# Determining the Remaining Useful Life of Lithium Ion Batteries using Machine Learning

**Sriya Gupta**

The Shri Ram School, Moulsari,
V-37, Moulsari Ave, DLF Phase 3, Sector 24, Gurugram, Haryana, 122002
*sriyasgupta@gmail.com*

**Abstract:** With a dramatic shift towards renewable energy in the energy sector, the global demand for batteries has exponentially increased. It has become imperative to reliably assess and predict the remaining lifespan of the lithium ion batteries (LIB). This paper works towards examining the battery chemistry of lithium ion batteries and their working mechanism in addition to a brief literature review about them. Descriptions of popular ML algorithms like linear regression, decision trees and extreme gradient boosting (XGBoost Regressor) have also been presented and these models have then been used to predict the remaining useful life (RUL) of NMC-LCO batteries, a type of LIB, using a publicly-available dataset. It was found that, out of these three models, XGBoost Regressor performed the best and was able to predict values for RUL with an accuracy of 99.93%. The paper discusses these results and observations.

Keywords: Machine Learning, Lithium Ion Batteries, Remaining Useful Lifespan

## 1. Introduction

Machine learning (ML) has played a critical role in several technological advancements which are driving the economy today; undoubtedly, the energy sector is one such area. The rapid digitization and increased accessibility to appropriate data has opened up an array of uses for ML in the energy sector. Researchers are now using ML algorithms to produce accurate forecasts of electricity demand as well as maintain suitable operating conditions for energy generating systems. [1] As humanity gears towards a more technology-driven world, batteries have become the key to powering most devices that society is heavily dependent on. Due to their rechargeable capacity, lithium-ion batteries (LIB) are among the top choices for batteries across the globe.

The lithium-ion battery (LIB) is a type of rechargeable battery that utilizes the process of reversible reduction to generate electricity.(2) It has played a monumental role in powering both portable and stationary electronic devices during the last few years. Due to its high energy density, an LIB is lightweight and therefore popular in this era of portable devices such as laptops and mobile phones. Commercially, there are various types of lithium ion batteries available. The variations are mostly in the type of metal combinations used for anodic and cathodic materials or the electrolytes used. To mention a few, Lithium Cobalt Oxide (LCO), Lithium Iron Phosphate, Lithium Nickel Manganese Cobalt Oxide (NMC), Lithium Titanate batteries and the lithium polymer batteries (use a polymer electrolyte). One of the most successful combinations is the NMC-LCO battery. This battery type consists of a mix of nickel, manganese and cobalt in the cathode as well as a graphite anode and generates a voltage ranging from 3.0V to 4.2V. Its immense success in the battery field can be attributed to nickel's high specific energy while manganese's stable spinel structure compensates for nickel's poor stability. The combination leads to low internal resistance and high specific energy. [2]

While these batteries are extremely suitable for various purposes, many organizations or individuals are hindered from using these batteries due to their unpredictable lifespan. Thus, by making it possible for us to estimate the RUL of the battery at any given time, it becomes more practical to implement these batteries in a plethora of other functions and shift our society towards operating in a sustainable yet efficient manner. In this paper, I have discussed the working mechanism of LIBs, incorporated a brief literature review, provided short descriptions of popular ML algorithms as well as employed ML models to predict the remaining useful life (RUL) of NMC-LCO batteries using a publicly available dataset.

### 1.1 Working mechanism of Li-ion batteries

The main purpose of a battery is to generate electricity. The term electricity is commonly defined as the flow of electrons from the negative terminal of a battery to the positive terminal of a battery. In a lithium-ion battery, the electrons originate from the element Lithium (Li). Li is in group 1 of the periodic table, making it extremely reactive and ready to lose the electron that lies in its 2s orbital. This allows Li to have one of the highest electrochemical potentials.

Lithium-ion batteries consist of an anode, cathode, a separator and an electrolyte. The anode refers to the positively-charged electrode and the cathode refers to the negatively charged electrode. NMC-LCO batteries consist of nickel, manganese and cobalt cathode and a graphite anode. Li is stored within the layers of the graphite anode with the help of a phenomenon known as intercalation. Intercalation is the process through which an element's atoms are inserted into a host lattice. As shown in Figure 1, during the discharge cycle of the battery, Li atoms separate from the anode and its valence electrons migrate to the cathode through an external path, typically a wire, leaving behind positively-charged Li ions. The flow of these delocalised electrons cause the production of electricity in an LIB. Upon their arrival at the cathode, these electrons are used by the positively charged cobalt to regain a neutral charge and obtain a noble-gas structure. The lithium cations, on the other hand, leave the graphite layers at the same time as the electrons migrate. They travel through an electrolyte which allows only the ions to move through and blocks the valence electrons. Thus, when the Li ions reach the cathode and intercalate themselves between the cobalt oxide layers, the charge at the cathode is balanced out. As shown in Figure 2, in order to charge the battery, the lithium ions and delocalised electrons have to return to their original positions in between the graphite layers. Finally, during the next discharge cycle the entire process repeats itself.

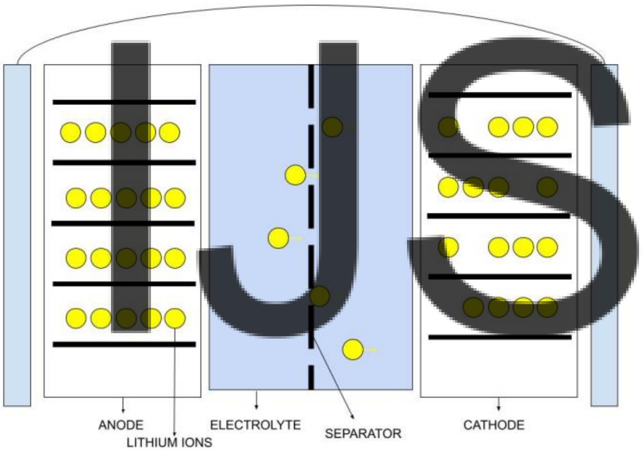


Figure 1. Illustration of the Discharging Process of Lithium Ion Batteries

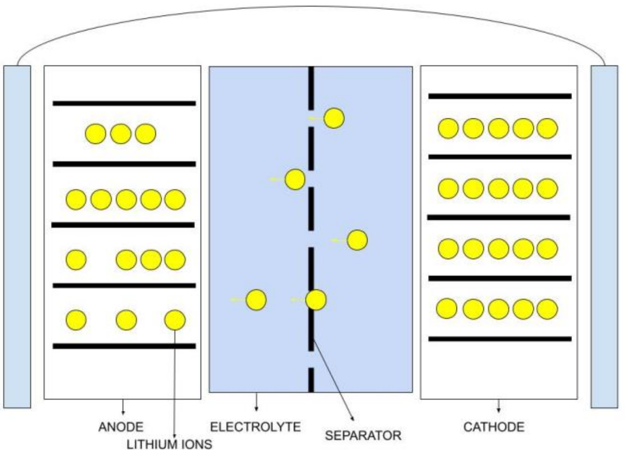


Figure 2, Illustration of the Charging Process of Lithium Ion Batteries

### 1.2. Introduction to Machine Learning

Machine Learning (ML) is a subset of artificial intelligence that focuses on managing data and algorithms to recreate the way that humans learn. ML algorithms are popularly used to make predictions or classification using a data set as well as identifying patterns in the data[3]. ML can further be classified into three more distinct types. The first type is commonly referred to as supervised learning. It involves algorithms that can accurately predict or make classifications from labeled data. On the other hand, unsupervised machine learning consists of algorithms that analyze hidden patterns in unlabelled data sets[4]. Finally, reinforcement learning, while it has similar outcomes as supervised machine learning, does not make use of training datasets; instead it employs the trial and error method to classify and predict data in a reliable and valid manner.

In this work, I have used three popular ML algorithms namely Linear Regression, Decision Trees and Extreme Gradient Boost (XGBoost Regressor). The following paragraphs briefly describe them.

#### 1.2.1. Linear Regression:

Linear regression is one of the most popular models in machine learning. This type of model works towards establishing a linear relationship between the independent and dependent variables. If a single independent variable is being used to predict the dependent variable, then an algorithm known as Simple Linear Regression is used. However, multiple linear regression is used when multiple independent variables are used to calculate a numerical dependent variable. This model uses the equation $y = mx+c+e$, wherein y is the dependent variable, m is the gradient, x is the independent variable, c is the y-intercept of the line, and e is the random error. An upward sloping regression line depicts direct proportionality while a downward sloping graph indicates inverse proportionality.[5]

#### 1.2.2. Decision Tree:

The algorithm starts by evaluating the input data and selecting the feature that is most informative for making predictions. It splits the data based on this feature into two or more subsets, where each subset represents a different branch of the tree. At each internal node of the tree, a condition is applied to the input data based on a specific feature which determines which branch to follow depending on whether the data satisfies the condition or not. The decision tree algorithm repeats the process of splitting and testing the data at each internal node, aiming to create a tree structure that accurately represents the patterns and relationships in the data.[6]

### 1.2.3. XGBoost Regressor:

XGBoost builds an ensemble of trees sequentially, where each new tree is trained to correct the mistakes made by the previous trees. It starts with a simple decision tree which predicts the target variable based on a single feature. Then, the differences between these predictions and the actual target values are calculated, which are called residuals. XGBoost builds a new decision tree to predict the residuals. This tree is designed to minimize the residuals, meaning it learns to predict the remaining errors from the previous trees. The new tree is added to the ensemble of trees, and the predictions from all the trees are combined. This combined prediction is an updated estimate of the target variable. This process is repeated iteratively, creating additional trees and adjusting the predictions. Each new tree focuses on correcting the mistakes made by the previous trees. The process continues until a predefined number of trees is reached or no further improvement is observed. The final prediction is obtained by combining the predictions from all the trees in the ensemble. [7]

### 1.3. Literature Review:

The concept of the first lithium battery was introduced in 1970 by Stanley Whittingham. He proposed the idea of a rechargeable battery that could lead to the mitigation of fossil fuel resources. This battery consisted of a lithium anode and a titanium disulphide cathode [8]. The electrons flowed from the metallic lithium anode powering an external device through an external circuit. While, the lithium cations flowed through the electrolyte from the anode to the cathode. The titanium disulfide cathode was composed of layers that housed the lithium ions between them. Recharge cycles forced lithium ions to return to their starting positions in the anode. Therefore, enabling lithium ions to be reused for future charge cycles. Despite achieving a high energy density and good working capability at room temperature, Whittingham's model harbored several safety issues. [9]. The main issue with Whittingham's structure was that over time, the surface of the battery would become rough and start to form long finger-like structures called dendrites. These dendrites possessed the ability to pierce through the separator of the battery and cause a short-circuit [10]. In an effort to mitigate these hazards, a lithium cobalt oxide cathode was proposed [11]. Similar to the titanium disulphide cathode, this new cathode allowed ions to exist between its layers enabling the battery to house a significantly larger amount of lithium ions, this idea led to the doubling of energy potential for the battery [12]. However, the battery still posed a fire hazard. To increase the battery's safety, Michael Armand proposed graphite as the anode material. His work demonstrated that graphite intercalates several alkali metals as it contains half-filled Pz orbitals to account for issues such as the growth of dendrites as well as volume expansion. However, the co-intercalation of propylene-carbonate soon led to the collapse of graphite as an anode material soon after [13]. In 1991, Sony went on to commercialize LIB. This updated structure was then adopted by larger industries and is now widely used throughout the world as a greener alternative to fossil fuels. Although many scientists have worked on the lithium-ion battery, issues with lifespan continue to exist. The reduction of mobile lithium ions continues to pose a great challenge to battery developers as side reactions occurring within the battery consume the mobile ions, thereby significantly reducing its capacity. Moreover, structural disordering can also lead to damage in the electrodes. This phenomenon causes the lifespan of an LIB to decrease substantially as well [14]. Thus, it is vital for us to be able to forecast and try to extend the remaining useful life of these batteries. Therefore, ML algorithms can prove to be extremely useful in order to discern the lifespan and capacity that is left within the battery. These algorithms have helped to accurately predict the RUL of batteries and do not require complex mathematical models. ML considers batteries a black box and by mapping a set of battery features can predict the battery lifetime through various ML models[15].

## 2. METHOD:

The three ML models discussed above namely linear regression, decision trees and XGBoost Regressor were used for predicting the Remaining Useful Life (RUL) of NMC-LCO batteries. The publicly available dataset was sourced from a reliable platform called Kaggle[16]. The dataset creator has provided the following description about the dataset: 14 NMC-LCO 18650 batteries with a nominal capacity of 2.8 Ah were investigated by the Hawaii Natural Energy Institute after being cycled more than 1000 times at a temperature of 25 °C with a CC-CV charge rate of C/2 rate and discharge rate of 1.5C. The creator of this dataset then developed characteristics from the source dataset that illustrate the behavior of the voltage and current during each cycle. These characteristics can be used to estimate the batteries' RUL. The 14 batteries' summaries are contained in the dataset as shown in Table 1. Also, Table 1 presents a description of the various features available in the dataset. Google Colaboratory was used for coding in Python.

