

STATISTICS SEMINAR

Tyler Bastian

MS Defense

Friday, April 29th, 2022

Dickens 109, 2:00 pm (Central Time)



Statistical Methods to Predict Future Risk of Suicidal Ideation from Social Media Data

Suicide, the act of taking one's own life, is a tragedy for all involved and a public health concern in the United States. Suicide is the tenth leading cause of death in the United States which makes the monitoring of suicide and suicidal ideation, or the process of thinking or ruminating one's own death, crucial in the interest of public health. With the rapid development of machine learning methods, new analyses of social media data to predict individual suicide risk and behavior have been reported. A recent study, "A Machine Learning Approach Predicts Future Risk to Suicidal Ideation From Social Media" by Roy et al. (2020) <https://doi.org/10.1038/s41746-020-0287-6>, showed promising results in classifying Twitter users into a suicidal category 4, 7, 14, and 21 days in advance of expressing suicidal ideation. Roy et al. propose training a set of neural networks to detect and score psychological constructs associated with suicidal thoughts. Using the obtained scores as inputs, they implement a Random Forest algorithm to determine individual Twitter users' risk of future suicidal ideation. In this report, we offer a detailed explanation of methodology used by Roy et al. alongside evaluating the reproducibility of their work. We first extract data from Twitter and train a series of neural networks to identify if a tweet expresses psychological constructs associated with suicidal thoughts which include; burden, stress, anxiety, loneliness, insomnia, hopelessness, and depression. Using 1.2 million tweets from N = 182 suicidal ideation (SI) cases and 30,648 tweets from 347 controls, we then train a Random Forest model using neural network outputs to predict a binary outcome of SI status. The model predicted within 7 days N = 78 SI events derived from an independent set of 342 suicidal ideators relative to N = 3,458 non-SI tweets with an AUC of 0.83, slightly lower than Roy et al.'s 7 day model prediction having an AUC of 0.88. Algorithmic approaches such as this could be applied to potentially identify an individual's future risk of suicidal ideation and could be integrated into medical technologies to aid in suicide screening and risk monitoring.