



Students' Performance (Multilabel Classification)

The goal of this study is to train a model in order to predict the grade class of high school students. The dataset used in this case study is found in <https://www.kaggle.com/datasets/rabieelkharoua/students-performance-dataset> and has 15 features and 2392 labelled samples. The dataset includes demographic details, study habits, parental involvement, extracurricular activities and academic performance.

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:

"GradeClass":

- 'A' - GPA ≥ 3.5 (0)
- 'B' - $3.0 \leq \text{GPA} < 3.5$ (1)
- 'C' - $2.5 \leq \text{GPA} < 3.0$ (2)
- 'D' - $2.0 \leq \text{GPA} < 2.5$ (3)
- 'F' - $\text{GPA} < 2.0$ (4)
-

"Gender":

- Male (0)
- Female (1)

"Ethnicity":

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

"ParentalEducation":

- None (0)
- High School (1)
- Some College (2)
- Bachelor's (3)
- Higher (4)

"ParentalSupport":

- None (0)
- Low (1)
- Moderate (2)
- High (3)
- Very High (4)

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 Students' Performance data.

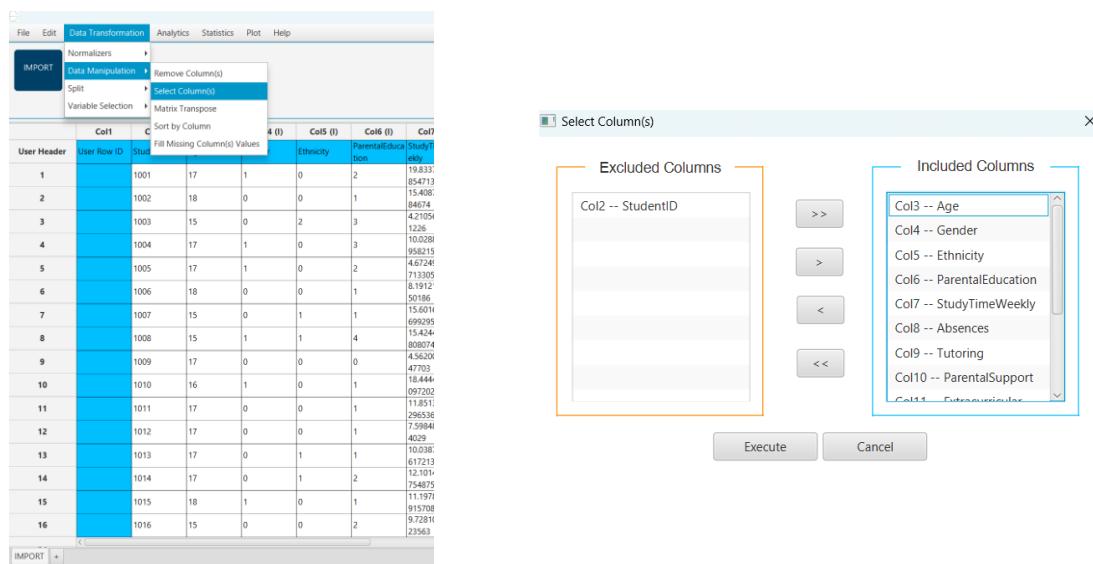
User Header	Col1	Col2	Col3	Col4	Col5	Col6
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						

User Header	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (D)	Col8 (I)	Col9 (I)	Col10 (I)	Col11 (I)	Col12 (I)	Col13 (I)
1	1001	17	1	0	2	19.831722807	7	1	2	0	0	0	1
2	1002	18	0	0	1	15.408756055	0	0	1	0	0	0	0
3	1003	15	0	2	3	4.2105697688	26	0	2	0	0	0	0
4	1004	17	1	0	3	10.028829473	14	0	3	1	0	0	0
5	1005	17	1	0	2	4.6724957279	17	1	3	0	0	0	0
6	1006	18	0	0	1	8.9171815452	0	0	1	1	0	0	0
7	1007	15	0	1	1	15.601680474	10	0	3	0	1	0	0
8	1008	15	1	1	4	15.424496305	22	1	1	1	0	0	0
9	1009	17	0	0	0	4.5620075580	1	0	2	0	1	0	0
10	1010	16	1	0	1	18.44466363	0	0	3	1	0	0	0
11	1011	17	0	0	1	11.851363655	11	0	1	0	0	0	0
12	1012	17	0	0	1	26.93348	15	0	2	0	0	0	0
13	1013	17	0	1	1	7.5984858192	15	0	3	1	0	0	0
14	1014	17	0	1	2	10.038711615	21	0	4	0	1	0	0
15	1015	18	1	0	1	12.101425068	21	0	1	2	0	0	0
16	1016	15	0	0	2	9.197810636	9	1	0	0	1	0	0

Step 2: Manipulate data

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

- On the menu click on "Data Transformation" → "Data Manipulation" → "Select Column(s)"
- Select all columns except the one that corresponds to the "StudentID".

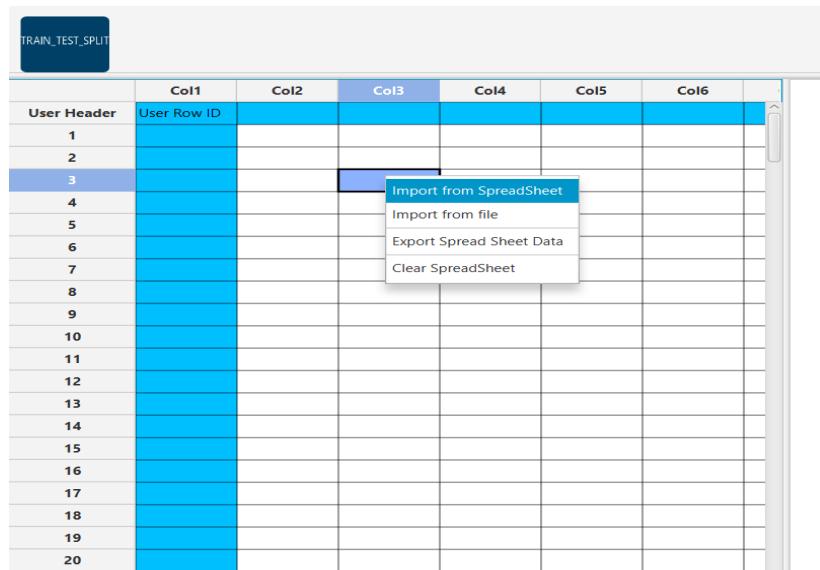


The data without the "StudentID" column will appear in the output spreadsheet.

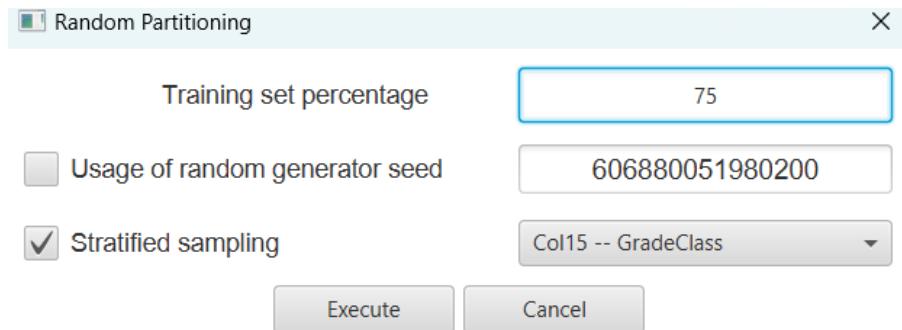
Step 3: Split data

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:



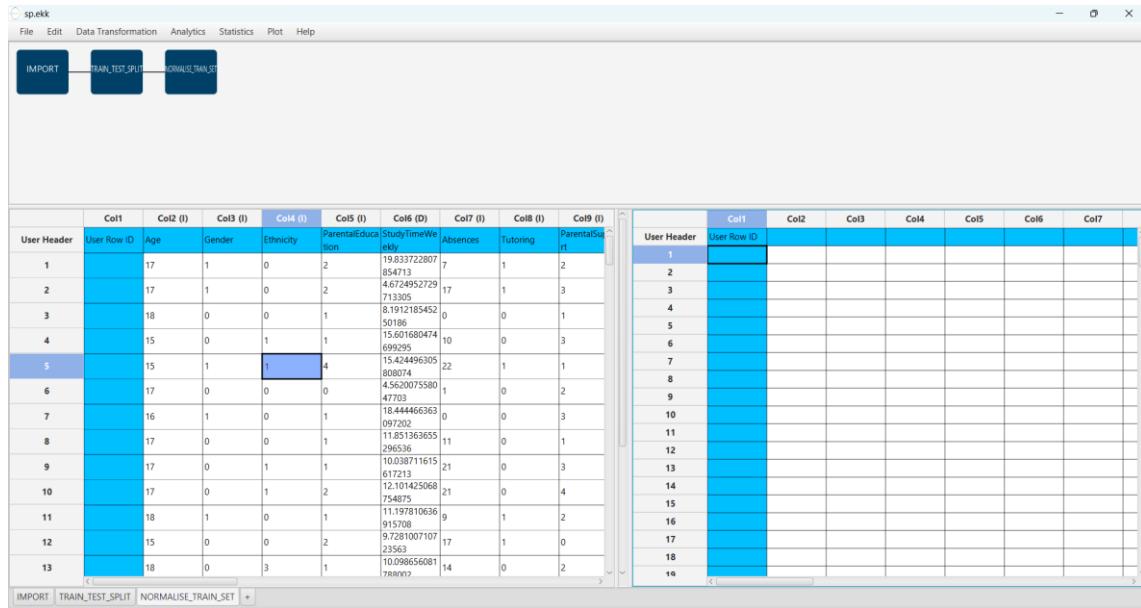
The results will appear on the output spreadsheet.

User Header	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (D)	Col7 (I)	Col8 (I)	Col9 (I)	User Header	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (D)	Col7 (I)	Col8 (I)	Col9 (I)
1		17	1	0	2	85433722807	7	1	2	1		17	1	0	2	19.833722807	7	1	
2		18	0	0	1	15.408756055	0	0	1	2		17	1	0	2	854713	17	1	
3		15	0	2	3	4.2105697688	26	0	2	3		18	0	0	1	4.6724952729	0	0	
4		17	1	0	3	10.038829473	14	0	3	4		15	0	1	1	8.1912185452	0	0	
5		17	1	0	2	4.6724952729	17	1	3	5		15	1	1	4	50186	10	0	
6		18	0	0	1	8.1912185452	0	0	1	6		17	0	0	0	15.424496305	22	1	
7		15	0	1	1	15.601680474	10	0	3	7		16	1	0	1	8.08074	0	0	
8		15	1	1	4	699295	22	1	1	8		17	0	0	1	11.851363655	21	0	
9		17	0	0	0	4.5620075580	1	0	2	9		17	0	1	1	296536	0	0	
10		16	1	0	1	18.444666362	0	0	3	10		17	0	1	2	10.038711615	21	0	
11		17	0	0	1	97202	11	0	1	11		18	1	0	1	617213	21	0	
12		17	0	0	1	7.5984858192	15	0	2	12		17	0	0	2	12.1014235068	17	1	
13		17	0	1	1	4029	21	0	3	13		18	0	3	1	754875	0	0	

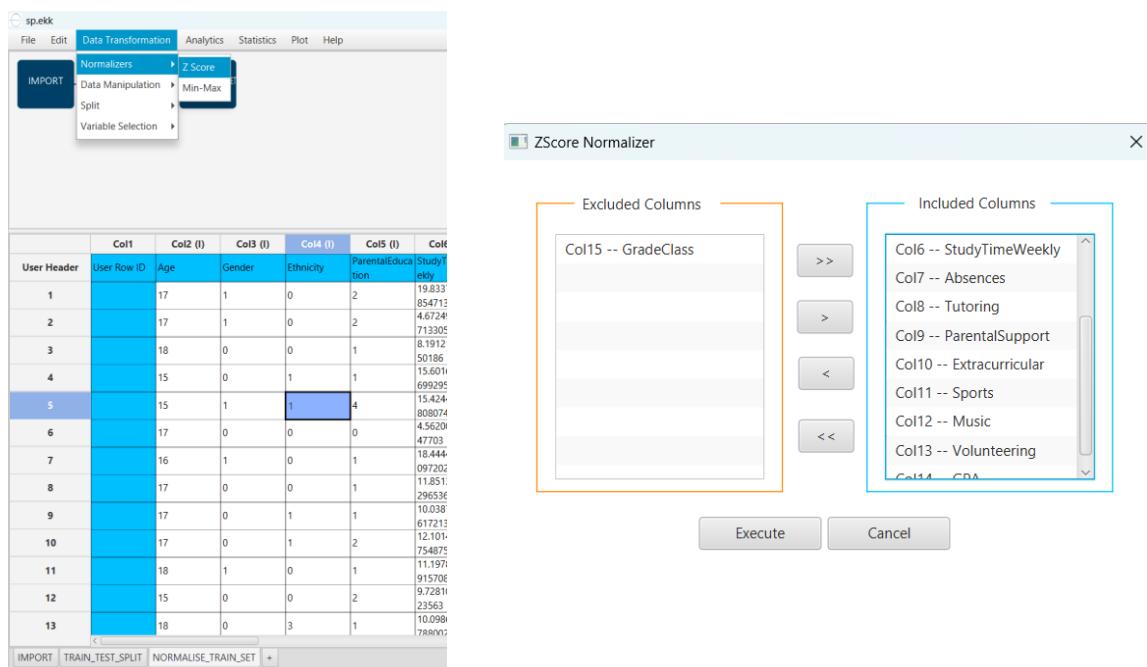
Step 4: Normalize the training set

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"



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



The results will appear on the output spreadsheet.

User Header	User Row ID	Age	Gender	Ethnicity	ParentalEducation	StudyTimeWeekly	Absences	Tutoring	ParentalSupport	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (D)	Col7 (I)	Col8 (I)	Col9 (I)
1	17	1	0	2	19.833722807	854713	7	1	2									
2	17	1	0	2	4.672495279	713305	17	1	3									
3	18	0	0	1	8.1912185452	90101	0	0	1									
4	15	0	1	1	15.801680474	699295	10	0	3									
5	15	1	1	4	15.424496305	808074	22	1	1									
6	17	0	0	0	4.5620075580	47703	1	0	2									
7	16	1	0	1	18.44446663	097202	0	0	3									
8	17	0	0	1	9.8601363655	11	0	1										
9	17	0	1	1	10.039711615	617213	21	0	3									
10	17	0	1	2	12.101425068	754875	21	0	4									
11	18	1	0	1	11.197810636	915708	9	1	2									
12	15	0	0	2	9.7281007107	23563	17	1	0									
13	18	0	3	1	10.998656081	788007	14	0	2									

User Header	User Row ID	Age	Gender	Ethnicity	ParentalEducation	StudyTimeWeekly	Absences	Tutoring	ParentalSupport	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Co
1	0.4881007033	0.9831409794	-0.832415968	0.2605448853	1.5												
2	0.4881007033	0.9831409794	-0.832415968	0.2605448853	1.5												
3	1.3752867843	-1.016581148	-0.832415968	-0.733960485	0.30006386												
4	1.286271458	0.9831409794	0.151996851	2.2495556268	0.9753186630	0.8821382341	1.5										
5	0.490501	0.788926	0.9023526	2.95642	0.650291	101976	1.5										
6	0.4881007033	0.9831409794	-0.16581148	-1.728465856	0.039971859	-1.607275685	0.286										
7	5.47801	0.788926	0.91213	4.161616	0.87634	3222355	0.286										
8	0.4881007033	-1.016581148	-0.832415968	-0.733960485	0.15078043726	-1.725819205	0.286										
9	0.490501	0.788926	0.9023526	2.95642	0.650291	101976	0.286										
10	0.4881007033	0.9831409794	-0.16581148	-1.728465856	0.039971859	-1.607275685	0.286										
11	0.490501	0.788926	0.9023526	2.95642	0.650291	101976	0.286										
12	0.4881007033	0.9831409794	-0.16581148	-0.832415968	0.3452984844	-0.421840485	0.286										
13	0.490501	0.788926	0.9023526	2.95642	0.650291	101976	0.286										

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	User Row ID	Age	Gender	Ethnicity	ParentalEducation	StudyTimeWeekly	Absences	Tutoring	ParentalSupport	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (D)	Col7 (I)	
1	18	0	0	1	15.08756055	84674	0	0	1								
2	15	0	2	3	4.2105697688	1226	26	0	2								
3	17	1	0	3	10.028829473	986215	14	0	3								
4	17	0	0	1	7.592858192	4029	15	0	2								
5	18	1	0	0	3.5282382085	577235	16	1	2								
6	18	1	0	1	7.3803546482	23455	15	0	2								
7	16	1	0	3	2.2521945869	844137	8	0	3								
8	18	0	0	1	5.3441979097	2446	26	1	1								
9	17	0	0	1	14.394046327	645153	7	0	2								
10	18	0	1	2	4.894319404	74019	7	0	4								
11	18	0	1	2	2.6313061327	870924	5	0	1								
12	16	1	1	1	19.441069492	469765	16	0	4								
13	17	1	2	2	17.186499021	anik2	5	0	3								

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 Isalos Analytics Platform interface. The top navigation bar includes File, Edit, Data Transformation, Analytics, Statistics, Plot, and Help. The Analytics menu is open, with Existing Model Utilization highlighted. Below the menu, a preview of a dataset is displayed in a spreadsheet format. The columns are labeled Col1 through Col6. The first row is labeled 'User Header'. The data rows show various values for each column. At the bottom of the preview, there are tabs for IMPORT, TRAIN_TEST_SPLIT, NORMALISE_TRAIN_SET, and NORMALISE_TEST_SET.

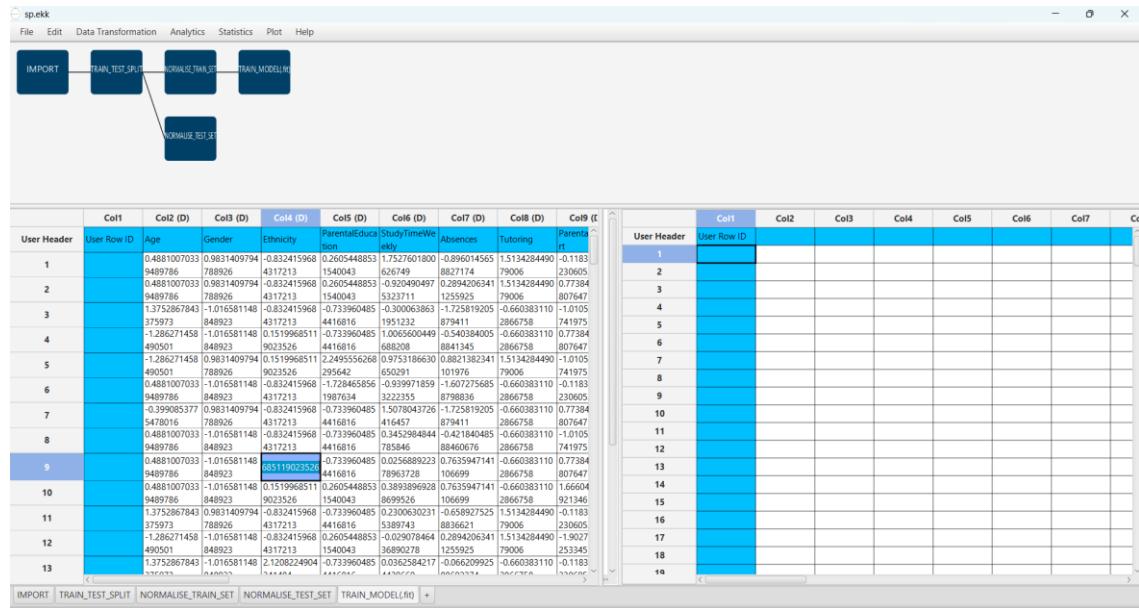
The results will appear on the output spreadsheet.

The screenshot shows the Isalos Analytics Platform interface with the NORMALISE_TRAIN_SET tab selected. The preview of the normalized training set shows the same structure as the previous screenshot, with columns Col1 through Col6 and a User Header row. The data values have been normalized. The bottom tabs remain the same: IMPORT, TRAIN_TEST_SPLIT, NORMALISE_TRAIN_SET, and NORMALISE_TEST_SET.

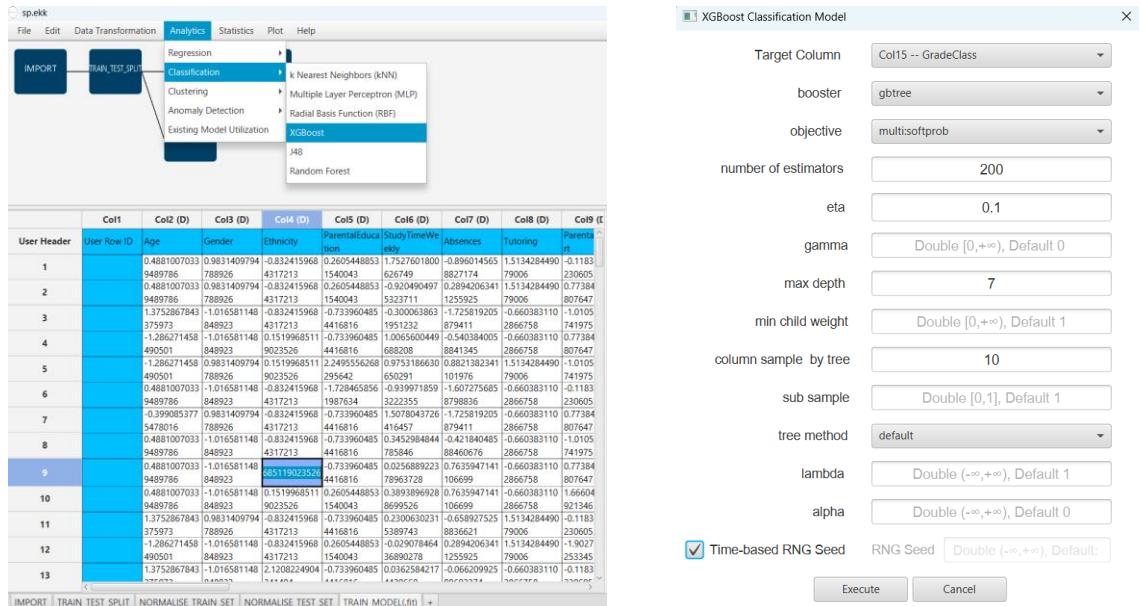
Step 6: Train the model

Create a new tab by pressing the "+" button on the bottom of the page with the name "TRAIN_MODEL(.fit)".

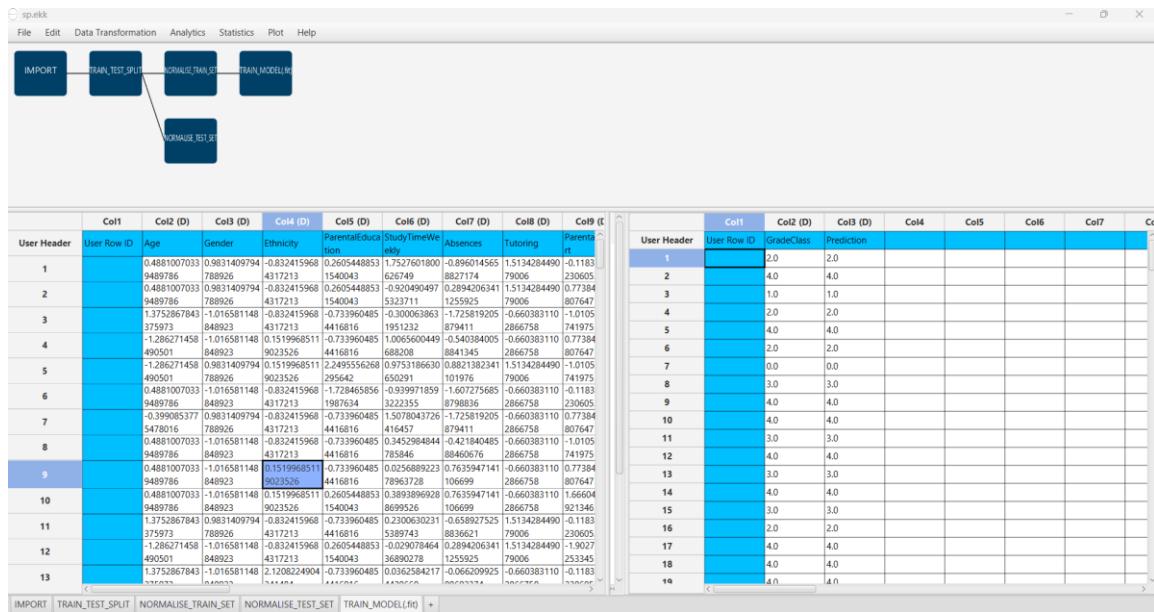
Import data into the input spreadsheet of the "TRAIN_MODEL(.fit)" tab from the output of the "NORMALISE_TRAIN_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".



Use the XGBoost Method to train and fit the model by browsing: "Analytics" → "Classification" → "XGBoost" and adjust model parameters based on training set performance.



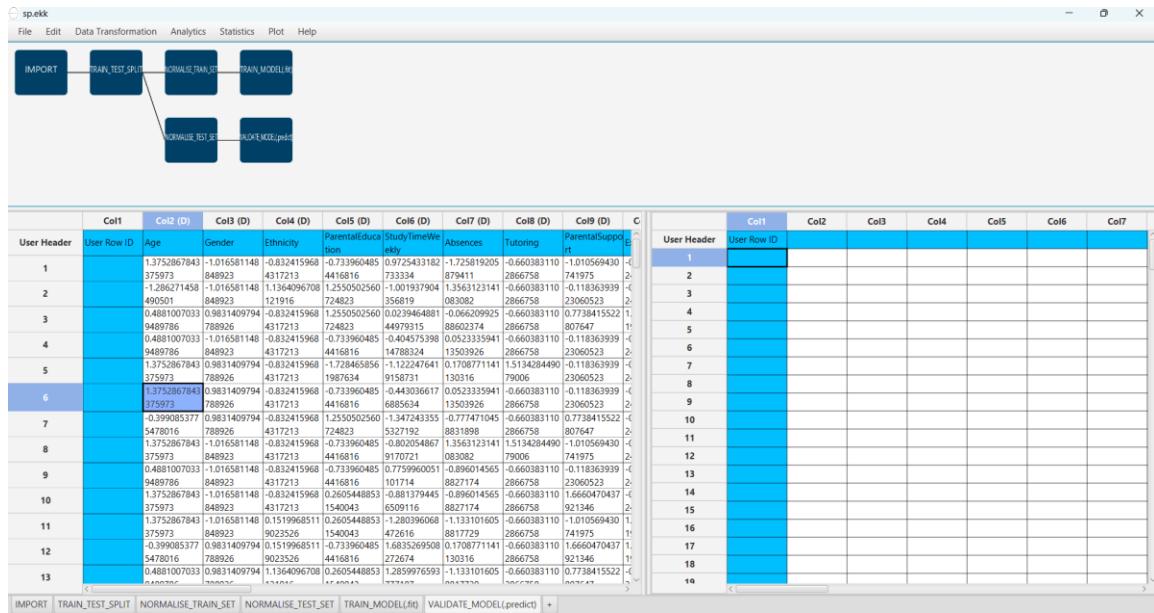
The predictions will appear on the output spreadsheet.



Step 7: Validate the model

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

Import data into the input spreadsheet of the "VALIDATE_MODEL(.predict)" tab from the output of the "NORMALISE_TEST_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".



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

The screenshot shows the Isalos Analytics Platform interface. The top navigation bar includes File, Edit, Data Transformation, Analytics, Statistics, Plot, and Help. The Analytics tab is currently selected. In the center, there is a flowchart with nodes: IMPORT, TRAIN_TEST_SPLIT, NORMALISE_TRAIN_SET, and TRAIN_MODEL(fit). A context menu is open over the 'Existing Model Utilization' node, listing options: Regression, Classification, Clustering, Anomaly Detection, and Existing Model Utilization. On the right side, a detailed configuration dialog for 'Existing Model Execution' is open. It shows the Model as '(from Tab:) TRAIN_MODEL(...)' and the Type as 'XGBoost classification mod'. The dialog includes sections for Model Input (with various columns mapped to Double types like Age, Gender, etc.) and Excluded Columns (listing columns like Col6 -- StudyTimeWeekly, Col7 -- Absences, etc.). The Included Columns section is empty. At the bottom are Execute and Cancel buttons.

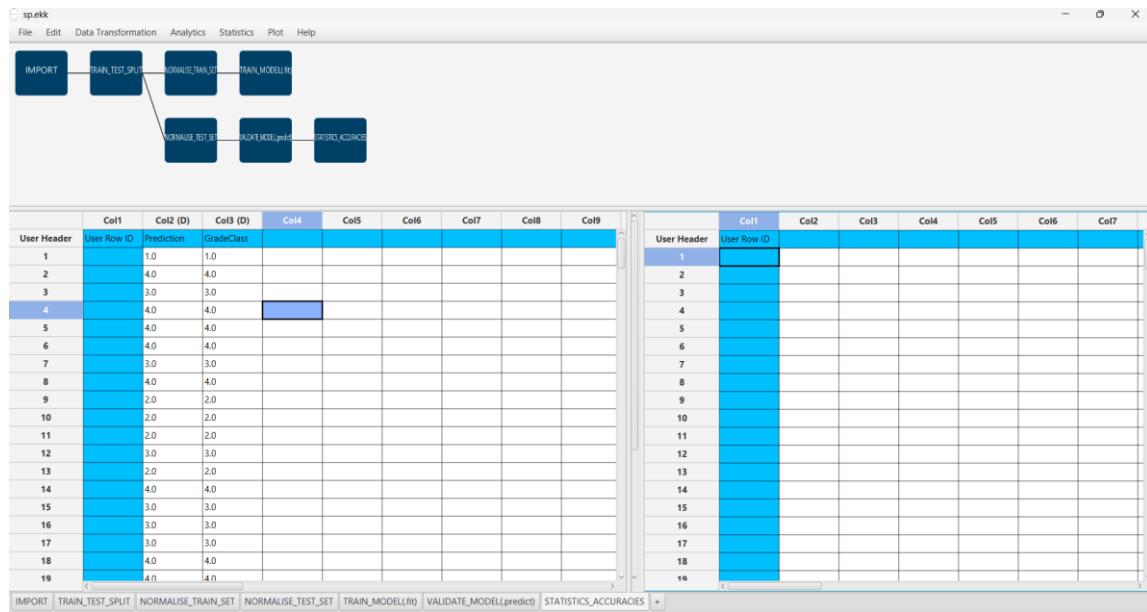
The predictions will appear on the output spreadsheet.

The screenshot shows the output spreadsheet from the Isalos Analytics Platform. The spreadsheet has 13 rows of data, each with the following columns: User Row ID, Age, Gender, Ethnicity, ParentalEducation, StudyTimeWeekly, Absences, Tutoring, ParentalSupport, Col6, Col7, Col8, Col9, and Col10. The last three columns (Col11, Col12, Col13) are empty. The data is identical to the input spreadsheet shown in the previous screenshot.

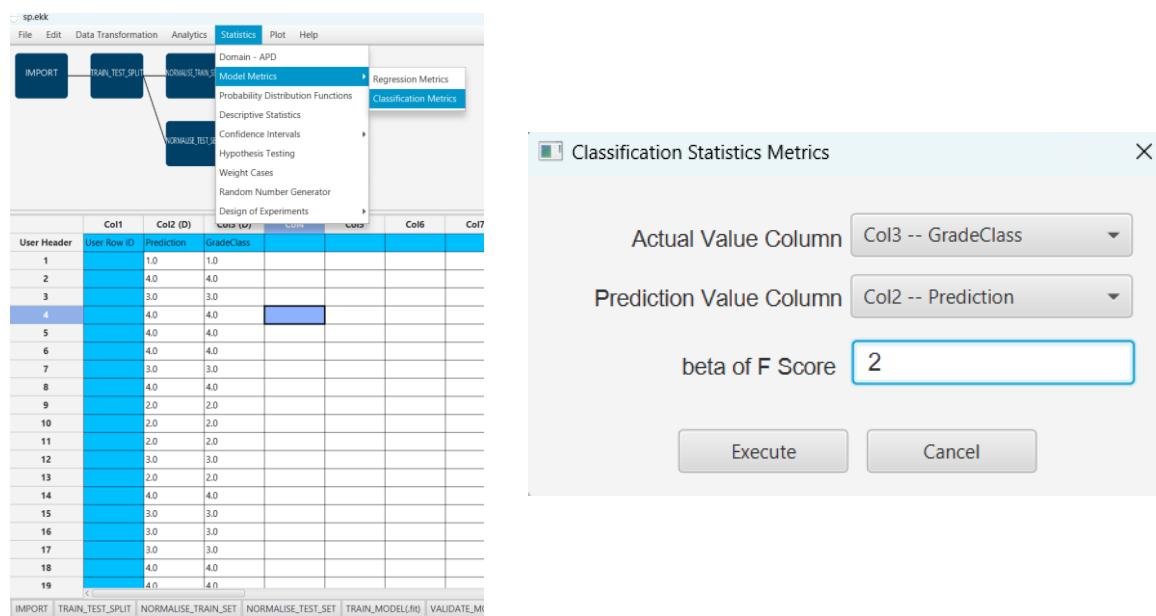
Step 8: Statistics calculation

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

Import data into the input spreadsheet of the "STATISTICS_ACCURACIES" tab from the output of the "VALIDATE_MODEL(.predict)" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".



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



The results will appear on the output spreadsheet.

Accuracy: 0.977

F1-Score = 0.954

The screenshot shows a software window titled "spekk" with a menu bar: File, Edit, Data Transformation, Analytics, Statistics, Plot, Help. Below the menu is a toolbar with buttons: IMPORT, TRAIN_TEST_SPLIT, NORMALISE_TRAIN_SET, and TRAIN_MODEL(R). The main area contains two tables side-by-side.

User Header	Col1	Col2 (D)	Col3 (D)	Col4	Col5	Col6	Col7
1		1.0	1.0				
2		4.0	4.0				
3		3.0	3.0				
4		4.0	4.0				
5		4.0	4.0				
6		4.0	4.0				
7		3.0	3.0				
8		4.0	4.0				
9		2.0	2.0				
10		2.0	2.0				
11		2.0	2.0				
12		3.0	3.0				
13		2.0	2.0				
14		4.0	4.0				
15		3.0	3.0				
16		3.0	3.0				
17		3.0	3.0				
18		4.0	4.0				
19		4.0	4.0				
20		2.0	2.0				
21		4.0	4.0				
22		4.0	4.0				
23		4.0	4.0				
24		4.0	4.0				
25		4.0	4.0				
26		4.0	4.0				

User Header	Col1 (S)	Col2 (D)	Col3 (S)	Col4 (S)	Col5 (S)	Col6 (S)	Col7 (S)	Col8	Col9
1			Predicted Class	Predicted Class	Predicted Class	Predicted Class			
2			1.0	4.0	3.0	2.0	0.0		
3	Actual Class	1.0	65	2	0	0	0		
4	Actual Class	4.0	0	302	0	1	0		
5	Actual Class	3.0	0	2	100	1	0		
6	Actual Class	2.0	0	0	1	97	0		
7	Actual Class	0.0	0	3	2	2	20		
8									
9									
10	Classification Accuracy	0.9765886287	625418						
11									
12	Precision		1.0	0.9773462783	0.9708737864	0.0603960396	1.0		
13									
14	Recall/Sensitivity	0.9701492537	0.9966996699	0.9708737864	0.9897959183	0.7407407407			
15									
16	Specificity	1.0	0.9762711864	0.9939393939	0.992	1.0			
17									
18	F1 Score	0.9848484848	0.9869281045	0.9708737864	0.074873787	0.8510638297			
19									
20	F(beta=2)	0.9759759759	0.9927679158	0.9708737864	0.0837728194	0.78125			
21									
22	MCC	0.9651351831	950721						

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

Step 9.a: Create the domain

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

Import data into the input spreadsheet of the "REMOVE_TARGET" 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 a software window titled "spekk" with a menu bar: File, Edit, Data Transformation, Analytics, Statistics, Plot, Help. Below the menu is a toolbar with buttons: IMPORT, TRAIN_TEST_SPLIT, NORMALISE_TRAIN_SET, and TRAIN_MODEL(R). The main area contains two tables side-by-side.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
1	User Row ID	Age	Gender	Ethnicity	ParentalEducation	StudyTimePerWeek	Absences	Tutoring
2	0.4881007033	0.9631409794	-0.832415960	0.2605448853	0.7527601000	-0.8960741565	1.5134284490	
3	0.4881007033	0.9831409794	-0.832415960	0.2605448853	0.920409497	0.2894206341	1.5134284490	
4	0.489786	0.789026	0.4317213	0.540403	0.5323711	0.255925	0.79006	
5	1.3752867843	-0.106581148	-0.832415960	-0.733960485	-0.000063863	-0.725819205	-0.660383110	
6	0.489786	0.849923	0.4317213	0.416816	0.195123	0.879411	0.2866758	
7	1.3752867843	-0.106581148	-0.832415960	-0.733960485	0.10656000449	0.540384005	-0.660383110	
8	0.489786	0.849923	0.4317213	0.416816	0.785846	0.88460676	0.2866758	
9	0.4881007033	-0.106581148	0.1519968511	-0.733960485	0.025689223	0.7635947141	-0.660383110	
10	0.489786	0.849923	0.4317213	0.416816	0.78963728	1.06699	0.2866758	
11	0.4881007033	-0.106581148	0.1519968511	0.2605448853	0.3893896928	0.7635947141	-0.660383110	
12	0.489786	0.849923	0.4317213	0.540403	0.36890278	0.255925	0.79006	
13	1.3752867843	-0.106581148	2.1208224904	0.733960485	0.0362584217	-0.066209925	-0.660383110	

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

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

User Header	User Row ID	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)	Col10 (D)	Col11 (D)	Col12 (D)	Col13 (D)	Col14 (D)	Col15 (D)	
1	0.4681007033	0.9831409794	0.832415968	0.2605448853	1.752761800	-0.896014565	1.51342										
2	0.4681007033	0.9831409794	0.832415968	0.2605448853	0.920490497	0.2894206341	1.51342										
3	1.3752867843	-1.016581148	0.832415968	-0.733960485	-0.300063863	-1.725819205	-0.603										
4	375973	848923	4317213	4416816	1951232	879411	286675										
5	1.286271458	-1.016581148	0.1519968511	-0.733960485	1.0065600449	0.54034005	-0.603										
6	490501	788926	9023526	9023526	688202	89345	286675										
7	0.4881007033	0.9831409794	0.832415968	-0.733960485	1.5078043725	1.725819205	-0.603										
8	5478016	788926	4317213	4416816	879411	286675											
9	0.4881007033	-1.016581148	0.832415968	-0.733960485	0.345294844	0.421840405	-0.603										
10	9489786	848923	4317213	4416816	789411	8846076	286675										
11	0.4881007033	-1.016581148	0.1519968511	-0.733960485	0.256889223	0.7635947141	-0.603										
12	490501	788926	9023526	9023526	86990278	1255925	79006										
13	1.3752867843	-1.016581148	2.120824904	-0.733960485	0.0362584217	0.066029925	-0.603										

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

User Header	User Row ID	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)	Col10 (D)	Col11 (D)	Col12 (D)	Col13 (D)	Col14 (D)	Col15 (D)	
1	0.4881007033	0.9831409794	0.832415968	0.2605448853	1.752761800	-0.896014565	1.51342										
2	0.4881007033	0.9831409794	0.832415968	0.2605448853	0.920490497	0.2894206341	1.51342										
3	1.3752867843	-1.016581148	0.832415968	-0.733960485	0.300063863	-1.725819205	-0.603										
4	375973	848923	4317213	4416816	1951232	879411	286675										
5	1.286271458	-1.016581148	0.1519968511	-0.733960485	0.0256889223	0.7635947141	-0.603										
6	490501	788926	9023526	9023526	688202	8841345	286675										
7	0.4881007033	-1.016581148	0.832415968	-0.733960485	0.256889223	0.7635947141	-0.603										
8	9489786	848923	4317213	4416816	879411	286675											
9	0.4881007033	-1.016581148	0.1519968511	-0.733960485	0.0256889223	0.7635947141	-0.603										
10	490501	788926	9023526	9023526	86990278	1255925	79006										
11	1.3752867843	-1.016581148	0.9831409794	-0.832415968	0.2300630231	0.658927525	1.51342										
12	375973	788926	4317213	4416816	5389743	8836621	286675										
13	1.286271458	-1.016581148	0.832415968	-0.733960485	0.0362584217	0.066029925	-0.603										

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

The screenshot shows the Isalos Analytics Platform interface. At the top, there is a navigation bar with tabs: File, Edit, Data Transformation, Analytics, Statistics, Plot, and Help. Below the navigation bar is a workflow diagram titled "Domain - APD". The diagram consists of several nodes connected by arrows: "IMPORT" leads to "TRAIN_TEST_SPLIT", which then connects to "NORMALISE_TRAIN_SET", "NORMALISE_TEST_SET", and "TRAIN_MODEL(R0)". "NORMALISE_TRAIN_SET" and "NORMALISE_TEST_SET" both lead to "DOMAIN". "DOMAIN" then connects to "REMOVE_TARGET", which finally connects to "NORMALISE_TEST_SET". A tooltip for the "DOMAIN" node says "APD = d + Zσ , Z= 0.5". Below the workflow diagram is a table with columns: Col1, Col2 (D), Col3 (D), Col4 (D), Col5 (D), Col6 (D), and Col7 (D). The table contains 11 rows of data, each with a User Row ID and various numerical values. At the bottom of the table are buttons for "IMPORT", "TRAIN_TEST_SPLIT", "NORMALISE_TRAIN_SET", "NORMALISE_TEST_SET", "TRAIN_MODEL(R0)", and "VALIDATE_MODEL(predict)".

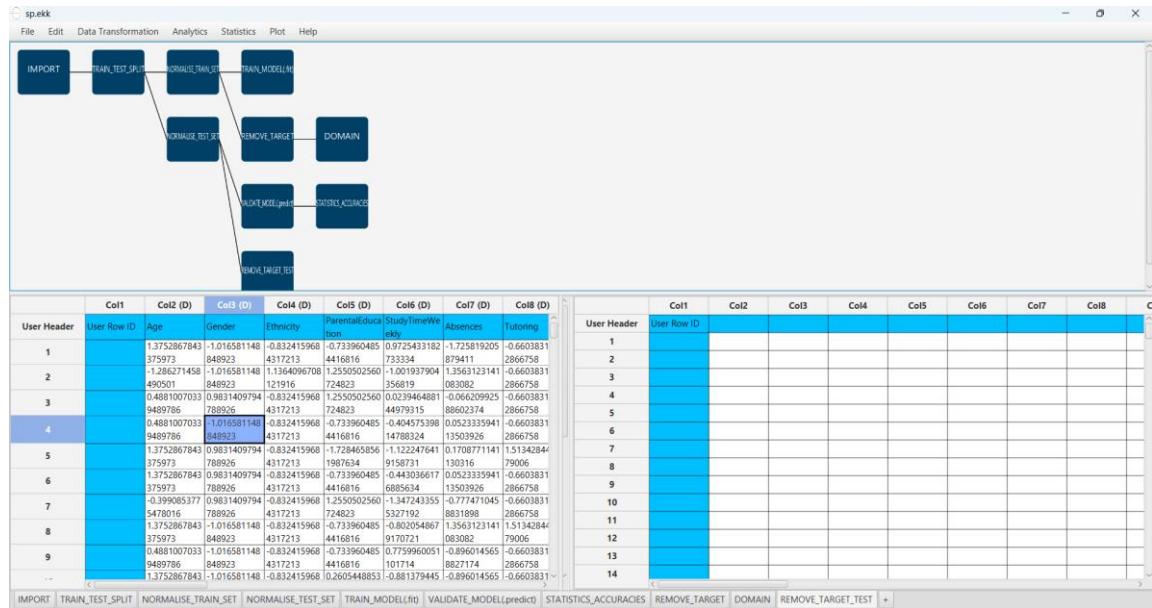
The results will appear on the output spreadsheet.

This screenshot shows a more detailed view of the Isalos Analytics Platform. The workflow diagram is identical to the one above, but the table below it is much larger and more detailed. It includes columns for "User Header" and "User Row ID", along with the same columns as the previous table (Col1 to Col7). The data rows are numbered 1 through 12, with row 12 being a continuation of the previous table's data. The table also includes a "Prediction" column at the end. At the bottom of the table are buttons for "IMPORT", "TRAIN_TEST_SPLIT", "NORMALISE_TRAIN_SET", "NORMALISE_TEST_SET", "TRAIN_MODEL(R0)", "VALIDATE_MODEL(predict)", "STATISTICS_ACCURACIES", and "REMOVE_TARGET".

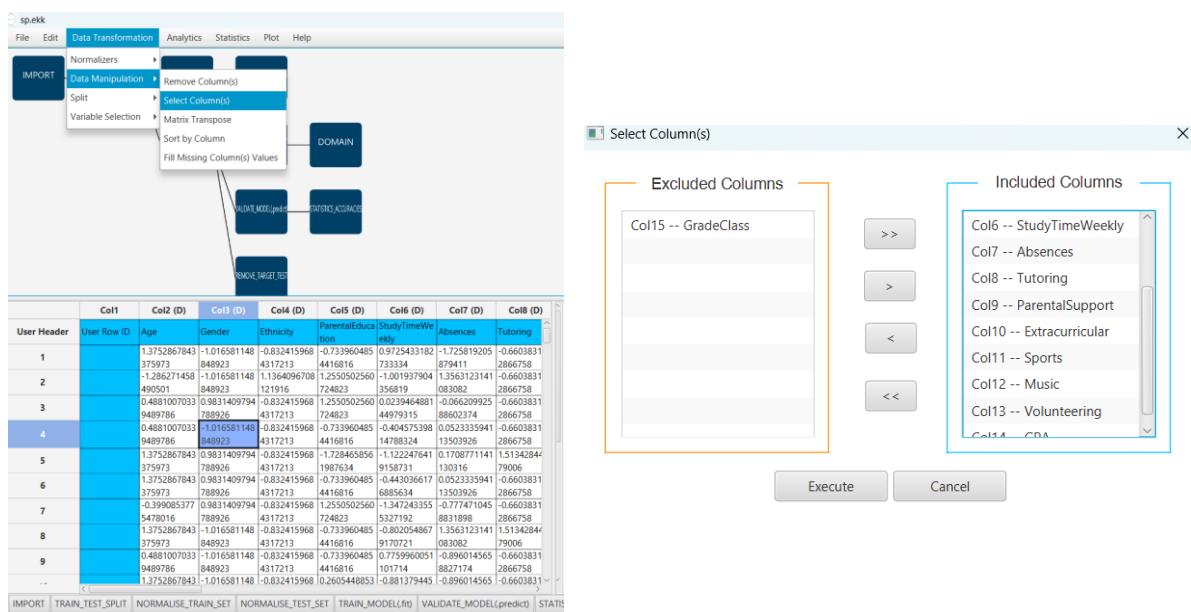
Step 9.b: Check the test set reliability

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

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



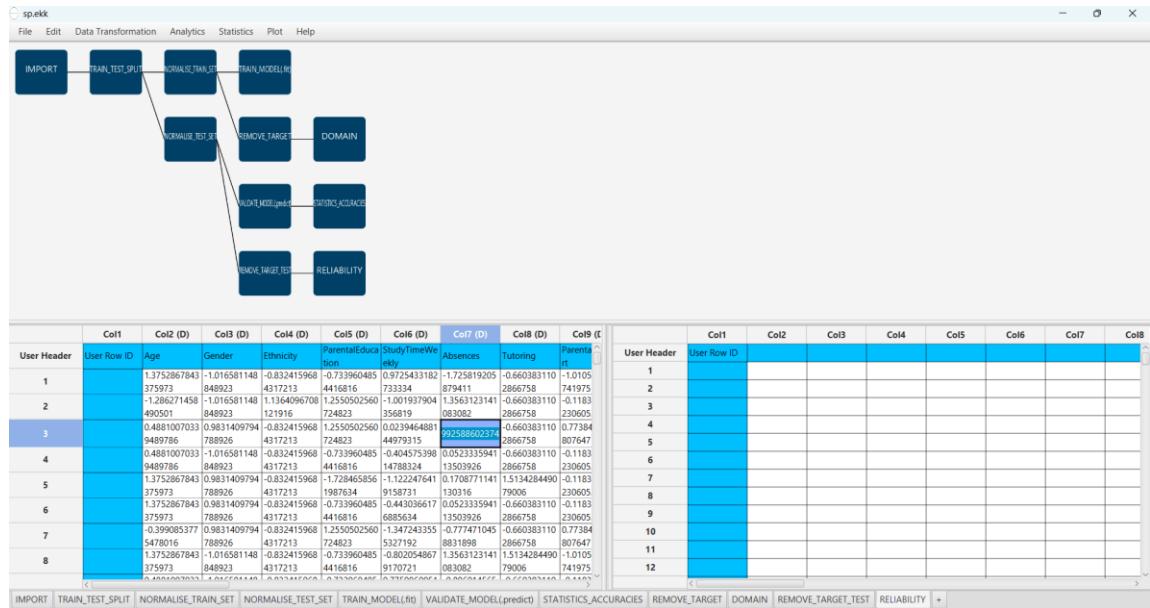
Filter the data to exclude the column that corresponds to the "GradeClass" by browsing: "Data Transformation" → "Data Manipulation" → "Select Column(s)". Then select all the columns except "GradeClass".



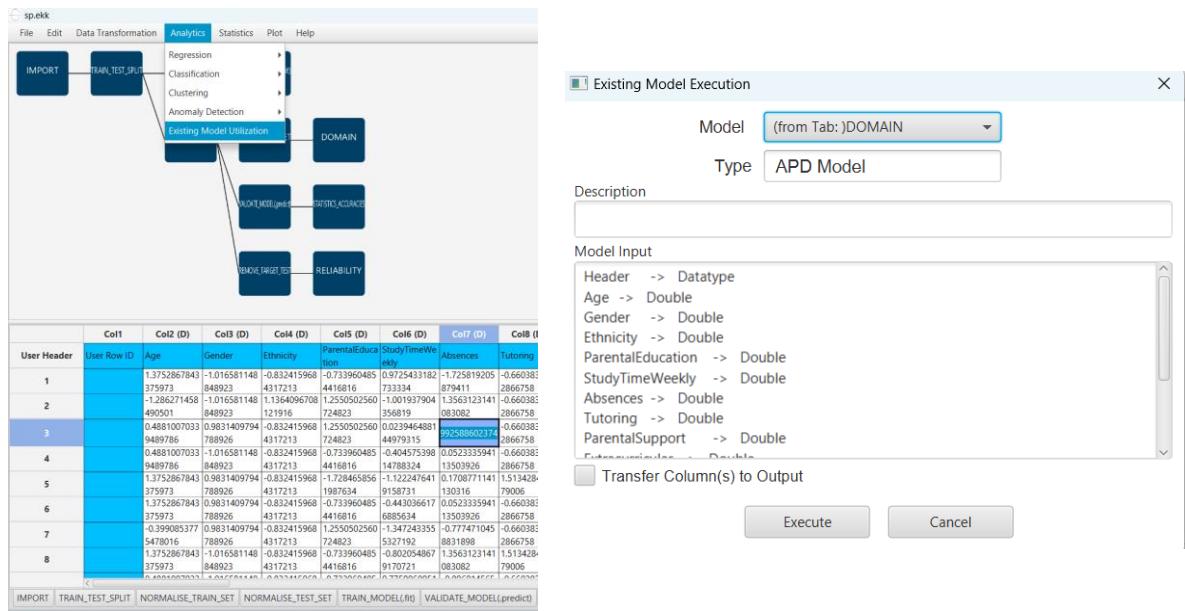
The results will appear on the output spreadsheet.

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



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



The results will appear on the output spreadsheet.

sp.ekk									
	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9
User Header	User Row ID	Age	Gender	Ethnicity	ParentalEducation	StudyTimePerWeek	Absences	Tutoring	FinalGrade
1	375973	0.9831409794	0.832415968	0.733960485	0.9725433182	1.72819205	0.660383110	1.01	7419
2	490501	0.9831409794	0.832415968	0.733960485	1.1364096708	1.255052560	1.001937904	0.660383110	0.11
3	9489786	0.9831409794	0.832415968	0.733960485	1.255052560	0.0239464881	99.538603374	0.660383110	0.77
4	9489786	0.9831409794	0.832415968	0.733960485	0.0239464881	99.538603374	0.660383110	0.77	8076
5	375973	0.9831409794	0.832415968	0.733960485	1.72819205	0.0239464881	99.538603374	0.660383110	0.11
6	375973	0.9831409794	0.832415968	0.733960485	1.72819205	0.0239464881	99.538603374	0.660383110	0.11
7	375973	0.9831409794	0.832415968	0.733960485	1.72819205	0.0239464881	99.538603374	0.660383110	0.11
8	375973	0.9831409794	0.832415968	0.733960485	1.72819205	0.0239464881	99.538603374	0.660383110	0.11
9	9489786	0.9831409794	0.832415968	0.733960485	1.72819205	0.0239464881	99.538603374	0.660383110	0.11
10	375973	0.9831409794	0.832415968	0.733960485	1.72819205	0.0239464881	99.538603374	0.660383110	0.11
11	375973	0.9831409794	0.832415968	0.733960485	1.72819205	0.0239464881	99.538603374	0.660383110	0.11
12	375973	0.9831409794	0.832415968	0.733960485	1.72819205	0.0239464881	99.538603374	0.660383110	0.11
13	490501	0.9831409794	0.832415968	0.733960485	1.1364096708	0.2655448853	1.285976593	-1.132101605	0.660383110
14	9489786	0.9831409794	0.832415968	0.733960485	1.72819205	0.0239464881	99.538603374	0.660383110	0.11
15	375973	0.9831409794	0.832415968	0.733960485	1.72819205	0.0239464881	99.538603374	0.660383110	0.11
16	375973	0.9831409794	0.832415968	0.733960485	1.72819205	0.0239464881	99.538603374	0.660383110	0.11
17	490501	0.9831409794	0.832415968	0.733960485	1.72819205	0.0239464881	99.538603374	0.660383110	0.11
18	375973	0.9831409794	0.832415968	0.733960485	1.72819205	0.0239464881	99.538603374	0.660383110	0.11
19	9489786	0.9831409794	0.832415968	0.733960485	1.72819205	0.0239464881	99.538603374	0.660383110	0.11
20	375973	0.9831409794	0.832415968	0.733960485	1.72819205	0.0239464881	99.538603374	0.660383110	0.11

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:

