Team 3 [Green] Phase 3

Srushtiben Kankotiya

Group 3 [Green] Phase 3 Target

- The goal is to final test the model that we chose in phase 2.
- Run the model to find the churned (50%) customers from the sample set provided.

Models finalized for the final phase

- Deep Learning
  - By using deep learning model, accuracy achieved was 95.73%.
- Generalized linear Model
  - Achieved an accuracy of 90.58%
- Logistic Regression Model
  - Fulfilled an accuracy of 90.43%

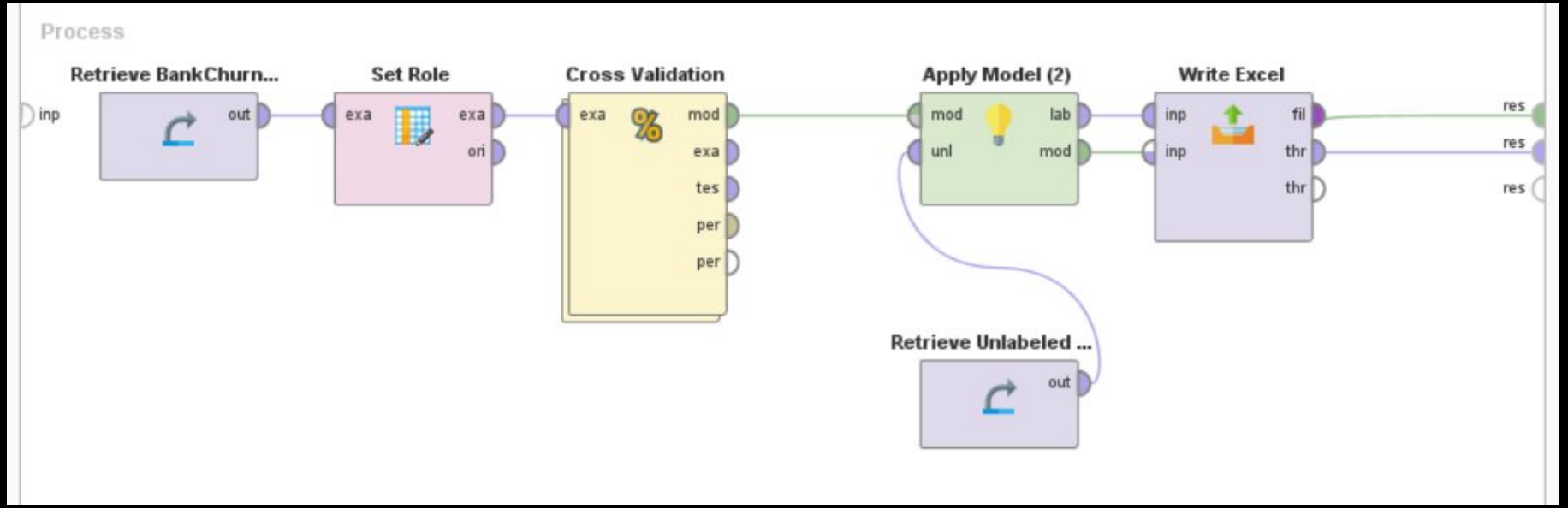
Generalized linear Model

The term "general" linear model (GLM) usually refers to conventional linear regression models for a continuous response variable given continuous and/or categorical predictors.

Selected Attributes

1. Average Utilisation Ratio
2. Education Level
3. Total Amount Chng Q4 to Q1
4. Total Ct Chng Q4 to Q1
5. Card Category
6. Income Category
7. ~~Credit Limit~~
8. Total Revol balance

How the Process Works?



Steps

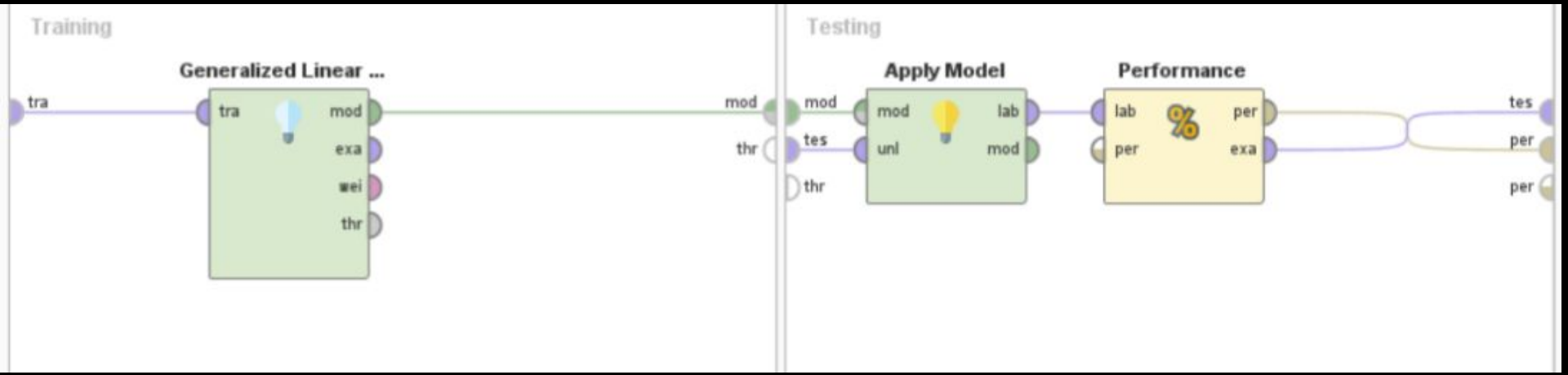
Step1- Importing of Data and setting role

The Data was imported from the Local Repository then, we associate it to set role

to indicate the job for the extraordinary attribute "attrition flag" which it "Label"

target role. The Cross Validation Operator was used.

How the model is Trained and Tested?



Steps (Cont.)

Step2- Training and Testing

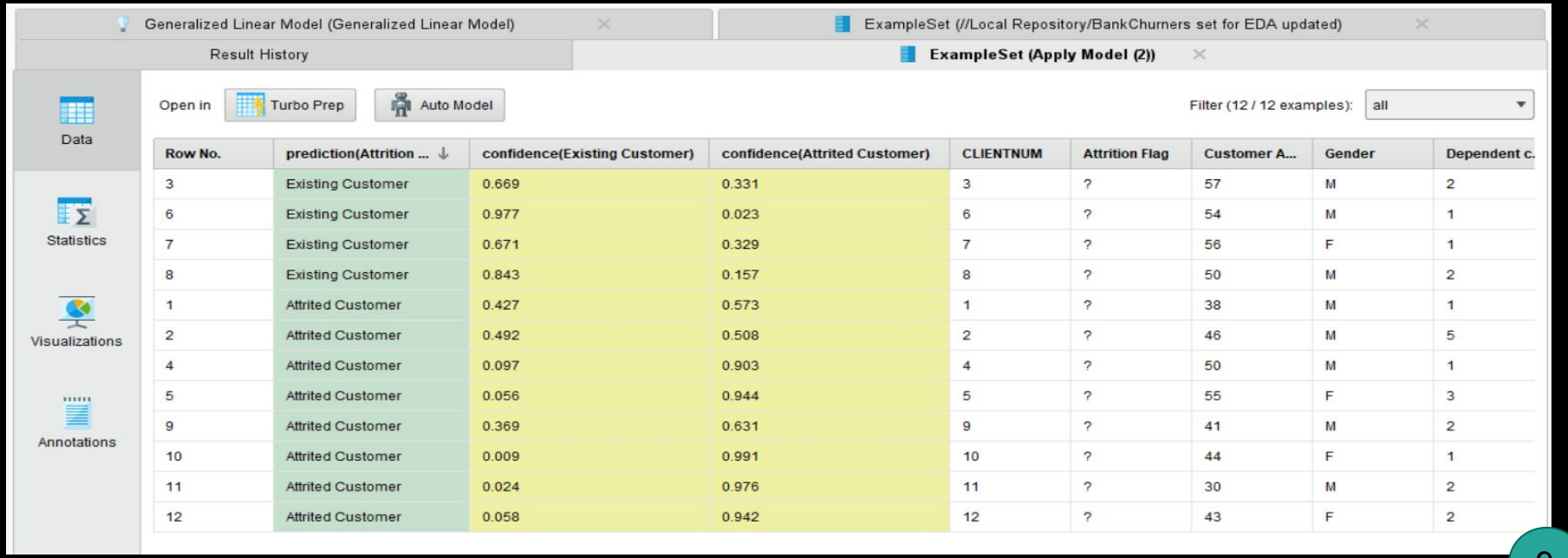
- In the cross validation the the generalized linear model operator was used for training purpose and for testing the into model we associated model to the apply model , In addition to it the apply model operator

- Subsequent to interfacing with apply model named information shared to the Performance operator. A model is first prepared on an ExampleSet by another Operator, which is much of the time a learning calculation.

- The Write Excel operator is used for writing ExampleSets into a Microsoft Excel spreadsheet file. Which gives us expected output.

Results

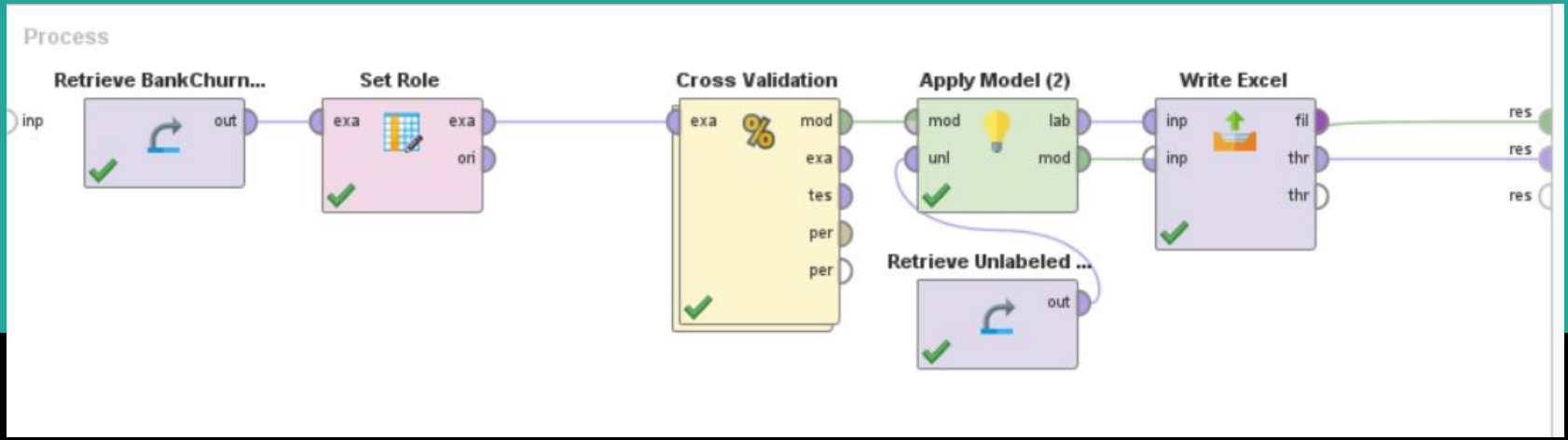
- After retraining with the generalized linear model it was found that the there were 8 Churned customers and 4 Existing customer. The proportion (67/33)



Row No.	prediction(Attrition ...	confidence(Existing Customer)	confidence(Attrited Customer)	CLIENTNUM	Attrition Flag	Customer A...	Gender	Dependent c.
3	Existing Customer	0.669	0.331	3	?	57	M	2
6	Existing Customer	0.977	0.023	6	?	54	M	1
7	Existing Customer	0.671	0.329	7	?	56	F	1
8	Existing Customer	0.843	0.157	8	?	50	M	2
1	Attrited Customer	0.427	0.573	1	?	38	M	1
2	Attrited Customer	0.492	0.508	2	?	46	M	5
4	Attrited Customer	0.097	0.903	4	?	50	M	1
5	Attrited Customer	0.056	0.944	5	?	55	F	3
9	Attrited Customer	0.369	0.631	9	?	41	M	2
10	Attrited Customer	0.009	0.991	10	?	44	F	1
11	Attrited Customer	0.024	0.976	11	?	30	M	2
12	Attrited Customer	0.058	0.942	12	?	43	F	2

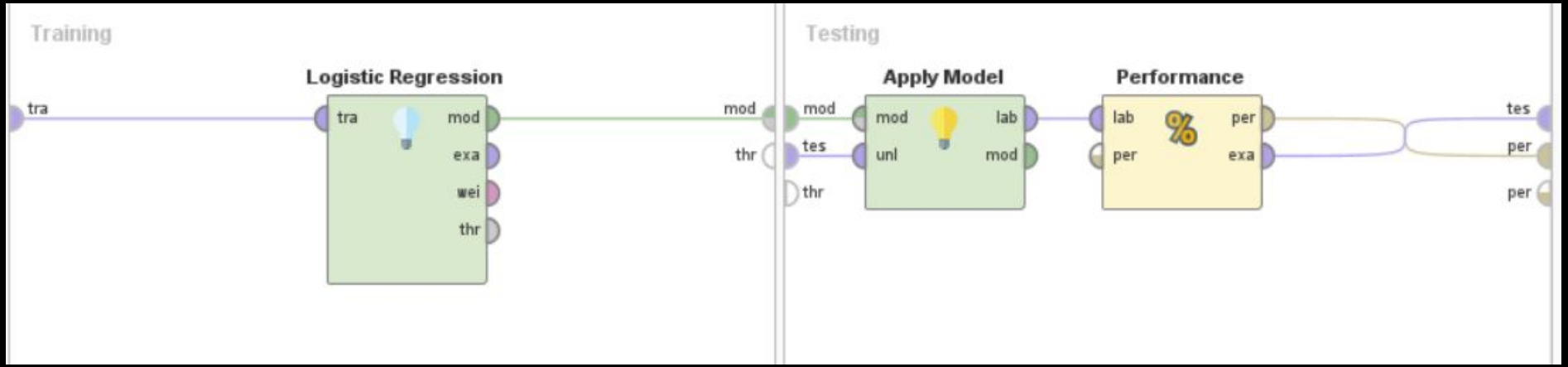
Logistic regression model

- Logistic Regression is a Supervised machine learning algorithm that can be used to model the probability of a certain class or event.
- It is used when the data is linearly separable and the outcome is binary.



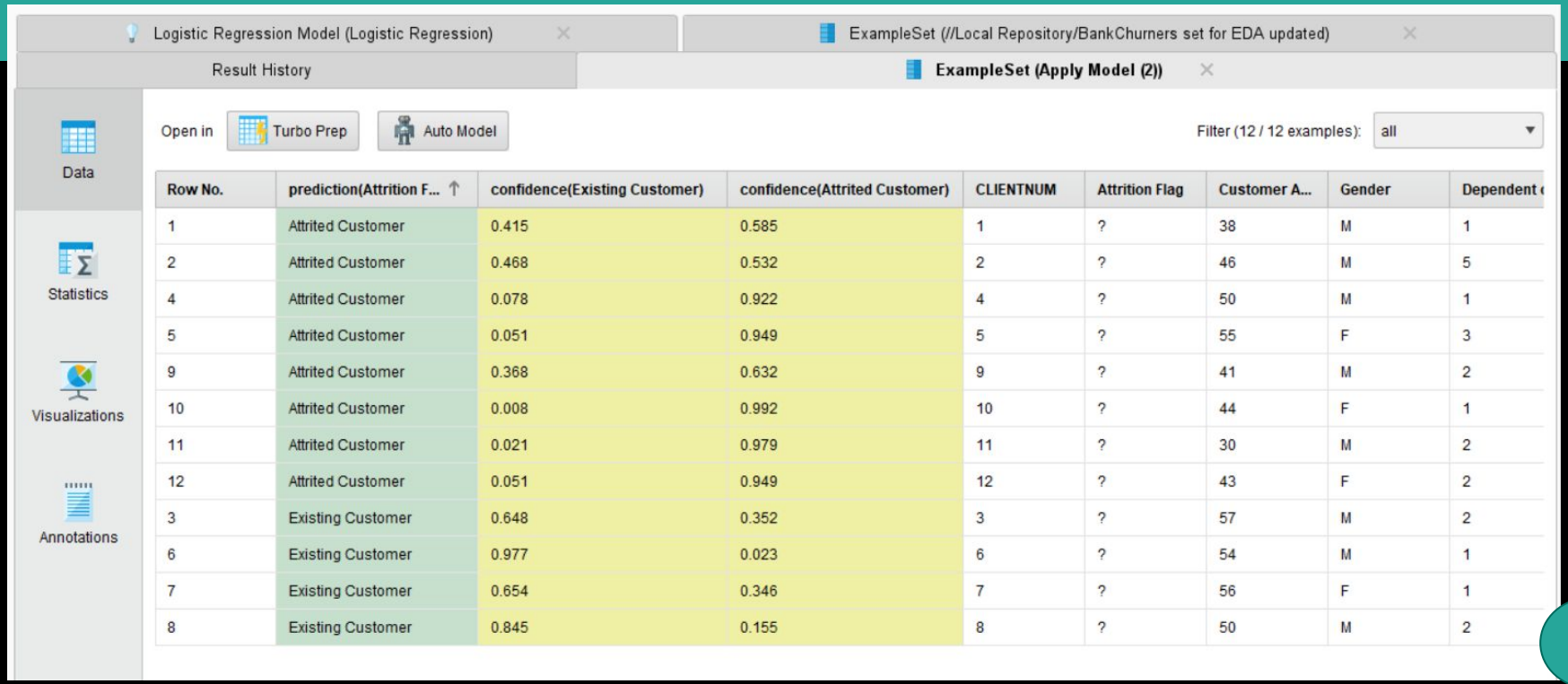
1. We imported the data set to the local repository.
2. Drag the data into the process.
3. Aim is to predict the target / dependent variable using the other variables. The target variable should be binomial, there must be only two possibilities.
4. In next step, for better results, we have used cross validation to divide the data.
5. After the cross validation we have connected the output with the model.
6. The Write Excel operator is used for writing ExampleSets into a Microsoft Excel spreadsheet file. The connected "fil" of Excel operator to "Results" instead of "thr" to "Res". Which gives us the expected output.

- After processing the data into model we connected model to the apply model.
- A model is first trained on an ExampleSet by another Operator. Afterwards, this model can be applied on another ExampleSet. Usually, the goal is to get a prediction on unseen data or to transform data by applying a preprocessing model.



Results

- After retraining with the Logistic Regression model it was found that the there were 8 Attrited customers and 4 Existing customer. the proportion (66.66/33.33)%.



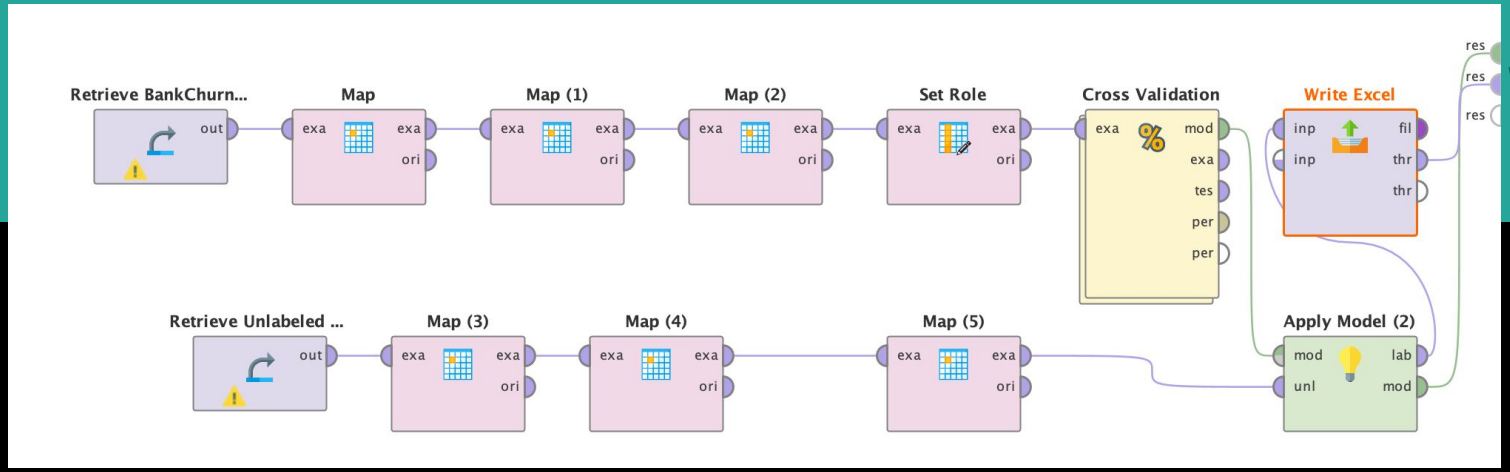
Row No.	prediction(Attrition F...	confidence(Existing Customer)	confidence(Attrited Customer)	CLIENTNUM	Attrition Flag	Customer A...	Gender	Dependent
1	Attrited Customer	0.415	0.585	1	?	38	M	1
2	Attrited Customer	0.468	0.532	2	?	46	M	5
4	Attrited Customer	0.078	0.922	4	?	50	M	1
5	Attrited Customer	0.051	0.949	5	?	55	F	3
9	Attrited Customer	0.368	0.632	9	?	41	M	2
10	Attrited Customer	0.008	0.992	10	?	44	F	1
11	Attrited Customer	0.021	0.979	11	?	30	M	2
12	Attrited Customer	0.051	0.949	12	?	43	F	2
3	Existing Customer	0.648	0.352	3	?	57	M	2
6	Existing Customer	0.977	0.023	6	?	54	M	1
7	Existing Customer	0.654	0.346	7	?	56	F	1
8	Existing Customer	0.845	0.155	8	?	50	M	2

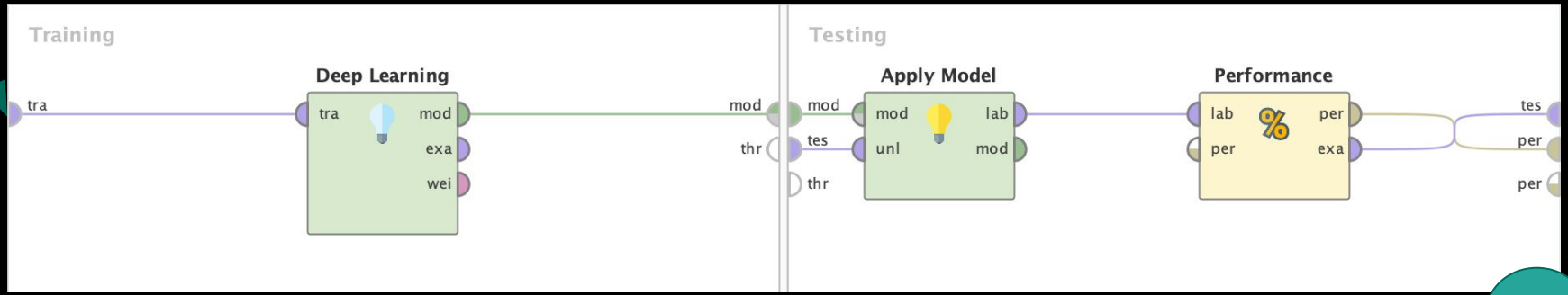
Deep Learning

Deep learning is a Machine learning technique that teaches computers to do what humans does naturally i.e. the algorithms are inspired by the structure of the human brain and are known as neural networks. The model learns to perform classification tasks directly from images, text or sound.

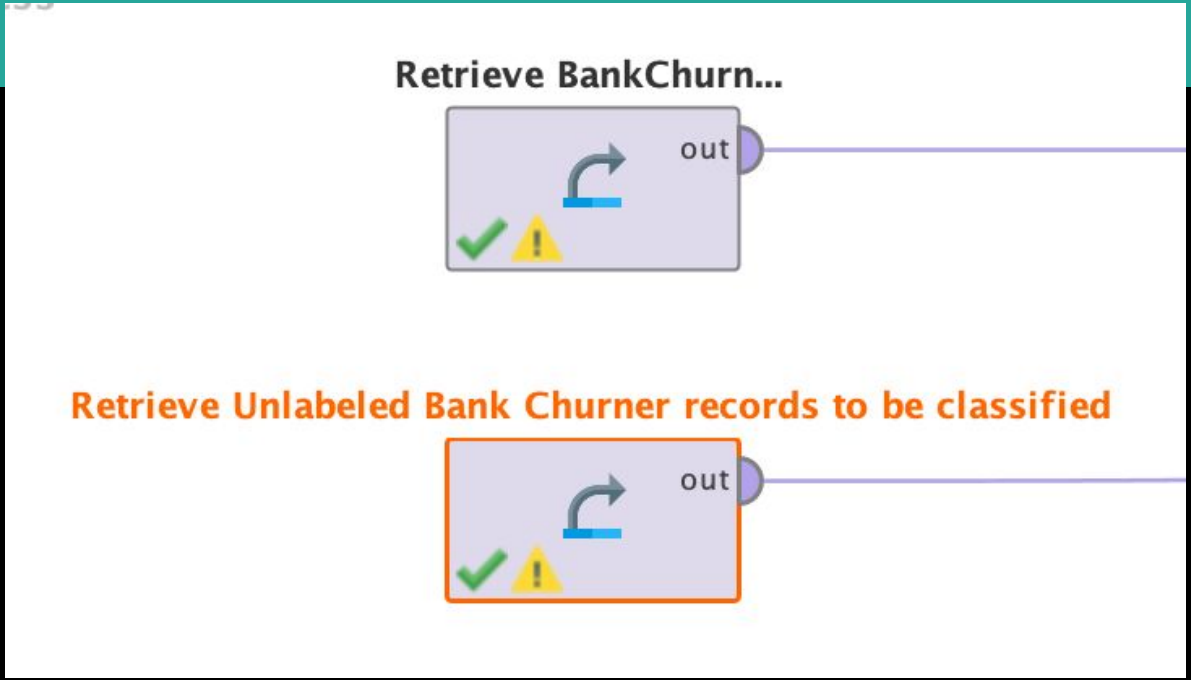
The neural networks are built from interconnected network switches designed to learn to recognize patterns in the same way the human brain and nervous system does.

Working on the Model





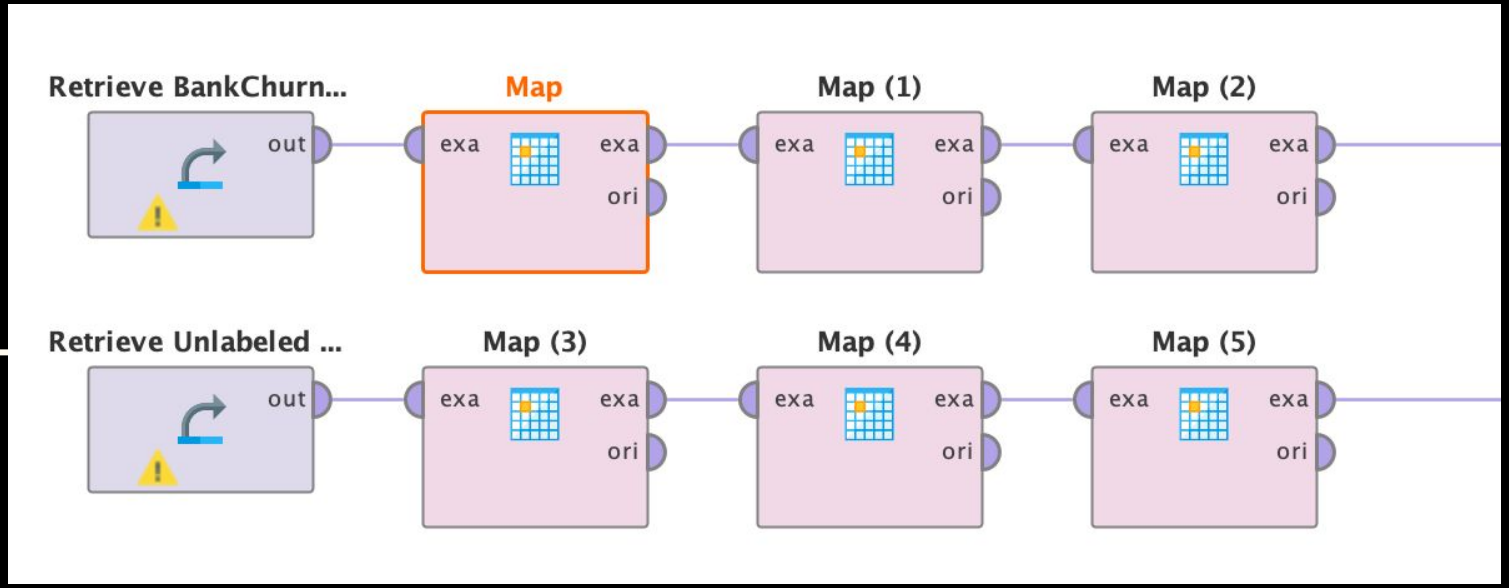
Retrieve Data



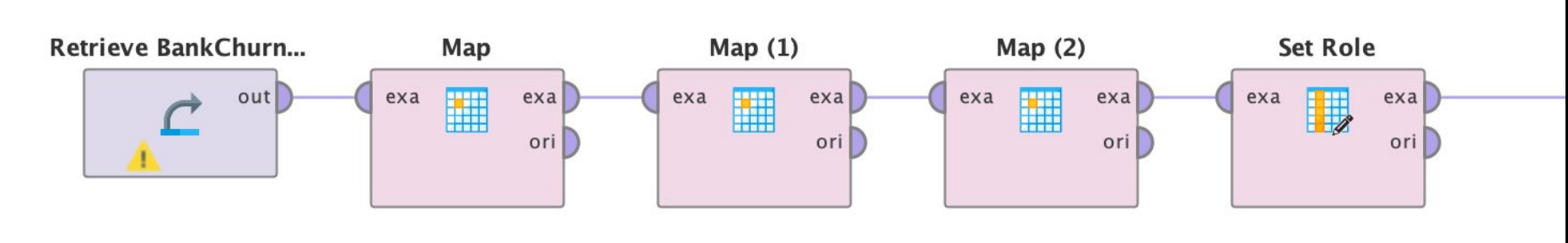
Selected Attributes

1. Average Utilisation Ratio
2. Education Level
3. Total Amnt Chng Q4 to Q1
4. Total Ct Chng Q4 to Q1
5. Card Category
6. Income Category
7. Credit Limit
8. Total Revol balance
9. Total Trans Amt
10. Total Trans Ct
11. Months on book

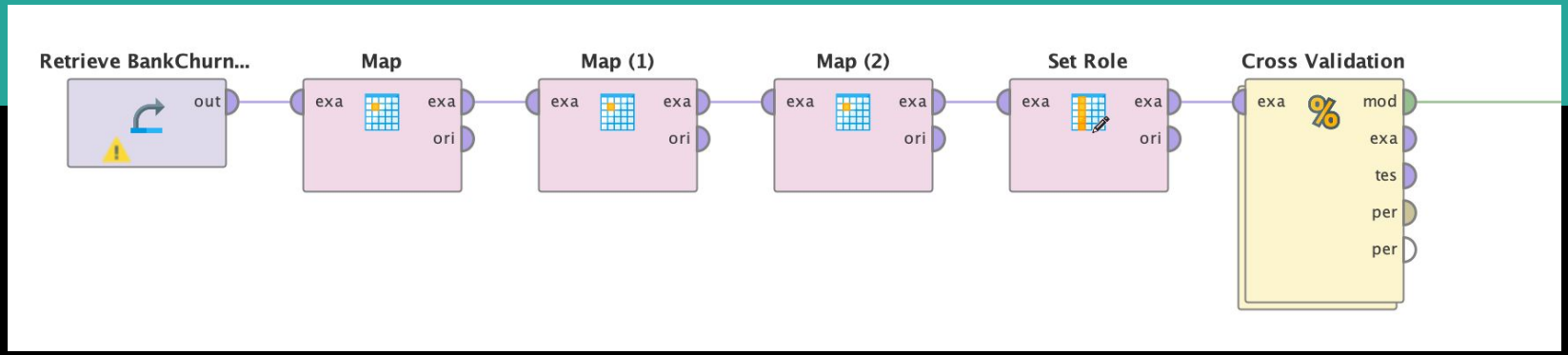
Replace Unknown Value

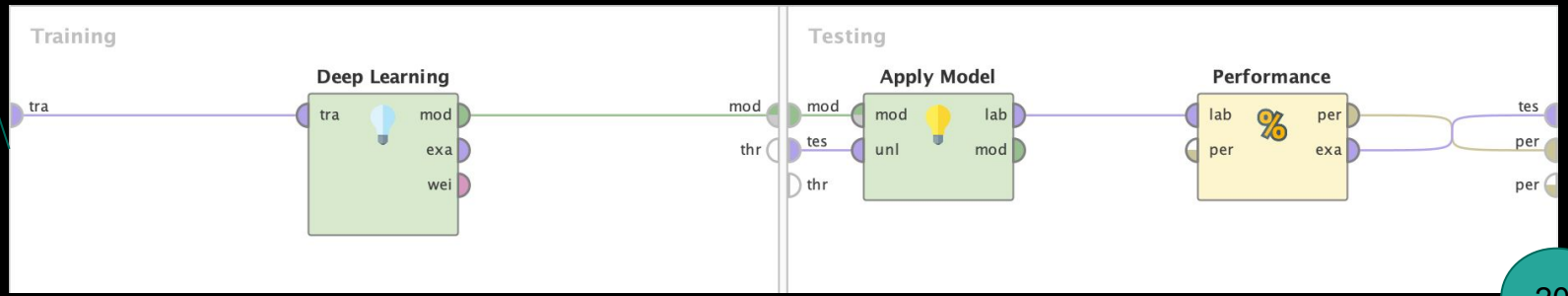


Set Role

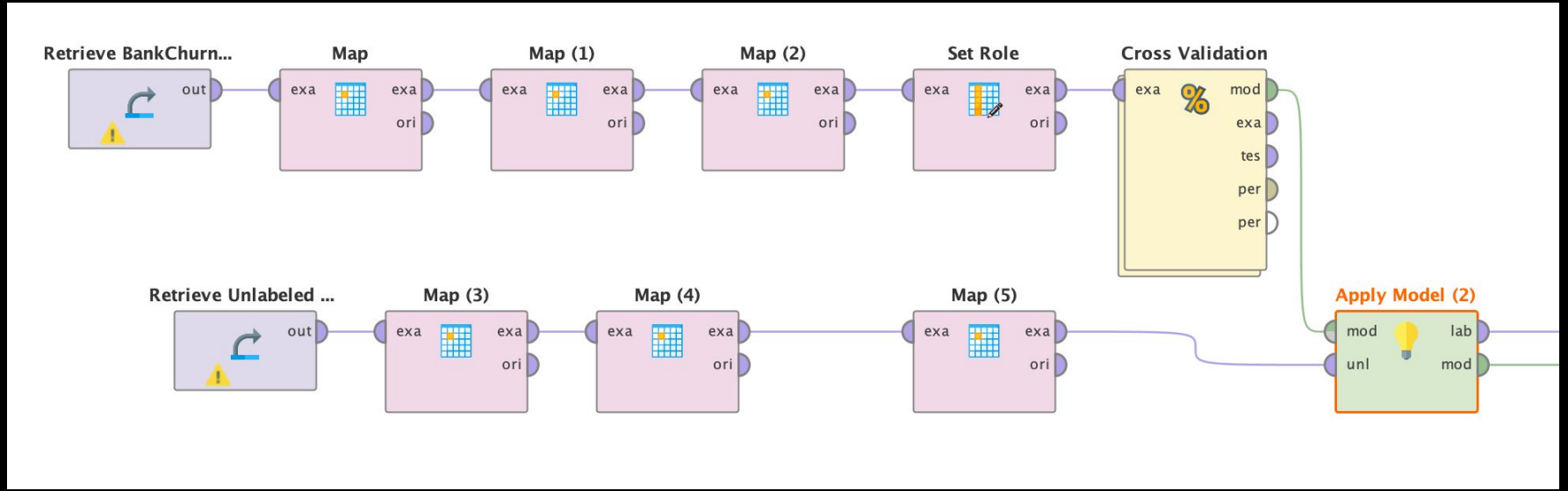

Retrieve BankChurn...
out
Map
exa
exa
ori
Map (1)
exa
exa
ori
Map (2)
exa
exa
ori
Set Role
exa
exa
ori

Cross Validation

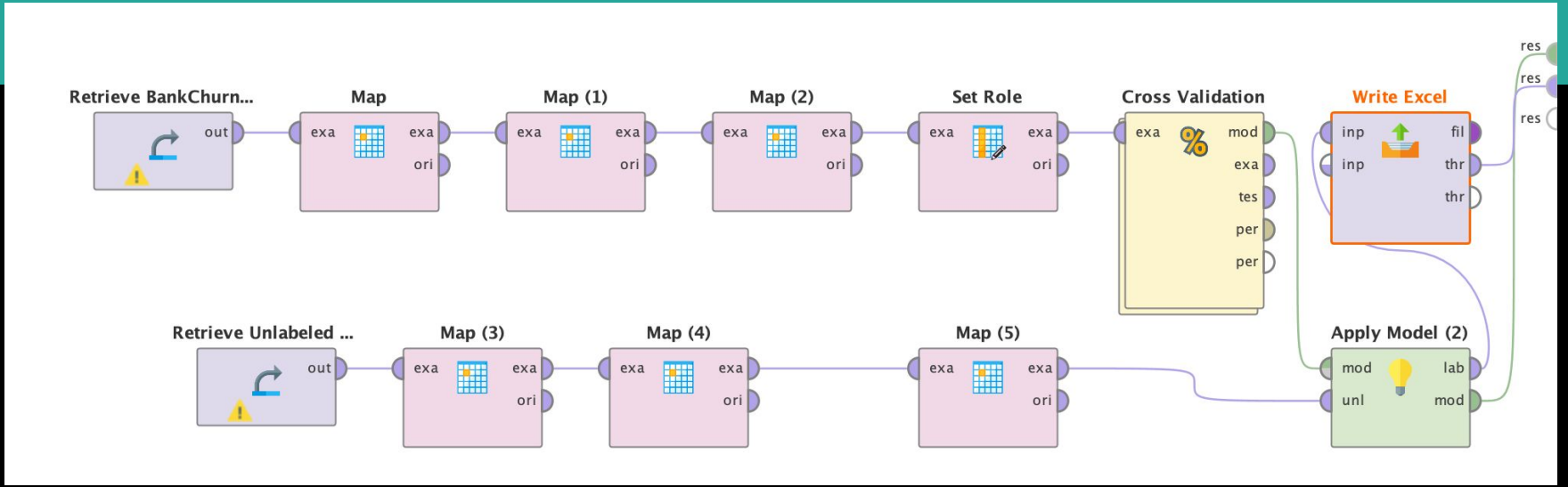




Apply Model to the sample set



Write Excel



The Write Excel operator used for writing ExampleSets into a Microsoft Excel spreadsheet file.

Result set

Row No.	prediction(Attrition Flag)	confidence(Existing Customer)	confidence(Attrited Customer)	Marital Status	Income Category	Education Level	Credit Limit	Total Trans Amt
1	Existing Customer	0.996	0.004	Single	$40K - $60K	Uneducated	5639	16628
2	Existing Customer	0.998	0.002	Single	$80K - $120K	College	13187	15354
3	Existing Customer	0.996	0.004	Married	$80K - $120K	Graduate	17925	17498
4	Attrited Customer	0.013	0.987	Married	$80K - $120K	Graduate	9959	10310
5	Attrited Customer	0.105	0.895	Single	Less Than $40k	Uneducated	14657	6009
6	Existing Customer	0.999	0.001	Single	$60K - $80K	High School	13940	15577
7	Existing Customer	0.995	0.005	Single	Less than $40K	Graduate	3688	14596
8	Existing Customer	0.998	0.002	Single	$40K - $60K	Graduate	4003	15476
9	Attrited Customer	0.217	0.783	Divorced	$40K - $60K	Graduate	4277	8764
10	Attrited Customer	0.007	0.993	Married	Less than $40K	High School	5409	10291
11	Attrited Customer	0.004	0.996	Married	$40K - $60K	Graduate	5281	8395
12	Attrited Customer	0.101	0.899	Married	Less than $40K	Graduate	10388	10294

There was the unknown values in pre-prediction testing set for column Marital Status, Income Category and Education Level and their values replaced with the max occurred value in it. Prediction of the attrition flag predicted with good confidence level.

Result set on Excel

Credit Limi	Total Revol	Avg Open T	Total Amt (	Total Trans	Total Trans	Total Ct Ch	Avg Utilizat	confidence(Existing Customer)	confidence(Attrited Customer)	prediction(Attrition Flag)
5639.0	1558.0	4081.0	.6	16628.0	109.0	.9	.3	1.0	.0	Existing Customer
13187.0	2241.0	10946.0	.7	15354.0	112.0	.9	.2	1.0	.0	Existing Customer
17925.0	1909.0	16016.0	.7	17498.0	111.0	.8	.1	1.0	.0	Existing Customer
9959.0	952.0	9007.0	.8	10310.0	63.0	1.1	.1	.0	1.0	Attrited Customer
14657.0	2517.0	12140.0	.2	6009.0	53.0	.5	.2	.1	.9	Attrited Customer
13940.0	2109.0	11831.0	.7	15577.0	114.0	.8	.2	1.0	.0	Existing Customer
3688.0	606.0	3082.0	.6	14596.0	120.0	.8	.2	1.0	.0	Existing Customer
4003.0	1851.0	2152.0	.7	15476.0	117.0	.9	.5	1.0	.0	Existing Customer
4277.0	2186.0	2091.0	.8	8764.0	69.0	.7	.5	.2	.8	Attrited Customer
5409.0	.0	5409.0	.8	10291.0	60.0	.8	.0	.0	1.0	Attrited Customer
5281.0	.0	5281.0	.5	8395.0	62.0	.7	.0	.0	1.0	Attrited Customer
10388.0	1961.0	8427.0	.7	10294.0	61.0	.6	.2	.1	.9	Attrited Customer

As we can see from the excel sheet export image the predictions of the attrition flag mostly predicted with the confidence level at 1.00 either of flag values. Values are rounded up automatically in excel spreadsheet..

Conclusion

- As we have seen the results of three models on which we have worked for pre-prediction testing the result itself says that among that 3 models the required result is only achieved through **deep learning** model instead of generalized linear and logistic regression model.

- We have found the expected result of prediction of attrition flag as churned (50%) and existing (50%) customers from the sample set of pre-prediction testing using deep learning model.

- Thus, due to the desire results, we can say it again that the **deep learning** model is the best suited model of our findings for the bank churners dataset.

Suggestions

- As i have mentioned in phase 1 that the customer who fall under the lower income brackets had low credit limit and the churning proportion of that customers is higher compared to the customer with the high credit limit as well as high income so my suggestion to this is that increase the a bit credit limit of customer who falls under the lower income brackets.

- Month of inactive in 12 months and total transaction counts are also playing the crucial role in the issue of customers leaving bank credit card services so my suggestion to this is that give some scheme like total transaction count reach at particular total in some month and month of inactive in 12 months decreased to 0 or 1 they should get some extra credit limit for having the good relationship with bank services.

Group 3 [Green]

Thank You For Your Precious Time