Predicting Hospital Readmissions at Berkshire Medical Center

Hallee Wong

Williams College

5/2/2018

Outline

Introduction
- The Data
- Accuracy Metrics

Model Fitting Methods
- Logistic Regression with Lasso
- Decision Tree
- Initial Results

Utilizing Medical Codes
- Ensemble Models
- Clustering Medical Codes
- Results

More Interpretable Models
- Mixed Integer Programming

Conclusions and Limitations

Hospital Readmission

Definition

30-Day readmission occurs when a patient returns to hospital within 30 days of being discharged from inpatient care and is again admitted to inpatient care.

Hospital Readmission

Definition

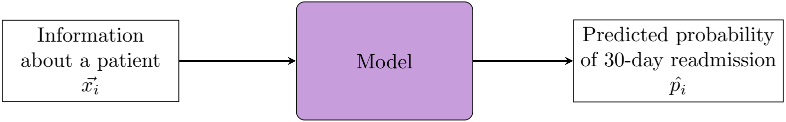
30-Day readmission occurs when a patient returns to hospital within 30 days of being discharged from inpatient care and is again admitted to inpatient care.

Healthcare Policy

- Frequent rehospitalizations are costly to patients' health and hospitals
- Affordable Care Act penalizes hospitals with higher than expect readmission rates
- The most effective interventions are time and resource intensive (e.g. follow-up home visits, inidividual case management) (Verhaegh et al., 2014)

The Prediction Problem

Goal



Berkshire Medical Center and the Data

- 17,093 rows (10,895 unique patients)
- September 1st 2015 - December 31st 2016
- 17% of patients readmitted within 30 days

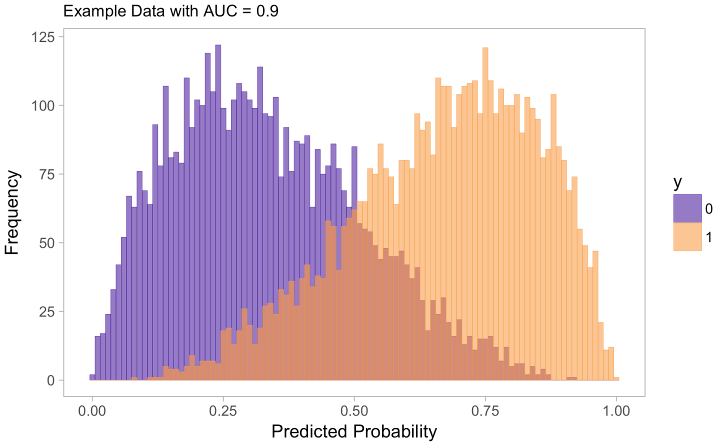
Available Variables

Demographic	Hospital Utilization	Clinical
-Age	-ED Admit	-Diagnosis codes
-Gender	-Surgery	-No. of diagnosis codes
	-Insurance	-No. of medications
	-Admit source	
	-Disposition	
	-Length of stay	
	-No. of previous hospitalizations	

Properties of an Ideal Model

- Rank accurate
- Interpretable
- Generalizable

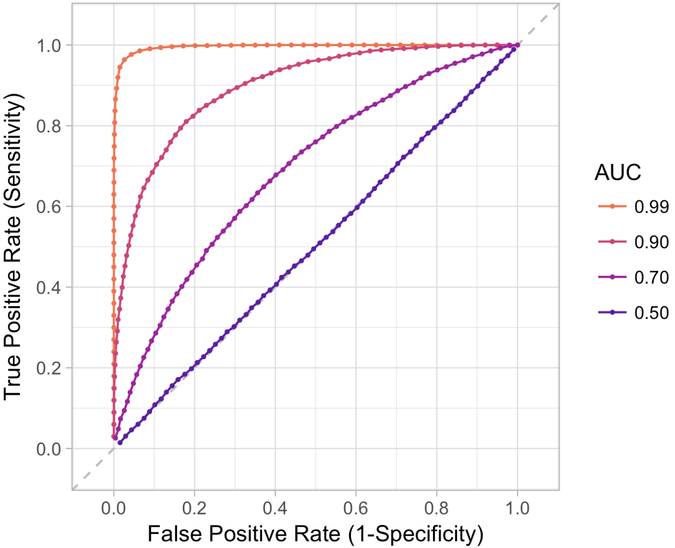
Measuring Rank Accuracy



Possible Metrics

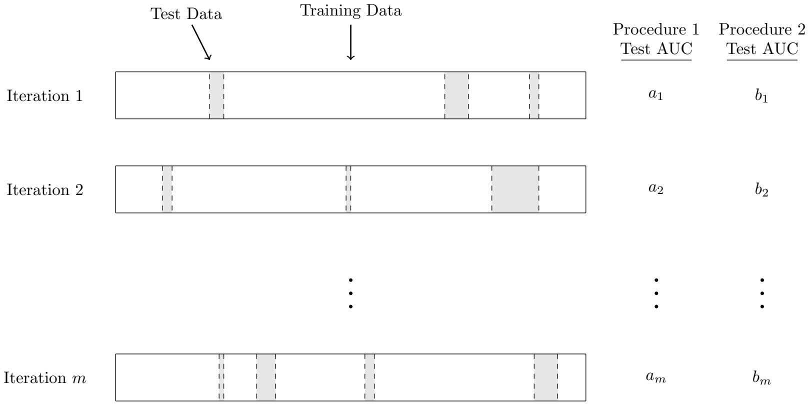
- Accuracy or Misclassification Rate
- Receiver Operator (ROC) Curve

The Receiver Operator (ROC) Curve



Estimating Performance on New Data

Repeated Random Sub-sampling



Modeling Methods

Now

- Logistic regression
  - with lasso feature selection
- Decision Tree

Later

- Mixed Integer Programming based methods

Logistic Regression

Assume $Y_i \sim \mathit{Bernoulli}(p_i)\ \forall i$ and

$$p_i = \frac{1}{1 + e^{-\vec{x_i}^T \vec{\beta}}}$$

then

$$P\big(\ Y = \vec{y} \mid p_i\big) = \prod_{i=1}^{n} {p_i}^{y_i}\ (1 - p_i)^{y_i}\ .$$

We find $\vec{\beta}$ such that

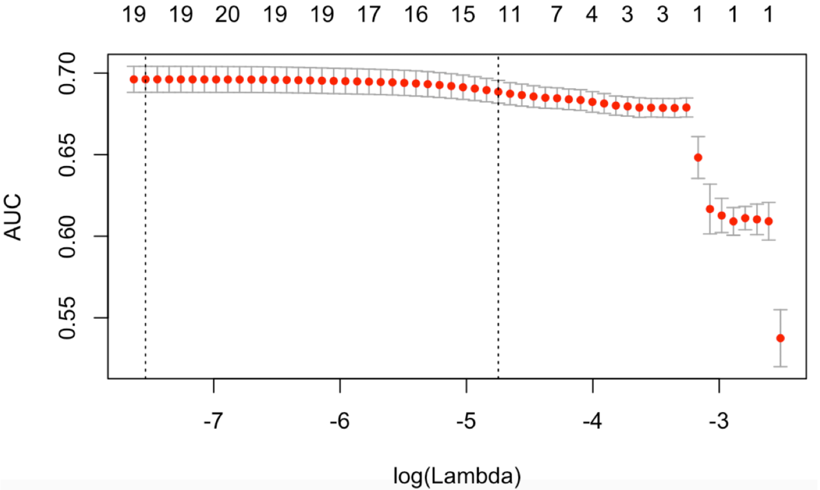
$$\max_{\vec{\beta} \in \mathbb{R}^k}\ L(\vec{\beta} \mid \vec{y}) = \max_{\vec{\beta} \in \mathbb{R}^k} P\big(\ Y = \vec{y} \mid p_i\big)\ .$$

Lasso Logistic Regression

$$\max_{\vec{\beta}\in\mathbb{R}^k} \ L(\vec{\beta} \mid \vec{y}) = \max_{\vec{\beta}\in\mathbb{R}^k} \ \prod_{i=1}^{n} \ p_i(1-p_i)^{1-y_i}$$

$$\text{subject to} \quad ||\vec{\beta}||_1 \quad = \sum_j |\beta_j| \ \leq \ \lambda$$

Lasso Logistic Regression - Choosing Lambda



Decision Tree

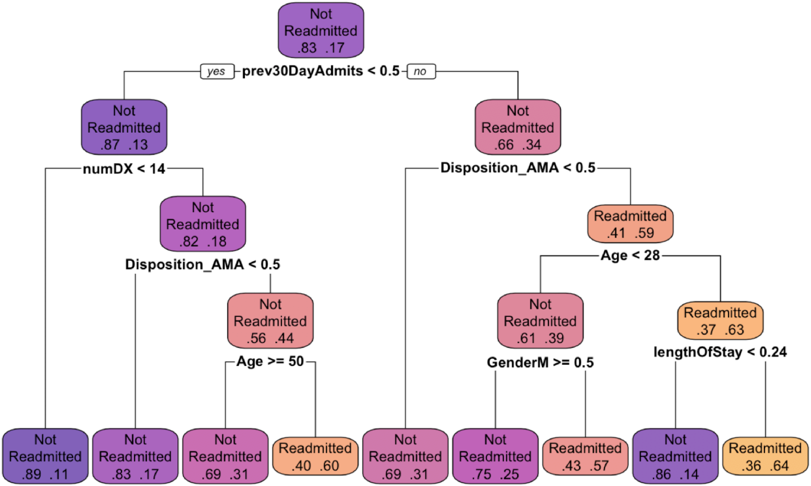
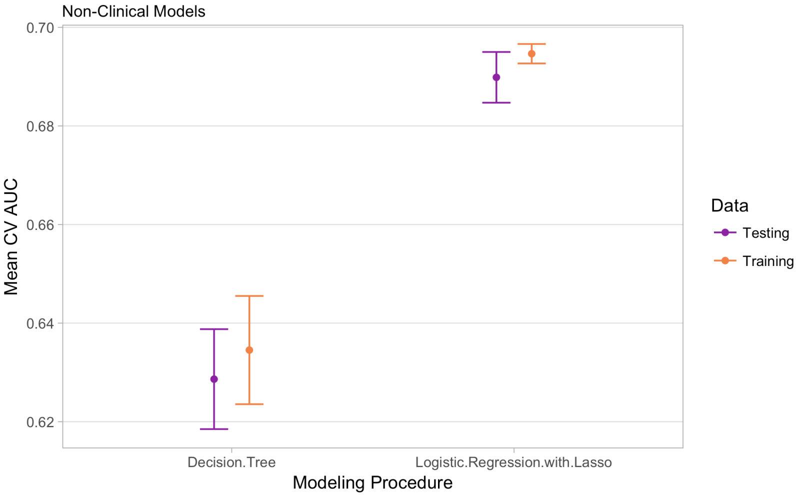


Figure 1: Example decision tree using the CART algorithm (Breiman et al., 1984)

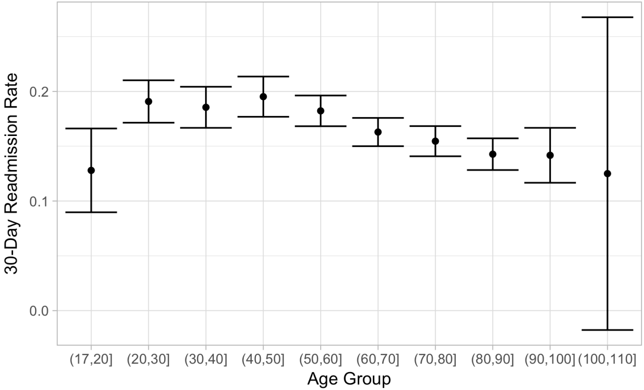
Results



Non-Clinical Models
0.70
0.68
0.66
0.64
0.62
Mean CV AUC
Decision.Tree
Logistic.Regression.with.Lasso
Modeling Procedure
Data
Testing
Training

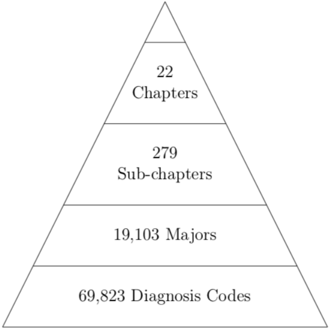
Limitations

- Patients may be readmitted to other hospitals
- Our readmission rate is confounded with mortality rate



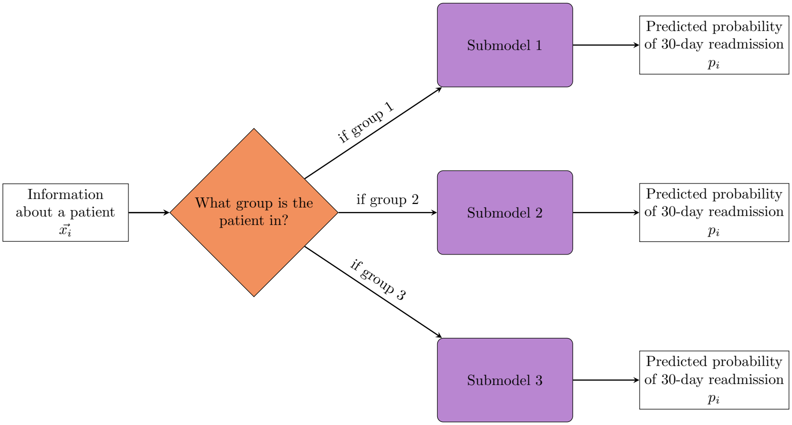
Incorporating High Dimensional Medical Code Data

Hierarchy of Diagnosis Codes (ICD-10)

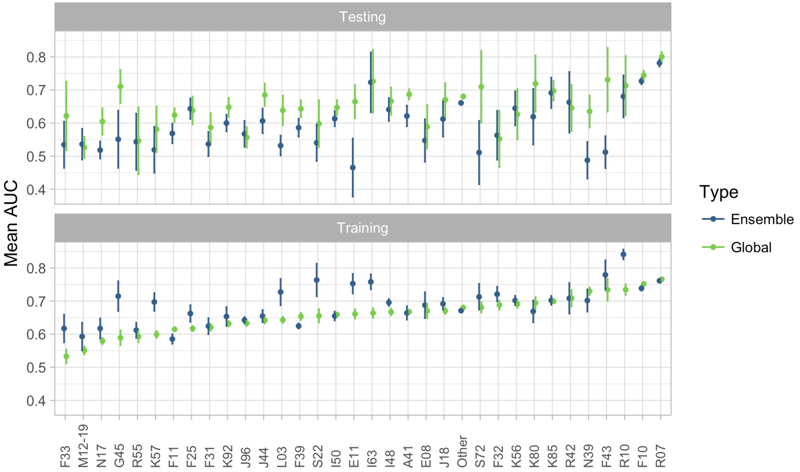


	Example
Chapter	Mental, Behavioral and Neurodevelopmental disorders
Sub-chapter	Mental And Behavioral Disorders Due To Psychoactive Substance Use
Major	Alcohol related disorders
Code	F10951
Diangosis	Alcohol dependence with alcohol-induced psychotic disorder with hallucinations

Ensemble Models



Results



Testing
Training
Mean AUC
0.8
0.7
0.6
0.5
0.4
F33
M12-19
N17
G45
R55
K57
F11
F25
F31
K92
J96
J44
L03
F39
S22
I50
E11
I63
I48
A41
E08
J18
Other
S72
F32
K56
K80
K85
R42
N39
F43
R10
F10
R07
Type
Ensemble
Global

Conclusions

Summary

- Ensemble models with 34 submodels generally performed worse than a global model

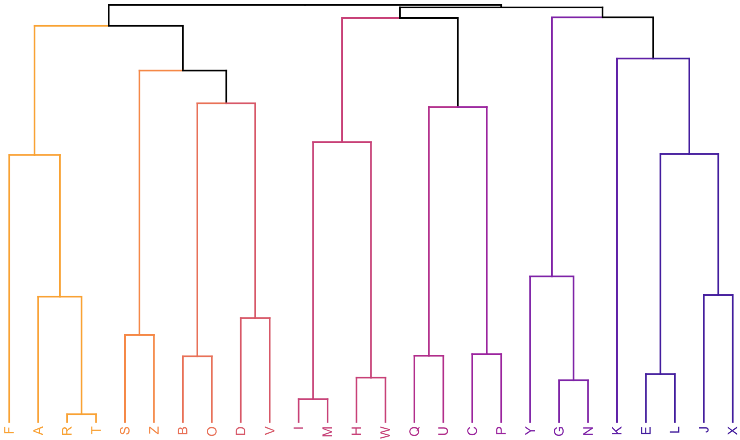
Possible Explanations

- this grouping is not reflective of the true structure in the patient population
- the patient population is homogenous

Problem

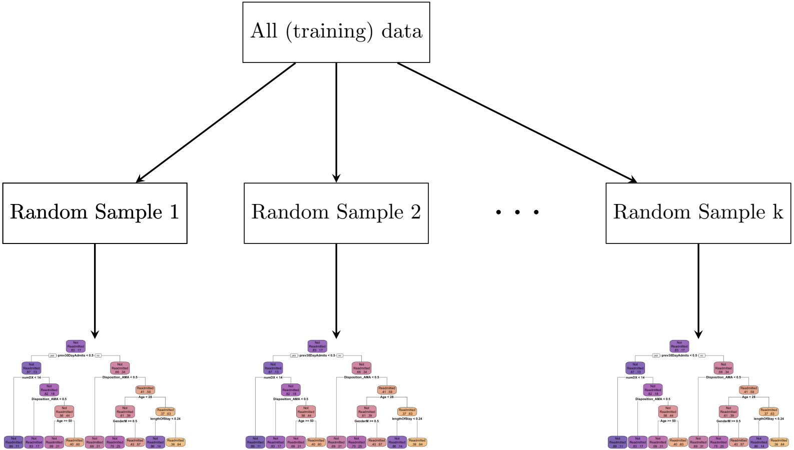
How can we identify groups which are signifigantly different enough that fitting submodels will improve the overall system's accuracy?

Hierarchical Clustering



*Requires a metric of similarity/distance between groups

Random Forests

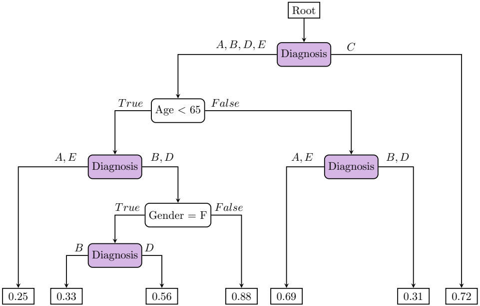


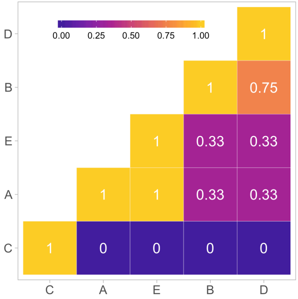
Categorical Co-frequency Analysis (CCFA)

Proposed Statistic

$$s_{i,j} = \frac{\#\text{ times label } i \text{ and label } j \text{ split in the } \textbf{same direction}}{\#\text{ times label } i \text{ and label } j \text{ are used as split criteria}}$$

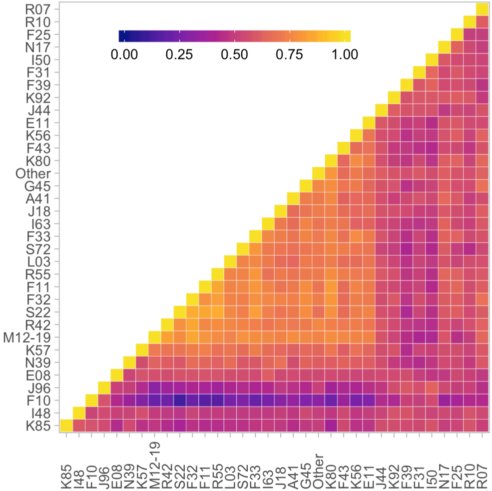
An Example


Root
A, B, D, E
Diagnosis
C
True
Age < 65
False
A, E
Diagnosis
B, D
A, E
Diagnosis
B, D
True
Gender = F
False
B
Diagnosis
D
0.25
0.33
0.56
0.88
0.69
0.31
0.72


0.00
0.25
0.50
0.75
1.00
D
B
E
A
C
C
A
E
B
D
1
1
0.75
1
0.33
0.33
1
1
0.33
0.33
1
0
0
0
0

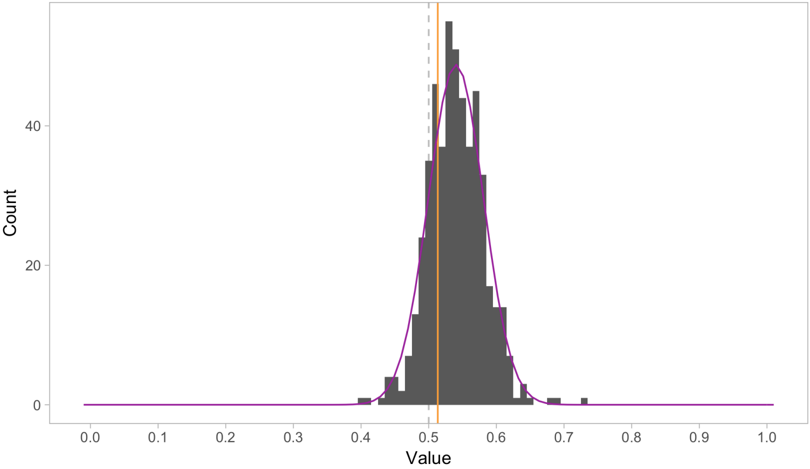
Categorical Co-Frequency Analysis (CCFA)

Figure 2: CCFA statistics calculated for each pair of labels in a forest of 100 trees

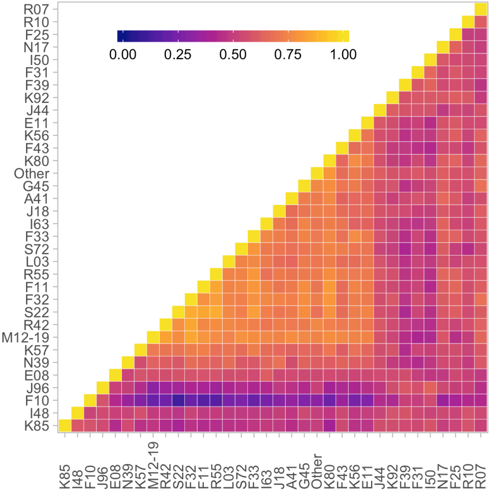


Are all of these statistics signifigant?

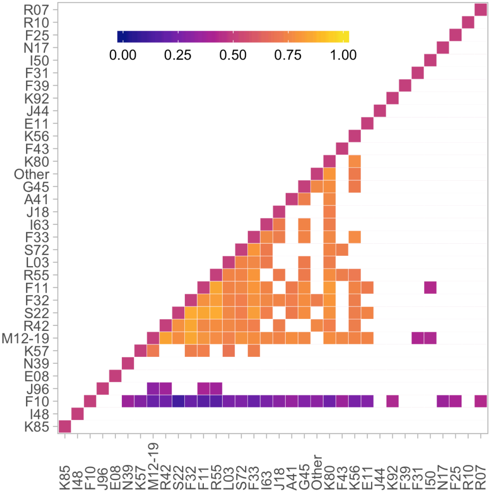
Figure 3: Type 2 diabetes mellitus (E11) and Bipolar disorder (F31)



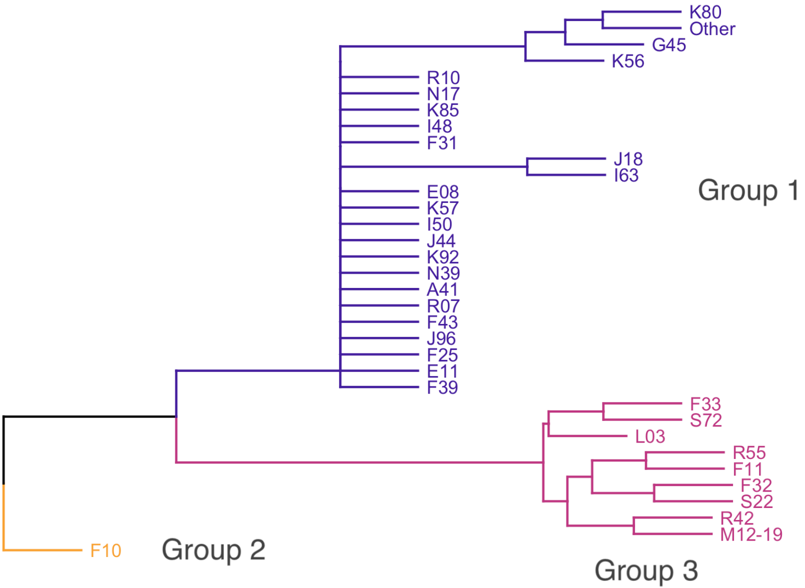
Performing Hypothesis Testing



After Hypothesis Testing



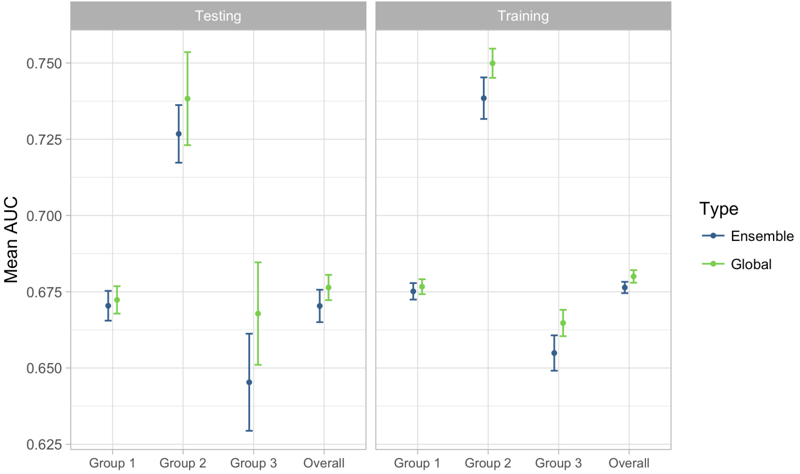
Hierarchical Clustering



CCFA Groups

	Group 1	Group 2	Group 3
Number of Patients	13,003	1,247	2,843
30-day Readmission Rate	16.87%	33.04%	9.85%

Results



Testing
Training
Mean AUC
0.750
0.725
0.700
0.675
0.650
0.625
Group 1
Group 2
Group 3
Overall
Group 1
Group 2
Group 3
Overall
Type
Ensemble
Global

Results

	Group 1	Group 2	Group 3	Global
(Intercept)	-2.075	-1.975	-2.083	-2.294
Age	-0.008			-0.006
GenderM	0.153			0.200
PayerCOM	-0.344		-0.549	-0.356
PayerMCD	0.111			0.095
PayerMCR		0.287		
EDAdmitY				0.131
SurgeryY			-0.856	-0.246
TotalMeds	0.020			0.020
prev30DayAdmits	0.648	0.707	0.632	0.705
numDX	0.028	0.051		0.032
Disposition_AMA	0.888	1.088	0.632	1.046
Disposition_BMC		-1.252		
Disposition_COMMFAC		-1.105		
Disposition_COMMRES	0.637			
Disposition_HOMEHS	0.377			0.198
Disposition_SNF				-0.235
Mean Test AUC	0.670	0.723	0.645	0.689

Table 1: Variables not used: *lengthOfStay, AdmitSource_ACHOSP, AdmitSource_BMC, AdmitSource_PHYSELF,AdmitSource_SNF, Disposition_HOME*

Conclusions

- Ensemble on 34 groups performed poorly
- Performance of ensemble on 3 CCFA groups was not signifigantly different than a single global model

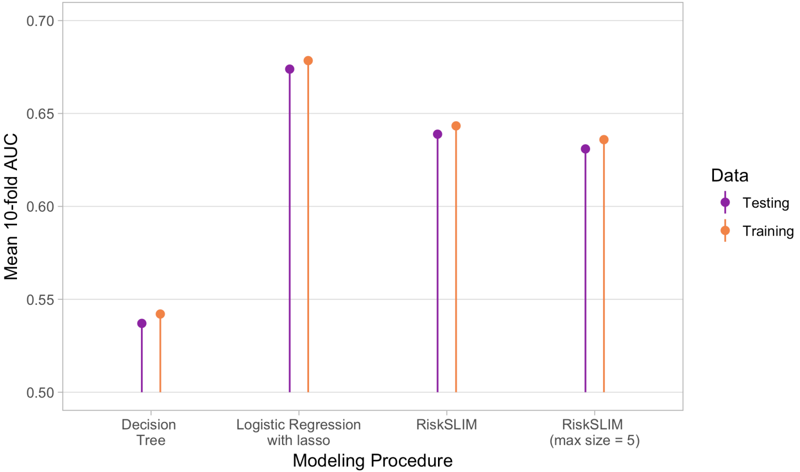
RiskSLIM

Table 2: RiskSLIM model (Ustun and Rudin, 2017) fit to all data with 30 minute time limit and a constraint of 5 non-zero coefficients. The model intercept is -2.

1. $\textit{NumDX} < 14$	-1 points	$\dots$
2. $\textit{CCoFA Group} = 2$	1 points	$+ \dots$
3-4. $\textit{Payer} =$ MCD or MCR	1 points	$+ \dots$
5. $\textit{Disposition} =$ AMA	1 points	$+ \dots$
	Total Score	$= \dots$

Total Score	-1	0	1	2	3
Predicted Risk	4.7 %	11.9 %	26.9 %	50.0 %	73.1 %

More Interpretable Models



Furture Work

Categorical Co-Frequency Analysis

- Simulation studies of the sensitivity of CCFA

Clustering for Ensemble Models

- Other similarity metrics:
  - co-frequency of diagnoses in the secondary diagnosis lists
  - embeddings
- An ensemble of RiskSLIM models

Acknowledgements

- Professor Miller and Professor Heggeseth
- Dr. Jeffrey Thomas and Alex Lopez (Lever Inc.)
- Dr. Gray Ellrodt, Dr. Mark Pettus and Berkshire Health Systems
- Clare Booth Luce Scholars Program
- Dr. Berk Ustun
- Family and friends

Selected References

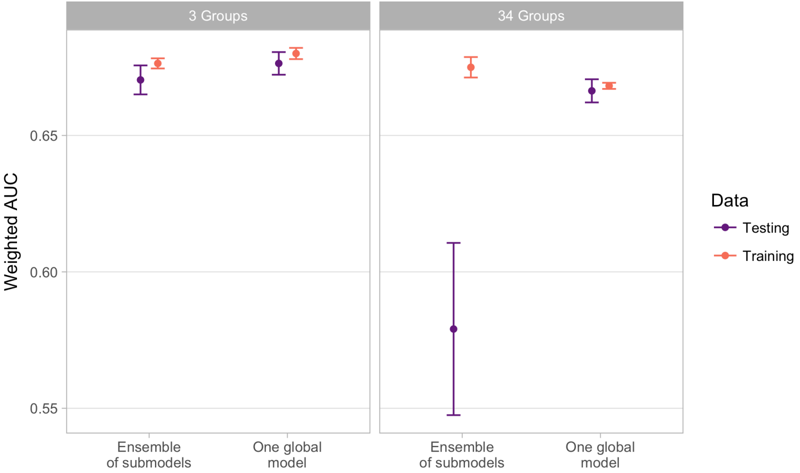
L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*. Chapman & Hall, New York, 1984.

J. Futoma, J. Morris, and J. Lucas. A comparison of models for predicting early hospital readmissions. *Journal of Biomedical Informatics*, 56:229–238, 2015.

B. Ustun and C. Rudin. Optimized risk scores. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2017.

K. J. Verhaegh, J. L. MacNeil-Vroomen, S. Eslami, S. E. Geerlings, S. E. de Rooij, and B. M. Buurman. Transitional care interventions prevent hospital readmissions for adults with chronic illnesses. *Health Affairs*, 33(9):1531–1539, 2014.

Summary of Ensemble Results



RiskSLIM (Ustun and Rudin, 2017)

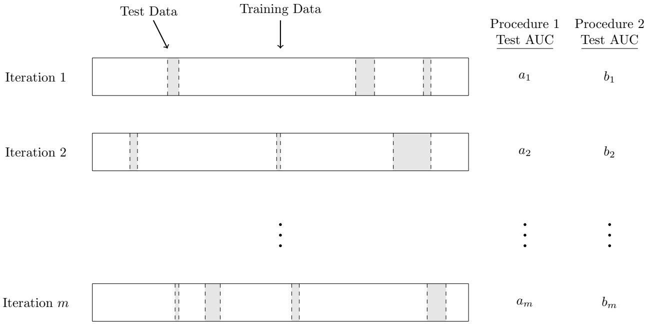
We assume

$$p_i = P(y_i = 1|\vec{x_i}) = \frac{1}{1 + e^{-\vec{x_i}^T\lambda}}$$

and find the *integer* coefficients λ from a set of integers $\mathcal{L}$ that satisfied

$$\min_{\lambda \in \mathcal{L}} \; \frac{1}{N}\sum_{i=1}^{N} \log(1 \; + e^{-y_i\lambda^T x_i}) + C_0||\lambda||_0 \; .$$

Corrected Resampled (Paired) T-test



$$t = \frac{\frac{1}{m}\sum_{i=1}^{n} x_i}{\hat{\sigma}\sqrt{\frac{1}{n} + \frac{n_{test}}{n_{train}}}}$$

$$\begin{aligned} \text{let } x_i &= a_i - b_i \\ m &= \text{number of repeated} \\ &\ \text{random sampling iter.} \\ n_{\text{test}} &= \text{No. test examples} \\ n_{\text{train}} &= \text{No. training examples} \end{aligned}$$

AUC vs. Readmission Rate (34 submodels)

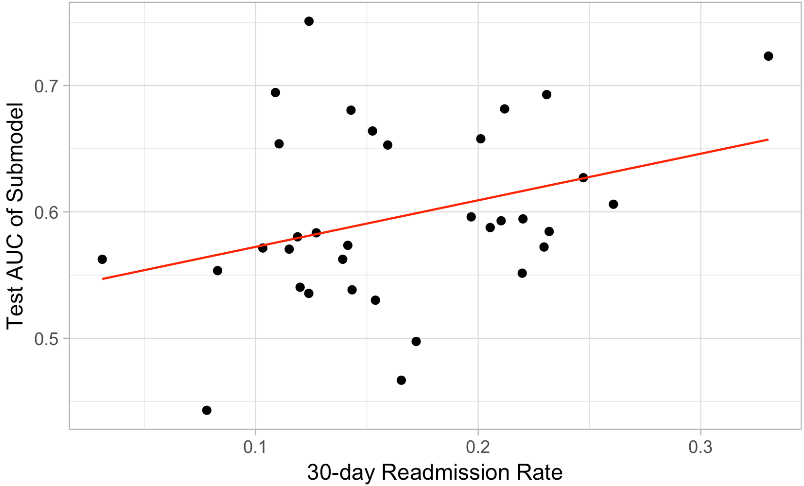


Figure 4: Slope is 0.37 (p-value = 0.06)

Random Forest vs. Non-Random Forest

Note: random forest implemented using dice rolling instead of subsetting at each node.

Similar Ensemble Models

(Futoma et al., 2015)

- Divided data into 560 groups by Diagnosis Related Group (DRG) codes
- Used 10-fold CV to assess models
- Found weak but signifigant corelation between 30-day readmission rate

Readmission Rates for Medicare Patients