1. Dv = pclass idv = remaining all

import numpy as np

import pandas as pd

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

dataset.columns

Out[3]:

Index(['PassengerId', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp', 'Parch',

'Ticket', 'Fare', 'Embarked'],

dtype='object')

from sklearn import preprocessing

from sklearn.cross\_validation import train\_test\_split

Traceback (most recent call last):

File "<ipython-input-5-d05cc6ee7707>", line 1, in <module>

from sklearn.cross\_validation import train\_test\_split

ModuleNotFoundError: No module named 'sklearn.cross\_validation'

from sklearn.model\_selection import train\_test\_split

from sklearn.naive\_bayes import GaussianNB

from sklearn.metrics import accuracy\_score

from sklearn.metrics import confusion\_matrix

le = preprocessing.LabelEncoder()

le.fit(dataset["sex"])

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: 'sex'

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

File "<ipython-input-11-372d7a335011>", line 1, in <module>

le.fit(dataset["sex"])

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: 'sex'

le.fit(dataset["Sex"])

Out[12]: LabelEncoder()

print(le.classes\_)

['female' 'male']

dataset["Sex"]=le.transform(dataset["Sex"])

y= dataset["Pclass"]

x =dataset.drop(["Pclass","PassengerId"],axis=1)

y.count()

Out[17]: 417

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.3,random\_state=0)

from sklearn.naive\_bayes import \*

clf = BernoulliNB()

y\_pred = clf.fit(x\_train,y\_train).predict(x\_test)

Traceback (most recent call last):

File "<ipython-input-21-421cf61e2543>", line 1, in <module>

y\_pred = clf.fit(x\_train,y\_train).predict(x\_test)

File "C:\ProgramData\Anaconda3\lib\site-packages\sklearn\naive\_bayes.py", line 588, in fit

X, y = check\_X\_y(X, y, 'csr')

File "C:\ProgramData\Anaconda3\lib\site-packages\sklearn\utils\validation.py", line 719, in check\_X\_y

estimator=estimator)

File "C:\ProgramData\Anaconda3\lib\site-packages\sklearn\utils\validation.py", line 536, in check\_array

array = array.astype(np.float64)

ValueError: could not convert string to float: 'Cavendish, Mrs. Tyrell William (Julia Florence Siegel)'

x=dataset.drop(["Name","Ticket","Embarked"],axis=1)

y\_pred = clf.fit(x\_train,y\_train).predict(x\_test)

Traceback (most recent call last):

File "<ipython-input-23-421cf61e2543>", line 1, in <module>

y\_pred = clf.fit(x\_train,y\_train).predict(x\_test)

File "C:\ProgramData\Anaconda3\lib\site-packages\sklearn\naive\_bayes.py", line 588, in fit

X, y = check\_X\_y(X, y, 'csr')

File "C:\ProgramData\Anaconda3\lib\site-packages\sklearn\utils\validation.py", line 719, in check\_X\_y

estimator=estimator)

File "C:\ProgramData\Anaconda3\lib\site-packages\sklearn\utils\validation.py", line 536, in check\_array

array = array.astype(np.float64)

ValueError: could not convert string to float: 'Cavendish, Mrs. Tyrell William (Julia Florence Siegel)'

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.3,random\_state=0)

from sklearn.naive\_bayes import \*

clf = BernoulliNB()

x=dataset.drop(["Name","Ticket","Embarked"],axis=1)

y\_pred = clf.fit(x\_train,y\_train).predict(x\_test)

accuracy\_score(y\_test,y\_pred,normalize=True)

Out[29]: 0.47619047619047616

confusion\_matrix(y\_test,y\_pred)

Out[30]:

array([[ 1, 0, 36],

[ 3, 0, 24],

[ 3, 0, 59]], dtype=int64)

2.Dv = sex, idv = remaining all

import numpy as np

import pandas as pd

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

from sklearn import preprocessing

from sklearn.cross\_validation import train\_test\_split

Traceback (most recent call last):

File "<ipython-input-4-d05cc6ee7707>", line 1, in <module>

from sklearn.cross\_validation import train\_test\_split

ModuleNotFoundError: No module named 'sklearn.cross\_validation'

from sklearn.model\_selection import train\_test\_split

from sklearn.naive\_bayes import GaussianNB

from sklearn.metrics import accuracy\_score

from sklearn.metrics import confusion\_matrix

le = preprocessing.LabelEncoder()

le.fit(dataset["Sex"])

Out[10]: LabelEncoder()

print(le.classes\_)

['female' 'male']

dataset["Sex"]=le.transform(dataset["Sex"])

y= dataset["Sex"]

x =dataset.drop(["Pclass","PassengerId","Name","Ticket"],axis=1)

y.count()

Out[15]: 417

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.3,random\_state=0)

from sklearn.naive\_bayes import \*

clf = BernoulliNB()

y\_pred = clf.fit(x\_train,y\_train).predict(x\_test)

Traceback (most recent call last):

File "<ipython-input-19-421cf61e2543>", line 1, in <module>

y\_pred = clf.fit(x\_train,y\_train).predict(x\_test)

File "C:\ProgramData\Anaconda3\lib\site-packages\sklearn\naive\_bayes.py", line 588, in fit

X, y = check\_X\_y(X, y, 'csr')

File "C:\ProgramData\Anaconda3\lib\site-packages\sklearn\utils\validation.py", line 719, in check\_X\_y

estimator=estimator)

File "C:\ProgramData\Anaconda3\lib\site-packages\sklearn\utils\validation.py", line 536, in check\_array

array = array.astype(np.float64)

ValueError: could not convert string to float: 'S'

x =dataset.drop(["Pclass","PassengerId","Name","Ticket","Embarked"],axis=1)

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.3,random\_state=0)

y\_pred = clf.fit(x\_train,y\_train).predict(x\_test)

accuracy\_score(y\_test,y\_pred,normalize=True)

Out[23]: 1.0

confusion\_matrix(y\_test,y\_pred)

Out[24]:

array([[55, 0],

[ 0, 71]], dtype=int64)

####################################################################

3 dv =sibsp idv = remaining all

y= dataset["SibSp"]

x =dataset.drop(["Pclass","PassengerId","Name","Ticket","Embarked"],axis=1)

y.count()

Out[28]: 417

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.3,random\_state=0)

y\_pred = clf.fit(x\_train,y\_train).predict(x\_test)

accuracy\_score(y\_test,y\_pred,normalize=True)

Out[31]: 0.9285714285714286

confusion\_matrix(y\_test,y\_pred)

Out[32]:

array([[83, 0, 0, 0, 0],

[ 0, 34, 0, 0, 0],

[ 0, 5, 0, 0, 0],

[ 0, 2, 0, 0, 0],

[ 0, 2, 0, 0, 0]], dtype=int64)

############################################

4 dv = parch , idv= remaining all

y= dataset["Parch"]

x =dataset.drop(["Pclass","PassengerId","Name","Ticket","Embarked"],axis=1)

y\_pred = clf.fit(x\_train,y\_train).predict(x\_test)

accuracy\_score(y\_test,y\_pred,normalize=True)

Out[37]: 0.9285714285714286

confusion\_matrix(y\_test,y\_pred)

Out[38]:

array([[83, 0, 0, 0, 0],

[ 0, 34, 0, 0, 0],

[ 0, 5, 0, 0, 0],

[ 0, 2, 0, 0, 0],

[ 0, 2, 0, 0, 0]], dtype=int64)

######################################

5 dv embarked, idv = remainng all

le.fit(dataset["Embarked"])

Out[40]: LabelEncoder()

print(le.classes\_)

['C' 'Q' 'S']

dataset["Embarked"]=le.transform(dataset["Embarked"])

y= dataset["Embarked"]

x =dataset.drop(["Pclass","PassengerId","Name","Ticket"],axis=1)

y.count()

Out[45]: 417

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.3,random\_state=0)

y\_pred = clf.fit(x\_train,y\_train).predict(x\_test)

accuracy\_score(y\_test,y\_pred,normalize=True)

Out[48]: 0.8888888888888888

confusion\_matrix(y\_test,y\_pred)

Out[49]:

array([[33, 0, 0],

[ 0, 0, 14],

[ 0, 0, 79]], dtype=int64)