



Credit Card (Binary Classification)

The goal of this study is to train a model in order to predict whether the application is Approved (0) or Rejected (1). The dataset used in this case study is found in https://www.kaggle.com/datasets/rohitudageri/credit-card-details?select=Credit_card_label.csv and has 20 features and 1458 labelled samples.

Step 1: Import Data from the file

Right click on the input spreadsheet and choose the option "Import from file". Then navigate through your files to find the one with the credit card data.

The screenshot shows the Isalos Analytics Platform interface. At the top, there is a toolbar with options: File, Edit, Data Transformation, Analytics, Statistics, Plot, Help, and IMPORT. Below the toolbar is a large table with 1458 rows and 20 columns. The first few columns are labeled Col1 through Col6. The first row is labeled "User Header". A context menu is open over the data area, with the "Import from Spreadsheet" option highlighted. To the left of the main table, another smaller table is visible, also labeled "User Header" and showing the first few rows of the dataset. The bottom left corner of the interface has an "IMPORT" button.

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	Col6	Col7	Col8	Col9	Col10	Col11	Col12
1	0	1	0	0	1	0	1	0	1	0	0	0
2	1	0	0	0	1	1	0	1	0	0	0	0
3	1	0	0	0	1	1	0	1	0	0	0	0
4	1	0	0	0	1	1	0	1	0	0	0	0
5	1	0	0	0	1	1	0	1	0	0	0	0
6	0	0	1	0	1	1	0	0	1	0	0	0
7	1	0	0	0	1	1	0	1	0	0	0	0
8	1	0	0	1	0	1	0	0	1	0	0	0
9	0	1	0	0	1	0	1	1	0	0	0	0
10	0	1	0	0	1	0	1	0	1	0	0	0
11	0	1	0	0	1	0	1	1	0	0	0	0
12	1	0	0	0	1	0	1	0	0	0	0	1
13	1	0	0	0	1	0	1	0	0	0	0	1
14	1	0	0	0	1	0	1	0	0	0	0	1
15	0	1	0	1	0	1	0	0	1	0	0	0
16	0	1	0	1	0	0	1	0	0	0	0	1
17	0	1	0	1	0	0	1	0	0	0	0	1
18	1	0	0	1	0	0	1	1	0	0	0	0
19	1	0	0	1	0	0	1	1	0	0	0	0
20	0	1	0	1	0	0	1	0	1	0	0	0

Step 2: Manipulate Data

In order to use the data for training we have to exclude any columns that do not represent factor, like Ind_ID. We follow these steps to execute this:

Data Transformation → Data Manipulation → Select Column(s)

Select all columns except the one that corresponds to the Ind_ID.

The screenshot shows the Isalos Analytics Platform interface. On the left, there's a sidebar with 'IMPORT' and 'Data Transformation' selected. Under 'Data Transformation', 'Data Manipulation' is expanded, showing options like 'Remove Column(s)', 'Select Column(s)', 'Split', 'Variable Selection', 'Matrix Transpose', 'Sort by Column', and 'Fill Missing Column(s) Values'. To the right of the sidebar is a spreadsheet-like table with columns labeled Col1 through Col8. Below the table is a 'User Header' row. The first column is 'User Row ID' and the second is 'User Header'. The remaining columns are categorical features: cat_GENDER_F, cat_GENDER_M, cat_GENDER_nan, cat_Car_Owner_N, cat_Car_Owner_Y, cat_Propert_Owner_N, and cat_Propert_Owner_Y. The table has 20 rows of data. To the right of the table is a 'Select Column(s)' dialog box. It has two main sections: 'Excluded Columns' (highlighted with an orange border) containing 'Col48 -- Ind_ID' and 'Included Columns' (highlighted with a blue border) containing 'Col2 -- cat_GENDER_F', 'Col3 -- cat_GENDER_M', 'Col4 -- cat_GENDER_nan', 'Col5 -- cat_Car_Owner_N', 'Col6 -- cat_Car_Owner_Y', 'Col7 -- cat_Propert_Owner_N', 'Col8 -- cat_Propert_Owner_Y', and 'Col9 -- cat_Type_Income_C'. At the bottom of the dialog are 'Execute' and 'Cancel' buttons.

The data without the Ind_ID column will appear in the output spreadsheet.

Step 3: Fill missing values

There are empty values in the Dataset. Specifically, we show below how many missing values there are for each feature:

```
Empty data:
Ind_ID          0
GENDER          7
Car_Owner        0
Propert_Owner    0
CHILDREN         0
Annual_income    23
Type_Income      0
EDUCATION        0
Marital_status   0
Housing_type     0
Birthday_count   22
Employed_days    0
Mobile_phone      0
Work_Phone        0
Phone             0
EMAIL_ID          0
Type_Occupation  488
Family_Members    0
dtype: int64
```

Create a new action by pressing the + button on the bottom of the page with the name FILL_MISSING_VALUES which will be used to fill the missing values.

Import Data into the input spreadsheet of the FILL_MISSING_VALUES action from the output of the IMPORT action by right-clicking on the input spreadsheet and then choosing Import from SpreadSheet.

User Header	User Row ID	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (I)	Col8 (I)	Col9 (I)	Col10 (I)
1		0	1	0	0	1	0	0	0	0	0
2		1	0	0	0	1	1	1	1	1	1
3		1	0	0	0	1	1	1	1	1	1
4		1	0	0	0	1	1	1	1	1	1
5		1	0	0	0	1	1	1	1	1	1
6		0	0	1	0	1	1	1	1	1	1
7		1	0	0	0	1	1	1	1	1	1
8		1	0	0	1	0	1	1	1	1	1
9		0	1	0	0	1	0	0	0	0	0
10		0	1	0	0	0	1	0	0	0	0
11		0	1	0	0	1	0	0	0	0	0
12		1	0	0	0	1	0	0	0	0	0
13		1	0	0	0	0	1	0	0	0	0
14		1	0	0	0	1	0	0	0	0	0
15		0	1	0	1	0	1	0	1	0	1
16		0	1	0	1	0	0	0	0	0	0
17		0	1	0	1	0	0	0	0	0	0
18		1	0	0	0	1	0	0	0	0	0
19		1	0	0	1	0	0	0	0	0	0
20		0	1	0	1	0	0	0	0	0	0
21		0	1	0	1	0	0	1	0	1	1

Handle missing columns values:

Data Transformation → Data Manipulation → Fill missing column(s) Values:

Then choose the Mean as the Numerical Method

User Header	User Row ID	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (I)	Col8 (I)	Col9 (I)	Col10 (I)
1		0	1	0	0	1	0	0	0	0	0
2		1	0	0	0	0	1	1	1	1	1
3		1	0	0	0	0	1	1	1	1	1
4		1	0	0	0	0	1	1	1	1	1
5		1	0	0	0	0	1	1	1	1	1
6		0	0	1	0	0	1	1	1	1	1
7		1	0	0	0	0	0	1	1	1	1
8		1	0	0	0	1	0	0	1	1	1
9		0	1	0	0	0	0	1	0	0	0
10		0	1	0	0	0	0	1	0	0	0
11		0	1	0	0	0	1	0	0	0	0
12		1	0	0	0	0	1	0	0	0	0
13		1	0	0	0	0	0	1	0	0	0
14		1	0	0	0	0	0	1	0	0	0
15		0	1	0	1	0	0	0	1	0	1
16		0	1	0	0	1	0	0	0	0	0
17		0	1	0	0	1	0	0	0	0	0
18		1	0	0	0	1	0	0	0	0	0
19		1	0	0	0	1	0	0	0	0	0
20		0	1	0	1	0	1	0	0	0	0
21		0	1	0	1	0	0	1	0	1	1

The results will appear on the output spreadsheet.

User Header	User Row ID	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (I)	Col8 (I)	Col9 (I)
1		0	1	0	0	1	0	1	0	0
2		1	0	0	0	1	1	0	1	0
3		1	0	0	0	1	1	0	1	0
4		1	0	0	0	1	1	0	1	0
5		1	0	0	0	1	1	0	1	0
6		0	0	1	0	1	1	0	0	0
7		1	0	0	0	1	1	0	1	0
8		1	0	0	1	0	1	0	0	0
9		0	1	0	0	1	0	1	1	0
10		0	1	0	0	1	0	1	0	0
11		0	1	0	0	1	0	1	1	0
12		1	0	0	0	1	0	1	0	0
13		1	0	0	0	1	0	1	0	0
14		1	0	0	0	1	0	1	0	0
15		0	1	0	1	0	1	0	0	0
16		0	1	0	1	0	0	1	0	0
17		0	1	0	1	0	0	1	0	0

Step 4: Split Data

Create a new action 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 action from the output of the FILL_MISSING_VALUES action by right-clicking on the input spreadsheet and then choosing Import from SpreadSheet.

User Header	User Row ID	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (I)	Col8 (I)	Col9 (I)
1		0	1	0	0	1	0	1	0	0
2		1	0	0	0	1	1	0	1	0
3		1	0	0	0	1	1	0	1	0
4		1	0	0	0	1	1	1	0	0
5		1	0	0	0	1	1	1	0	0
6		0	0	1	0	1	1	0	1	0
7		1	0	0	0	1	1	1	0	0
8		1	0	0	1	0	1	0	0	0
9		0	1	0	0	1	0	1	0	0
10		0	1	0	0	1	0	0	0	0
11		0	1	0	0	1	0	1	0	0
12		1	0	0	0	1	0	0	1	0
13		1	0	0	0	1	0	0	1	0
14		1	0	0	0	1	0	1	0	1
15		0	1	0	1	0	1	0	1	0
16		0	1	0	1	0	0	1	0	1
17		0	1	0	1	0	0	1	0	0

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. At the top, the 'Data Transformation' menu is open, with 'Split' selected. A sub-menu for 'Random Partitioning' is shown, containing 'Kennard-Stone' and 'Random Partitioning'. Below this, a spreadsheet view shows a dataset with columns: Col1, Col2 (I), Col3 (I), Col4 (I), Col5 (I), Col6 (I), and Col7 (I). The first row is labeled 'User Header'. To the right, a 'Random Partitioning' dialog box is open, with 'Training set percentage' set to 'Integer (0,100), Default: 40'. There are two checkboxes: 'Usage of random generator seed' (unchecked) and 'Stratified sampling' (checked). A dropdown menu next to 'Stratified sampling' shows 'Col57 -- label'. At the bottom of the dialog are 'Execute' and 'Cancel' buttons.

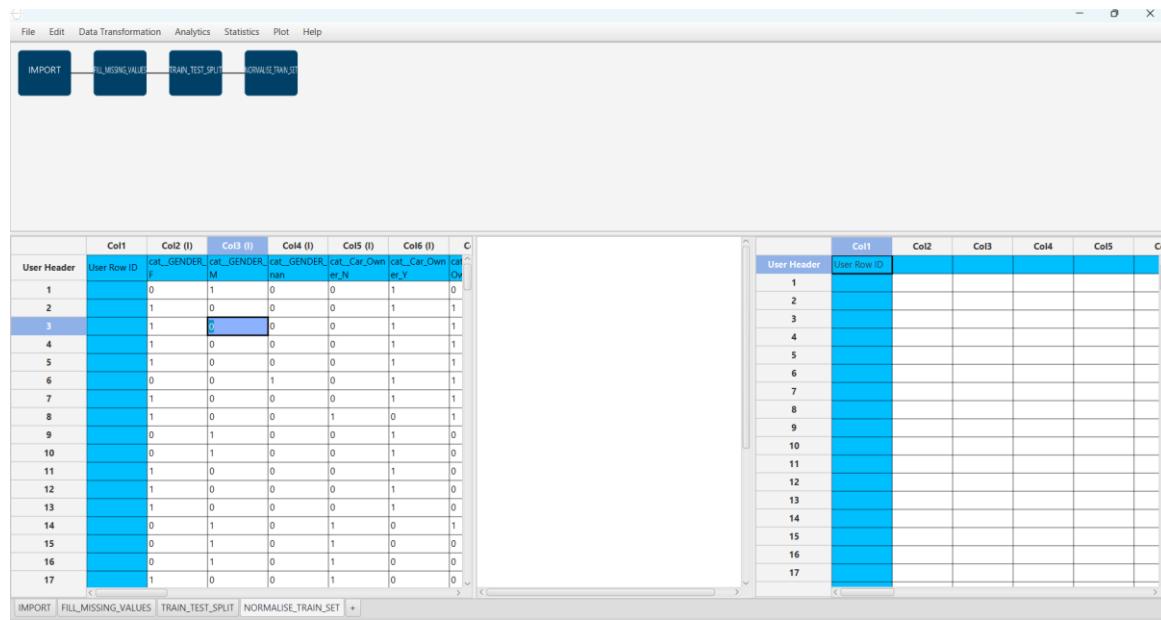
The results will appear on the output spreadsheet.

