**Project Proposal: Developing a Remaining Useful Life (RUL) Predictor for NMC-LCO 18650 Batteries**

**Problem statement:**

1. The reliable estimation of a battery's Remaining Useful Life (RUL) is critical for applications such as electric vehicles, renewable energy storage, and portable electronics. The ability to estimate the RUL might be further helpful in identifying an optimal charge discharge pattern to maximize the battery life. This project aims to develop an accurate RUL predictor for NMC-LCO 18650 batteries using a dataset provided by **Ignacio Vinuales**. The dataset includes detailed measurements from 14 batteries cycled over 1000 times at specific charge and discharge rates.

**Objectives**

1. **Develop an RUL Predictor:** Utilize the dataset to build a predictive model that accurately closely mimics the observed RUL of the batteries.
2. **Explore Various ML Algorithms:** Investigate different machine learning algorithms to identify the most effective method for RUL prediction.
3. **Align with Previous Research Experience:** This project will leverage my background of researcher in battery to transition into a data science career.

**Dataset Overview**

1. The dataset contains the following variables:

|  |  |
| --- | --- |
| * Cycle Index * Discharge Time (s) * Decrement 3.6-3.4V (s) * Max. Voltage Discharge (V) | * Min. Voltage Charge (V) * Time at 4.15V (s) * Time Constant Current (s) * Charging Time (s) * RUL |

**Methodology**

1. **Data Wrangling:**
   * Load and inspect the dataset.
   * Normalize or standardize features if necessary.
2. **Exploratory data analysis (EDA):**
   * Perform exploratory data analysis (EDA) to understand the data distribution and relationships.
3. **Feature Engineering:**
   * Extract additional features or create new ones if necessary.
   * Select the most relevant features using feature selection techniques.
4. **Model Development:**
   * Split the dataset into training and testing sets.
   * Implement and evaluate various machine learning algorithms.
5. **Model Evaluation:**
   * Compare the performance of different models using appropriate metrics (e.g., RMSE, MAE).
   * Select the best-performing model based on evaluation results.

**Potential Machine Learning Algorithms**

1. **Linear Regression:**
   * Simple and interpretable.
   * Useful as a baseline model to compare against more complex algorithms.
2. **K-Nearest Neighbors (KNN):**
   * Simple and intuitive.
   * Effective when the relationship between features and the target variable is highly non-linear.
3. **Random Forest:**
   * An ensemble method that can handle non-linear relationships.
   * Robust to overfitting due to the averaging of multiple decision trees.
4. **Neural Networks:**
   * Capable of capturing complex patterns in the data.
   * Requires careful tuning and larger computational resources.
5. **Support Vector Machines (SVM):**
   * Effective in high-dimensional spaces.
   * Suitable for regression tasks with non-linear relationships.

**Expected Outcomes**

1. A robust and accurate RUL predictor for NMC-LCO 18650 batteries.
2. A comprehensive comparison of various machine learning algorithms for RUL prediction.

**Conclusion**

1. This project not only aims to contribute to the field of battery technology but also serves as a critical step in my career transition to data science. By developing an RUL predictor using state-of-the-art machine learning techniques, the project will provide valuable insights and practical solutions for extending the life and reliability of battery systems.