### Lab\_2

#### October 31, 2023

**Aim:** 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

Name: Yash Ghorpade

Div: BE-A

**Roll No:** B211046

```
[1]: import pandas as pd
     import numpy as np
     import seaborn as sns
     import matplotlib.pyplot as plt
[3]: df=pd_read_csv("emails.csv")
     df
[3]:
              Email No.
                          the
                                     ect
                                           and
                                                 for
                                                      of
                                                                 you hou
                                to
                                                             a
                                                                                connevey
                                                        0
                                                             2
     0
                Email 1
                             0
                                 0
                                       1
                                             0
                                                                   0
                                                                         0
                                                                                        0
                                                        2
     1
                Email 2
                             8
                               13
                                      24
                                             6
                                                   6
                                                          102
                                                                   1
                                                                        27
                                                                                        0
     2
                Email 3
                             0
                                 0
                                      1
                                             0
                                                   0
                                                       0
                                                             8
                                                                   0
                                                                         0
                                                                                        0
     3
                             0
                                 5
                                      22
                                                   5
                                                            51
                                                                   2
                Email 4
                                             0
                                                                        10
                                                                                        0
     4
                Email 5
                             7
                                 6
                                      17
                                             1
                                                   5
                                                        2
                                                            57
                                                                   0
                                                                         9
                                                                                        0
                                 2
                                                            32
     5167
                             2
                                       2
                                             3
                                                   0
                                                        0
                                                                         0
                                                                                        0
            Email 5168
                                                                   0
                                27
     5168
            Email 5169
                            35
                                      11
                                             2
                                                   6
                                                        5
                                                          151
                                                                   4
                                                                         3
                                                                                        0
                            0
                                 0
                                                                                        0
     5169
            Email 5170
                                                            11
                             2
                                 7
                                                   2
                                                            28
                                                                   2
     5170
            Email 5171
                                             0
                                                        1
                                                                         0
                                                                                        0
                                       5
                            22
                                24
                                                        5 148
                                                                   8
                                                                         2
     5171
            Email 5172
                                                                           ff
            jay
                  valued
                            lay
                                 infrastructure
                                                    military
                                                               allowing
                                                                                dry
     0
               0
                        0
                              0
                                                                            0
                                                                                  0
                                                 0
                                                            0
                                                                        0
     1
               0
                        0
                              0
                                                 0
                                                            0
                                                                        0
                                                                            1
                                                                                  0
     2
               0
                        0
                              0
                                                 0
                                                            0
                                                                        0
                                                                            0
                                                                                  0
     3
               0
                        0
                              0
                                                 0
                                                            0
                                                                                  0
     4
                        0
                                                 0
                                                            0
                                                                        0
                                                                            1
               0
                              0
                                                                                  0
     5167
               0
                        0
                              0
                                                 0
                                                            0
                                                                        0
                                                                            0
                                                                                  0
                        0
                              0
                                                 0
                                                                        0
                                                                            1
                                                                                  0
     5168
               0
                                                            0
```

5169	0	0	0	0	0	0	0	0
5170	0	0	0	0	0	0	1	0
5171	0	0	0	0	0	0	0	0

#### Prediction

0	0
1	0
2	0

3	0
4	0
***	
5167	0
5168	0
5169	1
5170	1
5171	0

[5172 rows x 3002 columns]

# [4]: df.shape

[4]: (5172, 3002)

## [5]: df.size

[5]: 15526344

## [6]: df.head

[6] :			hod NDF			d of	I	Emai	l No	. th	e to	ect	a	nd f	or	of	a
	0		mail 1	0	0	1	0	0	0	2	0	0			(	)	
	1		mail 2	8	13	24	6	6	2	102	1	27			(	)	
	2		mail 3	0	0	1	0	0	0	8	0	0			(	)	
	3		mail 4	0	5	22	0	5	1	51	2	10			(	)	
	4		mail 5	7	6	17	1	5	2	57	0	9			(	)	
		<b>-</b>		•						22		^			,	•	
	5167		il 5168	2	2	2	3	0	0	32	0	0			(	)	
	5168		il 5169	35	27	11	2	6	5	151	4	3			(	)	
	5169		il 5170	0	0	1	1	0	0	11	0	0			(	)	
	5170		il 5171	2	7	<u> </u>	0	2	  -	28	2	0			(	)	
	5171	Ema	il 5172	22	24	5	ı	6	5	148	8	2			(	)	
		jay	valued	lay	in	frastrı	ucture	e m	nilita	ıry	allowi	ng	ff	dry	\		
	0	0	0	0			(	0		0		0	0	0			
	1	0	0	0			(	0		0		0	1	0			
	2	0	0	0			(	0		0		0	0	0			
	3	0	0	0			(	0		0		0	0	0			
	4	0	0	0			(	0		0		0	1	0			
	 5167	. 0		0				0		0		0	0	0			
	5168	0	0	0			Ì	0		0		0	1	0			
	5169	0	0	0			Ì	0		0		0	0	0			
	5170	0	0	0			Ì	0		0		0	1	0			
	5171	0	0	0			·	0		0		0	0	0			

