



# 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' -  $GPA \geq 3.5$  (0)
- 'B' -  $3.0 \leq GPA < 3.5$  (1)
- 'C' -  $2.5 \leq GPA < 3.0$  (2)
- 'D' -  $2.0 \leq GPA < 2.5$  (3)
- 'F' -  $GPA < 2.0$  (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.

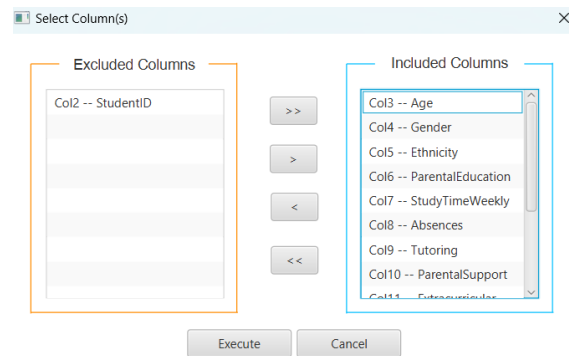
User Header	Col1	Col2	Col3	Col4	Col5	Col6
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						

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	User Row ID	StudentID	Age	Gender	Ethnicity	ParentalEducation	StudyTimeWeekly	Absences	Tutoring	ParentalSupport	Extracurricular	Sports	Music
2	1001	17	1	0	2	19.833722807854713	7	1	2	0	0	1	
3	1002	18	0	0	1	15.40875605584674	0	0	1	0	0	0	
4	1003	15	0	2	3	4.21056976881226	26	0	2	0	0	0	
5	1004	17	1	0	3	10.028829473958215	14	0	3	1	0	0	
6	1005	17	1	0	2	4.6724952729713205	17	1	3	0	0	0	
7	1006	18	0	0	1	8.191218545250186	0	0	1	1	0	0	
8	1007	15	0	1	1	15.60168047469295	10	0	3	0	1	0	
9	1008	15	1	1	4	15.424496305808074	22	1	1	1	0	0	
10	1009	17	0	0	0	4.562007558047705	1	0	2	0	1	0	
11	1010	16	1	0	1	18.44466363097202	0	0	3	1	0	0	
12	1011	17	0	0	1	11.851363655296536	11	0	1	0	0	0	
13	1012	17	0	0	1	7.59848581924029	15	0	2	0	0	0	
14	1013	17	0	1	1	10.038711615617213	21	0	3	1	0	0	
15	1014	17	0	1	2	12.101425068754875	21	0	4	0	1	0	
16	1015	18	1	0	1	11.197810636915708	9	1	2	0	0	0	
17	1016	15	0	0	2	9.728100710723563	17	1	0	0	1	0	

## Step 2: Manipulate data

In order to use the data for training we have to exclude any columns that do not contain features, 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”.

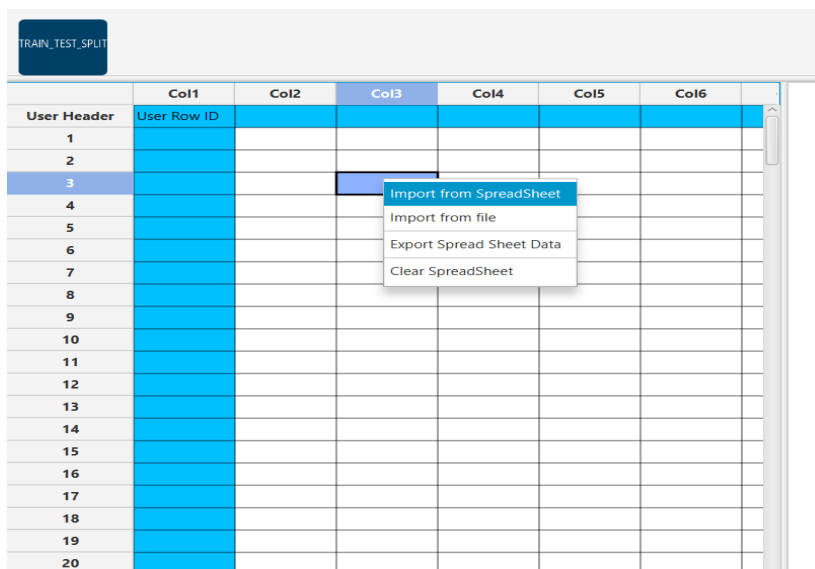


The data without the "StudentID" column will appear in the output spreadsheet.

### 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:

Random Partitioning
✕

Training set percentage

Usage of random generator seed

Stratified sampling

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)
1		17	1	0	2	19.833722807 854713	7	1	2
2		18	0	0	1	15.408756055 84674	0	0	1
3		15	0	2	3	4.2105697688 12.26	26	0	2
4		17	1	0	3	10.028829473 958215	14	0	3
5		17	1	0	2	4.6724952729 713309	17	1	3
6		18	0	0	1	8.1912185452 50186	0	0	1
7		15	0	1	1	15.601680474 699295	10	0	3
8		15	1	1	4	15.424496305 808074	22	1	1
9		17	0	0	0	4.5620075580 47703	1	0	2
10		16	1	0	1	18.444466363 097202	0	0	3
11		17	0	0	1	11.851363655 296536	11	0	1
12		17	0	0	1	7.5984858192 4029	15	0	2
13		17	0	1	1	10.038711615 617213	21	0	3

## 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"

User Header	Col1 (User Row ID)	Col2 (Age)	Col3 (Gender)	Col4 (Ethnicity)	Col5 (Parental Education)	Col6 (StudyTimeWeekly)	Col7 (Absences)	Col8 (Tutoring)	Col9 (ParentalSupport)
1	17	1	0	2	19.833722807	7	1	2	
2	17	1	0	2	854713	4.6724952729	17	1	3
3	18	0	0	1	713305	8.1912185452	0	0	1
4	15	0	1	1	50186	15.601680474	10	0	3
5	15	1	1	4	699295	15.424496305	22	1	1
6	17	0	0	0	808074	4.562007580	1	0	2
7	16	1	0	1	47703	12.444466363	0	0	3
8	17	0	0	1	097202	11.851363655	11	0	1
9	17	0	1	1	296536	10.038711615	21	0	3
10	17	0	1	2	617213	12.101425068	21	0	4
11	18	1	0	1	754875	11.197810636	9	1	2
12	15	0	0	2	915706	9.7281007107	17	1	0
13	18	0	3	1	23563	10.098656081	14	0	2

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	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (D)	Col7 (I)	Col8 (I)	Col9 (I)
1	17	1	0	2	19.833722807854713	7	1	2	
2	17	1	0	2	4.6724952729713305	17	1	3	
3	18	0	0	1	8.191218545250186	0	0	1	
4	15	0	1	1	15.601680474699295	10	0	3	
5	15	1	1	4	15.424496305808074	22	1	1	
6	17	0	0	0	4.562007558047703	1	0	2	
7	16	1	0	1	18.444466363097202	0	0	3	
8	17	0	0	1	11.851363655296536	11	0	1	
9	17	0	1	1	10.038711615915708	21	0	3	
10	17	0	1	2	12.101425066754875	21	0	4	
11	18	1	0	1	11.1978110636915708	9	0	1	
12	15	0	0	2	9.728100710723563	17	1	0	
13	18	0	3	1	10.098656081788907	14	0	2	

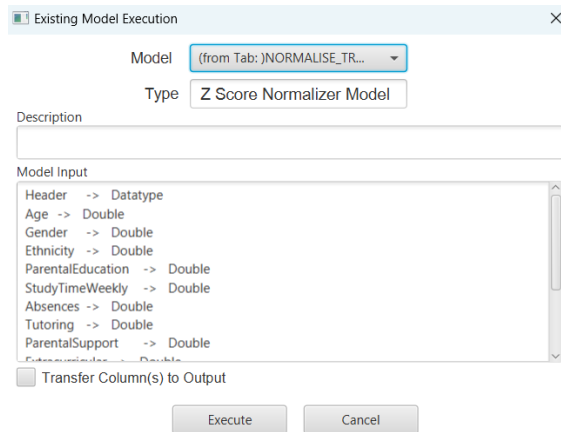
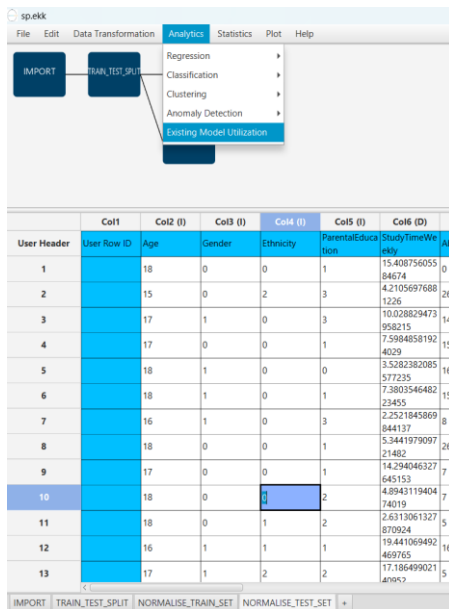
## Step 5: Normalize the test set

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

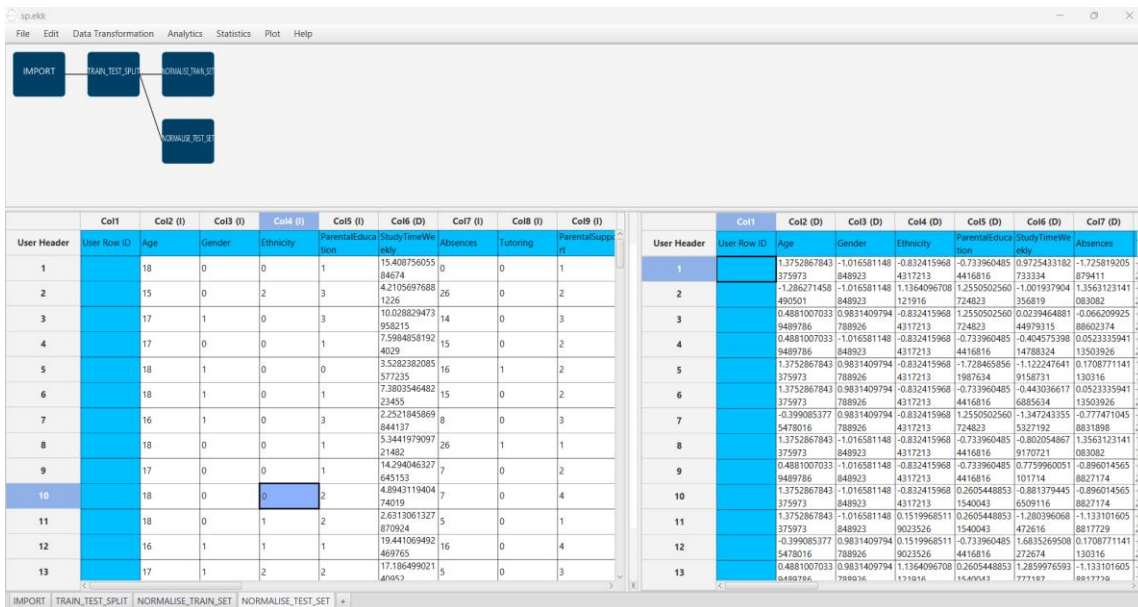
Import data into the input spreadsheet of the “NORMALISE\_TEST\_SET” tab the test set from the output of the “TRAIN\_TEST\_SPLIT” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”. From the available Select input tab options choose “TRAIN\_TEST\_SPLIT: Test Set”.

User Header	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (D)	Col7 (I)	Col8 (I)	Col9 (I)
1	18	0	0	1	15.40875605584674	0	0	1	
2	15	0	2	3	4.21056976881226	26	0	2	
3	17	1	0	3	10.028829473958215	14	0	3	
4	17	0	0	1	7.59848581924029	15	0	2	
5	18	1	0	0	3.5282382085577235	16	1	2	
6	18	1	0	1	7.380354648223455	15	0	2	
7	16	1	0	3	2.2521843869844337	8	0	3	
8	18	0	0	1	5.344197909721482	26	1	1	
9	17	0	0	1	14.294046327645153	7	0	2	
10	18	0	1	2	4.994311940474019	7	0	4	
11	18	0	1	2	2.6313061327870924	5	0	1	
12	16	1	1	1	19.44106492469765	16	0	4	
13	17	1	2	2	17.18649902149897	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 results will appear on the output spreadsheet.



## Step 6: Train the model

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

Import data into the input spreadsheet of the "TRAIN\_MODEL(.fit)" tab from the output of the "NORMALISE\_TRAIN\_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)	Col9 (D)
1	0.4881007033	0.9831409794	-0.832415968	0.2605448853	1.7527601800	-0.896014565	1.5134284490	-0.1183	
2	0.4881007033	0.9831409794	-0.832415968	0.2605448853	-0.920490497	0.2894206341	1.5134284490	0.77384	
3	1.3752867843	-1.016581148	-0.832415968	-0.733960485	-0.300063863	-1.725819205	-0.660383110	-1.0105	
4	375973	848923	4317213	4416816	1951232	879411	2866758	741975	
5	-1.286271458	-1.016581148	0.1519968511	-0.733960485	1.0065600449	-0.540384005	-0.660383110	0.77384	
6	490501	848923	9023526	295642	650291	101976	79006	741975	
7	0.4881007033	-1.016581148	-0.832415968	-1.728465856	-0.939971859	-1.607275685	-0.660383110	-0.1183	
8	0.4881007033	0.9831409794	-0.832415968	4416816	416457	879411	2866758	807647	
9	0.4881007033	-1.016581148	-0.832415968	4416816	785846	88460676	2866758	741975	
10	0.4881007033	-1.016581148	0.85119023526	-0.733960485	0.0256889223	0.7635947141	-0.660383110	0.77384	
11	0.4881007033	-1.016581148	0.1519968511	0.2605448853	0.3893896928	0.7635947141	-0.660383110	1.66604	
12	1.3752867843	0.9831409794	-0.832415968	-0.733960485	0.2300630231	-0.658927525	1.5134284490	-0.1183	
13	1.3752867843	-1.016581148	2.1208224904	-0.733960485	0.0362584217	-0.066209925	-0.660383110	-0.1183	

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 screenshot shows the 'Analytics' menu open, with 'Classification' selected. A sub-menu is visible, listing various machine learning methods: k Nearest Neighbors (kNN), Multiple Layer Perceptron (MLP), Radial Basis Function (RBF), XGBoost, J48, and Random Forest. The 'XGBoost' option is highlighted.

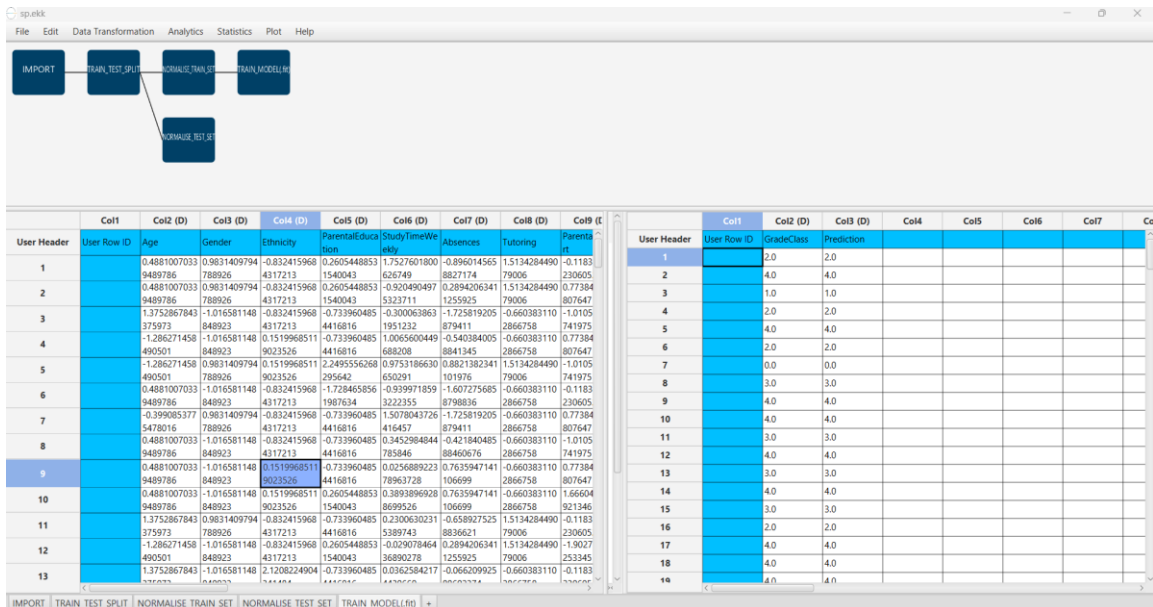
The screenshot shows the 'XGBoost Classification Model' configuration window. Key parameters are set as follows:

- Target Column: Col15 -- GradeClass
- booster: gbtree
- objective: multisoftprob
- number of estimators: 200
- eta: 0.1
- gamma: Double [0, +∞), Default 0
- max depth: 7
- min child weight: Double [0, +∞), Default 1
- column sample by tree: 10
- sub sample: Double [0,1], Default 1
- tree method: default
- lambda: Double (-∞, +∞), Default 1
- alpha: Double (-∞, +∞), Default 0
- Time-based RNG Seed: RNG Seed Double (-∞, +∞), Default:

Buttons for 'Execute' and 'Cancel' are visible at the bottom.

The predictions will appear on the output spreadsheet.

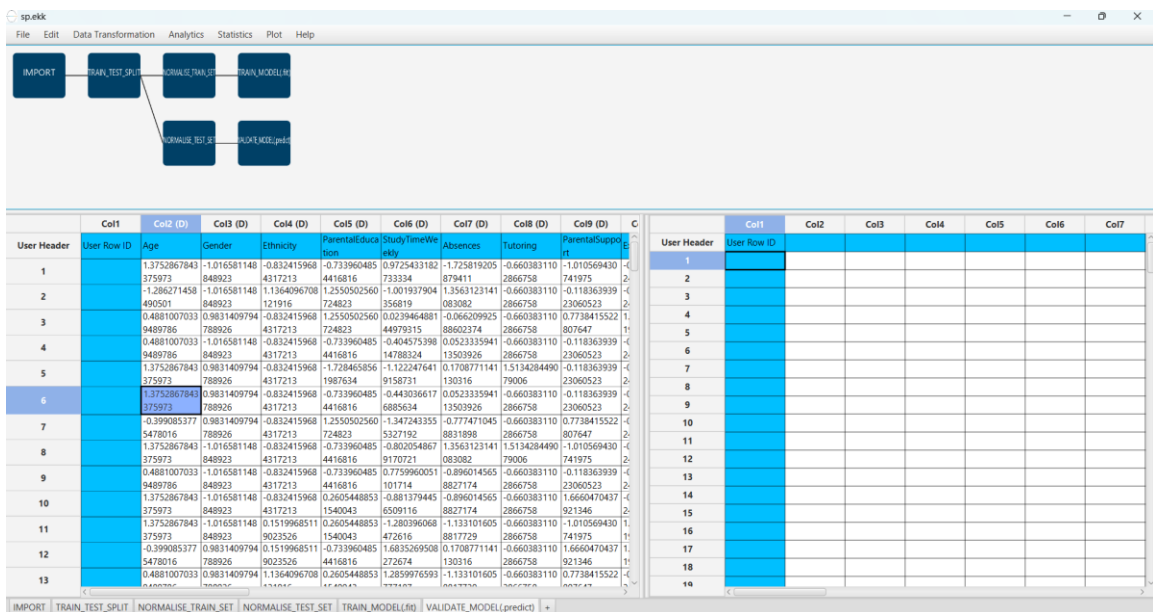




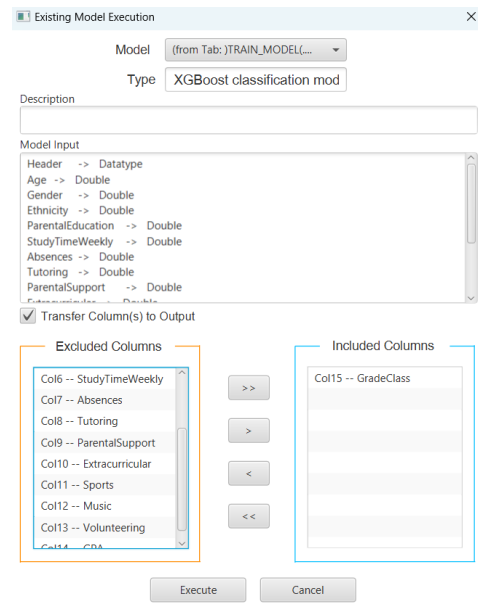
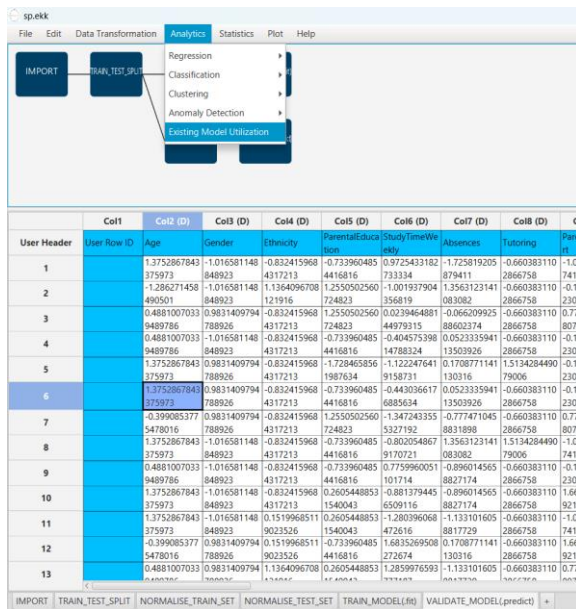
## Step 7: Validate the model

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

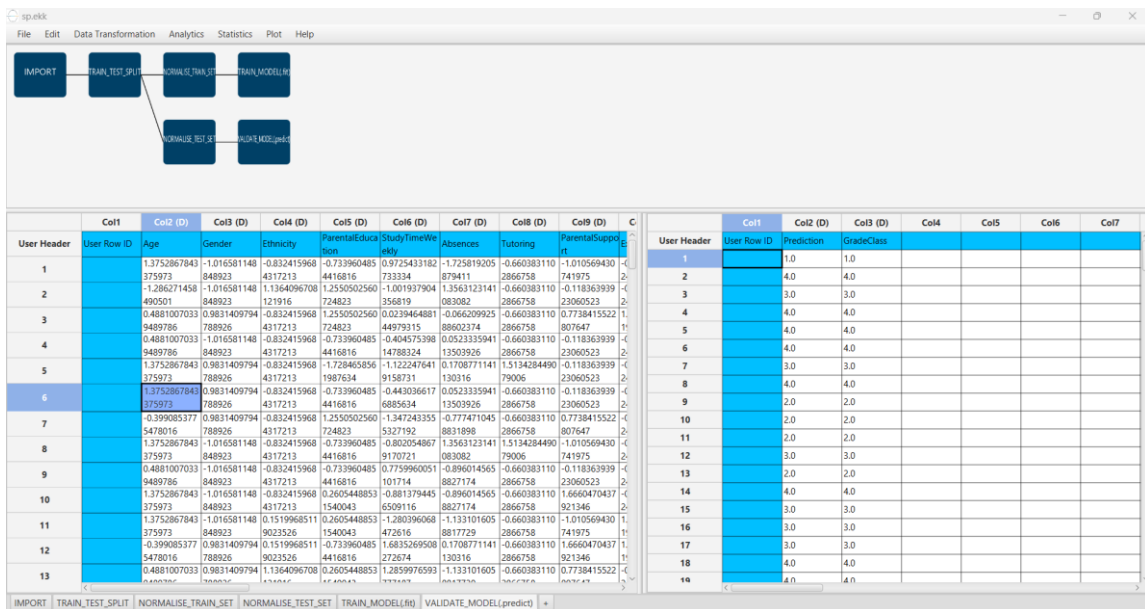
Import data into the input spreadsheet of the "VALIDATE\_MODEL(.predict)" tab from the output of the "NORMALISE\_TEST\_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet".



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



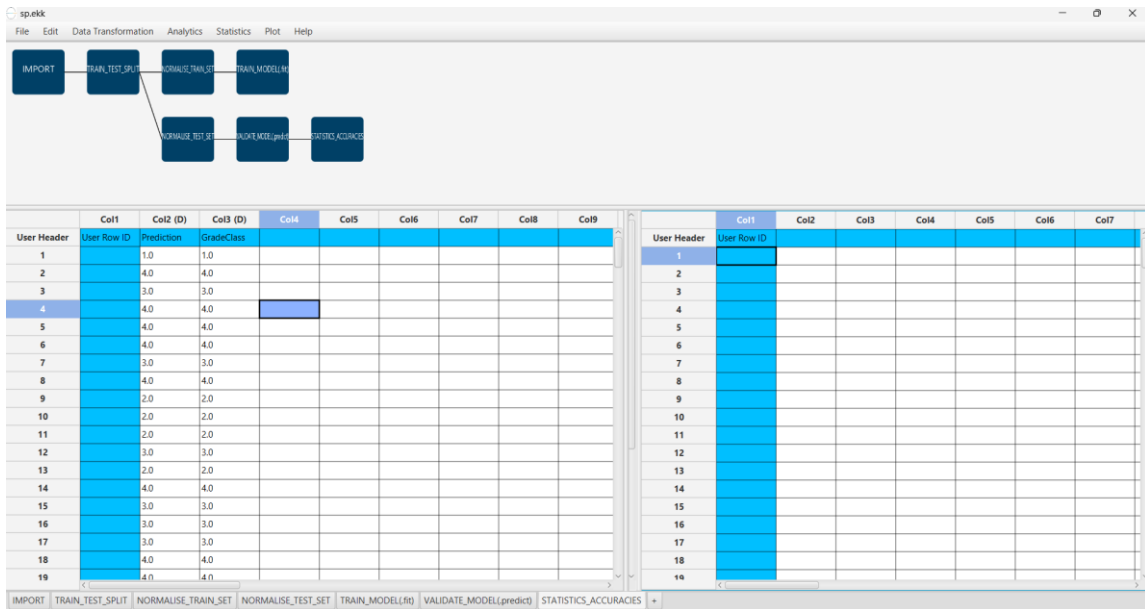
The predictions will appear on the output spreadsheet.



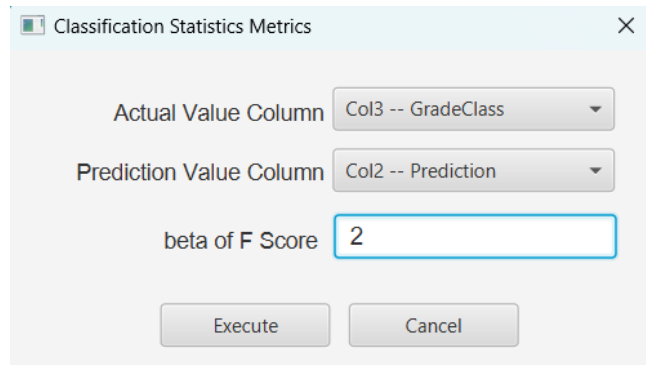
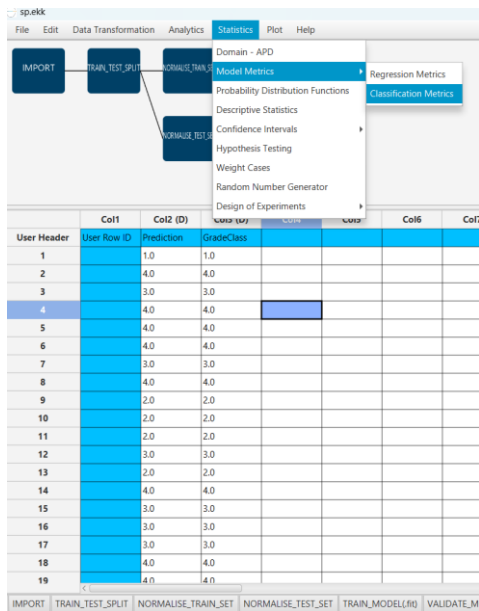
## Step 8: Statistics calculation

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

User Header	Col1 (S)	Col2 (D)	Col3 (S)	Col4 (S)	Col5 (S)	Col6 (S)	Col7 (S)	Col8	Col9
1									
2			Predicted Class	Predicted Class	Predicted Class	Predicted Class	Predicted Class		
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	0.9765886287625418							
11									
12	Precision		1.0	0.9773462783171522	0.970873786407767	0.9603960396039604	1.0		
13									
14	Recall/Sensitivity		0.9701492537313433	0.9966996699669967	0.970873786407767	0.9897959183673469	0.7407407407407407		
15									
16	Specificity		1.0	0.976271186440678	0.9939393939393939	0.992	1.0		
17									
18	F1 Score		0.9848484848484849	0.9869281045751634	0.970873786407767	0.9748743710592964	0.851063829787234		
19									
20	F (beta=2)		0.975975975975976	0.9927679158448389	0.970873786407767	0.9837728194726166	0.78125		
21									
22	MCC	0.9651351831950721							

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

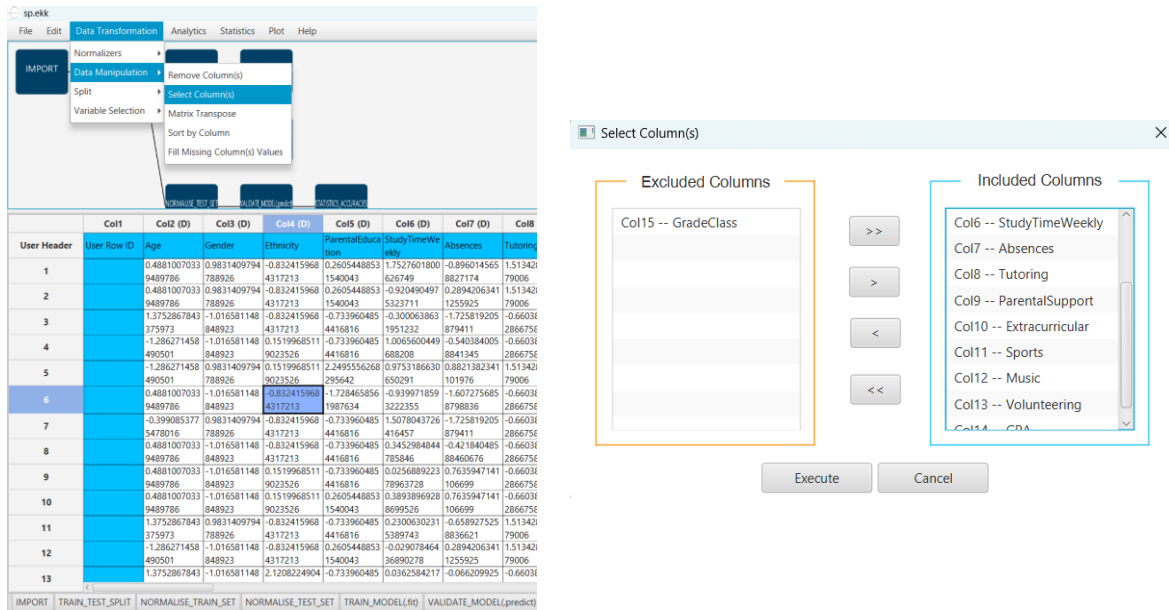
### Step 9.a: Create the domain

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

User Header	Col1	Col2	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8
1	0.4881007033	0.9831409794	-0.832415968	0.2605448853	1.7527601800	-0.899014565	1.5134284490	
2	9489786	788926	4317213	1540043	685749	8827174	79006	
3	1.3752867843	-1.016581148	-0.832415968	-0.733960485	-0.300063863	-1.725819205	-0.660383110	
4	375973	848923	4317213	4416816	1951232	879411	2866758	
5	-1.286271458	-1.016581148	0.1519968511	-0.733960485	1.0065400449	-0.540384005	-0.660383110	
6	490501	848923	9023526	4416816	688208	8841345	2866758	
7	-1.286271458	0.9831409794	0.1519968511	2.2495556268	0.9753186630	0.8821382341	1.5134284490	
8	490501	788926	9023526	295642	650291	101976	79006	
9	0.4881007033	-1.016581148	-0.832415968	1.728465856	-0.899791859	-1.607275685	-0.660383110	
10	9489786	848923	4317213	1987634	3223255	8798836	2866758	
11	-0.399085377	0.9831409794	-0.832415968	-0.733960485	1.5078043726	-1.725819205	-0.660383110	
12	5478016	788926	4317213	4416816	416457	879411	2866758	
13	0.4881007033	-1.016581148	-0.832415968	-0.733960485	0.3452984844	-0.421840485	-0.660383110	
14	9489786	848923	4317213	4416816	785846	88460676	2866758	
15	0.4881007033	-1.016581148	0.1519968511	-0.733960485	0.025689223	0.7635947141	-0.660383110	
16	9489786	848923	9023526	4416816	78963728	106699	2866758	
17	0.4881007033	-1.016581148	0.1519968511	0.2605448853	0.3893896928	0.7635947141	-0.660383110	
18	9489786	848923	9023526	1540043	8699526	106699	2866758	
19	1.3752867843	0.9831409794	-0.832415968	-0.733960485	0.2300630231	-0.658827325	1.5134284490	
20	375973	788926	4317213	4416816	5389743	8836621	79006	
21	-1.286271458	-1.016581148	-0.832415968	0.2605448853	-0.029078464	0.2894206341	1.5134284490	
22	490501	848923	4317213	1540043	36890278	1255925	79006	
23	1.3752867843	-1.016581148	2.1208224904	-0.733960485	0.0362584217	-0.066209925	-0.660383110	

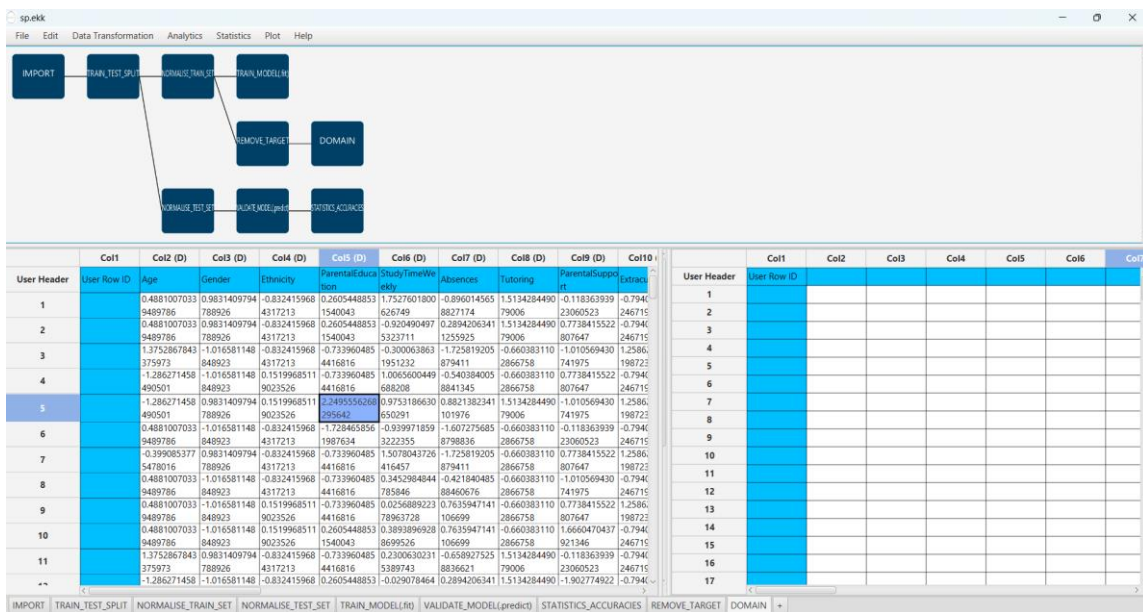
Manipulate the data to exclude the column that corresponds to the "GradeClass" by browsing: "Data Transformation" → "Data Manipulation" → "Select Columns". Then select all the columns except the "GradeClass".



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".



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

The screenshot shows the 'Domain - APD' dialog box with the equation  $APD = d + Z\sigma$  and a Z-value of 0.5. Below the equation is a dropdown menu for 'Perform Computations' set to 'CPU (double precision)'. There are 'Execute' and 'Cancel' buttons at the bottom.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)
1	0.4881007033	0.9831409794	-0.832415968	0.2605448853	1.7527601800	-0.896014565	
2	0.4881007033	0.9831409794	-0.832415968	0.2605448853	0.920490497	0.2894206341	
3	1.3752867843	-1.016581148	-0.832415968	-0.733960485	-0.300063863	1.1725819205	
4	-1.286271458	-1.016581148	0.1519968511	-0.733960485	1.0065600449	-0.540384005	
5	-1.286271458	0.9831409794	0.1519968511	2.2495556269	0.9753186630	0.8821382341	
6	0.4881007033	-1.016581148	-0.832415968	-0.733960485	-0.3452984844	-0.421840485	
7	-0.399085377	0.9831409794	-0.832415968	-0.733960485	1.5078043726	-1.725819205	
8	0.4881007033	-1.016581148	-0.832415968	-0.733960485	0.3452984844	-0.421840485	
9	0.4881007033	-1.016581148	0.1519968511	-0.733960485	0.0256889223	0.7635947141	
10	0.4881007033	-1.016581148	0.1519968511	0.2605448853	0.3893896928	0.7635947141	
11	1.3752867843	0.9831409794	-0.832415968	-0.733960485	0.2300630231	-0.658927525	
12	-1.286271458	-1.016581148	-0.832415968	0.2605448853	-0.029078464	0.2894206341	

The results will appear on the output spreadsheet.

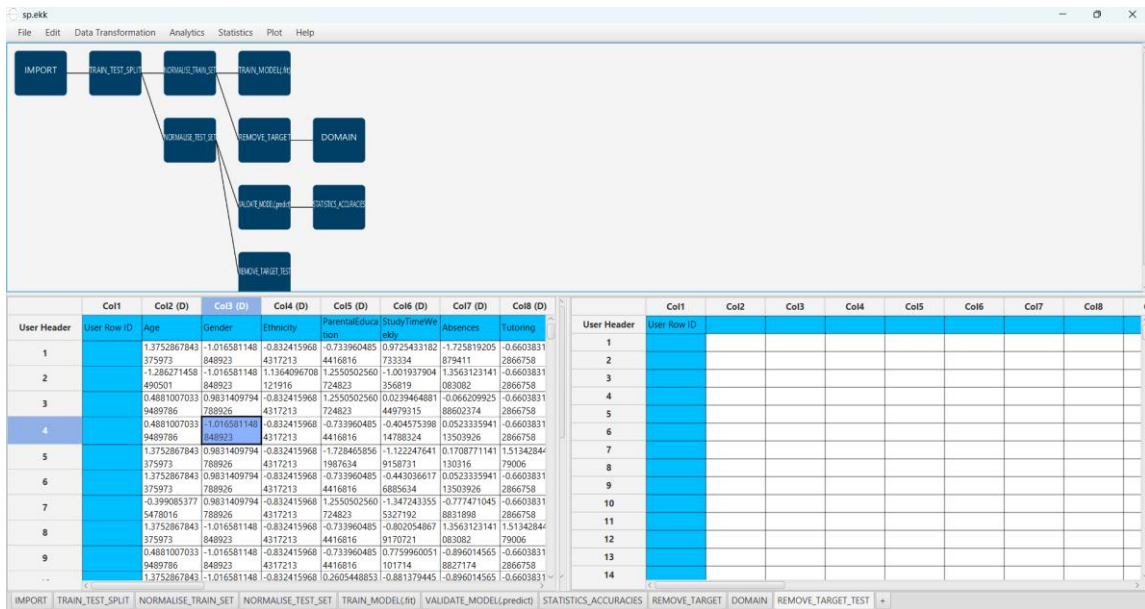
The screenshot shows the 'REMOVE\_TARGET\_TEST' tab in the Isalos Analytics Platform. The table displays results for APD, Domain, and Prediction across 12 rows. The 'APD' column contains values like 0.0, and the 'Domain' column contains values like 4.5313466798. The 'Prediction' column contains values like 'reliable'.

User Header	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)	Col16 (D)	Col17 (D)
1	0.4881007033	0.9831409794	-0.832415968	0.2605448853	1.7527601800	-0.896014565	1.5134284490	-0.118365939	-0.794	4.5313466798	reliable						
2	0.4881007033	0.9831409794	-0.832415968	0.2605448853	0.920490497	0.2894206341	1.5134284490	0.7738415322	-0.794	4.5313466798	reliable						
3	1.3752867843	-1.016581148	-0.832415968	-0.733960485	-0.300063863	1.1725819205	-0.660383110	-1.010569430	1.2506	4.5313466798	reliable						
4	-1.286271458	-1.016581148	0.1519968511	-0.733960485	1.0065600449	-0.540384005	-0.660383110	0.7738415322	-0.794	4.5313466798	reliable						
5	-1.286271458	0.9831409794	0.1519968511	2.2495556269	0.9753186630	0.8821382341	1.5134284490	-1.010569430	1.2506	4.5313466798	reliable						
6	0.4881007033	-1.016581148	-0.832415968	-0.733960485	-0.3452984844	-0.421840485	-0.660383110	-1.010569430	-0.794	4.5313466798	reliable						
7	-0.399085377	0.9831409794	-0.832415968	-0.733960485	1.5078043726	-1.725819205	-0.660383110	0.7738415322	1.2506	4.5313466798	reliable						
8	0.4881007033	-1.016581148	-0.832415968	-0.733960485	0.3452984844	-0.421840485	-0.660383110	-1.010569430	-0.794	4.5313466798	reliable						
9	0.4881007033	-1.016581148	0.1519968511	-0.733960485	0.0256889223	0.7635947141	-0.660383110	0.7738415322	1.2506	4.5313466798	reliable						
10	0.4881007033	-1.016581148	0.1519968511	0.2605448853	0.3893896928	0.7635947141	-0.660383110	1.6660470437	-0.794	4.5313466798	reliable						
11	1.3752867843	0.9831409794	-0.832415968	-0.733960485	0.2300630231	-0.658927525	1.5134284490	-0.118365939	-0.794	4.5313466798	reliable						
12	-1.286271458	-1.016581148	-0.832415968	0.2605448853	-0.029078464	0.2894206341	1.5134284490	-1.902774932	-0.794	4.5313466798	reliable						

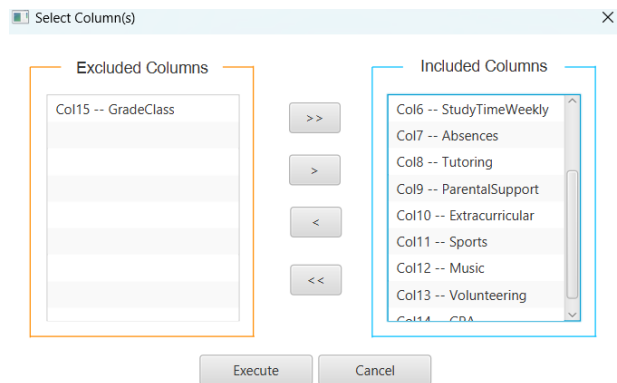
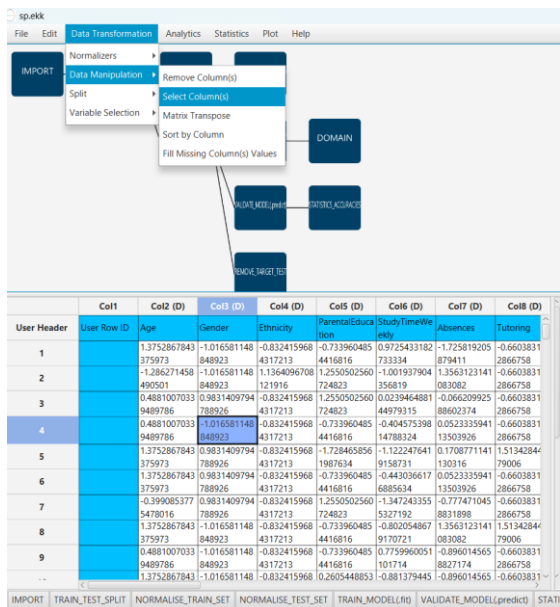
### Step 9.b: Check the test set reliability

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 Columns". 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".

The screenshot shows the 'sp.ekk' software interface. At the top, there is a menu bar with 'File', 'Edit', 'Data Transformation', 'Analytics', 'Statistics', 'Plot', and 'Help'. Below the menu is a workflow diagram with nodes: 'IMPORT', 'TRAIN\_TEST\_SPLIT', 'NORMALISE\_TRAIN\_SET', 'TRAIN\_MODEL(fit)', 'NORMALISE\_TEST\_SET', 'REMOVE\_TARGET', 'DOMAIN', 'VALIDATE\_MODEL(predict)', 'STATISTICS\_ACCURACIES', 'REMOVE\_TARGET\_TEST', 'RELIABILITY'. Below the diagram is a data table with columns: Col1, Col2 (D), Col3 (D), Col4 (D), Col5 (D), Col6 (D), Col7 (D), Col8 (D), Col9 (D). The table contains 8 rows of data. The bottom of the interface shows a row of buttons: 'IMPORT', 'TRAIN\_TEST\_SPLIT', 'NORMALISE\_TRAIN\_SET', 'NORMALISE\_TEST\_SET', 'TRAIN\_MODEL(fit)', 'VALIDATE\_MODEL(predict)', 'STATISTICS\_ACCURACIES', 'REMOVE\_TARGET', 'DOMAIN', 'REMOVE\_TARGET\_TEST', 'RELIABILITY'.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)
1	1.3752867843	-1.016581148	-0.832415968	0.733960485	0.9725433182	-1.725819205	-0.660383110	-1.0105	
2	375973	848923	4317213	4416816	724823	44979315	2866758	807647	
3	490501	848923	121916	724823	1.250502560	-1.001937904	1.3563123141	-0.660383110	-0.1183
4	0.4881007033	0.9831409794	-0.832415968	1.250502560	0.0239464881	892588602374	-0.660383110	0.77384	
5	9489786	788926	4317213	724823	44979315	2866758	807647		
6	0.4881007033	-1.016581148	-0.832415968	-0.733960485	-0.404575398	0.0523335941	-0.660383110	-0.1183	
7	9489786	848923	4317213	4416816	14788324	13503926	2866758	230605	
8	1.3752867843	0.9831409794	-0.832415968	-1.728465856	-1.122247641	0.1708771141	1.5134284490	-0.1183	
9	375973	788926	4317213	1987634	9158731	130316	79006	230605	
10	1.3752867843	0.9831409794	-0.832415968	-0.733960485	-0.443036617	0.0523335941	-0.660383110	-0.1183	
11	375973	848923	4317213	4416816	6885634	13503926	2866758	230605	
12	-0.399085377	0.9831409794	-0.832415968	1.250502560	-1.347243355	-0.77471045	-0.660383110	0.77384	

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

The screenshot shows the 'sp.ekk' software interface with the 'Analytics' menu open. The menu options are: Regression, Classification, Clustering, Anomaly Detection, Existing Model Utilization, and DOMAIN. The 'Existing Model Utilization' option is selected. Below the menu is a data table with columns: Col1, Col2 (D), Col3 (D), Col4 (D), Col5 (D), Col6 (D), Col7 (D), Col8 (D). The table contains 8 rows of data. The bottom of the interface shows a row of buttons: 'IMPORT', 'TRAIN\_TEST\_SPLIT', 'NORMALISE\_TRAIN\_SET', 'NORMALISE\_TEST\_SET', 'TRAIN\_MODEL(fit)', 'VALIDATE\_MODEL(predict)'.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
1	1.3752867843	-1.016581148	-0.832415968	-0.733960485	0.9725433182	-1.725819205	-0.660383110	-1.0105
2	375973	848923	4317213	4416816	724823	44979315	2866758	807647
3	490501	848923	121916	724823	1.250502560	-1.001937904	1.3563123141	-0.660383110
4	0.4881007033	0.9831409794	-0.832415968	1.250502560	0.0239464881	892588602374	-0.660383110	0.77384
5	9489786	788926	4317213	724823	44979315	2866758	807647	
6	0.4881007033	-1.016581148	-0.832415968	-0.733960485	-0.404575398	0.0523335941	-0.660383110	-0.1183
7	9489786	848923	4317213	4416816	14788324	13503926	2866758	230605
8	1.3752867843	0.9831409794	-0.832415968	-1.728465856	-1.122247641	0.1708771141	1.5134284490	-0.1183

The screenshot shows the 'Existing Model Execution' dialog box. It has a title bar with a close button. The 'Model' dropdown is set to '(from Tab:)DOMAIN'. The 'Type' field is 'APD Model'. There is a 'Description' field which is empty. The 'Model Input' section shows a list of variables and their datatypes: Age -> Double, Gender -> Double, Ethnicity -> Double, ParentalEducation -> Double, StudyTimeWeekly -> Double, Absences -> Double, Tutoring -> Double, ParentalSupport -> Double. There is a checkbox for 'Transfer Column(s) to Output' which is unchecked. At the bottom are 'Execute' and 'Cancel' buttons.

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:

