Machine Learning Assignment: Smart Manufacturing

Objective

Build a regression model to predict a key manufacturing metric (e.g., product quality score, energy consumption, machine wear level, or production output) using sensor and operational data from a smart manufacturing environment.

Dataset

You can use either real/Kaggle dataset /simulated dataset or to create one with features such as:

Temperature (°C)

Vibration level (units)

Motor speed (RPM)

Humidity (%)

Power consumption (kW)

Time of operation (hours)

Target variable: e.g., Product Quality Score or Energy Consumption.

Assignment Tasks

1. Data Loading and Exploration

Load the dataset into Python (CSV format).  
Perform exploratory data analysis (EDA): summary statistics, check for missing values, visualize distributions, and correlation matrix.  
Understand the relationships between features and the target variable.

2. Data Preprocessing

Handle missing data if any.  
Split the data into training and test sets (e.g., 80/20 split).  
Normalize or scale features if needed.

3. Model Building

Train a linear regression model using scikit-learn.  
Train a more advanced regression model (e.g., Random Forest Regressor/other if covered in class).

4. Model Evaluation

Evaluate model performance using metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and R² score.  
Compare the performance of different models.

5. Interpretation and Insights

Identify the most important features affecting the target variable.  
Discuss how the model predictions can help optimize manufacturing operations (e.g., reducing defects, saving energy).

Optional Bonus

Visualize actual vs predicted values.

Try hyperparameter tuning for the Random Forest model.

Discuss limitations of the models.

Deliverables

Python code (.ipynb or .py) with clear comments and dataset.  
A short report (1-2 pages) summarizing data analysis, model results, and manufacturing insights.