|  |  |
| --- | --- |
| Name: | Prerna Sunil Jadhav |
| Sap  Id: | 60004220127 |
| Class: | T. Y. B.Tech (Computer Engineering) |
| Course: | Data Mining and Warehouse Laboratory |
| Course Code: | DJ19CEL501 |
| Experiment  No.: | 01 |

**AIM:** Perform data Pre-processing task using Weka data mining tool

**THEORY**:

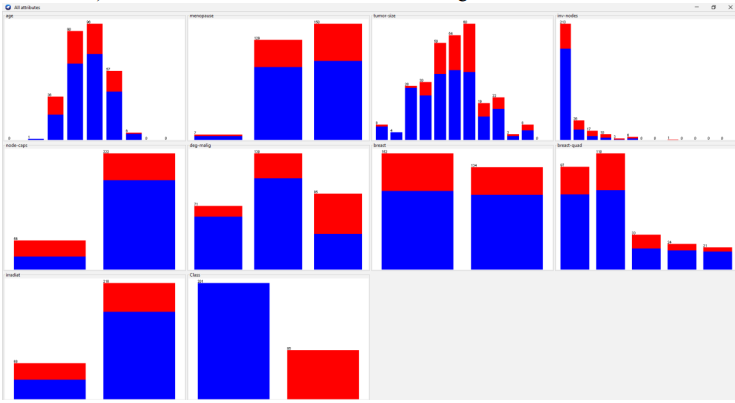
WEKA - an open source software provides tools for data preprocessing, implementation of several Machine Learning algorithms, and visualization tools so that you can develop machine learning techniques and apply them to real-world data mining problems

**TASKS PERFORMED THROUGH WEKA:**

* PREPROCESSING:

Procedure:

1. Visualize All: Select this button to visualize histograms of all attributes.



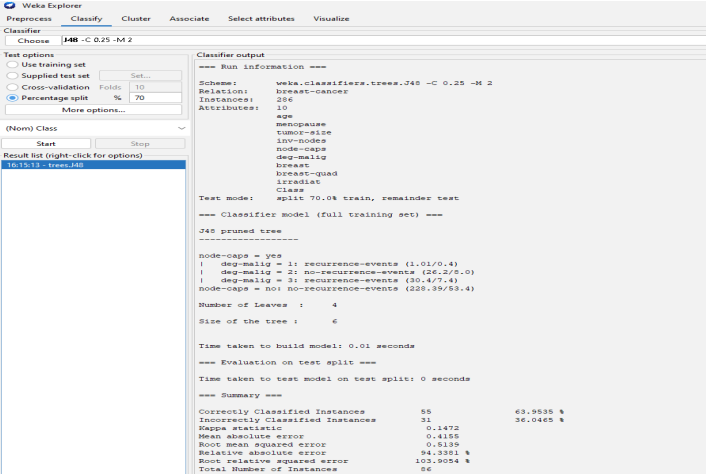
1. Filter: Choose Discretization under Unsupervised and Supervised methods. Observe the discretization and the outliers.
2. IQR: Observe the IQR values for a selected attribute. Observe the outlier and extreme values
3. Remove the value: Remove instances with outlier values and show the screenshots of dataset before and after the removal.

A screenshot of a computer

Description automatically generated

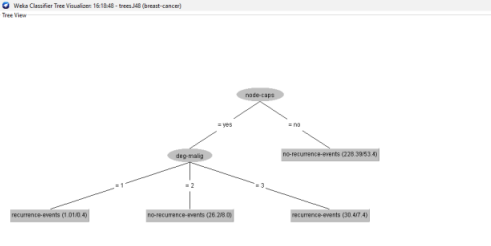
* CLASSIFICATION:

PROCEDURE: Perform NB, kNN and DT/rule based classification



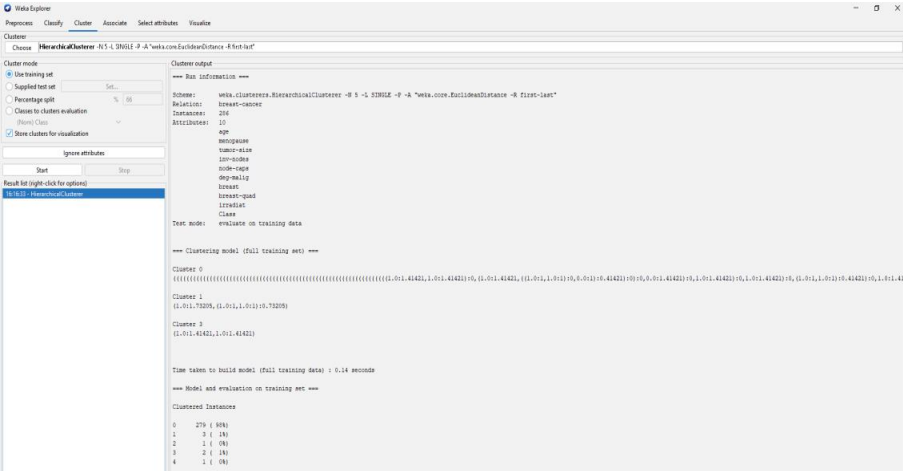
A screenshot of a computer

Description automatically generated



* CLUSTERING:

PROCEDURE: Perform kmeans, hierarchical clustering and explain the output



A screenshot of a computer

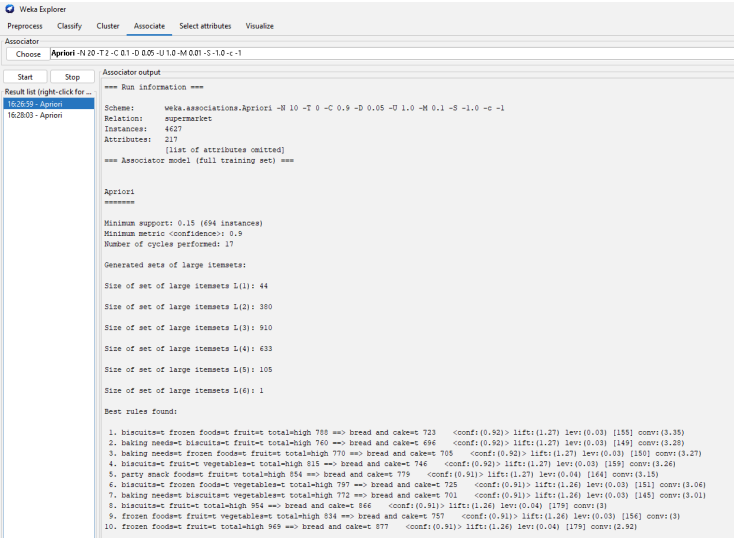
Description automatically generated

A diagram of a city

Description automatically generated

* ASSOCIATION RULE:

PROCEDURE: Perform apriori algorithm and show the rules created



* SELECT ATTRIBUTES:

PROCEDURE:

1. Apply suitable feature selection filter like GainRatio etc. to choose relevant attributes from the list of attributes.
2. Observe the ranks / priority provided by the filter.

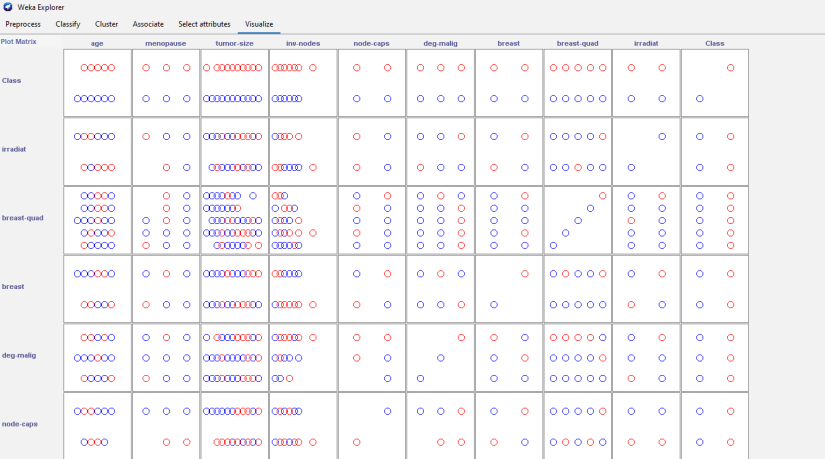
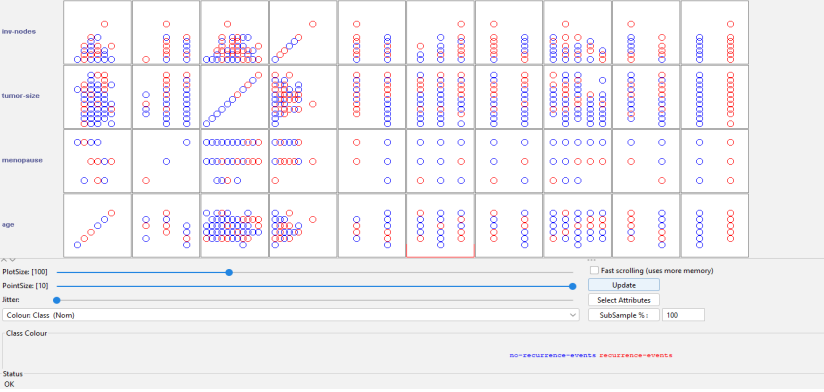
A screenshot of a computer

Description automatically generated

* VISUALIZATION:

PROCEDURE:

1. Visualize scatter plot for all the attributes from a dataset selected from Weka.
2. Determine correlation if any using these plots for different datasets



**CONCLUSION**:

* the experiment involving data pre-processing using the Weka data mining tool has been a valuable and essential step in preparing data for subsequent analysis.
* Weka provides a wide range of functionalities that aid in cleaning, transforming, and organizing data, making it more suitable for data mining and machine learning tasks.