



Salary Prediction (Regression)

The goal of this study is to train a model in order to predict a person's salary . The dataset used in this case study is found in <https://www.kaggle.com/datasets/mrsimple07/salary-prediction-data/data> and has 7 features and 1000 samples. It contains simulated data reflecting various factors influencing salary levels such as education, experience, location, job title, age, and gender.

The dataset contains no missing values and includes several categorical features. Categorical features contain multiple levels and the data was transformed to corresponding numeric codes, as detailed below:

Gender:

- Female (0)
- Male (1)

Education:

- High School (0)
- Bachelor (1)
- Master (2)
- PhD (3)

Location:

- Rural (0)
- Suburban (1)
- Urban (2)

Job_Title:

- Analyst (0)
- Director (1)
- Engineer (2)
- Manager (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 salary 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	Col3	Col4	Col5
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					

Step 2: Manipulate data

In order to use the data for training we have to exclude any columns that do not contain features. In our dataset there are no such columns. Therefore, we will include all columns in the training. We follow these steps to execute this:

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

The screenshot shows the Isalos Analytics Platform interface. In the top navigation bar, 'Data Transformation' is selected under the 'File' tab. A context menu is open over a spreadsheet titled 'IMPORT'. The 'Data Manipulation' option is highlighted, and its submenu includes 'Remove Column(s)', 'Select Column(s)', 'Matrix Transpose', 'Sort by Column', and 'Fill Missing Column(s) Values'. A separate 'Select Column(s)' dialog box is displayed, showing a list of columns from the spreadsheet. The 'Included Columns' section contains 'Col2 -- Education', 'Col3 -- Experience', 'Col4 -- Location', 'Col5 -- Job_Title', 'Col6 -- Age', 'Col7 -- Gender', and 'Col8 -- Salary'. The 'Excluded Columns' section is empty. Buttons for '>>', '>', '<', and '<<' are available between the two lists. At the bottom are 'Execute' and 'Cancel' buttons.

The data 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 Isalos Analytics Platform interface with two tabs visible: 'IMPORT' and 'TRAIN_TEST_SPLIT'. The 'IMPORT' tab is active, displaying a dataset with columns: Col1, Col2 (I), Col3 (I), Col4 (I), Col5 (I), Col6 (I), Col7 (I), Col8 (D), and Col9. The 'TRAIN_TEST_SPLIT' tab is also visible, showing the same dataset. The data consists of 21 rows, each with values for User Row ID, Education, Experience, Location, Job Title, Age, Gender, and Salary.

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 screenshot shows the Isalos Analytics Platform interface. In the top navigation bar, the 'Data Transformation' tab is selected. A context menu is open under the 'Data Manipulation' section, with 'Random Partitioning' highlighted. Below the menu, there is a data spreadsheet with columns labeled Col1 through Col6. A row is selected, and its values are displayed in a tooltip. To the right of the spreadsheet, a 'Random Partitioning' dialog box is open. It contains fields for 'Training set percentage' (set to 75), 'Usage of random generator seed' (with a value of 612642324635300), and 'Stratified sampling' (with a dropdown set to 'Col4 -- Location'). Buttons for 'Execute' and 'Cancel' are at the bottom.

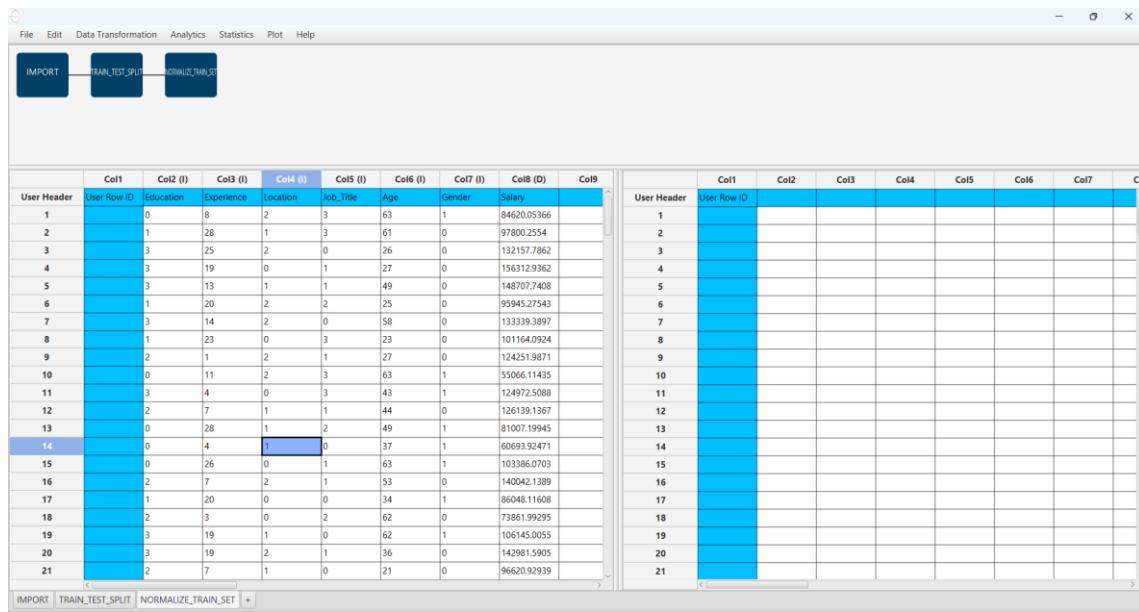
The results will appear on the output spreadsheet.

The screenshot shows the Isalos Analytics Platform interface with the 'TRAIN_TEST_SPLIT' tab selected. On the left, there is a small preview of the data. Two large data spreadsheets are displayed side-by-side. The left spreadsheet has columns Col1 through Col8. The right spreadsheet has columns User Header, Col1 through Col9. Both spreadsheets show the same data rows as the input spreadsheet, with some values highlighted in blue.

