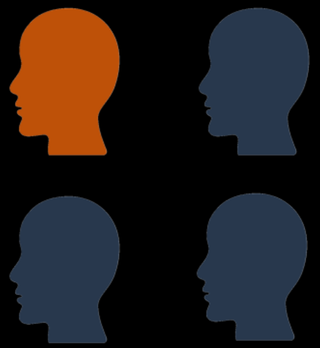
Artificial Intelligence and Mental Health

Dr. Ehi Nosakhare, Senior Data and Applied Science Manager at Microsoft AI Development Acceleration Program (MAIDAP)

5/12/2021

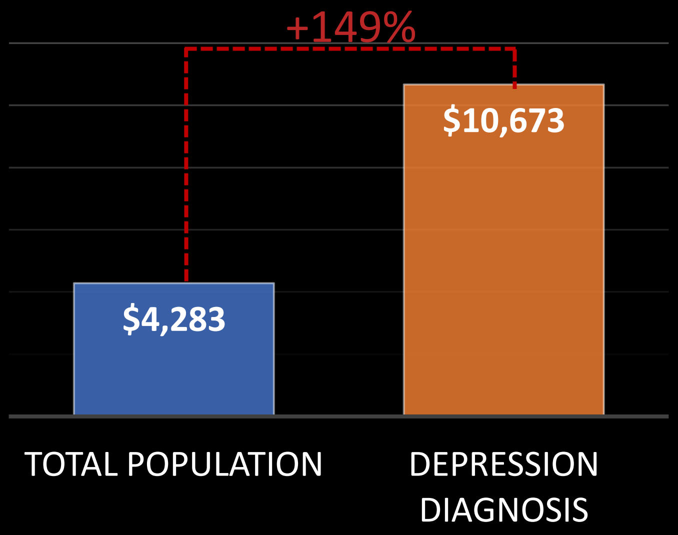
Growing Challenge

1 in 4 people in the world will be affected by a mental disorder at some point in their lives[1]



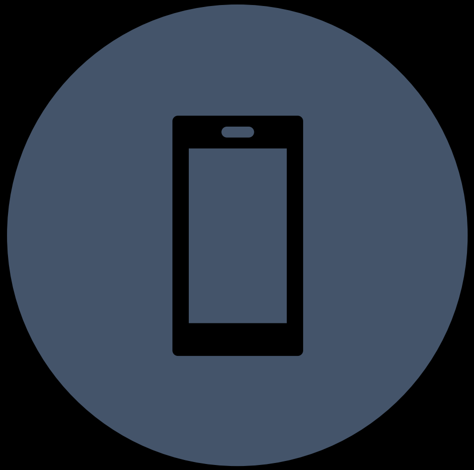
264 million people are affected by depression worldwide. More women are affected than men.[2] **People of color are disproportionately affected.**

US average healthcare cost per person per year (2016)[3]



Source:
1: World Health Organization, http://www.who.int/whr/2001/media_centre/press_release/en/. Accessed Aug. 10, 2018
2: World Health Organization, http://www.who.int/news-room/fact-sheets/detail/depression. Accessed Oct. 10, 2020
3: Blue Cross Blue Shield. Major Depression: The Impact on Overall Health. *2018 Health of America Report.* May 10, 2018

In addition to data collected in clinical settings, personal health-related mobile applications, fitness trackers and other sensors have made it feasible to collect data to better characterize environmental and lifestyle factors driving mental health outcomes

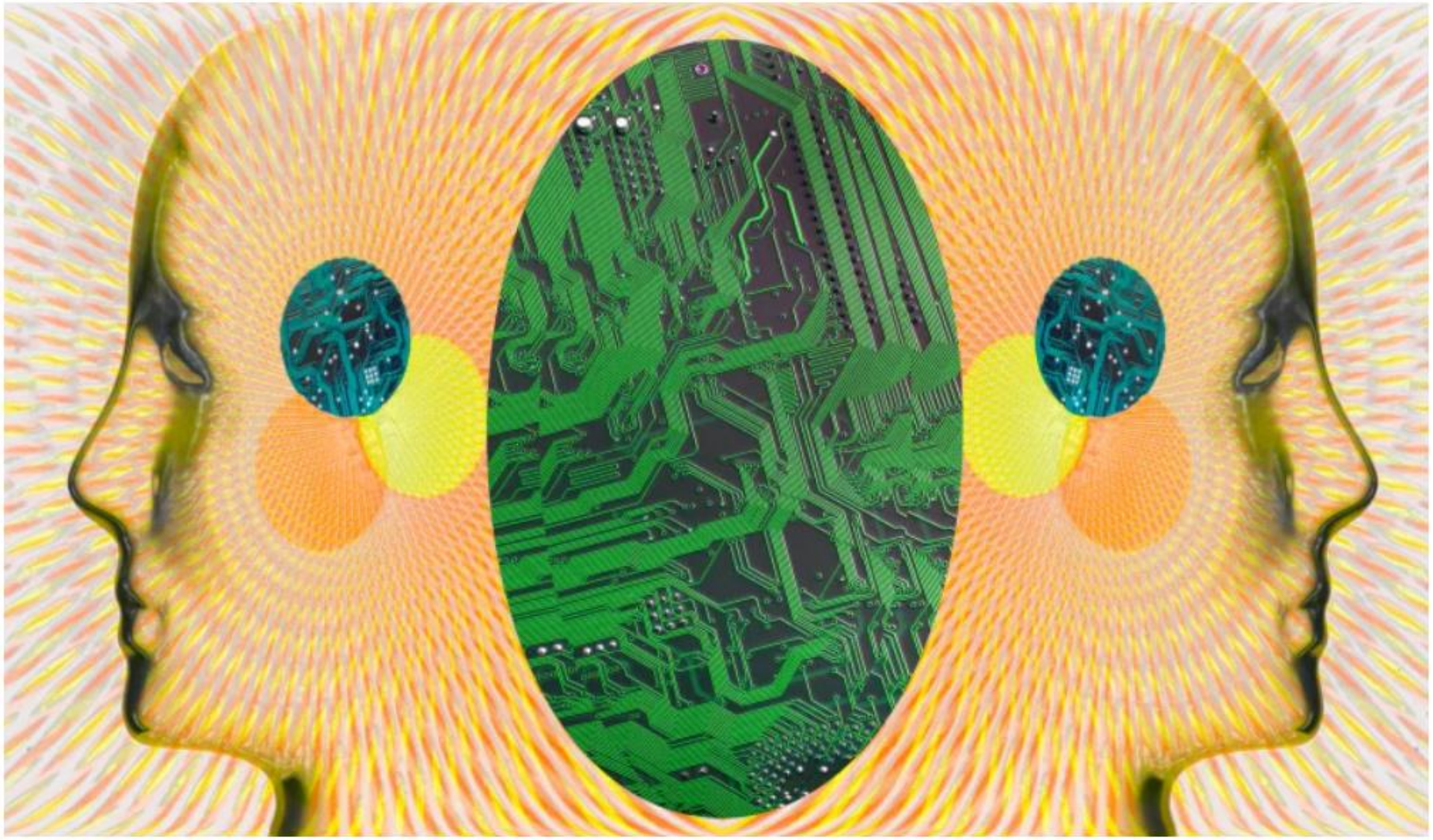


These devices can collect real-time information with the average user likely to produce over one million gigabytes of health-related data in his or her lifetime—the equivalent of about 300 million books[1]

[1]https://www-03.ibm.com/press/uk/en/pressrelease/46609.wss/ Accessed Dec. 15, 2018

Artificial Intelligence Could Help Solve America's Impending Mental Health Crisis

There is an increase in the exploration of how artificial intelligence can assist in the detection, diagnosis and treatment of mental health issues



Getty Images

Classes of Digital Health Technology

01

Behavioral intervention technologies

02

Mobile apps and wearable devices for symptom monitoring and health risk assessments

03

Computerized treatments

04

Platforms for peer or community support

Classes of Digital Health Technology

01

Behavioral intervention technologies

02

Mobile apps and wearable devices for symptom monitoring and health risk assessments

03

Computerized treatments

04

Platforms for peer or community support

The ability to model and predict self-reported stress, happiness, and health could be beneficial for an individual and in the treatment and **prevention** of mental illness

Self-reported health strongly relates to actual health and all-cause mortality

Stress increases susceptibility to infection and illness

Self-reported happiness is indicative of clinical depression and has as string an effect on longevity ad cigarette smoking

Our Approach

- Use multi-modal real-world data gathered from college students
- Identify objective correlates of self-reported mood, stress and health
- Bayesian Modeling: Employ interpretable models to
  - Predict future mood, stress and health
  - Study the influence of combinations of behaviors on well-being
  - Leverage insights from models to provide relevant personalized recommendations to individuals looking to improve well-being

Predict future mood, stress and health

Investigate the influence of combinations of behaviors on well-being

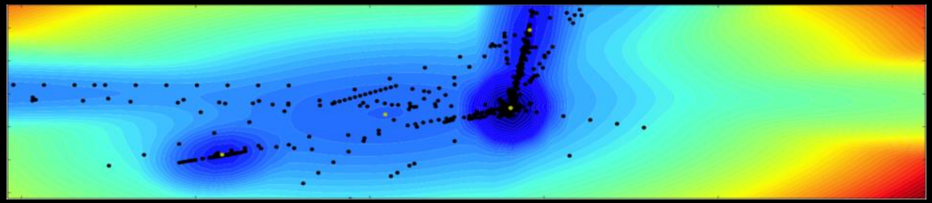
Predict future mood, stress and health

Investigate the influence of combinations of behaviors on well-being

with thanks to Sara Taylor, Natasha Jaques, Akane Sano, Rosalind Picard ML4H @ NeuRIPS 2016, IEEE TFAC 2017

Data Overview

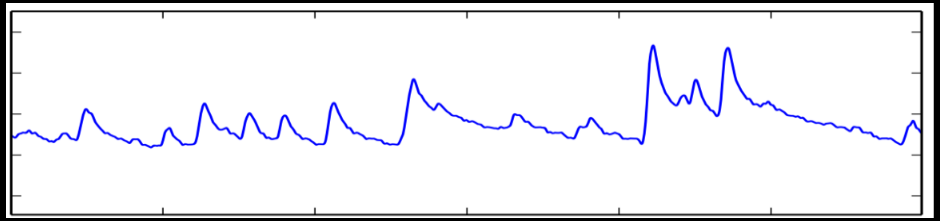
Daily features from 104 users, 1842 days of data



Location patterns modeled with a GMM

Smartphone logs
(call, sms, screen)

Weather



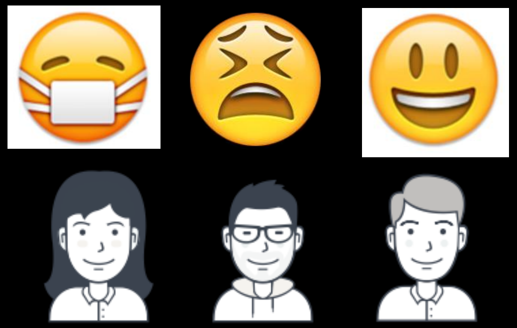
Physiology: Accelerometer, skin temperature, Electrodermal Activity (EDA)

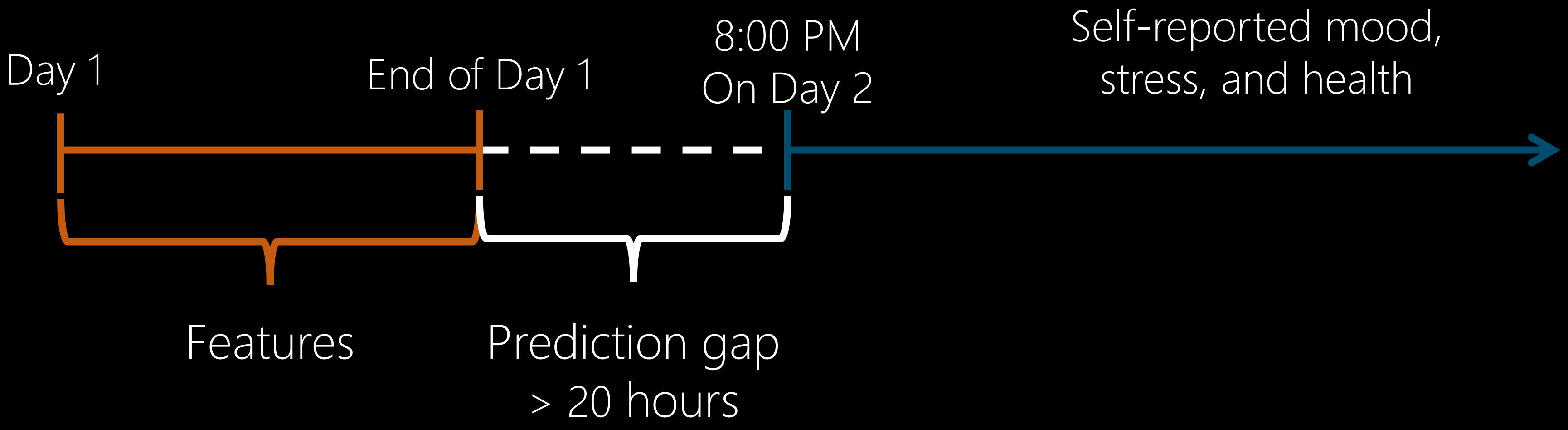
Behavioral surveys
Including self-reported stress, health and mood on a scale (0 - 100)

Predictive Modeling: Multi-Task Learning

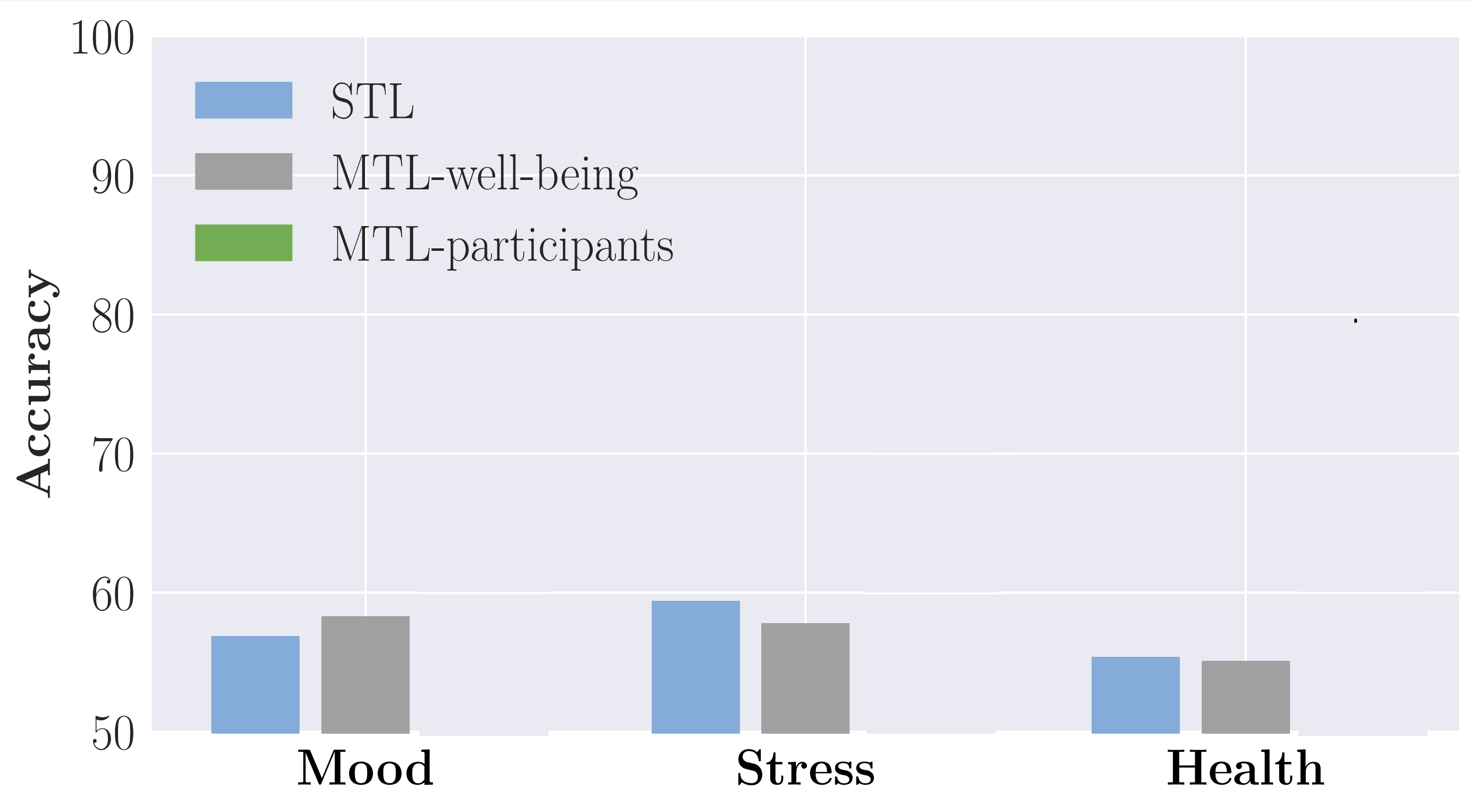
Two approaches: Well-being-as-tasks

Users-as-tasks

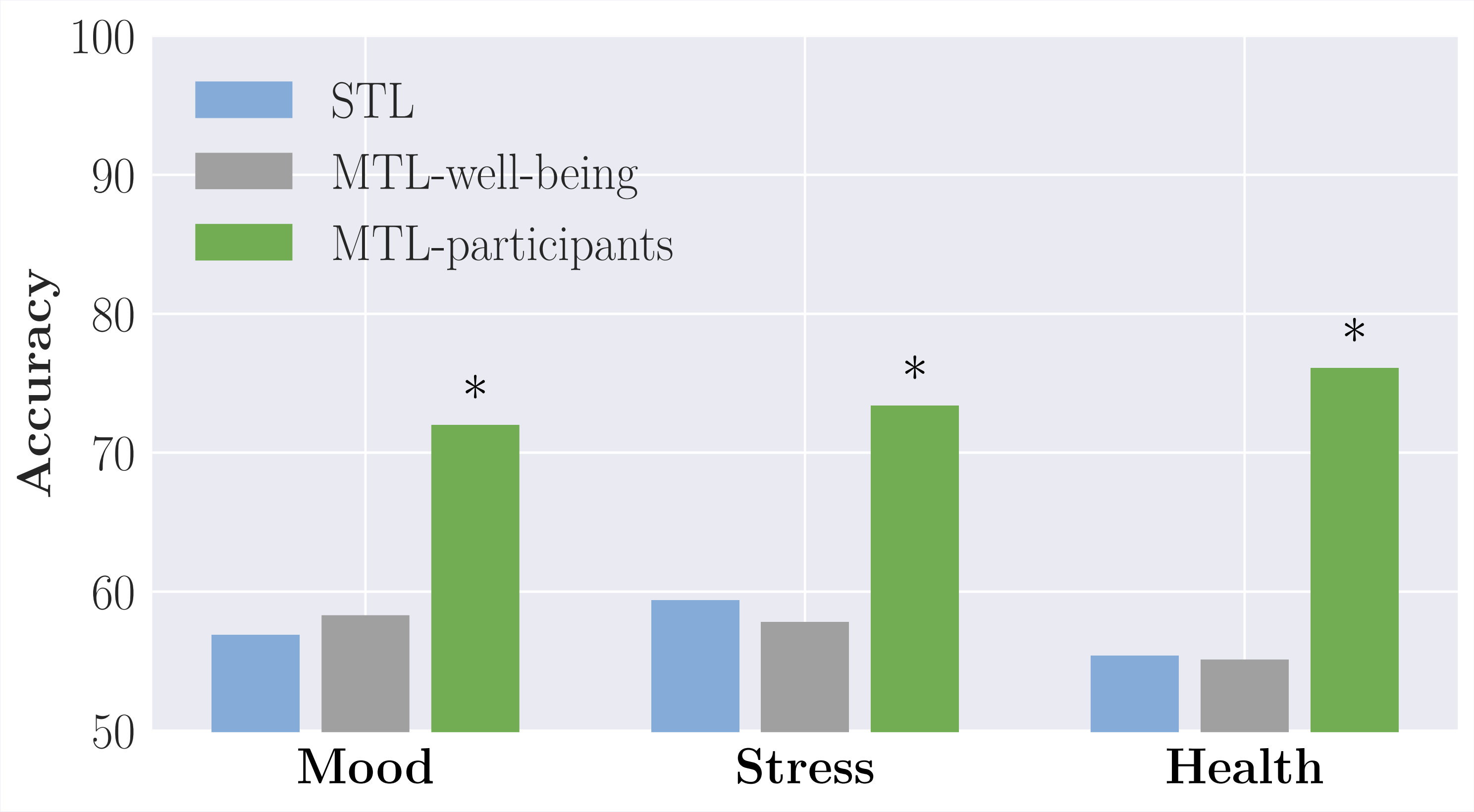




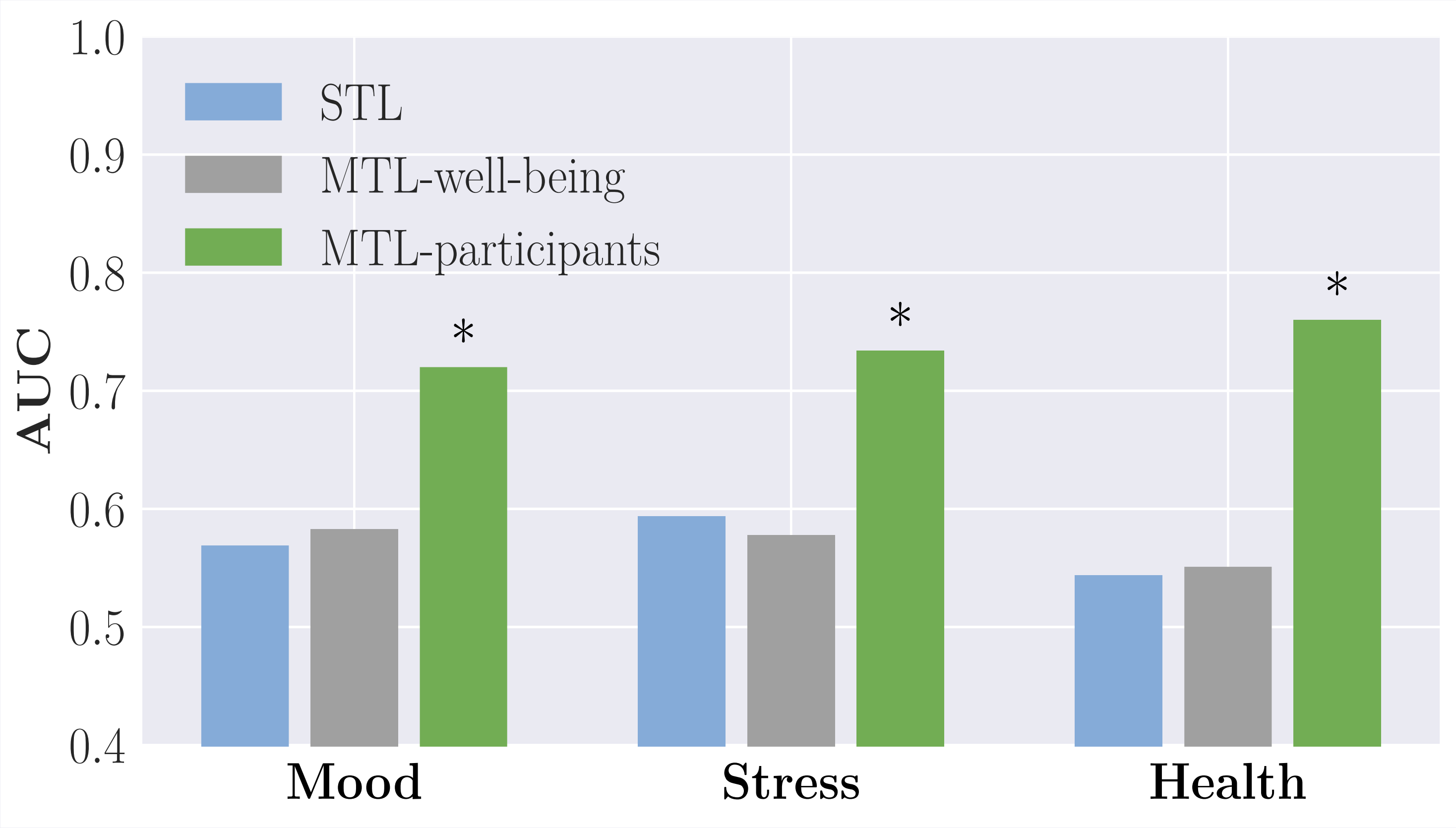
Personalized MTL approach significantly outperforms other approaches

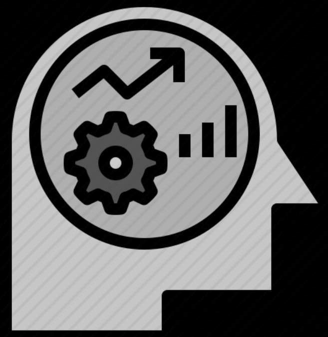


Personalized MTL approach significantly outperforms other approaches



Personalized MTL approach significantly outperforms other approaches





Predict future mood, stress and health

Investigate the influence of combinations of behaviors on well-being

with thanks to Rosalind Picard. KDD 2019, ACM Health 2020

Why is this relevant?

Studies have shown the effect of daily health behaviors on well-being *(Stubbe 2007, Wang 2014)*

Learning prediction rules is not enough. What group of behaviors influenced the outcome?
(Doshi-Velez and Kim 2017)

Understanding can help provide actionable insights and recommendations

Data Overview: Modifiable behaviors

Smartphone logs
(call, sms, screen)

Location patterns

Behavioral surveys
Including self-reported stress (0 - 100)

Wearable sensors
for sleep monitoring

Bag-of-behaviors

Example of behaviors in a participant's day

- 5 outgoing calls
- 2 mins. of incoming calls
- No SMS
- Studied for an hour
- Classes for 2 hours
- 7.5 hours of sleep the night before
- Bedtime at 12:23AM
- No bedtime deviation
- Worked out for an hour
- Extracurricular activities for an hour

Outgoing calls: 2 - 3
Outgoing calls: 4 - 5
Outgoing SMS: 0
Sleep duration: 6 - 7
Yesterday sleep duration: 7 - 8
Bedtime: 0 - 1

$\{\ldots\ 0\quad 1\ \ldots\ 1\ \ldots\ 0\quad 1\ \ldots\ 1\ \ldots\}$

Bag-of-behaviors

Example of behaviors in a participant's day

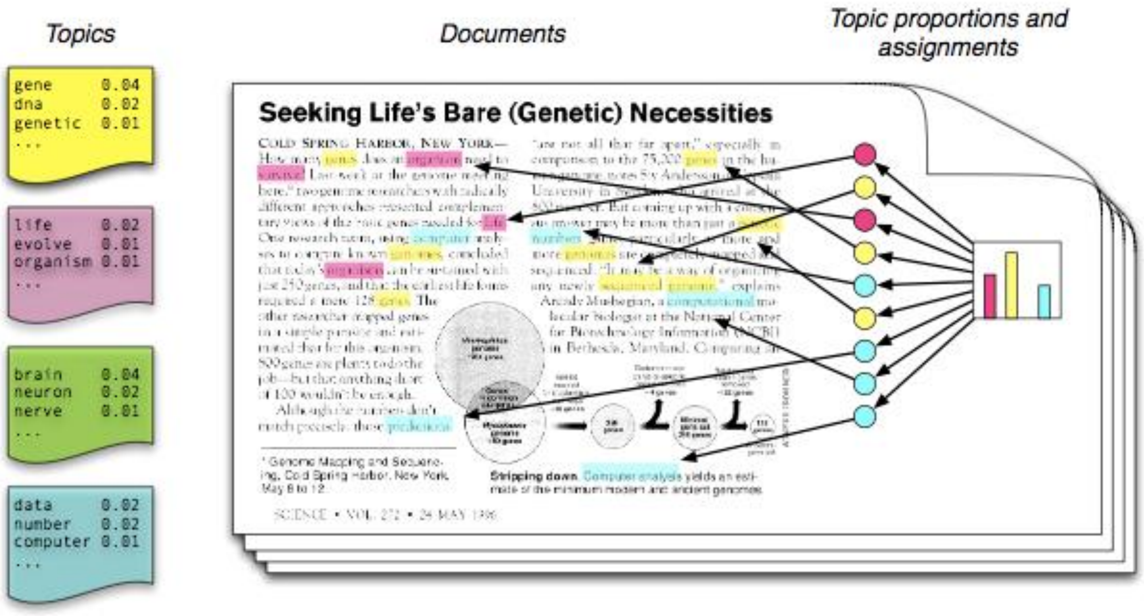
- 5 outgoing calls
- 2 mins. of incoming calls
- No SMS
- Studied for an hour
- Classes for 2 hours
- 7.5 hours of sleep the night before
- Bedtime at 12:23AM
- No bedtime deviation
- Worked out for an hour
- Extracurricular activities for an hour

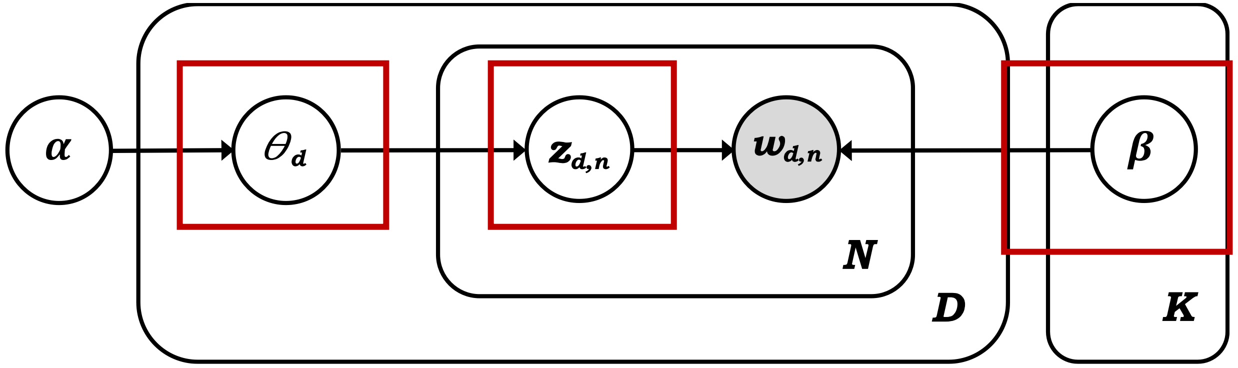
134 Behaviors
5397 days of data
115,659 observations
from 224 unique participants

Latent Dirichlet Allocation (LDA)

Given documents and model parameters, LDA learns:

- latent topics that are distributions of words in a documents
- personalized topic proportions for each the documents
- topic assignments for each word





(Blei 2003)

Supervised Latent Dirichlet Allocation

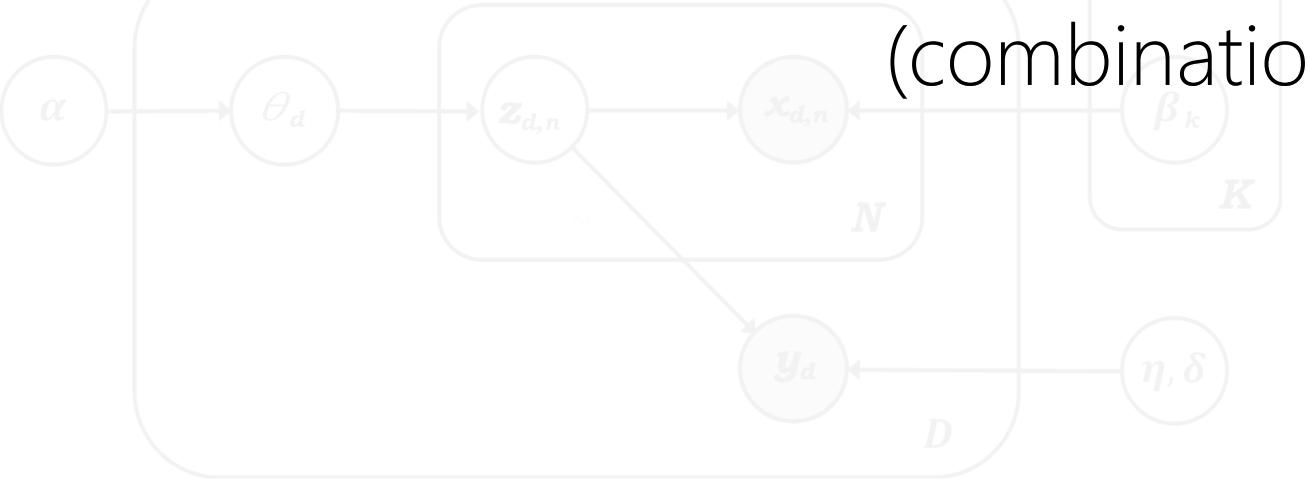
But sLDA supervises how each word is assigned to a topic by jointly modeling y as follows:

$$y \mid z, \eta, \delta \sim \mathcal{N}(\eta^T \bar{z}, \delta)$$

words ⟶ modifiable behaviors

response y ⟶ self-reported stress

latent topics ⟶ latent patterns (combinations of behaviors)



(Blei 2010)

Three Experiments

1. Show that the latent patterns are indeed predictive of self-reported stress when compared to patterns learned in an unsupervised way

2. Show that the sLDA model outperforms the LASSO when predicting self-reported stress

3. Show that the sLDA model uncovered meaningful latent patterns (or combinations) of health behavior

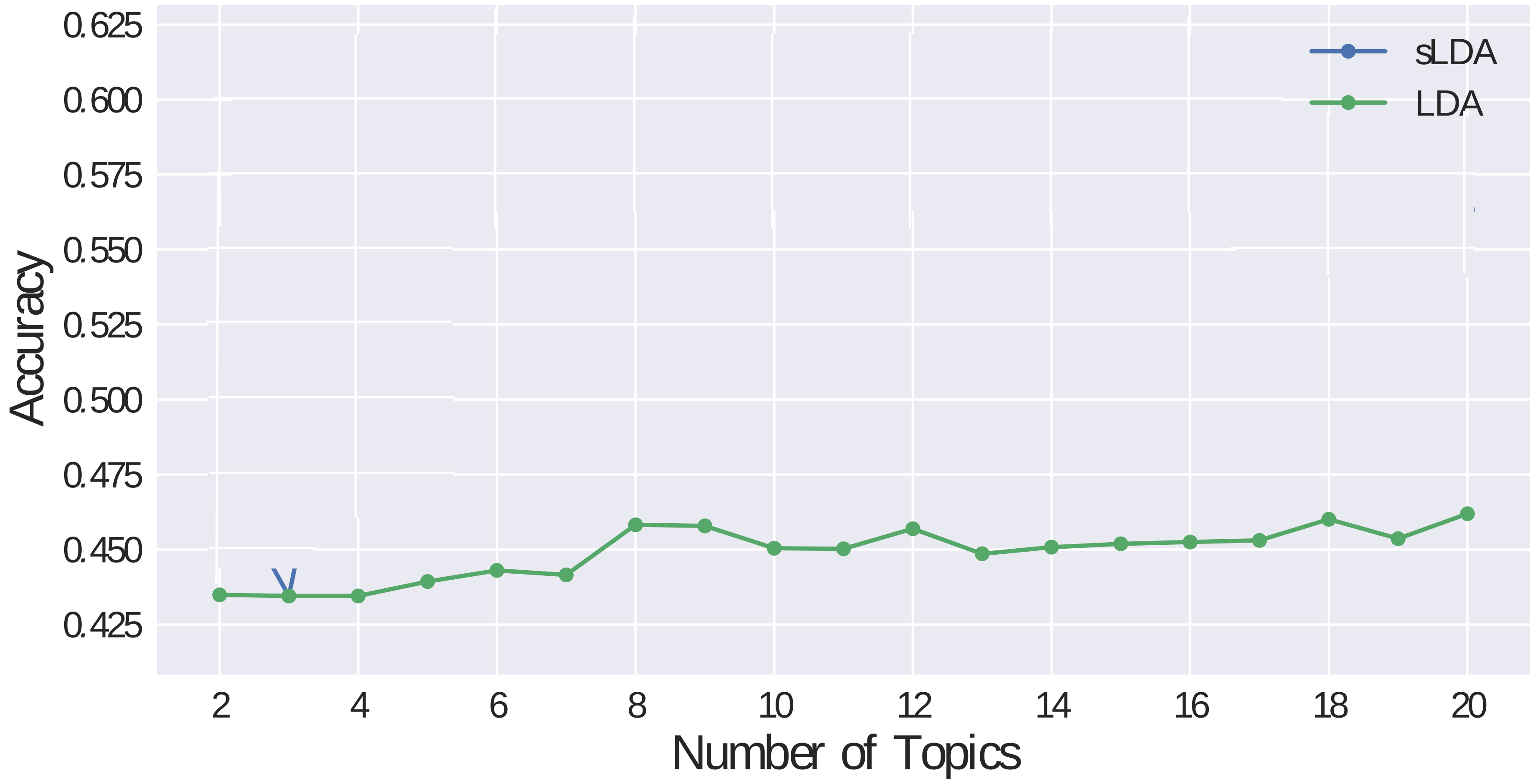
Three Experiments

1. Show that the latent patterns are indeed predictive of self-reported stress when compared to patterns learned in an unsupervised way

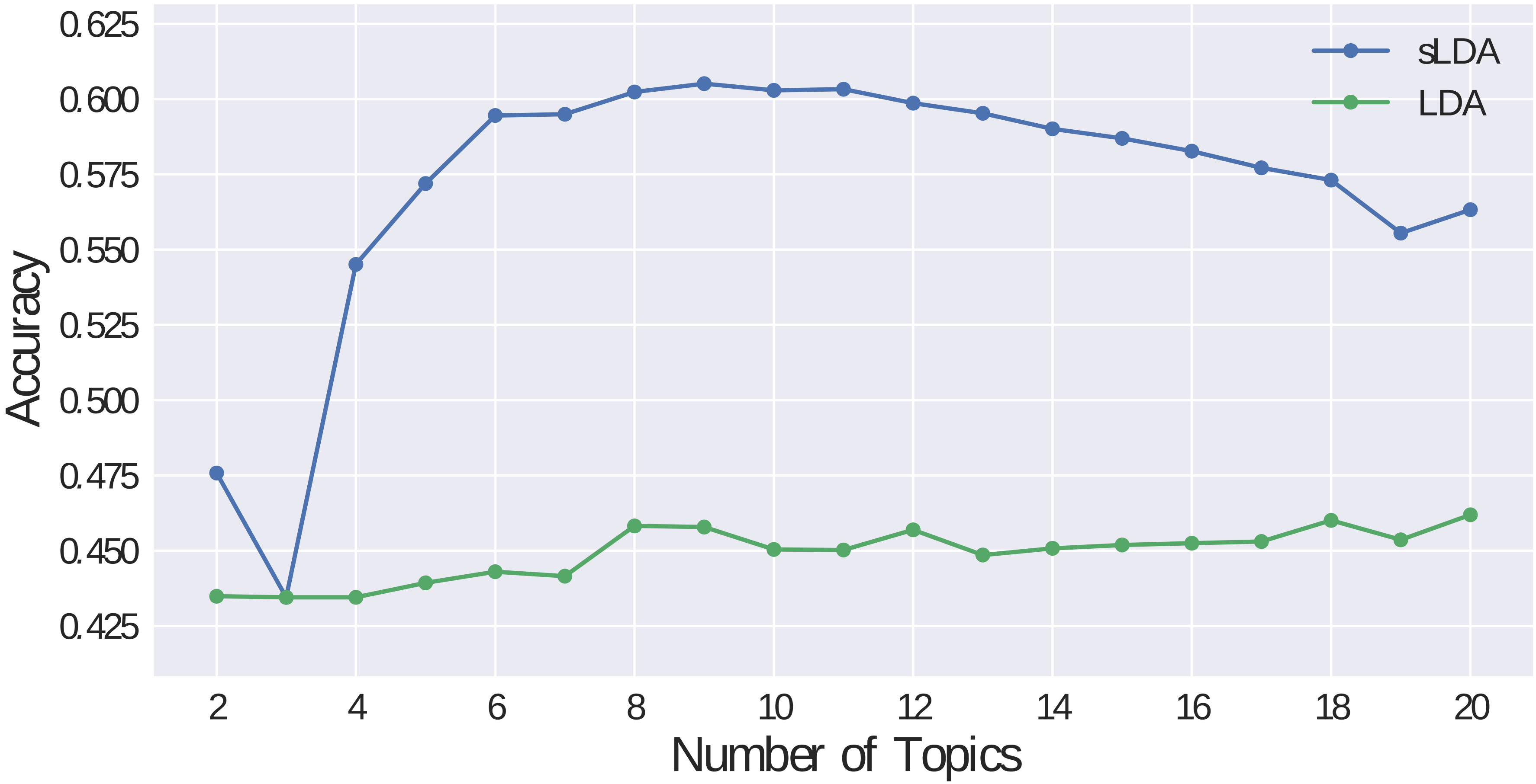
2. Show that the sLDA model outperforms the LASSO when predicting self-reported stress

3. Show that the sLDA model uncovered meaningful latent patterns (or combinations) of health behavior

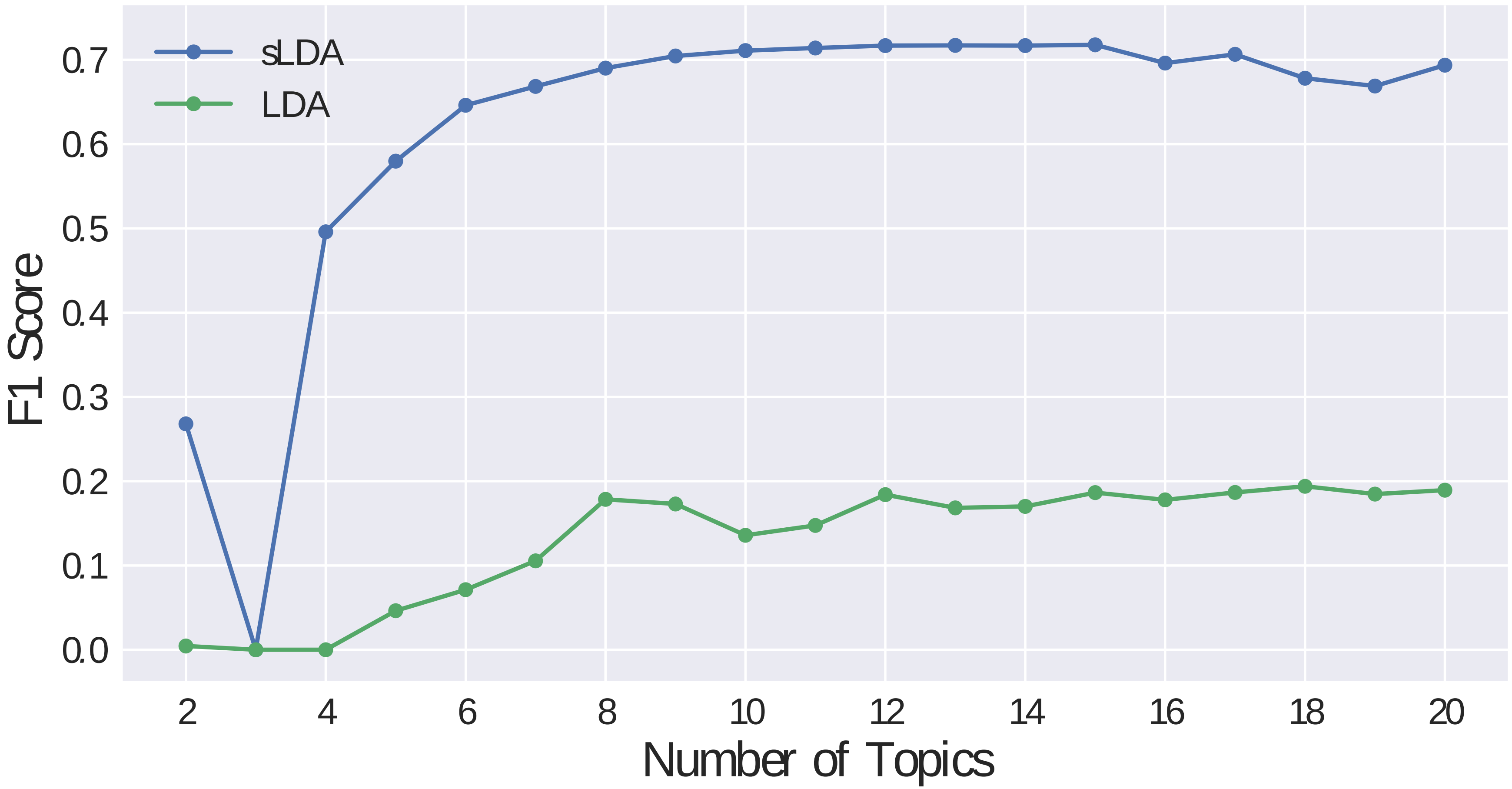
Supervised latent patterns significantly outperforms unsupervised patterns



Supervised latent patterns significantly outperforms unsupervised patterns



Supervised latent patterns significantly outperforms unsupervised patterns

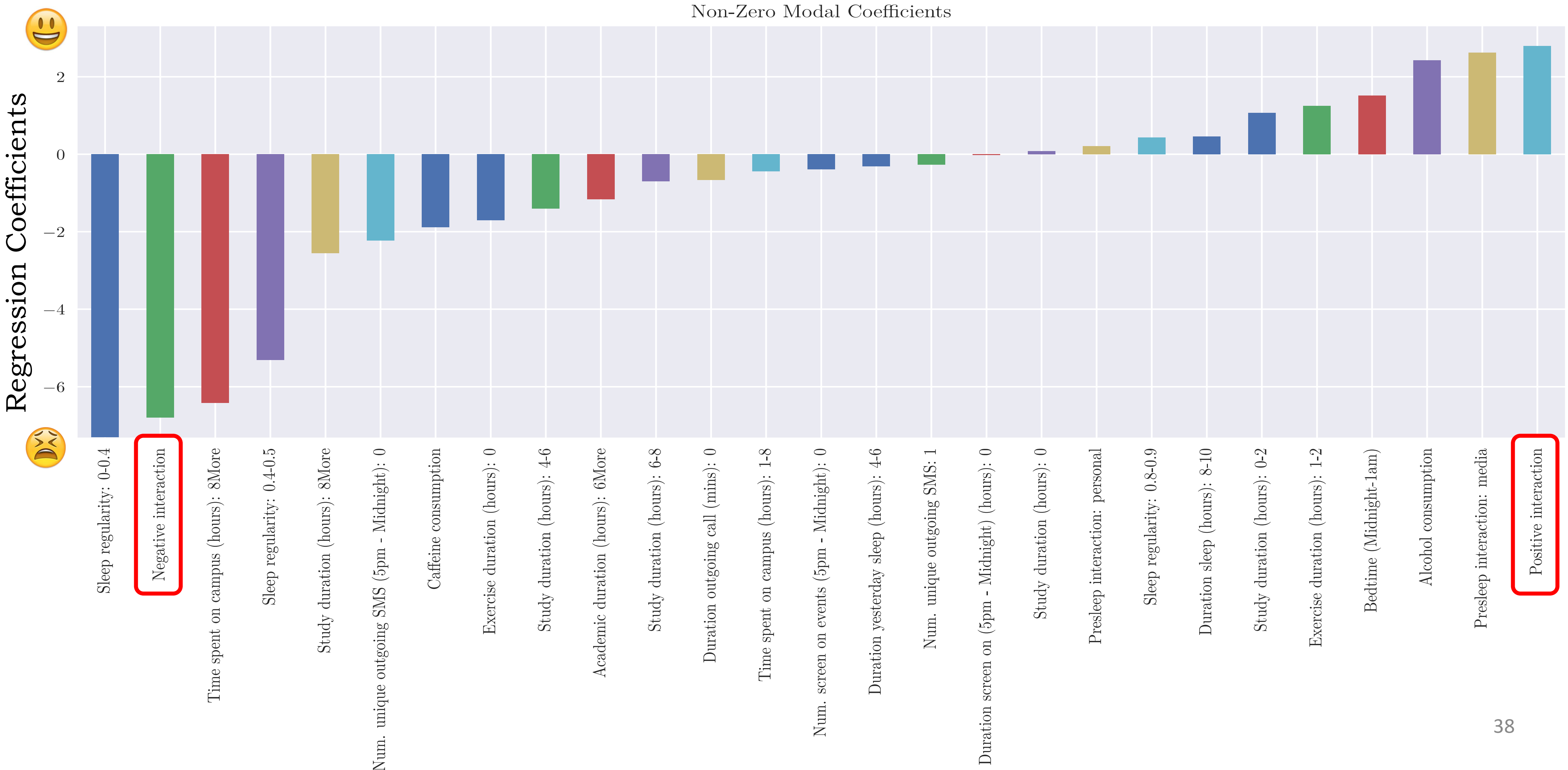


sLDA has better prediction performance compared to LASSO

	Models	
Perfomance Metrics	LASSO	sLDA
Binary Accuracy (thresh. = 50)	56.5% (± 1.0)	**60.5% (± 0.4)**
F1 Score (thresh. = 50)	0.48 (± 0.02)	**0.72 (± 0.01)**

*Bold entries: $p < 0.05$

LASSO Coefficients



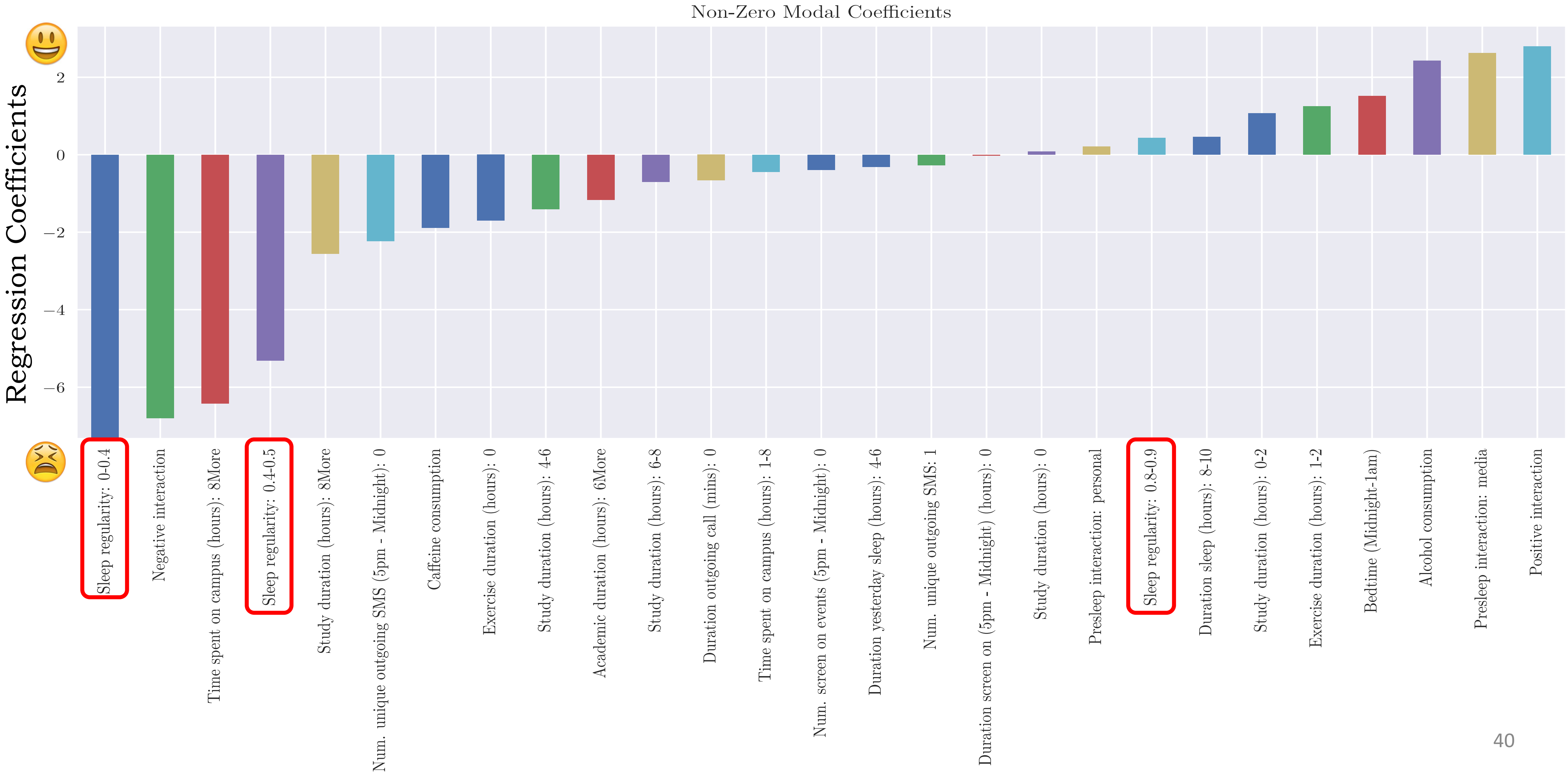
Three Experiments

1. Show that the latent patterns are indeed predictive of self-reported stress when compared to patterns learned in an unsupervised way

2. Show that the sLDA model outperforms the LASSO when predicting self-reported stress

3. Show that the sLDA model uncovered meaningful latent patterns (or combinations) of health behavior

LASSO Coefficients



Three Experiments

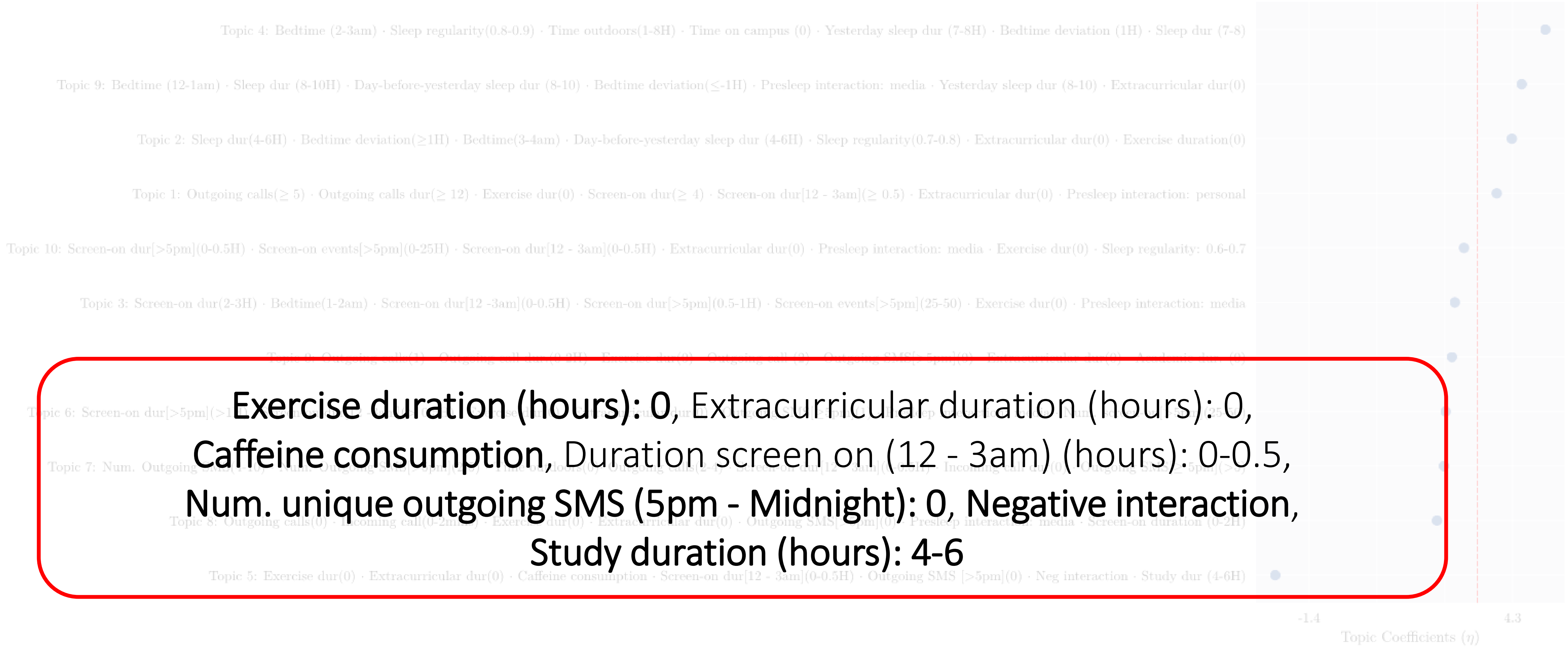
1. Show that the latent patterns are indeed predictive of self-reported stress when compared to patterns learned in an unsupervised way

2. Show that the sLDA model outperforms the LASSO when predicting self-reported stress

3. Show that the sLDA model uncovered meaningful latent patterns (or combinations) of health behavior

Output: 11-topic sLDA model fit to the data

Output: 11-topic sLDA model fit to the data

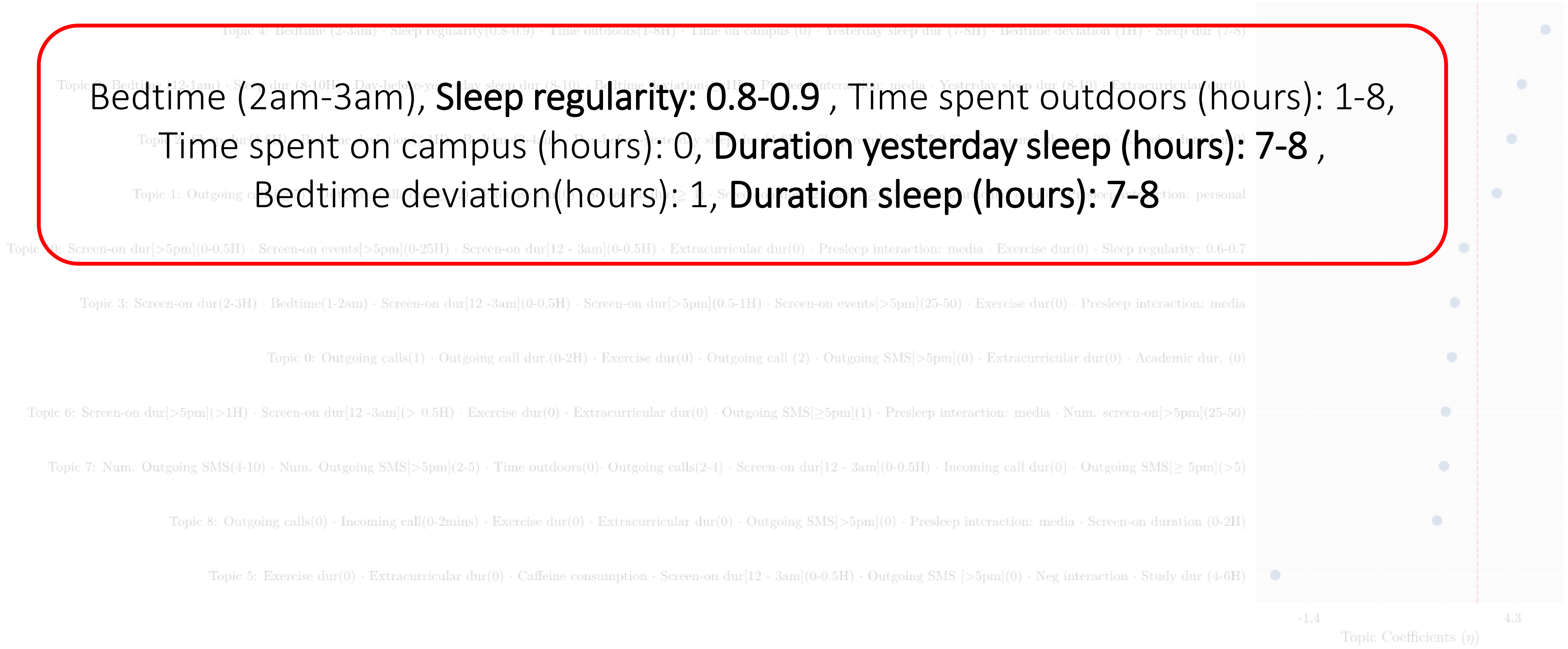


Exercise duration (hours): 0, Extracurricular duration (hours): 0, **Caffeine consumption**, Duration screen on (12 - 3am) (hours): 0-0.5, **Num. unique outgoing SMS (5pm - Midnight): 0**, **Negative interaction**, **Study duration (hours): 4-6**

Output: 11-topic sLDA model fit to the data

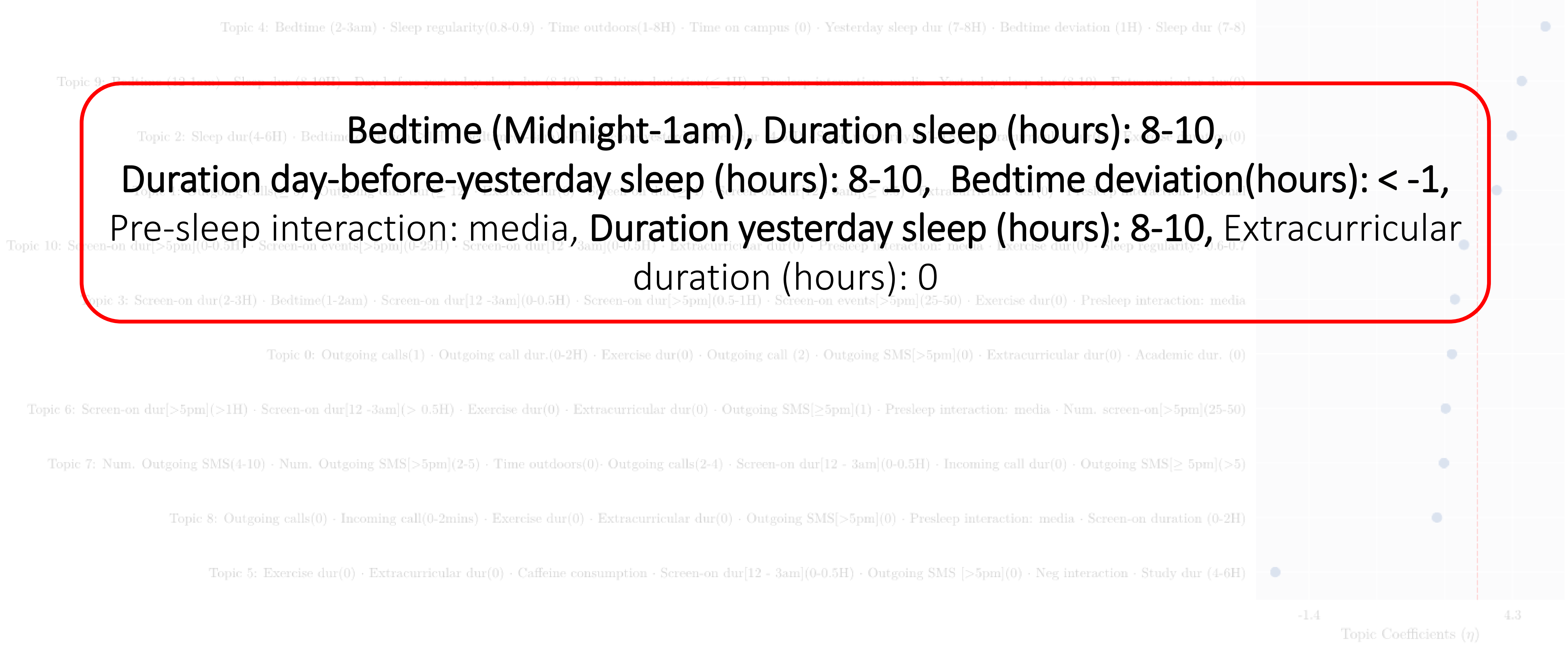
Output: 11-topic sLDA model fit to the data

Bedtime (2am-3am), **Sleep regularity: 0.8-0.9** , Time spent outdoors (hours): 1-8, Time spent on campus (hours): 0, **Duration yesterday sleep (hours): 7-8** , Bedtime deviation(hours): 1, **Duration sleep (hours): 7-8**



Output: 11-topic sLDA model fit to the data

Output: 11-topic sLDA model fit to the data

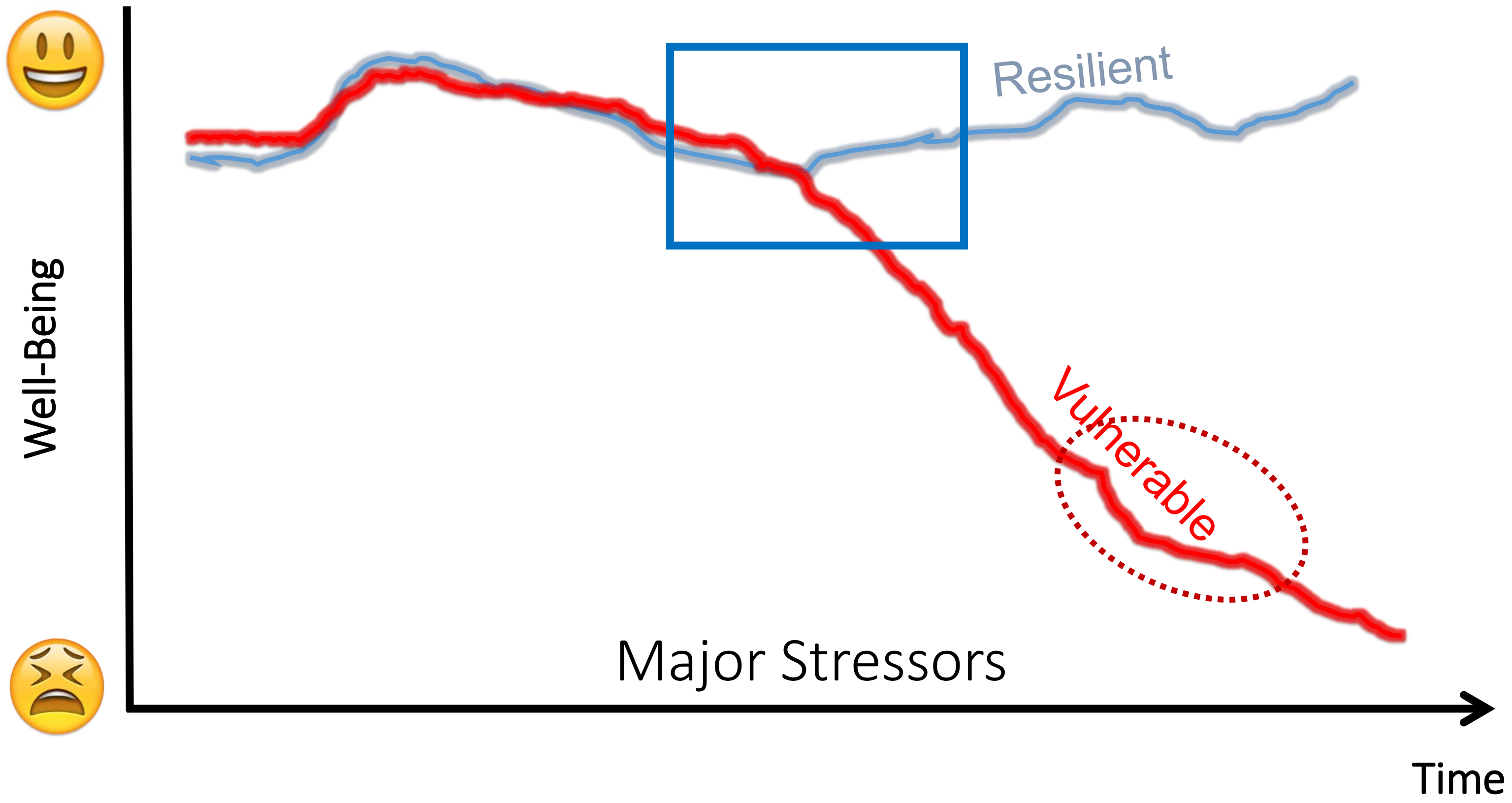


Bedtime (Midnight-1am), Duration sleep (hours): 8-10,
Duration day-before-yesterday sleep (hours): 8-10, Bedtime deviation(hours): < -1,
Pre-sleep interaction: media, **Duration yesterday sleep (hours): 8-10,** Extracurricular duration (hours): 0

Output: 11-topic sLDA model fit to the data

Output: 11-topic sLDA model fit to the data

Towards Personalized Well-Being Assessment



Research Groups working on AI and Mental Health

Affective Computing, MIT Media Lab

Computational Wellbeing group, Rice University

Human Understanding and Empathy, Microsoft Research

Healthcare Intelligence, Microsoft Research

People Aware Computing Lab, Cornell University

Student Life, Dartmouth University

Behavioral Data Science Group, University of Washington

What do I do now?

What type of problems do I solve?

- **Classification:** Will this user's PC fail? Identifying failure-prone PCs reduces employee downtime, especially during COVID-19 remote work.

- **Regression:** What size virtual machine does this task need? Choosing the right amount of memory helps Microsoft save on compute resources and offer lower prices to customers.

- **Grouping:** Are these Azure alerts from the same incident? Grouping notifications helps engineers narrow down root causes faster, making Azure more reliable.

MAIDAP Blog Post

Data Science Roles in Industry

- Machine Learning Researcher
- Data Analyst
- Machine Learning Scientist
- Machine Learning Engineer
- Machine Learning Program Manager

Questions?

@_ehinosa

/in/ehinosa