CS 334
Dr. Li Xiong
10/22/2021

Project Proposal

Title: Evaluation of the effectiveness of applying different ML algorithms on MIMIC-IV Dataset to reduce racial bias in ICU mortality prediction

Group Members: Elijah Chou, Tiantian Li, Tianqi Bao

Description of Problem:

Despite wide utilization of severity scoring systems for case-mix determination and benchmarking in the intensive care unit, the possibility of scoring bias across ethnicities has yet to be thoroughly examined. Recent guidelines on the use of illness severity scores to inform triage decisions for allocation of scarce resources such as mechanical ventilation during the current COVID-19 pandemic warrant examination for possible bias in these models.

Description of Dataset:

The Medical Information Mart for Intensive Care (MIMIC) dataset is a large, freely-available medical dataset that contains deidentified patient health data that was curated for research purposes. The dataset specifically contains ICU data from ICU units from the Beth Israel Deaconess Medical Center, Metavision, and CareVue. The dataset is categorized into five main modules, which consist of patient stay information, hospital level data, ICU level data, emergency department data, and metadata that links this database to a related database. For this project, we hope to use the information from the ICU level data to do our analyses.

References/Work Done so Far:

Previously, researchers have investigated the performance of three severity scoring systems across ethnic groups in two large ICU databases and found statistical evidence that suggested that illness severity scores did not discriminate for severity of disease, but were poorly calibrated for blacks and hispanics where these scores over predicted mortality. In this study, a simple logistic regression model was used to predict mortality using only the average scores as a model feature. In another retrospective study, researchers concluded that non-Hispanic Black patients were more likely to be denied medical resources if the SOFA score, a prominently used severity scoring system, is utilized. This further justifies the need for a better machine learning model to be implemented in limited hospital resource allocation.

Description of Tentative Plan:

We will first evaluate the effectiveness of different ML algorithms we have learned in class in mortality prediction, including *Knn* and *decision tree* on this dataset. We are going to apply

various feature selection methods that were introduced in class, including *filter methods, wrapper methods* and *embedded methods,* and we will evaluate the effectiveness of different feature selection methods probably on one of the ML algorithms or more. The effectiveness would be concerned with its accuracy and time spent.

Then we would like to determine if there are any racial biases in the prediction after we train the model with the most effective algorithm on the whole dataset, while excluding the "race" column and predict if a patient accepted into the ICU is likely to die based on the other features given or selected. We would split data into different race groups and evaluate the AUC on those groups as for the racial biases. During the testing phase, we will divide the test data by racial groups and apply the model on each group. We gather data on prediction accuracy for each racial group separately and determine if there exists a significant difference in the accuracy data among groups with calculated confidence intervals.

Citations for Mentioned Papers (In Order of Appearance):

Sarkar, Rahuldeb et al. "Performance of intensive care unit severity scoring systems across different ethnicities." *medRxiv : the preprint server for health sciences* 2021.01.19.21249222. 20 Jan. 2021, doi:10.1101/2021.01.19.21249222. Preprint.

Tolchin B, Oladele C, Galusha D, Kashyap N, Showstark M, et al. (2021) Racial disparities in the SOFA score among patients hospitalized with COVID-19. PLOS ONE 16(9): e0257608. https://doi.org/10.1371/journal.pone.0257608