Predicting 30-day hospital readmissions

By Zhanna Hakhverdyan, PhD

03.09.2020

Q: What's the problem?
A: Unplanned hospital readmissions are expensive

- Medicare spends $26B/ year on readmissions
  - Of which $17B are spent on avoidable readmissions
- In 2017 >2500 hospitals paid $564M in penalties for excessive 30-day readmissions
- Patient experience and outcome is worse for readmitted patients

Care management programs to the rescue

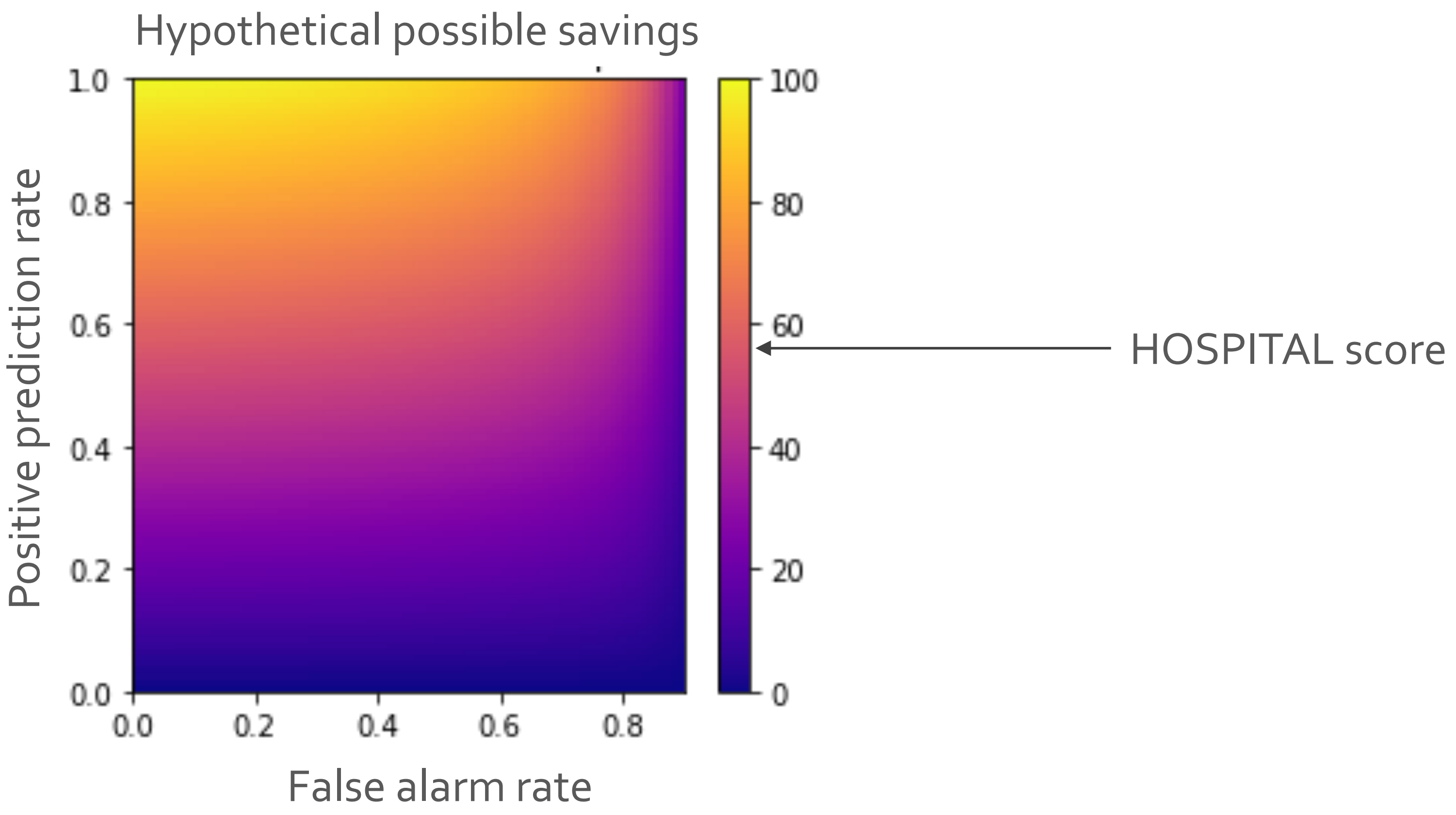
- Hospitals enroll patients at high risk of readmission into Care Management Programs
- Care management programs cost money and time
  - Hospital saves $s on readmission costs of correctly identified patients
  - Hospital wastes $s on care management programs targeted at incorrectly identified patients
- How do hospitals identify patients at high risk of readmission to optimally allocate limited resources?

Existing solutions vs proposed solution

- HOSPITAL score is currently used by many hospitals to score readmission risk
  - Uses 7 predictive variables for risk score assignment

- Proposed solution – machine learning model with a wider array of predictive variables

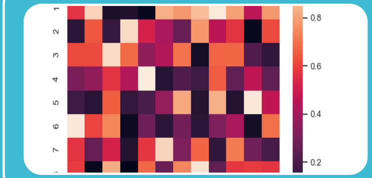
How much money should the proposed model save to be useful?



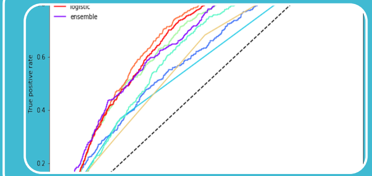
- The HOSPITAL score can save ~57% of unplanned 30-day readmission costs
- In order to be useful the machine learning model has to save more

Analysis workflow

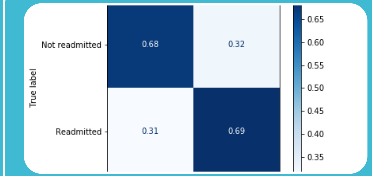
1. Data wrangling



2. Exploratory data analysis



3. Feature selection & machine learning



4. Final evaluation & interpretation

5. Hospital savings analysis

1. Data wrangling

Dataset used

MIMIC-III : electronic medical record (EMR) data

Tables	Variables	# of features
ADMISSIONS PATIENTS	demographic Hospital stay	22
SERVICES CPT_EVENTS	services procedures	36
LAB_EVENTS	Lab test results	69
DRG_CODES	diagnosis	187
Total		**314**

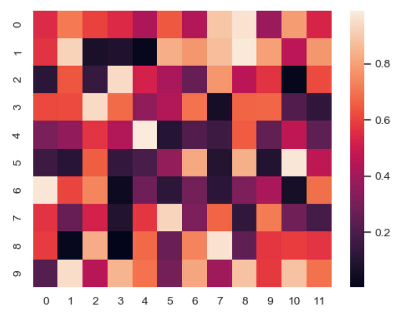
Python libraries and packages used:

Pandas, NumPy, NLTK, TF-IDF vectorizer

Feature engineering

Tables	Operations performed
ADMISSIONS PATIENTS	• Categories were grouped when possible, e.g. ethnicity • Categories with < 0.5% examples were grouped with most frequent category • One-hot-encoding of categorical variables • Calculate prior admissions and length of stay
SERVICES CPT_EVENTS	• Calculate the number and type of services • Select top 11 services (98%) • Calculate the number and type of procedures • Pick top 23 procedures (99.8%)
LAB_EVENTS	• 3 features per lab test per admission • Average, variance, count
DRG_CODES	• Concatenate diagnosis descriptions • Select words that occur in <50% of fields • Encode description field with top 200 words

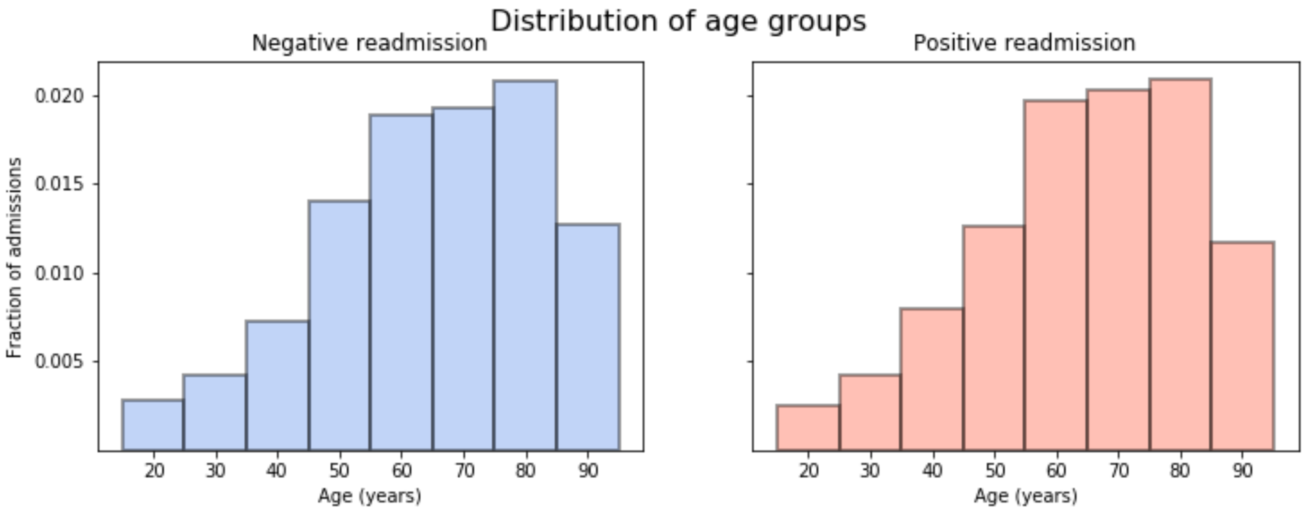
- A number of columns were used as is
- Median of the column was imputed for missing values
- Columns with > 20% missing values were discarded



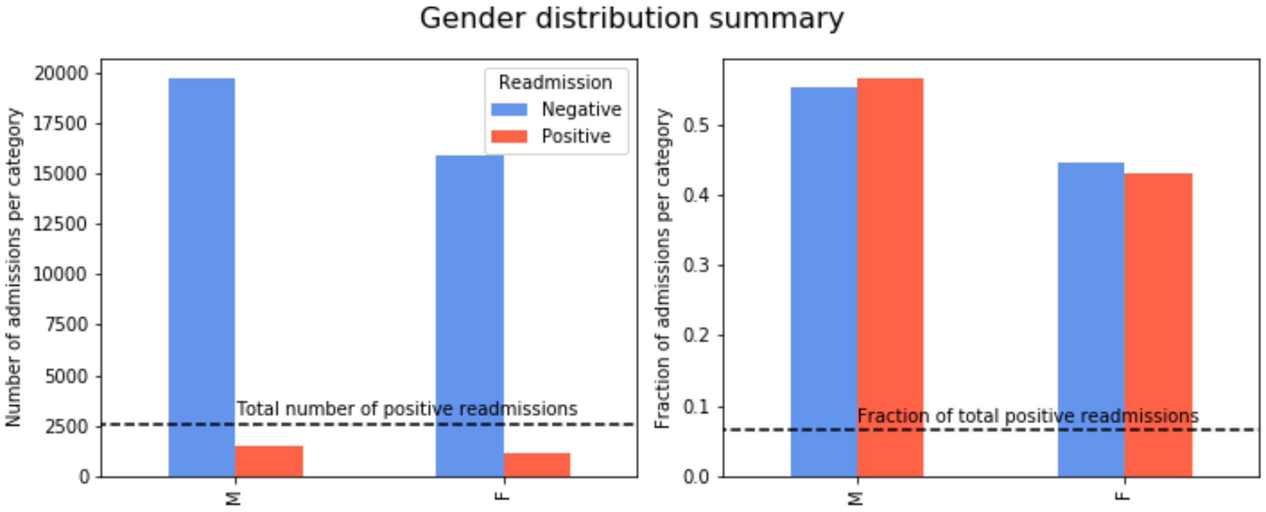
2. Exploratory data analysis

- Examine variable relationship to the output label – positive or negative for 30 day readmission
- Examine intervariable correlations and relationships
- Libraries and packages used: Matplotlib, Seaborn, SciPy(stats)

Age and gender distribution of readmitted patients



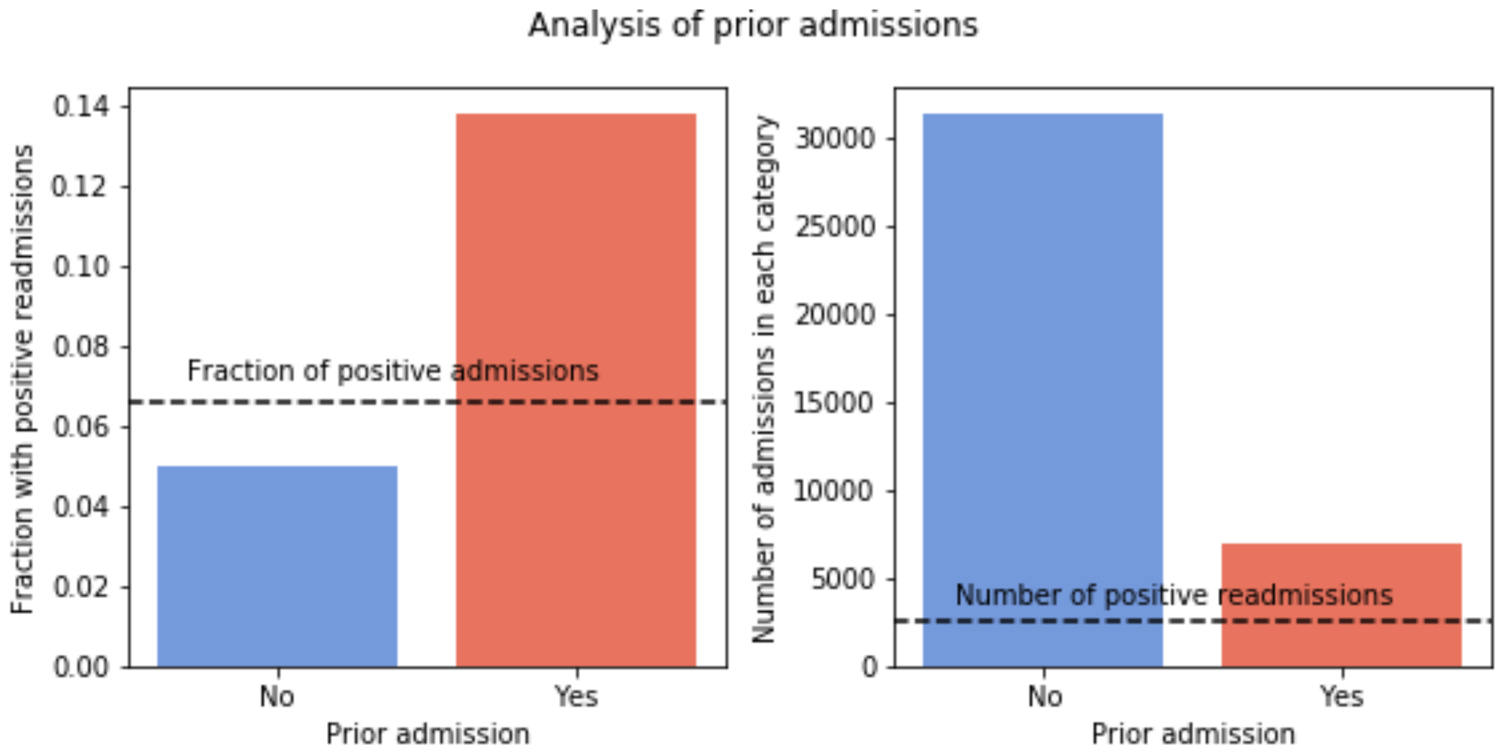
Patient age is similar in readmitted and non-readmitted groups



More men admitted to hospital overall

No difference in gender proportions in positive readmission group

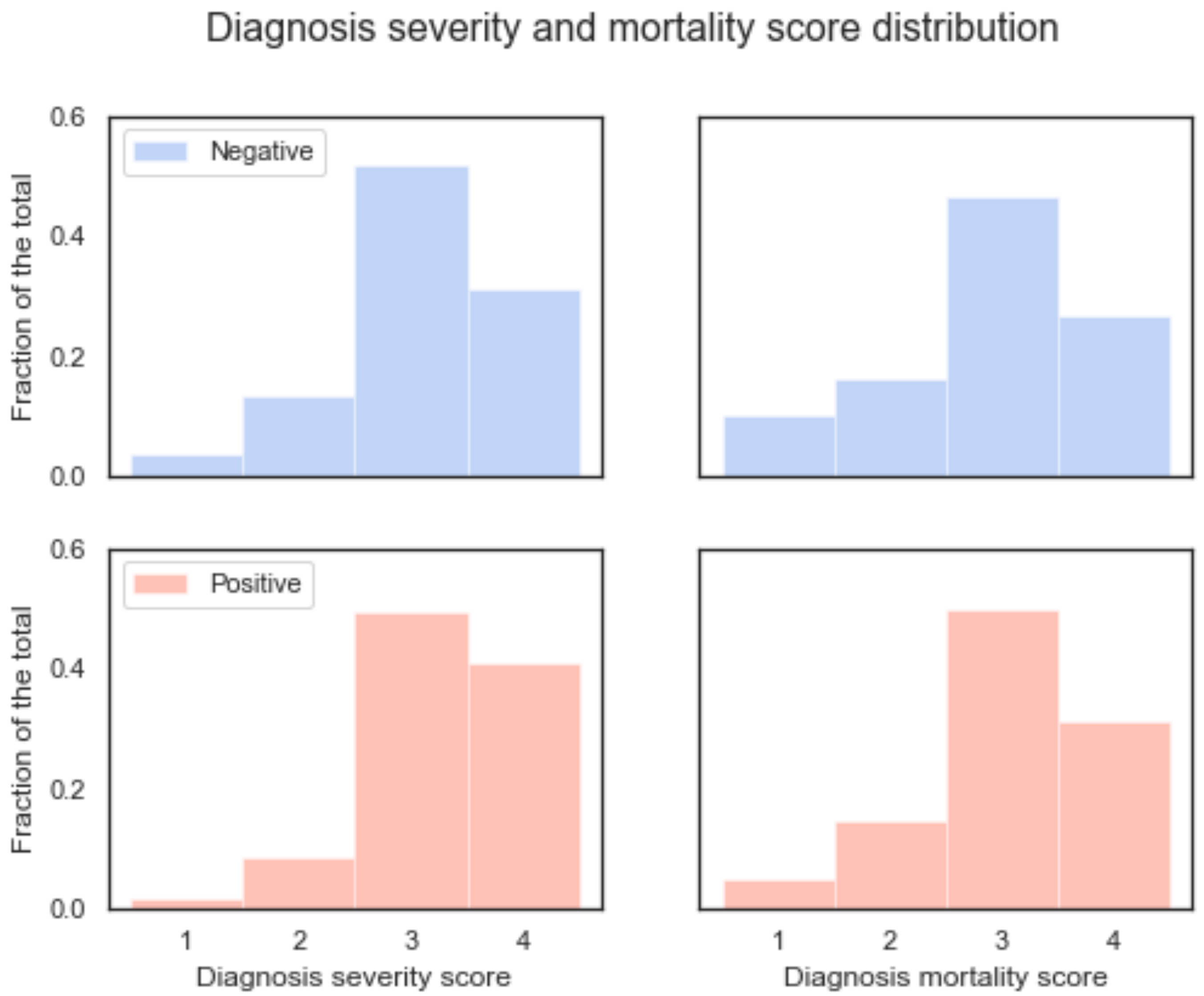
Positive readmissions are likely to have a prior admission within a year



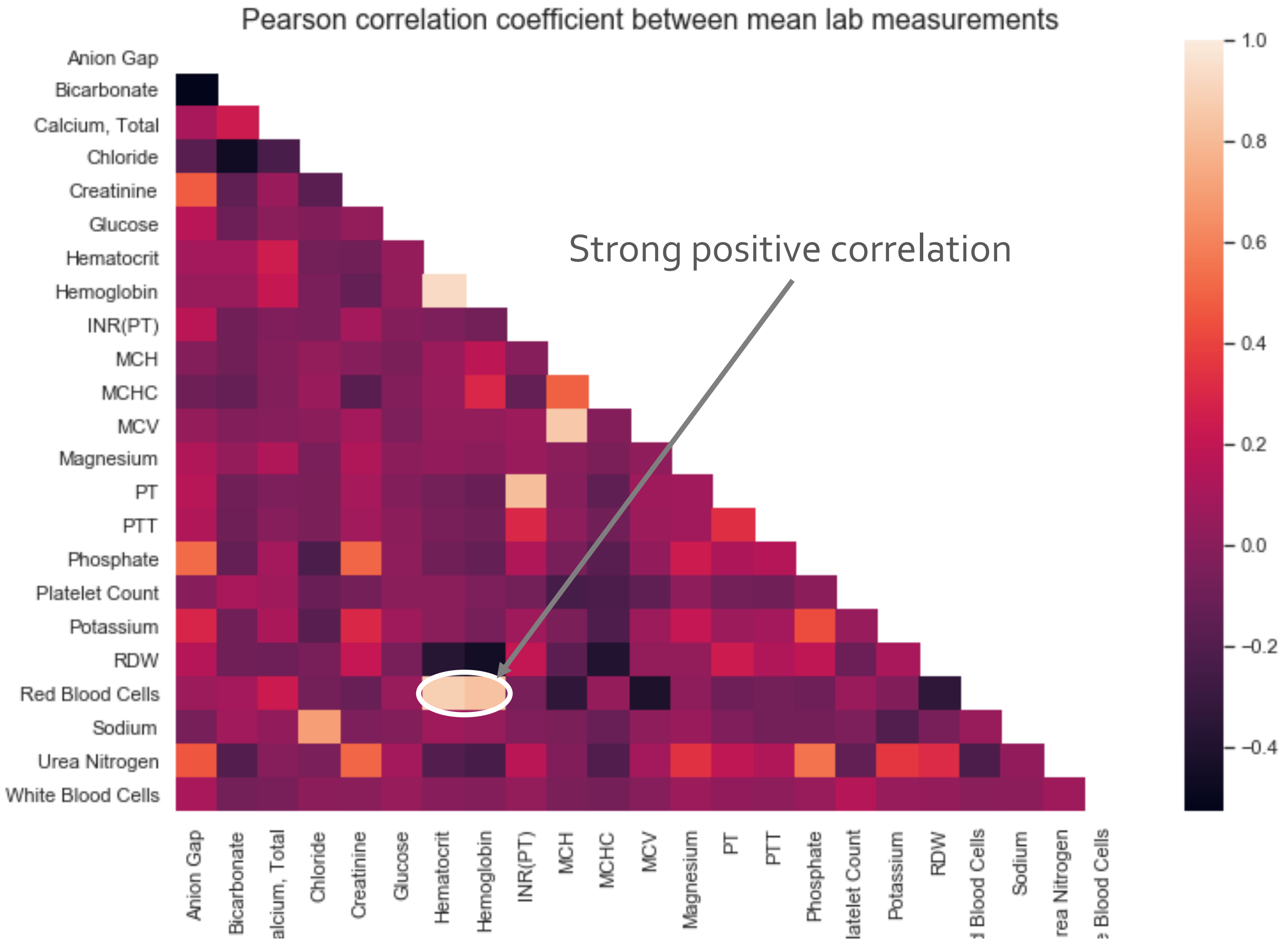
Readmission group	Chi square statistic	P value
Negative	47	E-12
Positive	657	E-145

- Patients in the positive readmission group have significantly more prior admissions
- Patients in the negative readmission group have significantly less prior admissions

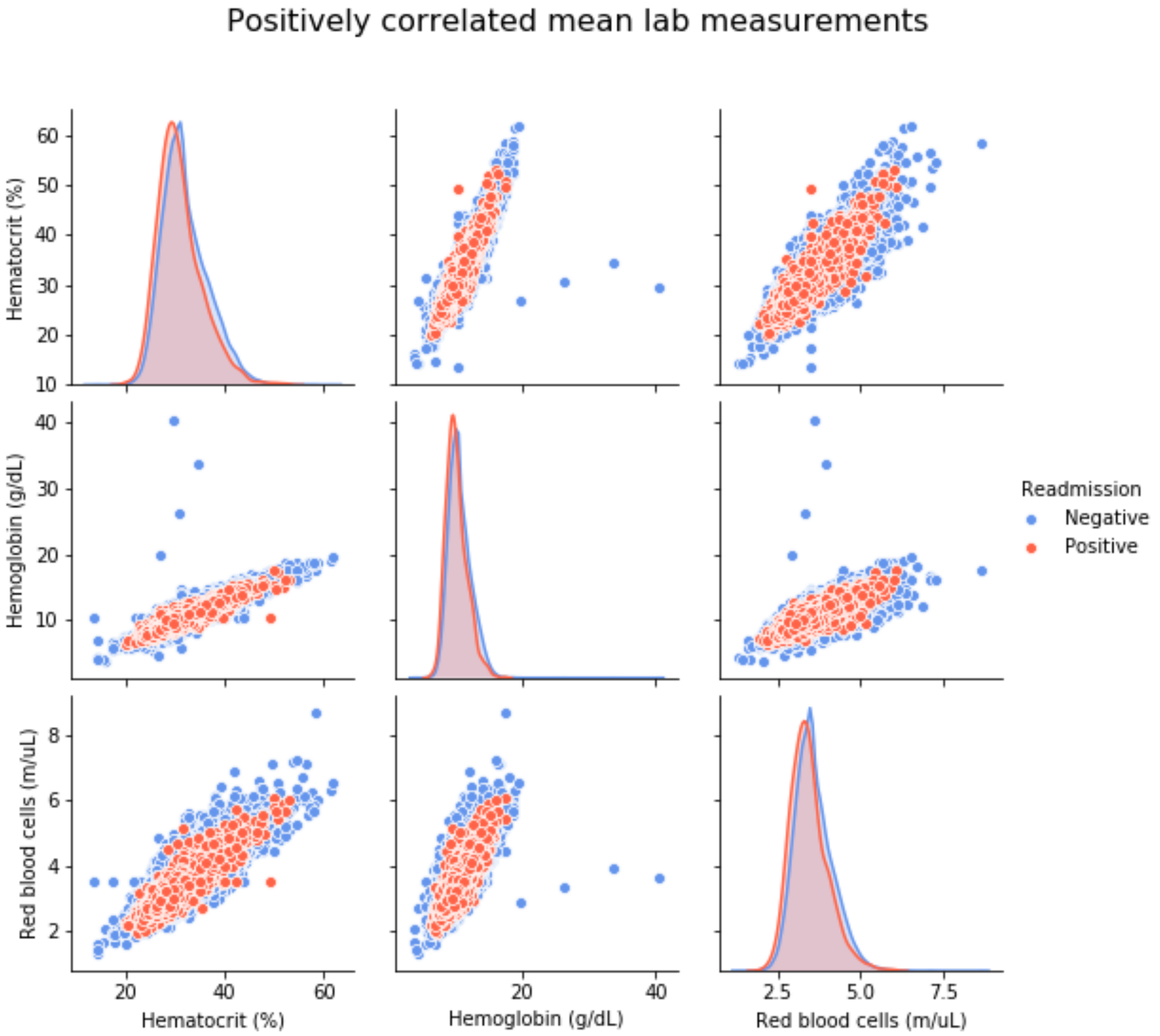
Diagnosis severity and mortality scores are higher for positive readmissions



Pearson correlation coefficient identifies linearly correlated columns

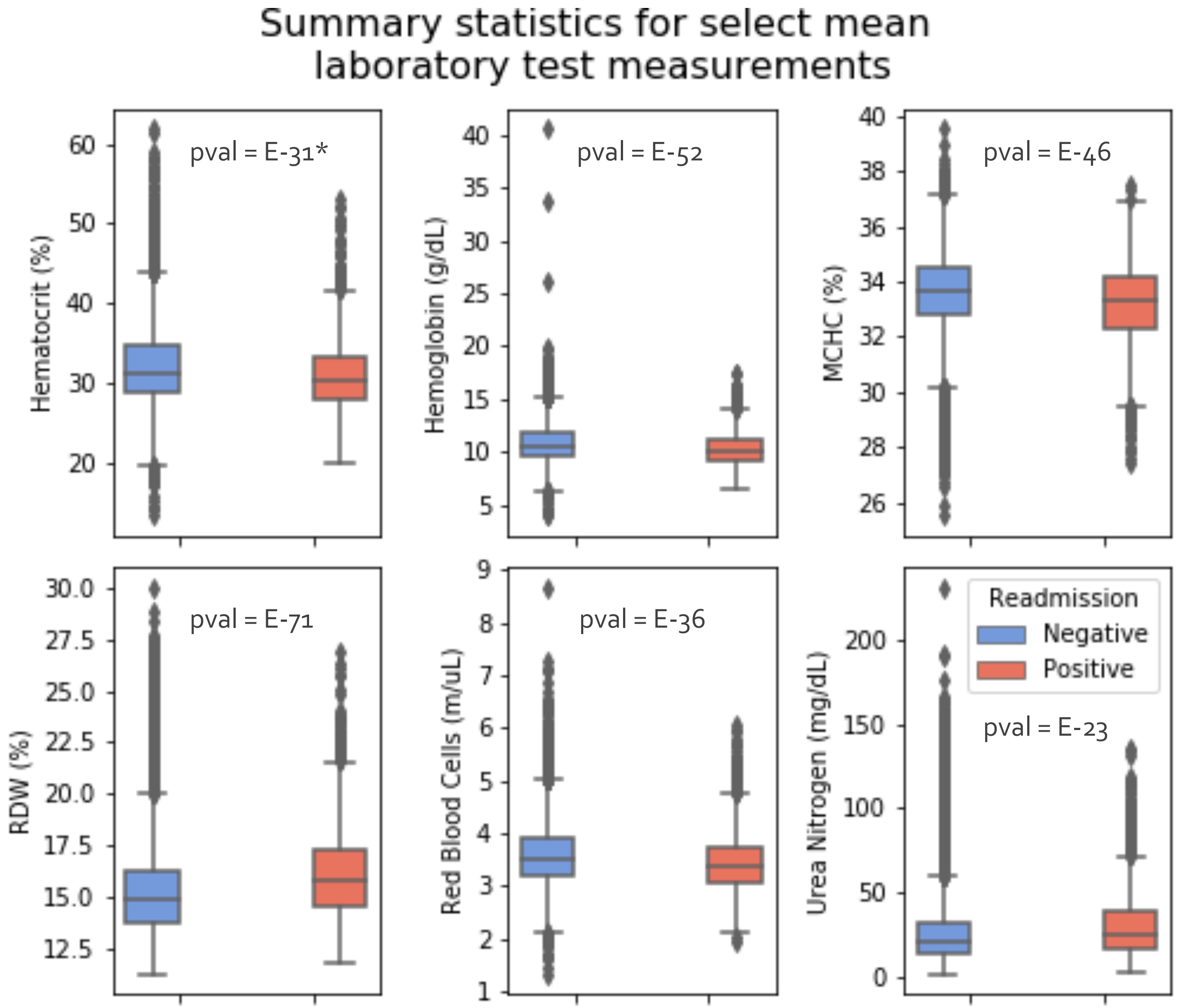


Inspection of positively correlated examples



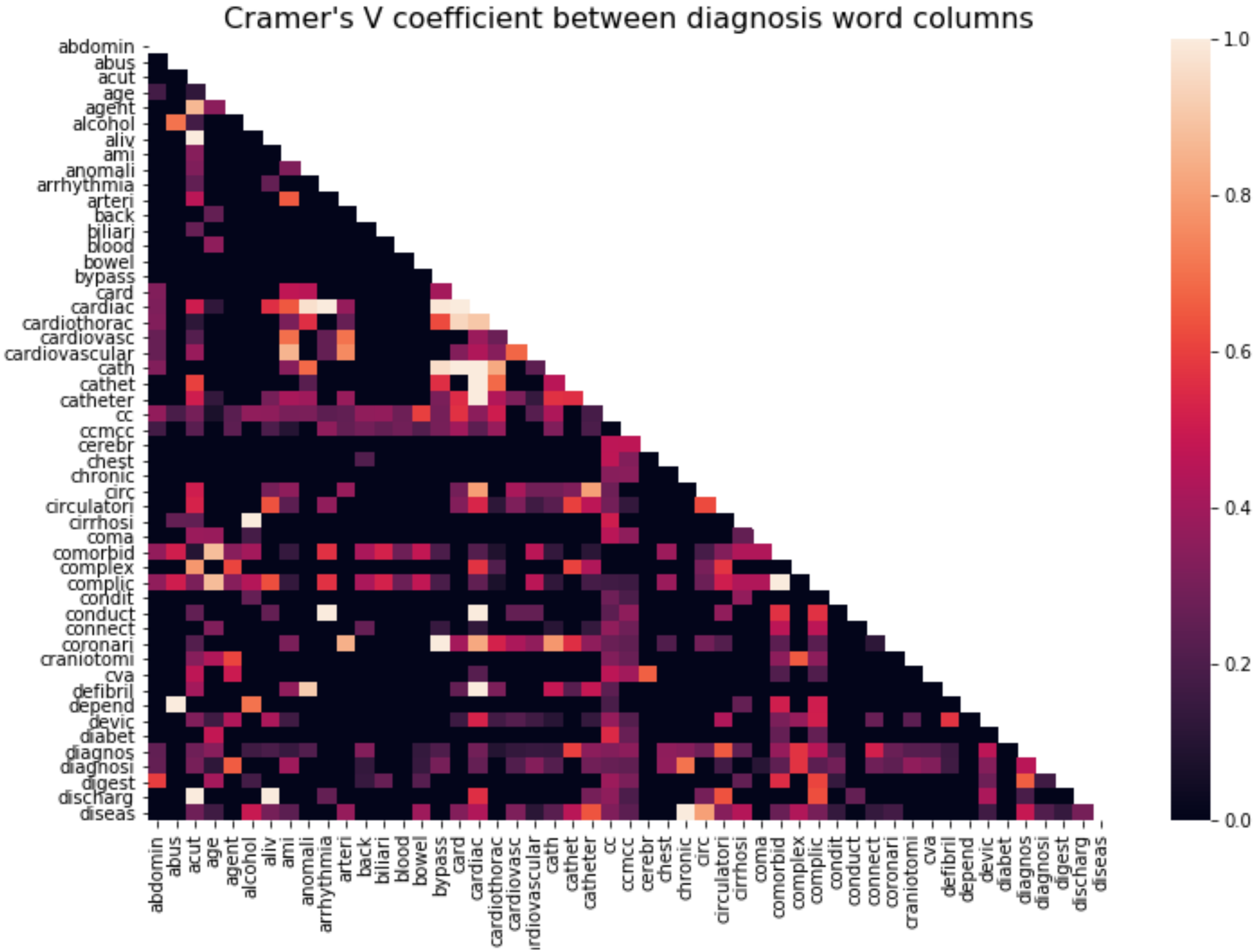
- Positively correlated lab measurements are meaningful
- On average, patients who will be readmitted in 30 days have lower hematocrit, hemoglobin and red blood cell measurements

Laboratory test results that are different for positive readmissions



* T test for mean difference of 2 independent samples with unequal variance.

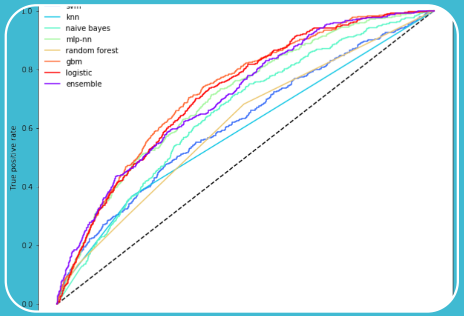
Cramer's V coefficient identifies strong associations between categorical variables



Manual inspection of diagnosis word associations

Word	Associated words
acut	agent, aliv, complex, discharg
age	comorbid, complic
alcohol	cirrhosi
anomali	cardiac, defibril
arrhythmia	cardiac, conduct
arteri	cardiovascular, coronari
bypass	cardiac, cath, coronari
card	cardiac, cardiothorac, cath

- Note, the incomplete words are result of text processing (stemming)
- The analysis looks reasonable, e.g. age is associated with comorbid(ities) and complic(ations), alcohol with cirrhosis(s) etc.



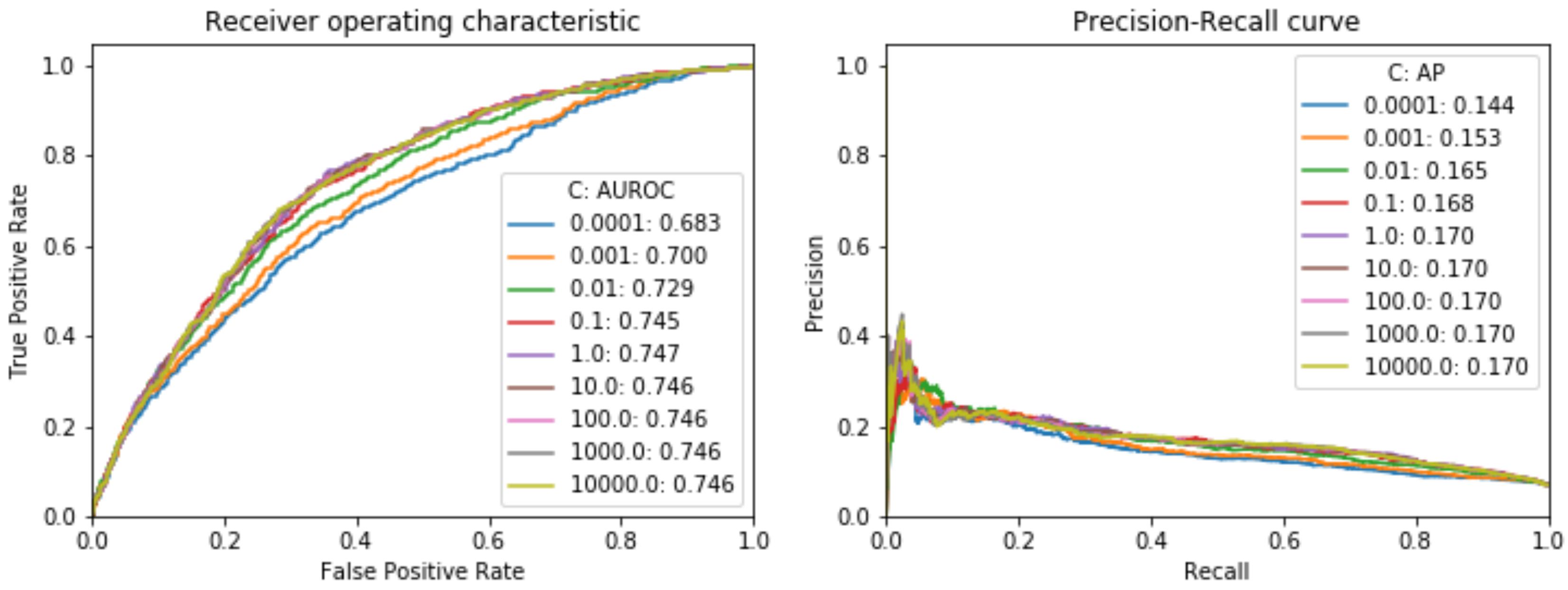
3. Feature selection and machine learning

- Define metrics for machine learning model evaluation
- Fit a machine learning model with and without feature selection
- Fit several machine learning models on reduced feature dataset
- Pick a final model based on evaluation metrics

Evaluation metric definition

- True positive rate (TPR) = positive prediction rate, the fraction of admissions followed by 30-day readmission that the model identifies
  - The higher the TPR, the more $s the hospital saves
- False positive rate (FPR) = False alarm rate, the fraction of incorrectly flagged positive readmissions out of total negative readmissions
  - The higher the FPR, the more $s hospital spends on wrong patients
- AUROC = area under receiver operating characteristic curve, a metric which combines TPR & FPR
  - Higher AUROC, more $s saved
- Precision = the fraction of admissions flagged by the model that will actually result readmission
  - The higher the precision, the less $s hospital spends on wrong patients
- Recall = TPR
- AP = Average precision, a metric to combine precision and recall
- **Since the dataset is imbalanced AUROC can give overly optimistic view, hence use both AUROC and AP**

Fit a logistic regression model without feature selection (314 features)

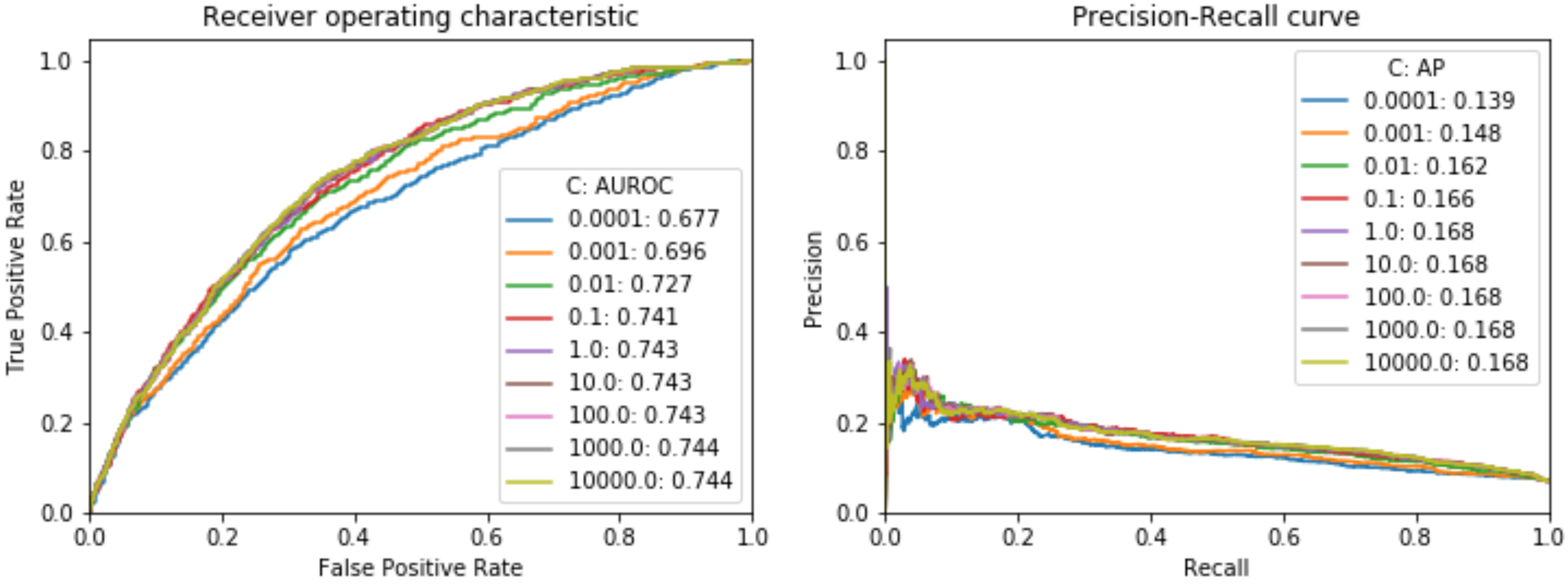


- Feature importance determined by the magnitude of coefficients
- C is the regularization parameter, high C value = larger coefficients
- Smaller C coefficients favor simpler models
- The best performance is achieved at C=1, after which there is no further improvement

Feature selection

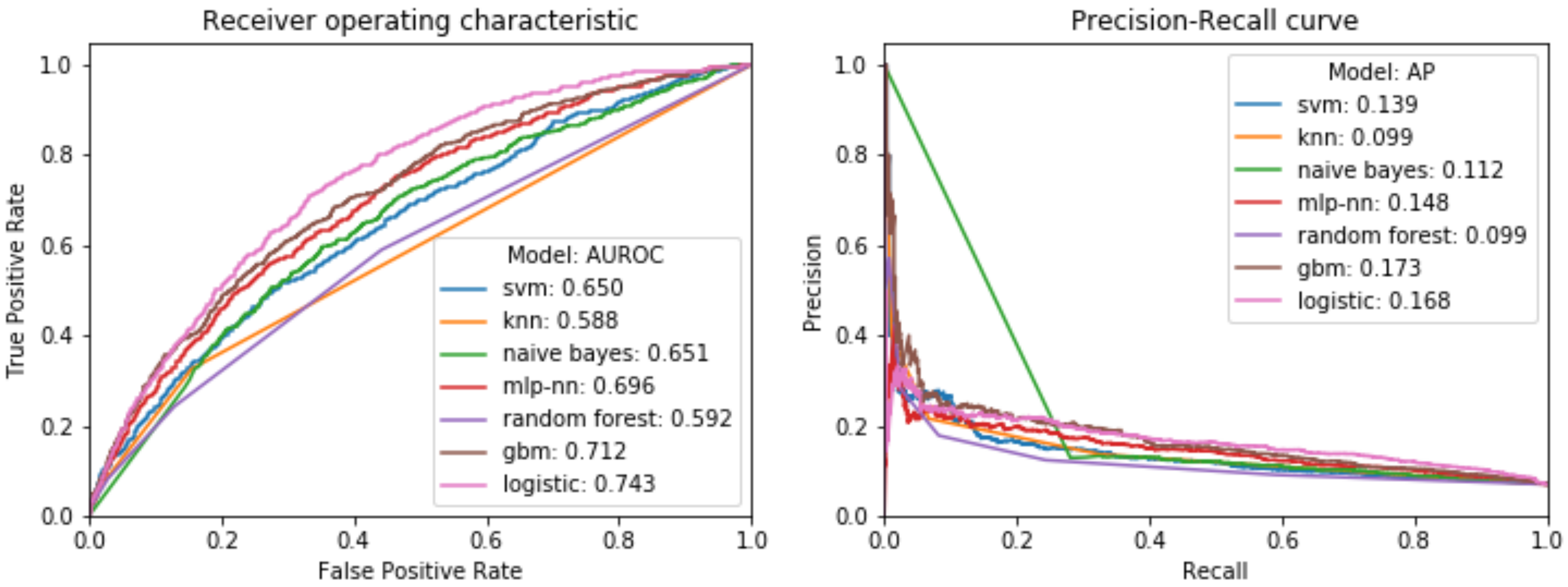
- Eliminate strongly associated features, helps model interpretation
  - Pearson correlation coefficient = strength of linear correlation between numerical features
    - Remove redundant variables |Pearson corr. Coef.| >=0.9
  - Cramer's C = strength of association between categorical variables
    - Remove redundant variables Cramer's C >= 0.9

Fit a logistic regression model after feature selection (188 features)

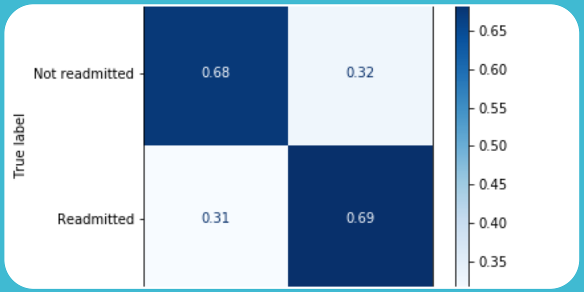


- The AUROC and AP are slightly smaller (by 0.04 and 0.02 respectively)
- Acceptable trade-off for eliminating ~40% of features
- Next, fit other models on reduced feature set

Logistic regression has the best predictive performance



- Gradient boosting machine (GBM) has better AP than logistic regression, also smaller AUROC
- Logistic regression is a preferable model to GBM (ensemble model)
- Proceed the analysis with the logistic regression model

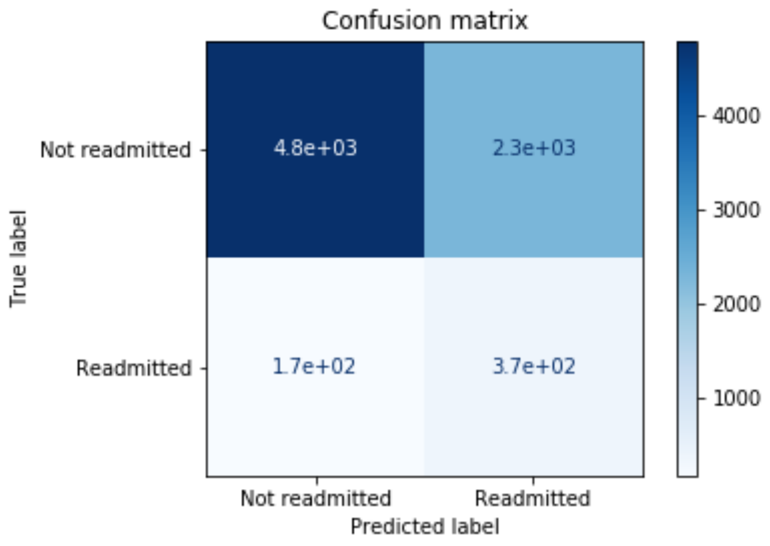


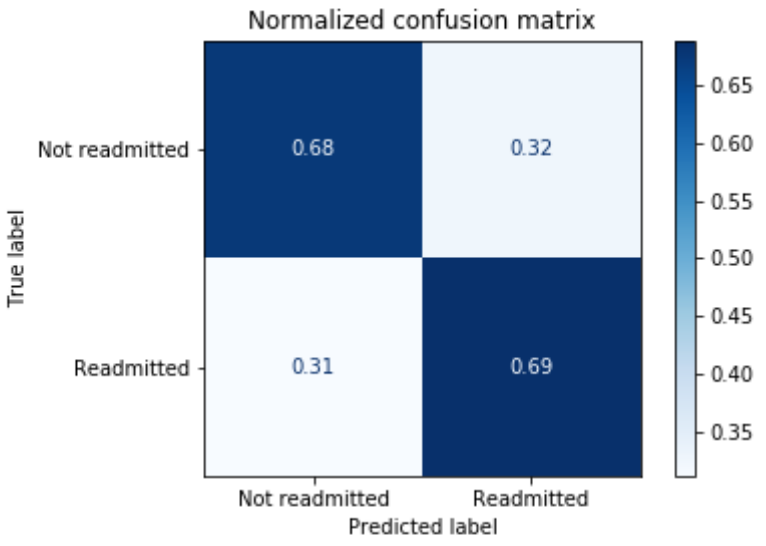
4. Final evaluation & interpretation

- Calculate the full classification report
- Plot the confusion matrix
  - Final tally of correct and missed predictions
- Interpret feature coefficients
  - Are the features with largest coefficients meaningful?
- Where to go next with this analysis

