Prepublished copy

Nature Environment & Pollution Technology

This is a peer-reviewed prepublished version of the paper

Original Research Paper

ConForMiSt: A Multi-model Dual-phase Framework utilizing Machine Learning for Carbon Footprint Prediction and Rein-forcement Learning for decision optimization

Sumukh R. Kashi[1], Yashas Surendra[1†], H.R Vineeth[1] and K.S Sowmya[1]

[1] Dept. of Information Science and Engineering, BMS College of Engineering, Bangalore, India

† Corresponding author: Yashas Surendra; yashassurendra.is20@bmsce.ac.in

ORCID IDs of Authors

Sumukh R. Kashi: *https://orcid.org/0009-0000-4580-617X*

Yashas Surendra: *https://orcid.org/0009-0005-5529-9737*

Vineeth H R: *https://orcid.org/0009-0001-7801-1549*

Key Words	Carbon Footprint, Emissions, Machine Learning, Reinforcement Learning, Q-learning, Touch Sensor, User Interface, Electrical Appliances, Transport Vehicles
DOI	https://doi.org/10.46488/NEPT.2025.v24i04.B4314 (DOI will be active only after the final publication of the paper)
Citation for the Paper	Sumukh R. Kashi, Yashas Surendra, H.R Vineeth and K.S Sowmya, 2025. ConForMiSt: A Multi-model Dual-phase Framework utilizing Machine Learning for Carbon Footprint Prediction and Rein-forcement Learning for decision optimization. *Nature Environment and Pollution Technology*, 24(4), p. B4314. https://doi.org/10.46488/NEPT.2025.v24i04.B4314

ABSTRACT

Over the past decade, there has been a significant surge in harmful waste emissions of greenhouse gases namely carbon dioxide, methane and fluorinated gases in the atmosphere. Two major categories of activities can be broadly identified which have contributed to this condition. The first is proliferation of world- wide industrial activity accounted by the industrial plants across all the major continents. Second is the human activity which also contributes to carbon emissions produced as a result of wide-ranging everyday activities that involve use of electricity, transportation, food consumption and other consumer-mindset driven activities. This article focuses on the second category to build a dual stage framework that will assess, evaluate, and recommend suitable mitigation measures to regulate usage patterns. The dual stage approach is a novelty based on sound engineering principles. Carbon emission data gathered by the system is analyzed to detect footprint generation patterns using mathematical models. Post-analysis, machine learning models selected from rigorous performance metrics (MAE, RMSE) are leveraged to make predictions of carbon footprint, in the first stage. The second stage employs a reinforcement learning framework that captures several aspects of emission in a 'state' and is used to analyze predictions and generate recommendations, considering user preferences. The ability to absorb user goal for emission data is a strength. This unique finer engineering of state representation exemplifies experimental data that shows minimal variation in state goal values within 2000 steps. A web application is developed to visualize various aspects like usage patterns, and predictions. The user interface provides interventional and specific and recommendations on a personalized level. These aspects are then utilized to provide insights at the aggregated level in the context of a group of individuals, which is yet another strength of the framework. The extensibility of the proposed methodology for carbon emission mitigation for higher aggregated levels is demonstrated by an exemplar 'location statistic' radar chart in the context of vehicle and electrical appliances categories.

INTRODUCTION

Over the past decade, there has been a meteoric rise in the number of industrial plants and the subsequent wastes expended by them. The pollution resulting from such plants can be mainly attributed to the harmful gases that are discharged by them. A significant increase in the presence of gases like carbon dioxide, methane nitrous oxides and fluorinated gases in the atmosphere is due to increasing air pollution (Ulfat and Noori 2024). These gases trap large quantities of heat constricting their circulation in the atmosphere. A direct consequence of this is increased carbon presence on a global scale.

Statistical data on greenhouse gas emissions in the Indian subcontinent provide valuable insights into various originators of carbon emissions across the country. India accounted for ~7.3% (3.9 billion metric tons (GtCO2) of the global greenhouse gas emissions in 2022 making it the third largest emitter after China (29.2%) and the United States of America (~11.2%) (Chandel 2022). G20 countries have been responsible for three-quarter of global warming as of 2022 (Jones et al. 2023).

The natural recovery process of the earth system to capture and store the CO2 will be exceeded unless an anthropogenic mitigation component is added to bring back the balance.

The objective of this article is to address the mitigation component by assessing the amount of carbon footprint contributions by human activities on a personal scale and appropriately provide behavioral modifications to mitigate the carbon footprint.

In the context of human activity, a personalized carbon footprint describes the impact of any given activity of a person or a group of people that results in the emission of greenhouse cases released into the atmosphere (Schwenkenbecher et al. 2014). It is an indicator of the amount of strain imposed on the environment. A personalized carbon footprint can result from a broad variety of activities ranging from electricity usage, personal vehicle usage for commuting, purchase of clothes, electronic products and various other activities like food consumption etc. Furthermore, buying food items and paper-based products add to the personal carbon footprint. A major contributor to carbon footprint in India is the domestic sector (Jain et al. 2021). Various sources of footprint include electricity consumption from home appliances like tube-lights, ceiling fans, air conditioners, washing machines. Additionally, fuel consumption from personal vehicles contributes to foot- print on a considerable level (Onat et al. 2015). The public transportation sector is yet another source of footprint emissions in a country. Footprint emissions are also contributed by livestock management and agricultural sectors (Ramachandra et al. 2015). Detailed assessments of life- cycles of carbon emitting substances provide a good understanding on their overall environmental impact. A recent study of various components of HCF (Household Carbon Footprint) of India is provided in (Huang et al. 2023). According to this study, 39% is contributed by energy consumption, 20% by travel, and 14% by food. Therefore, if these contributory components can be reduced, it will significantly improve the HCF.

Increased emissions of greenhouse gases result in increased atmospheric temperatures (Khan et al. 2024), altered weather patterns causing ice-caps to melt, rising sea-levels and potentially catastrophic events like hurricanes, floods and droughts (Kiehbadroudinezhad et al. 2024). Another notable consequence of increased carbon emissions is the rise of respiratory problems, cardiovascular diseases and premature death of flora and fauna. This can be mainly attributed to pollutants like Nitrogen oxides, Sulphur dioxide and particulate matter (PM) (Kumar et al. 2023).

Although the major contribution to the global carbon emissions problem stems from macro-industrial activities, it is important and cogent to address carbon emission contributions and problems from individual entities arising out of their activities and behaviors.

Section 1 provides an introduction to the problem of increasing carbon footprints from human activities and the need to appropriately modulate and control them to mitigate the adverse effects. Section 2 generates the problem statement and solution strategy for the work described in this article. Section 3 examines important work in this area and segregates literature in this area by various activities. Section 4 develops the system design and architecture for the solution conceptualized along with a description of a prototypical system that implements the ideas to showcase a working system. Section 5 discusses experiments and data gathered from experiments performed with this prototypical system with an analysis of the results obtained. Section 6 makes concluding remarks and provides pointers about how this work could be applied and taken to the next level of research and production.

2. MATERIALS AND METHODS

2.1 Problem Statement and solution strategy

The primary objective of this paper is to analyze carbon footprint generation patterns from various carbon-emitting devices and provide a mitigation solution (Goswami et al. 2024).

2.1.1. Summarized Problem statement

"Collect carbon footprint data from various categories of carbon emitters. Analyze the data and predict future emissions for the various activities in these categories. Using this information, provide recommendations to reduce carbon footprint emissions from various activities"

2.1.2. A Workable Solution Strategy

Our solution framework is ConForMiSt, an acronym for Carbon Footprint Mitigation System and shall consist of the following elements:

1. Collect raw consumption data from various categories of carbon emitters
2. Analyze carbon footprint generation patterns
3. Conceptualize machine learning models that will capture essential aspects of these patterns to formulate carbon footprint estimation and mitigation model
4. Use reinforcement learning models to provide recommendation to reduce carbon from various human activities – A framework
5. Provide a web-based and cloud-centric architecture for scalability and extensibility, so that this application may be used in a macro level setting like an institution or organization

2.2 Related Work

We now describe literature that provides description of carbon footprint measurement and the attendant mitigation mechanisms. In order to perform this study, we have captured and categorized the work into various categories driven by human activities that lead to generating the carbon footprint. This prior art study provides us a starting point to mitigate carbon footprint.

2.2.1. Mobile Phone Usage

A study that describes the usage of devices that have networking and communication capabilities has been succinctly described in (Lövehagen et al. 2023). Experiments were performed on smartphones, feature phones, tablets, laptops and personal computers in this study, revealing varied emissions of carbon on any given day. The equivalent carbon footprint (CO_2e) phone use through the factor 57 g Co_2 e per minute use. Study findings revealed that the respondents spent the most in texting with an average of 17.95 kg CO_2e per day. A general recommendation on reducing electronic usage was also described in this study.

2.2.2. Forest Activity

A categorization of the carbon emissions produced from different types of forests namely- primary forests, naturally regenerated forests and planted forests has been provided in (Mancini et al. 2016). The effect of forest wildfires and harvested wood product have been factored into the calculated of carbon footprint. A detailed account on forest carbon sequestration as a contributing factor to carbon footprint has been provided.

2.2.3. Organizational Activities

The concept of operational boundaries that define physical and geographical limits of an organization's activities for the purpose of measuring and reporting its greenhouse gas emissions have been detailed in (Gao et al. 2014). A study on carbon emissions generated in the corporate sector has been assessed by leveraging several machine learning models in (Musa et al. 2024).

2.2.4. Food Consumption and Behaviour

Analysis of changes in consumer behaviour in the context of food products that contain labels detailing carbon footprint output of food processing has been succented in (Rondoni et al. 2021). Labels enable food manufacturers to indicate the impact that their food production process has on the environment. Consequently, this allows consumers to make informed choice while purchasing processed food products. The findings of this work reflect the positive attitude of countries like Egypt and China towards carbon footprint information urging other emerging countries to develop a similar outlook on processed food products.

2.2.5. Approaches and Frameworks for estimation of carbon footprint

An approach revolving around statistical modelling has been employed in order to predict carbon footprints from a corporate standpoint for climate finance risk analysis (Nguyen et al. 2021). Models like linear regression, k-nearest neighbours, and decision tree ensembles have been leveraged for carrying out predictions. Model performance has been evaluated using mean absolute error (MAE) and multiple resampling techniques. A framework that targets systemic reporting, energy promotion and carbon efficiency using machine learning techniques has been described in (Henderson et al. 2020). The use of green defaults and effective component integration to achieve pro-environmental behaviour has been elaborated. An empirical framework comprising of elastic network regression models and machine learning models to arrive at effective carbon footprint mitigation strategies has been provided in (Dong et al. 2023). Energy consumption dynamics have been examined and correlation between carbon emissions and its causation factors have been analysed in order to provide suggestions for carbon emission mitigation measures (Liu et al. 2023).

