Machine Learning Techniques in Manufacturing

Applications & Caveats

Thomas Hill, PhD | Exec. Director Analytics | @DrTomHill

Overview

- Business drivers: Machine learning, big data, and a new paradigm for analytics in manufacturing
  - Who drives the projects?
- Machine learning vs. statistical analysis of data for manufacturing support
- Discussion of selected approaches and algorithms
  - Recursive partitioning or trees
  - Neural networks and deep learning
  - Proposed approaches to validation
- Some use cases in manufacturing
- Barriers to adoption
- Summary

Key Points

1

Technology is driving competitiveness. Machine learning can improve quality, processes, and deliver competitive advantage.

2

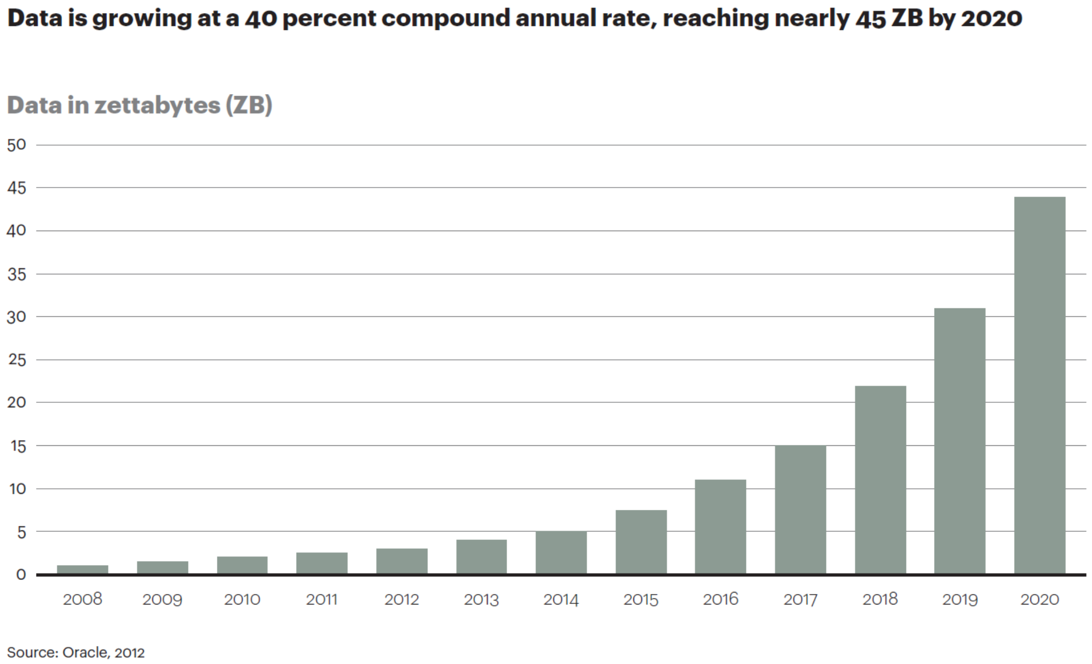
Machine learning algorithms can be "controlled." Best practices exist how to evaluate, validate, and interpret machine learning results.

3

Barriers can be overcome. Big-data-science skillsets are scarce; but configured role-based best-analytic practices will enable engineers, biologists, and Citizen Data Scientists to be successful.

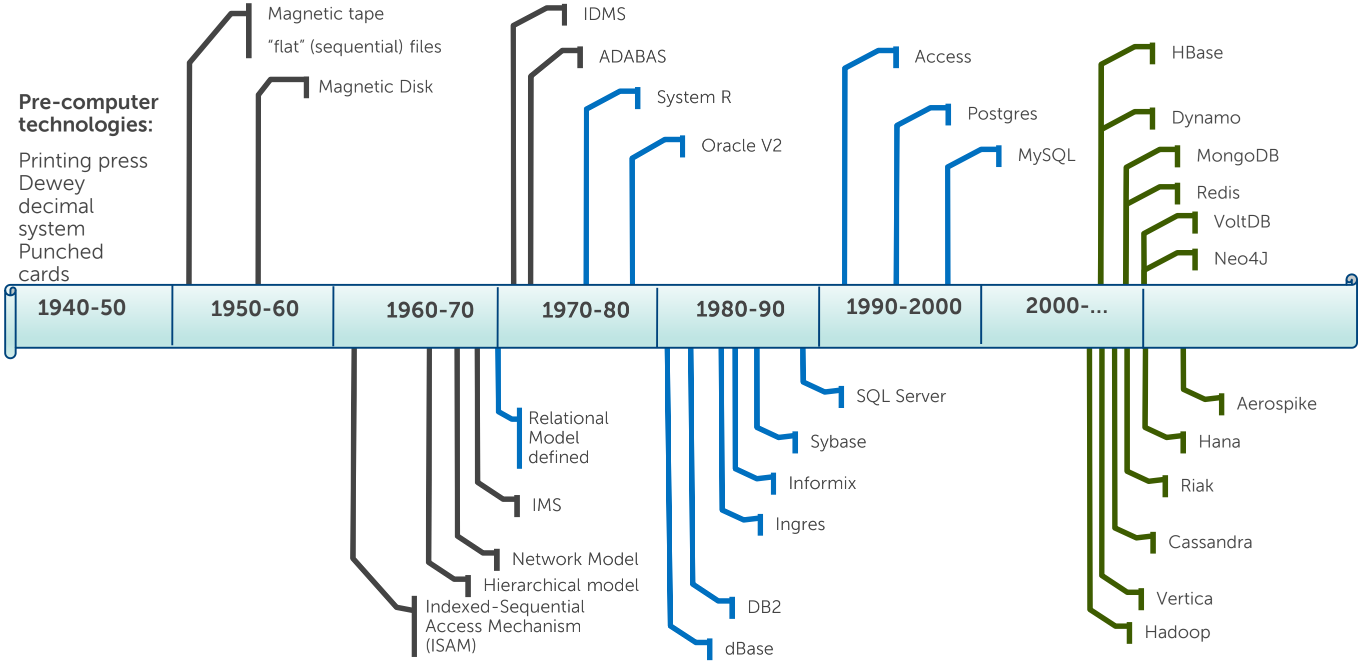
Drivers of Change

Data are Stored at Incredible Volumes and Rates



Statistica

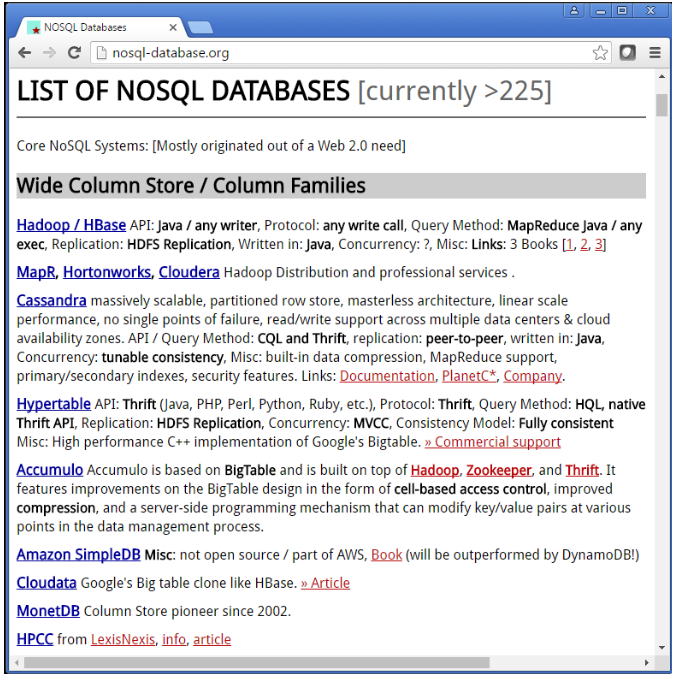
Paradigm Shift: Store Everything Now Because it May be Useful Later



Personal communication, Guy Harrison, Exec Dir, Prod/Solutions Mgmt, Dell Software, 2014

Paradigm Shift: Store Everything Now Because it May be Useful Later

Currently, there are **more than 225** NoSQL databases on the market



Statistica

The Next Wave: The Internet of Things (IoT)

50+ billion

"Things" to manage across the globe

$15 trillion

The amount the industrial Internet could add to **global GDP** over the next 20 years

- GE

"While smart homes and cool personal gadgets get the press, non-glamour applications get the traction. **Connected world will come first to warehouses, trucks, factories, and farms**"

- Forrester

The worldwide market for IoT solutions will grow to

$7.1 trillion in 2020

- IDC

By 2020, there will be over 100 million

Internet connected wireless light bulbs and lamps worldwide

up from 2.4 million in 2013

- OnWorld

55%

of discrete manufacturers are researching, piloting or in production with **IoT** initiatives

Service revenues for the IoT will reach:

$500 Billion by **2018**

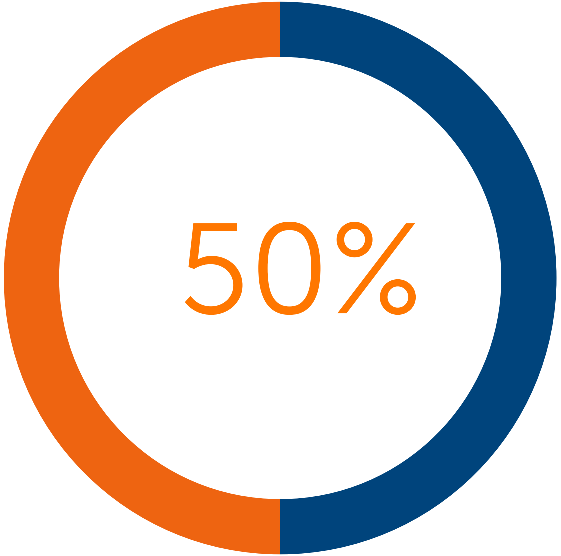
– Harbor Research

The number of IoT developers grow from 300k in 2014 to:

4.5 Million by 2020

– Vision Mobile

Organizations actively using data grow 50% faster than laggards.



The number of organizations who understand the benefits of big data grew slightly.

39% → **42%**

(2014) (2015)

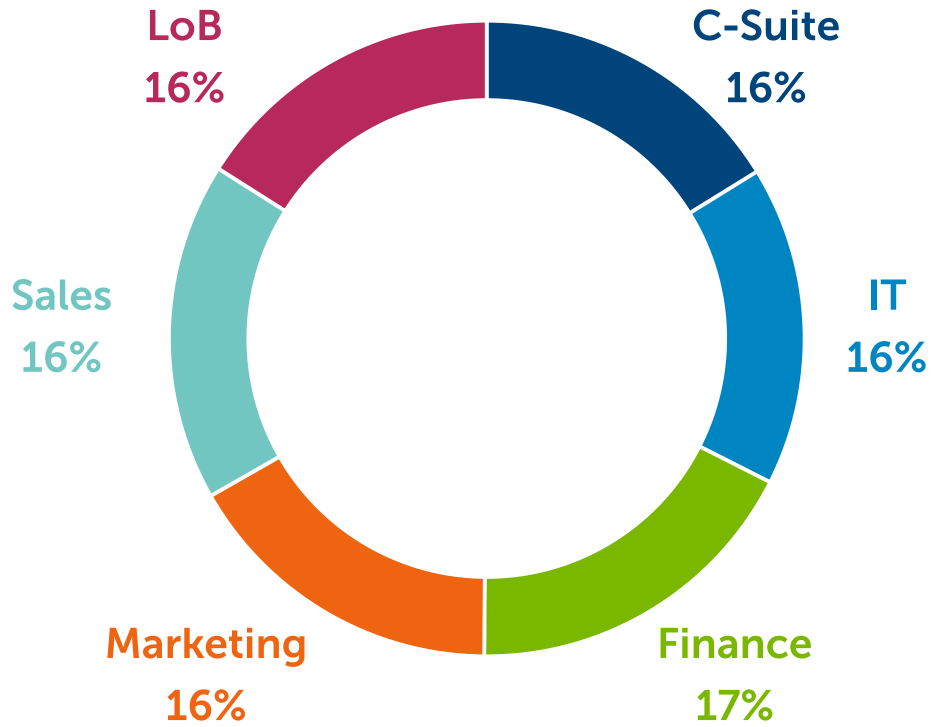
Source: 2015 Dell Global Technology Adoption Index

Statistica

By 2018 **more than half of large organizations around the globe** will compete using Advanced Analytics and proprietary algorithms, causing disruption on a grand scale.

Source: Gartner, Inc., Magic Quadrant for Advanced Analytics Platforms, Lisa Kart, Gareth Herschel, Alexander Linden, Jim Hare, 9 February 2016.

Only 16% of projects are championed by IT.



Source: 2015 Dell Global Technology Adoption Index

Statistica

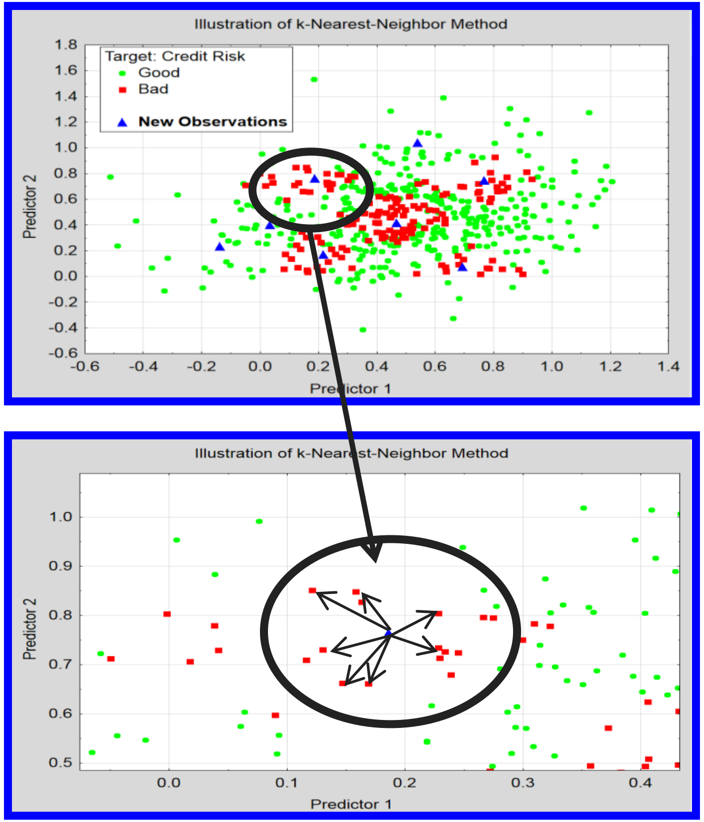
Machine Learning

(vs. Statistical Analysis)

Pattern Recognition Replacing Statistical Models

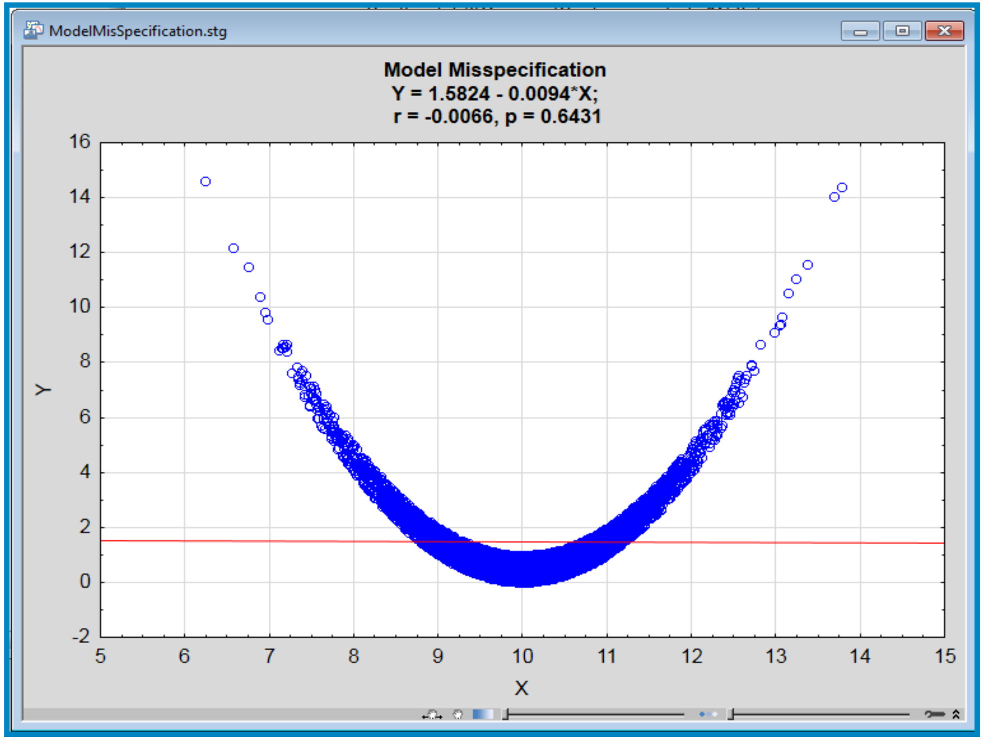
Knowledge Discovery vs. Statistical Analysis

- **Statistical Analysis**
  - Focuses on "hypothesis testing" and "parameter estimation"
  - Fits "parsimonious statistical models" with the goal to "explain" complex relationships with fewer parameters
  - Examples: Regression, PLS, PCA, nonparametric statistics, quality control
- **Pattern Recognition (Data Mining)**
  - **The data are your model!**
  - Algorithms include:
    - Trees, boosted & voted trees (forests), SVM, neural nets, ...
    - Deep learning, cognitive computing
  - Models are validated through
    - Statistics computed against test (hold-out) samples; Accuracy, MS Error, ROC, AUC, Lift,...
    - Established best practices
- See also: Leo Breiman (2001). Statistical Modeling: The Two Cultures *Statistical Science, 16*, 3, 199–231.



When statistical models are misspecified, you can get no results, wrong results

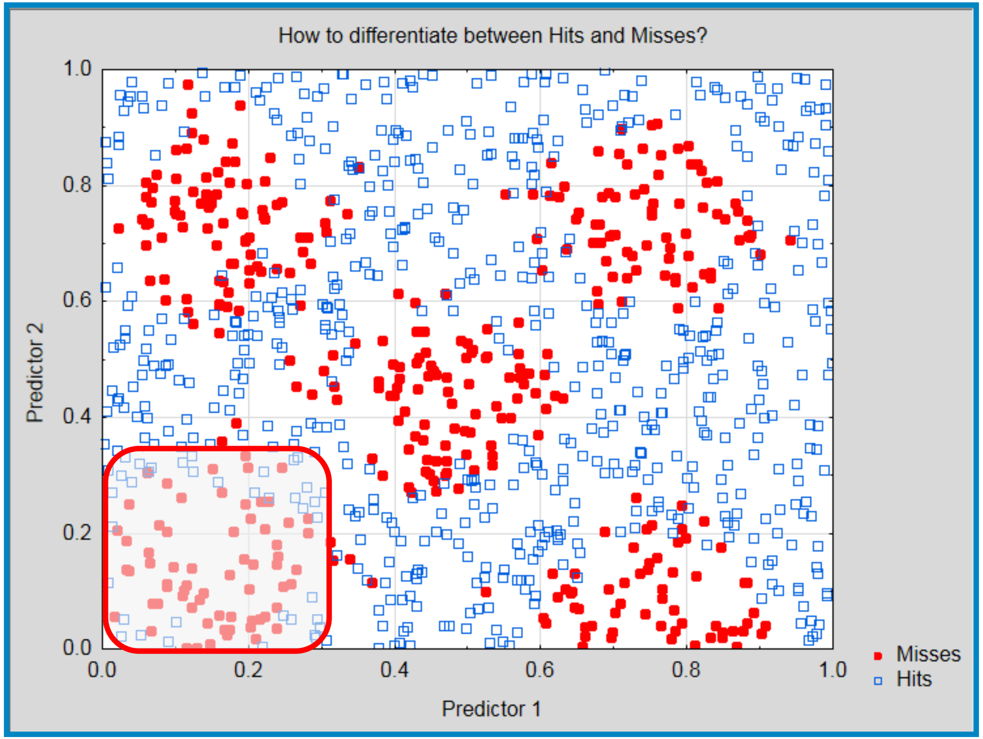
- Obviously, relationships can be
  - nonlinear
  - non-montonic
  - Non-additive, or with partial interactions



Dell Statistica

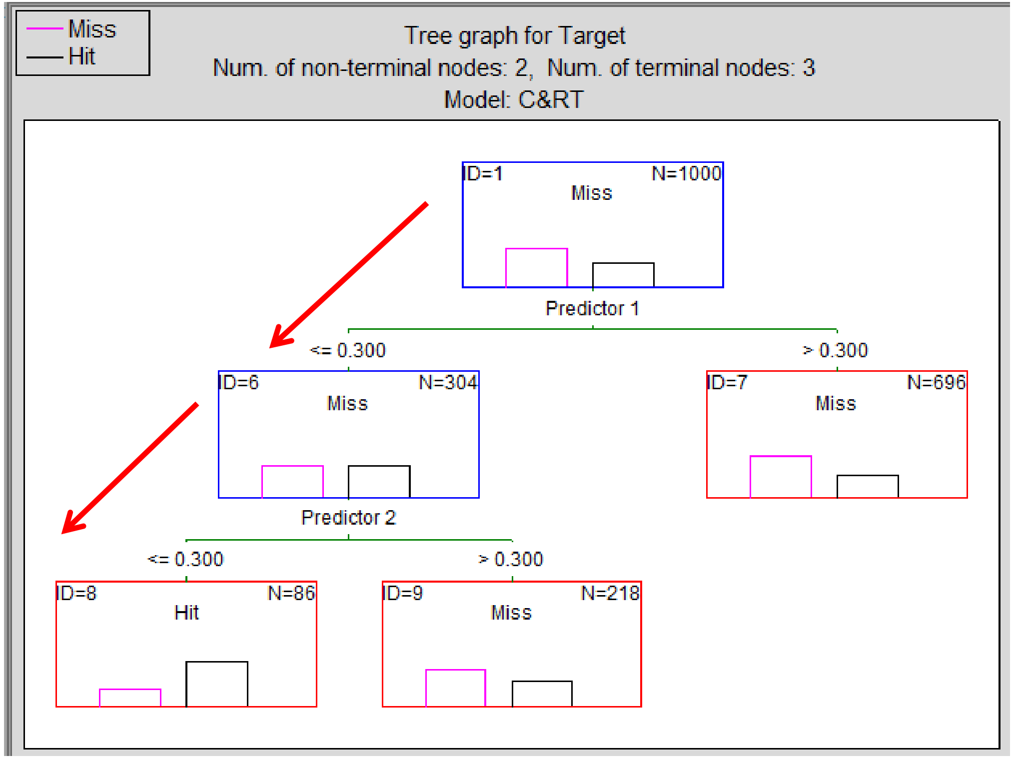
When statistical models are mis-specified, you may get no results, wrong results

- Relationships can be “difficult”
  - E.g., consider a linear model to predict Hits and Misses, when relationships are like this…
  - Machine learning will detect and model this relationship, and provide insight into the process

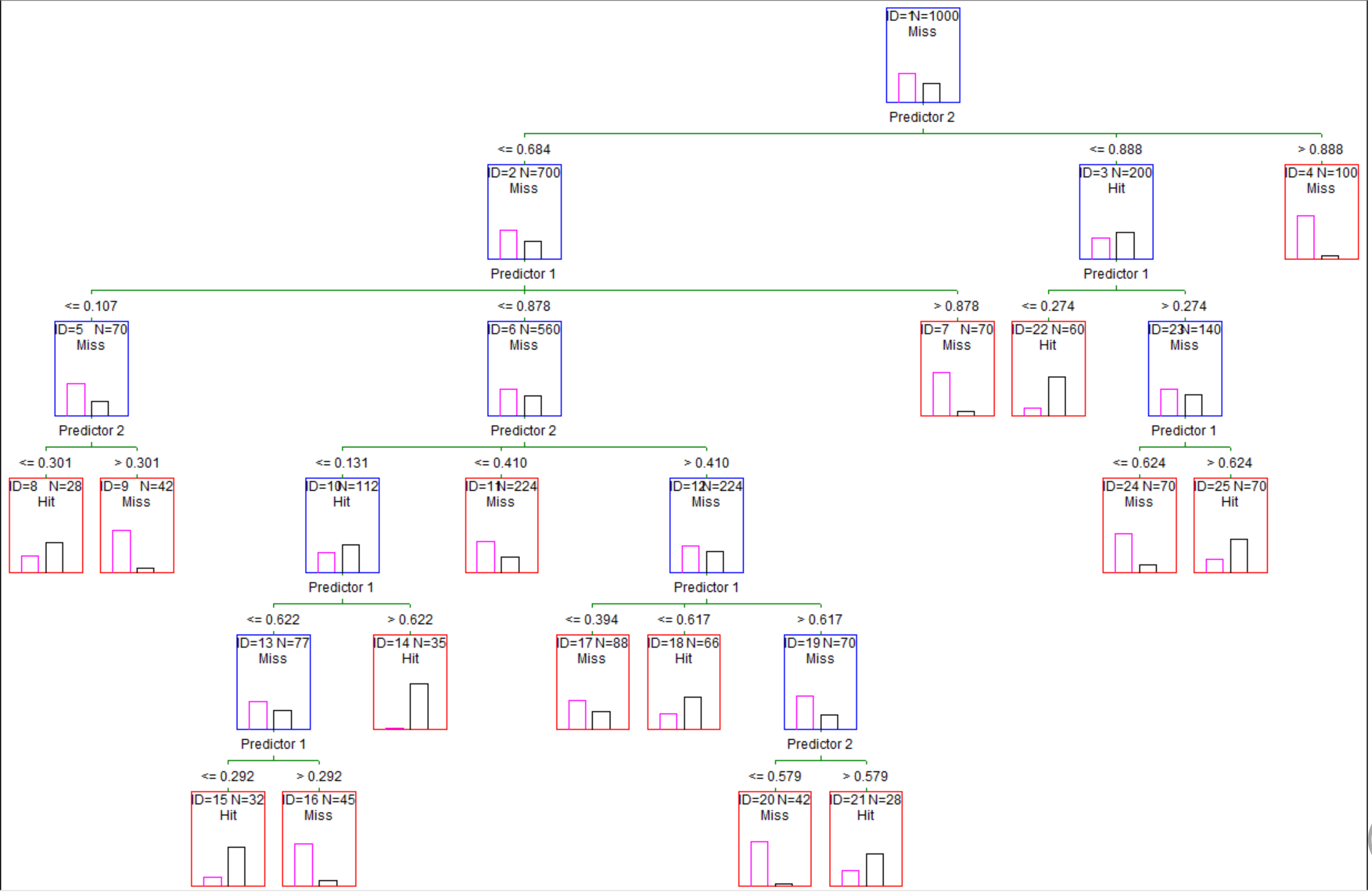


Tree algorithms will solve such problems

- CART, CHAID, etc. algorithms build will recursively partition data
- Goal is to find “pure nodes”



Statistica

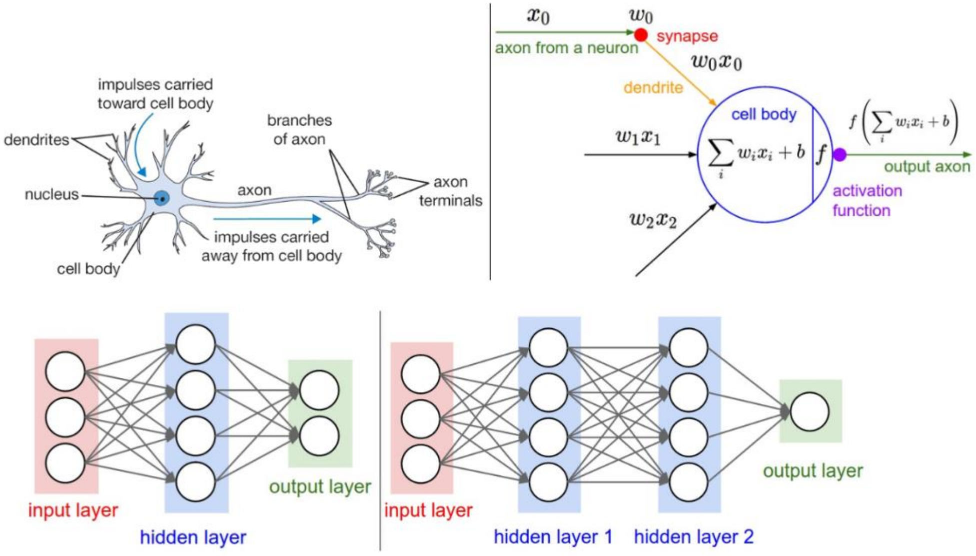

ID=1 N=1000
Miss
Predictor 2
<= 0.684
<= 0.888
> 0.888
ID=2 N=700
Miss
Predictor 1
ID=3 N=200
Hit
Predictor 1
ID=4 N=100
Miss
<= 0.107
<= 0.878
> 0.878
<= 0.274
> 0.274
ID=5 N=70
Miss
Predictor 2
ID=6 N=560
Miss
Predictor 2
ID=7 N=70
Miss
ID=22 N=60
Hit
ID=23 N=140
Miss
Predictor 1
<= 0.301
> 0.301
<= 0.131
<= 0.410
> 0.410
<= 0.624
> 0.624
ID=8 N=28
Hit
ID=9 N=42
Miss
ID=10 N=112
Hit
Predictor 1
ID=11 N=224
Miss
ID=12 N=224
Miss
Predictor 1
ID=24 N=70
Miss
ID=25 N=70
Hit
<= 0.622
> 0.622
<= 0.394
<= 0.617
> 0.617
ID=13 N=77
Miss
Predictor 1
ID=14 N=35
Hit
ID=17 N=88
Miss
ID=18 N=66
Hit
ID=19 N=70
Miss
Predictor 2
<= 0.292
> 0.292
<= 0.579
> 0.579
ID=15 N=32
Hit
ID=16 N=45
Miss
ID=20 N=42
Miss
ID=21 N=28
Hit

Statistica

Deep Learning becoming popular

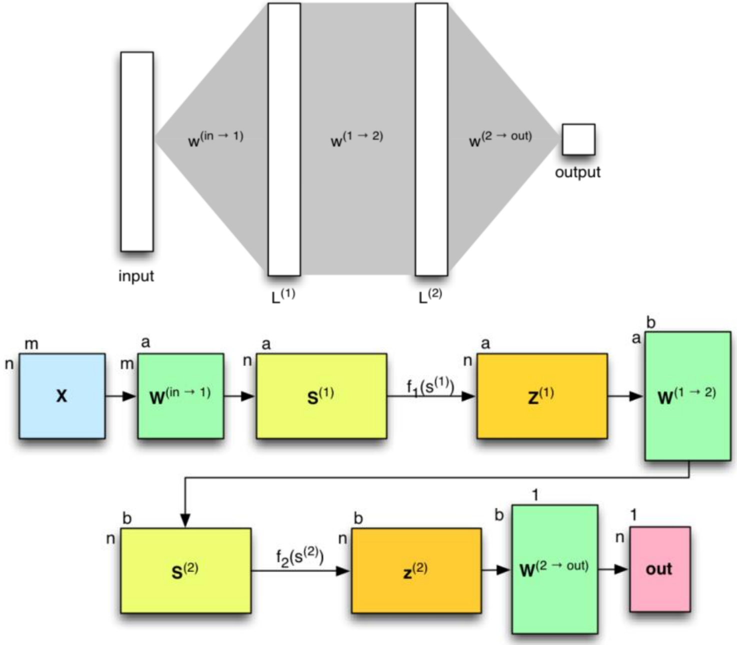
- Neural nets express some output as a nonlinear function (sets of equations) of some inputs

Biological Inspiration, Neurons



From: http://cs231n.github.io/neural-networks-1/
Extracted June 6, 2016

Flow of Vectors (*Tensors*)



From: http://briandolhansky.com/blog/2014/10/30/artificial-neural-networks-matrix-form-part-5
Extracted June 6, 2016

Google

TensorFlow

- Open source software library for machine intelligence
- “Tensors” are the vectors and matrices that are multiplied as data are transformed through multiple layers of a network
- These tensors/vectors can be thought of and arranged in a tree
- Google has built chips to perform tensor operations natively, presumably at an order of magnitude faster
- The “news are”: These computations, and the estimation of parameters/weights can be scaled to hundreds or thousands of processors in parallel

Paul King, Computational Neuroscientist, Data Scientist, Technology Entrepreneur
30k Views • Upvoted by Hadayat Seddiqi, engineering @ biotech startup and Bradley Voytek, Ph.D. neuroscience, UCSD Asst. Professor Cognitive Science
Most Viewed Writer in Neuroscience with 750+ answers

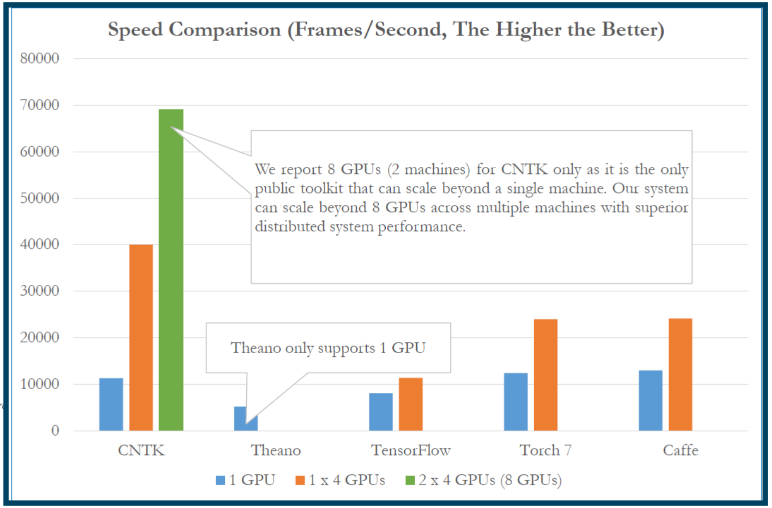
The most computationally intensive feed-forward network to date that does learning may be the visual object recognition network built by Le et al (2011) as a collaboration between Google and Andrew Ng's lab at Stanford. [1]

Their network comprised millions of neurons and 1 billion connection weights. They trained it on a dataset of 10 million 200x200 pixel RGB images to learn 20,000 object categories. The training simulation ran for three days on a cluster of 1,000 servers totaling 16,000 CPU cores. Each instantiation of the network spanned 170 servers.

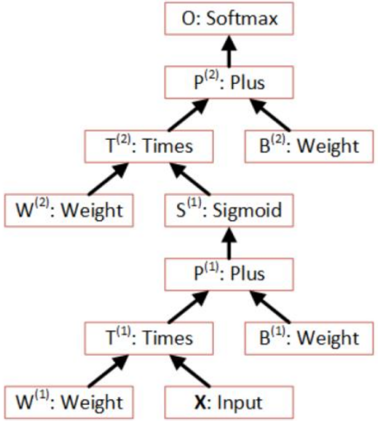
What is interesting about their model is that they used unsupervised learning, so the algorithm was not told in advance what the object categories were or where objects were located in the image. The main result they report is that the network spontaneously learned a reasonably performing face detector from YouTube video frames.

See also Microsoft CNTK, others

- Microsoft Computational Network Toolkit (CNTK)
  - Flexible toolkit based on "tensor-flow" architecture
  - Distributed computations
  - GPU support
- Very fast
- https://github.com/Microsoft/CNTK; https://www.cntk.ai/



From: Micr



Algorithm 2.1 Computational Steps Involved in an One-Hidden-Layer Sigmoid Neural Network

procedure ONEHIDDENLAYERNNCOMPUTATION($\mathbf{X}$) ▷ Each column of $\mathbf{X}$ is an observation vector

$\mathbf{T}^{(1)} \leftarrow \mathbf{W}^{(1)}\mathbf{X}$

$\mathbf{P}^{(1)} \leftarrow \mathbf{T}^{(1)} + \mathbf{B}^{(1)}$ ▷ Each column of $\mathbf{B}^{(1)}$ is the bias $\mathbf{b}^{(1)}$

$\mathbf{S}^{(1)} \leftarrow \sigma\left(\mathbf{P}^{(1)}\right)$ ▷ $\sigma(.)$ is the sigmoid function applied element-wise

$\mathbf{T}^{(2)} \leftarrow \mathbf{W}^{(2)}\mathbf{S}^{(1)}$

$\mathbf{P}^{(2)} \leftarrow \mathbf{T}^{(2)} + \mathbf{B}^{(2)}$ ▷ Each column of $\mathbf{B}^{(2)}$ is the bias $\mathbf{b}^{(2)}$

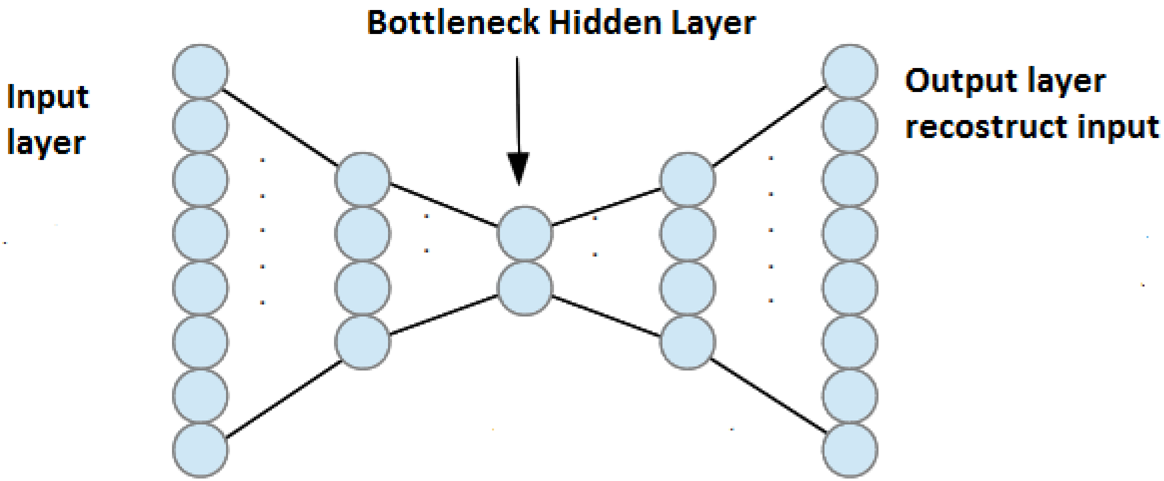
$\mathbf{O} \leftarrow \text{softmax}\left(\mathbf{P}^{(2)}\right)$ ▷ Apply softmax column-wise to get output $\mathbf{O}$

end procedure

Statistica

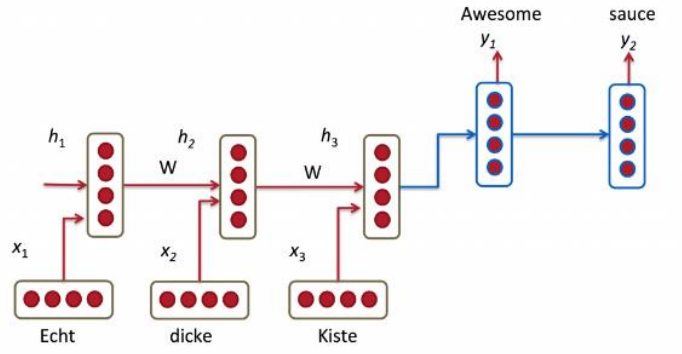
Auto-encoders for Dimensionality Reduction

- Think of it as factor analysis or classic PCA, but non-linear
- Allows you to de-noise signals
- Useful in many manufacturing contexts



Recursive Networks; Networks with Memory e.g., Recognizing Speech

- Delaying activation enables “memory”, context
- Necessary for understanding language



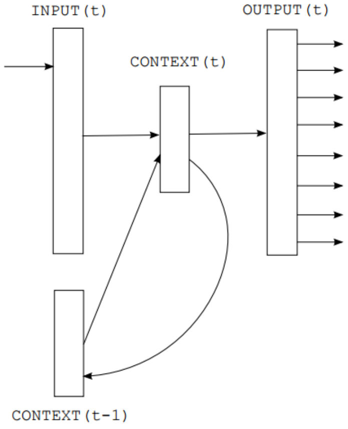
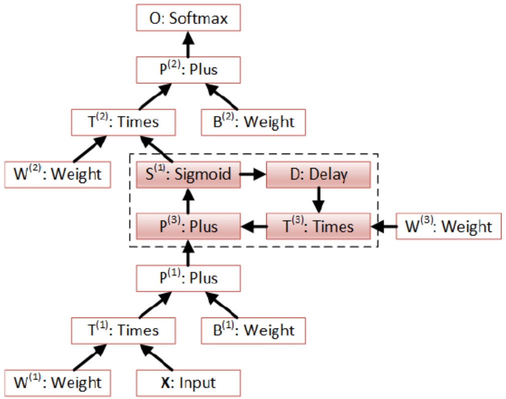


Figure 1: *Simple recurrent neural network.*



From: http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

Extracted June 6, 2016

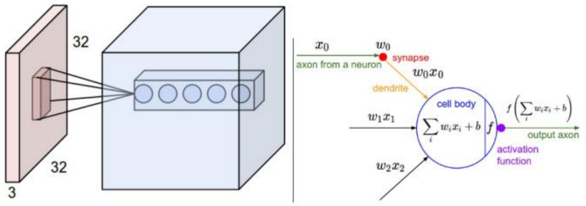
From: http://www.fit.vutbr.cz/research/groups/speech/publi/2010/mikolov_interspeech2010_IS100722.pdf

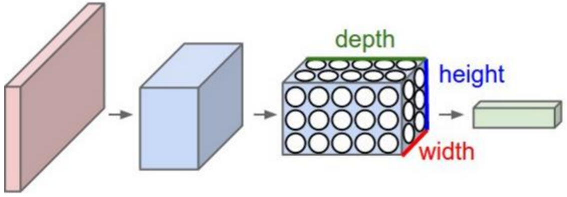
Mikolov, T., Karafiat, M., Burget, L., Cernocky, Khudanpur, S. (2010). Recurrent neural network based language model. Interpseech 2010.

From: An introduction to Computational Networks and the Computational Network Toolkit, 2016

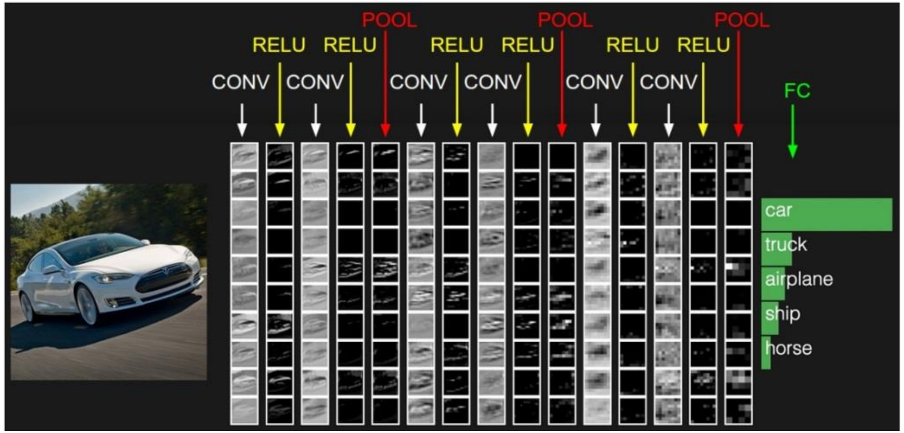
Statistica

Convolutional Networks
Am I about to hit a car or a horse?





- Converting bitmaps (width, height, color/depth)
- Into nodes that capture local activation (convolutional layer)
- Then "thresholding" to augment contrast: ReLU Rectified Linear Units; typically f(x)=max(0,x)
- Then pooling (down-sampling) to represent the "essence" of a pattern
- .....
- Then connecting the reduced representations to a fully connected layer for prediction



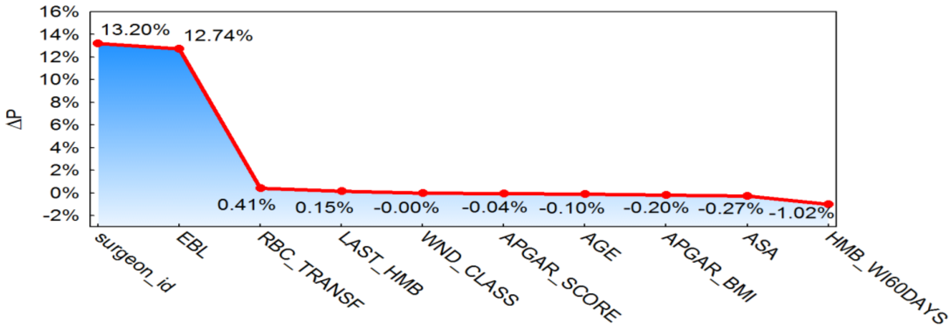
But can machine learning models be transparent?

Reason Scores via What-If or Scenario Analysis

- Given a prediction model of risk, “move” one-predictor-at-a-time by small amount
- Compute numeric derivatives which can be interpreted like parameters of linear models
- Like statistical models, machine learning models
  - Can be interpreted
  - Validated
  - Evaluated for accuracy, model mis-specification, quality of representation, …
- But, “importance” here can have different meaning than simple linear or one-variable-at-a-time importance

Table 1: Characteristics of a Patient

Parameter Name	Value	Parameter Name	Value
APGAR_SCORE	6	APGAR_BMI	35
AGE	83	HMB_WI60DAYS	13.8
RBC_TRANSF	1200	LAST_HMB	9.6
EBL	1500	WND_CLASS	20-Clean Contaminated
ASA	3	surgeon_id	33



Source: Hill, T, Rastunkov, V., Cromwell, J. W. (2013). *Predictive and Prescriptive Analytics for Optimal Decisioning: Hospital Readmission Risk Mitigation.*
Paper presented at the 2013 IEEE International Conference on Healthcare Informatics (ICHI), Philadelphia, PA.

Statistica

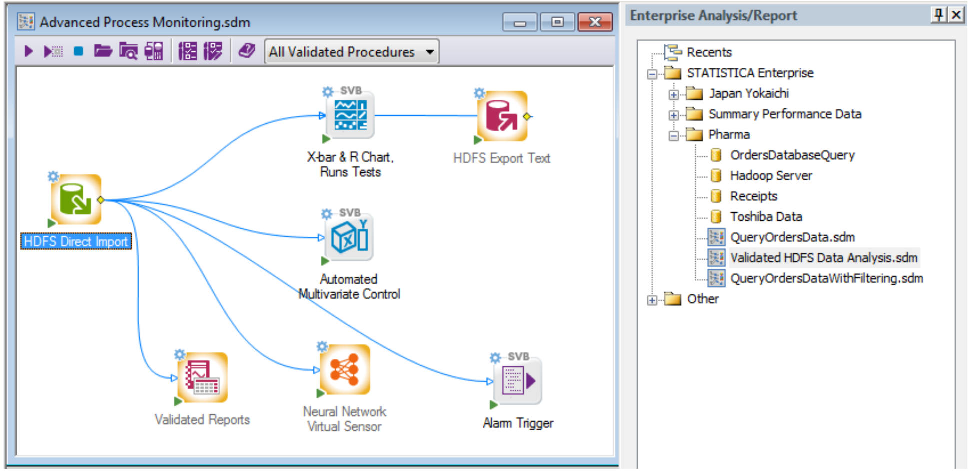
Approach to Validation

- Any analytic reporting in the regulated domain should have at least these attributes
  - Repeatability
  - Ability to Audit
  - Non-repudiation

- Regarding Big Data, Hadoop: "*The application should be validated; IT infrastructure should be qualified*"

  European Commission Health and Consumers Directorate – General, 4, Good Manufacturing Practice Medicinal Products for Human and Veterinary Use, Annex 11: Computerised Systems

- Pattern recognition algorithms are no different from (sometimes complex) statistical procedures
  - There are documented best practices
  - There are clear procedures for evaluating model quality
  - There is a huge amount of literature and documentation for how to apply and verify methods for process monitoring and control
  - E.g.:
    - Hastie, T., Tibshirani, R., & Friedman, J. (2013). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Cor. 3rd Printing). Springer Series in Statistics.
    - Miner, L., Bolding , P., Hilbe, J, Goldstein, M., Hill, T., Nisbett, R., Walton, N., & Miner, G. (2014). *Practical Predictive Analytics and Decisioning Systems for Medicine: Informatics Accuracy and Cost-Effectiveness for Healthcare Administration and Delivery Including Medical Research*. Academic Press.
    - Nisbett, R., Miner, G., & Elder, J. (2009). *Handbook of Statistical Analysis and Data Mining Applications.* Academic Press.
    - ....



Statistica

UNIVERSITY OF IOWA HOSPITALS & CLINICS
University of Iowa Health Care

Big data analytics transforms the operating room

Company: University of Iowa Hospitals and Clinics (UIHC) | Industry: Healthcare | Website: www.uihealthcare.org/

> "Predictive analytics is allowing us to deal with the ever-increasing types of data that healthcare institutions need to deal with."
>
> *Dr. John Cromwell, MD*
> *Director of Gastrointestinal Surgery*

Business challenge

UIHC surgeons needed to know the susceptibility of patients to infections in order to make critical treatment decisions in the operating room. Infection rates have major implications to overall patient health and cost savings.

Solution

The surgical team harnessed the power of big data analytics, coupled with other methods, to keep patients safe — reducing surgical site infections by 58 percent while decreasing the cost of care.

Results

- **Reduced surgical site infection** occurrence by 58 percent
- **Merged historical and live patient data** to predict likelihood of infection
- **Reduced cost** of patient care
- **Personalized care** based on patients' own characteristics
- **Improved efficiency** by enabling staff to run predictive models and access results with a mobile application or web browser

Read the full case study >

Published: March 2015 | Expires: March 2017

DELL Statistica

Applications to Manufacturing

Optimize processes, improve quality, monitor suppliers

Better error tracking, monitor supply chains, improve yields and reduce costs

- Manufacturing Optimization
- Predictive Failure Analysis
- Root Cause Analysis
- Process Optimization
- Statistical Process Control
- R&D
- Predictive Maintenance
- Design of Experiment
- Product Traceability
- Production Process

Applications in Manufacturing

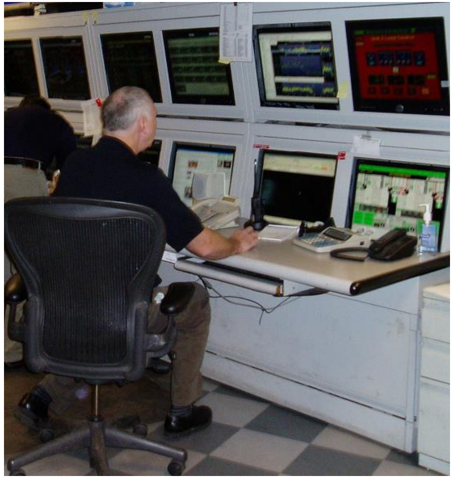
- Process monitoring:
  - Predictive failure analysis and preventive maintenance
  - "Virtual sensors" and non-linear model-based process monitoring (e.g., via Neural Nets, auto-encoders)
  - Anomaly detection in very high-dimensional data streams

- Root cause analysis and interaction detection
  - Automated root cause analysis and interaction detection in high-dimensional inputs

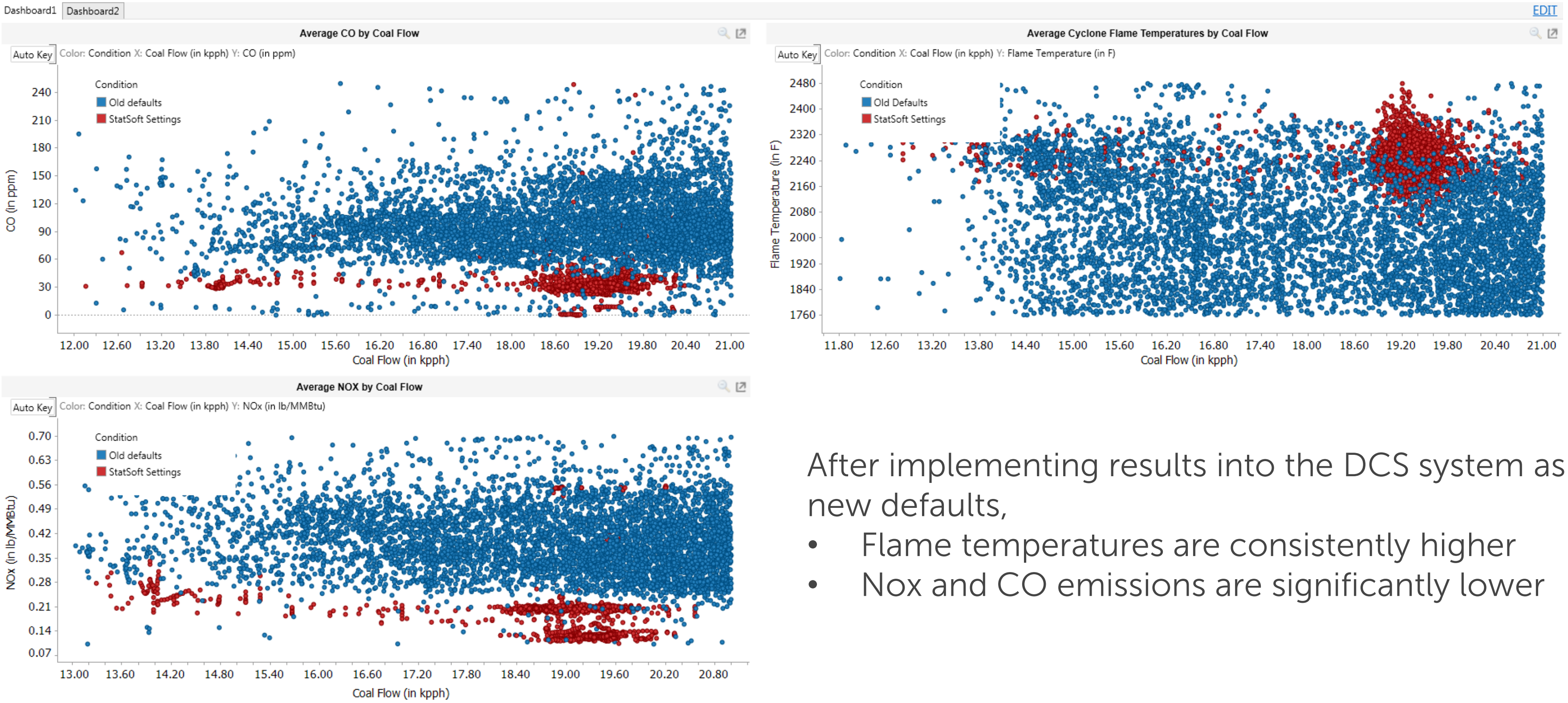
- Improving manufacturing processes through optimizing high-fidelity process models
  - Inverse prediction for robust processes, robust optimization
  - Warranty claim analytics, relating product performance in the field to manufacturing process parameters
  - What-if scenario analysis and simulation

Combustion Optimization from Historical Data

- Robust optimization of continuous high-dimensional process streams for competing goal functions
- Optimized combustion for reduced NOx, CO, and stable/sustainable temperatures



Robust Optimization of Continuous Processes

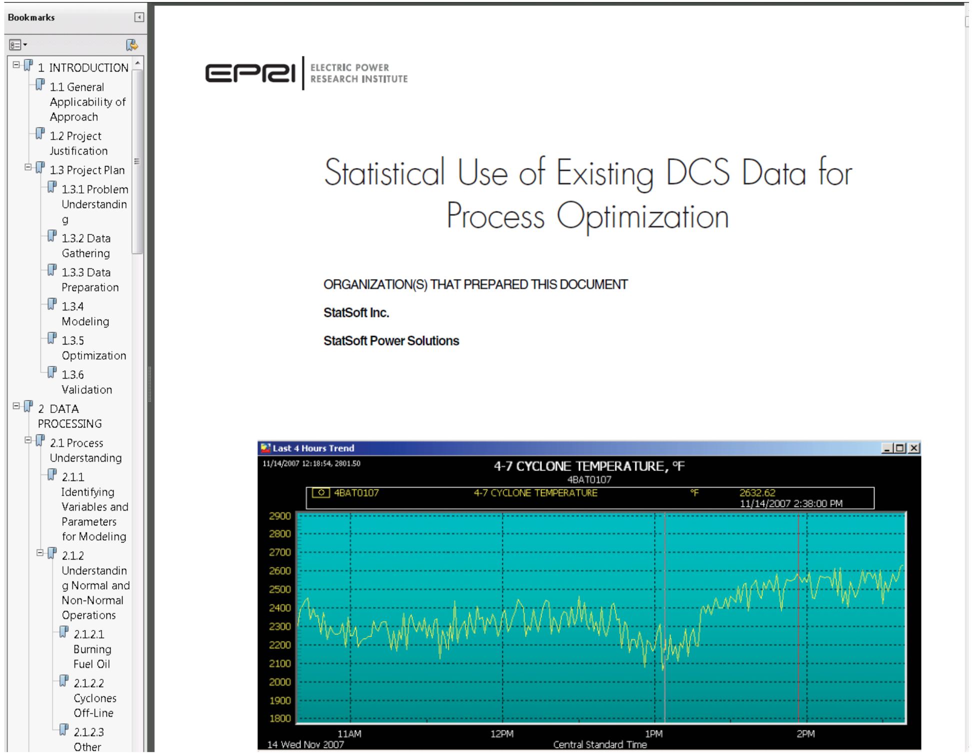


After implementing results into the DCS system as new defaults,

- Flame temperatures are consistently higher
- Nox and CO emissions are significantly lower

Optimization for Robust Continuous Performance and Competing Goal Functions

- Models were highly effective at
  - Lowering emissions
  - Stabilizing flame temperatures
  - Maintaining optimal conditions for robust uninterrupted operations (less downtime)



See : *EPRI/StatSoft Project 44771: Statistical Use of Existing DCS Data for Process Optimization*.

Statistica

Solar tech producer drives quality with predictive analytics

When your reputation relies on maintaining the highest standards of quality, performance and durability, Statistica shines.

Business challenge

Over 10,000 streaming, automated parameters required real-time monitoring and analysis to meet ever-higher demands of product quality—and to anticipate manufacturing issues—in this extremely competitive industry.

Solution

Statistica Enterprise integrated easily with the company's existing MES system and offered practical algorithmic capabilities in a scalable, web-enabled platform that maintains performance in the face of increasing complexity.

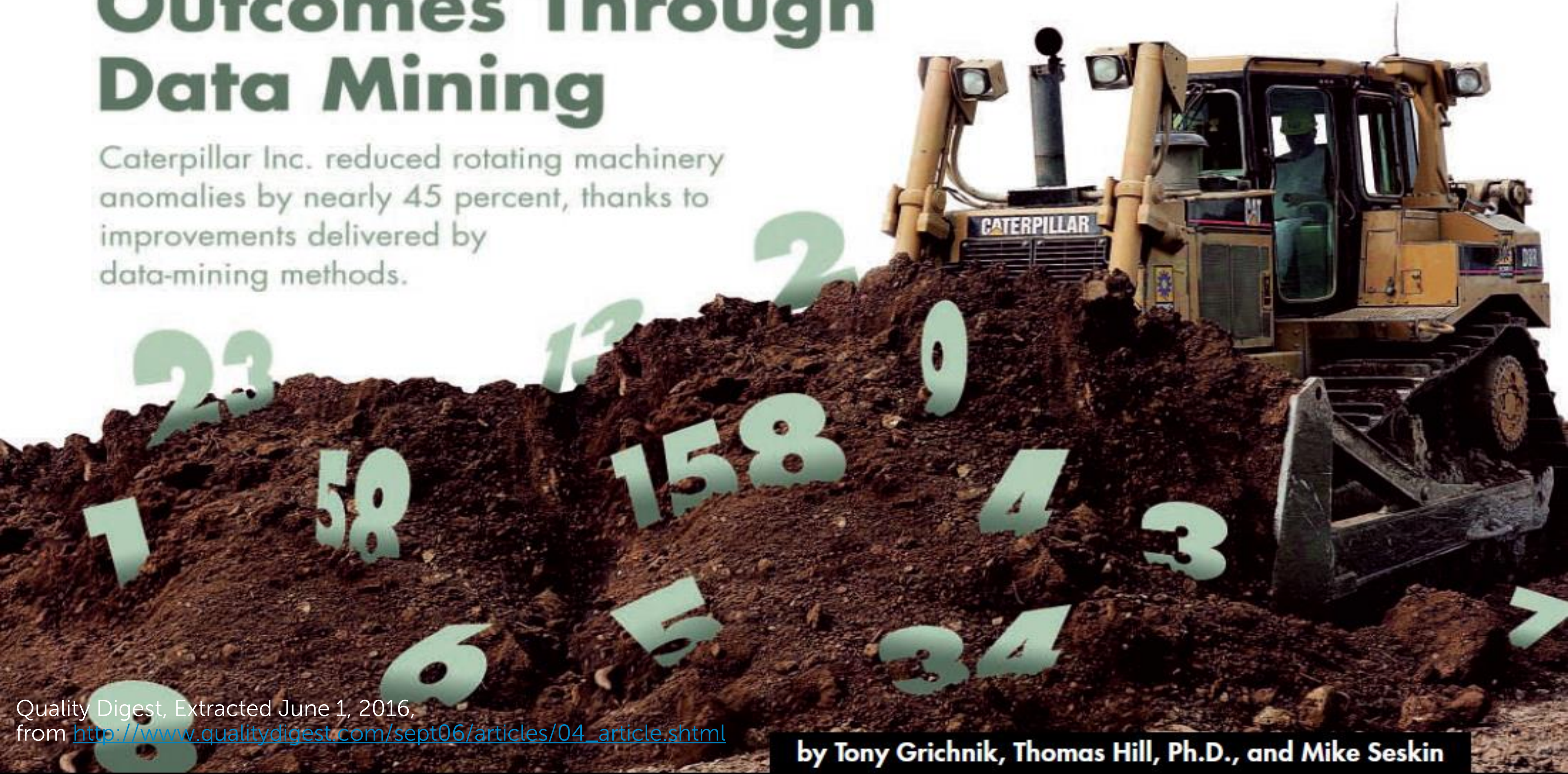
Results

- **Optimizes manufacturing efficiency** by enabling hundreds of end-users and engineers to monitor and respond to mission-critical data
- **Maintains company's competitive edge** through application of predictive process monitoring for potential quality issues
- **Supports real-time processes 24/7**

Published: June 2016 | Expires: June 2018

Predicting Quality Outcomes Through Data Mining

Caterpillar Inc. reduced rotating machinery anomalies by nearly 45 percent, thanks to improvements delivered by data-mining methods.

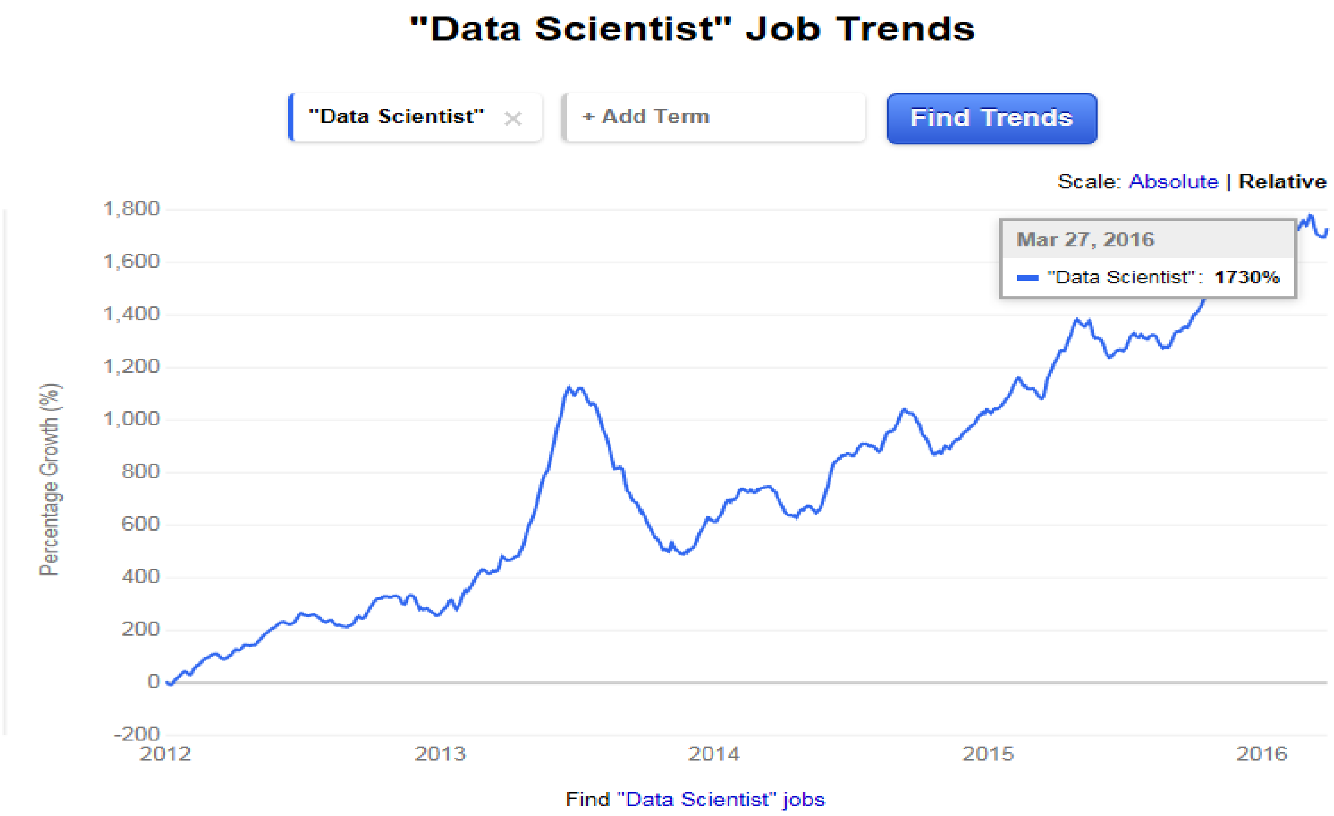


Quality Digest, Extracted June 1, 2016,
from http://www.qualitydigest.com/sept06/articles/04_article.shtml

by Tony Grichnik, Thomas Hill, Ph.D., and Mike Seskin

Barriers

“Data Scientists” are Difficult to Find

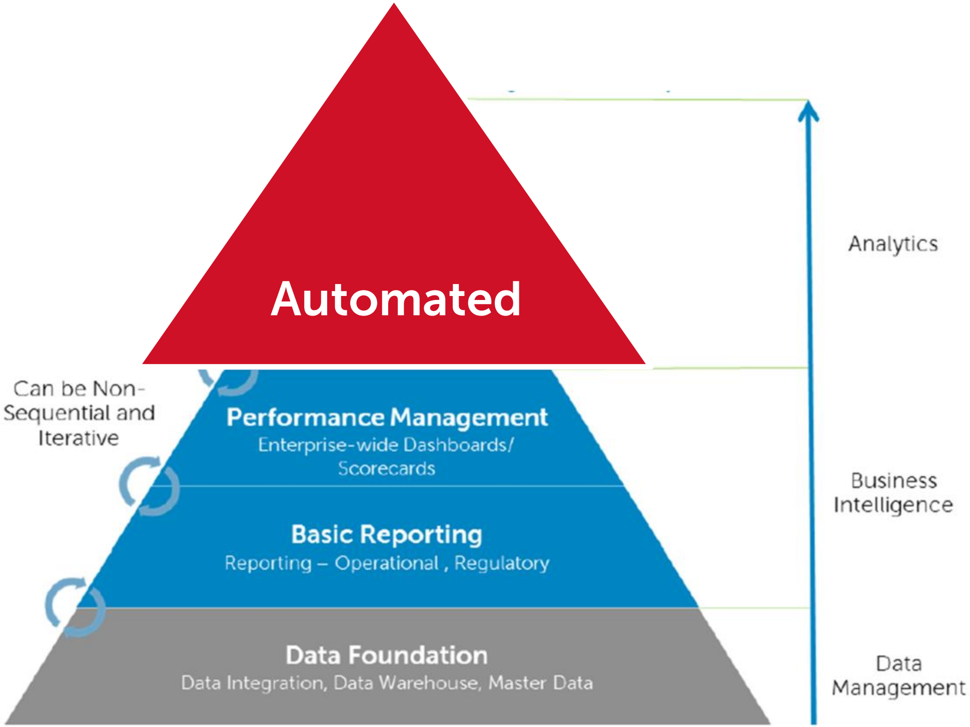


Source: www.Indeed.com; extracted May 23, 2016

DELL Statistica

Automated Analytics

- Analytic Maturity model ca. 2016 - ?



Automated Modeling
Automated Model Calibration
Automated Actions

CIO Journal.

CIO Report | Consumerization | Big Data | Cloud | Talent & Management | Security

January 14, 2015, 4:42 PM ET

The Rise of Automated Analytics

Article | Comments (3)

Email | Print

By THOMAS H. DAVENPORT

Guest Contributor

From Data Scientists to Citizen Data Scientists

- Gartner introduced the term “Citizen Data Scientist” to denote the business user empowered through simplified and automated data science tools, leveraging advanced analytics efficiently and effectively to augment their everyday work
- The current approach to analytics for validated manufacturing support will work:
  - Analytic workflows (recipes) are created by data scientists, statisticians
    - And validated
  - Workflows are made available to business end-users in role-based system
  - Empowering these Citizen Data Scientists through validated best analytic practices
  - (E.g., Dell Statistica Enterprise System)

Advanced Analytics Throughout the Enterprise: Personas

Citizen Data Scientists / Line of Business / Domain Expert

Business users who have little to no knowledge of analytics but understand the value of data. They need more than Excel and traditional BI tools but are challenged by data prep and visualization.

Line of Business Specialists, Consumer of Data Science

They use analytics for quality control, process monitoring and optimization. They generally have an scientific or engineering background but are not programmers.

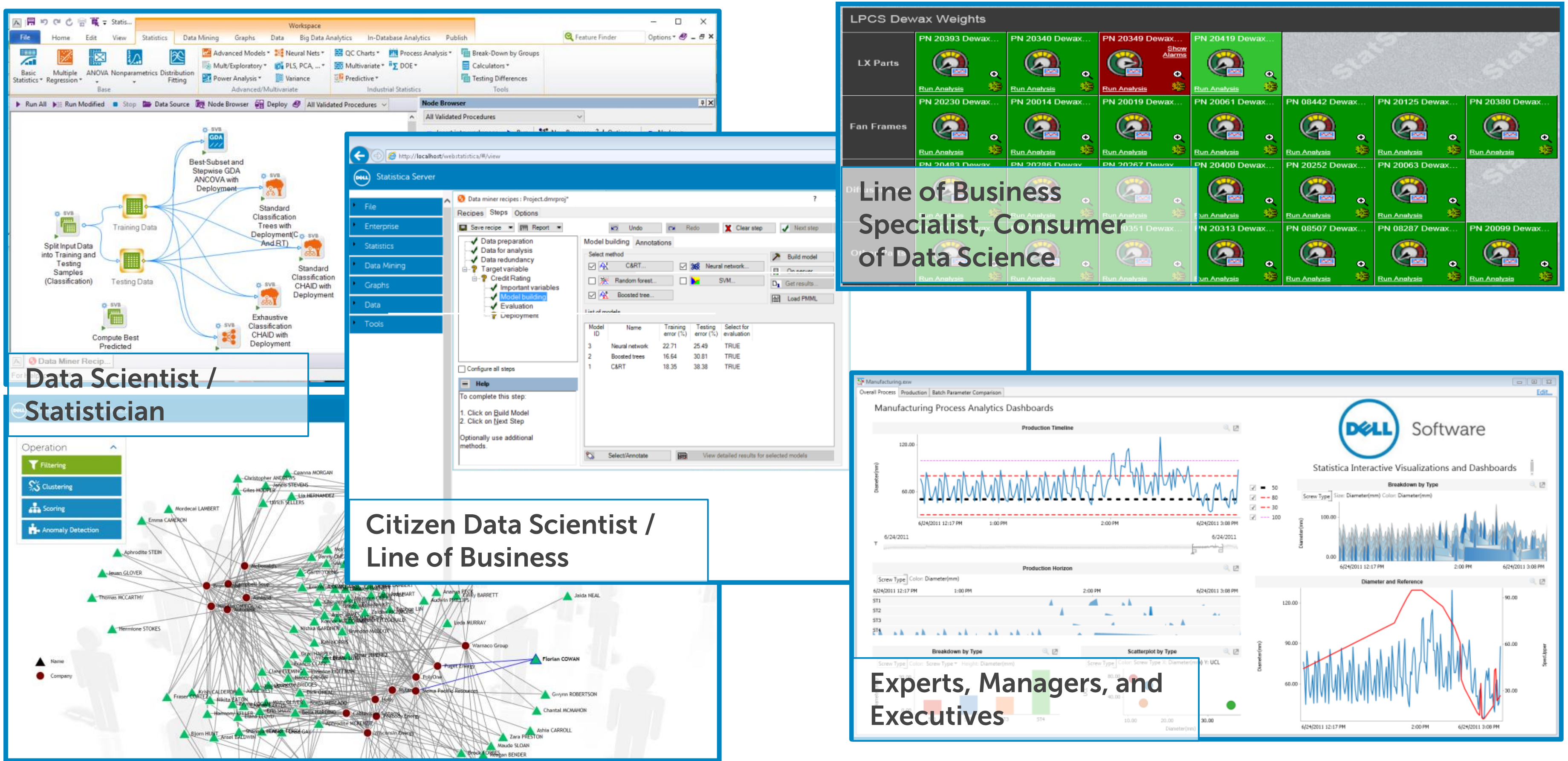
Data Scientists / Statisticians

If they are under 30 they call themselves data scientists. Over 30, they call themselves statisticians. Deep understanding of mathematics and statistics. May or may not be programmers.

Experts, Managers, and Executives

These individuals span the spectrum and industries but are generally in charge of applying analytics to run their businesses and make strategic decisions.

Empower More People – With the Right Tools



Shire

Specialty biopharmaceutical company reaps data analysis and process efficiencies

Company: Shire | Industry: Pharmaceutical | Country: United States | Website: www.shire.com

"Dell's validated data system for Internal Manufacturing represents a significant step toward better understanding of our process technology and higher business efficiency."

Gloria Gadea-Lopez
Director of Manufacturing Systems

Business challenge

Shire's Rare Diseases Business Unit needed a validated data capture and analysis tool to conduct statistical process control, monitor processes and identify areas for improvement.

Solution

Replacing a mix of in-house scripts and tools, Dell Statistica gave Shire Internal Manufacturing controlled access to validated, real-time process data in support of manufacturing operations management across the enterprise.

Results

- **Seamless, unified** data preparation
- **Flexible, scalable** data entry forms
- **Validated interface** with Laboratory Information Management System (LIMS)
- **Automated distribution** of predefined control charts
- **Efficient cross-site** process and product comparison

Read the full case study >

Published: March 2015 | Expires: March 2017

DELL Statistica

Some Cautions

- When predictions affect people's outcomes, the analytic process must support regulatory oversight

The Privacy Advisor

Original reporting and feature articles on the latest privacy developments

How Big Data Discriminates

…ress at this time last year with
…is item: "Big Data Doesn't

≡ LAW360 News, cases, companies, firms Advanced Search

Big Data's Potential Disparate Impact Problem

Law360, New York (August 21, 2014, 11:25 AM ET) -- If you have read any article the past year on a legal technology issue, you have undoubtedly heard about big data. Some very prominent technology executives believe that big data will have any even bigger

NYU | LAW

JD Admissions | Faculty & Scholarship | Global Opportunities
LLM/JSD Admissions | Academics & Courses | Law & Business
Executive Education | Current Students | Public Service
About NYU Law | Alumni & Giving | Centers & Institutes

Home › Centers & Institutes › Information Law Institute › Algorithms and Accountability Conference

INFORMATION LAW INSTITUTE

- Events
- People
- Research
- Privacy Research Group
- Security Seminar
- ILI Student Blog
- Contact Us

INFORMATION LAW INSTITUTE

Algorithms and Accountability Conference

Intro

Scholars, stakeholders, and policymakers question the adequacy of existing mechanisms governing algorithmic decision-making and grapple with new challenges presented by the rise of algorithmic power in terms of transparency, fairness, and equal treatment. Algorithms increasingly shape our news, economic options, and educational trajectories. The centrality and concerns about algorithmic decision making have only increased since we hosted the Governing Algorithms conference in May 2013. This event built upon that conversation to address legal, policy and ethical challenges related to algorithmic power in three specific contexts: media production and consumption, commerce, and education.

Date, Location, Organization

The Algorithms and Accountability conference took place on Saturday, February 28th, 2015, at New York University, Lipton Hall, NYC. Organized by the Information Law Institute, NYU School of Law, it was cosponsored by NYU Steinhardt Department of Media, Culture and Communications, the Intel Science & Technology Center for Social Computing and Microsoft. The organizing committee consists of Joris van Hoboken, Helen Nissenbaum and Elana Zeide.

Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights

Executive Office of the President

May 2016

THE WHITE HOUSE

Key Points

1

Technology is driving competitiveness. Machine learning can improve quality, processes, and deliver competitive advantage.

2

Machine learning algorithms can be "controlled." Best practices exist how to evaluate, validate, and interpret machine learning results.

3

Barriers can be overcome. Big-data-science skillsets are scarce; but configured role-based best-analytic practices will enable engineers, biologists, and Citizen Data Scientists to be successful.

The power to do more

Are you ready to predict the future?

/ Statistica /

Are you

/ future ready?/