Name:Islame,Rafiqul **Id:17-34438-1** **Sec:A(Datamining)**

Problem

Predicting the age of abalone from physical measurements. The age of abalone is determined by cutting the shell through the cone, staining it and counting the number of rings through a microscope -- a boring and time-consuming task. Other measurements, which are easier to obtain, are used to predict the age. Further information, such as weather patterns and location (hence food availability) may be required to solve the problem.

Dataset Preparation:

Here, I have declared a new attribute Age_class in my data set. Age_class has been defined according to the ring number.
Age= (Ring Number + 1.5) Years
I have classified age class as,

Ring no. <=10 | Young and
Ring no. >10 | Adult

No. Rings	Age	Age_class
1	2.5	Young
2	3.5	Young
3	4.5	Young
4	5.5	Young
5	6.5	Young

6	7.5	Young
7	8.5	Young
8	9.5	Young
9	10.5	Young
10	11.5	Young
11	12.5	Adult
12	13.5	Adult
13	14.5	Adult
14	15.5	Adult
15	16.5	Adult
16	17.5	Adult
17	18.5	Adult
18	19.5	Adult
19	20.5	Adult
20	21.5	Adult
21	22.5	Adult
22	23.5	Adult
23	24.5	Adult
24	25.5	Adult
25	26.5	Adult
26	27.5	Adult
27	28.5	Adult
29	30.5	Adult

Naïve Bayes Classifier

=== Run information ===

Scheme: weka.classifiers.bayes.NaiveBayes

Relation: abalone

Instances: 4177

Attributes: 10

Sex

Length

Diameter

Height

Whole_weight

Shucked_weight

Viscera_weight

Shell_weight

Rings

Age_Class

Test mode: evaluate on training data

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

Naive Bayes Classifier

	Class	
Attribute	Young	Adult
	(0.65)	(0.35)
Sex		
M	872.0	658.0
F	673.0	636.0
I	1188.0	156.0
[total]	2733.0	1450.0
Length		
mean	0.4883	0.591
std. dev.	0.1215	0.0832
weight sum	2730	1447
precision	0.0056	0.0056
Diameter		
mean	0.3772	0.4657

std. dev.	0.0993	0.0684
weight sum	2730	1447
precision	0.0054	0.0054

Height

mean	0.1274	0.1646
std. dev.	0.0427	0.0302
weight sum	2730	1447
precision	0.0226	0.0226

Whole_weight

mean	0.6721	1.1242
std. dev.	0.4307	0.4588
weight sum	2730	1447
precision	0.0012	0.0012

Shucked_weight

mean	0.3039	0.464
std. dev.	0.2034	0.2176
weight sum	2730	1447
precision	0.001	0.001

Viscera_weight

mean	0.1473	0.2434
std. dev.	0.0975	0.1035
weight sum	2730	1447
precision	0.0009	0.0009

Shell_weight

mean	0.188	0.3348
std. dev.	0.1145	0.1307
weight sum	2730	1447
precision	0.0011	0.0011

Rings

mean	8.4433	13.5467
std. dev.	1.6611	2.508
weight sum	2730	1447
precision	1.037	1.037

Time taken to build model: 0.19 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.26 seconds

=== Summary ===

Correctly Classified Instances	3232	77.3761 %
Incorrectly Classified Instances	945	22.6239 %
Kappa statistic	0.5369	
Mean absolute error	0.2281	
Root mean squared error	0.4385	
Relative absolute error	50.3704 %	
Root relative squared error	92.1544 %	
Total Number of Instances	4177	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.738	0.159	0.898	0.738	0.810	0.553	0.890	0.938	Young
	0.841	0.262	0.630	0.841	0.720	0.553	0.890	0.829	Adult
Weighted Avg.	0.774	0.195	0.805	0.774	0.779	0.553	0.890	0.900	

=== Confusion Matrix ===

a	b	<-- classified as
2015	715	a = Young
230	1217	b = Adult

VotedPerceptron Classifier:

=== Run information ===

Scheme: weka.classifiers.functions.VotedPerceptron -I 1 -E 1.0 -S 1 -M 10000

Relation: abalone

Instances: 4177

Attributes: 10

- Sex
- Length
- Diameter
- Height
- Whole_weight
- Shucked_weight
- Viscera_weight
- Shell_weight
- Rings

Age_Class

Test mode: evaluate on training data

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

VotedPerceptron: Number of perceptrons=981

Time taken to build model: 0.37 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.83 seconds

=== Summary ===

Correctly Classified Instances	4010	96.0019 %
Incorrectly Classified Instances	167	3.9981 %
Kappa statistic	0.9114	
Mean absolute error	0.04	
Root mean squared error	0.1999	
Relative absolute error	8.8282 %	
Root relative squared error	42.0195 %	
Total Number of Instances	4177	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.973	0.065	0.966	0.973	0.970	0.911	0.957	0.960	Young
0.935	0.027	0.949	0.935	0.942	0.911	0.967	0.931	Adult

Weighted Avg. 0.960 0.052 0.960 0.960 0.960 0.911 0.961 0.950

=== Confusion Matrix ===

a	b	<-- classified as
2657	73	a = Young
94	1353	b = Adult

BayesNet Classifiers:

=== Run information ===

Scheme: weka.classifiers.bayes.BayesNet -D -Q weka.classifiers.bayes.net.search.local.K2 -- -P 1 -S BAYES -E weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5

Relation: abalone

Instances: 4177

Attributes: 10

- Sex
- Length
- Diameter
- Height
- Whole_weight
- Shucked_weight
- Viscera_weight
- Shell_weight
- Rings
- Age_Class

Test mode: evaluate on training data

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

Bayes Network Classifier

not using ADTree

#attributes=10 #classindex=9

Network structure (nodes followed by parents)

Sex(3): Age_Class

Length(5): Age_Class

Diameter(6): Age_Class

Height(6): Age_Class

Whole_weight(6): Age_Class

Shucked_weight(5): Age_Class

Viscera_weight(5): Age_Class

Shell_weight(6): Age_Class

Rings(2): Age_Class

Age_Class(2):

LogScore Bayes: -48709.90597015288

LogScore BDeu: -48856.51113759298

LogScore MDL: -48858.1191709649

LogScore ENTROPY: -48562.14329692525

LogScore AIC: -48633.14329692525

Time taken to build model: 0.07 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.08 seconds

=== Summary ===

Correctly Classified Instances	4047	96.8877 %
Incorrectly Classified Instances	130	3.1123 %
Kappa statistic	0.9318	
Mean absolute error	0.0358	
Root mean squared error	0.159	
Relative absolute error	7.8961 %	
Root relative squared error	33.4211 %	
Total Number of Instances	4177	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.968	0.029	0.984	0.968	0.976	0.932	0.993	0.996	Young
	0.971	0.032	0.941	0.971	0.956	0.932	0.993	0.992	Adult
Weighted Avg.	0.969	0.030	0.969	0.969	0.969	0.932	0.993	0.994	

=== Confusion Matrix ===

a	b	<-- classified as
2642	88	a = Young
42	1405	b = Adult

Vote Classifiers:

=== Run information ===

Scheme: weka.classifiers.meta.Vote -S 1 -B "weka.classifiers.rules.ZeroR " -R AVG

Relation: abalone

Instances: 4177

Attributes: 10

Sex

Length

Diameter

Height

Whole_weight

Shucked_weight

Viscera_weight

Shell_weight

Rings

Age_Class

Test mode: evaluate on training data

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

Vote combines the probability distributions of these base learners:

weka.classifiers.rules.ZeroR

using the 'Average' combination rule

All the models:

ZeroR predicts class value: Young

Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.05 seconds

=== Summary ===

Correctly Classified Instances	2730	65.3579 %
Incorrectly Classified Instances	1447	34.6421 %
Kappa statistic	0	
Mean absolute error	0.4528	
Root mean squared error	0.4758	
Relative absolute error	100	%
Root relative squared error	100	%
Total Number of Instances	4177	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	1.000	0.654	1.000	0.791	?	0.500	0.654	Young
	0.000	0.000	?	0.000	?	?	0.500	0.346	Adult
Weighted Avg.	0.654	0.654	?	0.654	?	?	0.500	0.547	

