The High ROI of Machine Learning for Innovative Organizations

John F. Elder, Ph.D.

elder@elderresearch.com



Charlottesville, VA Washington, DC Baltimore, MD Raleigh, NC 434-973-7673 www.elderresearch.com

Before we get started...

Questions

- Please use the questions tab to enter questions throughout the webinar
- Questions will be answered during the Q&A session at the end

Slides

 A pdf of the slides will be posted on the webinar website next week or email <u>paul.derstine@elderresearch.com</u> to request a copy

Problem during the webinar?

- Simply refresh your browser screen

This webinar could be on...

- ROI = increased gain at what cost? = Δ Gain / Δ Cost
- Measuring current performance to establish the baseline
- Common myths about "current performance"
- Ways to increase performance
 - Improve profit, efficiency, accuracy, satisfaction
 - Reduce cost, delays
 - Find new good things
 - Eliminate bad things
- But many projects (and all in the webinar abstract) are "All or Nothing"; their gains are extreme, or zero depending only on *if the model is used*

Outline

- Data Science / Machine Learning / Artificial Intelligence hype
- Inductive modeling methods; which is best?
- 5 Key ERI Projects, and the technologies they triggered
 - Investment timing (target shuffling)
 - Credit scoring (ensembles)
 - Drug discovery (kernel density)
 - Health assessment (text mining, non-zero initialization)
 - Oil & Gas exploration (missing data imputation)
- Fruitful interaction with carbon-based life forms can be tough, but is essential to getting our work adopted





"Of course machines can think. After all, humans are just machines made of meat."

- MIT Comp Sci professor



"Of course machines can think. After all, humans are just machines made of meat."

- MIT Comp Sci professor

My view: Human and computer strengths are more complementary than alike



Al is getting a lot of attention lately...

Go: Super-human performance (2016)



Texas Hold-Em **Poker**: High-human (2015)

Al is getting a lot of attention lately... **Go**: Super-human performance (2016)





Texas Hold-Em **Poker**: High-human (2015)



Al is getting a lot of attention lately... **Go**: Super-human performance (2016)



Self-Driving **Cars**: Sub-human (2015) Recent pedestrian death. But 9/2014- 11/2015 in California, 272 failures and 13 crashes prevented by human attendants

Deep Learning = many-layered neural network has actually provided a lot of the breakthroughs attributed to AI



Image from:

https://cdn.edureka.co/blog/wp-content/uploads/2017/05/Deep-Neural-Network-What-is-Deep-Learning-Edureka.png



Decision Tree



Decision Tree

Nearest Neighbor



Decision Tree

Nearest Neighbor





Nearest Neighbor 0 : n

Neural Network (or Polynomial Network)



Relative Performance Example:

5 Algorithms on 6 Datasets (with Stephen Lee, U. Idaho, 1997)



Averaging Competing Models ~Does as well as Best



Voting Models Improved Also



All Ensemble Methods Improved Performance





Ensemble Methods in Data Mining

Improving Accuracy Through Combining Predictions

Giovanni Seni John Elder

Synthesis Lectures on Data Mining and Knowledge Discovery

Robert Grossman, Series Editor







ROBERT NISBET JOHN ELDER GARY MINER

HANDBOOK OF

Statistical Analysis



Target Shuffling

is an resampling solution to *p*-hacking, or the problem of *vast search*.

I first described it in this Handbook.

See also our 2-minute animated video at *www.ElderResearch.com*, and the explanatory white paper ("Are Orange Cars really not Lemons?") I wrote with Ben Bullard.

2) Credit Scoring Model Performance



Median (and Mean) Error Reduced with each new model Combined



No. Models in combination

...and in a multitude of counselors there is safety. Proverbs 24:6b

3) Drug Discovery



Placebo is a worthy foe (baseline)

3) Drug Discovery



Placebo is a worthy foe (baseline)

Stronger when it has side effects

Kernel Density plots (tectonic activity below San Francisco)



Grillenzoni, 2013: https://doi.org/10.1016/j.spasta.2013.11.003

Drug Effectiveness



Density surfaces enclose ascending quartiles of data (75% of data within red; 50% w/in green; 25% w/in blue)

Drug Effectiveness



Density surfaces enclose ascending quartiles of data (75% of data within red; 50% w/in green; 25% w/in blue)

• Pain: Approval process is long, bureaucratic



• Pain: Approval process is long, bureaucratic



1/2 of appeals overturn original decision

• Pain: Approval process is long, bureaucratic



- Goal: Fast-track "easy" cases
- Challenge: Free-text on disability application

• Pain: Approval process is long, bureaucratic



With Text Mining, 1/5 of cases approved immediately!

- Goal: Fast-track "easy" cases
- Challenge: Free-text on disability application
- Result: 20% of Approvals possible immediately and with greater consistency

5) Oil and Gas Exploration

Gas Production Saved



Workload (Wells Treated)

How we see ourselves

How our clients see us

ROI on the Five Examples

- On 4 of 5, the ROI was, or continues to be, incalculably large
- On 1, it is zero or negative (!) so far
- Surveys show that only ~33% of ML projects get implemented
- Implementation completion in our first decade was 65% (after project technical success of ~90%)
- The biggest reason for failure was social/political, not technical
- Implementation success in our 2nd decade increased to 93% (after project technical success approaching 100%)

Analytics is a Success only if it's Used

- Key Hard Issues:
 - Data Science technology
 - clear thinking
- Key Soft Issues:
 - trust
 - willingness and ability to change now



THE NEW YORK TIMES BESTSELLER



DANIEL

KAHNEMAN

WINNER OF THE NOBEL PRIZE IN ECONOMICS

"[A] masterpiece... This is one of the greatest and most engaging collections of insights into the human mind I have read." —WILLIAM BASTERLY, Financial Times Recent study in National Academy of Sciences: 8 Israeli parole judges Case review time averaged 6 minutes 35% of applicants approved overall Recent study in National Academy of Sciences: 8 Israeli parole judges Case review time averaged 6 minutes 35% of applicants approved overall























5 months

8 months





Effect size = (8-5)/(9-3) = 50%





5 months

8 months











Halo Effect 4% helped





4% helped 88% helped





4% helped 88% helped



Summary

- Data Science is powerful and proven (no hype needed)
- Companies in vastly different fields are achieving huge gains by extracting useful information from their data
- Human creativity and intuition flawed as it is is vital
- Our work is a succees only if it's used
- The client must trust *you* to trust your technology



John F. Elder IV Founder & Chair, Elder Research, Inc.

DR. JOHN ELDER HEADS THE USA'S MOST EXPERIENCED DATA MINING CONSULTING TEAM. FOUNDED IN 1995, ELDER RESEARCH, INC. HAS OFFICES IN CHARLOTTESVILLE, VA, WASHINGTON DC, BALTIMORE MD, AND RALEIGH NC (*WWW.DATAMININGLAB.COM*). ERI FOCUSES ON FEDERAL, COMMERCIAL, AND INVESTMENT APPLICATIONS OF ADVANCED ANALYTICS, INCLUDING TEXT MINING, CREDIT SCORING, PROCESS OPTIMIZATION, CROSS-SELLING, DRUG EFFICACY, INVESTMENT TIMING, AND FRAUD DETECTION.

JOHN EARNED ELECTRICAL ENGINEERING DEGREES FROM RICE UNIVERSITY, AND A PHD IN SYSTEMS ENGINEERING FROM THE UNIVERSITY OF VIRGINIA, WHERE HE'S BEEN AN ADJUNCT PROFESSOR TEACHING OPTIMIZATION OR DATA MINING. PRIOR TO 23 YEARS AT ERI, HE SPENT 5 YEARS IN AEROSPACE DEFENSE CONSULTING, 4 HEADING RESEARCH AT AN INVESTMENT MANAGEMENT FIRM, AND 2 IN RICE'S *COMPUTATIONAL & APPLIED MATHEMATICS* DEPARTMENT.

DR. ELDER HAS AUTHORED INNOVATIVE DATA MINING TOOLS, IS A FREQUENT KEYNOTE SPEAKER, AND HAS CHAIRED INTERNATIONAL ANALYTICS CONFERENCES. JOHN WAS HONORED TO SERVE FOR 5 YEARS ON A PANEL APPOINTED BY PRESIDENT BUSH TO GUIDE TECHNOLOGY FOR NATIONAL SECURITY. HIS BOOK WITH BOB NISBET AND GARY MINER, *HANDBOOK OF STATISTICAL ANALYSIS & DATA MINING APPLICATIONS*, WON THE PROSE AWARD FOR TOP BOOK IN MATHEMATICS FOR 2009. HIS BOOK WITH GIOVANNI SENI, *ENSEMBLE METHODS IN DATA MINING*, WAS PUBLISHED IN 2010, AND HIS BOOK WITH COLLEAGUE ANDREW FAST AND 4 OTHERS ON *PRACTICAL TEXT MINING* WON THE 2012 PROSE AWARD FOR COMPUTER SCIENCE.

JOHN IS GRATEFUL TO BE A FOLLOWER OF CHRIST AND FATHER OF 5.