**Project Proposal Template**

**Kaggle Competition Link:** https://www.kaggle.com/competitions/playground-series-s3e17/discussion

**Problem Description:**

This competition is a binary classification task. The goal is to predict whether a machine will fail or not, based on a set of features such as air temperature, process temperature, rotational speed, torque, tool wear, and product ID.

**Data:**

The dataset contains 136,429 rows and 14 features. The features are as follows:

* id: A unique identifier for each machine
* Product ID: The type of machine
* Type: The type of machine failure (or none)
* Air temperature [K]
* Process temperature [K]
* Rotational speed [rpm]
* Torque [Nm]
* Tool wear [min]
* Machine failure: A binary variable indicating whether or not the machine failed

**Machine Learning Task:**

This is a binary classification task. The goal is to train a machine learning model to predict whether a machine will fail or not, based on the set of features provided.

**Machine Learning Methods:**

I plan to use a variety of machine learning methods to solve this problem, including:

* **Random Forest Classifier:** Random forests are a type of ensemble learning method that combines the predictions of multiple decision trees to make a final prediction.
* **Gradient Boosting Classifier:** Gradient boosting is another type of ensemble learning method that combines the predictions of multiple weak learners to make a final prediction.
* **Support Vector Machine (SVM):** SVM is a machine learning algorithm that can be used for both classification and regression tasks. SVMs work by finding a hyperplane in the feature space that separates the data into two classes.

**Evaluation Methods:**

I will evaluate the performance of my models using the following metrics:

* **Accuracy:** Accuracy is the percentage of predictions that are correct.
* **Precision:** Precision is the percentage of positive predictions that are actually positive.
* **Recall:** Recall is the percentage of actual positive examples that are correctly predicted.
* **F1 Score:** The F1 score is a harmonic mean of precision and recall.

**Outcome of the Project:**

I hope to develop a machine learning model that can accurately predict whether a machine will fail or not. This model could be used to help prevent machine failures, which could save businesses time and money.

**Additional Considerations:**

One of the challenges of this problem is the class imbalance. The dataset contains significantly more negative examples (machines that did not fail) than positive examples (machines that failed). This can make it difficult for machine learning models to learn to identify the positive examples.

Another challenge is the lack of information about the machine failures. The dataset does not provide any information about the cause of the machine failures. This makes it difficult to develop a machine learning model that can accurately predict machine failures.

Despite these challenges, I believe that it is possible to develop a machine learning model that can accurately predict machine failures. I am excited to work on this project and to see what I can achieve.