Brown Datathon 2019

MINDGAP

Act fast, save lives

Kyi Goh
Columbia University

MOTIVATION

For me I naturally gravitated towards the topic given my personal experiences. Over the past two years, a handful of my close friends have been left debilitated by mental health issues - a few grapple with crippling anxiety while others bravely confront depression. Yet, as a friend, I did not notice warning signs. Indeed, many of them had not sought professional help prior to this, making it difficult to assess whether or not something is amiss.

DATASETS AND PROCESSING

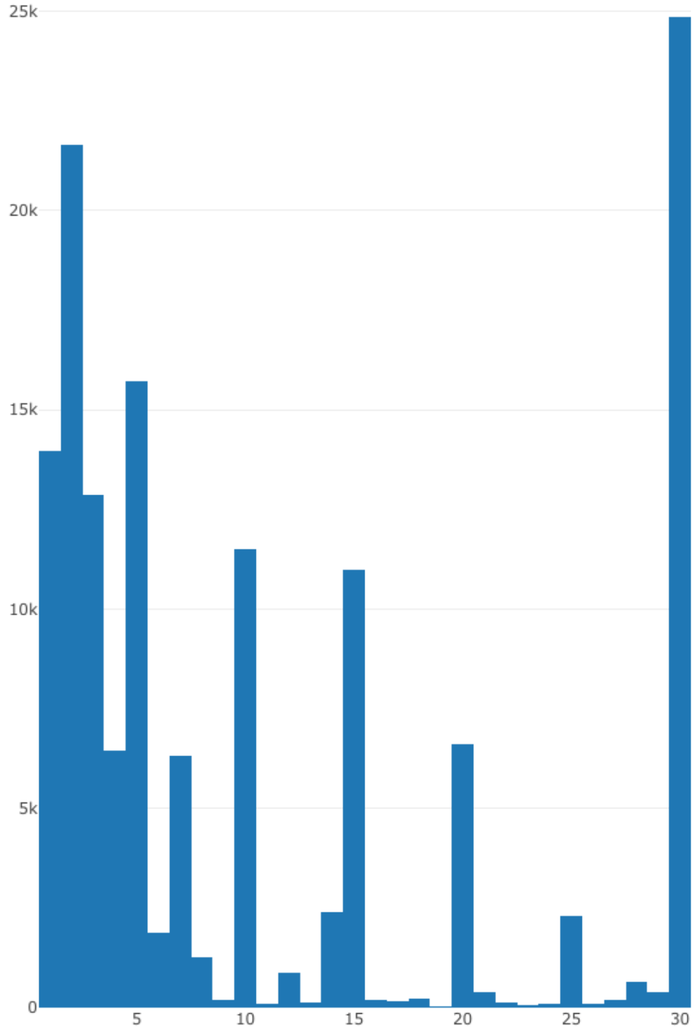
Combined two CDC datasets:

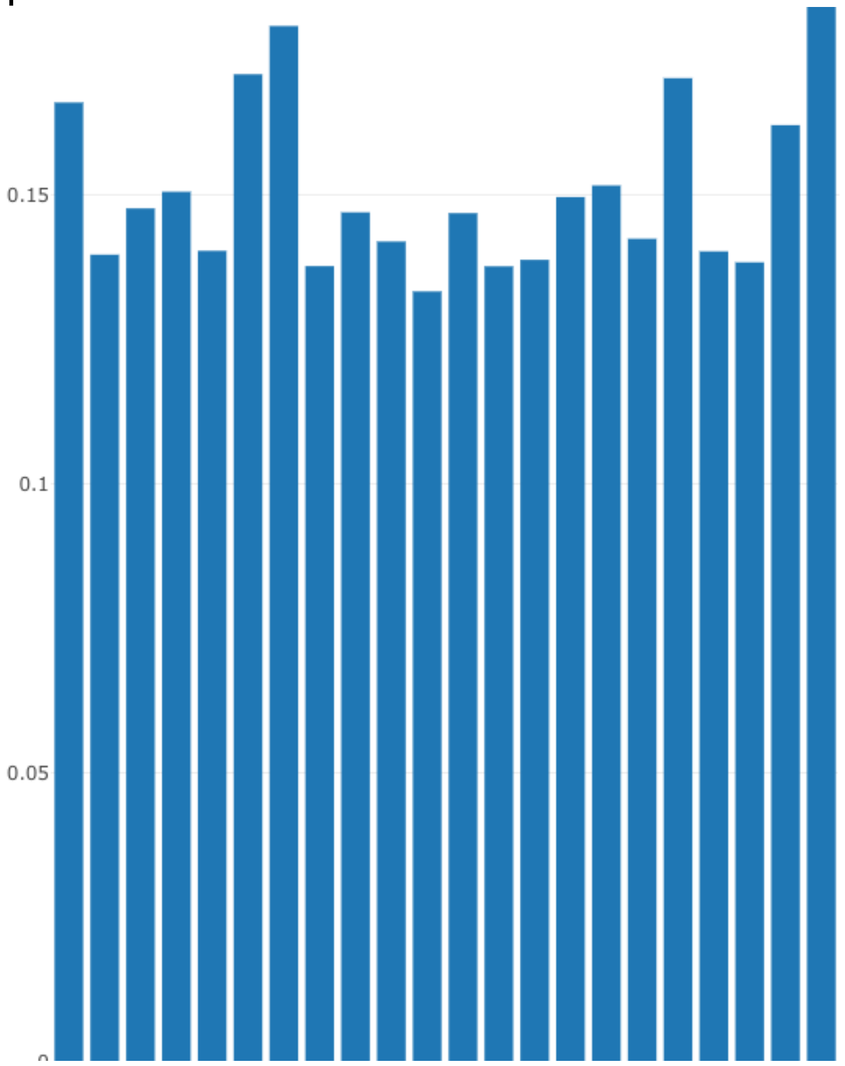
1. CDC Behavioural Risk Factor Surveillance System (BRFSS)
2. CDC Chronic Disease Indicators

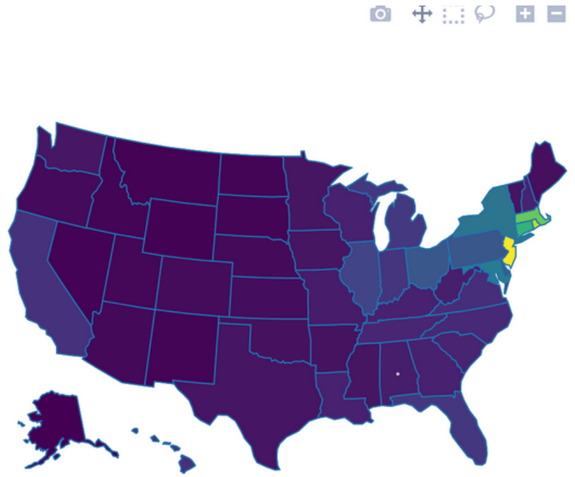
Target dependent variable, days that an individual had experienced mental distress in last 30 days. The dataset (after much cleaning) offered a total of 460+ variables, 17 of which were eventually included for testing.

EXPLORATORY ANALYSIS

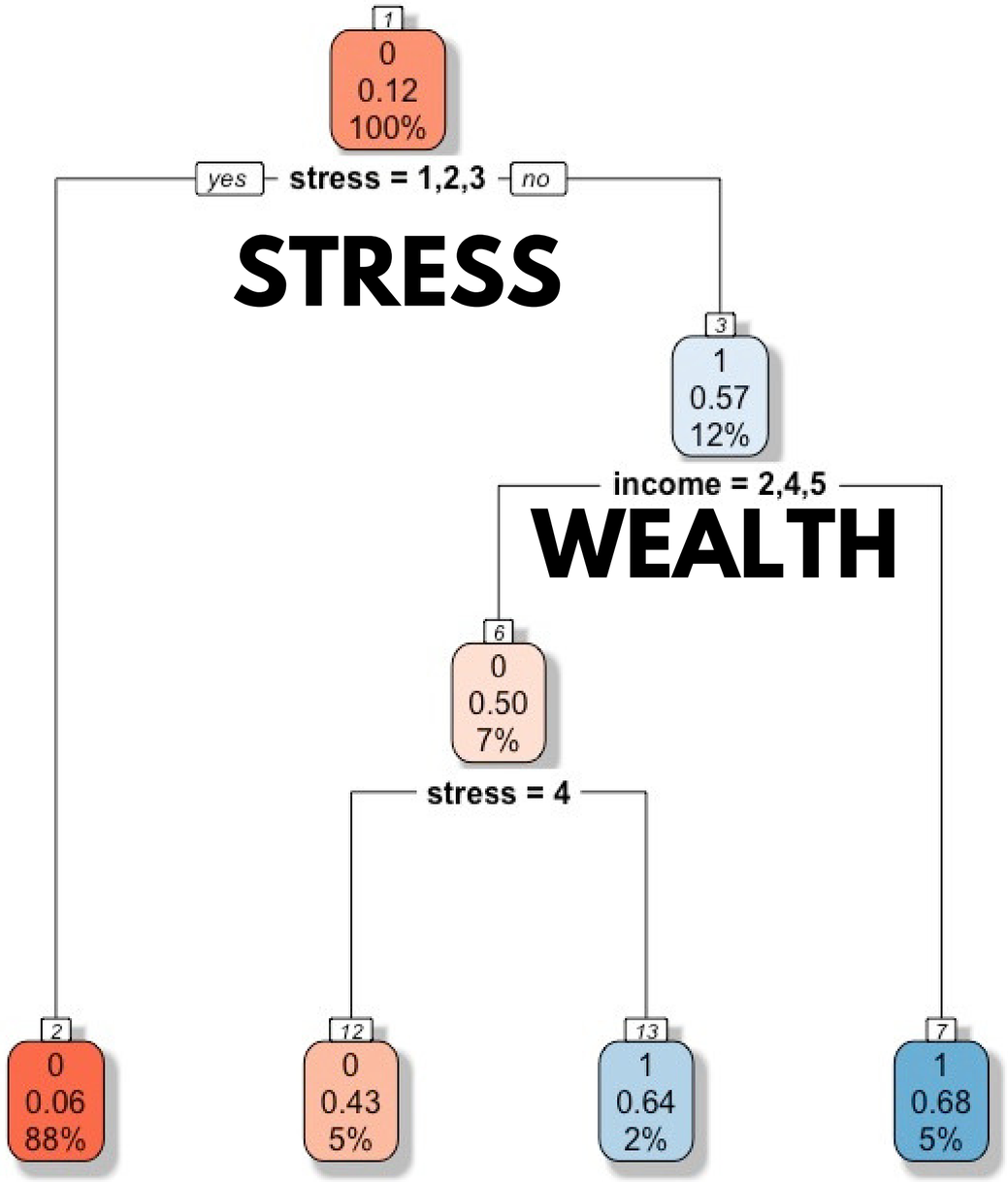
Interactive versions are included in a seperate html file







FINAL MODEL - PART 1



Decision tree ML model was ultimately the best performing, offering an overall prediction accuracy of around 92% and an accurate positive diagnosis rate of 50%.

An active decision was taken to remove previous medical records as a variable given that it would not uncover underreported cases. This would have given a much higher positive diagnosis score.

Stress is a very powerful predictor for how likely a person is undergoing soe kind of mental duress

Work appears to reduce stress, particularly amongst those with higher incomes

Interaction effects indicate that this wealth effect is offset as one ages around the 40 year old mark

FINAL MODEL - PART 2

Yet, we need to contextualise the results using a logistic regression model (though thinking in more probabilistic terms)

I also gave randomforest a shot though the results weren't much different from the two, albeit with much large false negatives.

VERY ROUGH PROJECT DEMO

A very rough guide of what this might look like from a user standpoint

	AGE	INCOME GROUP	Working?	Stress score	AGE*INCOME	Pr(follow-up)
INPUT	22	1	1	4	22	0.5156
WEIGHT	-0.113271344	-0.0653129	-0.05497119	0.48672	0.2624336	