Experiment 2.2

Student Name: Mayank Kumar **UID: 20BCS1353**
Branch: CSE **Section/Group:20BCS_DM-705 A**
Semester: 6th **Date of Performance:07/04/2023**
Subject Name: Data Mining Lab **Subject Code: 20CSP-376**

1. Aim:

To perform classification using Bayesian classification algorithm using R.

2. Objective:

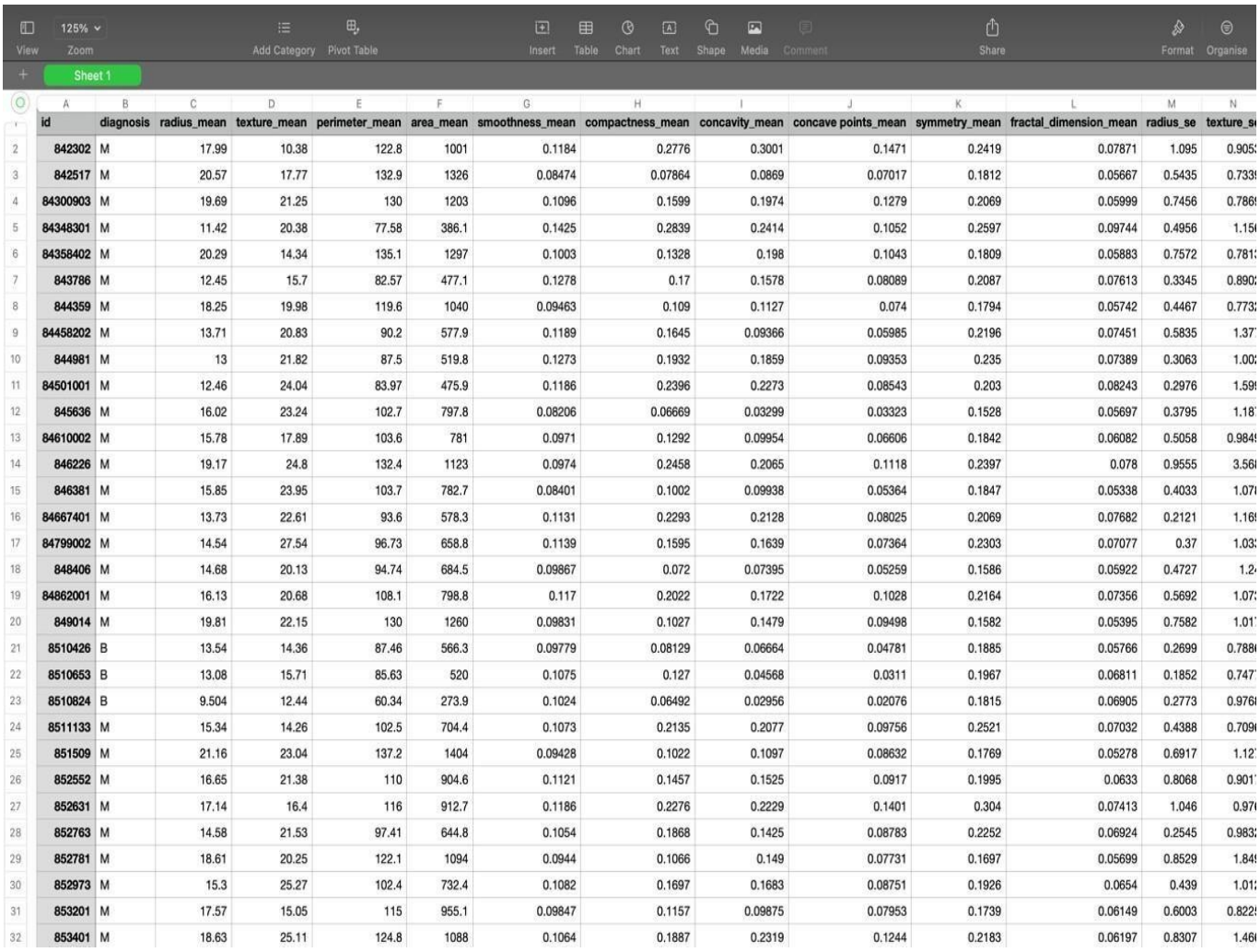
Performing Bayesian classification algorithm using R.

3. Script and Output:

```
library(e1071)
library(caTools)
library(caret)

bc=read.csv("/Users/Akanksha/Documents/Semester 6th/Data Mining
Lab/Breast_Cancer.csv")
nrow(bc)
train =bc[1:450,]
test =bc[451:569,]
head(test)
levels(train$diagnosis)

model=naiveBayes(diagnosis~.,data = train)
pred= predict(model,test)
table(pred)
pred
table(test$diagnosis)
```

```
#confusionMatrix
m1 <-table(test$diagnosis, pred)
library(scales)
accuracy=percent((85+25)/119)
accuracy

confusionMatrix(m1)
```

4. Dataset Used:



	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	symmetry_mean	fractal_dimension_mean	radius_se	texture_se
2	842302	M	17.99	10.38	122.8	1001	0.1184	0.2776	0.3001	0.1471	0.2419	0.07871	1.095	0.905
3	842517	M	20.57	17.77	132.9	1326	0.08474	0.07864	0.0869	0.07017	0.1812	0.05667	0.5435	0.733
4	84300903	M	19.69	21.25	130	1203	0.1096	0.1599	0.1974	0.1279	0.2069	0.05999	0.7456	0.786
5	84348301	M	11.42	20.38	77.58	386.1	0.1425	0.2839	0.2414	0.1052	0.2597	0.09744	0.4956	1.15
6	84358402	M	20.29	14.34	135.1	1297	0.1003	0.1328	0.198	0.1043	0.1809	0.05883	0.7572	0.781
7	843786	M	12.45	15.7	82.57	477.1	0.1278	0.17	0.1578	0.08089	0.2087	0.07613	0.3345	0.890
8	844359	M	18.25	19.98	119.6	1040	0.09463	0.109	0.1127	0.074	0.1794	0.05742	0.4467	0.773
9	84458202	M	13.71	20.83	90.2	577.9	0.1189	0.1645	0.09366	0.05985	0.2196	0.07451	0.5835	1.37
10	844981	M	13	21.82	87.5	519.8	0.1273	0.1932	0.1859	0.09353	0.235	0.07389	0.3063	1.00
11	84501001	M	12.46	24.04	83.97	475.9	0.1186	0.2396	0.2273	0.08543	0.203	0.08243	0.2976	1.59
12	845636	M	16.02	23.24	102.7	797.8	0.08206	0.06669	0.03299	0.03323	0.1528	0.05697	0.3795	1.18
13	84610002	M	15.78	17.89	103.6	781	0.0971	0.1292	0.09954	0.06606	0.1842	0.06082	0.5058	0.984
14	846226	M	19.17	24.8	132.4	1123	0.0974	0.2458	0.2065	0.1118	0.2397	0.078	0.9555	3.56
15	846381	M	15.85	23.95	103.7	782.7	0.08401	0.1002	0.09938	0.05364	0.1847	0.05338	0.4033	1.07
16	84667401	M	13.73	22.61	93.6	578.3	0.1131	0.2293	0.2128	0.08025	0.2069	0.07682	0.2121	1.16
17	84799002	M	14.54	27.54	96.73	658.8	0.1139	0.1595	0.1639	0.07364	0.2303	0.07077	0.37	1.03
18	848406	M	14.68	20.13	94.74	684.5	0.09867	0.072	0.07395	0.05259	0.1586	0.05922	0.4727	1.24
19	84862001	M	16.13	20.68	108.1	798.8	0.117	0.2022	0.1722	0.1028	0.2164	0.07356	0.5692	1.07
20	849014	M	19.81	22.15	130	1260	0.09831	0.1027	0.1479	0.09498	0.1582	0.05395	0.7582	1.01
21	8510426	B	13.54	14.36	87.46	566.3	0.09779	0.08129	0.06664	0.04781	0.1885	0.05766	0.2699	0.788
22	8510653	B	13.08	15.71	85.63	520	0.1075	0.127	0.04568	0.0311	0.1967	0.06811	0.1852	0.747
23	8510824	B	9.504	12.44	60.34	273.9	0.1024	0.06492	0.02956	0.02076	0.1815	0.06905	0.2773	0.976
24	8511133	M	15.34	14.26	102.5	704.4	0.1073	0.2135	0.2077	0.09756	0.2521	0.07032	0.4388	0.709
25	851509	M	21.16	23.04	137.2	1404	0.09428	0.1022	0.1097	0.08632	0.1769	0.05278	0.6917	1.12
26	852552	M	16.65	21.38	110	904.6	0.1121	0.1457	0.1525	0.0917	0.1995	0.0633	0.8068	0.901
27	852631	M	17.14	16.4	116	912.7	0.1186	0.2276	0.2229	0.1401	0.304	0.07413	1.046	0.97
28	852763	M	14.58	21.53	97.41	644.8	0.1054	0.1868	0.1425	0.08783	0.2252	0.06924	0.2545	0.983
29	852781	M	18.61	20.25	122.1	1094	0.0944	0.1066	0.149	0.07731	0.1697	0.05699	0.8529	1.84
30	852973	M	15.3	25.27	102.4	732.4	0.1082	0.1697	0.1683	0.08751	0.1926	0.0654	0.439	1.01
31	853201	M	17.57	15.05	115	955.1	0.09847	0.1157	0.09875	0.07953	0.1739	0.06149	0.6003	0.822
32	853401	M	18.63	25.11	124.8	1088	0.1064	0.1887	0.2319	0.1244	0.2183	0.06197	0.8307	1.46

5. Output:

```
R 4.1.2 · ~/
> library(e1071)
> library(gmodels)
> library(dplyr)
> iris <-read.csv('iris_1.csv')
> str(iris)
'data.frame':   150 obs. of  5 variables:
 $ sepal_length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ sepal_width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ petal_length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ petal_width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ class       : chr  "Iris-setosa" "Iris-setosa" "Iris-setosa" "Iris-setosa" ...
> summary(iris)
  sepal_length    sepal_width     petal_length    petal_width       class
 Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100   Length:150
 1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300   Class :character
 Median :5.800   Median :3.000   Median :4.350   Median :1.300   Mode  :character
 Mean   :5.843   Mean   :3.054   Mean   :3.759   Mean   :1.199
 3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
 Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
>
> index = sample(2,nrow(iris),prob=c(0.8,0.2),replace=TRUE)
>
> set.seed(1234)
> #training set
> train= iris[index==1,]
> #testing set
> test= iris[index==2,]
> #test_data will be given as an input to the model to predict species
> test_data = test[1:4]
> #test_labels are the actual values of species of the test data
> test_label=test[,5]
> print(test_label)
 [1] "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"
 [6] "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-versicolor" "Iris-versicolor"
[11] "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor"
[16] "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"
[21] "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"
[26] "Iris-virginica"  "Iris-virginica"


> model=naiveBayes(train$class~.,train)
> model

Naive Bayes Classifier for Discrete Predictors

Call:
naiveBayes.default(x = X, y = Y, laplace = laplace)

A-priori probabilities:
Y
    Iris-setosa Iris-versicolor  Iris-virginica
      0.3414634       0.3495935       0.3089431

Conditional probabilities:
                 sepal_length
Y                     [,1]      [,2]
  Iris-setosa     5.009524 0.3369950
  Iris-versicolor 5.986047 0.5138934
  Iris-virginica  6.586842 0.6597178

                 sepal_width
Y                     [,1]      [,2]
  Iris-setosa     3.419048 0.3690714
  Iris-versicolor 2.783721 0.2910819
  Iris-virginica  2.968421 0.3425700

                 petal_length
Y                     [,1]      [,2]
  Iris-setosa     1.473810 0.1767726
  Iris-versicolor 4.295349 0.4765718
  Iris-virginica  5.568421 0.5723849

                 petal_width
Y                     [,1]      [,2]
  Iris-setosa     0.250000 0.1109823
  Iris-versicolor 1.330233 0.1994455
  Iris-virginica  2.023684 0.2755184
```

```
> test_result=predict(model,test_data)
> test_result
 [1] Iris-setosa     Iris-setosa     Iris-setosa     Iris-setosa     Iris-setosa
 [6] Iris-setosa     Iris-setosa     Iris-setosa     Iris-versicolor Iris-versicolor
[11] Iris-versicolor Iris-versicolor Iris-versicolor Iris-versicolor Iris-versicolor
[16] Iris-virginica  Iris-virginica  Iris-virginica  Iris-virginica  Iris-virginica
[21] Iris-virginica  Iris-virginica  Iris-versicolor Iris-virginica  Iris-virginica
[26] Iris-virginica  Iris-virginica
Levels: Iris-setosa Iris-versicolor Iris-virginica
> #compare the predicted and actual values
> CrossTable(x=test_label, y=test_result)
```


```
> #compare the predicted and actual values
> CrossTable(x=test_label, y=test_result)
```


```
   Cell Contents
|-------------------------|
|                       N |
| Chi-square contribution |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|


Total Observations in Table:  27
```

test_label	test_result: Iris-setosa	Iris-versicolor	Iris-virginica	Row Total
Iris-setosa	8	0	0	8
	13.370	2.370	3.259	
	1.000	0.000	0.000	0.296
	1.000	0.000	0.000	
	0.296	0.000	0.000	
Iris-versicolor	0	7	0	7
	2.074	11.699	2.852	
	0.000	1.000	0.000	0.259
	0.000	0.875	0.000	
	0.000	0.259	0.000	
Iris-virginica	0	1	11	12
	3.556	1.837	7.639	
	0.000	0.083	0.917	0.444
	0.000	0.125	1.000	
	0.000	0.037	0.407	
Column Total	8	8	11	27
	0.296	0.296	0.407	

```
>
```