



# Parkinson's Disease (Binary Classification)

The goal of this study is to train a model in order to predict whether a patient has Parkinson's Disease or not. The dataset used in this case study is found in <https://www.kaggle.com/datasets/rabieelkharoua/parkinsons-disease-dataset-analysis/data>

and has 35 features and 2104 labelled samples. This dataset comprises comprehensive health information for 2,105 patients diagnosed with Parkinson's Disease, each uniquely identified with IDs ranging from 3058 to 5162. The dataset includes demographic details, lifestyle factors, medical history, clinical measurements, cognitive and functional assessments, symptoms, and a diagnosis indicator.

The dataset contains no missing values and includes several categorical features. Some of these features represent binary yes/no data, encoded as 0 for "No" and 1 for "Yes". Additionally, other categorical features contain multiple levels with corresponding numeric codes, as detailed below:

Gender:

- Male (0)
- Female (1)

Ethnicity:

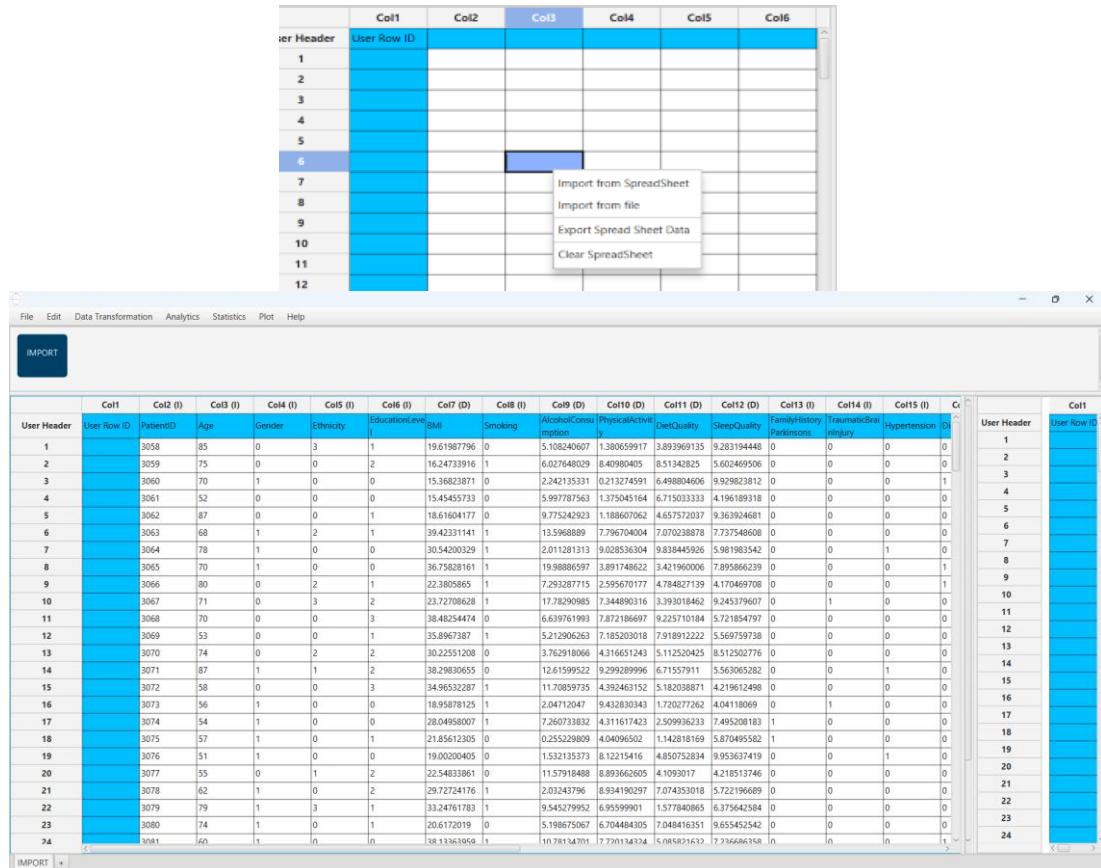
- Caucasian (0)
- African American (1)
- Asian (2)
- Other (3)

Education Level:

- None (0)
- High School (1)
- Bachelor's (2)
- Higher (3)

## 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 Parkinson's Disease data.



The screenshot shows a data spreadsheet interface. A context menu is open over the 6th row of the first column, containing options: "Import from Spreadsheet", "Import from file", "Export Spread Sheet Data", and "Clear SpreadSheet". The spreadsheet has columns labeled Col1 through Col16. The first row is labeled "User Header". The data rows contain various numerical values and some categorical entries like "n" and "1". The right side of the interface shows a vertical list of rows from 1 to 24, with "User Header" and "Col1" highlighted.

## Step 2: Manipulate data

In order to use the data for training we have to exclude any columns that do not contain features, like the "PatientID" and "DoctorInCharge" columns. We follow these steps to execute this:

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

- Select all columns except the ones that corresponds to the "PatientID" and "DoctorInCharge" columns.

The data without the "PatientID" and "DoctorInCharge" columns will appear in the output spreadsheet.

## Step 3: Split data

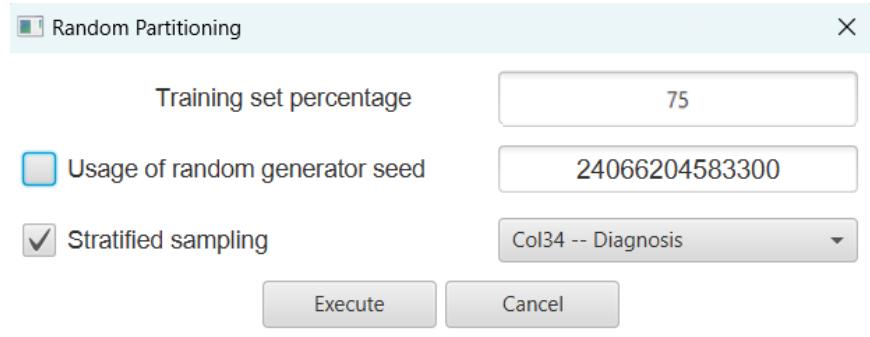
Create a new tab by pressing the "+" button on the bottom of the page with the name "TRAIN\_TEST\_SPLIT" which we will use for splitting to create the train and test set.

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".

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 appear on the output spreadsheet.

## Step 4: Normalize the training set

Create a new tab by pressing the "+" button on the bottom of the page with the name "NORMALISE\_TRAIN\_SET".

Import data into the input spreadsheet of the "NORMALISE\_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. 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 NORMALISE\_TRAIN\_SET buttons. The TRAIN\_TEST\_SPLIT tab is currently selected, showing a table with 24 rows and columns labeled Col1 through Col9. The NORMALISE\_TRAIN\_SET tab is also visible below the main window.

Normalize the data using Z-score:

"Data Transformation" → "Normalizers" → "Z-Score"

Then select all columns and click "Execute".

The screenshot shows the Data Transformation module. In the top navigation bar, 'Data Transformation' is selected. Under the 'Normalizers' section, 'Z Score' is chosen. A dialog box titled 'ZScore Normalizer' is open, showing a list of columns. On the left, 'Excluded Columns' is listed as 'Col34 -- Diagnosis'. On the right, 'Included Columns' is listed as 'Col26 -- FunctionalAssessment', 'Col27 -- Tremor', 'Col28 -- Rigidity', 'Col29 -- Bradykinesia', 'Col30 -- PosturalInstability', 'Col31 -- SpeechProblems', 'Col32 -- SleepDisorders', and 'Col33 -- Constipation'. At the bottom of the dialog are 'Execute' and 'Cancel' buttons.

The results will appear on the output spreadsheet.

IMPORT TRAIN\_TEST\_SPLIT NORMALISE\_TRAIN\_SET

User Header User Row ID Age Gender Ethnicity EducationLevel BMI Smoking AlcoholConsumption

1 85 0 3 1 19.61987796 0 5.1082400

2 75 0 0 2 16.24733916 1 6.0276480

3 52 0 0 0 15.45455733 0 5.9977875

4 67 0 0 1 18.61604177 0 9.7752425

5 78 1 0 0 30.54200329 1 2.0112815

6 70 1 0 0 36.75828161 1 19.98865

7 80 0 0 1 22.3805865 1 7.2928787

8 71 0 3 2 23.72708628 1 17.782905

9 70 0 0 3 38.42544744 0 6.6397615

10 74 0 2 2 30.2551206 0 3.7629180

11 67 1 1 2 38.29830655 0 12.615995

12 56 1 0 0 18.9578125 1 2.0471204

13 54 1 0 0 28.04958007 1 7.2607336

14 57 1 0 1 21.85612305 0 0.2552296

15 55 0 1 2 22.54833861 0 11.579184

16 62 1 0 2 29.72724176 0 2.0324375

17 79 1 3 1 33.24761783 1 9.5452795

18 74 1 0 1 20.6172019 0 5.198675C

19 60 1 0 0 38.13363959 1 10.781347

20 71 1 2 1 15.86360295 0 19.591718

21 79 1 1 2 36.90534340 0 9.8905975

22 66 0 0 2 23.88907318 0 11.423196

23 78 0 1 3 25.89636742 0 17.904796C

24 61 0 0 2 36.76177473 1 2.6088171

IMPORT TRAIN\_TEST\_SPLIT NORMALISE\_TRAIN\_SET +

User Header User Row ID Age Gender Ethnicity EducationLevel BMI Smoking AlcoholConsumption

1 13195736895 -0.990231301 2.2644699572 -0.401748895 -1.086608883 -0.666607948 -0.864918

2 44417 2191987 03018 3424679 1208031 9494684 2792321

3 04561887577 -0.990231301 0.693013965 0.7092169641 -1.560542520 1.4991820744 -0.704669

4 950968 2191987 3971186 071937 0707939 892366 6137969

5 -1.52595965 -0.990231301 0.693013965 -1.512714754 -1.671902964 -0.666607948 -0.709871

6 2283393 2191987 3971186 7921294 5032697 9494684 027559

7 14922506758 -0.990231301 0.693013965 -0.401748895 -1.227815702 -0.666607948 -0.051890

8 942808 2191987 3971186 3424679 4317398 9494684 85781620

9 07109442373 1092255077 0.693013965 1.512714754 0.401748895 1.086608883 1.404265

10 186929 64324 3971186 7921294 4647526 892366 51358

11 03044692919 1092255077 -0.693013965 1.512714754 1.2305976730 1.4991820744 1.717191

12 20436786 64324 3971186 7921294 3786635 892366 051028

13 08878812236 -0.990231301 1.2786419830 0.401748895 -0.699017793 1.4991820744 0.484212

14 679569 2191987 02972 3424679 152291 892366 9587895

15 0.1108347850 -0.990231301 2.2644699572 0.7092169641 -0.509877719 1.4991820744 1.342934

16 953688 2191987 03018 071937 231788 892366 2081

17 0.0244962919 -0.990231301 0.693013965 1.5267913931 -0.666607948 -0.598048

18 20436786 2191987 3971186 568551 657111 9494684 8826121

19 0.3698502646 -0.990231301 1.2786419830 0.7092169641 0.402482800 -0.666607948 -1.099154

20 2501917 2191987 02972 3424679 152291 892366 4439781

21 1.4922506758 1092255077 0.693013965 1.512714754 0.401748895 -0.72688102 -0.666607948 -1.710145

22 942808 64324 029674 071937 155732 9494684 2853529

23 7035434 2191987 0292674 071937 888083 9494684 404000

24 0.686211653 1092255077 -0.693013965 0.7092169641 0.401748895 1.404265

16 4790193 64324 3971186 071937 477253 892366 3442428

17 0.8015427304 1092255077 2.2644699572 -0.401748895 0.8274519022 1.4991820744 -0.091946

IMPORT TRAIN\_TEST\_SPLIT NORMALISE\_TRAIN\_SET NORMALISE\_TEST\_SET +

## Step 5: Normalize the test set

Create a new tab by pressing the "+" button on the bottom of the page with the name "NORMALISE\_TEST\_SET".

Import data into the input spreadsheet of the "NORMALISE\_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".

IMPORT TRAIN\_TEST\_SPLIT NORMALISE\_TRAIN\_SET

User Header User Row ID Age Gender Ethnicity EducationLevel BMI Smoking PhysicalActivity

1 70 1 0 1 15.36823871 0 2.242135331 0.2132745

2 68 1 2 1 39.42331141 1 13.5968889 7.7967040

3 53 0 0 1 35.9867387 1 5.212906263 7.1852030

4 58 0 0 3 34.96532287 1 11.70859735 4.3924631

5 51 1 0 0 19.0020405 0 1.532135373 8.1221541

6 63 1 1 3 25.3263836 0 1.094654253 8.0915424

7 56 0 1 0 16.11707674 0 4.473425263 9.6204920

8 69 1 1 1 18.7230873 0 0.543629553 9.3659498

9 72 1 0 1 29.65370805 1 5.418741531 1.0356313

10 60 0 0 0 31.93604901 0 8.241282371 7.9181040

11 73 1 0 1 30.70591465 0 16.91863358 4.2184886

12 58 1 0 1 17.5221794 0 11.6197039 6.9720368

13 58 1 1 1 17.61243635 0 8.239722525 1.9625576

14 75 0 3 1 30.00437964 0 3.737186934 6.2632734

15 57 1 1 1 15.84157373 1 17.09450292 5.1732513

16 72 0 3 1 16.02239896 0 6.469157837 3.3422069

17 50 0 1 3 32.01090917 0 0.606812251 2.5999024

18 55 1 3 0 15.7567057 0 12.02304347 1.8164269

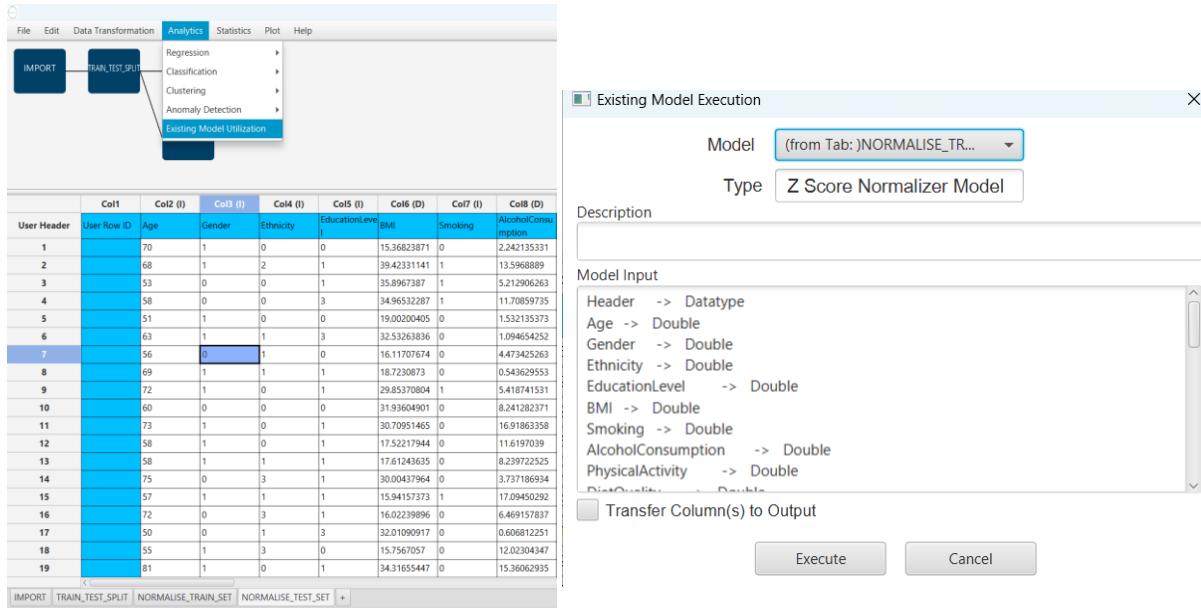
19 81 1 0 1 34.31655447 0 15.36062935 0.8390699

IMPORT TRAIN\_TEST\_SPLIT NORMALISE\_TRAIN\_SET NORMALISE\_TEST\_SET +

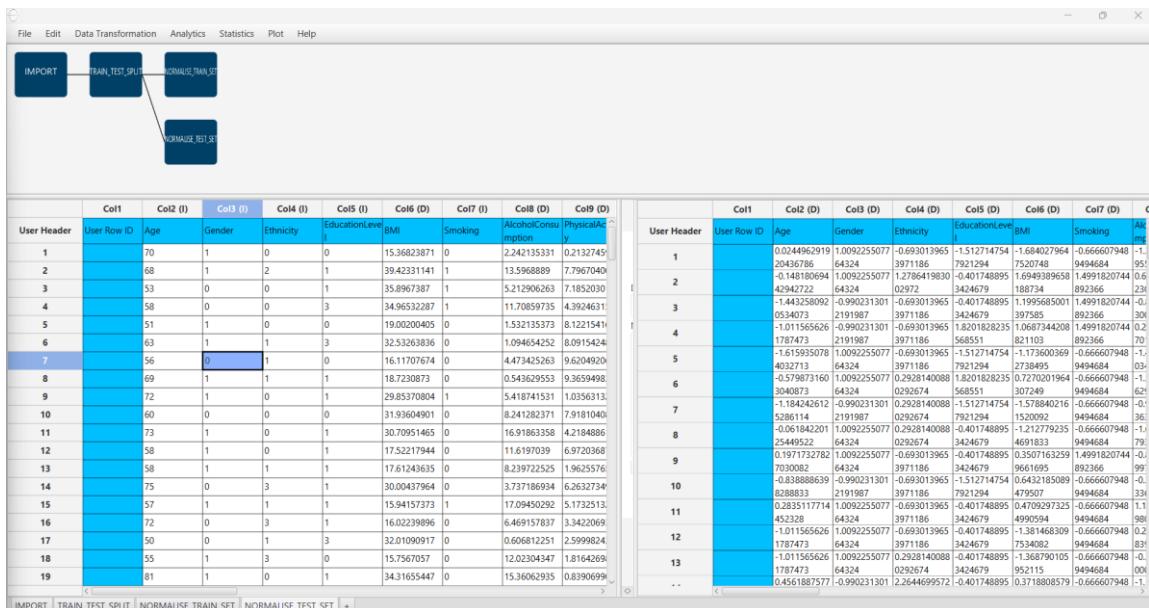
User Header User Row ID Age Gender Ethnicity EducationLevel BMI Smoking PhysicalActivity

Normalize the test set using the existing normalizer of the training set:

"Analytics" → "Existing Model Utilization" → "Model (from Tab: ) NORMALISE\_TRAIN\_SET".



The results will appear on the output spreadsheet.



## Step 6: Feature selection

Create a new tab by pressing the "+" button on the bottom of the page with the name "FEATURE\_SELECTION\_REGRESSION".

Import data into the input spreadsheet of the "FEATURE\_SELECTION\_REGRESSION" tab from the output of the "NORMALISE\_TRAIN\_SET" 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, and Help. Below the menu is a toolbar with icons for Import, Train/Test Split, Normalise Train Set, Normalise Test Set, and Feature Selection Regression. A main workspace contains a flowchart and a data table.

**Flowchart:**

```

graph TD
    IMPORT[IMPORT] --> TRAIN_TEST_SPLIT[TRAIN_TEST_SPLIT]
    TRAIN_TEST_SPLIT --> NORMALISE_TRAIN_SET[NORMALISE_TRAIN_SET]
    NORMALISE_TRAIN_SET --> FEATURE_SELECTION_REGRESSION[FEATURE_SELECTION_REGRESSION]
    NORMALISE_TRAIN_SET --> NORMALISE_TEST_SET[NORMALISE_TEST_SET]
    
```

**Data Table:**

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)
1	1.3195736895	-0.990231301	2.1644699572	-0.401748895	-0.886806883	-0.666607948	-0.864808001		
2	44417	2191987	03018	3424679	1208031	9494684	2702631	95139	
3	0.4561887577	-0.990231301	-0.693013965	-0.7092169641	1.560542520	1.4991820744	-0.704669906		1.111112045
4	950968	2191987	3971186	071937	0707939	892366	6137969	964801	
5	-1.529396585	-0.990231301	-0.693013965	-1.512714754	-1.671902964	-0.666607948	-0.709871188		-1.263601021
6	2283393	2191987	3971186	7921294	5032697	9494684	027559	409774	
7	1.4922506758	-0.990231301	-0.693013965	-0.401748895	-1.227815702	-0.666607948	-0.051890553		-1.328126798
8	942808	2191987	3971186	3424679	4317398	9494684	857816206	4204377	
9	0.7152042373	1.0092255077	-0.693013965	-1.512714754	0.0473997539	1.4991820744	1.404265656	1.385253785	
10	198929	64324	3971186	7921294	4647563	892366	510258	867963	
11	0.0243962919	1.0092255077	-0.693013965	-1.512714754	1.3205876730	1.4991820744	1.097817393		-0.39327603C
12	20436786	64324	3971186	7921294	4317398	892366	510258	867963	
13	0.0878012236	-0.990231301	1.2786419830	-0.401748895	0.699017793	1.4991820744	0.494212925		-0.841145651
14	697569	2191987	02972	3424679	152291	892366	9587895	1953941	
15	1.108374850	-0.990231301	459957203016	071937	2317884	892366	319518		0.802547973
16	953688	2191987	3971186	7921294	5032697	9494684	8826121	178595	
17	0.0244962919	-0.990231301	-0.693013965	1.8201828235	0.666607948	0.938048094	0.985043978		
18	20436786	2191987	3971186	568651	651711	9494684	8826121	178595	
19	0.3698502646	-0.990231301	1.2786419830	0.7092169641	0.0429428800	0.666607948	-1.09915461		-0.24551825E
20	2016480	2191987	02972	071937	9938967	9494684	4817781	97762694	
21	1.4922506758	1.0092255077	0.2928140008	0.7092169641	1.5369118309	0.666607948	0.4429293234		1.478960984
22	942808	64324	029674	071937	155732	9494684	839428	450777	
23	1.194518424612	1.0092255077	-0.693013965	-1.512714754	1.179817787	1.4991820744	-1.097817393		1.5231287976
24	530811	2191987	3971186	7921294	708758	892366	467864	472414	
25	1.356919598	1.0092255077	-0.693013965	-1.512714754	0.0972945140	1.4991820744	-0.247260451		
26	8784755	64324	3971186	7921294	1550046	892366	47284867	9277164	
27	-0.0973004119	1.0092255077	-0.693013965	-0.401748895	-0.722688102	-0.666607948	-1.710145561		-0.340933595

Choose the most important features for the classification using the Regression Analysis:

"Data Transformation" → "Variable Selection" → "Regression Analysis".

Then choose the "Diagnosis" column as the intercept column, the Significance level ( $\alpha$ ) as 0.05 and include all columns.

The screenshot shows the Isalos Analytics Platform interface. At the top, there is a menu bar with File, Edit, Data Transformation, Analytics, Statistics, Plot, and Help. Below the menu is a toolbar with icons for Import, Normalizers, Data Manipulation, Split, Variable Selection, Best First, Stepwise, and Regression Analysis. A main workspace contains a flowchart and a configuration dialog for the Regression Analysis Model.

**Regression Analysis Model Configuration:**

- Significance Level ( $\alpha$ ): 0.05
- Select Intercept Column: Col34 -- Diagnosis
- Excluded Columns: None
- Included Columns: Col2 -- Age, Col3 -- Gender, Col4 -- Ethnicity, Col5 -- EducationLevel, Col6 -- BMI, Col7 -- Smoking, Col8 -- AlcoholConsumption, Col9 -- PhysicalActivity

**Variable Selection Results:**

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)
1	1.3195736895	-0.990231301	2.1644699572	-0.401748895	-0.886806883	-0.666607948	-0.864808001		
2	44417	2191987	03018	3424679	1208031	9494684	2702631	95139	
3	0.4561887577	-0.990231301	-0.693013965	-0.7092169641	1.560542520	1.4991820744	-0.704669906		1.111112045
4	950968	2191987	3971186	071937	0707939	892366	6137969	964801	
5	1.529396585	-0.990231301	-0.693013965	-1.512714754	-1.671902964	-0.666607948	-0.709871188		
6	2283393	2191987	3971186	7921294	5032697	9494684	027559	409774	
7	1.4922506758	1.0092255077	-0.693013965	-1.512714754	0.0972945140	1.4991820744	-0.247260451		
8	953688	2191987	459957203016	071937	2317884	892366	319518		
9	0.0244962919	-0.990231301	-0.693013965	1.8201828235	1.5627913931	-0.666607948	-0.598048094		
10	20436786	64324	3971186	7921294	4317398	892366	510258	867963	
11	0.3698502646	-0.990231301	1.2786419830	0.7092169641	0.4029428800	-0.666607948	-1.099154636		
12	2016480	2191987	02972	071937	155732	9494684	839428	450777	
13	1.4922506758	1.0092255077	0.2928140008	0.7092169641	1.5369118309	-0.666607948	0.4429293234		

The results will appear on the output spreadsheet.

	Col1	Col2 (S)	Col3 (S)	Col4 (S)	Col5 (S)	Col6 (S)	Col7 (S)	Col8 (S)
User Header	User Row ID							
1		Regression Statistics						
2		Multiple R	0.6427533978429012					
3		R Square	0.41313193043859486					
4		Adjusted R Square	0.4009845965278801					
5		Standard Error	0.37590758946897873					
6		Observations	1579					
7								
8		Regression	df	SS	MS	F	Significance F	
9		Residual	32	153.78711830 863594	4.8058474471 44873	34.010090895 26929	4.0587245879 37507E-154	
10		Total	1546	218.45987345 83048	0.1413065158 2037826			
11			1578	372.24699176 69407				
12		Coefficients		Standard Error	t Stat	P-value	Lower 95.0%	Upper 95.0%
13		Diagnosis	0.6193793540215377	0.0094559758 7989092	65.473671591 2937	0.0	0.6008236148 982906	0.6379350931 447847
14		Age	0.034944578344872954	0.0095565389 61688421	3.6566144380 26531	2.6412976725 6839E-4	0.0161994307 7307164	0.0536897259 1667427
15		Gender	0.005905157055109047	0.0095891630 32430095	0.6158156906 017851	0.5381068188 904652	-0.012903982 619157102	0.0247142967 29375196
16		Ethnicity	-0.0051288114570836344	0.0095904324 85275994	-0.534784167 9671693	0.5928760144 825591	-0.023940441 1626286	0.0136828182 4846133
17		EducationLevel	-0.005631677102695146	0.0095434132 79615323	7773302	0.0552021963 015052	-0.024351078 654056724	0.0130877244 48666432
18		BMI	0.006842640052181735	0.0096094414 97752863	0.7120746875 645025	0.4765259676 099952	-0.012006275 824172727	0.0256915559 28536197
19		Smoking	-0.01068028611694388	0.0095538083 76531577	-1.117908764 339406	0.2637796445 924609	-0.029420077 64698911	0.0080595054 13101352
20		AlcoholConsumption	0.002461167657840956	0.0095636829 79952696	0.2573451737 1602894	0.7969465694 74437	-0.016297992 903093382	0.0212203282 18775293
21		PhysicalActivity	-0.00933461700271854	0.0095689073 55141496	-0.975488773 569992	0.3294704756 939527	-0.028103769 871170126	0.0094350464 70626419
22		DietQuality	-0.003050220046770803	0.0095468255 64645988	-0.319500971 93211996	0.7493898214 594091	-0.021776314 79393405	0.0156758747 0392445
23		SleepQuality	-0.021731024730421666	0.0095747622 37808862	-2.269615076 6658314	0.0233681600 34877315	-0.040511917 25146546	-0.002950132 2093778704
24		FamilyHistoryParkinsons	0.0101209008340667	0.0095524701 04518592	1.0595061511 1365	0.2895348303 630602	-0.008616265 675929415	0.0288580673 44062814
25		TraumaticBrainInjury	0.014589229732617963	0.0095958049 86224412	1.5203758052 150949	0.1286210332 7246324	-0.004232938 131512603	0.0334113975 9674853
26		Hypertension	0.006973590608428447	0.0095620828 29304448	0.7292961934 0431	0.4569309710 9648995	-0.011782431 257612396	0.0257296124 74469293
27		Diabetes	0.02562743849956757	0.0095871203 43858281	2.6731111721 138525	0.0075941977 85412139	-0.0068223055 58165209	0.0444325714 4096993
28		Depression	0.02723933473557805	0.0095605026 52271166	2.8491529918 7822	0.0044416491 5449658	-0.0084864123 86192469	0.045992570 8496363
29		Stroke	0.012415701060811338	0.0095488124 83364104	1.3002350902 211048	0.1937142092 2437116	-0.006314291 026668686	0.0311456931 4829136
30		SystolicBP	0.00528704609413583	0.0095272229 8344163	0.5549409416 918421	0.5790153341 23407	-0.013400598 197382648	0.0239746903 8565431
31		DiastolicBP	-0.003521415069752083	0.0095773887 41574176	-0.367680081 1546979	0.7131621231 346639	-0.022307459 476941696	0.0152646293 37437531
32		CholesterolTotal	0.007258067669105708	0.0095513189 59344359	0.7599021349 826152	0.4474289458 867108	-0.011476840 870064977	0.0259929762 80276394
33		CholesterolLDL	0.0015268790166764238	0.0095524401 46575041	0.1598417784 9762077	0.8730265930 105581	-0.017210228 855084888	0.0202639867 87645944
34		CholesterolHDL	5.411733944502038E-4	0.0095687599 01534996	0.0565562727 0607868	0.9549059762 495937	-0.018227945 54625449	0.0193102923 35154897
35		CholesterolTriglycerides	0.002268052516280527	0.0095630160 46575041	0.2371691634 9762077	0.8125569806 105581	-0.016489799 598043E-23	0.0210259048 4300933
36		UPDRS	0.18630288121195493	0.0096038801 88781513	19.398709433 045518	3.3396373538 02613E-75	0.1674648738 4105188	0.2051408885 8285798
37		MoCA	-0.07109273757277541	0.0095563178 9932587	7.439344140 884063	1.6703413366 346027E-13	-0.0089837451 53083589	-0.052348023 61471493
38		FunctionalAssessment	-0.11419317166623748	0.0095658700 57560244	-11.93756249 8665641	1.7403934101 474355E-31	-0.132956622 17907478	-0.095429721 15340018
39		Tremor	0.12423615764298186	0.0095515200 61017858	13.006951443 259872	8.8124775789 78414E-37	0.1055008546 4295439	0.1429714606 4300933
40		Rigidity	0.10336739579051336	0.00955057635 77577548	10.196505303 891393	1.1315824768 52925E-26	0.0786505923 3165018	0.1161182306 4937655
41		Bradykinesia	0.09738441151376426	0.00955057635 81646275	1.096505303 608197	1.1315824768 598043E-23	0.0786505923 4773905	0.1161182306 7978946
42		PosturalInstability	0.08795422774590572	0.0095343098 65763344	9.2250229942 43103	9.0117242673 54057E-20	0.0692526825 3738402	0.1066557729 5442742
43		SpeechProblems	-0.0053249866410449145	0.0095760103 43971408	-0.556075698 5185661	0.5782395755 011953	-0.024108327 321856172	0.0134583540 39766342
44		SleepDisorders	-0.01601764859514722	0.0095591144 87731708	-1.675641464 0398416	0.0940106527 4412763	-0.034767848 06031181	0.0027325508 70017371

The significant features according to the p-value are the following:

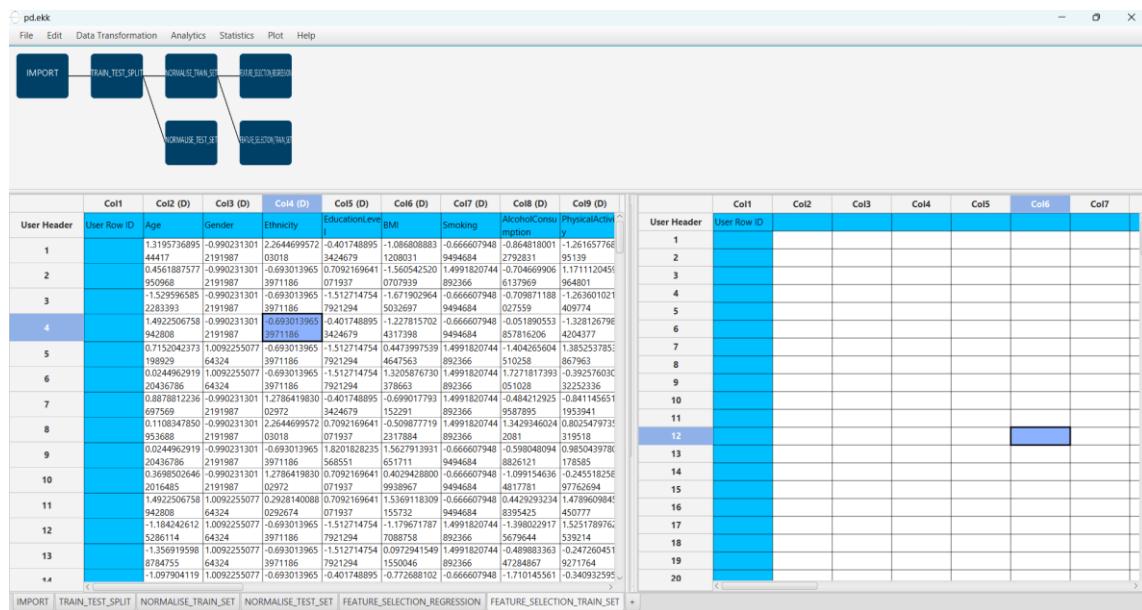
- Diagnosis (p-value = 0.0)

- Age (p-value = 2.64129767256839E-4)
- SleepQuality (p-value = 0.023368160034877315)
- Diabetes (p-value = 0.007594197785412139)
- Depression (p-value = 0.004441649154449658)
- UPDRS (p-value = 3.339637353802613E-75)
- MoCA (p-value = 1.6703413366346027E-13)
- FunctionalAssessment (p-value = 1.7403934101474355E-31)
- Tremor (p-value = 8.812477578978414E-37)
- Rigidity (p-value = 1.938920143652925E-26)
- Bradykinesia (p-value = 1.1315824768598043E-23)
- PosturalInstability (p-value = 9.011724267354057E-20)

## Step 7: Feature selection: train set

Create a new tab by pressing the "+" button on the bottom of the page with the name "FEATURE\_SELECTION\_TRAIN\_SET".

Import data into the input spreadsheet of the "FEATURE\_SELECTION\_TRAIN\_SET" tab from the output of the "NORMALISE\_TRAIN\_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".



Manipulate the data by choosing the columns that correspond to the significant features (from the previous step):

"Data Transformation" → "Data Manipulation" → "Select Column(s)".

User Header	User Row ID	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)
1	1.3195736895	0.990231301	2.2644699572	0.401748895	-1.086808883	0.666607948	-0.864818001	-1.261657776		
2	44417	2191987	03018	3424679	1208031	9494684	2792831	95139		
3	0.4561887577	0.990231301	-0.693013965	0.7092169641	-1.560542520	1.4991820744	-0.704669906	1.171112045		
4	950968	2191987	03971186	071937	0707939	892366	6137969	96480		
5	-1.529596585	0.990231301	-0.693013965	-1.512714754	-1.671902964	0.666607948	-0.709871188	-1.263601021		
6	283393	2191987	03971186	7921294	5032697	9494684	027559	409774		
7	1.4922506758	0.990231301	-0.693013965	-0.401748895	-1.227815703	0.666607948	-0.051890553	-1.328126796		
8	943636	2191987	03018	4317884	4317884	651208	0.401748895	0.202604		
9	0.0244962919	0.990231301	-0.693013965	1.2786419830	0.7092169641	0.403942880	0.666607948	0.98543978		
10	20436786	2191987	03971186	071937	0938067	9494684	4817781	0.245318256		
11	1.4922506758	0.990231301	-0.693013965	-1.512714754	-1.79671787	1.4991820744	0.442929324	1.47860984		
12	5286114	64324	0292674	071937	155732	9494684	8395425	450777		
13	-1.84242612	0.990231301	-0.693013965	-1.512714754	-0.097291549	1.4991820744	-0.48983363	-0.247260451		
14	8784755	64324	03971186	7921294	1550046	892366	47284867	9271764		
	-1.356919598	0.990231301	-0.693013965	-0.401748895	-0.72688102	0.666607948	-1.710145561	-0.340932595		
	-1.097904119	0.990231301	-0.693013965	-0.401748895	-0.72688102	0.666607948	-1.710145561	-0.340932595		

The results will appear on the output spreadsheet.

User Header	User Row ID	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)
1	1.3195736895	0.990231301	2.2644699572	0.401748895	-1.086808883	0.666607948	-0.864818001	-1.261657776		
2	44417	2191987	03018	3424679	1208031	9494684	2792831	95139		
3	0.4561887577	0.990231301	-0.693013965	0.7092169641	-1.560542520	1.4991820744	-0.704669906	1.171112045		
4	950968	2191987	03971186	7921294	5032697	9494684	027559	409774		
5	283393	2191987	03971186	071937	0938067	9494684	4817781	0.245318256		
6	1.4922506758	0.990231301	-0.693013965	-1.512714754	-1.79671787	1.4991820744	0.442929324	1.47860984		
7	5286114	64324	0292674	071937	155732	9494684	8395425	450777		
8	8784755	64324	03971186	7921294	1550046	892366	47284867	9271764		
9	-1.84242612	0.990231301	-0.693013965	-0.401748895	-0.72688102	0.666607948	-1.710145561	-0.340932595		
10	20436786	2191987	03018	4317884	4317884	651208	0.401748895	0.202604		
11	1.4922506758	0.990231301	-0.693013965	-0.401748895	-0.72688102	0.666607948	-1.710145561	-0.340932595		
12	5286114	64324	03971186	7921294	5032697	9494684	027559	409774		
13	-1.356919598	0.990231301	-0.693013965	-0.401748895	-0.72688102	0.666607948	-1.710145561	-0.340932595		
14	-1.097904119	0.990231301	-0.693013965	-0.401748895	-0.72688102	0.666607948	-1.710145561	-0.340932595		

## Step 8: Feature selection: test set

Create a new tab by pressing the "+" button on the bottom of the page with the name "FEATURE\_SELECTION\_TEST\_SET".

Import data into the input spreadsheet of the "FEATURE\_SELECTION\_TEST\_SET" tab from the output of the "NORMALISE\_TEST\_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".

The screenshot shows the pd.ekk Data Transformation interface. At the top, there is a menu bar with File, Edit, Data Transformation, Analytics, Statistics, Plot, and Help. The Data Transformation tab is selected. Below the menu is a flowchart with nodes: IMPORT, TRAIN\_TEST\_SPLIT, NORMALISE\_TRAIN\_SET, FEATURE\_SELECTION\_REGRESSION, and FEATURE\_SELECTION\_TEST\_SET. Arrows indicate the flow from IMPORT to TRAIN\_TEST\_SPLIT, then to NORMALISE\_TRAIN\_SET, then to FEATURE\_SELECTION\_REGRESSION, and finally to FEATURE\_SELECTION\_TEST\_SET. The FEATURE\_SELECTION\_TEST\_SET node has an arrow pointing down to a data grid. The data grid contains 11 rows of data with columns: Col1, Col2 (D), Col3 (D), Col4 (D), Col5 (D), Col6 (D), and Col7 (D). The first row is labeled 'User Header' and 'User Row ID'. The data includes various numerical values for age, gender, ethnicity, education level, BMI, smoking, alcohol consumption, and other medical metrics. The bottom of the interface shows tabs for IMPORT, TRAIN\_TEST\_SPLIT, NORMALISE\_TRAIN\_SET, NORMALISE\_TEST\_SET, FEATURE\_SELECTION\_REGRESSION, FEATURE\_SELECTION\_TRAIN\_SET, and FEATURE\_SELECTION\_TEST\_SET.

Manipulate the data by choosing the columns that correspond to the significant features (from the site):

"Data Transformation" → "Data Manipulation" → "Select Column(s)".

The screenshot shows the pd.ekk Data Transformation interface with a context menu open over the data grid. The menu path is Data Transformation > Data Manipulation > Select Column(s). A 'Select Column(s)' dialog box is displayed on the right. It has two main sections: 'Excluded Columns' (left) and 'Included Columns' (right). The 'Excluded Columns' section lists columns: Col17 -- Stroke, Col18 -- SystolicBP, Col19 -- DiastolicBP, Col20 -- CholesterolTotal, Col21 -- CholesterolLDL, Col22 -- CholesterolHDL, Col23 -- CholesterolTriglyc, and Col31 -- SpeechProblems. The 'Included Columns' section lists columns: Col24 -- UPDRS, Col25 -- MoCA, Col26 -- FunctionalAssessme, Col27 -- Tremor, Col28 -- Rigidity, Col29 -- Bradykinesia, Col30 -- PosturalInstability, and Col34 -- Diagnosis. 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 is a navigation bar with tabs: File, Edit, Data Transformation (which is selected), Analytics, Statistics, Plot, and Help. Below the navigation bar is a toolbar with icons for Import, Train/Test Split, Normalize Train Set, Feature Selection Regression, Feature Selection Train Set, Feature Selection Test Set, and Train Model. The main workspace contains a Data Transformation workflow diagram. The diagram starts with an 'IMPORT' node connected to a 'TRAIN\_TEST\_SPLIT' node. From 'TRAIN\_TEST\_SPLIT', two paths lead to 'NORMALISE\_TRAIN\_SET' and 'NORMALISE\_TEST\_SET' nodes. Both of these nodes then connect to a 'FEATURE\_SELECTION\_REGRESSION' node. Finally, the 'FEATURE\_SELECTION\_REGRESSION' node connects to a 'TRAIN\_MODEL' node. The bottom half of the screen displays the output spreadsheet. The first few rows of the spreadsheet are as follows:

User Header	User Row ID	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
1	0.0244692919	1.0092255077	-0.693013965	1.512714754	-1.684027954	-0.666607940	-1.36405017		
2	0.41918094	64324	3071106	7921294	7520348	9494684	9557914		
3	-1.443258092	-0.990231301	-0.693013965	0.401748895	1.694938958	1.4991802744	0.846586687		
4	0.534073	2191987	3971186	3424679	397585	892366	300567		
5	-1.615935078	1.0092255077	-0.693013965	-1.512714754	-1.173600369	-0.666607948	-1.48772622		
6	4032713	64324	3971186	7921294	273849	9494684	035291		
7	-0.579873160	1.0092255077	0.2928140086	1.820182835	0.7270201964	-0.666607948	-1.563923938		
8	1.184242612	-0.990231301	0.2928140086	-1.512714754	-1.578840216	-0.666607948	-0.97539408C		
9	0.061842201	1.0092255077	0.2928140086	0.401748895	-1.212779235	-0.666607948	-1.659910302		
10	0.2835117714	1.0092255077	-0.693013965	-0.401748895	0.470929735	0.3507163259	1.4991802744	0.81073021	
11	452328	64324	3971186	3424679	4990594	9494684	980105		

## Step 9: 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 "FEATURE\_SELECTION\_TRAIN\_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".

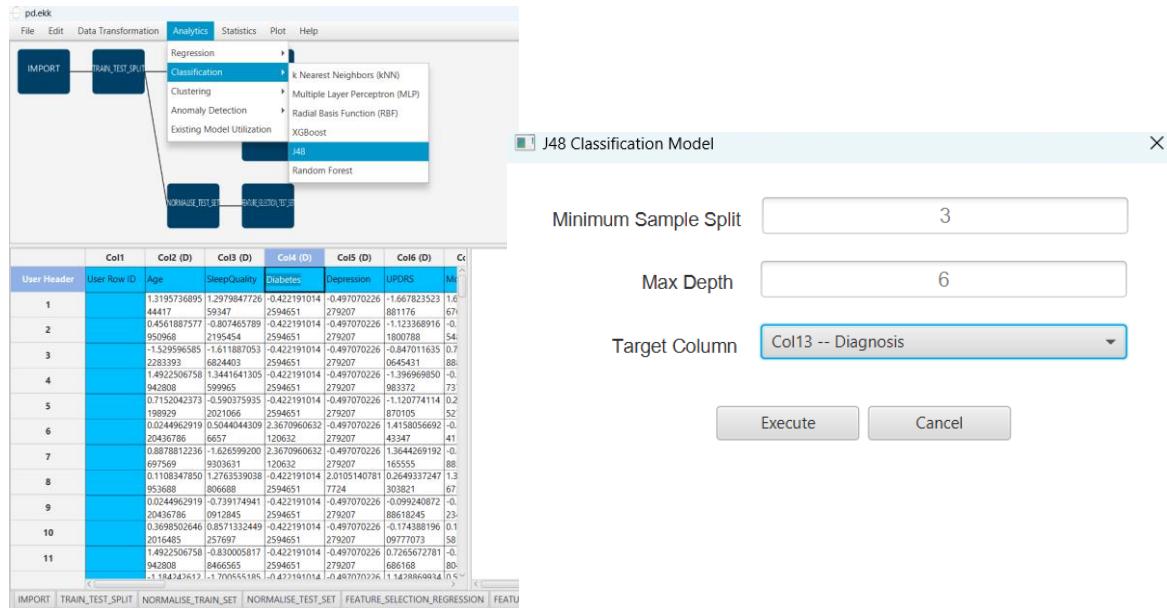
The screenshot shows the Isalos Analytics Platform interface. At the top, there is a navigation bar with tabs: File, Edit, Data Transformation (which is selected), Analytics, Statistics, Plot, and Help. Below the navigation bar is a toolbar with icons for Import, Train/Test Split, Normalize Train Set, Feature Selection Regression, Feature Selection Train Set, Feature Selection Test Set, and Train Model. The main workspace contains a Data Transformation workflow diagram. The diagram starts with an 'IMPORT' node connected to a 'TRAIN\_TEST\_SPLIT' node. From 'TRAIN\_TEST\_SPLIT', two paths lead to 'NORMALISE\_TRAIN\_SET' and 'NORMALISE\_TEST\_SET' nodes. Both of these nodes then connect to a 'FEATURE\_SELECTION\_REGRESSION' node. Finally, the 'FEATURE\_SELECTION\_REGRESSION' node connects to a 'TRAIN\_MODEL' node. The bottom half of the screen displays the output spreadsheet for the TRAIN\_MODEL tab. The first few rows of the spreadsheet are as follows:

User Header	User Row ID	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
1	1.3195736895	1.2979847726	-0.422191014	-0.497070226	-1.667623523	1.4			
2	44417	59347	2594651	279207	881176	67			
3	0.4545107577	-0.8074655789	-0.422191014	-0.497070226	-1.123368616	-0.1			
4	950968	219543	2594651	279207	1800788	54			
5	-1.52959658	-1.611887053	-0.422191014	-0.497070226	-0.497011635	0.7			
6	1.4922506758	1.3441641305	-0.422191014	-0.497070226	-1.396969850	-0.			
7	942808	599695	2594651	279207	983372	73			
8	0.7152024273	-0.590375935	-0.422191014	-0.497070226	-1.120774114	0.2			
9	198929	2021068	2594651	279207	870105	52			
10	0.0244692919	1.0092255077	0.044309	0.2928140086	0.470929735	1.4170556692	0.41		
11	0.0416768	6657	120632	279207	2000447	41			
12	0.887812236	-1.626599200	2.3670980632	0.497070226	1.3644269192	0.			
13	979549	9303631	120632	279207	165555	88			
14	0.110347850	1.276333908	-0.422191014	0.0105140781	0.2649337247	0.1			
15	953688	066688	2594651	7724	303821	67			
16	0.0244692919	-0.739174941	-0.422191014	-0.497070226	-0.099240872	-0.2			
17	2043678	091284	2594651	279207	88618245	23			
18	0.369852646	0.85711332449	-0.422191014	-0.497070226	-0.174888196	0.1			
19	2016485	257697	2594651	279207	09777073	58			
20	942808	846655	2594651	279207	886168	80			
21	-1.194742612	-1.709555185	-0.422191014	-0.497070226	-1.1439668074	0.5			

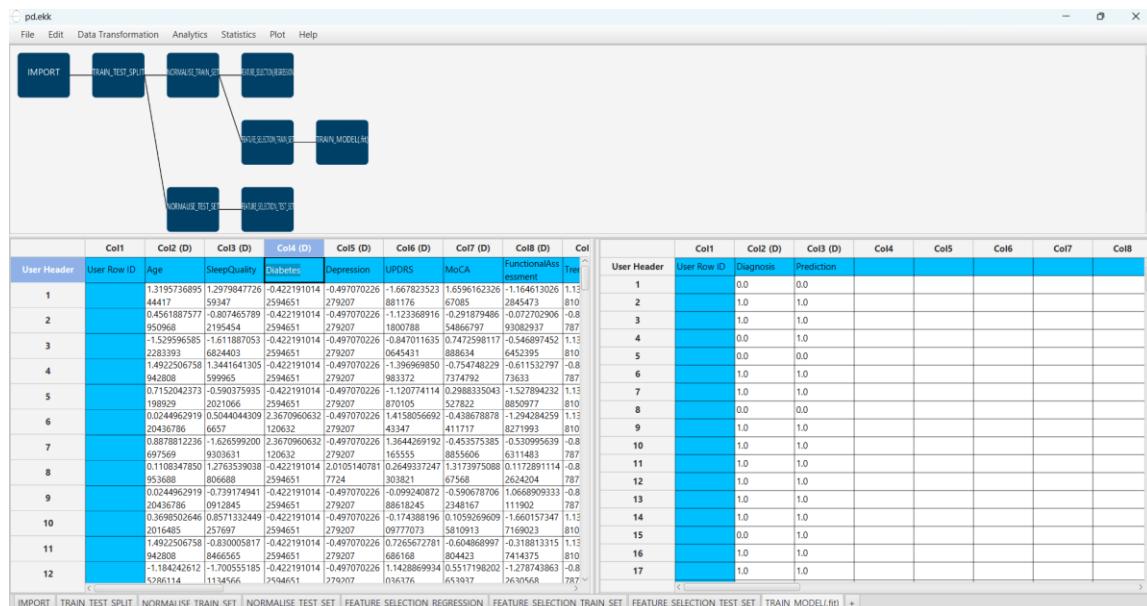
Use the J48 Method to train and fit the model:

"Analytics" → "Classification" → "J48"

and set the "Minimum Sample Split" as 3, the "Max Depth" as 6 and the "Target Column" as the column corresponding to "Diagnosis".



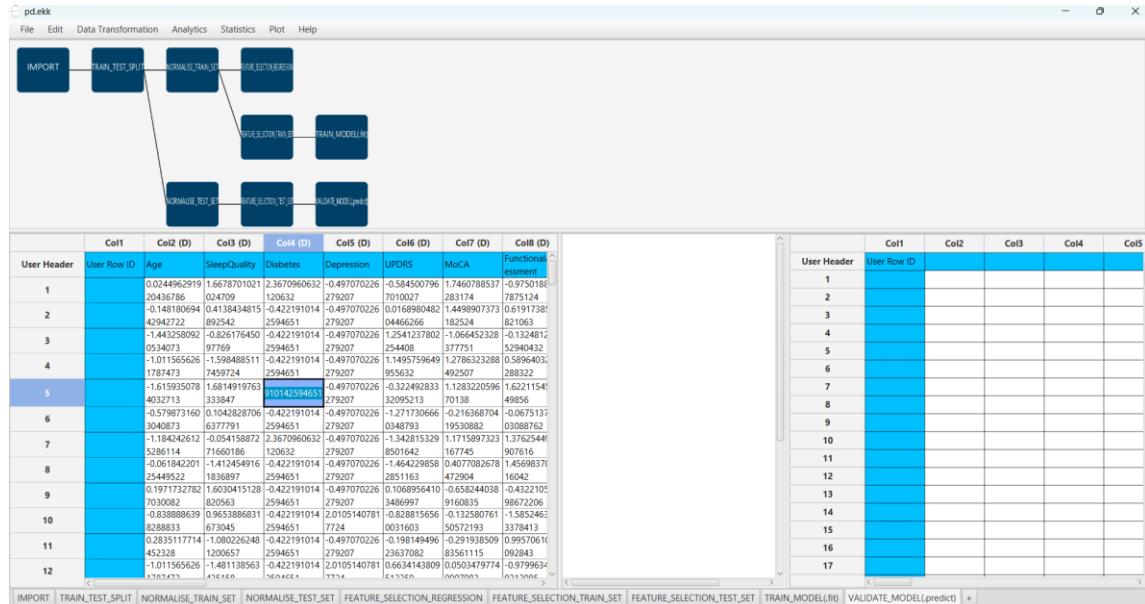
The predictions will appear on the output spreadsheet.



## Step 10: Validate the model

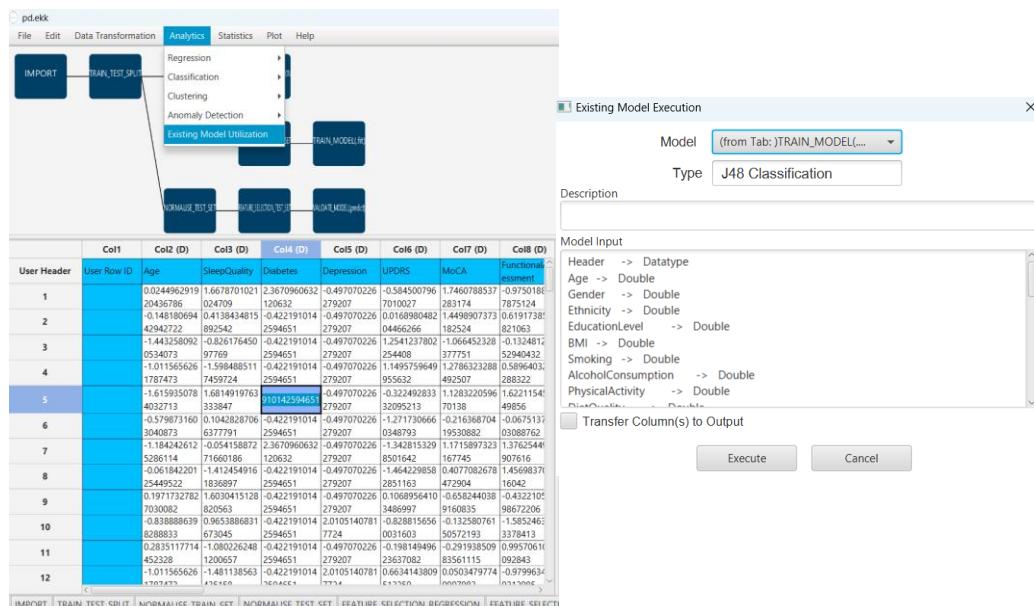
Create a new tab by pressing the "+" button on the bottom of the page with the name "VALIDATE\_MODEL(.predict)".

Import data into the input spreadsheet of the "VALIDATE\_MODEL(.predict)" tab from the output of the "FEATURE\_SELECTION\_TEST\_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".

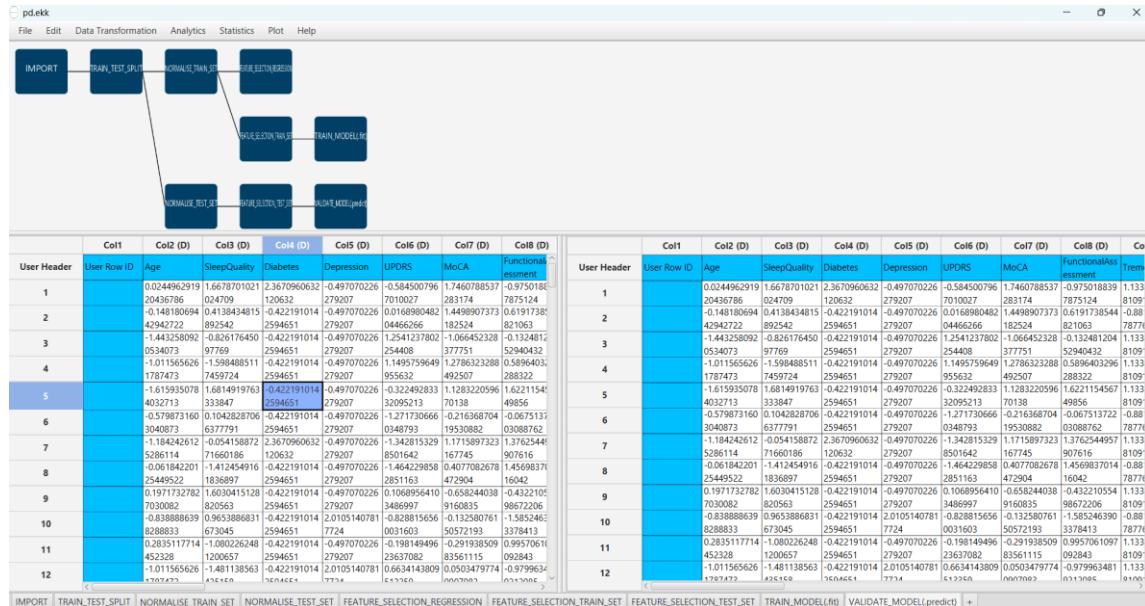


To validate the model:

"Analytics" → "Existing Model Utilization". Then choose Model "(from Tab: ) TRAIN\_MODEL (.fit)".



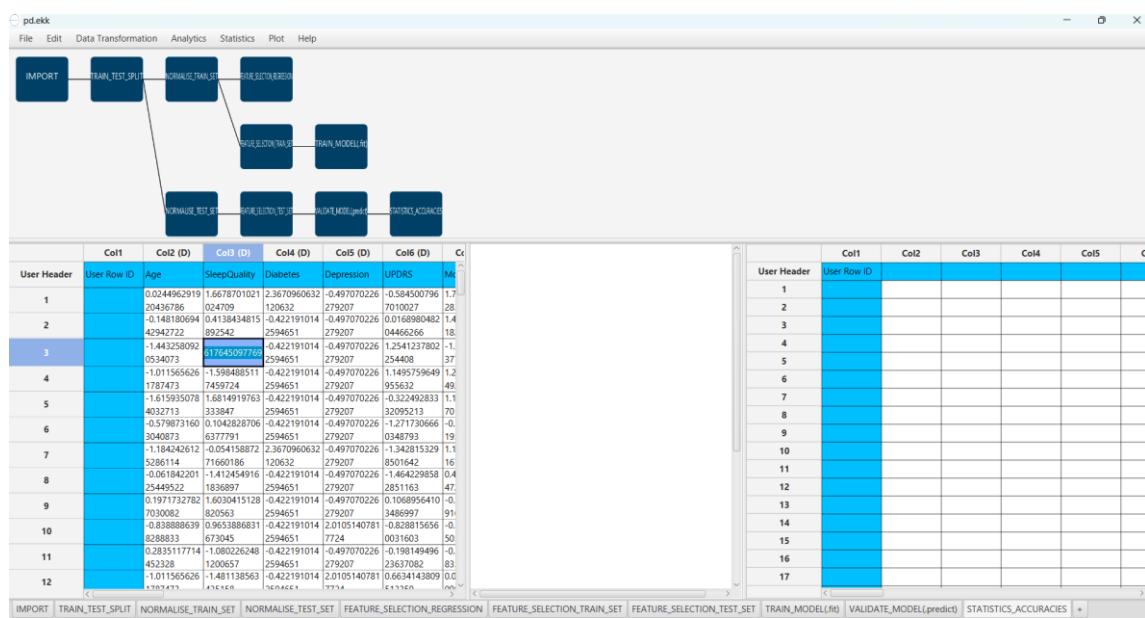
The predictions will appear on the output spreadsheet.



## Step 11: 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.

Accuracy: 0.93

F1-Score = 0.9199

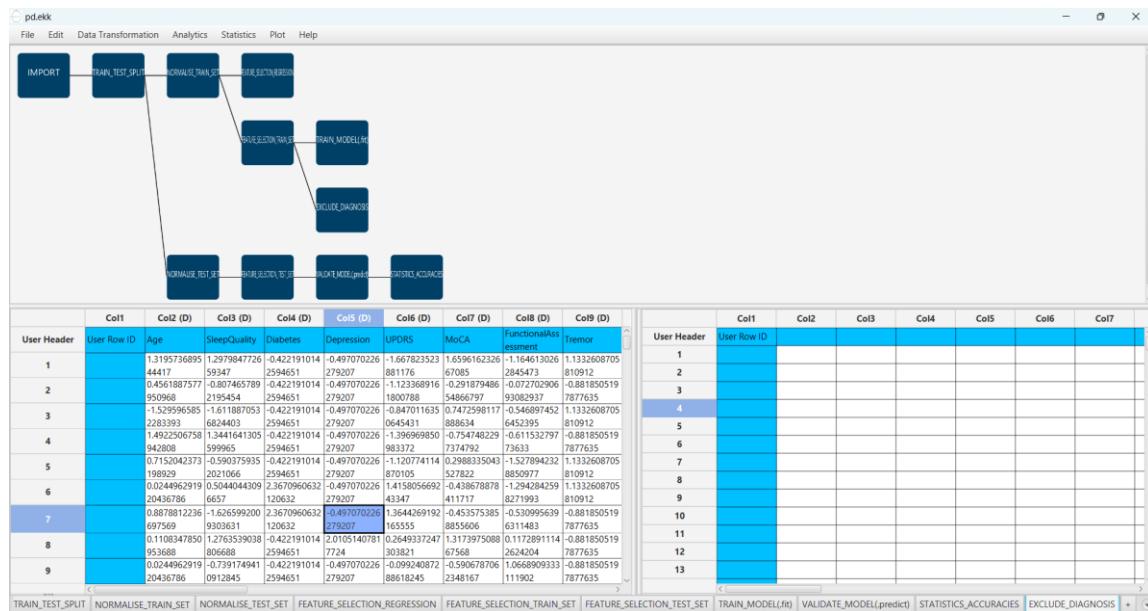
User Row ID	Col1	Col2	Col3	Col4	Col5	Col6	Col10
1	0.0244962919	1.6678701021	2.3670960632	-0.497070226	0.84500796	1.7	
2	0.0436786	0.04138434815	-0.42191014	-0.497070226	0.016890482	1.4	
3	-0.1443258092	0.1764509769	-0.42191014	-0.497070226	0.2512137802	-1.	
4	-1.011565626	0.0244962919	2.3670960632	-0.497070226	1.495759649	1.2	
5	-1.615935078	1.6814919763	-0.42191014	-0.497070226	0.3095219833	1.1	
6	0.403713	0.04138434815	2.3670960632	-0.497070226	1.211730666	-0.	
7	-0.5798317714	0.1764509769	-0.42191014	-0.497070226	1.342815329	1.1	
8	-0.061842201	-1.412454916	-0.42191014	-0.497070226	1.4464229658	0.4	
9	0.283517714	-1.080226248	-0.42191014	-0.497070226	0.198148496	-0.	
10	0.452328	1.6303415128	2.3670960632	-0.497070226	2.3637082	83	
11	-1.011565626	1.681138563	-0.42191014	-0.497070226	0.6634143809	0.0	
12	-1.707477	0.0244962919	2.3670960632	-0.497070226	0.0000000000	0.0	

# Step 12: Reliability check of each record of the test set

## Step 12.a: Create the domain

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

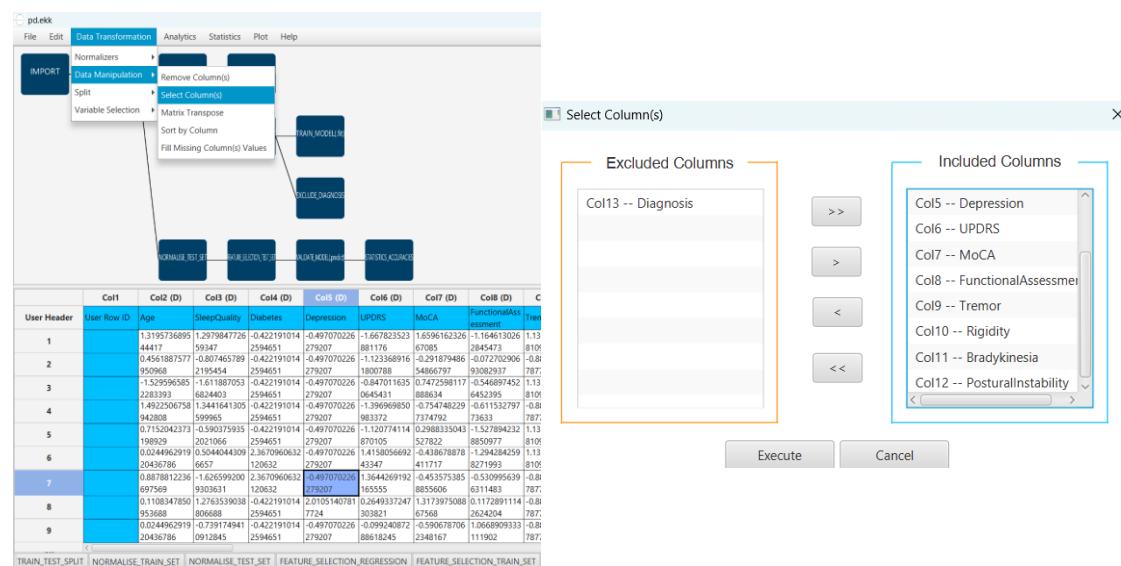
Import data into the input spreadsheet of the "EXCLUDE\_DIAGNOSIS" tab from the output of the "FEATURE\_SELECTION\_TRAIN\_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".



Manipulate the data to exclude the column that corresponds to the "Diagnosis"

"Data Transformation" → "Data Manipulation" → "Select Columns"

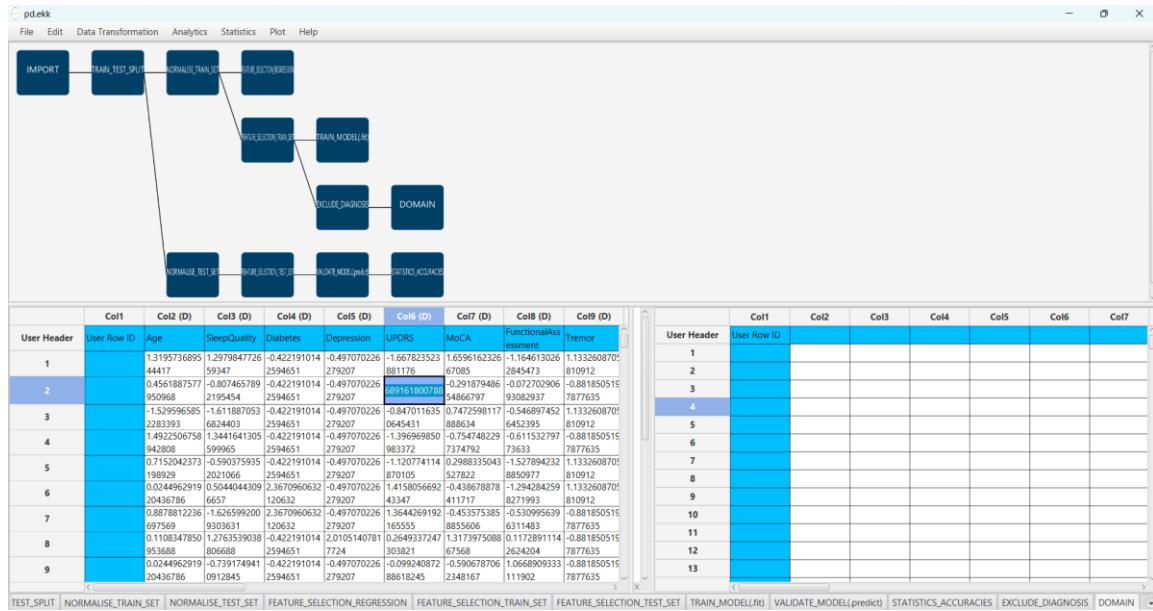
Then select all the columns except the "Diagnosis".



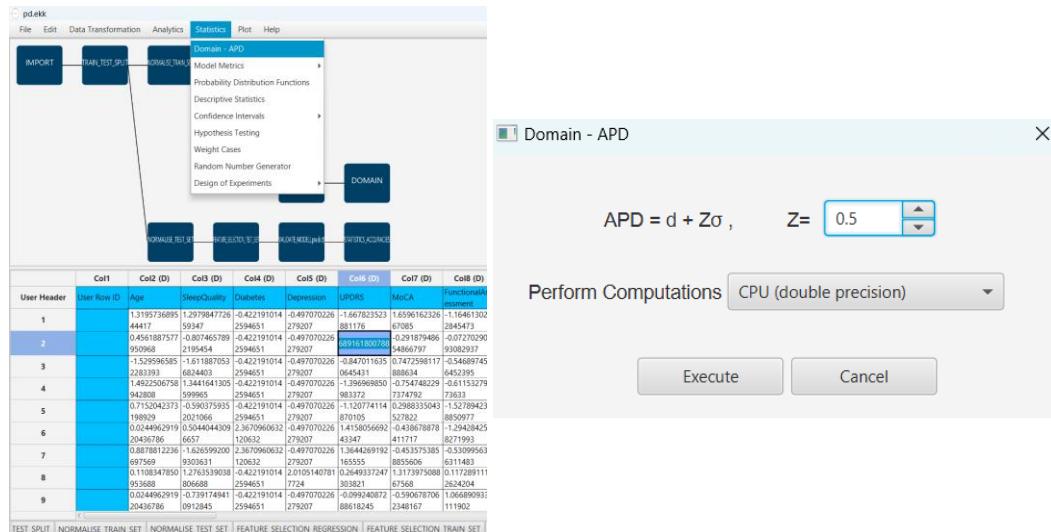
The results will appear on the output spreadsheet.

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

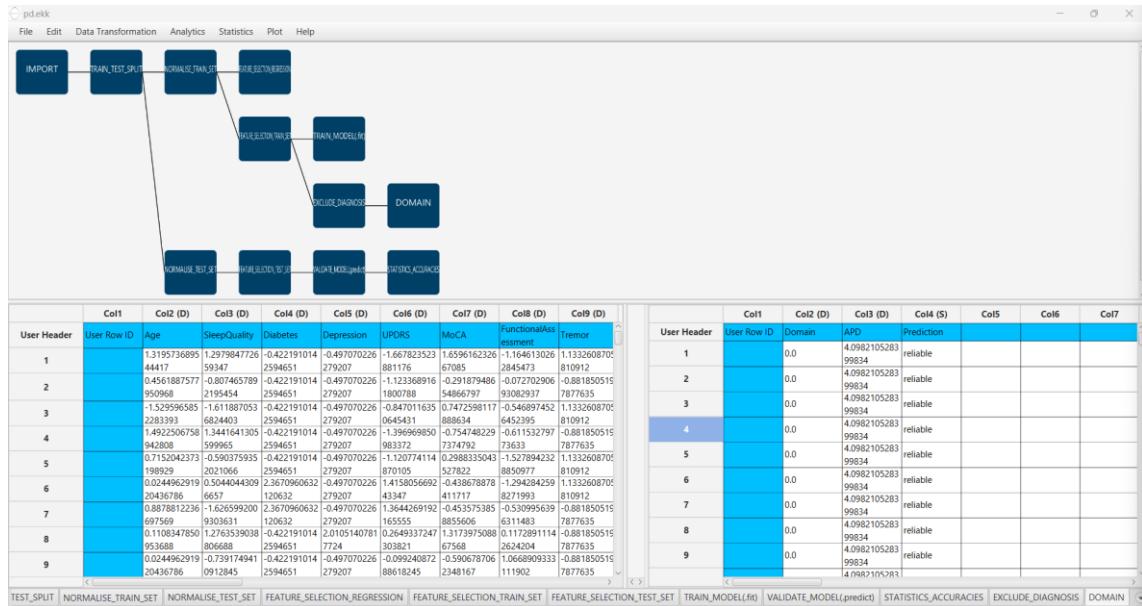
Import data into the input spreadsheet of the "DOMAIN" tab from the output of the "EXCLUDE\_DIAGNOSIS" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".



Create the domain: "Statistics" → "Domain APD"



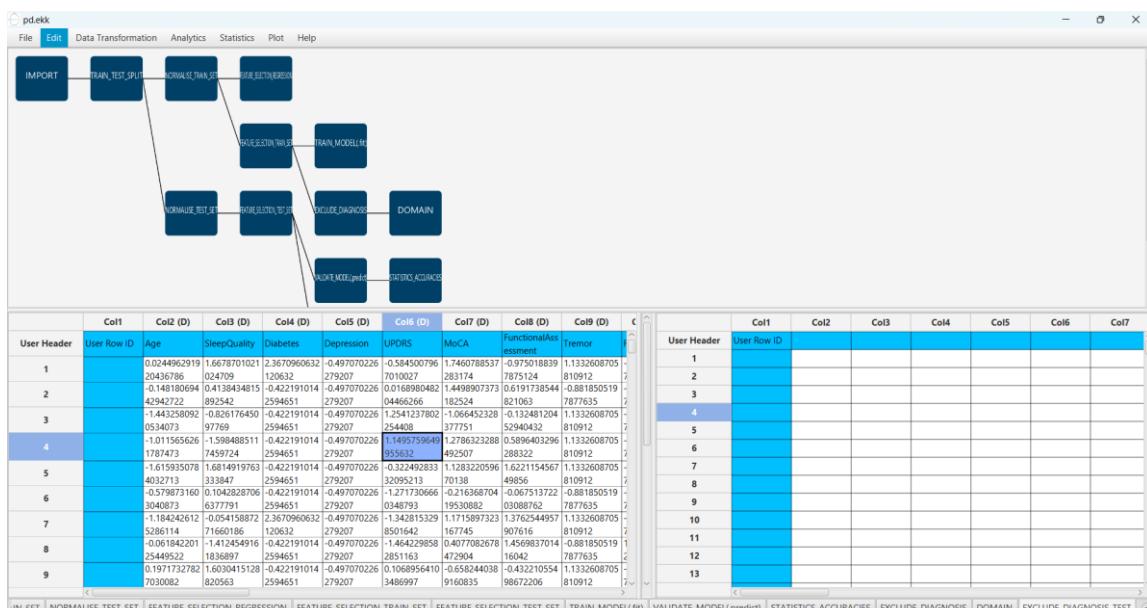
The results will appear on the output spreadsheet.



## Step 12.b: Check the test set reliability

Create a new tab by pressing the "+" button on the bottom of the page with the name "EXCLUDE\_DIAGNOSIS\_TEST\_SET".

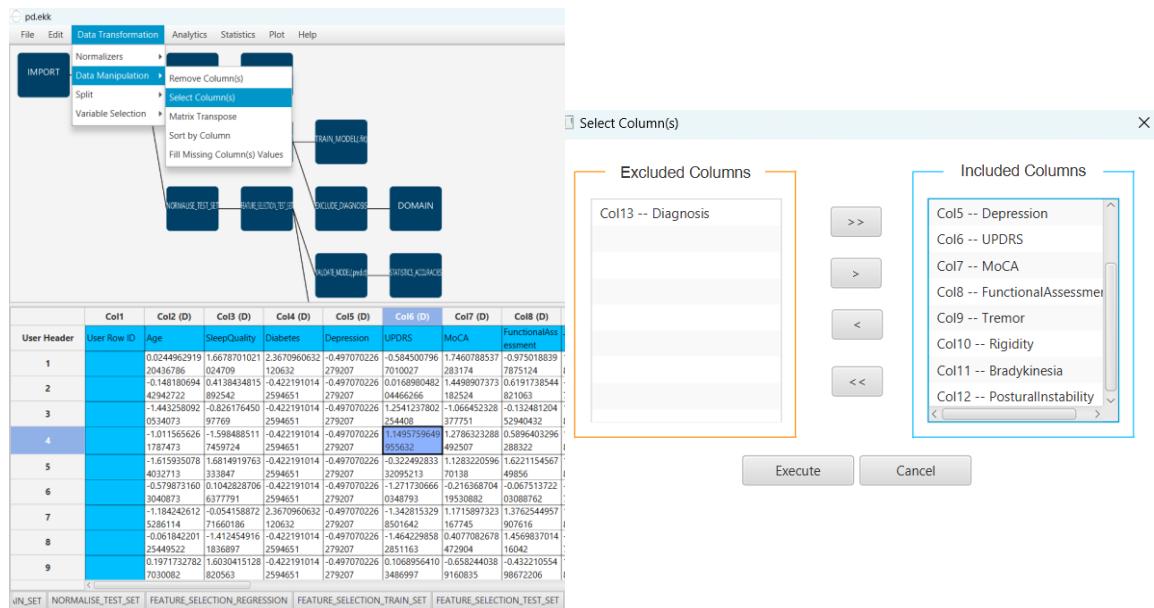
Import data into the input spreadsheet of the "EXCLUDE\_DIAGNOSIS\_TEST\_SET" tab from the output of the "FEATURE\_SELECTION\_TEST\_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".



Filter the data to exclude the column that corresponds to the "Diagnosis"

"Data Transformation" → "Data Manipulation" → "Select Columns".

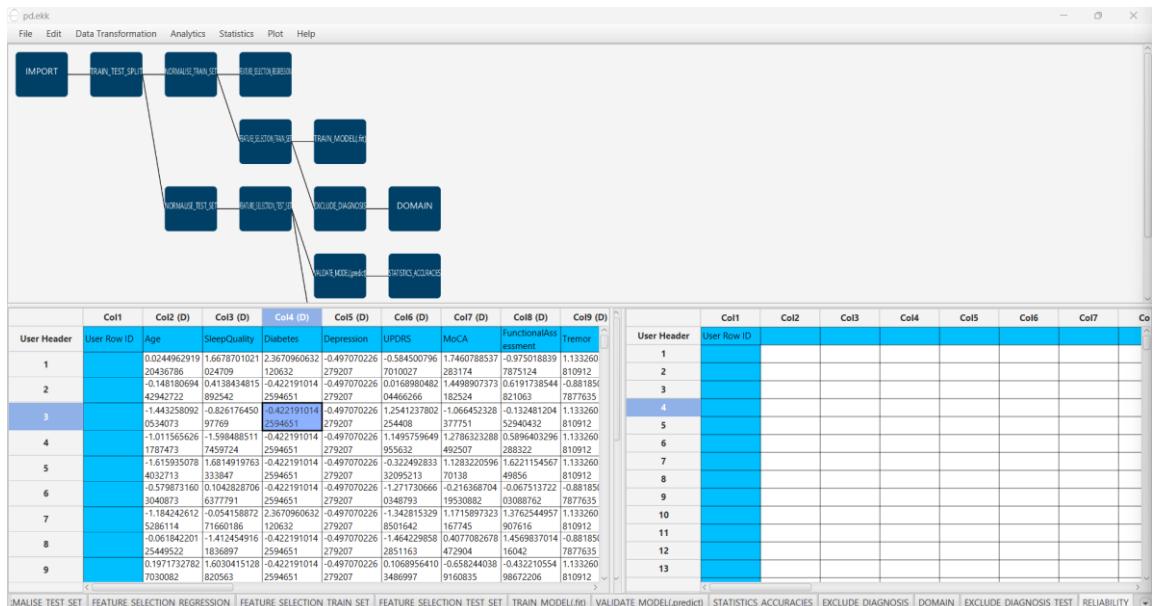
Then select all the columns except "Diagnosis".



The results will appear on the output spreadsheet.

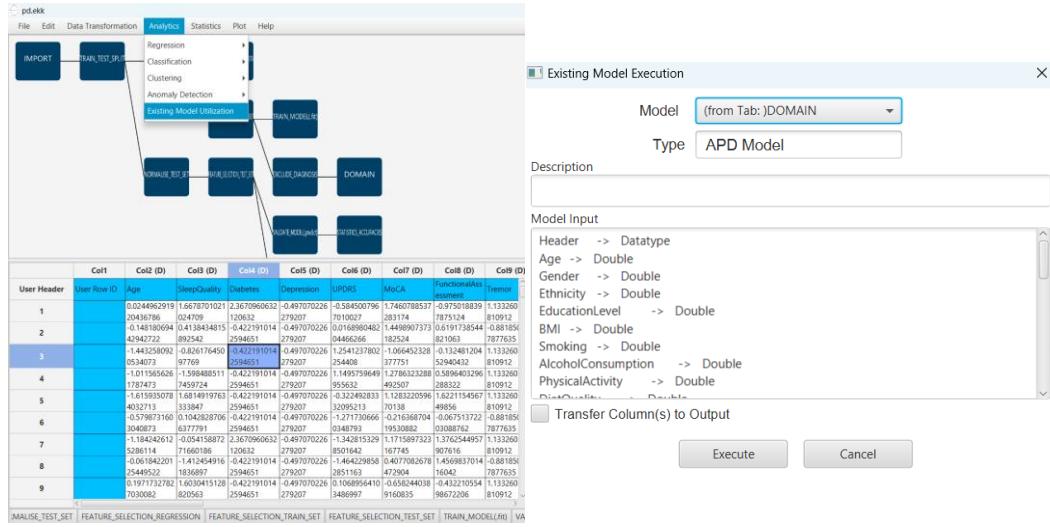
Create a new tab by pressing the "+" button on the bottom of the page with the name "RELIABILITY".

Import data into the input spreadsheet of the "RELIABILITY" tab from the output of the "EXCLUDE\_DIAGNOSIS\_TEST\_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".

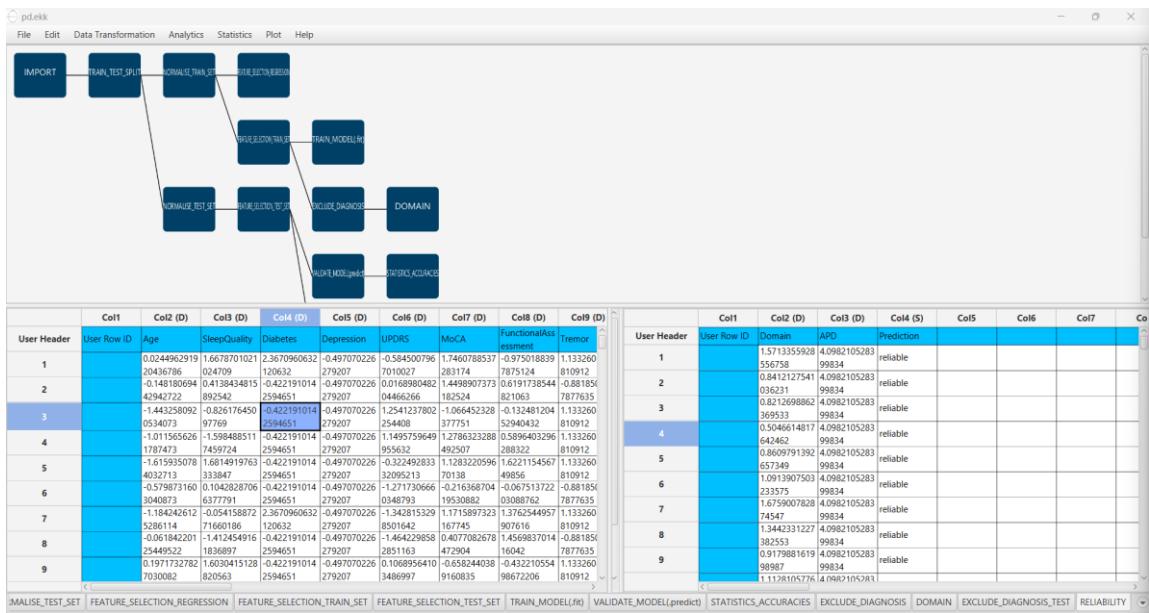


Check the Reliability: "Analytics" → "Existing Model Utilization".

Then select as Model "(from Tab: ) DOMAIN".



The results will appear on the output spreadsheet.



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:

