PREDICTING 30-DAY ALL-CAUSE READMISSION FROM HOSPITAL DISCHARGE SUMMARY

CYNTHIA PEDRASA

HOSPITAL READMISSIONS

- "Common, expensive and often preventable" occurrences in acute-care medicine
- Quality of care barometer – provide a basis to compare hospital performance measures
- A strategy to address this issue is to utilize risk stratification to better target patient engagement.

READMISSIONS: CLINICAL & FINANCIAL PROBLEM

2012

- CMS began penalizing hospitals for 30-day readmissions Oct. 1, 2012 at 1 percent, upping the penalty rate to 2 percent for fiscal year 2014

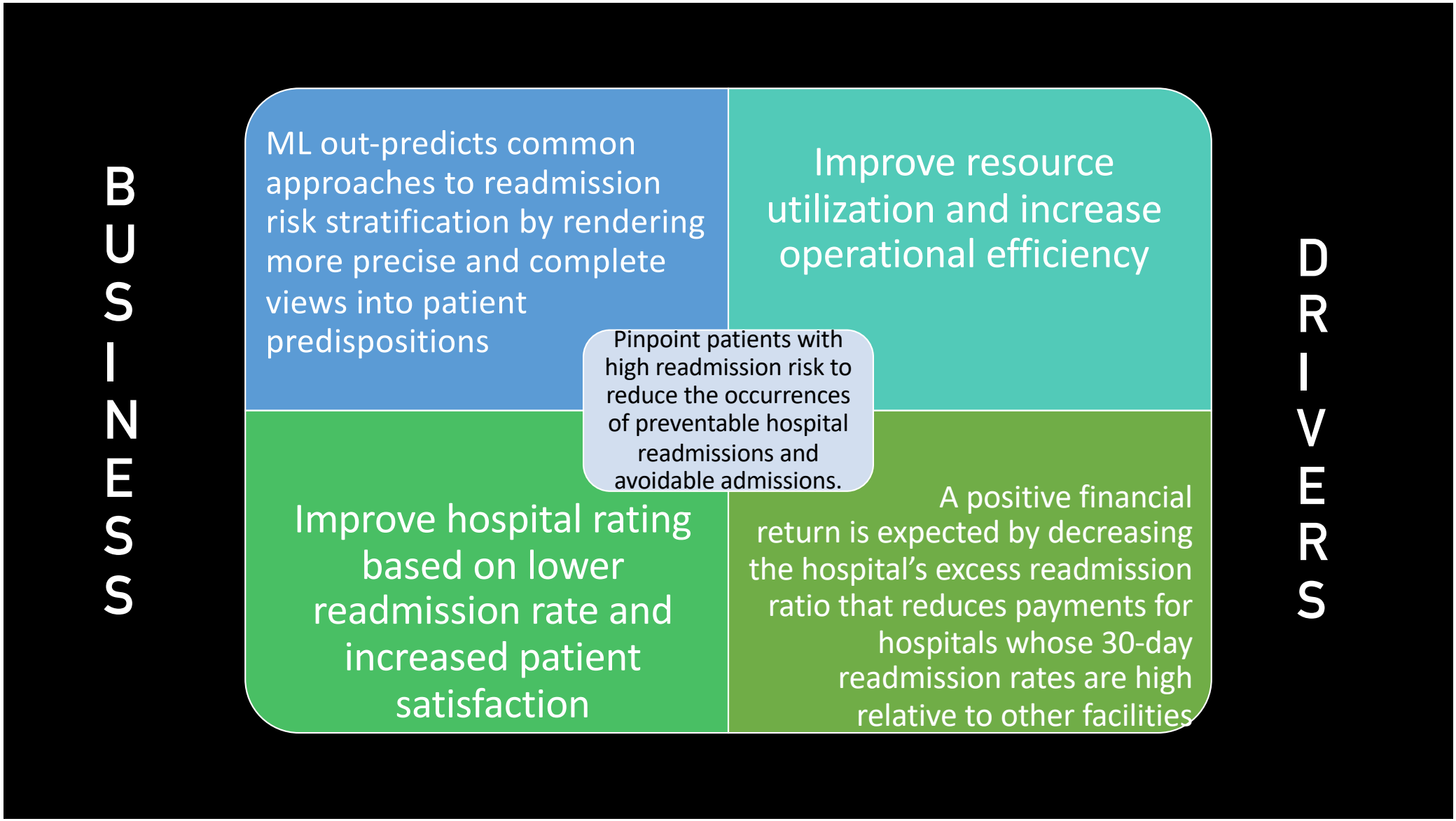
2020

- CMS will cut payments to the penalized hospitals by as much as 3 percent for each Medicare case during fiscal 2020, which runs Oct. 1 through September 2020

COST

- All-cause readmissions - The average cost of a readmission for any given cause is $11,200, with a 21.2 percent readmission rate.

https://www.beckershospitalreview.com/quality/6-stats-on-the-cost-of-readmission-for-cms-tracked-conditions.html
https://www.beckershospitalreview.com/finance/cms-penalizes-2-583-hospitals-for-high-readmissions-5-things-to-know.html

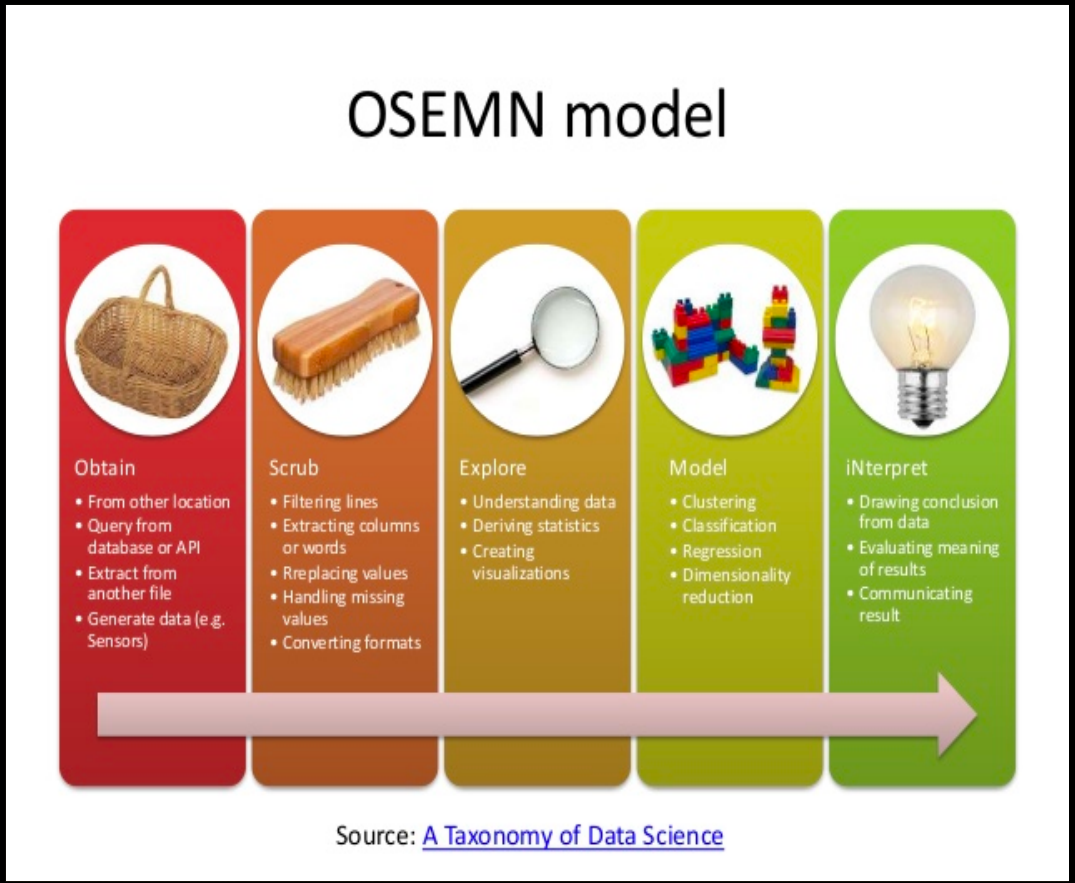

BUSINESS
ML out-predicts common approaches to readmission risk stratification by rendering more precise and complete views into patient predispositions
Improve resource utilization and increase operational efficiency
Pinpoint patients with high readmission risk to reduce the occurrences of preventable hospital readmissions and avoidable admissions.
Improve hospital rating based on lower readmission rate and increased patient satisfaction
A positive financial return is expected by decreasing the hospital's excess readmission ratio that reduces payments for hospitals whose 30-day readmission rates are high relative to other facilities
DRIVERS

PROBLEM

- How can we use unstructured data to predict the hospitalized patient's risk for 30-day readmissions?
- How can we risk stratify and rank patient/encounters with the highest probability of getting readmitted within 30 days?
- Which features are important predictors of 30-day all-cause readmission in patients?

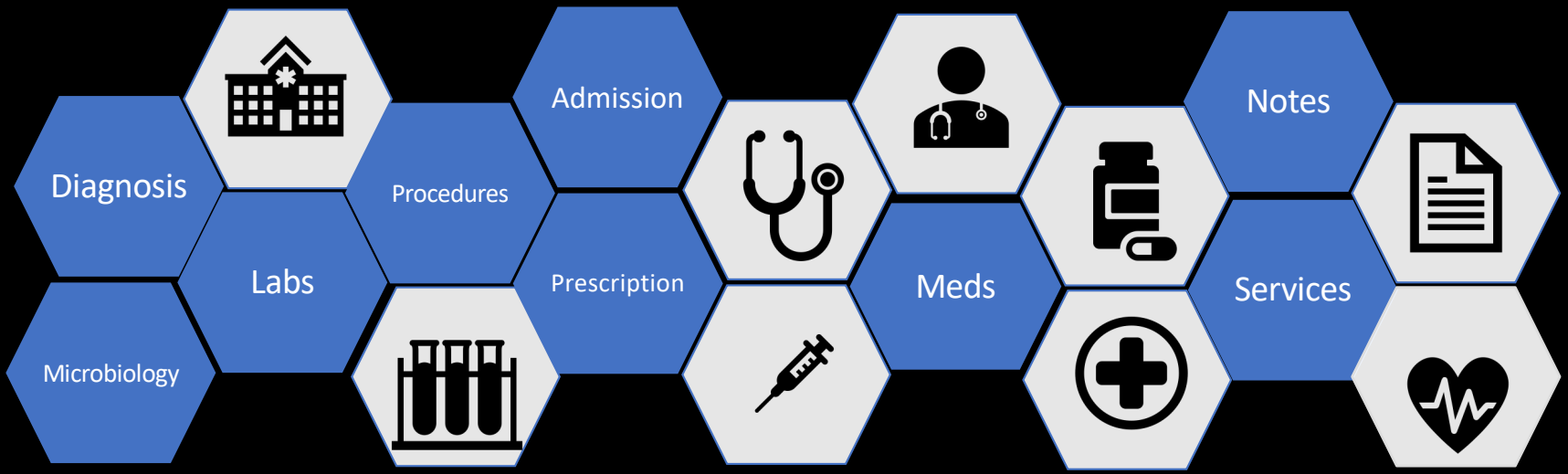
DATA SCIENCE PROCESS

- **Obtain**-Requirements and information gathering on the problem.
- **Scrub-** Pre-processing our data (removing nulls, outliers, normalization, feature selection)
- **Explore**-Understand the cohort characteristics and impactful predictors
- **Model**-Build and tune the model
- **iNterpret** - and communicate results to stakeholders.

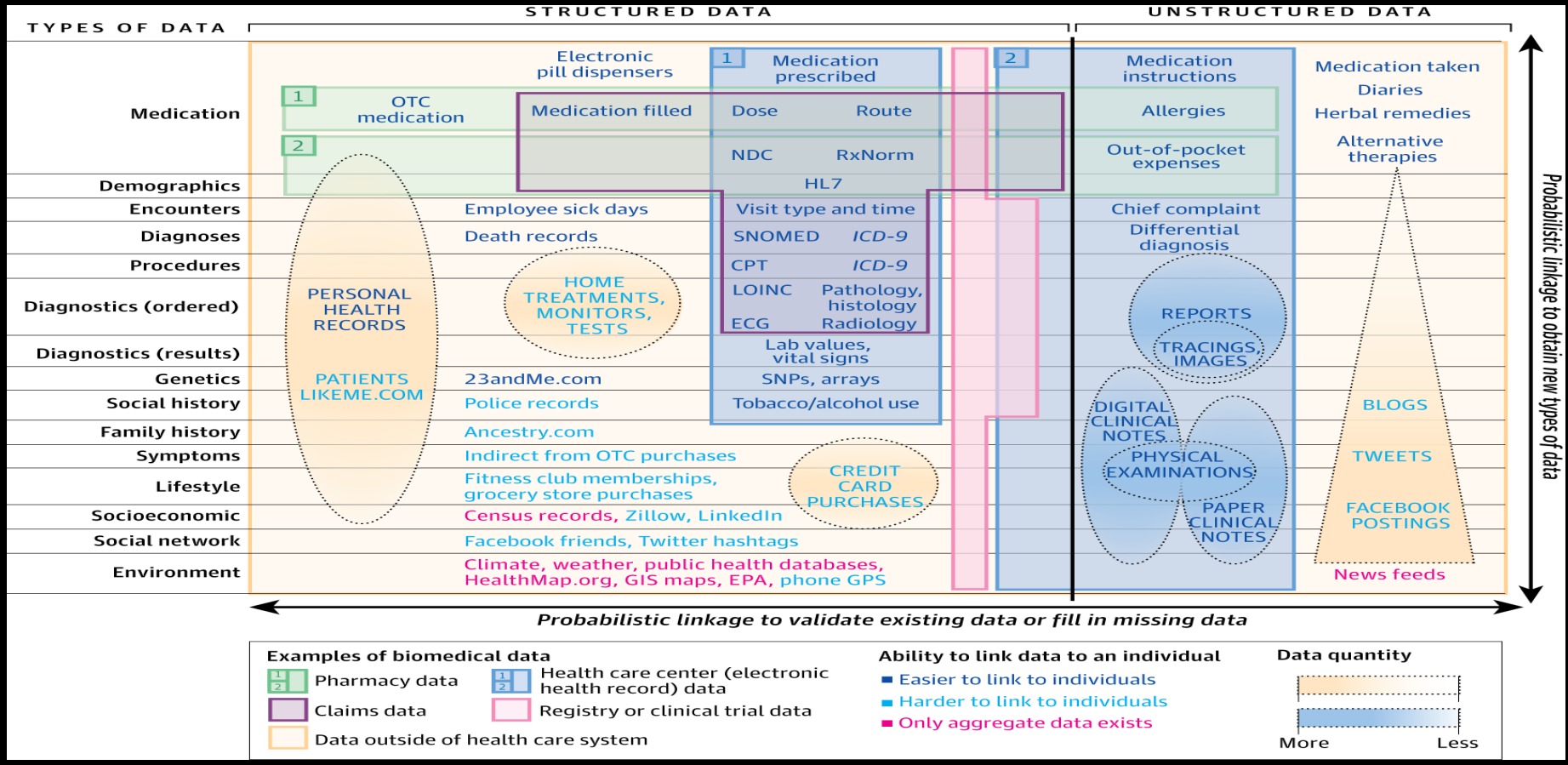


OBTAIN THE DATA: MIMIC III DATA SET

Deindentified health-related dataset represents 11 years (2001-2012) of clinical care at Beth Israel Deaconess Medical Center. It contains data regarding the care of over 40,000 real patients, and as such requires credentialing before access. MIMIC-III is a relational database consisting of 26 tables.

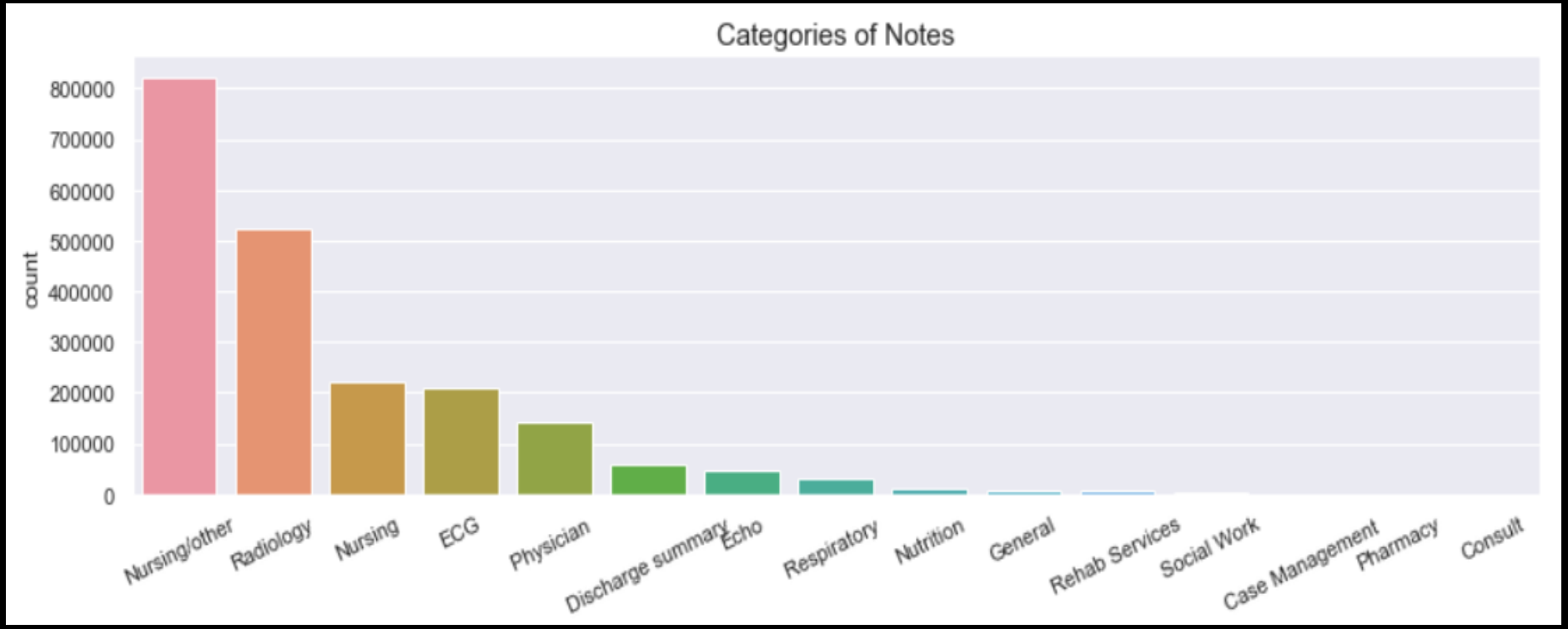


UNTAPPED UNSTRUCTURED DATA

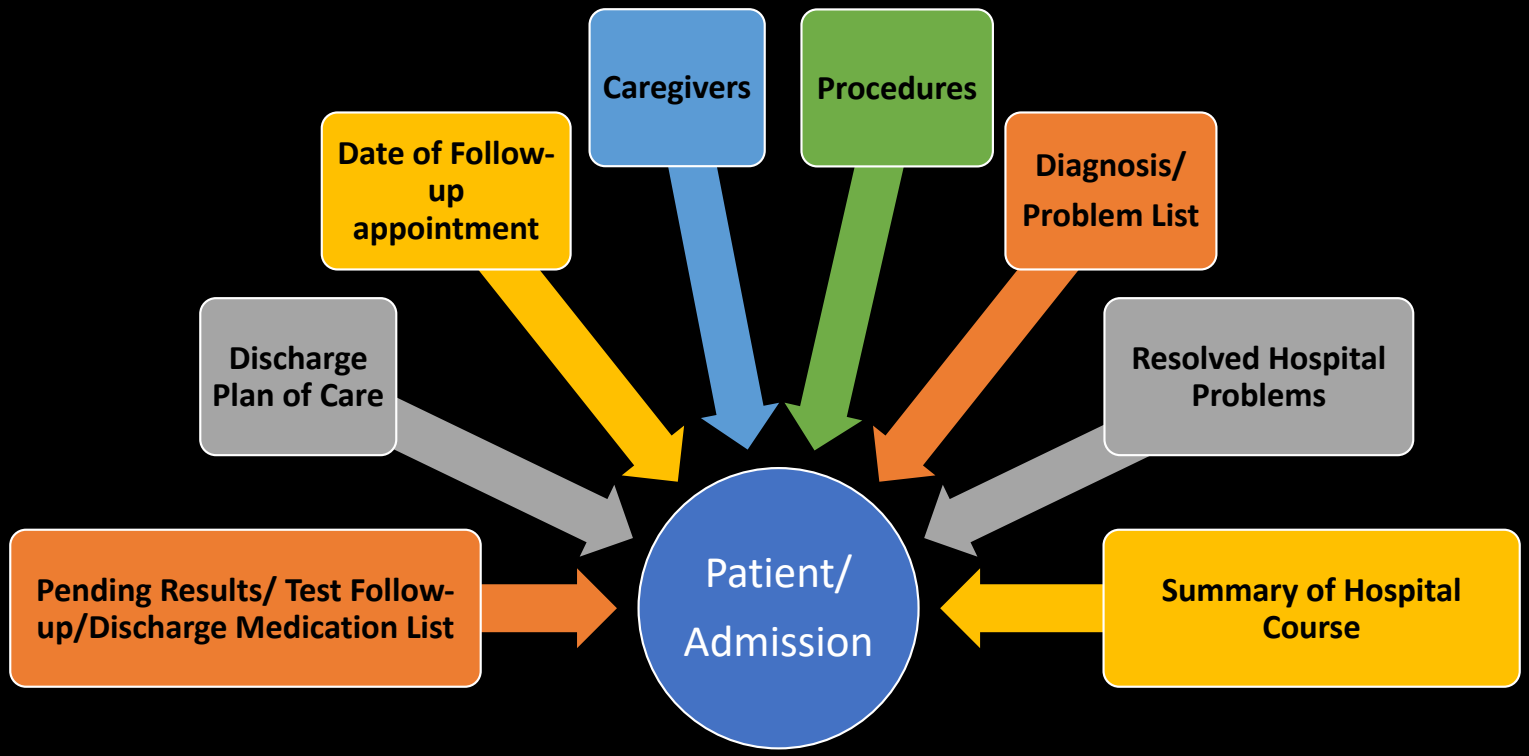


Source: Weber GM, et al., Finding the missing link for big biomedical data. JAMA 2014; 311(24):

UNSTRUCTURED CLINICAL NOTES



DISCHARGE SUMMARY COMPONENTS



Query the Admissions table

Convert strings to dates

Get the next unplanned admission

Calculate days until next admission

Filter out elective encounters

Query the Notes table

Filter to Discharge Summaries

Filter out Notes marked in error

Merge Notes & Admissions

Get one summary per admission

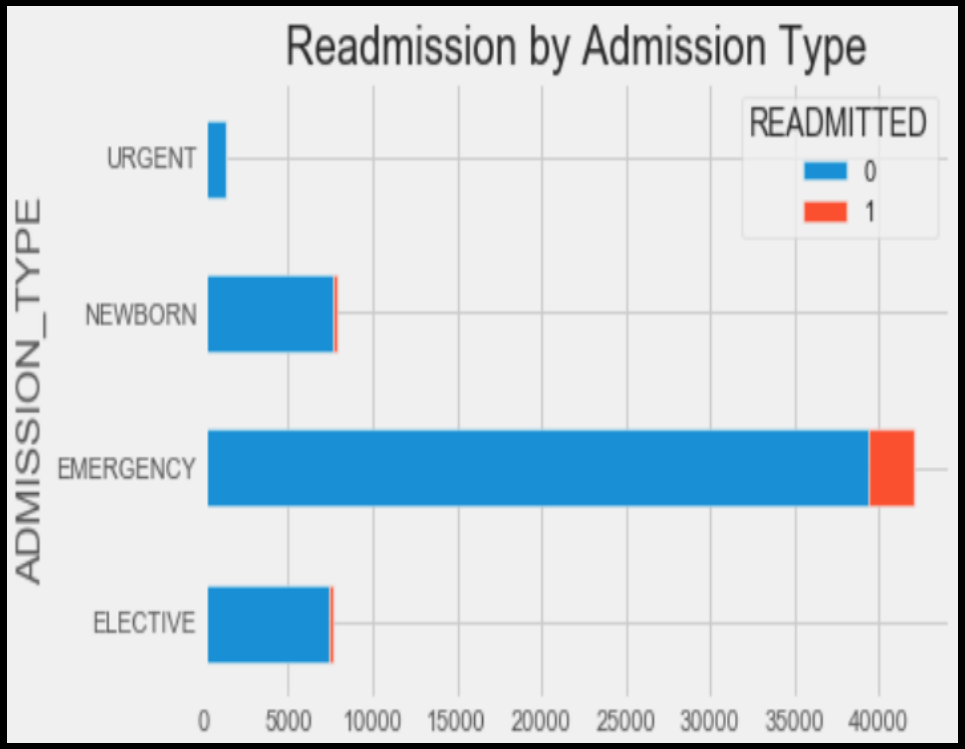
Remove Newborn encounters

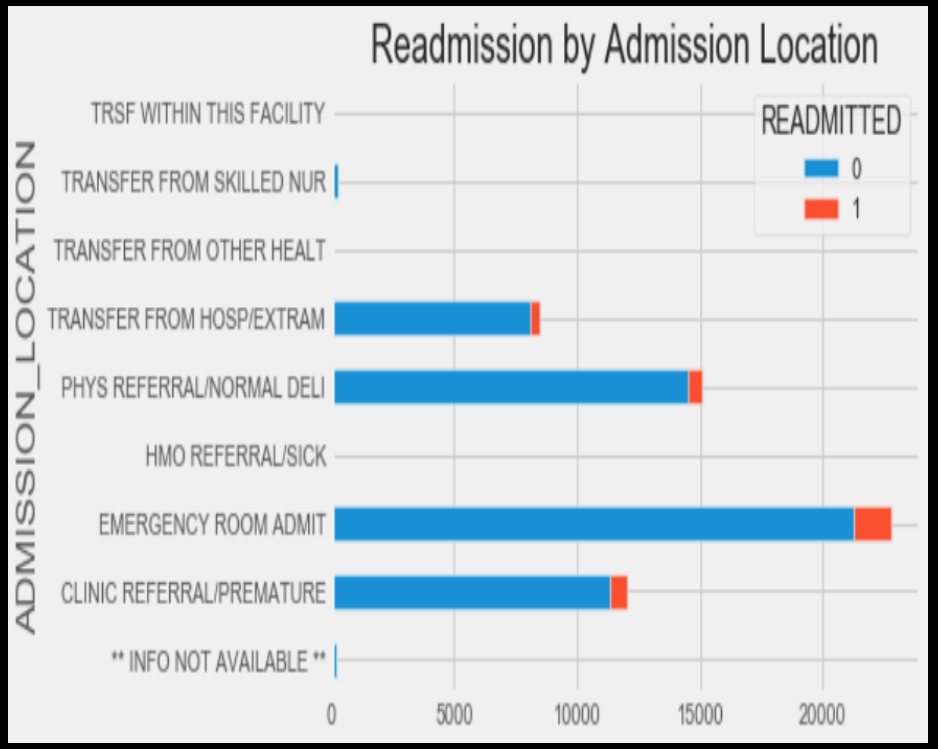
Create the class label

Create the training & test sets

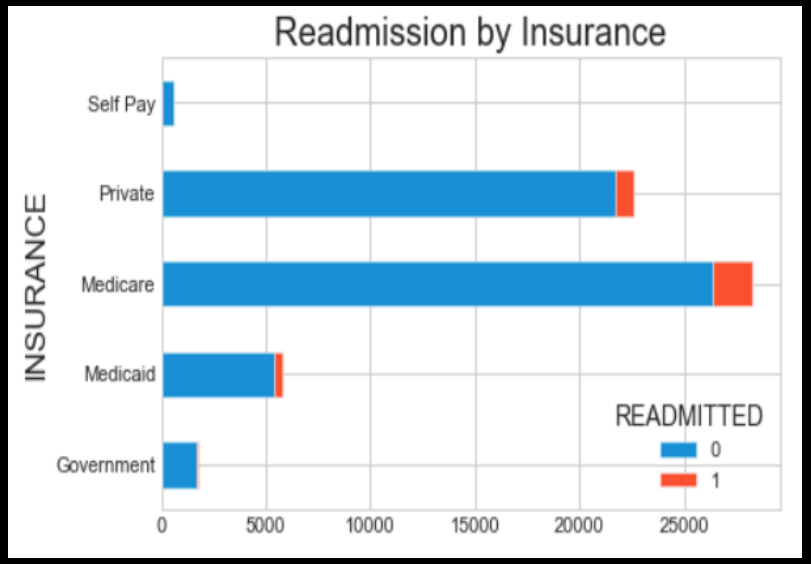
Scrub the Data

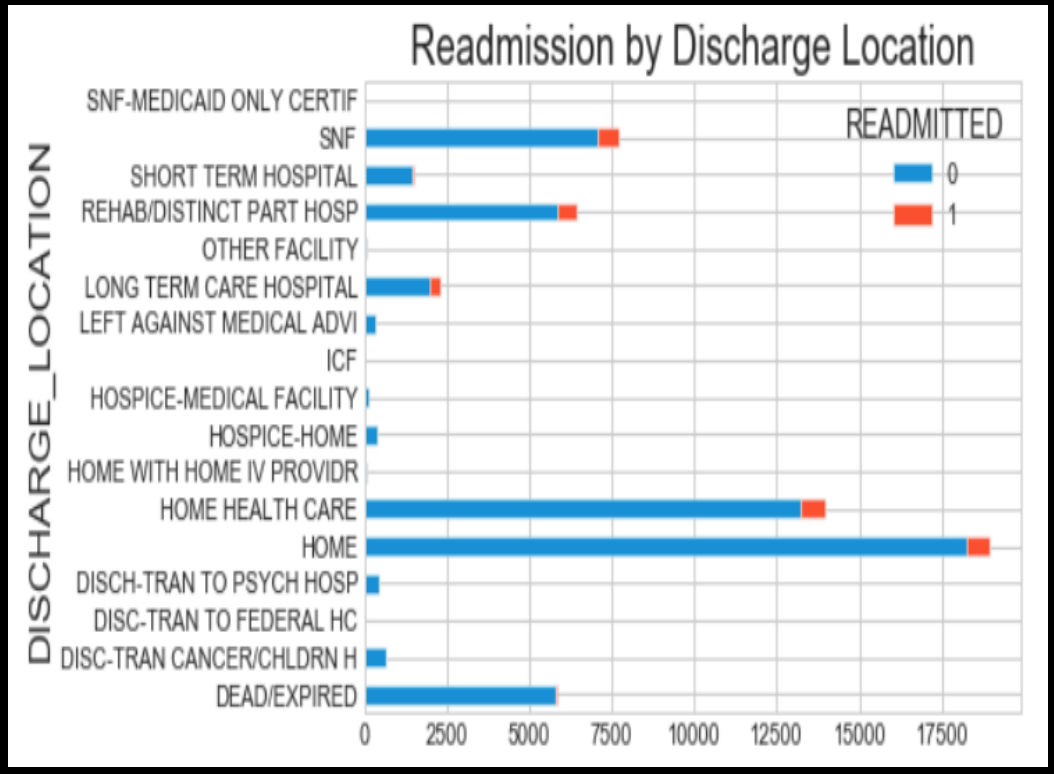
DATA EXPLORATION



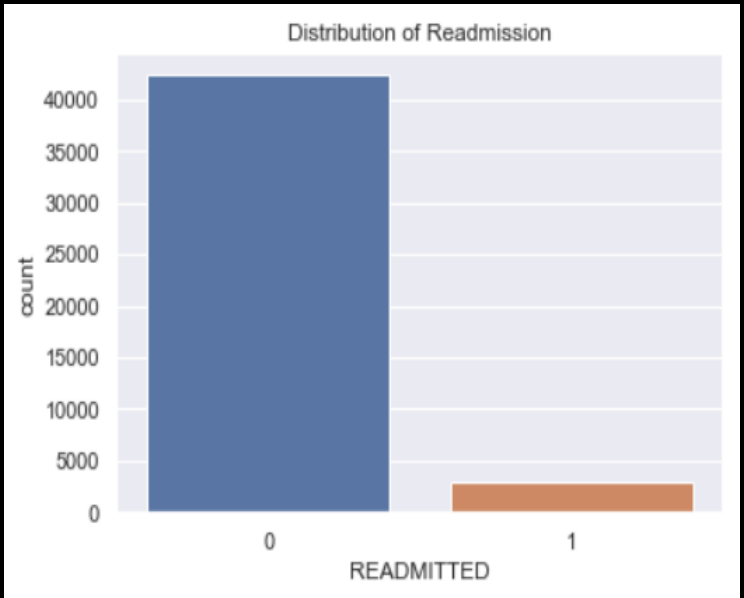


DATA EXPLORATION





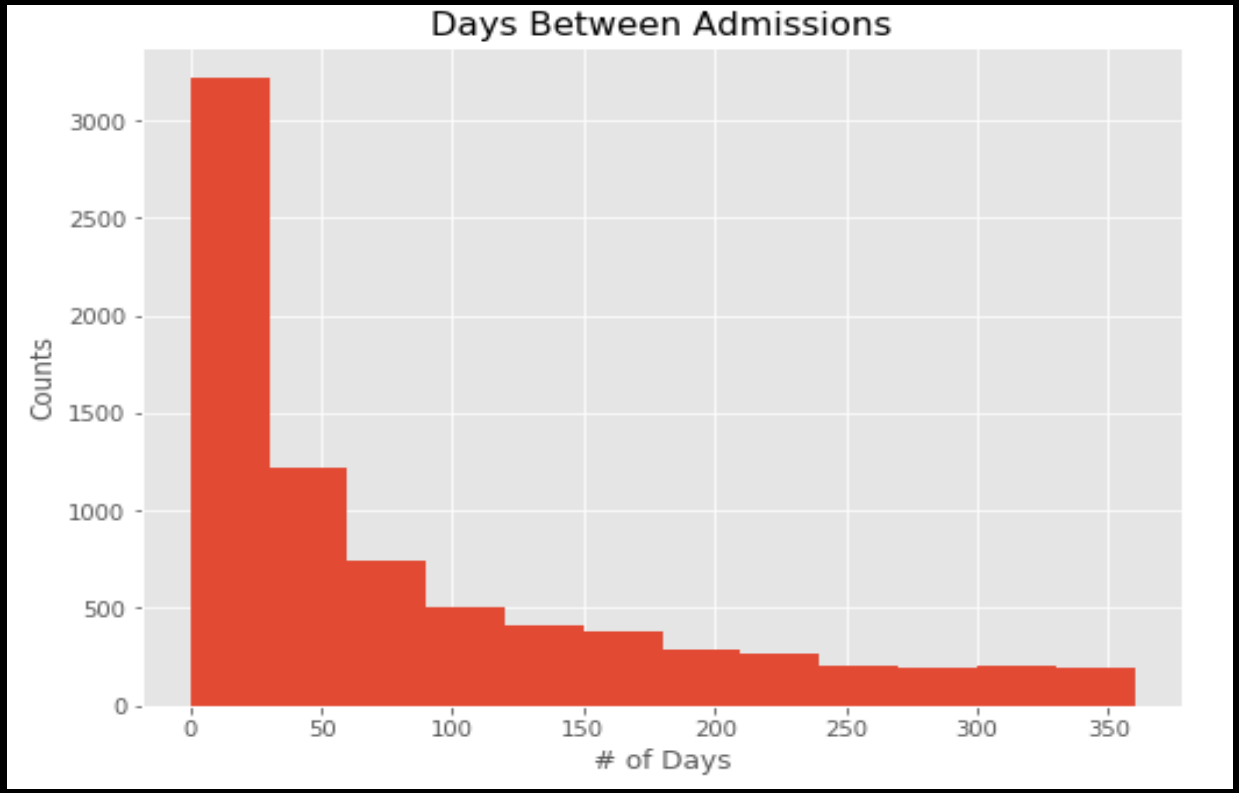
CLASS LABEL



Imbalanced Class

- Readmitted (minority) class is important
- Resample the training set
- Repeated Stratified Kfold Cross validation
- Resample with different ratios
- Use Tree-Based Algorithms
- Ensemble different resampled datasets
- Use the right evaluation metrics

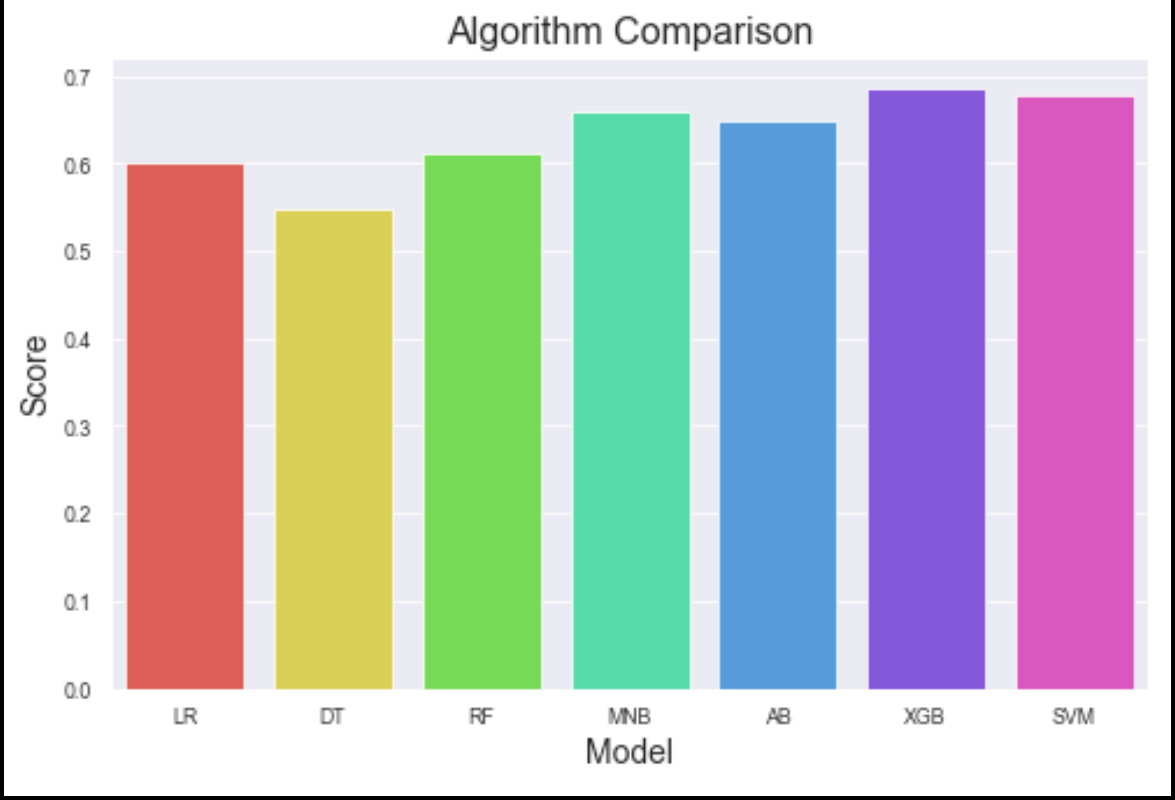
ALL CAUSE READMISSIONS - # DAYS BETWEEN



TEXT PROCESSING

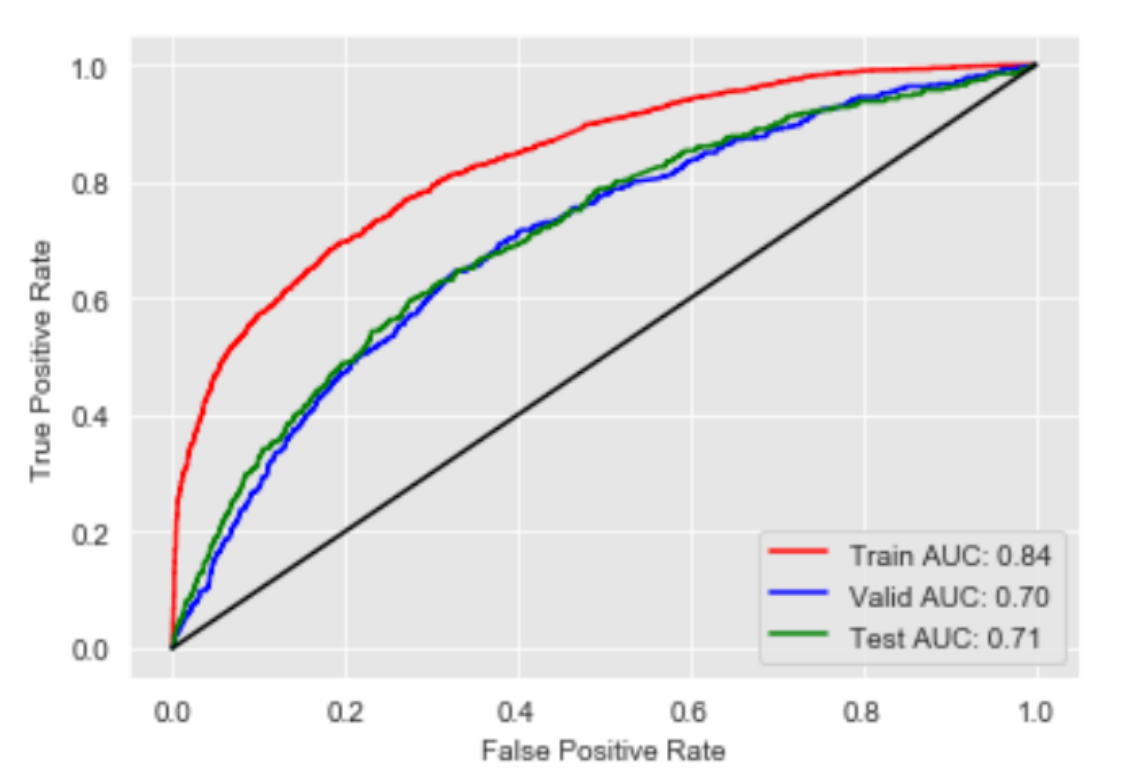
- Text cleaning is task specific – medical abbreviations
- Split discharge summary into words/tokens
- Remove punctuation
- Remove non-alphabetic characters
- Remove new line/carriage return/stop words
- Bag-of-Words Model (BoW)
- Count Vectorizer & TF-IDF

MODEL SELECTION



Model	ROCAUC
LR	0.600403
DT	0.548500
RF	0.612230
MNB	0.660130
AB	0.650057
XGB	**0.687193**
SVM	0.678143

TRAIN, VALIDATION, TEST SCORES



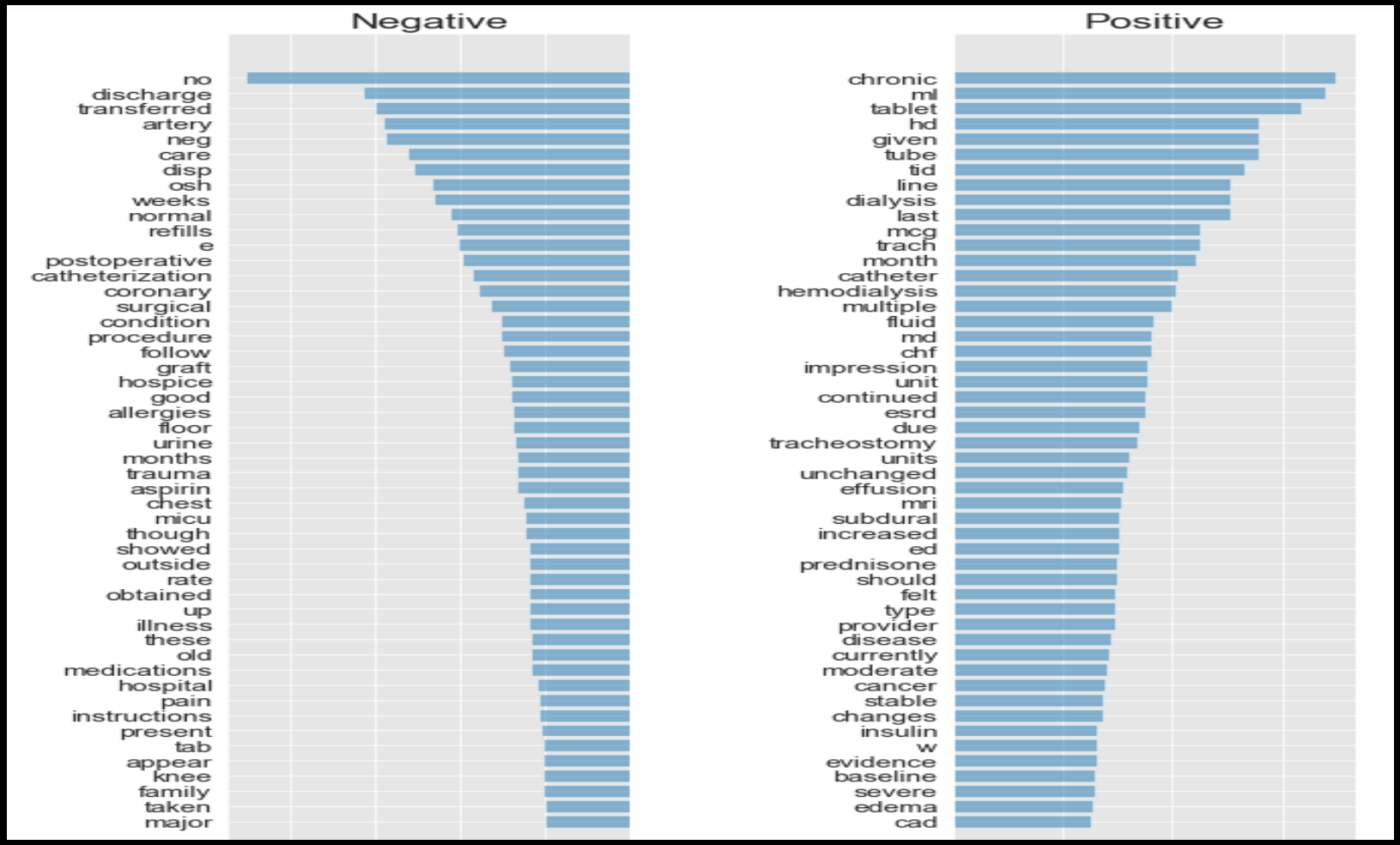
Train AUC:0.837
Valid AUC:0.700
Test AUC:0.708
Train accuracy:0.751
Valid accuracy:0.690
Test accuracy:0.680
Train recall:0.691
Valid recall:0.610
Test recall:0.631
Train precision:0.785
Valid precision:0.121
Test precision:0.111
Train specificity:0.811
Valid specificity:0.696
Test specificity:0.683

WORD CLOUD

Readmitted

Not Readmitted

FEATURE IMPORTANCE



PERFORMANCE METRICS

- Confusion Matrix - a table showing correct predictions and types of incorrect predictions.
- Precision - proportion of + identifications that are correct (TP/(TP+FP))
- Recall - proportion of actual positives identified correctly (TP/(TP+FN))
- Receiver operating characteristics curve - diagnostic ability of a binary classifier

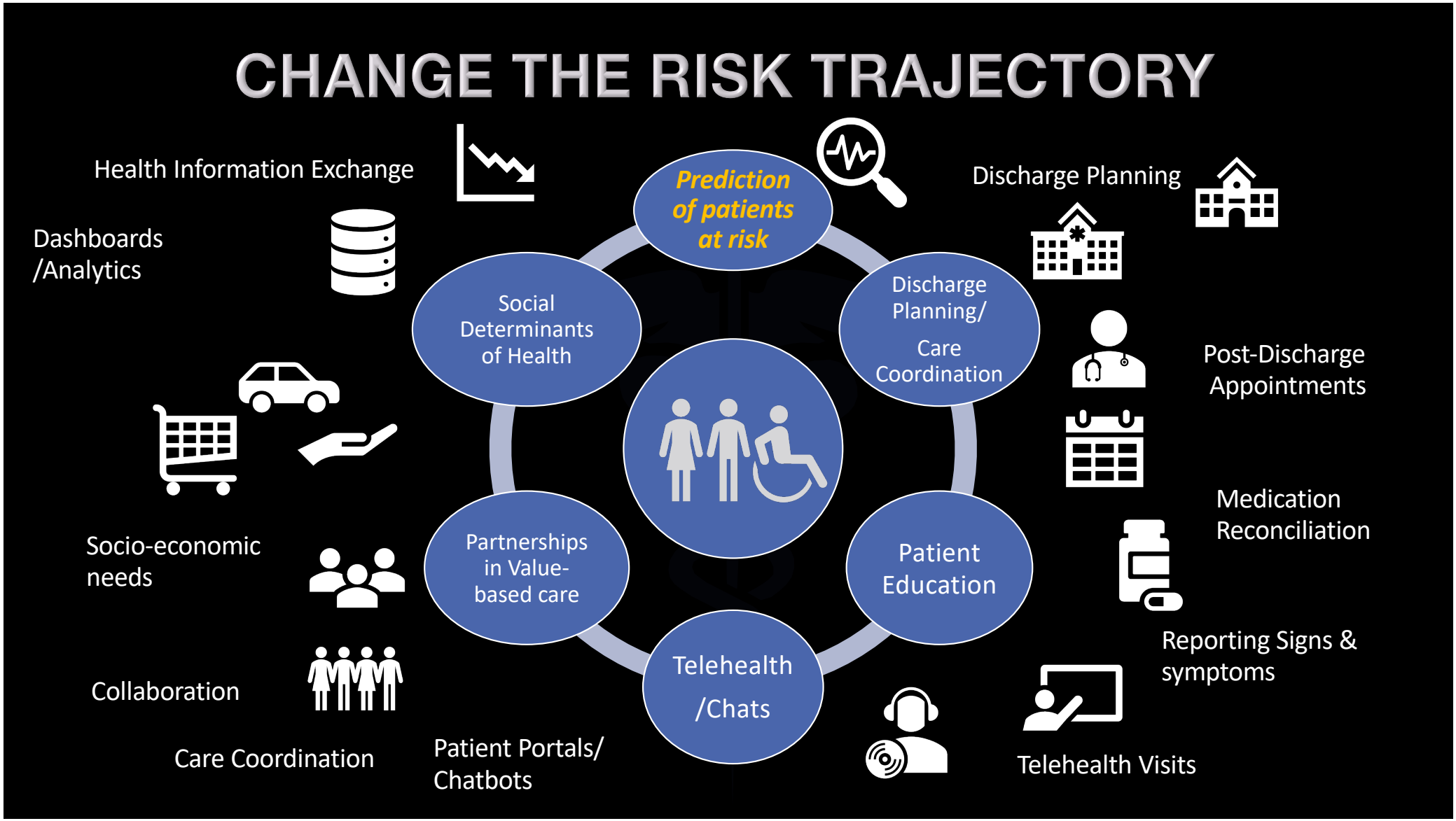
INTERPRETATION

- XGBoost Classifier was selected as the best model to predict readmissions from discharge notes.
- BoW with XGB model yields slightly better result than TF-IDF and the complex Neural Network model.
- Increasing n-gram range did not improve scores for BoW method.
- Reduce overfitting - early stop and regularization

Algorithm	Score
Basic XGB Initial AUC BoW	68.72
Tuned XGB BoW	**70.8**
XGBt BoW bi-gram	70.2
XGBt BoW tri-gram	69.70
XGBt TF-IDF tri-gram	68.06
Neural Networks	66.90

FUTURE WORK

- Compared to random predictions, results from our predictive model (AUC=.71) is a good baseline for further improving our model.
- Explore other unstructured notes or combine with structured clinical information to strengthen predictive scores.
- Feature engineering, ensemble of models and parameter tuning of the model will help the adoption of the model as a clinical decision system for evaluating readmission


CHANGE THE RISK TRAJECTORY
Health Information Exchange
Dashboards /Analytics
Socio-economic needs
Collaboration
Care Coordination
Patient Portals/ Chatbots
Telehealth Visits
Reporting Signs & symptoms
Medication Reconciliation
Post-Discharge Appointments
Discharge Planning
Prediction of patients at risk
Discharge Planning/ Care Coordination
Patient Education
Telehealth /Chats
Partnerships in Value-based care
Social Determinants of Health

THANK YOU!

- Keep in touch!