=== Confusion Matrix ===

a	b	<-- classified as
2730	0	a = Young
1447	0	b = Adult

RandomTree Classifiers

=== Run information ===

Scheme: weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1

Relation: abalone

Instances: 4177

Attributes: 10

Sex

Length

Diameter

Height

Whole_weight

Shucked_weight

Viscera_weight

Shell_weight

Rings

Age_Class

Test mode: evaluate on training data

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

RandomTree

==========

Shell_weight < 0.18

| Shell_weight < 0.11

| | Diameter < 0.23 : Young (258/0)

| | Diameter >= 0.23

| | | Sex = M

| | | | Whole_weight < 0.32

| | | | | Rings < 10.5 : Young (69/0)

| | | | | Rings >= 10.5 : Adult (14/0)

| | | | Whole_weight >= 0.32 : Young (17/0)

| | | Sex = F

| | | | Shell_weight < 0.09 : Young (24/0)

| | | | Shell_weight >= 0.09

| | | | | Rings < 10.5 : Young (15/0)

| | | | | Rings >= 10.5 : Adult (5/0)

| | | Sex = I

| | | | Rings < 10.5 : Young (409/0)

| | | | Rings >= 10.5 : Adult (8/0)

| Shell_weight >= 0.11

| | Shucked_weight < 0.23

| | | Height < 0.12

| | | | Rings < 10.5 : Young (258/0)

| | | | Rings >= 10.5 : Adult (49/0)

| | | Height >= 0.12

| | | | Viscera_weight < 0.09

| | | | | Shucked_weight < 0.2

| | | | | | Sex = M : Adult (6/0)

| | | | | | Sex = F

| | | | | | | Diameter < 0.35

| | | | | | | | Height < 0.13 : Young (1/0)

| | | | | | | | Height >= 0.13 : Adult (2/0)

| | | | | | | Diameter >= 0.35 : Young (2/0)

| | | | | | Sex = I

| | | | | | | Rings < 10 : Young (4/0)

| | | | | | | Rings >= 10 : Adult (7/0)

| | | | | Shucked_weight >= 0.2 : Young (2/0)

| | | | Viscera_weight >= 0.09

| | | | | Rings < 10.5 : Young (92/0)

| | | | | Rings >= 10.5 : Adult (39/0)

| | Shucked_weight >= 0.23

| | | Whole_weight < 0.58 : Young (122/0)

| | | Whole_weight >= 0.58

| | | | Rings < 10.5 : Young (131/0)

| | | | Rings >= 10.5 : Adult (13/0)

Shell_weight >= 0.18

| Rings < 10.5 : Young (1326/0)

| Rings >= 10.5 : Adult (1304/0)

Size of the tree : 47

Time taken to build model: 0.05 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.05 seconds

=== Summary ===

Correctly Classified Instances	4177	100 %
Incorrectly Classified Instances	0	0 %
Kappa statistic	1	

Mean absolute error 0

Root mean squared error 0

Relative absolute error 0 %

Root relative squared error 0 %

Total Number of Instances 4177

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Young
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Adult
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

