



Heart Disease

The goal of this study is to train machine learning models in order to predict whether a patient is likely to suffer from heart disease and to evaluate and compare the performance of different machine learning algorithms, specifically k-Nearest Neighbors (kNN), Logistic Regression, Random Forest, XGBoost, MLP and RBF.

The dataset, which can be found at <https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset>, contains 1025 observations and 14 variables in total, including 13 predictor variables and 1 target variable. The target variable, target, indicates the presence or absence of heart disease, where 0 denotes no heart disease and 1 denotes heart disease. The predictor variables include several demographic, clinical, and diagnostic factors that are commonly associated with cardiovascular risk, such as age, sex, chest pain type, resting blood pressure, cholesterol level, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression, slope of the peak exercise ST segment, number of major vessels, and thalassemia status. Since the dataset contains structured numerical and categorical-encoded medical information, it is well suited for the development and comparison of classification-based machine learning models for heart disease prediction.

Isalos version used: 2.0.2

Step 1: Import data from file

Right click on the input spreadsheet (left) and choose the option “Import from File”. Then navigate through your files to load the one with the heart disease data.

A screenshot of a spreadsheet application interface. The spreadsheet has 8 columns labeled Col1 through Col8 and 16 rows. The first row is a header row with "User Header" in Col1 and "User Row ID" in Col2. The following rows are numbered 1 through 16. A context menu is open over the cell at row 6, column 4. The menu items are: Show Spreadsheet Toolbar, Import from File, Import from Spreadsheet, Import from Multiple Spreadsheets, Adjust Spreadsheet Precision, Export Spreadsheet Data, and Clear Spreadsheet.

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
User Header	User Row ID							
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								

The data will appear on the left spreadsheet.

	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (I)	Col8 (I)	Col9 (I)	Col10 (I)	Col11 (D)	Col12 (I)	Col13 (I)	Col14 (I)	Col15 (I)
User Header	User Row ID	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
1		52	1	0	125	212	0	1	168	0	1	2	2	3	0
2		53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
3		70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
4		61	1	0	148	203	0	1	161	0	0	2	1	3	0
5		62	0	0	138	294	1	1	106	0	1.9	1	3	2	0
6		58	0	0	100	248	0	0	122	0	1	1	0	2	1
7		58	1	0	114	318	0	2	140	0	4.4	0	3	1	0
8		55	1	0	160	289	0	0	145	1	0.8	1	1	3	0
9		46	1	0	120	249	0	0	144	0	0.8	2	0	3	0
10		54	1	0	122	286	0	0	116	1	3.2	1	2	2	0
11		71	0	0	112	149	0	1	125	0	1.6	1	0	2	1
12		43	0	0	132	341	1	0	136	1	3	1	0	3	0
13		34	0	1	118	210	0	1	192	0	0.7	2	0	2	1
14		51	1	0	140	298	0	1	122	1	4.2	1	3	3	0
15		52	1	0	128	204	1	1	156	1	1	1	0	0	0
16		34	0	1	118	210	0	1	192	0	0.7	2	0	2	1
17		51	0	2	140	308	0	0	142	0	1.5	2	1	2	1
18		54	1	0	124	266	0	0	109	1	2.2	1	1	3	0
19		50	0	1	120	244	0	1	162	0	1.1	2	0	2	1
20		58	1	2	140	211	1	0	165	0	0	2	0	2	1
21		60	1	2	140	185	0	0	155	0	3	1	0	2	0
22		67	0	0	106	223	0	1	142	0	0.3	2	2	2	1
23		45	1	0	104	208	0	0	148	1	3	1	0	2	1

Step 2: Manipulate data

In this dataset there are not any empty values or categorical features, so we can select all the columns to be used. On the menu click on *Data Transformation* → *Data Manipulation* → *Select Column(s) and select all columns*.

The screenshot shows the software's menu system and a dialog box. On the left, the 'Data Manipulation' menu is open, listing options like 'Remove Column(s)', 'Select Column(s)', 'Matrix Transpose', 'Wide to Long Format', 'Sort by Column', and 'Fill Missing Column(s) Values'. On the right, the 'Select Column(s)' dialog box is displayed. It has two panes: 'Excluded Columns' (empty) and 'Included Columns' (containing a list of columns from 'Col2 -- age' to 'Col10 -- exang'). Navigation arrows (>>, >, <, <<) are between the panes. 'Execute' and 'Cancel' buttons are at the bottom.

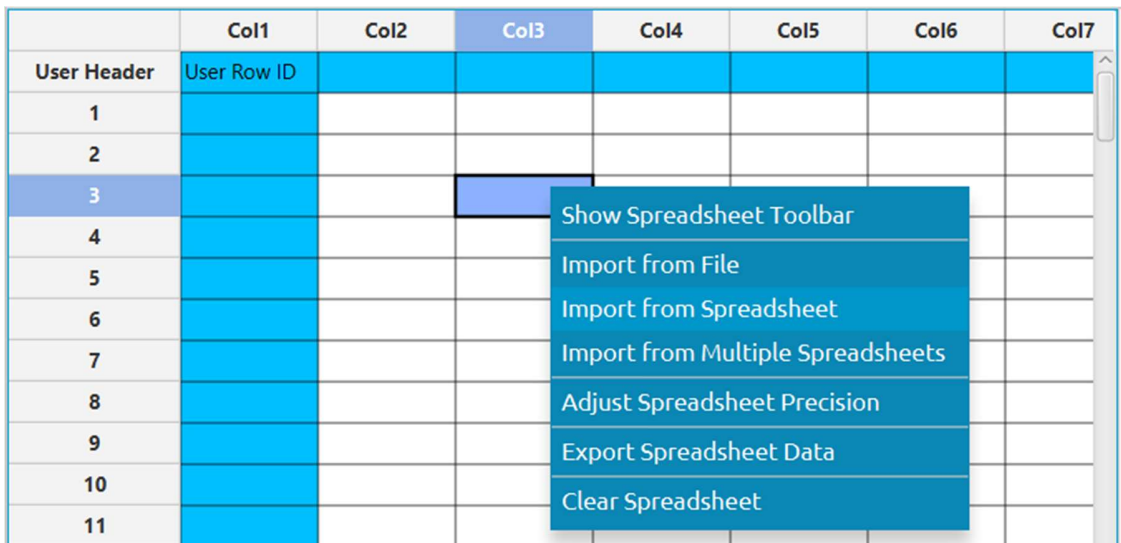
All the data will appear in the output (right) spreadsheet. This tab can be renamed “IMPORT” by right-clicking on it and choosing the “Rename” option.



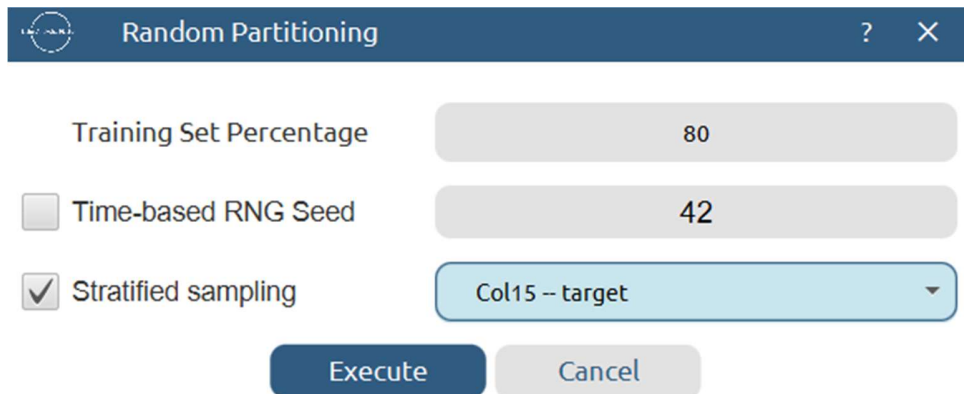
Step 3: Split data

Create a new tab by pressing the “+” button on the bottom of the page with the name “TRAIN_TEST_SPLIT” which we will use for splitting the train and test set.

Import data into the input spreadsheet of the “TRAIN_TEST_SPLIT” tab from the output of the “IMPORT” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.



Split the dataset by choosing *Data Transformation* → *Split* → *Random Partitioning*. Then choose the “Training set percentage” and the column for the sampling as shown below:



The results will be two separate spreadsheets, “TRAIN_TEST_SPLIT: Training Set” and “TRAIN_TEST_SPLIT: Test Set”, which will be available to import into the next tabs.

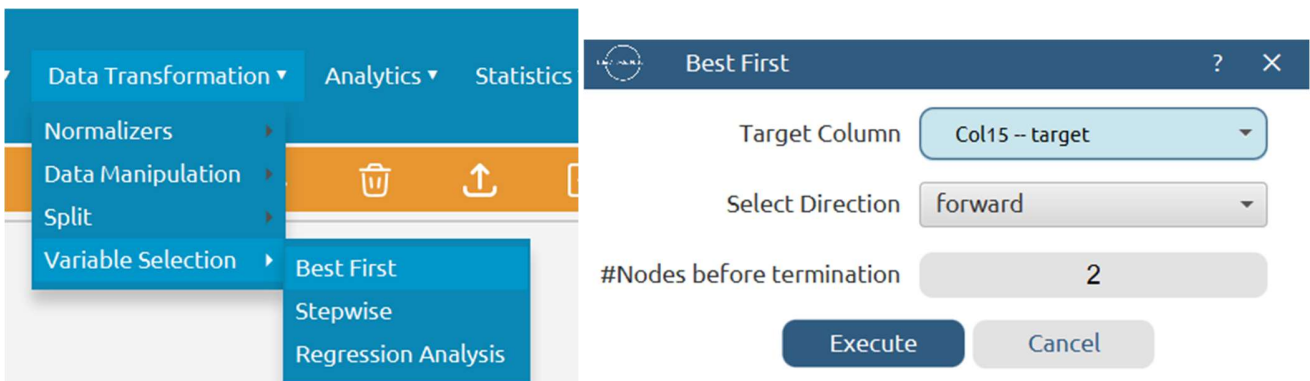
Step 4: Feature selection for training set

Create a new tab by pressing the “+” button on the bottom of the page with the name “FEATURE_SELECTION_TRAINING_SET”.

Import data into the input spreadsheet of the “FEATURE_SELECTION_TRAINING_SET” tab from the output of the “TRAIN_TEST_SPLIT” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”. From the available Select input tab options choose “TRAIN_TEST_SPLIT: Training Set”.

User Header	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (I)	Col8 (I)	Col9 (I)	Col10 (I)	Col11 (D)	Col12 (I)	Col13 (I)	Col14 (I)	Col15 (I)
	User Row ID	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
1		52	1	0	125	212	0	1	168	0	1	2	2	3	0
2		70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3		61	1	0	148	203	0	1	161	0	0	2	1	3	0
4		55	1	0	160	289	0	0	145	1	0.8	1	1	3	0
5		46	1	0	120	249	0	0	144	0	0.8	2	0	3	0
6		54	1	0	122	286	0	0	116	1	3.2	1	2	2	0
7		71	0	0	112	149	0	1	125	0	1.6	1	0	2	1
8		34	0	1	118	210	0	1	192	0	0.7	2	0	2	1
9		51	1	0	140	298	0	1	122	1	4.2	1	3	3	0
10		52	1	0	128	204	1	1	156	1	1	1	0	0	0
11		51	0	2	140	308	0	0	142	0	1.5	2	1	2	1
12		54	1	0	124	266	0	0	109	1	2.2	1	1	3	0
13		50	0	1	120	244	0	1	162	0	1.1	2	0	2	1
14		58	1	2	140	211	1	0	165	0	0	2	0	2	1
15		60	1	2	140	185	0	0	155	0	3	1	0	2	0
16		45	1	0	104	208	0	0	148	1	3	1	0	2	1
17		63	0	2	135	252	0	0	172	0	0	2	0	2	1
18		42	0	2	120	209	0	1	173	0	0	1	0	2	1

Then do Best First with the “Forward Selection” as the method: *Data Transformation* → *Variable Selection* → *Best First*



The results will appear on the right spreadsheet.

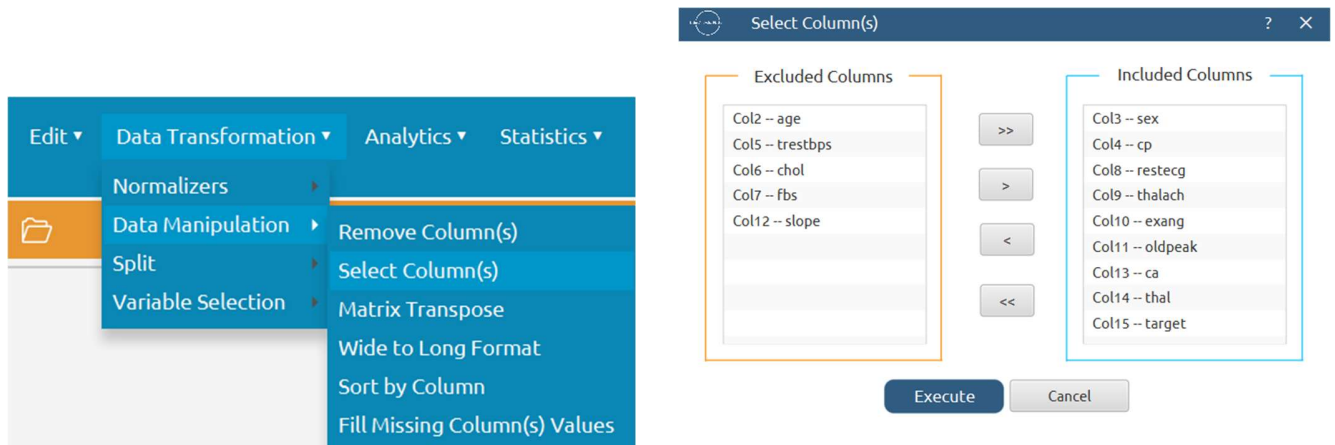
	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (D)	Col8 (I)	Col9 (I)	Col10 (I)
User Header	User Row ID	sex	cp	restecg	thalach	exang	oldpeak	ca	thal	target
1		1	0	1	168	0	1	2	3	0
2		1	0	1	125	1	2.6	0	3	0
3		1	0	1	161	0	0	1	3	0
4		1	0	0	145	1	0.8	1	3	0
5		1	0	0	144	0	0.8	0	3	0
6		1	0	0	116	1	3.2	2	2	0
7		0	0	1	125	0	1.6	0	2	1
8		0	1	1	192	0	0.7	0	2	1
9		1	0	1	122	1	4.2	3	3	0
10		1	0	1	156	1	1	0	0	0
11		0	2	0	142	0	1.5	1	2	1
12		1	0	0	109	1	2.2	1	3	0
13		0	1	1	162	0	1.1	0	2	1
14		1	2	0	165	0	0	0	2	1
15		1	2	0	155	0	3	0	2	0
16		1	0	0	148	1	3	0	2	1
17		0	2	0	172	0	0	0	2	1
18		0	2	1	173	0	0	0	2	1
19		0	0	0	146	1	1	0	3	0
20		1	2	1	179	1	0.4	0	2	1

Step 5: Feature selection for test set

Create a new tab by pressing the “+” button on the bottom of the page with the name “FEATURE_SELECTION_TEST_SET”.

Import data into the input spreadsheet of the “FEATURE_SELECTION_TEST_SET” tab from the output of the “TRAIN_TEST_SPLIT” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”. From the available Select input tab options choose “TRAIN_TEST_SPLIT: Test Set”

Knowing which features were selected for the training set, the same features should be selected for the test set as well, without applying any feature selection algorithm again. This is done simply by selecting the corresponding columns in the test dataset. On the menu click on *Data Transformation* → *Data Manipulation* → *Select Column(s)*



The results will appear on the right spreadsheet.

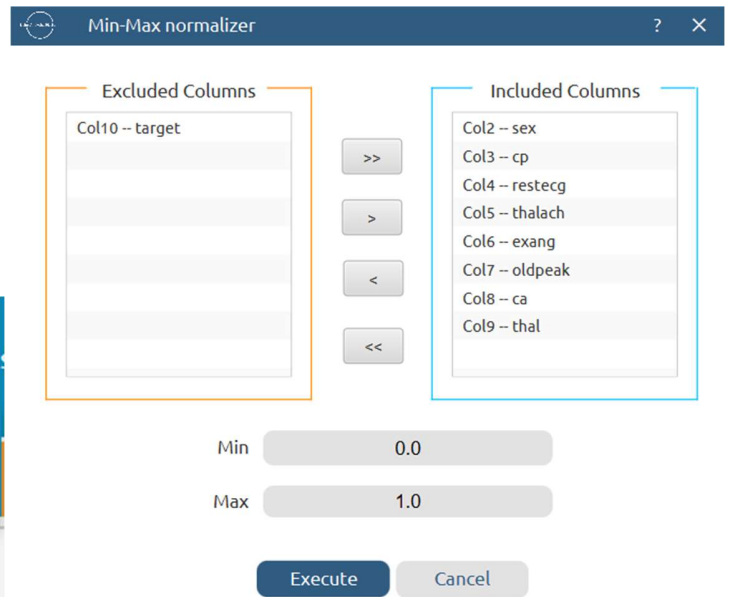
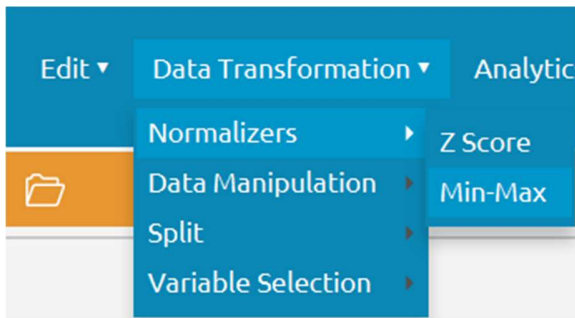
	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (D)	Col8 (I)	Col9 (I)	Col10 (I)
User Header	User Row ID	sex	cp	restecg	thalach	exang	oldpeak	ca	thal	target
1		1	0	0	155	1	3.1	0	3	0
2		0	0	1	106	0	1.9	3	2	0
3		0	0	0	122	0	1	0	2	1
4		1	0	2	140	0	4.4	3	1	0
5		0	0	0	136	1	3	0	3	0
6		0	1	1	192	0	0.7	0	2	1
7		0	0	1	142	0	0.3	2	2	1
8		1	2	1	112	1	2.9	1	3	0
9		0	2	0	152	0	0	1	2	1
10		1	0	1	160	1	1.4	2	3	0
11		0	3	0	162	0	1	0	2	1
12		1	0	1	111	1	5.6	0	3	0
13		1	1	0	202	0	0	0	2	1
14		0	2	0	152	0	0	1	2	1
15		1	2	1	157	0	1.6	0	2	1
16		1	0	1	125	1	1.8	0	1	0
17		1	0	0	140	1	3.4	0	3	0
18		1	0	0	144	1	1.2	1	2	0
19		1	0	0	150	1	0	2	3	0
20		1	1	1	158	0	0.8	1	2	1

Step 6: Normalize the training set

Create a new tab by pressing the “+” button on the bottom of the page with the name “NORMALIZE_TRAIN_SET”.

Import into the input spreadsheet of the “NORMALIZE_TRAIN_SET” tab the train set from the output of the “FEATURE_SELECTION_TRAINING_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Normalize the data using Min-Max normalizer: *Data Transformation* → *Normalizers* → *Min-Max* and select all columns except the “target” target column.



The results will appear on the output spreadsheet.

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)	Col10 (D)
User Header	User Row ID	sex	cp	restecg	thalach	exang	oldpeak	ca	thal	target
1		1.0	0.0	0.5	0.7404580	0.0	0.1612903	0.5	1.0	0.0
2		1.0	0.0	0.5	0.4122137	1.0	0.4193548	0.0	1.0	0.0
3		1.0	0.0	0.5	0.6870229	0.0	0.0	0.25	1.0	0.0
4		1.0	0.0	0.0	0.5648855	1.0	0.1290323	0.25	1.0	0.0
5		1.0	0.0	0.0	0.5572519	0.0	0.1290323	0.0	1.0	0.0
6		1.0	0.0	0.0	0.3435115	1.0	0.5161290	0.5	0.6666667	0.0
7		0.0	0.0	0.5	0.4122137	0.0	0.2580645	0.0	0.6666667	1.0
8		0.0	0.3333333	0.5	0.9236641	0.0	0.1129032	0.0	0.6666667	1.0
9		1.0	0.0	0.5	0.3893130	1.0	0.6774194	0.75	1.0	0.0
10		1.0	0.0	0.5	0.6488550	1.0	0.1612903	0.0	0.0	0.0
11		0.0	0.6666667	0.0	0.5419847	0.0	0.2419355	0.25	0.6666667	1.0
12		1.0	0.0	0.0	0.2900763	1.0	0.3548387	0.25	1.0	0.0
13		0.0	0.3333333	0.5	0.6946565	0.0	0.1774194	0.0	0.6666667	1.0
14		1.0	0.6666667	0.0	0.7175573	0.0	0.0	0.0	0.6666667	1.0
15		1.0	0.6666667	0.0	0.6412214	0.0	0.4838710	0.0	0.6666667	0.0
16		1.0	0.0	0.0	0.5877863	1.0	0.4838710	0.0	0.6666667	1.0
17		0.0	0.6666667	0.0	0.7709924	0.0	0.0	0.0	0.6666667	1.0
18		0.0	0.6666667	0.5	0.7786260	0.0	0.0	0.0	0.6666667	1.0
19		0.0	0.0	0.0	0.5725191	1.0	0.1612903	0.0	1.0	0.0

Step 7: Select Columns except Target

Create a new tab by pressing the “+” button on the bottom of the page with the name “SELECT_COLUMNS_EXCEPT_TARGET”.

Import into the input spreadsheet of the “SELECT_COLUMNS_EXCEPT_TARGET” tab the train set from the output of the “FEATURE_SELECTION_TEST_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Select columns of the data using Select Column(s): *Data Transformation* → *Data Manipulation* → *Select Column(s)* and select all columns except the “target” target column.

The image shows two parts of the software interface. On the left, a menu is open under 'Data Manipulation', with 'Select Column(s)' highlighted. On the right, the 'Select Column(s)' dialog box is shown. It has two columns: 'Excluded Columns' and 'Included Columns'. 'Col10 -- target' is listed in the 'Excluded Columns' list. The 'Included Columns' list contains: 'Col2 -- sex', 'Col3 -- cp', 'Col4 -- restecg', 'Col5 -- thalach', 'Col6 -- exang', 'Col7 -- oldpeak', 'Col8 -- ca', and 'Col9 -- thal'. There are navigation buttons (>>, >, <, <<) between the lists and 'Execute' and 'Cancel' buttons at the bottom.

The results will appear on the right spreadsheet.

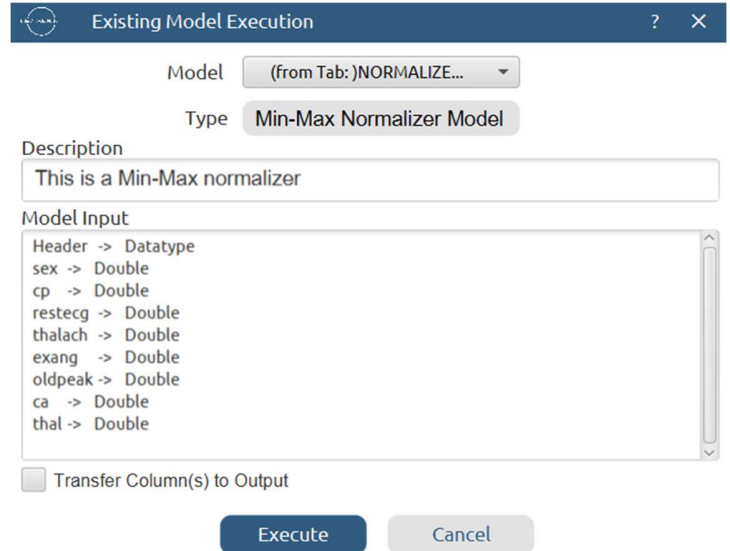
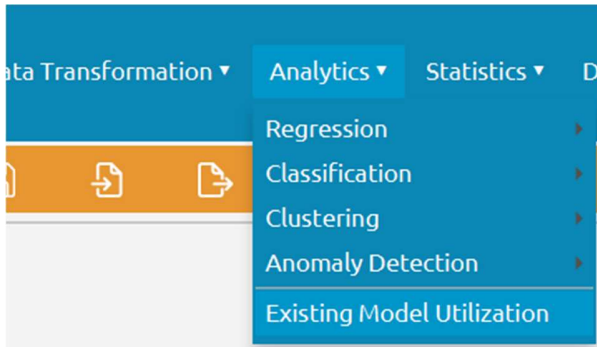
	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (D)	Col8 (I)	Col9 (I)
User Header	User Row ID	sex	cp	restecg	thalach	exang	oldpeak	ca	thal
1		1	0	0	155	1	3.1	0	3
2		0	0	1	106	0	1.9	3	2
3		0	0	0	122	0	1	0	2
4		1	0	2	140	0	4.4	3	1
5		0	0	0	136	1	3	0	3
6		0	1	1	192	0	0.7	0	2
7		0	0	1	142	0	0.3	2	2
8		1	2	1	112	1	2.9	1	3
9		0	2	0	152	0	0	1	2
10		1	0	1	160	1	1.4	2	3
11		0	3	0	162	0	1	0	2
12		1	0	1	111	1	5.6	0	3
13		1	1	0	202	0	0	0	2
14		0	2	0	152	0	0	1	2
15		1	2	1	157	0	1.6	0	2
16		1	0	1	125	1	1.8	0	1
17		1	0	0	140	1	3.4	0	3
18		1	0	0	144	1	1.2	1	2
19		1	0	0	150	1	0	2	3
20		1	1	1	158	0	0.8	1	2

Step 8: Normalize the test set

Create a new tab by pressing the “+” button on the bottom of the page with the name “NORMALIZE_TEST_SET”.

Import into the input spreadsheet of the “NORMALIZE_TEST_SET” tab the train set from the output of the “SELECT_COLUMNS_EXCEPT_TARGET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Normalize the test set using the existing normalizer of the training set: *Analytics* → *Existing Model Utilization* → *Model (from Tab:) NORMALIZE_TRAIN_SET*



The results will appear on the output spreadsheet.

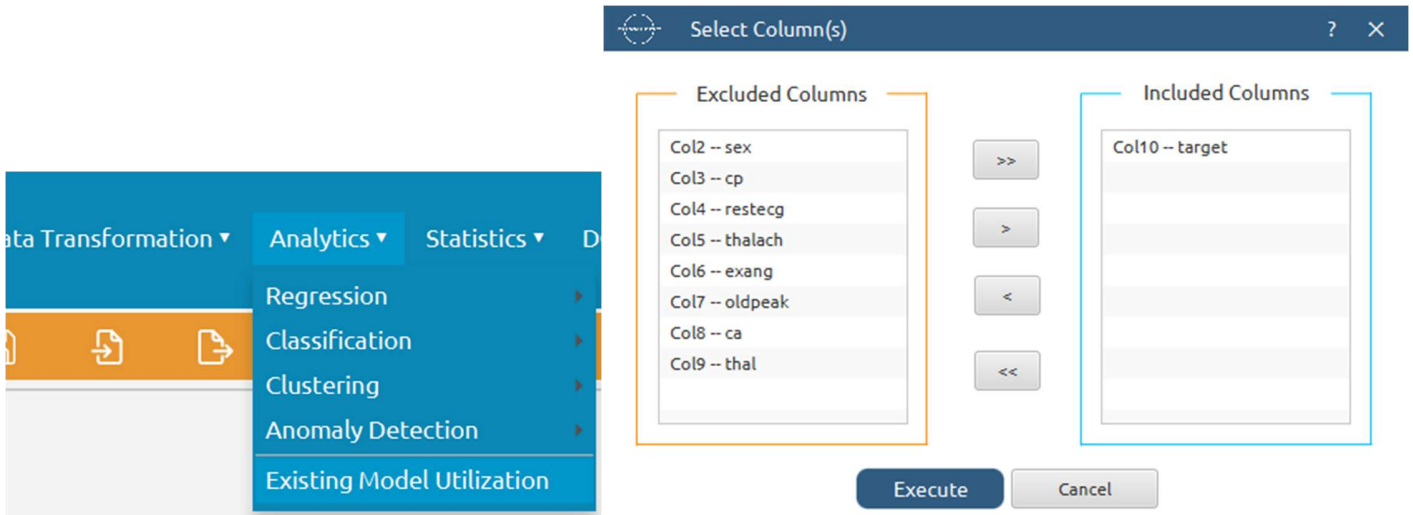
	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)
User Header	User Row ID	sex	cp	restecg	thalach	exang	oldpeak	ca	thal
1		1.0	0.0	0.0	0.6412214	1.0	0.5	0.0	1.0
2		0.0	0.0	0.5	0.2671756	0.0	0.3064516	0.75	0.6666667
3		0.0	0.0	0.0	0.3893130	0.0	0.1612903	0.0	0.6666667
4		1.0	0.0	1.0	0.5267176	0.0	0.7096774	0.75	0.3333333
5		0.0	0.0	0.0	0.4961832	1.0	0.4838710	0.0	1.0
6		0.0	0.3333333	0.5	0.9236641	0.0	0.1129032	0.0	0.6666667
7		0.0	0.0	0.5	0.5419847	0.0	0.0483871	0.5	0.6666667
8		1.0	0.6666667	0.5	0.3129771	1.0	0.4677419	0.25	1.0
9		0.0	0.6666667	0.0	0.6183206	0.0	0.0	0.25	0.6666667
10		1.0	0.0	0.5	0.6793893	1.0	0.2258065	0.5	1.0
11		0.0	1.0	0.0	0.6946565	0.0	0.1612903	0.0	0.6666667
12		1.0	0.0	0.5	0.3053435	1.0	0.9032258	0.0	1.0
13		1.0	0.3333333	0.0	1.0	0.0	0.0	0.0	0.6666667
14		0.0	0.6666667	0.0	0.6183206	0.0	0.0	0.25	0.6666667
15		1.0	0.6666667	0.5	0.6564885	0.0	0.2580645	0.0	0.6666667
16		1.0	0.0	0.5	0.4122137	1.0	0.2903226	0.0	0.3333333
17		1.0	0.0	0.0	0.5267176	1.0	0.5483871	0.0	1.0
18		1.0	0.0	0.0	0.5572519	1.0	0.1935484	0.25	0.6666667
19		1.0	0.0	0.0	0.6030534	1.0	0.0	0.5	1.0
20		1.0	0.3333333	0.5	0.6641221	0.0	0.1290323	0.25	0.6666667

Step 9: Select the target Column

Create a new tab by pressing the “+” button on the bottom of the page with the name “SELECT_TEST_SET_TARGET”.

Import into the input spreadsheet of the “SELECT_TEST_SET_TARGET” tab the train set from the output of the “FEATURE_SELECTION_TEST_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Select columns of the data using Select Column(s): *Data Transformation* → *Data Manipulation* → *Select Column(s)* and select only the “target” target column.



The results will appear on the output spreadsheet.

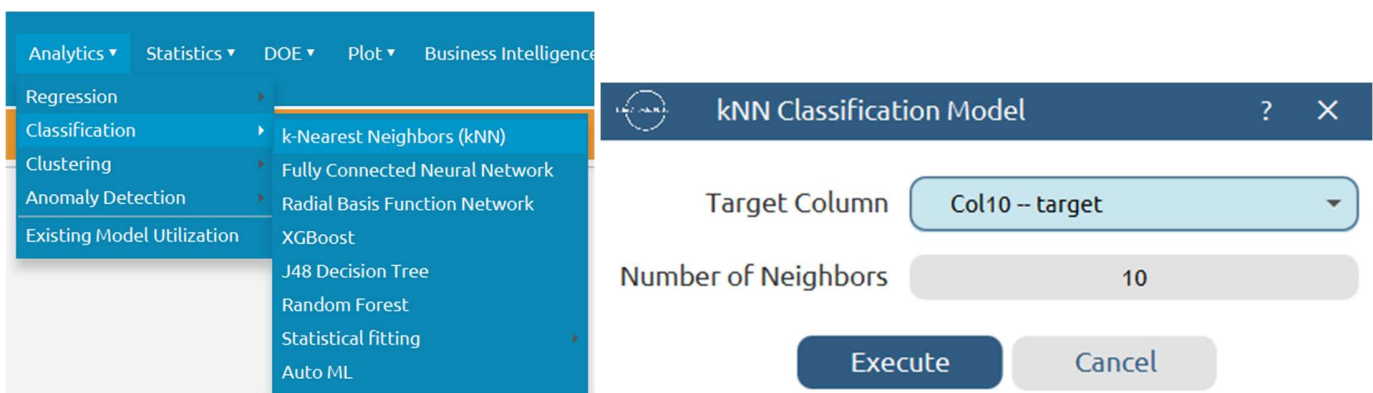
	Col1	Col2 (I)
User Header	User Row ID	target
1		0
2		0
3		1
4		0
5		0
6		1
7		1
8		0
9		1
10		0
11		1
12		0
13		1
14		1
15		1

Step 10: Train the Model-kNN

Create a new tab by pressing the “+” button on the bottom of the page with the name “kNN”.

Import data into the input spreadsheet of the “kNN” tab from the output of the “NORMALIZE_TRAIN_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Use the k-Nearest Neighbors (kNN) method to train and fit the model: *Analytics* → *Classification* → *kNearest Neighbors (kNN)*.



The screenshot shows the software's menu system and a configuration dialog. On the left, the 'Analytics' menu is expanded to show 'Classification', which is further expanded to show 'k-Nearest Neighbors (kNN)'. On the right, the 'kNN Classification Model' dialog box is open. It features a 'Target Column' dropdown menu with 'Col10 -- target' selected, and a 'Number of Neighbors' input field with the value '10'. At the bottom of the dialog are 'Execute' and 'Cancel' buttons.

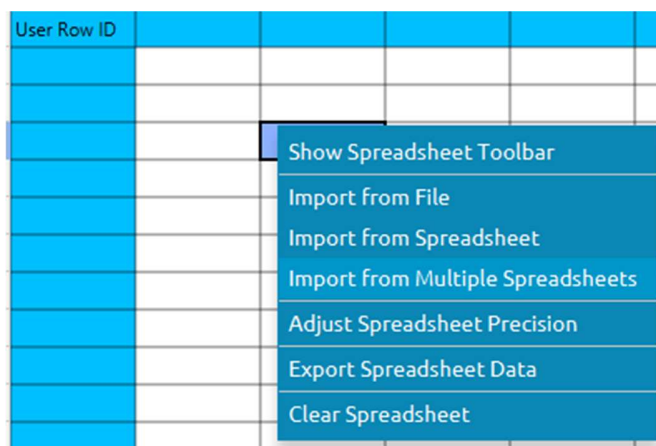
The predictions will appear on the output spreadsheet.

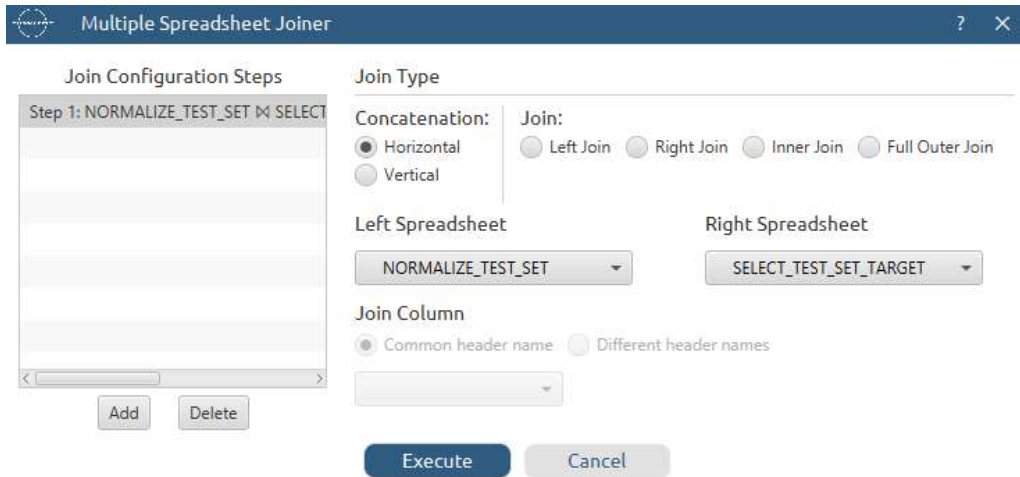
	Col1	Col2 (D)	Col3 (D)	Col4 (S)	Col5 (D)	Col6 (S)	Col7 (D)	Col8 (S)	Col9 (D)	Col10 (S)	Col11 (D)
User Header	User Row ID	target	kNN Prediction	Closest NN1	Distance from NN1	Closest NN2	Distance from NN2	Closest NN3	Distance from NN3	Closest NN4	Distance from NN4
1		0.0	0.0	Entry 672	0.0	Entry 520	0.0	Entry 491	0.0	Entry 1	0.0
2		0.0	0.0	Entry 747	0.0	Entry 691	0.0	Entry 2	0.0	Entry 716	0.1657449
3		0.0	0.0	Entry 461	0.0	Entry 406	0.0	Entry 3	0.0	Entry 524	0.0414359
4		0.0	0.0	Entry 606	0.0	Entry 211	0.0	Entry 4	0.0	Entry 790	0.0684600
5		0.0	0.0	Entry 718	0.0	Entry 36	0.0	Entry 5	0.0	Entry 814	0.1675427
6		0.0	0.0	Entry 455	0.0	Entry 424	0.0	Entry 167	0.0	Entry 6	0.0
7		1.0	1.0	Entry 504	0.0	Entry 388	0.0	Entry 311	0.0	Entry 7	0.0
8		1.0	1.0	Entry 8	0.0	Entry 589	0.1898837	Entry 231	0.1898837	Entry 233	0.2142319
9		0.0	0.0	Entry 608	0.0	Entry 370	0.0	Entry 9	0.0	Entry 566	0.4610607
10		0.0	0.0	Entry 700	0.0	Entry 572	0.0	Entry 533	0.0	Entry 10	0.0
11		1.0	1.0	Entry 800	0.0	Entry 732	0.0	Entry 11	0.0	Entry 757	0.1608050
12		0.0	0.0	Entry 728	0.0	Entry 642	0.0	Entry 12	0.0	Entry 313	0.2266005
13		1.0	1.0	Entry 712	0.0	Entry 25	0.0	Entry 13	0.0	Entry 233	0.0345325
14		1.0	1.0	Entry 802	0.0	Entry 315	0.0	Entry 64	0.0	Entry 14	0.0
15		0.0	0.0	Entry 532	0.0	Entry 186	0.0	Entry 15	0.0	Entry 772	0.1281254
16		1.0	1.0	Entry 204	0.0	Entry 16	0.0	Entry 657	0.3204543	Entry 638	0.3204543
17		1.0	1.0	Entry 639	0.0	Entry 17	0.0	Entry 759	0.0280105	Entry 740	0.0280105
18		1.0	1.0	Entry 433	0.0	Entry 298	0.0	Entry 18	0.0	Entry 494	0.0076336

Step 11: Validate the model-kNN

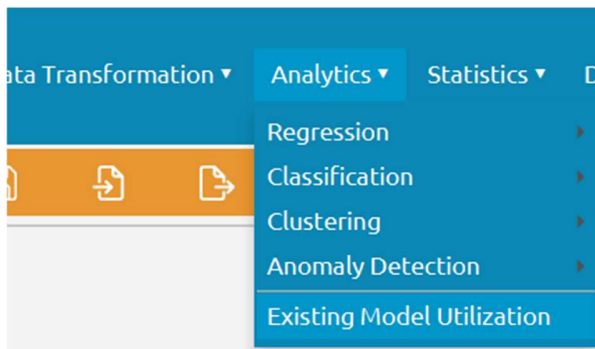
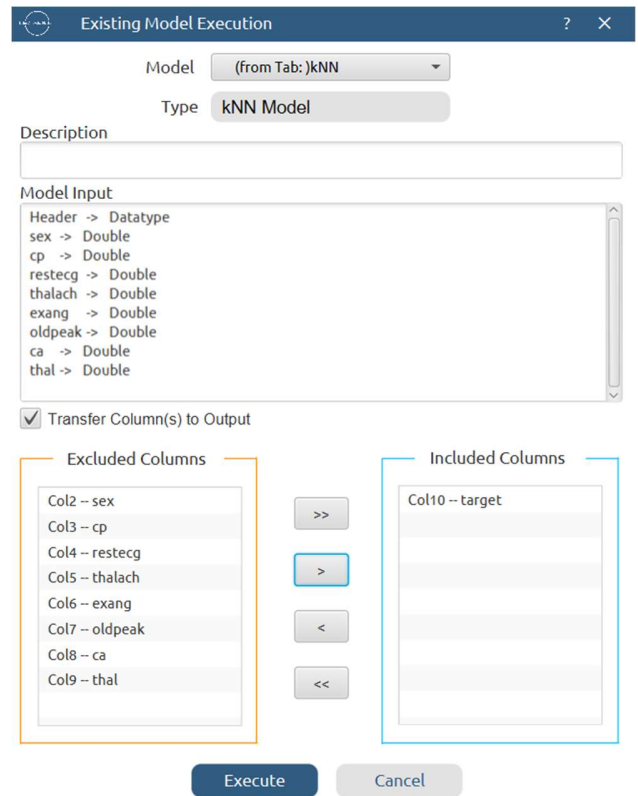
Create a new tab by pressing the “+” button on the bottom of the page with the name “VALIDATE_MODEL_KNN”.

Import data into the input spreadsheet of the “VALIDATE_MODEL_KNN” tab from the outputs of the “NORMALIZE_TEST_SET” tab and “SELECT_TEST_SET_TARGET” by right-clicking on the input spreadsheet and then choosing “Import from Multiple Spreadsheets”.





To validate the model: *Analytics* → *Existing Model Utilization* → *Model (from Tab:) kNN(.fit)*. Choose the column “target” to be transferred to the output spreadsheet.



The predictions will appear on the output spreadsheet.

	Col1	Col2 (I)	Col3 (S)	Col4 (D)	Col5 (S)	Col6 (D)	Col7 (S)	Col8 (D)	Col9 (S)	Col10 (D)
User Header	User Row ID	kNN Prediction	Closest NN1	Distance from NN1	Closest NN2	Distance from NN2	Closest NN3	Distance from NN3	Closest NN4	Distance from NN4
1		0	Entry 767	0.0	Entry 628	0.0	Entry 538	0.0	Entry 817	0.2761568
2		0	Entry 609	0.0	Entry 482	0.0	Entry 604	0.2787072	Entry 503	0.2787072
3		1	Entry 760	0.0	Entry 513	0.0	Entry 706	0.0645161	Entry 326	0.0645161
4		0	Entry 809	0.0	Entry 511	0.0	Entry 112	0.0	Entry 612	0.8848826
5		0	Entry 282	0.0	Entry 85	0.3314897	Entry 19	0.3314897	Entry 534	0.5258698
6		1	Entry 8	0.0	Entry 589	0.1898837	Entry 231	0.1898837	Entry 233	0.2142319
7		1	Entry 780	0.0	Entry 422	0.0	Entry 604	0.3138324	Entry 503	0.3138324
8		0	Entry 458	0.0	Entry 453	0.3391367	Entry 380	0.3391367	Entry 146	0.3391367
9		1	Entry 306	0.0	Entry 757	0.1345590	Entry 515	0.1345590	Entry 598	0.1679389
10		0	Entry 787	0.0	Entry 583	0.0	Entry 184	0.0	Entry 636	0.2782401
11		1	Entry 593	0.0	Entry 346	0.3416581	Entry 109	0.3416581	Entry 585	0.3452572
12		0	Entry 646	0.0	Entry 476	0.0	Entry 44	0.0	Entry 693	0.4642895
13		1	Entry 518	0.0	Entry 87	0.0	Entry 50	0.0	Entry 358	0.1068702
14		1	Entry 306	0.0	Entry 757	0.1345590	Entry 515	0.1345590	Entry 598	0.1679389
15		1	Entry 225	0.0	Entry 150	0.0	Entry 379	0.1297710	Entry 343	0.1297710
16		0	Entry 794	0.0	Entry 471	0.0	Entry 694	0.0489855	Entry 547	0.0489855
17		0	Entry 767	0.1243078	Entry 628	0.1243078	Entry 538	0.1243078	Entry 531	0.2106667
18		0	Entry 320	0.0	Entry 125	0.0	Entry 798	0.1996716	Entry 717	0.1996716

Step 12: Statistics calculation-kNN

Create a new tab by pressing the “+” button on the bottom of the page with the name “STATISTICS_ACCURACIES_kNN”.

Import data into the input spreadsheet of the “STATISTICS_ACCURACIES_kNN” tab from the output of the “VALIDATE_MODEL_kNN” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Calculate the statistical metrics for the regression: *Statistics* → *Model Metrics* → *Classification Metrics*

The results will appear on the output spreadsheet.

	Col1 (S)	Col2 (D)	Col3 (S)	Col4 (S)
User Header	User Row ID			
1			Predicted Class	Predicted Class
2			0.0	1.0
3	Actual Class	0.0	100	0
4	Actual Class	1.0	0	105
5				
6				
7	Classification Accuracy	1.0		
8				
9	Precision		1.0	1.0
10				
11	Recall/Sensitivity		1.0	1.0
12				
13	Specificity		1.0	1.0
14				
15	F1 Score		1.0	1.0
16				
17	F (beta=2)		1.0	1.0
18				
19	MCC	1.0		

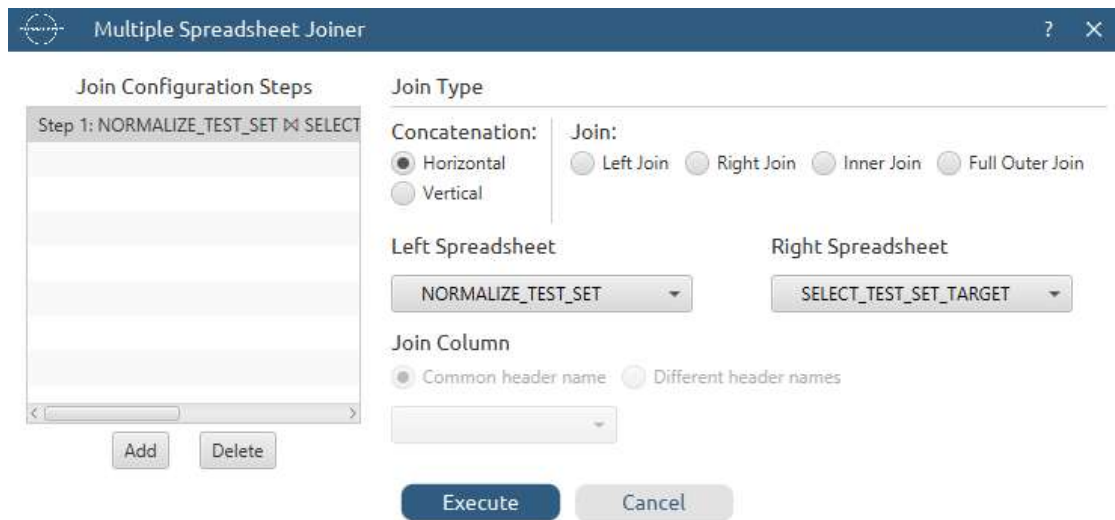
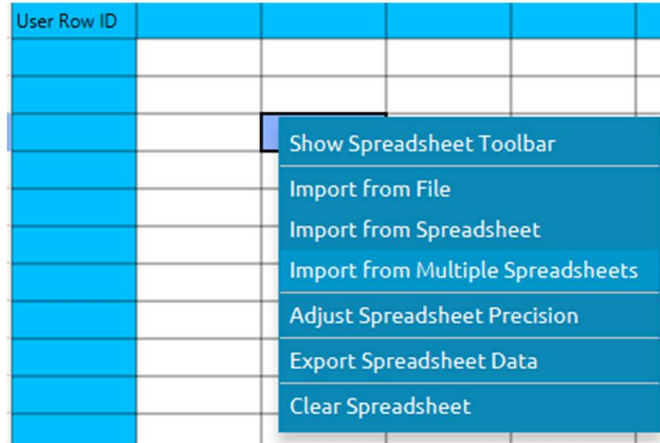
Step 13: Train the Model – Logistic Regression

Create a new tab by pressing the “+” button on the bottom of the page with the name “Logistic Regression”.

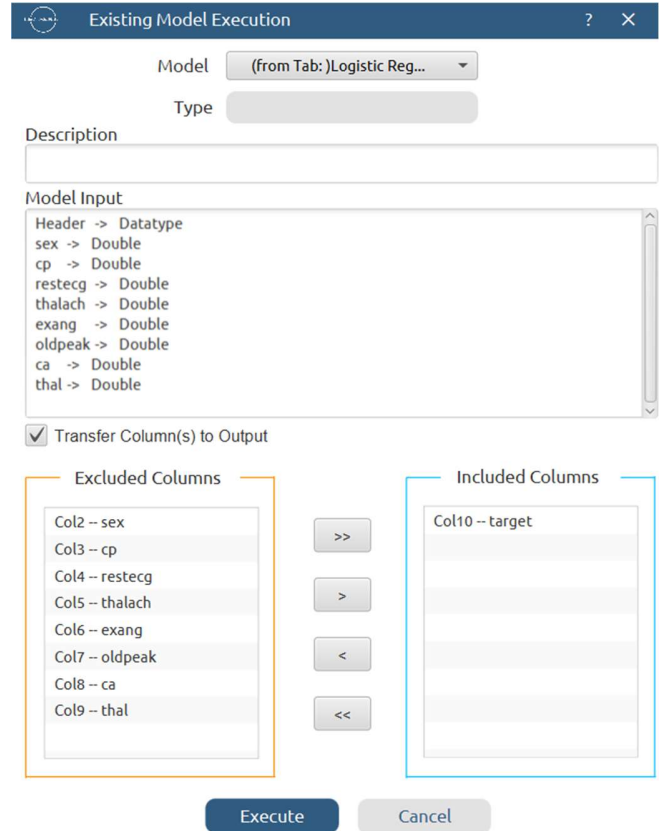
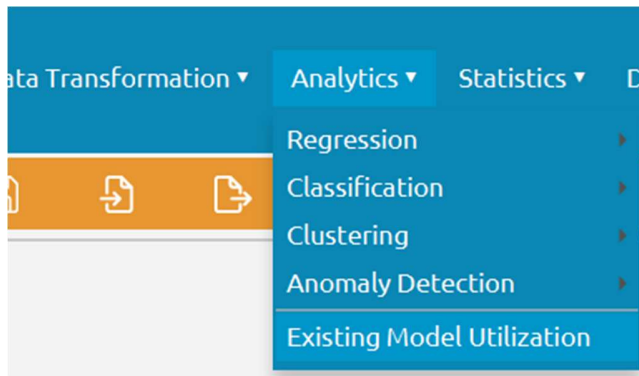
Import data into the input spreadsheet of the “Logistic Regression” tab from the output of the “NORMALIZE_TRAIN_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Use the Logistic Regression method to train and fit the model: Analytics → Classification → Statistical fitting → Generalized Linear Models.

The screenshot displays the 'Generalized Linear Models Classification' configuration window in the Isalos Analytics Platform. The window title is 'Generalized Linear Models Classification'. The 'Type' is set to 'Binary Logistic'. The 'Confidence Level (%)' is 95, 'Max Iterations' is 25, and 'Maximum Step-Halving' is 5. The 'Dependent Variable' is 'Col10 -- target' and the 'Reference Category' is '0.0'. The 'Parameter Estimation Method' is 'Newton-Ra...', 'Maximum Scoring Iterations' is 0, and 'Minimum Change in Parameter Estimates' is 1.0E-6. The 'Scale Parameter Method' is 'Fixed value' with a 'Value' of 1.0. Under 'Excluded Columns', there is an empty box. Under 'Factors', 'Col2 -- sex', 'Col3 -- cp', 'Col4 -- restecg', and 'Col6 -- exang' are listed. Under 'Covariates', 'Col5 -- thalach' and 'Col7 -- oldpeak' are listed. At the bottom, the 'Include All Main Effects' radio button is selected. The 'Formula' field contains 'Formula with Selected Column Names'. 'Execute' and 'Cancel' buttons are at the bottom.



To validate the model: *Analytics* → *Existing Model Utilization* → *Model (from Tab:) Logistic Regression (.fit)*.
 Choose the column “target” to be transferred to the output spreadsheet.



The predictions will appear on the output spreadsheet.

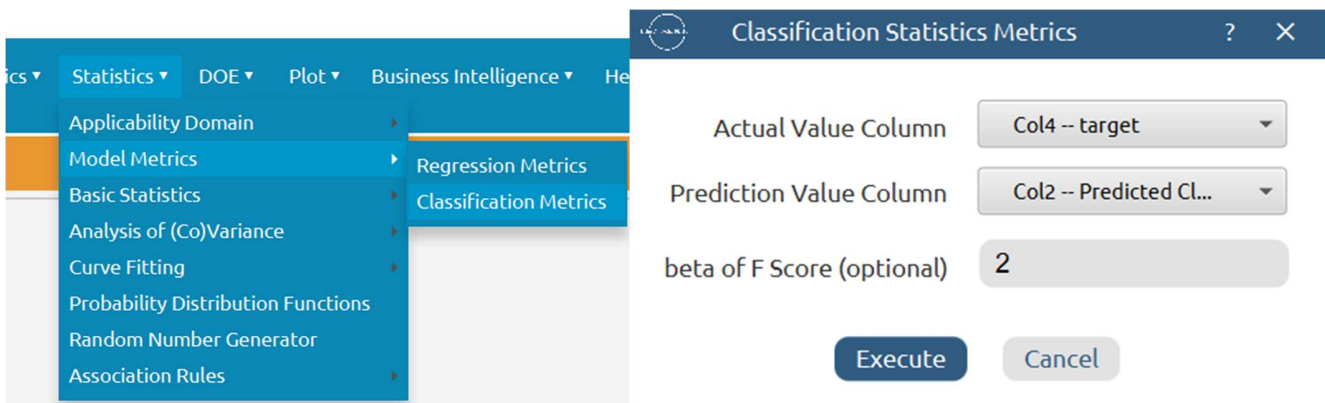
	Col1	Col2 (I)	Col3 (D)	Col4 (I)
User Header	User Row ID	Predicted Class	Probability	target
1		0	0.9723701	0
2		0	0.6877876	0
3		1	0.7160550	1
4		0	0.9878664	0
5		0	0.9077166	0
6		1	0.9876664	1
7		0	0.5582398	1
8		0	0.9797895	0
9		1	0.9023638	1
10		0	0.9878497	0
11		1	0.9837894	1
12		0	0.9964807	0
13		1	0.9445153	1
14		1	0.9023638	1
15		1	0.9003943	1
16		0	0.7715002	0
17		0	0.9838165	0
18		0	0.9584774	0

Step 15: Statistics calculation – Logistic Regression

Create a new tab by pressing the “+” button on the bottom of the page with the name “STATISTICS_ACCURACIES_ Logistic Regression”.

Import data into the input spreadsheet of the “STATISTICS_ACCURACIES_ Logistic Regression” tab from the output of the “VALIDATE_MODEL_ Logistic Regression” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Calculate the statistical metrics for the regression: *Statistics → Model Metrics → Classification Metrics*



The results will appear on the output spreadsheet.

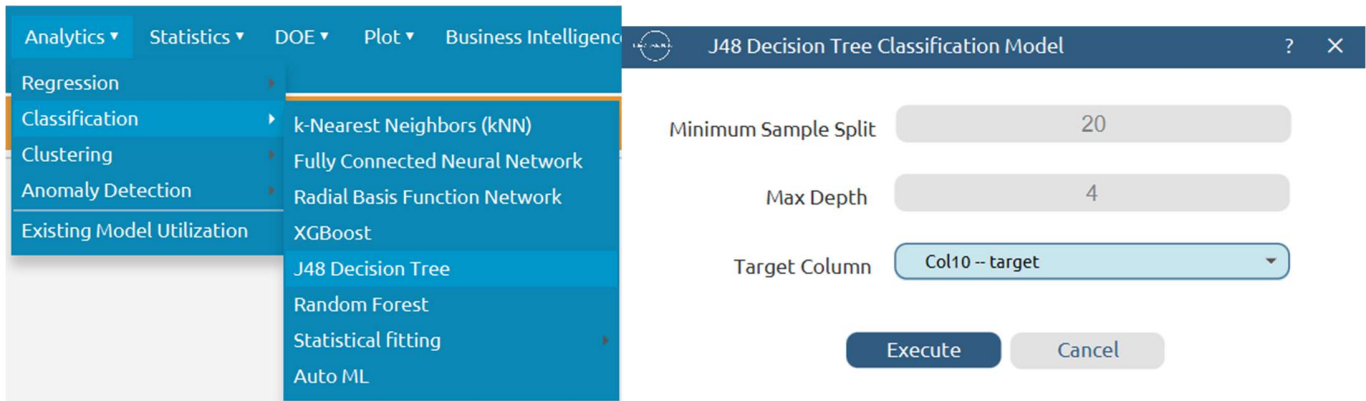
	Col1 (S)	Col2 (D)	Col3 (S)	Col4 (S)
User Header	User Row ID			
1			Predicted Class	Predicted Class
2			0.0	1.0
3	Actual Class	0.0	85	15
4	Actual Class	1.0	14	91
5				
6				
7	Classification Accuracy	0.8585366		
8				
9	Precision		0.8585859	0.8584906
10				
11	Recall/Sensitivity		0.85	0.8666667
12				
13	Specificity		0.8666667	0.85
14				
15	F1 Score		0.8542714	0.8625592
16				
17	F (beta=2)		0.8517034	0.8650190
18				
19	MCC	0.7168715		
--				

Step 14: Train the Model – Decision Tree

Create a new tab by pressing the “+” button on the bottom of the page with the name “Decision Tree”.

Import data into the input spreadsheet of the “Decision Tree” tab from the output of the “FEATURE_SELECTION_TRAINING_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Use the Decision Tree method to train and fit the model: *Analytics* → *Classification* → *Decision Tree*



The predictions will appear on the output spreadsheet.

	Col1	Col2 (I)	Col3 (I)
User Header	User Row ID	target	Prediction
1		0	0
2		0	0
3		0	0
4		0	0
5		0	0
6		0	0
7		1	1
8		1	1
9		0	0
10		0	0
11		1	1
12		0	0
13		1	1
14		1	1
15		0	0
16		1	0
17		1	1
18		1	1

Step 15: Validate the Model – Decision Tree

Create a new tab by pressing the “+” button on the bottom of the page with the name “VALIDATE_MODEL_ Decision Tree”.

Import data into the input spreadsheet of the “VALIDATE_MODEL_ Decision Tree” tab from the output of the “FEATURE_SELECTION_TEST_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

To validate the model: *Analytics* → *Existing Model Utilization* → *Model (from Tab:) Decision Tree (.fit)*.

The image shows two parts of the software interface. On the left is a partial view of the 'Analytics' menu, which includes options like 'Regression', 'Classification', 'Clustering', 'Anomaly Detection', and 'Existing Model Utilization'. On the right is the 'Existing Model Execution' dialog box. It has a title bar with a question mark and a close button. The 'Model' dropdown is set to '(from Tab:) Decision Tree' and the 'Type' is 'J48 Decision Tree Classific'. There is a 'Description' text area and a 'Model Input' list with the following items: sex -> Double, cp -> Double, restecg -> Double, thalach -> Double, exang -> Double, oldpeak -> Double, ca -> Double, thal -> Double. A checkbox for 'Transfer Column(s) to Output' is unchecked. At the bottom are 'Execute' and 'Cancel' buttons.

The predictions will appear on the output spreadsheet.

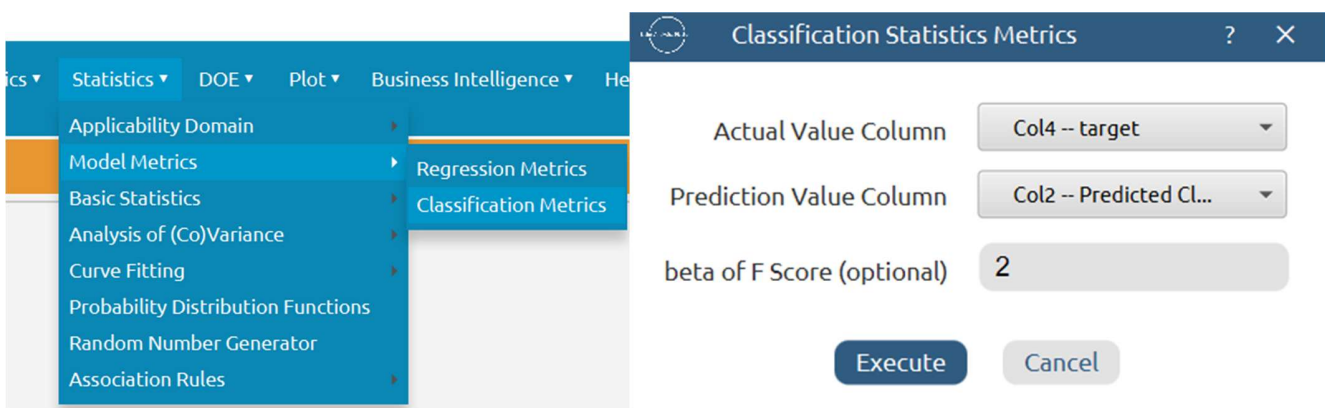
User Header	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (D)	Col8 (I)	Col9 (I)	Col10 (I)	Col11 (I)
User Row ID	sex	cp	restecg	thalach	exang	oldpeak	ca	thal	target	Prediction	
1	1	0	0	155	1	3.1	0	3	0	0	
2	0	0	1	106	0	1.9	3	2	0	0	
3	0	0	0	122	0	1	0	2	1	1	
4	1	0	2	140	0	4.4	3	1	0	0	
5	0	0	0	136	1	3	0	3	0	0	
6	0	1	1	192	0	0.7	0	2	1	1	
7	0	0	1	142	0	0.3	2	2	1	1	
8	1	2	1	112	1	2.9	1	3	0	0	
9	0	2	0	152	0	0	1	2	1	1	
10	1	0	1	160	1	1.4	2	3	0	0	
11	0	3	0	162	0	1	0	2	1	1	
12	1	0	1	111	1	5.6	0	3	0	0	
13	1	1	0	202	0	0	0	2	1	1	
14	0	2	0	152	0	0	1	2	1	1	
15	1	2	1	157	0	1.6	0	2	1	1	
16	1	0	1	125	1	1.8	0	1	0	0	
17	1	0	0	140	1	3.4	0	3	0	0	
18	1	0	0	144	1	1.2	1	2	0	0	
19	1	0	0	150	1	0	2	3	0	0	
20	1	1	1	158	0	0.8	1	2	1	1	
21	1	1	1	158	0	0.8	1	2	1	1	
22	1	2	0	126	0	0.8	3	2	0	1	

Step 16: Statistics calculation – Decision Tree

Create a new tab by pressing the “+” button on the bottom of the page with the name “STATISTICS_ACCURACIES_ Decision Tree”.

Import data into the input spreadsheet of the “STATISTICS_ACCURACIES_ Decision Tree” tab from the output of the “VALIDATE_MODEL_ Decision Tree” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Calculate the statistical metrics for the regression: *Statistics → Model Metrics → Classification Metrics*



The results will appear on the output spreadsheet.

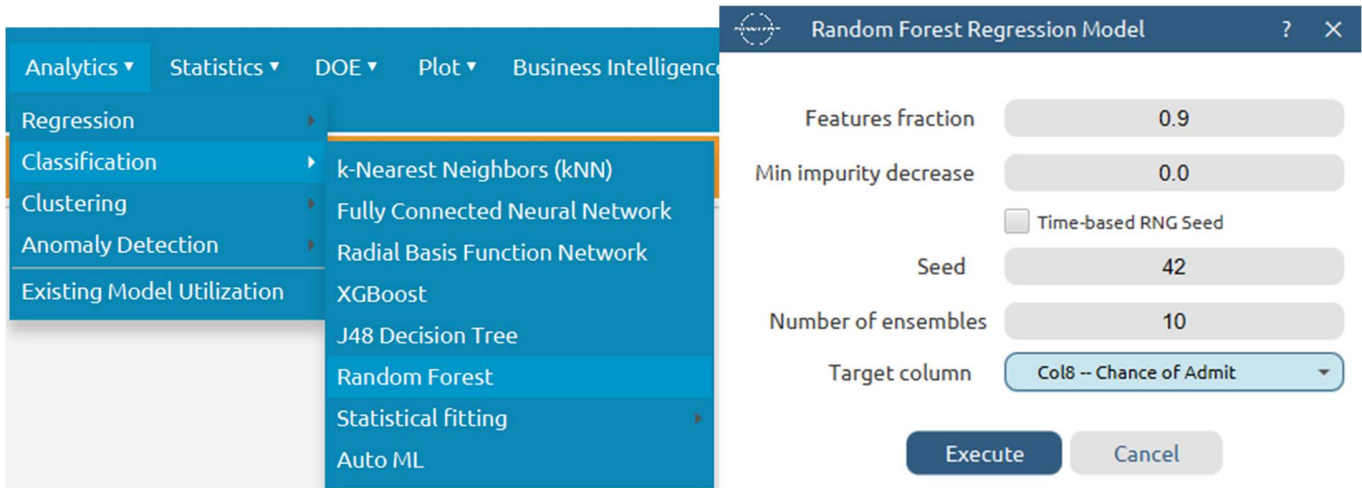
	Col1 (S)	Col2 (D)	Col3 (S)	Col4 (S)
User Header	User Row ID			
1			Predicted Class	Predicted Class
2			0	1
3	Actual Class	0	79	21
4	Actual Class	1	10	95
5				
6				
7	Classification Accuracy	0.8487805		
8				
9	Precision		0.8876404	0.8189655
10				
11	Recall/Sensitivity		0.79	0.9047619
12				
13	Specificity		0.9047619	0.79
14				
15	F1 Score		0.8359788	0.8597285
16				
17	F (beta=2)		0.8077710	0.8861940
18				
19	MCC	0.7006589		

Step 17: Train the Model – Random Forest

Create a new tab by pressing the “+” button on the bottom of the page with the name “Random Forest”.

Import data into the input spreadsheet of the “Random Forest” tab from the output of the “FEATURE_SELECTION_TRAINING_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Use the Random Forest method to train and fit the model: *Analytics* → *Classification* → *Random Forest*.



The predictions will appear on the output spreadsheet.

	Col1	Col2 (I)	Col3 (I)
User Header	User Row ID	target	Prediction
1		0	0
2		0	0
3		0	0
4		0	0
5		0	0
6		0	0
7		1	1
8		1	1
9		0	0
10		0	0
11		1	1
12		0	0
13		1	1
14		1	1
15		0	0
16		1	1
17		1	1
18		1	1

Step 18: Validate the Model – Random Forest

Create a new tab by pressing the “+” button on the bottom of the page with the name “VALIDATE_MODEL_Random Forest”.

Import data into the input spreadsheet of the “VALIDATE_MODEL_Random Forest” tab from the output of the “FEATURE_SELECTION_TEST_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

To validate the model: *Analytics* → *Existing Model Utilization* → *Model (from Tab:) Random Forest (.fit)*. Choose the column “target” to be transferred to the output spreadsheet.

The image shows a software interface with a menu on the left and a dialog box on the right. The menu path is: **Analytics** → **Existing Model Utilization**. The dialog box, titled "Existing Model Execution", has the following fields and options:

- Model:** (from Tab:)Random For...
- Type:** (empty)
- Description:** It is a Random forest model
- Model Input:** A list of variables and their datatypes:
 - Header -> Datatype
 - sex -> Double
 - cp -> Double
 - restecg -> Double
 - thalach -> Double
 - exang -> Double
 - oldpeak -> Double
 - ca -> Double
 - thal -> Double
- Transfer Column(s) to Output**
- Excluded Columns:** Col2 -- sex, Col3 -- cp, Col4 -- restecg, Col5 -- thalach, Col6 -- exang, Col7 -- oldpeak, Col8 -- ca, Col9 -- thal
- Included Columns:** Col10 -- target
- Buttons:** >>, >, <, <<, Execute, Cancel

The predictions will appear on the output spreadsheet.

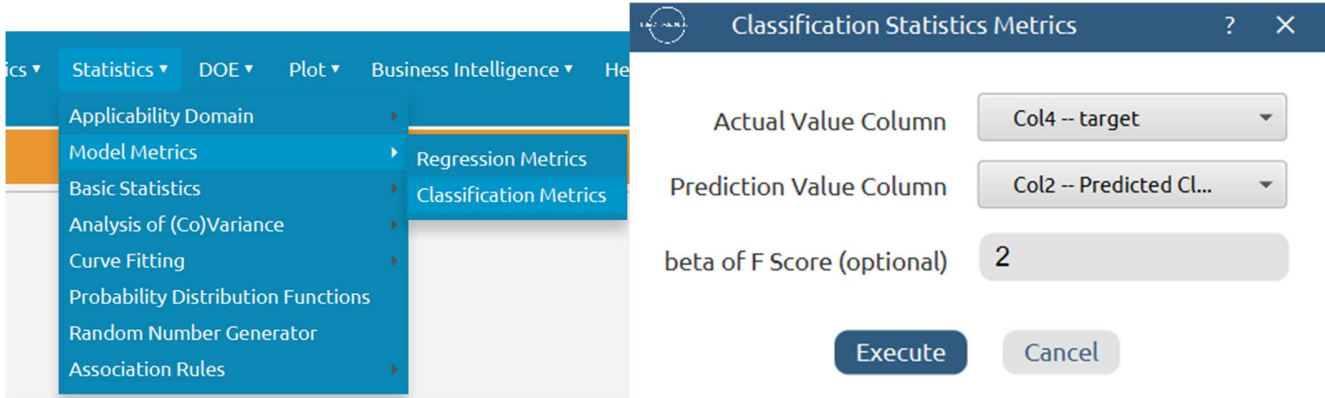
	Col1	Col2 (I)	Col3 (I)
User Header	User Row ID	Prediction	target
1		0	0
2		0	0
3		1	1
4		0	0
5		0	0
6		1	1
7		1	1
8		0	0
9		1	1
10		0	0
11		1	1
12		0	0
13		1	1
14		1	1
15		1	1
16		0	0
17		0	0
18		0	0
19		0	0

Step 19: Statistics calculation – Random Forest

Create a new tab by pressing the “+” button on the bottom of the page with the name “STATISTICS_ACCURACIES_Random Forest”.

Import data into the input spreadsheet of the “STATISTICS_ACCURACIES_Random Forest” tab from the output of the “VALIDATE_MODEL_Random Forest” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Calculate the statistical metrics for the regression: *Statistics → Model Metrics → Classification Metrics*



The results will appear on the output spreadsheet.

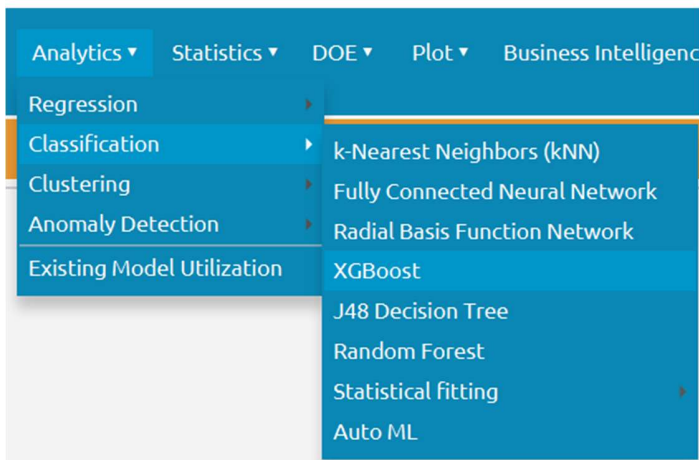
	Col1 (S)	Col2 (D)	Col3 (S)	Col4 (S)
User Header	User Row ID			
1			Predicted Class	Predicted Class
2			0	1
3	Actual Class	0	100	0
4	Actual Class	1	0	105
5				
6				
7	Classification Accuracy	1.0		
8				
9	Precision		1.0	1.0
10				
11	Recall/Sensitivity		1.0	1.0
12				
13	Specificity		1.0	1.0
14				
15	F1 Score		1.0	1.0
16				
17	F (beta=2)		1.0	1.0
18				
19	MCC	1.0		

Step 20: Train the Model – XGBoost

Create a new tab by pressing the “+” button on the bottom of the page with the name “XGBoost”.

Import data into the input spreadsheet of the “XGBoost” tab from the output of the “FEATURE_SELECTION_TRAINING_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Use the Linear SGD method to train and fit the model: *Analytics → Classification → XGBoost*.



XGBoost Classification Model ? X

Target Column: Col10 - target

Booster: gbtree

Objective: multisoftprob

Number of Trees: 4

Learning Rate: 0.01

Gamma: 0.0

Max Tree Depth: 6

Minimum Child Weight: 1.0

Column Sample by tree: 0.8

Subsample: 0.8

Tree Method: auto

Lambda: 1.0

Alpha: 0.0

Time-based RNG Seed RNG Seed: 42

Execute
Cancel

The predictions will appear on the output spreadsheet.

	Col1	Col2 (I)	Col3 (I)
User Header	User Row ID	target	Prediction
1		0	0
2		0	0
3		0	0
4		0	0
5		0	0
6		0	0
7		1	1
8		1	1
9		0	0
10		0	0
11		1	1
12		0	0
13		1	1
14		1	1
15		0	0
16		1	1
17		1	1
18		1	1

Step 21: Validate the Model – XGBoost

Create a new tab by pressing the “+” button on the bottom of the page with the name “VALIDATE_MODEL_Random Forest”.

Import data into the input spreadsheet of the “VALIDATE_MODEL_XGBoost” tab from the output of the “FEATURE_SELECTION_TEST_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

To validate the model: *Analytics* → *Existing Model Utilization* → *Model (from Tab:) XGBoost (.fit)*. Choose the column “target” to be transferred to the output spreadsheet.

The image shows two parts of the software interface. On the left, a portion of the 'Analytics' menu is visible, with 'Existing Model Utilization' highlighted. On the right, the 'Existing Model Execution' dialog box is open. The 'Model' dropdown is set to '(from Tab:)XGBoost' and the 'Type' is 'XGBoost classification mod'. The 'Description' field is empty. The 'Model Input' section lists the following columns and their datatypes: sex (Integer), cp (Integer), restecg (Integer), thalach (Integer), exang (Integer), oldpeak (Double), ca (Integer), and thal (Integer). The 'Transfer Column(s) to Output' checkbox is checked. Below this, there are two lists: 'Excluded Columns' containing Col2 -- sex, Col3 -- cp, Col4 -- restecg, Col5 -- thalach, Col6 -- exang, Col7 -- oldpeak, Col8 -- ca, and Col9 -- thal; and 'Included Columns' containing Col10 -- target. Navigation buttons (>>, >, <, <<) are positioned between the lists. At the bottom of the dialog are 'Execute' and 'Cancel' buttons.

The predictions will appear on the output spreadsheet.

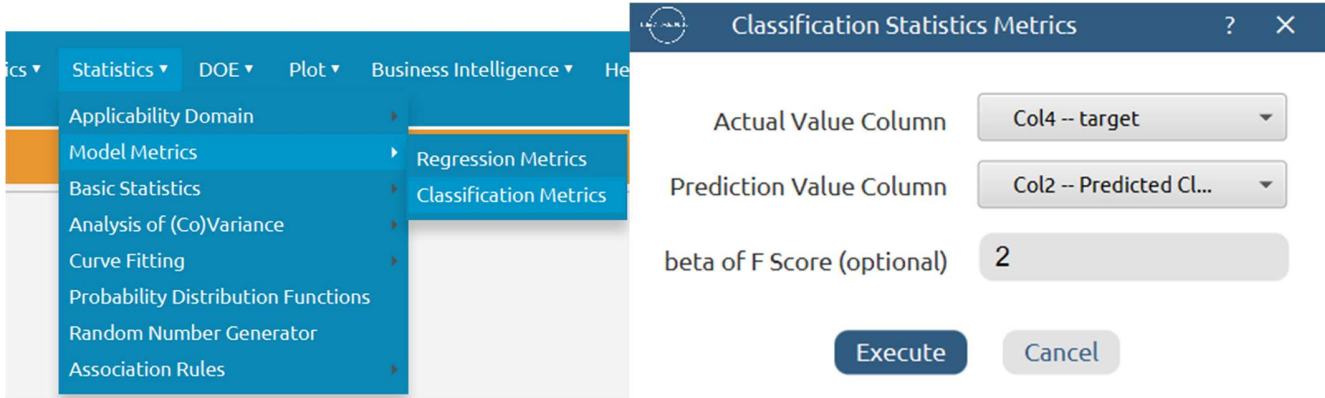
	Col1	Col2 (I)	Col3 (I)
User Header	User Row ID	Prediction	target
1		0	0
2		0	0
3		1	1
4		0	0
5		0	0
6		1	1
7		1	1
8		0	0
9		1	1
10		0	0
11		1	1
12		0	0
13		1	1
14		1	1
15		1	1
16		0	0
17		0	0
18		0	0

Step 22: Statistics calculation – XGBoost

Create a new tab by pressing the “+” button on the bottom of the page with the name “STATISTICS_ACCURACIES_XGBoost”.

Import data into the input spreadsheet of the “STATISTICS_ACCURACIES_XGBoost” tab from the output of the “VALIDATE_MODEL_XGBoost” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Calculate the statistical metrics for the regression: *Statistics → Model Metrics → Classification Metrics*



The results will appear on the output spreadsheet.

	Col1 (S)	Col2 (D)	Col3 (S)	Col4 (S)
User Header	User Row ID			
1			Predicted Class	Predicted Class
2			0	1
3	Actual Class	0	90	10
4	Actual Class	1	6	99
5				
6				
7	Classification Accuracy	0.9219512		
8				
9	Precision		0.9375	0.9082569
10				
11	Recall/Sensitivity		0.9	0.9428571
12				
13	Specificity		0.9428571	0.9
14				
15	F1 Score		0.9183673	0.9252336
16				
17	F (beta=2)		0.9072581	0.9357278
18				
19	MCC	0.8443058		

Step 23: Train the Model- MLP

Create a new tab by pressing the “+” button on the bottom of the page with the name “MLP”.

Import data into the input spreadsheet of the “MLP” tab from the output of the “NORMALIZE_TRAINING_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Use the MLP method to train and fit the model: *Analytics* → *Regression* → *Fully Connected Neural Network*

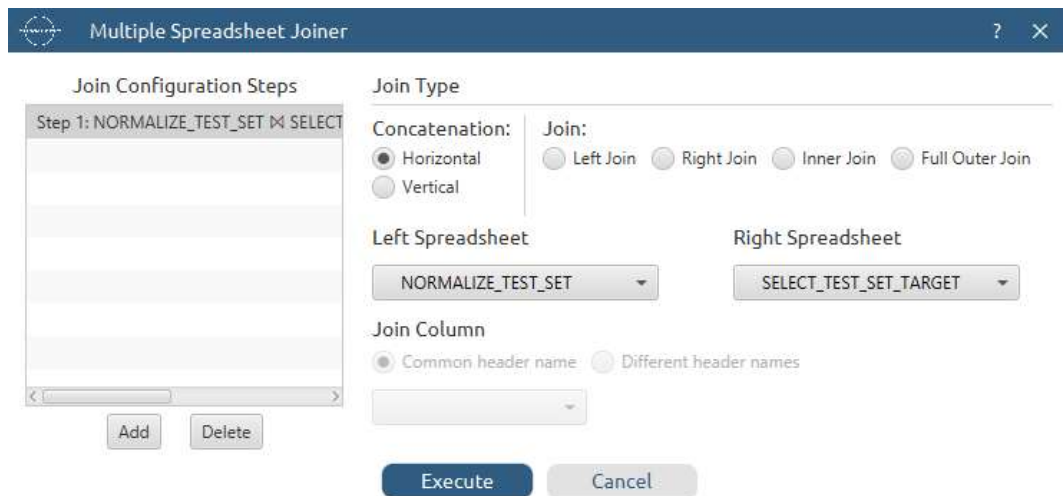
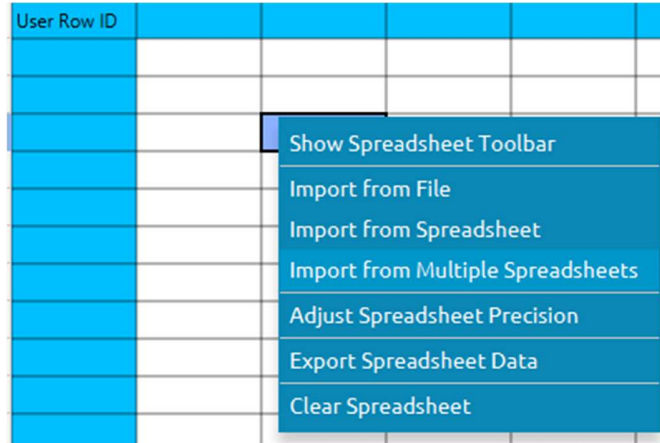
The predictions will appear on the output spreadsheet.

	Col1	Col2 (D)	Col3 (D)
User Header	User Row ID	target	Prediction
1		0.0	0.0
2		0.0	0.0
3		0.0	0.0
4		0.0	0.0
5		0.0	0.0
6		0.0	0.0
7		1.0	1.0
8		1.0	1.0
9		0.0	0.0
10		0.0	1.0
11		1.0	1.0
12		0.0	0.0
13		1.0	1.0
14		1.0	1.0
15		0.0	0.0
16		1.0	0.0
17		1.0	1.0
18		1.0	1.0

Step 24: Validate the Model – MLP

Create a new tab by pressing the “+” button on the bottom of the page with the name “VALIDATE_MODEL_MLP”.

Import data into the input spreadsheet of the “VALIDATE_MODEL_MLP” tab from the output of the “NORMALIZE_TEST_SET” tab and “SELECT_TEST_SET_TARGET” by right-clicking on the input spreadsheet and then choosing “Import from Multiple Spreadsheets”.



To validate the model: *Analytics* → *Existing Model Utilization* → *Model (from Tab:) MLP (.fit)*. Choose the column “target” to be transferred to the output spreadsheet.

Existing Model Execution

Model (from Tab:)MLP

Type MultiLayerPerceptron Mode

Description
It is a Fully Connected Neural Network Model

Model Input

Header	Datatype
sex	Double
cp	Double
restecg	Double
thalach	Double
exang	Double
oldpeak	Double
ca	Double
thal	Double

Transfer Column(s) to Output

Excluded Columns

- Col2 -- sex
- Col3 -- cp
- Col4 -- restecg
- Col5 -- thalach
- Col6 -- exang
- Col7 -- oldpeak
- Col8 -- ca
- Col9 -- thal

Included Columns

- Col10 -- target

Execute Cancel

The predictions will appear on the output spreadsheet.

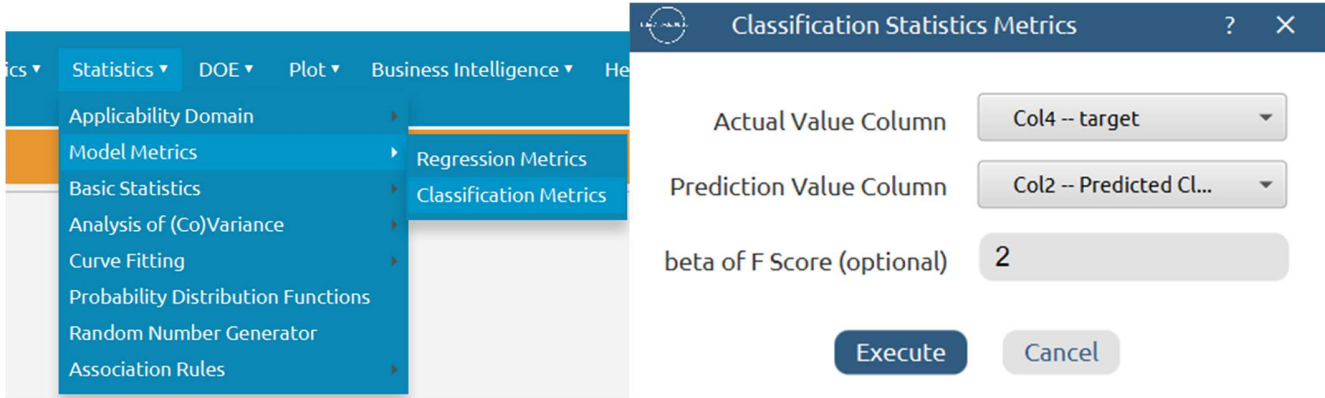
	Col1	Col2 (I)	Col3 (I)
User Header	User Row ID	Prediction	target
1		0	0
2		0	0
3		1	1
4		0	0
5		0	0
6		1	1
7		1	1
8		0	0
9		1	1
10		0	0
11		1	1
12		0	0
13		1	1
14		1	1
15		1	1
16		0	0
17		0	0
18		0	0

Step 25: Statistics calculation – MLP

Create a new tab by pressing the “+” button on the bottom of the page with the name “STATISTICS_ACCURACIES_MLP”.

Import data into the input spreadsheet of the “STATISTICS_ACCURACIES_MLP” tab from the output of the “VALIDATE_MODEL_MLP” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Calculate the statistical metrics for the regression: *Statistics* → *Model Metrics* → *Classification Metrics*



The results will appear on the output spreadsheet.

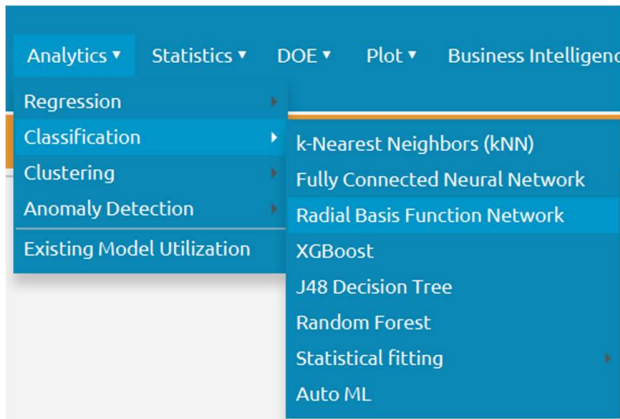
	Col1 (S)	Col2 (D)	Col3 (S)	Col4 (S)
User Header	User Row ID			
1			Predicted Class	Predicted Class
2			0.0	1.0
3	Actual Class	0.0	84	16
4	Actual Class	1.0	19	86
5				
6				
7	Classification Accuracy	0.8292683		
8				
9	Precision		0.8155340	0.8431373
10				
11	Recall/Sensitivity		0.84	0.8190476
12				
13	Specificity		0.8190476	0.84
14				
15	F1 Score		0.8275862	0.8309179
16				
17	F (beta=2)		0.8349901	0.8237548
18				
19	MCC	0.6588594		

Step 26: Train the Model – RBF

Create a new tab by pressing the “+” button on the bottom of the page with the name “RBF”.

Import data into the input spreadsheet of the “RBF” tab from the output of the “NORMALIZE_TRAINING_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Use the RBF method to train and fit the model: *Analytics* → *Classification* → *Radial Basis Function Network*



RBF Network Classification Model
?
×

Hidden Neurons

RBF Kernel

Epsilon

$\varphi(r) = e^{-(\epsilon r)^2}$

Point Selection Random Points from Training Set Use KMeans

Time-based RNG Seed

RNG Seed

Target Column

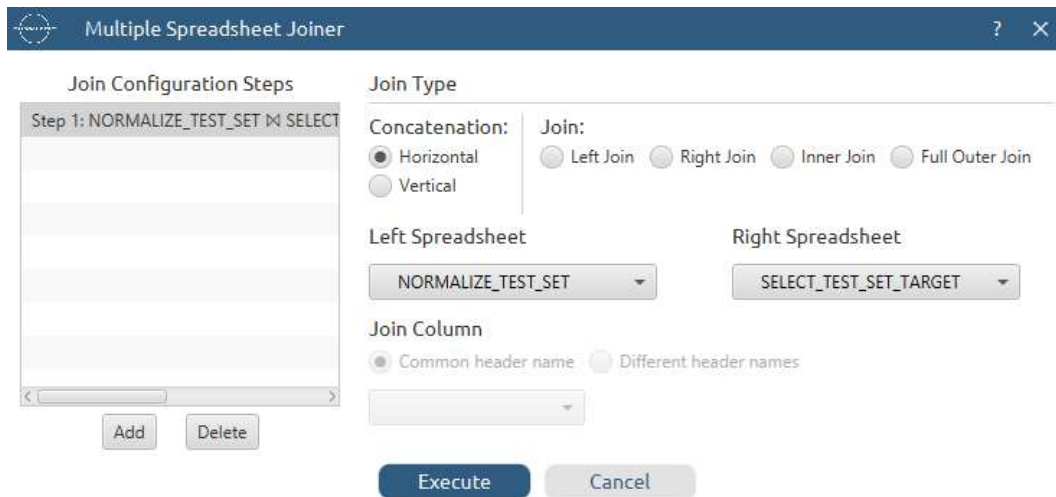
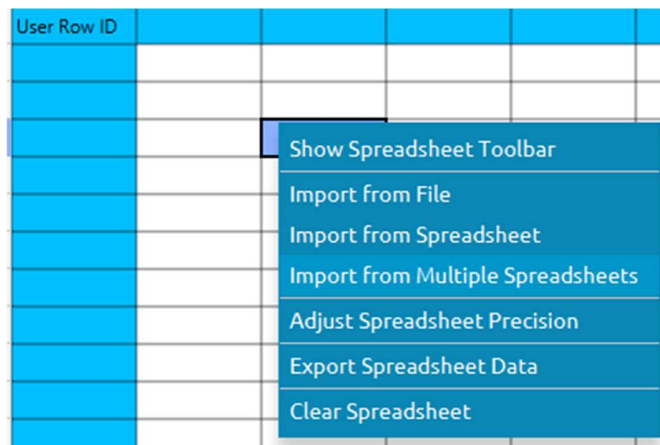
The predictions will appear on the output spreadsheet.

	Col1	Col2 (D)	Col3 (D)
User Header	User Row ID	target	Prediction
1		0.0	0.0
2		0.0	0.0
3		0.0	0.0
4		0.0	0.0
5		0.0	0.0
6		0.0	0.0
7		1.0	1.0
8		1.0	1.0
9		0.0	0.0
10		0.0	0.0
11		1.0	1.0
12		0.0	0.0
13		1.0	1.0
14		1.0	1.0
15		0.0	1.0
16		1.0	0.0
17		1.0	1.0
18		1.0	1.0

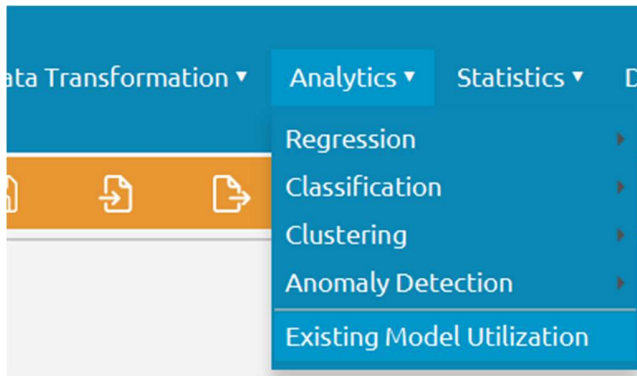
Step 27: Validate the Model – RBF

Create a new tab by pressing the “+” button on the bottom of the page with the name “VALIDATE_MODEL_RBF”.

Import data into the input spreadsheet of the “VALIDATE_MODEL_RBF” tab from the output of the “NORMALIZE_TEST_SET” tab and “SELECT_TEST_SET_TARGET” by right-clicking on the input spreadsheet and then choosing “Import from Multiple Spreadsheets”.



To validate the model: *Analytics* → *Existing Model Utilization* → *Model (from Tab:) RBF (.fit)*. Choose the column “target” to be transferred to the output spreadsheet.



The 'Existing Model Execution' dialog box is shown. It has a title bar with a refresh icon, a question mark, and a close button. The 'Model' dropdown is set to '(from Tab:)RBF'. The 'Type' field is empty. The 'Description' field contains the text: 'It is a Radial Basis Function model for classification'. The 'Model Input' section lists the following variables and their datatypes: sex -> Double, cp -> Double, restecg -> Double, thalach -> Double, exang -> Double, oldpeak -> Double, ca -> Double, and thal -> Double. A checkbox labeled 'Transfer Column(s) to Output' is checked. Below this, there are two columns: 'Excluded Columns' and 'Included Columns'. The 'Excluded Columns' list contains: Col2 -- sex, Col3 -- cp, Col4 -- restecg, Col5 -- thalach, Col6 -- exang, Col7 -- oldpeak, Col8 -- ca, and Col9 -- thal. The 'Included Columns' list contains: Col10 -- target. Between the two columns are four arrow buttons: '>>', '>', '<', and '<<'. At the bottom of the dialog are 'Execute' and 'Cancel' buttons.

The predictions will appear on the output spreadsheet.

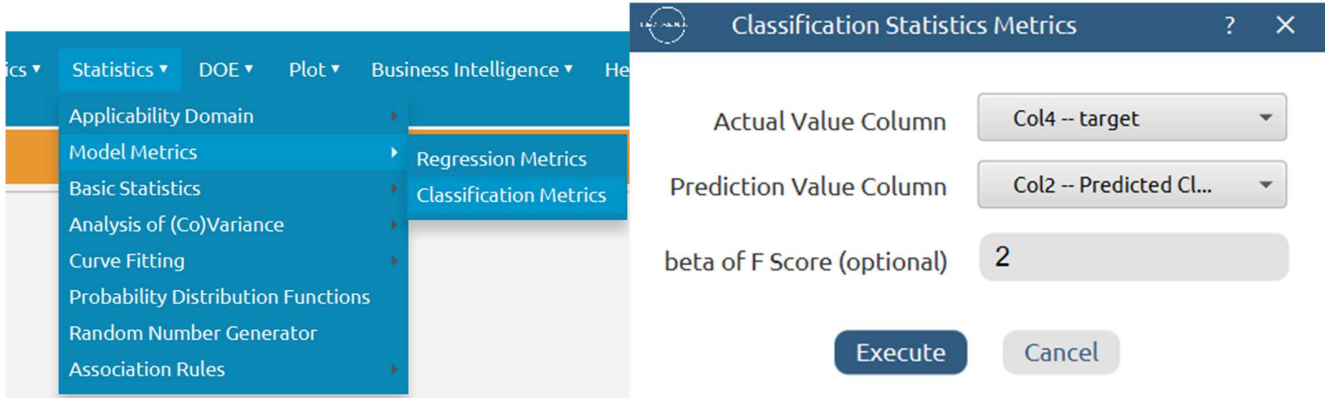
	Col1	Col2 (I)	Col3 (I)
User Header	User Row ID	Prediction	target
1		0	0
2		0	0
3		1	1
4		0	0
5		0	0
6		1	1
7		1	1
8		0	0
9		1	1
10		0	0
11		1	1
12		0	0
13		1	1
14		1	1
15		1	1
16		0	0
17		0	0
18		0	0

Step 28: Statistics calculation – RBF

Create a new tab by pressing the “+” button on the bottom of the page with the name “STATISTICS_ACCURACIES_RBF”.

Import data into the input spreadsheet of the “STATISTICS_ACCURACIES_RBF” tab from the output of the “VALIDATE_MODEL_RBF” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Calculate the statistical metrics for the regression: *Statistics → Model Metrics → Classification Metrics*



The results will appear on the output spreadsheet.

	Col1 (S)	Col2 (D)	Col3 (S)	Col4 (S)
User Header	User Row ID			
1			Predicted Class	Predicted Class
2			0.0	1.0
3	Actual Class	0.0	82	18
4	Actual Class	1.0	15	90
5				
6				
7	Classification Accuracy	0.8390244		
8				
9	Precision		0.8453608	0.8333333
10				
11	Recall/Sensitivity		0.82	0.8571429
12				
13	Specificity		0.8571429	0.82
14				
15	F1 Score		0.8324873	0.8450704
16				
17	F (beta=2)		0.8249497	0.8522727
18				
19	MCC	0.6779181		

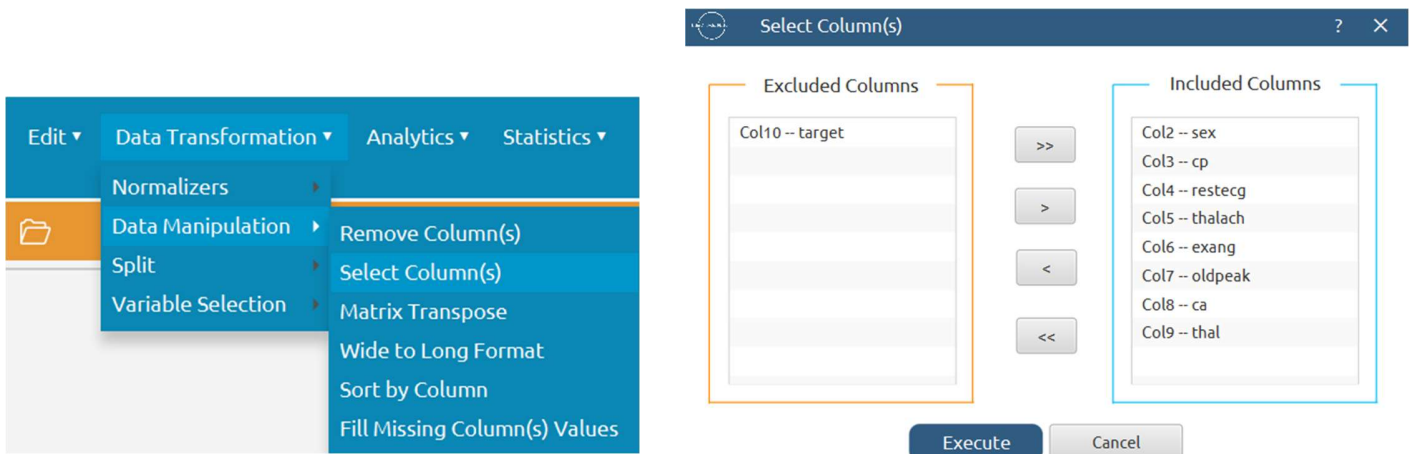
Step 29: Reliability check for each record of the test set

Step 29.a: Create the domain

Create a new tab by pressing the “+” button on the bottom of the page with the name “EXCLUDE_SCORES”.

Import data into the input spreadsheet of the “EXCLUDE_SCORES” tab from the output of the “NORMALIZE_TRAIN_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Manipulate the data to exclude the target column “Performance Index”: *Data Transformation → Data Manipulation → Select Column(s)*

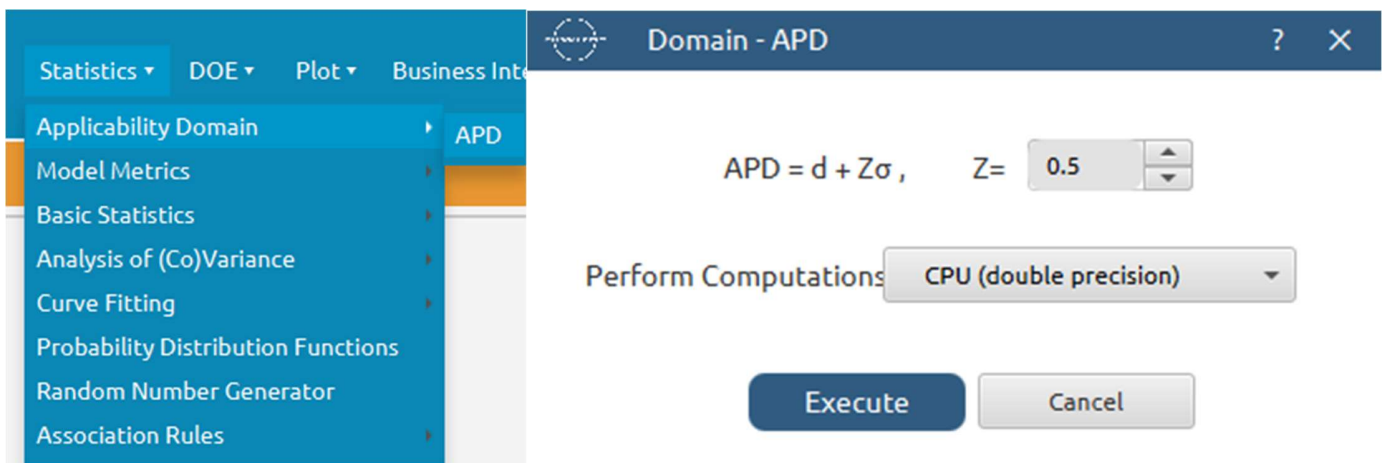


The results will appear on the output spreadsheet.

Create a new tab by pressing the “+” button on the bottom of the page with the name “DOMAIN”.

Import data into the input spreadsheet of the “DOMAIN” tab from the output of the “EXCLUDE_SCORES” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Create the domain: *Statistics → Applicability Domain → APD*



The results will appear on the output spreadsheet.

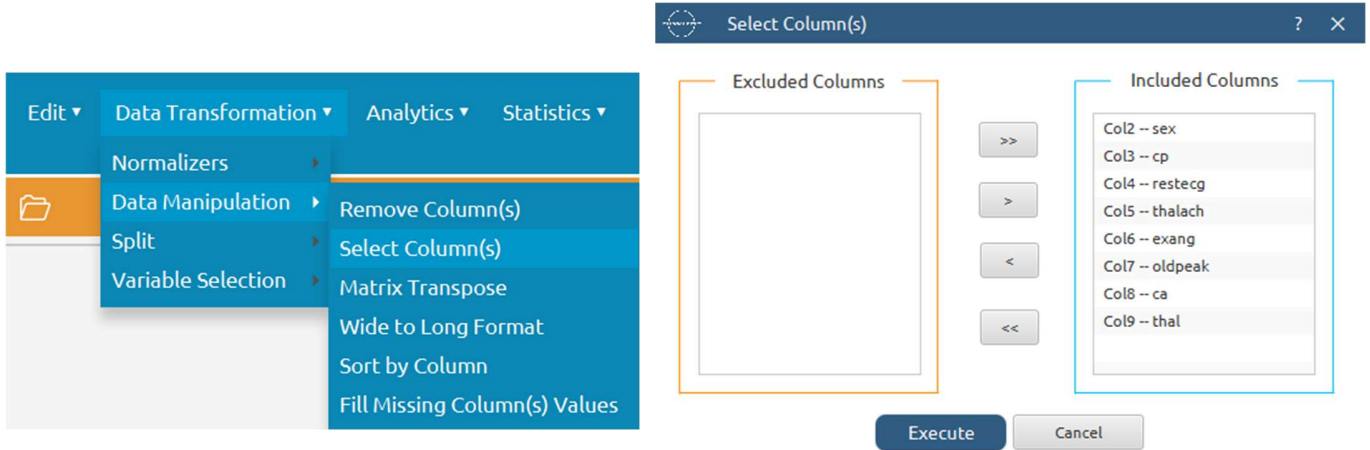
	Col1	Col2 (D)	Col3 (D)	Col4 (S)
User Header	User Row ID	Domain	APD	Prediction
1		0.0	1.0021128	reliable
2		0.0	1.0021128	reliable
3		0.0	1.0021128	reliable
4		0.0	1.0021128	reliable
5		0.0	1.0021128	reliable
6		0.0	1.0021128	reliable
7		0.0	1.0021128	reliable
8		0.0	1.0021128	reliable
9		0.0	1.0021128	reliable
10		0.0	1.0021128	reliable
11		0.0	1.0021128	reliable
12		0.0	1.0021128	reliable
13		0.0	1.0021128	reliable
14		0.0	1.0021128	reliable
15		0.0	1.0021128	reliable
16		0.0	1.0021128	reliable
17		0.0	1.0021128	reliable
18		0.0	1.0021128	reliable

Step 29.b: Check the test set reliability

Create a new tab by pressing the “+” button on the bottom of the page with the name “EXCLUDE_SCORES_TEST_SET”.

Import data into the input spreadsheet of the “EXCLUDE_SCORES_TEST_SET” tab from the output of the “NORMALIZE_TEST_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Manipulate the data to exclude the target column “Performance Index”: *Data Transformation → Data Manipulation → Select Column(s)*



The results will appear on the output spreadsheet.

Create a new tab by pressing the “+” button on the bottom of the page with the name “RELIABILITY”.

Import data into the input spreadsheet of the “RELIABILITY” tab from the output of the “EXCLUDE_SCORES_TEST_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Check the Reliability: *Analytics → Existing Model Utilization → Model (from Tab:) DOMAIN*

Existing Model Execution

Model (from Tab:)DOMAIN

Type APD Model

Description

Model Input

Header -> Datatype
sex -> Double
cp -> Double
restecg -> Double
thalach -> Double
exang -> Double
oldpeak -> Double
ca -> Double
thal -> Double

Transfer Column(s) to Output

Excluded Columns

Included Columns

- Col2 -- sex
- Col3 -- cp
- Col4 -- restecg
- Col5 -- thalach
- Col6 -- exang
- Col7 -- oldpeak
- Col8 -- ca
- Col9 -- thal

Execute Cancel

Transformation ▾ Analytics ▾ Statistics ▾ DC

- Regression
- Classification
- Clustering
- Anomaly Detection
- Existing Model Utilization

The results will appear on the output spreadsheet. There are no unreliable samples in the test set.

	Col1	Col2 (D)	Col3 (D)	Col4 (S)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)	Col10 (D)	Col11 (D)	Col12 (D)
User Header	User Row ID	Domain	APD	Prediction	sex	cp	restecg	thalach	exang	oldpeak	ca	thal
1		0.0	1.0021128	reliable	1.0	0.0	0.0	0.6412214	1.0	0.5	0.0	1.0
2		0.0	1.0021128	reliable	0.0	0.0	0.5	0.2671756	0.0	0.3064516	0.75	0.6666667
3		0.0	1.0021128	reliable	0.0	0.0	0.0	0.3893130	0.0	0.1612903	0.0	0.6666667
4		0.0	1.0021128	reliable	1.0	0.0	1.0	0.5267176	0.0	0.7096774	0.75	0.3333333
5		0.0	1.0021128	reliable	0.0	0.0	0.0	0.4961832	1.0	0.4838710	0.0	1.0
6		0.0	1.0021128	reliable	0.0	0.3333333	0.5	0.9236641	0.0	0.1129032	0.0	0.6666667
7		0.0	1.0021128	reliable	0.0	0.0	0.5	0.5419847	0.0	0.0483871	0.5	0.6666667
8		0.0	1.0021128	reliable	1.0	0.6666667	0.5	0.3129771	1.0	0.4677419	0.25	1.0
9		0.0	1.0021128	reliable	0.0	0.6666667	0.0	0.6183206	0.0	0.0	0.25	0.6666667
10		0.0	1.0021128	reliable	1.0	0.0	0.5	0.6793893	1.0	0.2258065	0.5	1.0
11		0.0	1.0021128	reliable	0.0	1.0	0.0	0.6946565	0.0	0.1612903	0.0	0.6666667
12		0.0	1.0021128	reliable	1.0	0.0	0.5	0.3053435	1.0	0.9032258	0.0	1.0
13		0.0	1.0021128	reliable	1.0	0.3333333	0.0	1.0	0.0	0.0	0.0	0.6666667
14		0.0	1.0021128	reliable	0.0	0.6666667	0.0	0.6183206	0.0	0.0	0.25	0.6666667
15		0.0	1.0021128	reliable	1.0	0.6666667	0.5	0.6564885	0.0	0.2580645	0.0	0.6666667
16		0.0	1.0021128	reliable	1.0	0.0	0.5	0.4122137	1.0	0.2903226	0.0	0.3333333
17		0.1243078	1.0021128	reliable	1.0	0.0	0.0	0.5267176	1.0	0.5483871	0.0	1.0
18		0.0	1.0021128	reliable	1.0	0.0	0.0	0.5572519	1.0	0.1935484	0.25	0.6666667
19		0.0	1.0021128	reliable	1.0	0.0	0.0	0.6030534	1.0	0.0	0.5	1.0
20		0.0	1.0021128	reliable	1.0	0.3333333	0.5	0.6641221	0.0	0.1290323	0.25	0.6666667
21		0.0	1.0021128	reliable	1.0	0.3333333	0.5	0.6641221	0.0	0.1290323	0.25	0.6666667

Final Isalos Workflow

Following the above-described steps, the final workflow on Isalos will look like this:

