2023

Interpreting Energy Utilisation With Shapley Additive Explanations by Defining a Synthetic Data Generator for Plausible Charging Sessions of Electric Vehicles

Prasant Kumar Mohanty
National Institute of Technology, India

Gayadhar Panda
National Institute of Technology, India

Malabika Basu
Technological University Dublin, malabika.basu@tudublin.ie

See next page for additional authors

Funder: Technological University Dublin, Ireland, under grant PB04433.

2023 5th International Conference on Energy, Power and Environment: Towards Flexible Green Energy Technologies (ICEPE) | 979-8-3503-1312-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICEPE57949.2023.10201554

Interpreting energy utilisation with Shapley additive explanations by defining a synthetic data generator for plausible charging sessions of electric vehicles

Prasant Kumar Mohanty
Department of CSE
National Institute of Technology
Meghalaya, India
prasantmohanty.r@gmail.com

Gayadhar Panda
Department of EE
National Institute of Technology
Meghalaya, India
gayadhar.panda@nitm.ac.in

Malabika Basu
Technological University Dublin
Ireland
malabika.basu@tudublin.ie

Diptendu Sinha Roy*
Department of CSE
National Institute of Technology
Meghalaya, India
diptendu.sr@gmail.com

***Abstract*—Electric vehicles (EVs) are an effective solution for reducing reliance on non-renewable energy sources. However, the lack of charging infrastructure and concerns over their range are some of the biggest hurdles to adopting EVs. Charging infrastructure for EVs is, however, on the rise. Proper planning of charging stations vis-à-vis road networks and related points of interest such as transportation hubs, schools, shopping centres, etc., alongside such roads become vital to laying out a plan for such infrastructure, particularly for developing countries like India where EV adoption is relatively in a nascent stage. Synthetic datasets can help overcome these hurdles and promote EV adoption. This article presents a synthetic dataset mechanism for EV charging infrastructure planning, taking the Indian city of Berhampur, Odisha with its existiing EV charging infrastructure as a reference. The dataset includes information on the number of charging sessions for EVs, allocation to chargers in EVCS, reach time, charging start and end time, waiting time, total time spent at EVCS, total charged amount, energy used, and cost for charging. This information can help city planners and utilities identify the optimal locations for charging stations and plan for future charging infrastructure augmentation. The dataset can also be used to predict energy usage for the near future and identify the key factors affecting the planning with the help of Explainable AI (XAI) techniques. This information can help forecast the demand for charging services and optimize energy usage in the city. The article contributes to the EV charging behaviour and infrastructure planning and aims to promote broader EV adoption for future sustainable transportation.**

***Index Terms*—Electric vehicles, Charging infrastructure, Synthetic datasets, Sustainable transportation, Explainable AI (XAI).**

I. Introduction

Humanity faces a significant problem by relying on non-renewable fossil fuels, which are finite and will eventually deplete. However, there are various ways that individuals, organizations, and governments can tackle this issue. One effective solution is to shift from using fossil fuels towards renewable energy sources like solar, wind, geothermal, and hydropower. These sources offer a cleaner and more sustainable energy option than fossil fuels. Measures like retrofitting buildings, using energy-efficient appliances, and reducing energy consumption in transportation can lower the need for fossil fuels. Alternative fuel vehicles, including electric vehicles, hydrogen fuel cell vehicles, and biofuels, can also help reduce reliance on nonrenewable energy sources. Electric vehicles produce zero emissions at the tailpipe, which can help reduce air pollution and combat climate change. As smart cities aim to be more sustainable, transitioning to EVs can significantly achieve this goal.

Despite the numerous benefits of electric vehicles (EVs), there are still some hurdles in adopting them for a new city. One of the biggest hurdles in adopting EVs is the lack of charging infrastructure in many cities. People may be hesitant to switch to EVs if they are worried about being able to find a charging station when they need it. But for petrol or diesel vehicle, such a problem does not exist. Another concern people have with EVs is range anxiety, which is the fear of running out of battery power before reaching a destination. Even though the range of EVs has been increasing in recent years, people may still be hesitant to switch to EVs if they are concerned about being able to travel long distances without a recharge, as there are very few charging stations compared to fuel stations.

Creating synthetic electric vehicle datasets can help remove hurdles to adopting EVs in a new city. Synthetic datasets can help predict and identify the locations where the demand for charging stations is high based on population density, demographics, and travel patterns. This can assist city planners and utility companies identify the optimal locations for charging stations and help plan future demanding EV charging infrastructure needs. Such synthetic datasets can help estimate EVs' range and battery performance based on various factors, such as vehicle type, driving conditions, and weather patterns. This can help address range anxiety by providing more accurate information about how far an EV can travel on a single charge and identifying the best route for EVs to take to optimise battery performance. These datasets can be used to create simulations and visualisations that help educate people about the benefits of EVs and how they work. This can help raise public awareness and dispel myths and misconceptions about EVs, making them more appealing and accessible to a broader audience.

The article contributes to the EV charging behaviour and infrastructure planning field by providing a synthetic dataset that can be used for research and analysis. It offers valuable insights into the challenges and opportunities associated with EV charging in a city like Berhampur, Odisha, with limited charging infrastructure, as only 4 EV charging stations are available. The dataset also includes information on the number of charging sessions for EVs, their allocation to chargers in EVCS, reach time, charging start and end time, waiting time, total time spent at EVCS, total charged amount, energy used, and cost for charging. This information can help city planners and utility companies identify the optimal locations for charging stations and plan for future charging infrastructure needs. It can help promote broader EV adoption and build a more sustainable transportation system. It also aims to predict energy usage for the near future and identify the topmost responsible factors with the help of explainable AI techniques. This information can help forecast the demand for charging services and optimise energy usage in the city.

The sections of this paper are organized as follows. Section II provides a brief overview of the existing analysis of the factors for electric vehicles and electric vehicle charging stations by refereeing some available data sets to predict their energy usage. The complete methodology for creating a synthetic dataset for electric vehicle charging sessions is briefly defined in Section III. The following section IV gives the exploration and analysis of the created synthetic dataset. The energy prediction is elaborated in section V for predicting, followed by understanding energy usage and its relationship with several other factors of electric vehicle charging stations in the created dataset. At last, the conclusion is derived.

II. Related work

Electric vehicles (EVs) are garnering popularity as a promising alternative to conventional fossil fuel-based vehicles due to their ability to reduce carbon emissions and reliance on nonrenewable energy sources. Increasing numbers of electric vehicle charging stations (EVCS) have been installed to accommodate the expanding EV market. The availability of a comprehensive dataset of EV charging sessions can facilitate the analysis of multiple factors and, ultimately, fulfil the demand for EV infrastructure. However, the lack of appropriate data sets and the energy consumption of EVCSs have been significant concerns as they contribute to the carbon footprint and operating costs of charging stations. Several works were devoted to electric vehicles, recharge stations, and their respective influences.

The article [1] provides an analysis of the electric vehicle (EV) market landscape in India and identifies opportunities for the adoption of EVs in the country. Specifically, the authors discuss the current market share of EVs in India, the policies and regulations governing EV adoption, and the challenges faced by the EV industry in India. Another article [2] analyzes user preferences related to electric vehicle (EV) charging decisions. The contribution of this article is to provide insights into user preferences for EV charging, pricing schemes, and charging infrastructure, which could help design policies and strategies for the adoption of EVs. The study in [3] examines the factors that influence the placement of EV charging stations and identifies the socio-economic and demographic characteristics of neighbourhoods under-served with EV charging infrastructure.

The authors in [4] highlight the importance of having accurate and reliable datasets to effectively design and deploy smart grid systems. They address the research gap of the lack of realistic datasets by proposing a novel dataset generation methodology that incorporates real-world data, such as vehicle travel data and charging infrastructure data, to generate a more representative and realistic dataset for smart grid systems with electric vehicles. Estimating EV charging demand that incorporates factors such as user behaviour, the spatial distribution of charging stations, and EV technology is provided in the article [5]. The authors use discrete choice models and numerical simulations to estimate the number of EV charging sessions and the energy required for charging during each session.

As mentioned above, the research gap addressed in these articles is the lack of proper datasets for EV charging sessions, making it challenging to accurately estimate charging infrastructure demand, energy utilization, and several other factors. It is noted that existing studies often rely on small or incomplete datasets, which may not accurately reflect EV charging behaviour.

Similarly, another study [6] assesses the influence of temperature and other trip characteristics on the energy consumption of electric vehicles. Electric vehicles may be refuelled at either public or private charging stations. Consequently, public charging facilities for electric vehicles face numerous challenges. There should be sufficient infrastructure for charging stations to meet the demand for electric vehicles. In [7], the profitability of charging stations influenced directly or indirectly by a list of factors is examined. Another article [8] calculates the electric energy consumption of charging stations by considering the charging duration and the number of charges.

