Part 1: Multilinear Regression & Polynomial Regression

Dataset: Student Performance Dataset:

https://www.kaggle.com/datasets/spscientist/students-performance-in-exams

Objective: Students will preprocess the data, apply Multilinear and Polynomial Regression models, and compare performance metrics like RMSE and R^2 score.

Assignment Instructions:

1. **Data Preprocessing - 4 marks**
    - Load the dataset.
    - Handle missing values (if any).
    - Encode categorical variables using one-hot encoding or label encoding.
    - Normalize or standardize features (optional but recommended).
2. **Feature Engineering - 4 marks**
    - Choose relevant features to predict math score.
    - Create interaction or polynomial features (only for Polynomial Regression part).
3. **Model Building - 4 marks**
    - Split the data into training and test sets (80/20).
    - Train two models:
        - Multilinear Regression
        - Polynomial Regression (degree 2 or 3)
4. **Model Evaluation - 5 marks**
    - Use metrics: RMSE, MAE, and R^2 score.
    - Visualize predicted vs actual values.

- Discuss overfitting or underfitting in Polynomial Regression.

5. **Submission Requirements - 3 marks**
    - Jupyter Notebook with code and markdown explanations.

Part 2: Naive Bayes, KNN, and Decision Tree

Dataset: Social Network Ads Dataset (from Kaggle)

https://www.kaggle.com/datasets/rakeshrau/social-network-ads

Objective: Students will predict whether a user purchases a product based on their age and estimated salary using Naive Bayes, KNN, and Decision Tree classifiers.

Assignment Instructions:

1. **Data Preprocessing - 4 marks**
    - Load the dataset (usually named Social_Network_Ads.csv).
    - Drop the User ID column (not a useful feature).
    - Encode the Gender column using Label Encoding or One-Hot Encoding.
    - Encode the target variable Purchased.
    - Standardize the feature columns (Age, EstimatedSalary).
2. **Model Building - 5 marks**
    - Split the dataset into training and test sets (75/25).
    - Train the following models:
        - Gaussian Naive Bayes
        - K-Nearest Neighbors (test with k=3, 5, and 7)

        - Decision Tree (try both Gini and Entropy as criteria)

3. **Model Evaluation - 8 marks**
    - Evaluate all models using:
        - Accuracy
        - Precision
        - Recall
        - F1-Score
        - Confusion Matrix
    - Plot decision boundaries (optional but encouraged for 2D features).
    - Compare models and identify which one performs best and why.
4. **Submission Requirements - 3 marks**
    - Submit a well-documented Jupyter Notebook (.ipynb).
    - Include comments, markdown cells, and graphs.
    - Include a 1-page summary analyzing the performance of all 3 models

Deliverables:

Only GitHub repository link will be accepted, containing both.ipynb notebooks with markdown explanations.