```python
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

from sklearn.model_selection import train_test_split

from sklearn.svm import SVC

from sklearn.metrics import accuracy_score

from sklearn.neighbors import KNeighborsClassifier


df = pd.read_csv("./emails.csv")


df.head()


df.isnull().sum()


X = df.iloc[:,1:3001]
X


Y = df.iloc[:,-1].values
Y


train_x,test_x,train_y,test_y = train_test_split(X,Y,test_size = 0.25)


svc = SVC(C=1.0,kernel='rbf',gamma='auto')
# C here is the regularization parameter. Here, L2 penalty is used(default). It is the inverse of the
strength of regularization.
# As C increases, model overfits.
# Kernel here is the radial basis function kernel.
# gamma (only used for rbf kernel) : As gamma increases, model overfits.
svc.fit(train_x,train_y)
y_pred2 = svc.predict(test_x)
print("Accuracy Score for SVC : ", accuracy_score(y_pred2,test_y))


X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size = 0.2, random_state=42)
```

```python
knn = KNeighborsClassifier(n_neighbors=7)

knn.fit(X_train, y_train)

print(knn.predict(X_test))

print(knn.score(X_test, y_test))
```