A logo of a company

Description automatically generated A close-up of a logo

Description automatically generated

**Joint Tech Internship Community Program**

**Assignment: Predicting Equipment Failure in a Manufacturing Plant**

**Problem Statement:**

A manufacturing plant is looking to predict equipment failures before they happen to minimize downtime and maintenance costs. Your task is to develop a machine learning model that predicts whether a piece of equipment will fail based on its operational data, usage history, and environmental conditions.

**Objective:**

Build a predictive model to classify whether a piece of equipment will fail within a specified time frame. Evaluate the model using appropriate metrics and provide insights into the factors contributing to equipment failure.

**Dataset:**

You are provided with a dataset containing the following columns:

1. **EquipmentID**: Unique identifier for each piece of equipment.
2. **Age**: Age of the equipment in years.
3. **UsageHours**: Total number of hours the equipment has been used.
4. **MaintenanceHistory**: Number of maintenance operations performed in the last year.
5. **Temperature**: Average operating temperature of the equipment (in degrees Celsius).
6. **Pressure**: Average operating pressure of the equipment (in psi).
7. **VibrationLevel**: Average vibration level of the equipment (in mm/s).
8. **OperatorExperience**: Experience level of the operator handling the equipment (in years).
9. **FailureHistory**: Number of failures the equipment has experienced in the past.
10. **Location**: The location of the equipment within the plant (e.g., Section A, Section B).
11. **Environment**: The environmental conditions where the equipment operates (e.g., Humid, Dry, Normal).
12. **Failure**: Whether the equipment has failed (Yes/No).

**Tasks:**

1. **Data Exploration and Preprocessing:**
   * Load the dataset and perform initial exploration to understand the data.
   * Identify and handle any missing values appropriately.
   * Detect and remove outliers using appropriate statistical methods (e.g., Z-score, IQR).
   * Convert categorical variables into numerical ones using techniques such as One-Hot Encoding.
2. **Feature Engineering:**
   * Perform feature scaling (e.g., Standardization or Normalization) on continuous variables.
   * Create new features if relevant, such as usage rate (UsageHours/Age).
   * Use dimensionality reduction techniques (e.g., PCA) to reduce the feature space if necessary.
3. **Model Building:**
   * Split the dataset into training and testing sets (e.g., 80-20 split).
   * Train different classification models (e.g., Logistic Regression, Decision Trees, Random Forest, Gradient Boosting).
   * Perform hyperparameter tuning using techniques like Grid Search or Random Search to optimize model performance.
4. **Model Evaluation:**
   * Evaluate your models using appropriate metrics such as Accuracy, Precision, Recall, F1-Score, and AUC-ROC.
   * Compare the performance of different models and select the best one.
   * Analyze feature importance to understand the most significant factors contributing to equipment failure.
5. **Insights and Recommendations:**
   * Provide insights based on your model analysis regarding the factors that influence equipment failure.
   * Suggest actionable strategies for preventive maintenance based on the findings.
6. **Documentation:**
   * Document your process, including data exploration, preprocessing steps, model selection, and evaluation.
   * Include visualizations where necessary to support your findings.

**Deliverables:**

* A Jupyter notebook (or Python script) with the entire workflow.
* A report summarizing your findings, including the model's performance and recommendations for the business.

This assignment problem focuses on predicting equipment failure in a manufacturing environment, covering various aspects of a typical machine learning workflow, including data preprocessing, outlier removal, feature engineering, dimensionality reduction, model training, and evaluation. It also emphasizes providing actionable insights for preventive maintenance.