PRACTICAL 2

Roll No.: K041	Name: Anish Sudhan Nair
Class: B.Tech Cybersecurity	Batch: K2/A2
Date of Practical: 15/01/2022	Date of Submission: 22/01/2022
Grade:	

Aim: To study various filters and knowledge flow feature in WEKA.

Prerequisite:

- KDD process and the importance of preprocessing in Data mining.

Outcome: After successful completion of this experiment, students will be able to

- Understand the usage of some basic and advanced filters for creating samples, randomizing, normalization etc. in preprocessing tab of WEKA Tool.
- Use the knowledge flow feature in WEKA tool

Theory:

Weka include many filters that can be used before invoking a classifier to clean up the dataset or alter it in some way. Filters help with data preparation. Like removing attributes sometimes leads to better classification and also simpler decision trees. These are accessible from the Explorer, and also from the Knowledge Flow and Experimenter interfaces. Few of the unsupervised filters are described below.

Add -An instance filter that adds a new attribute to the dataset.
AddValues - Adds the labels from the given list to an attribute if they are missing.
ReplaceMissingWithUserConstant - Replaces all missing values for nominal, string, numeric and date attributes in the dataset with user-supplied constant values.
StringToNominal - Converts a range of string attributes (unspecified number of values) to nominal (set number of values).
Normalize - Normalizes all numeric values in the given dataset (apart from the class attribute, if set).
RandomSubset - Chooses a random subset of attributes, either an absolute number or a percentage.
Randomize - This filter randomly shuffles the order of instances passed through it.

Knowledge Flow Interface:

The KnowledgeFlow provides an alternative to the Explorer as a graphical front end to WEKA's core algorithms. It presents a data-flow inspired interface to WEKA. The user can select WEKA components from a palette, place them on a layout canvas and connect them together in order to form a knowledge flow for processing and analyzing data. At present, all of WEKA's classifiers, filters, clusterers, associators, loaders and savers are available in the KnowledgeFlow along with some extra tools.

The KnowledgeFlow offers the following features:

- Intuitive data flow style layout
- Process data in batches or incrementally
- Process multiple batches or streams in parallel (each separate flow executes in its own thread)
- Process multiple streams sequentially via a user-specified order of execution
- Chain filters together
- View models produced by classifiers for each fold in a cross validation
- Visualize performance of incremental classifiers during processing (scrolling plots of classification accuracy, RMS error, predictions etc.)

(TO BE COMPLETED BY STUDENTS)

Roll No.: K041	Name: Anish Sudhan Nair
Class: B.Tech Cybersecurity	Batch: K2/A2
Date of Practical: 15/01/2022	Date of Submission: 22/01/2022
Grade:	

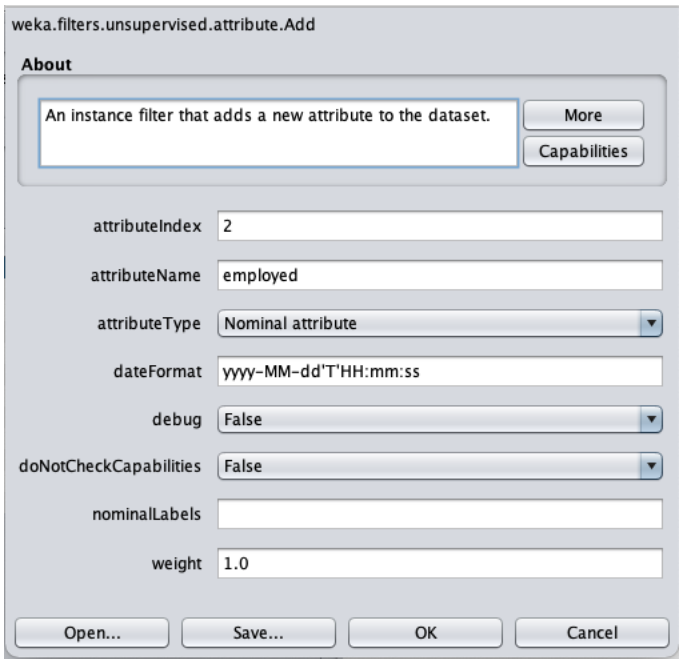
1. On the bank data set loaded earlier, apply the below mentioned filters.

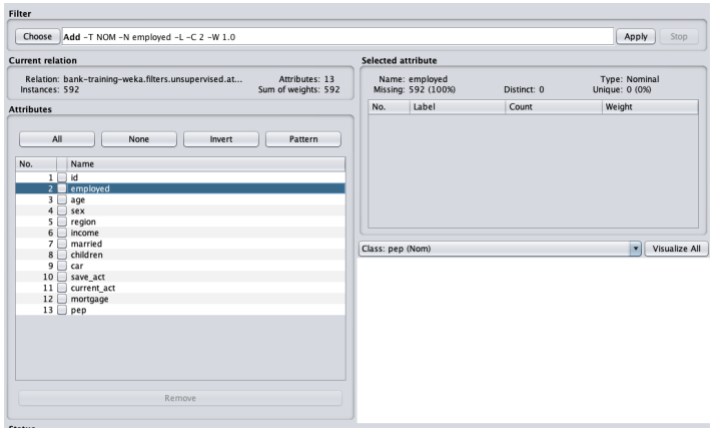
    - Use the filters *add attribute, add values* and insert a new column in the bank data set along with values.
    - Create a new data set with some missing values. Apply the *replace missing with user constant* filter.
    - Set the data type of the attribute ID to string in the data set and use the appropriate filter to convert it into a nominal value.
    - Apply the *normalization* filter to the attributes income and children and observe the results.
    - Create a training and a testing data set having a random subset of sample data using the appropriate filters.