The comparison of ConForMist framework with existing works has been provided in Table 1 to highlight the key features and differences of the proposed system.

Table 1: Comparison report of ConForMist with existing literature frameworks

Research Paper	Methodology	Key Results	Comparison
Lissa et al. (2021)	Deep Reinforcement Learning	Unique comfort factor v/s energy savings metric.	User's preferences are taken as an input and use of ML models with historical data to provide statistical data.

(Lee and Choi, 2019)	Deep Q-Network, Deep Learning	Forecast of indoor temperatures and ToU prices lead to real time decision making	Different models for different appliances hence providing more granularity.
(Mahmoud and Ben Slama, 2025)	Deep Learning, Reinforcement Learning	Vehicle to Home bilateral communication	Usage of carbon emission data over energy units to focus on home emission data.
(Solatidehkordi et al., 2023)	LSTM, IOT	Real time classification of appliance data usage	Microcontroller with timestamp recording, UI Dashboard provides yearly, location-based data and emission data statistics

2.2.6. Research gap in existing implementations of home energy management system (HEMS)

The existing frameworks in the HEM systems primarily focus on energy savings on electricity and money. The research gap being addressed in ConForMist is the carbon emission focus centric approach in homes. The proposed system will help the stakeholders to track and achieve reduced carbon emission goals. There is a need for developing sustainability solutions which can be easily integrated to existing architectures. The architecture functions without considering smart meters, smart appliances and smart grids and hence can be deployed faster, cheaper and provides a scalable solution in lesser developed electrical grids and homes.

We use carbon emission data directly to train our regression models. The outputs of carbon emissions are offset by a percentage (that is determined by the end-user) and fed as input to the reinforcement model as goal states. The novelty of the reinforcement model lies in its learning that is based on the concept of having the four phases in a day along with optimization on the usage of devices. Another research gap being addressed is the user flexibility on the input to their solutions and trade-offs on comfort versus energy savings.

2.3 Top Level Systems Approach System Design And Implementation

The first stage handles tasks ranging from collection of user consumption data to its cleaning and pre-processing. As a starting point, six different carbon emitting sources have been considered for measurement. Data related to carbon emission resulting from human activities are collected. These sources have been chosen to demonstrate proof of concept considering their consumption priority over low-usage devices in the Indian subcontinent. Two broad categories of carbon emission are considered (Transport and Electricity) to demonstrate the concept of operations and this can be easily extended to other categories as well. The transport component comprises of Four-wheeler and Two-wheeler vehicles which run on two types of fuel sources: petrol and diesel. Usage patterns for vehicles running on both types of fuel have been considered during the dataset formulation process. Household devices investigated include tube-light, ceiling fan, washing machine, and air-conditioner.

2.3.1 Data Collection and Pre-Processing

The top portion (a) of Figure 1 identifies the data collection and pre-processing step. The various sources of carbon emission across various individuals in each of these two broad categories of carbon emitters are collected, aggregated, and analyzed. The carbon emission factors of these appliances are used to compute their carbon footprint output for a specific duration of usage. The devices exhibit complete dependency on their corresponding carbon emission factors. The method of data collection involves recording the device readings during 4 phases of each day over a course of 12 months with the following apportionment. This is indicated in Table 2.

Table 2: Apportionment of Phases of a Day for Data Collection

Phase of Day	Time (24-hour Clock Format)
Morning	00:00 - 05:59
Afternoon	06:00 - 11:59
Evening	12:00 - 17:59
Night	18:00 - 23:59

Usage (in hours) is recorded during each phase of the day and the carbon footprint is computed. The usage-footprint pair values are stored in a dataset (CSV file) for further cleaning and processing. Subsequently, the values in the dataset are checked for their logical correctness in order to clean the data set from inaccurate measurements recorded by the sensors. If a user-input value exceeds the range provided in Table 2, it is perceived as an error input and subsequently discarded.

2.3.2 Initial Dataset Preparation

The left bottom portion (b) of Figure 1 identifies the data preparation step. In the second stage, the cleaned dataset is fed to machine learning models for training in the next stage of the implementation. The selection of a particular machine learning model for predicting carbon footprint output of a given device has been made after examining the input-output variable relationships. A linear regression model is trained for input-output data columns that exhibit a linear relationship (Maulud et al. 2020). Random Forest regression (Ali et al. 2012) and Gradient Boosting models have been trained for non-linear data points (Otchere et al. 2022) and the choice between these two models have been made by drawing a comparison between their mean absolute errors on training data.

2.3.3 Prediction of carbon footprint from activities monitored

Part (c) of Figure 1 identifies the prediction step. The third and final stage of the implementation involves generation of personalized recommendation reports based on device usage history of the individual. A reinforcement learning model that is customized for each type of vehicle or electrical appliance drives the recommendation generation engine. Additionally, data insights like average and total usage of appliances, major contributor by location of the users at an aggregated level are generated by leveraging NumPy, Pandas and Seaborn libraries. The stakeholders of the aggregated data insights are government organizations whose objective is to cut down carbon footprint emissions of areas under their jurisdiction by levying taxes and fines.

2.3.4 Generation of Recommendation, Strategies and on Dashboard

Part (d) of Figure 1 identifies the recommendation step. A Web-based tool with role-based access is developed to enable analysis, interpretation, and key strategies/actions to take. A key driver for the user interface is to enable individuals to access personalized recommendation reports and even larger entities like government organizations to access aggregated data insights from different regions of the state and city.

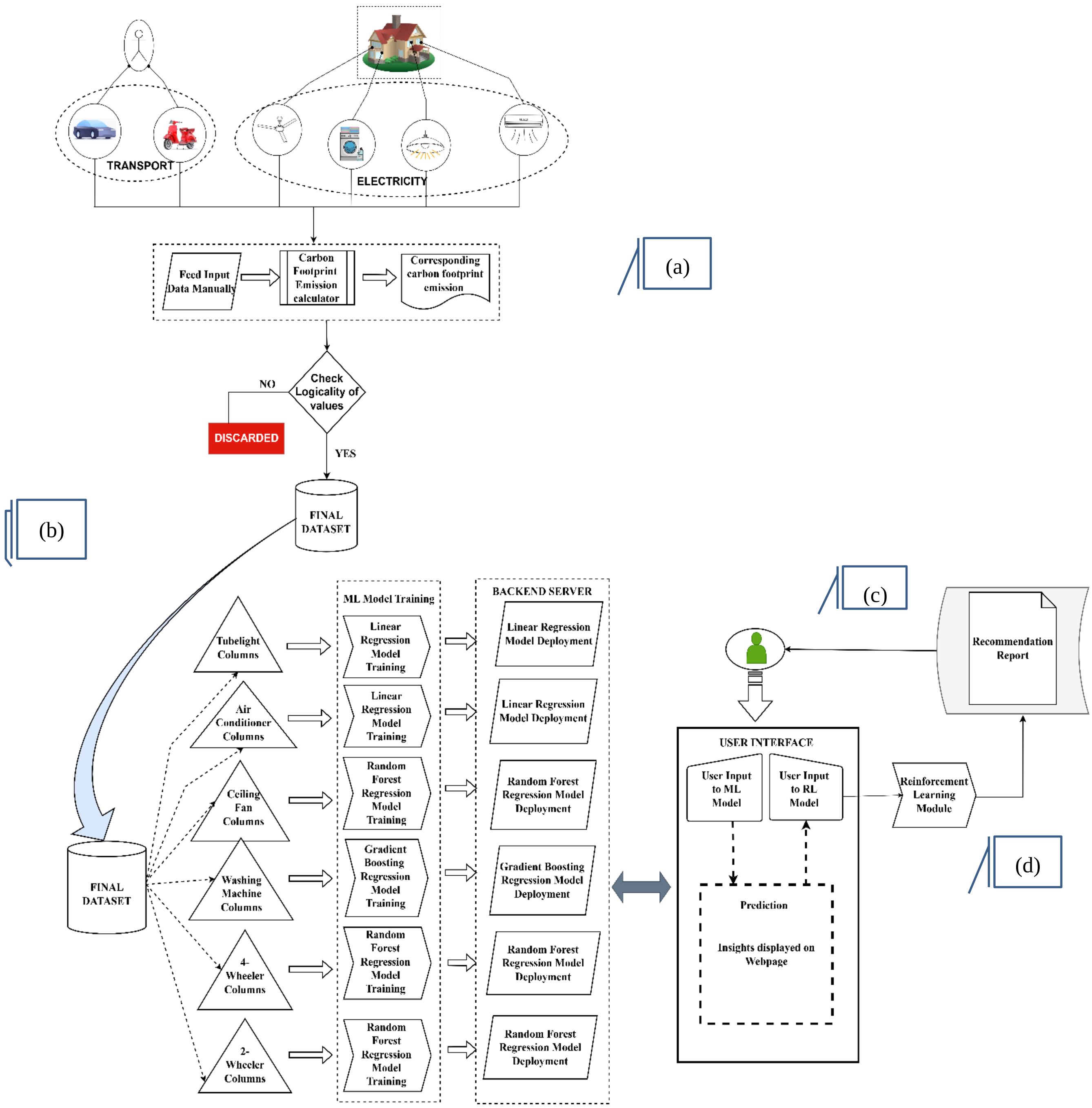

Fig. 1: A Bird's Eye View of ConForMiSt Process

2.4 System Design And Implementation

2.4.1 Engagement of Machine Learning Models

The structured dataset offers a rich training dataset for machine learning models. In the case of electrical appliance activity category, analyzing the correlations between time-dependent appliance usage, the models can learn to estimate carbon footprint with increasing accuracy. Similar is the case with the transport activity category which involves analyzing the correlation between vehicle travel patterns, fuel types, and other characteristics. This is a data-driven approach and focusses on capturing diurnal variations in user behavior. This approach is free of reliance on manual data input and paves the way for an automated and efficient system for carbon footprint assessment.

Several machine learning models are trained using the dataset that is produced following preprocessing. The approach begins by feeding the structured data of a certain appliance or vehicle to a machine learning model. The selection of a machine learning model for a certain appliance or vehicle is determined by empirical findings obtained from experimentation from a set of investigative trials (Varoquaux et al. 2023). The models were chosen after comparing each one of them individually based on mean absolute error. Machine Learning models which were considered for extracting empirical evidences were Linear Regression, Random Forest Regression and Gradient Boost Regression.

