

Green University of Bangladesh



Project Title

Mini Project on Weka Using Breast Cancer Dataset

Course Code: CSE 404

Course Title: Artificial Intelligent Lab

Submitted To:

MR. JARGIS AHMED

Lecturer

Department of Computer Science & Engineering

Submitted By:

Md. Imam Hasan

ID: 182002010

Section: DA

Department of Computer Science and Engineering

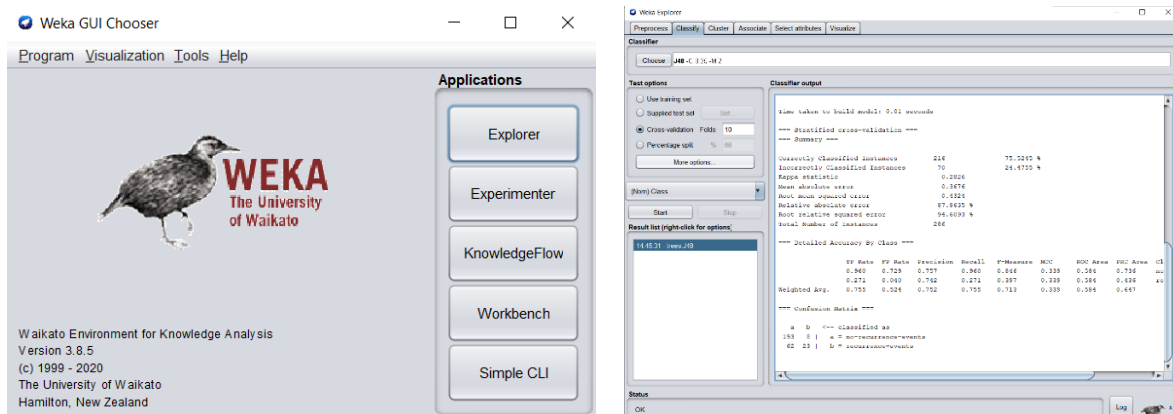
Date of Submission: 12th September 2021

Project Name: Mini Project on Weka Using Breast Cancer Dataset.

Introduction→

Weka is a collection of machine learning algorithms for solving real world data mining problems. It is written in Java and runs on almost any platform. Algorithms can be applied directly to a dataset or told from your own Java code.

- ✓ Open the Weka Application.
- ✓ Click the “Explorer” button to launch the Weka Explorer.
- ✓ Click the “Open file...” button to open a data set and double click on the “data” directory.
- ✓ Select the “**breast-cancer.arff**” file to load the Iris dataset.
- ✓ Click the “Classify” tab. This is the area for running algorithms against a loaded dataset in Weka.



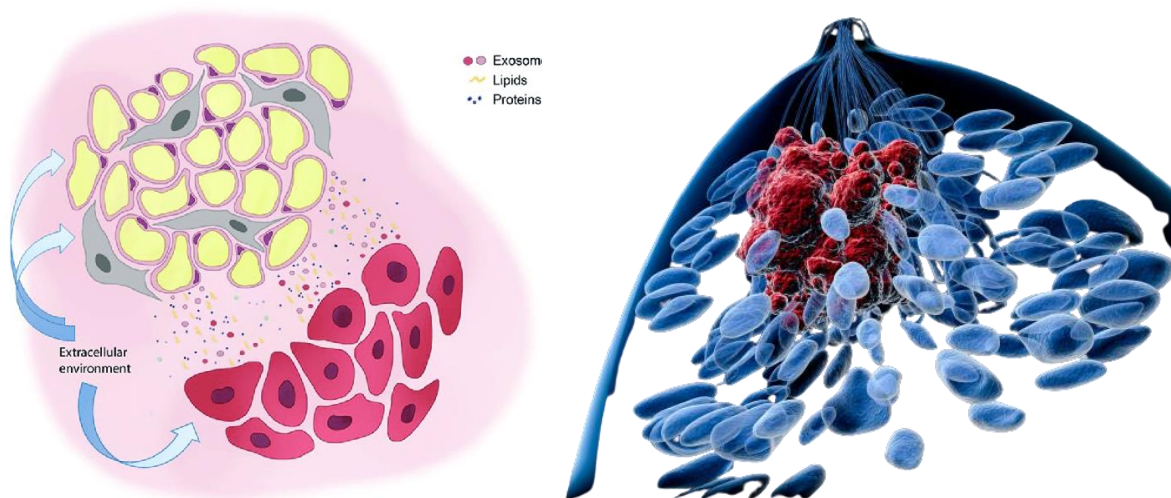
For Decision Tree:

- ✓ Click the “Choose” button in the “Classifier” section and click on “trees” and click on the “**J48**” algorithm.
- ✓ Click the “Start” button to run the algorithm.
- ✓ After running the J48 algorithm, we can note the results in the “Classifier output” section.
- ✓ The algorithm was run with 10-fold cross-validation and note the Classification Accuracy. WE can see that the model achieved a result of 216/286 correct or 75.5245%, which seems a lot better than the baseline of 94.6093%.

For Neural Network:

- ✓ Click the “Choose” button in the “Classifier” section and click on “functions” and click on the “**MultilayerPerceptron**” algorithm.
- ✓ After running the “MultilayerPerceptron” algorithm, we can note the results in the “Classifier output” section.
- ✓ The algorithm was run with 10-fold cross-validation, Hidden layers 'a', Training times 500 and note the Classification Accuracy. WE can see that the model achieved a result of 185/286 correct or 64.6853%, which seems a lot better than the baseline of 118.654%.

Breast Cancer is a disease in which cells in the breast grow out of control. It is caused when the DNA in breast cells mutate or change, disabling specific functions that control cell growth and division.

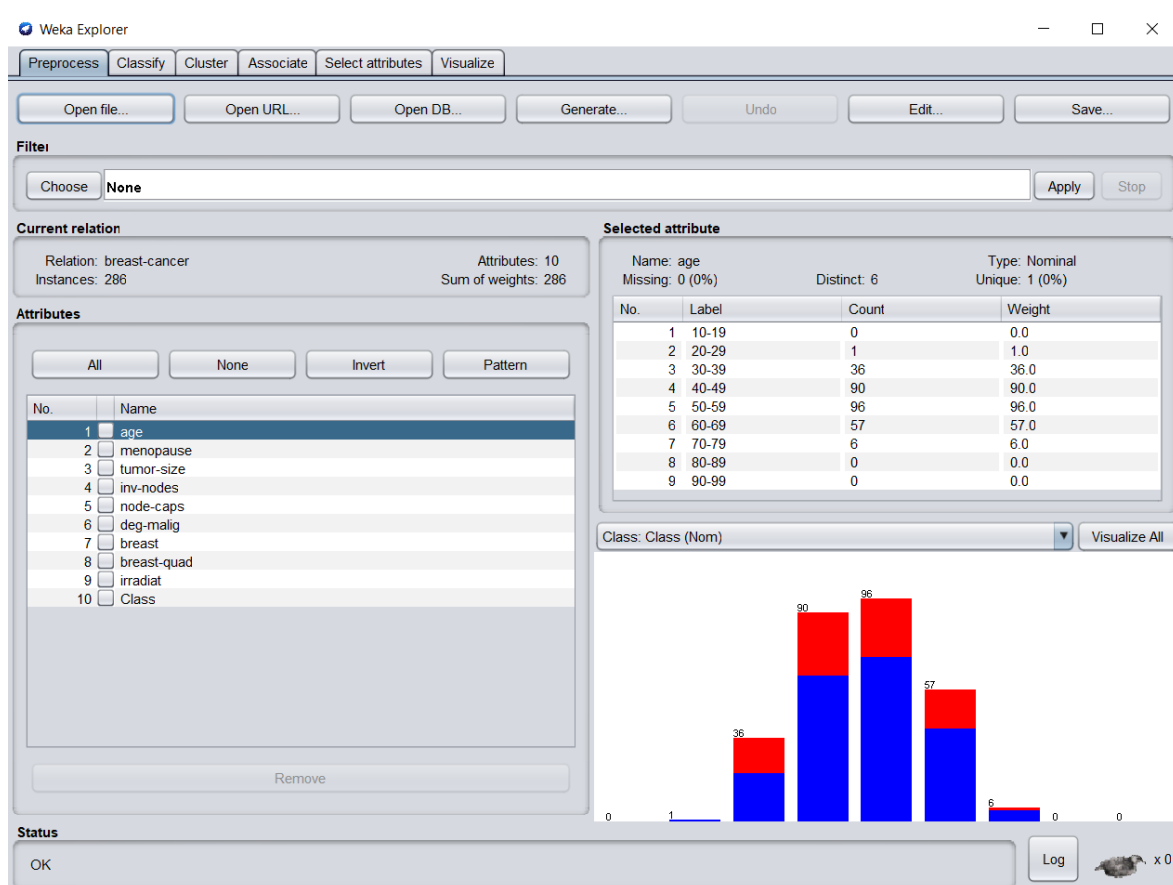


Breast cancer remains a leading dreadful cancer of women in Bangladesh. It has become a hidden burden which accounts **69%** death of women According to the latest **WHO** data published in 2018, Breast cancer accounted for **12764** new cases and deaths in Bangladesh reached **7,667** or 0.99% of total deaths. The age adjusted Death Rate is 11.47 per 100,000 of population ranks Bangladesh #139 in the world.

File Description→

This breast cancer domain was obtained from the University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia that has repeatedly appeared in the machine learning literature.

In breast cancer the data set includes 201 instances of one class and 85 instances of another class. The instances are described by 9+ attributes, some of which are linear, and some are nominal. The features are computed here from a digitized image of a fine needle aspirate (FNA) of a breast mass.



Best predictive accuracy obtained using one separating plane in the 3-D space of Worst Area, Worst Smoothness and Mean Texture. Estimated accuracy 97.5% using repeated 10-fold cross validations. Classifier has correctly diagnosed 176 consecutive new patients as of November 1995.

Using this model, a predictor with a 71,134 % effectiveness is obtained to estimate or not the recurrence of breast cancer.

Accuracy Vs Fold (Decision Tree)→

When the fold number is 2 then the accuracy is 70.2797% which is mentioned in the graph. After a certain period of time, increasing the fold number increases the accuracy. But after some time, the accuracy decreases instead of increasing the fold number.

Thus, after some time, the accuracy started to increase and decrease. But even if the fold number increases after a certain period of time, the accuracy remains the same.

Fold No	Accuracy
2	70.2797%
3	70.979%
4	69.9301%
5	74.1259%
6	73.7762%
7	73.0769%
8	74.1259%
9	75.5245%
10	75.5245%
11	73.0769%
12	74.1259%

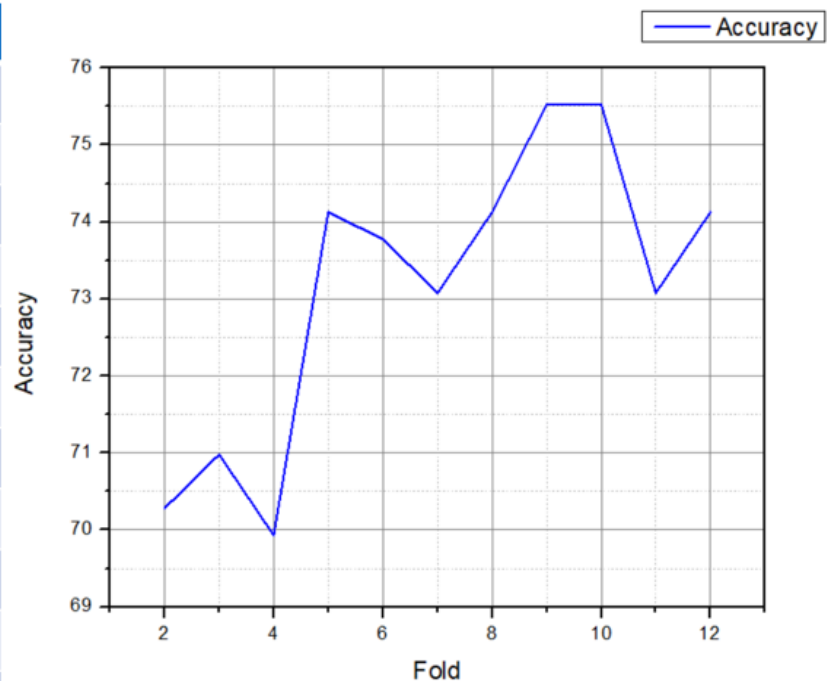


Fig: Accuracy Vs Fold

In the table, we can see when the fold numbers are 9 and 10, then the accuracy is 75.5245%. But after going to fold number 11, the accuracy has decreased to 73.0769%.

Accuracy Vs Hidden Layers (Neural network)→

When the fold number is 10 for all the hidden layers and for a specific one of hidden layers such as 6, the accuracy is 66.4336% which is mentioned in the graph. After a certain period of time, with the increasing number of hidden layers the accuracy is started to decrease.

Thus, accuracy increases a lot as the hidden layers grows. Then the accuracy decreases again.

Hidden Layers	Accuracy
6	66.4336%
7	65.3846%
8	66.0839%
9	66.4336%
10	69.5804%
11	63.2867%
12	66.0839%

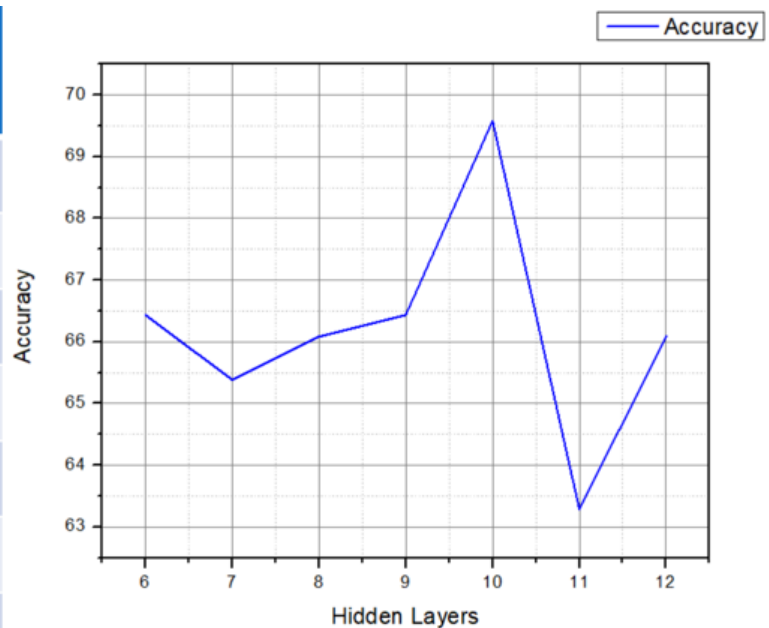


Fig: Accuracy Vs Hidden Layers

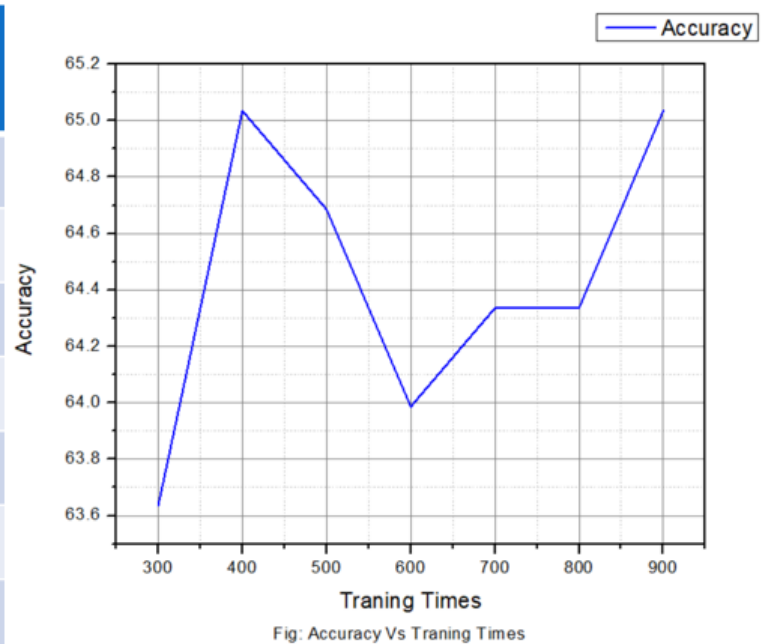
In the table, we can see when the hidden layers is 10, then the accuracy is maximum (69.5804%). But after going to the hidden layers is 11, the accuracy is minimum (decreased to 63.2867 %).

Accuracy Vs Training Times (Neural network)→

When the fold number is 10 for all the training times and for a specific one of training times such as 300, the accuracy is 63.6364% which is mentioned in the graph. After a certain period of time, with the increasing number of training times the accuracy is started increases.

Thus, accuracy decreases a lot as the training times grows. Then the accuracy increases again.

Training Times	Accuracy
300	63.6364%
400	65.035%
500	64.6853%
600	63.986%
700	64.33574%
800	64.3357%
900	65.035%



In the table, we can see when the training times are 400 and 900, then the accuracy remain same. Besides in this position the accuracy is maximum (65.035%).

References→

- ✓ The documents of Weka related things which is provided by our honorable course teacher **Mr. Jargis Ahmed Sir**.
- ✓ <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29>
- ✓ <https://revistas.utb.edu.ec/index.php/sr/article/download/405/294/>
- ✓ Google and YouTube

>>>THE END<<<