Binary Classification Assessment

Binary classification assessment: charts and tables with results of the performed classification based on the logistic regression are constructed in this visualizer.

The visualizer is divided into 3 areas:

- Area of settings is designated for the chart selection and configuration.
- Chart area displays the chart selected in the area of settings.
- Area of classification assessments contains the tables with classification assessments.

> **Note**: Sizes of *Area of classification assessments* can be changed by dragging the separator line with the left mouse button. The area will be hidden by double clicking on the separator line or single clicking on the middle part of the line. It is possible to hide the area also for *Area of settings*.

Area of Settings

It is located in the left part of the visualizer and contains three groups of parameters: *Chart type*, *Sets* and *Cutoff*.

Chart Type

The group contains nine switches each of which calls the definite chart in the visualizer center:

- **ROC curve**: dependence of *TPR* on *FPR*.
- **PR curve**: dependence of *TPR* on *PPV*.
- **Basic rates**: the following graphic charts are constructed in the chart: *TPR*, *TNR*, *FPR*, *FNR* and cutoff is displayed.
- **Precision chart**: the following graphic charts are constructed in the chart: *PPV*, *NPV*, *OPR*, *OCR* and cutoff is displayed.
- **Break even chart**: the following graphic charts are constructed in the chart: *PPV*, *TPR* and cutoff is displayed.
- **% captured response**: according to the selected *Cumulative* checkbox shows as follows:
  - ☑: cumulative % of events of the total number of events according to the sample size.
  - ☐: % of events that are in the bin of the total number of events according to the bin number.
- **Lift chart**: according to the selected *Cumulative* checkbox shows as follows:
  - ☑: cumulative Lift value according to the sample size.

    - ☐: Lift values for the bin according to the bin number.
- **Response chart**: according to the selected *Cumulative* checkbox shows as follows:
    - ☑: cumulative % of events in the sample according to the sample size.
    - ☐: % of events in the bin according to the bin number.
- **Gain chart**: Gain value according to the sample size.

> **Note**: Meanings of abbreviations are available on Terms page.

Cumulative checkbox becomes active when selecting the following charts: *% captured response*, *Lift chart* and *Response chart*. The checkbox is selected by default. When deselecting the checkbox, it becomes possible to select the bins in the drop-down list. Available values:

- **10 bins**: divide a set into 10 equal parts. This set of bins is used by default.
- **20 bins**: divide a set into 20 equal parts.
- **50 bins**: divide a set into 50 equal parts.

Sets

It contains two checkboxes:

- **Train**: when the checkbox is selected, it enables to display the graphic chart of the training set on the chart.
- **Test**: when the checkbox is selected, it enables to display the graphic chart of the test set on the chart.

It is possible to select only one set for the following charts: *Basic rates*, *Precision chart* and *Break even chart*.

Cutoff

It is a drop-down list with the following values:

- **From node settings**: the cutoff set by the *Logistic regression* node is used.
- **Set** enables to set the own cutoff. The cutoff is entered in the *Value* field, or it is set by moving the slider under the field.
- **Balance (TPR = TNR)** sets the cutoff with which *TPR* and *TNR* are equal.
- **Maximum (TPR + TNR)** sets the cutoff with which the sum of *TPR* and *TNR* has the maximum value.
- **Break even point (TPR = PPV)** sets the cutoff with which *TPR* and *PPV* are equal.
- **Highest overall accuracy** sets the cutoff with which *OCR* value is maximum.
- **Maximum F1 Score** sets the cutoff with which F_1 Score value is maximum.
- **Matthews coefficient (MCC)** sets the cutoff with which MCC value is maximum.

Chart Area

The area contains the following elements:

- **Header** shows the name of the displayed chart.
- **Event** displays the output field caption and the value that is an event.
- **Chart** places the chart into the center. It is always located in the area center and has the same height and width.
- **Legend** contains names of the series displayed on the chart. When pressing the series in the legend, it is possible to hide or display it on the chart.

> **Note**: Depending upon the free area space the legend will be always located under the chart or to the right from it.

Areas of Classification Assessments

It is located in the right part of the screen and contains the tables discribing classification results. Data in these tables is updated with *Cutoff* change.

Classification Scores

The table contains the columns:

- **Value** displays the name of the calculated item.
- **Sets**: the group from two columns:
  - **Train**: the assessment values of the indicators calculated for the training set are shown in the column.
  - **Test**: the assessment values of the indicators calculated for the test set are shown in the column.

The table is divided into two areas and includes the following rows:

- **Classification scores**: this area contains the scores that are calculated for the whole model and do not depend on the *Cutoff*:
  - **AUC ROC** shows the **area under the ROC curve**.
  - **AUC PR** shows the area under the PR curve. It is defined similar to ROC curve, and Precision and Recall but not FPR and TPR are set on the axes.
  - **Gini index**: **Gini index**.
  - **KS**: **Kolmogorov-Smirnov statistics**.
- **Cutoff:** the scores depending on the *Cutoff* are located in this part. The cutoff name is shown after the colon, for example *Cutoff: Set*:
  - **Value** shows the used *Cutoff* value.
  - **TPR (Sensitivity)** shows *TPR* value for the used cutoff.
  - **TNR (Specificity)** shows *TNR* value for the used cutoff.

- **FPR (1-Specificity)** shows *FPR* value for the used cutoff.
- **PPV** shows *PPV* value for the used cutoff.
- **F1 Score** shows *F1 Score* value for the used cutoff.
- **MCC** shows the Matthews correlation coefficient value for the used cutoff.

> **Note**: If there is no data set, its matching table cells will be grey and they will show nothing.

Confusion Matrices

It contains the table with error matrices for the training and test sets, and it is as follows:

Predicted	**Actually**		**Total**
	Event	**Non-event**	
Set	P	N	
Event	TP	FP	TP+FP
Non-event	FN	TN	FN+TN

> **Note**: The 1 %/1 % *Absolute or relative values* switch is located over the upper right table corner. When it is switched, the table data will be displayed in percentage terms or as number of the records included into the category.

Recognized

This table shows the ratio of correctly captured response to the total number of events in sets. The table consists of two fields:

- **Set**: the set type is displayed in this column.
- **Recognized**: the number of captured response from the total number of events. It is defined using the following formula: $\frac{TP+TN}{TP+TN+FP+FN}$

> **Note**: The 1 %/1 % *Absolute or relative values* switch is located over the upper right table corner. When it is switched, the table data will be displayed in percentage terms or as number of the records included into the category.

Articles in Section:

- Terms

- Error Matrix