EV Charging Demand Prediction using Machine Learning

A Report Submitted
in Partial Fulfillment of the Requirements
for the Degree of
Bachelor of Technology
in
Electrical Engineering

by
Chirag Srivastava (20202030)
Shreshtha Gupta (20206050)
Sandeep Oraon (20202072)

to the
ELECTRICAL ENGINEERING DEPARTMENT
MOTILAL NEHRU NATIONAL INSTITUTE OF TECHNOLOGY
ALLAHABAD PRAYAGRAJ
October, 2023

UNDERTAKING

I declare that the work presented in this report titled "*EV Charging Demand Prediction using Machine Learning*", submitted to the Electrical Engineering Department, Motilal Nehru National Institute of Technology Allahabad, for the award of the ***Bachelor of Technology*** degree in ***Electrical Engineering***, is my original work. I have not plagiarized or submitted the same work for the award of any other degree. In case this undertaking is found incorrect, I accept that my degree may be unconditionally withdrawn.

October, 2023
Allahabad

Chirag Srivastava (20202030)
Shreshtha Gupta (20206050)
Sandeep Oraon (20202072)

CERTIFICATE

Certified that the work contained in the report titled "*EV Charging Demand Prediction using Machine Learning*", by *Chirag Srivastava (20202030), Shreshtha Gupta (20206050), Sandeep Oraon (20202072)*, has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

(Dr. Navneet Kumar Singh)
Electrical Engineering Dept.
M.N.N.I.T, Allahabad

October, 2023

Preface

In this project, we will use to machine learning techniques to predict electric vehicle charging demand at charging stations. We will train different machine learning models on a dataset of past charging patterns, and then evaluate the model's performance on a held-out test set and compare their performances.

The goal of this project is to develop a robust and accurate model for predicting electric vehicle charging demand. This model can be used by charging station operators to better manage their resources and to ensure that they have enough electricity to meet the needs of their customers.

This project is important because it can help to make electric vehicles more reliable and convenient for consumers. Additionally, by predicting electric vehicle charging demand, we can help to reduce the load on the power grid and make the grid more efficient.

The project was undertaken with the help of Dr. Navneet Kumar Singh as our mentor. This research was difficult, but with his continued support and guidance, we could gradually grasp the important concepts and then move on to code, test, analyze and understand the project at hand. We are greatly indebted to him for all the support and motivation we have received in the process.

Acknowledgements

I am deeply honored to acknowledge the giants on whose shoulders I stand. I would like to express my sincere gratitude to Dr. Navneet Kumar Singh Singh, my project supervisor, for his invaluable guidance and support throughout this project. He was a constant source of encouragement and expertise, and I am grateful for his mentorship.

I would also like to thank our Director, Dr. R.S. Verma, for providing us with this opportunity to pursue our research interests. I am grateful for his vision and leadership.

Finally, I would like to thank all the panel members for their time and consideration in evaluating our project. I appreciate their valuable feedback and insights.

Contents

Chapter 1

Introduction

Electric vehicles (EVs) are becoming increasingly popular as a more sustainable and environmentally friendly transportation option. However, the growing number of EVs on the road puts a strain on the existing power grid. To ensure that the grid can meet the growing demand for electricity from EVs, it is important to accurately predict the electricity required for charging EVs at charging stations.

This project aims to develop a model to predict the electricity required for charging EVs at a charging station. The model will be trained on historical data of EV charging sessions, as well as other factors such as weather, traffic, and local events. The model will be used to predict the electricity required for charging EVs at a charging station on a periodic basis.

The predicted electricity demand can be used by utilities and charging station operators to plan and invest in the necessary infrastructure, such as charging stations and power lines. The predictions can also be used by grid operators to manage the power grid more efficiently and to implement demand response programs.

1.1 Motivation

The adoption of electric vehicles (EVs) is growing rapidly, driven by environmental concerns and government incentives. However, the increasing number of EVs on the road puts a strain on the existing power grid. To ensure that the grid can meet the growing demand for electricity from EVs, it is important to accurately predict the electricity required for charging EVs at charging stations.

Accurate prediction of electricity demand at stations can help in several ways:

- **Planning and investment**: Utilities and charging station operators can use electricity demand predictions to plan and invest in the necessary infrastructure, such as charging stations and power lines. This can help to ensure that there is enough capacity to meet the needs of EV drivers and avoid overloading the grid.

- **Grid Operation**: Electricity demand predictions can help grid operators to manage the power grid more efficiently. For example, grid operators can use this information to dispatch power plants and manage energy storage systems to ensure that there is enough electricity to meet demand at all times.

