Data Science Approaches to Assess Suicide Risk: Defining the Jobs Before Picking the Tools

Greg Simon
Kaiser Permanente Washington Health Research Institute

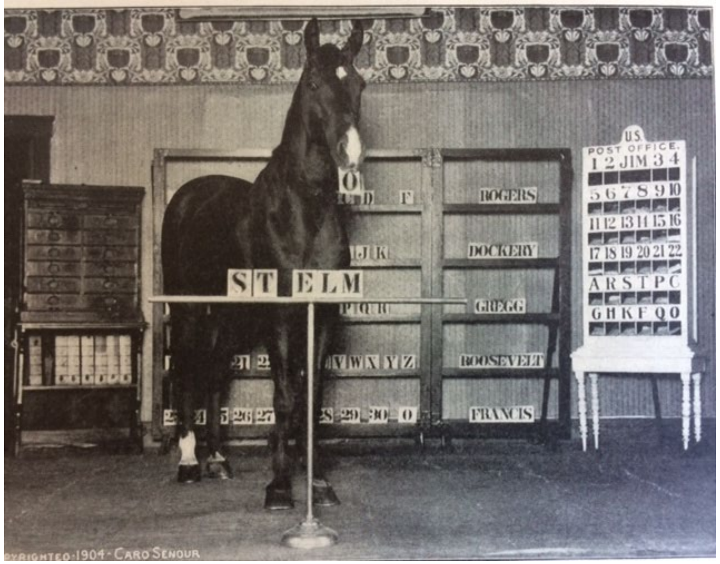
Support from:
NIMH: U19MH121738
FDA: HHS F223201810201C
No other disclosures

MHRN Mental Health Research Network

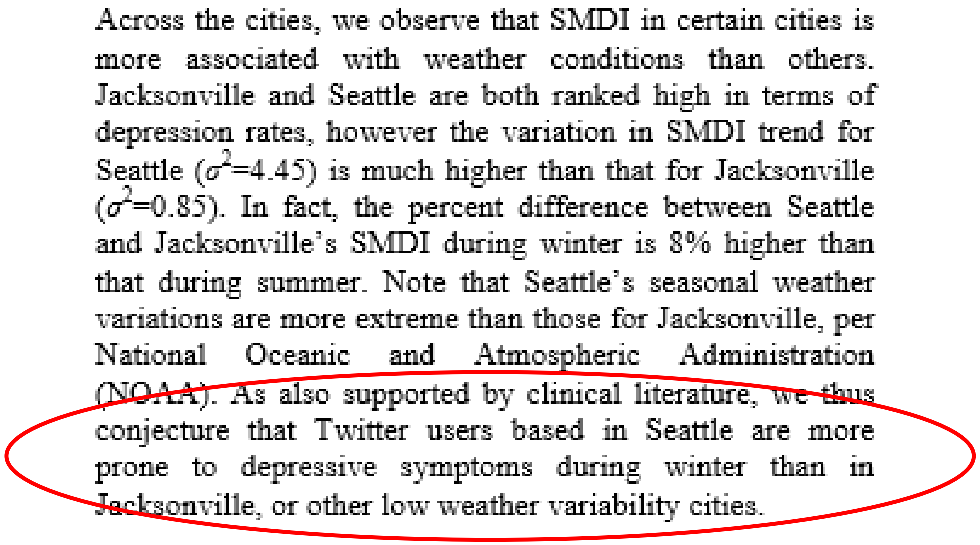
KAISER PERMANENTE®

Tools searching for jobs

1900: Clever Hans, the Calculating Horse



2020: Mining Twitter for winter depression



Across the cities, we observe that SMDI in certain cities is more associated with weather conditions than others. Jacksonville and Seattle are both ranked high in terms of depression rates, however the variation in SMDI trend for Seattle (σ^2=4.45) is much higher than that for Jacksonville (σ^2=0.85). In fact, the percent difference between Seattle and Jacksonville's SMDI during winter is 8% higher than that during summer. Note that Seattle's seasonal weather variations are more extreme than those for Jacksonville, per National Oceanic and Atmospheric Administration (NOAA). As also supported by clinical literature, we thus conjecture that Twitter users based in Seattle are more prone to depressive symptoms during winter than in Jacksonville, or other low weather variability cities.

Inference vs. Prediction

- Inference is about why (understanding)
  - We ask: Is it generally true?
- Prediction is about who and when (sorting)
  - We ask: Is it useful?

Prediction vs. Detection

- Prediction is about the future
  - Outcome is something that has not yet happened
- Detection is about the present
  - Outcome is something happening now (but we can't easily see it)

Four types of jobs:

- Inference for generalizable knowledge

  Example: Students who are subject to online bullying are at high risk.

- Detection of “hotspots” for community-level intervention

  Example: Students in this high school are at high risk now.

- Detection for individual-level intervention

  Example: This student is experiencing suicidal ideation now.

- Prediction for individual-level intervention

  Example: This student will more likely attempt suicide in the next month.

A tool should serve specific “customers”

- Clinical or public health decision
- Information available
- Potential actions
- Consequences of errors