Data Classification using K-NN: Personal Loan Acceptance Case-Study] **After dead-line not accepted in all cases]**

1) Analytic Problem

Universal Bank is a relatively young bank growing rapidly in terms of overall customer acquisition. The majority of these customers are liability customers (depositors) with varying sizes of relationship with the bank. The customer base of asset customers (borrowers) is quite small, and the bank is interested in expanding this base rapidly to bring in more loan business. In particular, it wants to explore ways of converting its liability customers to personal loan customers (while retaining them as depositors).

A campaign that the bank ran last year for liability customers showed a healthy conversion rate of over 9% success. This has encouraged the retail marketing department to devise smarter campaigns with better target marketing. The goal of our analysis is to model the previous campaign's customer behavior to analyze what combination of factors make a customer more likely to accept a personal loan. This will serve as the basis for the design of a new campaign.

2) Data Exploration

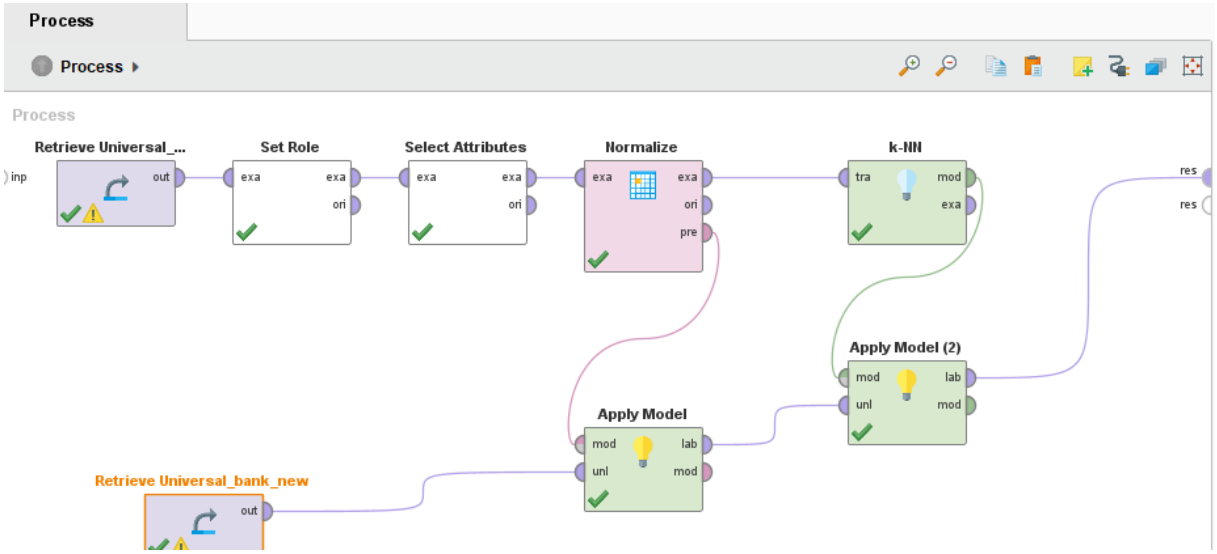
The file UniversalBank.xls contains data on 5000 customers. The data include customer demographic information (age, income, etc.), the customer's relationship with the bank (mortgage, securities account, etc.), and the customer response to the last personal loan campaign (*Personal Loan*). Among these 5000 customers, only 480(= 9.6%) accepted the personal loan that was offered to them in the earlier campaign.

3) Prediction using KNN with given K

Consider the following customer:

Age=40, Experience=10, Income=84, Family=2, CCAvg=2, Education_2=1, Education_3=0, Mortgage=0, Securities Account=0, CD Account=0, Online=1 and Credit card=1.

3.1) Perform a *k*-NN classification with **all predictors except ID and ZIP code** with using ***k* = 5**. Please do not forget to **normalize all datasets.** Show all steps of your work



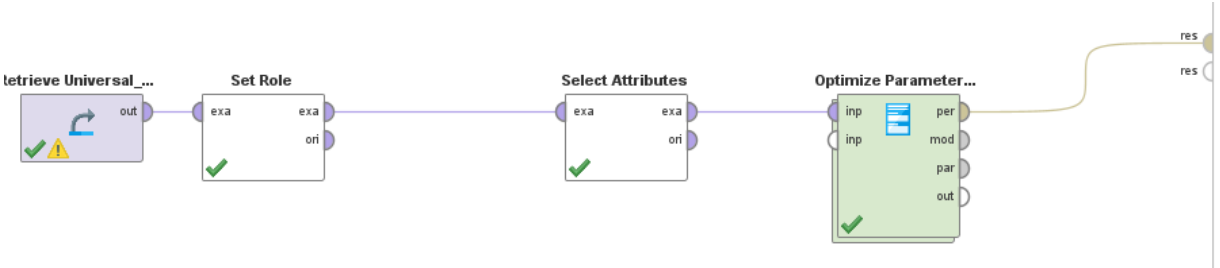
3.2) How would this customer be classified?

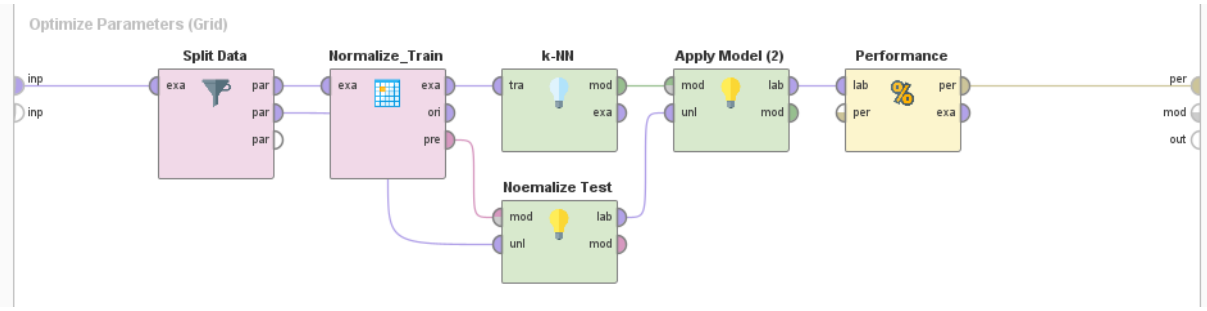
Row No.	prediction(P...	Age	Experience	Income	Family	CCAvg	Education	Mortgage	Securities A...	CD Account	Online
1	0	-0.466	-0.881	0.222	-0.345	0.036	0.142	-0.555	-0.341	-0.254	0.822

CreditCard	Personal Lo...
1.549	?

4) **Classifier construction to find Best K**

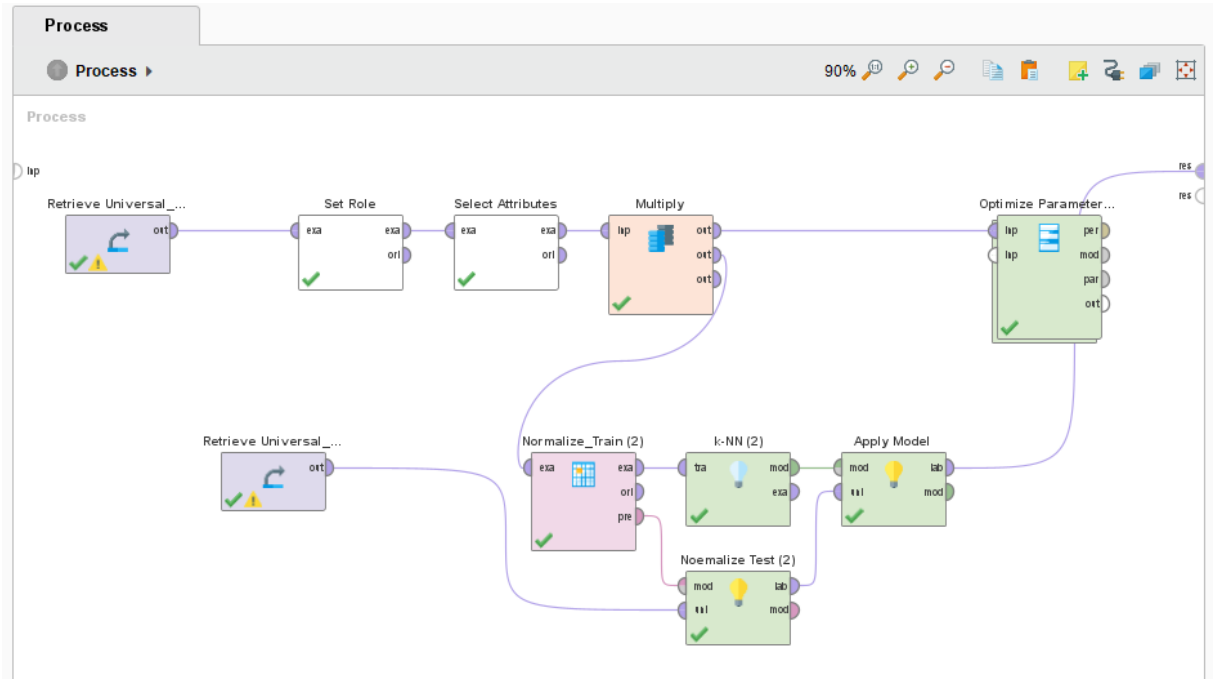
4.1) Partition the data into training (60%) and validation (40%) sets. Show the accuracy on the test data that results by varying *k*. Show final confusion matrix with your best K

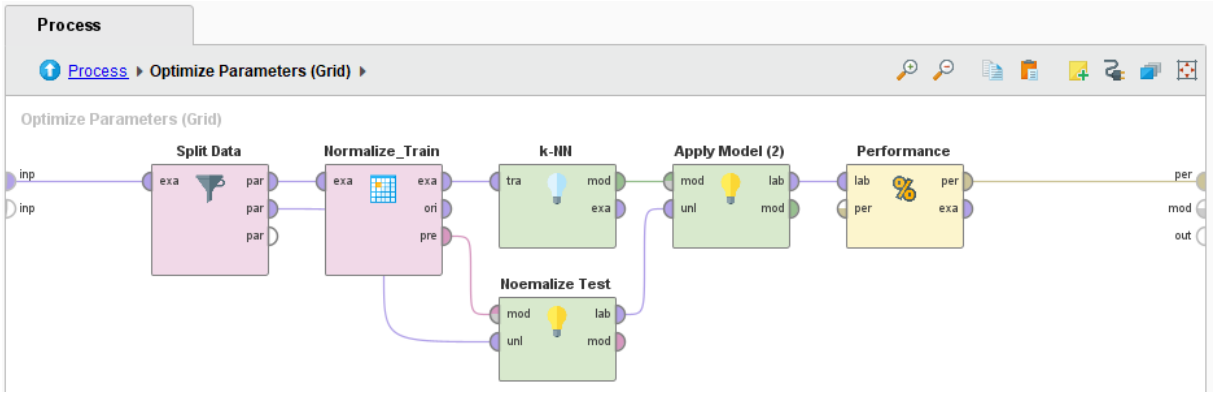




iteration	k-NN.k	root_m...
2	3	0.196
1	1	0.332
4	7	0.202
8	14	0.211
3	5	0.202
5	9	0.204
6	11	0.207
10	18	0.215
7	12	0.209
11	20	0.216
9	16	0.213

4.2) Using the best *k*, how would this customer be classified?





Row No.	prediction(P...	Age	Experience	Income	Family	CCAvg	Education	Mortgage	Securities A...	CD Account	Online	Credit
1	0	-0.466	-0.881	0.222	-0.345	0.036	0.142	-0.555	-0.341	-0.254	0.822	1.549

iteration	k-NN.k	root_m...
3	5	0.202
2	3	0.196
4	7	0.202
1	1	0.332
7	12	0.209
6	11	0.207
9	16	0.213
5	9	0.204
8	14	0.211
11	20	0.216
10	18	0.215

5) Submit All coding files