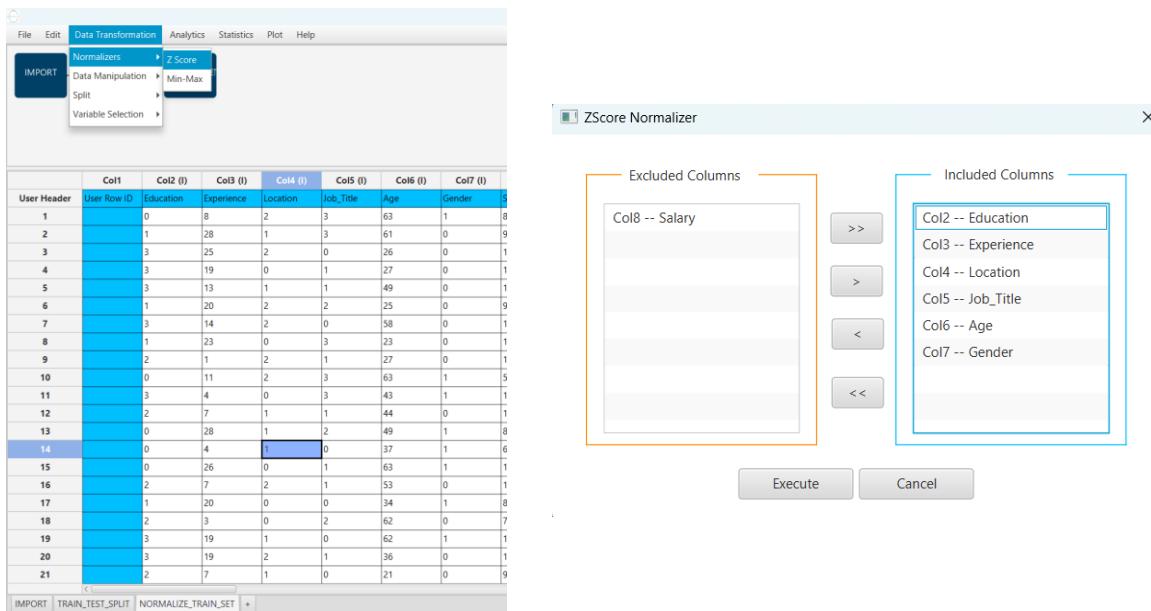
Step 4: Normalize the training set

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

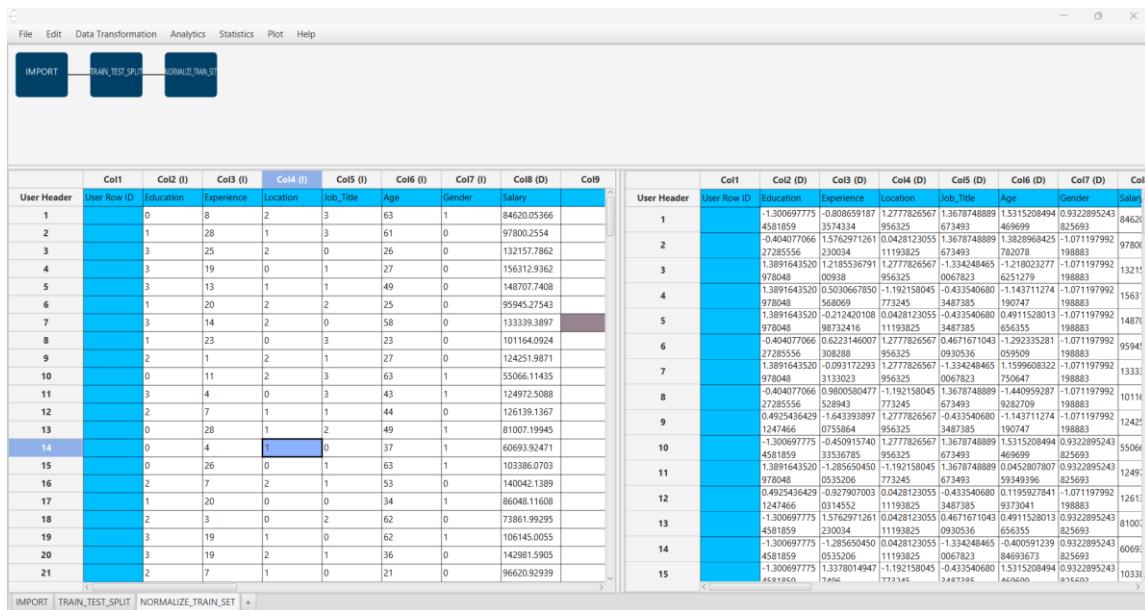
Import data into the input spreadsheet of the "NORMALIZE_TRAIN_SET" tab the train set from the output of the "TRAIN_TEST_SPLIT" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet". From the available Select input tab options choose "TRAIN_TEST_SPLIT : Training Set"



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



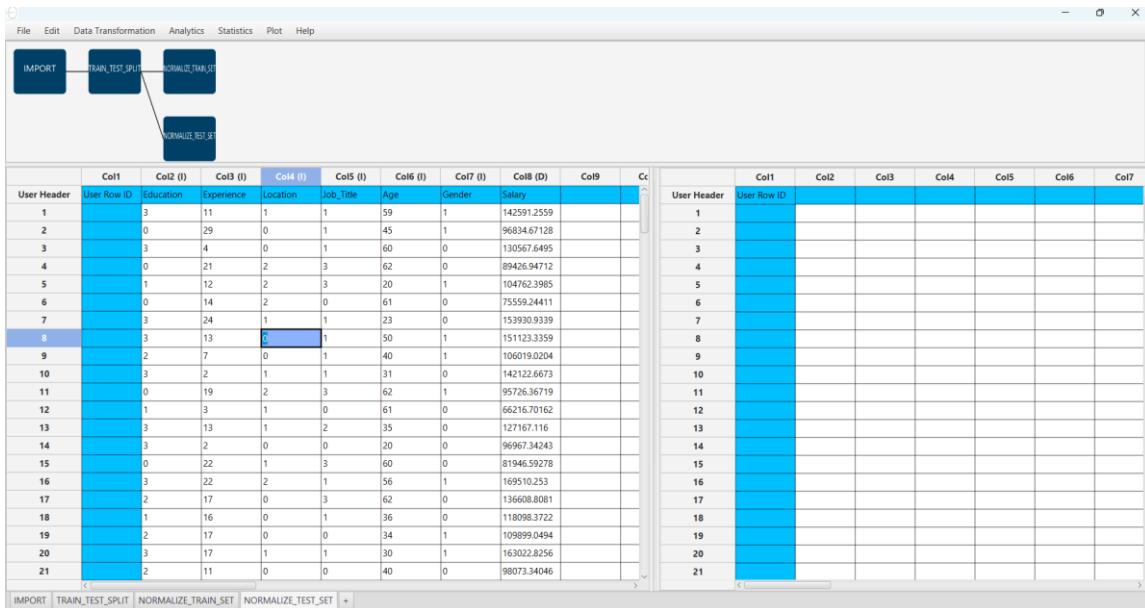
The results will appear on the output spreadsheet.



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 data 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".



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

The results will appear on the output spreadsheet.

Step 6: Feature selection

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

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

The screenshot shows the Isalos Analytics Platform interface. At the top, there is a menu bar with File, Edit, Data Transformation, Analytics, Statistics, Plot, and Help. Below the menu is a toolbar with buttons for IMPORT, TRAIN_TEST_SPLIT, NORMALIZE_TRAIN_SET, NORMALIZE_TEST_SET, and FEATURE_SELECTION_REGRESSION.

The main area contains a data flow diagram and a spreadsheet view. The data flow diagram consists of nodes connected by arrows: IMPORT → TRAIN_TEST_SPLIT → NORMALIZE_TRAIN_SET → FEATURE_SELECTION_REGRESSION. The NORMALIZE_TRAIN_SET node has a branch to NORMALIZE_TEST_SET. The FEATURE_SELECTION_REGRESSION node has a branch to the spreadsheet view.

The spreadsheet view displays two tables. The left table, titled "User Header", has columns Col1 through Col9. The right table, also titled "User Header", has columns Col1 through Col7. Both tables contain data rows numbered 1 through 21. The data includes columns for User Row ID, Education, Experience, Location, Job Title, Age, Gender, and Salary.

Choose the most important features using the Regression Analysis by browsing: "Data Transformation" → "Variable Selection" → "Regression Analysis". Then choose the "Salary" column as the intercept column, the Significance level (α) as 0.05 and include all columns.

The screenshot shows the Isalos Analytics Platform interface. At the top, there is a menu bar with File, Edit, Data Transformation, Analytics, Statistics, Plot, and Help. Below the menu is a toolbar with buttons for IMPORT, TRAIN_TEST_SPLIT, NORMALIZE_TRAIN_SET, NORMALIZE_TEST_SET, and FEATURE_SELECTION_REGRESSION.

The main area shows the Data Transformation menu open, with Variable Selection selected. Under Variable Selection, Regression Analysis is highlighted.

A dialog box titled "Regression Analysis Model" is open. It contains the following settings:

- Significance Level (α): 0.06
- Select Intercept Column: Col8 -- Salary
- Excluded Columns: None
- Included Columns: Col2 -- Education, Col3 -- Experience, Col4 -- Location, Col5 -- Job_Title, Col6 -- Age, Col7 -- Gender

At the bottom of the dialog are Execute and Cancel buttons.

The results will appear on the output spreadsheet.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8
User Header	User Row ID	Education	Experience	Location	Job_Title	Age	Gender	Salary
1		1.300697775	-0.008659187	1.2777826567	1.3678748889	1.5115208494	0.932895243	846.05366
2		4581859	3574334	956325	673493	469699	825693	846.05366
3		-0.040477066	1.5762971261	0.042812055	1.3678748889	1.3828968425	-0.071197992	978.002554
4		27285556	230034	11193825	673493	782078	198883	978.002554
5		1.3891643520	1.2185536791	1.2777826567	-1.334248465	-1.218023277	-0.071197992	132.1577862
6		1.3891643520	0.9303667850	1.92158045	0.4335406080	0.4335406080	0.4335406080	198883
7		79048	3133023	956325	067823	773245	198883	132.1577862
8		1.3891643520	0.9303667850	1.92158045	0.4335406080	0.4335406080	0.4335406080	198883
9		79048	3133023	956325	067823	3487385	190747	198883
10		1.3891643520	-0.121420108	0.042812055	0.4335406080	0.4671671043	-0.071197992	148.0004023
11		4581859	33536785	956325	673493	469699	825693	148.0004023
12		1.300697775	-0.045915740	1.2777826567	1.3678748889	1.5115208494	0.932895243	124.0004023
13		4581859	33536785	956325	673493	3487385	190747	198883
14		1.300697775	-1.2855605450	0.040477066	1.334248465	-0.40591239	0.932895243	606.0004023
15		4581859	7496	11193825	067823	3487385	198883	606.0004023
16		0.4925436429	-0.927907003	0.042812055	0.4335406080	0.4335406080	0.4335406080	198883
17		27285556	306288	773245	673493	031596	198883	140.0004023
18		0.4925436429	0.9303667850	1.92158045	0.4671671043	1.4572088460	-0.071197992	738.0004023
19		79048	3133023	11193825	067823	125887	825693	106.0004023
20		1.3891643520	0.9303667850	1.2777826567	0.4335406080	0.474903243	-0.071197992	141.0004023

IMPORT TRAIN_TEST_SPLIT NORMALIZE_TRAIN_SET NORMALIZE_TEST_SET FEATURE_SELECTION_REGRESSION +

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

- Salary (p-value = 0.0)
- Education (p-value = 1.1188205679825185E-203)
- Experience (p-value = 7.035817529201593E-54)
- Location (p-value = 1.7701240587180506E-13)
- Job_Title (p-value = 1.0075657611706519E-8)

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 "NORMALIZE_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)
User Header	User Row ID	Education	Experience	Location	Job_Title	Age	Gender	Salary
1		1.300697775	-0.008659187	1.2777826567	1.3678748889	1.5115208494	0.932895243	846.05366
2		4581859	3574334	956325	673493	469699	825693	846.05366
3		-0.040477066	1.5762971261	0.042812055	1.3678748889	1.3828968425	-0.071197992	978.002554
4		27285556	230034	11193825	673493	782078	198883	978.002554
5		1.3891643520	-0.9303667850	1.92158045	0.4335406080	0.4671671043	-0.071197992	148.0004023
6		79048	3133023	956325	067823	3487385	190747	198883
7		1.300697775	-0.991172293	1.2777826567	-1.334248465	1.1599608322	-0.071197992	133339.3897
8		4581859	33536785	956325	067823	750467	198883	101164.0924
9		0.4925436429	-1.6433939897	1.2777826567	-0.4335406080	0.4335406080	-1.43711274	12451.9671
10		27285556	306288	773245	673493	3487385	190747	198883
11		1.3891643520	-1.2855605450	1.92158045	0.3678748889	0.0452087087	0.932895243	14972.5088
12		79048	3133023	11193825	067823	125887	825693	106.0004023
13		1.3891643520	0.9303667850	1.2777826567	0.4335406080	0.474903243	-0.071197992	141.0004023

IMPORT TRAIN_TEST_SPLIT NORMALIZE_TRAIN_SET NORMALIZE_TEST_SET FEATURE_SELECTION_REGRESSION +

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

Select Column(s)

Included Columns

Col2 -- Education
Col3 -- Experience
Col4 -- Location
Col5 -- Job_Title
Col8 -- Salary

Excluded Columns

Col6 -- Age
Col7 -- Gender

The results will appear on the output spreadsheet.

User Header	User Row ID	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
1	-1.300697775	-0.450915740	1.277782567	1.3678748889	1.5315208494	0.9322895243	84620.05366		
2	4581859	3574334	0.956325	673493	469699	825693			
3	-0.404077066	1.5762071261	0.0420123055	1.3678748889	1.3828968425	-1.071197992	97800.2554		
4	27285556	230034	1.1193825	673493	782078	198883			
5	1.3891643520	0.5039067850	-1.192158045	-1.334248465	1.218023277	-1.071197992	132157.7862		
6	978048	00938	0.956325	0067823	6251278	198883			
7	1.3891643520	-0.212420108	0.0420123055	-0.433540680	0.4911528013	-1.071197992	148707.7408		
8	978048	568069	773245	3487385	211274190747	198883			
9	1.3891643520	-0.212420108	0.0420123055	-0.433540680	0.4911528013	-1.071197992	101164.0924		
10	978048	031452	11193825	3487385	656355	198883			
11	1.3891643520	0.5039067850	-1.192158045	-1.334248465	1.218023277	-1.071197992	124972.5088		
12	978048	031452	11193825	3487385	9373041	198883			
13	-1.300697775	1.5762071261	0.0420123055	0.4671671043	0.4911528013	0.9322895243	81007.1945		

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".

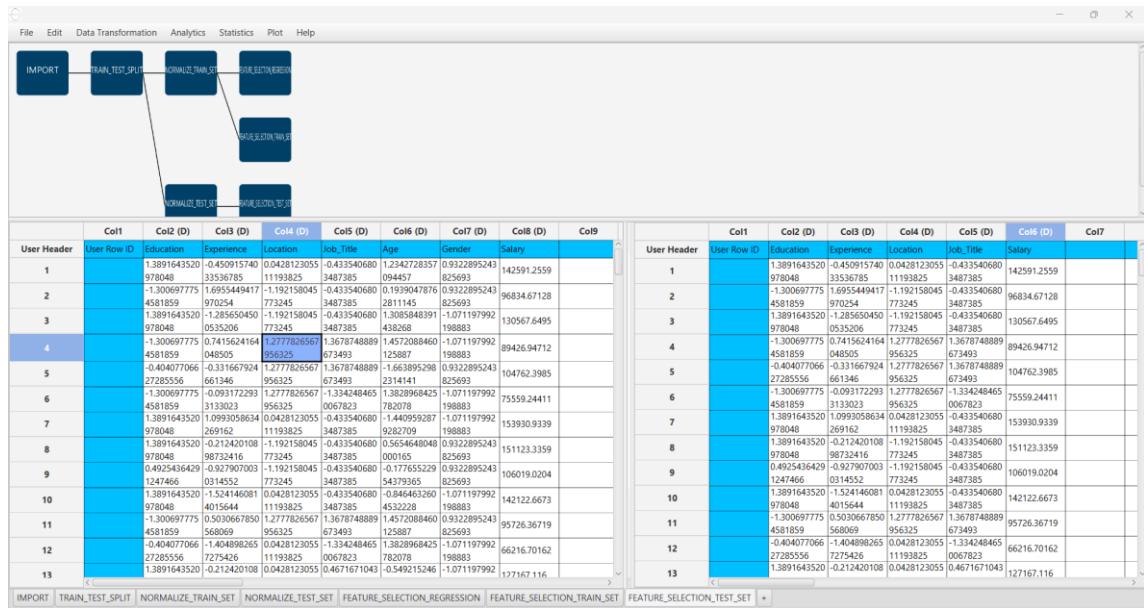
Import data into the input spreadsheet of the "FEATURE_SELECTION_TEST_SET" tab from the output of the "NORMALIZE_TEST_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".

The screenshot shows the Isalos Analytics Platform interface. At the top, there's a menu bar with File, Edit, Data Transformation, Analytics, Statistics, Plot, Help. Below the menu is a toolbar with buttons for IMPORT, TRAIN_TEST_SPLIT, NORMALIZE_TRAIN_SET, and NORMALIZE_TEST_SET. A data flow diagram is visible, showing nodes like 'IMPORT' connected to 'TRAIN_TEST_SPLIT', which then connects to 'NORMALIZE_TRAIN_SET' and 'NORMALIZE_TEST_SET'. These further connect to 'FEATURE_SELECTION_REGRESSION' and 'FEATURE_SELECTION_TRAIN_SET', respectively. Below the flow diagram is a table titled 'User Header' with columns Col1 through Col9. The data in the table consists of 13 rows of user information. To the right of the table is another table titled 'User Header' with columns Col1 through Col7, showing the first 18 rows of the data.

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

The screenshot shows the Isalos Analytics Platform interface with the 'Data Transformation' menu selected. Under 'Data Manipulation', the 'Select Column(s)' option is highlighted. A 'Select Column(s)' dialog box is open, showing two lists: 'Excluded Columns' (Col6 -- Age, Col7 -- Gender) and 'Included Columns' (Col2 -- Education, Col3 -- Experience, Col4 -- Location, Col5 -- Job_Title, Col8 -- Salary). There are buttons for >>, >, <, <<, Execute, and Cancel.

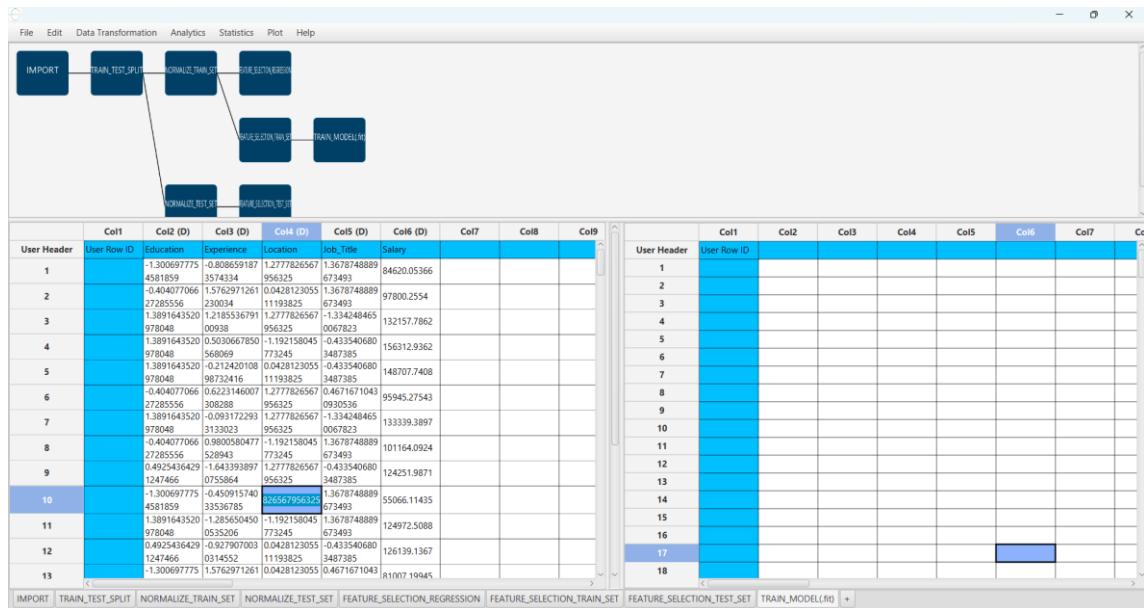
The results will appear on the output spreadsheet.



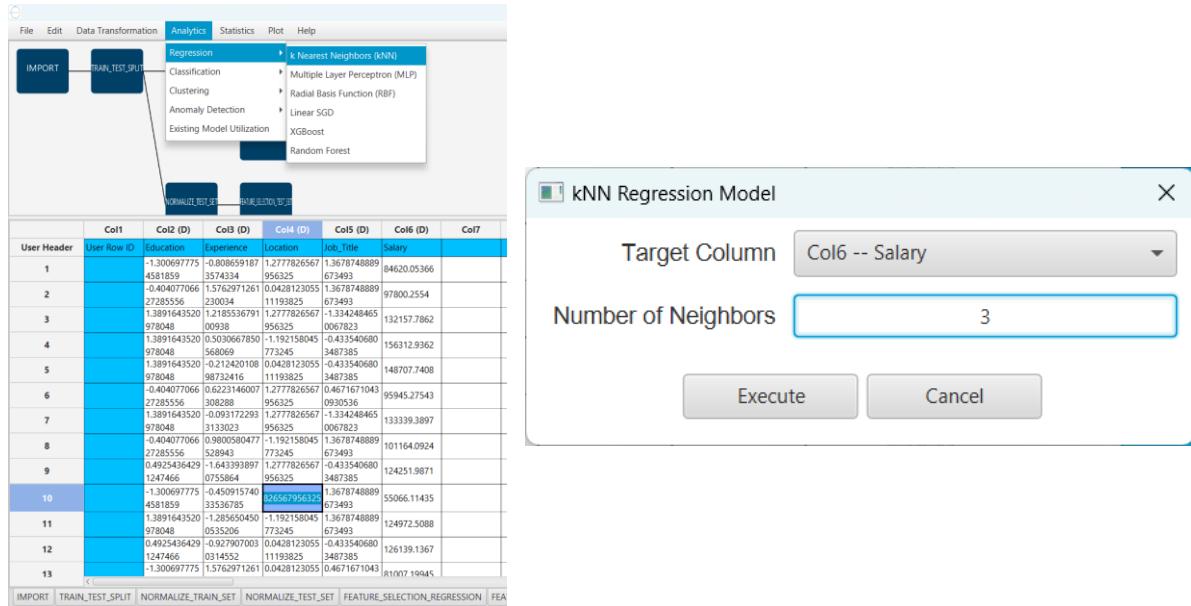
Step 9: Train the model

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

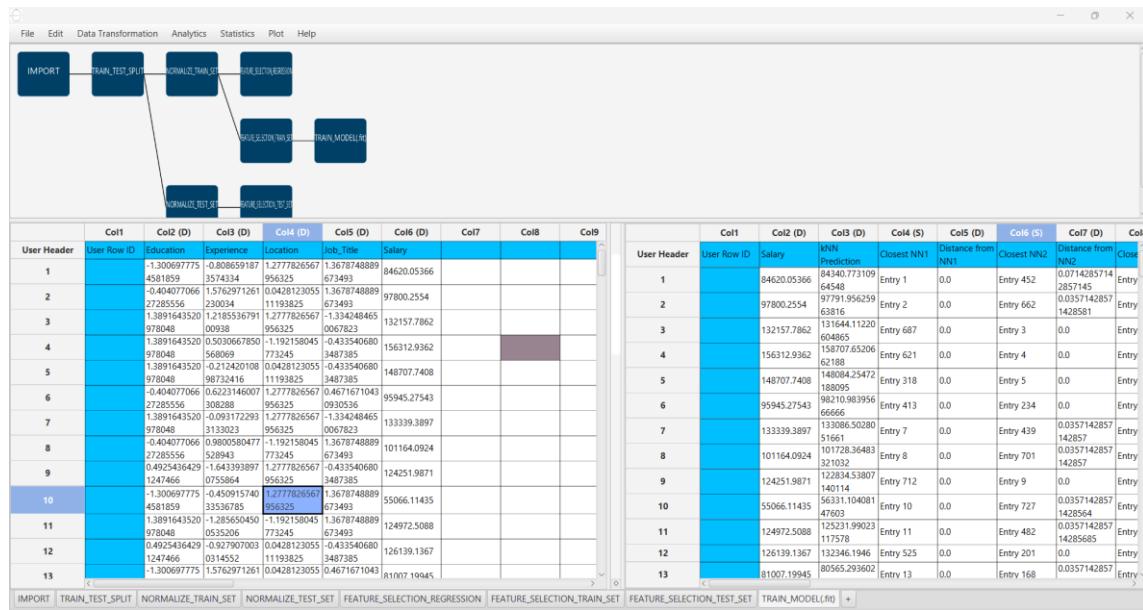
Import data into the input spreadsheet of the "TRAIN_MODEL(.fit)" tab from the output of the "FEATURE_SELECTION_TRAIN_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".



Use the k Nearest Neighbors (kNN) method to train and fit the model by browsing: "Analytics" → "Regression" → "k Nearest Neighbors (kNN)" and set the "Target Column" as the column corresponding to "Salary" and the "Number of Neighbors" to 3.



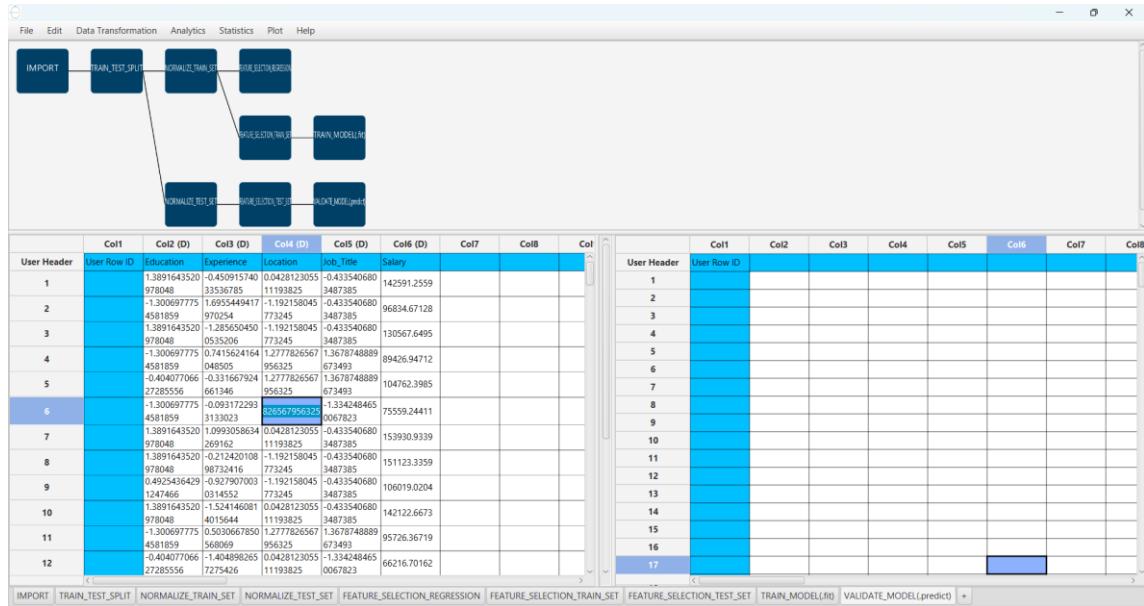
The predictions will appear on the output spreadsheet.



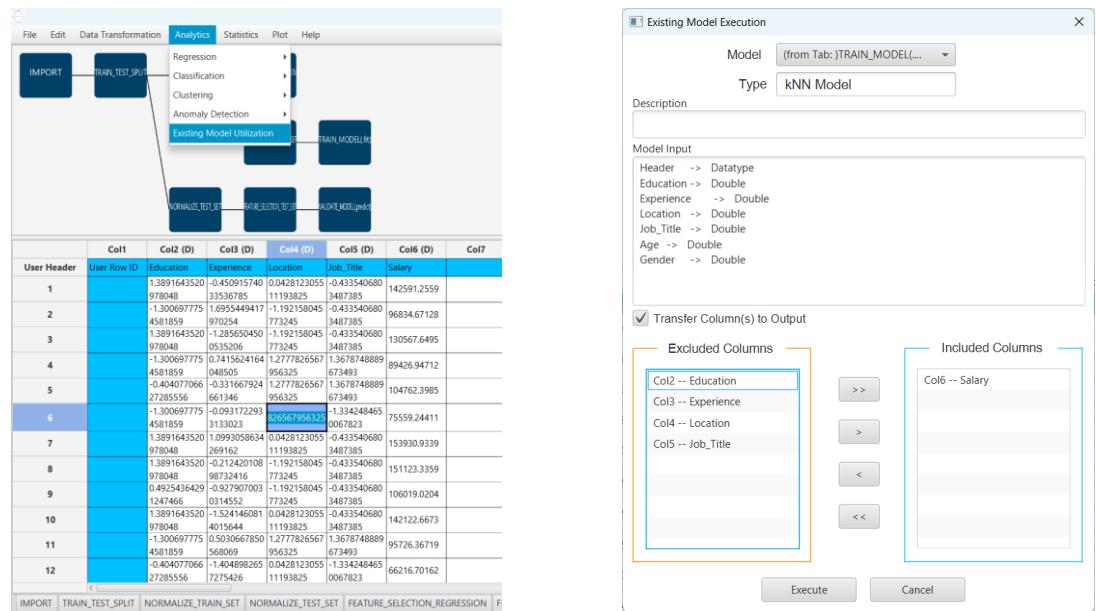
Step 10: Validate the model

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

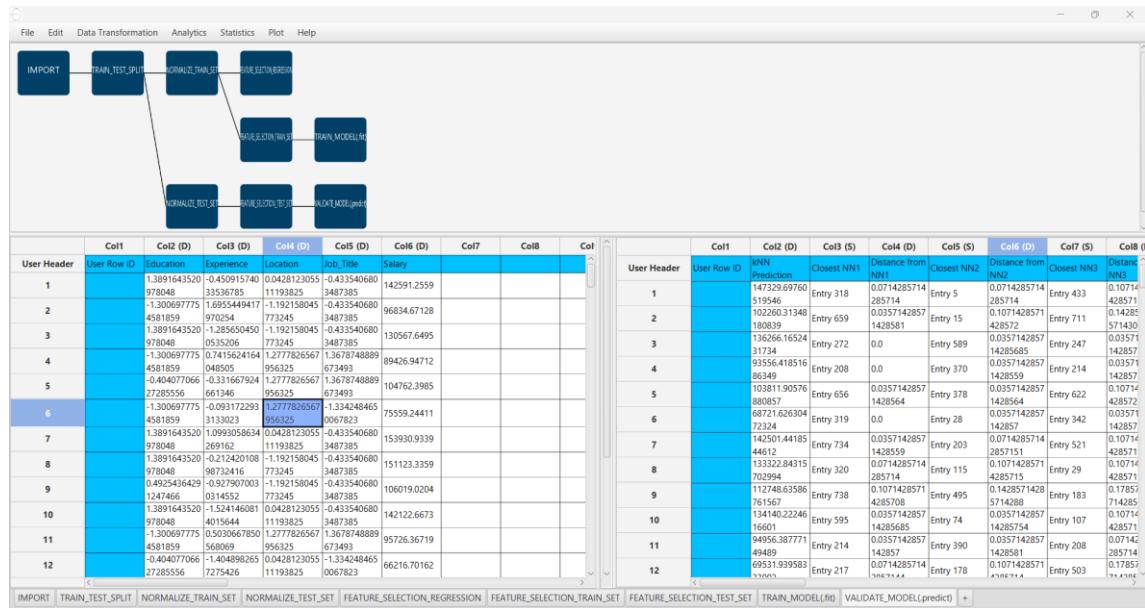
Import data into the input spreadsheet of the "VALIDATE_MODEL(.predict)" tab from the output of the "FEATURE_SELECTION_TEST_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".



To validate the model browse: "Analytics" → "Existing Model Utilization". Then choose Model "(from Tab:) TRAIN_MODEL (.fit)" and transfer the "Salary" column in the output.



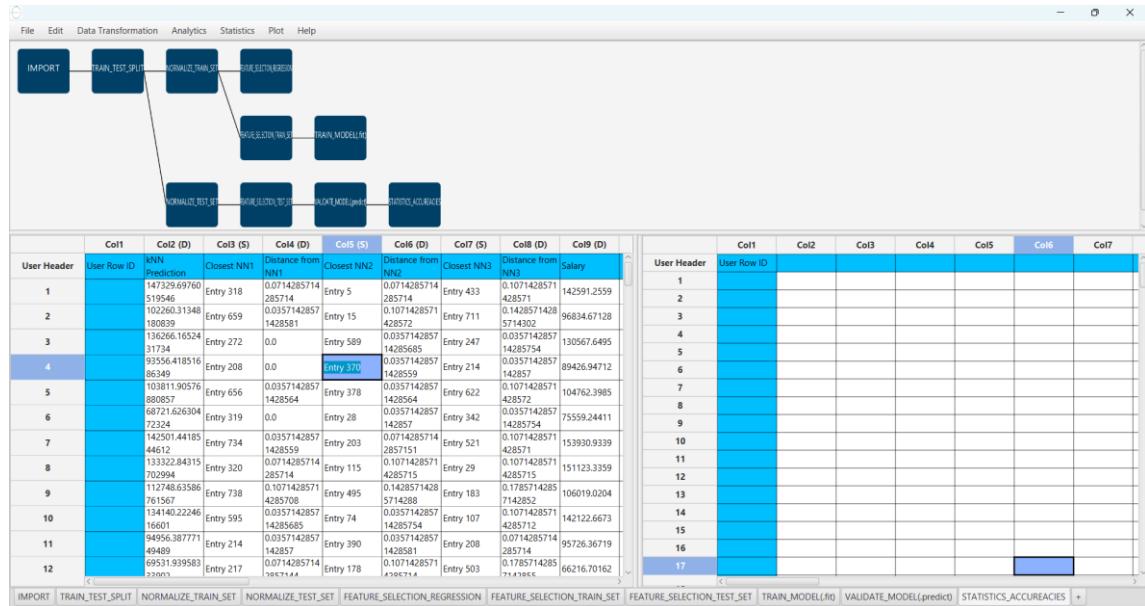
The predictions will appear on the output spreadsheet.