The screenshot shows the workflow setup and the resulting datasets. At the top, a workflow diagram consists of three actions: 'IMPORT', 'FILL_MISSING_VALUES', and 'TRAIN_TEST_SPLIT'. Below the workflow, two output spreadsheets are displayed side-by-side. Both spreadsheets have identical column headers: Col1, Col2 (I), Col3 (I), Col4 (I), Col5 (I), Col6 (I), Col7 (I), Col8 (I), and Col9 (I). The first spreadsheet is labeled 'User Header' and contains 17 rows of data. The second spreadsheet is also labeled 'User Header' and contains 17 rows of data. Both datasets include columns for gender ('cat_GENDER_F', 'cat_GENDER_M'), car ownership ('cat_Car_Own_N', 'cat_Car_Own_Y'), property ownership ('cat_Property_Owner_N', 'cat_Property_Owner_Y'), and other categorical variables.

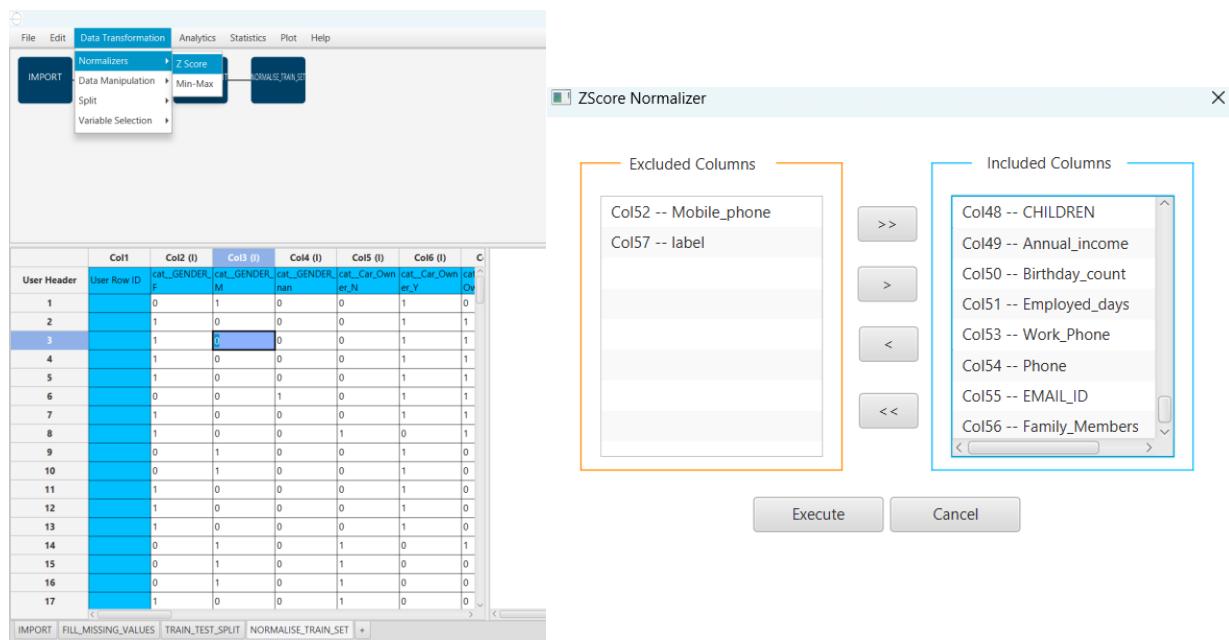
Step 5: Normalize the Training Set

Create a new action 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 action the train set from the output of the TRAIN_TEST_SPLIT action 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: Data Transformation → Normalize → Z-Score
Then select all columns excluding Mobile_phone and Label and click Execute.



The results will appear on the output spreadsheet.

User Header	User Row ID	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7
1		0	1	0	0	1	0	
2		1	0	0	0	1	0	
3		1	0	0	0	1	1	
4		1	0	0	0	1	1	
5		1	0	0	0	1	1	
6		0	0	1	0	1	1	
7		1	0	0	0	1	1	
8		1	0	0	1	0	1	
9		0	1	0	0	1	0	
10		0	1	0	0	1	0	
11		1	0	0	0	1	0	
12		1	0	0	0	1	0	
13		1	0	0	0	1	0	
14		0	1	0	1	0	1	
15		0	1	0	1	0	0	
16		0	1	0	1	0	0	
17		1	0	0	1	0	0	

Step 6: Normalize the Test Set

Create a new action 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 action the test set from the output of the TRAIN_TEST_SPLIT action 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	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7
1		0	1	0	0	1	0	
2		1	0	0	1	0	0	
3		1	0	0	0	1	0	
4		1	0	0	1	0	0	
5		1	0	0	1	0	0	
6		1	0	0	1	0	0	
7		0	1	0	0	1	1	
8		0	1	0	0	1	0	
9		1	0	0	0	1	1	
10		1	0	0	1	0	0	
11		1	0	0	1	0	0	
12		0	1	0	1	0	1	
13		0	1	0	0	1	1	
14		0	1	0	0	1	0	
15		1	0	0	0	1	0	
16		0	1	0	1	0	1	
17		1	0	0	1	0	0	

Normalize the test set using the existing normalizer of the training set:
 Analytics → Existing Model Utilization → Model: NORMALIZE_TRAIN_SET

The screenshot shows the 'Existing Model Execution' dialog. The 'Model' dropdown is set to '(from Tab:)NORMALISE_TRAIN_SET'. The 'Type' is 'Z Score Normalizer Model'. The 'Description' section lists the model's input headers and their data types. The 'Model Input' section shows the mapping from input columns to output columns. Below the dialog is a preview of the normalized data in a spreadsheet.

User Header	User Row ID	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (I)	Col8 (I)	Col9 (I)
1		cat_GENDER_F	cat_GENDER_M	cat_GENDER_nan	cat_Car_Own_N	cat_Car_Own_Y	cat_Property_Owner_N	cat_Property_Owner_Y	cat_Type_Income_Commercial_associate	cat_Type_Income_Domestic
2		0	1	0	0	1	0	0	0	0
3		1	0	0	0	0	1	0	0	0
4		1	0	0	0	1	0	0	0	0
5		1	0	0	0	1	0	0	0	0
6		1	0	0	0	1	0	0	0	0
7		0	1	0	0	0	1	1	0	0
8		0	1	0	0	0	0	1	0	0
9		1	0	0	0	0	0	1	1	0
10		1	0	0	0	1	0	0	0	0
11		1	0	0	0	1	0	0	0	0
12		0	1	0	0	1	0	0	1	0
13		0	1	0	0	0	1	1	0	0
14		0	1	0	0	0	0	1	0	0
15		1	0	0	0	0	0	1	0	0
16		0	1	0	0	1	0	0	1	0
17		1	0	0	0	1	0	0	0	0

The results will appear on the output spreadsheet.

The screenshot shows the 'NORMALISE_TEST_SET' step in the workflow. The workflow consists of steps: IMPORT, FILL_MISSING_VALUES, TRAIN_TEST_SPLIT, and NORMALISE_TRAIN_SET. The NORMALISE_TRAIN_SET step is connected to the output of TRAIN_TEST_SPLIT. The output of NORMALISE_TRAIN_SET is shown in a preview spreadsheet, which is identical to the normalized training data shown in the previous screenshot.

User Header	User Row ID	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)
1	-1.303278843	1.3202947062	-0.077850068	-1.222900234	1.2229002340	-0.724687135	0.7246871357	1.802	35875	
2	9345493	352057	73356258	0161226	161226	7407703	407701	35875		
3	791466	8421301	73356258	4353	2435302	7407703	407701	35875		
4	791466	8421301	73356258	0161226	161226	7407703	407701	35875		
5	791466	8421301	73356258	0161226	161226	7407703	407701	35875		
6	791466	8421301	73356258	4353	2435302	7407703	407701	35875		
7	791466	8421301	73356258	0161226	161226	7407703	407701	35875		
8	9345493	352057	73356258	0161226	161226	7407703	407701	35875		
9	791466	8421301	73356258	4353	2435302	7407703	407701	35875		
10	791466	8421301	73356258	0161226	161226	7407703	407701	35875		
11	791466	8421301	73356258	4353	2435302	7407703	407701	35875		
12	791466	8421301	73356258	0161226	161226	7407703	407701	35875		
13	791466	8421301	73356258	4353	2435302	7407703	407701	35875		
14	791466	8421301	73356258	0161226	161226	7407703	407701	35875		
15	791466	8421301	73356258	4353	2435302	7407703	407701	35875		
16	791466	8421301	73356258	0161226	161226	7407703	407701	35875		
17	791466	8421301	73356258	4353	2435302	7407703	407701	35875		
18	791466	8421301	73356258	0161226	161226	7407703	407701	35875		
19	791466	8421301	73356258	4353	2435302	7407703	407701	35875		
20	791466	8421301	73356258	0161226	161226	7407703	407701	35875		

Step 7: 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	User Row ID	Col1 (D)	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)	Col10 (D)
1	1303278843	-0.7246871357	-0.7246871357	-0.7246871357	-0.7246871357	-0.7246871357	-0.7246871357	-0.7246871357	-0.7246871357	-0.7246871357	-0.7246871357
2	9345493	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062
3	791466	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140
4	791466	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140
5	791466	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140
6	791466	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062
7	791466	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140
8	791466	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140
9	791466	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062
10	791466	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140
11	791466	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140
12	791466	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140
13	791466	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062
14	791466	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062
15	791466	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062
16	791466	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062

Use the XGBoost Method to train and fit the model:

"Analytics" → "Classification" → "XGBoost"

and set the "number of estimators" as 210, the "column sample by tree" as 10, the "Target Column" as the column corresponding to "Label" and use the following "RNG Seed":
1732285527644.

User Header	User Row ID	Col1 (D)	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9 (D)	Col10 (D)
1	1303278843	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062
2	9345493	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140
3	791466	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140
4	791466	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140
5	791466	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140
6	791466	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062
7	791466	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140
8	791466	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140	0.7666346140
9	791466	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062
10	791466	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062
11	791466	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062
12	791466	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062
13	791466	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062
14	791466	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062
15	791466	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062
16	791466	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062	0.1302947062

The predictions will appear on the output spreadsheet.

cc.ekk									
File Edit Data Transformation Analytics Statistics Plot Help									
IMPORT		FILL_MISSING_VALUES		TRAIN_TEST_SPLIT		NORMALISE_TRAIN_SET		TRAIN_MODEL(R)	
User Header	User Row ID	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
1	1	-1.303278843	1.3202947062	-0.077850068	-1.2229002340	0.724687135	0.724687135	0.724687135	0.724687135
	2	9345493	352057	73356258	0161226	161226	7407703	407701	1.0000000
2	3	0.7666346140	0.756754282	0.756754282	0.756754282	0.756754282	0.756754282	0.756754282	0.756754282
	4	791466	8421301	73356258	0161226	161226	466154	7468154	0.724687135
3	5	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	6	791466	8421301	73356258	0161226	161226	466154	7468154	0.724687135
4	7	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	8	791466	8421301	73356258	0161226	161226	466154	7468154	0.724687135
5	9	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	10	791466	8421301	73356258	0161226	161226	466154	7468154	0.724687135
6	11	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	12	791466	8421301	73356258	0161226	161226	466154	7468154	0.724687135
7	13	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	14	791466	8421301	73356258	0161226	161226	466154	7468154	0.724687135
8	15	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	16	791466	8421301	73356258	0161226	161226	466154	7468154	0.724687135
	17	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	18	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	19	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	20	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	21	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	22	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	23	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	24	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282

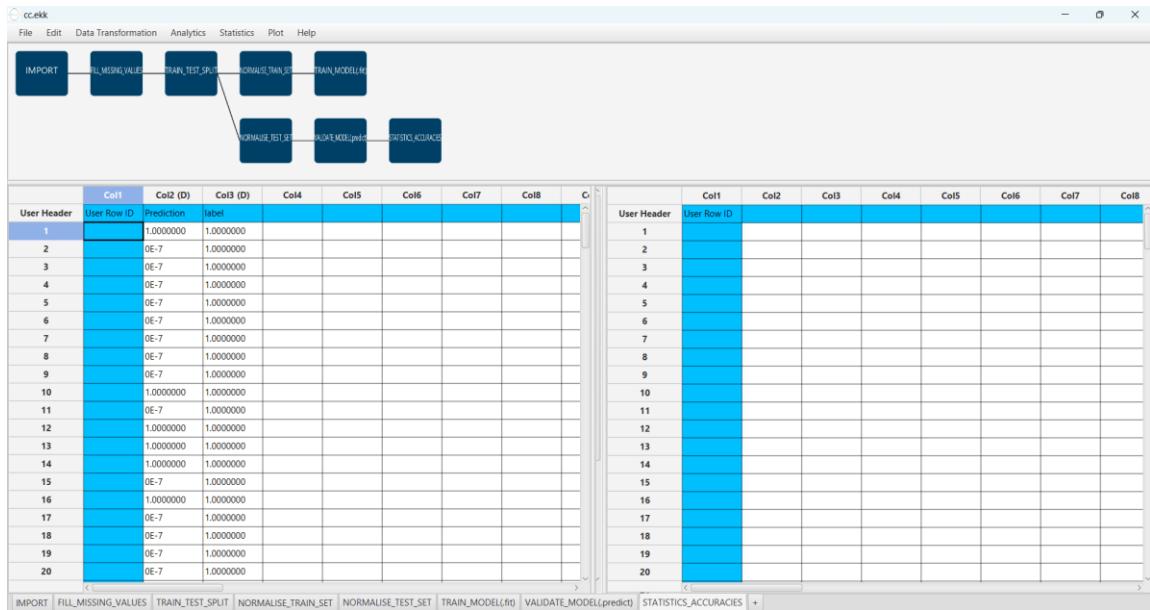
Step 8: Validate the model

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

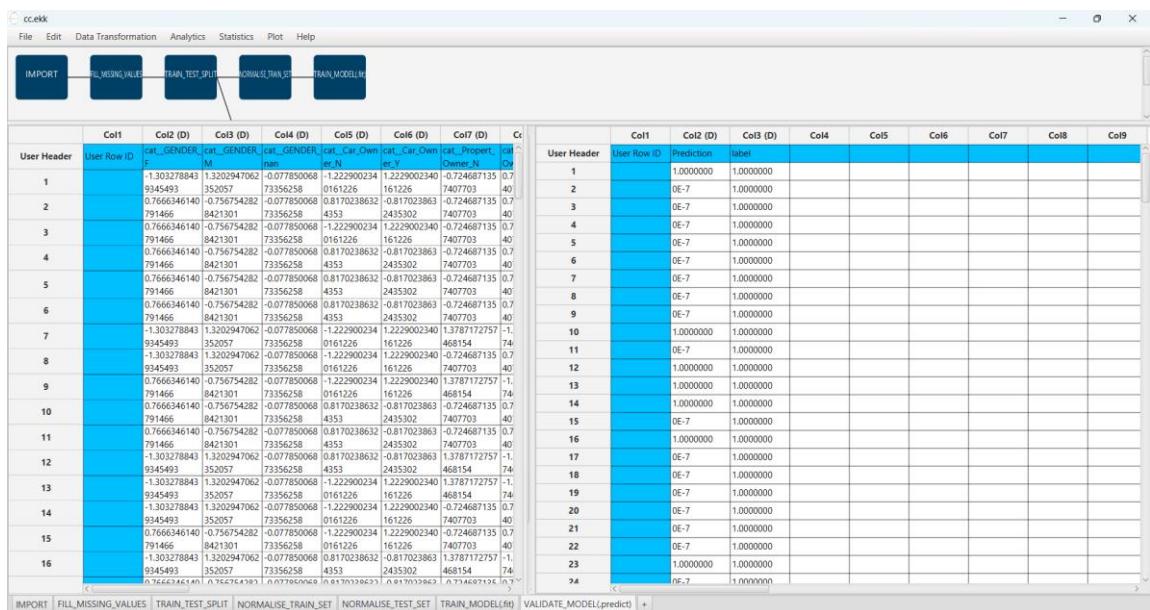
Import data into the input spreadsheet of the "VALIDATE_MODEL(.predict)" tab from the output of the "NORMALISE _TEST_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".

cc.ekk									
File Edit Data Transformation Analytics Statistics Plot Help									
IMPORT		FILL_MISSING_VALUES		TRAIN_TEST_SPLIT		NORMALISE_TRAIN_SET		NORMALISE_TEST_SET	
User Header	User Row ID	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
1	1	-1.303278843	1.3202947062	-0.077850068	-1.2229002340	0.724687135	0.724687135	0.724687135	0.724687135
	2	9345493	352057	73356258	0161226	161226	7407703	407701	1.0000000
2	3	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	4	791466	8421301	73356258	0161226	161226	466154	7468154	0.724687135
3	5	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	6	791466	8421301	73356258	0161226	161226	466154	7468154	0.724687135
4	7	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	8	791466	8421301	73356258	0161226	161226	466154	7468154	0.724687135
5	9	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	10	791466	8421301	73356258	0161226	161226	466154	7468154	0.724687135
6	11	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	12	791466	8421301	73356258	0161226	161226	466154	7468154	0.724687135
7	13	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	14	791466	8421301	73356258	0161226	161226	466154	7468154	0.724687135
8	15	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	16	791466	8421301	73356258	0161226	161226	466154	7468154	0.724687135
	17	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	18	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	19	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	20	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	21	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	22	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	23	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282
	24	0.7666346140	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282	-0.756754282

