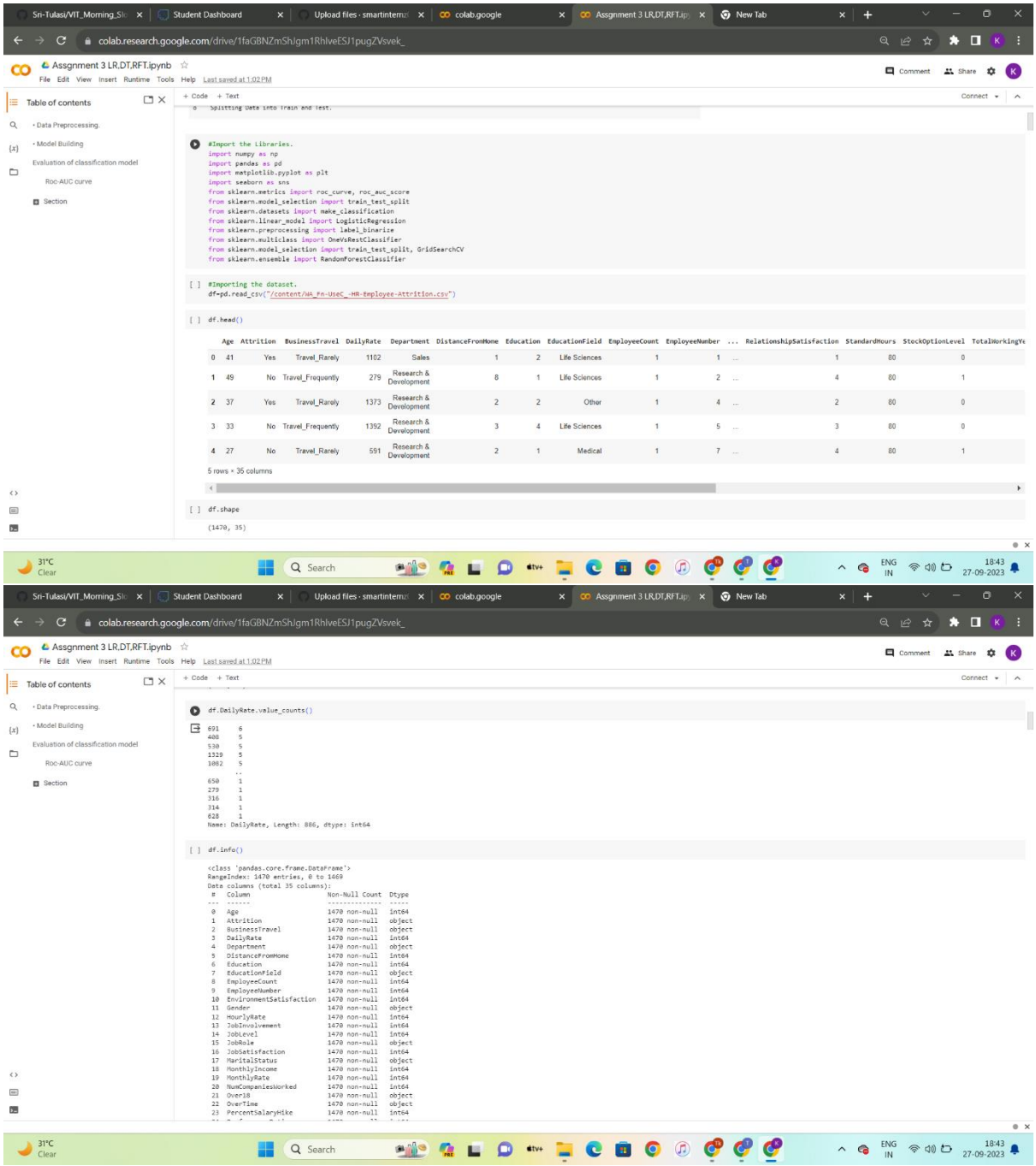
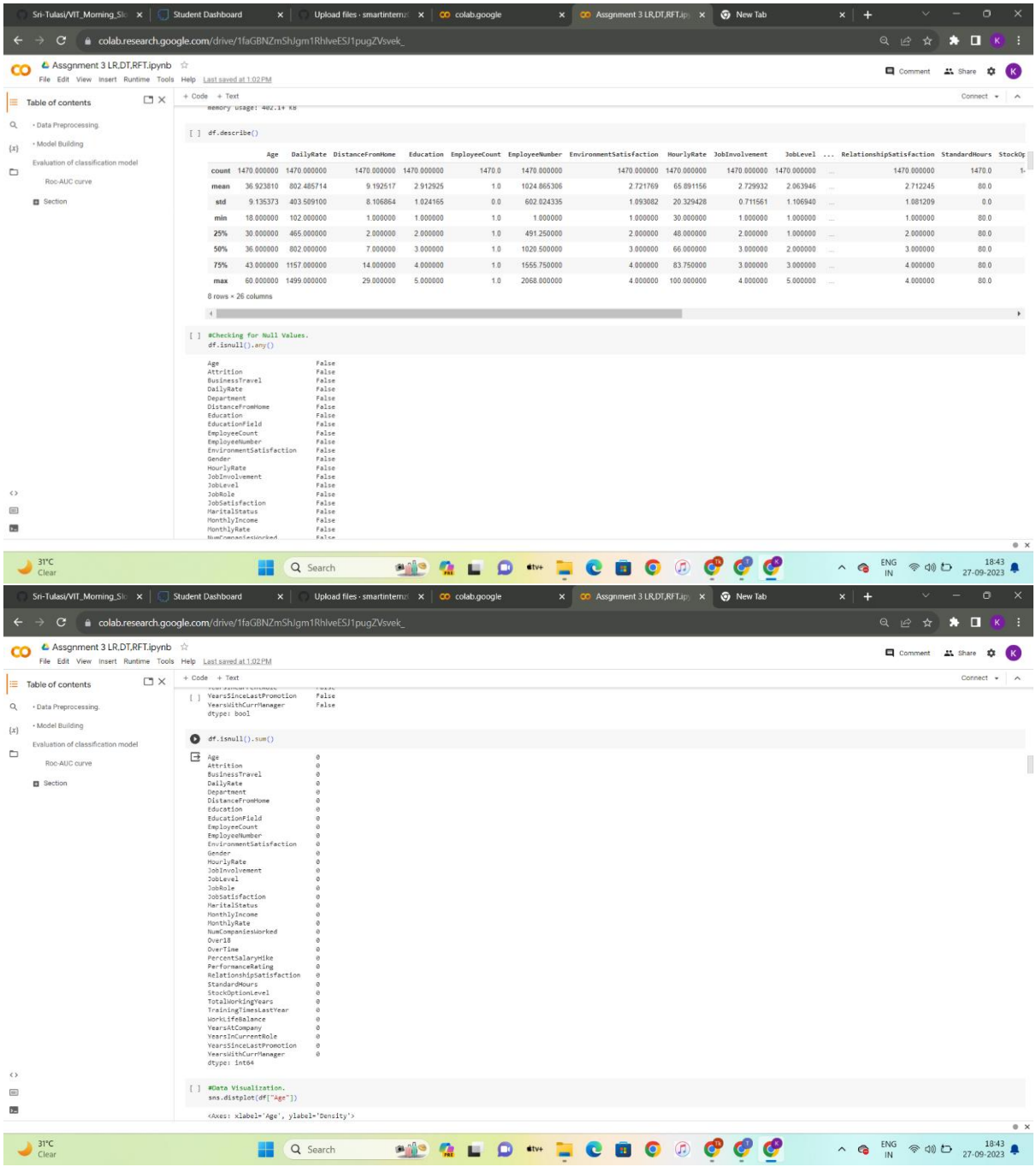
Assgnment_3_LR,DT,RFT

Thota Kankshith

21bce8879




Sri-Tulasi/VIT_Morning_Sl
Student Dashboard
Upload files - smartinternz
colab.google
Assignment 3 LR,DT,RFT.ipy
New Tab
colab.research.google.com/drive/1faGBNZmShJgm1RhlveESJ1pugZVsvek_
Assignment 3 LR,DT,RFT.ipynb
File Edit View Insert Runtime Tools Help Last saved at 1:02 PM
Comment
Share
Table of contents
+ Code + Text
Connect
• Data Preprocessing
• Model Building
Evaluation of classification model
Roc-AUC curve
Section
memory usage: 402.1+ KB
[ ] df.describe()
Age DailyRate DistanceFromHome Education EmployeeCount EmployeeNumber EnvironmentSatisfaction HourlyRate JobInvolvement JobLevel ... RelationshipSatisfaction StandardHours StockOp
count 1470.000000 1470.000000 1470.000000 1470.000000 1470.0 1470.000000 1470.000000 1470.000000 1470.000000 1470.000000 ... 1470.000000 1470.0 1.
mean 36.923810 802.485714 9.192517 2.912925 1.0 1024.865306 2.721769 65.891156 2.729932 2.063946 ... 2.712245 80.0
std 9.135373 403.509100 8.106864 1.024165 0.0 602.024335 1.093082 20.329428 0.711561 1.106940 ... 1.081209 0.0
min 18.000000 102.000000 1.000000 1.000000 1.0 1.000000 1.000000 30.000000 1.000000 1.000000 ... 1.000000 80.0
25% 30.000000 465.000000 2.000000 2.000000 1.0 491.250000 2.000000 48.000000 2.000000 1.000000 ... 2.000000 80.0
50% 36.000000 802.000000 7.000000 3.000000 1.0 1020.500000 3.000000 66.000000 3.000000 2.000000 ... 3.000000 80.0
75% 43.000000 1157.000000 14.000000 4.000000 1.0 1555.750000 4.000000 83.750000 3.000000 3.000000 ... 4.000000 80.0
max 60.000000 1499.000000 29.000000 5.000000 1.0 2068.000000 4.000000 100.000000 4.000000 5.000000 ... 4.000000 80.0
8 rows × 26 columns
[ ] #Checking for Null Values.
df.isnull().any()
Age False
Attrition False
BusinessTravel False
DailyRate False
Department False
DistanceFromHome False
Education False
EducationField False
EmployeeCount False
EmployeeNumber False
EnvironmentSatisfaction False
Gender False
HourlyRate False
JobInvolvement False
JobLevel False
JobRole False
JobSatisfaction False
MaritalStatus False
MonthlyIncome False
MonthlyRate False
31°C
Clear
Search
ENG
IN
18:43
27-09-2023

