In [8]:

```
import numpy as np
import pandas as pd
```

In [9]:

```
url = 'C:\\Users\\mohit\\Desktop\\LP III\\ML\\4\\emails.csv'
df = pd.read_csv(url)
```

In [10]:

```
df
```

Out[10]:

	Email No.	the	to	ect	and	for	of	a	you	hou	...	connevey	jay	valued	lay	infras
0	Email 1	0	0	1	0	0	0	2	0	0	...	0	0	0	0	
1	Email 2	8	13	24	6	6	2	102	1	27	...	0	0	0	0	
2	Email 3	0	0	1	0	0	0	8	0	0	...	0	0	0	0	
3	Email 4	0	5	22	0	5	1	51	2	10	...	0	0	0	0	
4	Email 5	7	6	17	1	5	2	57	0	9	...	0	0	0	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
5167	Email 5168	2	2	2	3	0	0	32	0	0	...	0	0	0	0	
5168	Email 5169	35	27	11	2	6	5	151	4	3	...	0	0	0	0	
5169	Email 5170	0	0	1	1	0	0	11	0	0	...	0	0	0	0	
5170	Email 5171	2	7	1	0	2	1	28	2	0	...	0	0	0	0	
5171	Email 5172	22	24	5	1	6	5	148	8	2	...	0	0	0	0	

5172 rows × 3002 columns

In [11]:

```
df.dropna(inplace = True)
```

In [14]:

```
x = df.drop(['Prediction', 'Email No.'], axis = 1)
y = df['Prediction']
```

In [15]:

```
from sklearn.preprocessing import scale
x = scale(x)
```

In [16]:

```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.2)
```

In [26]:

```
from sklearn.svm import SVC
svc = SVC(C = 0.5)

svc.fit(x_train, y_train)
y_pred = svc.predict(x_test)
```

In [27]:

```
from sklearn import metrics
print(metrics.confusion_matrix(y_test, y_pred))
```

```
[[718   1]
 [104 212]]
```

In [28]:

```
print(metrics.accuracy_score(y_test, y_pred))
```

```
0.8985507246376812
```