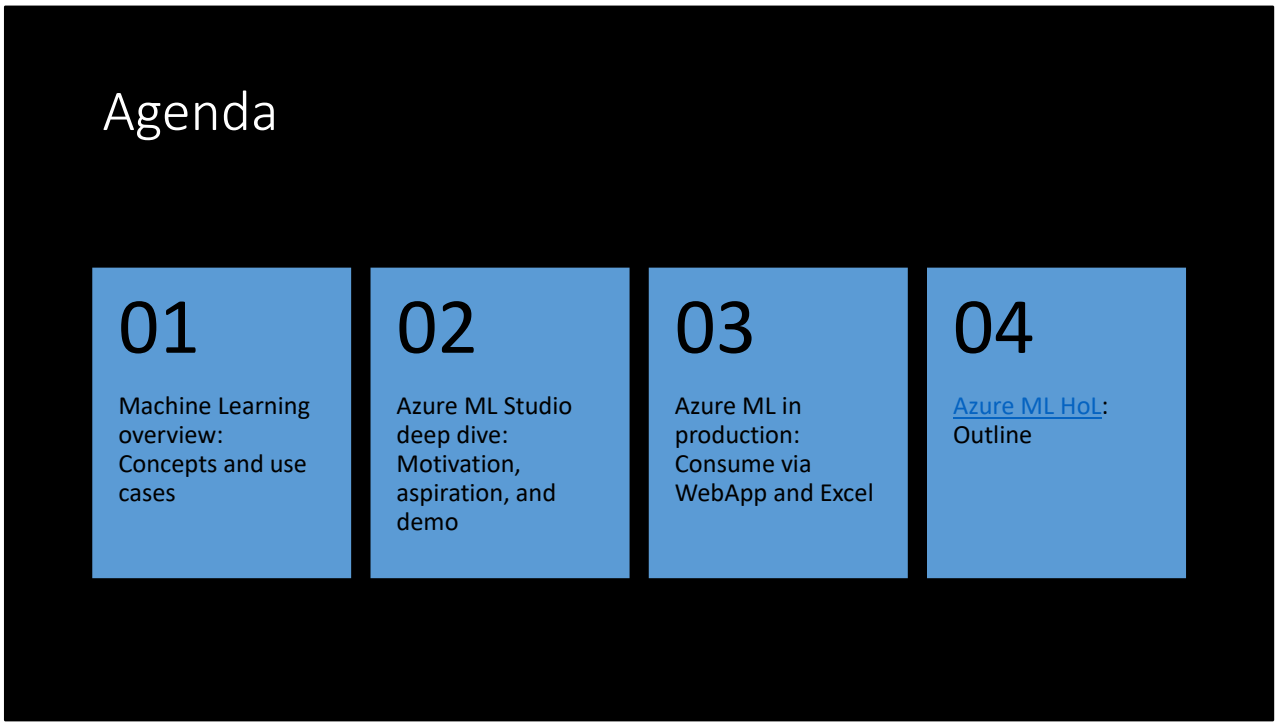
Microsoft Azure ML
Machine Learning for Everyone

Kay Apperson, PhD | kay.apperson@microsoft.com

Solution Architect | Resident Data Scientist
Microsoft St. Louis MO

Microsoft

Kay is an award-winning data scientist with a comprehensive experience in computer engineering, computer science, and statistics. She was a researcher and an Instructor in Medicine at Harvard Medical School before transferring her skills to work in the industry. Her noteworthy ML accomplishment is stabilizing $1.2 billion in cash flow for a Fortune 200 company. Some of the non-Microsoft tools she has expertise in include SAS, SPSS/IBM Clementine, open R, Weka, Apache Tomcat jsp, Oracle, MySQL, and Linux. At Microsoft, she specializes in Azure ML, Microsoft R, Cognitive Services APIs, Power BI, SQL Server, Azure Data Lake Store and Analytics (USQL), and Spark ML DStreams.



Azure ML and ML overview. Highlighting the values that Azure ML could bring to IT and business, where Azure ML fits in the entire BI stack, and ML solutions and use cases. This serves as a transition from the Cortana webinar to this one.
Azure ML Studio deep dive. Walking through the full machine learning process end-to-end using ML Studio. We'll demonstrate how easy it is to build a savvy ML model with no coding involved. We'll cover from data to ML model building, testing, and evaluation.
Azure ML in production. Realizing the values of Azure ML in business decision making when a predictive model is deployed. We'll end the session with what you could do more to operationalize ML.
Azure ML HoL. https://github.com/kayapperson/AzureMLHoL-CERN. Talk about the lab outline and prerequisites.

What is Machine Learning ?

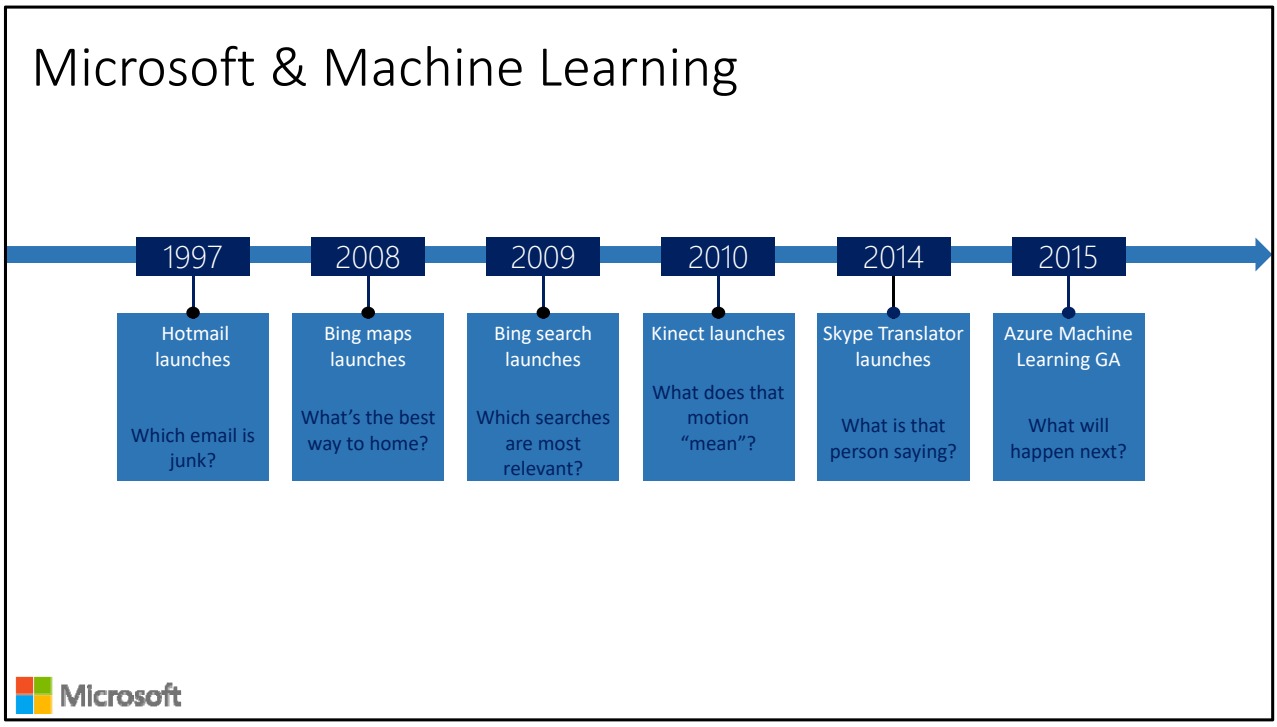
Using **known data**, develop a **model** to predict **unknown data**.

Microsoft

Basic definition:

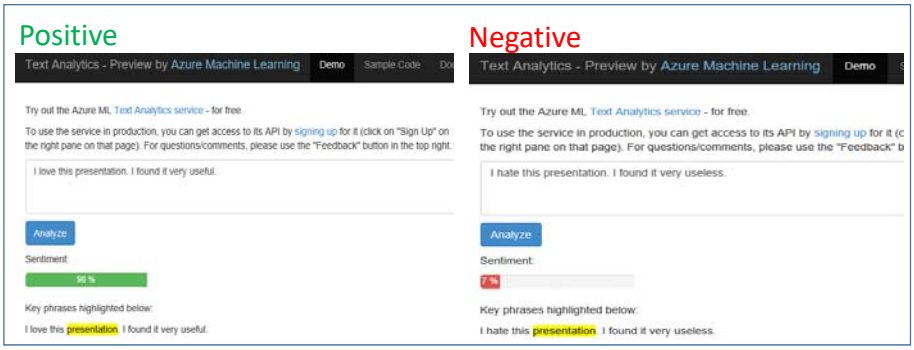
Machine learning develops algorithms for making **predictions (statistical sense)** from data *

Learning models from available training data, to make good predictions on unseen test data



More than a decade, Microsoft is using ML in its services.

Why Machine Learning?

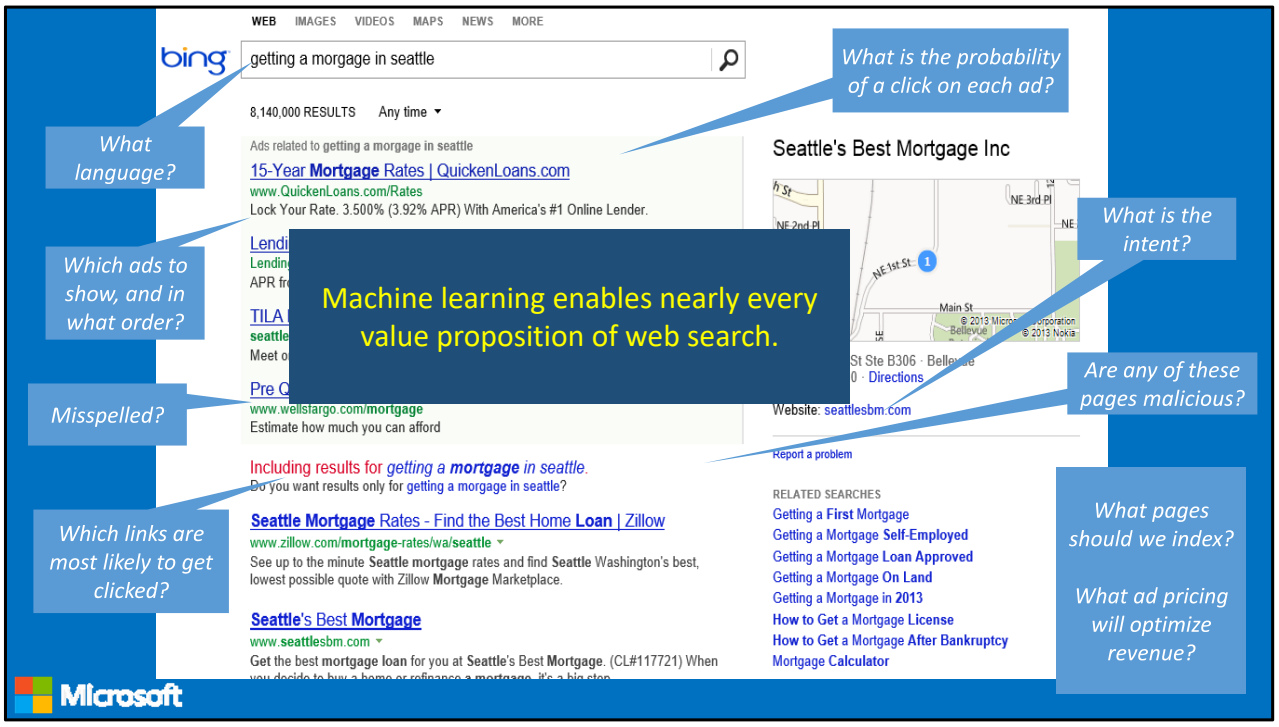


Microsoft

Smart applications, services, platforms…

Digital Smart Personal Assistant

Image / Text / Data analyze



The first "killer commercial application" of ML on Big Data was web search. With ML, we can use all of the billions of queries and associated clicks that we collect each month as training data for an ML system. That way, the system can learn, for each query, which web pages are likely the most relevant.

But the impact of ML goes beyond page relevance. With billions of queries, we can learn common misspellings and the native language of the user. We can discern the intent of the user – shopping? research? entertainment? ML can learn what kinds of pages are malicious.

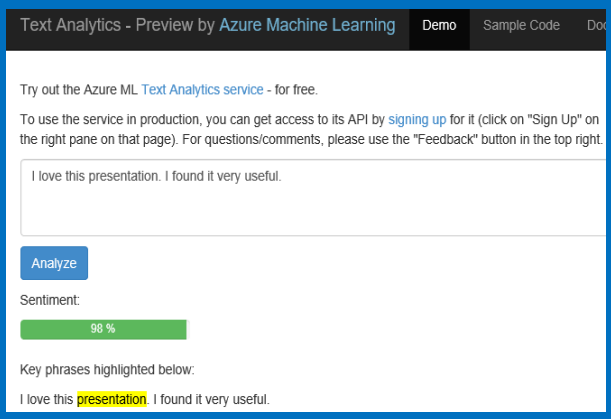
ML also is able to learn, for each query, what kinds of ads are most likely to be clicked, and also learn how best to show the ads.

And even before we get to a real user, ML can be used to figure out what parts of the web are worth scanning, and what pricing for ads will optimize revenues.

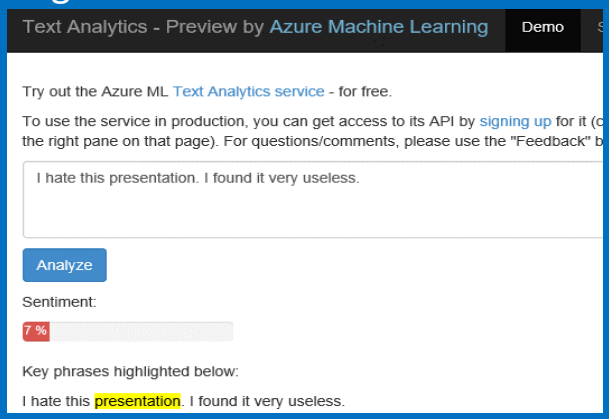
© 2012 Microsoft Corporation. All rights reserved. Microsoft, Windows, and other product names are or may be registered trademarks and/or trademarks in the U.S. and/or other countries.
The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation. MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.

Text Analytics: User reviews

Positive



Negative



Microsoft

Ready to use ML API published on Azure market place. A showcase of production ML solution in action (This can be used i.e. in shopping websites to see user reviews about a product etc.)

Test by yourself some of the reviews comments on Expedia or other online shopping sites. Copy/paste user comments to see the analyze result. If sentiment result shows higher percentage number, it means it is positive. If the percentage is low than it is a negative comment.

https://text-analytics-demo.azurewebsites.net/

Analyze Image

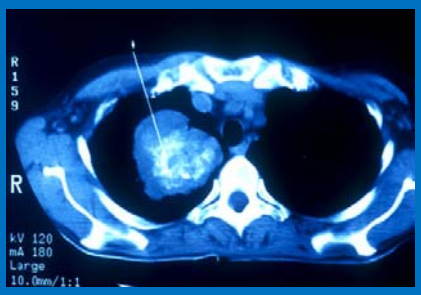
Features:

Feature Name	Value
Image Format	Jpeg
Image Dimensions	1500 x 1155
Clip Art Type	0 Non-clipart
Line Drawing Type	0 Non-LineDrawing
Black & White Image	False
Is Adult Content	False
Adult Score	0.02942577563226223
Is Racy Content	False
Racy Score	0.018232977017760277
Categories	[{ "name": "plant_flower", "score": 0.99609375 }]
Faces	[]
Dominant Color Background	
Dominant Color Foreground	
Dominant Colors	
Accent Color	#C8A403

Microsoft

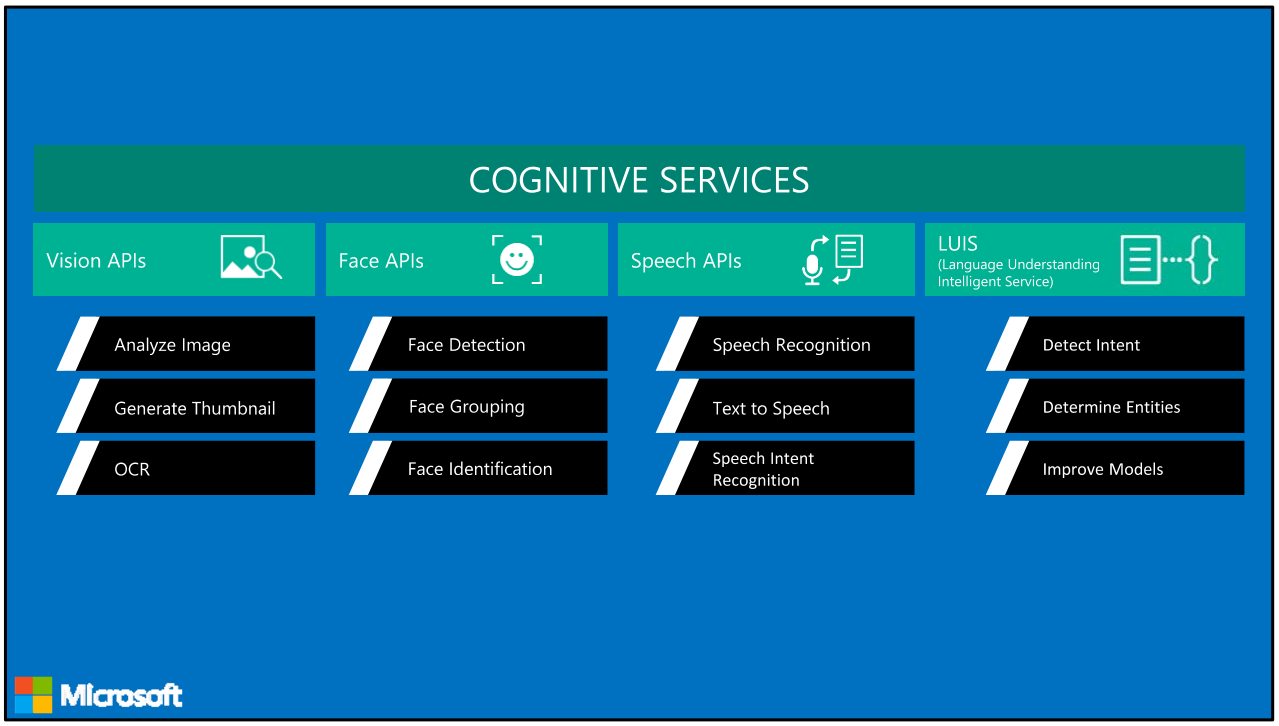
Go to https://www.projectoxford.ai/demo/visions#Analysis and analyze this image to find its accent color

Data is Hybrid



Microsoft

Data is hybrid, and many organizations have data silos. Organizing data and harvesting from it are equally essential.



These Microsoft Cognitive Services provide models that are sophisticated and high performance. We wouldn't advise that you build these models from the ground up, e.g. use Python CNTK or deep NNs. One reason is the amount and quality of the training data that are required to achieve this quality. Other reasons are the GPU computation and tuning model parameters that would be computationally expensive.

JJ Food Service Food Delivery Service: Shopping carts that already list what you're going to order?
Pier 1 Imports : predict which products customers will want in the future
Direct One : Non-sensitive data analysis, using Azure HDInsight for big data analysis and Azure Machine Learning for predictive analytics
Gjirafa: Search Engine for specific language that is not crawled well by major search engines.
JLL: Real Estate search Engine
Carnegie Mellon: more energy efficient buildings
eSmart Systems: automated energy management system
OSIsoft: collect, analyze, and share real-time sensor data to build their mission-critical operations intelligence infrastructure.
Mendeley: providing a faster and more flexible way to share and archive scientific findings and ideas.
Youboox: like Netflix and Spotify, French e-book subscription service: its recommendation engine
WASH Multifamily Laundry: measure company performance and drive productivity.
Kellogg School of Management : Real-world analytics for students
Optolexia: dyslexia screening tool for young childrenv

Rockwell Automation : Data Analysis

Tacoma Public Schools: the district is providing comprehensive data snapshots of student success indicators and has already helped to improve graduation rates from 55 to 78 percent.

Ziosk: predict your preferences and serve them up as part of a better dining experience.

North American Eagle: aggregate and analyze data from prior and current runs to create predictive models

Sun Branding Solutions: packaging text and artwork effectively in ongoing omnichannel campaigns.

The Next Up: predictive analytics about customer traffic based on variables such as weather, historical data, holidays, and major events.

Ultra Tendency: predict when airplanes will arrive at their designated airport gates.

Genscape: transform thousands of streams of energy supply data into consumable intelligence that its customers depend on.

xample use cases

Finance and risk	Sales and marketing	Customer and channel	Operations and workforce
Revenue Forecasting	Sales forecasting	User segmentation	Agent allocation
Portfolio optimization	Demand forecasting	Personalized offers	Warehouse effic
Investment modelling	Sales lead scoring	Product recommendation	Smart buildings
Fraud detection	Marketing mix optimization		Predictive maint
Risk management			Supply chain optimization

+ notes From the book: AzureMachineLearning – AzureFundamentals

Many examples of predictive analytics can be found literally everywhere today in our society:
Spam/junk email filters These are based on the content, headers, origins, and even user behaviors (for example, always delete emails from this sender).
Mortgage applications Typically, your mortgage loan and credit worthiness is determined by advanced predictive analytic algorithm engines.
Various forms of pattern recognition These include optical character recognition (OCR) for routing your daily postal mail, speech recognition on your smart phone, and even facial recognition for advanced security systems.
Life insurance Examples include calculating mortality rates, life expectancy, premiums, and payouts.
Medical insurance Insurers attempt to determine future medical expenses based on historical medical claims and similar patient backgrounds.
Liability/property insurance Companies can analyze coverage risks for automobile and home owners based on demographics.
Credit card fraud detection This process is based on usage and activity patterns. In the past year, the number of credit card transactions has topped 1 billion. The popularity of contactless payments via near-field communications (NFC) has also increased dramatically over the past year due to smart phone integration.
Airline flights Airlines calculate fees, schedules, and revenues based on prior air travel patterns and flight data.
Web search page results Predictive analytics help determine which ads, recommendations, and display sequences to render on the page.
Predictive maintenance This is used with almost everything we can monitor: planes, trains, elevators, cars, and yes, even data centers.
Health care Predictive analytics are in widespread use to help determine patient outcomes and future care based on historical data and pattern matching across similar patient data sets.

What is Machine Learning ?

Using **known data**, develop a **model** to predict **unknown data**.

Known Data: Big enough archive, previous observations, past data

Unknown Data: Unseen, not existing, future data

Model: Known data + Algorithms (ML algorithms)

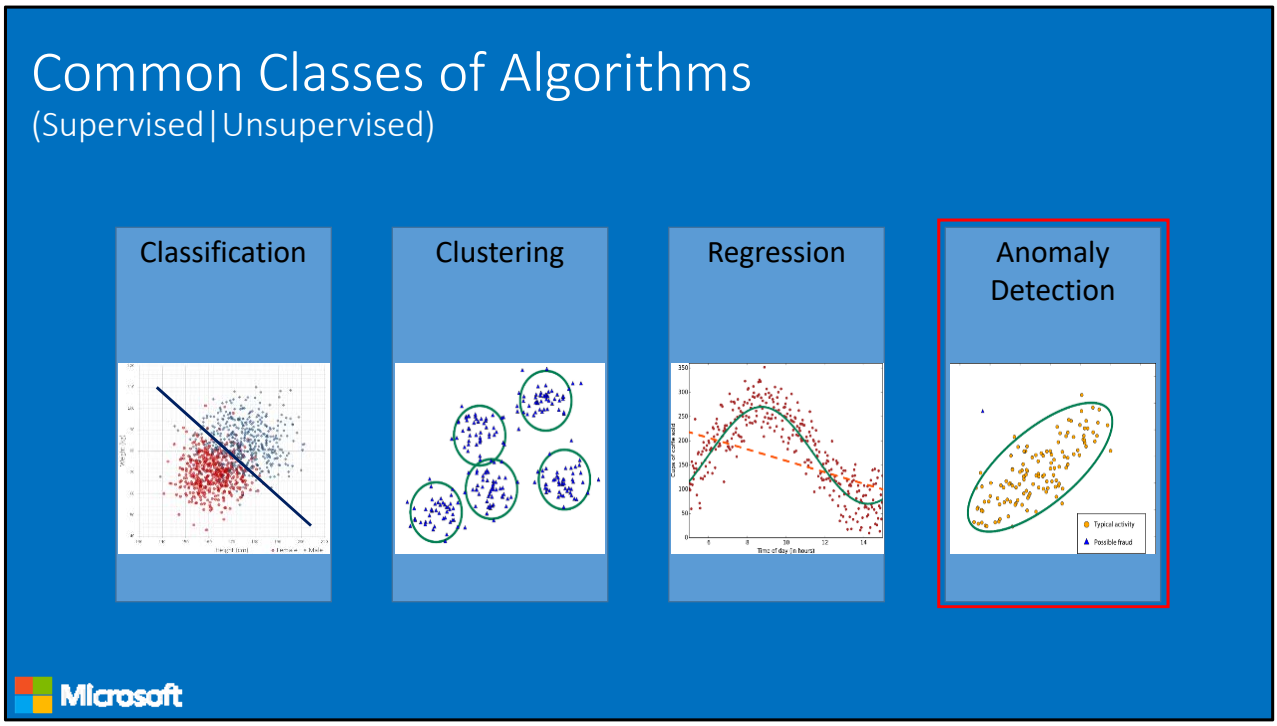
Microsoft

Basic definition:

Machine learning develops algorithms for making **predictions (statistical sense)** from data *

Learning models from available training data, to make good predictions on unseen test data

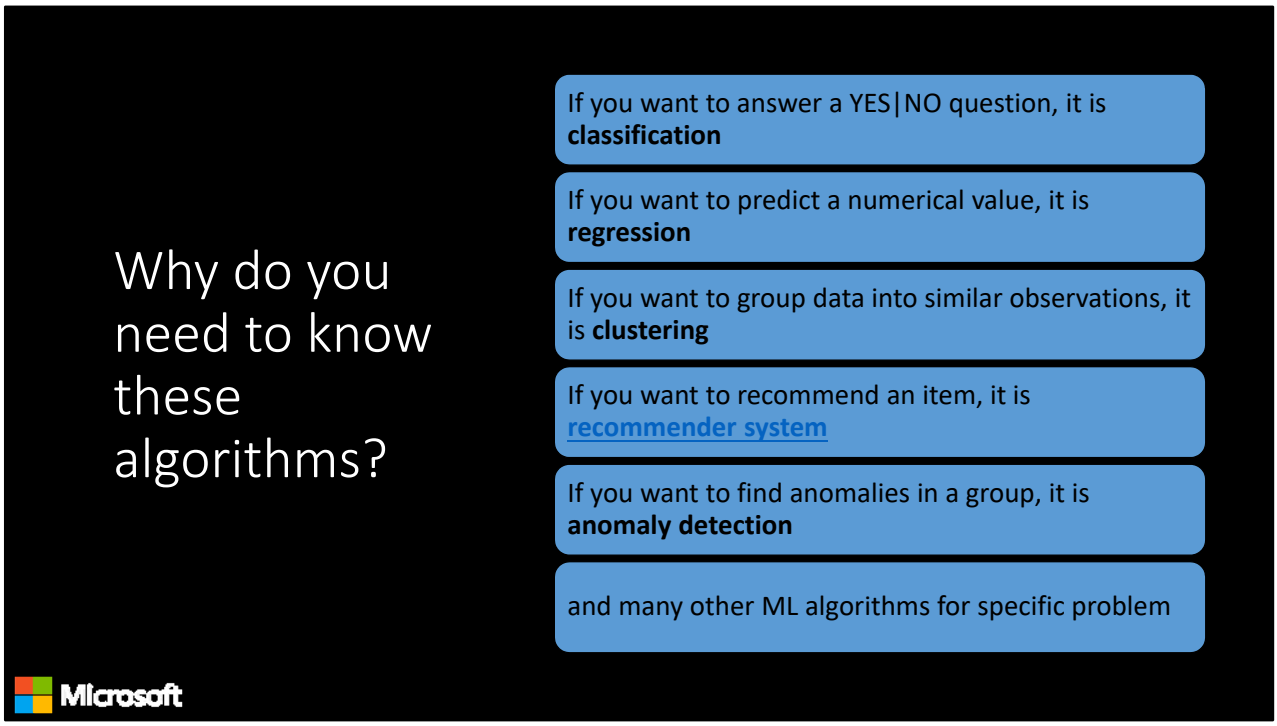
Highlight the keywords: Known Data, Model, Unknown Data and the Prediction (statistically…)



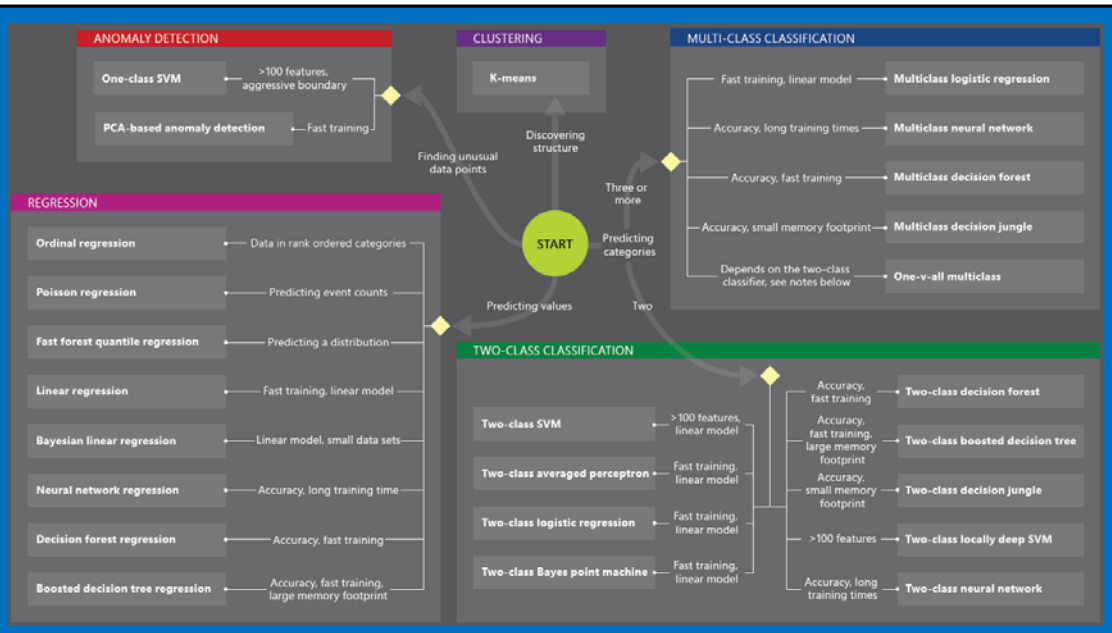
More samples on: https://azure.microsoft.com/en-us/documentation/articles/machine-learning-algorithm-choice/

Mention about Classification, Regression etc.

© 2012 Microsoft Corporation. All rights reserved. Microsoft, Windows, and other product names are or may be registered trademarks and/or trademarks in the U.S. and/or other countries.
The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation. MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.



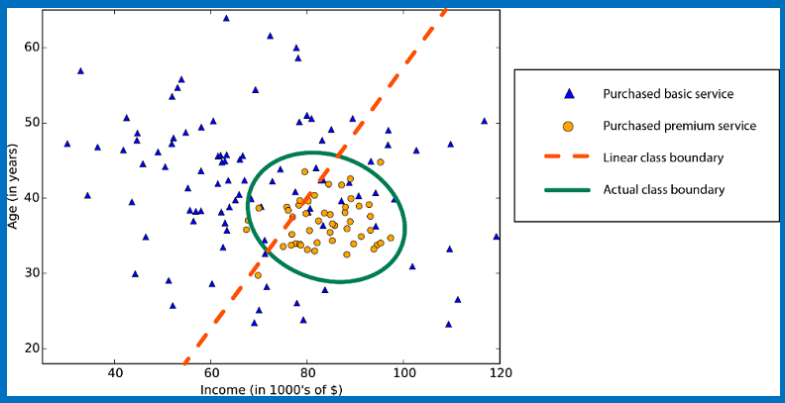
For a lab that is recommendation (recommender) system, see Lab 8 at https://github.com/Azure-Readiness/hol-azure-machine-learning/blob/master/008-lab-recommendation-system.md.



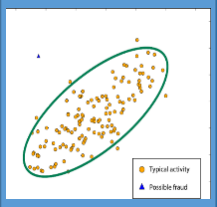
https://docs.microsoft.com/en-us/azure/machine-learning/machine-learning-algorithm-cheat-sheet

Microsoft

Anomaly Detection

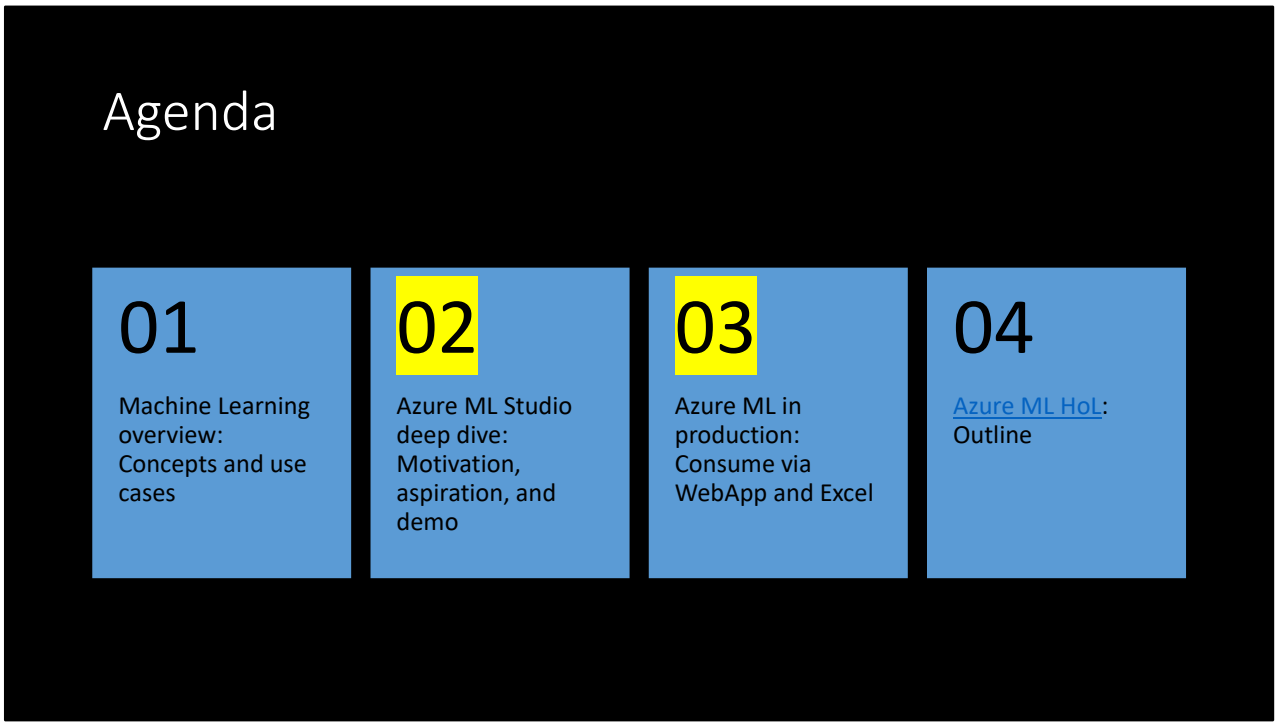


Anomaly Detection



https://docs.microsoft.com/en-us/azure/machine-learning/machine-learning-algorithm-choice

Microsoft



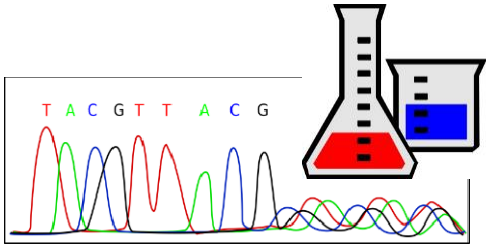
Azure ML and ML overview. Highlighting the values that Azure ML could bring to IT and business, where Azure ML fits in the entire BI stack, and ML solutions and use cases. This serves as a transition from the Cortana webinar to this one.
Azure ML Studio deep dive. Walking through the full machine learning process end-to-end using ML Studio. We'll demonstrate how easy it is to build a savvy ML model with no coding involved. We'll cover from data to ML model building, testing, and evaluation.
Azure ML in production. Realizing the values of Azure ML in business decision making when a predictive model is deployed. We'll end the session with what you could do more to operationalize ML.
Azure ML HoL. https://github.com/kayapperson/AzureMLHoL-CERN. Talk about the lab outline and prerequisites.

