide to use ensemble techniques to increase model performance.

Explain your step-by-step approach to:

- Choose between Bagging or Boosting
- Handle overfitting
- Select base models
- Evaluate performance using cross-validation
- Justify how ensemble learning improves decision-making in this real-world context.

(*Include your Python code and output in the code box below.*)

**Answer:**