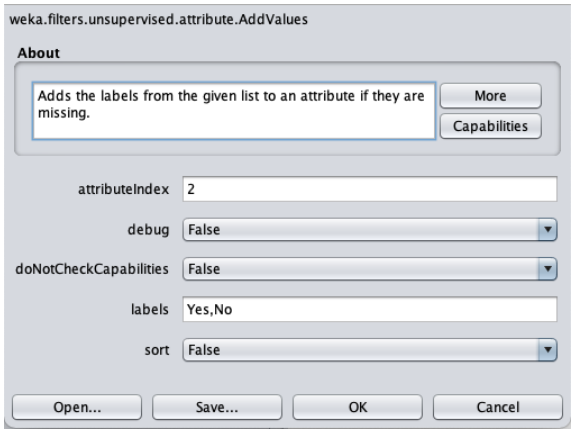
2. For the bank data set, create a knowledge flow to demonstrate the normalization.

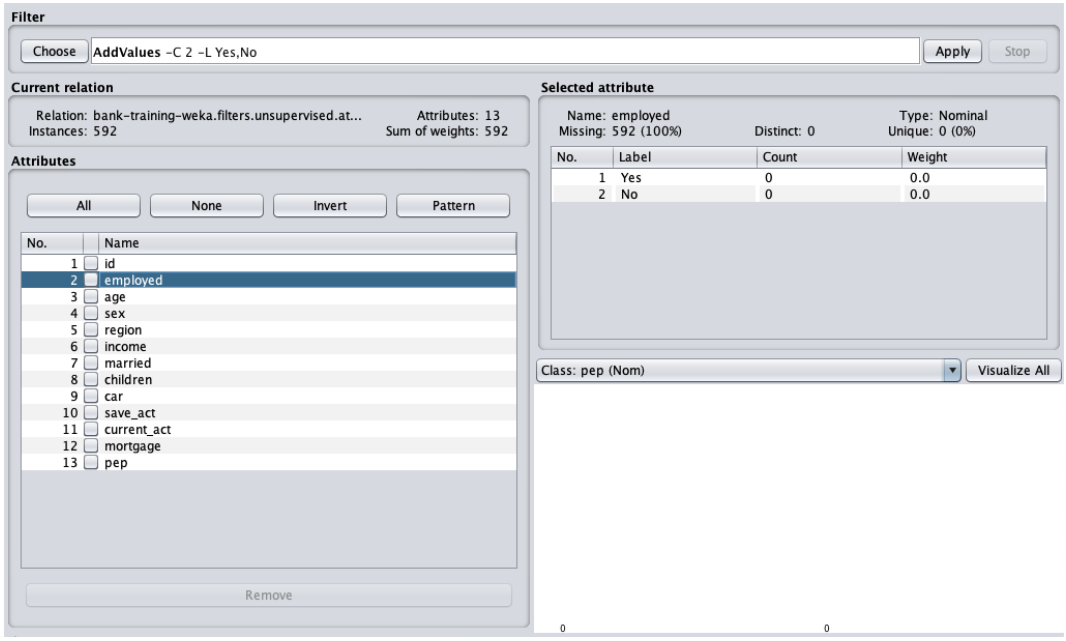
Observations:

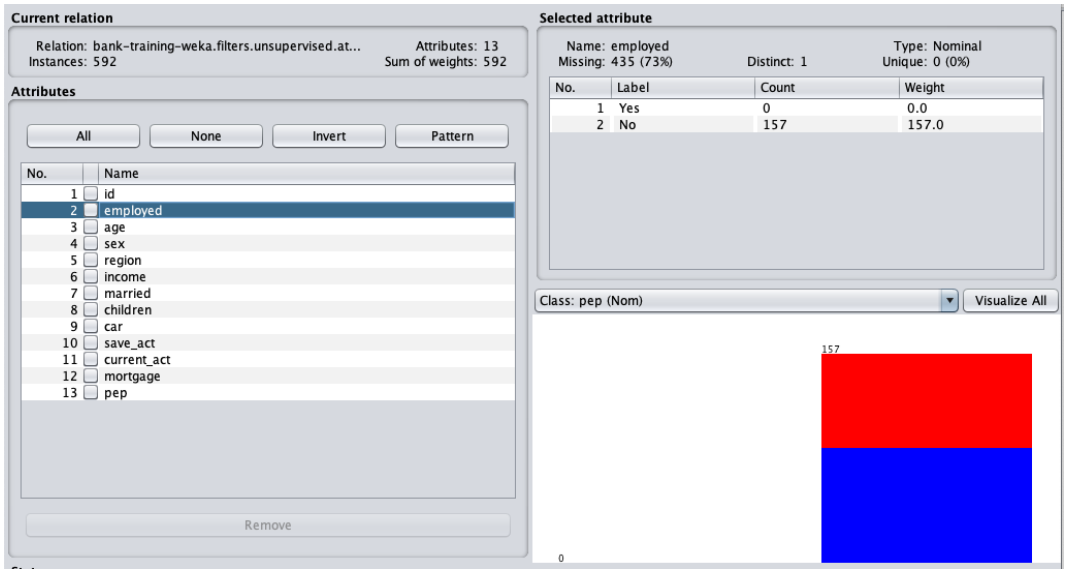
1.a







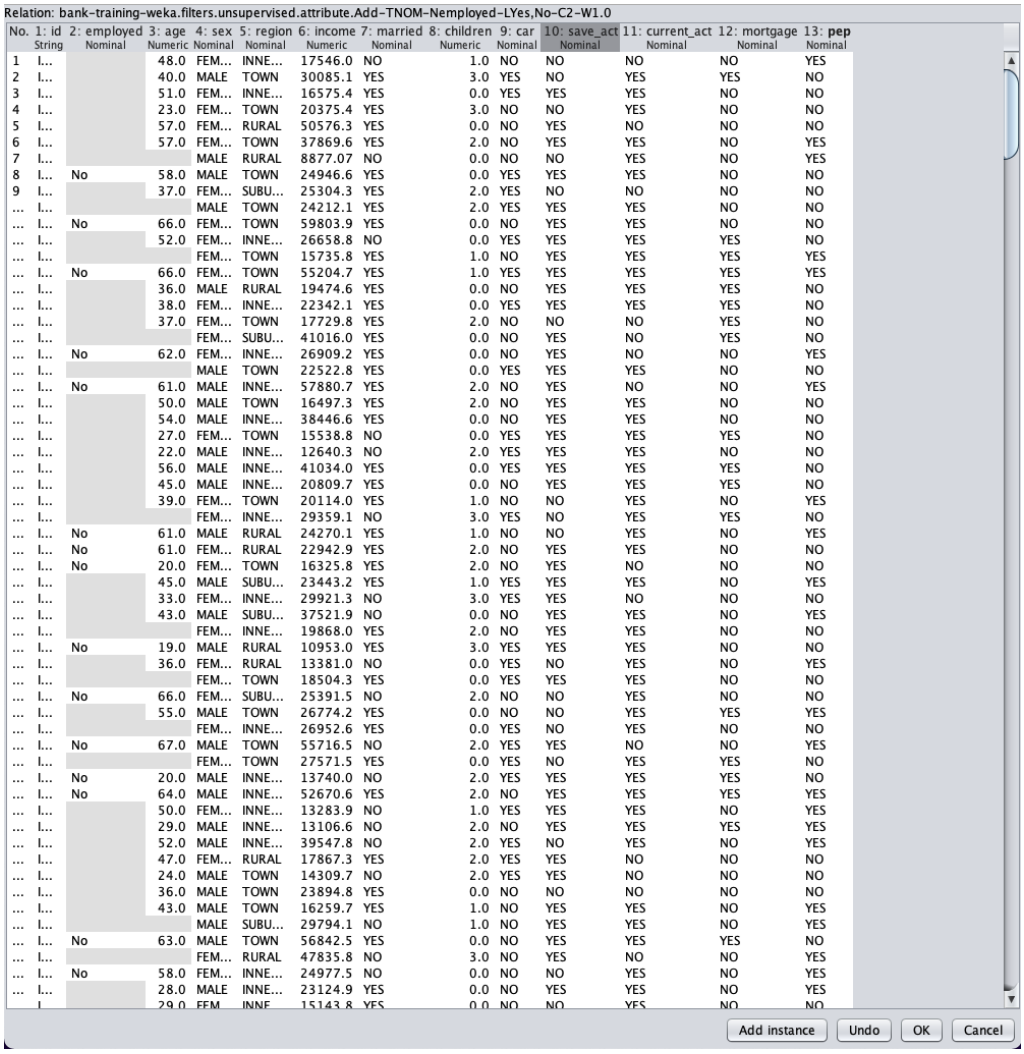




(Manually input values)

1.b

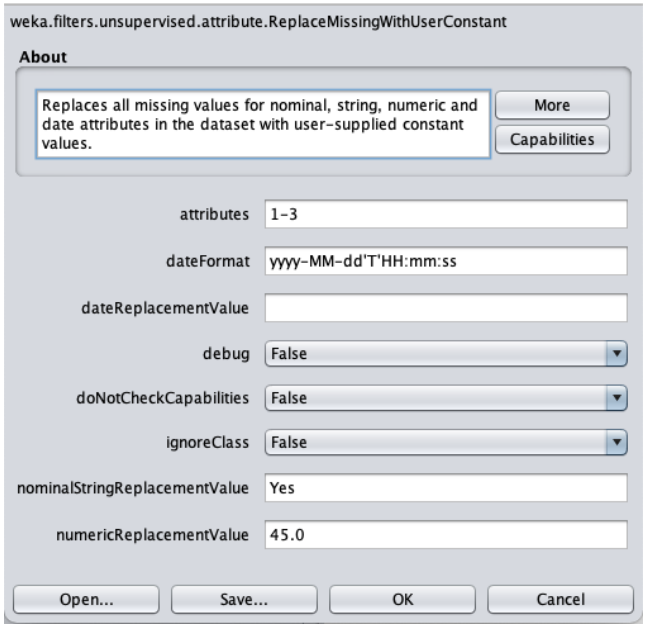
Missing values present in “employed”, “age”



Relation: bank-training-weka.filters.unsupervised.attribute.Add-TNOM-Nemployed-LYes,No-C2-W1.0

