

Azure Machine Learning (Data Camp)

Mostafa Elzoghbi

Sr. Technical Evangelist – Microsoft

@MostafaElzoghbi

<http://mostafa.rocks>

Session Objectives & Takeaways

- What is Data Science?
- What is Machine Learning?
- Why do we need Machine Learning?
- Azure ML studio platform capabilities.
- Machine Learning basic concepts & algorithms:
 - How to build a model
 - Supervised vs Unsupervised models
 - Evaluate a model.
 - ML algorithms: Regression, classification, clustering and recommendations.

Data Science Involves

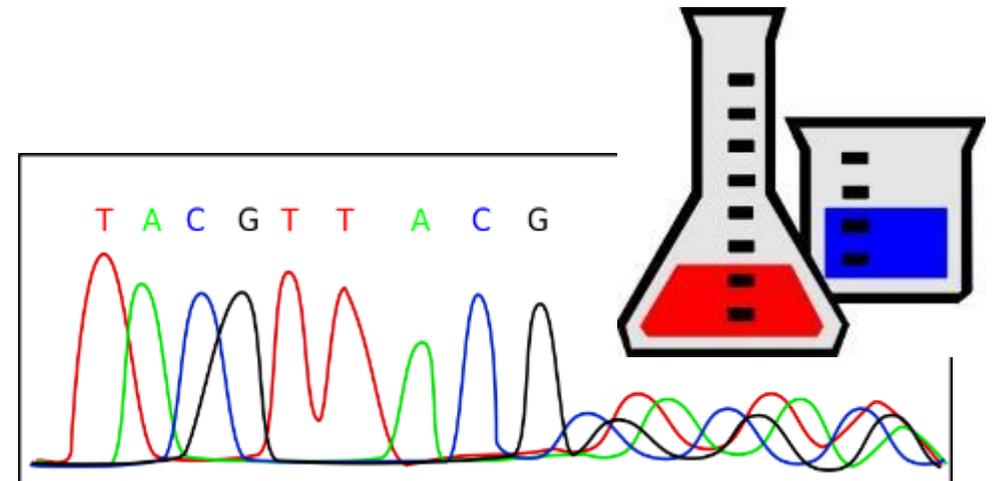
- Data science is about using **data** to **make decision** that **drive actions**.
- Data science process involves:
 - Data selection
 - Preprocessing
 - Transformation
 - Data Mining
 - Delivering value from data : Interpretation and evaluation

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



What is Machine Learning ?

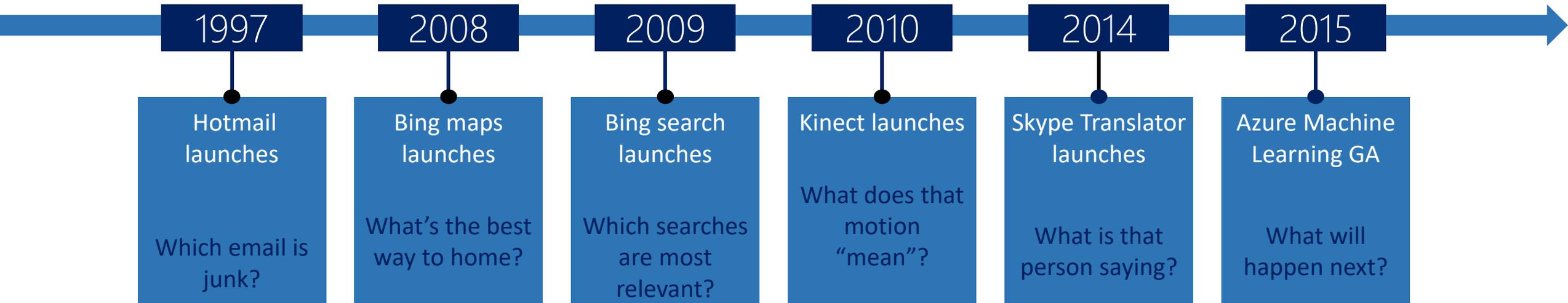
- Using **known data**, develop a **model** to predict **unknown data**.

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

Model: Known data + Algorithms (ML algorithms)

Unknown Data: Missing, Unseen, not existing, future data

Microsoft & Machine Learning



Why Machine Learning?

Is this where you work? If so, I can give you better directions and suggestions.

Yes No

How'd you guess?

ask me anything

Here's the forecast.

New York, NY

68° TUE 5/22 Low 58° Overcast 30%

70° 3:00 PM Cloudy 20%

71° 4:00 PM Cloudy 10%

73° 5:00 PM Partly Cloudy 10%

73° 6:00 PM Partly Cloudy 10%

71° 7:00 PM Partly Cloudy 20%

68° 8:00 PM Partly Cloudy 20%

Data from iMap Weather

Search for 'what's the weather toda ...

try Will it be warm next Tuesday?

Emotion Recognition

Identify emotions communicated by the facial expressions in techniques to provide these results. You can also click the op

- Image resolution $\geq 36 \times 36$ pixels and the file size $< 4M$
- The frontal and near-frontal faces have the best results
- Recognition is experimental, and not always accurate.

Anger	0.00000
Contempt	0.00000
Disgust	0.00000
Fear	0.00000
Happiness	0.99998
Neutral	0.00002
Sadness	0.00000
Surprise	0.00000

Image URL

39 40 43 45

Mark Carquero @reelofcarquero May 15
How the @HowOldRobot handled my dishwashing board. No facial hair - youth, mustache - old, ha

Stir fried baby vegetables
Stir fried lettuce
Braised Eggplant
Old kitchen Chinese cabbage
Scrambled egg炒鸡蛋

Tap to pause.



getting a morgage in seattle

8,140,000 RESULTS Any time

Ads related to getting a morgage in seattle

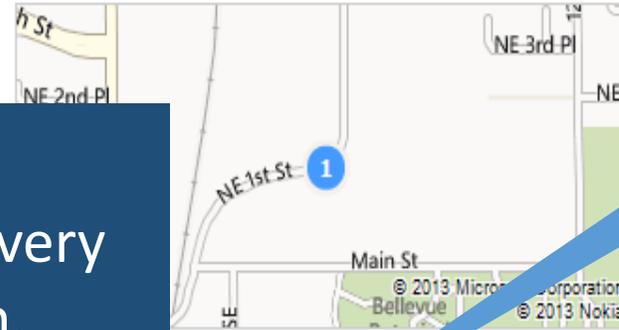
[15-Year Mortgage Rates | QuickenLoans.com](#)
www.QuickenLoans.com/Rates
 Lock Your Rate. 3.500% (3.92% APR) With America's #1 Online Lender.

[Lending](#)
 Lending
 APR fr

[TILA](#)
 seattle
 Meet o

[Pre Q](#)
www.wellstargo.com/mortgage
 Estimate how much you can afford

Seattle's Best Mortgage Inc



Website: seattlesbm.com

Report a problem

RELATED SEARCHES

- Getting a First Mortgage
- Getting a Mortgage Self-Employed
- Getting a Mortgage Loan Approved
- Getting a Mortgage On Land
- Getting a Mortgage in 2013
- How to Get a Mortgage License
- How to Get a Mortgage After Bankruptcy
- Mortgage Calculator

Ads related to getting a morgage in seattle

Machine learning enables nearly every value proposition of web search.

What language?

Which ads to show, and in what order?

Misspelled?

Which links are most likely to get clicked?

What is the probability of a click on each ad?

What is the intent?

Are any of these pages malicious?

What pages should we index?
What ad pricing will optimize revenue?

Including results for *getting a mortgage in seattle*.
Do you want results only for *getting a morgage in seattle*?

Seattle Mortgage Rates - Find the Best Home Loan | Zillow

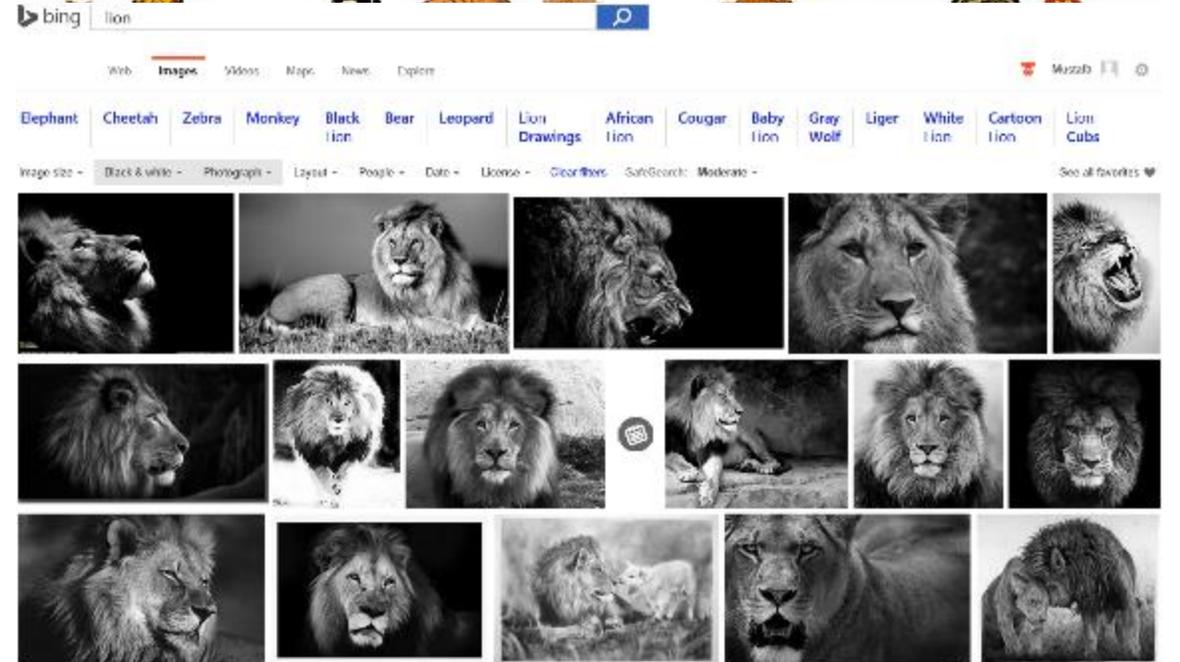
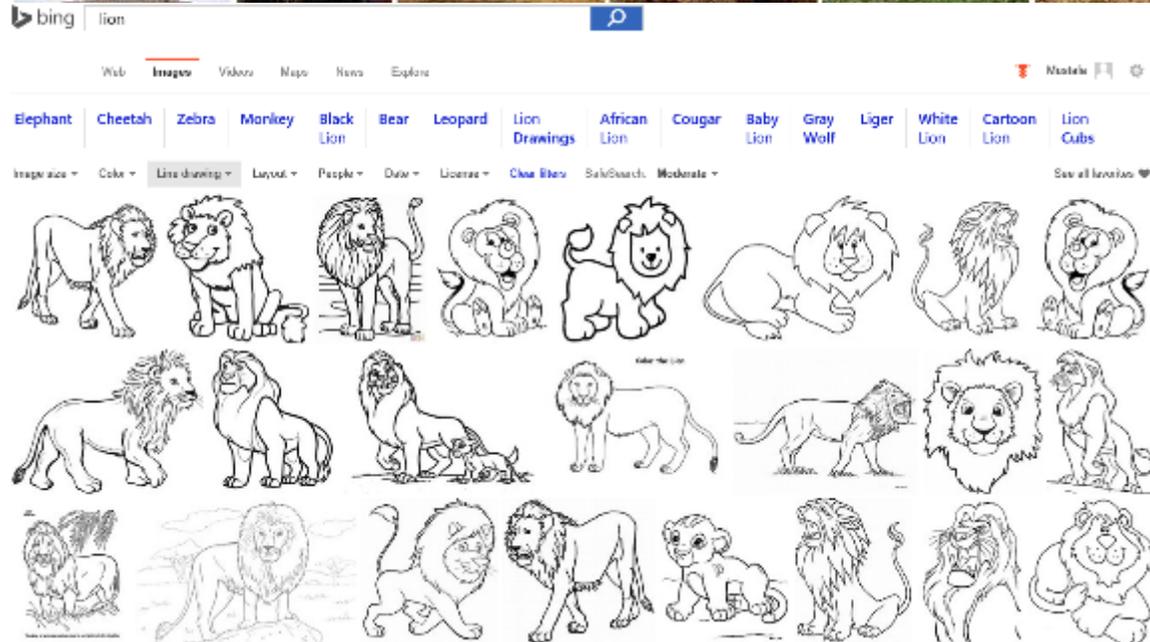
www.zillow.com/mortgage-rates/wa/seattle
See up to the minute Seattle mortgage rates and find Seattle Washington's best, lowest possible quote with Zillow Mortgage Marketplace.

Seattle's Best Mortgage

www.seattlesbm.com
Get the best mortgage loan for you at Seattle's Best Mortgage. (CL#117721) When you decide to buy a home or refinance a mortgage, it's a big step.

1 11911 Ne 1st St Ste B306, Bellevue · (425) 228-7000 · Directions · Bing Local

Image Analyze



Accent Color: Which border color is the best?



Accent Color: 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

Accent Color: Windows 10 Store

The screenshot shows the Windows 10 Store interface with a blue accent color. The top navigation bar includes 'Home', 'Apps', 'Games', 'Music', and 'Movies & TV'. A search bar is located on the right. The main content area features a large promotional image for 'Rihanna: Ultimate' by Ace of Dotcom, LLC, with a five-star rating and a 'Share' button. Below this, there is a 'Rihanna Fan App.' section with a 'More' link and a 'Free' button. The right side of the page displays a grid of music singles and YouTube videos related to Rihanna's work.

Store

Home Apps Games Music Movies & TV

Search

 **Rihanna: Ultimate**
Ace of Dotcom, LLC.
★★★★★
Share

Rihanna Fan App.
More
Free

Rihanna: Ultimate

Singles >

- RIHANNA WHAT NOW**
What Now (Remixes Part 2)
Release Date: 11/19/2012...
- Diamonds**
Diamonds (Remix)
Release Date: 11/19/2012...
- Where Have You Been**
Where Have You Been (The Calvin Harris Extended Remix)
Release Date: 6/13/2009...
- Where Have You Been (Dj Vlad Sawyer Remix)**
Release Date: 5/14/2013...
- Diamonds**
Release Date: 9/26/2012...
- Princess of China (Radio Edit)**
Release Date: 6/13/2009...

YouTube >

- Rihanna "What Now": Official Music Video Preview
- Styler To Rock Premieres October 25
- Making of Rihanna Pour It Up Video
- RIHANNA FOR RIVER ISLAND - WINTER COLLECTION
- Trailer: Rihanna 777
- Rihanna Pour It Up Audition Tape

PC

Accent Color: Windows 10 Store

Store

Home **Apps** Games Music Movies & TV

Search

 **Katy Perry: Ultimate**
Ace of Dotcom, LLC.
★★★★★
Share

Katy Perry Fan App.
More
Free

Katy Perry

Albums >

- PRISM**
Release Date: 10/21/2013 8:00:00 PM
- Katy Perry - Teenage Dream: The Complete Confection (Explicit Version)**
Release Date: 3/22/2012 8:00:00 PM
- Teenage Dream**
Release Date: 8/23/2010 8:00:00 PM

Singles >

- Roar**
Release Date: 8/15/2013 8:00:00 PM
- Part Of Me**
Release Date: 3/15/2012 8:00:00 PM
- Firework (feat. David Osmond, Aubree Oliverson, Salt Lake Pops Orchestra & Nathaniel Drew)**
Release Date: 6/26/2012 8:00:00 PM
- The One That Got Away**
Release Date: 12/10/2011 8:00:00 PM

PC

Text Analytics: User reviews

Positive

Text Analytics - Preview by [Azure Machine Learning](#) Demo Sample Code Doc

Try out the Azure ML [Text Analytics service](#) - for free.

To use the service in production, you can get access to its API by [signing up](#) for it (click on "Sign Up" on the right pane on that page). For questions/comments, please use the "Feedback" button in the top right.

I love this presentation. I found it very useful.

Analyze

Sentiment:

98 %

Key phrases highlighted below:

I love this **presentation**. I found it very useful.

Negative

Text Analytics - Preview by [Azure Machine Learning](#) Demo S

Try out the Azure ML [Text Analytics service](#) - for free.

To use the service in production, you can get access to its API by [signing up](#) for it (c the right pane on that page). For questions/comments, please use the "Feedback" b

I hate this presentation. I found it very useless.

Analyze

Sentiment:

7 %

Key phrases highlighted below:

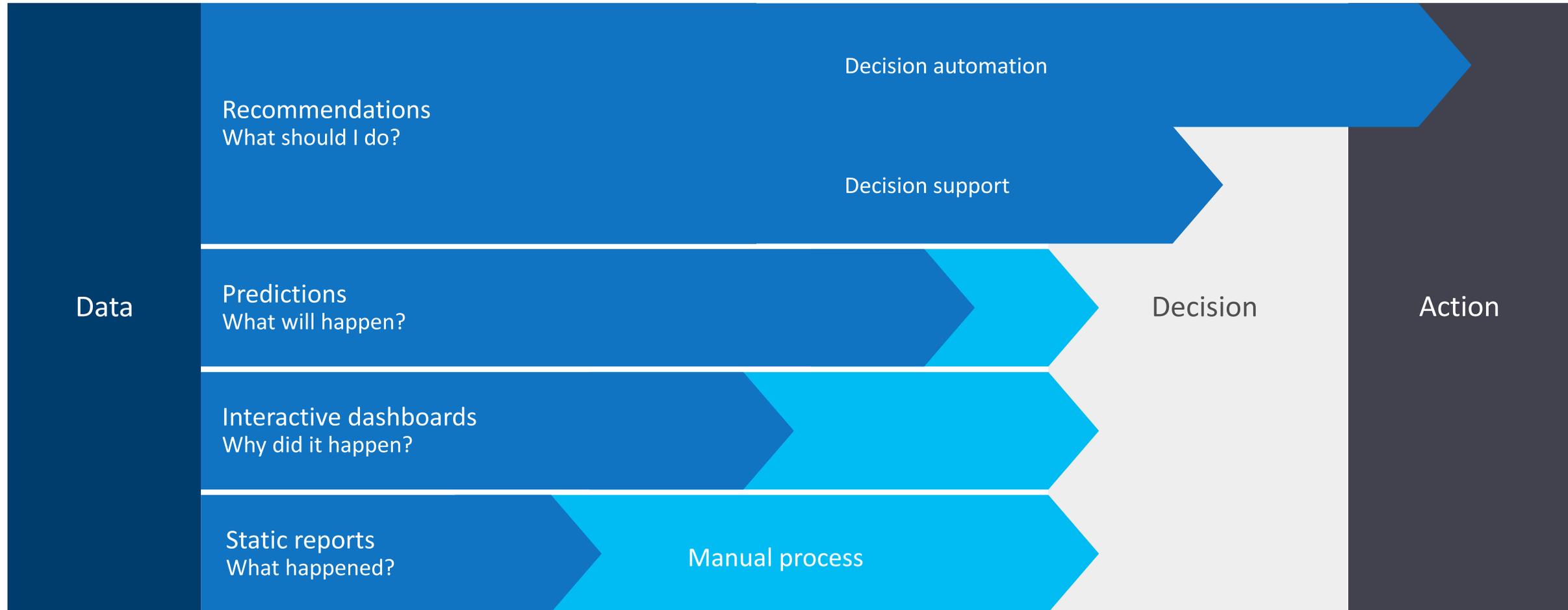
I hate this **presentation**. I found it very useless.

Microsoft Azure Machine Learning

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



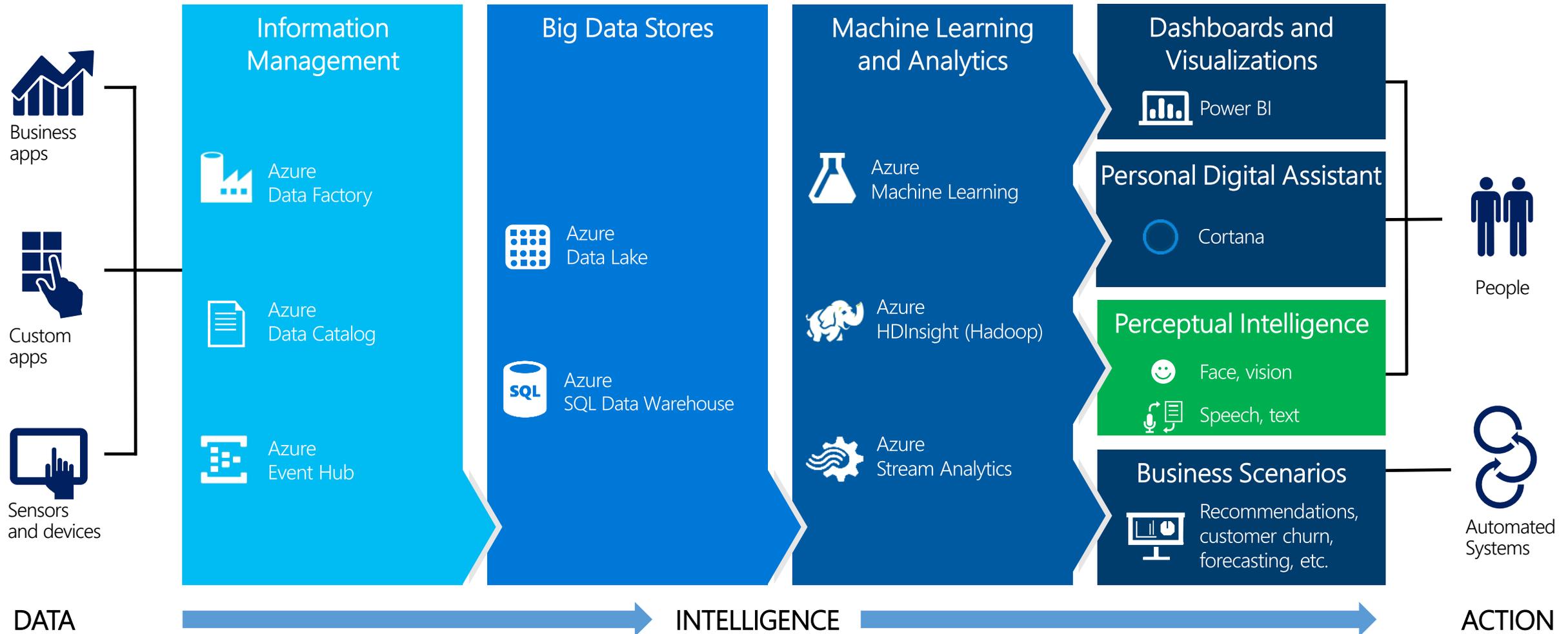
From data to decisions and actions



Value

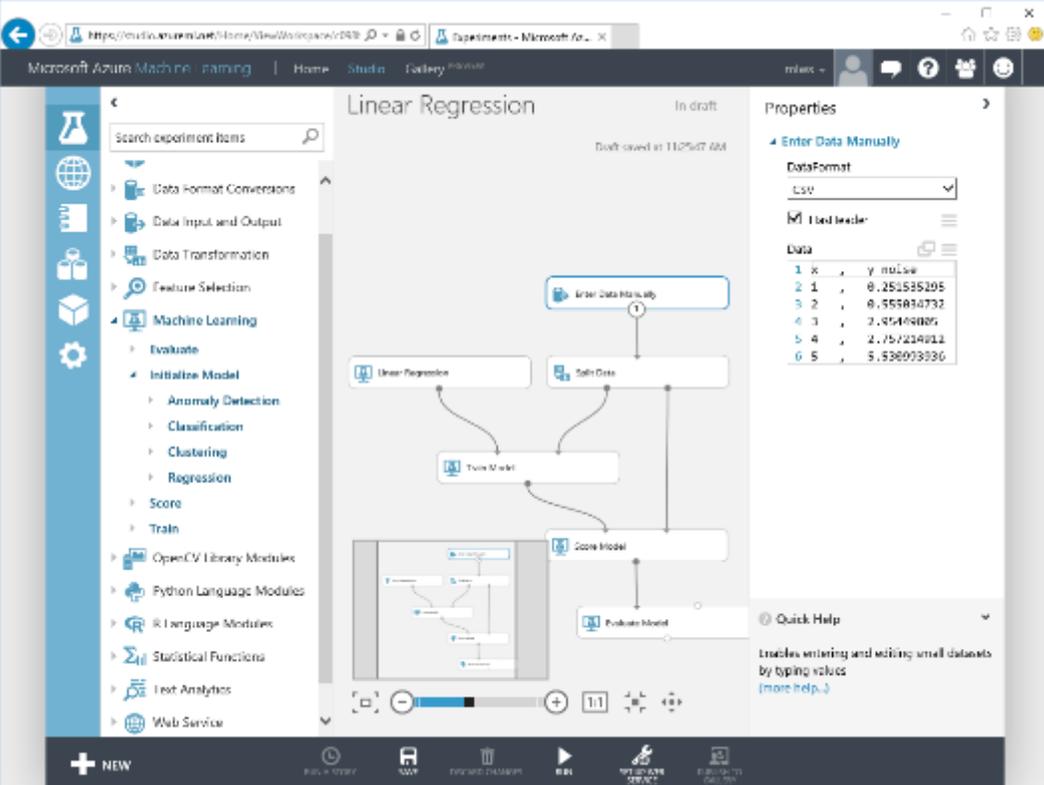


Transform data into intelligent action



Microsoft Azure Machine Learning

- 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**)



The screenshot displays the Microsoft Azure Machine Learning Studio interface. The main workspace shows a workflow for Linear Regression. The workflow consists of several steps: 'Enter Data Manually', 'Train Model', 'Score Model', and 'Evaluate Model'. The 'Enter Data Manually' step is currently selected, and its properties are shown in the right-hand pane. The 'Properties' pane for 'Enter Data Manually' shows the 'DataFormat' set to 'CSV' and a table of data with 6 rows and 2 columns (x and y).

x	y
1	0.251535295
2	0.555834732
3	2.85149885
4	2.751234811
5	5.53693336

The left-hand pane shows a search bar and a list of experiment items, including 'Data Format Conversions', 'Data Input and Output', 'Data Transformation', 'Feature Selection', 'Machine Learning', 'Evaluate', 'Initialize Model', 'Anomaly Detection', 'Classification', 'Clustering', 'Regression', 'Score', 'Train', 'OpenCV Library Modules', 'Python Language Modules', 'R Language Modules', 'Statistical Functions', 'Text Analytics', and 'Web Service'.

Microsoft Azure Machine Learning

Built for a cloud-first, mobile-first world

Fully managed

No software to install, no hardware to manage, all you need is an Azure subscription.

Integrated

Drag, drop and connect interface. Data sources with just a drop down; run across any data.

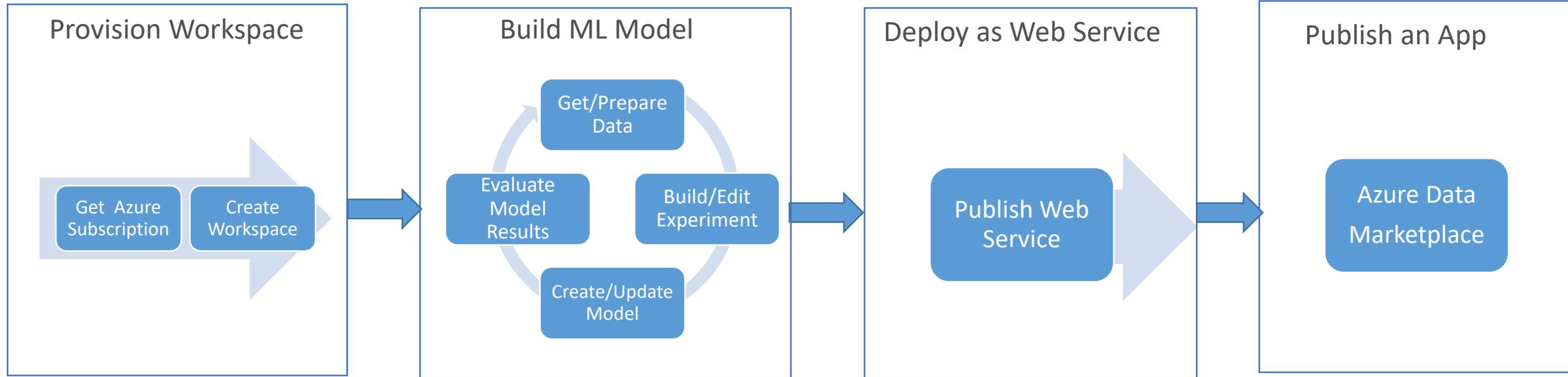
Flexible

Built-in collection of best of breed algorithms with no coding required. Drop in custom R or use popular CRAN packages.

Deploy in minutes

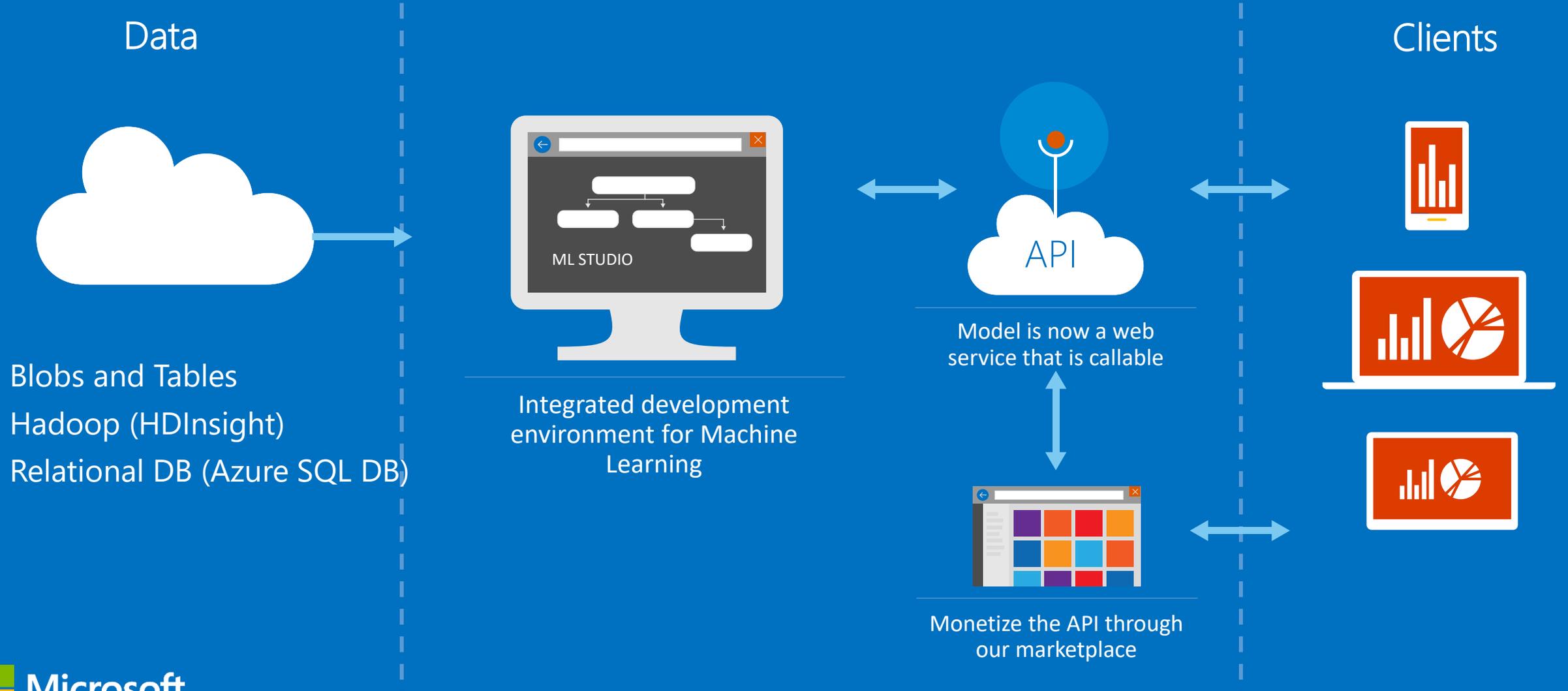
Operationalize models as web services with a single click. Monetize in Machine Learning Marketplace.

Azure Machine Learning Ecosystem



Azure Machine Learning Service

Data -> Predictive model -> Operational web API in minutes

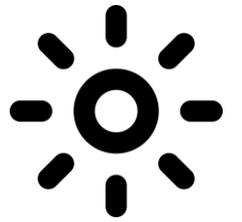
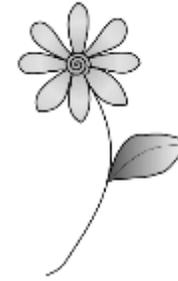


DEMO

Azure Machine Learning Studio

EXAMPLE

Known data
Model
Unknown data



...

...

...

...

...

1990

50°F

30°F

68°F

95°F

2000

Weather forecast sample

48°F

29°F

70°F

98°F

2010

49°F

27°F

67°F

96°F

2020

?

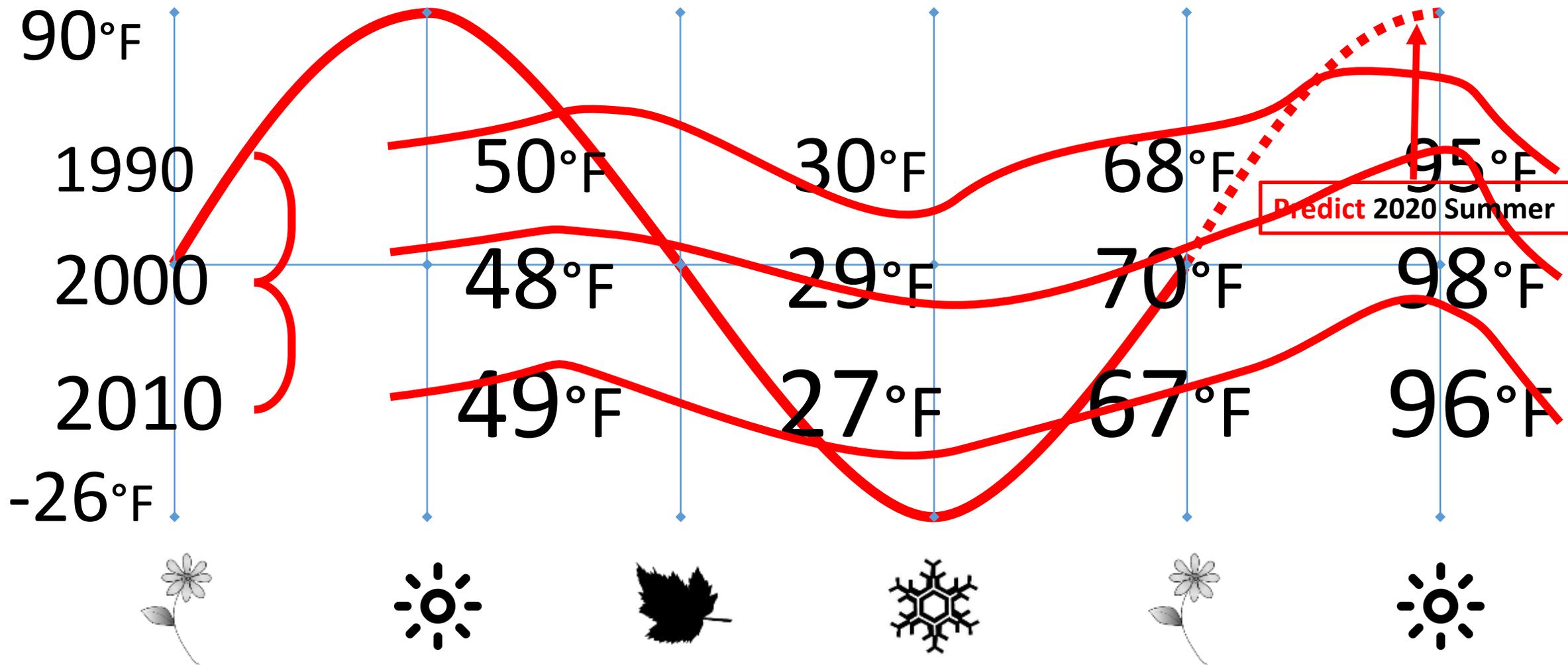
?

?

?

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

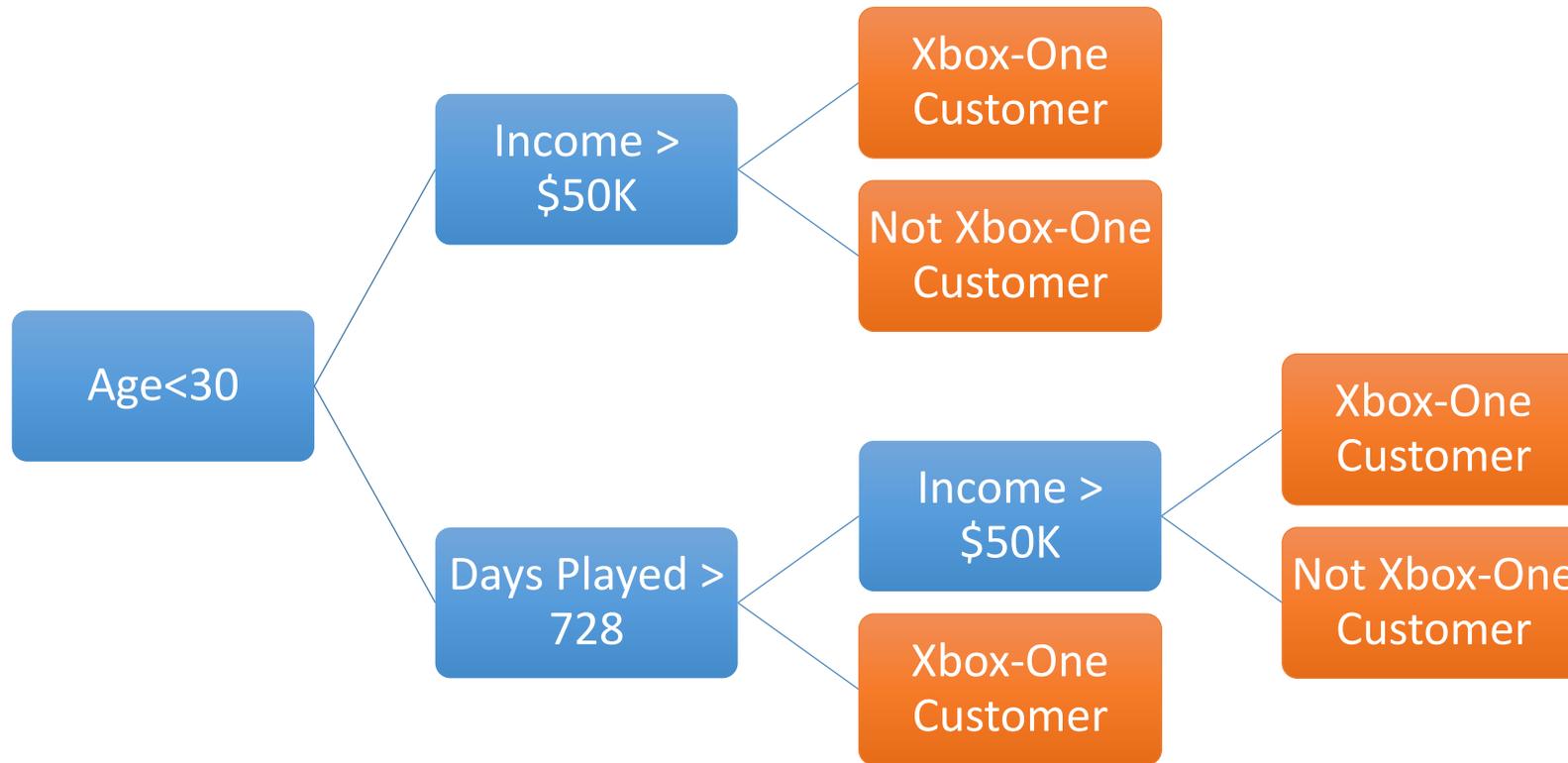
Model (Regression)



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

EXAMPLE

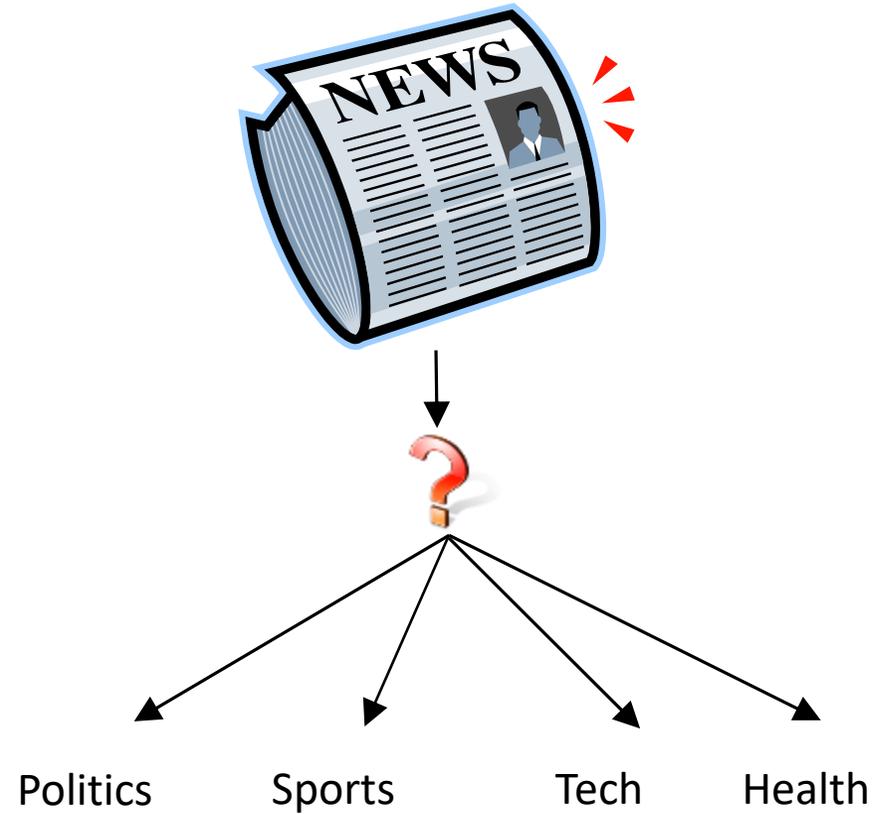
Model (Decision Tree)



EXAMPLE

Model (Classification)

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



Known data (Training data)

Documents Labels



Tech



Health



Politics



Politics



Sports

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

Process the raw data

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

Known data (Training data)

Documents

Labels

Process each data instance to represent it as a feature vector



Tech



Health



Politics



Politics



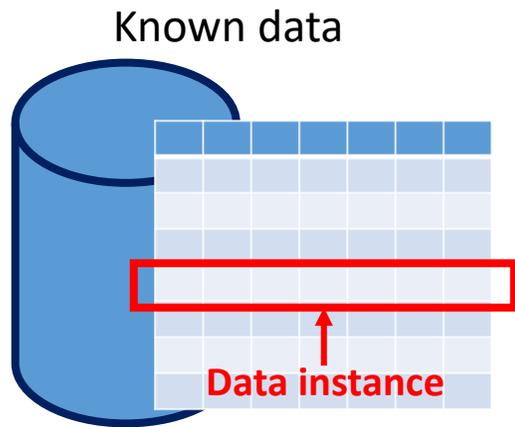
Sports

Feature

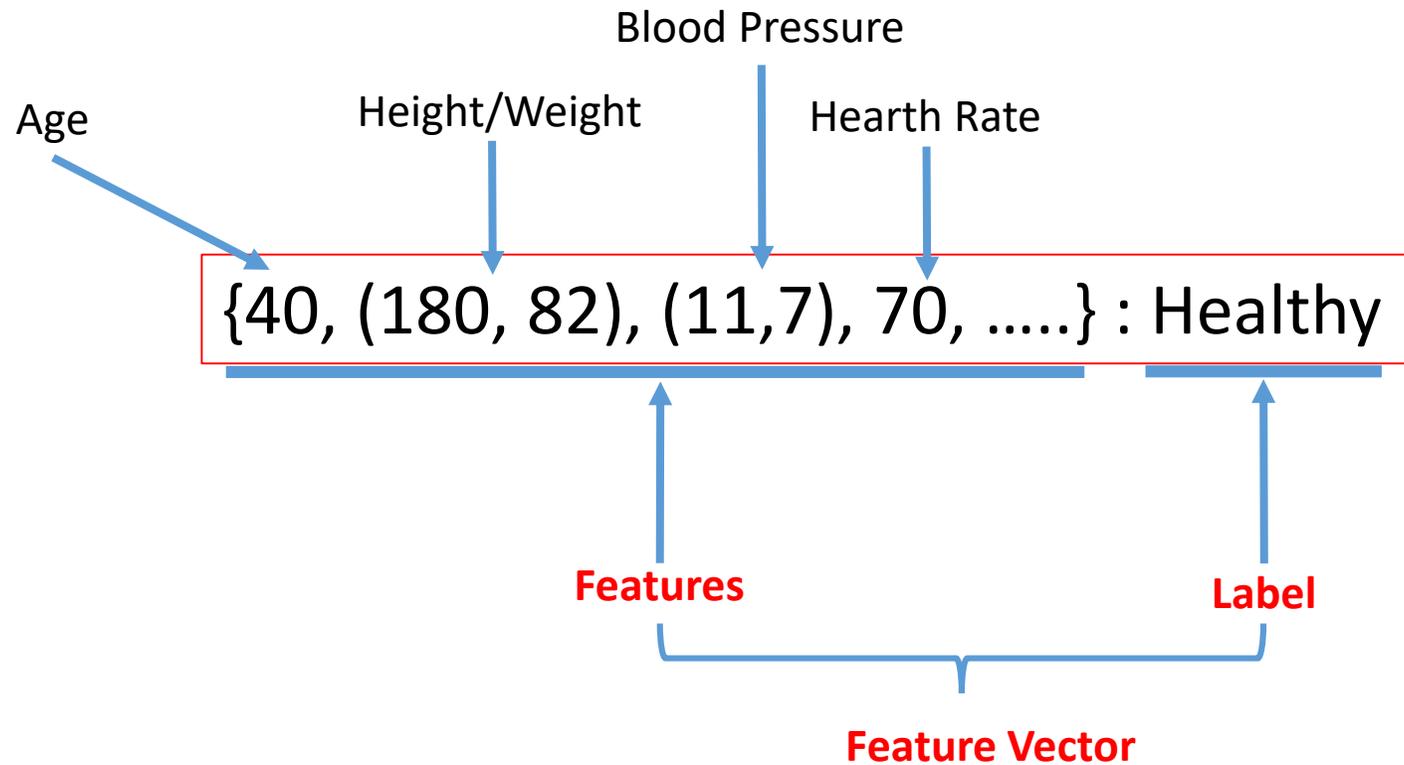


Documents															Labels		
DocumentID	Category	Subcategory	Section	Author	Headline	Text	Score	Score	Score								
1	Tech	AI	Microsoft	John Doe	Microsoft AI Research	Microsoft AI Research	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0	0.0
2	Health	Medical	Healthcare	Jane Smith	Healthcare Innovation	Healthcare Innovation	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0	0.0	0.0
3	Politics	Government	Policy	Bob Johnson	Government Policy	Government Policy	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0	0.0	0.0	0.0
4	Politics	Government	Policy	Bob Johnson	Government Policy	Government Policy	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0	0.0	0.0	0.0
5	Sports	Baseball	MLB	Alice Brown	MLB Season	MLB Season	0.6	0.5	0.4	0.3	0.2	0.1	0.0	0.0	0.0	0.0	0.0

Feature vector

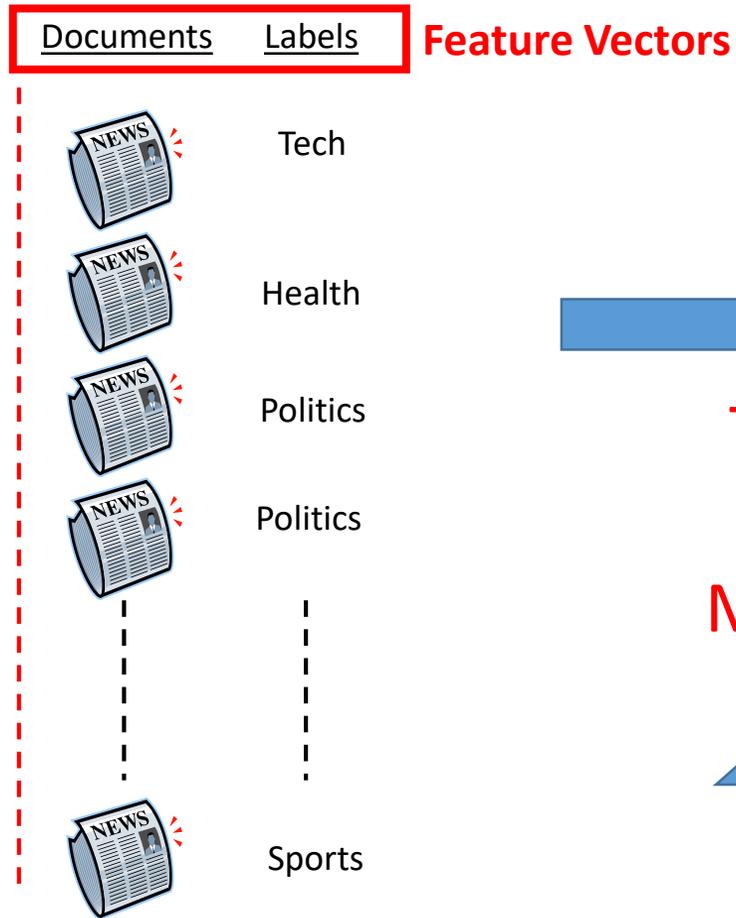


i.e.



Developing a Model

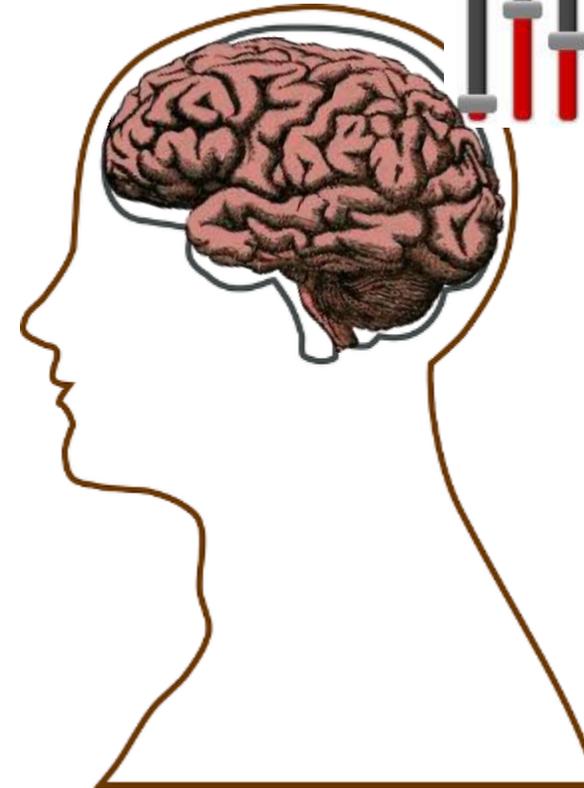
Training data



Train
the
Model



Base Model

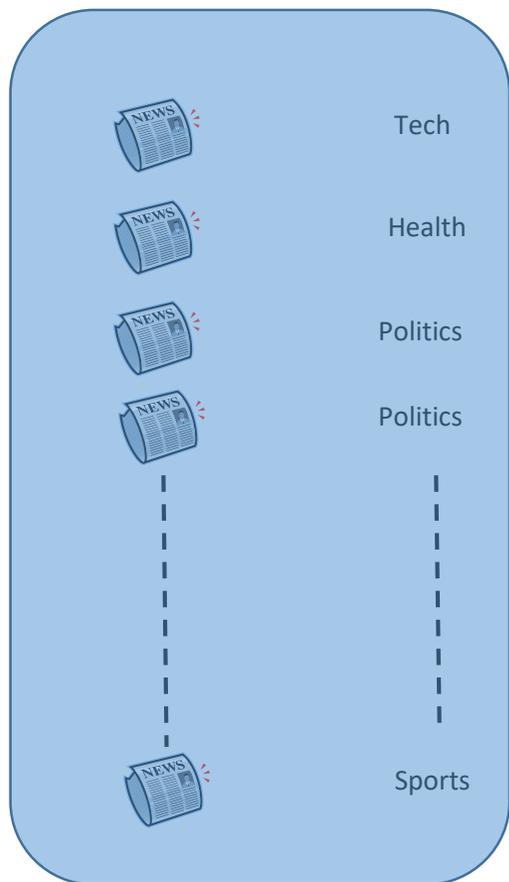


Adjust
Parameters

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

Model's Performance

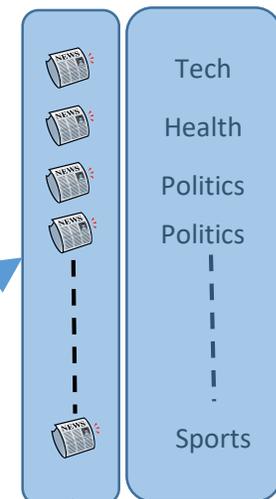
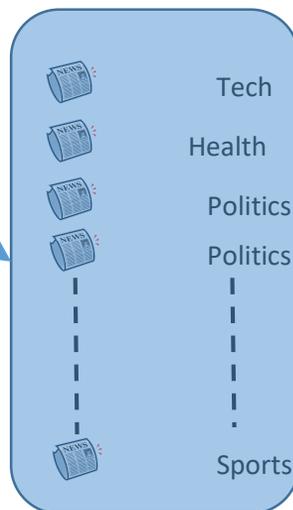
Known data with **true labels**



Split

Training data 80%

Test data 20%

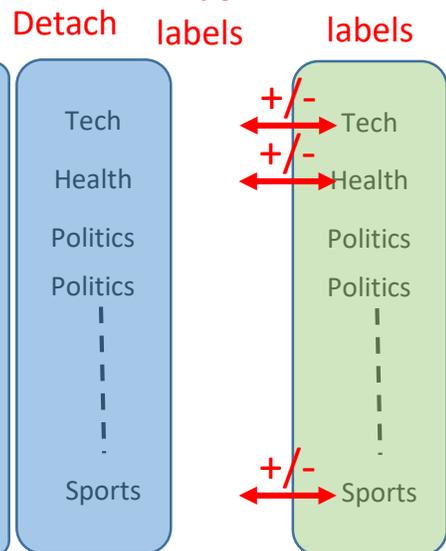
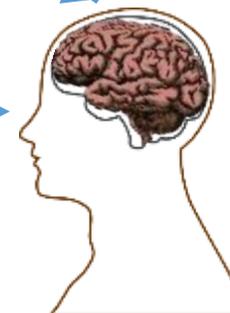


True labels

Train the Model

Test trained model with features

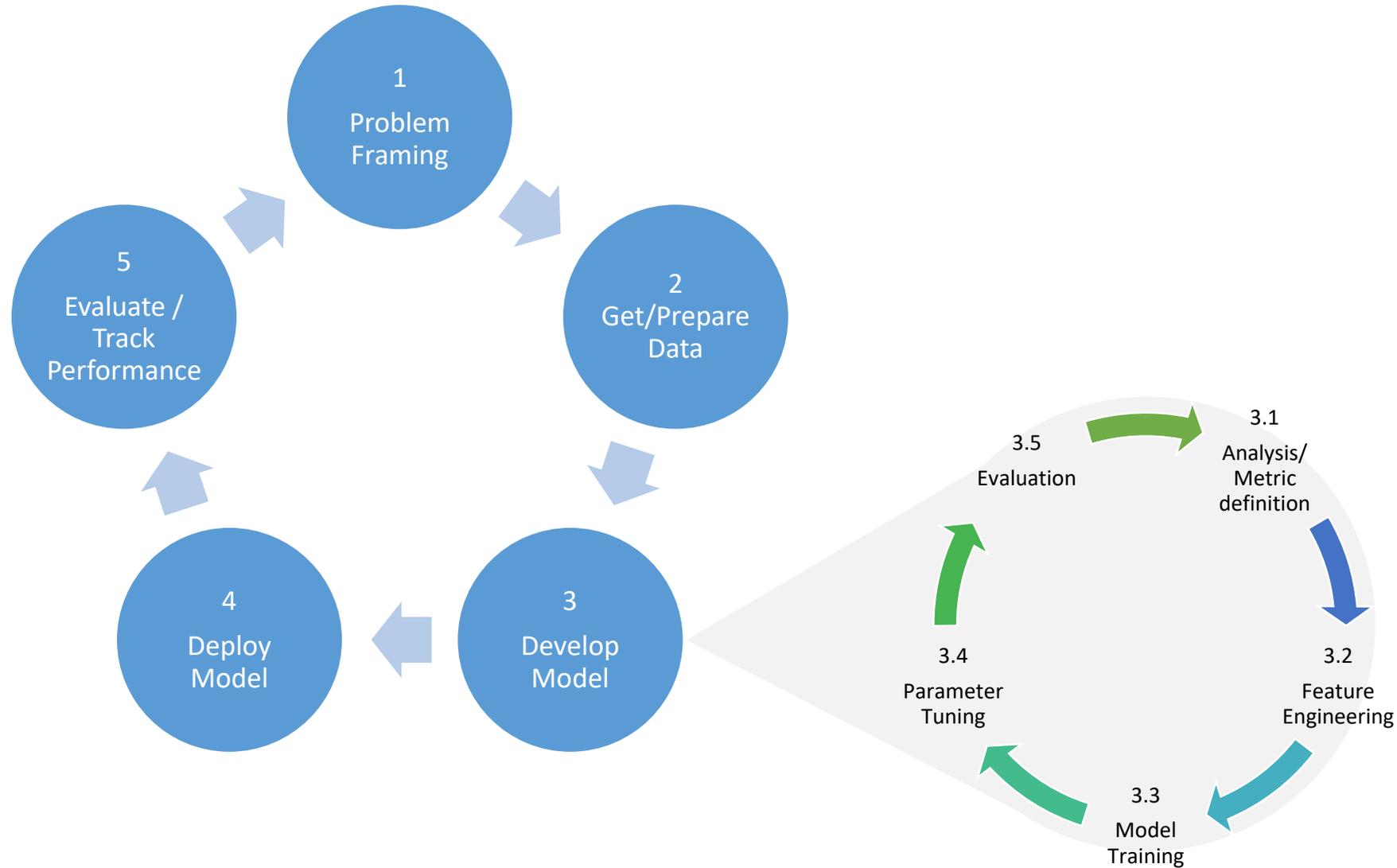
Compare prediction with true labels



Model's Performance

Difference between "True Labels" and "Predicted Labels"

Steps to Build a Machine Learning Solution

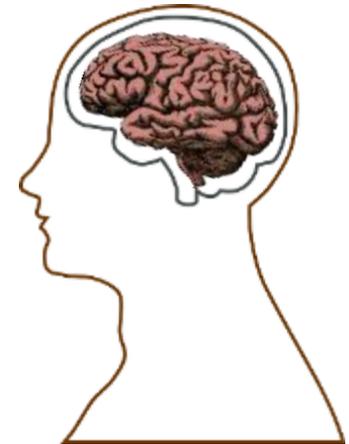
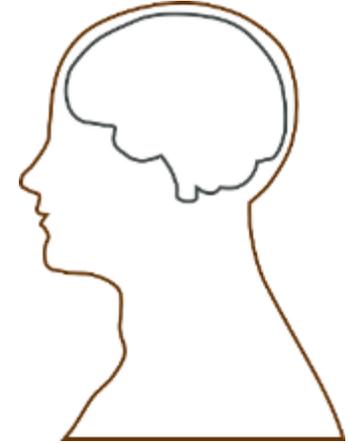


Example use cases

 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 efficiency
 Investment modelling	 Sales lead scoring	 Product recommendation	 Smart buildings
 Fraud detection	 Marketing mix optimization		 Predictive maintenance
 Risk management			 Supply chain optimization

Machine Learning Algorithms

- ML Algorithm defines how your **model** will react
- Which Algorithm to use? Depends on:
 - Data Quality
 - Data Size
 - What you want to predict
 - Time constraint
 - Computation power
 - Memory limits



Machine Learning Algorithms

You can develop solutions by using

- Custom algorithms written in R | Python
- Ready to use ML services from data market
- Existing algorithms

Machine Learning Algorithms

Two major category of algorithms

- Supervised
- Unsupervised

Most commonly used machine learning algorithms are **supervised** (requires **labels**)

- **Supervised** learning examples

- This customer will like *coffee*
- This network traffic indicates a denial of service attack

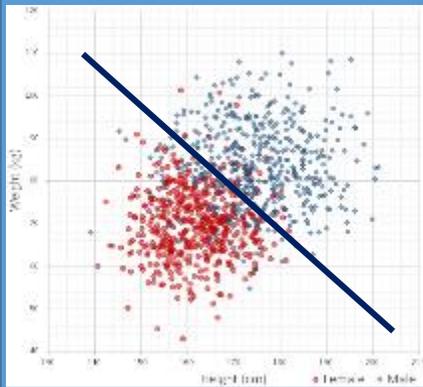
- **Unsupervised** learning examples

- These customers are similar
- This network traffic is unusual

Common Classes of Algorithms

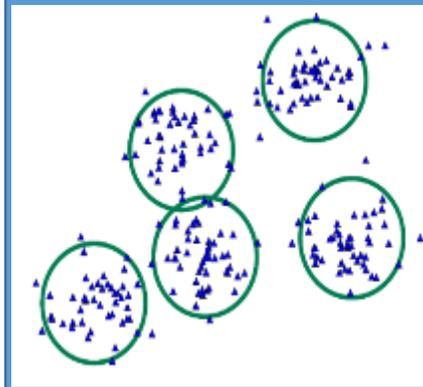
(Supervised | Unsupervised)

Classification



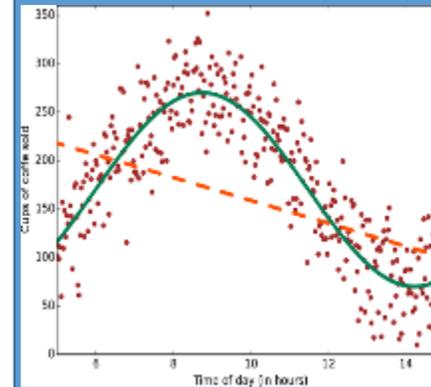
Supervised

Clustering



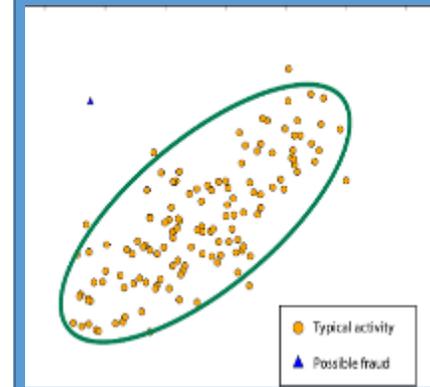
UnSupervised

Regression



Supervised

Anomaly Detection



Supervised

Why you need to know these algorithms?

- If you want to answer a **YES|NO** question, it is **classification**
- If you want to predict a **numerical value**, it is **regression**
- If you want to group data into similar observations, it is **clustering**
- If you want to recommend an item, it is **recommender system**
- If you want to find anomalies in a group, it is **anomaly detection**

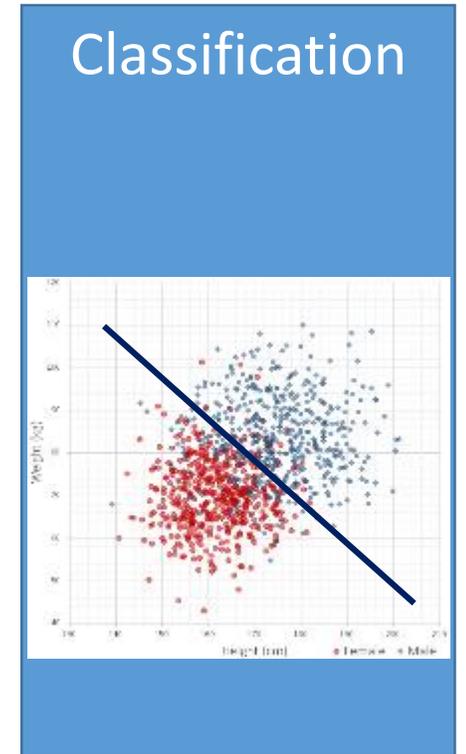
and many other ML algorithms for specific problem

Classification

Scenarios:

- Which customer are more likely to buy, stay, leave (churn analysis)
- Which transactions | actions are fraudulent
- Which quotes are more likely to become orders
- Recognition of patterns: speech, speaker, image, movement, etc.

Algorithms: Boosted Decision Tree, Decision Forest, Decision Jungle, Logistic Regression, SVM, ANN, etc.

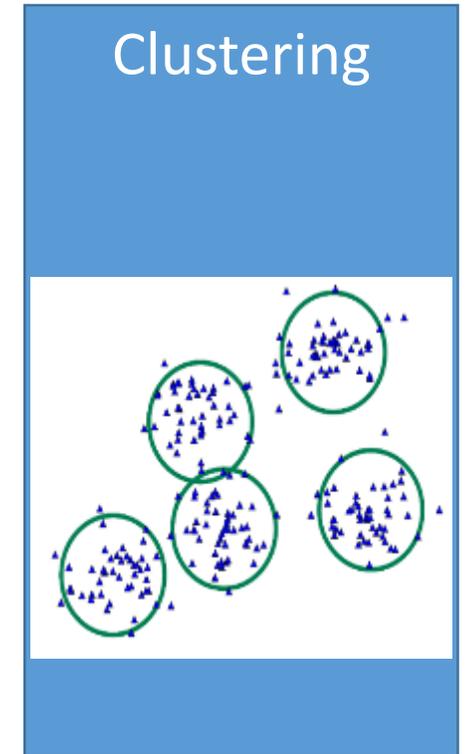


Clustering

Scenarios:

- Customer segmentation: divide a customer base into groups of individuals that are similar in specific ways relevant to marketing, such as age, gender, interests, spending habits, etc.
- Market segmentation
- Quantization of all sorts, such as, data compression, color reduction, etc.
- Pattern recognition

Algorithms: K-means

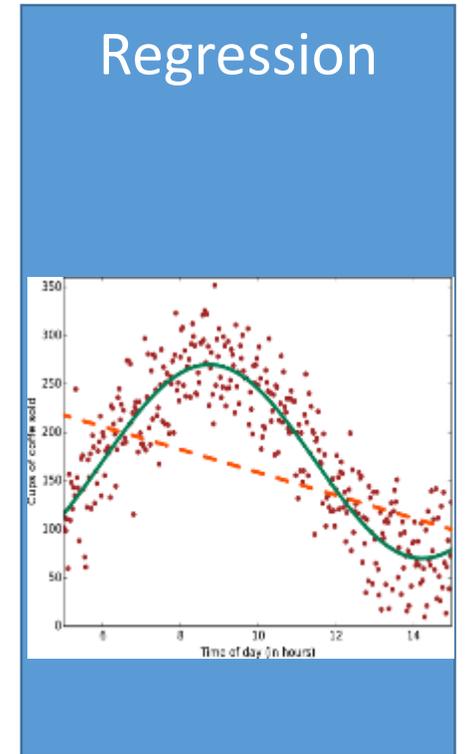


Regression

Scenarios:

- Stock prices prediction
- Sales forecasts
- Premiums on insurance based on different factors
- Quality control: number of complaints over time based on product specs, utilization, etc.
- Workforce prediction
- Workload prediction

Algorithms: Bayesian Linear, Linear Regression, Ordinal Regression, ANN, Boosted Decision Tree, Decision Forest



Regression versus Classification

Does your customer want to predict | estimate a number (regression) or apply a label | categorize (classification)?

- Regression problems

- Estimate household power consumption
- Estimate customer's income



- Classification problems

- Power station will | will not meet demand
- Customer will respond to advertising



Binary versus Multiclass Classification

Does your customer want a yes|no answer?

- Binary examples

- click prediction
- yes|no
- over|under
- win|loss



- Multiclass examples

- kind of tree
- kind of network attack
- type of heart disease



DEMO

Machine Learning Basics Infographic

References

- Free e-book “Azure Machine Learning”
- <https://mva.microsoft.com/ebooks#9780735698178>

- Azure Machine Learning documentation
- <https://azure.microsoft.com/en-us/documentation/services/machine-learning/>

- Data Science and Machine Learning Essentials
- www.edx.org

- Azure ML HOL (GitHub):
- <https://github.com/Azure-Readiness/hol-azure-machine-learning/>

HOL Document

- Access Azure HOL Doc: <https://aka.ms/azuremlhol>

Thank you

- Check out my blog for Azure ML articles: <http://mostafa.rocks>
- Follow me on Twitter: @MostafaElzoghbi