
Using a Naive Bayes Classifier to Identify Major Depression

Ivan Ching

Department of Computer Science
Stevens Institute of Technology
Hoboken, NJ 07030
iching@stevens.edu

Anthony Jin

Department of Computer Science
Stevens Institute of Technology
Hoboken, NJ 07030
ajin@stevens.edu

Brandon Cheung

Department of Computer Science
Stevens Institute of Technology
Hoboken, NJ 07030
btruong@stevens.edu

Abstract

There exists databases that contain information regarding the mental health and various possible indicators to mental diseases. In this study, we will utilize the NESARC dataset to investigate possible indicators of major depression. Due to the probabilistic nature of a Naive Bayes classifier, we will utilize one in order to classify survey participants into those with major depression and those without.

1 Introduction

Major depression is a mental disorder categorized by extended periods of low mood. It is often accompanied by low self-esteem and loss of interest in activities on used to fine enjoyable. Major depression, when left untreated, may negatively affect a person's daily life and those around them. Between 2-7% of adults suffering from major depression die by suicide. Common treatments for depression are usually after diagnosis of the mental illness.

We propose a different method of prevention by identifying risk factors of major depression before the subject allows the illness to progress to a dangerous stage. In order to accomplish this, we will use a Naive Bayes Classifier to identify people groups who display higher risk, or a higher probability of having, of major depression.

2 Data

The database we will be using is the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). It is a national organization that studies the occurrences of psychological disorders and substance abuses, including major depression. It is published and maintained by the National Institute of Health and Department of Health and Human Services. This survey contains data on more than 36,000 U.S. adults, which should be enough data to build a model that is sufficient for predicting major depression.

3 Naive Bayes Classifier

The Naive Bayes classifier is a supervised learning algorithm that applies Bayes' theorem with strong independent assumptions between features. Due to the strong probabilistic nature of Naive Bayes, we will be able to identify which features are the strongest indicators for major depression. Due to the assumption of conditional independence made by the Naive Bayes classifier, it scales very well as we increase the number of features we use to learn our model. Seeing as the NESARC dataset could possibly have a great number of features, this is imperative to efficiently train and predict a reliable model.