**ASSIGNMENT 4**

**Possible Points: 100 Deadline: April 9th, 2025**

Assignments should be done on individual basis.

Download the dataset **heart.csv** provided and perform the following tasks.

**Question 1 (50 Points)**

Use the heart.csv dataset to check whether the person has heart disease or not. Compare the performances based on the models below:

* Check for null values in the dataset if present remove them. Convert any categorical values to numerical values. Split the data into 70-30 training and testing. **(10 Points)**
* Train the data on Random Forest, Gradient Boosting, AdaBoost and XGBoost. Use n\_estimators as 50 for all the models. **(15 Points)**
* Evaluate the models based on accuracy, F-1 score, precision and recall. Explain which model performs best and why. **(15 Points)**
* Check for feature importance of all the models. Determine which features showcase more importance and which show least. **(10 Points)**

**Question 2 (50 Points)**

Use the heart.csv dataset. Use the attributes Age, Sex, ChestPainType, RestingBP, RestingECG and ExerciseAngina only. Compare the performances of models below:

* Check for null values in the dataset if present remove them. Convert any categorical values to numerical values. Split the data into 60-40 training and testing. **(10 Points)**
* Train the data on KNN, Decision Trees and Logistic regression individually. **(15 Points)**
* Train the data on stacking model (Final estimator as logistic regression) and voting classifier (voting type hard) based on the above three models. **(10 Points)**
* Evaluate the models based on accuracy, F-1 score, precision and recall. Explain which model performs best and why. **(10 Points)**
* Discuss whether stacking improves the predictive accuracy compared to individual models. **(5 Points)**