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



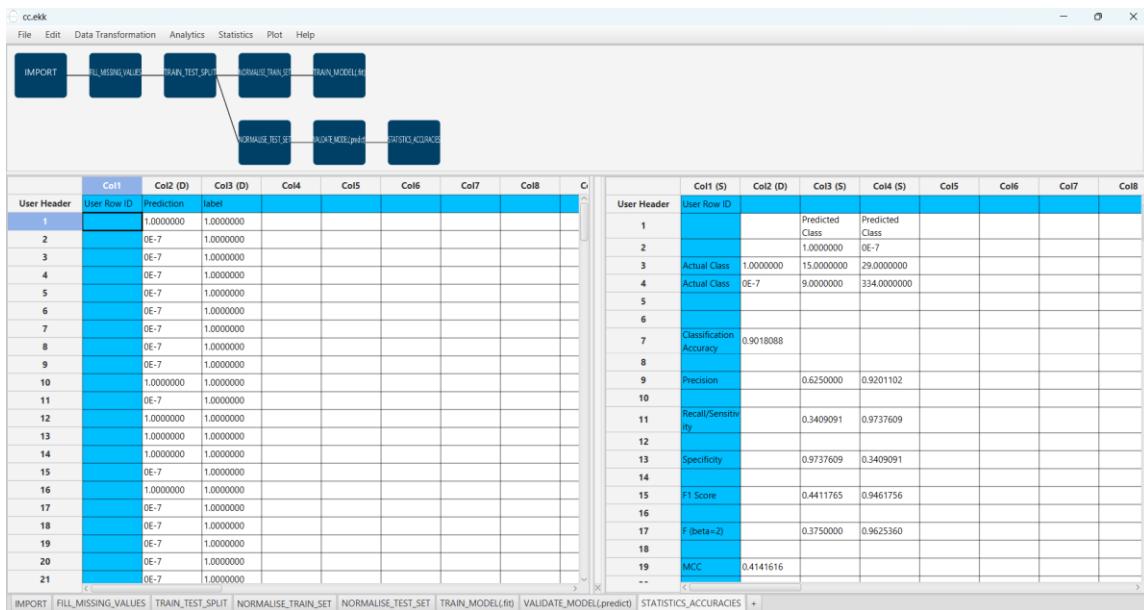
The predictions will appear on the output spreadsheet.



Step 9: 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 following steps:

- The **Analytics** tab is selected in the menu bar.
- A context menu is open over the **TRAIN_MODEL[fit]** node, with the **Classification Metrics** option highlighted.
- A configuration dialog titled **Classification Statistics Metrics** is displayed. It contains fields for **Actual Value Column** (set to **Col3 -- label**) and **Prediction Value Column** (set to **Col2 -- Prediction**). A **beta of F Score** input field is set to **2**.

The results will appear on the output spreadsheet.

Accuracy: 0.902

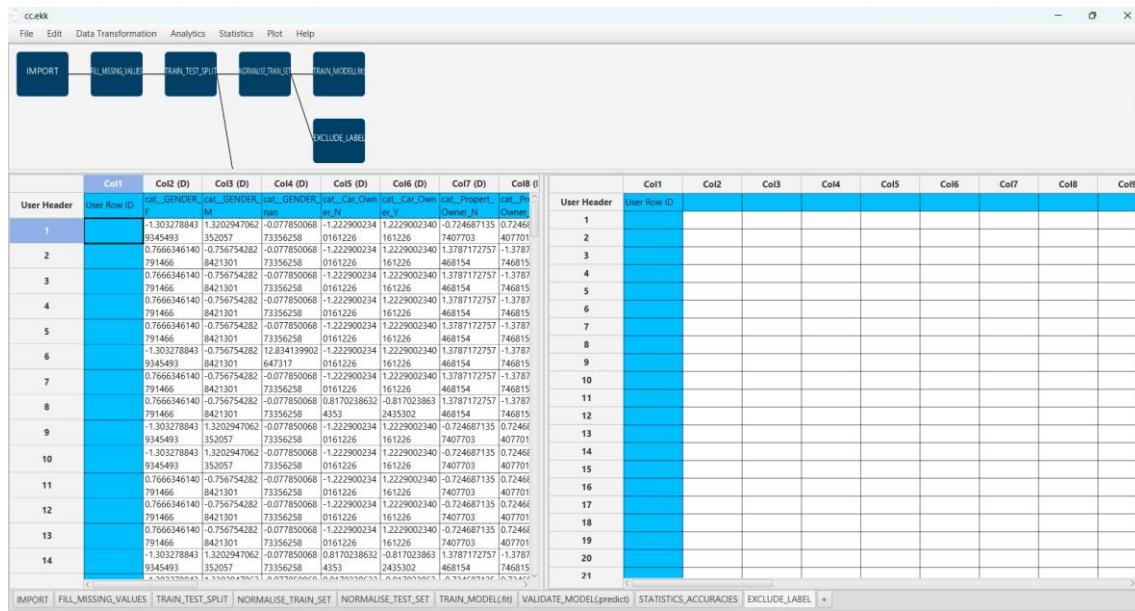
F1-Score = 0.694

Step 10: Reliability check of each record of the test set

Step 10.a: Create the domain

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

Import data into the input spreadsheet of the "EXCLUDE_LABEL" tab from the output of the "NORMALISE_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 "label"

"Data Transformation" → "Data Manipulation" → "Select Columns"

Then select all the columns except the "label".

