SmartBridge Internship Project

Project Title: **Car Performance Prediction Using IBM Watson Machine Learning**
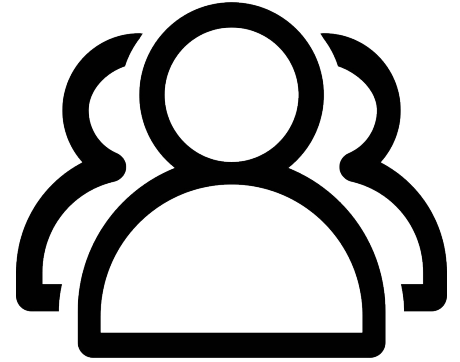
Group Members :

Achal(18BEC0605)

Aman(18BEC0648)

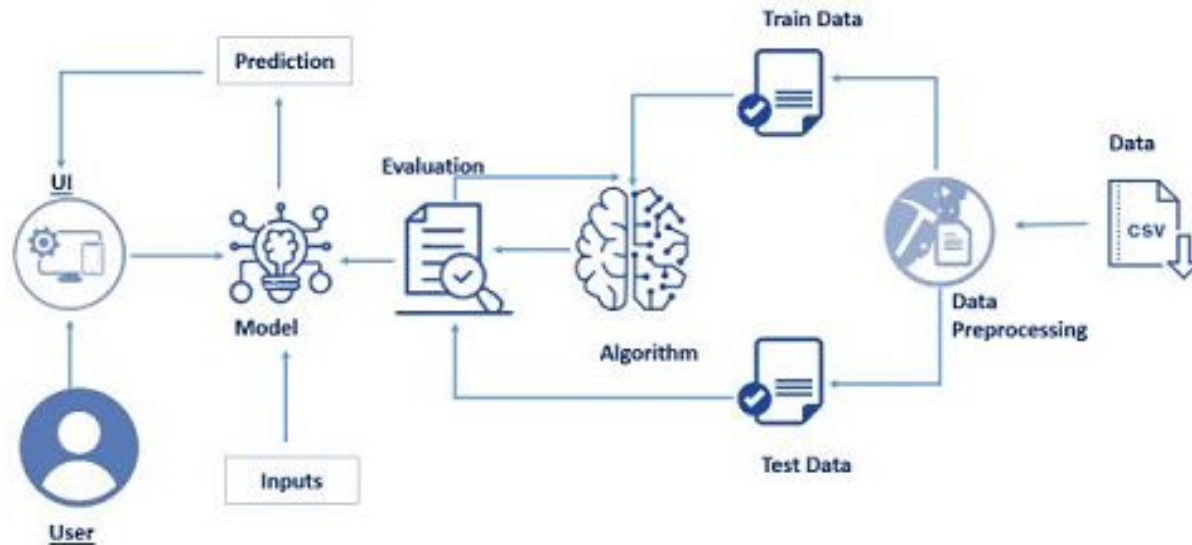Azeem Ullah Khan(19BIT0131)

Pranchal Sihare(19BIT0144)

Overview and Purpose

Predicting the performance level of cars is an important and interesting problem. The main goal is to predict the performance of the car to improve the certain behavior of the vehicle. This can significantly help to improve the system's fuel consumption and increase efficiency.
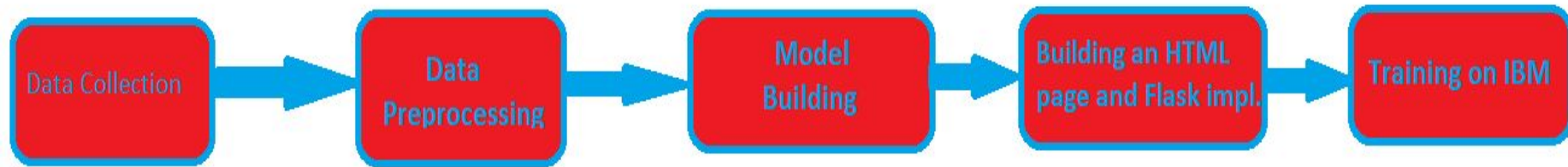
The performance analysis of the car is based on the engine type, no of engine cylinders, fuel type, and horsepower, etc. These are the factors on which the health of the car can be predicted. It is an on-going process of obtaining, researching, analyzing, and recording the health based on the above three factors. The performance objectives like mileage, dependability, flexibility, and cost can be grouped together to play a vital role in the prediction engine and engine management system. This approach is a very important step towards understanding the vehicle's performance.

Block Diagram

Flowchart

Data Collection → Data Preprocessing → Model Building → Building an HTML page and Flask impl. → Training on IBM

Random Forest Regression

Random Forest Regression is a supervised learning algorithm that uses **ensemble learning** method for regression. Ensemble learning method is a technique that combines predictions from multiple machine learning algorithms to make a more accurate prediction than a single model.
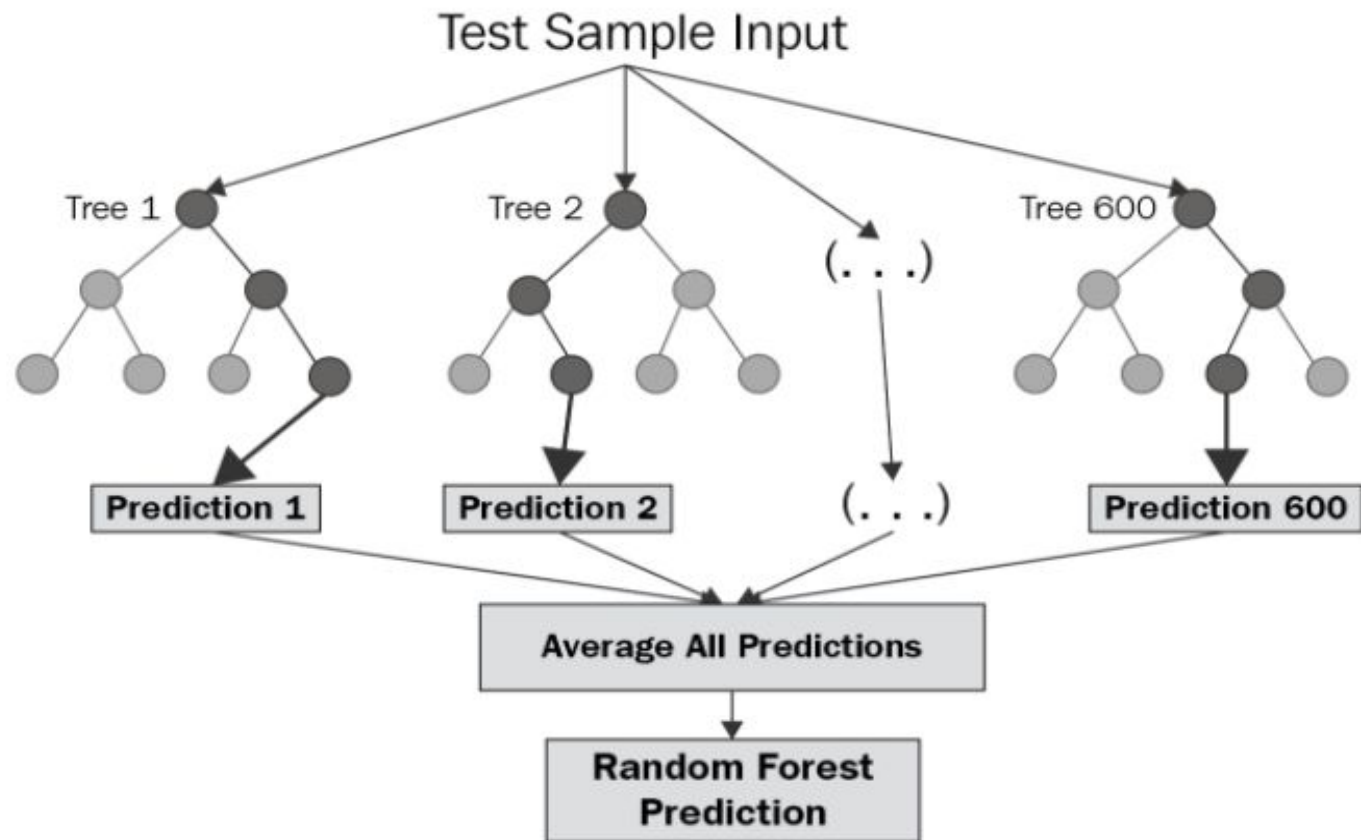
Random Forest Regression contd...

1. Pick at random k data points from the training set.

2. Build a decision tree associated to these k data points.

3. Choose the number N of trees you want to build and repeat steps 1 and 2.

4. For a new data point, make each one of your N-tree trees predict the value of y for the data point in question and assign the new data point to the average across all of the predicted y values.

Why random Forest Algorithm

A Random Forest Regression model is powerful and accurate. It usually performs great on many problems, including features with non-linear relationships. Disadvantages, however, include the following: there is no interpretability, overfitting may easily occur, we must choose the number of trees to include in the model.

Software Designing

1. **Data Collection :-**

   a. Collect the dataset or Create the dataset

2. **Data Preprocessing :-**

   a. Import the Libraries b. Importing the dataset.   c. Checking for Null Values.

   d. Data Visualization.   e. Taking care of Missing Data   f. Label encoding.

   g. One Hot Encoding. h. Feature Scaling.   i. Splitting Data into Train and Test.

Software designing Contd……

3. Model Building

      a. Training and testing the model

      b. Evaluation of Model

4. Application Building

      a. Create an HTML file

      b. Build a Python Code

Result

Advantages

We have a model that predicts fuel mileage for a variety of cars; We can use this to plan for cars that achieve desired levels of fuel efficiency. Additionally we can also observe that — per Figure 7 below — weight is the most influential variable in predicting mileage, with acceleration being second most. Horsepower, displacement, and acceleration are relatively close to each other in impact.With this detail, we can plan future car production or purchase plans.

Applications

1. Helps in knowing the maintenance cost of the car
2. Helps in knowing the approximate price of the car
3. Helps in knowing the approximate fuel efficiency of the car
4. Helps in predicting the price of the car

THANK YOU