1.

As for the weather.arff file, only the attributes of “temperature” and “humidity” are numetic.

Some statistics of “temperature”

Selected attribute

Name: temperature Type: Numeric
Missing: 0 (0%) Distinct: 12 Unique: 10 (71%)

Statistic	Value
Minimum	64
Maximum	85
Mean	73.571
StdDev	6.572

Some statistics of “humidity”

Selected attribute

Name: humidity Type: Numeric
Missing: 0 (0%) Distinct: 10 Unique: 7 (50%)

Statistic	Value
Minimum	65
Maximum	96
Mean	81.643
StdDev	10.285

As for the nominal attributes:

Selected attribute

Name: outlook Type: Nominal
Missing: 0 (0%) Distinct: 3 Unique: 0 (0%)

No.	Label	Count
1	sunny	5
2	overcast	4
3	rainy	5

Selected attribute

Name: windy Type: Nominal
Missing: 0 (0%) Distinct: 2 Unique: 0 (0%)

No.	Label	Count
1	TRUE	6
2	FALSE	8

Selected attribute

Name: play Type: Nominal
Missing: 0 (0%) Distinct: 2 Unique: 0 (0%)

No.	Label	Count
1	yes	9
2	no	5

2.

In this part, I used the weather.arff and just choose tree nominal attributes: “outlook”, “windy”, “play”. The result shows the classification is not very remarkable, that is because ZeroR is mainly based on the rule and predict the majority class. The minority class in the dataset sometimes cannot be predicted according to the rule (some combinations of attributes cannot found in the rule).

Output:

=== Run information ===

Scheme: weka.classifiers.rules.ZeroR
Relation: weather-weka.filters.unsupervised.attribute.Remove-R2-3
Instances: 14
Attributes: 3
outlook
windy
play
Test mode: 10-fold cross-validation

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

ZeroR predicts class value: yes

Time taken to build model: 0.02 seconds

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

Correctly Classified Instances	9	64.2857 %
Incorrectly Classified Instances	5	35.7143 %
Kappa statistic	0	
Mean absolute error	0.4762	
Root mean squared error	0.4934	
Relative absolute error	100	%
Root relative squared error	100	%
Total Number of Instances	14	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	1	1	0.643	1	0.783	0.178	yes
	0	0	0	0	0	0.178	no
Weighted Avg.	0.643	0.643	0.413	0.643	0.503	0.178	

=== Confusion Matrix ===

```
 a b   <-- classified as
 9 0 | a = yes
 5 0 | b = no
```

3.

I used the weather.nominal.arff file in this part.

Output:

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: weather.symbolic
Instances: 14
Attributes: 5
outlook
temperature
humidity
windy
play
Test mode: 10-fold cross-validation

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

J48 pruned tree

---

outlook = sunny
| humidity = high: no (3.0)
| humidity = normal: yes (2.0)
outlook = overcast: yes (4.0)
outlook = rainy
| windy = TRUE: no (2.0)
| windy = FALSE: yes (3.0)

Number of Leaves : 5

Size of the tree : 8

Time taken to build model: 0.02 seconds

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

Correctly Classified Instances	7	50 %
Incorrectly Classified Instances	7	50 %
Kappa statistic	-0.0426	
Mean absolute error	0.4167	
Root mean squared error	0.5984	
Relative absolute error	87.5 %	
Root relative squared error	121.2987 %	
Total Number of Instances	14	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.556	0.6	0.625	0.556	0.588	0.633	yes
	0.4	0.444	0.333	0.4	0.364	0.633	no
Weighted Avg.	0.5	0.544	0.521	0.5	0.508	0.633	

=== Confusion Matrix ===

 a b <-- classified as
 5 4 | a = yes
 3 2 | b = no

4.

In this part, I used weather.nominal.arff.

Rule-based:

1) ConjunctiveRule:

=== Run information ===

Scheme: weka.classifiers.rules.ConjunctiveRule -N 3 -M 2.0 -P -1 -S 1
Relation: weather.symbolic
Instances: 14
Attributes: 5
outlook
temperature
humidity
windy
play
Test mode: 10-fold cross-validation

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

Single conjunctive rule learner:

=> play = yes

Class distributions:
Covered by the rule:
yes no
0.6 0.4

Not covered by the rule:
yes no
0 0

Time taken to build model: 0 seconds

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

Correctly Classified Instances	9	64.2857 %
Incorrectly Classified Instances	5	35.7143 %
Kappa statistic	0	
Mean absolute error	0.4762	
Root mean squared error	0.5051	
Relative absolute error	100 %	
Root relative squared error	102.3787 %	
Total Number of Instances	14	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	1	1	0.643	1	0.783	0.333	yes
	0	0	0	0	0	0.333	no
Weighted Avg.	0.643	0.643	0.413	0.643	0.503	0.333	

=== Confusion Matrix ===

a	b	<-- classified as
9	0	a = yes
5	0	b = no

Analysis:

This algorithm generates a rule which consists of antecedents "AND"ed together and the consequent for the classification. In this case, the consequent is the distribution of the available classes in the dataset. If the test instance is not covered by this rule, then it's predicted using the default class value of the data not covered by the rule in the training data. We can see this kind of classification still cannot improve the performance for our dataset.

2) OneR

=== Run information ===

Scheme: weka.classifiers.rules.OneR -B 6
Relation: weather.symbolic
Instances: 14
Attributes: 5
outlook
temperature
humidity
windy

play
Test mode: 10-fold cross-validation

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

outlook:
sunny -> no
overcast -> yes
rainy -> yes
(10/14 instances correct)

Time taken to build model: 0 seconds

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

Correctly Classified Instances	6	42.8571 %
Incorrectly Classified Instances	8	57.1429 %
Kappa statistic	-0.1429	
Mean absolute error	0.5714	
Root mean squared error	0.7559	
Relative absolute error	120 %	
Root relative squared error	153.2194 %	
Total Number of Instances	14	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.444	0.6	0.571	0.444	0.5	0.422	yes
	0.4	0.556	0.286	0.4	0.333	0.422	no
Weighted Avg.	0.429	0.584	0.469	0.429	0.44	0.422	

=== Confusion Matrix ===

a	b	<-- classified as
4	5	a = yes
3	2	b = no

Analysis:

It builds and uses a 1R classifier; in other words, uses the minimum-error attribute for prediction, discretizing numeric attributes.

3) **Ridor**

=== Run information ===

Scheme: weka.classifiers.rules.Ridor -F 3 -S 1 -N 2.0
Relation: weather.symbolic
Instances: 14
Attributes: 5
outlook
temperature
humidity
windy
play
Test mode: 10-fold cross-validation

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

RIpple DOwn Rule Learner(Ridor) rules
---

play = yes (14.0/0.0)

Total number of rules (incl. the default rule): 1

Time taken to build model: 0 seconds

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

Correctly Classified Instances	5	35.7143 %
Incorrectly Classified Instances	9	64.2857 %
Kappa statistic	-0.4651	
Mean absolute error	0.6429	
Root mean squared error	0.8018	
Relative absolute error	135 %	
Root relative squared error	162.5137 %	
Total Number of Instances	14	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.556	1	0.5	0.556	0.526	0.278	yes
	0	0.444	0	0	0	0.278	no
Weighted Avg.	0.357	0.802	0.321	0.357	0.338	0.278	

=== Confusion Matrix ===

```
 a b   <-- classified as
 5 4 | a = yes
 5 0 | b = no
```

Analysis:

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. This algorithm has a worse performance than ZeroR for our dataset.

5.

NNge:

Nearest-neighbor method of generating rules using nonnested generalized exmplars.

Output:

=== Run information ===

```
Scheme:       weka.classifiers.rules.NNge -G 5 -I 5
Relation:     weather.symbolic
Instances:    14
Attributes:   5
              outlook
              temperature
              humidity
              windy
              play
Test mode:    10-fold cross-validation
```

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

NNGE classifier

Rules generated :

class no IF : outlook in {rainy} ^ temperature in {mild,cool} ^ humidity in {high,normal} ^ windy in {TRUE} (2)

class yes IF : outlook in {overcast,rainy} ^ temperature in {hot,mild,cool} ^ humidity in {high,normal} ^ windy in {FALSE} (5)

class yes IF : outlook in {overcast} ^ temperature in {mild,cool} ^ humidity in {high,normal} ^ windy in {TRUE} (2)

class yes IF : outlook in {sunny} ^ temperature in {mild,cool} ^ humidity in {normal} ^ windy in {TRUE,FALSE} (2)

class no IF : outlook in {sunny} ^ temperature in {hot,mild} ^ humidity in {high} ^ windy in {TRUE,FALSE} (3)

Stat :

class yes : 3 exemplar(s) including 3 Hyperrectangle(s) and 0 Single(s).

class no : 2 exemplar(s) including 2 Hyperrectangle(s) and 0 Single(s).

Total : 5 exemplars(s) including 5 Hyperrectangle(s) and 0 Single(s).

Feature weights : [0.24674981977443894 0.029222565658954577 0.15183550136234153 0.04812703040826924]

Time taken to build model: 0 seconds

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

Correctly Classified Instances	11	78.5714 %
Incorrectly Classified Instances	3	21.4286 %
Kappa statistic	0.5116	
Mean absolute error	0.2143	
Root mean squared error	0.4629	
Relative absolute error	45 %	
Root relative squared error	93.8273 %	
Total Number of Instances	14	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.889	0.4	0.8	0.889	0.842	0.744	yes
	0.6	0.111	0.75	0.6	0.667	0.744	no
Weighted Avg.	0.786	0.297	0.782	0.786	0.779	0.744	

=== Confusion Matrix ===

a	b	<-- classified as

8 1 | a = yes
2 3 | b = no

kStar:

It is an instance-based classifier, that is the class of a test instance is based upon the class of those training instances similar to it, as determined by some similarity function. It differs from other instance-based learners in that it uses an entropy-based distance function.

Output:

=== Run information ===

Scheme: weka.classifiers.lazy.KStar -B 20 -M a
Relation: weather.symbolic
Instances: 14
Attributes: 5
outlook
temperature
humidity
windy
play
Test mode: 10-fold cross-validation

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

KStar Beta Verion (0.1b).
Copyright (c) 1995-97 by Len Trigg (trigg@cs.waikato.ac.nz).
Java port to Weka by Abdelaziz Mahoui (am14@cs.waikato.ac.nz).

KStar options : -B 20 -M a

Time taken to build model: 0 seconds

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

Correctly Classified Instances	8	57.1429 %
Incorrectly Classified Instances	6	42.8571 %
Kappa statistic	0.0667	
Mean absolute error	0.4711	
Root mean squared error	0.5433	
Relative absolute error	98.9405 %	
Root relative squared error	110.1307 %	

Total Number of Instances 14

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.667	0.6	0.667	0.667	0.667	0.489	yes
	0.4	0.333	0.4	0.4	0.4	0.489	no
Weighted Avg.	0.571	0.505	0.571	0.571	0.571	0.489	

=== Confusion Matrix ===

a	b	<-- classified as
6	3	a = yes
3	2	b = no