**ML LAB ASSIGNMENT 2**

**Name:** Abhay Vijay Jagtap

**Class:** BE-B

**Problem Statement:**

Classify the email using the binary classification method. Email Spam detection has two states:

a) Normal State – Not Spam,

b) Abnormal State – Spam. Use K-Nearest Neighbors and Support Vector Machine for classification. Analyze their performance. Dataset link: The emails.csv dataset on the Kaggle <https://www.kaggle.com/datasets/balaka18/email-spam-classification-dataset-csv>

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

%matplotlib inline

import warnings

warnings.filterwarnings('ignore')

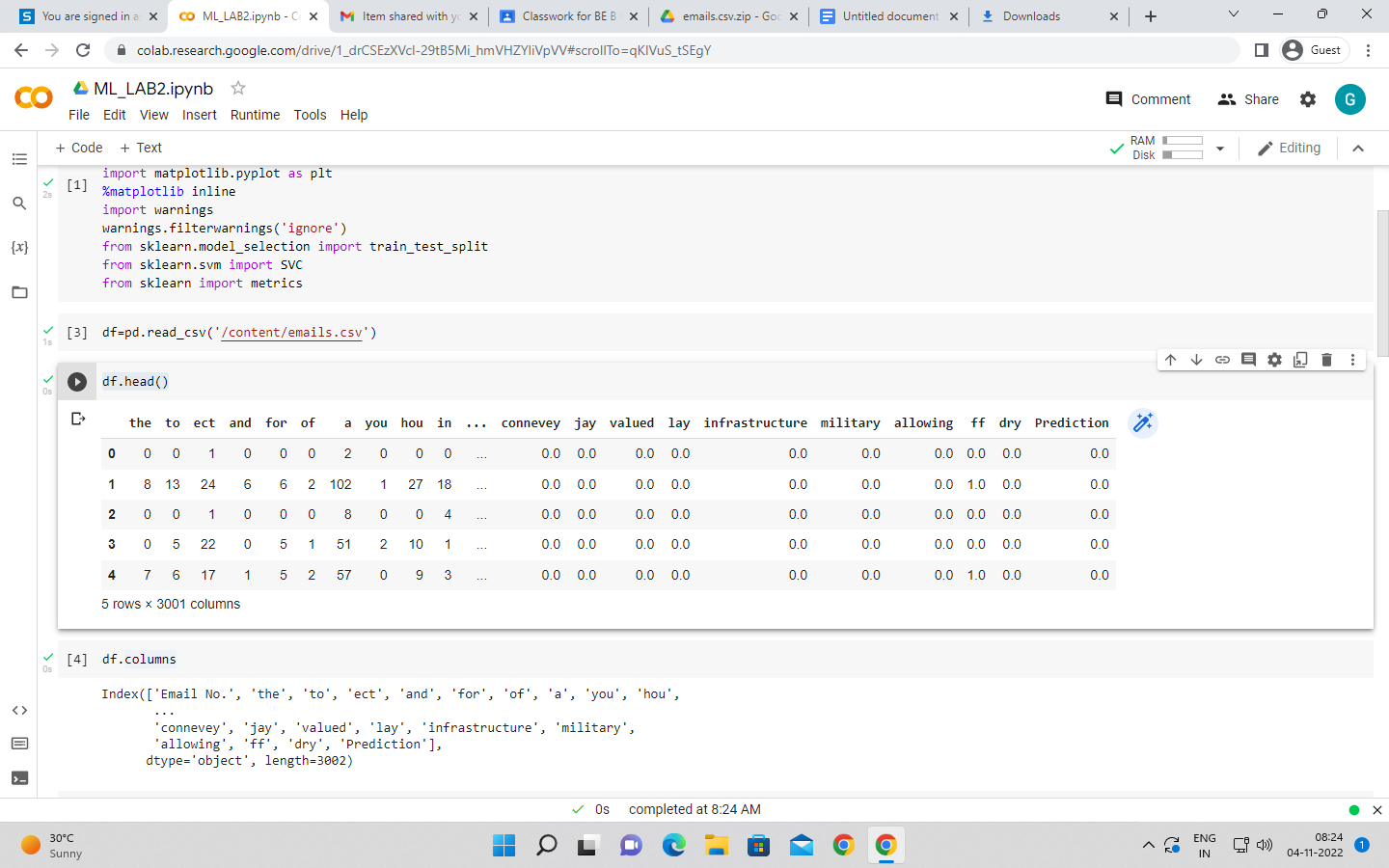
from sklearn.model\_selection import train\_test\_split

from sklearn.svm import SVC

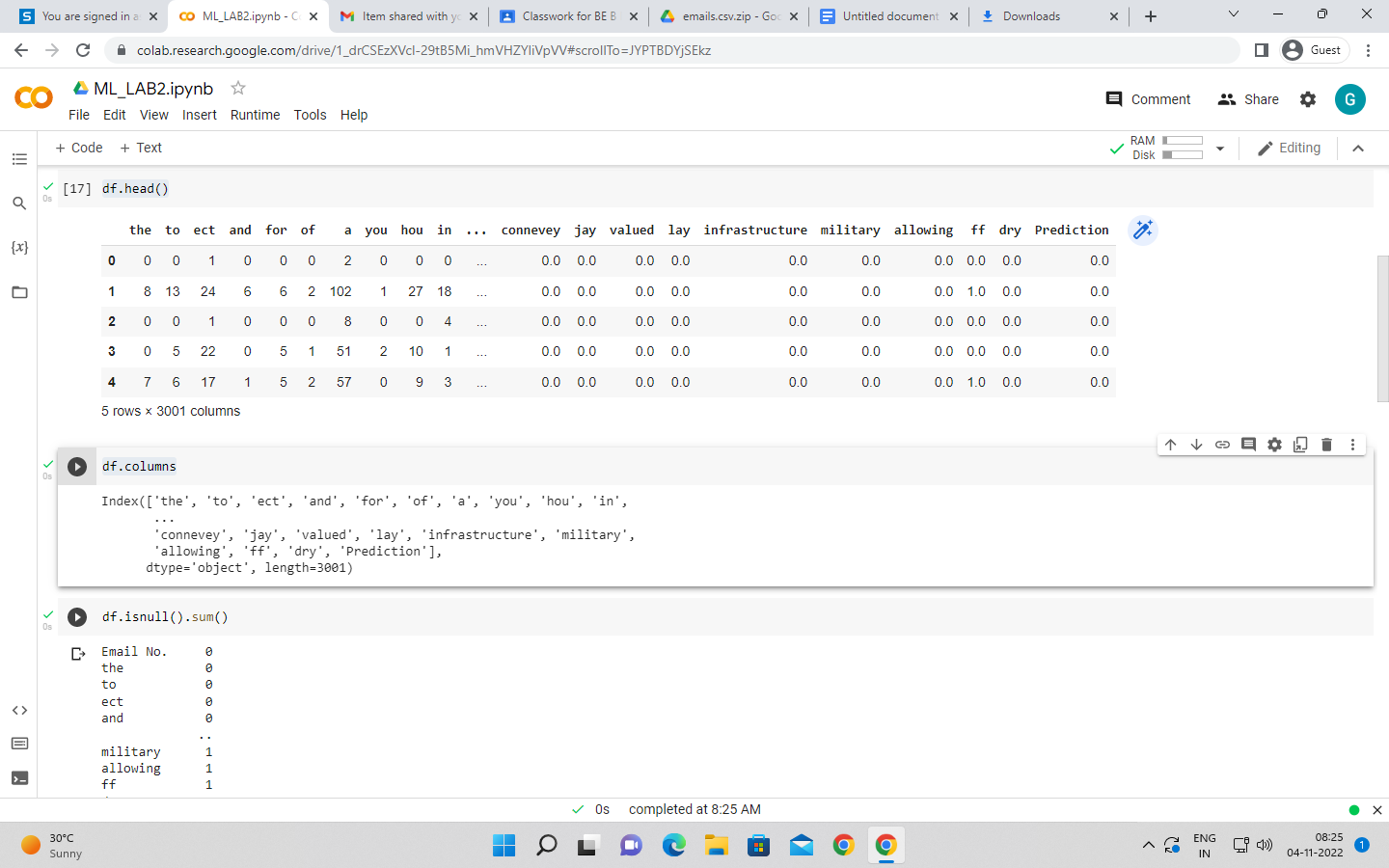
from sklearn import metrics

df=pd.read\_csv('/content/emails.csv')

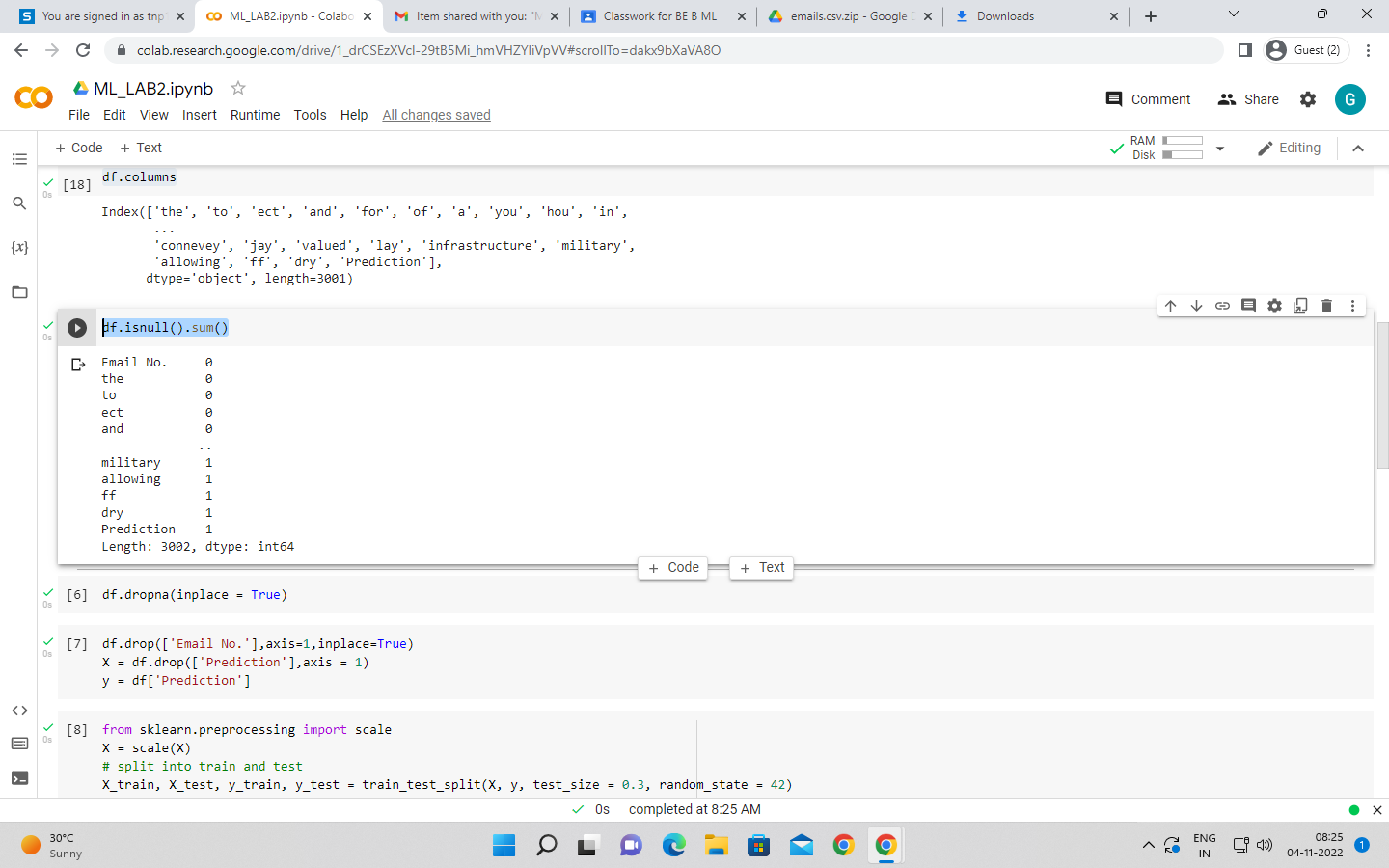
df.head()



df.columns



df.isnull().sum()



df.dropna(inplace = True)

df.drop(['Email No.'],axis=1,inplace=True)

