**CO2 EMISSION**

**By**

2103A52005 B. SHASHANKA

2103A52021 K. MANISWAR

2103A52020 K. RAHUL

**CONTENTS**

1. Abstract
2. Introduction
3. Literature Survey
4. Proposed work and Methodology
5. Experimental Work
6. Results
7. Conclusion and Future Scope
8. References

**ABSTRACT**

Personal vehicles play a vital role in emitting carbon dioxide in the atmosphere which causes global warming. In our work, we have taken certain features from the vehicles that affect the emission of CO2 and how the features will affect the quality of the vehicles. Here we have used machine learning models like knn, XGBoost, and random forest to find the root mean square error and the accuracy for the trained and the tested data. Based on the Canadian official data, our work quantifies such CO2 emissions for different types of model cars. By applying the machine learning models we got an accuracy of 97% and a less root mean square error of 0.919 and we predicted which type of feature should be used in certain types of vehicles so that the vehicle emits less carbon dioxide into the atmosphere.

**INTRODUCTION**

Global warming issues have become a universal problem for all nations nowadays. The Intergovernmental Panel on Climate Change (IPCC) reported that scientists were more than 95% certain that most of the global warming is caused by increasing concentrations of greenhouse gasses and other human (anthropogenic) activities [1]. That balance between the earth and atmosphere is affected by an increase in acid gasses charcoal or carbon dioxide (CO2), methane (CH4), nitrous oxide (N2O), hydrofluorocarbons (HFC), and perfluorocarbons (PFC) more commonly known by greenhouse gasses. In particular, CO2 is a major cause of Global Warming [2]. About eight billion tons per year of carbon in the form of CO2 is emitted globally through burning fossil fuels for transport and the production of heat and electricity around the world [3]. The emission of carbon dioxide is the result of the combustion of water (H2O) and carbon monoxide gas (CO) also called carbon dioxide (CO2) is a greenhouse gas. There’s a concept as a reference in the measurement of CO2 emission, namely carbon footprint [4]. This project addresses the urgent need to combat climate change by focusing on one of the largest carbon dioxide (CO2) emissions sources: transportation. Vehicles, particularly those powered by internal combustion engines, are major contributors to greenhouse gas emissions Due to the increase in the population, the public and personal transport increased on the roads, which contributes to the increase in carbon in the atmosphere thus deteriorating the air quality. The air polluted with carbon dioxide has harmful effects on human health. It causes serious health problems like inflammation, kidney and bone problems, inflammation and reduced cognitive performance. High levels of CO2 replace the oxygen carried by hemoglobin, which results in lower levels of oxygen that mostly cause headaches, dizziness, restlessness, tingling feeling, difficulty in breathing, sweating, tiredness, increased heart rate, elevated blood pressure, coma, convulsions. So, by using a machine learning model for the prediction of the rate of carbon emission, we can manage air pollution and the public health conditions affected by the high levels of carbon dioxide in the air.[15] Carbon footprint has become a widely used term and concept in the public debate on responsibility and abatement action against the threat of global climate change. It had a tremendous increase in public appearance over the last few months and years and is now a buzzword widely used across the media, the government, and the business world [5]. It is a measure of the total amount of carbon dioxide released into the atmosphere in a given time frame that is directly or indirectly caused by an activity to provide a service or product [6].

**LITERATURE SURVEY**

TABLE

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **S.NO** | **MODEL USED** | **PARAMETERS USED** | **MERITS** | **DEMERITS** | **FUTURE SCOPE** | **AUTHOR** |
| [1] | =>Lasso =>Random forest  => Support Vector regression | Accuracy=75.63% | => Innovative approach  => Data-Driven insights  => Predictive models  => Real-Time analysis  => Interdisciplinary collaboration | => Model selection  => Feature engineering  => Hyperparameter tuning  => Validation techniques  => Temporal considerations | => Carbon emission reduction strategies  => Extreme event prediction  => Climate finance and policy design  => Real-time monitoring and early warning systems | Harvey Zheng |
| [2] | =>Screening-based approach  =>Genetic algorithm  =>Genaretive model-based approach | Accuracy=74.52% | => Accelerated materials discovery  => High-dimensional data analysis  => Feature extraction and selection  => Customization for applications  => Reduced experimental costs | => Data quality and quantity  => Data imbalance  => Limited chemical insight  => Ethical considerations | => Advanced predictive models  => Multi-objective optimization  => Materials lifecycle analysis  => Real-world deployment | Geun Ho Gu , Juhwan Noh , Inkyung Kim  and Yousung Jung |
| [3] | => Artificial neural network  => Random forest | Accuracy=87% | => Climate change validation  => Hydrological understanding  => Long-term projections | => Spatial variability  => Extrapolation limitations  => Limited attribution  => Confounding variables | => Integration with ecosystem models  => Remote sensing data  => Hydrological modeling  => Human-climate interactions  => Climate change mitigation strategies | Ren Wang,Longhui Li,Pierre Gentine,Yao Zhang,Jianyao Chen,Xingwei Chen, Lijuan Chen,Liang Ning,Linwang Yuanand Guonian Lu |
| [4] | => Q-learning approach | Accuracy=82% | => Optimized design  => Data-driven decisions  => Adaptive strategies  => Policy support  => Reduced life-cycle impact | => Complex modeling  => Data requirements  => Policy generalization  => Continuous learning  => Model interpretability | => Lifecycle analysis  => Localized solutions  => Urban planning integration  => Policy recommendations  => Economic considerations | Sophie Renard, Benjamin Corbett, Omar Swei |
| [5] | => Support Vector Machine  => LASSO Regression  => Multiple Linear Regression  =>XG Boost Regression Algorithm | Accuracy=91% | => Environmental assessment  => Data-driven insights  => Simplified modeling  => Interpretability | => Limited complexity  => Assumption violations  => Collinearity  => Generalization issues | => Enhanced predictive models  => Integration of advanced techniques  => Material innovation  => Building-level analysis | [Ibrahim Tajuddeen](https://sciprofiles.com/profile/2633798), [Ibrahim Tajuddeen](https://sciprofiles.com/profile/2633798), [Mina Jafari](https://sciprofiles.com/profile/author/T0MzbFZXN3VLeWxKVEZ1cEVweldPdE02ZzBRMHFRalJ4R0VoNkE5Y1JaND0=) |
| [6] | => Artificial neural network  => Ridge regression | Accuracy=98.3% | => Process understanding  => Quantitative predictions  => Model flexibility  => Feature selection | => Limited physical understanding  => Bias and generalization  => Data quality | => Real-time forecasting  => Cross-disciplinary collaboration  => Data assimilation  => Regional and global scale applications  => Hybrid approaches | [D. J. Gagne](https://agupubs.onlinelibrary.wiley.com/authored-by/Gagne/D.+J.), [C.-C. Chen](https://agupubs.onlinelibrary.wiley.com/authored-by/Chen/C.%E2%80%90C.), [M. W. Christensen](https://agupubs.onlinelibrary.wiley.com/authored-by/Christensen/M.+W.), [Z. J. Lebo](https://agupubs.onlinelibrary.wiley.com/authored-by/Lebo/Z.+J.), [H. Morrison](https://agupubs.onlinelibrary.wiley.com/authored-by/Morrison/H.), [G. Gantos](https://agupubs.onlinelibrary.wiley.com/authored-by/Gantos/G.) |
| [7] | =>SVM  =>Lasso  => splines | Accuracy=89.5% | => [reduces the computational cost](https://www.mdpi.com/2075-5309/10/8/139)  =>Identifies energy patterns  => [predicts energy consumption and carbon emission of residential buildings](https://link.springer.com/chapter/10.1007/978-981-16-7160-9_19) | => Limited causality  => Overfitting  => Data availability | => Economic analysis  => Human-building interaction  => Smart grid integration  => Urban microclimate modelling  => Dynamic occupancy modeling | Wei Tiana,b, Elisabete A. Silvac , Ruchi Choudhary |
| [8] | =>ANN  =>  Runge–Kutta technique | Accuracy=85.9% | =>fisheries management  =>carbon footprint reduction  =>biodiversity conservation  =>food security  =>human health | => Causality vs. correlation  => Human expertise  => Scaling challenges  => Computational resources | => Climate change mitigation  => Remote sensing  => Ocean acidification  => Multi-species Modelling | [Hosam Alhakami](https://sciprofiles.com/profile/1045833), [Abdullah Baz](https://sciprofiles.com/profile/884381) |
| [9] | =>gradient boosting regression  =>ANN  =>SVM | Accuracy=87% | =>accuracy and efficiency  => Develop a rapid predictive model  => Locate spatial hot spots | =>require large sample sizes  =>lack interpretability  =>not capture all the uncertainties | => Extend to other corps  => Incorporate more features  => Evaluate the trade-offs  =>tests different scenarios | [Xiaobo Xue, Romeiko](https://sciprofiles.com/profile/152254) ,  [Yulei ,Pang](https://sciprofiles.com/profile/author/Z01wa2F2ZDVhTVE5Nm1GT3pBWFoxWmp0VjFXdnFSb0dQM1hsdVdUZStaND0=), [Xuesong Zhang](https://sciprofiles.com/profile/246352) |
| [10] | =>linear regression  =>ANN | Accuracy=98.99% | =>climate prediction  =>predicts extreme weather events  =>reliable statistical model | =>large amount of data required  =>climate models  => Public and policymakers | =>adaption strategies  =>support mitigation  => Predicts impact of globally warmed  =>improves the reliability of climate models | [M. Purushotham Reddy](https://ieeexplore.ieee.org/author/37089201850), [A. Aneesh](https://ieeexplore.ieee.org/author/37089202697), [K. Praneetha](https://ieeexplore.ieee.org/author/37089852132). [S. Vijay](https://ieeexplore.ieee.org/author/37089203781) |
