**LAPORAN TUGAS PEMBELAJARAN MESIN**

**MENJELASKAN, IMPLEMENTASI, ANALISIS, SERTA DESAIN TEKNIK PEMBELAJARAN MESIN SUPERVISED LEARNING**

laporan ini dibuat untuk memenuhi kebutuhan nilai mata kuliah Pembelajaran Mesin

Dosen mata kuliah : **Tjokorda Agung Budi Wirayuda, S.M., M.T.**

**CII3C3-IF-44-11**



disusun oleh

**KIAN NAILAIZZA - 1301204455**

**MUHAMMAD ABID DARMAWAN - 1301204380**

**NAUFAL ADRIAN HIDAYAT – 1301204056**

**RHEYFAN SYAFDANI - 1301204364**

**IF-44-11**

**FAKULTAS INFORMATIKA**

**TELKOM UNIVERSITY**

**2022**

# BAB 1 PENDAHULUAN

## Project Based

## Formulasi Masalah

Formulasi Anda dalam tugas ini adalah untuk memprediksi tingkat kehematan bahan bakar kendaraan/MPG (miles per gallon: rataan jarak tempuh mobil dalam mil untuk setiap galon bahan bakar yang dikonsumsi) berdasarkan profil mobil yang diberikan yang diwakili oleh atribut-atribut seperti silinder, daya (tenaga kuda), tahun keluaran, dll. Dataset autos MPG beserta deskripsi atributnya bisa diakses di tautan berikut (login SSO):

*https://drive.google.com/drive/folders/1HmKavcNCij76k02nCYQdP012cepM2nmz?usp=sharing*

## Tugas Anda

Untuk setiap tipe tugas, mahasiswa diminta melakukan enam hal berikut:

a. Formulasi masalah:  
 Jelaskan apa permasalahan yang akan diselesaikan.

b. Eksplorasi dan pra-pemrosesan data (termasuk data splitting):  
Lakukan semua teknik eksplorasi dan pra-pemrosesan data yang menurut Anda perlu dilakukan. Jelaskan mengapa perlu melakukan teknik tersebut, dan lakukan analisis terhadap hasilnya.

c. Pemodelan:

Bangunlah model menggunakan data hasil pra-pemrosesan di poin (b), dan lakukan proses pembelajaran mesin untuk mendapatkan hasil terbaik. Jelaskan secara detail semua proses yang Anda lakukan dilengkapi dengan justifikasi dan analisis hasilnya.

d. Evaluasi:

Pilih metode evaluasi yang sesuai beserta justifikasinya. Lakukan evaluasi terhadap model yang telah dihasilkan. Berikan analisis terhadap hasil evaluasi.

e. Eksperimen:

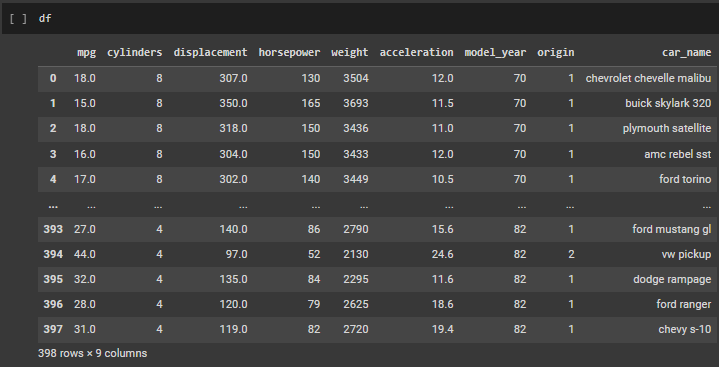
Lakukan berbagai eksperimen yang melibatkan tahapan eksplorasi dan pra pemrosesan data, pemodelan, dan evaluasi untuk mendapatkan hasil terbaik. Laporkan semua eksperimen yang Anda lakukan beserta analisis hasil dan perbandingannya.

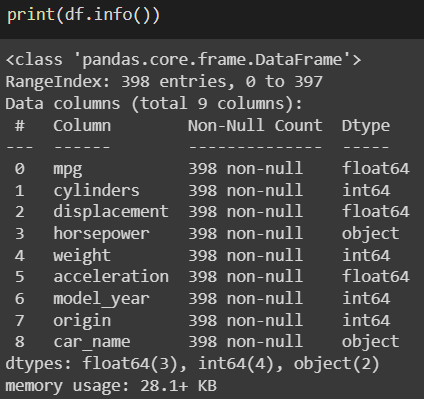
f. Kesimpulan:

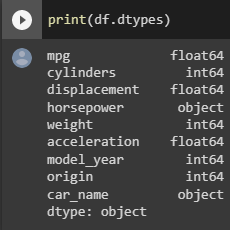
Berikan kesimpulan dari semua proses yang dijalankan serta hasil akhir dari berbagai eksperimen yang telah dilakukan.

# BAB 2 PEMBAHASAN

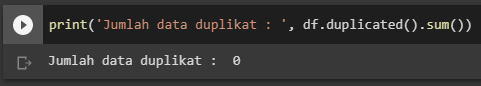
1. **Eksplorasi dan Pra-Pemrosesan Data**  
    Pada project based yang kami pakai adalah memakai NIM dari Naufal Adrian Hidayat yaitu 1301204056 dengan NIM terkecil di kelompok. Maka jika di modulo 4 akan mendapatkan hasil 0 yang artinya kami mendapatkan tugas tipe 0 yang harus mendeskripsikan tingkat kehematan bahan bakar kendaraan/MPG (miles per gallon : rataan jarak tempuh mobil dalam mil untuk setiap galon bahan bakar yang dikonsumsi) berdasarkan profil mobil yang diberikan yang diwakili oleh atribut-atribut seperti silinder, daya (tenaga kuda), tahun keluaran, dll.). Pada bagian ini, akan dijelaskan mengenai eksplorasi data serta Pra-Pemrosesan datanya.
2. **Bentuk Data**

Berikut adalah dataset dari *autos\_mpg.csv* :

Kemudian, pada tabel *autos\_mpg.csv* memiliki baris sebanyak 398 record dan juga kolom sebanyak 9 atribut serta pada dataset ini kita mendapat informasi terdiri sebanyak 3 kolom dengan tipe data float, sebanyak 4 kolom data integer, dan 2 kolom data object.

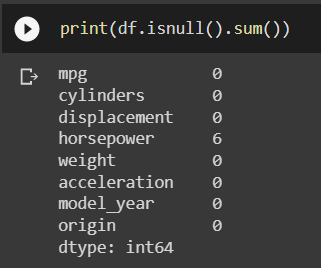


1. **Kualitas Data**
   1. **Mengecek Duplikasi Data**

Pada dataset *autos\_mpg.csv*, tidak ada data yang terduplikasi. Ketika dilakukan pengecekan data yang duplikasi, keluar *outputnya* ialah “0” yang menandakan tidak adanya data yang terduplikat pada *autos\_mpg.csv*, tersebut.

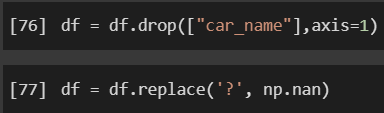
* 1. **Mengecek *Missing Value* Data**

Pada dataset *autos\_mpg.csv*,, terdapat data yang memiliki nilai kosong ataupun NaN Value. Hal tersebut bisa dibuktikan dengan mengecek datanya menggunakan code sepertinya berikut. Data yang memiliki nilai kosong akan mengeluarkan output seberapa banyak nilai kosong pada suatu kolom tersebut. Pada hal ini, jika diterapkan pada gambar, maka column horsepower yang memiliki data kosong sebanyak 6 data kosong di dalam dataset *auts\_mpg.csv*.

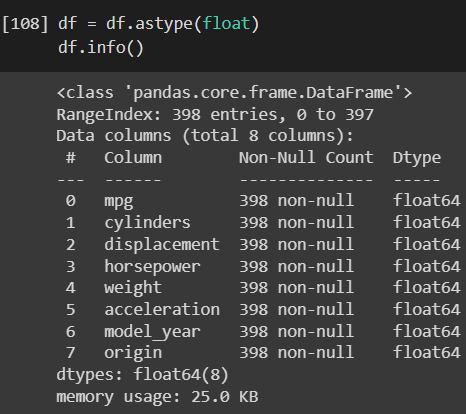


1. **Splitting dan Pra-Pemrosesan Data**
   1. **Drop Column dan Replace Nilai Kosong**

Pada bagian ini, kami ingin membuat column car\_name dihapus/didrop dikarenakan nilai atau bentuk dari isi column tersebut adalah string, maka dari itu column car\_name di drop. Kemudian kami akan mengganti nilai yang kosong atau tanda tanya “?” menjadi Nan.. Bisa dilihat pada gambar di bawah, bahwa kami melakukan drop column serta replace tanda tanya menjadi NaN.

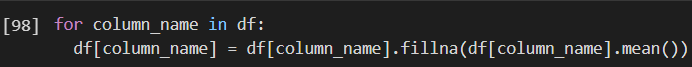
****

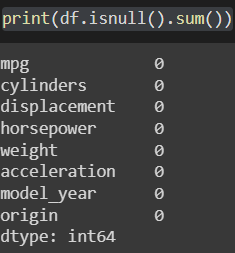
* 1. **Ubah Tipe Data Column**

Pada bagian ini, setelah menghapus column dan mengganti nilai kosong kami juga mengganti tipe data setiap column menjadi bentuk desimal atau float. Hal tersebut dikarenakan untuk mempermudah kami dalam membuat penghitungan dan pengerjaan pada tahap berikutnya. Pada gambar tersebut, bisa dilihat kami mengganti seluruh tipe data dalam column menjadi float. Jadi, ada 8 record bentuk float pada dataset *autos\_mpg.csv.*

* 1. **Memasukkan Nilai Kosong Menjadi Nilai Mean**

Pada bagian ini, kami mengisi nilai kosong atau yang diberi tanda tanya “?” dengan nilai Mean. Bisa dilihat pada gambar berikut, seperti ini kami memberikan nilai mean dari keseluruhan column ke dalam column Horsepower atau column yang memiliki nilai kosong tersebut. Jika dicek kembali, maka benar-benar data yang kosong tersebut terganti dengan nilai mean.



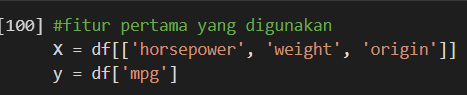


* 1. **Splitting Data**

Pada bagian ini, setelah memberikan nilai yang kosong tersebut menjadi nilai mean. Kami baru melakukan splitting data yakni membagi dataset tersebut menjadi data train dan data test. Hal ini, kami memiliki tiap nilai pada splitting data tersebut, yaitu x\_train, x\_test, y\_train, y\_test. Seperti yang sudah dijelaskan, berikut adalah codingan pada pembuatan splitting data. Artinya data kami membagi data test 0.2 dan data train sebesar 0.8.

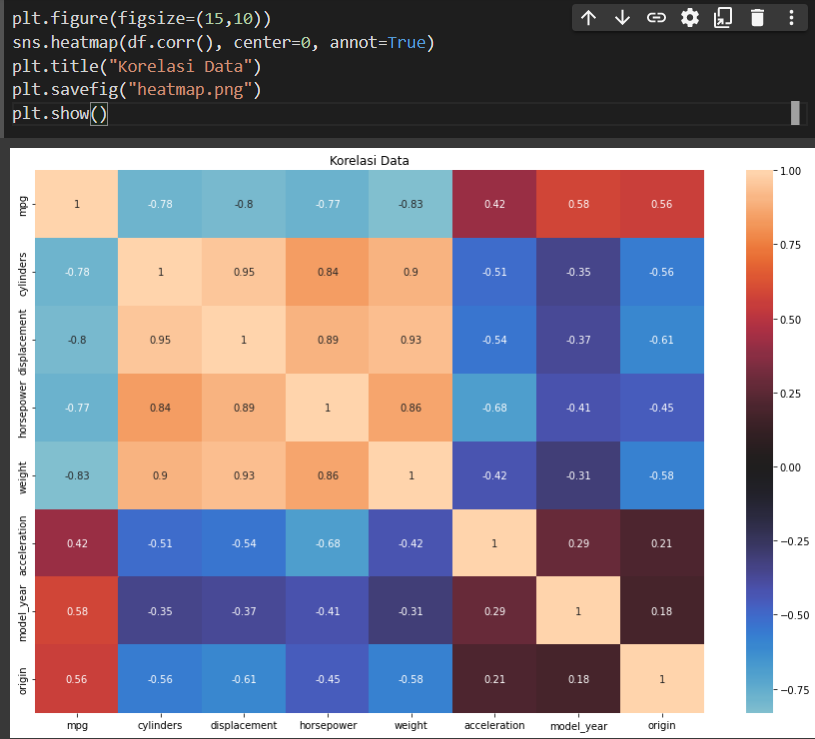


Pada splitting data, untuk data yang x kami memakai fitur horsepower, weight, origin dan data yang menjadi target ataupun y adalah mpg.



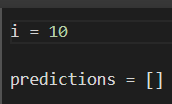
1. **Visualisasi Data**
   1. **Visualisasi Heatmap**

Kami memakai Heatmap untuk mengecek korelasi antar data. Korelasi

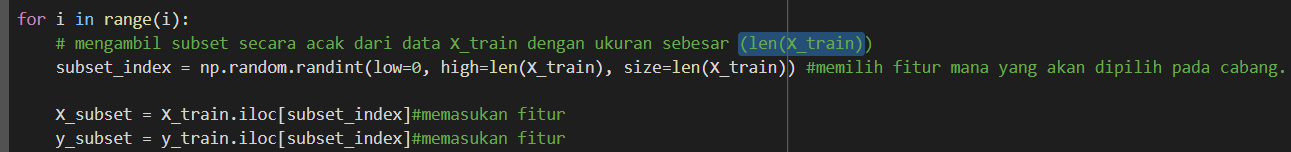


1. **Pemodelan**

Pada pemodelan ini, kami membuat 10 model yang bisa dilihat dari “i” yang diatur atau diset dengan nilai 10. Kemudian sebelum melakukan modelling, kami membuat array kosong yang dinamai prediction yang artinya list kosong tersebut yang akan dipakai untuk menyimpan prediksi dari masing-masing model.

****

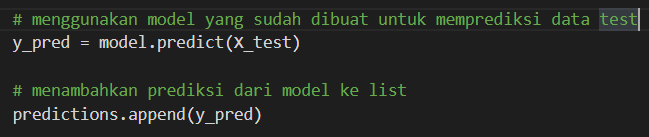
Kemudian setelah membuat array atau list kosong dan i = 10, kami melakukan looping dari i. Pertama yang dilakukan adalah mengambil subset dengan acak dari data x\_train dengan ukuran yang sebesar dari len(x\_train).



Setelah mendapatkan subset indeksnya, kami memasukkan subset indeks tersebut ke dalam variabel x\_subset dan y\_subset. Selanjutnya, kami membuat variabel model yang berisi decision tree regressor. Setelah itu, kami memakai model tersebut ke dalam x\_subset dan y\_subset. Berikut implementasi codingan pada pembuatan model decision tree regressor

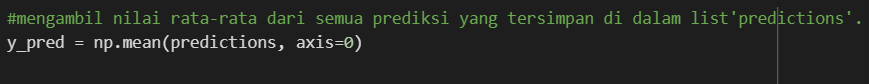


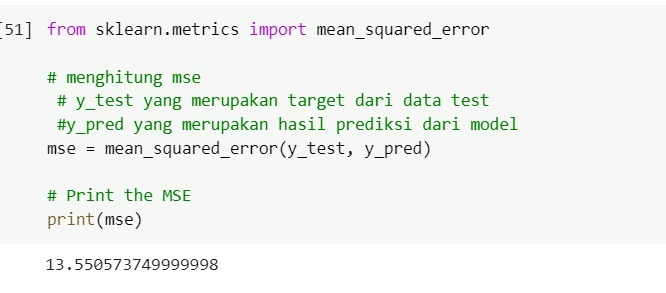
Setelah memasukkan x\_subset dan y\_subset ke dalam model, selanjutnya menggunakan model tersebut untuk mempredict x\_test dan dimasukkan ke dalam y\_pred. Setelah mendapatkan hasil prediksi, hasil tersebut akan dimasukkan ke dalam list prediction yang dibuat pada tahap sebelumnya. Berikut adalah contoh implementasi codingannya.



1. **Evaluasi**

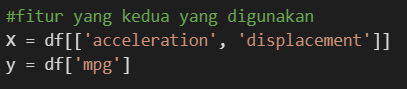
Pada evaluasi ini kami memakai MSE (Mean Squared Error). MSE adalah sebuah ukuran yang digunakan untuk mengevalusai kinerja sebuah pembelajaran mesin. Sebelum menghitung MSE, yang pertama dilakukan adalah mengambil nilai rata-raya dari semua prediksi yang tersimpan pada list prediction.

****

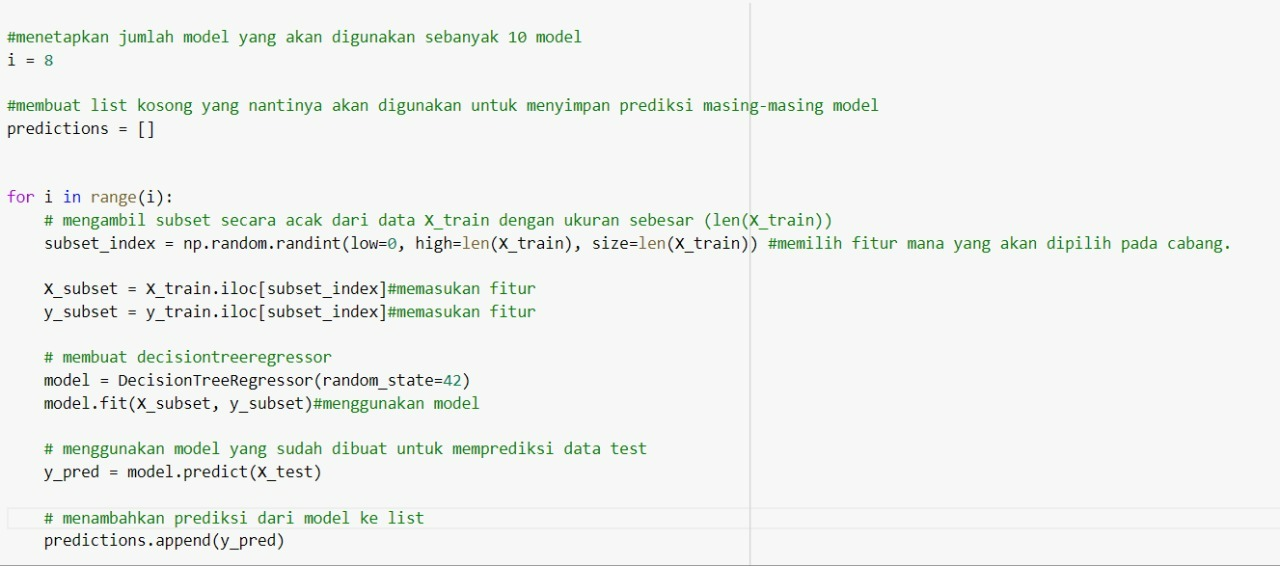
Sebelum memulai MSE, lakukanlah import library from sklearn.metrics import mean\_squared\_error. Selanjutnya menghitung MSE dengan y\_test merupakan target dari dataset y\_pred merupakan nilai rata-rata hasil prediksi. Berikut adalah implementasi codingan dan hasil dari evaluasinya.

1. **Eksperimen**

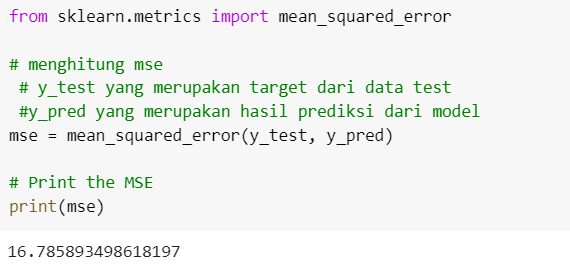
Eksperimen pertama yang kami lakukan ialah kami mengganti fitur menjadi berikut. yakni acceleration dan displacement dan yang menjadi target fitur tersebut adalah mpg.



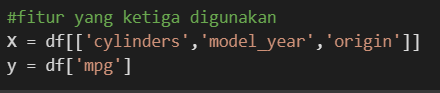
Selanjutnya kami mengganti random state pada bagian splitting data menjadi 30 dan mengubah model yang akan digunakan menjadi 8 yaitu variabel i = 8. 



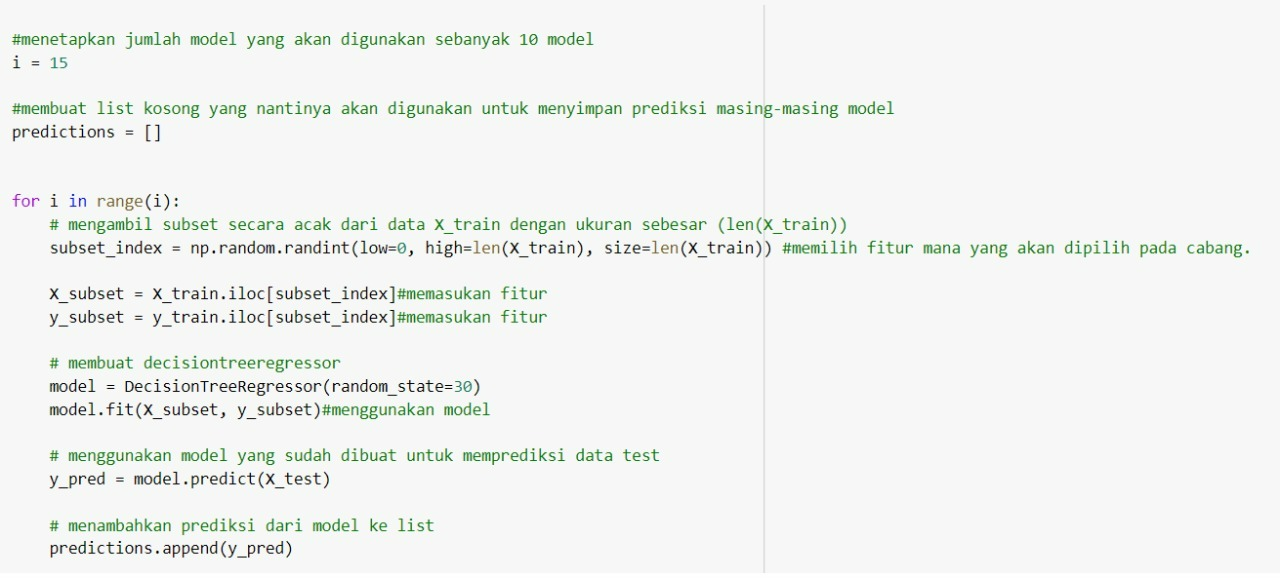
Pada pengubahan dari random state dan jumlah model yang dipakai, kami mendapatkan hasil MSE sebagai berikut.



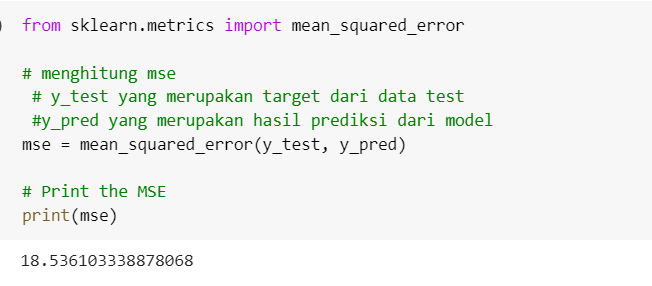
Kemudian kami melakukan eksperimen kembali dengan fitur yang ketiga yaitu column cylinders, model\_year, dan origin serta tagetnya yakni tetap column mpg.



Lalu, mengganti random state pada data split menjadi 50 dan model i yang digunakan menjadi 15 serta random state pada decision tree regressor menjadi 30. Berikut adalah implementasinya.

Pada pengubahan dari random state dan jumlah model yang dipakai, kami mendapatkan hasil MSE sebagai berikut.



1. **Kesimpulan**

Pada percobaan pertama dengan memakai fitur pertama yaitu housepower, weight, dan origin kami dengan random\_state data split 42 dan jumlah model adalah 10 serta decision tree regresor random\_state sebesar 42 mendapatkan hasil MSE 13.5

Pada percobaan kedua dengan memakai fitur kedua yaitu acceleration, displacement, dengan random\_state data split 30 dan jumlah model adalah 8 serta decision tree regresor random\_state sebesar 42 mendapatkan hasil MSE 16.7

Pada percobaan ketiga dengan memakai fitur ketiga yaitu cylinders, model\_year, dan origin, dengan random\_state data split 50 dan jumlah model adalah 15 serta decision tree regresor random\_state sebesar 30 mendapatkan hasil MSE 18.5

Pada ketiga tersebut dapat disimpulkan bahwa percobaan pertama mendapatkan hasil MSE lebih kecil dibandingkan percobaan kedua dan ketiga. Pada ketentuan MSE sendiri, nilai MSE yang lebih kecil menunjukkan bahwa model tersebut lebih akurat dalam memprediksi nilai yang sudah diharapkan. Maka model terbaik yang didapatkan adalah pada percobaan pertama dengan fitur housepower, weight, dan origin dengan random\_state data split 42 dan jumlah model adalah 10 serta decision tree regresor random\_state sebesar 42.

1. **Lampiran**

**Link Codingan (Google Collabs) :**

[*https://colab.research.google.com/drive/1UsBB2HKeRI8ktJ6WqSusmLzS\_WOxKAP1?usp=sharing*](https://colab.research.google.com/drive/1UsBB2HKeRI8ktJ6WqSusmLzS_WOxKAP1?usp=sharing)

**Link PowerPoint :**

[*https://docs.google.com/presentation/d/1uGncHJlJFmpygqe2ly8aIVgpG7B7IJnm/edit?usp=sharing&ouid=100359807533626157288&rtpof=true&sd=true*](https://docs.google.com/presentation/d/1uGncHJlJFmpygqe2ly8aIVgpG7B7IJnm/edit?usp=sharing&ouid=100359807533626157288&rtpof=true&sd=true)

**Link Laporan :**

[*https://docs.google.com/document/d/12ZLSbMNQA6uXDT\_xDHU3BBlz7E7Fpisv/edit?usp=sharing&ouid=100359807533626157288&rtpof=true&sd=true*](https://docs.google.com/document/d/12ZLSbMNQA6uXDT_xDHU3BBlz7E7Fpisv/edit?usp=sharing&ouid=100359807533626157288&rtpof=true&sd=true)

**Link Video (Youtube) :**

<https://youtu.be/euXYZvgnQpY>