“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
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
What are the Pieces? How do they work?

Analysis

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!!!
Predictive Analytics & AI are here & they are all Models
All algorithms (and by extension AI tools) rely on data.
Not All Data are Created Equal

"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.
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.

He disobeyed those orders and declared the alert a false alarm.

He was neither praised or punished.

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

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.
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
Good Analytics

- Provides the foundation for useful models and tools
- Is multi-disciplinary, drawing on expertise from across the spectrum of science and engineering
Sherman Kent, CIA

---

**Note**
the large number of dots outside the selected regions

<table>
<thead>
<tr>
<th>100% Certainty</th>
<th>The General Area of Possibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>93%</td>
<td>Almost certain</td>
</tr>
<tr>
<td>75%</td>
<td>give or take about 6%</td>
</tr>
<tr>
<td>50%</td>
<td>Probable</td>
</tr>
<tr>
<td>30%</td>
<td>give or take about 10%</td>
</tr>
<tr>
<td>7%</td>
<td>Almost certainly not</td>
</tr>
<tr>
<td>0%</td>
<td>Impossibility</td>
</tr>
</tbody>
</table>

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:

- Bias
- Veracity
- Observational Sensitivity
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
<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>01.</td>
<td>Develop a set of objectives that reflect organizational priorities and can be estimated (you can’t measure the future)</td>
</tr>
<tr>
<td>02.</td>
<td>Clearly describe your decision space</td>
</tr>
<tr>
<td></td>
<td>a. Alternatives</td>
</tr>
<tr>
<td></td>
<td>b. Outcomes</td>
</tr>
<tr>
<td></td>
<td>c. Risk</td>
</tr>
<tr>
<td>03.</td>
<td>Identify some way to estimate the relative contribution of your alternatives against the objectives (e.g. simulation, analytic models, expert opinion, etc.)</td>
</tr>
<tr>
<td>04.</td>
<td>Identify and consider your constraints explicitly (financial, physical, ..)</td>
</tr>
<tr>
<td>05.</td>
<td>Synthesize all of these factors into a model or models using optimization or other search technique.</td>
</tr>
<tr>
<td>06.</td>
<td>Exercise the heck out of the models so you understand what you see.</td>
</tr>
</tbody>
</table>

Then AND ONLY THEN are we ready to:

- Trust predictions
- Make recommendations
- Automate decisions (implement as AI)
You can’t forget the thinking part!!!


Decision Support for Magazine Cover Design
Document variables for 257 Issues

Normalize sales to 2016 levels

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

Learn the Forecasting Model
Machine Learned Bayesian Network

Next Step

Input Cover Attributes

Estimated sales/ With variance
“It’s tough to make predictions, especially about the future” – Yogi Berra
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
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.