IMPLEMENTATION OF
FLEXIBLE QUERY ANSWERING SYSTEM
PROJECT REPORT

SUBJECT: *knowledge Based Systems (SPRING'15)*

Submitted by

Samatha Kolagatla (800864273)

SaiVarun Reddy Patlori (800867606)

Venkata Sonika Desu (800866876)

INTRODUCTION:

This report exhibits the implementation of Flexible Query Answering System(FQAS) for the music data set.

Software used to implement:

"ORANGE" software.

About Orange Software:

Orange is a component-based data mining and machine learning software suite, featuring a visual programming front-end for explorative data analysis and visualization, and Python bindings and libraries for scripting. It includes a set of components for data preprocessing, feature scoring and filtering, modeling, model evaluation, and exploration techniques. It is implemented in C++ and Python. Its graphical user interface builds upon the cross-platform Qt framework.
It is maintained and developed at the Bioinformatics Laboratory of the Faculty of Computer and Information Science, University of Ljubljana, Slovenia.

PROCESS DESCRIPTION:

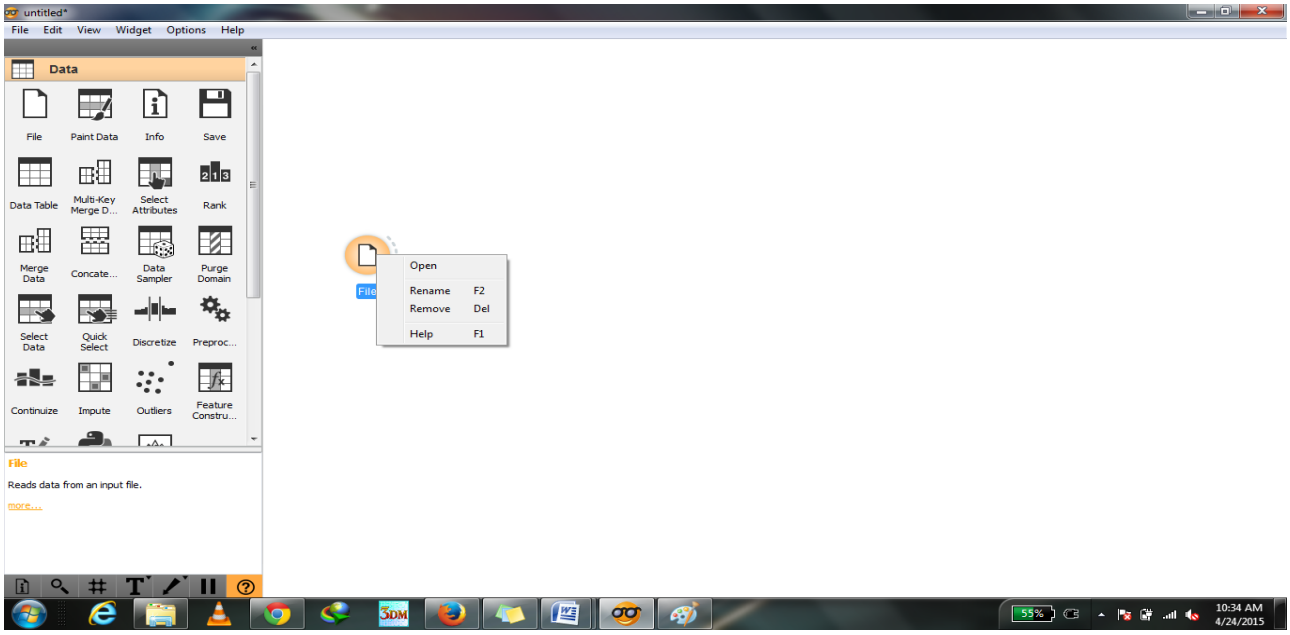
We have shown a stepwise report with details in every step.

Step-1: Loading the Dataset (MusicDataSet.txt) using the widget "**file**".

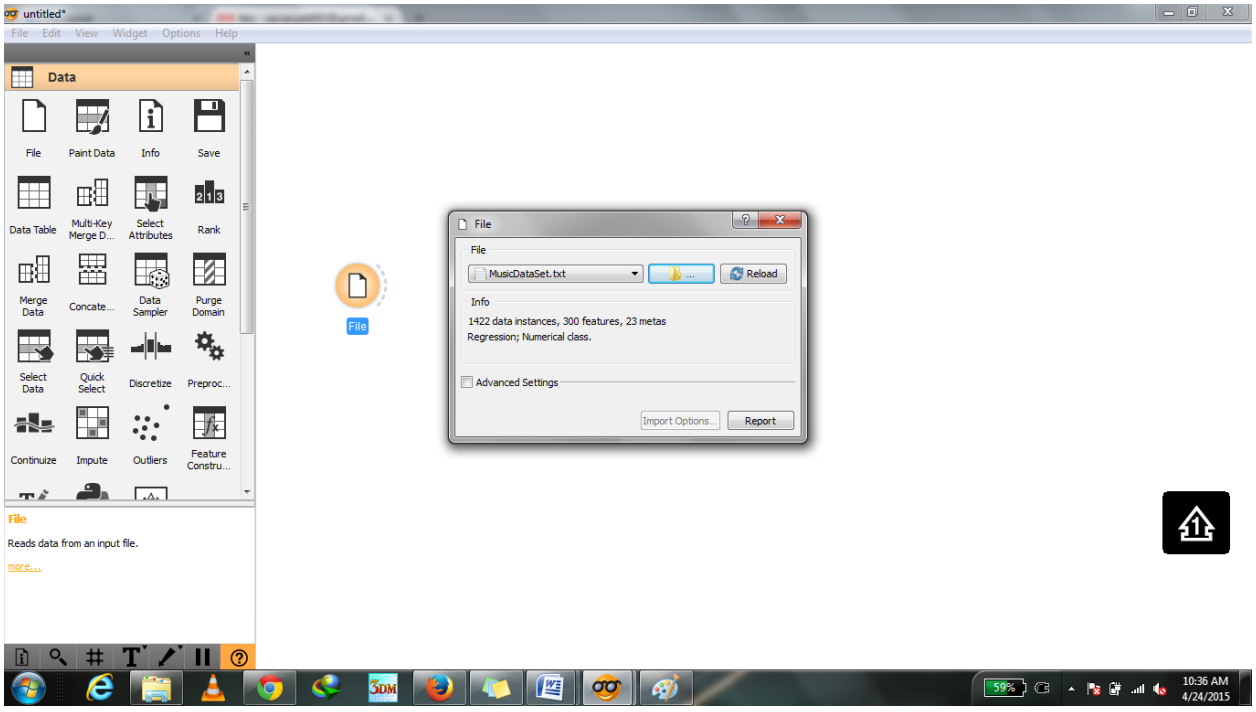
File:
Reads attribute-value data from an input file.

Signals:
Input: None
Output: Data

Data file consists of data table with data instances. File widget reads the input data file from the attribute-valued data set and sends the data set to its output channel. It maintains a history of most recently opened files.



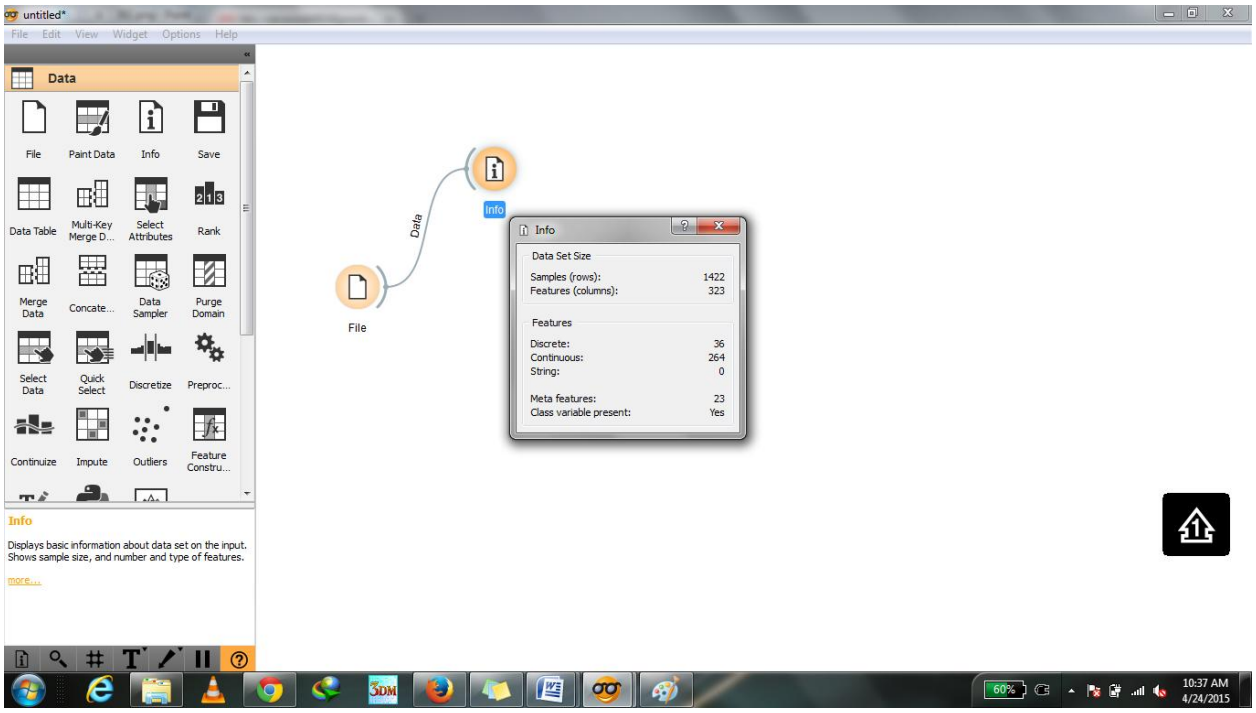
Step-2: Importing the MusicDataSet.txt file from the location it was stored in PC.



Step-3: Displaying the information about the file "MusicDataSet.txt" using the widget "Info"

Info:
In this step, basic information about data set (given as input) is displayed.
Sample size and number and type of features are shown

Signals:
Inputs: Data (File).
Output: No outputs.



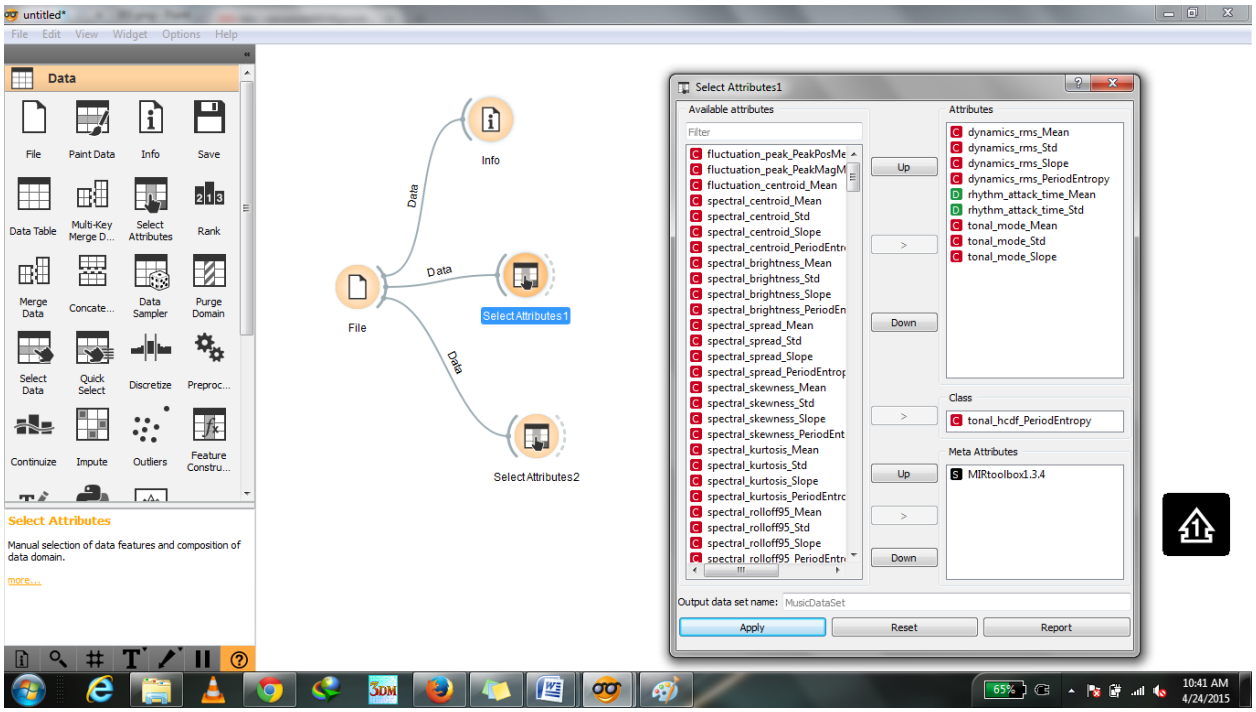
Step-4: Selecting attributes using the "Select Attributes" widget.

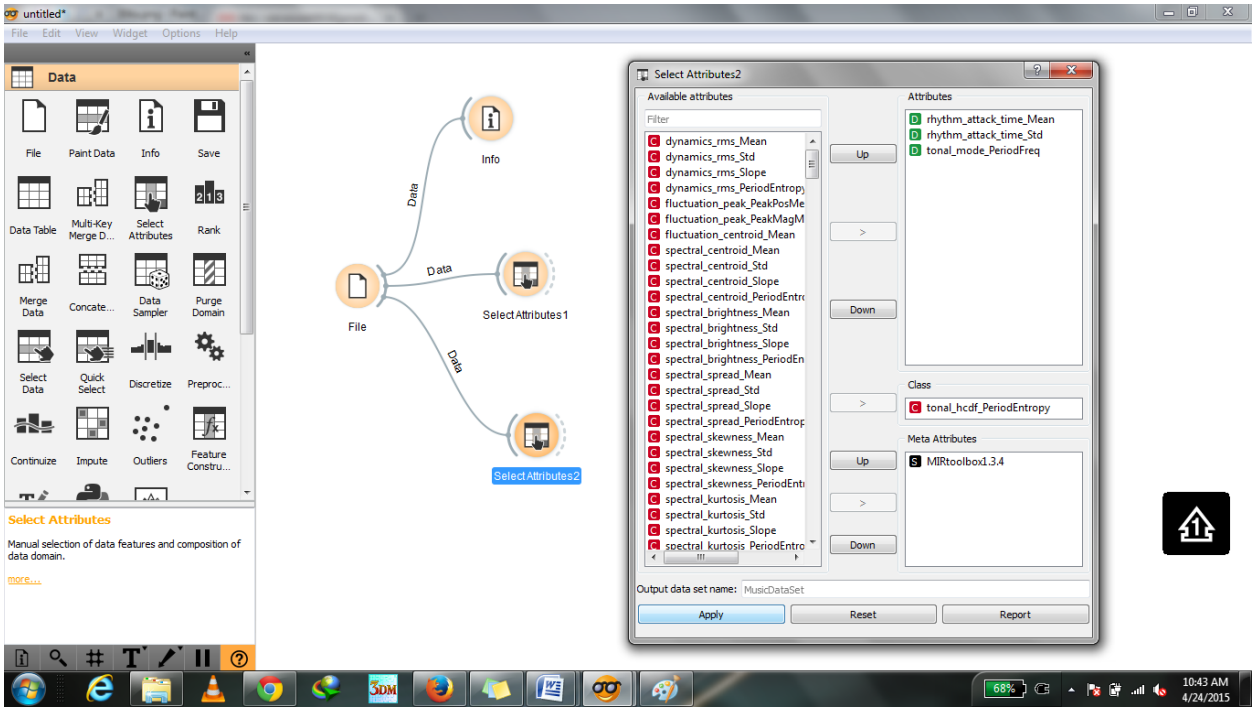
Select Attributes:
Select attributes is used in manual selection of data attributes and composition of data domain.

Signals:
Input: Data (Attribute-valued data set).
Output: Data (Attribute-valued data set composed using the domain specification from the widget.

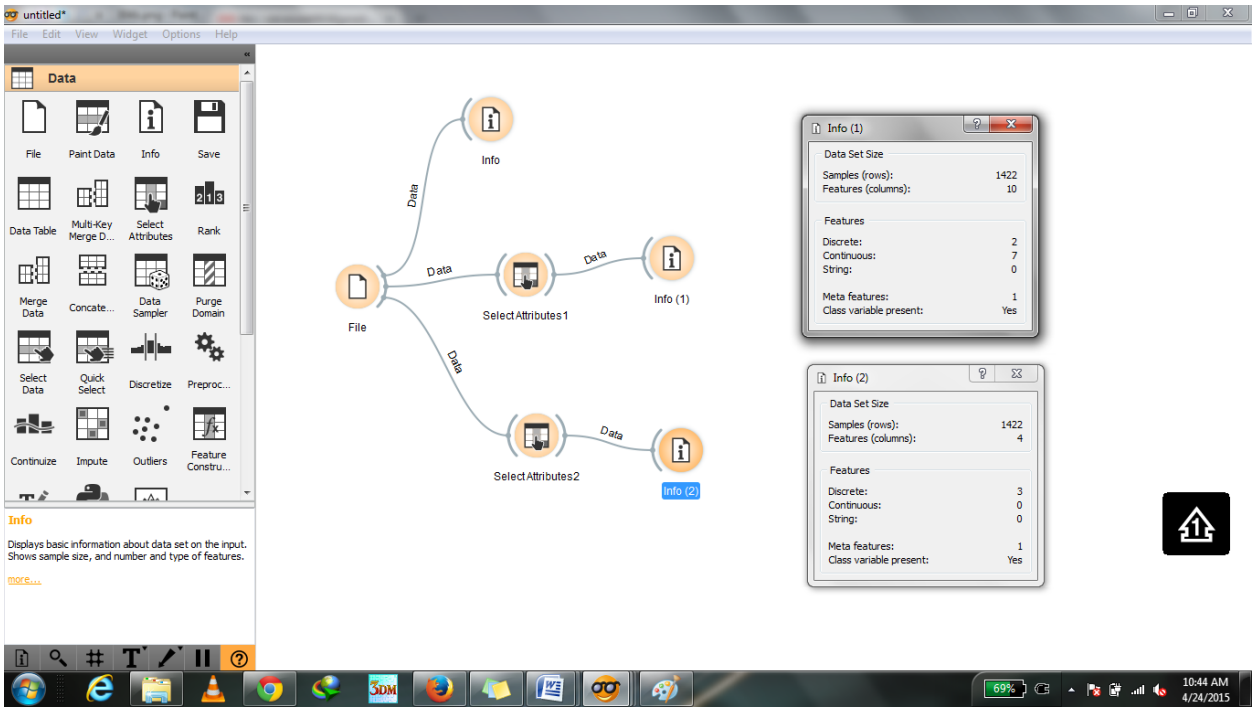
Select Attributes widget is used to manually compose your data domain. It is also used to manually select the data attributes .That is, user can decide which attributes will be used and how it is used. Orange distinguishes between ordinary attributes, (optional) class attributes and Meta attributes.

It can be seen that we have taken two sets of data to perform concatenation which will be described in the coming pages. Taking two sets of data to perform concatenation is an additional feature.





Step-5: Displaying the info of the selected attributes.

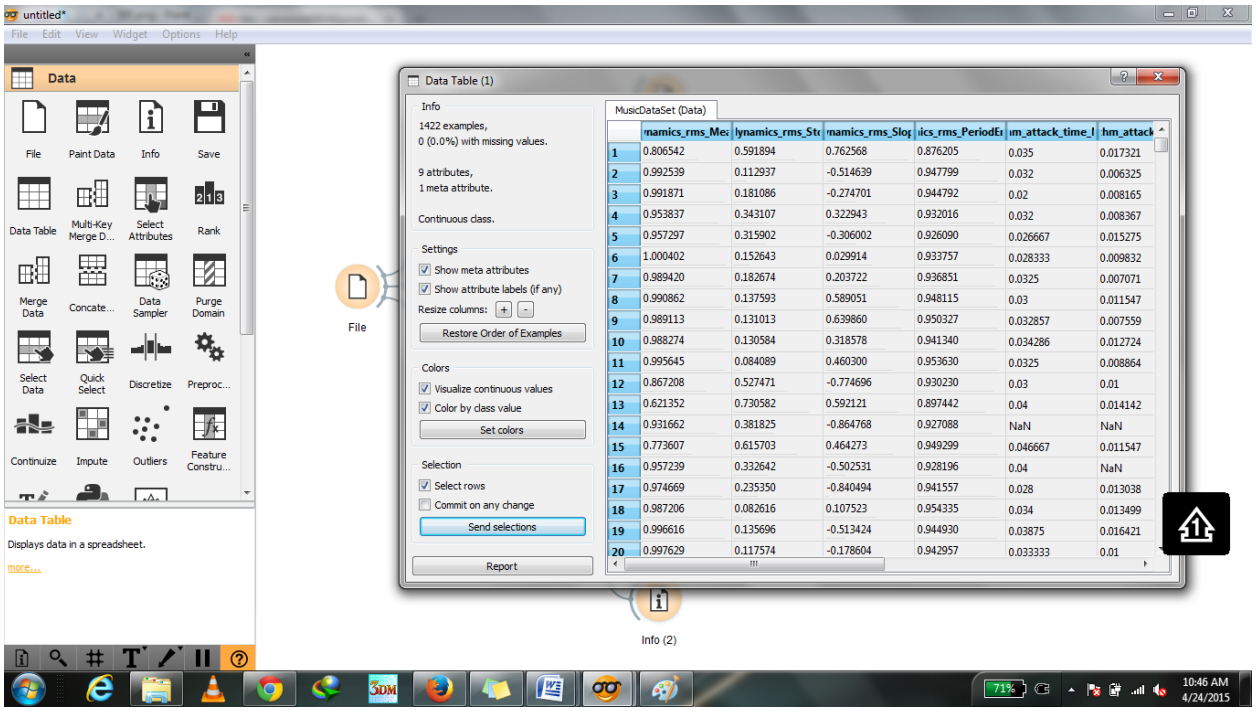


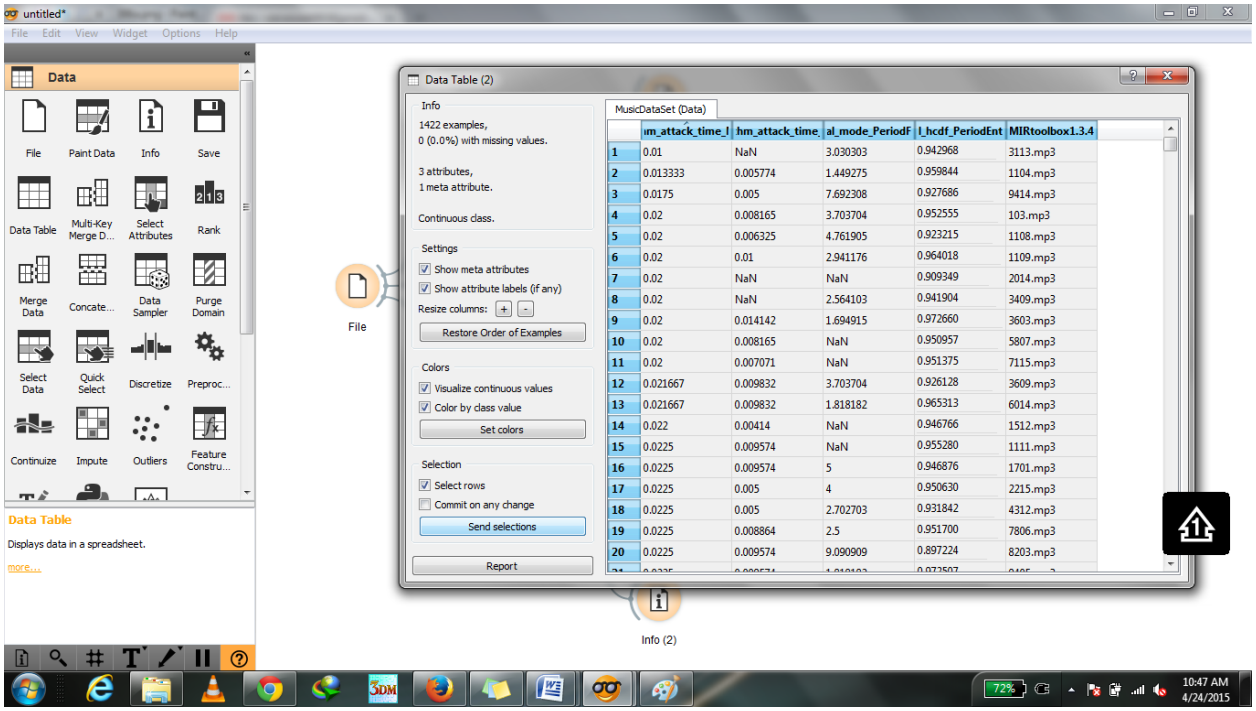
Step-6: Using "Data Table" widget to display attribute-value data in a spread sheet.

Data Table:
One or more data sets is received by Data Table widget on its input and displays them in a spreadsheet format. Data instances may be sorted by attribute values. Manual selection of data instances is also supported by this Widget.

Signals:
Input: Data (Attribute-valued data set).
Output: Selected Data (Selected data instances).

In this step, data is manually selected and send selections button is clicked, then it sends the data to Select Data Widget. This is done as shown in the following two screenshots.


untitled*
File Edit View Widget Options Help
Data
File
Paint Data
Info
Save
Data Table
Multi-Key Merge D...
Select Attributes
Rank
Merge Data
Concate...
Data Sampler
Purge Domain
Select Data
Quick Select
Discretize
Preproc...
Continuize
Impute
Outliers
Feature Constru...
Data Table
Displays data in a spreadsheet.
more...
Data Table (1)
Info
1422 examples,
0 (0.0%) with missing values.
9 attributes,
1 meta attribute.
Continuous class.
Settings
Show meta attributes
Show attribute labels (if any)
Resize columns:
Restore Order of Examples
Colors
Visualize continuous values
Color by class value
Set colors
Selection
Select rows
Commit on any change
Send selections
Report
MusicDataSet (Data)
namics_rms_Mea
lynamics_rms_Std
namics_rms_Slop
ics_rms_PeriodEn
im_attack_time_l
hm_attack
1 0.806542 0.591894 0.762568 0.876205 0.035 0.017321
2 0.992539 0.112937 -0.514639 0.947799 0.032 0.006325
3 0.991871 0.181086 -0.274701 0.944792 0.02 0.008165
4 0.953837 0.343107 0.322943 0.932016 0.032 0.008367
5 0.957297 0.315902 -0.306002 0.926090 0.026667 0.015275
6 1.000402 0.152643 0.029914 0.933757 0.028333 0.009832
7 0.989420 0.182674 0.203722 0.936851 0.0325 0.007071
8 0.990862 0.137593 0.589051 0.948115 0.03 0.011547
9 0.989113 0.131013 0.639860 0.950327 0.032857 0.007559
10 0.988274 0.130584 0.318578 0.941340 0.034286 0.012724
11 0.995645 0.084089 0.460300 0.953630 0.0325 0.008864
12 0.867208 0.527471 -0.774696 0.930230 0.03 0.01
13 0.621352 0.730582 0.592121 0.897442 0.04 0.014142
14 0.931662 0.381825 -0.864768 0.927088 NaN NaN
15 0.773607 0.615703 0.464273 0.949299 0.046667 0.011547
16 0.957239 0.332642 -0.502531 0.928196 0.04 NaN
17 0.974669 0.235350 -0.840494 0.941557 0.028 0.013038
18 0.987206 0.082616 0.107523 0.954335 0.034 0.013499
19 0.996616 0.135696 -0.513424 0.944930 0.03875 0.016421
20 0.997629 0.117574 -0.178604 0.942957 0.033333 0.01
File
Info (2)
71%
10:46 AM
4/24/2015


untitled*
File Edit View Widget Options Help
Data
File
Paint Data
Info
Save
Data Table
Multi-Key Merge D...
Select Attributes
Rank
Merge Data
Concate...
Data Sampler
Purge Domain
Select Data
Quick Select
Discretize
Preproc...
Continuize
Impute
Outliers
Feature Constru...
Data Table
Displays data in a spreadsheet.
more...
Data Table (2)
Info
1422 examples,
0 (0.0%) with missing values.
3 attributes,
1 meta attribute.
Continuous class.
Settings
Show meta attributes
Show attribute labels (if any)
Resize columns:
Restore Order of Examples
Colors
Visualize continuous values
Color by class value
Set colors
Selection
Select rows
Commit on any change
Send selections
Report
MusicDataSet (Data)
im_attack_time_l
hm_attack_time_
al_mode_PeriodF
l_hcdf_PeriodEnt
MIRtoolbox1.3.4
1 0.01 NaN 3.030303 0.942968 3113.mp3
2 0.013333 0.005774 1.449275 0.959844 1104.mp3
3 0.0175 0.005 7.692308 0.927686 9414.mp3
4 0.02 0.008165 3.703704 0.952555 103.mp3
5 0.02 0.006325 4.761905 0.923215 1108.mp3
6 0.02 0.01 2.941176 0.964018 1109.mp3
7 0.02 NaN NaN 0.909349 2014.mp3
8 0.02 NaN 2.564103 0.941904 3409.mp3
9 0.02 0.014142 1.694915 0.972660 3603.mp3
10 0.02 0.008165 NaN 0.950957 5807.mp3
11 0.02 0.007071 NaN 0.951375 7115.mp3
12 0.021667 0.009832 3.703704 0.926128 3609.mp3
13 0.021667 0.009832 1.818182 0.965313 6014.mp3
14 0.022 0.00414 NaN 0.946766 1512.mp3
15 0.0225 0.009574 NaN 0.955280 1111.mp3
16 0.0225 0.009574 5 0.946876 1701.mp3
17 0.0225 0.005 4 0.950630 2215.mp3
18 0.0225 0.005 2.702703 0.931842 4312.mp3
19 0.0225 0.008864 2.5 0.951700 7806.mp3
20 0.0225 0.009574 9.090909 0.897224 8203.mp3
File
Info (2)
72%
10:47 AM
4/24/2015

Step-7: Selection of data using "Data" widget for the attributes selected in the previous steps.

Data:
This selects data instances based on conditions over data features.

Signals:
Input: Data (Attribute-valued data set).
Output: Data (Matching Data and Non-matching Data).

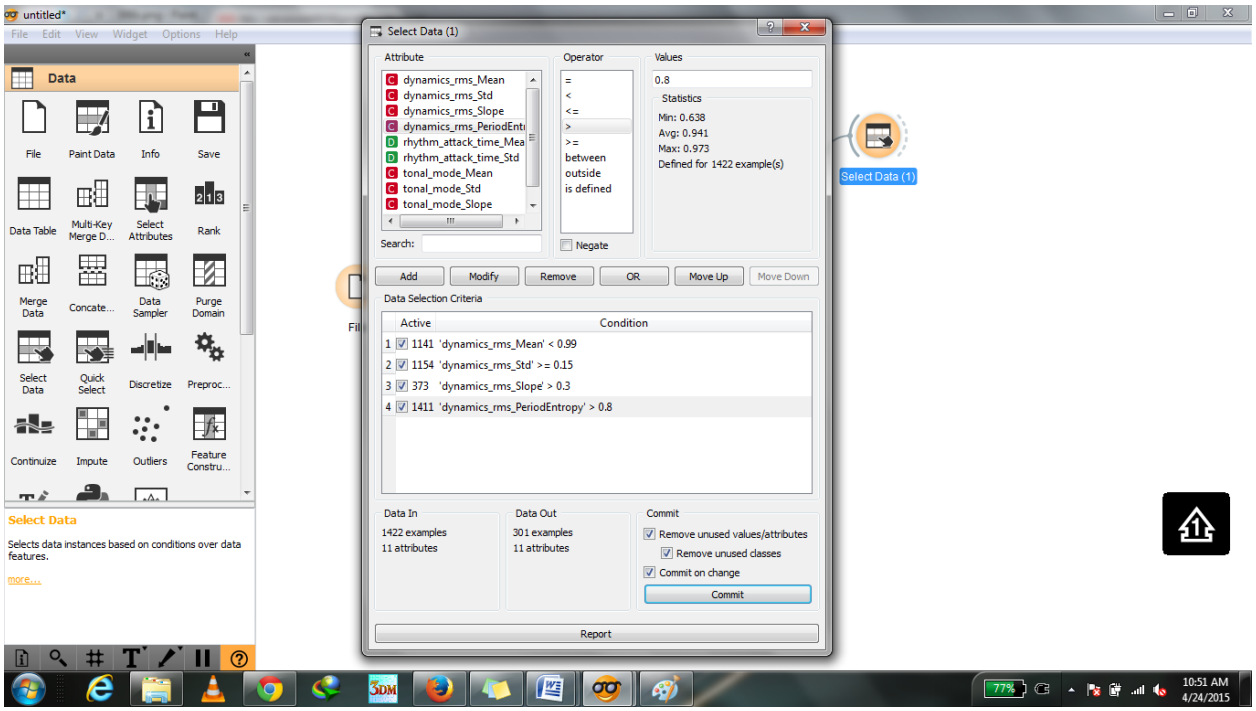
(Matching Data:
These are data instances that match the conditions.
Non-matching Data:
These are data instances that do not match the conditions.)

