Algorithmic and Data Transparency in NYC Agencies: Tools and Strategies

Julia Stoyanovich
Drexel University & Princeton CITP

data RESPONSIBLY
Outline

• Int. No. 1696-A: A Local Law in relation to automated decision systems used by agencies

• comments on the Law

• strategies for success
Form an automated decision systems (ADS) task force that surveys current use of algorithms and data in City agencies and develops procedures for:

- requesting and receiving an explanation of an algorithmic decision affecting an individual (3(b))
- interrogating ADS for bias and discrimination against members of legally-protected groups (3(c) and 3(d))
- allowing the public to assess how ADS function and are used (3(e)), and archiving ADS together with the data they use (3(f))
Mayor de Blasio Announces First-In-Nation Task Force To Examine Automated Decision Systems Used By The City

May 16, 2018

NEW YORK— Today, Mayor de Blasio announced the creation of the Automated Decision Systems Task Force which will explore how New York City uses algorithms. The task force, the first of its kind in the U.S., will work to develop a process for reviewing “automated decision systems,” commonly known as algorithms, through the lens of equity, fairness and accountability.

“As data and technology become more central to the work of city government, the algorithms we use to aid decision making must be aligned with our goals and values,” said Mayor de Blasio. “The establishment of the Automated Decision Systems Task Force is an important first step towards greater transparency and equity in our use of technology.”
algorithmic transparency is not synonymous with releasing the source code

publishing source code helps, but it is sometimes unnecessary and often insufficient

syntactic vs. semantic transparency

the interplay between code and data
Point 2

algorithmic transparency requires data transparency

data is used in training, validation, deployment

validity, accuracy, applicability can only be understood in the data context
data transparency is not synonymous with making all data public

release data whenever possible; also release:
data selection, collection and pre-processing methodologies; data provenance and quality;
dataset composition, statistical properties, sources of bias; validation methodologies
University Researchers Use 'Fake' Data for Social Good

Virtually every interaction we have with a public agency creates a data point. Amass enough data points and they can tell a story. However, factors like privacy, data storage and usability present challenges for local governments and researchers interested in helping improve services. In this installment of MetroLab's Innovation of the Month series, we highlight researchers at Data Responsibly are addressing those challenges by creating synthetic data sets for social good.

Since its development, the tool has been receiving a lot of attention. For example: T-Mobile is interested in generating synthetic data to better engage with researchers and improve transparency for customers, the Colorado Department of Education has asked relevant agencies to use the tool to experiment with sharing sensitive data, and Elsevier is interested in using the tool to generate synthetic citation networks for research.
actionable transparency requires interpretability

explain assumptions and effects, not details of operation

engage the public - technical and non-technical
Point 5

transparency by design, not as an afterthought

provision for transparency and interpretability at every stage of the data lifecycle

useful internally during development, for communication and coordination between agencies, and for accountability to the public
responsible data science requires a holistic view of the data lifecycle
Responsibility by design

Systems support for responsible data science

Responsibility by design, managed at all stages of the lifecycle of data-intensive applications

responsible data science requires a holistic view of the data lifecycle

Stoyanovich, Howe, Abiteboul, Miklau, Sahuguet, Weikum - SSDBM 2017
Point 6

transparency is a challenge and an opportunity

lots of ongoing research, but not a solved problem

will require time and resources to get right - we need all hands on deck

the GDPR is drawing tremendous technological investment in the EU, the NYC algorithmic transparency law should be our opportunity
Strategies

build on NYC Open Data Law
leverage public engagement
leverage the research community
learn from others