Table 1 - Description of Columns (features) present in the dataset.

| Column Name | Description |
|---|---|
| Discharge Time (s) | The time, in seconds, required to reach the minimum voltage |
| Decrement 3.6-3.4 V (s) | The time it takes, in seconds, for the voltage to decrease from 3.6V to 3.4V |
| Max. Voltage Discharge (V) | The initial and maximum value of voltage |
| Min. Voltage Discharge (V) | The initial and minimum value of voltage |
| Time at 4.15 V (s) | The time required for the battery to reach a voltage of 4.15V. |

| | |
|---|---|
| Time Constant Current (s) | The time at which the current remains constant at its maximum value. |
| Charging Time (s) | The amount of time it takes to charge the battery to its full capacity. |
| RUL | The number of cycles that the battery can continue to operate for. |

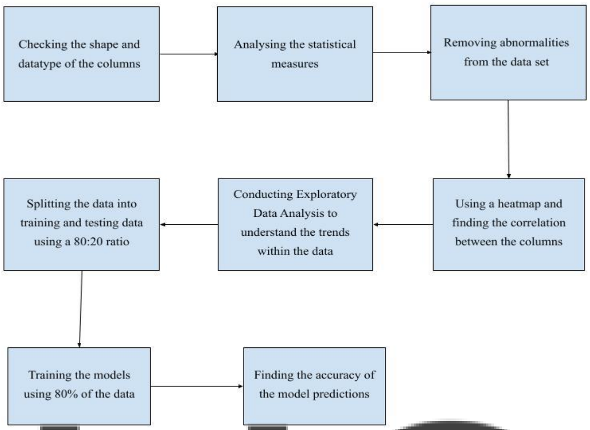


Figure 3, Summary of steps followed while training various models with the dataset

The steps mentioned in Figure 3 were performed and the observations have been discussed in Section 3.

## 3. Results and Discussion:

For understanding the data, various python functions are used to find the shape of the data, datatype of all the columns and any missing values for the columns are found. The dataset has 15,604 rows and 9 columns. No missing values were found in the dataset. Moreover, all the columns in the dataset are numerical with either integer or float data types. After this, the statistical characteristics of the data were computed and it was observed that there were negative values in the columns "Decrement 3.6-3.4V (s)" and "Time at 4.15V (s)". These were removed. There were extremely large positive values for certain columns as well, but due to the large number of rows that contained them, their treatment would severely bias the data. Hence they were not changed.

To further find out the relationships between different columns or features, a heatmap was created as shown in Figure XX. A heatmap depicts the correlation coefficients between the features of a dataset.