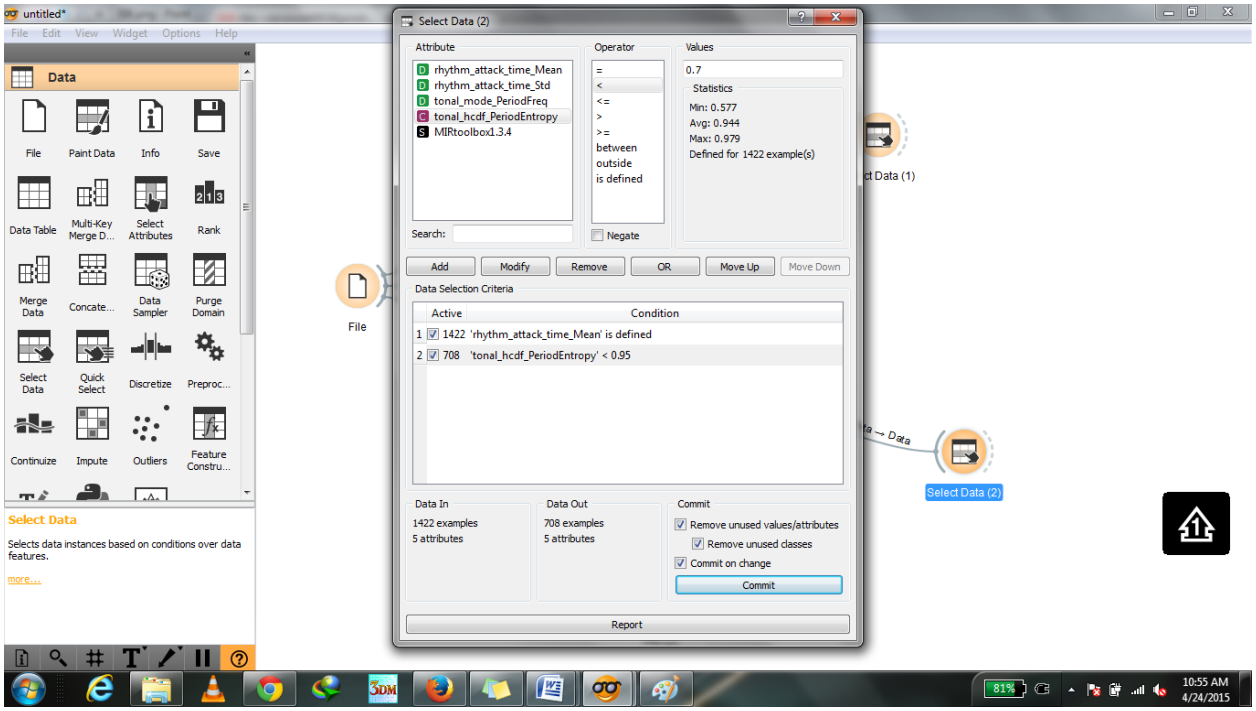
This widget allows the user to select a subset of the data from the input data set based on the condition which are defined over a set of data's attributes.
Data instances that match the selection rule are placed on the Output Matching Data channel

If condition does not include a disjunction (OR line in the condition) then selected items are those matching all terms in the condition. So we can say the criteria for data selection is presented in disjunctive normal form, as collection of conjunction terms (AND operator) with optional disjunction (OR operator).

Selecting an attribute, Selecting an operator from the list of operators define condition terms and these apply to attribute's type and if needed, defining the value to be used in condition term.

Operators are different for discrete, continuous and string attributes.


untitled*
File Edit View Widget Options Help
Data
File
Paint Data
Info
Save
Data Table
Multi-Key Merge D...
Select Attributes
Rank
Merge Data
Concate...
Data Sampler
Purge Domain
Select Data
Quick Select
Discretize
Preproc...
Continuize
Impute
Outliers
Feature Constru...
Select Data
Selects data instances based on conditions over data features.
more...
Select Data (1)
Attribute
dynamics_rms_Mean
dynamics_rms_Std
dynamics_rms_Slope
dynamics_rms_PeriodEntr
rhythm_attack_time_Mea
rhythm_attack_time_Std
tonal_mode_Mean
tonal_mode_Std
tonal_mode_Slope
Search:
Operator
=
<
<=
>
>=
between
outside
is defined
Negate
Values
0.8
Statistics
Min: 0.638
Avg: 0.941
Max: 0.973
Defined for 1422 example(s)
Add
Modify
Remove
OR
Move Up
Move Down
Data Selection Criteria
Active
Condition
1 1141 'dynamics_rms_Mean' < 0.99
2 1154 'dynamics_rms_Std' >= 0.15
3 373 'dynamics_rms_Slope' > 0.3
4 1411 'dynamics_rms_PeriodEntropy' > 0.8
Data In
1422 examples
11 attributes
Data Out
301 examples
11 attributes
Commit
Remove unused values/attributes
Remove unused classes
Commit on change
Commit
Report
Select Data (1)
77%
10:51 AM
4/24/2015

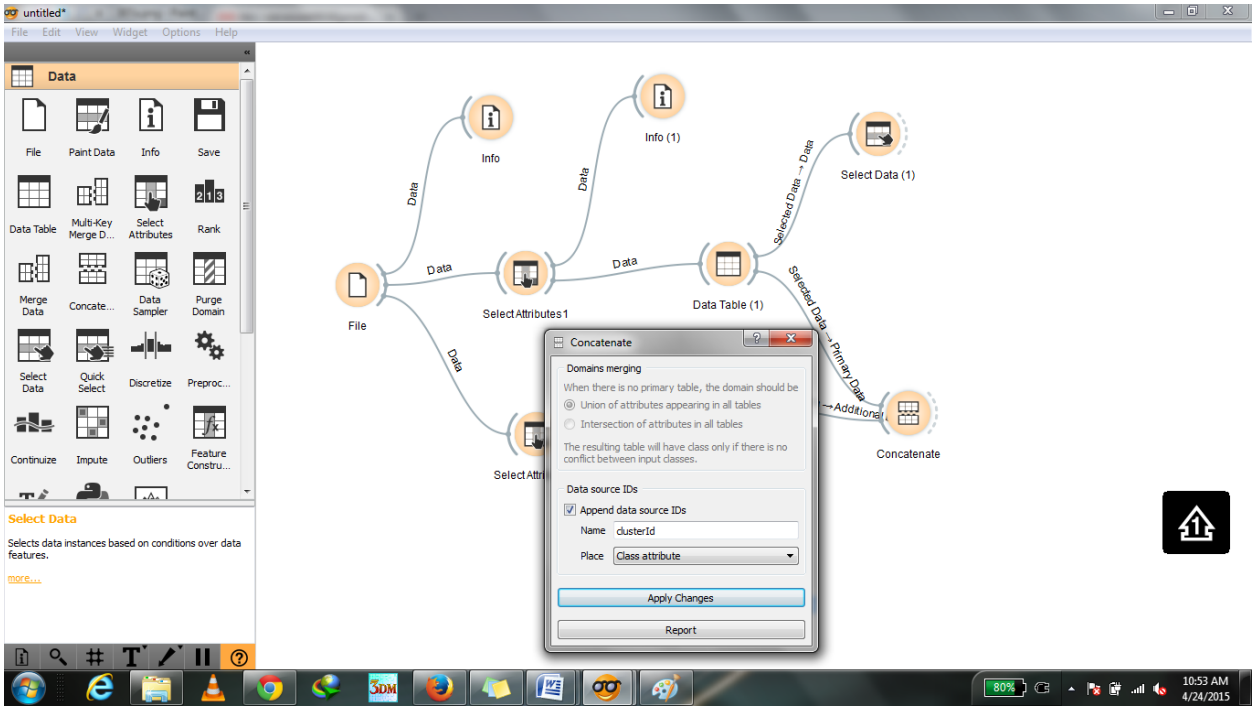

untitled*
File Edit View Widget Options Help
Data
File
Paint Data
Info
Save
Data Table
Multi-Key Merge D...
Select Attributes
Rank
Merge Data
Concate...
Data Sampler
Purge Domain
Select Data
Quick Select
Discretize
Preproc...
Continuize
Impute
Outliers
Feature Constru...
Select Data
Selects data instances based on conditions over data features.
more...
Select Data (2)
Attribute
rhythm_attack_time_Mean
rhythm_attack_time_Std
tonal_mode_PeriodFreq
tonal_hcdf_PeriodEntropy
MIRtoolbox1.3.4
Search:
Operator
=
<
<=
>
>=
between
outside
is defined
Negate
Values
0.7
Statistics
Min: 0.577
Avg: 0.944
Max: 0.979
Defined for 1422 example(s)
Add
Modify
Remove
OR
Move Up
Move Down
Data Selection Criteria
Active
Condition
1 1422 'rhythm_attack_time_Mean' is defined
2 708 'tonal_hcdf_PeriodEntropy' < 0.95
Data In
1422 examples
5 attributes
Data Out
708 examples
5 attributes
Commit
Remove unused values/attributes
Remove unused classes
Commit on change
Commit
Report
File
Select Data (2)
81%
10:55 AM
4/24/2015

Step-8: Concatenating data tables 1 and 2 using the "Concatenate" widget.

Concatenate: Concatenates data from multiple sources.

Signals:
Input: Primary Table (ExampleTable- A data set that defines the attribute set and gives the examples), Additional Tables (ExampleTable).
Output: Examples (ExampleTable).