Final logistic regression model metrics

	precision	recall	f1-score	support
Not readmitted	0.97	0.68	0.8	7109
Readmitted	0.14	0.69	0.23	534
accuracy			0.68	7643
macro avg	0.55	0.69	0.51	7643
weighted avg	0.91	0.68	0.76	7643





Model coefficient interpretation

feature	coefficient
SUBSECTIONHEADER_Pulmonary	3.532221
DISCHARGE_LOCATION_LONG TERM CARE HOSPITAL	3.052530
DISCHARGE_LOCATION_LEFT AGAINST MEDICAL ADVI	2.894396
DISCHARGE_LOCATION_REHAB/DISTINCT PART HOSP	2.838508
DISCHARGE_LOCATION_SNF	2.830531
DISCHARGE_LOCATION_HOME HEALTH CARE	2.554723
DISCHARGE_LOCATION_HOME	2.264512
DISCHARGE_LOCATION_SHORT TERM HOSPITAL	1.921868
DISCHARGE_LOCATION_DISCH-TRAN TO PSYCH HOSP	1.899075
Mean_RDW	1.575423
SUBSECTIONHEADER_TOTAL	1.483992
SUBSECTIONHEADER_Physical medicine and rehabil...	1.420191
Mean_Hematocrit	1.394244
Count_INR(PT)	1.309206
SUBSECTIONHEADER_Respiratory system	1.281702
Mean_White Blood Cells	1.221694
Mean_Glucose	1.190997
SUBSECTIONHEADER_Noninvasive vascular diagnost...	1.190089
craniotomi	1.153944
DISCHARGE_LOCATION_HOSPICE-HOME	1.123975
Mean_Creatinine	1.102897
minor	0.992975
spinal	0.926933
Mean_Bicarbonate	0.911841
Var_Sodium	0.817911
vascular	0.762537
Mean_Urea Nitrogen	0.726566
1Y_PRIOR_ADM	0.720474
heart	0.710269
DISCHARGE_LOCATION_DISC-TRAN CANCER/CHLDRN H	0.673262

Top 30 features for predicting readmission

Discharge locations – especially transfers to other healthcare facilities – indicates more acute condition

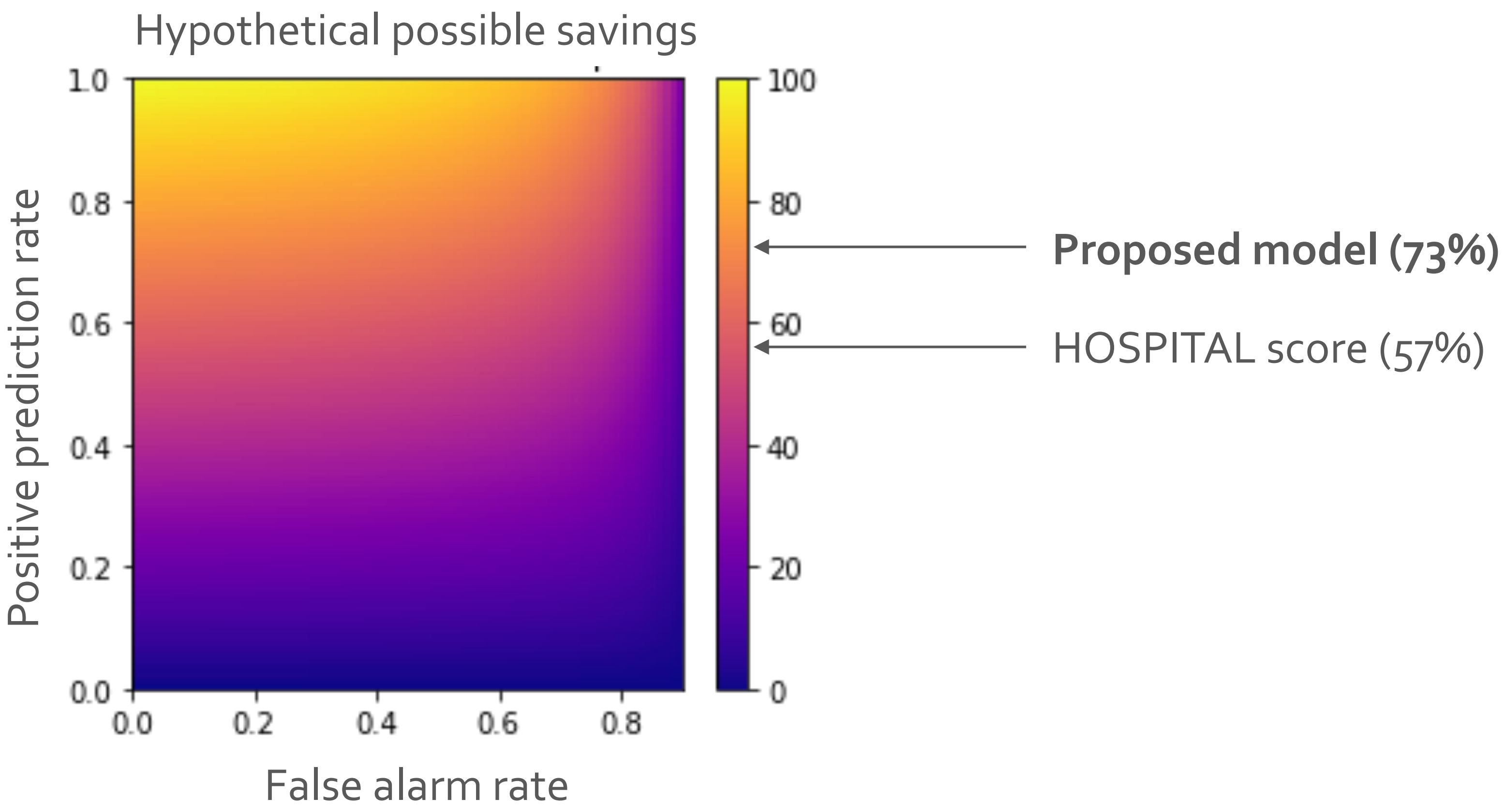
Procedures – especially pulmonary and respiratory

Mean lab measurements, indicating chronic conditions, e.g. glucose (diabetes), urea nitrogen (kidney malfunction), hematocrit (anemia), etc.

Diagnosis description words – brain surgery and cardiovascular condition

Prior admission in the proceeding year

5. Hospital savings analysis



The proposed logistic regression model will save the hospital additional 16%, on top of the HOSPITAL score, hence it is worth adopting!

Future work for improving the model

- The presented model suffers from high bias, and (to a much lower degree) high variance
  - Mean train accuracy 0.688, mean test accuracy 0.677
- Proposed solutions:
  - Add more features, e.g. vital signs, prescriptions
  - Gather more data
  - Fit non-linear models
  - Apply deep learning language models, e.g. BERT on clinical notes

Acknowledgements

- Thomas Blanchard (mentor)
- Springboard