A Prognostic approach to alleviate range anxiety for e- Vehicles using Machine Learning.

A PROJECT REPORT

Submitted by

DK VENKATA HEM
K SAI SRINIVAS
K SANDHYA SUSMITHA

COURSE CODE: CS6611

COURSE TITLE: CREATIVE AND INNOVATIVE PROJECT

in partial fulfillment for the award of the

degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

DEPARTMENT OF COMPUTER TECHNOLOGY
ANNA UNIVERSITY, MIT CAMPUS
CHENNAI 600044

MAY 2023

DEPARTMENT OF COMPUTER TECHNOLOGY

ANNA UNIVERSITY, MIT CAMPUS

CHROMEPET, CHENNAI – 600 044

BONAFIDE CERTIFICATE

Certified that the bonafide thesis report of the project work done by Mr/Ms. DK VENKATA HEM, K SAI SRINIVAS, K SANDHYA SUSMITHA of Register Number 2020503007, 2020503015, 2020503541 in the Creative and Innovative Project Laboratory subject code CS6611 during the period February 2023 to May 2023

SUPERVISOR **COURSE IN-CHARGE**

ABSTRACT

Rapid growth in the transportation sector is contributing to pollution. The conventional vehicles which are made of fossil fuels emit harmful gases that result in global warming. In this scenario e-Vehicles can be chosen as a good alternative but range anxiety is something that is stopping people from resorting to e-Vehicles. Range prediction is a crucial aspect of electric vehicles (EVs) as it directly impacts their usability and practicality. With the increasing adoption of EVs, accurately estimating the range has become essential for optimizing driving plans and addressing range anxiety. Range prediction refers to the estimation of how far an electric vehicle can travel on a single charge based on various factors such as battery capacity, driving conditions, weather, and driving style. It involves the utilization of data-driven models and algorithms that consider these variables to provide an estimate of the remaining range.

The development of accurate range prediction models requires a deep understanding of EV technology, battery behavior, and real-world driving conditions. By providing drivers with reliable range predictions, they can make informed decisions, plan their journeys more effectively, and enhance their overall experience with electric vehicles. Here we propose a machine learning approach that combines physical and environmental factors to accurately predict range and therefore minimize range anxiety.

ACKNOWLEDGEMENT

We thank teaching and non-teaching members who belong to the Computer Center,Applied Science & Humanities of Madras Institute of Technology, Anna Universityfor shaping our initial years of college life.

We are highly indebted to our respected Dean, **Dr. J Prakash,** for providing sufficient facilities that contributed to the success of the project.

We express our deep sense of gratitude to our respected Head of the Department **Dr. P. Jayashree**, Professor and Head, Department of Computer Technology, and our guide **Dr. S. Neelavathy Pari**, Assistant Professor, Department of Computer Technology, Madras Institute of Technology, Anna University for her guidance, vision and constant encouragement throughout our project.

We thank our panel members **Dr. C. Valliyammai**, Professor, **Dr. S. Muthuraj Kumar**, and **Dr. P. Pabitha**, for their valuable commentsand their different views on our project. Finally, we thank all the teaching and non-teaching members of the Department of Computer Technology, library staff, friends, seniors, juniors, and everyone who have bestowed us with their knowledgedirectly or indirectly, intentionally, or unknowingly.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

SOC	State of charge
SOE	State of energy
SOH	State of health
RUL	Remaining useful life
LSTM	Long short-term memory
ML	Machine learning
EV	Electric vehicles
ANN	Artificial neural networks
MAE	Mean absolute error
ReLU	Rectified linear unit

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Transport is a fundamental requirement of modern life, conventional combustion engines that use petrol and diesel are highly polluting and contribute to global warming. These conventional vehicles can be replaced with electric vehicles that have zero tailpipe emissions and are much better for the environment. They have relatively lower running costs, maintenance cost, convenience of charging at home, and are easy to drive and quiet. There is an increase in usage of electrical vehicles lately. It is expected that 145 million electric vehicles will be on the road by 2030. Range prediction is a critical factor for EV users to plan their trips effectively and alleviate range anxiety. Calculating the range of electric vehicles (EVs) using Machine Learning (ML) techniques has gained significant attention in recent years.

ML algorithms leverage historical data from EVs, such as driving patterns, battery characteristics, and external factors like weather and road conditions. By analyzing this data, ML models can learn complex relationships and patterns that impact the vehicle's range. These models can then predict the remaining range based on real-time data inputs.

ML-based range prediction systems can continuously adapt and improve their accuracy by learning from a large amount of data collected from various EVs and diverse driving scenarios. These models can account for factors like battery degradation, traffic conditions, driving behavior, and energy consumption patterns, providing more reliable and personalized range estimations. The integration of ML in range prediction for EVs has the potential to enhance the driving experience, optimize route planning, and increase confidence in electric vehicle technology. As research and development in this area advance, ML algorithms can contribute to maximizing the efficiency and usability of EVs, promoting their widespread adoption and contributing to a sustainable transportation future.

1.2 OBJECTIVES

1. To design a machine learning model that can accurately predict SOC (State of Charge), SOE (State of Energy) for e-vehicle batteries.

2. To estimate and display the remaining coverable range for an electric vehicle using machine learning to reduce range anxiety among users.

1.3 SUMMARY

Range estimation for electric vehicles (EVs) using ML algorithms involves collecting relevant data like battery characteristics, speed, temperature, and more. ML models are trained using historical data to predict remaining range based on inputs. Calibrating models with real-time data enhances accuracy. ML algorithms continuously learn and adapt to new data, improving range estimation over time. Predictions can be integrated into the EV's user interface, enabling informed decisions on driving behavior and charging. This advancement aims to provide EV owners with more reliable and convenient range estimation, enhancing the overall user experience.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

This chapter explores the various methods proposed for predicting State of charge (SOC), State of energy (SOE) and remaining estimated range.

2.2 STATE OF CHARGE (SOC):

The state of charge (SOC) is a measurement of the amount of energy available in a battery at a specific point in time expressed as a percentage.

$$SOC\ (\%) = (Q + Q(initial)) / Q(max) * 100 \qquad (2.1)$$

Q(initial)= Initial charge of the battery.
Q = The quantity of electricity delivered by or supplied to, the battery. It follows the convention of the current: it is negative during the discharge and positive during the charge.
Q(max) = The maximum charge that can be stored in the battery.

The paper proposes [4] a method that combines the strengths of AdaBoost.Rt, a boosting algorithm that can improve the accuracy of weak learners, and RNNs, a type of neural network that can model temporal dependencies in time-series data. In this paper [4], the SOC of lithium-ion power batteries was estimated by using a recurrent neural network. The RNN used in this study consists of 6 input neurons, 1 hidden layer and 1 output neuron.

2.3 STATE OF ENERGY (SOE):

The "state of energy" in electric vehicles (EVs) refers to the current level of stored energy in the vehicle's battery pack. It represents the amount of electrical energy available for powering the vehicle and is commonly measured using the State of Charge (SOC). The state of energy in EVs is a crucial piece of information for both the driver and the vehicle's onboard systems. It helps determine the remaining driving range, plan charging intervals, and optimize energy management.

The conventional methods for estimating state of charge (SoC) generally suffer from the drawback of accumulative error due to incorrect initial SOC. However, the proposed [13] combined SOC and state of energy (SOE) estimation framework using a multi-layer feedforward neural network has been found to be more accurate and robust under dynamic driving and temperature conditions. Also, the Mean Square Error (MSE) obtained during testing with various drive cycles is quite promising.

2.4 STATE OF HEALTH (SOH):

State of health (SOH) is the capability of the battery to retain charge now compared to its rated value. It is defined as the ratio of the maximum battery charge to its rated capacity. It is expressed as a percentage as seen below.

$$SOH\ (\%) = (Q(max)/C(rated)) * 100 \qquad (2.2)$$

Q(max) = The maximum charge available of the battery

C(rated) = The rated capacity

The paper [3] identifies a wide range of features that can be used to predict battery RUL, including electrical and thermal characteristics, cycle life data, vibration data, and environmental factors. The paper shows that ML techniques can achieve high accuracy in predicting battery RUL, with some studies reporting prediction errors of less than 5%. The paper explains the following ML models - Gaussian progress regression, XGBoost, AdaBoost, Boosted regression trees, Support vector regression.

The paper [5] proposes a variant LSTM model that incorporates attention mechanisms to emphasize the most relevant features and learn complex temporal dependencies in battery data. The tests are performed using a dataset of 30 li-ion batteries. In the proposed model the input and forget gate are coupled by a fixed connection, which causes a simultaneous selection of old information and new data and the element-wise product of the new inputs and the historical cell states is conducted for screening out more beneficial information.

2.5 RANGE ANXIETY:

Range Anxiety refers to worry on the part of a person driving an e-vehicle that the battery will run out of power before a suitable charging point or destination is reached.

The study [6] identifies that range anxiety is a significant concern for EV users in India and also identifies several factors that contribute to range anxiety.The survey [13] highlights that range anxiety is a significant obstacle to the adoption of EVs, and that addressing this issue is essential for the success of the personal transportation revolution. The authors suggest that a range of strategies, including technological advancements, government policies, and public education efforts, could be employed to address this issue and promote the widespread adoption of EVs.

2.6 INFLUENTIAL AUXILIARY PARAMETERS:

Voltage, Current and Temperature form the prime role in estimating the remaining range of an EV. But the paper [8] emphasizes these are essential but not sufficient to arrive at accurate estimation. The authors [8] emphasize the importance of understanding the various factors that influence EV range and highlights the need for accurate range prediction. A Critical Review of Factors considered in [12] show that, many factors, viz. vehicle characteristics, driving patterns, environmental conditions & battery aging, battery degradation can impact range over time. Factors such as temperature, charging patterns, and battery chemistry can all contribute to battery degradation.

2.7 REMAINING RANGE ESTIMATION:

Remaining range refers to the distance that the e-vehicle can still reach with the current charge or without recharging. The approach proposed in [11] uses data from the EV's BMS, GPS, and other sensors to develop a model for the vehicle's power consumption and range in real-time. The model is then updated continuously as new data is collected, allowing for more accurate predictions of the remaining range. The study made in [10] finds that range can depend on several factors, including the type of machine learning algorithm used, the input data used to train the algorithm, and the conditions under which the prediction is made. A ML based approach is used in [1] to find range, based on the identification and forecast of driving conditions. In addition to the driving conditions, model in [9] considers factors such as traffic conditions, and ambient temperature. The paper [2] proposes a hybrid machine learning mode by combining the RT and SOM algorithms. The input features (Trip - Feature vector) are fed into the hybrid model for training. Model mentioned in [7] involves a combination of physical and data-driven models which considers various factors such as vehicle speed, road gradient, air resistance, tire friction, SOC etc...

2.8 SUMMARY

Table 2.1 – Literature survey summary

REF NO.	Author	Topic	Method Used	Learnings
1	Zhigang He a, Xiaoyu Shen a, Yanyan Sun b, Shichao Zhao b, Bin Fan c, Chaofeng Pan	State-of-health estimation and remaining useful life prediction for lithium-ion battery based on a variant long short term memory neural network	The proposed method is a variant long short term memory neural network (AST-LSTM NN) for state-of-health estimation and remaining useful life prediction of lithium-ion batteries.	Long short term memory neural network (AST-LSTM NN) actively tracks cell states and screens out beneficial information using element-wise product of new inputs and historical cell states.. The model has a many-to-one structure and can accurately predict multi-step RUL.
2	Bohan Zheng, Peter He, Lian Zhao, Hongwei Li	A Hybrid Machine Learning Model for Range Estimation of Electric Vehicles	The proposed method is a modified Self-Organizing Maps (SOM) integrating Regression Trees (RT) to predict the power consumption of EV trips.	Experimental results, including both cross-validation and mathematical accuracy measuring criteria, demonstrate that our Hybrid Model could not only provide a better power consumption

				estimation of EV trips but also reveal the inherent nature of the EV Big Data. Favourable results are produced not only for short trips but more importantly, the prediction accuracy is satisfactory for long trips.
3	Prabhakar Sharma,Bhaskor J. Bora	A Review of Modern Machine Learning Techniques in the Prediction of Remaining Useful Life of Lithium-Ion Batteries	The investigators have extensively used adaptive neuro-fuzzy inference systems (ANFIS), regression trees (RTs), artificial neural networks (ANN), response surface methodology (RSM) and gene expression programming (GEP).	The paper discusses modern machine learning techniques used for predicting the remaining useful life (RUL) of lithium-ion batteries. It also provides insights into the advantages and limitations of each method and their potential applications in real-world battery management systems.
4	Ran Li, Hui Sun, Xue Wei, Weiwen Ta, Haiying Wang	Lithium Battery State-of-Charge Estimation Based on AdaBoost.Rt-RNN	The paper proposes an integrated learning algorithm of AdaBoost.Rt cyclical neural network model for state-of-charge (SOC) estimation of	The model addresses the low accuracy and poor generalization of neural network algorithms in SOC estimation research by using a chain-connected recurrent neural network

			lithium batteries.The AdaBoost.Rt ensemble learning algorithm is then used to construct a multi-RNN model integration method, with the RNN model as the base learner and AdaBoost.Rt-RNN strong learner model.	(RNN) model to address the correlation adaptability of sample data in the spatio-temporal dimension. Therefore, this proposed model can be useful for improving SOC estimation accuracy in lithium batteries under various complex operating conditions.
5	Ganesh Sankaran, S. Venkatesan, M. Prabhahar,	Range Anxiety on electric vehicles in India -Impact on customer and factors influencing range Anxiety	This paper discusses the issue of range anxiety in electric vehicles in India and its impact on customers. It explores the technical factors that contribute to range anxiety and the latest technologies and strategies available to address it.	Some of the strategies used to address range anxiety include increasing the battery capacity, improving the charging infrastructure, implementing smart charging solutions, and providing real-time information on the vehicle's range. The paper also discusses the use of range extenders, such as fuel cells or internal combustion engines, to increase the vehicle's

				range.
6	Llyès Miri, Abbas Fotouhi, Nathan Ewin	Electric vehicle energy consumption modelling and estimation—A case study	The paper proposes an accurate modelling approach for electric vehicle (EV) energy consumption estimation. The goal is to model the target EV, including its powertrain system and longitudinal dynamics, and then validate it using available data.	The current range estimators work on the basis of vehicle's historical data analysis and are therefore not very accurate. For range estimation, an accurate model of the EV's energy consumption is essential. Such a model can be implemented in EV range estimators to assess the energy consumption of any EV model.
7	Emilia M. Szumska and RafałS. Jurecki	Parameters Influencing on Electric Vehicle Range	The authors analyzed data from a real electric vehicle and used big data analysis to estimate the range of an electric vehicle. The big data analysis involved a mathematical model that examined large data sets to discover certain patterns, correlations, or trends	The article provides a detailed explanation of the parameters that affect the range of an electric vehicle. The range of electric vehicles generally depends on three main classes of factors: vehicle design, driver’s driving style and use conditions. Some parameters are invariable, e.g., vehicle type, gearbox type,

			that are used to estimate some assumed variable.	number of seats, weight of electric drive, weight and type of battery, road infrastructure, and availability of battery charging infrastructure. Other parameters such as the battery's state of charge (SOC), state of health (SOH), driver's behavior, traffic volume, weather factors, etc. are subject to change.
8	Jun Bi , Yongxing Wang, Qiuyue Sai, Cong Ding	Estimating remaining driving range of battery electric vehicles based on real-world data: A case study of Beijing, China	The nonlinear estimation models used for remaining driving range under different temperature conditions were established based on the data-driven method. The models consider the State of Charge (SOC), speed and temperature conditions as the impacting factors for remaining driving	The study highlights the importance of considering the practical travel situation in developing countries, such as Beijing, when estimating the driving range of BEVs, as the traffic condition and external environment in these cities are significantly different from those in developed countries.

			range	
9	JOUR, Sun, Shuai, Zhang, Jun, Bi, Jun, Wang, Yongxing	A Machine Learning Method for Predicting Driving Range of Battery Electric Vehicles	The proposed method is a new driving range prediction method that uses the gradient boosting decision tree (GBDT) algorithm as a machine learning technique. This method includes a large number of feature variables and takes into account real-world working conditions, battery status, and traffic environment to improve both the applicability and accuracy of the prediction.	This is a novel method for predicting the driving range of battery electric vehicles using machine learning. Specifically, the gradient boosting decision tree to develop a prediction model that takes into account a large number of feature variables, including real-world working conditions, battery status, and traffic environment. This method was found to be more accurate and reliable than conventional regression methods. The study has important implications for the future of sustainable transportation and could be useful for automotive manufacturers and policymakers.
10	J. Hong, S. Park	Accurate	The paper proposes a	the importance of accurate

	and N. Chang	remaining range estimation for Electric vehicles	hybrid modeling methodology for accurate remaining range estimation in electric vehicles. This methodology starts with a conventional physics equation-based model and then incorporates machine learning techniques to improve the accuracy of the power consumption model. The paper focuses on improving the accuracy of the power model, which is one of the two consecutive steps in a model-based remaining range estimation process.	remaining range estimation for electric vehicles and how it can help mitigate range anxiety. The paper proposes a hybrid modeling methodology that integrates conventional physics equation-based models and empirical models to improve the accuracy of the power consumption model. The proposed methodology is demonstrated to be superior to state-of-the-art remaining range estimation practices. By improving the accuracy of remaining range estimation, EV drivers can trust their vehicles more and potentially increase their driving range by up to 30%.
11	M. Smuts, B. Scholtz and J.	A critical review of	This paper provides a critical review of	From the paper one can understand the key factors

	Wesson	factors influencing the remaining driving range of electric vehicles	existing literature to identify the key factors that influence the RDR and suggest that accurate RDR estimations can reduce range anxiety for drivers, thereby improving the adoption rate of EVs.	that influence the remaining driving range (RDR) of electric vehicles (EVs). It provides a taxonomy of these factors and identify more than 50 data attributes that can be used as input during RDR calculations. By understanding these factors, researchers and engineers can develop more accurate models for estimating the RDR of EVs, which can reduce range anxiety for drivers and improve the adoption rate of EVs.

Chapter 2 provides an integral knowledge about the parameters SOC, SOE, SOH, auxiliary parameters that affect the remaining range of an electrical vehicle. The problems in the conventional methods are explored and how machine learning helps to overcome this is seen.

CHAPTER 3

PROPOSED WORK -REMAINING RANGE ESTIMATION

3.1 INTRODUCTION

This work proposal is to build a machine learning model that can predict the SOC, SOE and remaining range for an e-vehicle. At first, define the input variables/factors which have greater influence on the end result. This may include battery voltage, current, temperature, vehicle speed, and other sensor readings, as well as data on past charging and discharging cycles. Dataset has to be collected and preprocessed. This data should include a range of operating conditions and should be representative of the target application. The data should also be cleaned, filtered, and normalized to ensure that it is suitable for training the ML model. Using the preprocessed training data, the ML model can be trained using a supervised learning approach. Once the ML model has been trained, it should be evaluated to ensure that it provides accurate and reliable SOC and SOE estimates under a range of operating conditions. Once the ML model has been validated, it can be integrated into the BMS firmware and hardware. Once the ML algorithm has been integrated with the BMS, the entire system should be tested and validated to ensure that it provides accurate and reliable results, and it is important to continuously monitor and update the ML model to ensure that it remains accurate and reliable over time.

3.2 WORK PLAN

The module split up is as follows:

Module 1: Collection of data set from e-vehicle using various sensors and at various instances. (The collection of data sets will be from kaggle and from othersources).

Module 2: State of charge (SOC) calculation by training an ML model.
Module 3: State of energy (SOE) calculation.
Module 4: Training an ML model using calculated values of SOC, SOE and other influential parameters to find remaining coverable range.

3.3 ARTIFICIAL NEURAL NETWORK

An Artificial Neural Network (ANN) is a computational model inspired by the structure and functionality of the human brain. It is a powerful tool used in machine learning and artificial intelligence to process and analyze complex data patterns. At its core, an ANN consists of interconnected nodes, called neurons, which mimic the behavior of biological neurons. These neurons work together in layers to perform computations and make predictions. The network receives input data, processes it through the layers of neurons, and produces output predictions or classifications.

The structure of an ANN typically consists of an input layer, one or more hidden layers, and an output layer. The input layer receives the raw data, while the output layer provides the result or prediction. The hidden layers, located between the input and output layers, perform computations and extract features from the data. Each neuron in an ANN receives input signals from the previous layer or directly from the input data. The neuron applies a mathematical transformation to the inputs, known as an activation function, to compute its output. The activation function introduces non-linearity into the network, enabling it to learn complex relationships between inputs and outputs.

The connections between neurons in an ANN are represented by weights. These weights determine the strength of the signal transmitted from one neuron to another. During the learning process, the network adjusts these weights to optimize the model's performance. The weights are initially assigned random values and are updated iteratively using various learning algorithms, such as backpropagation. The learning process in an ANN involves training the network on a labeled dataset. The network compares its predicted output with the desired output and adjusts the weights accordingly. This iterative process continues until the network achieves a satisfactory level of accuracy or minimizes the error. Once trained, the ANN can make predictions or classify new, unseen data based on the patterns it has learned.

Activation functions:

ReLU - The rectifier or ReLU (rectified linear unit) activation function is an activation function defined as the positive part of its argument:

$$f(x) = x^+ = \max(0, x) = \begin{cases} x & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases} \qquad f'(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x < 0. \end{cases} \tag{3.1}$$

where x is the input to a neuron. This is also known as a ramp function and is analogous to half-wave rectification.

Linear - A linear function is also known as a straight-line function where the activation is proportional to the input i.e., the weighted sum from neurons. It has a simple function with the equation:

$$f(x) = ax + c. \tag{3.2}$$

3.4 LONG SHORT-TERM MEMORY (LSTM):

Long Short-Term Memory (LSTM) is a type of Recurrent Neural Network (RNN) that is specifically designed to handle sequential data, such as time series, speech, and text. LSTM networks are capable of learning long-term dependencies in sequential data, which makes them well suited for tasks such as language translation, speech recognition, and time series forecasting.

A traditional RNN has a single hidden state that is passed through time, which can make it difficult for the network to learn long-term dependencies. LSTMs address this problem by introducing a memory cell, which is a container that can hold information for an extended period. The memory cell is controlled by three gates: the input gate, the forget gate, and the output gate. These gates decide what information to add to, remove from, and output from the memory cell.

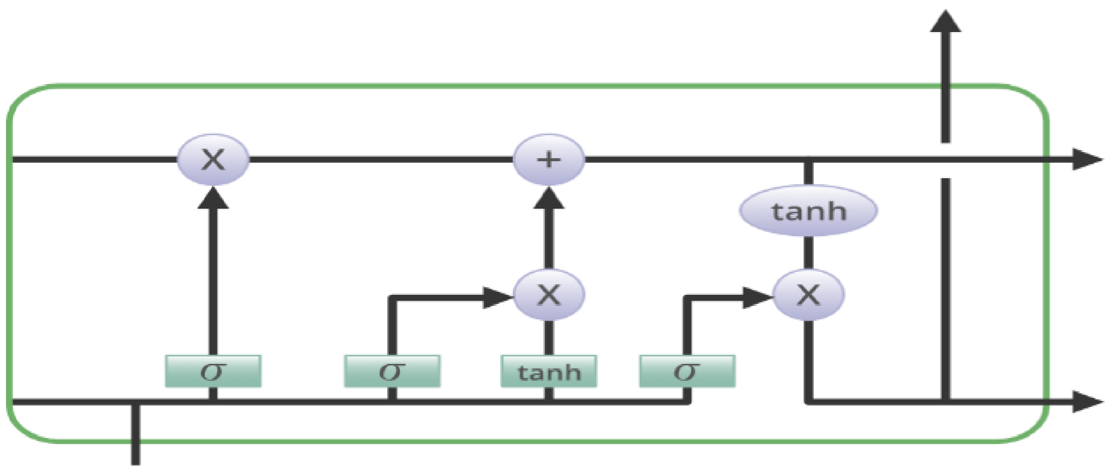

Fig 3.1 - Structure of LSTM

Activation function:

Tanh - The hyperbolic tangent activation function is also referred to simply as the Tanh (also “tanh” and “TanH“) function.

$$Tanh$$

$$f(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})} \tag{3.3}$$

The function takes any real value as input and outputs values in the range -1 to 1.

3.5 ARCHITECTURE DIAGRAM:

In fig 3.5.2, the data collected through various sources is preprocessed and divided into testing and training data. An ML model is designed to calculate SOC & SOE. Again, another ML model is built which takes these calculated SOC & SOH values and other additional parameters for training and predicts the remaining range.

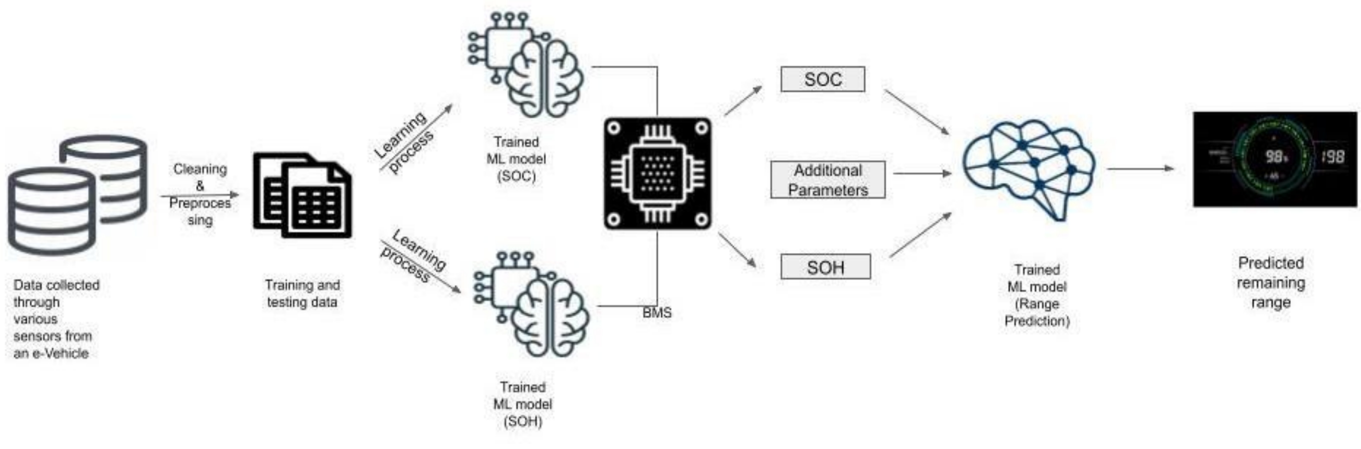

Fig 3.2 - Architecture diagram

3.6 DESIGN OF MODULES

Project Modules:

- EV data extraction
- SOC calculation
- SOE calculation
- Remaining range estimation

1. EV data extraction

Data such as current, voltage, speed, power, etc. are collected from electric vehicles using various sensors and at various instances. This data is later preprocessed.

2. SOC calculation

State of charge refers to the power remaining in the battery. An ML model is trained to calculate the SOC using the attributes current, temperature and voltage.

3. SOE calculation

SOE is defined as the ratio of the battery residual energy under specific operating conditions, e.g., varying load and temperature, over the total battery available energy.

4. Remaining range estimation

The remaining range of the vehicle is then predicted/estimated using the calculated SOC, SOE, driving style, road grade, weather conditions, power discharged and auxiliary devices.

3.7 SUMMARY

Chapter 3 sheds light on the workflow, current range estimation methods and problems involved in it. The various modules involved are detailed. The architecture of the work is illustrated. The implementation of the proposed work is discussed in the upcoming chapter.

CHAPTER 4

IMPLEMENTATION

4.1 INTRODUCTION

This chapter provides insight into various tools and libraries used, implementation details along with visualization.

4.1.1 TOOLS USED

Visual studio code

Visual Studio Code, also commonly referred to as VS Code, is a source-code editor made by Microsoft with the Electron Framework, for Windows, Linux and macOS. Features include support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded Git.

Jupyter notebooks

The Jupyter Notebook App is a server-client application that allows editing and running notebook documents via a web browser. The Jupyter Notebook App can be executed on a local desktop requiring no internet access (as described in this document) or can be installed on a remote server and accessed through the internet. VS Code supports this natively with help of an extension.

4.1.2 PROGRAMMING LANGUAGES USED

Python

The data preprocessing, machine learning model are built using python with the help of libraries.

4.2 LIBRARIES USED:

4.2.1 Matplotlib

Matplotlib is a cross-platform, data visualization and graphical plotting library for Python and its numerical extension NumPy. As such, it offers a viable open-source alternative to MATLAB. Developers can also use matplotlib's APIs (Application Programming Interfaces) to embed plots in GUI applications Matplotlib is a visualization library in Python for 2D plots of arrays. It is a multi-platform data visualization library built on NumPy arrays and designed to work with the broader SciPy stack. One of the greatest benefits of visualization is that it allows us visual access to huge amounts of data in easily digestible visuals. Matplotlib consists of several plots like line, bar, scatter, histogram etc.

4.2.2 Sklearn

Scikit-learn is probably the most useful library for machine learning in Python. The sklearn library contains a lot of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction.

4.2.3 Tensorflow

TensorFlow is an end-to-end open-source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML powered applications. TensorFlow is used to create large-scale neural networks with many layers. TensorFlow is mainly used for deep learning or machine learning problems such as Classification, Perception, Understanding, Discovering, Prediction and Creation.

4.2.4 Keras

Keras is a high-level, deep learning API developed by Google for implementing neural networks. It is written in Python and is used to make the implementation of neural networks easy. It also supports multiple backend neural network computation. TensorFlow is an open-sourced end-to-end platform, a library for multiple machine learning tasks, while Keras is a high-level neural network library that runs on top of TensorFlow. Both provide high-level APIs used for easily building and training models, but Keras is more user-friendly because it's built-in Python.

4.2.5 Numpy

NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays

4.2.6 Pandas

Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.

4.2.7 Seaborn

Seaborn is a library that uses Matplotlib underneath to plot graphs. It will be used to visualize random distributions.

4.2.8 SciPy

SciPy is a free and open-source Python library used for scientific computing and technical computing. SciPy contains modules for optimization, linear algebra, integration, interpolation, special functions, FFT, signal and image processing, ODE solvers and other tasks common in science and engineering.

4.3 Device specifications

Processor	**Operating system**	**Memory**
1.8 GHz or faster	*Windows, linux or macOs*	Minimum 8Gb of ram

Table 4.1 – System specifications

4.4 Machine learning model

4.4.1 SOC & SOE:

Data Collection: We have used drive cycle data from the Panasonic 18650PF Li-ion battery of an EV at normal temperature (25 °C). The drive cycle power profile is calculated for an electric Ford F150 truck.

Pre-processing: Required data is extracted, preprocessed, outliers are removed and is normalized before going for training the model. The dataset had the features Ah, Wh, Current, Voltage, Battery_temp, Chamber_temp, time & power. For calculating SOC current, temperature and voltage are taken into account and rest of the features are dropped. The target variable SOC can be calculated using amp-hour, which is the discharge from the battery with time. So, by adding the nominal amp-hour i.e. 2.5Ah, to all the values and dividing it again by 2.5Ah we will get out our target variable, SOC.

Data Analysis: There are no categorical values so encoding is not required and also no missing values, so imputation is not necessary. Pearson correlation coefficient is used to check the correlation between the features in the dataset.There are no features that are either positively or negatively correlated.Z-Score test is used to check for outliers and outlier records are dropped accordingly. So, we can continue with the dataset for training.

Data Splitting: 80 percent of the data is used for training and 20 percent is used for testing.

Model Building:

Code:

```
model_SOC = Sequential()
model_SOC.add(Dense(units=5, activation='relu',
input_dim=len(X_train_SOC.columns)))
model_SOC.add(Dense(units=5, activation='relu'))
model_SOC.add(Dense(units=1, activation='linear'))

es = tf.keras.callbacks.EarlyStopping(monitor='val_loss', mode='min',
verbose=1, patience=50)

opt = keras.optimizers.Adam(0.001)
model_SOC.compile(loss='mse', optimizer=opt, metrics='mse')
model_SOC.summary()
```

Visualization:

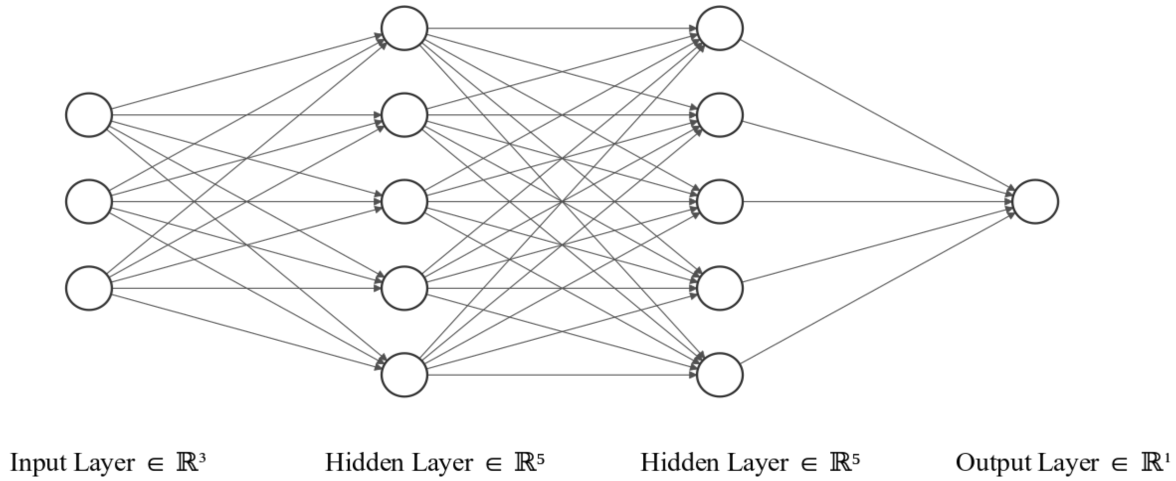

Fig 4.1 – NN for SOC calculation

In the ANN used, the input layer consists of 3 neurons, 2 hidden layers with 5 neurons each and output layer with a single neuron. For hidden layers relu activation function is used and for output layer linear activation function is used. To prevent the model from overfitting early stopping is used and adam optimiser is used.

Training & Testing: Finally, the model is trained and tested.

STATE OF ENERGY (SOE): Data collection and preprocessing involves the same steps as used above, but we consider SOC as an additional feature.

4.4.2 REMAINING RANGE ESTIMATION:

Data Collection: For training the model, we have used a dataset that contains real-life driving data of two Volkswagen e-Golf cars, with year of manufacture as 2014 and 2016 respectively. The file includes data about 3615 trips with a total travel distance of around 152167 kilometers.

Pre-processing: The data was preprocessed, necessary features were one- hot encoded and standardized before being used to train the model. The dataset had the features manufacturer, model, version, power, fuel_date, trip_distance, quantity, fuel_type, tire_type, city, motor_way, country_roads, driving_style, consumption, A/C, park_heating, avg_speed, ecr_derivation. Of which manufacturer, model, version, power, fuel_date, fuel_type are dropped and rest are retained. These attributes are converted to their respective compatible data types.

Data Analysis: Records having null values in the target variable are dropped and for other attributes missing values are filled using mean and median values.During analysis phase data is divided into continuous and discrete features and analysis is done accordingly. Continuous values such as trip_distance, quantity, consumption, avg_speed, ecr_derivation are analyzed by creating histograms. This helps in understanding their distribution. Whereas discrete features such as city, motor_way, country_roads, A/C, park_heating and their relation with the target variable is visualized with the help of graphs.

Encoding: Categorical features such as tyre_type and driving_style are encoded using one-hot encoding.

Outlier analysis: This analysis is performed on continuous features assuming that they are having Gaussian distribution. The data points having a z-score of more than 3 and less than -3 are removed from the dataset since they are considered as outliers.

Correlation b/w features: Correlations between different features are observed using heat-map. From the observation we can see that driving_style Normal & driving_style Moderate are possessing strong negative correlation, one of them needs to be dropped. Here we are dropping driving_style normal. Same goes for Summer & winter tire types and we drop tyre_type summer.

Data Splitting: 80 percent of the data is used for training and 20 percent is used for testing.

Model Building:

Code:

```
model = Sequential()
model.add(LSTM(16, input_shape=(len(X_train.columns),1),
return_sequences=True))
model.add(LSTM(32,return_sequences=False))
model.add(Dense(units=32, activation='relu'))
model.add(Dense(units=64, activation='relu'))
model.add(Dense(units=1, activation='linear'))
batch_size = 16     # batch size for model fitting
epochs = 1000       # number of epochs for model fitting

STEPS_PER_EPOCH = X_train.shape[0]/batch_size
es = tf.keras.callbacks.EarlyStopping(monitor='val_loss',
patience=50,verbose=1) # Early stopping to prevent overfitting
lr_schedule = tf.keras.optimizers.schedules.InverseTimeDecay(              #
Using a dynamic learning rate that decays a certain rate per 100 epochs
  0.001,
  decay_steps=STEPS_PER_EPOCH*100,
  decay_rate=0.98,
  staircase=False)
opt = tf.keras.optimizers.RMSprop(lr_schedule)   # Using RMSprop optimizer
with the learning rate schedule as mentioned above
model.compile(loss='mae', optimizer=opt, metrics=['mae'])
model.summary()
```

Visualization

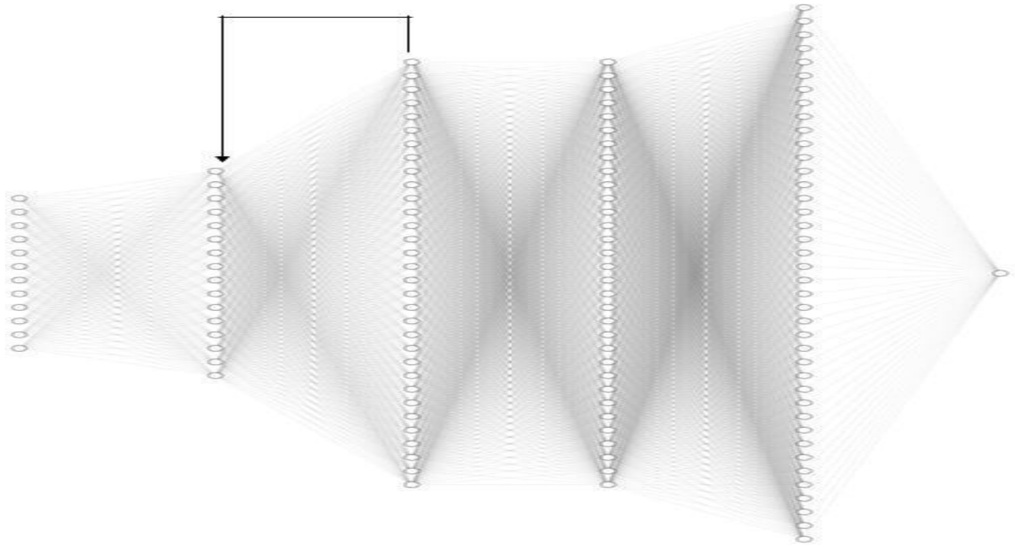

Fig 4.2 – NN for remaining range estimation

In the ANN used, the input layer consists of 12 nodes. There are 4 hidden layers. The first two are LSTM layers (layer1 -16 nodes & layer 2- 32 nodes) and the remaining two(layer 1- 32 nodes & layer 2 -64 nodes) are dense layers. For LSTM layers tanh activation function is used and for dense layers relu is used. The output layer consists of a single node which gives desired output. Linear activation function is used.

Training & Testing: Finally, the model for predicting remaining range is trained and tested.

