



Bodyfat Percentage (Regression)

The goal of this study is to train a model in order to predict the percentage of bodyfat in men. The dataset used in this case study is found in <https://www.kaggle.com/datasets/fedesoriano/body-fat-prediction-dataset> and has 14 features and 251 samples. This dataset contains information on certain factors like Density determined from underwater weighing, age, weight, height, neck circumference, Chest circumference, Abdomen 2 circumference, Hip circumference, Thigh circumference, Knee circumference, Ankle circumference, Biceps (extended) circumference, Forearm circumference, Wrist circumference. The dataset contains no missing values and also all features are numerical, therefore there is no need to for transforming any of them.

Step 1: Import data from file

Right click on the input spreadsheet and choose the option "Import from file". Then navigate through your files to load the one with the housing price data.

 A screenshot of the Isalos Analytics Platform software interface. At the top, a menu bar includes File, Edit, Data Transformation, Analytics, Statistics, Plot, and Help. Below the menu is a toolbar with an 'IMPORT' button. The main area shows a spreadsheet with columns labeled Col1 through Col6. Row 6 is highlighted in blue. A context menu is open over row 6, containing options: Import from SpreadSheet, Import from file, Export Spread Sheet Data, and Clear SpreadSheet. The bottom half of the screen displays two large tables of data. The left table is titled 'User Header' and lists rows 1 through 21. The right table is also titled 'User Header' and lists rows 1 through 21. Both tables have columns for User-Row ID, price, area, bedrooms, bathrooms, stories, mainroad, guestroom, and basement. The data in both tables is identical, showing various house prices and characteristics.

User Header	User-Row ID	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement
1	1.33E7	7420.0	4.0	2.0	3.0	1.0	0.0	0.0	
2	1.225E7	8960.0	4.0	4.0	4.0	1.0	0.0	0.0	
3	1.225E7	9960.0	3.0	2.0	2.0	1.0	0.0	1.0	
4	1.2215E7	7500.0	4.0	2.0	2.0	1.0	0.0	1.0	
5	1.141E7	7420.0	4.0	1.0	2.0	1.0	1.0	1.0	
6	1.085E7	7500.0	3.0	3.0	1.0	1.0	0.0	1.0	
7	1.015E7	8580.0	4.0	3.0	4.0	1.0	0.0	0.0	
8	1.015E7	16200.0	5.0	3.0	2.0	1.0	0.0	0.0	
9	8970000.0	8100.0	4.0	1.0	2.0	1.0	1.0	1.0	
10	8900000.0	5750.0	3.0	2.0	4.0	1.0	1.0	0.0	
11	9000000.0	13200.0	3.0	1.0	2.0	1.0	0.0	1.0	
12	9810000.0	6000.0	4.0	3.0	2.0	1.0	1.0	1.0	
13	9110000.0	6550.0	4.0	2.0	2.0	1.0	0.0	0.0	
14	5340000.0	3500.0	4.0	2.0	2.0	1.0	0.0	0.0	
15	6440000.0	7800.0	3.0	2.0	2.0	1.0	0.0	0.0	
16	7100000.0	6000.0	4.0	1.0	2.0	1.0	0.0	1.0	
17	9100000.0	6600.0	4.0	2.0	2.0	1.0	1.0	1.0	
18	2940000.0	8500.0	3.0	2.0	4.0	1.0	0.0	0.0	
19	8990000.0	4600.0	3.0	2.0	2.0	1.0	1.0	0.0	
20	8835000.0	6420.0	3.0	2.0	2.0	1.0	0.0	0.0	
21	8750000.0	4150.0	3.0	1.0	2.0	1.0	0.0	1.0	

Step 2: Manipulate data

In order to use the data for training we have to exclude any columns that do not contain features. In our dataset there are no such columns. Therefore, we will include all columns in the training. We follow these steps to execute this:

- On the menu click on "Data Transformation" → "Data Manipulation" → "Select Column(s)"
- Select all columns.

The screenshot shows the Isalos Analytics Platform interface. At the top, a menu bar includes 'File', 'Edit', 'Data Transformation' (which is highlighted in blue), 'Analytics', 'Statistics', 'Plot', and 'Help'. Under 'Data Transformation', the 'Data Manipulation' option is selected, and its submenu shows 'Remove Column(s)', 'Select Column(s)' (which is also highlighted in blue), 'Matrix Transpose', 'Sort by Column', and 'Fill Missing Column(s) Values'. Below the menu is a spreadsheet view with a header row labeled 'User Header' and columns 'Col1' through 'Col6'. The data rows are numbered 1 to 21. To the right of the spreadsheet is a 'Select Column(s)' dialog box. This dialog has two main sections: 'Excluded Columns' (empty) and 'Included Columns' (containing 'Col2 -- Density', 'Col3 -- BodyFat', 'Col4 -- Age', 'Col5 -- Weight', 'Col6 -- Height', 'Col7 -- Neck', 'Col8 -- Chest', 'Col9 -- Abdomen', and 'Col10 -- Icm'). Between these sections are four buttons: '>>', '>', '<', and '<<'. At the bottom of the dialog are 'Execute' and 'Cancel' buttons.

The data will appear in the output spreadsheet.

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 to create 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 interface with two tabs at the bottom: "IMPORT" and "TRAIN_TEST_SPLIT". The "TRAIN_TEST_SPLIT" tab is active, displaying a table with 21 rows of data. The columns are labeled Col1 through Col10. The first row contains column headers: User Header, User Row ID, Density, Bodyfat, Age, Weight, Height, Neck, Abdomen, and Hip. The data rows follow, showing numerical values for each column. To the right of the main table, there is a smaller, empty table with columns labeled Col1 through Col7, representing the test set.

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

The screenshot shows the software's ribbon menu with "Data Transformation" selected. Under "Data Transformation", the "Split" option is highlighted, and "Random Partitioning" is selected under it. Below the menu, a table of data is visible, identical to the one in the previous screenshot. On the right side, a "Random Partitioning" dialog box is open. It contains three fields: "Training Set Percentage" with the value "75", "Time-based RNG Seed" with the value "46934267465600", and a checked checkbox for "Stratified sampling" with the dropdown value "Col6 -- Height". At the bottom of the dialog are "Execute" and "Cancel" buttons.

The results will appear on the output spreadsheet.

The screenshot shows the Isalos Analytics Platform interface. At the top, there's a menu bar with File, Edit, Data Transformation, Analytics, Statistics, Plot, and Help. Below the menu is a toolbar with IMPORT, TRAIN_TEST_SPLIT, and NORMALIZE_TRAIN_SET buttons. The TRAIN_TEST_SPLIT tab is active, displaying a 21x9 grid of numerical data. The columns are labeled Col1 through Col9. The first row contains headers: User Header, User Row ID, Density, BodyFat, Age, Weight, Height, Neck, Chest, Abdomen, and Hips. The data rows follow, with values ranging from 1.0708000 to 1.0790000. The NORMALIZE_TRAIN_SET tab is also visible but is currently empty.

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 data 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".

The screenshot shows the Isalos Analytics Platform interface with three tabs: IMPORT, TRAIN_TEST_SPLIT, and NORMALIZE_TRAIN_SET. The TRAIN_TEST_SPLIT tab is active, displaying a 21x9 grid of numerical data. The NORMALIZE_TRAIN_SET tab is also visible and contains the same data as the TRAIN_TEST_SPLIT tab. The columns are labeled Col1 through Col9. The first row contains headers: User Header, User Row ID, Density, BodyFat, Age, Weight, Height, Neck, Chest, Abdomen, and Hips. The data rows follow, with values ranging from 1.0708000 to 1.0790000.

Normalize the data using Z-score by browsing: "Data Transformation" → "Normalizers" → "Z-Score". Then select all columns except "BodyFat" and click "Execute".

The results will appear on the output spreadsheet.

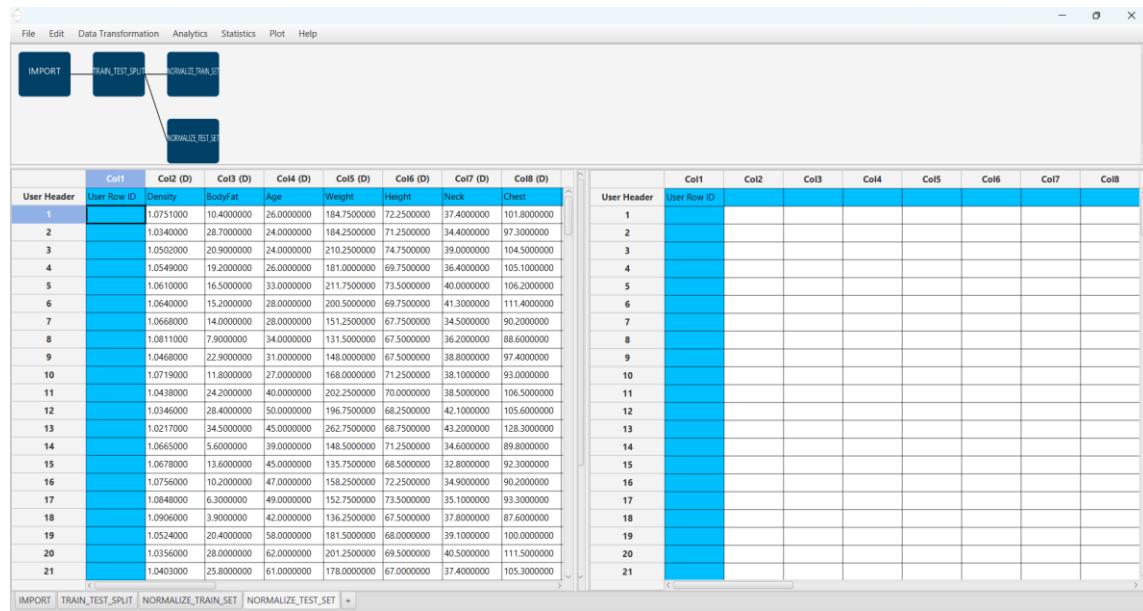
User Header	User Row ID	Density	BodyFat	Age	Weight	Height	Neck	Chest	Abdomen	Hip
1	1.0708000	12.300000	23.000000	154.250000	67.750000	36.200000	93.100000	85.200000	94	
2	1.0853000	6.100000	22.000000	173.250000	38.500000	93.600000	83.000000	98		
3	1.0414000	25.300000	22.000000	154.000000	66.250000	34.000000				
4	1.0704000	12.400000	25.000000	176.000000	72.500000	37.800000				
5	1.0900000	4.100000	25.000000	191.000000	74.000000	38.100000				
6	1.0722000	11.700000	23.000000	198.250000	73.500000	42.100000				
7	1.0830000	7.100000	26.000000	186.250000	74.500000	38.500000				
8	1.0812000	7.800000	27.000000	216.000000	76.000000	39.400000				
9	1.0513000	20.800000	32.000000	180.500000	69.500000	38.400000				
10	1.0505000	21.200000	30.000000	205.250000	71.250000	39.400000				
11	1.0484000	22.100000	35.000000	187.750000	69.500000	40.500000				
12	1.0512000	20.900000	35.000000	162.750000	66.000000	36.400000				
13	1.0330000	29.000000	34.000000	195.750000	71.000000	38.900000				
14	1.0468000	22.900000	32.000000	209.250000	71.000000	42.100000				
15	1.0622000	16.000000	28.000000	183.750000	67.750000	38.000000				
16	1.0551000	19.100000	28.000000	179.000000	68.000000	39.100000				
17	1.0631000	15.600000	31.000000	140.250000	68.250000	33.900000				
18	1.0584000	17.700000	32.000000	148.750000	70.000000	35.500000				
19	1.0911000	3.700000	27.000000	159.250000	71.500000	35.700000				
20	1.0910000	3.700000	27.000000	133.250000	64.750000	36.400000				
21	1.0790000	8.800000	29.000000	160.750000	69.000000	36.700000				

User Header	User Row ID	Density	BodyFat	Age	Weight	Height	Neck	Chest	Abdomen	Hip
1	0.8529956	12.300000	-1.7225084	-0.8604453	-0.5757451	-0.7801497				
2	1.6149861	6.100000	-1.7995809	-0.2182566	0.5630251	0.1666149				
3	-0.6920058	25.300000	-1.7995809	-0.8688952	-0.9553352	-1.6857505				
4	0.8319752	12.400000	-1.5683635	-0.1253082	0.6262902	-0.1215309				
5	1.8619762	4.100000	-1.5683635	0.3816829	1.0058802	0.0019602				
6	0.9265671	11.700000	-1.7225084	0.6267287	0.8793502	1.6485072				
7	1.4941187	7.100000	-1.4912910	0.2211358	1.1324103	0.1666149				
8	1.3995267	7.800000	-1.4142186	1.2266682	1.5120094	0.5370880				
9	-0.1717502	20.800000	-1.0288563	0.0267893	-0.1328900	0.1254512				
10	-0.2137911	21.200000	-1.1830012	0.8633245	0.3099651	0.5370880				
11	-0.3241483	22.100000	-0.7976389	0.2718349	-0.1132890	0.9699884				
12	-0.1770053	20.900000	-0.7976389	-0.5731503	-0.1078602	-0.6978223				
13	-1.1176695	29.000000	0.8747113	0.5422300	0.2467001	0.3312696				
14	-0.4082301	22.900000	-1.0288563	0.9985222	0.2467001	1.6485072				
15	0.4010564	16.000000	-1.3371461	0.1366372	-0.5757451	-0.0392035				
16	0.0279438	19.100000	-1.0288563	-0.0239100	-0.05124801	0.4135969				
17	0.4483524	15.600000	-1.1059287	-1.3336370	-0.4492151	-1.7269142				
18	0.2013624	17.700000	-1.0288563	-1.0463421	-0.0063600	-1.0682954				
19	1.9197823	3.700000	-1.4142186	-0.6914483	0.3732301	-0.9859681				
20	1.9145272	3.700000	-1.4142186	-1.5703229	-1.3349523	-0.6972223				
21	1.2839144	8.800000	-1.2607337	-0.6407492	-0.2594200	-0.5743313				

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".

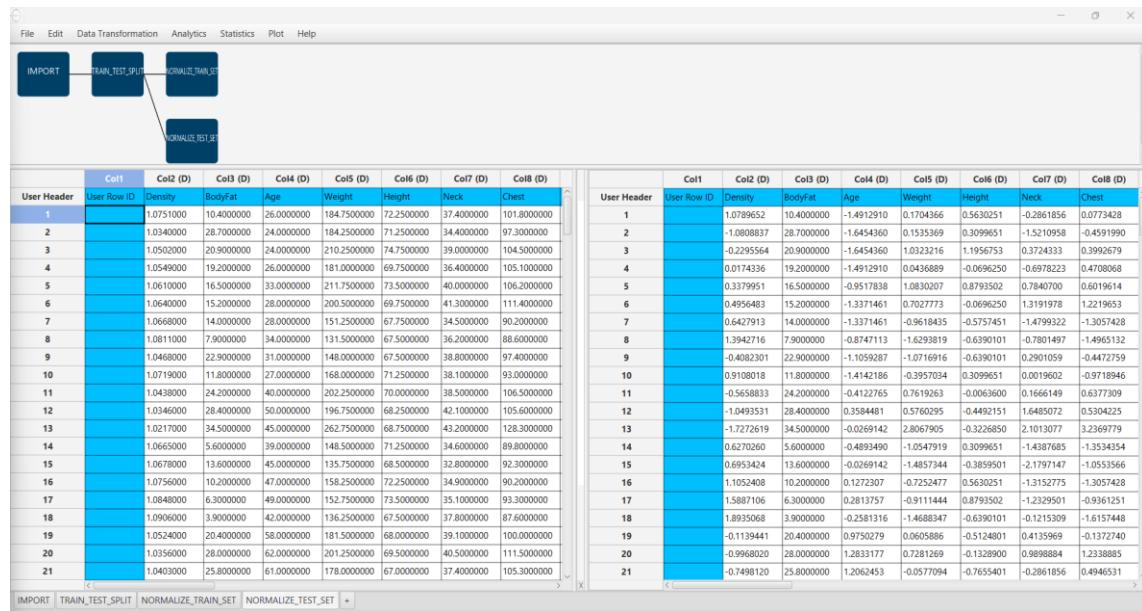
Import data into the input spreadsheet of the "NORMALIZE_TEST_SET" tab the test set from the output of the "TRAIN_TEST_SPLIT" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet". From the available Select input tab options choose "TRAIN_TEST_SPLIT: Test Set".



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

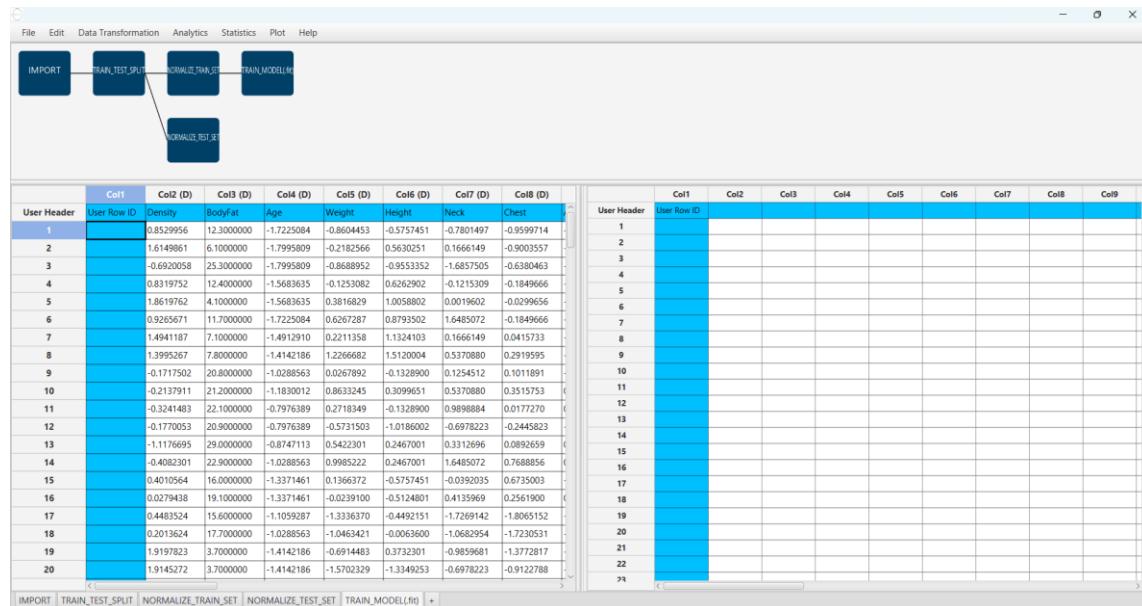
This screenshot shows the Isalos Analytics Platform interface with the Analytics menu selected. In the center, there is a 'Existing Model Execution' dialog box. The dialog has fields for 'Model' (set to '(from Tab:)NORMALIZE_TRAIN_SET') and 'Type' (set to 'Z Score Normalizer Model'). Below these are sections for 'Description' and 'Model Input'. The 'Model Input' section lists various columns and their data types: Header -> Datatype, Density -> Double, Age -> Double, Weight -> Double, Height -> Double, Neck -> Double, Chest -> Double, Abdomen -> Double, and Hip -> Double. At the bottom of the dialog are 'Execute' and 'Cancel' buttons. To the left of the dialog, there is a data table with a 'User Header' row and 21 data rows, identical to the one in the previous screenshot. The bottom of the screen shows a toolbar with IMPORT, TRAIN_TEST_SPLIT, NORMALIZE_TRAIN_SET, and NORMALIZE_TEST_SET buttons.

The results will appear on the output spreadsheet.



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 Neighbors (kNN) method to train and fit the model by browsing: "Analytics" → "Regression" → "k Nearest Neighbors (kNN)" and set the "Target Column" as the column corresponding to "BodyFat" and the "Number of Neighbors" to 5.

The screenshot shows the Isalos Analytics Platform interface. At the top, there's a navigation bar with File, Edit, Data Transformation, Analytics, Statistics, Plot, and Help. The Analytics tab is selected. Below the navigation bar, there's a process flow diagram with nodes: IMPORT, TRAIN_TEST_SPLIT, and several classification and regression models like k-Nearest Neighbors (kNN), Fully Connected Neural Network, Radial Basis Function Network, Linear SGD, XGBoost, and Random Forest. A callout box highlights the kNN node. To the right, a detailed configuration window for the kNN Regression Model is open. It shows the Target Column as Col3 -- BodyFat and the Number of Neighbors set to 5. Buttons for Execute and Cancel are at the bottom.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)
1	User Row ID	Density	BodyFat	Age	Weight	Height	Neck
2	0.8529956	12.300000	-1.7225084	-0.8604453	-0.5757451	-0.7801497	-0.9599714
3	1.6149861	6.1000000	-1.7995809	-0.2182566	0.5630251	0.1666149	-0.6920058
4	-0.6920058	25.300000	-1.7995809	-0.8688952	-0.9553352	-0.1215309	-0.8319752
5	0.8319752	4.1000000	-1.5683635	0.3816829	0.1005880	0.0019602	-0.8319752
6	0.9265671	11.700000	-1.7225084	0.6267287	0.8793502	1.6485072	-0.9265671
7	1.4941187	7.1000000	-1.4912910	0.2211358	1.1324103	0.1666149	-1.3995267
8	1.3995267	7.8000000	-1.4142186	1.2266682	1.5120004	0.5370880	-0.1717502
9	-0.1717502	20.800000	-1.0288563	0.0267892	-0.1328900	0.1254512	-0.2137911
10	-0.2137911	21.200000	-1.1830012	0.8633245	0.3099651	0.5370880	-0.3241483
11	-0.3241483	22.100000	-0.976389	0.2718349	-0.1328900	0.9898884	-0.1770053
12	-0.1770053	20.900000	-0.976389	-0.5731503	-0.10186002	-0.6978223	-1.1176695
13	-1.1176695	29.000000	-0.8747113	0.5422301	0.2467001	0.3312696	-0.4082301
14	-0.4082301	22.900000	-1.0288563	0.9985222	0.2467001	1.6485072	-0.4010564
15	0.4010564	16.000000	-1.3371461	0.1366372	-0.5757451	-0.0392035	-0.0279438
16	-0.0279438	19.100000	-1.3371461	-0.0239100	-0.5124801	0.4135969	-0.4483524
17	-0.4483524	15.600000	-1.1059287	-1.3336370	-0.4492151	-1.7269142	-0.2013624
18	-0.2013624	17.700000	-1.0288563	-1.0463421	-0.0063600	-1.0682954	-1.9197823
19	-1.9197823	3.700000	-1.4142186	-0.6914483	0.3732301	-0.9859681	-1.9145272
20	-1.9145272	3.700000	-1.4142186	-1.5702329	-1.3349253	-0.6978223	-1.3995267

The predictions will appear on the output spreadsheet.

The screenshot shows the Isalos Analytics Platform interface with the output spreadsheet. The process flow diagram includes nodes: IMPORT, TRAIN_TEST_SPLIT, NORMALIZE_TRAIN_SET, and TRAIN_MODEL(.fit). A callout box highlights the NORMALIZE_TEST_SET node. The output spreadsheet contains two tables: one for User Header and one for User Row ID. The User Header table has columns Col1 through Col8. The User Row ID table has columns Col1 through Col9. Both tables include columns for BodyFat, Prediction, Closest NN1, Distance from NN1, Closest NN2, Distance from NN2, Closest NN3, and Distance from NN3.

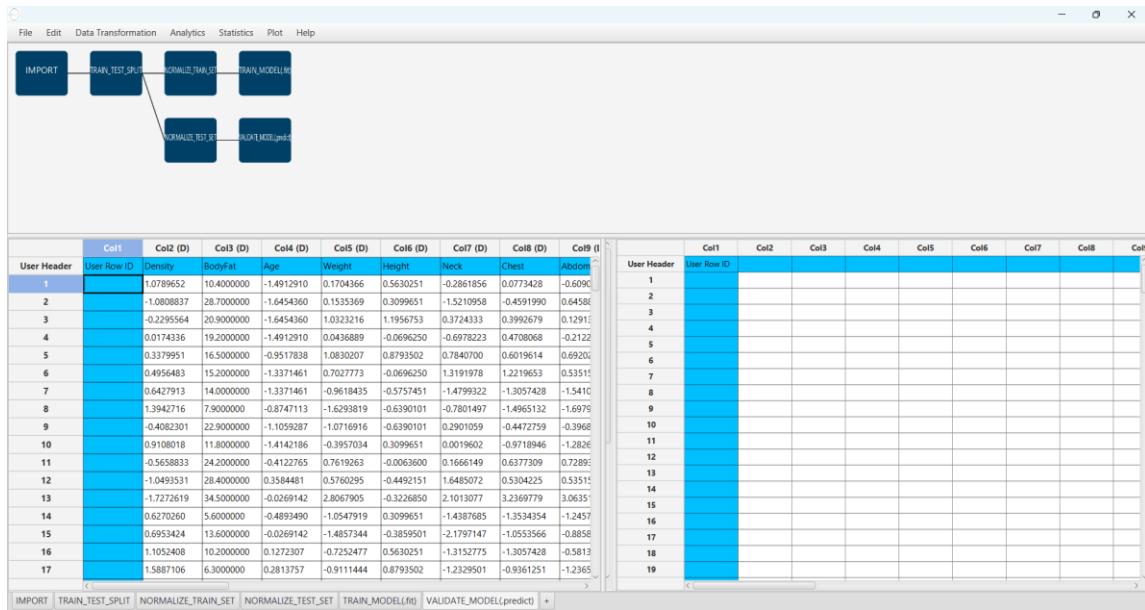
User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
1	User Row ID	Density	BodyFat	Age	Weight	Height	Neck	Chest
2	0.8529956	12.300000	-1.7225084	-0.8604453	-0.5757451	-0.7801497	-0.9599714	-0.6920058
3	1.6149861	6.1000000	-1.7995809	-0.2182566	0.5630251	0.1666149	-0.6920058	-0.8319752
4	-0.6920058	25.300000	-1.7995809	-0.8688952	-0.9553352	-0.1215309	-0.8319752	-0.8319752
5	0.8319752	4.1000000	-1.5683635	-0.1253082	0.6262902	-0.1215309	-0.1849666	-0.8319752
6	0.9265671	11.700000	-1.7225084	0.6267287	0.8793502	-0.1849666	-0.9265671	-0.9265671
7	1.4941187	7.1000000	-1.4912910	0.2211358	1.1324103	0.1666149	-1.3995267	-1.3995267
8	1.3995267	7.8000000	-1.4142186	1.2266682	1.5120004	0.5370880	-0.2919595	-0.2919595
9	-0.1717502	20.800000	-1.0288563	0.0267892	-0.1328900	0.1254512	0.1011891	-0.1717502
10	-0.2137911	21.200000	-1.1830012	0.8633245	0.3099651	0.5370880	0.3515753	-0.2137911
11	-0.3241483	22.100000	-0.976389	0.2718349	-0.1328900	0.9898884	0.0177270	-0.3241483
12	-0.1770053	20.900000	-0.976389	-0.5731503	-1.0186002	-0.6978223	-0.2445823	-0.1770053
13	-1.1176695	29.000000	-0.8747113	0.5422301	0.2467001	0.3132694	0.0892659	-1.1176695
14	-0.4082301	22.900000	-1.0288563	0.9985222	0.2467001	1.6485072	0.7688856	-0.4082301
15	0.4010564	16.000000	-1.3371461	0.1366372	-0.5757451	-0.0392035	0.6735003	-0.4010564
16	0.0279438	19.100000	-1.3371461	-0.0239100	-0.5124801	0.4135969	0.2561900	-0.0279438
17	-0.4483524	15.600000	-1.1059287	-1.3336370	-0.4492151	-1.7269142	-1.8065152	-0.4483524
18	-0.2013624	17.700000	-1.0288563	-1.0463421	-0.0063600	-1.0682954	-1.7230531	-0.2013624
19	-1.9197823	3.700000	-1.4142186	-0.6914483	0.3732301	-0.9859681	-1.3772817	-1.9197823
20	-1.9145272	3.700000	-1.4142186	-1.5702329	-1.3349253	-0.6978223	-0.9122788	-1.3995267

User Header	User Row ID	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)
1	12.300000	12.2966713	Entry 1	0.6-7	Entry 113	0.2416378	Entry 21	0.2563247	E	
2	6.1000000	6.1470758	Entry 2	0.6-7	Entry 107	0.2011506	Entry 19	0.2585959	E	
3	25.3000000	25.2212809	Entry 3	0.6-7	Entry 12	0.3686288	Entry 1	0.4269102	E	
4	12.4000000	12.3349552	Entry 4	0.6-7	Entry 116	0.2116042	Entry 107	0.2370151	E	
5	4.1000000	4.1758499	Entry 5	0.6-7	Entry 7	0.270779	Entry 107	0.3126455	E	
6	11.7000000	11.7075962	Entry 6	0.6-7	Entry 8	0.2971078	Entry 7	0.3487918	E	
7	7.1000000	7.1356214	Entry 7	0.6-7	Entry 116	0.2262800	Entry 4	0.2409434	E	
8	7.0000000	7.82151313	Entry 8	0.6-7	Entry 6	0.2971078	Entry 5	0.3351049	E	
9	20.8000000	20.7959210	Entry 9	0.6-7	Entry 129	0.1867059	Entry 128	0.2103977	E	
10	21.2000000	21.2425952	Entry 10	0.6-7	Entry 14	0.228561	Entry 122	0.2640104	E	
11	22.1000000	22.1003355	Entry 11	0.6-7	Entry 132	0.2263939	Entry 13	0.2882791	E	
12	20.9000000	20.8814447	Entry 12	0.6-7	Entry 82	0.2403556	Entry 131	0.2666028	E	
13	29.0000000	28.9322117	Entry 13	0.6-7	Entry 132	0.2744677	Entry 11	0.2882791	E	
14	22.9000000	22.8604494	Entry 14	0.6-7	Entry 10	0.2285561	Entry 125	0.2511103	E	
15	16.0000000	16.0657213	Entry 15	0.6-7	Entry 16	0.2810801	Entry 11	0.2920700	E	
16	19.1000000	19.1124245	Entry 16	0.6-7	Entry 9	0.2160155	Entry 15	0.2810801	E	
17	15.6000000	15.5719198	Entry 17	0.6-7	Entry 18	0.1965849	Entry 121	0.2580966	E	
18	17.7000000	17.6916994	Entry 18	0.6-7	Entry 17	0.1965848	Entry 121	0.2425329	E	
19	3.7000000	3.7519244	Entry 19	0.6-7	Entry 106	0.2030465	Entry 126	0.2065018	E	
20	3.7000000	3.7201349	Entry 20	0.6-7	Entry 126	0.2859354	Entry 19	0.291248	E	
21	8.0000000	8.7815776	Entry 21	0.6-7	Entry 106	0.2465051	Entry 126	0.2512018	E	
22	11.9000000	11.9023665	Entry 22	0.6-7	Entry 116	0.2630973	Entry 7	0.2630949	E	

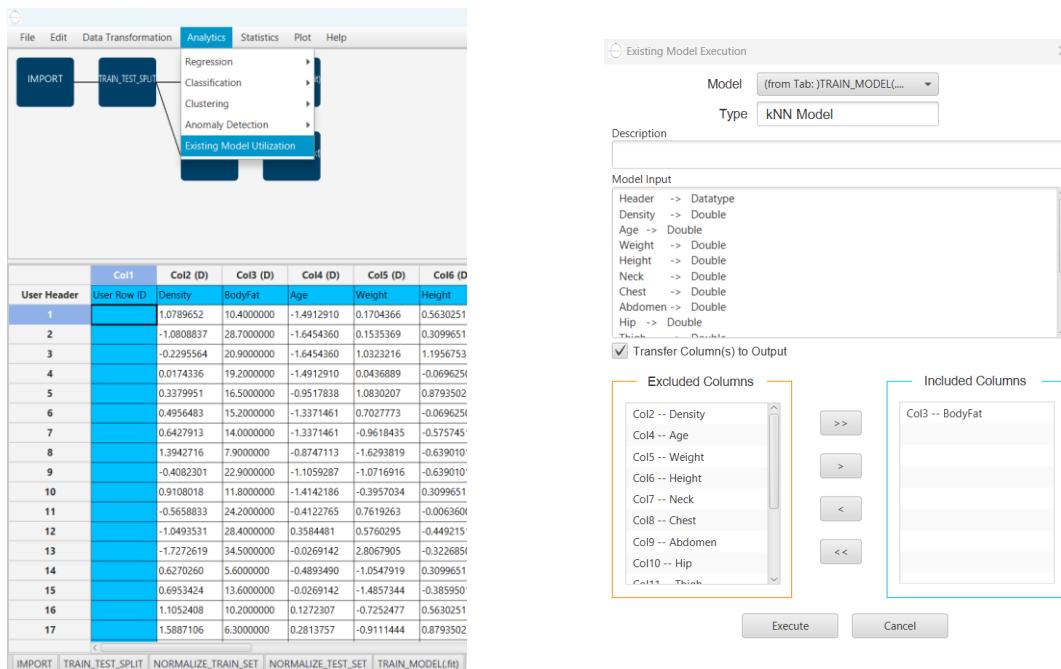
Step 7: Validate the model

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 browse: "Analytics" → "Existing Model Utilization". Then choose Model "(from Tab:) TRAIN_MODEL (.fit)". and transfer the "BodyFat" column in the output.



The predictions will appear on the output spreadsheet.

The screenshot shows the Isalos Analytics Platform interface. At the top, there is a menu bar with File, Edit, Data Transformation, Analytics, Statistics, Plot, Help, and a search bar. Below the menu is a workflow diagram:

```

graph LR
    IMPORT[IMPORT] --> TRAIN_TEST_SPLIT[TRAIN_TEST_SPLIT]
    TRAIN_TEST_SPLIT --> NORMALIZE_TRAIN_SET[NORMALIZE_TRAIN_SET]
    NORMALIZE_TRAIN_SET --> TRAIN_MODEL[TRAIN_MODEL(.fit)]
    TRAIN_MODEL --> VALIDATE_MODEL[VALIDATE_MODEL(.predict)]
    VALIDATE_MODEL --> NORMALIZE_TEST_SET[NORMALIZE_TEST_SET]
    NORMALIZE_TEST_SET --> PREDICTION[VALIDATE_MODEL(.predict)]
  
```

Below the workflow are two data tables:

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (I)
1	1.0789652	10.4000000	-1.4912910	0.1704366	0.5630251	-0.2861856	0.0773428	-0.0690	
2	-1.0800837	28.7000000	-1.6454360	0.1535369	0.3099651	-1.5210958	-0.4591990	0.64586	
3	0.2295564	20.9000000	-1.6454360	1.0323216	1.1956753	0.3724333	0.3992679	0.12915	
4	0.0174336	19.2000000	-1.4912910	0.0436804	-0.696250	-0.6978223	0.4708066	-0.2122	
5	0.3379951	16.5000000	-0.9517838	1.0830207	0.8793502	0.7840700	0.6019614	0.69202	
6	0.4956483	15.2000000	-1.3371461	0.7027773	-0.696250	1.3191978	1.2219653	0.53515	
7	0.6427913	14.0000000	-1.3371461	-0.9618435	-0.5757451	-1.4799322	-1.3057428	-1.5411	
8	1.3942716	7.9000000	-0.8747113	-1.6293819	-0.6390101	-0.7801497	-1.4965132	-1.6975	
9	-0.4082301	22.9000000	-1.1059287	-1.0716916	-0.6390101	0.2901059	-0.4472759	-0.3966	
10	0.9108018	11.8000000	-1.4142186	-0.3957034	0.3099651	0.0019602	-0.9718946	-1.2826	
11	-0.5658833	24.2000000	-0.422765	0.7619263	0.1666149	0.6377309	0.72895		
12	-1.0493531	28.4000000	0.3584481	0.5760295	-0.4926150	1.6485072	0.5304225	0.53515	
13	-1.7272619	34.5000000	-0.0269142	2.8057905	-0.3226850	2.1013077	3.2369779	3.0635	
14	0.6270260	5.6000000	-0.4893490	0.30547919	0.3099651	-1.3587685	-1.3534354	-1.2457	
15	0.6953424	13.6000000	-0.0269142	-1.4857344	-0.3859501	-2.1797147	-1.0535366	-0.8856	
16	1.1052408	10.2000000	-0.1272307	0.5630251	-1.3152775	0.2689456	0.2947428	0.5813	
17	1.5887106	6.3000000	0.2813757	-0.9111444	0.8793502	-1.2329500	-0.9361251	-1.2365	

User Header	Col1	Col2 (D)	Col3 (S)	Col4 (D)	Col5 (S)	Col6 (D)	Col7 (S)	Col8 (D)	Col9 (I)
1	9.3312024	MNN prediction	Closest NN1	Distance from NN1	Closest NN2	Distance from NN2	Closest NN3	Distance from NN3	Closest NN4
2	25.1974449	Entry 107	0.1676399	Entry 4	0.2219115	Entry 116	0.2350152	Entry 115	0.2203273
3	21.9519381	Entry 117	0.0316779	Entry 101	0.4301765	Entry 86	0.2789554	Entry 101	0.4301765
4	19.9563764	Entry 117	0.2294060	Entry 114	0.2233629	Entry 9	0.3104744	Entry 6	0.3104744
5	21.7408855	Entry 10	0.2132167	Entry 139	0.2689456	Entry 11	0.2902531	Entry 11	0.2902531
6	20.7413738	Entry 123	0.2990883	Entry 125	0.3280152	Entry 14	0.3642761	Entry 14	0.3642761
7	10.5655667	Entry 106	0.2276968	Entry 18	0.2392316	Entry 1	0.2475311	Entry 2	0.2475311
8	9.1975497	Entry 121	0.2373282	Entry 36	0.2524254	Entry 20	0.2558686	Entry 3	0.2558686
9	19.2633812	Entry 84	0.2662981	Entry 124	0.2760216	Entry 12	0.305974	Entry 1	0.305974
10	10.7318509	Entry 2	0.2552910	Entry 4	0.3100214	Entry 113	0.3174439	Entry 1	0.3174439
11	24.1306480	Entry 139	0.1389985	Entry 141	0.2086915	Entry 161	0.2242341	Entry 7	0.2242341
12	28.2789497	Entry 45	0.2569128	Entry 161	0.2940622	Entry 44	0.2973686	Entry 7	0.2973686
13	30.9301278	Entry 25	0.5081436	Entry 143	0.6147704	Entry 180	0.7325640	Entry 8	0.7325640
14	13.4702913	Entry 142	0.1610765	Entry 155	0.2203273	Entry 18	0.2401367	Entry 8	0.2401367
15	10.9296612	Entry 121	0.2692974	Entry 50	0.2829168	Entry 36	0.2904809	Entry 3	0.2904809
16	11.7469717	Entry 155	0.2156831	Entry 97	0.2948418	Entry 36	0.2995778	Entry 1	0.2995778
17	13.8738429	Entry 47	0.4003459	Entry 91	0.4195432	Entry 154	0.4256256	Entry 8	0.4256256
18	7.7430758	Entry 142	0.2621299	Entry 126	0.2912166	Entry 156	0.3335921	Entry 3	0.3335921
19	23.2826818	Entry 174	0.2240590	Entry 40	0.2317355	Entry 57	0.2438784	Entry 5	0.2438784

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".

The screenshot shows the Isalos Analytics Platform interface. At the top, there is a menu bar with File, Edit, Data Transformation, Analytics, Statistics, Plot, Help, and a search bar. Below the menu is a workflow diagram:

```

graph LR
    IMPORT[IMPORT] --> TRAIN_TEST_SPLIT[TRAIN_TEST_SPLIT]
    TRAIN_TEST_SPLIT --> NORMALIZE_TRAIN_SET[NORMALIZE_TRAIN_SET]
    NORMALIZE_TRAIN_SET --> TRAIN_MODEL[TRAIN_MODEL(.fit)]
    TRAIN_MODEL --> VALIDATE_MODEL[VALIDATE_MODEL(.predict)]
    VALIDATE_MODEL --> NORMALIZE_TEST_SET[NORMALIZE_TEST_SET]
    NORMALIZE_TEST_SET --> PREDICTION[VALIDATE_MODEL(.predict)]
    PREDICTION --> STATISTICS[STATISTICS_ACCURACIES]
  
```

Below the workflow are two data tables:

User Header	Col1	Col2 (D)	Col3 (S)	Col4 (D)	Col5 (S)	Col6 (D)	Col7 (S)	Col8 (D)	Col9 (I)
1	9.3312024	MNN prediction	Closest NN1	Distance from NN1	Closest NN2	Distance from NN2	Closest NN3	Distance from NN3	Closest NN4
2	25.1974449	Entry 107	0.1676399	Entry 4	0.2219115	Entry 116	0.2350152	Entry 115	0.2203273
3	21.9519381	Entry 117	0.0316779	Entry 101	0.4301765	Entry 86	0.2789554	Entry 101	0.4301765
4	19.9563764	Entry 117	0.2294060	Entry 114	0.2233629	Entry 9	0.3104744	Entry 6	0.3104744
5	21.7408855	Entry 10	0.2132167	Entry 139	0.2689456	Entry 11	0.2902531	Entry 11	0.2902531
6	20.7413738	Entry 123	0.2990883	Entry 125	0.3280152	Entry 14	0.3642761	Entry 14	0.3642761
7	10.5655667	Entry 106	0.2276968	Entry 18	0.2392316	Entry 1	0.2475311	Entry 2	0.2475311
8	9.1975497	Entry 121	0.2373282	Entry 36	0.2524254	Entry 20	0.2558686	Entry 10	0.2558686
9	19.2633812	Entry 84	0.2662981	Entry 124	0.2760216	Entry 12	0.305974	Entry 11	0.305974
10	10.7318509	Entry 2	0.2552910	Entry 4	0.3100214	Entry 113	0.3174439	Entry 12	0.3174439
11	24.1306480	Entry 139	0.1389985	Entry 141	0.2086915	Entry 161	0.2242341	Entry 13	0.2242341
12	28.2789497	Entry 45	0.2569128	Entry 161	0.2940622	Entry 44	0.2973686	Entry 14	0.2973686
13	30.9301278	Entry 25	0.5081436	Entry 143	0.6147704	Entry 180	0.7325640	Entry 15	0.7325640
14	13.4702913	Entry 142	0.1610765	Entry 155	0.2203273	Entry 18	0.2401367	Entry 16	0.2401367
15	10.9296612	Entry 121	0.2692974	Entry 50	0.2829168	Entry 36	0.2904809	Entry 17	0.2904809
16	11.7469717	Entry 155	0.2156831	Entry 97	0.2948418	Entry 36	0.2995778	Entry 18	0.2995778
17	13.8738429	Entry 47	0.4003459	Entry 91	0.4195432	Entry 154	0.4256256	Entry 19	0.4256256

User Header	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									

Calculate the statistical metrics for the regression by browsing: "Statistics" → "Model Metrics" → "Regression Metrics".

The screenshot shows the Isalos Analytics Platform interface. At the top, there's a menu bar with File, Edit, Data Transformation, Analytics, Statistics, Plot, and Help. The Analytics tab is selected. Below the menu is a workflow diagram with nodes: IMPORT, TRAIN_TEST_SPLIT, NORMALIZE_TRAIN_SET, and NORMALIZE_TEST_SET. A context menu is open over the NORMALIZE_TEST_SET node, showing options like Domain - APD, Model Metrics, Probability Distribution Functions, Descriptive Statistics, Confidence Intervals, Hypothesis Testing, Weight Cases, Random Number Generator, and Design of Experiments. The 'Model Metrics' option is highlighted. To the right, a modal window titled 'Regression Statistics Metrics' is open, showing fields for 'Actual Value Column' (set to Col13 -- BodyFat) and 'Prediction Value Column' (set to Col2 -- kNN Prediction). Buttons for 'Execute' and 'Cancel' are at the bottom.

User Header	Col1	Col2 (D)	Col3 (S)	Col4 (D)	Col5 (S)	Col6 (D)	Col7 (S)	Col8 (D)	Col9 (S)
1	9.3312024	Entry 107	0.1676399	Entry 4	0.2219115	Entry 116	0.2350152	Entry 10	Er
2	25.1974449	Entry 117	0.3516779	Entry 101	0.4301765	Entry 86	0.4085555	Entry 10	Er
3	21.9519381	Entry 10	0.2294060	Entry 114	0.2789554	Entry 6	0.3103474	Entry 10	Er
4	19.9563764	Entry 117	0.22660597	Entry 128	0.2333629	Entry 9	0.2394963	Entry 10	Er
5	21.740855	Entry 10	0.2132167	Entry 139	0.2689456	Entry 11	0.2902531	Entry 10	Er
6	20.7413738	Entry 123	0.2990883	Entry 125	0.3280152	Entry 14	0.3642761	Entry 10	Er
7	10.5655667	Entry 106	0.2276968	Entry 18	0.2392316	Entry 1	0.2475311	Entry 10	Er
8	9.1975497	Entry 121	0.2373282	Entry 36	0.2524254	Entry 20	0.2558688	Entry 10	Er
9	19.2633812	Entry 98	0.2662981	Entry 124	0.2760216	Entry 12	0.3058974	Entry 10	Er
10	10.7318509	Entry 2	0.2552910	Entry 4	0.3100214	Entry 113	0.3174439	Entry 10	Er
11	24.1308480	Entry 139	0.1389985	Entry 141	0.2086915	Entry 161	0.2242341	Entry 10	Er
12	28.2789497	Entry 45	0.2569126	Entry 161	0.2940622	Entry 44	0.2973686	Entry 10	Er
13	30.9301278	Entry 25	0.5081436	Entry 143	0.6147704	Entry 180	0.7325640	Entry 10	Er
14	13.4702913	Entry 142	0.1610765	Entry 155	0.2203273	Entry 18	0.2401367	Entry 10	Er
15	10.9296812	Entry 121	0.2692974	Entry 50	0.2829168	Entry 36	0.2904809	Entry 10	Er
16	11.7469717	Entry 155	0.2156831	Entry 97	0.2948418	Entry 36	0.2995778	Entry 10	Er
17	13.8738429	Entry 47	0.4003459	Entry 91	0.4195432	Entry 154	0.4256256	Entry 10	Er

The results will appear on the output spreadsheet.

The screenshot shows the Isalos Analytics Platform interface. At the top, there's a menu bar with File, Edit, Data Transformation, Analytics, Statistics, Plot, and Help. The Analytics tab is selected. Below the menu is a workflow diagram with nodes: IMPORT, TRAIN_TEST_SPLIT, NORMALIZE_TRAIN_SET, and NORMALIZE_TEST_SET. A context menu is open over the NORMALIZE_TEST_SET node, showing options like Domain - APD, Model Metrics, Probability Distribution Functions, Descriptive Statistics, Confidence Intervals, Hypothesis Testing, Weight Cases, Random Number Generator, and Design of Experiments. The 'Statistics' option is highlighted. To the right, a table titled 'STATISTICS ACCURACIES' is displayed, showing various statistical measures for each record. The columns include User Header, User Row ID, Col1, Col2 (D), Col3 (S), Col4 (D), Col5 (S), Col6 (D), Col7 (S), Col8 (D), Col9 (S), and Col10.

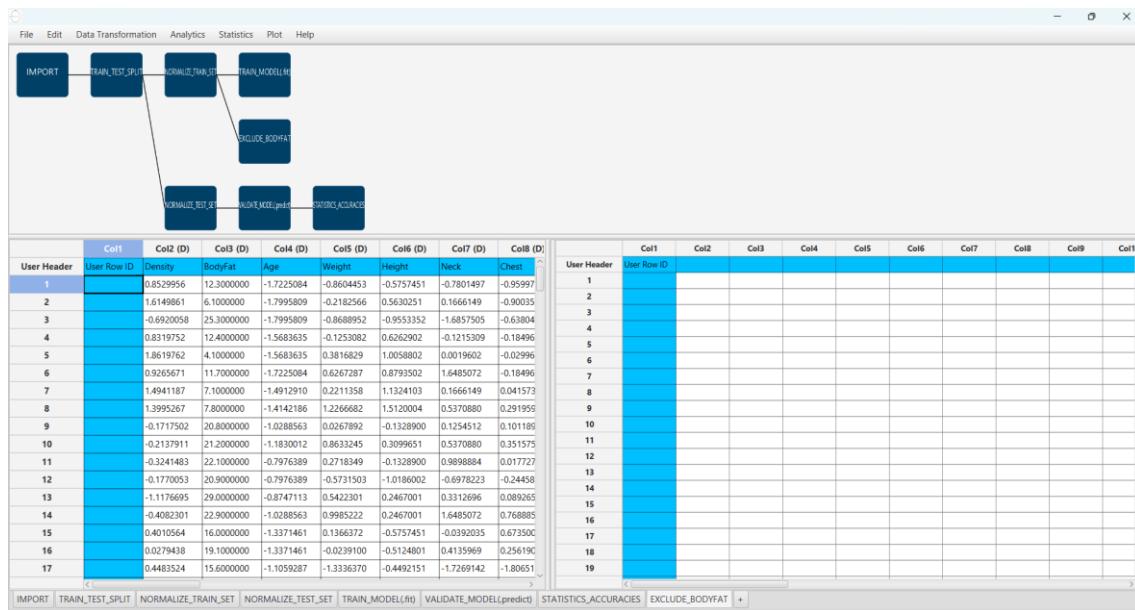
User Header	User Row ID	Col1	Col2 (D)	Col3 (S)	Col4 (D)	Col5 (S)	Col6 (D)	Col7 (S)	Col8 (D)	Col9 (S)	Col10
1	9.3312024	9.7045596	3.1152142	2.4368580	0.8697938						
2											
3											
4											
5											
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											

Step 9: Reliability check of 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".

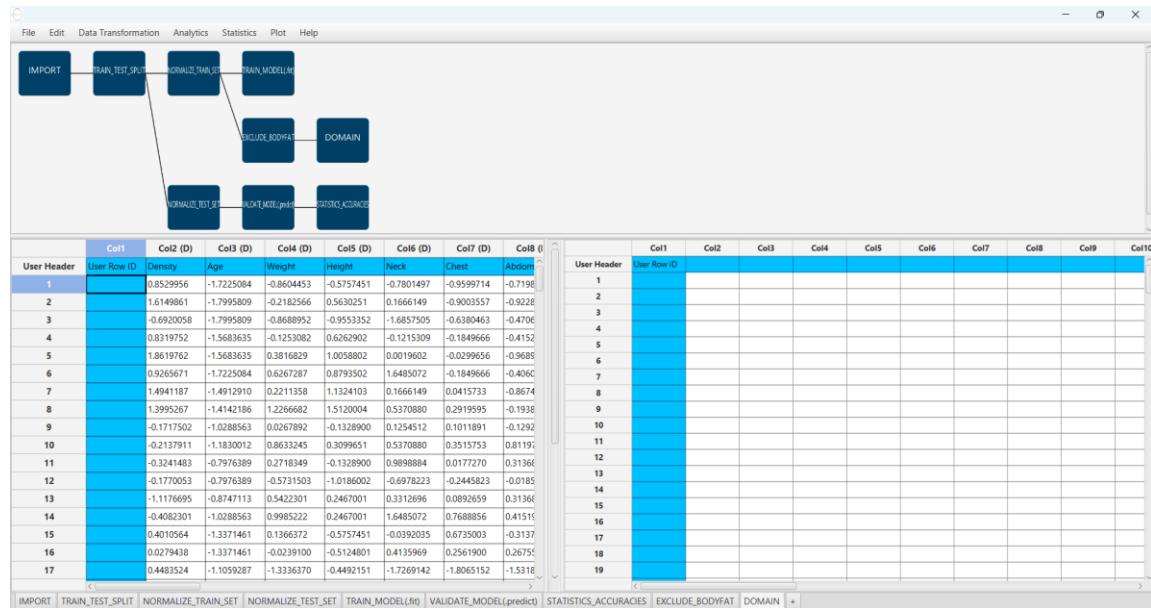


Manipulate the data to exclude the column that corresponds to the "BodyFat" by browsing: "Data Transformation" → "Data Manipulation" → "Select Column(s)". Then select all the columns except the "BodyFat". The filtered table will appear on the output spreadsheet.

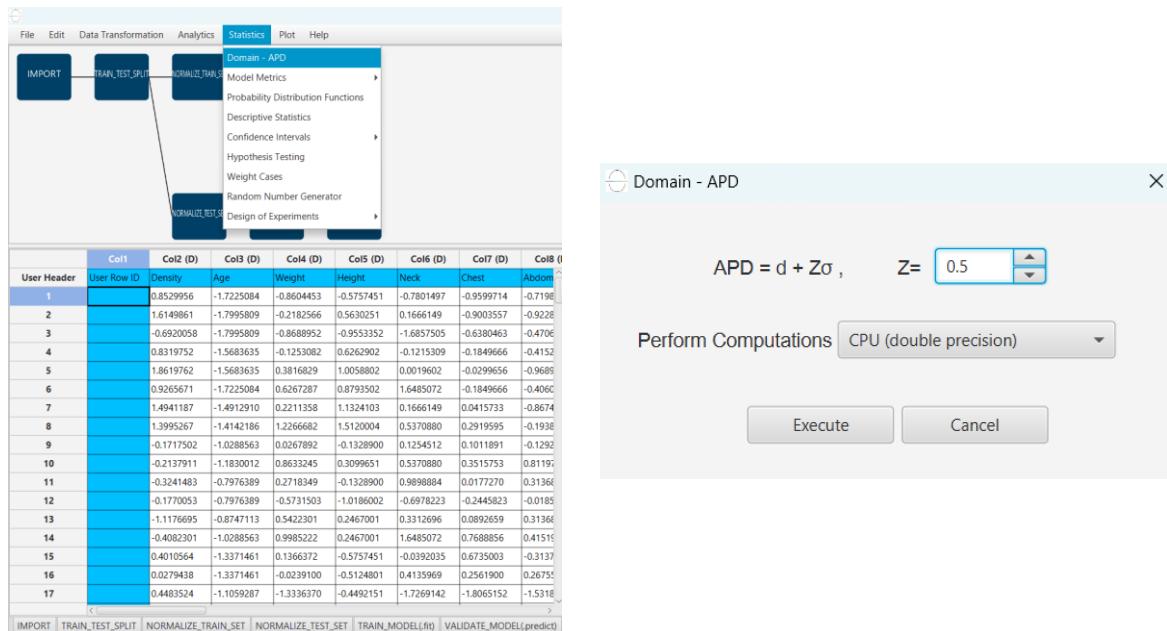
The screenshot shows the 'Data Transformation' menu open, with 'Data Manipulation' selected. A sub-menu 'Select Column(s)' is open. To the right, a 'Select Column(s)' dialog box is displayed. It has two sections: 'Excluded Columns' (containing 'Col3 -- BodyFat') and 'Included Columns' (containing 'Col2 -- Density', 'Col4 -- Age', 'Col5 -- Weight', 'Col6 -- Height', 'Col7 -- Neck', 'Col8 -- Chest', 'Col9 -- Abdomen', 'Col10 -- Hip', and 'Col11 -- Thigh'). At the bottom of the dialog are 'Execute' and 'Cancel' buttons.

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

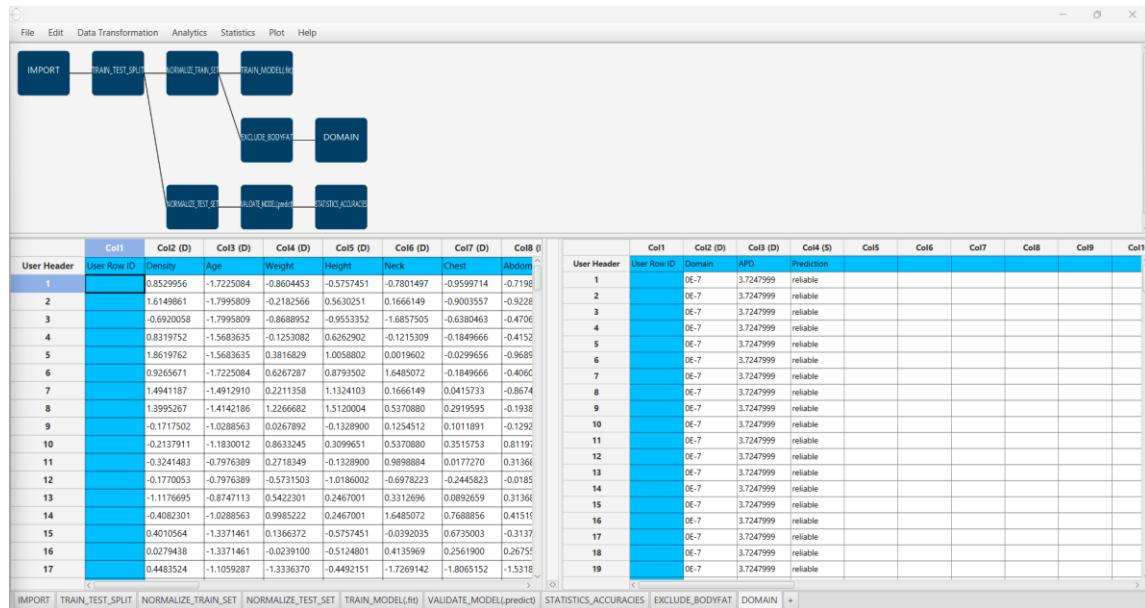
Import data into the input spreadsheet of the "DOMAIN" tab from the output of the "EXCLUDE_BODYFAT" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".



Create the domain of applicability by browsing: "Statistics" → "Domain APD".



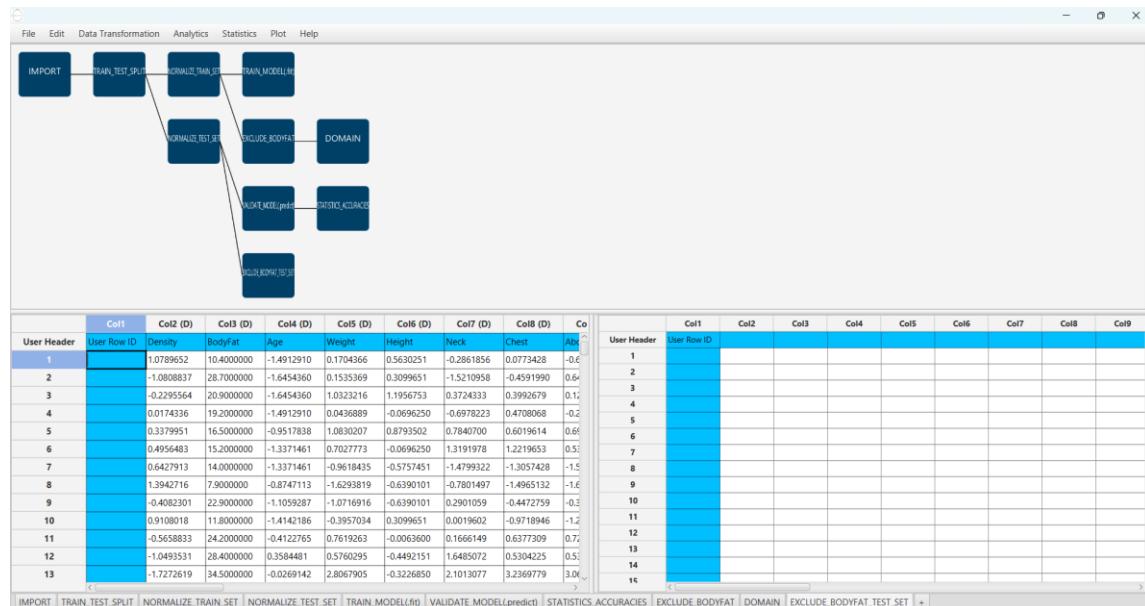
The results will appear on the output spreadsheet.



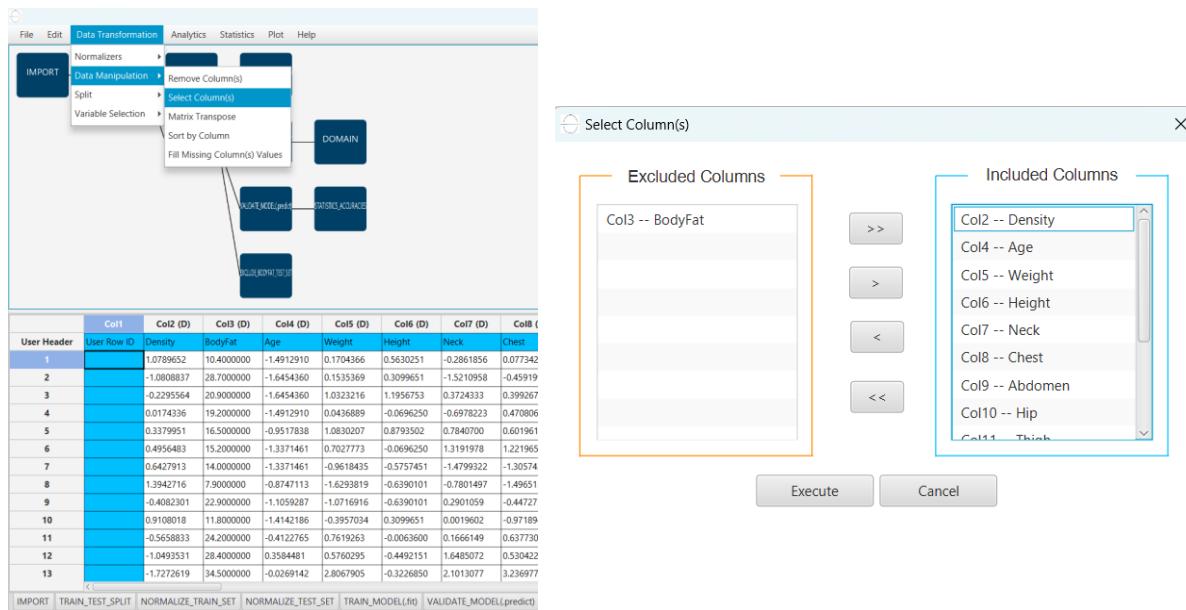
Step 9.b: Check the test set reliability

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".



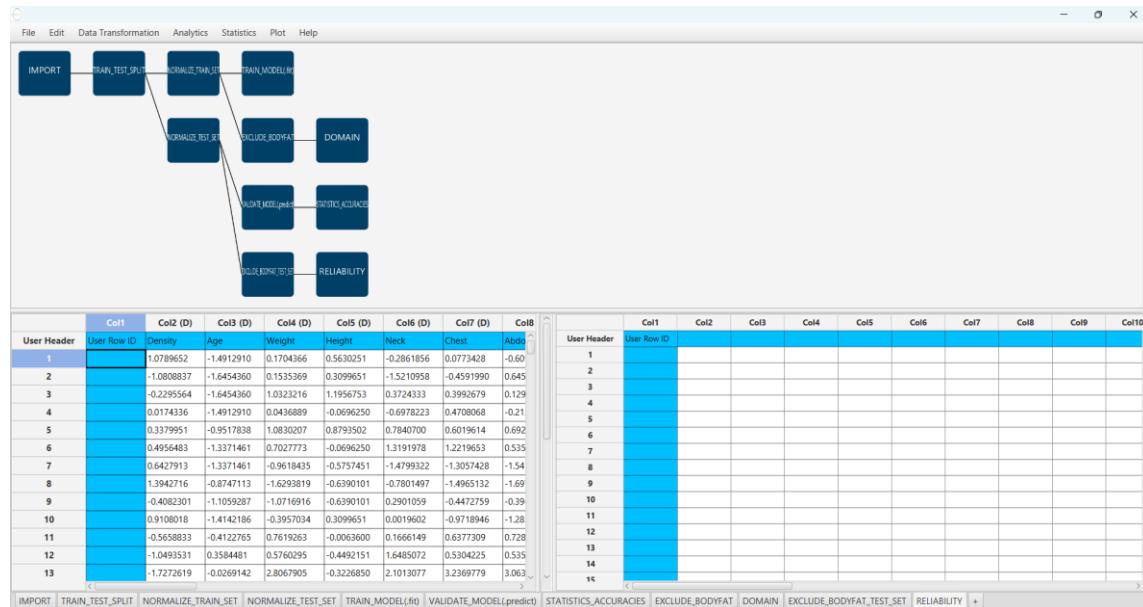
Filter the data to exclude the column that corresponds to the "BodyFat" by browsing: "Data Transformation" → "Data Manipulation" → "Select Columns". Then select all the columns except "BodyFat".



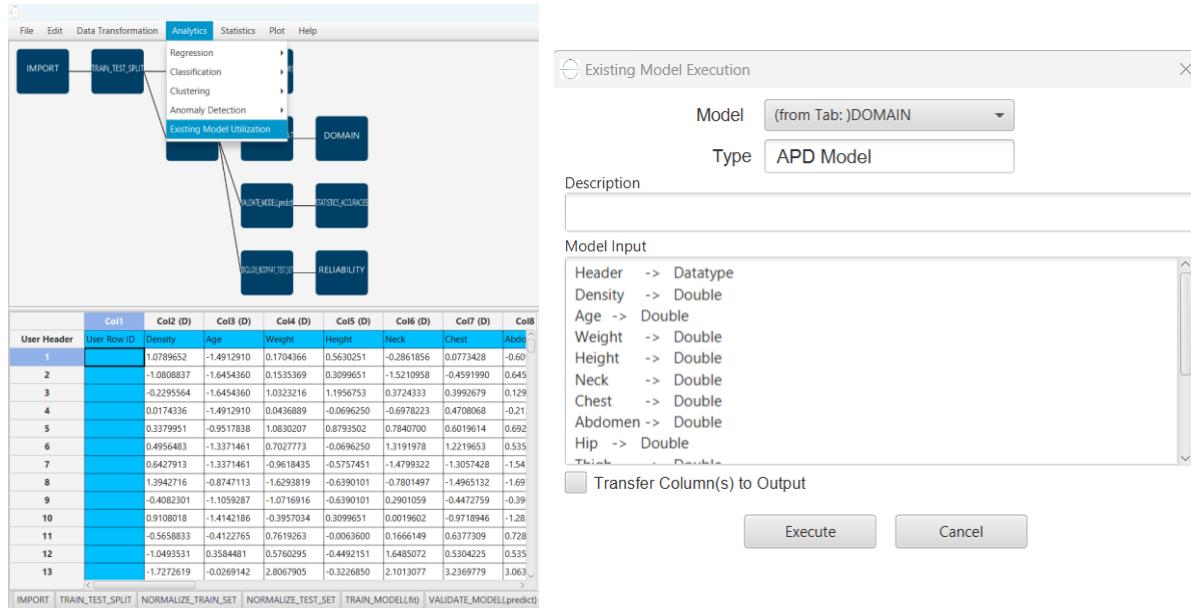
The results will appear on the output spreadsheet.

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

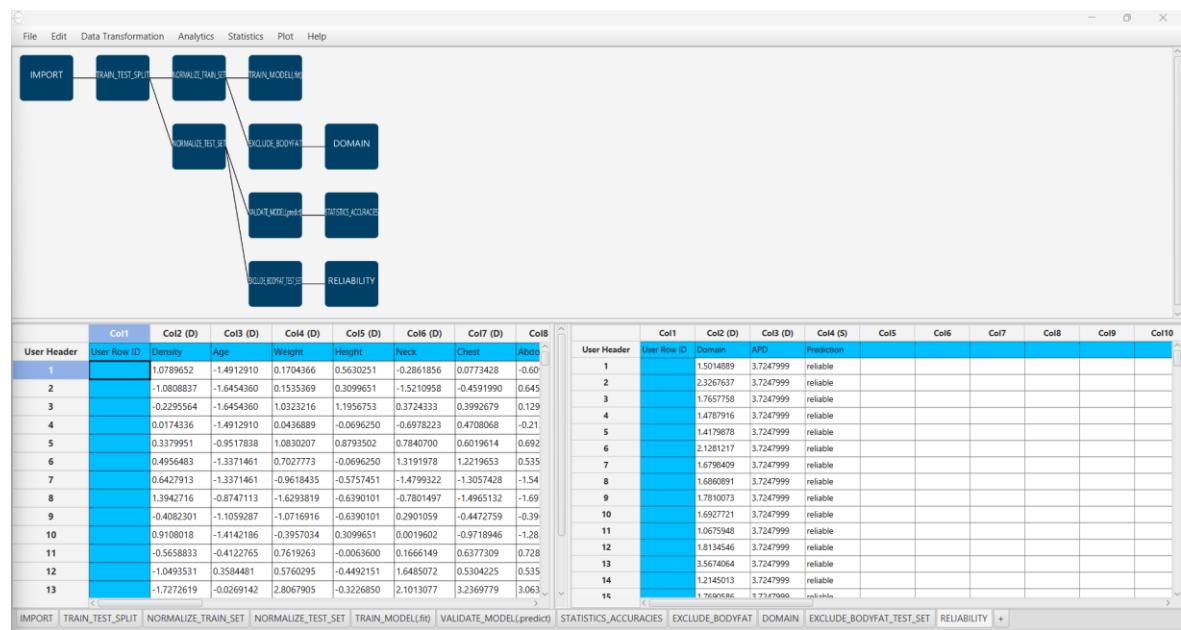
Import data into the input spreadsheet of the "RELIABILITY" tab from the output of the "EXCLUDE_BODYFAT_TEST_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".



Check the predictions' reliability by browsing: "Analytics" → "Existing Model Utilization". Then select as Model "(from Tab:) DOMAIN".



The results will appear on the output spreadsheet. We can observe that there are no unreliable samples in the test set.



Final Isalos Workflow

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

