

Introduction



to

DATA SCIENCE

Oct 4th 2021

Fauci Calls Merck's Experimental COVID-19 Pill 'Really Quite Impressive'

The White House chief medical adviser said on Sunday that the preliminary study results released over the weekend for drugmaker Merck's COVID-19 pill are "really quite impressive."

"You have now a small molecule, a drug that can be given orally. And the results of the trial that were just announced yesterday and the day before are really quite impressive. I mean, if you do a statistical significant analysis on it, it's very, very significant, cutting the deaths and hospitalization by 50%," Fauci told Jonathan Karl on ABC's "This Week."

"Importantly, in the placebo versus the drug group — in the drug group, there were zero deaths. In the placebo group, there were eight deaths. ... No matter how you slice that, that's impressive. So we're really looking forward to the implementation of this."

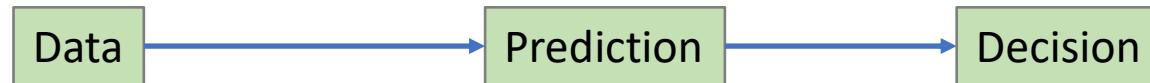
Merck and its partner Ridgeback Biotherapeutics said early results showed patients who received molnupiravir within five days of Covid symptoms had about a 50 percent reduction in the rate of hospitalization and death as patients who received a placebo pill. The study tracked 775 adults with mild-to-moderate Covid who were considered higher risk for severe disease due to health problems such as obesity, diabetes or heart disease.

None had been vaccinated.

Among patients taking molnupiravir, 7.3 percent were either hospitalized or died at the end of 30 days, compared with 14.1 percent of those getting the dummy pill. There were no deaths in the drug group after that time period compared with eight deaths in the placebo group, according to Merck.

Simply put, this course is about

predicting (the future) given data



About me

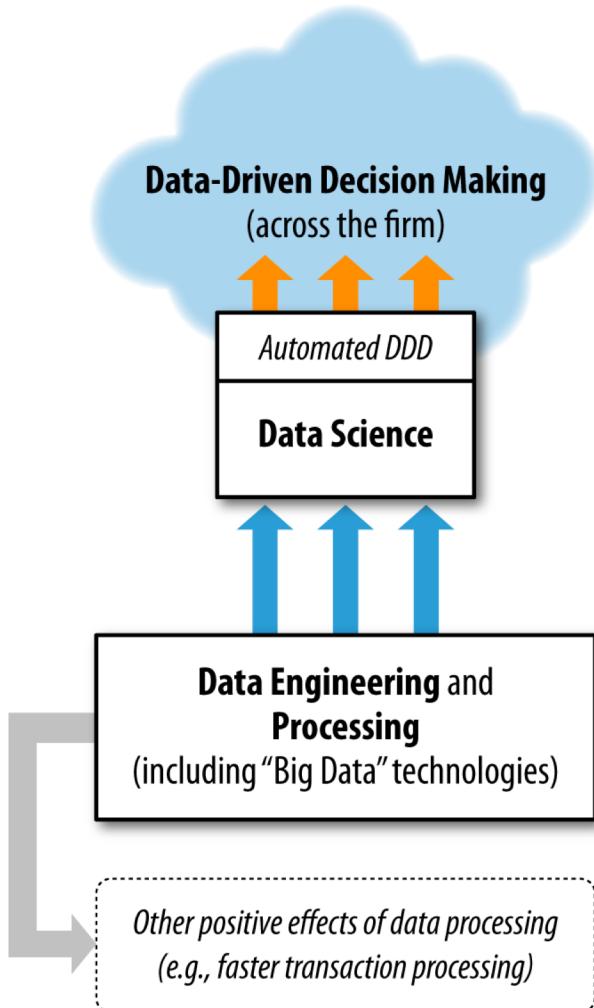
Home	Role
<u>InferLink Corporation</u>	<i>Chief Data Science Officer</i>
<u>Fred Hutch Cancer Center</u>	<i>Head of Data Science</i>
<u>Keck School of Medicine @ USC</u>	<i>Faculty</i>
<u>University of California Irvine</u>	<i>Faculty</i>
<u>DropThought Inc.</u>	<i>Consulting Chief Data Scientist</i>
<u>Xpresso.ai</u>	<i>Consulting Chief Data Scientist</i>
<u>NASA Ames Research Center</u>	<i>Research Scientist</i>
<u>USC Information Sciences Institute</u>	<i>Grad Student</i>

About you ...

