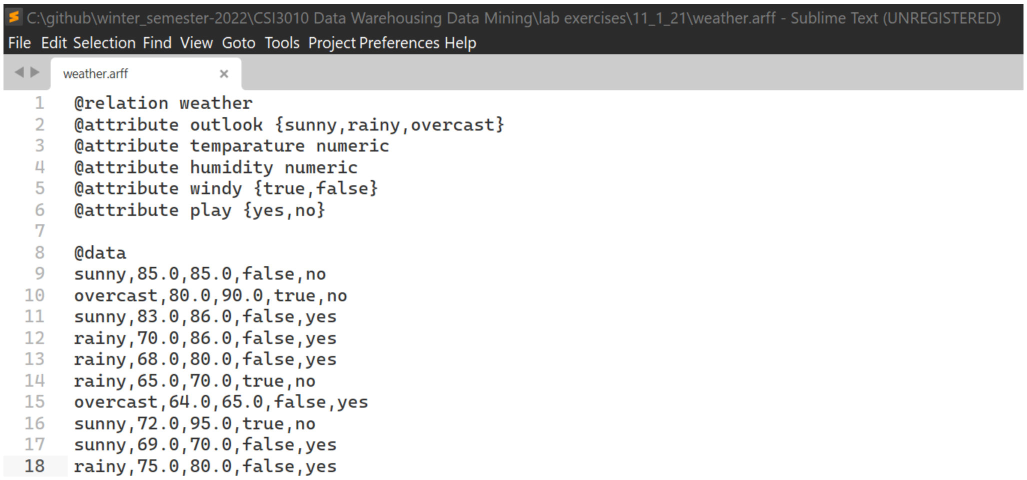
I) Data Preprocessing

19MID0020

Data-Set Creation using Editor

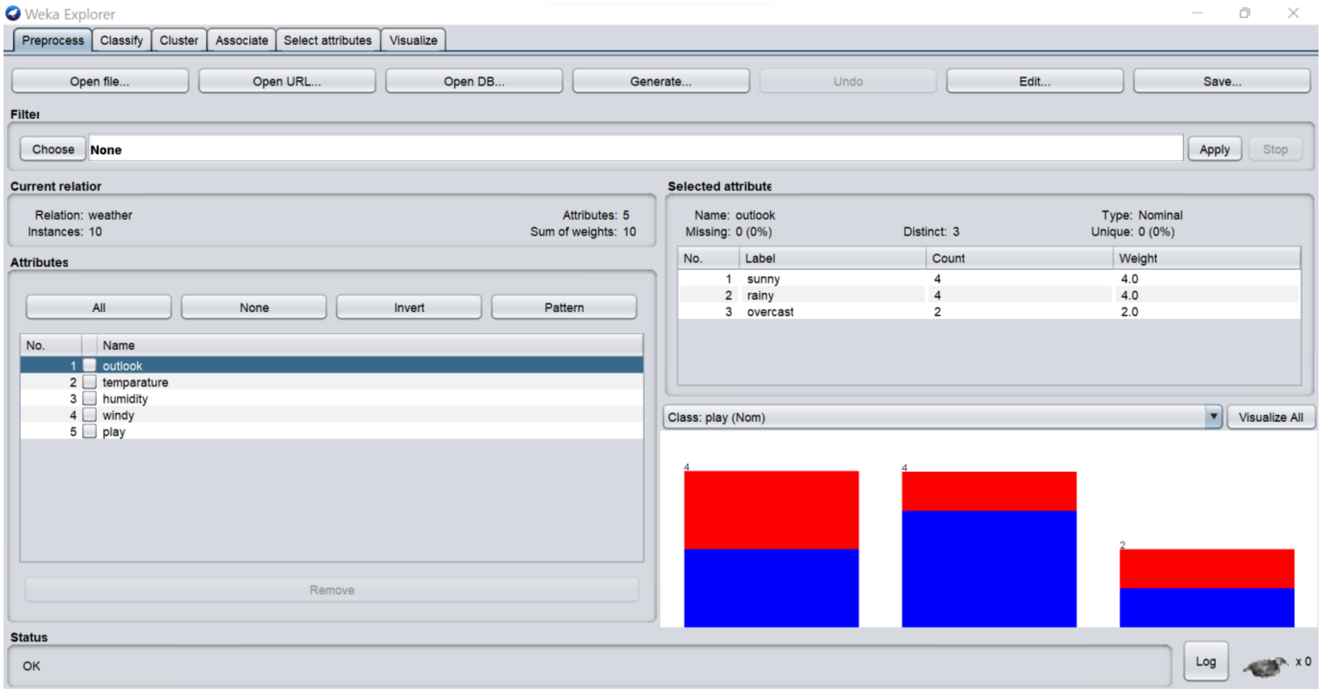


C:\github\winter_semester-2022\CSI3010 Data Warehousing Data Mining\lab exercises\11_1_21\weather.arff - Sublime Text (UNREGISTERED)
File Edit Selection Find View Goto Tools Project Preferences Help
weather.arff

```
@relation weather
@attribute outlook {sunny,rainy,overcast}
@attribute temparature numeric
@attribute humidity numeric
@attribute windy {true,false}
@attribute play {yes,no}

@data
sunny,85.0,85.0,false,no
overcast,80.0,90.0,true,no
sunny,83.0,86.0,false,yes
rainy,70.0,86.0,false,yes
rainy,68.0,80.0,false,yes
rainy,65.0,70.0,true,no
overcast,64.0,65.0,false,yes
sunny,72.0,95.0,true,no
sunny,69.0,70.0,false,yes
rainy,75.0,80.0,false,yes
```

Opening the created Data-Set using Weka



Filter: None

Relation: weather
Instances: 10
Attributes: 5
Sum of weights: 10

No.	Name
1	outlook
2	temparature
3	humidity
4	windy
5	play

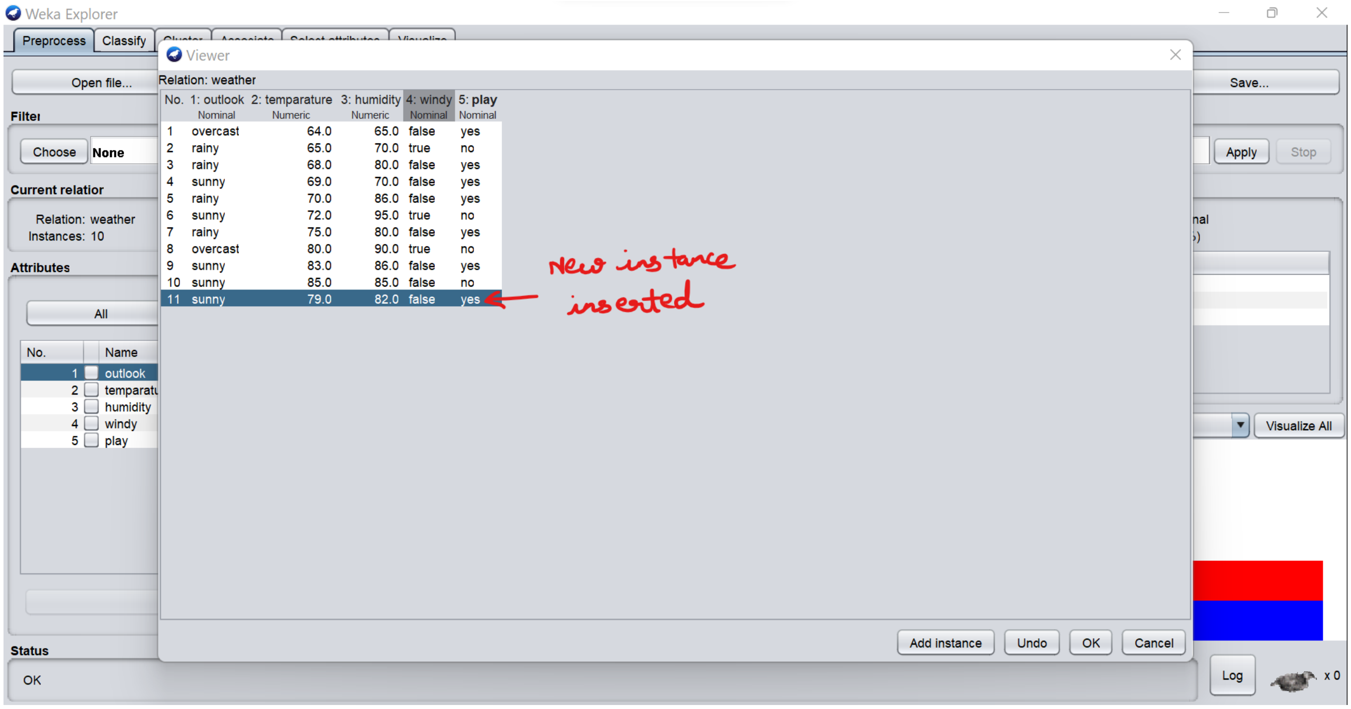
Name: outlook
Missing: 0 (0%)
Distinct: 3
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	sunny	4	4.0
2	rainy	4	4.0
3	overcast	2	2.0

Class: play (Nom)

Status: OK

Adding new Instances

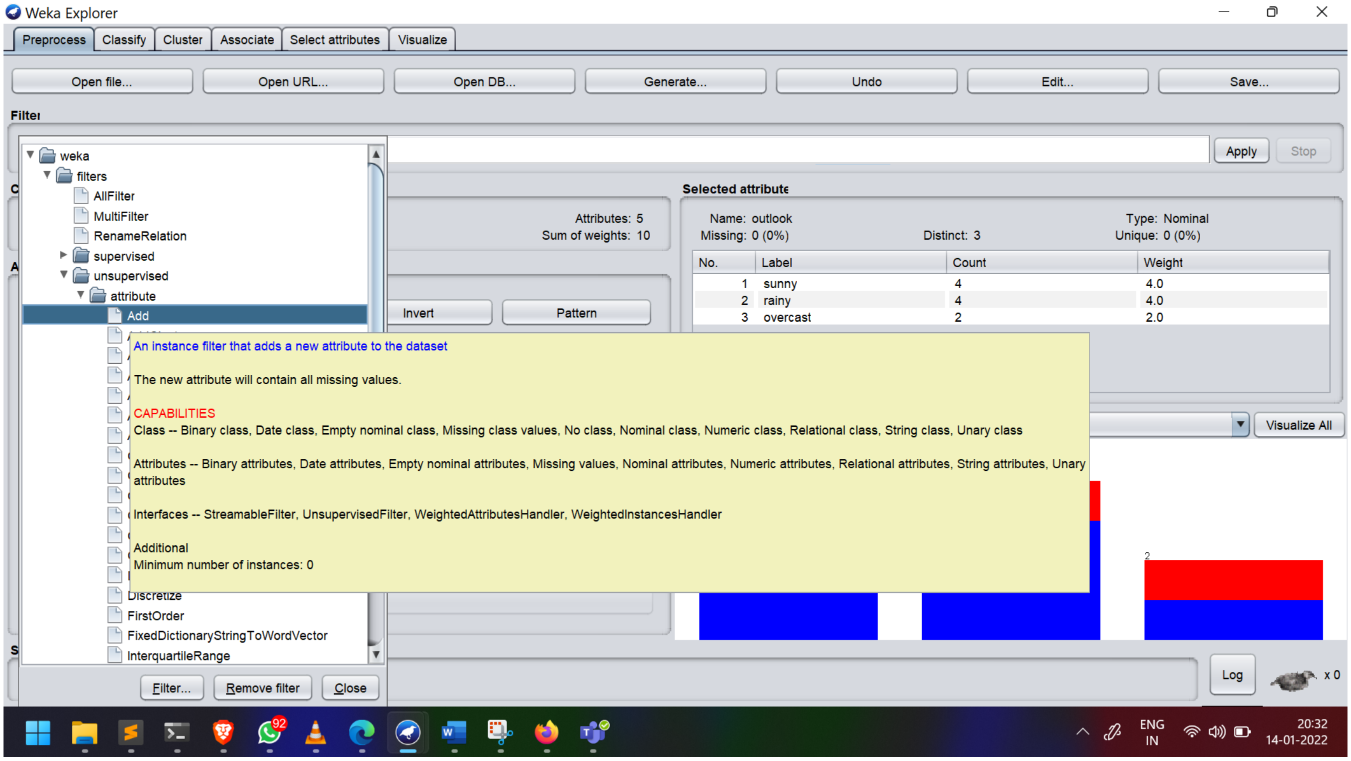


Relation: weather

No.	1: outlook Nominal	2: temparature Numeric	3: humidity Numeric	4: windy Nominal	5: play Nominal
1	overcast	64.0	65.0	false	yes
2	rainy	65.0	70.0	true	no
3	rainy	68.0	80.0	false	yes
4	sunny	69.0	70.0	false	yes
5	rainy	70.0	86.0	false	yes
6	sunny	72.0	95.0	true	no
7	rainy	75.0	80.0	false	yes
8	overcast	80.0	90.0	true	no
9	sunny	83.0	86.0	false	yes
10	sunny	85.0	85.0	false	no
11	sunny	79.0	82.0	false	yes

New instance inserted

Adding Attributes



Attributes: 5
Sum of weights: 10

Selected attribute

Name: outlook
Type: Nominal
Missing: 0 (0%)
Distinct: 3
Unique: 0 (0%)

No.	Label	Count	Weight
1	sunny	4	4.0
2	rainy	4	4.0
3	overcast	2	2.0

An instance filter that adds a new attribute to the dataset

The new attribute will contain all missing values.

CAPABILITIES
Class -- Binary class, Date class, Empty nominal class, Missing class values, No class, Nominal class, Numeric class, Relational class, String class, Unary class

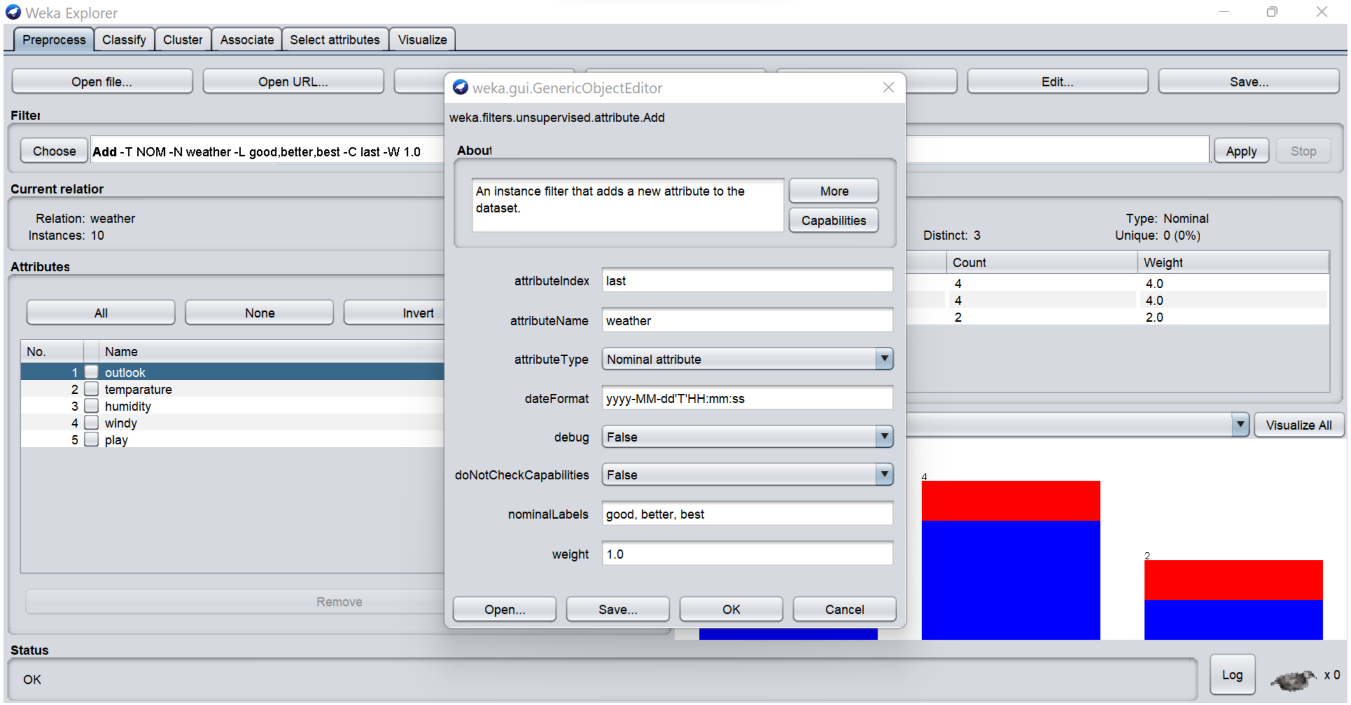
Attributes -- Binary attributes, Date attributes, Empty nominal attributes, Missing values, Nominal attributes, Numeric attributes, Relational attributes, String attributes, Unary attributes