4.5 SUMMARY

Chapter 4 provides comprehensive and detailed implementation details . The tools used, development environment and device specifications are discussed. How each module of the system is put together to make up the proposed solution is discussed. Each module is explained in a detailed way.

CHAPTER 5

RESULTS AND ANALYSIS

5.1 INTRODUCTION

The result obtained in the proposed solution is discussed in this chapter. Results are compared to the conventional solutions and an analysis is made for the same

5.2 DATASET ANALYSIS

5.2.1 The trip distance is compared with various parameters

1) Trip distance compared to city roads

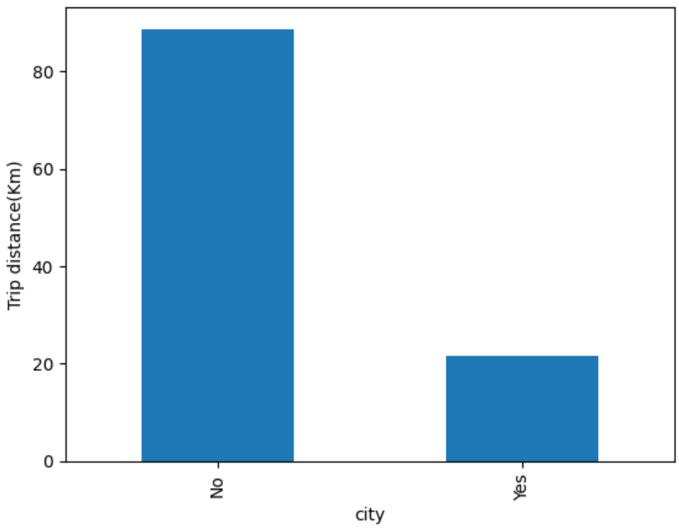

Fig 5.2.1 – City roads vs average trip distance

Trip distance is lower in cities (This may be due to traffic conditions causes wastage of energy as the brakes are used instead of regenerative braking.)

2) Trip distance compared to motor ways

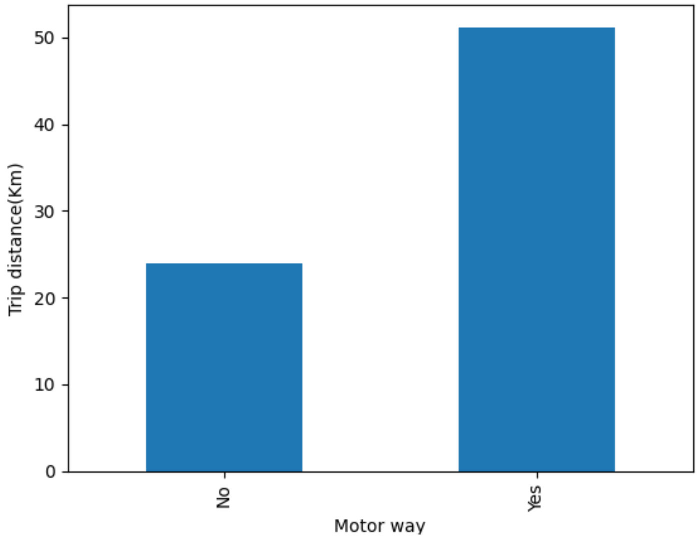

Fig 5.2.2 – motor way vs average trip distance

Trip distance is more in motor ways (This may be because motor ways are free therefore this leads to consistent speeds, reduced stops and reduced accelerations)

3) Trip distance compared to country roads

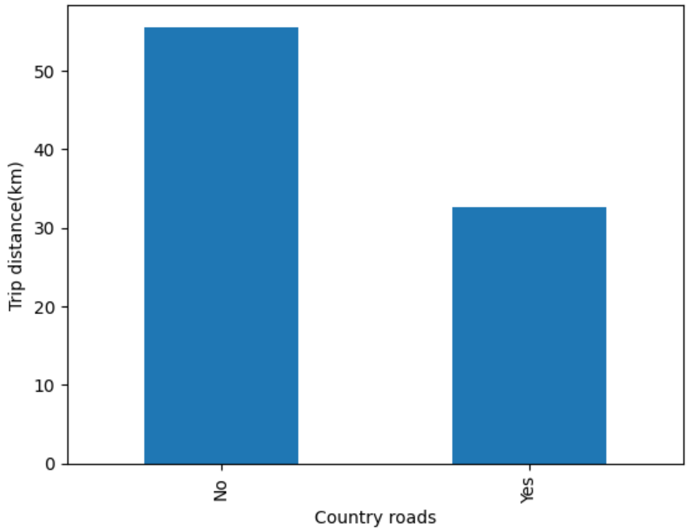

Fig 5.2.3 – country roads vs average trip distance

Trip distance is less on country roads (This maybe because country roads are uneven, have bumps and detours therefore there will be varied speed limits, frequent stops and acceleration and also limited charging infrastructure)

4) Trip distance compared to park heating

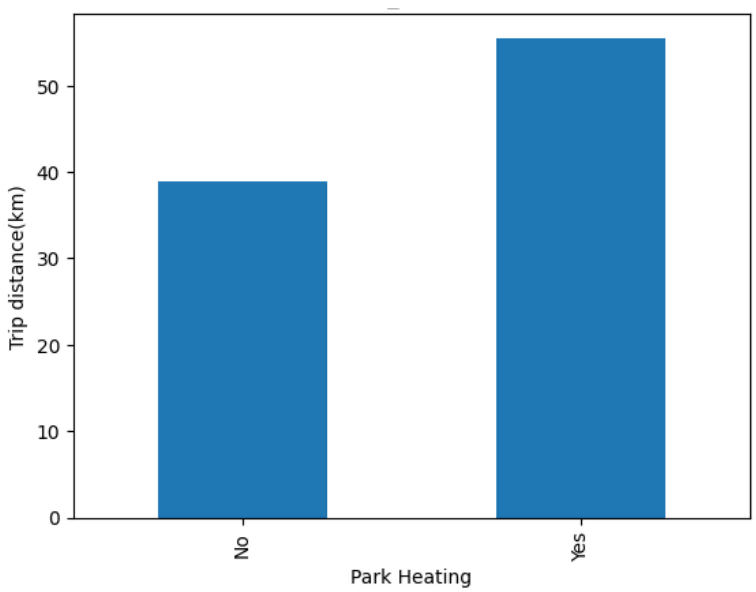

Fig 5.2.4 – Park heating vs average trip distance

Trip distance is more when park heating is used (park heating can help pre-warm the battery and improve its performance and efficiency)

5) Trip distance compared to air conditioning

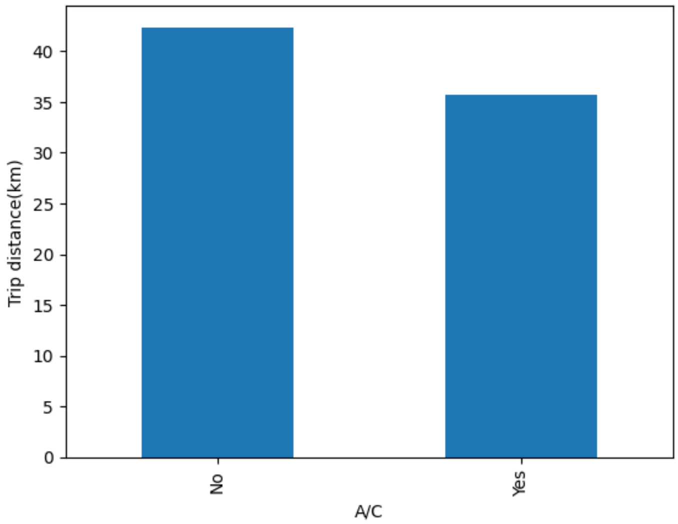

Fig 5.2.5 – A/C vs average trip distance

Trip distance is less when air conditions is used (air conditioning will result in increased energy consumption and load on the battery)

6) Car battery information comaperd to trip distance

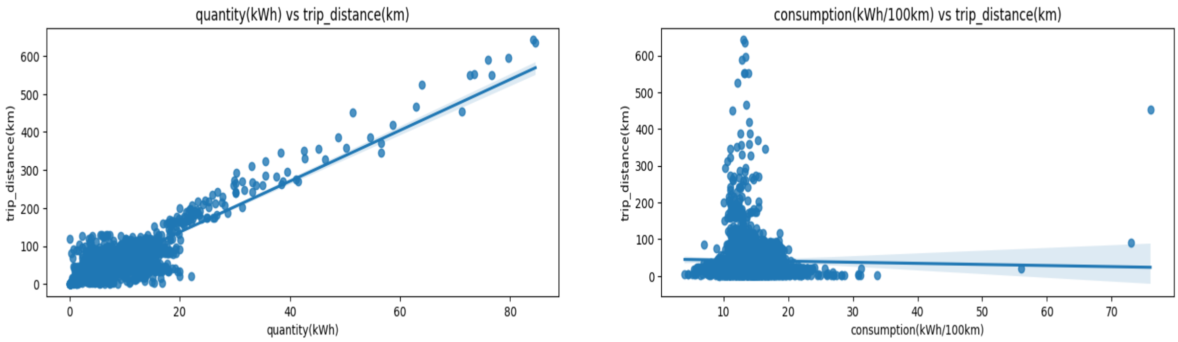

Fig 5.2.6 – trip distance vs quantity and consumption regression plot

Quantity above 20kwh is directly proportional to trip distance. Quantity in range 0 to 20, may not alone determine the trip distance, we need to combine some more features to determine the trip distance for quantity kwh. Trip distance is higher in EV's having energy consumption range in 10 to 20.