- **Demand response**: Electricity demand predictions can also be used to implement demand response programs. Demand response programs incentivize consumers to reduce their electricity consumption during peak demand periods. This can help to reduce the overall load on the grid and save money for consumers.

Overall, the accurate prediction of electricity required for charging EVs at charging stations is essential for the successful transition to a zero-emission transportation future.

1.2 Objective

- Train a machine learning model on historical data of EV charging sessions and other factors to predict electricity demand at charging stations:
  - Collect historical data of EV charging sessions, including the time and date of the charging session, the duration of the charging session, the amount of energy consumed, and other relevant data.
  - Clean and prepare the data for machine learning.
  - Select and train a machine learning algorithm to predict electricity demand based on the historical data and other factors.
  - Evaluate the performance of the machine learning model on a holdout test set
- Compare the performance of different machine learning algorithms to identify the best model for predicting electricity demand.
  - Select a variety of machine learning algorithms to compare.
  - Train each algorithm on the same training data.
  - Evaluate the performance of each algorithm on a holdout test set.
  - Select the algorithm with the best performance on the test set.

Chapter 2

Literature Survey

Reference No.	Techniques	Outcomes	Drawbacks
[1]	• SVM • Random Forest • Ensemble Techniques	• Using ensemble learning outperformed predictions made by individual ML models in both scenarios. • The results in this work outperformed all the previous works that reported similar evaluation metrics.	• Vehicle information such as vehicle model and type can improve predictions, because different vehicles have different energy consumption profiles. • The prediction of energy consumption is still challenging and proposed models are not very suitable for this task.
[2]	• K-Nearest Neighbors (KNN) • Decision Tree • Linear Regression	• Proposes model based on linear regression that predicts large-scale public building energy demand. • An attempt to use a hybrid approach to combine dimensionality reduction techniques with machine learning techniques.	• Hybrid models can be more difficult to interpret than traditional machine learning models. This can make it difficult to understand why the model makes the predictions it does. • Hybrid models can be more computationally expensive to train and predict than traditional machine learning models.
[3]	• ANN • LSTM	• This project aimed to predict energy consumption in Finland using an LSTM model, which was trained on a dataset containing 6 years of electricity consumption in Finland.	• The models are prone to overfitting the training data, especially when the training data is small.

Chapter 3

Machine Learning

Machine learning is a relatively old field with classical methods and algorithms since the 1960s. This field of study considers a subfield of Artificial Intelligence that allows computers to learn without being explicitly programmed. In simple terms, machine learning is a way to teach computers to do things by showing them examples, rather than by giving them explicit instructions.

3.1 Classical Machine Learning

Classical machine learning is a type of machine learning that relies on statistical methods and algorithms to learn from data. It is the foundation of many modern machine learning techniques, and it is used in a wide variety of applications, including fraud detection, medical diagnosis, product recommendation, image classification, and natural language processing.

3.1.1 Linear Regression

Linear regression is a statistical method that models the relationship between a dependent variable (y) and one or more independent variables (x). The model can be used to predict the value of y given the values of x. Linear regression is a simple but powerful method that can be used to solve a wide range of problems.

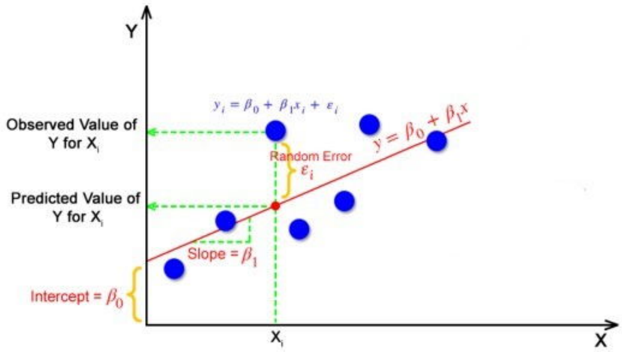


Fig.1 Linear regression

$$y^{(i)} = \beta_0 + \beta_1 x^{(i)}$$

3.1.2 Support Vector Machine

Support vector machines (SVMs) are a machine learning algorithm that can be used for both classification and regression tasks. They work by finding a hyperplane that separates the data into two classes, and then using that hyperplane to make predictions about new data.

