**Capstone Project 2 Details**

**1. Project Title:**  
Heart Disease Detection using Classification Algorithms

**2. Category:**  
Supervised Learning (Classification)

**3. Problem Statement:**

Heart disease is one of the leading causes of death worldwide. Early and accurate detection through diagnostic screening can significantly improve patient survival rates and reduce healthcare costs.

This project aims to build machine learning models that predict the presence of heart disease based on patients’ demographic, clinical, and diagnostic test data, enabling healthcare professionals to make timely and informed decisions.

**4. Objective:**

* Develop and compare multiple classification algorithms (Decision Tree, Random Forest, Logistic Regression, SVM) to predict heart disease.
* Present actionable findings to stakeholders in the healthcare domain.

**5. Dataset Information:**

* **Dataset Name/Source:** heart\_disease\_dataset
* **Size & Features:**
  + No of records: 400
  + Target variable: heart\_disease (0 = No, 1 = Yes)
  + Features: 13 Features + 1 ClassLabel ( heart\_desease)
  + Features are Age, Sex, Chest Pain Type, Resting Blood Pressure, Cholesterol, Fasting Blood Sugar, Resting ECG, Max Heart Rate, Exercise-Induced Angina, ST Depression, ST Slope, Num Major Vessels, Thalassemia
* **Pre-processing Required:**
  + There are no missing values in the dataset, so handling missing values is not required. The data is complete and ready for analysis
  + Data is already in numerical format, no encoding techniques are required for categorical variables
  + Split data into training (80%) and testing (20%) sets.

**6. Tools & Technologies:**

Python, Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, SciPy, FastAPI (for deployment), Docker (for containerization)

**7. Evaluation Criteria:**

* **Primary Metrics:** Recall (Sensitivity), Precision, F1-score, ROC-AUC
* **Secondary Metrics:** Specificity, Accuracy
* Model comparison report and recommendations
* Quality of EDA and feature analysis
* Clarity and professionalism of presentation and documentation

**8. Expected Deliverables:**

* **Code Files:** Python scripts/notebooks implementing EDA, model training and evaluation
* **Documentation/Report:** Detailed project report with insights, methodology, and results
* **Presentation Slides:** No of slides 15-20

**9. Timeline:**

Day 1: Data exploration, EDA, preprocessing

Day 2: Baseline model development & evaluation

Day 3: Model comparison, final evaluation, Deployment (FastAPI + Docker),

Day 4: final report & presentation