Ev Charging Cost Prediction and Nearest Charging Station Recommendation Using ML

Kanishka S
Department of Electrical and Electronics Engineering
Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham, India
cb.en.u4elc22019@cb.students.amrita.edu

Rahul B
Department of Electrical and Electronics Engineering
Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham, India
cb.en.u4elc22046@cb.students.amrita.edu

Madhu Priya Y V
Department of Electrical and Electronics Engineering
Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham, India
cb.en.u4elc22060@cb.students.amrita.edu

Dr.M.venkatesh Kumar
Department of Electrical and Electronics Engineering
Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham, India
m_venkateshkumar@cb.amrita.edu

Abstract — As a key pillar of smart transportation in smart city applications, electric vehicles (EVs) are becoming increasingly popular for their contribution to reducing greenhouse gas emissions. The article presents a novel information system for optimizing urban EV charging costs based on power consumption and providing the nearest charging station.To develop an economic charging cost algorithm, the first random forest [1] algorithm is used to do a regression of charging cost approximately, further improving system flexibility and assuring a better cost and availability of charging stations.The cost of charging EVs based on the demand is accurately predicted using random forest [1]. Algorithms for machine learning identify trends and make real-time decisions based on patterns in diverse data sources. The criteria for the decision model include electricity pricing, demand, peak hours, charging time, and battery capacity. The analysis employs public data on electricity pricing, energy consumption, availability, and nearest station.To develop a futuristic vehicle charging station that uses ML to predict charging cost, and time and provide best available option to the user, which helps in minimizing the rush at the charging station and maximizing the charging efficiency.

Keywords- EV charging cost prediction, Nearest charging station recommendation, linear regression [2], random forest [1], Multi-Layer Perceptron regression (MLP) [3], Haversine Distance calculation.

I INTRODUCTION

In response to EV users' requests, this study presents a modern charging recommender system for electric vehicles using machine learning, geospatial analysis, and optimization techniques. The system estimates the charging cost fare using Random Forest, Linear Regression, and Multi-Layer Perceptron (MLP) [3] algorithms, based on the location, battery capacity, and current internal charge level. The distance between the user and charging stations is computed using a Haversine formula [4], which helps to choose stations appropriately depending on distance. Z-score normalization [5] is employed in the data preprocessing pipeline to adjust features and ensure consistent and accurate inputs to the machine learning models. By providing cost estimates and smart station suggestions, our approach enhances the user experience whilst dealing with the central problem of the user's EV charging cost, convenience, and energy efficiency.

In order to achieve optimal prediction performance from the charge suggestion model, several machine learning methods, such as random forest [1] regression, linear regression, and MLP [3] neural network, are employed. After training these models on such variables as the latitude and longitude, the battery capacity, and the amount of charge needed, the accuracy of each model in predicting the costs is determined by the use of such metrics as Mean Squared Error (MSE) and R^2 [7]. Comparative assessments are graphically depicted by use of AI and Tal plots, which show the actual expenses and the expenses predicted by the model at a given time, and provide the information and the accuracy and reliability of the model. Random Forest and Neural Network models show more flexibility as they learn the nonlinear relationships present in the data to make predictions. Such as technique enables the development of a smart and data-centric charging suggestion system which is of great importance in the enhancement of the existing EV infrastructure.

In this approach, a user-oriented recommendation system is implemented, allowing for the selection of a charging station with the optimal distance and price range, putting the system above basic predictions. Visualizations that depict differences and similarities of actual costs as well as their estimates of various models are important additions aimed at improving understanding. It is in this respect that Customers are also educated on the costs and time involved in the process in form of statistics such as an average charge cost and how long it takes to charge fully. This technology further allows users to access cheap stations without going out of their way by providing an economic distance charging station optimization algorithm that determines the best to charge based on cost and distance. With its practical and user-friendly output, this recommendation system tremendously improves EV charging and contributes towards building a green EV ecosystem. The objective of this work is to minimize the recharge cost by proposing a function based on the distance constructed with Haversine Formula [14], which is connected to a cost function, and reaching the goal of best positions for charging stations.

II DATA PROCESSING

1 .1 Overview:

To improve the accuracy and efficiency of the machine learning models used in a cost prediction and nearest charging station finding including data processing and normalization are important techniques in this research. There are variables including geographic coordinates, charge required, battery level, and cost per unit The data must be pre-processed and standardized to speed up the machine-learning process before the prediction and classification process has been used.

