**Naïve Bayes-Dependent Variable: Survived**

import pandas as pd

dataset=pd.read\_csv("E:/Rutuja/LockDown/AI and ML/Day 25 Notes/train.csv")

pd.get\_dummies(dataset, columns = ["Sex"],prefix="Sex")

pd.get\_dummies(dataset, columns = ["Embarked"],prefix="Embarked", drop\_first=True)

dataset.head()

from sklearn import preprocessing

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

from sklearn.naive\_bayes import \*

le=preprocessing.LabelEncoder()

le.fit(dataset['Sex'])

print(le.classes\_)

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

y=dataset['Survived']

X=dataset.drop(["Survived","PassengerId","Name","Ticket","Cabin","Embarked"],axis=1)

y.count()

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

X\_train.head()

clf=BernoulliNB()

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

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

**0.7715355805243446**

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

**array([[163, 0],**

**[ 0, 104]], dtype=int64)**

**Naïve Bayes-Dependent Variable: Pclass**

import pandas as pd

dataset=pd.read\_csv("E:/Rutuja/LockDown/AI and ML/Day 25 Notes/train.csv")

from sklearn import preprocessing

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

from sklearn.naive\_bayes import \*

le=preprocessing.LabelEncoder()

le.fit(dataset['Sex'])

print(le.classes\_)

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

y=dataset['Pclass']

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

y.count()

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

X\_train.head()

clf=BernoulliNB()

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

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

**0.5842696629213483**

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

**array([[ 70, 0, 0],**

**[ 0, 49, 0],**

**[ 0, 0, 148]], dtype=int64)**

**Naïve Bayes-Dependent Variable: Gender**

import pandas as pd

dataset=pd.read\_csv("E:/Rutuja/LockDown/AI and ML/Day 25 Notes/train.csv")

from sklearn import preprocessing

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

from sklearn.naive\_bayes import \*

le=preprocessing.LabelEncoder()

le.fit(dataset['Sex'])

print(le.classes\_)

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

y=dataset['Sex']

X=dataset.drop(["Sex","PassengerId","Name","Ticket","Cabin","Embarked"],axis=1)

y.count()

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

X\_train.head()

clf=BernoulliNB()

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

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

**0.7752808988764045**

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

**array([[ 97, 0],**

**[ 0, 170]], dtype=int64)**

**Naïve Bayes-Dependent Variable: SibSp**

import pandas as pd

dataset=pd.read\_csv("E:/Rutuja/LockDown/AI and ML/Day 25 Notes/train.csv")

from sklearn import preprocessing

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

from sklearn.naive\_bayes import \*

le=preprocessing.LabelEncoder()

le.fit(dataset['Sex'])

print(le.classes\_)

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

y=dataset['SibSp']

X=dataset.drop(["SibSp","PassengerId","Name","Ticket","Cabin","Embarked"],axis=1)

y.count()

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

X\_train.head()

clf=BernoulliNB()

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

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

**0.6367041198501873**

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

**array([[177, 0, 0, 0, 0, 0, 0],**

**[ 0, 66, 0, 0, 0, 0, 0],**

**[ 0, 0, 10, 0, 0, 0, 0],**

**[ 0, 0, 0, 7, 0, 0, 0],**

**[ 0, 0, 0, 0, 3, 0, 0],**

**[ 0, 0, 0, 0, 0, 1, 0],**

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

**Naïve Bayes-Dependent Variable: Parch**

import pandas as pd

dataset=pd.read\_csv("E:/Rutuja/LockDown/AI and ML/Day 25 Notes/train.csv")

from sklearn import preprocessing

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

from sklearn.naive\_bayes import \*

le=preprocessing.LabelEncoder()

le.fit(dataset['Sex'])

print(le.classes\_)

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

y=dataset['Parch']

X=dataset.drop(["Parch","PassengerId","Name","Ticket","Cabin","Embarked"],axis=1)

y.count()

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

X\_train.head()

clf=BernoulliNB()

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

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

**0.6853932584269663**

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

**array([[193, 0, 0, 0],**

**[ 0, 47, 0, 0],**

**[ 0, 0, 25, 0],**

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