

Experiment-2.2

Student Name: Ashish Kumar

Branch: CSE AIML

Semester: 01

Subject Name: Artificial Intelligence Lab

UID: 23MAI10008

Section/Group: 23MAI-1

Date of Performance: 09/10/2023

Subject Code: 23CSH-621

Aim of the Experiment :

Aim of the Experiment is to apply different testing techniques for the UCI dataset and compare the performance of the best prediction model with the following techniques:

- a) 70/30 Split
- b) K-fold cross validation

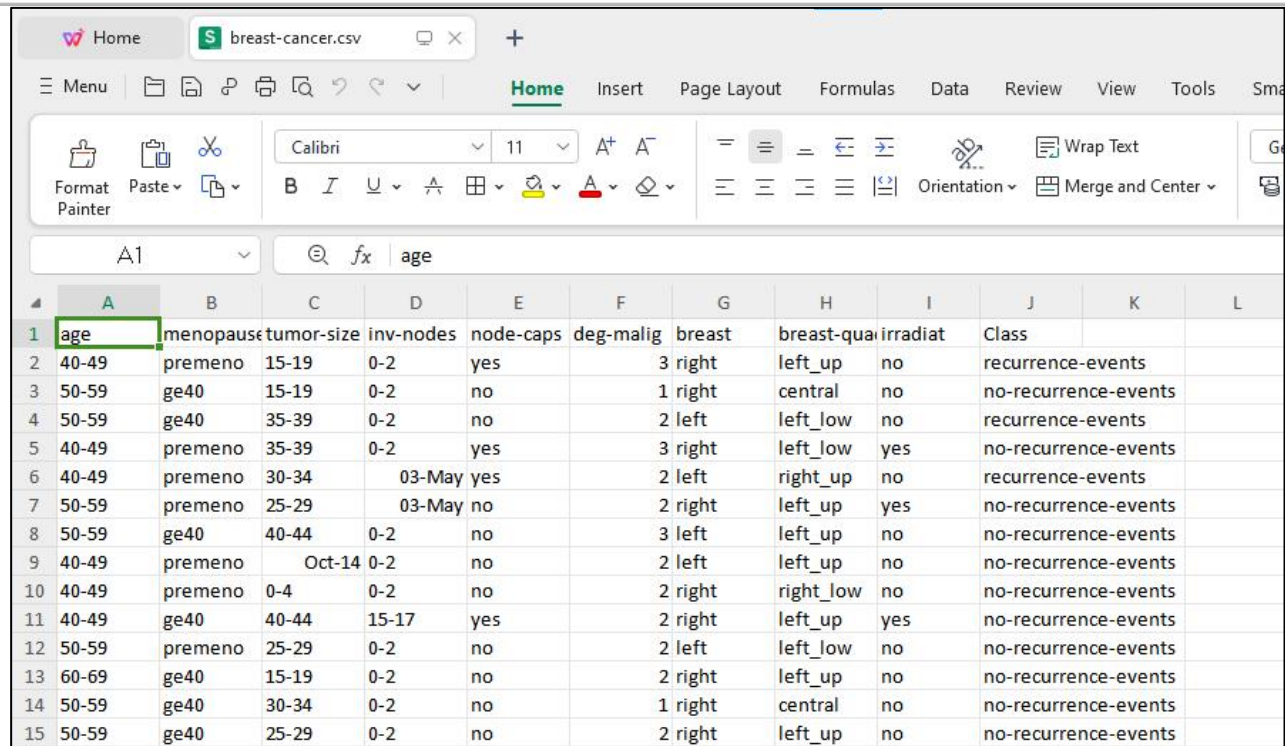
Objective of the Experiment :

Task to be done for this experiment is that we have to use the UCI dataset and find the best prediction model. Then we have to apply the different testing techniques for the UCI dataset and compare the performance of the best prediction model with the help of the following techniques:

- a) 70//30 Split
- b) K-fold cross validation

Algorithm/ Steps for Experiment :

Step 1: Download the **Cancer dataset** from UCI Machine Learning repository.

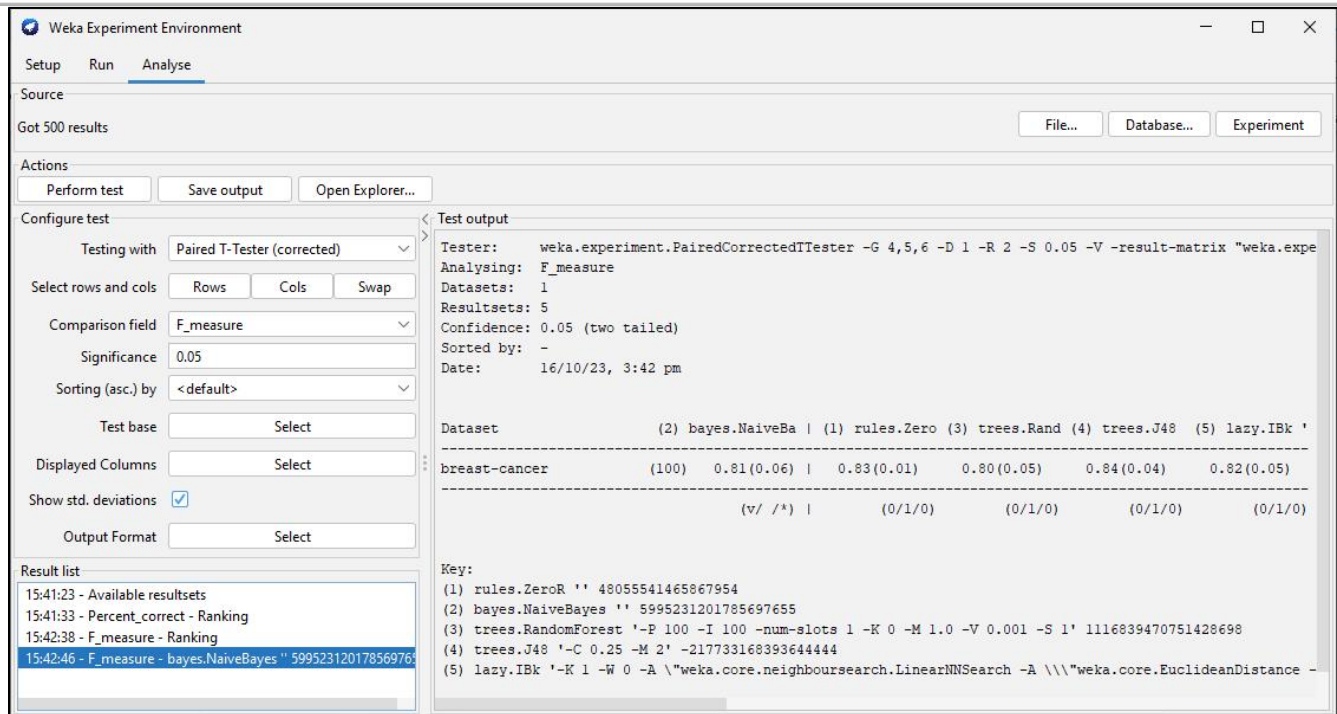


	A	B	C	D	E	F	G	H	I	J	K	L
1	age	menopause	tumor-size	inv-nodes	node-caps	deg-malig	breast	breast-qual	irradiat	Class		
2	40-49	premeno	15-19	0-2	yes		3 right	left_up	no	recurrence-events		
3	50-59	ge40	15-19	0-2	no		1 right	central	no	no-recurrence-events		
4	50-59	ge40	35-39	0-2	no		2 left	left_low	no	recurrence-events		
5	40-49	premeno	35-39	0-2	yes		3 right	left_low	yes	no-recurrence-events		
6	40-49	premeno	30-34	03-May	yes		2 left	right_up	no	recurrence-events		
7	50-59	premeno	25-29	03-May	no		2 right	left_up	yes	no-recurrence-events		
8	50-59	ge40	40-44	0-2	no		3 left	left_up	no	no-recurrence-events		
9	40-49	premeno	Oct-14	0-2	no		2 left	left_up	no	no-recurrence-events		
10	40-49	premeno	0-4	0-2	no		2 right	right_low	no	no-recurrence-events		
11	40-49	ge40	40-44	15-17	yes		2 right	left_up	yes	no-recurrence-events		
12	50-59	premeno	25-29	0-2	no		2 left	left_low	no	no-recurrence-events		
13	60-69	ge40	15-19	0-2	no		2 right	left_up	no	no-recurrence-events		
14	50-59	ge40	30-34	0-2	no		1 right	central	no	no-recurrence-events		
15	50-59	ge40	25-29	0-2	no		2 right	left_up	no	no-recurrence-events		

Step 2: Open WEKA and instead of 'Explorer' tab, open the 'Experimenter' tab.



Step 3: Find the best prediction model out of the applied 5 classifiers which is J48 Classifier.



Weka Experiment Environment

Setup Run **Analyse**

Source: Got 500 results

File... Database... Experiment

Actions: Perform test Save output Open Explorer...

Configure test

Testing with: Paired T-Tester (corrected)

Select rows and cols: Rows Cols Swap

Comparison field: F_measure

Significance: 0.05

Sorting (asc.) by: <default>

Test base: Select

Displayed Columns: Select

Show std. deviations: ☒

Output Format: Select

Result list

- 15:41:23 - Available resultsets
- 15:41:33 - Percent_correct - Ranking
- 15:42:38 - F_measure - Ranking
- 15:42:46 - F_measure - bayes.NaiveBayes "599523120178569765"

Test output

Tester: weka.experiment.PairedCorrectedTTester -G 4,5,6 -D 1 -R 2 -S 0.05 -V -result-matrix "weka.expe"

Analysing: F_measure

Datasets: 1

Resultsets: 5

Confidence: 0.05 (two tailed)

Sorted by: -

Date: 16/10/23, 3:42 pm

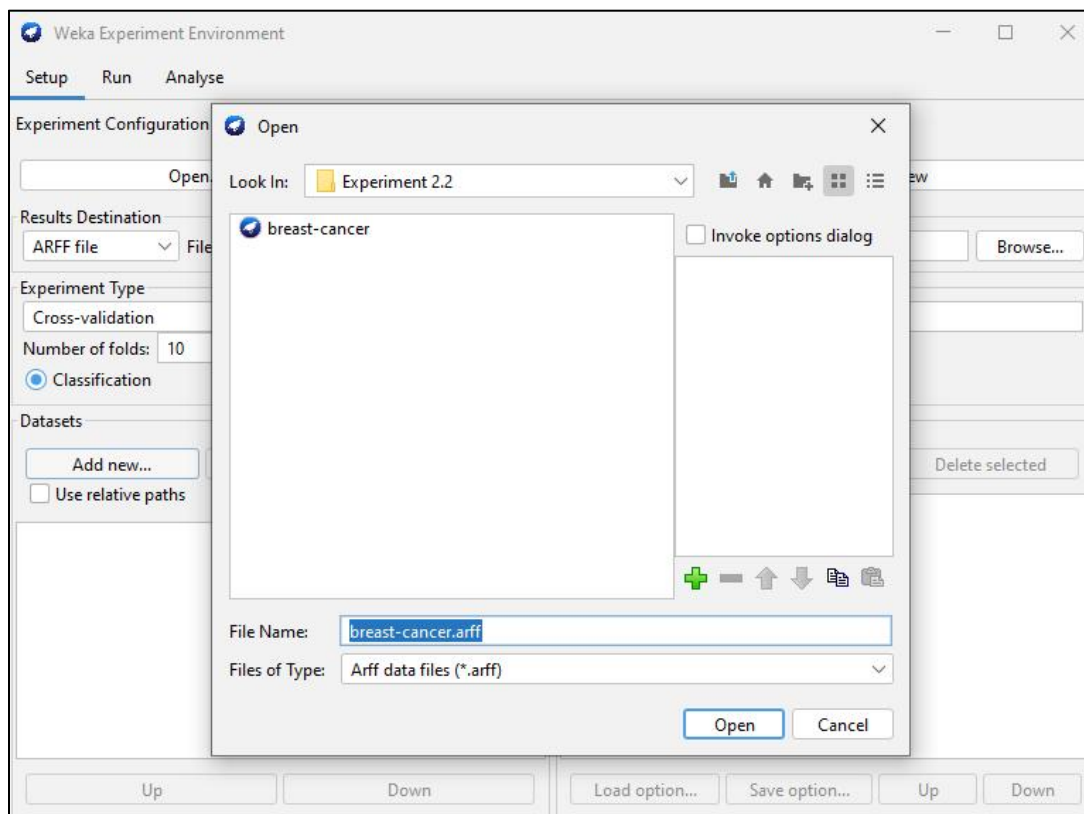
Dataset	(2) bayes.NaiveBa	(1) rules.Zero	(3) trees.Rand	(4) trees.J48	(5) lazy.IBk
breast-cancer	(100) 0.81(0.06)	0.83(0.01)	0.80(0.05)	0.84(0.04)	0.82(0.05)
	(v/ /*)	(0/1/0)	(0/1/0)	(0/1/0)	(0/1/0)

Key:

- (1) rules.ZeroR "48055541465867954"
- (2) bayes.NaiveBayes "5995231201785697655"
- (3) trees.RandomForest "-P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1" 1116839470751428698
- (4) trees.J48 "-C 0.25 -M 2" -217733168393644444
- (5) lazy.IBk "-K 1 -W 0 -A \"weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -"

Step 4: In 'Setup' tab, Click on the 'New' option to initiate the process of adding data.

Step 5: In 'Datasets' section, click on 'Add New' option and select arff file to import.



Weka Experiment Environment

Setup Run Analyse

Experiment Configuration

Open

Look In: Experiment 2.2

breast-cancer

Invoke options dialog

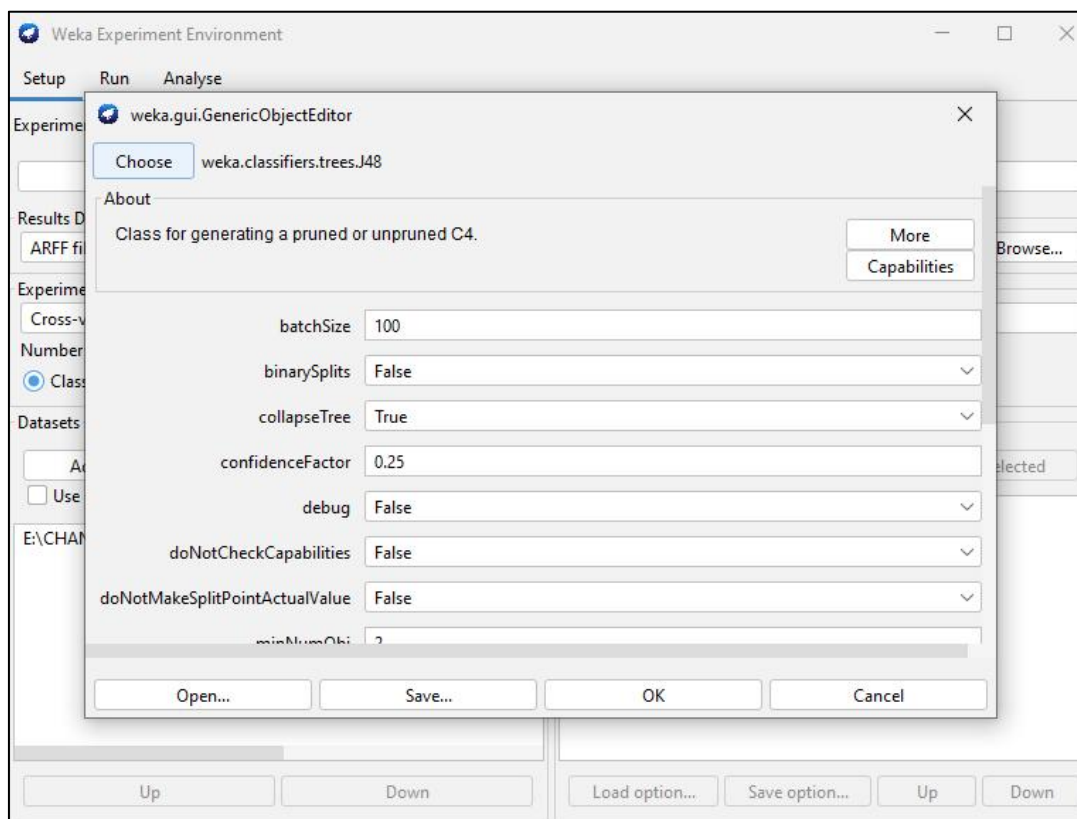
File Name: breast-cancer.arff

Files of Type: Arff data files (*.arff)

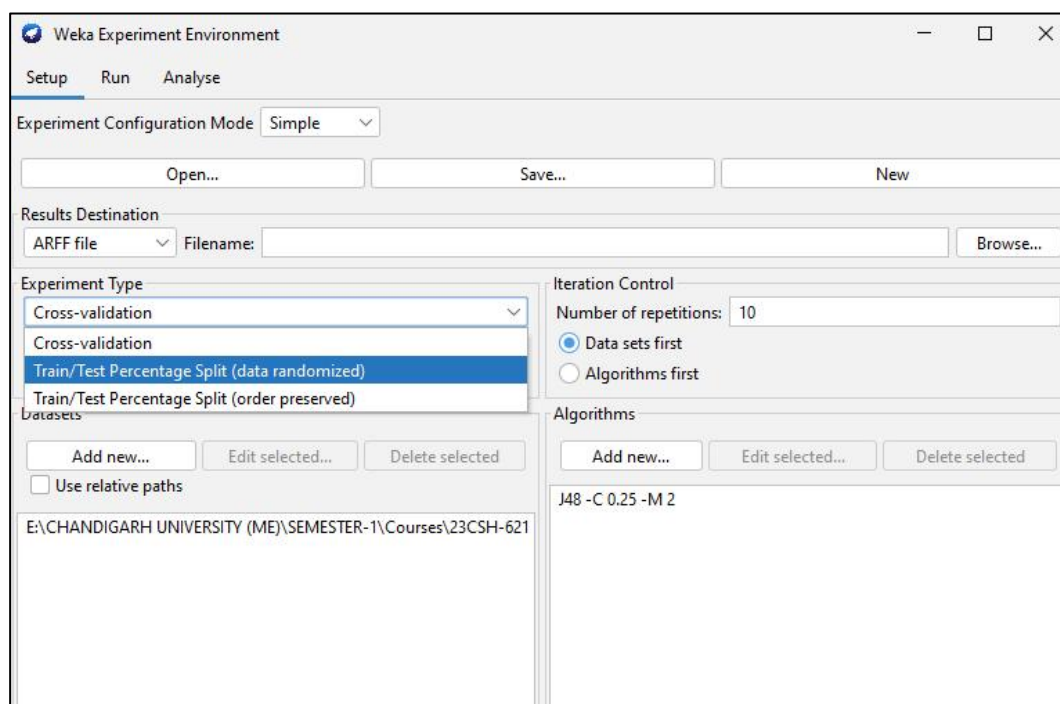
Open Cancel

Up Down Load option... Save option... Up Down

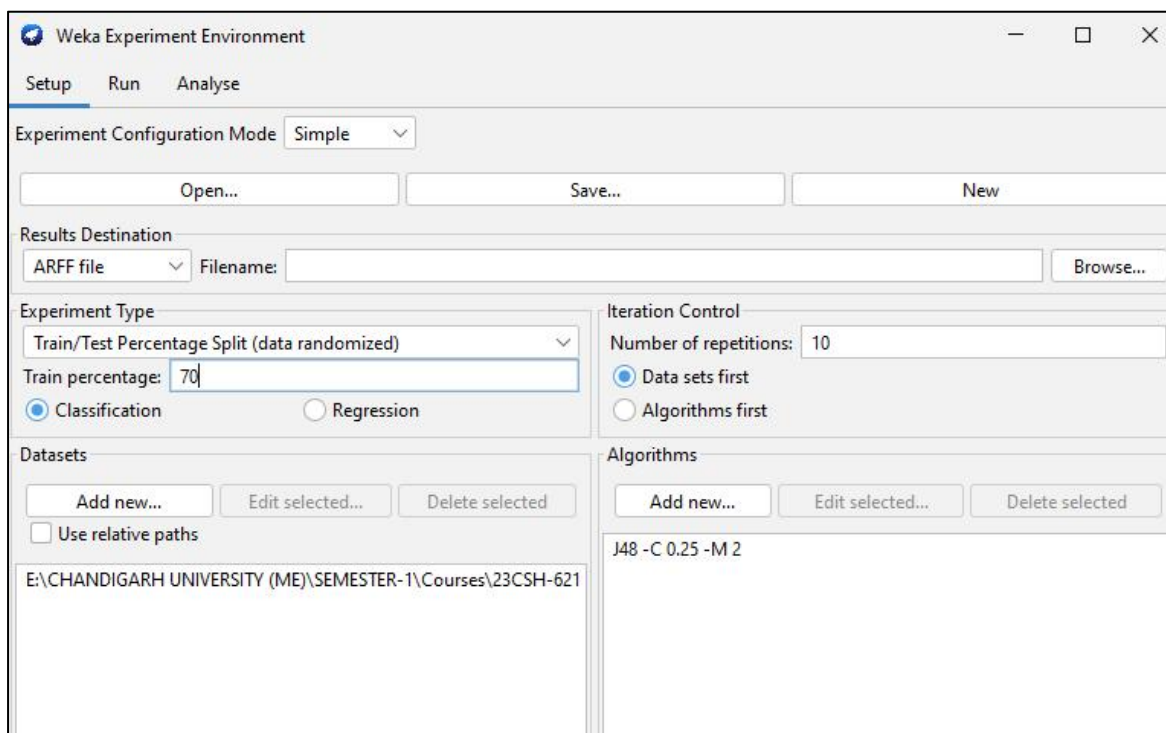
Step 6: In the **Algorithms** section, click on choose add new> choose>>**J48**.



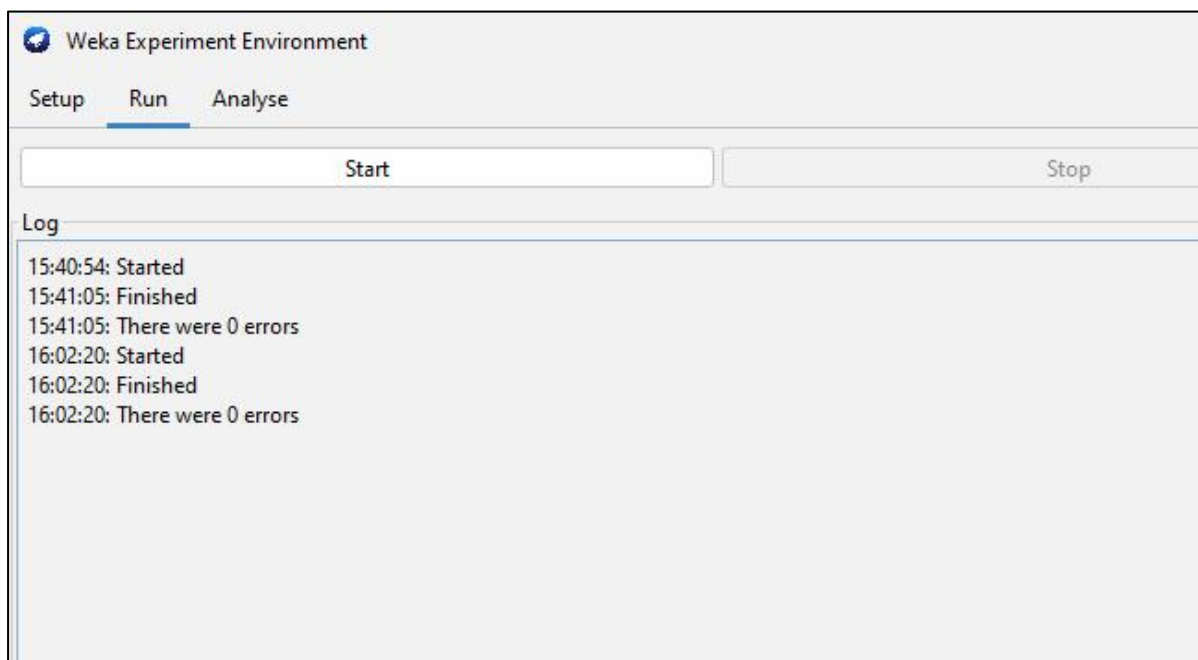
Step 7: In 'Experiment Type' section, click on 'Train/Test Percentage Split (data randomized)'.



Step 8: In 'Train Percentage' option, add the value 70 for 70/30 Split.

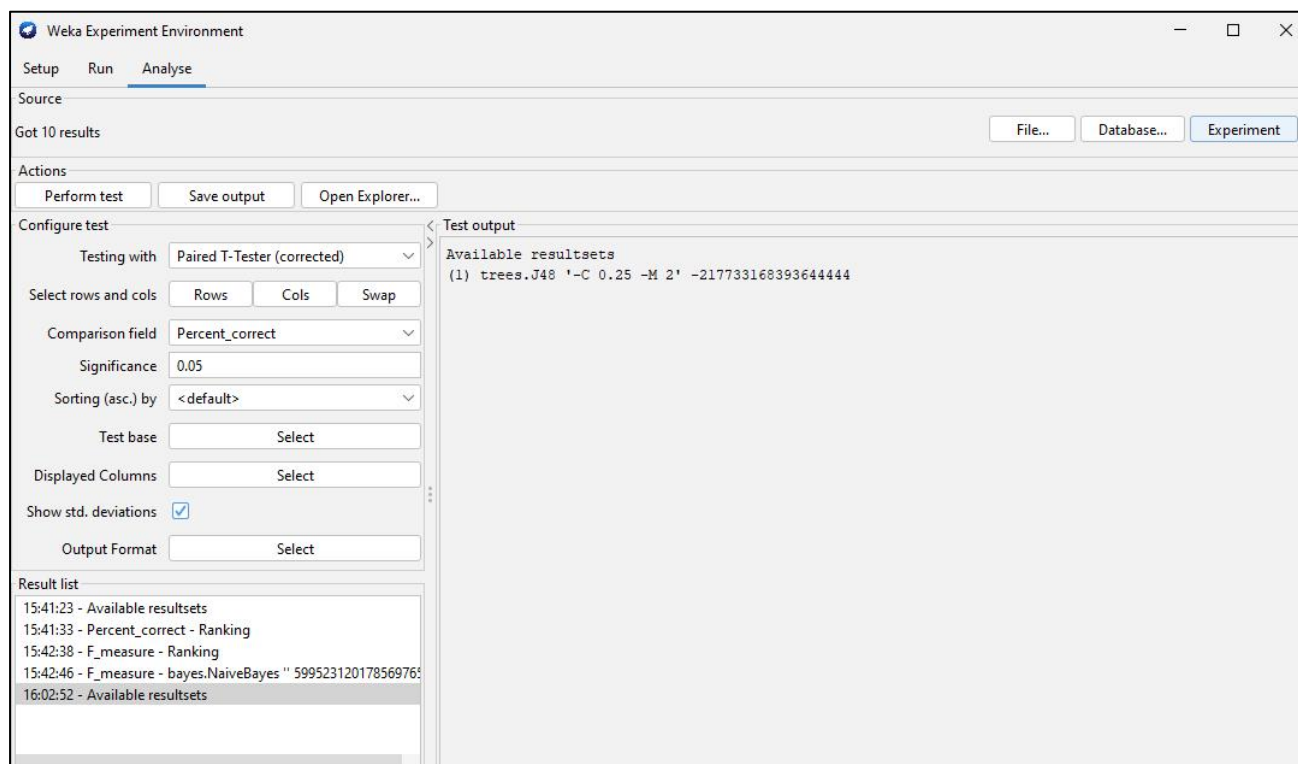


Step 9: Click on 'Run' tab >> click 'Start'.

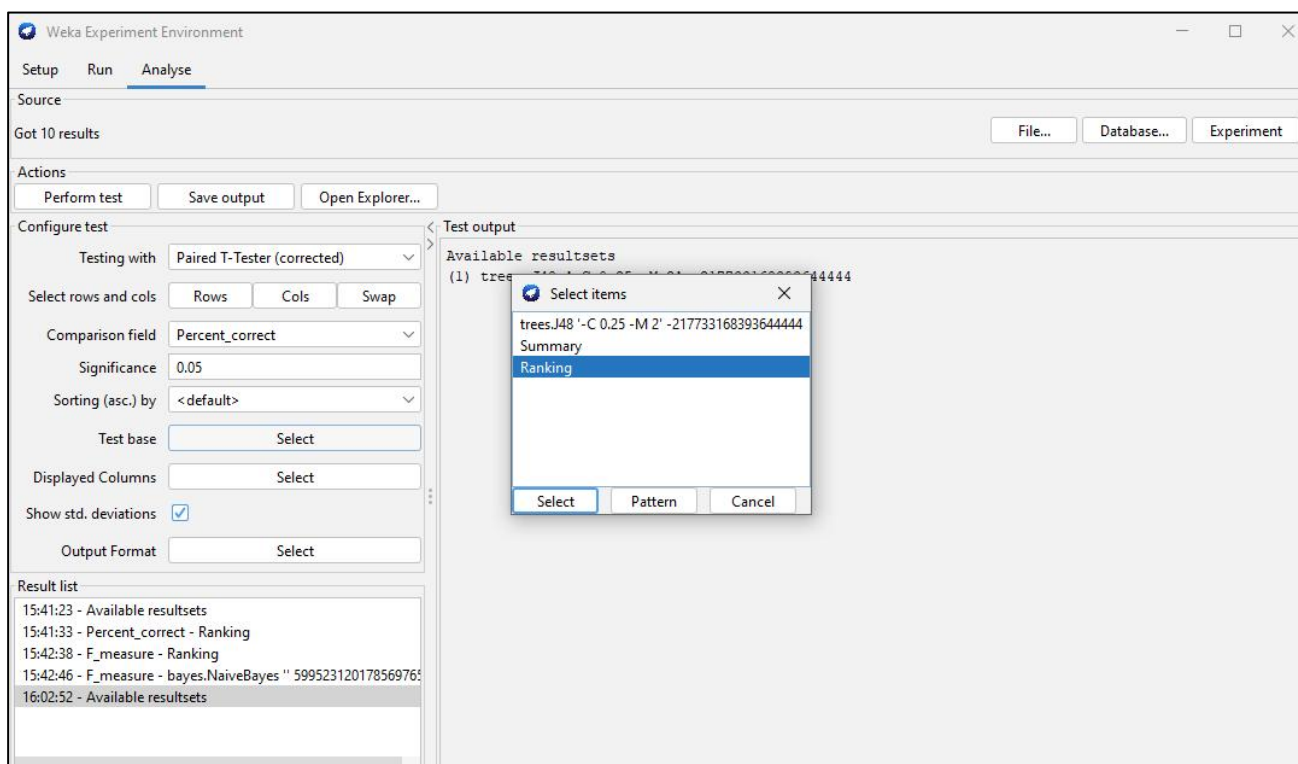


Step 10: After the execution is completed in 'Run' tab >>click on 'Analyse' tab.

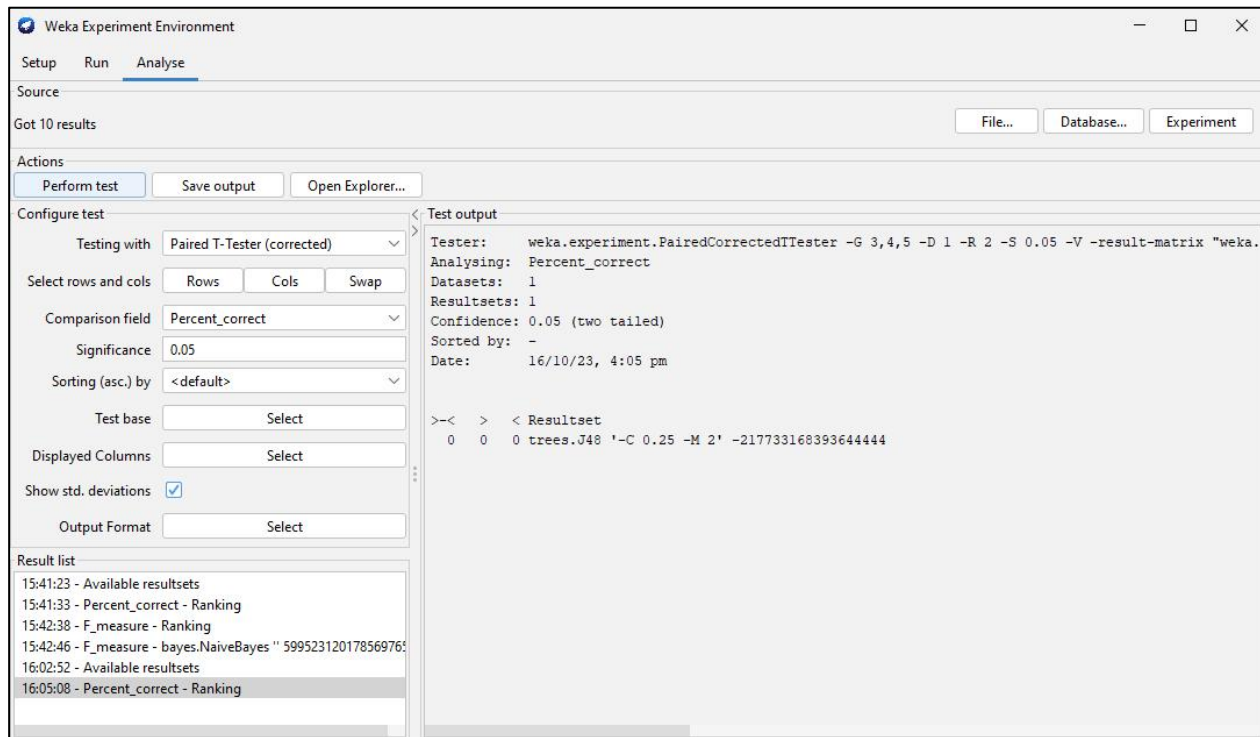
Step 11: In the ‘Analyse’ tab >> click on ‘Experiment’ tab >> all the options will become active.



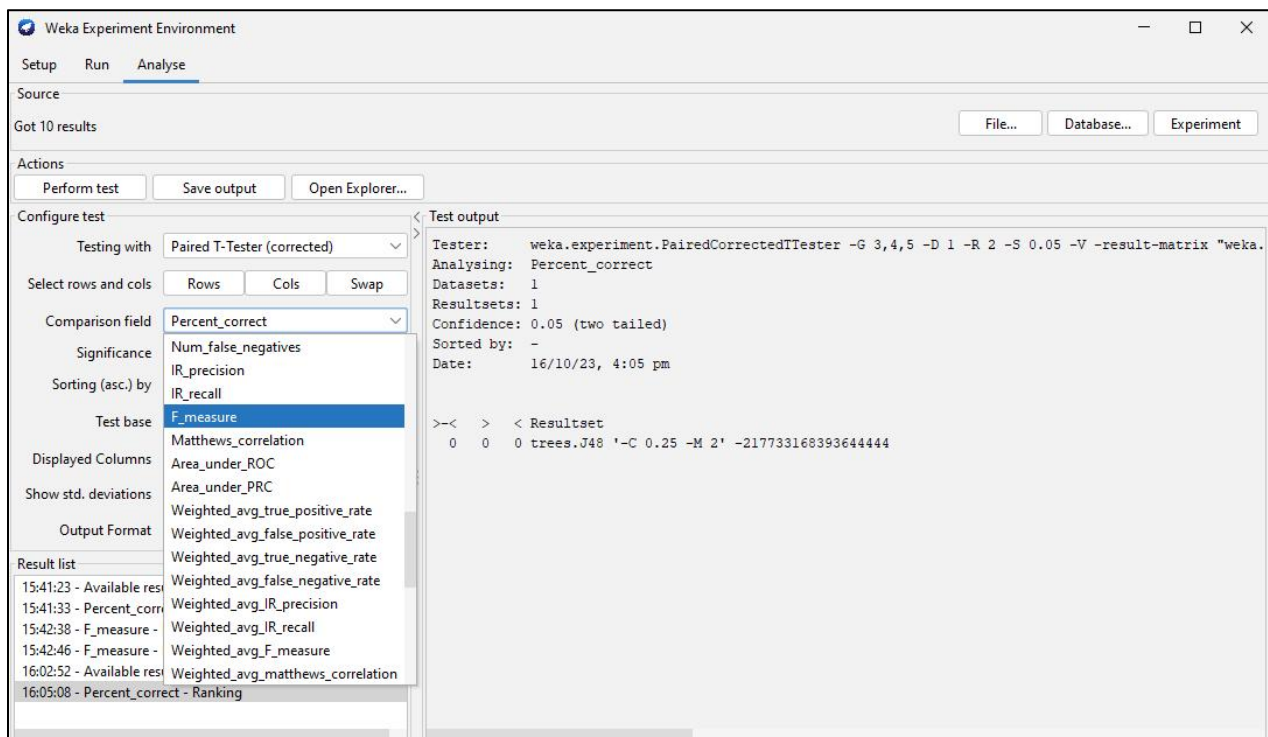
Step 12: In **Configure test** >> click on **Test base**>> and select **Ranking**>>click on **Perform test**.

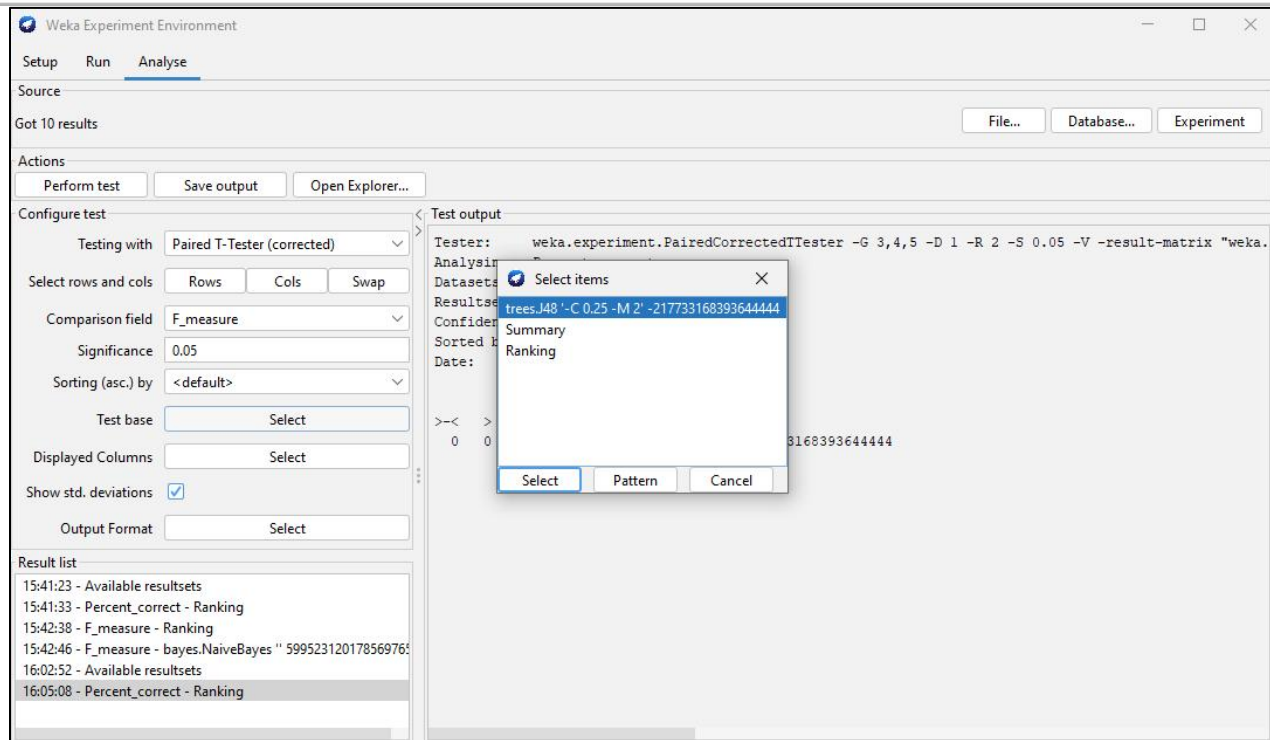


Step 13: In Test area ,it will show algorithms which have performed better and will display ranking.



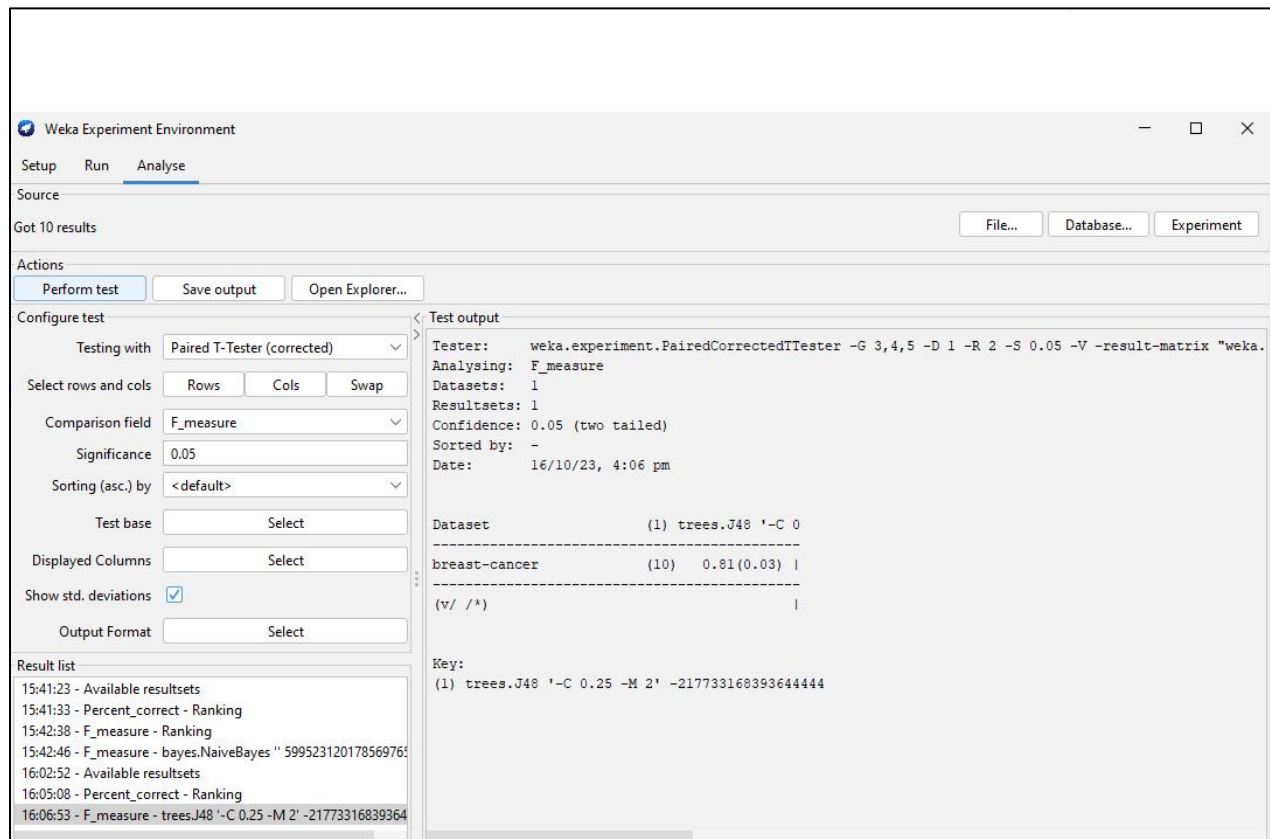
Step 14: To check the **f-measure** and std_dev of an particular algorithm >>click on std_tab >>then select the algorithm which you want to test >> select **f-measure** from the **comparison_field** tab.





The screenshot shows the Weka Experiment Environment interface. The 'Analyse' tab is selected. In the 'Actions' section, the 'Perform test' button is highlighted. The 'Configure test' section shows 'Paired T-Tester (corrected)' selected for testing, with 'F_measure' as the comparison field and '0.05' as the significance level. The 'Test output' section shows the command: `weka.experiment.PairedCorrectedTTester -G 3,4,5 -D 1 -R 2 -S 0.05 -V -result-matrix "weka.`

Step 15: Click on Perform test.



The screenshot shows the Weka Experiment Environment interface after the 'Perform test' action. The 'Test output' section displays the following information:

```

Tester: weka.experiment.PairedCorrectedTTester -G 3,4,5 -D 1 -R 2 -S 0.05 -V -result-matrix "weka.
Analysing: F_measure
Datasets: 1
Resultsets: 1
Confidence: 0.05 (two tailed)
Sorted by: -
Date: 16/10/23, 4:06 pm

Dataset (1) trees.J48 '-C 0
-----
breast-cancer (10) 0.81 (0.03) |
-----
(v/ /*) |

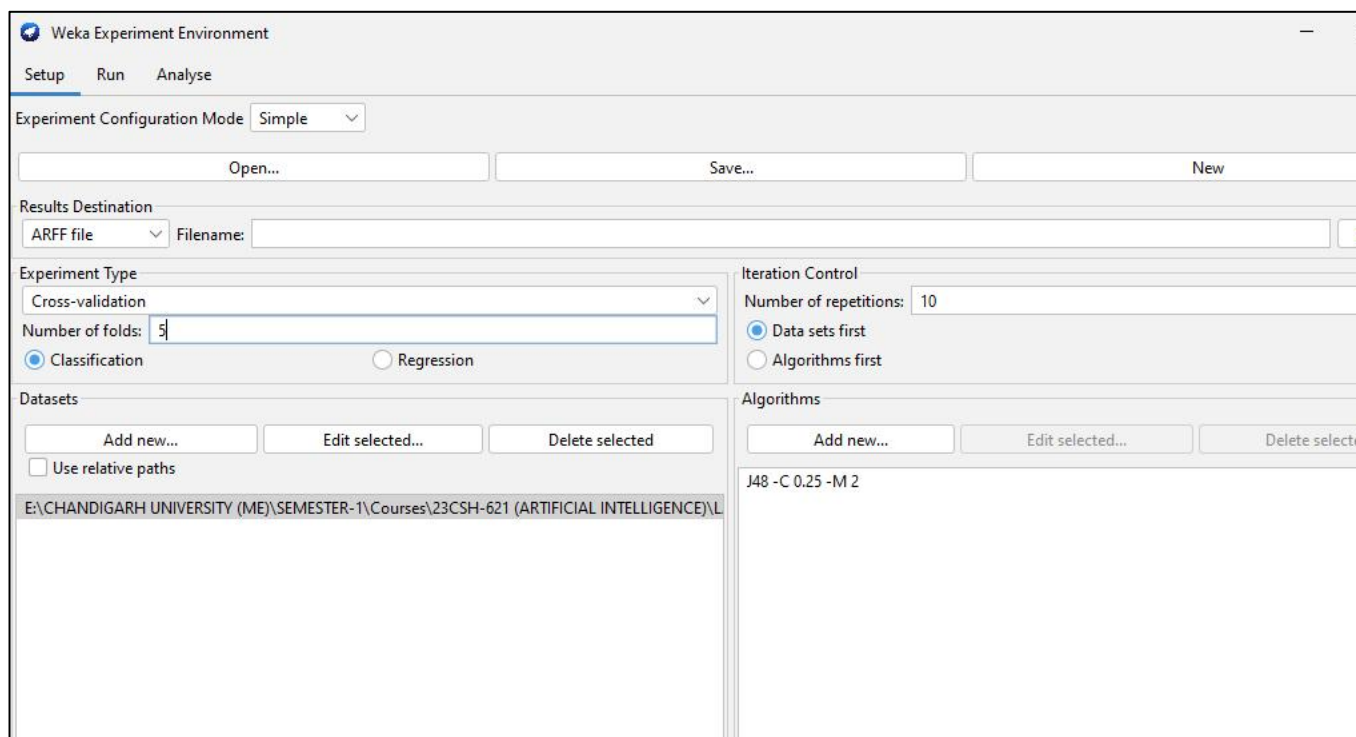
Key:
(1) trees.J48 '-C 0.25 -M 2' -217733168393644444

```

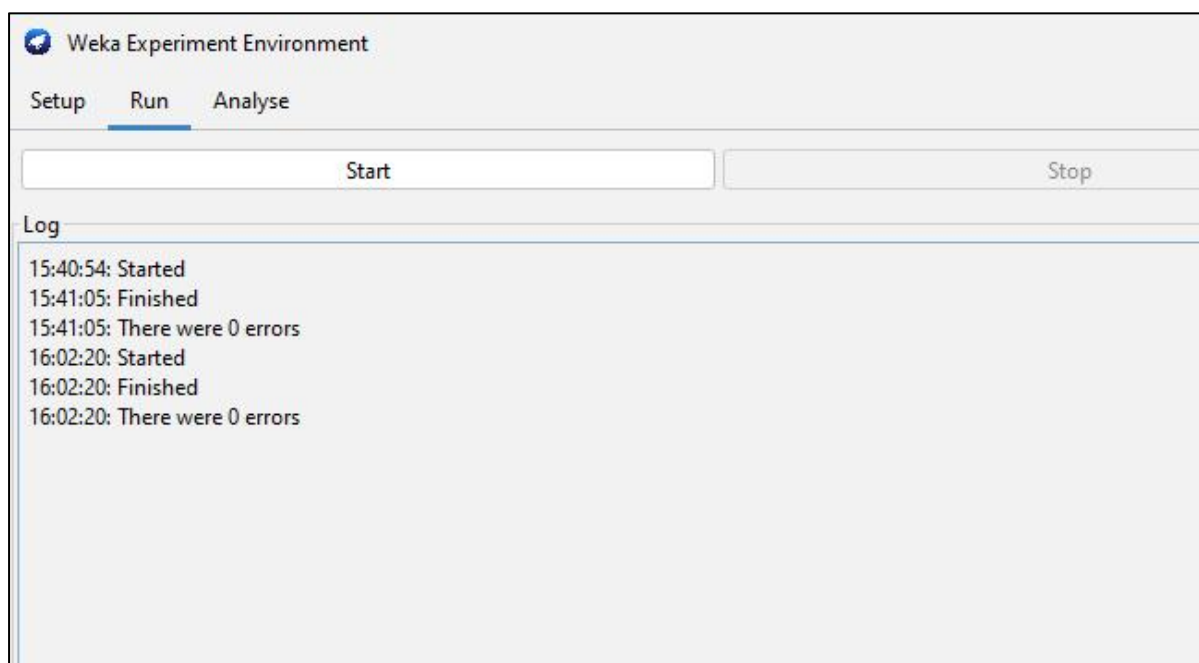
The 'Result list' section shows the following results:

- 15:41:23 - Available resultsets
- 15:41:33 - Percent_correct - Ranking
- 15:42:38 - F_measure - Ranking
- 15:42:46 - F_measure - bayes.NaiveBayes " 59952312017856976
- 16:02:52 - Available resultsets
- 16:05:08 - Percent_correct - Ranking
- 16:06:53 - F_measure - trees.J48 '-C 0.25 -M 2' -21773316839364

Step 16: In 'Setup' tab >> 'Experiment Type', click on 'Cross-validation' and set no of folds as 5.

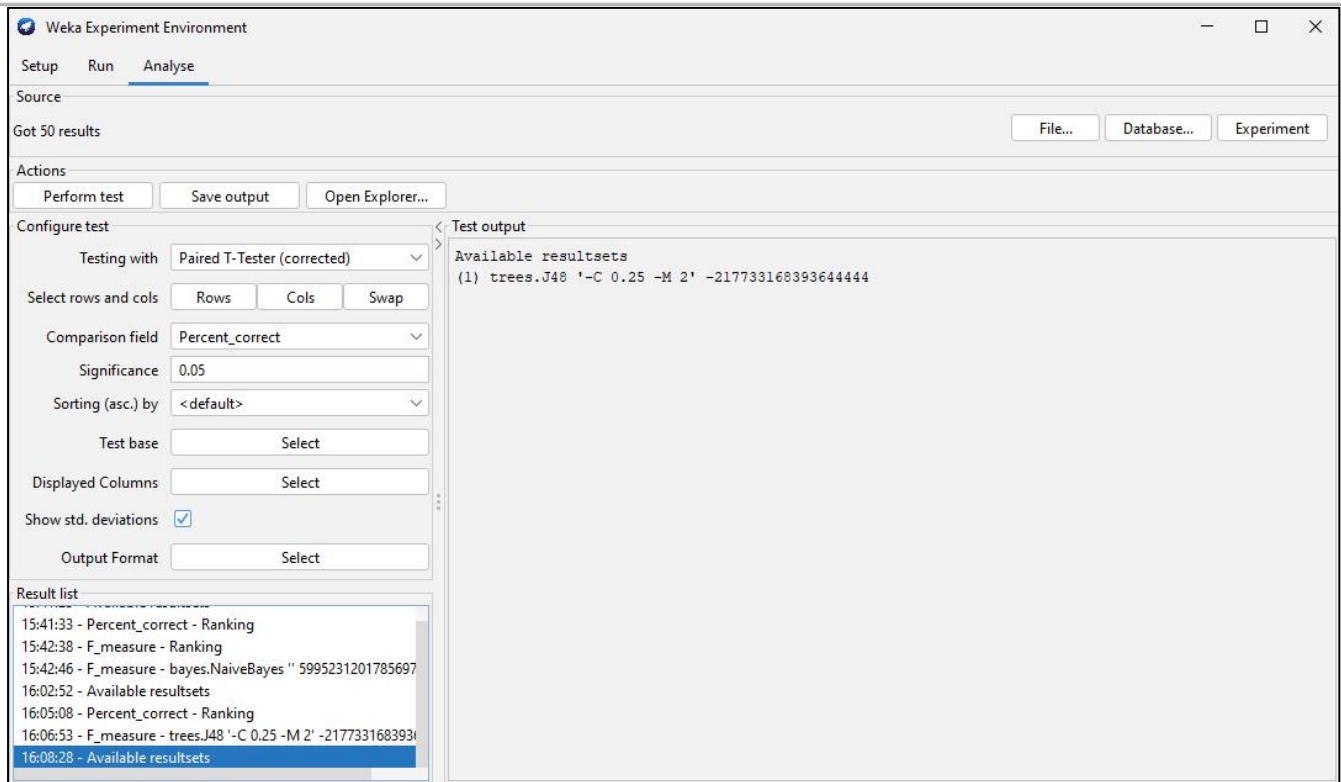


Step 17: Click on 'Run' tab >> click 'Start'.



Step 18: After the execution is completed in 'Run' tab >> click on 'Analyse' tab.

Step 19: In the 'Analyse' tab >> click on 'Experiment' tab >> all the options will become active.



Weka Experiment Environment

Setup Run **Analyse**

Source

Got 50 results

File... Database... Experiment

Actions

Perform test Save output Open Explorer...

Configure test

Testing with: Paired T-Tester (corrected)

Select rows and cols: Rows Cols Swap

Comparison field: Percent_correct

Significance: 0.05

Sorting (asc.) by: <default>

Test base: Select

Displayed Columns: Select

Show std. deviations: ☒

Output Format: Select

Test output

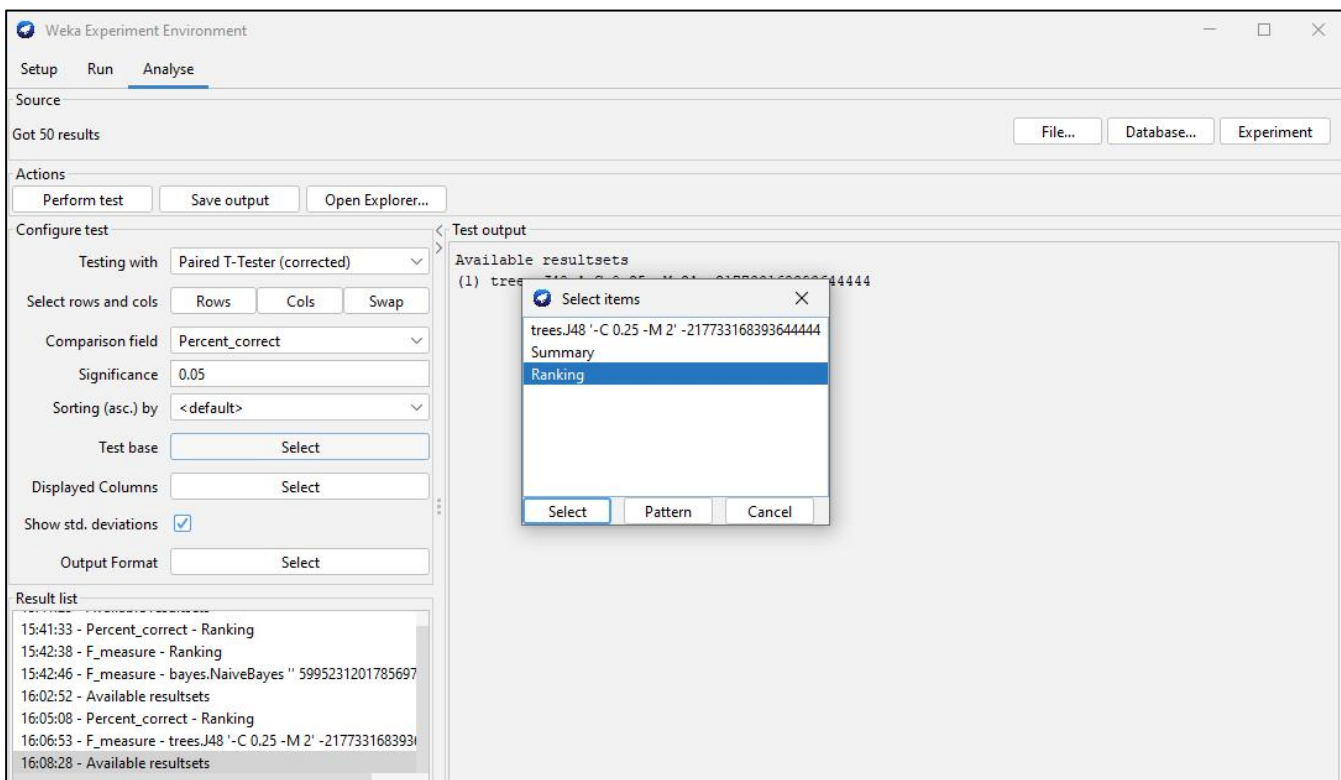
Available resultsets

(1) trees.J48 '-C 0.25 -M 2' -217733168393644444

Result list

- 15:41:33 - Percent_correct - Ranking
- 15:42:38 - F_measure - Ranking
- 15:42:46 - F_measure - bayes.NaiveBayes " 5995231201785697
- 16:02:52 - Available resultsets
- 16:05:08 - Percent_correct - Ranking
- 16:06:53 - F_measure - trees.J48 '-C 0.25 -M 2' -217733168393644444
- 16:08:28 - Available resultsets

Step 20: In Configure test >> click on Test base>> and select Ranking>>click on Perform test.



Weka Experiment Environment

Setup Run **Analyse**

Source

Got 50 results

File... Database... Experiment

Actions

Perform test Save output Open Explorer...

Configure test

Testing with: Paired T-Tester (corrected)

Select rows and cols: Rows Cols Swap

Comparison field: Percent_correct

Significance: 0.05

Sorting (asc.) by: <default>

Test base: Select

Displayed Columns: Select

Show std. deviations: ☒

Output Format: Select

Test output

Available resultsets

(1) trees.J48 '-C 0.25 -M 2' -217733168393644444

Select items

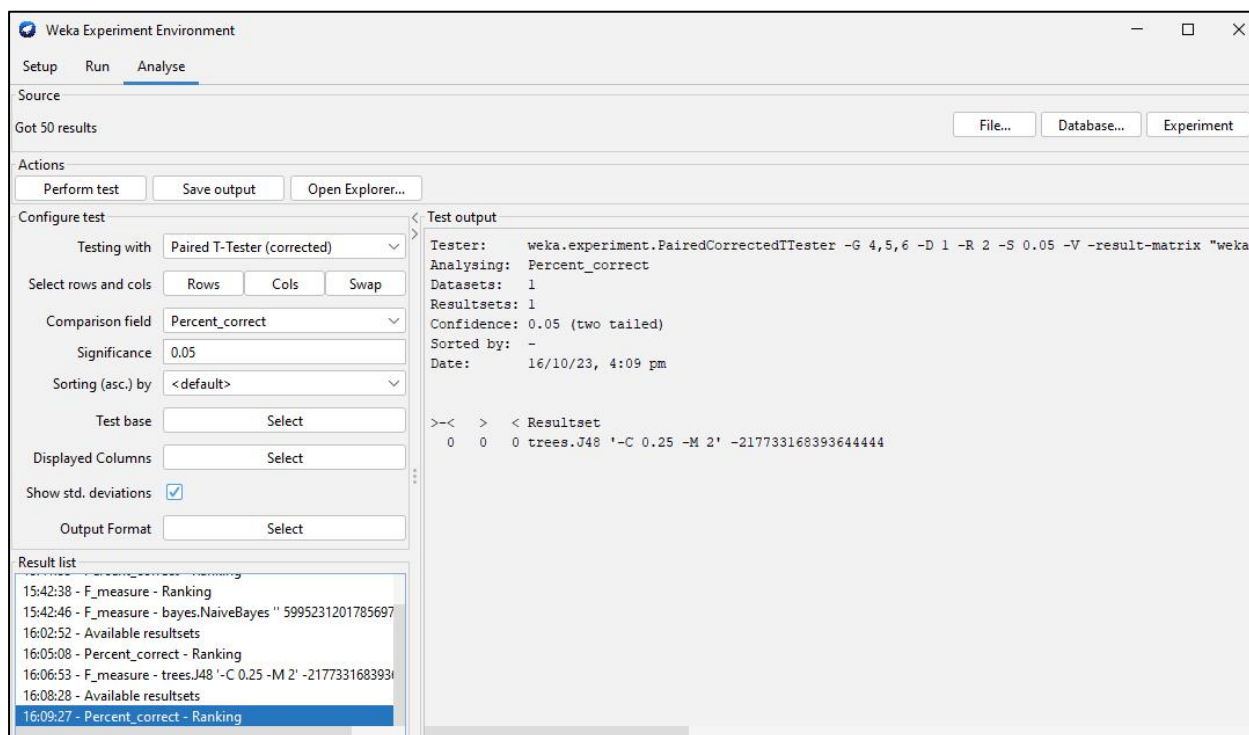
- trees.J48 '-C 0.25 -M 2' -217733168393644444
- Summary
- Ranking**

Select Pattern Cancel

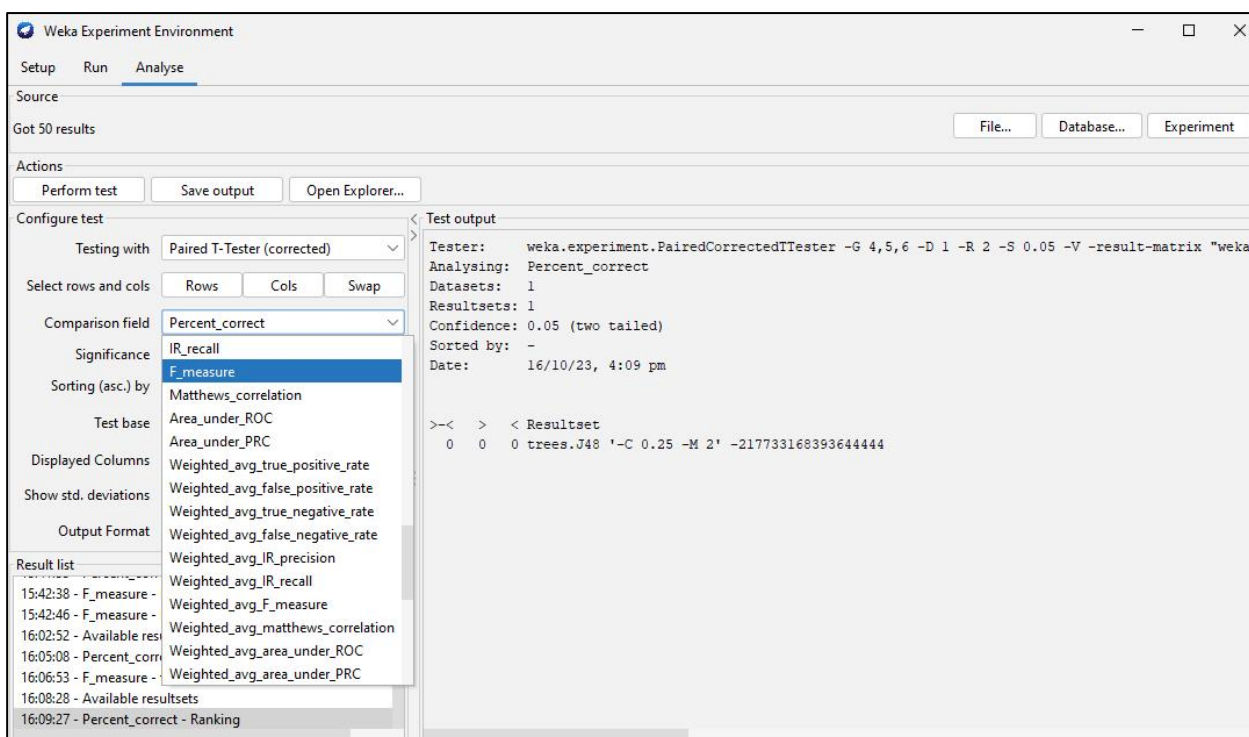
Result list

- 15:41:33 - Percent_correct - Ranking
- 15:42:38 - F_measure - Ranking
- 15:42:46 - F_measure - bayes.NaiveBayes " 5995231201785697
- 16:02:52 - Available resultsets
- 16:05:08 - Percent_correct - Ranking
- 16:06:53 - F_measure - trees.J48 '-C 0.25 -M 2' -217733168393644444
- 16:08:28 - Available resultsets

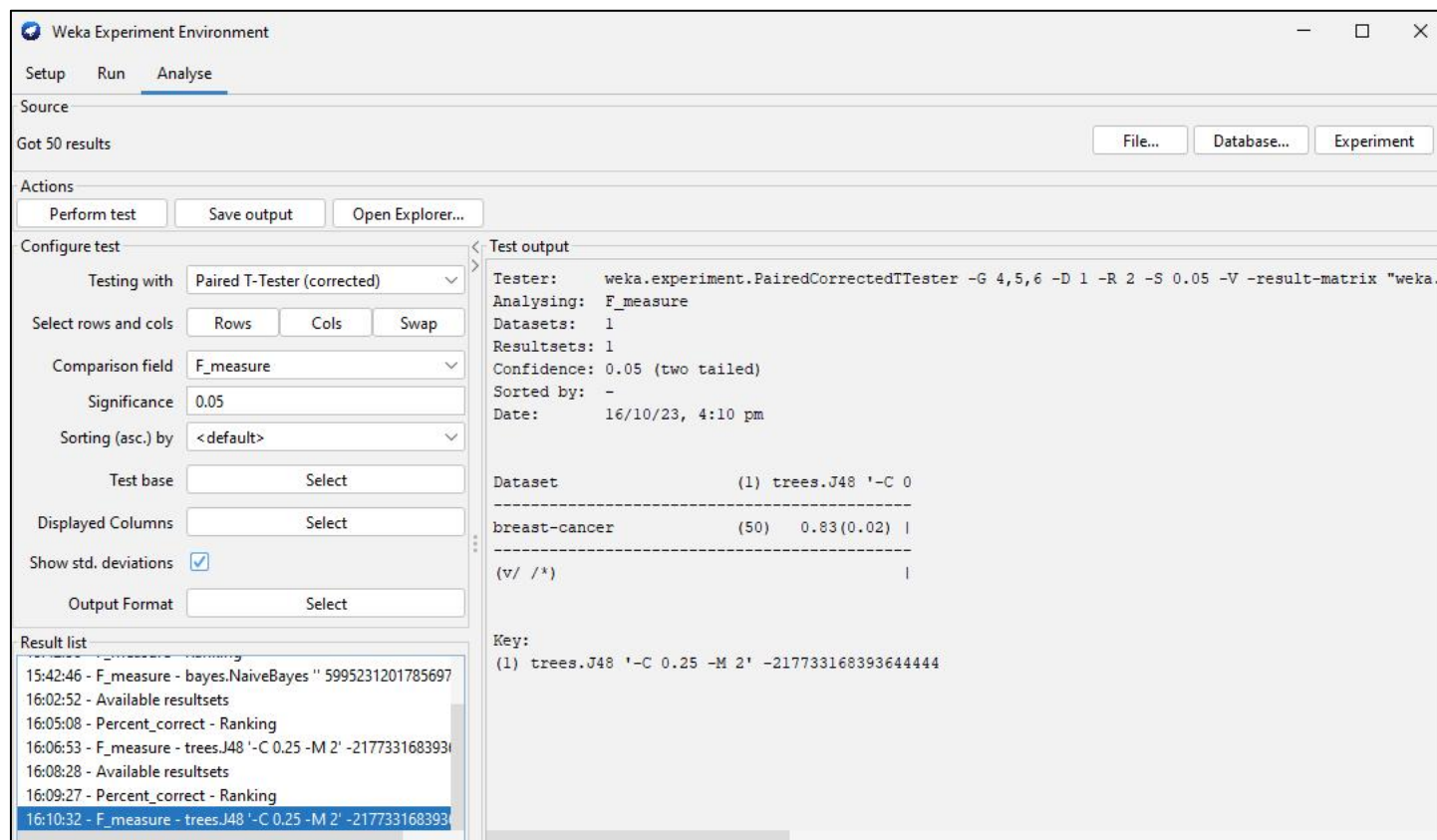
Step 21: In Test area ,it will show algorithms which have performed better and will display ranking.



Step 22: To check the **f-measure** and **std_dev** of an particular algorithm >>click on **std_tab** >>then select the algorithm which you want to test >> select **f-measure** from the **comparison_field** tab.



Step 23: Click on Perform test.



The screenshot shows the Weka Experiment Environment window. The 'Analyse' tab is selected. Under the 'Actions' section, the 'Perform test' button is highlighted. The 'Configure test' section shows the following settings:

- Testing with: Paired T-Tester (corrected)
- Select rows and cols: Rows, Cols, Swap
- Comparison field: F_measure
- Significance: 0.05
- Sorting (asc.) by: <default>
- Test base: Select
- Displayed Columns: Select
- Show std. deviations: ☒
- Output Format: Select

The 'Test output' section displays the following information:

```

Tester:   weka.experiment.PairedCorrectedTTester -G 4,5,6 -D 1 -R 2 -S 0.05 -V -result-matrix "weka.
Analysing: F_measure
Datasets: 1
Resultsets: 1
Confidence: 0.05 (two tailed)
Sorted by: -
Date:     16/10/23, 4:10 pm

Dataset      (1) trees.J48 '-C 0
-----
breast-cancer (50)  0.83(0.02) |
-----
(v/ /*)                                           |

Key:
(1) trees.J48 '-C 0.25 -M 2' -217733168393644444
  
```

The 'Result list' section shows a list of results, with the following entry highlighted:

```

16:10:32 - F_measure - trees.J48 '-C 0.25 -M 2' -217733168393644444
  
```

Learning outcomes (What I have learnt):

1. I learnt about the WEKA Tool and its applications.
2. I learnt about different machine learning classifiers in WEKA Tool.
3. I learnt about how to find the best prediction model in WEKA.
4. I learnt about the Train/Test Percentage Split Method to check model performance.
5. I learnt about the Cross-validation method to check model performance.