



# Students Performance Dataset

This dataset, which <https://www.kaggle.com/datasets/rabieelkharoua/students-performance-dataset>, contains information on 2392 high school students, detailing their demographics, study habits, parental involvement, extracurricular activities, and academic performance. There are 13 features in total, some of which are categorical and are encoded as:

- 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)
- Extracurricular: No (0), Yes (1)
- Sports: No (0), Yes (1)
- Music: No (0), Yes (1)
- Volunteering: No (0), Yes (1)

The target variable, GradeClass, classifies students' grades into distinct categories based on GPA: 0 ('A',  $\text{GPA} \geq 3.5$ ), 1 ('B',  $3.0 \leq \text{GPA} < 3.5$ ), 2 ('C',  $2.5 \leq \text{GPA} < 3.0$ ), 3 ('D',  $2.0 \leq \text{GPA} < 2.5$ ), 4 ('F',  $\text{GPA} < 2.0$ )

*Isalos version used: 2.0.6*

## Step 1: Import data from file

Right click on the input spreadsheet (left) and choose the option "Import from File". Then navigate through your files to load the one with the student performance data.

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

Import from File  
 Import from Spreadsheet  
 Import from Multiple Spreadsheets  
 Adjust Spreadsheet Precision  
 Export Spreadsheet Data  
 Clear Spreadsheet

The data will appear on the left spreadsheet.

	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (D)	Col7 (I)	Col8 (I)	Col9 (I)	Col10 (I)	Col11 (I)	Col12 (I)	Col13 (I)	Col14 (D)	Col15 (D)
User Header	User Row ID	Age	Gender	Ethnicity	ParentalEducation	StudyTimeWeekly	Absences	Tutoring	ParentalSupport	Extracurricular	Sports	Music	Volunteering	GPA	GradeClass
1	1001	17	1	0	2	19.833722807854713	7	1	2	0	0	1	0	2.929195591667681	2.0
2	1002	18	0	0	1	15.40875605584674	0	0	1	0	0	0	0	3.042914833436377	1.0
3	1003	15	0	2	3	4.21056976881226	26	0	2	0	0	0	0	0.1126022544661815	4.0
4	1004	17	1	0	3	10.028829473958215	14	0	3	1	0	0	0	2.0542181397029484	3.0
5	1005	17	1	0	2	4.6724952729713305	17	1	3	0	0	0	0	1.280611817953875	4.0
6	1006	18	0	0	1	8.191218545250186	0	0	1	1	0	0	0	3.0841836144863937	1.0
7	1007	15	0	1	1	15.601680474699295	10	0	3	0	1	0	0	2.748237414891583	2.0
8	1008	15	1	1	4	15.424496305808074	22	1	1	1	0	0	0	1.360142712316461	4.0
9	1009	17	0	0	0	4.562007558047703	1	0	2	0	1	0	1	2.896819189513569	2.0
10	1010	16	1	0	1	18.444466363097202	0	0	3	1	0	0	0	3.5734742103297656	0.0
11	1011	17	0	0	1	11.851363655296536	11	0	1	0	0	0	0	2.1471716250185144	3.0
12	1012	17	0	0	1	7.59848581924029	15	0	2	0	0	0	1	1.5595945190402773	4.0
13	1013	17	0	1	1	10.038711615617213	21	0	3	1	0	0	0	1.520077814874808	4.0
14	1014	17	0	1	2	12.101425068754875	21	0	4	0	1	0	0	1.7515809583340785	4.0
15	1015	18	1	0	1	11.197810636915708	9	1	2	0	0	0	0	2.396788117124796	3.0

## Step 2: Manipulate data

In our dataset there are not any empty values, so we can select all the columns to be used. On the menu click on *Data Transformation* → *Data Manipulation* → *Select Column(s)* and select all columns.

The screenshot shows the 'Select Column(s)' dialog box in the Isalos Analytics Platform. The 'Data Manipulation' menu is open, and the 'Select Column(s)' option is selected. The dialog box has two main sections: 'Excluded Columns' and 'Included Columns'. The 'Included Columns' list contains the following columns: Col2 -- Age, Col3 -- Gender, Col4 -- Ethnicity, Col5 -- ParentalEducation, Col6 -- StudyTimeWeekly, Col7 -- Absences, Col8 -- Tutoring, Col9 -- ParentalSupport, and Col10 -- Extracurricular. The 'Execute' button is highlighted at the bottom right of the dialog box.

All of the data will appear in the output (right) spreadsheet. This tab can be renamed “IMPORT” by right-clicking on it and choosing the “Rename” option.

The screenshot shows the 'Rename Tab' dialog box. The input field contains the text 'IMPORT'. The 'OK' button is highlighted at the bottom left of the dialog box.


## Step 3: Split data

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

Import data into the input spreadsheet of the “TRAIN\_TEST\_SPLIT” tab from the output of the “IMPORT” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

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

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

☐ Time-based RNG Seed

1573771614410400

☒ Stratified sampling

Col15 -- GradeClass

Execute

Cancel

The results will be two separate spreadsheets, “TRAIN\_TEST\_SPLIT: Training Set” and “TRAIN\_TEST\_SPLIT: Test Set”, which will be available to import into the next tabs.

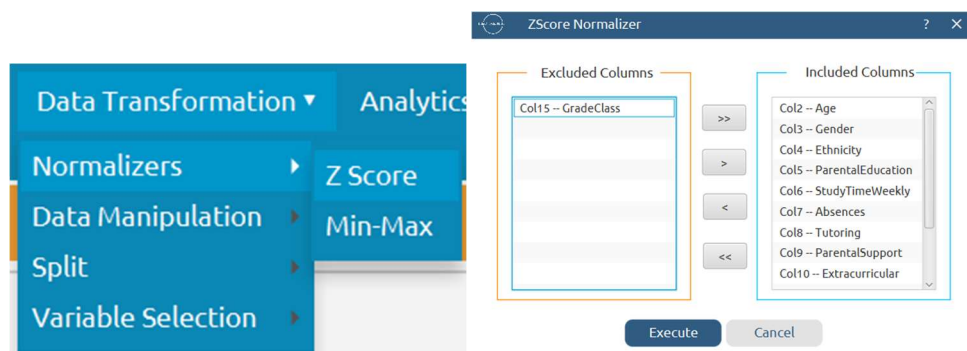
## Step 4: Normalize the training set

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

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

	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (D)	Col7 (I)	Col8 (I)	Col9 (I)	Col10 (I)	Col11 (I)	Col12 (I)	Col13 (I)	Col14 (D)	Col15 (D)
User Header	User Row ID	Age	Gender	Ethnicity	ParentalEducation	StudyTimeWeekly	Absences	Tutoring	ParentalSupport	Extracurricular	Sports	Music	Volunteering	GPA	GradeClass
1	1001	17	1	0	2	19.8337228	7	1	2	0	0	1	0	2.9291956	2.0
2	1002	18	0	0	1	15.4087561	0	0	1	0	0	0	0	3.0429148	1.0
3	1003	15	0	2	3	4.2105698	26	0	2	0	0	0	0	0.1126023	4.0
4	1004	17	1	0	3	10.0288295	14	0	3	1	0	0	0	2.0542181	3.0
5	1005	17	1	0	2	4.6724953	17	1	3	0	0	0	0	1.2880612	4.0
6	1006	18	0	0	1	8.1912185	0	0	1	1	0	0	0	3.0841836	1.0
7	1007	15	0	1	1	15.6016805	10	0	3	0	1	0	0	2.7482374	2.0
8	1008	15	1	1	4	15.4244963	22	1	1	1	0	0	0	1.3601427	4.0
9	1010	16	1	0	1	18.4444664	0	0	3	1	0	0	0	3.5734742	0.0
10	1012	17	0	0	1	7.5984858	15	0	2	0	0	0	1	1.5595945	4.0
11	1016	15	0	0	2	9.7281007	17	1	0	0	1	0	0	1.3415207	4.0
12	1017	18	0	3	1	10.0986561	14	0	2	1	1	0	0	2.2321753	3.0
13	1018	18	1	0	0	3.5282382	16	1	2	0	0	0	0	1.3844042	4.0
14	1020	17	0	0	1	10.8352064	9	0	2	0	0	1	0	2.3957841	3.0
15	1021	16	1	0	3	2.6215972	2	0	3	0	0	0	1	2.7784113	2.0

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



The results will appear on the output spreadsheet.

