PROJECT DEVELOPMENT PHASE

DELIVERY OF SPRINT-2

DATE	4 OCTOBER 2022
TEAM ID	PNT2022TMID43633
PROJECT NAME	Car Resale Value Prediction
MAXIMUM MARK	4 Marks

- **Label encoding the categorical data**

```
#label encoding the categorical data
labels = ['gearbox','notRepairedDamage','brand','fuelType','vehicleType']
```

In [24]:

```
mapper = {}
for i in labels:
        mapper[i] = LabelEncoder()
        mapper[i].fit(new_df[i])
        tr = mapper[i].transform(new_df[i])
        np.save(str('classes'+i+'.npy'),mapper[i].classes_)
        print(i,";",mapper[i])
        new_df.loc[:,i+'_labels'] = pd.Series(tr,index = new_df.index)
labeled = new_df[ ['price','yearOfRegistration','powerPS','kilometer','monthOfRegistration'] + [x+"_labels" for x in labels]]
```

```
gearbox ; LabelEncoder()
notRepairedDamage ; LabelEncoder()
brand ; LabelEncoder()
fuelType ; LabelEncoder()
vehicleType ; LabelEncoder()
```

In [25]:

```
print(labeled.columns)
```

```
Index(['price', 'yearOfRegistration', 'powerPS', 'kilometer',
       'monthOfRegistration', 'gearbox_labels', 'notRepairedDamage_labels',
       'brand_labels', 'fuelType_labels', 'vehicleType_labels'],
      dtype='object')
```

- **Choosing appropriate model**

In [46]:
```
Y = Y.reshape(-1,1)
from sklearn.model_selection import cross_val_score , train_test_split
X_train,X_test,Y_train,Y_test = train_test_split(X,Y,test_size=0.3,random_state=3)
```

In [47]:
```
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import r2_score
regressor = RandomForestRegressor(n_estimators = 1000,max_depth=10,random_state=34)
```

In [48]:
```
regressor.fit(X_train,np.ravel(Y_train,order='C'))
```

Out[48]:
```
RandomForestRegressor(max_depth=10, n_estimators=1000, random_state=34)
```

In [49]:
```
y_pred = regressor.predict(X_test)
print(r2_score(Y_test,y_pred))
```

```
0.8282067869983389
```

- **Dumping to pickle file**

In [50]:
```
filename = 'resale_value_pickle_file.sav'
pickle.dump(regressor,open(filename,'wb'))
```