October 22, 2017

All Models are Wrong; Some Models are Harmful

David Hunt, Oliver Wyman
Michael P Johnson, Jr.

Available at: https://works.bepress.com/michael_johnson/92/
ALL MODELS ARE WRONG; SOME MODELS ARE HARMFUL

OCTOBER 22, 2017
All models are wrong; some models are useful

– George E.P. Box

All models are wrong; some models are useful harmful
The INFORMS Ethics Guidelines contains several topics addressing the intentional misuse of models to mislead or to deceive…

INFORMS Members should aspire to be:

• **Forthcoming** about our assumptions, interests, sponsors, motivations, limitations, and potential conflicts of interest

• **Honest** in reporting our results, even when they fail to yield the desired outcome

• **Objective** in our assessments of facts, irrespective of our opinions or beliefs
INFORMS Members should aspire to be:

- **Accountable** for our professional actions and the impact of our work
- **Alert** to possible unintended or negative consequences that our results and recommendations may have on others
- **Rigorous** by adhering to proper professional practices in the development and reporting of our work
“But very often OR is more concerned with improving the workings of the current system than with exploring the fundamental ethical and moral issues involved.”¹

• OR professionals are often presented with a situation and asked to make it “better,” where “better” involves:
  – Cost reduction
  – Efficiency improvements
  – Increased production and output
  – Fewer failures
  – Revenue enhancements

• OR professionals are rarely presented with a situation and asked to make it “better,” where “better” involves:
  – Reduction of existing biases against humans in a current process
  – Fair treatment of all, free from any discrimination
  – Optimization based on equality, rather than the bottom line

To understand how the unintentional misuse of community-based operations research and analytical models can potentially cause harm, it is important to understand three basic ways that biases can enter models.

- Asking the wrong question, or failing to follow the generally accepted professional practices, can lead to harmful results.
- Unaddressed biases and omissions in the data will lead to biases and omissions in the results.
- Selection of the wrong type of mathematical model, or using improper assumptions in the model, can lead to misleading and potentially harmful results.
Case Study 1: Incorrect framing of the problem
Staying focused on the end, not the means. But what is the correct “end?”

<table>
<thead>
<tr>
<th>Employer Perspective</th>
<th>Employee Perspective</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The “end” is increased profits and growth</strong></td>
<td><strong>The “end” is a better quality of life for their family and for themselves</strong></td>
</tr>
<tr>
<td>• Adjustable schedules allow aligning employees with customer demand</td>
<td>• Adjustable schedules make it difficult to attend school if working toward a better job</td>
</tr>
<tr>
<td>• On-call allows adjusting to changing or unexpected demand</td>
<td>• On-call makes it difficult to plan child care</td>
</tr>
<tr>
<td>• Hours can be manipulated to avoid paying benefits and overtime</td>
<td>• Optimal solutions can have employees closing a store one night and opening it the next morning, with little time for rest</td>
</tr>
<tr>
<td>• Employees are often not allowed to provide input on schedules</td>
<td>However, the employee has little loyalty and will leave if a better opportunity is available</td>
</tr>
</tbody>
</table>

However, the employer must remain profitable to continue operating and creating jobs.

Seattle passed a new ordnance that requires, among other things: advance notice of work shift; employee input to schedules; sufficient rest between shift; on-call pay; and, access to more hours for existing employees before new employees can be hired.

Case Study 2: Use of biased data
Using historical data perpetuates historical patterns

Sexist Models
**Purpose:** training AI algorithms to recognize situations by studying large photo libraries.
**Bias:** the algorithm linked Images of shopping and washing to women, for example, while coaching and shooting were linked to men.
**Harm:** perpetuates historical stereotypes and, according to the researchers, in some cases amplifies them.

Predictive Policing
**Purpose:** predicts where police presence is needed, and allocates more resources to those locations.
**Bias:** often based on historical data; placing police where crimes have occurred, rather than where they might occur.
**Harm:** more police officers in an area increases likelihood that even minor crimes are detected, perpetuating “high-crime” areas.

Loan Approvals
**Purpose:** algorithms are used to approve/refuse loans, speeding up the process and cutting costs.
**Bias:** by using historic credit ratings, the models are biased toward past ability to repay the loan, rather than future ability.
**Harm:** makes it very difficult for people who have had past problems, but are trying hard for a better life.

Although the models are biased, do humans make less biased decisions?

Case Study 3: Improper model type and/or model assumptions
Desire to protect intellectual property can lead to black boxes, making it hard to detect both intentional and unintentional biases, and whether or not proper procedures and assumptions were used

**Granting Parole**
- A system called COMPAS offers to predict defendants’ likelihood of reoffending, and is used by some judges to determine whether an inmate is granted parole.
- The workings of COMPAS are kept secret, but an investigation by *ProPublica* found evidence that the model may be biased against minorities.

**Public school closings**
- Can be presented by government to parents, students and community members as a ‘done deal’
- Policy rationale often obscured, incomplete, misrepresented or absent, or conflated with and/or substituted for a political rationale

**Strategic plans**
- Can be sold on the basis of dubious, hidden, misrepresented or poorly-understood policy modeling assumptions.
- For example, an urban university’s ambitious but expensive and disruptive plans to re-make the physical infrastructure assumed it would be paid for by revenues arising from 4% enrollment growth.
- The justification supporting a 4% growth estimate was unclear.

Sources:  
(3) Walker, A., “UMass Boston’s Present is a Mess,” 2017, Boston Globe,
CONCLUSION: AVOIDING HARMFUL MODELS

- Adhere to proper professional practices
- Understand the subject matter, or work with an expert who does understand the subject matter
  - Seek the opinion of someone who will use the results of the model
- Trust your instincts
  - If it seems wrong, it most likely is wrong
- Be alert to possible biases
  - Question results that are unexpected
  - Question results that promote stereotypes
- Put yourself in the place of someone who might be impacted by the results
  - Are you still happy with the recommendations?
  - Consider Immanuel Kant’s Categorical Imperative “Act only according to that maxim whereby you can at the same time will that it should become a universal law.”

1. Immanuel Kant, *Grounding for the Metaphysics of Morals*
Be part of Pro Bono Analytics!

If you are interested in community-based operations research and analytics, sign up today as a volunteer on the Pro Bono Analytics web site, or contact a member of the Pro Bono Analytics Committee

www.probonoanalytics.org
probono@informatics.org

Thank You!