# Azure machine learning Practice 5

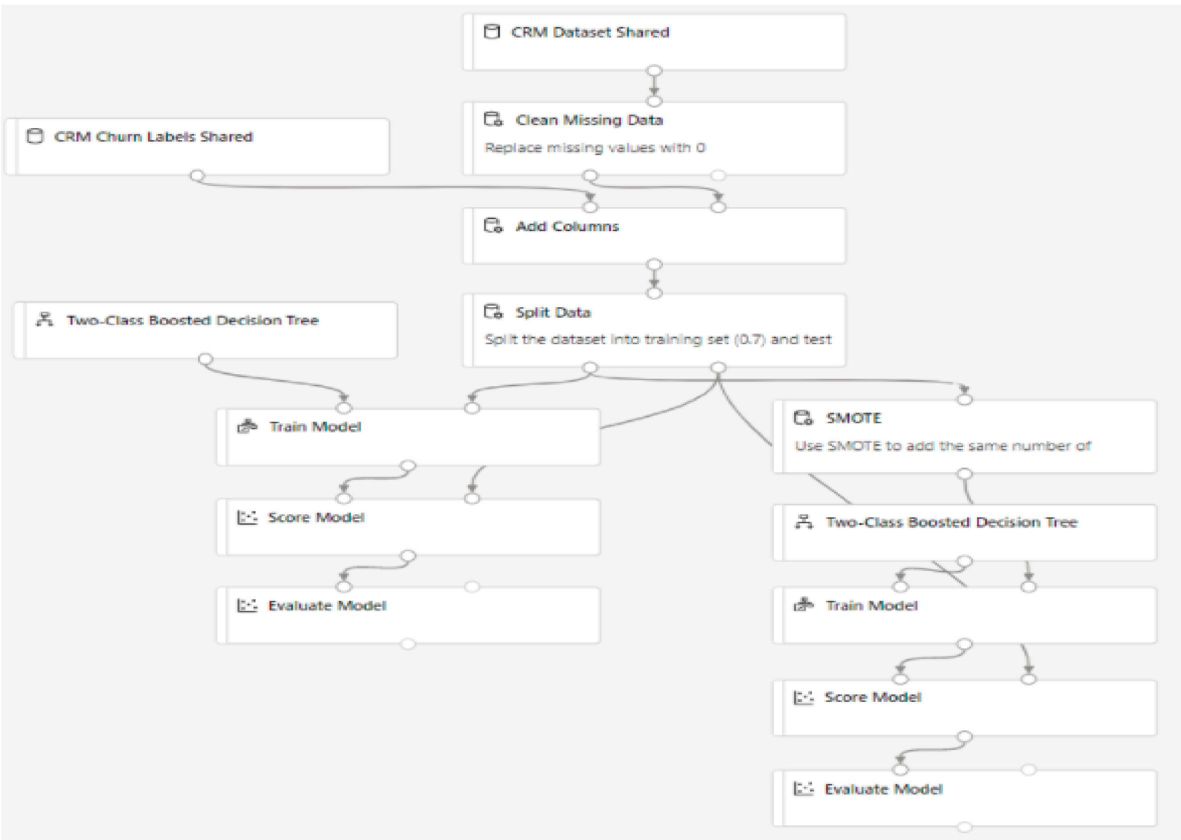
## 1. Questions: Decision tree to predict churn with Azure Machine Learning designer.

Introduction Decision Trees are a type of Supervised Machine Learning (that explains what the input is and what the corresponding output is in the training data) where the data is continuously split according to a certain parameter. The leaves are the decisions or the final outcomes.

Here we learn how to build a machine learning classifier without writing a single line of code using the designer. This sample trains a two-class boosted decision tree to predict adult census income (>=50Kor <=50K).

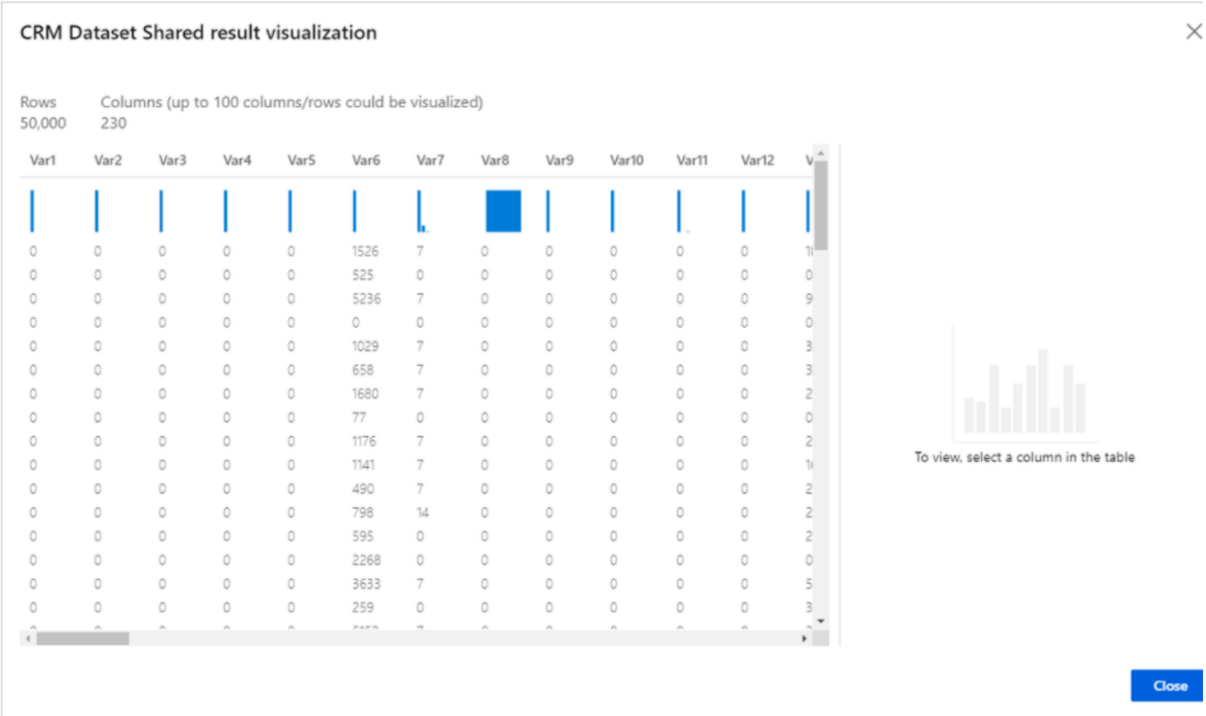
The Binary Classification - Customer Relationship Prediction sample pipeline on the designer homepage. This pipeline trains 2 two-class boosted decision tree classifiers to predict common tasks for customer relationship management (CRM) systems - customer churn. The data values and labels are split across multiple data sources and scrambled to anonymize customer information, however, we can still use the designer to combine data sets and train a model using the obscured values.

This is called a classification problem. We can apply the same fundamental process to tackle any type of machine learning problem - regression, classification, clustering, and so on.

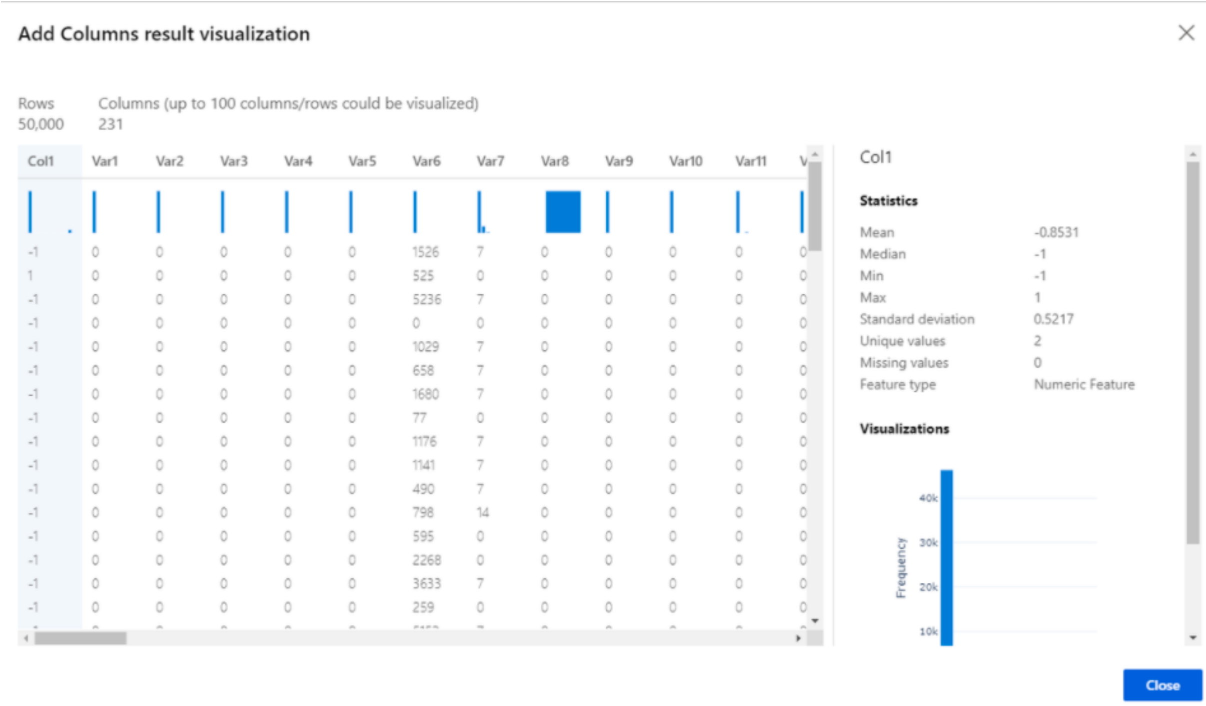


**Pipeline Summary:** This sample pipeline in the designer shows binary classifier prediction of churn, appetency, and up-selling, a common task for customer relationship management (CRM). First, some simple data processing.

- The raw dataset has many missing values. Use the Clean Missing Data module to replace the missing values with 0.




The features and the corresponding churn are in different datasets. Use the Add Columns module to append the label columns to the feature columns. The first column, Col1, is the label column. From the visualization result, we can see the dataset is unbalanced. There are way more negative (-1) examples than positive examples (+1). We will use SMOTE module to increase underrepresented cases later.

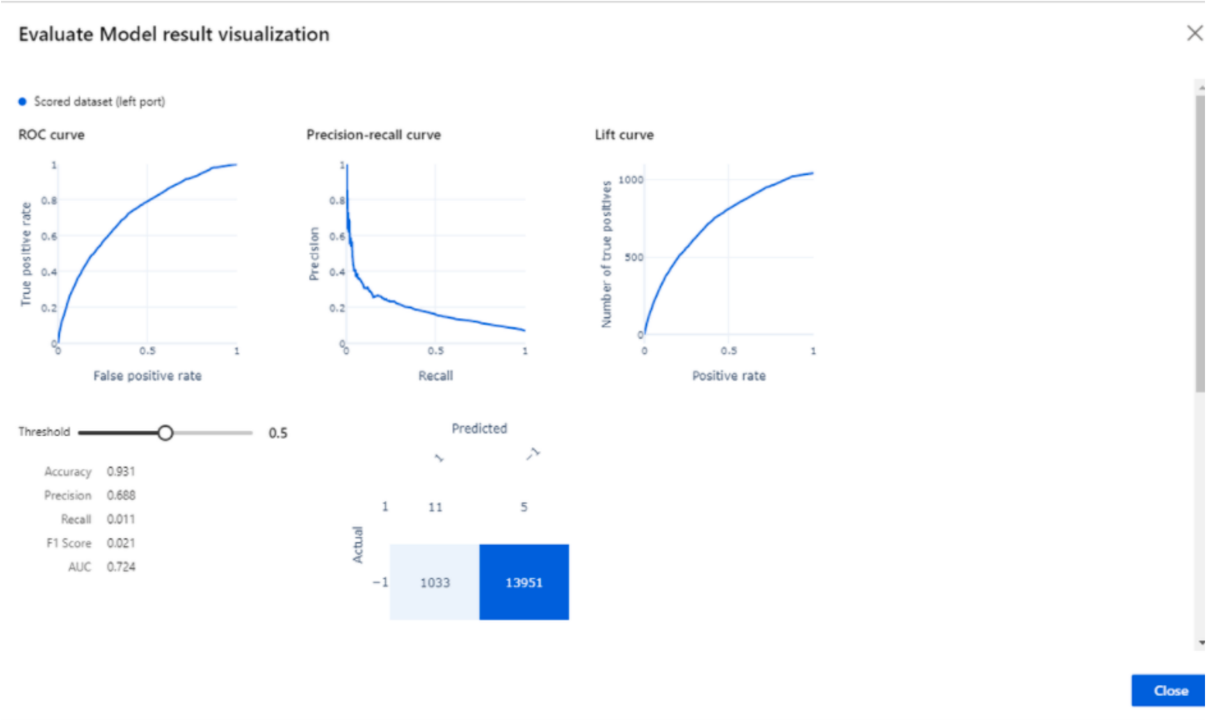



- Use the Split Data module to split the dataset into train and test sets.

- Then use the Boosted Decision Tree binary classifier with the default parameters to build the prediction models. Build one model per task, that is, one model each to predict up-selling, appetency, and churn.

- In the right part of the pipeline, we use SMOTE module to increase the percentage of positive examples. The SMOTE percentage is set to 100 to double the positive examples. Learn more on how SMOTE module works with SMOTE module reference0

## Results:




We can move the Threshold slider and see the metrics change for the binary classification task.

.

## 2. Questions: <u>Compose the list of possible predictors of a customer's churn</u>

**What is customer churn?**

Customer churn (or customer attrition) is a tendency of customers to abandon a brand and stop being a paying client of a particular business. The percentage of customers that discontinue using a company's products or services during a particular time period is called a customer churn (attrition) rate. One of the ways to calculate a churn rate is to divide the number of customers lost during a given time interval by the number of acquired customers and then multiply that number by 100 percent. For example, if we got 150 customers and lost three last month, then the monthly churn rate is 2 percent.

**The list of possible predictors of a customer's churn:** churn rate is one of the critical performance indicators for subscription businesses. The overall scope of work data scientists carry out to build ML-powered systems capable to forecast customer attrition may look like the following:

- Understanding a problem and final goal
- Data collection
- Data preparation and preprocessing

- Modeling and testing
- Model deployment and monitoring

The list of possible predictors of a customer's churn is Discussed below…

**Understanding a problem and a final goal**
It's important to understand what insights one needs to get from the analysis. In short, we must decide what question to ask and consequently what type of machine learning problem to solve: classification or regression. Sounds complicated, but bear with us.

**Data collection**
**Identifying data sources.** Identifying data sources. Once we've identified which kinds of insights to look for, we can decide what data sources are necessary for further predictive modeling. Assume the most common sources of data used for predicting churn.

**Data preparation and preprocessing**
Historical data that was selected for solving the problem must be transformed into a format suitable for machine learning. Since model performance and therefore the quality of received insights depend on the quality of data, the primary aim is to make sure all data points are presented using the same logic, and the overall dataset is free of inconsistencies.

Feature engineering, extraction, and selection. Feature engineering is a very important part of dataset preparation. During the process, data scientists create a set of attributes (input features) that represent various behavior patterns related to customer engagement level with a service or product.

**Modeling and testing**
The main goal of this project stage is to develop a churn prediction model. Specialists usually train numerous models, tune, evaluate, and test them to define the one that detects potential churners with the desired level of accuracy on training data.

Classic machine learning models are commonly used for predicting customer attrition, for example, logistic regression, decision trees, random forest, and others.

**Deployment and monitoring**
Predicting customer churn with machine learning and artificial intelligence is an iterative process that never ends. We monitor model performance and adjust features as necessary to improve accuracy when customer-facing teams give us feedback or new data becomes available. At the point of any human interaction – a support call, a CSM QBR [quarterly business review], a Sales discovery call – we monitor and log the human interpretation of customer help, which augments the machine learning models and increases the accuracy of our health prediction for each customer.