



# Azure Machine Learning

Instructor: Mithun Prasad, PhD  
Email: [miprasad@Microsoft.com](mailto:miprasad@Microsoft.com)

# WHAT YOU WILL BE ABLE TO DO AFTER THIS TRAINING

Build a Data Science experiment using ML studio.

Gain familiarity with Data Science components of the studio.

Customize Data Science components in the studio.

# MACHINE LEARNING 101

The first way of thinking about ML is by the type of information or input given to a system.

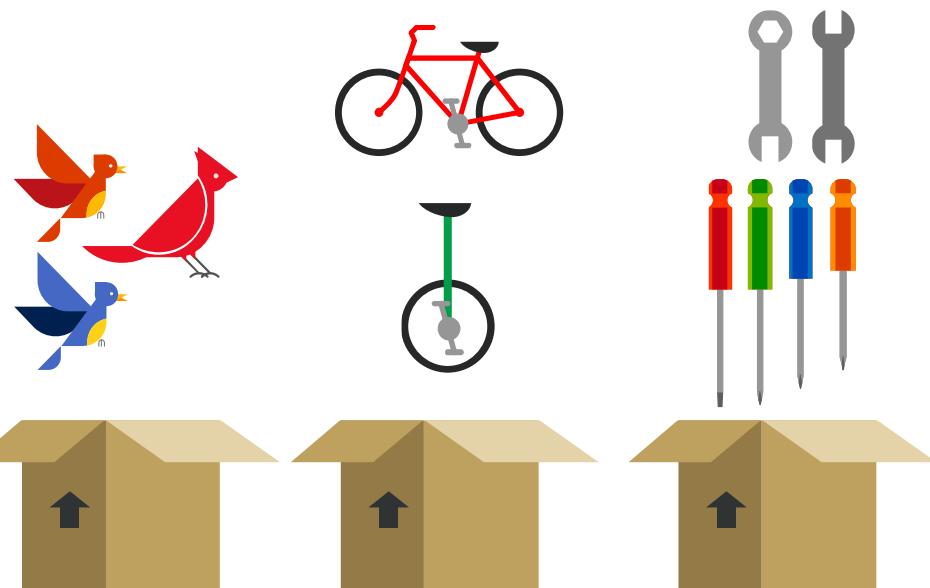
- 1. Supervised learning** – we get the data and the labels e.g. linear regression
- 2. Unsupervised learning** – only get the data (no labels) e.g. clustering
- 3. Reinforcement learning** – reward/penalty based information (feedback)

Another way of categorizing ML approaches, is to the desired output:

- 1. Classification** (e.g. decision tree)
- 2. Regression** (e.g. linear regression)
- 3. Clustering** (e.g. k-means)
- 4. Density estimation** (e.g. histograms)
- 5. Dimensionality reduction** (e.g. principal component analysis)

# MACHINE LEARNING CAPABILITIES

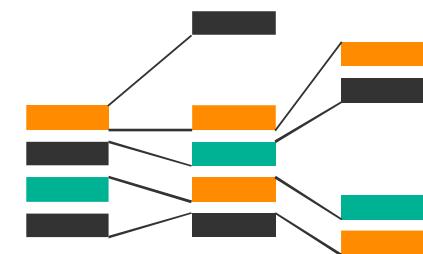
Which category  
*(Classification)*



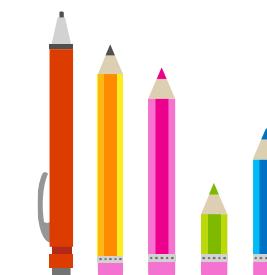
How much/many  
*(Regression)*



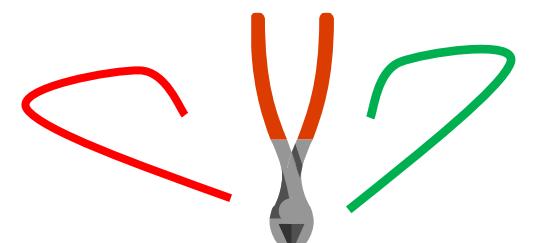
Which group  
*(Clustering,  
Recommender)*



Is it odd  
*(Anomaly)*



Which action  
*(Reinforcement  
Learning)*



# MACHINE LEARNING VISION

Make Machine Learning accessible to every enterprise, data scientist, developer, information worker, consumer, and device anywhere in the world

No need to be an expert

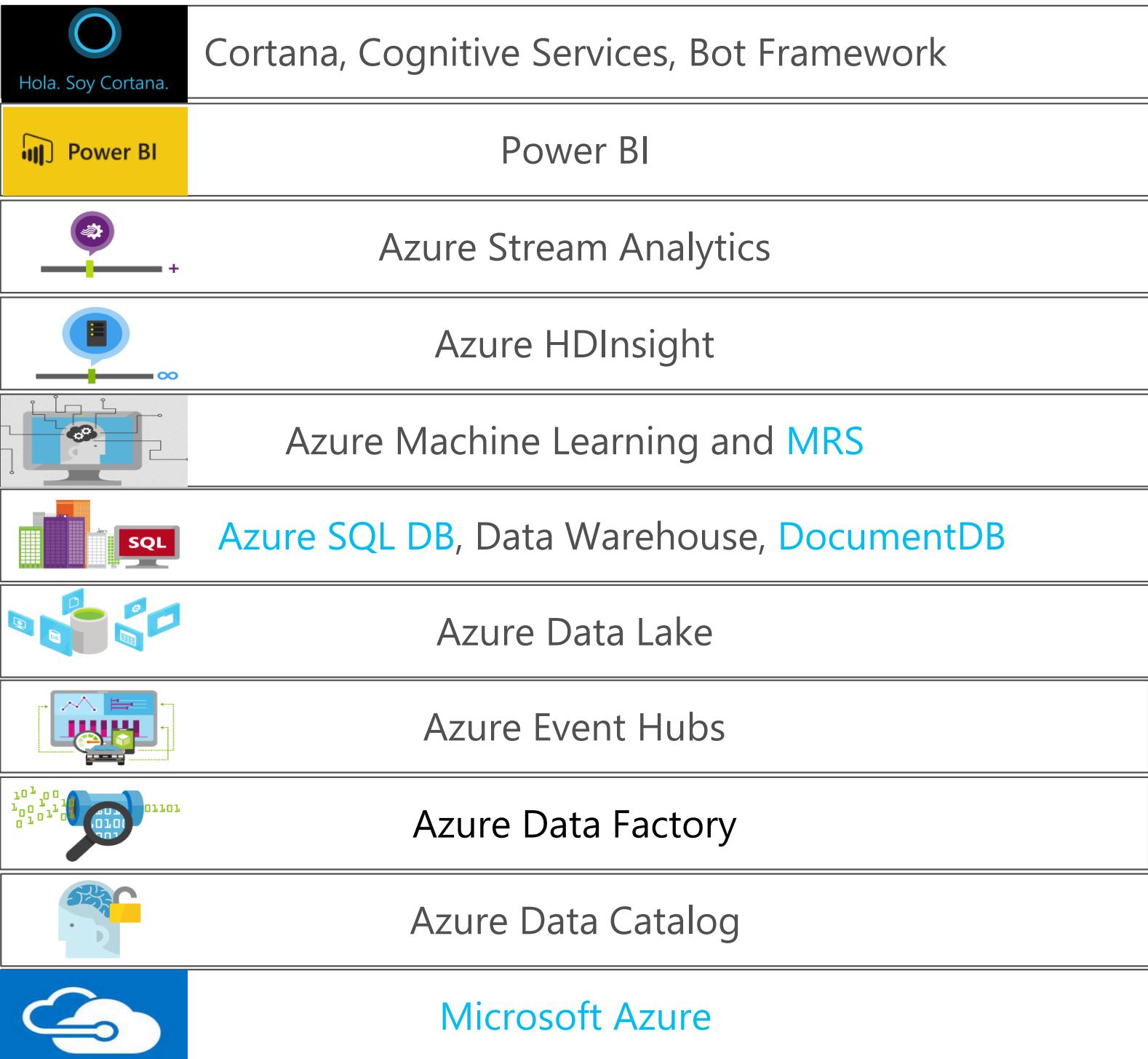
# MACHINE LEARNING TERMINOLOGY

Term	Definition
<b>Training set</b>	set of data used to learn a model
<b>Test set</b>	set of data used to test a model
<b>Feature</b>	a variable (continuous, discrete, categorical, etc.) aka column
<b>Target</b>	Label (associated with dependent variable, what we predict)
<b>Learner</b>	Model or algorithm
<b>Fit, Train</b>	Learn a model with an ML algorithm using a training set
<b>Predict</b>	w/ supervised learning, give a label to an unknown datum(data). w/unsupervised decide if new data is weird, in which group, or what to do next with the new data
<b>Accuracy</b>	percentage of correct predictions $((TP + TN) / \text{total})$
<b>Precision</b>	Percentage of correct positive predictions $(TP / (FP + TP))$
<b>Recall</b>	Percentage of positive cases caught $(TP / (FN + TP))$

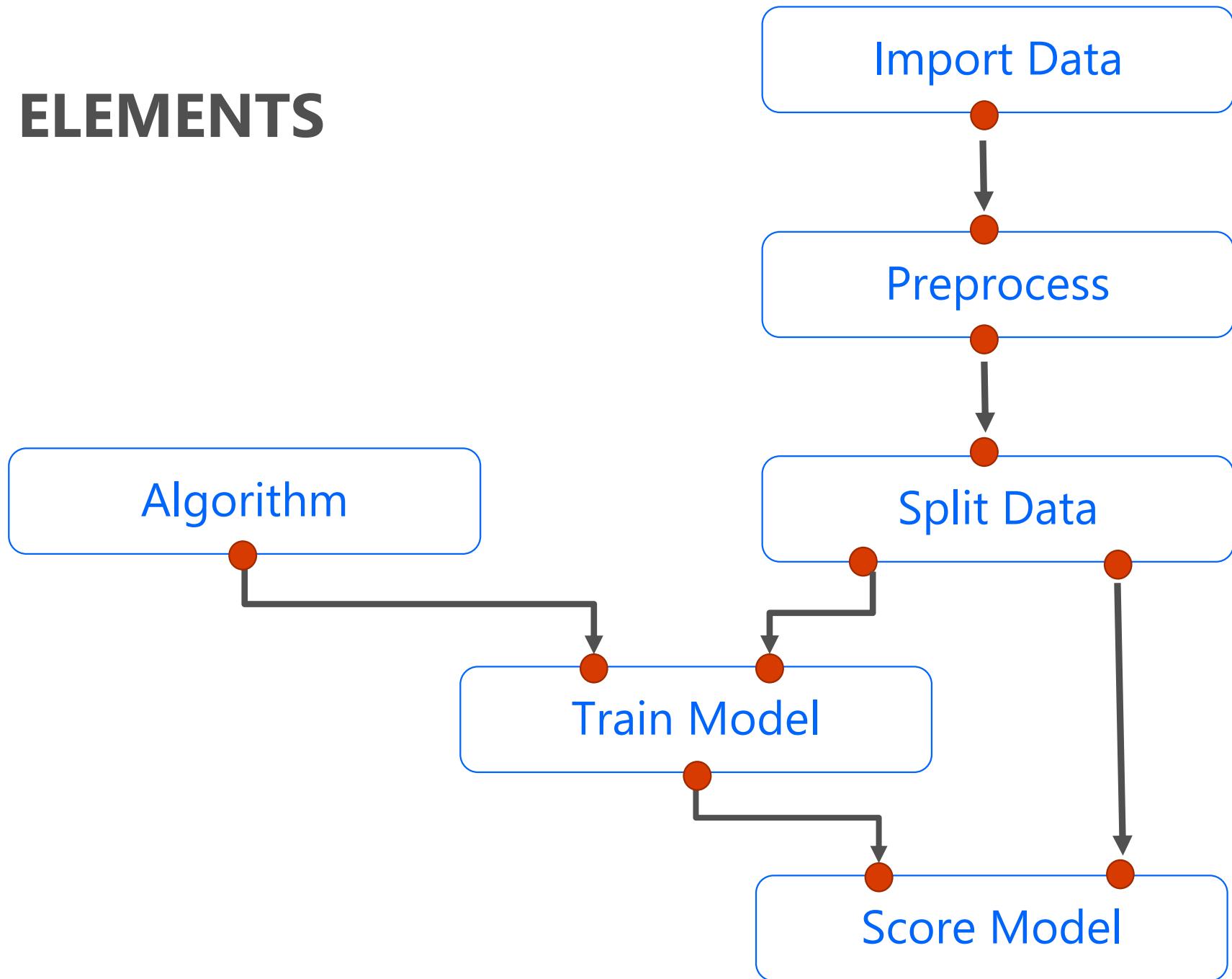
## CORTANA INTELIGENCE IN A SENTENCE

Cortana Intelligence is a **Platform** and a **Process** to perform advanced analytics from start to finish

# THE CORTANA INTELLIGENCE PLATFORM



# AZURE ML ELEMENTS



# KEY AZURE ML CONCEPTS

- **Workspace**

Roughly equivalent to an account. Cloud-based container for data, experiments, notebooks. Units for sharing.

- **Studio**

Web-based GUI for authoring experiments in a workspace.

- **Experiment**

A job, containing a data flow graph (DAG)

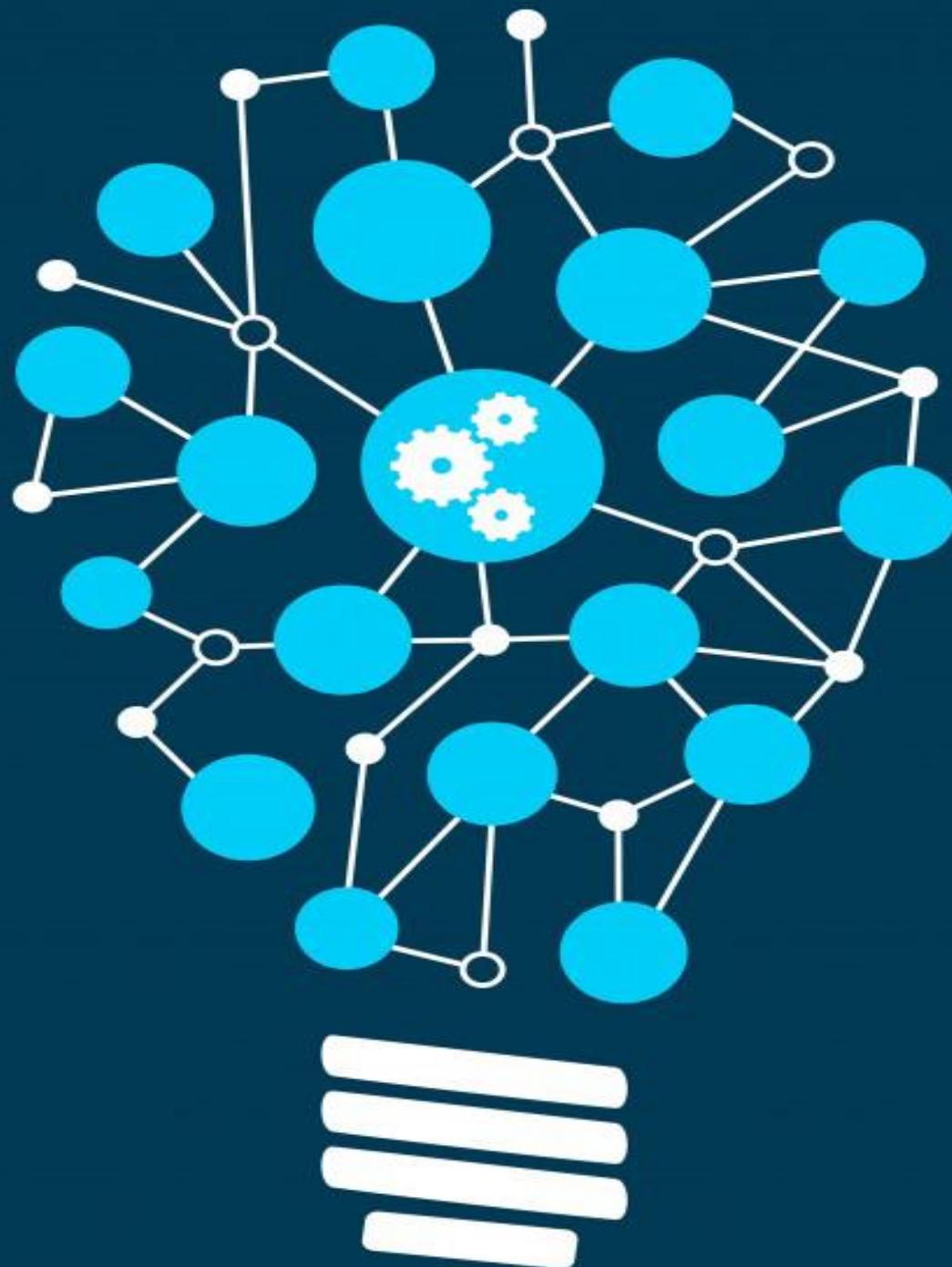
- **DataSet**

Rectangular data structure, similar to database table.

- **Module**

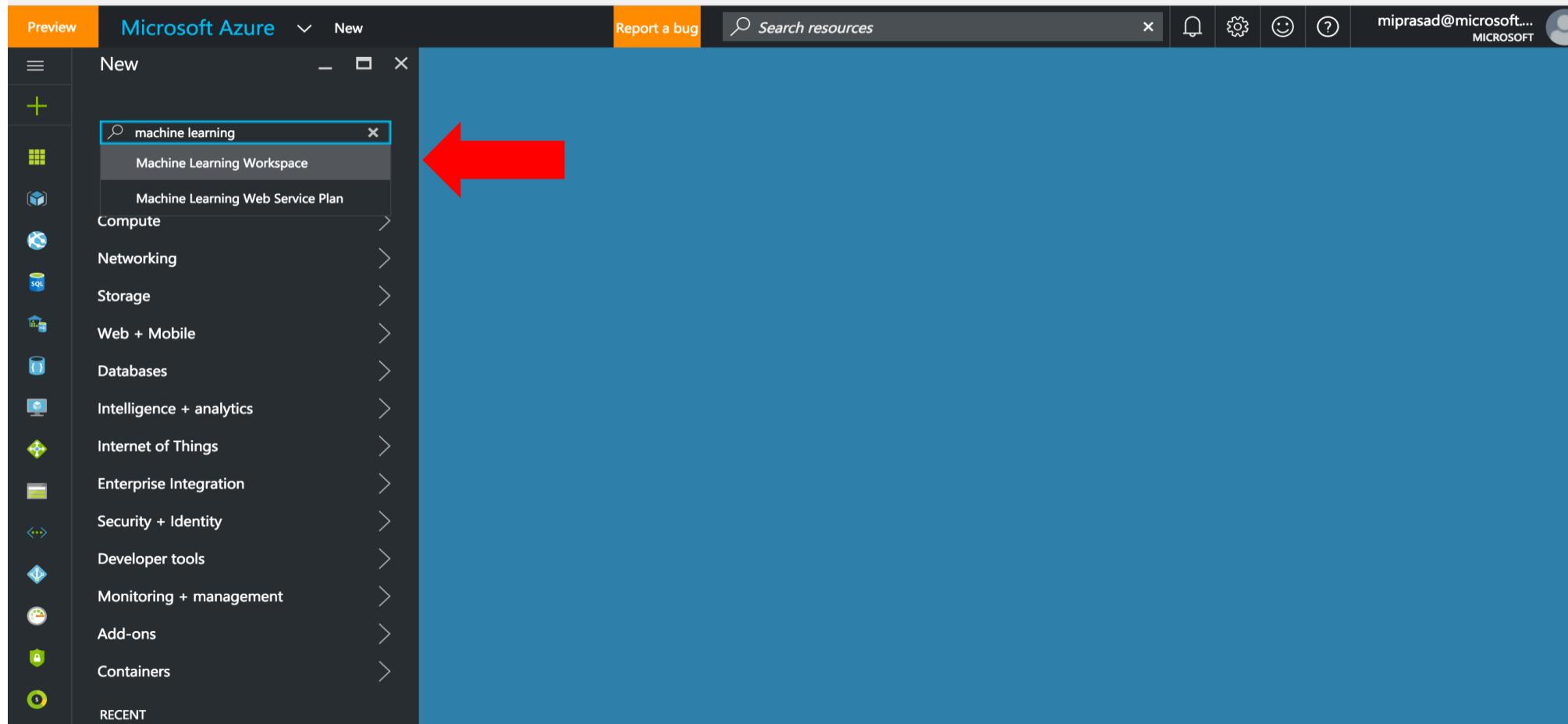
Units of work within a graph

# AZURE ML STUDIO AND THE TEAM DATA SCIENCE PROCESS



# WORKSPACE

portal.azure.com



# WORKSPACE

Preview Microsoft Azure Machine Learning Workspace Report a bug Search resources x

Machine Learning Wor... Machine Learning Workspace

Machine Learning Workspace

\* Workspace name Enter the workspace name

\* Subscription Azure Pass

\* Resource group ⓘ  
Create new Use existing  
Select a resource group !

\* Location South Central US

\* Storage account ⓘ  
Create new Use existing  
Enter the storage account name

Workspace pricing tier ⓘ Standard

\* Web service plan ⓘ  
Create new Use existing  
Enter the plan name

Pin to dashboard

Create Automation options

miprasad@microsoft... MICROSOFT



# AZURE ML STUDIO - SIGN UP

<https://studio.azureml.net/#>

The screenshot shows the Microsoft Azure Machine Learning Studio homepage. At the top, there's a navigation bar with the text "Microsoft Azure Machine Learning Studio" and icons for help, user profile, and sign-in. Below the header, a promotional banner says "Introducing: Competitions" with icons of a test tube and a trophy, and a "Learn More" button. The main content area features a photograph of four people working at desks with computers, with a "JUPYTER" and "MACHINE LEARNING" wall in the background. A large play button icon is overlaid on the photo. Below the photo, a section titled "Quick Tour of Azure ML" includes a thumbnail image of the tour.



Welcome to Azure  
Machine Learning

Try it for free

No [Azure subscription](#)? No credit card? No problem! Choose anonymous Guest Access, or sign in with your work or school account, or a Microsoft account.

[Sign Up](#)

Already an Azure ML User?

[Sign in here](#)

[Pricing & FAQ](#)

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



# TIERS

Microsoft Azure

SALES 000-800-100-3928 ▾ MY ACCOUNT PORTAL Search

Why Azure Solutions Products Documentation **Pricing** Partners Blog Resources Support FREE ACCOUNT >

Region:

South Central US

Currency:

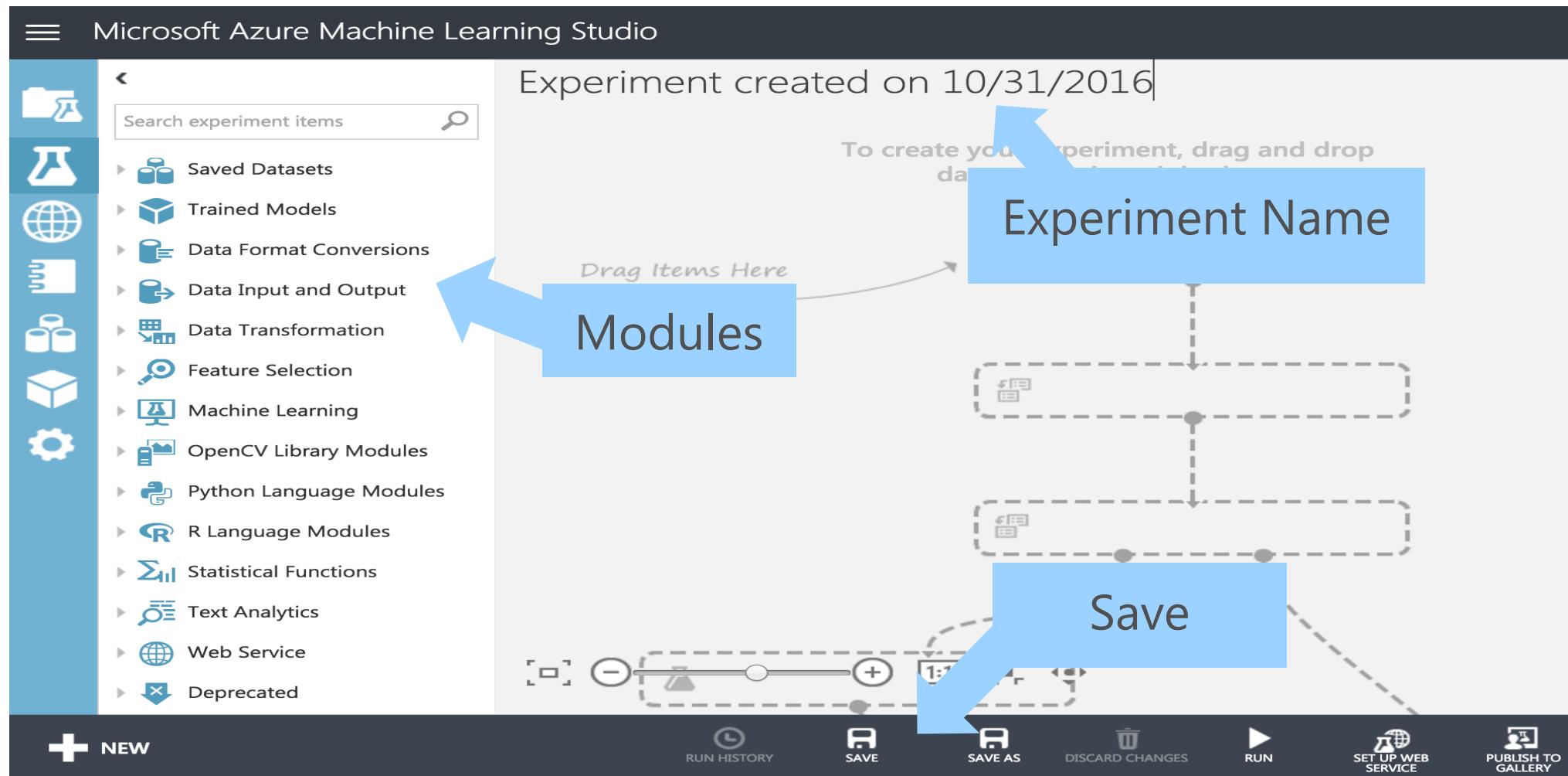
Indian Rupee (₹)

	<b>FREE</b>	<b>STANDARD</b>
Price	Free	₹600.27 per Seat per month ₹60.09 per Studio Experimentation Hour
Azure Subscription	Not Required	Required
Max Number of Modules per Experiment	100	Unlimited
Max Experiment Duration	1 hour per experiment	Up to 7 days per experiment with a maximum of 24 hours per module
Max Storage Space	10 GB	Unlimited - BYO
Read Data from On-Premises SQL <small>Preview</small>	No	Yes
Execution / Performance	Single Node	Multiple Nodes
Production Web API	No	Yes
SLA	No	Yes

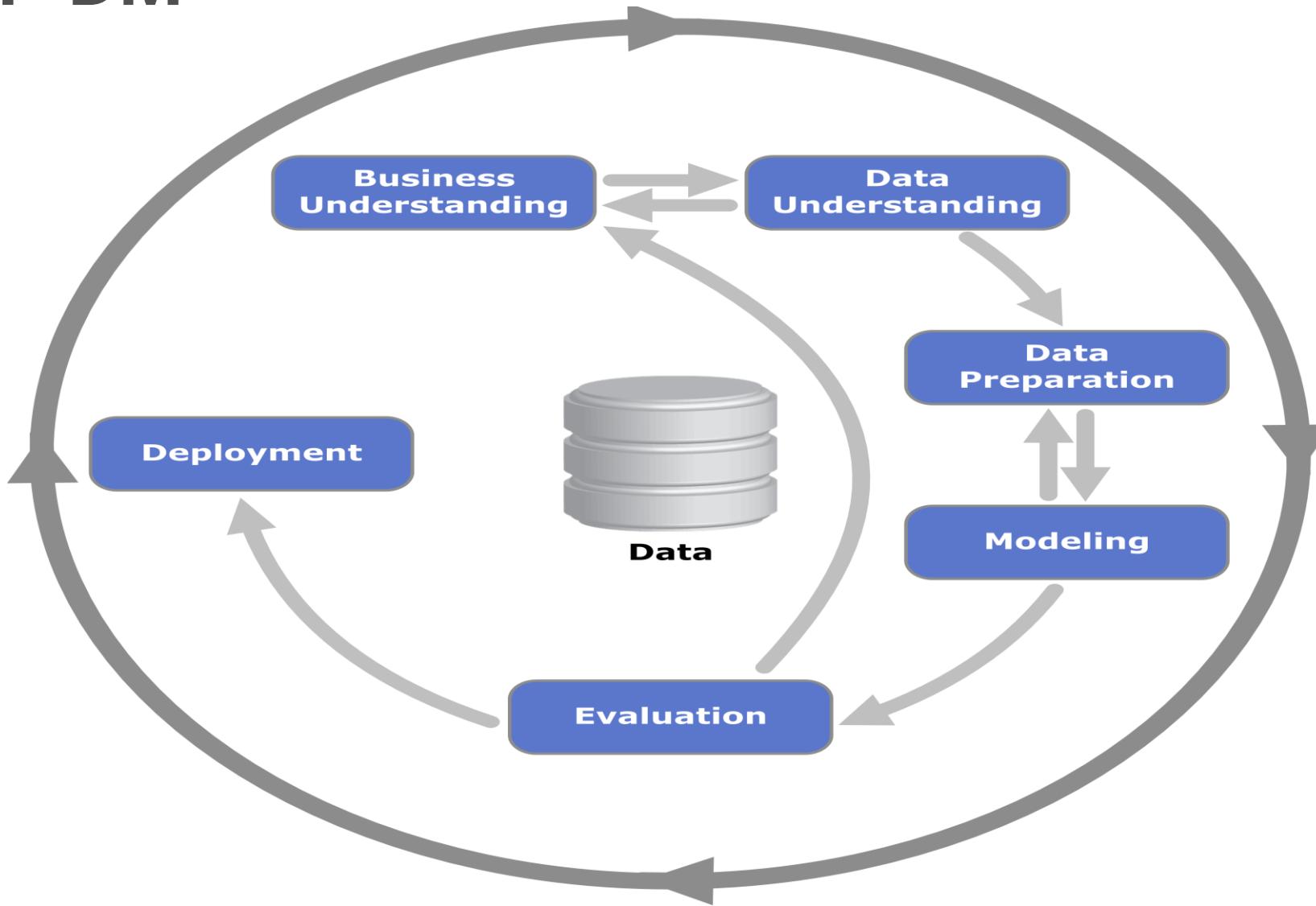
Hourly charges only apply to active use of the service. Where multiple meters are present they are applied concurrently.



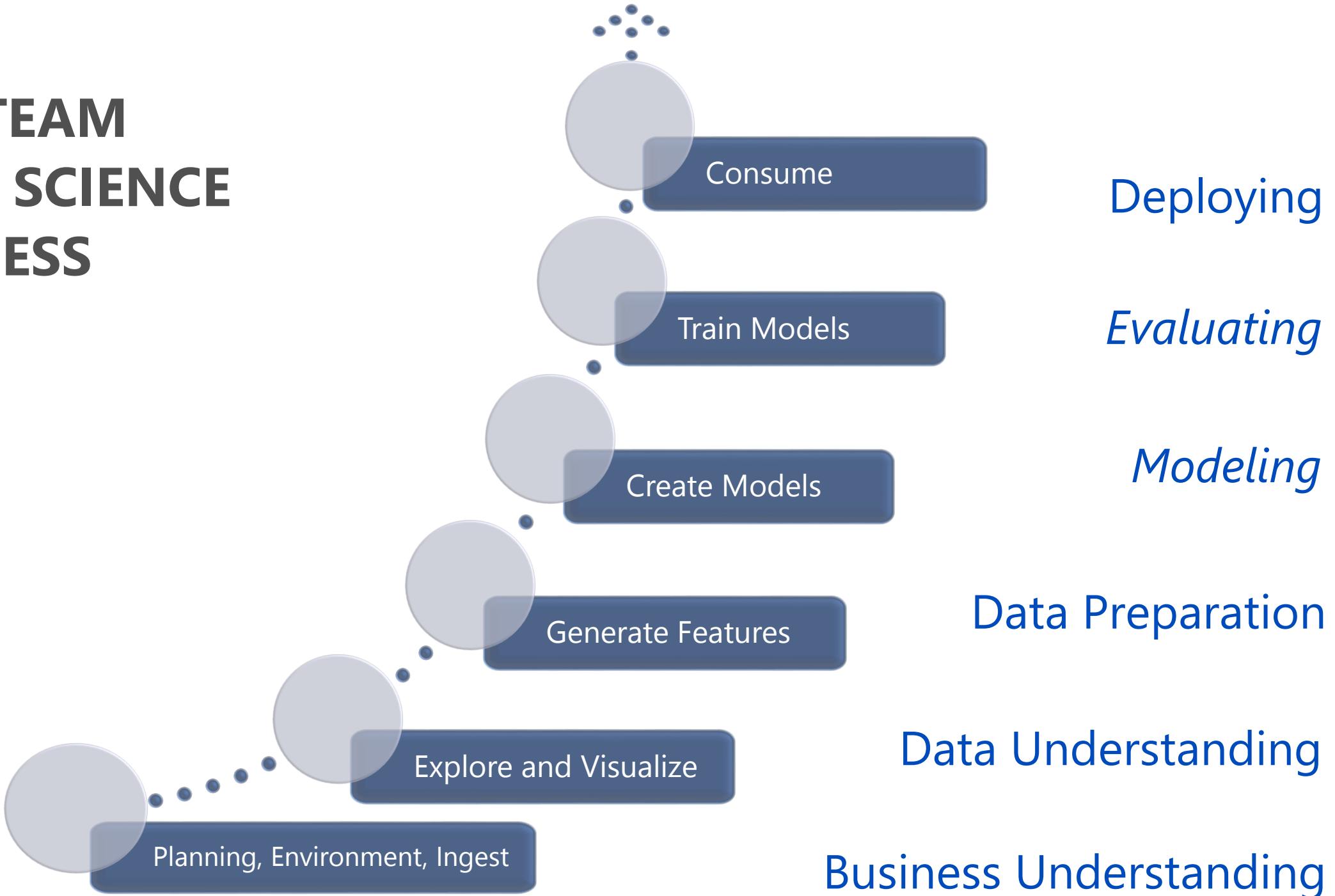
# AZURE ML STUDIO - NAVIGATION



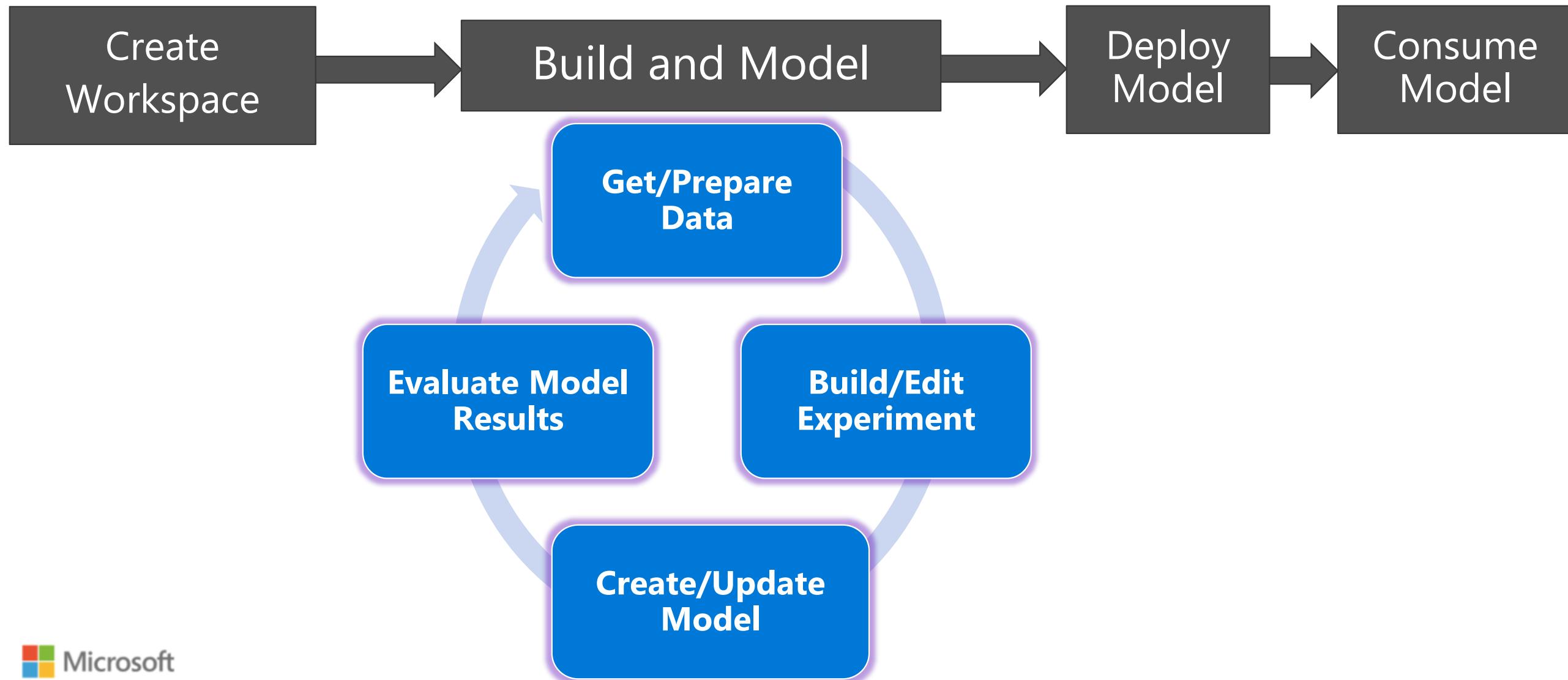
# CRISP-DM



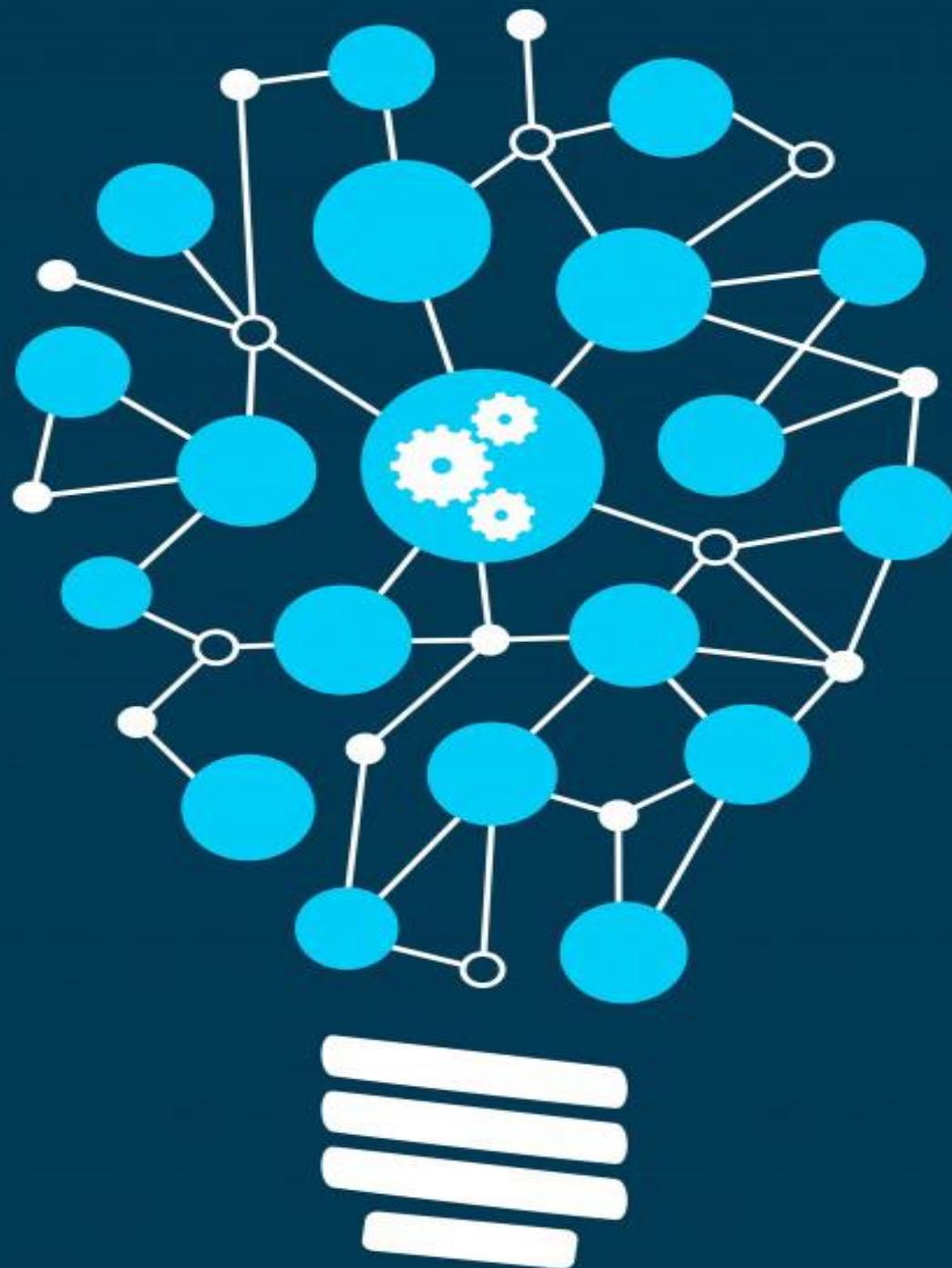
# THE TEAM DATA SCIENCE PROCESS



# CREATING AN EXPERIMENT



# LEARNING RESOURCES



# GALLERY

Microsoft Azure Machine Learning Studio

Mithun Prasad-Free-Work... ? ☰ 😊 🚙

Cortana Intelligence

Azure Machine Learning

Gallery

EXPERIMENTS SAMPLES

NAME	AUTHOR	STATUS	LAST EDITED	PROJECT
IrisExperimentCrossV...	miprasad	Finished	10/28/2016 1:02:29 PM	None
Lab1 [Predictive Exp.]	miprasad	Finished	10/28/2016 10:20:09 ...	None
Lab1	miprasad	Finished	10/28/2016 10:08:55 ...	None
Lab1	miprasad	Draft	10/27/2016 4:22:59 PM	None
IrisExperimentCrossV...	miprasad	Draft	10/24/2016 10:39:21 ...	None
IrisExperiment	miprasad	Draft	10/20/2016 3:37:13 PM	None
IrisExperimentSubset...	miprasad	Finished	10/20/2016 1:48:13 PM	None
IrisExperiment [Predi...	miprasad	Finished	10/19/2016 4:52:44 PM	None
Test	miprasad	Failed	10/19/2016 12:07:42 ...	Test

Import Data

Select Columns in Dataset

Multiclass Logistic Regression

Cross Validate Model

+ NEW

DELETE ADD TO PROJECT

# GALLERY

Cortana Intelligence Gallery

Browse all Industries Solutions Experiments Machine Learning APIs Custom Modules Competitions More

## Experiments

Explore predictive analytic experiments contributed by Microsoft and the data science community that solve interesting problems or demonstrate advanced machine learning techniques. Use these experiments as starting points to develop your own solutions.

How to contribute to the Gallery

NEWS

SPORTS

LIFESTYLE

MONEY

TECH

FRAUD

EXPERIMENT

Online Fraud Detection: Step 1 of 5: Generate tagged data

Microsoft

Sample 9: Split, partition and sample system

Microsoft

# DOCUMENTATION

A screenshot of the Microsoft Azure Machine Learning Studio interface. The top navigation bar shows the title "Microsoft Azure Machine Learning Studio" and the user "Mithun Prasad-Free-Work...". A red arrow points down to the user profile icon in the top right corner.

The main area displays the "experiments" section. On the left, a sidebar menu includes "PROJECTS", "EXPERIMENTS" (which is selected), "WEB SERVICES", "NOTEBOOKS", "DATASETS", "TRAINED MODELS", and "SETTINGS". Below the sidebar is a "NEW" button with a plus sign.

The central part of the screen shows a table of experiments:

	NAME	AUTHOR	STATUS	LAST EDITED	PROJECT
<input type="checkbox"/>	IrisExperimentCrossV...	miprasad	Finished	10/28/2016 1:02:29 PM	None
<input type="checkbox"/>	Lab1 [Predictive Exp.]	miprasad	Finished	10/28/2016 10:20:09 ...	None
<input type="checkbox"/>	Lab1	miprasad	Finished	10/28/2016 10:08:55 ...	None
<input type="checkbox"/>	Lab1	miprasad	Draft	10/27/2016 4:22:59 PM	None
<input type="checkbox"/>	IrisExperimentCrossV...	miprasad	Draft	10/24/2016 10:39:21 ...	None
<input type="checkbox"/>	IrisExperiment	miprasad	Draft	10/20/2016 3:37:13 PM	None
<input type="checkbox"/>	IrisExperimentSubset...	miprasad	Finished	10/20/2016 1:48:13 PM	None
<input type="checkbox"/>	IrisExperiment [Predi...	miprasad	Finished	10/19/2016 4:52:44 PM	None
<input type="checkbox"/>	Test	miprasad	Failed	10/19/2016 12:07:42 ...	Test

Below the table, there are buttons for "DELETE" and "ADD TO PROJECT". To the right of the table, a workflow diagram is shown:

```
graph TD; ImportData[Import Data] --> SelectColumns[Select Columns in Dataset]; SelectColumns --> MulticlassLogistic[Multiclass Logistic Regression]; MulticlassLogistic --> CrossValidate[Cross Validate Model]
```

The workflow consists of four steps: "Import Data", "Select Columns in Dataset", "Multiclass Logistic Regression", and "Cross Validate Model". Each step has a green checkmark indicating it is completed.

# DOCUMENTATION

Microsoft Azure

SALES 000-800-100-3928 | MY ACCOUNT | PORTAL | Search | [FREE ACCOUNT >](#)

Why Azure Solutions Products Documentation Pricing Partners Blog Resources Support

Search for docs

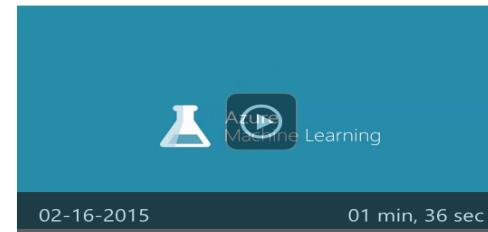
- ▼ Overview
  - [What's Machine Learning?](#)
  - [Data science for beginners](#)
  - [What's the Studio?](#)
  - [Studio capabilities](#)
  - [Frequently asked questions](#)
- Get started
  - [Data science process](#)
  - [Algorithms](#)
  - [Develop a model](#)
  - [Work with data](#)
  - [Extend with R & Python](#)

## Machine Learning documentation

Learn how data scientists and developers can embed predictive analytics into applications

Build your first experiment in Azure Machine Learning Studio [Start tutorial >](#)

## Machine Learning videos



[Overview of Azure Machine Learning](#)

[See more videos >](#)



[Getting Started with Azure Machine Learning Studio](#)



[Using R in Azure Machine Learning Studio](#)

Try Machine Learning Studio



# FORUMS

Microsoft Azure Machine Learning Studio

Mithun Prasad-Free-Work...    

experiments

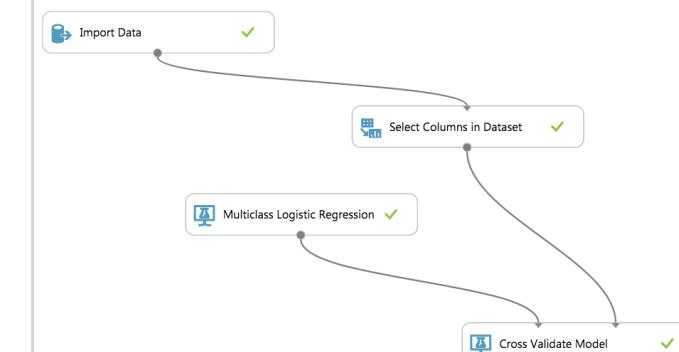
MY EXPERIMENTS SAMPLES

	NAME	AUTHOR	STATUS	LAST EDITED	PROJECT
<input type="checkbox"/>	IrisExperimentCrossV...	miprasad	Finished	10/28/2016 1:02:29 PM	None
<input type="checkbox"/>	Lab1 [Predictive Exp.]	miprasad	Finished	10/28/2016 10:20:09 ...	None
<input type="checkbox"/>	Lab1	miprasad	Finished	10/28/2016 10:08:55 ...	None
<input type="checkbox"/>	Lab1	miprasad	Draft	10/27/2016 4:22:59 PM	None
<input type="checkbox"/>	IrisExperimentCrossV...	miprasad	Draft	10/24/2016 10:39:21 ...	None
<input type="checkbox"/>	IrisExperiment	miprasad	Draft	10/20/2016 3:37:13 PM	None
<input type="checkbox"/>	IrisExperimentSubset...	miprasad	Finished	10/20/2016 1:48:13 PM	None
<input type="checkbox"/>	IrisExperiment [Predi...	miprasad	Finished	10/19/2016 4:52:44 PM	None
<input type="checkbox"/>	Test	miprasad	Failed	10/19/2016 12:07:42 ...	Test



Machine Learning Forums

Provide feedback and request features



Import Data ✓

Select Columns in Dataset ✓

Multiclass Logistic Regression ✓

Cross Validate Model ✓

 NEW

<https://qo.microsoft.com/fwlink/?LinkId=403669&clcid=0x409>

DELETE ADD TO PROJECT

# FORUMS

SALES: 1-800-867-1380 | MY ACCOUNT | PORTAL | SIGN IN | Search | FREE TRIAL | 

**Microsoft Azure**

Features Pricing Documentation Downloads Add-ons Community Support

**Ask a question**

Quick access ▾

Forums [ view all ]

Please select a forum ...

Selected forums

Clear

Machine Learning 

Top answerers Last 30 days

 Hilary Cotter	129
 Magnus (MM8)	77
 Dave Patrick	74
 Erland Sommarskog	63
 Dennis Guo	60

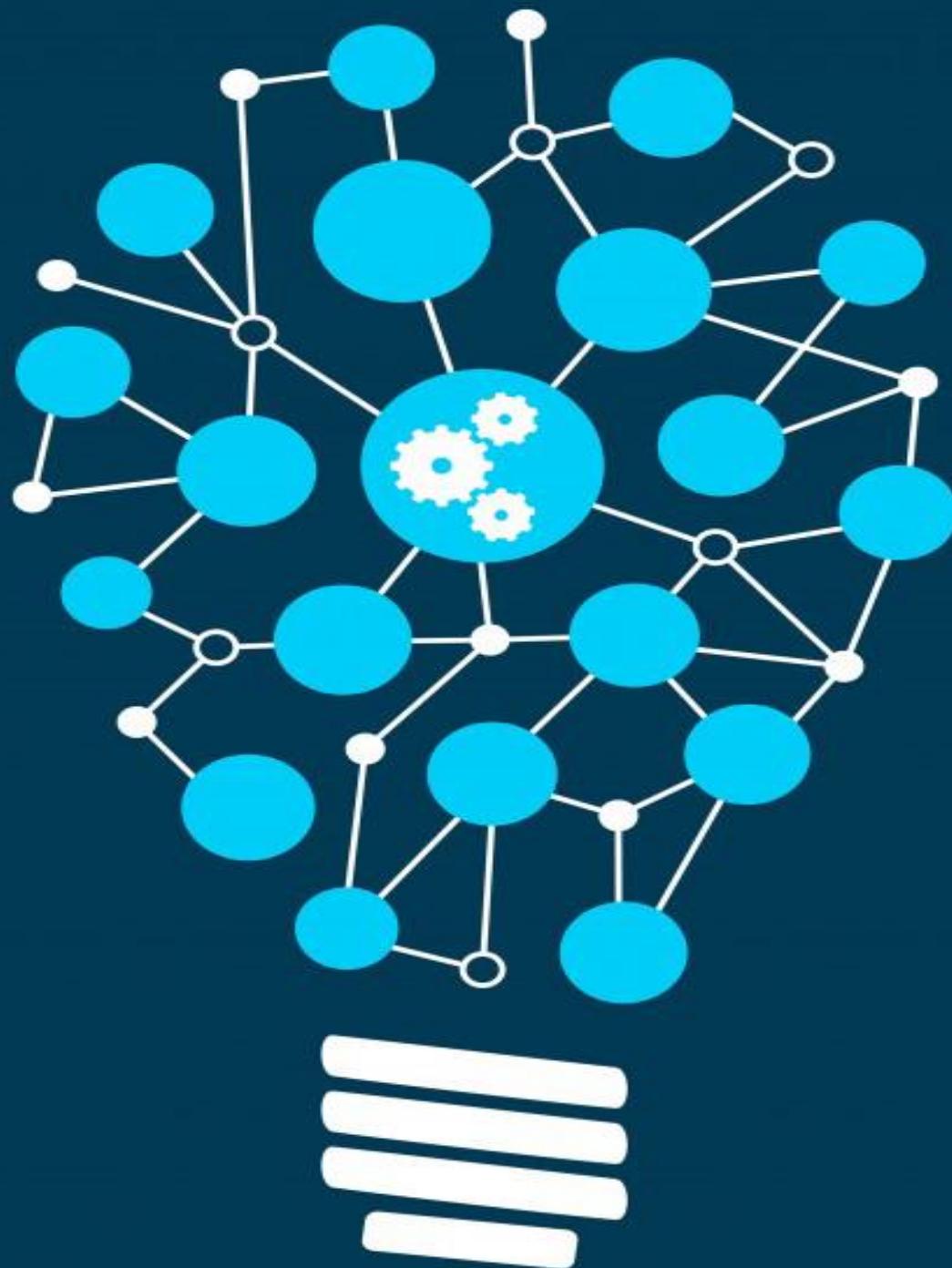
Announcement: 31

Search forum questions 

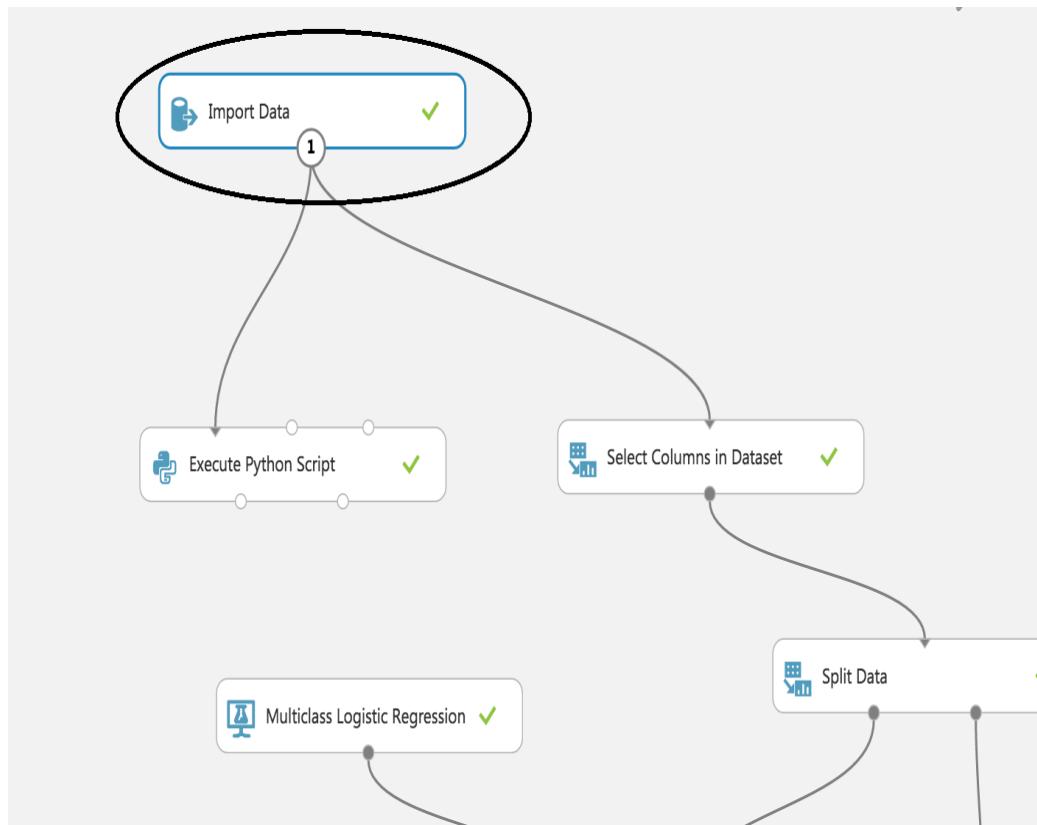
Filter : All threads ▾ Sort : Most recent post ▾

- ▶ **Is there any way to get a print friendly or large image version of an AML experiment?**  
Microsoft Azure > Machine Learning  
Greetings, I have a quite large AML experiment that I need to have a large format print out of it. At this point I have to use a screen capture tool and stitch ...  
Answered | 1 Replies | 40 Views | Created by amx2012 - Sunday, October 30, 2016 9:02 PM | Last reply by Hai Ning - 8 hours 30 minutes ago
- ▶ **Using the AzureML Services REST API instead of MLPS**  
Microsoft Azure > Machine Learning  
Hi, I was going through the MLPS source code. I am not able to run the MLPS module from Mac OS as the powershell core does not currently support the Web.Extensions. However ...  
Answered | 2 Replies | 36 Views | Created by vmandke - Tuesday, November 01, 2016 6:39 AM | Last reply by Hai Ning - 8 hours 32 minutes ago
- ▶ **Deploying a webservice with multiple models**  
Microsoft Azure > Machine Learning  
Hi, Can only one predictive webservice be associated with an experiment? I have an experiment which trains multiple binary classifiers. (Can only create a webservice for one of the ...  
Proposed | 8 Replies | 132 Views | Created by vmandke - Thursday, October 27, 2016 9:55 AM | Last reply by Hai Ning - 8 hours 34 minutes ago
- ▶ **Queuing of the tasks in AzureML**  
Microsoft Azure > Machine Learning  
Hi, I observed that training tasks are queued by default in AzureML. Would a queued task be executed with a guarantee or is it be dropped after a stipulated time ...  
Answered | 1 Replies | 24 Views | Created by vmandke - Tuesday, November 01, 2016 8:21 AM |

# DATA INGESTION AND PREPARATION



# DATA ACCESS (IMPORT)



## Import Data

Data source

Azure SQL Database

Database server name

irismlbd3.database.windows.net

Database name

irisMLDB

User name

miprasad@irismlbd3

Password

.....

Accept any server certificate (insecure)

Database query

```
1 Select * from iris
```

# IMPORT VIA URL

Microsoft Azure Machine Learning Studio

Mithun Prasad-Free-Work... ? 🔍 😊 🚙

Experiment created on 11... Finished running ✓

Import Data 1 ✓

Properties Project

Import Data

Data source: Web URL via HTTP

Data source URL: <https://raw.githubusercontent.com/mithun-prasad/azure-ml/master/Data/irisDataset.csv>

Data format: CSV

CSV or TSV has header row

Use cached results

START TIME: 11/8/2016 2:56:00 PM  
END TIME: 11/8/2016 2:56:10 PM  
ELAPSED TIME: 0:00:10.242  
STATUS CODE: Finished  
STATUS DETAILS: None

[View output log](#)

Quick Help

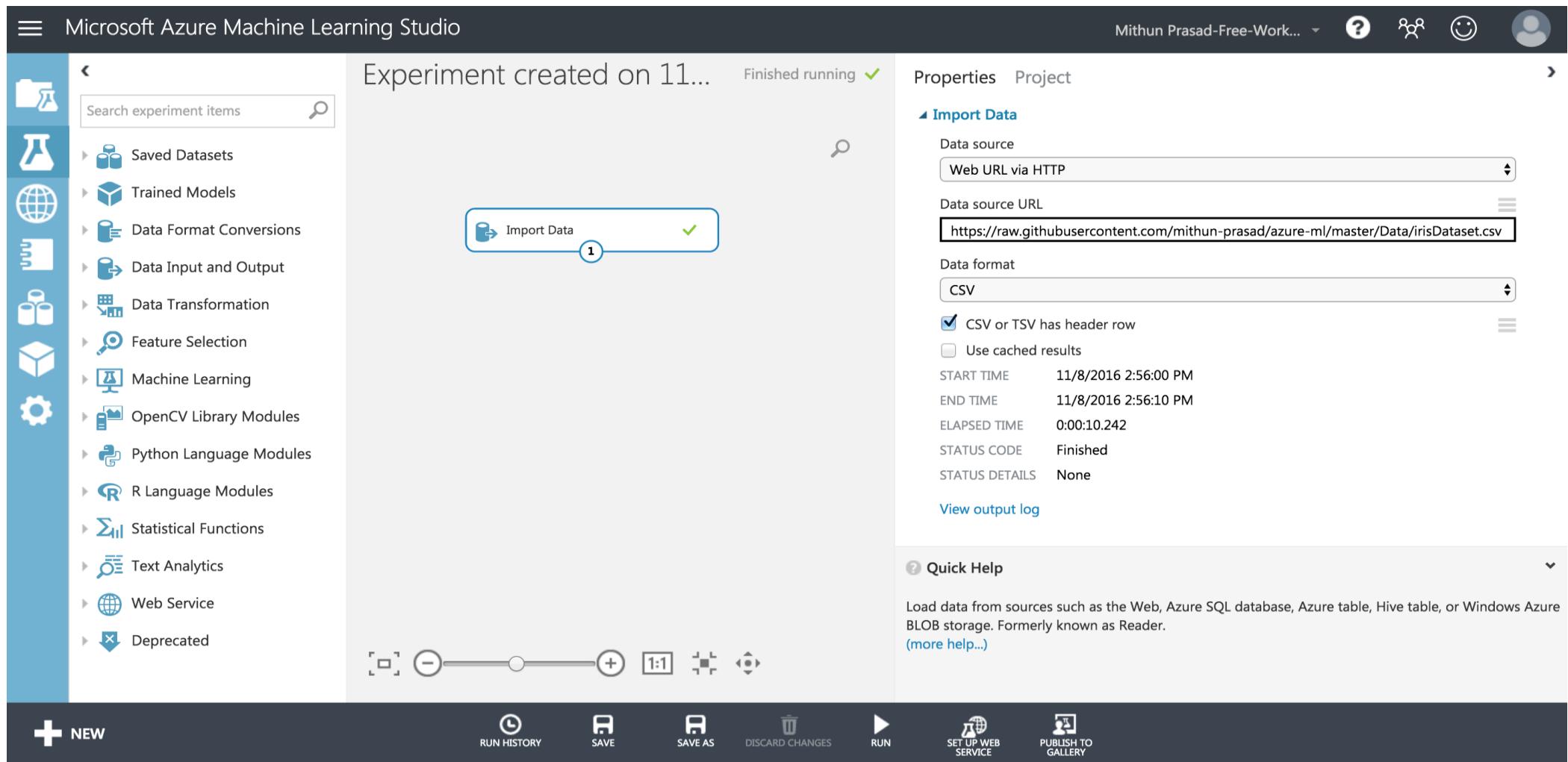
Load data from sources such as the Web, Azure SQL database, Azure table, Hive table, or Windows Azure BLOB storage. Formerly known as Reader.  
[\(more help...\)](#)

Search experiment items

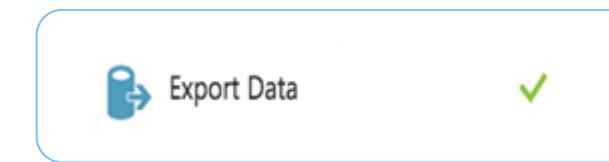
- Saved Datasets
- Trained Models
- Data Format Conversions
- Data Input and Output
- Data Transformation
- Feature Selection
- Machine Learning
- OpenCV Library Modules
- Python Language Modules
- R Language Modules
- Statistical Functions
- Text Analytics
- Web Service
- Deprecated

+

NEW RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY



# DATA ACCESS (EXPORT)



## Data Format Conversion

- Convert to ARFF
- Convert to CSV
- Convert to Dataset
- Convert to SVMLight
- Convert to TSV

### Export Data

Please specify data destination

Azure SQL Database

Database server name

irisMldb3.database.windows.net

Database name

irisMLDB

Server user account name

miprasad@irisMldb3

Server user account password

.....

Accept any server certificate (insecure)

Comma separated list of columns to be saved

sepallength, petallength, Scored Labels

Data table name

irisOutput

Comma separated list of datatable columns

sepallength, petallength, scoredclass

# ARFF FILE

```
% 1. Title: Iris Plants Database
%
% 2. Attribute Information:
%    1. sepal length in cm
%    2. sepal width in cm
%    3. petal length in cm
%    4. petal width in cm
%    5. class:
%       -- Iris Setosa
%       -- Iris Versicolour
%       -- Iris Virginica
|
@RELATION iris

@ATTRIBUTE sepallength REAL
@ATTRIBUTE sepalwidth REAL
@ATTRIBUTE petallength REAL
@ATTRIBUTE petalwidth REAL
@ATTRIBUTE class {Iris-setosa,Iris-versicolor,Iris-virginica}

@DATA
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
5.4,3.9,1.7,0.4,Iris-setosa
4.6,3.4,1.4,0.3,Iris-setosa
5.0,3.4,1.5,0.2,Iris-setosa
4.4,2.9,1.4,0.2,Iris-setosa
4.9,3.1,1.5,0.1,Iris-setosa
5.4,3.7,1.5,0.2,Iris-setosa
4.8,3.4,1.6,0.2,Iris-setosa
4.8,3.0,1.4,0.1,Iris-setosa
4.3,3.0,1.1,0.1,Iris-setosa
5.8,4.0,1.2,0.2,Iris-setosa
5.7,4.4,1.5,0.4,Iris-setosa
5.4,3.9,1.3,0.4,Iris-setosa
5.1,3.5,1.4,0.3,Iris-setosa
5.7,3.8,1.7,0.3,Iris-setosa
5.1,3.8,1.5,0.3,Iris-setosa
5.4,3.4,1.7,0.2,Iris-setosa
5.1,3.7,1.5,0.4,Iris-setosa
4.6,3.6,1.0,0.2,Iris-setosa
5.1,3.3,1.7,0.5,Iris-setosa
4.8,3.4,1.9,0.2,Iris-setosa
5.0,3.0,1.6,0.2,Iris-setosa
```

# IRIS DATASET



## Features

Sepallength

Sepalwidth

Petallength

Petalwidth

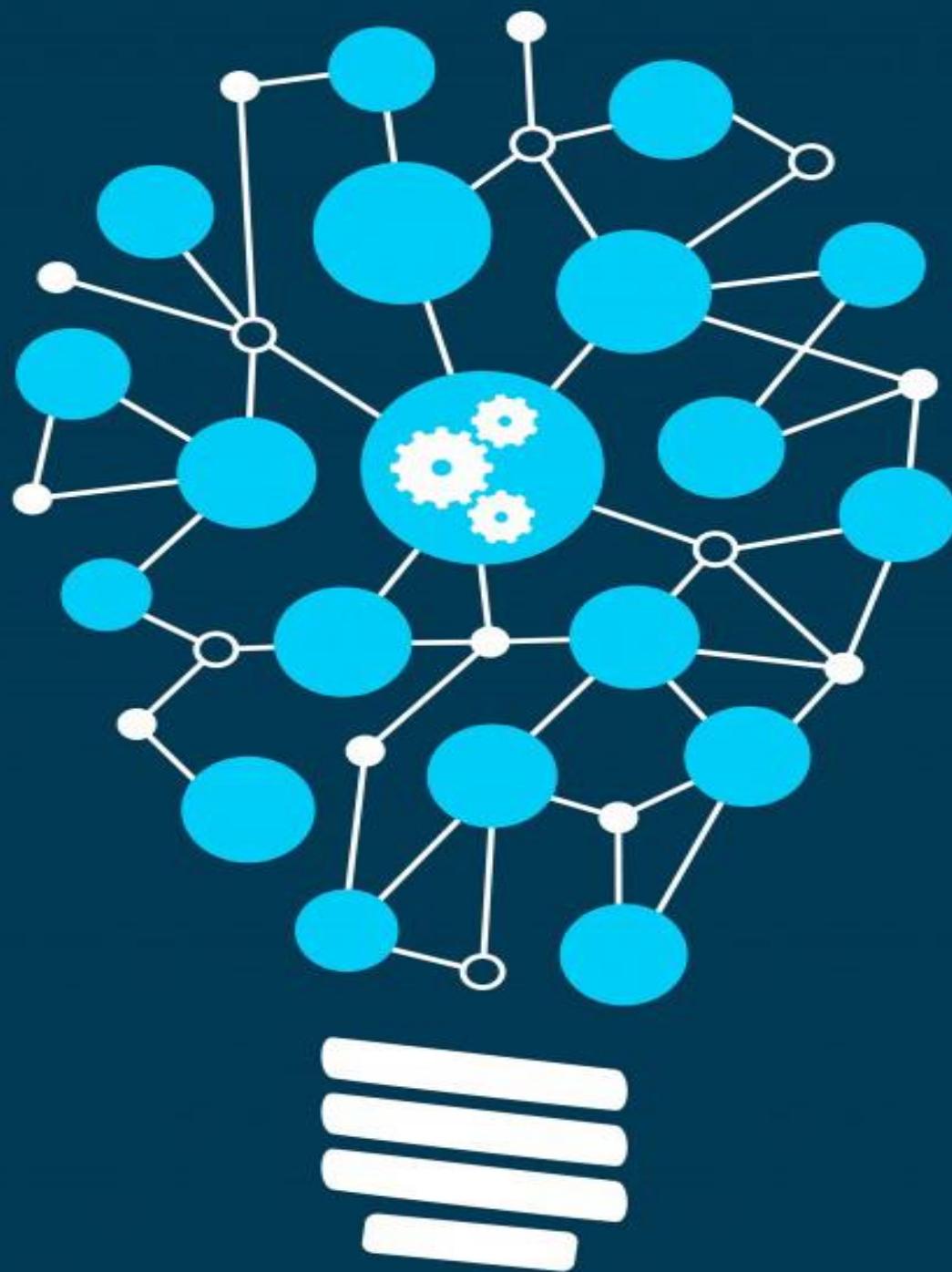
## Classes

Iris-setosa

Iris-versicolor

Iris-virginica

# Exploratory Statistics



# STATISTICS

Microsoft Azure Machine Learning Studio

Mithun Prasad-Free-Work... ? ☺ ☺ ☺

Training experiment Predictive experiment

Lab1

Finished running ✓

irisDataset.csv

Submitted by mprasad  
Size 4.50 KB  
Format GenericCSV  
Created on 10/28/2016...

Properties Project

dataset

1

Download Visualize Generate Data Access Code... Open in a new Notebook

Multiclass Decision Forest Train Model Score Model

Quick Help

Search experiment items

Saved Datasets Trained Models 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 Web Service Deprecated

NEW RUN HISTORY SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY

```
graph TD; DS[irisDataset.csv] --> MD[Multiclass Decision Forest]; MD --> TM[Train Model]; TM --> SM[Score Model];
```

# STATISTICS

Microsoft Azure Machine Learning Studio

Mithun Prasad-Free-Work... ? ☰ ☺ ☻ ☻

Lab1 > irisDataset.csv > dataset

rows 150 columns 5

sepallength sepalwidth petallength petalwidth class

sepallength	sepalwidth	petallength	petalwidth	class
5.1	3.5	1.4	0.2	Iris-setosa
4.9	3	1.4	0.2	Iris-setosa
4.7	3.2	1.3	0.2	Iris-setosa
4.6	3.1	1.5	0.2	Iris-setosa
5	3.6	1.4	0.2	Iris-setosa
5.4	3.9	1.7	0.4	Iris-setosa
4.6	3.4	1.4	0.3	Iris-setosa
5	3.4	1.5	0.2	Iris-setosa
4.4	2.9	1.4	0.2	Iris-setosa
4.9	3.1	1.5	0.1	Iris-setosa
5.4	3.7	1.5	0.2	Iris-setosa

view as

Statistics

Mean	5.8433
Median	5.8
Min	4.3
Max	7.9
Standard Deviation	0.8281
Unique Values	35
Missing Values	0
Feature Type	Numeric Feature

Visualizations

sepallength Histogram

compare to

Properties Project

Score Model

NEW RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY

# DESCRIPTIVE STATISTICS

Microsoft Azure Machine Learning Studio

Mithun Prasad-Free-Work... Properties Project

Experiment created on 11/8/2016

Experiment created on 11/8/2016 > Summarize Data > Results dataset

rows 5 columns 23

Feature	Count	Unique Value Count	Missing Value Count	Min	Max	Mean	Mean Deviation	1st Quartile
sepallength	150	35	0	4.3	7.9	5.843333	0.687556	5.1
sepalwidth	150	23	0	2	4.4	3.054	0.333093	2.8
petallength	150	43	0	1	6.9	3.758667	1.56192	1.6
petalwidth	150	22	0	0.1	2.5	1.198667	0.658933	0.3
class	150	3	0					

view as

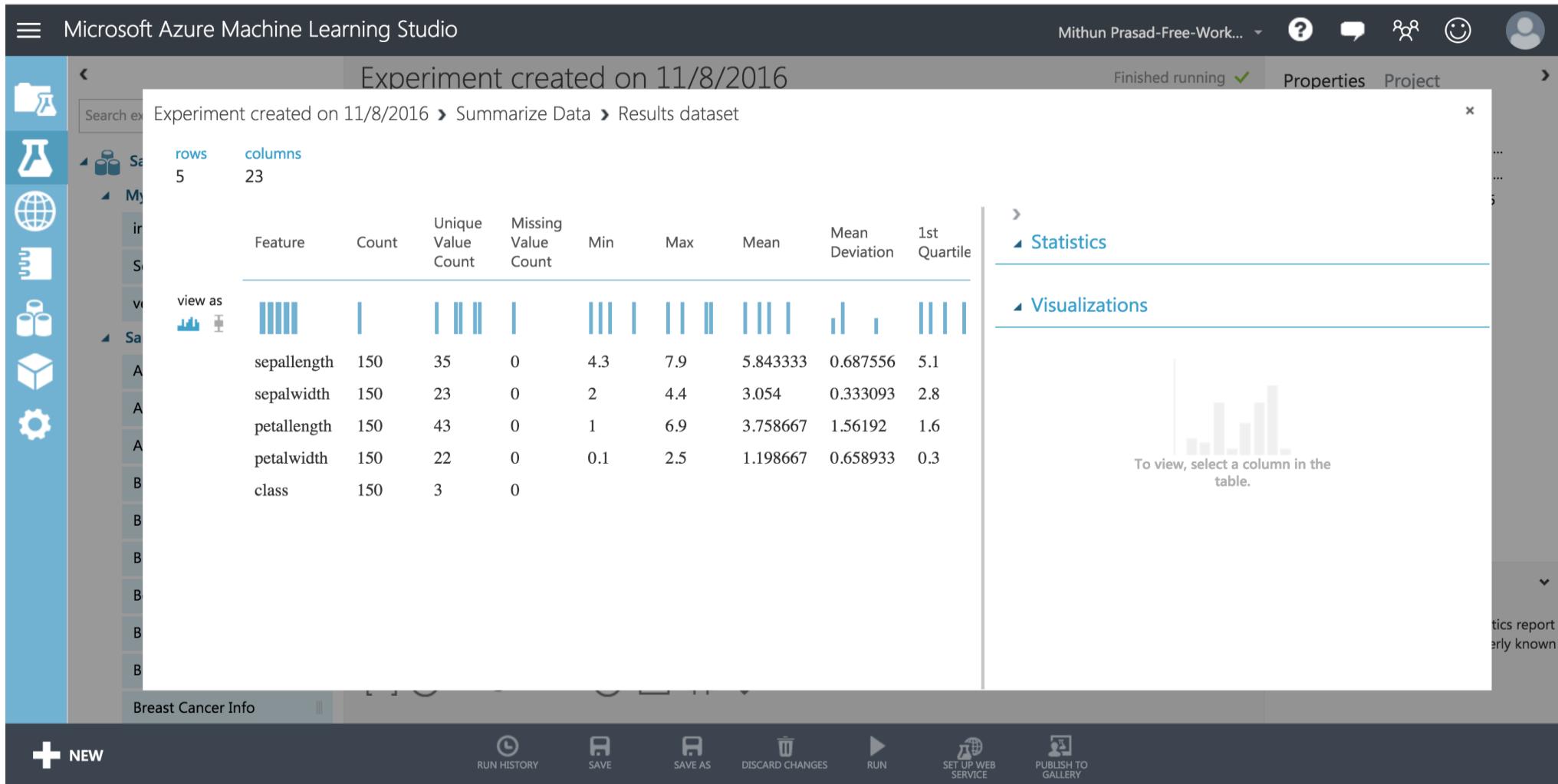
Statistics

Visualizations

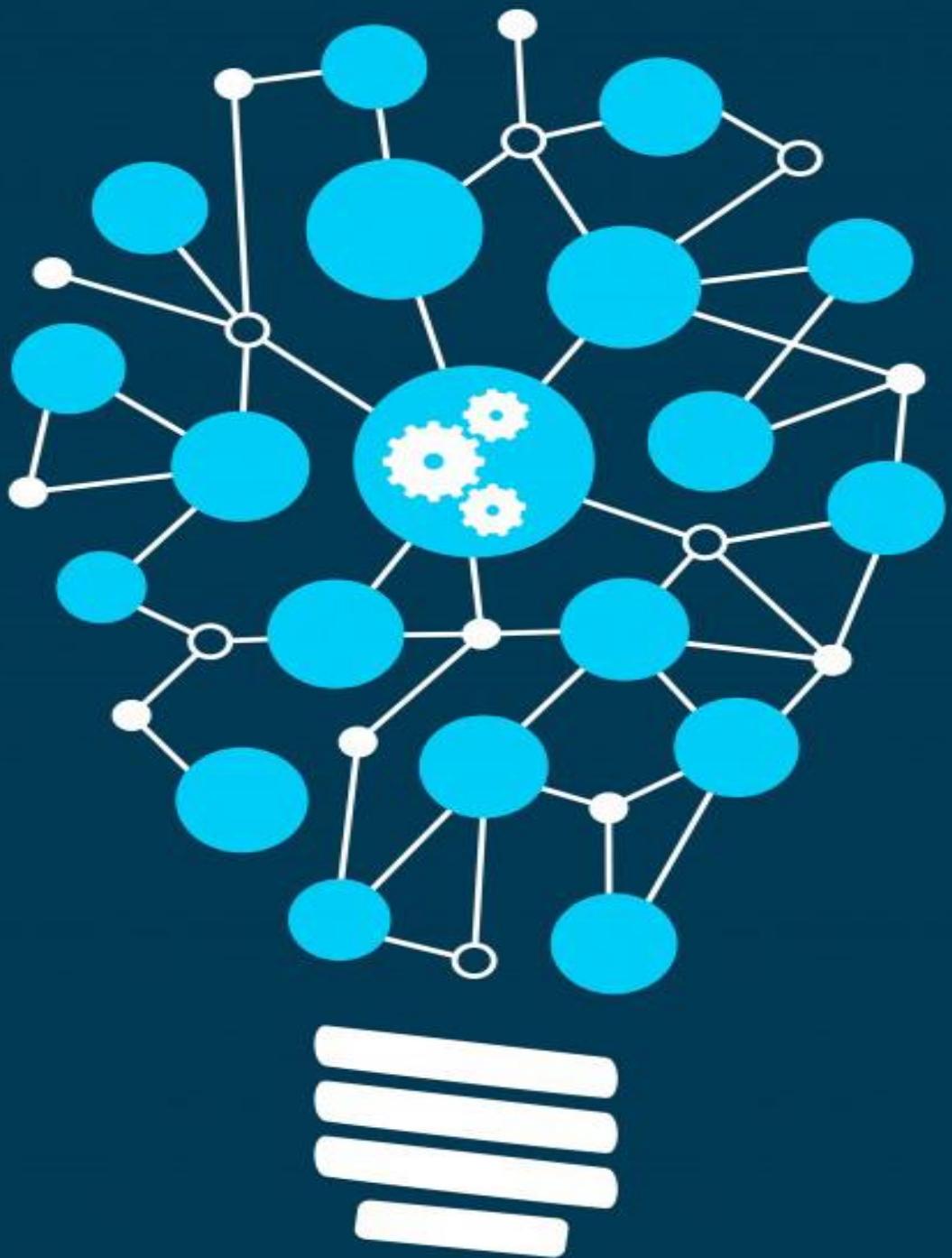
To view, select a column in the table.

Breast Cancer Info

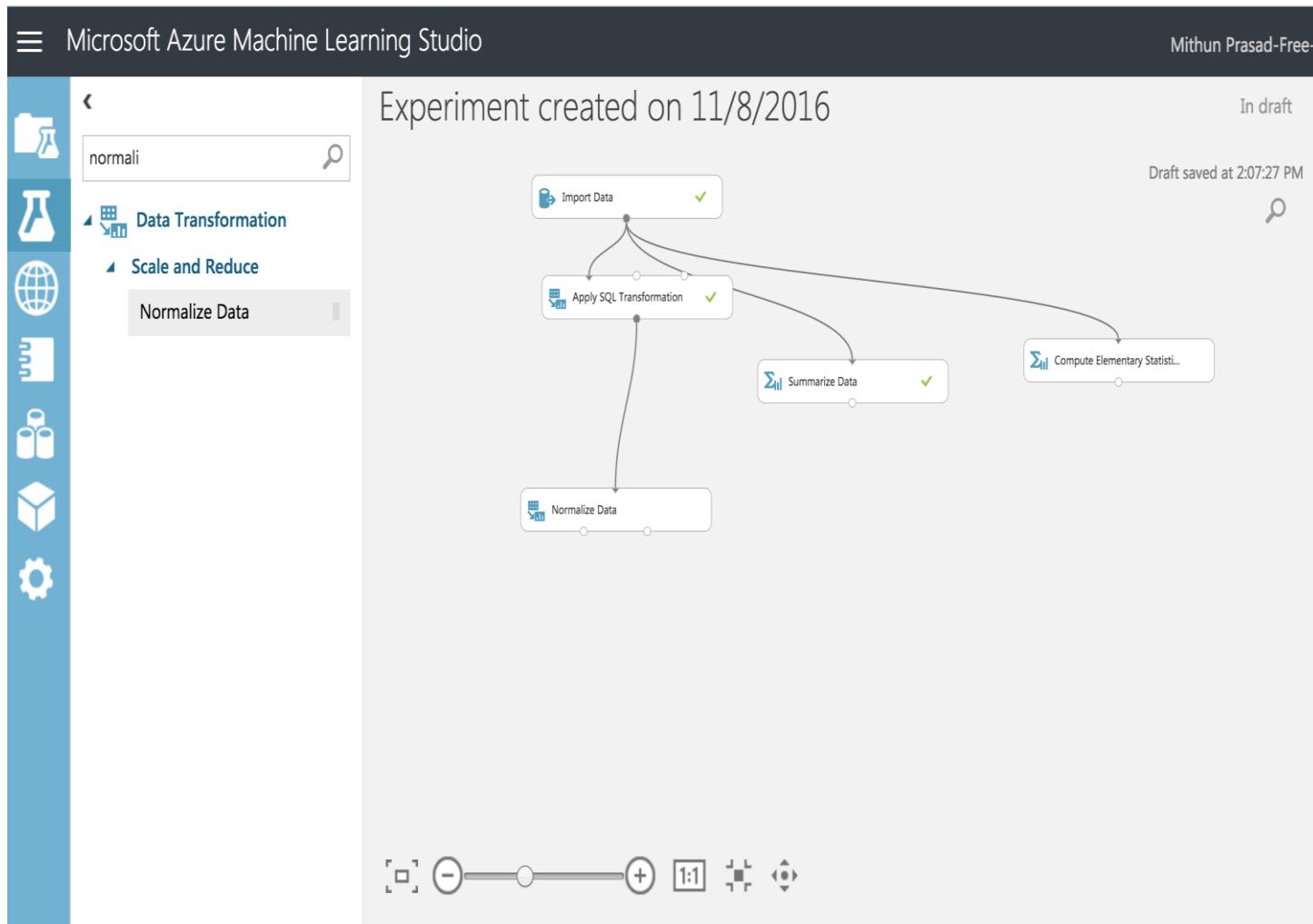
NEW RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY



# PREPROCESSING



# NORMALIZATION



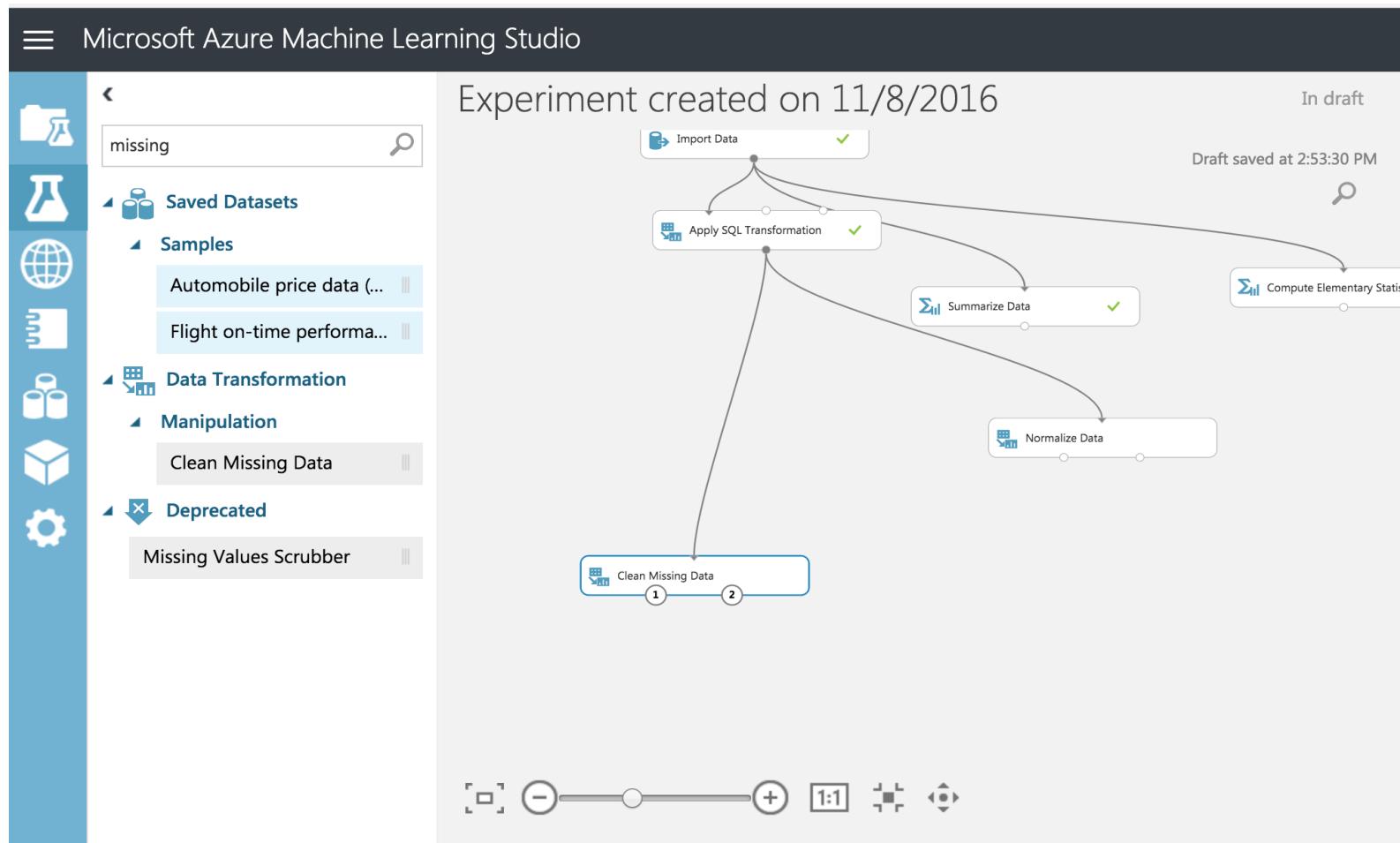
## Zscore

$$z = \frac{x - \text{mean}(x)}{\text{stdev}(x)}$$

## MinMax

$$z = \frac{x - \min(x)}{\left[ \max(x) - \min(x) \right]}$$

# MISSING DATA



## Cleaning Mode

- Replace using MICE
- ✓ Custom substitution value
- Replace with mean
- Replace with median
- Replace with mode
- Remove entire row
- Remove entire column
- Replace using Probabilistic PCA

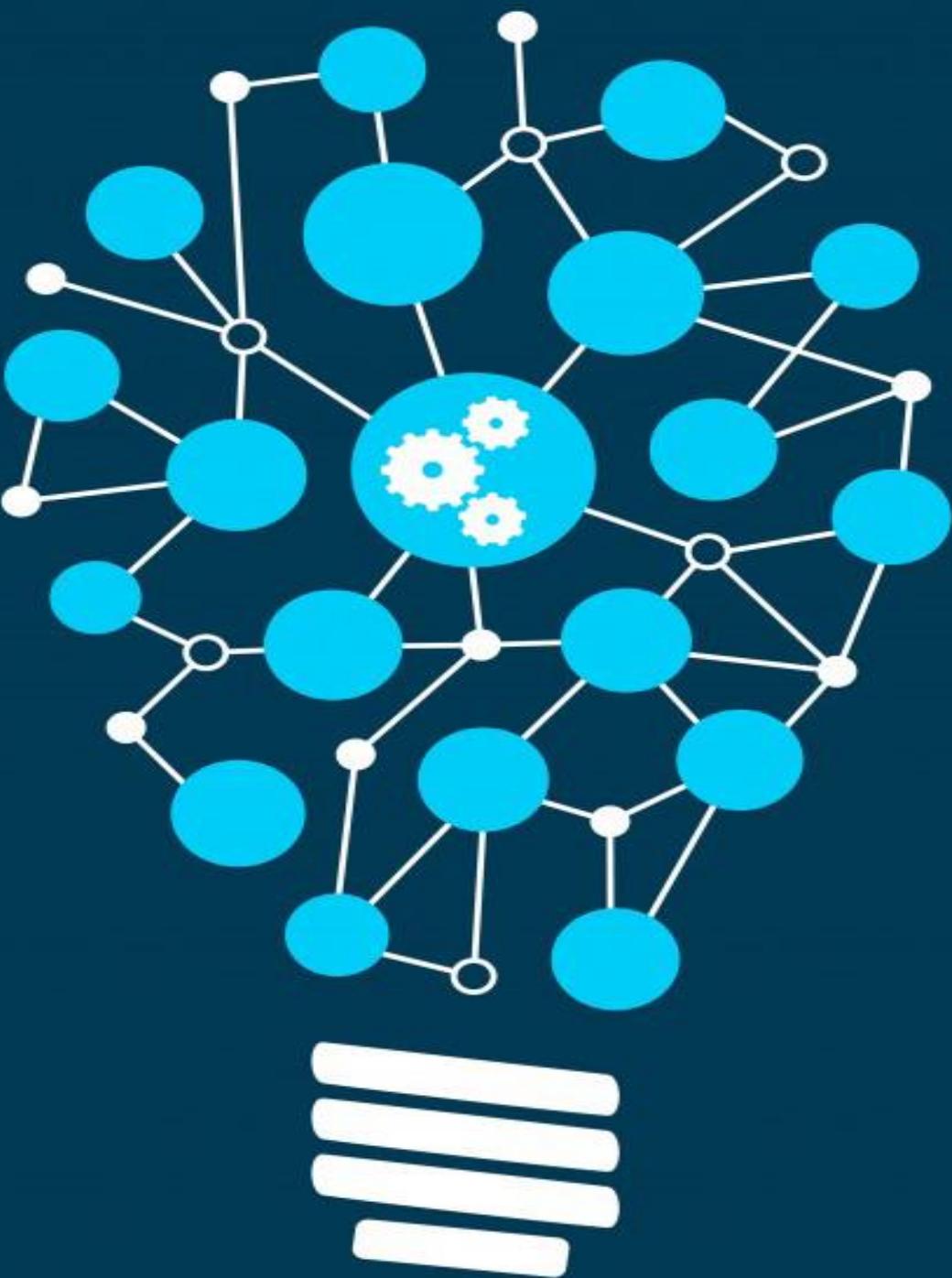
# FEATURE ENGINEERING

## Execute Python Script

Python script

```
1 # The script MUST contain a function named azureml_main
2 # which is the entry point for this module.
3
4 # imports up here can be used to
5 import pandas as pd
6 import numpy as np
7
8 # The entry point function can contain up to two input arguments:
9 #     Param<dataframe1>: a pandas.DataFrame
10 #    Param<dataframe2>: a pandas.DataFrame
11 def azureml_main(dataframe1 = None, dataframe2 = None):
12
13     # Execution logic goes here
14     # print('Input pandas.DataFrame #1:\r\n\r\n{0}'.format(dataframe1))
15
16     output_dataframe = dataframe1
17     output_dataframe['random'] = np.random.choice(range(1, 6), output_dataframe.shape[0])
18
19     return output_dataframe
```

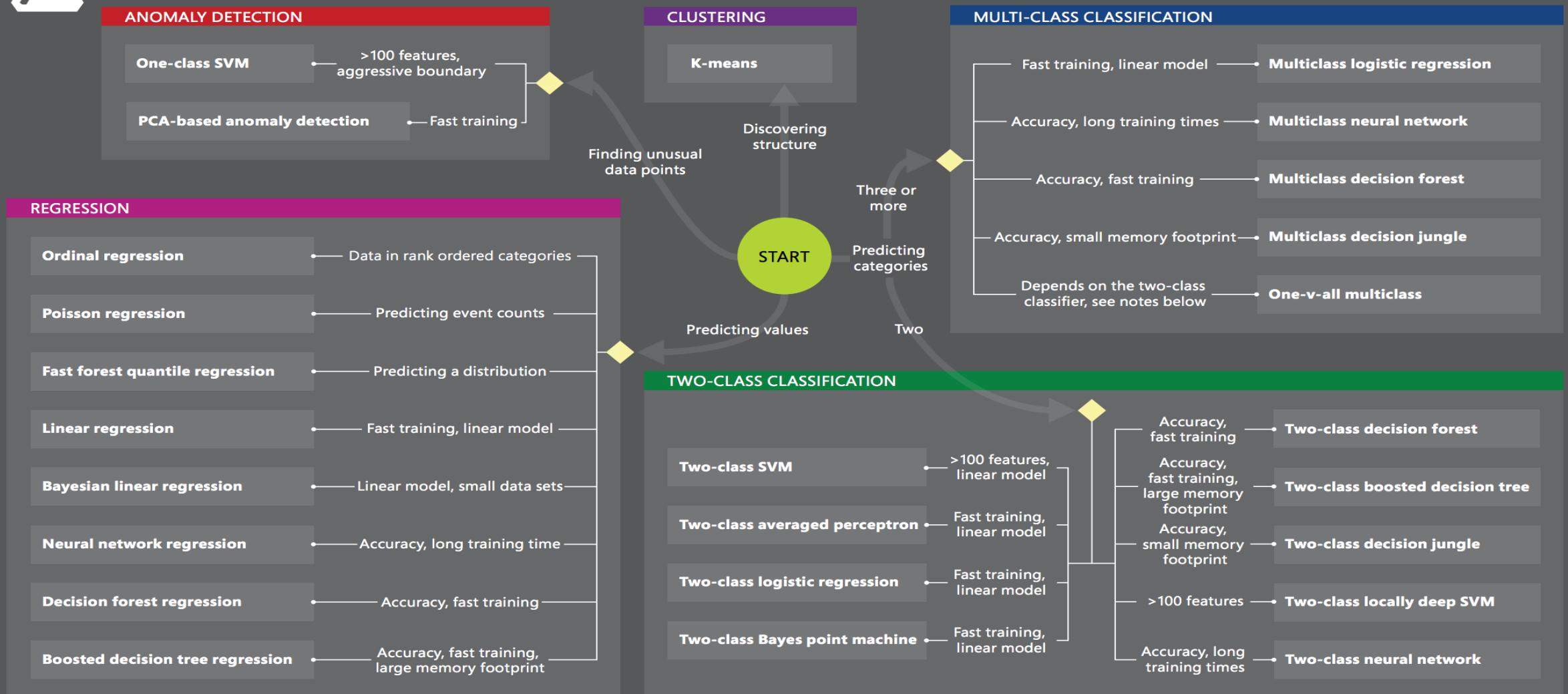
# ALGORITHMS





# Microsoft Azure Machine Learning: Algorithm Cheat Sheet

This cheat sheet helps you choose the best Azure Machine Learning Studio algorithm for your predictive analytics solution. Your decision is driven by both the nature of your data and the question you're trying to answer.



# CLUSTERING

Grouping items based on defined Features

- ▲  Machine Learning
- ▲ Initialize Model
- ▲ Clustering

K-Means Clustering



# CLASSIFICATION

Predicting the class or category for a single instance of data

## Initialize Model

## 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

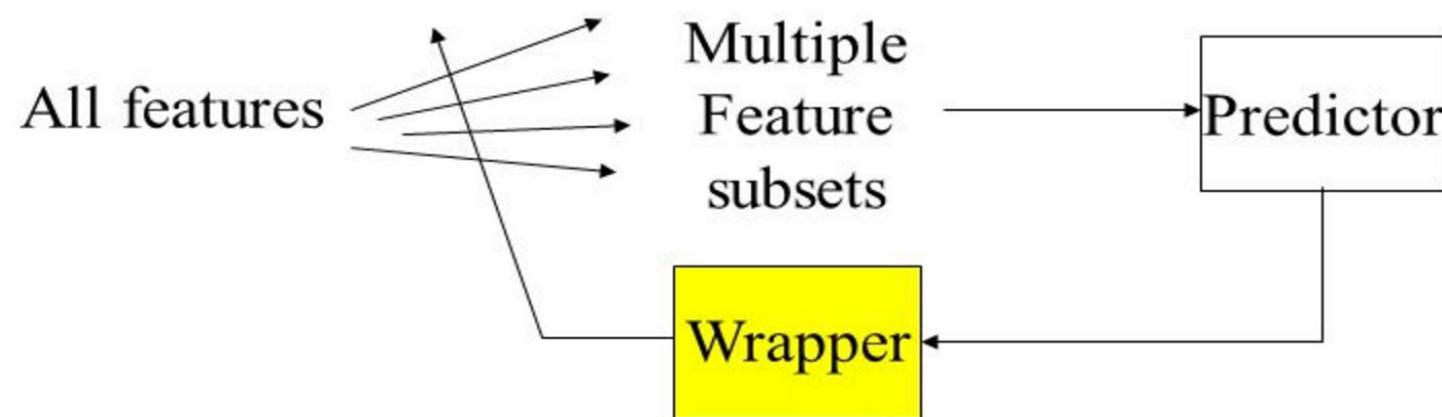
Two-Class Locally-Deep Support Vector Machine

Two-Class Logistic Regression

Two-Class Neural Network

Two-Class Support Vector Machine

# FEATURE SELECTION (FILTER)



## Filter Algorithms

- ✓ Pearson Correlation
- Mutual Information
- Kendall Correlation
- Spearman Correlation
- Chi Squared
- Fisher Score
- Count Based

# ANOMALY DETECTION

Selecting items based on unusual or suspicious patterns

- ▶  Machine Learning

- ▶ Initialize Model

- ▶ Anomaly Detection

- One-Class Support Vector Machine

- PCA-Based Anomaly Detection

# REGRESSION

Predicting the value of a datum given its history

- Initialize Model

- Classification

- Multiclass Logistic Regression

- Two-Class Logistic Regression

- Regression

- Bayesian Linear Regression

- Boosted Decision Tree Regression

- Decision Forest Regression

- Fast Forest Quantile Regression

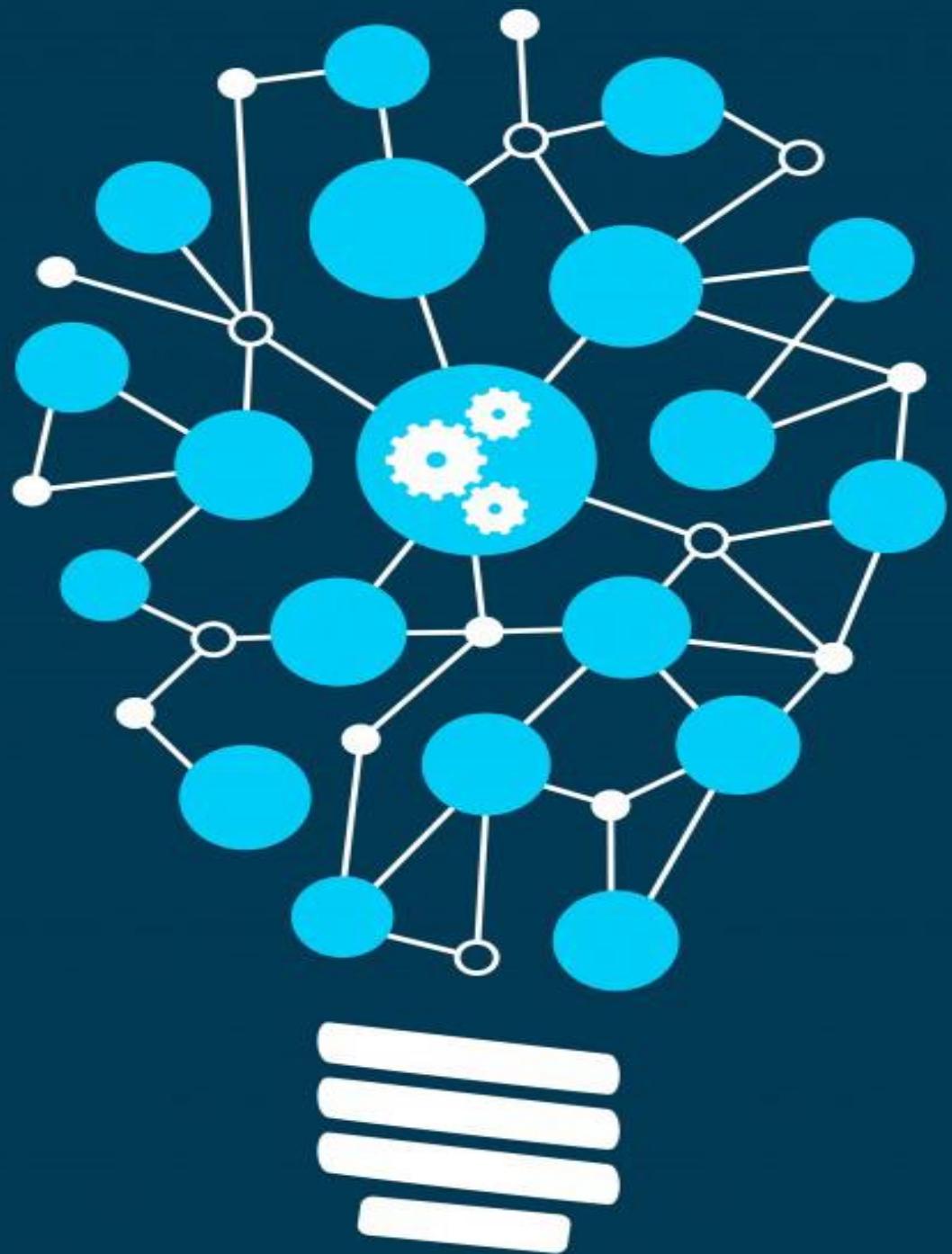
- Linear Regression

- Neural Network Regression

- Ordinal Regression

- Poisson Regression

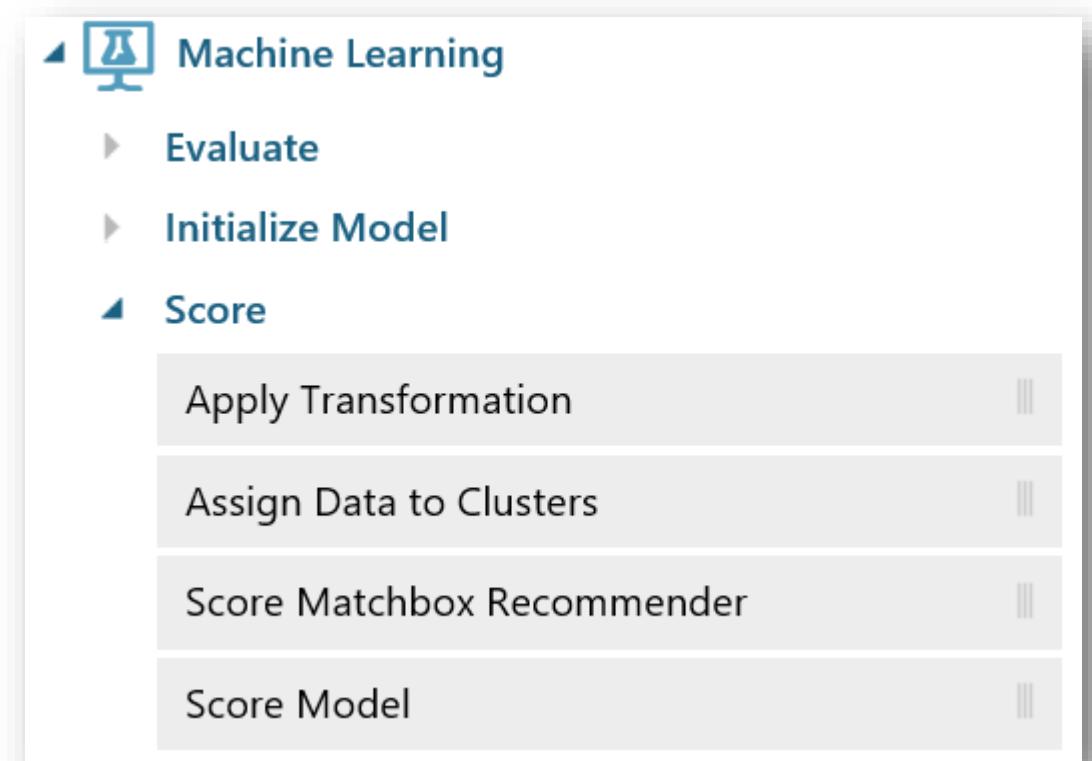
# MODEL SCORING AND EVALUATION



# SCORING A MODEL

Apply a trained model to:

- A list of recommended items
- Forecasts for time series models
- Estimates of projected demand, volume, or other numeric quantity, for regression models
- Cluster assignments
- A predicted class or outcome, for classification models
- Probability scores associated with these outputs



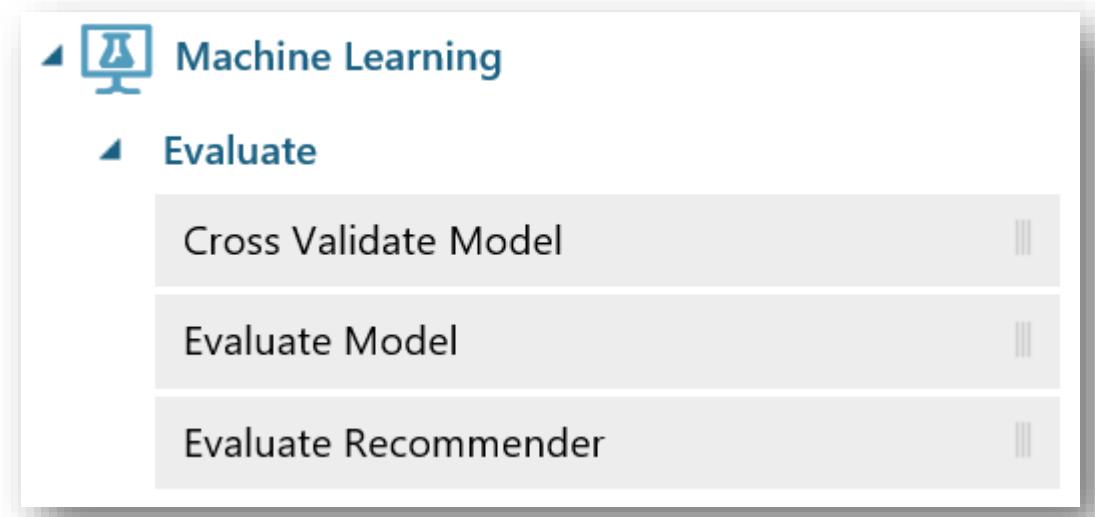
# EVALUATING A MODEL

## Metrics for Classification Models

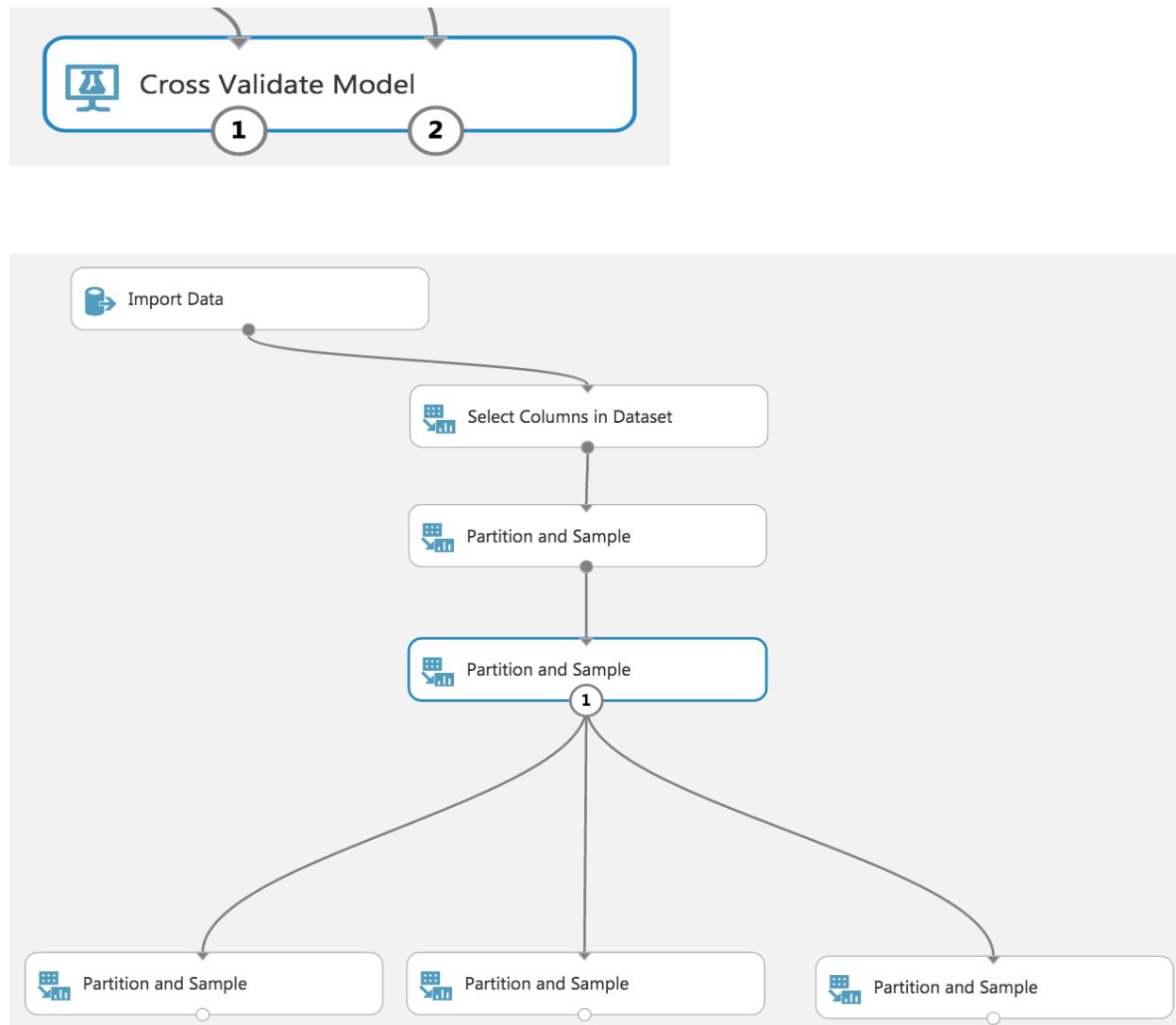
- Accuracy, Recall, Precision, F1-Score
- AUC
- Average Log Loss
- Training Log Loss

## Metrics for Regression Models

- Mean absolute error (MAE)
- Root mean squared error (RMSE)
- Relative absolute error (RAE)
- Relative squared error (RSE)
- Coefficient of determination



# CROSS VALIDATION



## Partition and Sample

Partition or sample mode

Assign to Folds

Use replacement in th...

Randomized split

Random seed

0

Specify the partitioner method

Partition evenly

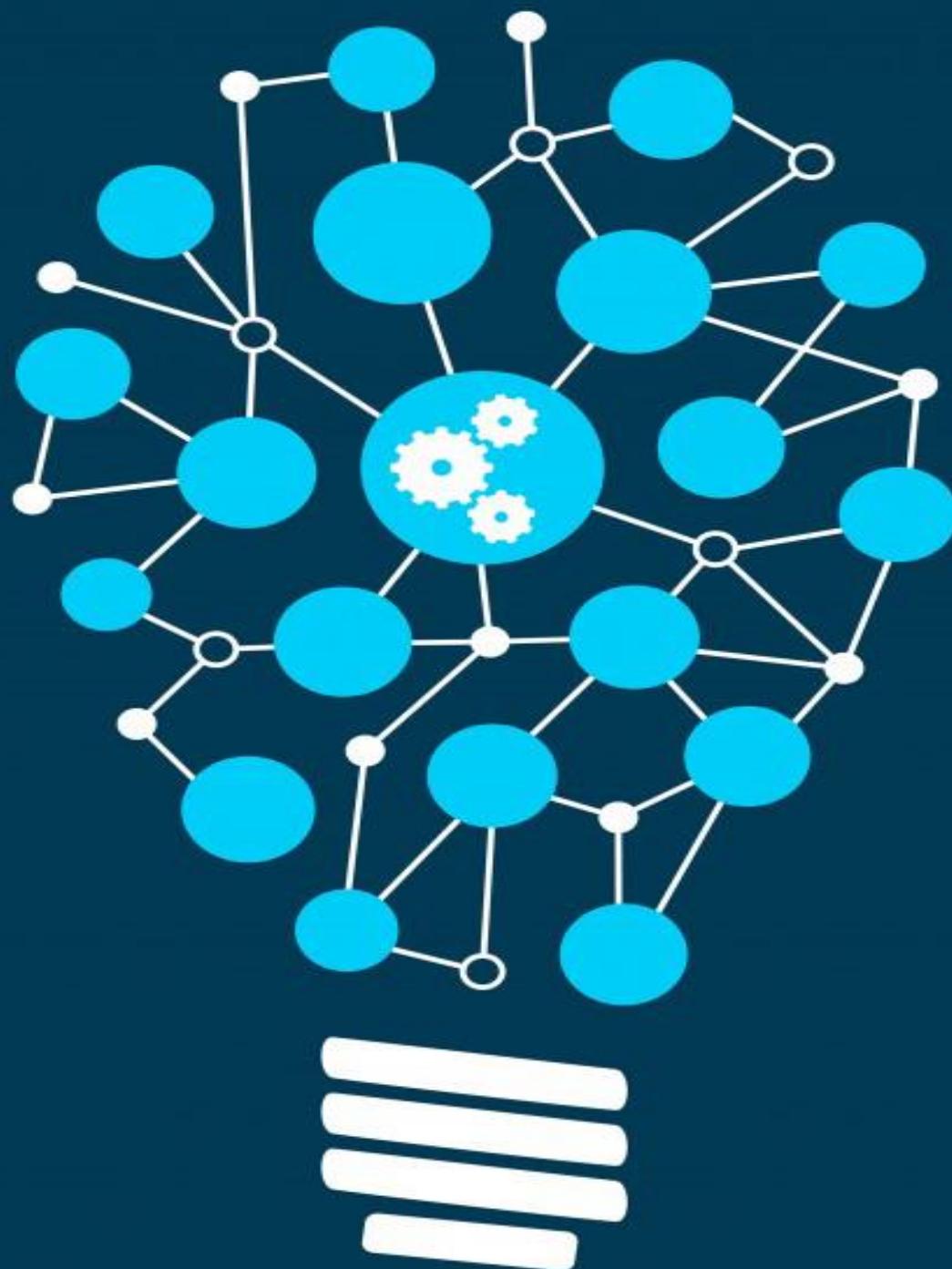
Specify number of folds to...

3

Stratified split

False

# CUSTOMIZATION



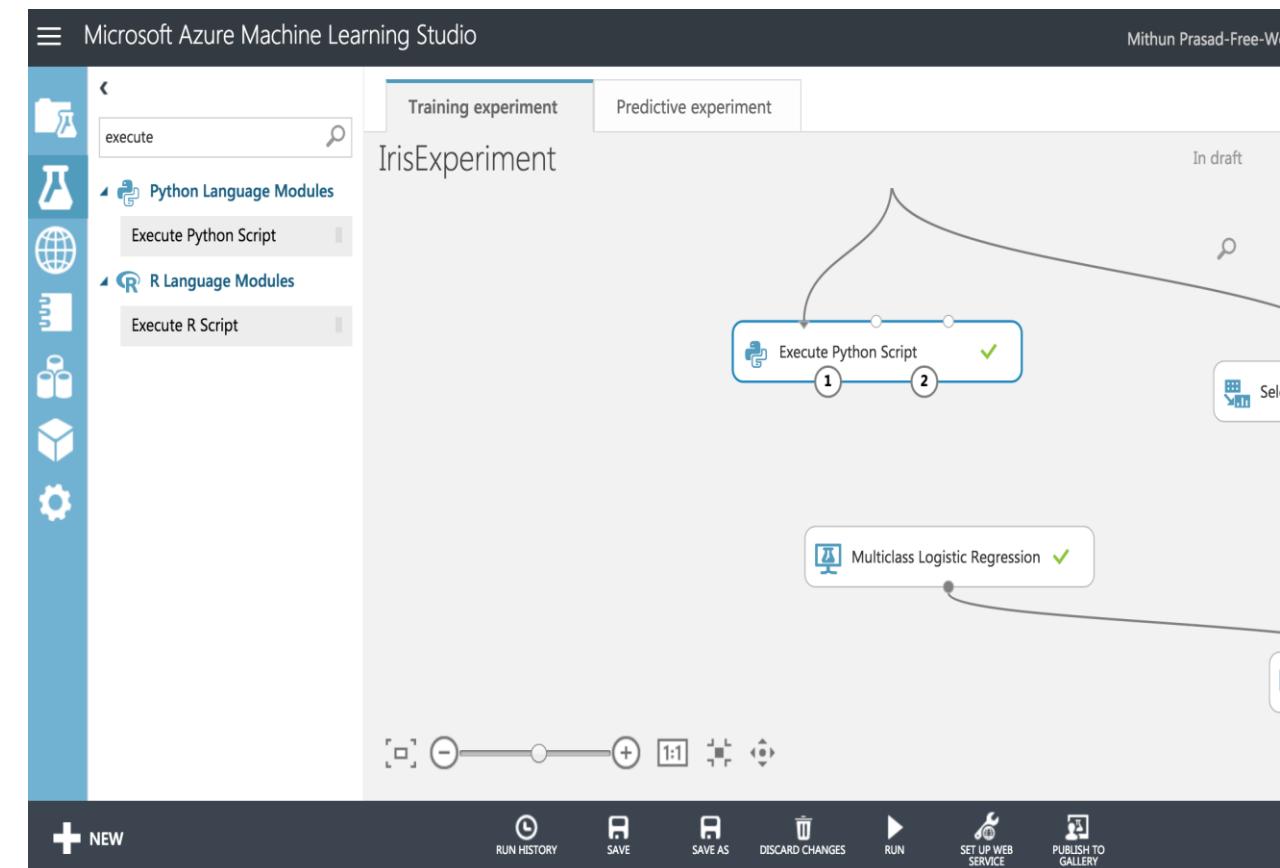
# HOW TO USE EXECUTE PYTHON SCRIPT

1. Add the **Execute Python Script** module to your experiment.
2. Connect any datasets that you want to use for input. You can also provide a zipped file containing custom resources.

**Dataset1.** An optional dataset from your Machine Learning Studio workspace, containing input data or values.

**Dataset2.** A second dataset, also optional.

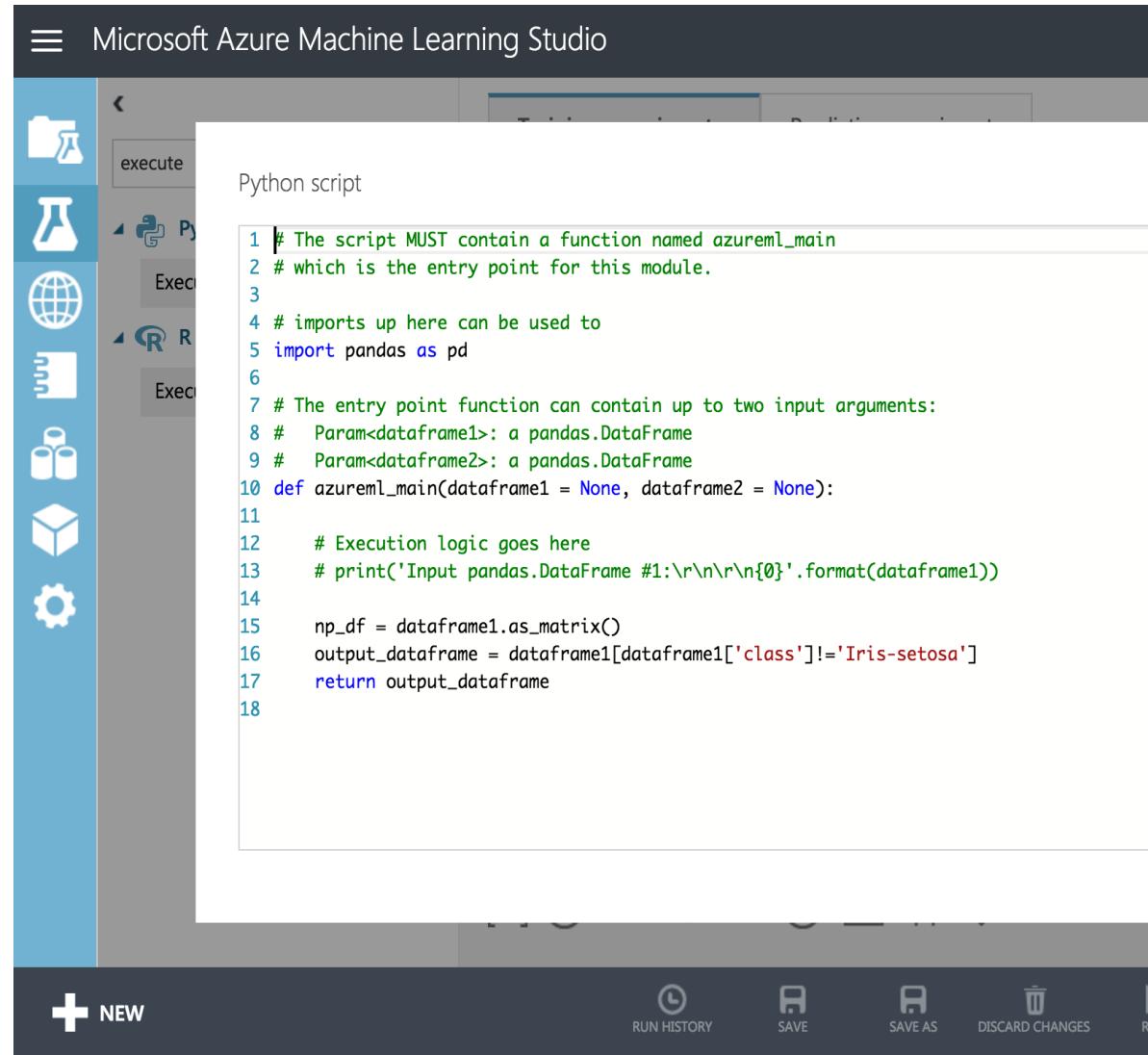
**Script bundle.** A zipped file containing custom resources.



# CUSTOM TASKS

Integrate Ipython notebooks with Azure Machine Learning to perform custom tasks:

- Visualization
- Use Python client libraries to enumerate datasets and models in your workspace
- Read, load, and manipulate data

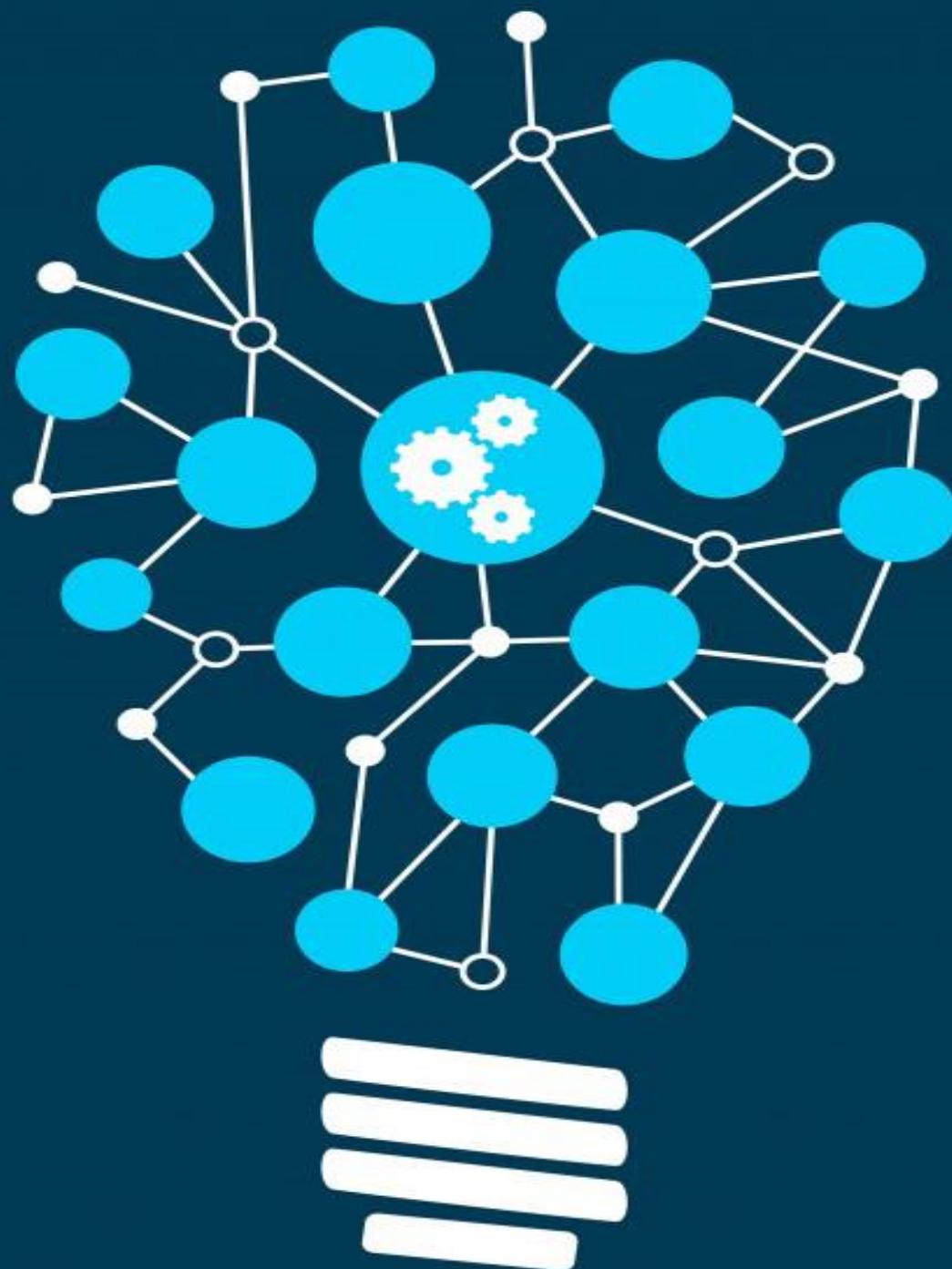


The screenshot shows the Microsoft Azure Machine Learning Studio interface. On the left, there is a vertical toolbar with icons for file operations, execute, Python, R, and other tasks. The main area is titled "Python script" and contains the following code:

```
1 # The script MUST contain a function named azureml_main
2 # which is the entry point for this module.
3
4 # imports up here can be used to
5 import pandas as pd
6
7 # The entry point function can contain up to two input arguments:
8 #     Param: a pandas.DataFrame
9 #     Param: a pandas.DataFrame
10 def azureml_main(dataframe1 = None, dataframe2 = None):
11
12     # Execution logic goes here
13     # print('Input pandas.DataFrame #1:\r\n\r\n{0}'.format(dataframe1))
14
15     np_df = dataframe1.as_matrix()
16     output_dataframe = dataframe1[dataframe1['class']!='Iris-setosa']
17     return output_dataframe
18
```

At the bottom of the studio window, there are buttons for "NEW", "RUN HISTORY", "SAVE", "SAVE AS", and "DISCARD CHANGES".

# DEPLOY AS A WEB SERVICE



# WEB SERVICE

## 1. Deploy the experiment as a web service

Microsoft Azure Machine Learning Studio

Mithun Prasad-Free-Work... ? ☰ ☺

lab1 [predictive exp.]

DASHBOARD CONFIGURATION

General New Web Services Experience [preview](#)

Published experiment

[View snapshot](#) [View latest](#)

Description

No description provided for this web service.

API key

TToiOovPXbyecWV1ISPLu9UUMKgBQqBaRVl3e8zOG7qVkJR6d59ozDjjdDSB9wrLe0qVs+jnGjLXEyXhfYQ==

Default Endpoint

API HELP PAGE TEST APPS LAST UPDATED

REQUEST/RESPONSE	Test <a href="#">Test preview</a>	Excel 2013 or later   Excel 2010 or earlier workbook	10/28/2016 10:21:15 AM
BATCH EXECUTION	<a href="#">Test preview</a>	Excel 2013 or later workbook	10/28/2016 10:21:15 AM

Enter data to predict

SEPALLLENGTH

5.1

SEPALWIDTH

3.5

PETALLLENGTH

1.4

PETALWIDTH

0.2

CLASS



# BATCH CLASSIFICATION

≡ Microsoft Azure Machine Learning Web Services

Quickstart    Dashboard    Batch Request Log    Configure    Consume    **Test**    Swagger API

← Lab1 [Predictive Exp.]

## default

[View in Studio](#)

Request-Response    **Batch**

✓ input1

irisDataset.csv    [Browse...](#)

Storage account    irisstorage3

✓ Test Batch Jobs

JOB ID	STATUS	RUN START	RUN END	DURATION	Result
db2c4ac1fea24d0e80cde9a83f817e75	Finished	10/28/2016 10:37 AM	10/28/2016 10:37 AM	2s	<a href="#">output1</a>

1 / 1

Note: We will enable CORS on your storage account to upload this file

[Test](#)