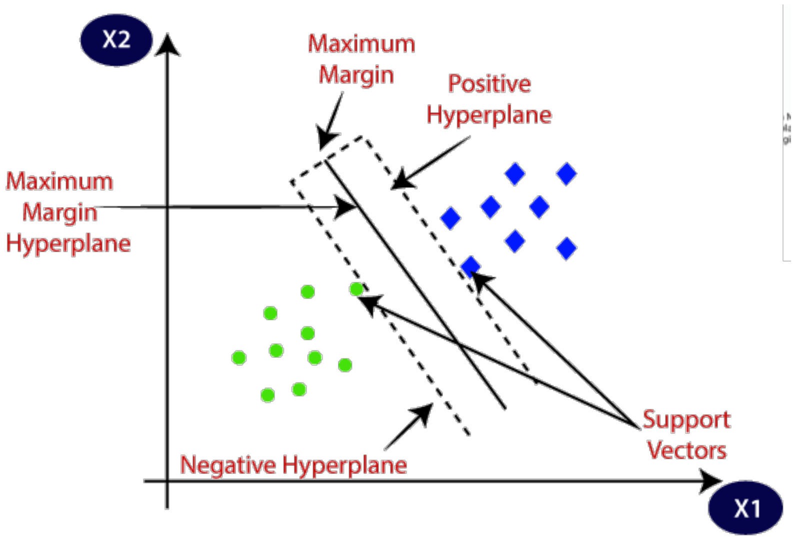


Fig.2 SVMs

3.1.3 Random Forest

Random forest is a machine learning algorithm that can be used for both classification and regression tasks. It works by building a large number of decision trees, and then averaging the predictions of the trees to produce a final prediction.

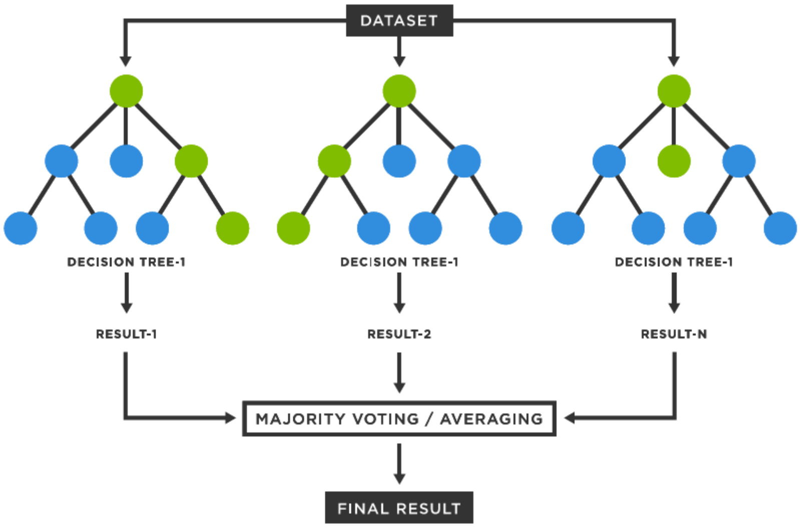


Fig.3 Random forest

$$Entropy = \sum_{i=1}^{C} -p_i * \log_2(p_i)$$

$$Gini = 1 - \sum_{i=1}^{C} (p_i)^2$$

Decision trees are a type of machine learning algorithm that works by splitting the data into different subsets, based on the values of the input variables. The tree is built recursively, until each subset contains only data points from the same class.

3.2 Evaluation Metrics

- *Mean Squared Error(MSE)*

MSE is a most used and very simple metric with a little bit of change in mean absolute error. Mean squared error states that finding the squared difference between actual and predicted value.It represents the squared distance between actual and predicted values. we perform squared to avoid the cancellation of negative terms and it is the benefit of MSE.

$$MSE = \frac{1}{n}\sum(y - \hat{y})^2$$

- *Root Mean Squared Error(RMSE)*

RMSE is the square root of MSE.It is also a good measure of the overall error of the model. It is less sensitive to outliers than MSE, but it is more difficult to interpret

$$RMSE = \sqrt{\frac{1}{n}\sum(y - \hat{y})^2}$$

- *R Squared (R2)*

R-squared is a measure of how well the model fits the data. It is calculated as follows:

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum(y - \hat{y})^2}{\sum(y - \bar{y})^2}$$

$$Where:$$

$(\bar{y})$: "*mean of the values of the outputs*"
$(\hat{y})$: "*predicted value of the output*"
(n) : "*total number of data points*"

3.3 Deep Learning

Deep learning is a type of machine learning that uses artificial neural networks to learn from data. Neural networks are inspired by the structure and function of the human brain, and they are able to learn complex patterns from data. Deep learning models, on the other hand, are able to learn features directly from the data without any human intervention. This is one of the reasons why deep learning models have been able to achieve such impressive results on a wide range of tasks.