Interfaces -- StreamableFilter, UnsupervisedFilter, WeightedAttributesHandler, WeightedInstancesHandler

Additional
Minimum number of instances: 0

DATA PREPROCESSING

19MID0020



Filter

Add -T NOM -N weather -L good,better,best -C last -W 1.0

Current relation

Relation: weather
Instances: 10

Distinct: 3
Type: Nominal
Unique: 0 (0%)

Count	Weight
4	4.0
4	4.0
2	2.0

Attributes

No.	Name
1	outlook
2	temparature
3	humidity
4	windy
5	play

weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.Add

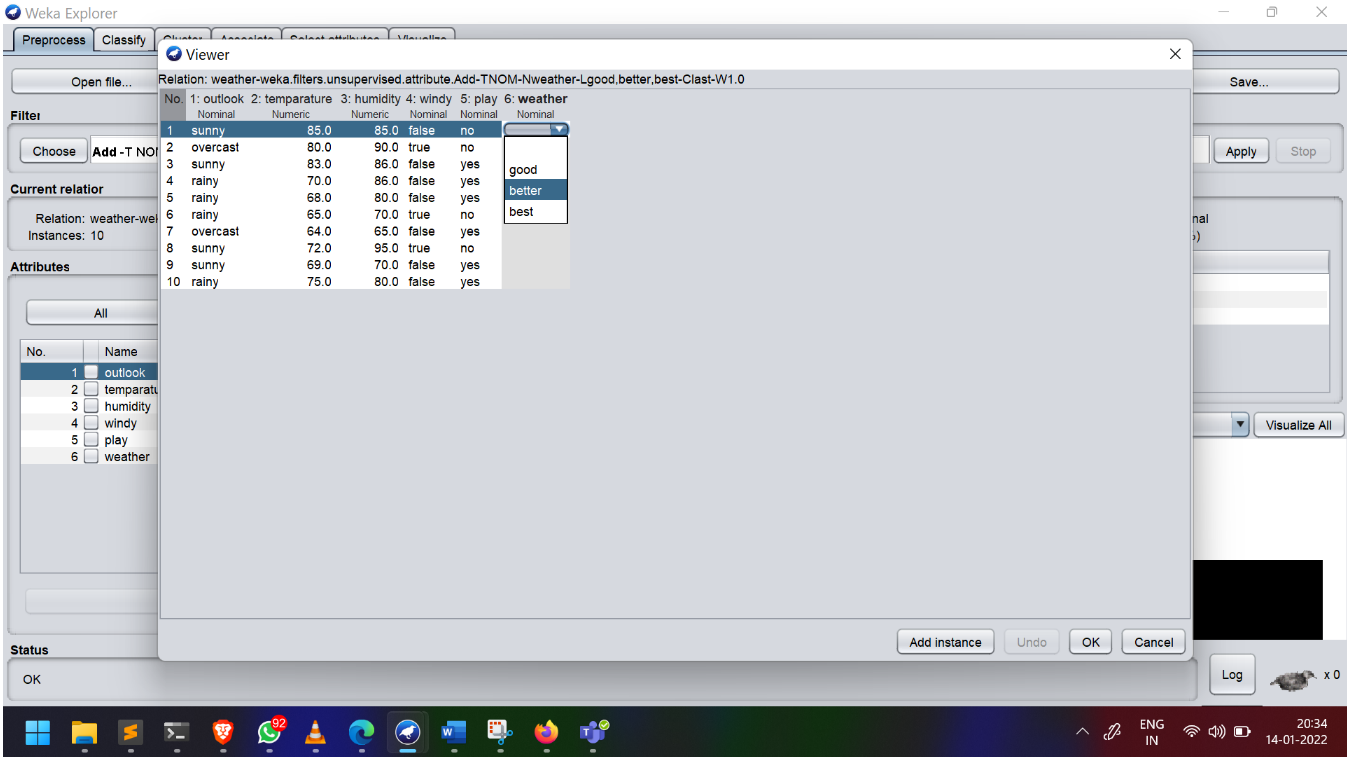
About

An instance filter that adds a new attribute to the dataset.

attributeIndex	last
attributeName	weather
attributeType	Nominal attribute
dateFormat	yyyy-MM-dd'T'HH:mm:ss
debug	False
doNotCheckCapabilities	False
nominalLabels	good, better, best
weight	1.0

Status

OK



Filter

Add -T NOI

Current relation

Relation: weather-wel
Instances: 10

Attributes

No.	Name
1	outlook
2	temparatu
3	humidity
4	windy
5	play
6	weather

Viewer

Relation: weather-weka.filters.unsupervised.attribute.Add-TNOM-Nweather-Lgood,better,best-Clast-W1.0

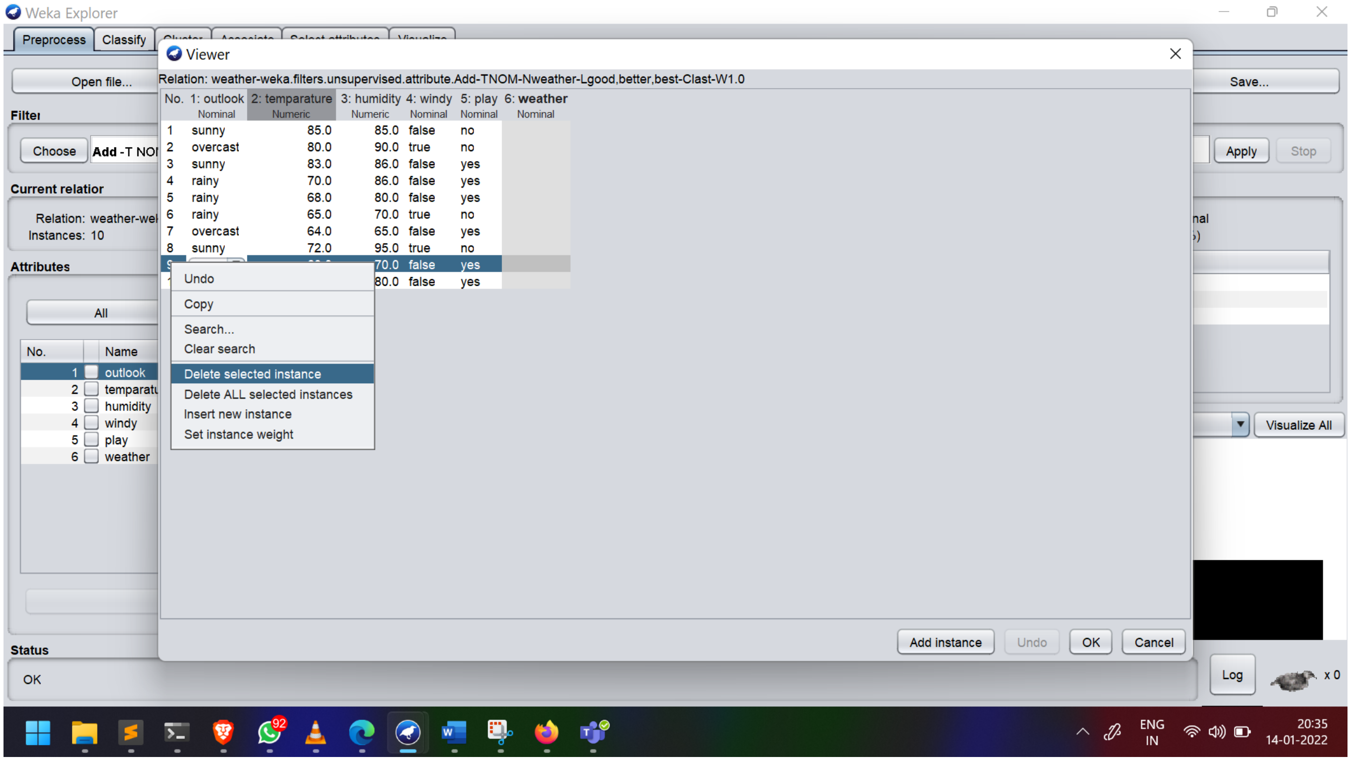
No.	1: outlook Nominal	2: temparature Numeric	3: humidity Numeric	4: windy Nominal	5: play Nominal	6: **weather** Nominal
1	sunny	85.0	85.0	false	no	
2	overcast	80.0	90.0	true	no	
3	sunny	83.0	86.0	false	yes	
4	rainy	70.0	86.0	false	yes	
5	rainy	68.0	80.0	false	yes	
6	rainy	65.0	70.0	true	no	
7	overcast	64.0	65.0	false	yes	
8	sunny	72.0	95.0	true	no	
9	sunny	69.0	70.0	false	yes	
10	rainy	75.0	80.0	false	yes	