=== Confusion Matrix ===

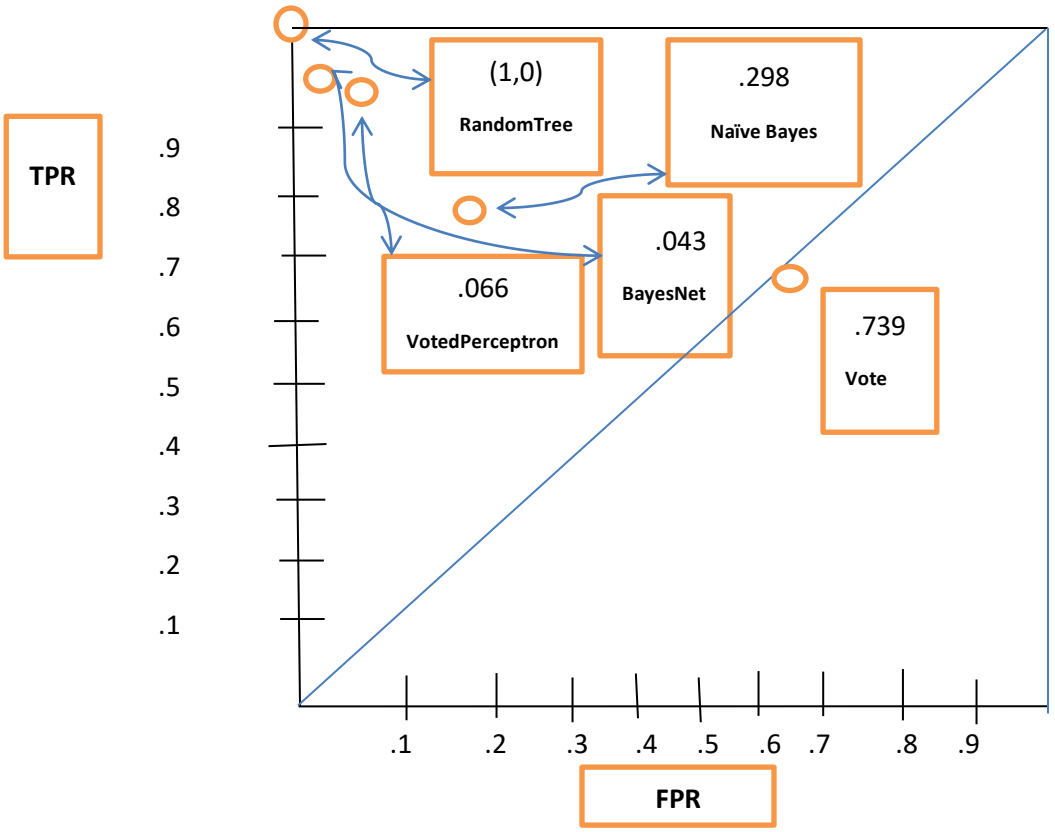
a	b	<-- classified as
2730	0	a = Young
0	1447	b = Adult

RESULT:

Items	FPR	TPR
Naïve Bayes	.195	.774
VotedPerceptron	.052	.960
BayesNet	.030	.969
Vote	.654	.654
RandomTree	.000	1.000

ROC GRAPH:

(0,1)



Analysis:

According to ROC GRAPH we see RandomTree classifier give best result from best case scenario.

Form the confusion matrix and ROC curve we can measure the performance of a classifier. Among the five classifiers RandomTree is the best classifier for this case. Moreover, RandomTree classifer is plotted to the best point of ROC curve. The false positive rate is zero for RandomTree Classifier. BayesNet and VotedPerceptron Classifier almost have the same result but in BayesNet, the percentage of correctly classified result is slightly better than VotedPerceptron and also the false positive rate is lower.

After analyzing all 5 different classifiers using the given dataset, it can be seen that RandomTree gives the best result among all other classifiers. It has the maximum number of correctly classified instances. Among 4177 instance, it can classify 4177 instances correctly which is the maximum. So, correctly classified instances is 4177 which is 100% .
Now 2nd best classifier is BayesNet which can correctly classify 4047 instances and incorrectly classify 130 instances out of 4177 instances. Its correctness is 96.8877% and incorrectness is 3.1123 % .
Then 3nd best classifier is VotedPerceptron which can correctly classify 4010 instances and incorrectly classify 167 instances out of 4177 instances. Its correctness is 96.0019% and incorrectness is 3.9981 %.

After that Naïve Bayes has 77.3761% accuracy. And Finally comes the Vote which has the lowest accuracy among these 5 which is 65.3579% .

So if we sort the classifiers according to their performances on the given dataset then the list becomes :

RandomTree > BayesNet > VotedPerceptron > Naïve Bayes >Vote.