“Digital Social Studies”

digital data and computational approaches in social studies

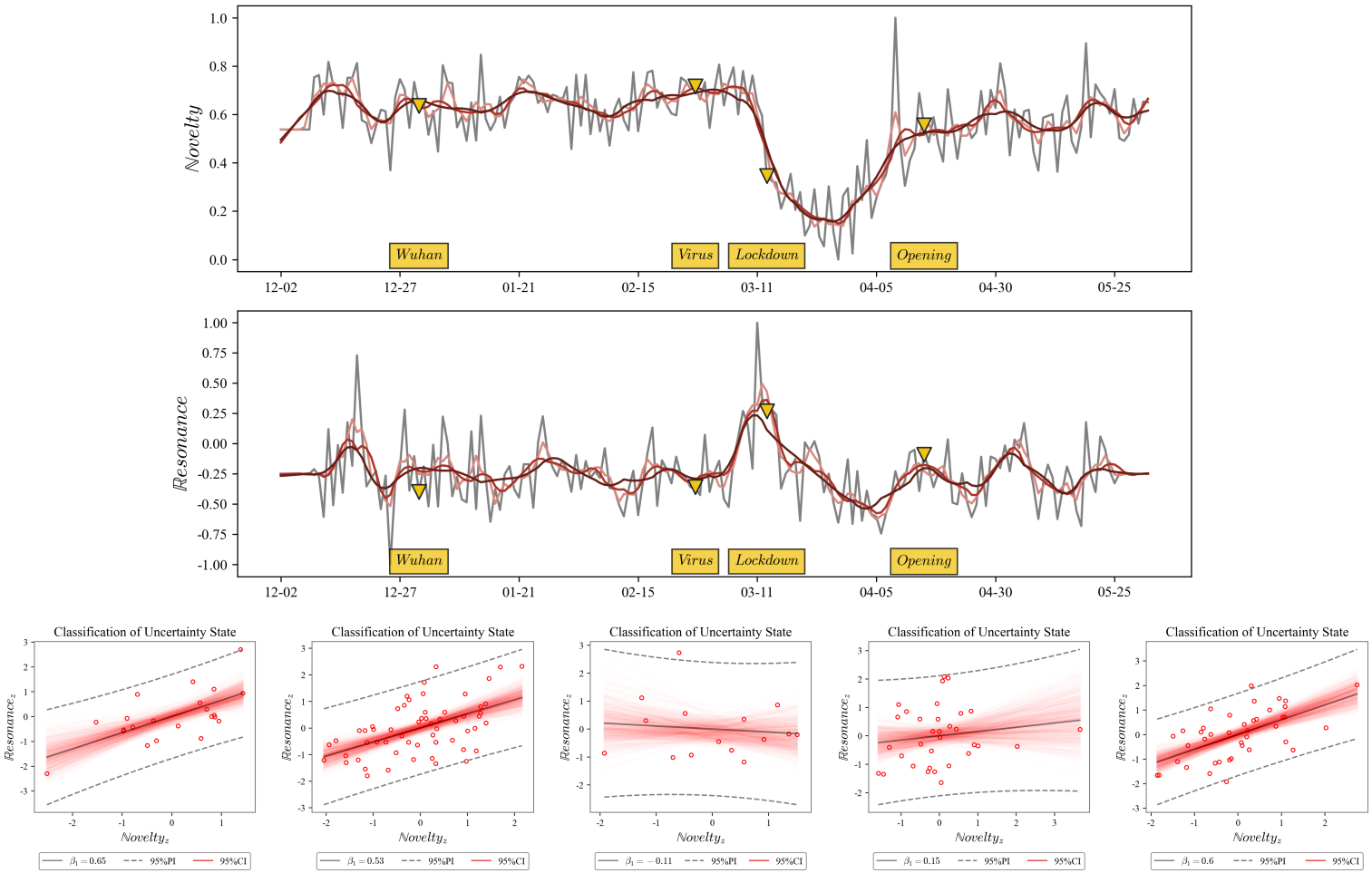
K. Nielbo
knielbo.github.io

Center for Humanities Computing Aarhus|chcaa.io
Aarhus University, Denmark

CENTER FOR HUMANITIES
COMPUTING AARHUS

Outline

News Information Decoupling

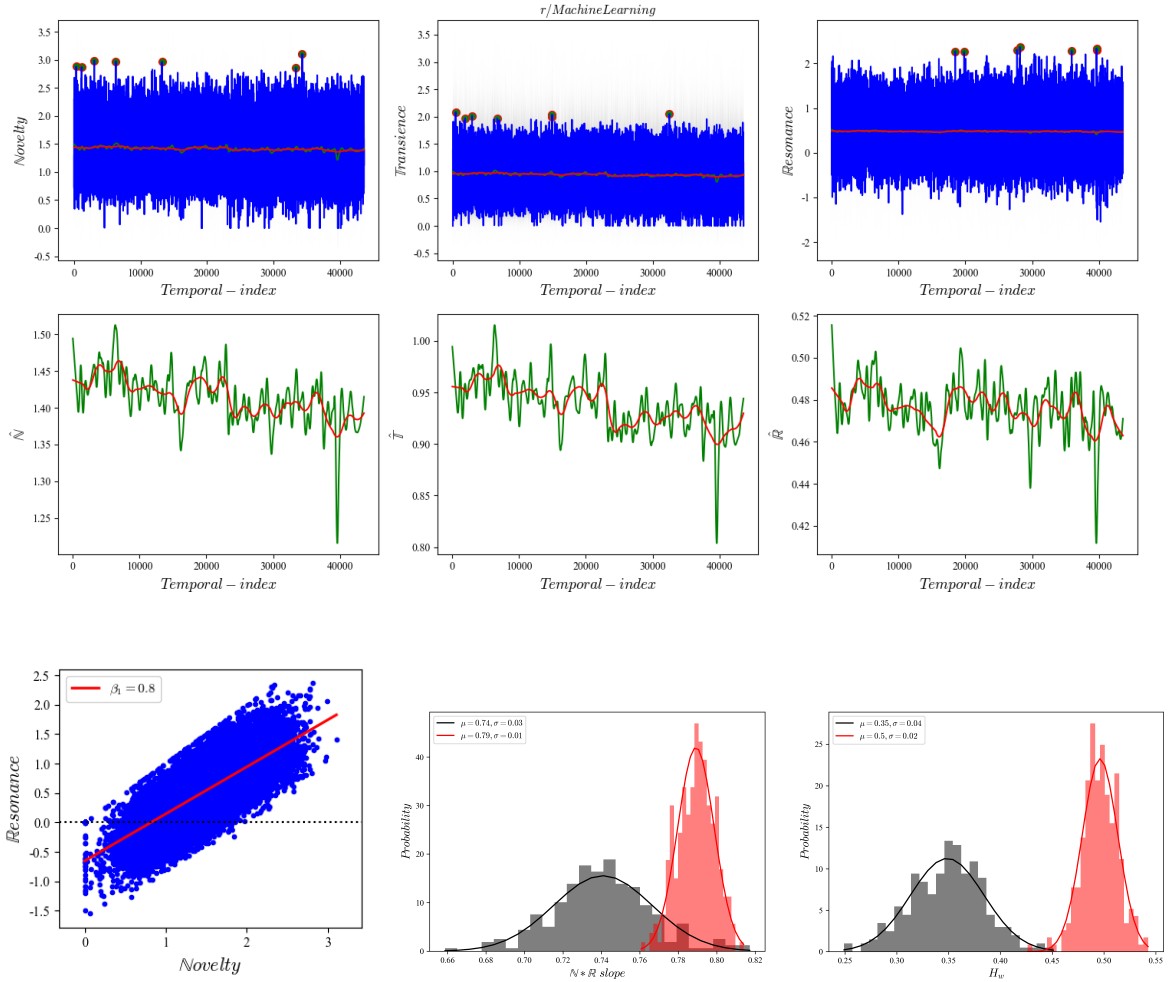


front pages from danish legacy print media *politiken* during Covid-19 phase 1

K. L. Nielbo, R. B. Baglini, P. B. Vahlstrup, K. C. Enevoldsen, A. Bechmann, and A. Roepstorff (2021) "News Information Decoupling: An Information Signature of Catastrophes in Legacy News Media," arXiv:2101.02956 [cs]

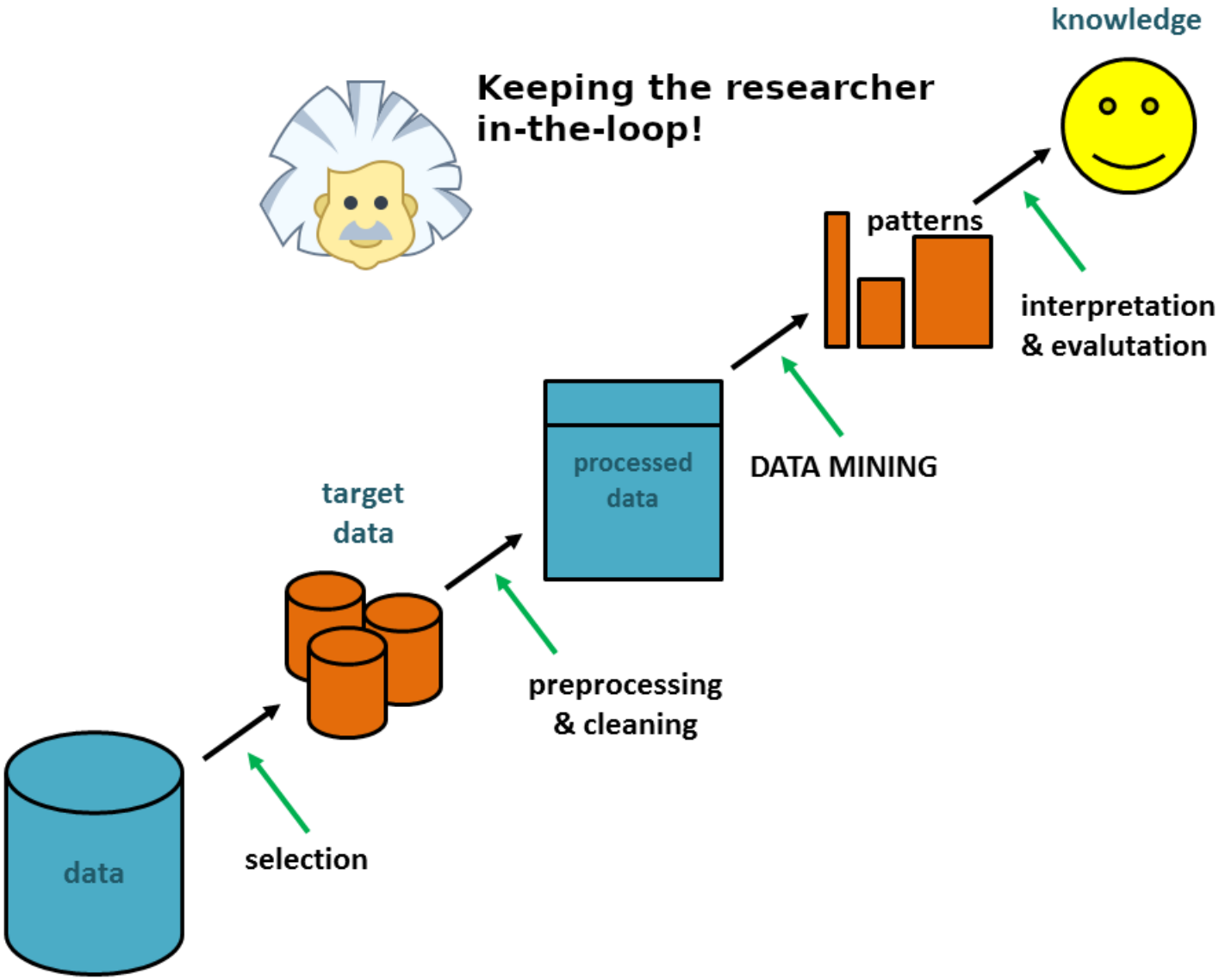
K. L. Nielbo, F. Haestrup, K. C. Enevoldsen, P. B. Vahlstrup, R. B. Baglini, A. Roepstorff (2021) "When no news is bad news – Detection of negative events from news media content," arXiv:2102.06505 [cs]

Trend Reservoir Detection



trend reservoirs (i.e., social media signals that display a high trend potential) can be identified by their relationship between novel and resonant behavior, and their minimal persistence.

Nielbo, K.L., Vahlstrup, P.B., Gao, J. & Bechmann A. (2020) Trend Reservoir Detection: Minimal Persistence and Resonant Behavior of Trends in Social Media. Proceedings of CHR 2020.


knowledge
Keeping the researcher in-the-loop!
patterns
interpretation & evalutation
processed data
DATA MINING
target data
preprocessing & cleaning
data
selection

Human-in-the-Loop Models

as task complexity increases, a need for (operational approaches to) leveraging human intelligence in the development of learning algorithms has become apparent

Type	Human Involvement	Resources	Relevance
Out-of-the-loop	not required	low	low
On-the-loop	checking	medium	medium↓
In-the-loop	required	high	medium↑

WHEN

algorithms are not understanding the input

data input is interpreted incorrectly

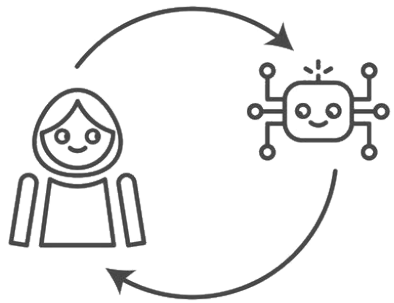
algorithms do not know how to perform the task

to make models more accurate

cost of errors is too high in development

data is rare or not available

THEN



HITL Models

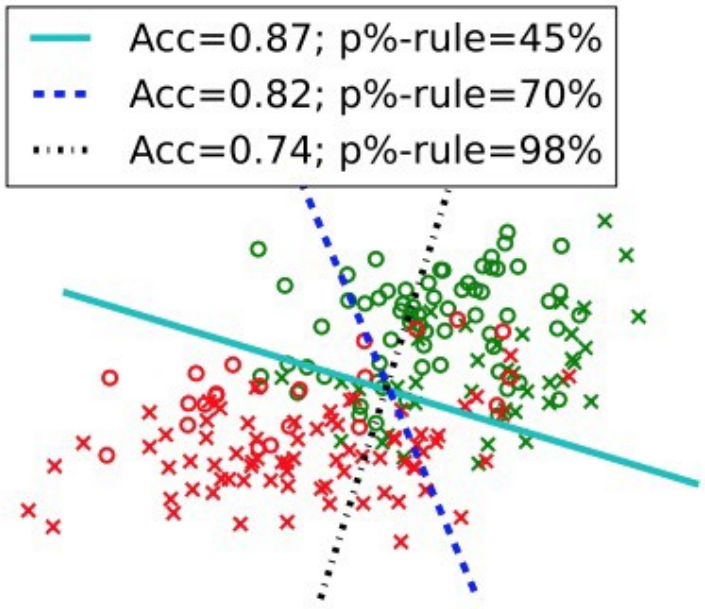
Statistical Parity

Assume differing base rates, $Pr_a(Y = 1) \neq Pr_b(Y = 1)$, and an imperfect learning algorithm, $C \neq Y$, then you cannot simultaneously achieve:

Precision parity $Pr_a(Y = 1 \mid C = 1) = Pr_b(Y = 1 \mid C = 1)$

True positive parity $Pr_a(C = 1 \mid Y = 1) = Pr_b(C = 1 \mid Y = 1)$

False positive parity $Pr_a(C = 1 \mid Y = 0) = Pr_b(C = 1 \mid Y = 0)$



`soft breakdown of performance` (precision, recall) in response to ethical and legal requirements, e.g., demographic parity at level x 'costs' $pr = y$

Kleinberg J., S. Mullainathan, & M. Raghavan (2016), Inherent Trade-Offs in the Fair Determination of Risk Scores, arXiv:1609.05807

THANKS

kln@au.dk
knielbo.github.io
chcaa.io

SLIDES

knielbo.github.io/files/kln_dstp21.pdf