```
#EMAIL SPAM DETECTION


#IMPORTING LIBRARIES


import numpy as np
import pandas as pd
import matplotlib.pyplot as plt



# ALGORITHMS
from sklearn.tree import DecisionTreeClassifier
from sklearn.svm import SVC
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score
from sklearn.model_selection import train_test_split
from sklearn.pipeline import Pipeline
from sklearn.feature_extraction.text import CountVectorizer


#LOADING THE DATASET


d = pd.read_csv("/content/spam.csv", encoding="ISO-8859-1")


d.head()
```

	v1	v2	Unnamed: 2	Unnamed: 3	Unnamed: 4
0	ham	Go until jurong point, crazy.. Available only ...	NaN	NaN	NaN
1	ham	Ok lar... Joking wif u oni...	NaN	NaN	NaN
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...	NaN	NaN	NaN
3	ham	U dun say so early hor... U c already then say...	NaN	NaN	NaN
4	ham	Nah I don't think he goes to usf, he lives aro...	NaN	NaN	NaN

```
d.tail()
d.columns
```

```
Index(['v1', 'v2', 'Unnamed: 2', 'Unnamed: 3', 'Unnamed: 4'], dtype='object')
```

```
#RENAME COLUMNS AS CATEGORY AND MESSAGE
d.rename(columns = {'v1':'Category', 'v2':'Message'}, inplace = True)
d.head(10)
```

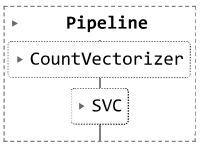
	Category	Message	Unnamed: 2	Unnamed: 3	Unnamed: 4
0	ham	Go until jurong point, crazy.. Available only ...	NaN	NaN	NaN
1	ham	Ok lar... Joking wif u oni...	NaN	NaN	NaN
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...	NaN	NaN	NaN
3	ham	U dun say so early hor... U c already then say...	NaN	NaN	NaN
4	ham	Nah I don't think he goes to usf, he lives aro...	NaN	NaN	NaN
5	spam	FreeMsg Hey there darling it's been 3 week's n...	NaN	NaN	NaN
6	ham	Even my brother is not like to speak with me. ...	NaN	NaN	NaN
7	ham	As per your request 'Melle Melle (Oru Minnamin...	NaN	NaN	NaN
8	spam	WINNER!! As a valued network customer you have...	NaN	NaN	NaN
9	spam	Had your mobile 11 months or more? U R entitle...	NaN	NaN	NaN

```
d.drop(['Unnamed: 2'],axis=1,inplace=True)


d.drop(['Unnamed: 3'],axis=1,inplace=True)
```

```
d.drop(['Unnamed: 4'],axis=1,inplace=True)


#RENAMING THE VALUES IN A EASIER WAY


#0: Ham, 1: Spam
d['Category']=d['Category'].apply(lambda x: 1 if x=='spam' else 0)
d.head()
```

	Category	Message
0	0	Go until jurong point, crazy.. Available only ...
1	0	Ok lar... Joking wif u oni...
2	1	Free entry in 2 a wkly comp to win FA Cup fina...
3	0	U dun say so early hor... U c already then say...
4	0	Nah I don't think he goes to usf, he lives aro...

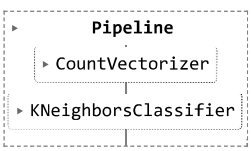
```
X=d['Message']
Y=d['Category']


#TRAINING THE MODEL


X_train, X_test, y_train, y_test = train_test_split(X,Y)


#SUPPORT VECTOR MACHINE ALGORITHM


clf_svm= Pipeline([
    ('vectorizer', CountVectorizer()),
    ('svc', SVC(kernel="rbf",C=1000,gamma=0.001))
])


clf_svm.fit(X_train,y_train)
```




```
y_pred_SVM=clf_svm.predict(X_test)


svm_acc=accuracy_score(y_test,y_pred_SVM)*100
print('THE ACCURACY SCORE IS ',svm_acc)
```

```
     THE ACCURACY SCORE IS  98.49246231155779
```

```
#KNN CLASSIFIER ALGORITHM


clf_knn= Pipeline([
    ('vectorizer', CountVectorizer()),
    ('knn', KNeighborsClassifier(n_neighbors=3))
])


clf_knn.fit(X_train,y_train)
```




```
y_pred_KNN=clf_knn.predict(X_test)
#HIDING WARNINGS REMOVE IF ANY ERROR OCCURS ON CHECKING THIS CODE OR WHILE VALIDATING THIS CODE
```

```
import warnings
warnings.filterwarnings('ignore')
```

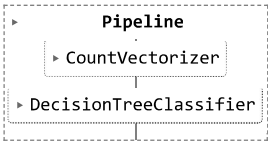
```
knn_acc=accuracy_score(y_test,y_pred_KNN)*100
print('THE ACCURACY SCORE IS ',knn_acc)
```

```
     THE ACCURACY SCORE IS  92.03158650394832
```

```
#DECISION TREE ALGORITHM
```

```
clf_DecisionTree= Pipeline([
    ('vectorizer', CountVectorizer()),
    ('dt',DecisionTreeClassifier())
])
```

```
clf_DecisionTree.fit(X_train,y_train)
```




```
y_pred_DT=clf_DecisionTree.predict(X_test)
```

```
dt_acc=accuracy_score(y_test,y_pred_DT)*100
print('THE ACCURACY SCORE IS ',dt_acc)
```

```
     THE ACCURACY SCORE IS  96.41062455132807
```