The dataset for feeding these models was divided into training and testing sets using various trial counts (Ranjan et al. 2019). Four training trials, each with a distinct training and test split of the data set, were used to train the machine learning models. In the initial experiment, 30% of the data was used for testing and 70% for training. The second trial used a 80% - 20% split for testing and training. The third trial used a 90% - 10% dataset split for the same while the fourth trial had a 75% - 25% spilt for the test-train dataset.

A grid-search approach to identify the right combination of hyper-parameters has been employed to methodically build and evaluate a model for each combination of algorithm parameters specified in a grid (Bhatti et al. 2000).

Table 3: Experimental ML Models for Third Stage (Prediction)

Design Rationale	**Carbon emitter Category/ Human Activity/**	**ML Model chosen for Analysis**	**Perf. Metric**	**Metric Value**	**Metric Value - RMSE**
Relationship between input-output variables	Transport (2Wheeler)	Linear Regression	MAE	19.566	24.53
To fit non-linear data points	Transport (2Wheeler)	**Random Forest Regression**	MAE	**2. 499**	**3.13**
To reduce mean absolute error	Transport (2Wheeler)	Gradient Boosting Regression	MAE	2. 614	3.28
Relationship between input-output variables	Transport (Four-Wheeler)	Linear Regression	MAE	29.077	36.45
To fit non-linear data points and	Transport (Four-Wheeler)	**Random Forest Regression**	MAE	**5.349**	**6.7**
To reduce mean absolute error	Transport (Four-Wheeler)	Gradient Boosting Regression	MAE	4.735	5.94
Relationship between input-output variables	Electricity (Tube Lights)	**Linear Regression**	MAE	**0.218**	**0.27**
To fit non-linear data points	Electricity (Tube Lights)	Random Forest Regression	MAE	2.757	3.46

To reduce mean absolute error	Electricity (Tube Lights)	Gradient Boosting Regression	MAE	8.583	10.76
Relationship between input-output variables	Electricity (Air conditioner)	**Linear Regression**	MAE	**0.175**	**0.22**
To fit non-linear data points	Electricity (Air conditioner)	Random Forest Regression	MAE	0.879	1.1
To reduce mean absolute error	Electricity (Air conditioner)	Gradient Boosting Regression	MAE	1.901	2.38
Relationship between input-output variables	Electricity (Ceiling Fan)	Linear Regression	MAE	9.163	11.48
To fit non-linear data points	Electricity (Ceiling Fan)	**Random Forest Regression**	MAE	**0.200**	**0.25**
To reduce mean absolute error	Electricity (Ceiling Fan)	Gradient Boosting Regression	MAE	0.631	0.79
Relationship between input-output variables	Electricity (Washing Machine)	Linear Regression	MAE	3.304	4.14
To fit non-linear data points	Electricity (Washing Machine)	Random Forest Regression	MAE	0.336	0.42
To reduce mean absolute error	Electricity (Washing Machine)	**Gradient Boosting Regression**	MAE	**0.189**	**0.24**

Comparing the metrics of the various models for the activities of transport and electricity usage from Table 3, the conclusions are as follows: Linear Regression model was found to give accurate results for data recorded by Tube light and Air Conditioner for the Electricity usage category. Similarly Random Forest Regression model was found to give accurate results for inputs provided by Ceiling Fan for the Electricity usage category wile the same model also was the model of choice for Two-Wheeler and Four-Wheeler usage in the transport category. Finally, the Gradient Boosting Regression model gave accurate results for inputs from Washing machine.

Output of the trained models:

The aforementioned machine learning models predict emission from the carbon emitter categories from usage input usage data.

Post Processing:

These trained models are serialized into a pickle file and later deployed on to a web server. Pickle files are generated by employing the technique of serialization. Serialization is the process of decomposition of complex data structures into a sequence of primitive data parts, which can be saved directly in a file or transferred over a network (Bellman 1957). On the receiver end, the pickled files are deserialized and fed with inputs from users to obtain prediction of the amount of carbon emission with respect to the amount of usage in hours or kilometers. The obtained prediction value is displayed to the user by the help of a responsive user interface. The user is offered a choice to offset their prediction by an amount of 5%, 10% or 15% respectively. The reduced value of carbon emission represents the users preferred amount of emission or a target which has to be achieved in a specific duration. The offset value along with preferences containing the amount of usage expected in all four time slots of a day, namely - morning, afternoon, evening and night are given as input to the corresponding reinforcement learning model deployed on the web server and tuned to that particular appliance takes the goal consumption input, preferences in a day and the fuel type in case of vehicles and gives personalized recommendation of the amount of usage in the hour-minute format in case of electrical appliances and Kilometer format in case of vehicles to achieve the goal consumption.

2.4.2 Engagement of Reinforcement Learning Models

The predicted carbon emissions from the usage of carbon emitters (devices in the various categories) are used as the base data for further processing after user interactions.

The user interaction involves the following steps:

- Present a summary of the carbon emission from the individual’s 12 month recorded data
- Provide the user three different reduction level choices.
- Accept the goal quantity for reduction in terms of percentage over the base data.
- This target is used to process the information through a reinforcement learning model to work out a suitable strategy for reducing the user carbon footprint. The reduced carbon emission from the offset value will be the goal state of the reinforcement model.

The reinforcement learning model is mechanized as follows:

Each state of the model is defined as a tuple indicative of:

- Phase of the day
- Device Operational Parameter
- Energy Consumption Level

An action set consisting of action values represented as an enumerative list:

Action_1Action2, ……

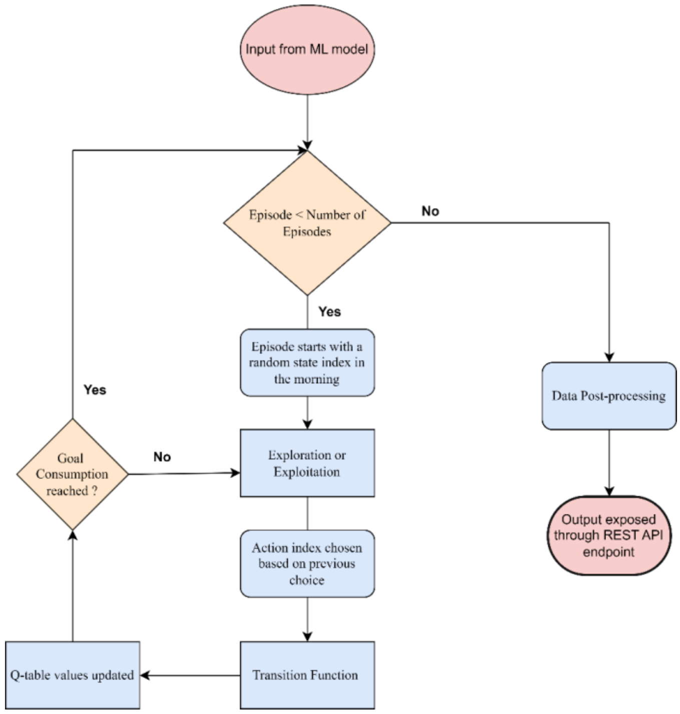

Fig. 2: Phase 2 Agent Mechanization using RL

A value function(q-value) that evaluates the value of a state and provides a value to the best state that generates the lowest consumption value. The reward value generation is driven by the simple principle of providing rewards for high carbon generation (Eg: higher temperature setting in case of an air conditioner).

The underlying principle of using reinforcement learning is to optimize and simulate the usage of the devices during a 24-hour period based on a percentage reduction from the previous machine learning algorithm prediction values. The output given to the user will be the apportionment of optimal usage of the devices among different phases of the day. The agent is given a goal consumption value as the input and learns until the consumption for a episode reaches the goal value. The agent mechanization is provided in Figure 2.

The reinforcement learning algorithm is devised for six different models corresponding to tube light, fan, washing machine, air conditioner, two-wheeler, and four-wheeler vehicles. Each of them has the same underlying principle which consists of a Markov decision process (Watkins et al. 1992) with a policy of q-learning (Babiuch et al. 2020). The idea behind considering the states in this process stems from the dataset structure which is the division of the 24 hours into 4 phases, namely - Morning, Afternoon, Evening and Night. Any action taken by the devices considered, leads to a result in a change of the carbon emission output from that corresponding device. The relation between the states and the actions must be established and defined in a manner so as to construct the q-table.

a. Actions

The actions define interaction of a user who makes use of the different functionalities of the devices under consideration. These actions are independent actions, i.e. one action has no effect on how the other action is performed. A brief outline of the actions for various device categories is provided below:

Light: [“Off”, “Low Brightness”, “Medium Brightness”, “High Brightness”]

These actions above have considered modern, present-day LED’s which have controls to change their brightness using a regulator component.

Fan: [“Off”, “Low Speed”, “Medium Speed”, “High Speed”]

The different speeds of the fan which can be set by rotating the dial are mapped to the actions above.

Air Conditioner: [“Off”, “Moderate Cooling”, “Cold Cooling”, “Coldest Cooling”, “Eco Mode”, “Turbo Mode”]

The actions for an air conditioner consider the choices that an appliance user generally selects: Ranging from a completely switched off state, degree of cooling (3 levels) to specific modes (Eco, Turbo). The Eco Mode refers to ideal cooling which will lead to a reduced consumption but achieve the required temperature at a slower pace and slower fan speed. Turbo Mode is quite the opposite, having higher consumption while reaching the required temperature faster along with a faster fan speed.

Washing Machine: [“Off”, “Quick Wash”, “Normal Wash”, “Heavy Duty”]

Actions mapped for a Washing Machine are those that represent general modes of the different types of wash in a general washing machine. If gradation of these modes is considered, the order of heavy, normal and quick wash indicates the descending order of consumption.

Transport: [“Off”, “Accelerate”, “Cruising”, “Speeding”, “Engine Braking”]

The actions for transport category device are indicated above. Although the actions are intuitive, we differentiate between accelerate and speeding as follows: accelerate action is a mediated increase in speed while the speeding action is going at a rapid pace at higher gear. Cruising and Engine Braking are ideal techniques which reduce the fuel consumption and emissions.

b. States

We need an accurate model of the state description. Therefore, a state is represented as a tuple consisting of ({phase of the day}, {Device Operational Parameter}, {Energy Consumption Level}).The state modelling is driven by three primary factors that determine a state: (a) The phase of the day influences the consumption profile (b) The device under consideration may operate in various modes (Eg: A vehicle operating in low and high gears, an air conditioner operating in eco mode or moderate-cooling or turbo cooling mode) (c) consumption level. The device operating parameter is then amenable for directly mapping user choice to the operating mode of the device. The Zero factor is equivalent to the OFF state and the device is not emitting footprint as provided in Table 5. The correlation between the device factor and the consumption level maybe either positive and negative depending on the action taken.

The various states are defined using Table 4, by choosing a discrete value from each column of the device.

Table 4: State "Tuple" Components for defining State Representation

Phase of the day	Device	Device Operational Parameter	Consumption
Morning, Afternoon, Evening, Night	Light	Zero, Low, Medium and High Brightness	High, Moderate, Low
	Fan	Zero, Low, Medium and High Speed	
	Air Conditioner	Zero, Low, Moderate and Coldest Temperature	
	Washing Machine	Zero, Quick, Normal and Heavy Wash	
	Transport	Zero, 1 ,3 and 5 gears	

c. Transition Functions

The recommender system uses Q-Learning category of reinforcement learning and therefore always chooses an action based on the Q-value state at a state, according to the equation (1).

$$TD(s_t, a_t) = r_t + \gamma \cdot \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \qquad \ldots(1)$$

where,

s_t: Current state at time step t

a_t: Action taken at time step t

r_t: Reward received after taking action a_t in state s_t

γ: Discount factor ($0 \leq \gamma \leq 1$) that determines the importance of future rewards

$Q(s_{t+1},a)$: Estimated future reward for taking action a in the next state s_{t+1}

$\max_a Q(s_{t+1}, a)$: Maximum Q-value over all possible actions in the next state s_{t+1}

$TD(s_t, a_t)$: Temporal Difference (TD) error, which measures the difference between the current estimate and the updated estimate of the Q-value.

A minor augmentation to this transition (which is based on Q-value) is the check for an "End State" at the end of every action. This is because the user is provided a choice to select the offset for total carbon footprint reduction and this is translated to a "maximum consumption limit" check. If this limit is reached in any state, then the learning process is stopped. An exemplar encoding of these end states for various device categories and transport vehicle category is provided in Table 5.

The transitions are based on a weight factor, the inputs to which are provided by the user. The structure of the weights is represented as [a, b, c, d] which represent the weights of usage for Morning, Afternoon, Evening and Night. Based on these weights, the reinforcement learning agent transitions to different phases. For example, if a user decides that the need for an LED would be higher in the evening and the night, he could increase the relative weightage of these two phases. A random number generator generates the next phase index based on the relative weights. The second part of the transition function is the action-transition function. This function defines the logic for device factor variation based on the action. The transitions vary based on the device.

The "Engine Braking" state can transition to three possible states making it a stochastic process. While considering the braking process, the vehicle will slow down to a lower gear. Hence the three states are justified. Another factor considered for the transport model is traffic probability. The higher the factor, the more probability of the agent transitioning to the state of the first gear. Under higher traffic conditions, the cars are slower and would be in the lower gears i.e. either the first or the second gear. The second gear state is not being considered in our model for simplicity and to avoid redundancy.

Another key factor to be considered will be the agent learning beyond a max carbon consumption. Hence a Max Consumption Limit has been taken into account.

$$\text{MaxConsmpCheck} = 6*(\text{device consmp factor kgCo2/hour}) \quad \ldots(2)$$

Table 5: End state encodings for devices and vehicles

Device: Light	
Action	Next State
"Off"	(Zero, Zero)
"Low Brightness"	(Low Brightness, Low Consumption)
"Medium Brightness"	(Medium Brightness, Medium Consumption)
"High Brightness"	(High Brightness, High Consumption)

Device: FAN	
Action	Next State
"Off"	(Zero, Zero)
"Low Speed"	(Low Speed, Low Consumption)
"Medium Speed"	(Medium Speed, Medium Consumption)
"High Speed"	(High Speed, High Consumption)
Device: Air Conditioner	
Action	End State
"Off"	(Zero, Zero)
"Moderate Cooling"	(Moderate Temperature, Low Consumption)
"Cold Cooling"	(Low Temperature, Moderate Consumption)
"Coldest Cooling"	(Coldest Temperature, High Consumption)
Device: Washing Machine	
Action	End State
"Off"	(Zero, Zero)
"Quick Wash"	(Quick Wash, Low Consumption)
"Normal Wash"	(Normal Wash, Moderate Consumption)
"Heavy Duty"	(Heavy Duty, High Consumption)
Transport: VEHICLE	
Action	End State
"Off"	(Zero, Zero)
"Accelerate"	(3, Moderate)
"Cruising"	(5, Low)
"Speeding"	(5, High)

Algorithm 1: Recommendation Algorithm

```
Require:
− reward_function (state)
− action_function (choice, action index)
− transition (cons, action, weights, max_consump_check)
− num_episodes
− epsilon
− Q_table
− actions
− states
− learning_rate
− discount_factor
− goal_consumption

Ensure:
− Q_table update
− Iteration for num_episodes
Begin
for each episode in range (num_episodes):
    if max_consump_check is true:
        Print "Max" and Break

    Initialize weights: [a, b, c, d]
    Initialize total_consumption: 0
    Choose random state index: state_index
    Initialize cons_values for times of day

    while true:
        if random number < epsilon:
            Choose random action_index
        else:
            Choose action_index with highest Q−value
        Simulate environment:
            − Get reward
            − Compute next_state_index, max_consump_check
            − Update Q−value in Q_table

        Update total_consumption
        Update consumption values
        Update state_index
        if total consumption >= goal_consumption:
            Break
End
```

2.4.3 System Implementation – Proof of Concept

The second category of carbon emitters in Table 3, 'Electricity' consumption driven by Tube/LED Light bulbs is chosen as a representative example for experimentation and demonstration of our solution. The experimental setup shown in Fig. 3 provides a test bed to collect usage data of individuals using the light appliance. The hardware circuit architecture consists of an ESP-32 WROOM 32D micro-controller, a TTP-223 touch sensor, a 10KΩ resistor and a LED bulb. The pin configurations for the micro-controller are provided in Fig. 3 to provide clarity of the interfaces. The choice of the ESP32 module is driven by the Wi-Fi connectivity feature, which eases integration with a web server.

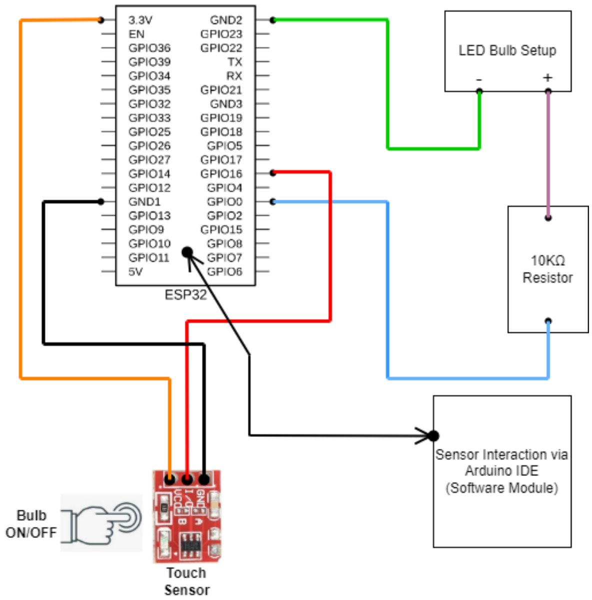

Fig. 3: Experimental setup for Lighting Consumption Data Gathering

a. Experimental setup of data gathering

The touch sensor is placed on the switch which is used to control the appliance. The touch sensor interfaced to the ESP32 module is used to record the electricity consumption by the individual. The implementation (code) that supports the touch sensor functionality has been configured as follows: the bulb is initially in an OFF state. On pressing the sensor (stimuli), the bulb transitions to ON state and continues to be in the same state until the sensor is pressed again. This subsequent action results in a transition of the bulb to OFF state. The timestamp at which the bulb transitions to the ON and OFF stats are recorded (Arduino IDE interface) from which the duration for which the bulb is utilized can be easily computed.

b. Data Collection and Processing from Sensor equipment

Fig. 4 provides an overview of the data processing pipeline showing how the information is gathered from devices with the help of touch sensors and a micro-controller (ESP32) (Hiraman 2018). The recorded data has the device usage data in milliseconds. The usage data obtained from the sensors is stored in a text file, with each device usage interval duration (each on-off condition) stored on a new line.

This usage device usage data is sent to an Apache Kafka platform for further processing. Apache Kafka is a platform for processing real time streaming data, using distributed publisher-subscriber messaging system. The significant advantage is its ability to handle large volumes of data (concurrent requests from users) and reduce latency in providing predictions to the users (Thein 2014). As noted earlier the device usage data is recorded with files and these files need to be uniquely identified. The identification process is enabled by the file's unique hash value called a message digest. This is created using a hashing algorithm like MD5, which is of fixed length and uniquely identifies the contents of the file.

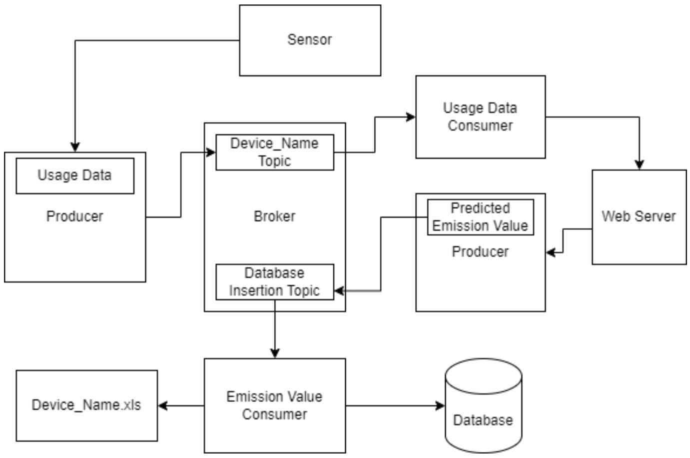

Fig. 4: Data Processing Pipeline in ConForMiSt

An update operation to the file will replace the pre-existing message digest with a new hash value, which now uniquely identifies the modified content of the file. A Python script, which indefinitely monitors the state of the message digests, triggers an event when the file is modified with the new sensor reading. The latest data obtained is passed to the Kafka broker with the corresponding device name (Eg: "Tube-light") as the topic through a Kafka Producer via a message after serialization. Another Python script indefinitely running the Kafka Consumer fetches the message from the broker and deserializes it, in order to obtain the usage data. Producers publish messages to Kafka topics, and consumers subscribe to these topics and consume the messages (Bahrami et al. 2018).

