# Deriving Population-Level Insights in Digital Mental Health: Opportunities and Challenges of Social Media

Munmun De Choudhury

munmund@gatech.edu

School of Interactive Computing | Georgia Tech

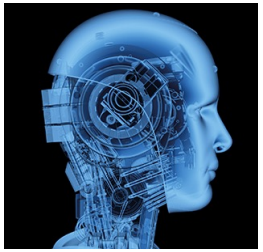
SocWeB

Georgia Tech College of Computing

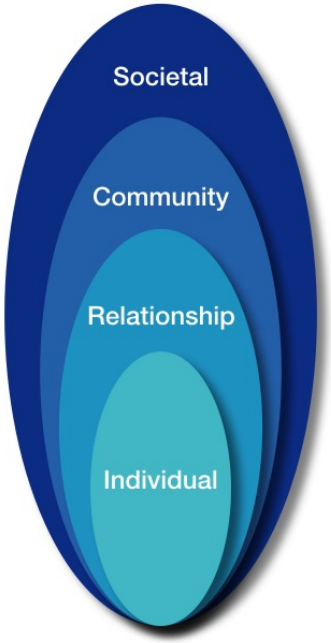
Social Dynamics and WellBeing Lab

Computational and AI artifacts for social good
Understand and improve wellbeing

+



+




Social Media

Human/Stakeholder-Centered
AI + Interdisciplinary

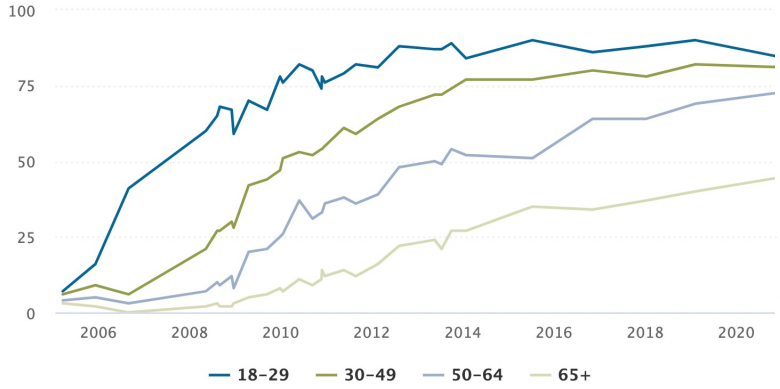
Theory Centered

No relevant disclosures or COIs

Sorrowing Old Man ('At Eternity's Gate'), by van Gogh

# A Revolution

*% of U.S. adults who say they use at least one social media site, by age*

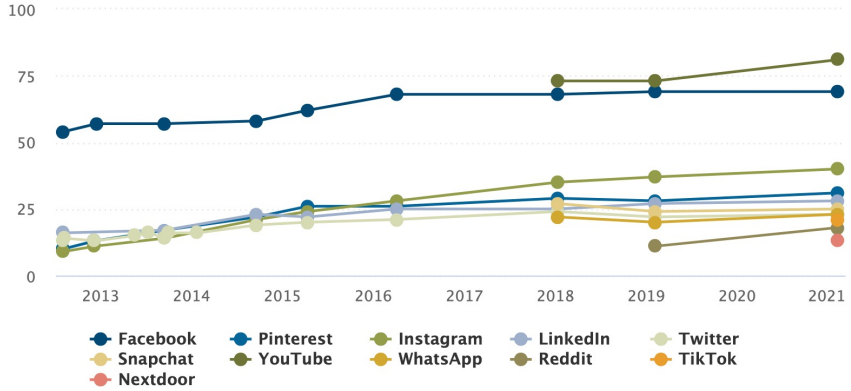



Note: Respondents who did not give an answer are not shown.

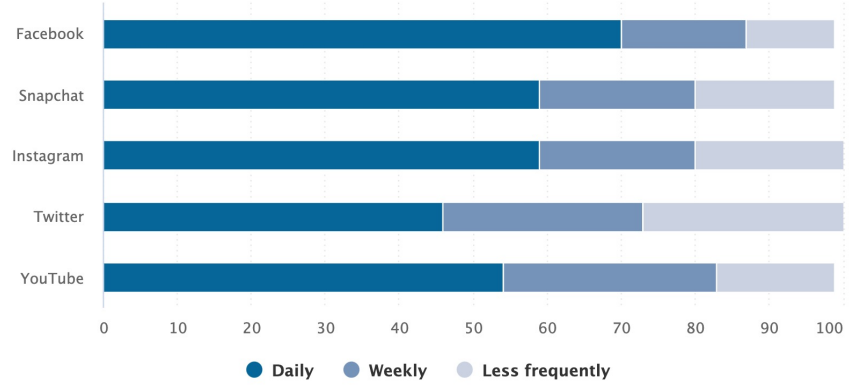
Source: Surveys of U.S. adults conducted 2005-2021.

PEW RESEARCH CENTER

*% of U.S. adults who say they ever use ...*




*Among U.S. adults who say they use ___, the % who use each site ...*

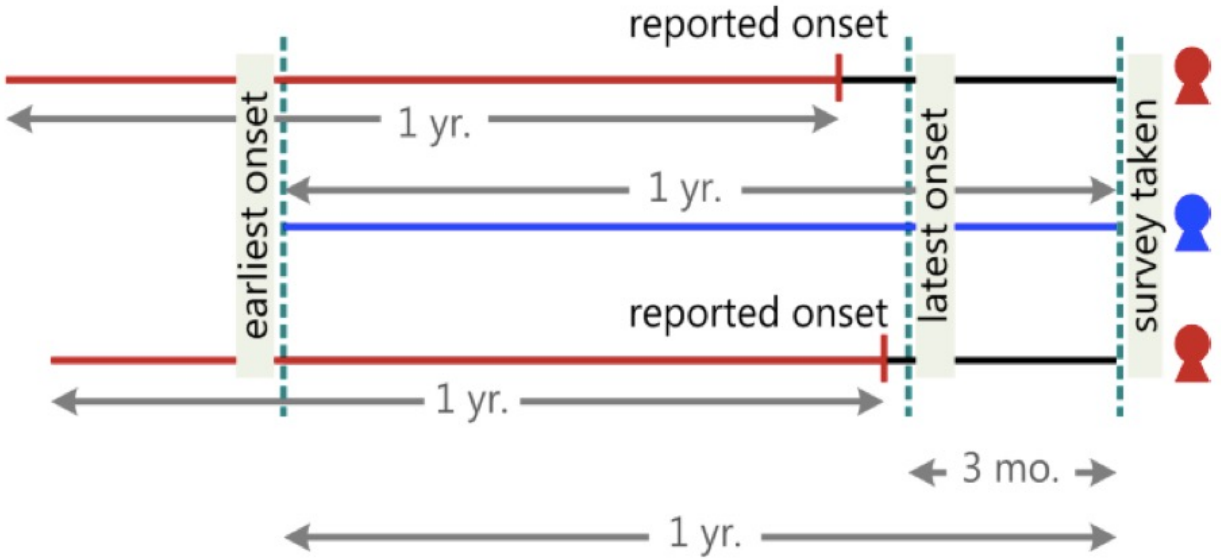


# Social Media – A New Potential in Digital Mental Health

(De Choudhury, Gamon, Counts, Horvitz, ICWSM2013)

# Social Media Language and Behavior Predicts Onset of Major Depressive Disorder

- 476 individuals (233 female) recruited via Amazon's Mechanical Turk; took CES-D/BDI and consented to accessing their Twitter data



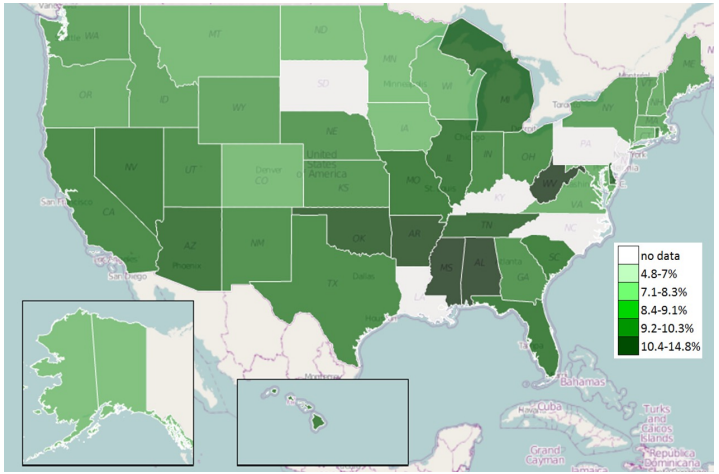
(De Choudhury, Counts, Horvitz, WebSci 2013)

# A Social Media Index

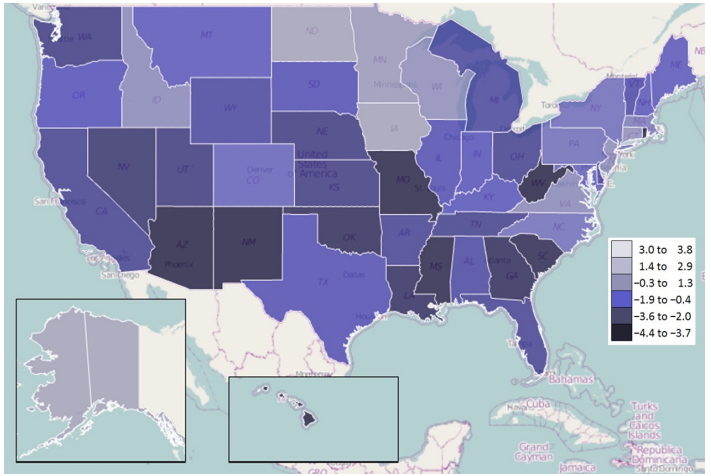
$$SMDI(t) = \frac{n_d(t) - \mu_d}{\sigma_d} - \frac{n_s(t) - \mu_s}{\sigma_s}$$

standardized difference between frequencies of depression-indicative and standard posts, compared to a period before between *k* and *t-1* ($1 \le k \le t\text{-}1$)

*actual (BRFSS data)*



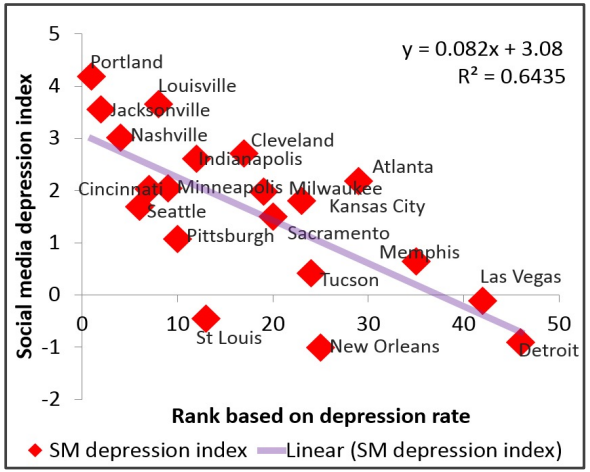

*predicted (SMDI)*

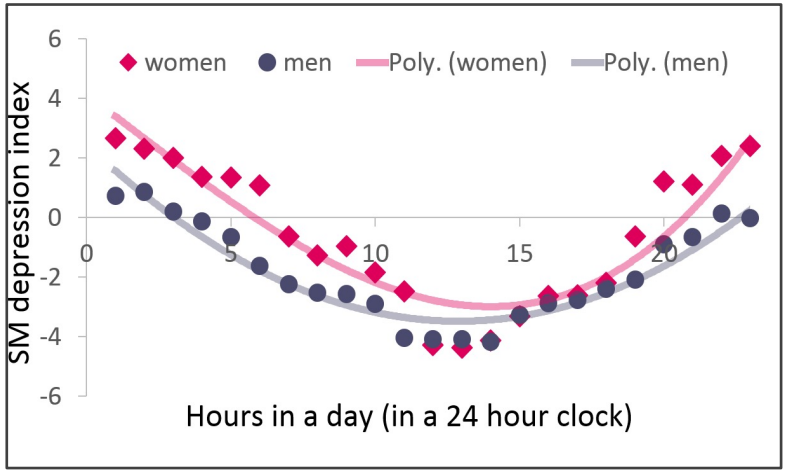


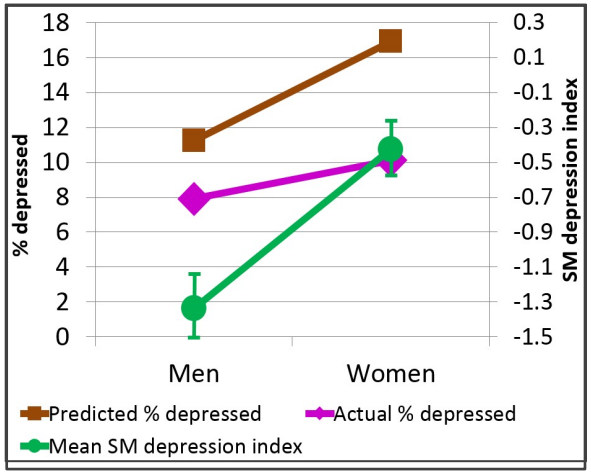

least squares regression fit yields correlation of 0.52

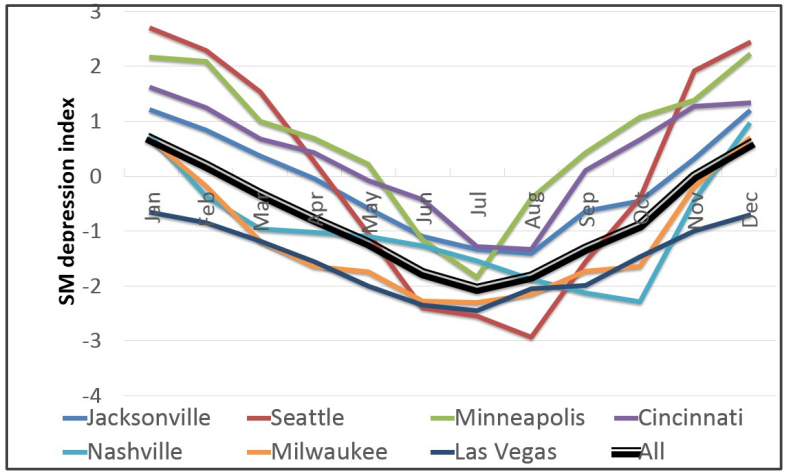
# Population Insights

(De Choudhury, Counts, Horvitz, WebSci 2013)



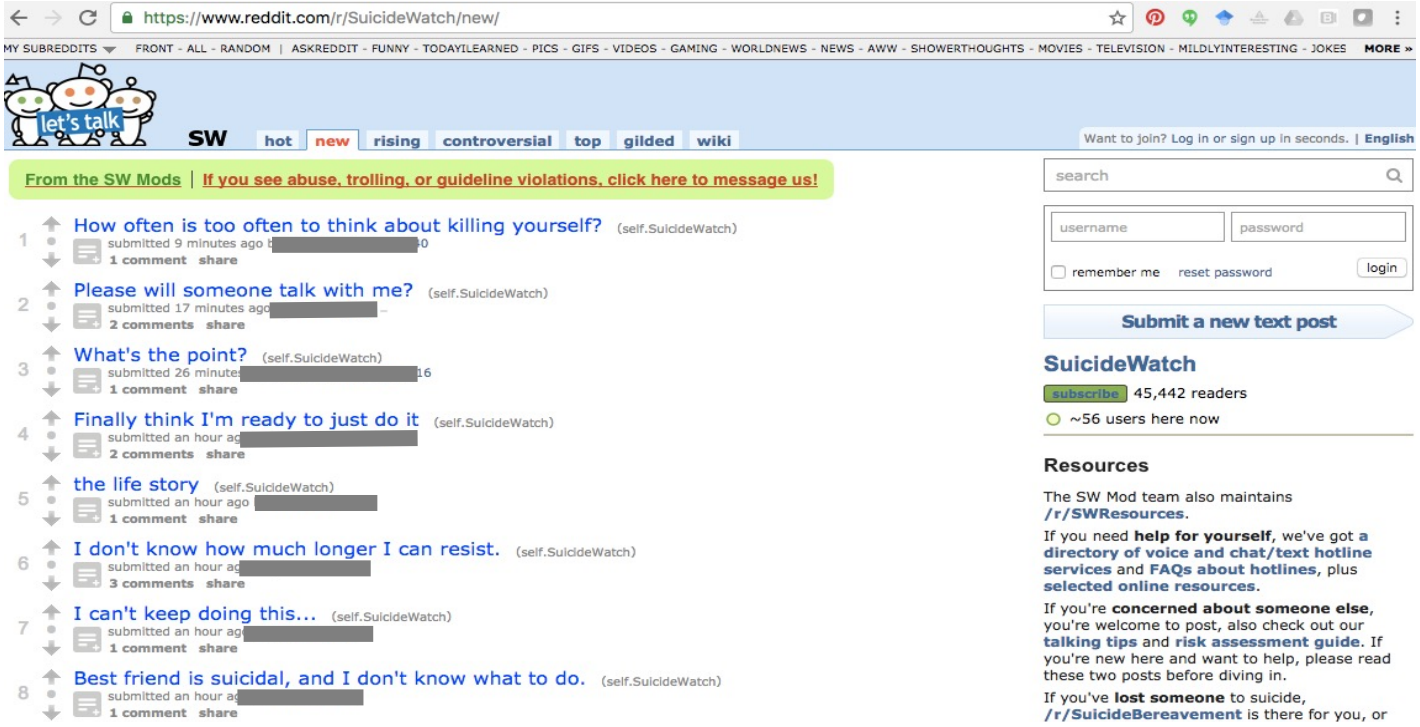



# Mental health risks as a precursor of suicidal thoughts and behaviors – where social media can help

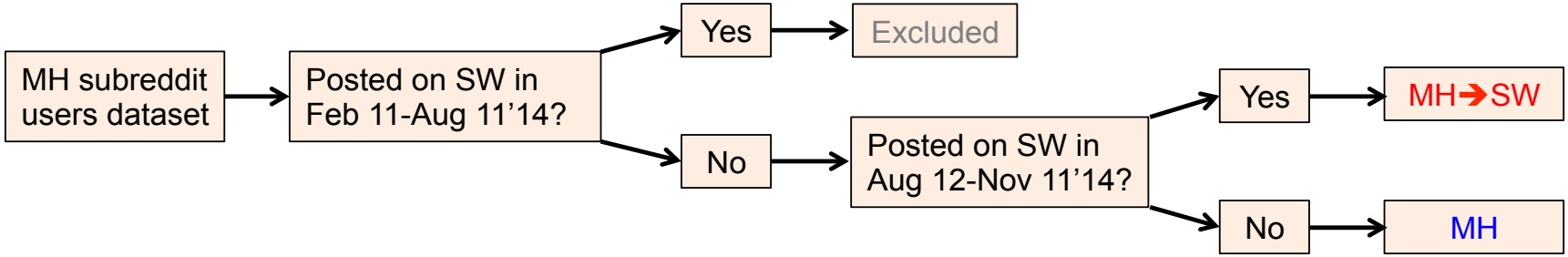
(De Choudhury, Kiciman, Dredze, Coppersmith, Kumar CHI 2016)

# Media Health Discourse in Social Media Predicts Shifts to Suicidal Ideation



(De Choudhury, Kiciman, Dredze, Coppersmith, Kumar CHI 2016)

# Propensity Score Matching




- Work within the potential outcomes framework for causal modeling, applying stratified propensity score matching to estimate causal effects

(De Choudhury, Kiciman, Dredze, Coppersmith, Kumar CHI 2016)

# Patterns of Mental Health Discourse on Social Media Predict Future Suicidal Ideation

| | |
|---|---|
| depression | +30% |
| useless | +51% |
| suicidal | +34% |
| medicine | +52% |
| locked | +51% |
| no friends | +51% |
| loneliness | +50% |
| alone I | +34% |

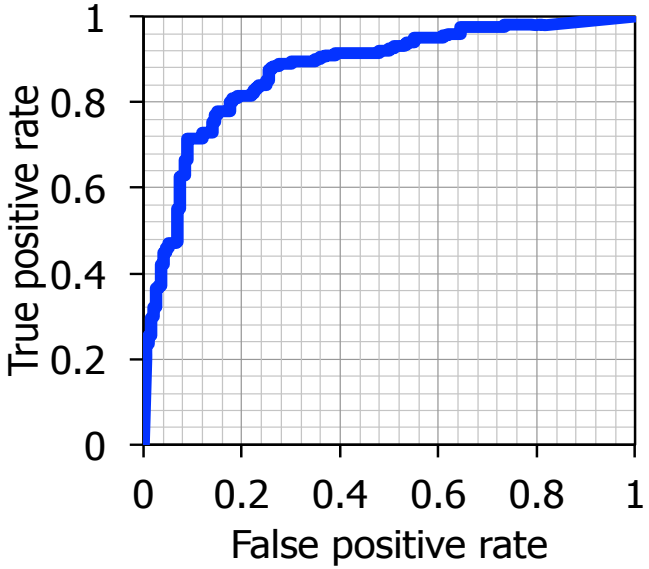
| | |
|---|---|
| differences | -57% |
| always a | -56% |
| be working | -56% |
| keep your | -56% |
| preferred | -56% |
| awesome I | -56% |
| and enjoy | -55% |
| it work | -55% |

(De Choudhury, Kiciman, Dredze, Coppersmith, Kumar CHI 2016)

# Prediction

| Model | Deviance | df | $\chi^2$ | $p$-value |
|---|---|---|---|---|
| Null | 9190.6 | 0 | | |
| Linguistic Structure | 5083.7 | 5 | 4106.9 | $< 10^{-6}$ |
| Interpers. Awareness | 7949.6 | 4 | 1241 | $< 10^{-9}$ |
| Interaction | 4429.2 | 8 | 4761.4 | $< 10^{-6}$ |
| Content | 2793.5 | 15000 | 6397.1 | $< 10^{-10}$ |
| Full | 1864.4 | 15017 | 7326.2 | $< 10^{-10}$ |

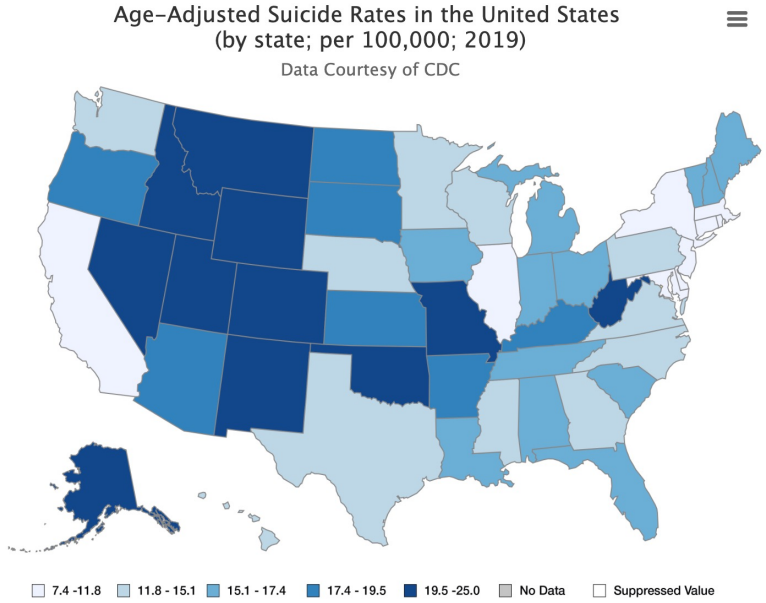
| Actual/Predicted | Class 0 | Class 1 | Total |
|---|---|---|---|
| Class 0 | 73 | 15 | 88 |
| Class 1 | 20 | 68 | 88 |
| Accuracy | 83.5% | 77.5% | 80% (mean) |
| Precision | .79 | .82 | .81 (mean) |
| Recall | .83 | .78 | .81 (mean) |
| F-1 | .81 | .8 | .80 (mean) |

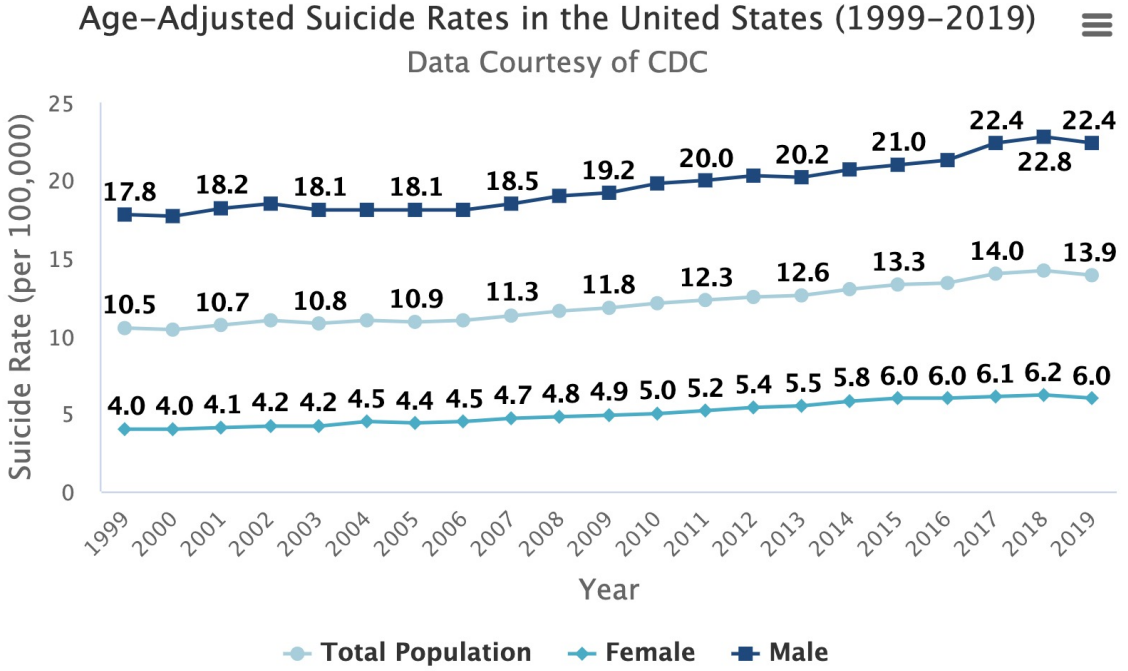


# Envisioning social media as part of a “data ecology” in suicidal risk prediction

(Choi, Sumner, Holland, Zwald, Bowen, Wang, Law, Taylor, Konjeti, De Choudhury; JAMA Network Open)

# Can We Accurately Forecast Nationwide Suicide Rates?



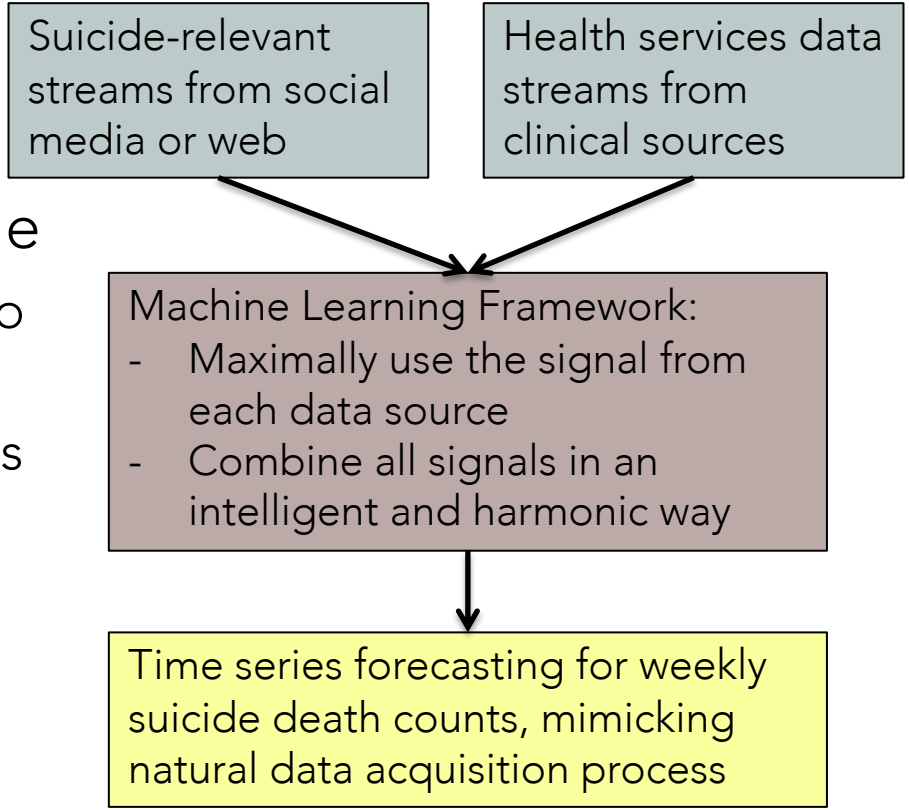




Partly supported by a contract from the CDC

(Choi, Sumner, Holland, Zwald, Bowen, Wang, Law, Taylor, Konjeti, De Choudhury; JAMA Network Open)

# Combining Real-Time Datasets

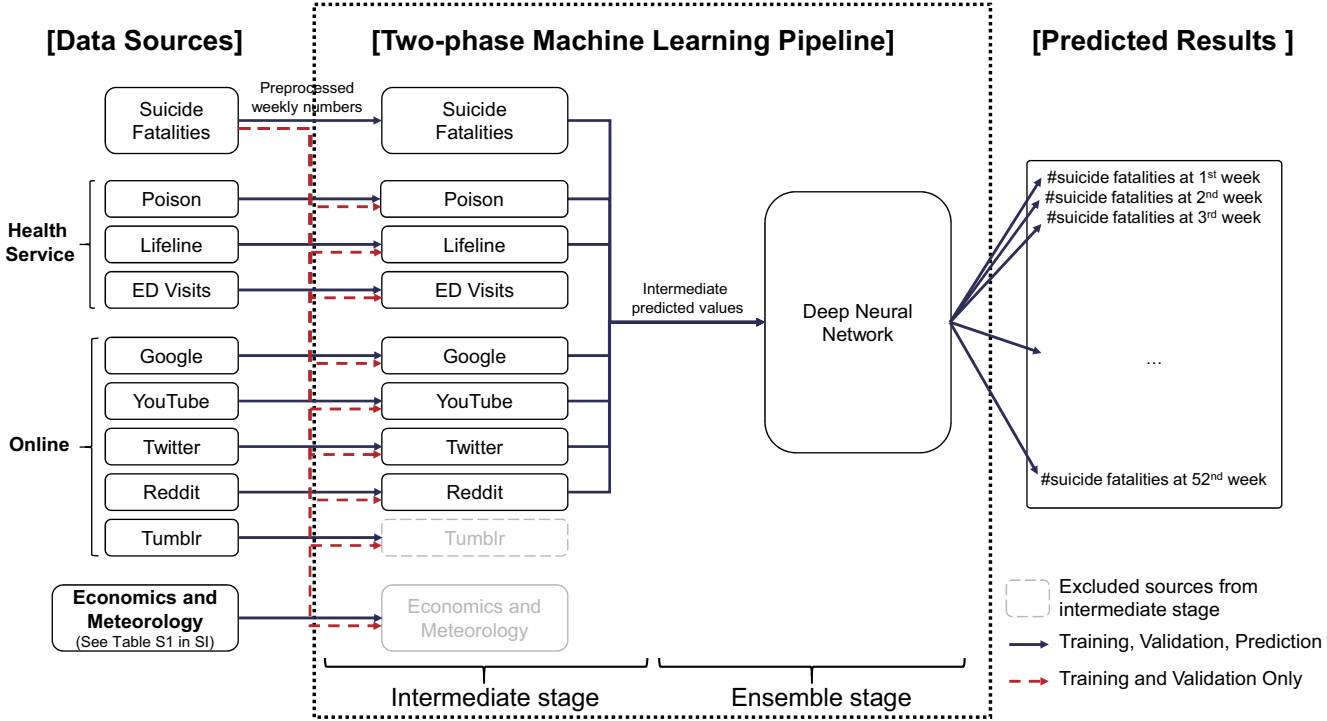
- Goal: Building a machine learning framework to predict suicide death counts in real-time
  - Time series forecasting problem to predict weekly number of suicide
  - Using multiple time series datasets collected and pre-processed from both social media and clinical sources




Partly supported by a contract from the CDC

(Choi, Sumner, Holland, Zwald, Bowen, Wang, Law, Taylor, Konjeti, De Choudhury; JAMA Network Open)

# Overall Architecture




Partly supported by a contract from the CDC

# Prediction Results (Ensemble Model)

Ground truth: 14.47

| Ensemble Type | Pearson Coeff. | RMSE | MAPE | SMAPE | Annual Estimated Rate (Error %) |
|---|---|---|---|---|---|
| Health Services Data Sources | 0.802 | 56.847 | 5.457 | 2.629 | 15.08 (4.14%) |
| Online Data Sources | 0.633 | 52.035 | 4.189 | 2.149 | 14.12 (2.49%) |
| Baseline + Health Services Sources | 0.832 | 46.096 | 4.367 | 2.123 | 14.91 (2.97%) |
| Baseline + Online Data Sources | 0.737 | 80.478 | 7.306 | 3.841 | 13.37 (7.67%) |
| Health Services + Online Data Sources | 0.791 | 43.239 | 3.806 | 1.916 | 14.31 (1.17%) |
| All Data Sources | 0.811 | 44.439 | 4.006 | 2.001 | 14.40 (0.55%) |

Partly supported by a contract from the CDC

(Choi, Sumner, Holland, Zwald, Bowen, Wang, Law, Taylor, Konjeti, De Choudhury; JAMA Network Open)
(Saha, Torous, Caine, De Choudhury, JMIR)

# Takeaways

SELF-CARE

The New York Times

## Is Everybody Doing … OK? Let's Ask Social Media

Researchers are looking at online behavior to gauge public mental health. The results aren't pretty.

Partly supported by a contract from the CDC; NSF RAPID 2027689

# The path forward…

# Thanks!

munmund@gatech.edu
http://www.munmund.net
@munmun10

and many more…