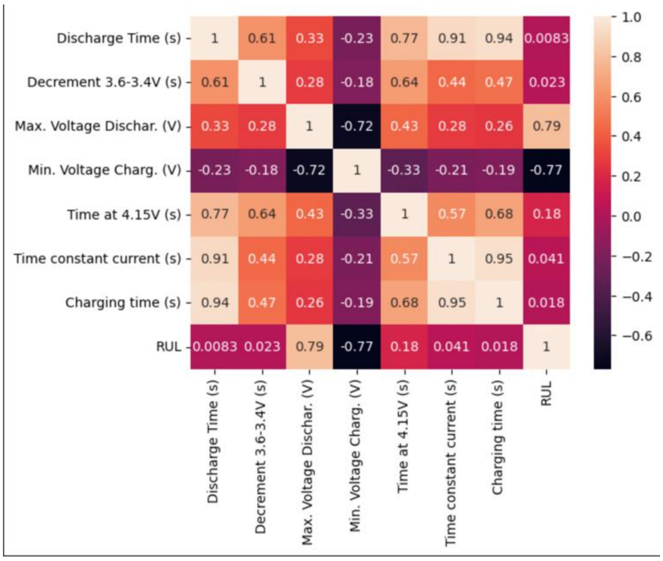


Figure 4. Heatmap illustrating correlations between the columns

A strong correlation was considered as one that is above 75 percent. The most notable correlation is the one that exists between the maximum voltage discharge and the RUL with a 0.79 correlation coefficient. As mentioned in section 2, the maximum voltage discharge acts as a measure for the value of the initial and maximum value of voltage. In most rechargeable battery chemistries, such as lithium-ion batteries, the maximum voltage is reached when the battery is fully charged. As the battery undergoes charge and discharge cycles over its lifetime, it gradually loses its ability to hold charge and deliver the same maximum voltage. As the battery degrades, its maximum voltage tends to decrease. This reduction in maximum voltage can be an indication of the battery's declining performance and its approaching end of life. Therefore, a lower maximum voltage may suggest that the battery's RUL is decreasing.
Similarly, due to the battery capacity deteriorating with higher voltages, the lower the minimum voltage charge in the battery, the higher is the RUL of the battery. As a result, the RUL of the battery and the minimum voltage charge share an inversely proportional relationship with a correlation coefficient of -0.77.

Moreover, the three features namely Discharge Time, Time Constant Current and Charging Time are all positively correlated with coefficients that have values above 0.9. The reasons behind this can help scientists gain further insights into the inner workings of battery chemistry. The time constant current refers to a constant discharge current that is applied to the battery during the discharging process. It represents the rate at which the battery's stored energy is being depleted. The discharging time, on the other hand, refers to the duration for which the battery can sustain the constant discharge current before reaching its minimum acceptable voltage or capacity limit. Thus, as the time constant current increases, the discharge and charging time increase as well. The Time at 4.15V and the Discharge Time also have a correlation coefficient of 0.77. This is because the longer it takes the battery to discharge overall, the longer it will take the battery to reach a voltage of 4.15V when charging.

Moving forward, the features were plotted in the form of graphs to understand them better. It was observed that the RUL column showed a normal distribution indicating it to be a good sample to reflect NMC-LCO batteries in general.

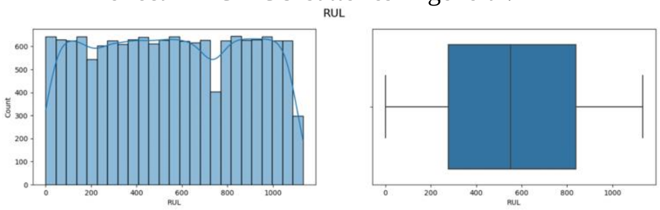


Figure 5. Normal Distribution Graph of RUL

The data was then split first into features and target, features comprising all the columns except the RUL column and the target data consisting of only the RUL column. Both these were further split into an 80:20 ratio with 80% forming the training data that will be used by the model to gain an understanding of the complex relationship between the target feature, RUL, and the other features of the dataset. The ML model will then use the information it has gathered from the training data on the testing data, which comprises 20% of the dataset. The ML models are then fed with the training data for both features and target. This allows the model to be trained with the training data. The accuracy is calculated in two phases: first for the training data, where after training, the model is asked to predict the target using the same training data features; and second for the testing data, when the model is fed the testing data containing the features and asked to predict the target. This allows us to check if the model is overfit. Overfitting is a concept that occurs when a statistical model fits exactly against its training data. If the accuracy of both the phases are close to each other varying less than 5%, then the model is not overfit.

Table 2. Comparing the models and their accuracies

| Model | Train Accuracy | Test Accuracy |
|---|---|---|
| Linear Regression | 0.80014 | 0.78028 |
| Decision Trees | 0.9999 | 0.78028 |
| XGBoostRegressor | 0.99870 | 0.99394 |

Moreover, the prediction of RUL may depend on some values being above certain thresholds and adhering to certain conditions. Decision trees and XGBoost Regressor can deduce these conditions and predict outcomes accordingly. However, it becomes difficult for the linear regression model to deduce conditions as it focuses mainly on establishing linearity between variables. This explains the accuracy observed during the training process. However, during the testing process, the greatest discrepancy in accuracy was seen for the decision tree model. This is because decision trees can be prone to overfitting if the tree becomes too complex, and they may not capture complex relationships as effectively as other algorithms. Moreover, the same discrepancy was not observed in XGBoost as it is often used to improve the performance and robustness of decision tree models. XGB is different from decision trees because it builds an ensemble of the trees independently using random subsets of the data and combines their predictions through majority voting or averaging. Also, here the trees are trained to make independent predictions without considering the errors of other trees. Thus, it can make the most accurate predictions, in this case.

