Predive Analytics of Hospital Readmissions

A healthcare organization lose over $30 million annually due to the readmission of patients who are discharged from the hospital prematurely. On the other hand, keeping all patients in the hospital longer, regardless of their condition, is a costly measure that is also inconvenient to patients.

You are hired by the hospital to use your skills in data analytics to develop a model that predicts the readmission risk. By doing this, doctors can use your predictive analytics to decide whether to discharge a patient without delay. The hospital shares a dataset of 10,000 patients with you in file named 10Kpatients.csv.

To help doctors understand your approach, you are asked to create a report that explains the techniques you used, the accuracy, and the confidence of your model. So, feel free to describe your techniques such as: the preprocessing procedures, consistency checks, draw scatter plots, show histograms, discuss correlations, outliers detected, machine learning models used / attempted, and interpret results.

Project Guide

Data Preprocessing (5 out of 20 points)

Did you perform any preprocessing techniques (column dropping, imputation, etc.)?

- What are the attributes (columns) that have missing data more than 50%? You can drop these columns as they might be beyond fixing. Check Weight, Payer Code, and medical Specialty.
- For the rest of the columns, how did you fix the missing/Not available data? Did you impute (assign values for the missing fields) these missing items? Did you use mean, median, or kNN imputation? (i.e., replacing missing values by the column's mean, or median, or using kNN to guess them).
- (Optional/bonus) feel free to share any other preprocessing techniques you applied.

Exploratory Analytics (Data Visualization and Plotting)

- Is the dataset skewed? What is the percentages of the following (plot each): (5 out of 20 points)
  - Patients readmitted vs non-readmitted patients?
  - Patients on Diabetic medication vs patients who are not?

- Patients demographics (race, age, gender)? Plot each.
- Patients passing away (discharge = Expired) vs staying alive?

- Are there any interesting patterns that involve patients' readmission or passing away (Expired)? Pattern mining and association rule learning (5 out of 20 points) tips: Apriori
  - What are the rules that cause a patient not to be readmitted? discuss how interesting is the rule and also reflect on its lift, support, and confidence values of the rules
  - What are the rules that cause a patient to be readmitted? discuss how interesting is the rule and also reflect on its lift, support, and confidence values of the rules
  - What are the rules that cause a patient to pass away (Expired)? discuss how interesting is the rule and also reflect on its lift, support, and confidence values of the rules

Predictive Analytics (5 out of 20 points)

Provided with a patient's attributes (for example, 20 attributes i.e., race, gender, ….., diabetesMed), can we predict whether or not the patient will be readmitted?

- What is the ratio of the training:testing data you chose? What is the percentage of the data you chose for training? Percentage of testing part?
- Which learning algorithm(s) did you use to build a predive model for patient's readmission.
- After building a predictive model to predict patients' readmissions (yes/no) using the provided data, how accurately can we predict whether patients will be readmitted? Reflect on the model's accuracy measures: True Positives, True Negatives, False Positives and False Negatives. What is the meaning of each value for your model.

Final Report Guidelines

Please submit a 2-15 page report describing the answers previous questions. Please do not answer the questions in a Question/Answer style, instead, write a report describing your process and steps while using the provided questions as a guide. Please include the required data visualization and plots (~6 Figures) in the report. The code/script used for these steps should be provided in separate file(s) with the submission package.