**Introduction:**

This assignment is aimed at understanding and implementing different classification and regression techniques to the dataset provided to us. The model needs to be trained using Logistic Regression and Naïve Bayes Classification, along with the EDA, model performance evaluation, make predictions and analyze the results and draw conclusions. The AI4I 2020 is a synthetic dataset that reflects real predictive maintenance data encountered in industry. It contains 10,000 data points stored as rows with 14 features in columns. Machine failure in the dataset has 5 independent failure modes.

**EDA:**

Exploratory Data Analysis (EDA) is carried out on the dataset to understand the data better by checking for missing or duplicate values in the dataset.

We check for duplicate values and drop them, convert the categorical columns to integer values as required. We then drop the UDI and product ID as it isn’t necessary for model training, check for missing values and find out the unique values for each column.

Following is the plot for the correlation matrix:

Chart

Description automatically generated

**Model Implementation and evaluation:**

**Conclusion:**