7) Average speed vs trip distance regression plot

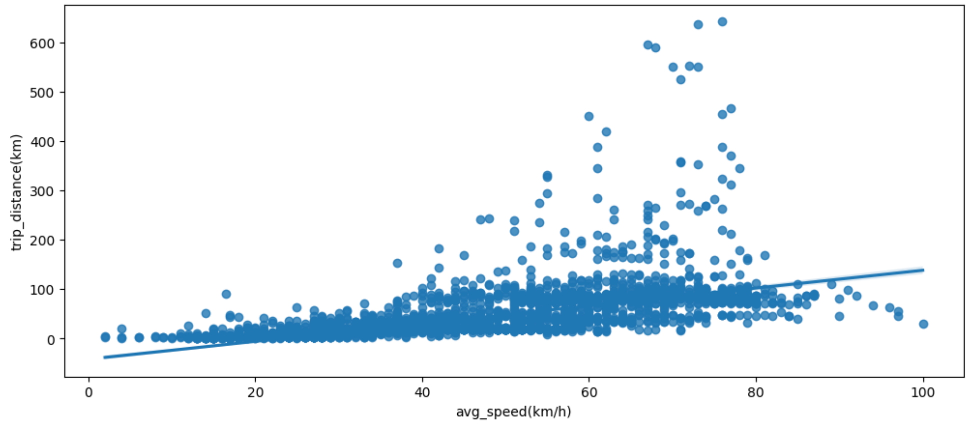

Fig 5.2.7 – trip distance vs average speed regression plot

We can see data points having an upward trend till 80 km/hr, so the avg_speed will be a feature to consider for distance range.

5.2.2 Average speed distribution

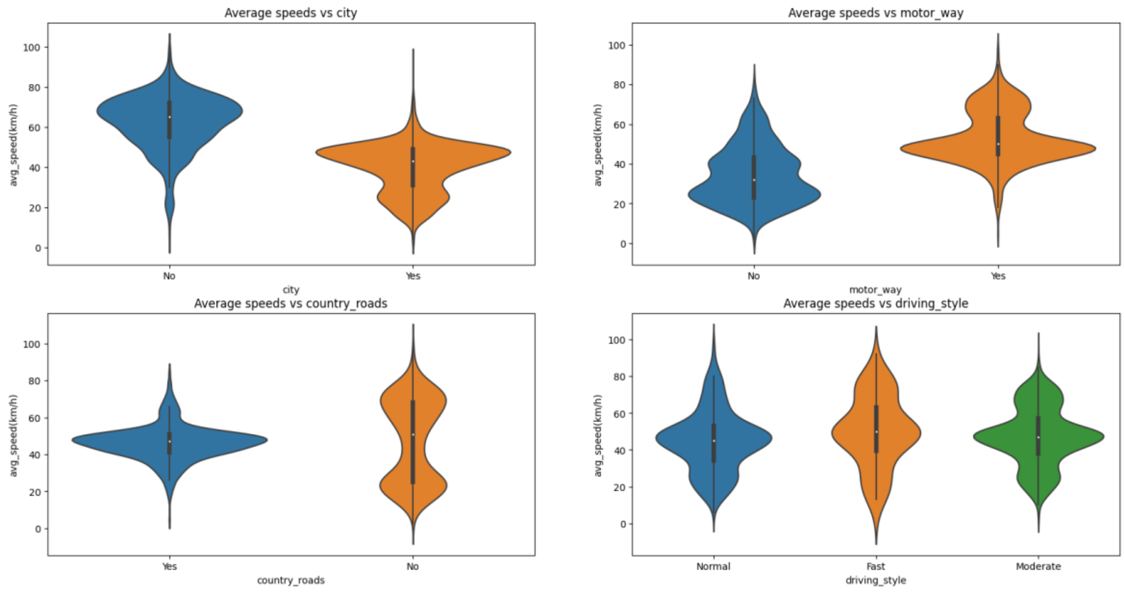

Fig 5.2.8 – Average speed compared with different road types and driving styles

InterQuartile Range of Avg_speed is clearly separated among city's. City 'No' have higher avg speeds. InterQuartile Range of Avg_speed is clearly separated among motor_way's. Motor_way 'Yes' have higher avg speeds. Avg speed range is widespread in country road 'No' compare to 'Yes'.

5.2.3 Effect of auxiliary devices on consumption

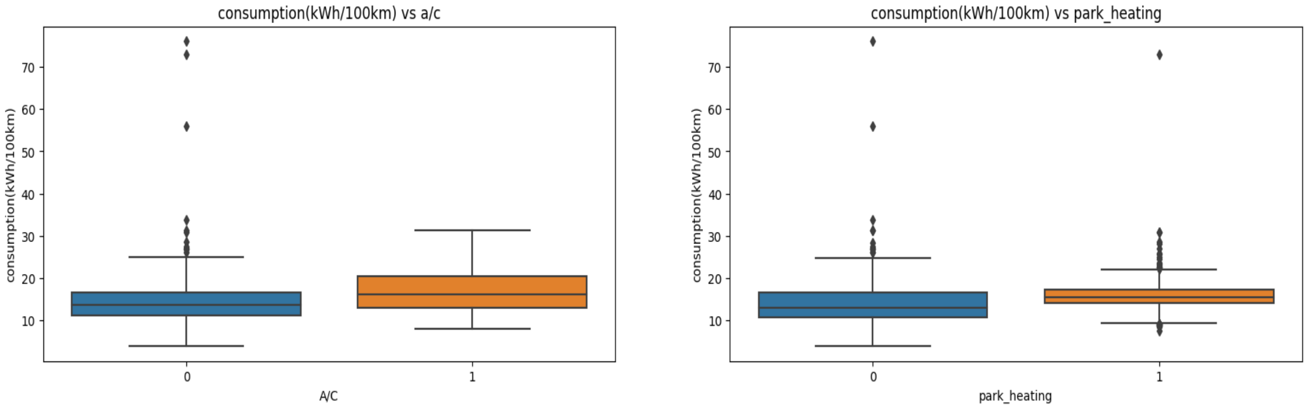

Fig 5.2.9 – Vehicle consumption compared with auxiliary devices

Energy consumption is comparatively higher when A/C is on, and park heat is used because of increased battery usage

5.2.4 Histograms for continuous variables to visualize their distribution

Histogram is a graphing tool used to summarize discrete or continuous data that are measured on an interval scale.

1) Trip distance

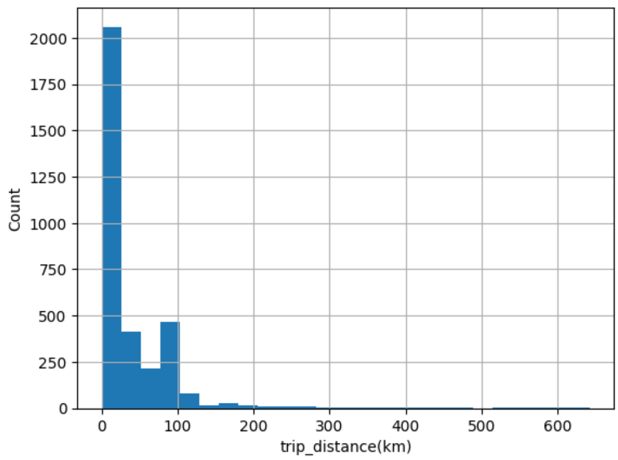

Fig 5.2.10 - Histogram for trip distance (target variable)

Skewness: positive, dispersion: smaller spread (Trip distance is spread in the region 1 to 100) and outliers: records after 150 .

2) Quantity

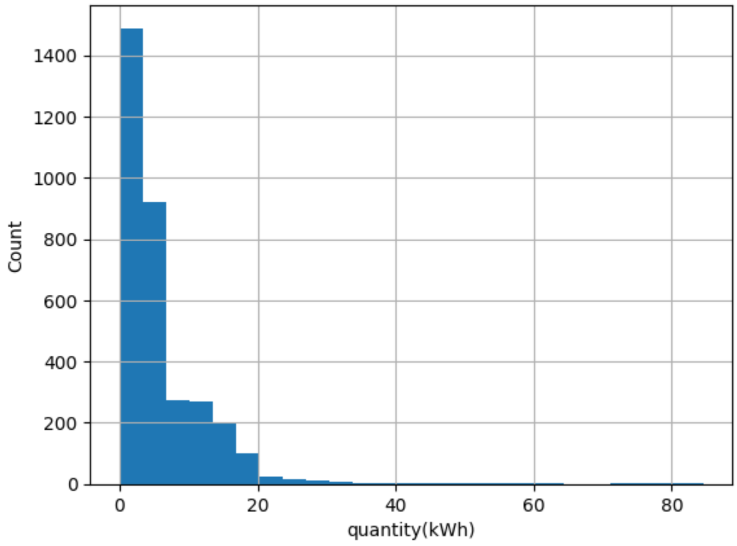

Fig 5.2.11 - Histogram for quantity

Skewness: positive, dispersion: smaller spread (Quantity is spread in the region 0 to 15) and Outliers: records after 20

3) Consumption

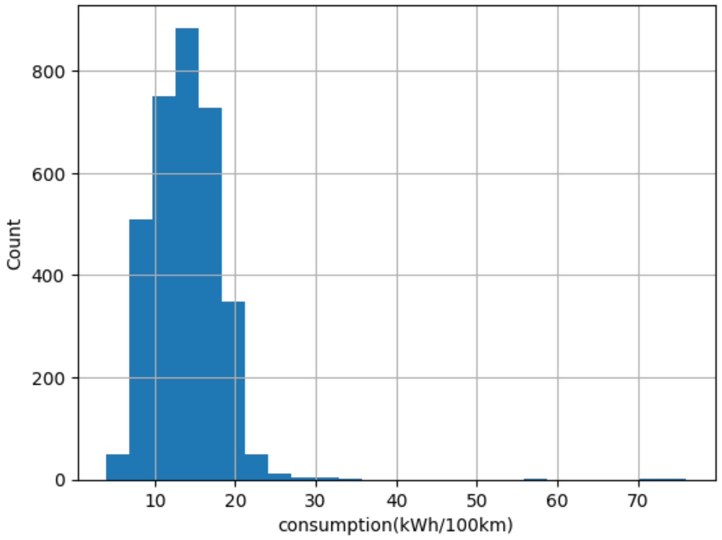

Fig 5.2.12 - Histogram for consumption

Skewness: positive, dispersion: smaller spread (Consumption is spread in the region 5 to 20) and outliers: records after 25.

