Capstone Project: *Predictive Maintenance Using Machine Telemetry Data*

Using the dataset: [Microsoft Azure Predictive Maintenance (kaggle.com)](https://www.kaggle.com/datasets/arnabbiswas1/microsoft-azure-predictive-maintenance), It consists of the following data:

**Machine conditions and usage:** The operating conditions of a machine e.g. data collected from sensors.

**Failure history:** The failure history of a machine or component within the machine.

**Maintenance history:** The **repair** history of a machine, e.g. error codes, previous maintenance activities or component replacements.

**Machine features:** The features of a machine, e.g. engine size, make and model, location.

**Details:**

Telemetry Time Series Data (PdM\_telemetry.csv): It consists of hourly average of voltage, rotation, pressure, vibration collected from 100 machines for the year 2015.

**Objective:**

The goal of this capstone project is to develop a predictive model that can forecast equipment failures using machine telemetry data. This model will help in scheduling maintenance more effectively, thereby reducing downtime and maintenance costs.

**Deliverables:**

1. A predictive model developed using machine learning techniques.

2. A technical write-up in a Jupyter Notebook, detailing the methodologies, code, and evaluations.

3. A non-technical README file that summarizes the findings and implications of the study.

**Initial Capstone Project Submission**

**General Question Idea**

*“How can machine telemetry data be used to predict equipment failures in order to schedule proactive maintenance and reduce operational downtime in manufacturing environments?"*

This question aims to explore the potential of utilizing existing telemetry data (such as temperature, pressure, vibration, etc.) from manufacturing equipment to predict potential failures before they occur. By predicting failures, maintenance can be scheduled at an optimal time, thus minimizing unexpected breakdowns and costly operational interruptions.

**Initial Guess at Required Data**

To answer this question, the following types of data will be required:

* Telemetry Data: Time-stamped machine data capturing various operational parameters like temperature, pressure, vibration, rotation, voltage, etc.
* Maintenance Records: Historical logs detailing maintenance activities, including dates, types of maintenance (corrective, preventive), and any part replacements.
* Failure Logs: Records of machine failures including the time of failure, type of failure, and any error codes associated.

**Initial Guess at Data Sourcing**

* Telemetry Data and Failure Logs: These can be sourced directly from the operational databases of manufacturing facilities. Most modern machinery is equipped with sensors that log this data continuously.
* Maintenance Records: These are typically kept in enterprise resource planning (ERP) systems or other maintenance scheduling software used by the facility.

This data will help in developing a predictive model that can identify patterns or anomalies that precede equipment failures, thereby enabling timely preventive measures. This submission will serve as the basis for discussions with the Learning Facilitator to refine the project focus, determine appropriate data handling methods, and explore potential analytical techniques for model development.

**Structure and Schedule:**

**Problem Statement and Project Prospectus**

**Problem Statement:**

Predictive maintenance (PdM) techniques are designed to help determine the condition of in-service equipment in order to predict when maintenance should be performed. This approach promises cost savings over routine or time-based preventive maintenance because tasks are performed only when warranted. The objective is to predict failures before they happen, based on the machinery's condition monitored through telemetry data.

**Project Prospectus:**

Question to Address: Can we predict machine failures in advance using historical telemetry data?

**Data Required:**

* Telemetry data capturing machine conditions at regular intervals (temperature, voltage, rotation, pressure, etc.).
* Maintenance records indicating past failures and repairs.
* Failures records
* Machine specifications including types, models, and installation details.

**Techniques to be Used:**

* Data Cleaning: Handling missing values, filtering noise.
* Feature Engineering: Generating meaningful features from raw telemetry data, like rolling averages, standard deviations, and more.
* Model Selection: Comparing different models to find the best predictor. Models could include:
* Decision Trees
* Random Forest
* Model Evaluation: Using confusion matrix, precision, recall, F1-score, and ROC curves to assess performance.
* Visualization: To illustrate the predictive capabilities and the importance of different features.

**Expected Outcome:**

A robust model that can accurately forecast upcoming failures, allowing preemptive maintenance actions to be scheduled at the optimal time.

**Data Preparation and Exploratory Analysis**

**1. Data Collection:**

- Import and combine telemetry data and maintenance records.

- Summarize data characteristics at initial stages.

**2. Data Cleaning and Preprocessing:**

- Handle missing data and anomalies.

- Normalize or standardize data as required.

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

- Visualize data distributions, correlations, and patterns.

- Initial hypothesis testing with simple statistical tests.

**Feature Engineering and Model Building**

**1. Feature Engineering:**

- Develop features that capture trends, cycles, and anomalies in the telemetry signals.

- Aggregate features over different time windows (hourly, daily).

**2. Model Development:**

- Split data into training and testing sets.

- Train various models and tune hyperparameters.

- Validate models using cross-validation techniques.

**Model Evaluation and Finalization**

**1. Testing:**

- Test the models on unseen data.

- Compare model outputs with actual outcomes.

**2. Evaluation:**

- Use various metrics to evaluate model performance.

- Refine models based on testing feedback.

**3. Visualization:**

- Create visualizations to demonstrate model effectiveness.

- Highlight critical features and prediction accuracies.

**Documentation and Presentation**

**1. Technical Write-Up:**

- Document the entire process in a Jupyter Notebook.

- Include code, comments, and visualizations.

**2. Non-Technical README:**

- Summarize the project's scope, methodologies, and findings in layman's terms.

- Post on GitHub along with the Jupyter Notebook.

**3. Presentation:**

- Prepare a presentation summarizing the project findings.

- Include insights for non-technical stakeholders.

**Milestone Reviews:**

- Review data collection, cleaning, and EDA findings.

- Evaluate initial model performances.

- Final model selection and evaluation.

- Complete documentation and prepare for final presentation.