
Moises Marin Martinez

<https://www.linkedin.com/in/moises-marin-517196113/>

mmarinm@okstate.edu | +1 669 246 3504 | +52 55 30704132

Improving Diagnoses of Depression With XGBOOST Machine Learning Model and a Large Biomarkers Dutch Dataset (n = 11,081)

4th December 2022

1. Machine learning to the rescue

Healthcare is one more area that can benefit from machine learning algorithms, in this paper the Extreme Gradient Boosting (XGBoosting) algorithm was used to classify depression cases in a sample dataset of 11k rows. Machine learning is used in the agnostic approach used to study the complexity of disorders in psychiatry, it is particularly useful to draw insight from variables of interest in big datasets that are not normally distributed.

In the case of depression, it is screened using interview questionnaires with one study showing that up to 27% of cases failed to be correctly classified in the clinical interview. The high amount of false negatives highlights the limitations of the classic interview methods, to overcome this its important to study the relationship between biomarkers and depression.

2. More than an interview

A biomarker is a characteristic that can be measured and evaluated in relationship with a process, examples of biomarkers are red cell distribution width, serum glucose and total bilirubin. The role of machine learning in this problem is to identify the biomarkers and help predict disease positive cases. Some studies have used SVM with whole brain images to classify patients. XGBoost was applied to identify the biomarkers that are most important for predicting depression, the dataset studied has a positive depression classification of 5.14%. To balance the data multiple resampling methods were used .

The best performance of this ensemble algorithm was with the dataset that was oversampled, it has the highest accuracy of 0.9729. Biomarkers present in blood and urine where the ones analyzed and used to predict mental illnesses.

3. What happens in the body doesn't stay in the body

I was not familiar with biomarkers and the way they were used by classification algorithms to help the healthcare domain. It is very interesting how images of the brain or blood and urine samples can be studied to identify mental illnesses unbeknown to the patient. I think this paper was published in an analytics journal because it is presenting an edge case where machine learning is being used as a tool by other disciplines. If I had participated in this research I would have presented more examples of important biomarkers found to be associated with mental illnesses, to show how machine learning is opening a door that was closed before, that of associating what happens in the body to what we see on the outside behavior.