No.	1: id String	2: employed Nominal	3: age Numeric	4: sex Nominal	5: region Nominal	6: income Numeric	7: married Nominal	8: children Numeric	9: car Nominal	10: save_act Nominal	11: current_act Nominal	12: mortgage Nominal	13: pep Nominal
1	I...		48.0	FEM...	INNE...	17546.0	NO	1.0	NO	NO	NO	NO	YES
2	I...		40.0	MALE	TOWN	30085.1	YES	3.0	YES	NO	YES	YES	NO
3	I...		51.0	FEM...	INNE...	16575.4	YES	0.0	YES	YES	YES	NO	NO
4	I...		23.0	FEM...	TOWN	20375.4	YES	3.0	NO	NO	YES	NO	NO
5	I...		57.0	FEM...	RURAL	50576.3	YES	0.0	NO	YES	NO	NO	NO
6	I...		57.0	FEM...	TOWN	37869.6	YES	2.0	NO	YES	YES	NO	YES
7	I...			MALE	RURAL	8877.07	NO	0.0	NO	NO	YES	NO	YES
8	I...	No	58.0	MALE	TOWN	24946.6	YES	0.0	YES	YES	YES	NO	NO
9	I...		37.0	FEM...	SUBU...	25304.3	YES	2.0	YES	NO	NO	NO	NO
...	I...			MALE	TOWN	24212.1	YES	2.0	YES	YES	YES	NO	NO
...	I...	No	66.0	FEM...	TOWN	59803.9	YES	0.0	NO	YES	YES	NO	NO
...	I...		52.0	FEM...	INNE...	26658.8	NO	0.0	YES	YES	YES	YES	NO
...	I...			FEM...	TOWN	15735.8	YES	1.0	NO	YES	YES	YES	YES
...	I...	No	66.0	FEM...	TOWN	55204.7	YES	1.0	YES	YES	YES	YES	YES
...	I...		36.0	MALE	RURAL	19474.6	YES	0.0	NO	YES	YES	YES	NO
...	I...		38.0	FEM...	INNE...	22342.1	YES	0.0	YES	YES	YES	YES	NO
...	I...		37.0	FEM...	TOWN	17729.8	YES	2.0	NO	NO	NO	YES	NO
...	I...			FEM...	SUBU...	41016.0	YES	0.0	NO	YES	NO	YES	NO
...	I...	No	62.0	FEM...	INNE...	26909.2	YES	0.0	NO	YES	NO	NO	YES
...	I...			MALE	TOWN	22522.8	YES	0.0	YES	YES	YES	NO	NO
...	I...	No	61.0	MALE	INNE...	57880.7	YES	2.0	NO	YES	NO	NO	YES
...	I...		50.0	MALE	TOWN	16497.3	YES	2.0	NO	YES	YES	NO	NO
...	I...		54.0	MALE	INNE...	38446.6	YES	0.0	NO	YES	YES	NO	NO
...	I...		27.0	FEM...	TOWN	15538.8	NO	0.0	YES	YES	YES	YES	NO
...	I...		22.0	MALE	INNE...	12640.3	NO	2.0	YES	YES	YES	NO	NO
...	I...		56.0	MALE	INNE...	41034.0	YES	0.0	YES	YES	YES	YES	NO
...	I...		45.0	MALE	INNE...	20809.7	YES	0.0	NO	YES	YES	YES	NO
...	I...		39.0	FEM...	TOWN	20114.0	YES	1.0	NO	NO	YES	NO	YES
...	I...			FEM...	INNE...	29359.1	NO	3.0	YES	NO	YES	YES	NO
...	I...	No	61.0	MALE	RURAL	24270.1	YES	1.0	NO	NO	YES	NO	YES
...	I...	No	61.0	FEM...	RURAL	22942.9	YES	2.0	NO	YES	YES	NO	NO
...	I...	No	20.0	FEM...	TOWN	16325.8	YES	2.0	NO	YES	NO	NO	NO
...	I...		45.0	MALE	SUBU...	23443.2	YES	1.0	YES	YES	YES	NO	YES
...	I...		33.0	FEM...	INNE...	29921.3	NO	3.0	YES	NO	NO	NO	NO
...	I...		43.0	MALE	SUBU...	37521.9	NO	0.0	NO	YES	YES	NO	YES
...	I...			FEM...	INNE...	19868.0	YES	2.0	NO	YES	YES	NO	NO
...	I...	No	19.0	MALE	RURAL	10953.0	YES	3.0	YES	YES	YES	NO	NO
...	I...		36.0	FEM...	RURAL	13381.0	NO	0.0	YES	NO	YES	NO	YES
...	I...			FEM...	TOWN	18504.3	YES	0.0	YES	YES	YES	NO	NO
...	I...	No	66.0	FEM...	SUBU...	25391.5	NO	2.0	NO	NO	YES	NO	NO
...	I...		55.0	MALE	TOWN	26774.2	YES	0.0	NO	NO	YES	YES	YES
...	I...			FEM...	INNE...	26952.6	YES	0.0	YES	NO	YES	NO	NO
...	I...	No	67.0	MALE	TOWN	55716.5	NO	2.0	YES	YES	NO	NO	YES
...	I...			FEM...	TOWN	27571.5	YES	0.0	YES	NO	YES	YES	NO
...	I...	No	20.0	MALE	INNE...	13740.0	NO	2.0	YES	YES	YES	YES	NO
...	I...	No	64.0	MALE	INNE...	52670.6	YES	2.0	NO	YES	YES	YES	YES
...	I...		50.0	FEM...	INNE...	13283.9	NO	1.0	YES	YES	YES	NO	YES
...	I...		29.0	MALE	INNE...	13106.6	NO	2.0	NO	YES	YES	YES	YES
...	I...		52.0	MALE	INNE...	39547.8	NO	2.0	YES	NO	YES	NO	YES
...	I...		47.0	FEM...	RURAL	17867.3	YES	2.0	YES	YES	NO	NO	NO
...	I...		24.0	MALE	TOWN	14309.7	NO	2.0	YES	YES	NO	NO	NO
...	I...		36.0	MALE	TOWN	23894.8	YES	0.0	NO	NO	NO	NO	NO
...	I...		43.0	MALE	TOWN	16259.7	YES	1.0	NO	YES	YES	NO	YES
...	I...			MALE	SUBU...	29794.1	NO	1.0	NO	YES	YES	NO	YES
...	I...	No	63.0	MALE	TOWN	56842.5	YES	0.0	NO	YES	YES	YES	NO
...	I...			FEM...	RURAL	47835.8	NO	3.0	NO	YES	NO	NO	YES
...	I...	No	58.0	FEM...	INNE...	24977.5	NO	0.0	NO	NO	YES	NO	YES
...	I...		28.0	MALE	INNE...	23124.9	YES	0.0	NO	YES	YES	NO	YES
...	I...		29.0	FEM...	INNE...	15143.8	YES	0.0	NO	NO	YES	NO	NO

Add instance | Undo | OK | Cancel



weka.filters.unsupervised.attribute.ReplaceMissingWithUserConstant

About

Replaces all missing values for nominal, string, numeric and date attributes in the dataset with user-supplied constant values.

More | Capabilities

attributes	1-3
dateFormat	yyyy-MM-dd'T'HH:mm:ss
dateReplacementValue	
debug	False
doNotCheckCapabilities	False
ignoreClass	False
nominalStringReplacementValue	Yes
numericReplacementValue	45.0

Open... | Save... | OK | Cancel

Relation: bank-training-weka.filters.unsupervised.attribute.Add-TNOM-Nemployed-LYes,No-C2-W1.0-weka.filters.unsupervised.attribute.ReplaceMissingWithUserConsta...

No.	1: id String	2: employed Nominal	3: age Numeric	4: sex Nominal	5: region Nominal	6: income Numeric	7: married Nominal	8: children Numeric	9: car Nominal	10: save_act Nominal	11: current_act Nominal	12: mortgage Nominal	13: pep Nominal
1	I...	Yes	48.0	FEM...	INNE...	17546.0	NO	1.0	NO	NO	NO	NO	YES
2	I...	Yes	40.0	MALE	TOWN	30085.1	YES	3.0	YES	NO	YES	YES	NO
3	I...	Yes	51.0	FEM...	INNE...	16575.4	YES	0.0	YES	YES	YES	NO	NO
4	I...	Yes	23.0	FEM...	TOWN	20375.4	YES	3.0	NO	NO	YES	NO	NO
5	I...	Yes	57.0	FEM...	RURAL	50576.3	YES	0.0	NO	YES	NO	NO	NO
6	I...	Yes	57.0	FEM...	TOWN	37869.6	YES	2.0	NO	YES	YES	NO	YES
7	I...	Yes	45.0	MALE	RURAL	8877.07	NO	0.0	NO	NO	YES	NO	YES
8	I...	No	58.0	MALE	TOWN	24946.6	YES	0.0	YES	YES	YES	NO	NO
9	I...	Yes	37.0	FEM...	SUBU...	25304.3	YES	2.0	YES	NO	NO	NO	NO
...	I...	Yes	45.0	MALE	TOWN	24212.1	YES	2.0	YES	YES	YES	NO	NO
...	I...	No	66.0	FEM...	TOWN	59803.9	YES	0.0	NO	YES	YES	NO	NO
...	I...	Yes	52.0	FEM...	INNE...	26658.8	NO	0.0	YES	YES	YES	YES	NO
...	I...	Yes	45.0	FEM...	TOWN	15735.8	YES	1.0	NO	YES	YES	YES	YES
...	I...	No	66.0	FEM...	TOWN	55204.7	YES	1.0	YES	YES	YES	YES	YES
...	I...	Yes	36.0	MALE	RURAL	19474.6	YES	0.0	NO	YES	YES	YES	NO
...	I...	Yes	38.0	FEM...	INNE...	22342.1	YES	0.0	YES	YES	YES	YES	NO
...	I...	Yes	37.0	FEM...	TOWN	17729.8	YES	2.0	NO	NO	NO	YES	NO
...	I...	Yes	45.0	FEM...	SUBU...	41016.0	YES	0.0	NO	YES	NO	YES	NO
...	I...	No	62.0	FEM...	INNE...	26909.2	YES	0.0	NO	YES	NO	NO	YES
...	I...	Yes	45.0	MALE	TOWN	22522.8	YES	0.0	YES	YES	YES	NO	NO
...	I...	No	61.0	MALE	INNE...	57880.7	YES	2.0	NO	YES	NO	NO	YES
...	I...	Yes	50.0	MALE	TOWN	16497.3	YES	2.0	NO	YES	YES	NO	NO
...	I...	Yes	54.0	MALE	INNE...	38446.6	YES	0.0	NO	YES	YES	NO	NO
...	I...	Yes	27.0	FEM...	TOWN	15538.8	NO	0.0	YES	YES	YES	YES	NO
...	I...	Yes	22.0	MALE	INNE...	12640.3	NO	2.0	YES	YES	YES	NO	NO
...	I...	Yes	56.0	MALE	INNE...	41034.0	YES	0.0	YES	YES	YES	YES	NO
...	I...	Yes	45.0	MALE	INNE...	20809.7	YES	0.0	NO	YES	YES	YES	NO
...	I...	Yes	39.0	FEM...	TOWN	20114.0	YES	1.0	NO	NO	YES	NO	YES
...	I...	Yes	45.0	FEM...	INNE...	29359.1	NO	3.0	YES	NO	YES	YES	NO
...	I...	No	61.0	MALE	RURAL	24270.1	YES	1.0	NO	NO	YES	NO	YES
...	I...	No	61.0	FEM...	RURAL	22942.9	YES	2.0	NO	YES	YES	NO	NO
...	I...	No	20.0	FEM...	TOWN	16325.8	YES	2.0	NO	YES	NO	NO	NO
...	I...	Yes	45.0	MALE	SUBU...	23443.2	YES	1.0	YES	YES	YES	NO	YES
...	I...	Yes	33.0	FEM...	INNE...	29921.3	NO	3.0	YES	YES	NO	NO	NO
...	I...	Yes	43.0	MALE	SUBU...	37521.9	NO	0.0	NO	YES	YES	NO	YES
...	I...	Yes	45.0	FEM...	INNE...	19868.0	YES	2.0	NO	YES	YES	NO	NO
...	I...	No	19.0	MALE	RURAL	10953.0	YES	3.0	YES	YES	YES	NO	NO
...	I...	Yes	36.0	FEM...	RURAL	13381.0	NO	0.0	YES	NO	YES	NO	YES
...	I...	Yes	45.0	FEM...	TOWN	18504.3	YES	0.0	YES	YES	YES	NO	NO
...	I...	No	66.0	FEM...	SUBU...	25391.5	NO	2.0	NO	NO	YES	NO	NO
...	I...	Yes	55.0	MALE	TOWN	26774.2	YES	0.0	NO	NO	YES	YES	YES
...	I...	Yes	45.0	FEM...	INNE...	26952.6	YES	0.0	YES	NO	YES	NO	NO
...	I...	No	67.0	MALE	TOWN	55716.5	NO	2.0	YES	NO	NO	NO	YES
...	I...	Yes	45.0	FEM...	TOWN	27571.5	YES	0.0	YES	NO	YES	YES	NO
...	I...	No	20.0	MALE	INNE...	13740.0	NO	2.0	YES	YES	YES	YES	NO
...	I...	No	64.0	MALE	INNE...	52670.6	YES	2.0	NO	YES	YES	YES	YES
...	I...	Yes	50.0	FEM...	INNE...	13283.9	NO	1.0	YES	YES	YES	NO	YES
...	I...	Yes	29.0	MALE	INNE...	13106.6	NO	2.0	NO	YES	YES	YES	YES
...	I...	Yes	52.0	MALE	INNE...	39547.8	NO	2.0	YES	NO	YES	NO	YES
...	I...	Yes	47.0	FEM...	RURAL	17867.3	YES	2.0	YES	NO	NO	NO	NO
...	I...	Yes	24.0	MALE	TOWN	14309.7	NO	2.0	YES	YES	NO	NO	NO
...	I...	Yes	36.0	MALE	TOWN	23894.8	YES	0.0	NO	NO	NO	NO	NO
...	I...	Yes	43.0	MALE	TOWN	16259.7	YES	1.0	NO	YES	YES	NO	YES
...	I...	Yes	45.0	MALE	SUBU...	29794.1	NO	1.0	NO	YES	YES	NO	YES
...	I...	No	63.0	MALE	TOWN	56842.5	YES	0.0	NO	YES	YES	YES	NO
...	I...	Yes	45.0	FEM...	RURAL	47835.8	NO	3.0	NO	YES	NO	NO	YES
...	I...	No	58.0	FEM...	INNE...	24977.5	NO	0.0	NO	NO	YES	NO	YES
...	I...	Yes	28.0	MALE	INNE...	23124.9	YES	0.0	NO	YES	YES	NO	YES
...	I...	Yes	29.0	FEM...	INNE...	15143.8	YES	0.0	NO	NO	YES	NO	NO

Add instance | Undo | OK | Cancel

Filter

Choose | ReplaceMissingWithUserConstant -A 1-3 -N Yes -R 45.0 -F "yyyy-MM-dd\'T\'HH:mm:ss" | Apply | Stop

Current relation

Relation: bank-training-weka.filters.unsupervised.at... Attributes: 13
Instances: 592 Sum of weights: 592

Attributes

All | None | Invert | Pattern

No.	Name
1	id
2	employed
3	age
4	sex
5	region
6	income
7	married
8	children
9	car
10	save_act
11	current_act
12	mortgage
13	pep

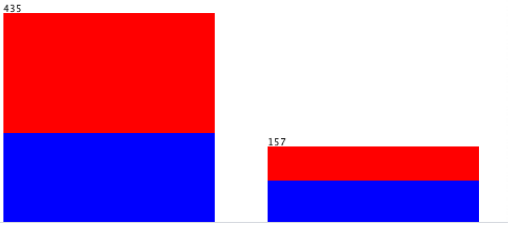
Remove

Selected attribute

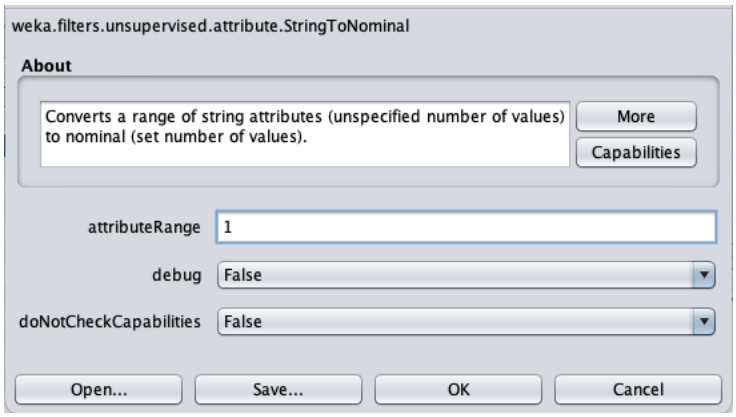
Name: employed Type: Nominal
Missing: 0 (0%) Distinct: 2 Unique: 0 (0%)

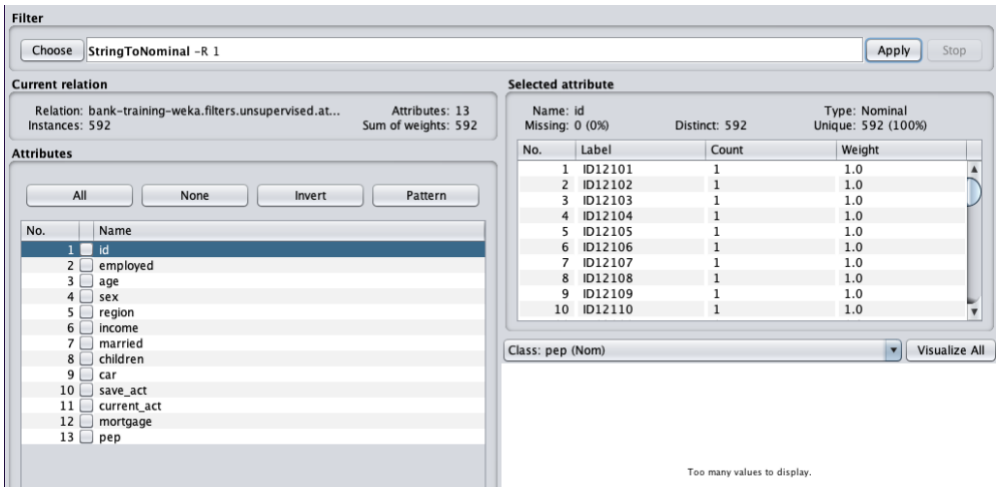
No.	Label	Count	Weight
1	Yes	435	435.0
2	No	157	157.0

Class: pep (Nom) | Visualize All

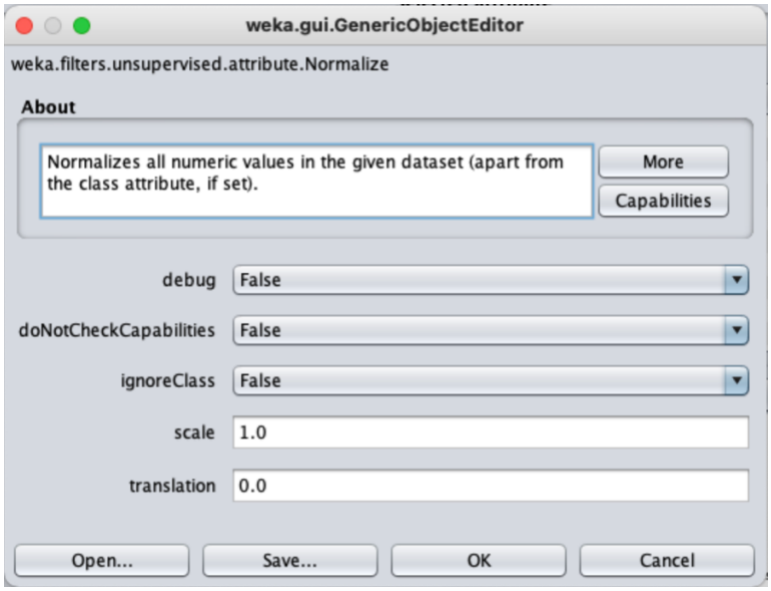


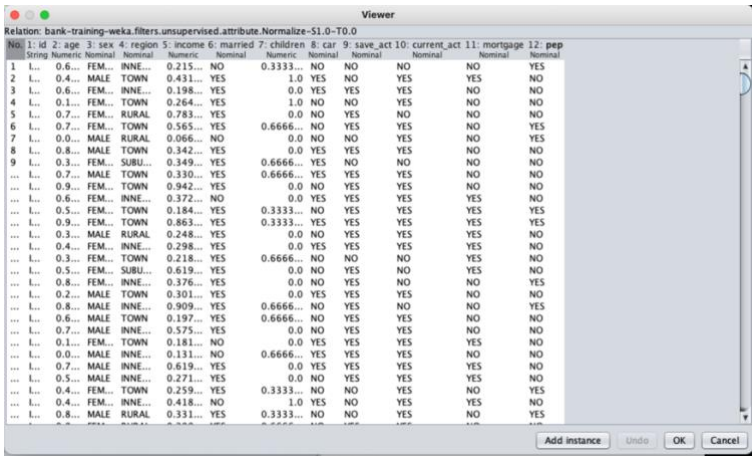
1.c


weka.filters.unsupervised.attribute.StringToNominal
About
Converts a range of string attributes (unspecified number of values) to nominal (set number of values).
More
Capabilities
attributeRange
1
debug
False
doNotCheckCapabilities
False
Open...
Save...
OK
Cancel


Filter
Choose
StringToNominal -R 1
Apply
Stop
Current relation
Relation: bank-training-weka.filters.unsupervised.at...
Instances: 592
Attributes: 13
Sum of weights: 592
Attributes
All
None
Invert
Pattern
No. Name
1 id
2 employed
3 age
4 sex
5 region
6 income
7 married
8 children
9 car
10 save_act
11 current_act
12 mortgage
13 pep
Selected attribute
Name: id
Missing: 0 (0%)
Distinct: 592
Type: Nominal
Unique: 592 (100%)
No. Label Count Weight
1 ID12101 1 1.0
2 ID12102 1 1.0
3 ID12103 1 1.0
4 ID12104 1 1.0
5 ID12105 1 1.0
6 ID12106 1 1.0
7 ID12107 1 1.0
8 ID12108 1 1.0
9 ID12109 1 1.0
10 ID12110 1 1.0
Class: pep (Nom)
Visualize All
Too many values to display.

1.d

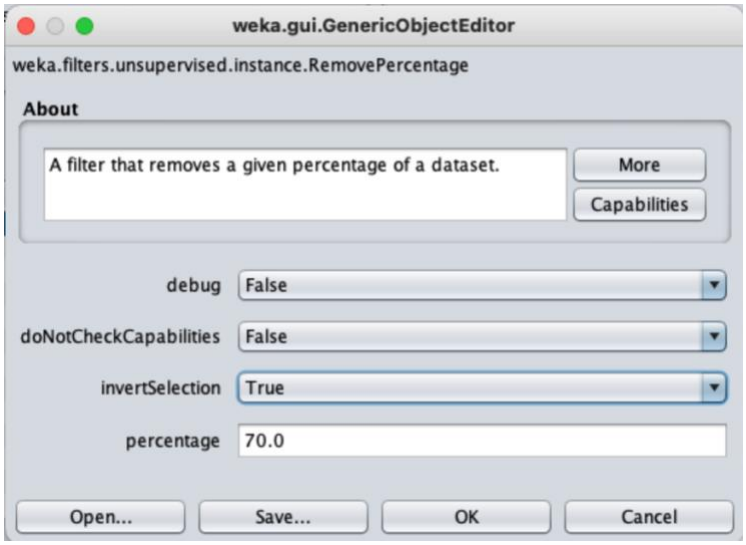

weka.gui.GenericObjectEditor
weka.filters.unsupervised.attribute.Normalize
About
Normalizes all numeric values in the given dataset (apart from the class attribute, if set).
More
Capabilities
debug
False
doNotCheckCapabilities
False
ignoreClass
False
scale
1.0
translation
0.0
Open...
Save...
OK
Cancel

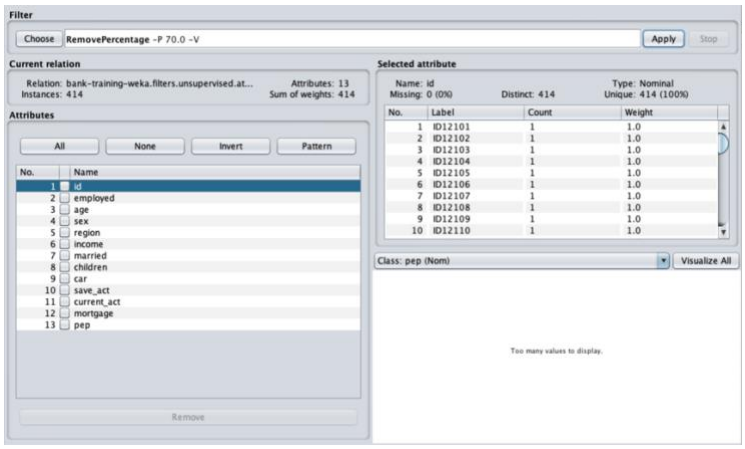


(Performed on unedited database)

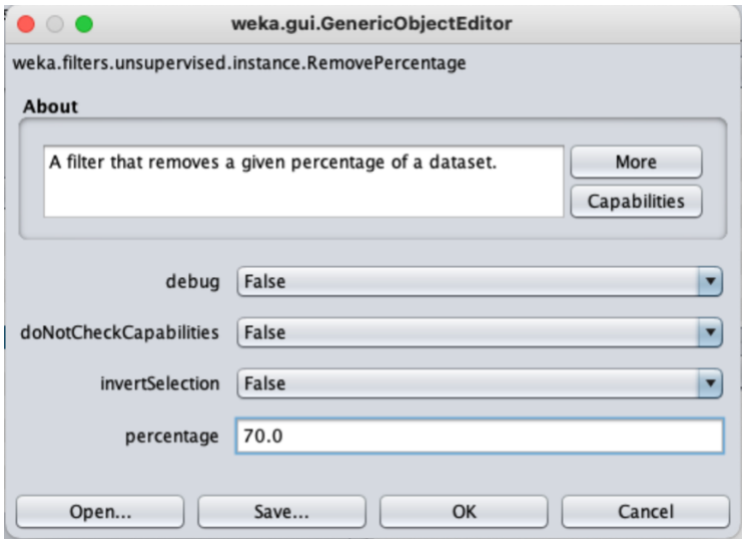
1.e

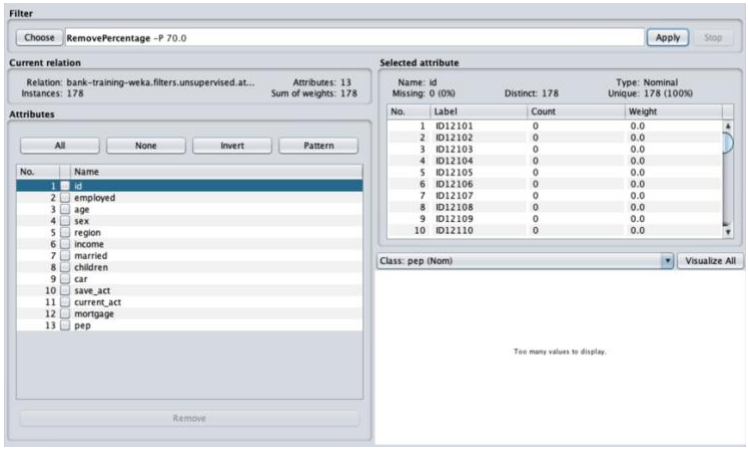
Training Data Set



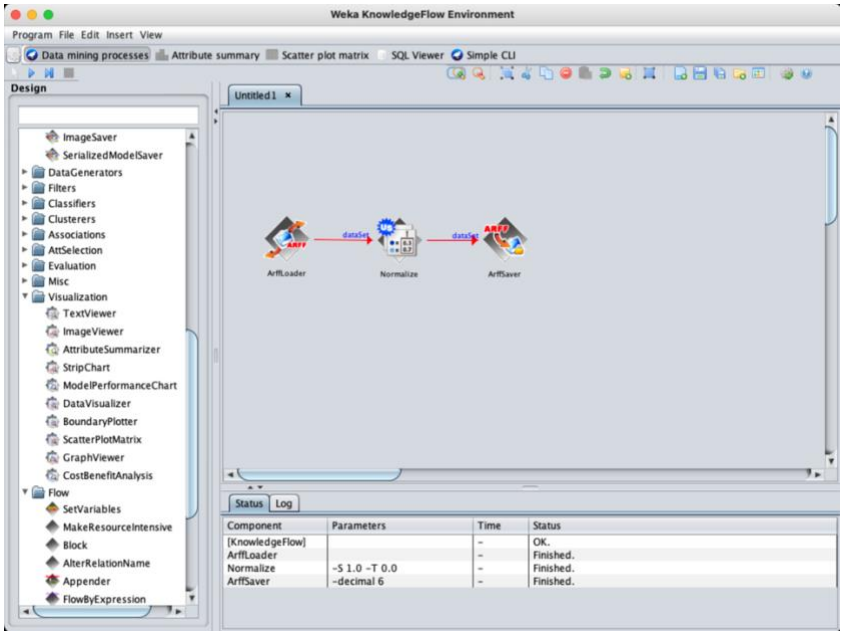


Testing Data Set





2.



Conclusion:

- **Through this practical, we were firstly introduced to a whole new slew of filters available in the Weka**
- **We learnt about adding data to the dataset by inserting columns and adding values**
- **We learnt to deal with missing data by using filters that either fill the data using average or the more common data or filling everything with user defined options**
- **We also learnt to convert strings into nominal data types using the StringToNominal filter**
- **Further, we learnt to normalize numeric data**
- **Additionally, we learnt to split the database into training and testing data using the RemovePercentage filter**
- **Finally, we learnt about knowledge flows and the means of creating one in Weka.**

****