	Prediction
0	0
1	0
2	0
3	0
4	0
5167	0
5168	0
5169	1
5170	1
5171	0

[5172 rows x 3002 columns]>

# [7]: df.tail

[7] :			thod ND			l of	Eı	mail	No	. th	ie to	ect	: a	nd f	or	of	a
	0		mail 1	0	Ò	1	0	0	0	2	0	0				0	
	1		mail 2	8	13	24	6	6	2	102	1	27				0	
	2		mail 3	0	0	1	0	0	0	8	0	0				0	
	3		mail 4	0	5	22	0	5	1	51	2	10				0	
	4		mail 5	7	6	17	1	5	2	57	0	9				0	
	•	_		_ *	Ū	.,			<u>-</u>	٠,		,				•	
	 5167	Emai	1 5168	2	2	2	3	0	0	32		0				0	
	5168		15169	35	27	11	2	6	5	151	4	3				0	
	5169	Emai	l 5170	0	0	1	1	0	0	11	0	0				0	
	5170		l 5171	2	7	1	0	2	1	28	2	0				0	
	5171		l 5172	22	24	5	1	6	5	148	8	2				0	
		iav	valued	lav	in	frastrı	ucture	m	ilita	arv	allow	/ina	ff	dry	\		
	0	0	0	0			0			0		0	0	0	,		
	1	0	0	0			0			0		0	1	0			
	2	0	0	0			0			0		0	0	0			
	3	0	0	0			0			0		0	0	0			
	4	0	0	0			0			0		0	1	0			
		•		Ū								·	•	Ū			
	5167	0	0	0			0			0		0	0	0			
	5168	0	0	0			0			0		0	1	0			
	5169	0	0	0			0			0		0	0	0			
	5170	0	0	0			0			0		0	1	0			
	5171	0	0	0			0			0		0	0	0			

Prediction

0 0

0
0
0
0
0
0
1
1
0

[5172 rows x 3002 columns]>

## [8]: df.describe

[8] :	<boun< th=""><th></th><th>hod NDF u co</th><th></th><th></th><th></th><th>1</th><th>Ema</th><th>il No</th><th>. the</th><th>to</th><th>ec</th><th>t and</th><th>d for</th><th>of</th></boun<>		hod NDF u co				1	Ema	il No	. the	to	ec	t and	d for	of	
	0		nail 1	0	0	` 1	0	0	0	2	0	0			0	
	1		nail 2	8	13	24	6	6	2	102	1	27			0	
	2	Er	mail 3	0	0	1	0	0	0	8	0	0			0	
	3	Er	nail 4	0	5	22	0	5	1	51	2	10			0	
	4	Er	mail 5	7	6	17	1	5	2	57	0	9			0	
	 5167	Emai	<b>.</b> . I 5168	 2	 2	<b></b>	 3	0	0	32		0			0	
	5168	Email	5169	35	27	11	2	6	5	151	4	3			0	
	5169	Emai	l 5170	0	0	1	1	0	0	11	0	0			0	
	5170	Emai	l 5171	2	7	1	0	2	1	28	2	0			0	
	5171	Email	5172	22	24	5	1	6	5	148	8	2			0	
		jay	valued	lay	in	frastru	ıcture	n	nilita	ary	allowi	ng	ff	dry	\	
	0	0	0	0			0			0		0	0	0		
	1	0	0	0			0			0		0	1	0		
	2	0	0	0			0			0		0	0	0		
	3	0	0	0			0			0		0	0	0		
	4	0	0	0			0			0		0	1	0		
	5167	0		0						0		0	0	0		
	5168	0	0	0			0			0		0	1	0		
	5169	0	0	0			0			0		0	0	0		
	5170	0	0	0			0			0		0	1	0		
	5171	0	0	0			0			0		0	0	0		

