

REPORT – CLASSIFICATION TASK (WEKA)

1. Dataset

The dataset used is the **Iris dataset**, which contains 3 classes:

- setosa
- versicolor
- virginica

The analysis was performed using **Weka**.

2. Objective

The objective of this experiment is to compare different classification models and evaluate their performance in terms of:

- accuracy
- precision
- recall
- confusion matrix

3. Methodology

Two classification algorithms were used:

- **J48 (Decision Tree)** → baseline model
- **RandomForest** → improved model

The evaluation was performed using **10-fold cross-validation**.

4. Results

◇ 4.1 J48

Confusion Matrix:

49	1	0
0	47	3
0	2	48

Performance:

- Setosa: Precision = 1.000, Recall = 0.980
- Versicolor: Precision = 0.940, Recall = 0.940
- Virginica: Precision = 0.941, Recall = 0.960

◇ 4.2 RandomForest

Confusion Matrix:

50	0	0
0	47	3
0	4	46

5. Model Comparison

Both models achieved high performance.

- **J48**: slightly more precise on some classes
- **RandomForest**: more stable and robust overall

There is no significant difference in overall accuracy, but RandomForest provides more consistent results.

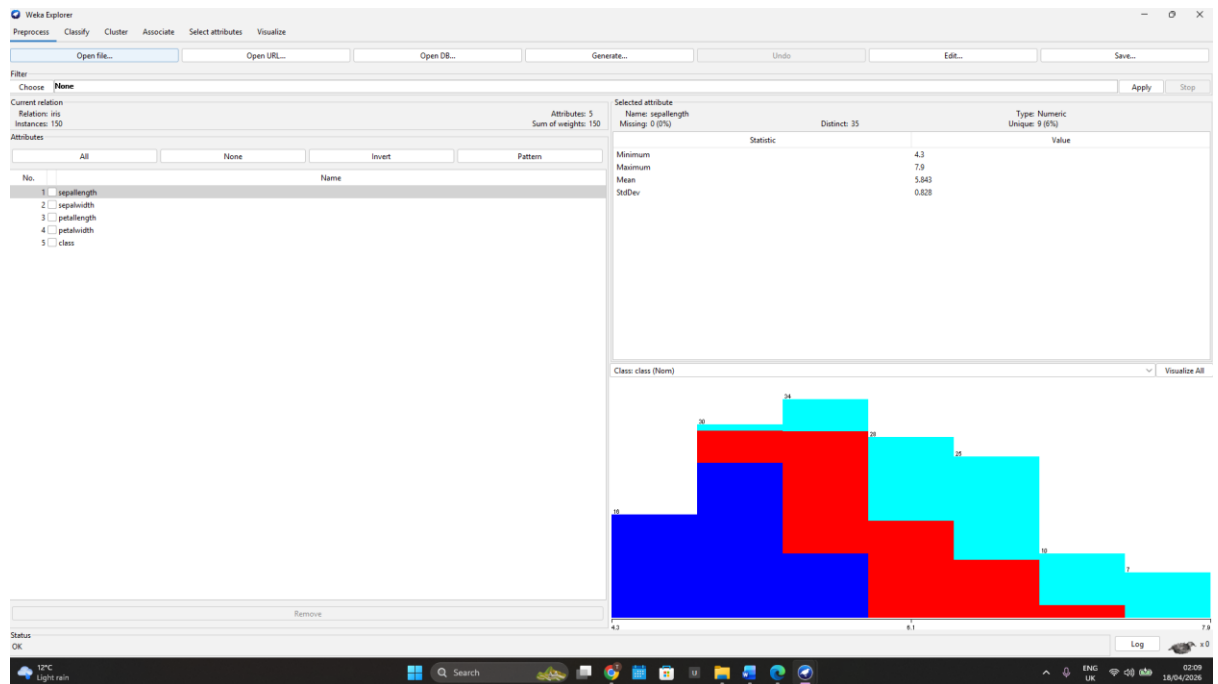
6. Conclusion

The experiment shows that both models are effective for classifying the Iris dataset. However, RandomForest performs slightly better in terms of stability and generalization.



SCREENSHOTS

1. Dataset loaded (Preprocess tab with Iris dataset)
2. J48 results (confusion matrix + metrics)
3. RandomForest results (confusion matrix + metrics)



Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier: Choose **RandomForest** -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Test options:
☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds: 10
☐ Percentage split %: 66
More options...

(Nom) class: Start Stop

Result list (right-click for options):
02:03:50 - trees.J48
02:07:45 - trees.RandomForest

Classifier output

=== Classifier model (full training set) ===
J48 pruned tree

petalwidth <= 0.6: Iris-setosa (50.0)
petalwidth > 0.6:
| petalwidth <= 1.7:
| | petallength <= 4.9: Iris-versicolour (45.0/1.0)
| | petallength > 4.9:
| | | petalwidth <= 1.5: Iris-virginica (3.0)
| | | petalwidth > 1.5: Iris-versicolour (3.0/1.0)
| petalwidth > 1.7: Iris-virginica (46.0/1.0)

Number of Leaves : 5
Size of the tree : 9
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances 144 96 %
Incorrectly Classified Instances 6 4 %
Kappa statistic 0.94
Mean absolute error 0.035
Root mean squared error 0.1866
Relative absolute error 7.8705 %
Root relative squared error 33.4333 %
Total Number of Instances 150

=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure ROC Area ROC Area PRC Area Class
0.950 0.000 1.000 0.950 0.950 0.945 0.950 0.907 Iris-setosa
0.940 0.030 0.940 0.940 0.940 0.910 0.952 0.880 Iris-versicolour
0.960 0.030 0.941 0.960 0.950 0.925 0.941 0.905 Iris-virginica
Weighted Avg. 0.960 0.020 0.960 0.960 0.960 0.945 0.963 0.924

=== Confusion Matrix ===
a b c <-- classified as
49 1 0 | a = Iris-setosa
0 47 3 | b = Iris-versicolour
0 2 48 | c = Iris-virginica

Status: OK Log

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier: Choose **RandomForest** -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Test options:
☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds: 10
☐ Percentage split %: 66
More options...

(Nom) class: Start Stop

Result list (right-click for options):
02:03:50 - trees.J48
02:07:45 - trees.RandomForest

Classifier output

Scheme: weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation: iris
Instances: 150
Attributes: 5
petalwidth
sepalwidth
petallength
class

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===
RandomForest
Bagging with 100 iterations and base learner
weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities
Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances 143 95.3333 %
Incorrectly Classified Instances 7 4.6667 %
Kappa statistic 0.93
Mean absolute error 0.0409
Root mean squared error 0.1621
Relative absolute error 9.19 %
Root relative squared error 34.3844 %
Total Number of Instances 150

=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure ROC Area ROC Area PRC Area Class
1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000 Iris-setosa
0.940 0.040 0.922 0.940 0.931 0.896 0.991 0.904 Iris-versicolour
0.920 0.030 0.939 0.920 0.929 0.895 0.951 0.902 Iris-virginica
Weighted Avg. 0.953 0.023 0.953 0.953 0.953 0.930 0.994 0.909

=== Confusion Matrix ===
a b c <-- classified as
50 0 0 | a = Iris-setosa
0 47 3 | b = Iris-versicolour
0 4 46 | c = Iris-virginica

Status: OK Log