Project 2 Proposal Team # 8

Predicting patient outcomes is one of the most important components of the global healthcare industry. Anticipating a patient's probability of survival is crucial because it helps doctors make optimal choices and elevates the standard of medical care in general. We started Project 1 to explore the exciting field of patient survival prediction by analyzing a dataset named "Patient Survival Prediction". This dataset contains approximately 92,000 rows with roughly 190 columns. Prior to moving on to feature importance, partitioning the data for training and testing, and model development, we preprocessed (EDA) to clean the dataset. This comprises taking care of missing values & outliers, understanding the distribution of data and analyzing relationships between columns. Based on a number of factors, including heart rate, creatinine, bilirubin levels, etc., we will be able to ascertain if a patient survives or not by evaluating the dataset. By using the data to train an ML model and varying the test data, we can estimate the survival rate while taking into account the medical history of the patients. Since, our target variable is 'hospital_death', which is binary (0&1), employing classification models such as logistic regression is appropriate. This study issue is a real-world use case scenario because it will enable medical professionals to diagnose patients more accurately and efficiently by pinpointing their precise health state. People's lives and time can be saved by using this data and the model that is based on it. We may also deduce the importance and influence of intensive care units from this data and determine whether or not they actually enhance patients' health.

Smart Questions :

1- How does the presence of diabetes mellitus correlate with age? Is diabetes a risk factor for hospital death in older patients, and does age significantly impact the chance of developing diabetes?

2- Are there any particular illnesses or risk factors (such as leukemia, cirrhosis, or diabetes mellitus) that are closely linked to hospital deaths?

3- Which underlying medical disorders account for the majority of hospital deaths among patients over the age of 70? Are there any trends in the kinds of illnesses that cause death in this particular age range?

4- What is the mortality rate in the dataset between patients under 65 and those above 65? Are the causes of death for these age groups significantly different from one another?

5- Which ICU types are most prevalent in the dataset, and does the type of ICU affect the risk of hospital death?

6- What differences exist between patients who survive and those who do not in terms of the distribution of "BMI" (body mass index)? Does a person's BMI have any bearing on hospital mortality?

Github link: https://github.com/SuryaVamsi-P/FA-23_DATS_6101-Team_8
Source of data : Kaggle, Patient Survival Prediction (kaggle.com)