Predicting the Treatment Completion for Psychological Disorders

Arman Jabbari, Dana Lansigan, Rebekah Tang, Stella Park, and Sue Kim

Overview

At the moment, studies show that just over one fifth of mental health patients do not complete treatment. This means that one fifth of patients do not recover even when they do seek treatment. Unsurprisingly, a big trend in the psychological health community has been to study appropriate treatment of mental illnesses, and the factors that can affect the likelihood of treatment completion. While this project initially started out with the goal to diagnose and differentiate between psychological mood disorders, further research into the trends of the psychological health community motivated us to pursue the issue of treatment completion instead. Harnessing the power of machine learning, we created a neural network that predicts the whether or not a patient is likely to complete treatment. We intend for our treatment completion likelihood predictor to add to the existing and ongoing conversation about treatment completion, which up until now consists mostly of traditional research studies using methods like surveys and interviews. Besides contributing to research, we intend for our predictor to also be able to assist mental health professionals on how to best optimize treatment completion based on the results of the predictor for each patient.

Using data from the National Comorbidity Survey Replication (NCS-R), our team has trained a relatively accurate dense neural network that can predict treatment completion based on numerous variables, which include sex, age, demographics, socioeconomic status, medication, treatment type, and symptoms. In order to do so, our team preprocessed a vast dataset with thousands of variables but also even more missing data points, using domain knowledge and statistical techniques. The team experimented with several machine learning models, before settling on a three-layer dense neural network as the final model, and ensured that best practices like cross-validation and hyperparameter tuning were employed.

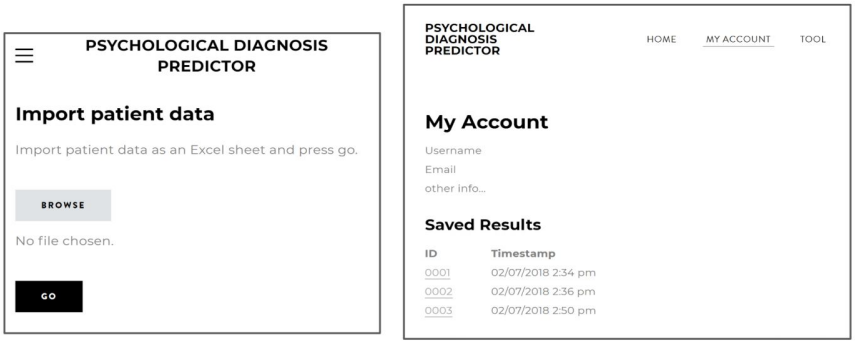
Problem Solved

Recent psychological research and studies are pivoted towards appropriate treatment of mental illnesses rather than the diagnosis. A meta analysis of 186 studies by Idaho University reports find average treatment refusal rate of 8.2% and average premature termination rate 21.9%. Treatment completion contributes not only to patient's recovery but also prevents of development of a new mental illness. Given this significance, it is crucial and necessary to understand the determinants of completion.

Our project aims to facilitate this process and aid professional psychologists devise treatments for patients by taking patient data and predicting the likelihood of treatment completion, identifying factors that significantly affect the likelihood. We do not intend to replace professional decision on treatment but propose a suggestion to be used as a tool. Extensive literature has investigated treatment completion using qualitative methods such as surveys and interviews (Jacobson, Robinson & Bluthenthal, 2007; Barnicot, Katsakou, Marougka & Priebe, 2011; Gors, Yoder, Tuerk, Lozano & Acierno, 2011). These analyses provide correlation between patient characteristics and treatment completion, but fail to give us causation. Applying machine learning techniques can thus give more comprehensive understanding of the topic and add strength the the findings.

As for the intended User Interface, we are constructing a jupyter notebook that will output a percentage given the input data. In the long term, we plan to build a website that is more user-friendly, and more accessible.

Future User Interface:



Solution

Using data from the National Comorbidity Survey Replication (NCS-R), we created a neural network to predict treatment completion for people with depression.

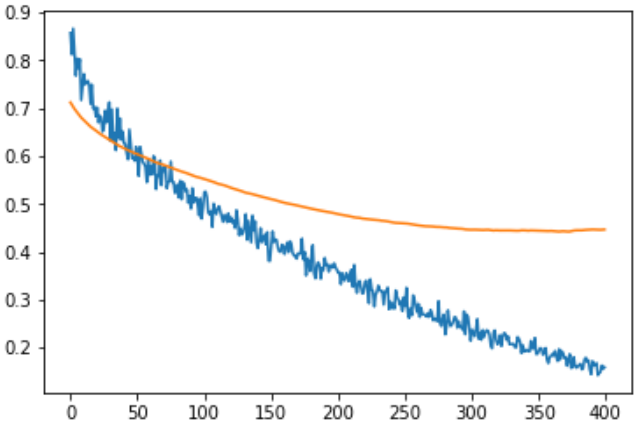
After testing many different machine learning models we learned in the class, we decided to use a dense neural network (DNN) with three layers and trained it on NCS-R data we believed were the most important factors for treatment completion. We found that the DNN was the best solution to our problem since it had the best accuracy results and smallest overfitting compared to other models we tried, among which were KNN, XGBoost, and Random Forest.

The DNN proved to be fairly accurate at predicting treatment completion. Our final testing accuracy was 88.09% and F1 score was 86.48%.

Our resulting solution is currently contained in a Jupyter Notebook, where professional psychologists or doctors can input patient information and get a prediction of whether or not that patient will complete his or her treatment. Specifically, the professional would input a patient's sex, age, height, weight, socioeconomic status, demographics, medication amount and type, treatment type, and symptoms (i.e., symptoms used for diagnosis from a typical structured interview), which were variables the network was trained on.

The resulting prediction would be very helpful for the professional to decide on the best course of treatment for the patient. Once we implement the solution on a website, we could also receive feedback from the professionals on how accurate the network's prediction was, and we can use this information to improve the network.

Test Accuracy Graph:



Loss Graph: