Addis Ababa Institute of Technology

School of Electrical and Computer Engineering

Telecommunication Engineering Graduate Program

Thesis on

Mobile Network Backup Power Supply Battery Remaining Useful Time Prediction Using Machine Learning Algorithms

By

Abel Hirpo

Mentor:

Dr. -Ing. Dereje Hailemariam

A Thesis Submitted to the School of Graduate Studies of Addis Ababa University in Partial Fulfillment of the Requirements for the Degree of Master of Science in Telecommunication Engineering

August 2023

Addis Ababa, Ethiopia

Addis Ababa Institute of Technology

School of Electrical and Computer Engineering

Telecommunication Engineering Graduate Program

Thesis on

Mobile Network Backup Power Supply Battery Remaining Useful Time Prediction Using Machine Learning Algorithms

By: Abel Hirpo

______________________________ ____________

Mentor Signature

______________________________ ____________

Mentor Signature

______________________________ ____________

Mentor Signature

______________________________ ____________

Evaluator Signature

______________________________ ____________

Evaluator Signature

Declaration

I, the undersigned, declare that this thesis is my original work, has not been presented for a degree in this or any other university, and all sources of materials used for the thesis have been fully acknowledged.

Abel Hirpo ____________________

Name Signature

Place: Addis Ababa

Date of Submission: August 31, 2023

This thesis has been submitted for examination with my approval as a university advisor.

Dr. -Ing. Dereje Hailemariam ________________

Advisor's Name Signature

Abstract

Base transceiver stations (BTSs) in mobile cellular systems are critical infrastructure for providing reliable service to mobile users. However, BTSs can be disrupted by electric power supply interruptions, which can lead to degraded quality of service (QoS) and quality of experience (QoE) for users. The reliability of the battery used in BTSs is affected by a number of factors, including: The instability of the primary power supply, Temperature fluctuations, Battery aging, the number of charging and discharging cycles (CDC) and the depth of discharge (DOD).As a result of these factors,the state of health (SOH) of a battery is impacted which will in turn affect the remaining useful time (RUT) of the battery. This can lead to disruptions in service for mobile users, as the BTS may not have available power to operate during a power outage.

To address this issue, the developed supervised machine learning (ML) techniques have predict the RUT of lithium iron phosphate (LFP) batteries installed in BTSs have used ML models and trained on data that has been extracted from power and environment (P&E) monitoring tool Net Eco(iManager NetEco data center infrastructure management system) .

The ML models can then be used to predict the RUT of a battery, which can help to ensure that batteries are replaced before they fail to deliver the designed capacity. In this study, three ML models were evaluated: linear regression, random forest regression, and support vector regression. The support vector regression model provided the best overall prediction performance, with a test error of 4.85%. This suggests that the support vector regression model is a promising tool for predicting the RUT of LFP batteries used in BTSs.

Keywords— CDC, DOD, RUT, QoS, QoE ….

Acknowledgment

First and foremost, I would like to thank my parents for their unconditional love and support for my dreams. They have always encouraged me to believe in myself since I was a child.

I would also like to take this opportunity to thank my advisor, Dr. -Ing. Dereje Hailemarimam for his continuous support and guidance on my TNE study.

My heartful appreciation to examiners member Dr. Bisrat Derebssa and Dr. Yalemzewd Negash for their invaluable feedback during Seminar 1 and Seminar 2 evaluation phases.

A special thanks to my friend Dr. Abebe Tamrat for his immense knowledge in machine learning and deep learning guidance and collaboration.

I am grateful to Mr. Gutema Tekon and Mr. Eyayew Werku for providing me an opportunity in collecting operational data .

I thank my beloved wife Roman Abebe for her love and support throughout my studies, especially in the period of my illness.

Last but not the least, I thank my brothers Tesfaye Hirpo, Capt. Berhanu Hirpo for filling my life outside schooling.

Table of Contents

Declaration ii

Abstract iii

Acknowledgment iv

Table of Contents v

List of Figures viii

List of Tables ix

List of Acronyms and Abbreviation x

Dedication xii

1. Introduction 7

1.1. Batteries in Cellular Networks 7

1.2. Impact of battery in cellular network 8

1.3. Facts about battery 9

1.4. Battery Models: From Analytical to Machine Learning 11

1.5. Statement of the Problem 11

1.6. Objectives of the Research 12

1.6.1. General Objective 12

1.6.2. Specific Objectives 12

1.7. Literature Review 13

1.8. Scope and Limitation of the Thesis 16

1.9. Contribution of the Thesis 17

1.10. Organization of the Thesis 18

2. Theoretical Background of Lithium-ion Batteries 19

2.1. Li-ion Batteries 19
2.1.1. History of Lithium-ion batteries 19
2.1.2. Fundamental Principles and Chemistry Behind Lithium-ion Batteries....... 21
2.2. General Properties of a Battery 24
2.2.1 . Voltage 24
2.2.2 . Capacity 25
2.2.3 . SOC and DOD 26
2.3. Machine Learning and Potential Application Areas 28
2.3.1 . Machine Learning Overview 28
2.3.2 . Machine Learning in Telecom 29
2.3.3 . Fundamental Principles of Machine Learning 29
2.3.4 . Data Preparation and Features 30
2.3.5 . Biase-variance Tradeoff 32
2.4. Regression Models 33
2.4.1. Linear regression 33
2.4.2. Random Forest Regression 34
2.4.3. Support Vector Regression (SVR) 34
2.4.4. Cross-Validation 36
2.4.5. Evaluation Method 37
2.4.6. Data Tools 39
3. Proposed Approach 40
3.1. Model Development from the Operational data 40
3.1.1. Data Description 41
3.1.2. Data Analysis 44

3.1.3. Domain-based Feature Generation ... 47
3.1.4. Labeling the target variable ... 48
4. Results and Discussion ... 50
4.1. Train-test and split ... 50
4.2. Discussion on models results ... 51
4.3. Feature contribution ... 54
4.4. Cross-validation ... 55
5. Conclusion and Future work ... 57
5.1. Conclusion ... 57
5.2. Future Work ... 58
References. ... 59

List of Figures

Figure 1.2: General View of Power System for BTS Operation [10]. 7

Figure 1.1: The charge or discharge process in an Li-ion battery [2]. 10

Figure 2.1: (a) and (b): Stanley Whittingham's and John B. Goodenough's Lithium-ion battery [15] 19

Figure 2.2: Akira Yoshino's Lithium-ion battery [15]. 20

Figure 2.3: Working principle of a Lithium-ion battery [15]. 23

Figure 2.4: Discharge voltage curves for different LiB cathode chemistries [15]. 24

Figure 2.5: Visualization of bias and variance [21]. 32

Figure 2.6: Illustration of k-fold cross-validation [22]. 37

Figure 3.1: Methodology and system model. 40

Figure 3.2: The charge voltage and charge current curve at 0.2C, 35ºC[https:Huawei.com]. .. 43

Figure 3.3: Shows the discharge performance at different discharge rate [https: Huawei.com]. 43

Figure 3.4: Capacity fade of 100 cycles in the P&E operational dataset. 45

Figure 3.5: Discharge capacity of 100 cycles in the P&E operational dataset. 45

Figure 3.6: Variance in discharge capacity 46

Figure 3.7: Ratio of change in capacity and discharge voltage. 47

Figure 3.8: Charging time at different operational cycles. 47

Figure 4.1: Predicted versus observed remaining useful time for SVR model. 52

Figure 4.2: Predicted versus observed remaining useful time for RFR model 52

Figure 4.3: Predicted versus observed remaining useful time for LR model. 52

Figure 4.4: Feature contribution by category using SVR model. 55

List of Tables

Table 2.1: Summary of cathode material [16]. 21

Table 2.2: Comparison of regression algorithm and their application area[22]........ 36

Table 3.1: Specification of LiFePo4 battery [https: www.huawei.com]. 42

Table 3.2: Correlation of discharge capacity at various cycles with other parameters. 45

Table 4.1: Evaluation of the three regression models using features by. 51

List of Acronyms and Abbreviation

AH	Ampere Hour
ANN	Artificial Neural Network
BMS	Battery Management System
BLVD	Battery Low Voltage Disconnection
BSC	Base Station Controller
BTS	Base Transceiver Station
CDC	Charging and Discharging Cycles
CM	Corrective Maintenance
CNN	Convolutional Neural Network
DC	Direct Current
DOD	Depth of Discharging
EU	Electric Utility
EMS	Element Management System
FFNN	Forward Feed Neural Networks
FMS	Fault Management System
GNB	Gaussian Naïve Bayes
K-NN	K-Nearest-Neighbors
KSVM	Kernel Space Vector Machine
LIB	Lithium Ion Battery
LLVD	Load Low Voltage Disconnection
LR	Linear Regression
ML	Machine Learning
MNN	Modern Neural Network
MSVM	Multi-Kernel SVM
NAR	Network Availability Rate

NetEco	iManager NetEco data center infrastructure management system
NMC	Lithium Nickel Cobalt Aluminum Oxide
NN	Neural Network
NOSM	Network Operation and Service Management
OSS	Operation Support Subsystem
PM	Preventive Maintenance
QoE	Quality of Experience
QoS	Quality of Service
RFR	Random Forest Regression
RBS	Radio Base Station
RNN	Recurrent Neural Network
RUL	Remaining Useful Life
RUT	Remaining Useful Time
RVM	Relevance Vector Machine
SEI	Solid Electrolyte Interface
SOC	State of Charge
SOF	State of Function
SOH	State of Health
SVM	Support Vector Machine
SVR	Support Vector Regression

Dedication

To my father **Hirpo Gurmessa Melka**, I dedicate this thesis to you, my biggest inspiration. You taught me the importance of hard work, perseverance, and never giving up on my dreams. I know that you would be so proud of me for completing this work, even though you are no longer here to see it.

1. Introduction

1.1. Batteries in Cellular Networks

Cellular network has been expanding rapidly and widely used by the public today. The communication signal service is provided by a base transceiver station (BTS) transmitter that directly serves to the user's handset involved in the telecommunications connection process. This operation of BTSs uses the main energy supply from electric utility (EU) that is converted to DC voltage by a rectifier [4]. To maintain the continuity of the electricity supply, BTS's are backed up by batteries as secondary energy. Furthermore, the availability of the BTS transmitter is expected to reach high availability value, and only a few percentages outage is allowed [9]. The below figure shows connection between the BTS and the battery.

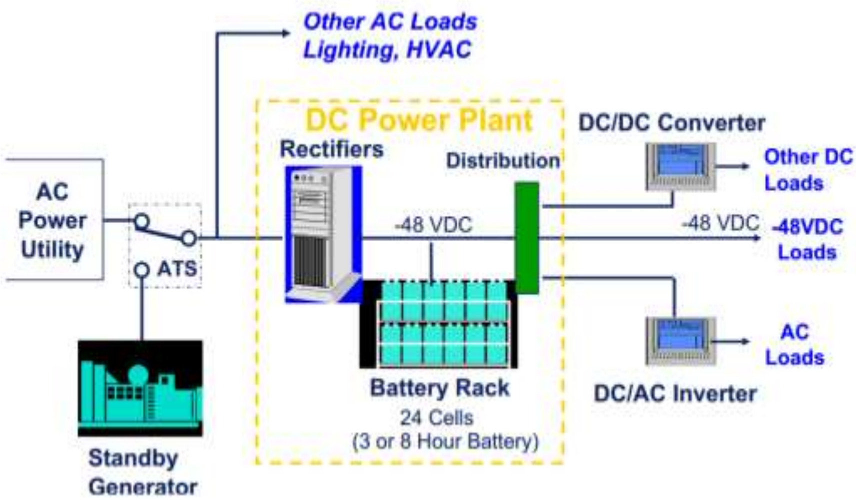

Figure 1.1: General View of Power System for BTS Operation [10].

A cellular BTS systems must be operated continuously in transmitting signals to the mobile stations. In contrary, this continuity is disrupted by many factors like mains failure, radio base station (RBS) problems, fiber cuts between base station controller (BSC) and BTSs. Among these, mains failure takes a lion share and contributes 67-70% of the total network outages [From National Network Operation Center Annual Summit]. Hence, to tackle such problem, an exhaustive preventive maintenance (PM) of the secondary power supply system should be carried out. To realize this maintenance activity, a prediction model that will tell battery’s remaining useful time (RUT) is important. In doing so, a better operation and maintenance activity can take place and as a result, the required QoS and QoE will be achieved.

1.2. Impact of battery in cellular network

Batteries play a vital role in cellular networks, particularly for Base Transceiver Stations (BTSs). BTSs are responsible for transmitting and receiving radio signals between mobile devices and the network. They are typically powered by grid electricity, but batteries are used to provide backup power in case of a power outage. Batteries also play an important role in enabling new cellular technologies, such as 5G. 5G networks require more power than previous generations of cellular networks, and batteries can help to provide this additional power. Overall, batteries have a number of important impacts on cellular networks, including:

- *Reliability*: Batteries provide backup power for BTSs in the event of a power outage. This helps to ensure that cellular service remains available even when there is a power disruption.
- *Coverage*: Batteries can be used to power BTSs in remote areas where there is no access to grid electricity. This helps to extend the coverage of cellular networks to more people.
- *Performance*: Batteries can be used to provide additional power for BTSs during peak usage periods. This helps to improve the performance of cellular networks during these times.
- *New technologies*: Batteries are essential for enabling new cellular technologies, such as 5G. 5G networks require more power than previous generations of cellular networks, and batteries can help to provide this additional power.

1.3. Facts about battery

Battery, as a secondary energy source, is a device that converts stored chemical energy into electrical energy. Each cell has a positive pole (cathode) and a negative pole (anode). A pole with a positive sign indicates that it has higher potential energy than a pole with a negative sign. A negative sign is a source of electrons that when connected to an external circuit will flow and provide energy to external equipment. When a battery is connected to an external circuit, electrolytes can move as ions inside it, so that chemical reactions occur in both poles. The transfer of ions in the battery will provide an electric current as an output that produces electrical energy [1].

Lithium-ion batteries are the most efficient type of battery for electrical systems, with a high energy and power density, a wide operating temperature range, a small size, a long lifespan, and fast recharging characteristics. They also have a low self-discharge rate [2]. In addition, lithium-ion batteries play an important role in optimizing the operation cost of energy storage systems in smart grids and micro grids [3].

Battery charge discharge cycle (CDC) is a process of discharging a battery from 100% to the lowest specified capacity limit. However, it is not recommended to discharge the battery until it is empty. Instead, a certain amount of charge, such as 10% or 20%, should be reserved, which is called the State of Charge (SOC). This helps to extend the battery life and improve its performance [3].

Maintaining certain SOC values is important to maintain the health of each battery cell . As the opposite of the SOC value, the value of the amount of battery discharge is usually referred to as the degree of discharge (DOD) which is presented as a percentage of discharge ampere-hour (AH) / Real Capacity [4]. The main factors of degradation are related to electrode corrosion and electrolyte degradation. Battery degradation is inevitable in practical applications, this is because of the irreversible electrochemical reactions inside an Li-ion battery [5].

Battery degradation under fast charging protocols cannot be captured by electrochemical models as this degradation is also due to battery materials [6]. There are four main components in an Li-ion battery, i.e., positive electrode (cathode), negative electrode (anode), conducting electrolyte, and separator as shown in *Figure 1.2*. An electrochemical reaction takes place to absorb energy when charging and release stored energy when discharging. *Figure 1.1* shows that when an Li-ion battery is charging, lithium (Li) ions (denoted as blue dots) move from the cathode through the separator to the anode. When an Li- ion battery is discharging, Li ions (denoted as red dots) move from the anode to the cathode through the separator. This charging and discharging continues until the battery reach its cycle life. It was reported in [7] that Li-ion batteries degrade due to the increase in internal resistance. The capacity degradation and increase in internal resistance affect the performance of these batteries [7].

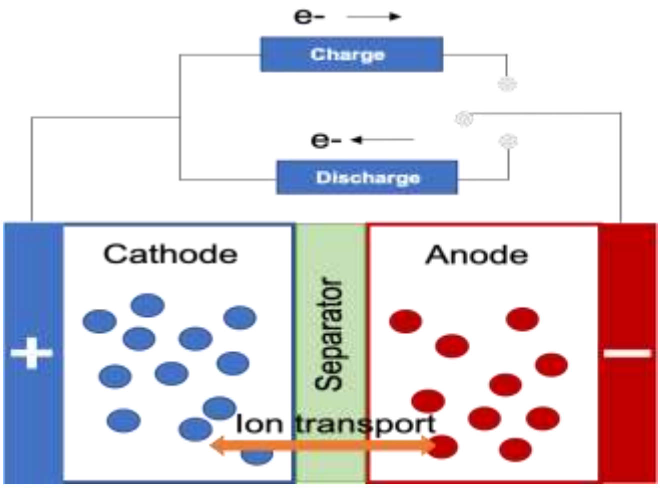

Figure 1.2: The charge or discharge process in an Li-ion battery [2].

Another factor which influences battery performance is an internal factor i.e., self-discharging which will reduce the battery capacity in a certain period [8].

In recent years, many studies on the degradation model and the early prediction of Li-ion batteries have been conducted. Various techniques and methods are proposed to improve the prediction accuracy in predicting capacity degradation,RUT and SOH of Li-ion battery. These techniques and methods can be divided into two categories, namely, model-based prognostics and data-driven prognostics [11].

1.4. Battery Models: From Analytical to Machine Learning

Model-based prediction requires a deep understanding of the composition of the model, in which mathematical expressions are used to describe the complex electrochemical process. Moreover, the explicit analysis of the solutions is difficult because the model is always non-linear due to lumped-element equivalent-circuit components, like resistors and capacitors, are used to represent the behavior of a battery cell [11]. To avoid such barriers, this research will only consider data-driven prediction approaches considering Li-ion battery as a black box without figuring out the complex electrochemical process in the model. Thus, history performance data of the LiFe Po4 battery should be explored to predict its RUT.

Algorithms or related parameters are the main factors that affect the accuracy of RUT prediction [11]. Since the accuracy of RUT prediction is a critical factor in ensuring the reliability and efficiency of energy storage systems (ESSs) , regression models have been preferred among different machine learning (ML) algorithms due to the following key features: Siutability for continuous target varables , interpretability ,versatility and work based on sound statistical principles. This will allow for more accurate and timely predictions of the RUT of ESSs, which can be used to improve the reliability and efficiency of these systems.

1.5. Statement of the Problem

In several cases BTSs are found failed due to several disturbances that occurred during operation. One of the disturbances is caused by electrical energy supply interruptions. ethio telecom has reported frequently that from the total network outage, the share of of the power utility outage is approximately 71%. Although the cellular BTS power supply system is usually equipped by battery as a secondary power supply, the reliability can be influenced by the health of the battery. Beside this, the RUT is not only affected by instability of main electricity supply but also temperature fluctuation of the battery environment, increment of battery aging, the number of CDC, and the number of DOD that occurred during operation.

There is a report called 'Battery Holding Capacity' is reported on a weekly basis by Network Operation and Service Management (NOSM) to functional team. This team will conduct maintenance activity by looking in to the bulk data and extract malfunctioning batteries in a traditional way due to the restriction imposed on the existing BMS in addressing the SOH and RUT functionality by Huawei . Hence, to resolve this problem a comprehensive prediction approach which will encompass key attributes of battery system is important. This will improve the quality of service for users while conserving network resources, Ehiotelecom can utilize predictive analytics along with insightful information about the performance of its network elements and make data-driven decisions by using operational performance data that can be extracted from the P&E monitoring tool . In doing so, the effectiveness of a battery system will be insured and this will have a positive impact on how satisfied the subscriber is with the offered services by keeping BTSs active in an on-demand basis and this in turn insures the required QoS, and QoE .

1.6. Objectives of the Research

1.6.1. General Objective

The general objective of this research is to develop and apply data-driven prediction approaches to predict the RUT of a battery system for conducting preventive maintenance in a scheduled manner to ensure the required QoS and QoE by reducing the downtime of BTSs due to secondary power system failure.

1.6.2. Specific Objectives

In order to accurately predict the remaining useful time of Lithium Iron Phosphate batteries, prediction model is formulated as a ML problem. Three ML models are considered, the support vector regression model, the linear regression model, and the random forest regression model. Furthermore, three metrics are used to evaluate the model performance. The objectives of this research are as follows.

- Obtain battery operational data of sites that are reported as low NAR and this will be used in construction of a dataset.

- Propose domain-based features that capture battery degradation.
- Identify, determine and discusses key (parameters), that are technically predictable and influencing the RUT of battery;
- Develop three ML models to predict the RUT using the domain-based features for the operational data that has been extracted from P&E monitoring tool.
- Implement three metrics to verify the performance of the three ML models.
- Draw some conclusions and recommendations.

1.7. Literature Review

Research on lithium-ion battery performance for BTS is limited compared to electric vehicles, as lack of standardized testing procedures for lithium-ion batteries in BTS applications ,BTS batteries are exposed to wider fluctuation in mains electricity supply [6] and the complexity of BTS systems makes it difficult to study the performance of lithium-ion batteries in isolation. The lack of research on BTS battery performance is a challenge worth exploring, as BTS batteries are critical infrastructure for telecommunications networks and their performance can have a significant impact on network reliability. Further research could develop more accurate models and monitoring methods, leading to improved reliability and efficiency. This would benefit both operators and end users.

Overall, the research on lithium-ion battery performance for BTS is still in its early stages. However, the results of the research that has been done suggest that lithium-ion batteries have a number of advantages over lead-acid batteries for BTS applications. As the cost of lithium-ion batteries continues to decline and the technology matures, it is likely that lithium-ion batteries will become more widely deployed in BTS applications in the future.

In [3], the authors propose a method for investigating battery health to ensure that the performance of BTSs meets the required level. The authors emphasize that the availability of the BTS transmitter is expected to be high, with only a few percent allowed to fail. Moreover, BTS systems must be operated continuously in transmitting signals to the mobile stations. To ensure the performance of BTSs, the battery system should work well during mains failure and this can be achieved by designing the MATLAB Simulink modeling, designing & running fuzzy logic function, and representing the result [3].From the analysis, it can be concluded that the prediction of effective capacity of BTS battery can be calculated by performing a fuzzy process of the input parameters in the form of current value criteria, temperature value criteria, voltage value criteria, and DOD value criteria and this will help in identifying the black spots as well as malfunctioning battery to ease the operation and maintenance work.

RUL prediction of Li-ion battery is crucial for the operation scheduling, spare parts management and maintenance decision for such kinds of systems are emphasized in [5]. Beside this, performance degradation, prognostics and RUL estimation is a focus area of the research concerned with Li-ion battery to ease maintenance decisions. The author describes methods that are helpful in achieving the above features are model-based approach, data-driven approach and hybrid approach. The hybrid approach, among the two, combines the advantages of model-based and data-driven approaches; as a result, it is effective, accurate and stable but demands high computational efficiency. For Li-ion battery, the key issues during RUL prediction include time-varying environment, random-variable current, self-healing feature, etc.Various factors caused by these issues have a significant impact on the accuracy of the prediction. Alongside, uncertainty has become a hotspot in the Li-ion battery RUL prediction. In the future, RUL estimation approaches for Li-ion battery can combine monitoring with prognostics techniques by considering RUL estimation under dynamic operating conditions [5].

Increasing the accuracy and efficiency of battery model is a hot research topic towards making battery state estimation, fault detection, monitoring and control tasks in [8]. A comprehensive study on the state-of-the-art of machine learning approaches on battery management system (BMS) is conducted in many literatures.Many scholars have proposed various models to estimate the SOC, SOH, RUL and state of function (SOF). However, inaccurate battery modeling is still exist [2]. The aforementioned are conducted based on Artificial Neural Networks (ANN) which comprises of Classic Neural Networks (CNN) &

Modern Neural Networks (MNN); a combination of the above methods have recently been investigated to improve their performances.Forward Feed Neural Networks (FFNN) and recurrent neural network (RNN), have been successfully applied to predict battery RUL .Among the neural network methods, RNN is considered to perform well comparing with SVM, relevance vector machine (RVM) provides comparable performance when using extremely sparsely defined kernel function. Support Vector Machine learning methods which are based on mathematical and statistical concepts are considered as reliable and convenient techniques to be used in BMS. The algorithms, that are evaluated for the diagnosis of battery cells are k-nearest neighbors (k-NN), logistic regression (LR), Gaussian naive Bayes (GNB), kernel Space Vector Machine (KSVM), and Neural Network (NN). Due to the variety of the available applications in terms of classification and regression in the BMS, each method can work well for a particular application and further studies are needed to reduce the computational complexity of machine learning approaches by means of different optimization techniques [2].

ML based methods have shown considerable promise on battery prognostics studies [12]. They are mechanism-free and do not require an explicit mathematical model to describe the complex nonlinear ageing dynamics of lithium-ion batteries and consider the battery as a black-box system and map a set of extracted battery features to the battery lifetime through various ML models [12]. Identified features from partial charging voltage curves, and utilized a least squares SVM model to estimate the battery SOH. Most ML-related studies first estimate the battery health indicators, such as the SOH, capacity, etc., and then predict the RUL by further calculations, such as by extrapolating the degradation model to a failure threshold [13].These existing methods have demonstrated satisfactory performance in the battery lifetime prediction; however, they usually perform battery lifetime predictions after accumulating 25% long term historical degradation data along the failure trajectory [14].

In summary,this research will use data-driven approach to predict the RUT of lithium-ion batteries used in cellular networks. The approach will use advanced sensor technology to extract operational data, which will be used to train an ML model that is selected based on its prediction accuracy, generalization ability, training time, and computational complexity.

To build the model, features such as capacity, state of charge (SOC), terminal voltage, temperature, and current should be extracted from the BMS and fed to the machine learning (ML) model. This can be done in a series of steps:

- *Characterizing battery parameters:* The battery parameters should be characterized to understand their behavior and how they relate to the RUT of the battery system.
- *Labeling the target variable:* The target variable, which is either the RUT or the SOH, should be labeled for each data point.
- *Building the ML model:* The ML model should be built using the labeled data points.
- *Evaluating the ML model:* The ML model should be evaluated to assess its accuracy in predicting the RUT of the battery system.

1.8. Scope and Limitation of the Thesis

The scope of this research is to develop a model that predicts the RUT of a battery that were deployed by Huawei in recent years which are located at Addis Ababa & Regional offices. The battery type which will be considered is Lithium-Iron Phosphate batteries that are connected as a secondary power source. This research will not only encompass sites which are deployed by other vendors battery system but also sites which were swapped to Huawei network are not considered due to functionality & integration problem with Net -Eco. Beside this, mains only sites are investigated for visualizing the impact of EU failure to determine the burden to which battery system exposed. Since the model is a data driven, it does not consider the electrochemical reactions and failure mechanisms inside Lithium-ion Battery(LiB). Along with this ,the performance of the model may degrade over time, as batteries age and experience different operating conditions.

1.9. Contribution of the Thesis

Prediction of the remaining useful time of a battery is a valuable tool for the telecom industry. By being able to predict when a battery is likely to fail, telecom operators can take steps to prevent outages and ensure that their networks remain reliable. Here are some of the contributions of predicting the RUT of a battery within the ESS for telecom industry are the following:

- *Improved reliability:* By knowing when a battery is likely to fail, telecom operators can take steps to prevent outages, such as by scheduling preventive maintenance or by replacing the battery with a new one. This can help to reduce the number of customer disruptions and improve the overall quality of service.
- *Reduced costs:* By knowing when a battery is likely to fail, telecom operators can plan for the replacement of the battery in advance. This can help to avoid the costs associated with unexpected outages, such as the cost of lost revenue and customer churn.
- *Optimized use of resources:* By knowing when a battery is likely to fail, telecom operators can optimize the use of their resources. For example, they can move traffic away from a degraded battery system and onto a newer, more reliable battery system. This can help to improve the overall performance of the network and reduce the risk of outages.
- *Increased efficiency:* By predicting the RUT of a battery, telecom operators can improve the efficiency of their operations. For example, they can use the predicted RUT to schedule preventive maintenance and to optimize the use of energy. This can help to reduce the operating costs of the network and improve the overall efficiency of the business.

1.10. Organization of the Thesis

This report is organized as follows:

Section ***1*** provides a brief overview of remaining useful time prediction for Li-ion batteries. The motivation of the research and related work were also discussed.

Section **2** describes theoretical background of Li-ion battery, basic principles of batteries operation along with their parts and the chemistory behind them .Finally, common terminologies are described and the ML model section discusses models that have been used in this work along with evaluation metrics and cross validation technique.

Section **3** introduces the development of ML models from operational data and discusses the following phases: Data description phases, data analysis, domain-based feature generations and finally, characterizing and labeling the target variable.

Section **4** result and discussion section briefly evaluates models' performance using the root mean squared error (RMSE) and mean absolute percentage error (MPE), and the results are discussed. Feature contribution is also explored along with partitioning of data (train and test) portions.

Section **5** concludes the report and provides suggestions for future work.

2. Theoretical Background of Lithium-ion Batteries

2.1. Li-ion Batteries

The theoretical information required to properly understand the work done with battery cycling data is covered in this part. First, a summary of the Lithium-ion battery's(LiB) development will be provided. Then, the basic principles of operation as well as the various parts of a battery and its functions are described. Finally, common battery terminologies and some commonly used terms will be discussed.

2.1.1. History of Lithium-ion batteries

Three of the main inventors of the LiB were awarded the chemistry Nobel Prize in 2019 [15]. Stanley Whittingham, the initially chosen recipient, served as the proof-of-concept. He developed the first operable LiB in the 1970s employing a layered titanium disulfide cathode and a metallic Lithium anode, as shown in *Figure 2.1:(*a). Through an electrolyte, Lithium ions were transported between the electrodes [15]

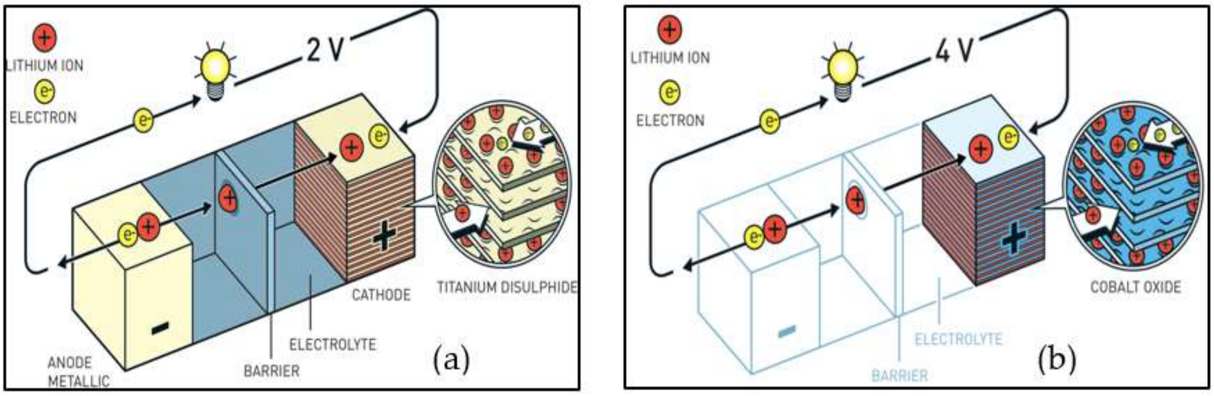

Figure 2.1: (a) and (b): Stanley Whittingham's and John B. Goodenough's Lithium-ion battery [15]

It was a significant development when John B. Goodenough improved on the idea because the battery voltage of 2 V is unsuitable for applications requiring for high power output. He believed that by swapping out the metal, the battery's voltage could be increased [15] . He developed a LiB with a cobalt oxide cathode that could reach a voltage of up to 4 V in 1980. The metallic Lithium anode, despite being a significant advance in battery power, had issues with safety and lifetime because the development of dendrites on the anode.

Akira Yoshino, the person who developed the first commercially successful LiB, was able to resolve this problem [15]. He utilized petroleum coke as an alternative anode in place of the battery's metallic Lithium anode. The material also assured that the batteries would last longer and was safer and lighter. *Figure 2.2* presents an illustration of the developed battery cell.

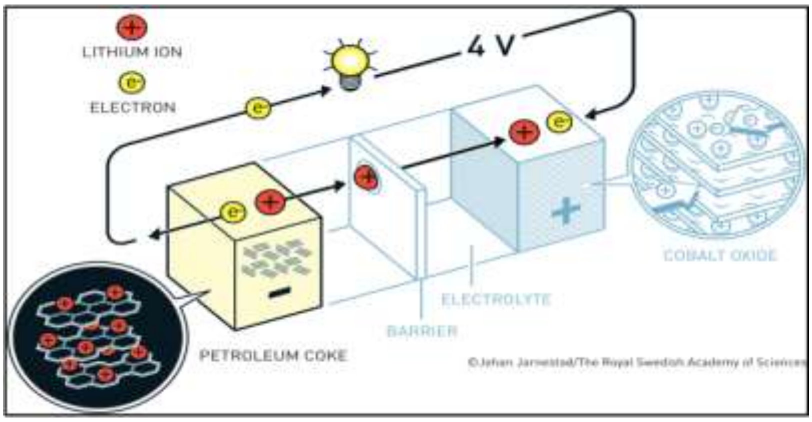

Figure 2.2: Akira Yoshino's Lithium-ion battery [15].

Modern Lithium-ion batteries (LiBs) still share the same general concept as the Akira Yoshino battery, despite the use of different materials. Natural or synthetic graphite is used in almost all commercial LiB anodes. However, there is also a lot of research on adding silicon to the anode to increase energy density, but the main challenge is that silicon expands much more than graphite. There is more variety in the materials used for cathodes.

The choice of cathode material for a LiB depends on the specific application [16]. Different cathode materials have different strengths and drawbacks, which must be carefully considered when making a selection. Nickel-based oxide cathodes, such as Lithium nickel manganese cobalt oxide (NMC) or Lithium nickel cobalt aluminum oxide, are the most popular choice for high-power applications where weight is crucial. These materials offer a high energy density and a high-power output, making them ideal for applications such as electric vehicles and power tools. However, nickel-based oxide cathodes are also more expensive than other cathode materials and can be less stable. Iron-based oxide cathodes, such as Lithium iron phosphate (LFP), are a more affordable and stable alternative to nickel-based oxide cathodes. LFP batteries offer a lower energy density and a lower power output than NMC or LiNiCoAlO2 batteries, but they are still suitable for many applications, such as stationary energy storage and low-power devices [16].

Other cathode materials that are being researched include manganese-based oxide cathodes and cobalt-free cathodes. Manganese-based oxide cathodes offer a high energy density and a long lifespan, but they are not as stable as iron-based oxide cathodes. Cobalt-free cathodes are being developed to address the high cost and limited availability of cobalt. Some promising cobalt-free cathode materials include Lithium-sulfur (Li-S) batteries and Lithium manganese oxide (LMO) batteries. The development of new cathode materials is an active area of research. The goal is to develop materials that have a high energy density, a long lifespan, and are safe and affordable. By carefully selecting the right cathode material for the specific application, Lithium-ion batteries can be optimized to meet the needs of a wide range of devices [16]. Here is a summary of cathode material along with their attribute as shows in *Table 2.1*.

Table 2.1: Summary of cathode material [16].

Description	LCO	LFP	NMC
Theoretical specific capacity (mAh/g)	274	170	280
Advantages	large specific capacity	good thermal stability long lifetime	high specific capacity long lifetime
Disadvantages	high cost poor thermal stability	low specific capacity	poor thermal stability

2.1.2. Fundamental Principles and Chemistry Behind Lithium-ion Batteries

There are two basic classifications of batteries: primary and secondary batteries. Primary batteries, such as the triple-A battery you put in your TV remote, can only be used once. Secondary batteries, which are rechargeable, are used in most larger applications, such as electric vehicles (EVs) and energy storage systems (ESSs) [15].All LIBs share some common components. These include two metal contacts that receive electrons and send them to an external circuit, a positive electrode and a negative electrode, an electrolyte that flows around the electrode particles, and a separator that keeps the electrodes apart from each other. Even though these names are only accurate during discharge and reversed during charge, the positive and negative electrodes are sometimes referred to as the cathode and anode, respectively [15]. The Lithium atom is the most basic component of the LiB. It has one electron in its outer shell because it has an atomic number of 3. Since this electron is only weakly attached to the nucleus, Lithium, or Li^2, is a very reactive metal. However, in a LiB, Lithium is combined with a metal oxide to form a stable compound. When an external voltage is applied to the LiB, the Lithium ions are delocalized from the cathode structure and flow into the electrolyte. The Lithium ions then diffuse through the separator and into the anode, where they intercalate, or enter between the layers of the graphite [15]. Since Lithium is a component of the cathode structure, the system's lowest energy is obtained in this condition, making it unstable.

When a Lithium-ion battery is discharged, Lithium ions migrate from the cathode to the anode through the electrolyte. This is because the anode has a lower concentration of negative charges than the cathode. When the battery is connected to an external load, electrons flow from the negative electrode (anode) to the positive electrode (cathode) through the external circuit. This flow of electrons provides power to the load [15].

The process of Lithium-ion migration and electron flow is reversible. When the battery is recharged, the Lithium ions migrate back from the anode to the cathode through the electrolyte [15]. This is because the cathode now has a lower concentration of negative charges than the anode. The flow of electrons in the external circuit is reversed, and the battery is recharged as shown in *Figure 2.3*.

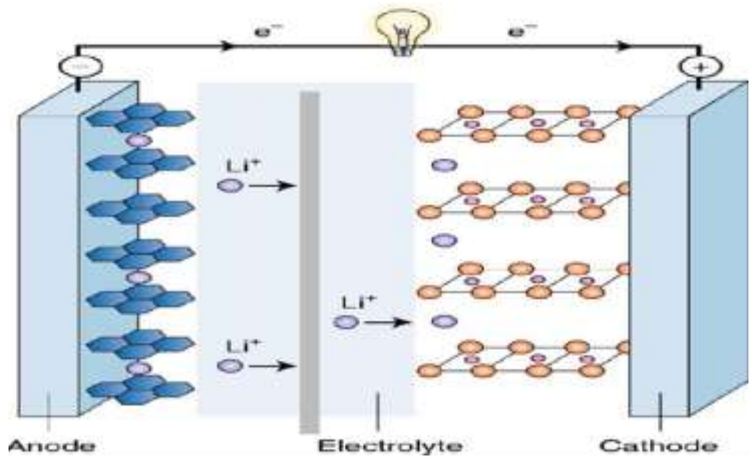

Figure 2.3: Working principle of a Lithium-ion battery [15].

The structure of the cathode affects the way Lithium is reduced. Layered cathodes, such as Lithium cobalt oxide (LCO), are made up of alternating layers of Lithium, cobalt, and oxygen. The Lithium atoms are positively charged because they have essentially donated their electron to the more electronegative cobalt and oxygen. This creates a weak repulsive force between the Lithium atoms. When a voltage is applied to the battery, the first Lithium atom is easily removed from the cathode. The remaining Lithium atoms are then more closely packed together and experience less repelling force. That is why the voltage required to charge the battery increases as the state of charge increases [15].

The process of Lithium reduction in layered cathodes is reversed during discharge. The initial cell voltage is high because the first Lithium atom is easily removed from the cathode. As the battery discharges, the voltage gradually drops because more power is needed to push Lithium into the structure. This is because the Lithium atoms are more closely packed together as the battery discharges. The discharge voltage curve of the NCA, a different layered structure, is shown in *Figure 2.4* in dark blue. The Lithium LFP cathode is another type of cathode structure. In this configuration, Lithium can only flow through the material in one dimension (as opposed to two dimensions in the layered structure). This means that LFP cathodes have a lower capacity than layered cathodes, but they are also more stable [15].

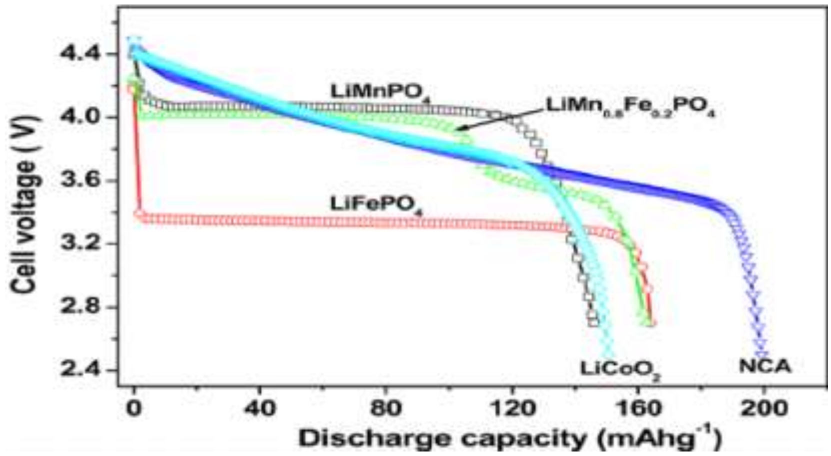

Figure 2.4: Discharge voltage curves for different LiB cathode chemistries [15].

In LFP cathodes, all Lithium atoms are subjected to the same conditions. This means that the voltage at which Lithium enters and exits the material is constant. This results in the flat voltage curve seen for LFP in red in above figure.

The flat voltage curve of LFP cathodes is due to the fact that Lithium can only flow through the material in one dimension. This means that the Lithium atoms are not as closely packed together as they are in other types of cathodes. This makes LFP cathodes more stable and less prone to dendrite formation. Dendrites are needle-like growths of Lithium that can form on the surface of the cathode during charging. Dendrite formation can lead to short-circuiting and fires. The flat voltage curve of LFP cathodes makes them a safer choice for applications where dendrite formation is a concern [15].

2.2. General Properties of a Battery

2.2.1 .Voltage

The voltage of a battery cell is a measure of the electrical potential difference between the two half-cells of the battery. Each half-cell consists of an electrode and the electrolyte that surrounds it. The potential difference between the two half-cells is created by a redox reaction, which is a reaction in which one substance is reduced, meaning it gains electrons, while the other is oxidized, meaning it loses electrons [15].

During discharge, the anode is oxidized and the cathode is reduced. This reaction must release energy in order for current to flow in the external circuit. The Gibbs free energy (ΔG) can be used to express the energy released by the reaction. The Gibbs free energy is a thermodynamic quantity that measures the energy available to do work. It is defined as the difference between the internal energy of a system and the product of the temperature and entropy of the system. The Gibbs free energy is always negative for a spontaneous reaction, such as the redox reaction that occurs in a battery cell during discharge [15].

The greater the negative value of ΔG, the more spontaneous the reaction and the greater the amount of energy that is released. The voltage of a battery cell is directly proportional to the negative value of ΔG and mathematically represented as follows:

$$\Delta G^0 = -nFE^0 \tag{2.1}$$

Where n is the number of electrons in the reaction, F is the Faraday constant (F = 96485 Cmol-1) and E^0 is the standard potential of the cell. The standard potential of a cell is the difference between the reduction potentials for the two half cells, expressed as.

$$E_0(\text{cell}) = E_0(\text{red cathode}) - E_0(\text{red anode}) \tag{2.2}$$

2.2.2 .Capacity

The amount of charge a battery can hold is measured by its capacity. In more concrete words, it is the maximum amount of current the battery is capable of discharging throughout a discharge cycle before the voltage falls below the cut-off value [15]. You can write the capacity as follows

$$C = i \cdot \Delta t \tag{2.3}$$

Where i is the current delivered by the battery during the time period Δt. It has the unit ampere- hours (Ah).

The quantity of active material in the positive and negative electrodes directly relates to a battery cell's capacity. A battery cell's capacity will decline over its lifespan as some active material is lost in irreversible reactions. Calculating the theoretical capacity is a standard practice when contrasting various battery materials. According to the amount of charge that may be stored per gram of active substance (gravimetric capacity) or per volume (volumetric capacity), are used. Since the active material cannot be used to its full potential in a real cell, the theoretical capacity should not be taken as the battery cell's actual capacity.It's customary to use the word "C-rate" to describe the current drawn from a battery. To normalize against various capacities, this is done. A C-rate of 1 C simply denotes the amount of current needed to completely drain the battery's capacity in an hour.

2.2.3 .SOC and DOD

State of charge (SOC) of the cells, which is defined as the available capacity (in Ah) and expressed as a percentage of its rated capacity. The SOC parameter can be viewed as a thermodynamic quantity enabling one to assess the potential energy of a battery. It is also important to estimate the state of health (SOH) of a battery, which represents a measure of the battery's ability to store and deliver electrical energy, compared with a new battery.
Depth of discharge (DOD) is the percentage of a battery's capacity that is discharged before it is recharged. For example, a battery with a capacity of 100 Ah that is discharged to 50 Ah has a DOD of 50%.The DOD of an Energy Storage System (ESS)is important because it affects the lifespan of the battery. A higher DOD will shorten the lifespan of the battery, while a lower DOD will extend the lifespan [3].

Technical principles:

The releasable capacity ($C_{releasable}$) of an operating battery is the released capacity when it is completely discharged. Accordingly, the SOC is defined as the percentage of the releasable capacity relative to the battery rated capacity (C_{rated}) given by the manufacturer.

$$SOC=\frac{C_{releasable}}{C_{rated}}*100\% \qquad (2.4)$$

A fully charged battery has the maximal releasable capacity(C_{max}), which can be different from the rated capacity. In general, $C_{max}$is to some extent different from C_{rated} for a newly used battery and will decline with the used time. It can be used for evaluating the SOH of a battery.

SOH = $\frac{C_{max}}{C_{rated}} * 100\%$ (2.5)

When a battery is discharging, the depth of discharge (DOD) can be expressed as the percentage of the capacity that has been discharged relative to C_{rated},

DOD = $\frac{C_{released}}{C_{rated}} * 100\%$ (2.6)

where $C_{released}$ is the capacity discharged by any amount of current

With a measured charging and discharging current (I_{b}), the difference of the DOD in an operating period (Ʈ) can be calculated by t_0

ΔDOD=$\frac{\int_{t_0}^{t_0+Ʈ} I_{\mathrm{b}}(t)dt}{C_{rated}}$*100% (2.7)

Where I_{b} is positive for charging and negative for discharging. As time elapsed, the DOD is accumulated

DOD(t)= DOD(t_0) + ΔDOD (2.8)

Without considering the operating efficiency and the battery aging, the SOC can be expressed as

SOC(t)= 100% – DOD(t) (2.9)

Considering the SOH, the SOC is estimated as

SOC(t)= SOH(t) – DOD(t) (2.10)

Accurate SOC estimation is the main tasks of battery management systems, which will help improve the system performance and reliability, and will also lengthen the lifetime of the battery. It is known that, precise SOC estimation of the battery can avoid unpredicted system interruption and prevent the batteries from being over charged and over discharged, which may cause permanent damage to the internal structure of batteries [3].

2.3. Machine Learning and Potential Application Areas

In order to provide the background information required to understand what has been done in this work, this part will discuss the basic ideas of machine learning (ML) and related issues. It starts with an overview of the various subfields, then moves on to a description of features and an introduction to data preparation. Next, the bias-variance tradeoff, a key ML challenges, is described. Following an explanation of the theory underlying a few distinct regression models, the support vector regression model that was employed in this study is presented. The section concludes with a definition of cross-validation and a discussion of several approaches used to assess ML models.

2.3.1 .Machine Learning Overview

Machine learning is a discipline that gives computers the ability to learn without being explicitly instructed. It is all about extracting knowledge from data. Machine learning algorithms are able to adjust and transform based on new data, enabling computers to extract valuable information from large amounts of data and identify correlations that would be impossible for humans to find on their own. For example, machine learning algorithms are used to recommend products to customers on Amazon, to predict customer churn at banks, and to diagnose diseases in medical images. Machine learning (ML) is a rapidly growing field with applications in a wide range of industries, including telecommunications. In cellular networks, ML can be used to improve the efficiency and reliability of power supply for base transceiver stations (BTSs) [17].

One way that ML can be used to improve BTS power supply is by optimizing the use of renewable energy sources. For example, ML can be used to predict the availability of solar and wind power, and then adjust the BTS power consumption accordingly. This can help to reduce the reliance on fossil fuels and improve the overall efficiency of the BTS power supply system [17].

2.3.2 .Machine Learning in Telecom

ML can also be used to improve the reliability of BTS power supply by detecting and mitigating faults in the system. For example, ML can be used to monitor the temperature of BTS batteries, and then predict when they are likely to fail. This information can then be used to take corrective action, such as replacing the batteries before they fail [18].

Overall, ML has the potential to significantly improve the efficiency and reliability of BTS power supply. By optimizing the use of renewable energy sources and detecting and mitigating faults in the system, ML can help to reduce operating costs and improve the overall performance of cellular networks [18].

Here are some specific examples of how ML is being used in the area of BTSs power supply:

- *Optimizing the use of renewable energy:* In a project by Huawei and China Mobile, ML was used to predict the availability of solar and wind power for BTSs in remote areas. This information was then used to adjust the BTS power consumption accordingly, which helped to reduce the reliance on fossil fuels and improve the overall efficiency of the BTS power supply system.
- *Detecting and mitigating faults in the system:* In a project by Nokia and Telefonica, ML was used to monitor the temperature of BTS batteries and predict when they were likely to fail. This information was then used to take corrective action, such as replacing the batteries before they failed. This helped to improve the reliability of the BTS power supply system and reduce the number of unplanned outages.

2.3.3 .Fundamental Principles of Machine Learning

The most popular approach to categorize the field of machine learning is into supervised learning and unsupervised learning, though there are other ways as well. In supervised learning, a model is developed using data for which the true value, also known as the response or target variable, is already known. The algorithm adjusts itself during the model's training phase to discover the logic in the data required to generate the known response variable. In the testing phase, the model will be given new data and asked to predict an unknown value. Unsupervised learning requires the model to discover patterns in the data on its own because the datasets do not contain a response variable [15].

Reinforcement learning, which is focused on rewarding desired conduct and punishing bad behavior, is the third category. The research is conducted on datasets that fall within the supervised learning category because the response variable is known. Therefore, further discussion of unsupervised and reinforcement learning will be ignored. In a supervised learning task, there are often pairs of training data points and response variables (x1, y1), (x2, y2), (xn, yn). The objective is to create a function f(x) = y that can anticipate what the value of y must be given a new data point of x [15].Based on the kind of prediction that is generated as an output, supervised learning can be further divided into categories.

The response variable in classification is one option from a list of class labels, which are predefined values. There are two types of classification: multi-class classification and binary classification. In contrast to multi-class questions, which can have a multitude of classes, binary classification problems only have two class labels based on the answer variability. This distinction is made due to many issues in the actual world are binary, such as "Is this email junk or legit?" or "Is a battery degraded or not"? The response variable in regression is often ($y \varepsilon \mathbb{R}$) and has a continuous value. For instance, this can involve forecasting a population's size or the direction and speed of the wind in the jungle [15].

This work considers the prediction of remaining useful time for batteries that have a secondary power system role, which is a continuous response variable. The type of machine learning that will be used is thus a regression.

2.3.4 .Data Preparation and Features

Rarely a large dataset is ideal and prepared to be fed directly into an ML model. Data cleaning is typically necessary to take into account discrepancies in the data. There are many various kinds of differences, including missing or duplicated numbers, inaccurate information, irrelevant information, and improperly formatted data. One of the more time-consuming steps in creating an ML model is the cleaning procedure, which can be tiresome.

Since important data may be lost, the procedure of eliminating unnecessary data is not straightforward. Instead, it calls for changing spelling, adding missing values, labeling, filtering, and many other things. Fortunately, there are lots of valuable tools that can help with this process [19].

ML models' input data are frequently referred to as features. In certain instances, they may be the complete collection of data, but most often they are representations of the data that have been explicitly created by humans to capture details that are required for the desired prediction. Large feature sets make ML models more likely to overfit and perform badly on unknown data.Additionally, it substantially raises the amount of memory storage needed and the cost of computation. On the other hand, a model that has too few features may not accurately represent all the significant characteristics of the data.Therefore, creating features that are descriptive of the dataset is an essential human role in ML, and it frequently involves knowledge of the industry for which the model was developed [19].

There are numerous techniques for reducing the quantity of attributes so that only the most important ones are included, in addition to human judgment. This practice is referred to as dimensionality reduction. Utilizing the Pearson correlation coefficient, which assesses the strength of the association between two variables, is a typical method [20].

It is written and defined as the covariance of two variables divided by the product of their standard deviation [20].

$$\rho = \frac{\mathrm{C_{\Theta}v(x,y)}}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})2}} \tag{2.11}$$

A number between -1 and 1 is always assigned to the correlation coefficient. -1 and 1 indicate perfect correlation, while 0 indicates no correlation at all. They can be ranked by highest score by computing for each feature, where x is the feature value and y is the observed response variable. The choice of features can then be made, for instance, by placing a limit on the total number and including those with the highest scores. The fundamental disadvantage of the correlation coefficient, while it is a straightforward and practical method for feature selection, is that correlation between the various features is not taken into account. Then, high-scoring but redundant features might replace other features that could capture additional data points [20].

2.3.5 .Biase-variance Tradeoff

When a model is created to approximate a function f(x) to a set of data, a trade-off must be made between the model's complexity, or how well the function approximates each data point, and its generalizability to additional data.

Bias is the error that results from the difference between the anticipated values and the related data. A model that predicts with good accuracy on a dataset is said to have a low bias. A set of datapoints and three distinct polynomial fits are shown below. The 30-degree polynomial on the right has a low bias since it tracks each data point very precisely. What happens, though, when this polynomial fit is presented with the-new data? Since it hasn't truly captured the trend in the data but has instead just followed every point, it will probably be unable to make reliable predictions. This is called overfitting. The variance describes how much forecasts differ from dataset to dataset around their average [21].

The 3rd degree polynomial fit in the *Figure 2.5* has a low variance since it is stiff and does not fluctuate significantly from dataset to dataset. However, it lacks any accurate predictions and has a strong bias.

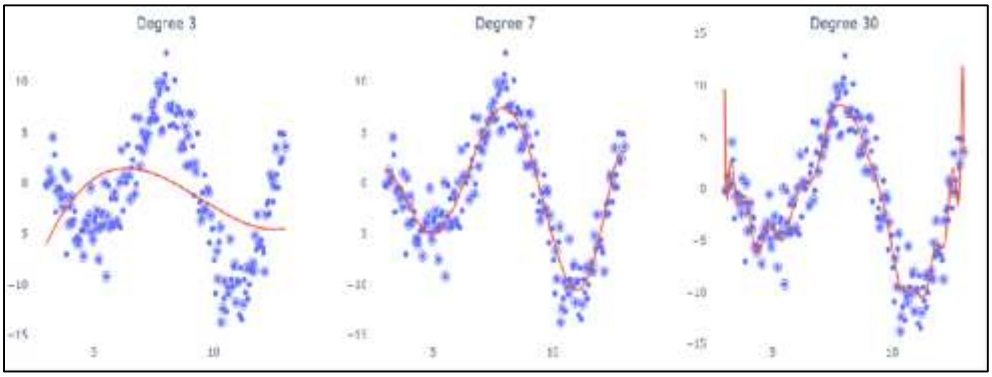

Figure 2.5: Visualization of bias and variance [21].

Underfitting is the term used to describe a model's inability to accurately represent the relationship between data points and the response variable. The 7-degree polynomial in the middle, which accurately reflects the trend in the data without losing the capacity to generalize to new data, has a decent balance between bias and variance. As can be seen, a regression problem's overall error can be broken down into three components: The irreducible error, bias, and variance are three important concepts in machine learning. Inherent data noise is the source of the variance, which cannot be taken into account in a model. The ratio of bias to variance in a model can be changed in a variety of ways, such as by the number of features it contains. Too few or too many features can result in overfitting or underfitting, respectively [21].

2.4. Regression Models

2.4.1. Linear regression

In practice, a class of models known as linear regression models is frequently employed [22]. They are popular in large part because they are simpler than other kinds of ML models. The following is a general formula for a linear regression:

$$\bar{y} = \omega_0 x_0 + w_1 x_1 + \cdots w_p x_P + b \tag{2.12}$$

Here, $\bar{y}$ is the predicted value, x_0 to x_P are the features used as input to the model, ω_0 to w_p are the weights assigned to each feature, and b is a constant. The values for $\boldsymbol{w}$ and $\boldsymbol{b}$ are what the model is changing as it gets trained. $\bar{y}$ and $\boldsymbol{xi}$ are vectors because they contain the values for each sample in the dataset. For example, in this work, $\bar{y}$ contain all the predicted cycle lives of the battery cells in the dataset, and each feature $\boldsymbol{xi}$ contain the feature value for each cell .

From the formula, it is clear that the fundamental unit of linear regression is a line, one for each feature that when combined, yields a prediction. While there are various varieties of linear regression, they all share this fundamental structure and vary in how **w** and **b** are discovered during model training.

2.4.2. Random Forest Regression

Random forest regression is a machine learning algorithm that builds multiple decision trees to make predictions. The values of the dependent and independent variables are passed to the random forest model. Each decision tree in the random forest makes a separate prediction. The final regression result is the average of the predictions made by each individual tree. In contrast, the output of a random forest classification algorithm is based on the mode of the class of the decision trees. This means that the algorithm chooses the class that is most common among the decision trees [23].

Linear regression and random forest regression are both regression algorithms, but they work in different ways. Linear regression uses a simple equation to predict the value of the dependent variable, while random forest regression uses multiple decision trees to make predictions. This makes random forest regression more complex than linear regression, but it can also make it more accurate. Despite having a similar concept, linear regression and random forest regression have different functions. $Y=wx + b$, where y is the dependent variable, x is the independent variable, **b** is the estimation parameter, is the formula for linear regression whereas the complicated random forest regression works similarly to a black box [22].

2.4.3. Support Vector Regression (SVR)

Support vector regression (SVR) and support vector machines (SVMs) are both supervised machine learning algorithms that can be used for both classification and regression tasks. However, there are some key differences between the two algorithms.

One of the main differences between SVR and SVM is that SVR has an additional tunable parameter called "epsilon" (ϵ). The value of ϵ controls the diameter of the tube that encloses the predicted function (hyperplane). Points that fall within this tube are considered to be accurate predictions, and the algorithm does not penalize these points. The support vectors are the points that fall on the edge of the tube, as well as any points that fall outside of the tube [22]

A simple way to think about SVR is to imagine a tube with an estimated function (hyperplane) in the middle and boundaries on either side defined by ϵ. The goal of the algorithm is to minimize the error by identifying a function that puts as many of the original points as possible inside the tube while at the same time reducing the "slack". The "slack" (ξ) measures the distance to points that fall outside of the tube. The amount of slack that the algorithm tolerates can be controlled by tuning a regularization parameter called C. A higher value of C will lead to a smaller amount of slack, which will result in a more accurate model but may also make the model more sensitive to outliers.

LR, SVR, and RFR are some of the most popular regression algorithms used in practice. There are a few reasons for this:They are all relatively simple to understand and implement, they are all very efficient, and they are all very versatile.In addition, each of these algorithms has its own specific strengths: LR is well-suited for linear regression problems, SVR is robust to outliers, and RFR is good at handling nonlinear regression problems.As a result of these advantages, LR, SVR, and RFR are widely used in a variety of applications, including machine learning competitions, natural language processing, computer vision, financial forecasting, and many more.

Overall, LR, SVR, and RFR are powerful and versatile regression algorithms that can be used to solve a wide range of problems. They are a good choice for many regression tasks, especially for beginners and those who need a simple and efficient solution.The below table shows a detail comparison of the three algorithms along with corresponding application areas.

Table 2.2: Comparison of regression algorithm and their application area[22]

Feature	SVR	Linear Regression	Random Forest Regression
Model	A set of hyperplanes	A linear equation	A set of decision trees
Accuracy	Can be more accurate than linear regression for non-linear data	Can be accurate for simple problems	Can be more accurate than linear regression for complex problems
Complexity	More complex than linear regression	Simple	More complex than linear regression
Interpretability	Difficult to interpret	Easy to interpret	Difficult to interpret
Outlier robustness	More robust to outliers than linear regression	Sensitive to outliers	Less robust to outliers than linear regression
Applications	Good for regression tasks where the data is not linearly separable	Good for regression tasks where the data is linearly separable	Good for regression tasks where the data is complex or has outliers

2.4.4. Cross-Validation

We typically divide the dataset into a training set, a test set, and occasionally a validation set when we train a machine learning model. Typically, a function will select random values from the data to accomplish this. A model's specific split isn't necessarily a good indicator of the data. If one is fortunate, the test data samples will be "easy" to predict, causing an artificially low-test error. If one is unlucky, the opposite scenario may occur [22].

Cross-validation is a method to split the data several times and for each split create a model and evaluate it. The most common way to do this is with k-fold cross-validation. This works by splitting the data k times into k separate folds, make a model, and evaluate it for each split. A k-fold cross-validation using k = 5 is illustrated in *Figure 2.6*.

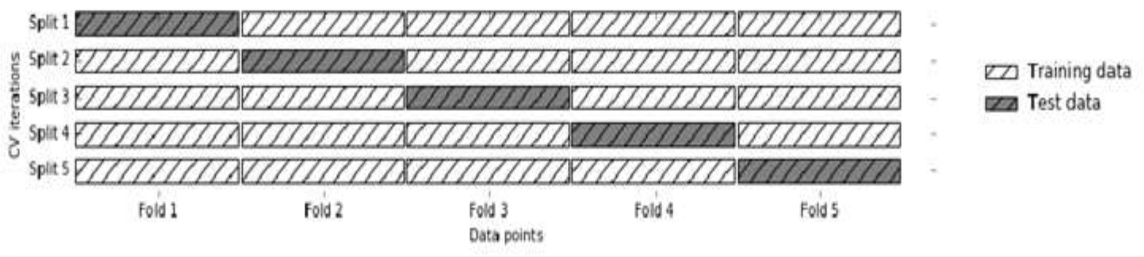

Figure 2.6: Illustration of k-fold cross-validation [22].

Because every value in the data is used exactly once in the test set, cross-validation has the advantage of detecting outlier values that, fortunately, were not included in the test set for the single split. It also provides an idea of the model's sensitivity to train/test splits. It is crucial to stress that because a new model is created for each split, this approach does not improve model training; rather, it merely provides insight [22].

Cross-validation can also be done for hyperparameter tuning. Instead of testing multiple train/test splits, it tests multiple combinations of hyperparameters and evaluates the error after each combination, returning the optimal choice.

2.4.5. Evaluation Method

To evaluate a regression model on a set of data, some way of quantifying how well each predicted value match with the corresponding observed value is required. The most widely used way to do this is by calculating the mean squared error (MSE) [24].

It measures the average squared distance between predicted and observed values, and is defined as:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y})^2 \qquad (2.13)$$

Where n is the number of samples, yi is the observed value for the ith sample and $\hat{y}i$ is the predicted value of the ith sample.

The difference between the observed and predicted value is referred to as the residual. Since the residual is squared, it is always a positive number.

The lower the number the better the prediction, and a MSE of 0 indicates a perfect fit. Taking the square root of the MSE can be done to obtain the root mean squared error (RMSE),

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2} \tag{2.14}$$

Which is a convenient method, since it results in a number with the same unit as the observation.

Another metric widely used is the coefficient of determination (R^2). It describes how much of the variation in yi can be described by the variation in $\hat{y}i$. Looking at the mathematical definition,

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2.15}$$

We have in the numerator the sum of squared residuals and in the denominator the total sum of squares, which is the deviation of each observation from the mean. The ratio then tells us the proportion of the variation in the predictions that cannot be described by the variation in observations.

Since we want to know the proportion that can be described by the predictions, we subtract from 1. When the residual approaches zero, the R2 becomes 1, which is a perfect fit.

The mean absolute error (MAE) is a similar evaluation metric to the MSE, but instead of taking the sum of squared residuals, it uses the absolute value, so the definition is

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}| \tag{2.16}$$

also known as the L1-norm. It is common to multiply by 100 to obtain the mean absolute percentage error, MAPE,

$$MAPE = 100 * \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}| \tag{2.17}$$

2.4.6. Data Tools

The steps provided in this chapter can be automated and simplified using a variety of tools available in Python programming. Pandas is a very helpful package for data preparation. Data is stored in containers known as DataFrames, and this is the core idea behind Pandas. Many different methods can be used to manipulate the data contained in a DataFrame object. It can be used for many different things, including data cleansing, graphing, sorting, and filtering.

The Scikit-learn library is another excellent resource for machine learning [25]. A variety of ML techniques that are easily implemented are included. The chosen hyperparameters are specified when the model is built. The model contains a variety of techniques for predicting and training on specific datasets. Other ML-related features such as cross-validation, hyperparameter optimization, and other metrics for errors are also included.

The above tools have been used to complete a data science workflow which includes the following steps:

- Importing data from a variety of sources
- Cleaning and transforming the data
- Visualizing the data to gain insights
- Selecting and training a machine learning model
- Evaluating the model's performance
- Deploying the model to FMS

3. Proposed Approach

This chapter describes the approach of developing machine learning models for operational data extracted from a power and environment (P&E)monitoring tool. The chapter is divided into two main sections. The first section describes data description, including specifications and how batteries have been cycled. The second section analyzes data and identify possible features for model input.The programming language used for this work was Python 3, using Pandas, NumPy, matplotlib, and built-in Python libraries for data processing and analysis. Machine learning(ML) modelling was implemented using Google Colab functionality. The workflow diagram shows the overall work process of the project.

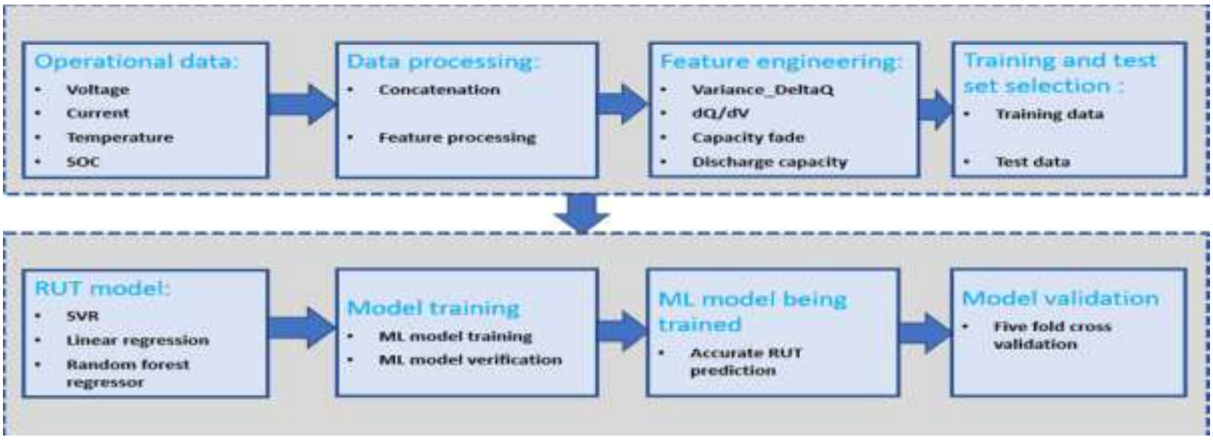

Figure 3.1: Methodology and system model.

3.1. Model Development from the Operational data

Data extracted from the advanced sensor technology did not adequately describe the degradation pattern of the operational batteries .This means that the data was not in a format that could be easily used to develop machine learning models to predict battery degradation. Therefore, the first step was to analyze the operational data in order to understand its structure and visualize the patterns of each variable against each other. This was done to gain insight into the batteries before developing the machine learning models. From the raw data files, a visualization tool was created to easily view the data and produce plots and figures using python, by utilizing Panda's library.

The second part of the data work was developing machine learning models to predict the remaining useful time (RUT) of the batteries. Three models were implemented and evaluated, all using the same set of features. The models were: Support vector regression (SVR), Random Forest Regression (RFR) and Linear Regression(LR).

The regression models was chosen as a model for three reasons: They have good prediction accuracy, easy to implement and interpretable; meaning that it was possible to understand how the models were making their predictions. The regression models were able to give valuable insights into how degradation manifests in ESS operational data. For example, the models were able to show that degradation is not a linear process, but rather it accelerates over time. Furthermore ,the models were also able to identify the specific features that are most important for predicting RUT.By considering the above key features of regression models,this work demonstrates the effectiveness of machine learning models for predicting the RUT of batteries. The regression model developed in this work is able to provide accurate, reliable, and interpretable predictions of RUT. This information can be used to improve the management and maintenance of batteries, which can lead to increased safety, reliability, and lifespan.

3.1.1. Data Description

The dataset used in this work consists of 10 lithium iron phosphate (LFP)/graphite batteries that are used as a secondary power supply.Batteries that have been selected for this research are categorized in to two groups .The first group of batterirs are located at Addis Ababa which have good functionality from the perspective of battery holding capacity evaluation and performance history from elemen management system(EMS) .The second group of batteries lie under the category of low NAR sites which were repeterdly reported due to lack of holding a load from the designed capacity and reported as a low network availability rate (NAR) sites .The batteries were subjected to different operational conditions, including varying charging C-rates since they are distribued all over the country and face different irregularities of mains supply.

Beside this,the provision period of all batteries were under the same period of deployement.*Table 3.1* provides the characteristics of the LFP batteries that were used to generate the dataset. The dataset includes four key data streams:

- *Capacity:* The amount of electrical charge that a battery can store.
- *Voltage:* The difference in electrical potential between the positive and negative terminals of a battery.
- *Current:* The rate at which electrical charge is flowing through a battery.
- *Operating temperature:* The temperature at which a battery is operating.

These data streams provide valuable information about the degradation of the batteries over time.

The dataset used in this work is a valuable resource for researchers and engineers who are developing new methods for predicting the remaining useful time of batteries. The dataset can also be used to develop new battery management strategies that can help to extend the lifespan of batteries.

Table 3.1: Specification of LiFePo4 battery [https: www.huawei.com].

Item		Description
Basic Parameters	Product model	ESM-48100B1
	Cathode material	$LiFePO_4$
	Nominal voltage	48 Vdc
	Nominal charging voltage	56.4 Vdc
	Max. charging / discharging current limited	100 A @ 35°C
	Max. charging / discharging power	4800 W
	Cycle life	3500 cycles @ 0.5C, 85% DOD, 70% EOL, 35°C
	Nominal capacity	100 Ah @ 0.2C, 35°C
	Weight	Approx. 43 kg
	Dimension (W×D×H)	442 mm×396 mm×130 mm (excluding mounting ear)
	Self discharge @ 25°C	Less than 5% after 90 days storage
	Communication interface	CAN / RS485; 2 dry contacts
	Max. quantity of parallel connection	CAN: 32; RS485: 16
	Terminal	M6, torque 4 N•m
	Installation type	Standard 19" rack, Air conditioning system or direct ventilation cabinet
	Operating condition	Air conditioner, direct ventilation in Class B environment
	Protection & alarm	Over temperature, overcurrent, short circuit, overcharge, overdischarge, etc.
	Certification	CE, UN38.3
	Design life	15 years

Key considerations from the datasheet:

1. Charging and discharging current may be derated or battery will be protected when

battery is out of the temperature range. *Figure 3.2* shows the charge voltage and charge current curve at 0.2C, 35ºC.

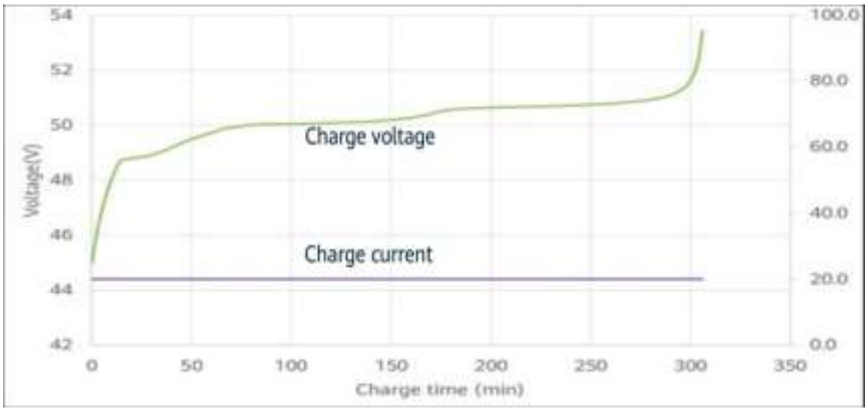

Figure 3.2: The charge voltage and charge current curve at 0.2C, 35ºC[https:Huawei.com].

2. When the output power of battery reaches to the maximum value, over temperature protection may be triggered, which shortens the battery discharging time. *Figure 3.3* shows the discharge performance at different discharge rate @ 35ºC

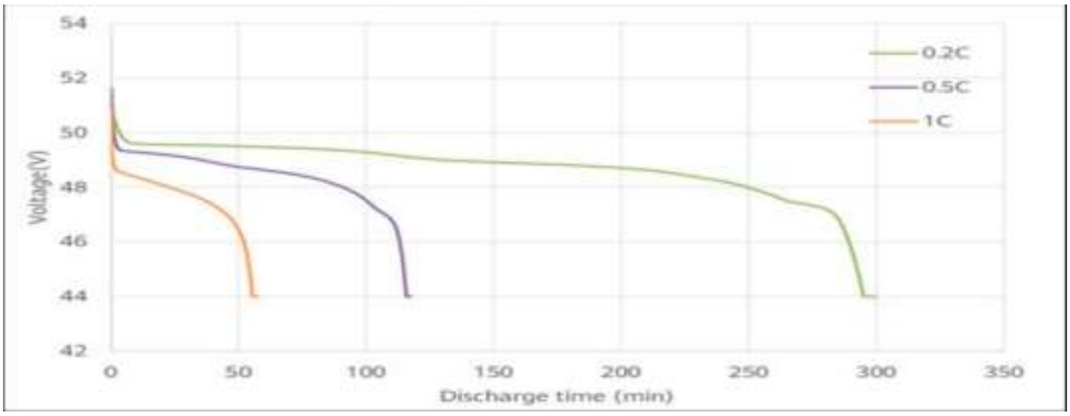

Figure 3.3: Shows the discharge performance at different discharge rate [https: Huawei.com].

3. Vmax and Vmin are determined by the BMS configuration.
4. Load low voltage disconnection (LLVD) and Battery low voltage disconnection (BLVD) effects are considered for modeling the working range.
5. The charge discharge cycle (CDC) count has been started at the time of extraction of data (i.e., without considering age related degradation and start up charging and

discharging pattern).

6. Frequency of mains outage has been taken from P&E monitoring tool to consider the extent to which the ESS has encountered a CDC.

3.1.2. Data Analysis

Figure 3.4 gives an overview of the capacity fade for all the batteries during the observation period. Two main trends can be observed from the capacity fade curves. The first trend is that battery degradation is typically insignificant in the early part of their operational time. This is because batteries are still relatively new and have not yet accumulated a significant number of charge-discharge cycles. However, as the batteries age and accumulate more cycles, the degradation rate starts to accelerate. This is due to a number of factors, including the growth of the solid electrolyte interphase (SEI) layer, structural degradation of the electrodes, and cyclic lithium loss.

The second trend is that the capacity fade trajectories of different batteries can cross each other as degradation sets in. This means that the rate of degradation of a battery is not necessarily constant over time. Some batteries may degrade more quickly than others, and the rate of degradation can also be affected by factors such as the operating temperature and the depth of discharge (DOD). This is verified by *Figure 3.5*. where it is attempted to correlate capacity at a given operational time with the remaining useful time. A linear trend would suggest a steady degradation towards every operational time, in which case remaining useful time prediction would be trivial and could just be extrapolated from early capacity fade.

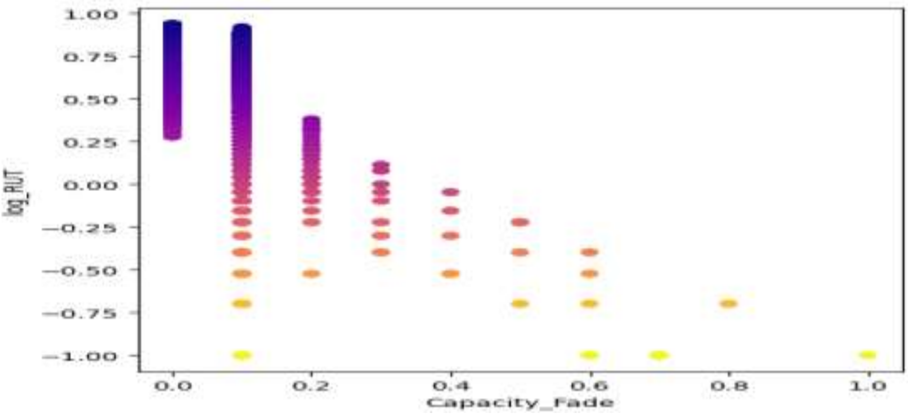

Figure 3.4: Capacity fade of 100 cycles in the P&E operational dataset.

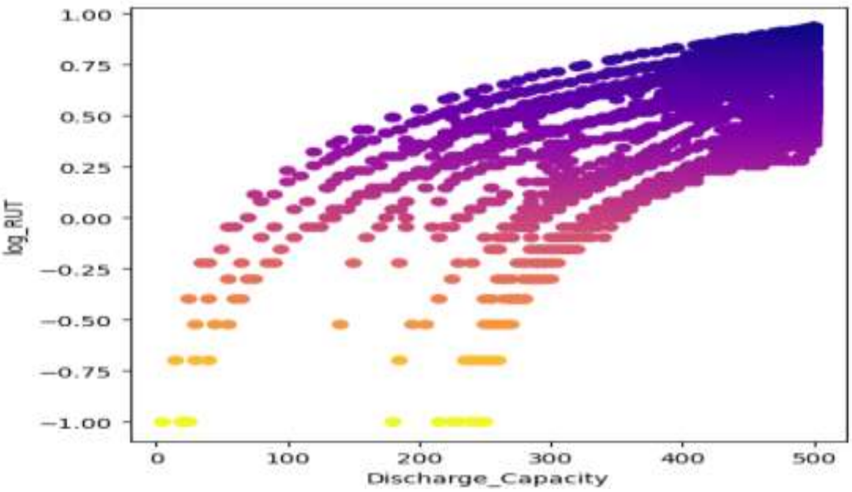

Figure 3.5: Discharge capacity of 100 cycles in the P&E operational dataset.

Table 3.2: Correlation of discharge capacity at various cycles with other parameters.

	Variance_DeltaQ	Discharge_Capacity	Capacity_Fade	dQ/dV	Charge_Time	Temperature_Integral
Variance_DeltaQ	1.000000	-0.853286	0.494137	-0.822732	0.038023	0.128783
Discharge_Capacity	-0.853286	1.000000	-0.711087	0.936724	-0.022326	-0.214344
Capacity_Fade	0.494137	-0.711087	1.000000	-0.649700	0.031448	0.246810
dQ/dV	-0.822732	0.936724	-0.649700	1.000000	-0.235601	-0.264505
Charge_Time	0.038023	-0.022326	0.031448	-0.235601	1.000000	0.215256
Temperature_Integral	0.128783	-0.214344	0.246810	-0.264505	0.215256	1.000000

To reinforce our model, in addition to capacity fade and discharge capacity at various cycles, there are other parameters that provide valuable input for predicting the remaining useful time, the battery cycling data is investigated further to see if information on degradation can be found prior to loss of capacity. Discharge capacity versus voltage-curves (discharge voltage curves for short) are known to possess useful information on degradation in batteries [15].

Two such curves can be viewed in *Figure 3.6* , where the cycle count is plotted against the variance in discharge capacity (variance_ΔQ) for batteries that have been selected for this research. In *Figure 3.7*, the ratio of change in capacity and discharge voltage (dQ/dV), is plotted. The trend suggests that the larger the value of dQ/dV, the larger the remaining useful time will be. Applying the variance to measure the spread of each curve reveals a linear relationship between variance in discharge capacity (variance_ΔQ) and RUT, especially when plotted on a logarithmic scale, shown in *Figure 3.6*. This shows that the discharge voltage curves possess important information about degradation of battery's terminal voltage, seeing this linear relationship also suggests that remaining useful time can be predicted with a linear regression model. Besides from the variance in discharge capacity, other features can be applied for capturing the degradation pattern like discharge voltage against discharge capacity (dV/dQ) which is a reciprocal of dQ/dV.

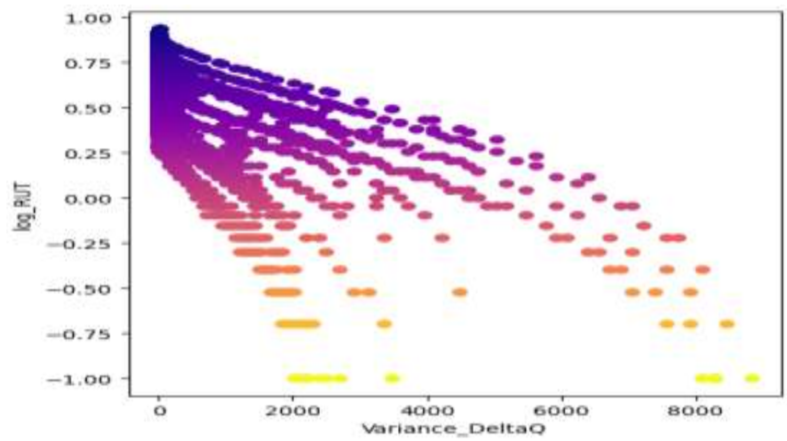

Figure 3.6: Variance in discharge capacity

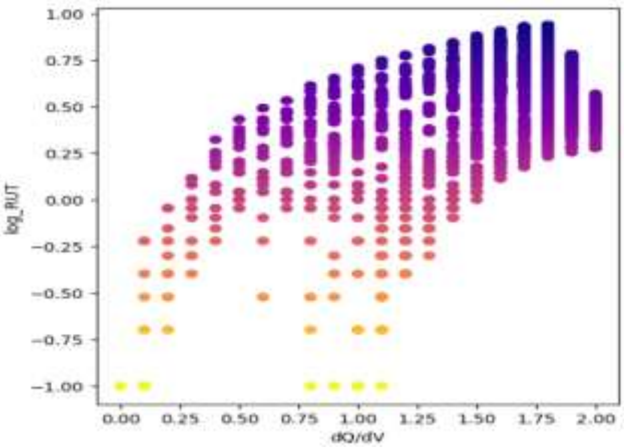

Figure 3.7: Ratio of change in capacity and discharge voltage.

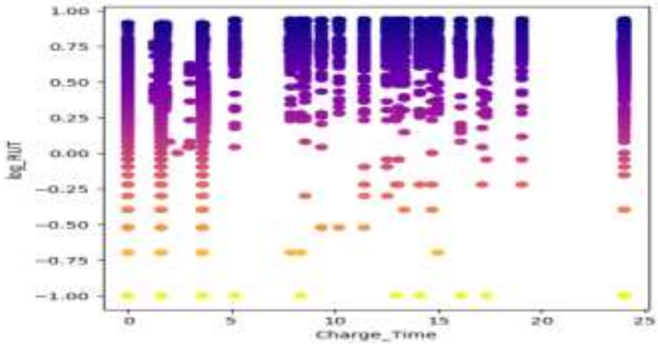

Figure 3.8: Charging time at different operational cycles.

3.1.3. Domain-based Feature Generation

Making a model that predicts the RUT of batteries with varying conditions would be a great achievement, since a more general model is desired in applications where the usage pattern is varying. This section presents the chosen features, which are based on the insight gained from the literature and characteristics of features against the target variable using P&E operational data. This set of batteries contain measurements of 2 months' operational data in a five minutes' basis in (i.e., voltage, current, temperature and SOC). These contain valuable information which is utilized to make features based on four categories:

- *Discharge voltage curves during the operational time*: These curves provide information about the voltage of the battery as it is being discharged. This information can be used to assess the health of the battery and also to predict how long it will last.

- *Differential capacity curves during the operational time:* These curves show the difference between the initial capacity of the battery and its capacity after each discharge cycle. This information can be used to track the degradation of the battery and to predict how much longer it will last.
- *Charging time during the operational time:* This information can be used to assess the efficiency of the battery charging process. It can also be used to predict how long it will take to charge the battery. *Figure 3.68* . shows the average time distribution during the operational time to fully charge the batteries from a depleted state. The charging time can vary depending on a number of factors, including the size of the ESS, the type of ESS, the charging source, and the charging rate.
- *Capacity fade curves:* These curves show how the capacity of the battery decreases over time. This information can also be used to predict the RUT of the battery.

3.1.4. Labeling the target variable

Based on the above categories, by visualizing the effect of different parameters against the target variable(i.e.characterization of input parameters against the target variable) as depicted in the above figures and the correlation among parameters discussed in *Table 3.3* , you can gain a better understanding of how the model works and how the parameters affect the model's output. This information can be used to label the target variable (RUT) and also improve model's performance.

Considering the above characteristics and correlation among each parameter, we can represent the below features in to the following form:

$\frac{dV}{dQ}$ (Differential voltage), $\frac{dQ}{dV}$(Differential capacity), Charging time, and Capacity fade can be related as a function for determining RUT.

The target variable will be formulated by encompassing the above key features as shown in the following way:

- RUT: It can be labeled as the time to voltage or the time to failure [until the cut off voltage reached].
- Here capacity as a function of voltage is represented to label the target variable [RUT]: dQ(V).
- dQ/dV: Is the differential capacity of a battery .
- $dV = Vt - Vcut\ off$ [LLVD]: Discharge voltage until cut off voltage reached.
- dQ=$Q - Qo = current\ capacity - previous\ capacity$.
- RUT~dQ= RUT is decreasing linearly with decreasing remaining capacity (dQ).
- RUT= dQ $*$ [$\frac{Vt - Vcutt\ off}{Vnominal - Vcutt\ off}$].

Finally, this mathematical formulation is fed to the dataset that has been constructed for predicting the remaining useful time of a battery under investigation.

4. Results and Discussion

In this chapter, the results from all models developed will be presented. All model results are included in the following form:

- An overview of the error metrics used to evaluate the models.
- A plot of predicted versus observed remaining useful time for each model.

The chapter is divided into three main sections based on the results and findings. The train-test and split are presented first. In the second section the brief discussion on models results along with prediction accuracy is presented. The third section shows a bar chart that shows the relative contribution of each feature has been added and discusses features contributions in detail. The fourth section describe how the models result have been validated using 5-fold cross validation .

4.1. Train-test and split

The regression models all share the same number of features, and all have a train- test split of the form   80% and 20% respectively and cross validated using 5-fold cross validation technique. SVR model, with the full set of features achieving a train and test RMSE of 0.825 and 0.67 , respectively(i.e. the average difference between the predicted RUT values and the actual RUT values is 1 hour). In contrast to other models RFR and LR, with a train and test RMSE of both with 1 and 1, respectively. Comparing the above results, it is noted that the SVR model got a much lower training error and a test error. This suggests that the SVR model is a best fit to the training set and test set amongst RFR and LR models. Here, from the configuration of battery holding capacity perspective , batteries are configured in hourly basis and most of them have configured as 5,6,7 and 8 hours of holding capacity .From this ,we can deduce that the holding capacity may vary but whether an RMSE of 1 is considered acceptable in the context of prediction of RUT of ESS depends on the specific application. For example, if the ESS is used to store energy for a critical application, such as a hospital, then an RMSE of 1 may not be acceptable. However, if the ESS is used to store energy for a non-critical application, such as a solar panel system, then an RMSE of 1 may be acceptable.

When summarizing the above results , the effect of dataset splitting on model performance was investigated. Random splitting with cross validation is used in this work with the same dataset for prediction of RUT using data for the operational time that ranges from Feb-2023 to March -2023. Hence using this technique, the MAPE and RMSE results obtained from the three models were evaluated. Based on this evaluation, SVR exceeds both LR and RFR models by 50% and 41.73% (i.e MAPE of SVR minus MAPE of LR and this will be divided by MAPE of LR). This will give a % age improvement of SVR over LR as well as RFR based on MAPE metric . The below table shows evaluation of the three regression models using three metrices .

Table 4.1: Evaluation of the three regression models using features by.

Algorithm	**Train**			**Test**			**Percentage Improvement of SVR**
	RMSE	MAPE	R^2	RMSE	MAPE	R^2	MAPE Improvement
Linear	1	8.99	0.86	1	9.04	0.85	50%
RF	1	7.62	0.88	1	7.66	0.88	41.73%
SVR	0.825	4.44	0.94	0.67	4.18	0.93	---

4.2. Discussion on models results

The predicted versus observed remaining useful time are plotted against each other in, *Figure 4.1-4.3* ,where a perfectly linear relationship between the two would indicate perfect predictions.

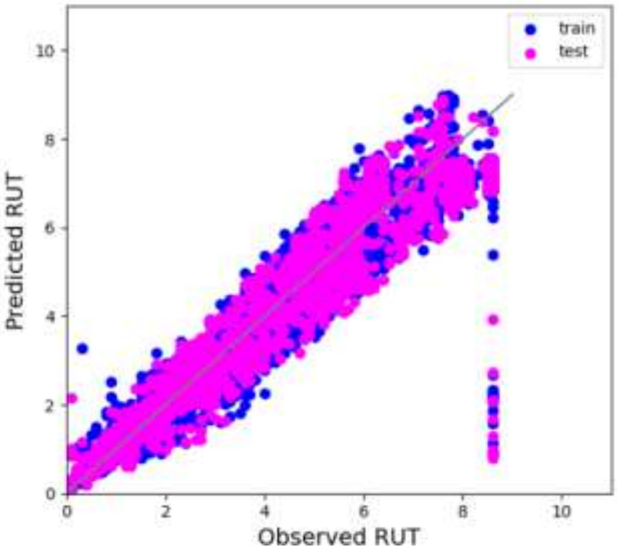

Figure 4.1: Predicted versus observed remaining useful time for SVR model.

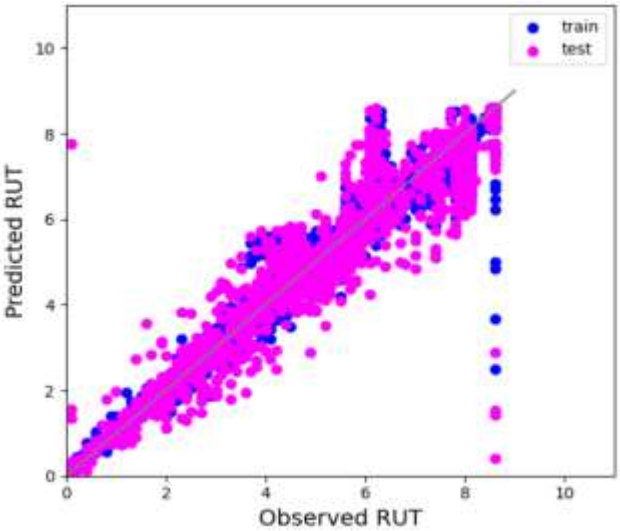

Figure 4.2: Predicted versus observed remaining useful time for RFR model

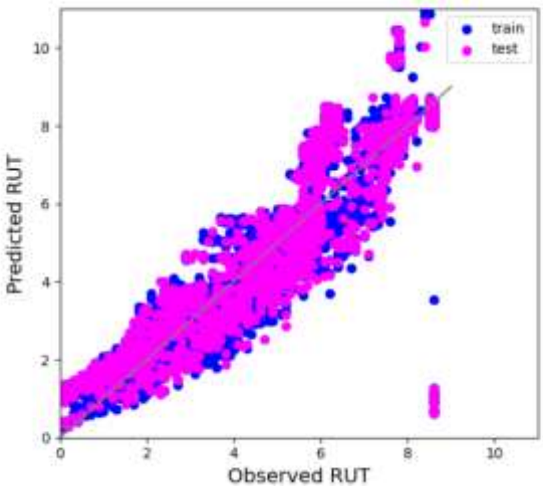

Figure 4.3: Predicted versus observed remaining useful time for LR model.

Model coefficients in the regression models reveal that capacity-related features have a high conditional dependence on the RUT of a battery. The correlation coefficients in Table 3.2 are a better measure of the actual predictive capability of each feature since they better show the marginal dependence. The discharge voltage curve features are clearly the highest performing parameters when looking at the correlation coefficents. Combining the insight gained from correlation coefficents and characteristics of parameters with respect to the target variable , it seems that both capacity and discharge voltage-related features are important for RUT prediction. A dQ/dV value, also known as differential capacity, provides information about the battery's operational condition. dQ/dV can also be a measure of the battery's capacity change across a relatively minor voltage shift. As we have already said above, a battery with a high dQ/dV value, shows a significant battery capacity increase with a relatively small change in voltage. From these observations, a dQ/dV tells the working status of a battery and it is another way to diagnose the functionality of a battery under a battery management system (BMS) platform.

Another is the log variance of ΔQ (V) is related to the dependence of the discharge energy dissipation on the discharge voltage, which is indicated by the blue dotted region between successive operational cycles as depicted in the *Figure 4.4* and the integral of this region is the total change in discharge capacity between cycles for a given voltage (V), which is also helpful in managing the discharge energy dissipation during the discharge process of a battery system. Beside this , Variance_ΔQ(V) is also a measure of how inconsistent the battery's discharge rate over time. A high Variance_ΔQ(V) indicates that the battery's discharge rate is inconsistent, while a low Variance_ΔQ(V) indicates that the battery's discharge rate is consistent.A consistent discharge rate is important for battery performance because it ensures that the battery can deliver the same amount of power over time. This is especially important for applications where the battery is used to provide backup power or to store energy for intermittent renewable energy sources which is feasible in our context.

4.3. Feature contribution

One of the most important things to understand in machine learning is the concept of feature contribution. Feature contribution is a measure of how important a particular feature is to the prediction of a machine learning model. It is calculated by measuring the change in the model's prediction accuracy when the feature is removed or modified. Feature contribution can be used to identify the most important features in a dataset, which can be helpful for model interpretation, feature selection, and improving model performance.

To analyze the above results further, each feature contribution shall be explored to extract main features out of eleven features. This technique is useful in minimizing the computational complexity of the model as well as operating time. From this technique, we can deduce that capacity-related features have the highest contribution in the model, with 50 % of the total weight(i.e. the waight of a parameter as depicted in the x-axis)with respect to 100% .More expectedly, the charging time-related feature has the second highest, with 35 % of the total weight. The remaining categories have negligible contributions. The two features with the clearly highest coefficients are:

- The charging time between operational cycles
- Discharge capacity between operational cycles.

These two features make up over 75 % of the model weights. What is particularly odd is that the charging time showed poor correlation with remaining useful time. Although this was not as expected, the relative contribution of each feature should not be directly interpreted as having the most impact on the predicted result. *Figure 4.5* shows the contribution of features in predicting the remaining useful time of batteries.

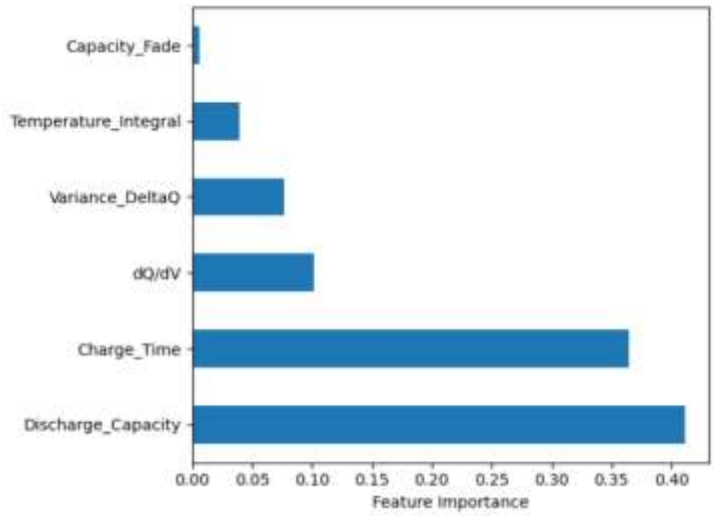

Figure 4.6: Feature contribution by category using SVR model.

4.4. Cross-validation

We have divided the input data into train and test datasets and use it for model training and testing respectively. This method is not very reliable as train and test data not always have same kind of variation like original data, which will affect the accuracy of the model. Cross validation solves this problem by dividing the input data into multiple groups instead of just two groups. There are multiple ways to split the data, in this research we have used 5-Fold cross validation technique. ***Error! Reference source not found.*** shows the inner working of K Fold cross validation:

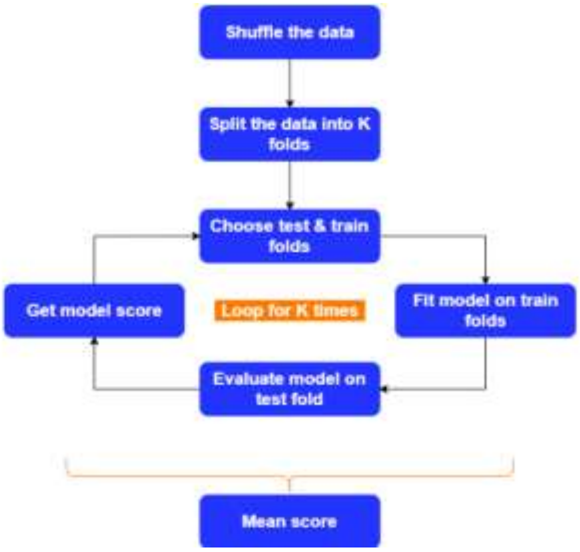

Figure 4.5: Inner working of K Fold cross validation [26].

In case of K Fold cross validation input data is divided into 'K' number of folds, hence the name K Fold. We have divided data into 5 folds i.e., K=5. Now we have 5 sets of data to train and test our model. So, the model will get trained and tested 5 times, but for every iteration we have used one-fold as test data and rest all as training data. Note that for every iteration, data in training and test fold changes which adds to the effectiveness of this method. This significantly reduces underfitting as we are using most of the data for training(fitting), and also significantly reduces overfitting as most of the data is also being used in validation set. 5- Fold cross validation helps to generalize the machine learning model, which results in better predictions on unknown data. The following **Error! Reference source not found.** shows 5-fold cross validation implementation to check model's performance and also help us in tunning the hyper parameter of SVR model.

Table 4.2: 5-fold cross validation implementation on SVR model.

Fold	**Training Set Size**	**Test Set Size**	**Mean Squared Error (MSE)**
1	60	40	0.0485
2	20	80	0.0486
3	40	60	0.0474
4	80	20	0.0487
5	0	100	0.0484
Average	**40**	**60**	**0.0483**

The SVR model was trained and evaluated using a five-fold cross-validation technique. The data was split into five folds, and the model was trained on four folds and evaluated on the remaining fold. This process was repeated five times, with a different fold being used for evaluation each time. The average mean squared error (MSE) across the five folds was 0.0483. This means that the SVR model was able to predict the values in the test set with an average error of 0.0483. By splitting the data into multiple folds and training and evaluating the model on each fold, cross-validation can give us a more accurate estimate of the model's performance on unseen data.

5. Conclusion and Future work

5.1. Conclusion

In this work, a data-driven approach was used to predict the remaining useful time (RUT) of 10 LFP batteries using operational performance data. The dataset constructed from 10 LFP batteries under different charging protocols defined by Huawei Technologies, one of the world's largest telecom operator.

Eleven domain-based features were considered, which fall into two categories: ΔQ (V) features and discharge capacity related to battery degradation. Six of these features were used as input to the support vector regression (SVR), linear regression, and random forest regression models. The log variance of ΔQ (V), discharge capacity, capacity fade, and dQ/dV were used in all models since they have a high correlation with RUT.

From model evaluation perespectives , the evaluation metrics that has been taken as a way of selecting models based on the error encountered during prediction of RUT was RMSE and MAPE . In this regard , SVR model achieved a train mean absolute percentage error (MAPE) of 4.4% and a train root mean squared error (RMSE) of 0.825% using the data for 60 days of operation. The random forest regression model achieved a train MAPE of 7.62% and a train and test RMSE of 1% using the data for 60 days of operation. Thus, the SVR model provided the best model performance with domain-based features. The results obtained confirm that machine learning (ML) modeling with domain-based features can be used to predict the RUT of Li-ion batteries and applied for predictive maintenance of ESS .

5.2. Future Work

For future work, feature selection shall be done using principal component analysis (PCA) to choose important features and improve model performance. The prediction results for RUT can be improved using a large amount of data along with optimization algorithms. In general, a combination of PCA and optimization algorithm will improve model's generalizability and predictability. Here are points shall be considered in the area of BMS that will improve the accuracy and efficiency of battery state prediction models.

Improved prediction using large datasets and optimization algorithms: larger datasets can provide more information for machine learning models, which can lead to improved performance. Optimization algorithms can be used to find the best parameters for a machine learning model, which can also lead to improved performance.

RUT prediction under dynamic operating conditions: RUT prediction is typically performed under static operating conditions, where the battery is subjected to a constant load. However, batteries are often subjected to dynamic operating conditions, where the load can vary over time. RUT prediction under dynamic operating conditions is a more challenging problem, but it is important for ensuring the reliability of battery systems.

Integrated prognostic framework: A prognostic framework is a system that combines monitoring and prognostics techniques. Monitoring techniques are used to collect data from a system, and prognostics techniques are used to predict the future state of the system. An integrated prognostic framework can be used to improve the accuracy of RUT prediction.

In the future, RUT prediction approaches for a battery system shall combine monitoring with prognostics techniques. The following aspects can be considered to achieve this functionality:

- RUT prediction under dynamic operating conditions.
- Integrated prognostic framework involving state monitoring, anomaly detection, intelligent diagnostics and prognostics.

References.

[1] Han, Xuebing and Lu, Languang and Zheng, Yuejiu and Feng, Xuning and Li, Zhe and Li, Jianqiu and Ouyang, Minggao, "A review on the key issues of the lithium ion battery degradation among the whole life cycle," *han2019review,* vol. 1, p. 100005, 2019.

[2] B. Balagopal, A. Alsabbagh, R. R. Ardeshiri,, C. Ma and M. Y. Chow, "Machine learning approaches in battery management systems: State of the art: Remaining useful life and fault detection.," in *In 2020 2nd IEEE International Conference on Industrial Electronics for Sustainable Energy Systems (IESES)*, 2020.

[3] M. B. Harijadi, T. Andromeda and M. Somantri, "Prediction effective capacity battery of cellular BTS for an effective maintenance," in *Materials Science and Engineering*, 2021.

[4] M. B. H. e. al, "Prediction effective capacity battery of cellular BTS," in *IOP Conf. Ser.: Mater. Sci. Eng. 1108 012013*, 2021.

[5] C. Su and H. J. Chen, "A review on prognostics approaches for remaining useful life of lithium-ion battery," in *Earth and Environmental Science*, 2017.

[6] D. Liu, Y. Luo, J. Liu, Y. Peng, L. Guo and M. Pecht, "Lithium-ion battery remaining useful life estimation based on fusion nonlinear degradation AR model and RPF algorithm," *Neural Computing and Applications,* vol. 25(3), pp. 557-572., 2014.

[7] L. Wu, X. Fu and Y. Guan, "Review of the remaining useful life prognostics of vehicle lithium-ion batteries using data-driven methodologies," *Applied Sciences,* vol. 6(6), p. 166, 2016.

[8] S. Prakash, W. E. Mustain and P. A. Kohl, "Performance of Li-ion secondary batteries in low power, hybrid power supplies," *Journal of Power Sources ,* vol. 2, pp. 1184-1189, 2009.

[9] "Green wireless network optimisation strategies within smart grid environments for Long Term Evolution (LTE) cellular networks in Malaysia," *alsharif2016green,* vol. 85, pp. 157--170, 2016.

[10] Pal, Nnoli Kosisochukwu and Ifeanyi, Bolu Jones, "Technical overview of all sources of electrical power used in BTSs in Nigeria," *International Research Journal of Engineering and Technology,* vol. 4, no. 2017, pp. 18--30, 2017.

[11] C. a. C. H. Su, "A review on prognostics approaches for remaining useful life of lithium-ion battery," *su2017review,* vol. 93, p. 012040, 2017.

[12] D. Gao and . M. Huang, "Prediction of remaining useful life of lithium-ion battery based on multi-kernel support vector machine with particle swarm optimization," *Journal of Power Electronics,* vol. 17(5), pp. 1288-1297, 2017.

[13] D. N. How, M. A. Hannan, M. H. Lipu and P. J. Ker, "State of charge estimation for lithium-ion batteries using model-based and data-driven methods," in *A review. Ieee Access, 7, 136116-136136*, 2019.

[14] Severson, Kristen A., et al., "Data-driven prediction of battery cycle life before capacity degradation.," *Nature Energy 4.5,* pp. 383-391, 2019.

[15] Dyve, Andreas, Cycle Life prediction of Lithium-ion Batteries Using Machine Learning, 2022.

[16] Liu, Wenmeng, Early Prediction of Battery Cycle Life Using Machine Learning, 2021.

[17] R. N. a. M. Mohammed H A, "Green wireless network optimisation," *Renewable Energy Journal,* vol. 85, pp. 157-170, 2016.

[18] Zubi G, Dufo-López R, Carvalho M and Pasaoglu G, "The lithium-ion battery: State of the art and future perspectives Renewable and Sustainable Energy Reviews," *Energy Reviews,* vol. 89, pp. 292-308, 2018.

[19] J. L. e. al, "Feature Selection: A Data Perspective," *ACM Comput. Surv,* vol. 50, p. 1–45, 2018.

[20] https://www.statisticshowto.com/probability-and-statistics/correlation-coefficient-formula/, "Correlation coefficient: Simple Definition, Easy Steps," *Statistics How To,* May 18 ,2022.

[21] https://www.kdnuggets.com/2021/08/visualizing-bias-variance.html, "Visualizing Bias-Variance," May 16,2022.

[22] A. C. Müller and S. Guido, "Introduction to Machine Learning with Python: A Guide for," *Sebastopol,* no. O'Reilly Media, Incorporated, 2016.

[23] A. C. a. G. S. M{\"u}ller, Introduction to machine learning with Python: a guide for data scientists, O'Reilly Media, Inc., 2016.

[24] F. Sohil, M. U. Sohali, and J. Shabbir,Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, "An introduction to statistical learning with," Vols. vol. 6, no. 1, no. Springer Science and Business Media, p. pp. 87–87, Jan. 2022.

[25] Scikit-learn:https://scikit-learn.org/stable/, "Scikit-learn: Machine Learning in Python," accessed May 16, 2022.

[26] S. P. Standardscaler, Available online: https: https://scikit-learn. org/stable/modules/generated/sklearn. preprocessing, standardscaler20available, accessed on 20 May 2022.

[27] X. Han, L. Lu and Y. Zheng, "A review on the key issues of the lithium ion battery degradation among the whole life cycle," in *ETransportation, 1, 100005.*, 2019.

[28] S. M. Rezvanizaniani, Z. Liu, Y. Chen and J. Lee, "Review and recent advances in battery health monitoring and prognostics technologies for electric vehicle (EV) safety and mobility.," *Journal of power sources,* vol. 256, pp. 110-124, 2014.

[29] H. Sak, A. W. Senior and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling.," 2014.

[30] Z. Xuegong, "ntroduction to statistical learning theory and support vector machines," in *Acta automatica sinica*, 2000.

[31] Varini, Maria and Campana, Pietro Elia and Lindbergh, G{\"o}ran, "A semi-empirical, electrochemistry-based model for Li-ion battery performance prediction over lifetime," *Journal of Energy Storage,* vol. 25, p. 100819, 2019.

[32] SvenSSon, Martin and S{\"o}derberg, JoakiM, "Machine-learning technologies in telecommunications," *svensson2008machine,* vol. 3, pp. 29--33, 2008.

[33] Fahrmeir, Ludwig and Kneib, Thomas and Lang, Stefan and Marx, Brian and Fahrmeir, Ludwig and Kneib, Thomas and Lang, Stefan and Marx, Brian, Regression models, Springer, 2013.

[34] Chen Y, Huang D, Zhu Q, Liu W, Liu C and Xiong N, "A new state of charge estimation," *Energies 10,* p. 1313, 2017.

[35] M B Harijadi*, T Andromeda and M Somantri, "Prediction effective capacity battery of cellular BTS for an," *IOP Conf. Series: Materials Science and Engineering* , 2021.