MVP: PREDICTIVE MAINTENANCE DATASET

WORK DONE:

1. **Exploratory Data Analysis (EDA):**

- EDA involves exploring and understanding the dataset before modeling. It typically includes summarizing the main characteristics of the data, checking for patterns, and identifying potential challenges or interesting features.

2**. Data Cleaning:**

- Removing Null Values: Addressing missing data by either imputation or removing rows/columns with null values, ensuring data completeness.
- Removing Duplicate Values: Identifying and eliminating duplicated rows to maintain dataset integrity.
- Handling Entry Errors: Correcting any errors or inconsistencies in the data, ensuring accurate and reliable information.
- Outlier Removal: Identifying and addressing outliers that might skew the analysis or modeling results.

3**. Data Visualization:**

- Distribution Plot: Visualizing the distribution of variables using plots like histograms or kernel density plots to understand the spread and shape of the data.
- Correlation Plot: Displaying correlations between different variables, providing insights into potential relationships and dependencies.

4. **Column Removal (UDI and PRODUCTID):**

- Dropping Unique Identifiers: If UDI and PRODUCTID columns contain unique identifiers that don't contribute meaningfully to the predictive modeling, removing them can enhance model efficiency and interpretability.

5**. Binary Classification:**

- Implementing a binary classification task involves assigning instances to one of two classes (positive or negative) based on predictive models.
- This could include selecting appropriate features, splitting the dataset into training and testing sets, choosing a classification algorithm (e.g., Logistic Regression, Decision Trees, Random Forests), and evaluating model performance using metrics like accuracy, precision, recall, and F1-score.

CONFLICTS:

1. **Choosing the right data preprocessing techniques:** Handling categories, encoding variables, scaling etc. require certain prior knowledge.

2. **Selecting appropriate machine learning algorithms:** There are many ML algorithms and as a beginner it is very difficult to choose the appropriate ML model. With experience one learns which algorithm works better for different problems.
3. **Interpreting model outcomes:** Making sense of accuracy metrics, confusion matrices, feature importances etc. require practice.

NEXT STEPS:

1. **Hyperparameter Tuning for Binary Classification:**

- Hyperparameter tuning involves optimizing the settings of a machine learning model to improve its performance.
- Common techniques include grid search or randomized search over a predefined hyperparameter space.

2. **Studying the Best Model for Binary Classification:**

- After hyperparameter tuning, evaluate and compare the models based on performance metrics like accuracy, precision, recall, and F1-score
- Choose the model with the best balance of these metrics.

3. **Perform Multiclass Classification on the Data by Considering Machine Failure Type**
- Extend the binary classification task to multiclass by predicting the specific type of machine failure.
- Adjust the target variable and model accordingly to accommodate multiple classes.

4. **Perform Exploratory Data Analysis (EDA) on Multiclass Dataset:**
- Analyze the distribution of features and target classes.
- Explore relationships between variables.
- Visualize patterns and trends in the data to gain insights.

5. **Perform Train-Test Split on Dataset:**
- Split the dataset into training and testing sets to assess the model's generalization performance.
- Ensure that the split maintains the distribution of classes to avoid bias.

6. **Define Logistic Regression, Decision Tree, and Random Forest Models:**
- Instantiate machine learning models (Logistic Regression, Decision Tree, Random Forest) for multiclass classification.
- Configure the models with default hyperparameters or the tuned ones from the binary classification phase.

7. **Plot Decision Tree in Decision Tree Model for Multiclass Classification Section:**
- For Decision Tree models, visualize the tree structure to understand how the model makes decisions.

- The plot helps interpret feature importance and decision pathways in the context of multiclass classification.

8. **Hyperparameter Tuning for Multiclass Classification:**
- Similar to binary classification, optimize hyperparameters for multiclass models to enhance performance.
- Use techniques like grid search or randomized search.

9. **Studying the Best Model for Multiclass Classification:**
- Evaluate models on multiclass metrics such as accuracy, precision, recall, and F1-score.
- Select the model that performs well across all classes.