# Prediction 0

U	U
1	0
2	0
3	0

4	0
5167	0
5168	0
5169	1
5170	1
5171	0

[5172 rows x 3002 columns]>

#### [9]: df.dtypes

[9] : Email No. object the int64 int64 to ect int64 and int64 int64 military allowing int64 ff int64 dry int64 Prediction int64 Length: 3002, dtype: object

#### [10]: df

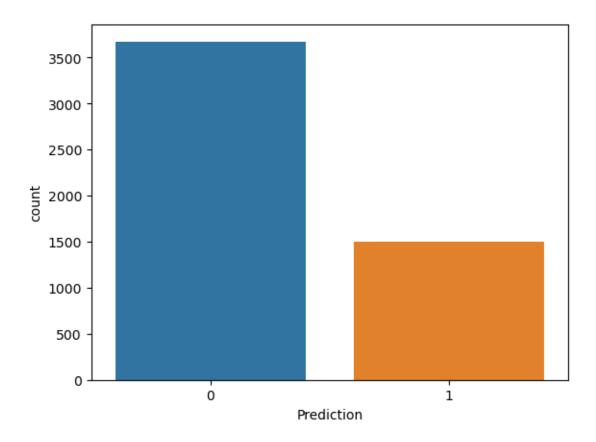
Email No. for [10]: the and of hou connevey to ect a you Email 1 Email 2 Email 3 Email 4 Email 5 **Email 5168 Email 5169 Email 5170** Email 5171 Email 5172 valued lay infrastructure military allowing ff dry \ 

	5167 5168 5169 5170 5171	0 0 0 0		0 0 0 0	0 0 0 0			0 0 0 0		0 0 0 0		0	1 0 1	0 0 0 0			
	0	Pred	dictio	n 0													
	1			0													
	2			0													
	4			0													
	5167 5168			0													
	5169 5170			1 1													
	5171			0													
	[5172	rows	x 30	02 cc	lumn	s]											
[12]:	X=df	f.dro	p(["E	mai <b>l</b>	No.",	"Pr	edict	ion"]	axis,	<b>s</b> =1)							
[13]:	X																
[13]:	•	the	to	ect	and			a	you	hou	in	en	hance	ements	\		
[13]:	0 1	the 0 8	to 0 13	ect 1 24	and 0 6	for 0 6	of 0 2	a 2 102	you 0 1	hou 0 27	in 0 18	en	hance	ements 0 0	\		
[13]:	1 2	0 8 0	0 13 0	1 24 1	0 6 0	0 6 0	0 2 0	2 102 8	0 1 0	0 27 0	0 18 4		hance	0 0 0	\		
[13]:	1	0 8	0 13	1 24	0 6	0 6	0 2	2 102	0 1	0 27	0 18		hance	0	\		
[13]:	1 2 3 4 	0 8 0 0 7	0 13 0 5 6 	1 24 1 22 17  2	0 6 0 0 1 3	0 6 0 5 5	0 2 0 1 2 	2 102 8 51 57 32	0 1 0 2	0 27 0 10 9	0 18 4 1 3  5		hance	0 0 0	\		
[13]:	1 2 3 4  5167 5168	0 8 0 7 2 35	0 13 0 5 6 2 27	1 24 1 22 17  2	0 6 0 0 1 3 2	0 6 0 5 5 0 6	0 2 0 1 2  0 5	2 102 8 51 57 32 151	0 1 0 2 0	0 27 0 10 9	0 18 4 1 3  5 23		hance	0 0 0 0 0	\		
[13]:	1 2 3 4 	0 8 0 0 7	0 13 0 5 6 	1 24 1 22 17  2	0 6 0 0 1 3	0 6 0 5 5	0 2 0 1 2 	2 102 8 51 57 32	0 1 0 2 0	0 27 0 10 9	0 18 4 1 3  5		hance	0 0 0 0 0	\		
[13]:	1 2 3 4  5167 5168 5169	0 8 0 0 7 2 35 0	0 13 0 5 6 	1 24 1 22 17  2 11	0 6 0 0 1 3 2	0 6 0 5 5 0 6 0	0 2 0 1 2  0 5 0	2 102 8 51 57  32 151	0 1 0 2 0	0 27 0 10 9	0 18 4 1 3  5 23 1		hance	0 0 0 0 0	\		
[13]:	1 2 3 4  5167 5168 5169 5170	0 8 0 0 7 2 35 0 2 22	0 13 0 5 6  2 27 0 7 24	1 24 1 22 17  2 11 1	0 6 0 0 1  3 2 1 0	0 6 0 5 5 0 6 0 2 6	0 2 0 1 2  0 5 0	2 102 8 51 57  32 151 11 28 148	0 1 0 2 0  0 4 0 2 8	0 27 0 10 9 0 3 0	0 18 4 1 3  5 23 1 8 23			0 0 0 0 0	•	dry	
[13]:	1 2 3 4  5167 5168 5169 5170	0 8 0 0 7 2 35 0 2 22	0 13 0 5 6  2 27 0 7 24	1 24 1 22 17 2 11 1 5 5 y jay 0	0 6 0 0 1 3 2 1 0	0 6 0 5 5 0 6 0 2 6	0 2 0 1 2  0 5 0 1 5	2 102 8 51 57  32 151 11 28 148	0 1 0 2 0  0 4 0 2 8	0 27 0 10 9 0 3 0 0 2	0 18 4 1 3  5 23 1 8 23	     	all	0 0 0 0 0 0 0 0 0 0 o	ff 0	0	
[13]:	1 2 3 4  5167 5168 5169 5170 5171	0 8 0 0 7 2 35 0 2 22	0 13 0 5 6  2 27 0 7 24	1 24 1 22 17 2 11 1 5 5 jay 0 0	0 6 0 0 1 3 2 1 0	0 6 0 5 5 0 6 0 2 6	0 2 0 1 2  0 5 0 1 5 0	2 102 8 51 57  32 151 11 28 148	0 1 0 2 0  0 4 0 2 8	0 27 0 10 9 0 3 0 0 2	0 18 4 1 3  5 23 1 8 23	     	all	0 0 0 0 0 0 0 0 0 o o	ff 0 1	0	
[13]:	1 2 3 4  5167 5168 5169 5170 5171	0 8 0 0 7 2 35 0 2 22	0 13 0 5 6  2 27 0 7 24 nevey 0 0	1 24 1 22 17 2 11 1 5 5	0 6 0 0 1 3 2 1 0	0 6 0 5 5 0 6 0 2 6	0 2 0 1 2  0 5 0 1 5	2 102 8 51 57  32 151 11 28 148	0 1 0 2 0  0 4 0 2 8	0 27 0 10 9 0 3 0 0 2	0 18 4 1 3  5 23 1 8 23	       (C	all	0 0 0 0 0 0 0 0 0 0 0 0	ff 0	0 0 0 0	
[13]:	1 2 3 4  5167 5168 5169 5170 5171	0 8 0 0 7 2 35 0 2 22	0 13 0 5 6  2 27 0 7 24 nevey 0 0	1 24 1 22 17 2 11 1 5 5 o 0 0 0 0 0 0	0 6 0 1 3 2 1 0 1	0 6 0 5 5  0 6 0 2 6	0 2 0 1 2  0 5 0 1 5 V	2 102 8 51 57  32 151 11 28 148	0 1 0 2 0 4 0 2 8	0 27 0 10 9 0 3 0 0 2 ture 0 0	0 18 4 1 3  5 23 1 8 23	       (0	all	0 0 0 0 0 0 0 0 0 0 0 0	ff 0 1 0	0 0 0	
[13]:	1 2 3 4  5167 5168 5169 5170 5171	0 8 0 0 7 2 35 0 2 22	0 13 0 5 6  2 27 0 7 24 nevey 0 0	1 24 1 22 17 2 11 1 5 5	0 6 0 0 1 3 2 1 0	0 6 0 5 5 0 6 0 2 6	0 2 0 1 2  0 5 0 1 5 V 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	2 102 8 51 57  32 151 11 28 148	0 1 0 2 0 4 0 2 8	0 27 0 10 9 0 3 0 0 2 ture 0 0	0 18 4 1 3  5 23 1 8 23 mi	       0	all(	0 0 0 0 0 0 0 0 0 0 0 0	ff 0 1 0 0	0 0 0 0	

	5169 5170 5171		0 0 0	0 0 0	0 0 0	0 0 0		0 0 0	0 0 0	0 0 0	0 1 0	0 0 0	
	[5172 r	ows x	3000	) colum	ns]								
[14]:	Y=df["P	redict	ion"	']									
[15]:	Υ												
[15]:	0	0											,
	1	0											
	3	0											
	4	0											
	5167	0											
	5168	0											
	5169	1											
	5170	1											
	5171	0											
	Name: P	redict	ion,	Length	ı: 51 <i>7</i> 2	2, dtype	: int64						

[16]: <Axes: xlabel='Prediction', ylabel='count'>

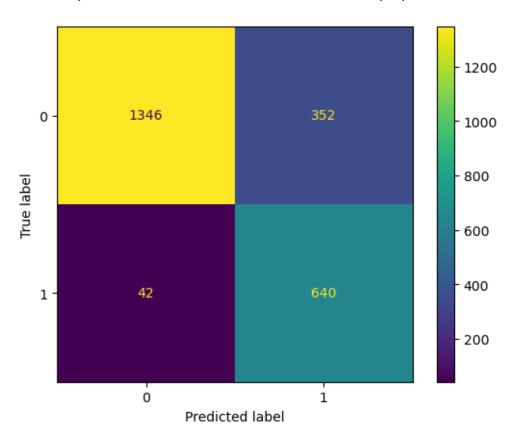
[16]: sns\_countplot(x=Y)



```
[17]: Y.value_counts
[17]: <bound method IndexOpsMixin.value_counts of 0
                                                            0
      2
              0
      3
              0
      4
              0
      5167
              0
      5168
              0
      5169
              1
      5170
              1
      5171
              0
      Name: Prediction, Length: 5172, dtype: int64>
[18]: from sklearn.preprocessing import MinMaxScaler
      scaler=MinMaxScaler()
      X_Scale=scaler_fit_transform(X)
[19]: X_Scale
```

```
[19]: array([[0.
                        , 0.
                                    , 0. , ..., 0.
                                                               , 0.
              0.
             [0.03809524, 0.09848485, 0.06705539, ..., 0.
                                                               . 0.00877193,
              0.
             [0.
                        , 0.
                                          , ..., 0.
                                    , 0.
                                                               , 0.
              0.
                        ],
             [0.
                        , 0.
                                    , 0.
                                                , ..., 0.
                                                                , 0.
              0.
                        ],
             [0.00952381, 0.0530303, 0.
                                                . ..., 0.
                                                               . 0.00877193.
              0.
             [0.1047619, 0.18181818, 0.01166181, ..., 0.
                                                               , 0.
              0.
[20]: from sklearn.model_selection import train_test_split
      X_train, X_test, Y_train, Y_Test=train_test_split(X_Scale,Y,test_size=0.
       46,random_state = 46)
      X_train.shape
[20]: (2792, 3000)
[21]: X_Scale.shape
[21]: (5172, 3000)
[22]: X_test.shape
[22]: (2380, 3000)
[23]: Y_train.shape
[23]: (2792,)
[24]: from sklearn.neighbors import KNeighborsClassifier
      knn= KNeighborsClassifier()
      knn.fit(X_train,Y_train)
[24]: KNeighborsClassifier()
[25]: Y_pred=knn_predict(X_test)
      Y_pred
[25]: array([0, 0, 1, ..., 0, 1, 0], dtype=int64)
       from sklearn.metrics import
[26]:
       -accuracy_score,classification_report,ConfusionMatrixDisplay
      ConfusionMatrixDisplay.from_predictions(Y_Test,Y_pred)
```

[26]: <sklearn.metrics.\_plot.confusion\_matrix.ConfusionMatrixDisplay at 0x208d3e813c0>



#### [27]: accuracy\_score(Y\_Test,Y\_pred)

[27]: 0.8344537815126051

#### [28]: print(classification\_report(Y\_Test,Y\_pred))

	precision	recall	f1-score	support
0	0.97	0.79	0.87	1698
1	0.65	0.94	0.76	682
accuracy			0.83	2380
macro avg	0.81	0.87	0.82	2380
weighted avg	0.88	0.83	0.84	2380

[]: