Predicting Mortality in the ICU:
XGBoost Modeling and Deployment

Asa Wilks, UCLA Statistics 418

MIMIC Data Summary and Goals

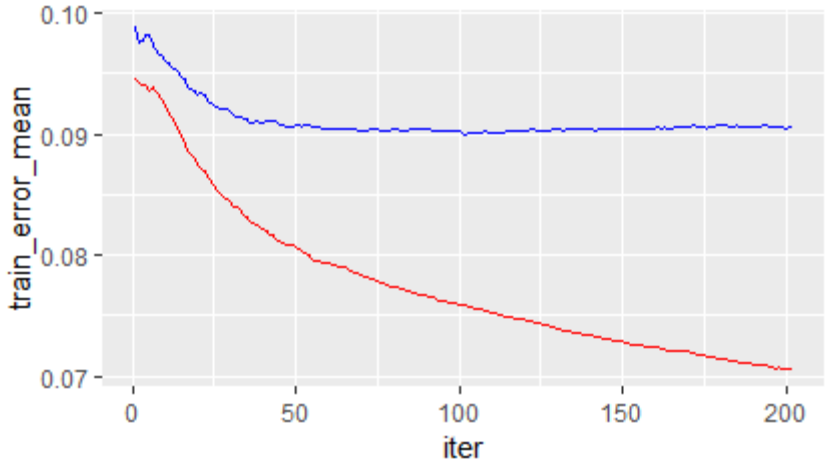
- Medical Information Mart for Intensive Care (MIMIC) is Electronic Medical Record data.
- Built as a postgres data base using a docker container to read and write to local.
- There are 58,976 ICU admissions (ADMISSIONS Table) and 651,047 diagnoses (DIAGNOSES_ICD9 Table).
- Eleven percent of admissions result in death.
- Goal: Predict death using nothing but the information available at admission and the diagnoses throughout.

Model Building: Setup

- Passing the categorical diagnoses into the model requires one hot encoding – a binary indicator for each possible diagnosis.
- Collapsed 650k diagnosis table to the patient-admission level and join to the admission, and used top 5 diagnoses per admission with a predictor for each diagnosis having at least 25 occurrences.
- Combined with observable characteristics at admission there are 1,232 predictors.

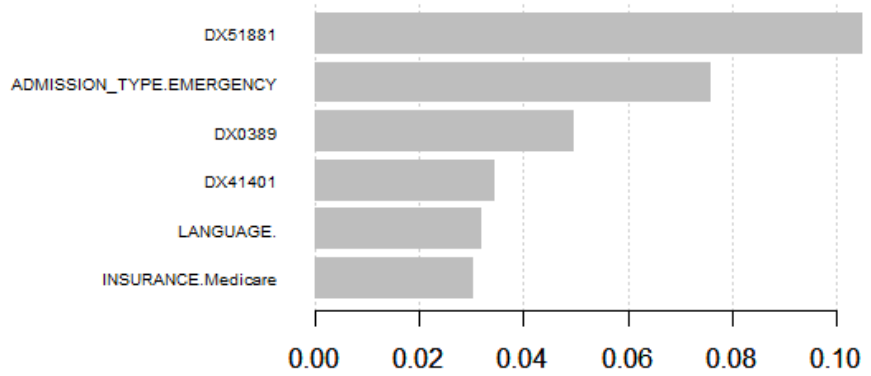
Model Building: Implementation

- Split data 90/10 into training and test sets.
- Used xgboost package in R.
- Ran 5-fold cross validation to select an appropriate number of trees without overfitting.



Model Building: Results

- Model had a CV test error of 9.0 percent with a slightly better 8.55 percent error on the 10 percent test sample that was held out entirely.
- The base rate of death was 11 percent.
- Variable importance mostly intuitive:
  - Respiratory Failure
  - Emergency
  - Unspecified Sepsis
  - Coronary Atherosclerosis of the Native Coronary Artery
  - No Language
  - Medicare

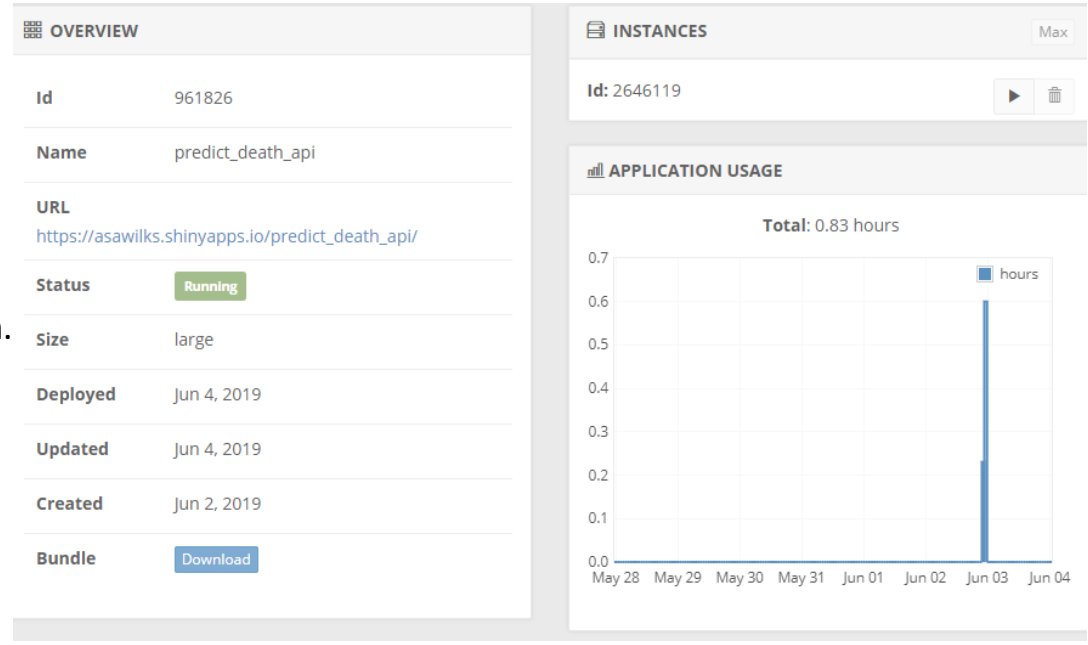


Model Deployment

- Wrote R function which allows few user inputs to be translated into a prediction.
  - Necessary to translate diagnosis code inputs into the “one hot” format necessary for prediction.

```
HAS_CHARTEVENTS_DATA=NA,  #1/0
ADMISSION_TYPE=NA,        #ELECTIVE / EMERGENCY / NEWBORN / URGENT
MARITAL_STATUS=NA,        #UNK / DIVORCED / LIFE.PARTNER / MARRIED / SEPARATED / SINGLE / WIDOWED
ADMISSION_LOCATION=NA, #UNK / EMERGENCY_ROOM /TRANSFER_WITHIN
INSURANCE=NA,                  #GOVERNMENT / MEDICAID / MEDICARE / PRIVATE / SELFPAY
ETHNICITY=NA,                  #WHITE / BLACK / ASIAN / LATINO
LANGUAGE=NA,                   #UNK / ENGLISH / SPANISH / KOREAN
RELIGION=NA,                    #UNK / CATHOLIC / JEWISH / MUSLIM
DX1-5=NA,                      #"DX___“
```

- Deployed to shinyapps.io with plummer
  - Accepts JSON of params and returns prediction.



<>