

Assessing Marriage Stability

Using the 1979 National Longitudinal Survey of Youth



Data & Objective

12,868 respondents \times **27** survey years \times \sim **2,000** questions per year

5y_fwd	age	sex	...	is_employed	num_kids
1	25	M	...	0	1
0	36	M	...	1	6
0	30	F	...	1	0
0	27	F	...	0	0

Y

X: shape = (98,117 samples \times 217 features)

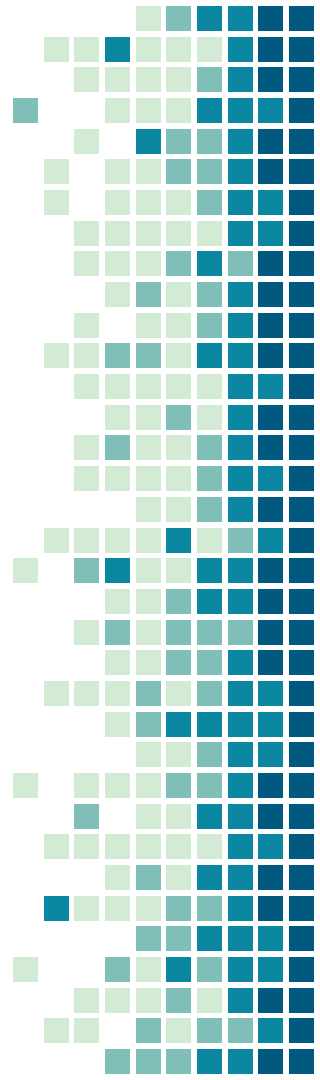
Considerations

Design matrix X : fairly high sparsity ratio (~72%)

Class imbalance (9:1)

Nonlinearity & feature interaction?

Effects of inflation & nonstationary in data's time-series component



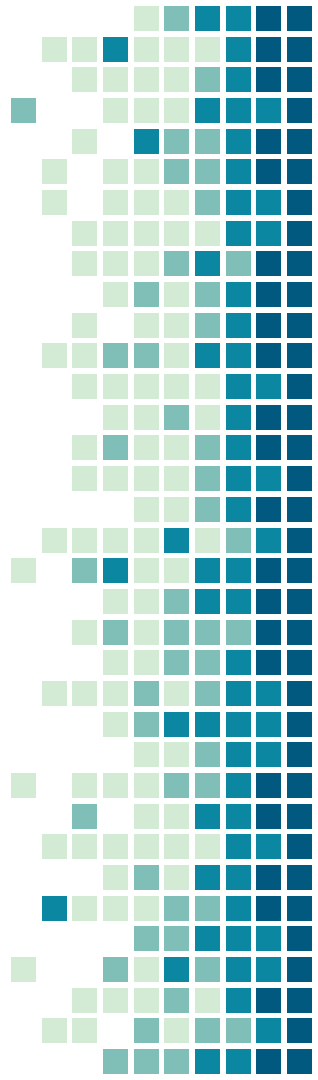
Implementation

Random Forest Classifier (2500 trees)

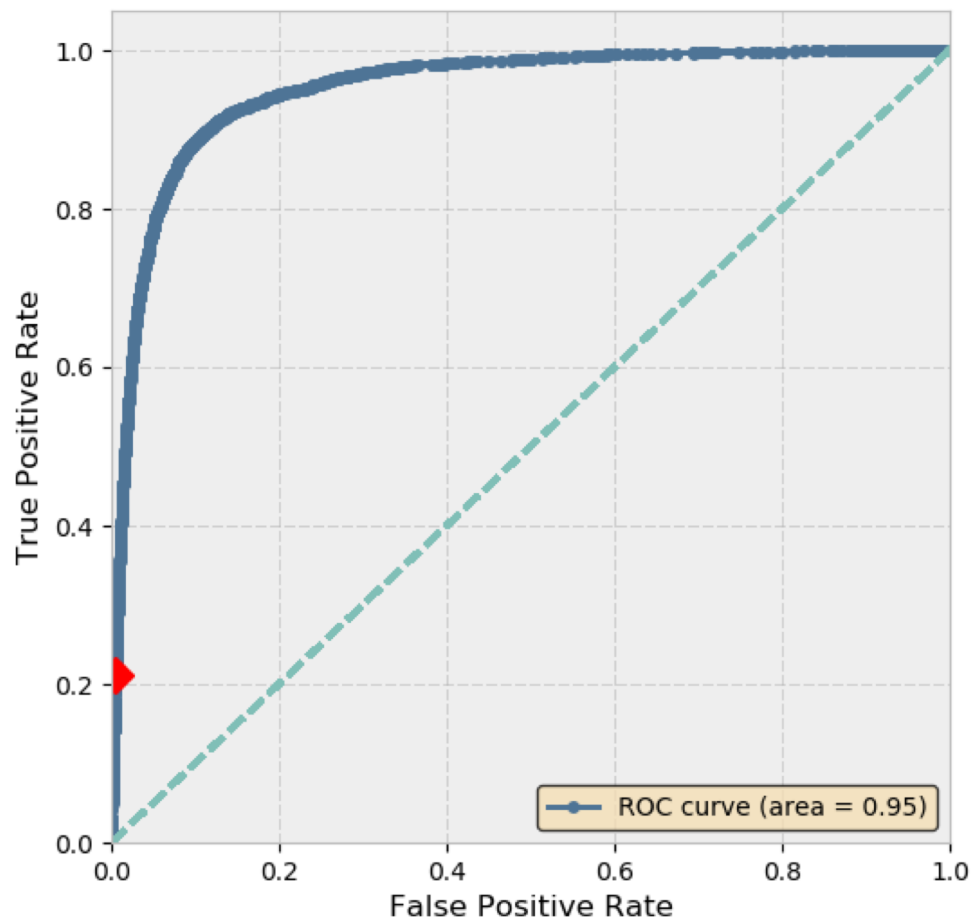
Grid search over $\{tree\ depth, feature\ space\}$

Optimize on *AUC* rather than accuracy

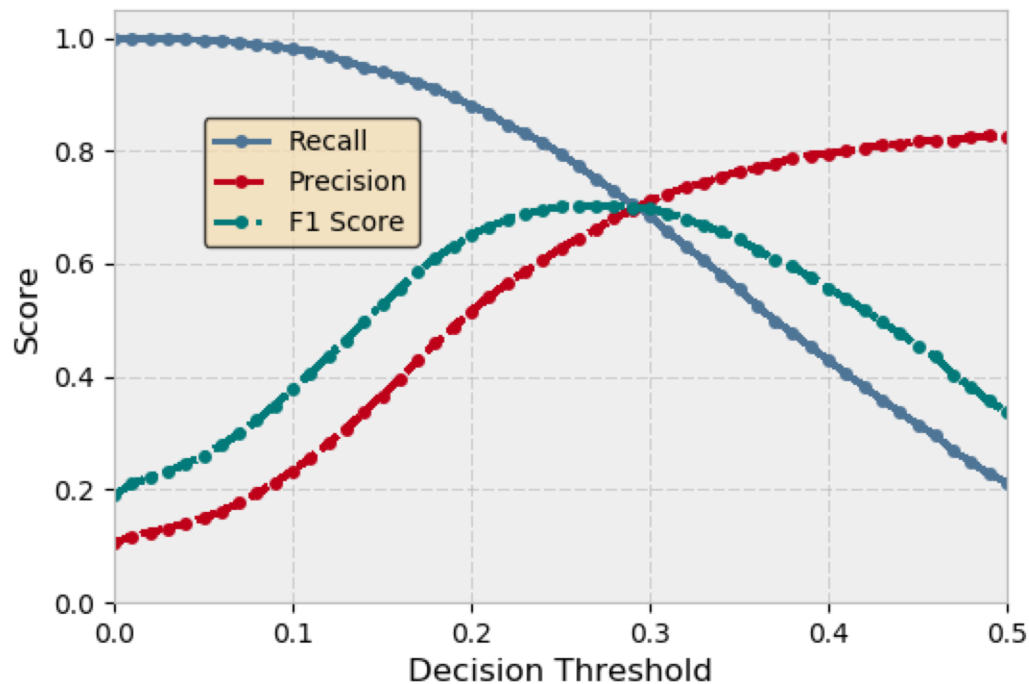
Control overfitting with minimum sample-split



Receiver Operating Characteristic



Recall, Precision, & F1 as a Function of Threshold



Notable Features

(From Feature Importance Scores)

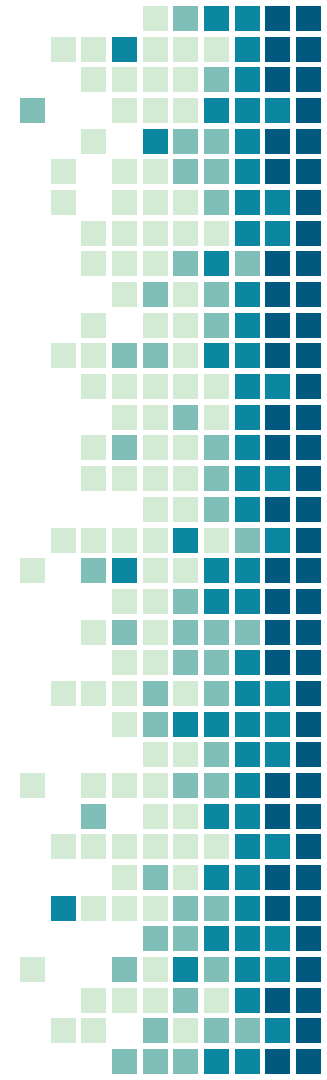
Status: is separated

Length of current marriage

Job turnover (cumulative jobs held)

Alcohol usage

Level of schooling completed



Example Cases

Person 541: $\rightarrow P(\text{divorce}) = 89.6\%$

2nd marriage at age 22; two children

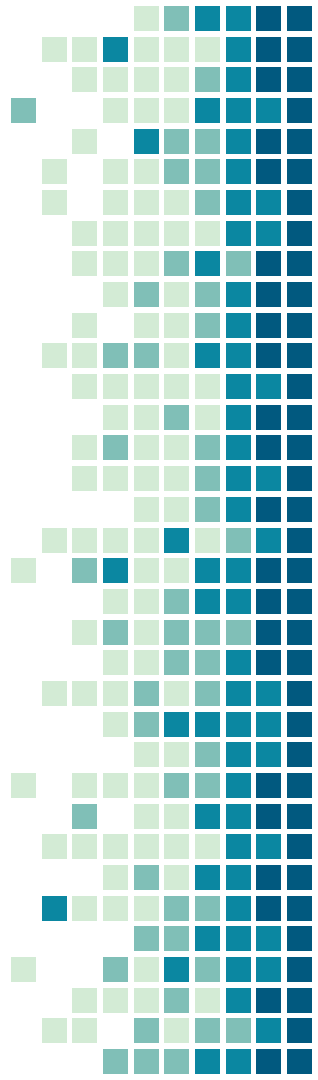
History of juvenile delinquency

Completed < 9th grade

Person 761: $\rightarrow P(\text{divorce}) = 80.2\%$

Extended unemployment & deteriorating net wealth

Very high self esteem and self-reported lifestyle expectations

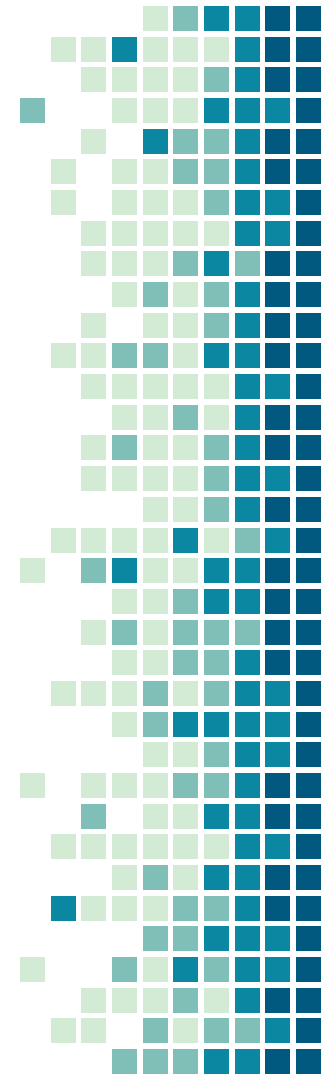


Next Steps

Merge with other cohorts (NLSY97; different feature sets)

Incorporate macroeconomic factors (coincident economic indicators)

Consider longer horizon (20 years versus 5)



Thank You

brad.solomon.1124@gmail.com

 bradleysolomon/

 bsolomon1124/nlsy79