KaDSci
KNOWLEDGE & DECISION SCIENCES

“Predictive Analytics”:

Understanding and Addressing The Power and Limits of Machines, and What We Should do about it

Daniel T. Maxwell, Ph.D.
President, KaDSci LLC

Copyright KaDSci LLC 2018 – All Rights Reserved

The Functions of Models in Analysis

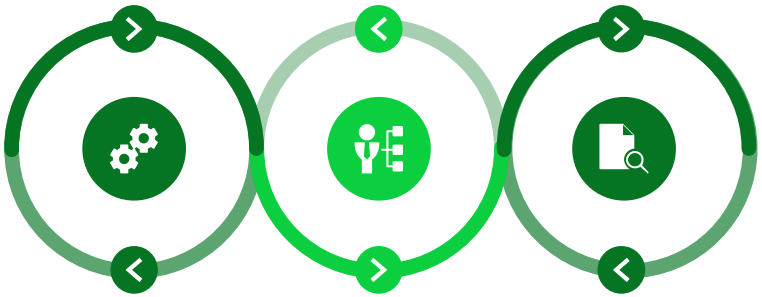
Models are used to:

01. Explain, account for, or describe a phenomenon (Diagnostic)

02. Predict, forecast, or estimate. (Predictive)

03. Recommend a course of action (Prescriptive)

Analysis uses models to



Synthesis

Tell the Story
"Narrative"

Structured Process

Analysis

What are the Pieces?
How do they work?

1 & 2. above are data focused – "Issues of fact"

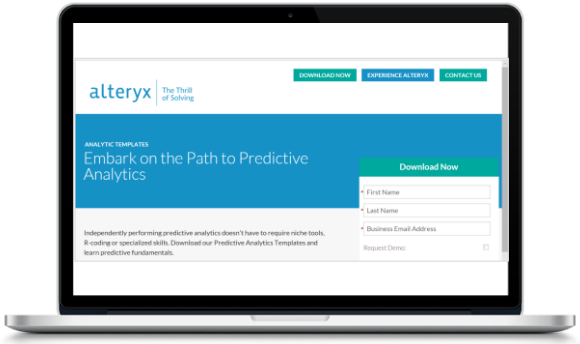
3 above is also predictive & includes decision maker (or decision making system)

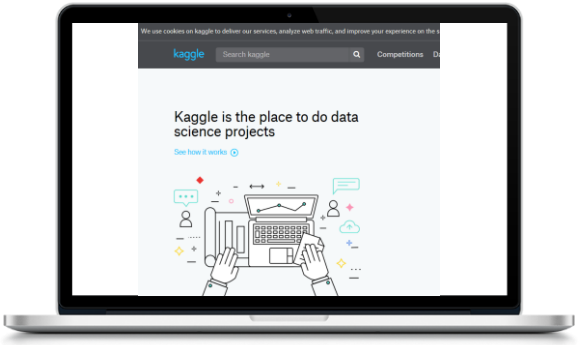
Goals and preferences
- Objectives
- Risk tolerance

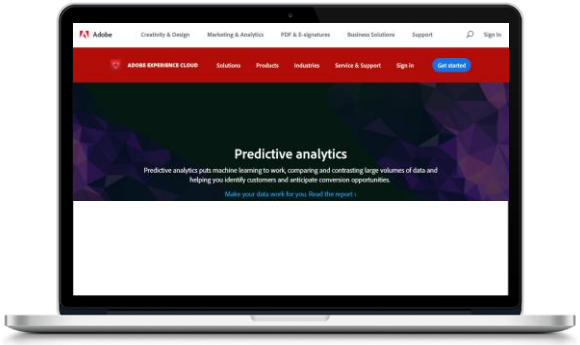
BLUF– You can't forget the thinking part!!!

Copyright KaDSci LLC 2018 – All Rights Reserved

Predictive Analytics & AI are here & they are all Models







Copyright KaDSci LLC 2018 – All Rights Reserved

ALGORITHMS!!

Regression
With all sorts of new and fancy names

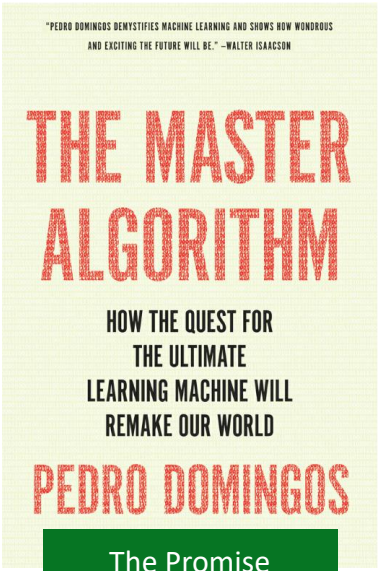
Bayesian Learning

Neural Nets

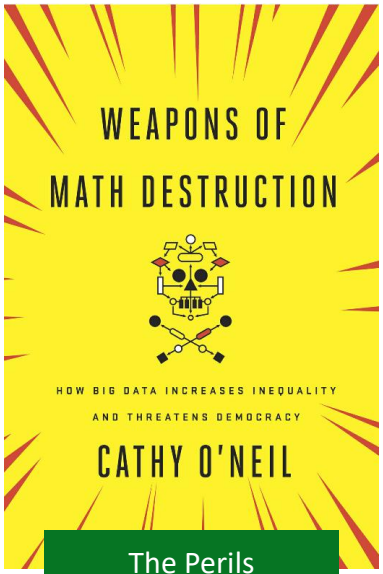
Case Based Reasoning

Influence Diagrams

……..



The Promise



The Perils

All algorithms (and by extension AI tools) rely on data.

Copyright KaDSci LLC 2018 – All Rights Reserved

Not All Data are Created Equal

Noisy → Accurate

Uncertain → Certain

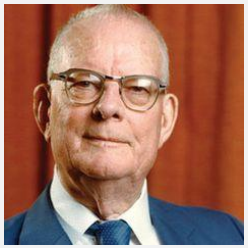
Sparse → Dense

Small → Big

Perishable → Persistent

Unstructured → Structured

Coarse → Precise

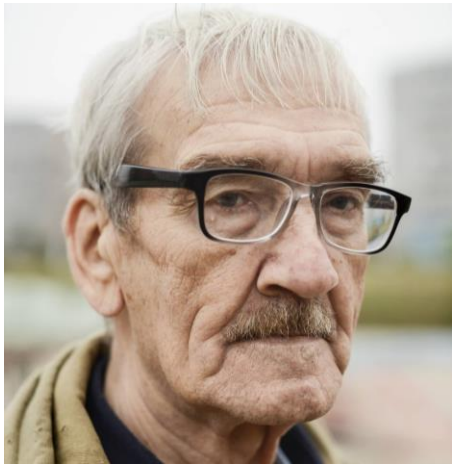


"Without data you're just another person with an opinion"
W. Edwards Deming

Data Science and Data Engineering are different things – The latter is necessary but not sufficient for providing effective analytics.

Copyright KaDSci LLC 2018 – All Rights Reserved

Stanislav Petrov – The Man Who Prevented Nuclear War



Time – September 26, 1983, Three weeks after KAL 007 was shot down.

Lt. Col. Stanislav Petrov (Russian Air Force) observed sensor alerts indicating the US launched an ICBM, followed by five more.

Russian standing orders called for an immediate counter strike against the US and NATO allies.

He disobeyed those orders and declared the alert a false alarm

He was neither praised or punished.

The cause – A rare alignment of clouds, sunlight, and Russian Satellites that watched North Dakota

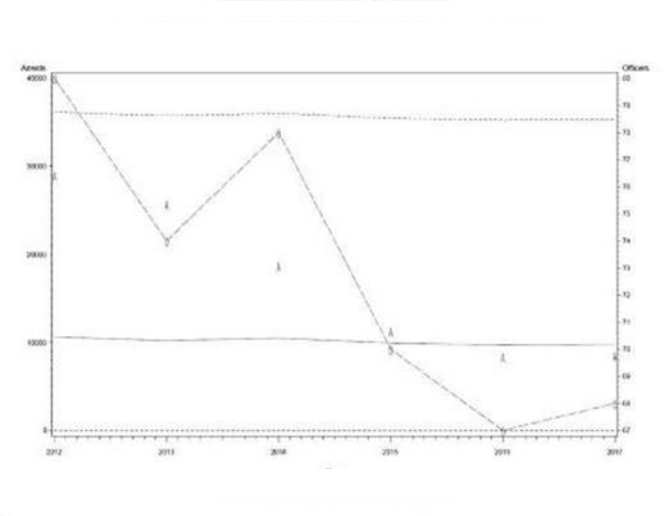
Why did he not report it:

- Five missiles were inconsistent with his understanding of how the US would attack
- Alert system was new
- Alert passed through 30 layers of verification too quickly
- Lack of corroborative evidence
- His civilian experience helped him to make that judgment

He believed if one of his pure military colleagues had been on duty, the outcome could have been very different.

Copyright KaDSci LLC 2018 – All Rights Reserved

An Attempt to Optimize Law Enforcement Officer Hiring

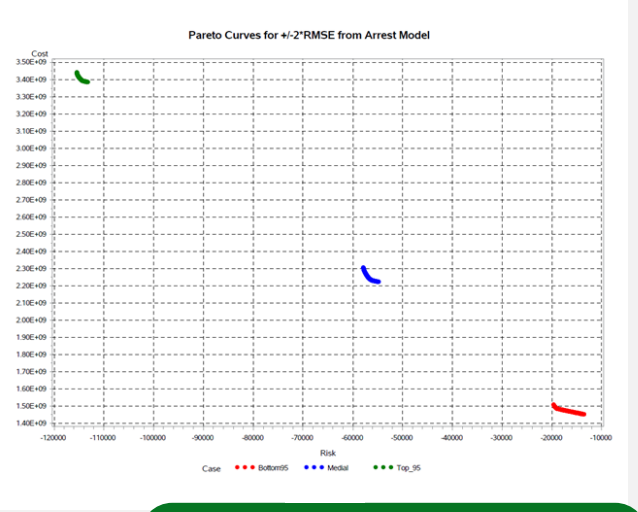


Hiring Strategy

Mean Squared Error is 10X bigger than the coefficient

Officer Coefficient ~ 100
Mean Squared Error ~ 1000
Y Intercept ~ -100

Could be inexpensive and effective

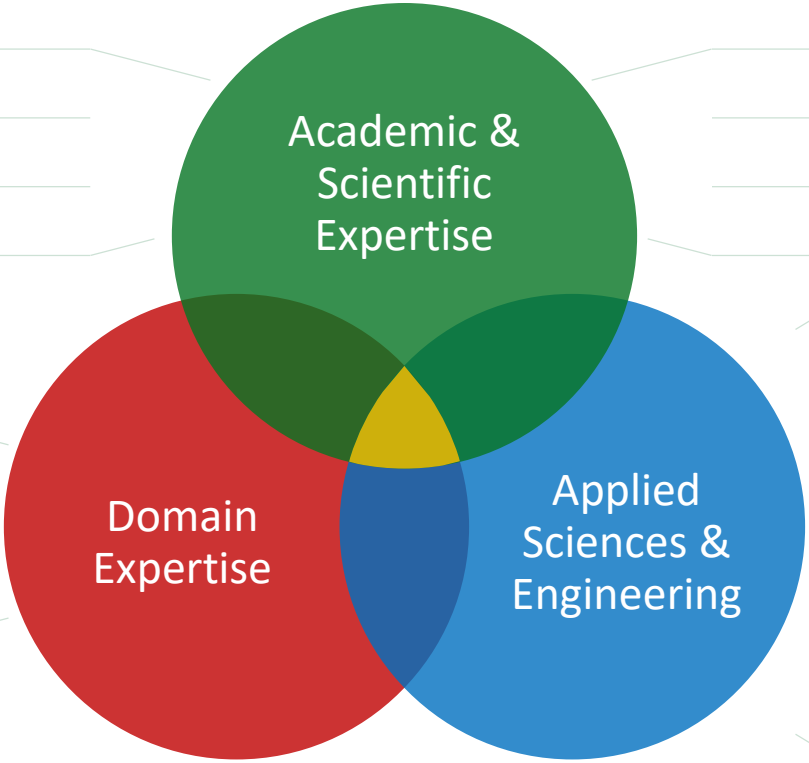


Could be expensive and ineffective

Real Situation | Notional Numbers | Anonymized Organization

Copyright KaDSci LLC 2018 – All Rights Reserved

Analytics (Operations Research)



- Mathematicians
- Psychologists
- Sociologists
- Theoreticians
- Political Scientists
- Historians
- Gov't Executives
- Gov't Managers
- Gov't Service Providers Military
- Technologists
- Software Engineers
- Hardware Engineers
- Systems Engineers
- Algorithm Developers
- Data Modelers
- Operations Researchers

Good Analytics

Provides the foundation for useful models and tools

Is multi-disciplinary, drawing on expertise from across the spectrum of science and engineering

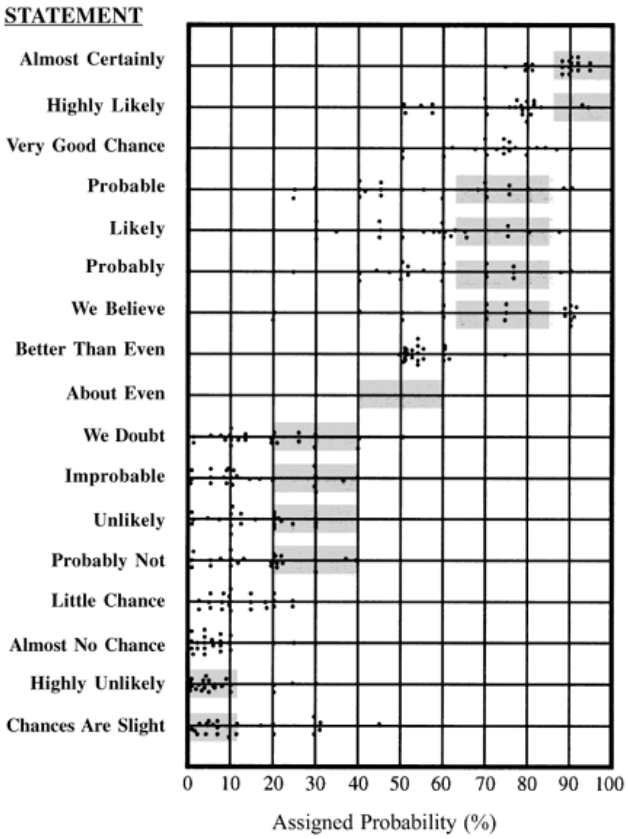
Drawing on Psychology --- Human Perceptions of Uncertainty

Sherman Kent,
CIA

Note
the large number of dots outside the selected regions

100% Certainty		
The General Area of Possibility		
93%	give or take about 6%	Almost certain
75%	give or take about 12%	Probable
50%	give or take about 10%	Chances about even
30%	give or take about 10%	Probably not
7%	give or take about 5%	Almost certainly not
0% Impossibility		

Figure 18: Measuring Perceptions of Uncertainty



Representing uncertainty is messy.
Ambiguity exists even in epistemic uncertainty

Assertion
Data is nothing more (or less) than evidence

Facts:

There is science behind evidence *(Schum, 1994)*

It has been applied (loosely) by the IC.

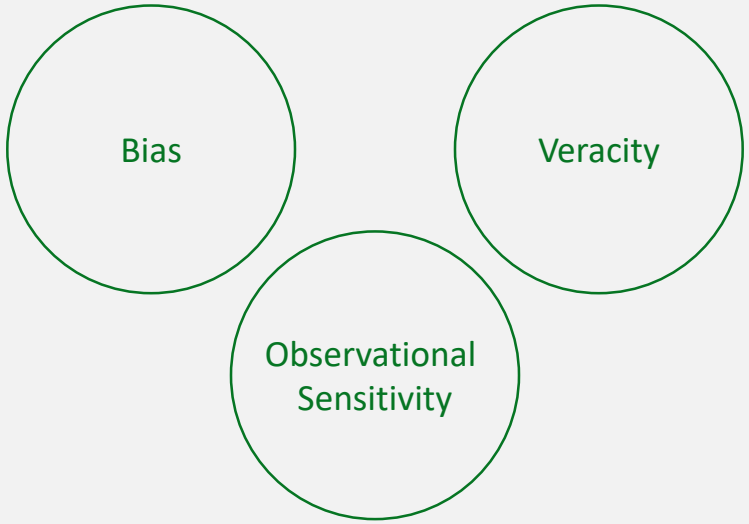
Assertion:

Failure to account for these evidential factors undermines the quality of machine reasoning.

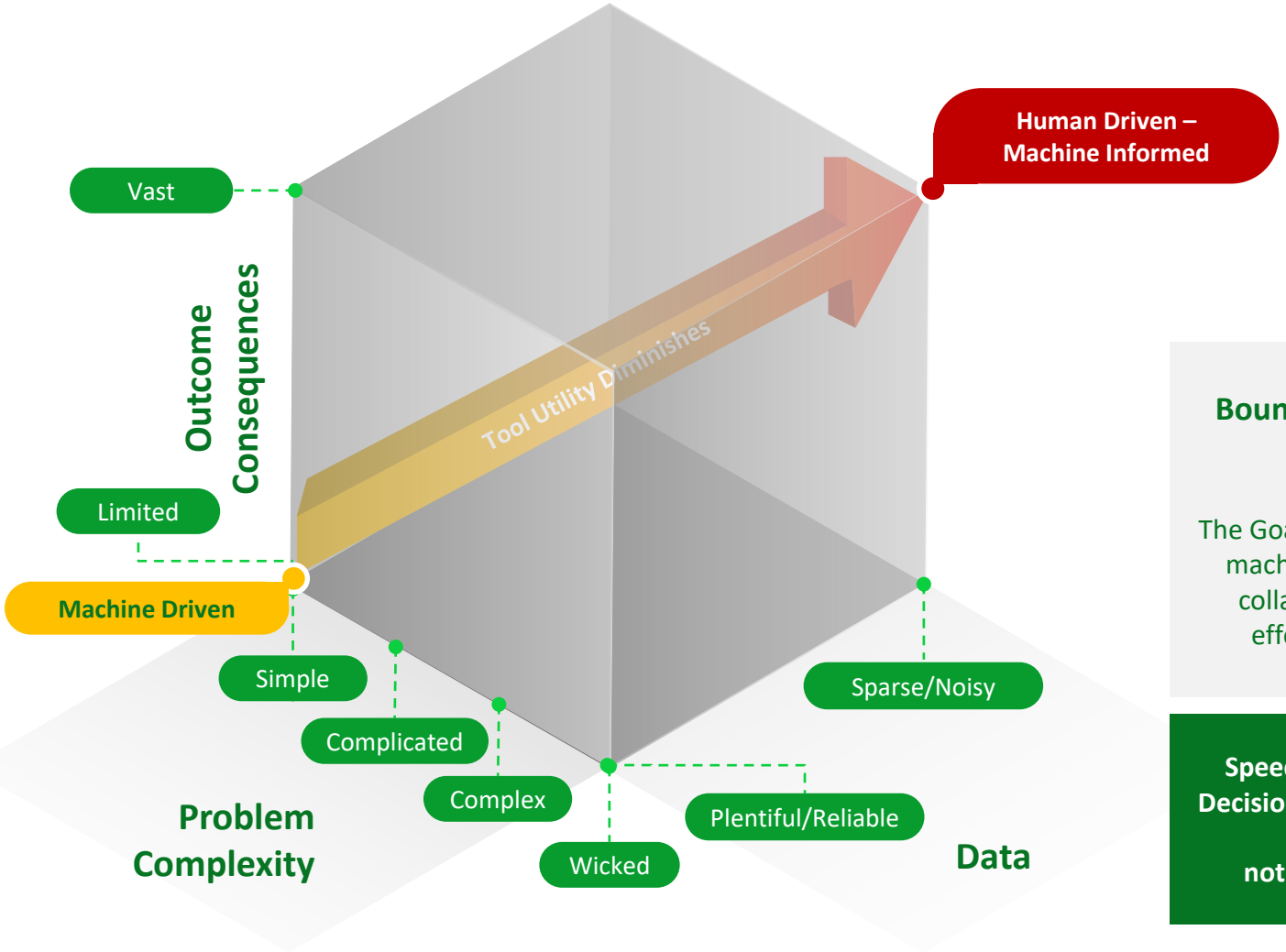
These factors can and should be systematically addressed (and reported)

The weight (value) of evidence is based on quality criteria.

Things like:



Pulling it all together – The Analytic Problem Space



Bounded Rationality is Real

-

The Goal should be to have machines and Humans collaborate for more effective decisions

Speed sometimes Kills
Decisions should be timely
–
not necessarily fast

The importance of human in the loop increases with situational complexity

Numbers and fluidity of objectives (wickedness of the problem)

Impact and size of irreducible uncertainty on the situation

There is settled science we can draw upon to improve predictive (inferential) performance

It appears the solution lies in improving Man – Machine Collaboration

Nugget – IARPA is already working the Problem (Hybrid Forecasting Program)

"There is no substitute for knowledgeable human eyes on the data"

Professor Loerch GMU

Copyright KaDSci LLC 2018 – All Rights Reserved

Bottom Line – What We Should Do for Every Analysis

01. Develop a set of objectives that reflect organizational priorities and can be estimated (you can't measure the future)

02. Clearly describe your decision space

- a. Alternatives
- b. Outcomes
- c. Risk

03. Identify some way to estimate the relative contribution of your alternatives against the objectives (e.g. simulation, analytic models, expert opinion, etc.)

04. Identify and consider your constraints explicitly (financial, physical, ..)

05. Synthesize all of these factors into a model or models using optimization or other search technique.

06. Exercise the heck out of the models so you understand what you see.

Then AND ONLY THEN are we ready to:

- Trust predictions
- Make recommendations
- Automate decisions (implement as AI)

Copyright KaDSci LLC 2018 – All Rights Reserved

You can't forget the thinking part!!!

Questions???

Comments

You can't forget the thinking part!!!

Copyright KaDSci LLC 2018 – All Rights Reserved

References

Hammond, K.R. (1996) Human Judgment and Social Policy, Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice. Oxford University Press: New York.

Kent, S. (1964) Words of estimative probability. *Studies in Intelligence*, 8, pp. 49-65.

Schum, D. (1994) The Evidential Foundations of Probabilistic Reasoning, John Wiley and Sons, New York.

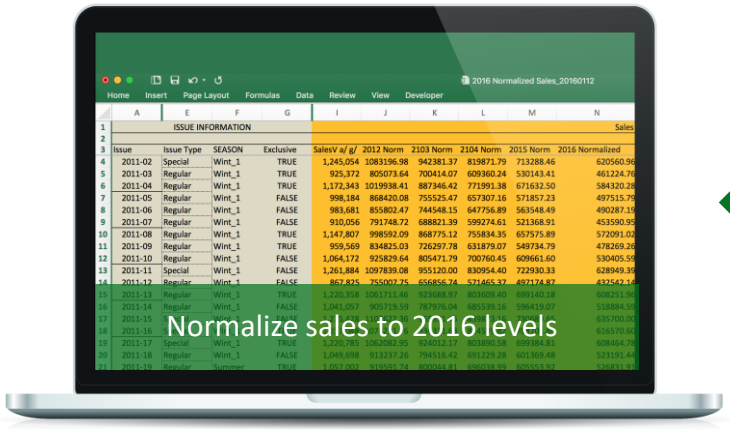
Copyright KaDSci LLC 2018 – All Rights Reserved

Decision Support for Magazine Cover Design

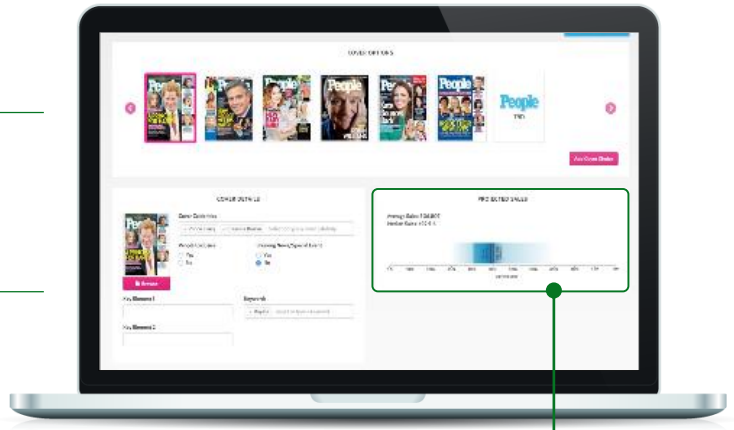
Copyright KaDSci LLC 2018 – All Rights Reserved

Modeling Approach

Document variables for 257 Issues



Next Step

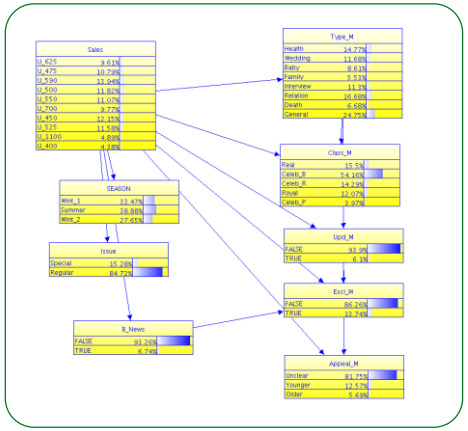


Estimated sales/
With variance

Input Cover Attributes

Generate synthetic data that
Expands 257 issues to 6,000+

Learn the Forecasting Model
Machine Learned Bayesian Network

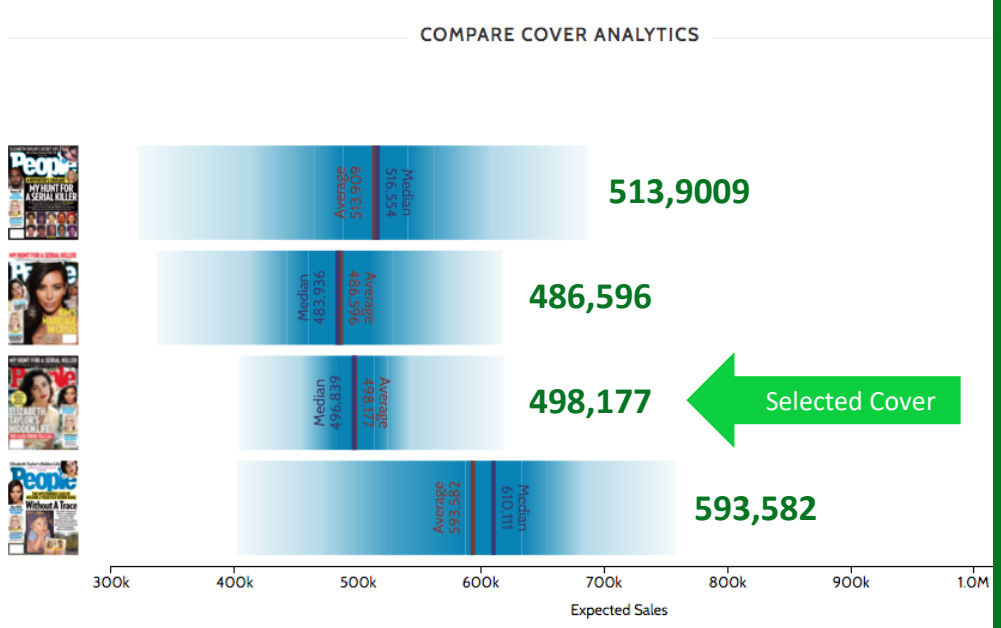


Copyright KaDSci LLC 2018 – All Rights Reserved

"It's tough to make predictions, especially about the future" – Yogi Berra

Copyright KaDSci LLC 2018 – All Rights Reserved

Forecasted Sales March 7th 2016 cover options



If we can believe the model & assuming a $3.00 profit per unit sold at the newsstand the opportunity cost of this decision is ~ $285K

Copyright KaDSci LLC 2018 – All Rights Reserved

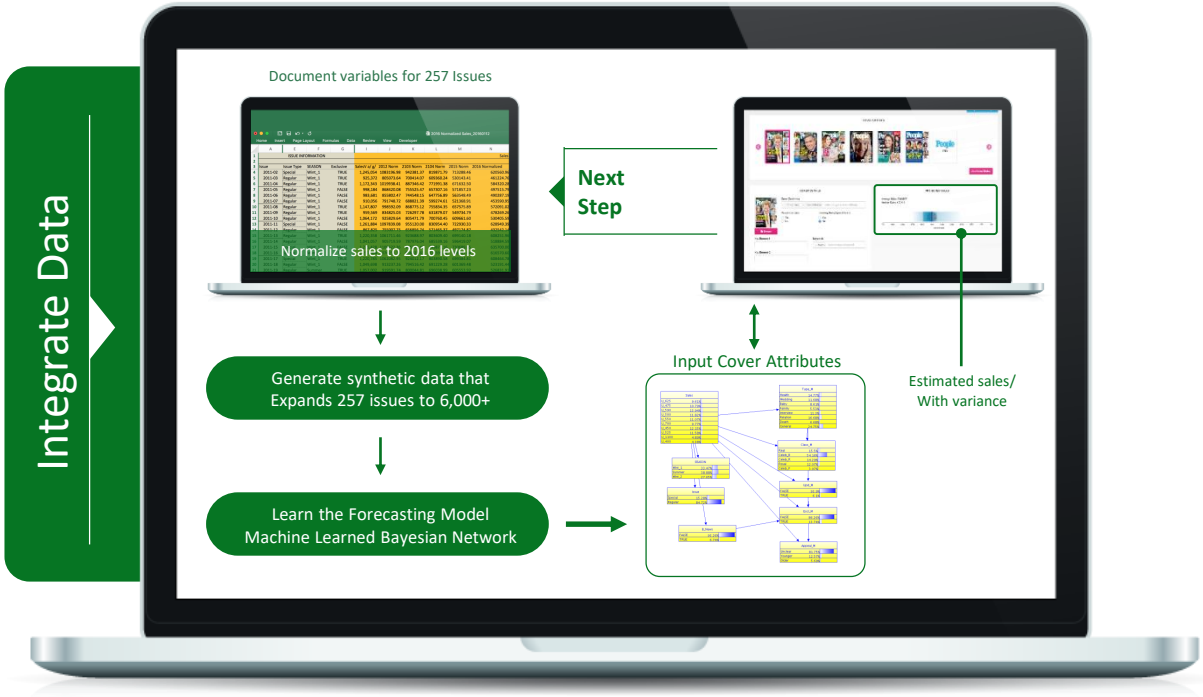
Current Model is at the summary level

Social Media Analysis is indicating that opinions about celebrities vary by location

For example Adelle is popular in the northeast US, weak negative perception in Pacific Northwest

Sales geographically

Celebrity Appeal Geographically



Copyright KaDSci LLC 2018 – All Rights Reserved