

ML practice

File Edit View Insert Runtime Tools Help Saving...

Files

sample_data

Travel.csv

cardekho_imputed.csv

+ Code + Text

[77] c=a.drop(['ProdTaken','CustomerID'],axis=1)

c

	Age	TypeofContact	CityTier	DurationOfPitch	Occupation	Gender	NumberOfPersonVisiting	NumberOfFollowups	ProductPitched
0	41.0	Self Enquiry	3	6.0	Salaried	Female	3	3.0	Deluxe
1	49.0	Company Invited	1	14.0	Salaried	Male	3	4.0	Deluxe
2	37.0	Self Enquiry	1	8.0	Free Lancer	Male	3	4.0	Basic
3	33.0	Company Invited	1	9.0	Salaried	Female	2	3.0	Basic
4	36.0	Self Enquiry	1	8.0	Small Business	Male	2	3.0	Basic
...
4883	49.0	Self Enquiry	3	9.0	Small Business	Male	3	5.0	Deluxe
4884	28.0	Company Invited	1	31.0	Salaried	Male	4	5.0	Basic

0s

completed at 11:28 AM

Disk

75.09 GB available

Search

11:29 08-12-2024

Chat | dhineshkannan65... x Machine Learning - Colab x ML practice - Colab x colsample_bytree - Google x GradientBoostingRegres... x

colab.research.google.com/drive/1DlDhy_mW7lfh5lQdniTEAqy7GJc1dCp9#scrollTo=GzzBld_T2Mda

ML practice

File Edit View Insert Runtime Tools Help Saving...

Files

- sample_data
- Travel.csv
- cardekho_imputed.csv

Code

```
d=a['ProdTaken']  
d
```

Next steps: [View recommended plots](#) [New interactive sheet](#)

	ProdTaken
0	1
1	0
2	1
3	0
4	0
...	...
4883	1
4884	1
4885	1
4886	1
4887	1

0s completed at 11:28 AM

Search

ENG IN 11:29 08-12-2024

Machine Learning - Colab ML practice - Colab

colab.research.google.com/drive/1DlDhy_mW7lfh5lQdniTEAqy7GJc1dCp9#scrollTo=GzzBld_T2Mda

ML practice

File Edit View Insert Runtime Tools Help Saving...

Files

- sample_data
- Travel.csv
- cardekho_imputed.csv

```
[79] from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(c,d,test_size=0.3)

[80] from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import ColumnTransformer
e=c.select_dtypes(include='object').columns
f=c.select_dtypes(exclude='object').columns
g=StandardScaler()
h=OneHotEncoder(drop='first')
ct=ColumnTransformer([('OneHotEncoder',h,e),('StandardScaler',g,f)])

[84] ct

[85] c=ct.fit_transform(c)
```

ColumnTransformer

- OneHotEncoder
- StandardScaler

completed at 11:28 AM

11:29 08-12-2024

Chat | dhineshkannan65... x Machine Learning - Colab x ML practice - Colab x colsample_bytree - Google x GradientBoostingRegres... x

colab.research.google.com/drive/1DlDhy_mW7lfh5lQdniTEAqy7GJc1dCp9#scrollTo=GzzBld_T2Mda

ML practice

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample_data
- Travel.csv
- cardekho_imputed.csv

Code

```
[85] c=ct.fit_transform(c)

[86] x_train,x_test,y_train,y_test=train_test_split(c,d,test_size=0.3)

[87] from sklearn.ensemble import RandomForestClassifier
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score,classification_report,f1_score,precision_score,recall_score,roc_auc_score

[88] ca={'RandomForestClassifier':RandomForestClassifier(),'GradientBoostingClassifier':GradientBoostingClassifier(),'LogisticRegression':LogisticRegression(),'DecisionTreeClassifier':DecisionTreeClassifier()}

[89] for i in range(len(ca)):
    l=list(ca.values())[i]
    l.fit(x_train,y_train)
    j=l.predict(x_train)
    k=l.predict(x_test)
    m=accuracy_score(y_train,j)
    n=f1_score(y_train,j,average='weighted')
    o=precision_score(y_train,j)
    p=recall_score(y_train,i)
```

completed at 11:28 AM

11:29 08-12-2024

Chat | dhineshkannan65... x Machine Learning - Colab x ML practice - Colab x colsample_bytree - Google x GradientBoostingRegressor x

colab.research.google.com/drive/1D1Dhy_mW7lfh5lQdniTEAqy7GJc1dCp9#scrollTo=GzzBld_T2Mda

ML practice

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample_data
- Travel.csv
- cardexho_imputed.csv

```
print('\n')

RandomForestClassifier
trainingset
accuracy_score 1.0
f1_score 1.0
precision_score 1.0
recall_score 1.0
roc_auc_score 1.0
-----
testset
accuracy_score 0.9052488070892979
f1_score 0.8957485409179543
precision_score 0.9248554913294798
recall_score 0.5594405594405595
roc_auc_score 0.7742164693900512
=====

GradientBoostingClassifier
trainingset
accuracy_score 0.8921368021046477
f1_score 0.8793707640794375
precision_score 0.8753541076487252
recall_score 0.48738170347003157
roc_auc_score 0.7357970591264762
-----
testset
```

0s completed at 11:28 AM

75.09 GB available

Search

ENG IN 11:29 08-12-2024

Machine Learning - Colab ML practice - Colab

colab.research.google.com/drive/1DlDhy_mW7lfh5lQdniTEAqy7GJc1dCp9#scrollTo=GzzBld_T2Mda

ML practice

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample_data
- Travel.csv
- cardekho_imputed.csv

```
[90] g={'learning_rate':[0.1,0.01],'max_depth':[5,8,12,20,30],'n_estimators':[100,200,300],'colsample_bytree':[0.5,0.8,1,0.3,0.4]}
      rf={'max_depths':[5,8,15,None,10],'max_features':[5,7,'auto',8],'min_samples_split':[2,8,15,20],'n_estimators':[100,200,500,1000]}

[92] la=[('gb',GradientBoostingClassifier(),g),('rf',RandomForestClassifier(),rf)]

[101] al={'RandomForestClassifier':RandomForestClassifier(max_depth=None,max_features=5,min_samples_split=2,n_estimators=100),
       'GradientBoostingClassifier':GradientBoostingClassifier(learning_rate=0.1,max_depth=5,n_estimators=100)}

for ii in range(len(al)):
    l=list(al.values())[ii]
    l.fit(x_train,y_train)
    j=l.predict(x_train)
    k=l.predict(x_test)
    m=accuracy_score(y_train,j)
    n=f1_score(y_train,j,average='weighted')
    o=precision_score(y_train,j)
    p=recall_score(y_train,j)
    q=roc_auc_score(y_train,j)
    mm=accuracy_score(y_test,k)
    nn=f1_score(y_test,k,average='weighted')
```

completed at 11:28 AM

75.09 GB available

11:29 08-12-2024

Chat | dhineshkannan65... x Machine Learning - Colab x ML practice - Colab x colsample_bytree - Google x GradientBoostingRegressor x

colab.research.google.com/drive/1DlDhy_mW7lfh5lQdniTEAqy7GJc1dCp9#scrollTo=GzzBld_T2Mda

ML practice

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample_data
- Travel.csv
- cardekho_imputed.csv

```
print('recall_score',pp)
print('roc_auc_score',qq)
print('-'*35)
print('\n')

RandomForestClassifier
trainingset
accuracy_score 1.0
f1_score 1.0
precision_score 1.0
recall_score 1.0
roc_auc_score 1.0
-----
testset
accuracy_score 0.907293796864349
f1_score 0.8972287483897368
precision_score 0.9518072289156626
recall_score 0.5524475524475524
roc_auc_score 0.7728368160205586
=====

GradientBoostingClassifier
trainingset
accuracy_score 0.9611224788073662
f1_score 0.9595766043679971
precision_score 0.9826589595375722
recall_score 0.804416403785489
```

0s completed at 11:28 AM

75.09 GB available

11:29 08-12-2024

Machine Learning - Colab ML practice - Colab colsample_bytree - Google GradientBoostingRegressor

colab.research.google.com/drive/1DlDhy_mW7lfh5lQdniTEAqy7GJc1dCp9#scrollTo=GzzBld_T2Mda

ML practice

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample_data
- Travel.csv
- cardexho_imputed.csv

```
import matplotlib.pyplot as plt
from sklearn.metrics import roc_curve
plt.figure()
dd=[{'label': 'GradientBoostingClassifier', 'model': GradientBoostingClassifier(learning_rate=0.1, max_depth=5, n_estimators=100), 'auc': 0.95}]
for roo in dd:
    ghf=roo['model']
    ghf.fit(x_train, y_train)
    fpr, tpr, thresholds=roc_curve(y_test, ghf.predict_proba(x_test)[:,1])
    plt.plot(fpr, tpr, label='%s roc(area=%0.2f)'%(roo['label'], roo['auc']))
    plt.plot([0,1],[0,1], 'r--')
    plt.xlim([0.0,1.0])
    plt.ylim([0.0,1.05])
    plt.xlabel('x')
    plt.ylabel('y')
    plt.legend()
    plt.show()
```

0s completed at 11:28 AM

75.09 GB available

11:29 08-12-2024

ML practice

File Edit View Insert Runtime Tools Help All changes saved

Files

sample_data

Travel.csv

cardekho_imputed.csv

+ Code + Text

plt.show()

GradientBoostingClassifier roc(area=0.76)

0s completed at 11:28 AM

Search

11:29 08-12-2024

Chat | dhineshkannan65...Machine Learning - ColabML practice - Colabcolsample_bytree - GoogleGradientBoostingRegres...

colab.research.google.com/drive/1DlDhy_mW7lfh5lQdniTEAqy7GJc1dCp9#scrollTo=GzzBld_T2Mda

ML practice

File Edit View Insert Runtime Tools Help All changes saved

Files

sample_dataTravel.csvcardekho_imputed.csv

75.09 GB available

+ Code + Text

```
import pandas as pd
a=pd.read_csv('cardekho_imputed.csv')
a
```

	Unnamed: 0	car_name	brand	model	min_cost_price	max_cost_price	vehicle_age	km_driven	seller_type	fuel_type	trans
0	0	Maruti Alto	Maruti	Alto	3.570039e+05	4.654015e+05	9	120000	Individual	Petrol	
1	1	Hyundai Grand	Hyundai	Grand	7.110000e+05	7.480000e+05	5	20000	Individual	Petrol	
2	2	Hyundai i20	Hyundai	i20	8.540829e+05	1.307926e+06	11	60000	Individual	Petrol	
3	3	Maruti Alto	Maruti	Alto	3.570039e+05	4.654015e+05	9	37000	Individual	Petrol	
4	4	Ford Ecosport	Ford	Ecosport	1.014000e+06	1.379000e+06	6	30000	Dealer	Diesel	
...
15406	19637	Hyundai i10	Hyundai	i10	6.983200e+05	7.293333e+05	9	10723	Dealer	Petrol	

0s completed at 11:28 AM

Search

11:29 08-12-2024

ML practice

File Edit View Insert Runtime Tools Help All changes saved

Files

sample_data
Travel.csv
cardekho_imputed.csv

+ Code + Text

15410 19543 City Honda City 1.3309/8e+08 1.70/43/e+06 2 13000 Dealer Petrol

15411 rows x 16 columns

Next steps: View recommended plots New interactive sheet

[106] a.drop(['Unnamed: 0', 'car_name', 'brand'],axis=1,inplace=True)

b=a.drop('selling_price',axis=1)
b

	model	min_cost_price	max_cost_price	vehicle_age	km_driven	seller_type	fuel_type	transmission_type	mileage	engine
0	Alto	3.570039e+05	4.654015e+05	9	120000	Individual	Petrol	Manual	19.70	796
1	Grand	7.110000e+05	7.480000e+05	5	20000	Individual	Petrol	Manual	18.90	1197
2	i20	8.540829e+05	1.307926e+06	11	60000	Individual	Petrol	Manual	17.00	1197
3	Alto	3.570039e+05	4.654015e+05	9	37000	Individual	Petrol	Manual	20.92	998
4	Ecosport	1.014000e+06	1.379000e+06	6	30000	Dealer	Diesel	Manual	22.77	1498

Disk 75.09 GB available

0s completed at 11:28 AM

Chat | dhineshkannan65...Machine Learning - ColabML practice - Colabcolsample_bytree - GoogGradientBoostingRegres...

colab.research.google.com/drive/1DlDhy_mW7lfh5lQdniTEAqy7GJc1dCp9#scrollTo=GzzBld_T2Mda

ML practice

File Edit View Insert Runtime Tools Help All changes saved

Files

sample_dataTravel.csvcardekho_imputed.csv

Disk75.09 GB available

+ Code + Text

Next steps: View recommended plotsNew interactive sheet

```
c=a['selling_price']
```

	selling_price
0	120000
1	550000
2	215000
3	226000
4	570000
...	...
15406	250000
15407	925000
15408	425000
15409	1225000

0s completed at 11:28 AM

Search

11:3008-12-2024

ML practice

File Edit View Insert Runtime Tools Help All changes saved

Files

sample_data

Travel.csv

cardekho_imputed.csv

+ Code + Text

dtype: int64

[138] x_train,x_test,y_train,y_test=train_test_split(b,c,test_size=0.3)

from sklearn.preprocessing import LabelEncoder

z=LabelEncoder()

b['model']=z.fit_transform(b['model'])

b

	model	min_cost_price	max_cost_price	vehicle_age	km_driven	seller_type	fuel_type	transmission_type	mileage	engine
0	7	3.570039e+05	4.654015e+05	9	120000	Individual	Petrol	Manual	19.70	796
1	54	7.110000e+05	7.480000e+05	5	20000	Individual	Petrol	Manual	18.90	1197
2	118	8.540829e+05	1.307926e+06	11	60000	Individual	Petrol	Manual	17.00	1197
3	7	3.570039e+05	4.654015e+05	9	37000	Individual	Petrol	Manual	20.92	998
4	38	1.014000e+06	1.379000e+06	6	30000	Dealer	Diesel	Manual	22.77	1498
...
15406	117	6.983200e+05	7.293333e+05	9	10723	Dealer	Petrol	Manual	19.81	1086
15407	118	8.540829e+05	1.307926e+06	11	60000	Individual	Petrol	Manual	17.00	1197

completed at 11:28 AM

75.09 GB available

11:30 08-12-2024

Machine Learning - Colab ML practice - Colab

colab.research.google.com/drive/1DlDhy_mW7lfh5lQdniTEAqy7GJc1dCp9#scrollTo=GzzBld_T2Mda

ML practice

File Edit View Insert Runtime Tools Help All changes saved

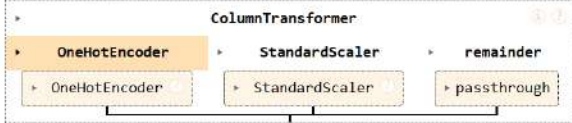
Files

- sample_data
- Travel.csv
- cardexho_imputed.csv

Next steps: View recommended plots New interactive sheet

```
[140] l=b.select_dtypes(exclude='object').columns
      k=['seller_type','fuel_type','transmission_type']
      s=StandardScaler()
      o=OneHotEncoder(drop='first')
      ct=ColumnTransformer([('OneHotEncoder',o,k),('StandardScaler',s,l)],remainder='passthrough')
```

ct



```
[141] b=ct.fit_transform(b)
```

```
[146] x_train,x_test,y_train,y_test=train_test_split(b,c,test_size=0.3)
```

```
[151] from sklearn.ensemble import RandomForestRegressor
```

completed at 11:28 AM

11:30 08-12-2024

Chat | dhineshkannan65... x Machine Learning - Colab x ML practice - Colab x colsample_bytree - Google x GradientBoostingRegressor x +

colab.research.google.com/drive/1DlDhy_mW7lfh5lQdniTEAqy7GJc1dCp9#scrollTo=GzzBld_T2Mda

ML practice

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample_data
- Travel.csv
- cardekho_imputed.csv

```
[141] b=ct.fit_transform(b)

[146] x_train,x_test,y_train,y_test=train_test_split(b,c,test_size=0.3)

from sklearn.ensemble import RandomForestRegressor
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.linear_model import LinearRegression,Lasso,Ridge
from sklearn.tree import DecisionTreeRegressor
from sklearn.neighbors import KNeighborsRegressor
from sklearn.metrics import r2_score,mean_squared_error,mean_absolute_error

[152] import numpy as np
def create(x,y):
    g=mean_squared_error(x,y)
    h=mean_absolute_error(x,y)
    i=r2_score(x,y)
    j=np.sqrt(mean_squared_error(x,y))
    return j,h,i

[153] s={'RandomForestRegressor':RandomForestRegressor(),'GradientBoostingRegressor':GradientBoostingRegressor(),'LinearRegression':Lir
```

completed at 11:28 AM

11:30 08-12-2024

Machine Learning - Colab ML practice - Colab

colab.research.google.com/drive/1DlDhy_mW7lfh5lQdniTEAqy7GJc1dCp9#scrollTo=GzzBld_T2Mda

ML practice

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample_data
- Travel.csv
- cardekho_imputed.csv

```
g=mean_squared_error(x,y)
h=mean_absolute_error(x,y)
i=r2_score(x,y)
j=np.sqrt(mean_squared_error(x,y))
return j,h,i
```

```
[153] s={'RandomForestRegressor':RandomForestRegressor(),'GradientBoostingRegressor':GradientBoostingRegressor(),'LinearRegression':Lir
      'KNeighborsRegressor':KNeighborsRegressor(),'DecisionTreeRegressor':DecisionTreeRegressor(),'Lasso':Lasso(),'Ridge':Ridge()}
```

```
for i in range(len(s)):
    gg=list(s.values())[i]
    gg.fit(x_train,y_train)
    p1=gg.predict(x_train)
    p2=gg.predict(x_test)
    m,n,o=create(y_train,p1)
    p,q,r=create(y_test,p2)
    print(list(s.keys())[i])
```

RandomForestRegressor
GradientBoostingRegressor
LinearRegression
KNeighborsRegressor
DecisionTreeRegressor

completed at 11:28 AM

Search

11:30 08-12-2024

Machine Learning - Colab ML practice - Colab colsample_bytree - Google GradientBoostingRegressor

colab.research.google.com/drive/1DlDhy_mW7Ifh5lQdniTEAqy7GJc1dCp9#scrollTo=GzzBld_T2Mda

ML practice

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample_data
- Travel.csv
- cardekho_imputed.csv

Code

```
KNeighborsRegressor
DecisionTreeRegressor
Lasso
Ridge
```

[155] m

520087.9942327551

[156] n

228680.93079013837

o

0.6538437945189679

[158] p

471168.7061891723

[159] q

completed at 11:28 AM

75.09 GB available

11:30 08-12-2024

Machine Learning - Colab ML practice - Colab

colab.research.google.com/drive/1D1Dhy_mW7lfh5lQdniTEAqy7GJc1dCp9#scrollTo=GzzBld_T2Mda

ML practice

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample_data
- Travel.csv
- cardekho_imputed.csv

Code

```
0.7360848932003553
```

```
[174] gr={'loss':['squared_error','huber','absolute_error'],'criterion':['friedman_mse','squared_error'],'min_samples_split':[2,8,15,20],  
      kk={'max_depth':[5,8,15,None,10],'max_features':[5,7,'auto',8],'min_samples_split':[2,8,15,20],'n_estimators':[100,200,500,1000]},  
      jj={'n_estimators':[50,60,70,80],'loss':['linear','square','exponential']}
```

```
[168] st=[('gb',GradientBoostingRegressor(),gr),('rf',RandomForestRegressor(),rf),('kn',KNeighborsRegressor(),'jj')]  
      st
```

```
[('gb',  
  GradientBoostingRegressor(),  
  {'loss':['log_loss','deviance','exponential'],  
   'criterion':['friedman_mse','squared_error'],  
   'min_samples_split':[2,8,15,20],  
   'n_estimators':[100,200,500],  
   'max_depth':[5,8,15,None,10]}),  
 ('rf',  
  RandomForestRegressor(),  
  {'max_depths':[5,8,15,None,10],  
   'max_features':[5,7,'auto',8],  
   'min_samples_split':[2,8,15,20],  
   'n_estimators':[100,200,500,1000]}),  
 ('kn',KNeighborsRegressor(),'jj')]
```

0s completed at 11:28 AM

75.09 GB available

11:30 08-12-2024

Machine Learning - Colab ML practice - Colab colsample_bytree - Google GradientBoostingRegressor

colab.research.google.com/drive/1DlDhy_mW7lfh5lQdniTEAqy7GJc1dCp9#scrollTo=GzzBld_T2Mda

ML practice

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample_data
- Travel.csv
- cardekho_imputed.csv

```
[176] v={'GradientBoostingRegressor':GradientBoostingRegressor(loss='squared_error',criterion='squared_error',min_samples_split=2,n_estimators=100,max_depth=5),'RandomForestRegressor':RandomForestRegressor(n_estimators=100,max_depth=None,max_features=5,min_samples
```

```
[177] for ii in range(len(v)):  
      gg=list(v.values())[ii]  
      gg.fit(x_train,y_train)  
      p1=gg.predict(x_train)  
      p2=gg.predict(x_test)  
      m,n=create(y_train,p1)  
      p,q,r=create(y_test,p2)  
      print(list(v.keys())[ii])
```

GradientBoostingRegressor
RandomForestRegressor
KNNNeighbors

```
[178] m
```

```
384116.3036442107
```

```
[179] n
```

completed at 11:28 AM

75.09 GB available

11:30 08-12-2024

Chat | dhineshkannan65... x Machine Learning - Colab x ML practice - Colab x colsample_bytree - Google x GradientBoostingRegres... x

colab.research.google.com/drive/1DlDhy_mW7Ifh5lQdniTEAqy7GJc1dCp9#scrollTo=GzzBld_T2Mda

ML practice

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample_data
- Travel.csv
- cardekho_imputed.csv

+ Code + Text

```
[179] n
103047.54449800686

[180] o
0.8111817595041859

[181] p
360457.69239219144

q
119344.27443771626

r
0.8455385986857666
```

0s completed at 11:28 AM

75.09 GB available

11:30 08-12-2024