The obtained usage data is sent to a web server running the trained Machine Learning models using a HTTP request to the API endpoints exposed by the web server (Raptis et al. 2022). The web server returns the predicted emission value, which is again sent to the broker using a Kafka Producer that is configured to a topic (Wu et al. 2019) named "Database Insertion" after serialization of the dictionary containing the user details, device name, sensor type, emission, and usage values. The Python script indefinitely running the Kafka consumer fetches the message (Narayanan 2024) under the "Database Insertion" topic in the broker and deserializes it to obtain the dictionary. The dictionary is saved to a database and to a CSV file along with timestamps, thus serving as a dataset to re-train the existing models.

3. RESULTS AND DISCUSSIONS

The subsection that follows describes inferences drawn from machine learning models of each type of device and vehicles belonging to the carbon emitter categories of electricity and transport. The consumption parameter has been examined using graphical plots.

3.1 Results from Machine Learning Models

The structured dataset offers a rich training ground for machine learning models. By analyzing the correlations between time-dependent appliance usage, vehicle travel patterns, fuel types, and the calculated carbon footprint, the models can learn to estimate carbon footprint with increasing accuracy. This data-driven approach, with its focus on capturing diurnal variations in user behavior, holds the potential to reduce reliance on manual data input in the future, paving the way for a more automated and efficient system for carbon footprint assessment. The results have been derived from models trained over a 12-month dataset. However, the models can be trained beyond the 12-month dataset.

3.1.1 Inferences drawn from Two-wheeler and Four-wheeler plots

The partial dependence plot in Figure 5 exhibits an overall positive correlation between Two-wheeler usage in kilometers and emission values, with a non-linear relationship characterized by multiple local maxima, minima, and periodic fluctuations. The plot describes the effect that the increasing number of kilometers has on the emissions produced by the two-wheeler. While the intricate shape indicates the significance of Two-wheeler usage as an influential feature for predicting emissions, it also highlights the trade-off between capturing complex non-linear relationships and the interpretability of the model's behavior.

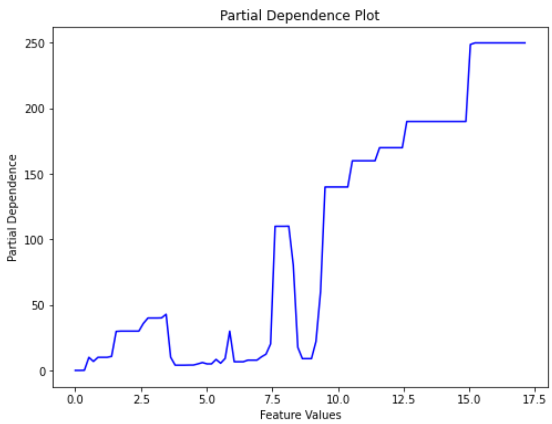

Figure 5: Partial Dependence Plot - 2-Wheeler

The partial dependence plot in Figure 6 represents a non-linear relationship between Four-wheeler usage in kilometers and emission values, with an overall positive correlation, and distinct patterns or regimes based on the fuel type (petrol or diesel), where the curves for petrol and diesel vehicles diverge at certain points, suggesting interactions between the features of Four-wheeler usage and fuel type in influencing emission values. The plot captures complex non-linear relationships and feature interactions, while the varying slopes and curvatures highlight the model's ability to capture these interactions.

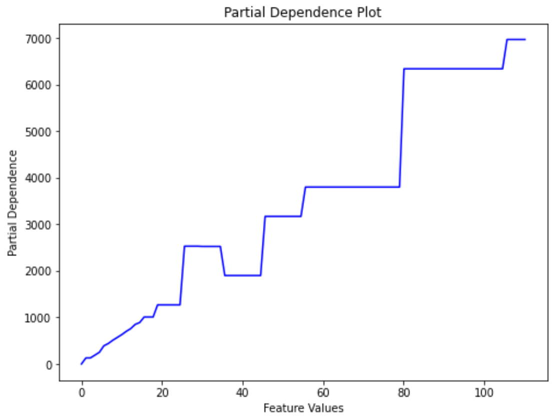

Figure 6: Partial Dependence Plot - 4-Wheeler

3.1.2 Inferences drawn from Tube-light and Air Conditioner plots

The plot in Figure 7 is a scatter plot showcasing the differences in the actual emission values and predicted emission values obtained from training the Linear Regression model on tube-light usage in hours and its corresponding emissions in kgCO2e. A scatter plot has been chosen to analyze the input-output relationship of variables as it provides a clear representation of data points and their alignment with predicted output data points. The simplicity of the scatter plot, in terms of visualization provides a clear rationale for employing it over a partial dependence plot. The scatter points follow a linear trend, indicating that the Linear Regression model is reasonably capturing the relationship between input features and target variables and hence indicates the overall good fit of the model. Points closer to the diagonal represents that the model is more accurate in its predictions.

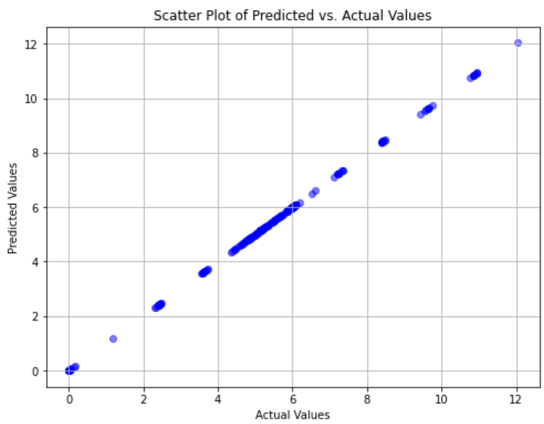

Figure 7: Scatter Plot - Tube Lights

The plot in Figure 8 is a scatter plot showcasing the differences in the actual emission values and predicted emission values obtained from training the Linear Regression model on an Air Conditioner usage data in hours and its corresponding emissions in kgCO2e. The scatter points follow a linear trend and hence indicates the Linear Regression model's ability to capture the relationship between the input and target features inferring an overall good fit. The scatter points are uniformly distributed indicating that the model is trained and tested using a large variety of data points. Points closer to the diagonal represents that the model makes forecasts that are more accurate.

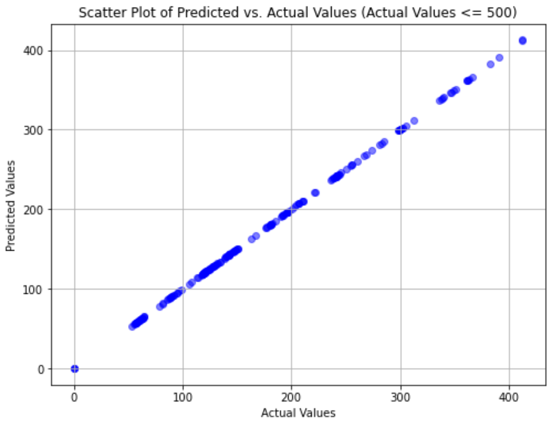

Figure 8: Scatter Plot - Air-Conditioner

3.1.3 Inferences drawn from Washing Machine and Ceiling Fan plots

The plot in Figure 9 represents a partial dependency plot obtained from Gradient Boost Regression after training it on usage data in hours and its corresponding emission values of a Washing Machine. The plot displays an overall increasing trend, indicating a positive correlation between the amount of usage and the emission values. The curve shows an increase in carbon emissions at higher fan usage hours, suggesting that longer washing machine usage times have a significantly larger impact on emissions. The plot displays a step-wise pattern, which is characteristic of tree-based models like Gradient Boost Regression.

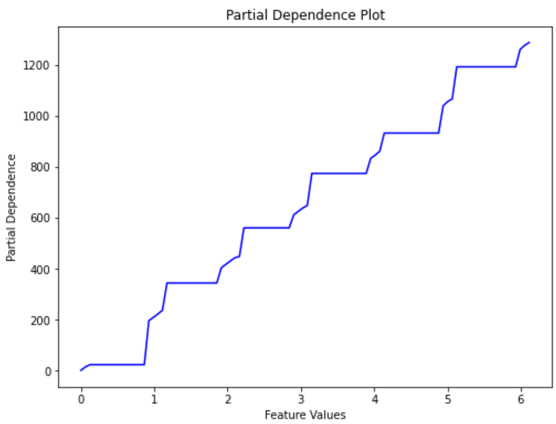

Figure 9: Partial Dependency Plot - Washing Machine

This pattern suggests that the model is capturing different regimes or thresholds in the relationship between washing machine usage hours and emissions.

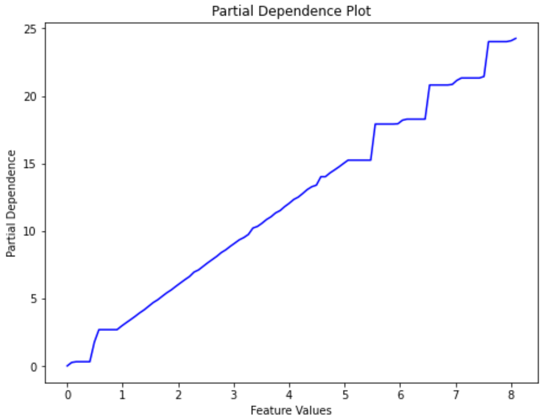

Figure 10: Partial Dependency Plot - Ceiling Fan

The plot in Figure 10 represents a partial dependency plot obtained from Random Forest Regression after training it on usage data in hours and its corresponding emission values of a ceiling fan. The plot displays a monotonically increasing trend, indicating a positive correlation between the number of hours used and the emission values. The curve shows a steeper increase in carbon emissions at higher fan usage hours, suggesting that longer fan usage times have a disproportionately larger impact on emissions.

3.2 Results from Reinforcement Learning Models

The following plots contain three distinct parameters of the reinforcement learning model that drive the performance of the model during its real-time deployment.

The following metrics have been graphically plotted to examine the performance of the reinforcement learning model:

- Steps
- Variation
- Consumption Dictionary Values

These metrics have been defined as follows:

- Steps: The steps are represented by the number of iterations taken by the agent to reach goal consumption in each episode.
- Variation: The precision of the findings with regard to the deviation from the intended state and the calculated overall consumption have been described by the variation between the final and goal consumption. The lower

the variation value, the better the model has performed indicated a more successful implementation of the proposed system.

