

1. Main page: <http://cortanaanalytics.com>
2. To begin this module, you should have:
 1. Basic Math and Stats skills
 2. Business and Domain Awareness
 3. General Computing Background

NOTE: These workbooks contain many resources to lead you through the course, and provide a rich set of references that you can use to learn much more about these topics. If the links do not resolve properly, type the link address in manually in your web browser. If the links have changed or been removed, simply enter the title of the link in a web search engine to find the new location or a corollary reference.

1. Understand Azure ML and how experiments are created
2. Understand how MRS can be used to perform Machine Learning experiments
3. Use ADF to schedule Azure ML Activities



1. At the end of this Module, you will:
 1. Understand Azure ML and how experiments are created
 2. Understand how MRS can be used to perform Machine Learning experiments
 3. Use ADF to schedule Azure ML Activities

The Data Science Process and Platform



The Team Data Science Process

Business Understanding

- Define Objectives
- Identify Data Sources

Data Acquisition and Understanding

- Ingest Data
- Explore Data
- Update Data

Modeling

- Feature Selection
- Create and Train Model

Deployment

- Operationalize












Customer Acceptance

- Testing and Validation
- Handoff
- Re-train and re-score

1. This process largely follows the CRISP-DM model:
<http://www.sv-europe.com/crisp-dm-methodology/>
2. It also references the Cortana Intelligence process:
<https://azure.microsoft.com/en-us/documentation/articles/data-science-process-overview/>
3. A complete process diagram is here:
<https://azure.microsoft.com/en-us/documentation/learning-paths/cortana-analytics-process/>
4. Some walkthrough's of the various services:
<https://azure.microsoft.com/en-us/documentation/articles/data-science-process-walkthroughs/>
5. An integrated process and toolset allows for a more close-to-intent deployment
6. Iterations are required to close in on the solution – but are

harder tio management and monitor

The Cortana Intelligence Platform

	Cortana, Cognitive Services, Bot Framework
	Power BI
	Stream Analytics
	HDInsight
	Azure Machine Learning (MRS)
	SQL Data Warehouse (SQL DB, Document DB)
	Data Lake
	Event Hubs
	Data Factory
	Data Catalog
	Microsoft Azure

1. Platform and Storage: Microsoft Azure – <http://microsoftazure.com> Storage: <https://azure.microsoft.com/en-us/documentation/services/storage/> (Host It)
2. Azure Data Catalog: <http://azure.microsoft.com/en-us/services/data-catalog> (Doc It)
3. Azure Data Factory: <http://azure.microsoft.com/en-us/services/data-factory/> (Move It)
4. Azure Event Hubs: <http://azure.microsoft.com/en-us/services/event-hubs/> (Bring It)
5. Azure Data Lake: <http://azure.microsoft.com/en-us/campaigns/data-lake/> (Store It)
6. Azure DocumentDB: <https://azure.microsoft.com/en-us/services/documentdb/> , Azure SQL Data Warehouse: <http://azure.microsoft.com/en-us/services/sql-data-warehouse/> (Relate It)
7. Azure Machine Learning: <http://azure.microsoft.com/en-us/services/machine-learning/> (Learn It)
8. Azure HDInsight: <http://azure.microsoft.com/en-us/services/hdinsight/> (Scale It)
9. Azure Stream Analytics: <http://azure.microsoft.com/en-us/services/stream-analytics/> (Stream It)
10. Power BI: <https://powerbi.microsoft.com/> (See It)
11. Cortana: <http://blogs.windows.com/buildingapps/2014/09/23/cortana-integration-and-speech-recognition-new-code-samples/> and <https://blogs.windows.com/buildingapps/2015/08/25/using-cortana-to-interact-with-your-customers-10-by-10/> and <https://developer.microsoft.com/en-us/Cortana> (Say It)
12. Cognitive Services: <https://www.microsoft.com/cognitive-services>
13. Bot Framework: <https://dev.botframework.com/>
14. All of the components within the suite: <https://www.microsoft.com/en-us/server-cloud/cortana-intelligence-suite/what-is-cortana-intelligence.aspx>
15. What can I do with it? <https://gallery.cortanaintelligence.com/>

16. Getting Started Quickly: <https://caqs.azure.net/#gallery>

Module 1: Azure ML



6

1. Example paths for using Azure ML:
<https://azure.microsoft.com/en-us/documentation/articles/machine-learning-data-science-plan-sample-scenarios/>

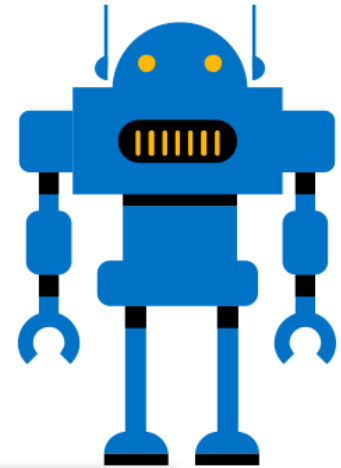
Machine Learning in 5 Minutes

The Formal one:

"A computer program is said to learn from experience **E** with respect to some class of tasks **T** and performance measure **P** if its performance at tasks in **T**, as measured by **P**, improves with experience **E**."

A Practical Example:

Look at data. Do the thing. Better? No? Look at the data. Do something different. Better? Yes? Do that again. (Repeat)



1. Choosing an Algorithm for Machine Learning:
<https://azure.microsoft.com/en-us/documentation/articles/machine-learning-algorithm-choice/>

Machine Learning Capabilities

Which category
(Classification)



How much/many
(Regression)



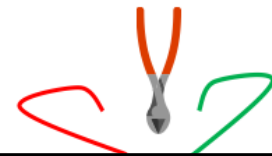
Which group
(Clustering,
Recommender)



Is it odd
(Anomaly)



Which action
(Reinforcement
Learning)



1. Regression: Predict a real value for each item (stock/currency value, temperature). – How much/how many?
2. Classification: Assign a category to each item (Chinese | French | Indian | Italian | Japanese restaurant). – Which Category?
3. Clustering/Recommendation: Partition items into homogeneous groups (clustering twitter posts by topic). – Which Groups?
4. Anomaly: Identify when something unexpected happens. – Is this weird?
5. Reinforcement Learning: Make an appropriate action for some new data. – Which action?

Machine Learning Algorithms

Split into two main categories:

- Supervised learning
 - Predicting the future
 - Learn from known past examples to predict future
 - Labels provided
- Unsupervised learning
 - Making sense of data
 - Understanding the past
 - Learning the structure of data
 - Labels no provided



1. Algorithm Documentation:
<https://msdn.microsoft.com/library/dn905974.aspx>
2. Exploring:
<https://azuremlsimples.azurewebsites.net/simples/>

The Azure ML Environment

Development Environment

- Creating Experiments
- Sharing a Workspace



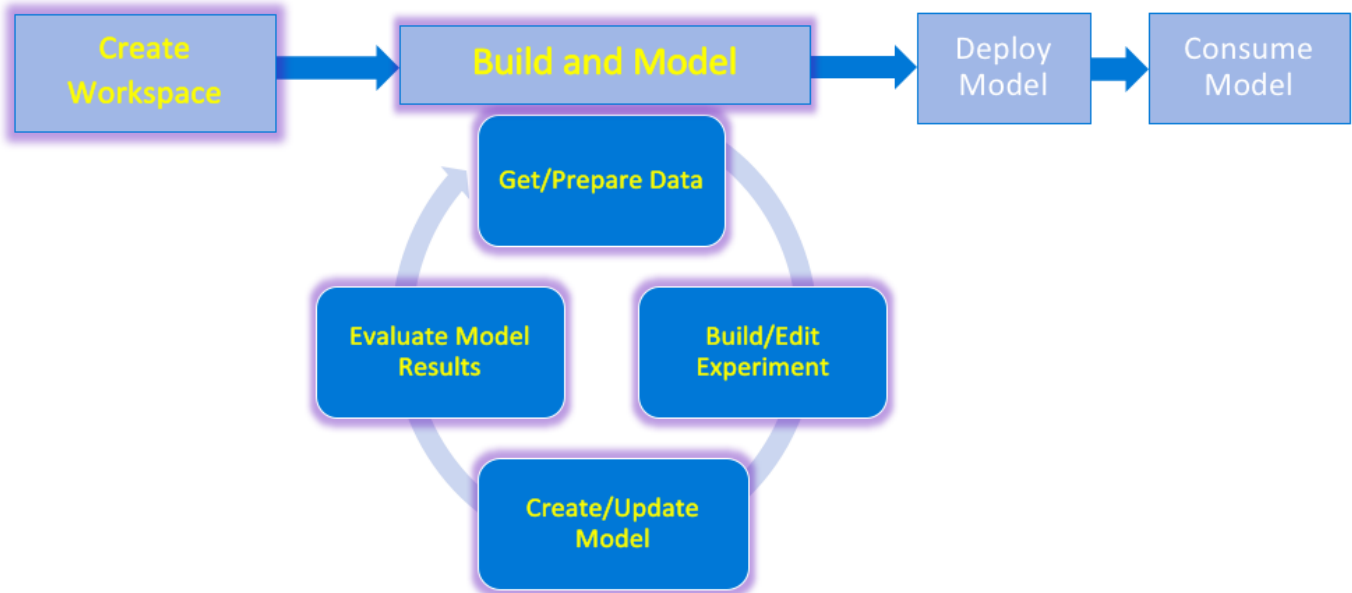
Deployment Environment

- Publishing the Model
- Using the API
- Consuming in various tools



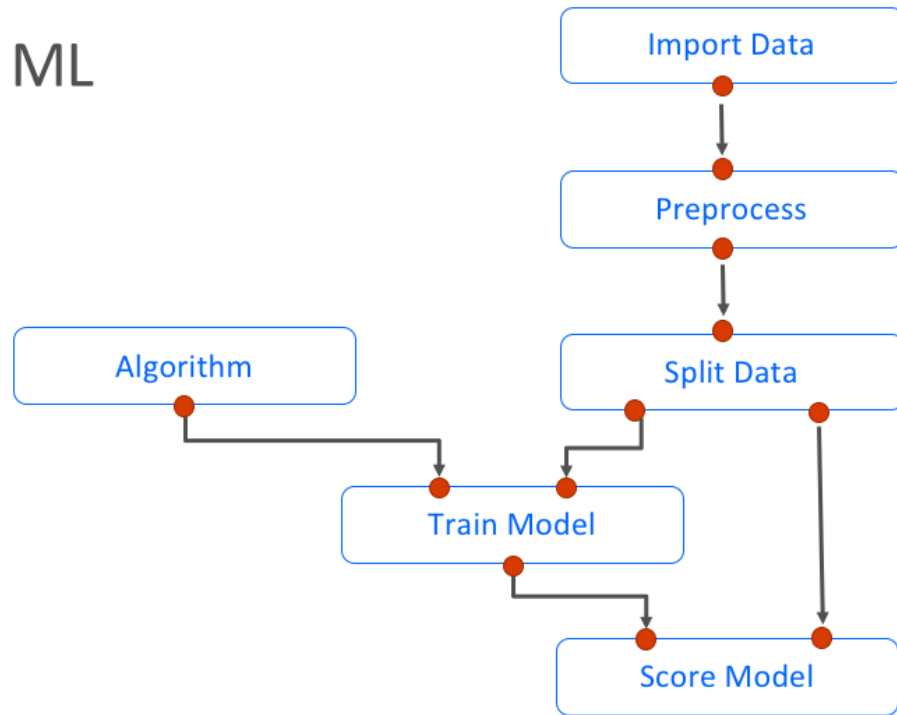
1. Guided tutorials: <https://azure.microsoft.com/en-us/documentation/services/machine-learning/>
2. Microsoft Azure Virtual Academy course: https://mva.microsoft.com/en-US/training-courses/microsoft-azure-machine-learning-jump-start-8425?l=ehQZFoKz_7904984382

Creating an Experiment



1. Beginning Series: <https://azure.microsoft.com/en-us/documentation/articles/machine-learning-data-science-for-beginners-the-5-questions-data-science-answers/>

Basic Azure ML Elements



1. Designing an experiment in the Studio:

<https://azure.microsoft.com/en-us/documentation/articles/machine-learning-what-is-ml-studio/>



1. Open the **AML Student Workbook** from your \Resources folder
2. Follow the instructions you find there

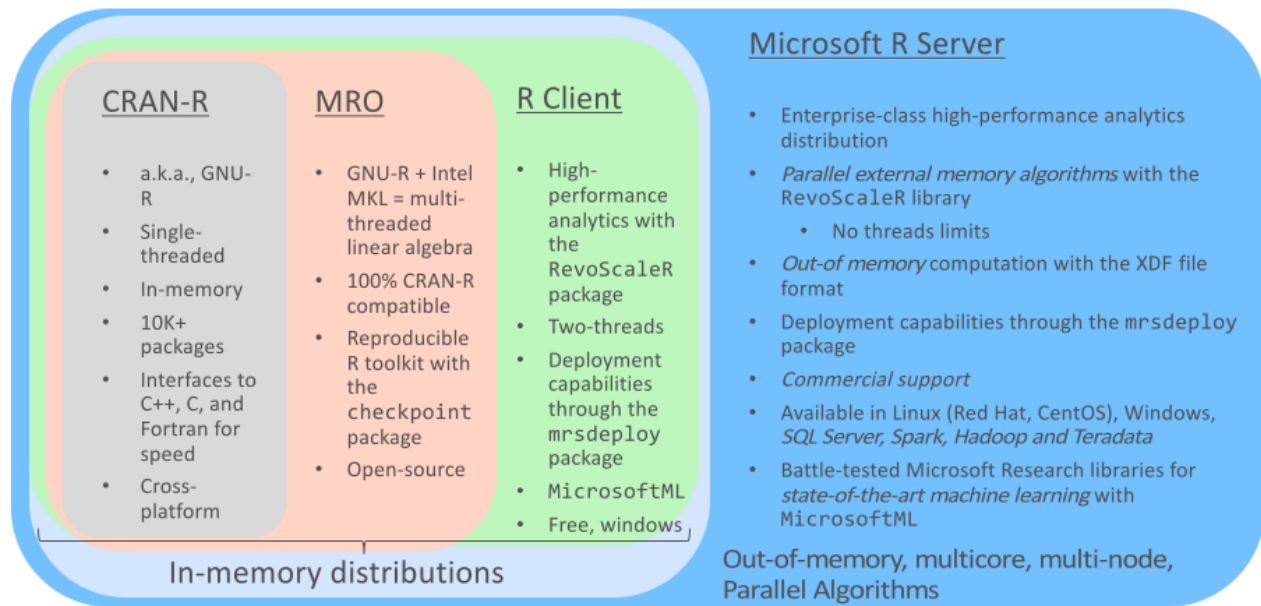
Module 2: Microsoft R Server for Machine Learning



14

1. Primary documentation: <https://www.microsoft.com/en-us/server-cloud/products/r-server/>

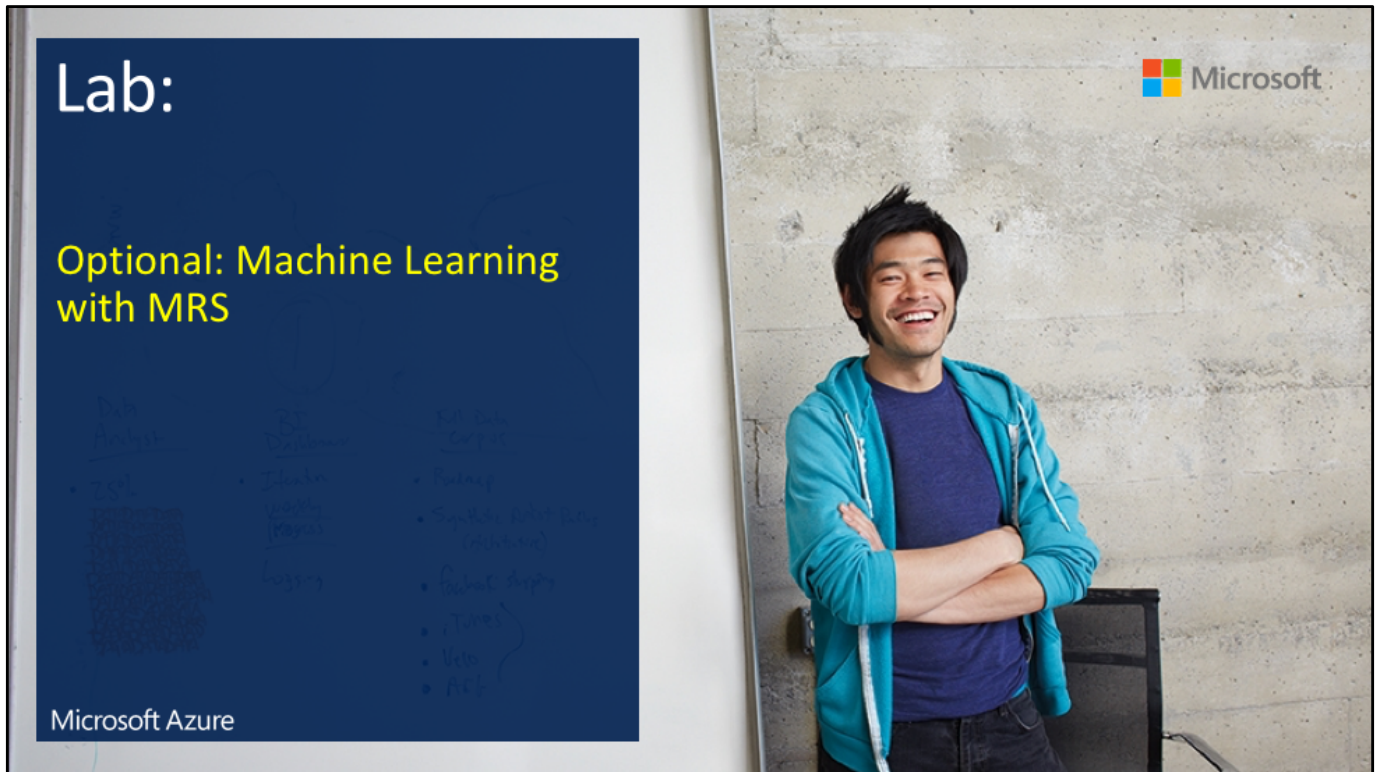
Distributions of R



Parallelized, Distributed Execution Algorithms

Data Step	Statistical Tests	Variable Selection
Data import – Delimited, Fixed, SAS, SPSS, ODBC	Chi Square Test	Stepwise Regression
Variable creation & transformation	Kendall Rank Correlation	
Recode variables	Fisher's Exact Test	
Factor variables	Student's t-Test	
Missing value handling		
Sort, Merge, Split		
Aggregate by category (means, sums)		
Descriptive Statistics	Sampling	Simulation
Min / Max, Mean, Median (approx.)	Subsample (observations & variables)	Simulation (e.g. Monte Carlo)
Quantiles (approx.)	Random Sampling	Parallel Random Number Generation
Standard Deviation		
Variance		
Correlation		
Covariance		
Sum of Squares (cross product matrix for set variables)		
Pairwise Cross tabs		
Risk Ratio & Odds Ratio		
Cross-Tabulation of Data (standard tables & long form)		
Marginal Summaries of Cross Tabulations		
	Predictive Models	Cluster Analysis
	Sum of Squares (cross product matrix for set variables)	K-Means
	Quantiles (approx.)	
	Generalized Linear Models (GLM) exponential family distributions: binomial, Gaussian, inverse Gaussian, Poisson, Tweedie. Standard link functions: cauchy, identity, log, logit, probit. User defined distributions & link functions.	
	Covariance & Correlation Matrices	
	Logistic Regression	
	Classification & Regression Trees	
	Predictions/scoring for models	
	Residuals for all models	
		Classification
		Decision Trees
		Decision Forests
		Gradient Boosted Decision Trees
		Naive Bayes
		Combination
		rxDataStep
		rxExec
		PEMA-R API Custom Algorithms





1. Open the **MRS Student Workbook** document from your \Resources file
2. Locate the section marked “**Predictive Modeling with MRS**” and follow the instructions there

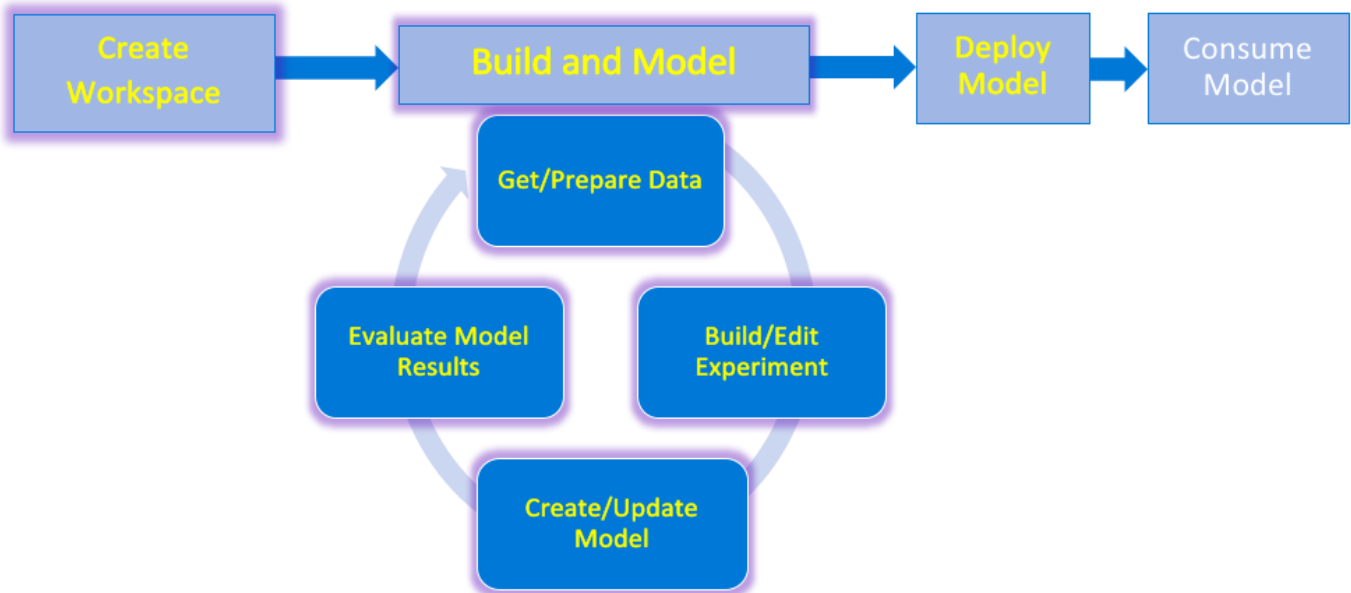
Module 3: Azure Data Factory and Azure ML



18

1. Create Predictive Pipelines using Azure ML Activities in ADF:
<https://azure.microsoft.com/en-us/documentation/articles/data-factory-azure-ml-batch-execution-activity/>

Deploying the Model



1. Deploying the Azure ML Model:

<https://azure.microsoft.com/en-us/documentation/articles/machine-learning-walkthrough-5-publish-web-service/>



1. Understand Azure ML and how experiments are created
2. Understand how MRS can be used to perform Machine Learning experiments
3. Use ADF to schedule Azure ML Activities

© 2018 Microsoft Corporation. All rights reserved.

Questions?