## Conclusion

This paper attempted to use machine learning models to determine the RUL of NMC-LCO batteries, a type of Li-ion batteries, with knowledge of pre-existing parameters. The main objective of this work was to make the use of batteries more reliable, encouraging consumers to switch to more sustainable ways of consuming energy. This objective was achieved by employing different machine learning models to the publicly available dataset for the RUL of NMC-LCO batteries. The RUL could be predicted with accuracies of above 99.93% by XGBoost Regressor model. Predictions using machine learning modes thus make the use of batteries more practical for consumers.

## References:


1. Blomgren, G.E. (2016). The Development and Future of Lithium Ion Batteries. *Journal of The Electrochemical Society*, [online] 164(1), pp.A5019–A5025. doi:https://doi.org/10.1149/2.0251701jes.
2. www.sciencedirect.com. (n.d.). *Lithium Ion Battery - an overview | ScienceDirect Topics*. [online] Available at: https://www.sciencedirect.com/topics/chemistry/lithium-ion-battery#:~:text=10.1%20Introduction- [Accessed 5 May 2022].
3. IBM (2023). *What is Machine Learning?* [online] www.ibm.com. Available at: https://www.ibm.com/topics/machine-learning.
4. ResearchGate. (n.d.). *(PDF) Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification*. [online] https://www.researchgate.net/publication/273246843_Comparison_of_Supervised_and_Unsupervised_Learning_Algorithms_for_Pattern_Classification.
5. www.javatpoint.com. (n.d.). *Linear Regression in Machine learning - Javatpoint*.[online] https://www.javatpoint.com/linear-regression-in-machine-learning.
6. scikit-learn. (n.d.). *1.10. Decision Trees*. [online] Available at: https://scikit-learn.org/stable/modules/tree.html#:~:text=Decision%20Trees%20(DTs)%20are%20a.
7. Brownlee, J. (2021). *XGBoost for Regression*. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/xgboost-for-regression/.
8. Reddy, Mogalahalli V., et al. "Brief History of Early Lithium-Battery Development." *PubMed Central (PMC)*, 17 Apr. 2020, https://doi.org/10.3390/ma13081884.
9. Liu, Zhao. "The History of the Lithium-Ion Battery." *Advancing Materials*, 11 Oct. 2019, www.thermofisher.com/blog/materials/the-history-of-the-lithium-ion-battery.
10. DNKPOWER. "Different Electrolytes Between Li-ion and Li-Po Batteries - Lithium Ion Battery Manufacturer and Supplier in China-DNK Power." *Lithium Ion Battery Manufacturer and Supplier in China-DNK Power*, 11 July 2022, www.dnkpower.com/different-electrolytes-li-ion-vs-lipo.
11. "Lithium Dendrite - an Overview | ScienceDirect Topics." *Lithium Dendrite - an Overview | ScienceDirect Topics*, https://doi.org/10.1016/B978-0-12-821434-3.00006-5.
12. "John B. Goodenough - the Inventor of the Li-ion Battery." *John B. Goodenough - the Inventor of the Li-ion Battery | FUERGY*, 12 Apr. 2018, www.fuergy.com/blog/john-b-goodenough-the-inventor-of-the-li-ion-battery.
13. Xie, Jing, and Yi-Chun Lu. "A Retrospective on Lithium-ion Batteries - Nature Communications." *Nature*, 19 May 2020, https://doi.org/10.1038/s41467-020-16259-9.
14. Browning, Beatrice. "Why Do Lithium-ion Batteries Degrade Over Time? - AirQualityNews." *AirQualityNews*, 12 Jan. 2021, airqualitynews.com/fuels/why-do-lithium-ion-batteries-degrade-over-time.
15. Fei, Z., Yang, F., Tsui, K.-L., Li, L. and Zhang, Z. (2021). Early prediction of battery lifetime via a machine learning based framework. *Energy*, [online] 225, p.120205. doi:https://doi.org/10.1016/j.energy.2021.120205.
16. www.kaggle.com. (n.d.). *Battery Remaining Useful Life (RUL)*. [online] Available at:https://www.kaggle.com/datasets/ignaciovinuales/battery-remaining-useful-life-rul.


## Author Profile

**Sriya Gupta is a 12th Grade student in The Shri Ram School. An internship with IISER instigated her interest in Lithium Ion batteries and she aims to integrate research on renewable energy, especially batteries with Artificial intelligence and Machine Learning.**