- Consumption Dictionary Values: Consumption dictionary values refer to the values for each consumption state which are listed in the consumption dictionary. These values provide a relative factor of consumption between the existing states and represents a unit of consumption consumed by the agent in that state in one iteration. The consumption dictionary has been represented as follows: {“Zero”: 0, “Low Temperature”: 1, “Moderate Temperature”: 2, “Coldest Temperature”: 5}

3.2.1 Steps vs Consumption Dictionary Values plot

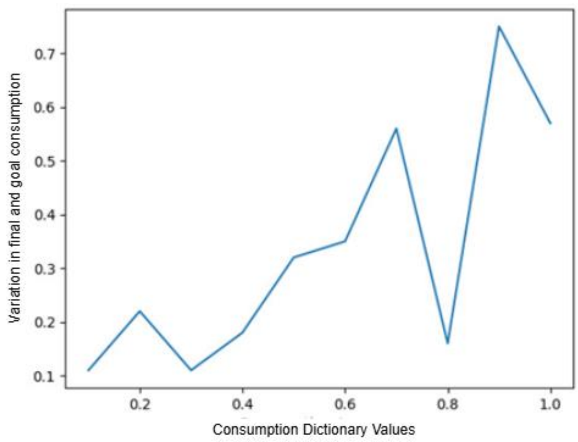

Figure 11: Line Plot - Steps v/s Consumption Dictionary Values

The horizontal axis represents a subset of the possible values of the consumption dictionary and the vertical axis represents the number of steps taken to achieve the values present in the matrix in Figure 11. A decreasing curve describes the relationship between consumption dictionary values and number of steps taken to achieve it. Lower values of consumption values in the dictionary lead to lower units of consumption consumed per iteration of learning. Therefore, a greater number of steps have been taken in order to achieve the final goal state.

3.2.2 Steps vs Variation in final and goal consumption plot

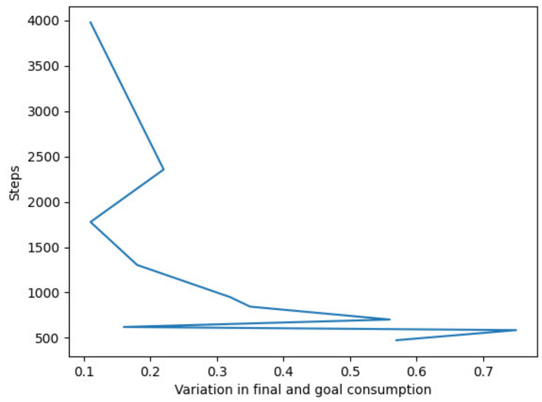

Figure 12: Line Plot **-** Steps v/s Variation in final and goal state consumption

The graph in Figure 12 shows the number of steps taken vs variation in final and goal state consumption. This variation is the result of the absolute difference between the final consumption and goal state consumption. The higher the number of steps indicates a lower variation between final and goal consumption. However, there is a trade-off between accuracy and the time taken for learning in a given episode. A relatively more accurate result is achieved when the algorithm expends more time reach the goal state and vice-versa. The sudden steeps and bends in the graph indicate the randomness of the algorithm which can be attributed to the exploration-exploitation trade-off in Q-learning.

3.2.3 Variation in final and goal consumption vs Consumption Dictionary Values plot

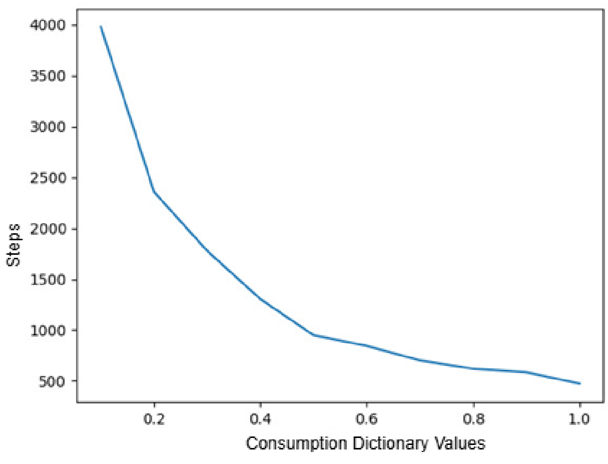

Figure 13: Line Plot **-** Consumption Dictionary Values v/s Variation in final and goal state consumption

Figure 13 connotes the variation in final and goal state consumption with respect to consumption dictionary values. Smaller values of present in the consumption dictionary will lead to smaller differences between final consumption and goal consumption. The choice of consumption values in the dictionary is very critical to the use-case presented in this paper. The objective is to reach a balanced value which is appropriate for the goal state, the device being used and level of accuracy required. Deployment of this model requires the values in the consumption dictionary to undergo optimization for achieving the desired performance.

The key results after user interaction with ConForMist framework have been recorded in Table 6.

Table 6: Key Results obtained using ConForMist framework

Users	Device	Input (km/ hour)	ML model output (kgCO2e)	Offset (%)	Phase of the day	User Prefer-ence (%)	RL model output
User 1	Tube light	5	6.02	5	Morning, Night	53, 74	2h 04m, 2h 40m
User 2	Ceiling fan	8	24.08	10	Afternoon, Evening	27, 80	1h 50m, 5h 22m
User 3	Washing machine	3	632.22	15	Evening, Night	90, 54	1h 25m, 0h 59m
User 4	Air Condi-tioner	9	541.89	10	Morning, Night	67, 61	4h 22m, 3h 38m
User 5	2-wheeler	20	249.56	15	Morning, Evening	42, 63	9.26km, 6.25km
User 6	4-wheeler	50	3169	5	Morning, Afternoon	78, 13	39.54km, 8.14km
User 7	4-wheeler	30	2249.97	10	Afternoon, Night	24, 70	8.33km, 22.6km
User 8	Air Condi-tioner	7	421.47	15	Morning, Evening	80, 49	3h 21m, 2h 37m
User 9	Washing Machine	2	421.48	5	Afternoon, Evening	55, 51	1h 02m, 0h 59m
User 10	Ceiling fan	10	1288.41	10	Evening, Night	63, 25	3h 30m, 1h 51m
User 11	Tubelight	6	7.22	15	Afternoon, Night	69, 49	2h 58m, 2h 07m
User 12	2-wheeler	15	189.81	10	Morning, Afternoon	50, 45	6.67km, 5.33km

3.3 Benchmarking and Validation with real-world data

The data from the ConForMist framework has been compared with real-world provided in a survey report (Hernandez et al. 2022). Table 7 provided in this report serves as a benchmark for carrying out validation with real-world data .

Table 7: Average Monthly Electricity Consumption for a household in Bangalore city

Appliance Type	Consensus Monthly Consumption (Units)	Consensus Monthly Carbon Emission (kgCo2e)
Lighting	11	9.02
Heating	0	0
Cooling	67	54.94
Appliance	45	36.9

The data in Table 8 provides user monthly aggragated carbon emissions and the variance between consensus data and ConForMist data. The benchmarked results show that the variance is minimal indicating the efficacy of the proposed system.

Table 8: ConForMist Monthly Carbon Emissions Benchmarked Against Consensus Table Values

Appliance Type	User Monthly Aggregated Carbon Emission (KgCo2e)	Variance Between Consensus data v/s ConForMist data
Lighting	10	0.98
Heating	0	0
Cooling	55	0.06
Appliance	38	1.1

3.4 The User Interface

GreenStride is the web-based tool developed to comprehensively record usage data across various carbon-emitter categories and elicit recommendations provided by the underlying machine learning data gathered. The front-end user interface developed provides two kinds of users to register and login.

The functionalities implemented in the web-based tool are portrayed in Figures 14 through 18. The key features of the web-based tool are detailed in Table 9.

Table 9: Website UI Functionality Description Table

Key Features	Description
User Roles and Analytics Features in the Emissions Monitoring Application	The application supports two distinct user types. The first comprises organizational or governmental entities that can monitor emissions and resource usage across regions such as constituencies, cities, or districts. These users gain insights through visual representations of emission and usage patterns over time. The second type includes individual users who can log the emissions and consumption associated with their appliances and vehicles. They receive personalized visual analytics, including charts and graphs, depicting their usage trends over defined time periods.
Emission Tracking and Optimization Features for Individual Users	Individual users can specify the type of vehicle or appliance they use along with the usage windows (in km or hours) and the date of usage to calculate the corresponding emissions in kilograms of CO_2 equivalent ($KgCO_2e$). The latter feature is represented in **Figure 14**. Users also have the option to reduce their estimated emissions by 5%, 10%, or 15%,upon which the system provides time-of-day–specific recommendations (morning, afternoon, evening, and night) for optimized usage aimed at lowering emissions which are represented in **Figure 15** and **Figure 16**.
Interactive Dashboard and Data Visualization for Emission Insights	Users are provided with a dashboard that displays comprehensive statistics, including average, minimum, maximum, and total emissions and usage data across all recorded vehicles and appliances. The platform enables visualization of usage and emission patterns through various filters such as month, appliance type, and vehicle category. Product-specific statistics through radar charts as shown in **Figure 17**, monthly trends via area charts as shown in **Figure 18** and daily statistics are presented using line charts as shown in **Figure 19**, and, facilitating a multi-dimensional understanding of consumption and emission behavior.
Advanced Dashboard and Regional Analytics for Governmental and Organizational Users	Governmental and organizational users have access to a dedicated dashboard that presents aggregated emissions and usage patterns. An additional location-based filter enables these users to visualize statistics specific to a selected region. Furthermore, an extended radar chart is provided as shown in **Figure 20** to depict average, minimum, and maximum emissions and usage metrics across all monitored locations, facilitating comparative analysis and informed decision-making.

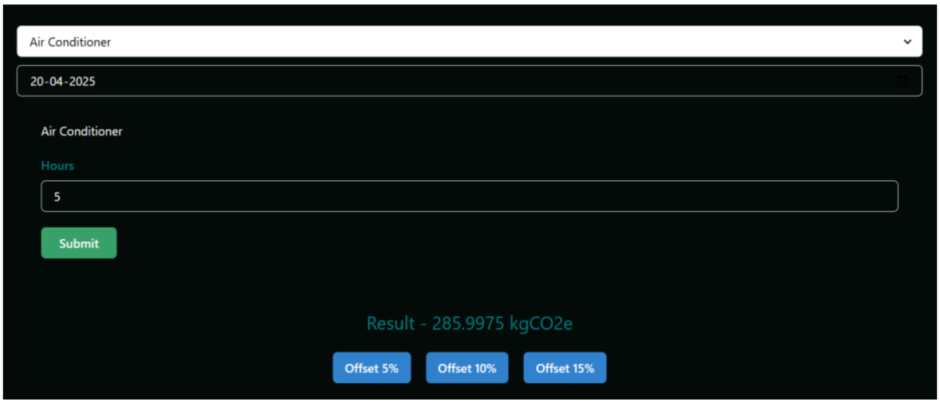

Figure 14: ConForMiSt Web User Interface (UI) - Emission calculator based on usage.

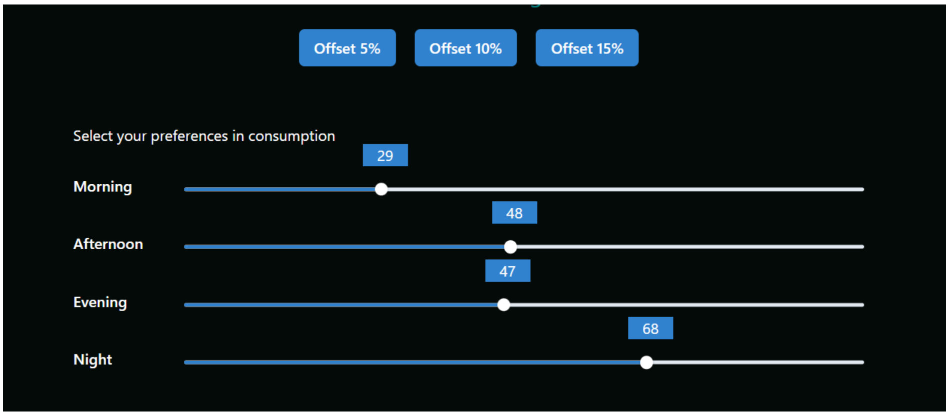

Figure 15: ConForMiSt UI Dashboard - User preferences for usage recommendations.

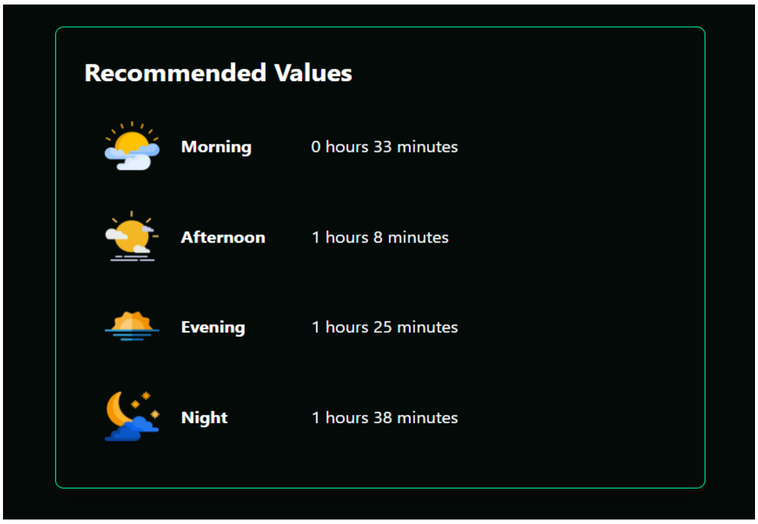

Figure 16: ConForMiSt UI Dashboard - Usage recommendation for lowering emissions

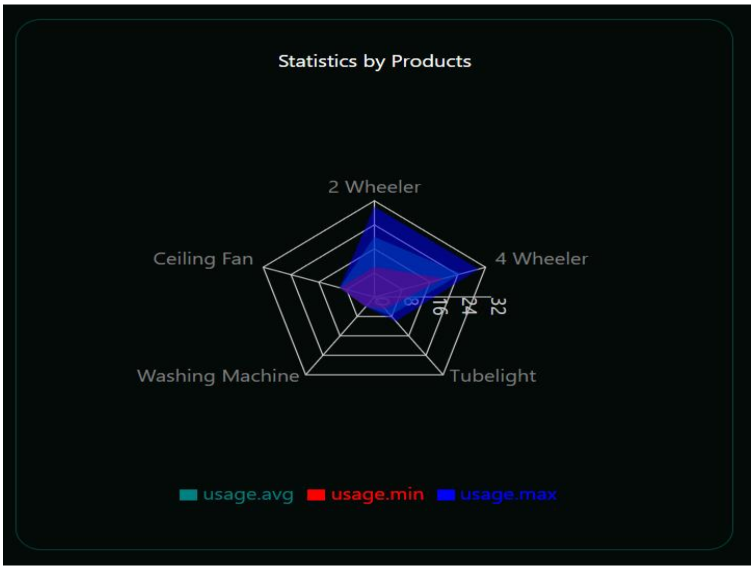

Figure 17: ConForMiSt UI Dashboard - Carbon Footprint Radar Plots for Appliances

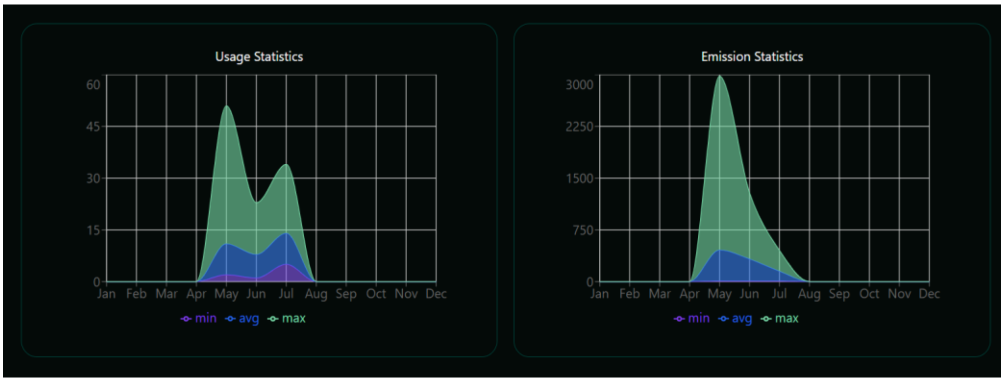

Figure 18: ConForMiSt UI Dashboard - Recommendations (Appliances)

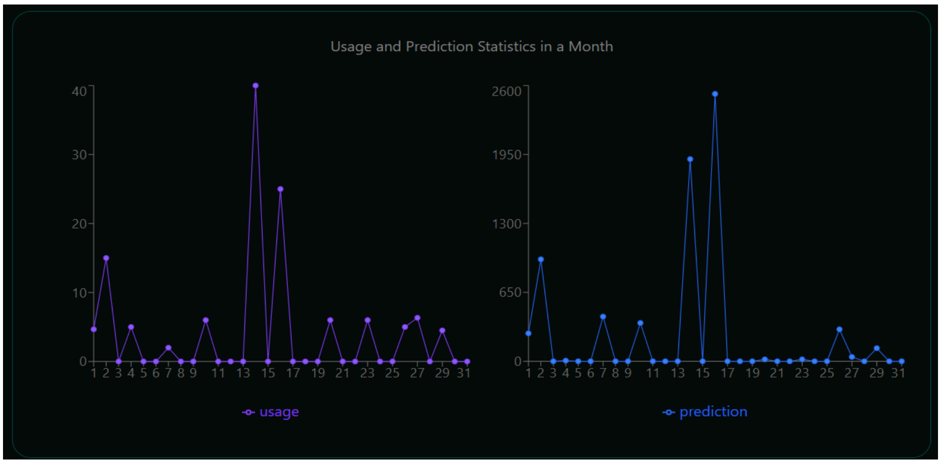

Figure 19: ConForMiSt UI Dashboard - Line charts for usage and emission patterns for a month

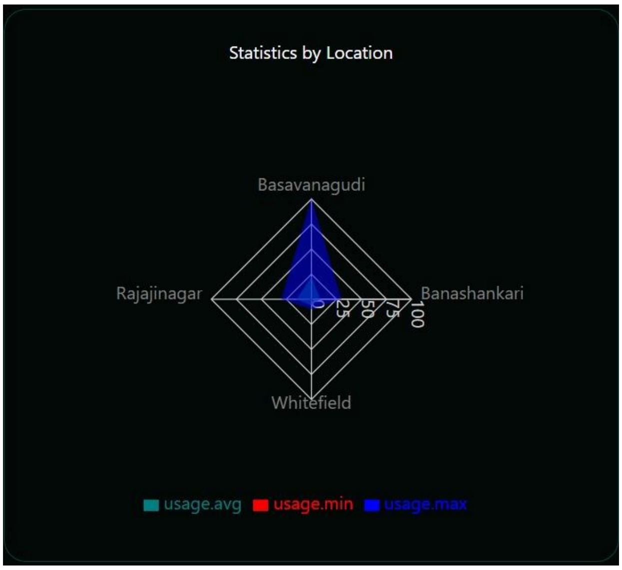

Figure 20: ConForMiSt UI Dashboard – Radar chart representing location-wise statistics

4.CONCLUSIONS

The proposed framework makes contributions in several areas:

- Data Gathering: The proposed system has presented a simple approach to gathering data from primary carbon emitter sources (exemplar ESP32 based hardware as the foundation, chosen for its computational and IoT features) and recording real-time data in a database enabled by Apache Kafka.
- Framework Design: A Dual phase framework: The design provides separation of concerns: first stage to understand the data and the second stage to drive recommendations with a model free approach.
- The first stage uses various ML models operating on a structured dataset to elicit insights about the carbon emission data. The models selected in this stage use the RMSE performance index.
- The second stage, a reinforcement learning model that uniquely models states and actions for various consumption categories drives the recommendations process. The unique state modeling capturing many aspects is a novelty.
- Experimentation and Result Interpretation: Inferences drawn from two exemplar categories: Electrical appliances category and Transport category, using scatter plots and partial dependence plots.

- User Interface (UI) presentation: These models are used to provide recommendations that will influence future consumption judiciously and also mitigate carbon emissions. Options exist to analyze current consumption and emissions data (Eg: Radar plots, statistical data)
- Comparison with other frameworks: The proposed ConForMiSt has been evaluated against models that are proposed for similar goals using deep learning, RL, Deep-Q learning, and LSTM with IoT. However, ConForMiSt has a broader scope to analyze different categories of data and provides a mechanism to aggregate data for higher organizational levels.
- An important aspect of hyper-parameter (consumption) optimization is elucidated.

4.1 LIMITATIONS

- Relatively limited data sample size used for training and testing: Machine Learning models operated on consumption and corresponding emission data for the span of a year, which was relatively small.
- A larger sample size spanning multiple years would help in reducing Mean Absolute Errors and increasing the accuracy of the models, thus achieving a better fit. Although the sample size considered gathering data in different time slots of a day for a multitude of carbon emission producers, it did not consider demographic factors such as age and gender, and location of the consumers.
- Incorporating parameters such as age, gender, and emission source can greatly help in detecting hidden patterns and improving predictions from Machine Learning models by providing better context.

4.2 FUTURE WORK

Smart home systems and designs can incorporate the implementation proposed in this study to encourage consumers to utilize resources effectively. Smart thermostats, lighting controls and automated energy-efficient appliance control systems can be implemented to monitor and reduce energy usage and resulting carbon footprint. Federated machine learning models can be trained and deployed across various devices to make predictions and provide recommendations thereby eliminating the need for a centralized training and deployment station. Edge Computing technologies can be paired with the sensor implementation proposed in the study for improved real-time data collection and processing.

Author Contributions: Conceptualization, Sumukh R. Kashi and Yashas Surendra.; methodology, Sumukh R. Kashi, Yashas Surendra and H.R Vineeth.; software, Sumukh R. Kashi, Yashas Surendra and H.R Vineeth.; validation, Sumukh R. Kashi, Yashas Surendra, H.R Vineeth and K.S Sowmya.; formal analysis, K.S Sowmya.; investigation, Sumukh R. Kashi, Yashas Surendra and H.R Vineeth.; resources, Sumukh R. Kashi, Yashas Surendra and H.R Vineeth.; data curation, Sumukh R. Kashi, Yashas Surendra and H.R Vineeth.; writing—original draft preparation, Sumukh R. Kashi, Yashas

Surendra and H.R Vineeth.; writing—review and editing, Sumukh R. Kashi, Yashas Surendra and H.R Vineeth.; visualization, Sumukh R. Kashi, Yashas Surendra and H.R Vineeth.; supervision, K.S Sowmya. ; project administration, K.S Sowmya.;. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Conflicts of Interest: The authors declare no conflicts of interest.

REFERENCES

Ulfat, W., and Noori, S., 2024. The climate and global warming responses to the changing greenhouse gas concentration in the atmosphere. *Nangarhar University International Journal of Biosciences*, pp.545-548.

Chandel, T.A., 2022. Green entrepreneurship and sustainable development. International Perspectives on Value Creation and Sustainability Through Social Entrepreneurship. *IGI Global*, pp.173-208.

Jones, M.W., Peters, G.P., Gasser, T., Andrew, R.M., Schwingshackl, C., Gütschow, J., Houghton, R.A., Friedlingstein, P., Pongratz, J., and Le Quéré, C., 2023. National contributions to climate change due to historical emissions of carbon dioxide, methane, and nitrous oxide since 1850. *Scientific Data*, 10(1), p.155.

Schwenkenbecher, A., 2014. Is there an obligation to reduce one's individual carbon footprint? Critical *Review of International Social and Political Philosophy*, 17(2), pp.168-188.

Jain, M., and Nagpal, A., 2021. Examining the role of principal contributors to ‘sectoral carbon emissions’: In the Indian context. *Vision*, 25(4), pp.399-406.

Onat, N.C., Kucukvar, M., and Tatari, O., 2015. Conventional, hybrid, plug-in hybrid or electric vehicles? State-based comparative carbon and energy footprint analysis in the United States. *Applied Energy*, 150, pp.36-49.

Ramachandra, T.V., Aithal, B.H., and Sreejith, K., 2015. GHG footprint of major cities in India. *Renewable and Sustainable Energy Reviews*, 44, pp.473-495.

Huang, L., et al., 2023. Sustainable lifestyle: Urban household carbon footprint accounting and policy implications for lifestyle-based decarbonization. *Energy Policy*, 181, p.113696.

Khan, A.A., 2024. An analytical study on the impact of global warming: Effects on environment. *Recent Trends in Commerce, Management, Accountancy and Business Economics*, p.54.

Kiehbadroudinezhad, M., Merabet, A., and Hosseinzadeh-Bandbafha, H., 2024. Health impacts of greenhouse gases emissions on humans and the environment. *Advances and Technology Development in Greenhouse Gases: Emission, Capture and Conversion. Elsevier*, pp.265-291.

Kumar, A., Mishra, S., Bakshi, S., Upadhyay, P., and Thakur, T.K., 2023. Response of eutrophication and water quality drivers on greenhouse gas emissions in lakes of China: A critical analysis. *Ecohydrology*, 16(1), e2483.

Goswami, M., Premkumar, A., Kishan, R., and Nautiyal, S., 2024. A methodological perspective on inclusive assessment of household carbon footprint. *Environment, Development and Sustainability*, pp.1-28.

Lövehagen, N., Malmodin, J., and Bergmark, P., 2023. Assessing embodied carbon emissions of communication user devices by combining approaches. *Renewable and Sustainable Energy Reviews*, 183, 113422.

Mancini, M.S., Galli, A., Niccolucci, V., Lin, D., Hanscom, L., and Marchettini, N., 2016. Ecological footprint: refining the carbon footprint calculation. *Ecological Indicators*, 61, pp.390-403.

Gao, T., Liu, Q., and Wang, J., 2014. A comparative study of carbon footprint and assessment standards. *International Journal of Low-Carbon Technologies*, 9(3), pp.237-243.

Musa, M., Rahman, M.M., Khan, M.A., and Alam, K., 2024. Sustainable development challenges in Bangladesh: an empirical study of economic growth, industrialization, energy consumption, foreign investment, and carbon emissions—using dynamic ARDL model and frequency domain causality approach. *Clean Technologies and Environmental Policy*, 26(6), pp.1799-1823.

Rondoni, A., and Grasso, S., 2021. Consumers behaviour towards carbon footprint labels on food: A review of the literature and discussion of industry implications. *Journal of Cleaner Production*, 301, 127031.

Nguyen, Q., Diaz-Rainey, I., and Kuruppuarachchi, D., 2021. Predicting corporate carbon footprints for climate finance risk analyses: a machine learning approach. *Energy Economics*, 95, 105129.

Henderson, P., Hu, J., and Romoff, J., 2020. Towards the systematic reporting of the energy and carbon footprints of machine learning. *Journal of Machine Learning Research*, 21(248), pp.1-43.

Dong, H., and Zhang, L., 2023. Transition towards carbon neutrality: Forecasting Hong Kong's buildings carbon footprint by 2050 using a machine learning approach. *Sustainable Production and Consumption*, 35, pp.633-642.

Liu, H., Wong, W.-K., Phan, T.C., Nassani, A.A., Haffar, M., and Abu-Rumman, A., 2023. Linkage among urbanization, energy consumption, economic growth and carbon emissions. Panel data analysis for China using ARDL model. *Fuel*, 332, 126122.

Lissa, P., Deane, C., Schukat, M., Seri, F., Keane, M. and Barrett, E., 2021. Deep reinforcement learning for home energy management system control. *Energy and AI*, 3, p.100043.

Lee, S. and Choi, D.-H., 2019. Reinforcement learning-based energy management of smart home with rooftop solar photovoltaic system, energy storage system, and home appliances. *Sensors*, 19(3937).

Mahmoud, M. and Ben Slama, S., 2025. Deep learning-based home energy management incorporating vehicle-to-home and home-to-vehicle technologies for renewable integration. *Energies*, 18(129).

Solatidehkordi, Z., Ramesh, J., Al-Ali, A.R., Osman, A. and Shaaban, M., 2023. An IoT deep learning-based home appliances management and classification system. *Energy Reports, 9(Supplement 3)*, pp.503–509.

Maulud, D.H., and Abdulazeez, A.M., 2020. A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trend*s, 1(2), pp.140-147.

Ali, J., Khan, R., Ahmad, N., and Maqsood, I., 2012. Random forests and decision trees. *International Journal of Computer Science Issues (IJCSI)*, 9(5), pp.272-278.

Otchere, D.A., Agyekum, E.B., and Adjei, E., 2022. Application of gradient boosting regression model for the evaluation of feature selection techniques in improving reservoir characterisation predictions. *Journal of Petroleum Science and Engineering*, 208, 109244.

Varoquaux, G., and Colliot, O., 2023. Evaluating machine learning models and their diagnostic value. *Machine Learning for Brain Disorders*, pp.601-630.

Ranjan, G.S.K., Verma, A.K., and Radhika, S., 2019. K-nearest neighbors and grid search CV based real-time fault monitoring system for industries. *2019 IEEE 5th International Conference for Convergence in Technology (I2CT).* IEEE, pp.1-5.

Bhatti, N., and Iqbal, M., 2000. Object serialisation and deserialisation using XML. *10th International Conference on Management of Data (COMAD 2000).*

Bellman, R., 1957. A Markovian decision process. *Journal of Mathematics and Mechanics*, 6(5), pp.679-684.

Watkins, C.J.C.H., and Dayan, P., 1992. Q-learning. *Machine Learning*, 8, pp.279-292.

Babiuch, M., and Postulka, J., 2020. Smart home monitoring system using ESP32 microcontrollers. *Internet of Things*, 12, 100314.

Hiraman, B.R., 2018. A study of Apache Kafka in big data stream processing. *2018 International Conference on Information, Communication, Engineering and Technology (ICICET).* IEEE, pp.1-4.

Thein, K.M.M., 2014. Apache Kafka: Next generation distributed messaging system. *International Journal of Scientific Engineering and Technology Research*, 3(47), pp.9478-9483.

Bahrami, M., Singhal, S., and Singhal, M., 2018. API learning: Applying machine learning to manage the rise of API economy. *Companion Proceedings of the The Web Conference 2018. International World Wide Web Conferences Steering Committee*, pp.1591-1598.

Raptis, T.P., and Passarella, A., 2022. On efficiently partitioning a topic in Apache Kafka. *2022 International Conference on Computer, Information and Telecommunication Systems (CITS).* IEEE, pp.1-5.

Wu, H., Shang, Z., and Wolter, K., 2019. Performance prediction for the Apache Kafka messaging system. *2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS).* IEEE, pp.1931-1938.

Narayanan, P.K., 2024. Engineering real-time data pipelines using Apache Kafka. *Data Engineering for Machine Learning Pipelines: From Python Libraries to ML Pipelines and Cloud Platforms. Berkeley,* CA: Apress, pp.277-322.

Hernandez, M., Bhagavatula, K., Cibi, S., R., K., Krishnan, S., Malaviya, S. and Jairaj, B., 2022. Shifting household energy use in Bangalore, India: Using behaviorally informed energy reports. *Working Paper*. Washington, DC: World Resources Institute.