**Carbon Emission Predictor**

For Cars

horizontal line

# 

# Introduction

The 21st century has seen a growing awareness of the need for environmental sustainability on a global scale. As a result, many human activities with significant environmental implications are being re-evaluated. One such activity that has come under scrutiny is the automotive industry, as it is a major contributor to carbon emissions.

This has created a demand for innovative approaches to tackle the environmental impact of vehicular transportation. As governments, industries, and consumers all work towards sustainable practices, it has become increasingly important to develop accurate predictive models for carbon emissions from automobiles.

The shift towards sustainable transportation has made the environmental impact of automobiles a crucial topic in global discourse. As the automotive industry grapples with the challenge of reducing carbon emissions, it is imperative to create precise predictive models.

This research aims to address this need by exploring the intricate dynamics of carbon emissions from vehicles, to informed decision-making in the pursuit of a greener automotive landscape.

## Problem Statement

The transportation sector is a major contributor to greenhouse gas emissions, with conventional combustion engine vehicles being one of the primary culprits. Various factors, such as vehicle characteristics, driving conditions, and fuel types, significantly impact carbon emissions.

Therefore, it is a complex challenge to predict the amount of CO2 that a car will emit using appropriate models. Additionally, it is necessary to determine whether the car is environmentally friendly or based on its carbon footprint.

## Solution

Our research aims to address the critical gap in estimating carbon emissions from vehicles. To achieve this, we will utilize a machine learning-centric approach and a diverse dataset encompassing various variables.

Our goal is to build predictive models using different algorithms, including ensemble methods, to estimate carbon emissions from vehicles. The dataset we will be using includes various factors such as make, model, vehicle class, engine size, cylinders, transmission, fuel type, fuel consumption in the city (L/100 km), fuel consumption on the highway (L/100 km), fuel consumption combined (L/100 km), and CO2 emissions.

We will use advanced regression methods, **Decision Trees**, artificial neural networks **(ANN)**, and the k-nearest neighbors algorithm **(KNN**) to explore the algorithmic capabilities of predicting carbon emissions.

## Methodology

We are using 3 methods namely

1. **Decision Tree**
2. **Artificial Neural Network**
3. **K- Nearest Neighbour**

We have trained the data set with these algorithms to predict the value of carbon emission that is released by the cars depending on fuel consumption and type of the fuel used.

1. **Decision Tree:-**

Model building is a crucial task in any data science project. After understanding the data, processing attributes, and analyzing their correlations and prediction power, the next step is to build a model. The decision trees algorithm is a supervised algorithm that uses a tree structure. It follows a divide-and-conquer strategy, which is recursive.

The decision tree algorithm is used to predict logical results by dividing the tree into subsets. It has an accuracy of about 93%, and it can be applied to both structured and unstructured data. Decision trees do not require feature scaling and are widely used to analyze datasets and make predictions. They are a valuable support tool in various fields.

As Machine Learning models become more popular in different use cases, such as credit risk assessment systems, diagnosing systems, or energy-saving applications, explaining the models' decisions becomes increasingly important. Decision trees are easier to understand than other algorithms and are easier to explain.

The decision tree model does not require much effort from the user as it does not require data normalization. It is essential to note that there are different kinds of decision trees, depending on what you are trying to predict. For instance, regression trees are used to predict continuous quantitative data.

To build a predictive model, you must first train the model (and build the tree) using known data. Once the model's accuracy and reliability have been verified, you can use it on test data to predict known outcomes.

In machine learning, decision trees offer simplicity and a visual representation of all possible outcomes.

1. **Artificial Neural Network:-**

An Artificial Neural Network (ANN) is a computational network, that is based on biological neural networks, that are similar to the structure and operations of the Human Brain. In the same way, ANN will have neurons, that are interconnected to one another in the networks at various levels. These neurons are known as nodes. Neural networks are the collection of neurons that work together.

It makes decisions the same human.

How does our brain work?

It takes input -> process -> gives the output

Similarly, Artificial Neural Network has three layers.

**1. Input Layer**

The input Layer is the first layer that receives input (data) from external sources and allows it to move to the hidden layer.

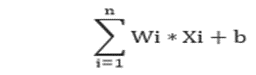
**2. Hidden Layer**

At this level, calculations are performed on the data received from the input layer to find the hidden features and patterns.

**3. Output Layer**

The process done by the hidden layer will finally result in output.

ANN takes the input and computes the weighted sum. This transfer function is represented in the form of a transfer function.

Transfer function = 

This ANN is trained by providing a data set so that it can process and identify the input that is given and give the accurate output. ANN can model complex and nonlinear relationships. This is because of real-life applications in which most of the relationships are very complex.

Most of the businesses use these technologies to solve problems like facial recognition.

It stores the data on the entire network is one of the benefits so that it can work even if there is some missing data.

1. **K- Nearest Neighbour:-**

K-Nearest Neighbors (KNN) is a simple yet powerful supervised machine learning algorithm used for classification and regression tasks. It falls under the category of instance-based learning, where the algorithm makes predictions based on the proximity of a new data point to existing labelled data points. KNN is non-parametric, meaning it doesn't make assumptions about the underlying data distribution.

The functioning of KNN is straightforward. For a given data point, the algorithm identifies its k-nearest neighbours in the training dataset based on a specified distance metric, commonly Euclidean distance. The predicted output of the new data point is then determined by the majority class (for classification) or the average of the neighbouring values (for regression) among its k-nearest neighbours.

Distance Calculation: The algorithm calculates the distance between the new data point and all other points in the training set.

Neighbour Selection: It selects the k data points with the shortest distances to the new point.

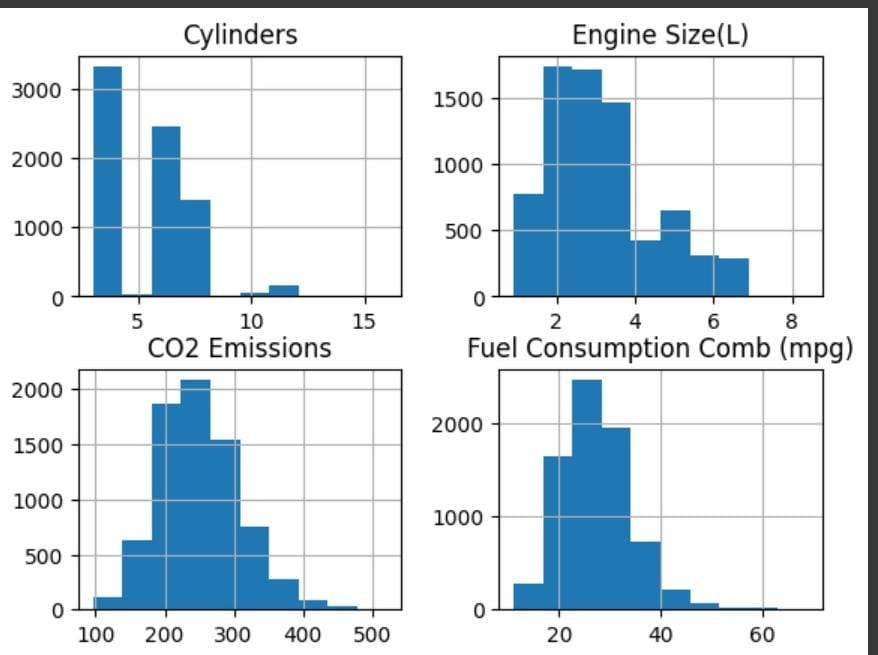
Majority Voting (Classification): For classification tasks, the algorithm assigns the class label that is most prevalent among the k-nearest neighbours.

Average (Regression): For regression tasks, the algorithm calculates the average value of the target variable for the k-nearest neighbours.

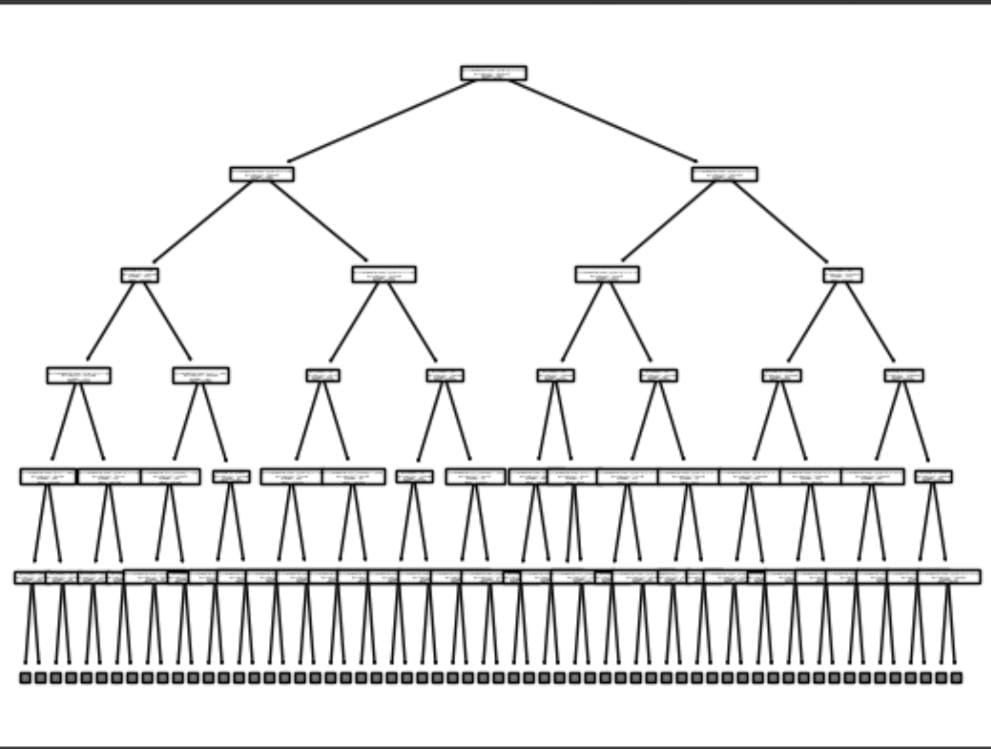
## Work Flow

## 

**Implementation**

 Analysis of

DataSet



Decision

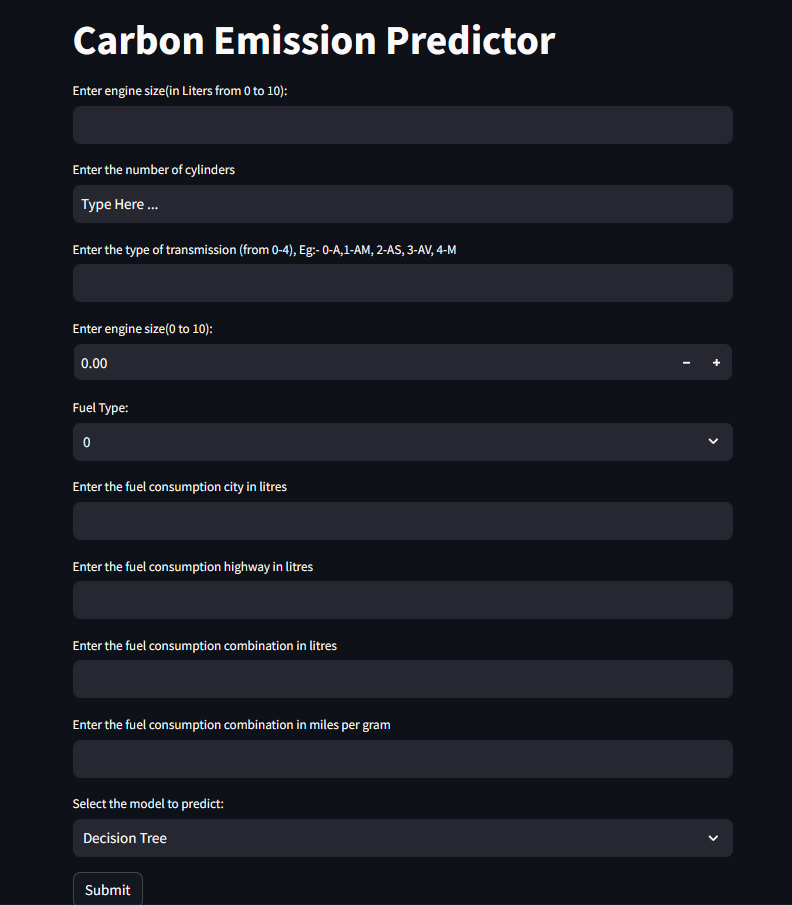
Tree

Future enhancement

* These calculate the carbon emission value only for cars.
* This can be further developed so that it can also done for different kinds of vehicles.

References

* <https://cdnsciencepub.com/doi/full/10.1139/er-2020-0019>
* <https://www.kaggle.com/datasets/midhundasl/co2-emission-of-cars-dataset>
* <https://www.kaggle.com/datasets/prathamtripathi/co2-emissions-by-cars-in-canada/data>
* <https://www.mdpi.com/2071-1050/15/9/7615>

**OUTPUT**