IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 4, May 2023

Estimation of Carbon Footprint for Home and Industry using Machine Learning

Gokulakrishnan V[1], Mohan V[2], Aahin K[3], Kamaleshwaran R[4], Baranikumar R[5]
Faculty Member, Department of Computer Science and Engineering[1]
Students, Department of Computer Science and Engineering[2,3,4,5]
Dhanalakshmi Srinivasan Engineering College, Perambalur, India

Abstract: *Carbon footprint has become a popular in recent years among Meteorologist's. The existing system predicts the carbon footprint using sparse regression algorithm. But it has the problem in subset selection and correlation between responder and predictor, so the system uses decision tree algorithm. Decision Trees are a sort of supervised machine learning in which the training data is continually segmented based on a particular parameter, with you describing the input and the associated output. Decision nodes and leaves are the two components that can be used to explain the tree.The choices or results are represented by the leaves .The decision tree divides the nodes .So the algorithm the system use is very efficient and time saving and both in accurate .So from this it can calculate the carbon footprint of home and industry in particular region .When it implement this can find the emitters who have huge amount of carbon Emission which will have huge impact on environment. Already scholars have raised hands on this carbon footprint.*

Keywords: Carbon footprint, Decision tree, Sparse regression, Subset selection

I. INTRODUCTION

Even so, it is a relatively straightforward calculation compared to evaluating the emissions involved in each stage. The emissions that occur at the assembly plant, the creation of the machinery used at those factories and at the assembly plant, the transport of all the component parts, the factories where the components were made, and so on, all the way back to the extraction of the miner. These calculators seek to provide you an estimate of the amount of greenhouse gases being emitted to support your way of life by asking you questions about your household's fuel use, and travel pattern. Decision Trees are a sort of supervised machine learning in which the training data is continually segmented based on a particular parameter, with you describing the input and the associated output. The two elements that can be utilised to explain the tree are decision nodes and leaves. The leaves stand in for the options or outcomes. The decision nodes divide the data. So, the algorithm we use is very efficient than existing system of sparse regression. The result we expect is very efficient and time saving and both in accurate. So from this we can calculate the carbon footprint of home and industrial in particular region. when we implement this we can find the emitters who have huge amount of carbon emission which will have huge impact on environment. Already scholars have raised hands on this carbon footprint. The algorithm we use for this decision tree.

As our proposed algorithms is efficient manner in terms of speeding up the process and predict the exact results in estimation of carbon footprint for home and industry. So, the system we have intend to find the exact emission for home and industry in specific region. A supervised machine learning algorithm called Decision Tree uses a set of principles to make judgments, much like how people do. A machine learning classification algorithm can be thought of as being created to make decisions.The model is typically said to forecast the class of the novel, previously unseen input, but in reality, the algorithm must choose the class to be assigned. You use a rule-based approach when organizing your upcoming vacation. Depending on how long you plan to spend on vacation, your budget, and whether or not your extended family is joining you, you might choose a different location. The answer to these queries influences the choice that is made. Additionally, This will be efficient in providing the exact information as we required. In this algorithm we have predict the exact emission outers in the specific region whether it is home user or industry user

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 4, May 2023

II. RELATED WORKS:

In this section, we evaluated some of the research similar to our project that has been conducted by various authors and researchers utilizing machine learning techniques to predict carbon footprint cases and the norm using user input.

Ana Radovanovic´, Ross Koningstein, Ian Schneider, Bokan Chen, Alexandre Duarte, Binz Roy, Diyue Xiao, Maya Haridasan, Patrick Hung, Nick Care, Saurav Talukdar, Eric Mullen, Kendal Smith, MariEllen Cottman,and Walfredo Cirne[1] verified the amount of CO2 emitted per kilowatt-hour inn Google data center.Due to the different forms of generation, the quantity of CO2 emitted per kilowatt-hour on an energy system varies significantly by location and by time of day.Many compute workloads are flexible in the times and locations where they run. Smaller peaks lessen the requirement for additional capacity because datacenters are designed based on peak power and resource demand.

Avita Katal, Susheela Dahiya, Tanupriya Choudhury [2] evaluated the carbon emission in cloud computing.By definition, the term "data center" is an assumption. It brings to mind a time when back-office computer systems for businesses were mostly used for data storage and were pieced together in a basement or closet. Nobody was supposed to see or be aware of "infrastructure" like a sewage system or the road's foundation under the potholes. These presumptions have all been proven false. A company's IT infrastructure consists of its computing resources, networking, and data storage. Additionally, it naturally tends to disperse, just like the enterprise. As with many other unique ideas in the IT sector, there is no agreed-upon description of what a hyperscale data center is.

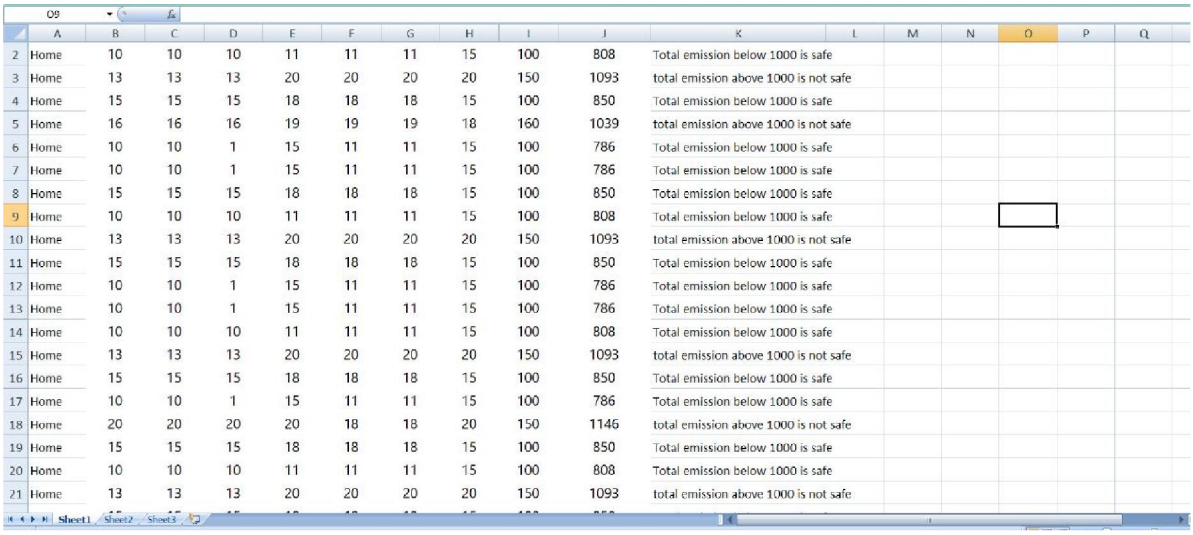
Samuel Asumadu Sarkodie a, Maruf Yakubu Ahmed a, Thomas Leirvik a, b, c,[3]calculated the trade volume of bitcoin energy consumption and carbon footprint.The biological interactions between species, when there is/is not a benefit for at least one species, are the inspiration for the framework. This kind of biological relationship emphasis the significance of interactions between measured variables and the accompanying externalities, whether they are symmetric or asymmetric (Von Jacobi, 2018). We hypothesis that bitcoin trade volume vs. carbon and energy footprint have mutualistic effects known as feedback interaction, simulating the relationship where both species gain. Energy and carbon footprint intensity are projected to raise bitcoin trade volume in conservation interactions, and any measure to lower energy and carbon intensity will result in a fall in bitcoin trading volume. On the other hand, it's possible that rising bitcoin trade volume will increase the carbon and energy footprint, which is known as the growth relationship.

Xinlu Sun a, Zhifu Mia, Andrew Sudmant b, D'Maris Coffmana, Pu Yanga, Richard Wood c [4] took time to done a report that mainly specifies the estimation of carbon footprints in global cities using crowdsourced data.To calculate the carbon footprints of world cities, this study employs a hybrid methodology that combines bottom-up crowdsourcing data with top-down input-output analyses. The term "carbon footprint" refers to the carbon dioxide emissions created during the production of goods and services used by consumers, governments, and other organisations. It includes some of the scope. 3 emissions (emissions beyond the city limit that are brought on by consumption inside the city), full scope 1 emissions (direct emissions excluding those brought on by exports), and full scope 2 emissions (emissions brought on by the use of grid-supplied energy, heating, and/or cooling). One of the key variables used to calculate cities' carbon footprints is their purchasing power, which is determined by their income level. confirms that income is a reliable predictor for calculating carbon footprints.

