Big Data & Machine Learning: What Does it Mean and How Do You Use it?

Ankur Teredesai, PhD – Chief Technology Officer, KenSci, Inc.
Whende M. Carroll, MSN, RN-BC – Director of Nursing Informatics, KenSci, Inc.

Agenda

- **The Basics of Big Data & Machine Learning: Definition, Type and Steps**

- **Case Study: Machine Learning Impacting Patient Care and Nursing Workforce**

- **Using Machine Learning in Healthcare**

- Getting Started Using Machine Learning in Your Organization

Questions we will answer together:

- What is Big Data and Machine Learning and why do we need them?
- What does this mean for our patients?
- How do we use it ? How does this impact me?
- Hope vs Hype. How do we know when it is real?
- Are we ready and how do we get started?

ABOUT ME | DR. ANKUR TEREDESAI

Ankur Teredesai, Ph.D.

@ankurt

Ankur@kensci.com

Co-Founder
KenSci
May 2015 – Present • 2 yrs 5 mos

Professor
University of Washington
Sep 2014 – Present • 3 yrs 1 mo

Information Director
SIGKDD
Aug 2011 – Present • 6 yrs 2 mos

Executive Director - Center for Data Science
University of Washington
Sep 2010 – Present • 7 yrs 1 mo
Greater Seattle Area

Research
IBM
May 2001 – Sep 2001 • 5 mos
Yorktown Heights, NY

Research Intern
Microsoft
May 2000 – Sep 2000 • 5 mos
Redmond, WA

11 Publications

Visual Tracking via Supervised Similarity Matching • Readmission Score as a Service(RaaS) • AMADEUS: A System for Monitoring Water Quality Parameters and Predicting Contaminant Paths - See more at: http://cwds.uw.edu/amadeus-system-monitoring-water-quality-parameters-and-predicting-contaminant-paths#sthash.4VyyT3a2.dpuf • Risk-O-Meter: an intelligent clinical risk calculator • Audience segment expansion using distributed in-database k-means clustering • CoMMA: A Framework for multimedia mining using multi relational associations • ACM SIGSPATIAL GIS Cup 2012 • ACM SIGSPATIAL GIS Cup 2012 • Computing Fuzzy Rough Approximations in Large Scale Information Systems • HealthSCOPE: An Interactive Distributed Data Mining Framework for Scalable Prediction of Healthcare Costs…

University at Buffalo
Ph.D., Computer Science
1998 – 2001

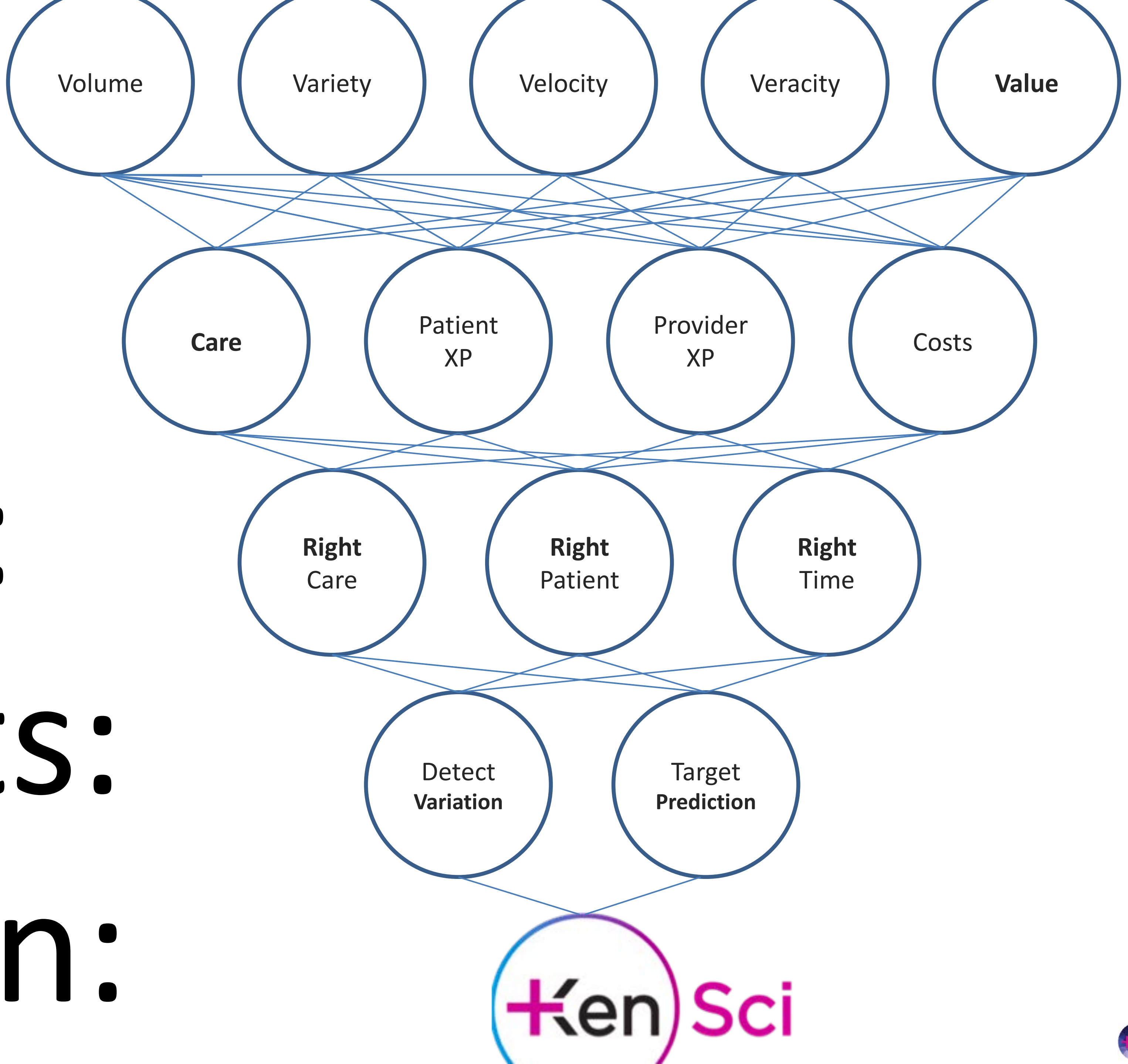
5 Vs:

4 Aims:

3 Rights:

2 Insights:

1 Mission:



KENSCI COMPANY PROFILE

MISSION

DEATH VS. DATA SCIENCE

Risk Prediction for Healthcare. Powered by Machine Learning.

RESEARCH PARTNERS

W UNIVERSITY of WASHINGTON

Microsoft Research

Microsoft Accelerator

TEAM

- DOCTORS & NURSES
- DATA SCIENTISTS
- DEVELOPERS

INVESTORS

ignition

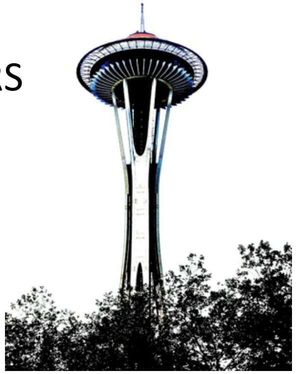
OSAGE UNIVERSITY PARTNERS

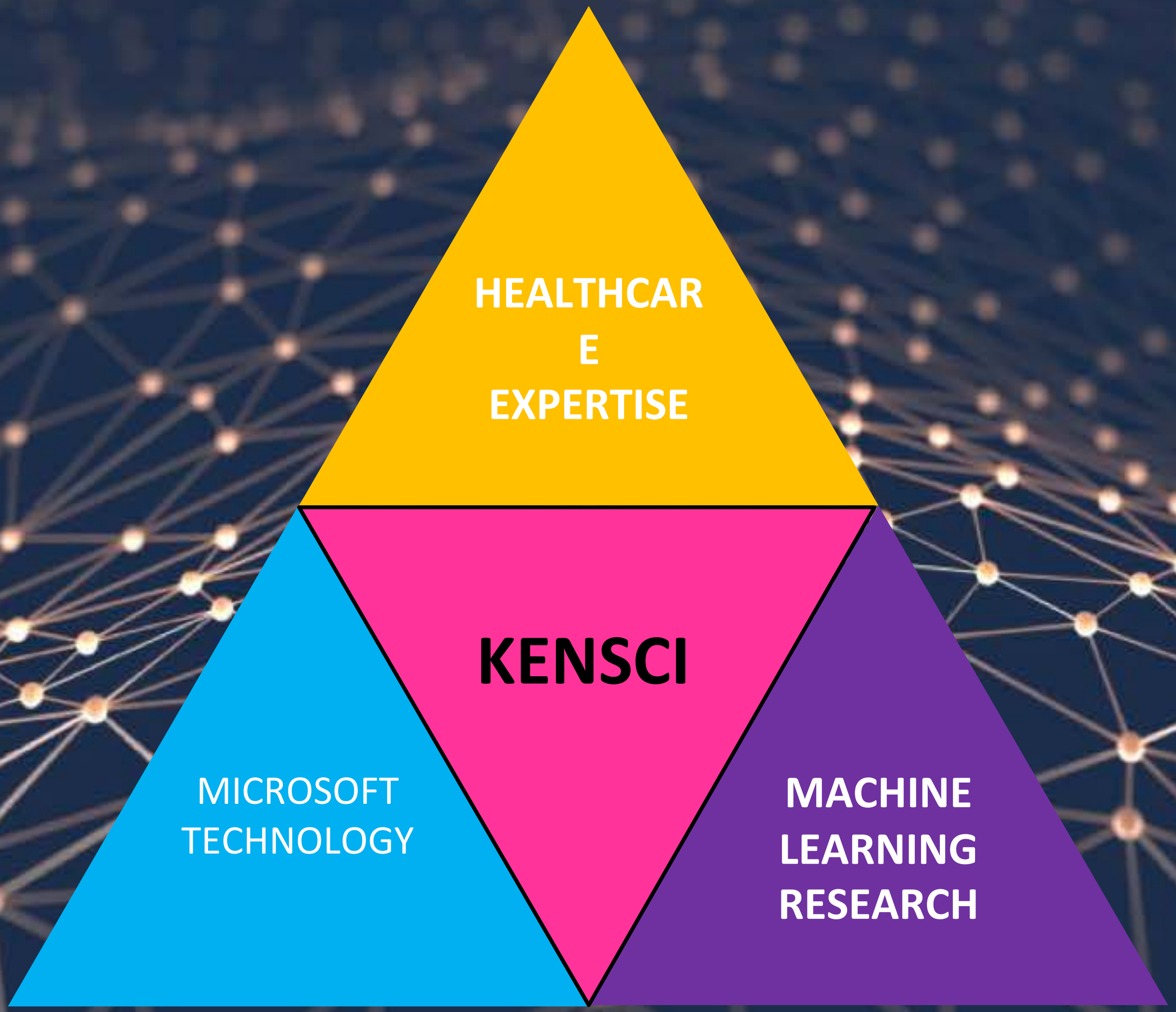
mindset ventures

HEADQUARTERS

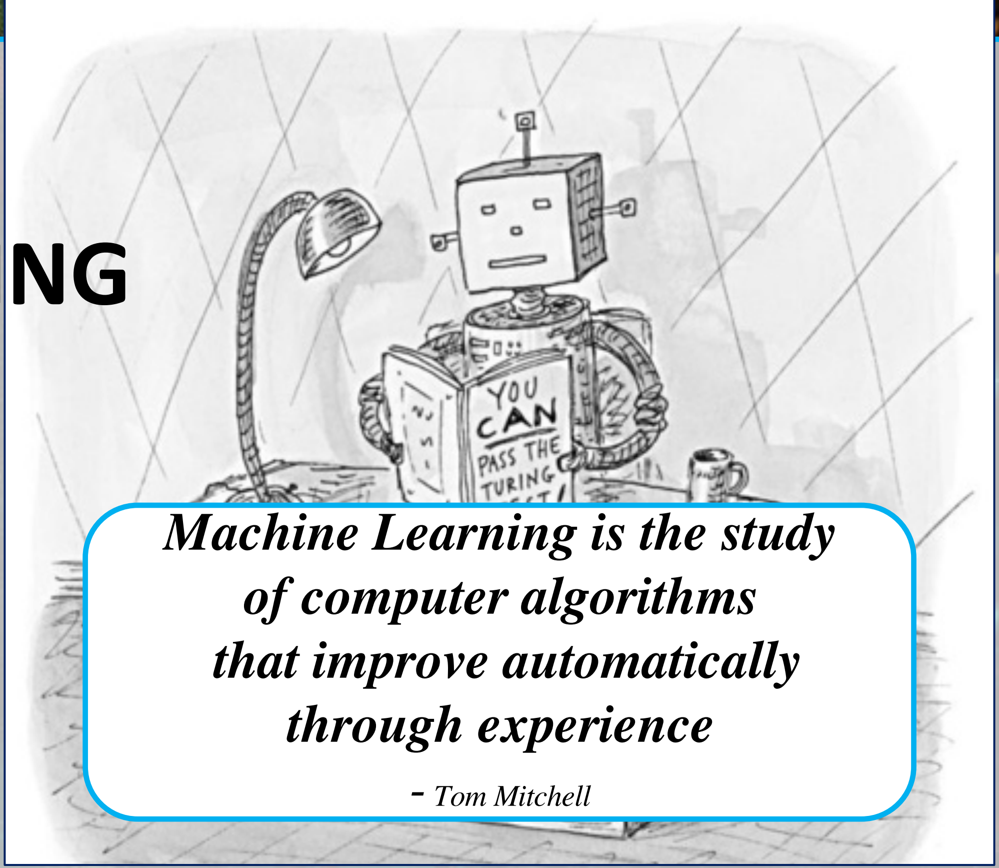
IN

SEATTLE



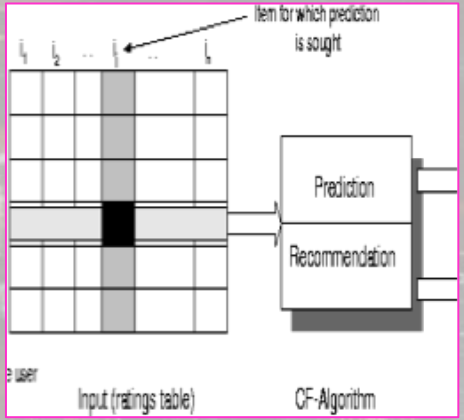

HEALTHCAR
E
EXPERTISE
KENSCI
MICROSOFT
TECHNOLOGY
MACHINE
LEARNING
RESEARCH

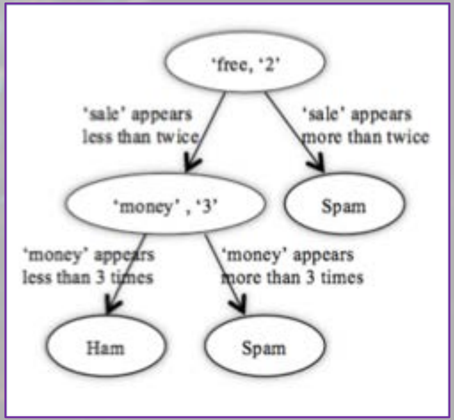
MACHINE LEARNING is Almost everywhere*

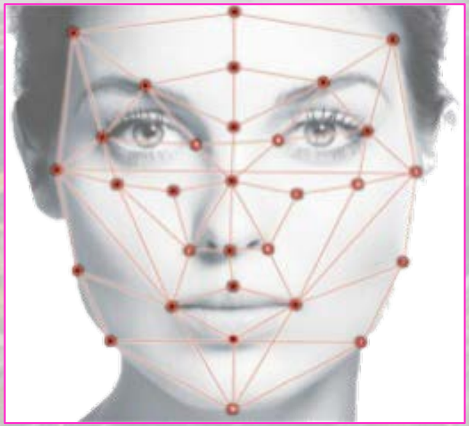


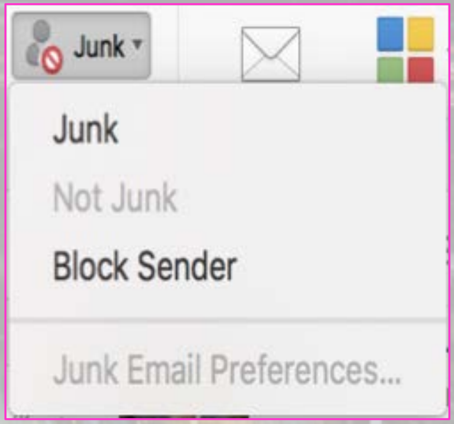
Machine Learning is the study of computer algorithms that improve automatically through experience

- Tom Mitchell

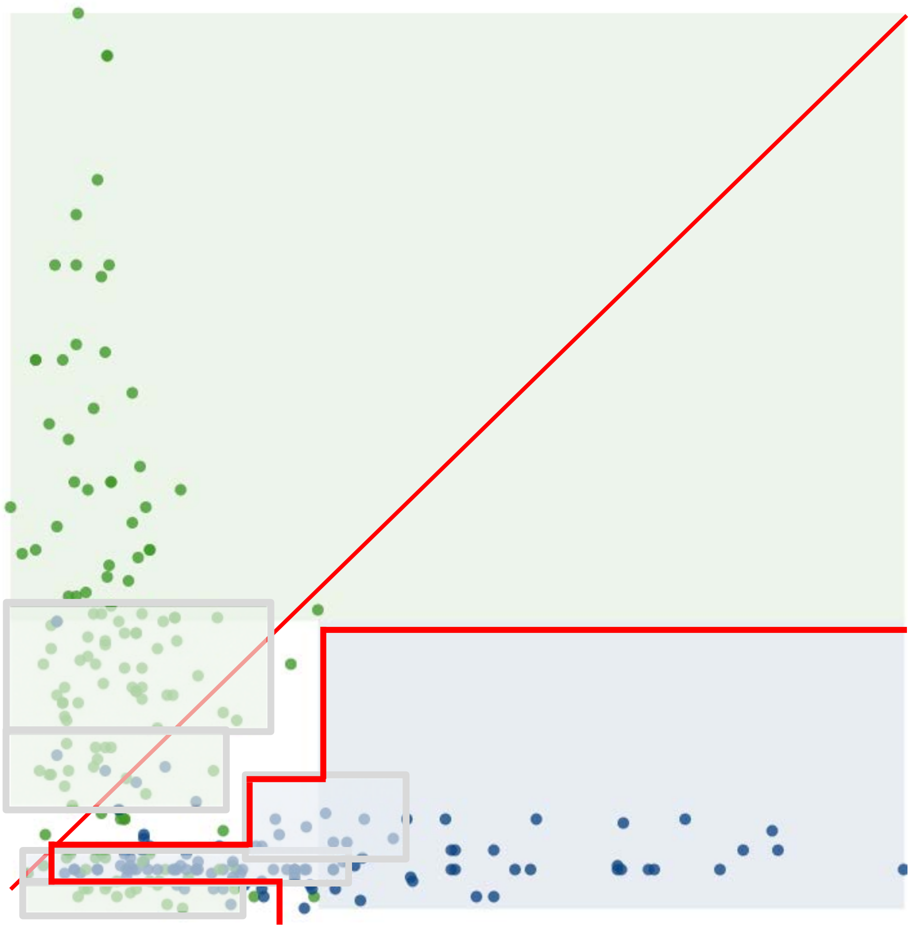


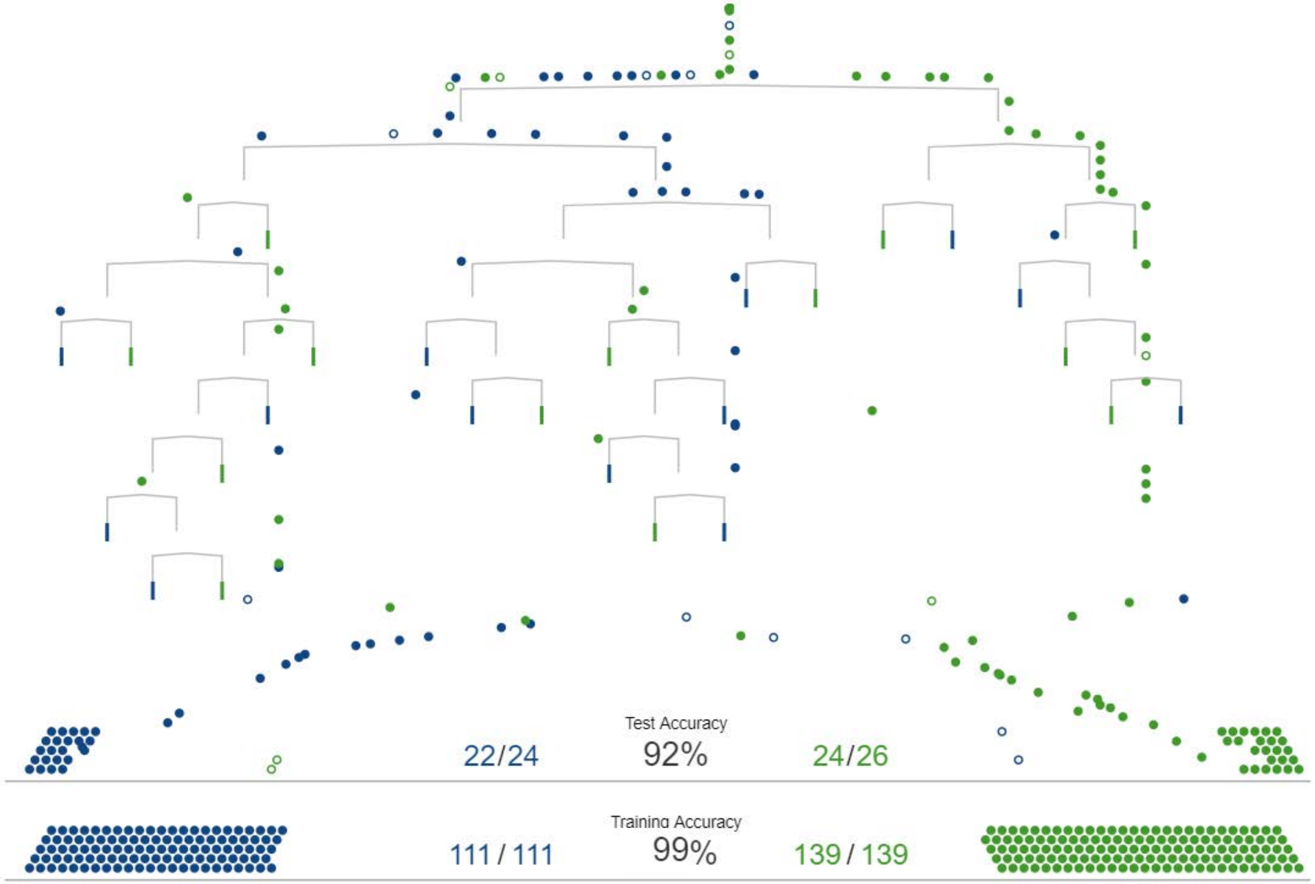






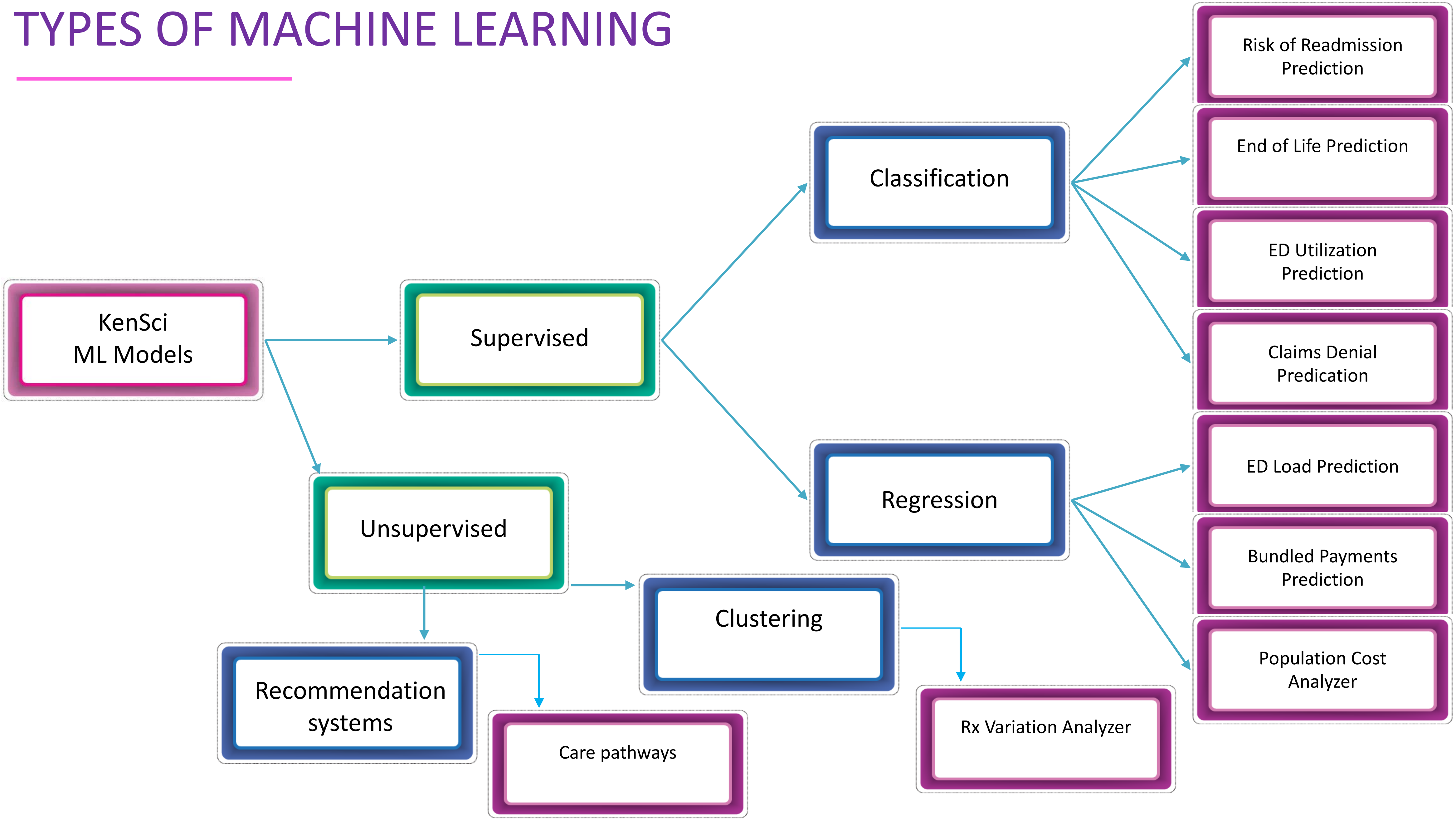
AN INTUITIVE WAY TO UNDERSTAND MACHINE LEARNING





Source : http://www.r2d3.us/visual-intro-to-machine-learning-part-1/

TYPES OF MACHINE LEARNING



STEPS TO PREDICTIVE MODELING

Get Data

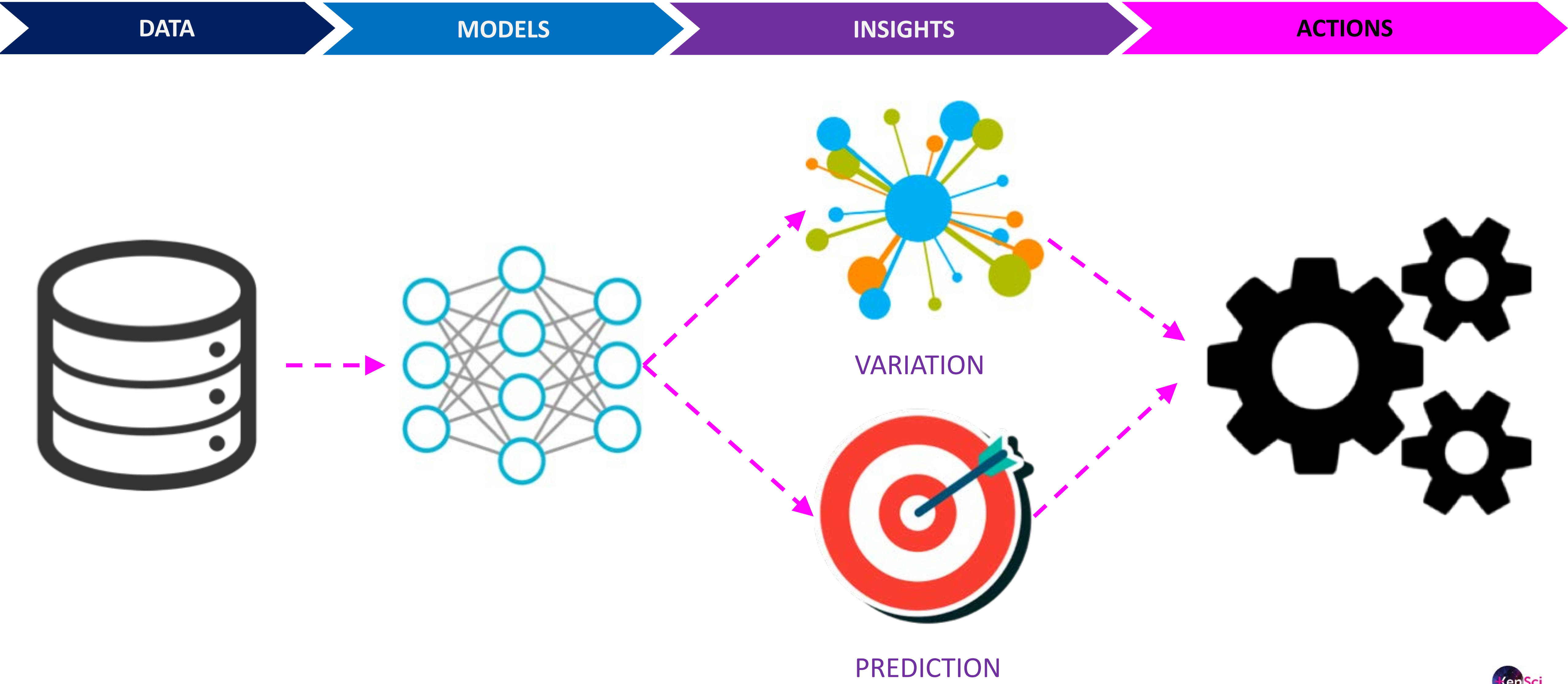
Clean, Prepare & Manipulate Data

Train Model

Test Data

Improve

STEPS IN ML: DATA to INSIGHT to IMPACT



ABOUT ME

| WHENDE M. CARROLL

Whende M. Carroll, MSN, RN-BC

@whendemcarroll

whende@kensci.com

Director of Nursing Informatics
KenSci
Mar 2017 – Present • 10 mos

Subject Matter Expert | Course Evaluator - Nursing Informatics
Western Governors University
Oct 2016 – Present • 1 yr 3 mos

Registered Nurse, Data Scientist
Vera Whole Health
Nov 2015 – Mar 2017 • 1 yr 5 mos
Seattle, Washington

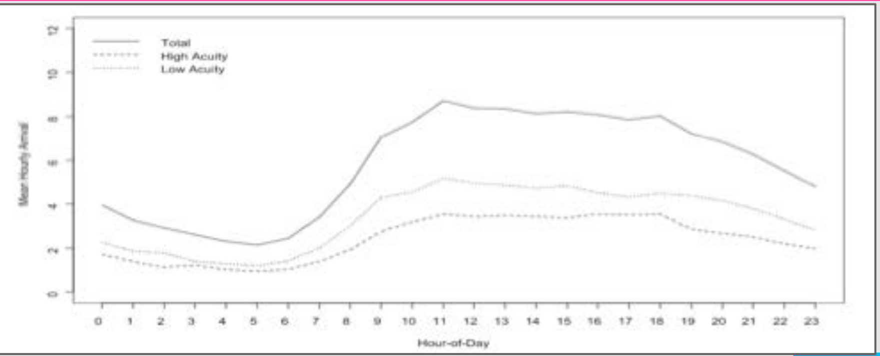
RN - Performance Improvement Project Manager / RN Informaticist
CHI Franciscan Health
Nov 2010 – Nov 2015 • 5 yrs 1 mo

Walden University
Master of Science, Nursing/Informatics

KenSci EvergreenHealth Microsoft

“Can our data help us predictively match ED staffing to demand?”

“ESPECIALLY ABOUT THE FUTURE…”



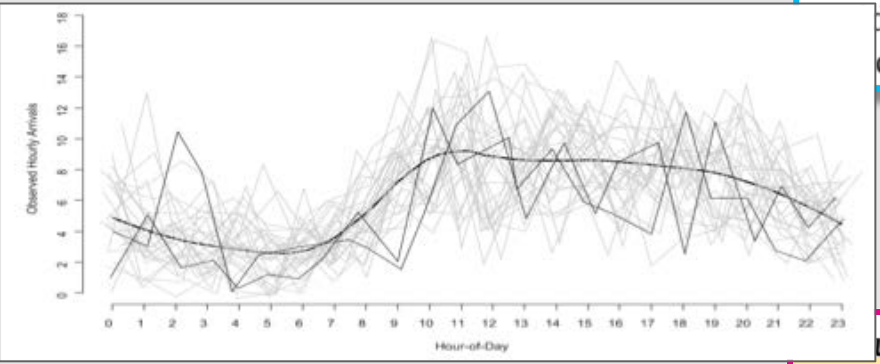
CLINICAL PRACTICE

Forecasting Daily Patient Volumes in the Emergency Department

Spencer S. Jones, MStat, Alun Thomas, PhD, R. Scott Evans, PhD, Shari J. Welch, MD, Peter J. Haug, MD, Gregory L. Snow, PhD

Conclusions: This study confirms the widely held belief that daily demand for ED services is characterized by seasonal and weekly patterns. The authors compared several time series forecasting methods to a benchmark multiple linear regression model. The results suggest that the existing methodology proposed in the literature, multiple linear regression based on calendar variables, is a reasonable approach to forecasting daily patient volumes in the ED. However, the authors conclude that regression-based models that incorporate calendar variables, account for site-specific special-day effects, and allow for dual autocorrelation provide a more appropriate, informative, and consistently accurate approach to casting daily ED patient volumes.

Good News:
Prediction is possible…



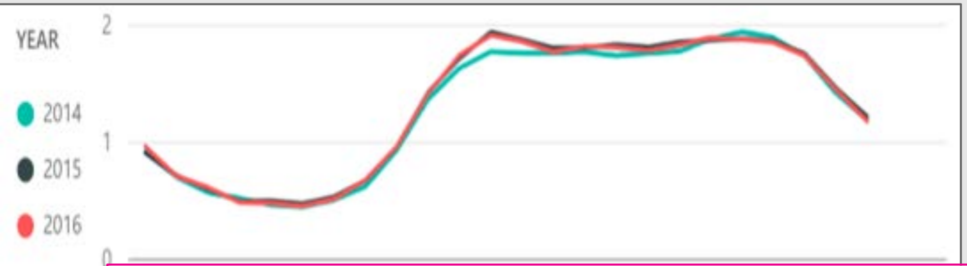
The Challenge of Predicting Demand for Emergency Department Services

Melissa L. McCarthy, MS, ScD, Scott L. Zeger, PhD, Ru Ding, MS, Dominik Aronsky, MD, PhD,

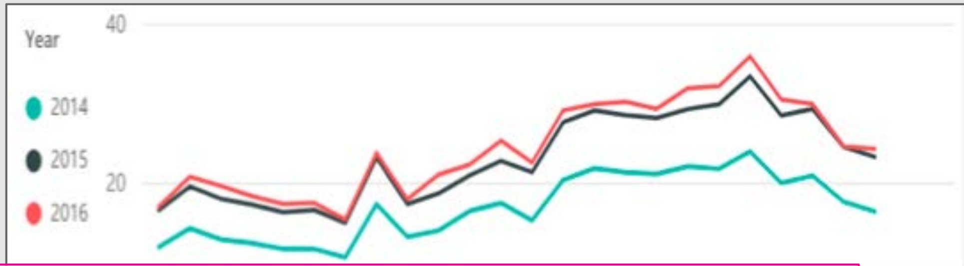
ults: Hourly ED arrivals were obtained for 8,760 study hours. Separate models were fit for high- versus low-acuity patients because of significant arrival pattern differences. The variance was approximately equal to the mean in the high- and low-acuity models. There was no residual autocorrelation ($r = 0$) present after controlling for temporal, climatic, and patient factors that influenced the arrival rate. The observed hourly count fell within the 50 and 90% prediction intervals 50 and 90% of the time, respectively. The observed histogram of arrival counts was nearly identical to the histogram predicted by a Poisson process.

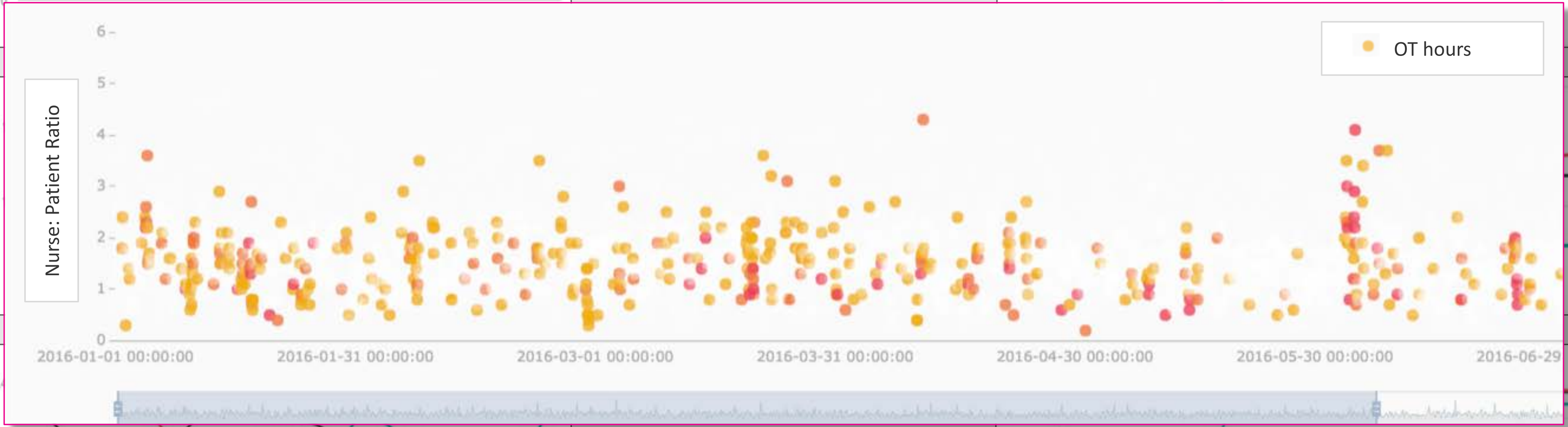
Less Good News:
Prediction is hard…

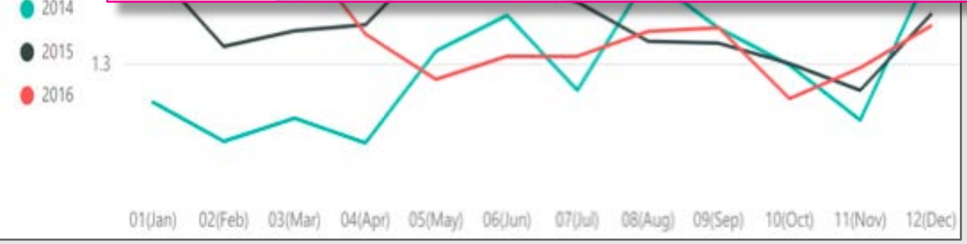
& STAFFING IS EVEN HARDER…



Hourly







Monthly

Evergreen Health: What We've Done With ML

Predict Emergency Department patient demand to lower costs, specifically nurse staffing

Situation

Evergreen's EDs often suffer from overcrowding, leading to poor patient and staff satisfaction

Unpredictable patient demand leading to the use of high cost staff and increased OT utilization

Tactics

Overlay multiple sources of data: EHR, historical traffic patterns with social and public data for weather, pollen, holidays

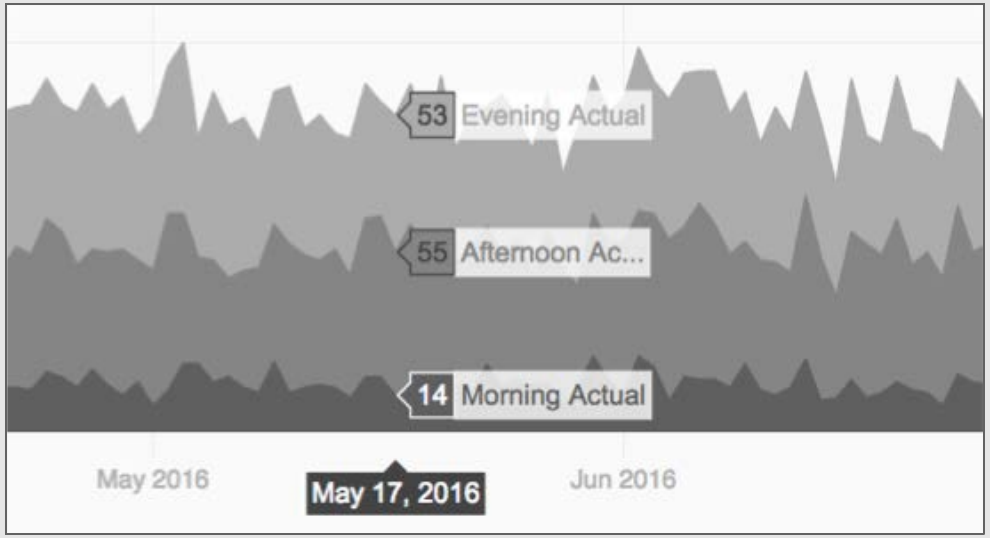
Use enriched data set to predict patient demand 2, 4
and 8 hours before

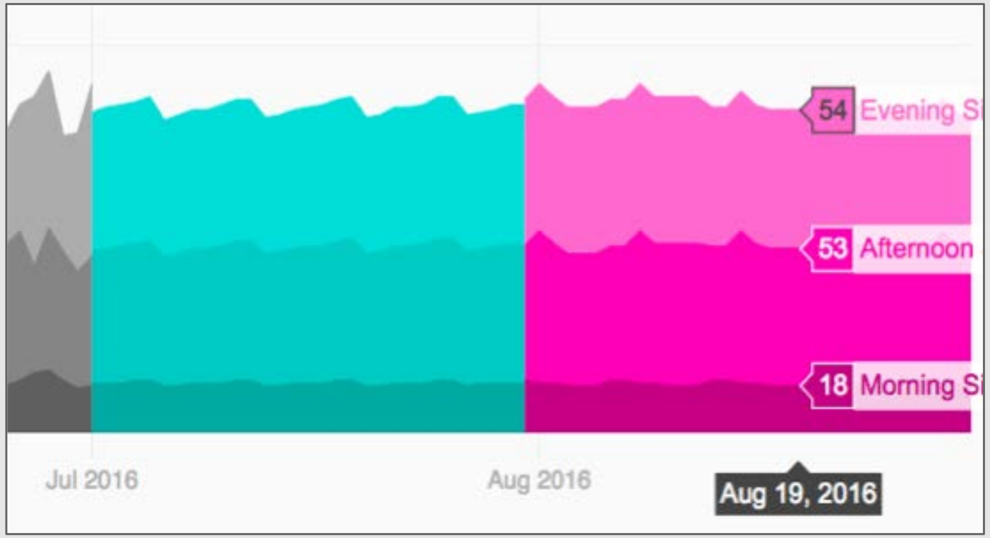
What Happened

- ✓ Weekly and monthly demand pattern predictions to optimize nursing staffing requirements
- ✓ ED leadership adjustment of staffing schedules and shifts based demand predictions

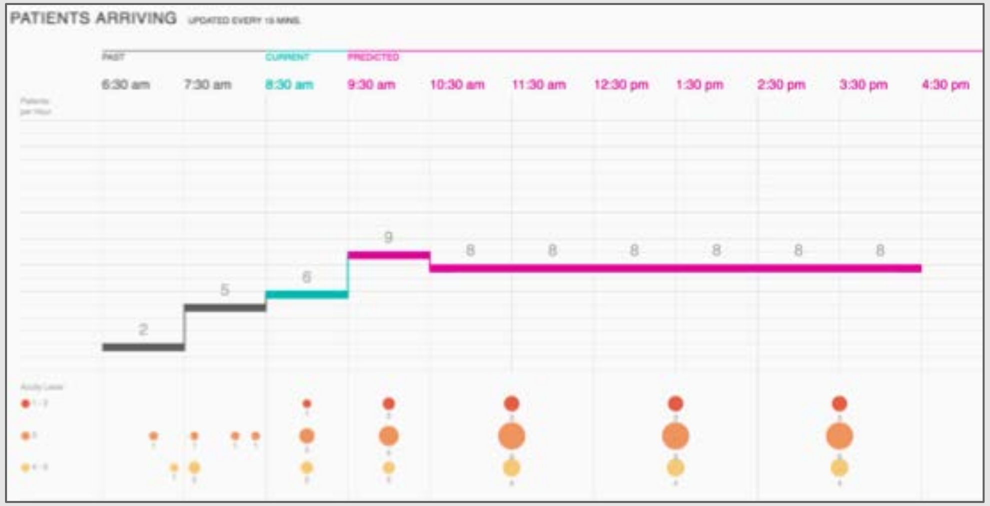
EvergreenHealth

From Noise to Signal…





Admin for monthly strategic planning



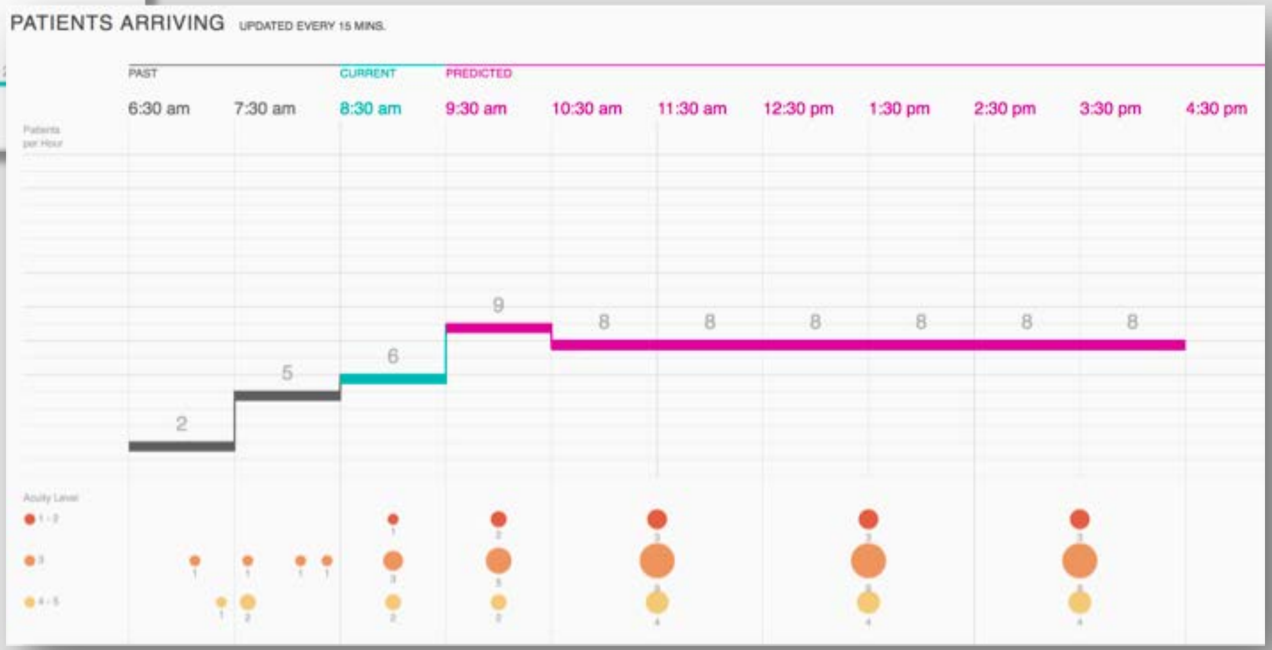
Ops for hourly tactical response

Prediction vs Historic Trend

1 hour	13.50%
2 hour	14%
4 hour	15%
6 hour	8.5%
8 hour	15.5%
1 month	17.3%
6 months	17.3%

From Noise to Signal…

Short-term, Real-time Forecast
= For hourly, tactical response



Long-term Forecast
= For monthly strategic planning

What does this mean for nurse staffing?

Less

- Waste
- Inappropriate nurse to patient ratios
- Staff burnout

More

- Productivity
- Time spent with patients for high-value interaction
- Staff satisfaction

Lower Costs

- **Control use of overtime and high-wage workers**
- **Lower Turnover**
- **Higher Retention**

"It's really exciting to find patterns in our massive data which lead to actionable change that can tangibly impact patients and their outcomes."

SVP and Chief Nursing Officer

EvergreenHealth

NCSBN: Use Cases to Consider

BONS

NCSBN

Variation 1:

The quality of nursing applicants?

Prediction 1:

Nurses who will be most fraudulent

Variation 2:

Nurses with violations?

Prediction:

Respondents likely to repeat serious errors

Evidence-based predictions =

Base licensure on multiple variables from several data sources to ensure quality and patient safety at the point of care

Reference: M Alexander. Oct. 2017. Precision Regulation. Journal Of

4 Steps to Applying ML in Healthcare

Identify an important problem

Get buy in on a data-driven approach

Move from variation to prediction to action

Deliver a quick win, rinse and repeat

KenSci

HEALTHCARE MACHINE LEARNING IS HARD

…and Healthcare is still the hardest part

ARE YOU READY FOR MACHINE LEARNING IN HEALTHCARE?

HERE'S A QUICK CHECKLIST TO EVALUATE YOUR READINESS

DATA DRIVEN DECISIONS

- ☐ Are nurse regulators and policy stakeholders convinced they need to make data driven decisions?
- ☐ Do we have use cases where data driven decisions can optimize the nursing workforce through regulation?

DATA READINESS

- ☐ Do you have a governance structure in place?
- ☐ Are you compliance ready?
- ☐ Way forward: Build vs Buy

ARE YOUR KEY STAKEHOLDERS CONVINCED?

Nurse Regulators / Policy Makers


The user


The critic


The navigator

BUILDING USE CASES FOR EACH OF THE STAKEHOLDERS

ARE YOU READY?

EVALUATING YOUR PARTNERS

FRAMEWORK TO EVALUATE ML VENDORS

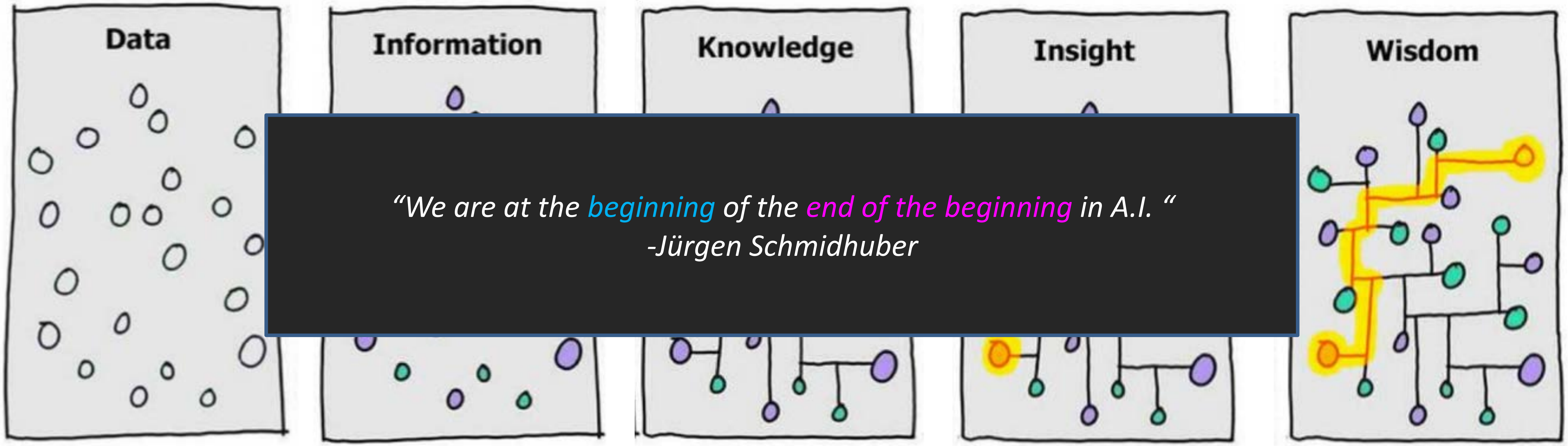
- ☐ Humility
- ☐ Expertise
- ☐ Bandwidth
- ☐ Platform

POWERING THROUGH

- ☐ Is the organization ready to change?
- ☐ Are your nurse regulators/policy makers supportive? Committed? Driven?
- ☐ Are you ready for immersion in 1-2 “quick win” use cases?
- ☐ Do you need a one stop solution or a transformational journey

ARE YOU READY TO TRANSFORM THE CARE CONTINUUM?

Start with a quick win



"We are at the beginning of the end of the beginning in A.I. "
-Jürgen Schmidhuber

Continuously Learn

Thank you for listening.

Download peer reviewed and published research pape http://bit.ly/kensci-research

Download Checklist to access your readiness for AI → http://bit.ly/kensci-guide

Learn more about Kensci → http://bit.ly/kensci-brochure