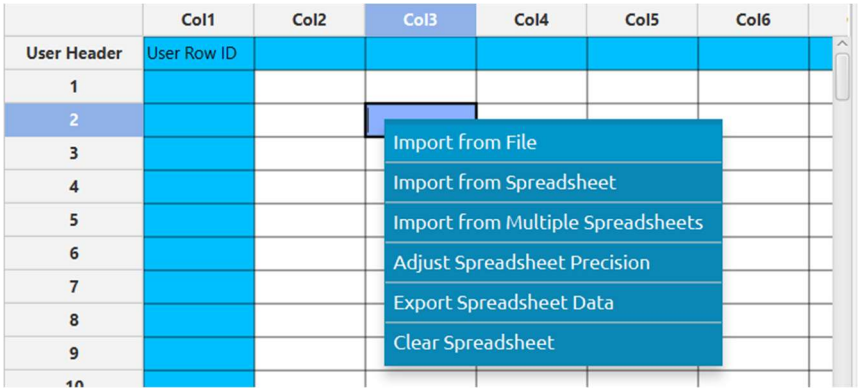
Fish Market Dataset

The csv file contains, which can be found in https://github.com/Ankit152/Fish-Market/blob/main/Fish.csv, records of fish samples from a fish market, where each row represents one fish. The dataset has 7 columns: the first column indicates the Species (e.g., Bream, Roach, Whitefish), while the remaining six are numerical features: Weight, Length1, Length2, Length3, Height, and Width. These physical measurements (in cm and grams) describe the dimensions of the fish and are used to predict the species or weight. The fish market contains 7 fish species and the main task is to predict each fish species.

Isalos version used: 2.0.6

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 fish market data.

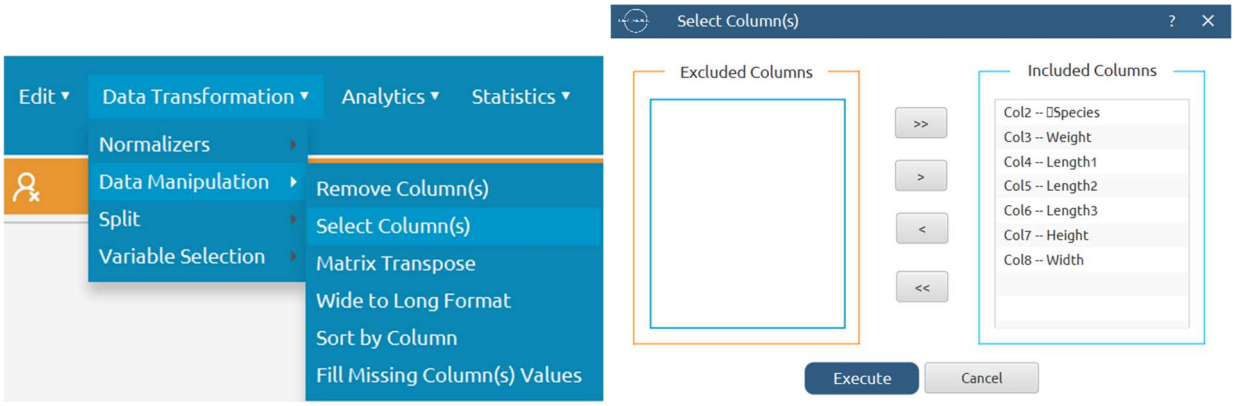


The data will appear on the left spreadsheet.

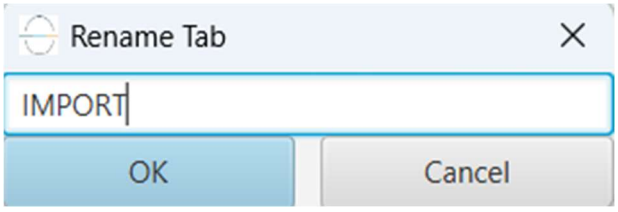
	Col1	Col2 (S)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
User Header	User Row ID	Species	Weight	Length1	Length2	Length3	Height	Width
1		Bream	242	23.2	25.4	30	11.52	4.02
2		Bream	290	24	26.3	31.2	12.48	4.3056
3		Bream	340	23.9	26.5	31.1	12.3778	4.6961
4		Bream	363	26.3	29	33.5	12.73	4.4555
5		Bream	430	26.5	29	34	12.444	5.134
6		Bream	450	26.8	29.7	34.7	13.6024	4.9274
7		Bream	500	26.8	29.7	34.5	14.1795	5.2785
8		Bream	390	27.6	30	35	12.67	4.69
9		Bream	450	27.6	30	35.1	14.0049	4.8438
10		Bream	500	28.5	30.7	36.2	14.2266	4.9594
11		Bream	475	28.4	31	36.2	14.2628	5.1042
12		Bream	500	28.7	31	36.2	14.3714	4.8146
13		Bream	500	29.1	31.5	36.4	13.7592	4.368
14		Bream	340	29.5	32	37.3	13.9129	5.0728
15		Bream	600	29.4	32	37.2	14.9544	5.1708

Step 2: Manipulate data

In our dataset there are not any empty values, 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.



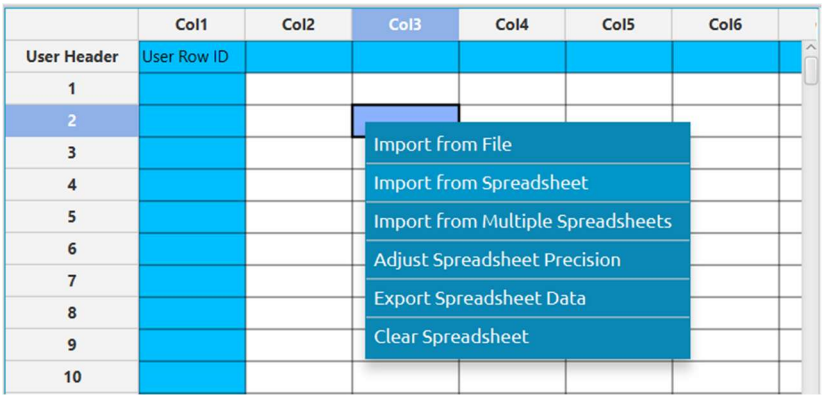
All of the data will appear in the output (right) spreadsheet. This tab can be renamed “IMPORT” by right-clicking on it and choosing the “Rename” option.



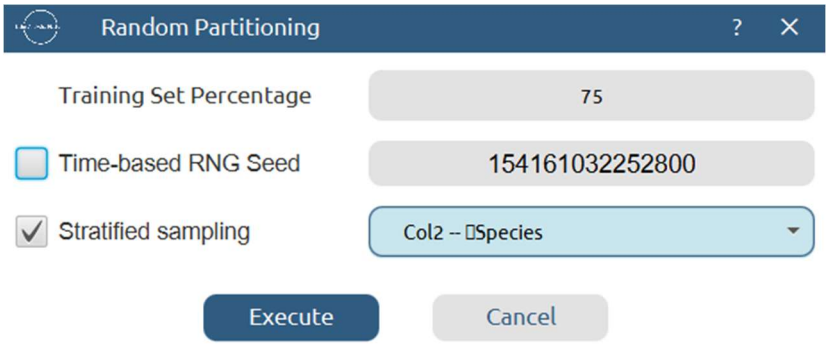
Step 3: Split data

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.

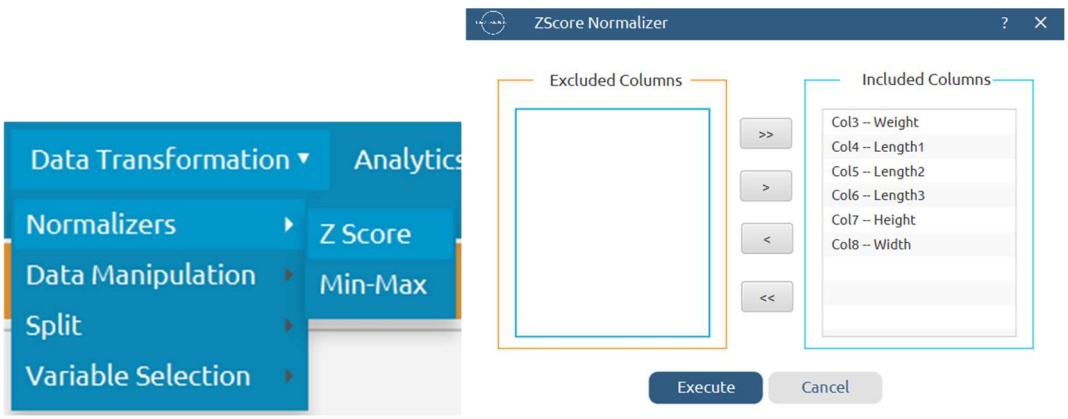
Step 4: Normalize the training set

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 "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".

	Col1	Col2 (S)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
User Header	User Row ID	Species	Weight	Length1	Length2	Length3	Height	Width
1		Bream	242	23.2	25.4	30	11.52	4.02
2		Bream	340	23.9	26.5	31.1	12.3778	4.6961
3		Bream	363	26.3	29	33.5	12.73	4.4555
4		Bream	450	26.8	29.7	34.7	13.6024	4.9274
5		Bream	500	26.8	29.7	34.5	14.1795	5.2785
6		Bream	390	27.6	30	35	12.67	4.69
7		Bream	500	28.5	30.7	36.2	14.2266	4.9594
8		Bream	475	28.4	31	36.2	14.2628	5.1042
9		Bream	500	28.7	31	36.2	14.3714	4.8146
10		Bream	500	29.1	31.5	36.4	13.7592	4.368
11		Bream	340	29.5	32	37.3	13.9129	5.0728
12		Bream	600	29.4	32	37.2	15.438	5.58
13		Bream	610	30.9	33.5	38.6	15.633	5.1338
14		Bream	650	31	33.5	38.7	14.4738	5.7276
15		Bream	685	31.4	34	39.2	15.9936	5.3704

Normalize the data using Z-score: *Data Transformation → Normalizers → Z Score* and select all columns except the "Species" target column.



The results will appear on the output spreadsheet.

	Col1	Col2 (S)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
User Header	User Row ID	Species	Weight	Length1	Length2	Length3	Height	Width
1		Bream	-0.4510066	-0.3051448	-0.2823356	-0.1107788	0.5702414	-0.2475009
2		Bream	-0.1921738	-0.2371443	-0.1826922	-0.0187677	0.7660333	0.1351813
3		Bream	-0.1314273	-0.0040000	0.0437700	0.1819838	0.8464225	-0.0010017
4		Bream	0.0983529	0.0445717	0.1071794	0.2823595	1.0455468	0.2661004
5		Bream	0.2304105	0.0445717	0.1071794	0.2656302	1.1772692	0.4648279
6		Bream	-0.0601162	0.1222865	0.1343549	0.3074535	0.8327276	0.1317286
7		Bream	0.2304105	0.2097156	0.1977643	0.4078292	1.1880197	0.2842128
8		Bream	0.1643817	0.2000013	0.2249398	0.4078292	1.1962823	0.3661717
9		Bream	0.2304105	0.2291443	0.2249398	0.4078292	1.2210702	0.2022540
10		Bream	0.2304105	0.2680017	0.2702322	0.4245585	1.0813362	-0.0505279
11		Bream	-0.1921738	0.3068591	0.3155247	0.4998403	1.1164181	0.3483988
12		Bream	0.4945257	0.2971447	0.3155247	0.4914757	1.4645204	0.6354812
13		Bream	0.5209372	0.4428599	0.4514020	0.6085807	1.5090289	0.3829257
14		Bream	0.6265833	0.4525743	0.4514020	0.6169454	1.2444429	0.7190249
15		Bream	0.7190236	0.4914317	0.4966944	0.6587686	1.5913355	0.5168446

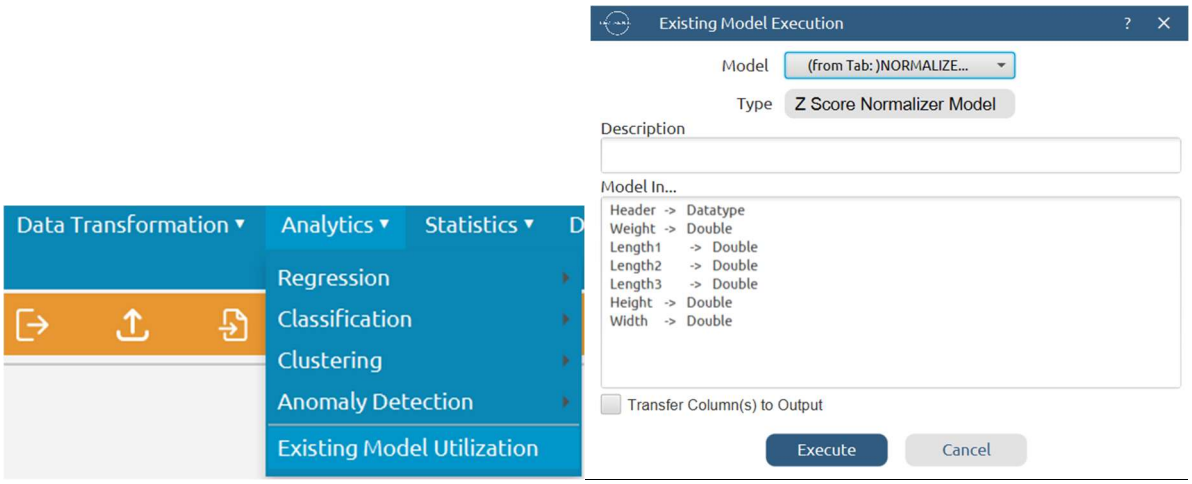
Step 5: Normalize the test set

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 test set 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".

	Col1	Col2 (S)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
User Header	User Row ID	Species	Weight	Length1	Length2	Length3	Height	Width
1		Bream	290	24	26.3	31.2	12.48	4.3056
2		Bream	430	26.5	29	34	12.444	5.134
3		Bream	450	27.6	30	35.1	14.0049	4.8438
4		Bream	600	29.4	32	37.2	14.9544	5.1708
5		Bream	700	30.4	33	38.3	14.8604	5.2854
6		Bream	700	30.4	33	38.5	14.938	5.1975
7		Bream	575	31.3	34	39.5	15.1285	5.5695
8		Bream	714	32.7	36	41.5	16.517	5.8515
9		Bream	955	35	38.5	44	18.084	6.292
10		Roach	40	12.9	14.1	16.2	4.1472	2.268
11		Roach	120	19.4	21	23.7	6.1146	3.2943
12		Roach	145	20.5	22	24.3	6.6339	3.5478
13		Roach	161	22	23.4	26.7	6.9153	3.6312
14		Roach	272	25	27	30.6	8.568	4.7736
15		Whitefish	270	23.6	26	28.7	8.3804	4.2476

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.

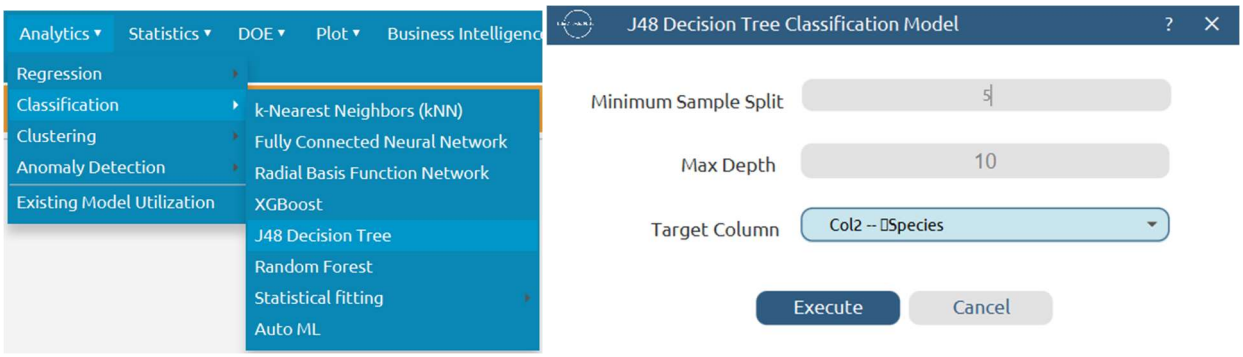
	Col1	Col2 (S)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
User Header	User Row ID	Species	Weight	Length1	Length2	Length3	Height	Width
1		Bream	-0.3242313	-0.2274300	-0.2008092	-0.0104031	0.7893603	-0.0858472
2		Bream	0.0455299	0.0154287	0.0437700	0.2238070	0.7811433	0.3830389
3		Bream	0.0983529	0.1222865	0.1343549	0.3158181	1.1374170	0.2187816
4		Bream	0.4945257	0.2971447	0.3155247	0.4914757	1.3541393	0.4038682
5		Bream	0.7586409	0.3942882	0.4061095	0.5834868	1.3326839	0.4687334
6		Bream	0.7586409	0.3942882	0.4061095	0.6002161	1.3503960	0.4189808
7		Bream	0.4284969	0.4817173	0.4966944	0.6838626	1.3938774	0.6295380
8		Bream	0.7956170	0.6177182	0.6778642	0.8511555	1.7108009	0.7891540
9		Bream	1.4321346	0.8411481	0.9043264	1.0602716	2.0684669	1.0384832
10		Roach	-0.9845193	-1.3057225	-1.3059447	-1.2651000	-1.1125917	-1.2391576
11		Roach	-0.7732272	-0.6742899	-0.6809091	-0.6377516	-0.6635349	-0.6582573
12		Roach	-0.7071984	-0.5674321	-0.5903242	-0.5875637	-0.5450053	-0.5147727
13		Roach	-0.6649399	-0.4217169	-0.4635053	-0.3868122	-0.4807761	-0.4675671
14		Roach	-0.3717721	-0.1302865	-0.1373998	-0.0605910	-0.1035492	0.1790474
15		Whitefish	-0.3770544	-0.2662874	-0.2279846	-0.2195192	-0.1463687	-0.1186760

Step 6: Train the model

Create a new tab by pressing the "+" button on the bottom of the page with the name "TRAIN_MODEL(.fit)".

Import data into the input spreadsheet of the "TRAIN_MODEL(.fit)" 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 J48 Decision tree method to train and fit the model: *Analytics → Classification → J48 Decision Tree*



The predictions will appear on the output spreadsheet.

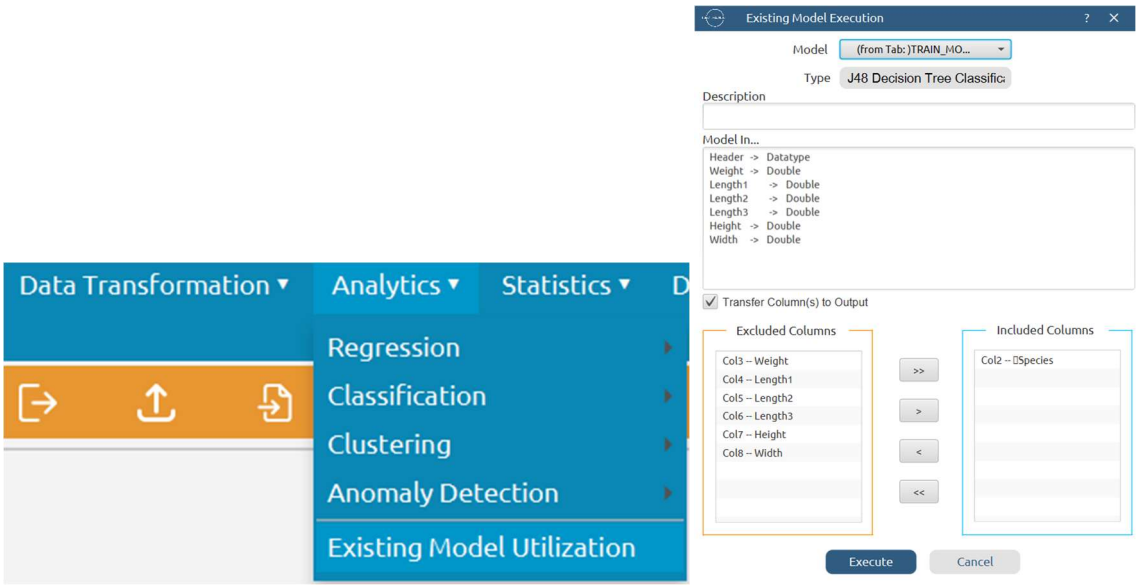
	Col1	Col2 (S)	Col3 (S)
User Header	User Row ID	Species	Prediction
1		Bream	Bream
2		Bream	Bream
3		Bream	Bream
4		Bream	Bream
5		Bream	Bream
6		Bream	Bream
7		Bream	Bream
8		Bream	Bream
9		Bream	Bream
10		Bream	Bream
11		Bream	Bream
12		Bream	Bream
13		Bream	Bream
14		Bream	Bream
15		Bream	Bream

Step 7: Validate the model

Create a new tab by pressing the "+" button on the bottom of the page with the name "VALIDATE_MODEL(.predict)".

Import data into the input spreadsheet of the "VALIDATE_MODEL(.predict)" tab from the output of the "NORMALIZE_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:) TRAIN_MODEL(.fit)*. Choose the column "Species" to be transferred to the output spreadsheet.



The predictions will appear on the output spreadsheet.

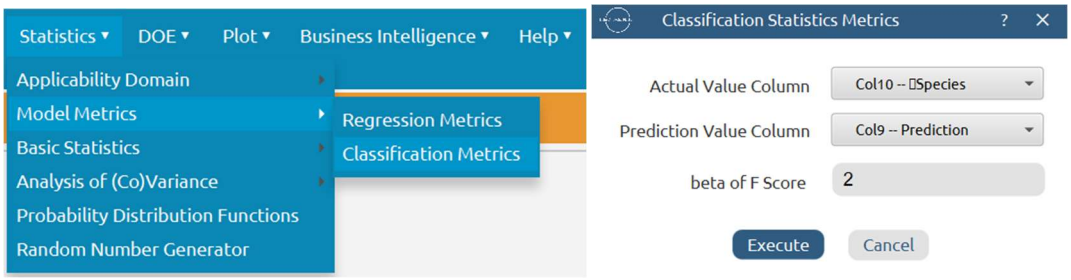
	Col9 (S)	Col10 (S)
User Header	Prediction	Species
1	Bream	Bream
2	Perch	Bream
3	Bream	Bream
4	Bream	Bream
5	Bream	Bream
6	Bream	Bream
7	Bream	Bream
8	Bream	Bream
9	Bream	Bream
10	Perch	Roach
11	Perch	Roach
12	Perch	Roach
13	Perch	Roach
14	Whitefish	Roach
15	Parkki	Whitefish

Step 8: Statistics calculation

Create a new tab by pressing the "+" button on the bottom of the page with the name "STATISTICS_ACCURACIES".

Import data into the input spreadsheet of the "STATISTICS_ACCURACIES" tab from the output of the "VALIDATE_MODEL(.predict)" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet".

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



The results will appear on the output spreadsheet.

	Col1 (S)	Col2 (S)	Col3 (S)	Col4 (S)	Col5 (S)	Col6 (S)	Col7 (S)	Col8 (S)	Col9 (S)
User Header	User Row ID								
1			Predicted Class	Predicted Class	Predicted Class	Predicted Class	Predicted Class	Predicted Class	Predicted Class
2			Bream	Roach	Whitefish	Parkki	Perch	Pike	Smelt
3	Actual Class	Bream	8	0	0	0	1	0	0
4	Actual Class	Roach	0	0	1	0	4	0	0
5	Actual Class	Whitefish	0	0	0	1	0	0	0
6	Actual Class	Parkki	0	0	0	3	0	0	0
7	Actual Class	Perch	0	1	2	0	11	0	0
8	Actual Class	Pike	0	0	0	0	1	4	0
9	Actual Class	Smelt	0	0	0	0	0	0	3
10									
11									
12	Classification Accuracy	0.725							
13									
14	Precision		1.0	0.0	0.0	0.75	0.6470588	1.0	1.0
15									
16	Recall/Sensitivity		0.8888889	0.0	0.0	1.0	0.7857143	0.8	1.0
17									
18	Specificity		1.0	0.9714286	0.9230769	0.9729730	0.7692308	1.0	1.0
19									
20	F1 Score		0.9411765	NaN	NaN	0.8571429	0.7096774	0.8888889	1.0
21									
22	F (beta=2)		0.9090909	0.0	0.0	0.9375	0.7534247	0.8333333	1.0
23									
24	MCC	0.6540609							

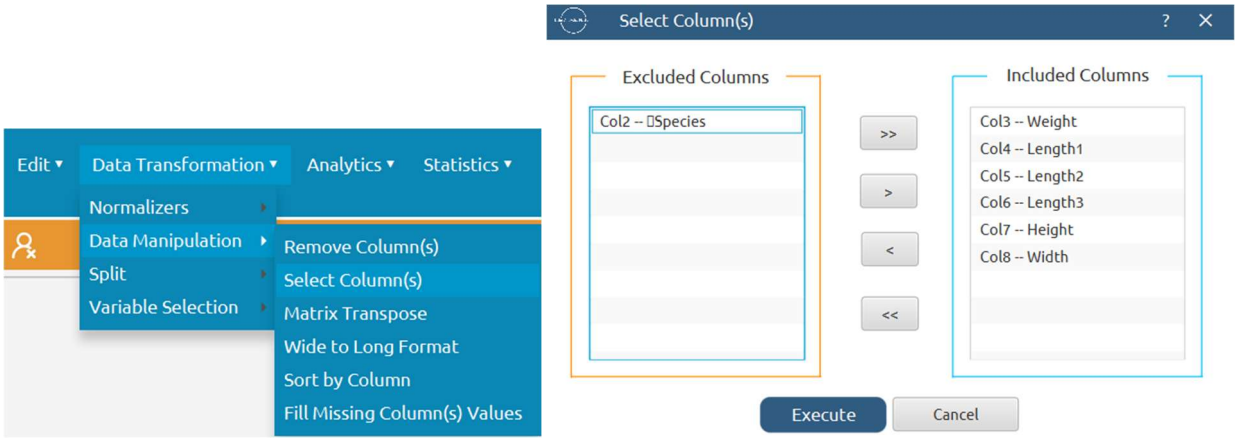
Step 9: Reliability check for each record of the test set

Step 9.a: Create the domain

Create a new tab by pressing the "+" button on the bottom of the page with the name "EXCLUDE_SPECIES".

Import data into the input spreadsheet of the "EXCLUDE_SPECIES" 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 "Species": *Data Transformation → Data Manipulation → Select Column(s)*

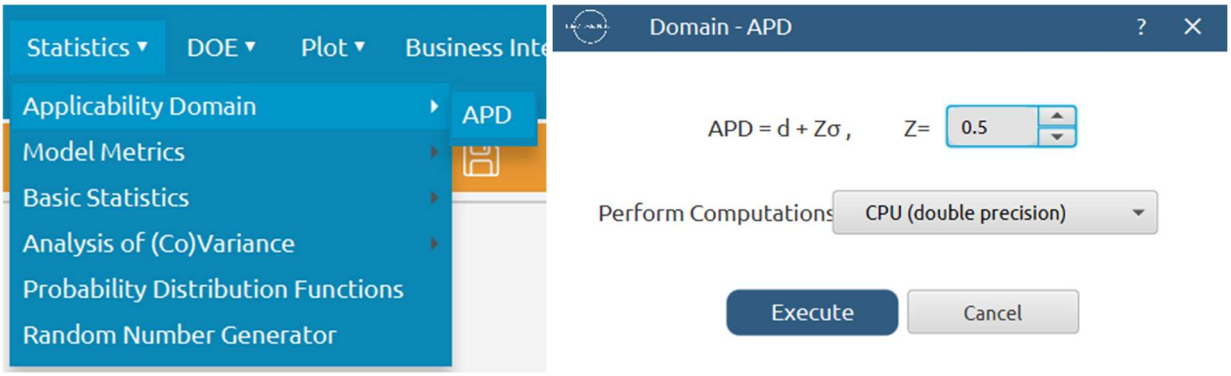


The results will appear on the output spreadsheet.

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_SPECIES" 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.

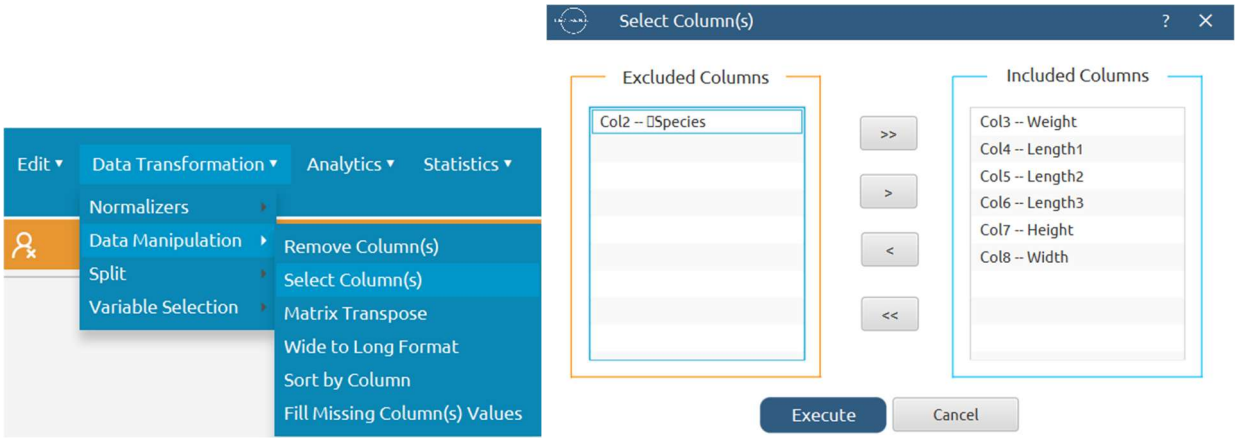
	Col1	Col2 (D)	Col3 (D)	Col4 (S)
User Header	User Row ID	Domain	APD	Prediction
1		0.0	1.9984306	reliable
2		0.0	1.9984306	reliable
3		0.0	1.9984306	reliable
4		0.0	1.9984306	reliable
5		0.0	1.9984306	reliable
6		0.0	1.9984306	reliable
7		0.0	1.9984306	reliable
8		0.0	1.9984306	reliable
9		0.0	1.9984306	reliable
10		0.0	1.9984306	reliable
11		0.0	1.9984306	reliable
12		0.0	1.9984306	reliable
13		0.0	1.9984306	reliable
14		0.0	1.9984306	reliable
15		0.0	1.9984306	reliable

Step 9.b: Check the test set reliability

Create a new tab by pressing the "+" button on the bottom of the page with the name "EXCLUDE_SPECIES_TEST_SET".

Import data into the input spreadsheet of the "EXCLUDE_SPECIES_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 "Species": *Data Transformation → Data Manipulation → Select Column(s)*

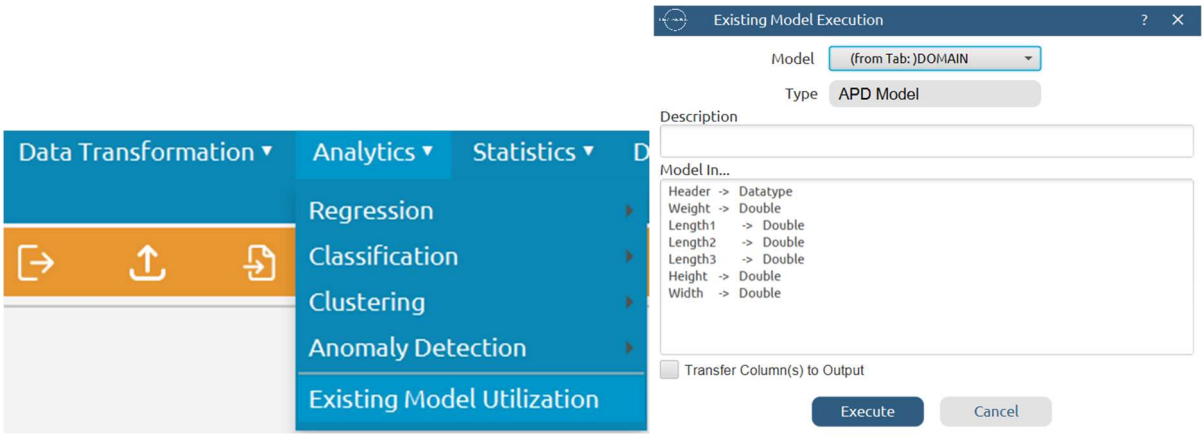


The results will appear on the output spreadsheet.

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_SPECIES_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*



The results will appear on the output spreadsheet.

	Col1	Col2 (D)	Col3 (D)	Col4 (S)
User Header	User Row ID	Domain	APD	Prediction
1		0.2594793	1.9984306	reliable
2		0.3076891	1.9984306	reliable
3		0.1362966	1.9984306	reliable
4		0.2565708	1.9984306	reliable
5		0.3067486	1.9984306	reliable
6		0.2957161	1.9984306	reliable
7		0.2275127	1.9984306	reliable
8		0.1831124	1.9984306	reliable
9		0.0937884	1.9984306	reliable
10		0.1692011	1.9984306	reliable
11		0.0885149	1.9984306	reliable
12		0.0775753	1.9984306	reliable
13		0.1960879	1.9984306	reliable
14		0.1579740	1.9984306	reliable
15		0.0990292	1.9984306	reliable

Final Isalos Workflow

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