Cesario Tavares a,, Xincheng Wang b, Sajib Saha c, Zachary Grasley d[5] designed a tool to minimize carbon footprint and cost of UHPC.This investigation was carried out utilising a novel test prototype that permits substantial data collecting to feed machine learning models, which was motivated by the resource and time-consuming nature of existing conventional methods. For the purpose of testing cement pastes and mortars with strengths up to the level of UHPC materials, a test procedure utilising cylindrical specimens with decreased dimensions was evaluated. In accordance with this process, cylinders with a height of around 50 mm were quickly produced and capped with commercially available felt cushion pads (commonly referred to as heavy duty pads). With this technique, big datasets may be produced quickly. The entire mixing and casting process takes around seven times as long as the ASTM C1856/1856 M standard procedure, which uses cylinders measuring 76 mm by 152 mm.

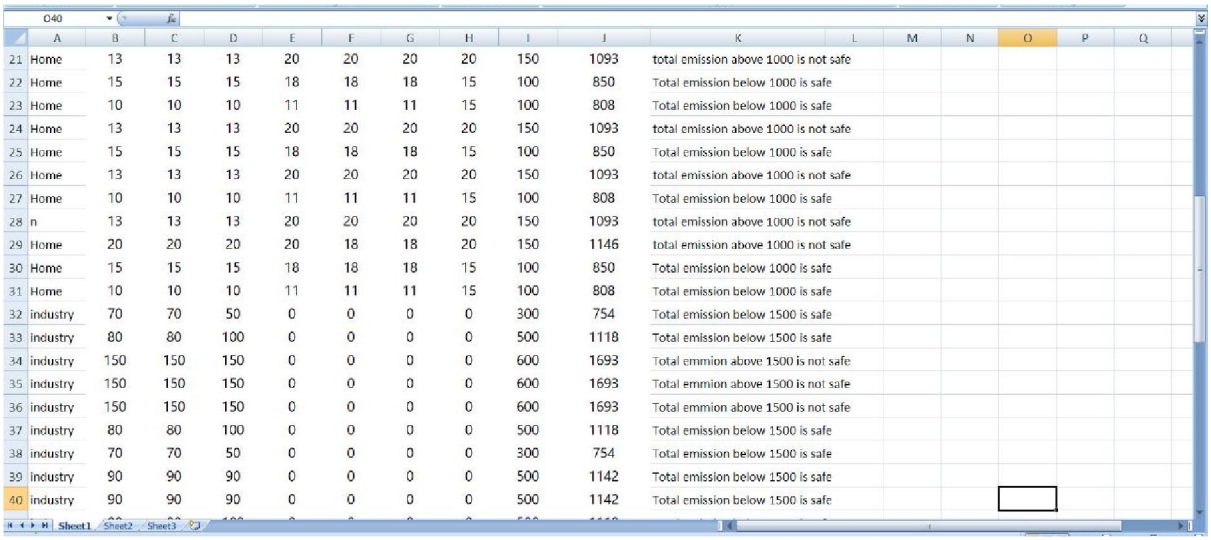
IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 4, May 2023

III. DATASET



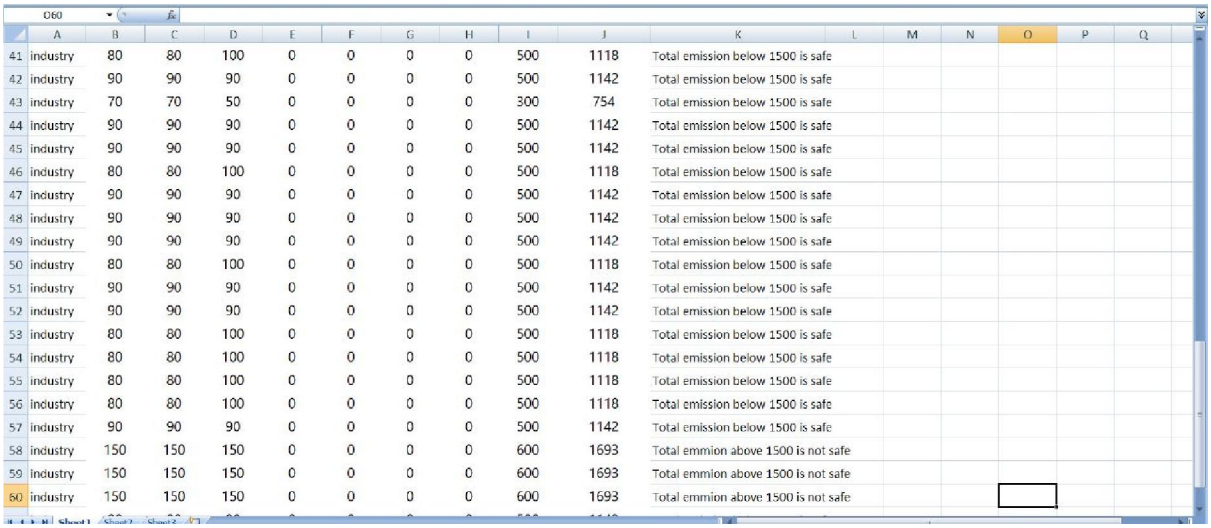
	A	B	C	D	E	F	G	H	I	J	K
2	Home	10	10	10	11	11	11	15	100	808	Total emission below 1000 is safe
3	Home	13	13	13	20	20	20	20	150	1093	total emission above 1000 is not safe
4	Home	15	15	15	18	18	18	15	100	850	Total emission below 1000 is safe
5	Home	16	16	16	19	19	19	18	160	1039	total emission above 1000 is not safe
6	Home	10	10	1	15	11	11	15	100	786	Total emission below 1000 is safe
7	Home	10	10	1	15	11	11	15	100	786	Total emission below 1000 is safe
8	Home	15	15	15	18	18	18	15	100	850	Total emission below 1000 is safe
9	Home	10	10	10	11	11	11	15	100	808	Total emission below 1000 is safe
10	Home	13	13	13	20	20	20	20	150	1093	total emission above 1000 is not safe
11	Home	15	15	15	18	18	18	15	100	850	Total emission below 1000 is safe
12	Home	10	10	1	15	11	11	15	100	786	Total emission below 1000 is safe
13	Home	10	10	1	15	11	11	15	100	786	Total emission below 1000 is safe
14	Home	10	10	10	11	11	11	15	100	808	Total emission below 1000 is safe
15	Home	13	13	13	20	20	20	20	150	1093	total emission above 1000 is not safe
16	Home	15	15	15	18	18	18	15	100	850	Total emission below 1000 is safe
17	Home	10	10	1	15	11	11	15	100	786	Total emission below 1000 is safe
18	Home	20	20	20	20	18	18	20	150	1146	total emission above 1000 is not safe
19	Home	15	15	15	18	18	18	15	100	850	Total emission below 1000 is safe
20	Home	10	10	10	11	11	11	15	100	808	Total emission below 1000 is safe
21	Home	13	13	13	20	20	20	20	150	1093	total emission above 1000 is not safe

Figure 1



	A	B	C	D	E	F	G	H	I	J	K
21	Home	13	13	13	20	20	20	20	150	1093	total emission above 1000 is not safe
22	Home	15	15	15	18	18	18	15	100	850	Total emission below 1000 is safe
23	Home	10	10	10	11	11	11	15	100	808	Total emission below 1000 is safe
24	Home	13	13	13	20	20	20	20	150	1093	total emission above 1000 is not safe
25	Home	15	15	15	18	18	18	15	100	850	Total emission below 1000 is safe
26	Home	13	13	13	20	20	20	20	150	1093	total emission above 1000 is not safe
27	Home	10	10	10	11	11	11	15	100	808	Total emission below 1000 is safe
28	n	13	13	13	20	20	20	20	150	1093	total emission above 1000 is not safe
29	Home	20	20	20	20	18	18	20	150	1146	total emission above 1000 is not safe
30	Home	15	15	15	18	18	18	15	100	850	Total emission below 1000 is safe
31	Home	10	10	10	11	11	11	15	100	808	Total emission below 1000 is safe
32	industry	70	70	50	0	0	0	0	300	754	Total emission below 1500 is safe
33	industry	80	80	100	0	0	0	0	500	1118	Total emission below 1500 is safe
34	industry	150	150	150	0	0	0	0	600	1693	Total emmion above 1500 is not safe
35	industry	150	150	150	0	0	0	0	600	1693	Total emmion above 1500 is not safe
36	industry	150	150	150	0	0	0	0	600	1693	Total emmion above 1500 is not safe
37	industry	80	80	100	0	0	0	0	500	1118	Total emission below 1500 is safe
38	industry	70	70	50	0	0	0	0	300	754	Total emission below 1500 is safe
39	industry	90	90	90	0	0	0	0	500	1142	Total emission below 1500 is safe
40	industry	90	90	90	0	0	0	0	500	1142	Total emission below 1500 is safe

Figure 2



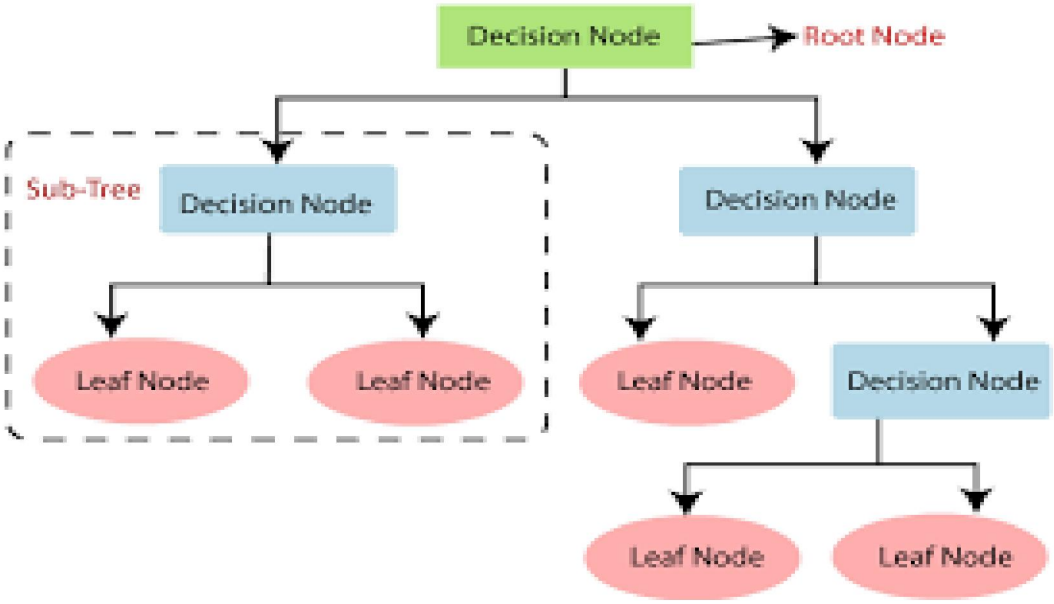
	A	B	C	D	E	F	G	H	I	J	K
41	industry	80	80	100	0	0	0	0	500	1118	Total emission below 1500 is safe
42	industry	90	90	90	0	0	0	0	500	1142	Total emission below 1500 is safe
43	industry	70	70	50	0	0	0	0	300	754	Total emission below 1500 is safe
44	industry	90	90	90	0	0	0	0	500	1142	Total emission below 1500 is safe
45	industry	90	90	90	0	0	0	0	500	1142	Total emission below 1500 is safe
46	industry	80	80	100	0	0	0	0	500	1118	Total emission below 1500 is safe
47	industry	90	90	90	0	0	0	0	500	1142	Total emission below 1500 is safe
48	industry	90	90	90	0	0	0	0	500	1142	Total emission below 1500 is safe
49	industry	90	90	90	0	0	0	0	500	1142	Total emission below 1500 is safe
50	industry	80	80	100	0	0	0	0	500	1118	Total emission below 1500 is safe
51	industry	90	90	90	0	0	0	0	500	1142	Total emission below 1500 is safe
52	industry	90	90	90	0	0	0	0	500	1142	Total emission below 1500 is safe
53	industry	80	80	100	0	0	0	0	500	1118	Total emission below 1500 is safe
54	industry	80	80	100	0	0	0	0	500	1118	Total emission below 1500 is safe
55	industry	80	80	100	0	0	0	0	500	1118	Total emission below 1500 is safe
56	industry	80	80	100	0	0	0	0	500	1118	Total emission below 1500 is safe
57	industry	90	90	90	0	0	0	0	500	1142	Total emission below 1500 is safe
58	industry	150	150	150	0	0	0	0	600	1693	Total emmion above 1500 is not safe
59	industry	150	150	150	0	0	0	0	600	1693	Total emmion above 1500 is not safe
60	industry	150	150	150	0	0	0	0	600	1693	Total emmion above 1500 is not safe

Figure 3

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 4, May 2023

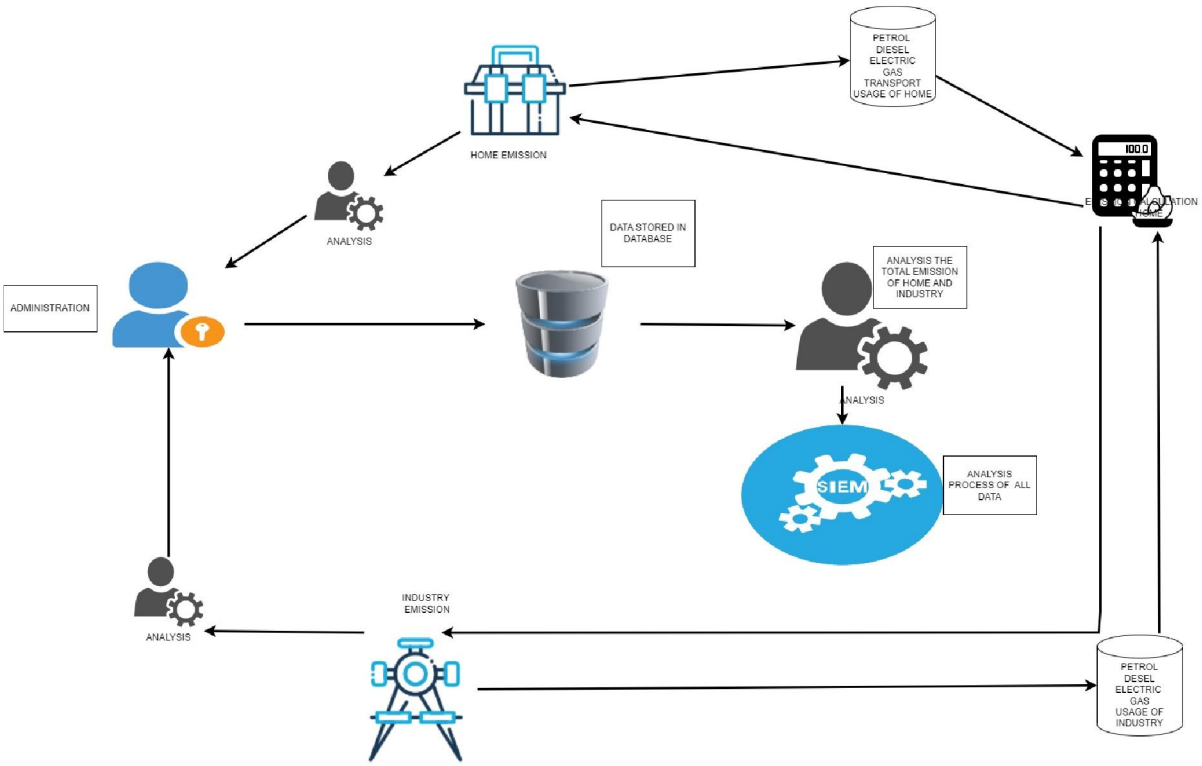
In this work, the model was trained using a dataset that includes inputs like location, vehicle use, petrol consumption, petrol consumption, electricity consumption, and total emission, all of which were acquired from publicly available sources. Additionally, 1200 publicavailable data were used to train the model.
Figures 1, 2, and 3 depict the dataset that was used to train the model

IV. DECISION TREE ARCHITECTURE



In this paper,we used the decision tree algorithm and the architecture for this algorithm is shown above.

V. PROPOSED SYSTEM ARCHITECTURE

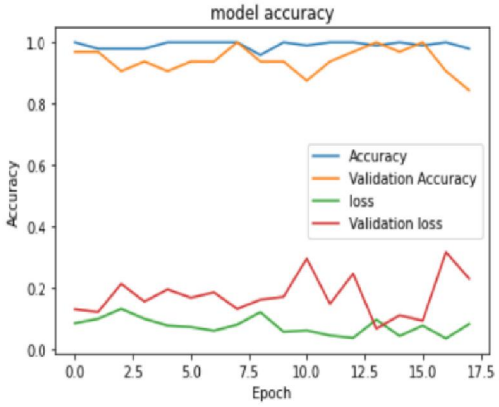


IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 4, May 2023

As the proposed algorithms is efficient manner in terms of speeding up the process and predict the exact results in estimation of carbon footprint for home and industry. So, the system we have intend to find the exact emission for home and industry in specific region. A supervised machine learning algorithm called Decision Tree uses a set of principles to make judgments, much like how people do. A machine learning classification algorithm can be thought of as being created to make decisions .The model is typically said to forecast the class of the novel, previously unseen input, but in reality, the algorithm must choose the class to be assigned. You use a rule-based approach when organizing your upcoming vacation.

VI . EVALUATION METRICS

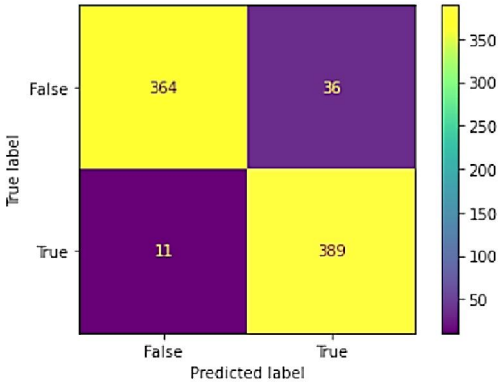
We achieved an accuracy of 94% after the model was fitted and trained and an accuracy of 84% after validation. Early stopping ended the training at the 18th epoch because the accuracy of the validation did not improve. And for the accuracy, loss, validation accuracy, and validation loss, we plotted a graph. Five evaluation measures were utilized to assess the performance of the suggested model: accuracy, precision, recall, specificity, and F1 Score The plotted graph for the model training is shown in Figure



Accuracy and loss graph

Confusion Matrix

The confusion matrix is an important matrix to create to find the precision, F1 score, specificity, sensitivity, and accuracy. A confusion matrix is used to find the False Negative, False positive, True Negative, and True positive values of the prediction of the trained model. To explain how effectively a categorization system performs, a confusion matrix is utilized. A confusion matrix displays and summarizes the effectiveness of a classification method. True positive: The emission where we predicted as carbon emission high and even actually it is the high emission. True positive prediction is 389.



True negative: The emission where we predict as carbon emission low and even actually it is a low emission. True negative prediction is 364.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 4, May 2023

False positive: The emission where we predict carbon emission high but it is a high emission. The false positive prediction is 36.
False negative: The emission where we predict carbon emission high, but it is a low emission. The false negative prediction is 11.

Confusion matrix of Proposed model

Accuracy

The statistics known as accuracy are used to evaluate the effectiveness of classification and regression algorithms. The accuracy value for the proposed model is 94 percent.

Accuracy = (TP + TN) / (TP + TN + FP +FN)
= (389 + 364) / (389 + 364 + 36 + 11)
= 0.94125

Precision

Precision is a measure of how well the suggested model classified the positive photos. The number of true positives divided by the number of true positive predictions is an indicator of the model's success and is explained by the number of positive predictions produced. The precision value of the proposed model is 91 percent.

Precision = TP / (TP + FP)
= 389 / (389 + 36)
= 0.915

Recall

Recall refers to positive findings that are correctly classified as positive. What it is referred to as is a real positive rate. The Recall value of the proposed model is 97 percent.

Recall = TP / (TP + FN)
= 389 / (389 + 11)
= 0.97

Specificity

It looks that how effectively the model predicts negative outcomes. Similar to sensitivity, but from the perspective of undesirable outcomes, is specificity. The specificity value of the proposed model is 91 percent

Specificity = TN / (TN + FP)
= 364 / (364 + 36)
= 0.91

VII. RESULT AND DISCUSSION

The pre-trained model that we employed in this study was imported from Decision tree classifier. More than 1500 rows of data from the dataset were used to pre-train this model. 800 rows from the training set and 800 rows from the validation set comprise the 1600 rows used to train this proposed Decision tree. This suggested model was trained using a hybrid dataset and achieved an accuracy of 94% when compared to Sparse regression model. With a 94% accuracy rate, this model is used with a hybrid dataset. Table provides values for recall, specificity, F1 Score, accuracy, and precision.

Network	Acc	Precision	Recall	Spec	F1
Decision Tree	0.94125	0.9152	0.9725	0.91	0.9430

Values of evaluation metrics

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 4, May 2023

REFERENCES:

[1] Ana Radovanovic´, Ross Koningstein, Ian Schneider, Bokan Chen , Alexandre Duarte, Binz Roy, Diyue Xiao, Maya Haridasan, Patrick Hung, Nick Care, Saurav Talukdar, Eric Mullen, Kendal Smith, MariEllen Cottman,and Walfredo Cirne,"Carbon -aware computing for datacenters" in IEEE journal.2021

[2] Avita Katal, Susheela Dahiya, Tanupriya Choudhury,"Energy efficiency in cloud computing data centers: a survey on software technologies",2022.

[3] Samuel Asumadu Sarkodie a, Maruf Yakubu Ahmed a, Thomas Leirvik a, b, c,"Trade volume affects bitcoin energy consumption and carbon footprint",2022.

[4] Xinlu Sun a, Zhifu Mia, Andrew Sudmant b, D'Maris Coffmana, Pu Yanga, Richard Wood c,"Using crowdsourced data to estimate the carbon footprints of global cities",2022.

[5] Cesario Tavares a,, Xincheng Wang b, Sajib Saha c, Zachary Grasley d,"Machine learning-based miX design tools to minimize carbon footprint and cost of UHPC. Part 1: Efficient data collection and modeling",2022.

[6] G. Liu, H. Chen, and H. Huang, "Sparse shrunk additive models," in Proc. Int. Conf. Mach. Learn., 2020, pp. 6194–6204.

[7] S. Zeng, B. Zhang, J. Gou, and Y. Xu, "Regularization on augmented data to diversify sparse representation for robust image classification," IEEE Trans. Cybern., early access, Oct. 21, 2020, doi: 10.1109/TCYB.2020.3025757.

[8] Y.-P. Zhao, L. Chen, and C. L. P. Chen, "Laplacian regularized nonnegative representation for clustering and dimensionality reduction," IEEE

[9] Y. Wang, Y. Y. Tang, L. Li, and H. Chen, "Modal regression-based atomic representation for robust face recognition and reconstruction," IEEE Trans. Cybern., vol. 50, no. 10, pp. 4393–4405, Oct. 2020.

[10] D. Cai, X. He, and J. Han, "Spectral regression: A unified approach for sparse subspace learning," in Proc. 7th IEEE Int. Conf. Data Mining

[11] S. Yi, Z. He, Y.-M. Cheung, and W.-S. Chen, "Unified sparse subspace learning via self-contained regression," IEEE Trans. Circuits Syst. Video Technol., vol. 28, no. 10, pp. 2537–2550, Oct. 2018.

[12] J. Zeng, Y. Liu, B. Leng, Z. Xiong, and Y.-M. Cheung, "Dimensionality reduction in multiple ordinal regression," IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 9, pp. 4088–4101, Sep. 2018.

[13] Y.-M. Cheung and H. Zeng, "Local kernel regression score for selecting features of high-dimensional data," IEEE Trans. Knowl. Data Eng., vol. 21, no. 12, pp. 1798–1802, Dec. 2009