Step 11: Statistics calculation

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

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



Calculate the statistical metrics for the regression by browsing: "Statistics" → "Model Metrics" → "Regression Metrics".

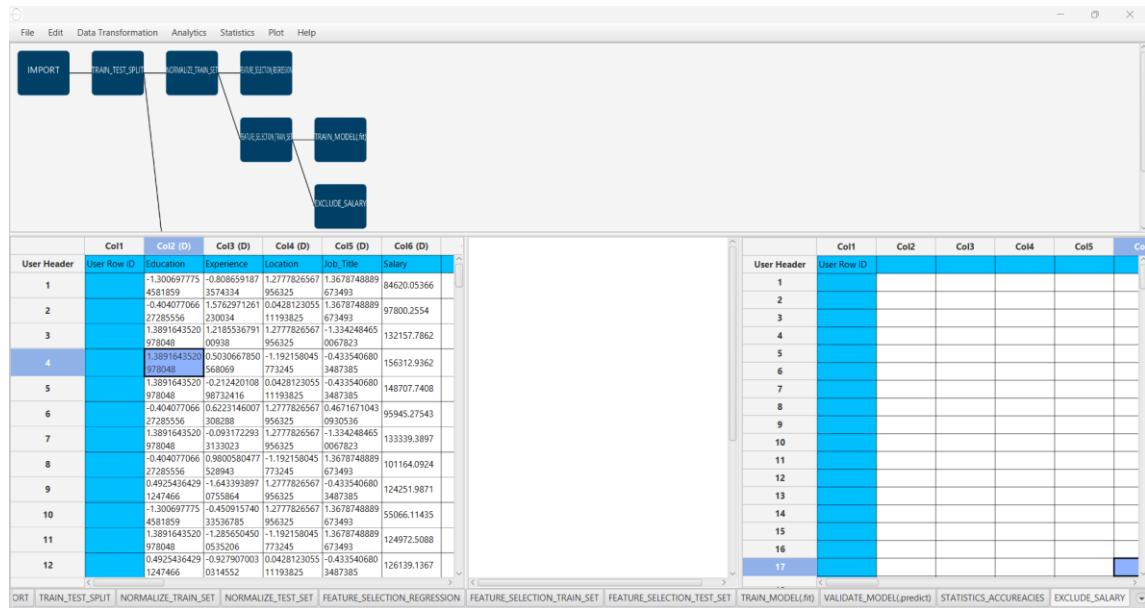
The results will appear on the output spreadsheet.

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_SALARY".

Import data into the input spreadsheet of the "EXCLUDE_SALARY" 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 "Salary" by browsing: "Data Transformation" → "Data Manipulation" → "Select Column(s)". Then select all the columns except the "Salary".

The screenshot shows the Isalos Analytics Platform interface with the Data Transformation menu selected. A context menu is open over a row in the spreadsheet, showing options like Normalizers, Data Manipulation, Split, and Variable Selection. The "Data Manipulation" option is highlighted, and its submenu is visible, showing "Remove Column(s)", "Select Column(s)", "Matrix Transpose", "Sort by Column", and "Fill Missing Column(s) Values".

Select Column(s) Dialog:

The dialog has two main sections: "Excluded Columns" and "Included Columns".

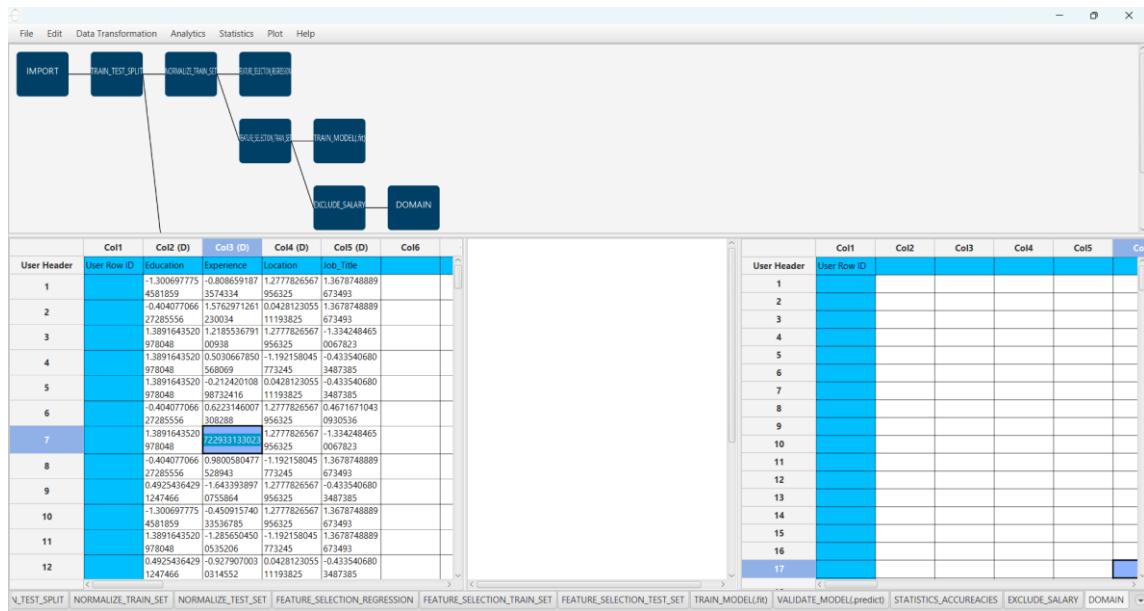
- Excluded Columns:** Contains the column "Col6 -- Salary".
- Included Columns:** Contains the columns "Col2 -- Education", "Col3 -- Experience", "Col4 -- Location", and "Col5 -- Job_Title".

At the bottom of the dialog are "Execute" and "Cancel" buttons.

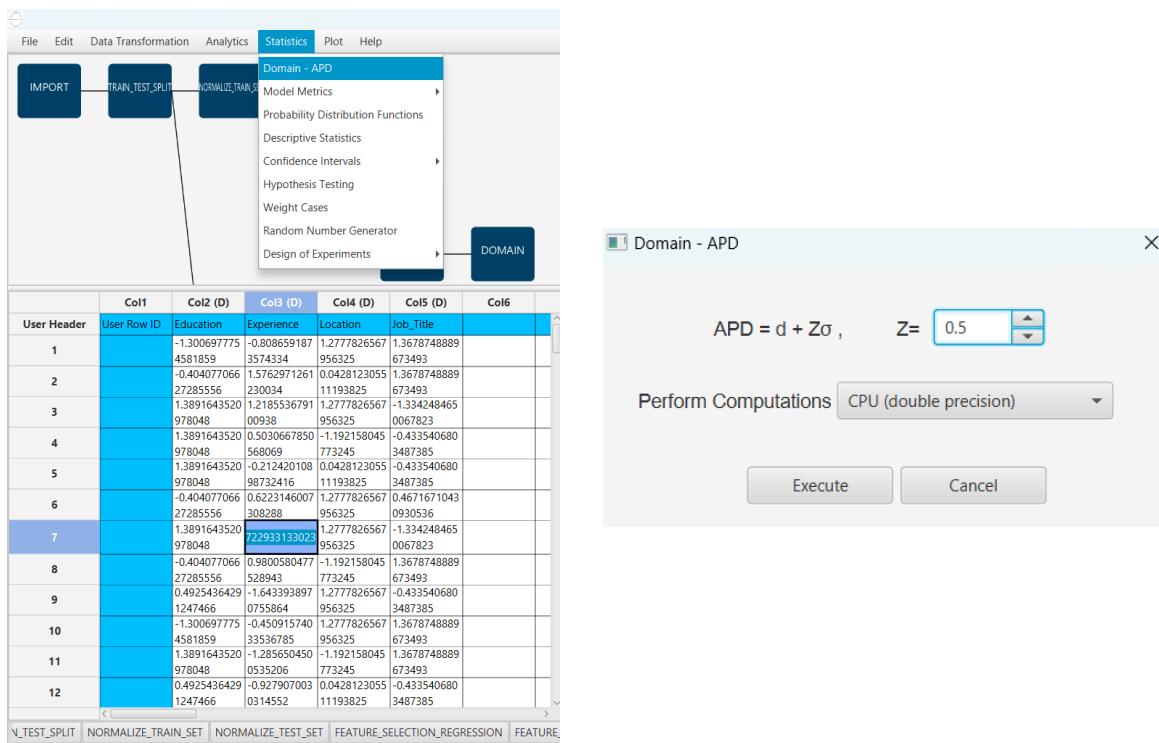
The results will appear on the output spreadsheet.

