



# 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 the Isalos Analytics Platform interface. At the top, a context menu is open over a spreadsheet, with the option "Import from file" selected. Below this, the main window displays a large data table with 15 columns and 24 rows. The columns are labeled as Col1 through Col15, and the rows are labeled as User Header and User Row ID. The data table contains various numerical and categorical values, including PatientID, Age, Gender, Ethnicity, EducationLevel, BMI, Smoking, AlcoholConsumption, PhysicalActivity, DietQuality, SleepQuality, FamilyHistoryParkinsons, TraumaticBrainInjury, and Hypertension.

User Header	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (D)	Col8 (I)	Col9 (D)	Col10 (D)	Col11 (D)	Col12 (D)	Col13 (I)	Col14 (I)	Col15 (D)	Col16 (I)
1	3058	85	0	3	1	19.61987796	0	5.108240607	1.380659917	3.893969135	9.283194448	0	0	0	0	0
2	3059	75	0	0	2	16.24733916	1	6.027648029	8.40980405	8.51342825	5.602469506	0	0	0	0	0
3	3060	70	1	0	0	15.36823871	0	2.242135331	0.213274591	6.498804606	9.929823812	0	0	0	0	1
4	3061	52	0	0	0	15.45455733	0	5.997787563	1.375045164	6.715033333	4.196189318	0	0	0	0	0
5	3062	87	0	0	1	18.61694177	0	9.775242923	1.188607062	4.657572037	9.363934681	0	0	0	0	0
6	3063	68	1	2	1	39.42331141	1	13.5968889	7.796704004	7.070238878	7.737548608	0	0	0	0	0
7	3064	78	1	0	0	30.54200329	1	2.011281313	9.028536304	9.838445926	5.981983542	0	0	1	0	0
8	3065	70	1	0	0	36.75828161	1	19.98886597	3.891748622	3.421960006	7.895866239	0	0	0	0	1
9	3066	80	0	2	1	22.3805865	1	7.293287715	2.595670177	4.784827139	4.170469708	0	0	0	0	1
10	3067	71	0	3	2	23.72708628	1	17.78290985	7.344890316	3.393018462	9.245379607	0	1	0	0	0
11	3068	70	0	0	3	38.48254474	0	6.639761993	7.872186697	9.225710184	5.721854797	0	0	0	0	0
12	3069	53	0	0	1	35.8967387	1	5.212906263	7.185203018	7.918912222	5.569759738	0	0	0	0	0
13	3070	74	0	2	2	30.22551208	0	3.762918066	4.316651243	5.112520425	8.512502776	0	0	0	0	0
14	3071	87	1	1	2	38.29830655	0	12.61599522	9.299289996	6.71557911	5.563065282	0	0	1	0	0
15	3072	58	0	0	3	34.96532287	1	11.70859735	4.392403152	5.182038871	4.219612498	0	0	0	0	0
16	3073	56	1	0	0	18.95878125	1	2.04712047	9.432830343	1.720277262	4.04118069	0	1	0	0	0
17	3074	54	1	0	0	28.04958007	1	7.260733832	4.311617423	2.509936233	7.495208183	1	0	0	0	0
18	3075	57	1	0	1	21.85612305	0	0.252529809	4.04096502	1.142818169	5.870495582	1	0	0	0	0
19	3076	51	1	0	0	19.00200405	0	1.532135373	8.12215416	4.850752834	8.953637419	0	0	1	0	0
20	3077	55	0	1	2	22.54033861	0	11.57918488	8.893636005	4.1093017	4.218513746	0	0	0	0	0
21	3078	62	1	0	2	29.72724176	1	2.03243796	8.934190297	7.074353018	5.722196689	0	0	0	0	0
22	3079	79	1	3	1	33.24761783	1	9.545279952	6.95599901	1.577840865	6.375642584	0	0	0	0	0
23	3080	74	1	0	1	20.6172019	0	5.198675067	6.704484305	7.048416351	9.655452542	0	0	0	0	0
24	3081	60	1	0	0	38.13361066	1	10.78134701	7.730134334	5.085816332	7.236686358	0	0	0	0	1

## 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 screenshot shows the 'Data Transformation' menu with options like 'IMPORT', 'Normalizers', 'Data Manipulation', 'Split', 'Variable Selection', 'Matrix Transpose', and 'Sort by Column'. The 'Split' option is selected, leading to the 'Select Column(s)' dialog box. This dialog has two sections: 'Excluded Columns' and 'Included Columns'. The 'Excluded Columns' section contains 'Col2 -- PatientID' and 'Col36 -- DoctorInCharge'. The 'Included Columns' section contains a list of medical conditions: 'Col28 -- Tremor', 'Col29 -- Rigidity', 'Col30 -- Bradykinesia', 'Col31 -- PosturalInstability', 'Col32 -- SpeechProblems', 'Col33 -- SleepDisorders', 'Col34 -- Constipation', and 'Col35 -- Diagnosis'. The 'Execute' button is highlighted.

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".

The screenshot shows the 'TRAIN\_TEST\_SPLIT' tab with a spreadsheet. The spreadsheet has columns labeled 'User Header', 'Col1', 'Col2', 'Col3', 'Col4', 'Col5', and 'Col6'. The 'User Header' column contains 'User Row ID' and numbers 1 through 20. A right-click context menu is open over the spreadsheet, showing options: 'Import from SpreadSheet', 'Import from file', 'Export Spread Sheet Data', and 'Clear SpreadSheet'. The 'Import from SpreadSheet' option is highlighted.

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:

Random Partitioning

Training set percentage

75

☐ Usage of random generator seed

24066204583300

☒ Stratified sampling

Col34 -- Diagnosis

Execute

Cancel

The results will appear on the output spreadsheet.

User Header	User Row ID	Age	Gender	Ethnicity	EducationLevel	BMI	Smoking	AlcoholConsumption
1	85	0	3	1	19.61987796	0	5.1982406	1.3806
2	75	0	0	2	16.24733916	1	6.027648	8.4096
3	70	1	0	0	15.36823871	0	2.2421352	1.3755
4	52	0	0	0	15.45455733	0	5.9977875	1.1886
5	87	0	0	1	18.61604177	0	9.7752425	9.0285
6	68	1	2	1	39.42331141	1	13.596886	3.8917
7	78	1	0	0	30.54200329	1	2.0112813	2.5956
8	70	1	0	0	36.75828161	1	19.988865	7.3446
9	80	0	2	1	22.3805865	1	7.2932877	7.8721
10	71	0	3	2	23.72708628	1	17.782905	9.4326
11	70	0	0	3	38.48254474	0	6.6397619	4.3116
12	53	0	0	1	35.8967387	1	5.2129062	9.2992
13	74	0	2	2	30.22551208	0	3.762918	4.3116
14	87	1	1	2	38.29830655	0	12.615995	0.4005
15	58	0	0	3	34.96532287	1	11.708597	8.8936
16	56	1	0	0	18.95878125	1	2.0471204	8.9341
17	54	1	0	0	28.04958007	1	7.2607136	6.9555
18	57	1	0	1	21.85612305	0	0.2552296	6.7044
19	51	1	0	0	19.00200405	0	1.5321353	7.7201
20	55	0	1	2	22.54833861	0	11.579184	7.2422
21	62	1	0	2	29.72724176	1	2.0324375	7.6781
22	79	1	3	1	33.24761783	1	9.5452795	0.9570
23	74	1	0	1	20.6172019	0	5.1986755	9.8692
24	60	1	0	0	38.13363958	1	10.781347	3.0874

## 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".

Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (D)	Col7 (I)	Col8 (D)	Col9 (D)
User Header	User Row ID	Age	Gender	Ethnicity	EducationLevel	BMI	Smoking	AlcoholConsumption
1	85	0	3	1	19.61987796	0	5.108240607	1.380659917
2	75	0	0	2	16.24733916	1	6.027648029	8.40980405
3	52	0	0	0	15.45455733	0	5.997787563	1.375045164
4	87	0	0	1	18.61604177	0	9.775242923	1.188607062
5	78	1	0	0	30.54200329	1	2.011281313	9.028536304
6	70	1	0	0	36.75828161	1	19.9886597	3.891748622
7	80	0	2	1	22.3805865	1	7.293287715	2.595670177
8	71	0	3	2	23.72708628	1	17.78290985	7.344890316
9	70	0	0	3	38.48254474	0	6.639761993	7.872186697
10	74	0	2	2	30.22551208	0	3.762918066	4.316651243
11	87	1	1	2	38.29830655	0	12.61599522	9.299289996
12	56	1	0	0	18.95878125	1	2.04712047	9.432830343
13	54	1	0	0	28.04958007	1	7.260733832	4.311617423
14	57	1	0	1	21.85612305	0	0.255229809	4.04096502
15	55	0	1	2	22.54833861	0	11.57918488	8.893662605
16	62	1	0	2	29.72724176	1	2.03243796	8.934190297
17	79	1	3	1	33.24761783	1	9.545279952	6.95599901
18	74	1	0	1	20.6172019	0	5.198675067	6.704484305
19	60	1	0	0	38.13363959	1	10.78134701	7.720134324
20	71	1	2	1	15.86360295	0	19.59171834	7.242315106
21	79	1	1	2	36.90543424	0	9.89059796	7.678179303
22	66	0	0	2	23.68997318	0	11.42519836	0.957084433
23	78	0	1	3	25.89636742	0	17.90479619	9.86938435
24	61	0	0	2	36.76373473	1	7.608817132	7.987486621

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

FileEditData TransformationAnalyticsStatisticsPlotHelp

NormalizersZ-Score

IMPORTData ManipulationSplitVariable Selection

	Col1 (I)	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (D)	Col7 (I)	Col8 (D)	Col9 (D)
User Header	User Row ID	Age	Gender	Ethnicity	EducationLevel	BMI	Smoking	AlcoholConsumption	PhysicalActivity
1	85	0	3	1	19.61987796	0	5.108240607	1.380659917	
2	75	0	0	2	16.24733916	1	6.027648029	8.40980405	
3	52	0	0	0	15.45455733	0	5.997787563	1.375045164	
4	87	0	0	1	18.61604177	0	9.775242923	1.188607062	
5	78	1	0	0	30.54200329	1	2.011281313	9.028536304	
6	70	1	0	0	36.75828161	1	19.9886597	3.891748622	
7	80	0	2	1	22.3805865	1	7.293287715	2.595670177	
8	71	0	3	2	23.72708628	1	17.78290985	7.344890316	
9	70	0	0	3	38.48254474	0	6.639761993	7.872186697	
10	74	0	2	2	30.22551208	0	3.762918066	4.316651243	
11	87	1	1	2	38.29830655	0	12.61599522	9.299289996	
12	56	1	0	0	18.95878125	1	2.04712047	9.432830343	
13	54	1	0	0	28.04958007	1	7.260733832	4.311617423	
14	57	1	0	1	21.85612305	0	0.255229809	4.04096502	
15	55	0	1	2	22.54833861	0	11.57918488	8.893662605	
16	62	1	0	2	29.72724176	1	2.03243796	8.934190297	
17	79	1	3	1	33.24761783	1	9.545279952	6.95599901	
18	74	1	0	1	20.6172019	0	5.198675067	6.704484305	
19	60	1	0	0	38.13363959	1	10.78134701	7.720134324	
20	71	1	2	1	15.86360295	0	19.59171834	7.242315106	
21	79	1	1	2	36.90543424	0	9.89059796	7.678179303	
22	66	0	0	2	23.68997318	0	11.42519836	0.957084433	
23	78	0	1	3	25.89636742	0	17.90479619	9.86938435	
24	61	0	0	2	36.76373473	1	7.608817132	7.987486621	

IMPORTTRAIN TEST COL1 COL2 COL3 COL4 COL5 COL6 COL7 COL8 COL9

**ZScore Normalizer**

Excluded Columns: Col34 -- Diagnosis

Included Columns: Col26 -- FunctionalAssessm, Col27 -- Tremor, Col28 -- Rigidity, Col29 -- Bradykinesia, Col30 -- PosturalInstability, Col31 -- SpeechProblems, Col32 -- SleepDisorders, Col33 -- Constipation

Execute Cancel

The results will appear on the output spreadsheet.

User Header	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (D)	Col7 (I)	Col8 (D)
User Header	User Row ID	Age	Gender	Ethnicity	Education	BMI	Smoking	AlcoholConsumption
1	85	0	3	1	19.61987796	0	5.1082406	
2	75	0	0	2	16.24733916	1	6.0276480	
3	52	0	0	0	15.45455733	0	5.9977875	
4	87	0	0	1	18.61604177	0	9.7752425	
5	78	1	0	0	30.54200329	1	2.0112812	
6	70	1	0	0	36.75828161	1	19.988865	
7	80	0	2	1	22.3805865	1	7.2932877	
8	71	0	3	2	23.72708628	1	17.782905	
9	70	0	0	3	38.48254474	0	6.6397615	
10	74	0	2	2	30.22551208	0	3.7629180	
11	87	1	1	2	38.29830655	0	12.615995	
12	56	1	0	0	18.95878125	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	1	2.0324375	
17	79	1	3	1	33.24761783	1	9.5452795	
18	74	1	0	1	20.6172019	0	5.1986755	
19	60	1	0	0	38.13363959	1	10.781347	
20	71	1	2	1	15.86360295	0	19.591716	
21	79	1	1	2	36.90543424	0	9.8905975	
22	66	0	0	2	23.68997318	0	11.425196	
23	78	0	1	3	25.89636742	0	17.904796	
24	61	0	0	2	36.76372473	1	7.6688171	

## 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".

User Header	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (D)	Col7 (I)	Col8 (D)	Col9 (D)
User Header	User Row ID	Age	Gender	Ethnicity	Education	BMI	Smoking	AlcoholConsumption	PhysicalActivity
1	70	1	0	0	15.36823871	0	2.242135331	0.2132745	
2	68	1	2	1	39.42331141	1	13.5968889	7.7967040	
3	53	0	0	1	35.8967387	1	5.212906263	7.1852030	
4	58	0	0	3	34.96532287	1	11.70859735	4.3924631	
5	51	1	0	0	19.00200405	0	1.532135373	8.1221541	
6	63	1	1	3	32.53263836	0	1.094654252	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	29.85370804	1	5.418741531	1.0356313	
10	60	0	0	0	31.93604901	0	8.241282371	7.9181040	
11	73	1	0	1	30.70951465	0	16.91863358	4.2184886	
12	58	1	0	1	17.52171944	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	15.94157373	1	17.09450292	5.1732513	
16	72	0	3	1	16.22239896	0	6.469157837	3.3422069	
17	50	0	1	3	32.01090917	0	0.606812251	2.5999824	
18	55	1	3	0	15.7567057	0	12.02304347	1.8164269	
19	81	1	0	1	34.31655447	0	15.36062935	0.8390699	

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

The screenshot shows the 'Analytics' menu with 'Existing Model Utilization' selected. The main window displays a data table with columns: User Header, User Row ID, Age, Gender, Ethnicity, EducationLevel, BMI, Smoking, and AlcoholConsumption. The table contains 19 rows of data.

User Header	User Row ID	Age	Gender	Ethnicity	EducationLevel	BMI	Smoking	AlcoholConsumption
1	70	1	0	0	0	15.36823871	0	2.242135331
2	68	1	2	1	1	39.42331141	1	13.5968889
3	53	0	0	1	1	35.8967387	1	5.212906263
4	58	0	0	3	3	34.96532287	1	11.70859735
5	51	1	0	0	0	19.00200405	0	1.532135373
6	63	1	1	3	3	32.53263836	0	1.094654252
7	56	0	1	0	1	16.11707674	0	4.473425263
8	69	1	1	1	1	18.7230873	0	0.543629553
9	72	1	0	1	1	29.85370804	1	5.418741531
10	60	0	0	0	0	31.93604901	0	8.241282371
11	73	1	0	1	1	30.70951465	0	16.91863358
12	58	1	0	1	1	17.52217944	0	11.6197039
13	58	1	1	1	1	17.61243635	0	8.23972525
14	75	0	3	1	3	30.00437964	0	3.737186934
15	57	1	1	1	1	15.94157373	1	17.09450292
16	72	0	3	1	1	16.02239896	0	6.469157837
17	50	0	1	3	3	32.01090917	0	0.606812251
18	55	1	3	0	0	15.7567057	0	12.02304347
19	81	1	0	1	1	34.31655447	0	15.36062935

The 'Existing Model Execution' dialog box shows the 'Model' dropdown set to '(from Tab: )NORMALISE\_TR...' and the 'Type' dropdown set to 'Z Score Normalizer Model'. The 'Description' field is empty. The 'Model Input' section lists various features mapped to 'Double' datatypes: Age, Gender, Ethnicity, EducationLevel, BMI, Smoking, AlcoholConsumption, and PhysicalActivity. The 'Transfer Column(s) to Output' checkbox is unchecked. 'Execute' and 'Cancel' buttons are at the bottom.

The results will appear on the output spreadsheet.

The screenshot shows the 'Analytics' menu with 'Existing Model Utilization' selected. The main window displays a data table with columns: User Header, User Row ID, Age, Gender, Ethnicity, EducationLevel, BMI, Smoking, and AlcoholConsumption. The table contains 19 rows of data. The 'Output' tab is selected, showing the results of the model execution.

User Header	User Row ID	Age	Gender	Ethnicity	EducationLevel	BMI	Smoking	AlcoholConsumption	PhysicalActivity
1	70	1	0	0	0	15.36823871	0	2.242135331	0.2132745
2	68	1	2	1	1	39.42331141	1	13.5968889	7.7967040
3	53	0	0	1	1	35.8967387	1	5.212906263	7.1852030
4	58	0	0	3	3	34.96532287	1	11.70859735	4.3924631
5	51	1	0	0	0	19.00200405	0	1.532135373	8.1221541
6	63	1	1	3	3	32.53263836	0	1.094654252	6.0915424
7	56	0	1	0	1	16.11707674	0	4.473425263	9.6204920
8	69	1	1	1	1	18.7230873	0	0.543629553	9.3659498
9	72	1	0	1	1	29.85370804	1	5.418741531	1.0356313
10	60	0	0	0	0	31.93604901	0	8.241282371	7.9181040
11	73	1	0	1	1	30.70951465	0	16.91863358	4.2184886
12	58	1	0	1	1	17.52217944	0	11.6197039	6.9720368
13	58	1	1	1	1	17.61243635	0	8.23972525	1.9625576
14	75	0	3	1	3	30.00437964	0	3.737186934	6.2622734
15	57	1	1	1	1	15.94157373	1	17.09450292	5.1732513
16	72	0	3	1	1	16.02239896	0	6.469157837	3.3422069
17	50	0	1	3	3	32.01090917	0	0.606812251	2.5999824
18	55	1	3	0	0	15.7567057	0	12.02304347	1.8164269
19	81	1	0	1	1	34.31655447	0	15.36062935	0.8390699

## 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".

Choose the most important features for the classification using the Regression Analysis by browsing: "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 results will appear on the output spreadsheet.

	Col1	Col2 (\$)	Col3 (\$)	Col4 (\$)	Col5 (\$)	Col6 (\$)	Col7 (\$)	Col8 (\$)
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	0.0094599758 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	-0.590111413 7773302	0.5552021963 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	-0.009334361700271854	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 00392445
23		SleepQuality	-0.021731024730421666	0.0095747622 37808862	-2.269615076 6658314	0.0233681600 34877315	-0.040511917 25146546	-0.022950132 2093778704
24		FamilyHistoryParkinsons	0.0101209008340667	0.0095524701 04518592	1.0595061511 1365	0.2895348303 630602	-0.008616265 675929415	0.0288580673 44062814
25		TraumaticBrainInjury	0.014589229732671963	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	0.4659309710 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 54449658	0.0084864123 86192469	0.0459922570 8496363
29		Stroke	0.012415701060811338	0.0095488124 83364104	1.3002350902 211048	0.1937142092 2437116	-0.006314291 026668886	0.0311456931 4829136
30		SystolicBP	0.00528704609413583	0.0095272229 8344163	0.5549409416 914821	0.5790153341 233407	-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 08276394
33		CholesterolLDL	0.0015268790166764238	0.0095524401 31189343	0.1598417781 9561138	0.8730265930 744	-0.017210228 700645668	0.0202639867 33998512
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 855084888	0.0210259048 87645944
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	-0.089837451 53083589	-0.052348023 61471493
38		FunctionalAssessment	-0.11419317166623748	0.0095658700 57560244	-11.93756249 866541	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	0.0095359870 77577548	10.839716428 891393	1.9389201436 52925E-26	0.0846625607 3165018	0.1220722308 4937655
41		Bradykinesia	0.09738441151376426	0.0095507635 81646275	10.196505303 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".

User Header	User Row ID	Age	Gender	Ethnicity	Education	BMI	Smoking	AlcoholConsumption	PhysicalActivity
1	1.3195736895	-0.990231301	2.2644699572	-0.401748895	-1.086808883	-0.666607948	-0.864818001	-1.261657766	
2	44417	2191987	03018	3424679	1200831	9494684	2792831	95139	
3	0.4561087577	-0.990231301	-0.693013965	0.7092169641	-1.560542520	1.4991820744	-0.704669996	-1.1711120455	
4	950968	2191987	3971186	071937	0707939	892366	6137969	964801	
5	-1.529596585	-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	-1.328126796	
8	942800	2191987	3971186	3424679	4317398	9494684	857816206	4204377	
9	0.7132042373	1.0092255077	-0.693013965	-1.512714754	0.4473997539	1.4991820744	1.404265604	1.3852537855	
10	198929	64324	3971186	7921294	4647563	892366	510258	867963	
11	0.0244962919	1.0092255077	-0.693013965	-1.512714754	1.3205876730	1.4991820744	1.7271817393	-0.392576030C	
12	20436786	64324	3971186	7921294	378663	892366	051028	32252336	
13	0.8878812236	-0.990231301	1.2786419830	-0.401748895	-0.699017793	1.4991820744	-0.484212935	-0.041145651	
14	697569	2191987	02972	3424679	152291	892366	9587895	1953941	
15	0.1108347850	-0.990231301	2.2644699572	0.7092169641	-0.509877719	1.4991820744	1.3429346024	0.8025479735	
16	953688	2191987	03018	071937	2317884	892366	2081	319518	
17	0.0244962919	-0.990231301	-0.693013965	1.8201828235	1.5627913931	-0.666607948	-0.598048094	0.985043978C	
18	20436786	2191987	3971186	568551	651711	9494684	8826121	178585	
19	0.3698502646	-0.990231301	1.2786419830	0.7092169641	0.4029428800	-0.666607948	-1.099154636	-0.245518258	
20	2016485	2191987	02972	071937	9938967	9494684	4817781	97762694	
21	1.4922506758	1.0092255077	0.2928140088	0.7092169641	1.5369118309	-0.666607948	0.4429293234	1.478960984C	
22	942800	64324	0292074	071937	155732	9494684	8395423	450777	
23	-1.184242612	1.0092255077	-0.693013965	-1.512714754	-1.179671787	1.4991820744	-1.398022917	1.525178976C	
24	5286114	64324	3971186	7921294	7088758	892366	5679644	539214	
25	-1.356919598	1.0092255077	-0.693013965	-1.512714754	0.0972941549	1.4991820744	-0.489883363	-0.247260451	
26	8784755	64324	3971186	7921294	1550046	892366	47284867	9271764	
27	-1.097904119	1.0092255077	-0.693013965	-0.401748895	-0.772688102	-0.666607948	-1.710145561	-0.340932595	

Manipulate the data by choosing the columns that correspond to the significant features (from the previous step): "Data Transformation" → "Data Manipulation" → "Select Column(s)".

pd.elek

File Edit Data Transformation Analytics Statistics Plot Help

Normalizers

IMPORT

Data Manipulation

Remove Column(s)

Select Column(s)

Split

Variable Selection

Matrix Transpose

Sort by Column

Fill Missing Column(s) Values

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)
User Header	User Row ID	Age	Gender	Ethnicity	EducationLevel	BMI	Smoking	AlcoholConsumption	PhysicalActivity
1	1.1915736895	-0.990231301	2.2644699572	-0.401748895	-1.086808883	-0.666607948	-0.864818001	-1.261657766	
2	44417	2191987	03018	3424679	1208031	9494684	2792831	95139	
3	0.4561887577	-0.990231301	-0.693013965	0.7092169641	-1.560542520	1.4991820744	-0.704669906	1.171120495	
4	950968	2191987	3971186	071937	0707939	892366	6137969	964801	
5	-1.529596585	-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	-1.328126796	
8	942808	2191987	3971186	3424679	4317398	9494684	857816206	4204377	
9	0.7152042373	1.0092255077	-0.693013965	-1.512714754	0.4473997539	1.4991820744	-1.404265604	1.385253785	
10	198929	64324	3971186	7921294	4647563	892366	510258	867963	
11	0.0244962919	1.0092255077	-0.693013965	-1.512714754	1.3205876730	1.4991820744	1.7271817393	-0.39257603C	
12	20436786	64324	3971186	7921294	378663	892366	051028	32252336	
13	0.8878812236	-0.990231301	1.2786419830	-0.401748895	-0.699017793	1.4991820744	-0.484212925	-0.841145651	
14	697569	2191987	02972	3424679	152291	892366	9587895	1953941	
15	0.1108347850	-0.990231301	2.2644699572	0.7092169641	-0.509877719	1.4991820744	1.3429346024	0.8025479731	
16	953688	2191987	03018	071937	2317884	892366	2081	319518	
17	0.0244962919	-0.990231301	-0.693013965	1.8201828235	1.5627913931	-0.666607948	-0.598048094	0.9850439786	
18	20436786	2191987	3971186	568551	651711	9494684	8826121	173585	
19	0.3698502646	-0.990231301	1.2786419830	0.7092169641	0.4029428800	-0.666607948	-1.099154636	-0.245518258	
20	2016485	2191987	02972	071937	9938967	9494684	4817781	97762694	
21	1.4922506758	1.0092255077	0.2928140088	0.7092169641	1.5369118309	-0.666607948	0.4429293234	1.4789609847	
22	942808	64324	0292674	071937	155732	9494684	8395425	450777	
23	-1.184242612	1.0092255077	-0.693013965	-1.512714754	-1.179671787	1.4991820744	-1.398022917	1.525178976	
24	5286114	64324	3971186	7921294	7088758	892366	5679644	539214	
25	-1.356919598	1.0092255077	-0.693013965	-1.512714754	0.0972941549	1.4991820744	-0.489883363	-0.247260451	
26	8784755	64324	3971186	7921294	1550046	892366	47284867	9271764	
27	-1.097904119	1.0092255077	-0.693013965	-0.401748895	-0.772688102	-0.666607948	-1.710145561	-0.340932595	

IMPORT TRAIN\_TEST\_SPLIT NORMALISE\_TRAIN\_SET NORMALISE\_TEST\_SET FEATURE\_SELECTION\_REGRESSION FEATURE\_SELECTION\_TRAIN\_SET

Select Column(s)

Excluded Columns

Col3 -- Gender

Col4 -- Ethnicity

Col5 -- EducationLevel

Col6 -- BMI

Col7 -- Smoking

Col8 -- AlcoholConsumption

Col9 -- PhysicalActivity

Col10 -- DietQuality

Included Columns

Col2 -- Age

Col11 -- SleepQuality

Col15 -- Diabetes

Col16 -- Depression

Col24 -- UPDRS

Col25 -- MoCA

Col26 -- FunctionalAssessment

Col27 -- Tremor

Execute Cancel

The results will appear on the output spreadsheet.

pd.elek

File Edit Data Transformation Analytics Statistics Plot Help

IMPORT

TRAIN\_TEST\_SPLIT

NORMALISE\_TRAIN\_SET

FEATURE\_SELECTION\_REGRESSION

NORMALISE\_TEST\_SET

FEATURE\_SELECTION\_TRAIN\_SET

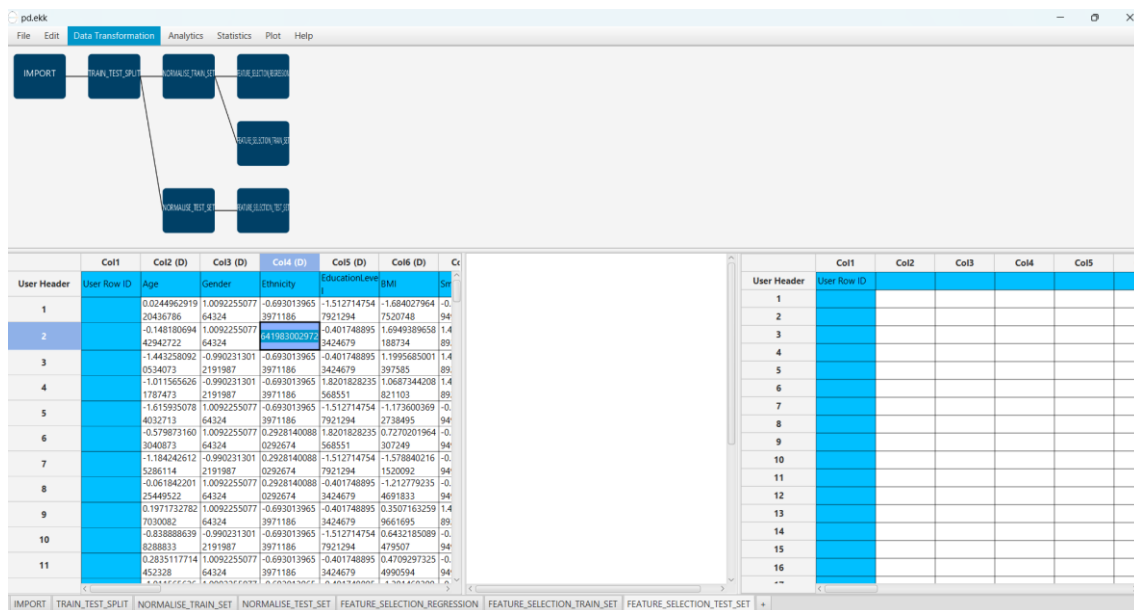
	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)
User Header	User Row ID	Age	Gender	Ethnicity	EducationLevel	BMI	Smoking	AlcoholConsumption	PhysicalActivity
1	1.1915736895	-0.990231301	2.2644699572	-0.401748895	-1.086808883	-0.666607948	-0.864818001	-1.261657766	
2	44417	2191987	03018	3424679	1208031	9494684	2792831	95139	
3	0.4561887577	-0.990231301	-0.693013965	0.7092169641	-1.560542520	1.4991820744	-0.704669906	1.171120495	
4	950968	2191987	3971186	071937	0707939	892366	6137969	964801	
5	-1.529596585	-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	-1.328126796	
8	942808	2191987	3971186	3424679	4317398	9494684	857816206	4204377	
9	0.7152042373	1.0092255077	-0.693013965	-1.512714754	0.4473997539	1.4991820744	-1.404265604	1.385253785	
10	198929	64324	3971186	7921294	4647563	892366	510258	867963	
11	0.0244962919	1.0092255077	-0.693013965	-1.512714754	1.3205876730	1.4991820744	1.7271817393	-0.39257603C	
12	20436786	64324	3971186	7921294	378663	892366	051028	32252336	
13	0.8878812236	-0.990231301	1.2786419830	-0.401748895	-0.699017793	1.4991820744	-0.484212925	-0.841145651	
14	697569	2191987	02972	3424679	152291	892366	9587895	1953941	
15	0.1108347850	-0.990231301	2.2644699572	0.7092169641	-0.509877719	1.4991820744	1.3429346024	0.8025479731	
16	953688	2191987	03018	071937	2317884	892366	2081	319518	
17	0.0244962919	-0.990231301	-0.693013965	1.8201828235	1.5627913931	-0.666607948	-0.598048094	0.9850439786	
18	20436786	2191987	3971186	568551	651711	9494684	8826121	173585	
19	0.3698502646	-0.990231301	1.2786419830	0.7092169641	0.4029428800	-0.666607948	-1.099154636	-0.245518258	
20	2016485	2191987	02972	071937	9938967	9494684	4817781	97762694	
21	1.4922506758	1.0092255077	0.2928140088	0.7092169641	1.5369118309	-0.666607948	0.4429293234	1.4789609847	
22	942808	64324	0292674	071937	155732	9494684	8395425	450777	
23	-1.184242612	1.0092255077	-0.693013965	-1.512714754	-1.179671787	1.4991820744	-1.398022917	1.525178976	
24	5286114	64324	3971186	7921294	7088758	892366	5679644	539214	
25	-1.356919598	1.0092255077	-0.693013965	-1.512714754	0.0972941549	1.4991820744	-0.489883363	-0.247260451	
26	8784755	64324	3971186	7921294	1550046	892366	47284867	9271764	
27	-1.097904119	1.0092255077	-0.693013965	-0.401748895	-0.772688102	-0.666607948	-1.710145561	-0.340932595	

IMPORT TRAIN\_TEST\_SPLIT NORMALISE\_TRAIN\_SET NORMALISE\_TEST\_SET FEATURE\_SELECTION\_REGRESSION FEATURE\_SELECTION\_TRAIN\_SET

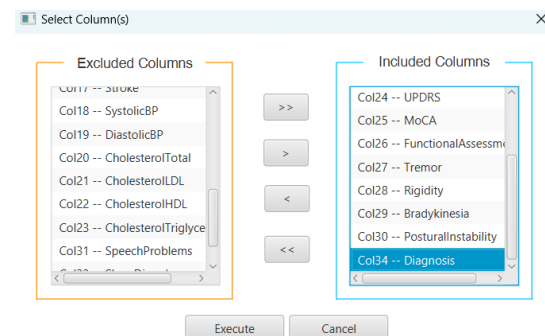
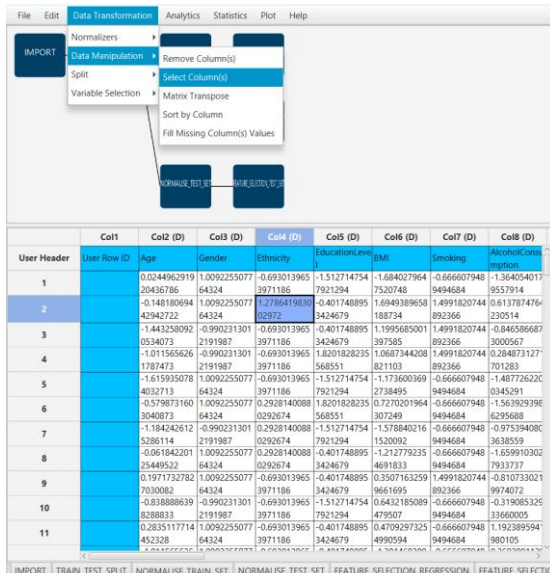
## Step 8: Feature selection: test set

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".



Manipulate the data by choosing the columns that correspond to the significant features (from step 7): "Data Transformation" → "Data Manipulation" → "Select Column(s)".



The results will appear on the output spreadsheet.

User Header	User Row ID	Age	Gender	Ethnicity	EducationLevel	BMI	Smoking	AlcoholConsumption
1	0.0244962919	1.0092255077	-0.693013965	-1.512714754	-1.684027964	-0.666607948	-1.364054017	0.0244962919
2	0.0436786	64324	3971186	7921294	7530748	9494604	9557914	0.0436786
3	0.148180694	1.0092255077	-0.693013965	-1.512714754	-1.684027964	-0.666607948	-1.364054017	0.148180694
4	0.2942722	64324	3971186	7921294	7530748	9494604	9557914	0.2942722
5	-1.443258092	-0.990231301	-0.693013965	-1.512714754	-1.684027964	-0.666607948	-1.364054017	-1.443258092
6	0.534073	2191987	3971186	7921294	7530748	9494604	9557914	0.534073
7	-1.011565626	-0.990231301	-0.693013965	-1.512714754	-1.684027964	-0.666607948	-1.364054017	-1.011565626
8	1.787473	2191987	3971186	7921294	7530748	9494604	9557914	1.787473
9	-1.615935078	1.0092255077	-0.693013965	-1.512714754	-1.684027964	-0.666607948	-1.364054017	-1.615935078
10	4032713	64324	3971186	7921294	7530748	9494604	9557914	4032713
11	-0.579873160	1.0092255077	-0.693013965	-1.512714754	-1.684027964	-0.666607948	-1.364054017	-0.579873160

## 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".

User Header	User Row ID	Age	Gender	Ethnicity	EducationLevel	BMI	Smoking	AlcoholConsumption
1	1.3195736895	1.2979847726	-0.422191014	-0.497070226	-1.667823523	1.2	67	1.3195736895
2	0.4561887577	59347	2594651	279207	881176	67	1.23368916	0.4561887577
3	950968	2195454	2594651	279207	1800788	54	1.39696850	950968
4	-1.529596585	-1.611887053	-0.422191014	-0.497070226	-0.847011635	0.7	1.39696850	-1.529596585
5	2.283393	6824403	2594651	279207	9645431	88	1.4158056692	2.283393
6	1.492506758	1.3441641305	-0.422191014	-0.497070226	-1.39696850	0.7	1.39696850	1.492506758
7	0.7152042373	-0.590375935	-0.422191014	-0.497070226	-1.120774114	0.2	1.39696850	0.7152042373
8	1.98929	2021066	2594651	279207	870105	52	1.4158056692	1.98929
9	0.0244962919	0.5044044309	2.3670960632	-0.497070226	-1.120774114	0.2	1.39696850	0.0244962919
10	20436786	6657	120632	279207	43347	41	1.39696850	20436786
11	0.8878812236	-1.626599200	2.3670960632	-0.497070226	-1.120774114	0.2	1.39696850	0.8878812236

Use the J48 Method to train and fit the model by browsing: "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 screenshot shows the Isalos Analytics Platform interface. On the left, a workflow diagram includes nodes for 'IMPORT', 'TRAIN\_TEST\_SPLIT', 'NORMALISE\_TRAIN\_SET', 'NORMALISE\_TEST\_SET', 'FEATURE\_SELECTION\_REGRESSION', and 'FEATURE\_SELECTION\_TRAIN\_SET'. The 'Analytics' menu is open, showing options like Regression, Classification, Clustering, Anomaly Detection, and Existing Model Utilization. The 'J48 Classification Model' dialog box is open, showing the following settings:

- Minimum Sample Split: 3
- Max Depth: 6
- Target Column: Col13 -- Diagnosis

Below the dialog box, there are 'Execute' and 'Cancel' buttons. The output spreadsheet is visible, showing columns for User Header, User Row ID, Age, SleepQuality, Diabetes, Depression, UPDRS, MoCA, and Functionality assessment. The 'Diagnosis' column (Col13) contains the predicted values for each row.

The predictions will appear on the output spreadsheet.

The screenshot shows the output spreadsheet with the following columns: User Header, User Row ID, Age, SleepQuality, Diabetes, Depression, UPDRS, MoCA, Functionality assessment, and Prediction. The 'Prediction' column (Col13) contains the predicted values for each row.

User Header	User Row ID	Age	SleepQuality	Diabetes	Depression	UPDRS	MoCA	Functionality assessment	Prediction
1	1.3195736895	1.2979847726	-0.422191014	-0.497070226	-1.667823523	1.6596162326	-1.164613026	1.13	0.0
2	44417	59347	2594651	279207	881176	67085	2845473	810	1.0
3	0.4561887577	-0.807465789	-0.422191014	-0.497070226	-1.123368916	-0.291879486	-0.072702906	-0.8	1.0
4	950968	2195454	2594651	279207	1800788	54866797	93082937	787	1.0
5	-1.529596585	-1.611887053	-0.422191014	-0.497070226	-0.847011635	0.7472598117	-0.546897452	1.13	0.0
6	2283393	6824403	2594651	279207	0645431	888634	6452395	810	0.0
7	1.4922506758	1.3441641305	-0.422191014	-0.497070226	-1.396969850	-0.754748229	-0.611532797	-0.8	1.0
8	942808	599965	2594651	279207	983372	7274792	73633	787	1.0
9	0.7152042373	-0.590375935	-0.422191014	-0.497070226	-1.120774114	0.2988335043	-1.527894232	1.13	1.0
10	198929	2021066	2594651	279207	870105	527822	8850977	810	0.0
11	0.0244962919	0.5044044309	2.3670960632	-0.497070226	1.158056692	-0.438678878	-1.294284259	1.13	0.0
12	20436786	6657	120632	279207	43347	411717	8271993	810	1.0
13	0.8878812236	-1.626599200	2.3670960632	-0.497070226	1.3644269192	-0.453575385	-0.53995639	-0.8	1.0
14	697569	9303631	120632	279207	165555	885566	6311483	787	1.0
15	0.1108347850	1.2763539038	-0.422191014	-0.497070226	0.2649337247	1.3173975088	0.1172891114	-0.8	1.0
16	953688	806688	2594651	279207	303821	67568	2624204	787	1.0
17	0.0244962919	-0.739174941	-0.422191014	-0.497070226	-0.099240872	-0.590678706	1.0668909333	-0.8	1.0
18	20436786	0912845	2594651	279207	88618245	2348167	111902	787	1.0
19	0.3698502646	0.8571332449	-0.422191014	-0.497070226	-0.174388196	0.1059269609	-1.660157347	1.13	1.0
20	2016485	257697	2594651	279207	09777073	5810913	7169023	810	0.0
21	1.4922506758	-0.830005817	-0.422191014	-0.497070226	0.7265672781	-0.604868997	-0.318813315	1.13	1.0
22	942808	8466565	2594651	279207	686168	804423	7414375	810	1.0
23	-1.184242612	-1.700555185	-0.422191014	-0.497070226	1.1428869934	0.5517198202	-1.278743863	-0.8	1.0
24	5286114	1134566	2594651	279207	076376	653817	2630568	787	1.0

## 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".

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
1	0.0244962919	1.6678701021	2.3670960632	-0.497070226	-0.584500796	1.7460788537	-0.9750181	
2	20436786	024709	120632	279207	7010027	283174	7875124	
3	-0.148180694	0.4138434815	-0.422191014	-0.497070226	0.0168980482	1.4489907373	0.61917381	
4	0.2942722	892542	2594651	279207	04466266	182524	821063	
5	-1.443258092	-0.826176450	-0.422191014	-0.497070226	1.2541237802	-1.066452328	-0.1324812	
6	0534073	97769	2594651	279207	254408	377751	52940432	
7	-1.011565626	-1.598488511	-0.422191014	-0.497070226	1.1495759649	1.2786323288	0.5896403	
8	1787473	7459724	2594651	279207	955632	492507	288322	
9	-1.615935078	1.6814919763	-0.497070226	-0.497070226	-0.322492833	1.1283220596	1.6221154	
10	4032713	333847	210142594653	279207	32095213	70138	49856	
11	-0.579873160	0.1042828706	-0.422191014	-0.497070226	-1.271730666	-0.216368704	-0.0675137	
12	3040873	6377791	2594651	279207	0348793	19530882	03088762	
13	-1.184242612	-0.054158872	2.3670960632	-0.497070226	-1.342815329	1.1715897323	1.3762544	
14	5286114	71660186	120632	279207	8501642	167745	907616	
15	-0.061842201	-1.412454916	-0.422191014	-0.497070226	-1.464229858	0.4077082678	1.45698371	
16	25449522	1838897	2594651	279207	2851163	472904	16042	
17	0.1971732782	1.6030415128	-0.422191014	-0.497070226	0.1068956410	-0.658244038	-0.4322105	
18	7030082	820563	2594651	279207	3486997	9160835	98672206	
19	-0.83888639	0.9653886831	-0.422191014	-0.497070226	2.0105140781	-0.828815656	-0.132580761	
20	8288833	673045	2594651	7724	0031603	50572193	3378413	
21	0.2835117714	-1.080226248	-0.422191014	-0.497070226	-0.198149496	-0.291938509	0.99570611	
22	452328	1200657	2594651	279207	23637082	83561115	092843	
23	-1.011565626	-1.481138563	-0.422191014	-0.497070226	0.6634143809	0.0503479774	-0.9799634	

To validate the model browse: "Analytics" → "Existing Model Utilization". Then choose Model "(from Tab:) TRAIN\_MODEL (.fit)".



The predictions will appear on the output spreadsheet.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)
1	0.0244962919	1.6678701021	2.3670960632	-0.497070226	-0.584500796	1.7460788537	-0.975018839	1.133	
2	20436786	024709	120632	279207	7010027	283174	7875124	8109	
3	-0.148180694	0.4138434815	-0.422191014	-0.497070226	0.0168980482	1.4498907373	0.6191738544	-0.88	
4	42942722	892542	2594651	279207	04466266	182524	821063	7877	
5	-1.443258092	-0.826176450	-0.422191014	-0.497070226	1.2541237802	-1.066452328	-0.132481204	1.133	
6	0534073	97769	2594651	279207	254408	377751	52940432	8109	
7	-1.011565626	-1.598488511	-0.422191014	-0.497070226	1.1495759649	1.2786323288	0.5896403296	1.133	
8	1787473	7459724	2594651	279207	955632	492507	288322	8109	
9	-1.615935078	1.6814919763	-0.422191014	-0.497070226	-0.322492833	1.1283220596	1.6221154567	1.133	
10	4032713	333847	2594651	279207	32095213	70138	49856	8109	
11	-0.579873160	0.1042828706	-0.422191014	-0.497070226	-1.271730666	-0.216368704	-0.067513722	-0.88	
12	3040873	6377791	2594651	279207	0348793	19530882	03088762	7877	

## 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".

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)
1	0.0244962919	1.6678701021	2.3670960632	-0.497070226	-0.584500796	1.7460788537	-0.975018839	1.133	
2	20436786	024709	120632	279207	7010027	283174	7875124	8109	
3	-0.148180694	0.4138434815	-0.422191014	-0.497070226	0.0168980482	1.4498907373	0.6191738544	-0.88	
4	42942722	892542	2594651	279207	04466266	182524	821063	7877	
5	-1.443258092	-0.826176450	-0.422191014	-0.497070226	1.2541237802	-1.066452328	-0.132481204	1.133	
6	0534073	97769	2594651	279207	254408	377751	52940432	8109	
7	-1.011565626	-1.598488511	-0.422191014	-0.497070226	1.1495759649	1.2786323288	0.5896403296	1.133	
8	1787473	7459724	2594651	279207	955632	492507	288322	8109	
9	-1.615935078	1.6814919763	-0.422191014	-0.497070226	-0.322492833	1.1283220596	1.6221154567	1.133	
10	4032713	333847	2594651	279207	32095213	70138	49856	8109	
11	-0.579873160	0.1042828706	-0.422191014	-0.497070226	-1.271730666	-0.216368704	-0.067513722	-0.88	
12	3040873	6377791	2594651	279207	0348793	19530882	03088762	7877	

Calculate the statistical metrics for the classification by browsing: "Statistics" → "Model Metrics" → "Classification Metrics".

The screenshot shows the 'Statistics' menu with options: 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' sub-menu is open, showing 'Regression Metrics' and 'Classification Metrics'. Below the menu is a data table with columns: User Header, Col1, Col2 (D), Col3 (D), Col4 (D), Col5 (D), Col6 (D), and Col7 (D). The table contains 12 rows of data.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)
1	0.0244962919	1.6678701021	2.3670960632	-0.497070226	-0.584500796	1.7	
2	20436786	024709	120632	279207	7010027	28	
3	-0.148180694	0.4138434815	-0.422191014	-0.497070226	0.0168980482	1.4	
4	42942722	892542	2594651	279207	0446266	18	
5	-1.443258092	517645097769	-0.422191014	-0.497070226	1.2541237802	-1	
6	0534073	517645097769	2594651	279207	254408	37	
7	-1.011565626	-1.598488511	-0.422191014	-0.497070226	1.1495759649	1.2	
8	1787473	7459724	2594651	279207	955632	49	
9	1.615935078	1.6814919763	-0.422191014	-0.497070226	0.3224902833	1.1	
10	4032713	333847	2594651	279207	32095213	70	
11	-0.579873160	0.1042828706	-0.422191014	-0.497070226	1.271730666	-0	
12	3040873	6377791	2594651	279207	0348793	19	
13	-1.184242612	-0.054158872	2.3670960632	-0.497070226	-1.342815329	1.1	
14	5286114	71660186	120632	279207	8501642	16	
15	-0.061842201	-1.412454916	-0.422191014	-0.497070226	-1.464229858	0.4	
16	25449522	1836897	2594651	279207	2851163	47	
17	0.1971732782	1.6030415128	-0.422191014	-0.497070226	0.1068956410	-0	
18	7030082	820563	2594651	279207	3486997	91	
19	-0.83888639	0.9653886831	-0.422191014	-0.497070226	0.828815656	-0	
20	8288833	673045	2594651	7724	0031603	50	
21	0.2835117714	-1.080226248	-0.422191014	-0.497070226	-0.198149496	-0	
22	452328	1200657	2594651	279207	23637082	83	
23	-1.011565626	-1.481138563	-0.422191014	-0.497070226	0.6634143809	0.0	
24	7030082	820563	2594651	7724	0031603	50	

The dialog box 'Classification Statistics Metrics' has the following settings:

- Actual Value Column: Col13 -- Diagnosis
- Prediction Value Column: Col14 -- Prediction
- beta of F Score: 2
- Buttons: Execute, Cancel

The results will appear on the output spreadsheet.

Accuracy: 0.93

F1-Score = 0.9199

The screenshot shows the 'Statistics' menu with options: 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' sub-menu is open, showing 'Regression Metrics' and 'Classification Metrics'. Below the menu is a data table with columns: User Header, Col1, Col2 (D), Col3 (D), Col4 (D), Col5 (D), Col6 (D), and Col7 (D). The table contains 12 rows of data.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)
1	0.0244962919	1.6678701021	2.3670960632	-0.497070226	-0.584500796	1.7	
2	20436786	024709	120632	279207	7010027	28	
3	-0.148180694	0.4138434815	-0.422191014	-0.497070226	0.0168980482	1.4	
4	42942722	892542	2594651	279207	0446266	18	
5	-1.443258092	517645097769	-0.422191014	-0.497070226	1.2541237802	-1	
6	0534073	517645097769	2594651	279207	254408	37	
7	-1.011565626	-1.598488511	-0.422191014	-0.497070226	1.1495759649	1.2	
8	1787473	7459724	2594651	279207	955632	49	
9	1.615935078	1.6814919763	-0.422191014	-0.497070226	0.3224902833	1.1	
10	4032713	333847	2594651	279207	32095213	70	
11	-0.579873160	0.1042828706	-0.422191014	-0.497070226	1.271730666	-0	
12	3040873	6377791	2594651	279207	0348793	19	
13	-1.184242612	-0.054158872	2.3670960632	-0.497070226	-1.342815329	1.1	
14	5286114	71660186	120632	279207	8501642	16	
15	-0.061842201	-1.412454916	-0.422191014	-0.497070226	-1.464229858	0.4	
16	25449522	1836897	2594651	279207	2851163	47	
17	0.1971732782	1.6030415128	-0.422191014	-0.497070226	0.1068956410	-0	
18	7030082	820563	2594651	279207	3486997	91	
19	-0.83888639	0.9653886831	-0.422191014	-0.497070226	0.828815656	-0	
20	8288833	673045	2594651	7724	0031603	50	
21	0.2835117714	-1.080226248	-0.422191014	-0.497070226	-0.198149496	-0	
22	452328	1200657	2594651	279207	23637082	83	
23	-1.011565626	-1.481138563	-0.422191014	-0.497070226	0.6634143809	0.0	
24	7030082	820563	2594651	7724	0031603	50	

## 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" by browsing: "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".

The screenshot shows the pd.ekk software interface. At the top, there is a menu bar with File, Edit, Data Transformation, Analytics, Statistics, Plot, and Help. Below the menu bar is a workflow diagram with several blue boxes connected by lines. The boxes are labeled: IMPORT, TRAIN\_SPLIT, NORMALISE\_TRAIN\_SET, NORMALISE\_TEST\_SET, FEATURE\_SELECTION\_REGRESSION, FEATURE\_SELECTION\_TRAIN\_SET, FEATURE\_SELECTION\_TEST\_SET, TRAIN\_MODEL\_LR, VALIDATE\_MODEL(predict), STATISTICS\_ACCURACIES, EXCLUDE\_DIAGNOSIS, and DOMAIN. The DOMAIN box is highlighted. Below the workflow diagram is a data table with 9 rows and 9 columns. The columns are labeled: User Header, User Row ID, Age, SleepQuality, Diabetes, Depression, UPDRS, MoCA, and FunctionalAssessment. The rows are numbered 1 to 9. The table contains numerical data for each row. To the right of the main table is a smaller table with 13 rows and 7 columns, labeled User Header, User Row ID, Col1, Col2, Col3, Col4, Col5, Col6, and Col7. This table is also numbered 1 to 13.

User Header	User Row ID	Age	SleepQuality	Diabetes	Depression	UPDRS	MoCA	FunctionalAssessment
1	1	1.3195736895	1.2979847726	-0.422191014	-0.497070226	-1.667823523	1.6596162326	-1.164613026
2	2	44417	59347	2594651	279207	881176	67085	2845473
3	3	0.4561887577	-0.807465789	-0.422191014	-0.497070226	0.99161800768	-0.291879486	-0.072702906
4	4	950968	2195454	2594651	279207	54866797	93082937	7877635
5	5	-1.529596585	-1.611887053	-0.422191014	-0.497070226	-0.847011635	0.7472598117	-0.546697452
6	6	2283393	6824403	2594651	279207	0645431	888634	6452395
7	7	1.4922506758	1.3441641305	-0.422191014	-0.497070226	-1.396969850	-0.754748229	-0.611532797
8	8	942808	599965	2594651	279207	983372	7374792	73633
9	9	0.7152042373	-0.590375935	-0.422191014	-0.497070226	-1.120774114	0.2968335043	-1.527894232

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

The screenshot shows the pd.ekk software interface. The 'Statistics' menu is open, showing options: Domain - APD, Model Metrics, Probability Distribution Functions, Descriptive Statistics, Confidence Intervals, Hypothesis Testing, Weight Cases, Random Number Generator, and Design of Experiments. The 'Domain - APD' option is selected. Below the menu is a data table with 9 rows and 9 columns, identical to the one in the previous screenshot.

User Header	User Row ID	Age	SleepQuality	Diabetes	Depression	UPDRS	MoCA	FunctionalAssessment
1	1	1.3195736895	1.2979847726	-0.422191014	-0.497070226	-1.667823523	1.6596162326	-1.164613026
2	2	44417	59347	2594651	279207	881176	67085	2845473
3	3	0.4561887577	-0.807465789	-0.422191014	-0.497070226	0.99161800768	-0.291879486	-0.072702906
4	4	950968	2195454	2594651	279207	54866797	93082937	7877635
5	5	-1.529596585	-1.611887053	-0.422191014	-0.497070226	-0.847011635	0.7472598117	-0.546697452
6	6	2283393	6824403	2594651	279207	0645431	888634	6452395
7	7	1.4922506758	1.3441641305	-0.422191014	-0.497070226	-1.396969850	-0.754748229	-0.611532797
8	8	942808	599965	2594651	279207	983372	7374792	73633
9	9	0.7152042373	-0.590375935	-0.422191014	-0.497070226	-1.120774114	0.2968335043	-1.527894232

The screenshot shows the 'Domain - APD' dialog box. It contains the formula  $APD = d + Z\sigma$ , where  $Z$  is set to 0.5. Below the formula is a dropdown menu labeled 'Perform Computations' with the option 'CPU (double precision)' selected. At the bottom 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 workflow diagram with nodes: IMPORT, TRAIN\_TEST\_SPLIT, NORMALISE\_TRAIN\_SET, FEATURE\_SELECTION\_REGRESSION, FEATURE\_SELECTION\_TRAIN\_SET, FEATURE\_SELECTION\_TEST\_SET, TRAIN\_MODEL, VALIDATE\_MODEL, STATISTICS\_ACCURACIES, EXCLUDE\_DIAGNOSIS, and DOMAIN. Below the diagram, there are two data tables. The left table is a spreadsheet with columns: User Header, Col1, Col2 (D), Col3 (D), Col4 (D), Col5 (D), Col6 (D), Col7 (D), Col8 (D), and Col9 (D). The right table is a similar spreadsheet with columns: User Header, User Row ID, Domain, APO, Prediction, Col5, Col6, and Col7. Both tables contain numerical data for 9 rows.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)
1	1.3195736895	1.2979847726	-0.422191014	-0.497070226	-1.667823523	1.8596162326	-1.164013026	1.1332608705	
2	44417	59347	2594651	279207	681176	67085	2845473	810912	
3	0.4561807577	-0.807465789	-0.422191014	-0.497070226	1.123368916	-0.291879486	-0.072702906	-0.881850519	
4	950968	2195454	2594651	279207	1800788	54866797	93082937	7877635	
5	-1.529596585	-1.611887053	-0.422191014	-0.497070226	-0.847011635	0.7472598117	-0.546897452	1.1332608705	
6	2.283393	6824403	2594651	279207	0645431	888634	6452395	810912	
7	1.4922506758	1.3441641305	-0.422191014	-0.497070226	-1.399696850	-0.754748229	-0.611532797	-0.881850519	
8	942808	599965	2594651	279207	983372	7214792	73633	7877635	
9	0.7152042373	-0.590375935	-0.422191014	-0.497070226	-1.120774114	0.2988335043	-1.527894232	1.1332608705	
10	198929	2021066	2594651	279207	870105	527822	8850977	810912	
11	0.0244962919	0.5044044309	2.3670960632	-0.497070226	-0.438678878	411717	-1.294284259	1.1332608705	
12	20436786	6657	120632	279207	42347	8271993	810912	7877635	
13	0.8878812236	-1.626599200	2.3670960632	-0.497070226	1.3644269192	-0.453575385	-0.530995639	-0.881850519	
14	697569	9303631	120632	279207	165555	8855606	6311483	7877635	
15	0.1108347850	1.2763539038	-0.422191014	-0.497070226	0.2649337247	1.3173975088	0.1172891114	-0.881850519	
16	953688	806688	2594651	279207	7724	303821	67568	7877635	
17	0.0244962919	-0.7301748491	-0.422191014	-0.497070226	0.096240872	-0.590678706	1.066899333	-0.881850519	
18	20436786	0912845	2594651	279207	88618245	2348167	111902	7877635	

## 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".

The screenshot shows the Isalos Analytics Platform interface. At the top, there is a workflow diagram with nodes: IMPORT, TRAIN\_TEST\_SPLIT, NORMALISE\_TRAIN\_SET, FEATURE\_SELECTION\_REGRESSION, FEATURE\_SELECTION\_TRAIN\_SET, FEATURE\_SELECTION\_TEST\_SET, TRAIN\_MODEL, VALIDATE\_MODEL, STATISTICS\_ACCURACIES, EXCLUDE\_DIAGNOSIS, and DOMAIN. Below the diagram, there are two data tables. The left table is a spreadsheet with columns: User Header, Col1, Col2 (D), Col3 (D), Col4 (D), Col5 (D), Col6 (D), Col7 (D), Col8 (D), and Col9 (D). The right table is a similar spreadsheet with columns: User Header, User Row ID, Domain, APO, Prediction, Col5, Col6, and Col7. Both tables contain numerical data for 13 rows.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)
1	0.0244962919	1.6678701021	2.3670960632	-0.497070226	-0.584500796	1.7460788537	-0.975018839	1.1332608705	
2	20436786	024709	120632	279207	7010027	283174	7875124	810912	
3	-0.148180694	0.4138434815	-0.422191014	-0.497070226	0.0168980482	1.4498907373	0.6191738544	-0.881850519	
4	42942722	892542	2594651	279207	04466266	182524	821063	7877635	
5	-1.443258092	-0.826176450	-0.422191014	-0.497070226	1.2541237802	-1.066453238	-0.132481204	1.1332608705	
6	0534073	97769	2594651	279207	254408	377751	5294032	810912	
7	-1.011565626	-1.598488511	-0.422191014	-0.497070226	1.1495759649	1.2786323288	0.5896403296	1.1332608705	
8	1787473	7459724	2594651	279207	955632	492507	288332	810912	
9	-1.615935078	1.6814919763	-0.422191014	-0.497070226	-0.322492833	1.283220596	1.6221154567	1.1332608705	
10	4032713	333847	2594651	279207	32095213	70138	49856	810912	
11	-0.579873160	0.1042828706	-0.422191014	-0.497070226	-1.271730666	-0.216368704	-0.067513722	-0.881850519	
12	3040873	6377791	2594651	279207	0348793	19530882	03088762	7877635	
13	-1.184242612	-0.054158872	2.3670960632	-0.497070226	-1.342815329	1.1715897323	1.3762544957	1.1332608705	
14	5286114	71660186	120632	279207	167745	8501642	907616	810912	
15	-0.061842201	-1.412454916	-0.422191014	-0.497070226	-1.464228658	0.4077082678	1.4568837014	-0.881850519	
16	25449522	1836897	2594651	279207	2851163	472904	16042	7877635	
17	0.1971732782	1.6030415128	-0.422191014	-0.497070226	0.1068956410	-0.658244038	-0.432210554	1.1332608705	
18	7030082	820563	2594651	279207	3486997	9160835	98672206	810912	

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

The screenshot shows the Isalos Analytics Platform interface. On the left, a workflow diagram includes steps like 'IMPORT', 'Data Manipulation', 'Split', 'Variable Selection', 'Matrix Transpose', 'Sort by Column', and 'Fill Missing Column(s) Values'. Below this is a data table with columns: User Header, User Row ID, Age, SleepQuality, Diabetes, Depression, UPDRS, MoCA, and FunctionalAssessment. A 'Select Column(s)' dialog is open on the right, showing 'Excluded Columns' (Col13 -- Diagnosis) and 'Included Columns' (Col5 -- Depression, Col6 -- UPDRS, Col7 -- MoCA, Col8 -- FunctionalAssessment, Col9 -- Tremor, Col10 -- Rigidity, Col11 -- Bradykinesia, Col12 -- PosturalInstability). The 'Execute' button is visible at the bottom of the dialog.

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”.

The screenshot shows the Isalos Analytics Platform interface. On the left, a workflow diagram includes steps like 'IMPORT', 'Data Manipulation', 'Split', 'Variable Selection', 'Matrix Transpose', 'Sort by Column', and 'Fill Missing Column(s) Values'. Below this is a data table with columns: User Header, User Row ID, Age, SleepQuality, Diabetes, Depression, UPDRS, MoCA, FunctionalAssessment, and Tremor. The table contains 13 rows of data. The 'RELIABILITY' tab is selected at the bottom of the interface.

Check the Reliability of the test set predictions by browsing: “Analytics” → “Existing Model Utilization”. Then select as Model “(from Tab:) DOMAIN”.



The screenshot shows the 'Existing Model Execution' dialog box on the right, which is open over a data table. The dialog has a 'Model' dropdown set to '(from Tab: JDOMAIN)' and a 'Type' dropdown set to 'APD Model'. Below these, there is a 'Description' field and a 'Model Input' section with a list of variables and their datatypes: Age (Double), Gender (Double), Ethnicity (Double), EducationLevel (Double), BMI (Double), Smoking (Double), AlcoholConsumption (Double), and PhysicalActivity (Double). At the bottom of the dialog are 'Execute' and 'Cancel' buttons.

The data table in the background has the following columns: User Header, User Row ID, Age, SleepQuality, Diabetes, Depression, UPDRS, MoCA, FunctionalAssessment, Tremor. The table contains 9 rows of data.

The results will appear on the output spreadsheet.

The screenshot shows the 'Existing Model Execution' dialog box on the right, which is open over a data table. The dialog has a 'Model' dropdown set to '(from Tab: JDOMAIN)' and a 'Type' dropdown set to 'APD Model'. Below these, there is a 'Description' field and a 'Model Input' section with a list of variables and their datatypes: Age (Double), Gender (Double), Ethnicity (Double), EducationLevel (Double), BMI (Double), Smoking (Double), AlcoholConsumption (Double), and PhysicalActivity (Double). At the bottom of the dialog are 'Execute' and 'Cancel' buttons.

The data table in the background has the following columns: User Header, User Row ID, Age, SleepQuality, Diabetes, Depression, UPDRS, MoCA, FunctionalAssessment, Tremor. The table contains 9 rows of data.

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:

