

# TOPICS IN PATTERN RECOGNITION ASSIGNMENT 1

PAARTH GUPTA

February 11, 2019

The path to the training files is of the form "E://dolphins.csv"  
Different Bayes algorithms are implemented in different files.  
Graphs have been uploaded in the folder

## **PART - 1**

Random Projections algorithm has been implemented and the results have been saved in corresponding files.

### **KEY POINTS :-**

## **PART - 2**

Bayes classifier and Nearest Neighbour classifier are implemented and the accuracies are noted in the files provided.

### **KEY POINTS :-**

## **PART - 3 and 4**

Accuracy and F1-score are calculated for all the values of D and on the whole dataset and the results are plotted in the form of graphs.

Test Accuracy on dolphin using inbuilt bayes and  $d = 2$ : 94.54545454545453

Test Accuracy on dolphin using inbuilt knn and  $d = 2$ : 96.36363636363635

Test Accuracy on dolphin using my bayes and  $d = 2$ : 94.54545454545453

Test Accuracy on dolphin using my knn and  $d = 2$ : 74.54545454545455

Test Macro F1 Score on dolphin using inbuilt bayes and  $d = 2$ : 0.9456613756613755

Test Macro F1 Score on dolphin using inbuilt knn and  $d = 2$ : 0.9661375661375662

Test Macro F1 Score on dolphin using my bayes and  $d = 2$ : 0.9456613756613755

Test Macro F1 Score on dolphin using my knn and  $d = 2$ : 0.6957671957671957

Test Micro F1 Score on dolphin using inbuilt bayes and  $d = 2$ : 0.9454545454545455

Test Micro F1 Score on dolphin using inbuilt knn and  $d = 2$ : 0.9636363636363636

Test Micro F1 Score on dolphin using my bayes and  $d = 2$ : 0.9454545454545455

Test Micro F1 Score on dolphin using my knn and  $d = 2$ : 0.7454545454545455

Test Accuracy on dolphin using inbuilt bayes and  $d = 4$ : 96.36363636363635

Test Accuracy on dolphin using inbuilt knn and  $d = 4$ : 96.36363636363636

Test Accuracy on dolphin using my bayes and  $d = 4$ : 96.36363636363635

Test Accuracy on dolphin using my knn and  $d = 4$ : 96.36363636363636

Test Macro F1 Score on dolphin using inbuilt bayes and  $d = 4$ : 0.9661375661375662

Test Macro F1 Score on dolphin using inbuilt knn and d =4::0.9638095238095239  
Test Macro F1 Score on dolphin using my bayes and d =4::0.9661375661375662  
Test Macro F1 Score on dolphin using my knn and d =4::0.9638095238095239

Test Micro F1 Score on dolphin using inbuilt bayes and d =4::0.9636363636363636  
Test Micro F1 Score on dolphin using inbuilt knn and d =4::0.9636363636363636  
Test Micro F1 Score on dolphin using my bayes and d =4::0.9636363636363636  
Test Micro F1 Score on dolphin using my knn and d =4::0.9636363636363636

Test Accuracy on dolphin using inbuilt bayes and d =8::92.72727272727272  
Test Accuracy on dolphin using inbuilt knn and d =8::92.72727272727272  
Test Accuracy on dolphin using my bayes and d =8::92.72727272727272  
Test Accuracy on dolphin using my knn and d =8::98.18181818181817

Test Macro F1 Score on dolphin using inbuilt bayes and d =8::0.9156613756613756  
Test Macro F1 Score on dolphin using inbuilt knn and d =8::0.9246560846560847  
Test Macro F1 Score on dolphin using my bayes and d =8::0.9156613756613756  
Test Macro F1 Score on dolphin using my knn and d =8::0.9830687830687831

Test Micro F1 Score on dolphin using inbuilt bayes and d =8::0.9272727272727274  
Test Micro F1 Score on dolphin using inbuilt knn and d =8::0.9272727272727274  
Test Micro F1 Score on dolphin using my bayes and d =8::0.9272727272727274  
Test Micro F1 Score on dolphin using my knn and d =8::0.9818181818181818

Test Accuracy on dolphin using inbuilt bayes and d =16::96.36363636363635  
Test Accuracy on dolphin using inbuilt knn and d =16::98.18181818181817  
Test Accuracy on dolphin using my bayes and d =16::96.36363636363635  
Test Accuracy on dolphin using my knn and d =16::96.36363636363635

Test Macro F1 Score on dolphin using inbuilt bayes and d =16::0.9661375661375662  
Test Macro F1 Score on dolphin using inbuilt knn and d =16::0.9830687830687831  
Test Macro F1 Score on dolphin using my bayes and d =16::0.9661375661375662  
Test Macro F1 Score on dolphin using my knn and d =16::0.9661375661375662

Test Micro F1 Score on dolphin using inbuilt bayes and d =16::0.9636363636363636  
Test Micro F1 Score on dolphin using inbuilt knn and d =16::0.9818181818181818  
Test Micro F1 Score on dolphin using my bayes and d =16::0.9636363636363636  
Test Micro F1 Score on dolphin using my knn and d =16::0.9636363636363636

Test Accuracy on dolphin using inbuilt bayes ::92.72727272727272  
Test Accuracy on dolphin using inbuilt knn ::94.54545454545455  
Test Accuracy on dolphin using my bayes ::92.72727272727272  
Test Accuracy on dolphin using my knn ::94.54545454545455

Test Macro F1 Score on dolphin using inbuilt bayes ::0.9261375661375663  
Test Macro F1 Score on dolphin using inbuilt knn ::0.9492063492063492  
Test Macro F1 Score on dolphin using my bayes ::0.9261375661375663  
Test Macro F1 Score on dolphin using my knn ::0.9453968253968255

Test Micro F1 Score on dolphin using inbuilt bayes ::0.9272727272727274  
Test Micro F1 Score on dolphin using inbuilt knn ::0.9454545454545455  
Test Micro F1 Score on dolphin using my bayes ::0.9272727272727274  
Test Micro F1 Score on dolphin using my knn ::0.9454545454545455

Test Accuracy on pubmed using inbuilt bayes and d =2::42.06330031408553  
Test Accuracy on pubmed using inbuilt knn and d =2::34.65571394056535  
Test Accuracy on pubmed using my bayes and d =2::42.06330031408553  
Test Accuracy on pubmed using my knn and d =2::37.78207296448418

Test Macro F1 Score on pubmed using inbuilt bayes and d =2::0.313926782028219  
Test Macro F1 Score on pubmed using inbuilt knn and d =2::0.33866254844166727  
Test Macro F1 Score on pubmed using my bayes and d =2::0.313926782028219  
Test Macro F1 Score on pubmed using my knn and d =2::0.34317179306145856

Test Micro F1 Score on pubmed using inbuilt bayes and d =2::0.42063300314085533  
Test Micro F1 Score on pubmed using inbuilt knn and d =2::0.3465571394056536  
Test Micro F1 Score on pubmed using my bayes and d =2::0.42063300314085533  
Test Micro F1 Score on pubmed using my knn and d =2::0.3778207296448418

Test Accuracy on pubmed using inbuilt bayes and d =4::42.20343078038174  
Test Accuracy on pubmed using inbuilt knn and d =4::33.79077071756463  
Test Accuracy on pubmed using my bayes and d =4::42.20343078038174  
Test Accuracy on pubmed using my knn and d =4::37.26020777965692

Test Macro F1 Score on pubmed using inbuilt bayes and d =4::0.3149518342588216  
Test Macro F1 Score on pubmed using inbuilt knn and d =4::0.329053289940673  
Test Macro F1 Score on pubmed using my bayes and d =4::0.3149518342588216  
Test Macro F1 Score on pubmed using my knn and d =4::0.33500719978015686

Test Micro F1 Score on pubmed using inbuilt bayes and d =4::0.4220343078038174  
Test Micro F1 Score on pubmed using inbuilt knn and d =4::0.3379077071756463  
Test Micro F1 Score on pubmed using my bayes and d =4::0.4220343078038174  
Test Micro F1 Score on pubmed using my knn and d =4::0.3726020777965692

Test Accuracy on pubmed using inbuilt bayes and d =8::41.425465088185554  
Test Accuracy on pubmed using inbuilt knn and d =8::34.81034066199565  
Test Accuracy on pubmed using my bayes and d =8::41.425465088185554  
Test Accuracy on pubmed using my knn and d =8::38.12998308770234

Test Macro F1 Score on pubmed using inbuilt bayes and d =8::0.3100502610100101  
Test Macro F1 Score on pubmed using inbuilt knn and d =8::0.33898740054610527  
Test Macro F1 Score on pubmed using my bayes and d =8::0.3100502610100101  
Test Macro F1 Score on pubmed using my knn and d =8::0.34283599968932915

Test Micro F1 Score on pubmed using inbuilt bayes and d =8::0.41425465088185554  
Test Micro F1 Score on pubmed using inbuilt knn and d =8::0.34810340661995653

Test Micro F1 Score on pubmed using my bayes and d =8::0.41425465088185554  
Test Micro F1 Score on pubmed using my knn and d =8::0.38129983087702346

Test Accuracy on pubmed using inbuilt bayes and d =16::42.89925102681807  
Test Accuracy on pubmed using inbuilt knn and d =16::34.98429572360474  
Test Accuracy on pubmed using my bayes and d =16::42.89925102681807  
Test Accuracy on pubmed using my knn and d =16::37.840057985020536

Test Macro F1 Score on pubmed using inbuilt bayes and d =16::0.3262901193346363  
Test Macro F1 Score on pubmed using inbuilt knn and d =16::0.3413616606302181  
Test Macro F1 Score on pubmed using my bayes and d =16::0.3262901193346363  
Test Macro F1 Score on pubmed using my knn and d =16::0.3427865652045024

Test Micro F1 Score on pubmed using inbuilt bayes and d =16::0.42899251026818075  
Test Micro F1 Score on pubmed using inbuilt knn and d =16::0.34984295723604736  
Test Micro F1 Score on pubmed using my bayes and d =16::0.42899251026818075  
Test Micro F1 Score on pubmed using my knn and d =16::0.37840057985020537

Test Accuracy on pubmed using inbuilt bayes and d =32::43.07320608842716  
Test Accuracy on pubmed using inbuilt knn and d =32::34.38994926310703  
Test Accuracy on pubmed using my bayes and d =32::43.07320608842716  
Test Accuracy on pubmed using my knn and d =32::37.52597245711525

Test Macro F1 Score on pubmed using inbuilt bayes and d =32::0.35645780338803  
Test Macro F1 Score on pubmed using inbuilt knn and d =32::0.3357954667031151  
Test Macro F1 Score on pubmed using my bayes and d =32::0.35645780338803  
Test Macro F1 Score on pubmed using my knn and d =32::0.33972046470389045

Test Micro F1 Score on pubmed using inbuilt bayes and d =32::0.4307320608842716  
Test Micro F1 Score on pubmed using inbuilt knn and d =32::0.3438994926310703  
Test Micro F1 Score on pubmed using my bayes and d =32::0.4307320608842716  
Test Micro F1 Score on pubmed using my knn and d =32::0.3752597245711524

Test Accuracy on pubmed using inbuilt bayes and d =64::42.5610050736893  
Test Accuracy on pubmed using inbuilt knn and d =64::34.844165257308525  
Test Accuracy on pubmed using my bayes and d =64::42.5610050736893  
Test Accuracy on pubmed using my knn and d =64::38.0333413868084

Test Macro F1 Score on pubmed using inbuilt bayes and d =64::0.38297110791003347  
Test Macro F1 Score on pubmed using inbuilt knn and d =64::0.3395359965431111  
Test Macro F1 Score on pubmed using my bayes and d =64::0.38297110791003347  
Test Macro F1 Score on pubmed using my knn and d =64::0.3452304378471532

Test Micro F1 Score on pubmed using inbuilt bayes and d =64::0.4256100507368929  
Test Micro F1 Score on pubmed using inbuilt knn and d =64::0.3484416525730853  
Test Micro F1 Score on pubmed using my bayes and d =64::0.4256100507368929  
Test Micro F1 Score on pubmed using my knn and d =64::0.3803334138680841

Test Accuracy on twitter using inbuilt bayes and d =2::51.58597662771287  
 Test Accuracy on twitter using inbuilt knn and d =2::39.348914858096826  
 Test Accuracy on twitter using my bayes and d =2::51.58597662771287  
 Test Accuracy on twitter using my knn and d =2::43.622704507512516

Test Macro F1 Score on twitter using inbuilt bayes and d =2::0.22687224669603526  
 Test Macro F1 Score on twitter using inbuilt knn and d =2::0.33002569449790814  
 Test Macro F1 Score on twitter using my bayes and d =2::0.22687224669603526  
 Test Macro F1 Score on twitter using my knn and d =2::0.3253593291381102

Test Micro F1 Score on twitter using inbuilt bayes and d =2::0.5158597662771286  
 Test Micro F1 Score on twitter using inbuilt knn and d =2::0.3934891485809683  
 Test Micro F1 Score on twitter using my bayes and d =2::0.5158597662771286  
 Test Micro F1 Score on twitter using my knn and d =2::0.4362270450751252

Test Accuracy on twitter using inbuilt bayes and d =4::51.569282136894834  
 Test Accuracy on twitter using inbuilt knn and d =4::38.03005008347246  
 Test Accuracy on twitter using my bayes and d =4::51.569282136894834  
 Test Accuracy on twitter using my knn and d =4::43.27212020033389

Test Macro F1 Score on twitter using inbuilt bayes and d =4::0.2268487841976028  
 Test Macro F1 Score on twitter using inbuilt knn and d =4::0.3223938374796925  
 Test Macro F1 Score on twitter using my bayes and d =4::0.2268487841976028  
 Test Macro F1 Score on twitter using my knn and d =4::0.3252741621287864

Test Micro F1 Score on twitter using inbuilt bayes and d =4::0.5156928213689482  
 Test Micro F1 Score on twitter using inbuilt knn and d =4::0.38030050083472455  
 Test Micro F1 Score on twitter using my bayes and d =4::0.5156928213689482  
 Test Micro F1 Score on twitter using my knn and d =4::0.43272120200333897

Test Accuracy on twitter using inbuilt bayes and d =8::47.51252086811352  
 Test Accuracy on twitter using inbuilt knn and d =8::40.63439065108514  
 Test Accuracy on twitter using my bayes and d =8::47.51252086811352  
 Test Accuracy on twitter using my knn and d =8::44.524207011686144

Test Macro F1 Score on twitter using inbuilt bayes and d =8::0.3055517831122264  
 Test Macro F1 Score on twitter using inbuilt knn and d =8::0.3480926537019829  
 Test Macro F1 Score on twitter using my bayes and d =8::0.3055517831122264  
 Test Macro F1 Score on twitter using my knn and d =8::0.3436185855931644

Test Micro F1 Score on twitter using inbuilt bayes and d =8::0.47512520868113517  
 Test Micro F1 Score on twitter using inbuilt knn and d =8::0.4063439065108514  
 Test Micro F1 Score on twitter using my bayes and d =8::0.47512520868113517  
 Test Micro F1 Score on twitter using my knn and d =8::0.44524207011686145

Test Accuracy on twitter using inbuilt bayes and d =16::45.425709515859765  
 Test Accuracy on twitter using inbuilt knn and d =16::41.81969949916528  
 Test Accuracy on twitter using my bayes and d =16::45.425709515859765

Test Accuracy on twitter using my knn and d =16::44.624373956594326

Test Macro F1 Score on twitter using inbuilt bayes and d =16::0.3238916591766737  
Test Macro F1 Score on twitter using inbuilt knn and d =16::0.35997469984743724  
Test Macro F1 Score on twitter using my bayes and d =16::0.3238916591766737  
Test Macro F1 Score on twitter using my knn and d =16::0.34629996905000104

Test Micro F1 Score on twitter using inbuilt bayes and d =16::0.4542570951585977  
Test Micro F1 Score on twitter using inbuilt knn and d =16::0.41819699499165275  
Test Micro F1 Score on twitter using my bayes and d =16::0.4542570951585977  
Test Micro F1 Score on twitter using my knn and d =16::0.4462437395659432

Test Accuracy on twitter using inbuilt bayes and d =32::41.26878130217029  
Test Accuracy on twitter using inbuilt knn and d =32::44.59098497495827  
Test Accuracy on twitter using my bayes and d =32::41.26878130217029  
Test Accuracy on twitter using my knn and d =32::48.8313856427379

Test Macro F1 Score on twitter using inbuilt bayes and d =32::0.35452496991094573  
Test Macro F1 Score on twitter using inbuilt knn and d =32::0.38924549916987267  
Test Macro F1 Score on twitter using my bayes and d =32::0.35452496991094573  
Test Macro F1 Score on twitter using my knn and d =32::0.381357470092503

Test Micro F1 Score on twitter using inbuilt bayes and d =32::0.4126878130217029  
Test Micro F1 Score on twitter using inbuilt knn and d =32::0.4459098497495827  
Test Micro F1 Score on twitter using my bayes and d =32::0.4126878130217029  
Test Micro F1 Score on twitter using my knn and d =32::0.4883138564273789

Test Accuracy on twitter using inbuilt bayes and d =64::36.6110183639399  
Test Accuracy on twitter using inbuilt knn and d =64::47.36227045075126  
Test Accuracy on twitter using my bayes and d =64::36.57762938230384  
Test Accuracy on twitter using my knn and d =64::51.151919866444075

Test Macro F1 Score on twitter using inbuilt bayes and d =64::0.3413914828373614  
Test Macro F1 Score on twitter using inbuilt knn and d =64::0.40318068161960186  
Test Macro F1 Score on twitter using my bayes and d =64::0.34111943787064003  
Test Macro F1 Score on twitter using my knn and d =64::0.40231609744408736

Test Micro F1 Score on twitter using inbuilt bayes and d =64::0.36611018363939896  
Test Micro F1 Score on twitter using inbuilt knn and d =64::0.47362270450751254  
Test Micro F1 Score on twitter using my bayes and d =64::0.36577629382303833  
Test Micro F1 Score on twitter using my knn and d =64::0.5115191986644406

Test Accuracy on twitter using inbuilt bayes and d =128::36.86143572621036  
Test Accuracy on twitter using inbuilt knn and d =128::46.76126878130217  
Test Accuracy on twitter using my bayes and d =128::36.644407345575964  
Test Accuracy on twitter using my knn and d =128::50.76794657762938  
Test Macro F1 Score on twitter using inbuilt bayes and d =128::0.3450536425073261  
Test Macro F1 Score on twitter using inbuilt knn and d =128::0.39595248167671077

Test Macro F1 Score on twitter using my bayes and d =128::0.34318193116911005  
Test Macro F1 Score on twitter using my knn and d =128::0.39977217050570973

Test Micro F1 Score on twitter using inbuilt bayes and d =128::0.3686143572621035  
Test Micro F1 Score on twitter using inbuilt knn and d =128::0.46761268781302173  
Test Micro F1 Score on twitter using my bayes and d =128::0.3664440734557596  
Test Micro F1 Score on twitter using my knn and d =128::0.5076794657762939

Test Accuracy on twitter using inbuilt bayes and d =256::32.053422370617696  
Test Accuracy on twitter using inbuilt knn and d =256::48.01335559265442  
Test Accuracy on twitter using my bayes and d =256::41.469115191986646  
Test Accuracy on twitter using my knn and d =256::51.135225375626035  
Test Macro F1 Score on twitter using inbuilt bayes and d =256::0.3158094152957128  
Test Macro F1 Score on twitter using inbuilt knn and d =256::0.4150725998149986  
Test Macro F1 Score on twitter using my bayes and d =256::0.41081336160351014  
Test Macro F1 Score on twitter using my knn and d =256::0.41154495418795956

Test Micro F1 Score on twitter using inbuilt bayes and d =256::0.3205342237061769  
Test Micro F1 Score on twitter using inbuilt knn and d =256::0.4801335559265443  
Test Micro F1 Score on twitter using my bayes and d =256::0.41469115191986646  
Test Micro F1 Score on twitter using my knn and d =256::0.5113522537562604

Test Accuracy on twitter using inbuilt bayes and d =512::32.454090150250416  
Test Accuracy on twitter using inbuilt knn and d =512::48.981636060100165  
Test Accuracy on twitter using my bayes and d =512::41.202003338898162  
Test Accuracy on twitter using my knn and d =512::52.07011686143572  
Test Macro F1 Score on twitter using inbuilt bayes and d =512::0.32106785883934874  
Test Macro F1 Score on twitter using inbuilt knn and d =512::0.42395446006357407  
Test Macro F1 Score on twitter using my bayes and d =512::0.409922619666259  
Test Macro F1 Score on twitter using my knn and d =512::0.4179460697145753

Test Micro F1 Score on twitter using inbuilt bayes and d =512::0.3245409015025042  
Test Micro F1 Score on twitter using inbuilt knn and d =512::0.48981636060100164  
Test Micro F1 Score on twitter using my bayes and d =512::0.4120200333889816  
Test Micro F1 Score on twitter using my knn and d =512::0.5207011686143572

Test Accuracy on twitter using inbuilt bayes and d =1024::31.986644407345572  
Test Accuracy on twitter using inbuilt knn and d =1024::48.898163606010016  
Test Accuracy on twitter using my bayes and d =1024::47.69616026711185  
Test Accuracy on twitter using my knn and d =1024::52.78797996661102  
Test Macro F1 Score on twitter using inbuilt bayes and d =1024::0.3129348915403053  
Test Macro F1 Score on twitter using inbuilt knn and d =1024::0.41684268329923835  
Test Macro F1 Score on twitter using my bayes and d =1024::0.30341482806635706  
Test Macro F1 Score on twitter using my knn and d =1024::0.42171788539663646

Test Micro F1 Score on twitter using inbuilt bayes and d =1024::0.31986644407345577  
Test Micro F1 Score on twitter using inbuilt knn and d =1024::0.4889816360601002  
Test Micro F1 Score on twitter using my bayes and d =1024::0.47696160267111853

Test Micro F1 Score on twitter using my knn and d =1024::0.5278797996661102

Test Accuracy on twitter using smoothing bayes::58.24707846410685

Test Macro F1 Score on twitter using smoothing bayes::0.510286500611228

Test Micro F1 Score on twitter using smoothing bayes::0.5824707846410685

## **KEY POINTS :-**

### **PART - 5**

## **KEY POINTS :-**

### **PART - 6**

## **KEY POINTS :-**

### **PART - 7**

Test Accuracy on dataset 1 using LSH on d = 2=67.27272727272728

Test Accuracy on dataset 1 using PCA on d = 2=94.54545454545453

Macro Score on dataset 1 using LSH on d = 2=0.5304088504088504

Macro Score on dataset 1 using PCA on d = 2=0.9066666666666666

Micro Score on dataset 1 using LSH on d = 2=0.6727272727272726

Micro Score on dataset 1 using PCA on d = 2=0.9454545454545455

Test Accuracy on dataset 1 using LSH on d = 4=69.0909090909091

Test Accuracy on dataset 1 using PCA on d = 4=94.54545454545455

Macro Score on dataset 1 using LSH on d = 4=0.546969696969697

Macro Score on dataset 1 using PCA on d = 4=0.9364021164021163

Micro Score on dataset 1 using LSH on d = 4=0.6909090909090909

Micro Score on dataset 1 using PCA on d = 4=0.9454545454545455

Test Accuracy on dataset 1 using LSH on d = 8=83.63636363636365

Test Accuracy on dataset 1 using PCA on d = 8=94.54545454545453

Macro Score on dataset 1 using LSH on d = 8=0.8133525733525733

Macro Score on dataset 1 using PCA on d = 8=0.9456613756613758

Micro Score on dataset 1 using LSH on d = 8=0.8363636363636365

Micro Score on dataset 1 using PCA on d = 8=0.9454545454545455

Test Accuracy on dataset 1 using LSH on d = 16=96.36363636363635

Test Accuracy on dataset 1 using PCA on d = 16=92.72727272727272

Macro Score on dataset 1 using LSH on d = 16=0.9623280423280424

Macro Score on dataset 1 using PCA on d = 16=0.9301587301587301

Micro Score on dataset 1 using LSH on d = 16=0.9636363636363636

Micro Score on dataset 1 using PCA on d = 16=0.9272727272727274

Test Accuracy on dataset 1 using LSH on d = 32=90.9090909090909

Test Accuracy on dataset 1 using PCA on d = 32=92.72727272727272

Macro Score on dataset 1 using LSH on d = 32=0.8980952380952381

Macro Score on dataset 1 using PCA on d = 32=0.9293314093314095

Micro Score on dataset 1 using LSH on d = 32=0.909090909090909

Micro Score on dataset 1 using PCA on d = 32=0.9272727272727274

## **KEY POINTS :-**