PROJECT DEVELOPMENT PHASE

DELIVERY OF SPRINT – 3

Date	17 November 2022
Team Id	PNT2022TMID21221
Project Name	Car Resale Value Prediction

In sprint-3, the model is built. The machine learning model that we have used to predict the car resale value is Random Forest Classifier. Here, data is encoded using label encoder and the dataset is split into training and testing test upon which the ml model is applied and finally the model is saved for future purposes. This sprint is the continuation of the previous sprint sprint-2.

CODE:

```
#Label Encoding the categorical data

labels = ['gearbox', 'notRepairedDamage', 'model', 'brand', 'fuelType', 'vehicleType']

mapping = {}

for i in labels:

    mapping[i] = LabelEncoder()

    mapping[i].fit(newdf[i])

    trans = mapping[i].transform(newdf[i])

    np.save(str('classes'+i+'.npy'),mapping[i].classes_)

    print(i,":",mapping[i])

    newdf.loc[:,i+'_labels'] = pd.Series(trans, index = newdf.index)


#final data is put inside a new dataframe called labeled

labeled = newdf[["price",

        "yearOfRegistration",

        "powerPS",

        "kilometer",

        "monthOfRegistration"]

        + [x+"_labels" for x in labels]]

print(labeled.columns)


#sorting price in Y and rest of the data in X
```

```python
Y = labeled.iloc[:,0].values

X = labeled.iloc[:,1:].values

Y = Y.reshape(-1,1)


#splitting the dataset into testing and training set

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)


from sklearn.ensemble import RandomForestRegressor

from sklearn.metrics import r2_score

regressor = RandomForestRegressor(n_estimators=1000, max_depth = 10, random_state=34)

#fitting the model

regressor.fit(X_train, np.ravel(Y_train,order='C'))


#predicting the values of test test

y_pred = regressor.predict(X_test)

#predicting the accuracy for test set

print(r2_score(Y_test, y_pred))


#saving the model for future use

filename = 'resale_model.sav'

pickle.dump(regressor, open(filename,'wb'))
```