Predicting In-Hospital Mortality using the eICU Dataset

Ankita Singh, Purav Zumkhawala, Joel Dsouza
Under guidance of Dr. Bobak Mortazavi
Department of Computer Science, Texas A&M University

Objective

- Predicting in-hospital mortality rate of patients upon ICU discharge using two different models
- To explore possibility of patient re-admissions
- To analyze hospital-wise mortality rate

Introduction

- Patient is placed in an Intensive Care Unit (ICU) when the patient's condition is extremely critical
- Incorrect readings of vital signs or mistakes by medical staff can cost a persons life
- Aim of the project is to predict patients in-hospital mortality rate upon ICU discharge using supervised and unsupervised machine learning techniques
- Dataset used is eICU Collaborative Research Database [4]
- Skewed data with 200,859 patient unit encounters for 139,367 unique patients
- Admission Baseline model and Retrospective Derived Features model

Results

The following results were obtained

In-hospital Mortality Prediction

- L1 converged faster
- No change in performance on tuning C, PaO2, pH, sodium - most important features

Re-admissions Prediction

- AUROC = 0.575

Hospital-wise Mortality Prediction

- Average f1 score = 82

Methods

In-hospital Mortality Prediction

Target - Status on discharge

- Admission Baseline model - age, gender, and features needed to calculate the SAPS II score at admission (3 features)
- Logistic Regression - weighted labels for higher penalty on misprediction
- XGBoost - Tree depth = 3
- Misprediction penalty causes overall loss to keep decreasing
- Retrospective Derived Features Model - age, gender, features needed to calculate the SAPS II score, and all Elixhauser comorbidities (36 features)

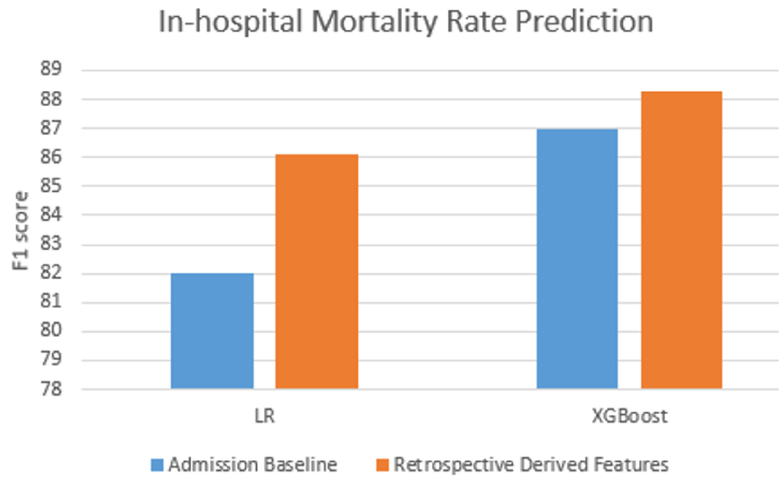
Re-admissions Prediction

Target - Type of ICU visit

- Logistic Regression with 19 predictors - age, gender, features for SAPS-II score

Hospital-wise Mortality Prediction

- Leave-one-out cross validation using the hospitals



Related Work

- Earlier works [3] [1] utilized outdated MIMIC-III dataset [2]
- [3] performs re-admissions predictions
- AUROC 0.71
- [1] predicts in-hospital, 30-day and 1 year post-discharge mortality rate
- AUROC 0.771, Sensitivity 0.999, Specificity 0.010

Discussions and Future Work

- Mortality rate and re-admissions is predicted more accurately using the eICU dataset
- Selection of hospital has no effect on mortality rate
- Future work - Perform mortality rate and re-admissions predictions using notes made by doctors and nurses

References

[1] Marzyeh Ghassemi, Tristan Naumann, Finale Doshi-Velez, Nicole Brimmer, Rohit Joshi, Anna Rumshisky, and Peter Szolovits.
Unfolding physiological state: Mortality modelling in intensive care units.
In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 75–84. ACM, 2014.

[2] Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark.
Mimic-iii, a freely accessible critical care database.
Scientific data, 3:160035, 2016.

[3] Arash Pakbin, Parvez Rafi, Nate Hurley, Wade Schulz, M Harlan Krumholz, and J Bobak Mortazavi.
Prediction of icu readmissions using data at patient discharge.
In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4932–4935. IEEE, 2018.

[4] Tom J Pollard, Alistair EW Johnson, Jesse D Raffa, Leo A Celi, Roger G Mark, and Omar Badawi.
The eicu collaborative research database, a freely available multi-center database for critical care research.
Scientific data, 5, 2018.