Mesopotamian journal of Computer Science
Vol. **(2023)**, 2023, **pp**. 53-63
DOI: https://doi.org/10.58496/MJCSC/2023/009; ISSN: 2958-6631
https://mesopotamian.press/journals/index.php/cs

Research Article

Predicting Carbon Dioxide Emissions with the Orange Application: An Empirical Analysis

Israa Ezzat Salem[1,2,*], Alaa Wagih Abdulqader[1],

[1] *Computer Techniques Engineering Department, Baghdad College of Economic Sciences University, Baghdad, Iraq*

[2] *Computer Science Department, College of Science, Mustansiriyah University, Baghdad, Iraq*

ARTICLE INFO

Article History
Received 26 Jan 2023
Accepted 09 Mar 2023
Published 30 Mar 2023

Keywords
Machine learning
Data Mining
Orange Application
Predicting
CO2

ABSTRACT

The effects of climate change, such as droughts, storms, and extreme weather, are increasingly being felt around the world. Greenhouse gases are the primary contributors to climate change, with carbon dioxide (CO2) being the most significant. In fact, CO2 accounts for a significant percentage of all greenhouse gas emissions. As a result, reducing CO2 emissions has become a critical priority for mitigating the impacts of climate change and preserving our planet for future generations. Based on simulation and data mining technologies that use historical data, CO2 is expected to continue to rise. Around the world, 80% of CO2 emissions come from burning fossil fuels, mostly in the automotive or manufacturing industries. Governments have created policies to control CO2 emissions by focusing them on either consumers or manufacturers, in both developed and developing nations. Within the scope of this project, an investigation of vehicle emissions will be carried out using various attributes included within the vehicle dataset, as well as the use of many data mining techniques via the utilization of an orange application. The practical program is an example of organization, and the example will be about cars, exploring data, and figuring out how much gas will be needed. CO2 is taken away from cars, and we will use the CARS.csv file, which has data for a group of car types. It has a table with 36 records that shows the model, weight, and amount of carbon dioxide based on the car's size and weight.

1. INTRODUCTION

Climate change has become a pressing topic of contemporary discourse due to its potential impact on future generations [1]. Researchers from around the world have conducted extensive studies on the topic, revealing that the effects of climate change are already being felt in the form of extreme weather conditions and other environmental challenges [2]. As a result, there is a growing sense of urgency to take action to mitigate these impacts and preserve the planet for future generations. Hot and dry conditions, together with recent storms, have dominated news broadcasts. Given the current situation, it is imperative that scientists and policymakers establish standards for protecting the planet for the sake of future generations. The primary reasons leading to climate change must be identified for solutions offering sustainability to be successful. Greenhouse gases, specifically carbon dioxide (CO2), have been identified as the primary climatic contributors by several studies [3]. Machine learning and data mining are both related to the analysis of data. They both involve using algorithms to uncover patterns and insights from data. However, there are some critical differences between the two: Machine learning is a subset of artificial intelligence that involves developing algorithms that enable machines to learn from data and make predictions or decisions. It is typically operated for tasks such as classification, regression, and clustering. At the same time, data mining is extracting useful information or knowledge from large amounts of data. It involves applying algorithms to identify data patterns, trends, and relationships. Data mining can be used for various tasks, including fraud detection, marketing analysis, and scientific research. Moreover, machine learning is generally more focused on building predictive models, while data mining is concentrated on discovering insights from data. Both machine learning and data mining require a strong foundation in statistics, mathematics, and computer science. Both are becoming increasingly important in a wide range of industries, including healthcare, finance, and retail.

This study uses global datasets, and data mining to extract information from the databases that may be used in the study. The main data mining issues affecting CO2 emissions and climate change in general are explored in the sections below. The paper provides a summary of current initiatives to apply data mining to this area with several examples of studies

*Corresponding author. Email: farg.israa@uomustansiriyah.edu.iq

pertaining to climate change. This paper starts with a general summary of earlier research on data mining-related topics before addressing application of data mining for predicting climate change and CO2 emissions by using orange application.

2. LITERATURE REVIEW

In South Korea, a benchmark for CO2 emissions was established using data mining techniques and a decision tree algorithm [4]. Jeslet and Jeevanandham [5] also used machine learning methods to anticipate diverse weather patterns, including the artificial neural network (ANN) and decision tree. The research included analyzing factors such as wind speed, air temperature, amount of rain, and water evaporation. The team built a meteorological data model and utilized it to train a classifier. The study demonstrated strong performance in predicting weather occurrences and climate change forecasts in their area, indicating the potential for machine learning to be a valuable tool in climate research. In a study conducted by Somu et al. in 2020 [6], two machine learning methods, artificial neural network (ANN) and decision tree, were employed to anticipate various weather patterns. The study utilized factors such as wind speed, air temperature, rainfall, and water evaporation to build a meteorological data model, which was used to train the classifier. The study showed good performance in predicting weather occurrences and climate change forecasts in the area. Farhate et al. [7] aimed to predict the quantity of CO2 emissions from soil caused by crop management in Brazilian sugarcane fields using data mining and predictive analytic tools. The study demonstrated great performance and accuracy in forecasting CO2 emissions while incurring lower computational costs, using the multilayer perceptron classifier and logistic regression bagging classifier. The COVID-19 pandemic, officially declared a global pandemic by the World Health Organization on March 11, 2020, has had a profound impact on many aspects of our lives, including CO2 emissions [8]. The pandemic led to widespread lockdowns, travel restrictions, and reduced economic activity, resulting in a significant decline in global emissions in 2020. This unprecedented reduction in emissions has demonstrated the potential for immediate and widespread action to mitigate the impacts of climate change. According to a study, there was a 1% rise in CO2 emissions by the year 2019 [9]. However, measurements revealed that there was no increase in the proportion of carbon dioxide (CO2) emissions after COVID-19 in late 2020 [10]. In a study by Le Quéré et al. [11], daily worldwide CO2 emissions declined by 17% during the COVID-19 confinement and government choices, falling to 83 MtCO2 from an average of 100 MtCO2 released on a typical day in 2019. The study analyzed and estimated the daily change in CO2 emissions and its ramifications during the COVID-19 confinement based on official data associated with the nature of energy resources, daily activities, and policies. Global CO2 emissions decreased as a result of the lockdown in most countries. It is important to note that the percentage of daily worldwide carbon emissions from surface transport that have decreased by this amount is almost exactly half. Big data applications for climate change are limited. The Hassani et al., 2019 article[12] examines and summarizes the methods for applying big data technology to climate change. The study's main objective is to leverage big data to further future scientific investigations into climate change. Utilizing a data mining approach, researchers examined fuel usage and passenger car emissions to investigate how automobiles contribute to climate change [13]. Research efforts are underway to advance technologies and smart energy infrastructures that enable the widespread adoption of renewable and non-polluting energy sources, while enhancing safety and traffic fluidity [3][14]. The MERGE initiative is preparing the European electrical infrastructure for the increased adoption of electric vehicles. The goal of this initiative is to find solutions that reduce the need to upgrade electric grid infrastructures and power systems, thereby avoiding additional expenses for electric vehicle users. This research project is a key component of the 7th EU Framework Programme and is expected to have a significant impact. Real-time data mining systems, such as IBM's InfoSphere Streams, are finding increasing use in this sector, while emerging standards for wireless emission monitoring in automobiles are being developed. NASA's Advanced Integrated Vehicle Health Monitoring (IVHM) program is also underway to support research and other resources. Commercial systems are also being developed. Agnik's MineFleet system, which won the 2010 Frost & Sullivan North American Enabling Technology of the Year Award in commercial telematics, is a decision support platform that enables modeling, benchmarking, and monitoring of vehicle health, emissions, driver behavior, fuel consumption, and fleet characteristics. MineFleet is a self-sufficient onboard data stream mining platform that performs continuous monitoring of data produced by the vehicle's sensors, analyses the data streams (including manufacturer-improved parameters, generic parameters, and fault codes), and creates predictive models, reports, and warnings. Additionally, MineFleet allows for the sharing of data between vehicles in the fleet.

3. KNOWLEDGE DISCOVERY AND DATA MINING

KDD, or Knowledge Discovery in Databases, refers to the process of extracting useful knowledge from large volumes of data. It involves identifying novel and valuable patterns in data that can help inform decision-making. According to [15], the KDD process is nontrivial and participatory, involving several iterative phases as outlined in Figure 1 [16]. These phases typically include data selection, preprocessing, transformation, mining, evaluation, and interpretation. By following this structured process, KDD can yield insights that are both valid and ultimately understandable to stakeholders.

1- Knowledge of the application domain: This involves understanding the application domain and defining the objectives of the data mining process based on applicable previous knowledge.
2- Extraction of the target data set: This involves selecting a relevant data set or subset of variables to focus on.
3- Data cleansing and pre-processing: This involves performing fundamental operations such as noise reduction and missing data management to cleanse poor-quality data from real-world sources.
4- Data integration: This involves merging diverse, heterogeneous data sources to create a unified data set.
5- Data reduction and projection: This involves selecting useful features to describe the data and employing techniques such as dimensionality reduction or transformation.
6- Selecting the function of data mining: This involves determining the intent of the model derived by the data mining algorithm, such as summarization, classification, regression, clustering, web mining, image retrieval, discovering association rules and functional dependencies, rule extraction, or a combination of these.
7- Selecting the data mining algorithm(s): This involves deciding on the method(s) to be used for finding patterns in the data, including which model and parameters are appropriate.
8- Data mining: This involves searching for patterns of interest in a certain representational form or a group of such representations.
9- Interpretation: This involves interpreting the identified patterns and visualizing the extracted patterns, if possible. One can examine patterns automatically or semiautomatically to determine whether they are interesting or beneficial to the user.
10- Using discovered information: This involves incorporating the knowledge gained through data mining into the performance system and acting on it.

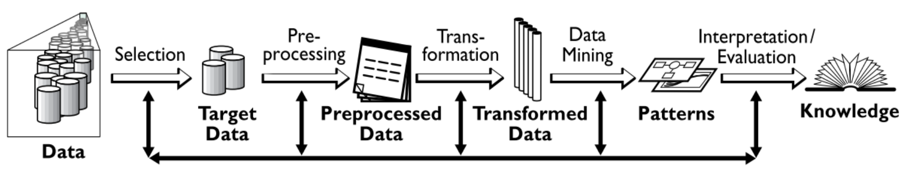


Fig. 1. Overview of the steps constituting the KDD process Data

4. DATA MINING

The data mining approach is a strong tool that allows us to extract valuable insights and knowledge from complex and large datasets. By leveraging machine-learning algorithms, data mining enables us to identify patterns, associations, anomalies, and statistically significant structures in the data that may not be immediately visible to the naked eye. The data mining process consists of several phases, beginning with data preparation. In this phase, the data is cleaned by removing noisy data and filling in missing values, eliminating, or reducing redundant data. Once the data has been prepared, we move on to selecting the exciting data for manipulation. This is followed by formatting and transforming the data into other forms, which is the output of the data preparation model. The second model is named the Data Mining model, which includes different techniques and algorithms for classification, clustering, and association to search and identify interesting patterns in the data. Once the patterns have been identified, they are visualized in different ways to make them more accessible to the user. Finally, data mining ends with an evaluation model, where users can predict, validate, and interpret the results to confirm or prove some results or hypotheses. Altogether, the data mining approach is a powerful tool for making scientific discoveries and gaining fundamental insights from complex datasets. By leveraging machine-learning algorithms, we can extract valuable knowledge from data that may have previously been hidden or difficult to analyze. The different phases of data preparation and mining, along with the evaluation model, provide a comprehensive framework for uncovering new patterns and associations in data. Data mining is a process of discovering patterns and insights in large datasets using various statistical and machine learning techniques. In order to perform data mining, it is important to have access to data in a suitable format. There are several ways to import data into a data mining tool, depending on the format of the data and the tool being used. For example, data can be imported from a CSV or Excel file, a SQL database, a NoSQL database, or from web scraping tools. Many data mining tools have built-in data import functionality, making it easy to load data directly into the tool for analysis. In addition to importing data, it is also important to preprocess and clean the data before performing data mining tasks. This may involve tasks such as removing missing values, normalizing data, and transforming data into a suitable format for analysis. Once the data is cleaned and preprocessed, various data mining techniques can be applied to discover patterns, relationships, and insights in the data.

5. THE ROLE OF DATA MINING IN TRANSPORTATION SYSTEMS

Data mining has the potential to play a crucial role in mitigating greenhouse gas (GHG) emissions and understanding their impact on the environment. Innovative solutions resulting from data mining could help meet the Kyoto Protocol's goal of replacing 30% of fossil fuel usage and bring economic benefits. The following are key agenda items for data miners in this area [17]:

- Developing technological and socio-economic solutions to reduce GHG emissions.
- Creating clean and efficient engines and powertrains, including hybrid technologies.
- Exploring alternative fuels for transportation, particularly hydrogen and fuel cells.
- Considering cost-efficiency and energy-efficiency in all decision-making.
- Developing end-of-life strategies for vehicles and vessels to reduce waste and emissions.

To achieve these goals, data miners can utilize various supervised techniques, such as classification and regression, to analyze large datasets and identify patterns and correlations. These techniques can aid in predicting future trends and making informed decisions that will reduce GHG emissions and promote sustainable development. By leveraging data mining, we can take important steps towards meeting our environmental and economic goals. The term "data mining" is used to describe many modeling techniques [18][19]. In this section, we'll take a quick look at a few of these methods. Our choice was based on the models that are used the most now. In supervised or predictive modeling, the goal is to predict an event or estimate the values of a continuous numerical attribute. In these models, there are fields for the input attributes and an output area or target. Input fields are also called predictors because the model uses them to figure out how to predict the value of an output field. You can think of the predictors as the X part of the function and the target area as the Y part, or the result.
There are two types of predictive models: classification models and estimation models [20]:

- The classification model: In these models, the groups or target classes are known from the start. The goal is to put each case into one of these groups or to predict what will happen. The generated template can be used as a "tagging engine" to add new cases to classes that have already been set up. He also gives each case a score for how likely it is to happen. The propensity score shows how likely it is that the target group or event will occur.
- Estimation models are like classification models, but there is one big difference. They are used to figure out what the value of a continuous field will be based on the values of the attributes that have already been seen.
- The supervised learning strategies for data mining that have been chosen are: Decision Trees, linear Regression, Support Vector Machines (SVM), Random Forest, K-Nearest Neighbor (KNN), Gradient Boosting and Adaboost [21].

6. ORANGE DATA MINING TOOL

Orange is a free and open-source data mining and artificial intelligence application that can be found at https://orange.biolab.si (written in Python). It may be used as a Python library since it offers a visual programming front-end for exploratory data study and depiction. The Bioinformatics Research Facility of the Workforce of PC and Data Science at the College of Ljubljana maintains and produces the program. Orange is a visual programming language for data mining, artificial intelligence, and data analysis. Contraptions are parts that range from fundamental functions like subset determination and preprocessing to observational evaluation functions like learning computations and perceptive illustrative functions [22][23]. In addition, Orange joins a large number of graphical devices that employ the Focus Library and Orange Module techniques. A visual programming tool called Orange Canvas allows devices to be put together into an application using visual programming [24].

7. BACKGROUND OF ORANGE

The authors of this article started working on Orange in 1997. Both the Laboratory of Bioinformatics and the Artificial Intelligence Laboratory at the University of Ljubljana are now working on its continued development. Orange is a collection of command-line tools and C++ components. In the beginning, Orange was intended to be a C++ library of machine learning algorithms and related procedures, such as data preparation, sampling, and other data management techniques. Our goal was to provide a platform that would allow advanced users to design their own C++ components. It turns out that this wasn't the case. Instead, Orange was mostly utilized for data exploration, testing, and scoring, utilizing various combinations of available preprocessing and learning algorithms. The parts were incorporated into applications having command-line interfaces. This was too constricting, so we chose to expose these components to Python in order to provide them with a scripting interface. As a Python module, Orange Python was chosen as the current scripting language for a number of reasons.It features a very clear and straightforward syntax that is simple for beginners as well as programmers to master. (Many top colleges, including CMU, MIT, Berkeley, Rice, and Caltech, are choosing to teach programming using Python.) Despite being simple, Python is a powerful industrial language. For instance, many of Google's technologies are powered by Python, which is also why Google is one of the main funders of Python's development. Python is a quick programming language, making it ideal for quickly developing new ideas. Adding C or C++ modules to Python is a fairly simple process. Python has even been referred to as a "glue language" since it is used to join libraries written in C or Fortran. Due to the availability of top-notch libraries like NumPy and SciPy, one seldom needs to create specialized procedures in low-level languages today. Orange has been nearly solely used as a Python module since 1999. The majority of developers have been adding to the C++ core's Python modules rather than creating their own C++ classes, despite the fact that the C++ core finally grew to reach roughly 140,000 lines. The switch to Python has made possible a number of significant advancements. In order to combine the quick functions offered by the C++ Orange core with the Python-written bridge code, more and more of Orange's functionality is being rewritten in pure Python. Because Python scripts are so readable, they make it possible for larger teams to collaborate without having to coordinate development and establish a set of coding standards. The team that develops the system now comprises between 10 and 15 individuals, the majority of whom are from the Laboratory of Bioinformatics. Most notably, switching to Python made creating the graphical user interface simpler [25-35].

8. ORANGE VISUAL PROGRAMMING

Our group has a proven track record of collaborating with scientists and industry experts from diverse fields, particularly in the area of biomedicine. In order to provide people with an accessible data exploration tool that doesn't require scripting or Python programming, we developed Orange. To build the graphical user interface for Orange, we chose the Qt library, a powerful cross-platform toolkit that offers both GPL and commercial licenses. You can see some examples of Orange's interface in Figure 1, Figure 2, and Figure 3. Most Orange users utilize its graphical interface (Figure 1). It features a canvas where users can drag and drop widgets that serve as building blocks for constructing data analysis pipelines. These widgets offer basic functionality such as reading data, displaying data tables, selecting data features based on various scoring systems, training predictors, and cross-validating models. The user can connect these widgets to establish communication channels between them. Orange's greatest strength lies in the versatility of its widgets, which can be combined in numerous ways to create custom analysis pipelines. We put a lot of effort into designing widgets that support data visualization and interaction. For example, our classification tree viewer (Figure 2) allows the user to click on a node of the tree to send the corresponding data samples to any connected widgets. This way, the user can easily explore the data associated with a particular node by examining a data table or creating scatter plots (Figure 3). When samples are chosen, they are highlighted in a scatter plot and displayed in a table (see Figure 3).

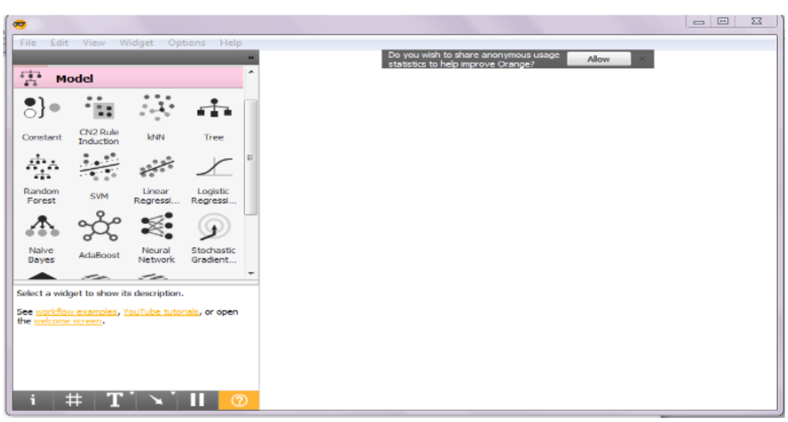


Fig. 2. Orange's main interface, which demonstrates how the models may use KNN, trees, naive bayes, and other techniques

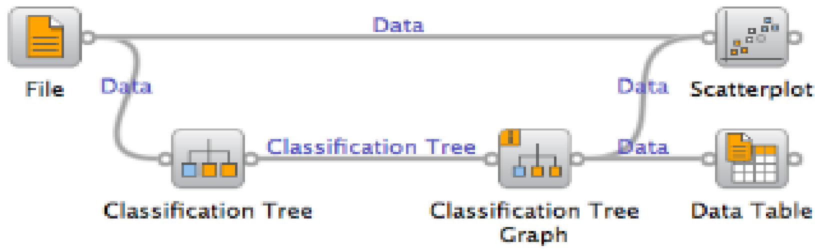


Fig. 3. shows a schema where the user may view data related to particular decision tree nodes.

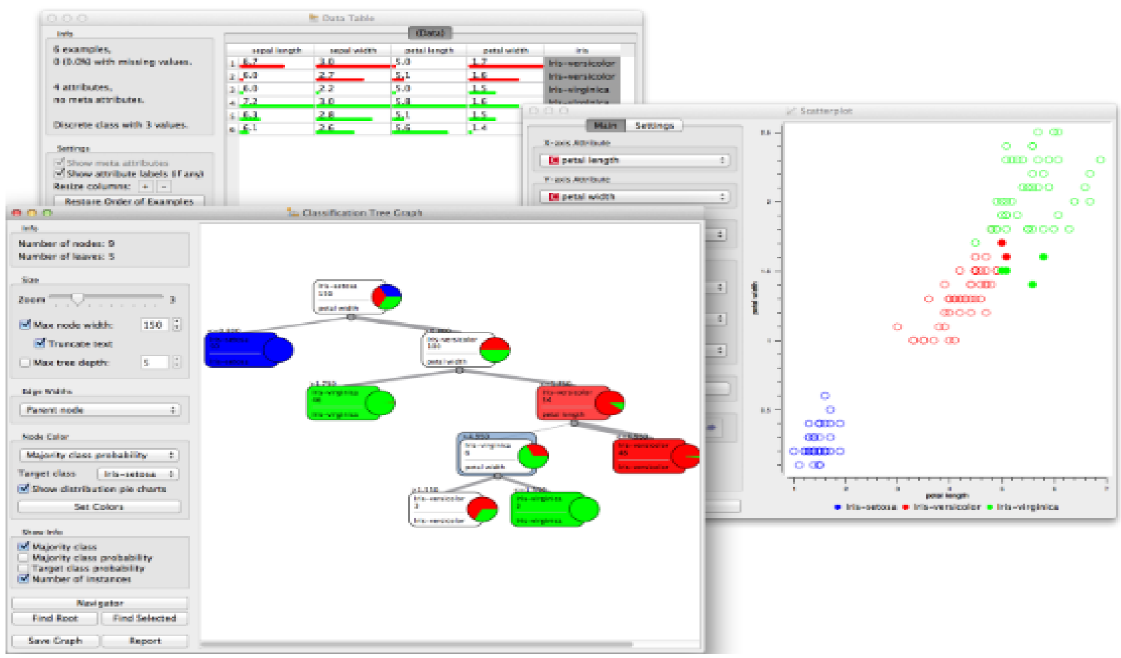


Fig. 4. The classification tree with a chosen node of six examples is displayed using widgets.

From the Figure 2 paradigm. Both the scatterplot and the table show them and highlight them. The user may alter the selection and tell Orange to update the table and scatterplot widget as well as disseminate the change across the schema. The collection of widgets transforms into a tool for exploratory data analysis by fusing interactivity and signal propagation.

9. PRACTICXAL PART

The Orange program is a set of open-source tools for displaying data, learning from machines, and mining data. It has a visual interface for exploring data and visualizing data in an interactive way. It is also a software package that uses software components for data visualization, machine learning, extraction, and analysis. The parts are called widgets, and they range from simple data visualization to subset selection and preprocessing. They are based on empirical evaluations of learning algorithms and predictive modeling. Visual programming is done through an interface, and workflows are made by connecting user interface elements that you already know or have designed before. The user interface can be changed by using the Python library to manipulate data and change the user interface. The practical program is an example of organization, and the example will be about cars, exploring data, and figuring out how much gas will be needed. CO_2 is taken away from cars, and we'll use the CARS.csv file, which has data for a group of car types. It has a table with 36 records that lists the model, weight, and amount of CO_2 based on the car's size and weight. Below, in Table 3-1, you'll find all the information.

TABLE I. CARS.CVS

Car	Model	Volume	Weight	CO2
Toyoty	Aygo	1000	790	99
Mitsubishi	Space Star	1200	1160	95
Skoda	Citigo	1000	929	95
Fiat	500	900	865	90
Mini	Cooper	1500	1140	105
VW	Up!	1000	929	105
Skoda	Fabia	1400	1109	90
Mercedes	A-Class	1500	1365	92
Ford	Fiesta	1500	1112	98
Audi	A1	1600	1150	99
Hyundai	I20	1100	980	99
Suzuki	Swift	1300	990	101
Ford	Fiesta	1000	1112	99
Honda	Civic	1600	1252	94
Hundai	I30	1600	1326	97
Opel	Astra	1600	1330	97
BMW	1	1600	1365	99
Mazda	3	2200	1280	104
Skoda	Rapid	1600	1119	104
Ford	Focus	2000	1328	105
Ford	Mondeo	1600	1584	94
Opel	Insignia	2000	1428	99
Mercedes	C-Class	2100	1365	99
Skoda	Octavia	1600	1415	99
Volvo	S60	2000	1415	99
Mercedes	CLA	1500	1465	102
Audi	A4	2000	1490	104
Audi	A6	2000	1725	114
Volvo	V70	1600	1523	109
BMW	5	2000	1705	114
Mercedes	E-Class	2100	1605	115
Volvo	XC70	2000	1746	117
Ford	B-Max	1600	1235	104
BMW	216	1600	1390	108
Opel	Zafira	1600	1405	109
Mercedes	SLK	2500	1395	120

We will train on the weights and volume of the car, and we will fit the model based on the co2 and we will use the mini data set through which we will predict two values or two records as an example If the volume is 1001 and the weight is 970, what is the amount of CO2? If the volume is 1001 and the weight is 2300, what is the amount of CO2 gas? through the attached table below

TABLE II. WEIGHTS AND VOLUME OF THE CAR

volume	Weight	co2
1001	790	?
1300	2300	?

10. THE AIM OF PROJECT

Drought, storms, and severe weather are caused by climate change. Greenhouse gases and CO2 cause climate change (CO2). emits more greenhouse gases. Simulations and data mining using past data predict CO2 will climb. 80% of CO2 emissions originate from burning fossil fuels, largely in the car and industrial sectors. Both industrialized and developing countries have implemented strategies to decrease CO2 emissions by targeting consumers or industry. This research will investigate vehicle emissions utilizing vehicle dataset characteristics and data mining methods through an orange application. The practical program operates as follows: see figure 4 The program will be built using tools that allow data entry using the tools:

- File (CSV) (CSV) It is a widget into which we will enter the automobile data via a file (cars.csv).
- The Data Table. We will employ a tool which includes the data table (cars.csv).
- Choose Columns. This tool is used to choose particular columns since we will be focusing on the value and weight and removing some columns to focus on the objective, which is CO2 gas.
- Exam and grade: It is a tool that displays the outcome of what we input from the value and weight data, as well as the amount of CO2 gas released.
- Data Table: When we used this tool a second time, it displayed the data that we had picked using the choose column tool.
- linear regression This tool is used to test the results. We connected this tool to select columns and Test and score, and the results were displayed in Test and score.
- SVM: We linked this tool with chosen columns and test and score results, and when the test and score results arrived, we compared them to the results of linear regression (see Figure 5).
- csv file: With this tool, we'll give you a file with just the value, weight, and CO2 gas.
- Tree: Some of the columns in this tool will be linked to the test, the score, and the prediction.
- Random forest: We'll connect this tool to tests, scores, and predictions.
- KNN
- Gradient boosting
- Adaboost.

Note: (All the tools that were used are different linear regression algorithms.) (All the tools that were used are different linear regression algorithms). Figure 6 shows the working mechanism in the orange program. Figure 7 shows test and scores.

	Linear Regression	SVM	Tree	Random Forest	kNN	Gradient Boosting	AdaBoost	Volume	Weight	CO2
1	93.473	99.4...	94.5	97.289	97.6	99.1011	99	1001	790	?
2	107.209	101....	115	113.164	10...	109.9	109	1300	2300	?

Fig. 5. Models results

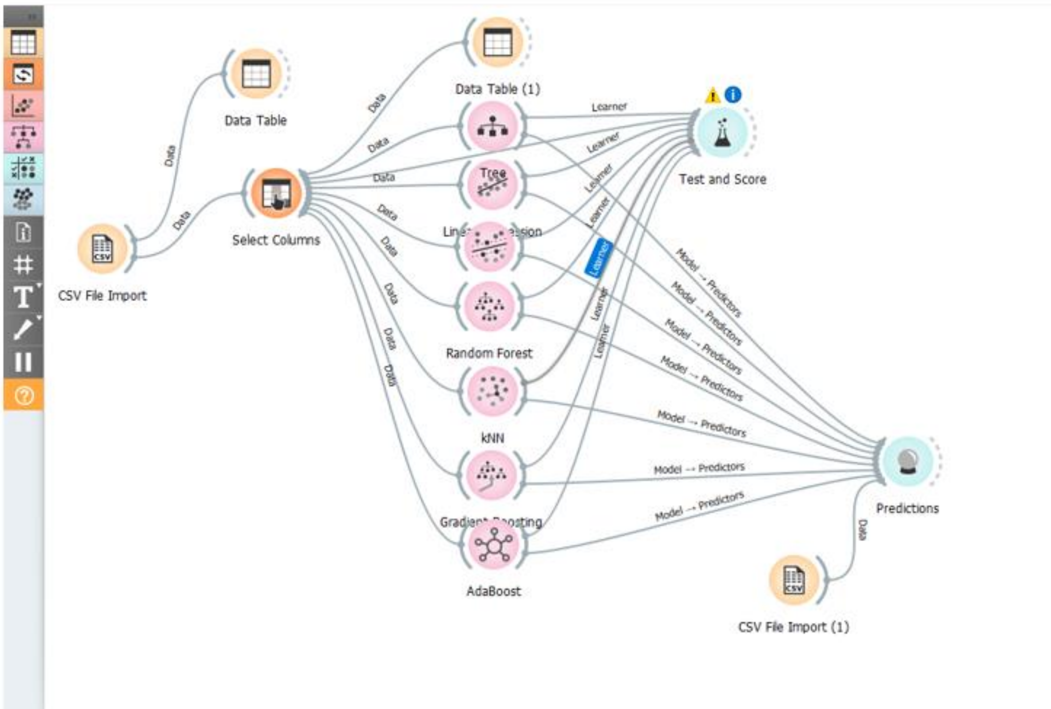
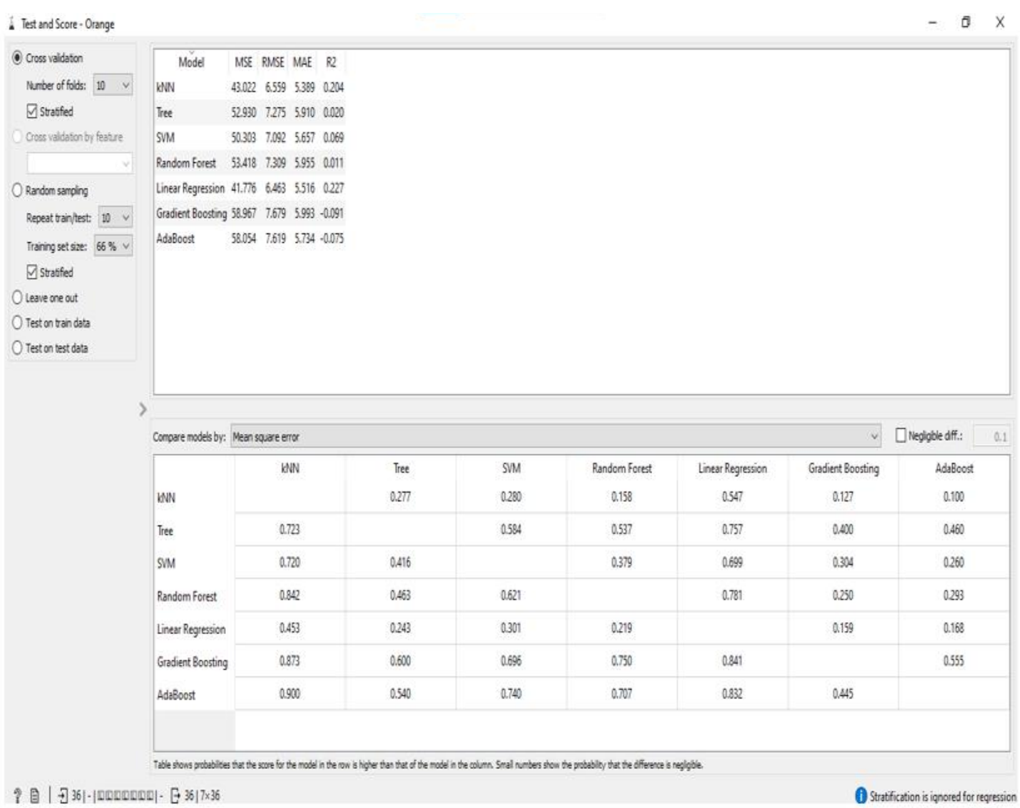


Fig. 6. Working mechanism in the orange program



Model	MSE	RMSE	MAE	R2
kNN	43.022	6.559	5.389	0.204
Tree	52.930	7.275	5.910	0.020
SVM	50.303	7.092	5.657	0.069
Random Forest	53.418	7.309	5.955	0.011
Linear Regression	41.776	6.463	5.516	0.227
Gradient Boosting	58.967	7.679	5.993	-0.091
AdaBoost	58.054	7.619	5.734	-0.075

	kNN	Tree	SVM	Random Forest	Linear Regression	Gradient Boosting	AdaBoost
kNN		0.277	0.280	0.158	0.547	0.127	0.100
Tree	0.723		0.584	0.537	0.757	0.400	0.460
SVM	0.720	0.416		0.379	0.699	0.304	0.260
Random Forest	0.842	0.463	0.621		0.781	0.250	0.293
Linear Regression	0.453	0.243	0.301	0.219		0.159	0.168
Gradient Boosting	0.873	0.600	0.696	0.750	0.841		0.555
AdaBoost	0.900	0.540	0.740	0.707	0.832	0.445	

Fig. 7. Test and scores.

11. CONCLUSION

The repercussions of climate change, which include droughts, hurricanes, and other forms of extreme weather, are having an influence on the whole planet. Greenhouse gases are the primary agents that are driving climate change, and carbon dioxide (CO2) accounts for a larger proportion of the total amount of greenhouse gases that are being emitted. It is anticipated that the concentration of CO2 will continue to rise because of simulation and data mining technologies that make use of historical data. Burning fossil fuels is the primary source of 80 percent of the world's CO2 emissions, the majority of which come from the manufacturing and automotive industries. Both developed and developing countries' governments have enacted policies to control CO2 emissions, with the goal of shifting the responsibility for doing so from consumers to manufacturers or vice versa. A group of methods have been tested by dripping data on the cars file, and it was the best method for diagnosing the amount of carbon monoxide emission in ADAboost, with an accuracy of 0.90.

Funding
The paper states that the author independently carried out the research without any financial support from institutions or sponsors.

Conflicts of Interest
The author's paper clearly states that no conflicts of interest exist in relation to the research or its publication.

Acknowledgment
The author acknowledges the institution for their commitment to fostering a research-oriented culture and providing a platform for knowledge dissemination.

References

[1] J. H. Faghmous and V. Kumar, "A Big Data Guide to Understanding Climate Change: The Case for Theory-Guided Data Science," *Big Data*, vol. 2, no. 3, pp. 155–163, 2014, doi: 10.1089/big.2014.0026.

[2] M. Burhanuddin, "Epigenetic Modifications as Predictive Biomarkers for Cardiovascular Disease: A Longitudinal Study on DNA Methylation Patterns and Disease Risk", SHIFAA, vol. 2023, pp. 59–65, Jun. 2023, doi: 10.70470/SHIFAA/2023/007.

[3] H. Kargupta, J. Gama, and W. Fan, "The next generation of transportation systems, greenhouse emissions, and data mining," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, no. July, pp. 1209–1211, 2010, doi: 10.1145/1835804.1835956.

[4] K. Jeong, T. Hong, J. Kim, and J. Lee, "A data-driven approach for establishing a CO2 emission benchmark for a multi-family housing complex using data mining techniques," *Renew. Sustain. Energy Rev.*, vol. 138, no. March, p. 110497, 2021, doi: 10.1016/j.rser.2020.110497.

[5] D. S. Jeslet and S. Jeevanandham, "Climate Change Analysis using Data Mining Techniques," *Int. J. Adv. Res. Sci. Eng.*, vol. 8354, no. 4, pp. 46–53, 2015.

[6] T. Addison and A. Bhadrashetty, "Advanced Predictive model for heart disease in clinical decision support systems", EDRAAK, vol. 2023, pp. 11–15, Feb. 2023, doi: 10.70470/EDRAAK/2023/003.

[7] C. V. V. Farhate, Z. M. De Souza, S. R. De Medeiros Oliveira, R. L. M. Tavares, and J. L. N. Carvalho, "Use of data mining techniques to classify soil CO2 emission induced by crop management in sugarcane field," *PLoS One*, vol. 13, no. 3, pp. 1–18, 2018, doi: 10.1371/journal.pone.0193537.

[8] O. Abbas, M. Elhafiz, and S. Balal, "The Role of Data Mining in Information Security," *Int. J. Comput. Sci. Inf. Technol.*, vol. 17, no. 1, pp. 1–20, 2015.

[9] H. K. Al-Mahdawi , Trans., "Analysis and prediction of rainfall using support vector machine (SVM) in the city of Najaf", Babylonian Journal of Machine Learning, vol. 2023, pp. 46–54, Sep. 2023, doi: 10.58496/BJML/2023/009.

[10] G. P. Peters *et al.*, "Carbon dioxide emissions continue to grow amidst slowly emerging climate policies," *Nat. Clim. Chang.*, vol. 10, no. 1, pp. 3–6, 2020, doi: 10.1038/s41558-019-0659-6.

[11] C. Le Quéré *et al.*, *Temporary reduction in daily global CO2 emissions during the COVID-19 forced confinement*, vol. 10, no. 7. 2020. doi: 10.1038/s41558-020-0797-x.

[12] H. Hassani, X. Huang, and A. E. Silva, "Big data and climate change," *Big Data Cogn. Comput.*, vol. 3, no. 1, pp. 1–17, 2019, doi: 10.3390/bdcc3010012.

[13] S. Gupta, S. Roy, and K. Maji , Trans., "Integrated Learning Paradigm for Ecological Predictive Modeling", Babylonian Journal of Artificial Intelligence, vol. 2023, pp. 64–73, Oct. 2023, doi: 10.58496/BJAI/2023/010.

[14] P. K. Dutta, S. M. El-kenawy, M. Abotaleb, and M. M. Eid , Trans., "AI-Driven Marketplaces and Price Prediction Tools for Rag Pickers: Enhancing Economic Opportunities in Africa's Circular Economy", Babylonian Journal of Artificial Intelligence, vol. 2023, pp. 33–42, Jul. 2023, doi: 10.58496/BJAI/2023/007.

[15] S. Mitra, S. K. Pal, and P. Mitra, “Data mining in soft computing framework: A survey,” *IEEE Trans. Neural Networks*, vol. 13, no. 1, pp. 3–14, 2002, doi: 10.1109/72.977258.
[16] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “The KDD Process for Extracting Useful Knowledge from Volumes of Data,” *Commun. ACM*, vol. 39, no. 11, pp. 27–34, 1996, doi: 10.1145/240455.240464.
[17] S. M. Khazaal and H. Maarouf , Trans., “Predicting Coronary Artery Disease Utilizing Support Vector Machines: Optimizing Predictive Model ”, MJAIH , vol. 2023, pp. 21–26, Mar. 2023, doi: 10.58496/MJAIH/2023/004.
[18] S. J. Lee and K. Siau, “A review of data mining techniques,” *Ind. Manag. Data Syst.*, vol. 101, no. 1, pp. 41–46, 2001, doi: 10.1108/02635570110365989.
[19] B. Khalid, C. Rachid, and B. Mounir, “Data Mining: Analysis and Comparative Study of Supervised Techniques,” *Int. J. Comput. Appl.*, vol. 178, no. 41, pp. 22–25, 2019, doi: 10.5120/ijca2019919300.
[20] R. K. Dwivedi, “Data Mining Techniques for Dynamically Classifying and Analyzing Library Database,” pp. 8–10, 2007.
[21] Fabiana Meijon Fadul, *済無No Title No Title No Title*. 2019.
[22] J. Demšar, B. Zupan, G. Leban, and T. Curk, “Orange : From Experimental Machine Learning,” *Knowl. Discov. Databases PKDD 2004*, pp. 537–539, 2004.
[23] B. Kaur and W. Singh, “Review on Heart Disease Prediction System using Data Mining Techniques,” *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 2, no. 10, pp. 3003–3008, 2014.
[24] A. H, Q. A., M. N., and E. M., “A Comparison Study between Data Mining Tools over some Classification Methods,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 1, no. 3, pp. 18–26, 2011, doi: 10.14569/specialissue.2011.010304.
[25] J. Demšar and B. Zupan, “Orange: Data mining fruitful and fun - A historical perspective,” *Inform.*, vol. 37, no. 1, pp. 55–60, 2013.
[26] Mijwil M. M., Doshi R., Hiran K. K., Al-Mistarehi AH, and Gök M., “Cybersecurity Challenges in Smart Cities: An Overview and Future Prospects,” *Mesopotamian journal of cybersecurity*, vol.2022, pp:1-4, 2022. https://doi.org/10.58496/MJCS/2022/001
[27] Mijwil M. M., Sadıkoğlu E., Cengiz E., and Candan H., “Siber Güvenlikte Yapay Zekanın Rolü ve Önemi: Bir Derleme,” *Veri Bilimi*, vol.5, no.2 pp:97-105, December 2022
[28] Mijwil M. M., Aljanabi M., and Ali A. H., “ChatGPT: Exploring the Role of Cybersecurity in the Protection of Medical Information,” *Mesopotamian journal of cybersecurity*, vol.2023, pp:18-21, 1 February 2023. https://doi.org/10.58496/MJCS/2023/004
[29] Mijwil M. M., Aljanabi M., and ChatGPT, “Towards Artificial Intelligence-Based Cybersecurity: The Practices and ChatGPT Generated Ways to Combat Cybercrime,” *Iraqi Journal For Computer Science and Mathematics*, vol.4, no.1, pp:65-70, January 2023. https://doi.org/10.52866/ijcsm.2023.01.01.0019
[30] Mijwil M. M., Salem I. E., and Ismaeel M. M., “The Significance of Machine Learning and Deep Learning Techniques in Cybersecurity: A Comprehensive Review,” *Iraqi Journal For Computer Science and Mathematics,* vol.4 no.1, pp:87-101, January 2023, https://doi.org/10.52866/ijcsm.2023.01.01.008
[31] Mijwil M. M., Filali Y., Aljanabi M., Bounabi M., Al-Shahwani H., and ChatGPT, “The Purpose of Cybersecurity in the Digital Transformation of Public Services and Protecting the Digital Environment,” *Mesopotamian journal of cybersecurity*, vol.2023, pp:1-6, January 2023. https://doi.org/10.58496/MJCS/2023/001
[32] Salem I. E., Mijwil M. M., Abdulqader A. W., Ismaeel M. M., Alkhazraji A., and Alaabdin A. M. Z., “Introduction to The Data Mining Techniques in Cybersecurity,” *Mesopotamian journal of cybersecurity*, vol.2022, pp:28-37, 30 May 2022. https://doi.org/10.58496/MJCS/2022/004
[33] Mijwil M. M., Unogwu O. J., Filali Y., Bala I., and Al-Shahwani H., “Exploring the Top Five Evolving Threats in Cybersecurity: An In-Depth Overview,” *Mesopotamian journal of cybersecurity*, vol.2023, pp:57-63, March 2023. https://doi.org/10.58496/MJCS/2023/010
[34] Aljanabi M. and ChatGPt, “ChatGPT: Future Directions and Open possibilities,” *Mesopotamian Journal of Cybersecurity*, vol. 2023, pp:16–17, January 2023. https://doi.org/10.58496/MJCS/2023/003
[35] Mijwil M. M., Hiran K. K., Doshi R., and Unogwu O. J., “Advancing Construction with IoT and RFID Technology in Civil Engineering: A Technology Review,” *Al-Salam Journal for Engineering and Technology*, vol. 02, no. 02, pp:54-62, March 2023. https://doi.org/10.55145/ajest.2023.02.02.007
[36] T. Curk *et al.*, “Microarray data mining with visual programming,” *Bioinformatics*, vol. 21, no. 3, pp. 396–398, 2005, doi: 10.1093/bioinformatics/bth474.