| [11] | =>linear regression  => neural networks | Accuracy=90% | =>broader effort  => Awareness and accountability  => Reduce  carbon footprint | => Approximation and assumptions  =>Root cause  => Not provide concrete solutions | => Research and innovation  => Collaboration and communication  => Sustainable and ethical development | [Alexandra Lacoste](https://arxiv.org/search/cs?searchtype=author&query=Lacoste%2C+A), [Alexandra Luccioni](https://arxiv.org/search/cs?searchtype=author&query=Luccioni%2C+A), [Victor Schmidt](https://arxiv.org/search/cs?searchtype=author&query=Schmidt%2C+V), [Thomas Dandres](https://arxiv.org/search/cs?searchtype=author&query=Dandres%2C+T) |
| [12] | =>neural networks  =>SVM | Accuracy=88.86% | =>Uses real-time data  =>Compares the solution  =>Accuracy | => Scalability is not shown  =>Neglects more factors  =>Environmental conditions | => Improving data quality  =>security  => Extending the model to other cities  =>evaluating the model | [Usha Mahalingam](https://ieeexplore.ieee.org/author/37531093300), [Kirthiga Elangovan](https://ieeexplore.ieee.org/author/37088337170), [Himanshu Dobhal](https://ieeexplore.ieee.org/author/37088335495), [Chocko Valliappa](https://ieeexplore.ieee.org/author/37088338099), [Sindhu Shrestha](https://ieeexplore.ieee.org/author/37088337859) |
| [13] | =>Ordinary Least Squares regression =>Decision Tree  =>Deep Neural Networks | Accuracy=81% | =>Scientific advancement  =>Informed decision-making  =>Climate justice advocacy  =>Prediction uncertainty quantification | =>Computational demands  => Overfitting risks  =>Limited socio-economic detail  =>Transferability issues | =>Fine-tuning ensemble techniques  =>Extreme event attribution  =>Multi-scenario exploration  =>Feedback mechanisms | Piaoyin Zhang, Jianzhong Lu, Xiaoling Chen |
| [14] | =>deep learning-based approaches  =>U-Net  =>Mask-CNN | Accuracy=87.97% | =>Environmental monitoring and management  =>flexible tools  =>implements different augmentation techniques | =>Misinterpretation of patterns  => Data noise  => Change detection sensitivity  => Dependency on training data | => Fine-grained analysis  => Multi-sensor fusion  => Dynamic change detection  => Hydrological modeling  => Education and outreach | [Rajdeep Chatterjee](file:///C:\Users\SHASHANKA\OneDrive\Desktop\shashi%20dafe.docx#auth-Rajdeep-Chatterjee-Aff1),  [Ankita Chatterjee](file:///C:\Users\SHASHANKA\OneDrive\Desktop\shashi%20dafe.docx#auth-Ankita-Chatterjee-Aff2) , [SK Hafizul Islam](file:///C:\Users\SHASHANKA\OneDrive\Desktop\shashi%20dafe.docx#auth-SK_Hafizul-Islam-Aff3) |
| [15] | =>knn  =>random forest | Accuracy=>67.568% | => Climate change ontologies  => Issues related to global warming  => Enhance the conversational capabilities | => Persuade strongly biased users  => Handle complex or ambiguous questions  => Counter all the false claims | => Handle complex or ambiguous questions  => More effective communication strategies | Diana Toniuc, Adrian Groza. |
| [16] | => k-Nearest Neighbour (KNN)  => Support Vector Machine (SVM)  => Adaptive Boosting (AdB)  => Random Forest | Accuracy=76.34% | => Daily temperature records  => optimal parameters  => high accuracy | => Validation and robustness  => Model selection  => Lack of domain expertise  => Climate non-stationarity | => Global coverage  => Online Learning  => Real-time forecasting  => Uncertainty estimation | Babak Azari , Khairul Hassan1 , Joel Pierce1 , Saman Ebrahimi2 |
| [17] | => linear regression =>Pearson correlation | Accuracy=95.24% | => Reliability and accuracy of the analysis  => enhances the understanding  => provides insights | => only focuses on maximum temperature  => does not consider other factors that may influence | => Increased frequency and intensity of extreme weather  => Adverse health impacts due to heat stress  => Loss of biodiversity | S. K. Dash, S. K. Mishra, A. K. Singh |
| [18] | => Lasso regression  => Support vector regression (SVR)  => Multi-regression tree  => Pearson correlation coefficient | Accuracy=68.57% | => Capture the complex and nonlinear relationships  => Provide accurate and reliable forecasts  => handle large and heterogeneous data sets | => require a lot of computational resources  => suffer from overfitting or underfitting problems  => depend on the quality and availability | => improve their performance and accuracy  => enhance their interpretability  => Support decision-making and policy-making | D. Deva Hema, Anirban Pal, Vineet Loyer, Rajeev Gaurav |
| [19] | => Principal component analysis  => Principal component analysis  => Random forest | Accuracy=85.6574% | => Dentify the key drivers and factors  => Quantify the impact and cost of global warming on human lives  => Support decision-making and policy-making by providing evidence-based and data-driven recommendations | => lack of interpretability and transparency  => can face ethical and social challenges in dealing with sensitive  => depend on the quality and availability | => can incorporate more data and variables  => transparency by using explainable artificial intelligence (XAI)  => Empower the stakeholders | [C. Orsenigo](file:///C:\Users\SHASHANKA\OneDrive\Desktop\shashi%20dafe.docx#auth-C_-Orsenigo-Aff1)  [C. Vercellis](file:///C:\Users\SHASHANKA\OneDrive\Desktop\shashi%20dafe.docx#auth-C_-Vercellis-Aff1) |
| [20] | => Linear regression  => Random forest | Accuracy=94.356% | =>Can improve the representation and accuracy  =>Can provide insights and understanding  =>Can reduce the computational cost and complexity | => Can lack interpretability and transparency  => Can face ethical and social challenges  => Depend on the quality and availability of the training data | => Engage and empower the stakeholders  => Enhance their interpretability and transparency  => Incorporate more data and variables | Peer Nowack1,2,3, Peter Braesicke4, Joanna Haigh1,2, Nathan Luke Abraham5,6, John Pyle5,6 and Apostolos Voulgarakis2 |

SUMMARY:

The relation with temperature and other factors such as concentrations of carbon dioxide, nitrous oxide, and methane[1]. Demonstrate applications of machine learning methods for theoretical approaches in key renewable energy technologies [2]. Estimates of change in global land evapotranspiration (ET)and consistent with the idea that ET would increase with climate warming[3]. Subsequently tests the performance of the Q-learning approach from three representative case studies with varying traffic volumes [4] selection of effective low-energy-intensive thermal insulation, thus mitigating environmental impacts [5]. get simulations that perform as the detailed model with the computational cost of the control simulation [6]. optimizing the energy performance and sustainability of buildings [7]. quantitative analysis of the impact of global warming on marine ecosystems, which assures changes in ocean temperature and acidity [8] forecasting spatially different life cycle global warming and eutrophication impacts from corn production [9] useful insights and forecasts of the global warming phenomenon and its impacts.[10] machine learning practitioners should be aware of the carbon footprint of their experiments and adopt best practices to reduce it.[11] air quality prediction for smart cities is a crucial task for improving the health and well-being of urban residents[12]. Extreme changes under different socio-economic pathways [13] Detect and monitor the changes in shape and size of the water bodies over time [14]. agent can effectively engage users in a dialogue about climate change and provide arguments based on facts and logic.[15] air pressure is the most important feature for temperature prediction.[16] positive correlation between maximum temperature and solar radiation, and a negative correlation between maximum temperature and rainfall, humidity, and wind speed.[17] annual global warming prediction, from previously measured values over India.[18] marginal damage cost of carbon dioxide emissions and suggests that it is higher than the current carbon price.[19] algorithm on preindustrial and abrupt 4x co2 simulations and compares it with a standard ozone parameterization[20]

PROBLEM STATEMENT:

* Carbon dioxide is a major component of greenhouse gases responsible for global warming. The objective for predict CO2 emissions in different types of vehicles and reduce carbon dioxide levels.
* [Predictive artificial intelligence can forecast future emissions across a company’s CO2 emissions, about current reduction efforts, new carbon reduction methodologies, and future demand.](https://www.bcg.com/publications/2021/ai-to-reduce-carbon-emissions)
* To identify some features that are causing more carbon dioxide emissions.
* And to reduce their usage in the future.

**PROPOSED WORK AND METHODOLOGY**

Procedure to solve the given problem:

A diagram of a structure

Description automatically generated

Fig 1

For predicting the CO2 emissions from the vehicles we use different K-nearest neighbor, XGBoost, and Random forest models for predicting the emissions of the data set

Steps followed during model implementation

* Importing the raw data
* Extracting the features from the data
* Describing the data set
* Perform feature engineering
* Splitting data
* Extracting the score from methods

KNN:

We used the k-nearest neighbour method to solve the data set. The K-nearest neighbour method is a key component of machine learning. It is based on the technique of supervised machine learning. The k-nearest neighbour strategy implies that the new case/data and previous cases are comparable and it assigns the new case to the category that is closest to the previous categories. The k-nearest neighbour algorithm keeps all available data and categorizes categorized data points depending on how comparable they are to earlier classified data. This implies that fresh data may be rapidly categorized into a well-defined category using the k-nearest neighbour technique. While the k-nearest neighbour technique applies to both regression and classification issues, it is often employed for classification problems. This study uses the k-nearest neighbour classifier, which is one of the most frequently used classification algorithms in machine learning. The k-nearest neighbor approach is a nonparametric method for classifying data. This classifier classifies objects according to their proximity and “k” closest neighbours. It is concerned with the immediate surroundings of the item rather than with the required data distribution.

XGBoost:

**Gradient boosting** refers to a class of ensemble machine learning algorithms that can be used for classification or regression predictive modelling problems. Ensembles are constructed from decision tree models. Trees are added one at a time to the ensemble and fit to correct the prediction errors made by prior models. This is a type of ensemble machine-learning model referred to as boosting. Models have been fitted using any arbitrary differentiable loss function and gradient descent optimization algorithm. This gives the technique its name, “gradient boosting,” as the loss gradient is minimized as the model is fit, much like an artificial neural network.

Random Forest:

Random forests or random decision forests are an ensemble learning method for classification, regression, and other tasks that operate by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most of the trees. For regression tasks, the mean or average prediction of the individual trees will be returned. A random forest is a collection of Decision Trees, Each Tree independently makes a prediction, and the values are then averaged/Max voted to arrive at the final value. The strength of this model lies in creating different trees with different sub-features from the different features. The Features selected for each tree are at random, so the trees do not get deep and are focused only on the set of features.

**EXPERIMENTAL WORK**

The experimental work undertaken for the research is as follows

Setting up an infrastructure

google colab notebook

importing necessary libraries like NumPy, pandas, matplotlib, warnings, seaborn, and sklearn

reading the data using the Pandas library

defining data set and describing about it

making different types of graphs like heat map and bar graph using matplotlib and seaborn libraries

descriptive statistics

scaling data into testing set and training set

finding the accuracy for the data using knn, random forest, and XGBoost simply by importing the particular libraries from sklearn library

and lastly printing the scores in a perfect table format.

# Description of The Data:

# This dataset captures the details of CO2 emissions by a vehicle that can differ with the different features. The dataset is taken from the Kaggle website. This is a collected interpretation. This contains data around a period of 7 years. There are total 7385 rows and 12 columns. There are few abbreviations that are used to describe the features.

# The data used in the project is collected from the Canadian Government's Official website

## About Data:

### It includes the following attributes:

* *Make*: Car brands
* *Model*: Model of the car.
* *Vehicle\_class:*Car body type.
* *Engine\_size:* Size of the engine in liters.
* *Cylinders:* No of cylinders.
* *Transmission:* Type of transmission
* *Fuel\_type:* Type of fuel used by the car.
* *Fuel\_consumption\_city:* Fuel consumption in city ratings in liters per 100 kilometers.
* *Fuel\_consumption\_hwy:* Fuel consumption on highway ratings in liters per 100 kilometers.
* *Fuel\_consumption\_comb(l/100km)*: Combined fuel consumption rating (city and highway) in liters per 100 kilometers.
* *Fuel\_consumption\_comb(mpg):* Combined fuel consumption rating in miles per gallon (mpg).
* *Co2\_emissions*: Tailpipe emissions of carbon dioxide for combined city and highway driving, in grams per kilometer.