1.2 Cleaning of Column Names:

Raw data often has missing columns with a white foreground or background, which can cause problems when displaying or manipulating data. The first step in data cleaning is to remove trailing spaces from column names to ensure consistency and improve model training to provide information and prevent possible errors.

1.3 Feature Selection:

Feature engineering encompasses transforming raw data into useful features for the model, including selecting the most pertinent predictor variables. The model comprises both outcome and predictor variables

To enhance the model performance and interpretability required features are filtered from the dataset like charging cost, demand, charging speed, and so on. Selecting features enhances model accuracy by training train and testing data fast.

1.4: Z-Score Normalization:

To standardize the data provided a certain number of scores were subjected to Z-score normalization [5]. This method uses the statistics' comparability across several factors and units by transforming the data into a format with a zero mean and one standard deviation. Used to find the Z-score normalized form

$$Z = (X-\mu) \div (\sigma)\ldots\ldots\ldots\ldots\ldots..(1)$$

where Z is the Z-score value, x is a data point, μ is the mean of the values, and σ is the standard deviation of the values.

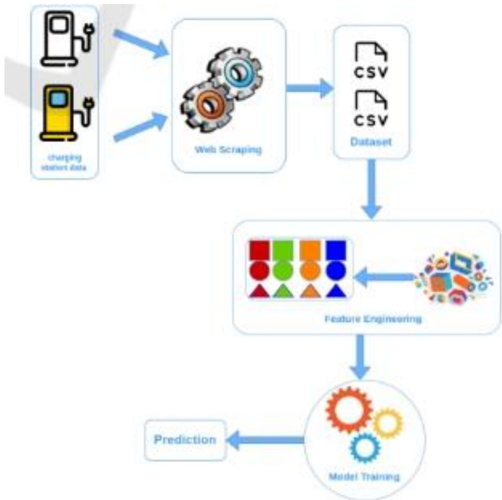

Fig-1 Data processing

III METHODOLOGY

I. IN THIS SECTION, THE MAIN FOUNDATIONS OF THE NEWLY PROPOSED NOVEL ARE:

1.) EV Charging Cost Prediction,
2.) Nearest Station Recommendation.

This section describes the construction of the proposed model.

1.EV CHARGING COST PREDICTION

2.1 Linear Regression:

Linear regression (LR) can be used to model the mathematical relationship between the output variable and one or multiple input variables (multiple LR). Linear regression shows a linear connection between a dependent (y) and one or more independent (y) variables. Linear regression shows a linear connection; therefore, it may be used to identify how the dependent variable's value varies with the independent variable's value.

$$y=(bo+b1*x1+b2*x2+\cdots+bn*xn)\ \ldots\ldots\ldots\ldots\ (2)$$

To apply linear regression [2] using X_train and y_train datasets and evaluate its performance, a linear regression model was first created as LR_Model. LR is particularly useful when the dataset is linearly separable and the algorithm itself is very simple to implement. Overfitting is a major challenge in training ML algorithms, it occurs when a given model performs exceptionally well during the training phase, The custom function evaluate model was created to calculate the performance metrics for the X_test and y_test set of tests. This function used the y_pred = model. predict(X_test) to make a prediction. Analytical metrics Mean Squared Error (MSE) [6] and R-squared (R2) were calculated using the functions mean_squared_error and r2_score the MSE and R2 values were published to give a quantitative sense of the quality of the model. A line plot (plt.plot) was also used to visualize the predictions, complete with lines and colors for easy differentiation. In addition to contributing to iterative machine learning analysis, this evaluation framework provides a systematic approach for evaluating linear regression performance.

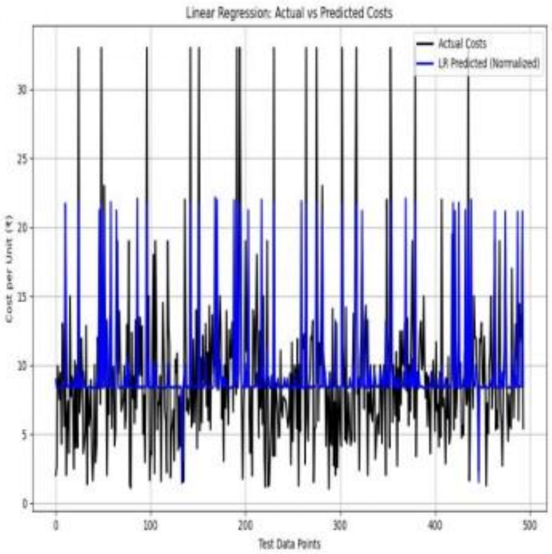

Fig-2
Plot for linear regression

2.2 *Multi-Layer Perceptron (MLP)*

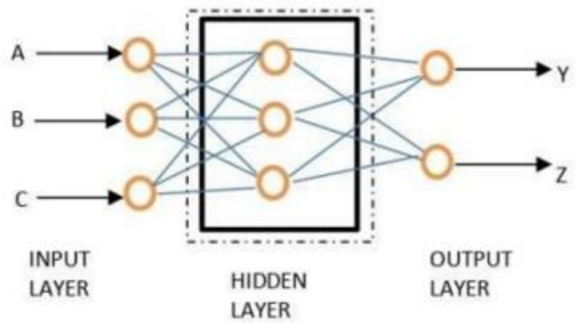

Fig-3
Neural Network

This study employs a multi-layer perceptron (MLP) [3] neural community model to address and study complicated and non-linear connections inside the records. This choice was made due to the MLP's capability to apprehend complex styles in information that a linear model. The model can successfully research from non-linear information without requiring immoderate schooling time or sources. It balances computational effort with version complexity...

This method ensures that the model reaches a very good answer without taking too long to system. To prevent overloading the model with computations, a limit of 1000 training cycles becomes set. Additionally, a random set of 42 situations turned into used across numerous training tiers to ensure the outcomes might be reproduced. This consistency in random beginning points made the effects dependable and more desirable, as well as the robustness of comparing the model's predictive energy.

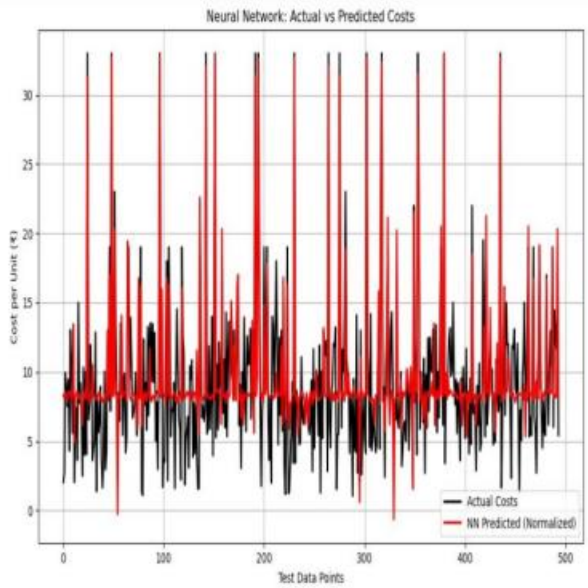

Fig-4
Plot for MLP regression

2.3 Random Forest

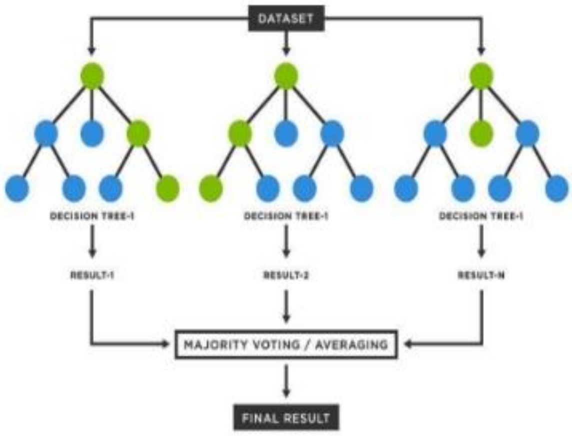

Fig-5
Random Forest

RF represents a versatile machine learning algorithm widely employed for classification and regression tasks. In this study, the exceptional capability of RF model for regression is leveraged to predict the cost of charging electric vehicle based on the electricity demand during the peak hours.

Random forest uses a part of the same dataset that can be utilized to train and test different decision trees after which every decision tree gives its output and by majority voting and averaging, RF produces the output. more number of trees helps in reducing the problem of over-fitting [10], [11]. pros of the Random Forest algorithm:

• less training time than other algorithms.
• High accuracy even for large datasets.

• good accuracy even with missing data.

In order to predict the charging cost, a Random Forest Regressor version is used the dataset becomes pre-processed through isolating it into training and testing sets. Random Forest is used as it uses less training time as compared to other algorithms. High accuracy even for large dataset. • good accuracy even with missing data

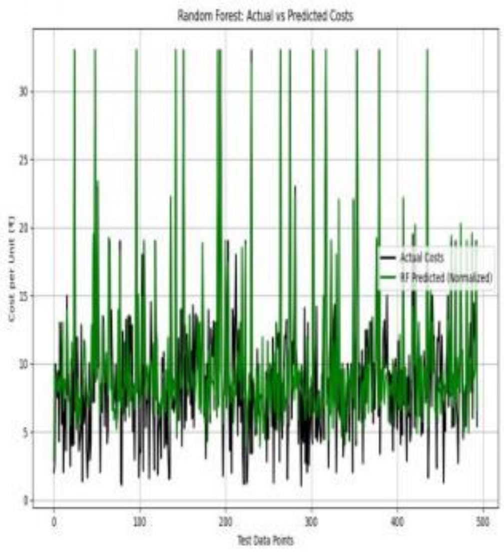

Fig-6
Plot for Random Forest

2.NEAREST STATION RECOMMENDATION:

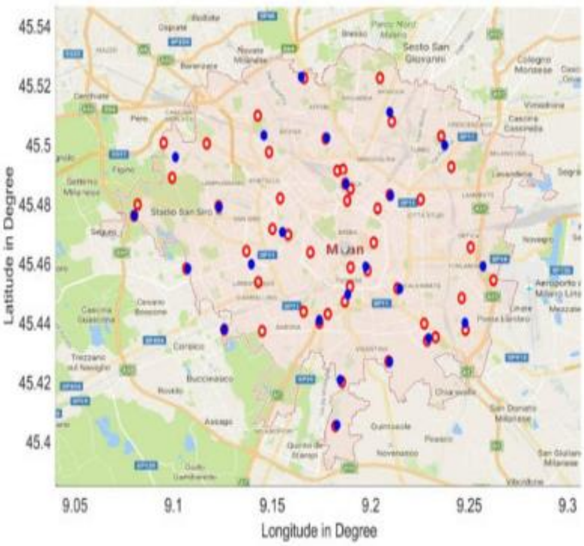

Fig-7
Geolocations for nearest station recommendation

Haversine distance formula:
Haversine formula [4] is used to calculate the distance of a particular point by using the latitude and longitude of user's location to the desired location by using earths radius which is approximately R=6371km.

$$a = \sin^2\left(\frac{\Delta Lat}{2}\right) + \cos(Lat_1)\cdot\cos(Lat_2)\cdot\sin^2\left(\frac{\Delta lon_0}{2}\right) \ldots(3)$$

$$c = 2\cdot\{atan2\}\left(\sqrt{\{a\}}, \sqrt{\{1-a\}}\right) \ldots(4)$$

$$d = R\cdot c \ldots(5)$$

To find the distance of EV charging station from the location of the user Haversine formula is used and the top 3 stations with less distance is shown to the user. As the Haversine formula is used it uses trigonometric functions, so convert the calculated values to radians before processing on to the trigonometric functions. The calculated value is finally multiplied with the earth's distance. The process is repeated for the entire row by using the latitude and longitude of each location.

When the battery charge level is less than 30 percent, the intimation of the 3 best stations pops up as an alert. When the battery charge level is more than 30 percent, only 1 best station pops up for information. The maximum limit of the charging station is fixed within a 50km radius from user location, the car can go to the minimum distance of 50km using 30 percent of charge based on the battery capacity of the vehicle.

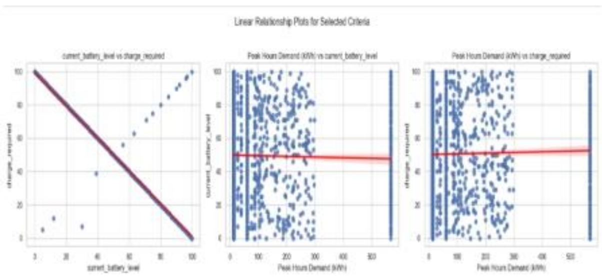

Fig-5
Analysis of Battery Level, Charge Required, and Peak Hours Demand for EV Charging Optimization using linear regression

IV EV CHARGING DATASET

The success of a good predictive ML model depends on the quality of the dataset. In this section, the commonly used datasets for studying EV charging behaviour are discussed.
The dataset contains the charging station details of New Delhi, Nagpur, Kolkata, Chennai, Coimbatore, Palakkad, Thiruvananthapuram, Kochi, Malappuram, Kottayam, Thrissur, Panaji, Raipur, Ahmedabad, Kasaragod, Mapusa Dehradun, Gurgaon, Dhanbad. The dataset contains up to 2700 charging stations in India. This provides a good platform for studying EV charging stations, power supply demand, and variations of cost due to it and charging behaviour.

IV RESULTS

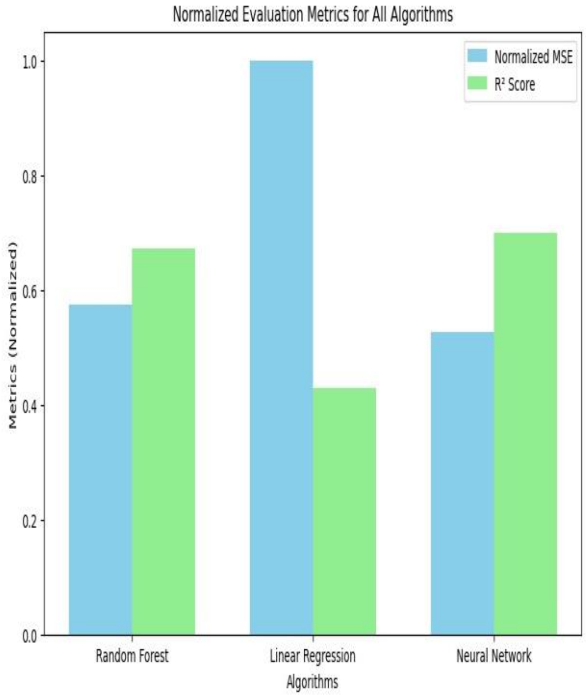

Fig-9
Results of regression

```
# Example Usage
latitude = 11.001994
longitude = 76.974032
battery_level = 20.8  # Example battery level
battery_capacity = 45.0
```

[Fig-9] User input

```
PREDICTED AVERAGE COST OF CHARGING 'EV' :
₹498.96

ESTIMATED CHARGING TIME TO FULL FROM (20.8% to 100%) :
2.38 hours
RECOMMENDED CHARGING STATIONS:
                                   address        city  capacity  \
2653         12/88, Race Course, Gopalapuram  Coimbatore      15.0
2645  Race Course Rd, Near Park, Race Course  Coimbatore      11.0
2655                       278/74, Ram Nagar,  Coimbatore       7.0

      cost_per_unit  distance_km  predicted_cost
2653           12.0     0.000000          427.68
2645           15.0     0.479970          534.60
2655           15.0     1.665035          534.60

BEST USER/COST-FRIENDLY CHARGING STATION:
Address: 12/88, Race Course, Gopalapuram
City: Coimbatore
Capacity: 15.0 kW
Cost per Unit: ₹12.00
Distance: 0.00 km
Predicted Cost: ₹427.68
```

Fig-10
Recommendation of charging cost and stations(output)

With the user's current Battery status and charging needs this paper provides a complete solution to estimate EV charging cost and recommend local charging stations. The dataset must first be pre-processed. This involved editing for consistency and removing extra columns… The system determines the difference between the customer's location and a specific station using the haversine attribute, then filters stations within the desired range for access. To estimate the unit cost` at the charging station. Three prediction models are trained on a special task: random forest [1]. linear, multilevel regression; Mean Squared Error [6] (MSE) and R-squared (R2) metrics are used to evaluate perceptron's (MLP) [3] and models, and the accuracy of each model is judged by comparing the visual predictions with the results.

This idea device depends on the level and capacity of the battery. The buyer can predict the charging time and calculate the price. Top stations are highlighted with important information such as address, and probability. and expected value Nearby stations to Peak 3 are recommended based on proximity and cost-effectiveness. Normal charge display and total station options. The system will highlight a charging recommendation if the battery phase is below 30%, otherwise it will send the user to the most economical station. This method combines proximity cost estimation and charging time calculations into a smooth, easy-to-use solution, making it easier for EV customers to make informed decisions.

V CONCLUSION

The process begins with raw data input that includes vehicle battery capacity and battery status information. Z score normalization [5] is used to normalize the data to a consistent scale across symptoms. Prediction of cost is done through a random forest [1] regression by analyzing the charge of EV according to specific behaviors, such as current charge level, battery capacity, and cost of electricity based on the demand during peak hours. The time to charge the EV from current battery level to 100% and also the best cost-efficient station from a list of 3 stations within 50 kilowmeters were also found. The cost was predicted using random forest, MLP [3], and linear regression [2]. MLP and linear regression are used for comparison. Add that when comparing the normalized evaluation metrics and also the plots between cost predicted from algorithms and actual cost. The MSE value of MLP and Random Forest is less than linear regression, and the R^2 [7] value of MLP and Random Forest is closer to 1. Also comparing the plots of predicted cost, the cost predicted from NLP and Random Forest aligns best with the actual cost. So compared to linear regression, MLP is the best, and random forest is next to it. Whereas linear regression is not capable of finding the charging cost accurately.

VI REFERENCES

[1] S. Shahriar, A. R. Al-Ali, A. H. Osman, S. Dhou, and M. Nijim, "Prediction of EV Charging Behavior Using Machine Learning," *IEEE Access*, vol. 9, pp. 111576-111586, Aug. 2021, doi: 10.1109/ACCESS.2021.3103119.

[2] M.A.Elkasrawy, S.O.Abdellatif, G.A.Ebrahim, H.A.Ghali, "Real-Time Optimization in Electric Vehicle Stations Using Artificial Neural Networks," Electrical Engineering (2023) 105:79–89 doi:/10.1007/s00202-022-01647-9

[3] M. A. Elkasrawy, S. O. Abdellatif, G. A. Ebrahim, and H. A. Ghali, "Real-time optimization in electric vehicle stations using artificial neural networks," *Electrical Engineering*, vol. 105, pp. 79-89, 2023, doi: 10.1007/s00202-022-01647-9.

[4] Y. Mohammed, A. Manoharan, S. Kappagantula, and H. Manoharan, "Optimizing Electric Vehicle Charging Costs Using Machine Learning," *Ingénierie des Systèmes d'Information*, vol. 29, no. 3, pp. 1085-1095, June 2024, doi: 10.18280/isi.290326.

[5]Y.Yang, Y.Zhang, and X.Meng, "A Data-Driven Approach for Optimizing the EV Charging Stations Network,"DoP:June 2020, doi:10.1109/ACCESS.2020.3004715

[6]K.C.Akshay, G.H.Grace, K.Gunasekaran, R.Samikannu, "Power Consumption Prediction for Electric Vehicle Charging Stations and Forecasting Income," Sci Rep. 2024 Mar 18; 14:6497. doi: 10.1038/s41598-024-56507-2

[7] S. Shahriar, A. R. Al-Ali, A. H. Osman, S. Dhou, and M. Nijim, "Machine Learning Approaches for EV Charging Behavior," *IEEE Access*, vol. 9, Aug. 2021

.

[8] M.Gjelaj, N.B.Arias, C.Traeholt, and S.Hashemi, "Multifunctional Applications of Batteries Within Fast-Charging Stations Based on EV Demand-Prediction of the User's Behavior," The Journal of Engineering, 2019(18):4869–4873, 2019.

[9] H.Rathor, H.K.Meena, P.Jain "Prediction of EV Consumption Using Random Forest and XGBoost," 2023 International Conference on Power Electronics and Energy (ICPEE) 978-1-6654-7058-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICPEE54198.2023.10060798

[10] M. Majidpour, C. Qiu, P. Chu, R. Gadh, and H. R. Pota, "A novel forecasting algorithm for electric vehicle charging stations," in 2014 International Conference on Connected Vehicles and Expo (ICCVE). IEEE, 2014, pp. 1035–1040.

[11] A. Liaw, M. Wiener et al., "Classification and regression by randomforest," R news, vol. 2, no. 3, pp. 18–22, 2002.

[12] Y.Lu, Y.Li, D.Xie, E.Wei, X.Bao, H.Chen, X.Zhong, "The Application of Improved Random Forest Algorithm on the Prediction of Electric Vehicle Charging Load," *Energies* **2018**, *11*(11), 3207; **doi.org/10.3390/en11113207**

[13] M.B.Arias, S.Bae, "Electric Vehicle Charging Demand Forecasting Model Based on Big Data Technologies," september 2016

[14] M. M. Morcos, Curtis R. Mersman, G.D. Sugavanam and Norman G. Dillman, "Battery Chargers for Electric Vehicles", *IEEE Power Engineering Review*, vol. 20, no. 11, pp. 8-18, Nov 2000.

[15] M.E.Amoli, K.Choma and J.Stefani, "Rapid-charge electric-vehicle stations", *IEEE Trans. Power Del.*, vol. 25, no. 3, pp. 1883-1887, June 2010.