Assignment 7

Objective: Develop machine learning models to predict the income status of customers based on their spending score and demographic information.

Dataset: Use the 'Adult Income' dataset from the UCI Machine Learning Repository. This dataset is more comprehensive and includes variables such as age, workclass, education, occupation, relationship, race, sex, capital gain/loss, hours per week, and income. It provides a broader perspective for modeling income. The dataset is available at UCI Machine Learning Repository – Adult Income Dataset.

Tasks:

1. Data Preparation:
    - Load and preprocess the dataset. Handle missing values, encode categorical variables, and normalize the data as needed.
    - Split the dataset into training and test sets.
2. Model Implementation:
    - Implement at least three different machine learning models (e.g., linear regression, decision tree, random forest, logistic regression, SVM).
    - You can treat income as a continuous, binary (e.g., '>50K' or '<=50K'), or ordinal variable.
3. Model Tuning:
    - Tune the hyperparameters of each model for optimal performance.
    - Use cross-validation to ensure robustness.
4. Model Evaluation:
    - Evaluate the models using appropriate metrics (e.g., MAE, RMSE for continuous; accuracy, F1-score for classification).
    - Compare the performance of the models.
5. Interpretation and Reporting:
    - Briefly interpret the results of the models. Discuss which model performed best and why.
    - Prepare a one-page report summarizing your findings, methodology, and insights.
6. Coding and Documentation:
    - Ensure the code is well-documented and legible. Include comments and explanations to make it understandable.

Submission: A one-page written report in PDF format detailing the results and a Jupyter Notebook containing the code.