**Research Proposal**

**Group information:**  
Group number: Group 4  
Group members: Li Minghao, Lu Yung-Ting, Pan Yumeng, Shi Yilin, Yang Xinyu

**Dataset selection and motivation:**  
We have chosen the Machine failure dataset as the basis of our research. By analyzing machine operation data, we want to understand how machine failure takes place and predict when they are likely to occur, ultimately enhancing machine operation efficiency and reducing costs in industrial settings.

**Research questions:**

1. What are the different failure modes of machines, and what characteristics define each mode?
2. Based on the state of various features, can we determine whether a machine is operating normally?
3. Can we predict if and when a machine will fail based on the current feature states?

**Proposed methods and motivation:**  
To address these research questions, we propose the following machine learning approaches:

* **K-Means Clustering**: To identify distinct machine failure modes, we will apply K-Means clustering to group failures into different categories based on their feature patterns.
* **Regression models**: For predicting machine status and failure timelines, we will develop regression models. The models will be trained using feature data to estimate the probability of machine failure and predict the remaining life of a machine. Regression models are chosen due to their ability to capture relationships between numerical feature states and machine failure occurrences.

**Experimental plan:**  
To test and evaluate our methods, we plan to:

1. **Validate the K-Means model** by analyzing cluster quality and interpreting feature distributions within each cluster.
2. **Evaluate regression models** using performance metrics such as R-squared (R²), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) to assess predictive accuracy.
3. **Compare multiple regression algorithms**, such as Random Forest regression and Neural Networks, to determine the best-performing model for failure prediction.
4. **Conduct feature importance analysis** to identify the most significant features contributing to machine failures.

Through these experiments, we aim to develop robust models that provide insights for predictive maintenance, improving machine reliability and efficiency.