

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import os
from sklearn import datasets, linear_model
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import mean_squared_error
from sklearn.tree import export_graphviz
from subprocess import check_call
import pydot as pydot

#master_input = pd.read_csv("C:/Users/manav.gupta/OneDrive - Antuit India Private
Limited/Attachments/Mahindra/Pricing/Manav_Tests/feature_input_manav.csv")
#master_input = pd.read_csv("C:/Users/balaji.suresh/Desktop/Mahindra Pricing/Model
Iterations/Manav Final 2 - New Variant Test/feature_input_manav.csv")
master_input = pd.read_csv("C:/Users/balaji.suresh/Desktop/Mahindra Pricing/Demand Estimation/PE
and MS data files/feature_input_manav_pricing.csv")

master = master_input.drop(['Unique_Identity','Make','Model','Version','Month','Reason','Overall length
(mm)'
                        , 'Overall width (mm)', 'Overall height (mm)', 'Ground clearance unladen (mm)'
                        , 'Maximum_Torque_Nm', 'bulb type (high beam)'
                        , 'Electric trunk/hatch pull down', 'Roof rails', 'Powered Tailgate'], axis = 1)

master.shape

#####TRAINING THE
MODEL#####
X_train = master[master.columns[1:]]
y_train = master['Price']

rf_1 = RandomForestRegressor(n_estimators = 200, criterion='mse', max_features='auto',
max_depth=5, random_state = 1729,min_samples_split = 2 )
rf_1.fit(X_train, y_train)
rf_1.feature_importances_
features_contrib_rf = pd.DataFrame(data = list(zip(X_train.columns,rf_1.feature_importances_)),
columns = ['features','coefficients'])
features_contrib_rf.sort_values(by = ['coefficients'], ascending = False, inplace = True)
#features_contrib_rf.to_csv("C:/Users/manav.gupta/OneDrive - Antuit India Private
Limited/Attachments/Mahindra/Pricing/Manav_Tests/Model_2.csv", index = False)
features_contrib_rf.to_csv("C:/Users/balaji.suresh/Desktop/Mahindra Pricing/Model Iterations/Manav
Final 2 - New Variant Test/feat_imp.csv", index = False)

```

```
features_contrib_rf.head(20)
```

###TEST PREDICTION OF MODEL

```
test_input = pd.read_csv("C:/Users/manav.gupta/OneDrive - Antuit India Private
Limited/Attachments/Mahindra/Pricing/Manav_Tests/prediction_check_data.csv")
#test_input = pd.read_csv("C:/Users/balaji.suresh/Desktop/Mahindra Pricing/Model Iterations/Manav
Final 2 - New Variant Test/prediction_check_data.csv")
#test_input = pd.read_csv("C:/Users/balaji.suresh/Desktop/Mahindra Pricing/Demand Estimation/PE
and MS data files/w6 w8 w10 ++.csv")

# test_input = test_input[~test_input.Unique_Identity.isin(variant_drop_list)]
test_input.head()
test = test_input.drop(['Price','Unique_Identity','Make','Model','Version','Month','Reason','Overall
length (mm)'
                        , 'Overall width (mm)', 'Overall height (mm)', 'Ground clearance unladen (mm)'
                        , 'Maximum_Torque_Nm', 'bulb type (high beam)'
                        , 'Electric trunk/hatch pull down', 'Roof rails', 'Powered Tailgate'], axis = 1)

predict_output = pd.DataFrame(data = list(zip(test_input.Unique_Identity, rf_1.predict(test))), columns =
['unique_identifier', 'predicted_price'])
predict_output.head(10)
#predict_output.to_csv("C:/Users/manav.gupta/OneDrive - Antuit India Private
Limited/Attachments/Mahindra/Pricing/Manav_Tests/prediction_output_1Nov_v1.csv", index=False)
predict_output.to_csv("C:/Users/balaji.suresh/Desktop/Mahindra Pricing/Model Iterations/Manav Final
2 - New Variant Test/prediction_output_4.csv", index=False)
```