good
better
best

Status

OK

Removing Instances



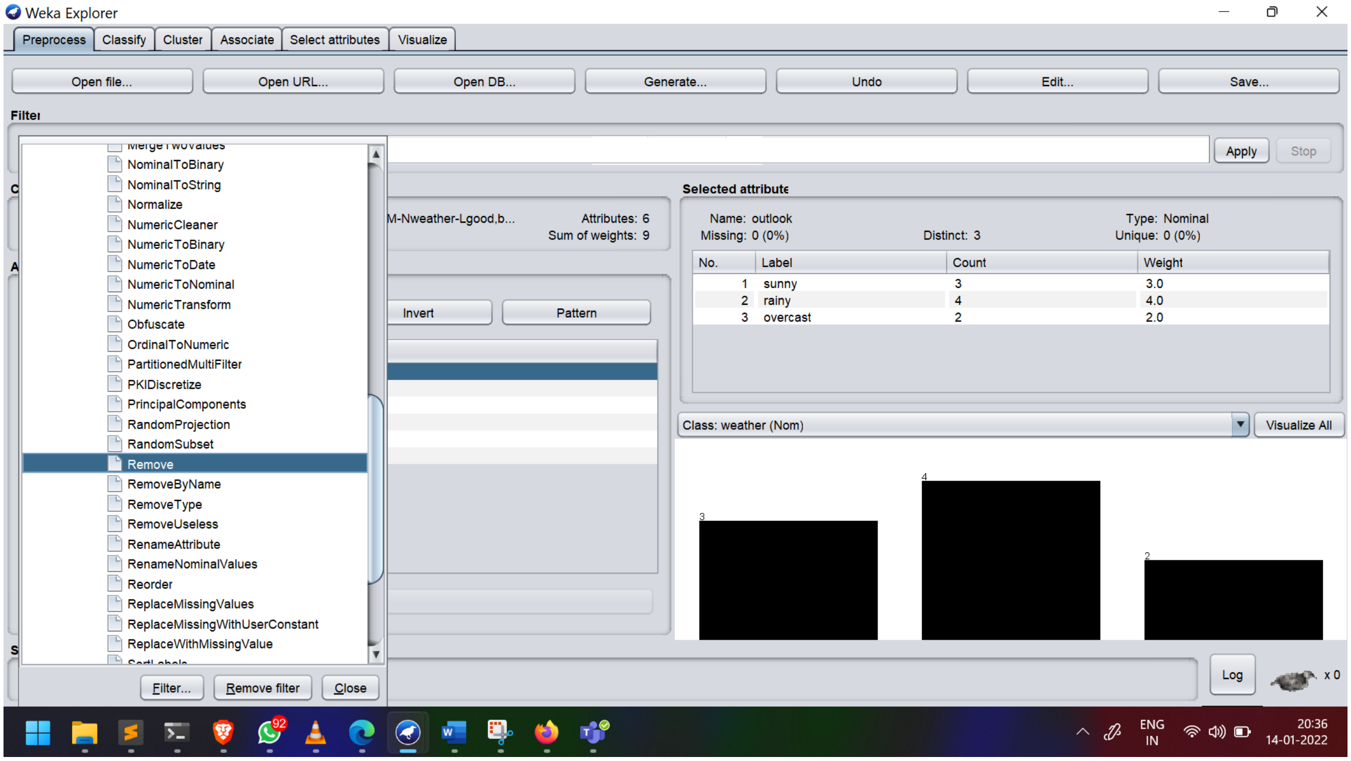
Relation: weather-wel
Instances: 10

No.	Name
1	outlook
2	temparatu
3	humidity
4	windy
5	play
6	weather

Relation: weather-weka.filters.unsupervised.attribute.Add-TNOM-Nweather-Lgood,better,best-Clast-W1.0

No.	1: outlook Nominal	2: temparature Numeric	3: humidity Numeric	4: windy Nominal	5: play Nominal	6: **weather** Nominal
1	sunny	85.0	85.0	false	no	
2	overcast	80.0	90.0	true	no	
3	sunny	83.0	86.0	false	yes	
4	rainy	70.0	86.0	false	yes	
5	rainy	68.0	80.0	false	yes	
6	rainy	65.0	70.0	true	no	
7	overcast	64.0	65.0	false	yes	
8	sunny	72.0	95.0	true	no	
9			70.0	false	yes	
1			80.0	false	yes	

Removing Attributes



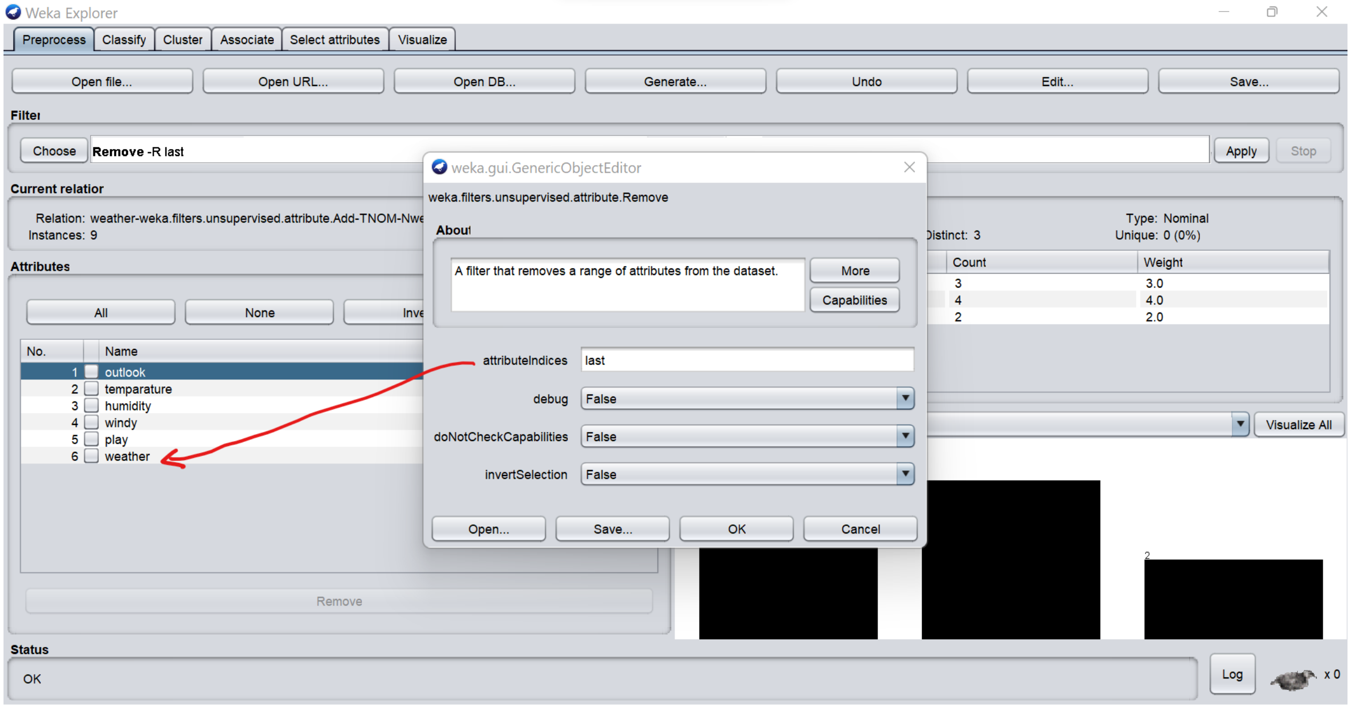
MergeTwoValues
NominalToBinary
NominalToString
Normalize
NumericCleaner
NumericToBinary
NumericToDate
NumericToNominal
NumericTransform
Obfuscate
OrdinalToNumeric
PartitionedMultiFilter
PKIDiscretize
PrincipalComponents
RandomProjection
RandomSubset
Remove
RemoveByName
RemoveType
RemoveUseless
RenameAttribute
RenameNominalValues
Reorder
ReplaceMissingValues
ReplaceMissingWithUserConstant
ReplaceWithMissingValue
SortLabels

M-Nweather-Lgood,b... Attributes: 6
Sum of weights: 9

Selected attribute
Name: outlook Type: Nominal
Missing: 0 (0%) Distinct: 3 Unique: 0 (0%)

No.	Label	Count	Weight
1	sunny	3	3.0
2	rainy	4	4.0
3	overcast	2	2.0

Class: weather (Nom)



Remove -R last

Current relatior
Relation: weather-weka.filters.unsupervised.attribute.Add-TNOM-Nwe
Instances: 9

Attributes

No.	Name
1	outlook
2	temparature
3	humidity
4	windy
5	play
6	weather

weka.gui.GenericObjectEditor
weka.filters.unsupervised.attribute.Remove

About
A filter that removes a range of attributes from the dataset.

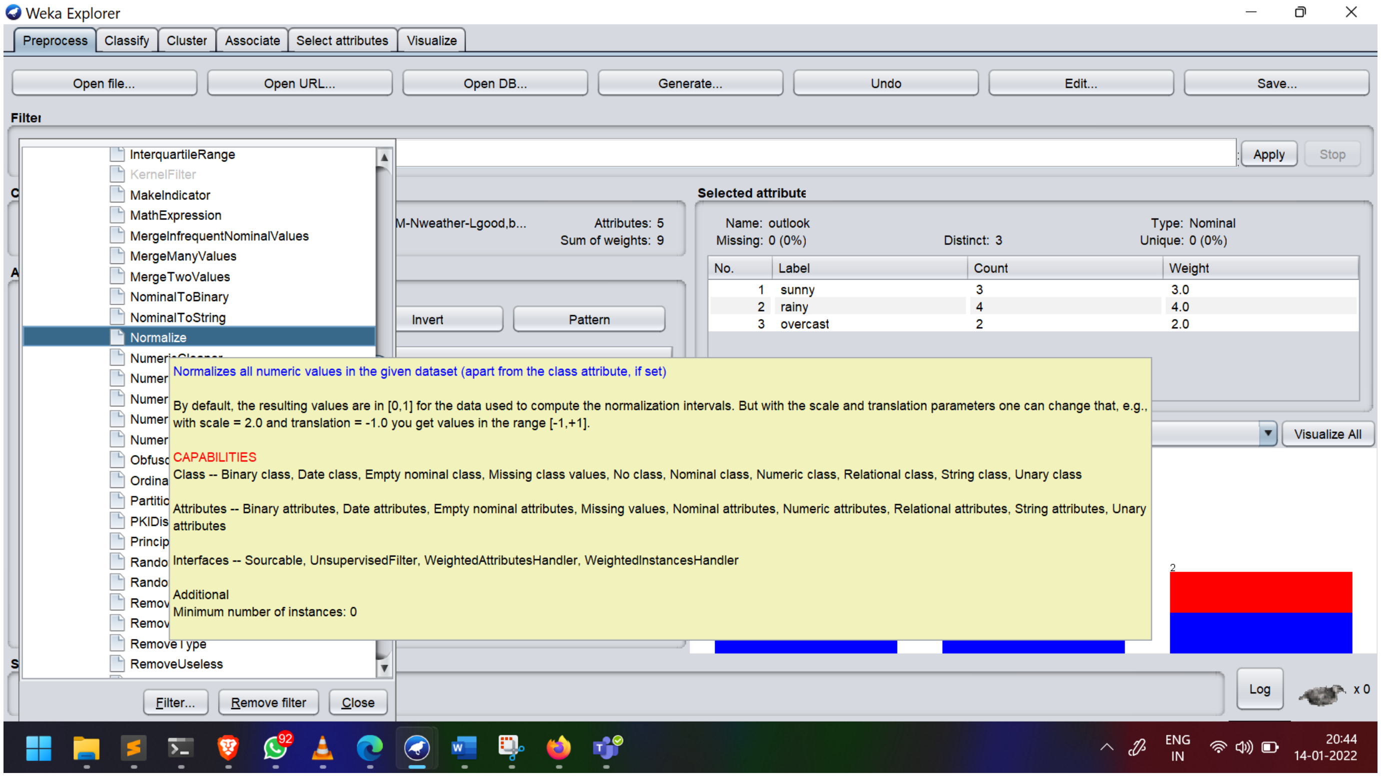
attributeIndices: last
debug: False
doNotCheckCapabilities: False
invertSelection: False

Distinct: 3 Type: Nominal
Unique: 0 (0%)

Count	Weight
3	3.0
4	4.0
2	2.0

Status
OK

Normalize



M-Nweather-Lgood,b... Attributes: 5
Sum of weights: 9

Selected attribute

Name: outlook Type: Nominal
Missing: 0 (0%) Distinct: 3 Unique: 0 (0%)

No.	Label	Count	Weight
1	sunny	3	3.0
2	rainy	4	4.0
3	overcast	2	2.0

Normalizes all numeric values in the given dataset (apart from the class attribute, if set)

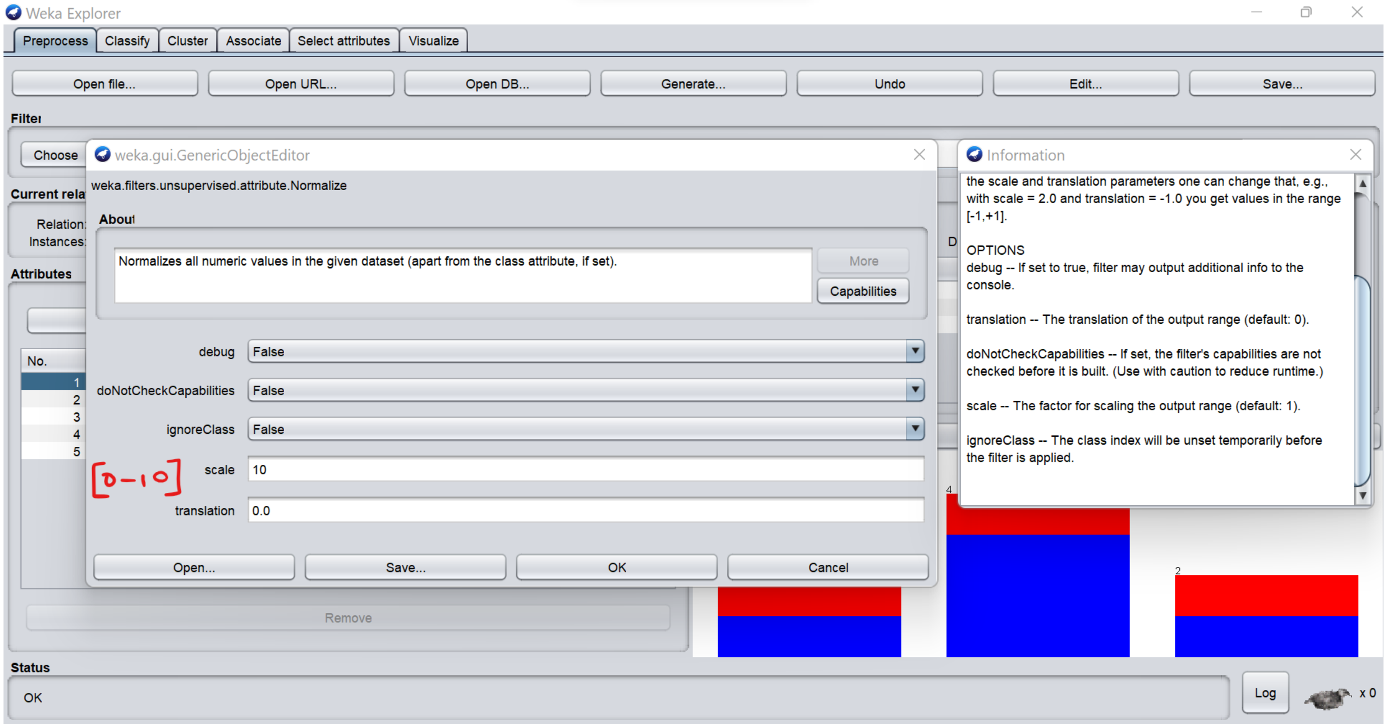
By default, the resulting values are in [0,1] for the data used to compute the normalization intervals. But with the scale and translation parameters one can change that, e.g., with scale = 2.0 and translation = -1.0 you get values in the range [-1,+1].

