**Project 1**

**AnomaData** **(Automated Anomaly Detection for Predictive Maintenance)**

### Problem Statement

Predictive maintenance is crucial across various industries to minimize risks and enhance operational efficiency by analyzing equipment data to foresee potential failures. While equipment failures are common across different machines, the ability to predict and prevent such failures is essential. Predictive maintenance involves evaluating equipment conditions through continuous monitoring to preemptively address issues before they escalate.

This Capstone project aims to forecast machine breakdowns by identifying anomalies within provided data. The dataset comprises over 18,000 rows collected over a few days. The 'y' column in the dataset contains binary labels where 1 signifies an anomaly. Other columns serve as predictors for the model.

**Steps for Solution:**

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

Perform a thorough analysis to uncover patterns, relationships, and trends.

Utilize descriptive statistics and visualizations to understand the dataset.

· 2. **Data Cleaning:**

Standardize the data and handle missing values and outliers.

Ensure data quality and consistency.

· 3. **Feature Engineering:**

Develop new features or transform existing ones to enhance model performance.

· 4. **Model Selection:**

Identify and select the most suitable machine learning model for predicting anomalies.

· 5. **Model Training:**

Split the dataset into training and testing subsets.

Use the training set to optimize model parameters.

·6. **Model Validation:**

Evaluate model performance on the test set.

Assess the model's ability to generalize to unseen data and identify any overfitting issues.

· 7. **Model Deployment:**

Plan and execute the deployment of the trained model in a production environment.

## Getting Started with Libraries:

## library(RandomizedSearchCV), library(RandomForestClassifier), library(LabelEncoder)

## Models:

These models are prominently used for the project

1. **Logistic Regression**

### 2. **Decision Trees**

### 3. **Random Forest**

### 4. **Gradient Boosting Machines (GBM)**

### 5. **Support Vector Machines (SVM)**

### 6. **Neural Networks**

## Environment Settings:

Anaconda Jupyter Notebook  
Python 7.0.8

**Benefits to company:**

* **Improved Equipment Reliability:** By predicting and preventing machine breakdowns, the company can reduce downtime and maintenance costs.
* **Operational Efficiency:** Enhanced predictive maintenance can lead to more efficient use of resources and better planning.
* **Data-Driven Decision Making:** The model provides actionable insights based on data, enabling more informed decisions.
* **Cost Savings:** Preventative maintenance can lower repair and replacement costs, contributing to overall cost savings.
* **Competitive Advantage:** Implementing advanced predictive maintenance solutions can give the company a competitive edge by improving service quality and reliability.