

PROJECT DEVELOPMENT PHASE (DELIVERY OF SPRINT-2)

Date	10th November 2022
Team ID	PNT2022TMID01716
Project Name	Car Resale Value prediction
Maximum marks	4 Marks

- **Choosing Appropriate Model**

```
[ ] from sklearn.model_selection import cross_val_score, train_test_split

#Split into train and validation
X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.33, random_state = 3)
print(X_train.shape, X_test.shape, y_train.shape, y_test.shape)
```

```
(203769, 7) (100364, 7) (203769,) (100364,)
```

```
from sklearn.ensemble import HistGradientBoostingRegressor
from sklearn.model_selection import GridSearchCV

hr = HistGradientBoostingRegressor()

param_grid = { "loss" : ['squared_error']
               , "max_leaf_nodes" : [31]
               , "min_samples_leaf": [20]
               , "max_depth": [None]
               , "max_iter": [500]}

gs = GridSearchCV(estimator=hr, param_grid=param_grid, cv=2, n_jobs=-1, verbose=1)
gs = gs.fit(X_train, y_train)
print('Score: %.2f' % gs.score(X_test, y_test))
```

➡ Fitting 2 folds for each of 1 candidates, totalling 2 fits
Score: 0.78

- **Dumping into pickle file:**

```
[ ] print(gs.best_score_)  
    print(gs.best_params_)
```

```
0.7707290246596359
```

```
{'loss': 'squared_error', 'max_depth': None, 'max_iter': 500, 'max_leaf_nodes': 31, 'min_samples_leaf': 20}
```

```
[ ] import pickle
```

```
pickle.dump(gs, open('histmodel.pkl', 'wb'))
```