However, the factors influencing the energy consumption of electric vehicle charging stations are rarely examined in academic literature. Along with the involved factors, it is necessary to comprehend how they affect energy consumption and how it is possible to regulate it by updating the relevant factors. And based on the interpretability of these variables, charging facilities for electric vehicles can optimise their energy consumption. Consequently, it is essential to understand various XAI techniques to explain this energy consumption. Shapley Additive exPlanations (SHAP), one of the XAI techniques, is a model-agnostic framework introduced by [9] designed to visualise explanations of ML algorithms. It is based on the concept of game theory. Using the Shapley value, the marginal contribution of each participant in the game is determined and explained. This explanation method is used to analyse the involved factors and assess their relationship to the energy consumption of charging stations for electric vehicles.

III. Synthetic Dataset creation methodology for EV charging sessions

Creating a synthetic dataset for electric vehicle charging sessions can have several benefits. This can be particularly useful for testing and optimizing charging infrastructure and developing load balancing and flexibility analysis strategies. Synthetic datasets can also solve the problem of the limited availability of real-world data for a new city. The use of synthetic data generators can provide a way to model EV charging sessions based on a large real-world dataset, allowing for the evaluation and optimization of different charging scenarios.

In this work, the dataset is created for the Berhampur city of Odisha state, where there is a significantly less number of electric vehicle movements and the number of EV charging stations is also very few, i.e. only four charging stations are available. The data set is generated for 2022, i.e. for all 365 days of the year. The historical population data for the city is collected from the govt. and other census site [10], [11]. Similarly, the number of EV sales data for the last 20 years is collected from the Ministry of Road Transport and Highway website [12]. The electric vehicle charging infrastructure available in India is collected from the government. sources [13]. The data regarding details of the different electric vehicles currently running in the Indian market are collected from their manufacturer. The LSTM model is known for capturing long-term dependencies and temporal patterns, making it suitable for predicting time-series data. Here in the case of population and EV sales, the next five years are predicted using the historical data collected from different sources, and it is shown in the following figures 1 and 2. In figure 2, the prediction is made for cumulative EV sales of the city, as all the previously purchased electric vehicles will be used for charging in the current scenario.

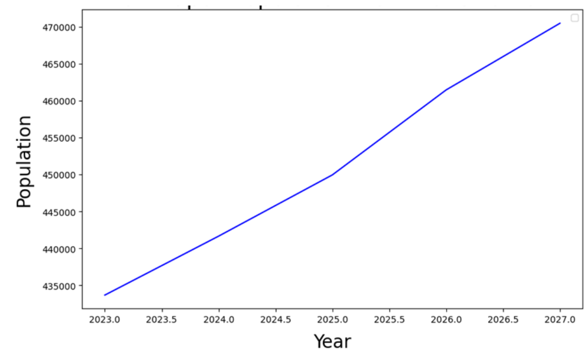

Fig. 1. Berhampur Population Prediction from 2023 to 2027

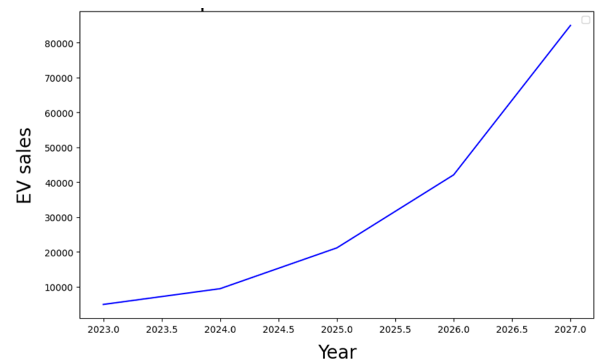

Fig. 2. Berhampur EV SALES (cumulative) Prediction from 2023 to 2027

A. Assumptions used in Dataset creation

After predicting the population and EV sales, the dataset will be used for day-wise growth for the following year. Every day, many electric vehicles (EVs) will search for electric vehicle charging stations (EVCSs) in the city. It is assumed that, out of the total available EVs, 10% are opting for EVCS in the city, as some EVs may choose to use the home chargers, and others may use chargers outside of the town. And all EVs may not need charging every day. This data set is generated for the whole of 2022. The initial population and EV sales data are used from the above sources, whereas the final data is the one-year predicted data that is to be used. As per the city's official record, only 4 EVCSs were available in 2022. By keeping the same number of EVCSs, it is assumed that at each EVCS, three different types of charges are used, where each charger is suitable for different types of EVs. The details of EV charging stations are given in the following table I.

As per the records of each EV, Level-3 is suitable for only 4-Wheeler, whereas level-2 and 1 are suitable for all three types of EVs, i.e. 4,3,2-Wheelers. It is also assumed that all the chargers are available 24*7 for charging. As each day, several vehicles are opting for charging, for each charging, a unique charging id is assigned. As per the vehicle sales data available on the Vahan website, out of the total EVs available, 16.3% are 2-wheelers, 74.4% are 3-wheelers, and the remaining are 4-wheelers. This rule is also followed in this dataset. An EV can arrive for charging at any time of the day. If the required charger is unavailable for charging, the EV must wait.

B. Process of Dataset creation

After collecting all the pre-requisite data from different sources and pre-processing them, they will be used for cre-

TABLE I
DETAILS OF EV CHARGING STATIONS

EVCS Name	charger type	Level	charger name	Charger_ID
EVCS-1	fast	3	CCS-2	101
EVCS-1	fast	2	Type-2	102
EVCS-1	slow	1	Bharat DC-001	103
EVCS-2	fast	3	CHAdeMO	201
EVCS-2	fast	2	Type-2	202
EVCS-2	slow	1	Bharat DC-001	203
EVCS-3	fast	3	CHAdeMO	301
EVCS-3	fast	2	Type-2	302
EVCS-3	slow	1	Bharat DC-001	303
TATA Charging Station	fast	3	CCS-2	401
TATA Charging Station	fast	2	Type-2	402
TATA Charging Station	slow	1	Bharat DC-001	403

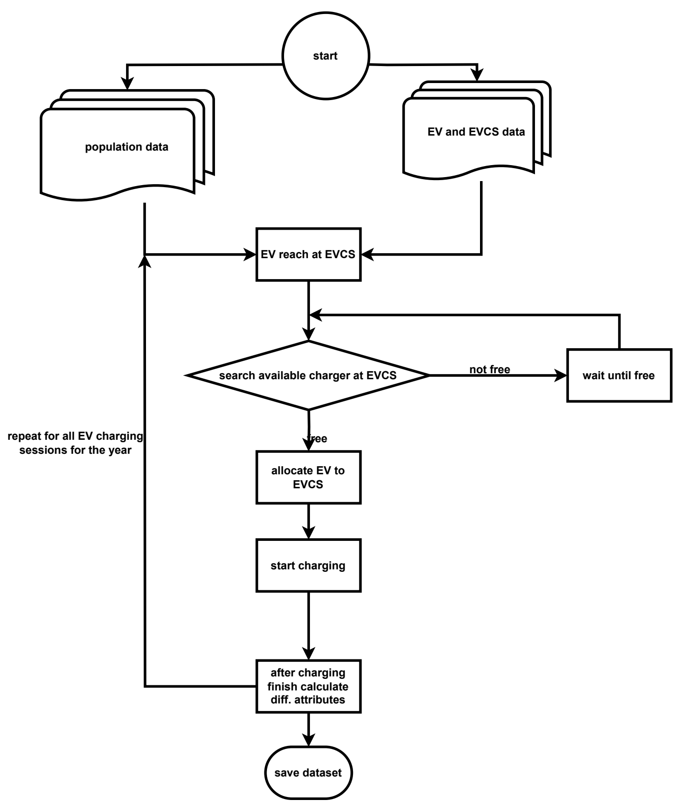

Fig. 3. process of the dataset creation

ating numerous charging sessions daily. The detailed dataset creation process is described in Figure 3.

When an EV reaches a charging station for each charging session, it first searches for the charger suitable for its EV. Then it checks whether the charger from the list of suitable chargers is free to charge or some other charger already occupied it. If it is free, then the EV start charging at that charger. If none of the suitable chargers is available, then wait until any charger is available. Any charger is available; the waited EV will first move to that charger to start charging. When the charging process is complete, it evaluates different attributes such as total charging reading, the total energy used, duration of charging, the cost to be paid for charging, total waiting time for charging, total time spent for charging, and many more. In this way, the data is generated for each charging session for a day, and the same process is repeated for all the days of the year.

Here several attributes are dependent on some other attributes. The waiting time is calculated from the time difference between the EV's arrival at EVCS and the charger assignment time. The total charging reading of an EV is calculated from the reading when EV start charging to the reading when EV stop charging. And based on this total charging reading, the time duration, energy utilized for charging and cost for charging are evaluated. The time an EV spends at a charger is calculated from the time difference between the EV's arrival at EVCS and the charging end time. Based on these attributes, several ways can be adopted to establish and enhance the EV infrastructure by making proper exploratory data analysis of such features.

IV. EXPLORING THE SYNTHETIC DATASET

It is possible to gain valuable insights into EV charging patterns and energy utilisation by generating a dataset that accurately describes assigning EVs to different charging stations. With this information, stakeholders in the EV industry, such as policymakers, charging station operators, and car manufacturers, can make informed decisions about charging infrastructure planning, energy management, and customer needs. Such synthetic datasets can conduct simulations and modelling exercises to explore different scenarios and potential outcomes. These simulations can help identify potential challenges and opportunities in the EV charging ecosystem and aid in designing efficient charging networks. Analysing synthetic datasets for EV charging sessions can provide valuable insights and aid in decision-making, planning, and optimization of the EV charging infrastructure.

As there are four different charging stations, an EV may arrive at any of these charging stations. From these created synthetic data, the average waiting time for each type of EV is shown in figure 4. The figure shows that the waiting time for a 4-wheeler is significantly shorter than for two and 3-wheelers; this is because the number of 4-wheelers is much less (only 10%) than other vehicles.

Similarly, if each type of charger is taken into consideration, then for chargers like CHAdeMO and CCS-2, the waiting time is significantly less than for other slow chargers because slow chargers take more time to charge. The vehicles only waiting for such chargers have to wait for their turn. This analysis is wholly reflected in figure 5.

Figure 6 shows the charge duration for each type of charger. It also shows that chargers like CHAdeMO and CCS-2 take less time to charge as they are fast chargers.

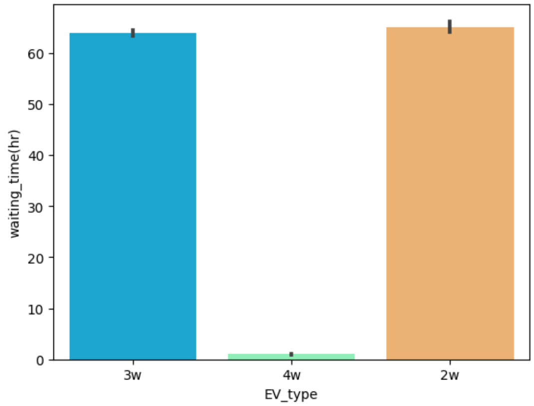

Fig. 4. Waiting time for each charging station

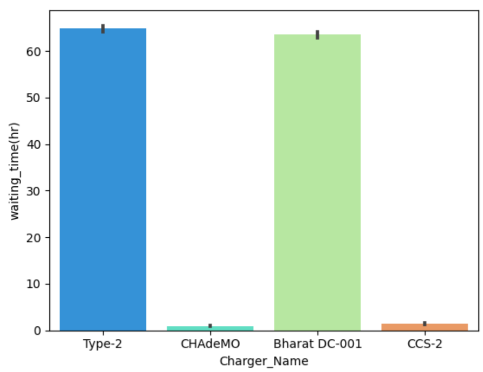

Fig. 5. Waiting time for each type of charger

Suppose the monthly wise average waiting time is used for analysis. In that case, it gradually increases the waiting time from the first month to the last, and this is because the number of EVs is increasing gradually. It is shown in the figure 7.

Similarly, several features can be used for analyzing data, and various conclusions regarding the behaviour of the charging sessions can be generated. Out of all the features, energy is one of the essential features for any EVCS infrastructure. The prediction of energy requirements can be made using the black box model. Still, it is also necessary to understand the model instead of interpreting the factors responsible for generating such predictions with the help of explainable AI (XAI) techniques. Based on this, stack holders can make several policies to strengthen the EV infrastructure.

V. Prediction of Energy utilization

A black-box machine learning model, such as the Random Forest model, is used to predict the energy consumption of each EV after learning from the synthetic dataset and analysing the relationship between several features and the total amount of energy consumed. Before prediction, the dataset is divided into test and training datasets with a ratio of 30:70. The Random Forest Regressor model is then implemented by configuring the model's hyperparameters. The predicted total energy values are then generated based on the test dataset. The

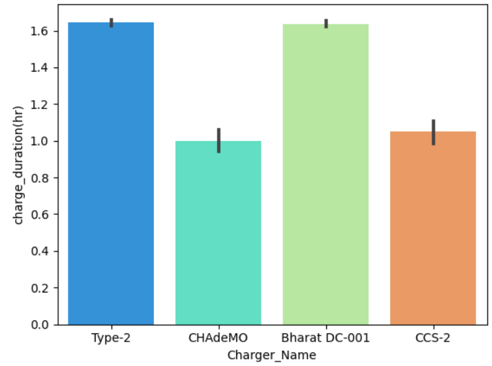

Fig. 6. charging duration for each type of charger

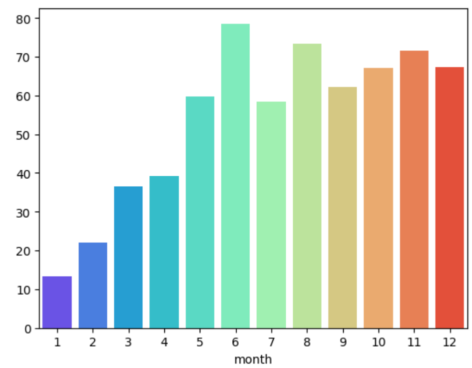

Fig. 7. monthly average waiting time

following figure 8 depicts the evaluation graph between actual and predicted values.

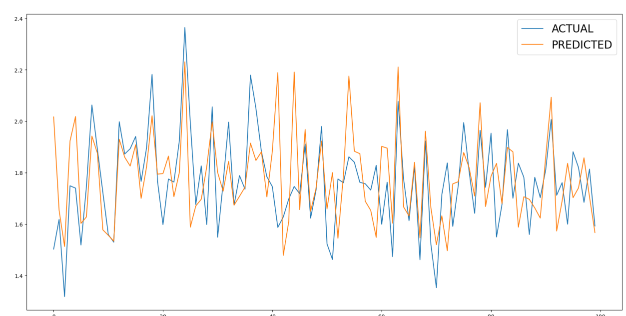

Fig. 8. Evaluation Graph between actual and predicted energy usage

for this prediction model, the evaluation matrix for different scaling parameters is as follows.
MSE : 0.030589808637766044
RMSE : 0.17489942434944158
R2 : 0.4244798876622954
Adjusted R2 : 0.38498340936460984
Accuracy: 92.12%

Along with the involved factors, it is necessary to comprehend how they affect energy consumption and how it is possible to regulate it by updating the relevant factors.

And based on the interpretability of these variables, charging facilities for electric vehicles can optimise their energy consumption. Consequently, it is essential to understand various XAI techniques to explain this energy consumption. Shapley Additive exPlanations (SHAP), one of the XAI techniques, is a model-agnostic framework introduced by [9] designed to visualise explanations of ML algorithms.

After importing the SHAP module, the TreeExplainer is added to the previous black-box model. SHAP values are generated using the explainer, but these values are challenging to understand. However, these values can be used to generate the force plot, which provides a straightforward explanation. The following figure 9 provides the force plot. It provides both local and global explanations. When a specific instance is highlighted, local explanations are provided. Taken as a whole, the force plot provides global explanations for these 50 instances from the test dataset. It provides the function variable, total energy consumed, on the y-axis.

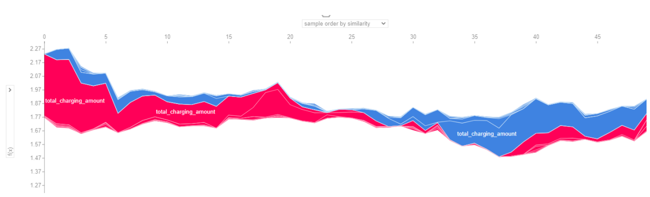

Fig. 9. Force plot for the Global Explanation

Here the feature value in blue forces the function value toward up, and the value in red forces the function value toward down. It can also be possible to select one particular feature to see its behaviour concerning functional value i.e. energy consumption. Here the total charging amount is considered in Figure 10. From this figure, it can be understood that as the charging amount increases from 25 to 29, the energy consumption also increases from 1.77 to 2.27, and decrease for the lesser value of the charging amount.

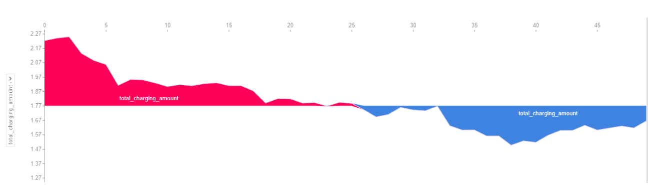

Fig. 10. Force plot concerning total charging amount

The summary plot for all instances of the test dataset is explained in Figure 11; it demonstrates that the total charging amount has the greatest impact on the total energy consumption at each charging station, followed by charging time. In contrast, the remainder of the characteristics have little to no impact on the functional value, i.e. energy consumption. Red is responsible for increasing the energy, whereas blue is responsible for decreasing it.

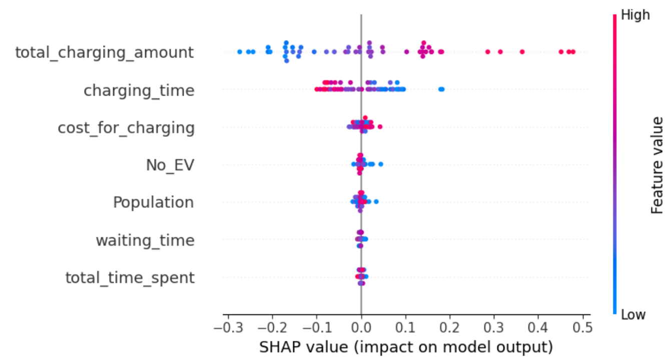

Fig. 11. Summary Plot

VI. Conclusion

The synthetic dataset for EV charging sessions is generated by considering all conceivable scenarios for a new city like Berhampur, Odisha, where very little EV infrastructure currently exists. The projected data indicates that electric vehicles will increase from 3,000 to 80,000 over the next five years. Consequently, all stack proprietors directly or indirectly impacted by the new establishment of EV infrastructure are interested in analysing this dataset type. From these datasets, they can analyse their impact on multiple factors. Here One of the most important factors is the energy consumption of each charging station. This predicts energy consumption with an accuracy of 92.12%. With the assistance of the SHAP explanation, it is possible to comprehend how numerous factors, including total charging amount, charging duration and charging cost, influence energy consumption. This information will assist policymakers, electric vehicle owners, and charging station administrators in making more informed decisions regarding energy consumption and constructing a more efficient charging infrastructure.

References

[1] Hema, R., and M. J. Venkatarangan. "Adoption of EV: Landscape of EV and opportunities for India." Measurement: Sensors 24 (2022): 100596.
[2] Visaria, Anant Atul, et al. "User preferences for EV charging, pricing schemes, and charging infrastructure." Transportation Research Part A: Policy and Practice 165 (2022): 120-143.
[3] Roy, Avipsa, and Mankin Law. "Examining spatial disparities in electric vehicle charging station placements using machine learning." Sustainable Cities and Society 83 (2022): 103978.
[4] Charalambidis, Georgios, et al. "A realistic dataset generator for smart grid ecosystems with electric vehicles." Proceedings of the Thirteenth ACM International Conference on Future Energy Systems. 2022.
[5] Sica, Lorenzo, and Francesco Deflorio. "Estimation of charging demand for electric vehicles by discrete choice models and numerical simulations: application to a case study in Turin." Green Energy and Intelligent Transportation (2023): 100069.
[6] Al-Wreikat, Yazan, Clara Serrano, and José Ricardo Sodré. "Effects of ambient temperature and trip characteristics on the energy consumption of an electric vehicle." Energy 238 (2022): 122028.
[7] Zhang, Qi, et al. "Factors influencing the economics of public charging infrastructures for EV–A review." Renewable and Sustainable Energy Reviews 94 (2018): 500-509.
[8] Cheon, Seongpil, and Suk-Ju Kang. "An electric power consumption analysis system for the installation of electric vehicle charging stations." Energies 10.10 (2017): 1534.

[9] Lundberg, Scott M., and Su-In Lee. ”A unified approach to interpreting model predictions.” Advances in neural information processing systems 30 (2017).
[10] https://www.berhampur.gov.in/demographic-feature/
[11] https://www.macrotrends.net/cities/21207/brahmapur/population
[12] https://vahan.parivahan.gov.in/
[13] https://www.smev.in/charging-infrastructure