Conducting an Analysis of a Qualitative Dataset Using the Waikato Environment for Knowledge Analysis (WEKA)

by Robert A Sottolare

ARL-TR-7182

February 2015

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory

Aberdeen Proving Ground, MD 21005-5425

ARL-TR-7182**February 2015**

Conducting an Analysis of a Qualitative Dataset Using the Waikato Environment for Knowledge Analysis (WEKA)

Robert A Sottolare
Human Research and Engineering Directorate, ARL

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
<p>Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>					
1. REPORT DATE (DD-MM-YYYY) February 2015		2. REPORT TYPE Final		3. DATES COVERED (From - To) November 2014–December 2014	
4. TITLE AND SUBTITLE Conducting an Analysis of a Qualitative Dataset Using the Waikato Environment for Knowledge Analysis (WEKA)				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Robert A Sottolare				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) US Army Research Laboratory ATTN: RDRL-HRT-T Orlando, FL 32826				8. PERFORMING ORGANIZATION REPORT NUMBER ARL-TR-7182	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT The purpose of this technical report is to provide an exemplar for conducting an analysis of a qualitative dataset using machine learning techniques. Qualitative data are measured or expressed as a natural language description (e.g., category or attribute) rather than numbers as in quantitative datasets. It is often difficult to evaluate the relationship between qualitative variables of interest and outcomes (dependent variables), but machine learning techniques offer simplified methods to classify these outcomes. WEKA, the Waikato Environment for Knowledge Analysis, is a popular set of machine learning algorithms developed at the University of Waikato in New Zealand, which can be used to analyze both qualitative and quantitative data. To illustrate the use of WEKA on a qualitative dataset, we selected a known set of primate species with the desire to classify them into 1 of 3 classes (prosimians, monkeys, and apes) based on 7 qualitative attributes. When an existing dataset is used with known relationships, it allows us to evaluate a large number of WEKA algorithms in a relatively short time and validate their accuracy with the goal of identifying best practices for analyzing qualitative data.					
15. SUBJECT TERMS machine learning techniques, qualitative datasets, machine learning classifiers, validation, data mining					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 62	19a. NAME OF RESPONSIBLE PERSON Robert A Sottolare
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (Include area code) (407) 208-3007

Contents

List of Figures	iv
List of Tables	iv
1. Introduction	1
2. Description of the Analysis Task	1
2.1 Performance Measure, P	3
2.2 Training Experience, E	3
2.3 Target Function, H	3
2.4 Target Function Representation, $h(x)$	3
3. Results and Discussion	4
4. Conclusions	6
5. References	8
Appendix A. Training Examples	9
Appendix B. Test Data	17
Appendix C. Classifier Performance Using Training Data and Test Data	21
Distribution List	55

List of Figures

Fig. 1	Primate classes: prosimians (left), monkeys (center), and apes (right)	1
Fig. 2	Primate classifier	6

List of Tables

Table 1	Classifier accuracy and speed	4
Table 2	Ranking of classifier models.....	5

1. Introduction

The purpose of this technical report is to provide an exemplar for conducting an analysis of a qualitative dataset using machine learning techniques. Qualitative data are measured or expressed as a natural language description (e.g., category or attribute) rather than numbers as in quantitative datasets. It is often difficult to evaluate the relationship between qualitative variables of interest and outcomes (dependent variables), but machine learning techniques offer simplified methods to classify these outcomes. WEKA, the Waikato Environment for Knowledge Analysis, is a popular set of machine learning algorithms developed at the University of Waikato in New Zealand (Witten and Frank 2005), which can be used to analyze both qualitative and quantitative data. To illustrate the use of WEKA on a qualitative dataset, we selected a known set of primate species with the desire to classify them into 1 of 3 classes (prosimians, monkeys, and apes) based on 7 qualitative attributes. When an existing dataset is used with known relationships, it allows us to evaluate a large number of WEKA algorithms in a relatively short time and validate their accuracy with the goal of identifying best practices for analyzing qualitative data.

2. Description of the Analysis Task

The task, T, was to accurately classify primates into 1 of 3 classes (prosimians, monkeys, or apes) based on 7 attributes identified through a review of zoological references (Goodman et al. 1990; McKenna and Bell 1997; Groves 2001, 2005) that indicated the major attributes and differences among the various species of the order primates. A sample of each class of the primate order is shown in Fig. 1.

Fig. 1 Primate classes: prosimians (left), monkeys (center), and apes (right)

The context of the task follows: A robot was sent to the zoo to visit the primate area and collect data on the attributes of different types of primates. WEKA 3.6 (Witten and Frank 2005) was used to analyze the 7 attributes selected and to classify each primate using several candidate machine-learning algorithms. The performance of each algorithm was then compared to determine the best classifiers for this qualitative dataset. Primates at the zoo are described by the following 7 attributes:

1. Tail attribute: 2 possible values (tail/no tail) to indicate the presence or absence of a tail.
2. Chest size attribute: 3 possible values (broad, deep, or equal) to indicate the relationship between the breadth and depth of the chest. Broad indicates the chest is broader than it is deep, deep indicates the chest is deeper than it is broad, and equal indicates that the chest is roughly the same breadth as depth.
3. Arm-leg ratio attribute: 3 possible values (shorter, longer, or equal) to indicate the relationship between the length of the arms and the length of the legs. Shorter indicates that the arms are shorter than the legs, longer indicates that the arms are longer than the legs, and equal indicates that the arms and legs are approximately the same length.
4. Size attribute: 5 possible values (very small, small, medium, large, or very large) to indicate the relative size of primates based on ranges of weight and body length.
5. Sleep habits attribute: 2 possible values (nocturnal or diurnal) to indicate the primary sleeping habits of various species of primates. Nocturnal indicates that the primate is primarily awake at night and asleep during the day; diurnal animals generally sleep at night and are awake during the day.
6. Nose attribute: 2 possible values (wet or dry) to indicate the nose and snout type. Wet indicates the primate has a wet nose and a snout; dry indicates the primate has a dry nose and no snout.
7. Nails attribute: 3 possible values (nails, claws, or both) to indicate whether the primate possesses nails on their digits, claws on their digits, or some combination of both nails and claws.

The problem for classifying each primate at the zoo may be determined by the following given information:

- Instances, X: each of the primates at the local zoo, described by the previously mentioned attributes.
- Hypotheses, H: each hypothesis is described by a conjunction of constraints on the attributes (tail, chest size, arm-leg ratio, size, sleep habits, nose, and nails). The constraints may be “?” (any value is acceptable), “0” (no value is acceptable), or a specific value as noted in the possible values of the 7 attributes listed.

- Target Concept, c : PrimateType: $X \rightarrow (\text{prosimian}, \text{monkey}, \text{ape})$.
- Training Examples, D : 318 training examples of 3 classes of primates were provided and are listed in Appendix A. The training examples are approximately 75% of the total sample of 423 primates.

The goal of this analysis is to determine a hypothesis, h , in H such that $h(x) = c(x)$ for all x in X . A more detailed description of the primate classifier task in terms of performance, training experience, target function, and target function representation follows.

2.1 Performance Measure, P

The performance measure, P , for each of the classifiers tested was the percentage of primates accurately identified as prosimians, monkeys, or apes during the test runs. The speed with which each classifier generated the models was also evaluated as a performance factor which might not be critical to this small dataset but will be in larger, more complex datasets (e.g., datasets with large numbers of attributes or multiple outcomes). Finally, the number of attributes required to classify a primate was evaluated as an efficiency factor where lower numbers of attributes to gain the same accuracy are desirable.

2.2 Training Experience, E

The examples used to train the robot to classify and identify prosimians, monkeys, and apes are listed in Appendix A. The training examples cover all the possible values for the 7 attributes selected to classify the primate species into the 3 possible outcomes (prosimians, monkeys, and apes). Approximately 24% of all primate species are prosimians. Approximately 67% of all primate species are monkeys, and approximately 9% are apes. These proportions are reflected in the training examples.

2.3 Target Function, H

For the target function, H , the valuation of each hypothesis was chosen as the indicator of the performance of the algorithms tested. If some instances, x , satisfied all the constraints of the hypothesis, h , for that outcome (i.e., monkeys), then h , classifies x as a positive example ($h(x) = 1$).

2.4 Target Function Representation, $h(x)$

The target function representation is in the form of rules that define the combination of attribute values needed to define each specific outcome. Since we have 3 possible outcomes (prosimians, monkeys, and apes), the expectation is that we will have a minimum of 3 rules. The rules may or may not require all attributes. For the 7 attributes evaluated, an example of a rule that classifies x as a positive example of ape may be

$$h(x)_{ape} = (\text{no tail}, ?, ?, ?, ?, \text{nails}) = 1 . \quad (1)$$

This hypothesis states that any primate that lacks a tail and has nails will be classified as an ape in 100% of the instances presented to the classifier model. It also indicates that no other attributes besides the tail and nails attributes influence this hypothesis.

3. Results and Discussion

Fifteen classifiers were trained using the training data in Appendix A and validated with the testing data in Appendix B. The results of each of the classifier runs are presented in Appendix C. As noted earlier, the measures to evaluate the performance of each algorithm include classifier accuracy, classifier speed, and classifier efficiency.

The classifier accuracy rates are shown in Table 1. Eleven of the 15 classifiers evaluated were 100% accurate in classifying primates for both the training data and the test data. The number of accurate classifiers indicates that the data set (attributes and values) describing the primate group are very representative of the characteristics of and differences between the 3 classes within the primate group.

Table 1 Classifier accuracy and speed

Classifier	Training Data		Test Data		Time to Build Data (s)
	% Correct	% Incorrect	% Correct	% Incorrect	
Naïve bayes classifier	100.000	0.000	100.000	0.000	<0.01
Random tree	100.000	0.000	100.000	0.000	0.01
LWL classifier	100.000	0.000	100.000	0.000	0.01
Non-nested generalized exemplars	100.000	0.000	100.000	0.000	0.02
PART decision list	100.000	0.000	100.000	0.000	0.02
Prism classifier	100.000	0.000	100.000	0.000	0.02
Bayes network classifier	100.000	0.000	100.000	0.000	0.02
RiDoR learner	100.000	0.000	100.000	0.000	0.04
JRip	100.000	0.000	100.000	0.000	0.07
Logistic function	100.000	0.000	100.000	0.000	0.26
Multilayer perceptron function	100.000	0.000	100.000	0.000	10.12
OneR classifier	91.195	8.805	91.429	8.571	0.01
Decision stump	91.195	8.805	91.429	8.571	0.01
Conjunctive rule	91.195	8.805	91.429	8.571	0.03
ZeroR classifier	67.296	32.704	67.619	32.381	0.00

Generally, the errors in the 4 less accurate classifiers were due to an insufficient number of rules to cover the all the possible outcomes. For example, the OneR Classifier generated 2 rules related to the nose attribute: 1) wet -> Prosimian and 2) dry -> Monkey. The outcome, “Apes”, is not covered by these rules. For the training data set, 290 of 318 instances were correct; over 91%. This might be considered fairly accurate, but the accuracy of the OneR Classifier for this data set is dependent on the number of apes in the sample. Since apes are not covered in the OneR Classifier model, the higher the number of instances of apes in the data set, the higher the error rate.

In reviewing the influence of each attribute on the 3 hypotheses, the attributes for tail presence, nose type, arm-leg ratio, and sleep habits appeared most often with nose type and tail presence appearing in almost every model. The attributes for nail type, size, and chest size did not seem to have as much of an influence on classification.

Time to build the model was compared (Table 1) with the thought that time and accuracy would be discerning factors in selecting the best (most efficient) model. However, the time to build a classifier model from such a small dataset was not an indicator of a well-performing model. Instead, a more representative factor in evaluating a model was the number of attributes needed to predict a class. The fewer attributes needed to classify indicated a more efficient model requiring fewer rules and thereby less processing cycles. The classifiers were ranked based on accuracy of predictions and number of attributes as shown in Table 2.

Table 2 Ranking of classifier models

Classifier	Accuracy (%)	No. of Attributes
JRip	100.000	2
Part decision list	100.000	2
RiDoR learner	100.000	3
Random tree	100.000	3
Prism classifier	100.000	4
Nonnested generalized exemplars	100.000	7
Bayes network classifier	100.000	7
Naïve bayes classifier	100.000	7
Logistic function	100.000	7
Multilayer perceptron function	100.000	7
LWL classifier	100.000	7
Conjunctive rule	91.429	1
OneR classifier	91.429	1
Decision stump	91.429	1
ZeroR classifier	67.619	0

4. Conclusions

Based on the rules/weights generated by the 15 classifiers evaluated, the following 3 most generalized hypotheses are validated for $h(x) = (\text{tail, chest size, arm-leg ratio, size, sleep habits, nose, and nails})$:

- $h(x)_{\text{prosimian}} = (?, ?, ?, ?, ?, \text{wet}, ?) = 1$
- $h(x)_{\text{ape}} = (\text{no tail}, ?, ?, ?, ?, ?, ?) = 1$
- $h(x)_{\text{monkey}} = (\text{tail}, ?, ?, ?, ?, \text{dry}, ?) = 1$

Based on the previously stated validated hypotheses, 3 rules were developed:

- IF Tails = “no tail” THEN Outcome = Ape;
- IF Nose = “wet” THEN Outcome = Prosimian;
- IF Nose = “dry” AND Tails = “tail” THEN Outcome = Monkey

In other words, these may be combined into one compound rule where all primates are monkeys unless they have no tail (in which case they are an ape) or they have a wet nose (in which case they are a prosimian). This simplified model is shown in Fig. 2. There are no primates that have no tail and a wet nose.

		Nose Attribute	
		Wet	Dry
Tail Attribute	No Tail		Apes
	Tail	Prosimians	Monkeys

Fig. 2 Primate classifier

Only 2 (nose type and tail presence) of the 7 attributes evaluated were actually required to classify a primate. Five (size, arm-leg ratio, chest size, sleeping habits, and nail types) of the 7 attributes evaluated are unnecessary to classify primates. For classification of primates, the most accurate (percent of classifications correct) and efficient models (fewest number of attributes to define an outcome) were generated using the Java Repeated Incremental Pruning algorithm and Partial Decision List classifiers.

This technical report has illustrated methods for analyzing qualitative datasets and also shows that different methods (e.g., machine-learning algorithms) can vary in accuracy, speed, and efficiency even with the identical datasets. While scalability was not a major point of emphasis in this report, it is a factor in analyzing larger qualitative datasets where near real-time response is required (e.g., training systems).

5. References

- Goodman M, Tagle DA, Fitch DH, Bailey W, Czelusniak J, Koop BF, Benson PJ, Slightom L. Primate evolution at the DNA level and a classification of hominoids. *Journal of Molecular Evolution*. 1990;30:260–266.
- Groves C. *Primate taxonomy*. Washington (DC): Smithsonian Institution Press; 2001. ISBN 1-56098-872-X.
- Groves C. In Wilson DE, Reeder DM. ed. *Mammal Species of the World*, 3rd ed. Baltimore (MD): Johns Hopkins University Press; 2005. 111–184. ISBN 0-801-88221-4.
- McKenna MC, Bell SK. *Classification of mammals above the species level*. New York (NY): Columbia University Press; 1997. p. 329.
- Witten IH, Frank E. *Data mining: practical machine learning tools and techniques*. 2nd. ed. San Francisco (CA): Morgan Kaufmann; 2005.

Appendix A. Training Examples

This appendix appears in its original form, without editorial change.

Example	Tail	Chest Size	Arm-Leg Ratio	Size	Sleep Habits	Nose	Nails	Outcome
56	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
57	tail	deep	shorter	small	nocturnal	wet	claws and nails	Prosimian
58	tail	deep	shorter	small	nocturnal	wet	claws and nails	Prosimian
59	tail	deep	shorter	small	nocturnal	wet	claws and nails	Prosimian
60	tail	deep	shorter	small	nocturnal	wet	claws and nails	Prosimian
61	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
62	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
63	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
64	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
65	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
66	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
67	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
68	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
69	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
70	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
71	tail	equal	shorter	very small	nocturnal	wet	claws and nails	Prosimian
72	tail	equal	shorter	very small	nocturnal	wet	claws and nails	Prosimian
73	tail	equal	shorter	very small	nocturnal	wet	claws and nails	Prosimian
74	tail	equal	shorter	very small	nocturnal	wet	claws and nails	Prosimian
75	tail	equal	shorter	very small	nocturnal	wet	claws and nails	Prosimian
76	tail	equal	shorter	very small	nocturnal	wet	claws and nails	Prosimian
77	tail	deep	shorter	small	diurnal	dry	claws	Monkey
78	tail	deep	shorter	small	diurnal	dry	claws	Monkey
79	tail	deep	shorter	small	diurnal	dry	claws	Monkey
80	tail	deep	shorter	small	diurnal	dry	claws	Monkey
81	tail	deep	shorter	small	diurnal	dry	claws	Monkey
82	tail	deep	shorter	small	diurnal	dry	claws	Monkey
83	tail	deep	shorter	small	diurnal	dry	claws	Monkey
84	tail	deep	shorter	small	diurnal	dry	claws	Monkey
85	tail	deep	shorter	small	diurnal	dry	claws	Monkey
86	tail	deep	shorter	small	diurnal	dry	claws	Monkey
87	tail	deep	shorter	small	diurnal	dry	claws	Monkey
88	tail	deep	shorter	small	diurnal	dry	claws	Monkey
89	tail	deep	shorter	small	diurnal	dry	claws	Monkey
90	tail	deep	shorter	small	diurnal	dry	claws	Monkey
91	tail	deep	shorter	small	diurnal	dry	claws	Monkey
92	tail	deep	shorter	small	diurnal	dry	claws	Monkey
93	tail	deep	shorter	medium	diurnal	dry	claws	Monkey
94	tail	deep	shorter	medium	diurnal	dry	claws	Monkey
95	tail	deep	shorter	medium	diurnal	dry	claws	Monkey
96	tail	deep	shorter	medium	diurnal	dry	claws	Monkey
97	tail	deep	shorter	medium	diurnal	dry	claws	Monkey
98	tail	deep	shorter	medium	diurnal	dry	claws	Monkey
99	tail	deep	shorter	medium	diurnal	dry	claws	Monkey
100	tail	deep	shorter	medium	diurnal	dry	claws	Monkey
101	tail	deep	shorter	medium	diurnal	dry	claws	Monkey
102	tail	deep	shorter	medium	diurnal	dry	claws	Monkey
103	tail	deep	shorter	medium	diurnal	dry	claws	Monkey
104	tail	deep	shorter	medium	diurnal	dry	claws	Monkey
105	tail	deep	shorter	medium	diurnal	dry	claws	Monkey
106	tail	deep	shorter	medium	diurnal	dry	claws	Monkey
107	tail	deep	shorter	medium	diurnal	dry	claws	Monkey
108	tail	deep	shorter	medium	diurnal	dry	claws	Monkey
109	tail	deep	equal	small	diurnal	dry	nails	Monkey
110	tail	deep	equal	small	diurnal	dry	nails	Monkey
111	tail	deep	equal	small	diurnal	dry	nails	Monkey
112	tail	deep	equal	small	diurnal	dry	nails	Monkey
113	tail	deep	equal	small	diurnal	dry	nails	Monkey
114	tail	deep	equal	small	diurnal	dry	nails	Monkey
115	tail	deep	equal	medium	diurnal	dry	nails	Monkey
116	tail	deep	equal	medium	diurnal	dry	nails	Monkey
117	tail	deep	equal	medium	diurnal	dry	nails	Monkey
118	tail	deep	equal	medium	diurnal	dry	nails	Monkey
119	tail	deep	shorter	medium	nocturnal	dry	nails	Monkey
120	tail	deep	shorter	medium	nocturnal	dry	nails	Monkey

Example	Tail	Chest Size	Arm-Leg Ratio	Size	Sleep Habits	Nose	Nails	Outcome
316	no tail	broad	longer	large	diurnal	dry	nails	Ape
317	no tail	broad	longer	large	diurnal	dry	nails	Ape
318	no tail	broad	longer	large	diurnal	dry	nails	Ape

INTENTIONALLY LEFT BLANK.

Appendix B. Test Data

This appendix appears in its original form, without editorial change.