# Tools and Technologies:

# Developed using various tools and technologies, including:

* Python programming language
* Libraries like NumPy, Matplotlib, pandas and sklearn
* Linear Regression, Random Forest, K-Nearest Neighbors (KNN), and XG BOOST models for predicting the score

Key words and abbreviations:

1. *Model*

4WD/4X4 Four-wheel Drive

AWD-All-wheel Dive

FFV-Flexible-Fuel Vehicle

SWB-Short Wheel Base

LWB-Long Wheel Base

EWB-Extended wheel base

*2. Transmission*

A-Automatic

AM-Automated Manual

AS-Automatic with Select shift

AV-Continuously Variable

M-Manual

3-10 Number of gears

*3. Fuel type*

X-Regular gasoline

Z-Premium gasoline

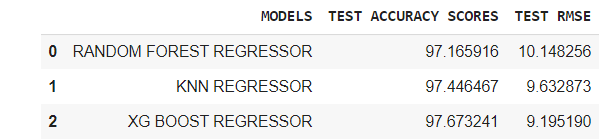
D-Diesel

E-Ethanol (E85)

N-Natural gas

**RESULTS**

The result table given below shows the test accuracy scores for the respective methodologies along with the RMSE score. Where RMSE refers to root mean square error. As we can see in below table with each methodology the accuracy scores have reached to the 97.00000.

****

The heat map given below shows the correlation between the features, coming to our work with the help of this graph we can able to predict the relation among the features and which feature from the dataset has more effect in the emission of the co2 the part which is highlighted shows the related results. Heat map is taken with the help of the seaborn library.**A screenshot of a computer

Description automatically generated**

Fig 2

A screenshot of a computer

Description automatically generated

Fig 3

The above table shows the average count, mean, standard deviation, minimum, 25 percentile,50 percentile,75 percentile, and maximum of the respective features.

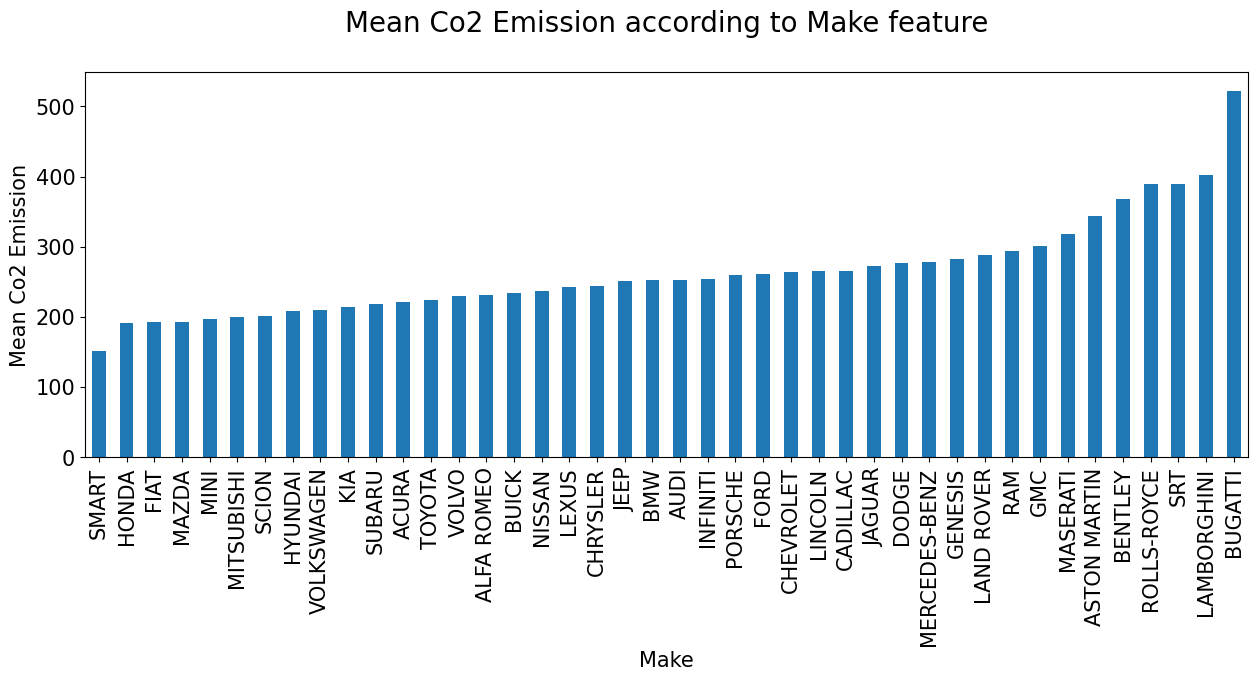
Fig3(a)

Fig3(b) A graph of a graph showing the number of vehicles

Description automatically generated

The graph (a) given above shows about the mean CO2 emission with respect to the make feature which type of vehicle emits more amount of co2 and which is less according to the graph smart vehicle emits less CO2 and Bugatti emits more amount of co2.

Graph (b) given above shows about the mean CO2 emission concerning the vehicle class feature.

Fig3(c)A graph of a graph

Description automatically generated

Fig3(d)A graph of a graph

Description automatically generated

The graph (c) represents the mean of co2 emissions to the engine size means how the engine size will affect the co2 emission from the graph we came to know that the smaller the size the lower the emission and the bigger size more emission. That is the CO2 emissions are directly proportional to the engine size.

The graph (d) represents the co2 mean emission to the cylinder feature

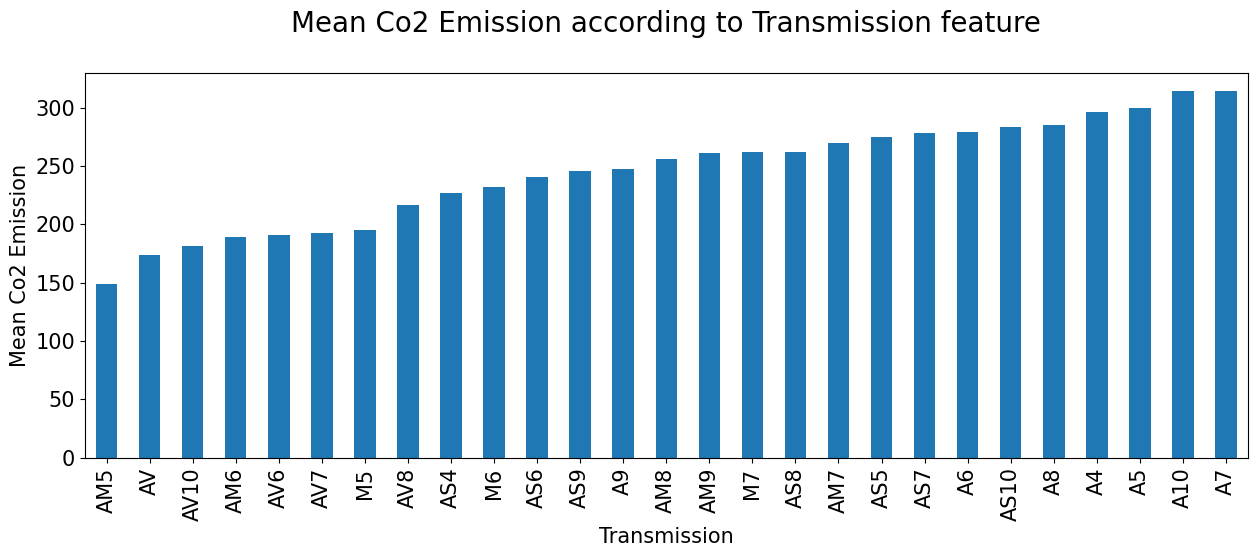
Fig3(e)

Fig3(f)A graph of blue rectangles

Description automatically generated

Graph (e) shows the relation between transmission to the CO2 emissions which type of transmission shows mean CO2 emission.

Graph (f) represents which type of fuel emits more CO2 according to the graph we can say that the E fuel type emits more CO2

*Github link:*

https://github.com/Shashankabasani/dafe/blob/main/dafe\_c02\_emmitions.ipynb

**CONCLUSION AND FUTURE SCOPE**

We have already discussed all the important features of the datasets and their visualization in the above sections. But in order to conclude our report we would choose XGBoost regressor. As we are satisfied by seeing how closely we have predicted the root causes of increasing CO2 emissions and it has the lowest Root Mean Square error of 0.919.

There are still many technical indicators and feature variables that we have not included in our project, maybe there are some other indicators that we haven’t explored that would perform better.

There are lots of Machine Learning algorithms that we haven’t tried and maybe a neural network or gradient boost would perform better than our solution.

Last but not least we have used data from the Canadian environment if we increase the data, we think the performance of our solution models may be improved.

**REFERENCES**

*Dataset link:*

https://www.kaggle.com/datasets/bhuviranga/co2-emissions

[1] Analysis of Global Warming Using Machine Learning: Zheng, H. (2018). Analysis of global warming using machine learning. Computational Water, Energy, and Environmental Engineering, 7(03), 127

[2] Gu, G. H., Noh, J., Kim, I., & Jung, Y. (2019). Machine learning for renewable energy materials. Journal of Materials Chemistry A, 7(29), 17096-17117.

[3] Wang, R., Li, L., Gentine, P., Zhang, Y., Chen, J., Chen, X., ... & Lü, G. (2022). The recent increase in observation-derived land evapotranspiration is due to global warming. Environmental Research Letters, 17(2), 024020.

[4] Renard, S., Corbett, B., & Swei, O. (2021). Minimizing the global warming impact of pavement infrastructure through reinforcement learning. Resources, conservation and recycling, 167, 105240.

[5] Tajuddeen, I., Sajjadian, S. M., & Jafari, M. (2023). Regression models for predicting the global warming potential of thermal insulation materials. Buildings, 13(1), 171.

[6] Gettelman, A., Gagne, D. J., Chen, C. C., Christensen, M. W., Lebo, Z. J., Morrison, H., & Gantos, G. (2021). Machine learning the warm rain process. Journal of Advances in Modeling Earth Systems, 13(2), e2020MS002268.

[7] Wei, L., Tian, W., Silva, E. A., Choudhary, R., Meng, Q., & Yang, S. (2015). Comparative study on machine learning for urban building energy analysis. Procedia Engineering, 121, 285-292.

[8] Alhakami, H., Kamal, M., Sulaiman, M., Alhakami, W., & Baz, A. (2022). A machine learning strategy for the quantitative analysis of the global warming impact on marine ecosystems. Symmetry, 14(10), 2023.

[9] Romeiko, X. X., Guo, Z., Pang, Y., Lee, E. K., & Zhang, X. (2020). Comparing machine learning approaches for predicting spatially explicit life cycle global warming and eutrophication impacts from corn production. Sustainability, 12(4), 1481.

[10] Reddy, M. P., Aneesh, A., Praneetha, K., & Vijay, S. (2021, November). Global Warming Analysis and Prediction Using Data Science. In 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) (pp. 1055-1059). IEEE.

[11] Lacoste, A., Luccioni, A., Schmidt, V., & Dandres, T. (2019). Quantifying the carbon emissions of machine learning. arXiv preprint arXiv:1910.09700.

[12] Mahalingam, U., Elangovan, K., Dobhal, H., Valliappa, C., Shrestha, S., & Kedam, G. (2019, March). A machine learning model for air quality prediction for smart cities. In 2019 International Conference on Wireless Communications Signal Processing and networking (WiSPNET) (pp. 452-457). IEEE.

[13] Zhang, P., Lu, J., & Chen, X. (2022). Machine-learning ensembled CMIP6 projection reveals socio-economic pathways will aggravate global warming and precipitation extreme. Hydrology and Earth System Sciences Discussions, 2022, 1-40.

[14] Chatterjee, R., Chatterjee, A., & Islam, S. H. (2022). Deep learning techniques for observing the impact of global warming from satellite images of water bodies. Multimedia Tools and Applications, 81(5), 6115-6130.

[15] Toniuc, D., & Groza, A. (2017, September). Clime Bot: An argumentative agent for climate change. In 2017 13th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP) (pp. 63-70). IEEE.

[16] Azari, B., Hassan, K., Pierce, J., & Ebrahimi, S. (2022). Evaluation of machine learning methods application in temperature prediction. Environ Eng, 8, 1-12.

[17] Thirumalai, C., Saikrishna, G., Raju, C. S., & Senthilkumar, M. (2017, May). Analysis of global warming in India over maximum temperature using Pearson and machine learning. In 2017 International Conference on Trends in Electronics and Informatics (ICEI) (pp. 496-499). IEEE.

[18] Hema, D. D., Pal, A., Loyer, V., & Gaurav, R. (2019). Global Warming Prediction in India Using Machine Learning. International Journal of Engineering and Advanced Technology (IJEAT), 9(1).

[19] Orsenigo, C., & Vercellis, C. (2018). Anthropogenic influence on global warming for effective cost-benefit analysis: a machine learning perspective. Economia e Politica Industriale, 45(3), 425-442.

[20] Nowack, P., Braesicke, P., Haigh, J., Abraham, N. L., Pyle, J., & Voulgarakis, A. (2018). Using machine learning to build temperature-based ozone parameterizations for climate sensitivity simulations. Environmental Research Letters, 13(10), 104016.