Introduction to Machine Learning

Manoj Kumar Chinnakotla

Senior Applied Scientist

AI and Research

Microsoft, India

Microsoft

Welcome to the field of Machine Learning! ☺

"A breakthrough in Machine Learning would be worth ten Microsofts"

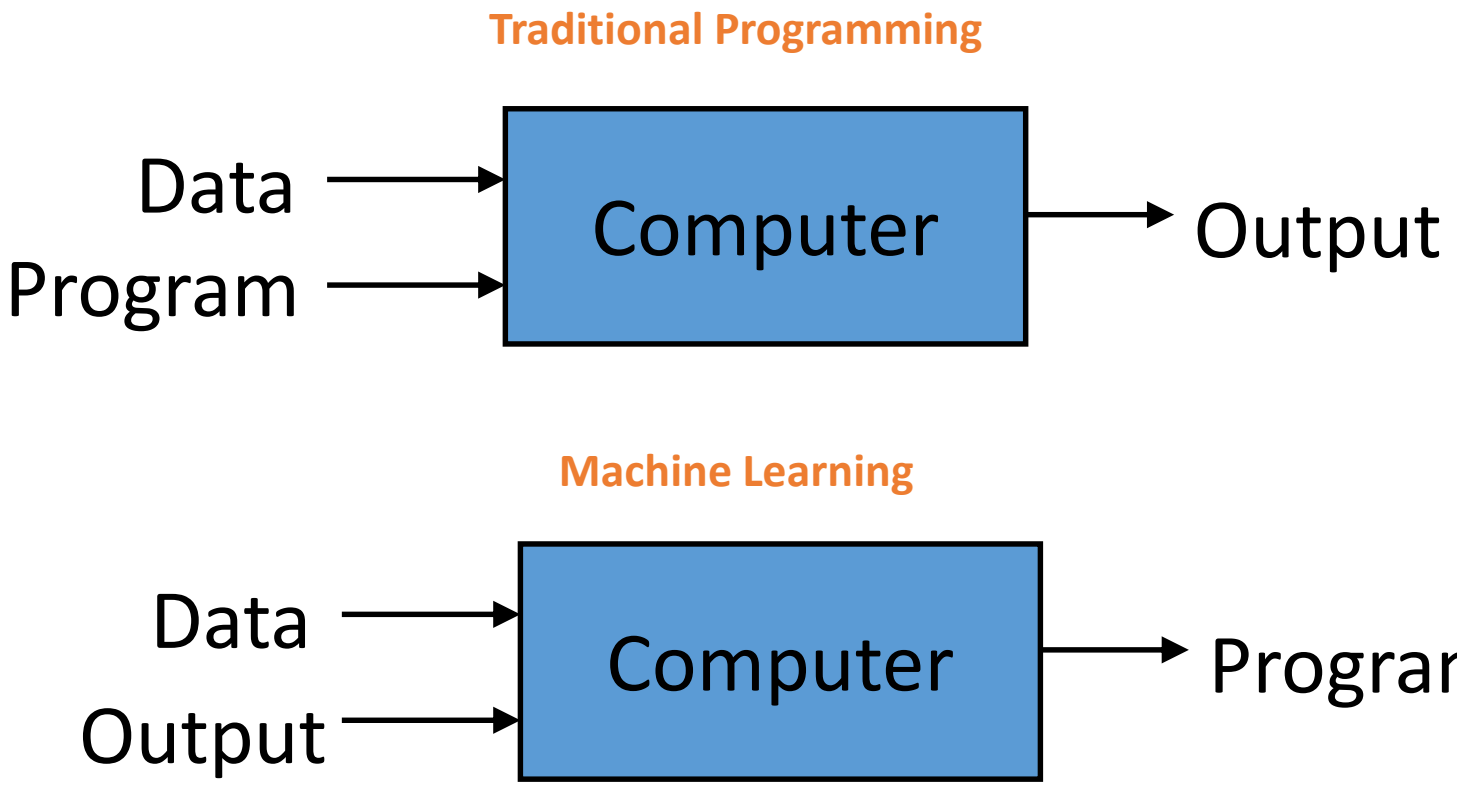
– Bill Gates

"I think Machine Learning may be the single most important technology for the next generation of innovation"

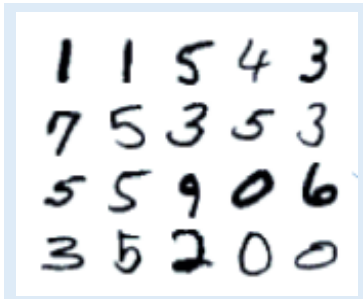
– SteveB (MS Company Meeting, 2013)

What is Machine Learning (ML)?

- Automating automation
- Getting computers to program themselves
- Whenever writing software is the bottleneck
  - Let the data do the work instead!



Handwritten Digit Recognition



Training examples

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

Training labels

Accurate digit classifier

When to use ML?

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)

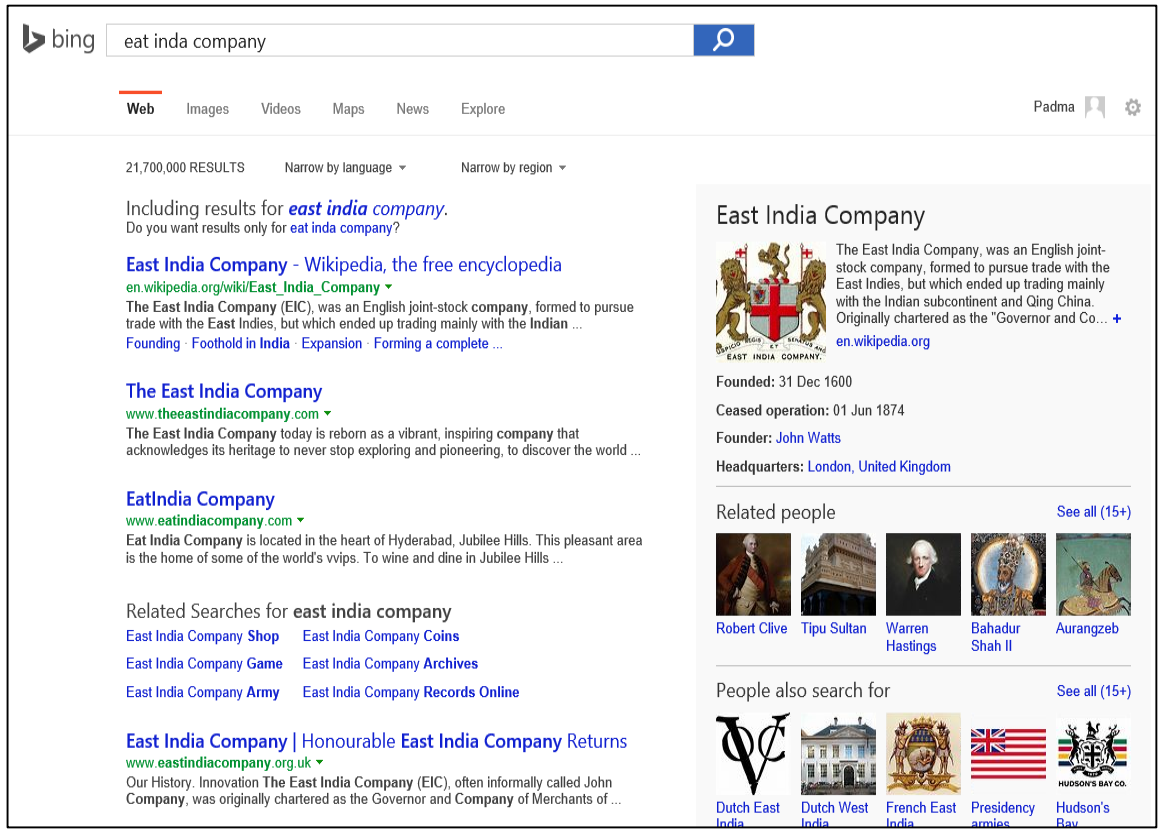
Sounds like Magic?

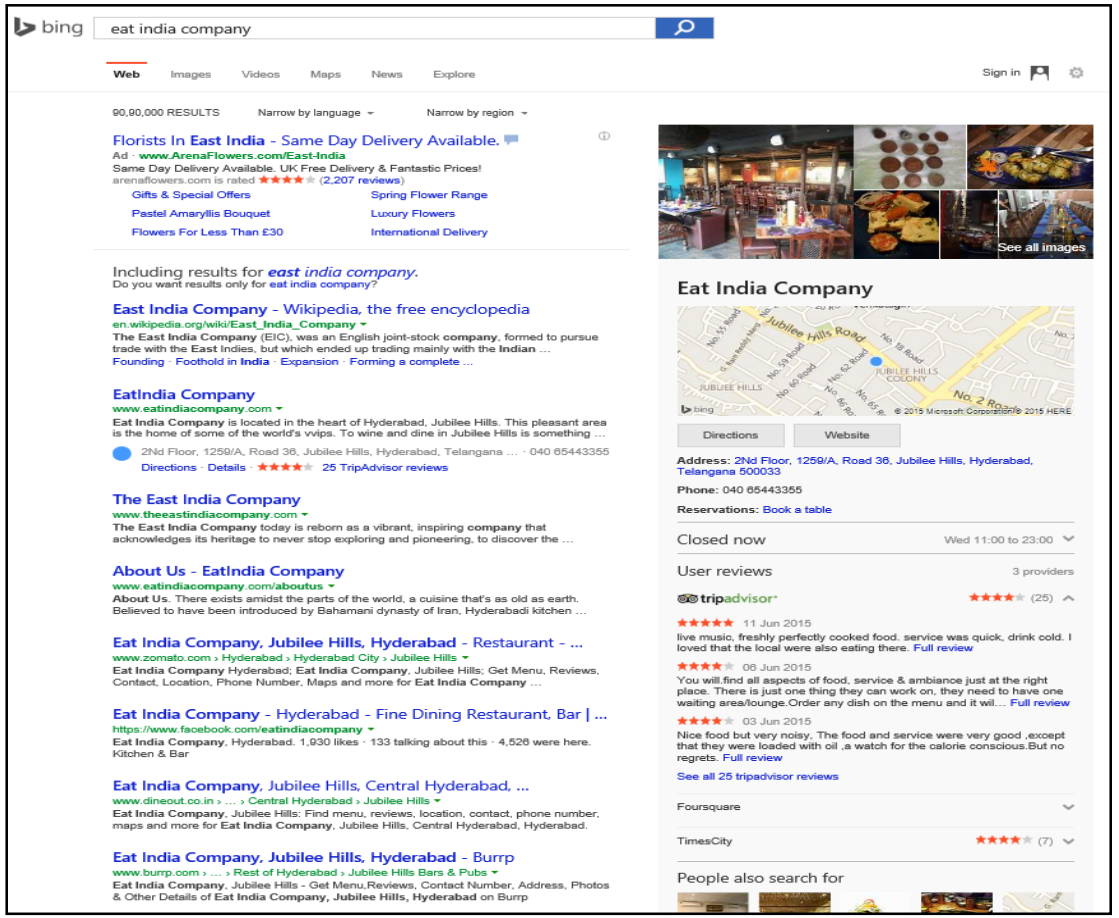
Not really!, it's more like gardening

- **Seeds** = Algorithms
- **Nutrients** = Data
- **Gardener** = You
- **Plants** = Programs

Where is ML applied today?

- Information search, extraction on the web

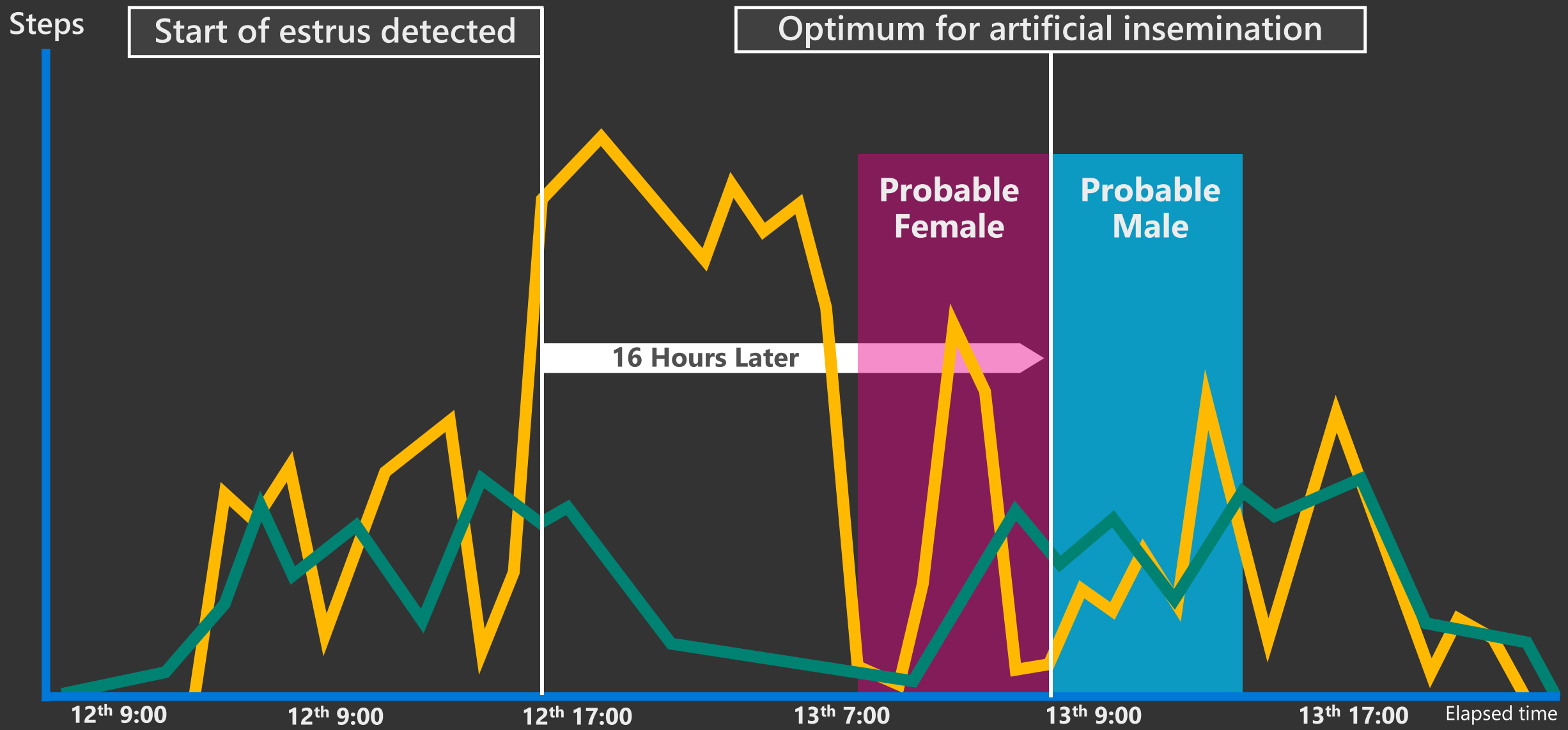




Where is ML applied today?

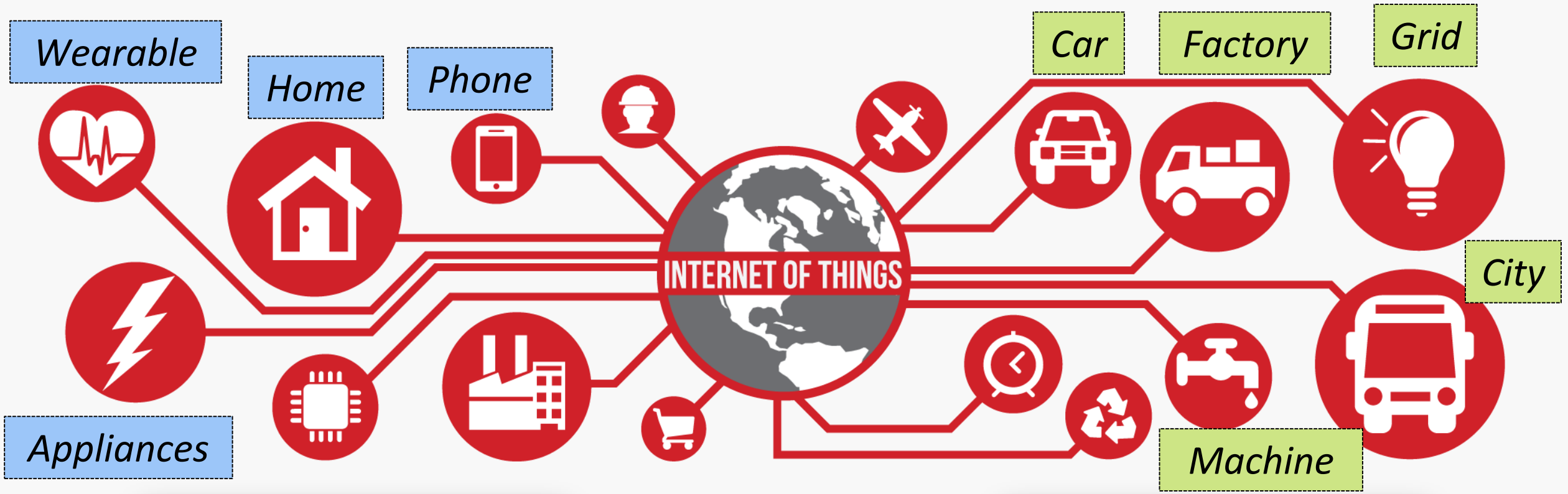
- Speech Recognition
- Natural Language Understanding/Processing

The Connected Cow



Source: http://www.fujitsu.com/global/about/resources/news/press-releases/2013/1015-01.html

Internet of Things (IOT)



CONSUMER

Internet of Things

INDUSTRIAL

Internet of Things

THE VALUE FROM INTELLIGENT APPS IS HUGE

AVIATION	POWER GEN	HEALTH CARE	RAIL: FREIGHT
$30B	**$66B**	**$63B**	**$27B**
1% fuel savings	1% fuel savings	1% reduction inefficiency	1% reduction inefficiency




The Power of 1 %

Source: GE - Industrial Internet: Pushing the Boundaries of Minds and Machines

Overview of ML Paradigms

Classification Task

Tid	Home Owner (categorical)	Marital Status (categorical)	Taxable Income (continuous)	Cheat (class)
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Training Data

Given historical data on loan payments, learn a model to predict a likely **"defaulter"**?

$$Y = \boldsymbol{f}(A_1, A_2, .. A_n)$$

Learn function $\boldsymbol{f}$ where $A_1, A_2, .. A_n$ are attributes

The label to be predicted is ***"Discrete"***
In the above case, $Y \in \{\text{"Yes"}, \text{"}No\text{"}\}$

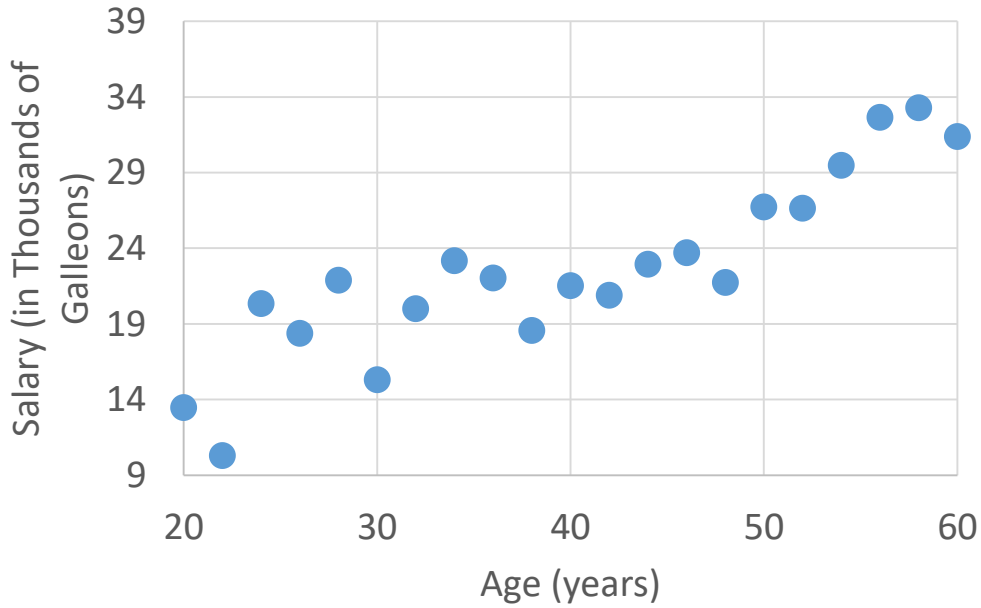
Test Data

Home Owner	Marital Status	Taxable Income	Cheat
No	Married	80K	?

$= \boldsymbol{f}(A_1, A_2, .. A_n)$

Regression Task

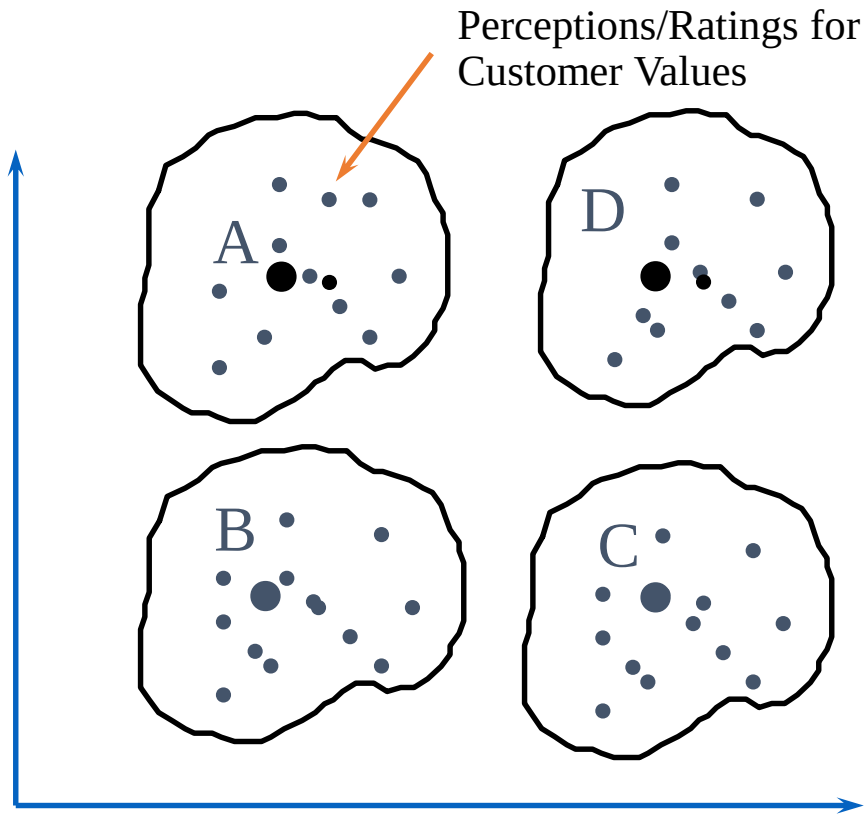
Age	Salary (in Galleons)
20	15416
22	17140
24	18292
26	12723
28	18698
30	18164
32	23577
34	20154
36	21957
38	22244
40	23300
42	20765
44	22106
46	25414
48	22295
56	23494
58	32170
60	30461



- In general, salary increases with age.
- Can we quantify this relationship? How strong is this dependence of salary in age?
- Given an age (say 31.5), can we predict the salary?

Clustering

Market Segmentation Problem – Carpet Fiber



A,B,C,D:
Location of segment centers.

Typical Members:

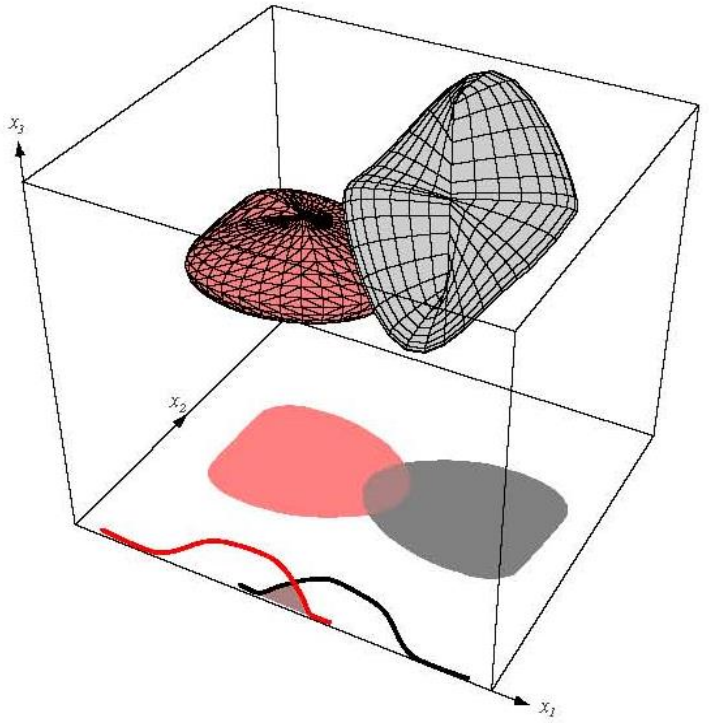
A: Schools
B: Light commercial
C: Indoor/outdoor Carpeting
D: Health clubs

The Task
Automatically Discover the underlying structure in the data

The Challenge
No Training Data available

Dimensionality Reduction

- If data x lies in high dimensional space, then an enormous amount of data is required to learn distributions or decision rules.
- Example: 50 dimensions. Each dimension has 2 levels. This gives a total of 2^{50}cells. But the no. of data samples will be far less.
- The Main Idea
  - Reduce the dimensionality of the space
  - Project the d-dimensional points in a k-dim space
  - k << d
  - distances are preserved as well as possible
- Solve the problem in low dimensions



Sequence Learning

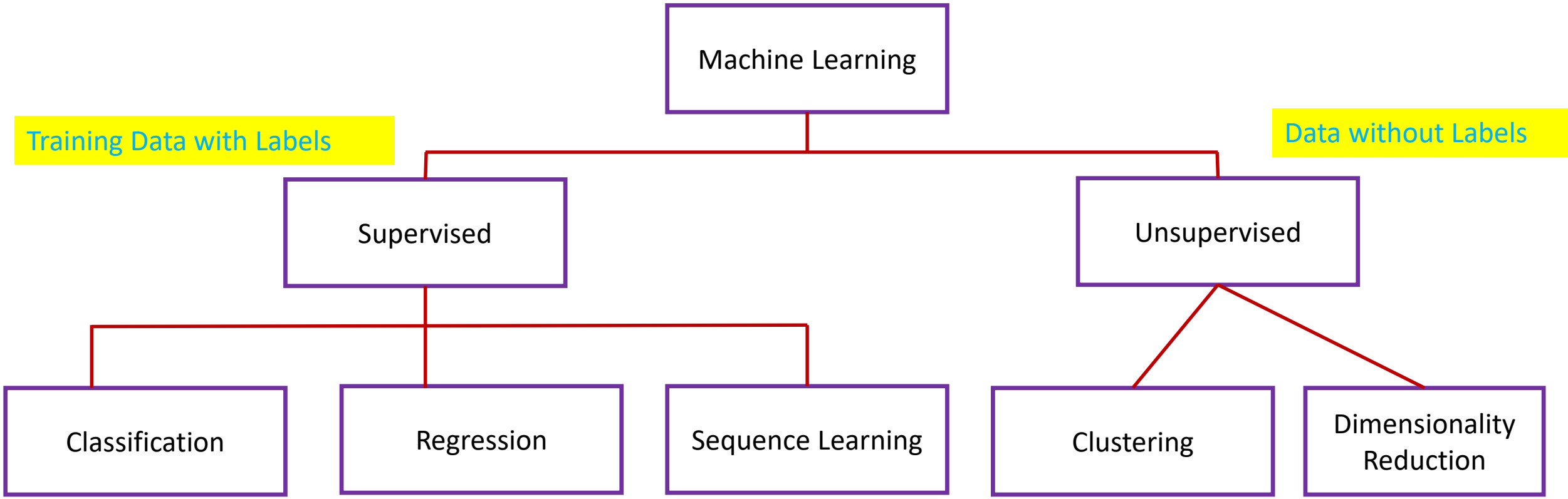
Parts of Speech Tagging

DT NN VBD IN DT NN . ← States (S)

The cat sat on the mat . ← Observations (O)

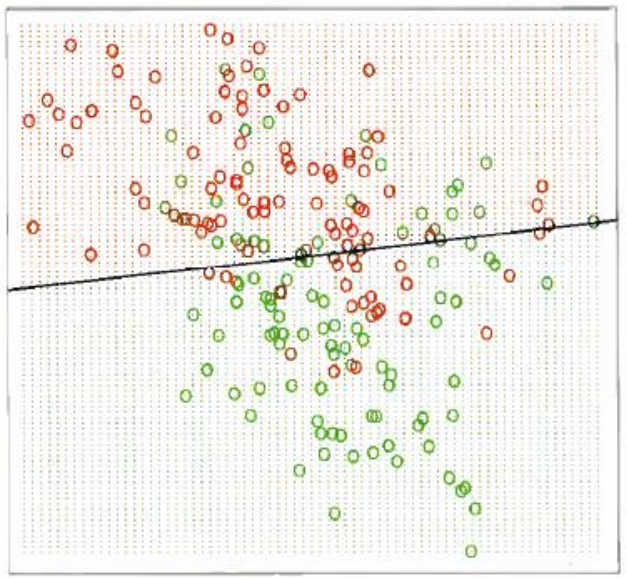
NN=noun, VBD=verb (past tense),
DT=determiner, IN=preposition or subordinating conjunction

Basic Classification of ML Problems



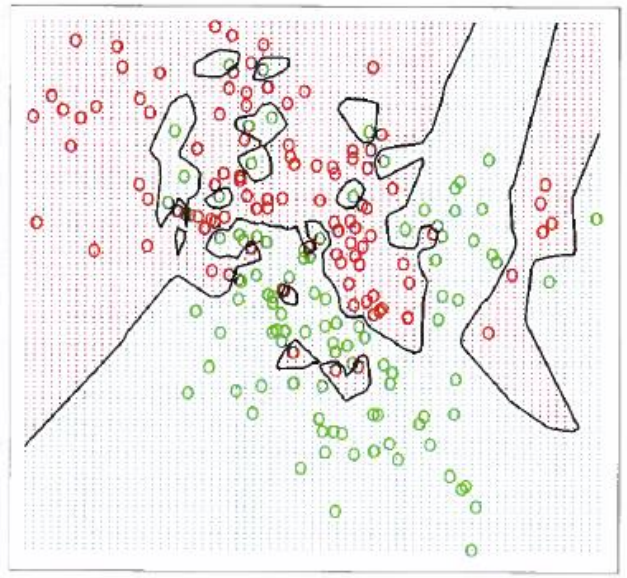
Its Generalization That COUNTS

- Generalize beyond examples in training set



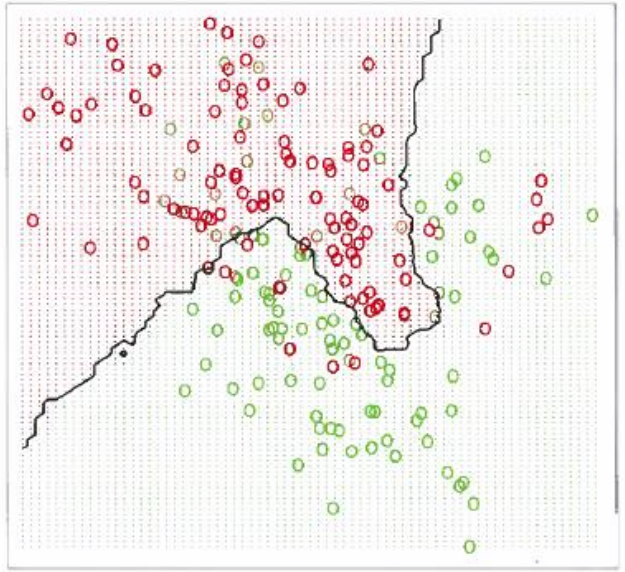
Under-fitting!

Insufficient flexibility/power in the model



Over-fitting!

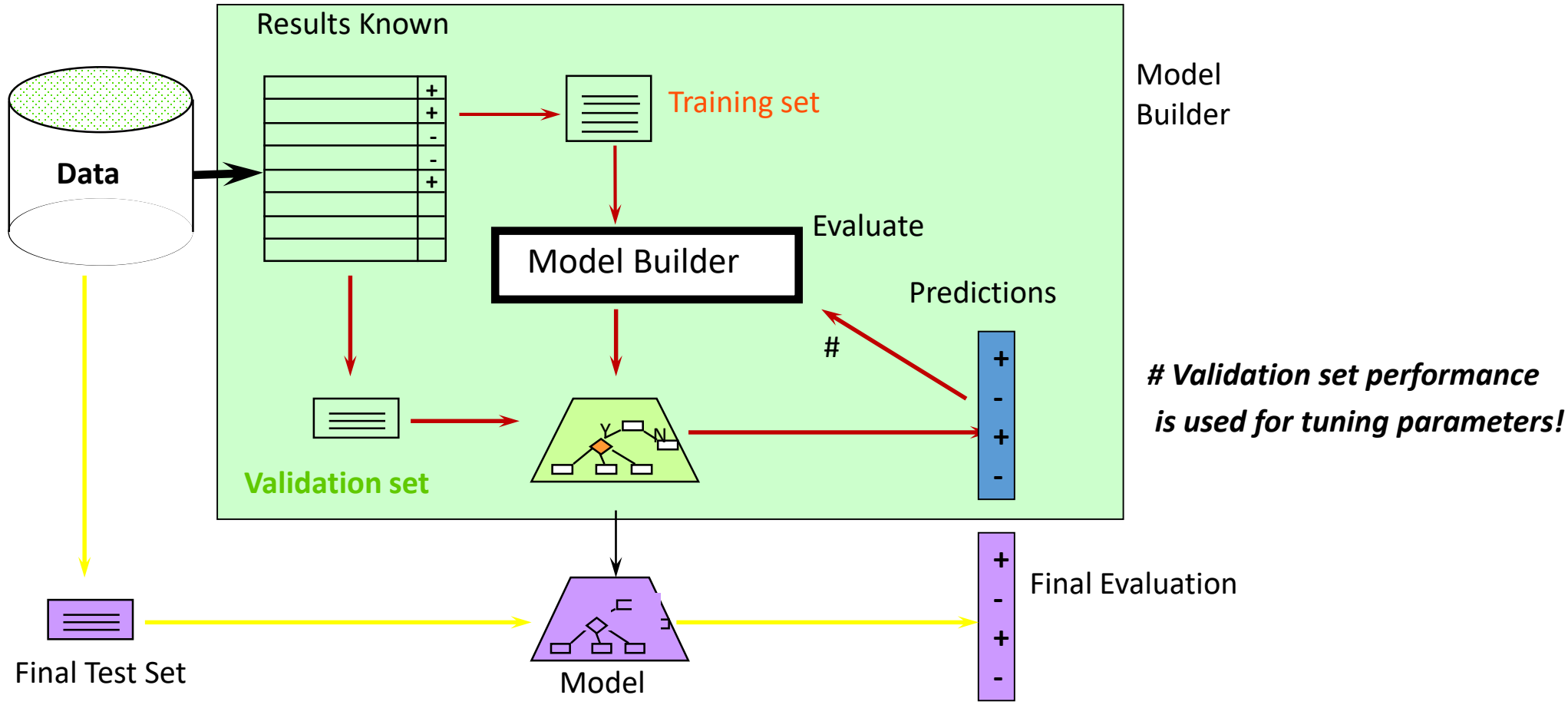
Too much specialization on training data



Better!

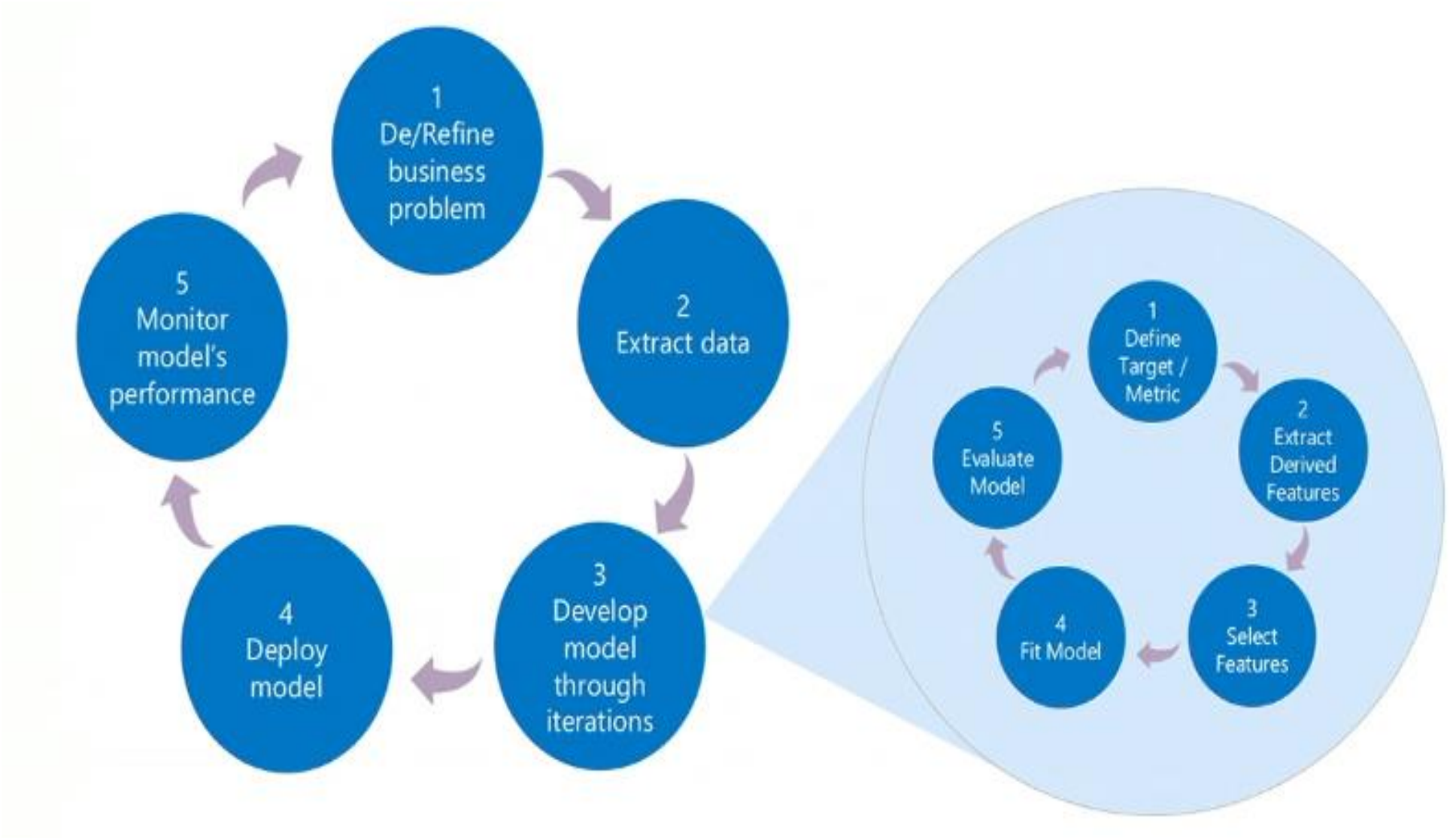
Is this the best or can we do better ?

The Model Building Process



The test data shouldn't be used for any parameter tuning!

The Model Building Process



Measurement

- Why?
  - To check the quality of Model
- Each task will have it’s own measurement story
- For Classification
  - Precision
  - Recall
  - F Measure

Confusion Matrix and Precision/Recall

Predicted

Actual

	Positive	Negative
Positive	A	B
Negative	C	D

- Total Positive Instances =
  - A + B
- Total Positive Predicted Instances by Model =
  - A + C
- Precision =
  - A / (A + C)
- Recall =
  - A / (A + B)

Some Practical Aspects of ML

- Feature Engineering is the key step
- Data visualization and representation is crucial in coming up with features
- More data usually beats a clever algorithm

Thanks!