Example	Tails	Chest Size	Arm-Leg Ratio	Size	Sleep Habits	Nose	Nails	Outcome
1	tail	equal	shorter	very small	nocturnal	wet	claws	Prosimian
2	tail	equal	shorter	very small	nocturnal	wet	claws	Prosimian
3	tail	equal	shorter	very small	nocturnal	wet	claws	Prosimian
4	tail	equal	shorter	very small	nocturnal	wet	claws	Prosimian
5	tail	equal	shorter	very small	nocturnal	wet	claws	Prosimian
6	tail	equal	shorter	very small	nocturnal	wet	claws	Prosimian
7	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
8	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
9	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
10	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
11	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
12	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
13	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
14	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
15	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
16	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
17	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
18	tail	equal	shorter	small	diurnal	wet	claws	Prosimian
19	tail	deep	shorter	small	nocturnal	wet	claws and nails	Prosimian
20	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
21	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
22	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
23	tail	equal	shorter	small	nocturnal	wet	claws and nails	Prosimian
24	tail	equal	shorter	very small	nocturnal	wet	claws and nails	Prosimian
25	tail	equal	shorter	very small	nocturnal	wet	claws and nails	Prosimian
26	tail	deep	shorter	small	diurnal	dry	claws	Monkey
27	tail	deep	shorter	small	diurnal	dry	claws	Monkey
28	tail	deep	shorter	small	diurnal	dry	claws	Monkey
29	tail	deep	shorter	small	diurnal	dry	claws	Monkey
30	tail	deep	shorter	small	diurnal	dry	claws	Monkey
31	tail	deep	shorter	medium	diurnal	dry	claws	Monkey
32	tail	deep	shorter	medium	diurnal	dry	claws	Monkey
33	tail	deep	shorter	medium	diurnal	dry	claws	Monkey
34	tail	deep	shorter	medium	diurnal	dry	claws	Monkey
35	tail	deep	shorter	medium	diurnal	dry	claws	Monkey
36	tail	deep	shorter	small	diurnal	dry	claws	Monkey
37	tail	deep	equal	small	diurnal	dry	nails	Monkey
38	tail	deep	equal	small	diurnal	dry	nails	Monkey
39	tail	deep	equal	medium	diurnal	dry	nails	Monkey
40	tail	deep	shorter	medium	nocturnal	dry	nails	Monkey
41	tail	deep	shorter	medium	nocturnal	dry	nails	Monkey
42	tail	deep	equal	medium	diurnal	dry	nails	Monkey
43	tail	deep	equal	medium	diurnal	dry	nails	Monkey
44	tail	deep	equal	medium	diurnal	dry	nails	Monkey
45	tail	deep	equal	medium	diurnal	dry	nails	Monkey
46	tail	deep	equal	medium	diurnal	dry	nails	Monkey
47	tail	deep	equal	medium	diurnal	dry	nails	Monkey
48	tail	deep	equal	medium	diurnal	dry	nails	Monkey
49	tail	deep	equal	medium	diurnal	dry	nails	Monkey
50	tail	deep	equal	medium	diurnal	dry	nails	Monkey

Example	Tails	Chest Size	Arm-Leg Ratio	Size	Sleep Habits	Nose	Nails	Outcome
51	tail	deep	equal	medium	diurnal	dry	nails	Monkey
52	tail	deep	equal	large	diurnal	dry	nails	Monkey
53	tail	deep	equal	large	diurnal	dry	nails	Monkey
54	tail	deep	equal	large	diurnal	dry	nails	Monkey
55	tail	deep	equal	large	diurnal	dry	nails	Monkey
56	tail	deep	equal	medium	diurnal	dry	nails	Monkey
57	tail	deep	equal	medium	diurnal	dry	nails	Monkey
58	tail	deep	equal	medium	diurnal	dry	nails	Monkey
59	tail	deep	equal	medium	diurnal	dry	nails	Monkey
60	tail	deep	equal	medium	diurnal	dry	nails	Monkey
61	tail	deep	equal	medium	diurnal	dry	nails	Monkey
62	tail	deep	equal	medium	diurnal	dry	nails	Monkey
63	tail	deep	equal	medium	diurnal	dry	nails	Monkey
64	tail	deep	equal	medium	diurnal	dry	nails	Monkey
65	tail	deep	equal	medium	diurnal	dry	nails	Monkey
66	tail	deep	equal	medium	diurnal	dry	nails	Monkey
67	tail	deep	equal	medium	diurnal	dry	nails	Monkey
68	tail	deep	equal	medium	diurnal	dry	nails	Monkey
69	tail	deep	equal	medium	diurnal	dry	nails	Monkey
70	tail	deep	equal	medium	diurnal	dry	nails	Monkey
71	tail	deep	equal	medium	diurnal	dry	nails	Monkey
72	tail	deep	equal	medium	diurnal	dry	nails	Monkey
73	tail	deep	equal	medium	diurnal	dry	nails	Monkey
74	tail	deep	equal	medium	diurnal	dry	nails	Monkey
75	tail	deep	equal	medium	diurnal	dry	nails	Monkey
76	tail	broad	equal	large	diurnal	dry	nails	Monkey
77	tail	deep	equal	medium	diurnal	dry	nails	Monkey
78	tail	deep	equal	medium	diurnal	dry	nails	Monkey
79	tail	broad	equal	large	diurnal	dry	nails	Monkey
80	tail	broad	equal	large	diurnal	dry	nails	Monkey
81	tail	broad	equal	large	diurnal	dry	nails	Monkey
82	tail	broad	equal	large	diurnal	dry	nails	Monkey
83	tail	broad	equal	large	diurnal	dry	nails	Monkey
84	tail	broad	equal	large	diurnal	dry	nails	Monkey
85	tail	broad	equal	large	diurnal	dry	nails	Monkey
86	tail	deep	equal	medium	diurnal	dry	nails	Monkey
87	tail	deep	equal	medium	diurnal	dry	nails	Monkey
88	tail	deep	equal	medium	diurnal	dry	nails	Monkey
89	tail	deep	equal	medium	diurnal	dry	nails	Monkey
90	tail	deep	equal	medium	diurnal	dry	nails	Monkey
91	tail	deep	equal	medium	diurnal	dry	nails	Monkey
92	tail	deep	equal	medium	diurnal	dry	nails	Monkey
93	tail	deep	equal	medium	diurnal	dry	nails	Monkey
94	tail	deep	equal	medium	diurnal	dry	nails	Monkey
95	tail	deep	equal	medium	diurnal	dry	nails	Monkey
96	tail	deep	equal	medium	diurnal	dry	nails	Monkey
97	no tail	broad	longer	medium	diurnal	dry	nails	Ape
98	no tail	broad	longer	medium	diurnal	dry	nails	Ape
99	no tail	broad	longer	medium	diurnal	dry	nails	Ape
100	no tail	broad	longer	medium	diurnal	dry	nails	Ape
101	no tail	broad	longer	medium	diurnal	dry	nails	Ape
102	no tail	broad	longer	medium	diurnal	dry	nails	Ape
103	no tail	broad	longer	very large	diurnal	dry	nails	Ape
104	no tail	broad	longer	very large	diurnal	dry	nails	Ape
105	no tail	broad	longer	large	diurnal	dry	nails	Ape

INTENTIONALLY LEFT BLANK.

Appendix C. Classifier Performance Using Training Data and Test Data

Conjunctive Rule

SYNOPSIS: This class implements a single conjunctive rule learner that can predict for numeric and nominal class labels.

A rule consists of antecedents "AND"ed together and the consequent (class value) for the classification/regression. In this case, the consequent is the distribution of the available classes (or mean for a numeric value) in the dataset. If the test instance is not covered by this rule, then it's predicted using the default class distributions/value of the data not covered by the rule in the training data. This learner selects an antecedent by computing the Information Gain of each antecedent and prunes the generated rule using Reduced Error Pruning (REP) or simple pre-pruning based on the number of antecedents.

For classification, the Information of one antecedent is the weighted average of the entropies of both the data covered and not covered by the rule.

For regression, the Information is the weighted average of the mean-squared errors of both the data covered and not covered by the rule.

In pruning, weighted average of the accuracy rates on the pruning data is used for classification while the weighted average of the mean-squared errors on the pruning data is used for regression.

=== Run information ===

```
Scheme:      weka.classifiers.rules.ConjunctiveRule -N 3 -M 2.0 -P -1 -S 1
Relation:    training data
Instances:   318
Attributes:  8
              Tails
              Chest Size
              Arm-Leg Ratio
              Size
              Sleep Habits
              Nose
              Nails
              Outcome
Test mode:    evaluate on training data
```

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

Single conjunctive rule learner:

(Nose = wet) => Outcome = Prosimian

Class distributions:

Covered by the rule:

Prosimian	Monkey	Ape
1	0	0

Not covered by the rule:

Prosimian	Monkey	Ape
0	0.888199	0.111801

Time taken to build model: 0.03 seconds

=== Evaluation on training set ===

=== Summary ===

Correctly Classified Instances	290	91.195 %
Incorrectly Classified Instances	28	8.805 %
Kappa statistic	0.7956	
Mean absolute error	0.1023	
Root mean squared error	0.2279	
Relative absolute error	31.7043 %	
Root relative squared error	56.8283 %	
Total Number of Instances	318	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0.269	0.884	1	0.939	0.865	Monkey
0	0	0	0	0	0.631	Ape

=== Confusion Matrix ===

a	b	c	<-- classified as
76	0	0	a = Prosimian
0	214	0	b = Monkey
0	28	0	c = Ape

=== Re-evaluation on test set ===

User supplied test set
Relation: test data
Instances: 105
Attributes: 8

=== Summary ===

Correctly Classified Instances	96	91.4286 %
Incorrectly Classified Instances	9	8.5714 %
Kappa statistic	0.7998	
Mean absolute error	0.1012	
Root mean squared error	0.2252	
Total Number of Instances	105	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0.265	0.888	1	0.94	0.868	Monkey
0	0	0	0	0	0.63	Ape

=== Confusion Matrix ===

a	b	c	<-- classified as
25	0	0	a = Prosimian

0	71	0		b = Monkey
0	9	0		c = Ape

JRip

SYNOPSIS: This class implements a propositional rule learner, Repeated Incremental Pruning to Produce Error Reduction (RIPPER), which was proposed by Cohen¹ as an optimized version of IREP.

The algorithm is briefly described as follows:

Initialize RS = {}, and for each class from the less prevalent one to the more frequent one, DO:

1. Building stage:

Repeat 1.1 and 1.2 until the description length (DL) of the ruleset and examples is 64 bits greater than the smallest DL met so far, or there are no positive examples, or the error rate $\geq 50\%$.

1.1. Grow phase:

Grow one rule by greedily adding antecedents (or conditions) to the rule until the rule is perfect (i.e., 100% accurate). The procedure tries every possible value of each attribute and selects the condition with highest information gain: $p(\log(p/t) - \log(P/T))$.

1.2. Prune phase:

Incrementally prune each rule and allow the pruning of any final sequences of the antecedents; The pruning metric is $(p-n)/(p+n)$ -- but it's actually $2p/(p+n) - 1$, so in this implementation we simply use $p/(p+n)$ (actually $(p+1)/(p+n+2)$, thus if $p+n$ is 0, it's 0.5).

2. Optimization stage:

after generating the initial ruleset $\{R_i\}$, generate and prune two variants of each rule R_i from randomized data using procedure 1.1 and 1.2. But one variant is generated from an empty rule while the other is generated by greedily adding antecedents to the original rule. Moreover, the pruning metric used here is $(TP+TN)/(P+N)$. Then the smallest possible DL for each variant and the original rule is computed. The variant with the minimal DL is selected as the final representative of R_i in the ruleset. After all the rules in $\{R_i\}$ have been examined and if there are still residual positives, more rules are generated based on the residual positives using Building Stage again.

3. Delete the rules from the ruleset that would increase the DL of the whole ruleset if it were in it, and add resultant ruleset to RS.

ENDDO

=== Run information ===

```

Scheme:      weka.classifiers.rules.JRip -F 3 -N 2.0 -O 2 -S 1
Relation:    training data
Instances:    318
Attributes:   8
              Tails
              Chest Size
              Arm-Leg Ratio
              Size
              Sleep Habits
              Nose

```

¹ Cohen W. Fast effective rule induction. Published in Proceedings of the 12th International Conference on Machine Learning; 1995 Jul 9–12; Lake Tahoe, CA. San Francisco (CA): Morgan Kaufmann; c1995. p. 115–123.

```

        Nails
        Outcome
Test mode:  evaluate on training data

```

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

JRIP rules:
=====

```

(Tails = no_tail) => Outcome=Ape (28.0/0.0)
(Nose = wet)      => Outcome=Prosimian (76.0/0.0)
=> Outcome=Monkey (214.0/0.0)

```

Number of Rules : 3

Time taken to build model: 0.07 seconds

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances	318	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	318		

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0	1	1	1	1	Monkey
1	0	1	1	1	1	Ape

=== Confusion Matrix ===

a	b	c	<-- classified as
76	0	0	a = Prosimian
0	214	0	b = Monkey
0	0	28	c = Ape

=== Re-evaluation on test set ===

```

User supplied test set
Relation:      test data
Instances:    105
Attributes:    8

```

=== Summary ===

Correctly Classified Instances	105	100	%
Incorrectly Classified Instances	0	0	%

```

Kappa statistic          1
Mean absolute error      0
Root mean squared error  0
Total Number of Instances 105

```

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0	1	1	1	1	Monkey
1	0	1	1	1	1	Ape

=== Confusion Matrix ===

```

  a  b  c  <-- classified as
25  0  0 | a = Prosimian
 0 71  0 | b = Monkey
 0  0  9 | c = Ape

```

Non-Nested Generalized Exemplars

SYNOPSIS: Nearest-neighbor-like algorithm using non-nested generalized exemplars (which are hyper-rectangles that can be viewed as if-then rules). For more information, see Martin² and Roy.³

=== Run information ===

```

Scheme:      weka.classifiers.rules.NNge -G 5 -I 5
Relation:    training data
Instances:   318
Attributes:  8
              Tails
              Chest Size
              Arm-Leg Ratio
              Size
              Sleep Habits
              Nose
              Nails
              Outcome
Test mode:   evaluate on training data

```

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

NNGE classifier

Rules generated :

```

class Ape IF : Tails in {no_tail} ^ Chest Size in {broad} ^ Arm-Leg
Ratio in {longer} ^ Size in {medium,large,very_large} ^ Sleep Habits in
{diurnal} ^ Nose in {dry} ^ Nails in {nails}  (28)
class Monkey IF : Tails in {tail} ^ Chest Size in {deep,broad} ^ Arm-

```

²Martin B. Instance-based learning: Nearest neighbor with generalization. [dissertation]. [Hamilton (New Zealand)]: University of Waikato; 1995.

³Roy S. Nearest neighbor with generalization. [unpublished]. [Christchurch (New Zealand)]: University of Canterbury, 2002.


```

Leg Ratio in {shorter,equal} ^ Size in {small,medium,large} ^ Sleep Habits in
{nocturnal,diurnal} ^ Nose in {dry} ^ Nails in {claws,nails} (214)
class Prosimian IF : Tails in {tail} ^ Chest Size in {equal,deep} ^
Arm-Leg Ratio in {shorter} ^ Size in {very_small,small,medium} ^ Sleep Habits
in {nocturnal,diurnal} ^ Nose in {wet} ^ Nails in {claws,claws_and_nails}
(76)

```

Stat :

```

class Prosimian : 1 exemplar(s) including 1 Hyperrectangle(s) and 0
Single(s).
class Monkey : 1 exemplar(s) including 1 Hyperrectangle(s) and 0
Single(s).
class Ape : 1 exemplar(s) including 1 Hyperrectangle(s) and 0
Single(s).

```

Total : 3 exemplars(s) including 3 Hyperrectangle(s) and 0 Single(s).

```

Feature weights : [0.42992832030968897 0.932127623605122
0.857511606191957 0.6290251428980856 0.6703867660729255 0.7933628469756622
0.6511518907599008]

```

Time taken to build model: 0.02 seconds

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances	318	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	318		

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0	1	1	1	1	Monkey
1	0	1	1	1	1	Ape

=== Confusion Matrix ===

a	b	c	<-- classified as
76	0	0	a = Prosimian
0	214	0	b = Monkey
0	0	28	c = Ape

=== Re-evaluation on test set ===

```

User supplied test set
Relation:      test data
Instances:    105
Attributes:    8

```

=== Summary ===

Correctly Classified Instances	105	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0		
Root mean squared error	0		
Total Number of Instances	105		

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0	1	1	1	1	Monkey
1	0	1	1	1	1	Ape

=== Confusion Matrix ===

a	b	c	<-- classified as
25	0	0	a = Prosimian
0	71	0	b = Monkey
0	0	9	c = Ape

OneR Classifier

SYNOPSIS: Class for building and using a 1R classifier; in other words, uses the minimum-error attribute for prediction, discretizing numeric attributes. For more information, see Holte.⁴

=== Run information ===

Scheme: weka.classifiers.rules.OneR -B 6
Relation: training data
Instances: 318
Attributes: 8
Tails
Chest Size
Arm-Leg Ratio
Size
Sleep Habits
Nose
Nails
Outcome
Test mode: evaluate on training data

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

Nose:
wet -> Prosimian
dry -> Monkey
(290/318 instances correct)

Time taken to build model: 0.01 seconds

⁴ Holte RC. Very simple classification rules perform well on most commonly used datasets. Machine Learning. 1993;11:63–91.

```

=== Evaluation on training set ===
=== Summary ===

```

```

Correctly Classified Instances      290          91.195 %
Incorrectly Classified Instances    28           8.805 %
Kappa statistic                    0.7956
Mean absolute error                 0.0587
Root mean squared error             0.2423
Relative absolute error             18.1929 %

Root relative squared error        60.4273 %
Total Number of Instances          318

```

```

=== Detailed Accuracy By Class ===

```

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0.269	0.884	1	0.939	0.865	Monkey
0	0	0	0	0	0.5	Ape

```

=== Confusion Matrix ===

```

```

  a   b   c   <-- classified as
76   0   0   |   a = Prosimian
 0 214   0   |   b = Monkey
 0  28   0   |   c = Ape

```

```

=== Re-evaluation on test set ===

```

```

User supplied test set
Relation:      test data
Instances:     105
Attributes:    8

```

```

=== Summary ===

```

```

Correctly Classified Instances      96          91.4286 %
Incorrectly Classified Instances     9           8.5714 %
Kappa statistic                    0.7998
Mean absolute error                 0.0571
Root mean squared error             0.239
Total Number of Instances          105

```

```

=== Detailed Accuracy By Class ===

```

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0.265	0.888	1	0.94	0.868	Monkey
0	0	0	0	0	0.5	Ape

```

=== Confusion Matrix ===

```

```

  a   b   c   <-- classified as
25   0   0   |   a = Prosimian
 0  71   0   |   b = Monkey
 0   9   0   |   c = Ape

```

PART Decision List

SYNOPSIS: Class for generating a PART decision list. Uses separate-and-conquer. Builds a partial C4.5 decision tree in each iteration and makes the "best" leaf into a rule.

For more information, see: Frank.⁵

=== Run information ===

```
Scheme:      weka.classifiers.rules.PART -M 2 -C 0.25 -Q 1
Relation:    training data
Instances:   318
Attributes:  8
              Tails
              Chest Size
              Arm-Leg Ratio
              Size
              Sleep Habits
              Nose
              Nails
              Outcome
Test mode:    evaluate on training data
```

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

PART decision list

Nose = dry AND

Tails = tail: Monkey (214.0)

Tails = tail: Prosimian (76.0)

: Ape (28.0)

Number of Rules : 3

Time taken to build model: 0.02 seconds

=== Evaluation on training set ===

=== Summary ===

Correctly Classified Instances	318	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	318		

=== Detailed Accuracy By Class ===

⁵ Frank E, Witten IH. Generating accurate rule sets without global optimization. Published in 15th International Conference on Machine Learning; 1998 Jul 24–27; Madison, WI. San Mateo (CA): Morgan Kaufmann; c1998. p. 144–151.

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0	1	1	1	1	Monkey
1	0	1	1	1	1	Ape

=== Confusion Matrix ===

a	b	c	<-- classified as
76	0	0	a = Prosimian
0	214	0	b = Monkey
0	0	28	c = Ape

=== Re-evaluation on test set ===

User supplied test set
Relation: test data
Instances: 105
Attributes: 8

=== Summary ===

Correctly Classified Instances	105	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0		
Root mean squared error	0		
Total Number of Instances	105		

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0	1	1	1	1	Monkey
1	0	1	1	1	1	Ape

=== Confusion Matrix ===

a	b	c	<-- classified as
25	0	0	a = Prosimian
0	71	0	b = Monkey
0	0	9	c = Ape

Prism Classifier

SYNOPSIS

Class for building and using a PRISM rule set for classification. Can only deal with nominal attributes. Can't deal with missing values. Doesn't do any pruning.

For more information, see Cendrowska.⁶

=== Run information ===

Scheme: weka.classifiers.rules.Prism
Relation: training data
Instances: 318

⁶Cendrowska J. PRISM: An algorithm for inducing modular rules. International Journal of Man-Machine Studies. 1987;27(4):349-370.

```

Attributes:      8
                Tails
                Chest Size
                Arm-Leg Ratio
                Size
                Sleep Habits
                Nose
                Nails
                Outcome

```

Test mode: evaluate on training data

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

Prism rules

```

If Nose = wet then Prosimian
If Arm-Leg Ratio = equal then Monkey
If Chest Size = deep
    and Nose = dry then Monkey
If Tails = no_tail then Ape

```

Time taken to build model: 0.02 seconds

=== Evaluation on training set ===

=== Summary ===

Correctly Classified Instances	318	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	318		

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0	1	1	1	1	Monkey
1	0	1	1	1	1	Ape

=== Confusion Matrix ===

a	b	c	<-- classified as
76	0	0	a = Prosimian
0	214	0	b = Monkey
0	0	28	c = Ape

=== Re-evaluation on test set ===

```

User supplied test set
Relation:      test data
Instances:     105
Attributes:    8

```

```
=== Summary ===
```

Correctly Classified Instances	105	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0		
Root mean squared error	0		
Total Number of Instances	105		

```
=== Detailed Accuracy By Class ===
```

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0	1	1	1	1	Monkey
1	0	1	1	1	1	Ape

```
=== Confusion Matrix ===
```

a	b	c	<-- classified as
25	0	0	a = Prosimian
0	71	0	b = Monkey
0	0	9	c = Ape

Ripple-DOWN Rule learner (RIDOR)

SYNOPSIS: The implementation of a Ripple-DOWN Rule learner. It generates a default rule first and then the exceptions for the default rule with the least (weighted) error rate. Then it generates the "best" exceptions for each exception and iterates until pure. Thus it performs a tree-like expansion of exceptions. The exceptions are a set of rules that predict classes other than the default. IREP is used to generate the exceptions.

For more information about Ripple-Down Rules, see Gaines and Compton.⁷

```
=== Run information ===
```

```
Scheme:      weka.classifiers.rules.Ridor -F 3 -S 1 -N 2.0
Relation:    training data
Instances:   318
Attributes:  8
              Tails
              Chest Size
              Arm-Leg Ratio
              Size
              Sleep Habits
              Nose
              Nails
              Outcome
Test mode:   evaluate on training data
```

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

```
RIpple DOWn Rule Learner(Ridor) rules
-----
```

⁷Gaines BR, Compton P. Induction of ripple-down rules applied to modeling large databases. Joint Intelligent Inference Systems. 1995;5(3):211–228.

```

Outcome = Monkey (318.0/104.0)
      Except (Nose = wet) => Outcome = Prosimian (50.0/0.0) [26.0/0.0]
      Except (Chest Size = broad) and (Arm-Leg Ratio = longer) =>
Outcome = Ape (19.0/0.0) [9.0/0.0]

```

Total number of rules (incl. the default rule): 3

Time taken to build model: 0.04 seconds

=== Evaluation on training set ===

=== Summary ===

Correctly Classified Instances	318	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	318		

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0	1	1	1	1	Monkey
1	0	1	1	1	1	Ape

=== Confusion Matrix ===

a	b	c	<-- classified as
76	0	0	a = Prosimian
0	214	0	b = Monkey
0	0	28	c = Ape

=== Re-evaluation on test set ===

User supplied test set
Relation: test data
Instances: 105
Attributes: 8

=== Summary ===

Correctly Classified Instances	105	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0		
Root mean squared error	0		
Total Number of Instances	105		

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0	1	1	1	1	Monkey
1	0	1	1	1	1	Ape

=== Confusion Matrix ===

a	b	c	<-- classified as
25	0	0	a = Prosimian
0	71	0	b = Monkey
0	0	9	c = Ape

ZeroR Classifier

SYNOPSIS: Class for building and using a 0-R classifier. Predicts the mean (for a numeric class) or the mode (for a nominal class).

=== Run information ===

Scheme: weka.classifiers.rules.ZeroR
Relation: training data
Instances: 318
Attributes: 8
Tails
Chest Size
Arm-Leg Ratio
Size
Sleep Habits
Nose
Nails
Outcome
Test mode: evaluate on training data

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

ZeroR predicts class value: Monkey

Time taken to build model: 0 seconds

=== Evaluation on training set ===

=== Summary ===

Correctly Classified Instances	214	67.2956 %
Incorrectly Classified Instances	104	32.7044 %
Kappa statistic	0	
Mean absolute error	0.3227	
Root mean squared error	0.4009	
Relative absolute error	100	%
Root relative squared error	100	%
Total Number of Instances	318	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0	0	0	0	0	0.5	Prosimian
1	1	0.673	1	0.805	0.5	Monkey
0	0	0	0	0	0.5	Ape

=== Confusion Matrix ===

a	b	c	<-- classified as
0	76	0	a = Prosimian
0	214	0	b = Monkey
0	28	0	c = Ape

=== Re-evaluation on test set ===

User supplied test set
 Relation: test data
 Instances: 105
 Attributes: 8

=== Summary ===

Correctly Classified Instances	71	67.619 %
Incorrectly Classified Instances	34	32.381 %
Kappa statistic	0	
Mean absolute error	0.3215	
Root mean squared error	0.3995	
Total Number of Instances	105	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0	0	0	0	0	0.5	Prosimian
1	1	0.676	1	0.807	0.5	Monkey
0	0	0	0	0	0.5	Ape

=== Confusion Matrix ===

a	b	c	<-- classified as
0	25	0	a = Prosimian
0	71	0	b = Monkey
0	9	0	c = Ape

Decision Stump

SYNOPSIS: Class for building and using a decision stump. Usually used in conjunction with a boosting algorithm. Does regression (based on mean-squared error) or classification (based on entropy). Missing is treated as a separate value.

=== Run information ===

Scheme: weka.classifiers.trees.DecisionStump
 Relation: training data
 Instances: 318
 Attributes: 8
 Tails

```

        Chest Size
        Arm-Leg Ratio
        Size
        Sleep Habits
        Nose
        Nails
        Outcome
Test mode:    evaluate on training data

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

Decision Stump

Classifications

Nose = wet : Prosimian
Nose != wet : Monkey
Nose is missing : Monkey

Class distributions

Nose = wet
Prosimian    Monkey    Ape
1.0    0.0    0.0
Nose != wet
Prosimian    Monkey    Ape
0.0    0.8842975206611571    0.11570247933884298
Nose is missing
Prosimian    Monkey    Ape
0.2389937106918239    0.6729559748427673    0.0880503144654088

Time taken to build model: 0.01 seconds

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances      290                91.195 %
Incorrectly Classified Instances    28                8.805 %
Kappa statistic                    0.7956
Mean absolute error                 0.1038
Root mean squared error             0.2278
Relative absolute error             32.1758 %
Root relative squared error         56.824 %
Total Number of Instances          318

=== Detailed Accuracy By Class ===

TP Rate    FP Rate    Precision    Recall    F-Measure    ROC Area    Class
1          0          1          1          1          1          Prosimian
1          0.269      0.884      1          0.939      0.865      Monkey
0          0          0          0          0          0.631      Ape

=== Confusion Matrix ===

  a    b    c  <-- classified as
76    0    0 |    a = Prosimian

```

```

0 214 0 | b = Monkey
0 28 0 | c = Ape

```

=== Re-evaluation on test set ===

```

User supplied test set
Relation:      test data
Instances:    105
Attributes:    8

```

=== Summary ===

```

Correctly Classified Instances      96          91.4286 %
Incorrectly Classified Instances     9          8.5714 %
Kappa statistic                     0.7998
Mean absolute error                  0.1027
Root mean squared error              0.2252
Total Number of Instances           105

```

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0.265	0.888	1	0.94	0.868	Monkey
0	0	0	0	0	0.63	Ape

=== Confusion Matrix ===

```

a  b  c  <-- classified as
25  0  0 | a = Prosimian
0 71  0 | b = Monkey
0  9  0 | c = Ape

```

RandomTree

SYNOPSIS: Class for constructing a tree that considers K randomly chosen attributes at each node. RandomTree performs no pruning.

=== Run information ===

```

Scheme:      weka.classifiers.trees.RandomTree -K 1 -M 1.0 -S 1
Relation:    training data
Instances:   318
Attributes:  8
              Tails
              Chest Size
              Arm-Leg Ratio
              Size
              Sleep Habits
              Nose
              Nails
              Outcome
Test mode:   evaluate on training data

```

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

RandomTree

=====

```
Arm-Leg Ratio = shorter
|   Sleep Habits = nocturnal
|   |   Nose = wet : Prosimian (75/0)
|   |   Nose = dry : Monkey (6/0)
|   Sleep Habits = diurnal
|   |   Nose = wet : Prosimian (1/0)
|   |   Nose = dry : Monkey (32/0)
Arm-Leg Ratio = equal : Monkey (176/0)
Arm-Leg Ratio = longer : Ape (28/0)
```

Size of the tree : 10

Time taken to build model: 0.01 seconds

=== Evaluation on training set ===

=== Summary ===

Correctly Classified Instances	318	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	318		

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0	1	1	1	1	Monkey
1	0	1	1	1	1	Ape

=== Confusion Matrix ===

a	b	c	<-- classified as
76	0	0	a = Prosimian
0	214	0	b = Monkey
0	0	28	c = Ape

=== Re-evaluation on test set ===

User supplied test set
Relation: test data
Instances: 105
Attributes: 8

=== Summary ===

Correctly Classified Instances	105	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0		
Root mean squared error	0		
Total Number of Instances	105		

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0	1	1	1	1	Monkey
1	0	1	1	1	1	Ape

=== Confusion Matrix ===

a	b	c	<-- classified as
25	0	0	a = Prosimian
0	71	0	b = Monkey
0	0	9	c = Ape

Bayes Network Classifier

SYNOPSIS: Bayes Network learning using various search algorithms and quality measures. Base class for a Bayes Network classifier. Provides datastructures (network structure, conditional probability distributions, etc.) and facilities common to Bayes Network learning algorithms like K2 and B.
For more information see:

<http://www.cs.waikato.ac.nz/~remco/weka.pdf>

=== 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:        training data
Instances:        318
Attributes:       8
                  Tails
                  Chest Size
                  Arm-Leg Ratio
                  Size
                  Sleep Habits
                  Nose
                  Nails
                  Outcome
Test mode:        evaluate on training data
```

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

```
Bayes Network Classifier
not using ADTree
#attributes=8 #classindex=7
Network structure (nodes followed by parents)
```

```

Tails(2): Outcome
Chest Size(3): Outcome
Arm-Leg Ratio(3): Outcome
Size(5): Outcome
Sleep Habits(2): Outcome
Nose(2): Outcome
Nails(3): Outcome
Outcome(3):
LogScore Bayes: -966.7300307896232
LogScore BDeu: -1041.4313183663835
LogScore MDL: -1081.2851986418486
LogScore ENTROPY: -963.163145294855
LogScore AIC: -1004.163145294855

```

Time taken to build model: 0.02 seconds

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances	318	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0		
Root mean squared error	0.0001		
Relative absolute error	0.0098 %		
Root relative squared error	0.0286 %		
Total Number of Instances	318		

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0	1	1	1	1	Monkey
1	0	1	1	1	1	Ape

=== Confusion Matrix ===

a	b	c	<-- classified as
76	0	0	a = Prosimian
0	214	0	b = Monkey
0	0	28	c = Ape

=== Re-evaluation on test set ===

```

User supplied test set
Relation:      test data
Instances:    105
Attributes:    8

```

=== Summary ===

Correctly Classified Instances	105	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0		

```
Root mean squared error          0.0001
Total Number of Instances       105
```

```
=== Detailed Accuracy By Class ===
```

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0	1	1	1	1	Monkey
1	0	1	1	1	1	Ape

```
=== Confusion Matrix ===
```

```

a  b  c  <-- classified as
25  0  0 | a = Prosimian
 0 71  0 | b = Monkey
 0  0  9 | c = Ape
```

Naive Bayes Classifier

SYNOPSIS: Class for a Naive Bayes classifier using estimator classes. Numeric estimator precision values are chosen based on analysis of the training data. For this reason, the classifier is not an `UpdateableClassifier` (which in typical usage are initialized with zero training instances)—if you need the `UpdateableClassifier` functionality, use the `NaiveBayesUpdateable` classifier. The `NaiveBayesUpdateable` classifier will use a default precision of 0.1 for numeric attributes when `buildClassifier` is called with zero training instances.

For more information on Naive Bayes classifiers, see John.⁸

```
=== Run information ===
```

```

Scheme:      weka.classifiers.bayes.NaiveBayes
Relation:    training data
Instances:   318
Attributes:  8
              Tails
              Chest Size
              Arm-Leg Ratio
              Size
              Sleep Habits
              Nose
              Nails
              Outcome
Test mode:   evaluate on training data
```

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

```
Naive Bayes Classifier
```

```
Class Prosimian: Prior probability = 0.24
```

```

Tails:  Discrete Estimator. Counts =  77 1  (Total = 78)
Chest Size:  Discrete Estimator. Counts =  73 5 1  (Total = 79)
Arm-Leg Ratio:  Discrete Estimator. Counts =  77 1 1  (Total = 79)
```

⁸John GH, Langley P. Estimating continuous distributions in Bayesian classifiers. Published in 11th Conference on Uncertainty in Artificial Intelligence; 1995 Aug 18–20; Montreal, Quebec, Canada. San Mateo (CA): Morgan Kaufmann; c.1995. p. 338–345.

Size: Discrete Estimator. Counts = 27 50 2 1 1 (Total = 81)
 Sleep Habits: Discrete Estimator. Counts = 76 2 (Total = 78)
 Nose: Discrete Estimator. Counts = 77 1 (Total = 78)
 Nails: Discrete Estimator. Counts = 22 56 1 (Total = 79)

Class Monkey: Prior probability = 0.67

Tails: Discrete Estimator. Counts = 215 1 (Total = 216)
 Chest Size: Discrete Estimator. Counts = 1 190 26 (Total = 217)
 Arm-Leg Ratio: Discrete Estimator. Counts = 39 177 1 (Total = 217)
 Size: Discrete Estimator. Counts = 1 25 156 36 1 (Total = 219)
 Sleep Habits: Discrete Estimator. Counts = 7 209 (Total = 216)
 Nose: Discrete Estimator. Counts = 1 215 (Total = 216)
 Nails: Discrete Estimator. Counts = 33 1 183 (Total = 217)

Class Ape: Prior probability = 0.09

Tails: Discrete Estimator. Counts = 1 29 (Total = 30)
 Chest Size: Discrete Estimator. Counts = 1 1 29 (Total = 31)
 Arm-Leg Ratio: Discrete Estimator. Counts = 1 1 29 (Total = 31)
 Size: Discrete Estimator. Counts = 1 1 18 6 7 (Total = 33)
 Sleep Habits: Discrete Estimator. Counts = 1 29 (Total = 30)
 Nose: Discrete Estimator. Counts = 1 29 (Total = 30)
 Nails: Discrete Estimator. Counts = 1 1 29 (Total = 31)

Time taken to build model: 0 seconds

=== Evaluation on training set ===
 === Summary ===

Correctly Classified Instances	318	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0.0001		
Root mean squared error	0.0004		
Relative absolute error	0.0327	%	
Root relative squared error	0.0961	%	
Total Number of Instances	318		

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0	1	1	1	1	Monkey
1	0	1	1	1	1	Ape

=== Confusion Matrix ===

a	b	c	<-- classified as
76	0	0	a = Prosimian
0	214	0	b = Monkey
0	0	28	c = Ape

=== Re-evaluation on test set ===

User supplied test set
 Relation: test data
 Instances: 105
 Attributes: 8

=== Summary ===

Correctly Classified Instances	105	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0.0001		
Root mean squared error	0.0004		
Total Number of Instances	105		

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0	1	1	1	1	Monkey
1	0	1	1	1	1	Ape

=== Confusion Matrix ===

a	b	c	<-- classified as
25	0	0	a = Prosimian
0	71	0	b = Monkey
0	0	9	c = Ape

Logistic Function

SYNOPSIS: Class for building and using a multinomial logistic regression model with a ridge estimator.

There are some modifications, however, compared to the paper of le Cessie and van Houwelingen⁹ :

If there are k classes for n instances with m attributes, the parameter matrix B to be calculated will be an m*(k-1) matrix.

The probability for class j with the exception of the last class is

$$P_j(X_i) = \exp(X_i B_j) / ((\sum_{j=1..(k-1)} \exp(X_i B_j)) + 1)$$

The last class has probability

$$1 - (\sum_{j=1..(k-1)} P_j(X_i)) \\ = 1 / ((\sum_{j=1..(k-1)} \exp(X_i B_j)) + 1)$$

The (negative) multinomial log-likelihood is thus:

$$L = -\sum_{i=1..n} \{ \\ \sum_{j=1..(k-1)} (Y_{ij} * \ln(P_j(X_i))) \\ + (1 - (\sum_{j=1..(k-1)} Y_{ij})) \\ * \ln(1 - \sum_{j=1..(k-1)} P_j(X_i)) \}$$

⁹le Cessie S, van Houwelingen JC. Ridge estimators in logistic regression. Applied Statistics. 1992;41(1):191–201.

} + ridge * (B^2)

To find the matrix B for which L is minimised, a Quasi-Newton Method is used to search for the optimized values of the $m \times (k-1)$ variables. Before we use the optimization procedure, we 'squeeze' the matrix B into a $m \times (k-1)$ vector. For details of the optimization procedure, please check `weka.core.Optimization` class.

Although original Logistic Regression does not deal with instance weights, we modify the algorithm a little bit to handle the instance weights.

For more information see le Cessie, et al. (1992).

=== Run information ===

```
Scheme:      weka.classifiers.functions.Logistic -R 1.0E-8 -M -1
Relation:    training data
Instances:   318
Attributes:  8
              Tails
              Chest Size
              Arm-Leg Ratio
              Size
              Sleep Habits
              Nose
              Nails
              Outcome
Test mode:   evaluate on training data
```

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

Logistic Regression with ridge parameter of 1.0E-8
Coefficients...

Variable	Coeff.	
1	-9.9896	-32.249
2	5.5378	-7.7552
3	-1.6877	3.6214
4	-4.0848	3.5597
5	-5.814	6.7355
6	8.6558	4.2102
7	-9.9896	-32.249
8	-0.9676	-4.5626
9	-4.0745	2.2912
10	2.8524	0.2208
11	0.6734	-5.8543
12	0.6297	28.4311
13	-0.6473	3.3401
14	-39.8208	38.7989
15	-5.4798	9.456
16	-1.9953	-8.5087
17	4.666	-0.4291
Intercept	31.9474	-24.6568

Odds Ratios...

Variable	O.R.	
1	0	0
2	254.1236	0.0004
3	0.185	37.3885

4	0.0168	35.1512
5	0.003	841.7932
6	5743.3444	67.3675
7	0	0
8	0.38	0.0104
9	0.017	9.8864
10	17.3287	1.2471
11	1.9608	0.0029
12	1.8771	2.225633855741724E12
13	0.5234	28.2217
14	0	7.0819733759335056E16
15	0.0042	12785.2051
16	0.136	0.0002
17	106.2759	0.6511

Time taken to build model: 0.26 seconds

=== Evaluation on training set ===
 === Summary ===

Correctly Classified Instances	318	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	318		

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0	1	1	1	1	Monkey
1	0	1	1	1	1	Ape

=== Confusion Matrix ===

a	b	c	<-- classified as
76	0	0	a = Prosimian
0	214	0	b = Monkey
0	0	28	c = Ape

=== Re-evaluation on test set ===

User supplied test set
 Relation: test data
 Instances: 105
 Attributes: 8

=== Summary ===

Correctly Classified Instances	105	100	%
Incorrectly Classified Instances	0	0	%

```

Kappa statistic          1
Mean absolute error      0
Root mean squared error  0
Total Number of Instances 105

```

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0	1	1	1	1	Monkey
1	0	1	1	1	1	Ape

=== Confusion Matrix ===

```

  a  b  c  <-- classified as
25  0  0 | a = Prosimian
 0 71  0 | b = Monkey
 0  0  9 | c = Ape

```

Multilayer Perceptron Function

SYNOPSIS: A Classifier that uses back propagation to classify instances. This network can be built by hand, created by an algorithm, or both. The network can also be monitored and modified during training time. The nodes in this network are all sigmoid (except for when the class is numeric, in which case the output nodes become unthresholded linear units).

=== Run information ===

```

Scheme:      weka.classifiers.functions.MultilayerPerceptron -L 0.3 -M 0.2 -
N 500 -V 0 -S 0 -E 20 -H a
Relation:    training data
Instances:   318
Attributes:  8
              Tails
              Chest Size
              Arm-Leg Ratio
              Size
              Sleep Habits
              Nose
              Nails
              Outcome
Test mode:   evaluate on training data

```

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

```

Sigmoid Node 0
  Inputs      Weights
Threshold    1.3012106967650572
Node 3       -0.15071256073059616
Node 4       3.5665363147322036
Node 5       -1.8776292168358872
Node 6       -0.9832971903489884
Node 7       -0.9279063157966061
Node 8       -0.5714146675677457
Node 9       -2.3367795016783064
Node 10      -0.8921068780522895

```

```

Node 11      -2.3980561283521236
Node 12      0.905513496995433
Sigmoid Node 1
Inputs      Weights
Threshold   -4.591049245230357
Node 3      1.6087556954669266
Node 4      -2.3043762155765473
Node 5      -0.41420818708580687
Node 6      2.7632927457519703
Node 7      1.9784174901212792
Node 8      2.1694662343240463
Node 9      0.4527091278606214
Node 10     2.603402907190577
Node 11     -0.6199390532758464
Node 12     0.5846311391177982
Sigmoid Node 2
Inputs      Weights
Threshold   0.3271651330385988
Node 3      -1.9851678817289902
Node 4      -3.039920034682528
Node 5      1.4622001852234694
Node 6      -1.9047348840695637
Node 7      -1.2867292171442253
Node 8      -1.9403459222178623
Node 9      1.2612518617806905
Node 10     -1.909182187805521
Node 11     2.192002256276242
Node 12     -2.6010184279203687
Sigmoid Node 3
Inputs      Weights
Threshold   -0.22868683815479499
Attrib Tails -0.8706728859494896
Attrib Chest Size=equal -0.0448688366456124
Attrib Chest Size=deep  0.7324142188639325
Attrib Chest Size=broad -0.38996975745642043
Attrib Arm-Leg Ratio=shorter 0.029756656298761813
Attrib Arm-Leg Ratio=equal 1.0638679287290083
Attrib Arm-Leg Ratio=longer -0.8998606930210152
Attrib Size=very_small -0.04307226810555337
Attrib Size=small 0.04110750389722187
Attrib Size=medium 0.2616382076998021
Attrib Size=large 0.4738579902658915
Attrib Size=very_large 0.06414998512816784
Attrib Sleep Habits 0.24063898379633883
Attrib Nose 0.46129527392087954
Attrib Nails=claws 0.23387049471950103
Attrib Nails=claws_and_nails -0.1835628828345494
Attrib Nails=nails 0.15821876113733002
Sigmoid Node 4
Inputs      Weights
Threshold   -0.024454725426920874
Attrib Tails -0.40528918682444504
Attrib Chest Size=equal 0.8616469066329208
Attrib Chest Size=deep -0.4550516047893378
Attrib Chest Size=broad -0.3635436102502261
Attrib Arm-Leg Ratio=shorter 0.2727172286177625
Attrib Arm-Leg Ratio=equal 0.24719466201347665

```

```

Attrib Arm-Leg Ratio=longer      -0.34887607805371423
Attrib Size=very_small          0.37507041507268357
Attrib Size=small                0.21197841440576198
Attrib Size=medium              -0.3267426356024103
Attrib Size=large                0.03617658165743892
Attrib Size=very_large          -0.038110161865268125
Attrib Sleep Habits              -0.7437405467706905
Attrib Nose                      -1.0707113102263948
Attrib Nails=claws              -0.18177186021291622
Attrib Nails=claws_and_nails     0.7137651410337604
Attrib Nails=nails              -0.47167319926321416
Sigmoid Node 5
Inputs      Weights
Threshold   0.03857162395900296
Attrib Tails 0.18601616251264821
Attrib Chest Size=equal -0.5748860630358589
Attrib Chest Size=deep  0.35350457393749635
Attrib Chest Size=broad 0.2275878584849743
Attrib Arm-Leg Ratio=shorter -0.4484319571515167
Attrib Arm-Leg Ratio=equal 0.1782546520944032
Attrib Arm-Leg Ratio=longer 0.17832927445606872
Attrib Size=very_small -0.2569496693896704
Attrib Size=small -0.24253501956668033
Attrib Size=medium 0.2553808277244045
Attrib Size=large 0.11587673467161907
Attrib Size=very_large 0.09867814113033117
Attrib Sleep Habits 0.6500393525438719
Attrib Nose 0.7161299822285648
Attrib Nails=claws 0.011965781339222407
Attrib Nails=claws_and_nails -0.5104637106198827
Attrib Nails=nails 0.5089780614365306
Sigmoid Node 6
Inputs      Weights
Threshold   -0.4407940458466727
Attrib Tails -0.8768496017447188
Attrib Chest Size=equal -0.2993926991762047
Attrib Chest Size=deep  1.1793734360421806
Attrib Chest Size=broad -0.3687176505825691
Attrib Arm-Leg Ratio=shorter 0.010619147442384345
Attrib Arm-Leg Ratio=equal 1.3625094153376334
Attrib Arm-Leg Ratio=longer -0.902362677180076
Attrib Size=very_small 0.032285101994000004
Attrib Size=small -0.025119315809754093
Attrib Size=medium 0.47299591823879933
Attrib Size=large 0.6695639364535472
Attrib Size=very_large 0.24581818084348261
Attrib Sleep Habits 0.4179863681333151
Attrib Nose 0.9087377781035069
Attrib Nails=claws 0.5949414111242165
Attrib Nails=claws_and_nails -0.4480911911524315
Attrib Nails=nails 0.3408618390285688
Sigmoid Node 7
Inputs      Weights
Threshold   -0.42889286121398607
Attrib Tails -0.7500162424221544
Attrib Chest Size=equal -0.31209774601128293
Attrib Chest Size=deep  1.0178835714822518

```

```

Attrib Chest Size=broad      -0.349325471643971
Attrib Arm-Leg Ratio=shorter  0.00598709422663726
Attrib Arm-Leg Ratio=equal    1.1502510722402293
Attrib Arm-Leg Ratio=longer   -0.6871970988951007
Attrib Size=very_small        0.04678787258870635
Attrib Size=small             0.0035010585112271527
Attrib Size=medium            0.41566061117006115
Attrib Size=large             0.5711758656053683
Attrib Size=very_large        0.20456425031913883
Attrib Sleep Habits           0.42634221939740335
Attrib Nose                   0.7094277002076911
Attrib Nails=claws            0.48645528059175724
Attrib Nails=claws_and_nails  -0.3870580834621535
Attrib Nails=nails            0.24700015118118937

Sigmoid Node 8
Inputs      Weights
Threshold   -0.36941389560302135
Attrib Tails -0.8453340218650478
Attrib Chest Size=equal      -0.202256639703651
Attrib Chest Size=deep       0.9614405088301267
Attrib Chest Size=broad      -0.3767362058390345
Attrib Arm-Leg Ratio=shorter  0.0027711478149257426
Attrib Arm-Leg Ratio=equal    1.2519135151823058
Attrib Arm-Leg Ratio=longer   -0.8831122803646008
Attrib Size=very_small        0.05449735546384061
Attrib Size=small             -0.028370360938928713
Attrib Size=medium            0.3920810595802077
Attrib Size=large             0.5905557583568866
Attrib Size=very_large        0.18816156653425414
Attrib Sleep Habits           0.3548643253823544
Attrib Nose                   0.7561898134279856
Attrib Nails=claws            0.45810266923853177
Attrib Nails=claws_and_nails  -0.3142752996651219
Attrib Nails=nails            0.2979108556784972

Sigmoid Node 9
Inputs      Weights
Threshold    0.03820276970562375
Attrib Tails  -0.01123011608833772
Attrib Chest Size=equal      -0.756261241901086
Attrib Chest Size=deep       0.6316286927466682
Attrib Chest Size=broad      0.13943799252273947
Attrib Arm-Leg Ratio=shorter  -0.35428175030019193
Attrib Arm-Leg Ratio=equal    0.4455810613680112
Attrib Arm-Leg Ratio=longer   -0.025092368483231424
Attrib Size=very_small        -0.34397155585794514
Attrib Size=small             -0.2640088488200906
Attrib Size=medium            0.3537371978616421
Attrib Size=large             0.1375478167250311
Attrib Size=very_large        0.10486496035855912
Attrib Sleep Habits           0.7984657225897356
Attrib Nose                   0.9878054437272061
Attrib Nails=claws            0.17109204698447372
Attrib Nails=claws_and_nails  -0.6646202815697826
Attrib Nails=nails            0.505596379678196

Sigmoid Node 10
Inputs      Weights
Threshold    -0.4279610177424282

```



```

Attrib Tails      -0.8555043686551411
Attrib Chest Size=equal    -0.31896739070126745
Attrib Chest Size=deep     1.128377494236569
Attrib Chest Size=broad    -0.39480327170169116
Attrib Arm-Leg Ratio=shorter  0.05974792636865227
Attrib Arm-Leg Ratio=equal   1.3239447319338782
Attrib Arm-Leg Ratio=longer  -0.8997271829758088
Attrib Size=very_small     0.04605954948148706
Attrib Size=small         -0.061164288925263384
Attrib Size=medium        0.4436682083935165
Attrib Size=large         0.6725321321722918
Attrib Size=very_large     0.27823415177038313
Attrib Sleep Habits       0.4608065164708468
Attrib Nose              0.8265080080813495
Attrib Nails=claws        0.5572608055533798
Attrib Nails=claws_and_nails -0.4192955619948751
Attrib Nails=nails        0.3536017576701385

Sigmoid Node 11
Inputs      Weights
Threshold   -0.006815893586848453
Attrib Tails  0.2995164285845169
Attrib Chest Size=equal    -0.6590786212943847
Attrib Chest Size=deep     0.3450280887297489
Attrib Chest Size=broad    0.33332317704047076
Attrib Arm-Leg Ratio=shorter -0.5028573164224909
Attrib Arm-Leg Ratio=equal  0.18563330908873857
Attrib Arm-Leg Ratio=longer 0.26434020567495936
Attrib Size=very_small     -0.2653476914665689
Attrib Size=small         -0.24983172019176017
Attrib Size=medium        0.2107053119125198
Attrib Size=large         0.07724611111975499
Attrib Size=very_large     0.0397859326235434
Attrib Sleep Habits       0.7510339172304787
Attrib Nose              0.8023427967140727
Attrib Nails=claws        0.0034222325005445926
Attrib Nails=claws_and_nails -0.49747421996597047
Attrib Nails=nails        0.4772318405942686

Sigmoid Node 12
Inputs      Weights
Threshold    0.1375267582521297
Attrib Tails  -1.0559935859455922
Attrib Chest Size=equal    0.03197272950039177
Attrib Chest Size=deep     0.3299648939964681
Attrib Chest Size=broad    -0.6484497712771019
Attrib Arm-Leg Ratio=shorter 0.28535546672457196
Attrib Arm-Leg Ratio=equal  0.6735763642494895
Attrib Arm-Leg Ratio=longer -1.1306099287755755
Attrib Size=very_small     -0.1339336694249159
Attrib Size=small         0.07538523227217672
Attrib Size=medium        -0.19396330812424623
Attrib Size=large         0.025155585757356783
Attrib Size=very_large     -0.32484897247477534
Attrib Sleep Habits       -0.09150058326118905
Attrib Nose              0.0044122906645481024
Attrib Nails=claws        -0.031248739211253762
Attrib Nails=claws_and_nails 0.031606951232683596
Attrib Nails=nails        -0.19904197135905405

```

```

Class Prosimian
  Input
    Node 0
Class Monkey
  Input
    Node 1
Class Ape
  Input
    Node 2

```

Time taken to build model: 10.12 seconds

=== Evaluation on training set ===
 === Summary ===

Correctly Classified Instances	318	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0.003		
Root mean squared error	0.0041		
Relative absolute error	0.9363	%	
Root relative squared error	1.0128	%	
Total Number of Instances	318		

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0	1	1	1	1	Monkey
1	0	1	1	1	1	Ape

=== Confusion Matrix ===

a	b	c	<-- classified as
76	0	0	a = Prosimian
0	214	0	b = Monkey
0	0	28	c = Ape

=== Re-evaluation on test set ===

```

User supplied test set
Relation:      test data
Instances:    105
Attributes:    8

```

=== Summary ===

Correctly Classified Instances	105	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0.003		
Root mean squared error	0.004		
Total Number of Instances	105		

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0	1	1	1	1	Monkey
1	0	1	1	1	1	Ape

=== Confusion Matrix ===

a	b	c	<-- classified as
25	0	0	a = Prosimian
0	71	0	b = Monkey
0	0	9	c = Ape

Locally Weighted Learning (LWL) Classifier

SYNOPSIS: Locally weighted learning. It uses an instance-based algorithm to assign instance weights, which are then used by a specified WeightedInstancesHandler. Can do classification (e.g. using naive Bayes) or regression (e.g. using linear regression).

For more info, see Frank¹⁰ and Atkeson.¹¹

=== Run information ===

```
Scheme:          weka.classifiers.lazy.LWL -U 0 -K -1 -A
"weka.core.neighboursearch.LinearNNSearch -A weka.core.EuclideanDistance" -W
weka.classifiers.trees.DecisionStump
Relation:        training data
Instances:       318
Attributes:      8
                  Tails
                  Chest Size
                  Arm-Leg Ratio
                  Size
                  Sleep Habits
                  Nose
                  Nails
                  Outcome
Test mode:       evaluate on training data
```

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

```
Locally weighted learning
=====
Using classifier: weka.classifiers.trees.DecisionStump
Using linear weighting kernels
Using all neighbours
Time taken to build model: 0.01 seconds
```

=== Evaluation on training set ===

=== Summary ===

¹⁰ Frank E, Hall M, Pfahringer B. Locally Weighted Naive Bayes. Published In: 19th Conference in Uncertainty in Artificial Intelligence; 2003 Aug 7–10; Acapulco, NM. San Francisco (CA): Morgan Kaufmann; c2003. p. 249–256.

¹¹ Atkeson C, Moore A, Schaal S. Locally weighted learning. AI Review. 1997;11(1–5):11–73.

Correctly Classified Instances	318	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0.0027		
Root mean squared error	0.0072		
Relative absolute error	0.8347	%	
Root relative squared error	1.7957	%	
Total Number of Instances	318		

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0	1	1	1	1	Monkey
1	0	1	1	1	1	Ape

=== Confusion Matrix ===

a	b	c	<-- classified as
76	0	0	a = Prosimian
0	214	0	b = Monkey
0	0	28	c = Ape

=== Re-evaluation on test set ===

User supplied test set
 Relation: test data
 Instances: 105
 Attributes: 8

=== Summary ===

Correctly Classified Instances	105	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0.0025		
Root mean squared error	0.0067		
Total Number of Instances	105		

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0	1	1	1	1	Prosimian
1	0	1	1	1	1	Monkey
1	0	1	1	1	1	Ape

=== Confusion Matrix ===

a	b	c	<-- classified as
25	0	0	a = Prosimian
0	71	0	b = Monkey
0	0	9	c = Ape

1 DEFENSE TECHNICAL
(PDF) INFORMATION CTR
DTIC OCA

2 DIRECTOR
(PDF) US ARMY RESEARCH LAB
RDRL CIO LL
IMAL HRA MAIL & RECORDS
MGMT

1 GOVT PRINTG OFC
(PDF) A MALHOTRA

1 ARMY RSCH LAB – HRED
(PDF) RDRL HRM D
T DAVIS
BLDG 5400 RM C242
REDSTONE ARSENAL AL
35898-7290

1 ARMY RSCH LAB – HRED
(PDF) RDRL HRS EA DR V J RICE
BLDG 4011 RM 217
1750 GREELEY RD
FORT SAM HOUSTON TX
78234-5002

1 ARMY RSCH LAB – HRED
(PDF) RDRL HRM DG J RUBINSTEIN
BLDG 333
PICATINNY ARSENAL NJ
07806-5000

1 ARMY RSCH LAB – HRED
(PDF) ARMC FIELD ELEMENT
RDRL HRM CH C BURNS
THIRD AVE BLDG 1467B RM 336
FORT KNOX KY 40121

1 ARMY RSCH LAB – HRED
(PDF) AWC FIELD ELEMENT
RDRL HRM DJ D DURBIN
BLDG 4506 (DCD) RM 107
FORT RUCKER AL 36362-5000

1 ARMY RSCH LAB – HRED
(PDF) RDRL HRM CK J REINHART
10125 KINGMAN RD BLDG 317
FORT BELVOIR VA 22060-5828

1 ARMY RSCH LAB – HRED
(PDF) RDRL HRM AY M BARNES
2520 HEALY AVE
STE 1172 BLDG 51005
FORT HUACHUCA AZ
85613-7069

1 ARMY RSCH LAB – HRED
(PDF) RDRL HRM AP D UNGVARSKY
POPE HALL BLDG 470
BCBL 806 HARRISON DR
FORT LEAVENWORTH KS
66027-2302

1 ARMY RSCH LAB – HRED
(PDF) RDRL HRM AT J CHEN
12423 RESEARCH PKWY
ORLANDO FL 32826-3276

1 ARMY RSCH LAB – HRED
(PDF) HUMAN SYSTEMS
INTEGRATION ENGR
TACOM FIELD ELEMENT
RDRL HRM CU P MUNYA
6501 E 11 MILE RD
MS 284 BLDG 200A
WARREN MI 48397-5000

1 ARMY RSCH LAB – HRED
(PDF) FIRES CTR OF EXCELLENCE
FIELD ELEMENT
RDRL HRM AF C HERNANDEZ
3040 NW AUSTIN RD RM 221
FORT SILL OK 73503-9043

1 ARMY RSCH LAB – HRED
(PDF) RDRL HRM AV
W CULBERTSON
91012 STATION AVE
FORT HOOD TX 76544-5073

1 ARMY RSCH LAB – HRED
(PDF) RDRL HRM DE A MARES
1733 PLEASANTON RD BOX 3
FORT BLISS TX 79916-6816

8 ARMY RSCH LAB – HRED
(PDF) SIMULATION & TRAINING
TECHNOLOGY CENTER
RDRL HRT COL G LAASE
RDRL HRT I MARTINEZ
RDRL HRT T R SOTTILARE
RDRL HRT B N FINKELSTEIN
RDRL HRT G A RODRIGUEZ
RDRL HRT I J HART
RDRL HRT M C METEVIER
RDRL HRT S B PETTIT
12423 RESEARCH PARKWAY
ORLANDO FL 32826

1 ARMY RSCH LAB – HRED
(PDF) HQ USASOC
RDRL HRM CN R SPENCER
BLDG E2929 DESERT STORM DR
FORT BRAGG NC 28310

1 ARMY G1
(PDF) DAPE MR B KNAPP
300 ARMY PENTAGON RM 2C489
WASHINGTON DC 20310-0300

ABERDEEN PROVING GROUND

12 DIR USARL
(PDF) RDRL HR
L ALLENDER
P FRANASZCZUK
RDRL HRM
P SAVAGE-KNEPSHIELD
RDRL HRM AL
C PAULILLO
RDRL HRM B
J GRYNOVICKI
RDRL HRM C
L GARRETT
RDRL HRS
J LOCKETT
RDRL HRS B
M LAFIANDRA
RDRL HRS C
K MCDOWELL
RDRL HRS D
A SCHARINE
RDRL HRS E
D HEADLEY
RDRL HRT T
R SOTTILARE