



Hands on tutorial

A Classification Model in Azure ML

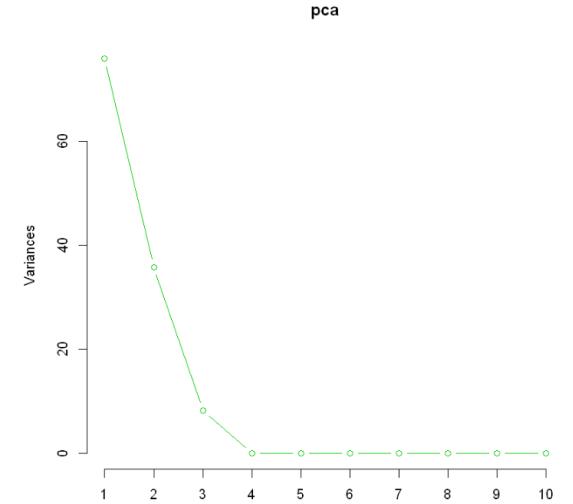
Tutorial #2 Questions:

- What are we trying to accomplish?
 - Given these dynamometer card readings, can we classify the operations into failed or not??
- How can we use Azure Machine Learning to answer these questions?



Scenario Data:

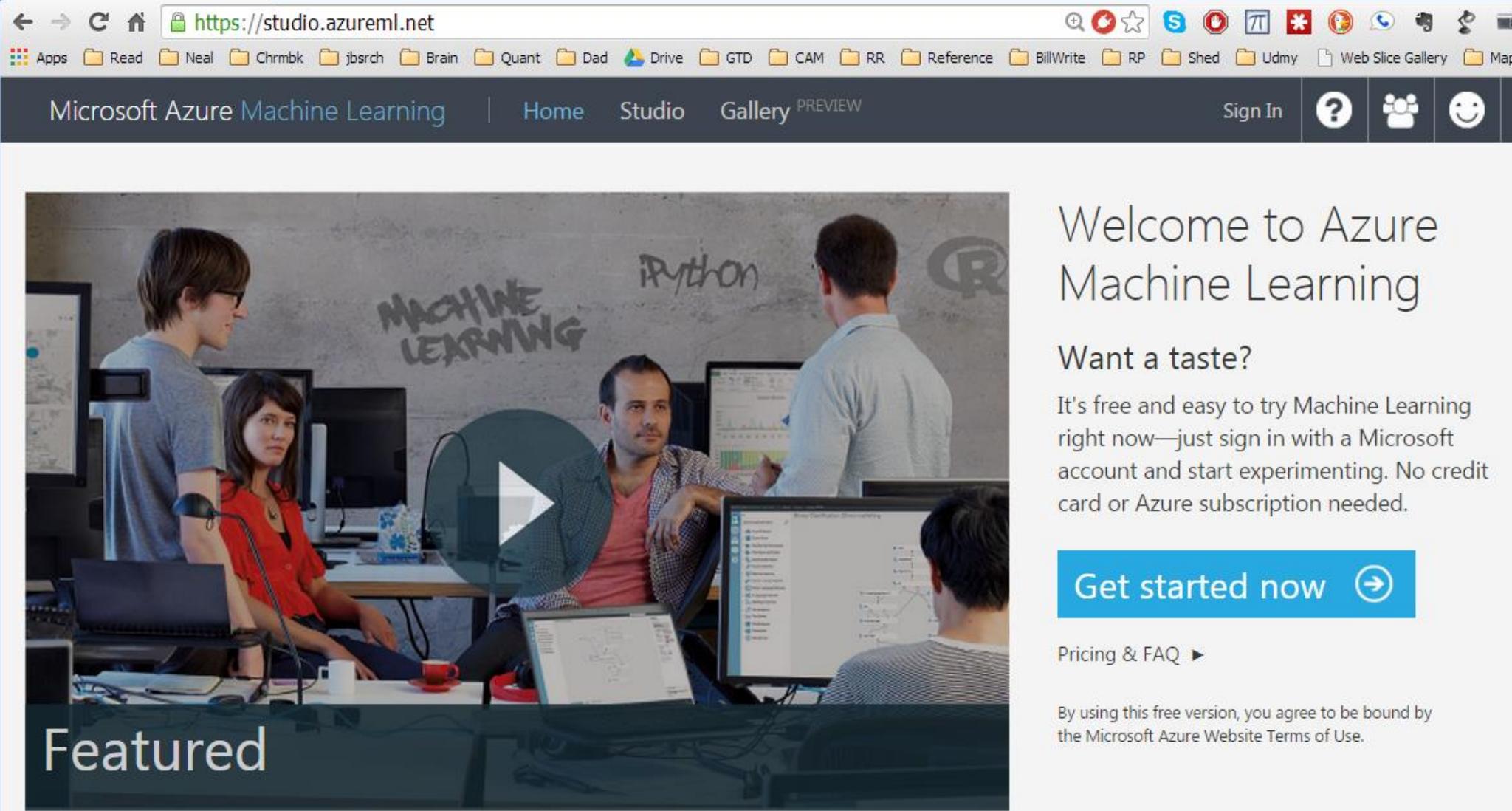
- After the PCA was performed we ended up with ~57K rows of data with 3 Principal Components (PC1, PC2, PC3)
- 263 unique wells
- We merged the new dataset with the operation failure dataset. (Already done that for the purpose of this tutorial)
- Training data now comprises of the following:
 - Date-Time at which the readings were taken
 - PC1
 - PC2
 - PC3
 - Assignments



rows 57883 columns 51

	well	date	PC1	PC2	PC3	Assignments
view as	Alabama	2013-08-16T01:55:22	-8.243341	2.918071	2.023029	35
	Alabama	2013-08-16T23:49:23	-8.243341	2.918071	2.023029	35
	Alabama	2013-09-03T00:17:00	-8.495328	2.311737	2.154067	35
	Alabama	2013-09-06T09:38:47	-8.148383	2.731808	2.341957	35
	Alabama	2013-09-06T16:13:14	-8.148383	2.731808	2.341957	35

Step 1 : Go to <https://studio.azureml.net/>

The screenshot shows the Microsoft Azure Machine Learning Studio homepage. At the top, there's a navigation bar with links for 'Home', 'Studio', and 'Gallery PREVIEW'. On the right side of the nav bar are 'Sign In' and three icons: a question mark, a user group, and a smiley face. Below the nav bar is a large image of four people working at desks in an office, with a chalkboard in the background featuring the words 'MACHINE LEARNING', 'Python', and 'R'. To the left of this image, the word 'Featured' is displayed. To the right of the image, the text 'Welcome to Azure Machine Learning' is followed by 'Want a taste?'. Below this is a paragraph about the free trial. A blue button labeled 'Get started now' with a right-pointing arrow is centered. At the bottom, there's a link to 'Pricing & FAQ' with a right-pointing arrow, and a note about agreeing to the Terms of Use.

Welcome to Azure Machine Learning

Want a taste?

It's free and easy to try Machine Learning right now—just sign in with a Microsoft account and start experimenting. No credit card or Azure subscription needed.

Get started now →

Pricing & FAQ ►

By using this free version, you agree to be bound by the Microsoft Azure Website Terms of Use.

Step 2 : Log in to your account

The screenshot shows the Microsoft Azure Machine Learning Studio interface. The title bar reads "Experiments - Microsoft Azur" and the URL is "https://studio.azureml.net/Home/ViewWorkspace/c1e94b3c2eaa4ec5a64888cba93e1fcb#Worl". The top navigation bar includes links for Home, Studio, and Gallery, along with a "PREVIEW" button. The main content area is titled "experiments" and features a sidebar with icons for EXPERIMENTS, WEB SERVICES, DATASETS, TRAINED MODELS, SETTINGS, and a NEW button. The main pane displays a table header with columns for NAME, AUTHOR, STATUS, and LA... (with a dropdown arrow), followed by a search icon. Below the header, the message "No experiments found" is displayed, and to the right, it says "0 items selected". A "DELETE" button with a trash can icon is located at the bottom center.

- ▶  Saved Datasets
- ▶  Data Format Conversions
- ▶  Data Input and Output
- ▶  Data Transformation
- ▶  Feature Selection
- ▶  Machine Learning
- ▶  OpenCV Library Modules
- ▶  Python Language Modules
- ▶  R Language Modules
- ▶  Statistical Functions
- ▶  Text Analytics
- ▶  Deprecated
- ▶  Web Service

Step 3.1 : Create New Experiment

- AML modelling ... a checklist approach
 - Create new experiment

Step 3.2 : Create a new experiment

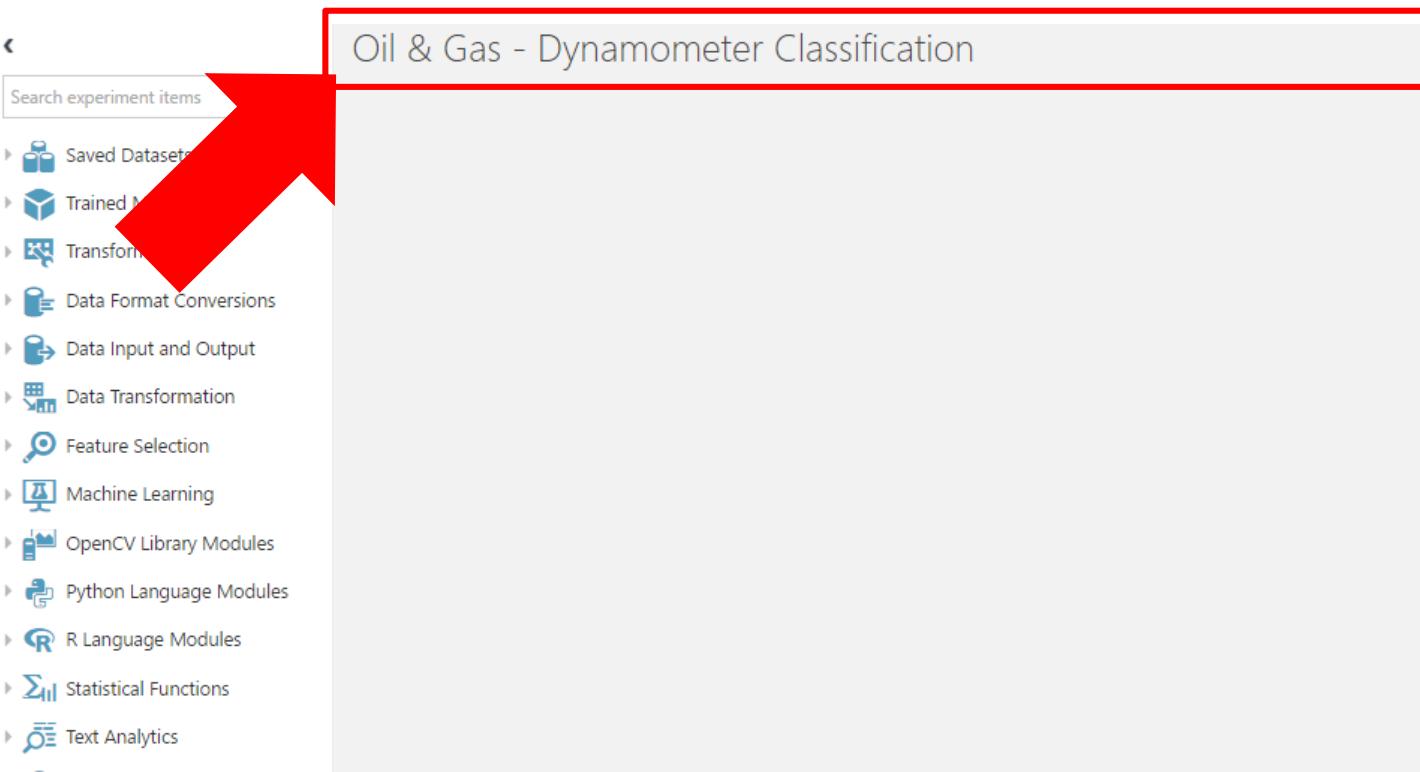
The screenshot shows the Microsoft Azure Machine Learning Studio interface. The title bar reads "Experiments - Microsoft Azur" and the URL is "https://studio.azureml.net/Home/ViewWorkspace/c1e94b3c2eaa4ec5a64888cba93e1fc#.Worl". The top navigation bar includes links for Apps, Read, Neal, Chrbmk, jbsrch, Brain, Quant, Dad, Drive, GTD, CAM, RR, Reference, BillWrite, RP, Shed, Udmv, Web Slice Gallery, and Maps. Below the navigation bar is a dark header with "Microsoft Azure Machine Learning" and "PREVIEW" buttons, along with icons for Home, Studio, and Gallery.

The main content area is titled "experiments". It features a sidebar on the left with icons for EXPERIMENTS (test tube), WEB SERVICES (globe), DATASETS (cylinders), TRAINED MODELS (cube), SETTINGS (gear), and NEW (plus sign). A large red arrow points to the "NEW" button at the bottom of the sidebar. The main panel has tabs for "MY EXPERIMENTS" and "SAMPLES". A table header with columns "NAME", "AUTHOR", "STATUS", and "LA... ▾" is shown, followed by a message "No experiments found" and a count "0 items selected".

Step 3.3 : Create a new experiment

The screenshot shows the Microsoft Azure Machine Learning interface. At the top, there's a navigation bar with 'Microsoft Azure Machine Learning' and a menu icon. Below it, a sidebar has 'PROJECTS' and 'EXPERIMENTS' tabs, with 'EXPERIMENTS' selected. The main area is titled 'experiments' and shows 'MY EXPERIMENTS' and 'SAMPLES' sections. A search bar says 'Search experiment templates'. On the left, a vertical sidebar lists 'NEW' items: 'DATASET', 'MODULE', 'PROJECT PREVIEW', 'EXPERIMENT' (which is highlighted in grey), and 'NOTEBOOK PREVIEW'. Two red arrows point to the 'EXPERIMENT' item in the sidebar: arrow 1 points to the sidebar itself, and arrow 2 points to the 'Blank Experiment' button.

Step 3.4 : Name your experiment



The screenshot shows the Azure Machine Learning studio interface. On the left, there's a sidebar with various module categories like Saved Datasets, Trained Models, and Data Format Conversions. The main area displays an experiment titled "Oil & Gas - Dynamometer Classification". A large red arrow points from the top-left towards the experiment title, and a red rectangular box highlights the title itself.

- Click on the title box at the top that says "Experiment Created on"
- Give the experiment the following title:
"Oil & Gas –
Dynamometer
Classification"

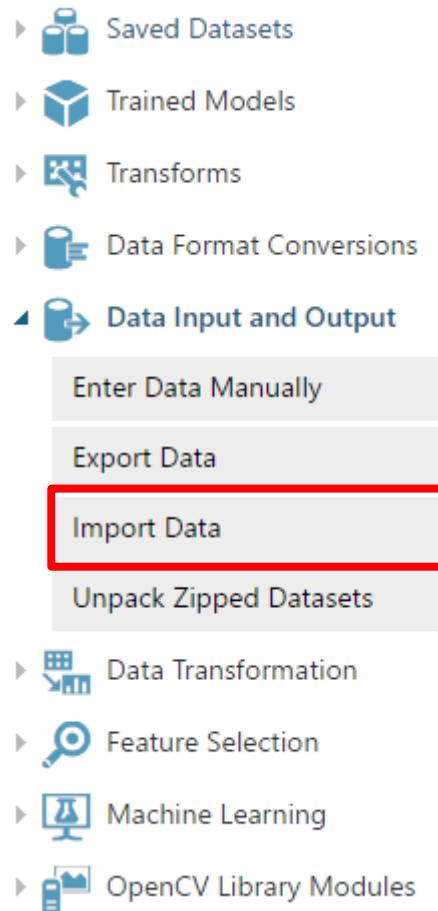
- ▶  Saved Datasets
- ▶  Data Format Conversions
- ▶  Data Input and Output
- ▶  Data Transformation
- ▶  Feature Selection
- ▶  Machine Learning
- ▶  OpenCV Library Modules
- ▶  Python Language Modules
- ▶  R Language Modules
- ▶  Statistical Functions
- ▶  Text Analytics
- ▶  Deprecated
- ▶  Web Service

Step 4.1 : Import the dataset

- AML modelling ... a checklist approach
 - Create new experiment
 - Import data set

Option 1 -
Import the training dataset
from the Azure Blob Storage. If
option 1 did not work, skip to
option 2.

Step 4.2 : Q? How to import data from Azure Blob?

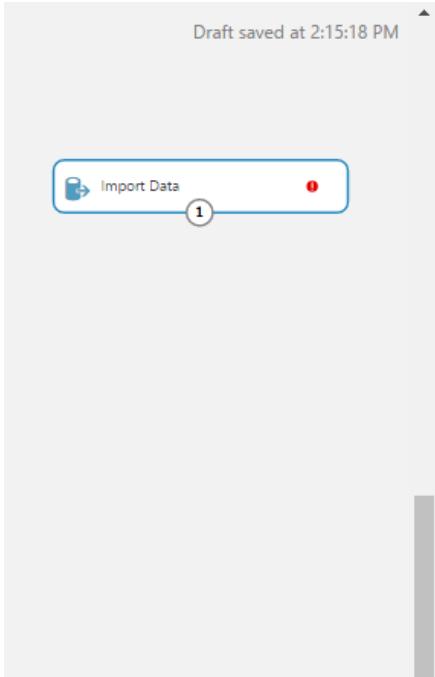


- Open “Data Input and Output” from the navigation pane at the left
- Drag “Import Data” to the canvas
 - “Import Data” loads data from sources such as the Web, Azure SQL, Windows Azure Blob storage, etc



Step 4.3 : Q? How to import data from Azure Blob?

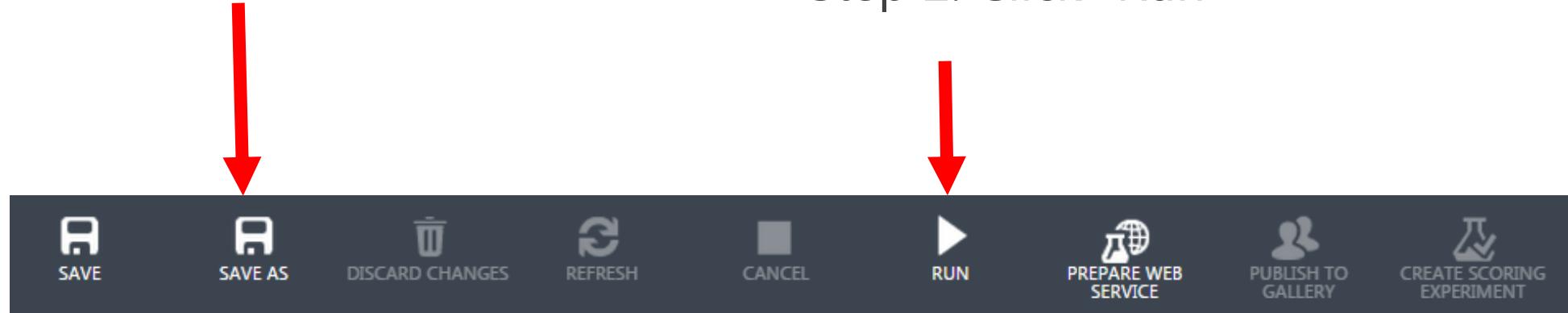
- Machine Learning
 - Click on Import Data
 - Chose “Azure Blob Storage” for Data source
 - Chose “Storage Account” for Authentication type
 - Copy and Paste the following information without the quotes in the Import Data module
 - Account Name – **“nealworkshop”**
 - Account key –
“RER9c7kfM1e67p7p7gl+TbkE5Y6alzURg4PQc9Iew+l8O+ZfU58gFjNgBW/WQm0u8N0YZQUG+wlalzfWKxyljA==”
 - Path to container – **“/oilandgas/Dynamometer Clusters.csv”**
 - Check the File has header row check box



Step 4.4 : Q? How to import data from Azure Blob?

Step 1: Click
"Save"

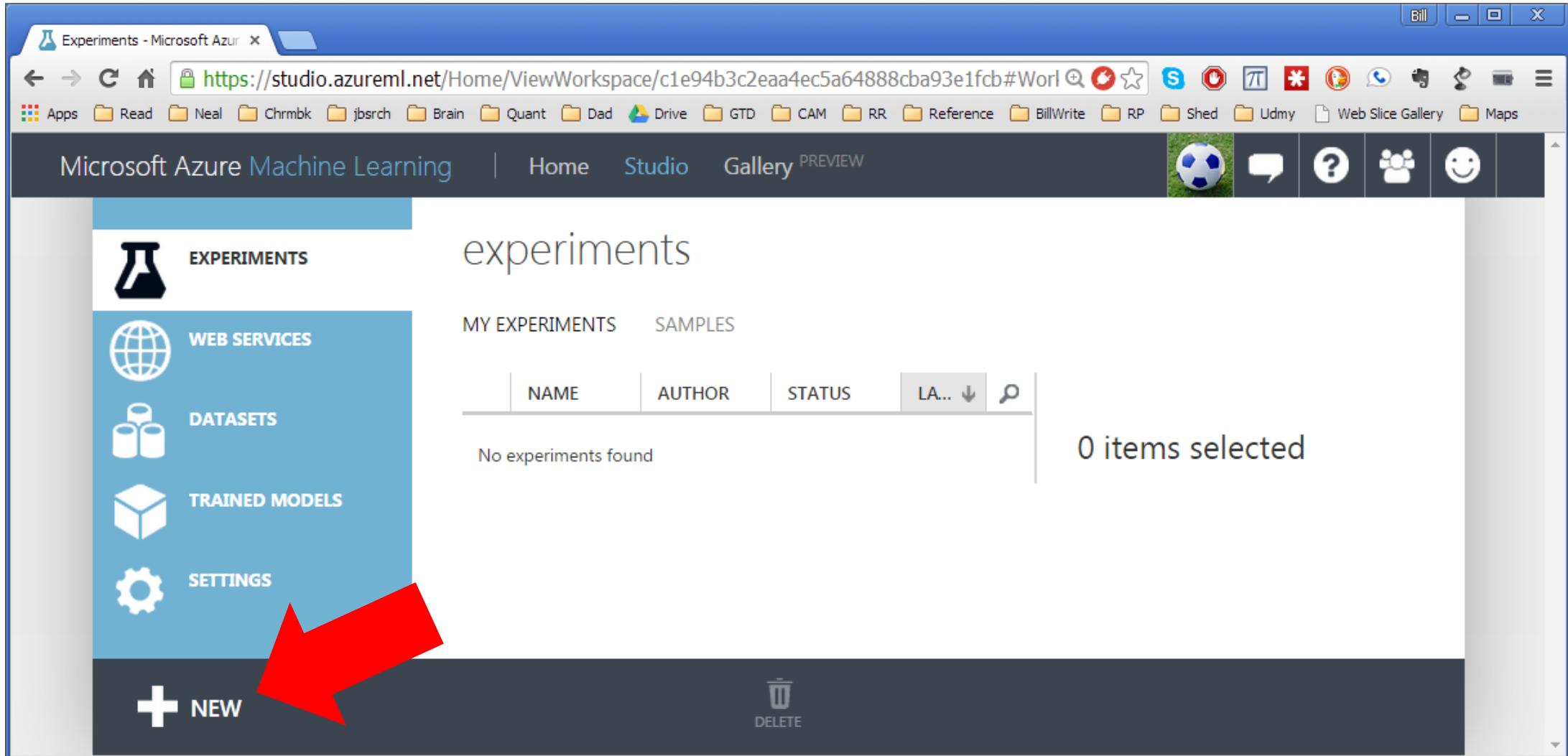
Step 2: Click "Run"



If Option 1 worked, move on to
Step 4.13

Option 2 -
Import the training dataset
from the saved datasets.

Step 4.5 : Import this tutorial's training dataset



Step 4.6 : Import this tutorial's training dataset

The screenshot shows the Microsoft Azure Machine Learning interface. At the top, there is a dark header bar with the text "Microsoft Azure Machine Learning". Below this is a navigation bar with two main tabs: "PROJECTS" (highlighted in blue) and "EXPERIMENTS" (highlighted in white). Underneath the tabs, there are two sub-tabs: "MY EXPERIMENTS" and "SAMPLES". The main content area is titled "experiments". It features a search bar with the placeholder "NAME" and a "FROM LOCAL FILE" button. On the left, there is a sidebar with the heading "NEW" followed by a list of options: "DATASET" (selected), "MODULE", "PROJECT PREVIEW", "EXPERIMENT", and "NOTEBOOK PREVIEW". A large red arrow points from the text "Import this tutorial's training dataset" in the previous slide towards the "DATASET" option in the sidebar.

Step 4.7 : Import this tutorial's training dataset

The screenshot shows the Microsoft Azure Machine Learning Studio interface. A red arrow points to the 'DATASETS' icon in the left sidebar, which is highlighted. A modal window titled 'Upload a new dataset' is open, prompting the user to 'SELECT THE DATA TO UPLOAD' via a 'Choose File' button. Below this, there is a checkbox for 'This is the new version of an existing dataset'. The main workspace shows a list of datasets:

Name	Date modified	Type	Size
Dynamometer Clusters	9/27/2016 11:40 AM	Microsoft Excel Comma Separated Values File	47,740 KB
2. Azure Machine Learning Building A Classification Model	9/26/2016 2:15 PM	Microsoft PowerPoint Presentation	4,536 KB

Step 4.8 : Import this tutorial's training dataset

Microsoft Azure Machine Learning Studio

Advanced-Workshop ? 🔍 😊 🙋

experiments

MY EXPERIMENTS SAMPLES

	NAME	AUTHOR	STATUS	LAST EDITED	PROJECT
Oil & Gas - Dynamometer Clusters	achal_mallaya	Draft	9/28/2016 11:15:40 AM	None	
Oil & Gas - Brine Analysis	sailaja.karthik	Finished	9/28/2016 9:53:03 AM	None	
OK Training - Tank Level	sailaja.karthik	Finished	9/28/2016 9:24:37 AM	None	
Oil & Gas - Binary Classification	achal_mallaya	Draft	9/28/2016 5:33:52 AM	Oklahoma Training Work...	
Oil & Gas - Linear Regression	achal_mallaya	Finished	9/28/2016 5:26:28 AM	Oklahoma Training Work...	
Oil & Gas - Linear Regression	achal_mallaya	Draft	9/28/2016 5:25:59 AM	Oklahoma Training Work...	
Oil & Gas - Dynamometer Clusters	achal_mallaya	Finished	9/27/2016 6:33:37 PM	None	
Oil & Gas - Dynamometer Clusters	achal_mallaya	Draft	9/27/2016 6:33:25 PM	None	
Oil & Gas - Linear Regression	Ryan	Draft	9/27/2016 4:33:20 PM	None	
OK Cluster, Classify, Reg...	eric.hullander	Finished	9/26/2016 12:32:05 PM	None	
Oil & Gas - Dynamometer Clusters	achal_mallaya	Draft	9/26/2016 12:30:40 PM	Oklahoma Training Work...	
Brine Analysis_K Means Clustering	sailaja.karthik	Draft	9/26/2016 4:00:55 AM	None	
OK - K Means Clustering	sailaja.karthik	Finished	9/26/2016 3:54:17 AM	None	
OK Predictive Maintenance	eric.hullander	Draft	9/22/2016 2:02:02 PM	OK Oil and Gas Workshop	
Oil & Gas - Binary Classification	sailaja.karthik	Failed	9/20/2016 12:30:00 PM	None	
Predictive Maintenance	microsoft	Draft	9/14/2016 4:14:28 PM	None	
Aerospace - Linear Regression	microsoft	Draft	9/14/2016 12:20:47 PM	OK Oil and Gas Workshop	
Predict Fracking Success	Tyler Chessman	Finished	9/14/2016 12:11:35 PM	Aerospace Workshop	
Two-Class Decision Forest			9/8/2016 3:24:39 PM	None	

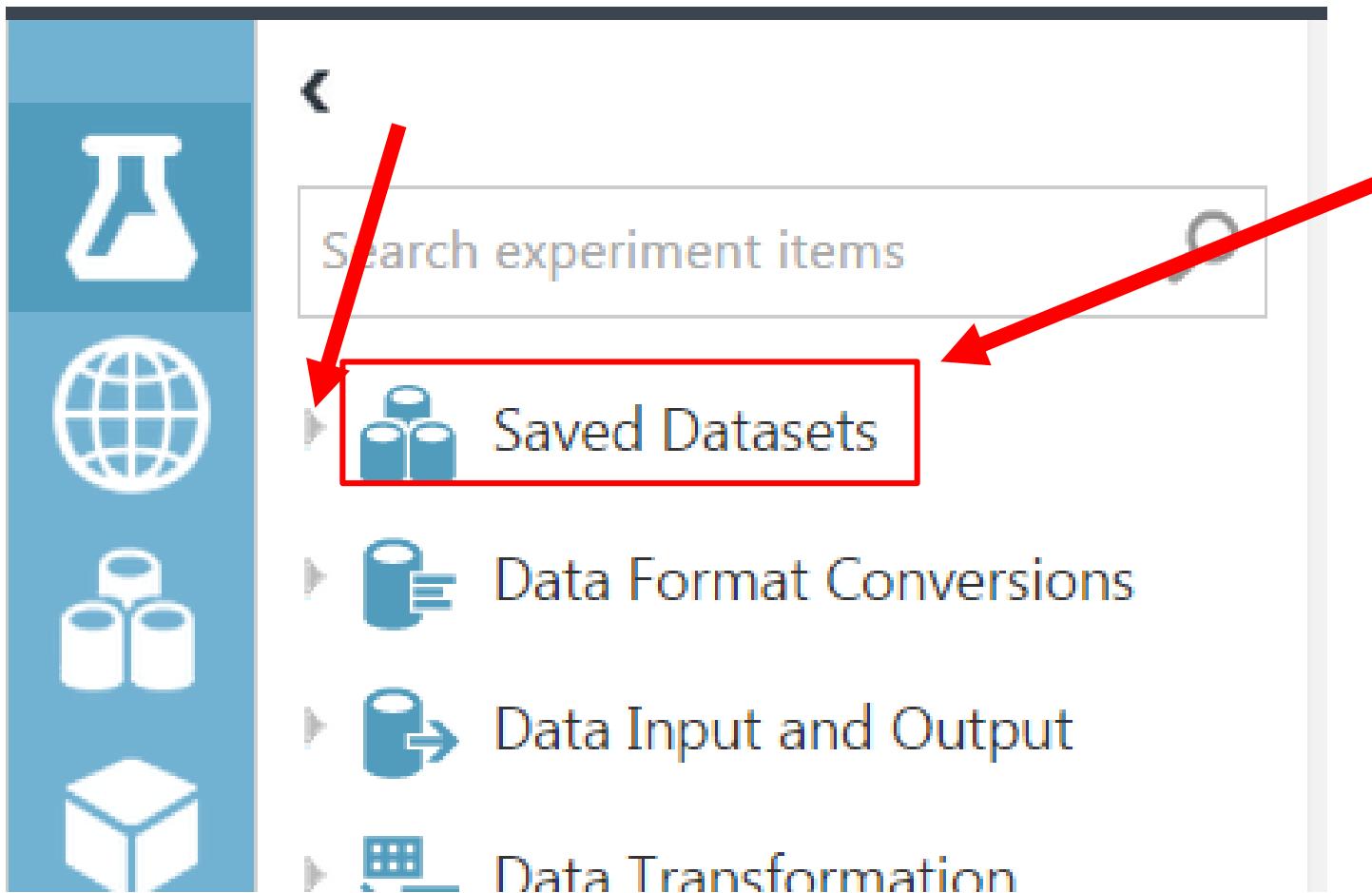
Convert to CSV Convert to Indicator Values Select Columns in Dataset Edit Metadata Two-Class Decision Forest Split Data Two-Class Neural Network Tune Model Hyperparameters Tune Model Hyperparameters Permutation Feature Importance Score Model Score Model Evaluate Model

Upload of the dataset 'Dynamometer Clusters.csv' has completed.

OK

NEW DELETE COPY TO WORKSPACE ADD TO PROJECT

Step 4.9 : Open “Saved Datasets”



- By clicking on the triangle at the left of “Saved Datasets”

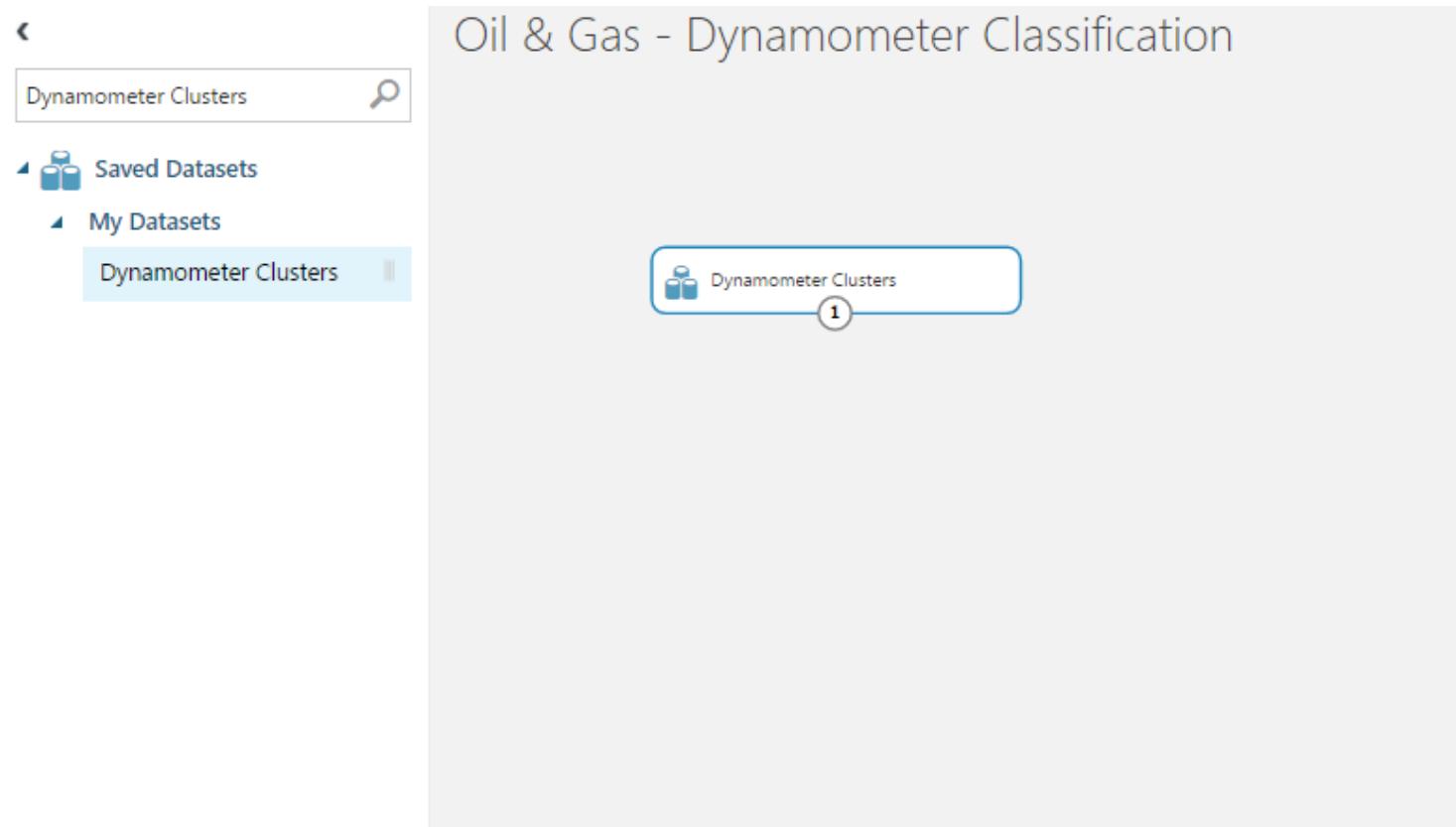
Step 4.10 : Take a second to notice the MANY datasets

A screenshot of the Azure Machine Learning Studio interface. On the left, there's a vertical toolbar with icons for experiment, workspace, datasets, and settings. The main area shows an experiment created on 3/11/20. A sidebar on the left lists 'Saved Datasets' with items like 'Adult Census Income Bin...', 'Airport Codes Dataset', 'Automobile price data (R...)', 'Bike Rental UCI dataset', 'Bill Gates RGB Image', 'Blood donation data', 'Book Reviews from Amaz...', and 'Breast cancer data'. A red arrow points from the text 'To filter to the data set for this tutorial ...' to the search bar at the top of the sidebar.

- To filter to the data set for this tutorial ...
- Type "Dynamometer Clusters" in the "Search experiment items" dialog box
- The data set list will reduce to our data set for this tutorial

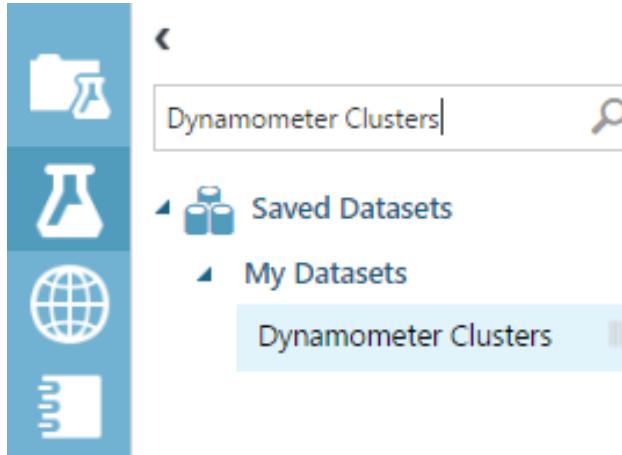
A screenshot of the Azure Machine Learning Studio interface. The search bar at the top now contains "Dynamometer Clusters". The sidebar on the left shows 'Saved Datasets' with 'My Datasets' expanded, and 'Dynamometer Clusters' is listed under it. A red box highlights the search bar, and a red arrow points from the text "Type 'Dynamometer Clusters' in the 'Search experiment items' dialog box" to the search bar.

Step 4.11 : Drag the data set to the experiment

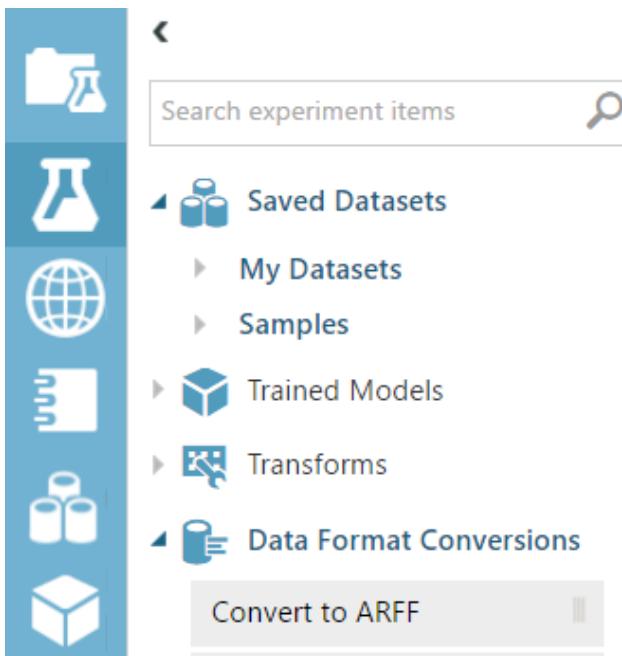


- *Note* when you drag the first element of your model to the canvas ... all the guides disappear
- Now, where are all the tools that were at the left Azure Machine Learning?
- They are still there, ... but we need to un-filter to see them

Step 4.12 : Backspace over “Dynamometer Clusters”



- Backspace to remove “Dynamometer Clusters” from the “Search experiment items” dialog



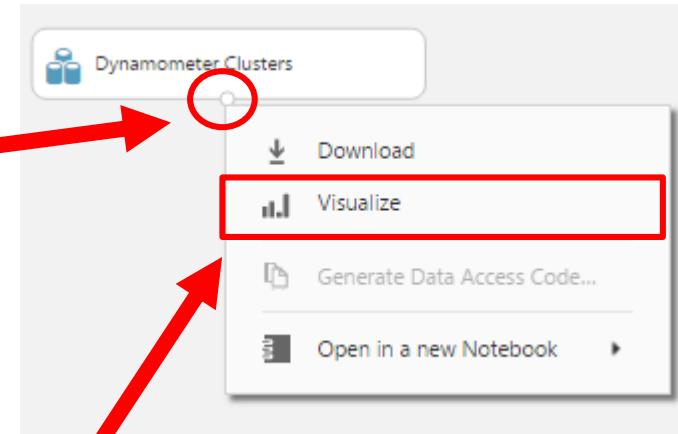
- Then click the triangle at left of “Saved Datasets” to close the dataset list

Step 4.13 : Admire your data set living in your experiment

The screenshot shows the Microsoft Machine Learning studio interface. The title bar reads "Oil & Gas - Dynamometer Classification". On the left, there is a navigation pane with a search bar containing "Dynamometer Clusters" and a magnifying glass icon. Below the search bar, there is a tree view of saved datasets under "Saved Datasets": "My Datasets" has a single item, "Dynamometer Clusters", which is highlighted with a blue selection bar at the bottom of the list. The main workspace is currently empty, showing a placeholder icon and the text "Dynamometer Clusters".

Step 4.14 : Visualize the data set

Right click on the
bottom-middle
circle of the
module



Then click "Visualize"

Step 4.15 : Now look at the data for 4 attributes

#1 Row count

#2 Variable count

#3 Variables

#4 Variable histogram density plots

#5 Variable density plot style

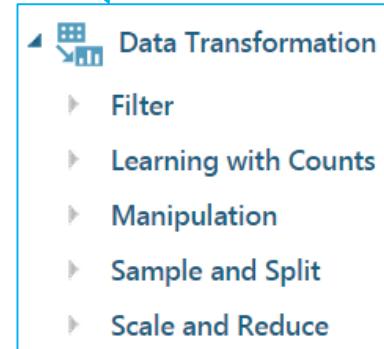
- Histogram
- Boxplot

rows	57883	columns	51			
well date PC1 PC2 PC3 Assignments DistancesToClusterCenter no.0						
view as						
Alabama		2013-08-16T01:55:48.243341	2.918071	2.023029	35	6.71486
Alabama		2013-08-16T23:49:23.8243341	2.918071	2.023029	35	6.71486
Alabama		2013-09-03T17:00:48.495328	2.311737	2.154067	35	6.53259
Alabama		2013-09-06T09:38:47.8148383	2.731808	2.341957	35	6.693467
Alabama		2013-09-06T16:13:14.8148383	2.731808	2.341957	35	6.693467
Alabama		2013-09-13T03:48:51.8206878	2.583768	2.282913	35	6.598962

- ▶  Saved Datasets
- ▶  Data Format Conversions
- ▶  Data Input and Output
- ▶  Data Transformation
- ▶  Feature Selection
- ▶  Machine Learning
- ▶  OpenCV Library Modules
- ▶  Python Language Modules
- ▶  R Language Modules
- ▶  Statistical Functions
- ▶  Text Analytics
- ▶  Deprecated
- ▶  Web Service

Step 5.1 : In addition to data what are the other components we need for an AML model?

- AML modelling ... a checklist approach
 - Create new experiment
 - Import data set
 - Data transformation



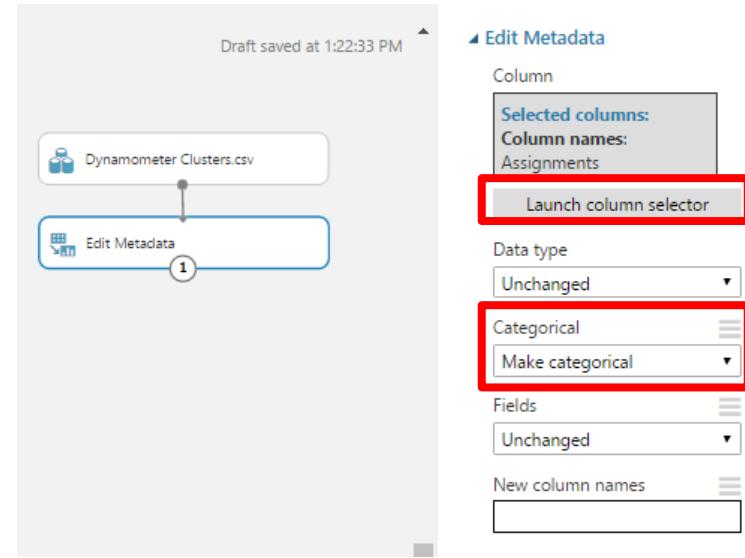
Step 5.2 : Q? Edit Metadata

- ▶ Data Format Conversions
- ▶ Data Input and Output
- ◀ Data Transformation
 - ▶ Filter
 - ▶ Learning with Counts
 - ◀ Manipulation
 - Add Columns
 - Add Rows
 - Apply SQL Transform...
 - Clean Missing Data
 - Convert to Indicator ...
 - Group Categorical Va...
 - Join Data
 - Metadata Editor**
 - Project Columns
 - Remove Duplicate Ro...
 - Select Columns Trans...
 - SMOTE



- Open “Data Transformation” and then “Manipulation” from the navigation pane at the left
- Drag “Metadata Editor” to the canvas
 - “Metadata Editor” edits metadata associated with the columns in a dataset
- Next, click and hold on the bottom middle circle of your “Dataset” module
- While holding down the mouse button, drag the line to the top middle circle of “Metadata Editor” module

Step 5.3 : Q? Edit Metadata



Select columns



Select "With Rules"

- Machine Learning
 - Click on Metadata Editor
 - Choose “Make categorical” from the drop down menu under Categorical
 - Then click on “Launch column selector”

Select “No Columns”

Select “Include” from the drop down menu

Step 5.4 : Q? Edit Metadata

Select columns

BY NAME Allow duplicates and preserve column order in selection

WITH RULES **Begin With**

ALL COLUMNS **NO COLUMNS**

Include column names Assignments

x

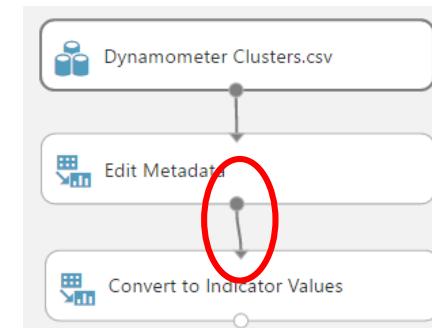
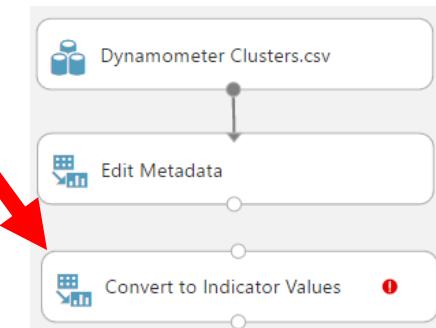
- Click here and a drop down list appears
- Select one variable at a time
 - Assignments

Step 1: Next,
click on the
check circle

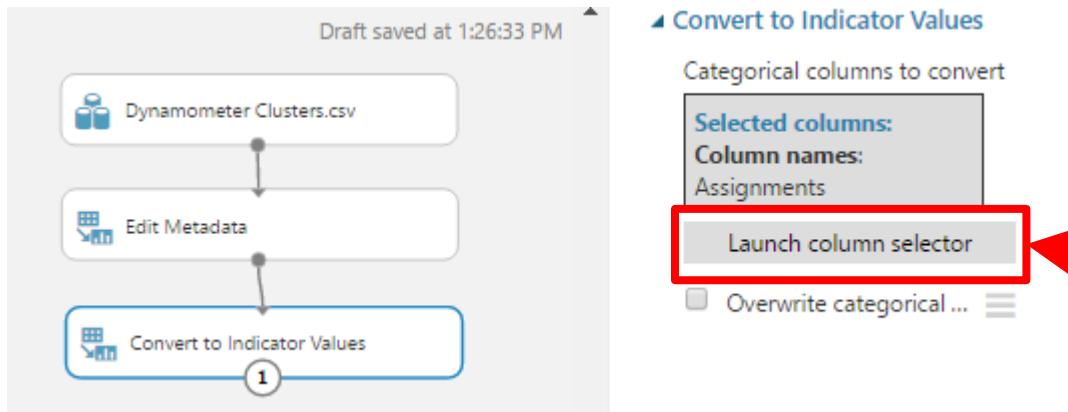
Step 5.5 : Q? How do we exclude variables?

- ▶ Data Format Conversions
- ▶ Data Input and Output
- ◀ Data Transformation
 - ▶ Filter
 - ▶ Learning with Counts
- ◀ Manipulation
 - Add Columns
 - Add Rows
 - Apply SQL Transform...
 - Clean Missing Data
 - Convert to Indicator ...**
 - Edit Metadata
 - Group Categorical Va...
 - Join Data
 - Remove Duplicate R...
 - Select Columns in Da...
 - Select Columns Trans...
 - SMOTE
- ▶ Sample and Split
- ▶ Scale and Reduce

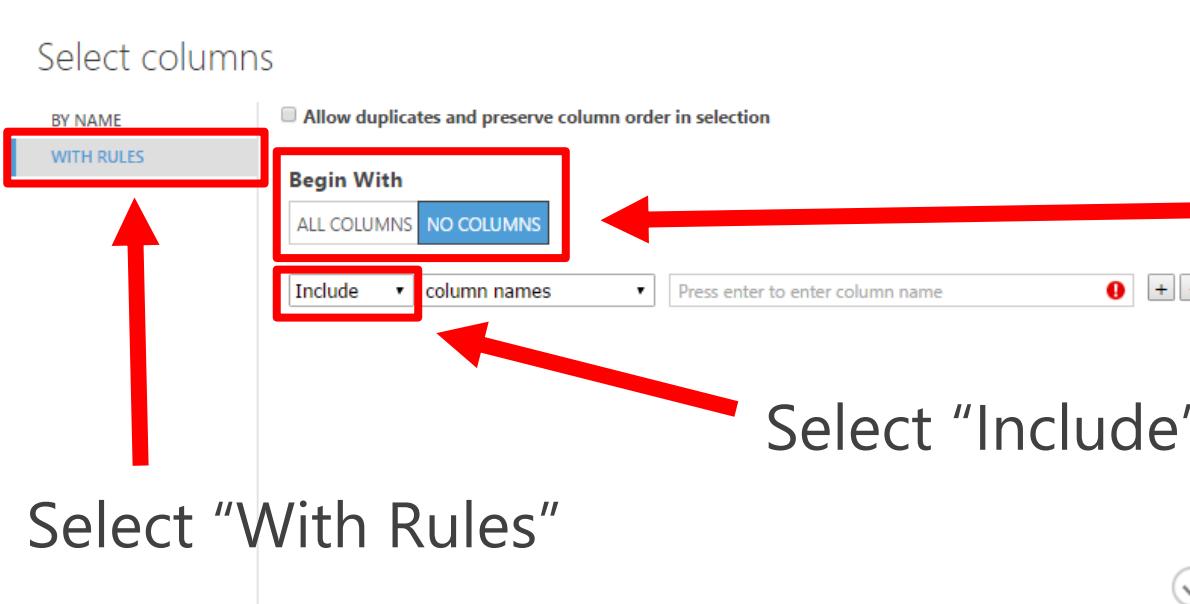
- Open “Data Transformation” and then “Manipulation” from the navigation pane at the left
- Drag “Convert to Indicator Values” to the canvas
 - “Convert to Indicator Values” converts categorical values in columns to indicator values
- Next, click and hold on the bottom middle circle of your “Edit Metadata” module
- While holding down the mouse button, drag the line to the top middle circle of “Convert to Indicator Values” module



Step 5.6 : Q? How do we exclude variables?



- Click on “Select Columns in Dataset” on your canvas, and make sure it has a blue outline
- Then click on “Launch column selector” in the menu at right

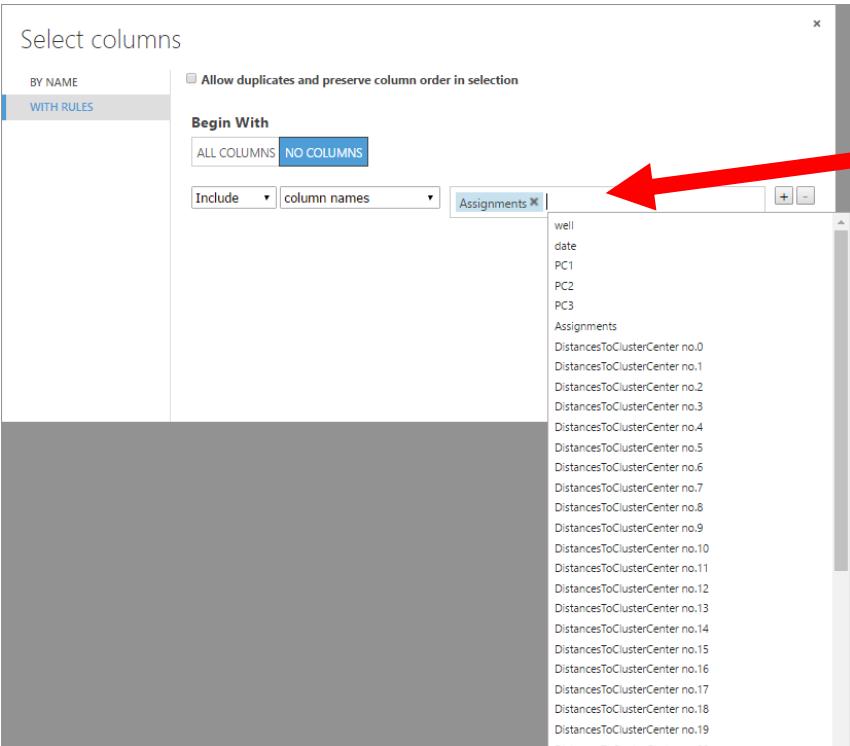


Select “With Rules”

Select “Include” from the drop down menu

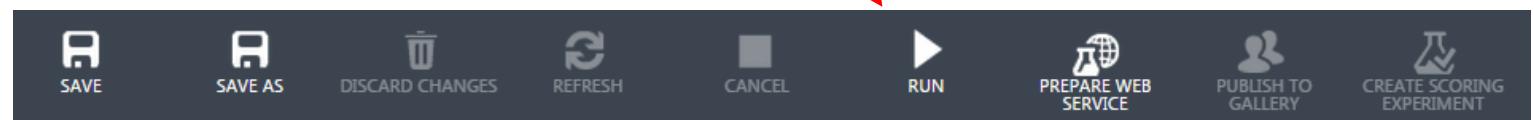
Select “No Columns”

Step 5.7 : Q? How do we exclude variables?



- Click here and a drop down list appears
- Select one variable at a time
 - Assignments

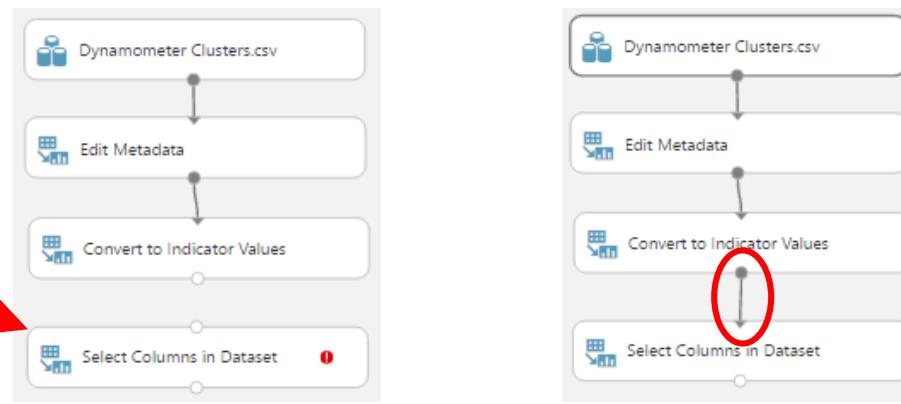
Click "Run"



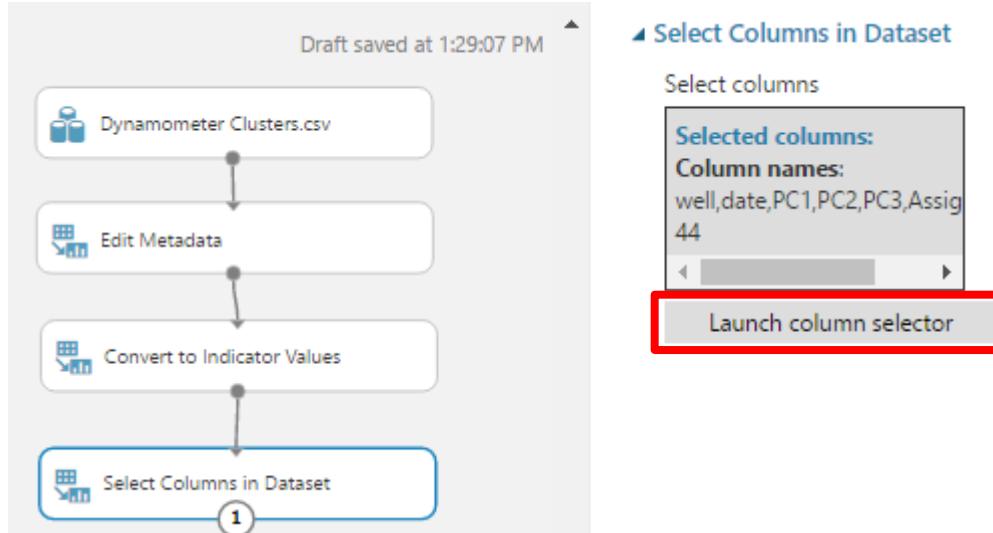
Step 5.8 : Q? How do we exclude variables?

- ▶ Data Format Conversions
- ▶ Data Input and Output
- ◀ Data Transformation
- ▶ Filter
- ▶ Learning with Counts
- ◀ Manipulation
- Add Columns
- Add Rows
- Apply SQL Transform...
- Clean Missing Data
- Convert to Indicator ...
- Edit Metadata
- Group Categorical Va...
- Join Data
- Remove Duplicate R...
- Select Columns in Da...**
- Select Columns Trans...
- SMOTE
- ▶ Sample and Split
- ▶ Scale and Reduce

- Open “Data Transformation” and then “Manipulation” from the navigation pane at the left
- Drag “Select Columns in Dataset” to the canvas
 - “Select Columns in Dataset” selects columns to include or exclude from a dataset in an operation
- Next, click and hold on the bottom middle circle of your “Convert to Indicator Values” module
- While holding down the mouse button, drag the line to the top middle circle of “Select Columns in Dataset” module



Step 5.9 : Q? How do we exclude variables?



- Click on "Select Columns in Dataset" on your canvas, and make sure it has a blue outline
- Then click on "Launch column selector" in the menu at right

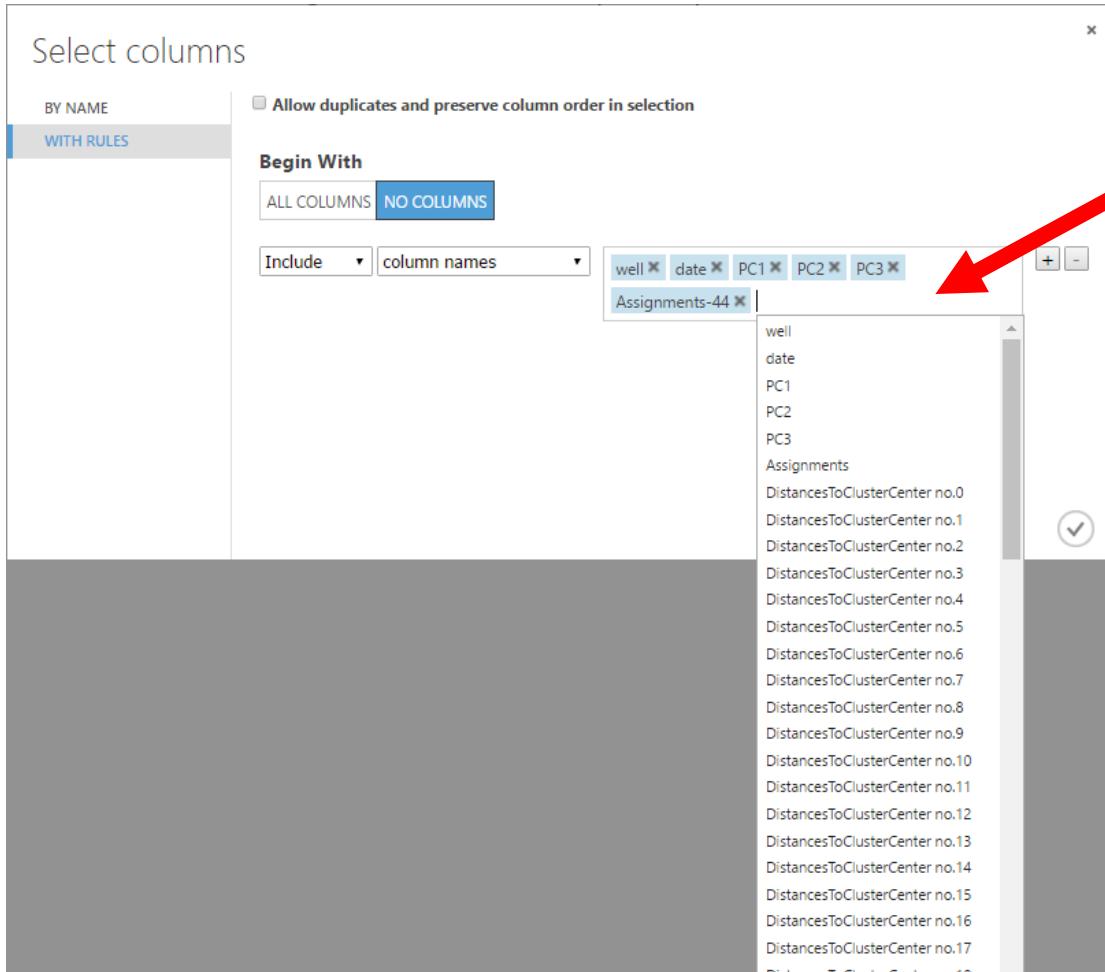


Select "With Rules"

Select "No Columns"

Select "Include" from the drop down menu

Step 5.10 : Q? How do we exclude variables?



- Click here and a drop down list appears
- Select one variable at a time
 - Well
 - Date
 - PC1
 - PC2
 - PC3
 - Assignments-44

Step 5.11 : Q? How do we exclude variables?

Select columns

BY NAME WITH RULES

Allow duplicates and preserve column order in selection

Begin With

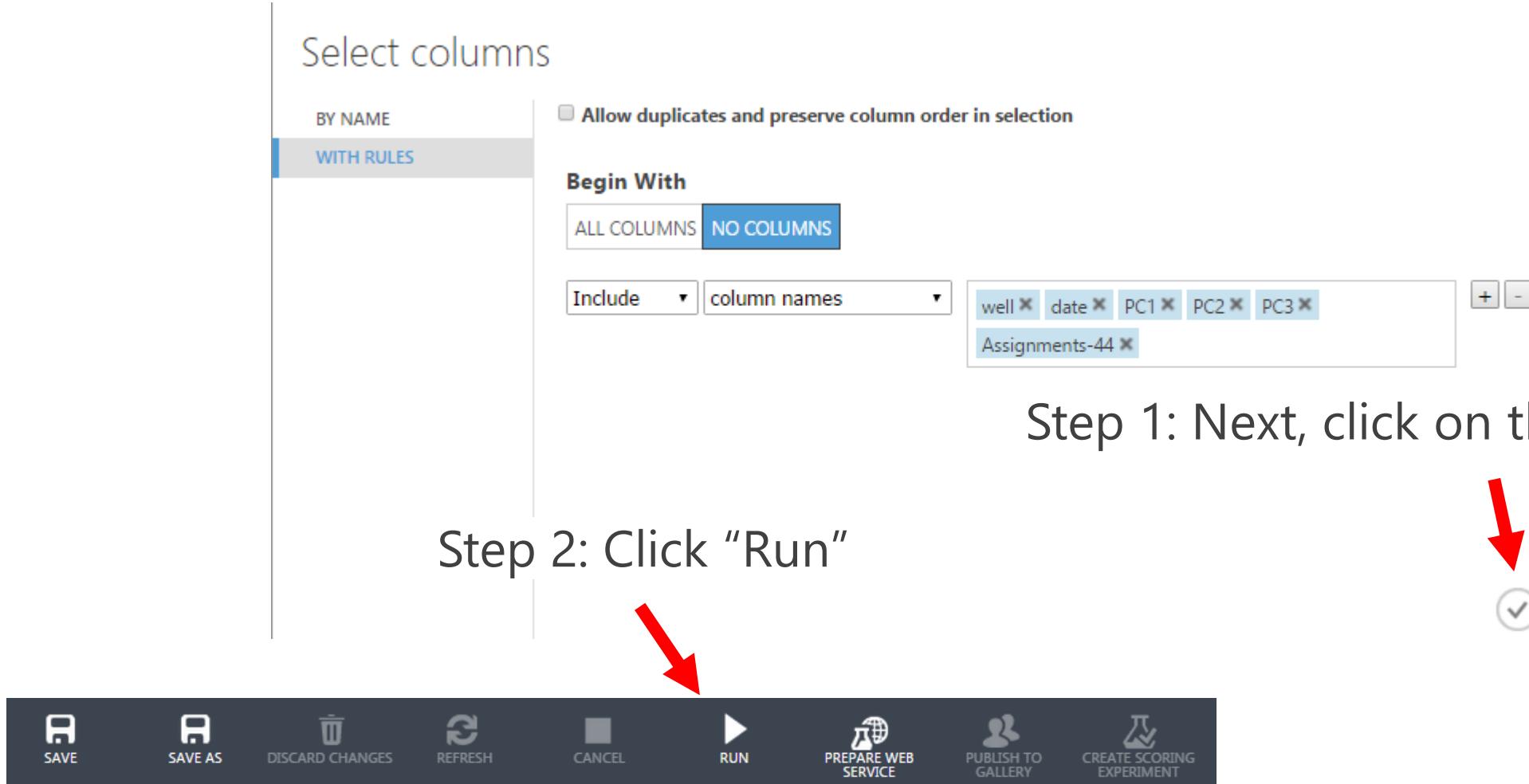
ALL COLUMNS NO COLUMNS

Include column names well date PC1 PC2 PC3 Assignments-44

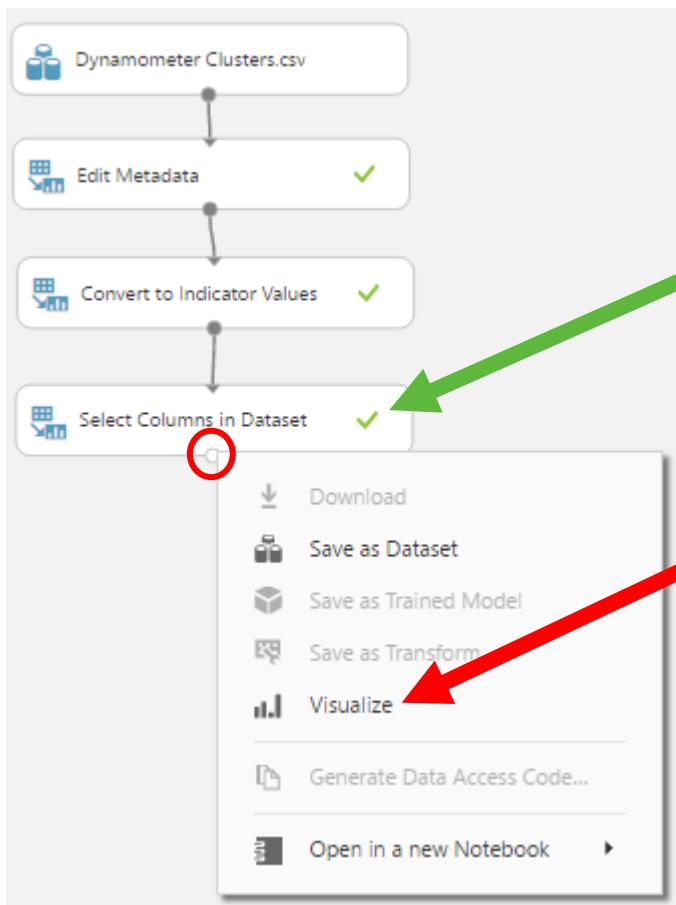
Step 1: Next, click on the check circle

Step 2: Click "Run"

SAVE SAVE AS DISCARD CHANGES REFRESH CANCEL RUN PREPARE WEB SERVICE PUBLISH TO GALLERY CREATE SCORING EXPERIMENT



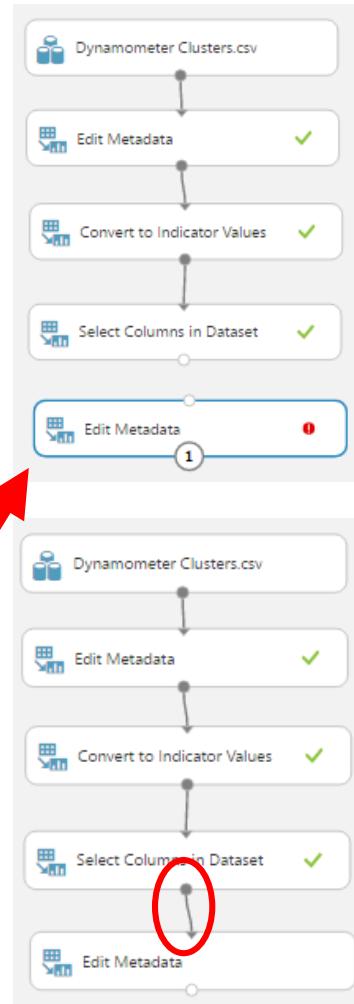
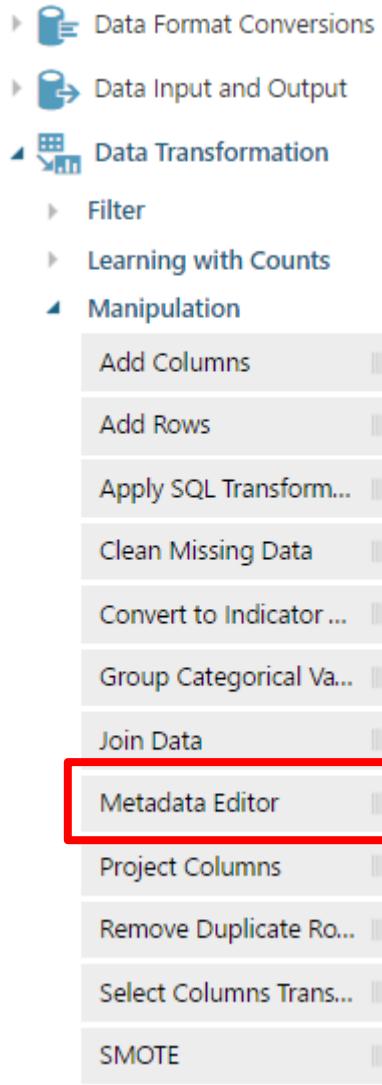
Step 5.12 : Q? How do we exclude variables?



- After Azure Machine Learning “runs” all the steps in your model, you will see a green check mark in the “Select Columns in Dataset” box
 - Next right click on the bottom-middle circle in “Select Columns in Dataset” and select “Visualize”

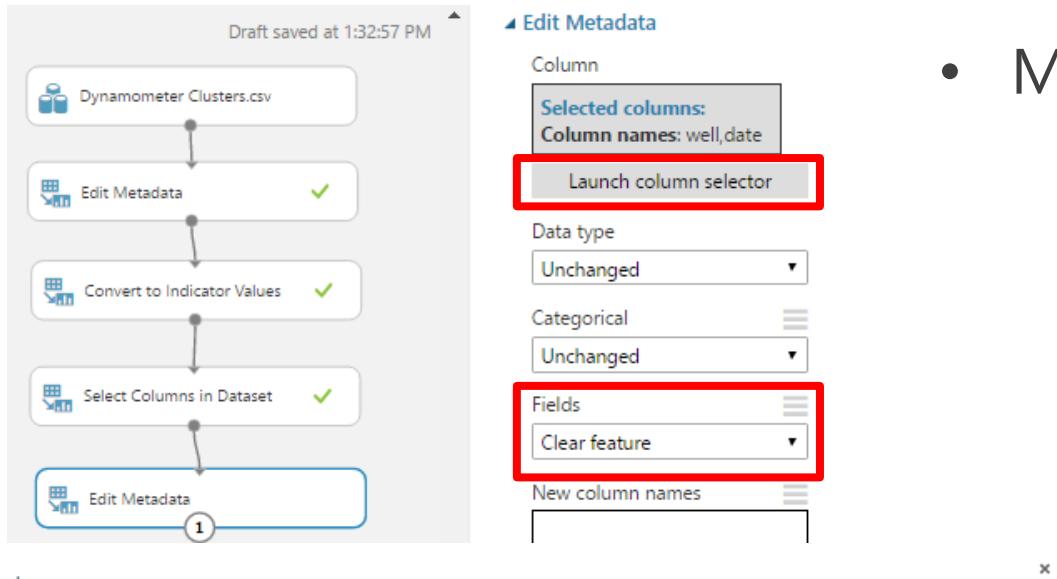
rows	columns					
57883	6					
view as	well	date	PC1	PC2	PC3	Assignments-44
	Alabama	2013-08-16T01:55:22	-8.243341	2.918071	2.023029	0
	Alabama	2013-08-16T23:49:23	-8.243341	2.918071	2.023029	0
	Alabama	2013-09-03T00:17:00	-8.495328	2.311737	2.154067	0
	Alabama	2013-09-06T09:38:47	-8.148383	2.731808	2.341957	0
	Alabama	2013-09-06T16:13:14	-8.148383	2.731808	2.341957	0
	Alabama	2013-09-13T03:48:51	-8.206878	2.583768	2.282913	0
	Alabama	2013-09-13T20:28:44	-8.206878	2.583768	2.282913	0
	Alabama	2013-11-05T13:17:14	-8.640142	2.146508	2.971267	0

Step 5.13 : Q? Edit Metadata



- Open “Data Transformation” and then “Manipulation” from the navigation pane at the left
- Drag “Metadata Editor” to the canvas
 - “Metadata Editor” edits metadata associated with the columns in a dataset
- Next, click and hold on the bottom middle circle of your “Select Columns in Dataset” module
- While holding down the mouse button, drag the line to the top middle circle of “Metadata Editor” module

Step 5.14 : Q? Edit Metadata



Select columns



Select "With Rules"

Select "Include" from the drop down menu

Step 5.15 : Q? Edit Metadata

Select columns

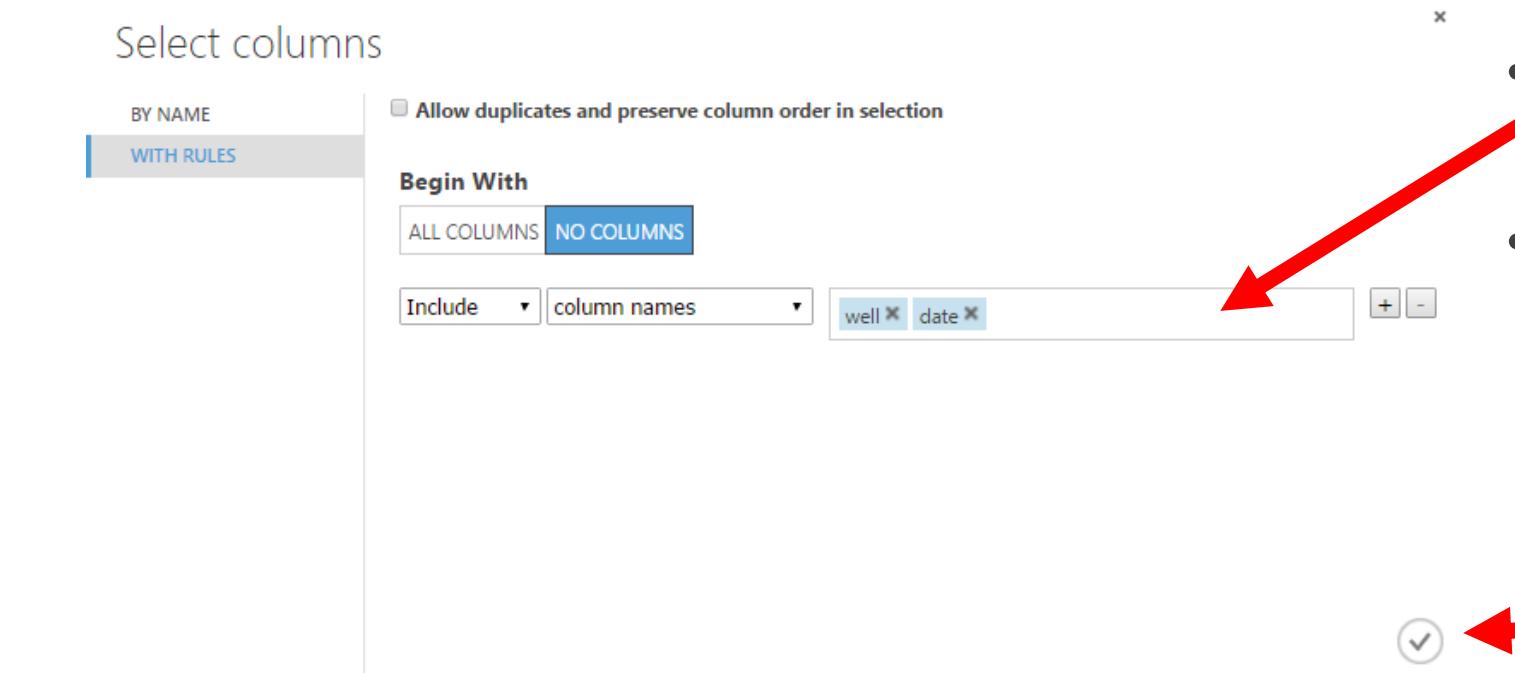
BY NAME WITH RULES Allow duplicates and preserve column order in selection

Begin With

ALL COLUMNS NO COLUMNS

Include column names well date

Step 2: Click "Run"



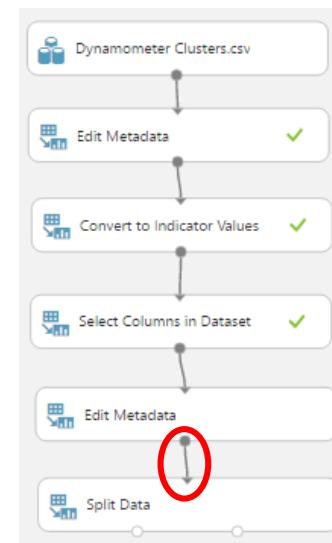
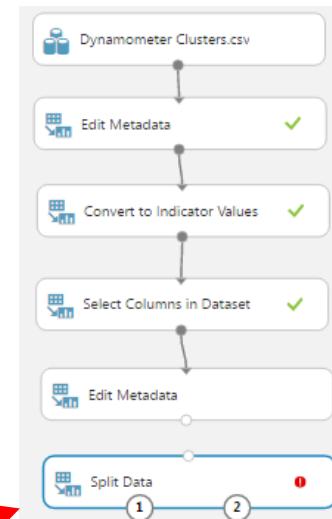
SAVE SAVE AS DISCARD CHANGES REFRESH CANCEL RUN PREPARE WEB SERVICE PUBLISH TO GALLERY CREATE SCORING EXPERIMENT

- Click here and a drop down list appears
- Select one variable at a time
 - well
 - date

Step 1: Next, click on the check circle

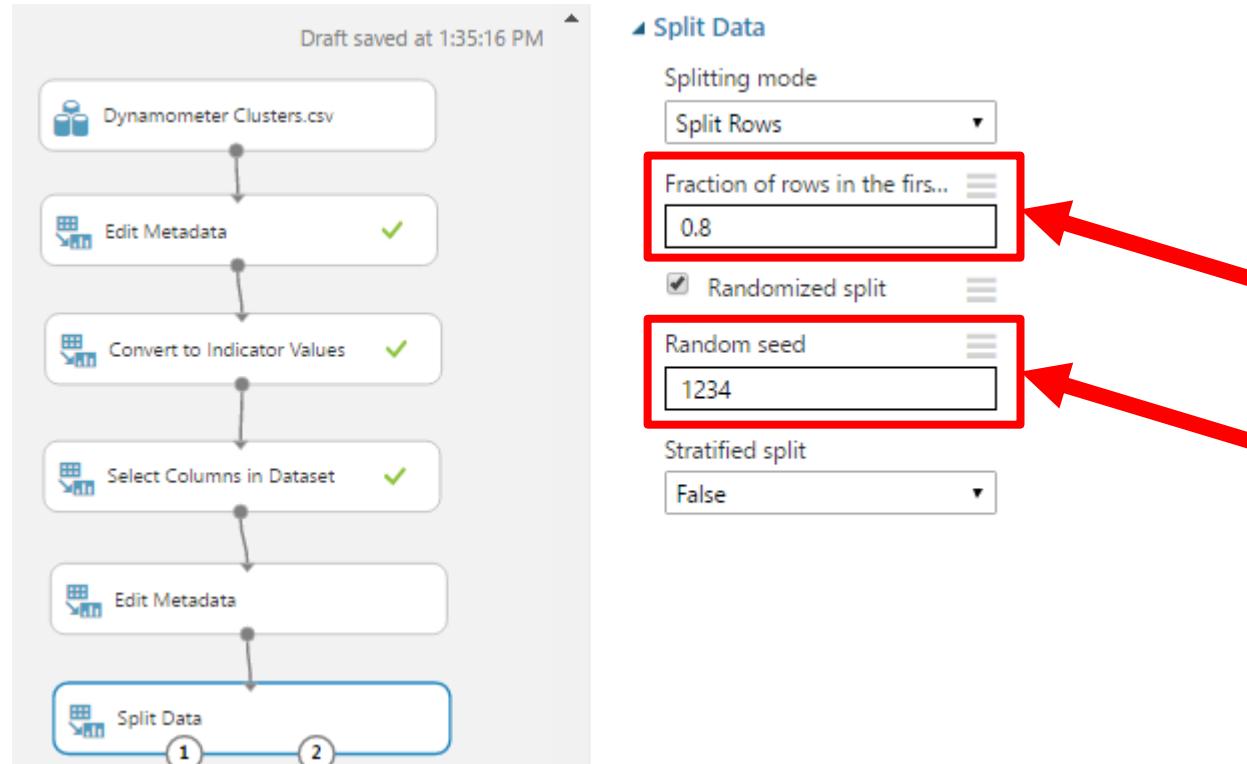
Step 5.16 : Q? How to split the dataset for training and scoring?

- ▲ **Data Transformation**
 - ▶ Filter
 - ▶ Learning with Counts
 - ▶ Manipulation
 - ◀ **Sample and Split**
 - Partition and Sample
 - Split Data**



- Open “Data Transformation” and then “Sample and Split” from the navigation pane at the left
- Drag “Split Data” to the canvas
 - “Split Data” splits the rows of a dataset into two distinct sets
- Next, click and hold on the bottom middle circle of your “Metadata Editor” module
- While holding down the mouse button, drag the line to the top middle circle of “Split Data” module

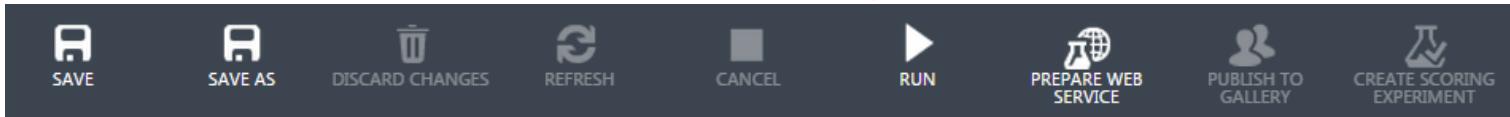
Step 5.17 : Q? How to split the dataset for training and scoring?



- Click on “Split Data” on your canvas, and make sure it has a blue outline
- Change the “Fraction of rows in the first output dataset” to 0.8
- Set the Random seed to 1234

Step 5.18 : Run the Experiment

Finally, click
"Run"



- All the data transformation needed for this tutorial is complete.
 - It was all
 - Drag and Drop
 - Every step preserved on the canvas
 - Data and programs are just simple shapes
- We are ready for the next component of our model

- ▶  Saved Datasets
- ▶  Data Format Conversions
- ▶  Data Input and Output
- ▶  Data Transformation
- ▶  Feature Selection
- ▶  Machine Learning
- ▶  OpenCV Library Modules
- ▶  Python Language Modules
- ▶  R Language Modules
- ▶  Statistical Functions
- ▶  Text Analytics
- ▶  Deprecated
- ▶  Web Service

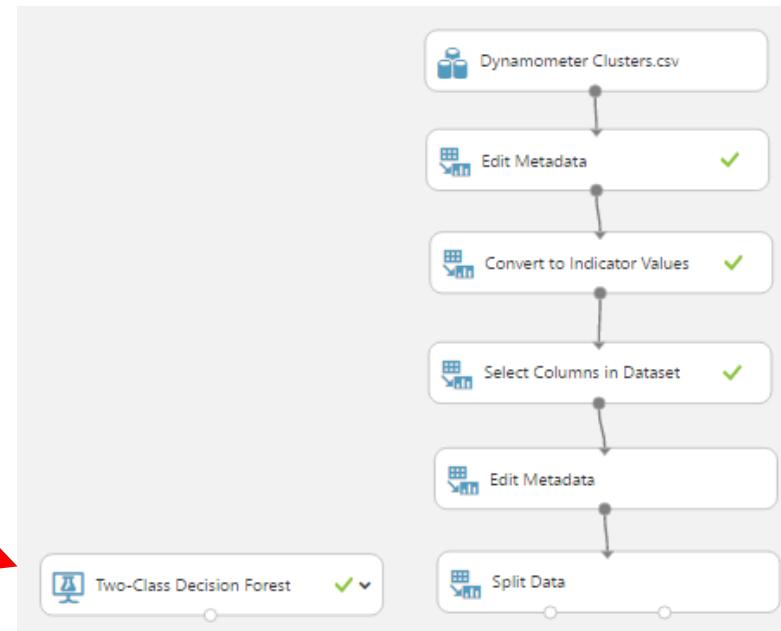
Step 6.1 : AML Modeling Components

- AML modelling ... a checklist approach
 - Initial data set
 - Create new experiment
 - Import data set
 - Data transformation
 - Machine Learning

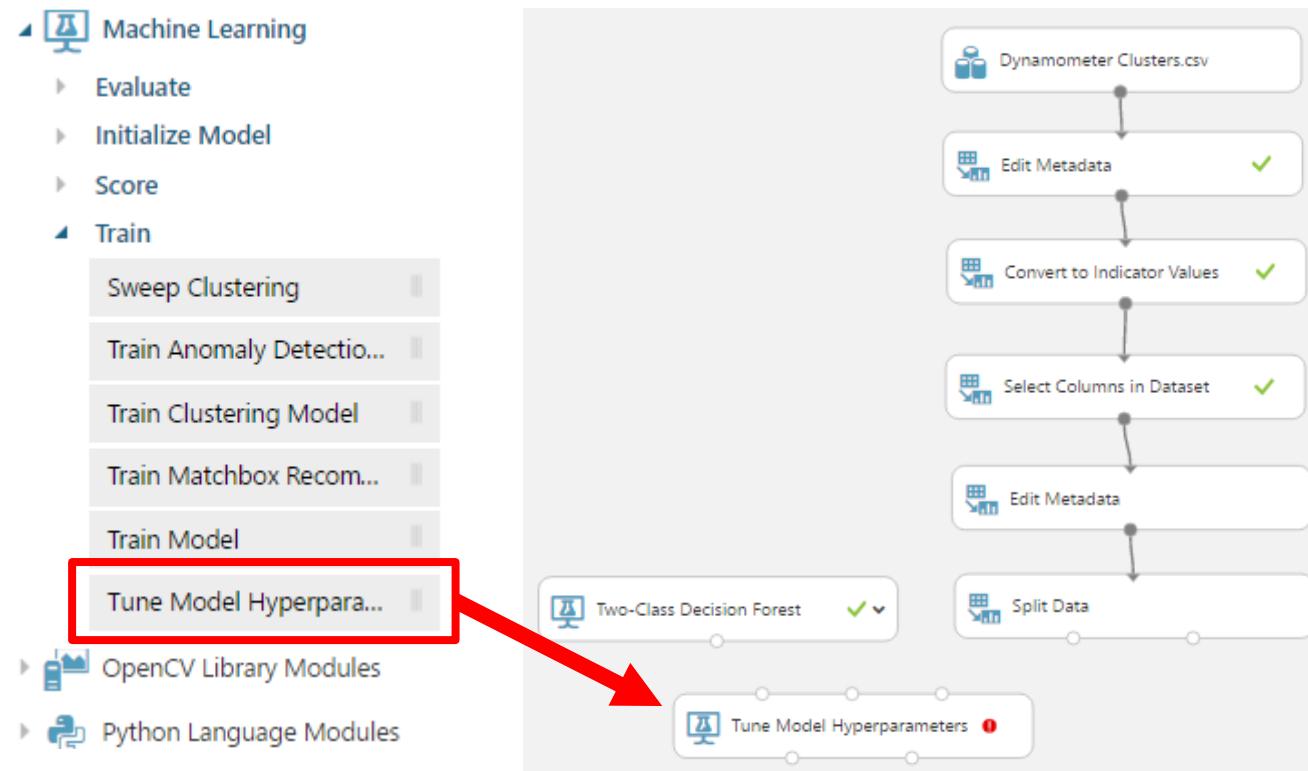
Step 6.2 : Q? How to develop the Machine Learning Algorithm?

- ▲ Machine Learning
 - ▶ Evaluate
 - ◀ Initialize Model
 - ◀ Anomaly Detection
 - One-Class Support Vector Machine
 - PCA-Based Anomaly Detection
- ◀ Classification
 - Multiclass Decision Forest
 - Multiclass Decision Jungle
 - Multiclass Logistic Regression
 - Multiclass Neural Network
 - One-vs-All Multiclass
 - Two-Class Averaged Perceptron
 - Two-Class Bayes Point Machine
 - Two-Class Boosted Decision Tree
 - Two-Class Decision Forest**
 - Two-Class Decision Jungle
 - Two-Class Locally-Deep Support V...
 - Two-Class Logistic Regression
 - Two-Class Neural Network

- Open “Machine Learning” -> “Initialize Model” -> “Classification” from the navigation pane at the left
- Drag “Two-Class Decision Forest” to the canvas

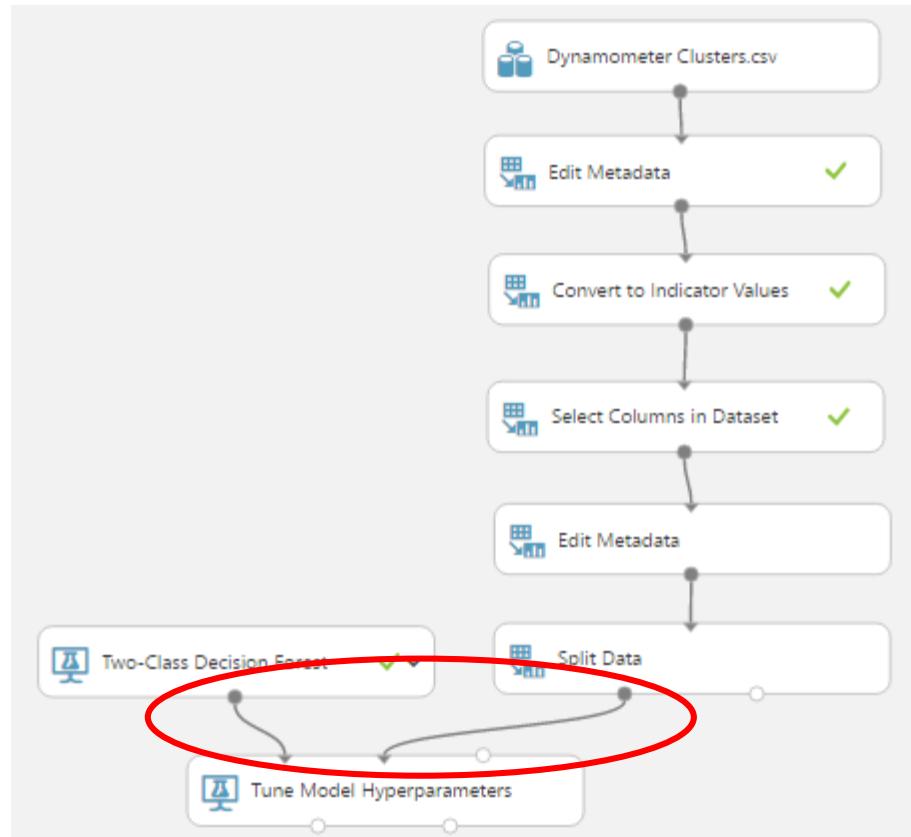


Step 6.3 : Q? How to develop the Machine Learning Algorithm?



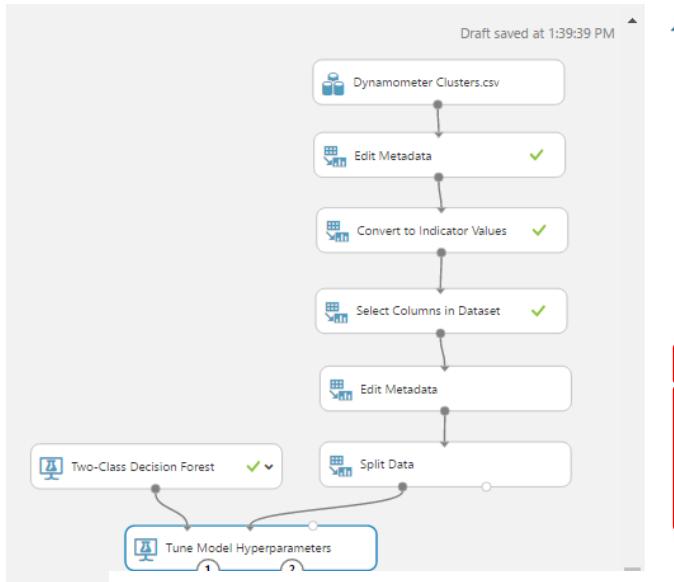
- Open “Machine Learning” and then “Train” from the navigation pane at the left
- Drag “Tune Model Hyperparameters” module to the canvas
 - “Tune Model Hyperparameters” perform a parameter sweep on the model to determine the optimum parameter settings

Step 6.4 : Q? How to develop the Machine Learning Algorithm?

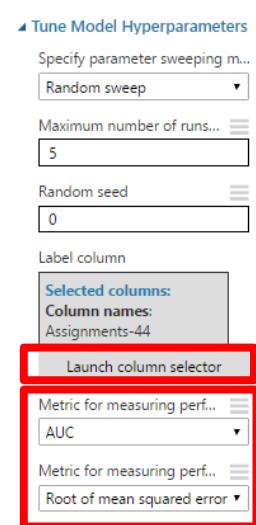


- Next, click and hold on the bottom middle circle of your “Two-Class Decision Forest” module and drag the line to the top left circle of the “Tune Model Hyperparameters” module.
- Similarly, click and hold on the bottom left circle of your “Split Data” module and drag the line to the top right circle of the “Tune Model Hyperparameters” module.

Step 6.5 : Q? How to develop the Machine Learning Algorithm



Select a single column



- Machine Learning
 - Click on “Tune Model Hyperparameters”
 - Set Metric for measuring performance for classification to “AUC”
 - Set Metric for measuring performance for regression to “Root of mean squared error”
 - Then click on “Launch column selector”



Select “With Rules”

Select “Include” from the drop down menu

Step 6.6 : Q? How to develop the Machine Learning Algorithm?

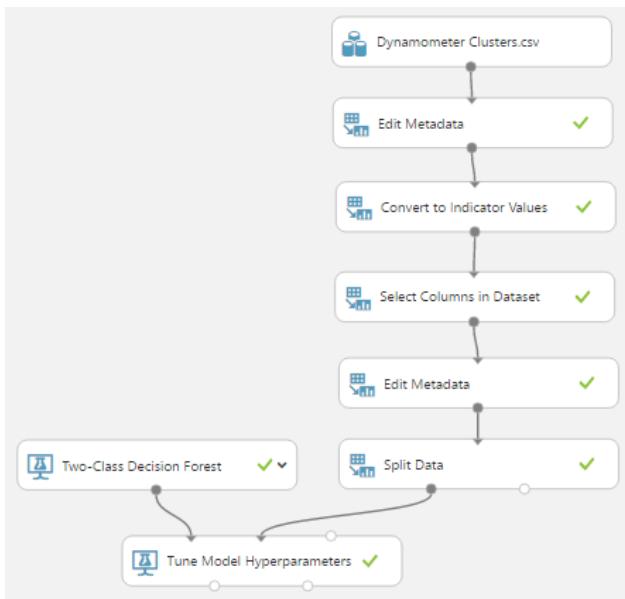
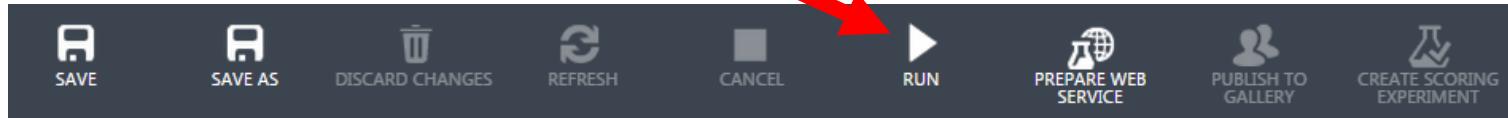
Select a single column

The screenshot shows a software interface for selecting a single column. On the left, there's a sidebar with 'BY NAME' and 'WITH RULES' buttons. In the center, there are two dropdown menus: 'Include' and 'column names'. Below them is a list titled 'Assignments-44' which contains the following items: 'well', 'date', 'PC1', 'PC2', 'PC3', and 'Assignments-44'. A red arrow points from the top right towards the 'Assignments-44' item in the list. Another red arrow points from the bottom right towards a circular checkmark icon at the bottom of the screen.

- Click here and a drop down list appears
- Select one variable at a time
 - Assignments-44
- Next, click on the check circle

Step 6.7 : Run the Experiment

Click "Run"



- After Azure Machine Learning “runs”, you will see a green check mark in all the modules
- All it took to train a machine learning model was drag and drop

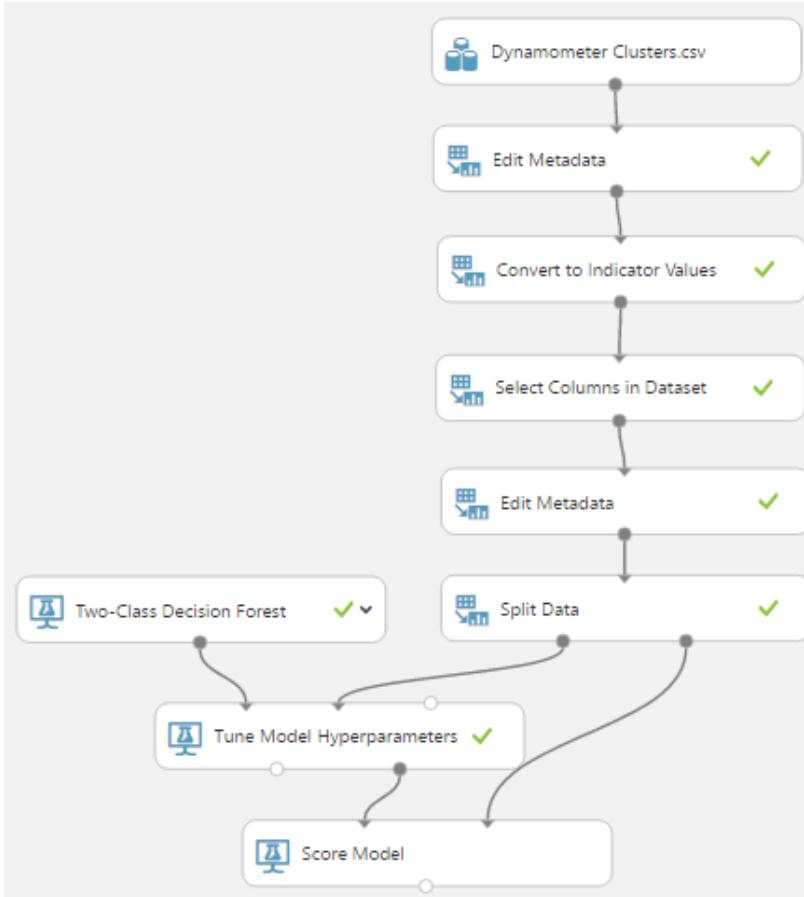
Step 6.8 : Add Score Model

- ▶  Data Format Conversions
- ▶  Data Input and Output
- ▶  Data Transformation
- ▶  Feature Selection
- ▶  Machine Learning
 - ▶ Evaluate
 - ▶ Initialize Model
 - ▶ **Score**
 - Apply Transformation**
 - Assign Data to Cluste...**
 - Score Matchbox Reco..**
 - Score Model**
- ▶ Train
- ▶  OpenCV Library Modules



- Open “Machine Learning” and then “Score” from the navigation pane at the left
- Drag two “Score Model” to the canvas
 - “Score Model” scores a trained classification or a regression model

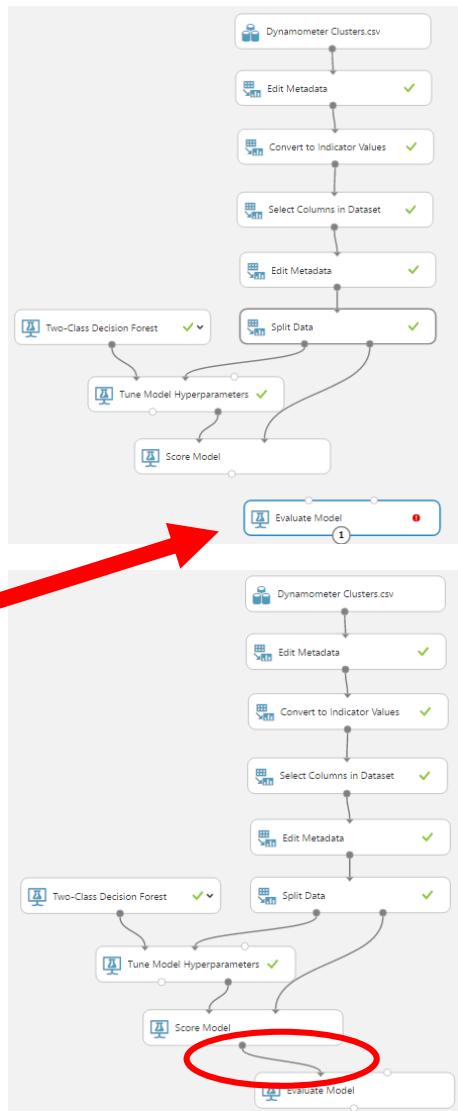
Step 6.9 : Q? How to develop the Machine Learning Algorithm?



- Next, click and hold on the bottom right circle of your “Tune Model Hyperparameters” module and drag the line to the top left circle of the “Score Model” module.
- Similarly, click and hold on the bottom right circle of your “Split Data” module and drag the line to the top right circle of the “Score Model” module.

Step 6.10 : Add Evaluate Model

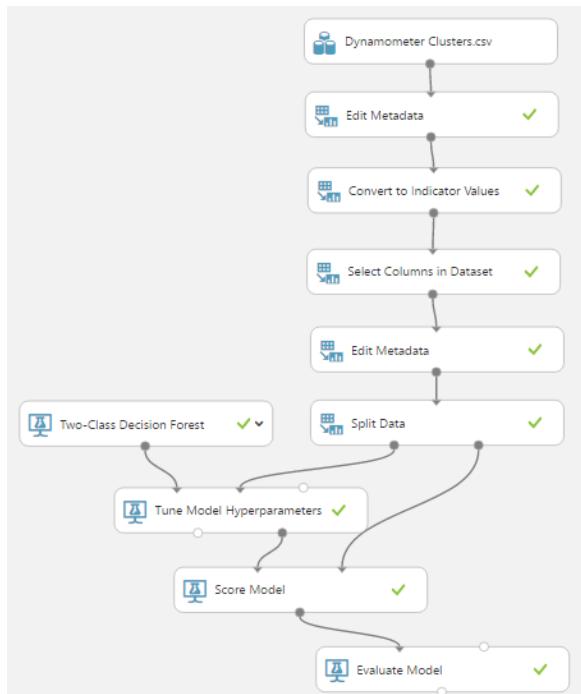
- ▶ Saved Datasets
- ▶ Trained Models
- ▶ Transforms
- ▶ Data Format Conversions
- ▶ Data Input and Output
- ▶ Data Transformation
- ▶ Feature Selection
- ▶ Machine Learning
 - ◀ Evaluate
 - Cross Validate Model
 - Evaluate Model**
 - Evaluate Recommend...
 - ▶ Initialize Model
 - ▶ Score
 - ▶ Train
- ▶ OpenCV Library Modules



- Open “Machine Learning” and then “Evaluate” from the navigation pane at the left
- Drag “Evaluate Model” to the canvas
 - “Evaluate Model” evaluates the scored classification or regression model with standard metrics
- Next, click and hold on the bottom middle circle of your “Score Model” module
- While holding down the mouse button, drag the line to the top left circle of “Evaluate Model” module

Step 6.11 : Run the Experiment

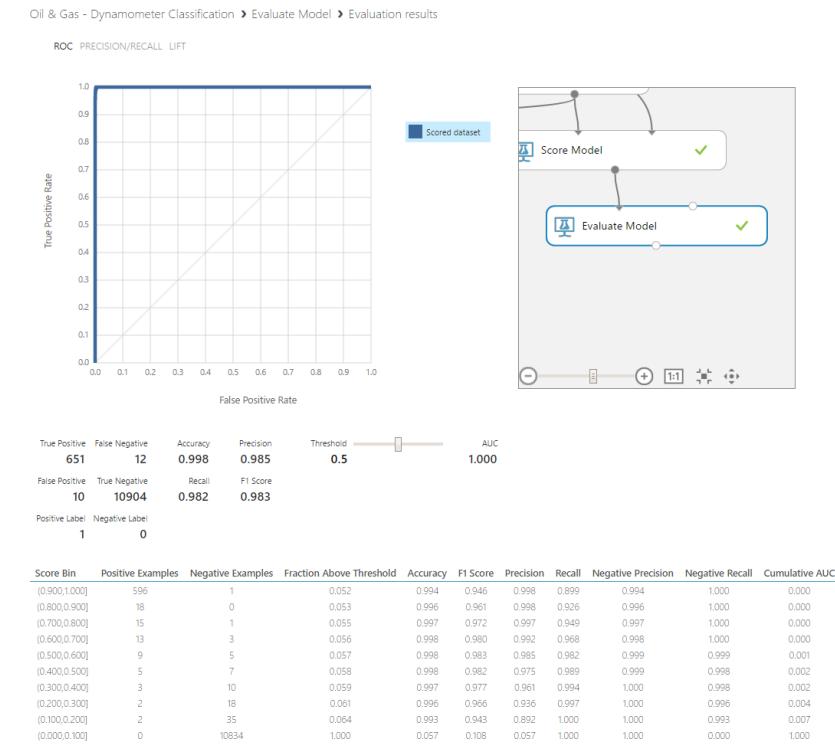
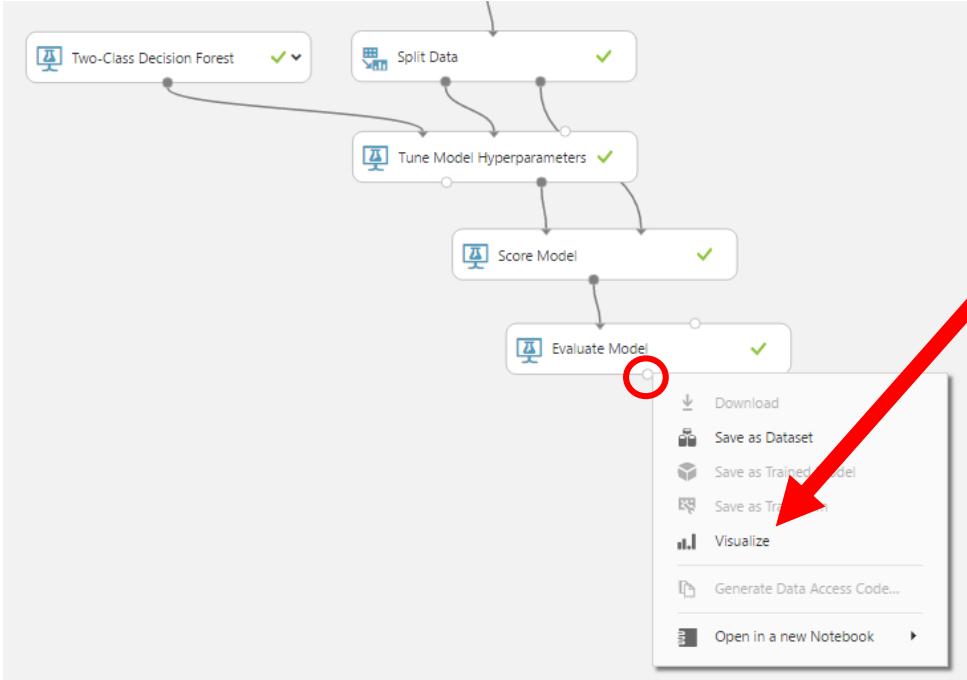
Click "Run"



- After Azure Machine Learning “runs”, you will see a green check mark in all the modules
- All it took to train a machine learning model was drag and drop

Step 6.12 : Visualize the Results

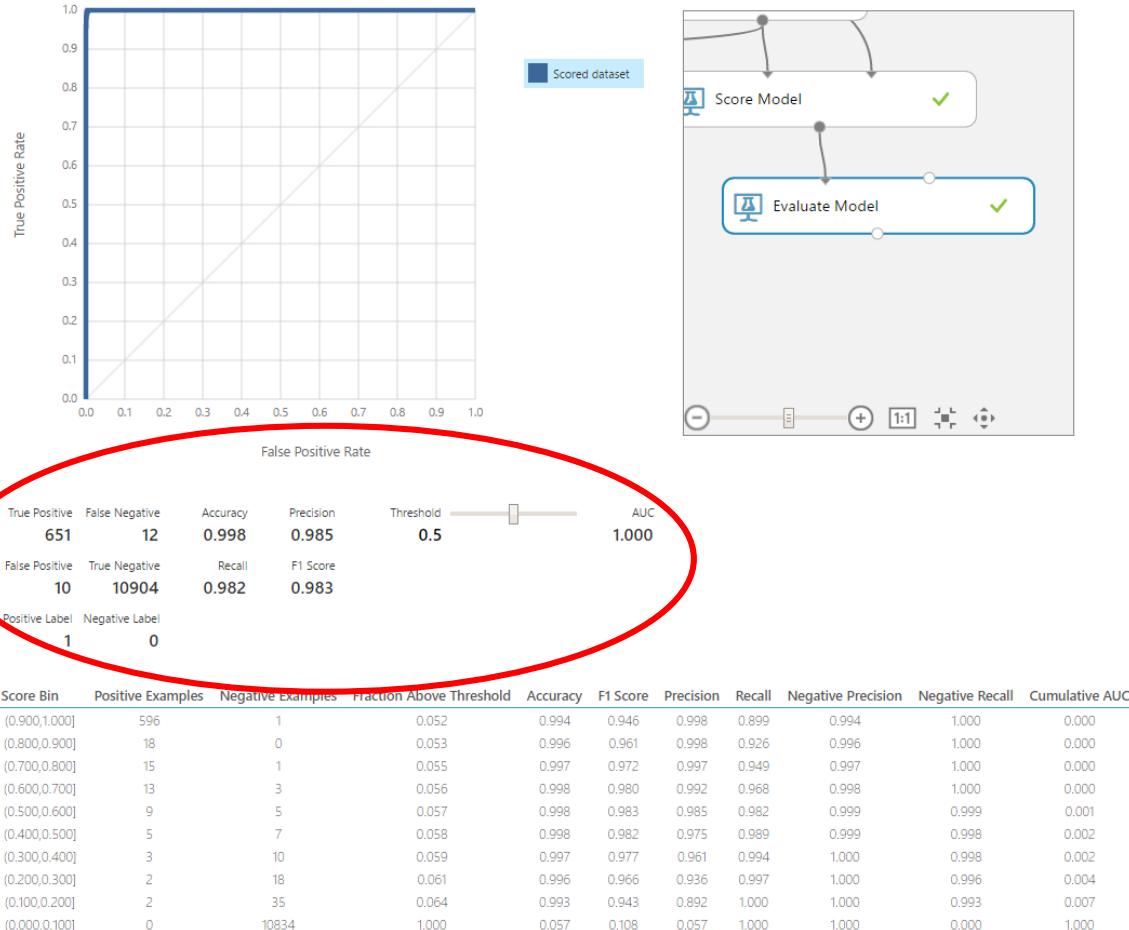
- Right click on the bottom-middle circle of "Evaluate Model" and select "Visualize"



Step 6.13 : Interpret the Results

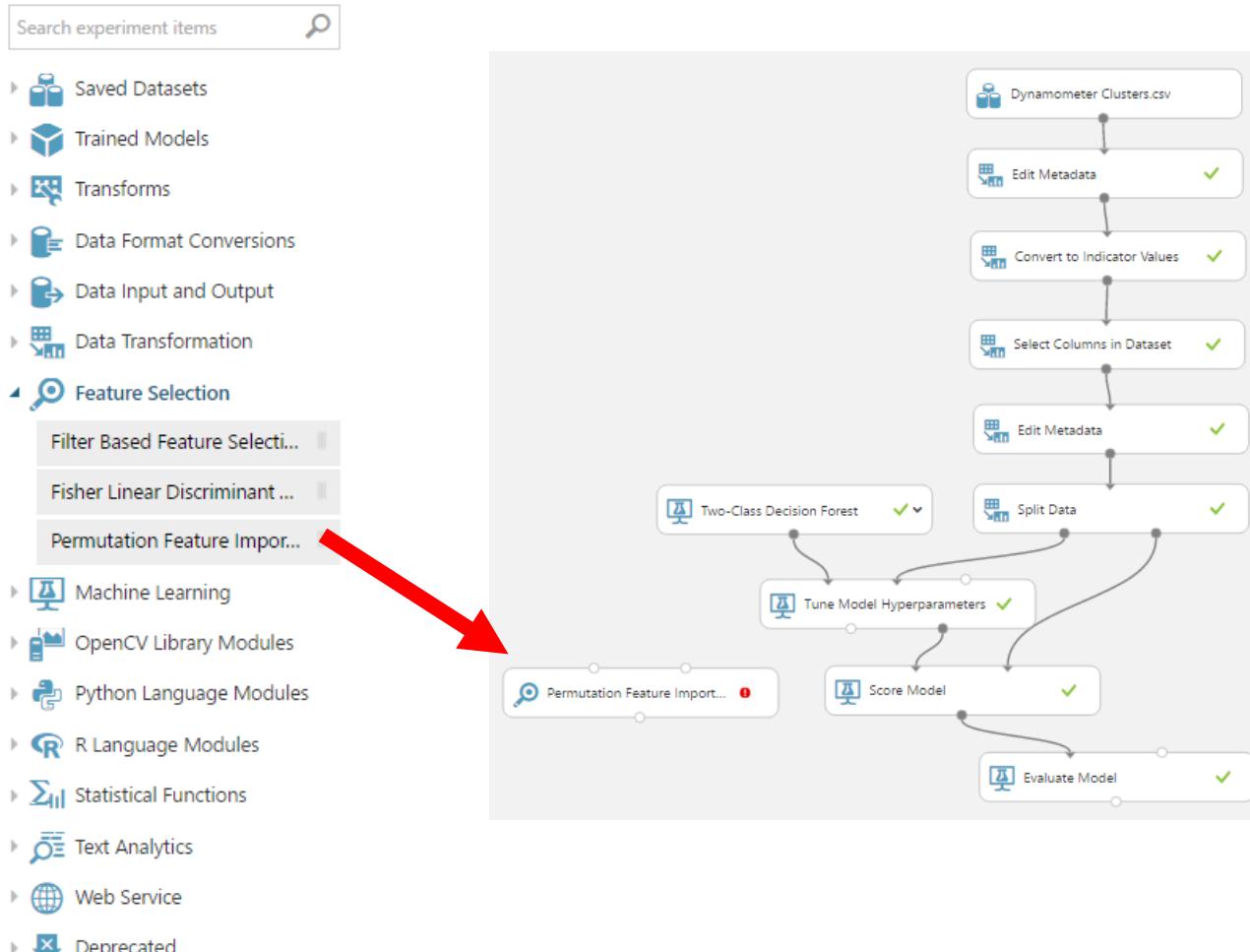
Oil & Gas - Dynamometer Classification > Evaluate Model > Evaluation results

ROC PRECISION/RECALL LIFT



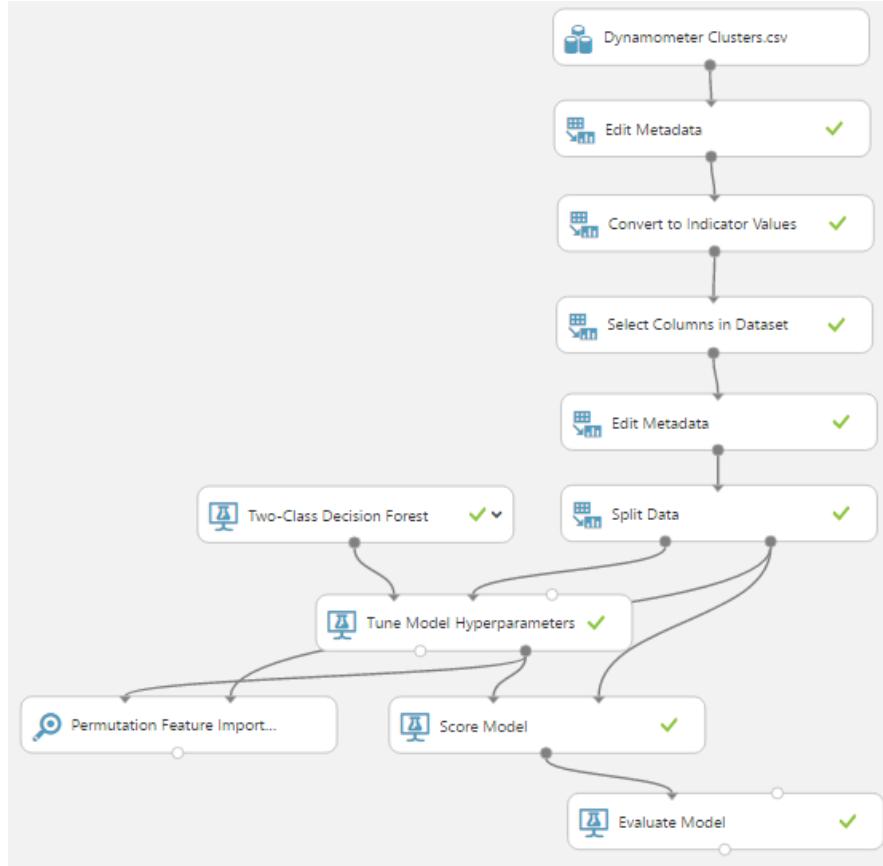
- ROC - Accuracy is measured by the area under the ROC curve. An area of 1 represents a perfect test; an area of .5 represents a worthless test
 - .90-1 = excellent
 - .80-.90 = good
 - .70-.80 = fair
 - .60-.70 = poor
 - .50-.60 = fail
- Model performance metrics: "Accuracy", "Precision", "Recall", and "F-Score".

Step 6.14 : Add Permutation Feature Importance Module



- Open “Feature Selection” from the navigation pane at the left
- Drag “Permutation Feature Importance” to the canvas
 - “Permutation Feature Importance” computes the permutation feature importance of feature variables given a trained model and a test dataset

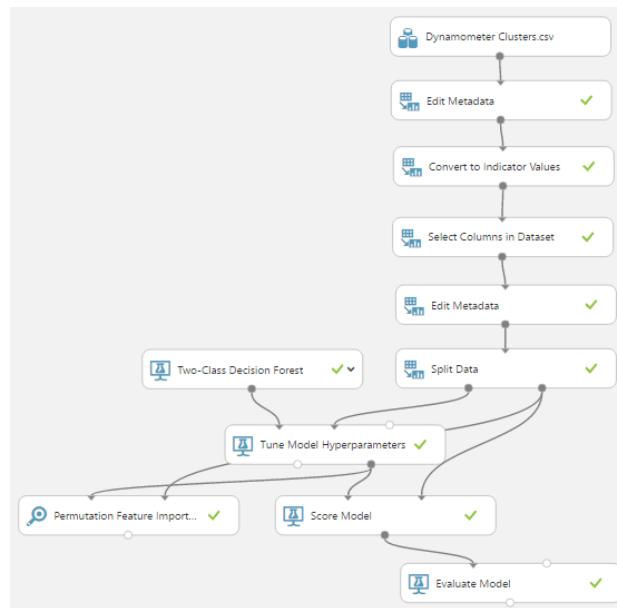
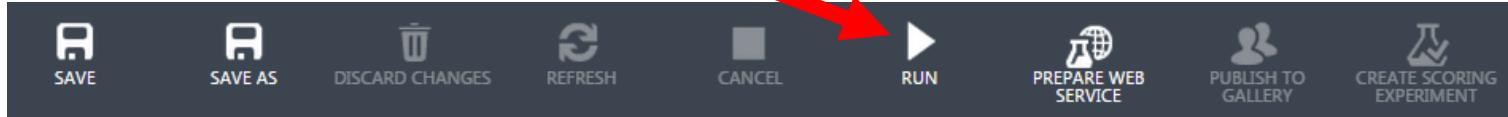
Step 6.15 : Add Permutation Feature Importance Module



- Next, click and hold on the bottom right circle of your “Tune Model Hyperparameters” module
- While holding down the mouse button, drag the line to the top left circle of “Permutation Feature Importance” module
- Similarly, click and hold on the bottom right circle of your “Split Data” module and drag the line to the top right circle of the “Permutation Feature Importance” module.

Step 6.16 : Run the Experiment

Click "Run"

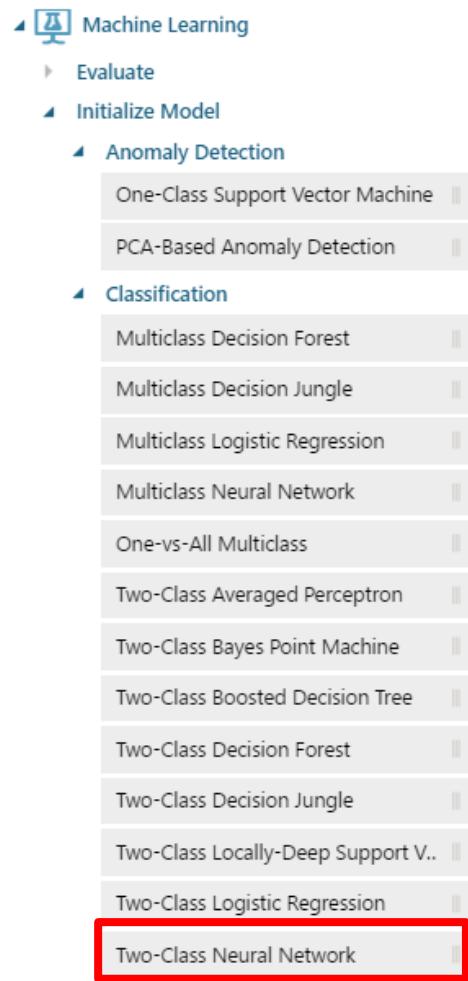


- After Azure Machine Learning “runs”, you will see a green check mark in all the modules
- All it took to train a machine learning model was drag and drop

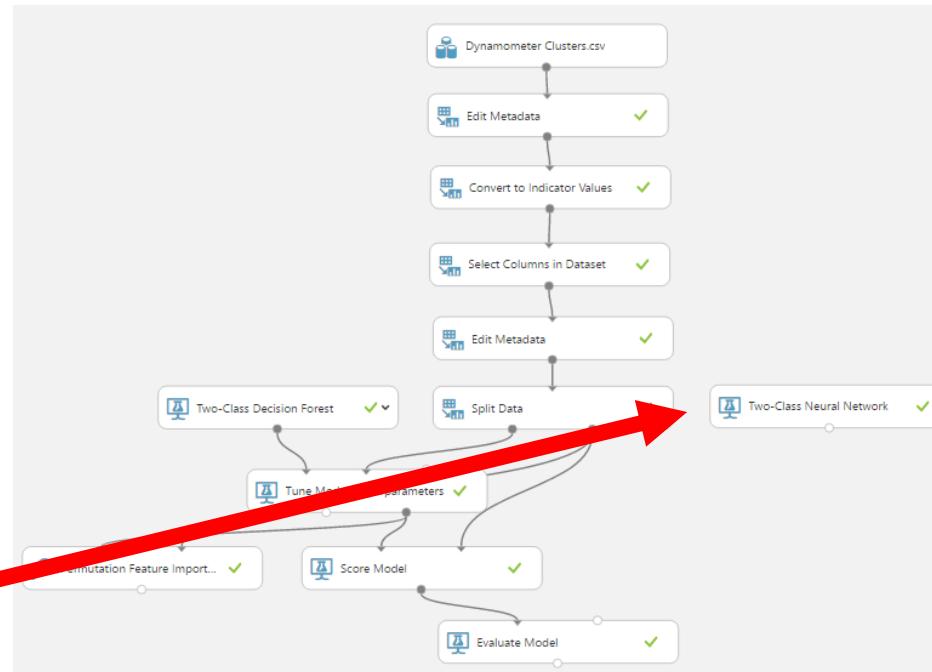


If you have time left then let's try a comparison between two classification algorithms

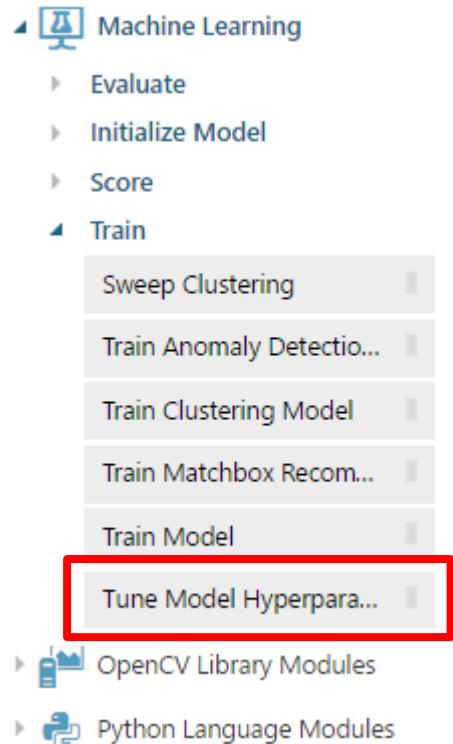
Step 7.1 : Adding the Two-Class Neural Network



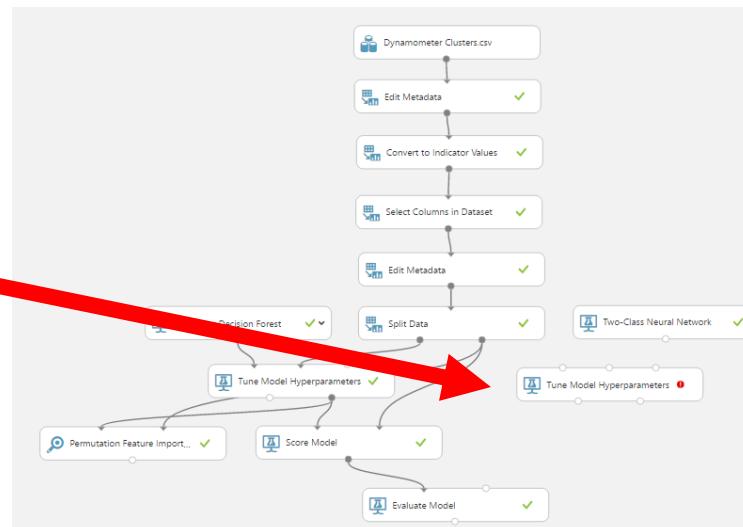
- Open “Machine Learning” -> “Initialize Model” -> “Classification” from the navigation pane at the left
- Drag “Two-Class Neural Network” to the canvas



Step 7.2 : Q? Train the Model



- Open "Machine Learning" and then "Train" from the navigation pane at the left
- Drag "Tune Model Hyperparameters" module to the canvas
 - "Tune Model Hyperparameters" perform a parameter sweep on the model to determine the optimum parameter settings

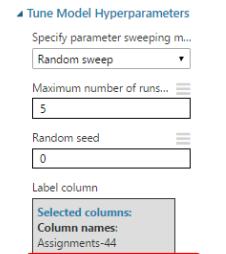
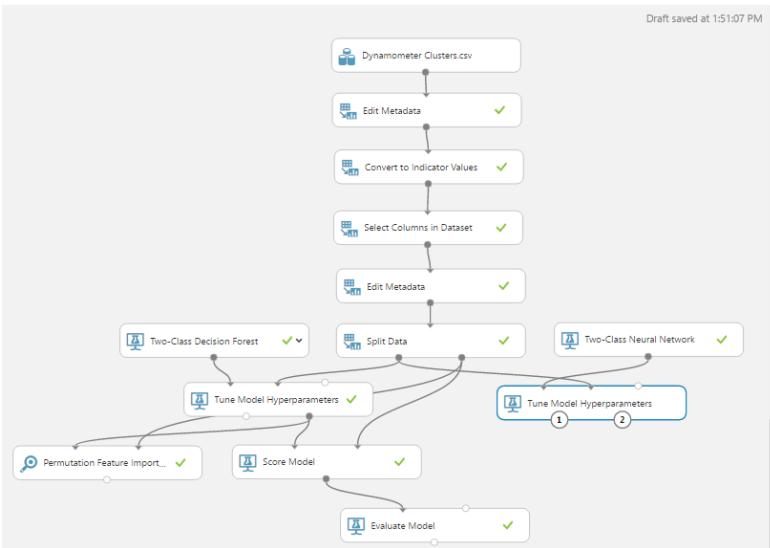


Step 7.3 : Q? Train the Model



- Next, click and hold on the bottom middle circle of your “Two-Class Neural Network” module and drag the line to the top left circle of the new “Tune Model Hyperparameters” module.
- Similarly, click and hold on the bottom left circle of your “Split Data” module and drag the line to the top right circle of the new “Tune Model Hyperparameters” module.

Step 7.4 : Q? Train the Model



- Machine Learning
 - Click on “Tune Model Hyperparameters”
 - Set Metric for measuring performance for classification to “AUC”
 - Set Metric for measuring performance for regression to “Root of mean squared error”
 - Then click on “Launch column selector”

Select a single column



Select “With Rules”

Select “Include” from the drop down menu

Step 7.5 : Q? Train the Model

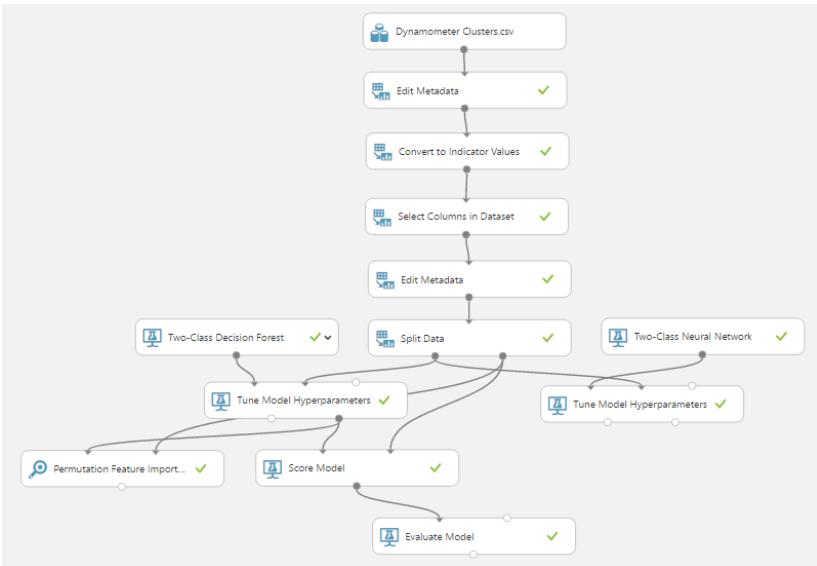
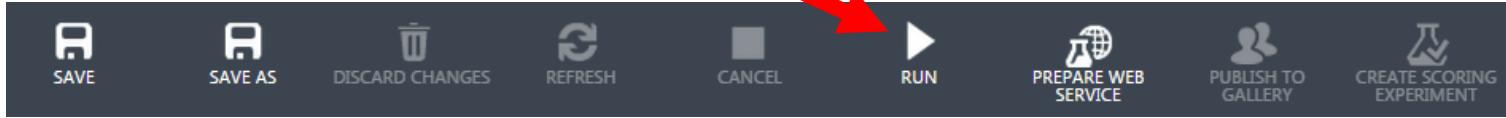
Select a single column

The screenshot shows a software interface for selecting a single column. On the left, there's a sidebar with 'BY NAME' and 'WITH RULES' buttons. Below them are two dropdown menus: 'Include' and 'column names'. The 'column names' dropdown is open, showing a list of variables: 'well', 'date', 'PC1', 'PC2', 'PC3', and 'Assignments-44'. A red arrow points from the text 'Click here and a drop down list appears' to the 'column names' dropdown button. Another red arrow points from the text 'Next, click on the check circle' to the checkmark icon at the bottom right of the dropdown list.

- Click here and a drop down list appears
- Select one variable at a time
 - Assignments-44
- Next, click on the check circle

Step 7.6 : Run the Experiment

Click "Run"

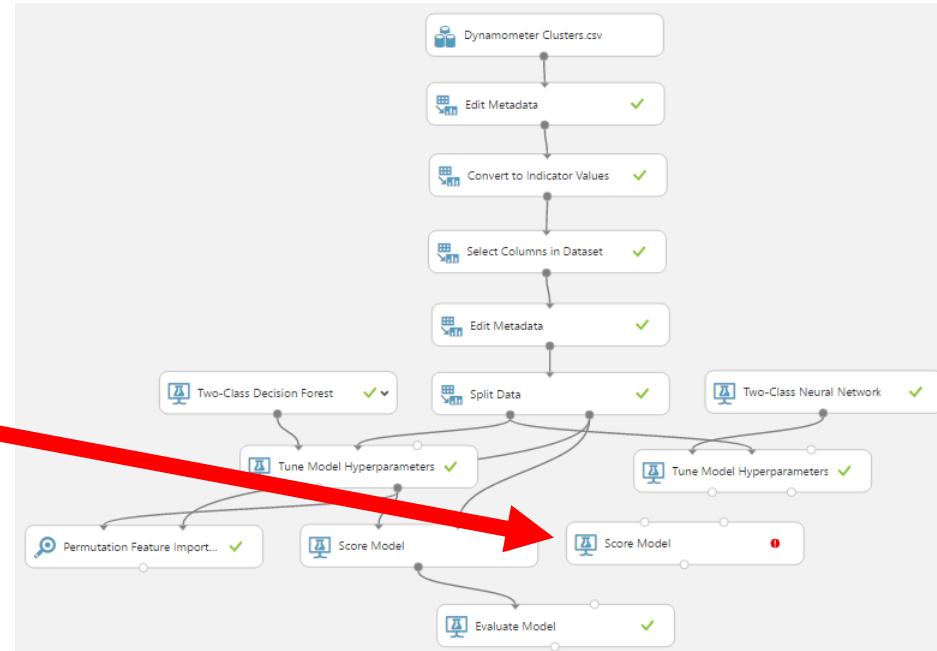


- After Azure Machine Learning “runs”, you will see a green check mark in all the modules
- All it took to train a machine learning model was drag and drop

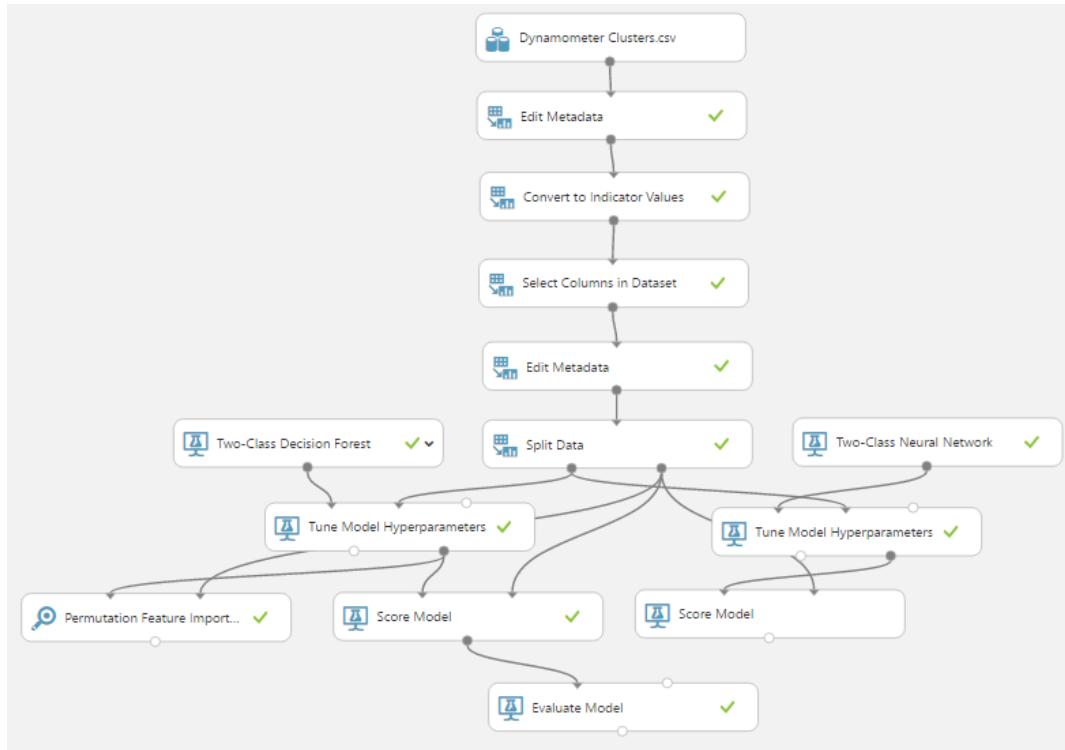
Step 7.7 : Adding the Score Model Module

- ▶ Data Format Conversions
- ▶ Data Input and Output
- ▶ Data Transformation
- ▶ Feature Selection
- ▶ Machine Learning
 - ▶ Evaluate
 - ▶ Initialize Model
 - ▶ **Score**
 - Apply Transformation**
 - Assign Data to Cluste...**
 - Score Matchbox Reco..**
 - Score Model**
 - ▶ Train
- ▶ OpenCV Library Modules

- Open “Machine Learning” and then “Score” from the navigation pane at the left
- Drag “Score Model” module to the canvas
 - “Score Model” scores a trained classification or a regression model

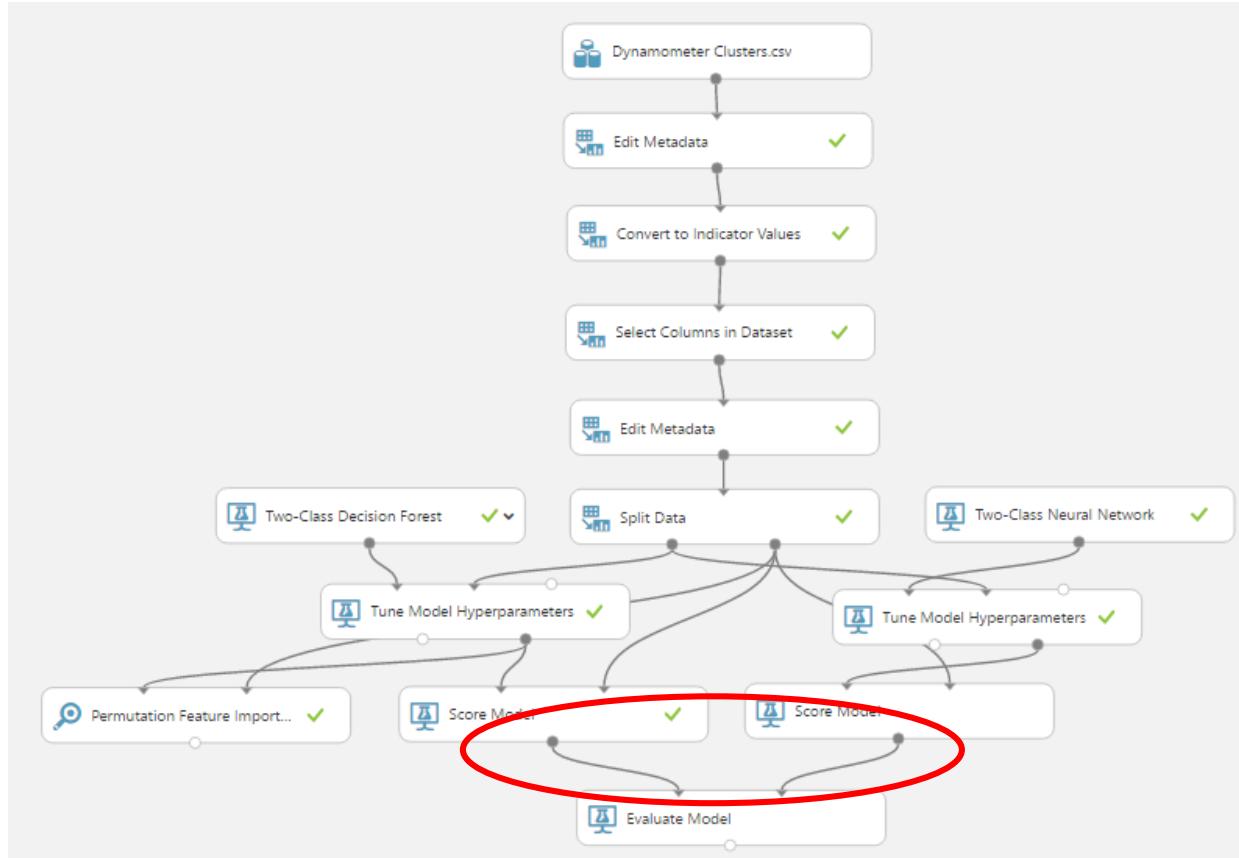


Step 7.8 : Adding the Score Model Module



- Next, click and hold on the bottom right circle of your new “Tune Model Hyperparameters” module and drag the line to the top left circle of the “Score Model” module.
- Similarly, click and hold on the bottom right circle of your “Split Data” module and drag the line to the top right circle of the “Score Model” module.

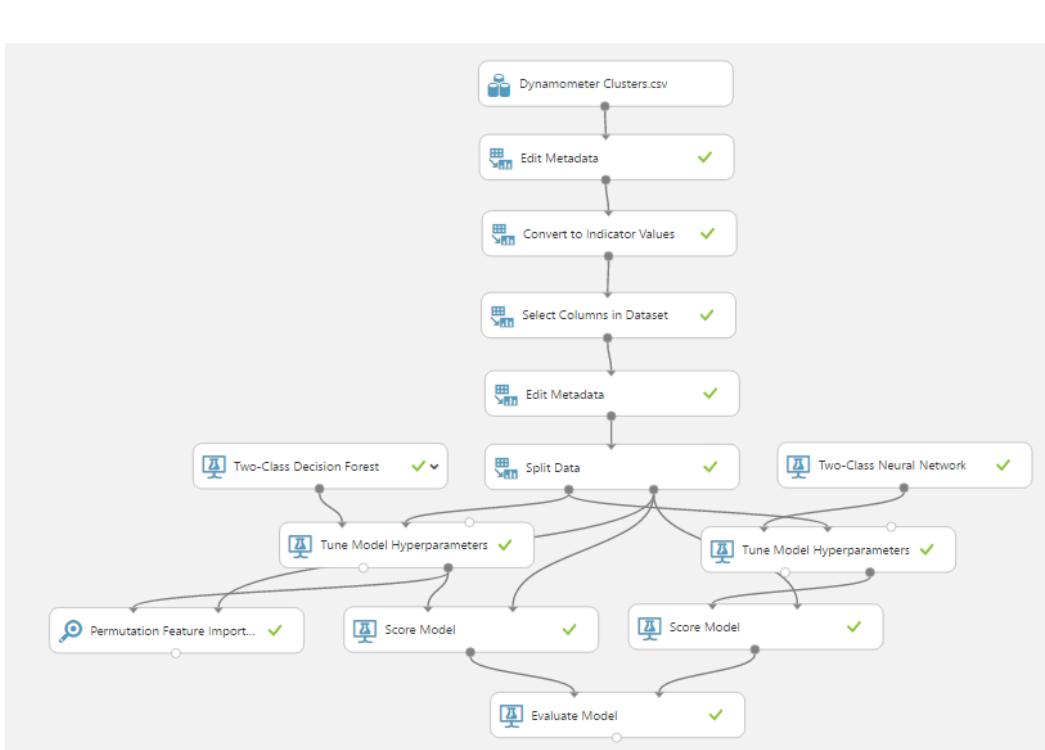
Step 7.9 : Connect the to the Evaluate Model



- Click and hold on the bottom middle circle of your “Score Model” module on the right side
- While holding down the mouse button, drag the line to the top right circle of “Evaluate Model” module

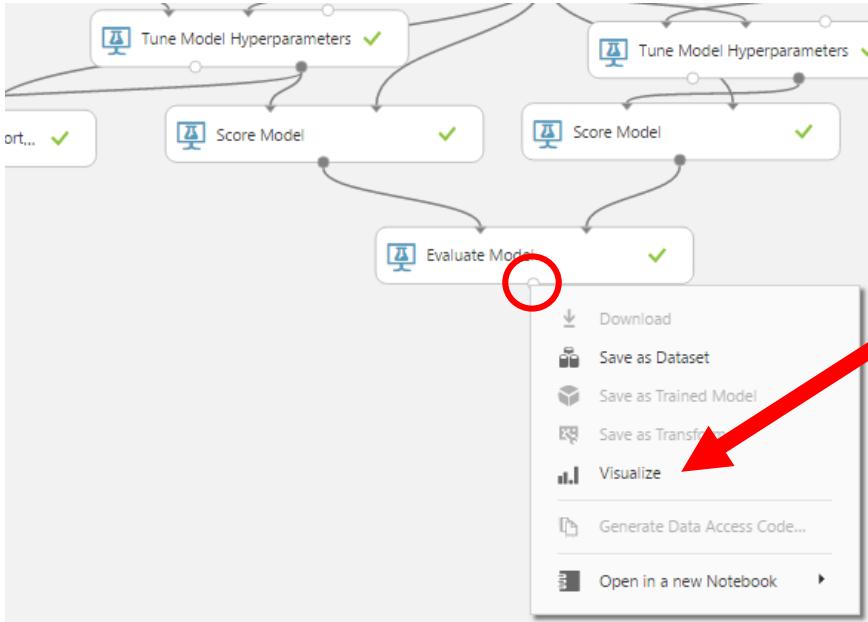
Step 7.10 : Run the Experiment

Click "Run"

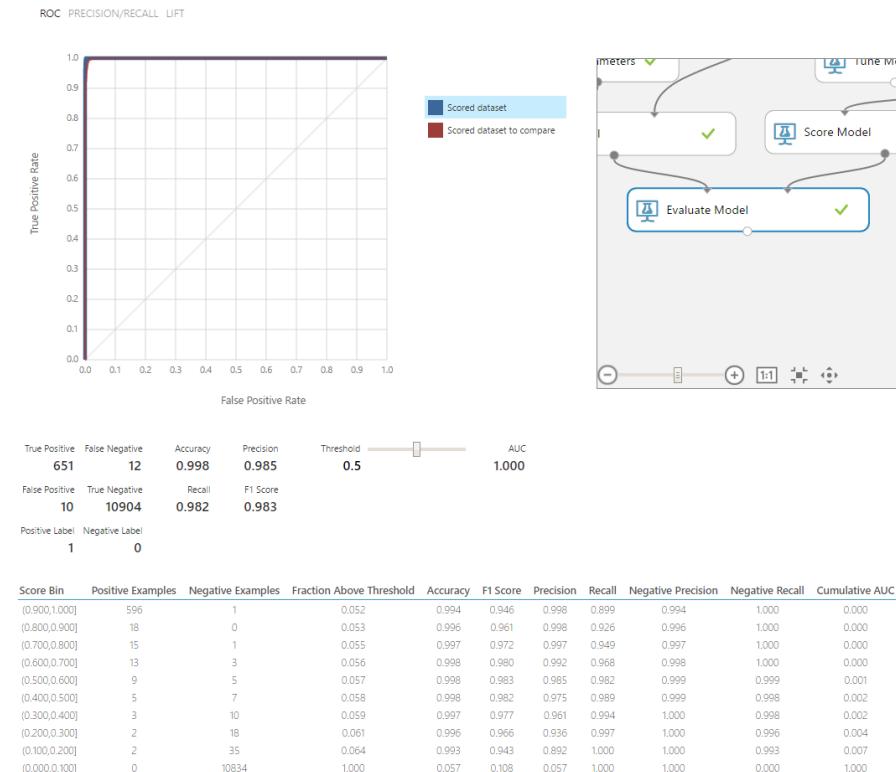


- After Azure Machine Learning “runs”, you will see a green check mark in all the modules
- All it took to train a machine learning model was drag and drop

Step 7.11 : Visualize the Results

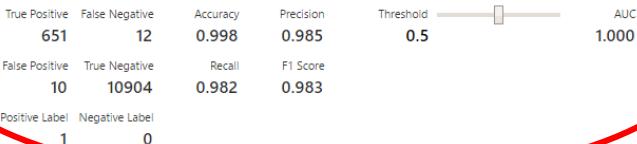
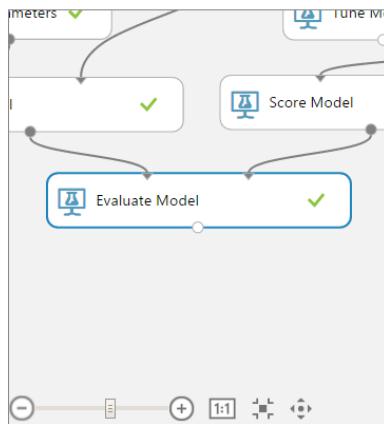
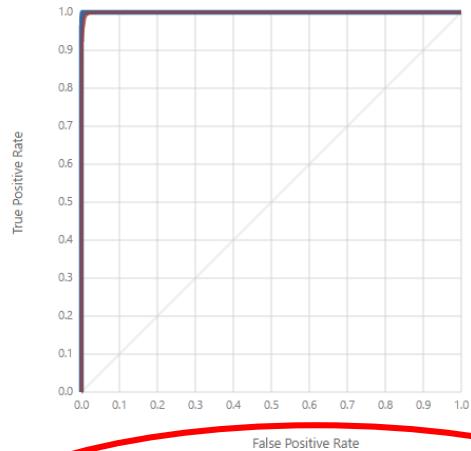


- Right click on the bottom-middle circle of "Evaluate Model" and select "Visualize"



Step 7.12 : Interpret the Results

ROC PRECISION/RECALL LIFT



Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	596	1	0.052	0.994	0.946	0.998	0.899	0.994	1.000	0.000
(0.800,0.900]	18	0	0.053	0.996	0.961	0.998	0.926	0.996	1.000	0.000
(0.700,0.800]	15	1	0.055	0.997	0.972	0.997	0.949	0.997	1.000	0.000
(0.600,0.700]	13	3	0.056	0.998	0.980	0.992	0.968	0.998	1.000	0.000
(0.500,0.600]	9	5	0.057	0.998	0.983	0.985	0.982	0.999	0.999	0.001
(0.400,0.500]	5	7	0.058	0.998	0.982	0.975	0.989	0.999	0.998	0.002
(0.300,0.400]	3	10	0.059	0.997	0.977	0.961	0.994	1.000	0.998	0.002
(0.200,0.300]	2	18	0.061	0.996	0.966	0.936	0.997	1.000	0.996	0.004
(0.100,0.200]	2	35	0.064	0.993	0.943	0.892	1.000	1.000	0.993	0.007
(0.000,0.100]	0	10834	1.000	0.057	0.108	0.057	1.000	1.000	0.000	1.000

- ROC - Accuracy is measured by the area under the ROC curve. An area of 1 represents a perfect test; an area of .5 represents a worthless test
 - .90-1 = excellent
 - .80-.90 = good
 - .70-.80 = fair
 - .60-.70 = poor
 - .50-.60 = fail
- The algorithm "Two-Class Decision Forest" performs best in terms of four metrics: "Accuracy", "Precision", "Recall", and "F-Score".