Normalization of Continuous Data

Let us assume that it is required to cast values to [a, b] range. The following formula is used for this purpose:

$x^* = \frac{x - x_{min}}{x_{max} - x_{min}} * (b - a) + a$,

where x is a current value, x_{min}, x_{max} are minimum and maximum field values. This is how the linear normalizer is used.

The following data normalization options are available in Megaladata:

- **No**: there is no normalization. In this case, the data is sent to the main algorithm without preprocessing.
- **Scaling [min; max]**: the data reduction to [min; max] range set by a user by means of the linear transformation:
  - **Minimum**: the minimum value.
  - **Maximum**: the maximum value.
- **Scaling [-1;1]**: the data reduction to [-1;1] range by means of the linear transformation.
- **Scaling [0;1]**: the data reduction to [0;1] range by means of the linear transformation.
- **Absolute scaling**: each value is divided by the maximum absolute value.
- **Standardization**: the average value is subtracted from each value, and then it is divided by the standard deviation.
- **Quotient**: each value is divided by the statistical indicator or the value set by a user.
  - Divisor:
    - Statistical indicator (Minimum, Maximum, Average, Sum, Standard deviation).
    - Set value.

Range Control

Control of values range does not let the value go out of the specified range saved while training when executing the node.

- Control method:
  - **No**: there is no control method.
  - **Error**: if the value that goes out of the given range is detected, the following error occurs when the node is started: *Normalizing value exceeds the values range maximum*.
  - **Winsorization**: if the value that goes out of the given range is detected, it is reduced to the maximum/minimum value.

- **Manual setting**: when this checkbox is selected, it becomes possible to select the range of values.
- **Source Range**: the values of the input field minimum and maximum. It is set when retraining the node.
  - Minimum.
  - Maximum.
- **Controlled Range**: the manual setting of values range.
  - Minimum.
  - Maximum.