NAÏVE BAYERS ALOGITHM

Accuracy is 80% in this algorithm

Full code:

# -\*- coding: utf-8 -\*-

"""

Created on Fri Nov 6 12:04:09 2020

@author: sunil

"""

import pandas as pd

import numpy as np

from sklearn.naive\_bayes import GaussianNB

from sklearn.naive\_bayes import MultinomialNB

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

from sklearn.metrics import confusion\_matrix

salary\_train = pd.read\_csv("SalaryData\_Train.csv")

salary\_test = pd.read\_csv("SalaryData\_Test.csv")

string\_columns=["workclass","education","maritalstatus","occupation","relationship","race","sex","native"]

from sklearn import preprocessing

number = preprocessing.LabelEncoder()

for i in string\_columns:

salary\_train[i] = number.fit\_transform(salary\_train[i])

salary\_test[i] = number.fit\_transform(salary\_test[i])

colnames = salary\_train.columns

len(colnames[0:13])

trainX = salary\_train[colnames[0:13]]

trainY = salary\_train[colnames[13]]

testX = salary\_test[colnames[0:13]]

testY = salary\_test[colnames[13]]

sgnb = GaussianNB()

smnb = MultinomialNB()

spred\_gnb = sgnb.fit(trainX,trainY).predict(testX)

confusion\_matrix(testY,spred\_gnb)

print ("Accuracy",(10759+1209)/(10759+601+2491+1209)) # 80%

spred\_mnb = smnb.fit(trainX,trainY).predict(testX)

confusion\_matrix(testY,spred\_mnb)

print("Accuracy",(10891+780)/(10891+780+2920+780)) # 75%



