



CloudTech

Marrakesh 2016

Introducing Microsoft Azure Machine Learning

Claudio Rossi



Agenda:

Overview of Machine Learning

Overview of Azure ML

Guided Demo

Hands-on Lab

CloudTech

Marrakesh 2016

Overview of Machine Learning

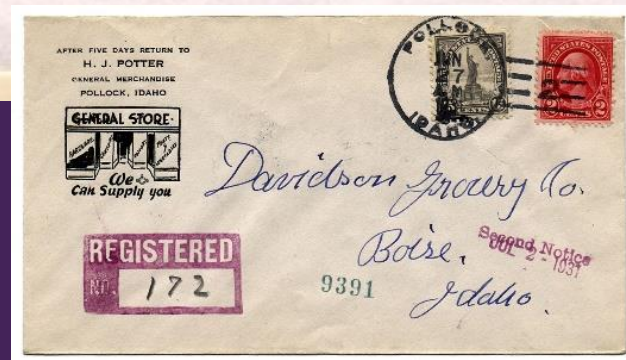
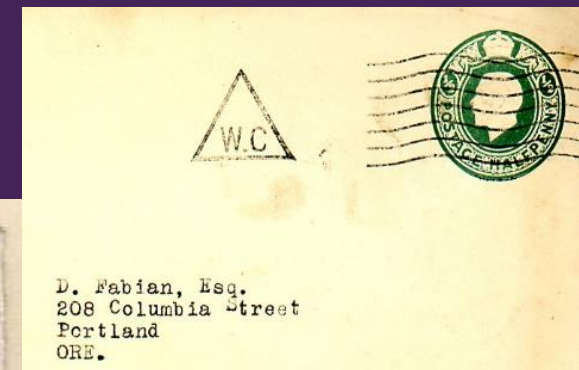
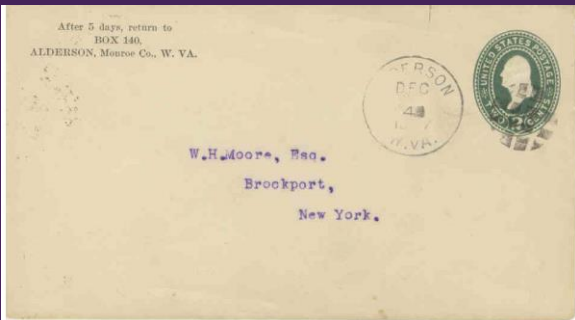


I believe over the next decade computing will become even more ubiquitous and **intelligence will become ambient**...This will be made possible by an ever-growing network of connected devices, incredible computing capacity from the cloud, insights from big data, and **intelligence from machine learning**.

Satya Nadella
CEO @ Microsoft

"If you invent a
breakthrough in
Artificial
Intelligence, so
machines can learn,
that is worth 10
Microsofts"







Accurate digit
classifier

2

1	1	5	4	3
7	5	3	5	3
5	5	9	0	6
3	5	2	0	0

Training examples

1	1	5	4	3
7	5	3	5	3
5	5	9	0	6
3	5	2	0	0

Training labels



Machine learning system

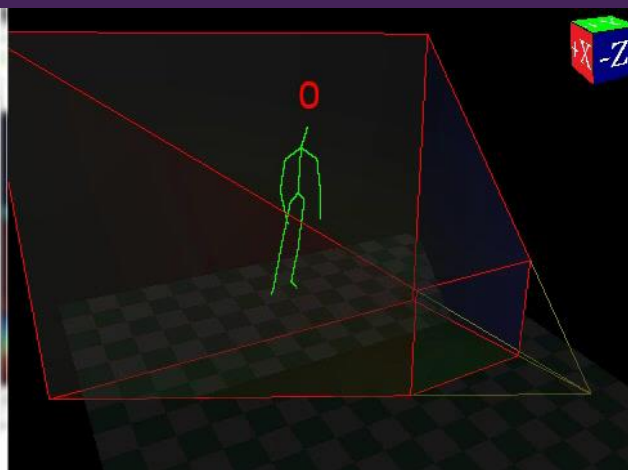




training data (expensive)



synthetic training data (cheaper)



IVL system

Machine Learning (ML)

“The goal of machine learning is to build computer systems that can adapt and learn from their experience.”

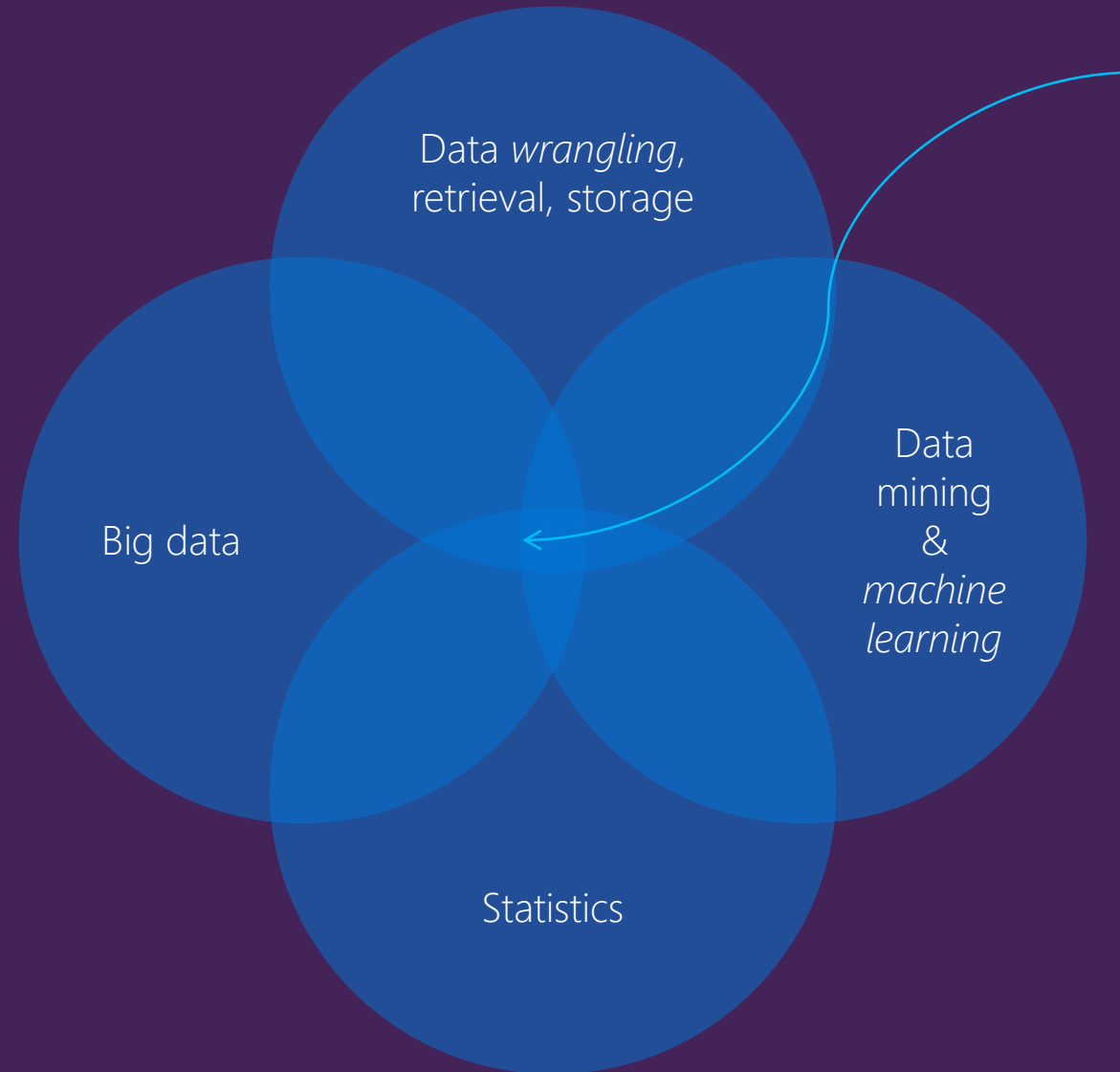
– Tom Dietterich

ML is Data Science?

Data science

- is the study of the generalizable extraction of [knowledge](#) from [data](#) *(Wikipedia)*
- is getting predictive and/or actionable insight from data *(Neil Raden)*
- involves extracting, creating, and processing data to turn it into business value. – Vincent Granville *(Developing Analytic Talent: Becoming a Data Scientist)*

Where Data Science lies?



Data
science

Why learn it?

1. Learn it when you can't code it
(e.g. Recognizing Speech/image/gestures)
2. Learn it when you can't scale it
(e.g. Recommendations, Spam & Fraud detection)
3. Learn it when you have to adapt/personalize
(e.g. Predictive typing)
4. Learn it when you can't track it
(e.g. AI gaming, robot control)

ML Cycle

1. **Select** & initialise a model
2. **Train** model (process cases)
3. **Validate** model

...by **scoring** (making predictions) a test data set and **evaluating** the results

4. Use it: **Explore** or **Deploy**

...visualise and study

...deploy as a (web) service

5. **Update** and **revalidate**

Algorithm Classes

Supervised

Ground truth known in the data set (regression, classifiers, ...)

Un-Supervised

Ground truth not known (clustering, dimensionality reduction)

Classifiers

Assign a category to each item

Clustering

Discover natural groupings of cases

Regression

Predict numerical outcomes

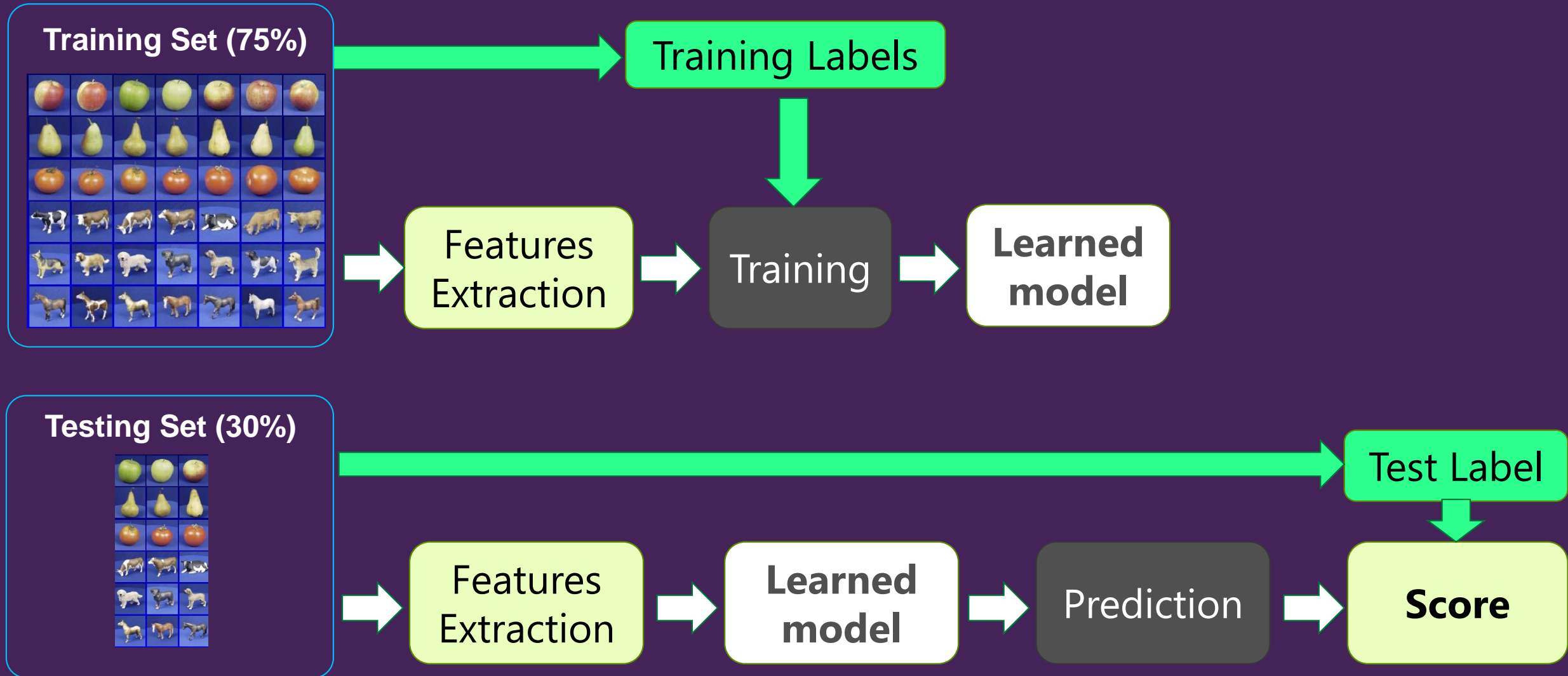
Recommenders

Explore associations between cases

Ensembles

mix them up

Training & Testing a Model



Scoring a Model

➔ Regression: Mean Square Error (MSE)

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

➔ Classification: Confusion Matrix

➔ Accuracy = (TP + TN) / TOTAL

➔ Precision = TP / (TP + FP)

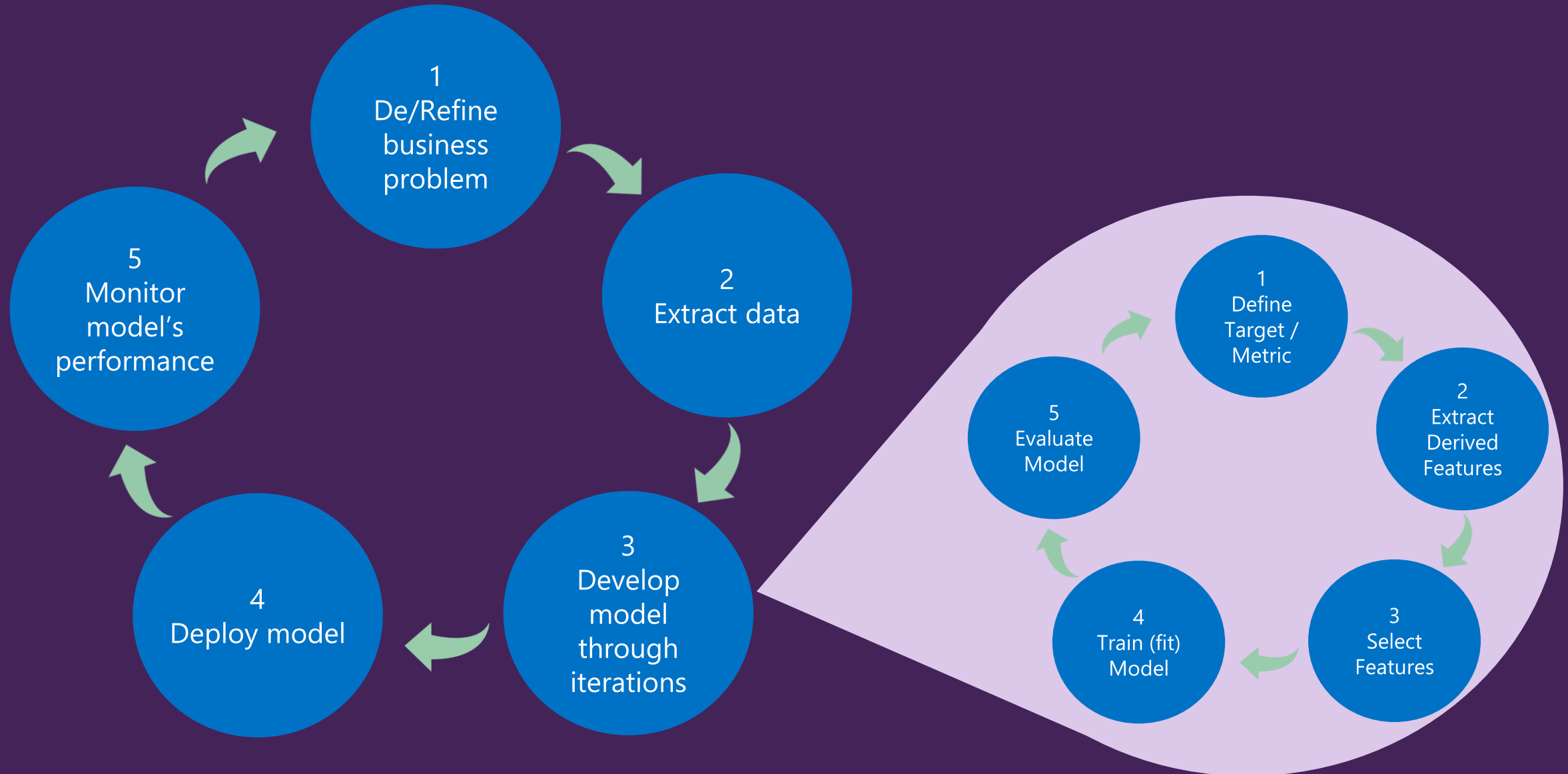
➔ Specificity = TN / RN

➔ ...

➔ Full @ https://en.wikipedia.org/wiki/Sensitivity_and_specificity

<u>Confusion Matrix</u>		<i>Predicted</i>	
		TRUE	FALSE
<i>Real</i>	TRUE	True Positive (TP)	False Negative (FN)
	FALSE	False Positive (FP)	True Negative (TN)

Steps to build a Machine Learning Solution



Feature engineering is the **key**...

Feature engineering: when you use your knowledge about the data to create fields that make machine learning algorithms work better.

It is easily the most important factor in determining the success of a machine learning project

How does one engineer a good feature? Rule of thumb is to try to design features where the likelihood of a certain class goes up monotonically with the value of the field.

Great things happen in machine learning when human and machine work together, **combining a person's knowledge of how to create relevant features from the data with the machine's talent for optimization..**

More data beats a cleverer algorithm...

More data wins. There's increasingly good evidence that, in a lot of problems, very simple machine learning techniques can be levered into incredibly powerful classifiers with the addition of loads of data.

Computer algorithms trying to learn models have only a relatively few tricks they can do efficiently, and many of them are not so very different. Performance differences between algorithms are typically not large.

Thus, if you want better performances:

1. Engineer better features
2. Get your hands on more high-quality data

Tools & Salaries

SQL (any): #1 data science tool

Love it, or...use it.

R rulez

Getting very important

SQL Server!

SAS: most likely BI to be discontinued
by Gartner clients (Oct 2014 MQ)

Sometimes



Chart from "2014 Data Science Salary Survey" (ISBN 978-1-491-91842-5)

© 2015 O'Reilly Media, used with permission. Arrows mine.

For more info, and great titles on data science, visit oreilly.com