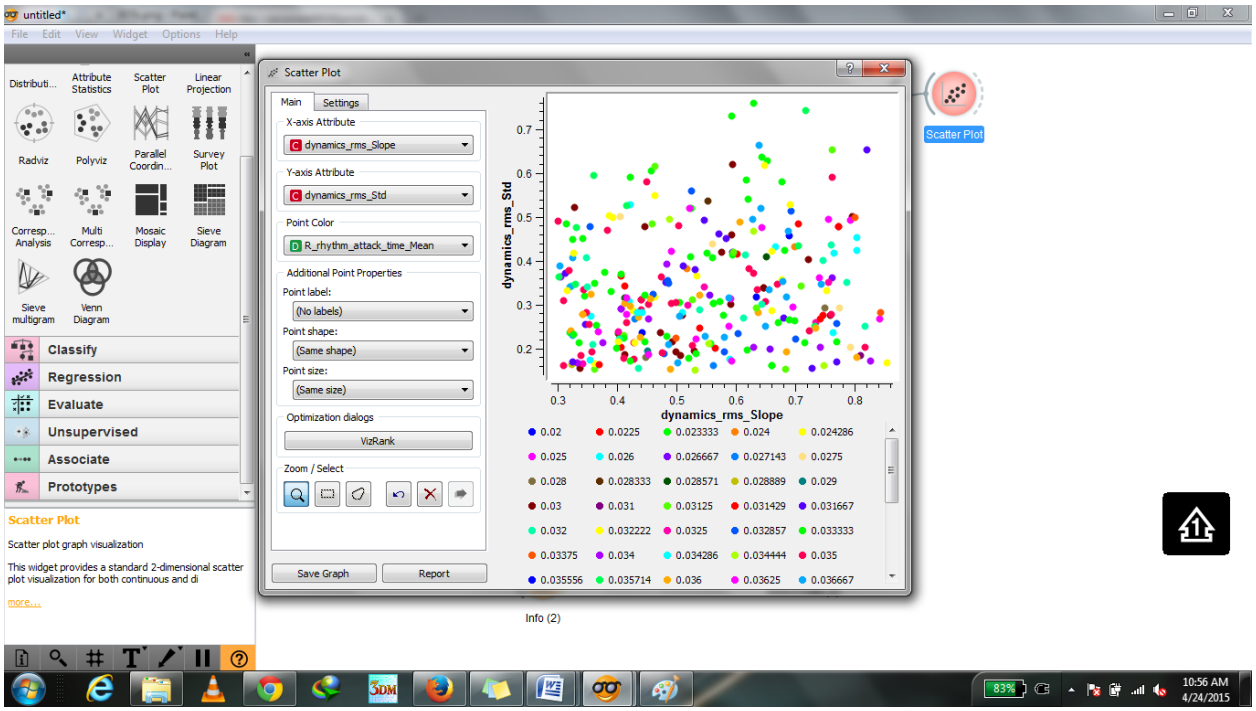
The widget concatenates multiple sets of examples. The merge is "vertical".
In a sense that two sets of 10 and 5 examples yield a new set of 15 examples.



Step-9: Scatter Plot Visualization using the "Scatter Plot" widget.

Scatter Plot:
Using this standard scatter plot visualization with explorative analysis and intelligent data visualization enhancements is done.

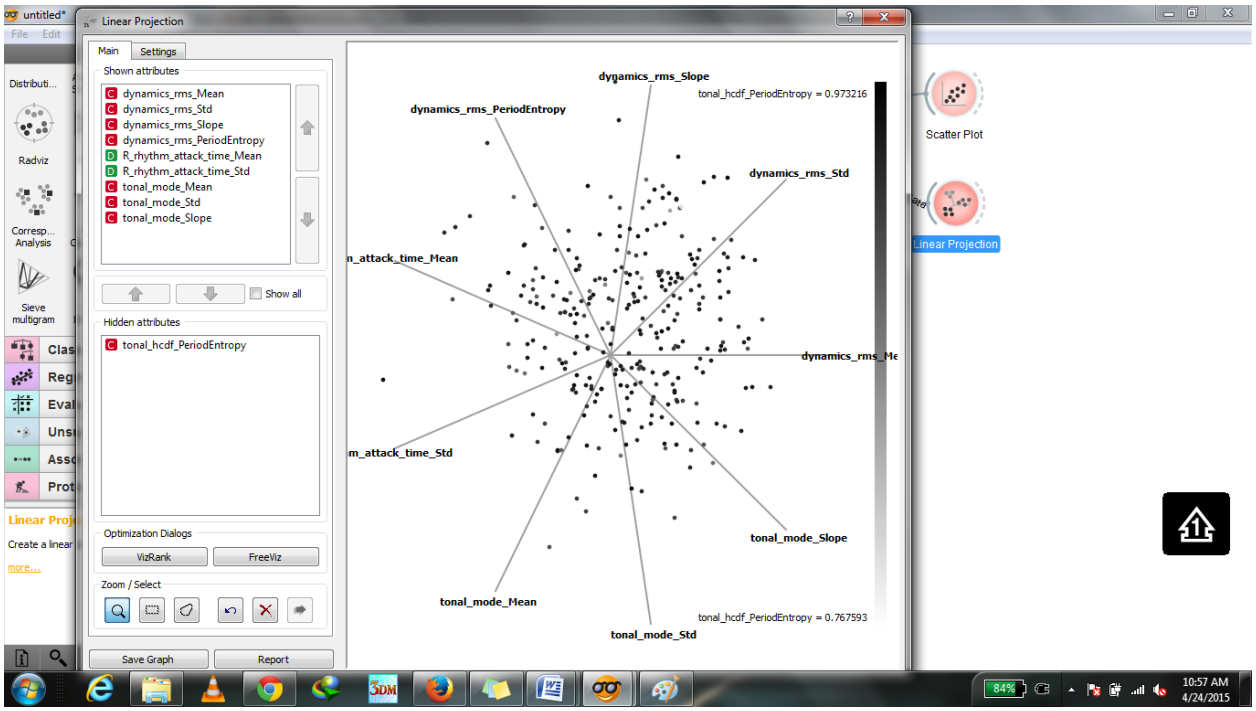
Signals:
Input: Examples (Example Table), Example Subset (A subset of data instances from Examples).
Output: Selected Examples (Example Table) and Unselected Examples (Example Table).



Step-10: Linear Projection Visualization using the "Linear Projection" widget.

Liner Projection:
Various linear projection methods with explorative data analysis and intelligent data visualization enhancements.

Signals:
Input: Examples, Example Subset, and Attribute Selection List.
Output: Selected Examples, Unselected Examples.
The below image shows the list of attributes used in the visualization.



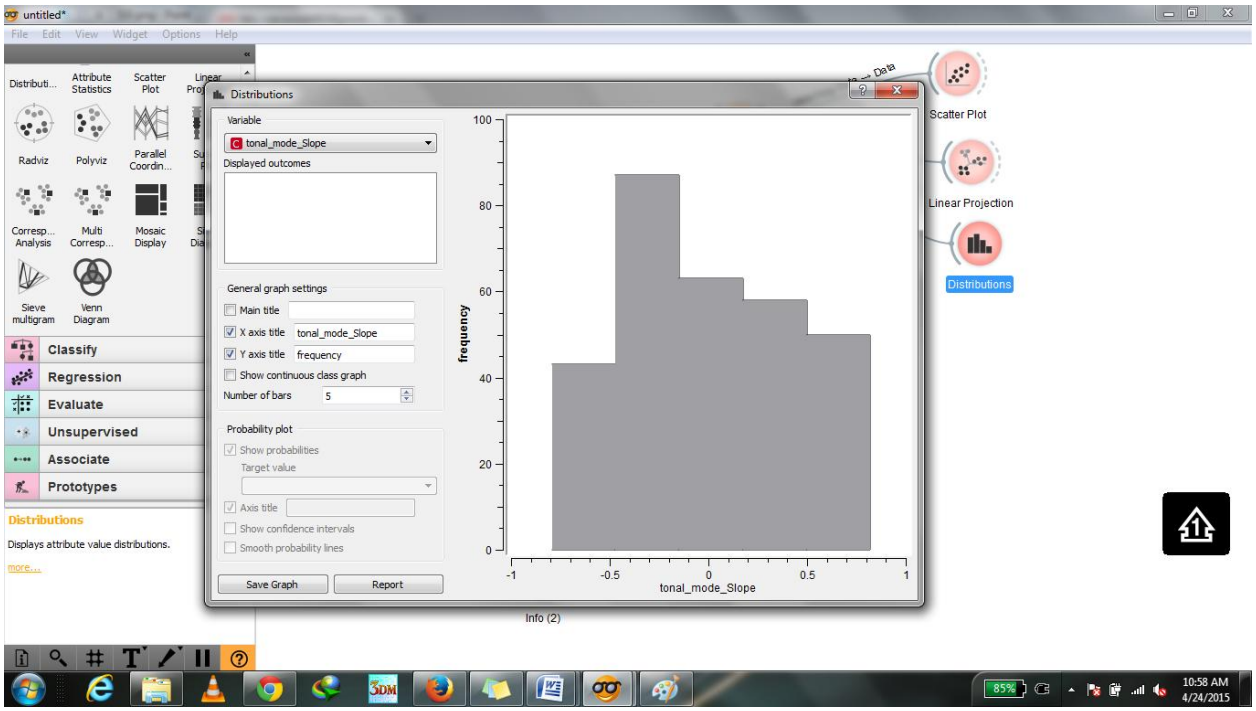
Step-11: Distributions Visualization (Bar Graph) using the "Distributions" widget.

Distributions:
This widget displays value distributions for a single attribute.

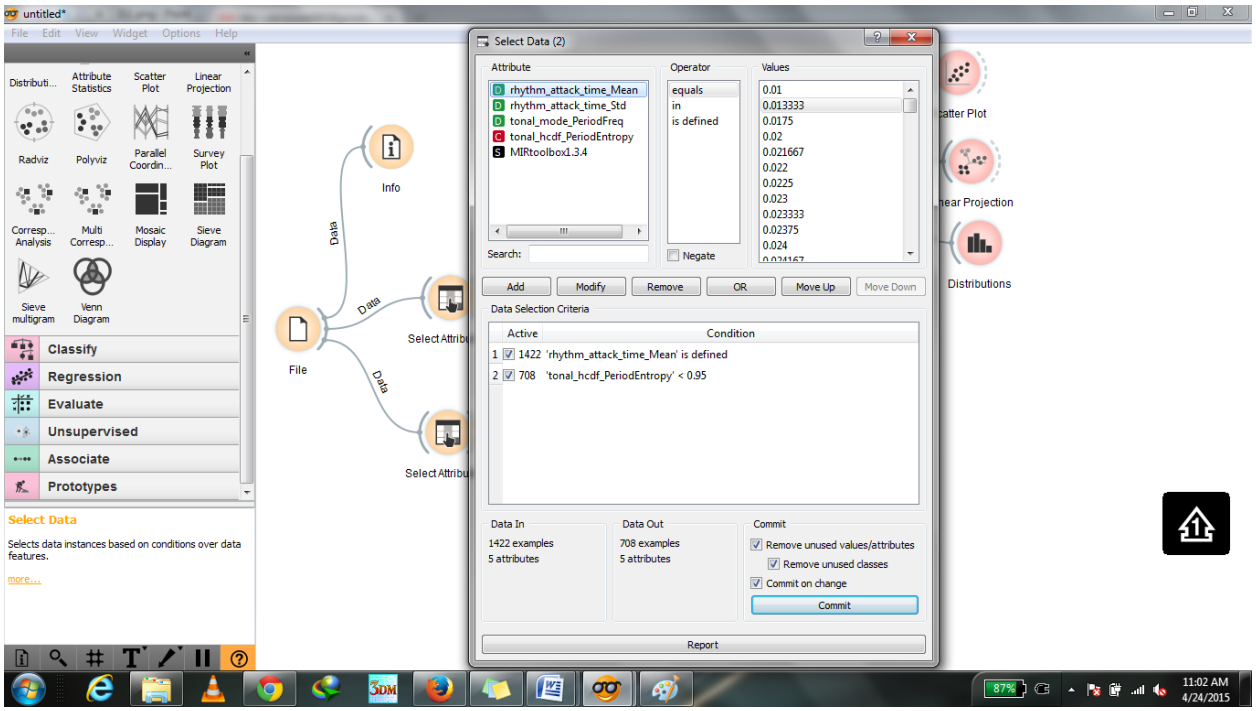
Signals:

Input: Examples (Example Table) - Input data set.
Output: None.

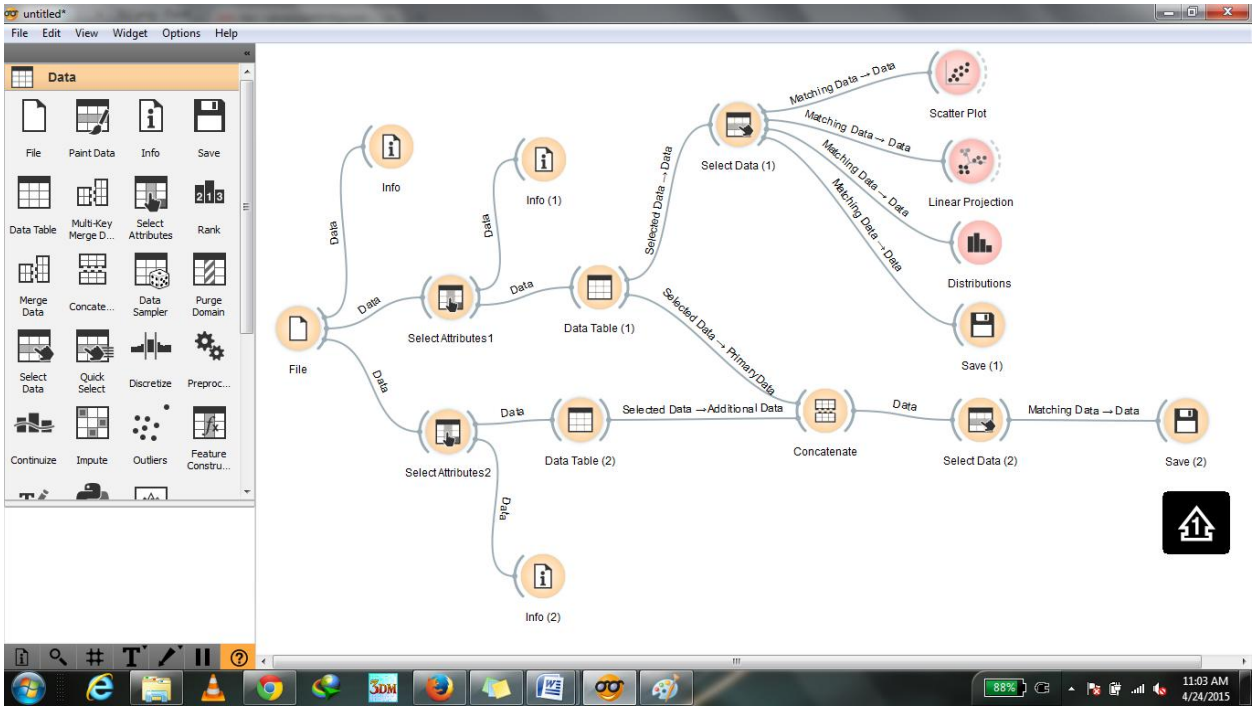
Distributions display the value distribution of either discrete or continuous attributes. Distributions are conditioned on the class. That is in case of data consisting of class.



Step-12: Obtained the data after concatenation.

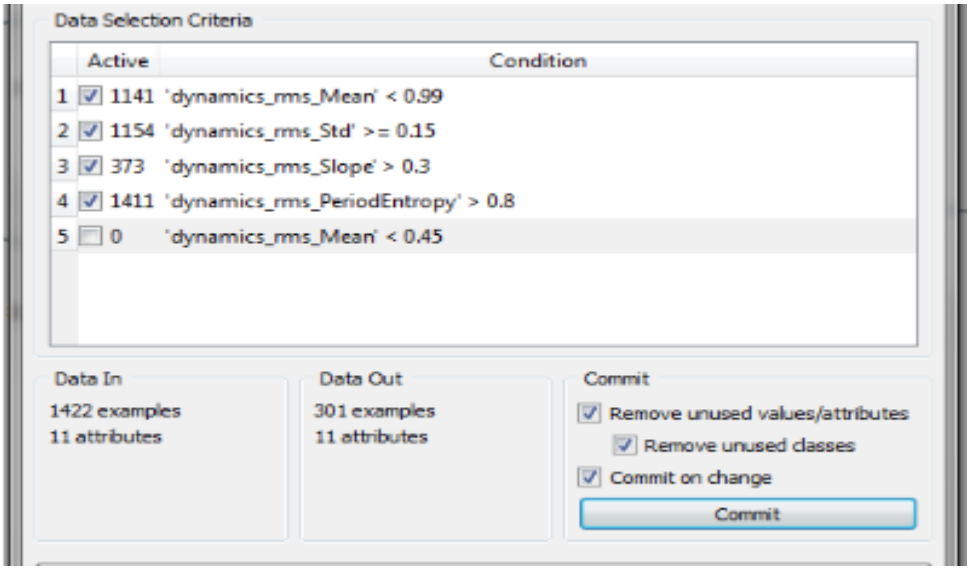


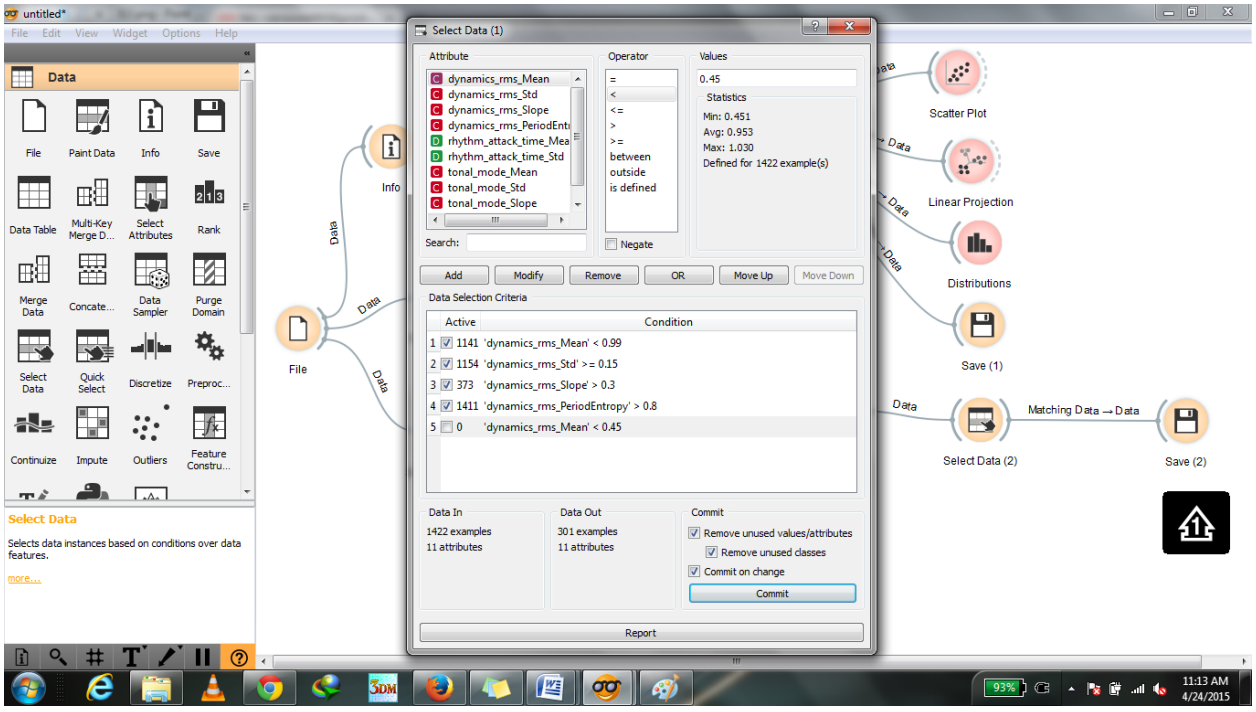
Step-13: Overview of the above steps. The scheme of all the above steps are merged and displayed in the below screenshot.



Failed Query Demonstration:

Consider the successful query, the conditions included in the Data selection criteria are shown below. (Conditions can be selected or unselected). The condition “dynamics_rms_PeriodEntropy > 0.8” is considered and “dynamics_rms_Mean < 0.45” is excluded here. We can observe the “Data Out” column section that consists of 301 examples and 11 attributes which implies that there is no failing query.



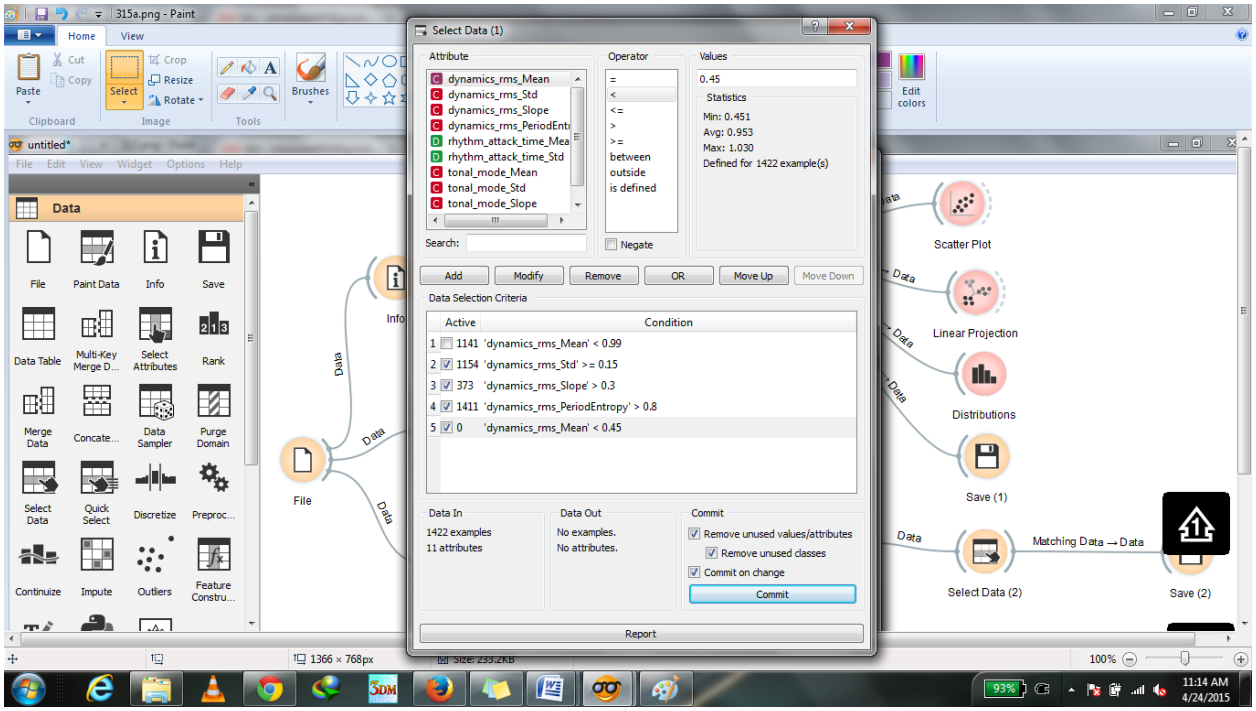


Any change in composition of the condition will trigger updates in information panel that displays the number of data instances being selected (**Data Out**).

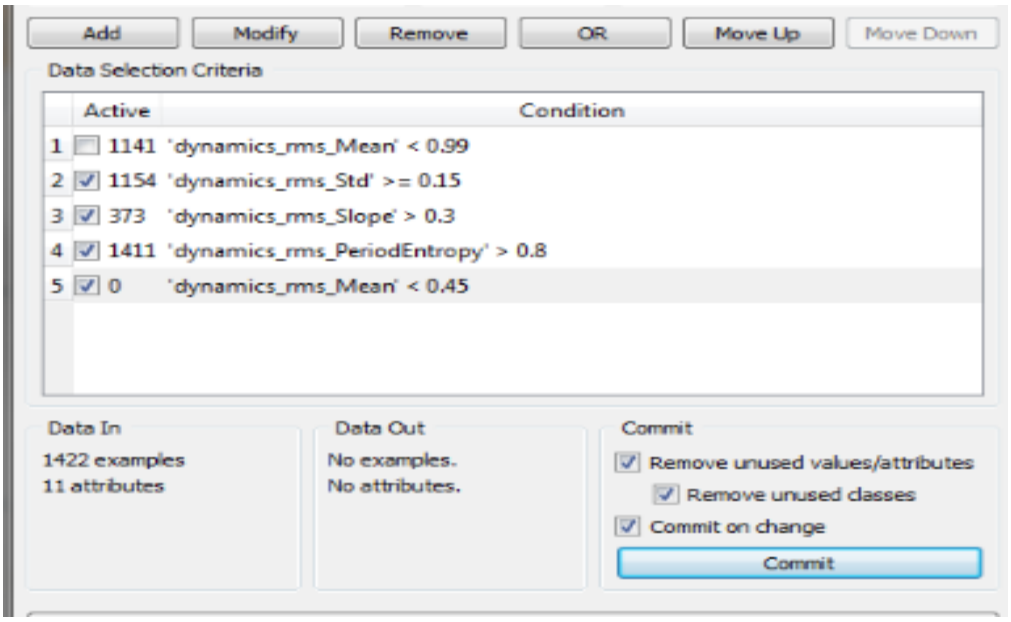
Number of instances that match specific condition terms are also displayed at the start of the condition term line.

Data set composed of instances that match the defined condition is placed on the output channel.

The below image shows the failed query.



The attributes that are included and excluded are shown in the following screenshot.



The condition “dynamics_rms_Mean < 0.45” is considered and “dynamics_rms_Slope <0.99” is excluded here.

Under the “Data Out” section, we can observe that we have not obtained any examples or attributes and this implies a “failed query”.
In this way it displays the failing queries that do not satisfy the tuples.