

Experiment 2.2

Student Name:- Nikhil Kumar

Branch: B.E CSE

Semester: 6

Subject Name: Data Mining Lab

UID 20BCS1817

Section/Group: 716/B

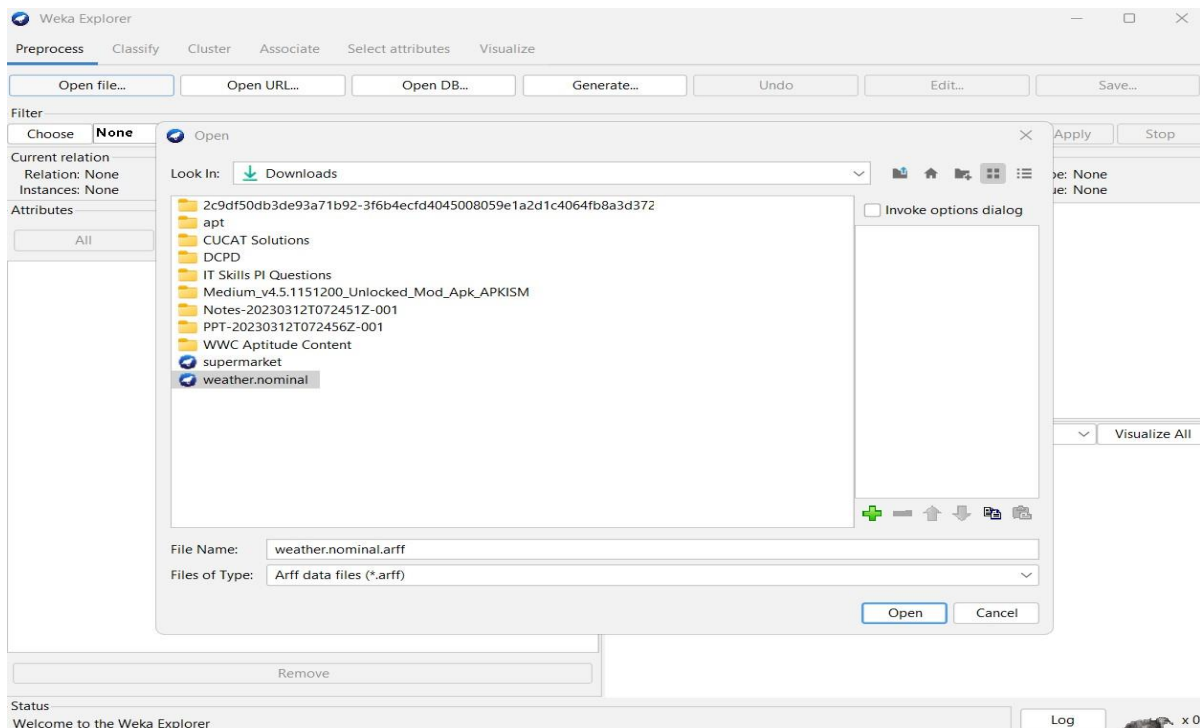
Date of Performance: 05/04/23

Subject Code: 20CSP-351

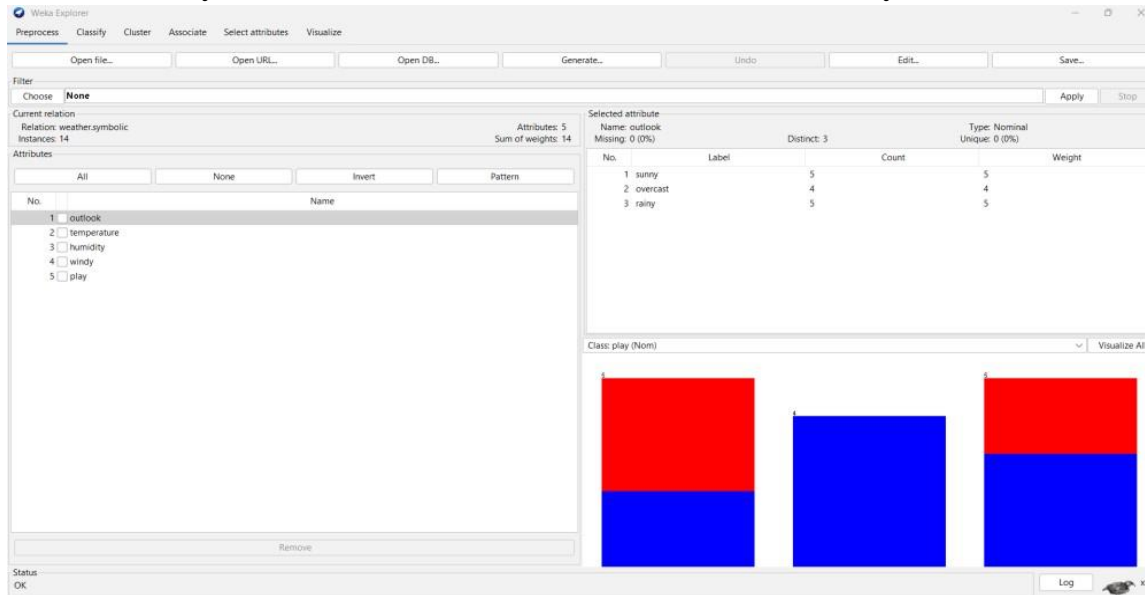
Aim: To perform classification using Bayesian classification algorithm using Rweka.

Procedure:

1. Load the weather Dataset.



2. Identify the attributes and decide a method to classify the dataset.



Weka Explorer - Preprocess

Filter: Choose **None** [Apply] [Stop]

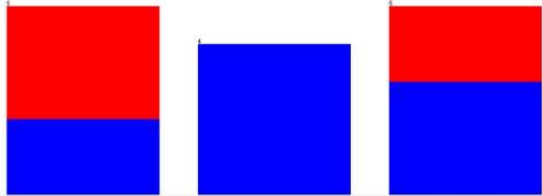
Current relation: weather.symbolic
Instances: 14
Attributes: 5
Sum of weights: 14

Attributes: outlook, temperature, humidity, windy, play

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

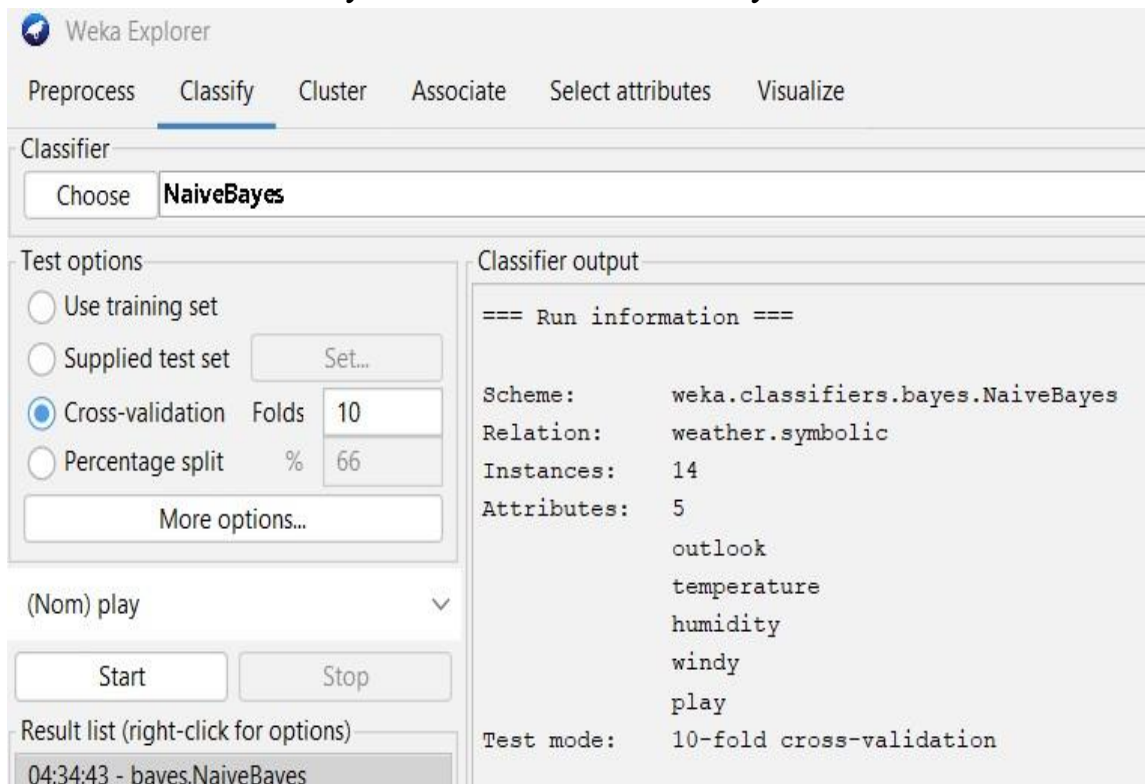
No.	Label	Count	Weight
1	sunny	5	5
2	overcast	4	4
3	rainy	5	5

Class: play (Nom) [Visualize All]



Status: OK [Log]

3. Go to the classify tab and choose “NaiveBayes Classifier”



Weka Explorer - Classify

Classifier: Choose **NaiveBayes**

Test options:

- ☐ Use training set
- ☐ Supplied test set [Set...]
- ☒ Cross-validation Folds: 10
- ☐ Percentage split %: 66

[More options...]

(Nom) play [v]

[Start] [Stop]

Result list (right-click for options)

04:34:43 - baves.NaiveBayes

Classifier output:

```

=== Run information ===

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

4. Start the classification in order to display the results of the classification.

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

```
Naive Bayes Classifier
```

```
Attribute      Class  
              yes    no  
              (0.63) (0.38)  
=====
```

```
outlook  
  sunny          3.0    4.0  
  overcast       5.0    1.0  
  rainy          4.0    3.0  
  [total]       12.0    8.0
```

```
temperature  
  hot            3.0    3.0  
  mild           5.0    3.0  
  cool           4.0    2.0  
  [total]       12.0    8.0
```

```
humidity  
  high           4.0    5.0  
  normal         7.0    2.0  
  [total]       11.0    7.0
```

```
windy  
  TRUE           4.0    4.0  
  FALSE          7.0    3.0  
  [total]       11.0    7.0
```

5. The confusion matrix and statistics are evaluated as follows.

```
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                    -0.0244
Mean absolute error                 0.4374
Root mean squared error             0.4916
Relative absolute error             91.8631 %
Root relative squared error         99.6492 %
Total Number of Instances          14

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

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.778	0.800	0.636	0.778	0.700	-0.026	0.578	0.697	yes
	0.200	0.222	0.333	0.200	0.250	-0.026	0.578	0.557	no
Weighted Avg.	0.571	0.594	0.528	0.571	0.539	-0.026	0.578	0.647	

```


=== Confusion Matrix ===

a b  <-- classified as
7 2 | a = yes
4 1 | b = no
```