Sri-Tulasi/VIT_Morning_Sl
Student Dashboard
Upload files - smartinternz
colab.google
Assignment 3 LR,DT,RFT.ipy
New Tab
colab.research.google.com/drive/1faGBNZmShJgm1RhlveESJ1pugZVsvek_
Assignment 3 LR,DT,RFT.ipynb
File Edit View Insert Runtime Tools Help Last saved at 1:02 PM
Comment
Share
Table of contents
+ Code + Text
Connect
• Data Preprocessing
• Model Building
Evaluation of classification model
Roc-AUC curve
Section
YearsSinceLastPromotion False
YearsWithCurrManager False
dtype: bool
df.isnull().sum()
Age 0
Attrition 0
BusinessTravel 0
DailyRate 0
Department 0
DistanceFromHome 0
Education 0
EducationField 0
EmployeeCount 0
EmployeeNumber 0
EnvironmentSatisfaction 0
Gender 0
HourlyRate 0
JobInvolvement 0
JobLevel 0
JobRole 0
JobSatisfaction 0
MaritalStatus 0
MonthlyIncome 0
MonthlyRate 0
NumCompaniesWorked 0
Over18 0
OverTime 0
PercentSalaryHike 0
PerformanceRating 0
RelationshipSatisfaction 0
StandardHours 0
StockOptionLevel 0
TotalWorkingYears 0
TrainingTimesLastYear 0
WorkLifeBalance 0
YearsAtCompany 0
YearsInCurrentRole 0
YearsSinceLastPromotion 0
YearsWithCurrManager 0
dtype: int64
[ ] #Data Visualization.
sns.distplot(df["Age"])
<Axes: xlabel='Age', ylabel='Density'>
31°C
Clear
Search
ENG
IN
18:43
27-09-2023

Sri-Tulasi/VIT_Morning_Sl | Student Dashboard | Upload files - smartinternz | colab.google | Assgnment 3 LR,DT,RFT.ipy | New Tab

colab.research.google.com/drive/1faGBNZmShJgm1RhlveESJ1pugZVsvek_

Assgnment 3 LR,DT,RFT.ipynb

File Edit View Insert Runtime Tools Help Last saved at 1:02 PM

Comment Share

Table of contents

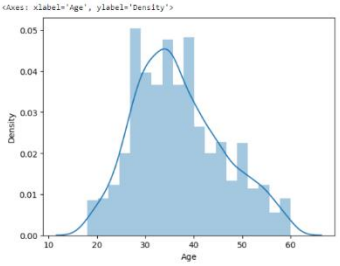
- Data Preprocessing.
- Model Building
- Evaluation of classification model
  - Roc-AUC curve
- Section

+ Code + Text

Connect

```
[ ] sns.distplot(df["Age"])
```

<Axes: xlabel='Age', ylabel='Density'>




```
[ ] df.corr()
```

	Age	DailyRate	DistanceFromHome	Education	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	HourlyRate	JobInvolvement	JobLevel	...	RelationshipSatisfaction	StandardHours
Age	1.000000	0.010661	-0.001686	0.208034	NaN	-0.010145	0.010146	0.024287	0.029820	0.509604	...	0.053535	NaN
DailyRate	0.010661	1.000000	-0.004985	-0.016806	NaN	-0.050990	0.018355	0.023381	0.046135	0.002966	...	0.007846	NaN
DistanceFromHome	-0.001686	-0.004985	1.000000	0.021042	NaN	0.032916	-0.016075	0.031131	0.008783	0.005303	...	0.006557	NaN
Education	0.208034	-0.016806	0.021042	1.000000	NaN	0.042070	-0.027128	0.016775	0.042438	0.101589	...	-0.009118	NaN
EmployeeCount	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN
EmployeeNumber	-0.010145	-0.050990	0.032916	0.042070	NaN	1.000000	0.017621	0.035179	-0.006888	-0.018519	...	-0.069861	NaN
EnvironmentSatisfaction	0.010146	0.018355	-0.016075	-0.027128	NaN	0.017621	1.000000	-0.049857	-0.008278	0.001212	...	0.007665	NaN
HourlyRate	0.024287	0.023381	0.031131	0.016775	NaN	0.035179	-0.049857	1.000000	0.042861	-0.027853	...	0.001330	NaN
JobInvolvement	0.029820	0.046135	0.008783	0.042438	NaN	-0.006888	-0.008278	0.042861	1.000000	-0.012630	...	0.034297	NaN

31°C Clear | Search | ENG IN | 18:43 27-09-2023

Sri-Tulasi/VIT_Morning_Sl | Student Dashboard | Upload files - smartinternz | colab.google | Assgnment 3 LR,DT,RFT.ipy | New Tab

colab.research.google.com/drive/1faGBNZmShJgm1RhlveESJ1pugZVsvek_

Assgnment 3 LR,DT,RFT.ipynb

File Edit View Insert Runtime Tools Help Last saved at 1:02 PM

Comment Share

Table of contents

- Data Preprocessing.
- Model Building
- Evaluation of classification model
  - Roc-AUC curve
- Section

+ Code + Text

Connect

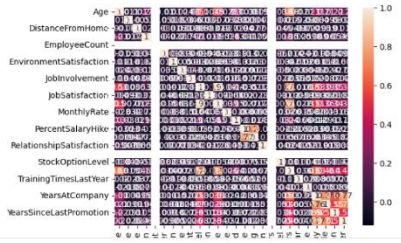
```
df.head()
```

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	...	RelationshipSatisfaction	StandardHours	StockOptionLevel	TotalWorkingYe
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	...	1	80	0	
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	...	4	80	1	
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	...	2	80	0	
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	...	3	80	0	
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	...	4	80	1	

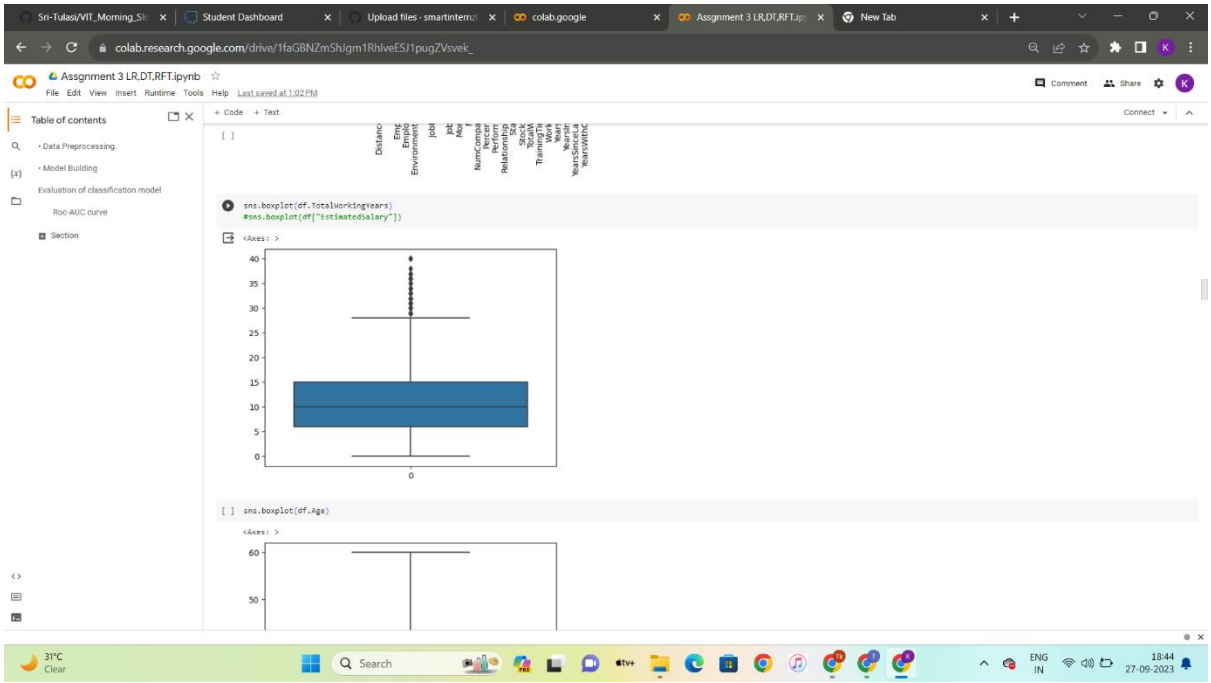
5 rows × 35 columns

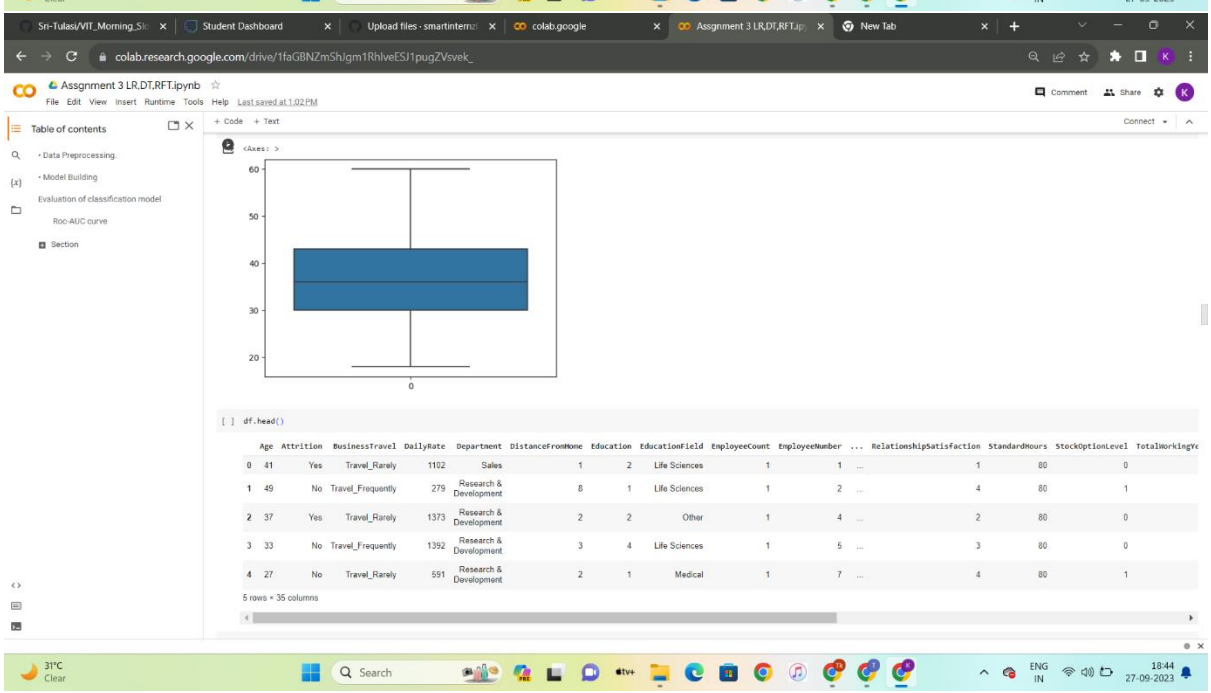
```
[ ] sns.heatmap(df.corr(),annot=True)
```

<Axes: >



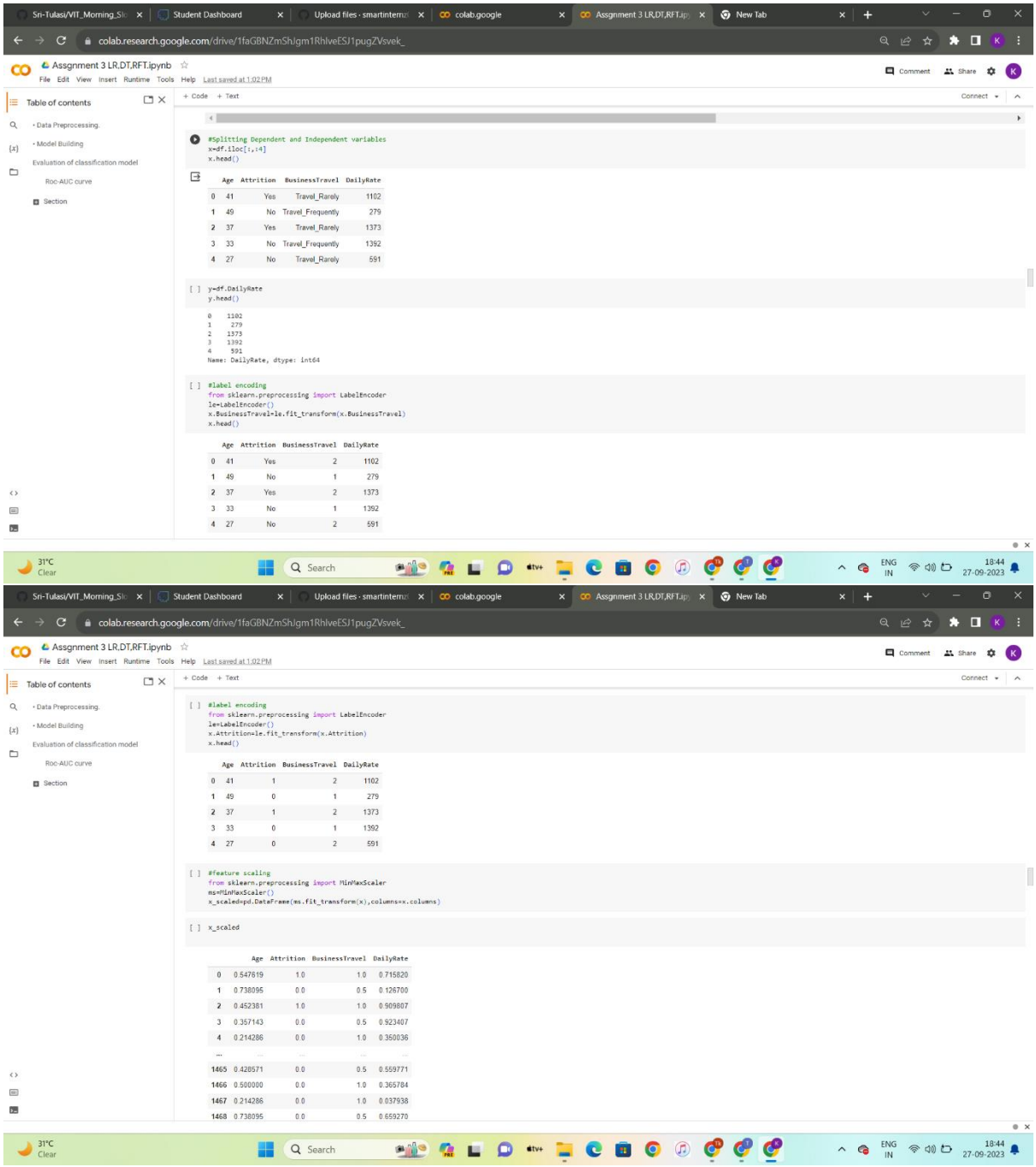
31°C Clear | Search | ENG IN | 18:44 27-09-2023



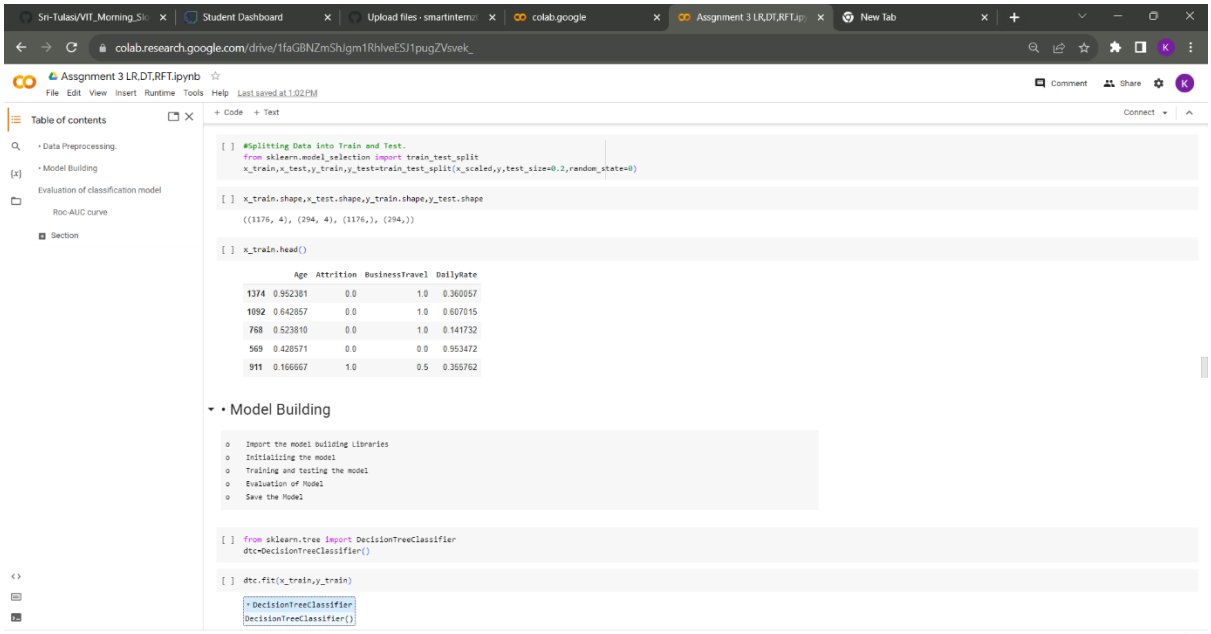


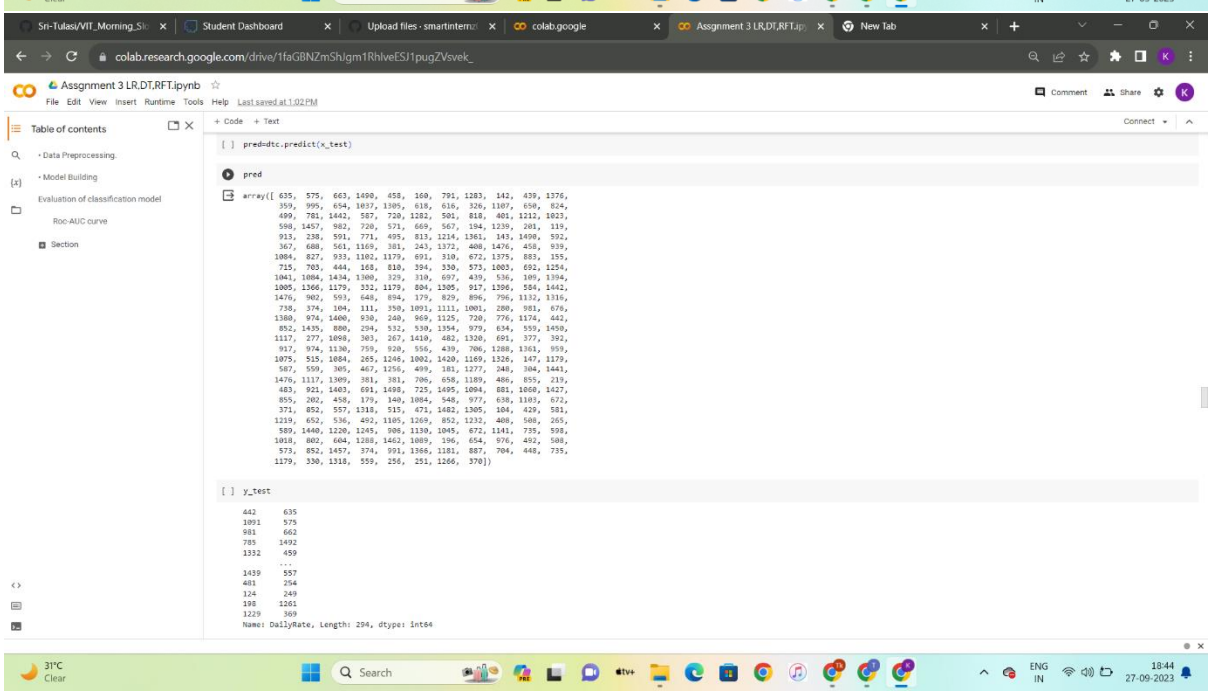
	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	...	RelationshipSatisfaction	StandardHours	StockOptionLevel	TotalWorkingYe
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	...	1	80	0	
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	...	4	80	1	
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	...	2	80	0	
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	...	3	80	0	
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	...	4	80	1	

