Assignment – Group Work

Name: Gulam Ali Khorajiya
Course: AI for Manufacturing
Date of Submission: 15-03-2025

Week 4 Assignment: Data Handling and Predictive Modeling

Introduction

This assignment involves analyzing a dataset from a manufacturing unit, performing data cleaning, exploratory data analysis (EDA), and developing a predictive model to estimate the time required for new product manufacturing.

Step 1: Exploratory Data Analysis (EDA)

1.1 Data Inspection

- Load the dataset and check for missing values, data types, and inconsistencies.
- Identify key features relevant to the problem.
- Summarize numerical and categorical variables.

1.2 Data Visualization

- Use histograms, box plots, and scatter plots to understand data distribution.
- Identify patterns, trends, and outliers affecting the prediction model.
- Correlation analysis between variables.

Step 2: Data Cleaning and Preprocessing

2.1 Handling Missing Values

- Use mean/median/mode for numerical variables.
- Use the most frequent category or create an "Unknown" category for categorical data.

2.2 Outlier Detection and Treatment

- Use Z-score or IQR method to detect and remove or cap outliers.

2.3 Encoding Categorical Variables

- Convert categorical variables into numerical values using One-Hot Encoding or Label Encoding.

2.4 Splitting Data

- Divide the dataset into training (80%) and testing (20%) sets to validate the model.

Step 3: Feature Engineering

- Select relevant features using correlation analysis and domain knowledge.
- Create new meaningful features, if applicable.
- Standardize or normalize data if needed.

Step 4: Model Building

4.1 Selecting the Model

- Consider models like Linear Regression, Decision Trees, or Random Forest.
- Compare model performances using evaluation metrics.

4.2 Model Training and Evaluation

- Train the model on the training dataset.
- Evaluate performance using RMSE, MAE, and R-squared.
- Optimize hyperparameters to improve accuracy.

Step 5: Conclusion

- Summarize key findings from EDA and model performance.
- Discuss any limitations and potential improvements.
- Provide recommendations for future enhancements.