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)	Col10 (D)	Col11 (D)	Col12 (D)	Col13 (D)	Col14 (D)	Col15 (D)
User Header	User Row ID	Age	Gender	Ethnicity	ParentalEducation	StudyTimeWeekly	Absences	Tutoring	ParentalSupport	Extracurricular	Sports	Music	Volunteering	GPA	GradeClass
1	1001	0.4884409	0.9668266	-0.8470386	0.2688991	1.7714100	-0.8900786	1.5035150	-0.0989572	-0.7987492	-0.6525608	2.0092472	-0.4281064	1.1157883	2.0
2	1002	1.3825869	-1.0337351	-0.8470386	-0.7445569	0.9877228	-1.7177429	-0.6647373	-0.9910637	-0.7987492	-0.6525608	-0.4974214	-0.4281064	1.2400206	1.0
3	1003	-1.2998510	-1.0337351	1.0949259	1.2823552	-0.9955407	1.3564389	-0.6647373	-0.0989572	-0.7987492	-0.6525608	-0.4974214	-0.4281064	-1.9611932	4.0
4	1004	0.4884409	0.9668266	-0.8470386	1.2823552	0.0349067	-0.0624142	-0.6647373	0.7931493	1.2512595	-0.6525608	-0.4974214	-0.4281064	0.1599210	3.0
5	1005	0.4884409	0.9668266	-0.8470386	0.2688991	-0.9137310	0.2922990	1.5035150	0.7931493	-0.7987492	-0.6525608	-0.4974214	-0.4281064	-0.6770656	4.0
6	1006	1.3825869	-1.0337351	-0.8470386	-0.7445569	-0.2905448	-1.7177429	-0.6647373	-0.9910637	1.2512595	-0.6525608	-0.4974214	-0.4281064	1.2851046	1.0
7	1007	-1.2998510	-1.0337351	0.1239437	-0.7445569	1.0218909	-0.5353653	-0.6647373	0.7931493	-0.7987492	1.5315700	-0.4974214	-0.4281064	0.9181009	2.0
8	1008	-1.2998510	0.9668266	0.1239437	2.2958112	0.9905105	0.8834878	1.5035150	-0.9910637	1.2512595	-0.6525608	-0.4974214	-0.4281064	-0.5983202	4.0
9	1010	-0.4057050	0.9668266	-0.8470386	-0.7445569	1.5253647	-1.7177429	-0.6647373	0.7931493	1.2512595	-0.6525608	-0.4974214	-0.4281064	1.8196291	0.0
10	1012	0.4884409	-1.0337351	-0.8470386	-0.7445569	-0.3955212	0.0558235	-0.6647373	-0.0989572	-0.7987492	-0.6525608	-0.4974214	2.3345658	-0.3804295	4.0
11	1016	-1.2998510	-1.0337351	-0.8470386	0.2688991	-0.0183541	0.2922990	1.5035150	-1.8831701	-0.7987492	1.5315700	-0.4974214	-0.4281064	-0.6186638	4.0
12	1017	1.3825869	-1.0337351	2.0659082	-0.7445569	0.0472734	-0.0624142	-0.6647373	-0.0989572	1.2512595	1.5315700	-0.4974214	-0.4281064	0.3543299	3.0
13	1018	1.3825869	0.9668266	-0.8470386	-1.7580129	-1.1163856	0.1740613	1.5035150	-0.0989572	-0.7987492	-0.6525608	-0.4974214	-0.4281064	-0.5718159	4.0
14	1020	0.4884409	-1.0337351	-0.8470386	-0.7445569	0.1777207	-0.6536031	-0.6647373	-0.0989572	-0.7987492	-0.6525608	2.0092472	-0.4281064	0.5330640	3.0
15	1021	-0.4057050	0.9668266	-0.8470386	1.2823552	-1.2769569	-1.4812674	-0.6647373	0.7931493	-0.7987492	-0.6525608	-0.4974214	2.3345658	0.9510643	2.0

## Step 5: Normalize the test set

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

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

	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (D)	Col7 (I)	Col8 (I)	Col9 (I)	Col10 (I)	Col11 (I)	Col12 (I)	Col13 (I)	Col14 (D)	Col15 (D)
User Header	User Row ID	Age	Gender	Ethnicity	ParentalEducation	StudyTimeWeekly	Absences	Tutoring	ParentalSupport	Extracurricular	Sports	Music	Volunteering	GPA	GradeClass
1	1009	17	0	0	0	4.5620076	1	0	2	0	1	0	1	2.8968192	2.0
2	1011	17	0	0	1	11.8513637	11	0	1	0	0	0	0	2.1471716	3.0
3	1013	17	0	1	1	10.0387116	21	0	3	1	0	0	0	1.5200778	4.0
4	1014	17	0	1	2	12.1014251	21	0	4	0	1	0	0	1.7515810	4.0
5	1015	18	1	0	1	11.1978106	9	1	2	0	0	0	0	2.3967881	3.0
6	1019	18	0	1	3	16.2546581	29	0	2	1	0	0	1	0.4695533	4.0
7	1023	16	1	1	0	18.6488796	29	1	1	0	0	0	0	0.3125462	4.0
8	1024	18	1	3	4	18.9461380	20	0	2	1	0	0	0	1.7701319	4.0
9	1029	18	0	0	0	18.6797484	10	0	3	1	0	0	0	2.8548039	2.0
10	1034	16	0	3	3	15.8932578	21	0	0	1	1	0	0	0.9494061	4.0
11	1038	15	0	2	1	10.1822677	21	1	1	0	1	0	0	1.0677636	4.0
12	1040	17	1	2	1	12.2331135	21	1	3	1	1	0	0	1.5268678	4.0
13	1046	18	0	0	2	4.8943119	7	0	4	0	0	1	0	2.5153003	2.0
14	1047	18	0	1	2	2.6313061	5	0	1	1	0	0	0	2.5637654	2.0
15	1048	17	0	0	2	6.0455340	15	0	2	0	0	0	0	1.7036118	4.0

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

Data Transformation ▾

Analytics ▾

Statistics ▾

➡

⬆

↺

Regression

Classification

Clustering

Anomaly Detection

Existing Model Utilization

Existing Model Execution

Model (from Tab:) NORMALIZE...

Type Z Score Normalizer Model

Description

Model In...

Header → Datatype

Age → Double

Gender → Double

Ethnicity → Double

ParentalEducation → Double

StudyTimeWeekly → Double

Absences → Double

Tutoring → Double

ParentalSupport → Double

Extracurricular → Double

Transfer Column(s) to Output

Execute

Cancel

The results will appear on the output spreadsheet.

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)	Col10 (D)	Col11 (D)	Col12 (D)	Col13 (D)	Col14 (D)	Col15 (D)
User Header	User Row ID	Age	Gender	Ethnicity	ParentalEducation	StudyTimeWeekly	Absences	Tutoring	ParentalSupport	Extracurricular	Sports	Music	Volunteering	GPA	GradeClass
1	1009	0.4884409	-1.0337351	-0.8470386	-1.7580129	-0.9332990	-1.5995051	-0.6647373	-0.0989572	-0.7987492	1.5315700	-0.4974214	2.3345658	1.0804187	2.0
2	1011	0.4884409	-1.0337351	-0.8470386	-0.7445569	0.3576881	-0.4171275	-0.6647373	-0.9910637	-0.7987492	-0.6525608	-0.4974214	-0.4281064	0.2614678	3.0
3	1013	0.4884409	-1.0337351	0.1239437	-0.7445569	0.0366569	0.7652501	-0.6647373	0.7931493	1.2512595	-0.6525608	-0.4974214	-0.4281064	-0.4235995	4.0
4	1014	0.4884409	-1.0337351	0.1239437	0.2688991	0.4019754	0.7652501	-0.6647373	1.6852558	-0.7987492	1.5315700	-0.4974214	-0.4281064	-0.1706944	4.0
5	1015	1.3825869	0.9668266	-0.8470386	-0.7445569	0.2419401	-0.6536031	1.5035150	-0.0989572	-0.7987492	-0.6525608	-0.4974214	-0.4281064	0.5341608	3.0
6	1019	1.3825869	-1.0337351	0.1239437	1.2823552	1.1375370	1.7111522	-0.6647373	-0.0989572	1.2512595	-0.6525608	-0.4974214	2.3345658	-1.5712427	4.0
7	1023	-0.4057050	0.9668266	0.1239437	-1.7580129	1.5615674	1.7111522	1.5035150	-0.9910637	-0.7987492	-0.6525608	-0.4974214	-0.4281064	-1.7427648	4.0
8	1024	1.3825869	0.9668266	2.0659082	2.2958112	1.6142136	0.6470123	-0.6647373	-0.0989572	1.2512595	-0.6525608	-0.4974214	-0.4281064	-0.1504284	4.0
9	1029	1.3825869	-1.0337351	-0.8470386	-1.7580129	1.5670345	-0.5353653	-0.6647373	0.7931493	1.2512595	-0.6525608	-0.4974214	-0.4281064	1.0345193	2.0
10	1034	-0.4057050	-1.0337351	2.0659082	1.2823552	1.0735309	0.7652501	-0.6647373	-1.8831701	1.2512595	1.5315700	-0.4974214	-0.4281064	-1.0470286	4.0
11	1038	-1.2998510	-1.0337351	1.0949259	-0.7445569	0.0620815	0.7652501	1.5035150	-0.9910637	-0.7987492	1.5315700	-0.4974214	-0.4281064	-0.9177292	4.0
12	1040	0.4884409	0.9668266	1.0949259	-0.7445569	0.4252982	0.7652501	1.5035150	0.7931493	1.2512595	1.5315700	-0.4974214	-0.4281064	-0.4161817	4.0
13	1046	1.3825869	-1.0337351	-0.8470386	0.2688991	-0.8744460	-0.8900786	-0.6647373	1.6852558	-0.7987492	-0.6525608	2.0092472	-0.4281064	0.6636292	2.0
14	1047	1.3825869	-1.0337351	0.1239437	0.2688991	-1.2752374	-1.1265541	-0.6647373	-0.9910637	1.2512595	-0.6525608	-0.4974214	-0.4281064	0.7165749	2.0
15	1048	0.4884409	-1.0337351	-0.8470386	0.2688991	-0.6705579	0.0558235	-0.6647373	-0.0989572	-0.7987492	-0.6525608	-0.4974214	-0.4281064	-0.2230982	4.0



## 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 “NORMALIZE\_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: *Analytics → Classification → XGBoost*

The image shows two parts of the software interface. On the left is a menu titled 'Analytics' with a dropdown arrow. The 'Classification' option is selected, opening a submenu with the following items: 'k-Nearest Neighbors (kNN)', 'Fully Connected Neural Network', 'Radial Basis Function Network', 'XGBoost', 'J48 Decision Tree', 'Random Forest', 'Statistical fitting', and 'Auto ML'. On the right is the 'XGBoost Classification Model' configuration window. It contains the following settings: Target Column (Col15 - GradeClass), Booster (gbtree), Objective (multisoftprob), Number of Trees (200), Learning Rate (0.1), Gamma (0.0), Max Tree Depth (7), Minimum Child Weight (1.0), Column Sample by tree (1.0), Subsample (1.0), Tree Method (auto), Lambda (1.0), Alpha (0.0), and a checkbox for 'Time-based RNG Seed' which is unchecked. The 'RNG Seed' is set to 1765897528250. At the bottom are 'Execute' and 'Cancel' buttons.

The predictions will appear on the output spreadsheet.

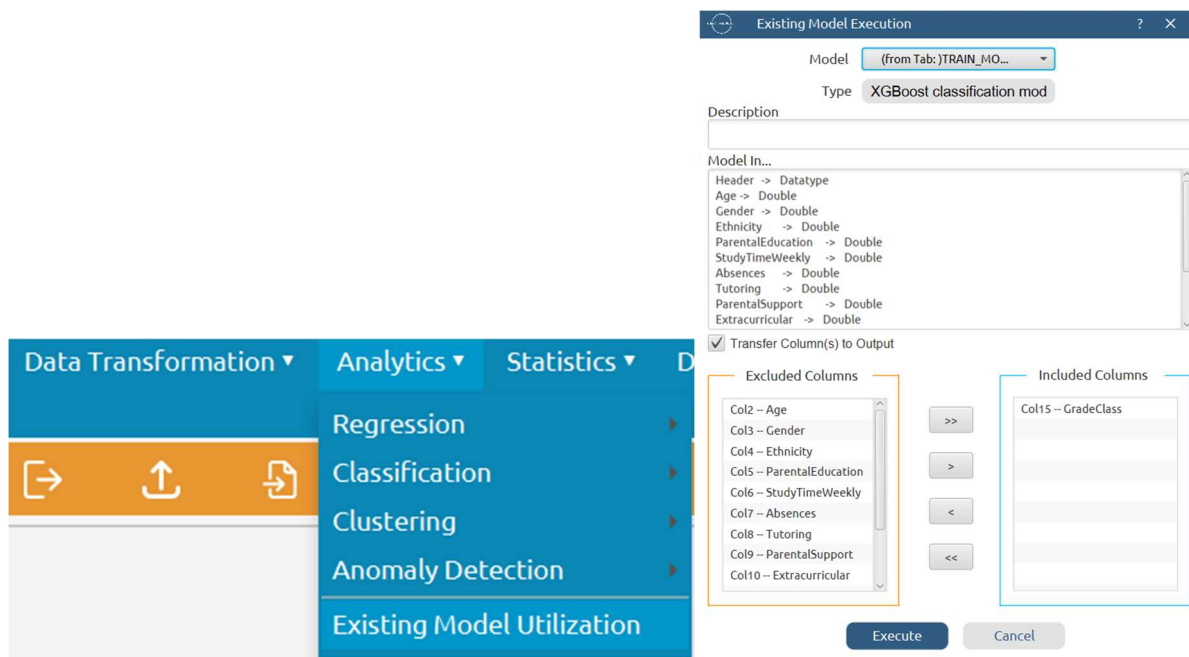
	Col1	Col2 (D)	Col3 (D)
User Header	User Row ID	GradeClass	Prediction
1	1001	2.0	2.0
2	1002	1.0	1.0
3	1003	4.0	4.0
4	1004	3.0	3.0
5	1005	4.0	4.0
6	1006	1.0	1.0
7	1007	2.0	2.0
8	1008	4.0	4.0
9	1010	0.0	0.0
10	1012	4.0	4.0
11	1016	4.0	4.0
12	1017	3.0	3.0
13	1018	4.0	4.0
14	1020	3.0	3.0
15	1021	2.0	2.0

## 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 “NORMALIZE\_TEST\_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

To validate the model: *Analytics → Existing Model Utilization → Model (from Tab:) TRAIN\_MODEL(.fit)*. Choose the column “GradeClass” to be transferred to the output spreadsheet.



The predictions will appear on the output spreadsheet.

	Col1	Col2 (D)	Col3 (D)
User Header	User Row ID	Prediction	GradeClass
1	1009	2.0	2.0
2	1011	3.0	3.0
3	1013	4.0	4.0
4	1014	4.0	4.0
5	1015	3.0	3.0
6	1019	4.0	4.0
7	1023	4.0	4.0
8	1024	4.0	4.0
9	1029	2.0	2.0
10	1034	4.0	4.0
11	1038	4.0	4.0
12	1040	4.0	4.0
13	1046	2.0	2.0
14	1047	2.0	2.0
15	1048	4.0	4.0

## 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: *Statistics → Model Metrics → Classification Metrics*

The screenshot shows the 'Classification Statistics Metrics' dialog box. On the left, a sidebar menu is open, showing the path: Statistics > Model Metrics > Classification Metrics. The main dialog area has the following settings:

- Actual Value Column: Col3 -- GradeClass
- Prediction Value Column: Col2 -- Prediction
- beta of F Score: 2
- Buttons: Execute, Cancel

The results will appear on the output spreadsheet.

	Col1 (S)	Col2 (D)	Col3 (S)	Col4 (S)	Col5 (S)	Col6 (S)	Col7 (S)
User Header	User Row ID						
1			Predicted Class	Predicted Class	Predicted Class	Predicted Class	Predicted Class
2			2.0	3.0	4.0	0.0	1.0
3	Actual Class	2.0	96	2	0	0	0
4	Actual Class	3.0	1	97	5	0	0
5	Actual Class	4.0	0	0	303	0	0
6	Actual Class	0.0	2	0	6	18	1
7	Actual Class	1.0	0	1	6	0	60
8							
9							
10	Classification Accuracy	0.9598662					
11							
12	Precision		0.9696970	0.97	0.946875	1.0	0.9836066
13							
14	Recall/Sensitivity		0.9795918	0.9417476	1.0	0.6666667	0.8955224
15							
16	Specificity		0.994	0.9939394	0.9423729	1.0	0.9981168
17							
18	F1 Score		0.9746193	0.9556650	0.9727127	0.8	0.9375
19							
20	F (beta=2)		0.9775967	0.9472656	0.9889034	0.7142857	0.9118541
21							
22	MCC	0.9402038					



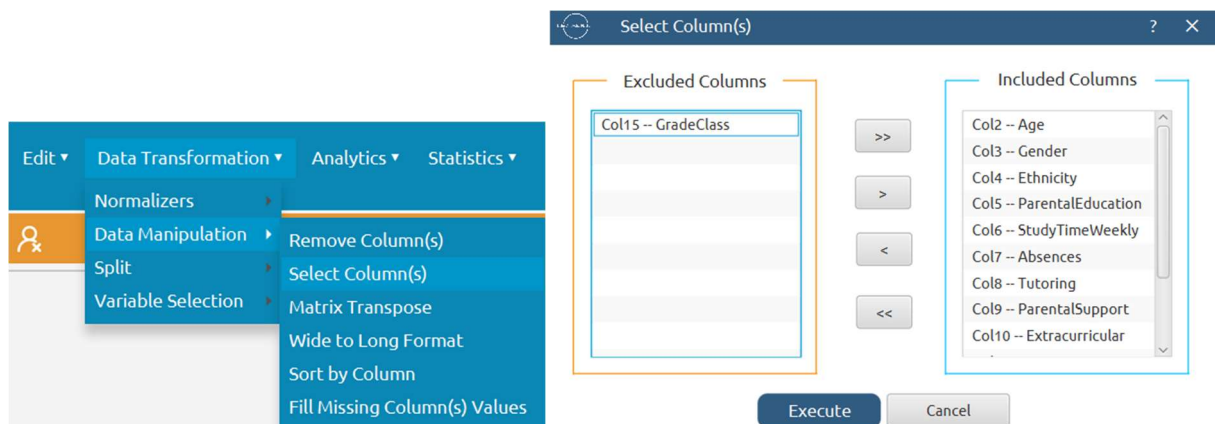
## Step 9: Reliability check for each record of the test set

### Step 9.a: Create the domain

Create a new tab by pressing the “+” button on the bottom of the page with the name “EXCLUDE\_GRADE”.

Import data into the input spreadsheet of the “EXCLUDE\_GRADE” tab from the output of the “NORMALIZE\_TRAIN\_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Manipulate the data to exclude the target column “GradeClass”: *Data Transformation → Data Manipulation → Select Column(s)*

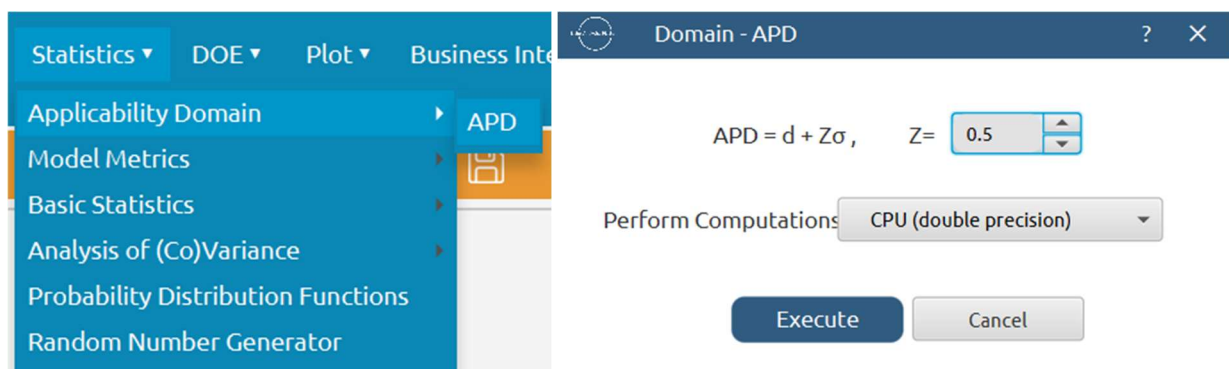


The results will appear on the output spreadsheet.

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\_GRADE” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

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



The results will appear on the output spreadsheet.

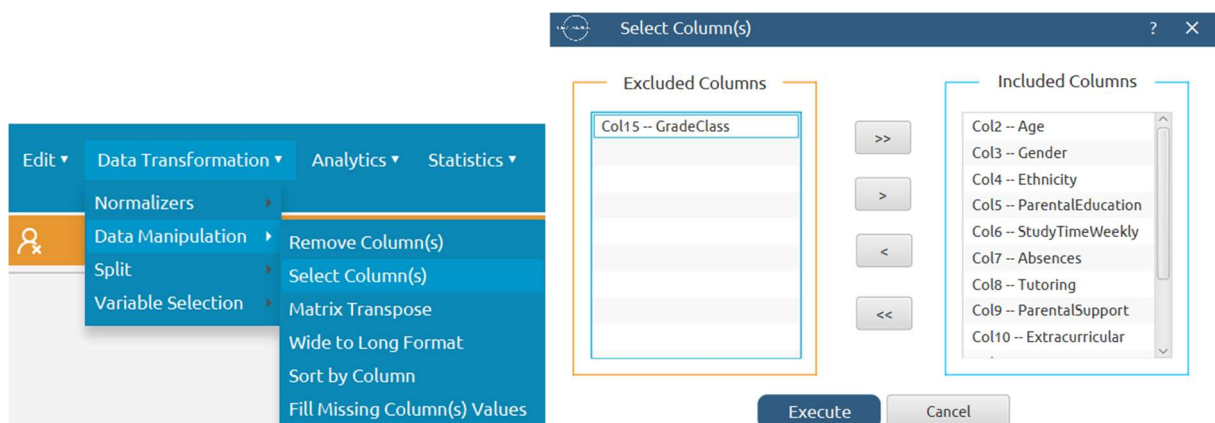
	Col1	Col2 (D)	Col3 (D)	Col4 (S)
User Header	User Row ID	Domain	APD	Prediction
1	1001	0.0	4.5346279	reliable
2	1002	0.0	4.5346279	reliable
3	1003	0.0	4.5346279	reliable
4	1004	0.0	4.5346279	reliable
5	1005	0.0	4.5346279	reliable
6	1006	0.0	4.5346279	reliable
7	1007	0.0	4.5346279	reliable
8	1008	0.0	4.5346279	reliable
9	1010	0.0	4.5346279	reliable
10	1012	0.0	4.5346279	reliable
11	1016	0.0	4.5346279	reliable
12	1017	0.0	4.5346279	reliable
13	1018	0.0	4.5346279	reliable
14	1020	0.0	4.5346279	reliable
15	1021	0.0	4.5346279	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 “EXCLUDE\_GRADE\_TEST\_SET”.

Import data into the input spreadsheet of the “EXCLUDE\_GRADE\_TEST\_SET” tab from the output of the “NORMALIZE\_TEST\_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Manipulate the data to exclude the target column “GradeClass”: *Data Transformation → Data Manipulation → Select Column(s)*

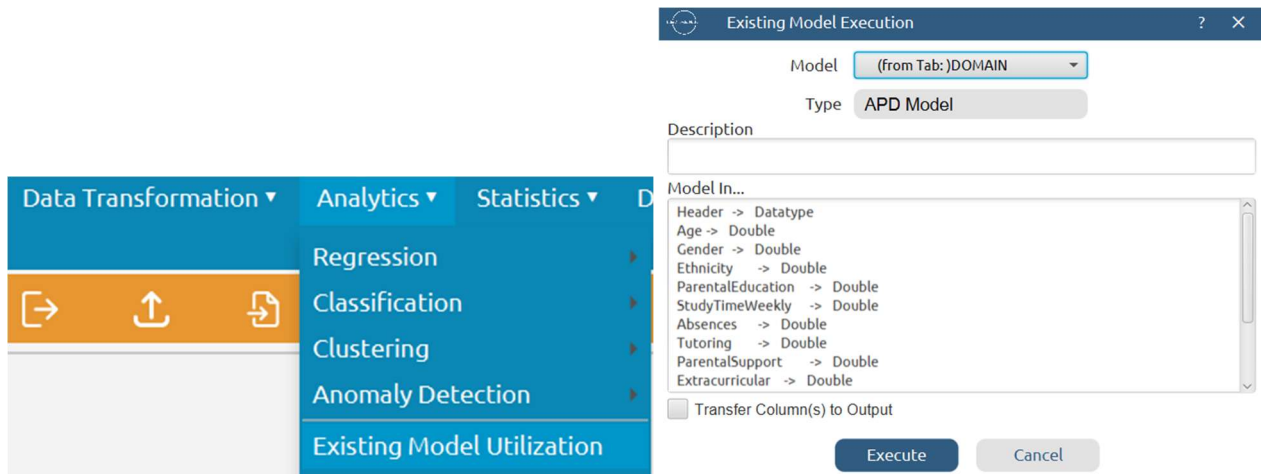


The results will appear on the output spreadsheet.

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\_GRADE\_TEST\_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from Spreadsheet”.

Check the Reliability: *Analytics → Existing Model Utilization → Model (from Tab:) DOMAIN*



The results will appear on the output spreadsheet.

	Col1	Col2 (D)	Col3 (D)	Col4 (S)
User Header	User Row ID	Domain	APD	Prediction
1	1009	2.1101625	4.5346279	reliable
2	1011	1.0392272	4.5346279	reliable
3	1013	1.4837259	4.5346279	reliable
4	1014	1.6691902	4.5346279	reliable
5	1015	1.3690005	4.5346279	reliable
6	1019	1.8302651	4.5346279	reliable
7	1023	2.2178098	4.5346279	reliable
8	1024	1.7138957	4.5346279	reliable
9	1029	1.2254525	4.5346279	reliable
10	1034	2.2981795	4.5346279	reliable
11	1038	1.9729790	4.5346279	reliable
12	1040	1.7055963	4.5346279	reliable
13	1046	1.3620969	4.5346279	reliable
14	1047	1.4001214	4.5346279	reliable
15	1048	1.0061933	4.5346279	reliable

## Final Isalos Workflow

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

