[2]: import numpy as np

import pandas as pd

from sklearn import preprocessing

dataset=pd.read\_excel("machine assign.xlsx",sheetname=0)

Traceback (most recent call last):

File "<ipython-input-4-2285733c5ac3>", line 1, in <module>

dataset=pd.read\_excel("machine assign.xlsx",sheetname=0)

File "C:\ProgramData\Anaconda3\lib\site-packages\pandas\io\excel\\_base.py", line 301, in read\_excel

raise TypeError(f"read\_excel() got an unexpected keyword argument `{arg}`")

TypeError: read\_excel() got an unexpected keyword argument `sheetname`

dataset=pd.read\_excel("machine assign.xlsx",sheet\_name=0)

from sklearn.ensemble import RandomForestClassifier

dataset.columns

Out[7]:

Index(['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',

'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender',

'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome',

'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours',

'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',

'YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager'],

dtype='object')

label\_encoder=preprocessing.LabelEncoder()

dataset["Attrition"]= label\_encoder.fit\_transform(dataset["Attrition"])

rf\_model=RandomForestClassifier(n\_estimators=1000,max\_features=3,oob\_score=True)

features=['Age','DistanceFromHome','MonthlyIncome']

rf\_model.fit(X=dataset[features],y=dataset["Attrition"])

Out[12]:

RandomForestClassifier(bootstrap=True, class\_weight=None, criterion='gini',

max\_depth=None, max\_features=3, max\_leaf\_nodes=None,

min\_impurity\_decrease=0.0, min\_impurity\_split=None,

min\_samples\_leaf=1, min\_samples\_split=2,

min\_weight\_fraction\_leaf=0.0, n\_estimators=1000,

n\_jobs=None, oob\_score=True, random\_state=None,

verbose=0, warm\_start=False)

print("OOB Accuracy:")

OOB Accuracy:

print(rf\_model.oob\_score\_);

1.0

for features,imp in zip(features,rf\_model.feature\_importances\_):

print(features,imp);

Age 0.24737103811711225

DistanceFromHome 0.20544536409120318

MonthlyIncome 0.5471835977916846

dataset["Age"]=label\_encoder.fit\_transform(dataset["Age"])

dataset["DistanceFromHome"]=label\_encoder.fit\_transform(dataset["DistanceFromHome"])

dataset["MonthlyIncome"]=label\_encoder.fit\_transform(dataset["MonthlyIncome"])

rf\_model.fit(X=dataset[features],y=dataset["Personal Loan"])

Traceback (most recent call last):

File "C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\indexes\base.py", line 2646, in get\_loc

return self.\_engine.get\_loc(key)

File "pandas\\_libs\index.pyx", line 111, in pandas.\_libs.index.IndexEngine.get\_loc

File "pandas\\_libs\index.pyx", line 138, in pandas.\_libs.index.IndexEngine.get\_loc

File "pandas\\_libs\hashtable\_class\_helper.pxi", line 1618, in pandas.\_libs.hashtable.PyObjectHashTable.get\_item

File "pandas\\_libs\hashtable\_class\_helper.pxi", line 1626, in pandas.\_libs.hashtable.PyObjectHashTable.get\_item

KeyError: 'Personal Loan'

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

File "<ipython-input-19-0ac3e9719c3f>", line 1, in <module>

rf\_model.fit(X=dataset[features],y=dataset["Personal Loan"])

File "C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py", line 2800, in \_\_getitem\_\_

indexer = self.columns.get\_loc(key)

File "C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\indexes\base.py", line 2648, in get\_loc

return self.\_engine.get\_loc(self.\_maybe\_cast\_indexer(key))

File "pandas\\_libs\index.pyx", line 111, in pandas.\_libs.index.IndexEngine.get\_loc

File "pandas\\_libs\index.pyx", line 138, in pandas.\_libs.index.IndexEngine.get\_loc

File "pandas\\_libs\hashtable\_class\_helper.pxi", line 1618, in pandas.\_libs.hashtable.PyObjectHashTable.get\_item

File "pandas\\_libs\hashtable\_class\_helper.pxi", line 1626, in pandas.\_libs.hashtable.PyObjectHashTable.get\_item

KeyError: 'Personal Loan'

rf\_model.fit(X=dataset[features],y=dataset["Attrition"])

Traceback (most recent call last):

File "<ipython-input-20-2f4d26b5ccbe>", line 1, in <module>

rf\_model.fit(X=dataset[features],y=dataset["Attrition"])

File "C:\ProgramData\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py", line 249, in fit

X = check\_array(X, accept\_sparse="csc", dtype=DTYPE)

File "C:\ProgramData\Anaconda3\lib\site-packages\sklearn\utils\validation.py", line 521, in check\_array

"if it contains a single sample.".format(array))

ValueError: Expected 2D array, got 1D array instead:

array=[1166. 505. 1318. ... 434. 164. 744.].

Reshape your data either using array.reshape(-1, 1) if your data has a single feature or array.reshape(1, -1) if it contains a single sample.

titanic\_train = pd.read\_excel("machine assign.xlsx",sheet\_name=0)

new\_age\_var = np.where(titanic\_train["Age"].isnull(),28,titanic\_train["Age"])

new\_monthly\_var = np.where(titanic\_train["MonthlyIncome"].isnull(),45000,titanic\_train["MonthlyIncome"])

new\_distance\_var = np.where(titanic\_train["DistanceFromHome"].isnull(),10,titanic\_train["DistanceFromHome"])

titanic\_train["Age"]=new\_age\_var

titanic\_train["DistanceFromHome"]=new\_distance\_var

label\_encoder=preprocessing.LabelEncoder()

encoded\_age=label\_encoder.fit\_transform(titanic\_train['Age'])

encoded\_distance=label\_encoder.fit\_transform(titanic\_train['DistanceFromHome'])

from sklearn import tree

tree\_model =tree.DecisionTreeClassifier()

tree\_model.fit(X=pd.DataFrame(encoded\_age,encoded\_distance),y=titanic\_train["Attrition"])

Out[32]:

DecisionTreeClassifier(class\_weight=None, criterion='gini', max\_depth=None,

max\_features=None, max\_leaf\_nodes=None,

min\_impurity\_decrease=0.0, min\_impurity\_split=None,

min\_samples\_leaf=1, min\_samples\_split=2,

min\_weight\_fraction\_leaf=0.0, presort=False,

random\_state=None, splitter='best')

f=tree.export\_graphviz

with open("Dtree1.dot",'w') as f:

f=tree.export\_graphviz(tree\_model,feature\_names=["Age","DistanceFromHome"],out\_file=f);

Traceback (most recent call last):

File "<ipython-input-34-54968c4c19d5>", line 2, in <module>

f=tree.export\_graphviz(tree\_model,feature\_names=["Age","DistanceFromHome"],out\_file=f);

File "C:\ProgramData\Anaconda3\lib\site-packages\sklearn\tree\export.py", line 775, in export\_graphviz

exporter.export(decision\_tree)

File "C:\ProgramData\Anaconda3\lib\site-packages\sklearn\tree\export.py", line 401, in export

decision\_tree.n\_features\_))

ValueError: Length of feature\_names, 2 does not match number of features, 1

with open("Dtree1.dot",'w') as f:

f=tree.export\_graphviz(tree\_model,feature\_names=["Age"],out\_file=f);

f=tree.export\_graphviz(tree\_model,feature\_names=["DistanceFromHome"],out\_file=f);

predictors=pd.DataFrame([encoded\_age,encoded\_distance,titanic\_train['Age','DistanceFromHome']]).T

Traceback (most recent call last):

File "C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\indexes\base.py", line 2646, in get\_loc

return self.\_engine.get\_loc(key)

File "pandas\\_libs\index.pyx", line 111, in pandas.\_libs.index.IndexEngine.get\_loc

File "pandas\\_libs\index.pyx", line 138, in pandas.\_libs.index.IndexEngine.get\_loc

File "pandas\\_libs\hashtable\_class\_helper.pxi", line 1618, in pandas.\_libs.hashtable.PyObjectHashTable.get\_item

File "pandas\\_libs\hashtable\_class\_helper.pxi", line 1626, in pandas.\_libs.hashtable.PyObjectHashTable.get\_item

KeyError: ('Age', 'DistanceFromHome')

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

File "<ipython-input-36-ac9f5f40a04b>", line 1, in <module>

predictors=pd.DataFrame([encoded\_age,encoded\_distance,titanic\_train['Age','DistanceFromHome']]).T

File "C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py", line 2800, in \_\_getitem\_\_

indexer = self.columns.get\_loc(key)

File "C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\indexes\base.py", line 2648, in get\_loc

return self.\_engine.get\_loc(self.\_maybe\_cast\_indexer(key))

File "pandas\\_libs\index.pyx", line 111, in pandas.\_libs.index.IndexEngine.get\_loc

File "pandas\\_libs\index.pyx", line 138, in pandas.\_libs.index.IndexEngine.get\_loc

File "pandas\\_libs\hashtable\_class\_helper.pxi", line 1618, in pandas.\_libs.hashtable.PyObjectHashTable.get\_item

File "pandas\\_libs\hashtable\_class\_helper.pxi", line 1626, in pandas.\_libs.hashtable.PyObjectHashTable.get\_item

KeyError: ('Age', 'DistanceFromHome')

predictors=pd.DataFrame([encoded\_age,titanic\_train['Age']]).T

predictors=pd.DataFrame([encoded\_distance,titanic\_train['DistanceFromHome']]).T

tree\_model.fit(X=predictors,y=titanic\_train["Attrition"])

Out[39]:

DecisionTreeClassifier(class\_weight=None, criterion='gini', max\_depth=None,

max\_features=None, max\_leaf\_nodes=None,

min\_impurity\_decrease=0.0, min\_impurity\_split=None,

min\_samples\_leaf=1, min\_samples\_split=2,

min\_weight\_fraction\_leaf=0.0, presort=False,

random\_state=None, splitter='best')

tree\_model.score(X=predictors,y=titanic\_train["Attrition"])

Out[40]: 0.8387755102040816