

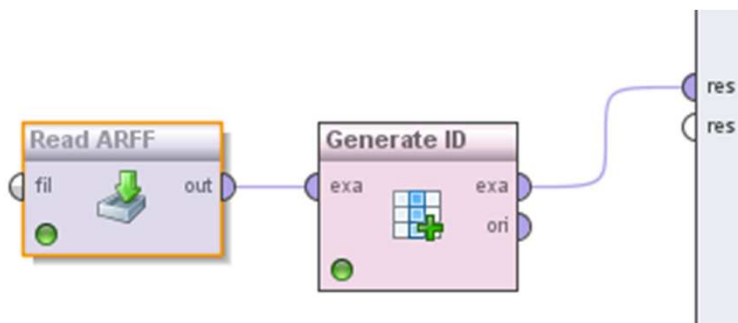


Wstępne przetwarzanie danych

Rapid miner

GenerateID

- Data Transformation\Attribute Set Reduction and transformation\Generation



☐ create nominal ids

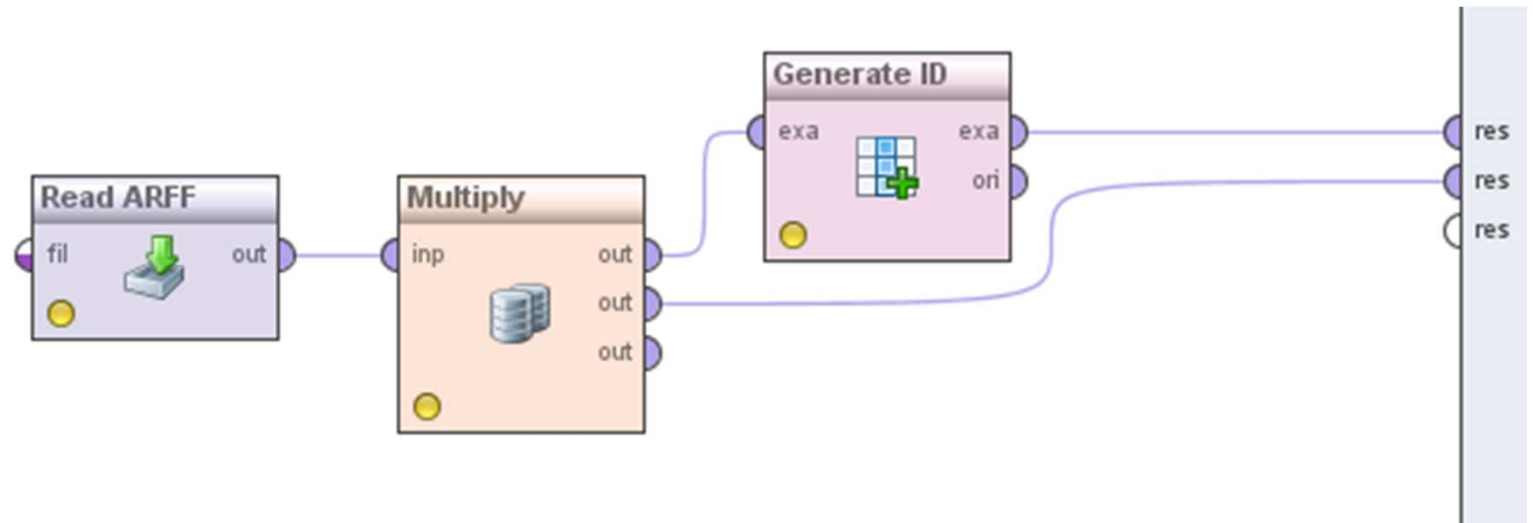
offset

Czy ID ma być tworzony jako atrybut nominalny?

Wartość dodawana do każdej liczby początkowej

Multiply

- Process Control



Generate Attribute

- Data Transformation\Attribute Set Reduction and transformation\Generation

The image displays the Orange3 data mining software interface. At the top, a workflow is shown with a 'Read ARFF' widget connected to a 'Generate Attribute' widget. The 'Generate Attribute' widget has two output ports labeled 'res' and 'ori'. Below the workflow, the 'Generate Attributes' dialog box is open, showing options for 'function descriptions', 'use standard constants', and 'keep all'. The 'keep all' option is checked. Below this, the 'Edit Parameter List: function descriptions' dialog box is open, showing a table with two columns: 'attribute name' and 'function expressions'. The 'attribute name' column is highlighted with a red box and an arrow pointing to it from the text 'Nazwa nowego atrybutu'. The 'function expressions' column is highlighted with a red box and an arrow pointing to it from the text 'Wyrażenie matematyczne'. Another red box with the text 'Zachowaj niemodyfikowane atrybuty' (Keep unmodified attributes) has an arrow pointing to the 'keep all' checkbox in the 'Generate Attributes' dialog.

Read ARFF

Generate Attri...

function descriptions

☒ use standard constants

☒ keep all

Edit List (0)...

Edit Parameter List: function descriptions

Edit Parameter List: **function descriptions**
List of functions to generate.

attribute name	function expressions

Nazwa nowego atrybutu

Wyrażenie matematyczne

Zachowaj niemodyfikowane atrybuty

Add Entry

Add Noise

- Utility\Data Generation

The diagram illustrates the 'Add Noise' utility workflow and its configuration options. On the left, a 'Read ARFF' node is connected to an 'Add Noise' node, which is then connected to a 'res' node. The 'Add Noise' node has ports for 'exa', 'ori', and 'pre'. On the right, the 'Add Noise' configuration panel is shown with various settings.

Configuration panel settings:

- attribute filter type: single
- attribute: age
- ☐ invert selection
- ☐ include special attributes
- random attributes: 0
- label noise: 0.05
- default attribute noise: 10.0
- noise: Edit List (0)...
- offset: 0.0
- linear factor: 1.0
- ☐ use local

Annotations in red boxes with arrows pointing to the configuration panel:

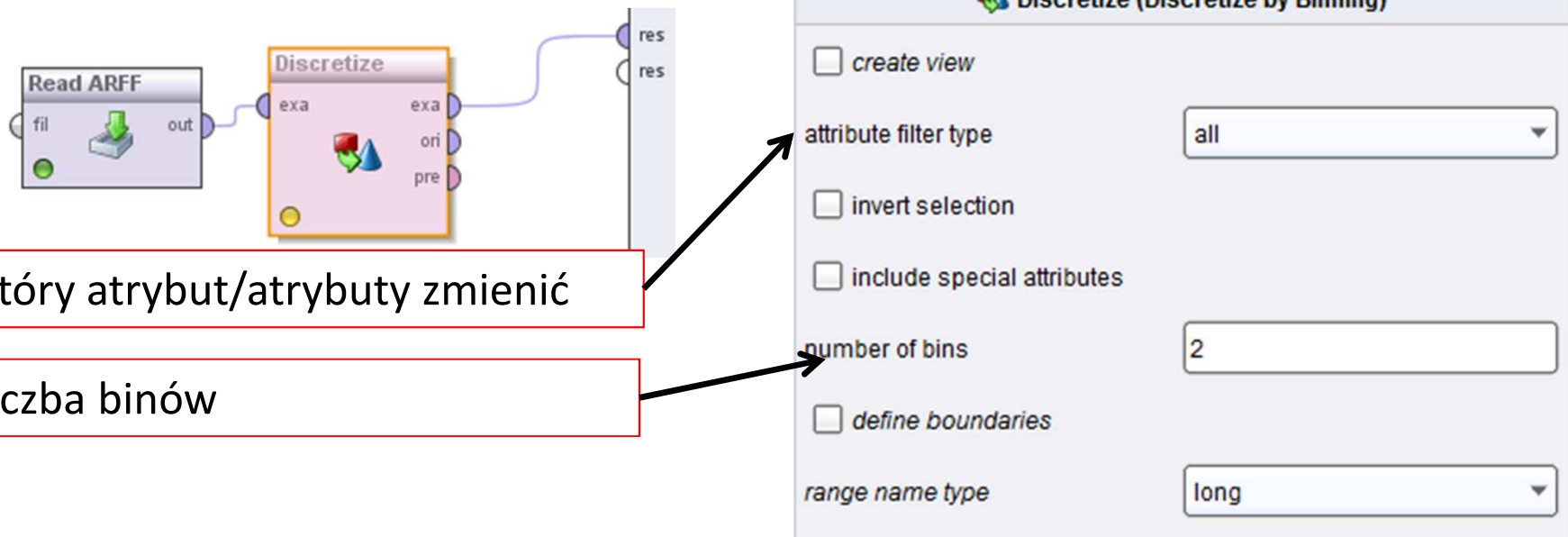
- Który atrybut zmienić (Which attribute to change) - points to the 'attribute' dropdown.
- Ile nowych atrybutów z szumem stworzyć? (How many new attributes with noise to create?) - points to the 'random attributes' input.
- Domyślna „Siła” szumu (Default „Strength” of noise) - points to the 'default attribute noise' input.
- Można tutaj definiować siłę szumu dla wielu atrybutów równocześnie. (You can define the strength of noise for multiple attributes simultaneously.) - points to the 'noise' list editor.
- Liniowa transformacja każdego losowego atrybutu (Linear transformation of each random attribute) - points to the 'offset' and 'linear factor' inputs.

„Siła” szumu

- Dla atrybutów nominalnych
 - Prawdopodobieństwo zmiany wartości na inną
- Dla atrybutów liczbowych
 - Odchylenie standardowe rozkładu normalnego

Discretize by binning (equalWidth)

- Data Transformation \ Type conversion \ Discretization



Discretize by frequency (equal frequency)

- Data Transformation \ Type conversion \ Discretization

The diagram illustrates the process of discretizing data using the Weka software. It starts with a 'Read ARFF' node connected to a 'Discretize' node. The 'Discretize' node has an 'exa' input and 'exa', 'ori', and 'pre' outputs. A 'res' node is also shown. A red box highlights the 'Discretize' node, with an arrow pointing to its configuration window. The configuration window, titled 'Discretize (Discretize by Frequency)', shows the following settings:

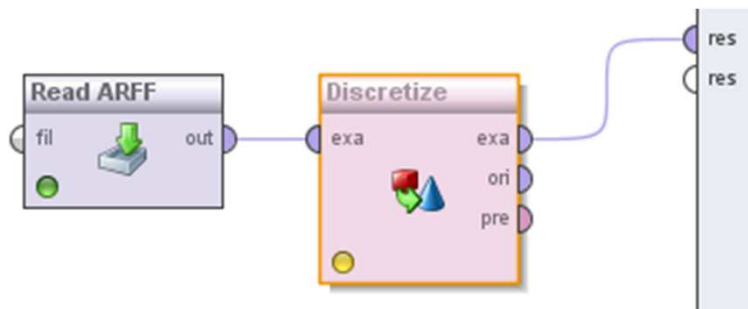
- ☐ create view
- attribute filter type: all
- ☐ invert selection
- ☐ include special attributes
- ☐ use sqrt of examples
- number of bins: 2
- range name type: long

Two red boxes with arrows point to the configuration window:

- Który atrybut/atomy zmieni
- Liczba binów

Discretize by size (equal frequency)

- Data Transformation \ Type conversion \ Discretization



Który atrybut/atomy zmieni

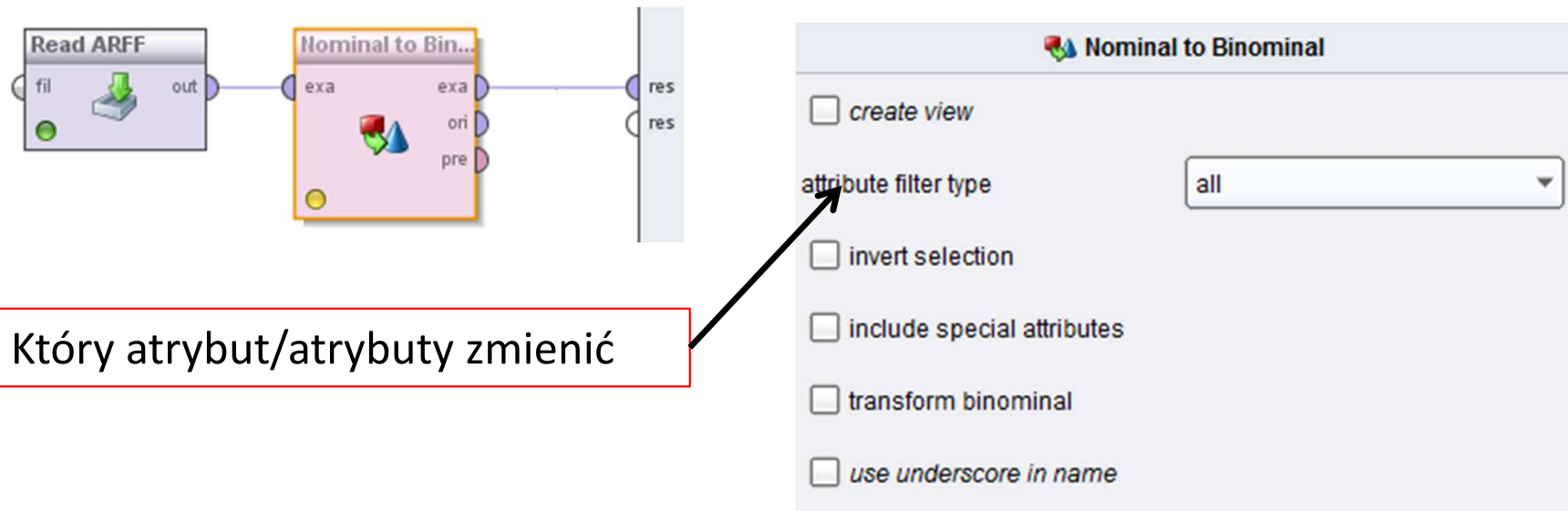
Rozmiar binu

The screenshot shows the 'Discretize (Discretize by Size)' dialog box. It contains the following options and settings:

- ☐ create view
- attribute filter type: all (dropdown)
- ☐ invert selection
- ☐ include special attributes
- size of bins: (empty text box)
- sorting direction: decreasing (dropdown)
- range name type: long (dropdown)

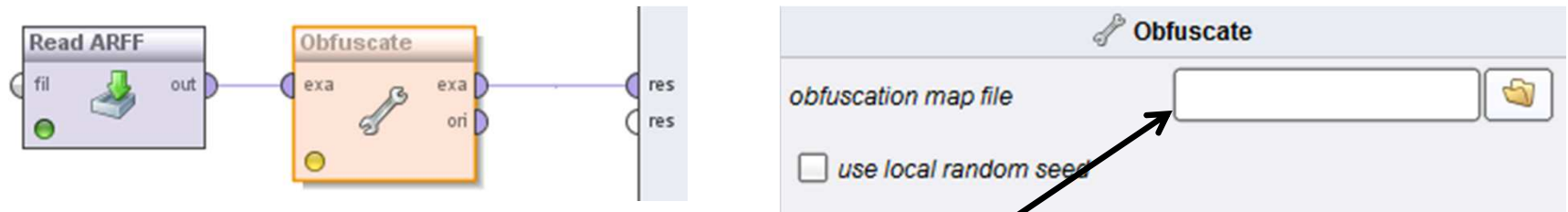
Nominal to Binomial

- Data transformation \ Type conversion \



Obfuscate

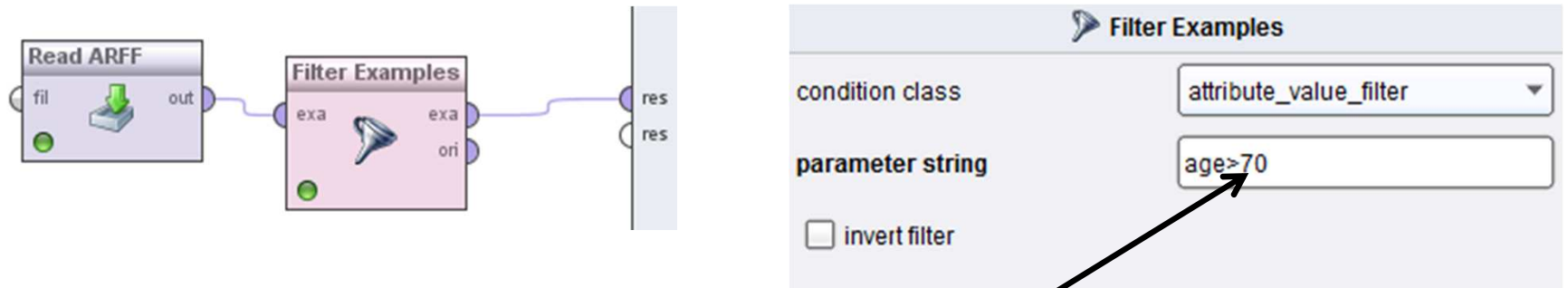
- Utility \ Miscellaneous



Gdzie zapisać plik z wykonanymi podmianami

Filter Examples

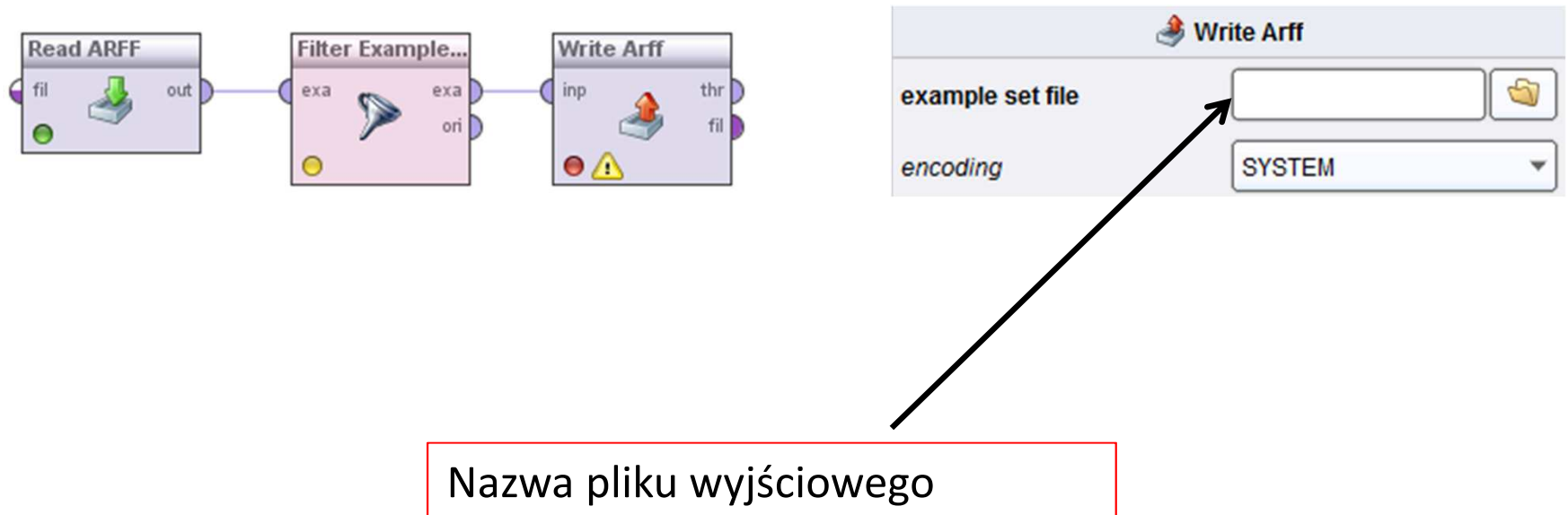
- Data transformation \ Filtering



Tylko wiersze spełniające ten warunek pozostają

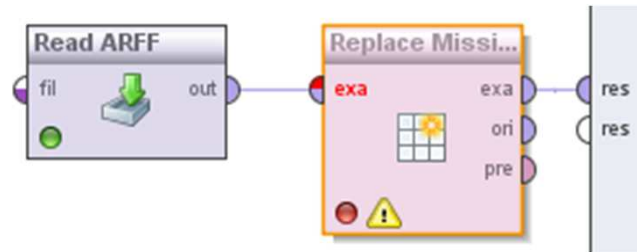
Write ARFF

- Export\Data



Replace Missing Values

- Data Transformation\Data Cleansing



The 'Replace Missing Values' configuration window shows the following settings:

- ☐ create view
- attribute filter type: single
- attribute: age
- ☐ invert selection
- ☐ include special attributes
- default: average
- columns: Edit List (0)...

Który atrybut

Jak wyliczyć brakującą wartość

Ćwiczenie

- Przygotuj proces, który wczytuje plik heart_c.arff a następnie zapisuje 3 pliki z danymi dotyczącymi kolejno: pacjentów młodszych niż 50 lat, pacjentów 51-60 i pacjentów ponad 60 lat. Dane powinny być „Zaciemnione”.
- Wczytaj plik ARFF a następnie zamień atrybut AGE na atrybut nominalny. Użytkany w ten sposób atrybut nominalny przekonwertuj na atrybut binarny(binominalny). Wyświetl: oryginalne dane, dane po 1 etapie i dane po 2 etapie transformacji.