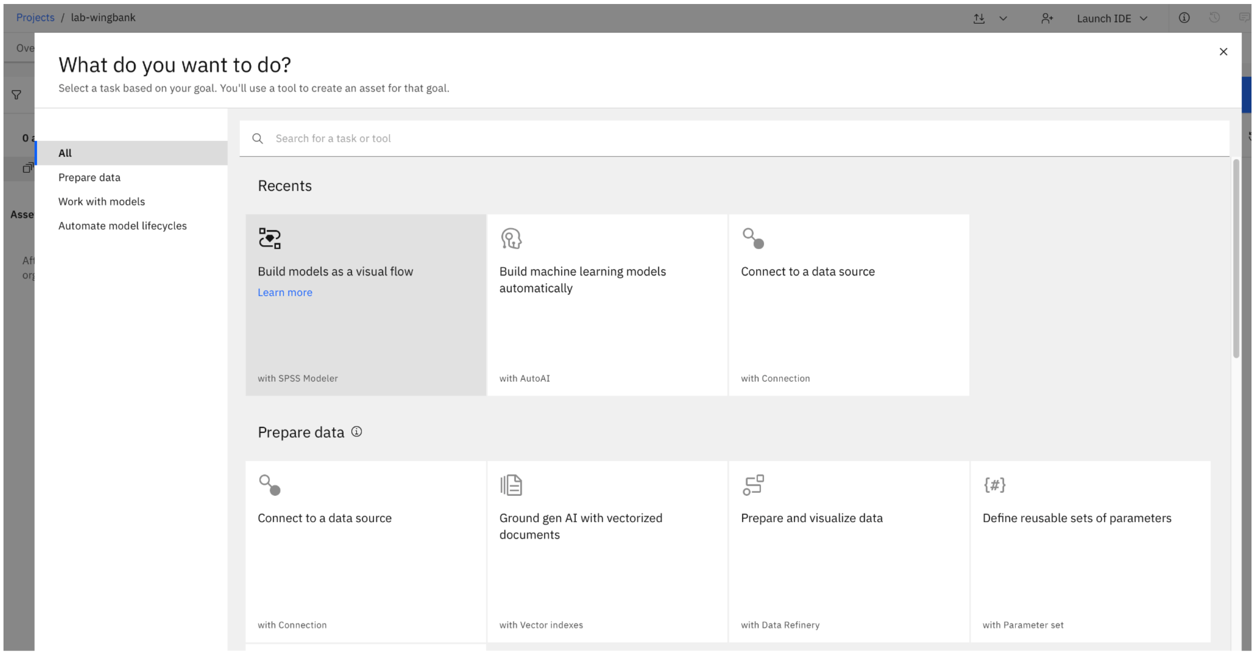
Lab 4 SPSS Modeler

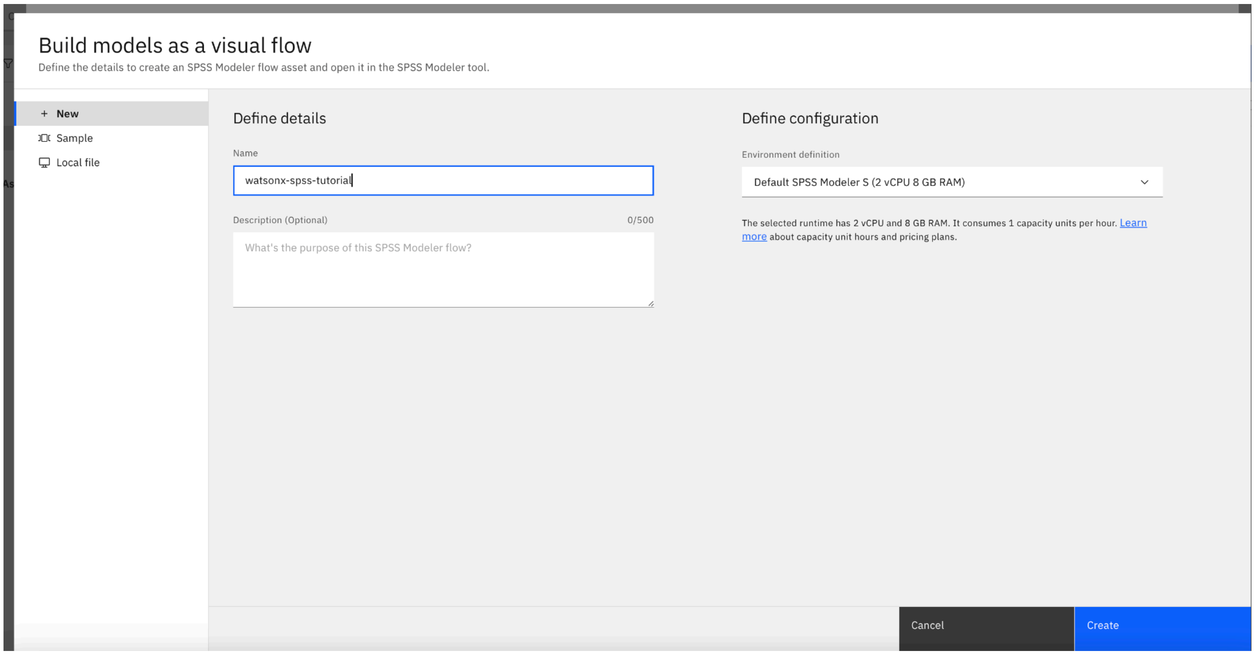
Step 1) Select the SPSS Modeler, “Build models as a visual flow”.



What do you want to do?

Select a task based on your goal. You'll use a tool to create an asset for that goal.

Step 2) Type in the name of the visual flow. This will be the name of your spss modeler asset.



Build models as a visual flow

Define the details to create an SPSS Modeler flow asset and open it in the SPSS Modeler tool.

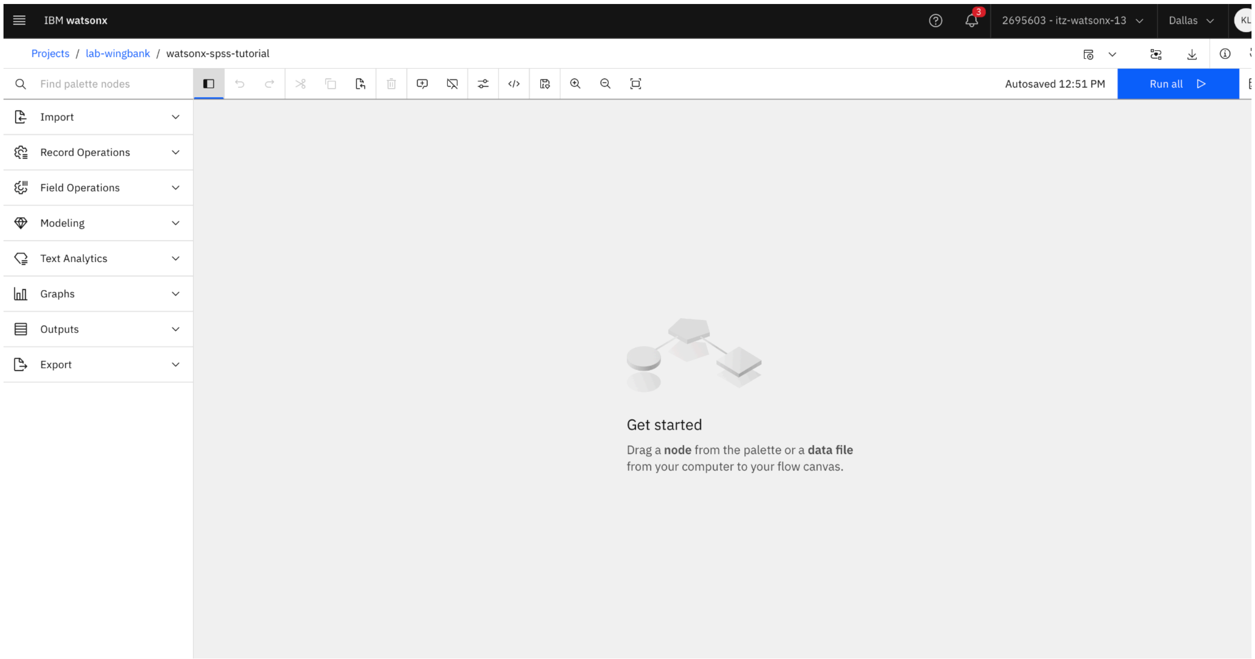
Name: watsonx-spss-tutorial

Description (Optional) 0/500: What's the purpose of this SPSS Modeler flow?

Environment definition: Default SPSS Modeler S (2 vCPU 8 GB RAM)

The selected runtime has 2 vCPU and 8 GB RAM. It consumes 1 capacity units per hour. Learn more about capacity unit hours and pricing plans.

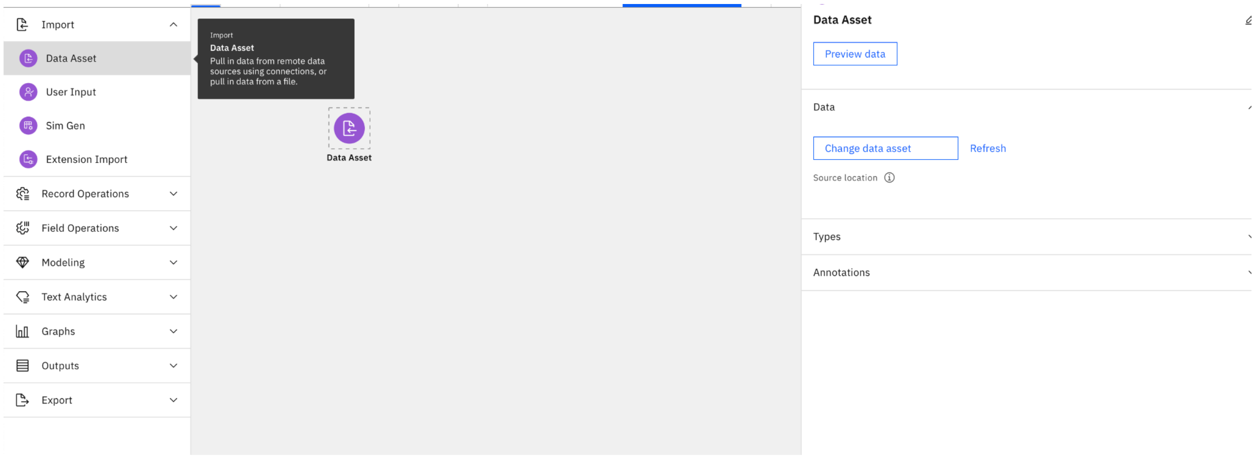
Once the resource is created, you should see this empty canvas.



Get started

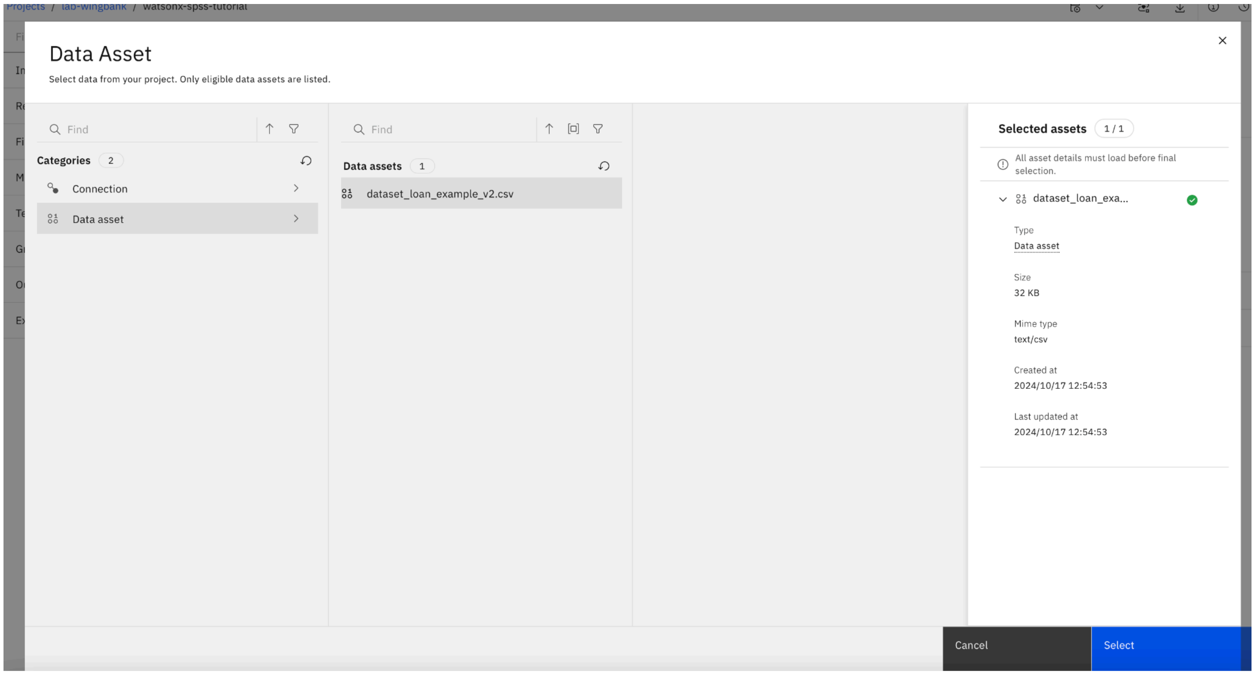
Drag a **node** from the palette or a **data file** from your computer to your flow canvas.

Step 3.1) Create a data asset by selecting “Data Asset” under “Import”.



Import
Data Asset
Pull in data from remote data sources using connections, or pull in data from a file.

Step 3.2) Select the change data asset button, then locate and select your `dataset_loan_example_v2.csv` file.



Projects / lab-wingbank / watsonx-spss-tutorial

Data Asset

Select data from your project. Only eligible data assets are listed.

Find

Categories 2

- Connection
- Data asset

Find

Data assets 1

- dataset_loan_example_v2.csv

Selected assets 1 / 1

All asset details must load before final selection.

dataset_loan_exa...

Type
Data asset

Size
32 KB

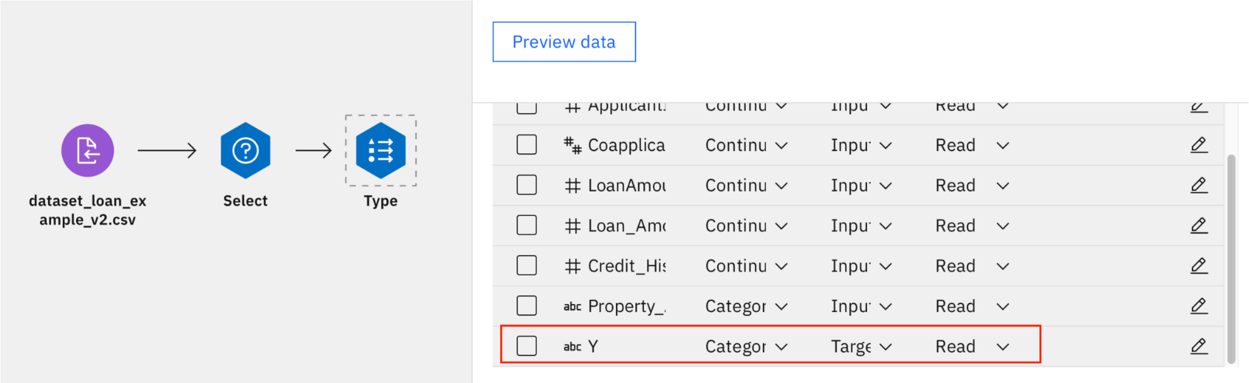
Mime type
text/csv

Created at
2024/10/17 12:54:53

Last updated at
2024/10/17 12:54:53

Cancel Select

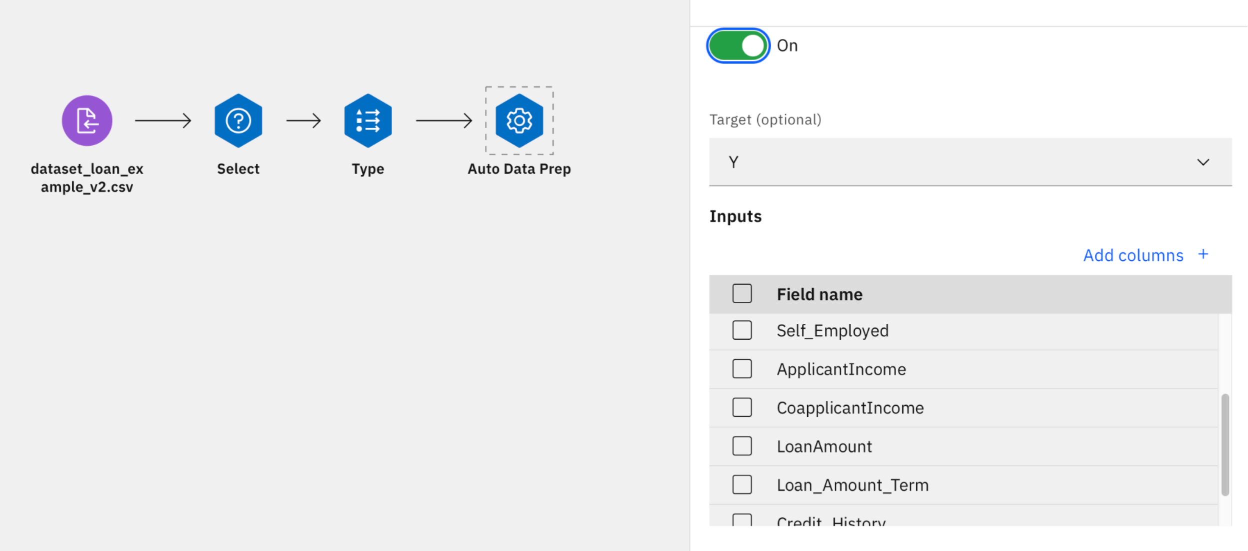
Step 4) Please drag and drop the “Select” icon under “Record Operations” to select the columns you want for Auto Data Preparation. Then drag the type Column and select “Y” as the target variable.



Preview data

□	# Applicant	Continu	Inpu	Read
□	# Coapplica	Continu	Inpu	Read
□	# LoanAmou	Continu	Inpu	Read
□	# Loan_Amo	Continu	Inpu	Read
□	# Credit_His	Continu	Inpu	Read
□	abc Property_	Categor	Inpu	Read
□	abc Y	Categor	Targe	Read

Step 5) Drag and drop “Auto Data Prep” under “Field Operations”.



On

Target (optional)

Y

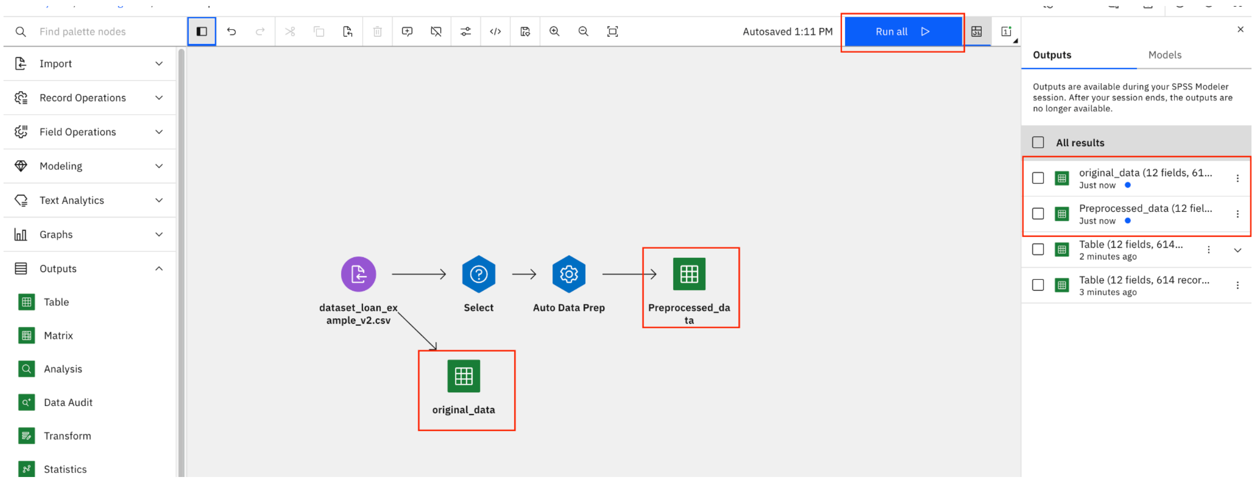
Inputs

Add columns +

□	Field name
□	Self_Employed
□	ApplicantIncome
□	CoapplicantIncome
□	LoanAmount
□	Loan_Amount_Term
□	Credit_History

Step 6.1) Now, we can add tables to visualize the preprocessed and un-preprocessed data. Be sure to rename the tables to avoid confusion

Step 7) Press on run all.

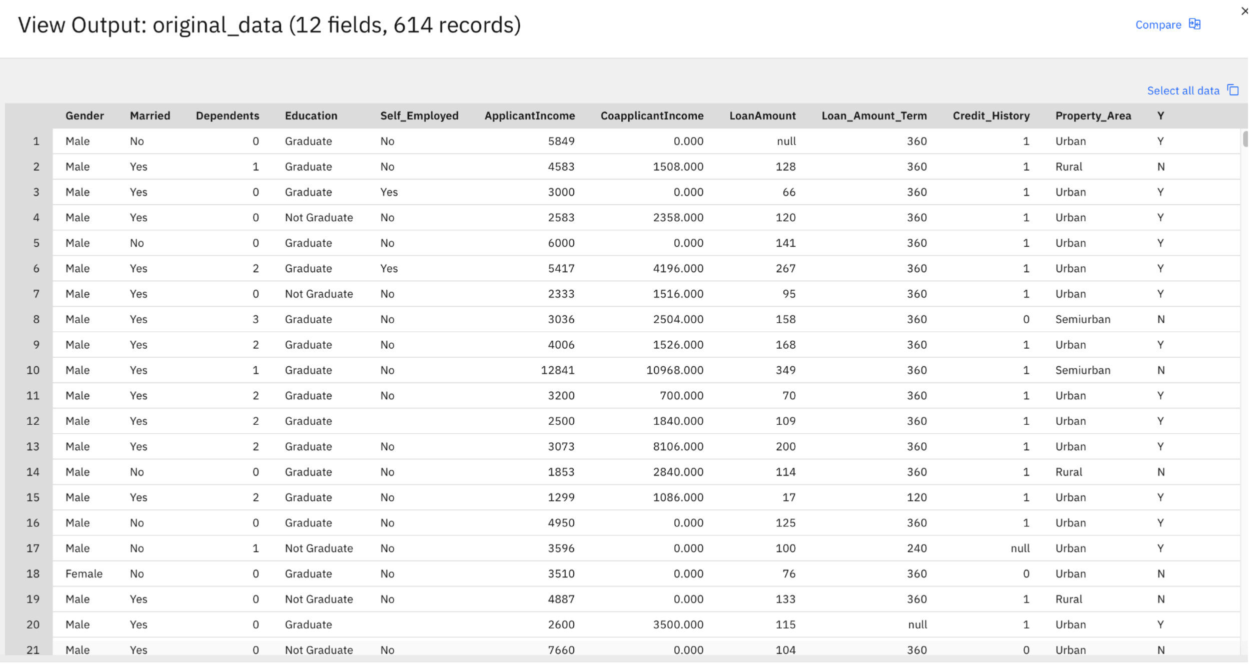


Outputs are available during your SPSS Modeler session. After your session ends, the outputs are no longer available.

□ All results

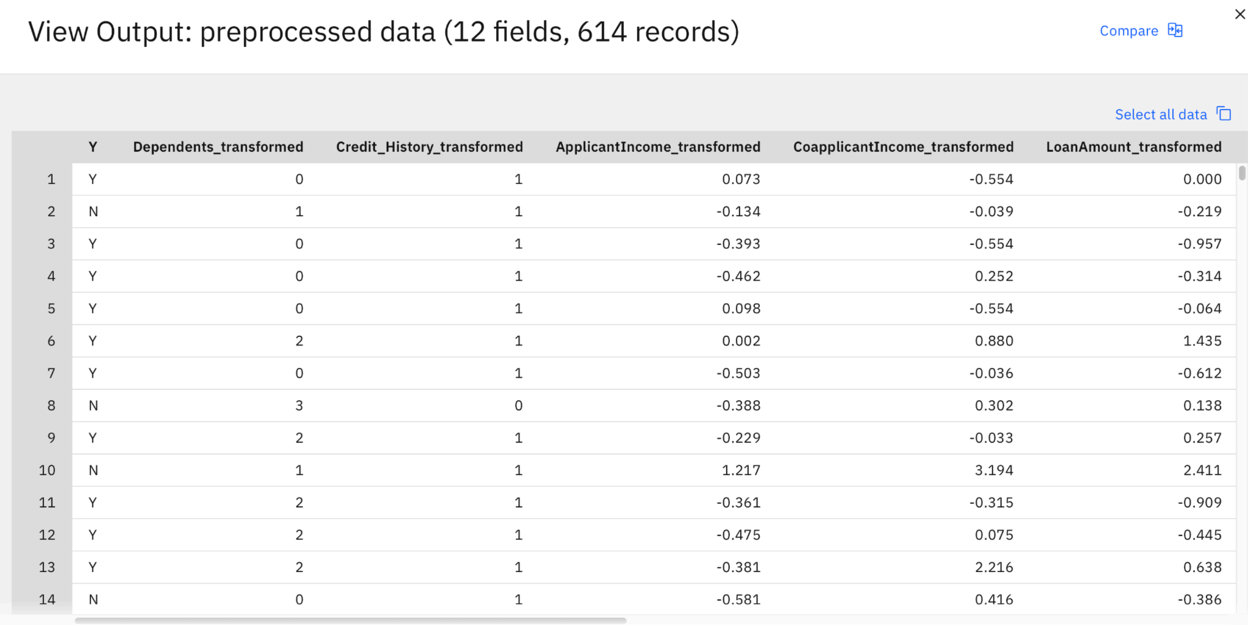
- □ original_data (12 fields, 61... Just now
- □ Preprocessed_data (12 fiel... Just now
- □ Table (12 fields, 614... 2 minutes ago
- □ Table (12 fields, 614 recor... 3 minutes ago

Step 8) Now, you should be able to view the preprocessed data and the origin data tables.



View Output: original_data (12 fields, 614 records)

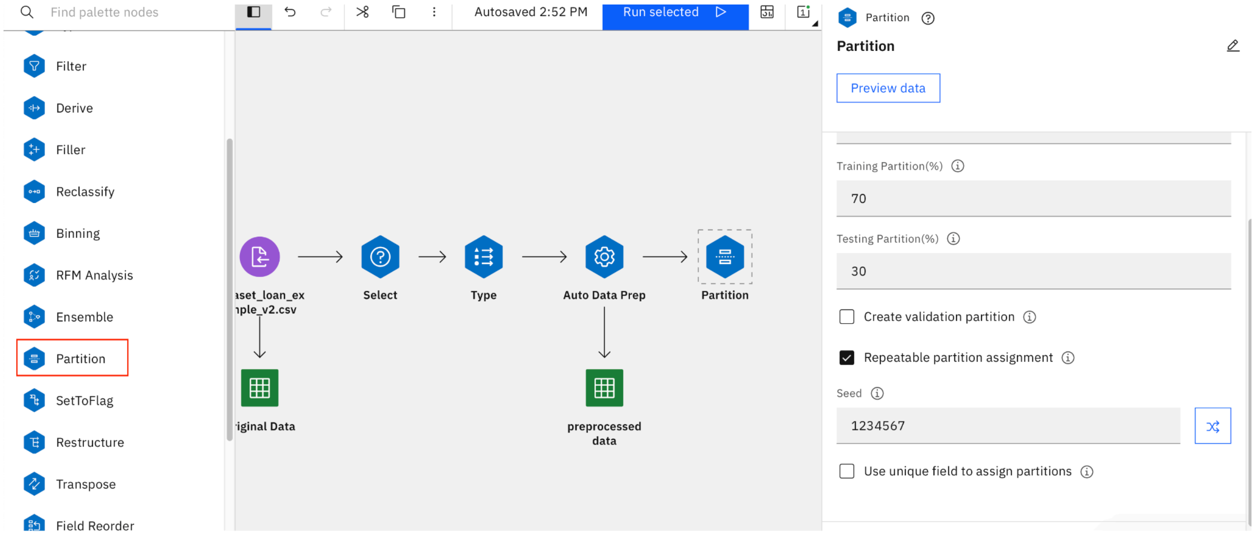
	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area	Y
1	Male	No	0	Graduate	No	5849	0.000	null	360	1	Urban	Y
2	Male	Yes	1	Graduate	No	4583	1508.000	128	360	1	Rural	N
3	Male	Yes	0	Graduate	Yes	3000	0.000	66	360	1	Urban	Y
4	Male	Yes	0	Not Graduate	No	2583	2358.000	120	360	1	Urban	Y
5	Male	No	0	Graduate	No	6000	0.000	141	360	1	Urban	Y
6	Male	Yes	2	Graduate	Yes	5417	4196.000	267	360	1	Urban	Y
7	Male	Yes	0	Not Graduate	No	2333	1516.000	95	360	1	Urban	Y
8	Male	Yes	3	Graduate	No	3036	2504.000	158	360	0	Semiurban	N
9	Male	Yes	2	Graduate	No	4006	1526.000	168	360	1	Urban	Y
10	Male	Yes	1	Graduate	No	12841	10968.000	349	360	1	Semiurban	N
11	Male	Yes	2	Graduate	No	3200	700.000	70	360	1	Urban	Y
12	Male	Yes	2	Graduate		2500	1840.000	109	360	1	Urban	Y
13	Male	Yes	2	Graduate	No	3073	8106.000	200	360	1	Urban	Y
14	Male	No	0	Graduate	No	1853	2840.000	114	360	1	Rural	N
15	Male	Yes	2	Graduate	No	1299	1086.000	17	120	1	Urban	Y
16	Male	No	0	Graduate	No	4950	0.000	125	360	1	Urban	Y
17	Male	No	1	Not Graduate	No	3596	0.000	100	240	null	Urban	Y
18	Female	No	0	Graduate	No	3510	0.000	76	360	0	Urban	N
19	Male	Yes	0	Not Graduate	No	4887	0.000	133	360	1	Rural	N
20	Male	Yes	0	Graduate		2600	3500.000	115	null	1	Urban	Y
21	Male	Yes	0	Not Graduate	No	7660	0.000	104	360	0	Urban	N



View Output: preprocessed data (12 fields, 614 records)

	Y	Dependents_transformed	Credit_History_transformed	ApplicantIncome_transformed	CoapplicantIncome_transformed	LoanAmount_transformed
1	Y	0	1	0.073	-0.554	0.000
2	N	1	1	-0.134	-0.039	-0.219
3	Y	0	1	-0.393	-0.554	-0.957
4	Y	0	1	-0.462	0.252	-0.314
5	Y	0	1	0.098	-0.554	-0.064
6	Y	2	1	0.002	0.880	1.435
7	Y	0	1	-0.503	-0.036	-0.612
8	N	3	0	-0.388	0.302	0.138
9	Y	2	1	-0.229	-0.033	0.257
10	N	1	1	1.217	3.194	2.411
11	Y	2	1	-0.361	-0.315	-0.909
12	Y	2	1	-0.475	0.075	-0.445
13	Y	2	1	-0.381	2.216	0.638
14	N	0	1	-0.581	0.416	-0.386

Step 9: Next we want to partition the data in order to split between train and test.



Partition

Partition

Preview data

Training Partition(%)

70

Testing Partition(%)

30

☐ Create validation partition

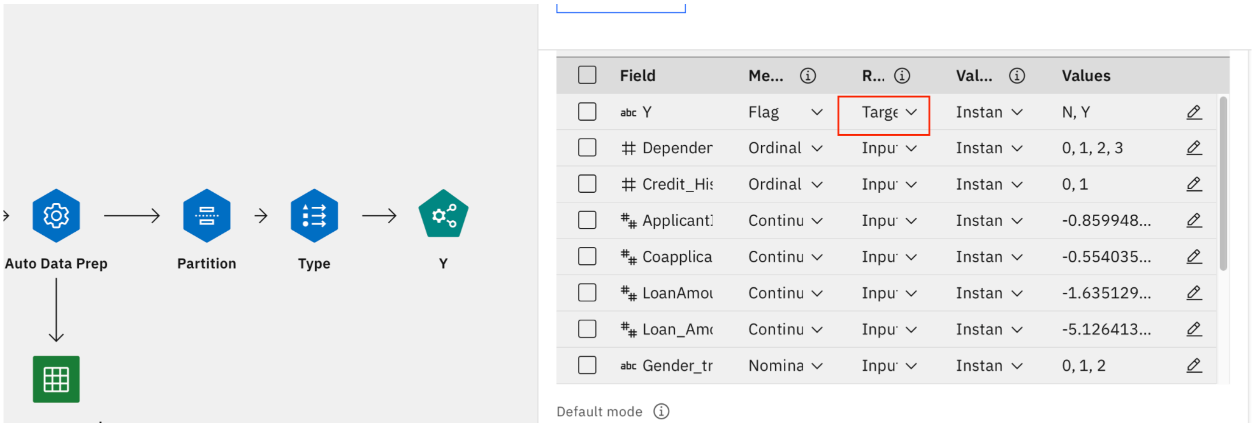
☑ Repeatable partition assignment

Seed

1234567

☐ Use unique field to assign partitions

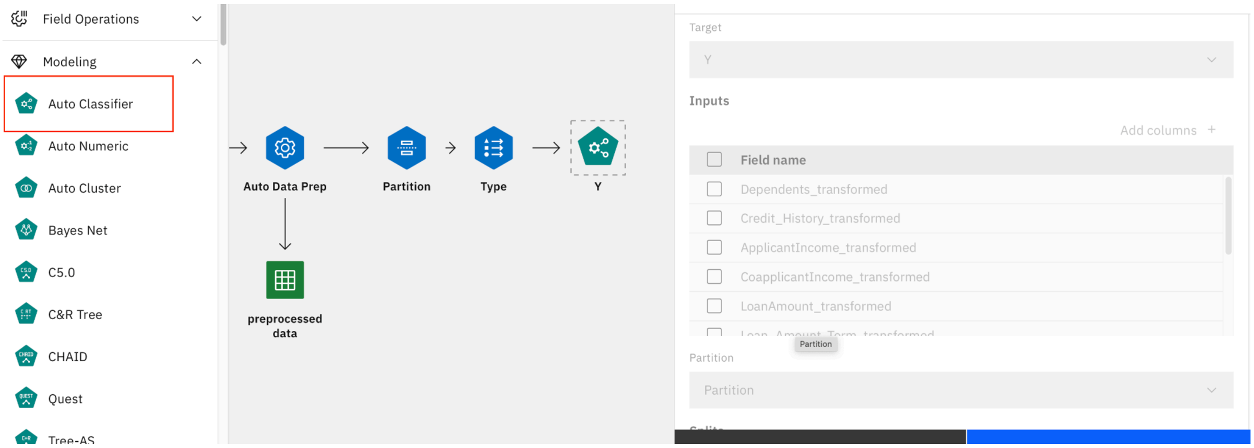
Step 10) Double tap on the “Type” icon, then scroll down the column fields until you see the “Y” column field. Here, please change the field type, the third parameter, from “Input” to “Target”.



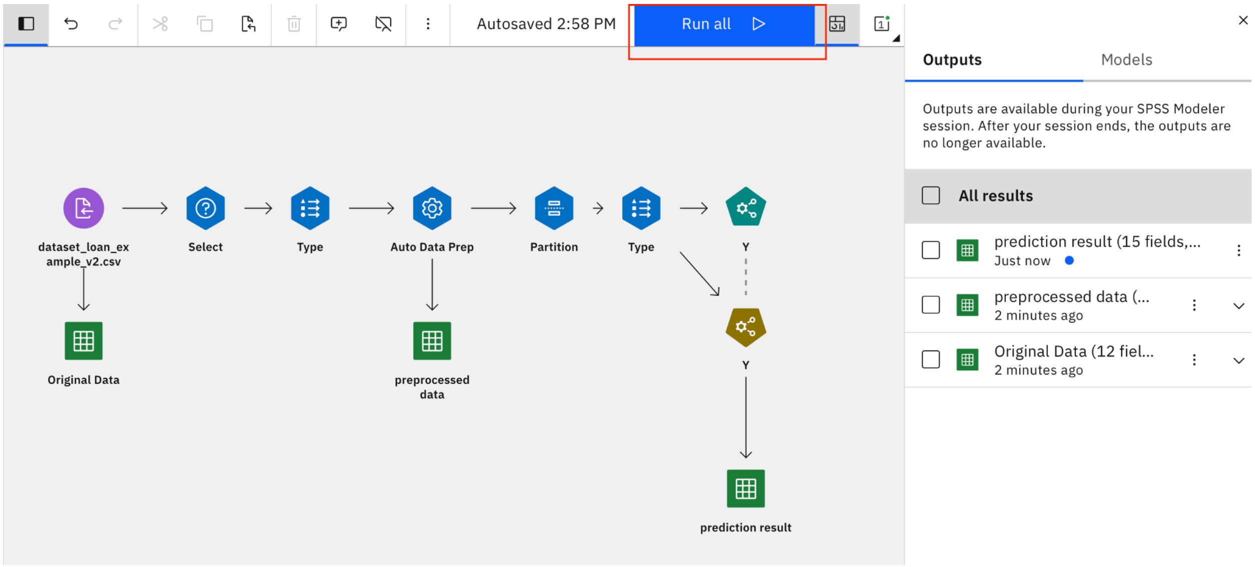
☐	Field	Me...	R...	Val...	Values
☐	abc Y	Flag	Targe	Instan	N, Y
☐	# Depender	Ordinal	Inpu	Instan	0, 1, 2, 3
☐	# Credit_His	Ordinal	Inpu	Instan	0, 1
☐	# Applicant	Continu	Inpu	Instan	-0.859948...
☐	# Coapplica	Continu	Inpu	Instan	-0.554035...
☐	# LoanAmou	Continu	Inpu	Instan	-1.635129...
☐	# Loan_Amo	Continu	Inpu	Instan	-5.126413...
☐	abc Gender_tr	Nomina	Inpu	Instan	0, 1, 2

Default mode

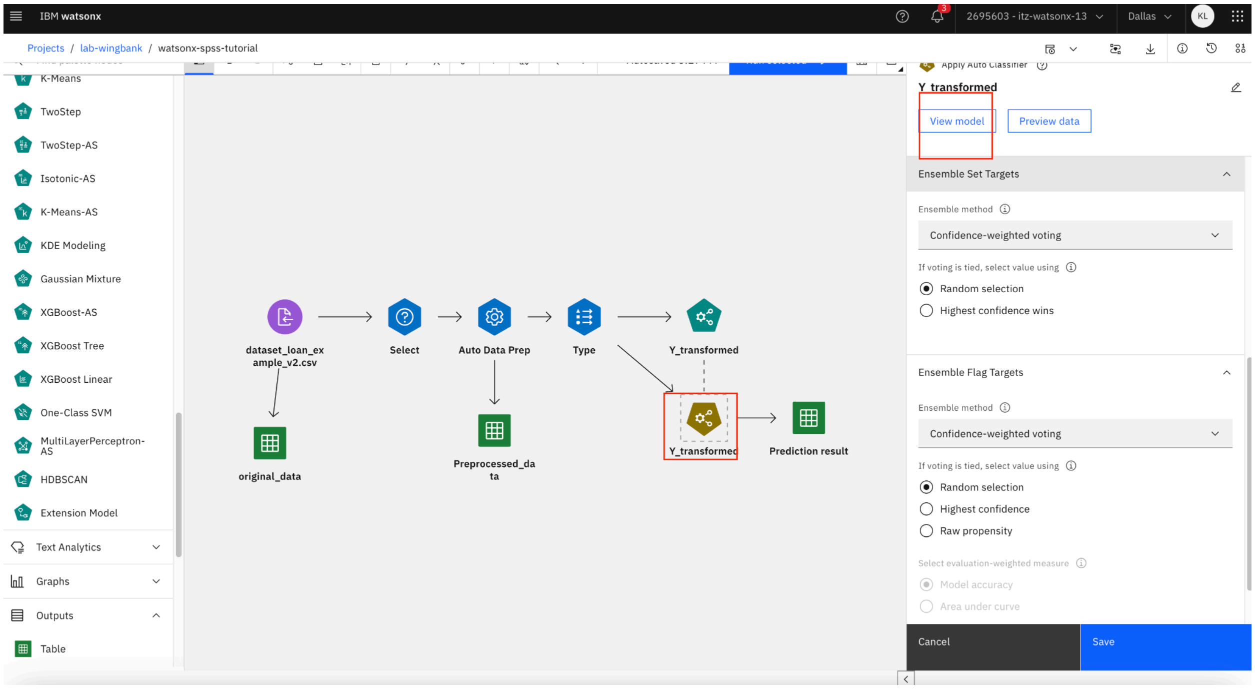
Step 11) Next we will be creating an Auto AI Classifier. Drag and drop and a line between auto data prep and auto classifier. The “Type” icon will automatically be created if not already added.



Step 12) Be sure to add a table output called "prediction result". Here we run all again, to run the auto ai classifier


Autosaved 2:58 PM
Run all
Outputs
Models
Outputs are available during your SPSS Modeler session. After your session ends, the outputs are no longer available.
All results
prediction result (15 fields,...
Just now
preprocessed data (...
2 minutes ago
Original Data (12 fiel...
2 minutes ago
dataset_loan_example_v2.csv
Select
Type
Auto Data Prep
Partition
Type
Y
Original Data
preprocessed data
Y
prediction result

Step 13) Now we can view the auto AI pipeline by clicking View Mode.



Ensemble Set Targets

Ensemble method: Confidence-weighted voting

If voting is tied, select value using

- ● Random selection
- ○ Highest confidence wins

Ensemble Flag Targets

Ensemble method: Confidence-weighted voting

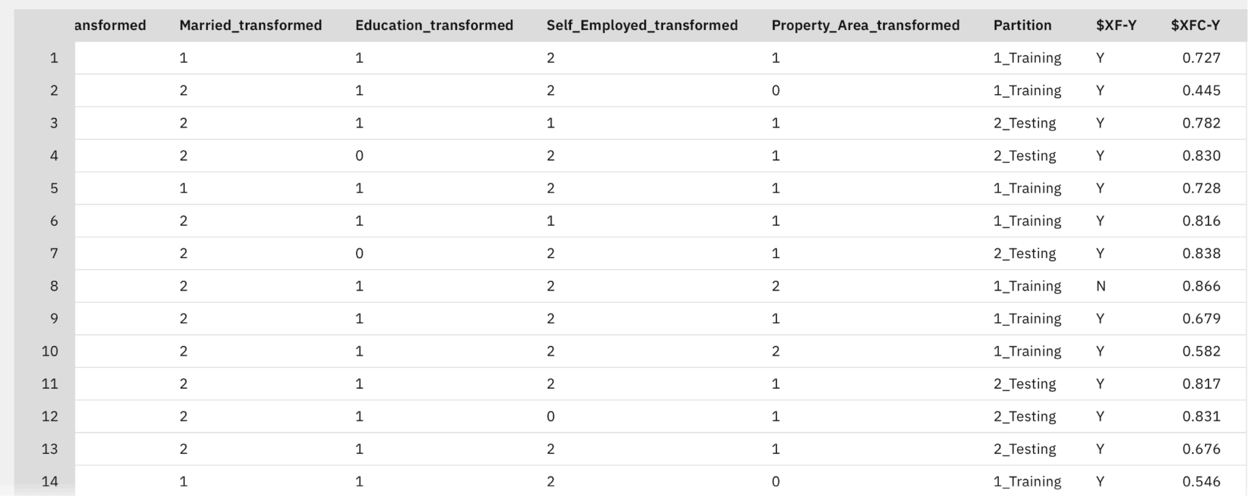
If voting is tied, select value using

- ● Random selection
- ○ Highest confidence
- ○ Raw propensity

Select evaluation-weighted measure

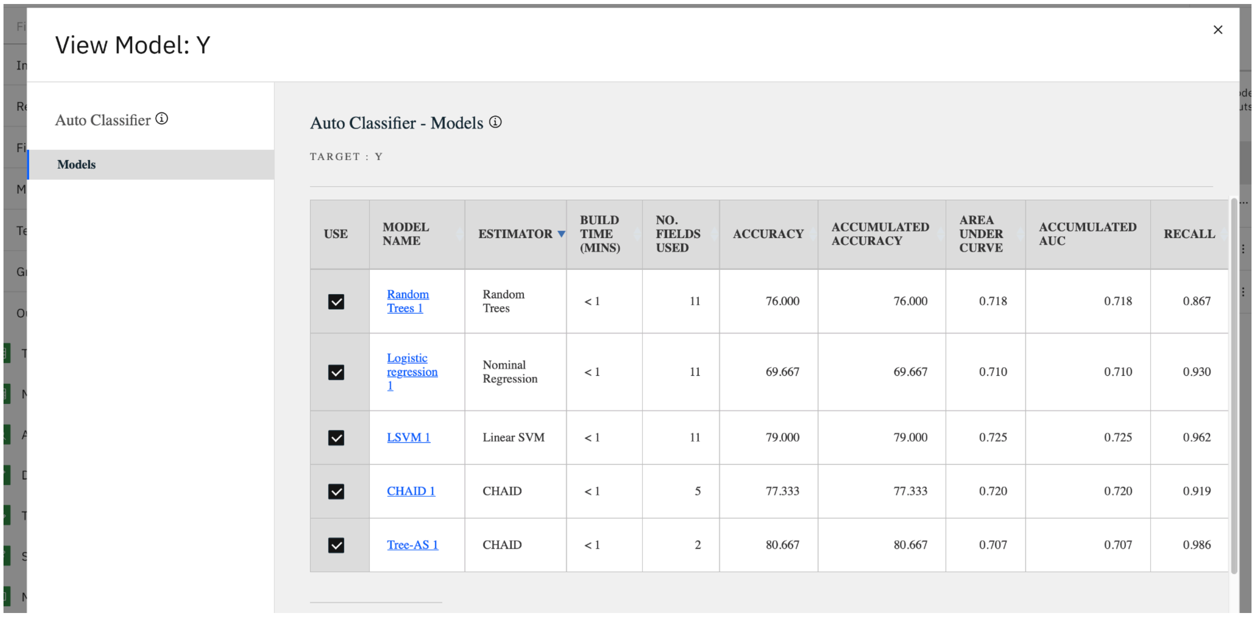
- ○ Model accuracy
- ○ Area under curve

Step 14) View the Prediction result table for the ensemble model (Note we can also view individual predictions of each model by turning “Filter out fields generated by ensemble models to be false).



	ansformed	Married_transformed	Education_transformed	Self_Employed_transformed	Property_Area_transformed	Partition	$XF-Y	$XFC-Y
1		1	1	2	1	1_Training	Y	0.727
2		2	1	2	0	1_Training	Y	0.445
3		2	1	1	1	2_Testing	Y	0.782
4		2	0	2	1	2_Testing	Y	0.830
5		1	1	2	1	1_Training	Y	0.728
6		2	1	1	1	1_Training	Y	0.816
7		2	0	2	1	2_Testing	Y	0.838
8		2	1	2	2	1_Training	N	0.866
9		2	1	2	1	1_Training	Y	0.679
10		2	1	2	2	1_Training	Y	0.582
11		2	1	2	1	2_Testing	Y	0.817
12		2	1	0	1	2_Testing	Y	0.831
13		2	1	2	1	2_Testing	Y	0.676
14		1	1	2	0	1_Training	Y	0.546

The model trained and ranked.

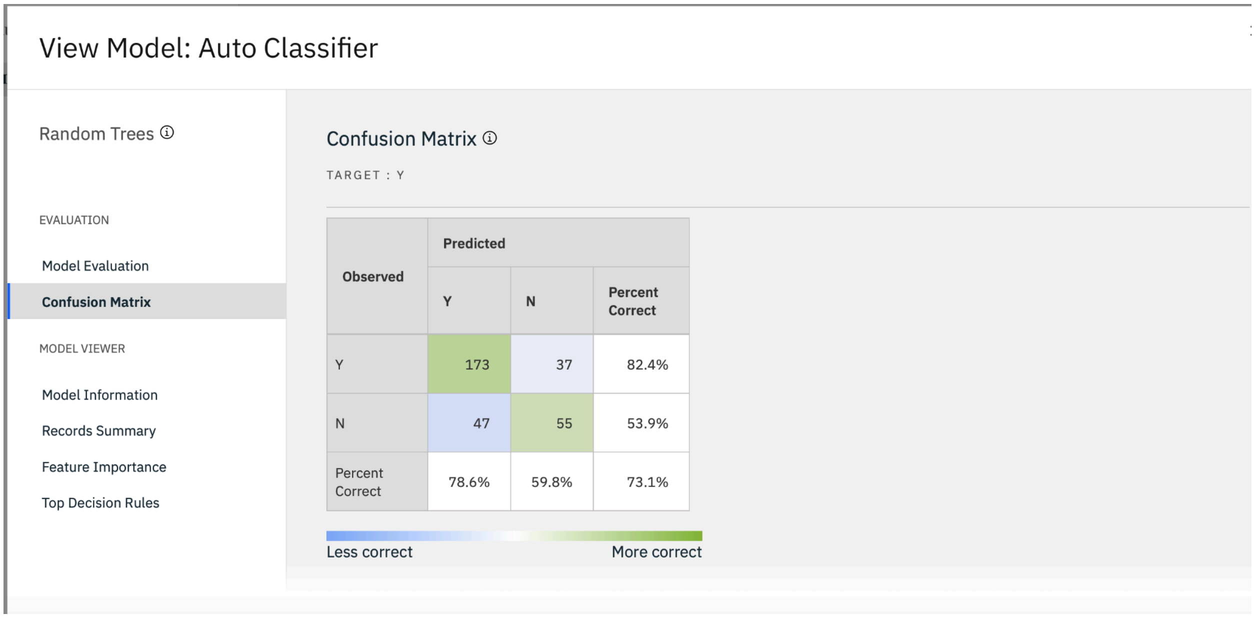


Auto Classifier - Models ⓘ

TARGET : Y

USE	MODEL NAME	ESTIMATOR	BUILD TIME (MINS)	NO. FIELDS USED	ACCURACY	ACCUMULATED ACCURACY	AREA UNDER CURVE	ACCUMULATED AUC	RECALL
☑	Random Trees 1	Random Trees	< 1	11	76.000	76.000	0.718	0.718	0.867
☑	Logistic regression 1	Nominal Regression	< 1	11	69.667	69.667	0.710	0.710	0.930
☑	LSVM 1	Linear SVM	< 1	11	79.000	79.000	0.725	0.725	0.962
☑	CHAID 1	CHAID	< 1	5	77.333	77.333	0.720	0.720	0.919
☑	Tree-AS 1	CHAID	< 1	2	80.667	80.667	0.707	0.707	0.986

Finally we can click into each model to see the performance of the model:



View Model: Auto Classifier

Confusion Matrix ⓘ

TARGET : Y

Observed	Predicted		
	Y	N	Percent Correct
Y	173	37	82.4%
N	47	55	53.9%
Percent Correct	78.6%	59.8%	73.1%