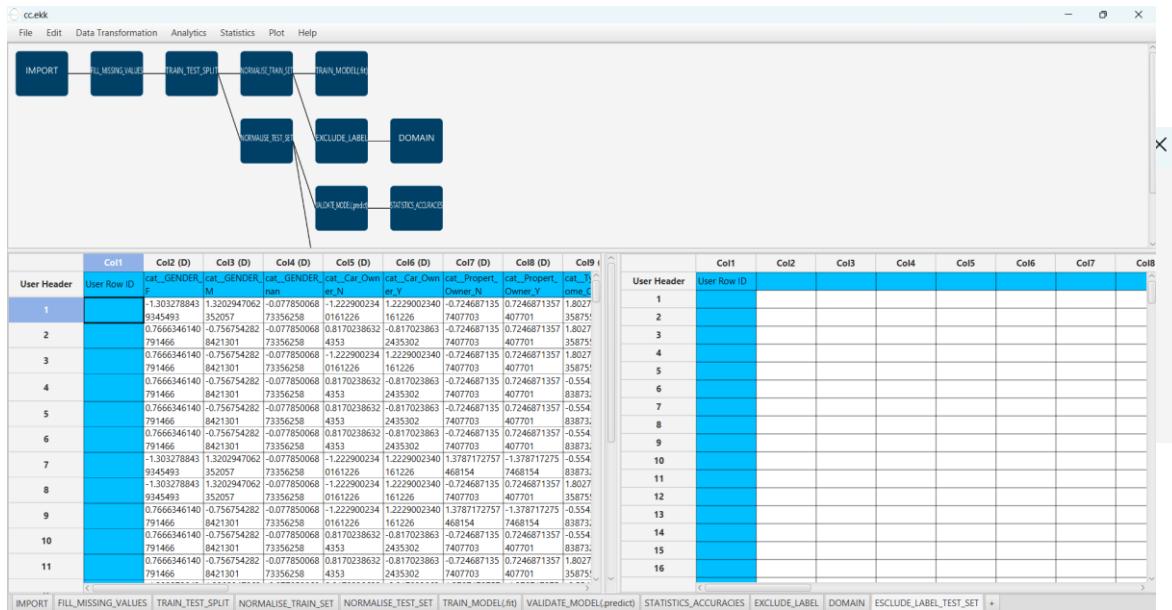
The results will appear on the output spreadsheet.

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

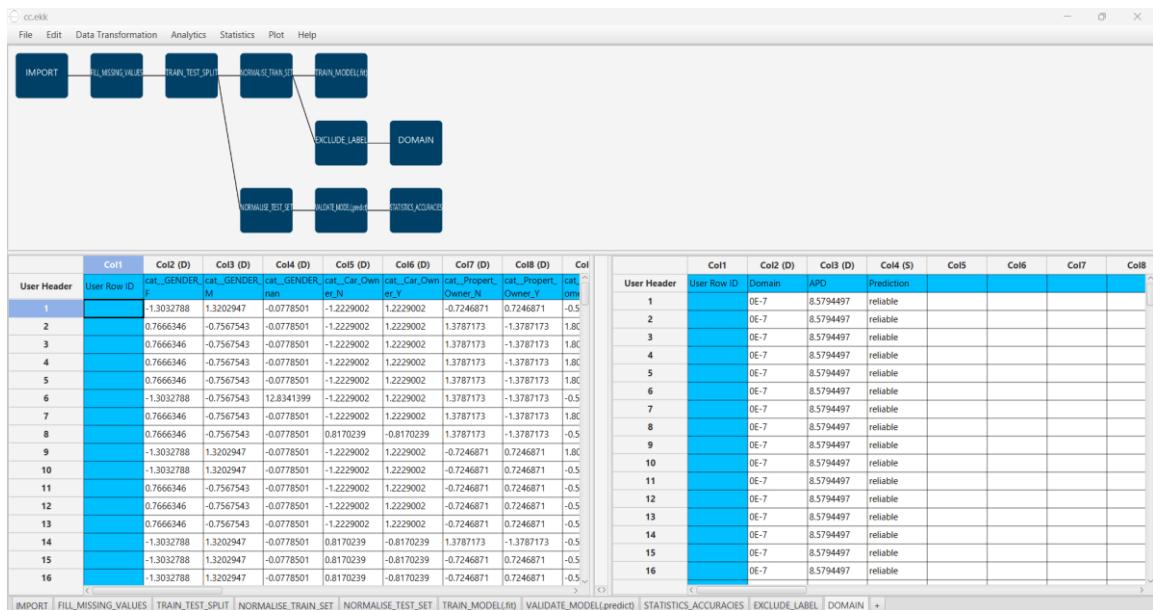
Import data into the input spreadsheet of the "DOMAIN" tab from the output of the "EXCLUDE_LABEL" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".

Create the domain:

"Statistics" → "Domain APD"



The results will appear on the output spreadsheet.



Step 10.b: Check the test set reliability

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

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

Filter the data to exclude the column that corresponds to the "label"

"Data Transformation" → "Data Manipulation" → "Select Column(s)".

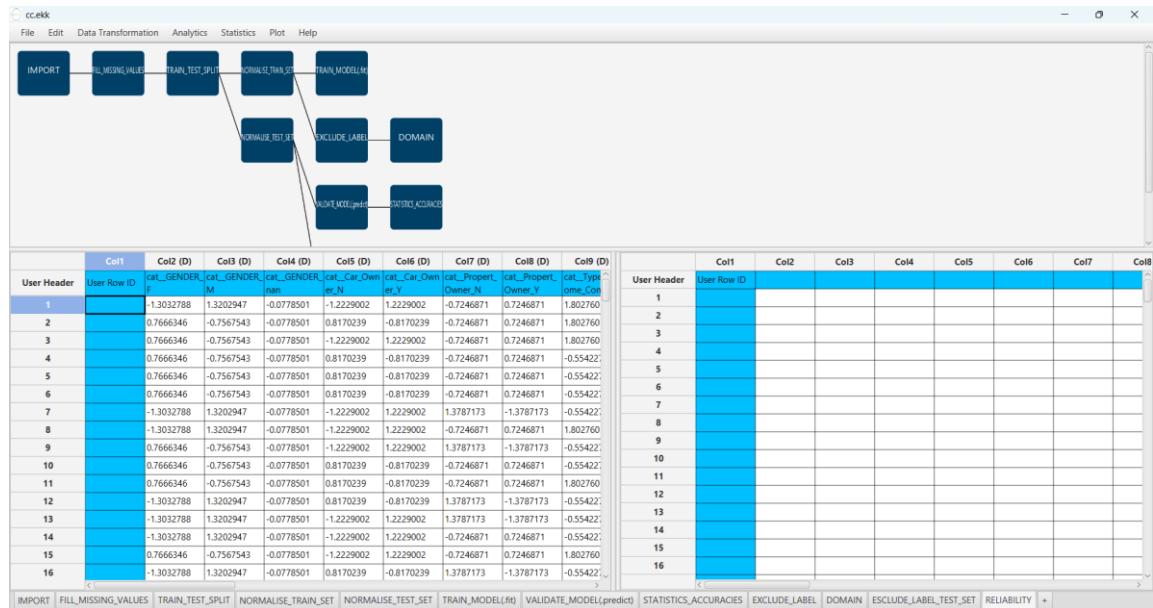
Then select all the columns except "label".

User Header	User Row ID	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)
1	9345493	-1.303278843	1.3202947062	-0.077850068	-1.2229002340	1.2229002340	0.724687135	
2	791466	0.7666346140	0.756754282	-0.077850068	0.8170238632	-0.817023863	0.724687135	
3	791466	0.7666346140	0.756754282	-0.077850068	0.8170238632	-0.817023863	0.724687135	
4	791466	0.7666346140	0.756754282	-0.077850068	0.8170238632	-0.817023863	0.724687135	
5	791466	0.7666346140	0.756754282	-0.077850068	0.8170238632	-0.817023863	0.724687135	
6	791466	0.7666346140	0.756754282	-0.077850068	0.8170238632	-0.817023863	0.724687135	
7	9345493	1.3202947062	-0.077850068	-1.2229002340	1.2229002340	0.724687135		
8	791466	0.7666346140	0.756754282	-0.077850068	0.8170238632	-0.817023863	0.724687135	
9	791466	0.7666346140	0.756754282	-0.077850068	0.8170238632	-0.817023863	0.724687135	
10	791466	0.7666346140	0.756754282	-0.077850068	0.8170238632	-0.817023863	0.724687135	
11	791466	0.7666346140	0.756754282	-0.077850068	0.8170238632	-0.817023863	0.724687135	

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".

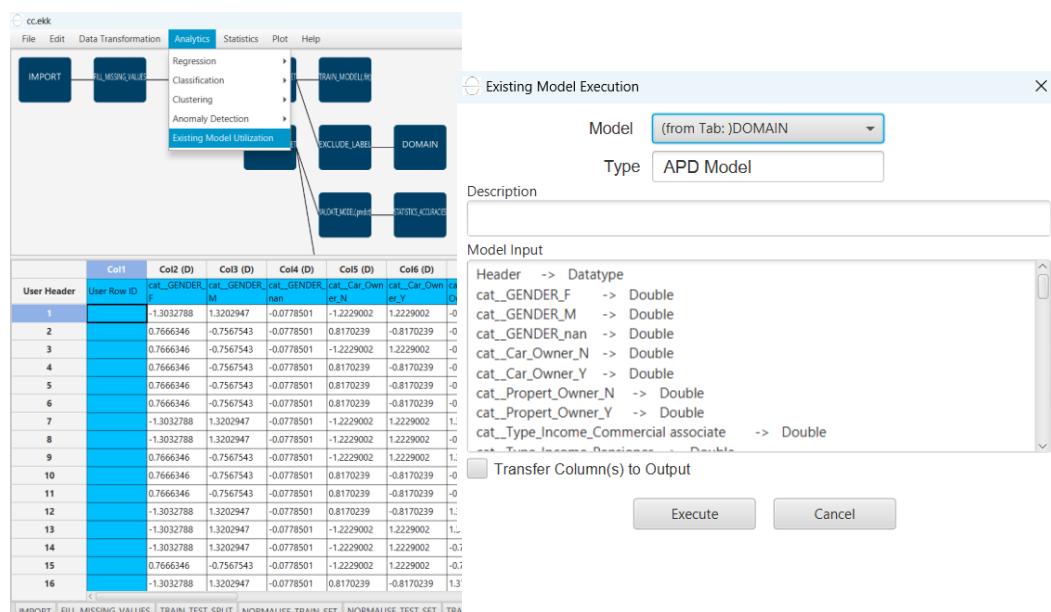
Import data into the input spreadsheet of the "RELIABILITY" tab from the output of the "EXCLUDE_LABEL_TEST_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".



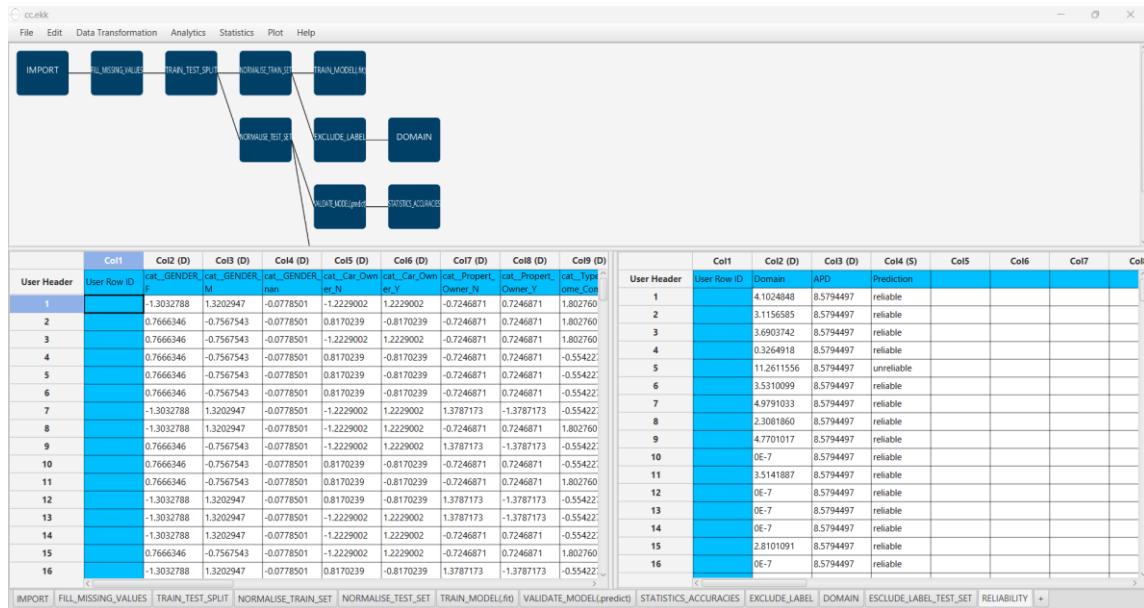
Check the Reliability:

"Analytics" → "Existing Model Utilization".

Then select as Model "(from Tab:) DOMAIN".



The results will appear on the output spreadsheet.



There are four unreliable samples in the test set.

Final Isalos Workflow

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