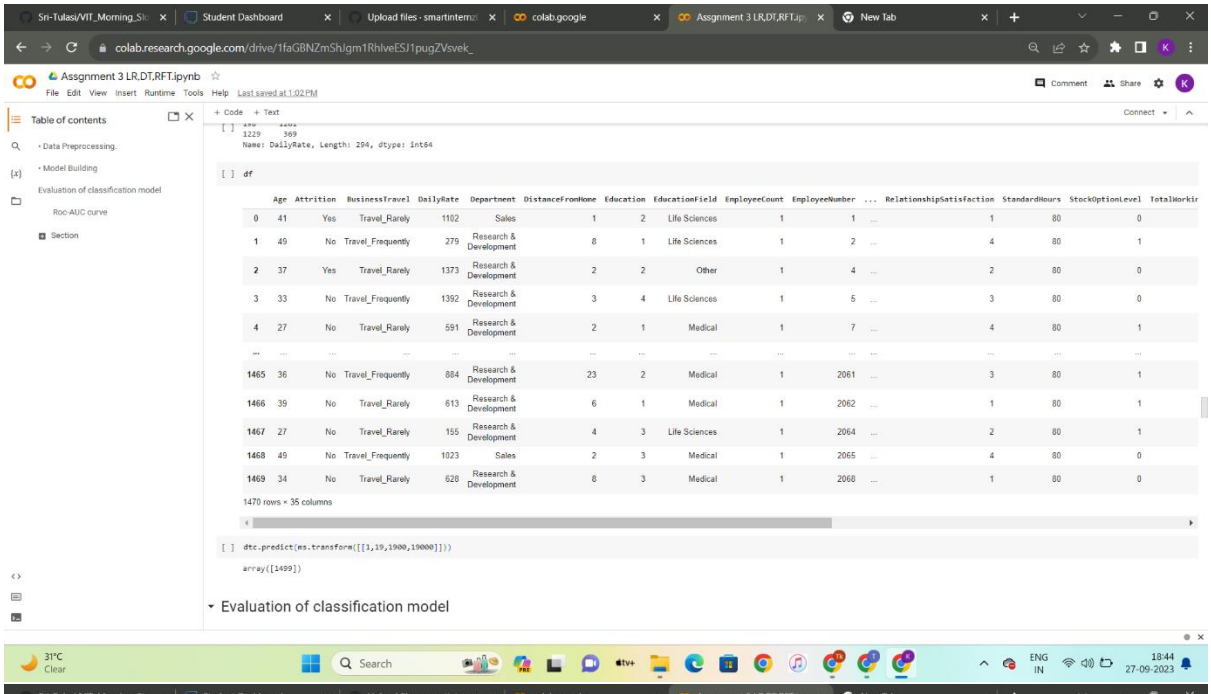
5 rows × 35 columns

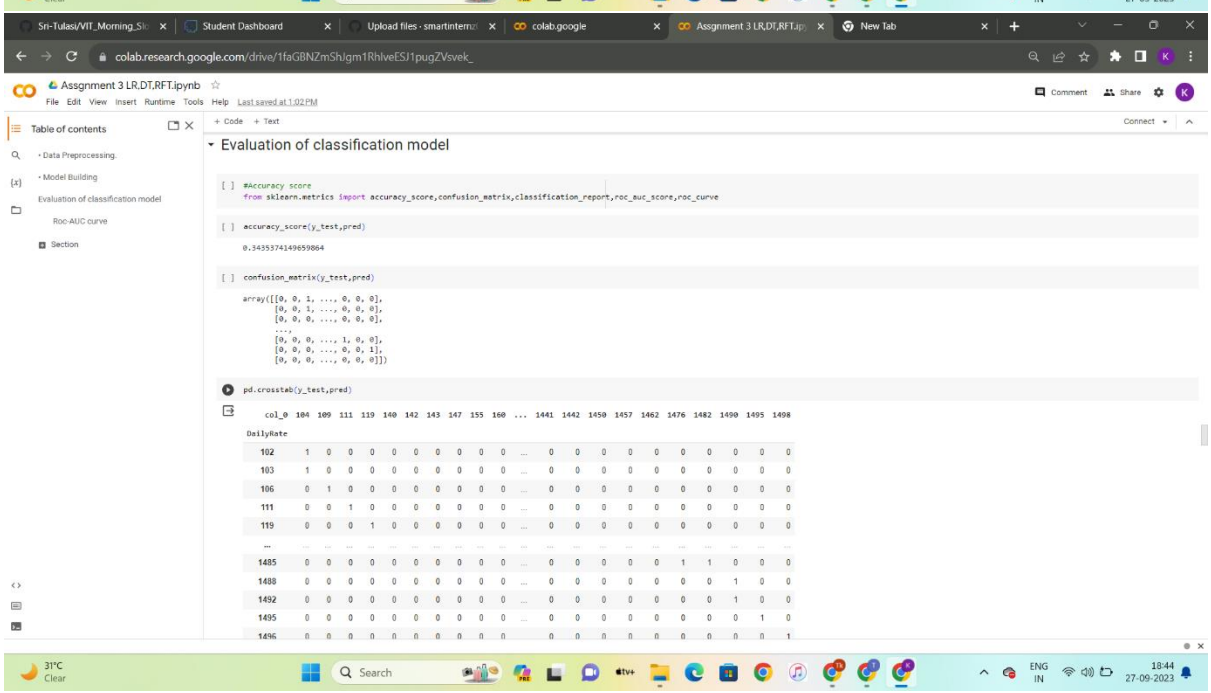

Sri-Tulasi/VIT_Morning_Sk
Student Dashboard
Upload files - smartinternz
colab.google
Assgnment 3 LR,DT,RFT.ipy
New Tab
colab.research.google.com/drive/1faGBNZmShJgm1RhlveESJ1pugZVsvek_
Assgnment 3 LR,DT,RFT.ipynb
File Edit View Insert Runtime Tools Help Last saved at 1:02 PM
Comment
Share
Table of contents
+ Code + Text
Connect
• Data Preprocessing.
• Model Building
Evaluation of classification model
Roc-AUC curve
Section
#Splitting Dependent and Independent variables
x=df.iloc[:,:4]
x.head()
Age Attrition BusinessTravel DailyRate
0 41 Yes Travel_Rarely 1102
1 49 No Travel_Frequently 279
2 37 Yes Travel_Rarely 1373
3 33 No Travel_Frequently 1392
4 27 No Travel_Rarely 591
y=df.DailyRate
y.head()
0 1102
1 279
2 1373
3 1392
4 591
Name: DailyRate, dtype: int64
#label encoding
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
x.BusinessTravel=le.fit_transform(x.BusinessTravel)
x.head()
Age Attrition BusinessTravel DailyRate
0 41 Yes 2 1102
1 49 No 1 279
2 37 Yes 2 1373
3 33 No 1 1392
4 27 No 2 591
31°C
Clear
Search
ENG
IN
18:44
27-09-2023

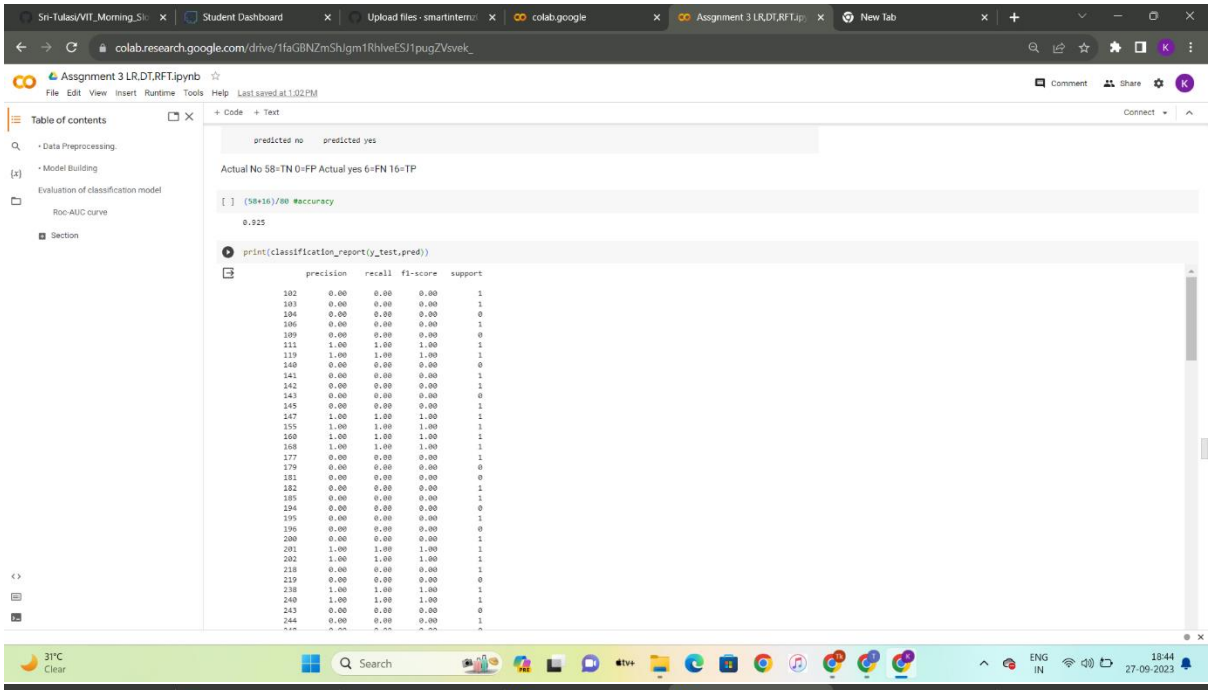
Sri-Tulasi/VIT_Morning_Sk
Student Dashboard
Upload files - smartinternz
colab.google
Assgnment 3 LR,DT,RFT.ipy
New Tab
colab.research.google.com/drive/1faGBNZmShJgm1RhlveESJ1pugZVsvek_
Assgnment 3 LR,DT,RFT.ipynb
File Edit View Insert Runtime Tools Help Last saved at 1:02 PM
Comment
Share
Table of contents
+ Code + Text
Connect
• Data Preprocessing.
• Model Building
Evaluation of classification model
Roc-AUC curve
Section
#label encoding
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
x.Attrition=le.fit_transform(x.Attrition)
x.head()
Age Attrition BusinessTravel DailyRate
0 41 1 2 1102
1 49 0 1 279
2 37 1 2 1373
3 33 0 1 1392
4 27 0 2 591
#feature scaling
from sklearn.preprocessing import MinMaxScaler
ms=MinMaxScaler()
x_scaled=pd.DataFrame(ms.fit_transform(x),columns=x.columns)
x_scaled
Age Attrition BusinessTravel DailyRate
0 0.547619 1.0 1.0 0.715820
1 0.738095 0.0 0.5 0.126700
2 0.452381 1.0 1.0 0.909807
3 0.357143 0.0 0.5 0.923407
4 0.214286 0.0 1.0 0.350036
1465 0.428571 0.0 0.5 0.559771
1466 0.500000 0.0 1.0 0.365784
1467 0.214286 0.0 1.0 0.037938
1468 0.738095 0.0 0.5 0.659270
31°C
Clear
Search
ENG
IN
18:44
27-09-2023

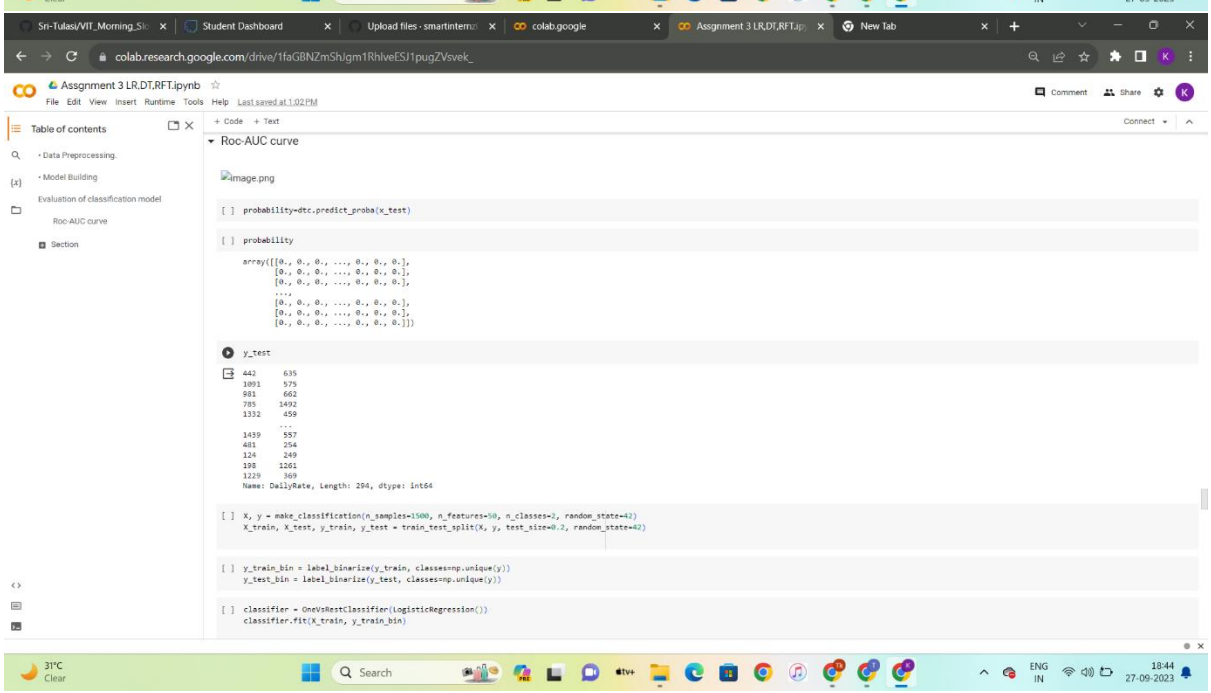

Sri-Tulasi/VIT_Morning_Sk
Student Dashboard
Upload files - smartinternz
colab.google
Assgnment 3 LR,DT,RFT.ipy
New Tab
colab.research.google.com/drive/1faGBNZmShJgm1RhlveESJ1pugZVsvek_
Assgnment 3 LR,DT,RFT.ipynb
File Edit View Insert Runtime Tools Help Last saved at 1:02 PM
Comment
Share
Table of contents
+ Code + Text
Connect
• Data Preprocessing.
• Model Building
Evaluation of classification model
Roc-AUC curve
Section
[ ] #Splitting Data into Train and Test.
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x_scaled,y,test_size=0.2,random_state=0)
[ ] x_train.shape,x_test.shape,y_train.shape,y_test.shape
((1176, 4), (294, 4), (1176,), (294,))
[ ] x_train.head()
Age Attrition BusinessTravel DailyRate
1374 0.952381 0.0 1.0 0.360057
1092 0.642857 0.0 1.0 0.607015
768 0.523810 0.0 1.0 0.141732
569 0.428571 0.0 0.0 0.953472
911 0.166667 1.0 0.5 0.355762
• Model Building
o Import the model building Libraries
o Initializing the model
o Training and testing the model
o Evaluation of Model
o Save the Model
[ ] from sklearn.tree import DecisionTreeClassifier
dtc=DecisionTreeClassifier()
[ ] dtc.fit(x_train,y_train)
DecisionTreeClassifier
DecisionTreeClassifier()
31°C
Clear
Search
ENG
IN
18:44
27-09-2023


Sri-Tulasi/VIT_Morning_Sk
Student Dashboard
Upload files - smartinternz
colab.google
Assgnment 3 LR,DT,RFT.ipy
New Tab
colab.research.google.com/drive/1faGBNZmShJgm1RhlveESJ1pugZVsvek_
Assgnment 3 LR,DT,RFT.ipynb
File Edit View Insert Runtime Tools Help Last saved at 1:02 PM
Comment
Share
Table of contents
+ Code + Text
Connect
• Data Preprocessing.
• Model Building
Evaluation of classification model
Roc-AUC curve
Section
[ ] pred=dtc.predict(x_test)
pred
array([635, 575, 663, 1490, 458, 160, 791, 1283, 142, 439, 1376,
359, 995, 654, 1037, 1305, 618, 616, 326, 1107, 650, 824,
499, 781, 1442, 587, 720, 1282, 501, 818, 401, 1212, 1023,
598, 1457, 982, 720, 571, 669, 567, 194, 1239, 201, 119,
913, 238, 591, 771, 495, 813, 1214, 1361, 143, 1490, 592,
367, 608, 561, 1169, 381, 243, 1372, 408, 1476, 458, 939,
1084, 827, 933, 1102, 1179, 691, 310, 672, 1375, 883, 155,
715, 703, 444, 168, 810, 394, 330, 573, 1003, 692, 1254,
1041, 1084, 1434, 1300, 329, 310, 697, 439, 536, 109, 1394,
1005, 1366, 1179, 552, 1179, 804, 1305, 917, 1396, 584, 1442,
1476, 902, 593, 648, 894, 179, 829, 896, 796, 1132, 1316,
738, 374, 104, 111, 350, 1091, 1111, 1001, 280, 981, 676,
1380, 974, 1400, 930, 240, 969, 1125, 720, 776, 1174, 442,
852, 1435, 880, 294, 532, 530, 1354, 979, 634, 559, 1450,
1117, 277, 1098, 303, 267, 1410, 482, 1320, 691, 377, 392,
917, 974, 1130, 759, 920, 556, 439, 706, 1288, 1361, 959,
1075, 515, 1084, 265, 1246, 1002, 1420, 1169, 1326, 147, 1179,
587, 559, 305, 467, 1256, 499, 181, 1277, 348, 304, 1441,
1476, 1117, 1309, 381, 381, 706, 658, 1189, 486, 855, 219,
483, 921, 1403, 691, 1498, 725, 1495, 1094, 881, 1060, 1427,
855, 202, 458, 179, 140, 1084, 548, 977, 638, 1103, 672,
371, 852, 557, 1318, 515, 471, 1482, 1305, 104, 429, 581,
1219, 652, 536, 492, 1105, 1269, 852, 1232, 408, 508, 265,
589, 1440, 1220, 1245, 906, 1139, 1045, 672, 1141, 735, 598,
1018, 802, 604, 1288, 1462, 1089, 196, 654, 976, 492, 508,
573, 852, 1457, 374, 991, 1366, 1181, 887, 704, 448, 735,
1179, 330, 1318, 559, 256, 251, 1266, 370])
[ ] y_test
442 635
1091 575
981 662
785 1492
1332 459
...
1439 557
481 254
124 249
198 1261
1229 369
Name: DailyRate, Length: 294, dtype: int64
31°C
Clear
Search
ENG
IN
18:44
27-09-2023


Assgnment 3 LR,DT,RFT.ipynb
Table of contents
• Data Preprocessing.
• Model Building
Evaluation of classification model
Roc-AUC curve
Section
Name: DailyRate, Length: 294, dtype: int64
df
Age Attrition BusinessTravel DailyRate Department DistanceFromHome Education EducationField EmployeeCount EmployeeNumber ... RelationshipSatisfaction StandardHours StockOptionLevel TotalWorkir
0 41 Yes Travel_Rarely 1102 Sales 1 2 Life Sciences 1 1 ... 1 80 0
1 49 No Travel_Frequently 279 Research & Development 8 1 Life Sciences 1 2 ... 4 80 1
2 37 Yes Travel_Rarely 1373 Research & Development 2 2 Other 1 4 ... 2 80 0
3 33 No Travel_Frequently 1392 Research & Development 3 4 Life Sciences 1 5 ... 3 80 0
4 27 No Travel_Rarely 591 Research & Development 2 1 Medical 1 7 ... 4 80 1
1465 36 No Travel_Frequently 884 Research & Development 23 2 Medical 1 2061 ... 3 80 1
1466 39 No Travel_Rarely 613 Research & Development 6 1 Medical 1 2062 ... 1 80 1
1467 27 No Travel_Rarely 155 Research & Development 4 3 Life Sciences 1 2064 ... 2 80 1
1468 49 No Travel_Frequently 1023 Sales 2 3 Medical 1 2065 ... 4 80 0
1469 34 No Travel_Rarely 628 Research & Development 8 3 Medical 1 2068 ... 1 80 0
1470 rows × 35 columns
dtc.predict(ms.transform([[1,19,1900,19000]]))
array([1499])
Evaluation of classification model


Assgnment 3 LR,DT,RFT.ipynb
Table of contents
• Data Preprocessing.
• Model Building
Evaluation of classification model
Roc-AUC curve
Section
Evaluation of classification model
#Accuracy score
from sklearn.metrics import accuracy_score,confusion_matrix,classification_report,roc_auc_score,roc_curve
accuracy_score(y_test,pred)
0.3435374149659864
confusion_matrix(y_test,pred)
array([[0, 0, 1, ..., 0, 0, 0],
[0, 0, 1, ..., 0, 0, 0],
[0, 0, 0, ..., 0, 0, 0],
...,
[0, 0, 0, ..., 1, 0, 0],
[0, 0, 0, ..., 0, 0, 1],
[0, 0, 0, ..., 0, 0, 0]])
pd.crosstab(y_test,pred)
col_0 104 109 111 119 140 142 143 147 155 160 ... 1441 1442 1450 1457 1462 1476 1482 1490 1495 1498
DailyRate
102 1 0 0 0 0 0 0 0 0 0 ... 0 0 0 0 0 0 0 0 0 0
103 1 0 0 0 0 0 0 0 0 0 ... 0 0 0 0 0 0 0 0 0 0
106 0 1 0 0 0 0 0 0 0 0 ... 0 0 0 0 0 0 0 0 0 0
111 0 0 1 0 0 0 0 0 0 0 ... 0 0 0 0 0 0 0 0 0 0
119 0 0 0 1 0 0 0 0 0 0 ... 0 0 0 0 0 0 0 0 0 0
1485 0 0 0 0 0 0 0 0 0 0 ... 0 0 0 0 0 1 1 0 0 0
1488 0 0 0 0 0 0 0 0 0 0 ... 0 0 0 0 0 0 0 1 0 0
1492 0 0 0 0 0 0 0 0 0 0 ... 0 0 0 0 0 0 0 1 0 0
1495 0 0 0 0 0 0 0 0 0 0 ... 0 0 0 0 0 0 0 0 1 0
1496 0 0 0 0 0 0 0 0 0 0 ... 0 0 0 0 0 0 0 0 0 1


Sri-Tulasi/VIT_Morning_Sh
Student Dashboard
Upload files - smartinternz
colab.google
Assgnment 3 LR,DT,RFT.ipy
New Tab
colab.research.google.com/drive/1faGBNZmShJgm1RhlveESJ1pugZVsvek_
Assgnment 3 LR,DT,RFT.ipynb
File Edit View Insert Runtime Tools Help Last saved at 1:02 PM
Comment
Share
Table of contents
+ Code + Text
Connect
• Data Preprocessing.
• Model Building
Evaluation of classification model
Roc-AUC curve
Section
predicted no predicted yes
Actual No 58=TN 0=FP Actual yes 6=FN 16=TP
[ ] (58+16)/80 #accuracy
0.925
print(classification_report(y_test,pred))
precision recall f1-score support
102 0.00 0.00 0.00 1
103 0.00 0.00 0.00 1
104 0.00 0.00 0.00 0
106 0.00 0.00 0.00 1
109 0.00 0.00 0.00 0
111 1.00 1.00 1.00 1
119 1.00 1.00 1.00 1
140 0.00 0.00 0.00 0
141 0.00 0.00 0.00 1
142 0.00 0.00 0.00 1
143 0.00 0.00 0.00 0
145 0.00 0.00 0.00 1
147 1.00 1.00 1.00 1
155 1.00 1.00 1.00 1
160 1.00 1.00 1.00 1
168 1.00 1.00 1.00 1
177 0.00 0.00 0.00 1
179 0.00 0.00 0.00 0
181 0.00 0.00 0.00 0
182 0.00 0.00 0.00 1
185 0.00 0.00 0.00 1
194 0.00 0.00 0.00 0
195 0.00 0.00 0.00 1
196 0.00 0.00 0.00 0
200 0.00 0.00 0.00 1
201 1.00 1.00 1.00 1
202 1.00 1.00 1.00 1
218 0.00 0.00 0.00 1
219 0.00 0.00 0.00 0
238 1.00 1.00 1.00 1
240 1.00 1.00 1.00 1
243 0.00 0.00 0.00 0
244 0.00 0.00 0.00 1
31°C
Clear
Search
ENG
IN
18:44
27-09-2023


Sri-Tulasi/VIT_Morning_Sh
Student Dashboard
Upload files - smartinternz
colab.google
Assgnment 3 LR,DT,RFT.ipy
New Tab
colab.research.google.com/drive/1faGBNZmShJgm1RhlveESJ1pugZVsvek_
Assgnment 3 LR,DT,RFT.ipynb
File Edit View Insert Runtime Tools Help Last saved at 1:02 PM
Comment
Share
Table of contents
+ Code + Text
Connect
• Data Preprocessing.
• Model Building
Evaluation of classification model
Roc-AUC curve
Section
Roc-AUC curve
image.png
[ ] probability=dtc.predict_proba(x_test)
[ ] probability
array([[0., 0., 0., ..., 0., 0., 0.],
[0., 0., 0., ..., 0., 0., 0.],
[0., 0., 0., ..., 0., 0., 0.],
...,
[0., 0., 0., ..., 0., 0., 0.],
[0., 0., 0., ..., 0., 0., 0.],
[0., 0., 0., ..., 0., 0., 0.]])
y_test
442 635
1093 575
981 662
785 1492
1332 459
...
1439 557
481 254
124 249
198 1261
1229 369
Name: DailyRate, Length: 294, dtype: int64
[ ] X, y = make_classification(n_samples=1500, n_features=50, n_classes=2, random_state=42)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
[ ] y_train_bin = label_binarize(y_train, classes=np.unique(y))
y_test_bin = label_binarize(y_test, classes=np.unique(y))
[ ] classifier = OneVsRestClassifier(LogisticRegression())
classifier.fit(X_train, y_train_bin)
31°C
Clear
Search
ENG
IN
18:44
27-09-2023

Sri-Tulasi/VIT_Morning_Sl | Student Dashboard | Upload files - smartinternz | colab.google | Assgnment 3 LR,DT,RFT.ipy | New Tab

colab.research.google.com/drive/1faGBNZmShJgm1RhlveESJ1pugZVsvek_

Assgnment 3 LR,DT,RFT.ipynb

File Edit View Insert Runtime Tools Help Last saved at 1:02 PM

Comment Share

Table of contents

- Data Preprocessing.
- Model Building
- Evaluation of classification model
  - Roc-AUC curve
- Section

+ Code + Text

Connect

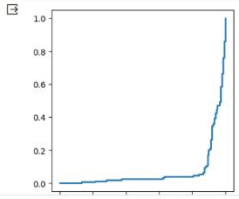
```
classifier = OneVsRestClassifier(LogisticRegression())
classifier.fit(X_train, y_train_bin)
```

OneVsRestClassifier
estimator: LogisticRegression
LogisticRegression

```
y_probs = classifier.predict_proba(X_test)
```

```
fpr = dict()
tpr = dict()
roc_auc = dict()
for i in range(y_train_bin.shape[1]):
    fpr[i], tpr[i], _ = roc_curve(y_test_bin[:, i], y_probs[:, i])
    roc_auc[i] = roc_auc_score(y_test_bin[:, i], y_probs[:, i])
```

```
plt.figure(figsize=(4,4))
for i in range(y_train_bin.shape[1]):
    plt.plot(fpr[i], tpr[i], lw=2, label=f'ROC curve (Class {i}, AUC = {roc_auc[i]:.2f})')
```



31°C Clear | Search | ENG IN | 18:45 27-09-2023

Sri-Tulasi/VIT_Morning_Sl | Student Dashboard | Upload files - smartinternz | colab.google | Assgnment 3 LR,DT,RFT.ipy | New Tab

colab.research.google.com/drive/1faGBNZmShJgm1RhlveESJ1pugZVsvek_

Assgnment 3 LR,DT,RFT.ipynb

File Edit View Insert Runtime Tools Help Last saved at 1:02 PM

Comment Share

Table of contents

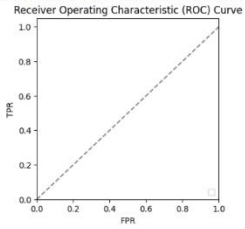
- Data Preprocessing.
- Model Building
- Evaluation of classification model
  - Roc-AUC curve
- Section

+ Code + Text

Connect

```
plt.figure(figsize=(4,4))
plt.plot([0, 1], [0, 1], color='gray', linestyle='--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('FPR')
plt.ylabel('TPR')
plt.title('Receiver Operating Characteristic (ROC) Curve')
plt.legend(loc='lower right')
plt.show()
```

```
WARNING:matplotlib.legend:No artists with labels found to put in legend.  Note that artists whose label start with an underscore are ignored when legend() is called with no argument.
```




```
probability=dtc.predict_proba(x_test)[:,1]
```

Hyper parameter tuning

```
from sklearn import tree
plt.figure(figsize=(25,15))
tree.plot_tree(dtc,filled=True)
```

31°C Clear | Search | ENG IN | 18:45 27-09-2023

Sri-Tulasi/VIT_Morning_Sh | Student Dashboard | Upload files - smartinternz | colab.google | Assgnment 3 LR,DT,RFT.ipy | New Tab

colab.research.google.com/drive/1faGBNZmShJgm1RhlveESJ1pugZVsvek_

Assgnment 3 LR,DT,RFT.ipynb

File Edit View Insert Runtime Tools Help Last saved at 1:02 PM

Comment Share Connect

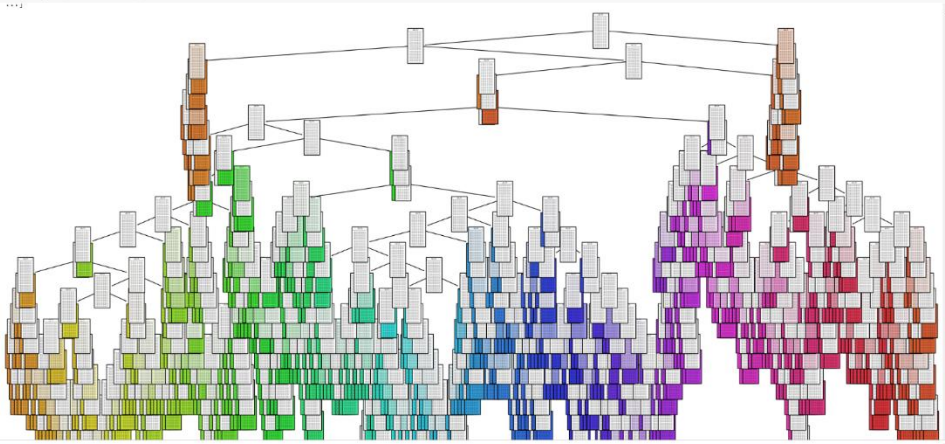
Table of contents

- Data Preprocessing.
- Model Building
- Evaluation of classification model
  - Roc-AUC curve
- Section

+ Code + Text

Hyper parameter tuning

```
[ ] from sklearn import tree
    plt.figure(figsize=(25,15))
    tree.plot_tree(dtc,filled=True)
```



31°C Clear

Search

ENG IN 18:45 27-09-2023

Sri-Tulasi/VIT_Morning_Sh | Student Dashboard | Upload files - smartinternz | colab.google | Assgnment 3 LR,DT,RFT.ipy | New Tab

colab.research.google.com/drive/1faGBNZmShJgm1RhlveESJ1pugZVsvek_

Assgnment 3 LR,DT,RFT.ipynb

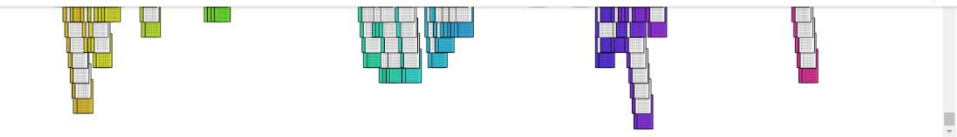
File Edit View Insert Runtime Tools Help Last saved at 1:02 PM

Comment Share Connect

Table of contents

- Data Preprocessing.
- Model Building
- Evaluation of classification model
  - Roc-AUC curve
- Section

+ Code + Text



```
[ ] from sklearn.model_selection import GridSearchCV
    from sklearn.tree import DecisionTreeClassifier
```

```
[ ] # Define the decision tree classifier
    dt_classifier = DecisionTreeClassifier(random_state=42)
```

```
[ ] # Define the parameter grid
    parameter_grid = {
        'criterion': ['gini', 'entropy'],
        'splitter': ['best', 'random'],
        'max_depth': [None, 10, 20, 30, 40],
        'max_features': ['auto', 'sqrt', 'log2']
    }
```

```
[ ] # Create GridSearchCV object
    grid_search = GridSearchCV(dt_classifier, param_grid=parameter_grid, cv=10, scoring='accuracy')
```

```
[ ] # Fit the GridSearchCV to your training data (X_train, y_train)
    grid_search.fit(X_train, y_train)
```

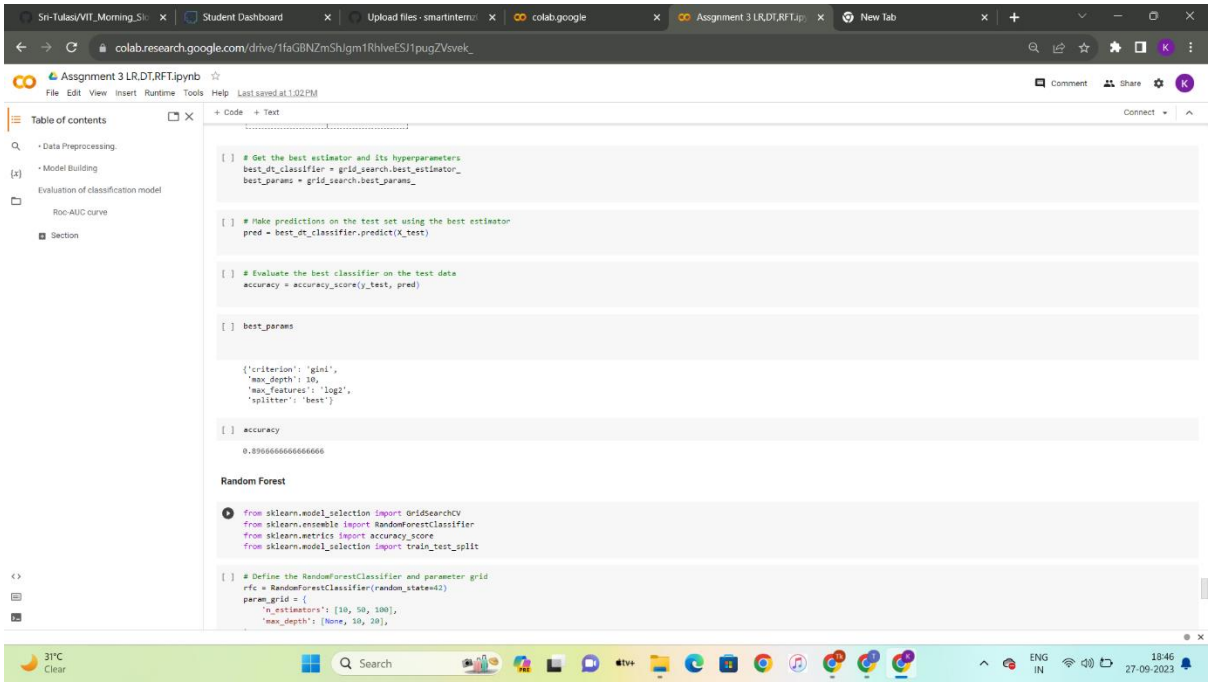
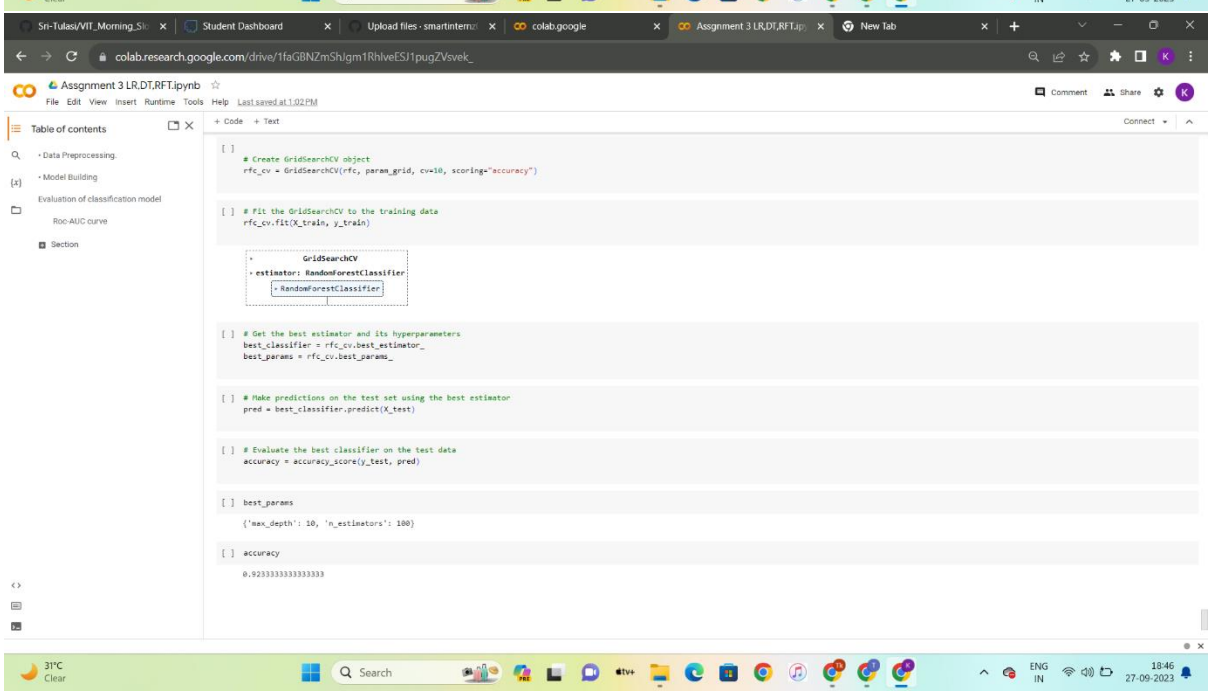
GridSearchCV
estimator: DecisionTreeClassifier
DecisionTreeClassifier

```
[ ] # Get the best estimator and its hyperparameters
```

31°C Clear

Search

ENG IN 18:46 27-09-2023


Sri-Tulasi/VIT_Morning_Sl
Student Dashboard
Upload files - smartinternz
colab.google
Assgnment 3 LR,DT,RFT.ip
New Tab
colab.research.google.com/drive/1faGBNZmShJgm1RhlveESJ1pugZVsvek_
Assgnment 3 LR,DT,RFT.ipynb
File Edit View Insert Runtime Tools Help Last saved at 1:02 PM
Comment
Share
Table of contents
+ Code + Text
Connect
• Data Preprocessing.
• Model Building
Evaluation of classification model
Roc-AUC curve
Section
# Get the best estimator and its hyperparameters
best_dt_classifier = grid_search.best_estimator_
best_params = grid_search.best_params_
# Make predictions on the test set using the best estimator
pred = best_dt_classifier.predict(X_test)
# Evaluate the best classifier on the test data
accuracy = accuracy_score(y_test, pred)
best_params
{'criterion': 'gini',
'max_depth': 10,
'max_features': 'log2',
'splitter': 'best'}
accuracy
0.8966666666666666
Random Forest
from sklearn.model_selection import GridSearchCV
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
from sklearn.model_selection import train_test_split
# Define the RandomForestClassifier and parameter grid
rfc = RandomForestClassifier(random_state=42)
param_grid = {
'n_estimators': [10, 50, 100],
'max_depth': [None, 10, 20],
31°C
Clear
Search
ENG
IN
18:46
27-09-2023

Sri-Tulasi/VIT_Morning_Sl
Student Dashboard
Upload files - smartinternz
colab.google
Assgnment 3 LR,DT,RFT.ip
New Tab
colab.research.google.com/drive/1faGBNZmShJgm1RhlveESJ1pugZVsvek_
Assgnment 3 LR,DT,RFT.ipynb
File Edit View Insert Runtime Tools Help Last saved at 1:02 PM
Comment
Share
Table of contents
+ Code + Text
Connect
• Data Preprocessing.
• Model Building
Evaluation of classification model
Roc-AUC curve
Section
# Create GridSearchCV object
rfc_cv = GridSearchCV(rfc, param_grid, cv=10, scoring="accuracy")
# Fit the GridSearchCV to the training data
rfc_cv.fit(X_train, y_train)
GridSearchCV
estimator: RandomForestClassifier
RandomForestClassifier
# Get the best estimator and its hyperparameters
best_classifier = rfc_cv.best_estimator_
best_params = rfc_cv.best_params_
# Make predictions on the test set using the best estimator
pred = best_classifier.predict(X_test)
# Evaluate the best classifier on the test data
accuracy = accuracy_score(y_test, pred)
best_params
{'max_depth': 10, 'n_estimators': 100}
accuracy
0.9233333333333333
31°C
Clear
Search
ENG
IN
18:46
27-09-2023