Data Science

- Data Science is far too complex
  - Cost of accessing/using efficient ML algorithms is high
  - Comprehensive knowledge required on different tools/platforms to develop a complete ML project
  - Difficult to put the developed solution into a scalable production stage

- Need a simpler/scalable method:

  Azure Machine Learning Service



Microsoft

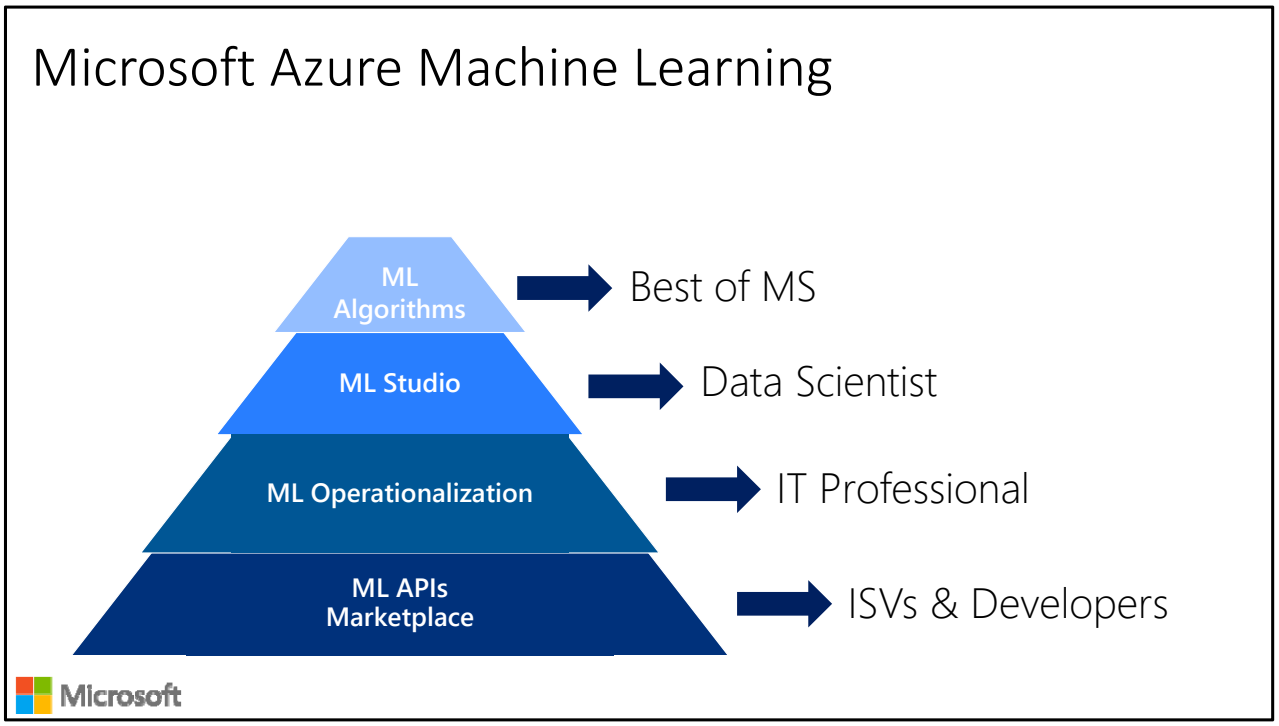
Actually it is the science on data.
Data science is far too complex. For example, I spent years to understand it, so that I could explain it clearly and simply. However, once you grasped the concepts by understanding it simply, you may not need all the bagages, maths and stats, to realize the values of data science.

Microsoft Azure Machine Learning

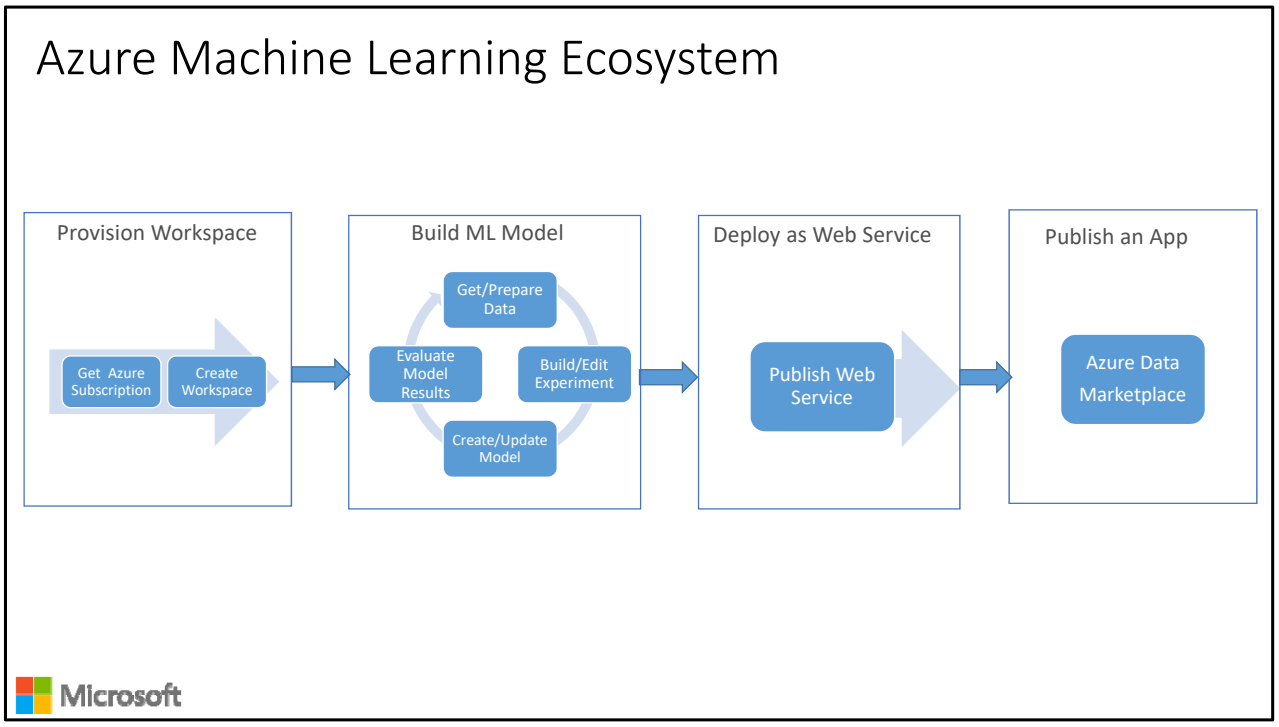
Make machine learning accessible to every enterprise, data scientist, developer, information worker, consumer, and device anywhere in the world.

Microsoft

Data science is far too complex. For example, I spent years to understand it, so that I could explain it clearly and simply. However, once you grasped the concepts by understanding it simply, you may not need all the bagages, maths and stats, to realize the values of data science.

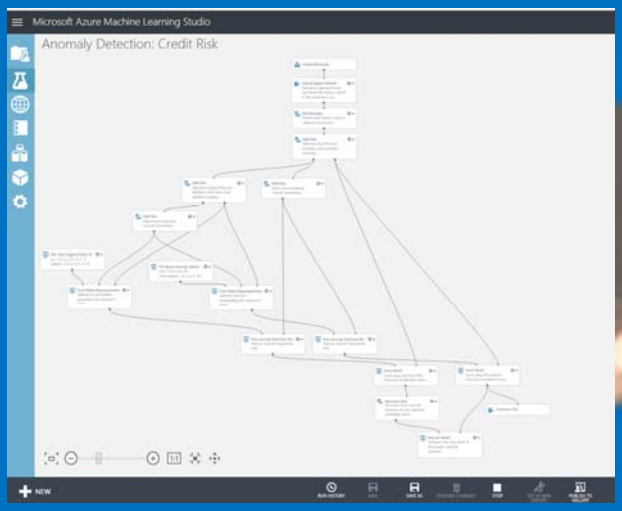


How customers generally benefit from Azure ML



Stages to develop end to end Azure ML solution. First you need an AML workspace... then you build a model under AML Studio, then publish it as web service and as an app (see text analytics demo in prior slides)

Demo Anomaly Detection Credit Risk



First time users: https://portal.azure.com to setup ML workspace
Returning users: https://studio.azureml.net
Cortana Intelligence Gallery: https://gallery.cortanaintelligence.com/Experiment/Anomaly-Detection-Credit-Risk-5

Microsoft

ML Studio:

Web based UI accessible from different browsers

Share|collaborate to any other ML workspace

Drag&Drop visual design|development

Wide range of ML Algorithms catalog

Extend with OSS R|Python scripts

Share|Document with IPython|Jupyter

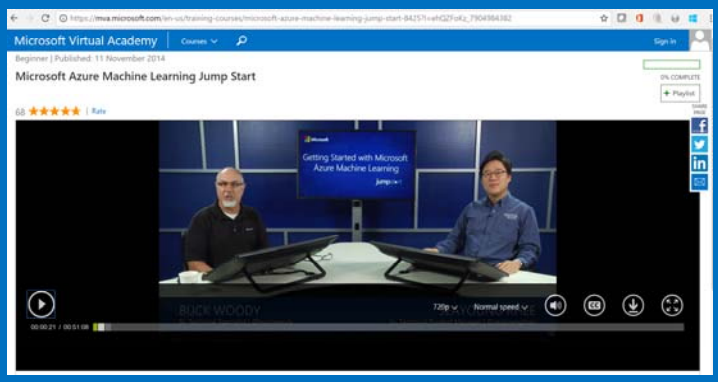
Deploy|Publish|Scale rapidly (APIs)

Data is credit risk from Kaggle competition.

https://www.kaggle.com/dalpozz/creditcardfraud/version/2

Azure Machine Learning Jump Start (highly rated)

- @Microsoft Virtual Academy



Other courses on edX with Azure ML:
Principles of Machine Learning dat203-2x-4

Data Science Essential dat203-1x-4

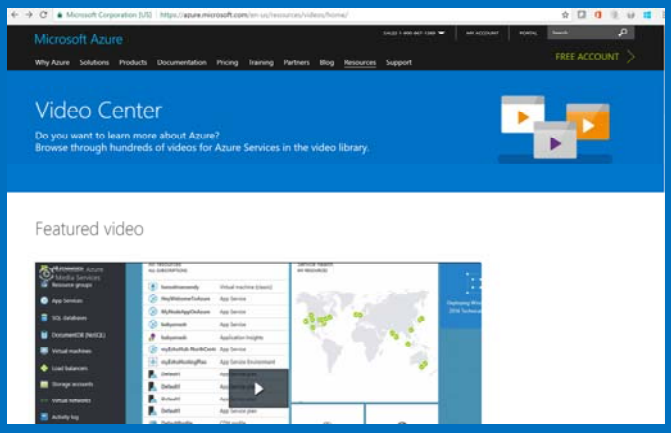
Applied Machine Learning dat203-3x-2

Microsoft

Thank you!
kay.apperson@microsoft.com

3 hours. 4 modules. 4 assessment.

Microsoft Azure Video Center



There are more videos at https://azure.microsoft.com/en-us/resources/videos/home/.

Learn How to Work with Large Datasets to Build Predictive Models with Microsoft's Analytics Toolkit

Cloud Cover 151: Azure Machine Learning with Parmita Mehta

Enhancing Your Application with Machine Learning Through APIs

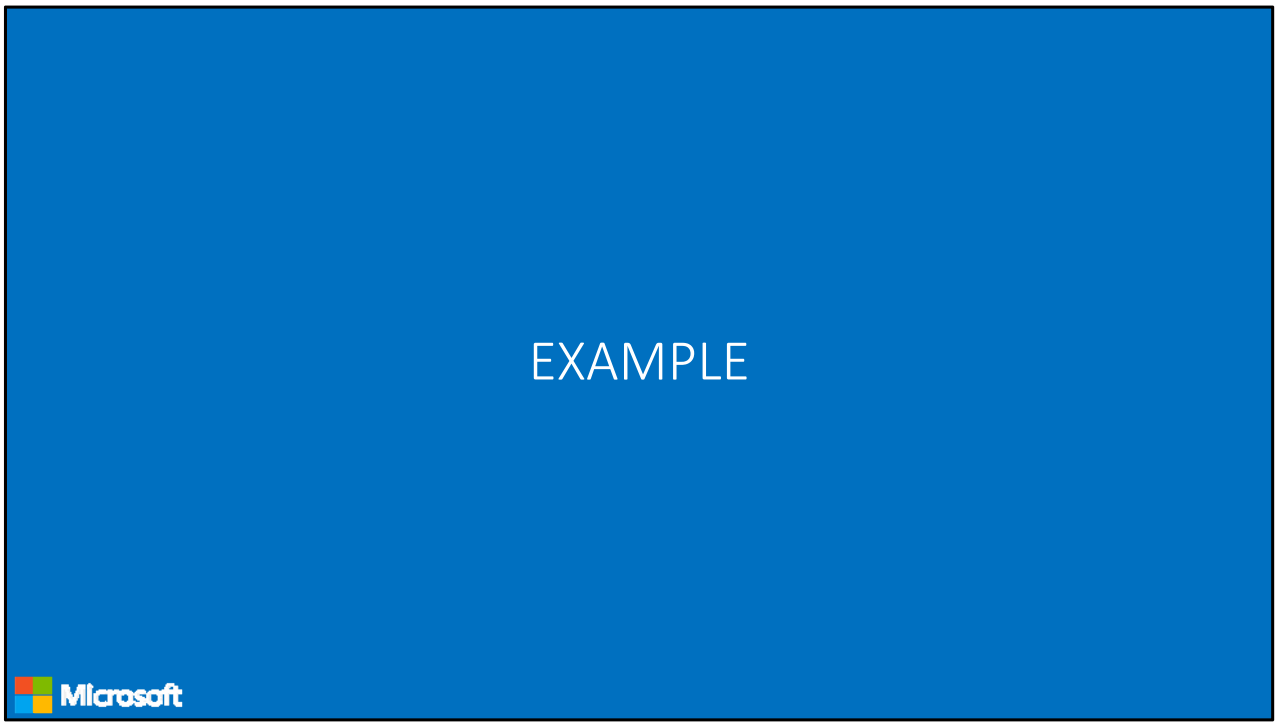
Microsoft Cognitive Services: Give Your Apps a Human Side

Integrating Azure Machine Learning with Azure SQL Data Warehouse

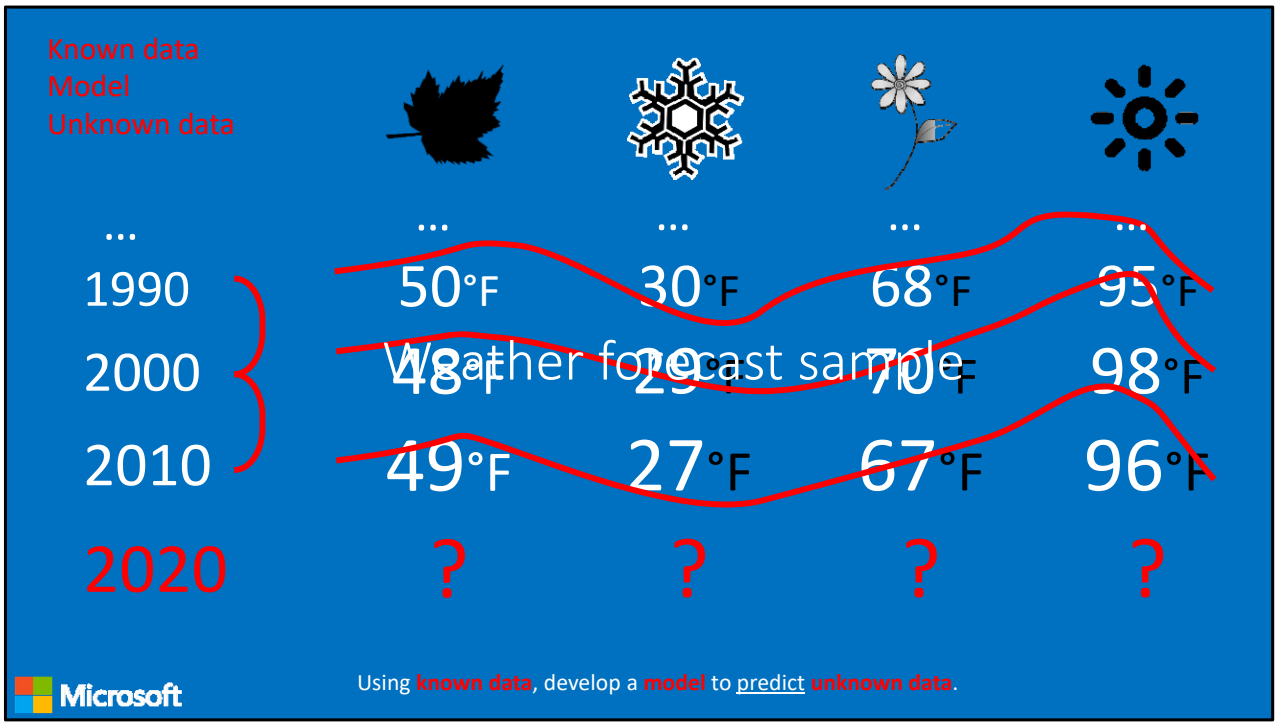
Building a Recommender System in Azure Machine Learning Studio

Microsoft

There are more videos at https://azure.microsoft.com/en-us/resources/videos/home/. To narrow down, go to All videos and choose Machine Learning in the SERVICES drop down menu.

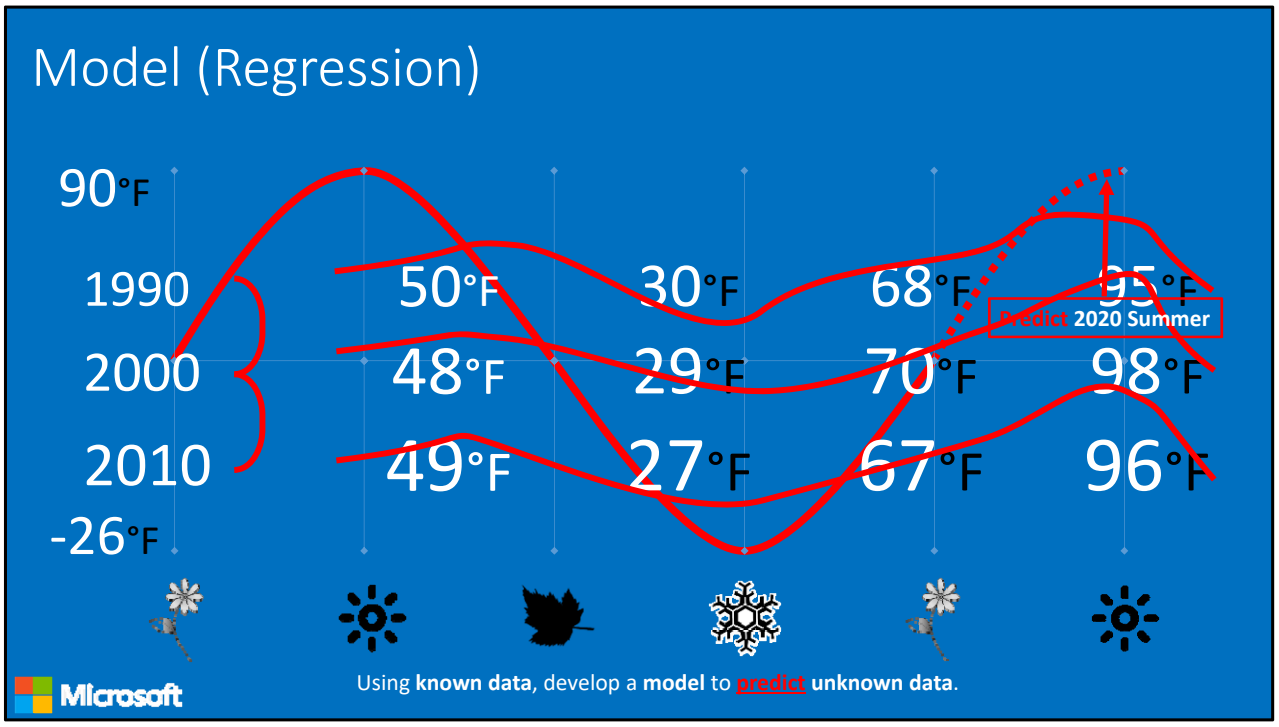


See the “ML algorithms” section of this file.



Weather forecast sample:

- You have previous weather data
- You build your ML model based on existing data
- Based on the model you developed, make good predictions about future

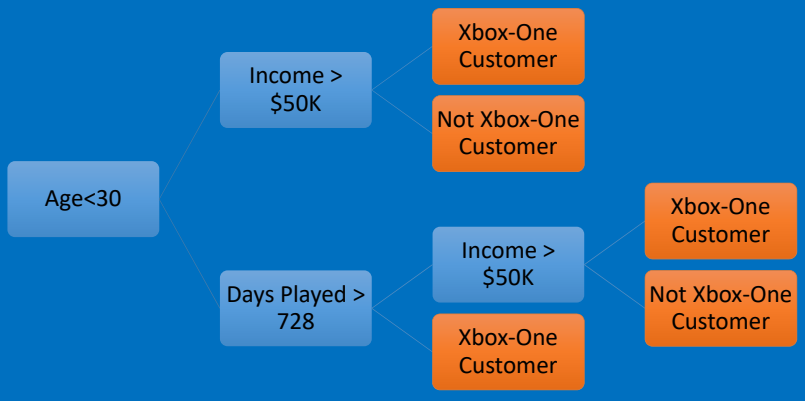


Using appropriate algorithms with the existing data, we generate a mathematical model, formulation or define a pattern to predict future, unknown values

Highlight the keywords: Known Data, Model, Unknown Data and the Prediction (statistically…)

EXAMPLE

Model (Decision Tree)



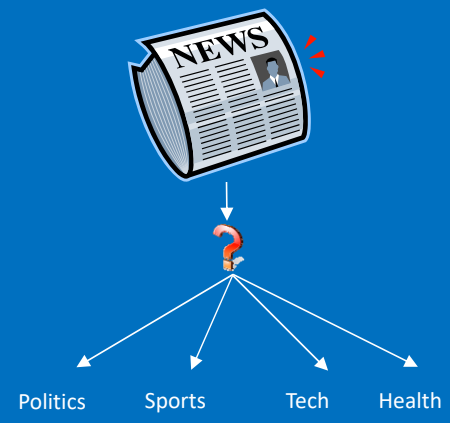
Microsoft

More samples on: https://azure.microsoft.com/en-us/documentation/articles/machine-learning-algorithm-choice/

EXAMPLE

Model (Classification)

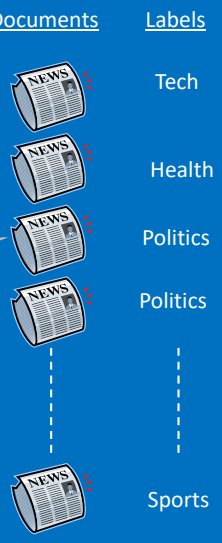
Classify a news article as (politics, sports, technology, health, …)



Microsoft

Using **known data**, develop a **model** to predict **unknown data**.

Known data (Training data)



Documents consist of unstructured text. Machine learning typically assumes a more structured format of examples

Process the raw data

Microsoft

Using **known data**, develop a **model** to predict **unknown data**.

Known data (Training data)

Documents Labels

Tech

Health

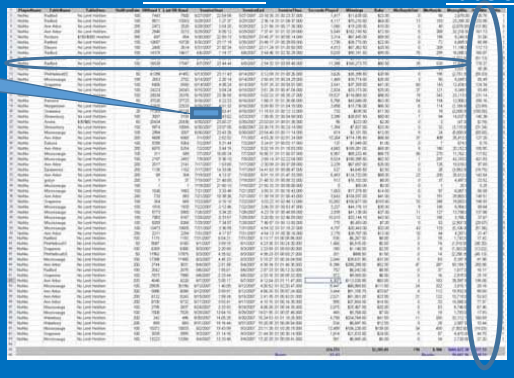
Politics

Politics

Sports

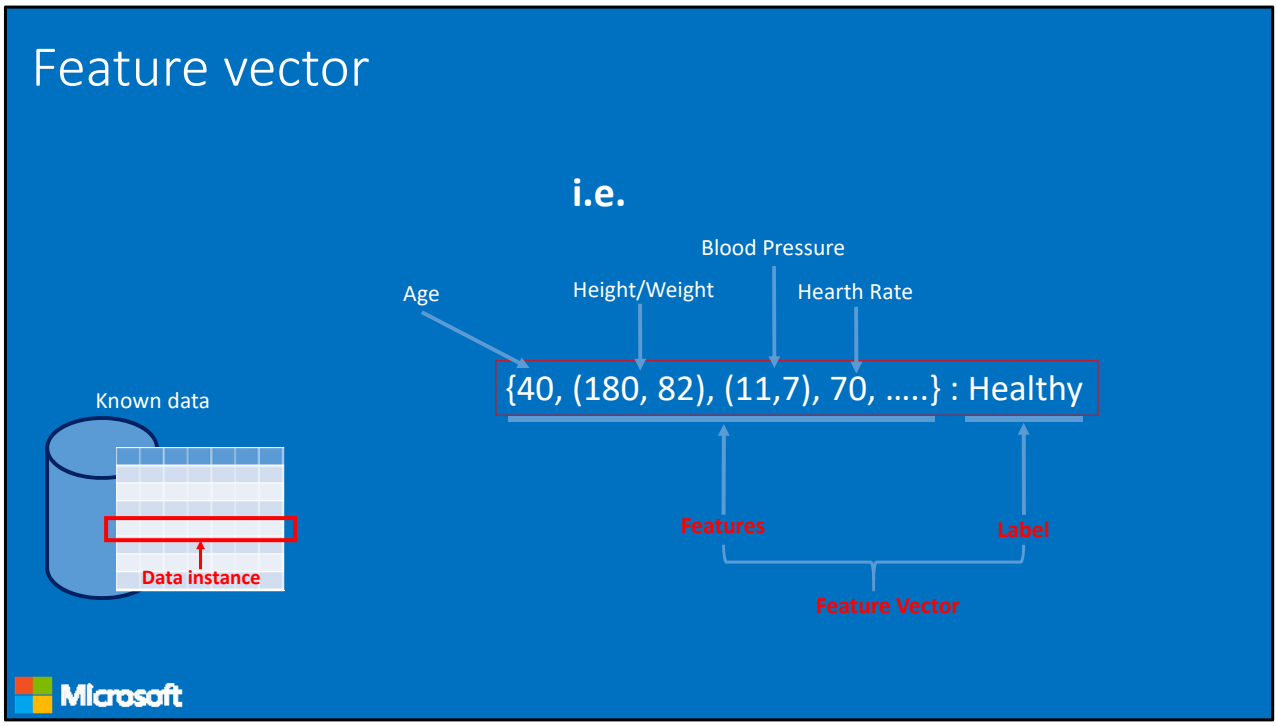
Process each data instance to represent it as a feature vector

Feature



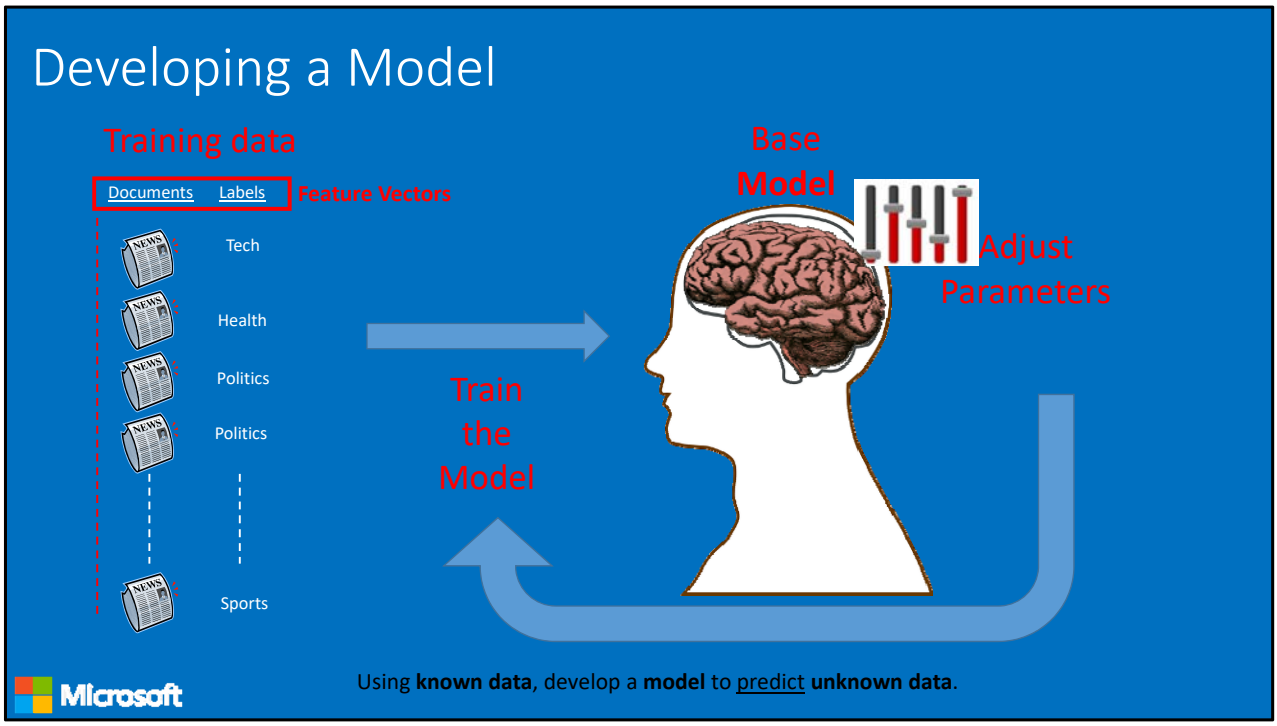
Microsoft

Using **known data**, develop a **model** to predict **unknown data**.

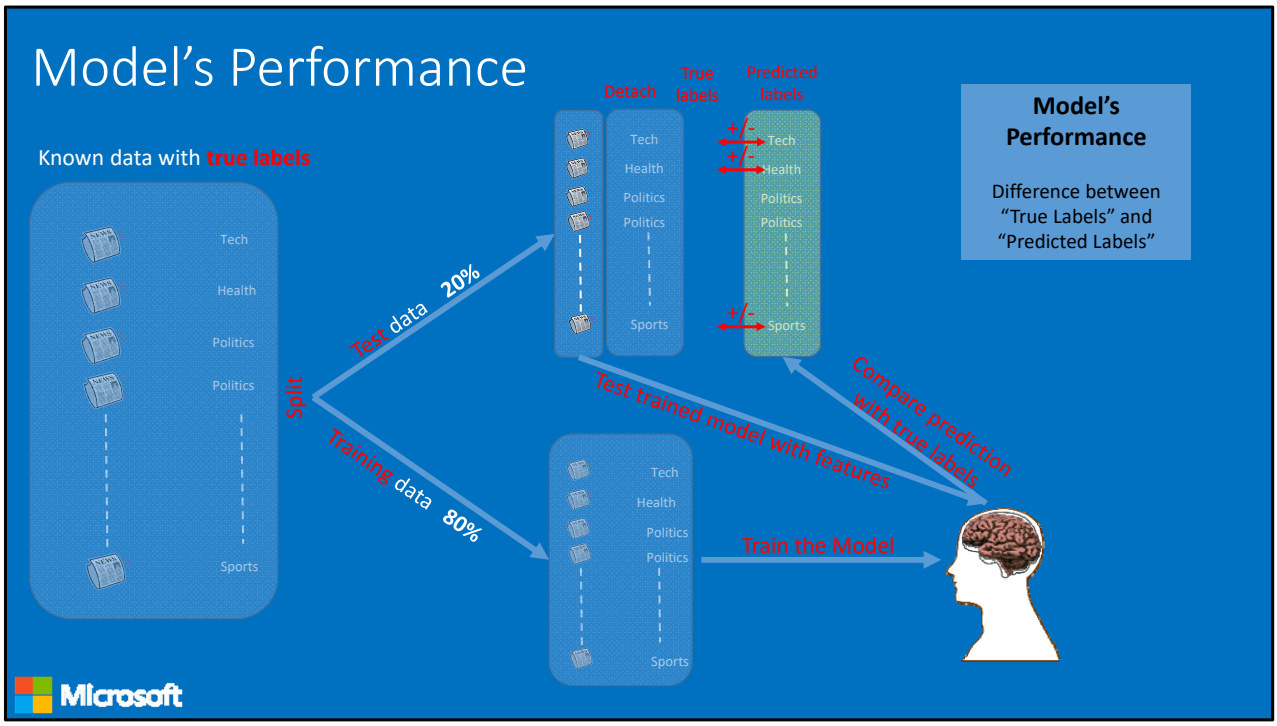


Just a brief info about some keywords used in ML

Developing a Model



Using **known data**, develop a **model** to predict **unknown data**.



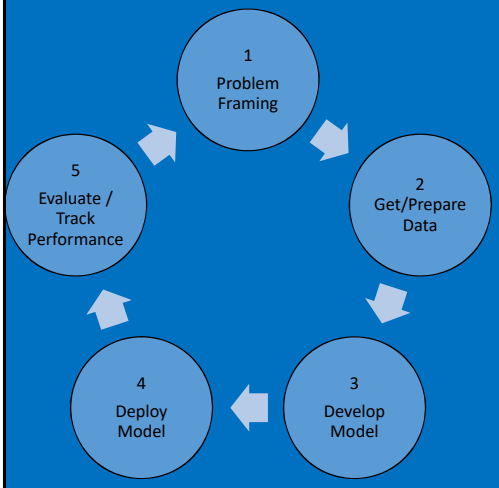
Ratio of the Test/Training data is trivial. Depends on model, data size, quality etc.

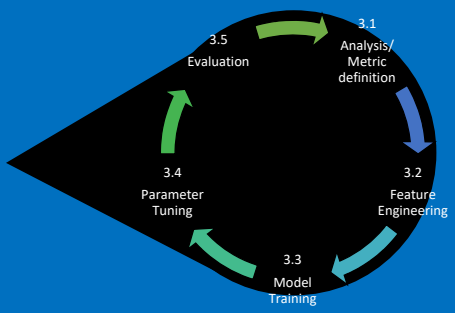
Split is random (i.e. which %80 portion of the data?)

There is no rule to split into %80 & %20 … it might be %50 & %50 etc depending on the problem, model, case.

If you are not happy with the Model's performance than adjust the parameters and re-train the model. Or change the algorithm (training approach) behind the method…

Steps to Build a Machine Learning Solution





Microsoft

Confusion

- Statistics

  Develops methods|models that explains data

- Data Mining

  Find previously unknown schema, patterns in data

- Artificial Intelligence

  Procedure of programing a computer to behave intelligent

- Predictive Analytics

  Using Machine Learning techniques to predict future

Microsoft