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

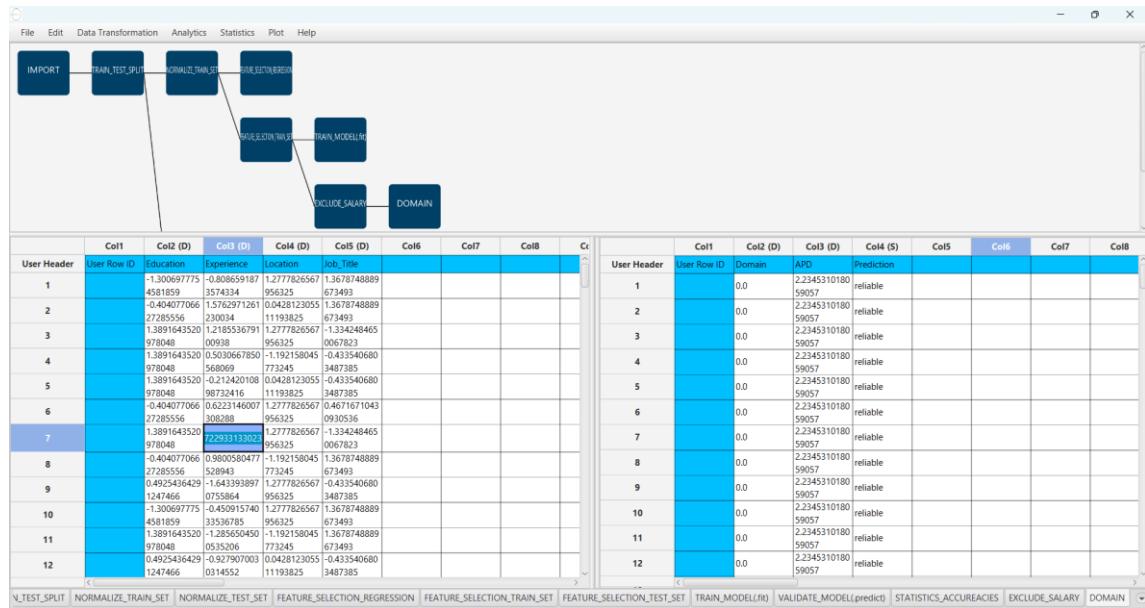
Import data into the input spreadsheet of the "DOMAIN" tab from the output of the "EXCLUDE_SALARY" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".



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



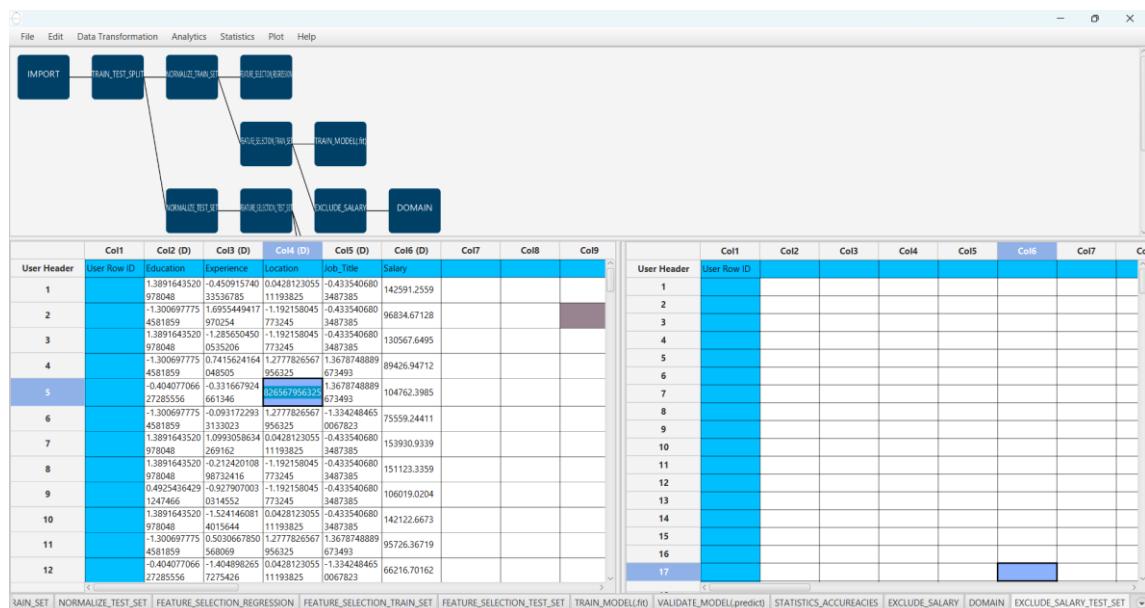
The results will appear on the output spreadsheet.



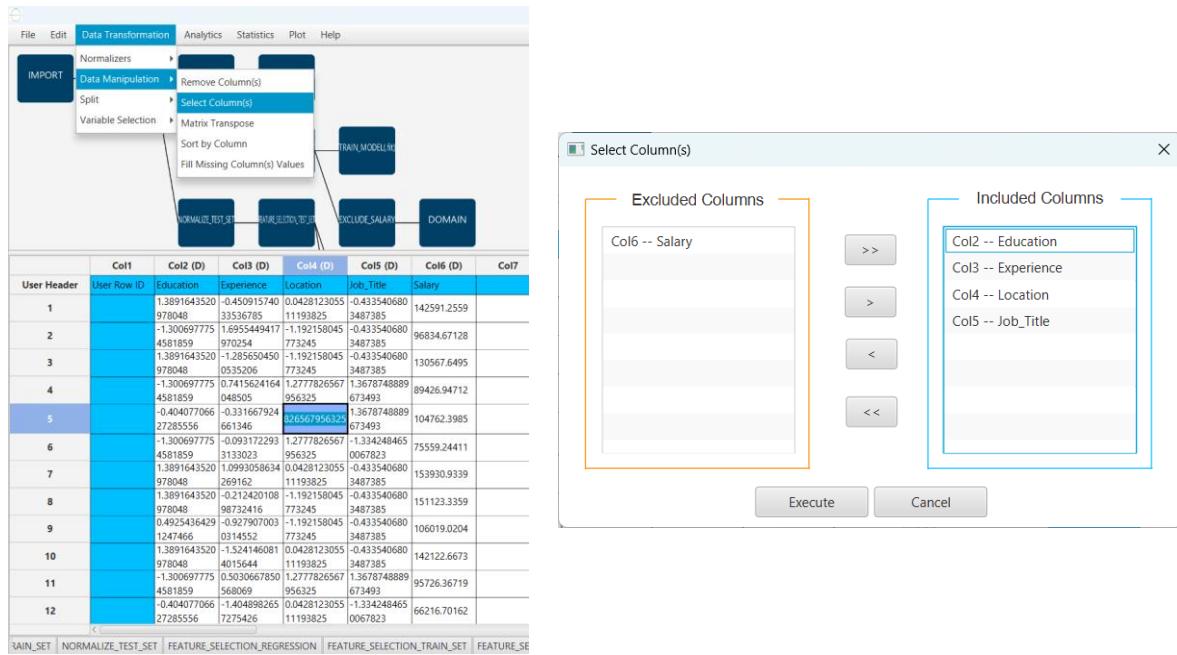
Step 12.b: Check the test set reliability

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

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



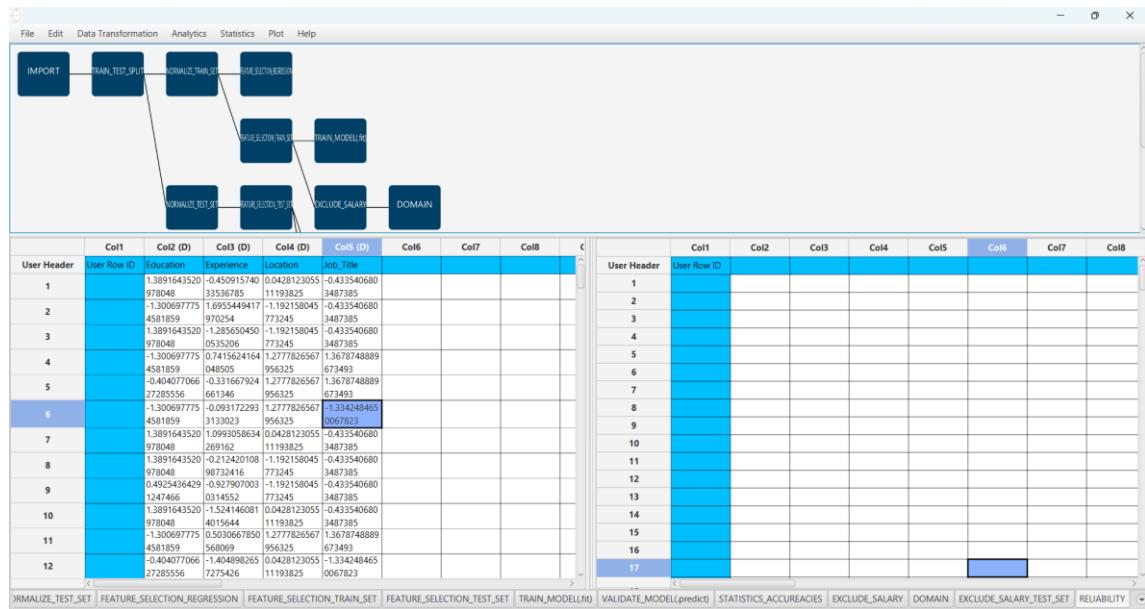
Filter the data to exclude the column that corresponds to the "Salary" by browsing: "Data Transformation" → "Data Manipulation" → "Select Columns". Then select all the columns except "Salary".



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_SALARY_TEST_SET" 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 screenshot shows a data flow interface and a corresponding spreadsheet output.

Data Flow Diagram:

```

graph LR
    IMPORT[IMPORT] --> TRAIN_TEST_SPLIT[TRAIN_TEST_SPLIT]
    TRAIN_TEST_SPLIT --> NORMALIZE_TRAIN_SET[NORMALIZE_TRAIN_SET]
    NORMALIZE_TRAIN_SET --> FEATURE_SELECTION_REGRESSION[FEATURE_SELECTION_REGRESSION]
    FEATURE_SELECTION_REGRESSION --> DOMAIN[DOMAIN]
    NORMALIZE_TRAIN_SET --> FEATURE_SELECTION_TRAIN_SET[FEATURE_SELECTION_TRAIN_SET]
    FEATURE_SELECTION_TRAIN_SET --> TRAIN_MODEL[TRAIN_MODEL]
    TRAIN_MODEL --> DOMAIN
    NORMALIZE_TEST_SET[NORMALIZE_TEST_SET] --> FEATURE_SELECTION_TEST_SET[FEATURE_SELECTION_TEST_SET]
    FEATURE_SELECTION_TEST_SET --> DOMAIN
    DOMAIN --> EXCLUDE_SALARY[EXCLUDE_SALARY]
    EXCLUDE_SALARY --> DOMAIN
  
```

Spreadsheet View:

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6	Col7
1	1.3891643520	0.450915740	0.042812055	-0.433540680	978048	33536785	1193825
2	-1.300697775	1.6955449417	-1.192158045	-0.433540680	4581859	970254	773245
3	1.3891643520	-1.285650450	-1.192158045	-0.433540680	978048	0535206	773245
4	-1.300697775	0.7415624164	1.2777826567	1.3678748889	4581859	048505	956325
5	-0.404077066	-0.331667924	1.2777826567	1.3678748889	27285556	661346	673493
6	-1.300697775	-0.093172295	1.2777826567	-1.33428465	4581859	310203	1193825
7	1.3891643520	0.9993058634	0.042812055	-0.433540680	978048	269162	11193825
8	1.3891643520	0.2112420108	-1.192158045	-0.433540680	978048	07275426	773245
9	0.4925436429	0.927907003	-1.192158045	-0.433540680	1247466	014552	773245
10	1.3891643520	-1.5224146081	0.042812055	-0.433540680	978048	4015644	11193825
11	-1.300697775	0.5030567850	1.2777826567	1.3678748889	4581859	568069	956325
12	-0.404077066	-1.404898265	0.042812055	-1.33428465	27285556	7275426	11193825

Below the table, there are tabs for: NORMALIZE_TEST_SET, FEATURE_SELECTION_REGRESSION, FEATURE_SELECTION_TRAIN_SET, and FEATURE_SELECTION_TEST_SET.

The results will appear on the output spreadsheet.

The screenshot shows a data flow interface and a detailed spreadsheet view of the model execution.

Data Flow Diagram:

```

graph LR
    IMPORT[IMPORT] --> TRAIN_TEST_SPLIT[TRAIN_TEST_SPLIT]
    TRAIN_TEST_SPLIT --> NORMALIZE_TRAIN_SET[NORMALIZE_TRAIN_SET]
    NORMALIZE_TRAIN_SET --> FEATURE_SELECTION_REGRESSION[FEATURE_SELECTION_REGRESSION]
    FEATURE_SELECTION_REGRESSION --> DOMAIN[DOMAIN]
    NORMALIZE_TRAIN_SET --> FEATURE_SELECTION_TRAIN_SET[FEATURE_SELECTION_TRAIN_SET]
    FEATURE_SELECTION_TRAIN_SET --> TRAIN_MODEL[TRAIN_MODEL]
    TRAIN_MODEL --> DOMAIN
    NORMALIZE_TEST_SET[NORMALIZE_TEST_SET] --> FEATURE_SELECTION_TEST_SET[FEATURE_SELECTION_TEST_SET]
    FEATURE_SELECTION_TEST_SET --> DOMAIN
    DOMAIN --> EXCLUDE_SALARY[EXCLUDE_SALARY]
    EXCLUDE_SALARY --> DOMAIN
  
```

Model Execution Details:

- Model:** (from Tab:)DOMAIN
- Type:** APD Model
- Description:** (empty)
- Model Input:**
 - Header -> Datatype
 - Education -> Double
 - Experience -> Double
 - Location -> Double
 - Job_Title -> Double
 - Age -> Double
 - Gender -> Double
- Transfer Column(s) to Output:** (checkbox checked)

Spreadsheet View:

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6	Col7
1	1.3891643520	0.450915740	0.042812055	-0.433540680	978048	33536785	1193825
2	-1.300697775	1.6955449417	-1.192158045	-0.433540680	4581859	970254	773245
3	1.3891643520	-1.285650450	-1.192158045	-0.433540680	978048	0535206	773245
4	-1.300697775	0.7415624164	1.2777826567	1.3678748889	4581859	048505	956325
5	-0.404077066	-0.331667924	1.2777826567	1.3678748889	27285556	661346	673493
6	-1.300697775	-0.093172295	1.2777826567	-1.33428465	4581859	310203	1193825
7	1.3891643520	0.9993058634	0.042812055	-0.433540680	978048	269162	11193825
8	1.3891643520	0.2112420108	-1.192158045	-0.433540680	978048	07275426	773245
9	0.4925436429	0.927907003	-1.192158045	-0.433540680	1247466	014552	773245
10	1.3891643520	-1.5224146081	0.042812055	-0.433540680	9581859	4015644	11193825
11	-1.300697775	0.5030567850	1.2777826567	1.3678748889	4581859	568069	673493
12	-0.404077066	-1.404898265	0.042812055	-1.33428465	27285556	7275426	11193825

Below the table, there are tabs for: NORMALIZE_TEST_SET, FEATURE_SELECTION_REGRESSION, FEATURE_SELECTION_TRAIN_SET, FEATURE_SELECTION_TEST_SET, TRAIN_MODEL(D), VALIDATE_MODEL(predict), STATISTICS_ACCURACIES, EXCLUDE_SALARY, DOMAIN, EXCLUDE_SALARY_TEST_SET, and RELIABILITY.

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:

