Final Project

Azure Machine Learning

Vaithilingam, Sudha

**Deep Azure@McKesson**

**Dr. Zoran B. Djordjević**

**Problem Statement**

In this competitive world, for companies to stay profitable it is important to have a strategy focusing not only on winning new customers but retaining existing customers. If companies can predict which customers are at a risk of leaving (churn), which customers are good candidates for add-ons (upselling) and new products/services (appetency), it would help them optimize their customer relationship and sales/marketing effort and dollars. In this project, I have used Azure Machine Learning to build a Customer Relationship Prediction model to predict churn, appetency and upselling using dataset modeled from KDD Cup 2009.

**Description of Dataset**

Sample datasets available in Azure Machine Learning Studio modeled from

<http://www.kdd.org/kdd-cup/view/kdd-cup-2009/Data>

CRM dataset Shared, Size:25 MB, Format: GenericTSV

CRM upselling Labels Shared, Size:192KB, Format: GenericTSVNoHeader

CRM Appetency Labels Shared, Size:194KB, Format: GenericTSVNoHeader

CRM Churn Labels Shared, Size:192KB, Format: GenericTSVNoHeader

**Description of Hardware**

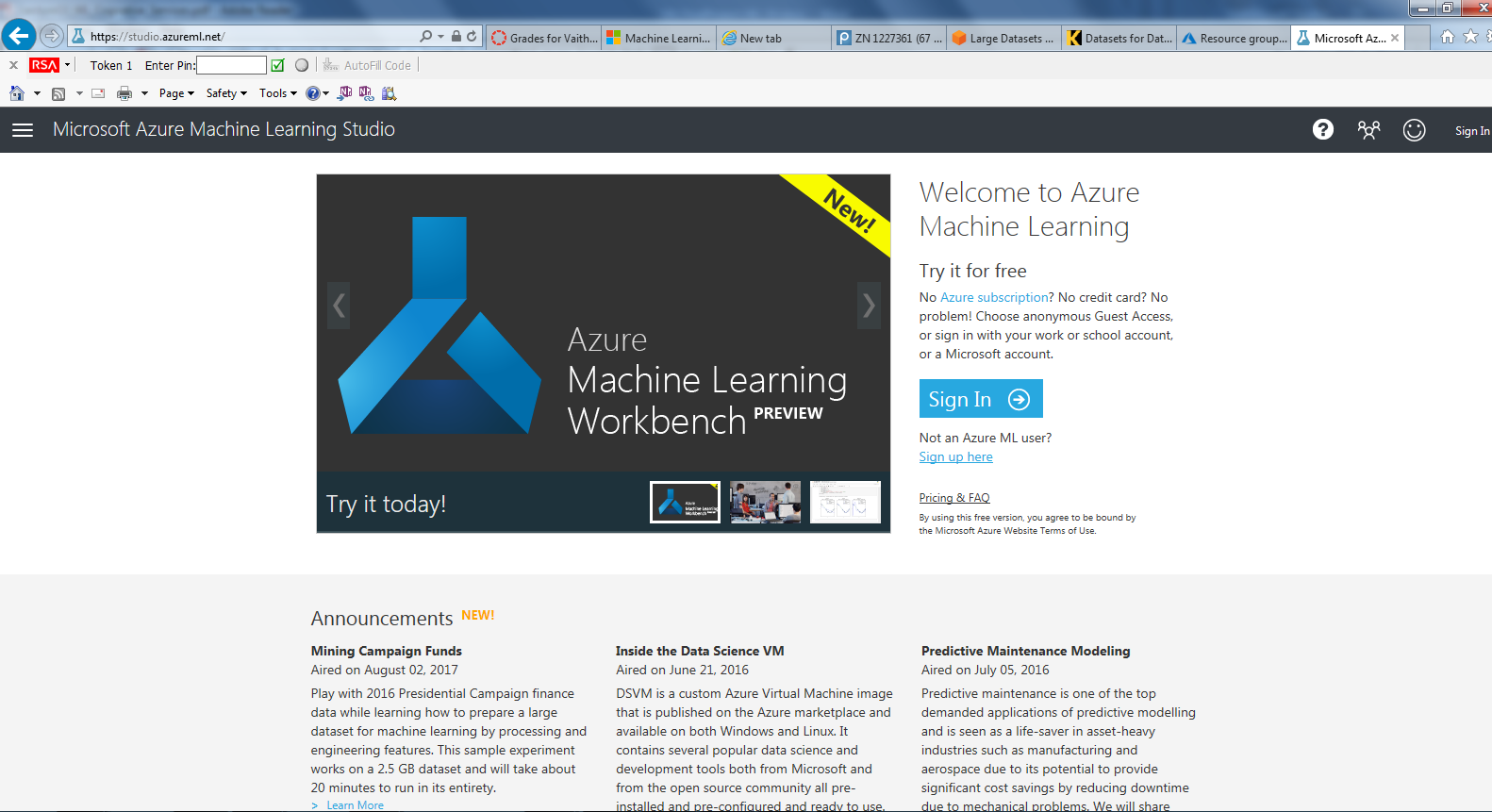
Windows 7 64 bit on Intel Core i5 CPU @2.40Ghz, 16.0 GB RAM

**Description of Software** (Provide URLs of downloads)

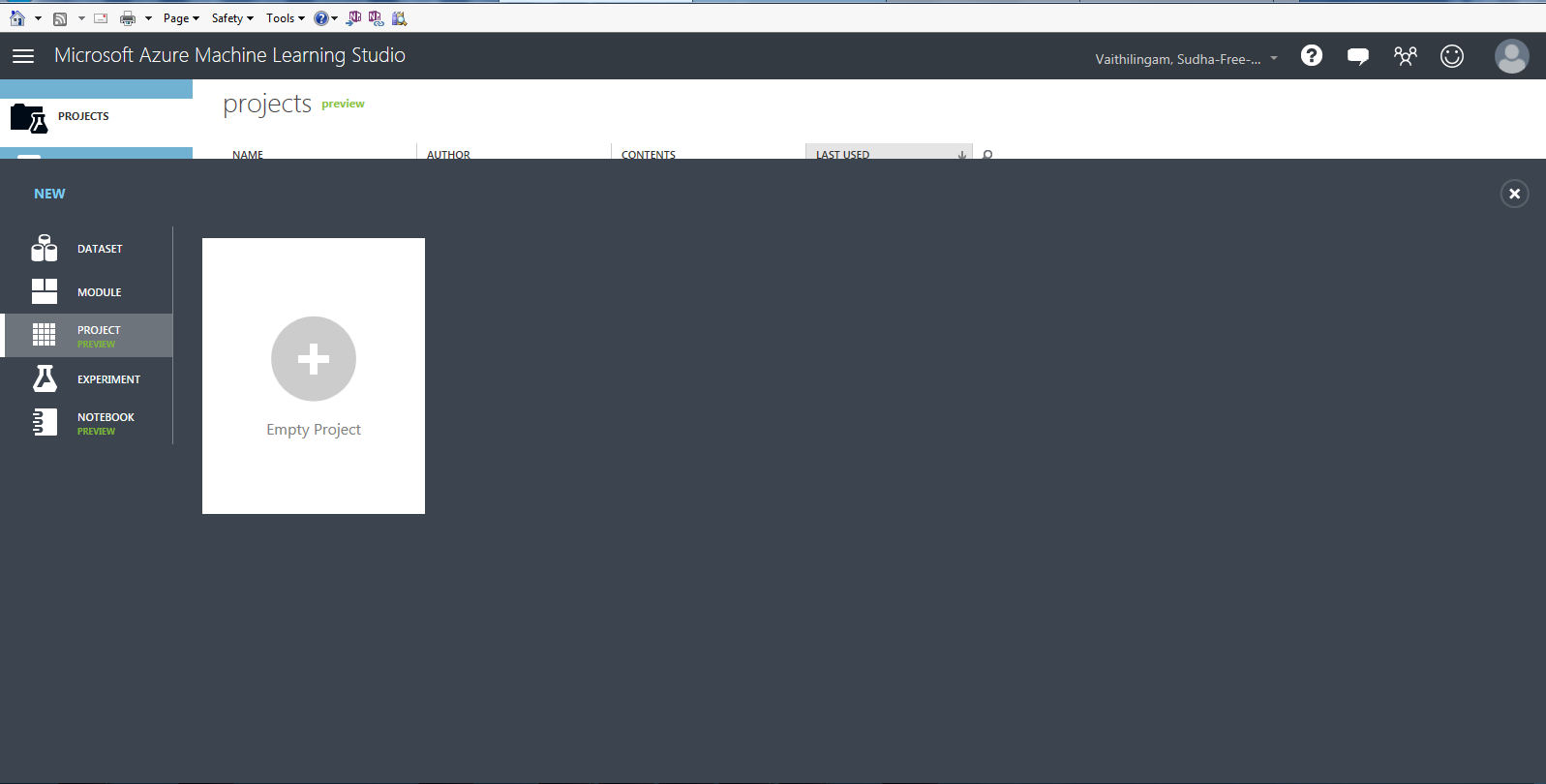
Azure Machine Learning Studio <https://studio.azureml.net>

**Installation** & **configuration**

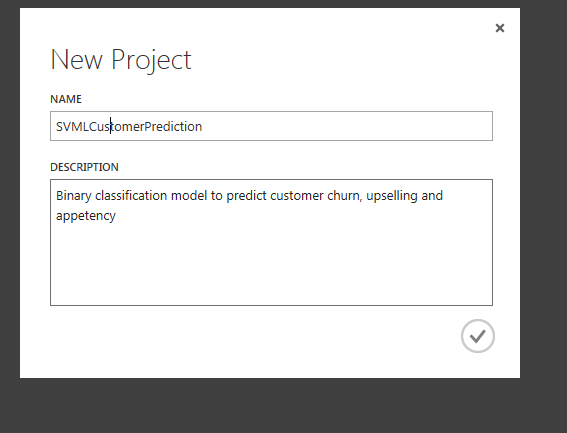
1. Type <https://studio.azureml.net> in your browser. Sign in to Azure Machine Learning Studio. A workspace is created in your name



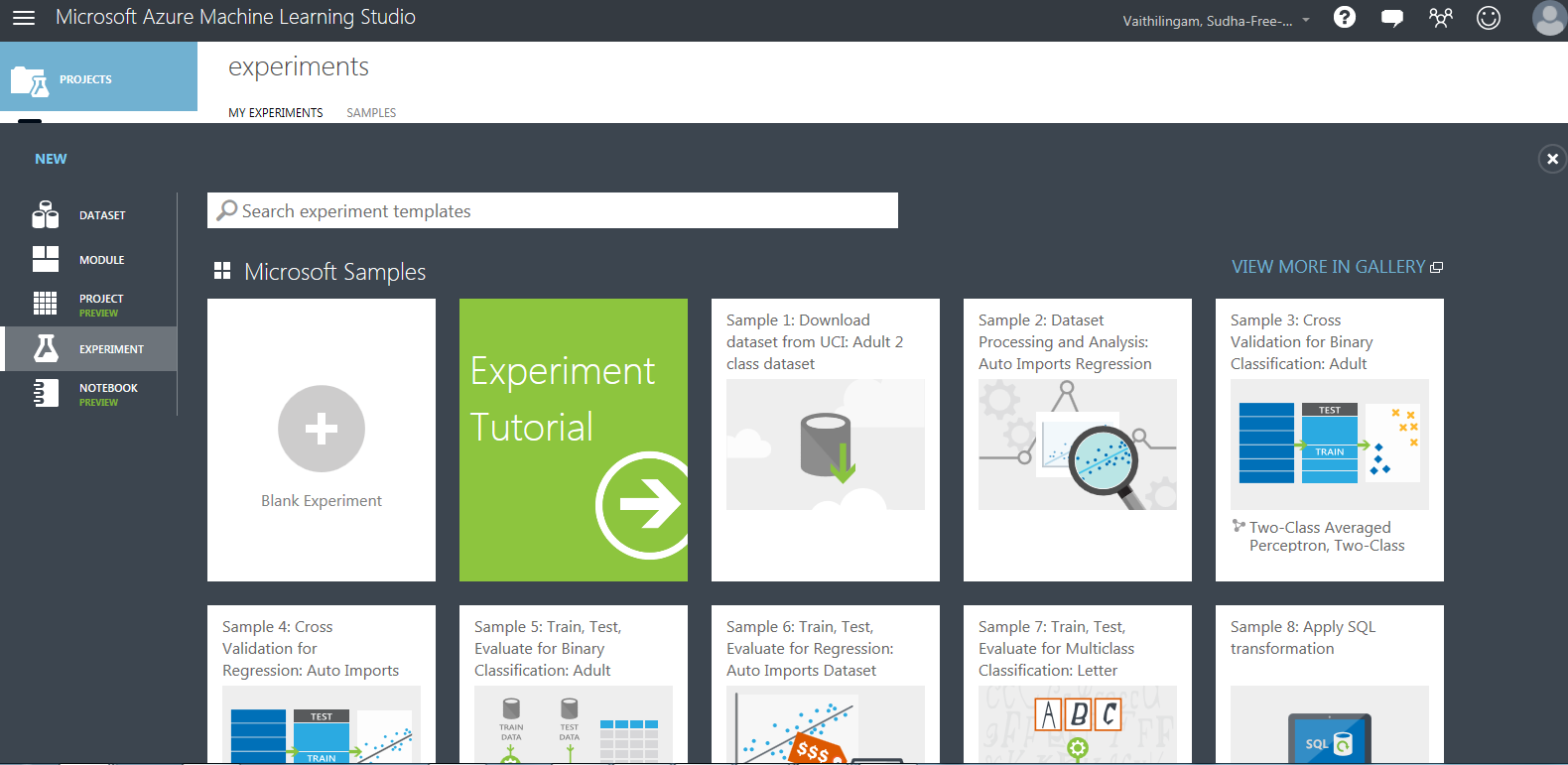
1. Create New project - Select Projects in the left pane and + New in the bottom of the window. Select Empty project



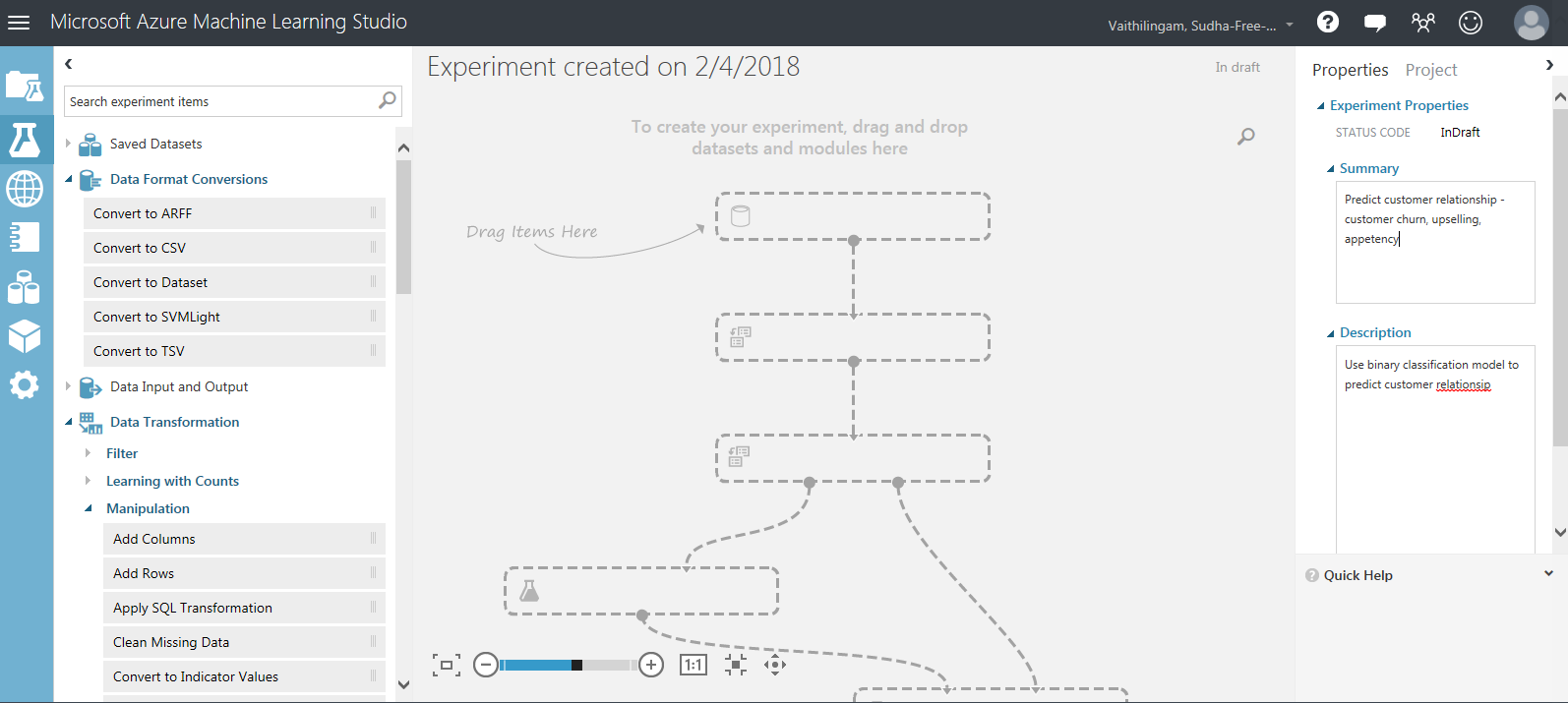
1. Enter a name and description for your project and click OK



1. Create New Experiment – In the left pane, select Experiment and click + New at the bottom of the window. Select Blank experiment.

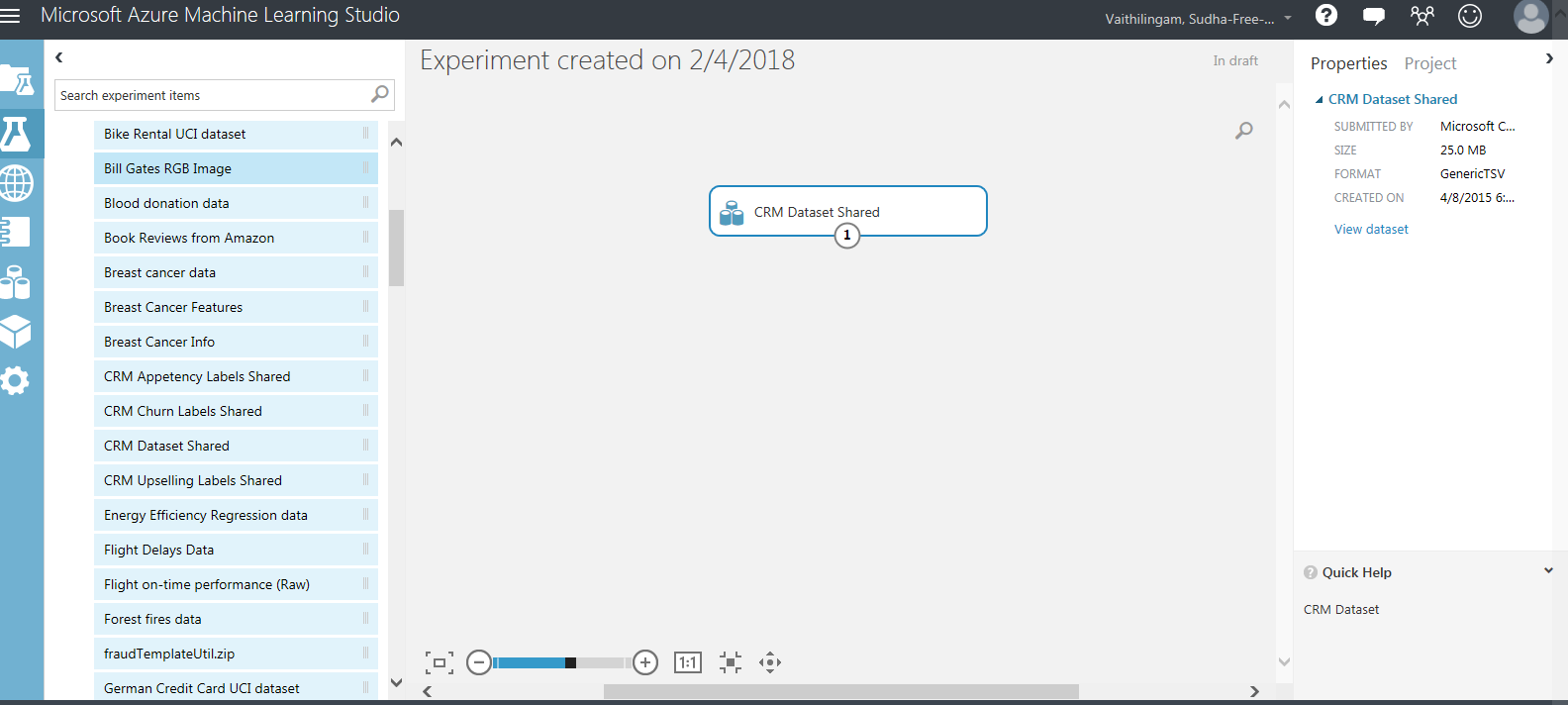


1. The experiment canvas is displayed. Now you can create your experiment by dragging and dropping modules from the left pane on to the canvas. The right pane allows you to configure parameters based on the module selected.



**Data**

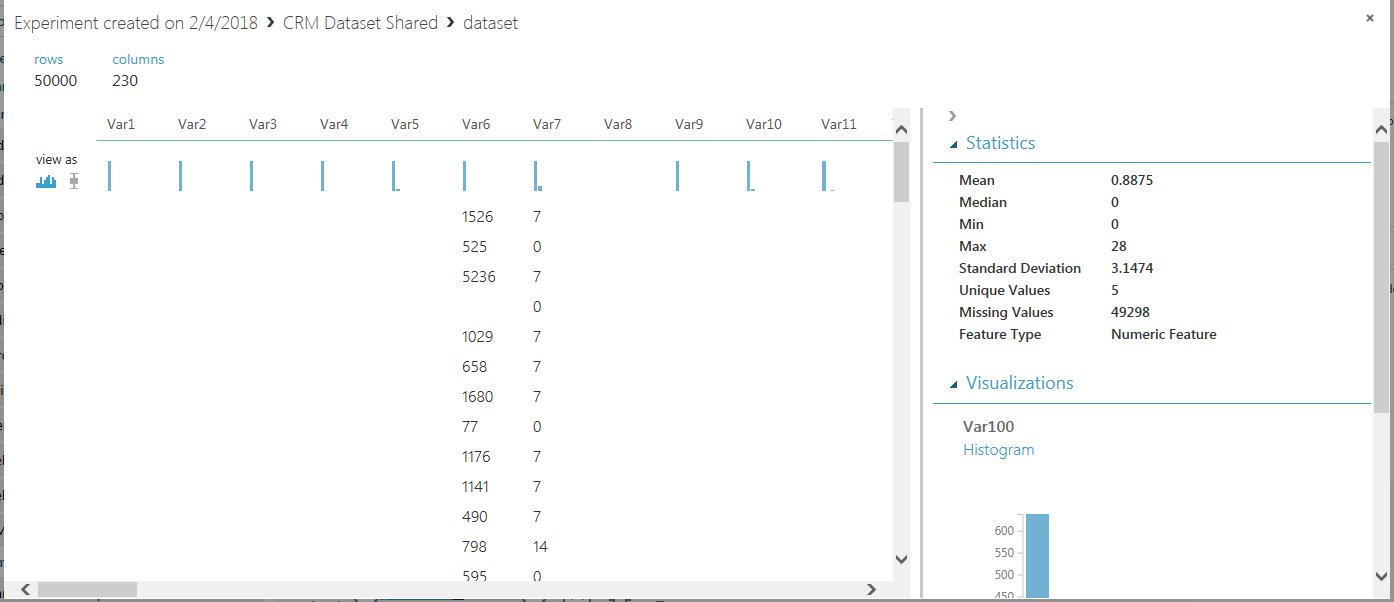
1. First, we need to select the dataset that we will be using to train the model. From the Samples Data set, select **CRM dataset Shared**, drag and drop it on the canvas



1. Right-click on the dataset and select visualize. The dataset has lot of missing values. We need to clean the dataset. The dataset has **50,000 rows** and **230 columns** with lot of missingvalues.

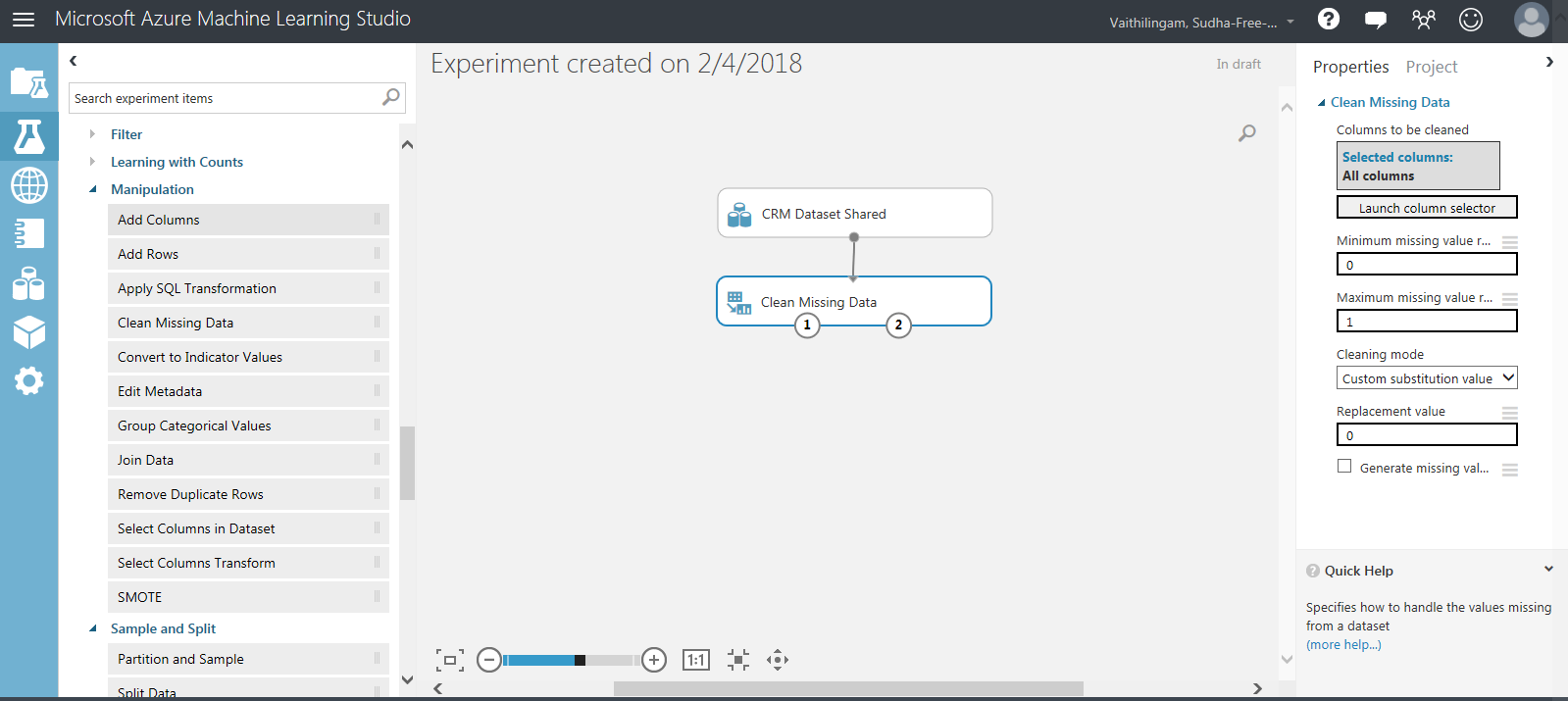
By clicking on a column, you can see the statistics for the column such as missing values, min, max, mean, median etc.

Sample data before cleaning

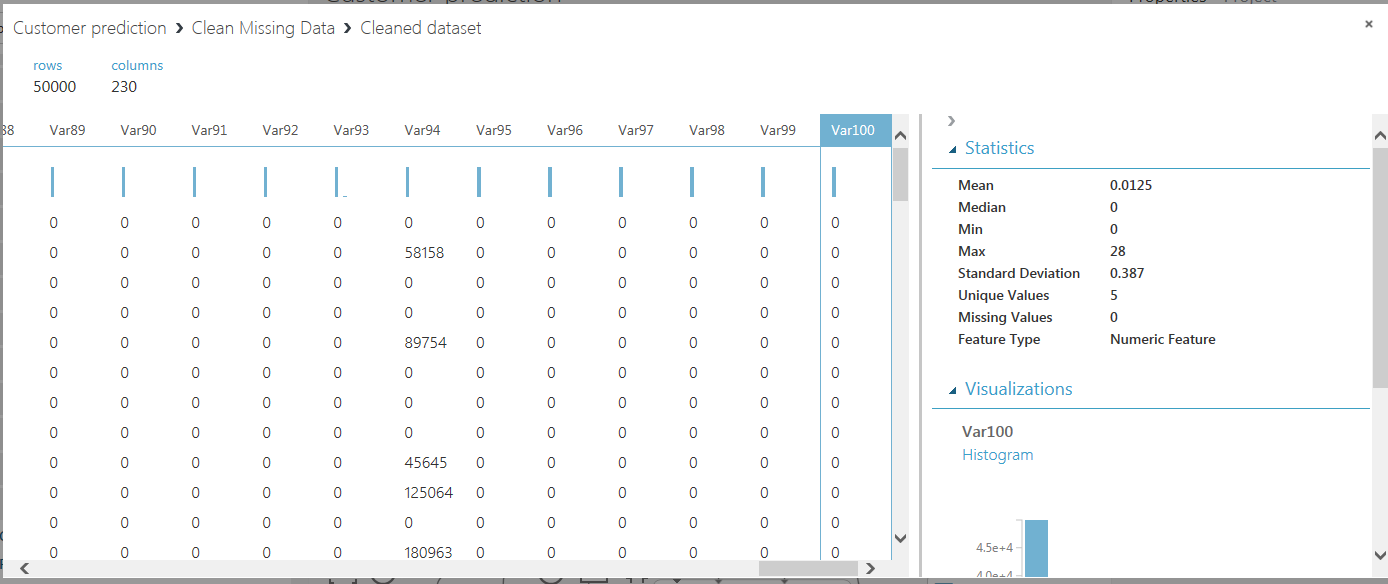


**Clean data**

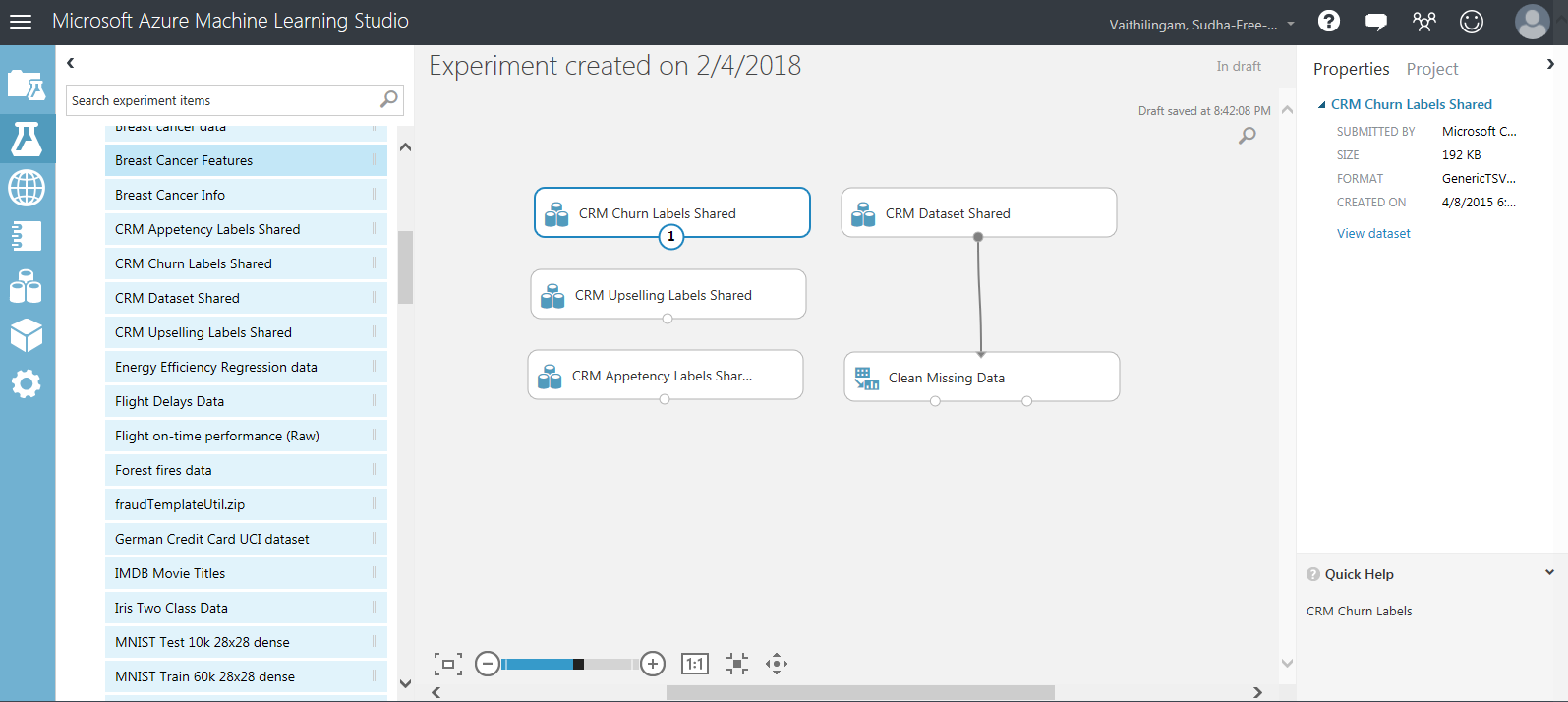
1. Add module **Clean Missing Data** and on the right pane select to substitute 0 for missing columns. Connect **Clean Missing Data** to **CRM Dataset shared**.



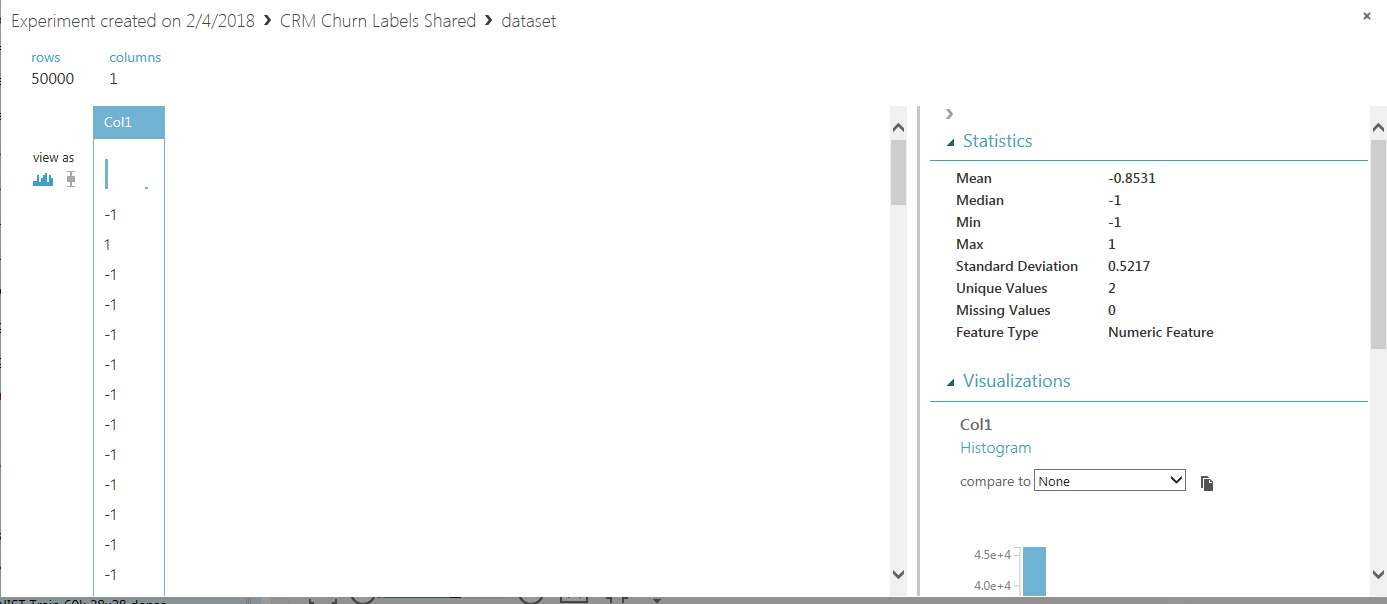
After **Clean missing data** module is run, Right click on **Clean missing data**, cleaned dataset -> visualize to see the data. The cleaned data set looks as below. The missing values are replaced with 0.



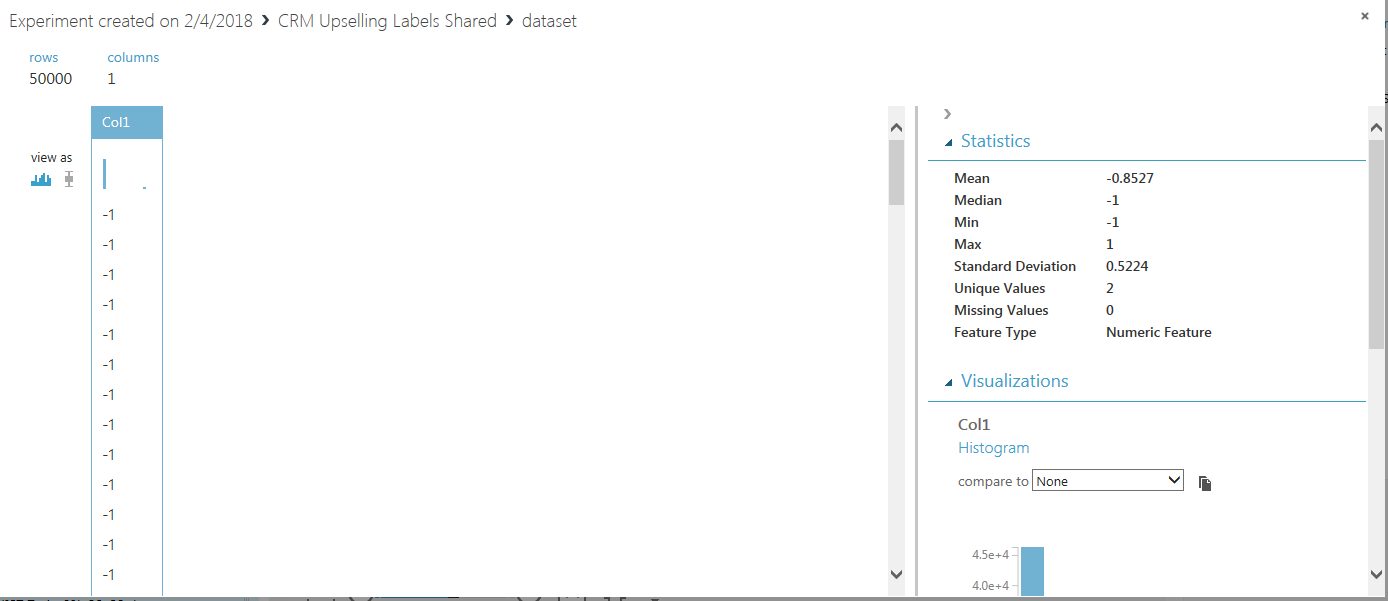
1. Add the datasets **CRM Churn Labels Shared**, **CRM Upselling Labels Shared** and **CRM Appetency Labels shared** by dragging from the left pane and dropping on the canvas



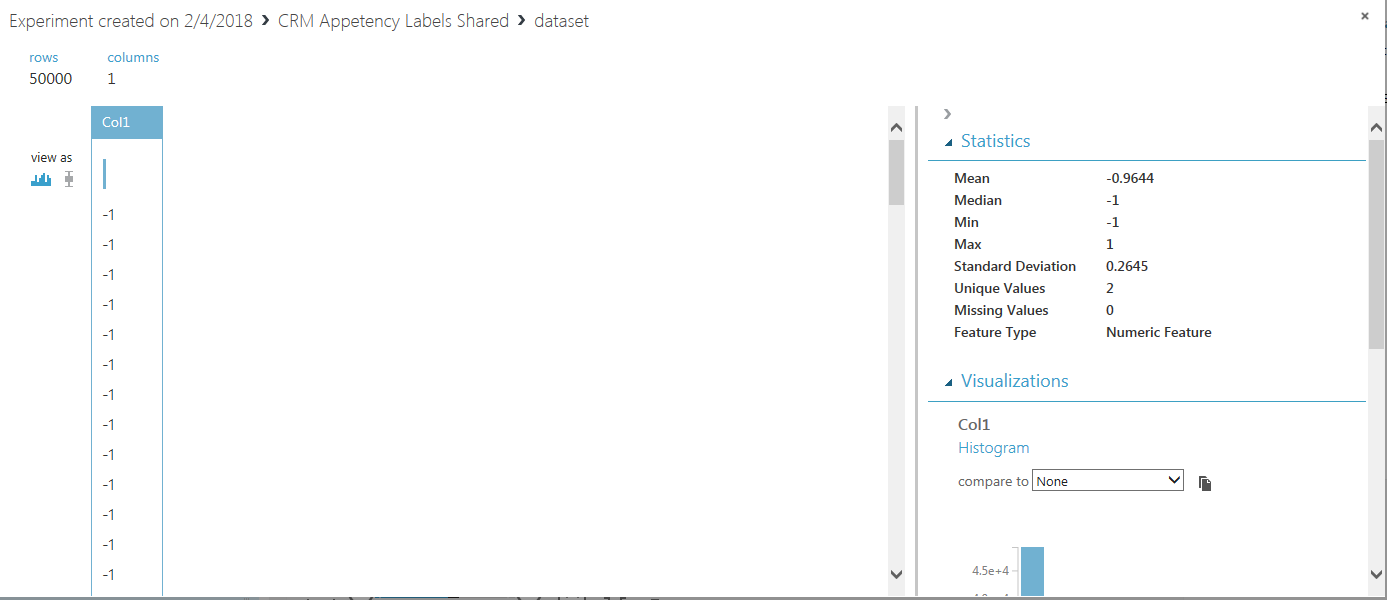
1. Right-click on **CRM Churn Labels shared** and select Visualize. There are **50,000 rows** corresponding to the CRM dataset Shared with 2 values **– 1 and 1**. There are no missing values.



1. Right-click on **CRM Upselling labels shared** dataset and select Visualize. There are **50,000 rows** corresponding to the CRM dataset Shared w**ith 2 values – 1 and 1**. There are no missing values.

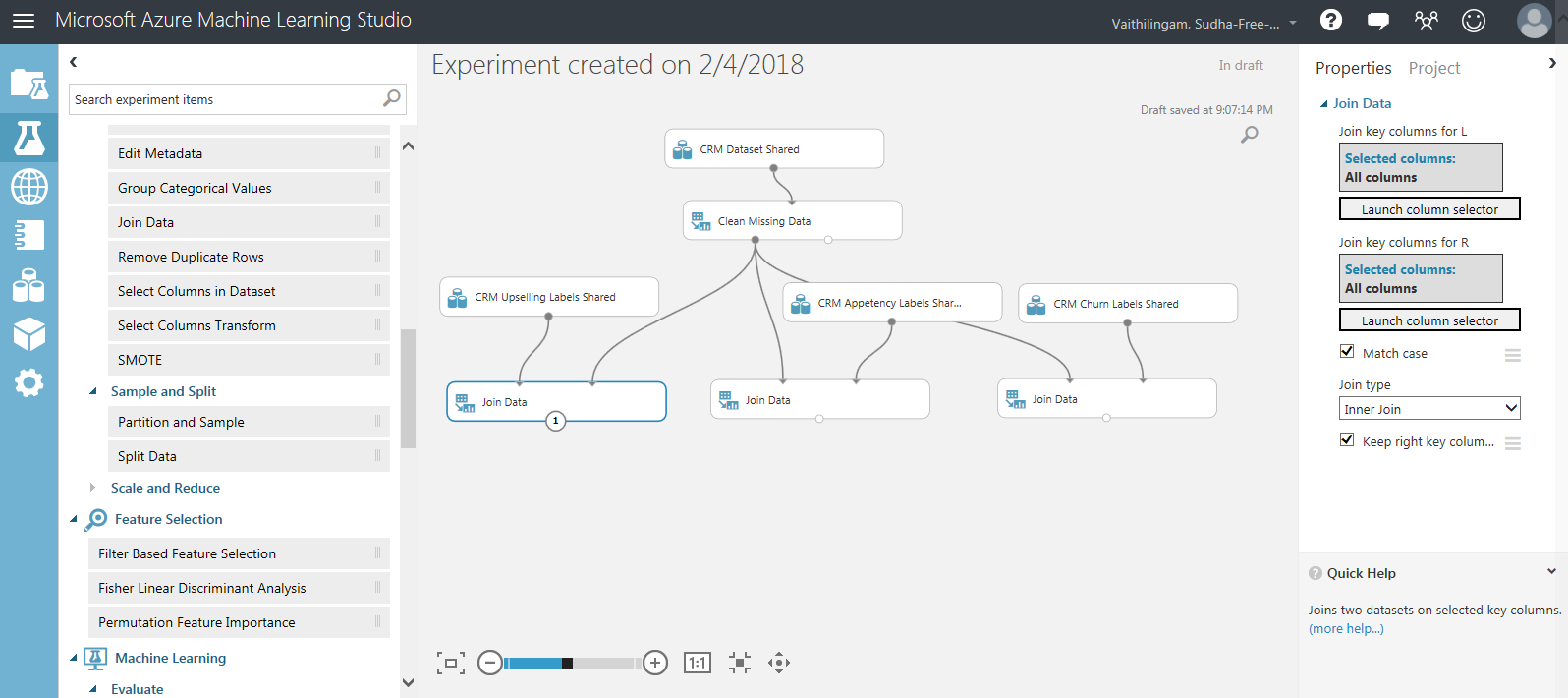


1. Right-click on **CRM Appetency labels** shared dataset and select Visualize. There are **50,000 rows** corresponding to the CRM dataset Shared w**ith 2 values – 1 and 1**. There are no missing values.

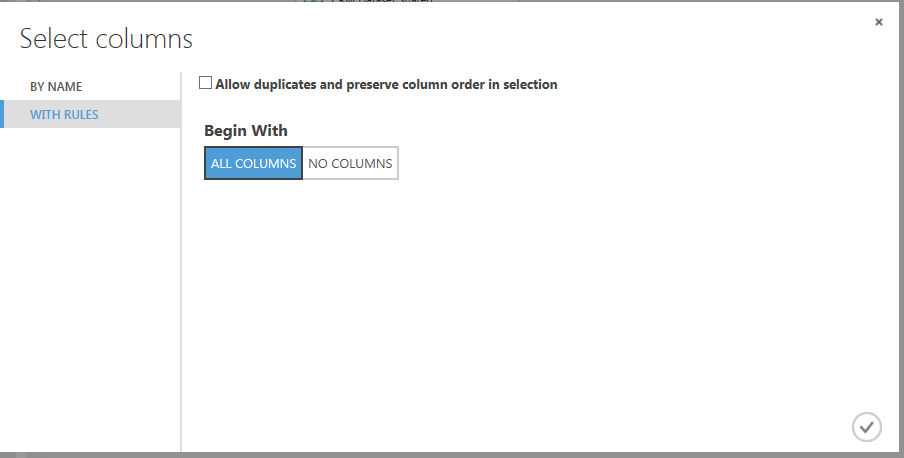


**Transform data**

1. We will use **join data** module to join the cleaned CRM dataset shared to the CRM Churn labels shared, CRM Appetency labels shared and CRM upselling labels shared datasets



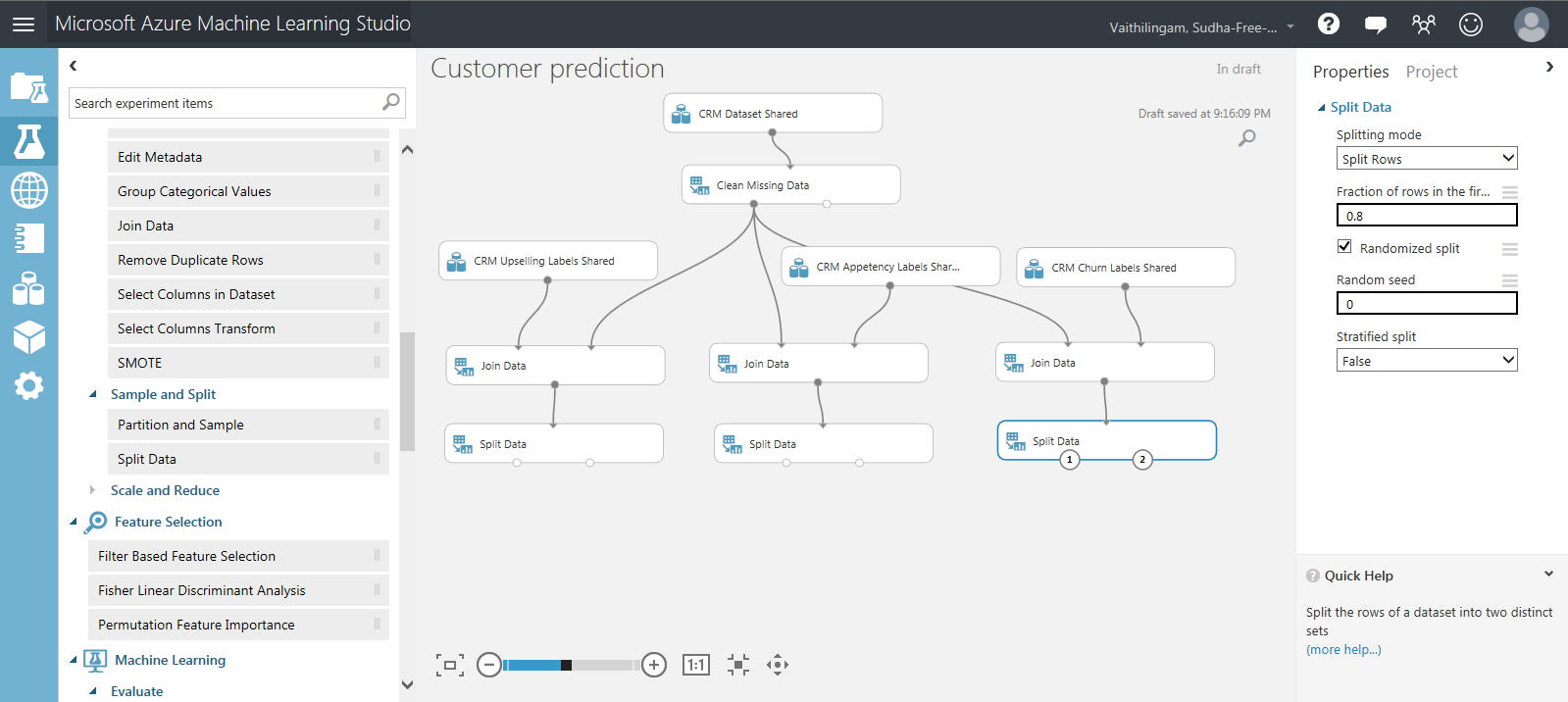
1. For every join Data, click on Launch column selector and select all columns



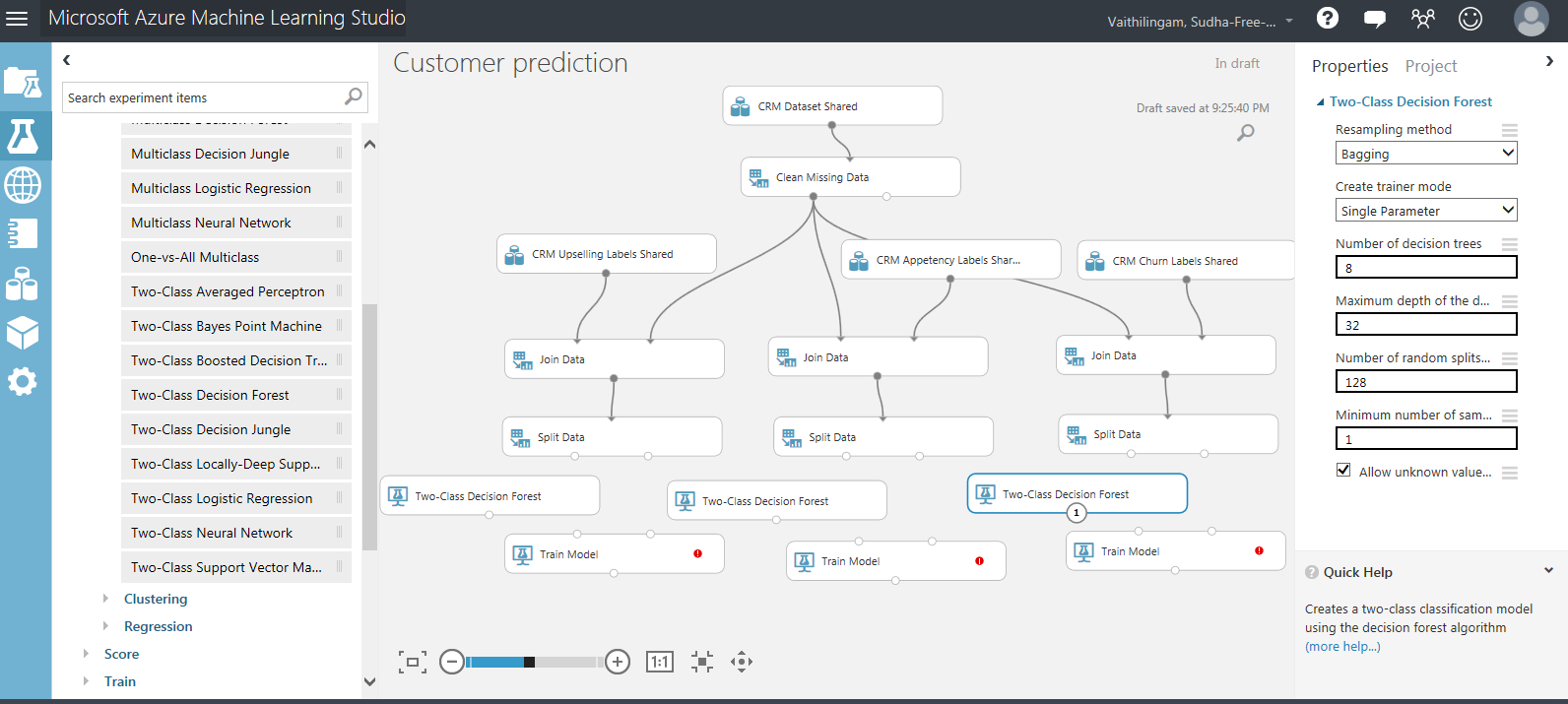
**Choose machine learning algorithm and Train the model**

1. In machine learning, you need to use the dataset for both training and validation.

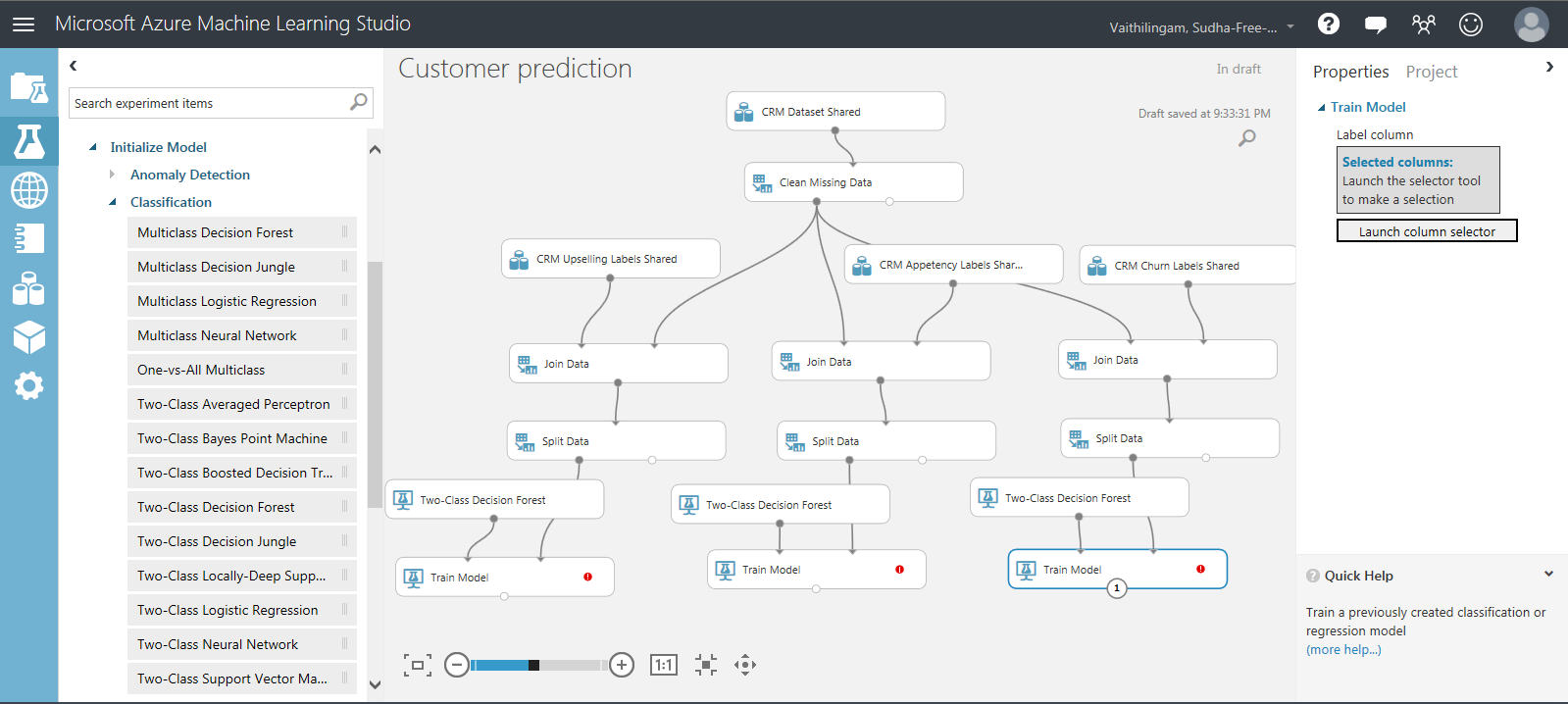
Add module **Split Data** and on the right pane update parameters to split the data set to 80% training, 20% test. Connect **Split Data** to **Join Data.**



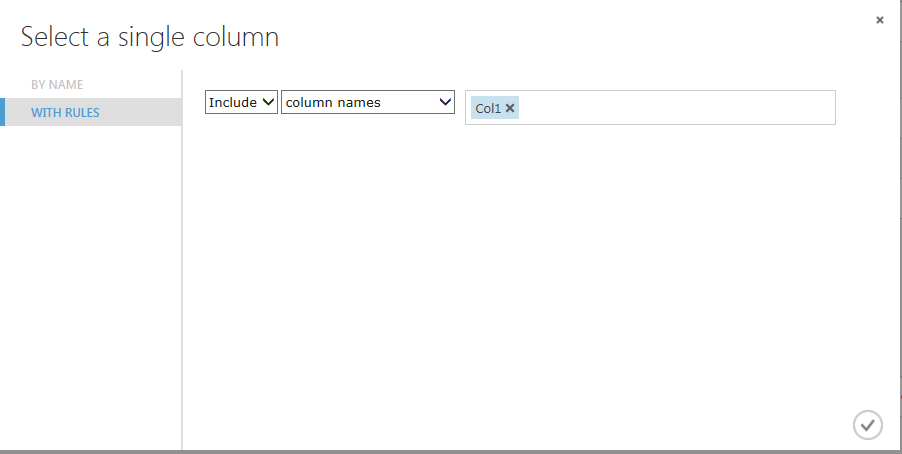
1. We are building a model to predict 3 values - customer churn, customer upselling and customer appetency. Since each of these labels has 2 values -1 or 1, our prediction is a binary classification problem. We will use the machine learning algorithm of Two-class decision forest for our prediction model. Add module **Two-class decision forest** to the canvas – one for CRM churn, one for CRM upselling. One for CRM Appetency



1. Add module **Train Model** and connect it to modules **Two-class decision forest** and **Split data.** Do this for each of the 3 values.

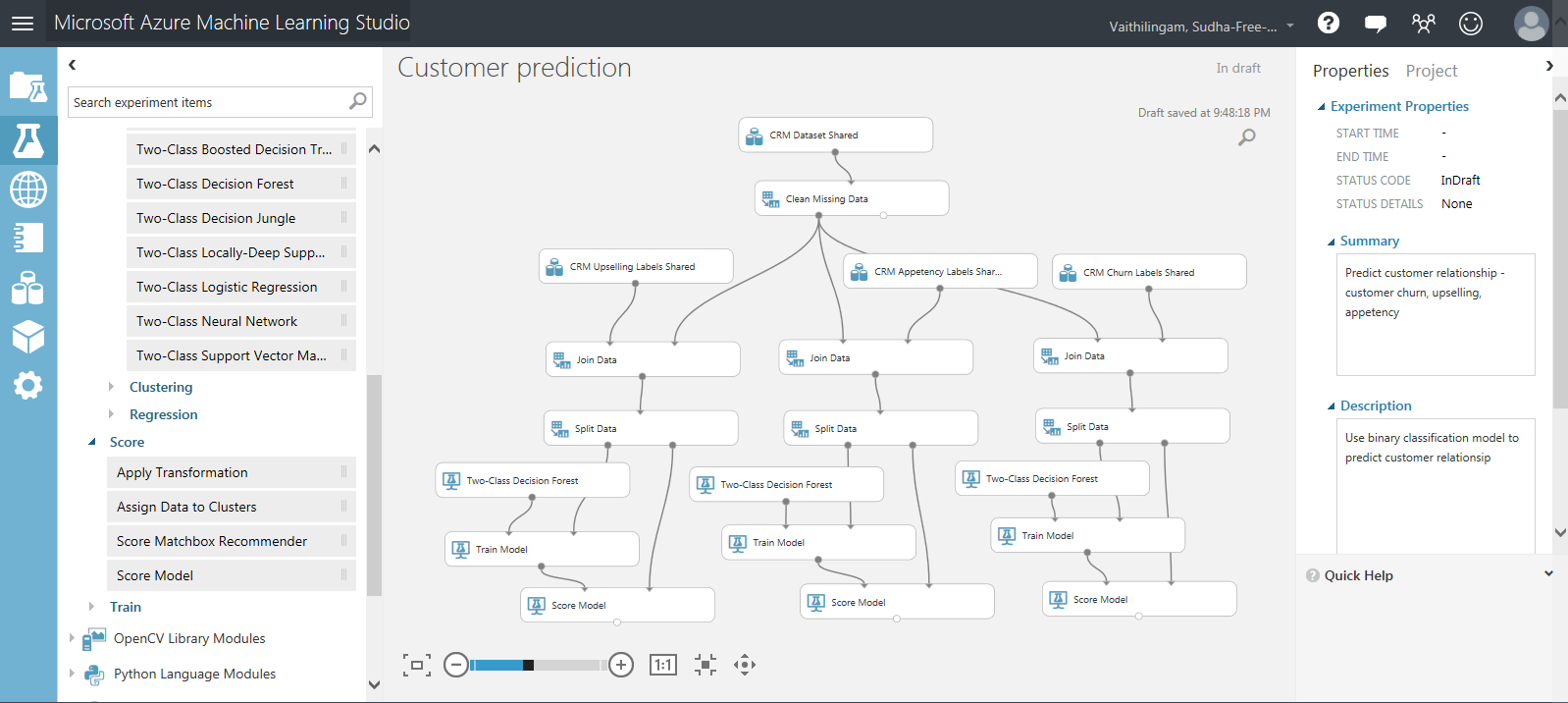


1. Select train model, on the right pane select Launch column selector and select Col1. This is the column we are predicting. (the value with 1 or -1). Do this for all the 3 train model modules.



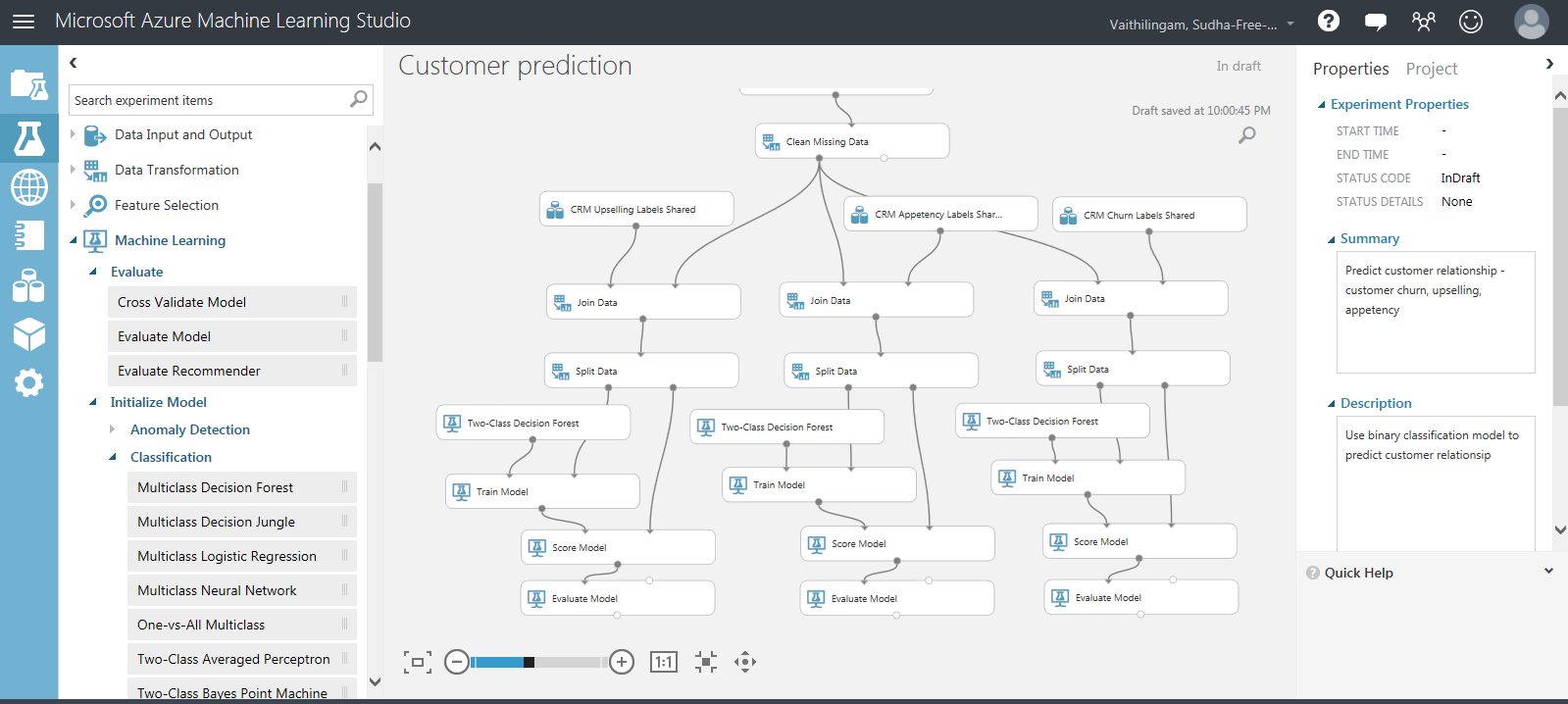
**Validate model**

1. Add module **Score Model** and connect it to **Split Data** and **Train Model**. Do this for each of the 3 values.

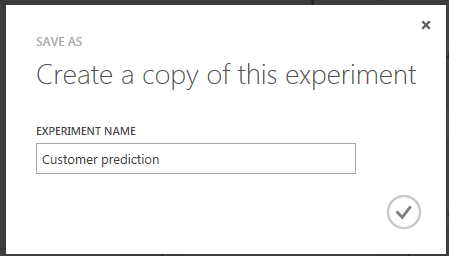


**Evaluate model**

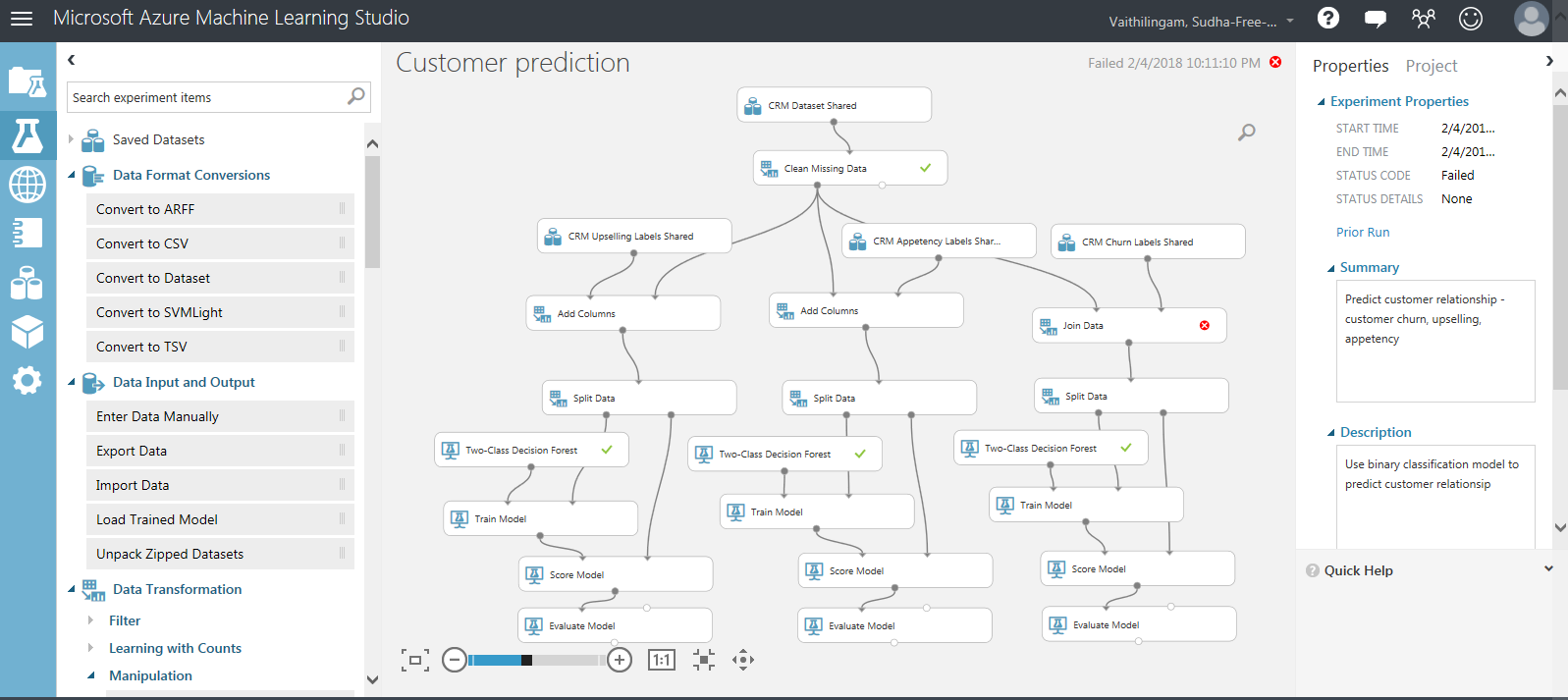
1. Add module **Evaluate model** and connect to Score Model. Do this for each of the 3 values.



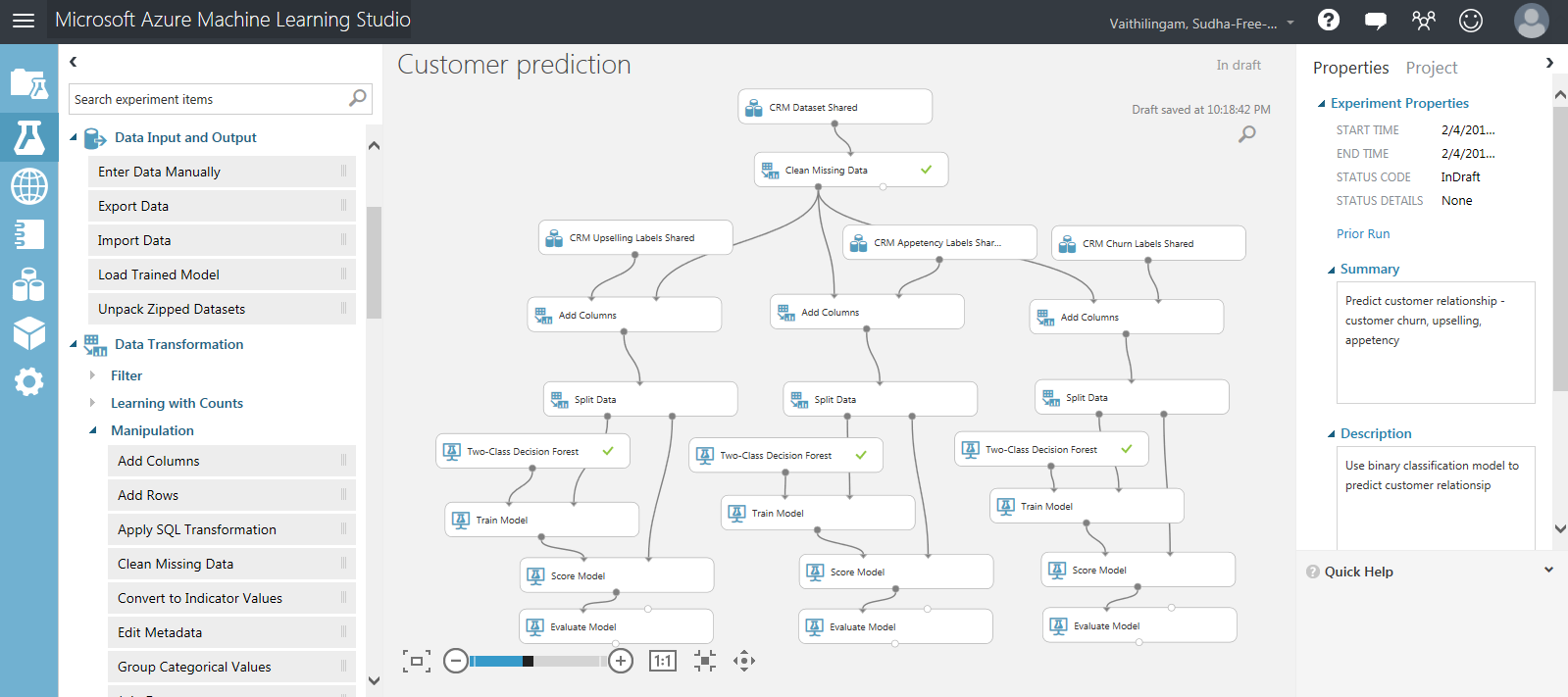
1. Save the experiment – Click on Save as at the bottom of the window and enter a name in the Save as dialog window and click OK



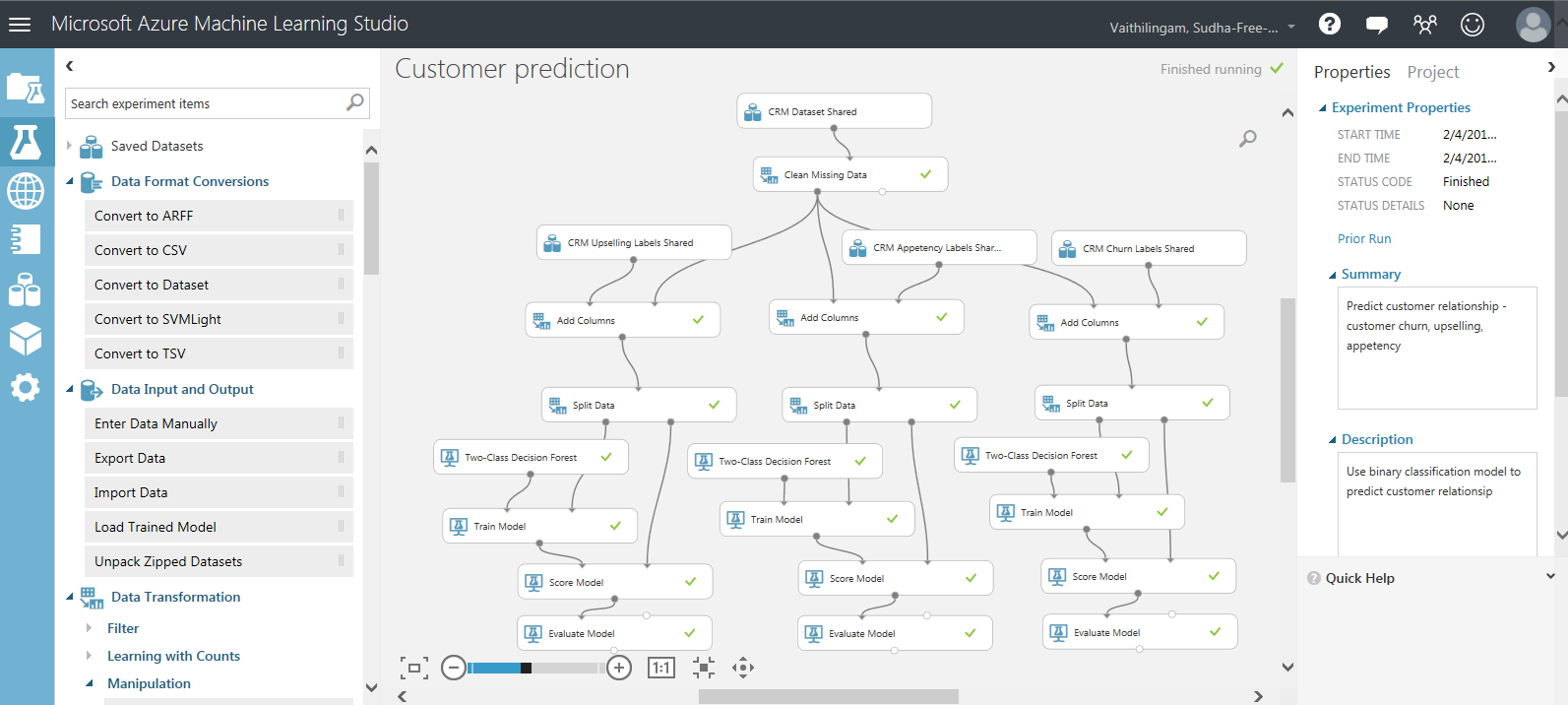
1. Run the experiment – Click on Run at the bottom of the window
2. When the experiment runs, a green check mark is displayed if the module ran successfully or a Red cross mark is displayed if there is an error.
3. When running the experiment, we see an error in Join Data module.



25.To eliminate the error, Remove Join Data module. Add **Add columns** module in its place.

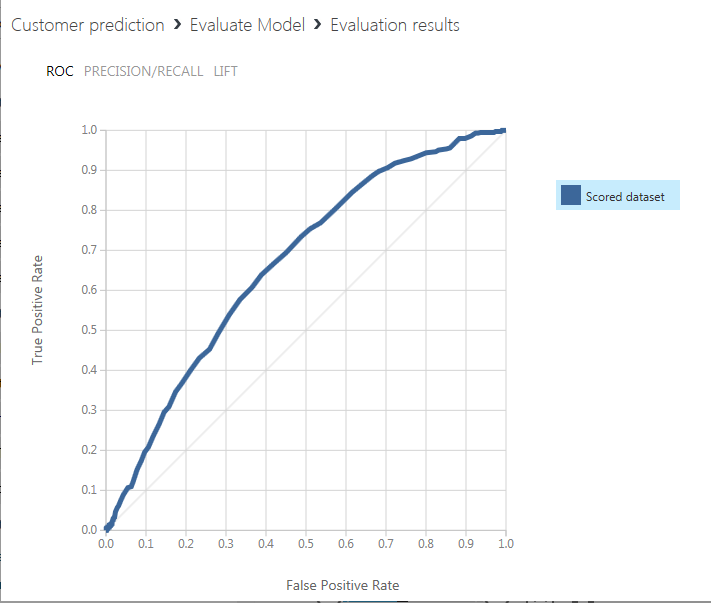
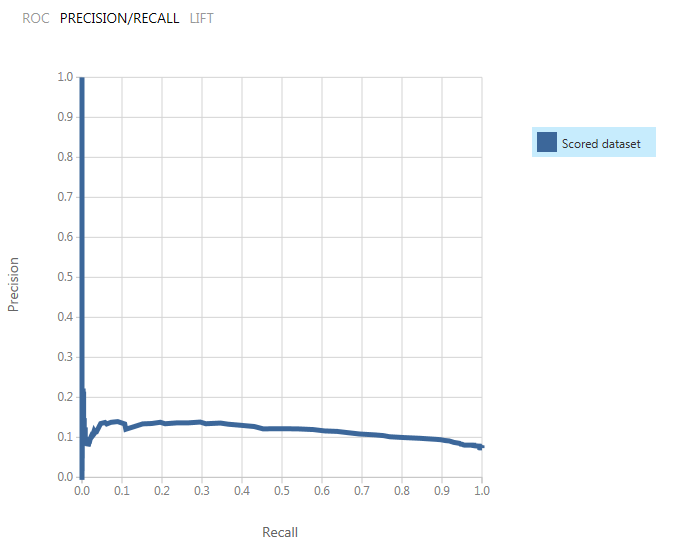


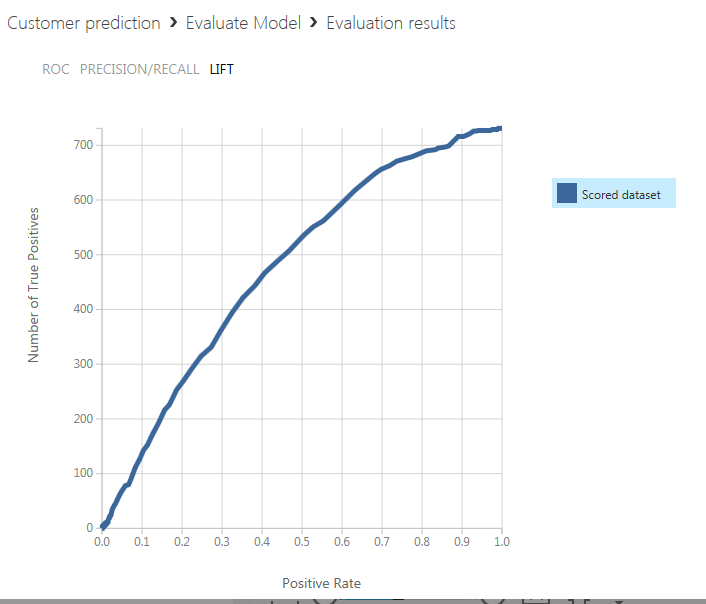
1. Click on Run to run the experiment. The experiment runs successfully. All the modules show green check mark.

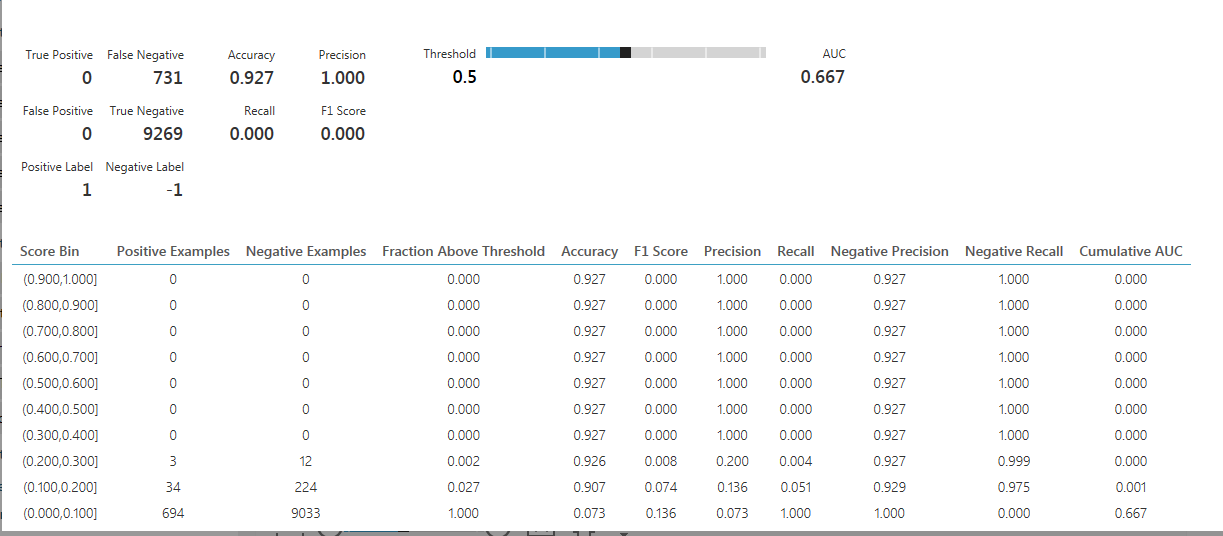


**Results**

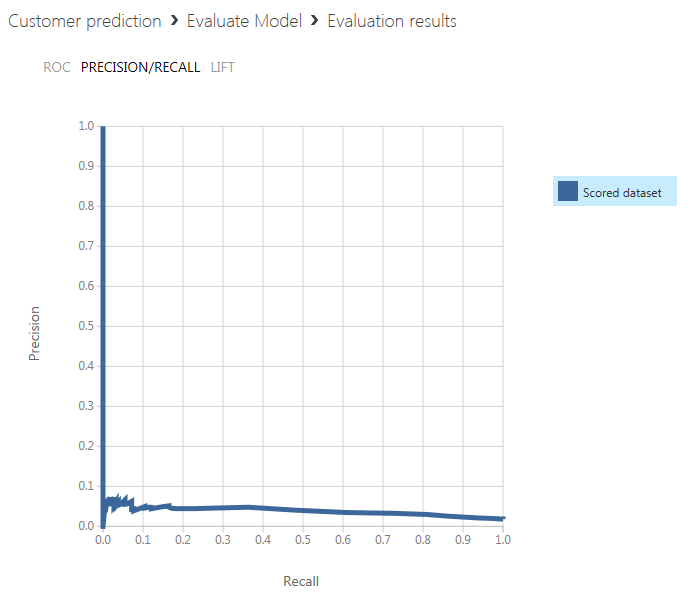
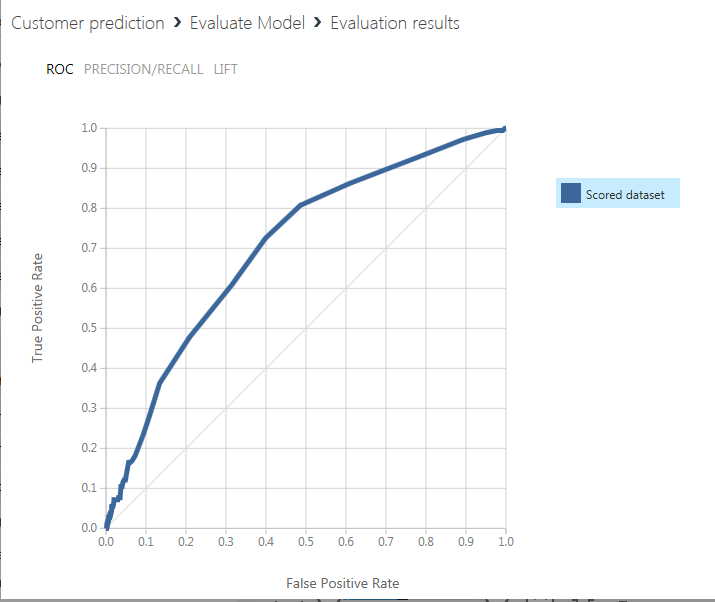
1. **For CRM Upselling**, right click on Evaluate model -> evaluation results->visualize. The metrics for the model is displayed.

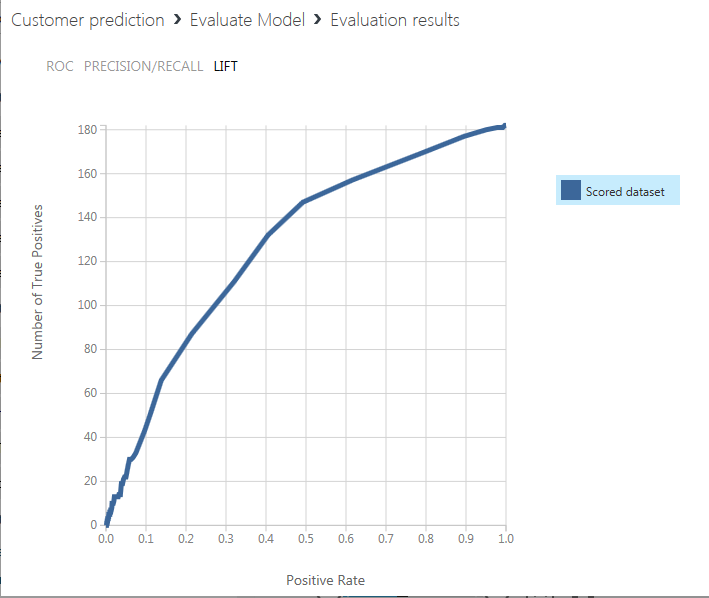
 

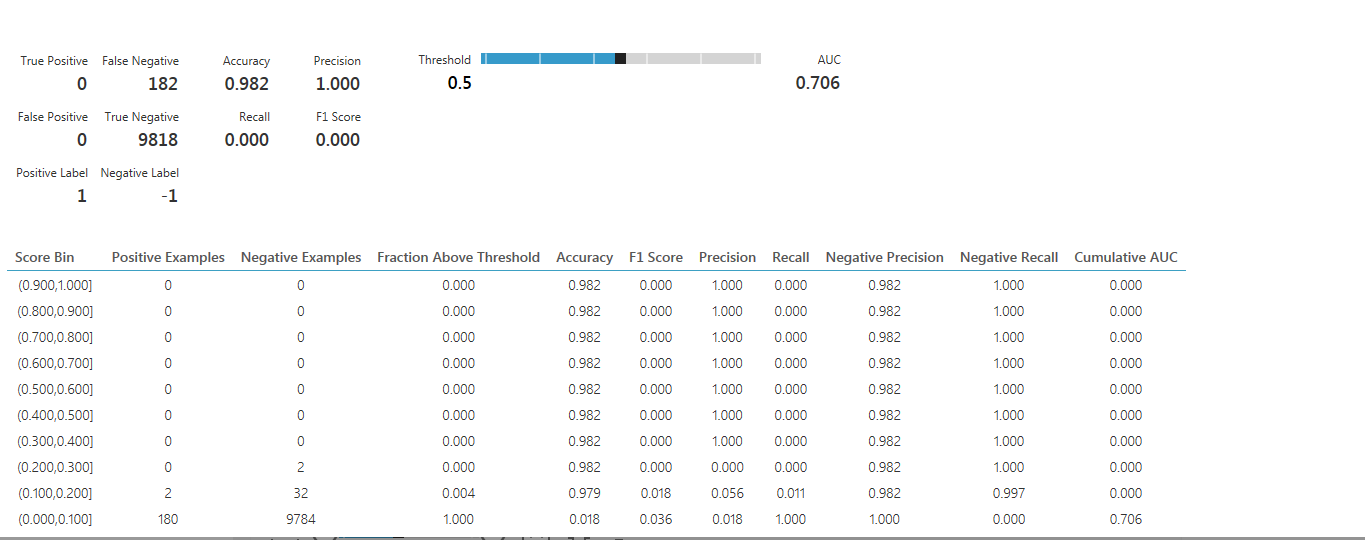




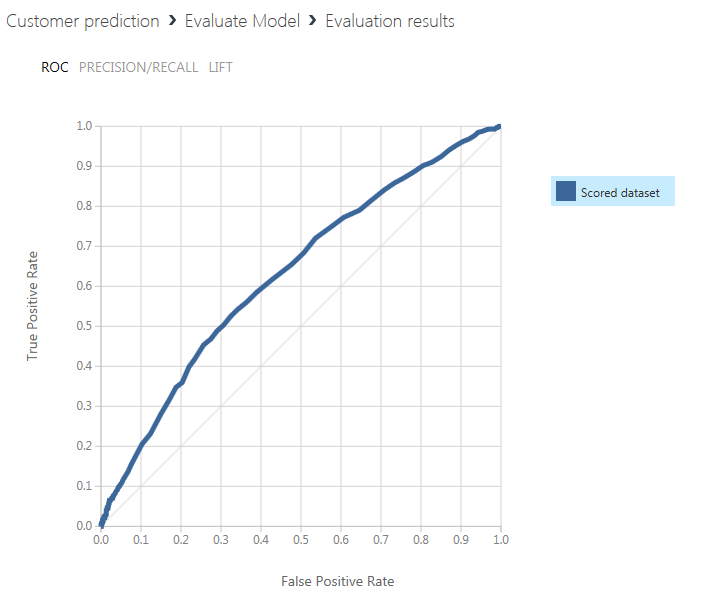
1. **For CRM Appetency**, right click on Evaluate model -> evaluation results->visualize. The metrics for the model is displayed.

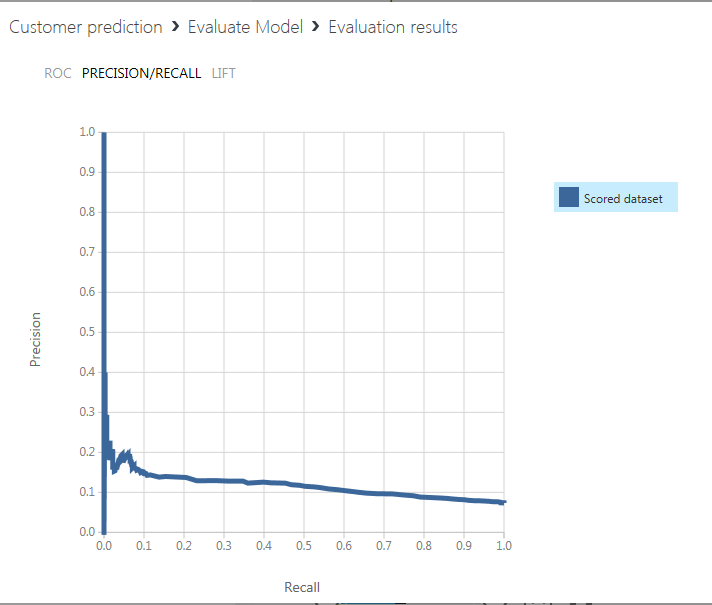


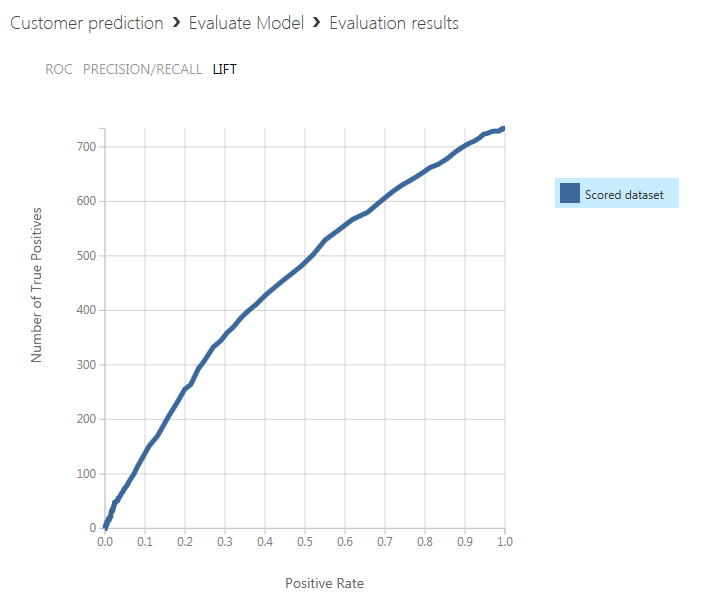


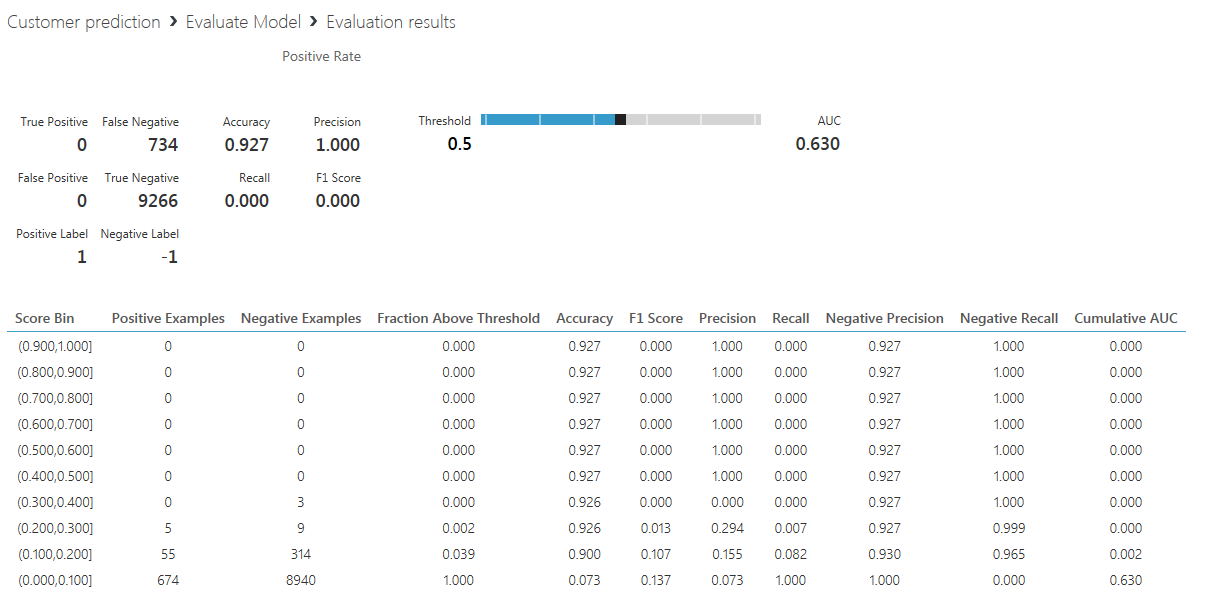


1. **For customer churn**, right click on Evaluate model -> evaluation results->visualize





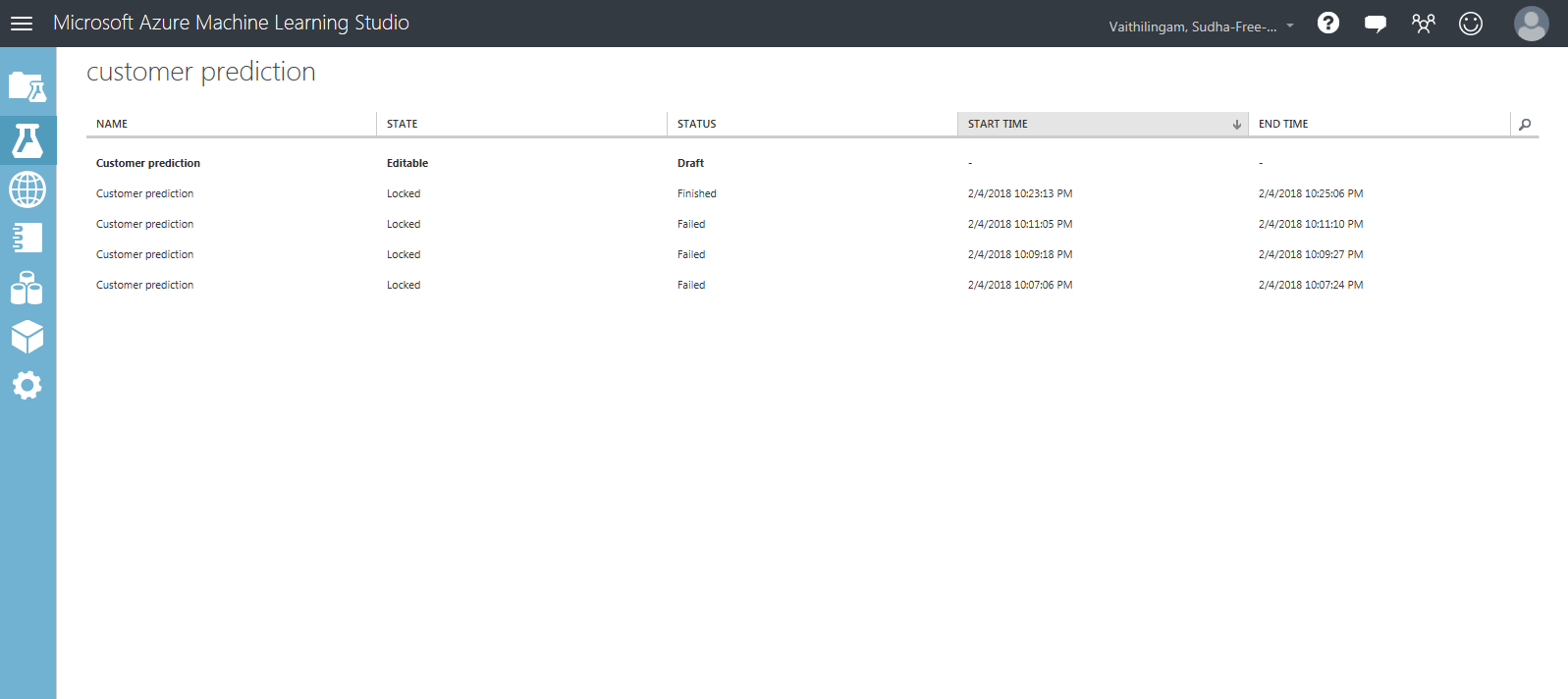




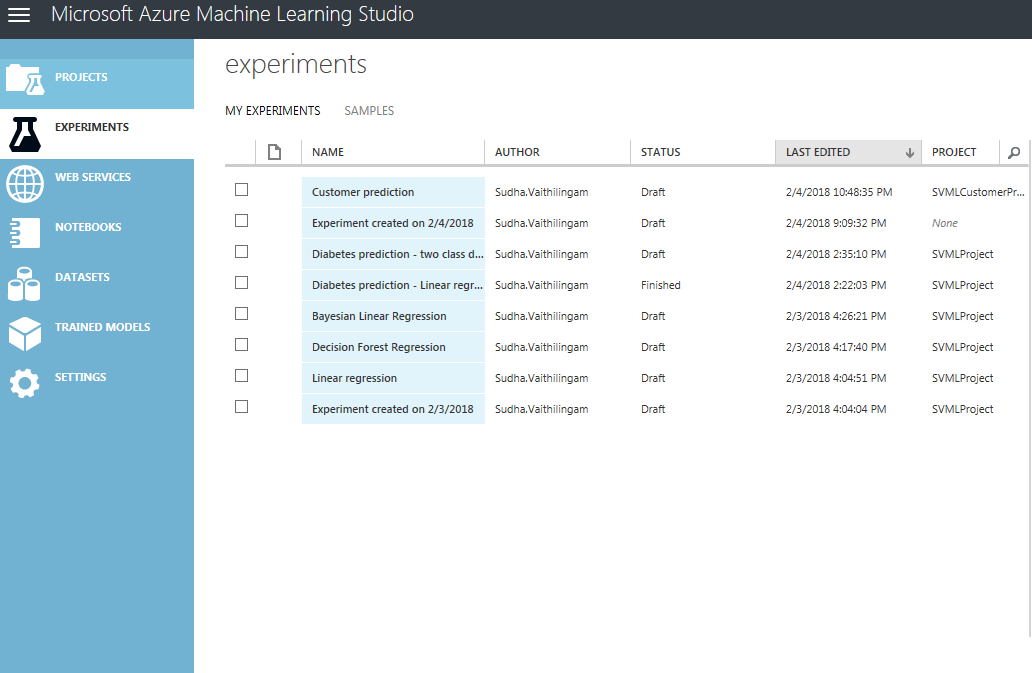
1. You can run different experiments using different algorithms (e.g. Two class boosted decision tree, two class neural network etc.), evaluate your model and select the model which has the most accuracy based on the metrics used for evaluation. I used Two class decision forest.

To run different experiments with the same dataset, it is very easy to save your experiment as a new experiment. Remove the algorithm, drag and drop a new algorithm, connect it to train model and run the experiment. It is as simple as that.

You can see all the different runs of your experiment by selecting Run History



31. You can see all your experiments by selecting the Experiments



**Lessons Learned & Pros/Cons**

In summary, I was able to create a Customer Relationship Prediction model successfully using Azure Machine Learning studio.

**Pros:**

* Azure Machine learning Studio has Easy to use UI which allows you create experiments easily by dragging and dropping modules to canvas. No coding is required.
* Several modules are available for data transformation, machine learning, evaluating model
* Running experiments was fast – experiments ran in a matter for minutes.
* You can change the algorithm and run experiments easily to select the best model for the problem statement.
* Definitely a good tool for beginners in Machine Learning.
* Good tool for data scientists who may not be technologists.

**Cons:**

* Modules are a black box.
* Deploying trained model as a web service did not work as I expected.

**Next steps**

As a next step, I will deploy my predictive model as a web service. The customer prediction web service will take customer profile as input and using the predictive model that I have built will predict customer churn (Yes or No), upselling (Yes or No) and Appetency (Yes or No)

**YouTube URLs:**

**GitHub:**

**References:**

<https://docs.microsoft.com/en-us/azure/machine-learning/studio/>

<https://docs.microsoft.com/en-us/azure/machine-learning/studio/basics-infographic-with-algorithm-examples>

<https://gallery.cortanaintelligence.com/>

<http://www.kdd.org/kdd-cup/view/kdd-cup-2009/Data>

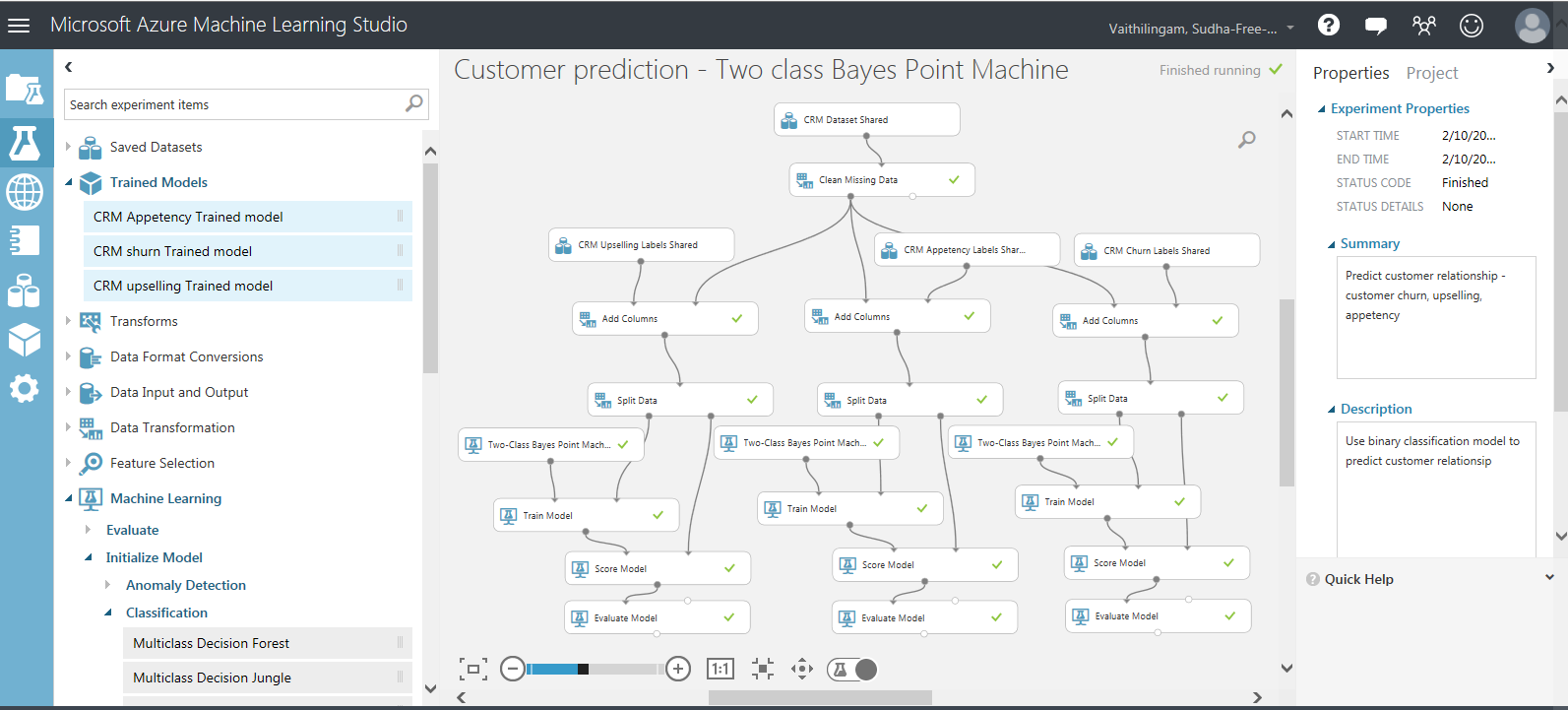
David Chapell.(2015). Introducing Azure Machine Learning A guide for technical professionals. Retrieved from <http://download.microsoft.com/download/3/B/9/3B9FBA69-8AAD-4707-830F-6C70A545C389/Introducing_Azure_Machine_Learning.pdf>

AzureML team for Microsoft. (2014). Binary classification: Customer relationship prediction. Retrieved from <https://gallery.cortanaintelligence.com/Experiment/Binary-Classification-Customer-relationship-prediction-1>

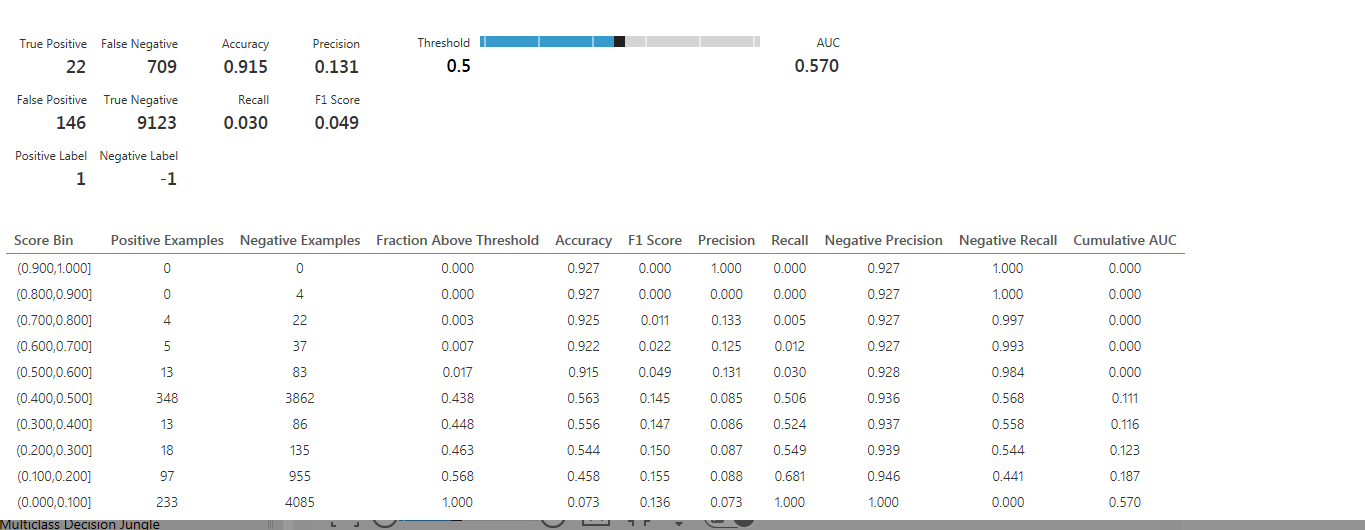
**Appendix**

Other algorithms explored

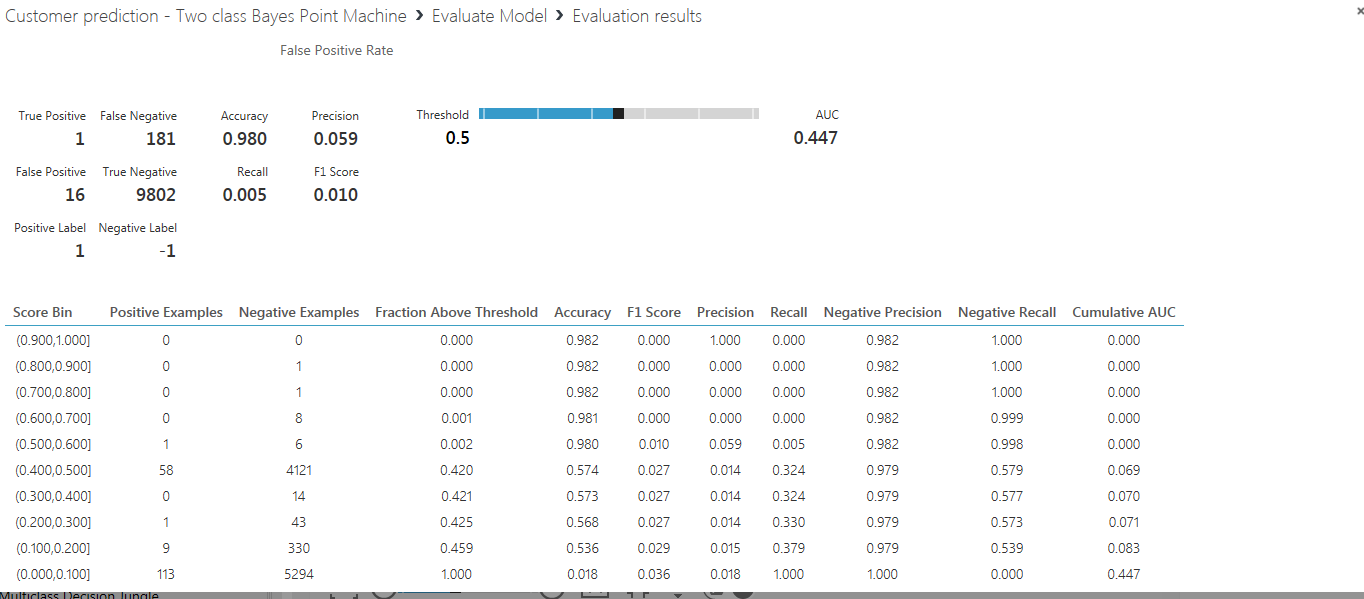
**Two class Bayes point Machine**



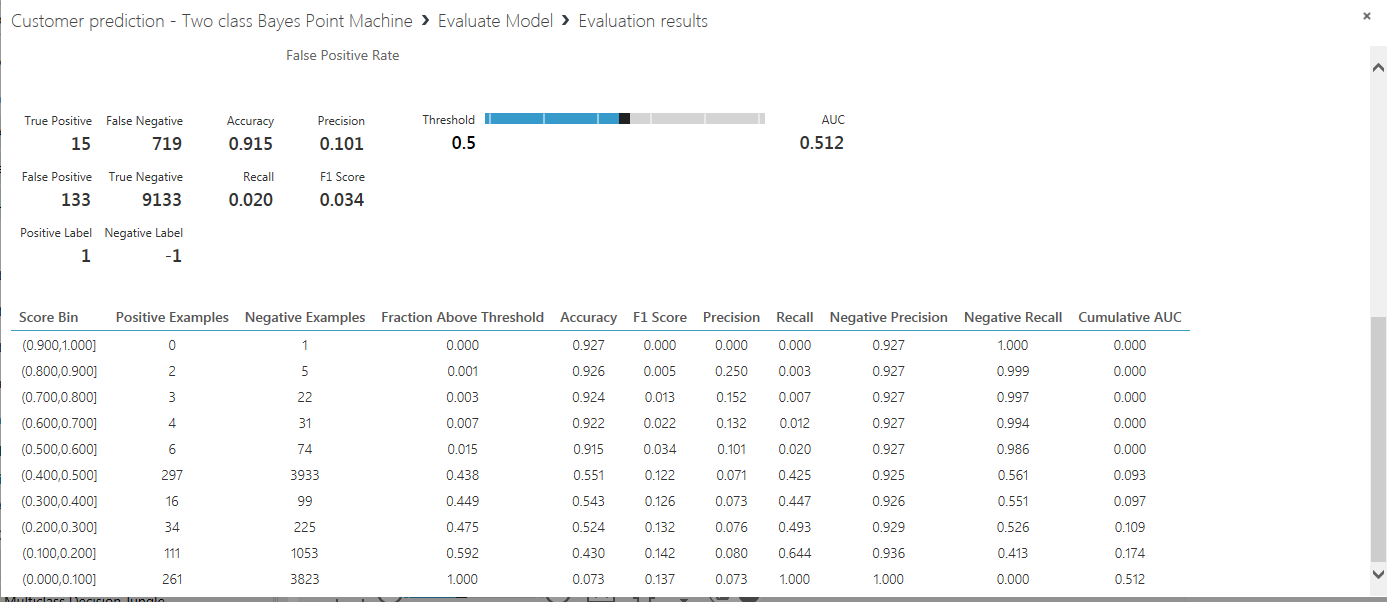
Evaluate Model – Results - CRM upselling



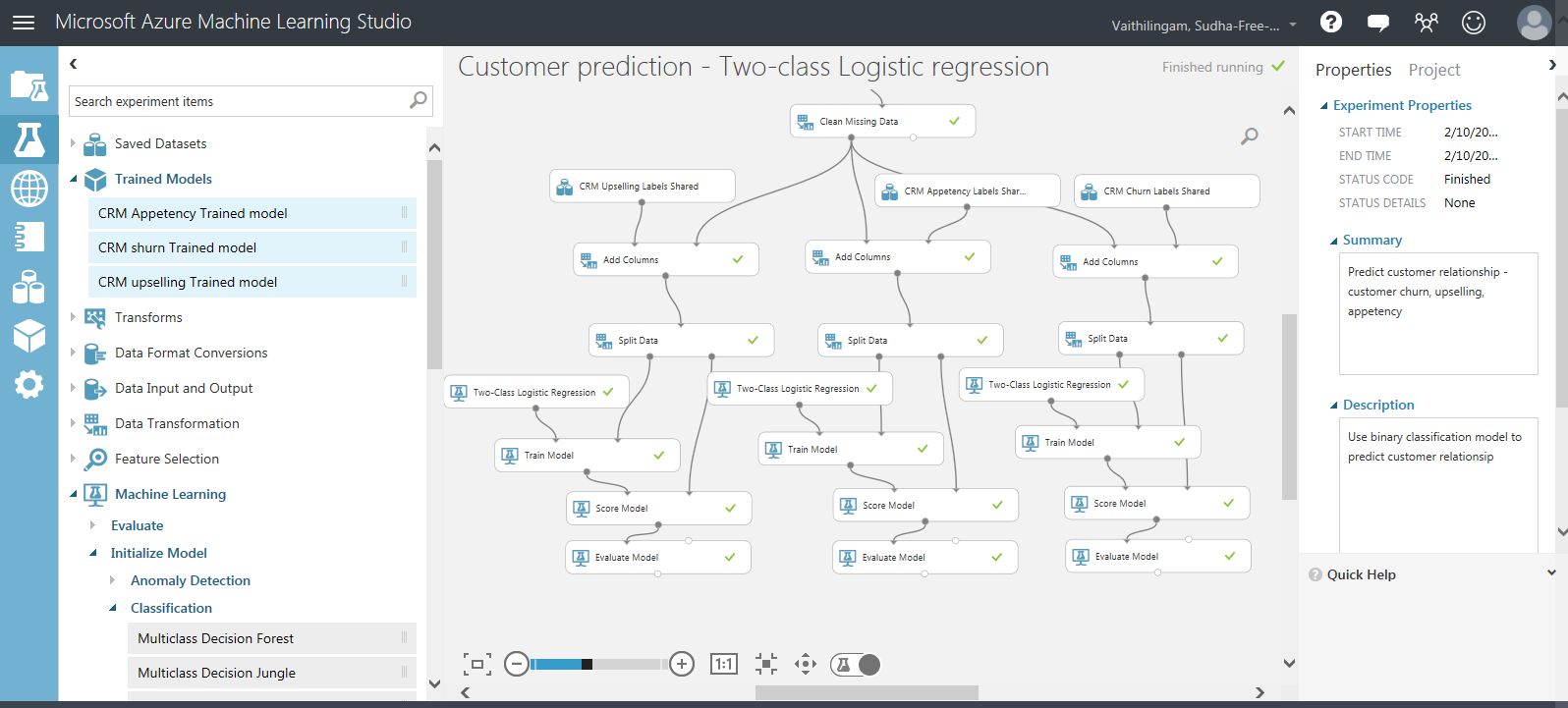
Evaluate Model – Results - CRM Appetency



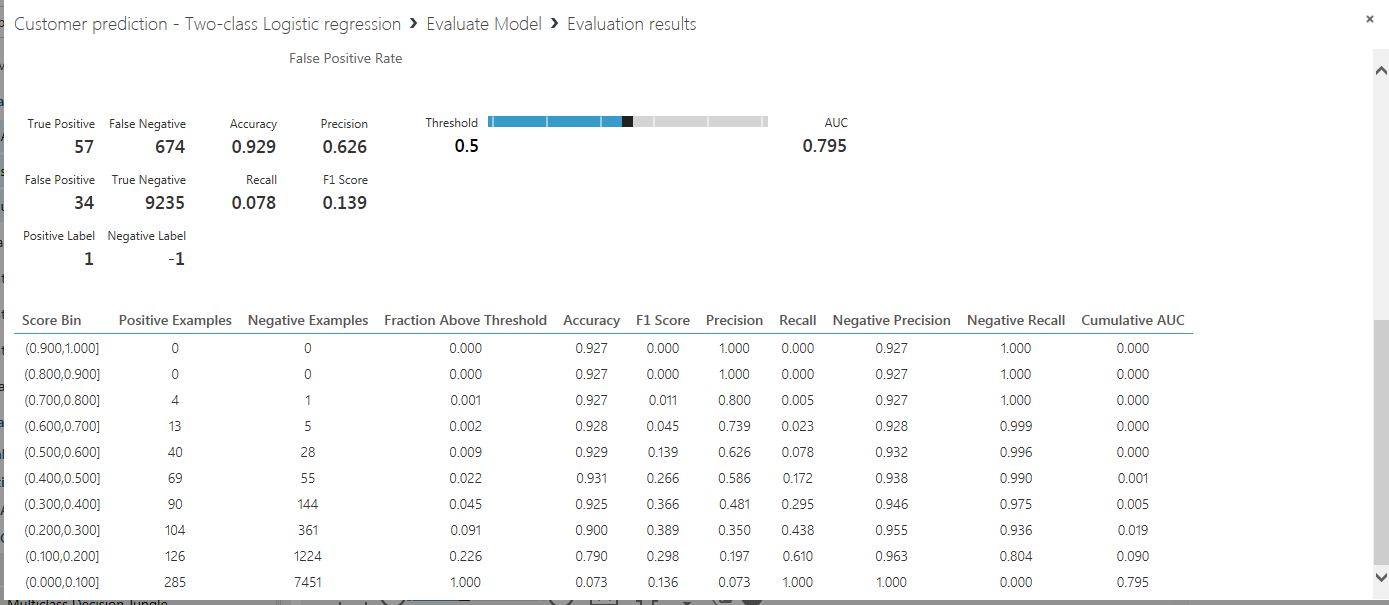
Evaluate Model – Results - CRM Churn



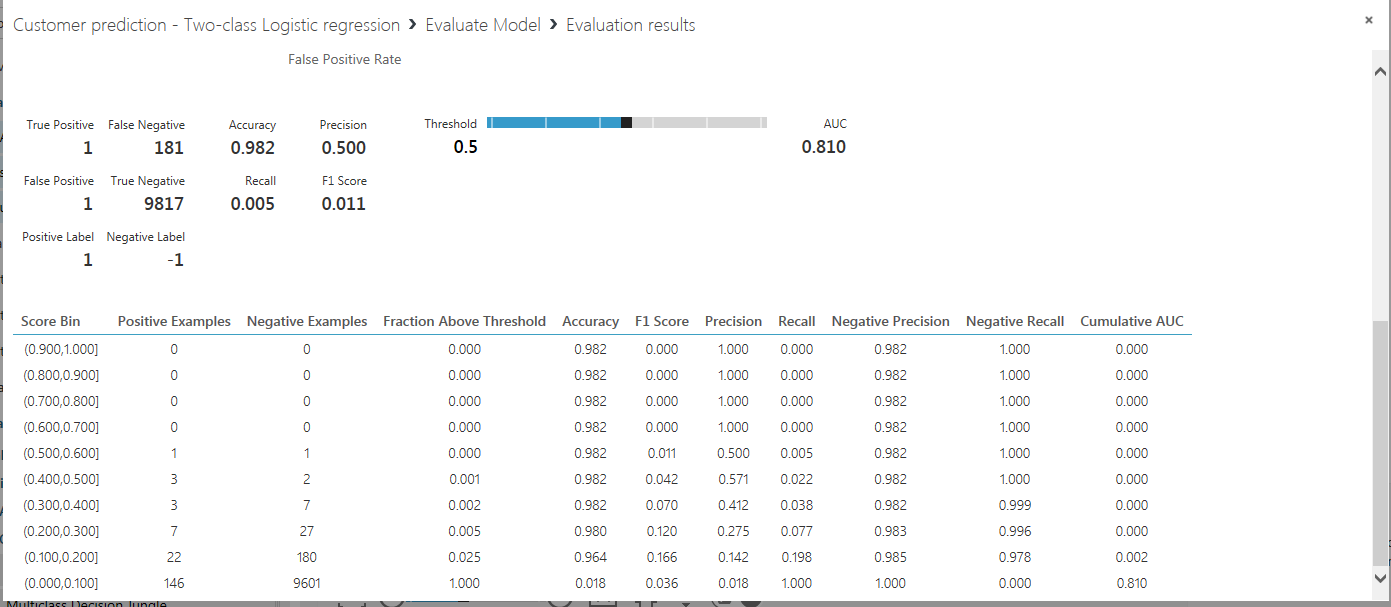
**Two class Logistic regression**



Evaluate Model – Results – CRM upselling



Evaluate Model – Results – CRM Appetency



Evaluate Model – Results – CRM Churn