4) Average speed

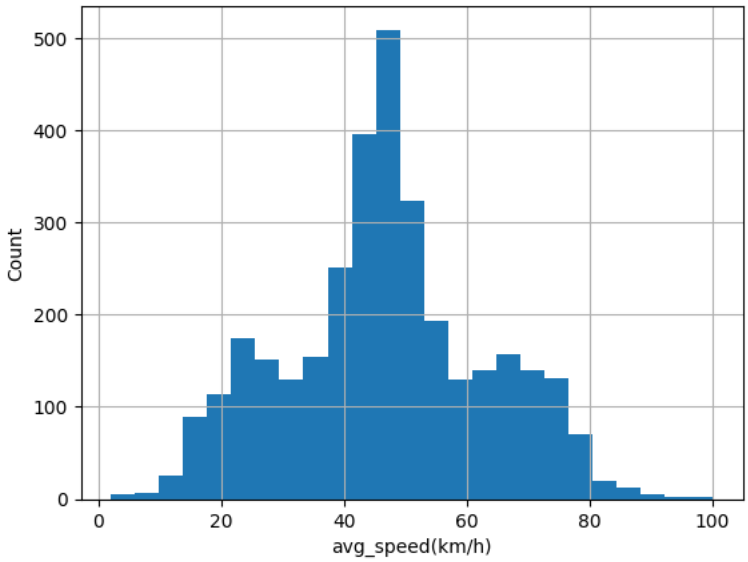

Fig 5.2.13 - Histogram for average speed

Skewness: none, dispersion: evenly spread and outliers: records between 0 to 15 and after 80

5) ECR deviation

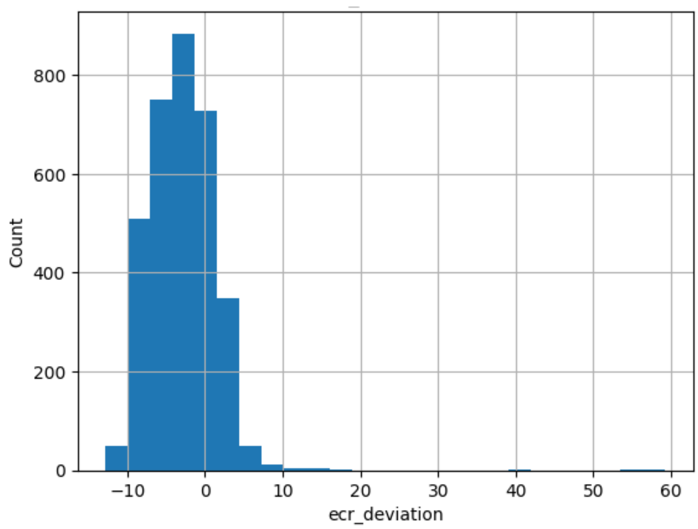

Fig 5.2.14 - Histogram for ecr deviation

Skewness: positive, dispersion: smaller spread (ECR deviation is accumulated around the region -10 to 5) and outliers: records after 5

5.3 TEST RESULTS

5.3.1 SOC

Error:

```
641/641 [==============================] - 1s 763us/step
Mean absolute error: 0.014113650812155867
```

Fig 5.3.1 – Mean absolute error for SOC model

Test: Prediction vs Actual values:

Fig 5.3.2 – Predicted vs actual SOC values

	Predictions	Actual Output
0	0.135118	0.167040
1	0.793014	0.791524
2	0.249730	0.254800
3	0.279199	0.282960
4	0.816404	0.790392
...	...	...
20484	0.147491	0.162160
20485	0.984723	0.928672
20486	0.316771	0.315720
20487	0.878058	0.884896
20488	0.284465	0.284760

20489 rows × 2 columns

5.3.2 SOE

Error:

```
641/641 [==============================] - 0s 600us/step
Mean absolute error: 0.0008240243712780419
```

Fig 5.3.3 – Mean absolute error for SOE model

Test: prediction vs Actual values:

	Predictions	Actual Output
0	0.161298	0.161857
1	0.777971	0.777673
2	0.244460	0.244004
3	0.271357	0.271443
4	0.777336	0.776499
...	...	...
20484	0.157434	0.157908
20485	0.920528	0.921808
20486	0.302690	0.302908
20487	0.874473	0.874161
20488	0.272930	0.273210

20489 rows × 2 columns

Fig 5.3.4 – Predicted vs actual SOE values

5.3.3 Range Estimation:

Error:

```
from sklearn.metrics import mean_absolute_error
predictions = (model.predict(X_test))
Y_test.to_csv('.//y_test.csv', index= False)
pd.DataFrame(predictions).to_csv('.//predictions.csv', index= False)
print(mean_absolute_error(Y_test, predictions))
✓ 2.4s
21/21 [==============================] - 2s 6ms/step
5.967760939558764
```

Fig 5.3.5 – Mean absolute error for Range estimation model

Test: Prediction vs Actual values:

	Predictions	Actual Output
0	19.913588	20.0
1	19.177187	19.0
2	20.933769	22.0
3	20.233088	20.0
4	19.742170	20.0
...	...	...
662	142.805740	143.0
663	23.974070	12.0
664	4.279182	5.0
665	20.014788	20.0
666	31.427786	33.0

667 rows × 2 columns

Fig 5.3.6 – Predicted vs actual range values

Comparison with other models

Table 5.2 – Comparison of test results

Model	MAE Value	Parameters
SVR	15.25	Kernel = rbf
Linear regression	10.28	Multi valued
Random forest	6.78	n_estimators = 12
MLP	5.96	batchsize = 16 optimizer = RMSprop callbacks = earlystopping

From the above table it can be observed that the mean absolute error is least for MLP. So MLP is most reliable for remaining range prediction.

Validation loss vs epochs:

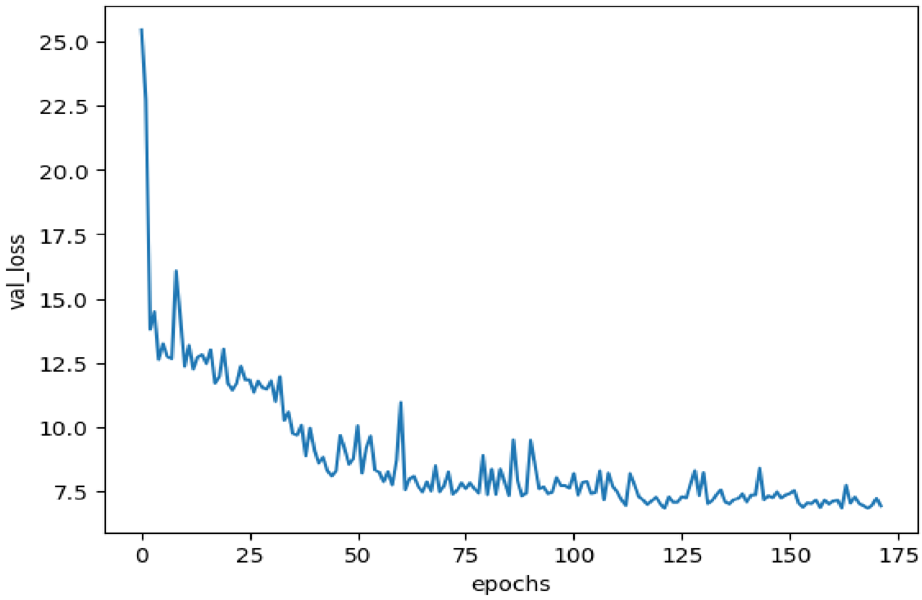

Fig 5.3.7 – Validation loss over time

Validation loss was around 25.0 initially and after 175 epochs it is reduced to 7.

Mean absolute error over time:

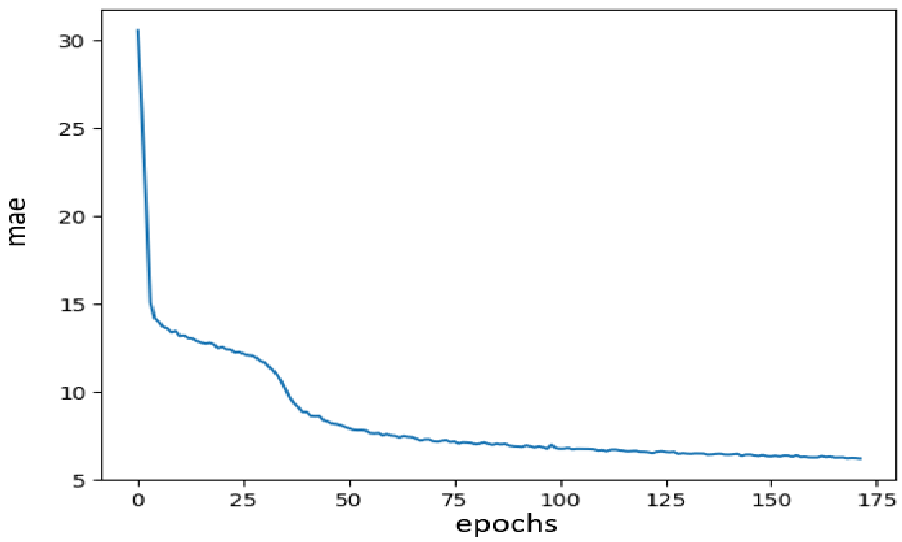

Fig 5.3.8 – MAE over time

MAE was around 30.0 initially and after 175 epochs it is reduced to 5.

5.3.4 Webapp

A website with simple UI is developed, where the user can enter the required details viz., Qunatity in KWh, City (yes:0, no:1), Motorway,country roads, Consumption per 100 km,n auxiliary devices details like A/C,park_heating, Driving style, Avg speed. After entering the details, estimated remaining range is shown.

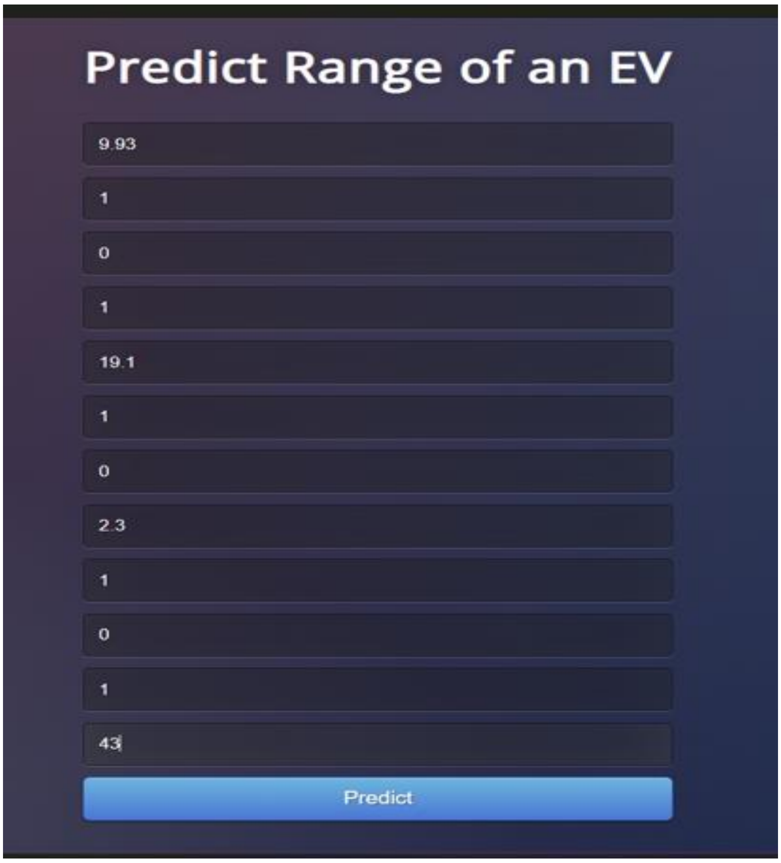

Fig 5.3.9 – Webapp UI

Fill all the required fields with suitable values in order to estimate remaining range.

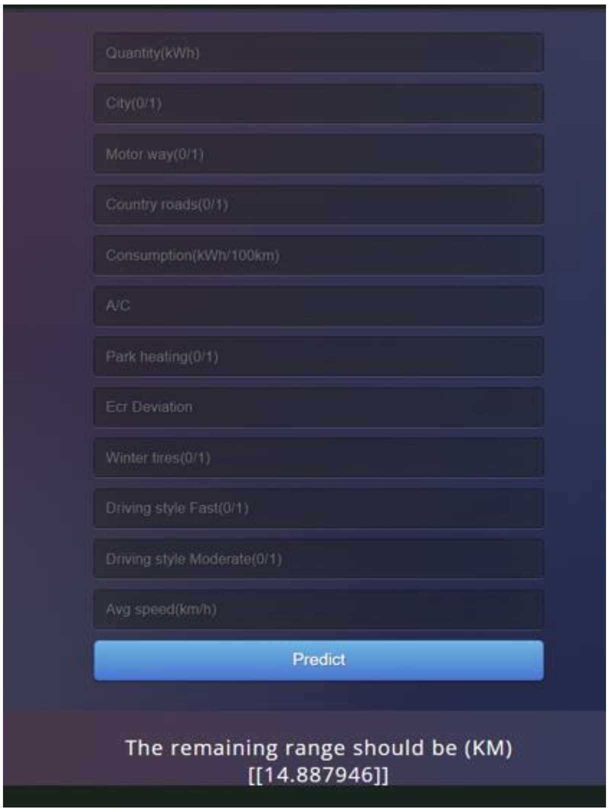

Fig 5.3.10 – Webapp UI

Once the values are entered the model estimates range and displays the same.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 CONCLUSION:

The results obtained from the application of the machine learning (ML) model for calculating State of Charge (SOC) have shown promising results. By utilizing updated SOC values, the model demonstrates accurate predictions for State of Energy (SOE). The Mean Absolute Error (MAE) observed during the calculation of the remaining range is 5.96, indicating a close match between the predicted and actual output. These findings establish the model's suitability for practical use.

6.2 FUTURE WORK:

The proposed model focuses on predicting the remaining range of an electric vehicle (e-vehicle). It considers various parameters including road grade, driver behavior, environmental factors, and auxiliary devices. Additionally, the model recognizes that seating capacity can impact the range and can be incorporated as a constraint for future training. This allows the model to improve its accuracy in predicting the range over time by considering these additional factors.

REFERENCES

1. Penghua Li, Zijian Zhang, Qingyu Xiong, Baocang Ding, Jie Hou, Dechao Luo, Yujun Rong, Shuaiyong Li, (2020) "State-of-health estimation and remaining useful life prediction for the lithium-ion battery based on a variant long short term memory neural network",Journal of Power Sources,Volume 459

2. B. Zheng, P. He, L. Zhao and H. Li, (2016) "A Hybrid Machine Learning Model for Range Estimation of Electric Vehicles," 2016 IEEE Global Communications Conference (GLOBECOM), Washington, DC, USA, pp.1-6

3. P. Sharma and B. J. Bora, (2022) "A Review of Modern Machine Learning Techniques in the Prediction of Remaining Useful Life of Lithium-Ion Batteries," Batteries, vol. 9, no. 1, p. 13

4. R. Li, H. Sun, X. Wei, W. Ta, and H. Wang, "Lithium Battery State-of-Charge Estimation Based on AdaBoost.Rt-RNN," Energies, vol. 15, no. 16, p. 6056, Aug. 2022

5. Ganesh Sankaran, S. Venkatesan, M. Prabhahar,Range Anxiety on electric vehicles in India -Impact on customer and factors influencing range Anxiety,Materials Today: Proceedings,Volume 33, Part 1,2020

6. Miri, I, Fotouhi, A, Ewin, N. Electric vehicle energy consumption modelling and estimation—A case study. Int J Energy Res. 2021

7. Szumska, E.M.; Jurecki, R.S. Parameters Influencing on Electric Vehicle Range. Energies 2021, 14, 4821.

8. Jun Bi, Yongxing Wang, Qiuyue Sai, Cong Ding, estimating remaining driving range of battery electric vehicles based on real-world data: A case study of Beijing, China, Energy, Volume 169, 2019, Pages 833-843, ISSN 0360-5442

9. JOUR, Sun, Shuai, Zhang, Jun, Bi, Jun, Wang, Yongxing, 2019/01/09, A Machine Learning Method for Predicting Driving Range of Battery Electric Vehicles, 4109148

10. J. Hong, S. Park and N. Chang, "Accurate remaining range estimation for Electric vehicles," 2016 21st Asia and South Pacific Design Automation Conference (ASP-DAC), Macao, China, 2016

11. M. Smuts, B. Scholtz and J. Wesson, "A critical review of factors influencing the remaining driving range of electric vehicles," 2017 1st International Conference on Next Generation Computing Applications (NextComp), Mauritius, 2017

12. S. Sharma, P. D. Achlerkar, P. Shrivastava, A. Garg and B. K. Panigrahi, "Combined SoC and SoE Estimation of Lithium-ion Battery using Multi-layer Feedforward Neural Network," 2022 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES), Jaipur, India, 2022, pp. 1-6

13. O. Rahbari, N. Omar, P. Van Den Bossche and J. Van Mierlo, "A centralized state of charge estimation technique for electric vehicles equipped with lithium-ion batteries in smart grid environment," 2018 IEEE International Conference on Industrial Technology (ICIT), Lyon, France, 2018, pp. 1721-1725

14. P. A. Topan, M. N. Ramadan, G. Fathoni, A. I. Cahyadi and O. Wahyunggoro, "State of Charge (SOC) and State of Health (SOH)

estimation on lithium polymer battery via Kalman filter," 2016 2nd International Conference on Science and Technology-Computer (ICST), Yogyakarta, Indonesia, 2016, pp. 93-96

15. M. Xu, Q. Zhu and M. Zheng, "A State of Charge Estimation Approach Based on Fractional Order Adaptive Extended Kalman Filter for Lithium-ion Batteries," 2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS), Enshi, China, 2018, pp. 271-276

16. S. Barcellona, D. D. Simone and S. Grillo, "Real-time Electric Vehicle Range Estimation Based on a Lithium-Ion Battery Model," 2019 International Conference on Clean Electrical Power (ICCEP), Otranto, Italy, 2019, pp. 351-357

17. S. Sautermeister, M. Falk, B. Bäker, F. Gauterin and M. Vaillant, "Influence of Measurement and Prediction Uncertainties on Range Estimation for Electric Vehicles," in IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 8, pp. 2615-2626, Aug. 2018

18. A. Bolovinou, I. Bakas, A. Amditis, F. Mastrandrea and W. Vinciotti, "Online prediction of an electric vehicle remaining range based on regression analysis," 2014 IEEE International Electric Vehicle Conference (IEVC), Florence, Italy, 2014, pp. 1-8

19. K. Sarrafan, K. M. Muttaqi, D. Sutanto and G. E. Town, "A Real-Time Range Indicator for EVs Using Web-Based Environmental Data and Sensorless Estimation of Regenerative Braking Power," in IEEE Transactions on Vehicular Technology, vol. 67, no. 6, pp. 4743-4756, June 2018

20. P. Tyagi, L. Padmavathi and A. Abhishek, "Performance Analysis of State of Charge Estimation Techniques for Two-Wheeler Electric Vehicles," 2022 International Virtual Conference on Power Engineering Computing and Control: Developments in Electric Vehicles and Energy Sector for Sustainable Future (PECCON), Chennai, India, 2022, pp. 1-5, doi: 10.1109/PECCON55017.2022.9851113.

21. R. Zhi-yong, "Research on influence factors affecting driving range of flame-proof battery electric vehicles," 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chengdu, China, 2019, pp. 1982-1986

22. K. S. Grewal and P. M. Darnell, "Model-based EV range prediction for Electric Hybrid Vehicles," IET Hybrid and Electric Vehicles Conference 2013 (HEVC 2013), London, 2013, pp. 1-6

23. J. Poxon, A. McGordon, G. Muraleedharakurup and P. Jennings, "Determining a suitable all electric range for a light weight plug-in hybrid electric vehicle," 2010 IEEE Vehicle Power and Propulsion Conference, Lille, France, 2010, pp. 1-8

24. A. Enthaler, T. Weustenfeld, F. Gauterin and J. Koehler, "Thermal management consumption and its effect on remaining range estimation of electric vehicles," 2014 International Conference on Connected Vehicles and Expo (ICCVE), Vienna, Austria, 2014, pp. 170-177