

Data Science Case Study

Background:

We have data regarding patients with COVID-19 and a hospital has asked us to predict the need for mechanical ventilation among patients. The goal is to reduce patient mortality by identifying high risk patients based on the available data and devoting the limited ventilators to them.

Data:

baselines.csv – Contains patient level data regarding patient characteristics

labs and vitals.csv – Contains time series biometric information for patients

Exercise:

1. Perform data cleaning and feature engineering
2. Explore the data and perform appropriate data visualization to extract insights
3. Use statistical methods to make inferences
4. Select ML classifiers to predict the “event” (need for mechanical ventilation among patients). Note: Check assumptions and provide reasoning for this model selection
5. Tune hyperparameter to improve any chosen metric (Eg: misclassification error/ accuracy/ precision/ recall) and discuss the results.

Please include explanations wherever necessary. Feel free to use either R or Python and make sure you submit your solution as an HTML file. Using online resources for reference is permissible however plagiarism will not be tolerated.