Data Science, Big Data, Data Analytics

- Unprecedented amount of DATA
- Competitive advantage
- And

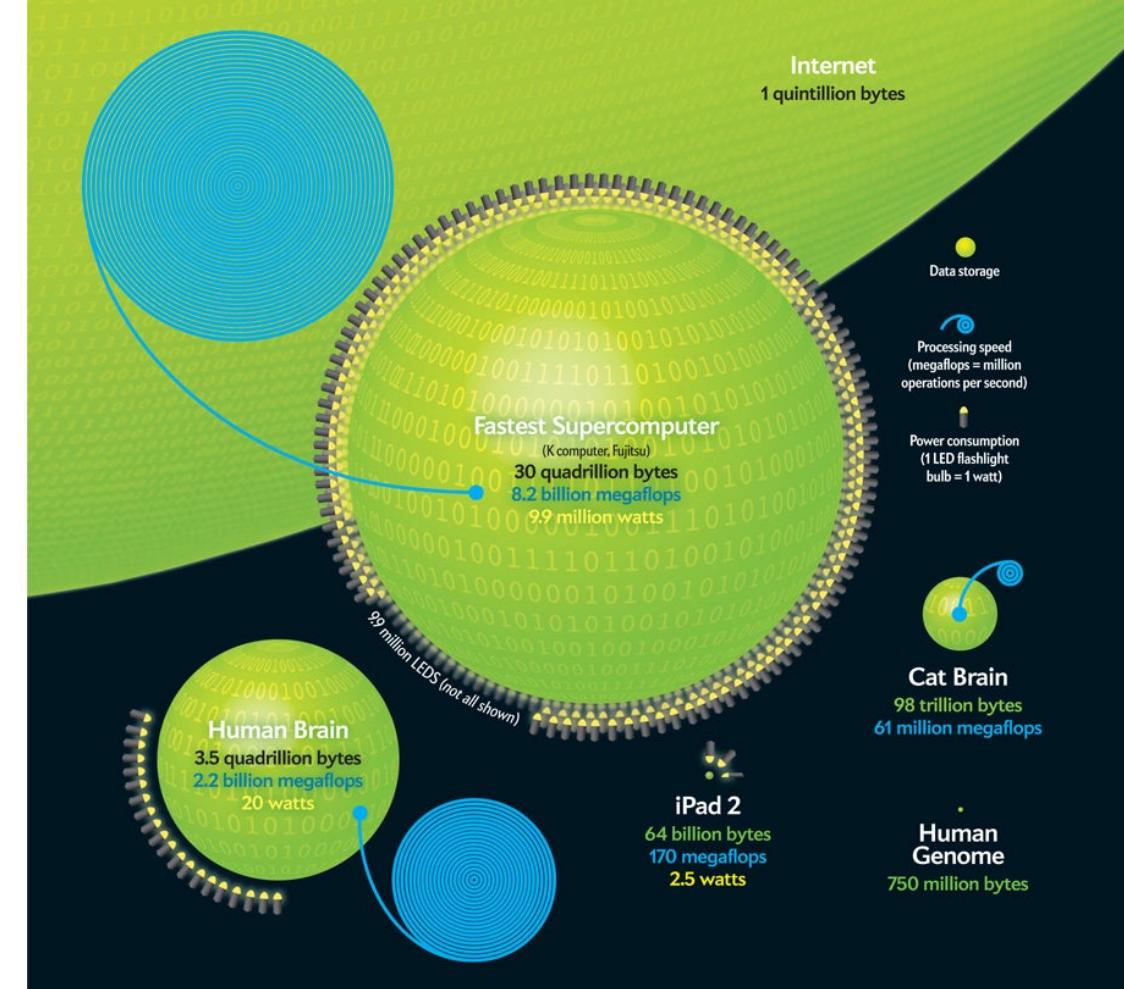
Data Driven Decision-making (DDD)



- **DDD** benefits objectively demonstrated
 - (even) one standard deviation → 4-6% productivity increase
- Automation vs augmentation

Big-Data in Data Science

- One standard deviation → 1-3% change in productivity



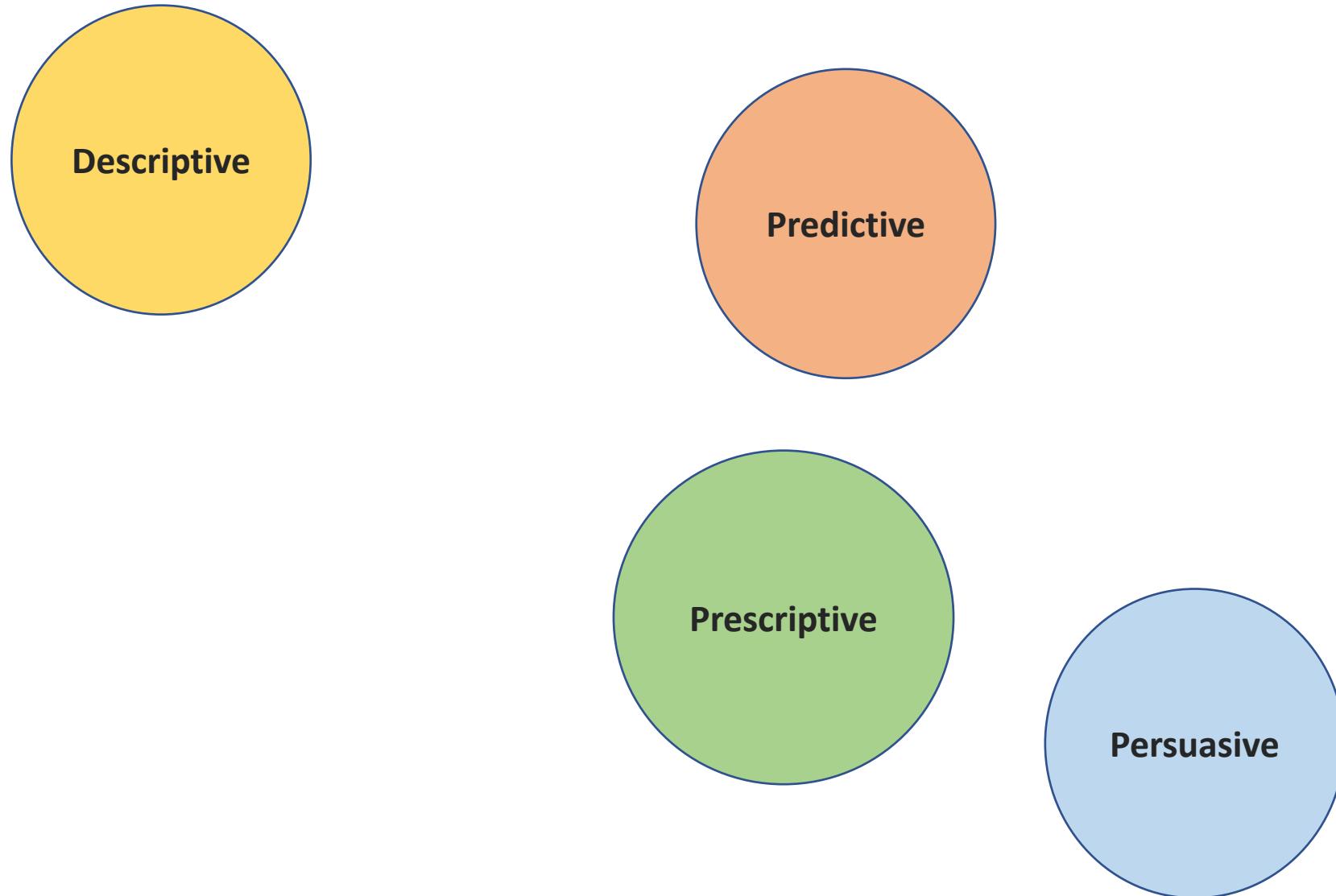
85% of big data projects fail !

Data, and Data Science as an **Asset**

- The value of data
- Data is a plenty, and it is not
- Who has data
- How does one get data

Data Analytic Thinking

- Understand data science, even if you will never do it yourself
- Data science reaches multiple business units
- Close interaction between business and data science



Objectives

How should I think about the problem ?

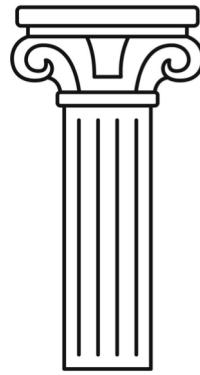
How should I organize my data ?

How do I do this ? (what tools etc.)

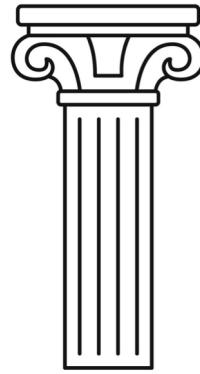
How well have we done here ?

Four Fundamental Pillars of Data Science

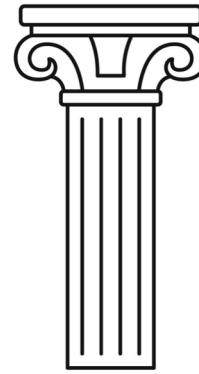
Concepts



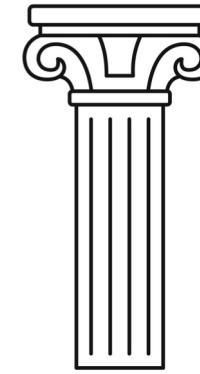
Coding



Math & Stats



AI: Machine Learning



Resources

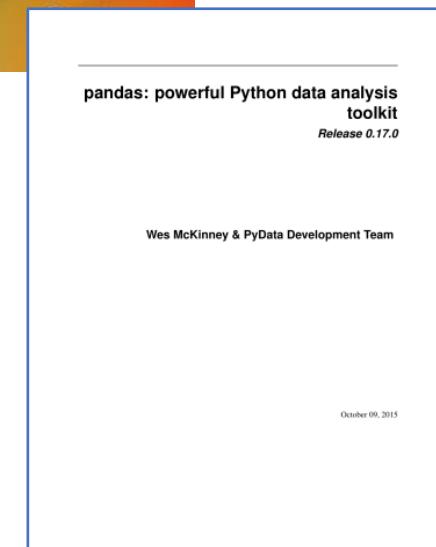
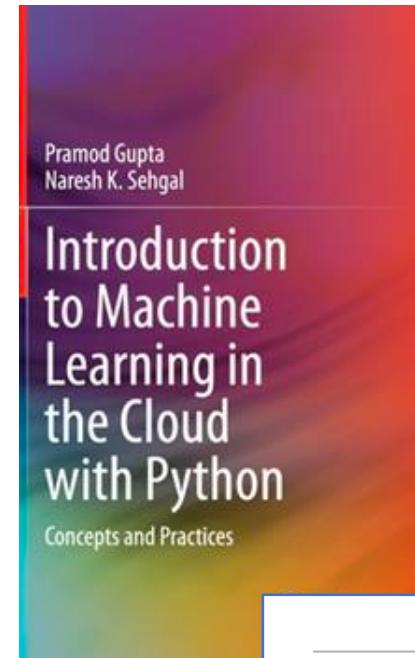
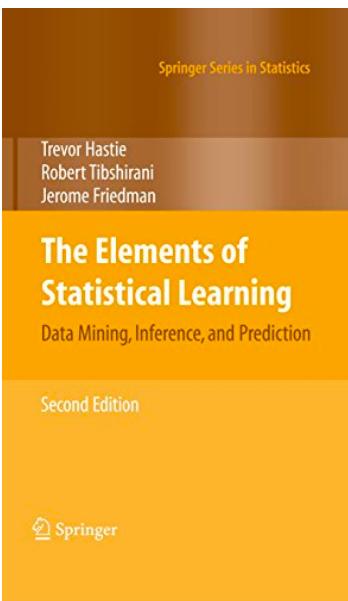
"A must-read resource for anyone who is serious about embracing the opportunity of big data." —Craig Vaughan, Global Vice President, SAP

Data Science for Business

What You Need to Know
About Data Mining and
Data-Analytic Thinking



Foster Provost & Tom Fawcett



This is *all* online

Business Problems to Data Science

- “Decompose”
- Model
- A unit can be mapped to a **familiar** problem
- Few techniques

The Few **Algorithms**

Classification

Regression

Similarity Matching

Clustering

Co-occurrence grouping

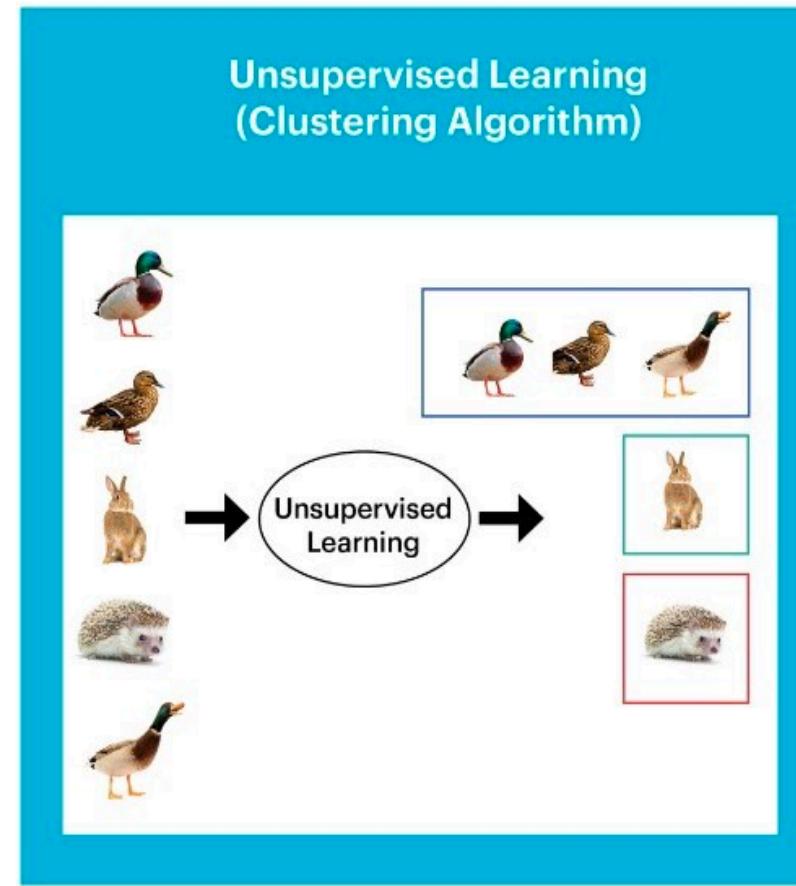
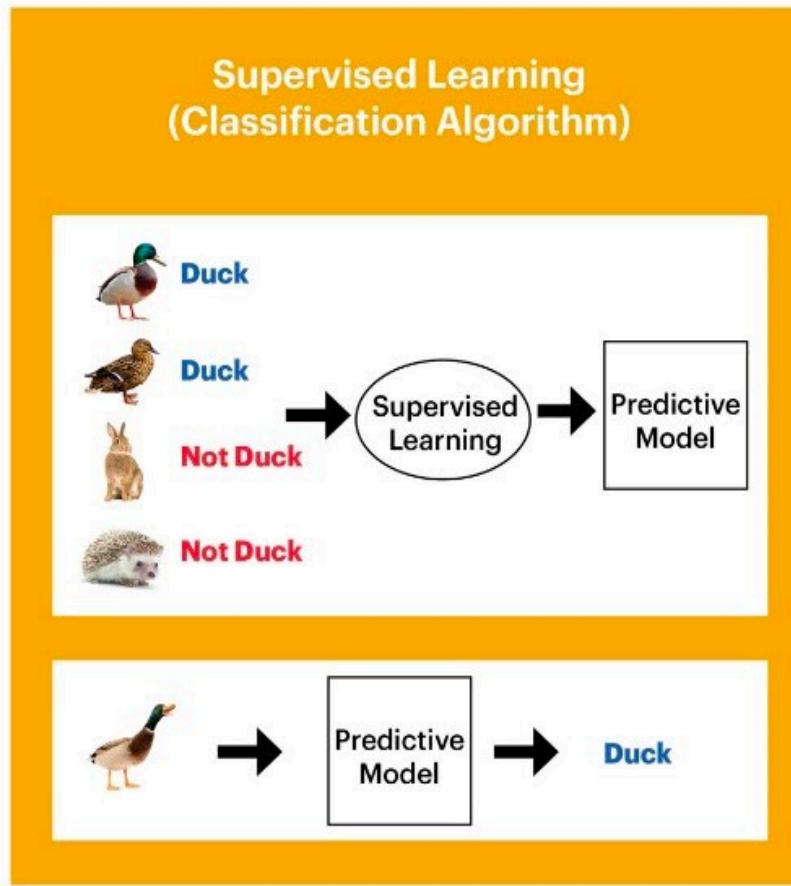
Profiling

Link prediction

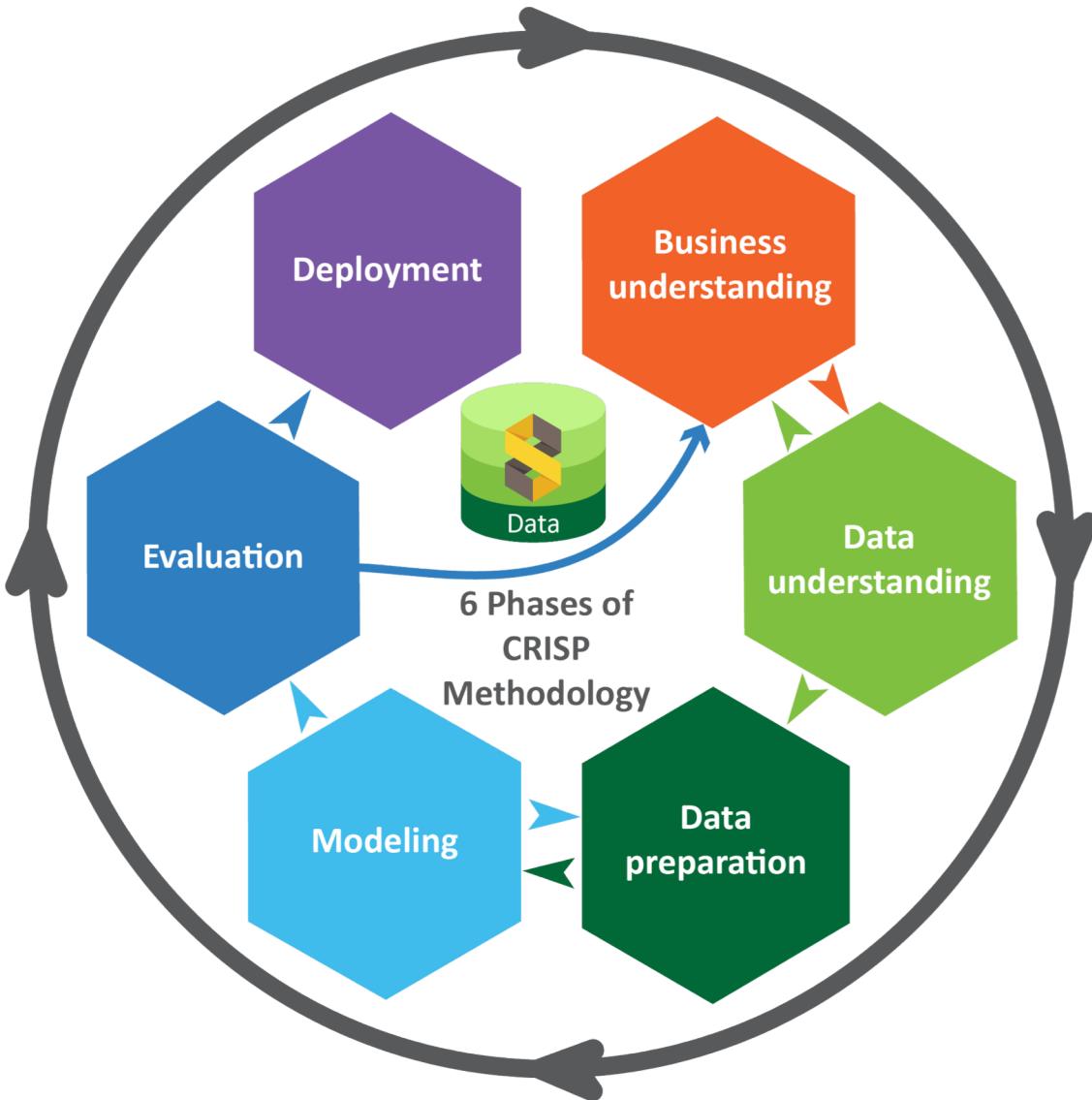
Explanation

Examples

Supervised vs Unsupervised



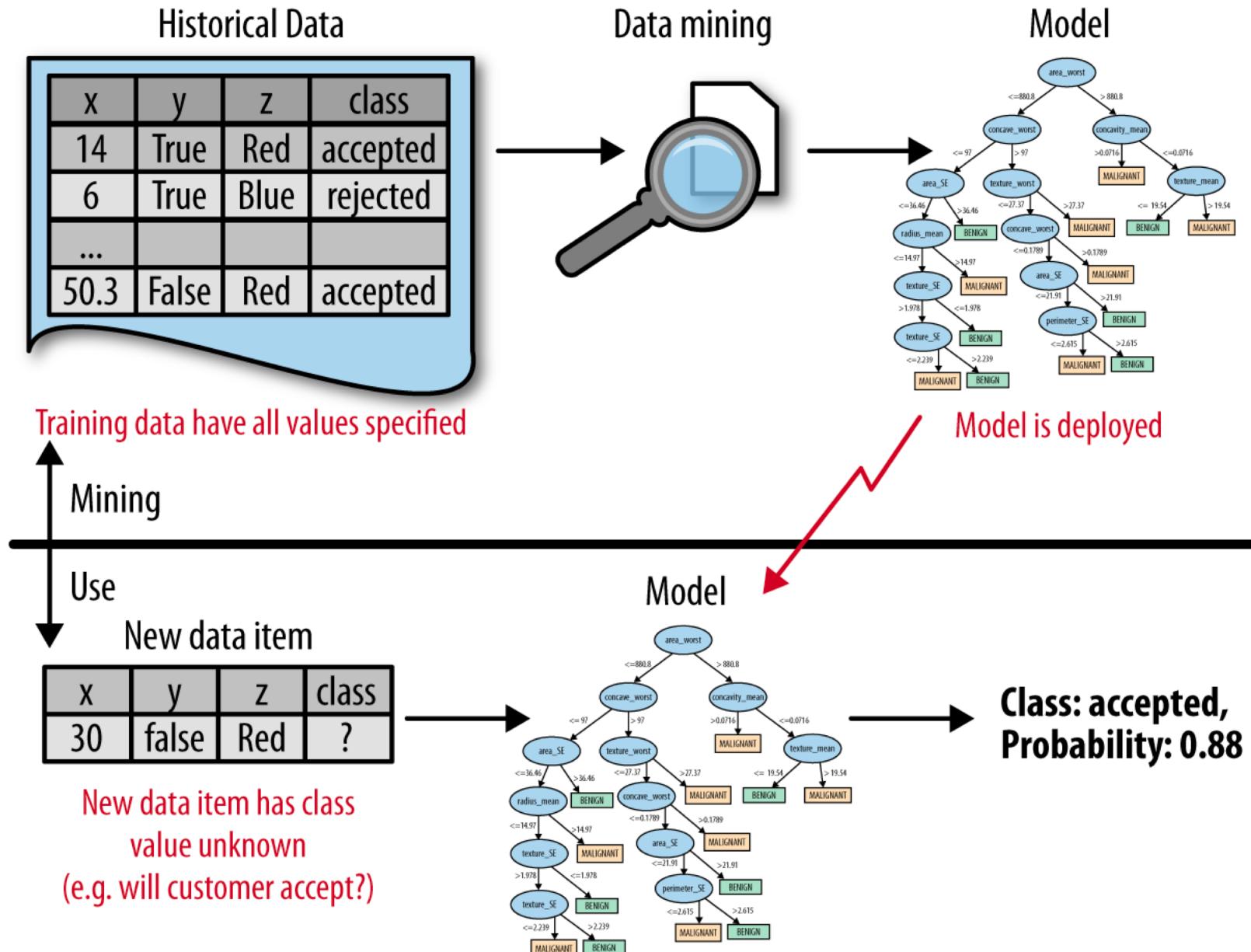
Data Mining Process : CRISP-DM



CRISP-DM Process Steps

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Model: Learn, Apply

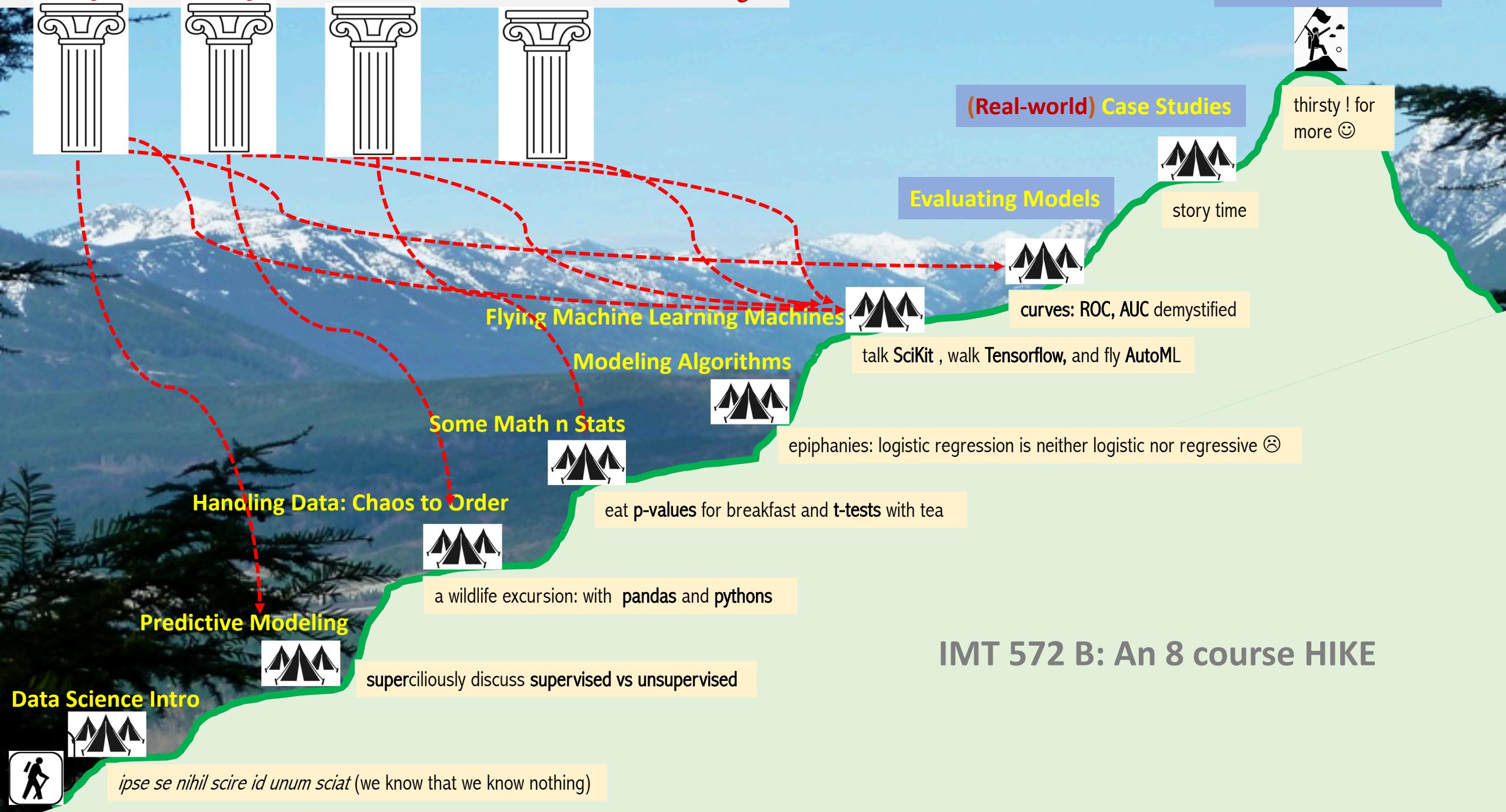


Success

		Medical	Software Engineering	Stock Prediction
What do we want to achieve ?	What am I trying to achieve ?	Predict an onset Raise a alarm Screen Treatment Intervene Help research (genetics, marrow match)		
	What do I know			
	What is my question of the data			
Data	What is my data assumption			
Approach	What is my approach selection			
Team	My team			
Quantify: risk, success	Risks			
	Success criteria			
	My metric ?			

Success

- Define outcomes
- DATA !
- Approach
- Team



IMT 572 B: An 8 course HIKE