17-537 Project Proposal:

Modelling and Predicting Mental Health Issues

Lee Wei Xin (weixinl), Andey Ng (andeyn), Claudia Kho (ckho)

Pin down the societal challenge(s) (0.5 point)

Mental health issues affect one in five adults[1] - this makes it a pressing social problem, yet it is often not openly discussed. Given that full-time professionals spend on average 42.5 hours a week working[2], more research is needed to find out how employers can most effectively improve the environment at the office, so as to ensure the mental well-being of their employees. More specifically, the tech industry is a high-stress environment, due to the fast-paced nature of the field. In the United Kingdom, people working in the tech industry are 5 times more likely to suffer from depression than those who do not[2].

In this paper, we will analyze various indicators and identify those that have the most significant impact on an individual's mental wellbeing. Next, we will build a model to predict the state of an individual's mental health, which will determine whether they require intervention or help.

Pin down the AI method(s) (0.5 point)

We will employ a two-step approach in this paper. The first step involves using logistic regression to determine the most important predictors in mental health illness. This can be done by using a likelihood-ratio test, or alternatively, by the Wald statistic.

Once we have an idea of which indicators have a more significant impact on a professional's mental health, we will proceed to construct and train a multilayer perceptron network that predicts whether the individual is likely to experience a mental health issue, based on his/her indicators. We will use the dataset that we have, splitting them into a train and test set, so as to analyze the accuracy of our model.

---

[1] https://www.nami.org/learn-more/mental-health-by-the-numbers

[2] https://diginomica.com/mental-health-awareness-week-tech-industry-crisis-organizations-need-tackle

Provide a list of references, including news articles or reports that describes the societal challenge, relevant papers, and data source if data-centric (0.5 points)

Data Sources:

- **https://osmihelp.org/research** (planning on using 2019 data, and possibly earlier years too if needed)
- Kaggle links:
    - 2019 - Osmi. “OSMI Mental Health In Tech Survey 2019.” *Kaggle*, 2 Jan. 2020, https://www.kaggle.com/osmihelp/osmi-mental-health-in-tech-survey-2019/data
    - 2018 - https://www.kaggle.com/osmihelp/osmi-mental-health-in-tech-survey-2018
    - 2017 - https://www.kaggle.com/osmihelp/osmi-mental-health-in-tech-survey-2017

Articles:

- https://diginomica.com/mental-health-awareness-week-tech-industry-crisis-organizations-need-tackle
- https://www.forbes.com/sites/forbestechcouncil/2018/08/08/getting-honest-about-mental-health-in-the-world-of-tech-startups/#522fe212641a
- Everett, Cath. “Mental Health Awareness Week - the Tech Industry Crisis That Organizations Need to Tackle.” Diginomica, Diginomica, 26 June 2019, diginomica.com/mental-health-awareness-week-tech-industry-crisis-organizations-need-tackle.
- Snobar, Abdullah. “Getting Honest About Mental Health In The World Of Tech Startups.” Forbes, 8 August 2018, https://www.forbes.com/sites/forbestechcouncil/2018/08/08/getting-honest-about-mental-health-in-the-world-of-tech-startups/#1803b4d8641a

Describe 2~3 envisioned milestones of the proposed project, i.e., the important checkpoints that demonstrate the progress of the project (1 point)

1. Perform logistic regression to figure out which indicators are more significant.
2. Train a multilayer perceptron to predict whether an individual is currently experiencing a mental health issue, and will require intervention.
3. Write a report based on our findings, as well as prepare a poster presentation.

Describe the tentative plan of action, including the steps and the expected time needed for each step (1 point)

Action	Expected Time	Comments
Clean and preprocess data	1 week	
Explore suitability of models	1 week	Eg. logistic regression assumes that the log-odds are linear
Train logistic regression model	1 week	To determine strongest predictors
Train multilayer perceptron model	1 week	To predict whether an individual should receive treatment
Tune models if needed	1 week	Model parameters etc.
Create data visualizations	1 week	Charts and graphs etc.
Evaluate models and collect results	1 week	
Write final project report	1 week	
Make final project poster	1 week	

Describe the tentative plan of distributing workload among team members (0.5 points)

Everyone	Wei Xin	Andey	Claudia
- Evaluate models - Write project report - Make project poster	- Train multilayer perceptron model - Tune models	- Clean data - Create data visualizations	- Perform logistic regression - Tune models

Provide reasonable comments and constructive feedback to the proposals assigned in peer review (1 point)