Experiment 4

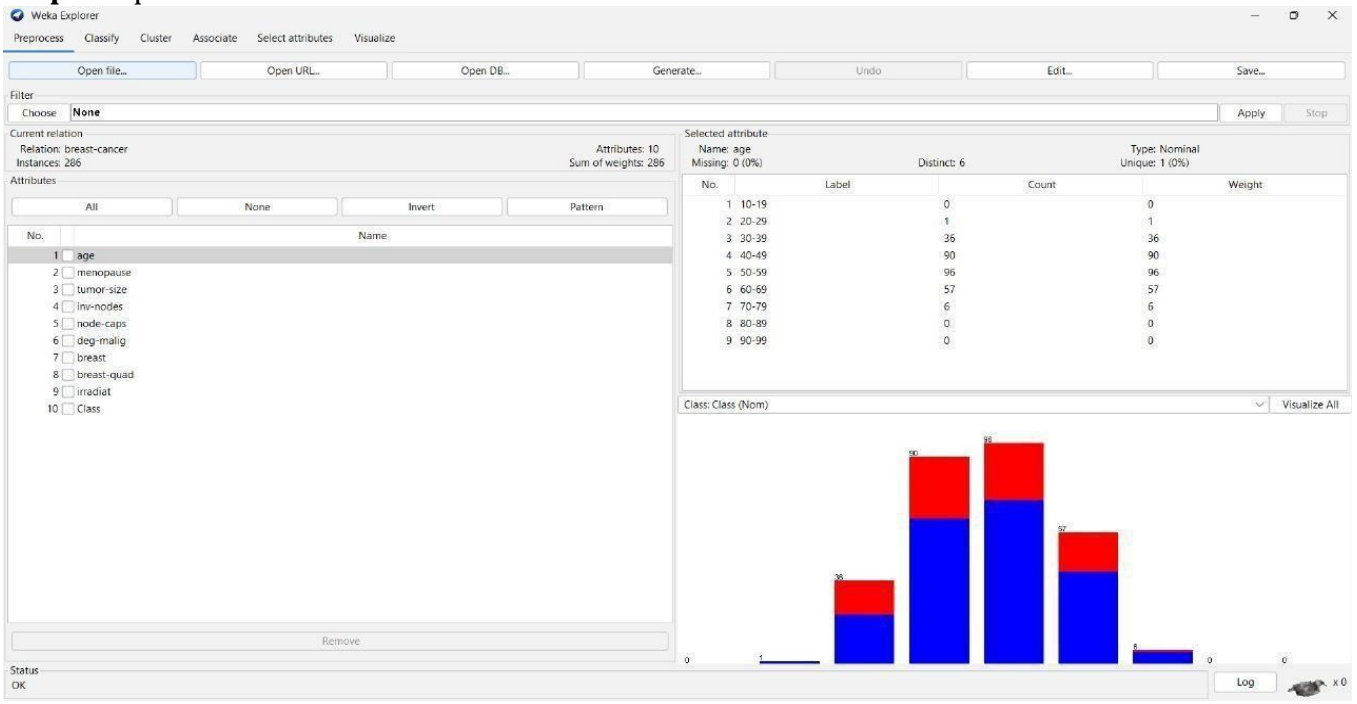
Title: Apply Preprocessing techniques on dataset using filters: Remove, ReplaceMissingValues, ReplaceMissingWithUserConstant, ReplaceWithMissingValue, Descritize. Also do the result analysis before and after preprocessing.

Filter 1: Remove

The Remove filter in Weka is an unsupervised attribute filter used to delete specific attributes (columns) from a dataset. It is commonly applied during data preprocessing to eliminate irrelevant, redundant, or non-informative features such as ID numbers or metadata that do not contribute to the learning process. Located under filters → unsupervised → attribute → Remove, this filter allows users to specify which attributes to remove using the -R option, where attributes are indexed starting from 1. For example, using -R 1,3 will remove the first and third attributes from the dataset. The Remove filter is essential for simplifying the dataset and improving model performance by focusing only on the most relevant attributes.

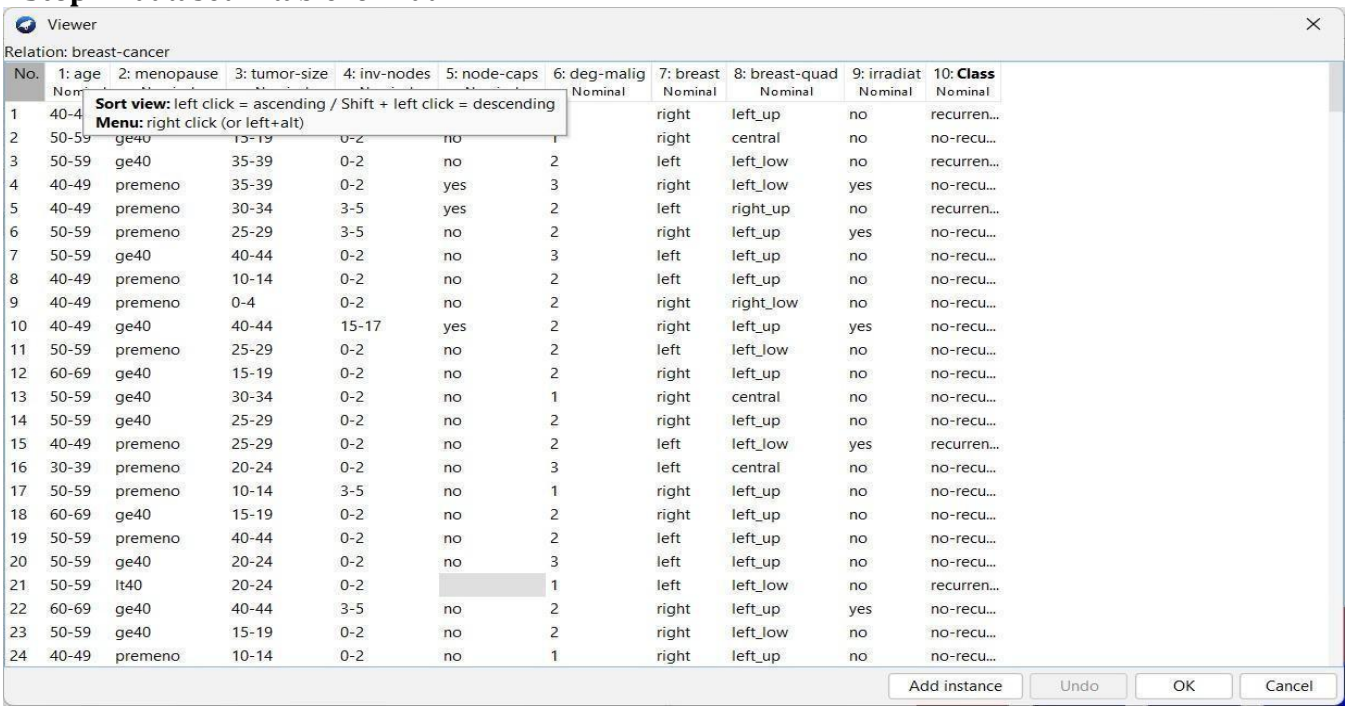
Dataset: breast-cancer.arff

Step-1: Upload dataset in Weka.



This image shows the **Preprocess tab** of the **Weka Explorer** interface, a data mining tool. The dataset in use is titled **"breast-cancer"**, containing **286 instances** and **10 attributes**. The selected attribute is **"age"**, which is of **Nominal** type with **6 distinct values** (e.g., 30–39, 40–49, etc.).
On the right side, a bar chart visualizes the distribution of the "age" attribute against the "Class" label (likely recurrence vs. no recurrence). Blue and red segments in bars represent the counts for different class labels. Most instances are concentrated in the **40–49**, **50–59**, and **60–69** age ranges.

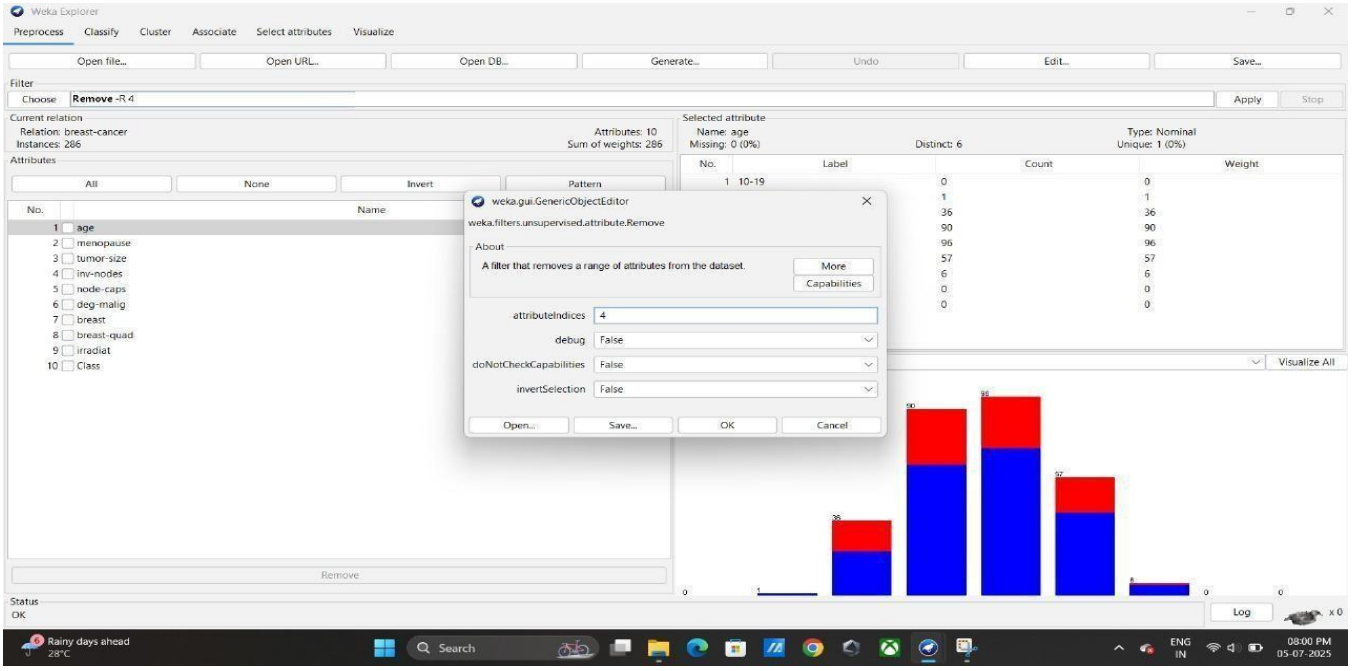
Step-2: dataset in table format.



No.	1: age	2: menopause	3: tumor-size	4: inv-nodes	5: node-caps	6: deg-malig	7: breast	8: breast-quad	9: irradiat	10: Class
	Nominal					Nominal	Nominal	Nominal	Nominal	Nominal
1	40-4						right	left_up	no	recurren...
2	50-59	ge40	15-19	0-2	no	1	right	central	no	no-recu...
3	50-59	ge40	35-39	0-2	no	2	left	left_low	no	recurren...
4	40-49	premeno	35-39	0-2	yes	3	right	left_low	yes	no-recu...
5	40-49	premeno	30-34	3-5	yes	2	left	right_up	no	recurren...
6	50-59	premeno	25-29	3-5	no	2	right	left_up	yes	no-recu...
7	50-59	ge40	40-44	0-2	no	3	left	left_up	no	no-recu...
8	40-49	premeno	10-14	0-2	no	2	left	left_up	no	no-recu...
9	40-49	premeno	0-4	0-2	no	2	right	right_low	no	no-recu...
10	40-49	ge40	40-44	15-17	yes	2	right	left_up	yes	no-recu...
11	50-59	premeno	25-29	0-2	no	2	left	left_low	no	no-recu...
12	60-69	ge40	15-19	0-2	no	2	right	left_up	no	no-recu...
13	50-59	ge40	30-34	0-2	no	1	right	central	no	no-recu...
14	50-59	ge40	25-29	0-2	no	2	right	left_up	no	no-recu...
15	40-49	premeno	25-29	0-2	no	2	left	left_low	yes	recurren...
16	30-39	premeno	20-24	0-2	no	3	left	central	no	no-recu...
17	50-59	premeno	10-14	3-5	no	1	right	left_up	no	no-recu...
18	60-69	ge40	15-19	0-2	no	2	right	left_up	no	no-recu...
19	50-59	premeno	40-44	0-2	no	2	left	left_up	no	no-recu...
20	50-59	ge40	20-24	0-2	no	3	left	left_up	no	no-recu...
21	50-59	lt40	20-24	0-2		1	left	left_low	no	recurren...
22	60-69	ge40	40-44	3-5	no	2	right	left_up	yes	no-recu...
23	50-59	ge40	15-19	0-2	no	2	right	left_low	no	no-recu...
24	40-49	premeno	10-14	0-2	no	1	right	left_up	no	no-recu...

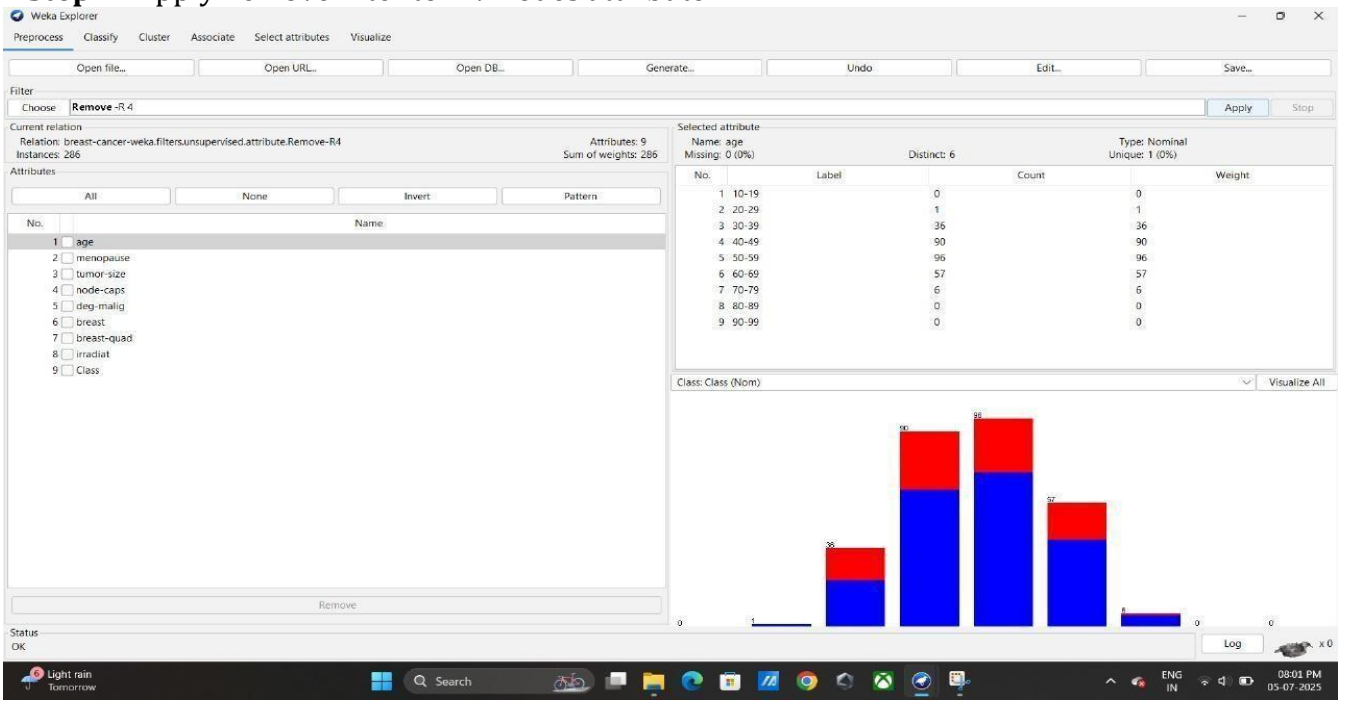
This image shows the **Viewer window** of the **Weka Explorer**, displaying raw data from the **breast-cancer dataset**. Each row represents an instance, and columns represent attributes like **age, menopause, tumor-size**, and **Class** (recurrence or no recurrence).

Step-3: Apply Remove filter to inv-nodes attribute.



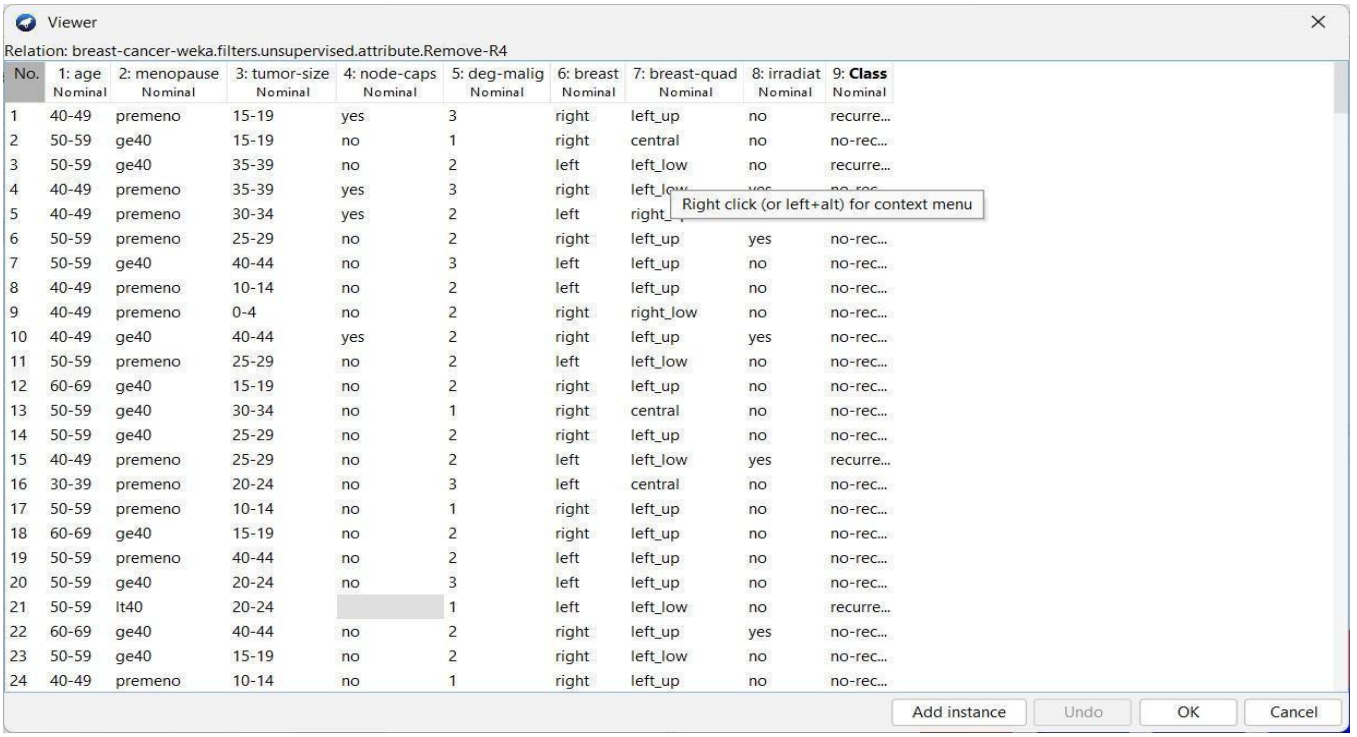
This image shows Weka Explorer's **Preprocess tab**, where the user is applying a **Remove filter** to delete attribute index **4 (inv-nodes)** from the **breast-cancer dataset** using the **GenericObjectEditor**.

Step-4: Apply Remove filter to inv-nodes attribute.



This Screenshot shows that inv-nodes attribute is removed after applying Remove filter.

Step-5: Dataset in table format after applying Remove filter .



No.	1: age Nominal	2: menopause Nominal	3: tumor-size Nominal	4: node-caps Nominal	5: deg-malig Nominal	6: breast Nominal	7: breast-quad Nominal	8: irradiat Nominal	9: Class Nominal
1	40-49	premeno	15-19	yes	3	right	left_up	no	recurre...
2	50-59	ge40	15-19	no	1	right	central	no	no-rec...
3	50-59	ge40	35-39	no	2	left	left_low	no	recurre...
4	40-49	premeno	35-39	yes	3	right	left_low		
5	40-49	premeno	30-34	yes	2	left	right_		
6	50-59	premeno	25-29	no	2	right	left_up	yes	no-rec...
7	50-59	ge40	40-44	no	3	left	left_up	no	no-rec...
8	40-49	premeno	10-14	no	2	left	left_up	no	no-rec...
9	40-49	premeno	0-4	no	2	right	right_low	no	no-rec...
10	40-49	ge40	40-44	yes	2	right	left_up	yes	no-rec...
11	50-59	premeno	25-29	no	2	left	left_low	no	no-rec...
12	60-69	ge40	15-19	no	2	right	left_up	no	no-rec...
13	50-59	ge40	30-34	no	1	right	central	no	no-rec...
14	50-59	ge40	25-29	no	2	right	left_up	no	no-rec...
15	40-49	premeno	25-29	no	2	left	left_low	yes	recurre...
16	30-39	premeno	20-24	no	3	left	central	no	no-rec...
17	50-59	premeno	10-14	no	1	right	left_up	no	no-rec...
18	60-69	ge40	15-19	no	2	right	left_up	no	no-rec...
19	50-59	premeno	40-44	no	2	left	left_up	no	no-rec...
20	50-59	ge40	20-24	no	3	left	left_up	no	no-rec...
21	50-59	lt40	20-24		1	left	left_low	no	recurre...
22	60-69	ge40	40-44	no	2	right	left_up	yes	no-rec...
23	50-59	ge40	15-19	no	2	right	left_low	no	no-rec...
24	40-49	premeno	10-14	no	1	right	left_up	no	no-rec...

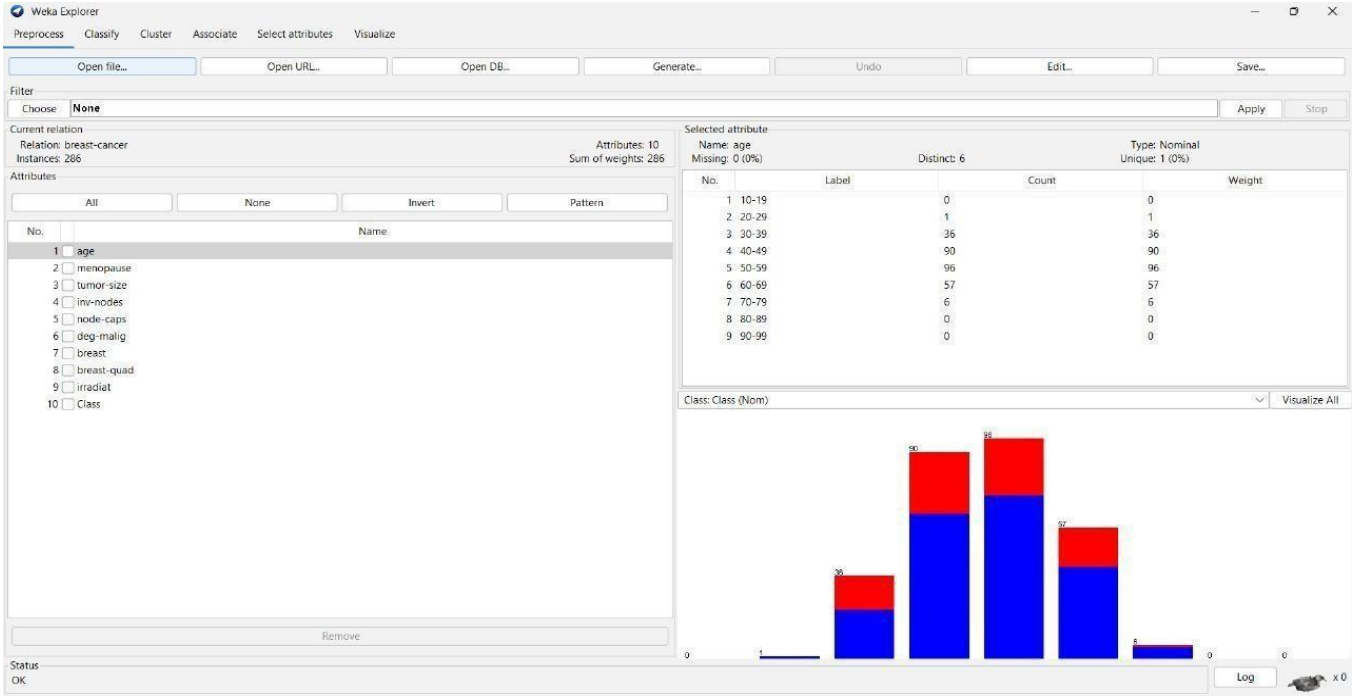
This screenshot shows the table format of dataset after removing inv-nodes attribute.

Filter 2: ReplaceMissingValues

The **ReplaceMissingValues** filter in Weka is an unsupervised filter used to automatically handle missing data in a dataset. When applied, it replaces any missing values in numeric attributes with the **mean** of the non-missing values, and for nominal (categorical) attributes, it replaces missing values with the **mode** (most frequent value). This filter is located under filters → unsupervised → attribute → ReplaceMissingValues. It is commonly used during the preprocessing stage to ensure that machine learning algorithms receive complete input data, as many models cannot handle missing values directly. This filter helps maintain dataset integrity while avoiding the loss of valuable data due to deletion of incomplete records.

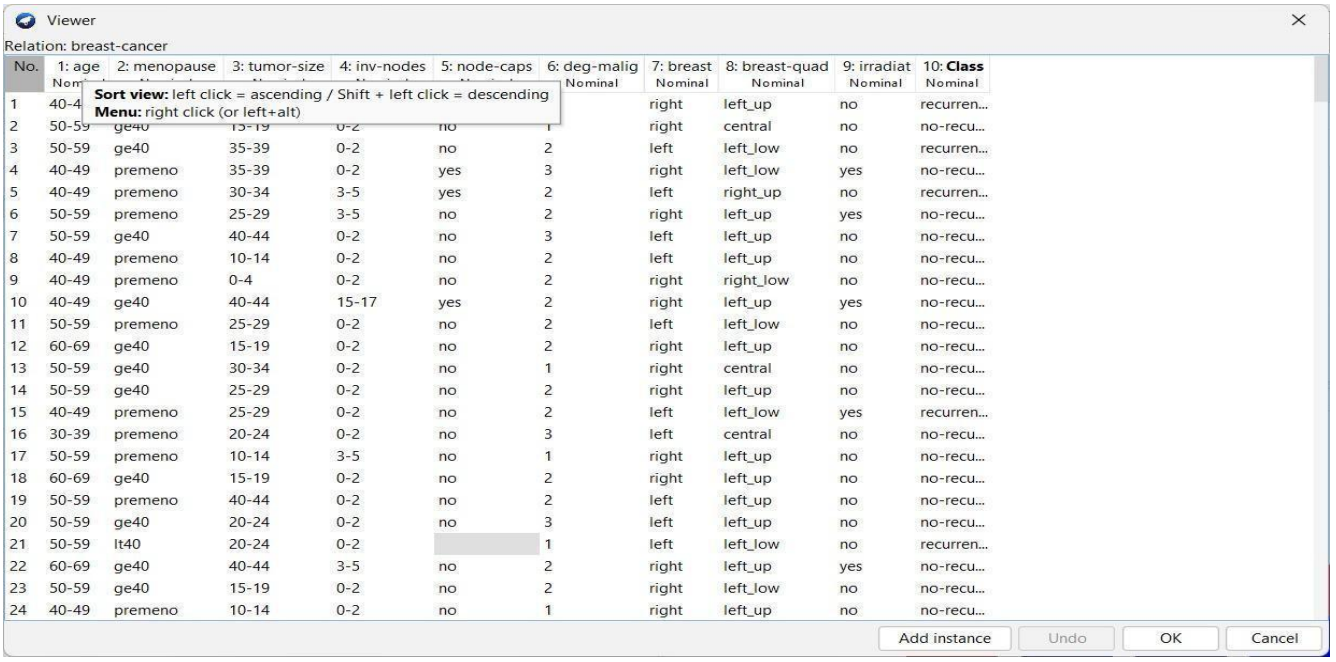
Dataset: breast-cancer.arff

Step-1: Upload dataset in Weka.



This image shows the **Preprocess tab** of the **Weka Explorer** interface, a data mining tool. The dataset in use is titled **"breast-cancer"**, containing **286 instances** and **10 attributes**. The selected attribute is **"age"**, which is of **Nominal** type with **6 distinct values** (e.g., 30–39, 40–49, etc.).
On the right side, a bar chart visualizes the distribution of the "age" attribute against the "Class" label (likely recurrence vs. no recurrence). Blue and red segments in bars represent the counts for different class labels. Most instances are concentrated in the **40–49**, **50–59**, and **60–69** age ranges.

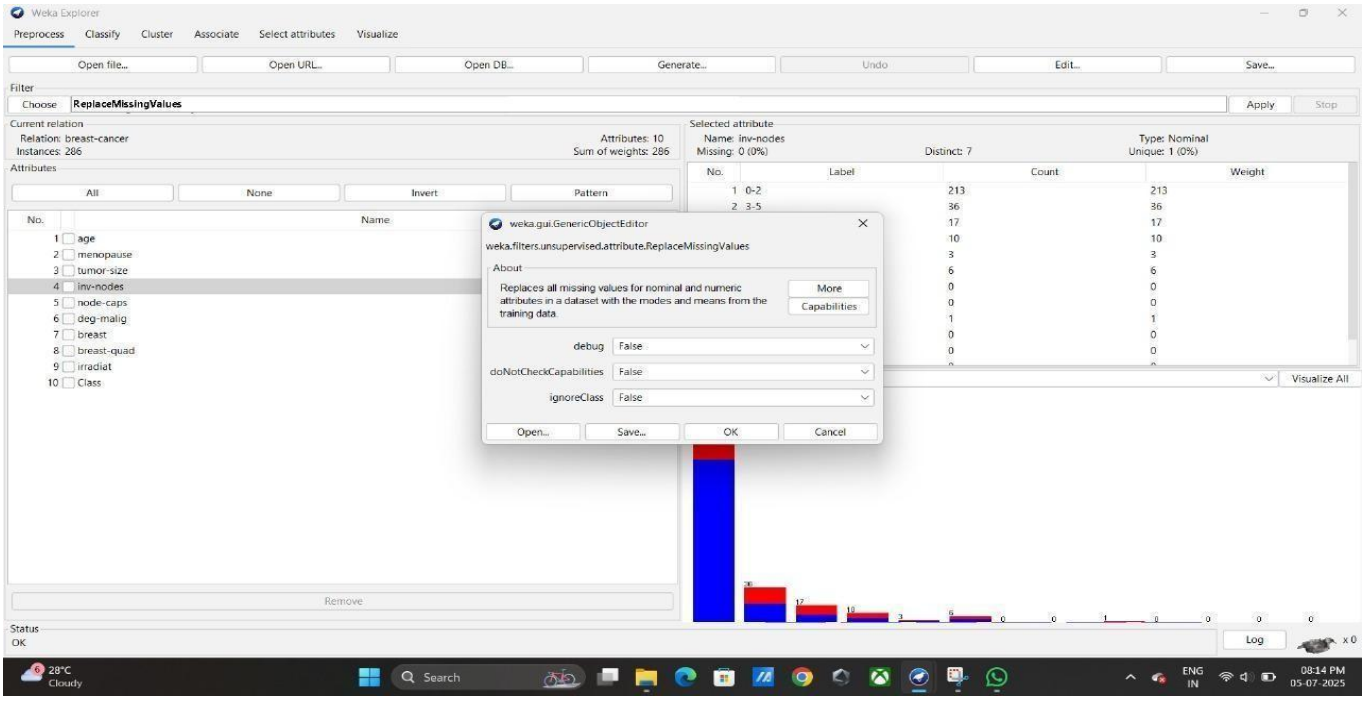
Step-2: dataset in table format.



No.	1: age	2: menopause	3: tumor-size	4: inv-nodes	5: node-caps	6: deg-malig	7: breast	8: breast-quad	9: irradiat	10: Class
	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal
1	40-4						right	left_up	no	recurren...
2	50-59	ge40	15-19	0-2	no	1	right	central	no	no-recu...
3	50-59	ge40	35-39	0-2	no	2	left	left_low	no	recurren...
4	40-49	premeno	35-39	0-2	yes	3	right	left_low	yes	no-recu...
5	40-49	premeno	30-34	3-5	yes	2	left	right_up	no	recurren...
6	50-59	premeno	25-29	3-5	no	2	right	left_up	yes	no-recu...
7	50-59	ge40	40-44	0-2	no	3	left	left_up	no	no-recu...
8	40-49	premeno	10-14	0-2	no	2	left	left_up	no	no-recu...
9	40-49	premeno	0-4	0-2	no	2	right	right_low	no	no-recu...
10	40-49	ge40	40-44	15-17	yes	2	right	left_up	yes	no-recu...
11	50-59	premeno	25-29	0-2	no	2	left	left_low	no	no-recu...
12	60-69	ge40	15-19	0-2	no	2	right	left_up	no	no-recu...
13	50-59	ge40	30-34	0-2	no	1	right	central	no	no-recu...
14	50-59	ge40	25-29	0-2	no	2	right	left_up	no	no-recu...
15	40-49	premeno	25-29	0-2	no	2	left	left_low	yes	recurren...
16	30-39	premeno	20-24	0-2	no	3	left	central	no	no-recu...
17	50-59	premeno	10-14	3-5	no	1	right	left_up	no	no-recu...
18	60-69	ge40	15-19	0-2	no	2	right	left_up	no	no-recu...
19	50-59	premeno	40-44	0-2	no	2	left	left_up	no	no-recu...
20	50-59	ge40	20-24	0-2	no	3	left	left_up	no	no-recu...
21	50-59	lt40	20-24	0-2		1	left	left_low	no	recurren...
22	60-69	ge40	40-44	3-5	no	2	right	left_up	yes	no-recu...
23	50-59	ge40	15-19	0-2	no	2	right	left_low	no	no-recu...
24	40-49	premeno	10-14	0-2	no	1	right	left_up	no	no-recu...

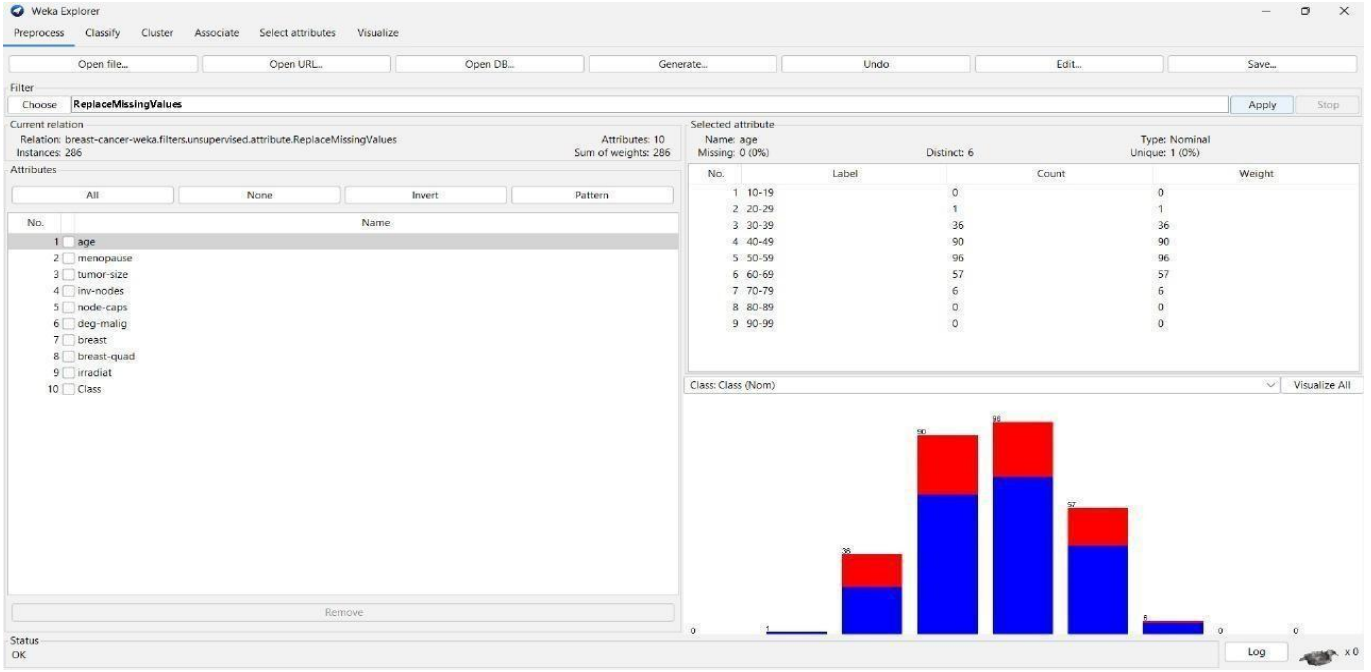
This image shows the **Viewer window** of the **Weka Explorer**, displaying raw data from the **breast-cancer dataset**. Each row represents an instance, and columns represent attributes like **age, menopause, tumor-size**, and **Class** (recurrence or no recurrence).

Step-3: Configure the parameters of ReplaceMissingValues filter.



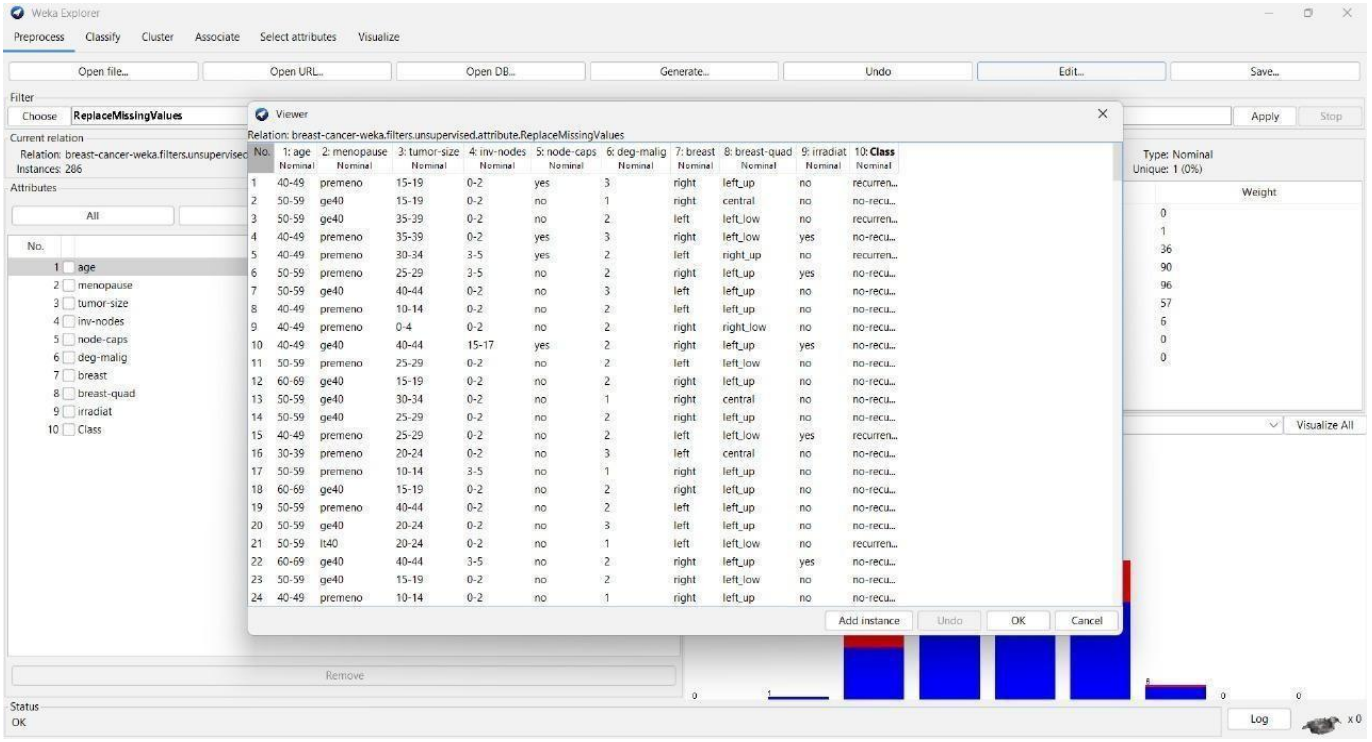
This image shows Weka Explorer with the **ReplaceMissingValues** filter selected, aiming to handle missing data in the **inv-nodes** attribute by replacing them with the **mode (for nominal)** or **mean (for numeric)** values.

Step-4: Apply ReplaceMissingValues filter to inv-nodes attribute.



This Screenshot shows that before applying filter inv-nodes has 42 missing values and after applying filter it becomes 0.

Step-5: Dataset in table format after applying Remove filter .



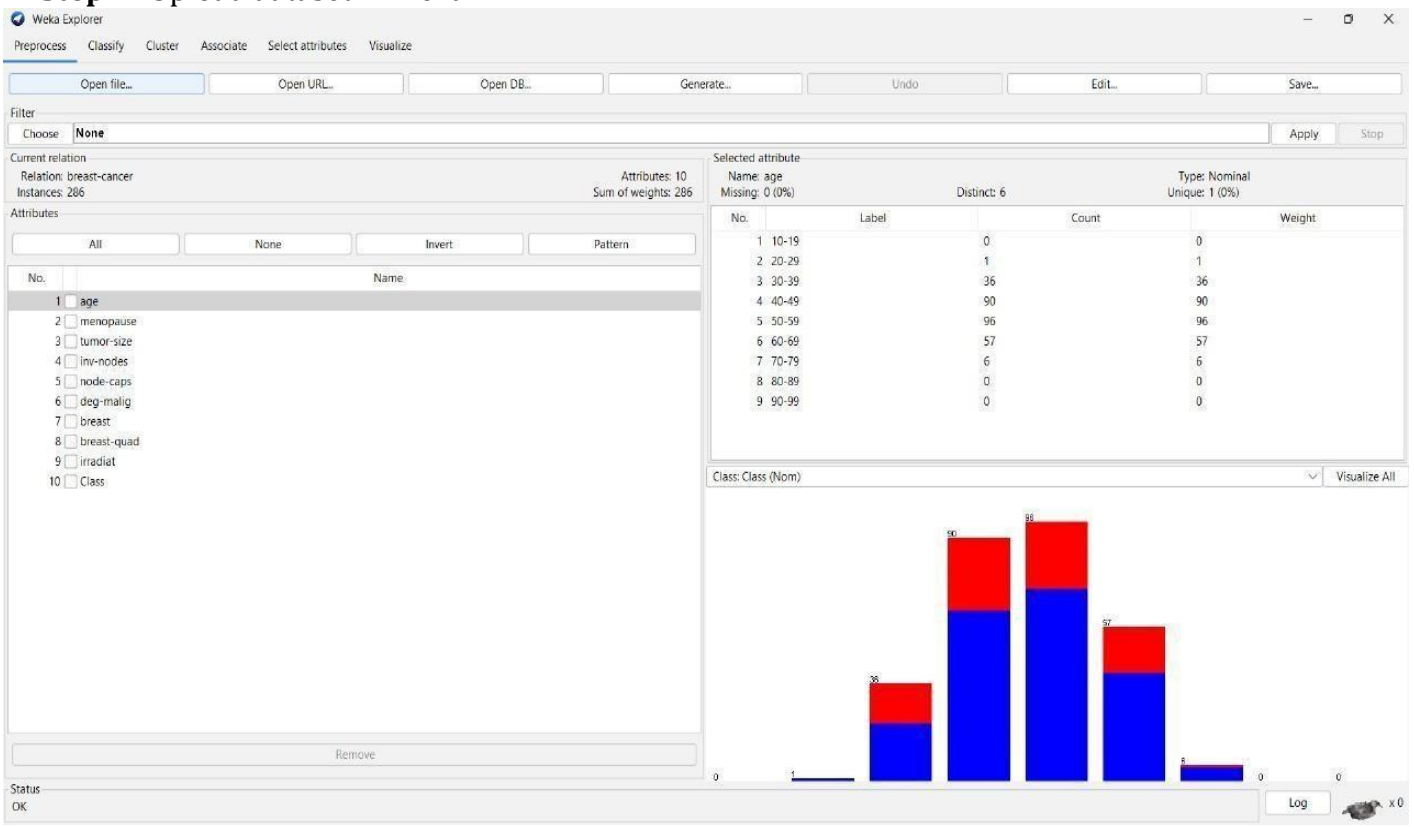
This screenshot shows the table format of dataset after Replacing missing values of inv-nodes attribute.

Filter 3: ReplaceMissingWithUserConstant

The ReplaceMissingWithUserConstant filter in WEKA replaces all missing values in a dataset with a user-specified constant. It allows setting different constants for nominal and numeric attributes, providing control over how missing data is handled.

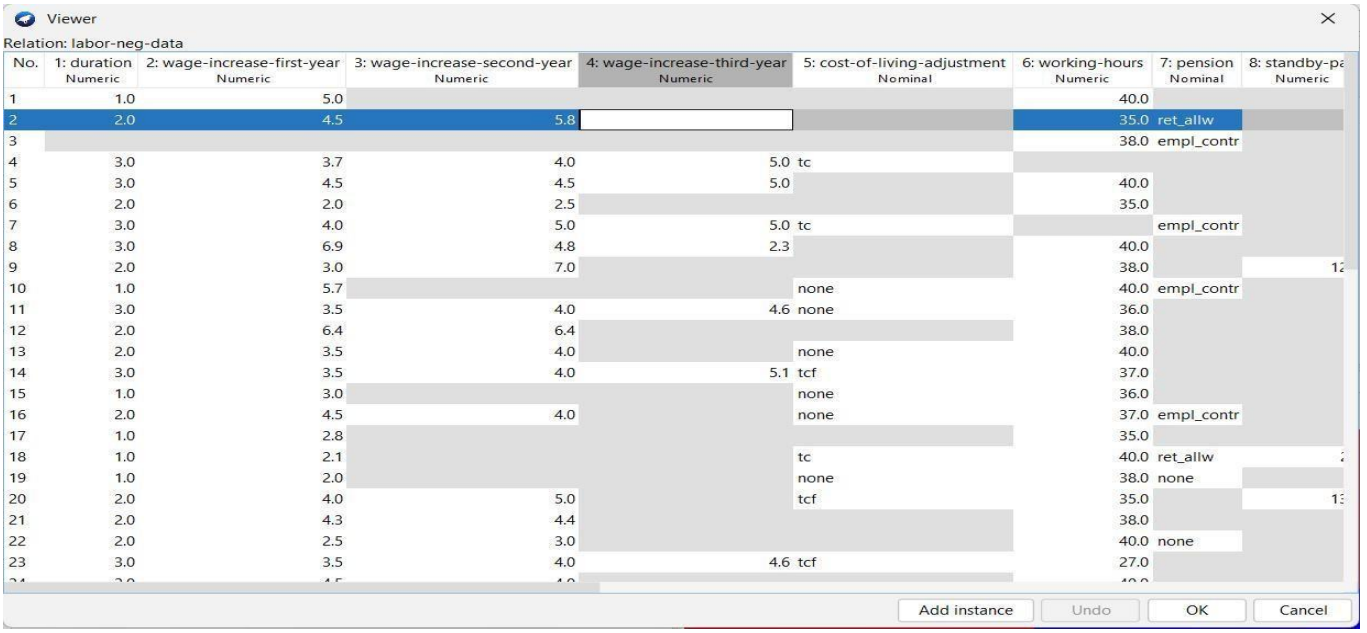
Dataset: labor.arff

Step-1: Upload dataset in Weka.



This image shows the Preprocess tab of the Weka Explorer interface, a data mining tool. The dataset in use is titled "breast-cancer", containing 286 instances and 10 attributes. The selected attribute is "age", which is of Nominal type with 6 distinct values (e.g., 30–39, 40–49, etc.).
On the right side, a bar chart visualizes the distribution of the "age" attribute against the "Class" label (likely recurrence vs. no recurrence). Blue and red segments in bars represent the counts for different class labels. Most instances are concentrated in the 40–49, 50–59, and 60–69 age ranges.

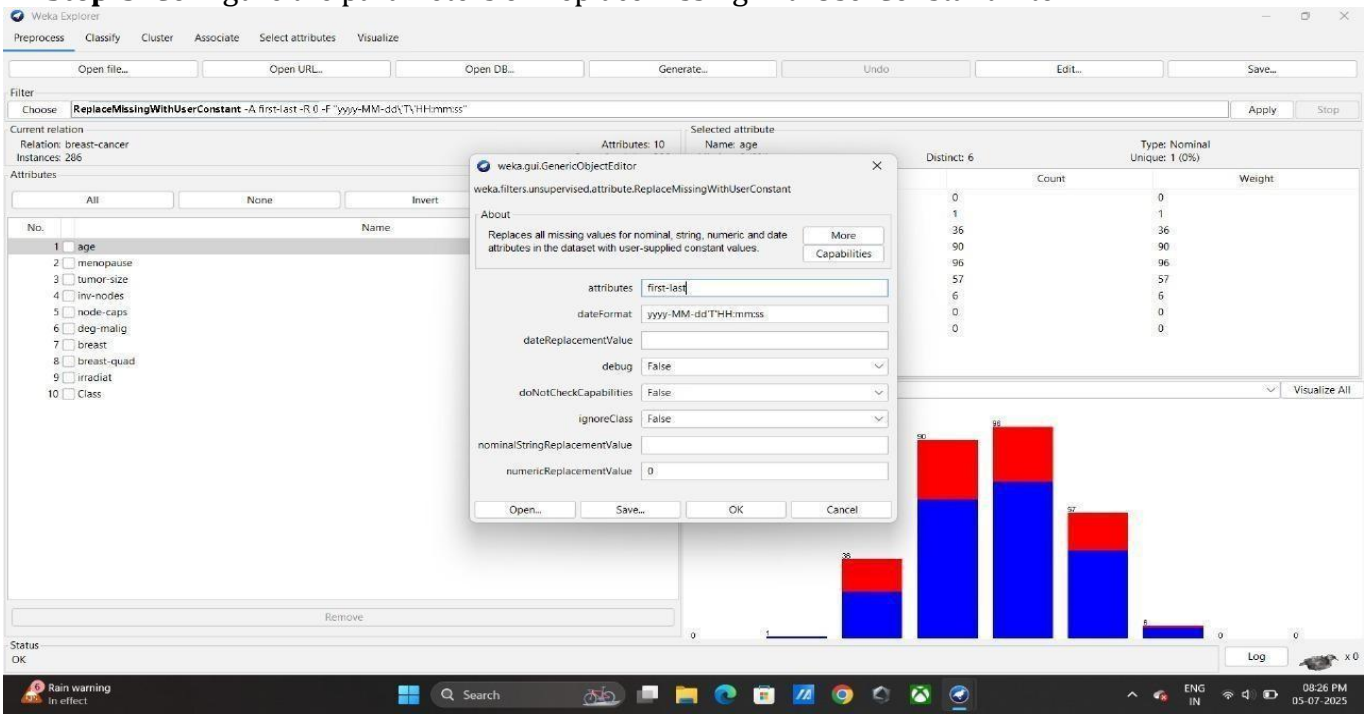
Step-2: dataset in table format.



No.	1: duration Numeric	2: wage-increase-first-year Numeric	3: wage-increase-second-year Numeric	4: wage-increase-third-year Numeric	5: cost-of-living-adjustment Nominal	6: working-hours Numeric	7: pension Nominal	8: standby-pa Numeric
1	1.0	5.0				40.0		
2	2.0	4.5	5.8			35.0	ret_allw	
3						38.0	empl_contr	
4	3.0	3.7	4.0	5.0	tc			
5	3.0	4.5	4.5	5.0		40.0		
6	2.0	2.0	2.5			35.0		
7	3.0	4.0	5.0	5.0	tc		empl_contr	
8	3.0	6.9	4.8	2.3		40.0		
9	2.0	3.0	7.0			38.0		12
10	1.0	5.7			none	40.0	empl_contr	
11	3.0	3.5	4.0	4.6	none	36.0		
12	2.0	6.4	6.4			38.0		
13	2.0	3.5	4.0		none	40.0		
14	3.0	3.5	4.0	5.1	tcf	37.0		
15	1.0	3.0			none	36.0		
16	2.0	4.5	4.0		none	37.0	empl_contr	
17	1.0	2.8				35.0		
18	1.0	2.1			tc	40.0	ret_allw	2
19	1.0	2.0			none	38.0	none	
20	2.0	4.0	5.0		tcf	35.0		13
21	2.0	4.3	4.4			38.0		
22	2.0	2.5	3.0			40.0	none	
23	3.0	3.5	4.0	4.6	tcf	27.0		

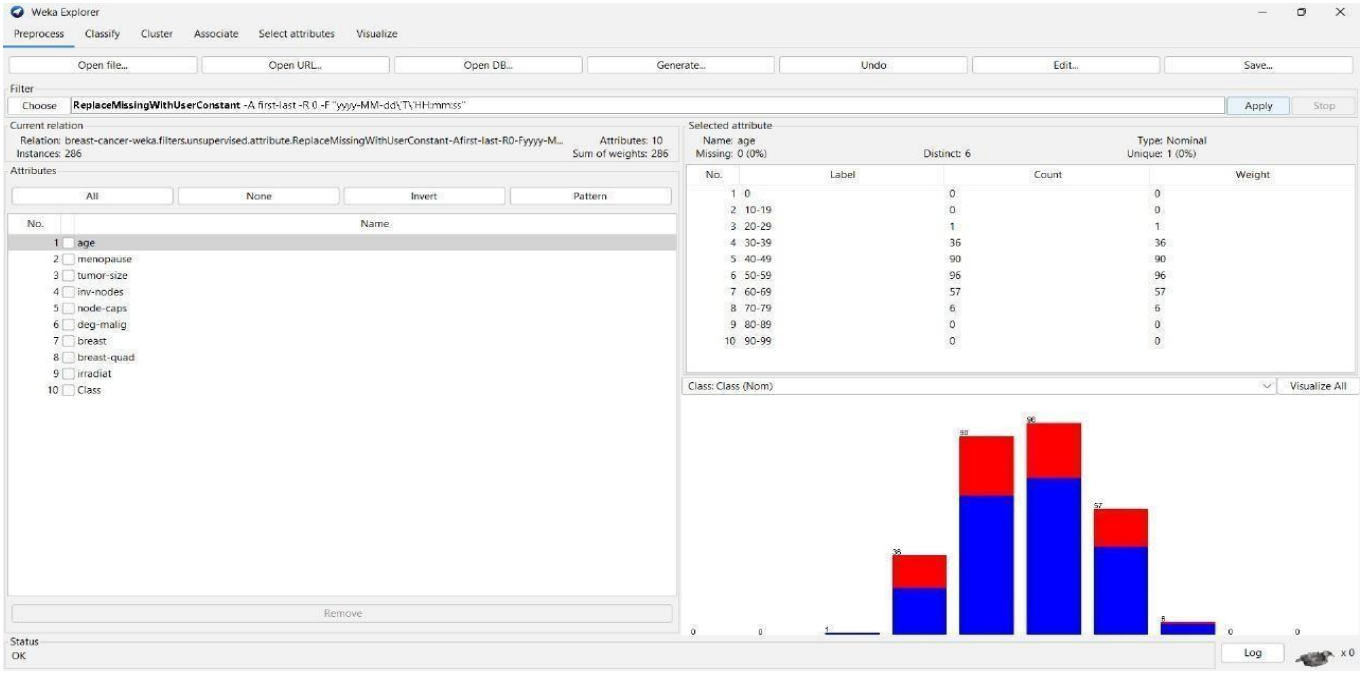
This image shows the Viewer window of the Weka Explorer, displaying raw data from the breast-cancer dataset. Each row represents an instance, and columns represent attributes like age, menopause, tumor-size, and Class (recurrence or no recurrence).

Step-3: Configure the parameters of ReplaceMissingWithUserConstant filter.



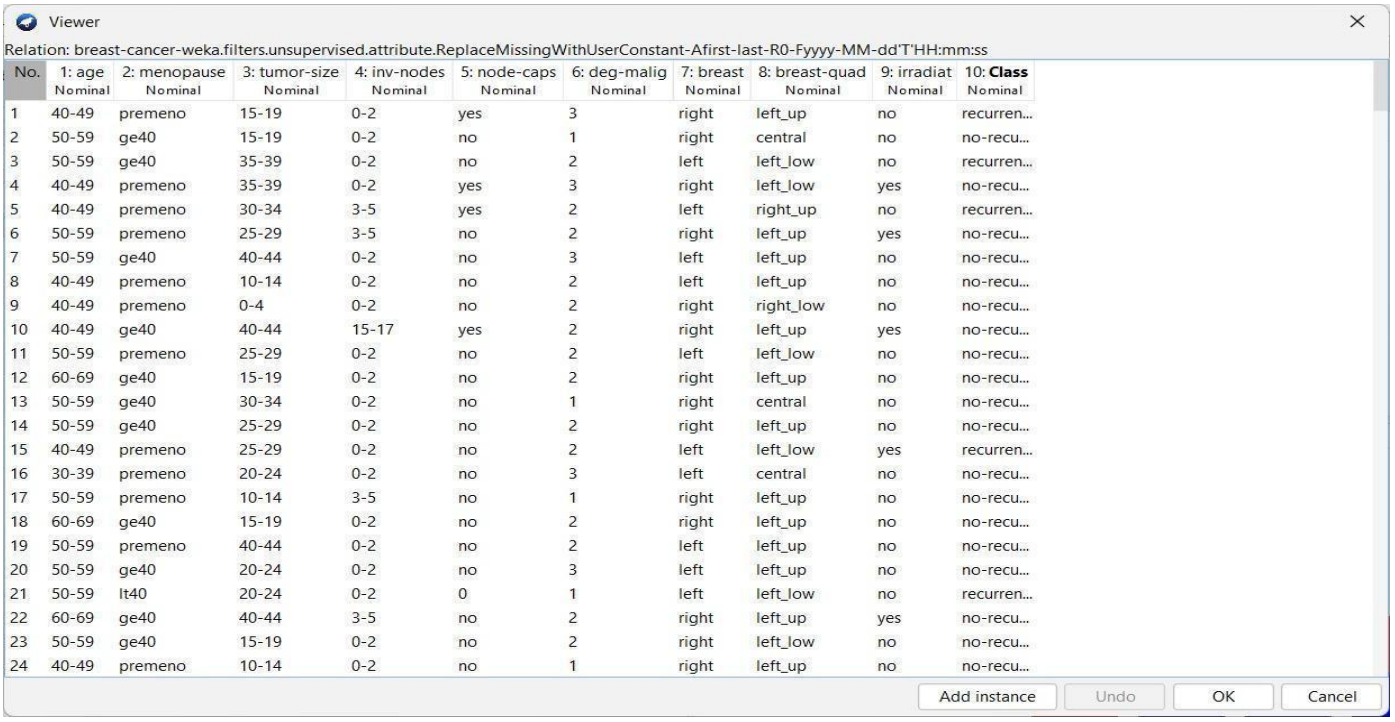
This image shows Weka Explorer using the **ReplaceMissingWithUserConstant** filter, which replaces all missing values in the dataset with user-defined constants for string, numeric, and date attributes.

Step-4: Apply ReplaceMissingWithUserConstant filter to all attributes.



This Screenshot shows that ReplaceMissingWithUserConstant filter applied to all attributes and when I click on Apply then missing values are replaced by ABC.

Step-5: Dataset in table format after applying Remove filter .



No.	1: age Nominal	2: menopause Nominal	3: tumor-size Nominal	4: inv-nodes Nominal	5: node-caps Nominal	6: deg-malig Nominal	7: breast Nominal	8: breast-quad Nominal	9: irradiat Nominal	10: **Class** Nominal
1	40-49	premeno	15-19	0-2	yes	3	right	left_up	no	recurren...
2	50-59	ge40	15-19	0-2	no	1	right	central	no	no-recu...
3	50-59	ge40	35-39	0-2	no	2	left	left_low	no	recurren...
4	40-49	premeno	35-39	0-2	yes	3	right	left_low	yes	no-recu...
5	40-49	premeno	30-34	3-5	yes	2	left	right_up	no	recurren...
6	50-59	premeno	25-29	3-5	no	2	right	left_up	yes	no-recu...
7	50-59	ge40	40-44	0-2	no	3	left	left_up	no	no-recu...
8	40-49	premeno	10-14	0-2	no	2	left	left_up	no	no-recu...
9	40-49	premeno	0-4	0-2	no	2	right	right_low	no	no-recu...
10	40-49	ge40	40-44	15-17	yes	2	right	left_up	yes	no-recu...
11	50-59	premeno	25-29	0-2	no	2	left	left_low	no	no-recu...
12	60-69	ge40	15-19	0-2	no	2	right	left_up	no	no-recu...
13	50-59	ge40	30-34	0-2	no	1	right	central	no	no-recu...
14	50-59	ge40	25-29	0-2	no	2	right	left_up	no	no-recu...
15	40-49	premeno	25-29	0-2	no	2	left	left_low	yes	recurren...
16	30-39	premeno	20-24	0-2	no	3	left	central	no	no-recu...
17	50-59	premeno	10-14	3-5	no	1	right	left_up	no	no-recu...
18	60-69	ge40	15-19	0-2	no	2	right	left_up	no	no-recu...
19	50-59	premeno	40-44	0-2	no	2	left	left_up	no	no-recu...
20	50-59	ge40	20-24	0-2	no	3	left	left_up	no	no-recu...
21	50-59	lt40	20-24	0-2	0	1	left	left_low	no	recurren...
22	60-69	ge40	40-44	3-5	no	2	right	left_up	yes	no-recu...
23	50-59	ge40	15-19	0-2	no	2	right	left_low	no	no-recu...
24	40-49	premeno	10-14	0-2	no	1	right	left_up	no	no-recu...

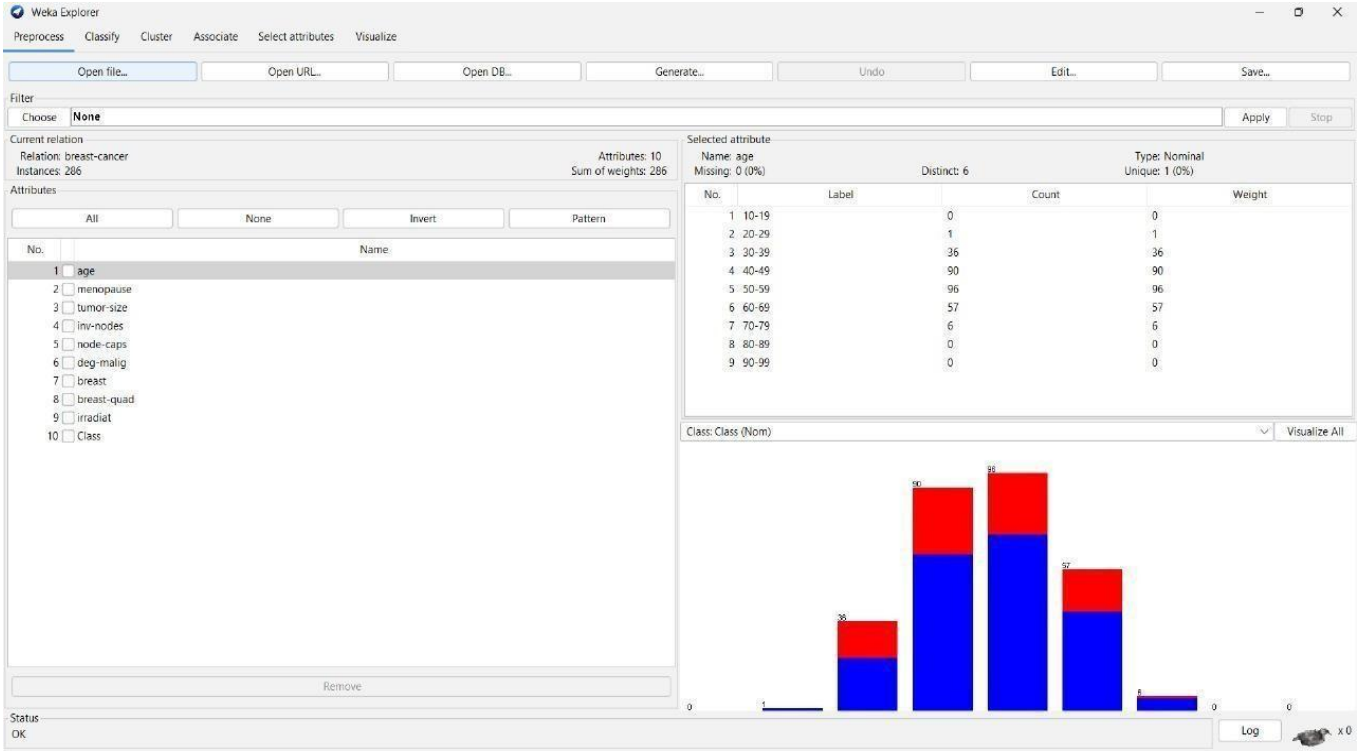
This screenshot shows the table format of dataset after Replacing missing values of inv-nodes attribute.

Filter 4: ReplaceWithMissingValue

The ReplaceWithMissingValue filter in WEKA replaces specified attribute values with missing values. It is useful for simulating missing data or reverting imputed values back to missing for testing or preprocessing purposes.

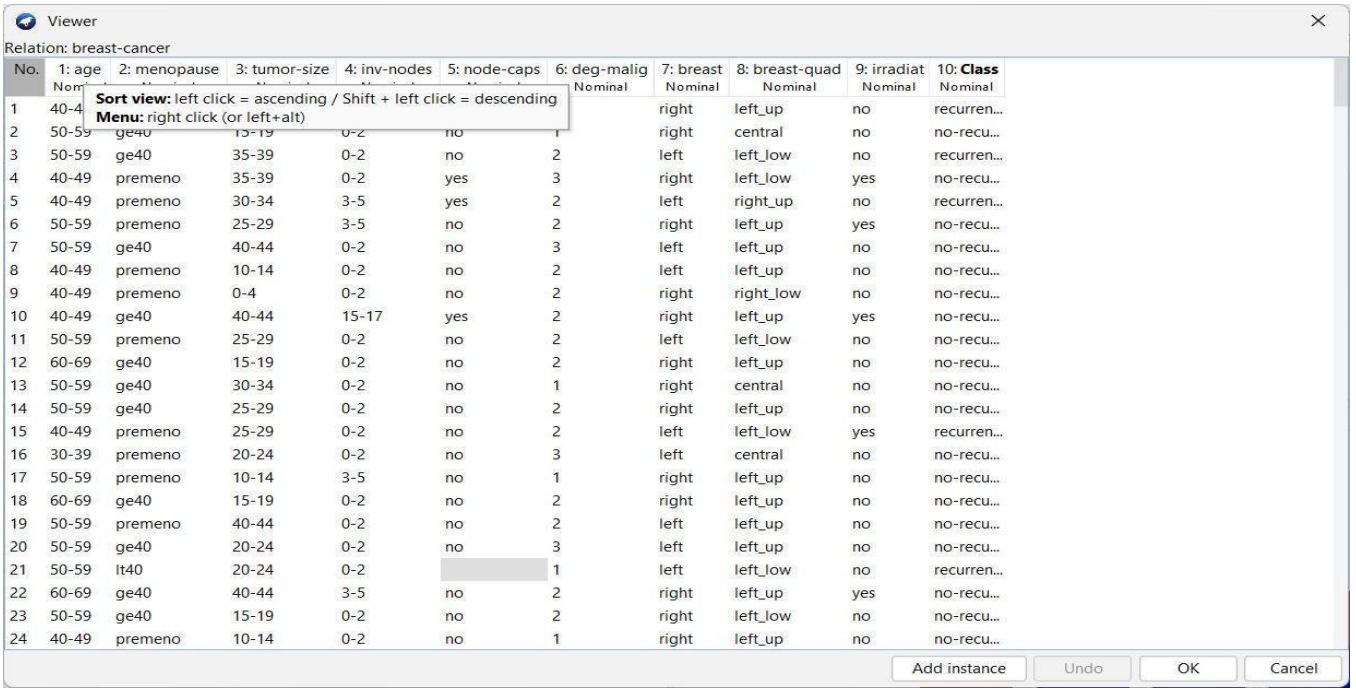
Dataset: breast-cancer.arff

Step-1: Upload dataset in Weka.



This image shows the Preprocess tab of the Weka Explorer interface, a data mining tool. The dataset in use is titled "breast-cancer", containing 286 instances and 10 attributes. The selected attribute is "age", which is of Nominal type with 6 distinct values (e.g., 30–39, 40–49, etc.).
On the right side, a bar chart visualizes the distribution of the "age" attribute against the "Class" label (likely recurrence vs. no recurrence). Blue and red segments in bars represent the counts for different class labels. Most instances are concentrated in the 40–49, 50–59, and 60–69 age ranges.

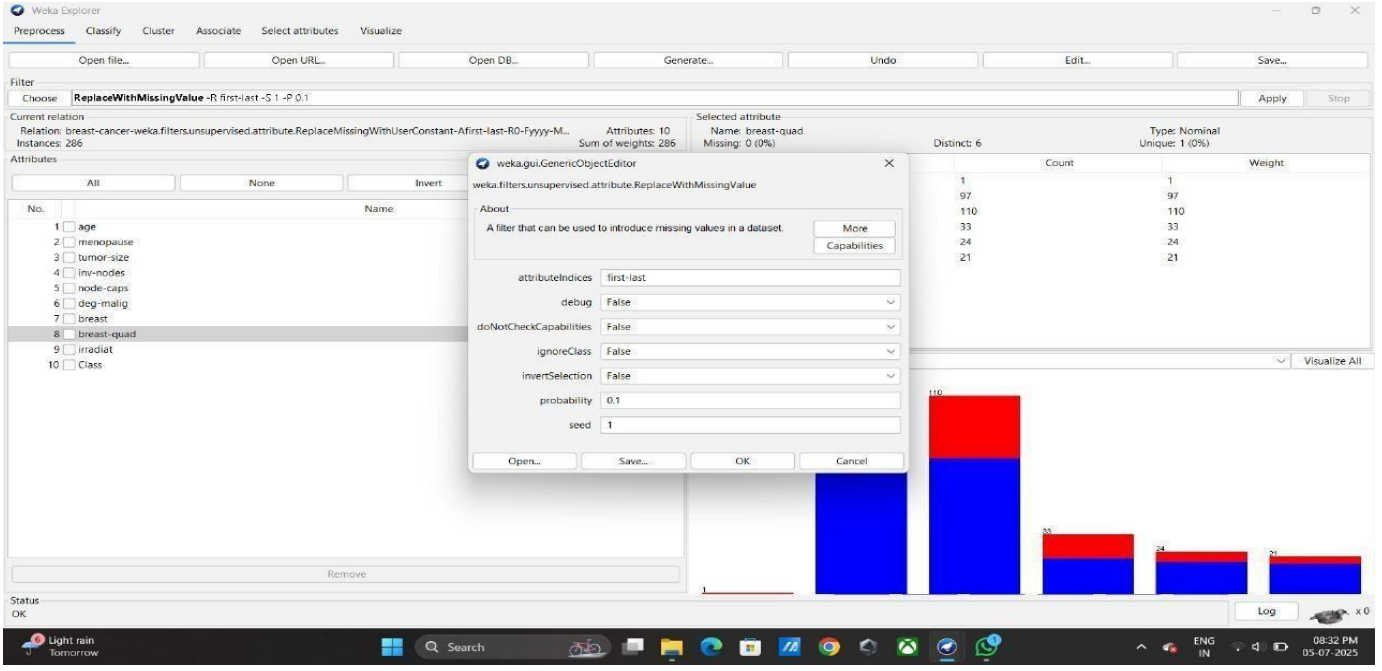
Step-2: dataset in table format.



No.	1: age	2: menopause	3: tumor-size	4: inv-nodes	5: node-caps	6: deg-malig	7: breast	8: breast-quad	9: irradiat	10: Class
	Nominal					Nominal	Nominal	Nominal	Nominal	Nominal
1	40-4						right	left_up	no	recurren...
2	50-59	ge40	15-19	0-2	no	1	right	central	no	no-recu...
3	50-59	ge40	35-39	0-2	no	2	left	left_low	no	recurren...
4	40-49	premeno	35-39	0-2	yes	3	right	left_low	yes	no-recu...
5	40-49	premeno	30-34	3-5	yes	2	left	right_up	no	recurren...
6	50-59	premeno	25-29	3-5	no	2	right	left_up	yes	no-recu...
7	50-59	ge40	40-44	0-2	no	3	left	left_up	no	no-recu...
8	40-49	premeno	10-14	0-2	no	2	left	left_up	no	no-recu...
9	40-49	premeno	0-4	0-2	no	2	right	right_low	no	no-recu...
10	40-49	ge40	40-44	15-17	yes	2	right	left_up	yes	no-recu...
11	50-59	premeno	25-29	0-2	no	2	left	left_low	no	no-recu...
12	60-69	ge40	15-19	0-2	no	2	right	left_up	no	no-recu...
13	50-59	ge40	30-34	0-2	no	1	right	central	no	no-recu...
14	50-59	ge40	25-29	0-2	no	2	right	left_up	no	no-recu...
15	40-49	premeno	25-29	0-2	no	2	left	left_low	yes	recurren...
16	30-39	premeno	20-24	0-2	no	3	left	central	no	no-recu...
17	50-59	premeno	10-14	3-5	no	1	right	left_up	no	no-recu...
18	60-69	ge40	15-19	0-2	no	2	right	left_up	no	no-recu...
19	50-59	premeno	40-44	0-2	no	2	left	left_up	no	no-recu...
20	50-59	ge40	20-24	0-2	no	3	left	left_up	no	no-recu...
21	50-59	lt40	20-24	0-2		1	left	left_low	no	recurren...
22	60-69	ge40	40-44	3-5	no	2	right	left_up	yes	no-recu...
23	50-59	ge40	15-19	0-2	no	2	right	left_low	no	no-recu...
24	40-49	premeno	10-14	0-2	no	1	right	left_up	no	no-recu...

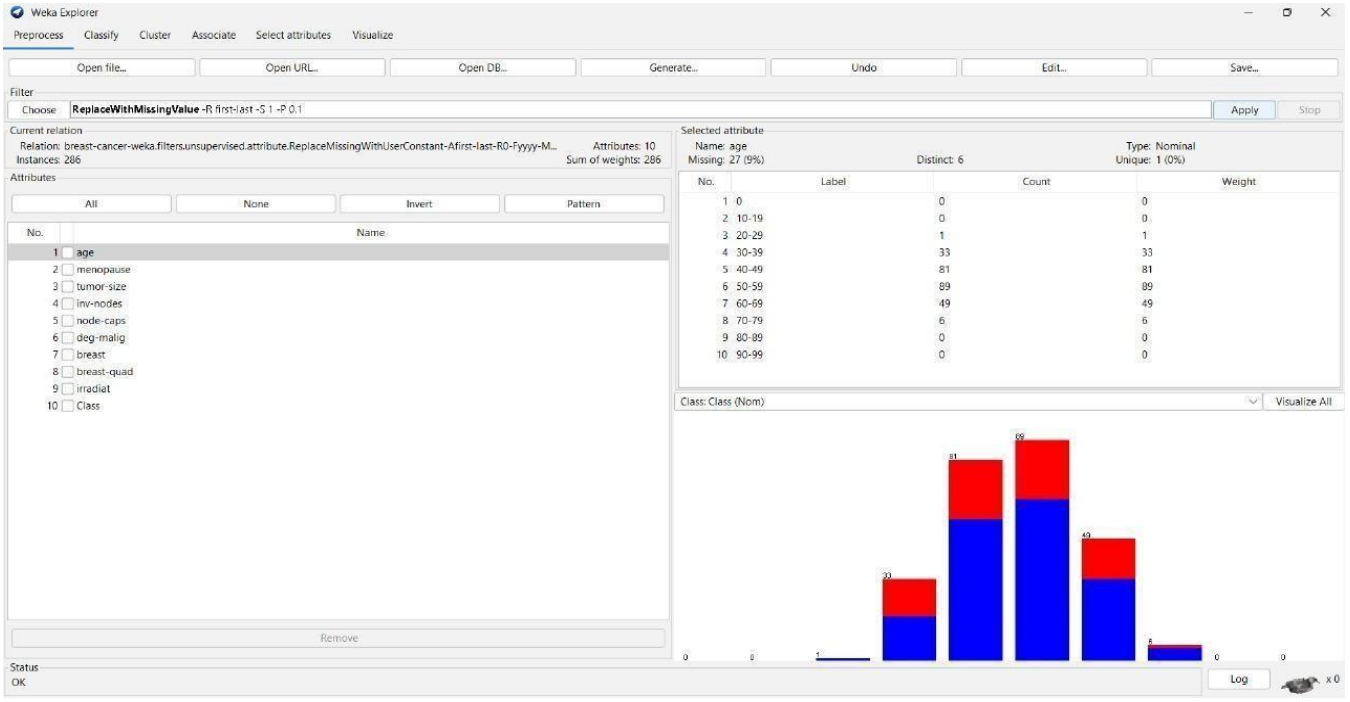
This screenshot shows the **tabular view of the labor.arff dataset** in Weka's **Instance Viewer**. Each row represents an instance (or record) from labor negotiations data, and each column is an attribute related to employment terms.

Step-3: Configure the parameters of ReplaceWithMissingValue filter.



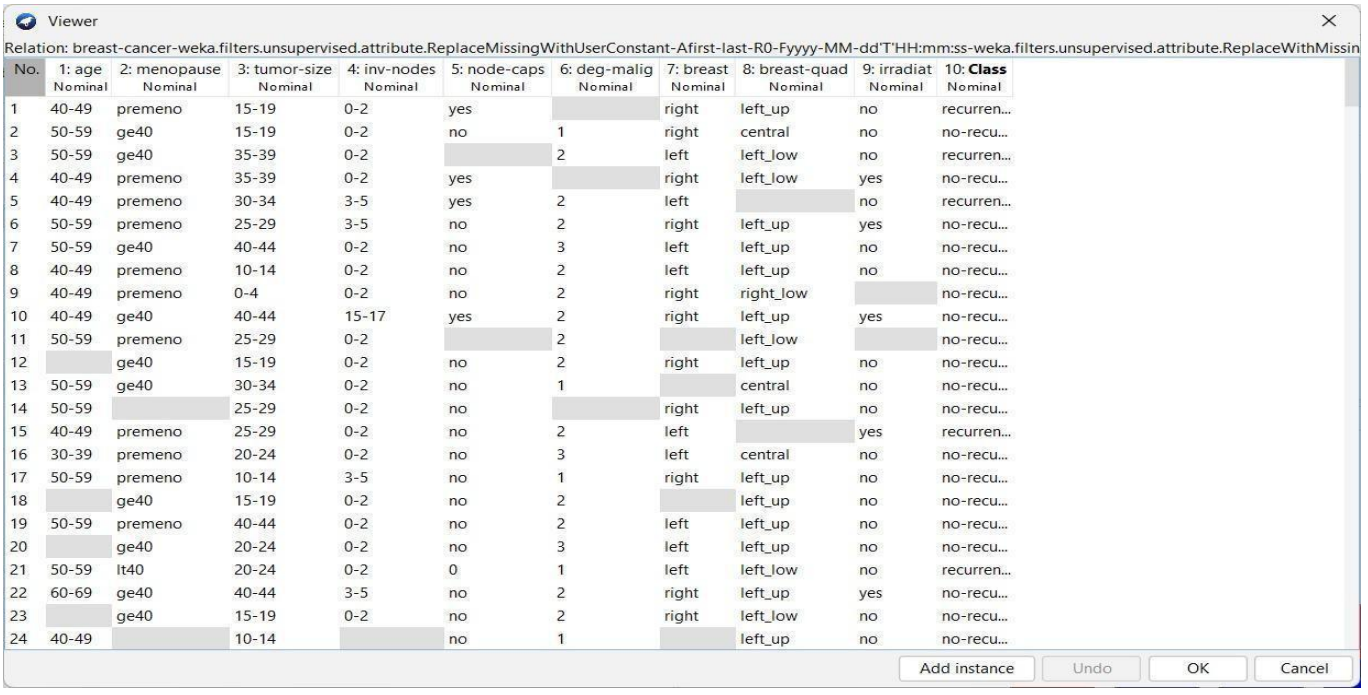
This image shows Weka Explorer using the **ReplaceWithMissingValue** filter, which artificially introduces missing values into the dataset based on a specified probability (0.1 in this case) for testing imputation methods.

Step-4: Apply ReplaceWithMissingValue filter from 1 to 6 attribute.



This Screenshot shows that ReplaceWithMissingValue to all attributes and when I click on Apply then values are replaced by missing values.

Step-5: Dataset in table format after applying Remove filter .



No.	1: age Nominal	2: menopause Nominal	3: tumor-size Nominal	4: inv-nodes Nominal	5: node-caps Nominal	6: deg-malig Nominal	7: breast Nominal	8: breast-quad Nominal	9: irradiat Nominal	10: **Class** Nominal
1	40-49	premeno	15-19	0-2	yes		right	left_up	no	recurren...
2	50-59	ge40	15-19	0-2	no	1	right	central	no	no-recu...
3	50-59	ge40	35-39	0-2		2	left	left_low	no	recurren...
4	40-49	premeno	35-39	0-2	yes		right	left_low	yes	no-recu...
5	40-49	premeno	30-34	3-5	yes	2	left		no	recurren...
6	50-59	premeno	25-29	3-5	no	2	right	left_up	yes	no-recu...
7	50-59	ge40	40-44	0-2	no	3	left	left_up	no	no-recu...
8	40-49	premeno	10-14	0-2	no	2	left	left_up	no	no-recu...
9	40-49	premeno	0-4	0-2	no	2	right	right_low		no-recu...
10	40-49	ge40	40-44	15-17	yes	2	right	left_up	yes	no-recu...
11	50-59	premeno	25-29	0-2		2		left_low		no-recu...
12		ge40	15-19	0-2	no	2	right	left_up	no	no-recu...
13	50-59	ge40	30-34	0-2	no	1		central	no	no-recu...
14	50-59		25-29	0-2	no		right	left_up	no	no-recu...
15	40-49	premeno	25-29	0-2	no	2	left		yes	recurren...
16	30-39	premeno	20-24	0-2	no	3	left	central	no	no-recu...
17	50-59	premeno	10-14	3-5	no	1	right	left_up	no	no-recu...
18		ge40	15-19	0-2	no	2		left_up	no	no-recu...
19	50-59	premeno	40-44	0-2	no	2	left	left_up	no	no-recu...
20		ge40	20-24	0-2	no	3	left	left_up	no	no-recu...
21	50-59	lt40	20-24	0-2	0	1	left	left_low	no	recurren...
22	60-69	ge40	40-44	3-5	no	2	right	left_up	yes	no-recu...
23		ge40	15-19	0-2	no	2	right	left_low	no	no-recu...
24	40-49		10-14		no	1		left_up	no	no-recu...

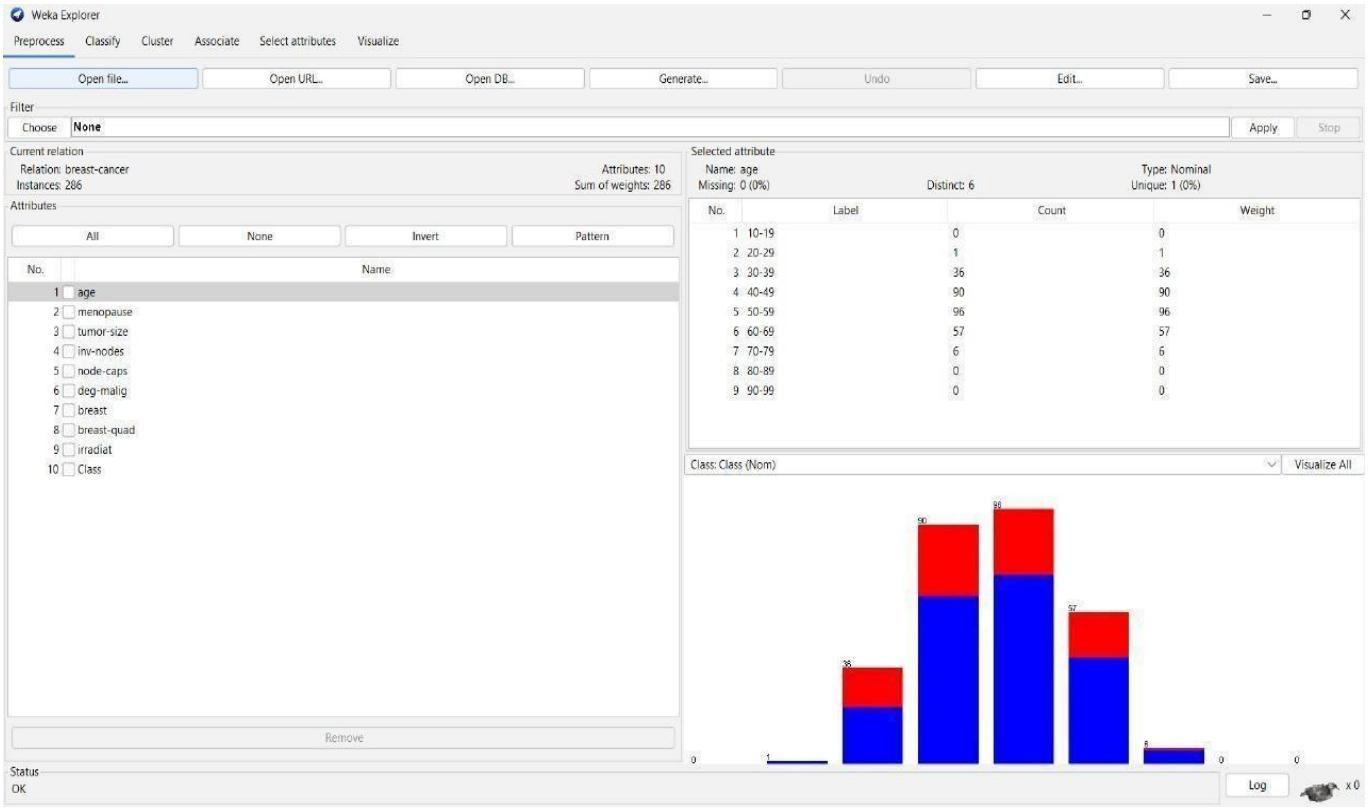
This screenshot shows the table format of dataset in which values are replaced by missing values.

Filter 5: Descritize

The Discretize filter in WEKA converts numeric attributes into nominal ones by dividing their range into intervals or bins. This is useful for algorithms that require categorical input or for simplifying data analysis. Binning can be done using equal-width or equal-frequency methods.

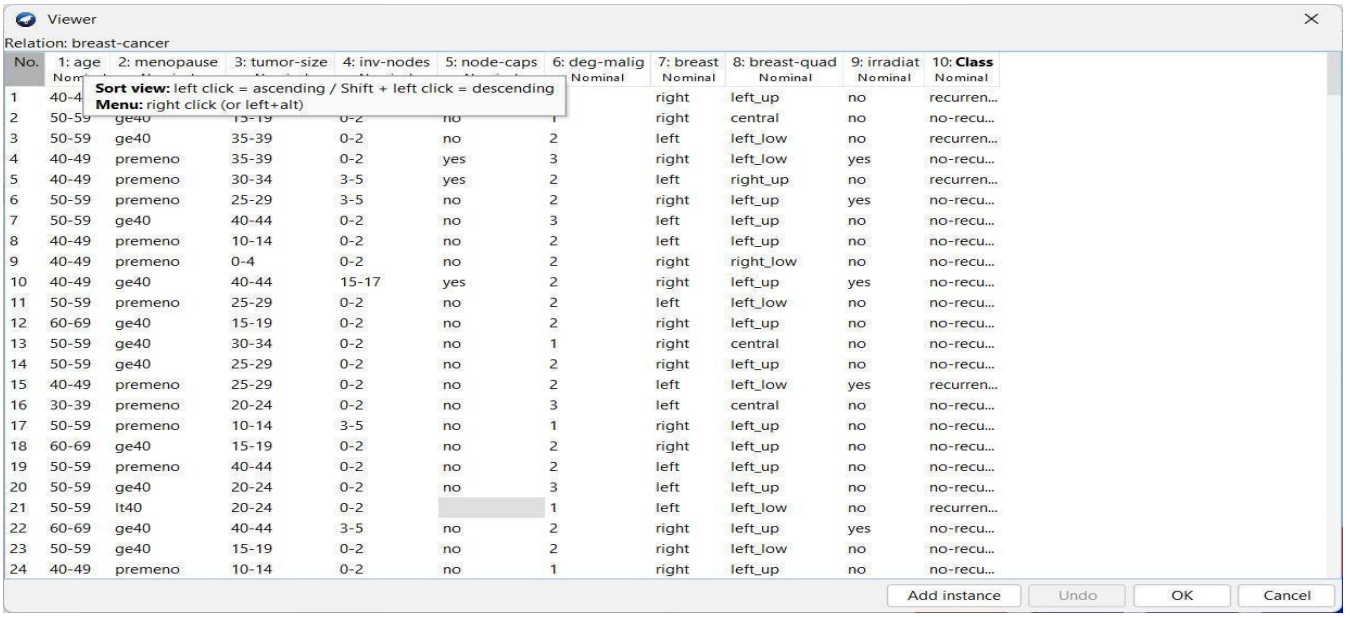
Dataset: breast-cancer.arff

Step-1: Upload dataset in Weka.



This image shows the Preprocess tab of the Weka Explorer interface, a data mining tool. The dataset in use is titled "breast-cancer", containing 286 instances and 10 attributes. The selected attribute is "age", which is of Nominal type with 6 distinct values (e.g., 30–39, 40–49, etc.).
On the right side, a bar chart visualizes the dstribution of the "age" attribute against the "Class" label (likely recurrence vs. no recurrence). Blue and red segments in bars represent the counts for different class labels. Most instances are concentrated in the 40–49, 50–59, and 60–69 age ranges.

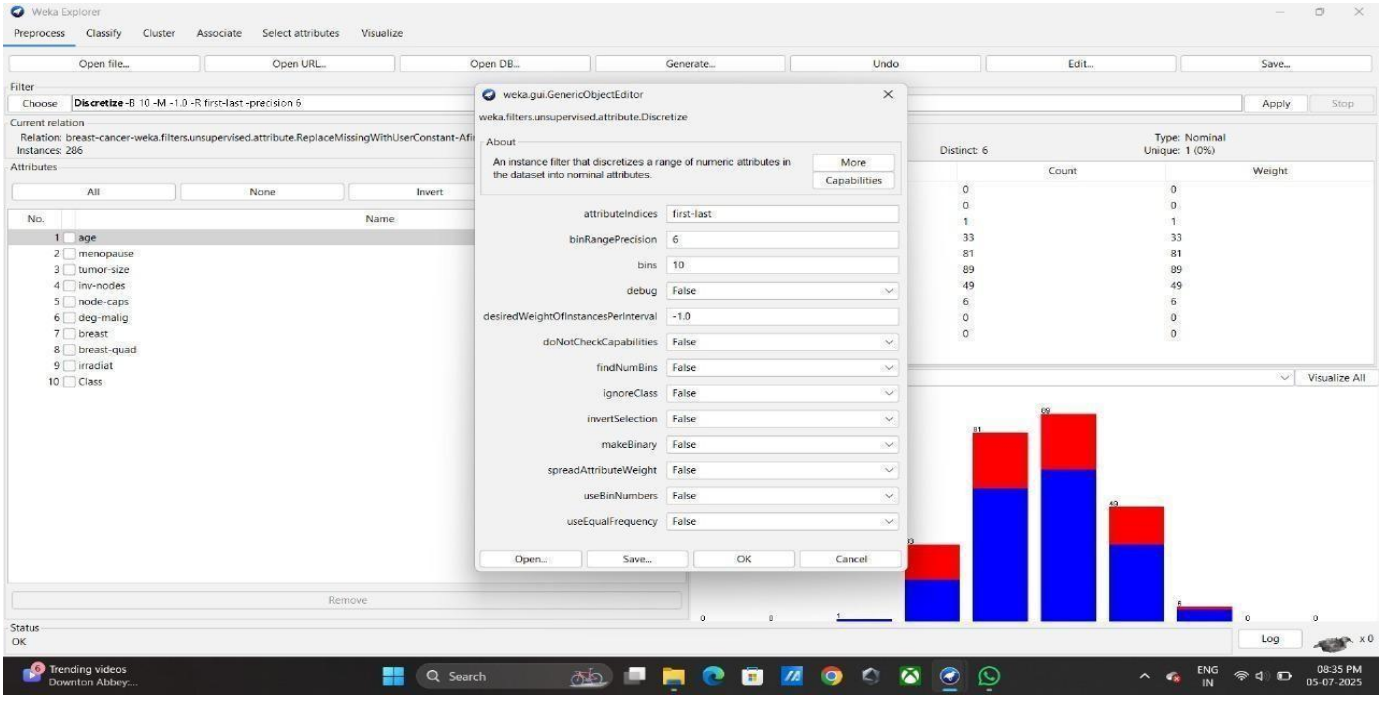
Step-2: dataset in table format.



No.	1: age	2: menopause	3: tumor-size	4: inv-nodes	5: node-caps	6: deg-malig	7: breast	8: breast-quad	9: irradiat	10: Class
	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal
1	40-4						right	left_up	no	recurren...
2	50-59	ge40	15-19	0-2	no	1	right	central	no	no-recu...
3	50-59	ge40	35-39	0-2	no	2	left	left_low	no	recurren...
4	40-49	premeno	35-39	0-2	yes	3	right	left_low	yes	no-recu...
5	40-49	premeno	30-34	3-5	yes	2	left	right_up	no	recurren...
6	50-59	premeno	25-29	3-5	no	2	right	left_up	yes	no-recu...
7	50-59	ge40	40-44	0-2	no	3	left	left_up	no	no-recu...
8	40-49	premeno	10-14	0-2	no	2	left	left_up	no	no-recu...
9	40-49	premeno	0-4	0-2	no	2	right	right_low	no	no-recu...
10	40-49	ge40	40-44	15-17	yes	2	right	left_up	yes	no-recu...
11	50-59	premeno	25-29	0-2	no	2	left	left_low	no	no-recu...
12	60-69	ge40	15-19	0-2	no	2	right	left_up	no	no-recu...
13	50-59	ge40	30-34	0-2	no	1	right	central	no	no-recu...
14	50-59	ge40	25-29	0-2	no	2	right	left_up	no	no-recu...
15	40-49	premeno	25-29	0-2	no	2	left	left_low	yes	recurren...
16	30-39	premeno	20-24	0-2	no	3	left	central	no	no-recu...
17	50-59	premeno	10-14	3-5	no	1	right	left_up	no	no-recu...
18	60-69	ge40	15-19	0-2	no	2	right	left_up	no	no-recu...
19	50-59	premeno	40-44	0-2	no	2	left	left_up	no	no-recu...
20	50-59	ge40	20-24	0-2	no	3	left	left_up	no	no-recu...
21	50-59	lt40	20-24	0-2		1	left	left_low	no	recurren...
22	60-69	ge40	40-44	3-5	no	2	right	left_up	yes	no-recu...
23	50-59	ge40	15-19	0-2	no	2	right	left_low	no	no-recu...
24	40-49	premeno	10-14	0-2	no	1	right	left_up	no	no-recu...

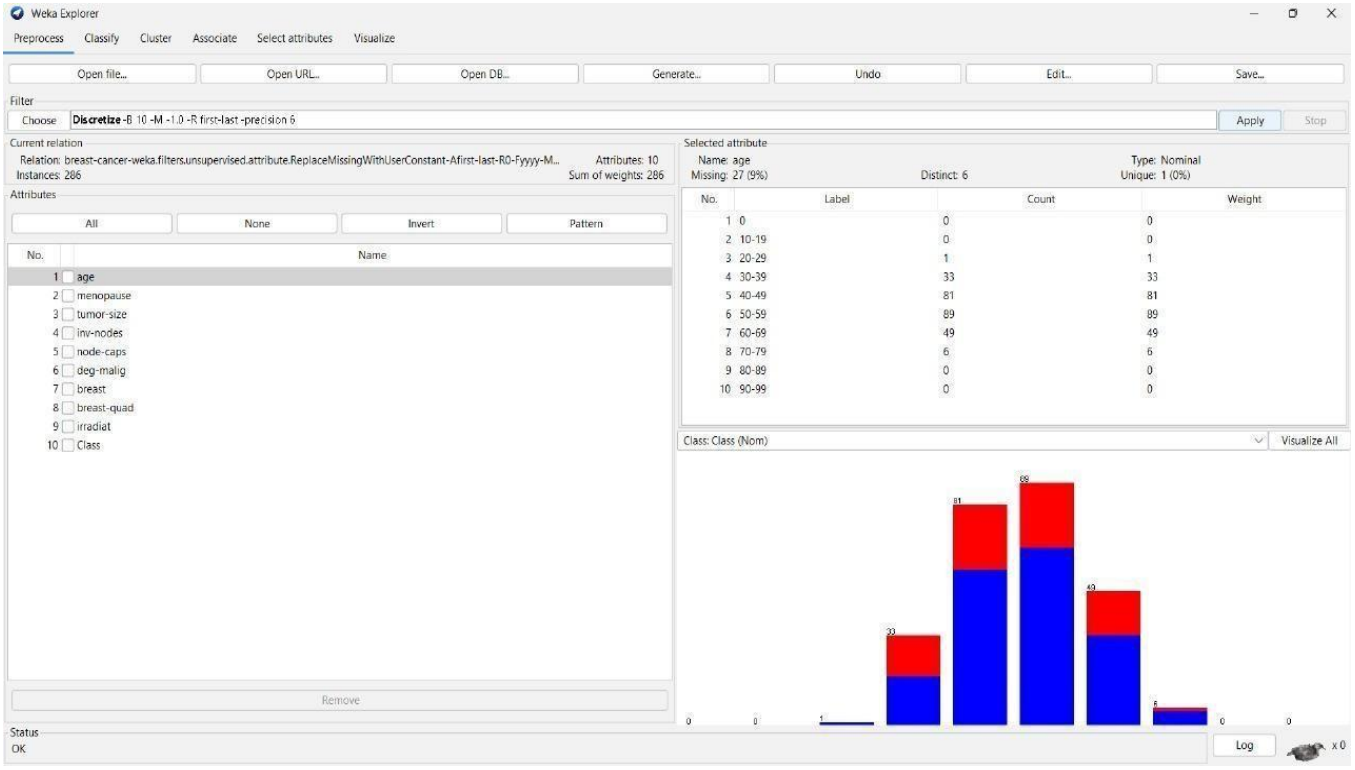
This screenshot shows the **tabular view of the labor.arff dataset** in Weka's **Instance Viewer**. Each row represents an instance (or record) from labor negotiations data, and each column is an attribute related to employment terms.

Step-3: Configure the parameters of Descritize filter.



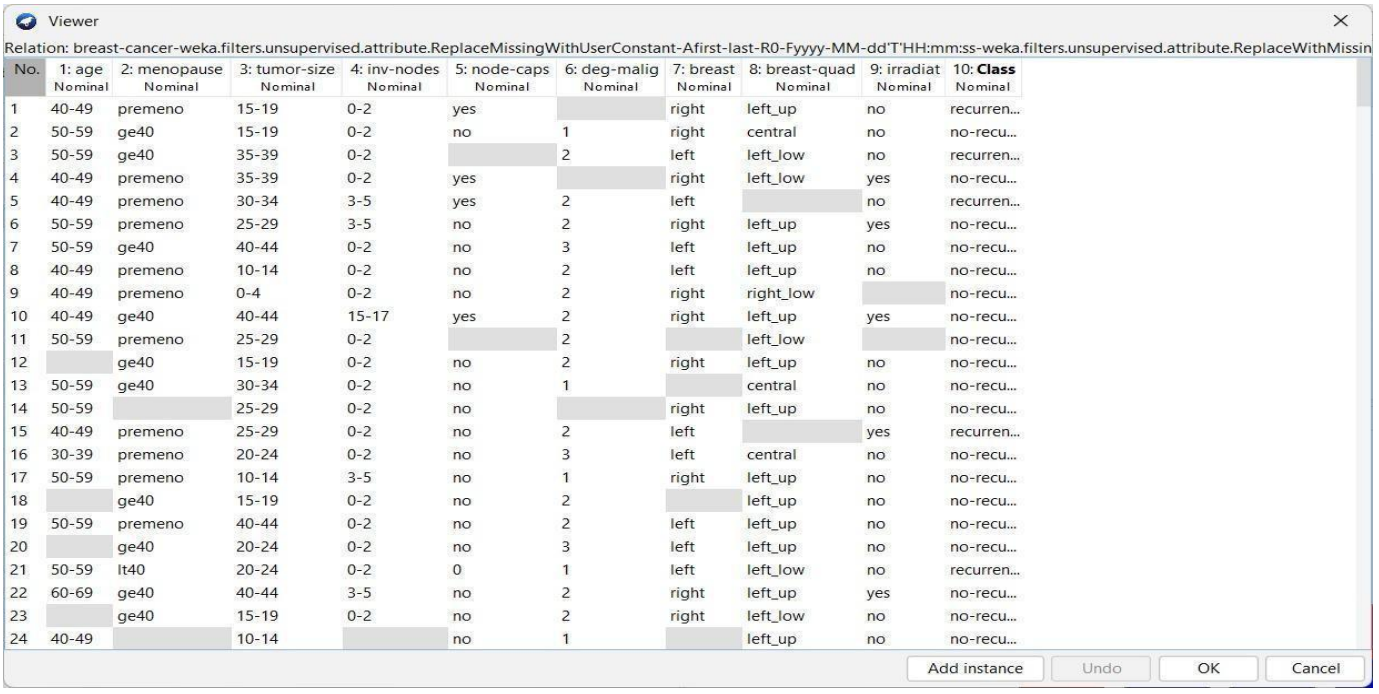
This image shows the use of the **Discretize** filter in Weka Explorer, which converts numeric attributes into nominal ones by dividing them into 10 bins with equal width, applied across all attributes (first-last).

Step-4: Apply descritize filter from to all attribute.



This Screenshot shows that descritize to all attributes and attributes are divided in 10 bins.

Step-5: Dataset in table format after applying Remove filter .



No.	1: age Nominal	2: menopause Nominal	3: tumor-size Nominal	4: inv-nodes Nominal	5: node-caps Nominal	6: deg-malig Nominal	7: breast Nominal	8: breast-quad Nominal	9: irradiat Nominal	10: **Class** Nominal
1	40-49	premeno	15-19	0-2	yes		right	left_up	no	recurren...
2	50-59	ge40	15-19	0-2	no	1	right	central	no	no-recu...
3	50-59	ge40	35-39	0-2		2	left	left_low	no	recurren...
4	40-49	premeno	35-39	0-2	yes		right	left_low	yes	no-recu...
5	40-49	premeno	30-34	3-5	yes	2	left		no	recurren...
6	50-59	premeno	25-29	3-5	no	2	right	left_up	yes	no-recu...
7	50-59	ge40	40-44	0-2	no	3	left	left_up	no	no-recu...
8	40-49	premeno	10-14	0-2	no	2	left	left_up	no	no-recu...
9	40-49	premeno	0-4	0-2	no	2	right	right_low		no-recu...
10	40-49	ge40	40-44	15-17	yes	2	right	left_up	yes	no-recu...
11	50-59	premeno	25-29	0-2		2		left_low		no-recu...
12		ge40	15-19	0-2	no	2	right	left_up	no	no-recu...
13	50-59	ge40	30-34	0-2	no	1		central	no	no-recu...
14	50-59		25-29	0-2	no		right	left_up	no	no-recu...
15	40-49	premeno	25-29	0-2	no	2	left		yes	recurren...
16	30-39	premeno	20-24	0-2	no	3	left	central	no	no-recu...
17	50-59	premeno	10-14	3-5	no	1	right	left_up	no	no-recu...
18		ge40	15-19	0-2	no	2		left_up	no	no-recu...
19	50-59	premeno	40-44	0-2	no	2	left	left_up	no	no-recu...
20		ge40	20-24	0-2	no	3	left	left_up	no	no-recu...
21	50-59	lt40	20-24	0-2	0	1	left	left_low	no	recurren...
22	60-69	ge40	40-44	3-5	no	2	right	left_up	yes	no-recu...
23		ge40	15-19	0-2	no	2	right	left_low	no	no-recu...
24	40-49		10-14		no	1		left_up	no	no-recu...

This screenshot shows the table format of dataset in which Range is provided to all attributes.

Experiment 5

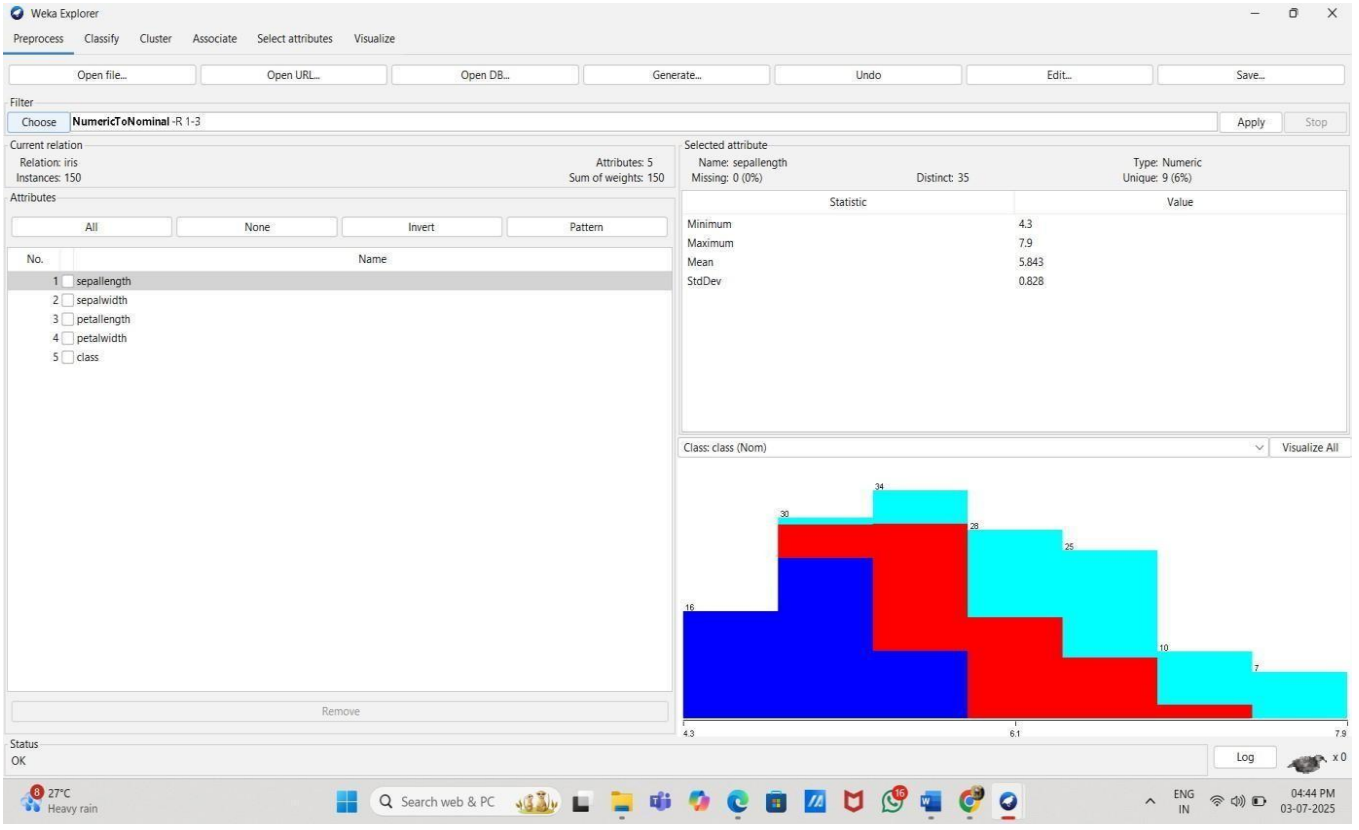
Title: Apply Preprocessing techniques on dataset using filters: NumericToNominal, StringToNominal, NominalToBinary, Normalize. Also do the result analysis before and after preprocessing.

Filter 1: NumericToNominal

The **NumericToNominal** filter in **Weka** is used to convert one or more **numeric attributes** in a dataset into **nominal (categorical)** attributes. This is useful when a numeric attribute actually represents categories (like codes for classes or labels), and should be treated as such during data mining or machine learning processes.

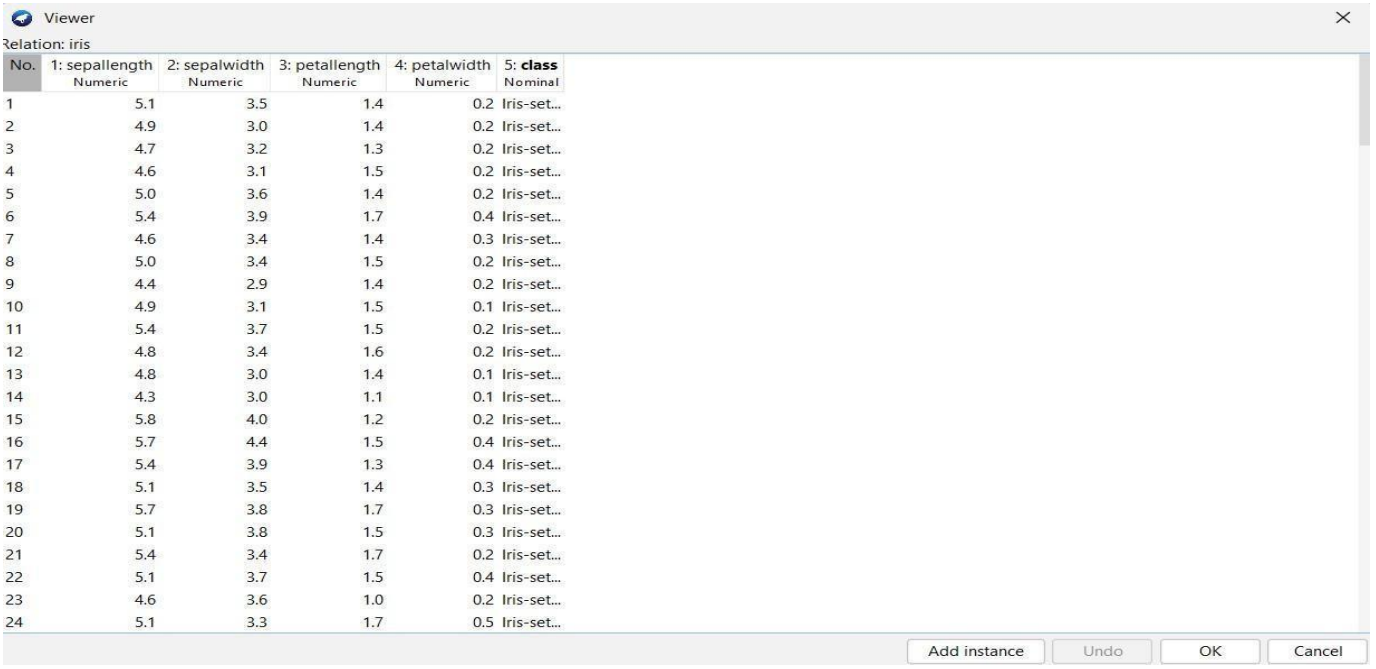
Dataset: iris.arff

Step-1: Upload dataset in Weka.



The iris.arff dataset is a well-known and widely used dataset in machine learning and pattern recognition. It consists of 150 instances, each representing a sample of an iris flower. The dataset includes five attributes: sepallength, sepalwidth, petallength, petalwidth, and class. The first four attributes are numeric and represent the physical dimensions of the flower's sepals and petals in centimeters. The fifth attribute, class, is nominal and indicates the species of the iris flower, which can be one of three categories: *Iris-setosa*, *Iris-versicolor*, or *Iris-virginica*.

Step-2: dataset in table format.



Viewer

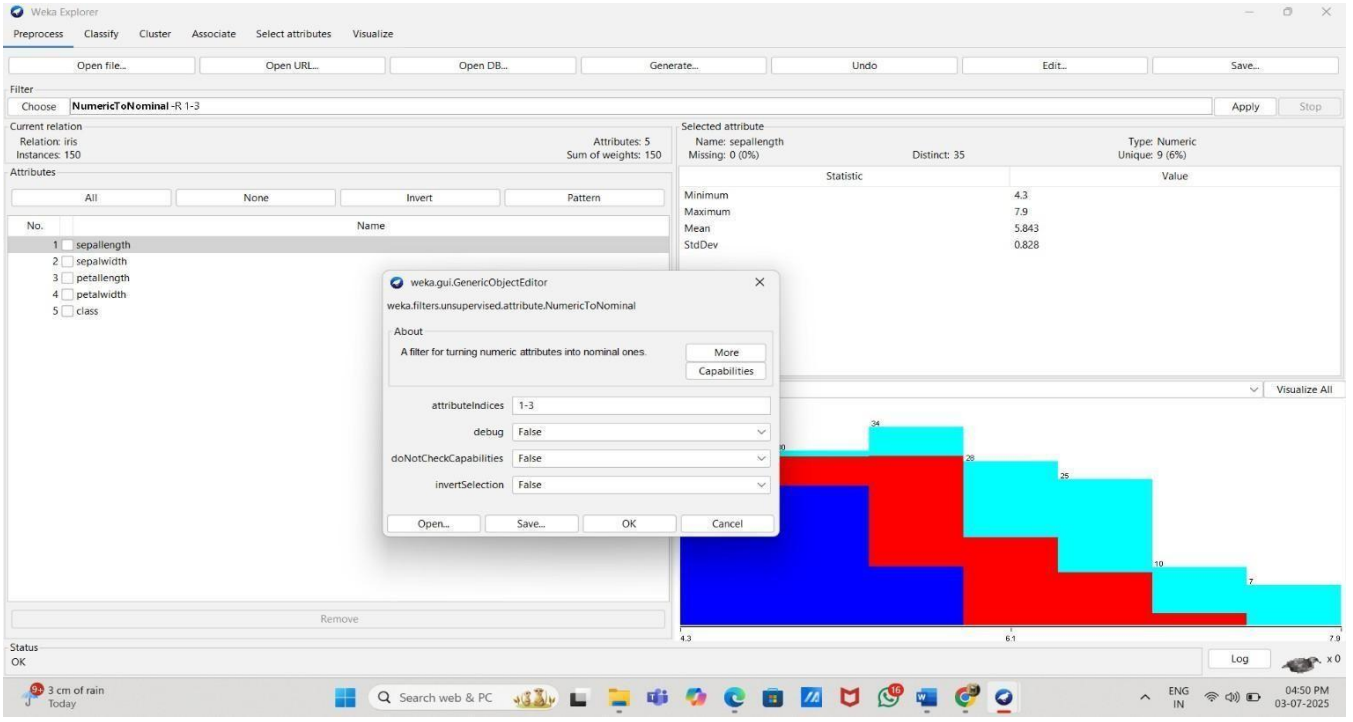
Relation: iris

No.	1: sepallength Numeric	2: sepalwidth Numeric	3: petallength Numeric	4: petalwidth Numeric	5: class Nominal
1	5.1	3.5	1.4	0.2	Iris-set...
2	4.9	3.0	1.4	0.2	Iris-set...
3	4.7	3.2	1.3	0.2	Iris-set...
4	4.6	3.1	1.5	0.2	Iris-set...
5	5.0	3.6	1.4	0.2	Iris-set...
6	5.4	3.9	1.7	0.4	Iris-set...
7	4.6	3.4	1.4	0.3	Iris-set...
8	5.0	3.4	1.5	0.2	Iris-set...
9	4.4	2.9	1.4	0.2	Iris-set...
10	4.9	3.1	1.5	0.1	Iris-set...
11	5.4	3.7	1.5	0.2	Iris-set...
12	4.8	3.4	1.6	0.2	Iris-set...
13	4.8	3.0	1.4	0.1	Iris-set...
14	4.3	3.0	1.1	0.1	Iris-set...
15	5.8	4.0	1.2	0.2	Iris-set...
16	5.7	4.4	1.5	0.4	Iris-set...
17	5.4	3.9	1.3	0.4	Iris-set...
18	5.1	3.5	1.4	0.3	Iris-set...
19	5.7	3.8	1.7	0.3	Iris-set...
20	5.1	3.8	1.5	0.3	Iris-set...
21	5.4	3.4	1.7	0.2	Iris-set...
22	5.1	3.7	1.5	0.4	Iris-set...
23	4.6	3.6	1.0	0.2	Iris-set...
24	5.1	3.3	1.7	0.5	Iris-set...

Add instance | Undo | OK | Cancel

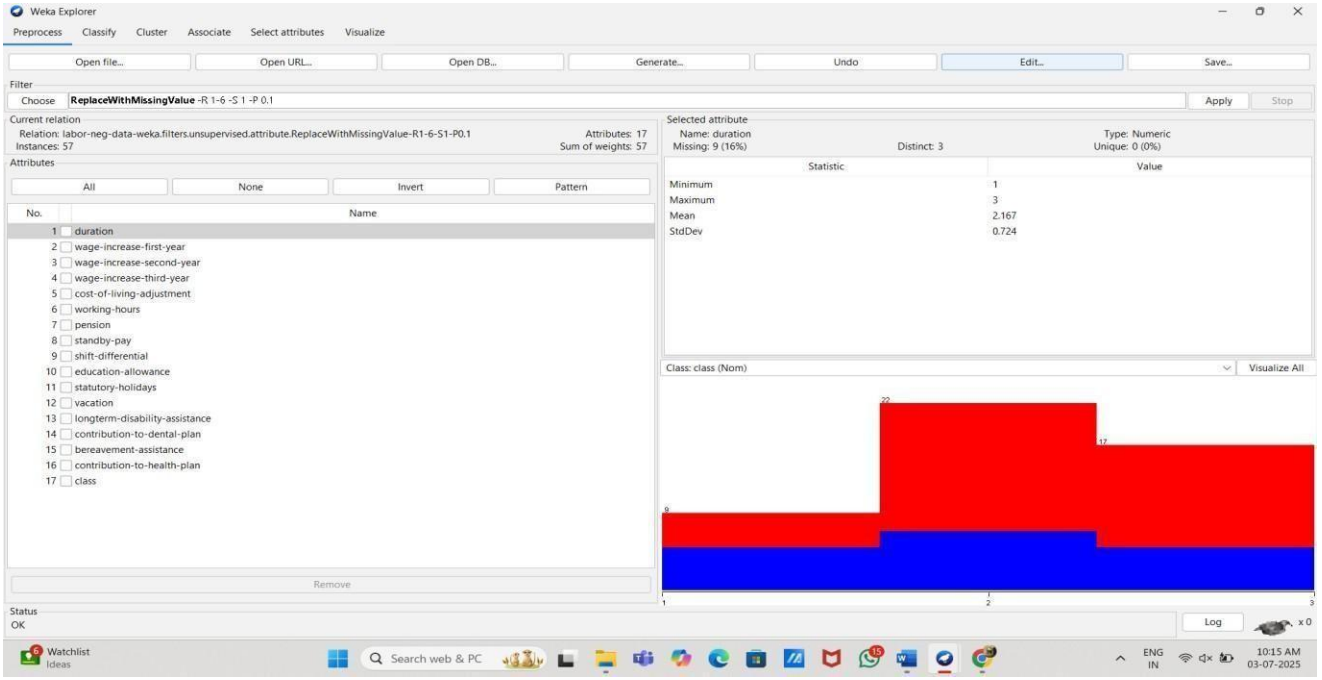
The screenshot shows the **tabular view** of the iris.arff dataset in Weka. It contains five attributes: sepallength, sepalwidth, petallength, petalwidth (all numeric), and class (nominal), which indicates the iris flower species such as *Iris-setosa*. Each row represents one flower instance with its measured values.

Step-3: Configure the parameters of NumericToNominal filter.



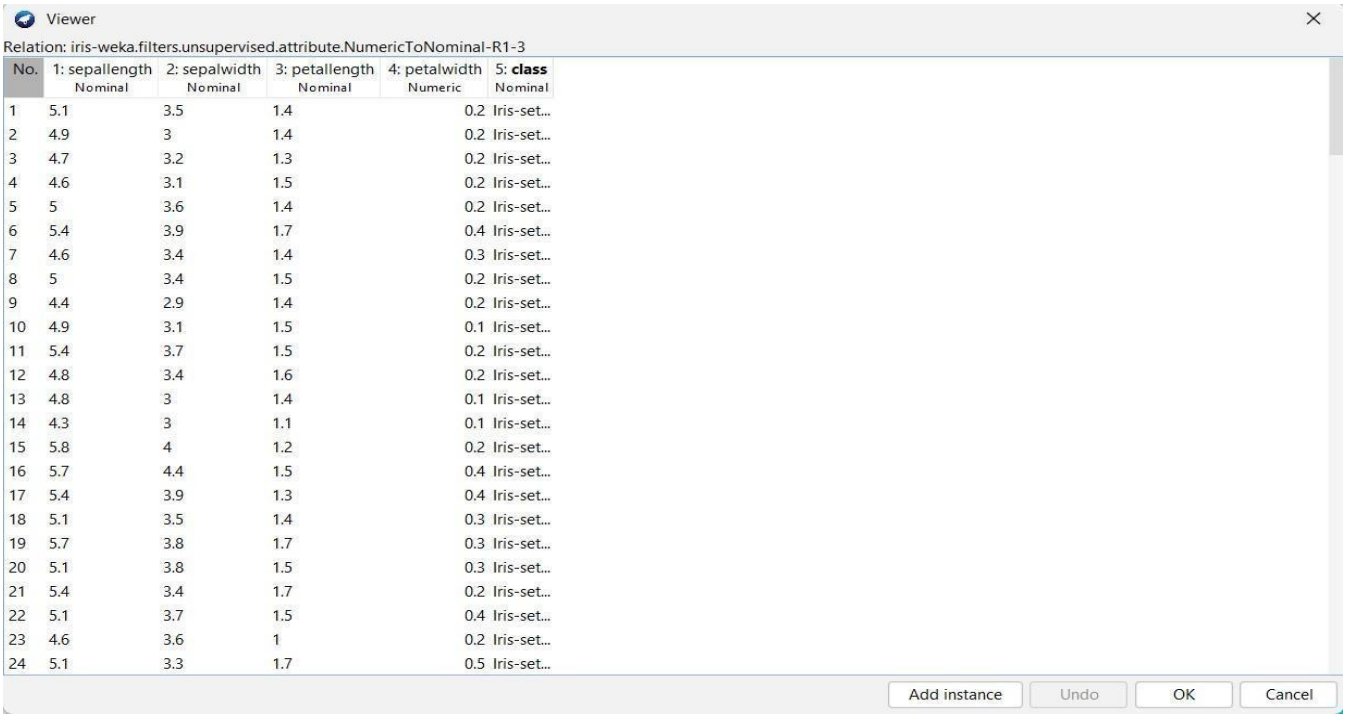
This image shows the Weka Explorer interface applying the NumericToNominal filter to convert numeric attributes (1–3) in the Iris dataset to nominal. A histogram and attribute statistics are also visible.

Step-4: Apply NumericToNominal filter from 1 to 3 attribute.



This image shows Weka Explorer after applying the NumericToNominal filter to attributes 1–3 of the Iris dataset. The sepallength attribute is now treated as nominal, with distinct value counts and a class distribution histogram displayed.

Step-5: Dataset in table format after applying NumericToNominal filter .



Viewer

Relation: iris-weka.filters.unsupervised.attribute.NumericToNominal-R1-3

No.	1: sepallength Nominal	2: sepalwidth Nominal	3: petallength Nominal	4: petalwidth Numeric	5: **class** Nominal
1	5.1	3.5	1.4	0.2	Iris-set...
2	4.9	3	1.4	0.2	Iris-set...
3	4.7	3.2	1.3	0.2	Iris-set...
4	4.6	3.1	1.5	0.2	Iris-set...
5	5	3.6	1.4	0.2	Iris-set...
6	5.4	3.9	1.7	0.4	Iris-set...
7	4.6	3.4	1.4	0.3	Iris-set...
8	5	3.4	1.5	0.2	Iris-set...
9	4.4	2.9	1.4	0.2	Iris-set...
10	4.9	3.1	1.5	0.1	Iris-set...
11	5.4	3.7	1.5	0.2	Iris-set...
12	4.8	3.4	1.6	0.2	Iris-set...
13	4.8	3	1.4	0.1	Iris-set...
14	4.3	3	1.1	0.1	Iris-set...
15	5.8	4	1.2	0.2	Iris-set...
16	5.7	4.4	1.5	0.4	Iris-set...
17	5.4	3.9	1.3	0.4	Iris-set...
18	5.1	3.5	1.4	0.3	Iris-set...
19	5.7	3.8	1.7	0.3	Iris-set...
20	5.1	3.8	1.5	0.3	Iris-set...
21	5.4	3.4	1.7	0.2	Iris-set...
22	5.1	3.7	1.5	0.4	Iris-set...
23	4.6	3.6	1	0.2	Iris-set...
24	5.1	3.3	1.7	0.5	Iris-set...

Add instance | Undo | OK | Cancel

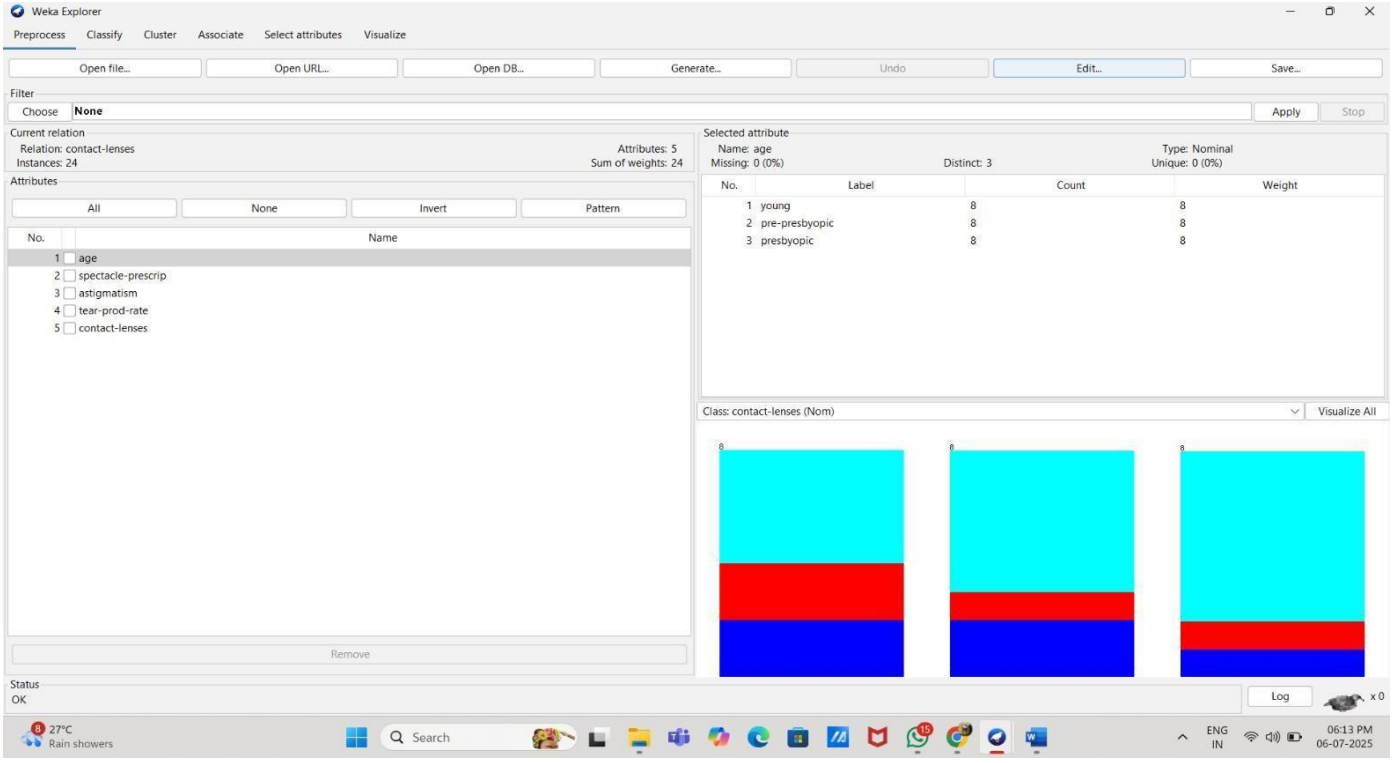
This screenshot shows the table format of dataset after applying NumericToNominal filter.

Filter 2: StringToNominal

The **StringToNominal** filter in Weka is used to convert string attributes into nominal attributes. It is particularly useful when the dataset contains categorical data represented as text (strings) that needs to be transformed into discrete values (nominal). This conversion helps in applying machine learning algorithms that require nominal data as input.

Dataset: contact-lenses.arff

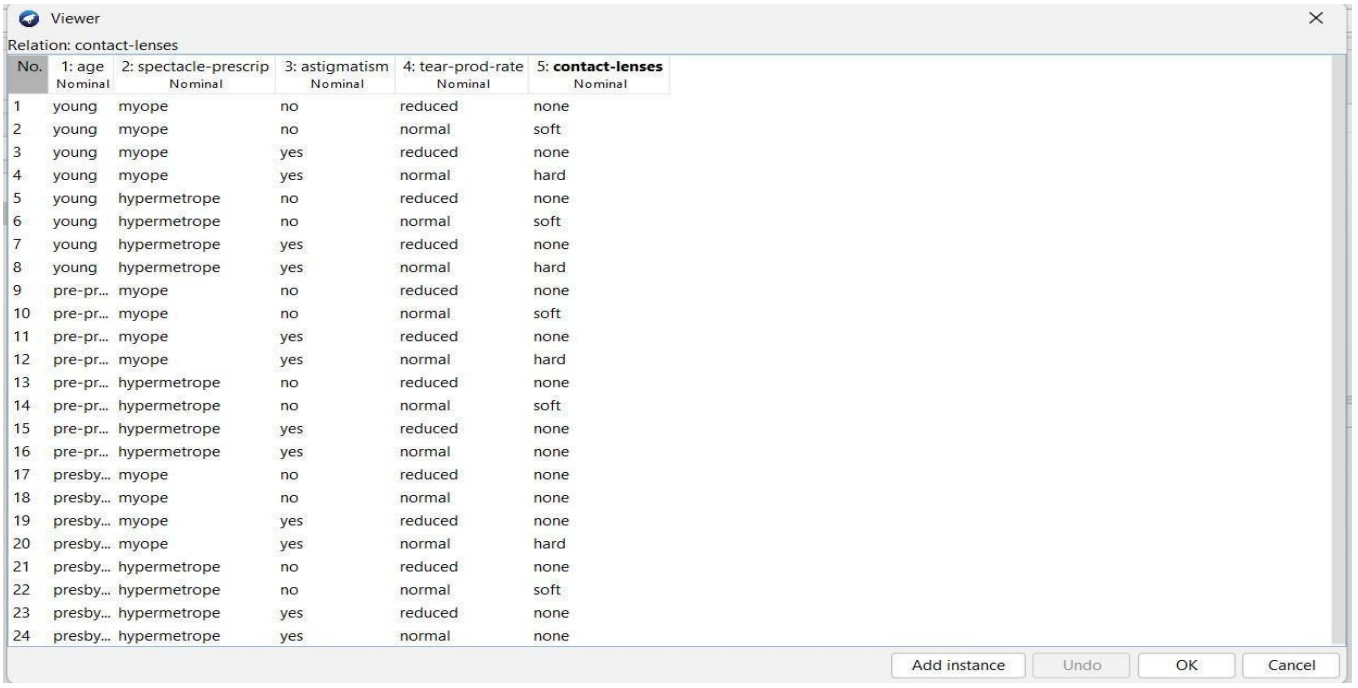
Step-1: Upload dataset in Weka.



This image shows the Preprocess tab of the Weka Explorer interface, where the "contact-lenses" dataset is currently loaded. The dataset contains 24 instances and 5 attributes: age, spectacle-prescrip, astigmatism, tear-prod-rate, and contact-lenses. The selected attribute in this view is "age," which is a nominal attribute with three distinct values: young, pre-presbyopic, and presbyopic. Each of these age categories contains an equal count of 8 instances, indicating a balanced distribution across the dataset.

On the right side, a detailed summary of the selected attribute is displayed, showing the label names, counts, and weights. Below that, a bar chart visualizes how the values of the class attribute "contact-lenses" are distributed across each age category. Each color in the bars represents a different class label (such as "no lenses," "soft," or "hard" lenses). The visualization allows users to observe how the lens recommendations vary based on age groups, providing insights into the relationship between age and lens type. This setup is part of the preprocessing phase in Weka, often used to explore and understand the structure of the dataset before applying machine learning algorithms.

Step-2: dataset in table format.



Viewer

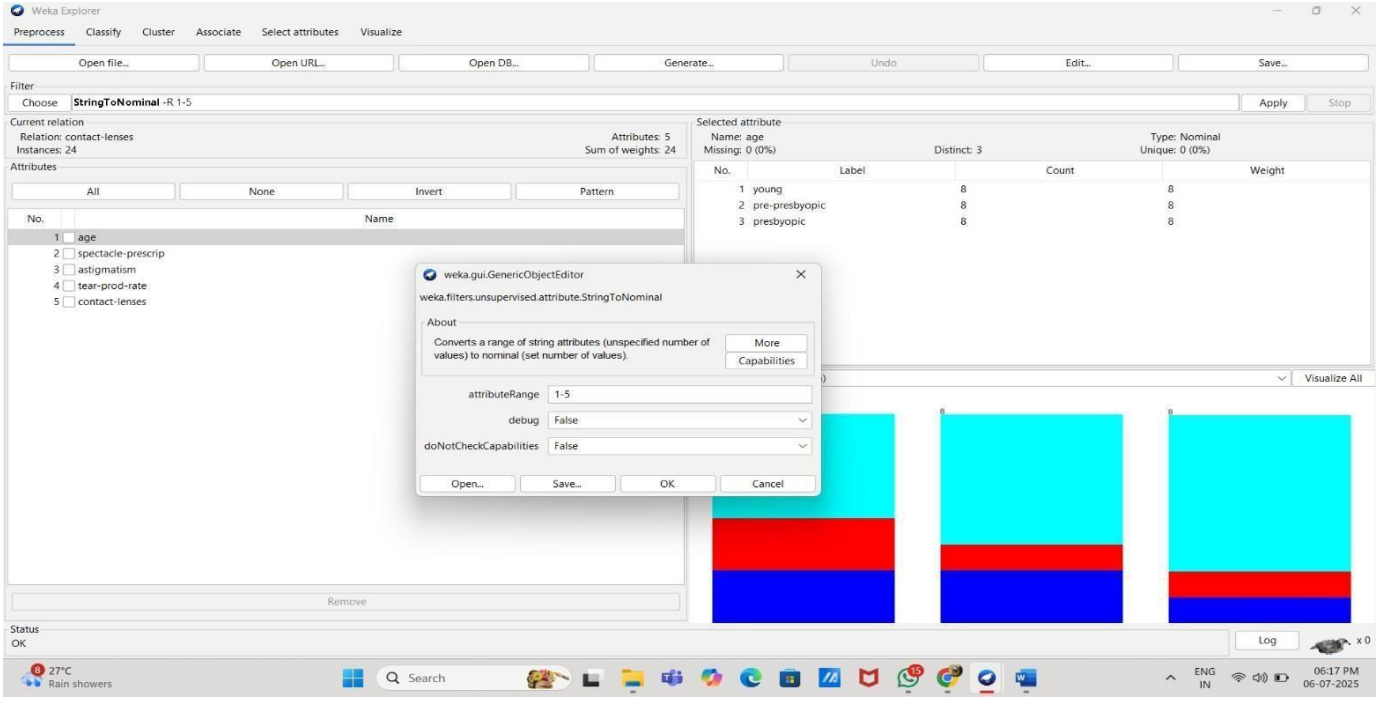
Relation: contact-lenses

No.	1: age Nominal	2: spectacle-prescrip Nominal	3: astigmatism Nominal	4: tear-prod-rate Nominal	5: **contact-lenses** Nominal
1	young	myope	no	reduced	none
2	young	myope	no	normal	soft
3	young	myope	yes	reduced	none
4	young	myope	yes	normal	hard
5	young	hypermetrope	no	reduced	none
6	young	hypermetrope	no	normal	soft
7	young	hypermetrope	yes	reduced	none
8	young	hypermetrope	yes	normal	hard
9	pre-pr...	myope	no	reduced	none
10	pre-pr...	myope	no	normal	soft
11	pre-pr...	myope	yes	reduced	none
12	pre-pr...	myope	yes	normal	hard
13	pre-pr...	hypermetrope	no	reduced	none
14	pre-pr...	hypermetrope	no	normal	soft
15	pre-pr...	hypermetrope	yes	reduced	none
16	pre-pr...	hypermetrope	yes	normal	none
17	presby...	myope	no	reduced	none
18	presby...	myope	no	normal	none
19	presby...	myope	yes	reduced	none
20	presby...	myope	yes	normal	hard
21	presby...	hypermetrope	no	reduced	none
22	presby...	hypermetrope	no	normal	soft
23	presby...	hypermetrope	yes	reduced	none
24	presby...	hypermetrope	yes	normal	none

Add instance | Undo | OK | Cancel

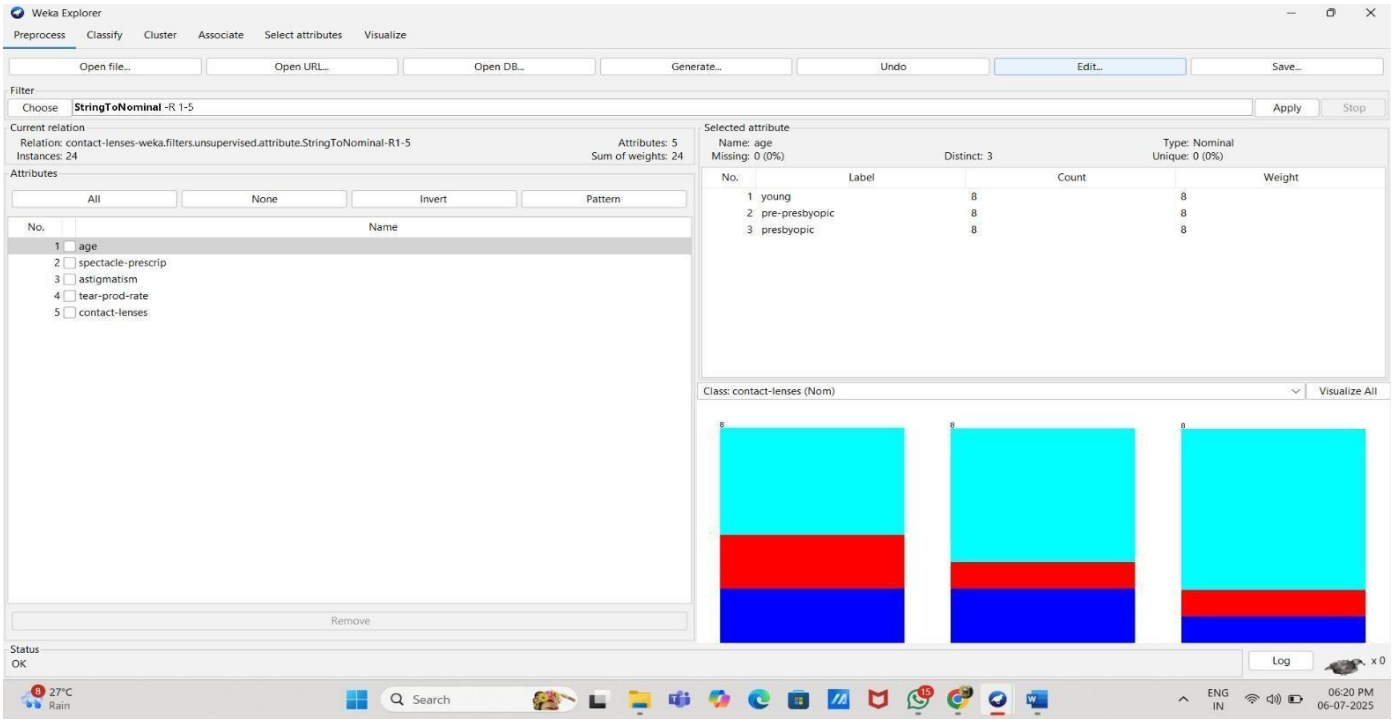
This image displays the data viewer in Weka for the "contact-lenses" dataset. It shows 24 instances with 5 nominal attributes: age, spectacle prescription, astigmatism, tear production rate, and contact lens recommendation. The table provides a clear view of how different attribute combinations influence the contact lens type prescribed (none, soft, or hard).

Step-3: Configure the parameters of StringToNominal filter.



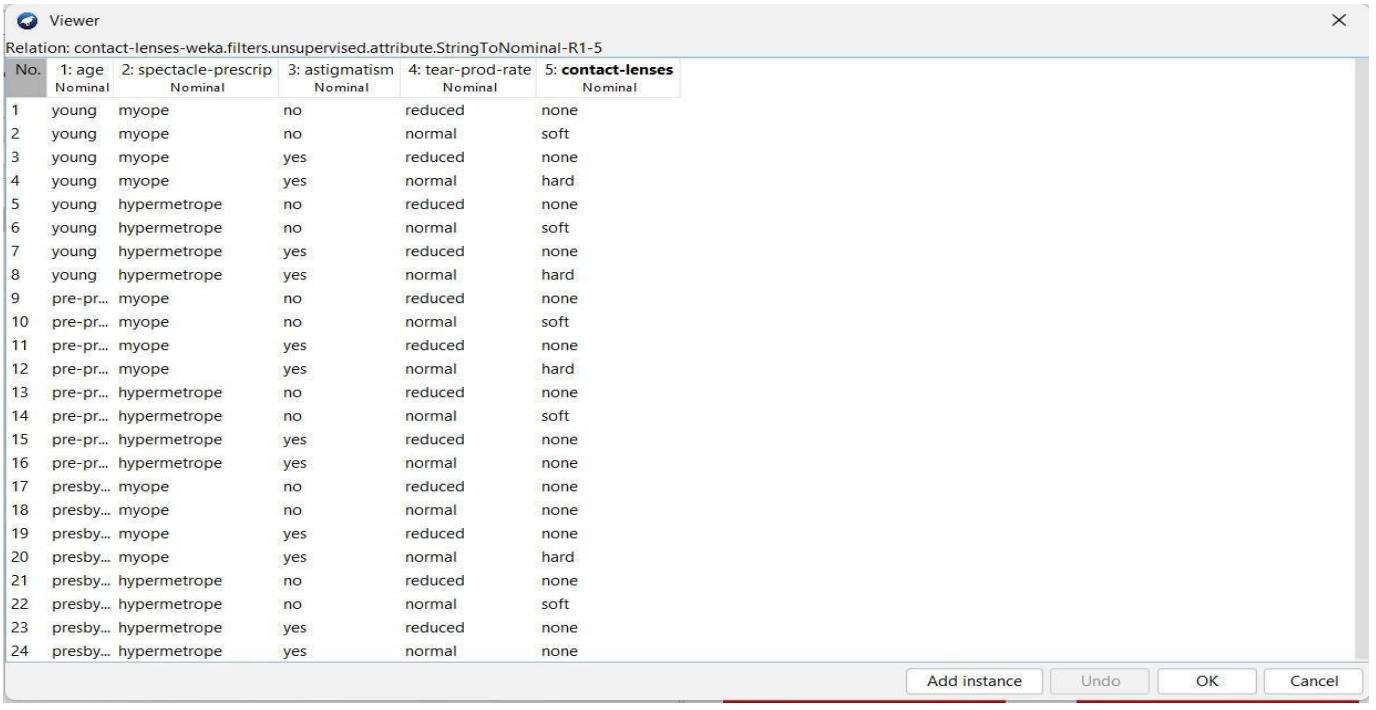
This image shows Weka Explorer with the **StringToNominal** filter selected. The filter is set to convert string attributes in the range **1–5** to nominal. A pop-up window displays filter settings, including attribute range and debugging options.

Step-4: Apply StringToNominal filter from 1 to 5 attribute.



This image shows the Weka Explorer after applying the **StringToNominal** filter to the contact-lenses dataset. All five attributes (age, spectacle-prescrip, astigmatism, tear-prod-rate, and contact-lenses) have been successfully converted to **nominal** type, enabling categorical data analysis. The visualization panel displays class distribution for the contact-lenses attribute across the three age groups (young, pre-presbyopic, and presbyopic), each with equal instance counts.

Step-5: Dataset in table format after applying StringToNominal filter .



No.	1: age Nominal	2: spectacle-prescrip Nominal	3: astigmatism Nominal	4: tear-prod-rate Nominal	5: **contact-lenses** Nominal
1	young	myope	no	reduced	none
2	young	myope	no	normal	soft
3	young	myope	yes	reduced	none
4	young	myope	yes	normal	hard
5	young	hypermetrope	no	reduced	none
6	young	hypermetrope	no	normal	soft
7	young	hypermetrope	yes	reduced	none
8	young	hypermetrope	yes	normal	hard
9	pre-pr...	myope	no	reduced	none
10	pre-pr...	myope	no	normal	soft
11	pre-pr...	myope	yes	reduced	none
12	pre-pr...	myope	yes	normal	hard
13	pre-pr...	hypermetrope	no	reduced	none
14	pre-pr...	hypermetrope	no	normal	soft
15	pre-pr...	hypermetrope	yes	reduced	none
16	pre-pr...	hypermetrope	yes	normal	none
17	presby...	myope	no	reduced	none
18	presby...	myope	no	normal	none
19	presby...	myope	yes	reduced	none
20	presby...	myope	yes	normal	hard
21	presby...	hypermetrope	no	reduced	none
22	presby...	hypermetrope	no	normal	soft
23	presby...	hypermetrope	yes	reduced	none
24	presby...	hypermetrope	yes	normal	none

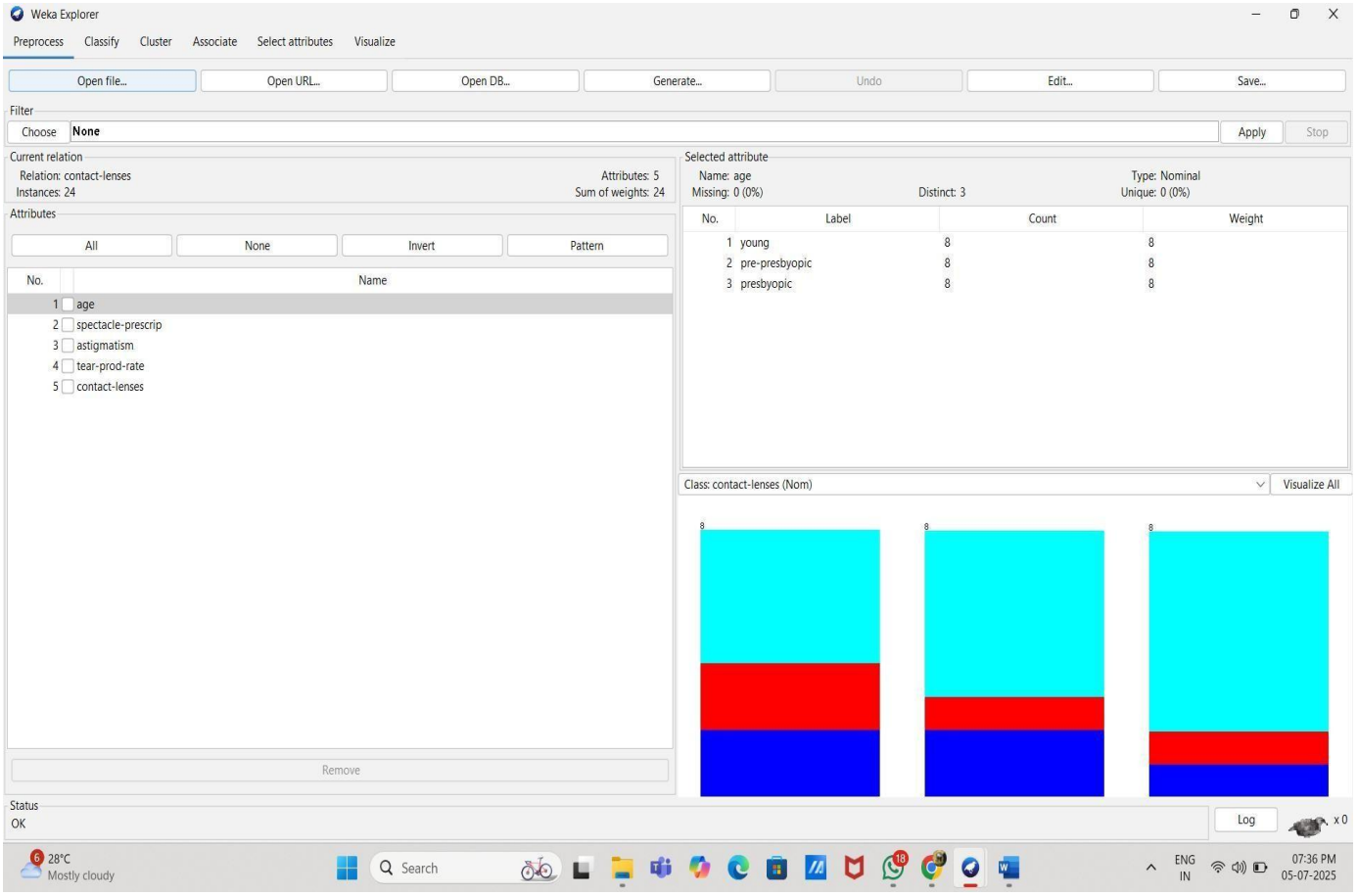
This screenshot shows the table format of dataset after applying StringToNominal filter.

Filter 3: NominalToBinary

The **NominalToBinary** filter in Weka is an unsupervised attribute filter that transforms nominal (categorical) attributes into binary (numeric) form. This process, known as one-hot encoding, creates a separate binary attribute for each possible value of a nominal attribute. For example, if an attribute "Color" has values like Red, Green, and Blue, the filter converts it into three new binary attributes: "Color=Red", "Color=Green", and "Color=Blue", with values of 0 or 1 indicating the presence of each category. This conversion is particularly useful for machine learning algorithms in Weka that require numerical input rather than categorical data.

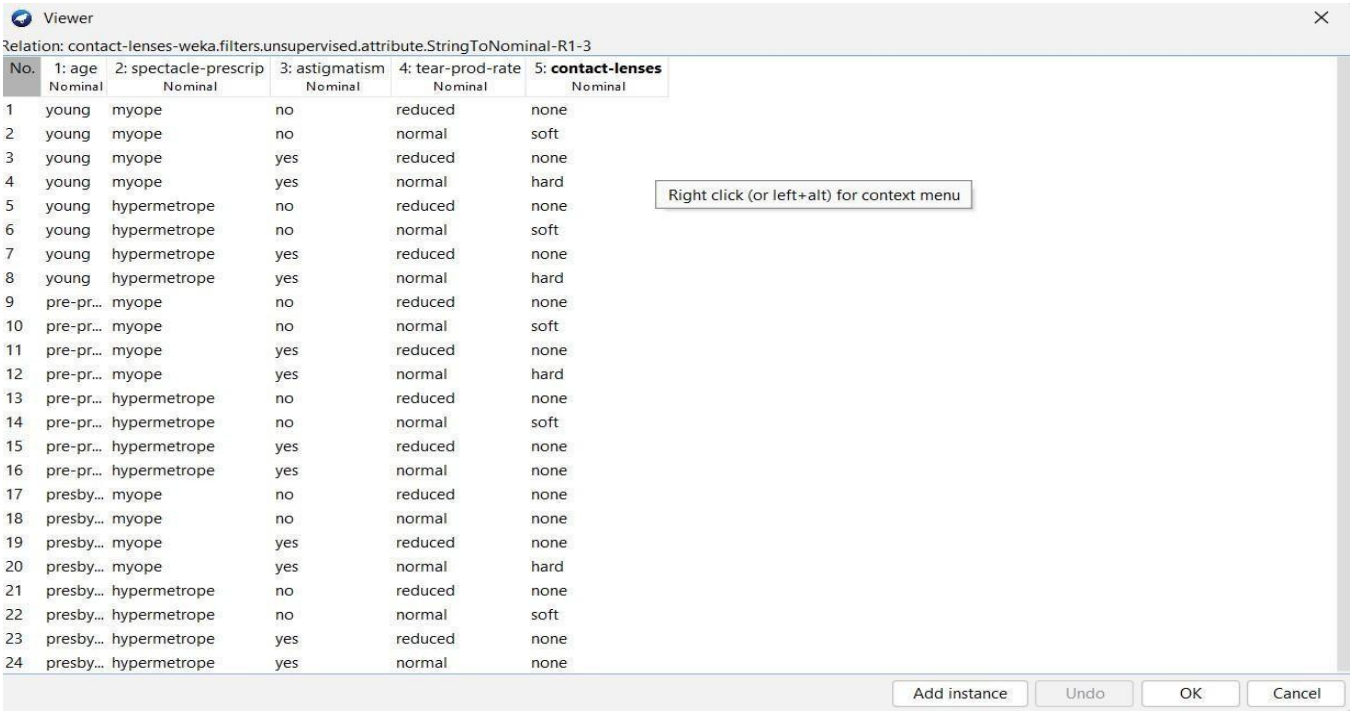
Dataset: contact-lenses.arff

Step-1: Upload dataset in Weka.



This image shows the **Weka Explorer – Preprocess tab** with the **contact-lenses** dataset loaded. It contains 5 nominal attributes and 24 instances. The attribute **"age"** is selected, showing three distinct values: *young*, *pre-presbyopic*, and *presbyopic*, each with 8 instances. A bar chart displays the distribution of the target class **contact-lenses** across the different age groups.

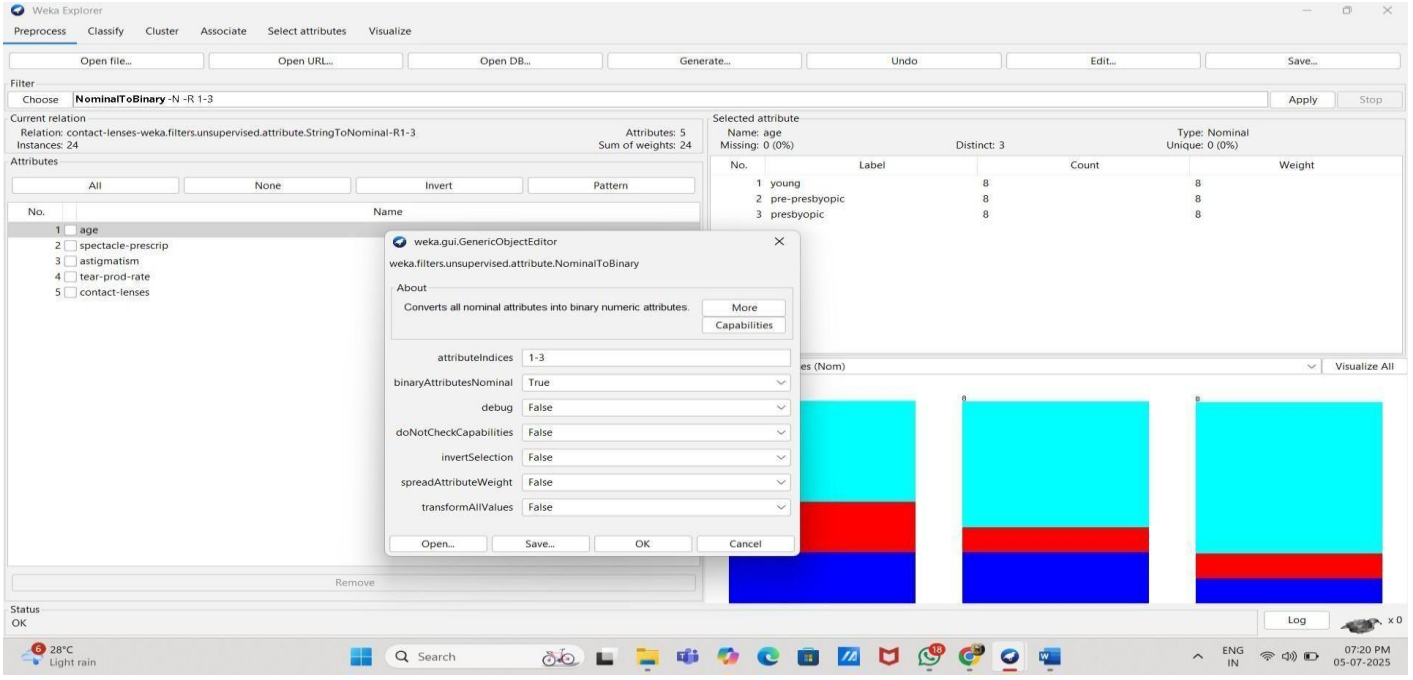
Step-2: dataset in table format.



No.	1: age Nominal	2: spectacle-prescrip Nominal	3: astigmatism Nominal	4: tear-prod-rate Nominal	5: **contact-lenses** Nominal
1	young	myope	no	reduced	none
2	young	myope	no	normal	soft
3	young	myope	yes	reduced	none
4	young	myope	yes	normal	hard
5	young	hypermetrope	no	reduced	none
6	young	hypermetrope	no	normal	soft
7	young	hypermetrope	yes	reduced	none
8	young	hypermetrope	yes	normal	hard
9	pre-pr...	myope	no	reduced	none
10	pre-pr...	myope	no	normal	soft
11	pre-pr...	myope	yes	reduced	none
12	pre-pr...	myope	yes	normal	hard
13	pre-pr...	hypermetrope	no	reduced	none
14	pre-pr...	hypermetrope	no	normal	soft
15	pre-pr...	hypermetrope	yes	reduced	none
16	pre-pr...	hypermetrope	yes	normal	none
17	presby...	myope	no	reduced	none
18	presby...	myope	no	normal	none
19	presby...	myope	yes	reduced	none
20	presby...	myope	yes	normal	hard
21	presby...	hypermetrope	no	reduced	none
22	presby...	hypermetrope	no	normal	soft
23	presby...	hypermetrope	yes	reduced	none
24	presby...	hypermetrope	yes	normal	none

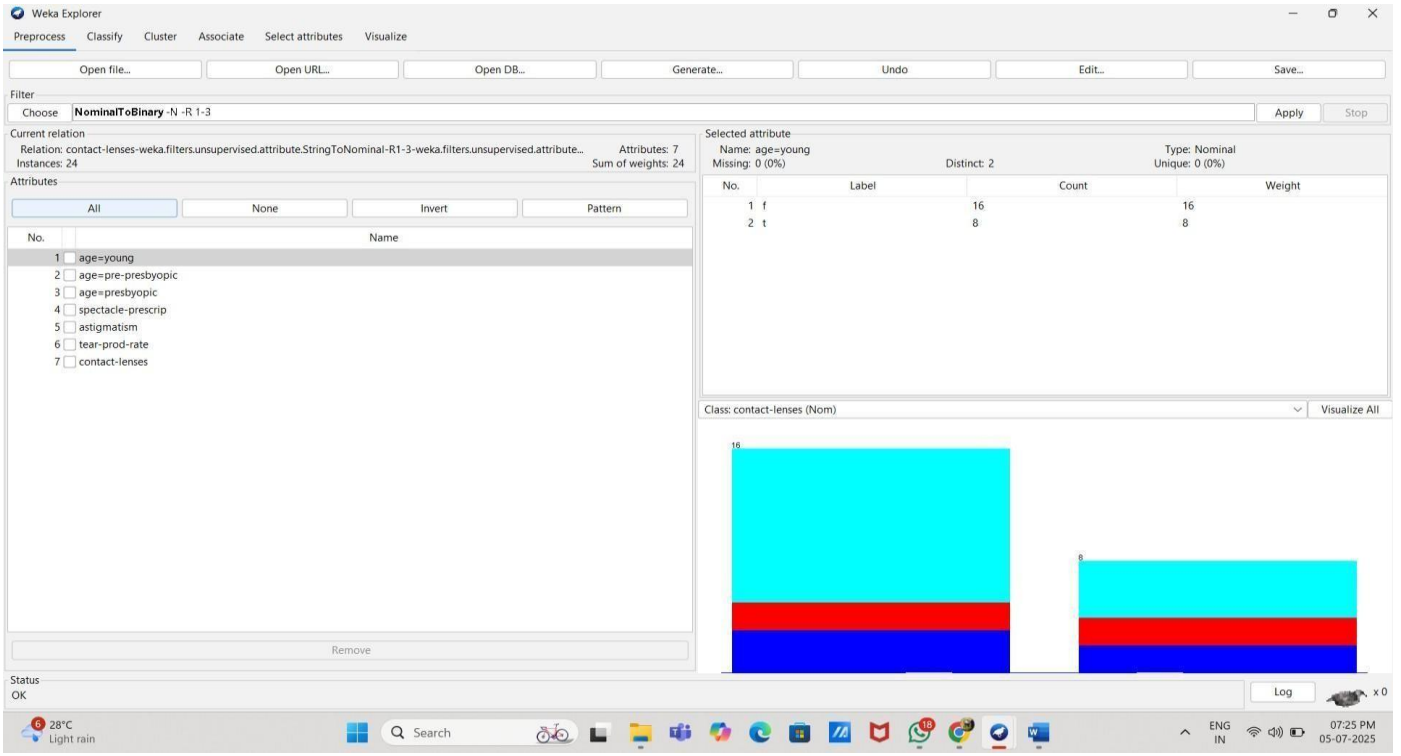
The screenshot shows the **tabular view** of the contact-lenses.arff dataset in Weka. It contains five attributes: age, spectacle-prescrip, astigmatism, tear-prod-rate and contact-lense. All the attributes having type of string.

Step-3: Configure the parameters of NominalToBinary filter.



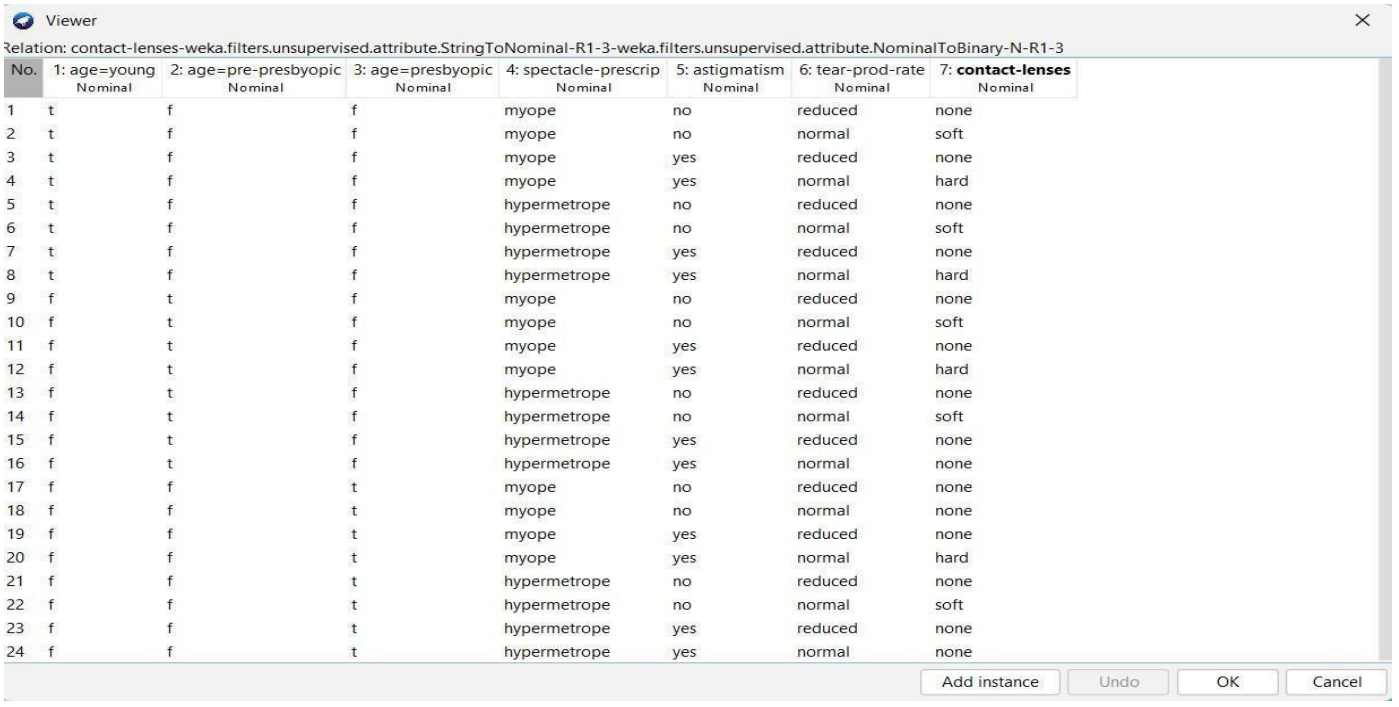
This image shows the **Weka Explorer** interface with the **NominalToBinary filter** selected. The filter is configured to convert nominal attributes in indices **1 to 3** into binary numeric attributes. The filter options window is open, allowing customization of parameters before applying the transformation to the dataset.

Step-4: Apply NominalToBinary filter from 1 to 3 attribute.



This image shows that first three attributes values are replaced by t(true) or f(false).

Step-5: Dataset in table format after applying NominalToBinary filter .



No.	1: age=young Nominal	2: age=pre-presbyopic Nominal	3: age=presbyopic Nominal	4: spectacle-prescrip Nominal	5: astigmatism Nominal	6: tear-prod-rate Nominal	7: **contact-lenses** Nominal
1	t	f	f	myope	no	reduced	none
2	t	f	f	myope	no	normal	soft
3	t	f	f	myope	yes	reduced	none
4	t	f	f	myope	yes	normal	hard
5	t	f	f	hypermetrope	no	reduced	none
6	t	f	f	hypermetrope	no	normal	soft
7	t	f	f	hypermetrope	yes	reduced	none
8	t	f	f	hypermetrope	yes	normal	hard
9	f	t	f	myope	no	reduced	none
10	f	t	f	myope	no	normal	soft
11	f	t	f	myope	yes	reduced	none
12	f	t	f	myope	yes	normal	hard
13	f	t	f	hypermetrope	no	reduced	none
14	f	t	f	hypermetrope	no	normal	soft
15	f	t	f	hypermetrope	yes	reduced	none
16	f	t	f	hypermetrope	yes	normal	none
17	f	f	t	myope	no	reduced	none
18	f	f	t	myope	no	normal	none
19	f	f	t	myope	yes	reduced	none
20	f	f	t	myope	yes	normal	hard
21	f	f	t	hypermetrope	no	reduced	none
22	f	f	t	hypermetrope	no	normal	soft
23	f	f	t	hypermetrope	yes	reduced	none
24	f	f	t	hypermetrope	yes	normal	none

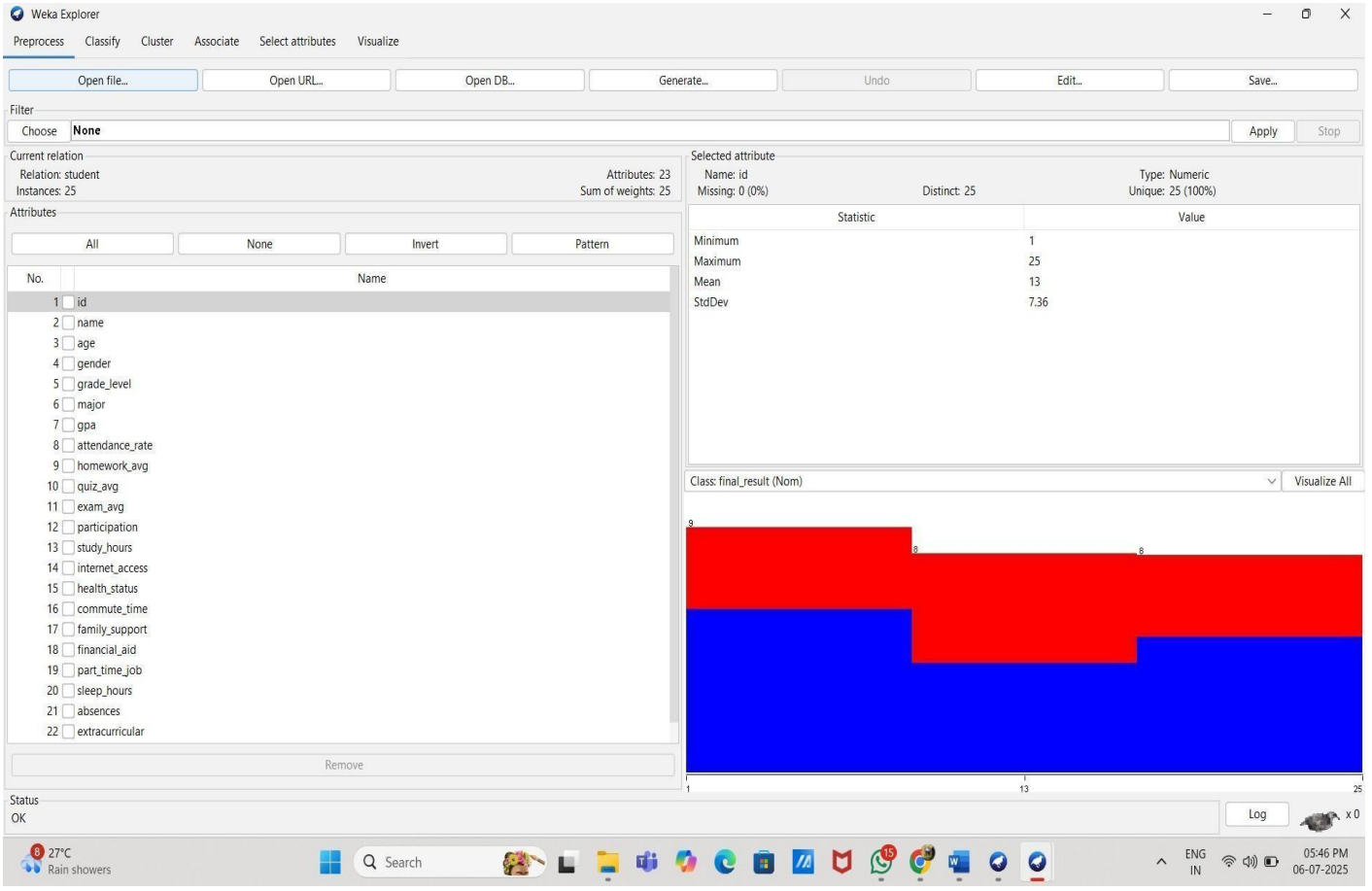
This screenshot shows the table format of dataset after applying NominalToBinary filter.

Filter 4: Normalize

The **Normalize** filter in Weka is a data preprocessing tool that scales numeric attribute values to a specified range, typically **[0, 1]**. This is useful for improving the performance of machine learning algorithms that are sensitive to the scale of input data, such as k-NN or neural networks. It ensures that all numeric attributes contribute equally to the model.

Dataset: student.arff

Step-1: Upload dataset in Weka.



This image shows the **Preprocess tab** of the Weka Explorer with the **student dataset** loaded. It contains 25 instances and 23 attributes. The selected attribute is **"id"**, which is numeric with distinct values ranging from 1 to 25. At the bottom, a **class distribution histogram** for the attribute **"final_result"** (a nominal class) is displayed in red and blue, indicating the frequency of each class value.

Step-2: dataset in table format.

Viewer

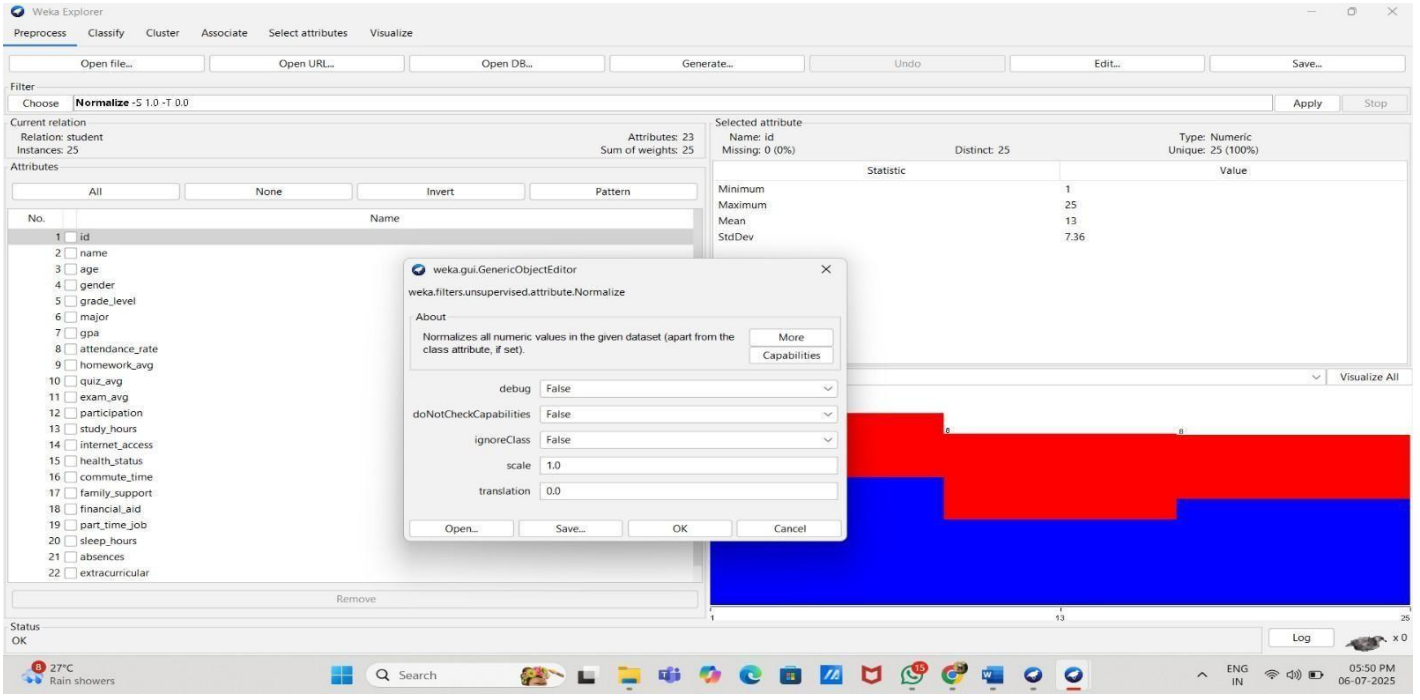
Relation: student

No.	1: id Numeric	2: name String	3: age Numeric	4: gender Nominal	5: grade_level Nominal	6: major String	7: gpa Numeric	8: attendance_rate Numeric	9: homework_avg Numeric	10: quiz_avg Numeric	11: exam_avg Numeric	12: participation Nominal	13: study_hours Numeric
1	1.0	Alice	18.0	female	freshman	Comput...	3.6	95.0	92.0	88.0	91.0	high	3.5
2	2.0	Bob	19.0	male	freshman	Mechan...	2.4	75.0	65.0	60.0	62.0	medium	2.0
3	3.0	Cara	20.0	female	sophomore	Math	3.1	85.0	78.0	81.0	84.0	high	4.0
4	4.0	David	21.0	male	junior	Physics	2.8	80.0	72.0	70.0	69.0	medium	3.0
5	5.0	Emma	22.0	female	senior	Biology	3.9	98.0	95.0	94.0	93.0	high	5.5
6	6.0	Frank	20.0	male	sophomore	History	2.7	78.0	68.0	65.0	66.0	low	1.5
7	7.0	Grace	18.0	female	freshman	English	3.5	92.0	89.0	91.0	90.0	high	3.8
8	8.0	Hank	19.0	male	freshman	IT	3.0	84.0	80.0	78.0	79.0	medium	3.2
9	9.0	Ivy	20.0	female	sophomore	Psychol...	3.2	86.0	82.0	85.0	83.0	high	4.5
10	10.0	Jake	21.0	male	junior	Business	2.5	70.0	60.0	55.0	58.0	low	2.0
11	11.0	Karen	22.0	female	senior	Chemist...	3.7	96.0	93.0	90.0	91.0	high	5.0
12	12.0	Leo	20.0	male	sophomore	Math	2.9	77.0	70.0	68.0	69.0	medium	2.5
13	13.0	Mia	18.0	female	freshman	CS	3.4	90.0	88.0	85.0	87.0	high	3.7
14	14.0	Ned	19.0	male	freshman	Electro...	2.6	73.0	65.0	62.0	63.0	medium	2.0
15	15.0	Olive	21.0	female	junior	IT	3.3	88.0	85.0	82.0	83.0	high	4.0
16	16.0	Paul	22.0	male	senior	Finance	2.7	76.0	68.0	66.0	67.0	low	2.2
17	17.0	Quinn	20.0	female	sophomore	Biology	3.6	94.0	90.0	89.0	90.0	high	4.8
18	18.0	Ray	21.0	male	junior	History	2.8	79.0	70.0	71.0	70.0	medium	3.0
19	19.0	Sophia	22.0	female	senior	Psychol...	3.8	97.0	94.0	93.0	94.0	high	5.2
20	20.0	Tom	20.0	male	sophomore	Physics	2.9	81.0	74.0	72.0	73.0	medium	3.1
21	21.0	Uma	18.0	female	freshman	English	3.5	91.0	88.0	86.0	87.0	high	4.0
22	22.0	Victor	19.0	male	freshman	CS	2.3	68.0	60.0	59.0	60.0	low	1.8
23	23.0	Wendy	20.0	female	sophomore	Math	3.1	85.0	81.0	80.0	81.0	medium	3.6

Add instance | Undo | OK | Cancel

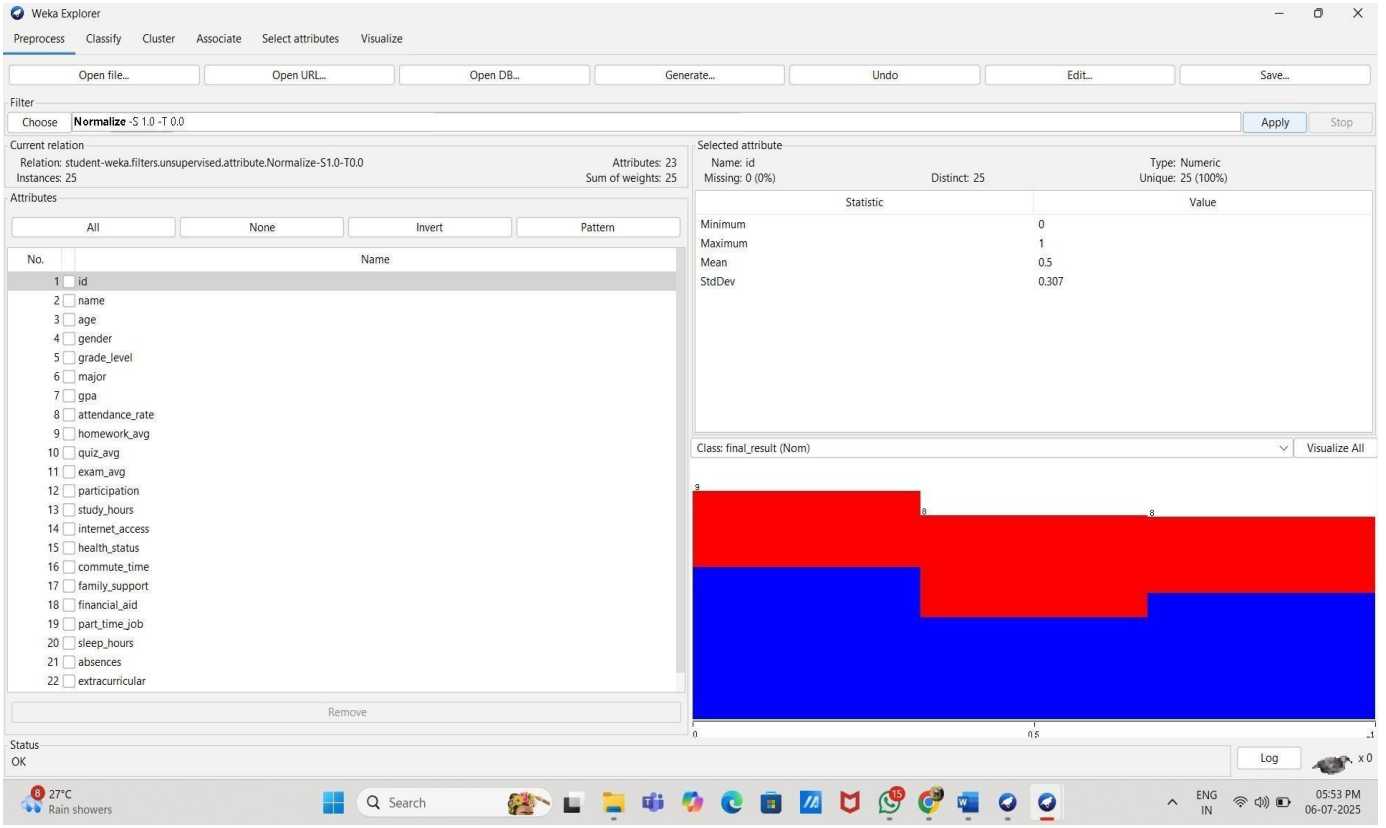
This image shows the **data viewer** window in Weka, displaying the **student dataset** with 23 attributes, including both numeric and nominal types such as id, name, age, gender, gpa, and participation. Each row represents a student instance with corresponding values.

Step-3: Configure the parameters of Normalize filter.



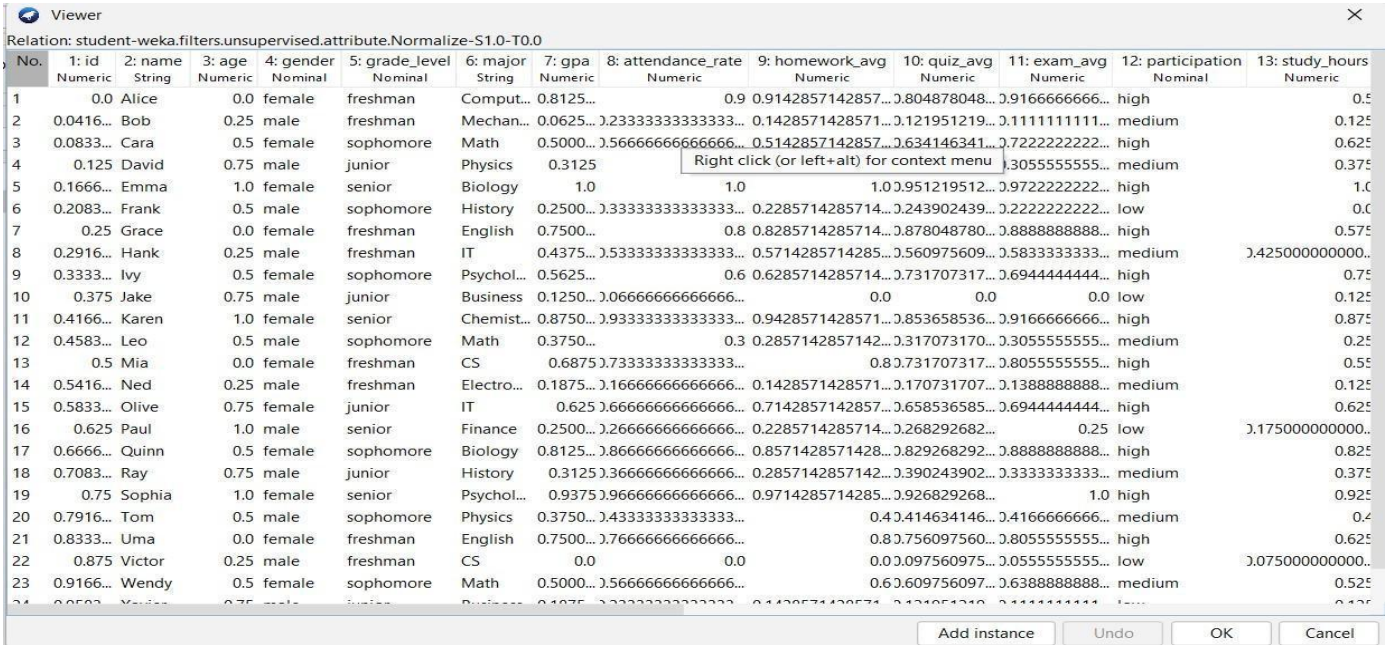
This image shows the Normalize filter settings in Weka, where numeric attributes are scaled between 0.0 and 1.0 using the weka.filters.unsupervised.attribute.Normalize filter.

Step-4: Apply Normalize filter from to all attributes.



This screenshot shows the Weka Explorer after applying the Normalize filter, where all numeric attributes have been scaled to the [0, 1] range.

Step-5: Dataset in table format after applying Normalize filter .



No.	1: id Numeric	2: name String	3: age Numeric	4: gender Nominal	5: grade_level Nominal	6: major String	7: gpa Numeric	8: attendance_rate Numeric	9: homework_avg Numeric	10: quiz_avg Numeric	11: exam_avg Numeric	12: participation Nominal	13: study_hours Numeric
1	0.0	Alice	0.0	female	freshman	Comput...	0.8125...	0.9	0.9142857142857...	0.804878048...	0.9166666666...	high	0.5
2	0.0416...	Bob	0.25	male	freshman	Mechan...	0.0625...	0.23333333333333...	0.1428571428571...	0.121951219...	0.1111111111...	medium	0.125
3	0.0833...	Cara	0.5	female	sophomore	Math	0.5000...	0.56666666666666...	0.5142857142857...	0.634146341...	0.7222222222...	high	0.625
4	0.125	David	0.75	male	junior	Physics	0.3125				0.3055555555...	medium	0.375
5	0.1666...	Emma	1.0	female	senior	Biology	1.0	1.0	1.0	0.951219512...	0.9722222222...	high	1.0
6	0.2083...	Frank	0.5	male	sophomore	History	0.2500...	0.33333333333333...	0.2285714285714...	0.243902439...	0.2222222222...	low	0.0
7	0.25	Grace	0.0	female	freshman	English	0.7500...	0.8	0.8285714285714...	0.878048780...	0.8888888888...	high	0.575
8	0.2916...	Hank	0.25	male	freshman	IT	0.4375...	0.53333333333333...	0.5714285714285...	0.560975609...	0.5833333333...	medium	0.425000000000...
9	0.3333...	Ivy	0.5	female	sophomore	Psychol...	0.5625...	0.6	0.6285714285714...	0.731707317...	0.6944444444...	high	0.75
10	0.375	Jake	0.75	male	junior	Business	0.1250...	0.06666666666666...	0.0	0.0	0.0	low	0.125
11	0.4166...	Karen	1.0	female	senior	Chemist...	0.8750...	0.93333333333333...	0.9428571428571...	0.853658536...	0.9166666666...	high	0.875
12	0.4583...	Leo	0.5	male	sophomore	Math	0.3750...	0.3	0.2857142857142...	0.317073170...	0.3055555555...	medium	0.25
13	0.5	Mia	0.0	female	freshman	CS	0.6875	0.73333333333333...	0.8	0.731707317...	0.8055555555...	high	0.55
14	0.5416...	Ned	0.25	male	freshman	Electro...	0.1875...	0.16666666666666...	0.1428571428571...	0.170731707...	0.1388888888...	medium	0.125
15	0.5833...	Olive	0.75	female	junior	IT	0.625	0.66666666666666...	0.7142857142857...	0.658536585...	0.6944444444...	high	0.625
16	0.625	Paul	1.0	male	senior	Finance	0.2500...	0.26666666666666...	0.2285714285714...	0.268292682...	0.25	low	0.175000000000...
17	0.6666...	Quinn	0.5	female	sophomore	Biology	0.8125...	0.86666666666666...	0.8571428571428...	0.829268292...	0.8888888888...	high	0.825
18	0.7083...	Ray	0.75	male	junior	History	0.3125	0.36666666666666...	0.2857142857142...	0.390243902...	0.3333333333...	medium	0.375
19	0.75	Sophia	1.0	female	senior	Psychol...	0.9375	0.96666666666666...	0.9714285714285...	0.926829268...	1.0	high	0.925
20	0.7916...	Tom	0.5	male	sophomore	Physics	0.3750...	0.43333333333333...	0.4	0.414634146...	0.4166666666...	medium	0.4
21	0.8333...	Uma	0.0	female	freshman	English	0.7500...	0.76666666666666...	0.8	0.756097560...	0.8055555555...	high	0.625
22	0.875	Victor	0.25	male	freshman	CS	0.0	0.0	0.0	0.097560975...	0.0555555555...	low	0.075000000000...
23	0.9166...	Wendy	0.5	female	sophomore	Math	0.5000...	0.56666666666666...	0.6	0.609756097...	0.6388888888...	medium	0.525

This screenshot shows the table format of dataset after applying Normalize filter. where all numeric attributes have been scaled to the [0, 1] range.