

Tp Preprocessing

jeux de données	Que représentent les données	Quels sont les attributs	Quel est leur type	Y a t'il un attribut classe	Y a t'il des données manquantes (%)
iris	Ce sont des données sur différentes variété de la plante Iris	sepalwidth petalwidth	Numeric Numeric Numeric	Oui	0
vote	Elles représentent le résultat de votes politiques.	handicapped-infants water-project-cost-sharing adoption-of-the-budget-resolution physician-fee-freeze el-salvador-aid religious-groups-in-schools anti-satellite-test-ban aid-to-nicaraguan-contras mx-missile immigration synfuels-corporation-cutback education-spending superfund-right-to-sue crime-duty-free-exports export-administration act-south-africa	Nominal Nominal Nominal Nominal Nominal Nominal Nominal Nominal Nominal Nominal Nominal	Oui	3%
Breast Cancer	Elles représentent les données de personnes ayant contracté du cancer.	age menopause tumor-size inv-nodes node-caps deg-malg breast-breast-quadrant irradiated	Nominal Nominal Nominal Nominal Nominal Nominal Nominal	Oui	3%
Labor	Données sur	duration wage-hour	Numeric	Oui	2%

différent	increase-first-	Numeric
règlements de	year wage-	Numeric
l'industrie	increase-	Numeric
	second-year	Nominal
	wage-increase-	Numeric
	third-year cost-	Nominal
	of-living-	Numeric
	adjustment	Numeric
	working-hours	Nominal
	pension	Numeric
	standby-pay	Nominal
	shift-differential	Nominal
	education-	Nominal
	allowance	Nominal
	statutory-	Nominal
	holidays	
	vacation	
	longterm-	
	disability-	
	assistance	
	contribution-to-	
	dental-plan	
	bereavement-	
	assistance	
	contribution-to-	
	health-plan	

2) Remplacement des valeur manquantes :

Filter

Choose **ReplaceMissingValues** Apply Stop

Current relation
 Relation: labor-neg-data-weka.filters.unsupervis...
 Instances: 57
 Attributes: 17
 Sum of weights: 57

Selected attribute
 Name: contribution-to-health-plan
 Missing: 0 (0%)
 Distinct: 3
 Type: Nominal
 Unique: 0 (0%)

Attributes

All None Invert Pattern

No.	Name
1	<input type="checkbox"/> duration
2	<input type="checkbox"/> wage-increase-first-year
3	<input type="checkbox"/> wage-increase-second-year
4	<input type="checkbox"/> wage-increase-third-year
5	<input type="checkbox"/> cost-of-living-adjustment
6	<input type="checkbox"/> working-hours
7	<input type="checkbox"/> pension
8	<input type="checkbox"/> standby-pay
9	<input type="checkbox"/> shift-differential
10	<input type="checkbox"/> education-allowance
11	<input type="checkbox"/> statutory-holidays
12	<input type="checkbox"/> vacation
13	<input type="checkbox"/> longterm-disability-assistance
14	<input type="checkbox"/> contribution-to-dental-plan
15	<input type="checkbox"/> bereavement-assistance
16	<input checked="" type="checkbox"/> contribution-to-health-plan
17	<input type="checkbox"/> class

Remove

No.	Label	Count	Weight
1	none	8	8
2	half	9	9
3	full	40	40

Class: class (Nom) Visualize All

Status: OK Log x 0

Filter

Choose **ReplaceMissingValues** Apply Stop

Current relation
 Relation: vote-weka.filters.unsupervised.attribu...
 Instances: 435
 Attributes: 17
 Sum of weights: 435

Selected attribute
 Name: handicapped-infants
 Missing: 0 (0%)
 Distinct: 2
 Type: Nominal
 Unique: 0 (0%)

Attributes

All None Invert Pattern

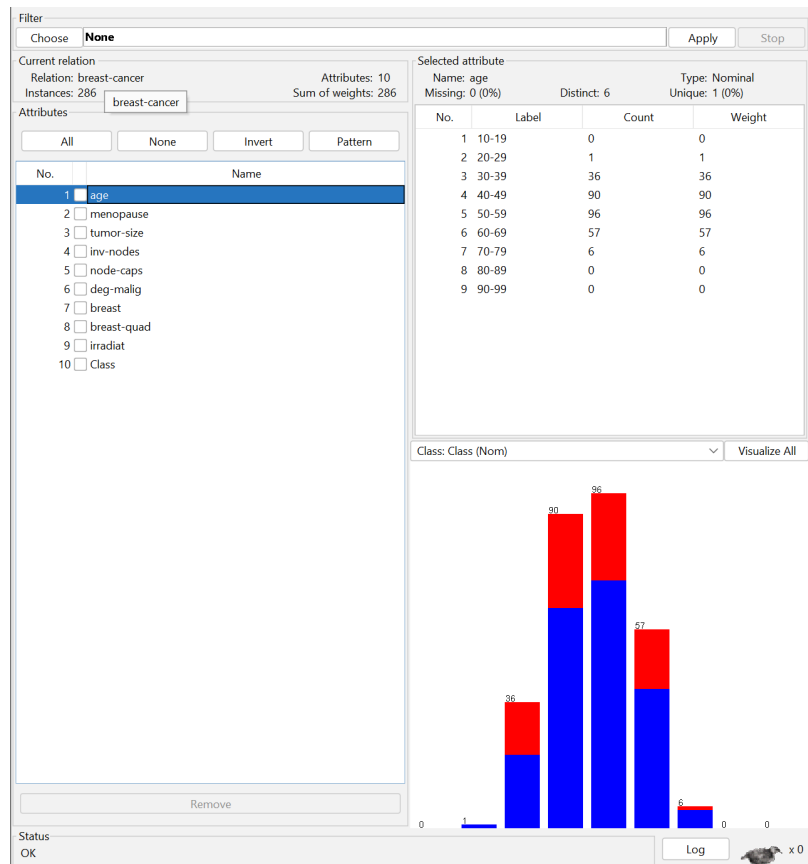
No.	Name
1	<input checked="" type="checkbox"/> handicapped-infants
2	<input type="checkbox"/> water-project-cost-sharing
3	<input type="checkbox"/> adoption-of-the-budget-resolution
4	<input type="checkbox"/> physician-fee-freeze
5	<input type="checkbox"/> el-salvador-aid
6	<input type="checkbox"/> religious-groups-in-schools
7	<input type="checkbox"/> anti-satellite-test-ban
8	<input type="checkbox"/> aid-to-nicaraguan-contras
9	<input type="checkbox"/> mx-missile
10	<input type="checkbox"/> immigration
11	<input type="checkbox"/> synfuels-corporation-cutback
12	<input type="checkbox"/> education-spending
13	<input type="checkbox"/> superfund-right-to-sue
14	<input type="checkbox"/> crime
15	<input type="checkbox"/> duty-free-exports
16	<input type="checkbox"/> export-administration-act-south-africa
17	<input type="checkbox"/> Class

Remove

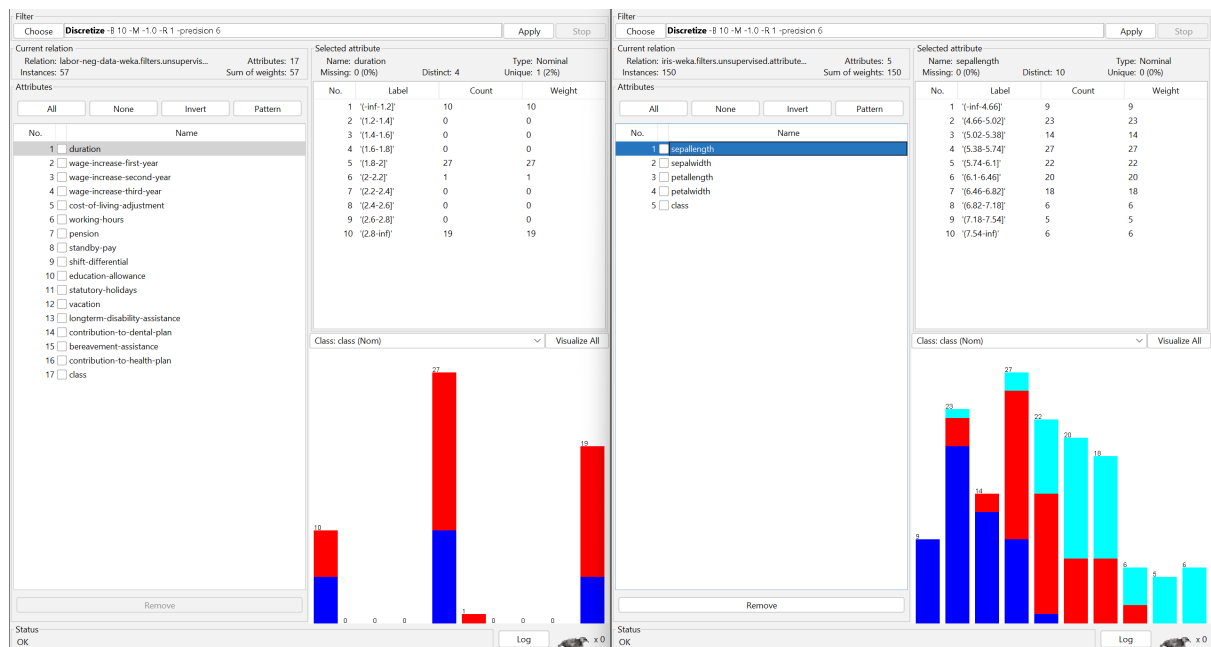
No.	Label	Count	Weight
1	n	248	248
2	y	187	187

Class: Class (Nom) Visualize All

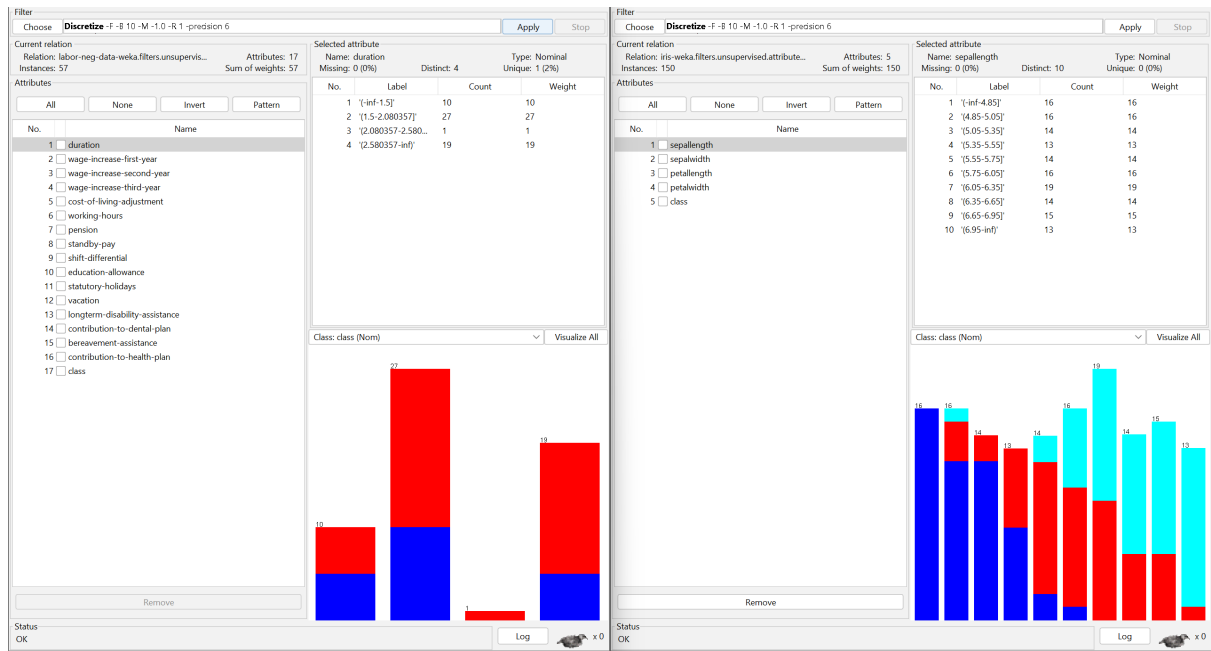
Status: OK Log x 0



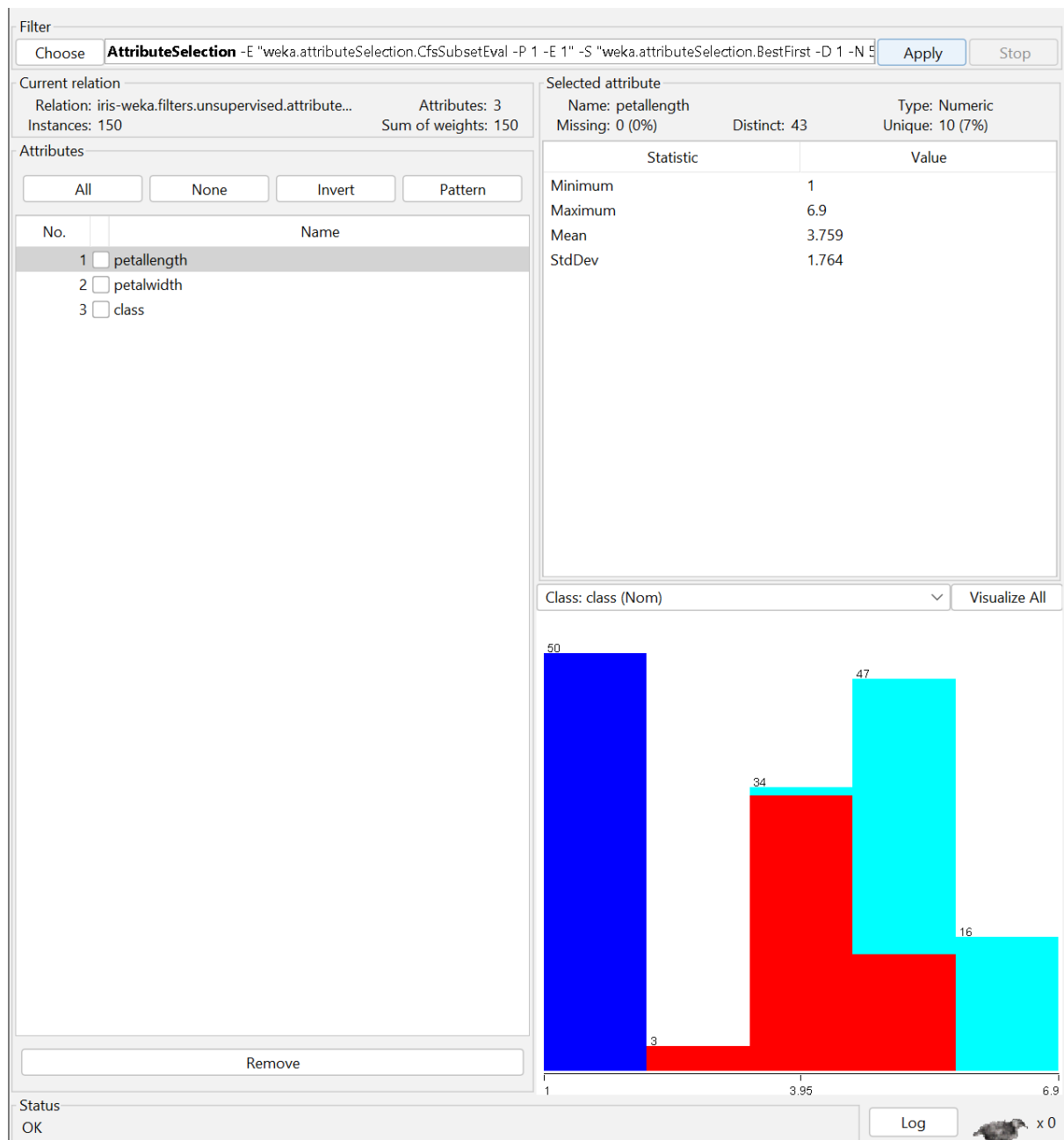
3) Discrétisations du premier attribut numérique des jeux de données en intervalles réguliers



Discrétisations du premier attribut numérique des jeux de données en fréquences



4) Visualisation du nouveau jeu de données après sélection d'attribut iris :



labor :

Filter

Choose

AttributeSelection -E "weka.attributeSelection.CfsSubsetEval -P 1 -E 1" -S "weka.attributeSelection.BestFirst -D 1 -N 5"

Apply

Stop

Current relation

Relation: labor-neg-data-weka.filters.unsupervis...

Attributes: 8

Instances: 57

Sum of weights: 57

Attributes

All

None

Invert

Pattern

No.	Name
1	<input checked="" type="checkbox"/> wage-increase-first-year
2	<input type="checkbox"/> wage-increase-third-year
3	<input type="checkbox"/> pension
4	<input type="checkbox"/> standby-pay
5	<input type="checkbox"/> statutory-holidays
6	<input type="checkbox"/> longterm-disability-assistance
7	<input type="checkbox"/> contribution-to-dental-plan
8	<input type="checkbox"/> class

Remove

Selected attribute

Name: wage-increase-first-year

Type: Numeric

Missing: 0 (0%)

Distinct: 18

Unique: 8 (14%)

Statistic	Value
Minimum	2
Maximum	7
Mean	3.804
StdDev	1.358

Class: class (Nom) Visualize All

Attribute Value	Red Count	Blue Count
2	21	21
4.5	23	8
7	8	5

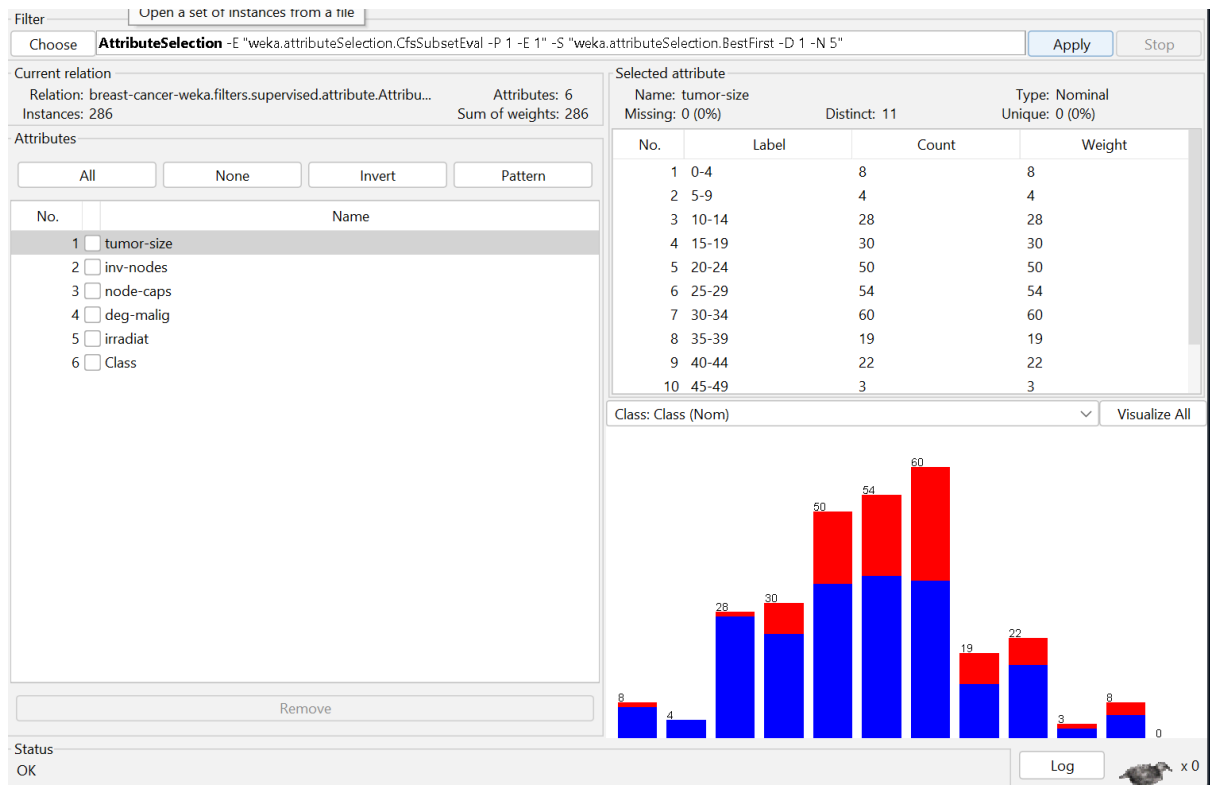
Status

OK

Log

x 0

breast cancer :



vote :

Filter

Choose
AttributeSelection -E "weka.attributeSelection.CfsSubsetEval -P 1 -E 1" -S "weka.attributeSelection.BestFirst -D 1 -N 5
Apply
Stop

Current relation
Relation: vote-weka.filters.unsupervised.attribu...
Instances: 435
Attributes: 6
Sum of weights: 435

Attributes

All
None
Invert
Pattern

No.	Name
1	<input type="checkbox"/> adoption-of-the-budget-resolution
2	<input type="checkbox"/> physician-fee-freeze
3	<input type="checkbox"/> immigration
4	<input type="checkbox"/> synfuels-corporation-cutback
5	<input type="checkbox"/> duty-free-exports
6	<input type="checkbox"/> Class

Remove

Selected attribute
Name: adoption-of-the-budget-resolution
Missing: 0 (0%)
Type: Nominal
Distinct: 2
Unique: 0 (0%)

No.	Label	Count	Weight
1	n	171	171
2	y	264	264

Class: Class (Nom)
Visualize All

Label	Count	Weight
n	171	171
y	264	264

Status

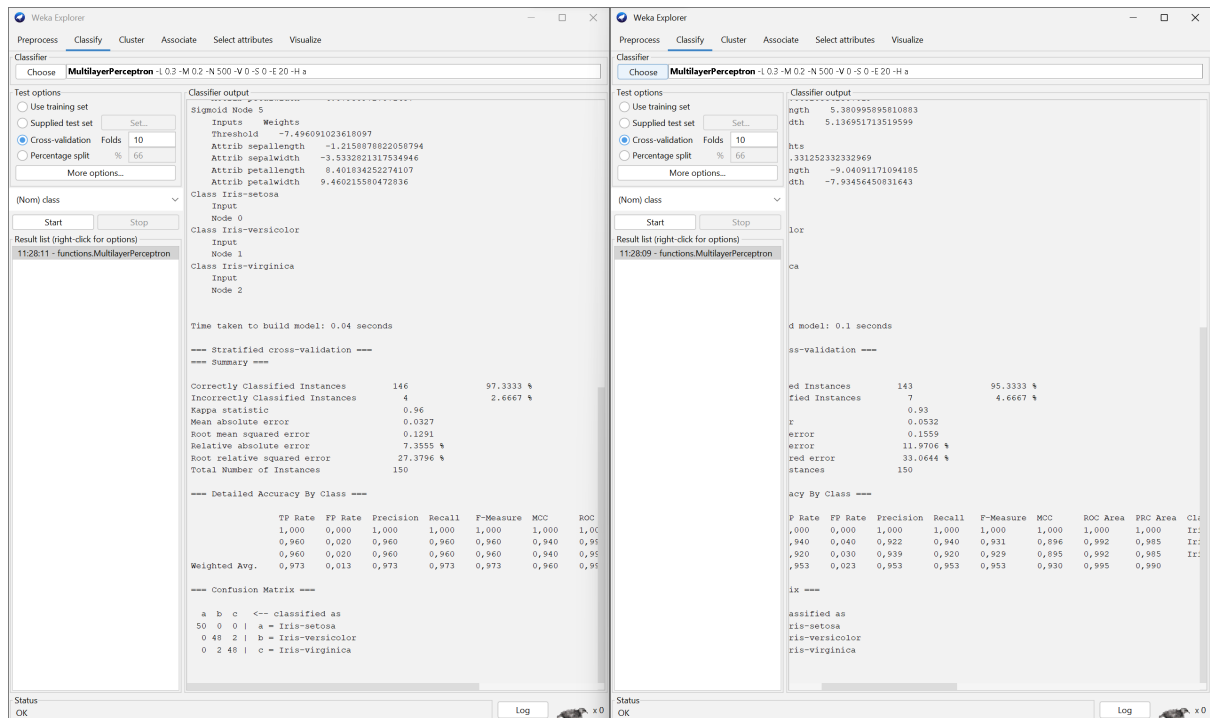
OK

Log

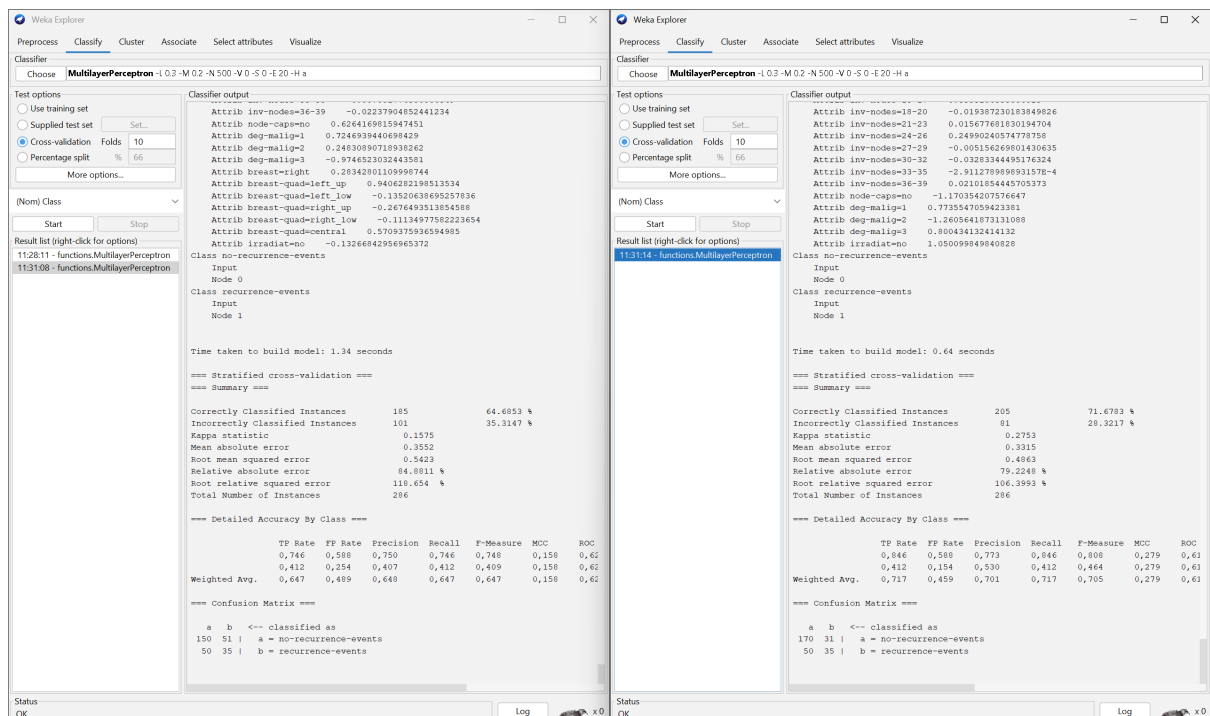
x 0

5)

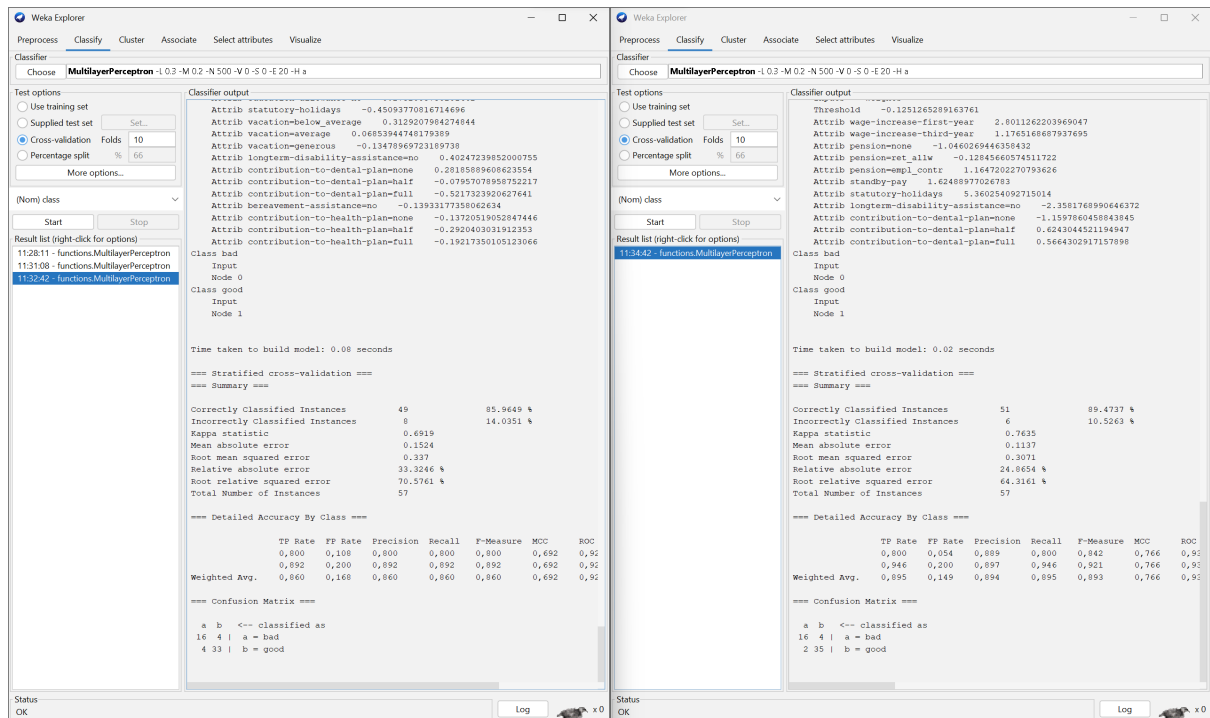
Comparaison des résultat obtenus :



Nous observons une diminution de 2% des instances de classe correct du jeux de données Iris.



Nous observons une augmentation de 7% des instances de classe correct du jeux de données Breast-cancer.



Nous observons une augmentation de 4,5% des instances de classe correct du jeu de données Labor.

Nous observons une augmentation de 0,2% des instances de classe correct du jeu de données Vote.