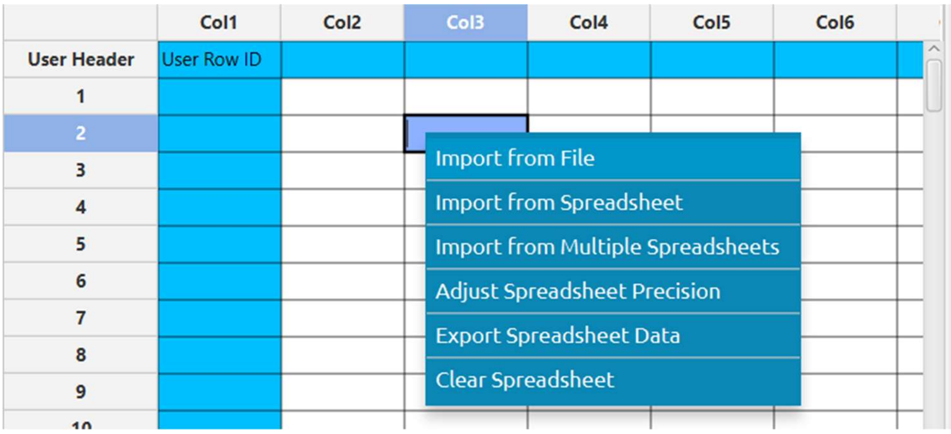
Pima Indians Diabetes Database

This dataset, which can be found in https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database, is originally from the National Institute of Diabetes and Digestive and Kidney Diseases and it contains 8 features and 768 samples. The objective of the dataset is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage. The included features are Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, Age. The target is the column "Outcome", which takes values 0 or 1 as it is a binary classification.

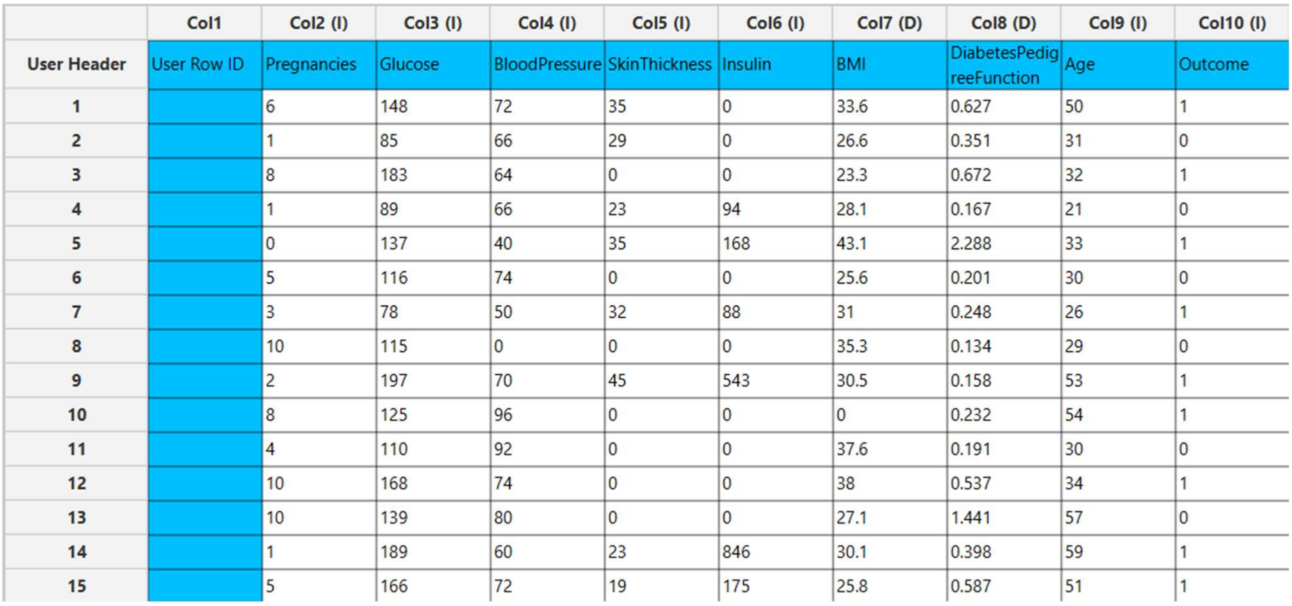
Isalos version used: 2.0.6

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 diabetes data.



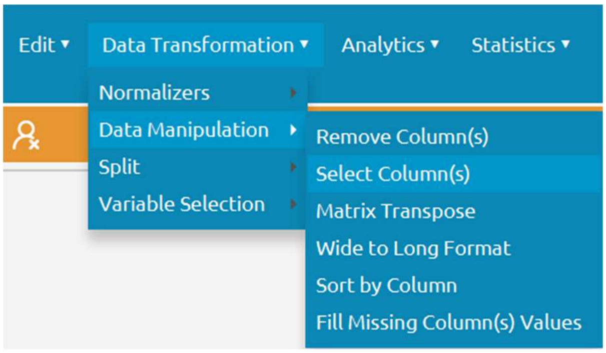
The data will appear on the left spreadsheet.

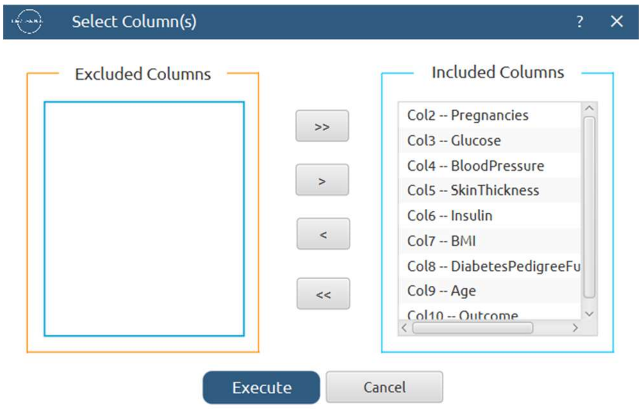


	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (D)	Col8 (D)	Col9 (I)	Col10 (I)
User Header	User Row ID	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
1		6	148	72	35	0	33.6	0.627	50	1
2		1	85	66	29	0	26.6	0.351	31	0
3		8	183	64	0	0	23.3	0.672	32	1
4		1	89	66	23	94	28.1	0.167	21	0
5		0	137	40	35	168	43.1	2.288	33	1
6		5	116	74	0	0	25.6	0.201	30	0
7		3	78	50	32	88	31	0.248	26	1
8		10	115	0	0	0	35.3	0.134	29	0
9		2	197	70	45	543	30.5	0.158	53	1
10		8	125	96	0	0	0	0.232	54	1
11		4	110	92	0	0	37.6	0.191	30	0
12		10	168	74	0	0	38	0.537	34	1
13		10	139	80	0	0	27.1	1.441	57	0
14		1	189	60	23	846	30.1	0.398	59	1
15		5	166	72	19	175	25.8	0.587	51	1

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.

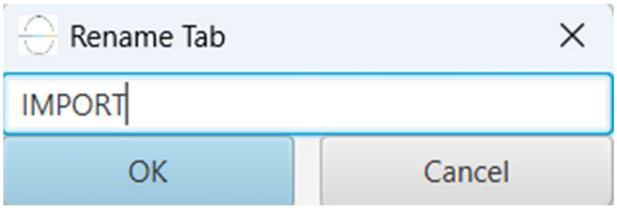




Included Columns:

- Col2 -- Pregnancies
- Col3 -- Glucose
- Col4 -- BloodPressure
- Col5 -- SkinThickness
- Col6 -- Insulin
- Col7 -- BMI
- Col8 -- DiabetesPedigreeFu
- Col9 -- Age
- Col10 -- Outcome

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.

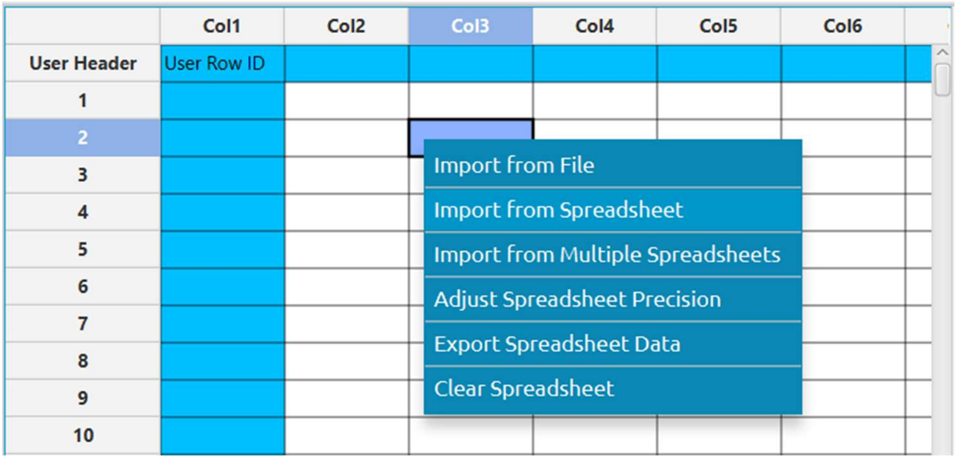


IMPORT

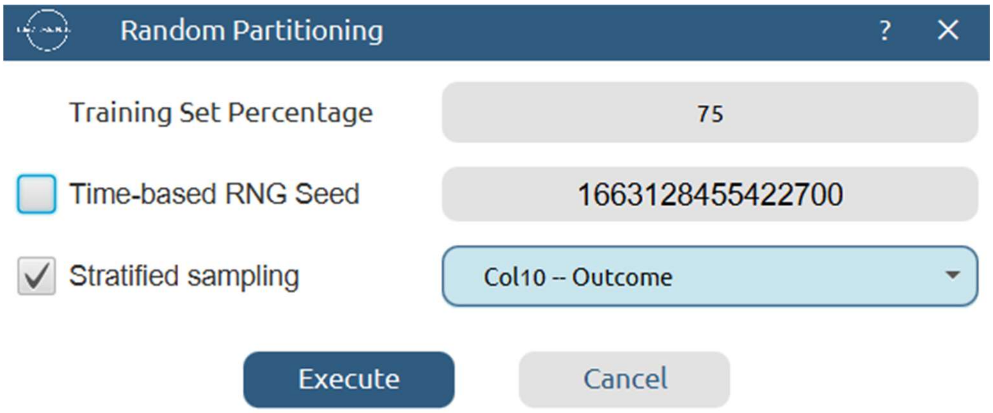
Step 3: Split data

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:



Training Set Percentage 75

☐ Time-based RNG Seed 1663128455422700

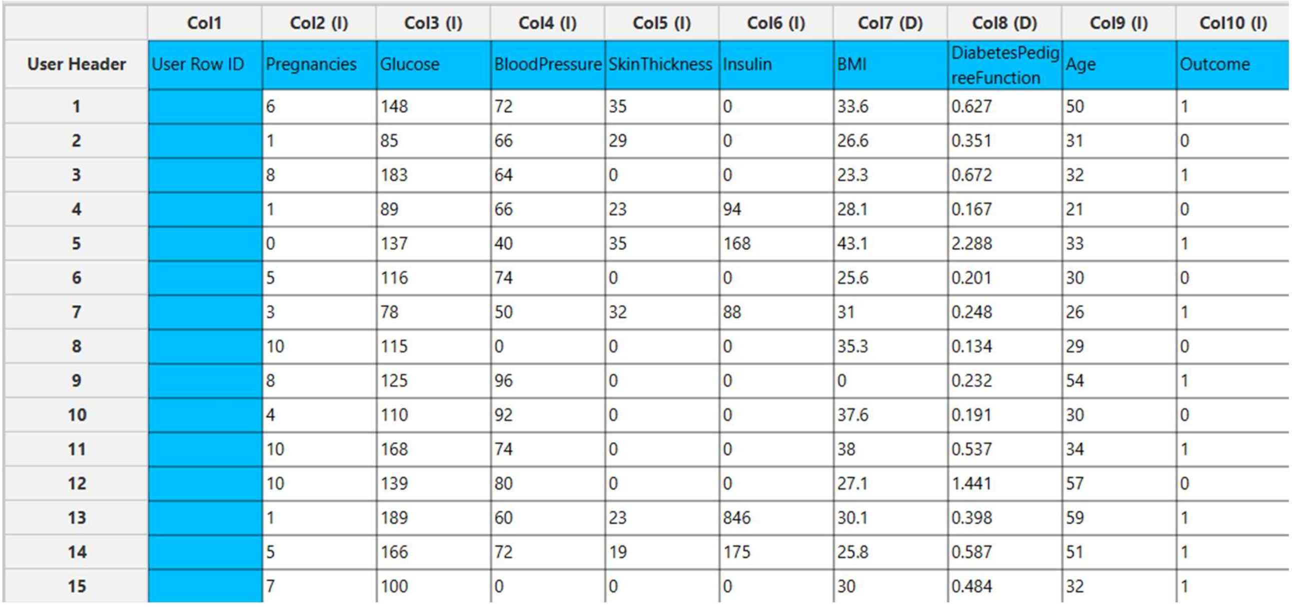
☑ Stratified sampling Col10 -- Outcome

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

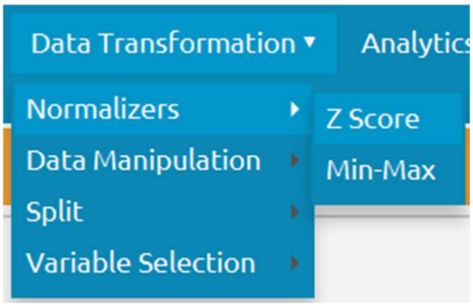
Create a new tab by pressing the "+" button on the bottom of the page with the name "NORMALIZE_TRAIN_SET".

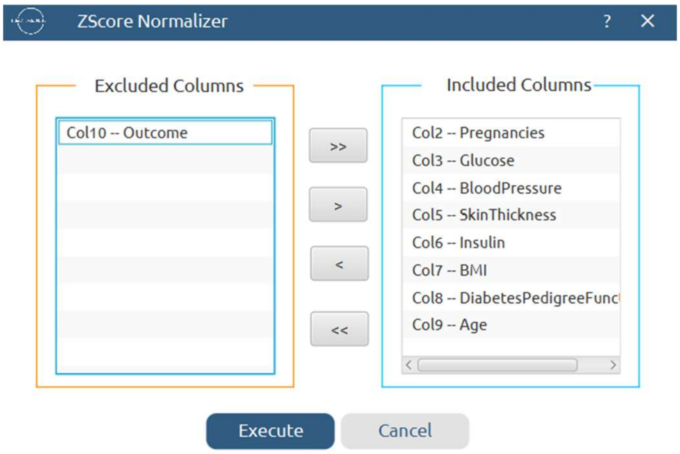
Import into the input spreadsheet of the "NORMALIZE_TRAIN_SET" tab the train set from the output of the "TRAIN_TEST_SPLIT" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet". From the available Select input tab options choose "TRAIN_TEST_SPLIT: Training Set".



	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (D)	Col8 (D)	Col9 (I)	Col10 (I)
User Header	User Row ID	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
1		6	148	72	35	0	33.6	0.627	50	1
2		1	85	66	29	0	26.6	0.351	31	0
3		8	183	64	0	0	23.3	0.672	32	1
4		1	89	66	23	94	28.1	0.167	21	0
5		0	137	40	35	168	43.1	2.288	33	1
6		5	116	74	0	0	25.6	0.201	30	0
7		3	78	50	32	88	31	0.248	26	1
8		10	115	0	0	0	35.3	0.134	29	0
9		8	125	96	0	0	0	0.232	54	1
10		4	110	92	0	0	37.6	0.191	30	0
11		10	168	74	0	0	38	0.537	34	1
12		10	139	80	0	0	27.1	1.441	57	0
13		1	189	60	23	846	30.1	0.398	59	1
14		5	166	72	19	175	25.8	0.587	51	1
15		7	100	0	0	0	30	0.484	32	1

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





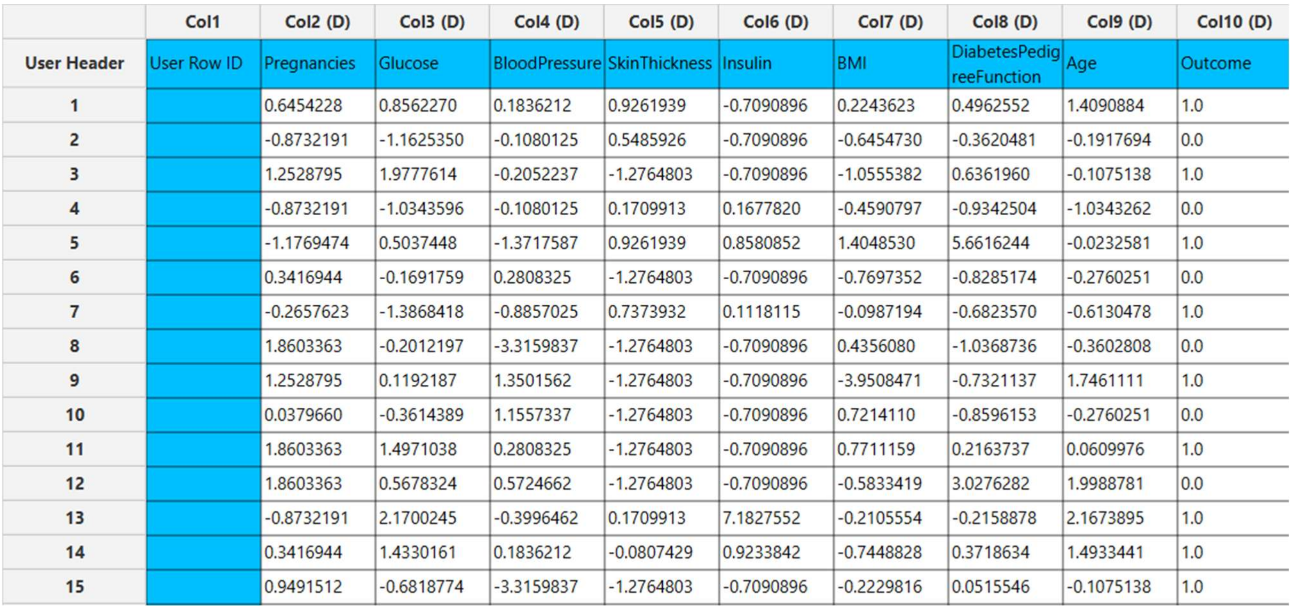
Excluded Columns

Col10 -- Outcome

Included Columns

Col2 -- Pregnancies
Col3 -- Glucose
Col4 -- BloodPressure
Col5 -- SkinThickness
Col6 -- Insulin
Col7 -- BMI
Col8 -- DiabetesPedigreeFunc
Col9 -- Age

The results will appear on the output spreadsheet.

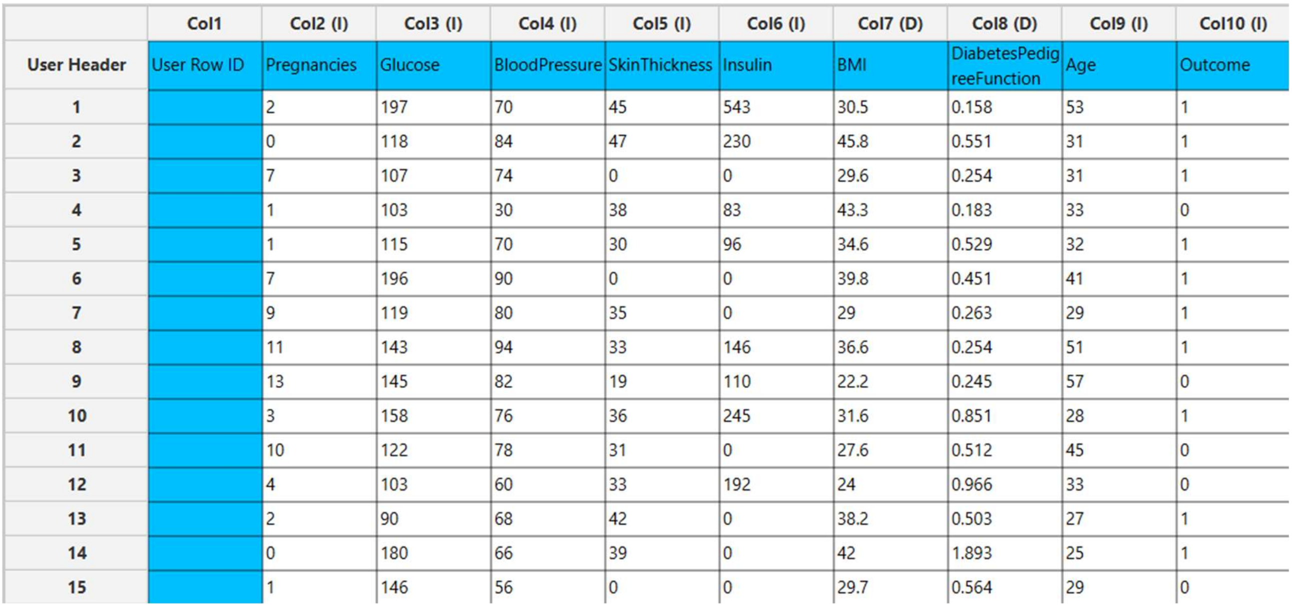


	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)	Col10 (D)
User Header	User Row ID	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
1		0.6454228	0.8562270	0.1836212	0.9261939	-0.7090896	0.2243623	0.4962552	1.4090884	1.0
2		-0.8732191	-1.1625350	-0.1080125	0.5485926	-0.7090896	-0.6454730	-0.3620481	-0.1917694	0.0
3		1.2528795	1.9777614	-0.2052237	-1.2764803	-0.7090896	-1.0555382	0.6361960	-0.1075138	1.0
4		-0.8732191	-1.0343596	-0.1080125	0.1709913	0.1677820	-0.4590797	-0.9342504	-1.0343262	0.0
5		-1.1769474	0.5037448	-1.3717587	0.9261939	0.8580852	1.4048530	5.6616244	-0.0232581	1.0
6		0.3416944	-0.1691759	0.2808325	-1.2764803	-0.7090896	-0.7697352	-0.8285174	-0.2760251	0.0
7		-0.2657623	-1.3868418	-0.8857025	0.7373932	0.1118115	-0.0987194	-0.6823570	-0.6130478	1.0
8		1.8603363	-0.2012197	-3.3159837	-1.2764803	-0.7090896	0.4356080	-1.0368736	-0.3602808	0.0
9		1.2528795	0.1192187	1.3501562	-1.2764803	-0.7090896	-3.9508471	-0.7321137	1.7461111	1.0
10		0.0379660	-0.3614389	1.1557337	-1.2764803	-0.7090896	0.7214110	-0.8596153	-0.2760251	0.0
11		1.8603363	1.4971038	0.2808325	-1.2764803	-0.7090896	0.7711159	0.2163737	0.0609976	1.0
12		1.8603363	0.5678324	0.5724662	-1.2764803	-0.7090896	-0.5833419	3.0276282	1.9988781	0.0
13		-0.8732191	2.1700245	-0.3996462	0.1709913	7.1827552	-0.2105554	-0.2158878	2.1673895	1.0
14		0.3416944	1.4330161	0.1836212	-0.0807429	0.9233842	-0.7448828	0.3718634	1.4933441	1.0
15		0.9491512	-0.6818774	-3.3159837	-1.2764803	-0.7090896	-0.2229816	0.0515546	-0.1075138	1.0

Step 5: Normalize the test set

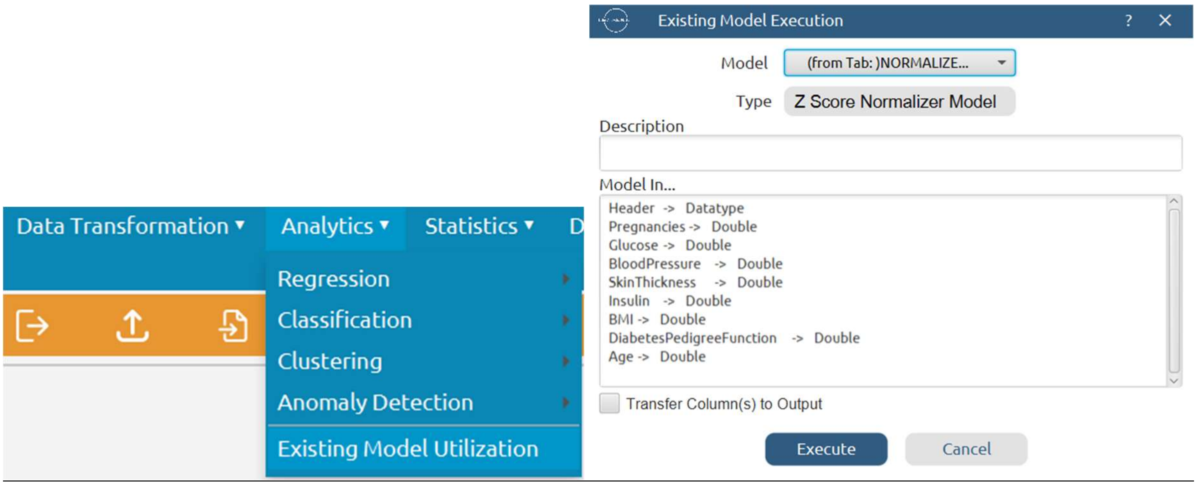
Create a new tab by pressing the "+" button on the bottom of the page with the name "NORMALIZE_TEST_SET".

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 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (D)	Col8 (D)	Col9 (I)	Col10 (I)
User Header	User Row ID	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
1		2	197	70	45	543	30.5	0.158	53	1
2		0	118	84	47	230	45.8	0.551	31	1
3		7	107	74	0	0	29.6	0.254	31	1
4		1	103	30	38	83	43.3	0.183	33	0
5		1	115	70	30	96	34.6	0.529	32	1
6		7	196	90	0	0	39.8	0.451	41	1
7		9	119	80	35	0	29	0.263	29	1
8		11	143	94	33	146	36.6	0.254	51	1
9		13	145	82	19	110	22.2	0.245	57	0
10		3	158	76	36	245	31.6	0.851	28	1
11		10	122	78	31	0	27.6	0.512	45	0
12		4	103	60	33	192	24	0.966	33	0
13		2	90	68	42	0	38.2	0.503	27	1
14		0	180	66	39	0	42	1.893	25	1
15		1	146	56	0	0	29.7	0.564	29	0

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



Model (from Tab:)NORMALIZE...

Type Z Score Normalizer Model

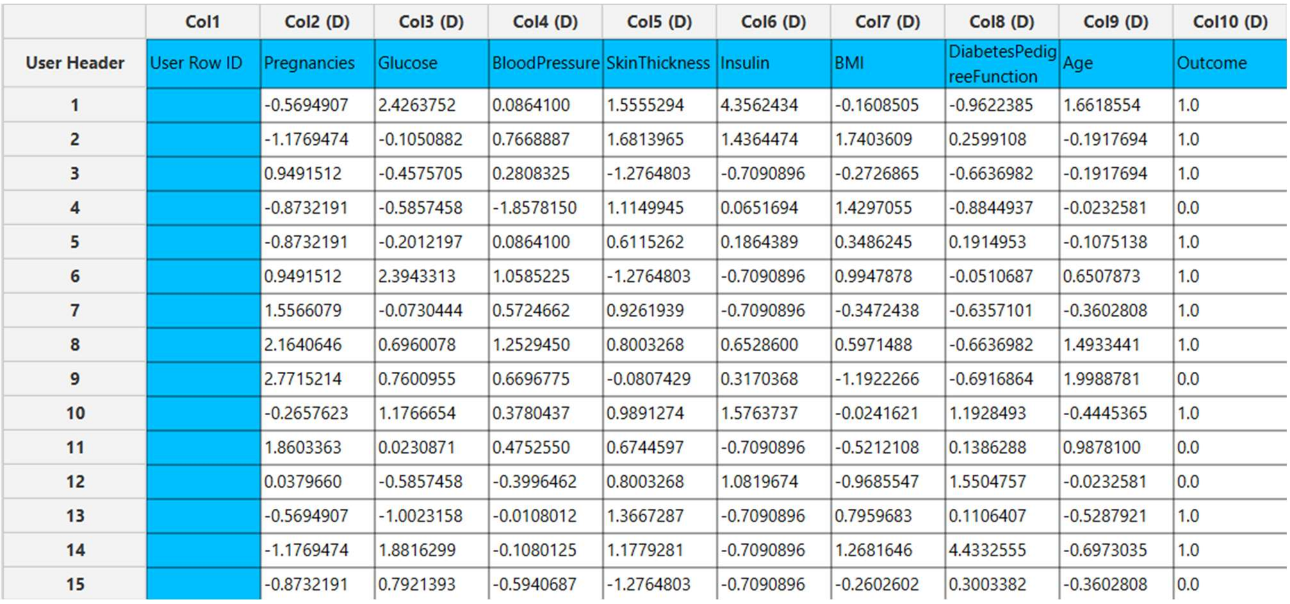
Description

Model In...

Header -> Datatype
Pregnancies -> Double
Glucose -> Double
BloodPressure -> Double
SkinThickness -> Double
Insulin -> Double
BMI -> Double
DiabetesPedigreeFunction -> Double
Age -> Double

□ Transfer Column(s) to Output

The results will appear on the output spreadsheet.



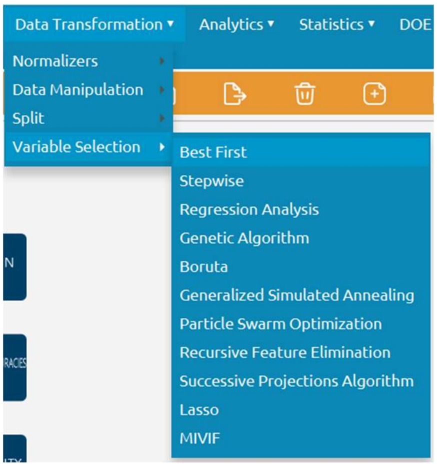
	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)	Col10 (D)
User Header	User Row ID	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
1		-0.5694907	2.4263752	0.0864100	1.5555294	4.3562434	-0.1608505	-0.9622385	1.6618554	1.0
2		-1.1769474	-0.1050882	0.7668887	1.6813965	1.4364474	1.7403609	0.2599108	-0.1917694	1.0
3		0.9491512	-0.4575705	0.2808325	-1.2764803	-0.7090896	-0.2726865	-0.6636982	-0.1917694	1.0
4		-0.8732191	-0.5857458	-1.8578150	1.1149945	0.0651694	1.4297055	-0.8844937	-0.0232581	0.0
5		-0.8732191	-0.2012197	0.0864100	0.6115262	0.1864389	0.3486245	0.1914953	-0.1075138	1.0
6		0.9491512	2.3943313	1.0585225	-1.2764803	-0.7090896	0.9947878	-0.0510687	0.6507873	1.0
7		1.5566079	-0.0730444	0.5724662	0.9261939	-0.7090896	-0.3472438	-0.6357101	-0.3602808	1.0
8		2.1640646	0.6960078	1.2529450	0.8003268	0.6528600	0.5971488	-0.6636982	1.4933441	1.0
9		2.7715214	0.7600955	0.6696775	-0.0807429	0.3170368	-1.1922266	-0.6916864	1.9988781	0.0
10		-0.2657623	1.1766654	0.3780437	0.9891274	1.5763737	-0.0241621	1.1928493	-0.4445365	1.0
11		1.8603363	0.0230871	0.4752550	0.6744597	-0.7090896	-0.5212108	0.1386288	0.9878100	0.0
12		0.0379660	-0.5857458	-0.3996462	0.8003268	1.0819674	-0.9685547	1.5504757	-0.0232581	0.0
13		-0.5694907	-1.0023158	-0.0108012	1.3667287	-0.7090896	0.7959683	0.1106407	-0.5287921	1.0
14		-1.1769474	1.8816299	-0.1080125	1.1779281	-0.7090896	1.2681646	4.4332555	-0.6973035	1.0
15		-0.8732191	0.7921393	-0.5940687	-1.2764803	-0.7090896	-0.2602602	0.3003382	-0.3602808	0.0

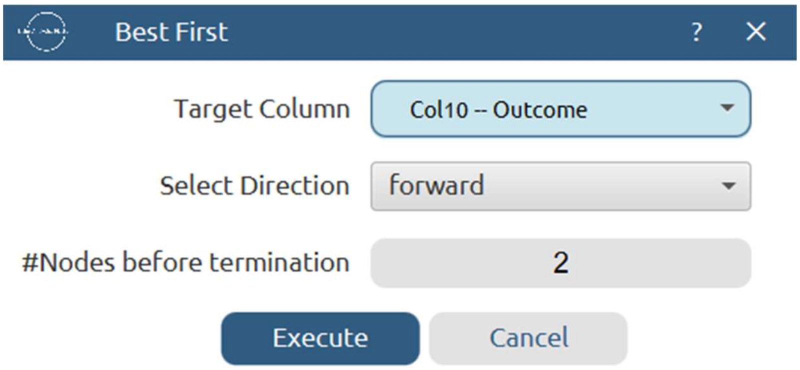
Step 6: Best First Algorithm

We want to choose the features that will be the most useful for predicting the diabetes outcome. Create a new tab by pressing the "+" button on the bottom of the page with the name "BEST_FIRST".

Import data into the input spreadsheet of the "BEST_FIRST" 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 best first algorithm by choosing: *Data Transformation → Variable Selection → Best First*



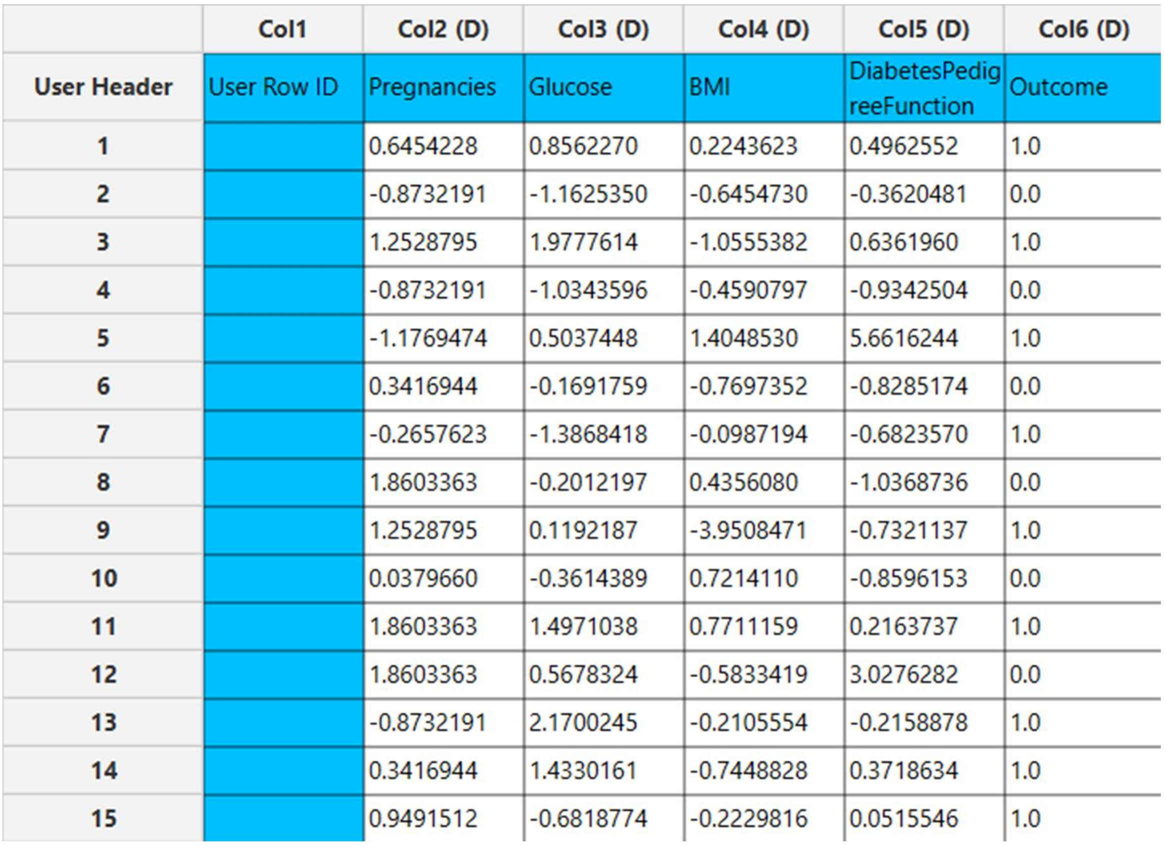


Target Column: Col10 -- Outcome

Select Direction: forward

#Nodes before termination: 2

The results will appear on the output spreadsheet.



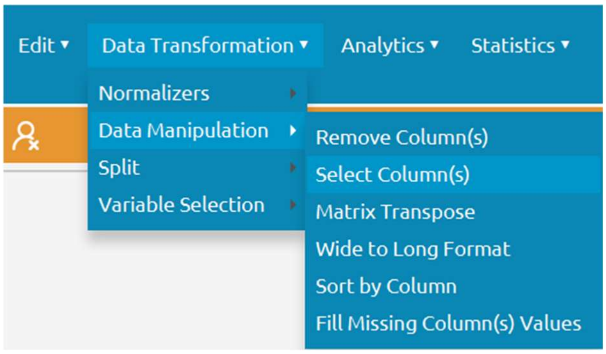
	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)
User Header	User Row ID	Pregnancies	Glucose	BMI	DiabetesPedigreeFunction	Outcome
1		0.6454228	0.8562270	0.2243623	0.4962552	1.0
2		-0.8732191	-1.1625350	-0.6454730	-0.3620481	0.0
3		1.2528795	1.9777614	-1.0555382	0.6361960	1.0
4		-0.8732191	-1.0343596	-0.4590797	-0.9342504	0.0
5		-1.1769474	0.5037448	1.4048530	5.6616244	1.0
6		0.3416944	-0.1691759	-0.7697352	-0.8285174	0.0
7		-0.2657623	-1.3868418	-0.0987194	-0.6823570	1.0
8		1.8603363	-0.2012197	0.4356080	-1.0368736	0.0
9		1.2528795	0.1192187	-3.9508471	-0.7321137	1.0
10		0.0379660	-0.3614389	0.7214110	-0.8596153	0.0
11		1.8603363	1.4971038	0.7711159	0.2163737	1.0
12		1.8603363	0.5678324	-0.5833419	3.0276282	0.0
13		-0.8732191	2.1700245	-0.2105554	-0.2158878	1.0
14		0.3416944	1.4330161	-0.7448828	0.3718634	1.0
15		0.9491512	-0.6818774	-0.2229816	0.0515546	1.0

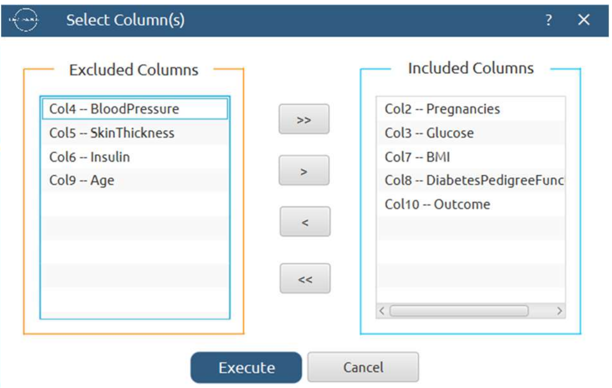
Step 7: Feature Selection: Test set

We need to select the features of the test set that the best first algorithm indicated. Create a new tab by pressing the "+" button on the bottom of the page with the name "FEATURE_SELECTION".

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

Select the columns that correspond to the important features: *Data Transformation → Data Manipulation → Select Column(s)*

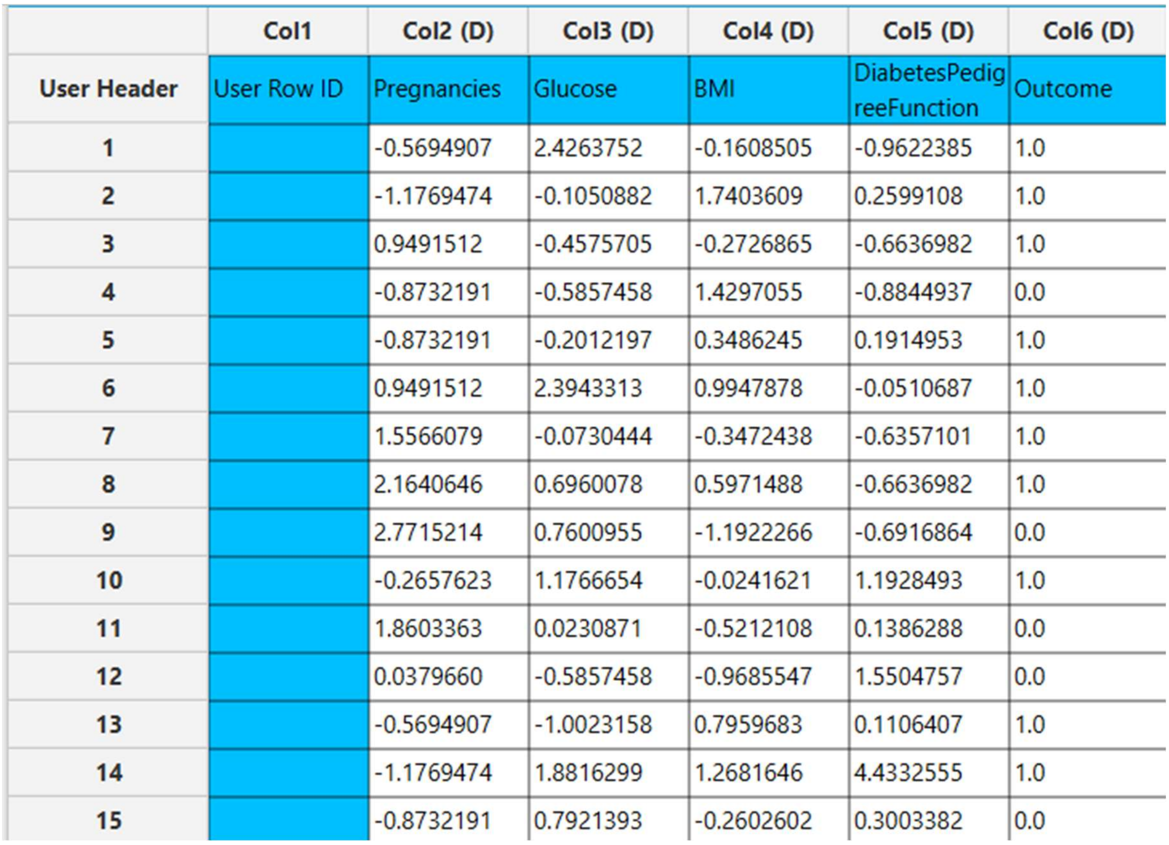




Excluded Columns
- Col4 -- BloodPressure
- Col5 -- SkinThickness
- Col6 -- Insulin
- Col9 -- Age

Included Columns
- Col2 -- Pregnancies
- Col3 -- Glucose
- Col7 -- BMI
- Col8 -- DiabetesPedigreeFunc
- Col10 -- Outcome

The results will appear on the output spreadsheet.



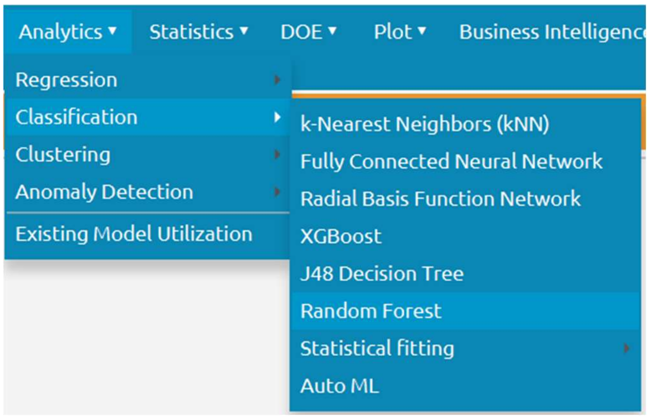
	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)
User Header	User Row ID	Pregnancies	Glucose	BMI	DiabetesPedigreeFunction	Outcome
1		-0.5694907	2.4263752	-0.1608505	-0.9622385	1.0
2		-1.1769474	-0.1050882	1.7403609	0.2599108	1.0
3		0.9491512	-0.4575705	-0.2726865	-0.6636982	1.0
4		-0.8732191	-0.5857458	1.4297055	-0.8844937	0.0
5		-0.8732191	-0.2012197	0.3486245	0.1914953	1.0
6		0.9491512	2.3943313	0.9947878	-0.0510687	1.0
7		1.5566079	-0.0730444	-0.3472438	-0.6357101	1.0
8		2.1640646	0.6960078	0.5971488	-0.6636982	1.0
9		2.7715214	0.7600955	-1.1922266	-0.6916864	0.0
10		-0.2657623	1.1766654	-0.0241621	1.1928493	1.0
11		1.8603363	0.0230871	-0.5212108	0.1386288	0.0
12		0.0379660	-0.5857458	-0.9685547	1.5504757	0.0
13		-0.5694907	-1.0023158	0.7959683	0.1106407	1.0
14		-1.1769474	1.8816299	1.2681646	4.4332555	1.0
15		-0.8732191	0.7921393	-0.2602602	0.3003382	0.0

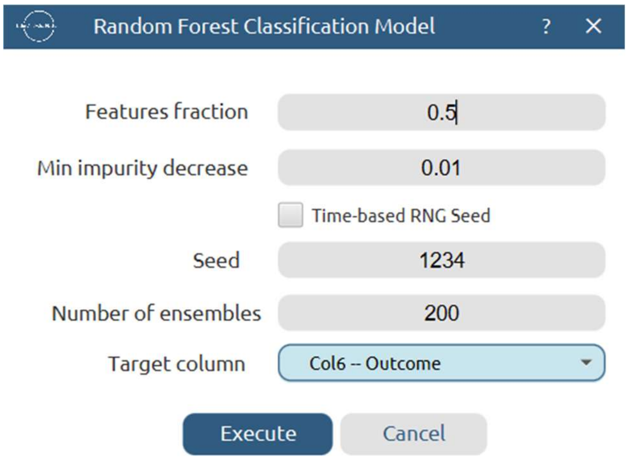
Step 8: Train the model

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 "BEST_FIRST" 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*





Features fraction 0.5

Min impurity decrease 0.01

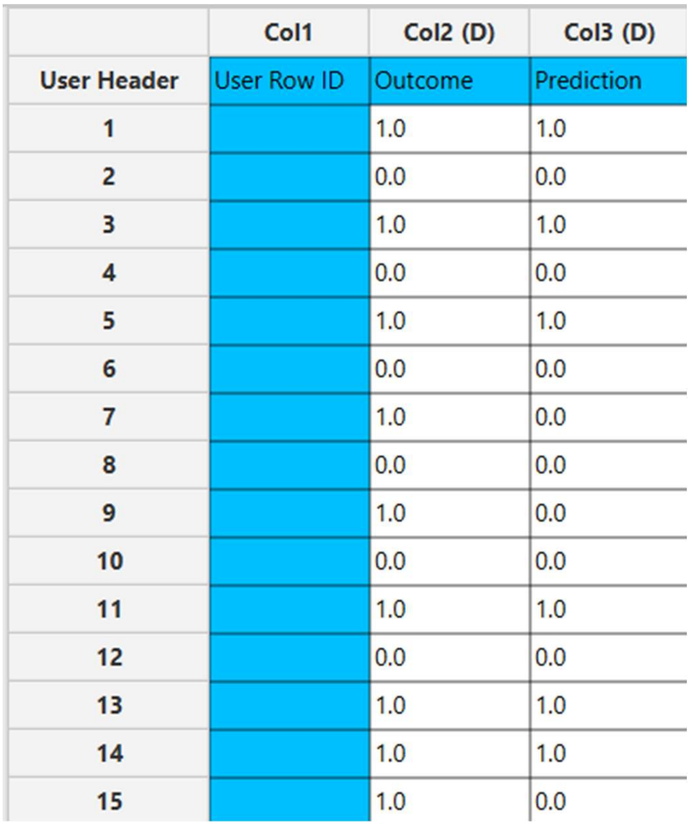
□ Time-based RNG Seed

Seed 1234

Number of ensembles 200

Target column Col6 -- Outcome

The predictions will appear on the output spreadsheet.



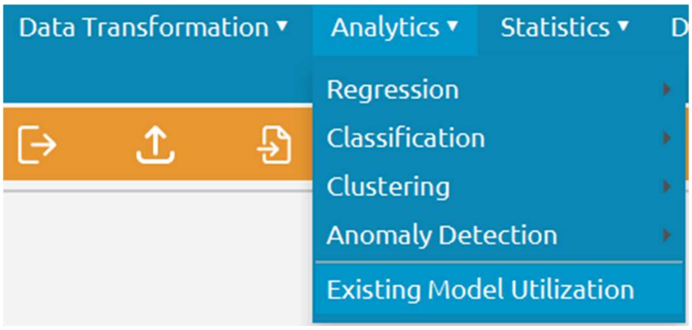
	Col1	Col2 (D)	Col3 (D)
User Header	User Row ID	Outcome	Prediction
1		1.0	1.0
2		0.0	0.0
3		1.0	1.0
4		0.0	0.0
5		1.0	1.0
6		0.0	0.0
7		1.0	0.0
8		0.0	0.0
9		1.0	0.0
10		0.0	0.0
11		1.0	1.0
12		0.0	0.0
13		1.0	1.0
14		1.0	1.0
15		1.0	0.0

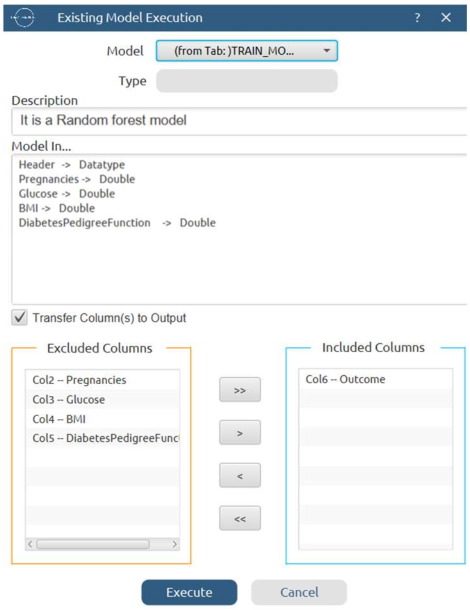
Step 9: Validate the model

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 "FEATURE_SELECTION" 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 "Outcome" to be transferred to the output spreadsheet.





Model (from Tab:)TRAIN_MO...

Type

Description

It is a Random forest model

Model In...

Header -> Datatype
Pregnancies -> Double
Glucose -> Double
BMI -> Double
DiabetesPedigreeFunction -> Double

☑ Transfer Column(s) to Output

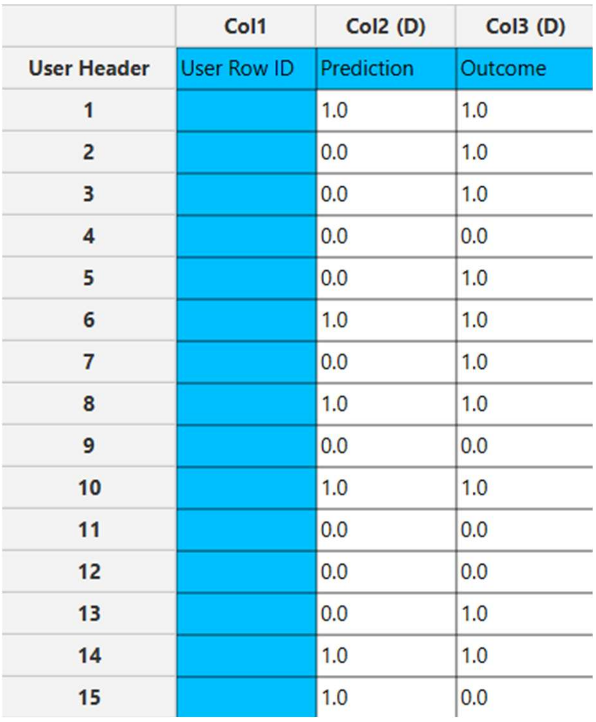
Excluded Columns

Col2 -- Pregnancies
Col3 -- Glucose
Col4 -- BMI
Col5 -- DiabetesPedigreeFunc

Included Columns

Col6 -- Outcome

The predictions will appear on the output spreadsheet.



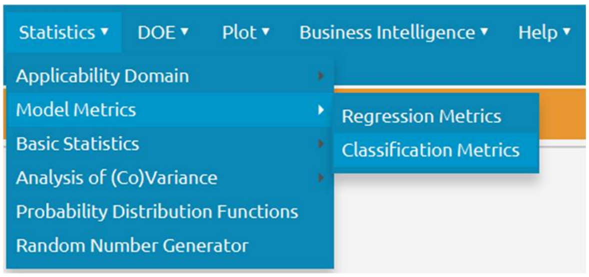
	Col1	Col2 (D)	Col3 (D)
User Header	User Row ID	Prediction	Outcome
1		1.0	1.0
2		0.0	1.0
3		0.0	1.0
4		0.0	0.0
5		0.0	1.0
6		1.0	1.0
7		0.0	1.0
8		1.0	1.0
9		0.0	0.0
10		1.0	1.0
11		0.0	0.0
12		0.0	0.0
13		0.0	1.0
14		1.0	1.0
15		1.0	0.0

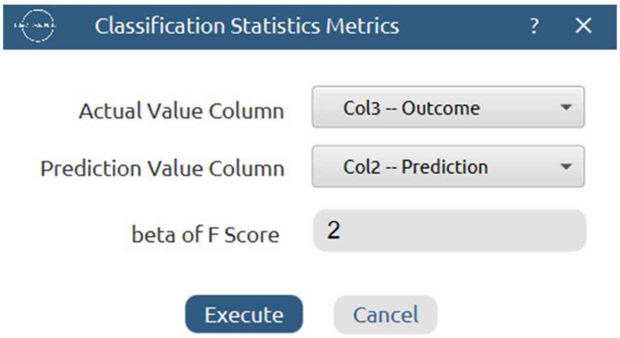
Step 10: Statistics calculation

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*



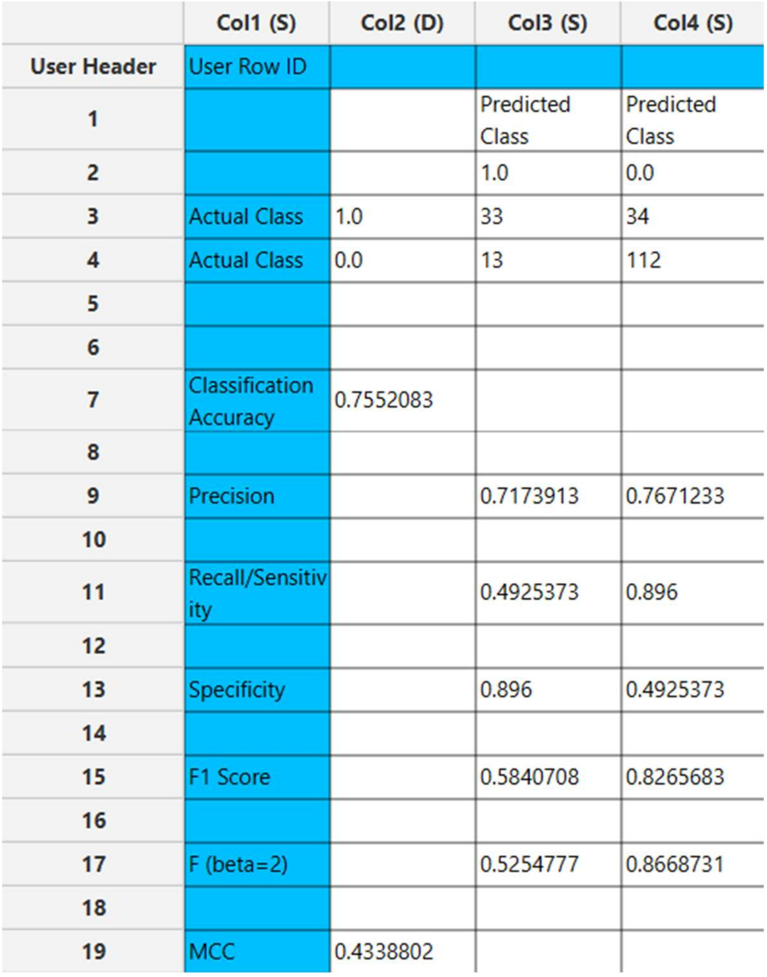


Actual Value Column: Col3 -- Outcome

Prediction Value Column: Col2 -- Prediction

beta of F Score: 2

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			1.0	0.0
3	Actual Class	1.0	33	34
4	Actual Class	0.0	13	112
5				
6				
7	Classification Accuracy	0.7552083		
8				
9	Precision		0.7173913	0.7671233
10				
11	Recall/Sensitivity		0.4925373	0.896
12				
13	Specificity		0.896	0.4925373
14				
15	F1 Score		0.5840708	0.8265683
16				
17	F (beta=2)		0.5254777	0.8668731
18				
19	MCC	0.4338802		

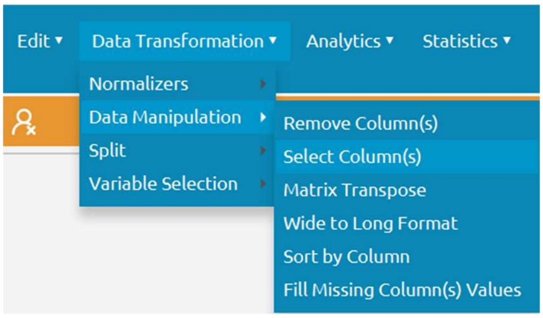
Step 11: Reliability check for each record of the test set

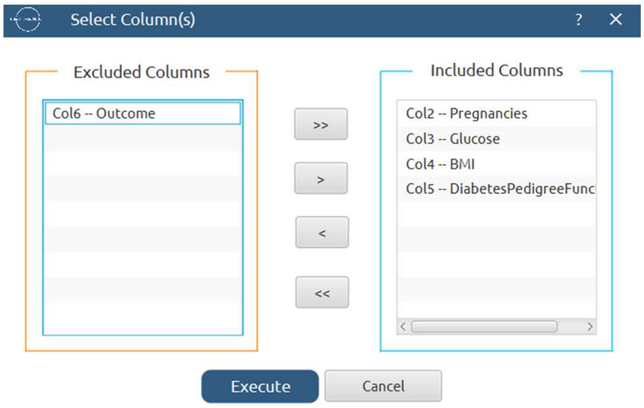
Step 11.a: Create the domain

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

Import data into the input spreadsheet of the "EXCLUDE_OUTCOME" tab from the output of the "BEST_FIRST" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet".

Manipulate the data to exclude the target column "Outcome": *Data Transformation → Data Manipulation → Select Column(s)*





Excluded Columns

Col6 -- Outcome

Included Columns

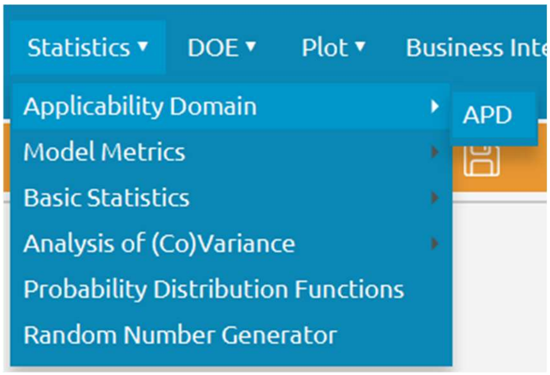
Col2 -- Pregnancies
Col3 -- Glucose
Col4 -- BMI
Col5 -- DiabetesPedigreeFunc

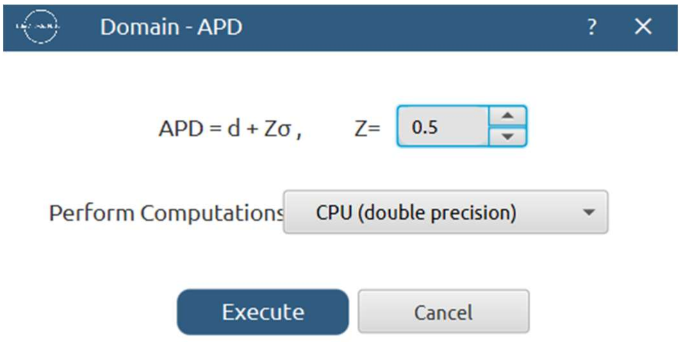
The results will appear on the output spreadsheet.

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_OUTCOME" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet".

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

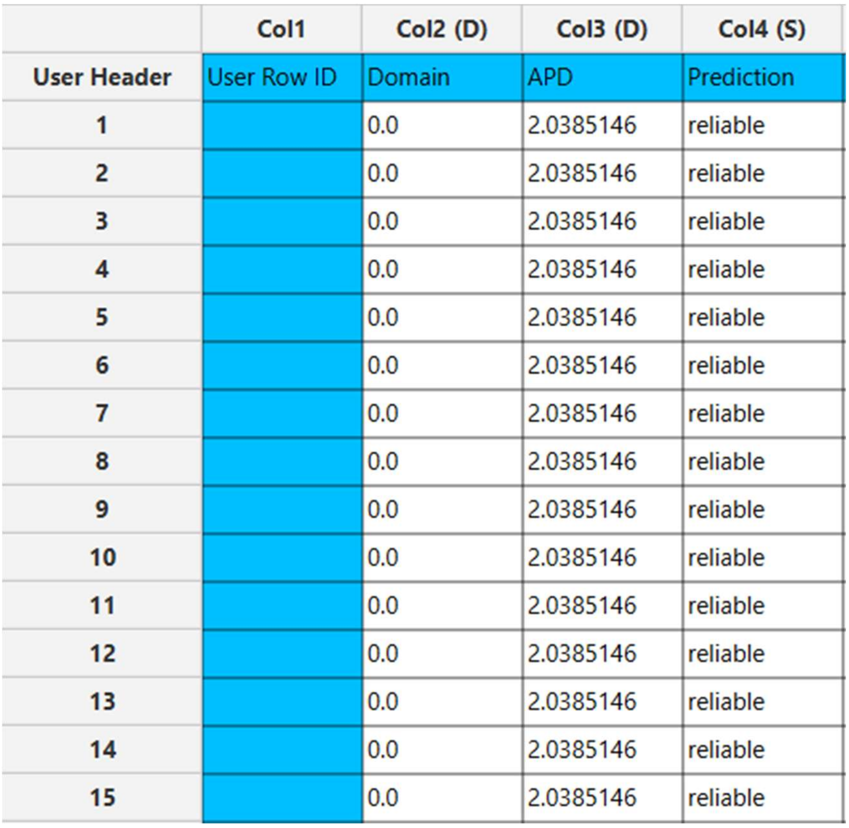




APD = d + Zσ , Z= 0.5

Perform Computations CPU (double precision)

The results will appear on the output spreadsheet.



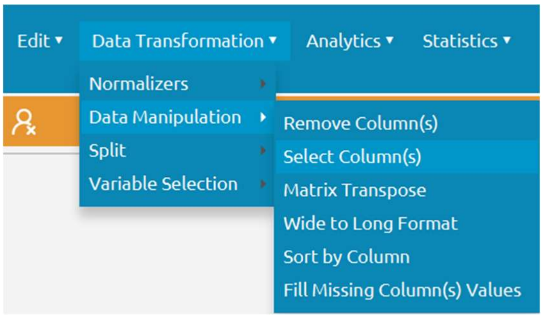
	Col1	Col2 (D)	Col3 (D)	Col4 (S)
User Header	User Row ID	Domain	APD	Prediction
1		0.0	2.0385146	reliable
2		0.0	2.0385146	reliable
3		0.0	2.0385146	reliable
4		0.0	2.0385146	reliable
5		0.0	2.0385146	reliable
6		0.0	2.0385146	reliable
7		0.0	2.0385146	reliable
8		0.0	2.0385146	reliable
9		0.0	2.0385146	reliable
10		0.0	2.0385146	reliable
11		0.0	2.0385146	reliable
12		0.0	2.0385146	reliable
13		0.0	2.0385146	reliable
14		0.0	2.0385146	reliable
15		0.0	2.0385146	reliable

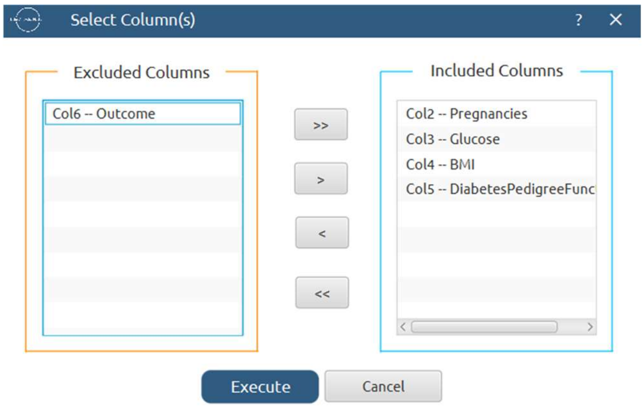
Step 11.b: Check the test set reliability

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

Import data into the input spreadsheet of the "EXCLUDE_OUTCOME_TEST_SET" tab from the output of the "FEATURE_SELECTION" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet".

Manipulate the data to exclude the target column "Outcome": *Data Transformation → Data Manipulation → Select Column(s)*





Excluded Columns

Col6 -- Outcome

Included Columns

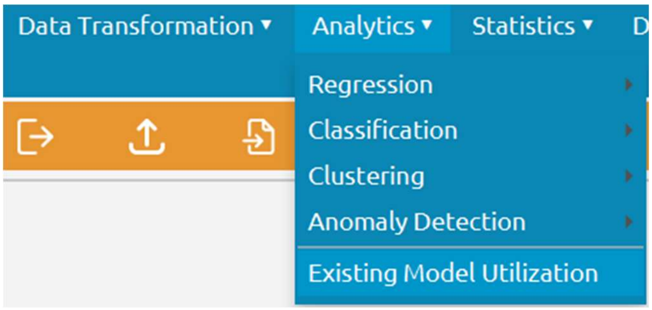
Col2 -- Pregnancies
Col3 -- Glucose
Col4 -- BMI
Col5 -- DiabetesPedigreeFunc

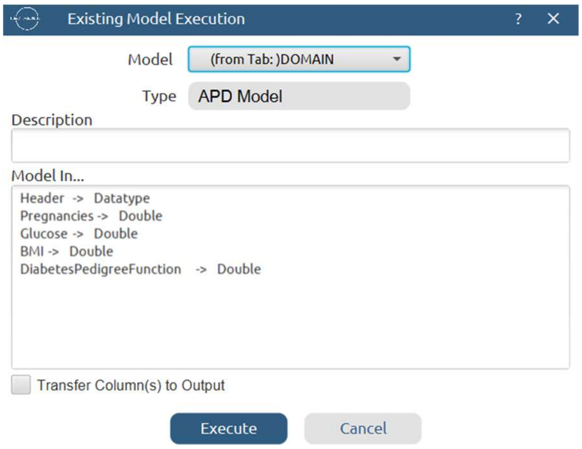
The results will appear on the output spreadsheet.

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_OUTCOME_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*





Model (from Tab:)DOMAIN

Type APD Model

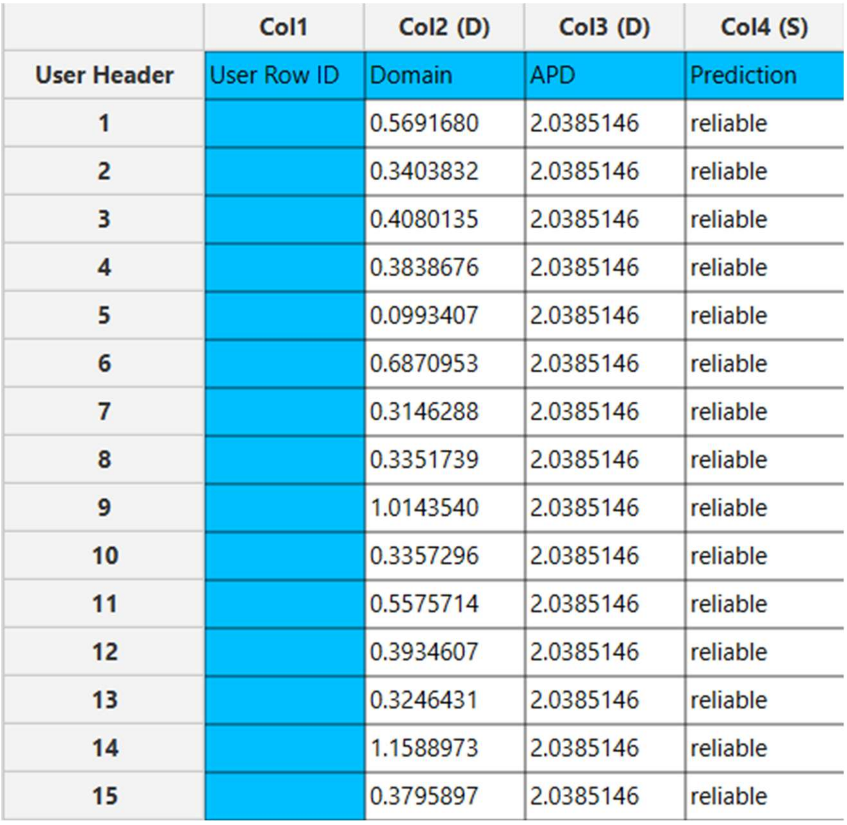
Description

Model In...

```
Header -> Datatype
Pregnancies -> Double
Glucose -> Double
BMI -> Double
DiabetesPedigreeFunction -> Double
```

□ Transfer Column(s) to Output

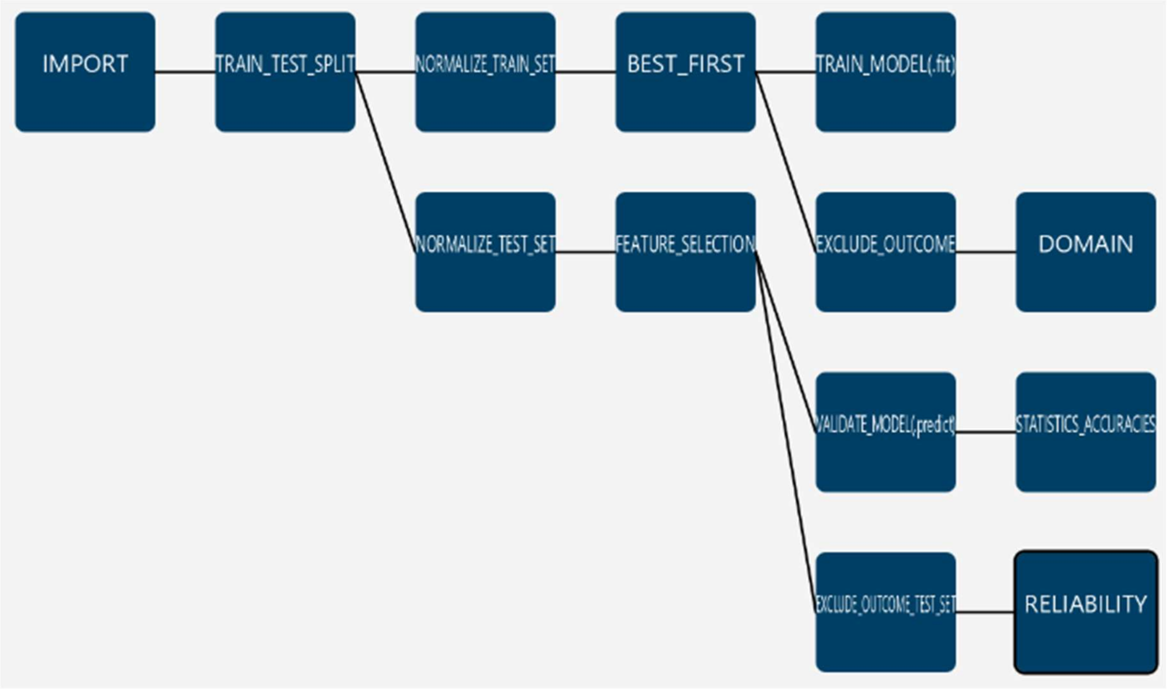
The results will appear on the output spreadsheet.



	Col1	Col2 (D)	Col3 (D)	Col4 (S)
User Header	User Row ID	Domain	APD	Prediction
1		0.5691680	2.0385146	reliable
2		0.3403832	2.0385146	reliable
3		0.4080135	2.0385146	reliable
4		0.3838676	2.0385146	reliable
5		0.0993407	2.0385146	reliable
6		0.6870953	2.0385146	reliable
7		0.3146288	2.0385146	reliable
8		0.3351739	2.0385146	reliable
9		1.0143540	2.0385146	reliable
10		0.3357296	2.0385146	reliable
11		0.5575714	2.0385146	reliable
12		0.3934607	2.0385146	reliable
13		0.3246431	2.0385146	reliable
14		1.1588973	2.0385146	reliable
15		0.3795897	2.0385146	reliable

Final Isalos Workflow

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


IMPORT
TRAIN_TEST_SPLIT
NORMALIZE_TRAIN_SET
BEST_FIRST
TRAIN_MODEL(.fit)
NORMALIZE_TEST_SET
FEATURE_SELECTION
EXCLUDE_OUTCOME
DOMAIN
VALIDATE_MODEL(.predict)
STATISTICS_ACCURACIES
EXCLUDE_OUTCOME_TEST_SET
RELIABILITY