Applied Machine Learning for IoT

Hank Roark

13-April-2016

Data Science Pop-up Austin

WHO AMI I

Lead, Customer Data Science @ H2O.ai

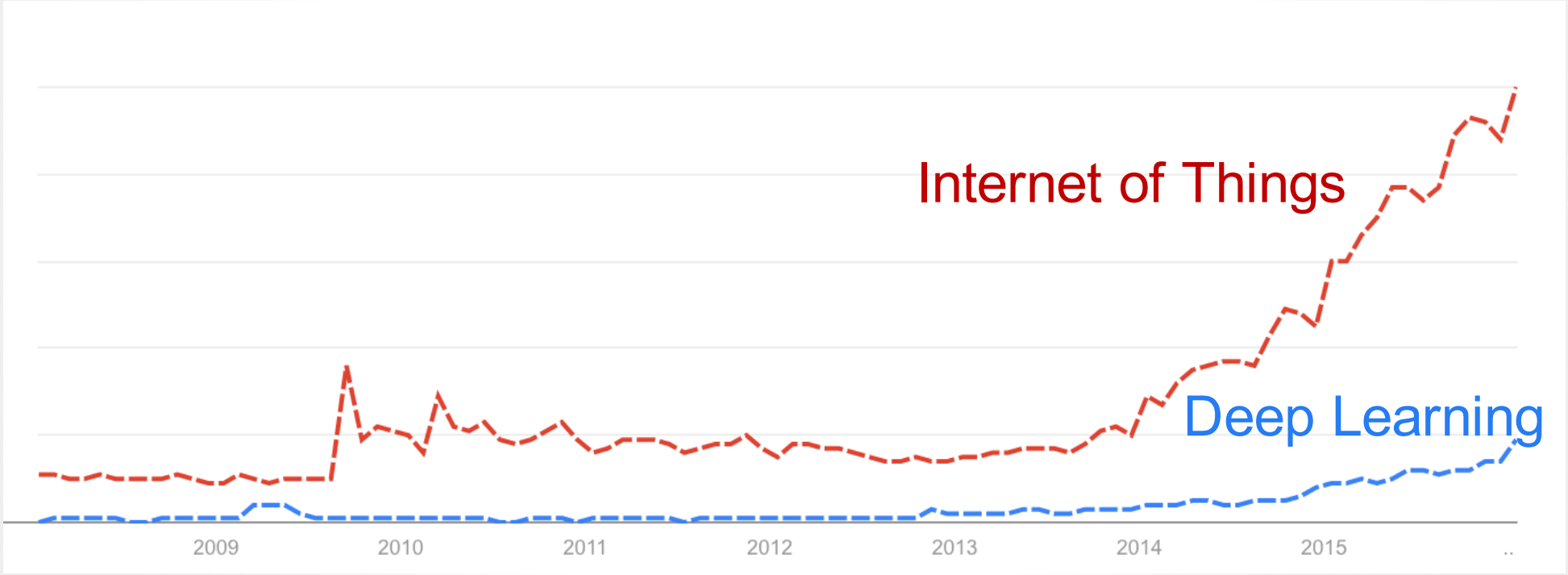
John Deere: Research, Software Product Development, High Tech Ventures
Lots of time dealing with data off of machines, equipment, satellites, weather, radar, hand sampled, and on.
Geospatial, temporal / time series data almost all from sensors.

Systems Design and Management: MIT
Physics: Georgia Tech

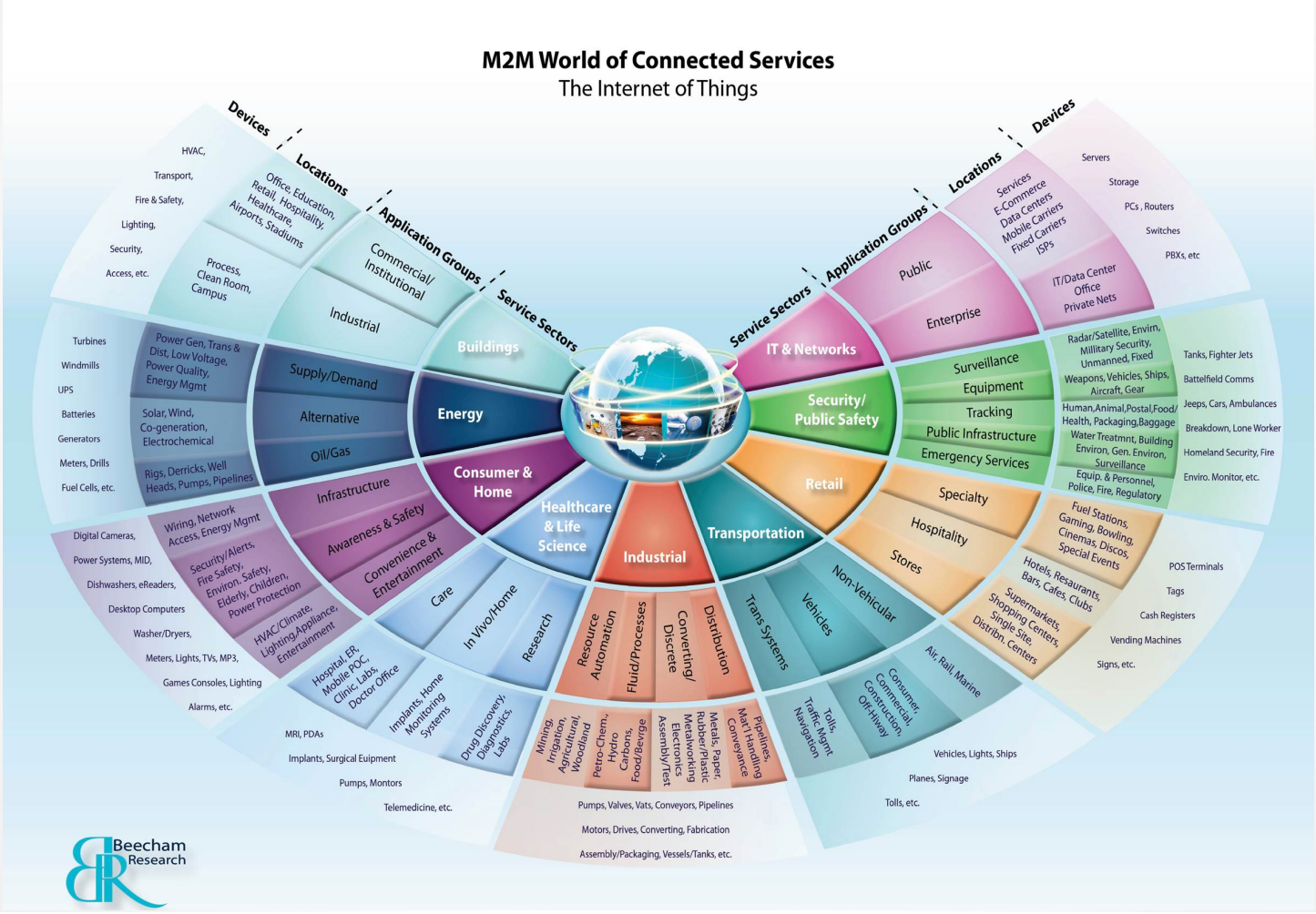
hank@h2o.ai
@hankroark
https://www.linkedin.com/in/hankroark

WHY IOT?

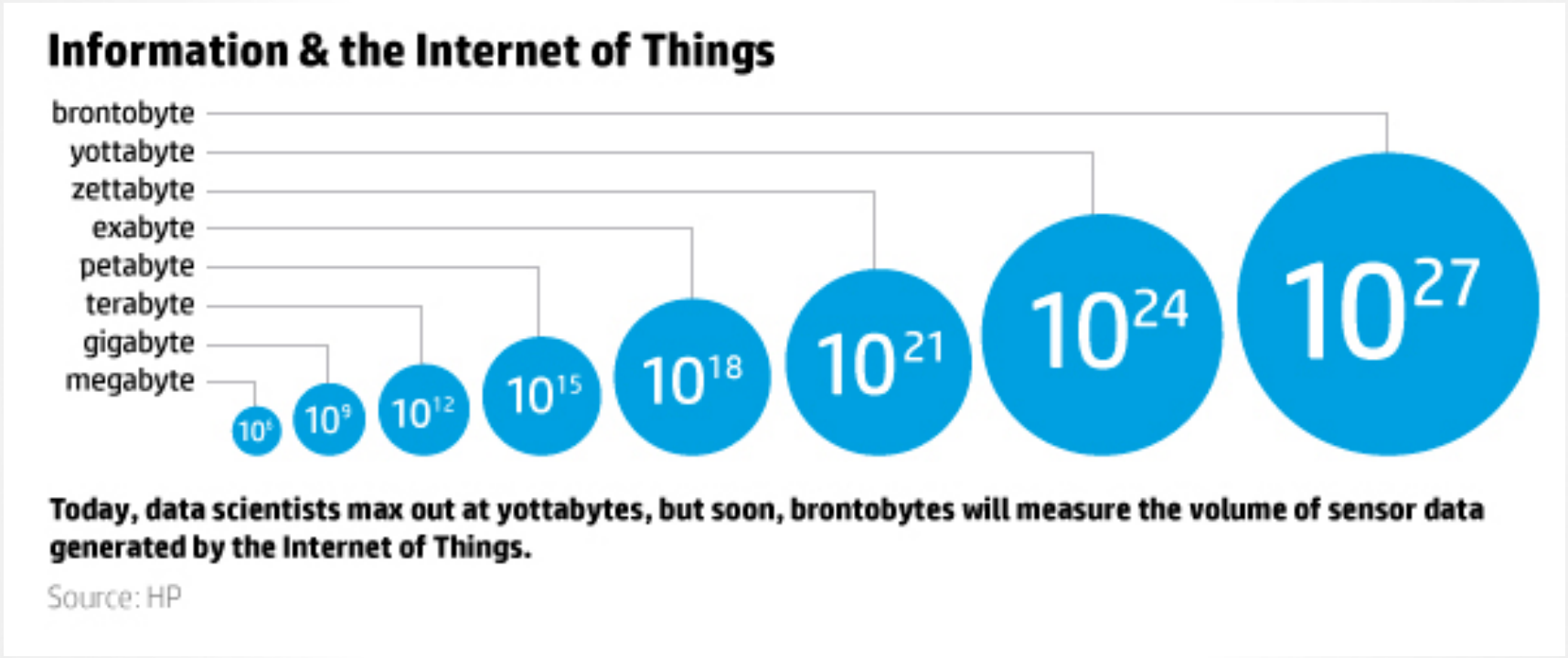
Because it's trending (so it must be promising)!



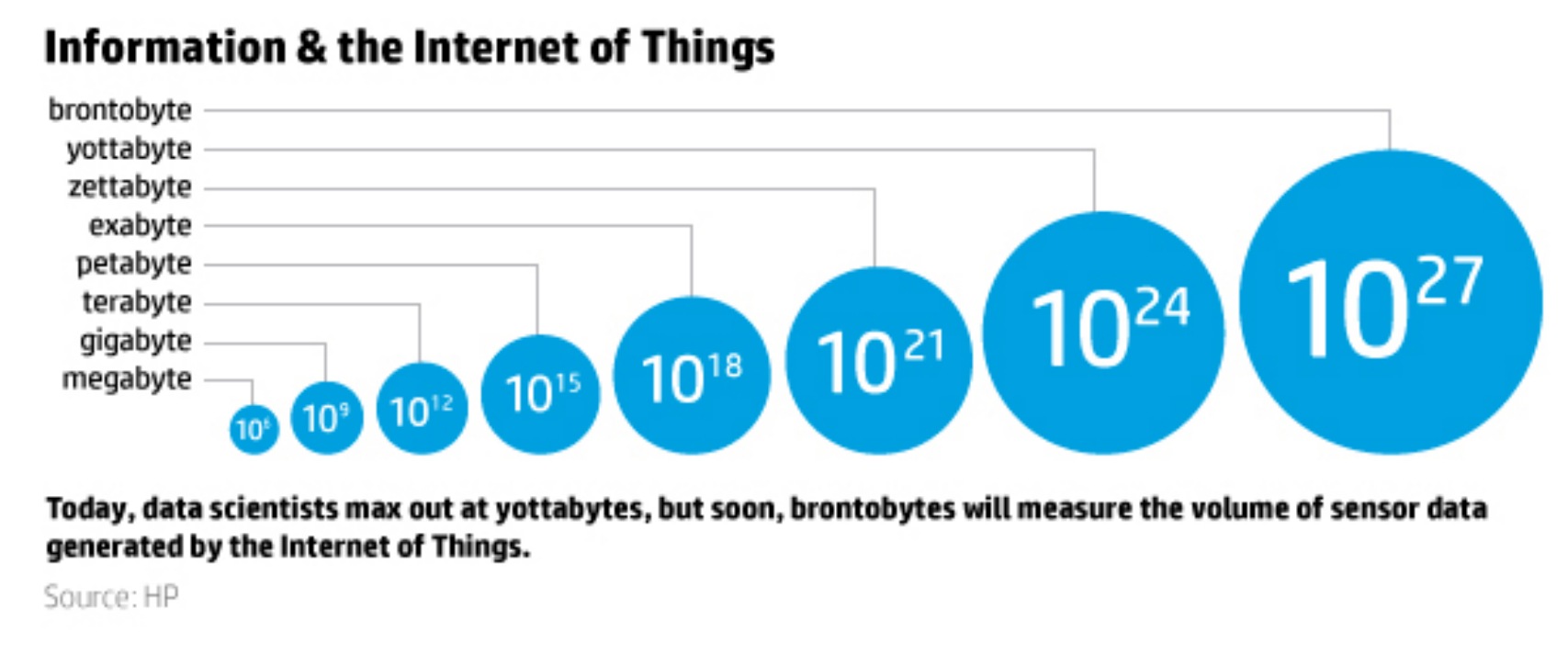
IF YOU ARE INTO DATA, THE IOT HAS IT



LOTS OF DATA


Information & the Internet of Things
brontobyte
yottabyte
zettabyte
exabyte
petabyte
terabyte
gigabyte
megabyte
10^{6}
10^{9}
10^{12}
10^{15}
10^{18}
10^{21}
10^{24}
10^{27}
Today, data scientists max out at yottabytes, but soon, brontobytes will measure the volume of sensor data generated by the Internet of Things.
Source: HP

LOTS OF DATA



GET READY FOR BRONTOBYTES!!

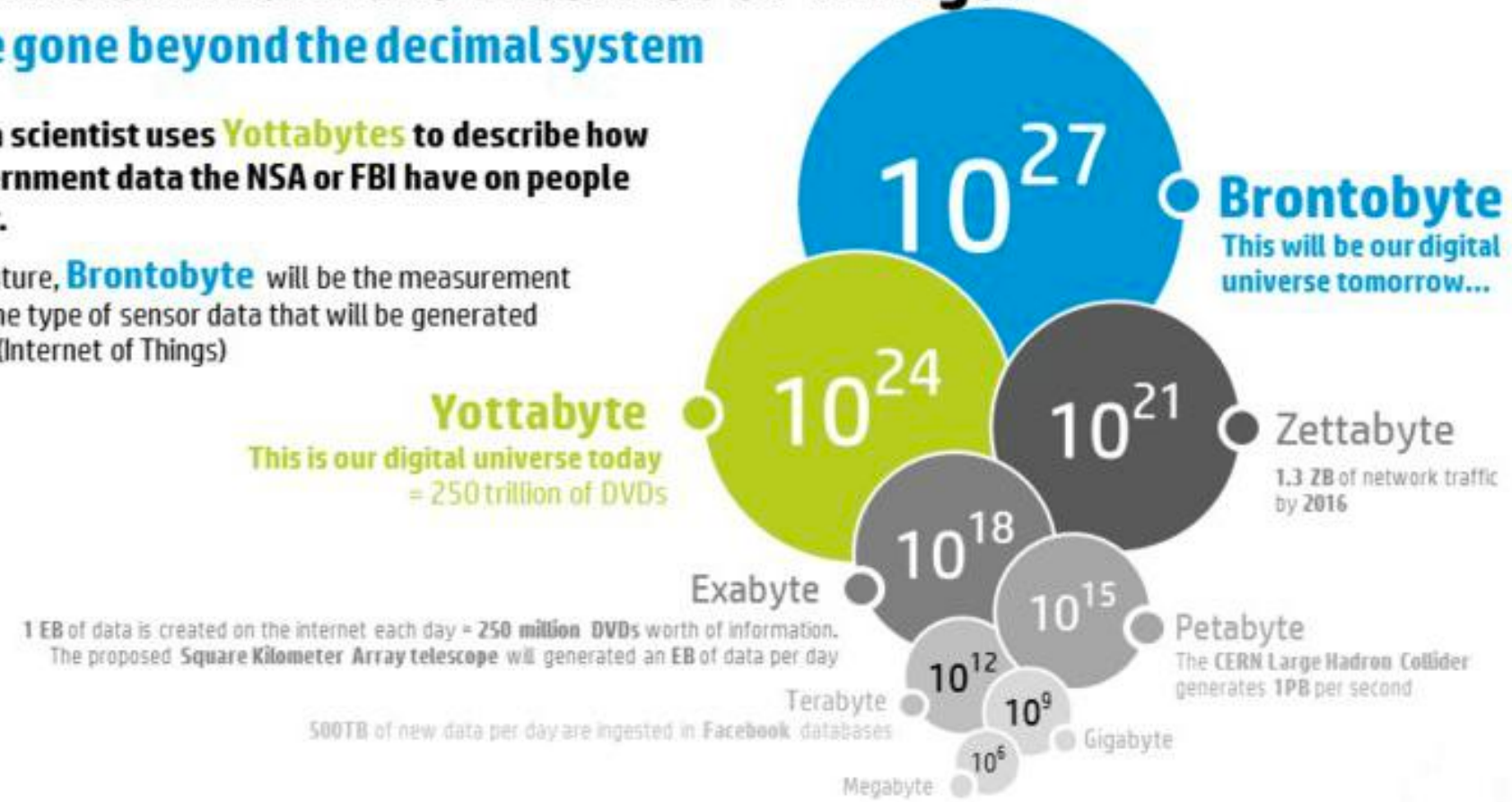
WOW, HOW BIG IS A BRONTOBYTE?

Information from the Internet of Things:

We have gone beyond the decimal system

Today data scientist uses Yottabytes to describe how much government data the NSA or FBI have on people altogether.

In the near future, **Brontobyte** will be the measurement to describe the type of sensor data that will be generated from the IoT (Internet of Things)



IOT ARCHITECTURE IS SETUP FOR REINFORCING, EXPONENTIAL GROWTH

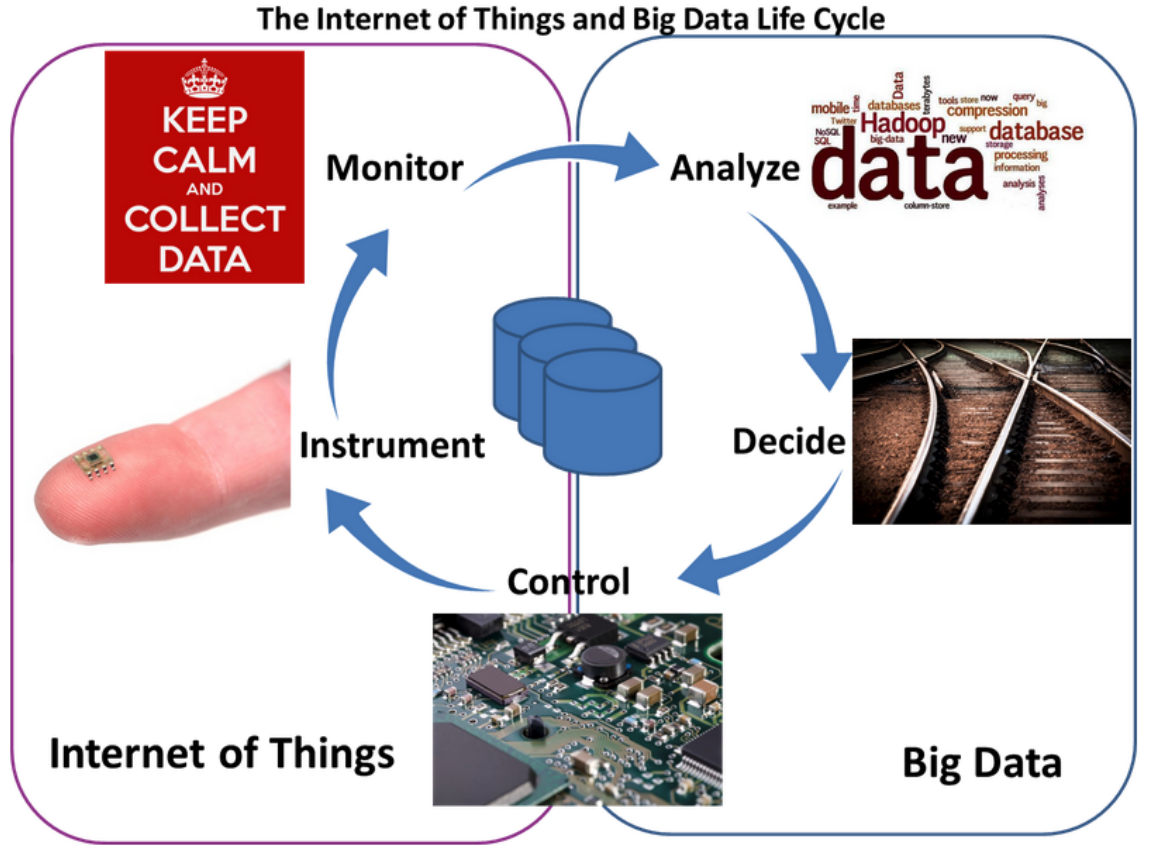
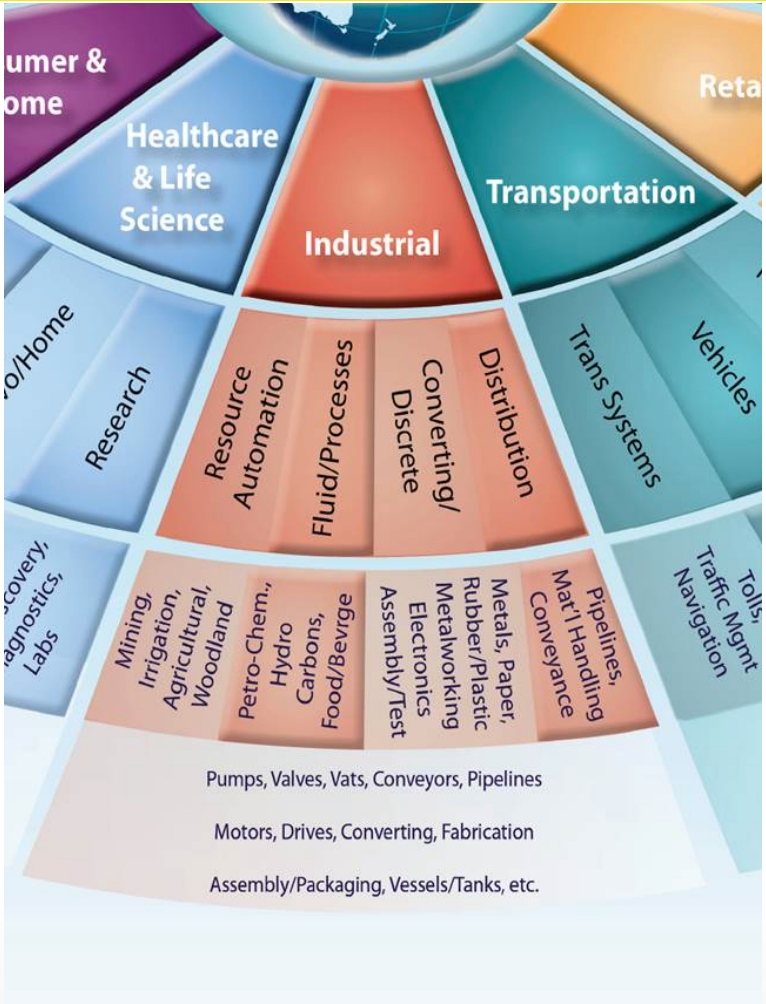


Image courtesy http://www.telecom-cloud.net/wp-content/uploads/2015/05/Screen-Shot-2015-05-27-at-3.51.47-PM.png

MY PARTICULAR CURIOSITY



EXAMPLE FROM THE IOT

Domain: Prognostics and Health Management

Machine: Turbofan Jet Engines

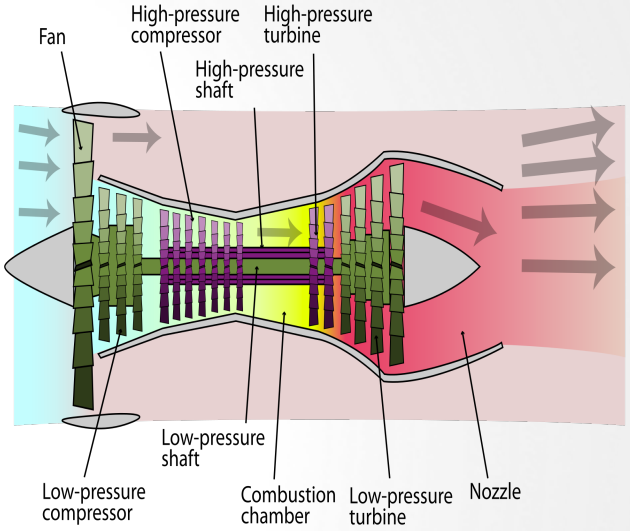
Data Set: A. Saxena and K. Goebel (2008). "Turbofan Engine Degradation Simulation Data Set", NASA Ames Prognostics Data Repository

Predict **Remaining Useful Life** from Partial Life Runs

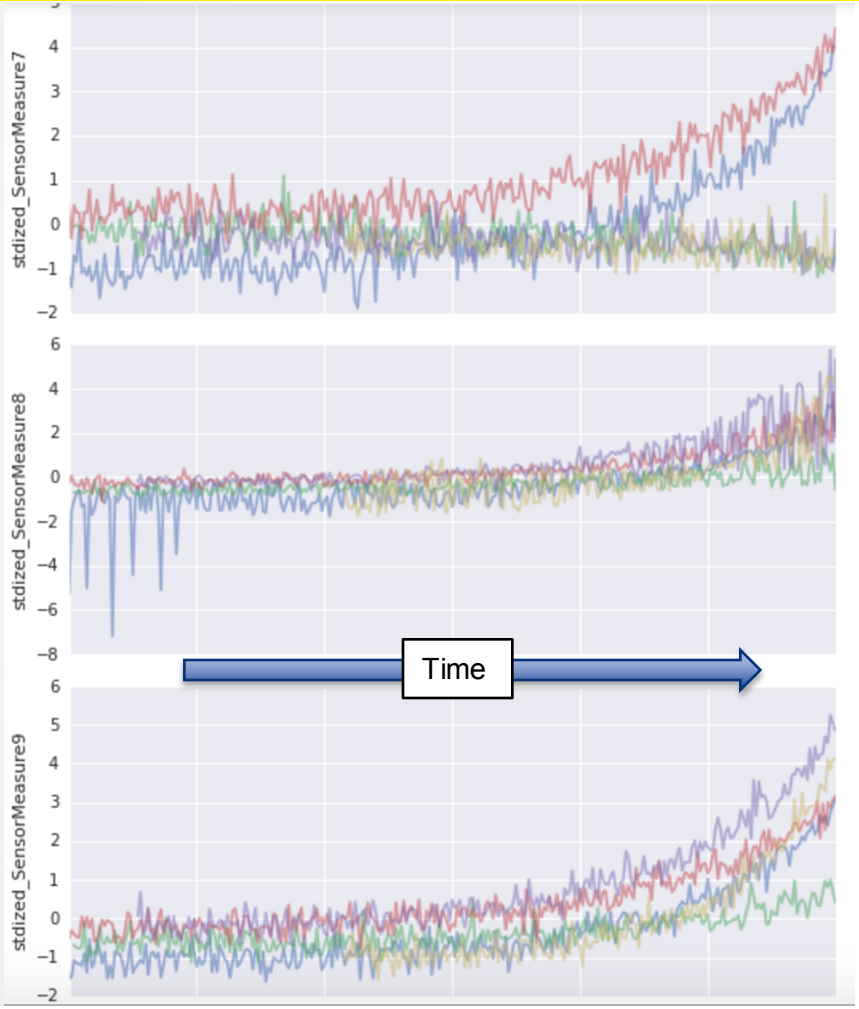
Six operating modes, two failure modes, manufacturing variability

Training: 249 jet engines run to failure

Test: 248 jet engines



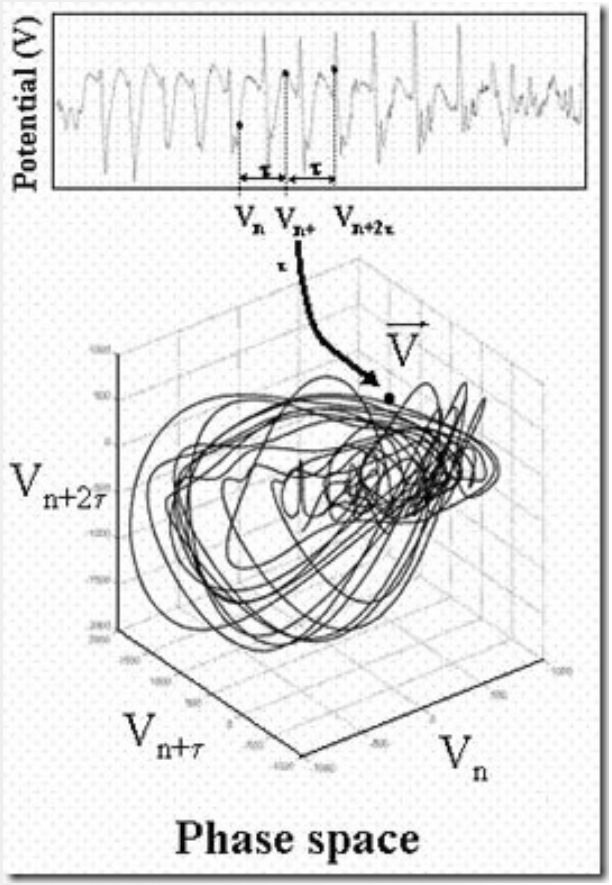
SENSORS SHOW TRENDS OVERTIME



We also know remaining useful life decreases by at least one (1) cycle each operation.

How can one take advantage of this knowledge?

INCORPORATING PRIOR STATE



One option:
Phase Space Embedding

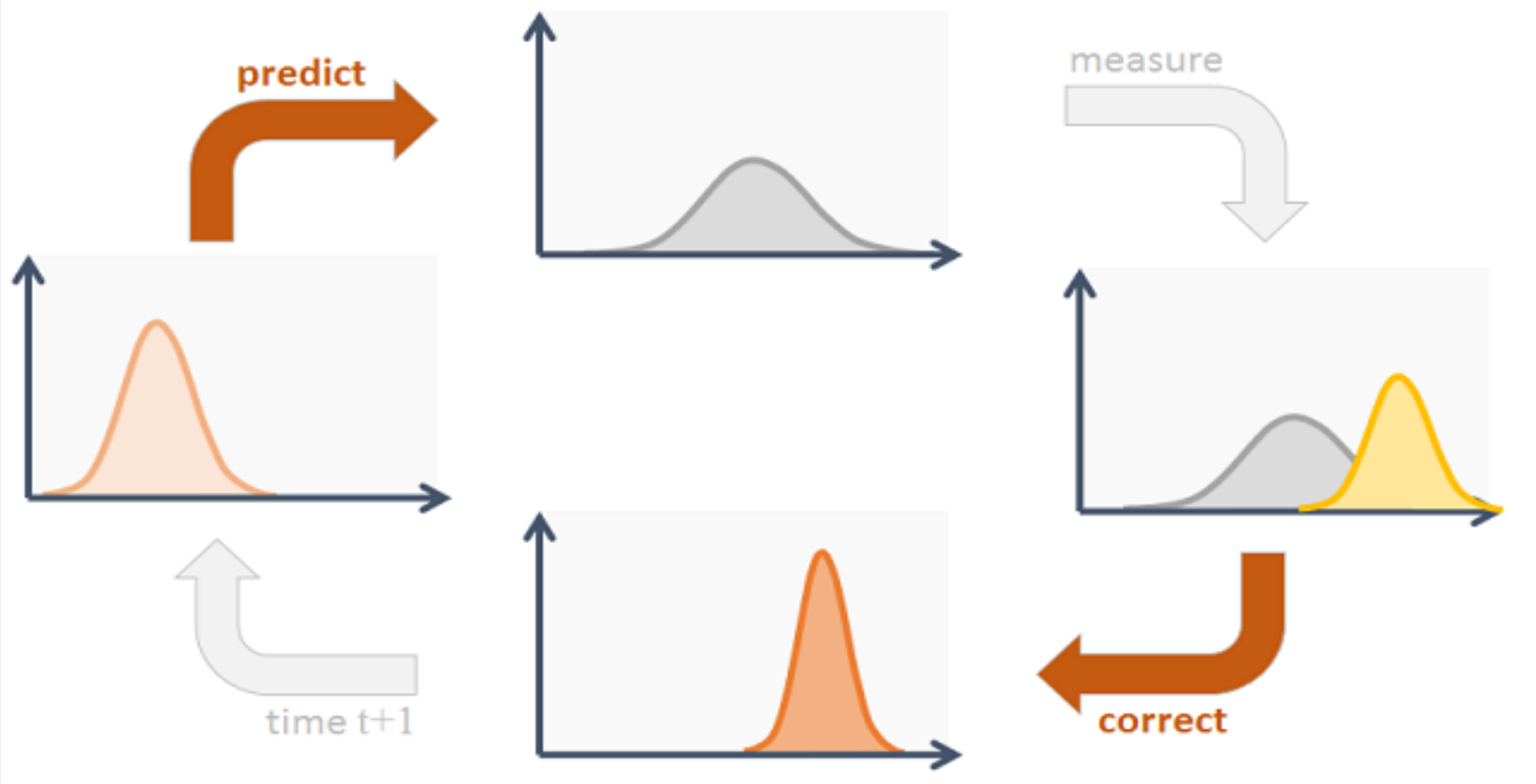
Drawbacks:
Incorporates knowledge from only small number of prior states

Curse of dimensionality

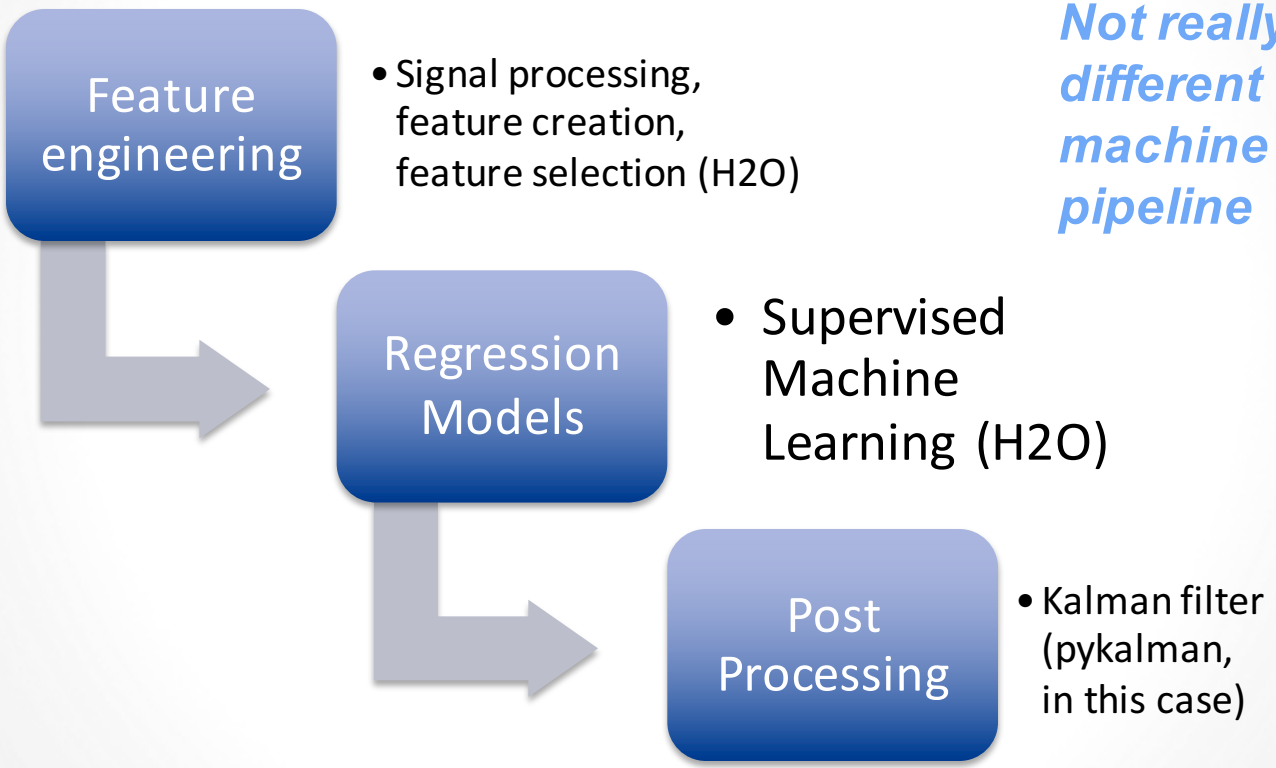
Another parameter (tau)

Disentangling the dynamic core: a research program for a neurodynamics at the large-scale
MICHEL LE VAN QUYEN
Biol. Res. v.36 n.1 Santiago 2003
http://dx.doi.org/10.4067/S0716-97602003000100006

KALMAN FILTER

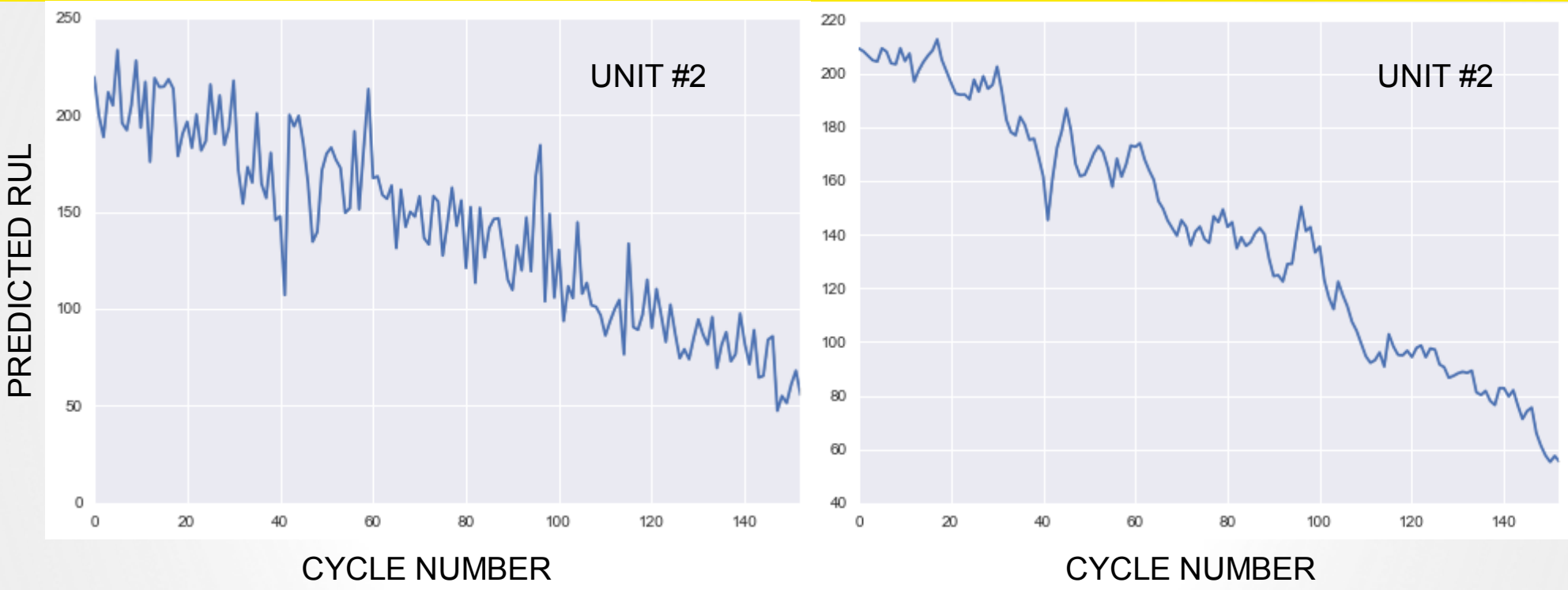


FINAL PIPELINE



Not really that much different than typical machine learning pipeline

BEFORE AND AFTER POST-PROCESSING



This presentation, data, and associated Jupyter notebooks (look in 2016_04_13_AppliedMachineLearningForIoT): https://github.com/h2oai/h2o-meetups/

FLEET TELEMATICS: *PREVENTIVE MAINTENANCE*

"Annual Savings are **$7M**"
– Anonymized, Member Technical Staff

Leading Mobile Telecom Operator

PROBLEM

- Fleet telematics—analyze maintenance records and vehicle performance to make predictions on when to do preventive maintenance
- Couldn't scale by sampling data
- Took days to create models

WHY H2O

- H2O support for customer's Kerberos authentication mechanism for Hadoop
- Support for MapReduce, YARN, R, Python and Spark in Hadoop
- In-memory, distributed architecture
- Rapid deployment to production with POJO
- Quick prototyping with H2O Flow

IMPACT

- When you look at the cost of towing a stranded vehicle, technician loss of productivity, and the customer lifetime value, the annual savings is $7M."
  – Anonymized, Lead Member Technical Staff

Telecommunications

TAKE-AWAYS

- Know your “physics”
  - ”Physics” is will be like a fish finder in this sea of Brontobytes
  - Linear dynamic systems, Queues, Signals, etc.
- Big data *and* Big models *and* Small models
- Keeping and eye on promising new methods (e.g., Convolution Neural Nets and LSTM-RNN)

H2O RESOURCES

- Download and go: http://www.h2o.ai/download
- Documentation: http://docs.h2o.ai/
- Booklets, Datasheet: http://www.h2o.ai/resources/
- Github: http://github.com/h2oai/
- Training: http://learn.h2o.ai/

THANK YOU