From Algorithmic Transparency to Explanation

Nicholas Diakopoulos
University of Maryland College Park
@ndiakopoulos | nad@umd.edu
Live Presidential Forecast

Donald J. Trump has won the presidency. Below, our estimates of three key indicators:

- **Chance of Winning Presidency**: >95% Trump
- **Popular vote margin**: Clinton +1.2
- **Electoral votes**: 306 Trump

The projections for each state are based on the votes reported so far and how those places have voted in previous elections.

We’re showing the closest states by default.  

<table>
<thead>
<tr>
<th>State</th>
<th>Est. pct. of votes</th>
<th>Reported margin</th>
<th>NYT projection</th>
<th>NYT win prob.</th>
<th>Even</th>
<th>+2%</th>
<th>+4%</th>
<th>+6%</th>
<th>+8%</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Mexico</td>
<td>100%</td>
<td>Clinton +8.2</td>
<td><strong>Clinton +8.2</strong></td>
<td>✔ WON</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virginia</td>
<td>&gt;95%</td>
<td>Clinton +4.8</td>
<td><strong>Clinton +4.9</strong></td>
<td>✔ WON</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Transparency is defined as “the ways in which people both inside and external to journalism are given a chance to monitor, check, criticize and even intervene in the journalistic process.”

– Mark Deuze, University of Amsterdam
What could be made **public** about the algorithm?
What are **limits** of algorithmic transparency?
What are comprehensive **disclosure** mechanisms?
BUSINESS CONCERNS

Costs of information production (data prep, documentation, benchmarking)

Privacy issues

Legal concerns over admitting errors / uncertainties

Trade secrets & competitive advantage

Manipulation & gaming
Transparency information needs to be **accessible** and **understandable**

Needs to be **explained** in context.

What’s the right vehicle for this information?

“If it’s detailed enough for somebody to understand the science of algorithms to really use critically, maybe 1 percent of the readership would be able to use it.”
OPEN QUESTIONS

User Modeling / User Studies
What are the decisions the public or other stakeholders would make based on the algorithm in question? What could a user know about an algorithm that would change their interaction, decision, or outcome? How would users respond to this information?

Manipulation – “Threat modeling”
If a piece of information were made available about the algorithm, how might it be gamed, manipulated, or circumvented? Who would stand to gain or lose from that manipulation? What are the consequences of that manipulation to the public or organization? Can we build manipulation resistance?

Cost
What degree of time/money/effort would be required by the creator to publish the relevant transparency info?

Presentation
What would be an appropriate vehicle or format to publish this info? How could this information be presented in a salient way to users? How often would this information need to be disclosed?
How BuzzFeed News Used Betting Data To Investigate Match-Fixing In Tennis

With GIFs.

posted on Jan. 17, 2016, at 5:02 p.m.

John Templon
BuzzFeed News Reporter

I’m John Templon, an investigative data reporter for BuzzFeed News. I spent the past 15 months analyzing tennis betting data to see if I could figure out whether players were fixing matches.
Methodology and Code: Detecting Match-Fixing Patterns In Tennis

A closer look at the data analysis behind BuzzFeed News’ investigation into corruption in tennis.

General Notes

In “The Tennis Racket,” a yearlong investigation into match-fixing in professional tennis, BuzzFeed News published findings from an original data analysis we performed. That analysis revealed many examples of one particularly suspicious pattern: heavy betting against a player, followed by that player's loss.

Betting patterns alone aren’t proof of fixing. Players can underperform for all sorts of reasons — injury, fatigue, bad luck — and sometimes that underperformance will just happen to coincide with heavy betting against them. But it’s extremely unlikely for a player to underperform repeatedly in matches on which people just happen to be betting massive sums against him.
Data and Analysis: Detecting Match-Fixing Patterns In Tennis

The Python code below runs the anonymized implementation of the methodology described here that was used in “The Tennis Racket”. The methodology contains many important details. Please read it before continuing here.

Importing The Data

```python
In [1]: import pandas as pd
   import random

In [2]: betting_data = pd.read_csv("../data/anonymous_betting_data.csv")
```

Match Selection

The code below excludes opening odds that implied probabilities more than 10 percentage points higher or lower than the median of all bookmakers’ opening odds for the match. (Otherwise the return of these odds toward the consensus could be mistaken for a sign of suspicious betting.) The code also excludes matches that were noted as “canceled” — typically a result of pre-match withdrawals — or “walkover” on OddsPortal.

```python
In [3]: def get_outlier_openings(match_books):
   median = match_books['implied_prob_winner_open'].median()
   return match_books[(match_books['implied_prob_winner_open'] - median).abs() > 0.1]

In [4]: outlier_openings = betting_data['match_uid'].apply(get_outlier_openings)

In [5]: selected_betting_data = betting_data[
   ~betting_data['match_book_uid'].isin(outlier_openings['match_book_uid']) &
   ~betting_data['is_canceled_or_walkover'])
```

Transparency
About the “what” behind an algorithm
Information to inform an explanation

Vs.

Explanation
About the “how” of an algorithmic decision / outcome
Causality and delivering on question of “why”
May incorporate human intent in sociotechnical
Adapted to stakeholders and what they care about
Thanks! Questions?

Contact
Nick Diakopoulos
Twitter: @ndiakopoulos
Email: nad@umd.edu
Web: http://www.nickdiakopoulos.com
Web: http://www.github.com/comp-journalism

Related Papers