LabAssignment\_6 - Jupyter Notebook 07/12/22, 11:11 PM



In [84]: **import** pandas **as** pd

df **=** pd.read\_csv('SmokingDataSet.csv')

df.head()



. . .



In [85]: obj\_list **=** list(df.select\_dtypes(include**=**'object'))

obj\_list

Out[85]: ['gender', 'ever\_married', 'work\_type', 'Residence\_type', 'smoking \_status']



In [86]: **from** sklearn **import** preprocessing **for** i **in** obj\_list:

Encoder **=** preprocessing.LabelEncoder()

df[i]**=** Encoder.fit\_transform(df[i])



In [87]: df.head()

Out[87]: **gender age hypertension heart\_disease ever\_married work\_type Residence\_type av**



|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 1 | 67.0 |  | 0 | 1 | 1 | 1 | 1 |  |
| **1** | 1 | 80.0 |  | 0 | 1 | 1 | 1 | 0 |  |
| **2** | 0 | 49.0 |  | 0 | 0 | 1 | 1 | 1 |  |
| **3** | 0 | 79.0 |  | 1 | 0 | 1 | 2 | 0 |  |
| **4** | 1 | 81.0 |  | 0 | 0 | 1 | 1 | 1 |  |
| In [88]: x **=** df.drop(columns**=**['stroke'],axis**=**1) | | | | | | |  |  |  |
| y **=** df['stroke'] | | | |  |  |  |  |  |  |
| In [89]: x,y |  |  |  |  |  |  |  |  |  |
| Out[89]: ( | gender | | age | hypertension | | heart\_disease | ever\_married | wo |  |
| rk\_type | | \ |  |  |  |  |  |  |  |
| 0 |  | 1 | 67.0 |  | 0 | 1 | 1 |  |  |
| 1 |  |  |  |  |  |  |  |  |  |
| 1 |  | 1 | 80.0 |  | 0 | 1 | 1 |  |  |
| 1 |  |  |  |  |  |  |  |  |  |
| 2 |  | 0 | 49.0 |  | 0 | 0 | 1 |  |  |
| 1 |  |  |  |  |  |  |  |  |  |
| 3 |  | 0 | 79.0 |  | 1 | 0 | 1 |  |  |
| 2 |  |  |  |  |  |  |  |  |  |
| 4 |  | 1 | 81.0 |  | 0 | 0 | 1 |  |  |
| 1 |  |  |  |  |  |  |  |  |  |
| ... |  | ... ... | |  | ... | ... | ... |  |  |
| ... |  |  |  |  |  |  |  |  |  |
| 4976 |  | 1 | 41.0 |  | 0 | 0 | 0 |  |  |
| 1 |  | 1 | 40.0 |  | 0 | 0 | 1 |  |  |
| 4977 |  |  |  |  |



http://localhost:8888/notebooks/Documents/MTECH(DSA)/MyLabWork/AML\_Lab/LabAssignment\_6.ipynb Page 1 of 5

LabAssignment\_6 - Jupyter Notebook 07/12/22, 11:11 PM

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 1 |  |  |  |  |  |  |
| 4978 | 0 | 45.0 | 1 |  | 0 | 1 |
| 0 |  |  |  |  |  |  |
| 4979 | 1 | 40.0 | 0 |  | 0 | 1 |
| 1 |  |  |  |  |  |  |
| 4980 | 0 | 80.0 | 1 |  | 0 | 1 |
| 1 |  |  |  |  |  |  |
|  | Residence\_type | | avg\_glucose\_level bmi | |  | smoking\_status |
| 0 |  | 1 | 228.69 | 36.6 |  | 1 |
| 1 |  | 0 | 105.92 | 32.5 |  | 2 |
| 2 |  | 1 | 171.23 | 34.4 |  | 3 |
| 3 |  | 0 | 174.12 | 24.0 |  | 2 |
| 4 |  | 1 | 186.21 | 29.0 |  | 1 |
| ... |  | ... | ... ... | |  | ... |
| 4976 |  | 0 | 70.15 | 29.8 |  | 1 |
| 4977 |  | 1 | 191.15 | 31.1 |  | 3 |
| 4978 |  | 0 | 95.02 | 31.8 |  | 3 |
| 4979 |  | 0 | 83.94 | 30.0 |  | 3 |
| 4980 |  | 1 | 83.75 | 29.1 |  | 2 |

[4981 rows x 10 columns],

1. 1
2. 1
3. 1
4. 1
5. 1

..

|  |  |
| --- | --- |
| 4976 | 0 |
| 4977 | 0 |
| 4978 | 0 |
| 4979 | 0 |
| 4980 | 0 |

Name: stroke, Length: 4981, dtype: int64)



In [90]: **from** imblearn.over\_sampling **import** RandomOverSampler

over\_sampler **=** RandomOverSampler(sampling\_strategy**=**'minority')

x,y **=** over\_sampler.fit\_resample(x,y)



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| In [91]: x,y |  |  |  |  |  |
| Out[91]: ( | gender | age | hypertension | heart\_disease | ever\_married wo |
| rk\_type | \ |  |  |  |  |
| 0 | 1 | 67.0 | 0 | 1 | 1 |
| 1 |  |  |  |  |  |
| 1 | 1 | 80.0 | 0 | 1 | 1 |
| 1 |  |  |  |  |  |
| 2 | 0 | 49.0 | 0 | 0 | 1 |
| 1 |  |  |  |  |  |
| 3 | 0 | 79.0 | 1 | 0 | 1 |
| 2 |  |  |  |  |  |
| 4 | 1 | 81.0 | 0 | 0 | 1 |
| 1 |  |  |  |  |  |

http://localhost:8888/notebooks/Documents/MTECH(DSA)/MyLabWork/AML\_Lab/LabAssignment\_6.ipynb Page 2 of 5

LabAssignment\_6 - Jupyter Notebook 07/12/22, 11:11 PM

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1 |  |  |  |  |  |
| ... | ... ... | | ... | ... | ... |
| ... |  |  |  |  |  |
| 9461 | 0 | 70.0 | 0 | 1 | 1 |
| 1 |  |  |  |  |  |
| 9462 | 1 | 54.0 | 0 | 0 | 1 |
| 1 |  |  |  |  |  |
| 9463 | 0 | 75.0 | 0 | 0 | 1 |
| 1 |  |  |  |  |  |
| 9464 | 0 | 61.0 | 0 | 0 | 1 |
| 2 |  |  |  |  |  |
| 9465 | 0 | 66.0 | 1 | 0 | 1 |
| 0 |  |  |  |  |  |
|  | Residence\_type | | avg\_glucose\_level bmi | | smoking\_status |
| 0 |  | 1 | 228.69 | 36.6 | 1 |
| 1 |  | 0 | 105.92 | 32.5 | 2 |
| 2 |  | 1 | 171.23 | 34.4 | 3 |
| 3 |  | 0 | 174.12 | 24.0 | 2 |
| 4 |  | 1 | 186.21 | 29.0 | 1 |
| ... |  | ... | ... ... | | ... |
| 9461 |  | 0 | 239.07 | 26.1 | 2 |
| 9462 |  | 1 | 71.22 | 28.5 | 2 |
| 9463 |  | 0 | 199.20 | 26.6 | 0 |
| 9464 |  | 0 | 202.21 | 31.6 | 2 |
| 9465 |  | 0 | 116.55 | 31.1 | 1 |

[9466 rows x 10 columns],

1. 1
2. 1
3. 1
4. 1
5. 1

..

9461 1

9462 1

9463 1

9464 1

9465 1

Name: stroke, Length: 9466, dtype: int64)



In [92]: **from** sklearn.model\_selection **import** train\_test\_split

x\_train,x\_test,y\_train,y\_test **=** train\_test\_split(x,y,test\_size**=**0.25

Random Forest Classifier



In [93]: **from** sklearn.ensemble **import** RandomForestClassifier RF\_model **=** RandomForestClassifier(n\_estimators**=**1000) RF\_model.fit(x\_train,y\_train)

y\_pred **=** RF\_model.predict(x\_test)

http://localhost:8888/notebooks/Documents/MTECH(DSA)/MyLabWork/AML\_Lab/LabAssignment\_6.ipynb Page 3 of 5

LabAssignment\_6 - Jupyter Notebook 07/12/22, 11:11 PM



In [94]: **from** sklearn.metrics **import** precision\_score,recall\_score,roc\_auc\_sc **print**("Precision Score for RandomForest model",precision\_score(y\_te

Precision Score for RandomForest model 0.9809286898839138

DecisionTreeClassifier



In [95]: **from** sklearn.tree **import** DecisionTreeClassifier DT\_model **=** DecisionTreeClassifier() DT\_model.fit(x\_train,y\_train)

tree\_pred **=** DT\_model.predict(x\_test)



In [96]: **print**("Precision Score for Decision model",precision\_score(y\_test,t

Precision Score for Decision model 0.9441340782122905



In [97]: **print**("Recall Score for Decision model",recall\_score(y\_test,tree\_pr **print**("Recall Score for RandomForest model",recall\_score(y\_test,y\_p

Recall Score for Decision model 1.0

Recall Score for RandomForest model 1.0



In [98]: **print**("ROC-AUC Score for Decision model",roc\_auc\_score(y\_test,tree\_ **print**("ROC-AUC Score for RandomForest model",roc\_auc\_score(y\_test,y

ROC-AUC Score for Decision model 0.9704391891891893 ROC-AUC Score for RandomForest model 0.9902871621621622

Confusion Matrix



In [100]: **from** sklearn.metrics **import** confusion\_matrix DT\_model.fit(x\_train,y\_train)

DT\_model **=** DT\_model.predict(x\_test)

cm\_log **=** confusion\_matrix(y\_test,tree\_pred)

cm\_log

|  |  |  |
| --- | --- | --- |
| Out[100]: array([[1114, | | 70], |
| [ | 0, | 1183]]) |

http://localhost:8888/notebooks/Documents/MTECH(DSA)/MyLabWork/AML\_Lab/LabAssignment\_6.ipynb Page 4 of 5

LabAssignment\_6 - Jupyter Notebook 07/12/22, 11:11 PM



In [101]: **from** prettytable **import** PrettyTable

Comparision\_table **=** PrettyTable(["Model", "Precision Score", "Re Comparision\_table.add\_row(["Decision Model","0.94", "1.0", "0.96"]) Comparision\_table.add\_row(["RandomForest Model","0.98", "1.0", "0.9 **print**(Comparision\_table)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| + | -------------------- | + | ----------------- | + | -------------- | + | ----------- |
| ----+ | |  |  |  |  |  |  |
| | | Model | | Precision Score | Recall Score | ROC-AUC Sc | | | | | |
| ore | | |  |  |  |  |  |  |
| + | -------------------- | + | ----------------- | + | -------------- | + | ----------- |
| ----+ | |  |  |  |  |  |  |
| | | Decision Model | | | 0.94 | | | 1.0 | | | 0.96 |
| | |  |  |  |  |  |  |  |
| | RandomForest Model | | | | 0.98 | | | 1.0 | | | 0.99 |
| | |  |  |  |  |  |  |  |
| + | -------------------- | + | ----------------- | + | -------------- | + | ----------- |
| ----+ | |  |  |  |  |  |  |

http://localhost:8888/notebooks/Documents/MTECH(DSA)/MyLabWork/AML\_Lab/LabAssignment\_6.ipynb Page 5 of 5