



Body Fat Prediction Dataset

The goal of this study is to train a model in order to predict the percentage of bodyfat in men. The dataset, which can be found in <https://www.kaggle.com/datasets/fedesoriano/body-fat-prediction-dataset>, has 14 features and 251 samples. It contains information on certain factors like density determined from underwater weighing, age, weight, height, neck circumference, chest circumference, abdomen circumference, hip circumference, thigh circumference, knee circumference, ankle circumference, biceps (extended) circumference, forearm circumference and wrist circumference. The dataset contains no missing values and also all features are numerical.

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 body fat data.

The screenshot shows a web-based spreadsheet interface. The spreadsheet has columns labeled Col1 through Col6 and rows labeled 1 through 10. A right-click context menu is open over the spreadsheet, displaying the following options: "Import from File", "Import from Spreadsheet", "Import from Multiple Spreadsheets", "Adjust Spreadsheet Precision", "Export Spreadsheet Data", and "Clear Spreadsheet". The "Import from File" option is highlighted.

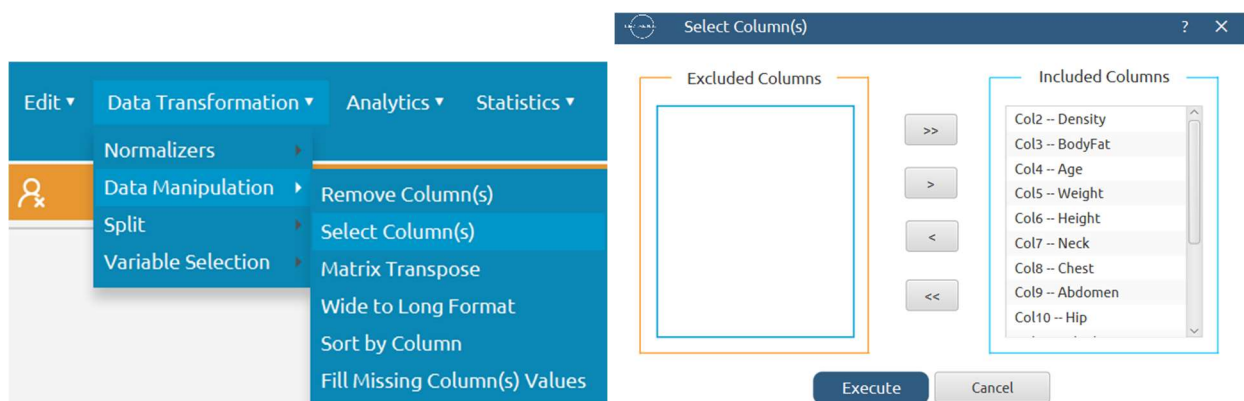
	Col1	Col2	Col3	Col4	Col5	Col6
User Header	User Row ID					
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						

The data will appear on the left spreadsheet.

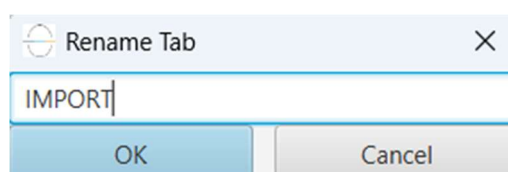
	Col1	Col2 (D)	Col3 (D)	Col4 (I)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
User Header	User Row ID	Density	BodyFat	Age	Weight	Height	Neck	Chest
1		1.0708	12.3	23	154.25	67.75	36.2	93.1
2		1.0853	6.1	22	173.25	72.25	38.5	93.6
3		1.0414	25.3	22	154.0	66.25	34.0	95.8
4		1.0751	10.4	26	184.75	72.25	37.4	101.8
5		1.034	28.7	24	184.25	71.25	34.4	97.3
6		1.0502	20.9	24	210.25	74.75	39.0	104.5
7		1.0549	19.2	26	181.0	69.75	36.4	105.1
8		1.0704	12.4	25	176.0	72.5	37.8	99.6
9		1.09	4.1	25	191.0	74.0	38.1	100.9
10		1.0722	11.7	23	198.25	73.5	42.1	99.6
11		1.083	7.1	26	186.25	74.5	38.5	101.5
12		1.0812	7.8	27	216.0	76.0	39.4	103.6
13		1.0513	20.8	32	180.5	69.5	38.4	102.0
14		1.0505	21.2	30	205.25	71.25	39.4	104.1
15		1.0484	22.1	35	187.75	69.5	40.5	101.3
16		1.0512	20.9	35	162.75	66.0	36.4	99.1
17		1.0333	29.0	34	195.75	71.0	38.9	101.9
18		1.0468	22.9	32	209.25	71.0	42.1	107.6
19		1.0622	16.0	28	183.75	67.75	38.0	106.8
20		1.061	16.5	33	211.75	73.5	40.0	106.2

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.



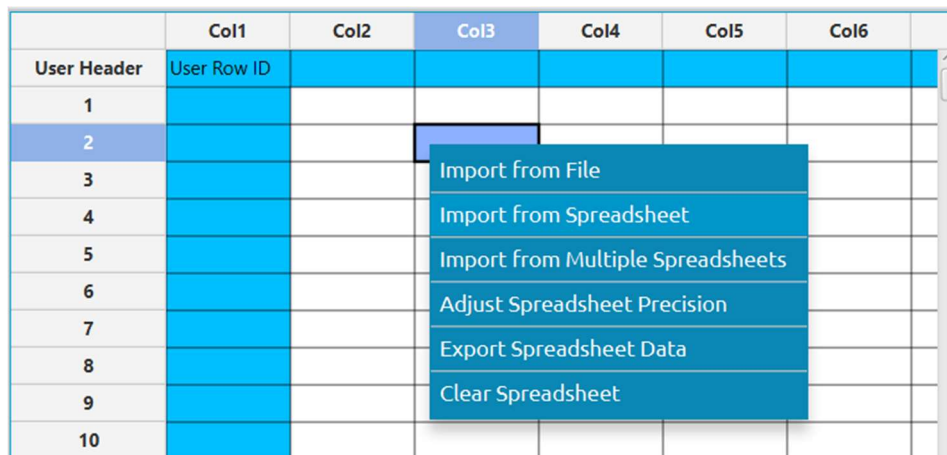
All of the data will appear in the output (right) spreadsheet. This tab can be renamed “IMPORT” by right-clicking on it and choosing the “Rename” option.



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.

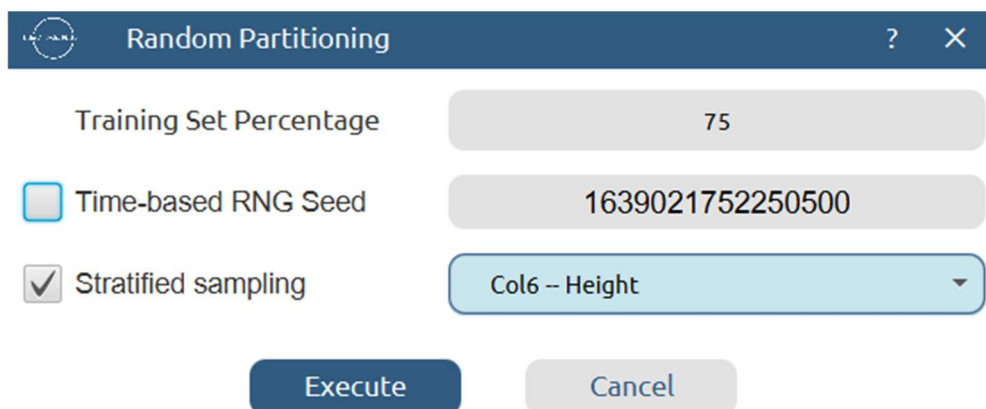
Import data into the input spreadsheet of the “TRAIN_TEST_SPLIT” tab from the output of the “IMPORT” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.



The screenshot shows a spreadsheet with 7 columns (Col1 to Col6) and 11 rows (User Header to 10). A context menu is open over cell C3, listing the following options: 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
User Header	User Row ID					
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						

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 'Random Partitioning' dialog box contains the following settings:

- Training Set Percentage: 75
- ☐ Time-based RNG Seed: 1639021752250500
- ☒ Stratified sampling: Col6 -- Height

Buttons: Execute, Cancel

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 (D)	Col3 (D)	Col4 (I)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
User Header	User Row ID	Density	BodyFat	Age	Weight	Height	Neck	Chest
1		1.0708	12.3	23	154.25	67.75	36.2	93.1
2		1.0853	6.1	22	173.25	72.25	38.5	93.6
3		1.0414	25.3	22	154.0	66.25	34.0	95.8
4		1.034	28.7	24	184.25	71.25	34.4	97.3
5		1.0502	20.9	24	210.25	74.75	39.0	104.5
6		1.09	4.1	25	191.0	74.0	38.1	100.9
7		1.0722	11.7	23	198.25	73.5	42.1	99.6
8		1.083	7.1	26	186.25	74.5	38.5	101.5
9		1.0513	20.8	32	180.5	69.5	38.4	102.0
10		1.0512	20.9	35	162.75	66.0	36.4	99.1
11		1.0333	29.0	34	195.75	71.0	38.9	101.9
12		1.0468	22.9	32	209.25	71.0	42.1	107.6
13		1.0622	16.0	28	183.75	67.75	38.0	106.8
14		1.061	16.5	33	211.75	73.5	40.0	106.2
15		1.0551	19.1	28	179.0	68.0	39.1	103.3
16		1.064	15.2	28	200.5	69.75	41.3	111.4
17		1.0668	14.0	28	151.25	67.75	34.5	90.2
18		1.0468	22.9	31	148.0	67.5	38.8	97.4
19		1.079	8.8	29	160.75	69.0	36.7	97.4
20		1.0716	11.9	32	182.0	73.75	38.7	100.5

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

ZScore Normalizer
?
X

Excluded Columns

Col3 -- BodyFat

Included Columns

Col2 -- Density
Col4 -- Age
Col5 -- Weight
Col6 -- Height
Col7 -- Neck
Col8 -- Chest
Col9 -- Abdomen
Col10 -- Hip
Col11 -- Thigh

>>
>
<
<<

Execute
Cancel

The results will appear on the output spreadsheet.

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
User Header	User Row ID	Density	BodyFat	Age	Weight	Height	Neck	Chest
1		0.8093508	12.3	-1.7699357	-0.8540362	-0.9707136	-0.7873960	-0.9451989
2		1.5525574	6.1	-1.8486829	-0.2381360	0.7124504	0.1260842	-0.8882411
3		-0.6975647	25.3	-1.8486829	-0.8621401	-1.5317683	-1.6611598	-0.6376265
4		-1.0768564	28.7	-1.6911886	0.1184377	0.3384139	-1.5022936	-0.4667530
5		-0.2465152	20.9	-1.6911886	0.9612484	1.6475415	0.3246669	0.3534400
6		1.7934589	4.1	-1.6124414	0.3372443	1.3670141	-0.0327819	-0.0566565
7		0.8811087	11.7	-1.7699357	0.5722589	1.1799959	1.5558794	-0.2047469
8		1.4346695	7.1	-1.5336943	0.1832693	1.5540324	0.1260842	0.0116929
9		-0.1901340	20.8	-1.0612115	-0.0031215	-0.3161498	0.0863677	0.0686508
10		-0.1952595	20.9	-0.8249700	-0.5785019	-1.6252774	-0.7079630	-0.2617047
11		-1.1127353	29.0	-0.9037172	0.4912194	0.2449048	0.2849504	0.0572592
12		-0.4207843	22.9	-1.0612115	0.9288326	0.2449048	1.5558794	0.7065787
13		0.3685524	16.0	-1.3762000	0.1022298	-0.9707136	-0.0724984	0.6154461
14		0.3070456	16.5	-0.9824643	1.0098721	1.1799959	0.7218322	0.5470967
15		0.0046374	19.1	-1.3762000	-0.0517452	-0.8772045	0.3643834	0.2167412
16		0.4608125	15.2	-1.3762000	0.6451944	-0.2226407	1.2381472	1.1394583
17		0.6043283	14.0	-1.3762000	-0.9512836	-0.9707136	-1.4625771	-1.2755544
18		-0.4207843	22.9	-1.1399586	-1.0566349	-1.0642227	0.2452338	-0.4553614
19		1.2296469	8.8	-1.2974529	-0.6433335	-0.5031681	-0.5888134	-0.4553614
20		0.8503553	11.9	-1.0612115	0.0455022	1.2735050	0.2055173	-0.1022228

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 (D)	Col3 (D)	Col4 (I)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
User Header	User Row ID	Density	BodyFat	Age	Weight	Height	Neck	Chest
1		1.0751	10.4	26	184.75	72.25	37.4	101.8
2		1.0549	19.2	26	181.0	69.75	36.4	105.1
3		1.0704	12.4	25	176.0	72.5	37.8	99.6
4		1.0812	7.8	27	216.0	76.0	39.4	103.6
5		1.0505	21.2	30	205.25	71.25	39.4	104.1
6		1.0484	22.1	35	187.75	69.5	40.5	101.3
7		1.0631	15.6	31	140.25	68.25	33.9	86.0
8		1.0584	17.7	32	148.75	70.0	35.5	86.7
9		1.0911	3.7	27	159.25	71.5	35.7	89.6
10		1.0811	7.9	34	131.5	67.5	36.2	88.6
11		1.091	3.7	27	133.25	64.75	36.4	93.5
12		1.0258	32.6	50	203.0	67.0	40.2	114.8
13		1.025	32.9	44	205.0	29.5	36.6	106.0
14		1.067	13.9	43	164.25	73.25	35.7	96.6
15		1.0665	5.6	39	148.5	71.25	34.6	89.8
16		1.0678	13.6	45	135.75	68.5	32.8	92.3
17		1.0903	4.0	47	127.5	66.75	34.0	83.4
18		1.0756	10.2	47	158.25	72.25	34.9	90.2
19		1.0807	8.0	51	137.25	67.75	36.5	89.7
20		1.0473	22.6	54	198.0	72.0	39.9	107.6

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

Data Transformation ▾

Analytics ▾

Statistics ▾

Regression

Classification

Clustering

Anomaly Detection

Existing Model Utilization

Existing Model Execution

Model (from Tab:)NORMALIZE...

Type Z Score Normalizer Model

Description

Model In...

Header -> Datatype

Density -> Double

Age -> Double

Weight -> Double

Height -> Double

Neck -> Double

Chest -> Double

Abdomen -> Double

Hip -> Double

Thigh -> Double

☐ Transfer Column(s) to Output

Execute

Cancel

The results will appear on the output spreadsheet.

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
User Header	User Row ID	Density	BodyFat	Age	Weight	Height	Neck	Chest
1		1.0297500	10.4	-1.5336943	0.1346456	0.7124504	-0.3107976	0.0458676
2		-0.0056137	19.2	-1.5336943	0.0130864	-0.2226407	-0.7079630	0.4217894
3		0.7888485	12.4	-1.6124414	-0.1489926	0.8059595	-0.1519315	-0.2047469
4		1.3424093	7.8	-1.4549472	1.1476393	2.1150870	0.4835330	0.2509159
5		-0.2311385	21.2	-1.2187057	0.7991694	0.3384139	0.4835330	0.3078737
6		-0.3387753	22.1	-0.8249700	0.2318930	-0.3161498	0.9204149	-0.0110902
7		0.4146824	15.6	-1.1399586	-1.3078573	-0.7836954	-1.7008763	-1.7540004
8		0.1737810	17.7	-1.0612115	-1.0323231	-0.1291316	-1.0654118	-1.6742594
9		1.8498400	3.7	-1.4549472	-0.6919572	0.4319230	-0.9859787	-1.3439039
10		1.3372838	7.9	-0.9037172	-1.5914956	-1.0642227	-0.7873960	-1.4578195
11		1.8447145	3.7	-1.4549472	-1.5347679	-2.0928229	-0.7079630	-0.8996326
12		-1.4971525	32.6	0.3562371	0.7262339	-1.2512409	0.8012653	1.5267717
13		-1.5381570	32.9	-0.1162458	0.7910655	-15.2776074	-0.6285299	0.5243136
14		0.6145794	13.9	-0.1949929	-0.5298782	1.0864868	-0.9859787	-0.5464940
15		0.5889516	5.6	-0.5099815	-1.0404270	0.3384139	-1.4228606	-1.3211207
16		0.6555839	13.6	-0.0374986	-1.4537284	-0.6901863	-2.1377582	-1.0363315
17		1.8088355	4.0	0.1199956	-1.7211587	-1.3447501	-1.6611598	-2.0501812
18		1.0553778	10.2	0.1199956	-0.7243730	0.7124504	-1.3037110	-1.2755544
19		1.3167815	8.0	0.4349842	-1.4051047	-0.9707136	-0.6682464	-1.3325123
20		-0.3951565	22.6	0.6712256	0.5641549	0.6189413	0.6821157	0.7065787

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 k-Nearest Neighbours (kNN) method to train and fit the model: *Analytics* → *Regression* → *k-Nearest Neighbors (kNN)*

The screenshot displays the 'Analytics' menu with 'Regression' selected, leading to a submenu where 'k-Nearest Neighbors (kNN)' is chosen. To the right, a configuration window titled 'kNN Regression Model' is open. It features a 'Target Column' dropdown set to 'Col3 – BodyFat' and a 'Number of Neighbors' input field set to '5'. At the bottom of the window 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)	Col12 (S)	Col13 (D)
User Header	User Row ID	BodyFat	kNN Prediction	Closest NN1	Distance from NN1	Closest NN2	Distance from NN2	Closest NN3	Distance from NN3	Closest NN4	Distance from NN4	Closest NN5	Distance from NN5
1		12.3	12.3180174	Entry 1	0.0	Entry 110	0.2452996	Entry 17	0.2510007	Entry 103	0.2642572	Entry 19	0.2753285
2		6.1	6.1349145	Entry 2	0.0	Entry 101	0.2653769	Entry 22	0.2664016	Entry 21	0.3034601	Entry 8	0.3317935
3		25.3	25.2273072	Entry 3	0.0	Entry 10	0.3857529	Entry 1	0.4344751	Entry 18	0.4454168	Entry 124	0.4548256
4		28.7	28.6456751	Entry 4	0.0	Entry 11	0.4583022	Entry 9	0.4639845	Entry 97	0.4720301	Entry 119	0.4844415
5		20.9	20.8562403	Entry 5	0.0	Entry 14	0.3065005	Entry 7	0.3219477	Entry 128	0.3891621	Entry 114	0.3909337
6		4.1	4.1450844	Entry 6	0.0	Entry 8	0.2675016	Entry 102	0.3946333	Entry 2	0.3947714	Entry 68	0.4006126
7		11.7	11.7350069	Entry 7	0.0	Entry 5	0.3219477	Entry 8	0.3562581	Entry 114	0.3801029	Entry 14	0.3933264
8		7.1	7.1028393	Entry 8	0.0	Entry 6	0.2675016	Entry 102	0.3291884	Entry 2	0.3317935	Entry 68	0.3385871
9		20.8	20.7559106	Entry 9	0.0	Entry 121	0.2357195	Entry 15	0.2378501	Entry 119	0.2460424	Entry 77	0.2890749
10		20.9	20.8780139	Entry 10	0.0	Entry 124	0.2755293	Entry 18	0.3252402	Entry 141	0.3657199	Entry 77	0.3770280
11		29.0	28.9419037	Entry 11	0.0	Entry 126	0.3021431	Entry 31	0.3709831	Entry 26	0.3727252	Entry 14	0.3832615
12		22.9	22.9025004	Entry 12	0.0	Entry 117	0.2634361	Entry 134	0.3034564	Entry 133	0.3182151	Entry 31	0.3230688
13		16.0	16.0628423	Entry 13	0.0	Entry 15	0.2815535	Entry 126	0.3300654	Entry 9	0.3786682	Entry 26	0.4023803
14		16.5	16.5685965	Entry 14	0.0	Entry 133	0.2974149	Entry 5	0.3065005	Entry 134	0.3243685	Entry 128	0.3380295
15		19.1	19.0985038	Entry 15	0.0	Entry 9	0.2378501	Entry 13	0.2815535	Entry 119	0.3366300	Entry 126	0.3377107
16		15.2	15.2657347	Entry 16	0.0	Entry 12	0.3745238	Entry 117	0.3766968	Entry 134	0.4377824	Entry 26	0.4424501
17		14.0	13.9649215	Entry 17	0.0	Entry 107	0.2496449	Entry 1	0.2510007	Entry 19	0.2693057	Entry 110	0.2854921
18		22.9	22.8497149	Entry 18	0.0	Entry 116	0.2731866	Entry 80	0.3039519	Entry 110	0.3115761	Entry 10	0.3252402
19		8.8	8.8582013	Entry 19	0.0	Entry 17	0.2693057	Entry 1	0.2753285	Entry 107	0.2965405	Entry 110	0.3058657
20		11.9	11.9012073	Entry 20	0.0	Entry 8	0.6322888	Entry 114	0.6661730	Entry 84	0.7057182	Entry 66	0.7193273

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 “BodyFat” to be transferred to the output spreadsheet.

The screenshot displays the 'Existing Model Execution' dialog box and the 'Analytics' menu. The dialog box is titled 'Existing Model Execution' and has a 'Model' dropdown set to '(from Tab:) TRAIN_MO...'. The 'Type' is 'kNN Model'. The 'Description' field is empty. The 'Model In...' section lists various input columns: Density, Age, Weight, Height, Neck, Chest, Abdomen, Hip, and Thigh, all with a 'Double' datatype. The 'Transfer Column(s) to Output' checkbox is checked. Below the dialog, the 'Analytics' menu is open, showing options: Regression, Classification, Clustering, Anomaly Detection, and Existing Model Utilization. The 'Existing Model Utilization' option is selected, and a sub-menu is visible showing 'Col2 - Density', 'Col4 - Age', 'Col5 - Weight', 'Col6 - Height', 'Col7 - Neck', 'Col8 - Chest', 'Col9 - Abdomen', 'Col10 - Hip', and 'Col11 - Thigh'. The 'Included Columns' section shows 'Col3 - BodyFat'.

The predictions will appear on the output spreadsheet.

	Col1	Col2 (D)	Col3 (S)	Col4 (D)	Col5 (S)	Col6 (D)	Col7 (S)	Col8 (D)	Col9 (S)	Col10 (D)	Col11 (S)	Col12 (D)	Col13 (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	Closest NN5	Distance from NN5	BodyFat
1		8.6739348	Entry 2	0.2376811	Entry 121	0.2973315	Entry 8	0.3003420	Entry 68	0.3004789	Entry 6	0.3114649	10.4
2		18.0773696	Entry 9	0.2450304	Entry 119	0.2832257	Entry 121	0.3126304	Entry 110	0.3302626	Entry 15	0.3399715	19.2
3		9.6634543	Entry 2	0.2619037	Entry 8	0.2782243	Entry 114	0.3016951	Entry 68	0.3054300	Entry 22	0.3210950	12.4
4		12.1196450	Entry 7	0.3462686	Entry 5	0.3565994	Entry 6	0.3623463	Entry 14	0.3756025	Entry 8	0.3879748	7.8
5		21.7424994	Entry 12	0.2299381	Entry 14	0.2620999	Entry 133	0.2942386	Entry 26	0.2966448	Entry 106	0.3075840	21.2
6		22.2804982	Entry 126	0.2286120	Entry 26	0.3051129	Entry 11	0.3066401	Entry 13	0.3228296	Entry 15	0.3277619	22.1
7		11.9914365	Entry 33	0.2843194	Entry 17	0.3020503	Entry 115	0.3030724	Entry 34	0.3393229	Entry 124	0.3490364	15.6
8		15.8859309	Entry 115	0.2274359	Entry 80	0.2699224	Entry 145	0.2730918	Entry 136	0.2855435	Entry 17	0.2855944	17.7
9		8.1376459	Entry 101	0.2105990	Entry 21	0.2651548	Entry 2	0.2668557	Entry 109	0.2955704	Entry 107	0.3007535	3.7
10		8.7834653	Entry 33	0.2642223	Entry 123	0.2742615	Entry 34	0.2772419	Entry 115	0.3153220	Entry 120	0.3389159	7.9
11		8.9540426	Entry 120	0.3935641	Entry 130	0.4129916	Entry 123	0.4297267	Entry 107	0.4333125	Entry 17	0.4360075	3.7
12		31.1545331	Entry 45	0.3370163	Entry 44	0.3403798	Entry 27	0.3490828	Entry 42	0.3859836	Entry 148	0.3994689	32.6
13		33.3782775	Entry 157	2.5981095	Entry 25	2.6291593	Entry 150	2.6812946	Entry 39	2.7148829	Entry 10	2.7244571	32.9
14		14.4349796	Entry 81	0.2243528	Entry 62	0.2629808	Entry 61	0.2692064	Entry 155	0.2777116	Entry 66	0.2800563	13.9
15		14.1793653	Entry 136	0.2132248	Entry 153	0.2338953	Entry 131	0.2781394	Entry 80	0.2907887	Entry 145	0.3016448	5.6
16		10.0023662	Entry 34	0.2955041	Entry 33	0.3181009	Entry 115	0.3185526	Entry 32	0.3435308	Entry 123	0.3466104	13.6
17		5.3390562	Entry 33	0.2633937	Entry 34	0.2903347	Entry 129	0.3113036	Entry 120	0.3145016	Entry 32	0.3453461	4.0
18		13.0426088	Entry 153	0.2393637	Entry 49	0.2494100	Entry 93	0.2992296	Entry 160	0.3461151	Entry 52	0.3599674	10.2
19		9.3032603	Entry 164	0.2333632	Entry 34	0.3204073	Entry 160	0.3225913	Entry 169	0.3345394	Entry 33	0.3353881	8.0
20		20.1394224	Entry 159	0.2273176	Entry 71	0.2575408	Entry 75	0.2577765	Entry 56	0.2730527	Entry 170	0.2957201	22.6

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 regression: *Statistics* → *Model Metrics* → *Regression Metrics*

The screenshot shows the 'Regression Statistics Metrics' dialog box in the Isalos Analytics Platform. The 'Actual Value Column' is set to 'Col13 -- BodyFat' and the 'Prediction Value Column' is set to 'Col2 -- kNN Predict...'. The 'Statistics' menu is open, showing 'Model Metrics' and 'Regression Metrics' selected. The 'Execute' button is highlighted.

The results will appear on the output spreadsheet.

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)
User Header	User Row ID	Mean Squared Error	Root Mean Squared Error	Mean Absolute Error	R Squared
1		9.6614934	3.1082943	2.4670728	0.8657049

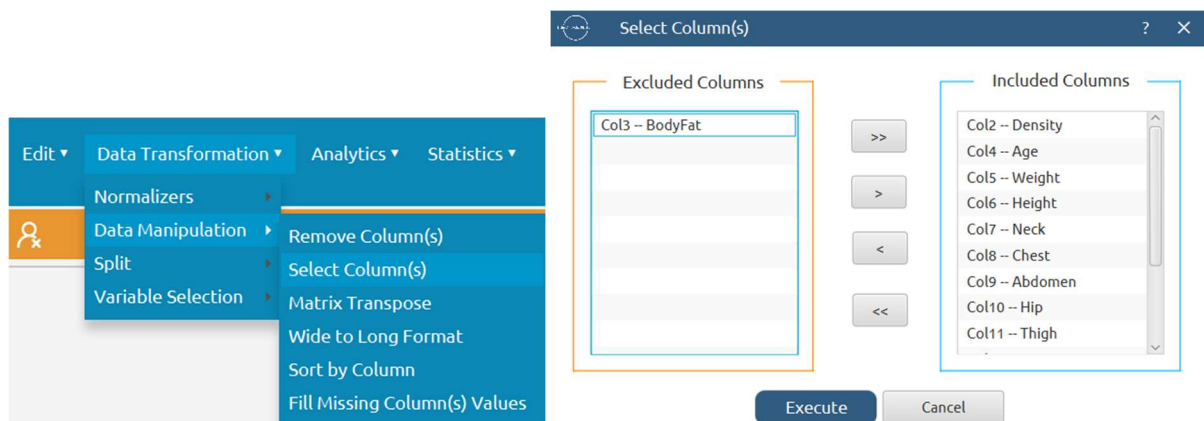
Step 9: Reliability check for each record of the test set

Step 9.a: Create the domain

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

Import data into the input spreadsheet of the “EXCLUDE_BODYFAT” 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 “BodyFat”: *Data Transformation → Data Manipulation → Select Column(s)*

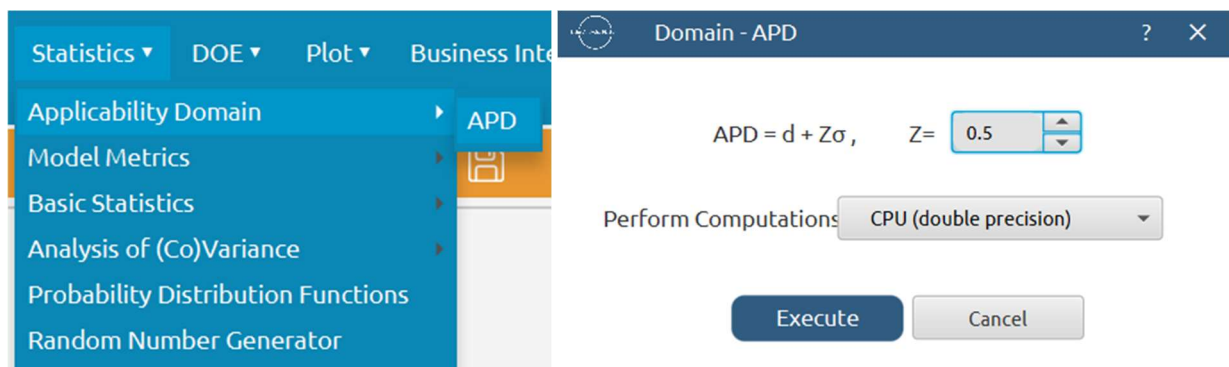


The results will appear on the output spreadsheet.

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_BODYFAT” 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.

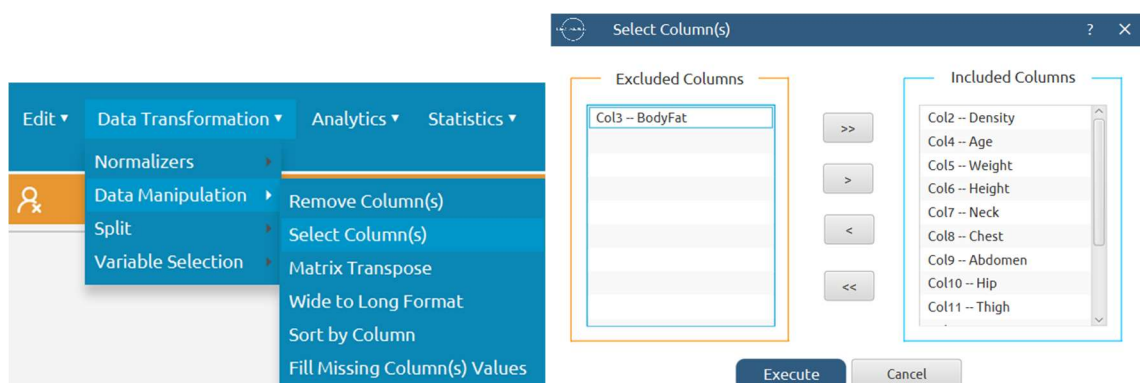
	Col1	Col2 (D)	Col3 (D)	Col4 (S)
User Header	User Row ID	Domain	APD	Prediction
1		0.0	3.8212817	reliable
2		0.0	3.8212817	reliable
3		0.0	3.8212817	reliable
4		0.0	3.8212817	reliable
5		0.0	3.8212817	reliable
6		0.0	3.8212817	reliable
7		0.0	3.8212817	reliable
8		0.0	3.8212817	reliable
9		0.0	3.8212817	reliable
10		0.0	3.8212817	reliable
11		0.0	3.8212817	reliable
12		0.0	3.8212817	reliable
13		0.0	3.8212817	reliable
14		0.0	3.8212817	reliable
15		0.0	3.8212817	reliable
16		0.0	3.8212817	reliable
17		0.0	3.8212817	reliable
18		0.0	3.8212817	reliable
19		0.0	3.8212817	reliable
20		0.0	3.8212817	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_BODYFAT_TEST_SET”.

Import data into the input spreadsheet of the “EXCLUDE_BODYFAT_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 “BodyFat”: *Data Transformation → Data Manipulation → Select Column(s)*

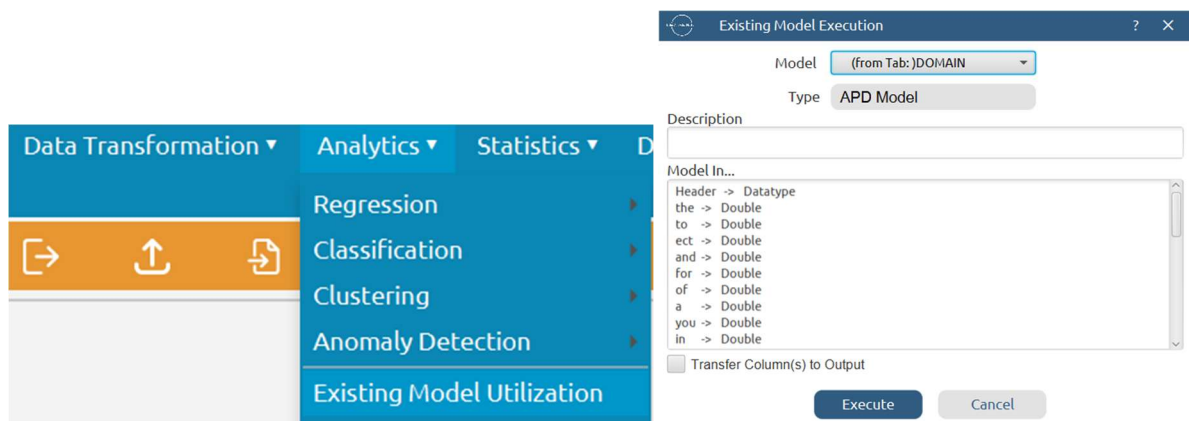


The results will appear on the output spreadsheet.

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_BODYFAT_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		1.5401686	3.8212817	reliable
2		1.6668716	3.8212817	reliable
3		1.5910133	3.8212817	reliable
4		2.1297127	3.8212817	reliable
5		1.5119937	3.8212817	reliable
6		1.7048114	3.8212817	reliable
7		1.6947442	3.8212817	reliable
8		1.4142838	3.8212817	reliable
9		1.2517821	3.8212817	reliable
10		1.7239406	3.8212817	reliable
11		2.2796717	3.8212817	reliable
12		2.0076306	3.8212817	reliable
13		13.6500954	3.8212817	unreliable
14		1.4192282	3.8212817	reliable
15		1.3068319	3.8212817	reliable
16		1.7885726	3.8212817	reliable
17		1.5285037	3.8212817	reliable
18		1.5031340	3.8212817	reliable
19		1.6151654	3.8212817	reliable
20		1.5067090	3.8212817	reliable

Final Isalos Workflow

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

