Final Project Proposal

###PPOL 670 Final Project Proposal

Introduction

In 2015, over 43 million adults had a mental illness, and nearly 10 million had a serious mental illness, such as depression, bipolar disorder, or schizophrenia. People with mental health conditions often have chronic medical conditions, significant health care services utilization, and barriers to employment, and are frequently involved with the criminal justice system. The National Health Interview Survey (NHIS) has monitored the health of the nation since 1957. NHIS data on a broad range of health topics, including mental health conditions, are collected through personal household interviews. The U.S. Census Bureau has been the data collection agent for the National Health Interview Survey. Survey results have been instrumental in providing data to track health status, health care access, and progress toward achieving national health objectives.

For the Final Project, I will analyze the National Health Interview Survey data to generate a machine learning model that predicts the occurrence of depression, anxiety, or emotional problems among respondents.

Data

I will obtain the relevant data needed from The National Health Interview Survey (NHIS) Website. Several datasets contain the data on occurrence of depression, anxiety, and emotional problems as well as variables that impact it. I will trace the relevant datasets and combine the data to form a dataset that can be used for my analyses. The datasets are available in a .csv format. An overview and dictionary of the variables is also provided.

Methods

I will manipulate the variables and change the unit of analyses to make it appropriate to my analyses. I plan to clean and manipulate the data as well as merge the datasets using the tidyverse package and the data wrangling methods learned in class. I will be employing the ggplot2 and other packages to create professional-quality visualizations to present various aspects of the data and my results in a visually appealing and easily understandable way. I will be using the concepts and supervised learning techniques learned in class to create a machine learning model for my analyses. I plan to run different models such as linear model, KNN model, Regression trees, and Random forest to tune the parameters as much as possible to determine the best model for supervised learning. I will present the results and my prediction from the model using visualizations.

Success Criteria

As previously mentioned, I plan to create a machine learning model to predict the occurrence of depression, anxiety, or emotional problems. I plan to use data from 2008 to 2018. I will use data from 2008 to 2017 as training data and data from 2018 as test data. I will then build a machine learning model using the data on the variables available in the dataset to predict the occurrence of depression, anxiety or emotional problems 2018. I will consider the project to be successful if the predicted readmission rate is within 10% of the actual readmission rate as available in the dataset.

Using the skills and concepts learned in class along with my background in working with mental wellness programs, I am confident I will be able to produce a meaningful analysis.

Correction from Previous Proposal

As per my discussion with the professor, I have identified a new dataset for the project with a higher number of observations. The current dataset has around 24,000 observations.

References

+KFF. (2019). Facilitating Access to Mental Health Services: A Look at Medicaid, Private Insurance, and the Uninsured. [online] Available at: https://www.kff.org/medicaid/fact-sheet/facilitating-access-to-mental-health-services-a-look-at-medicaid-private-insurance-and-the-uninsured/:~:targetText=The%2 [Accessed 12 Nov. 2019].