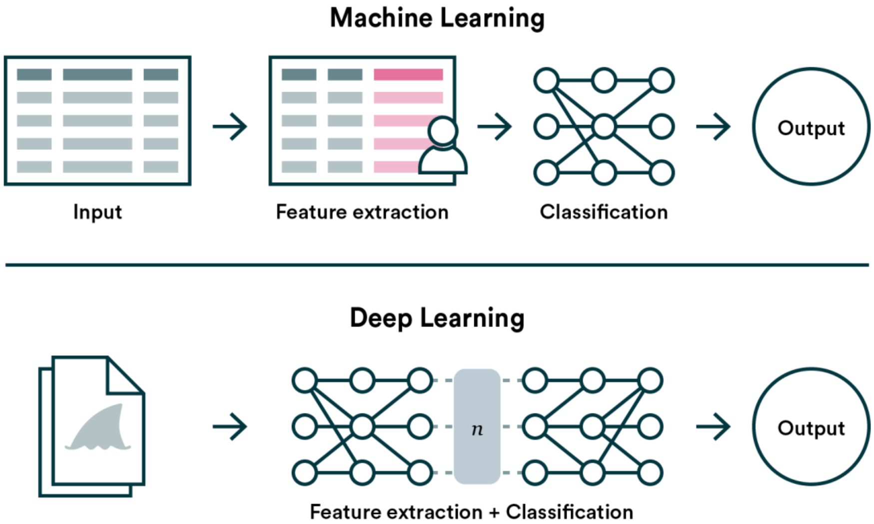
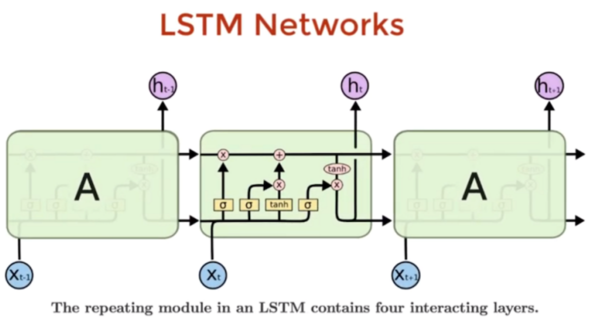


Fig.4 Classical ML vs Deep Learning

Deep learning models are able to learn hierarchical representations of data. This means that deep learning models are able to learn how different features are related to each other, and how they can be used to make predictions. This ability to learn hierarchical representations of data is one of the reasons why deep learning models have been so successful in complex tasks such as image classification and natural language processing.

3.3.1 Long Short-Term Memory (LSTM) networks

Long short-term memory is a special kind of RNN, specially made for solving vanishing gradient problems. They are capable of learning long-term dependencies. They are type of recurrent neural network (RNN) that are specifically designed to model long-range dependencies in sequential data.



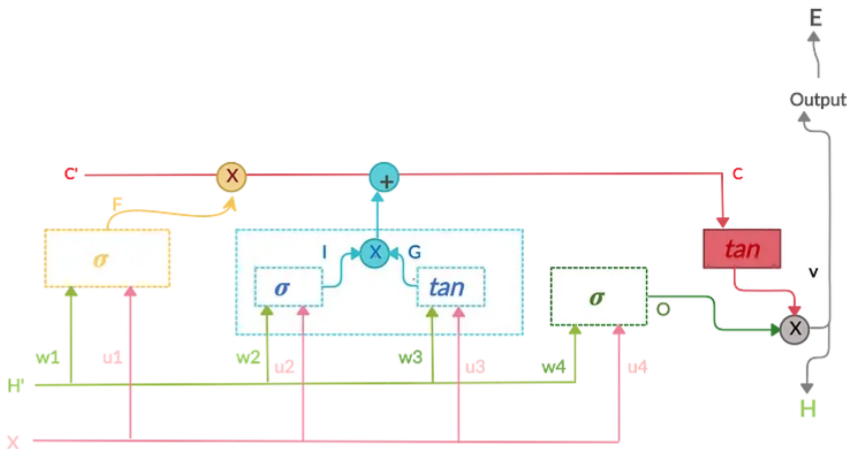


Fig.5 LSTMs

$$F = \sigma\left(u_1 X + \omega_1 H'\right) + \tan() + C'F$$

$$I = \sigma\left(u_2 X + \omega_2 H'\right)$$

$$G = \tan\left(u_3 X + \omega_3 H'\right)$$

$$C = C'F + G$$

$$O = \sigma\left(u_4 X + \omega_4 H'\right)$$

$$Output = v.O.\tan\left(C\right)$$

3.3.2 Transformers

Transformers work by using a self-attention mechanism to learn long-range dependencies in sequential data. The self-attention mechanism allows the network to learn how different parts of the sequence are related to each other, even if those parts are separated by many time steps

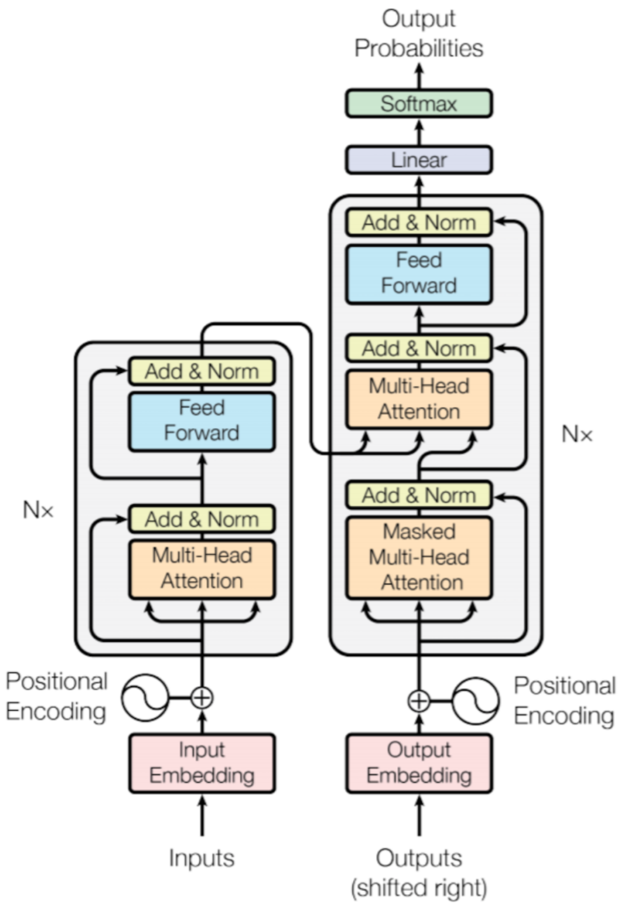


Fig.6 Transformer Architecture

Chapter 4

Experimental setup

4.1 Data

This file contains data from around 4000 high-resolution electric vehicle charging sessions. The data includes interviews with 85 EV drivers who repeatedly use 105 charging stations at 25 locations on Charging Station. Offices include facilities such as research and development centers, manufacturing facilities, laboratories , and corporate headquarters of companies participating in the U.S. Department of Energy (DOE) Workplace Charging Challenge. The file is available in human- and machine-readable *.CSV format. The resolution of the data is correct for the second, the same resolution used in the analysis paper.

sessionId	kwhTotal	dollars	created	ended	startTime	endTime	chargeTim	weekday	platform	distance	userId	stationId	locationId	managerV	facilityTyp	Mon	Tues	Wed	Thurs	Fri
1366563	7.78	0	0014-11-1	0014-11-1	15	17	1.510556	Tue	android	NA	35897499	582873	461655	0	3	0	1	0	0	
3075723	9.74	0	0014-11-1	0014-11-1	17	19	2.177222	Wed	android	NA	35897499	549414	461655	0	3	0	0	1	0	
4228788	6.76	0.58	0014-11-2	0014-11-2	12	16	4.671667	Fri	android	NA	35897499	129465	461655	0	3	0	0	0	0	
3173284	6.17	0	0014-12-0	0014-12-0	19	21	1.768333	Wed	android	NA	35897499	569889	461655	0	3	0	0	1	0	
3266500	0.93	0	0014-12-1	0014-12-1	20	21	0.298611	Thu	android	NA	35897499	414088	566549	0	3	0	0	0	1	
4099366	2.14	0	0014-12-1	0014-12-1	14	15	0.422222	Fri	android	NA	35897499	911231	202527	0	3	0	0	0	0	
5084244	0.3	0	0014-12-1	0014-12-1	15	15	0.64	Fri	android	NA	35897499	920264	461655	0	3	0	0	0	0	
2948436	1.82	0	0014-12-1	0014-12-1	20	21	1.010833	Wed	android	NA	35897499	431796	461655	0	3	0	0	1	0	
3515913	0.81	0	0014-12-1	0014-12-1	17	18	0.179167	Thu	android	NA	35897499	134427	620906	0	3	0	0	0	1	
8490014	1.98	0	0014-12-1	0014-12-1	18	18	0.387778	Thu	android	NA	35897499	207262	928191	0	3	0	0	0	1	
7075912	4.56	0	0014-12-1	0014-12-1	18	20	2.419444	Thu	android	NA	35897499	280221	976902	0	3	0	0	0	1	
5226095	2.05	0	0014-12-1	0014-12-1	14	15	0.690833	Fri	android	NA	35897499	371335	461655	0	3	0	0	0	0	
9342496	5.3	0	0015-01-0	0015-01-0	17	20	3.019444	Wed	android	NA	35897499	405157	976902	0	3	0	0	1	0	
1853945	0	0	0015-01-0	0015-01-0	13	14	0.365	Fri	android	NA	35897499	355208	976902	0	3	0	0	0	0	
2113485	7.85	0.5	0015-01-1	0015-01-1	17	21	4.141111	Mon	android	NA	35897499	944575	928191	0	3	1	0	0	0	
7492587	8.49	0	0015-01-1	0015-01-1	17	19	1.658056	Fri	android	NA	35897499	582873	461655	0	3	0	0	0	0	
1111437	7.52	0	0015-01-1	0015-01-1	19	21	2.157222	Fri	android	NA	35897499	500856	976902	0	3	0	0	0	0	
4716378	5.38	0	0015-01-2	0015-01-2	18	20	2.294167	Wed	android	NA	35897499	250527	976902	0	3	0	0	1	0	

Fig.7 Dataset view in CSV format

4.2 Data Exploration

Data exploration is the process of examining and analyzing data to discover patterns, trends, and insights. It is the first step in any data analysis project, and it is essential for understanding the data and developing hypotheses. Data exploration is an iterative process. As you explore the data, you may discover new patterns and trends that lead you to ask new questions. Data exploration is a powerful tool that can be used to gain insights from data and to solve complex problems.

	sessionId	kwhTotal	dollars	created	ended	startTime	endTime	chargeTimeHrs	weekday	platform	...	managerVehicle	facilityType
0	1366563	7.78	0.00	0014-11-18 15:40:26	0014-11-18 17:11:04	15	17	1.510556	Tue	android	...	0	3
1	3075723	9.74	0.00	0014-11-19 17:40:26	0014-11-19 19:51:04	17	19	2.177222	Wed	android	...	0	3
2	4228788	6.76	0.58	0014-11-21 12:05:46	0014-11-21 16:46:04	12	16	4.671667	Fri	android	...	0	3
3	3173284	6.17	0.00	0014-12-03 19:16:12	0014-12-03 21:02:18	19	21	1.768333	Wed	android	...	0	3
4	3266500	0.93	0.00	0014-12-11 20:56:11	0014-12-11 21:14:06	20	21	0.298611	Thu	android	...	0	3

5 rows × 24 columns

Fig.8 Data head

Outliers that have large or small values compared to most observations have a large impact on the model results. After exploring the data, the max values looked too far from reasonable values compared to the mean values, so outliers were present in the data

	kwhTotal	dollars	startTime	endTime	chargeTimeHrs	distance
count	3395.000000	3395.000000	3395.000000	3395.000000	3395.000000	2330.000000
mean	5.809629	0.118268	13.743446	16.455965	2.841488	18.652378
std	2.892727	0.492562	3.204370	3.406732	1.507472	11.420571
min	0.000000	0.000000	0.000000	0.000000	0.012500	0.856911
25%	4.350000	0.000000	11.000000	14.000000	2.110278	5.135871
50%	6.230000	0.000000	13.000000	16.000000	2.808889	21.023826
75%	6.830000	0.000000	17.000000	20.000000	3.544167	27.285053
max	23.680000	7.500000	23.000000	23.000000	55.238056	43.059292

Fig.9 Data description

The data which contain null values will make the analysis more complicated and affect the result of the model. Null values were found in this data.

```
kwhTotal       0
distance    1065
dollars        0
dtype: int64
```

Fig.10 Missing values

Various methods are present for imputing the null values in the dataset, such as Mode imputation, Median imputation, K-nearest neighbors imputation, Bayesian imputation.

We are using KNN imputation for the project.K-nearest neighbors imputation (KNN imputation) is a method for imputing missing values in a dataset by using the values of the k most similar rows in the dataset.

4.3 Data Visualization

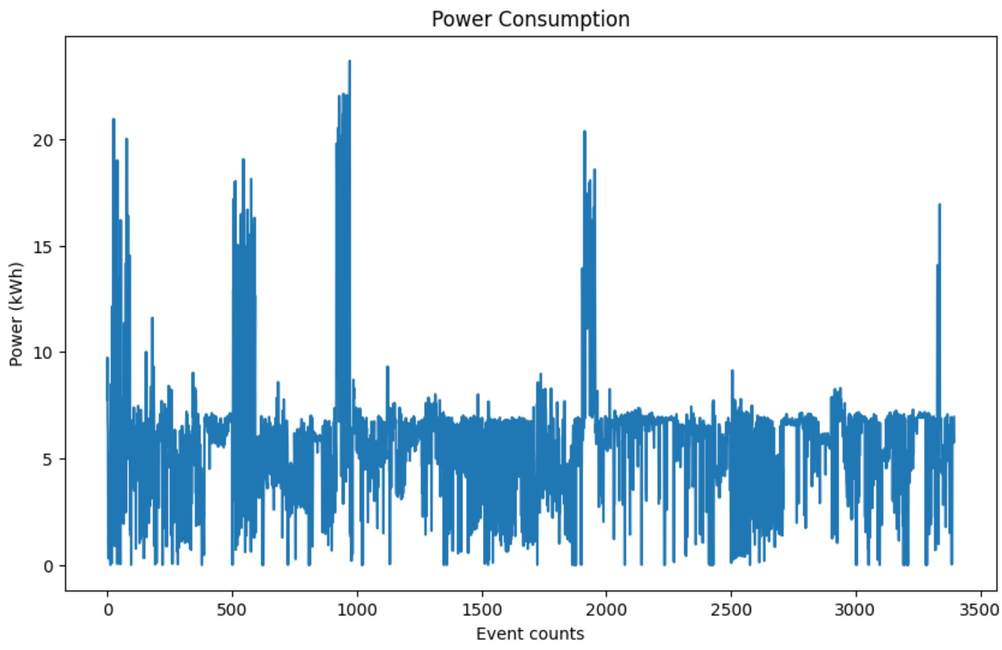


Fig.11 Power consumed vs Event

4.4 Training and Evaluation

4.4.1 Data Preparation

To train the model we split the data into a training set (80%) and a testing set (20 %). Also, we split training data into 20 % for the validation set and 80 % for training. The validation set is used to validate the model performance during training.The input layer of the LSTM model requires 3D input, Where the first dimension represents the sample size, and the second represents the number of time steps. And the third dimension represents the number of features used to train the model [number of samples, time steps, and number of features].

4.4.2 Model Structure

Stacked LSTM is used which contains four hidden LSTM layers one on top of another. The model building starts by creating a sequential model using Keras. Then we added the hidden LSTM layers to the model. Each layer had 50 units. A dropout layer was added to help in preventing overfitting. The last layer was the dense layer, which does the below operation on the input and returns the output. To update network weights iterative based on training data we used the Adam optimization algorithm which was used famously instead of the classical gradient descent procedure. Also, we used root mean squared error (RMSE) to test the model performance. A smaller RMSE means that our model is performing better.

4.4.3 Model Training

Random Forest and SVM was trained for the training set and validation set for testing the result through the training process. Training the LSTM model was done using the training set and the validation dataset for testing the results through the training process. The learning algorithm worked through the entire training dataset 60 times (Epoch), and the model weights were updated after each batch where the batch size is 20.

```
Epoch 43/60
65/65 [==============================] - 12s 182ms/step - loss: 0.0028 - val_loss: 0.0016
Epoch 44/60
65/65 [==============================] - 12s 182ms/step - loss: 0.0024 - val_loss: 0.0016
Epoch 45/60
65/65 [==============================] - 12s 181ms/step - loss: 0.0022 - val_loss: 0.0017
Epoch 46/60
65/65 [==============================] - 13s 198ms/step - loss: 0.0023 - val_loss: 0.0015
Epoch 47/60
65/65 [==============================] - 12s 181ms/step - loss: 0.0022 - val_loss: 0.0017
Epoch 48/60
65/65 [==============================] - 13s 195ms/step - loss: 0.0023 - val_loss: 0.0020
Epoch 49/60
65/65 [==============================] - 12s 181ms/step - loss: 0.0018 - val_loss: 0.0015
Epoch 50/60
65/65 [==============================] - 13s 197ms/step - loss: 0.0017 - val_loss: 0.0016
Epoch 51/60
65/65 [==============================] - 12s 185ms/step - loss: 0.0020 - val_loss: 0.0012
Epoch 52/60
65/65 [==============================] - 12s 183ms/step - loss: 0.0017 - val_loss: 0.0011
Epoch 53/60
65/65 [==============================] - 12s 184ms/step - loss: 0.0015 - val_loss: 0.0010
Epoch 54/60
65/65 [==============================] - 13s 198ms/step - loss: 0.0016 - val_loss: 0.0013
Epoch 55/60
65/65 [==============================] - 12s 182ms/step - loss: 0.0015 - val_loss: 0.0011
Epoch 56/60
65/65 [==============================] - 12s 183ms/step - loss: 0.0015 - val_loss: 0.0010
Epoch 57/60
65/65 [==============================] - 13s 195ms/step - loss: 0.0016 - val_loss: 0.0010
Epoch 58/60
65/65 [==============================] - 12s 182ms/step - loss: 0.0014 - val_loss: 0.0012
Epoch 59/60
65/65 [==============================] - 12s 183ms/step - loss: 0.0015 - val_loss: 0.0013
Epoch 60/60
65/65 [==============================] - 12s 185ms/step - loss: 0.0016 - val_loss: 0.0014
```

Fig.14 Training the LSTM model

The check after training was to compare the training loss against the validation loss. The result shows that the two values were low, and no overfitting was detected.

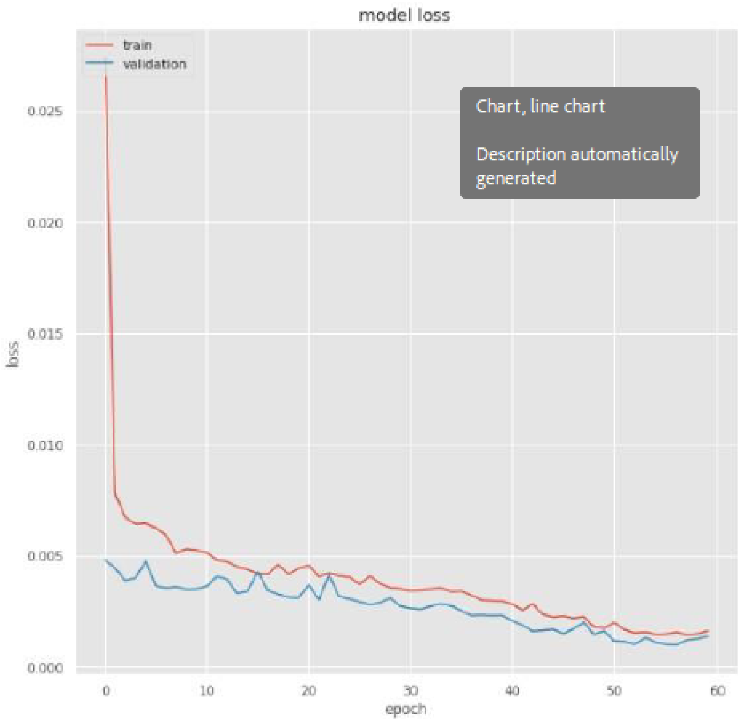


Fig.14 Loss comparison for training and validation set

4.5 Conclusion

The Random Forest and SVM model gave good result for forecasting, the Random Forest model performed better compared to SVM on the dataset.We can use the models to form an ensemble model and use it for prediction and compare the results. The deep learning model was built successfully using four LSTM layers, one Dropout, and one output layer. This project used the Adam optimization algorithm and root mean squared error (RMSE) to test the model performance. The result has shown good results when applying the stacked LSTM for short-term predictions. The model was tested first on training, validation, and test datasets.

Chapter 5

Work Plan

- **August**: Select a problem statement, read previous works and solutions proposed.
- **September**: Data gathering, exploring, preprocessing and model selections.
- **October**: Training and tune the forecasting models and work exploring advanced modern models.
- **November**: Evaluate and comparing the various models and identifying the best solution.

References

[1] Prediction of EV Charging Behavior Using Machine Learning by SAKIB SHAHRIAR published August, 2021

[2] Electricity Consumption Prediction using Machine Learning Models Article · June 2020 by Swati Sucharita Barik and Sasmita Nayak

[3] Agnihotri, N. (n.d.). Classification of machine learning algorithms. Retrieved from engineersgarage.com: https://www.engineersgarage.com/machine-learning-algorithms-classification/

[4] Energy Consumption Prediction using Machine Learning by MAIS NIJIM

[5] Olah, C. (2015, August 27). Understanding LSTM Networks. Retrieved from colah.github.io: https://colah.github.io/posts/2015-08-Understanding-LSTMs/

[6] Peixeiro, M. (2019, August 7). The Complete Guide to Time Series Analysis and Forecasting. Retrieved from towardsdatascience.com: https://towardsdatascience.com/the-complete-guide-to-time-series-analysis-and-forecasting-70d476bfe775

[7] Time Series - LSTM Model. (2022, 10 16). Retrieved from tutorialspoint.com: https://www.tutorialspoint.com/time_series/time_series_lstm_model.htm

[8] Brownlee, J. (2016, August 6). 11 Classical Time Series Forecasting Methods in Python (Cheat Sheet). Retrieved from machinelearningmastery.com: https://machinelearningmastery.com/time-series-forecasting-methods-in-python-cheat-sheet/