proposal

We are using an online dataset to predict the outcome of different therapy of depression, and the time until depression recurrence under different kinds of therapy. To do so, we are planning to do k-fold cross validation on the dataset, which can help us do better in prediction instead of explanation.

Content:

Clinical depression is the most common mental illness in the United States, affecting 19 million adults each year. Nearly 50% of individuals who experience a major episode will have a recurrence within 2-3 years. Researchers are interested in comparing therapeutic solutions that could delay or reduce the incidence of recurrence.

In a study conducted by the National Institutes of Health, 109 clinically depressed patients were separated into three groups, and each group was given one of two active drugs (imipramine or lithium) or no drug at all. For each patient, the dataset contains the treatment used, the outcome of the treatment, and several other interesting characteristics.

Variables information:

Hospt: The patient’s hospital, represented by a code for each of the 5 hospitals (1, 2, 3, 5, or 6)

Treat: The treatment received by the patient (Lithium, Imipramine, or Placebo)

Outcome: Whether or not a recurrence occurred during the patient’s treatment (Recurrence or No Recurrence)

Time: Either the time (days) till recurrence, or if no recurrence, the length (days) of the patient’s participation in the study.

AcuteT: The time (days) that the patient was depressed prior to the study.

Age: The age of the patient in years, when the patient entered the study.

Gender: The patient’s gender (1 = Female, 2 = Male)

Link: <https://www.kaggle.com/everseek/depression>