2019
Research Experience for Undergraduates

Predicting Readmissions: Cohort Selection

Ayzhamal Zhamangaraeva
Advisor: Dr. Ioannis A. Kakadiaris

Final Presentation
August 9th, 2019

UNIVERSITY of HOUSTON

Motivation

- “Hospital Readmission Reduction Program” established by Medicaid since Oct 1, 2012
- About half of US registered hospitals are penalized for **excess** readmission rates
- Historically, **nearly 20%** of all Medicare discharges had a readmission **within 30 days**
- Congestive Heart Failure (CHF) is the **most frequent** diagnosis for 30-day readmissions

Accurate prediction model can help to readjust health care, reduce readmission rates and health care costs.

Goal

To select cohort for prediction model of hospital readmissions for Heart Failure patients

Objectives

1. Define the cohort of CHF patients
2. Compute new features
3. Partition to training, testing, and validation datasets
4. Build a machine learning prediction model

Objective 1: Tasks

Define the cohort of CHF patients:

1. Explanatory Data Analysis
2. Define criteria for CHF cohort
3. Write R script to filter the cohort
4. Compare positives and negatives

Objective 1: Accomplishments

Criteria: length of stay>0, non-deceased, discharge status code to home or self-care, ICD-9 version code, CHF diagnosis

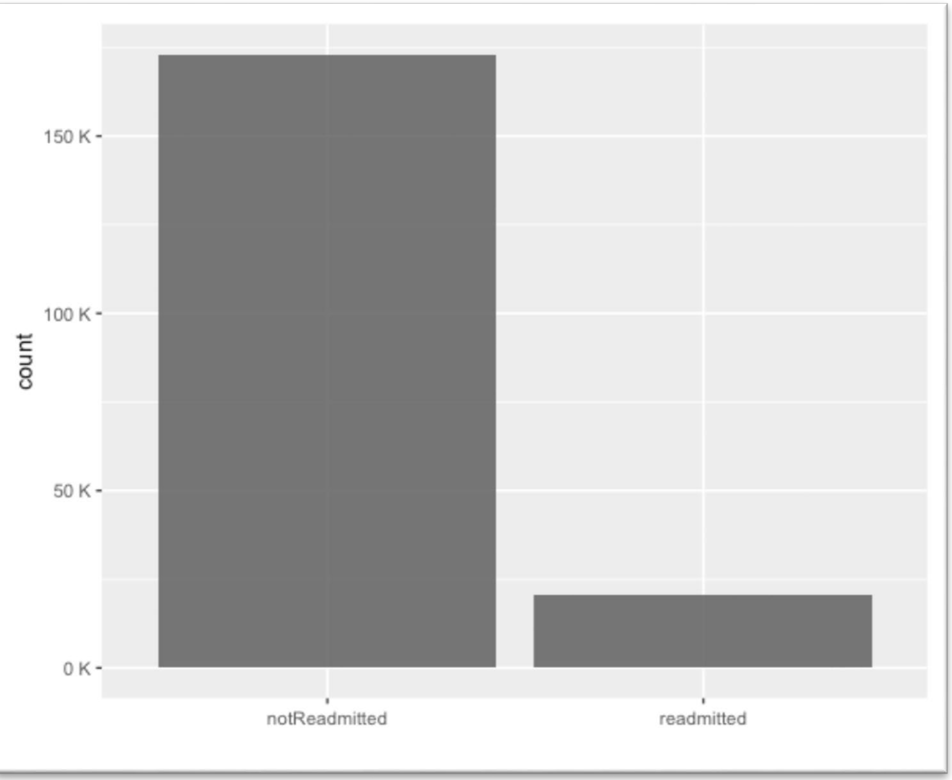
Total 193,419 CHF hospital admissions related to 108,453 patients id
20,415 positive readmissions = 10.55%

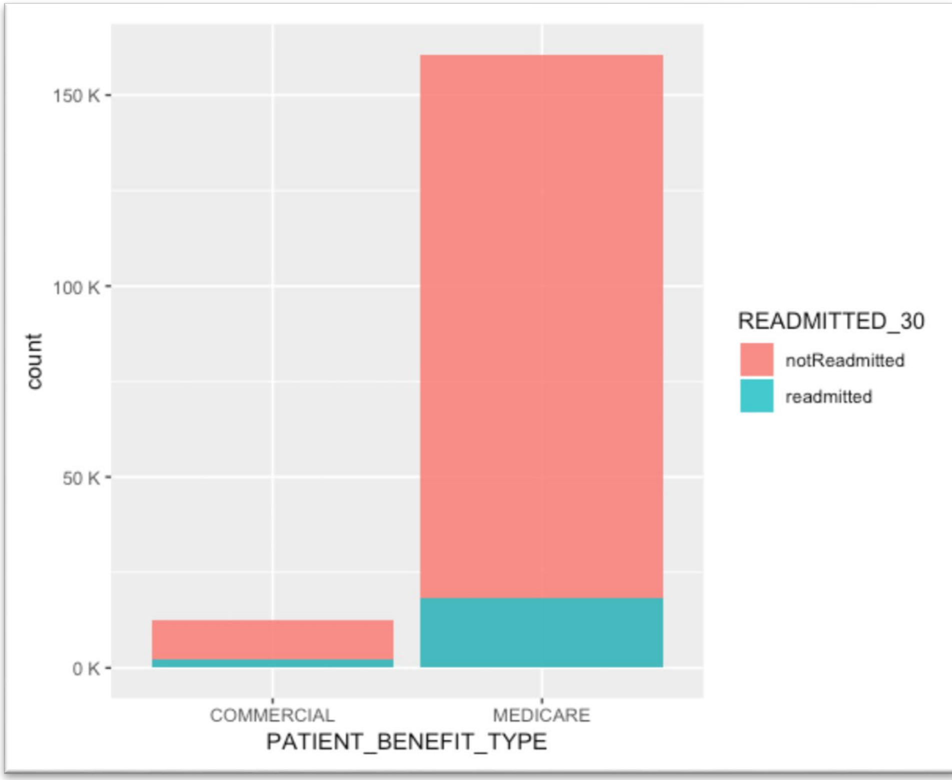
Objective 1: Methodology

Conceptual challenge: presence of different classification systems for diagnoses, ICD structure (428.21), repetitive records with overlapping dates

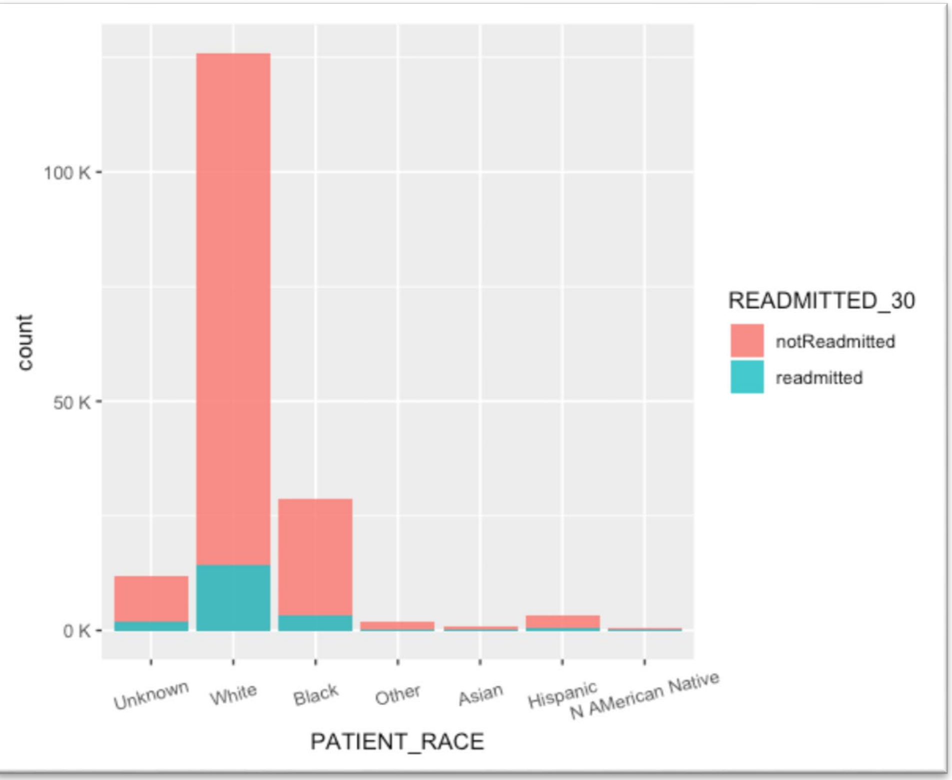
Solution: setting new assumptions and restricting criteria

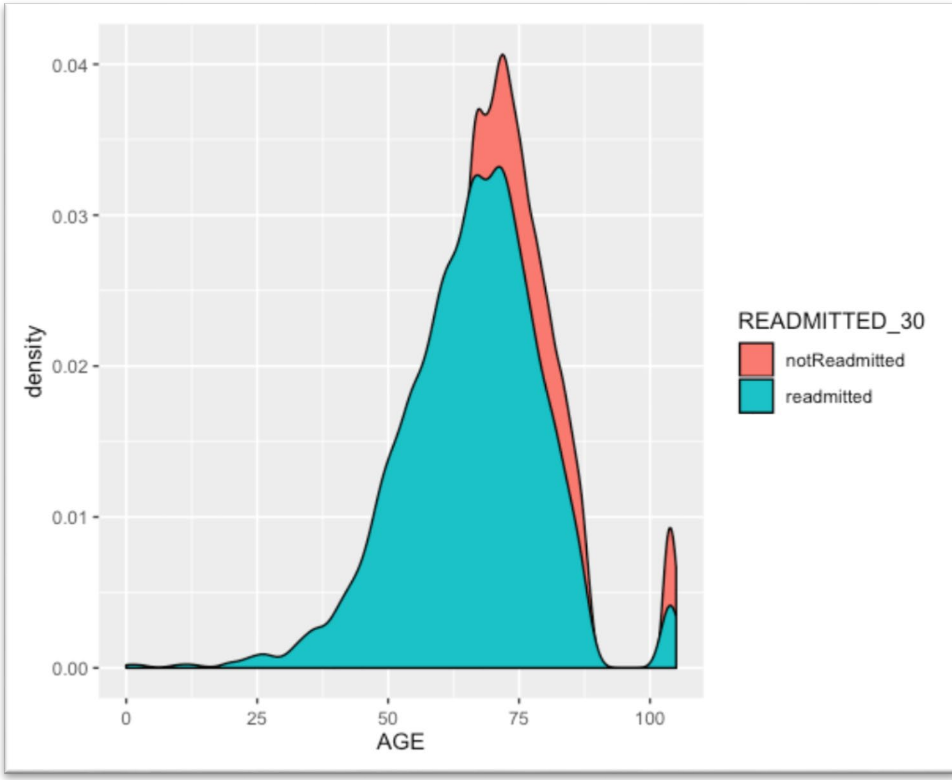
Objective 1: Results





Objective 1: Results (2)





Objective 2: Tasks

Compute new features:

1. Define new features from LR
2. Write R script to derive new features
3. Compute possible features

Objective 2: Accomplishments

New computed features:

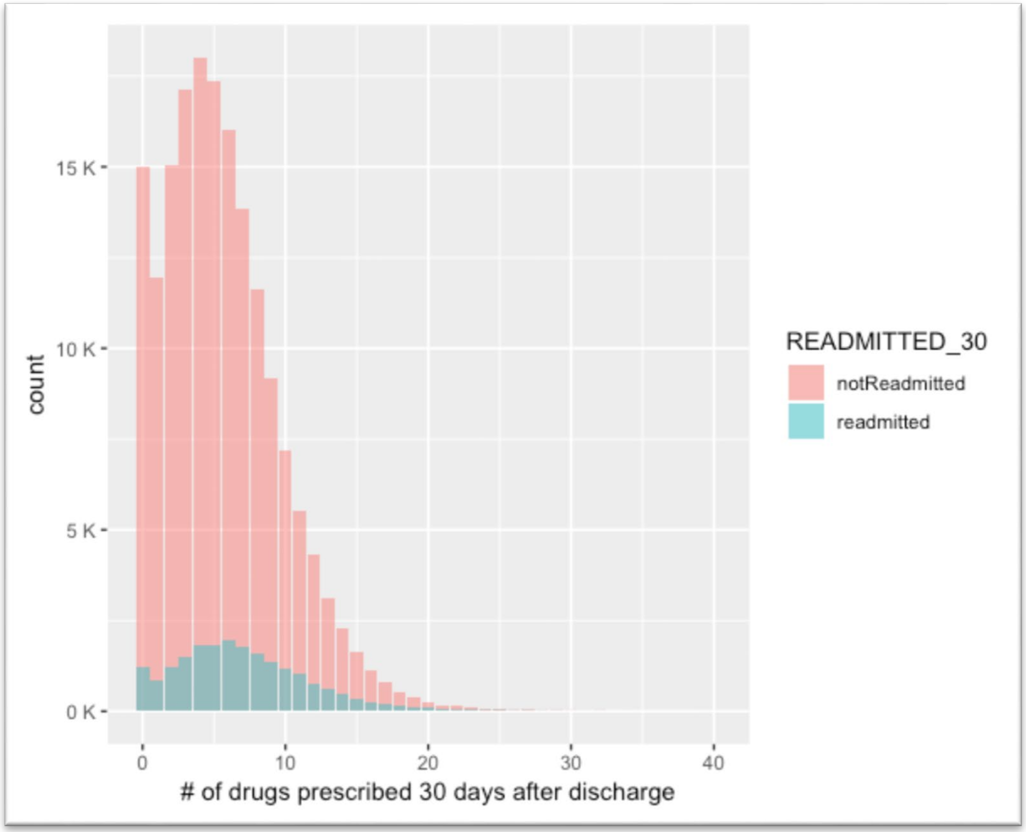
- # of diagnoses for all period
- # of chronic conditions for all period
- # of generic drug names 30 days before admission, during hospital stay, and 30 days after discharge
- # of unique lab orders for different timeframes
- age

Objective 2: Methodology

Technical challenge: long I/O process time, RAM memory exhaust, long computation processing time

Solution: “fst” package for I/O, partition a file into chunks and submitting separate slurm jobs

Objective 2: Results



rxAfter: mean=5.8, max=40

labDuring: mean=6.7, max=240, 85% of CHF admissions have 0 lab orders during hospital stay-> incomplete data

Deliverables

1. Dataset with computed features
2. Source code for cohort selection
3. Literature review XLS
4. Final report

Limitations and Future Work

1. Different ICD version codes – develop an equivalence table
2. Negative readmission intervals – clarify with the creator of dataset
3. Lab file has parsing problems and incomplete - reduces cohort size
4. Need to incorporate physician’s intuition on feature set

Future Work

1. Derive more features
2. Partition to training, testing, and validation datasets
3. Build a machine learning prediction model

Self-Reflection

1. Coding is not enough
2. It’s very important to **understand** the data and **analyze** it
3. Data pre-processing is FUN ☺

Acknowledgements

The REU project is sponsored by NSF under award NSF-1659755. Special thanks to the following UH offices for providing financial support to the project: Department of Computer Science; College of Natural Sciences and Mathematics; Dean of Graduate and Professional Studies; VP for Research; and the Provost's Office. The views and conclusions contained in this presentation are those of the author and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the sponsors.