A depression prediction model based on causal inference and machine learning

Xilin Zhang¹, Tiantian Wang², and Chuang Xue³

¹University of Science and Technology of China School of Mathematical Sciences
²Hunan University of Technology and Business
³Hangzhou Seventh People’s Hospital

March 13, 2023

Abstract

Background: Depression is one of the most common psychological disorders nowadays, with continuous and prolonged low mood as the main clinical feature, and it is the most important type of psychological disorders in modern people. The aim of this study is to develop a depression prediction model based on causal inference and machine learning. Methods: This case study included 7000 subjects. A feature selection model was built based on a causal inference algorithm. The selected features were entered as variables in seven machine learning (ML) models built to create a predictive model for the diagnosis of depression. Results: Among the seven ML models, the random forest model (RF) showed the best performance. For the prediction of depression, the area under receiver operating characteristic (AUC) of the RF model was 0.908 (0.810-1.00) in 10-fold stratified cross-validation and 0.901 (0.893-0.91) in external validation.

Hosted file