# 

**CSE523 : Machine Learning**

Winter 2021 - 2022

**Weekly Report - 9**

Dt : 06-04-2022

**Group Name : Discover Decipher**

**Group Members**

Nimisha Patel - AU1940146 - Btech Computer Science and Engineering.

Sakshi Shah - AU1940213 - Btech Computer Science and Engineering.

Astha Patel - AU1940312 - Btech Computer Science and Engineering.

Kareena Matwani - AU1940314 - Btech Computer Science and Engineering.

**Task performed this week**

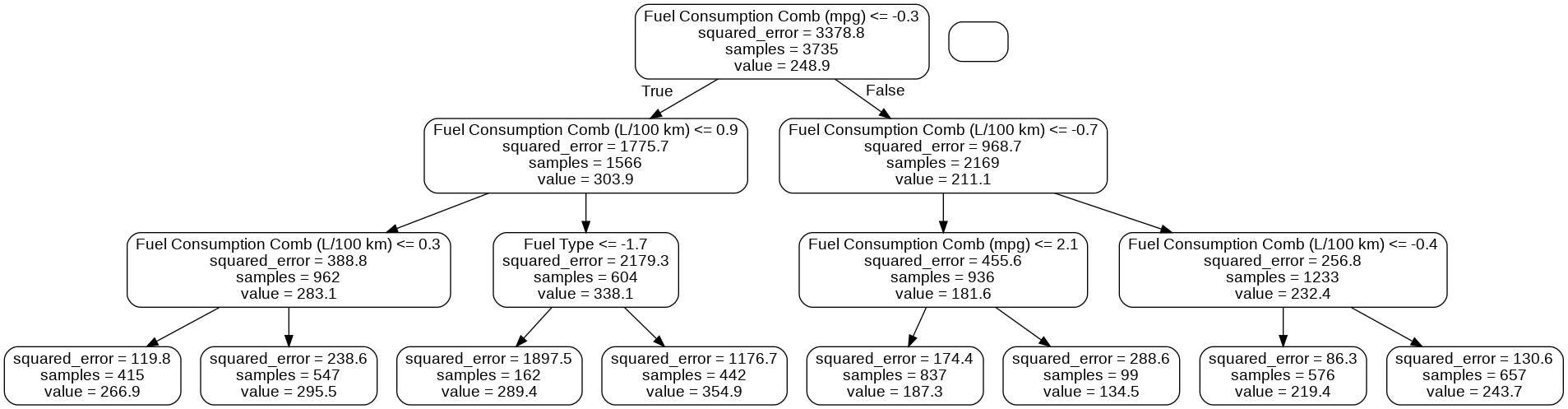
* **Performed Random Forest Regression method for prediction of CO2 emission.**

The averaging results of each single decision tree makes a random forest better than a single decision tree and reduces overfitting, thus overall leads to better accuracy and thus we performed random forest regression method after the decision tree implementation for classification and regression. Also the decision tree model gives better interpretability of the model, the negative point is that it works well only when the depth of the data is very small. For the data with the higher depth , the random forest algorithm can be used. The random forest method also provides a feature importance method to find the relevant and important features in prediction, which in turn can reduce the complexity and give the same accuracy.

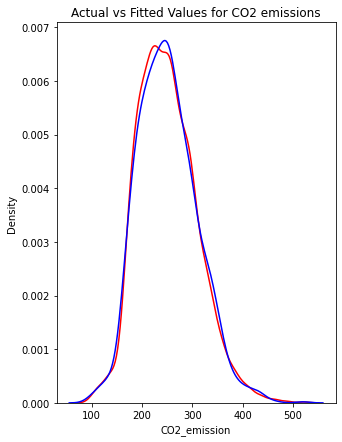
**Data set split :** Test : 20%, Train : 80%

**The random forest tree :**

* Small branched decision tree from the random forest implementation

****

**Outcome**



**RMSE: 4.5640 , MAE: 1.8745, R2\_score: 0.9939, Accuracy = 99.21%**

* **Performed Feature Importance on the 11 features (including non - numerical features) using random forest importance score feature**

**Outcome**

| **Variable** | **Importance Score** |
| --- | --- |
| **Fuel Consumption Comb (L/100 km)** | 0.74 |
| **Fuel Consumption Comb (mpg)** | 0.18 |
| **Fuel Type** | 0.05 |
| **Fuel Consumption City (L/100 km)** | 0.02 |
| **Engine Size(L)** | 0.01 |
| **Make**  **Model**  **Vehicle Class**  **Cylinders**  **Transmission**  **Fuel Consumption Hwy (L/100 km)** | 0.0 |

**Thus, now we performed the random forest model on the important features only to know the accuracy difference**

1. Most Important Feature : **Fuel Consumption Comb (L/100 km)** as a feature vector and CO2 emission as label vector.

RMSE: 19.6899

MAE: 8.0198

R2\_score: 0.8868

1. Most Important Feature : **Fuel Consumption Comb (L/100 km),Fuel Consumption Comb (mpg)** as a feature vector and CO2 emission as label vector.

RMSE: 19.6421

MAE: 8.0110

R2\_score: 0.8874

1. Most Important Feature : **Fuel Consumption Comb (L/100 km),Fuel Consumption Comb (mpg),Fuel Type** as a feature vector and CO2 emission as label vector.

RMSE: 6.2537

MAE: 2.4257

R2\_score: 0.9886

1. Most Important Feature : **Fuel Consumption Comb (L/100 km), Fuel Consumption Comb (mpg), Fuel Type,**  **Fuel Consumption City (L/100 km)** as a feature vector and CO2 emission as label vector.

RMSE: 5.7825

MAE: 2.1441

R2\_score: 0.9902

1. Most Important Feature : **Fuel Consumption Comb (L/100 km), Fuel Consumption Comb (mpg), Fuel Type,**  **Fuel Consumption City (L/100 km), Engine Size(L)** as a feature vector and CO2 emission as label vector.

**RMSE: 4.5461**

**MAE: 1.9547**

**R2\_score: 0.9940**

Thus, out of 11 features, only 5 important features can be used for prediction which helps decrease complexity of the model and are responsible for CO2 Emission.

Performed The feature selection by cross validation method and the Chosen important features are

['Engine Size(L)', 'Cylinders', 'Fuel Type', 'Fuel Consumption City (L/100 km)', 'Fuel Consumption Comb (L/100 km)', 'Fuel Consumption Comb (mpg)'].

**Task to be performed next week.**

* Hyperparameter tuning and evaluation of models on the basis of important hyperparameters.
* Random Forest Classification and dimensionality reduction using LDA approach for classification.
* Model Comparison and inferences.