CAPABILITIES
Class -- Binary class, Date class, Empty nominal class, Missing class values, No class, Nominal class, Numeric class, Relational class, String class, Unary class

Attributes -- Binary attributes, Date attributes, Empty nominal attributes, Missing values, Nominal attributes, Numeric attributes, Relational attributes, String attributes, Unary attributes

Interfaces -- Sourcable, UnsupervisedFilter, WeightedAttributesHandler, WeightedInstancesHandler

Additional
Minimum number of instances: 0



weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.Normalize

About

Normalizes all numeric values in the given dataset (apart from the class attribute, if set).

debug: False
doNotCheckCapabilities: False
ignoreClass: False
[0-10] scale: 10
translation: 0.0

the scale and translation parameters one can change that, e.g., with scale = 2.0 and translation = -1.0 you get values in the range [-1,+1].

OPTIONS
debug -- If set to true, filter may output additional info to the console.

translation -- The translation of the output range (default: 0).

doNotCheckCapabilities -- If set, the filter's capabilities are not checked before it is built. (Use with caution to reduce runtime.)

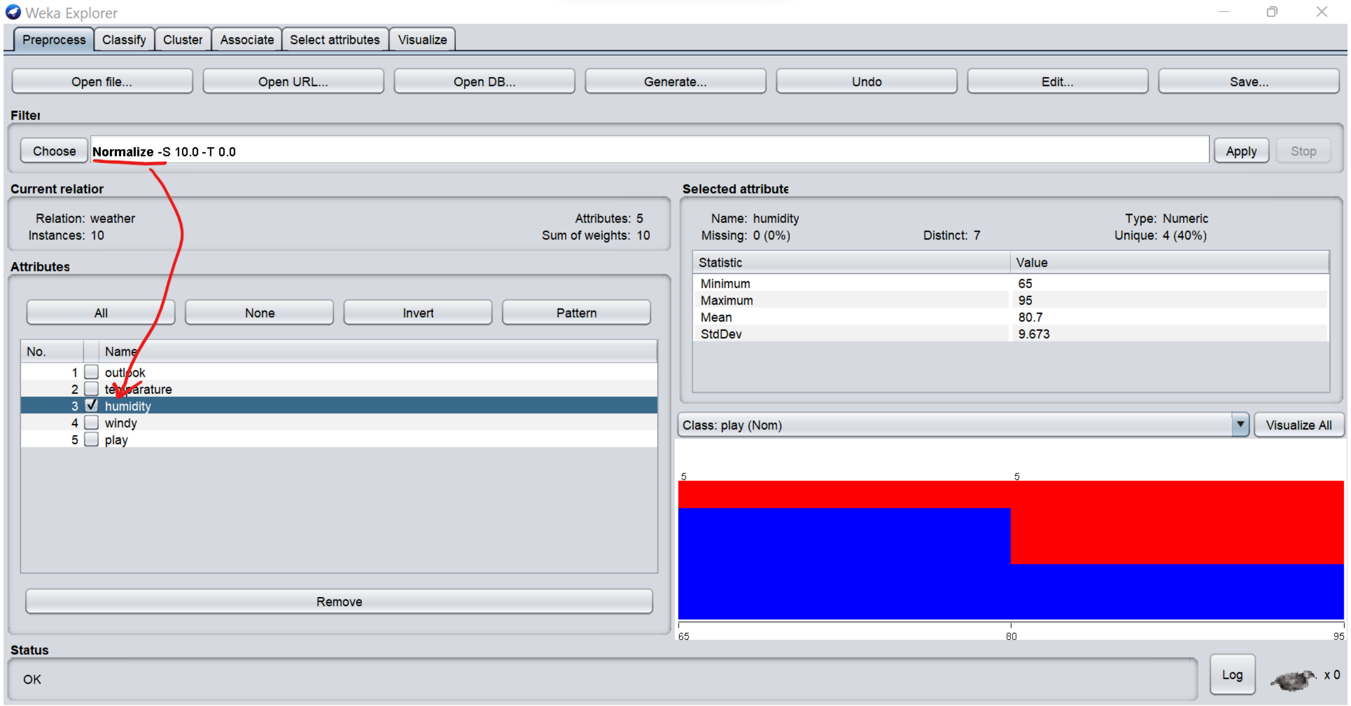
scale -- The factor for scaling the output range (default: 1).

ignoreClass -- The class index will be unset temporarily before the filter is applied.

Status
OK

DATA PREPROCESSING

19MID0020



Normalize -S 10.0 -T 0.0

Current relation

Relation: weather
Instances: 10
Attributes: 5
Sum of weights: 10

Attributes

No.		Name
1	□	outlook
2	□	temparature
3	☑	humidity
4	□	windy
5	□	play

Selected attribute

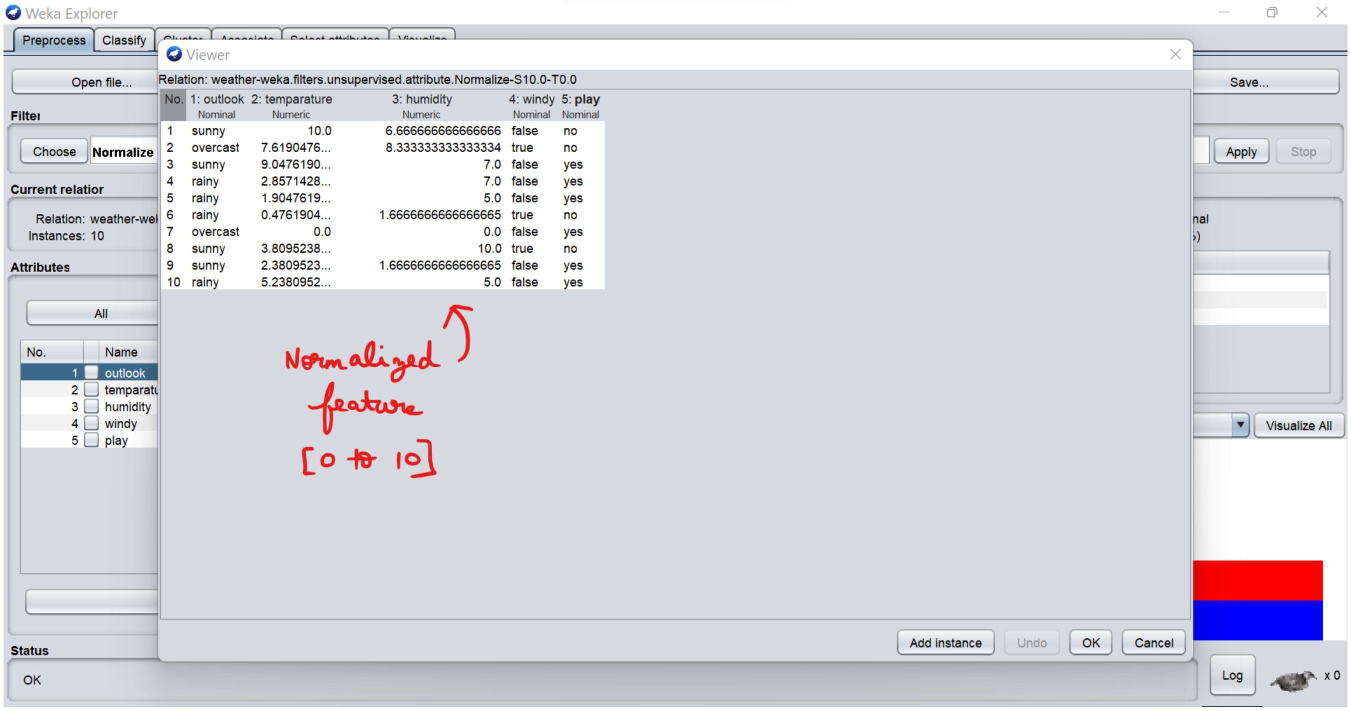
Name: humidity
Missing: 0 (0%)
Distinct: 7
Type: Numeric
Unique: 4 (40%)

Statistic	Value
Minimum	65
Maximum	95
Mean	80.7
StdDev	9.673

Class: play (Nom)

Status

OK



No.	Name
1	outlook
2	temparatu
3	humidity
4	windy
5	play

Relation: weather-weka.filters.unsupervised.attribute.Normalize-S10.0-T0.0

No.	1: outlook Nominal	2: temparature Numeric	3: humidity Numeric	4: windy Nominal	5: **play** Nominal
1	sunny	10.0	6.666666666666666	false	no
2	overcast	7.6190476...	8.333333333333334	true	no
3	sunny	9.0476190...	7.0	false	yes
4	rainy	2.8571428...	7.0	false	yes
5	rainy	1.9047619...	5.0	false	yes
6	rainy	0.4761904...	1.6666666666666665	true	no
7	overcast	0.0	0.0	false	yes
8	sunny	3.8095238...	10.0	true	no
9	sunny	2.3809523...	1.6666666666666665	false	yes
10	rainy	5.2380952...	5.0	false	yes

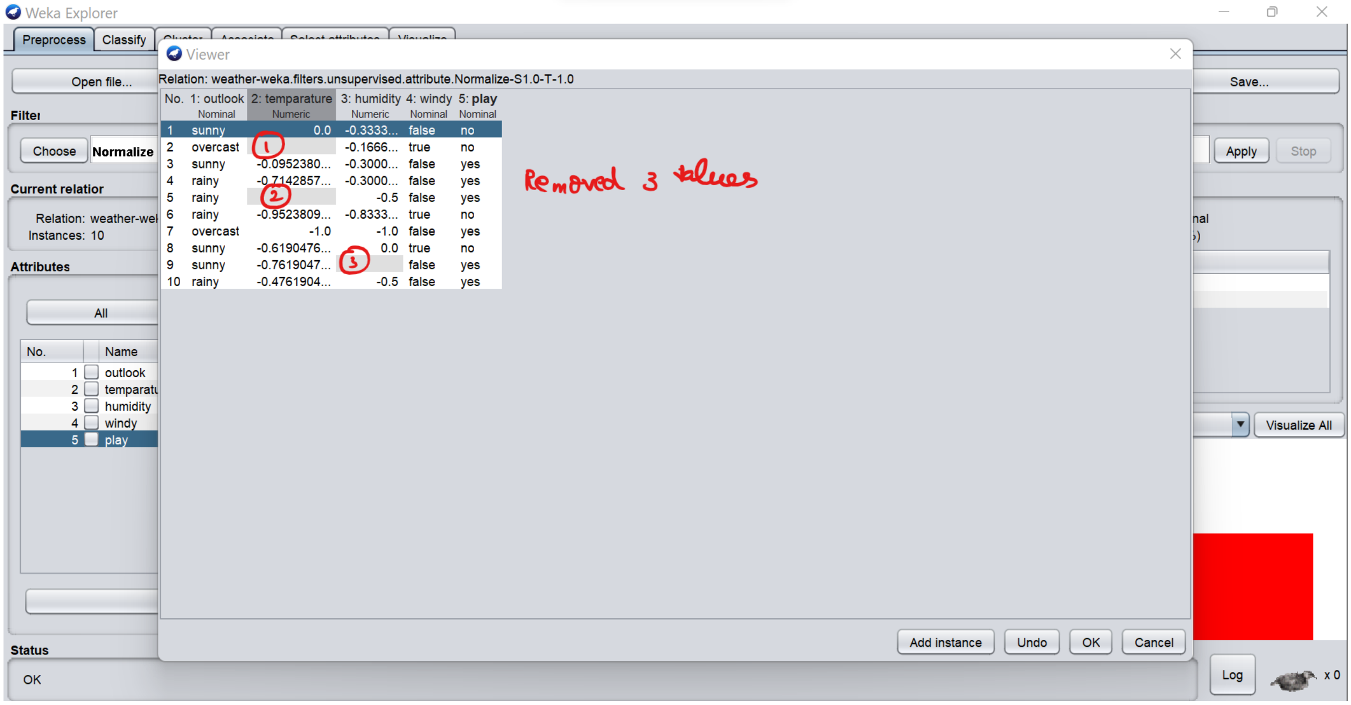
Normalized feature
[0 to 10]

Status

OK

Replacing the Missing Values

Since there are no missing values in the data-set, I am deleting some records intentionally.



Filter: Normalize

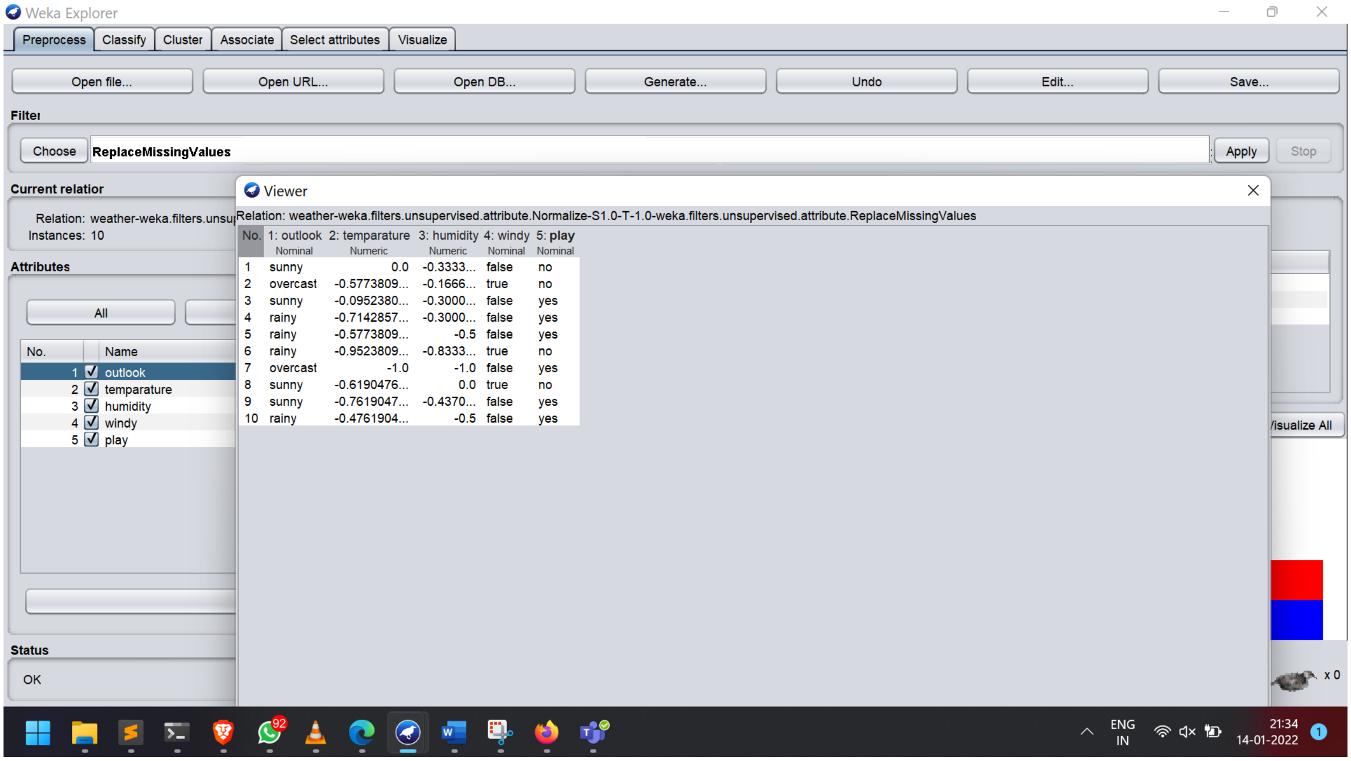
Relation: weather-weka.filters.unsupervised.attribute.Normalize-S1.0-T-1.0

No.	1: outlook Nominal	2: temparature Numeric	3: humidity Numeric	4: windy Nominal	5: **play** Nominal
1	sunny	0.0	-0.3333...	false	no
2	overcast	1	-0.1666...	true	no
3	sunny	-0.0952380...	-0.3000...	false	yes
4	rainy	-0.7142857...	-0.3000...	false	yes
5	rainy	2	-0.5	false	yes
6	rainy	-0.9523809...	-0.8333...	true	no
7	overcast	-1.0	-1.0	false	yes
8	sunny	-0.6190476...	0.0	true	no
9	sunny	-0.7619047...	3	false	yes
10	rainy	-0.4761904...	-0.5	false	yes

Removed 3 values

Relation: weather-wel...
Instances: 10

Status: OK



Filter: ReplaceMissingValues

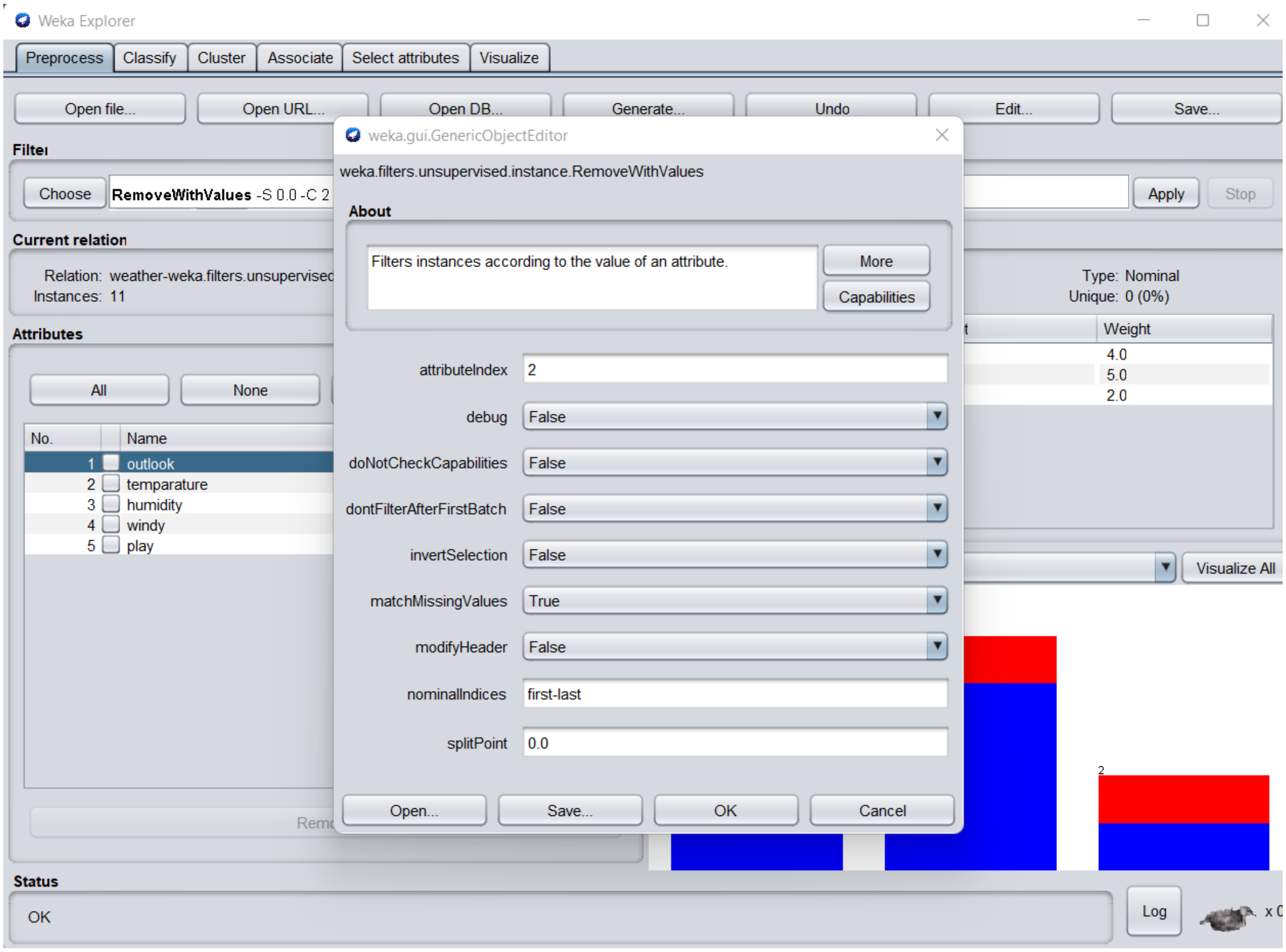
Relation: weather-weka.filters.unsupervised.attribute.Normalize-S1.0-T-1.0-weka.filters.unsupervised.attribute.ReplaceMissingValues

No.	1: outlook Nominal	2: temparature Numeric	3: humidity Numeric	4: windy Nominal	5: **play** Nominal
1	sunny	0.0	-0.3333...	false	no
2	overcast	-0.5773809...	-0.1666...	true	no
3	sunny	-0.0952380...	-0.3000...	false	yes
4	rainy	-0.7142857...	-0.3000...	false	yes
5	rainy	-0.5773809...	-0.5	false	yes
6	rainy	-0.9523809...	-0.8333...	true	no
7	overcast	-1.0	-1.0	false	yes
8	sunny	-0.6190476...	0.0	true	no
9	sunny	-0.7619047...	-0.4370...	false	yes
10	rainy	-0.4761904...	-0.5	false	yes

Relation: weather-weka.filters.unsu...
Instances: 10

Status: OK

Remove the values filter



RemoveWithValues -S 0.0 -C 2

Relation: weather-weka.filters.unsupervised
Instances: 11

Type: Nominal
Unique: 0 (0%)

No.	Name
1	outlook
2	temparature
3	humidity
4	windy
5	play

Weight
4.0
5.0
2.0

weka.gui.GenericObjectEditor

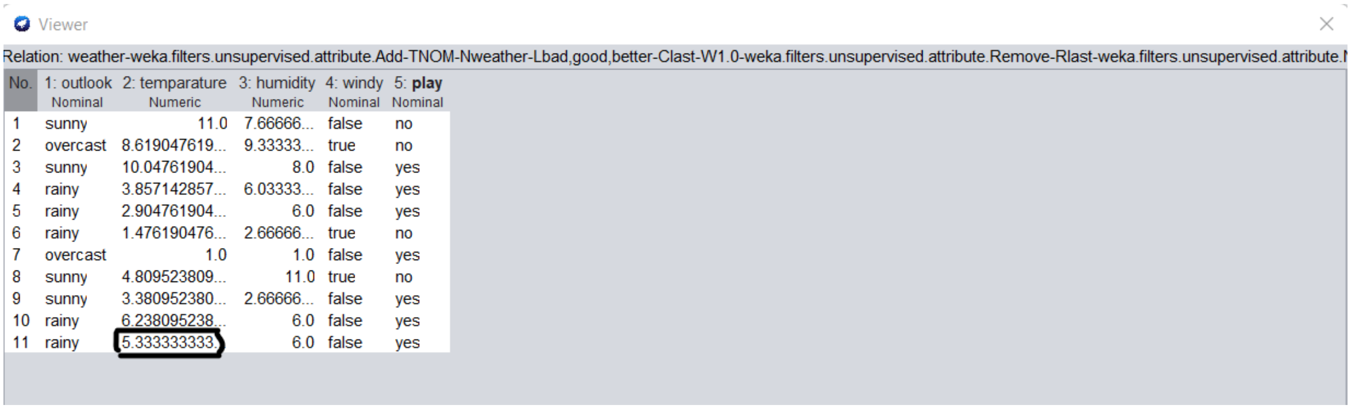
weka.filters.unsupervised.instance.RemoveWithValues

About

Filters instances according to the value of an attribute.

attributeIndex: 2
debug: False
doNotCheckCapabilities: False
dontFilterAfterFirstBatch: False
invertSelection: False
matchMissingValues: True
modifyHeader: False
nominalIndices: first-last
splitPoint: 0.0

OK



Relation: weather-weka.filters.unsupervised.attribute.Add-TNOM-Nweather-Lbad,good,better-Clast-W1.0-weka.filters.unsupervised.attribute.Remove-Rlast-weka.filters.unsupervised.attribute.

No.	1: outlook Nominal	2: temparature Numeric	3: humidity Numeric	4: windy Nominal	5: **play** Nominal
1	sunny	11.0	7.66666...	false	no
2	overcast	8.619047619...	9.33333...	true	no
3	sunny	10.04761904...	8.0	false	yes
4	rainy	3.857142857...	6.03333...	false	yes
5	rainy	2.904761904...	6.0	false	yes
6	rainy	1.476190476...	2.66666...	true	no
7	overcast	1.0	1.0	false	yes
8	sunny	4.809523809...	11.0	true	no
9	sunny	3.380952380...	2.66666...	false	yes
10	rainy	6.238095238...	6.0	false	yes
11	rainy	5.333333333	6.0	false	yes