Universidade Federal do Sul e Sudeste do Pará
Faculdade de Computação e Engenharia Elétrica
Atividade Avaliativa 05

Aluno: Iago Costa das Flores
Matrícula: 201840601017
Curso: Engenharia da Computação 2018
Disciplina: Inteligência Artificial
Professor: Dr. Elton Alves

Marabá, PA - 2022

Questão 01:

• Utilizando a base de dados “breast-câncer”, aplique “Naives Bayes, Árvore de decisão e Random Florest” para classificar os registros.

• Os resultados obtidos devem ser avaliados e comparados em termos de: ROC, F1, Especificidade, Precisão, acurácia e Recall.

• Os resultados devem ser comentados em um pequeno relatório.

Relatório:

Algoritmos	ROC	F1	Especificidade	Precisão	Acurácia	Recall
Naives Bayes	0,701	0,708	0,435	0,704	0,716	0,717
Árvore de decisão	0,584	0,713	0,270	0,752	0,755	0,755
Random Forest	0,634	0,669	0,282	0,664	0,695	0,696

Pelo valor de ROC o Naives Bayes é destaque com a melhor performance e a Árvore de Decisão tem a pior performance.

Pelo valor de F1 o Árvore de decisão é destaque com a melhor performance e o Random Forest tem a pior performance.

Pelo valor de Especificidade o Naives Bayes é destaque com a melhor performance e a Árvore de Decisão tem a pior performance.

Pelo valor de Precisão o Árvore de decisão é destaque com a melhor performance e a Random Forest tem a pior performance.

Pelo valor de Acurácia o Árvore de decisão é destaque com a melhor performance e a Random Forest tem a pior performance.

Pelo valor de Recall o Árvore de decisão é destaque com a melhor performance e a Random Forest tem a pior performance.

Log pelo Weka:

Naive Bayes

=== Run information ===

Scheme: weka.classifiers.bayes.NaiveBayes
Relation: breast-cancer
Instances: 286
Attributes: 10
age
menopause
tumor-size
inv-nodes
node-caps
deg-malig
breast
breast-quad
irradiat
Class
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Naive Bayes Classifier

Attribute	Class no-recurrence-events (0.7)	 recurrence-events (0.3)
age		
10-19	1.0	1.0
20-29	2.0	1.0
30-39	22.0	16.0
40-49	64.0	28.0
50-59	72.0	26.0
60-69	41.0	18.0
70-79	6.0	2.0
80-89	1.0	1.0
90-99	1.0	1.0
[total]	210.0	94.0
menopause		
lt40	6.0	3.0
ge40	95.0	36.0
premeno	103.0	49.0
[total]	204.0	88.0
tumor-size		
0-4	8.0	2.0
5-9	5.0	1.0
10-14	28.0	2.0
15-19	24.0	8.0
20-24	35.0	17.0

25-29	37.0	19.0
30-34	36.0	26.0
35-39	13.0	8.0
40-44	17.0	7.0
45-49	3.0	2.0
50-54	6.0	4.0
55-59	1.0	1.0
[total]	213.0	97.0
inv-nodes		
0-2	168.0	47.0
3-5	20.0	18.0
6-8	8.0	11.0
9-11	5.0	7.0
12-14	2.0	3.0
15-17	4.0	4.0
18-20	1.0	1.0
21-23	1.0	1.0
24-26	1.0	2.0
27-29	1.0	1.0
30-32	1.0	1.0
33-35	1.0	1.0
36-39	1.0	1.0
[total]	214.0	98.0
node-caps		
yes	26.0	32.0
no	172.0	52.0
[total]	198.0	84.0
deg-malig		
1	60.0	13.0
2	103.0	29.0
3	41.0	46.0
[total]	204.0	88.0
breast		
left	104.0	50.0
right	99.0	37.0
[total]	203.0	87.0
breast-quad		
left_up	72.0	27.0
left_low	76.0	36.0
right_up	21.0	14.0
right_low	19.0	7.0
central	18.0	5.0
[total]	206.0	89.0

irradiat

yes	38.0	32.0
no	165.0	55.0
[total]	203.0	87.0

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	205	71.6783 %
Incorrectly Classified Instances	81	28.3217 %
Kappa statistic	0.2857	
Mean absolute error	0.3272	
Root mean squared error	0.4534	
Relative absolute error	78.2086 %	
Root relative squared error	99.1872 %	
Total Number of Instances	286	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,836	0,565	0,778	0,836	0,806	0,288	0,701	0,837	no-recurrence-events
	0,435	0,164	0,529	0,435	0,477	0,288	0,701	0,514	recurrence-events
Weighted Avg.	0,717	0,446	0,704	0,717	0,708	0,288	0,701	0,741	

=== Confusion Matrix ===

a	b	<-- classified as
168	33	a = no-recurrence-events
48	37	b = recurrence-events

Árvore de Decisão

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: breast-cancer
Instances: 286
Attributes: 10

```
            age
            menopause
            tumor-size
            inv-nodes
            node-caps
            deg-malig
            breast
            breast-quad
            irradiat
            Class
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
------------------

node-caps = yes
|   deg-malig = 1: recurrence-events (1.01/0.4)
|   deg-malig = 2: no-recurrence-events (26.2/8.0)
|   deg-malig = 3: recurrence-events (30.4/7.4)
node-caps = no: no-recurrence-events (228.39/53.4)

Number of Leaves  :   4

Size of the tree :        6


Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         216               75.5245 %
Incorrectly Classified Instances        70               24.4755 %
Kappa statistic                          0.2826
Mean absolute error                      0.3676
Root mean squared error                  0.4324
Relative absolute error                 87.8635 %
Root relative squared error             94.6093 %
Total Number of Instances              286

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area
Class
               0,960    0,729    0,757      0,960    0,846      0,339    0,584     0,736
no-recurrence-events
```

0,271 0,040 0,742 0,271 0,397 0,339 0,584 0,436
recurrence-events
Weighted Avg. 0,755 0,524 0,752 0,755 0,713 0,339 0,584 0,647

=== Confusion Matrix ===

a b <-- classified as
193 8 | a = no-recurrence-events
62 23 | b = recurrence-events

Random Forest

=== Run information ===

Scheme: weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation: breast-cancer
Instances: 286
Attributes: 10
age
menopause
tumor-size
inv-nodes
node-caps
deg-malig
breast
breast-quad
irradiat
Class
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.13 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances 199 69.5804 %

Incorrectly Classified Instances 87 30.4196 %
Kappa statistic 0.1736
Mean absolute error 0.3727
Root mean squared error 0.4613
Relative absolute error 89.0857 %
Root relative squared error 100.9171 %
Total Number of Instances 286

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,871	0,718	0,742	0,871	0,801	0,184	0,634	0,798	no-recurrence-events
0,282	0,129	0,480	0,282	0,356	0,184	0,634	0,409	recurrence-events
Weighted Avg. 0,696	0,543	0,664	0,696	0,669	0,184	0,634	0,682	

=== Confusion Matrix ===

a	b	<-- classified as
175	26	a = no-recurrence-events
61	24	b = recurrence-events