

## Practical : 1

**Aim:** Discretization on diabetes file

**Output:**

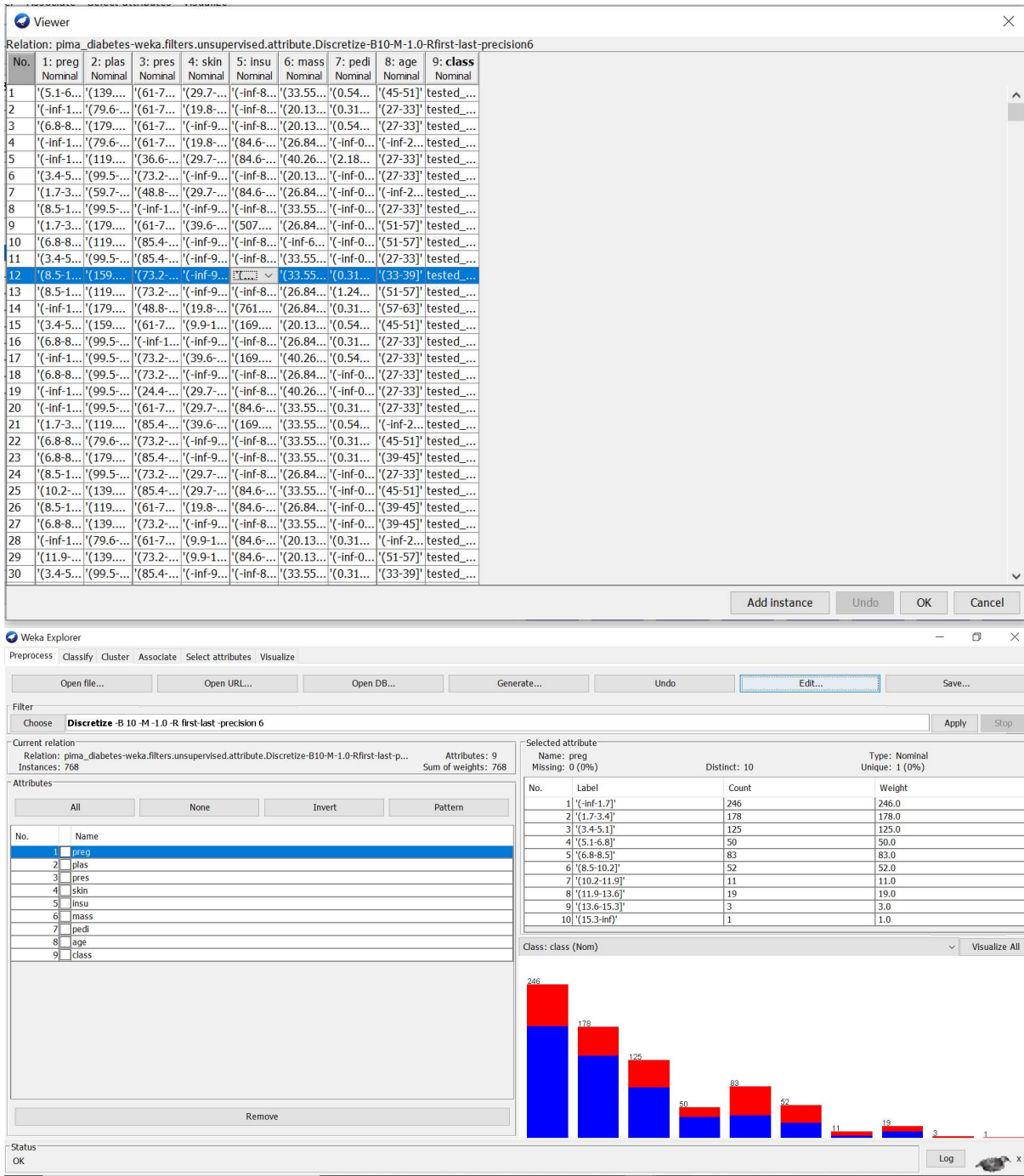
The screenshot shows the Weka Explorer interface with the 'NominalToBinary' filter applied to the 'duration' attribute. The 'duration' attribute is selected, and its statistics are displayed: Minimum: 1, Maximum: 3, Mean: 2.161, StdDev: 0.708. The bar chart shows the distribution of the attribute values across the class, with a red bar for '1' and a blue bar for '0'.

Below the main interface, a table of data is shown, representing the output of the filter. The table has 10 columns: No., 1: duration, 2: wage-increase-first-year, 3: wage-increase-second-year, 4: wage-increase-third-year, 5: cost-of-living-adjustment=none, 6: cost-of-living-adjustment=tc, 7: cost-of-living-adjustment=tc, 8: working-hours, 9: pension=none, and 10: pension=ret\_allw. The data is as follows:

No.	1: duration	2: wage-increase-first-year	3: wage-increase-second-year	4: wage-increase-third-year	5: cost-of-living-adjustment=none	6: cost-of-living-adjustment=tc	7: cost-of-living-adjustment=tc	8: working-hours	9: pension=none	10: pension=ret_allw
1	1.0	5.0						40.0	0	
2	2.0	4.5						35.0	0	
3			5.8					38.0	0	
4	3.0	3.7	4.0	5.0	0.0	0.0	1.0	40.0	0	
5	3.0	4.5	4.5	5.0				35.0	0	
6	2.0	2.0	2.5					40.0	0	
7	3.0	4.0	5.0	5.0	0.0	0.0	1.0	38.0	0	
8	3.0	6.9	4.8	2.3				40.0	0	
9	2.0	3.0	7.0					38.0	0	
10	1.0	5.7			1.0	0.0	0.0	40.0	0	
11	3.0	3.5	4.0	4.6	1.0	0.0	0.0	36.0	0	
12	2.0	6.4	6.4					38.0	0	
13	2.0	3.5	4.0		1.0	0.0	0.0	40.0	0	
14	3.0	3.5	4.0	5.1	0.0	1.0	0.0	37.0	0	
15	1.0	3.0			1.0	0.0	0.0	36.0	0	
16	2.0	4.5	4.0		1.0	0.0	0.0	37.0	0	
17	1.0	2.8						35.0	0	
18	1.0	2.1			0.0	0.0	1.0	40.0	0	
19	1.0	2.0			1.0	0.0	0.0	38.0	1	
20	2.0	4.0	5.0		0.0	1.0	0.0	35.0	0	
21	2.0	4.3	4.4					38.0	0	
22	2.0	2.5	3.0					40.0	1	
23	3.0	3.5	4.0	4.6	0.0	1.0	0.0	27.0	1	
24	2.0	4.5	4.0					40.0	0	
25	1.0	6.0						38.0	0	
26	3.0	2.0	2.0	2.0	1.0	0.0	0.0	40.0	1	
27	2.0	4.5	4.5		0.0	1.0	0.0	33.0	0	
28	2.0	3.0			1.0	0.0	0.0	37.0	0	
29	2.0	5.0	4.0		1.0	0.0	0.0	37.0	0	

2. Dummy variables on Labor file

output



3. Find out weka database for discrreatization  
output:-

Viewer

Relation: cpu-weka.filters.unsupervised.attribute.NominalToBinary-Rfirst-last-weka.filters.unsupervised.attribute.NominalToBinary-Rfirst-last-weka.filters.unsupervised.attribute.Discretize-B10-M-1.0...

No.	1: MYCT Nominal	2: MMIN Nominal	3: MMAX Nominal	4: CACH Nominal	5: CHMIN Nominal	6: CHMAX Nominal	7: class Numeric
1	'(-inf-1...	'(-inf-3...	'(-inf-64...	'(230.4-...	'(15.6-2...	'(123.2-1...	198.0
2	'(-inf-1...	'(6451...	'(25638...	'(25.6-5...	'(5.2-10...	'(17.6-35...	269.0
3	'(-inf-1...	'(6451...	'(25638...	'(25.6-5...	'(5.2-10...	'(17.6-35...	220.0
4	'(-inf-1...	'(6451...	'(25638...	'(25.6-5...	'(5.2-10...	'(17.6-35...	172.0
5	'(-inf-1...	'(6451...	'(12851...	'(25.6-5...	'(5.2-10...	'(-inf-17.6)	132.0
6	'(-inf-1...	'(6451...	'(25638...	'(51.2-7...	'(5.2-10...	'(17.6-35...	318.0
7	'(-inf-1...	'(12838...	'(25638...	'(51.2-7...	'(15.6-2...	'(17.6-35...	367.0
8	'(-inf-1...	'(12838...	'(25638...	'(51.2-7...	'(15.6-2...	'(17.6-35...	489.0
9	'(-inf-1...	'(12838...	'(57606...	'(51.2-7...	'(15.6-2...	'(17.6-35...	636.0
10	'(-inf-1...	'(28806...	'(57606...	'(102.4-...	'(31.2-3...	'(52.8-70...	1144.0
11	'(313.6...	'(-inf-3...	'(-inf-64...	'(-inf-2...	'(-inf-5.2)	'(-inf-17.6)	38.0
12	'(313.6...	'(-inf-3...	'(-inf-64...	'(-inf-2...	'(-inf-5.2)	'(-inf-17.6)	40.0
13	'(-inf-1...	'(-inf-3...	'(6457...	'(51.2-7...	'(-inf-5.2)	'(-inf-17.6)	92.0
14	'(-inf-1...	'(3257...	'(12851...	'(51.2-7...	'(-inf-5.2)	'(-inf-17.6)	138.0
15	'(313.6...	'(-inf-3...	'(-inf-64...	'(-inf-2...	'(-inf-5.2)	'(-inf-17.6)	10.0
16	'(165.3...	'(-inf-3...	'(12851...	'(-inf-2...	'(-inf-5.2)	'(17.6-35...	35.0
17	'(165.3...	'(-inf-3...	'(-inf-64...	'(-inf-2...	'(-inf-5.2)	'(-inf-17.6)	19.0
18	'(-inf-1...	'(-inf-3...	'(-inf-64...	'(-inf-2...	'(5.2-10...	'(17.6-35...	28.0
19	'(-inf-1...	'(-inf-3...	'(-inf-64...	'(-inf-2...	'(-inf-5.2)	'(-inf-17.6)	31.0
20	'(-inf-1...	'(3257...	'(-inf-64...	'(128-1...	'(5.2-10...	'(52.8-70...	120.0
21	'(-inf-1...	'(-inf-3...	'(-inf-64...	'(-inf-2...	'(-inf-5.2)	'(17.6-35...	30.0
22	'(-inf-1...	'(-inf-3...	'(-inf-64...	'(-inf-2...	'(-inf-5.2)	'(17.6-35...	33.0
23	'(-inf-1...	'(-inf-3...	'(-inf-64...	'(-inf-2...	'(5.2-10...	'(52.8-70...	61.0
24	'(-inf-1...	'(-inf-3...	'(-inf-64...	'(-inf-2...	'(5.2-10...	'(52.8-70...	76.0
25	'(313.6...	'(-inf-3...	'(-inf-64...	'(-inf-2...	'(-inf-5.2)	'(-inf-17.6)	23.0
26	'(313.6...	'(-inf-3...	'(-inf-64...	'(-inf-2...	'(-inf-5.2)	'(-inf-17.6)	69.0
27	'(313.6...	'(-inf-3...	'(-inf-64...	'(-inf-2...	'(-inf-5.2)	'(-inf-17.6)	33.0
28	'(313.6...	'(-inf-3...	'(-inf-64...	'(-inf-2...	'(-inf-5.2)	'(-inf-17.6)	27.0
29	'(313.6...	'(-inf-3...	'(-inf-64...	'(-inf-2...	'(-inf-5.2)	'(-inf-17.6)	77.0
30	'(313.6...	'(-inf-3...	'(-inf-64...	'(-inf-2...	'(-inf-5.2)	'(-inf-17.6)	27.0

Add instance Undo OK Cancel

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter  
Choose Discretize-B10-M-1.0-R first-last-precision 6 Apply Stop

Current relation  
Relation: cpu-weka.filters.unsupervised.attribute.NominalToBinary-Rfirst-last-weka.filters.uns...  
Instances: 209  
Attributes: 7  
Sum of weights: 209

Attributes  
All None Invert Pattern

No.	Name
1	MYCT
2	MMIN
3	MMAX
4	CACH
5	CHMIN
6	CHMAX
7	class

Remove

Selected attribute  
Name: MYCT  
Missing: 0 (0%)  
Distinct: 8  
Type: Nominal  
Unique: 0 (0%)

No.	Label	Count	Weight
1	'(-inf-165.3)	137	137.0
2	'(165.3-313.6)	34	34.0
3	'(313.6-461.9)	15	15.0
4	'(461.9-610.2)	4	4.0
5	'(610.2-758.5)	2	2.0
6	'(758.5-906.8)	13	13.0
7	'(906.8-1055.1)	0	0.0
8	'(1055.1-1203.4)	2	2.0
9	'(1203.4-1351.7)	0	0.0
10	'(1351.7-inf)	2	2.0

Class: class (Num) Visualize All

Status OK Log x0

4. Find out weka dataset for nominal to binary output :-

**Viewer**

Relation: cpu-weka.filters.unsupervised.attribute.NominalToBinary-Rfirst-last

No.	1: vendor=adviser Numeric	2: vendor=amdahl Numeric	3: vendor=apollo Numeric	4: vendor=basf Numeric	5: vendor=btl Numeric	6: vendor=burroughs Numeric	7: vendor=c.r.d Numeric	8: vendor=cdc Numeric	9: vendor=cambex Numeric	10: vendor=dec Numeric
1	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
12	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
13	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
14	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
15	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
16	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
17	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
18	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
19	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
20	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
21	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
22	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
23	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
24	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
25	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
26	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
27	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
28	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
29	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0

**Weka Explorer**

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: Choose **NominalToBinary-R first-last** Apply Stop

Current relation: Relation: cpu-weka.filters.unsupervised.attribute.NominalToBinary-Rfirst-last  
Instances: 209

Attributes: 37 Sum of weights: 209

Attributes: All None Invert Pattern

No.	Name
1	vendor=adviser
2	vendor=amdahl
3	vendor=apollo
4	vendor=basf
5	vendor=btl
6	vendor=burroughs
7	vendor=c.r.d
8	vendor=cdc
9	vendor=cambex
10	vendor=dec
11	vendor=dg
12	vendor=formation
13	vendor=four-phase
14	vendor=gould
15	vendor=hp
16	vendor=harris
17	vendor=honeywell
18	vendor=ibm
19	vendor=ipl
20	vendor=maauson

Remove

Selected attribute: Name: vendor=adviser  
Missing: 0 (0%) Distinct: 2 Type: Numeric  
Unique: 1 (0%)

Statistic	Value
Minimum	0
Maximum	1
Mean	0.005
StdDev	0.069

Class: class (Num) Visualize All

209

Log