X = df.drop(['Prediction'],axis = 1)

y = df['Prediction']

from sklearn.preprocessing import scale

X = scale(X)

# split into train and test

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size = 0.3, random\_state = 42)

## KNN classifier

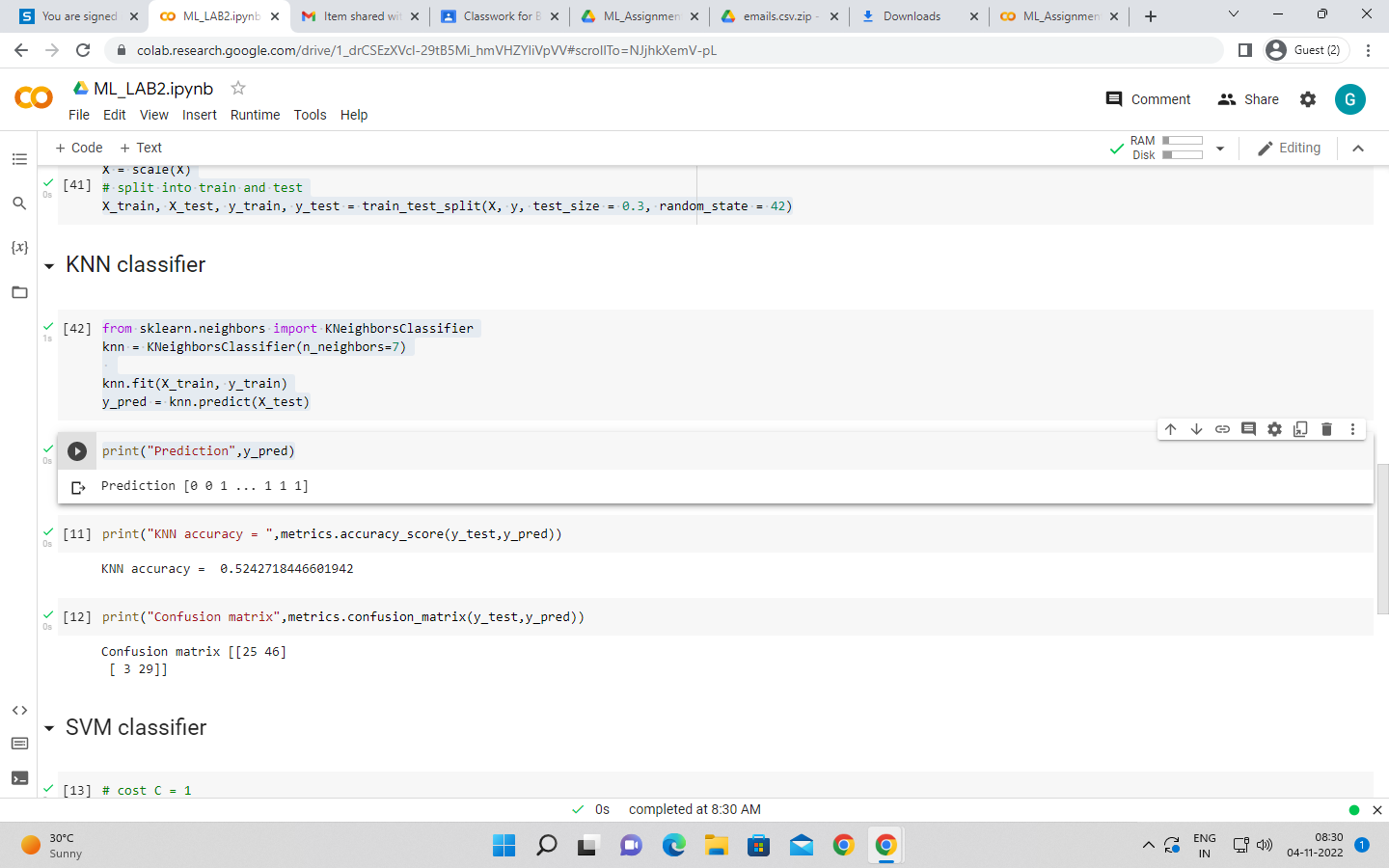
from sklearn.neighbors import KNeighborsClassifier

knn = KNeighborsClassifier(n\_neighbors=7)

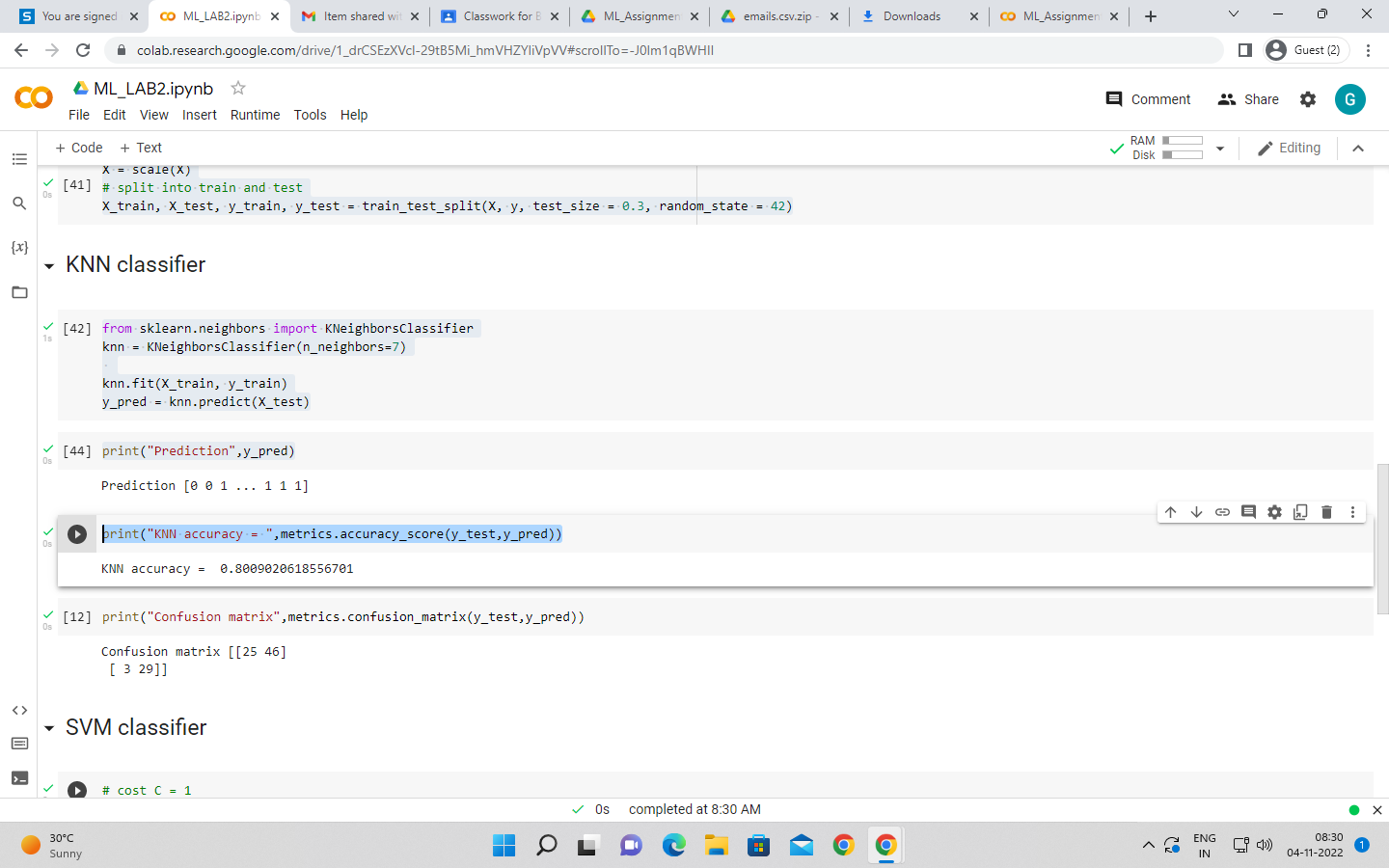
knn.fit(X\_train, y\_train)

y\_pred = knn.predict(X\_test)

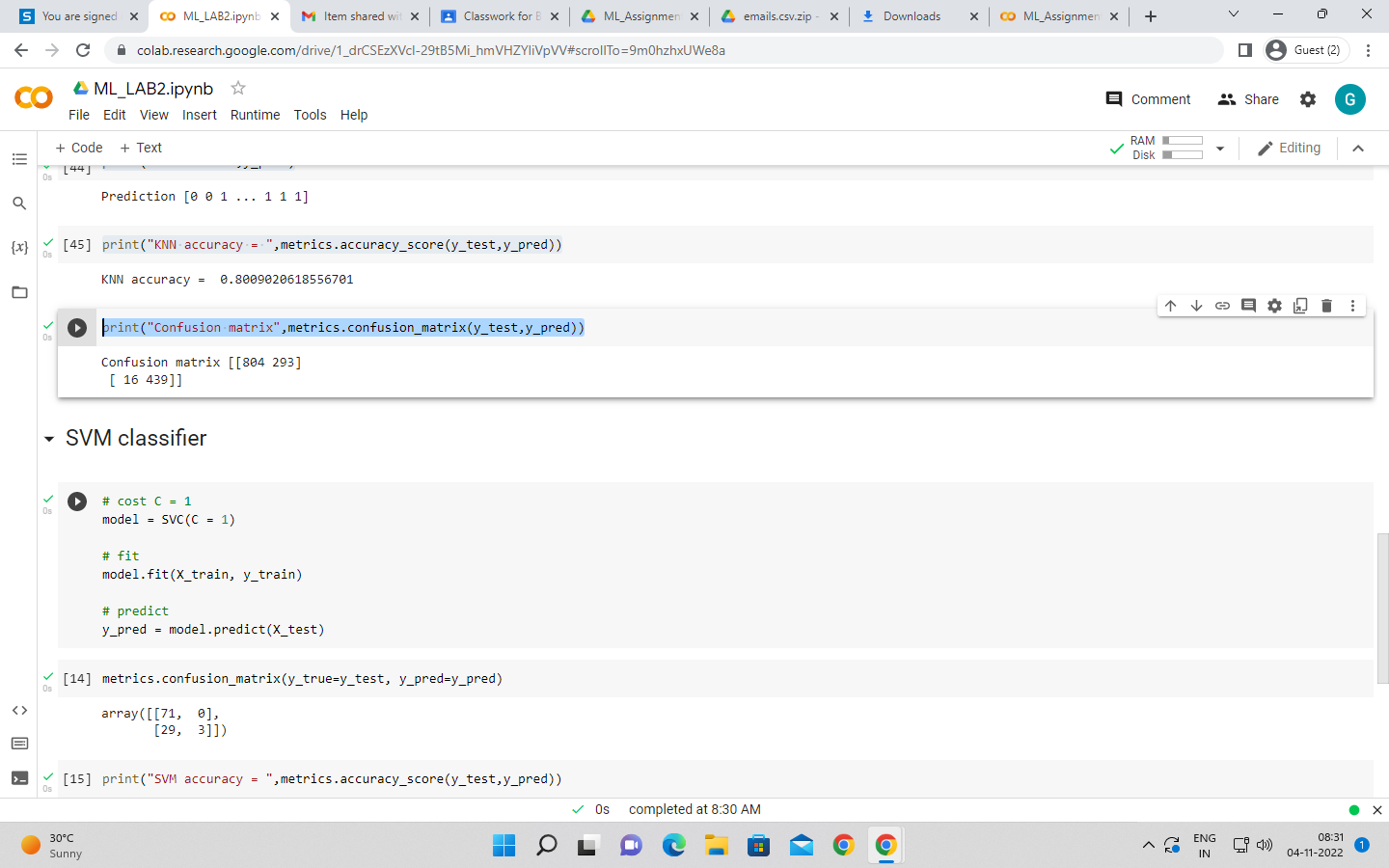
print("Prediction",y\_pred)



print("KNN accuracy = ",metrics.accuracy\_score(y\_test,y\_pred))



print("Confusion matrix",metrics.confusion\_matrix(y\_test,y\_pred))



## SVM classifier

# cost C = 1

model = SVC(C = 1)

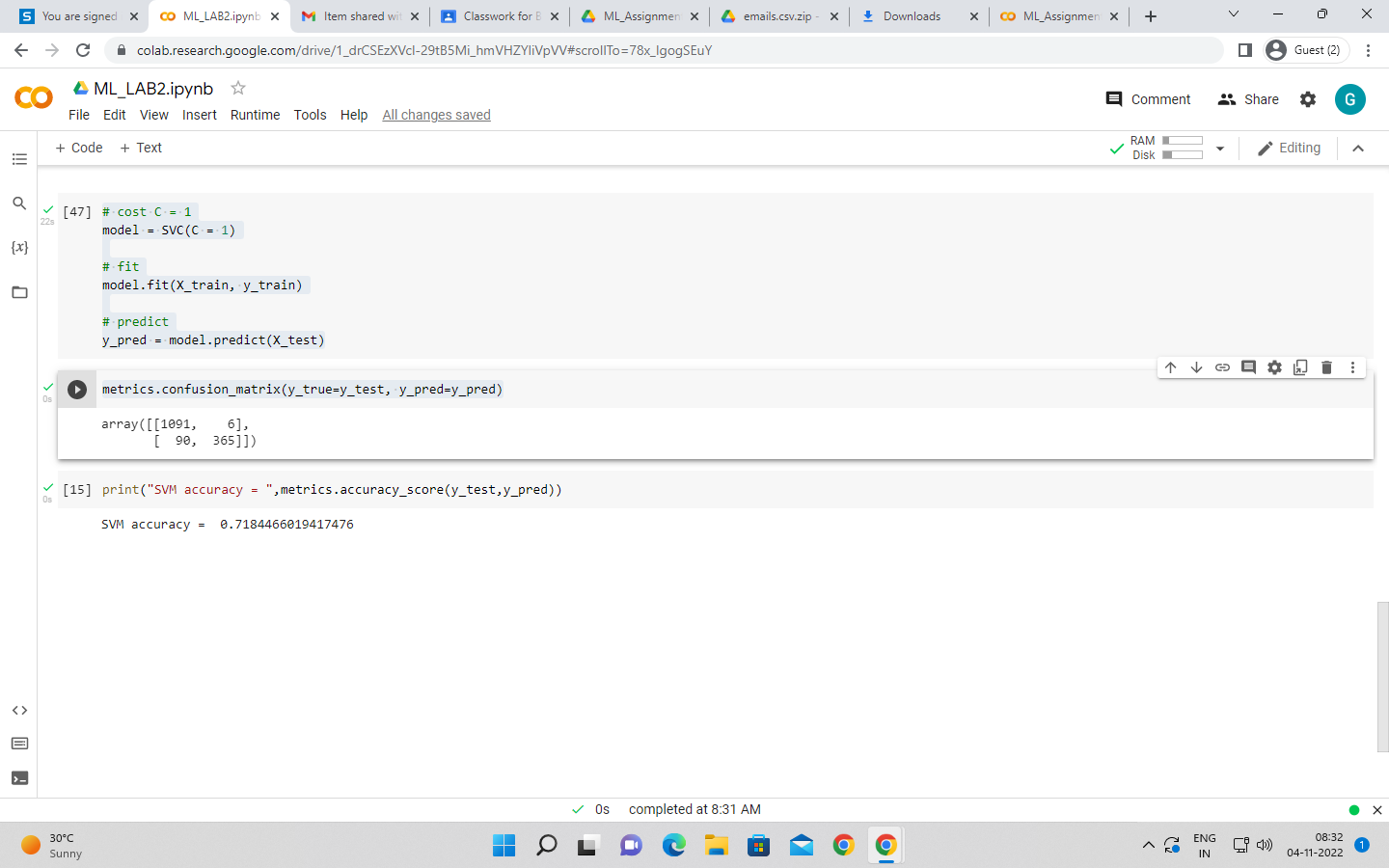
# fit

model.fit(X\_train, y\_train)

# predict

y\_pred = model.predict(X\_test)

metrics.confusion\_matrix(y\_true=y\_test, y\_pred=y\_pred)



print("SVM accuracy = ",metrics.accuracy\_score(y\_test,y\_pred))

