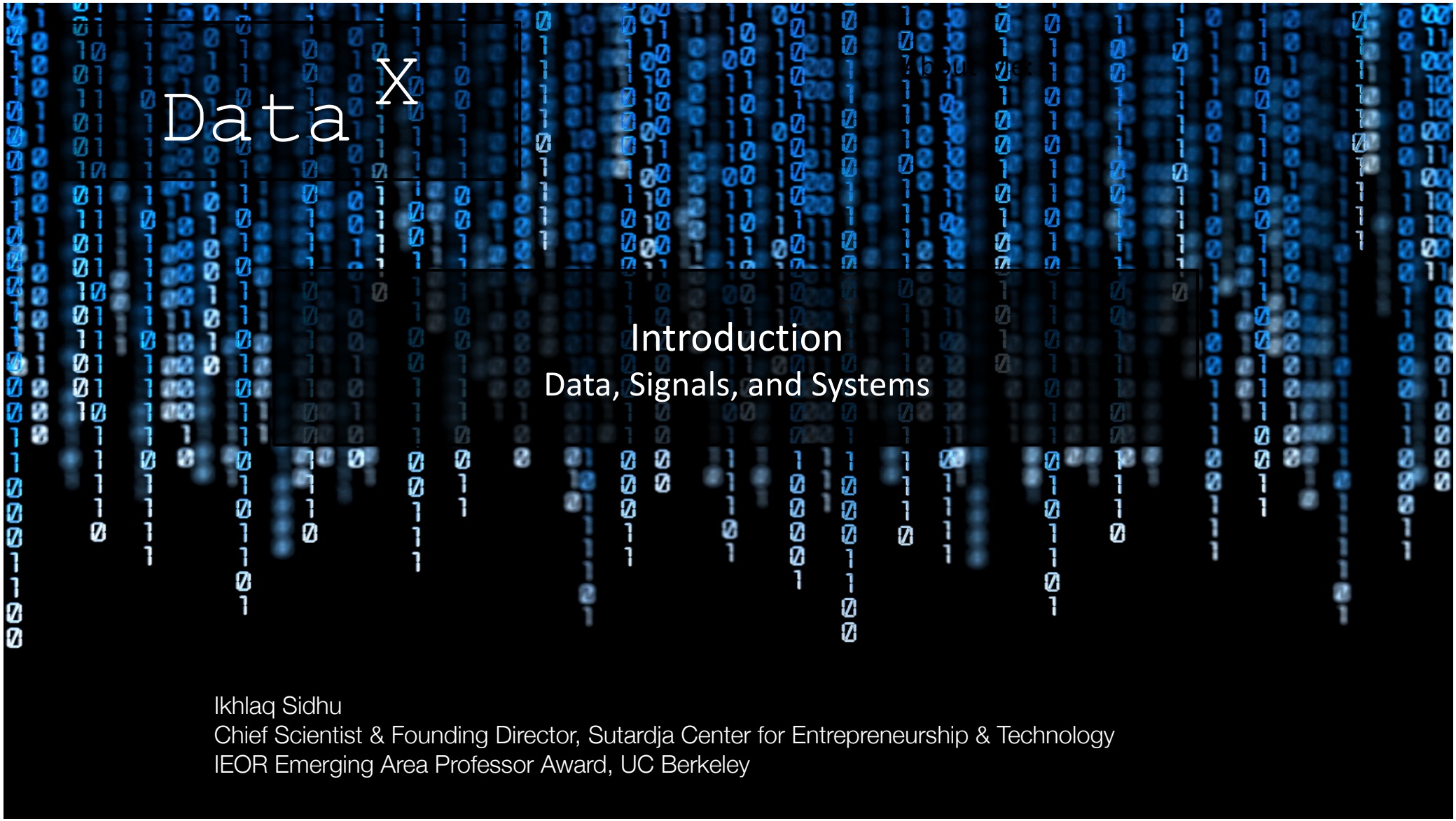


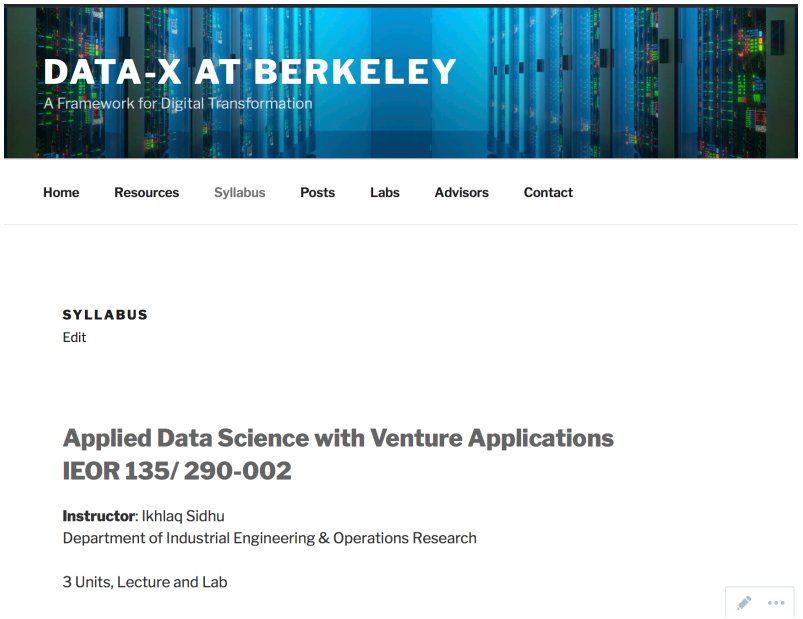
Introduction

Data, Signals, and Systems

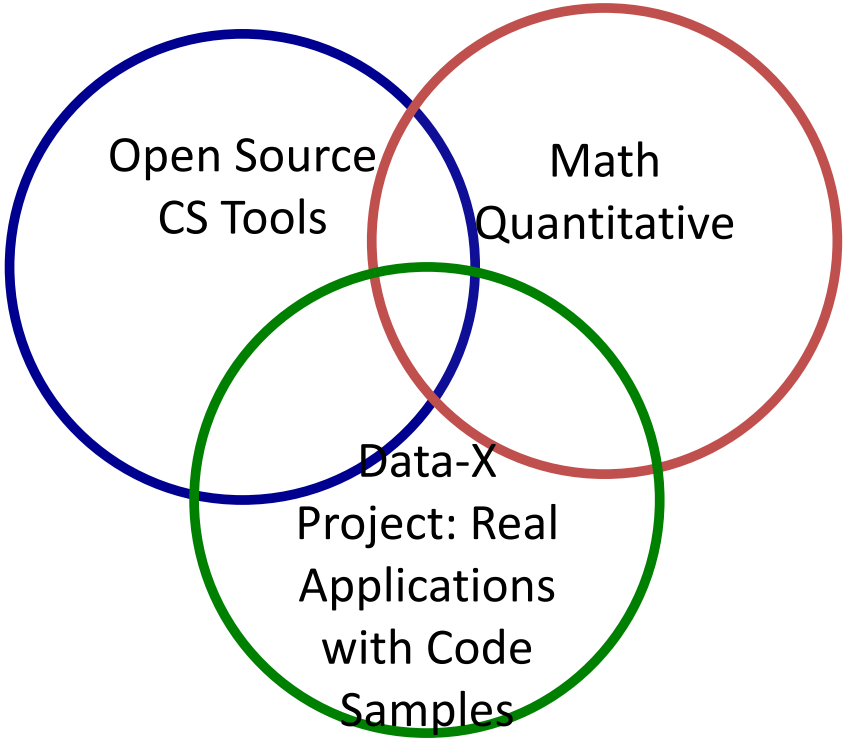
Ikhlaq Sidhu
Chief Scientist & Founding Director, Sutardja Center for Entrepreneurship & Technology
IEOR Emerging Area Professor Award, UC Berkeley

Most Resources Are Available at data-x.blog

1. Go to Data-X.blog
   - Syllabus
   - Instructions for SW Install
   - Link to GitHub with Cookbook Code Samples and Slides
2. Download Instructions to Install Python 3.x Anaconda Environment. For now you only need Anaconda, don't worry about other packages that are not already included.
3. Be able to create your own Jupyter notebook
4. Self-Review Python references as needed. See Ref CS01 and as needed BIDS Python Bootcamp.
5. HW to be emailed by Bcourses lists



What is in this course



Holistic Perspective: Industry, Social Applications, Customer Driven

What is in this class?

Common Open Source CS Tools:

- Numpy, SciPy
- Pandas
- TensorFlow, Sklearn
- SQL to Pandas
- NLP / NLTK
- Matplotlib

Often: Working Code First
Fill In Theory After

Quantitative

- Prediction: Regression
- ML Classification: Logistic, SVM.. Trees, Forests, Bagging, Boosting,..
- Entropy / Information Topics
- Deep Learning examples, including CCNs
- Correlations
- Markov Processes
- LTI Systems: Fourier, Filters where applicable
- Control Models where applicable

Building Block Code Samples

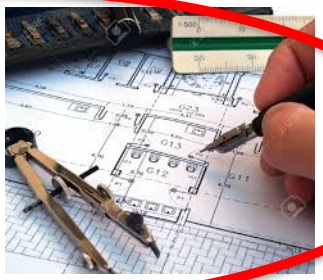
- Webscraping
- Stock market live download, simple trading
- Convolutional Neural Networks
- Next Word Predictor, Spell Checking
- Recommendation
- Web Crawler
- Chatbot, E-mail
- Social net interfaces including twitter

This class will help you combine math and data concepts
The course updates with new tools to stay current. You may learn and use tools not presented in the class project.

What is actually in this class?

- The ML stack use most commonly used in creating ML/AI/Data applications
- Application and systems viewpoint of data and ML
- Implementation, architecture, and relevant process to build anything
- Statistical, rule based, and hybrid decision systems
- Connection with relevant mathematical foundations (entropy, correlation, spectral, LTI, basic prediction, classification)
- Practical insight into advanced techniques and tools: (eg. CNNs, NLP, scraping, recurrent networks, etc.)
- System modeling for data applications

Where we will focus:



Make the Tools

Use the Tools (Optimally)

Architect the System

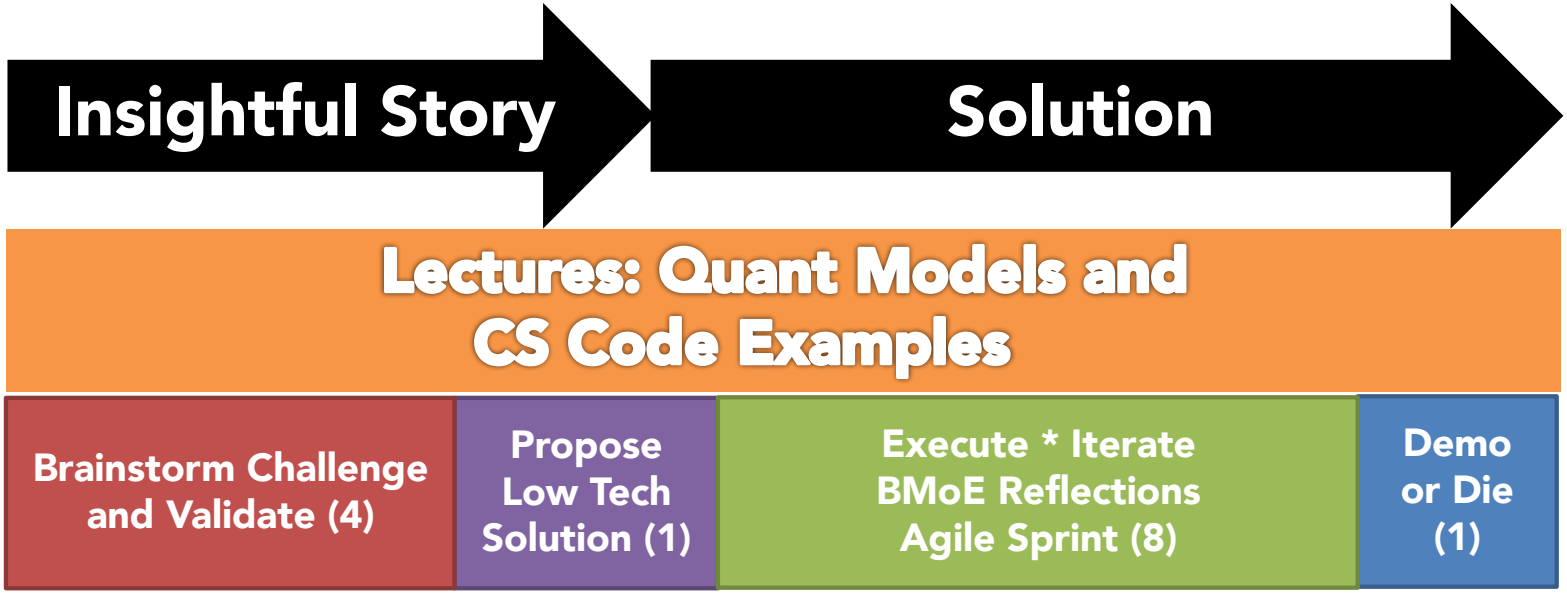
Why and how you build

Most CS

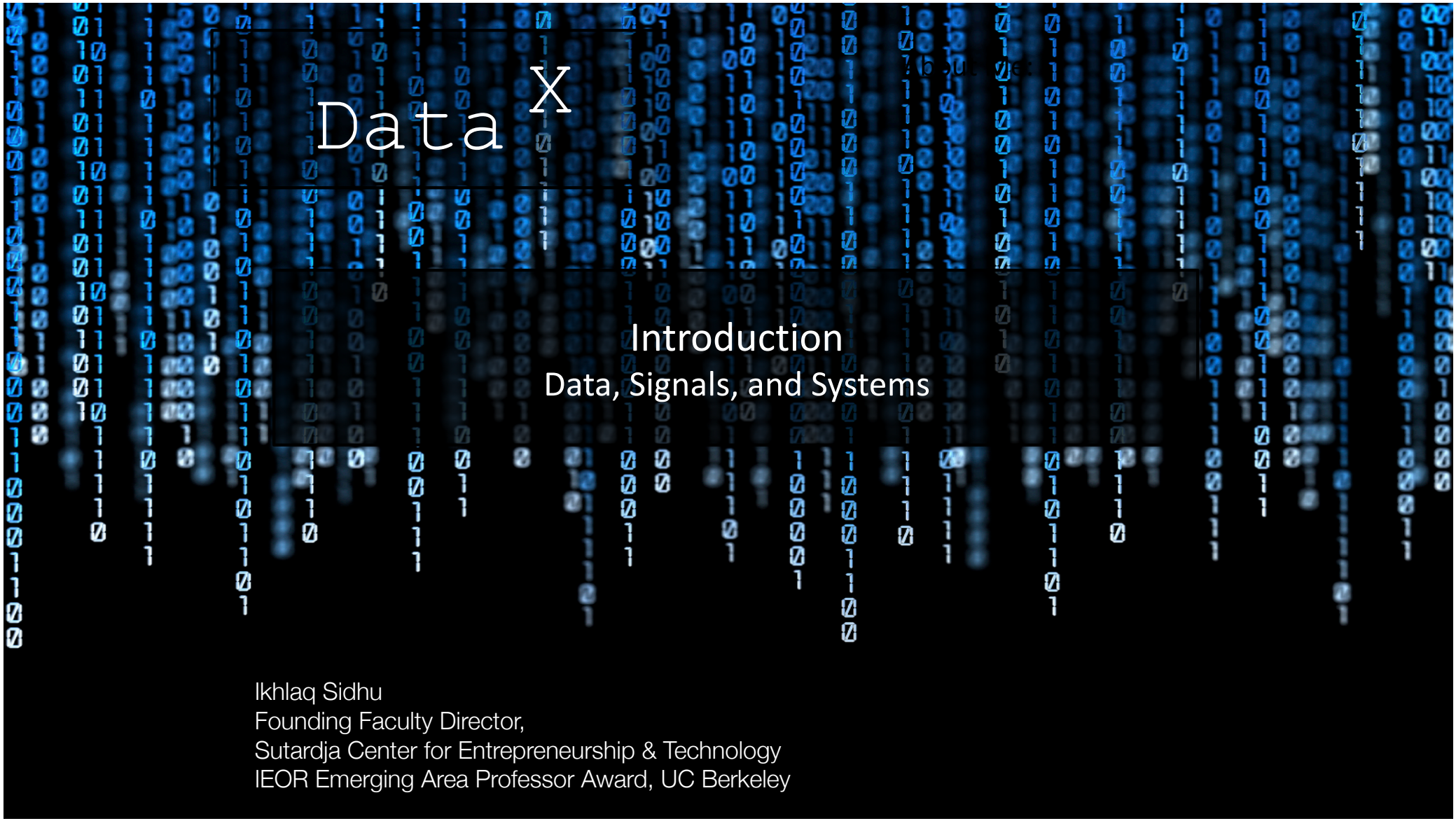
This Course

Sutardja Center

How the Data-X Course Works:



Team: typically 5 students, with available advisor network



Ikhlaq Sidhu
Founding Faculty Director,
Sutardja Center for Entrepreneurship & Technology
IEOR Emerging Area Professor Award, UC Berkeley

A High Level Overview of Data

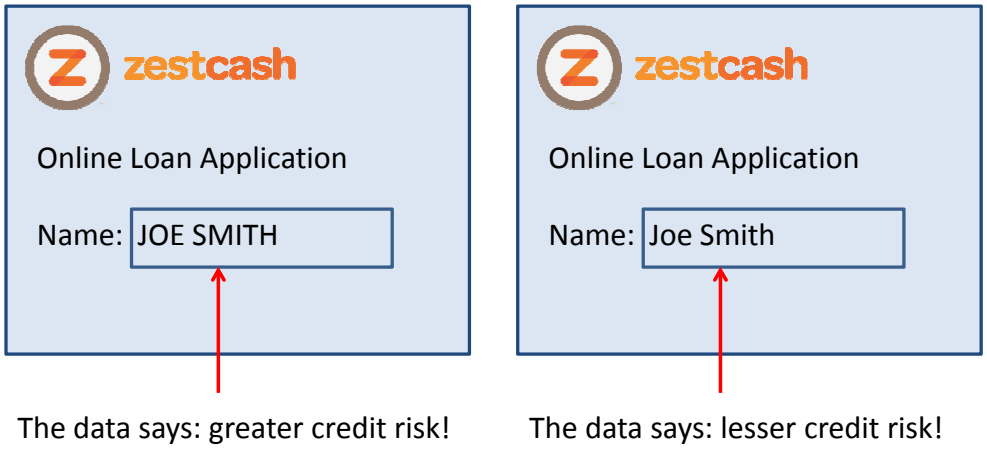
Basic Concept of Working with Data

- Data Wrangling
- In Production

Example: Data and Information is a competitive advantage

Real-life Example: ZestCash

- "All data is credit data"



Reference: Shomit Ghose


• Service provider of Gambling and Casinos
• Entry Card
• Pain points
• Intervention
Harrah's Casino: Knowing your customer
PLAY & WIN
Reference: Supercrunchers

Top 8 Business Models Using Data

1. Knowing your customer, better targeting and relationship.
   E.g. Target, Disney, Netflix

2. Improving physical product or servicer with complimentary information:
   E.g. UPS, FedEx

3. Data-driven reliability or security
   E.g. GE, BMW, Siemens

4. Information Brokers, Arbitrage, and Trading Opportunities:
   E.g. Investment funds.

5. Improving the customer journey/experience..
   E.g. Harrah's

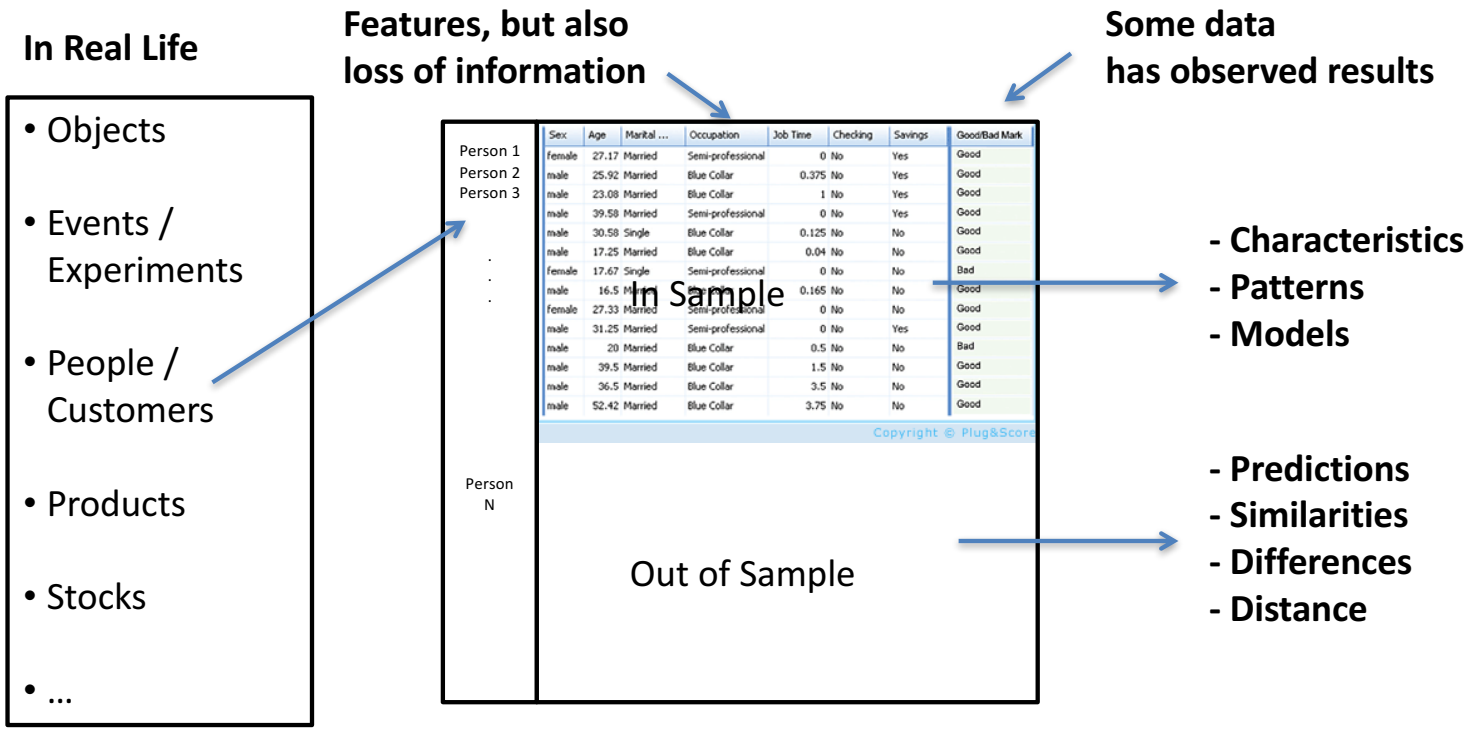
---

6. Functional Applications: HR/Hiring, Operations etc..
   Eg Walmart, Baseball, Sports

7. Efficiency or better performance per dollar cost.
   E.G. General IT, SAP, etc

8. Risk Management, regulation, and compliance
   Eg. Compliance 360

An ML High Level Framework



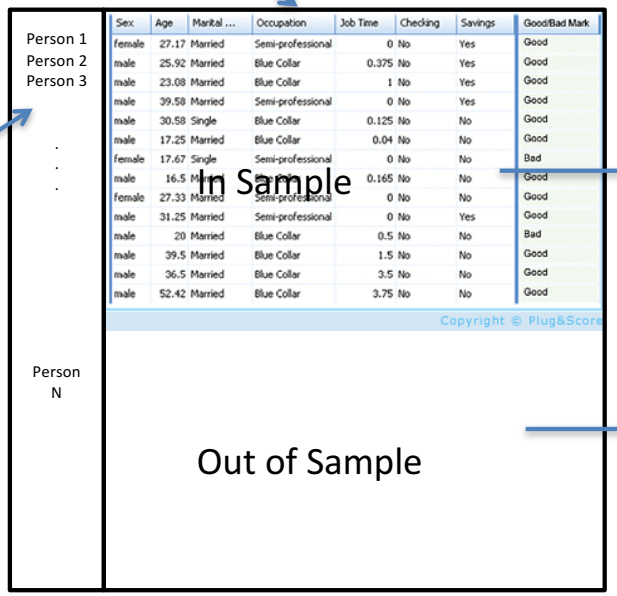
An ML High Level Framework

In Real Life

- Objects
- Events / Experiments
- People / Customers
- Products
- Stocks
- …

Features, but also loss of information

Some data has observed results



- Characteristics
- Patterns
- Models

- Predictions
- Similarities
- Differences
- Distance

CS: Table

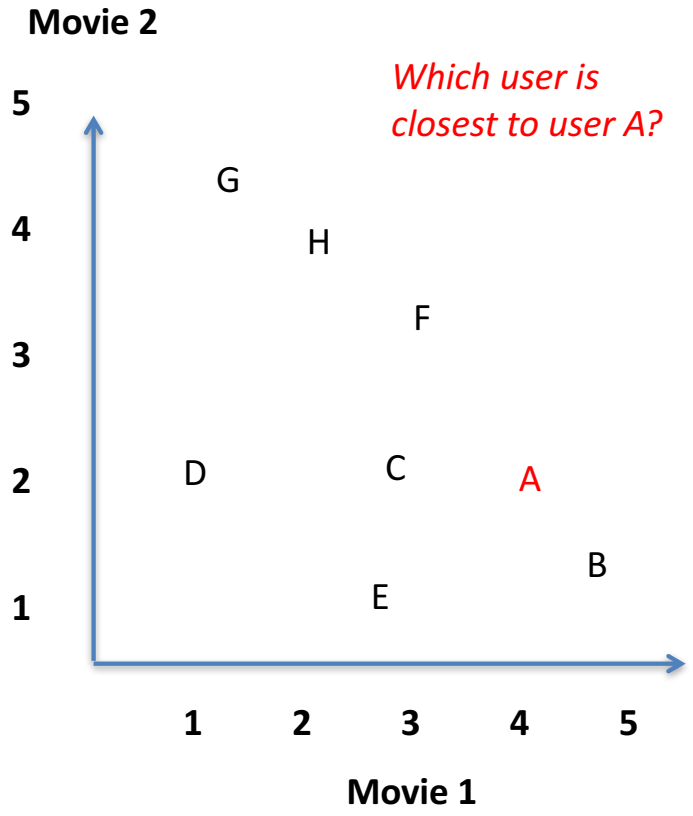
Math: Matrix **X**, with N rows – each person
m columns, each feature (age, salary, ..)

$$\mathbf{X} = \begin{bmatrix} -2 & 4 & 7 & 31 \\ 6 & 9 & 12 & 6 \\ 12 & 11 & 0 & 1 \\ 9 & 10 & 2 & 3 \end{bmatrix}$$

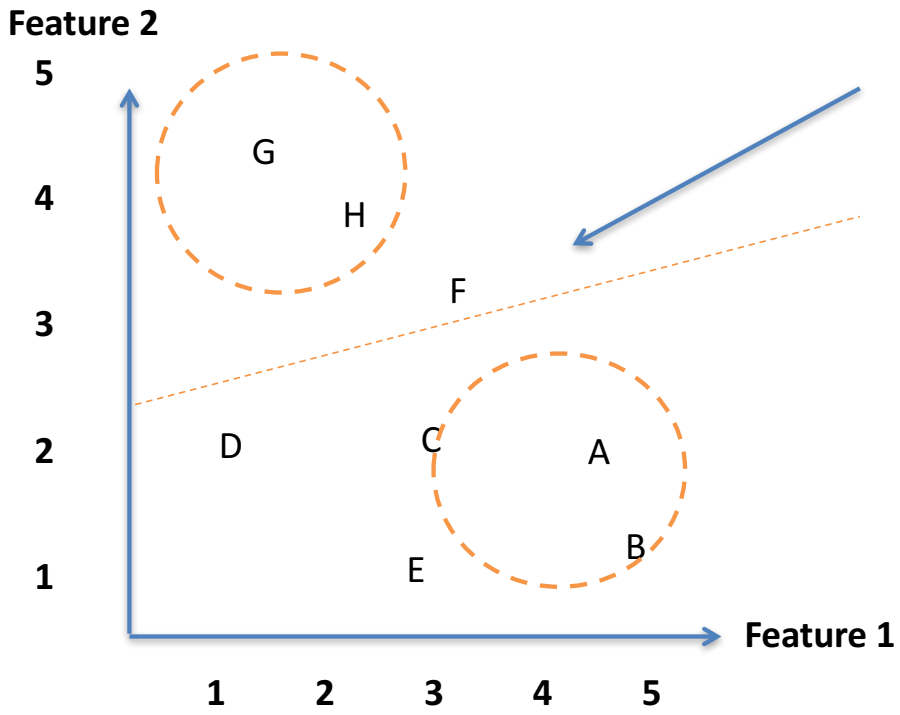
A Fundamental Idea: From Table to N- Dimensional Space

X =

Element	F1	F2	F3
A	4	2	2
B	4.5	1.5	3
C	3	3	5
D	1	2	2
E	3	1.5	5
F	3.5	3.5	1
..	..	..	..



Clustering to Classification



- Target customers?
- Pictures of Cats and Dogs
- Speech recognition
- Recognize Letters: A, B, C..

A Fundamental Idea: From Table to Score

X =

Cust	F1	F2	F3
A	4	2	2
B	4.5	1.5	3
C	3	3	5
D	1	2	2
E	3	1.5	5
F	3.5	3.5	1
..	..	..	..

F(X)

Cust	Credit Score
A	552
B	381
C	760
D	330
E	452
F	678
..	..

Machine Learning: Learning from Data

Input Data = Matrix X

Customer 1: [Name, income, x, y, .. Features ..z]
Customer 2: [Name, income, x, y, .. Features ..z]
Customer N: [Name, income, x, y, .. Features ..z]

Output Data = Column Vector Y

Customer 1: [20]
Customer 2: [60]
Customer N: [05]

Purchases/year, repaid loan, …

Target: What is F(X) = Y a formula that we don't know

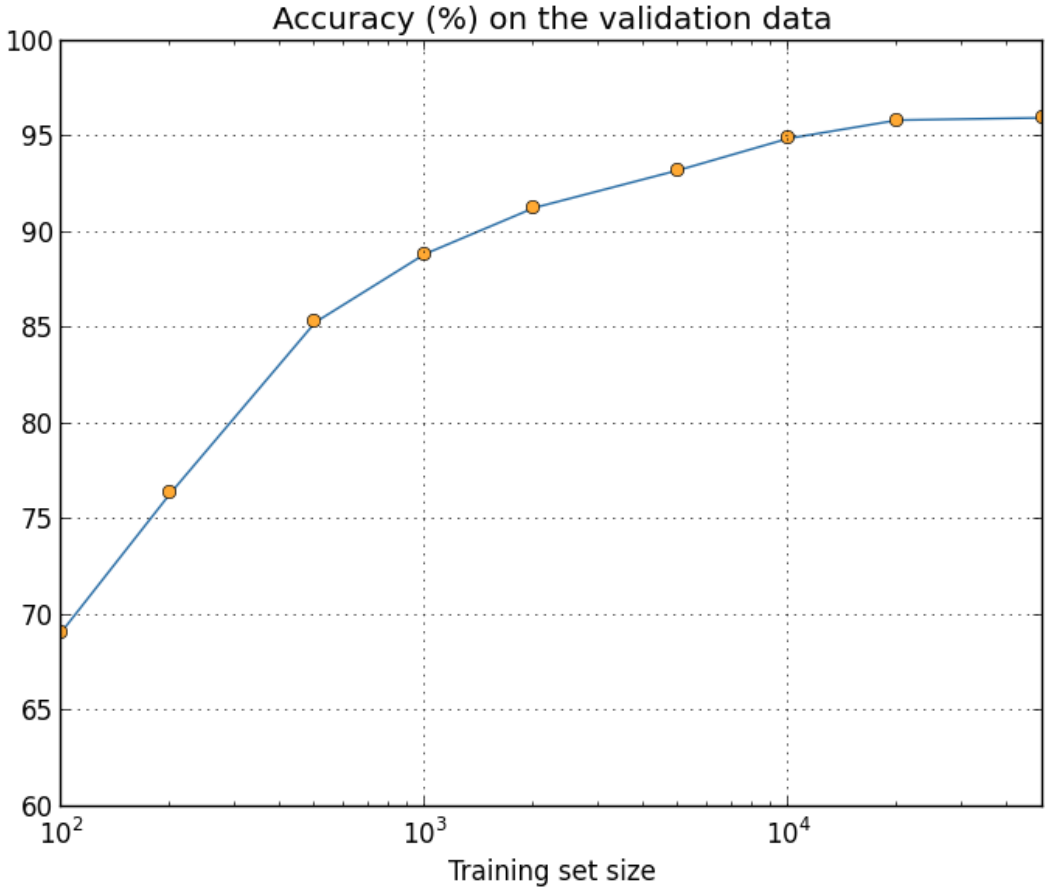
Sample data (training): (x1,y1) (x2,y2) … (xm,ym) we have this

Algorithm A from H

Find G(x) which is approx. F(x)

H: Hypothesis Set:
All possible algorithms or formulas

a) Supervised ML – as shown

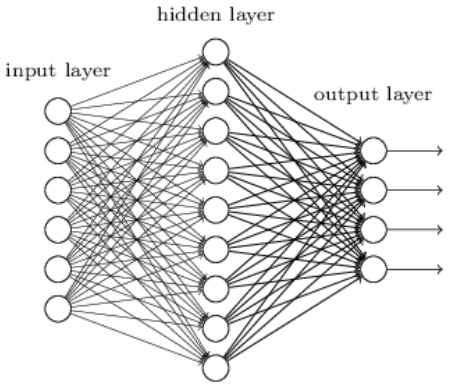

Accuracy (%) on the validation data
100
95
90
85
80
75
70
65
60
10^2
10^3
10^4
Training set size

Neural Networks as Function Approximators

X → ML Algorithms Guess this function F(x) → Y

"Non-deep" feedforward neural network

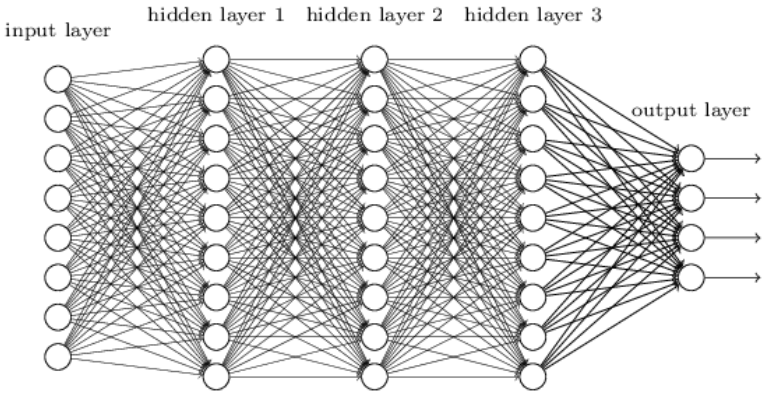
X



Y

Deep neural network

X



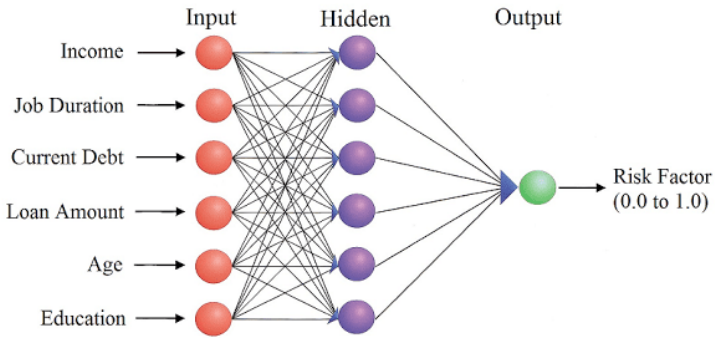
Y

Neural Networks can also Perform regression

X → ML Algorithms Guess this function F(x) → Y

"Non-deep" feedforward neural network

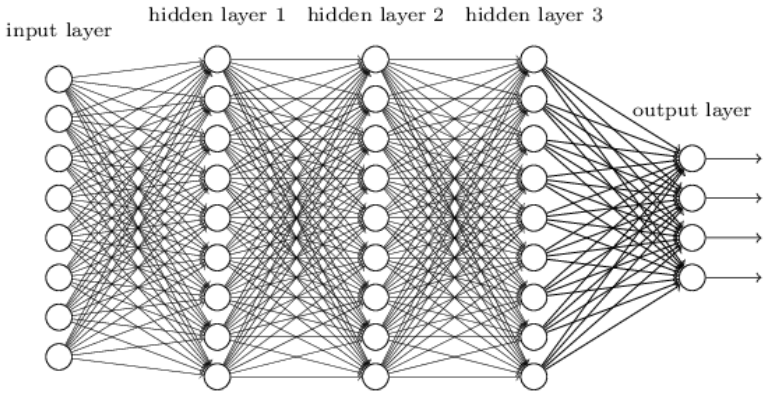
X



Y

Deep neural network

X



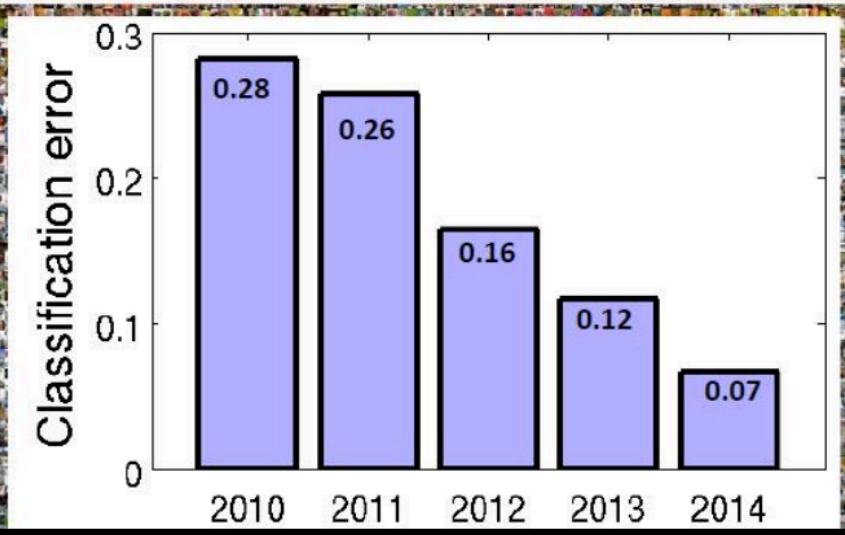
Y

IMAGENET **Large Scale Visual Recognition Challenge**

The Image Classification Challenge:

1,000 object classes

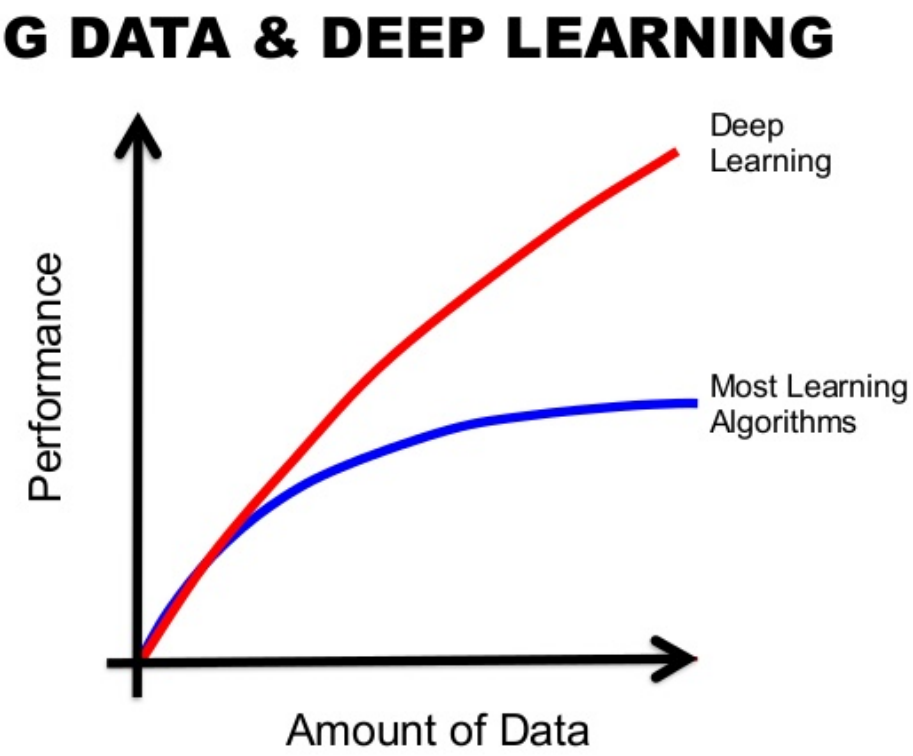
1,431,167 images



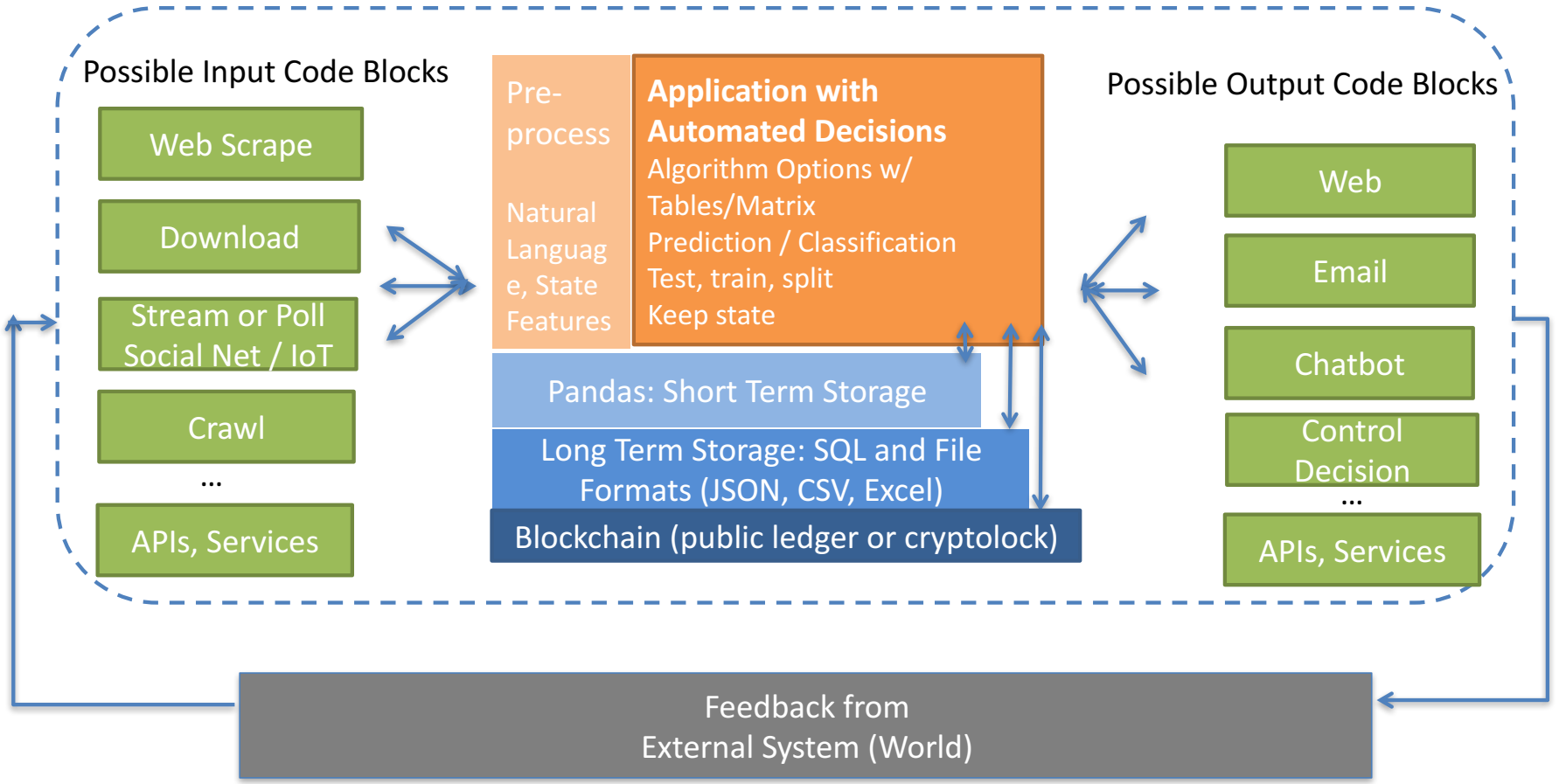
Neural net results are close to human results

Data

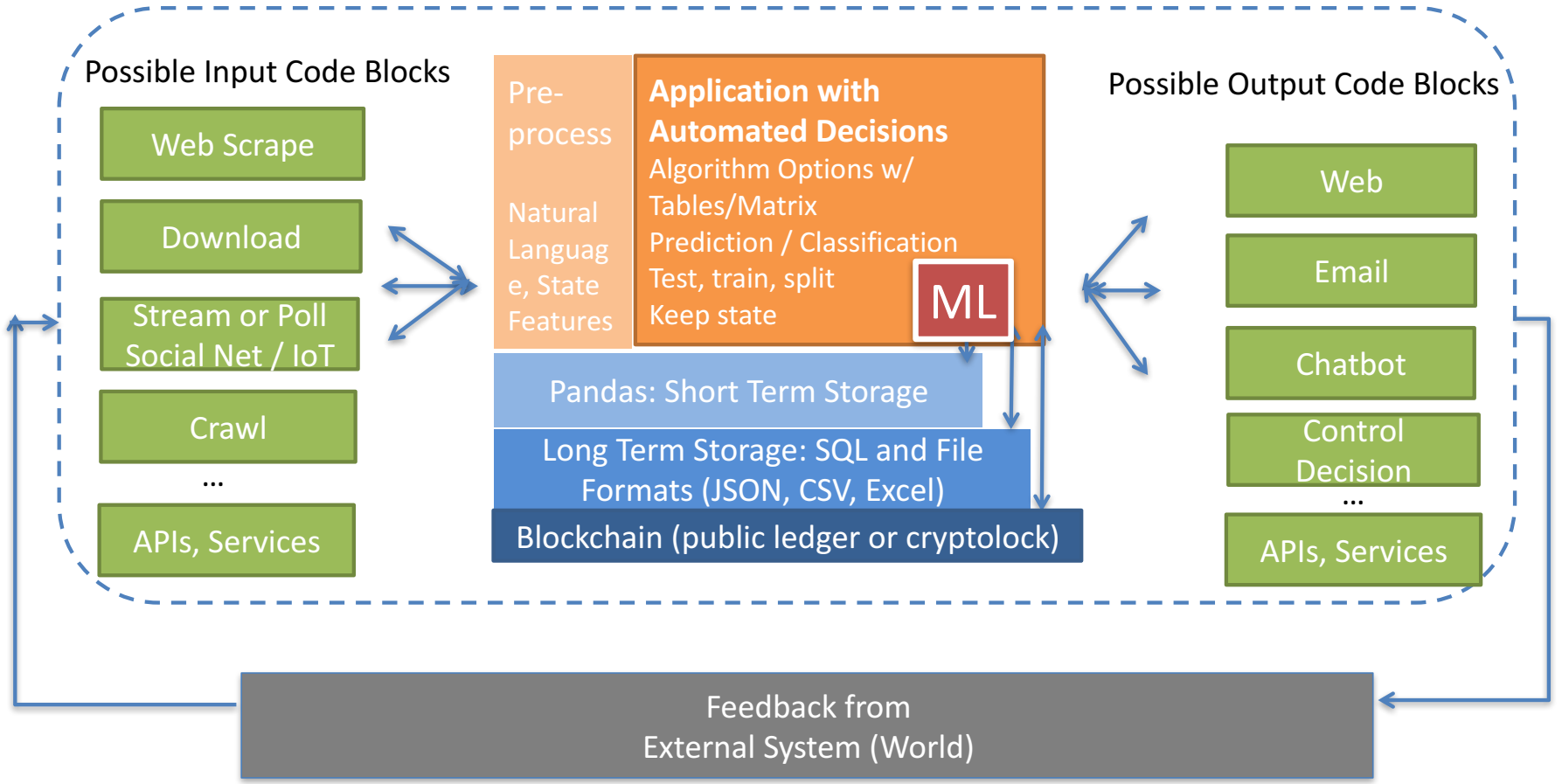
This means
Accuracy



The Data-X System View



The Data-X System View: It's more than ML, it's also systems and models



Project Types

Business or Consumer Use Case

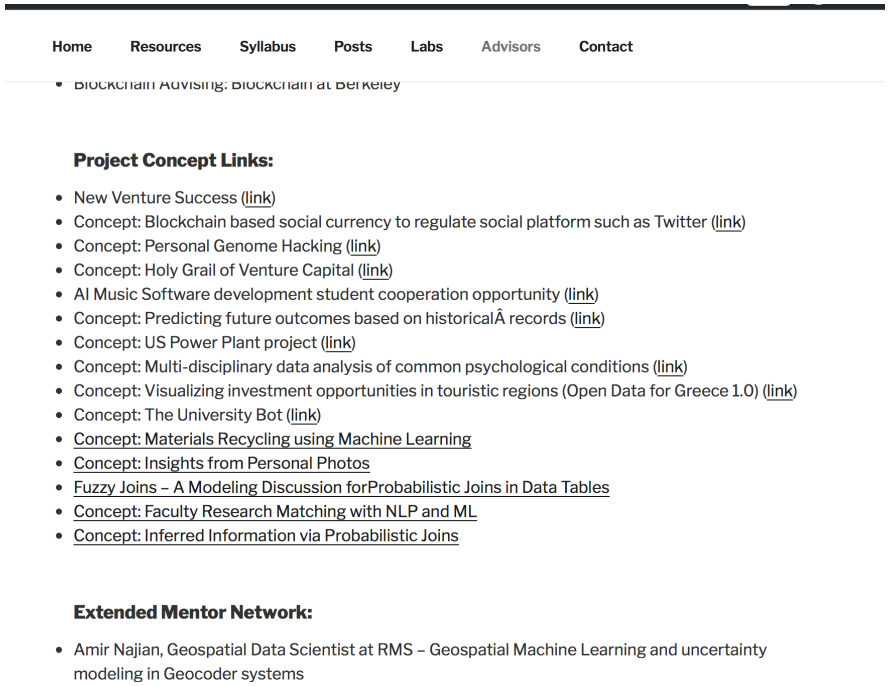
Social Impact

(or improve part of a data pipeline or work towards a research result)

Its Just Cool

Project Ideation

- Past Projects Concepts:
  - See the Advisor's Tab of data-x.blog
- Past Projects:
  - See the archive on the Posts page and on the Labs page of Data-x.blog
- Combine ideas or extend previous work
- You can also choose to build part of a system,
  - Ie, just the part that automatically collects data by web scraping, or
  - just the part that makes a decision based on data already available



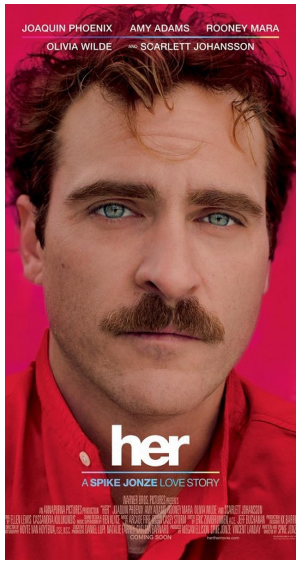
Project Concept: Predicting Energy Prices

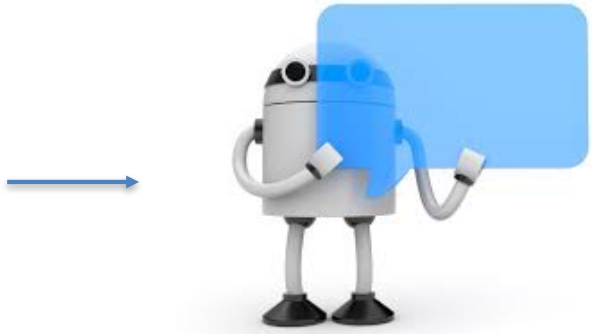
- The goal for this project would be to predict future energy prices on a price per kilowatt hour basis for a specific gas fired electric power plant, preferably located in the USA but also open to other countries, assuming fixed competitive supply in its local market over the next 5 years.
- Introduced by Antonio Vitti, Chief Financial Officer and Senior Technology Executive formerly with Merchant Atlas, Inc. See post at data-x.blog

Project Concept: Chatbot with Personality

- How would you go about creating a chatbot that mimics your grandmother.

- Stephen Torres, SCET





End of Section