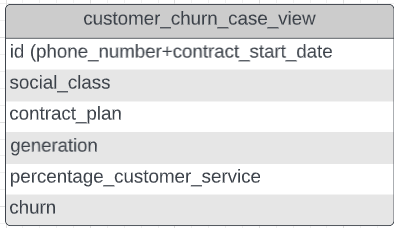
# Intro

This document shows the usage of two different ML algorithms using business-defined parameters that were traditionally considered relevant for the business to understand the risk of churn from a customer. Those parameter were defined in the table below:



The details of how this work was done are contained in the file Section\_C\_-\_ML.sql

### **Decision tree model**

We’ve created a decision tree model with auto prep on. With the auto prep parm ON, the algorithms considered that the only variable important on the customer\_churn\_case\_view was percentage\_customer\_service which is the percentage of all user activity that is customer services call. That is the model target and independent variables:

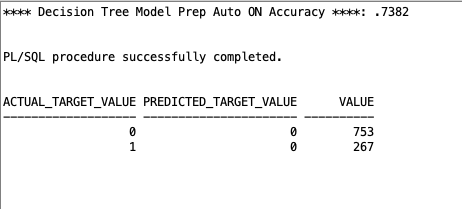
ATTRIBUTE\_NAME ATTRIBUTE\_T USAGE\_TY TAR

-----------------------------------------------------------

PERCENTAGE\_CUSTOMER\_SERVICE NUMERICAL ACTIVE NO

CHURN CATEGORICAL ACTIVE YES

The limitation from this model is that it predicted always not churn (0) as most likely result. The accuracy of 0.7382 is merely the percentage of customer who didn’t churn 73.82% on the test dataset



Given those issues, we’ve decided to use a different algorithm.

### **Naive Bayes model**

We’ve created a decision tree model with auto prep OFF. With the auto prep parm OFF, the algorithm considered all the variables on customer\_churn\_case\_view to build the model. That is the model target and independent variables:

ATTRIBUTE\_NAME ATTRIBUTE\_T USAGE\_TY TAR

-----------------------------------------------------------

PERCENTAGE\_CUSTOMER\_SERVICE NUMERICAL ACTIVE NO

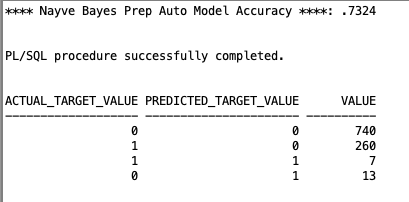
CHURN CATEGORICAL ACTIVE YES

GENERATION CATEGORICAL ACTIVE NO

SOCIAL\_CLASS CATEGORICAL ACTIVE NO

CONTRACT\_PLAN\_NAME CATEGORICAL ACTIVE NO

Although this model didn’t always predict 0, the accuracy of this model, 0.7324, is inferior to the previous one. It predicted 7 churns right but predicted 13 churns wrongly.



# Conclusion

Given the poor performance of two different ML algorithms, we might conclude that the business defined important fields are not that important to predict customer churn. A new analysis with new fields would be required to build a good model.