Predicting Depression Using Machine Learning

Vivian Dong 300525329, Saumya Sajwan, Samanalie Perera

November 1, 2022

Contents

1 Executive Summary

2 Background

A study found that depression increased from 9% in 2017–2018 to over 14% in April 2020 among US adults during the COVID-19 pandemic. Among different age groups, the mental health of young adults is most affected by the pandemic. This increased rate was not normal compared to other years, and there was not enough attention paid to data examining tools either. (Daly, Sutin, and Robinson 2021)

There are blood tests, brain scans and other medical examining methods for a depression diagnosis. In the end, the most affected way is to let the patients describe their symptoms. To achieve this, patients can answer a questionnaire like determining the frequency of depression symptoms over the past two weeks. This method may lead to subjective bias and imperfections in the diagnostic capabilities. (Buntinx et al. 2004) However, the cause of depression could be from many other things, such as sleep disorders, drug use, alcohol use and weight loss. For example, if a person has been working in a stressful environment with low income and not enough sleep, that person may have a higher chance than other people to have depression.

Healthcare data from the National Health and Nutrition Examination Survey database includes a wide range of concepts, like health records, genetic information and even demographic data. Furthermore, machine learning tools tend to perform better than humans at processing these big data sets and making use of it.(Beam and Kohane 2018)

Our primary goal in this project was to train a machine learning classification model to identify patients who suffer from depression using demographics and healthcare data from the NHANES database.

3 Data Description [1-2 pages]

- State the types of data in the dataset(s) and the structure of the dataset(s). Are the data nu- merical, categorical, or both? Time series? Coordinates? Diagnostic categories? This does NOT need to be an exhaustive list of every variable, just a few comments on the overall types.
- Specifically state the attributes of any variables that are used in the Detailed Analysis, in partic- ular for any categorical variables explain the meaning of their categories
- State how complete the dataset(s) are (i.e.how many missing, any structure to the missing data, whether there are errors in the data)

we are using a questionnaire and calculate score.

4 Ethics, Privacy and Security

5 Exploratory Data Analysis

6 Detailed Analysis Results

7 Conclusions and Recommendations

Reference List

Beam, Andrew L., and Isaac S. Kohane. 2018. “Big Data and Machine Learning in Health Care.” *JAMA* 319 (13): 1317. https://doi.org/10.1001/jama.2017.18391.

Buntinx, Frank, Jan De Lepeleire, Jan Heyrman, Benjamin Fischler, Dirk Vander Mijnsbrugge, and Marjan Van den Akker. 2004. “Diagnosing Depression: What's in a Name?” *European Journal of General Practice* 10 (4): 162–65. https://doi.org/10.3109/13814780409044305.

Daly, Michael, Angelina R. Sutin, and Eric Robinson. 2021. “Depression Reported by US Adults in 2017–2018 and March and April 2020.” *Journal of Affective Disorders* 278 (January): 131–35. https://doi.org/10.1016/j.jad.2020.09.065.