seismos

Predicting Non-For-Profit Health Risks

Alden Golab, Paul Mack, Sam Sun

Chicago Tribune

Questions? Call 1-800-Tribune | Saturday, January 21, 2012 | Breaking news at chicagotribune

The fall of Hull House

Jane Addams' dream of helping Chicago's needy dates to the 19th century. Now the charity has run out of time and money.

In 2012, the collapse of Hull House unexpectedly left 60,000 at-risk Chicagoans without services

By Liam Ford and Kate Thayer
Tribune

departed from Addams' reliance on private donors and turned instead

for foster children, that brought in only about a quarter of Hull House's

ers housing prominent Chicagoans at places like the Museum of Science and Industry, it's been hard

Seismos seeks to develop risk scores for declining not-for-profit financial performance, predicting the fiscal stability of NFP firms for the year ahead so that they can get support before it's too late

Our Data

Tax information for the Universe of US Non-Profits filling 990s in 2012-2015

Every year, an extract of the tax information for all qualifying non-profit organizations is released

Data is available from CY 2012 through CY 2015; 990 filings cover organizations with more than $500k in assets or $200k in income

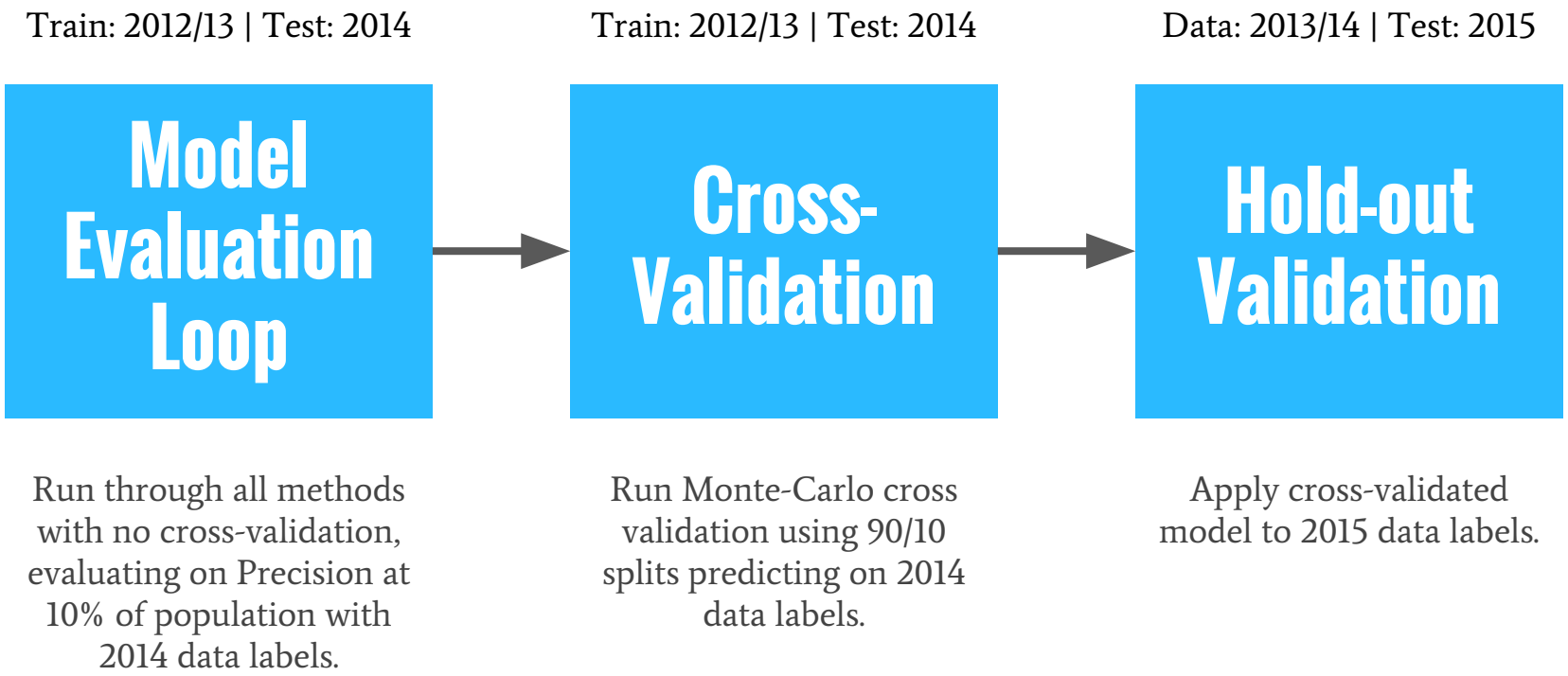
Approximately 300,000 organizations

Feature Generation

154 Generated **98** Utilized in Model

We include a variety of fiscal performance features like change in total functional expenses, percent of revenue from programs, asset-debt ratios, and others, in addition to some macro level features for the MSA

Model Development Flow

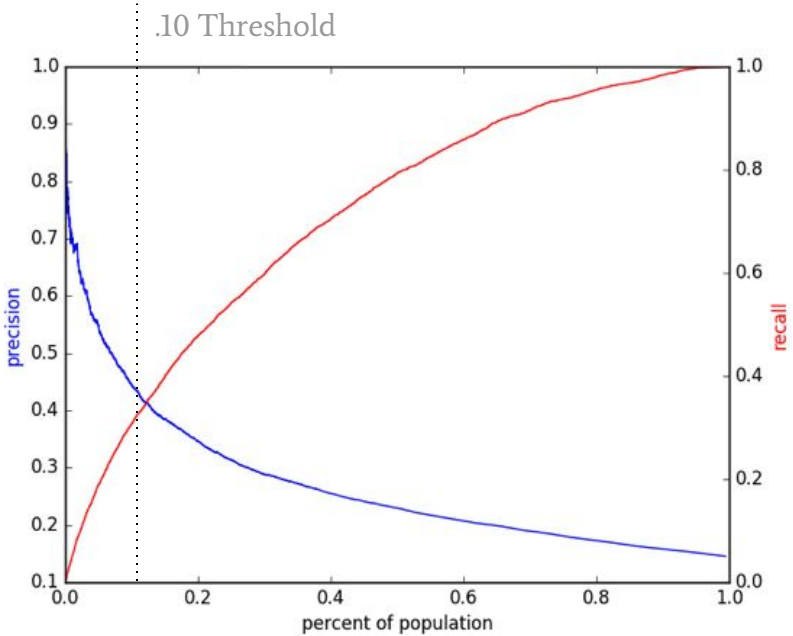


Modeling and Evaluation

Models: KNN, LR, DT, ET, NB, SVM, AB, RF

Random Forests Out-performed

1) Highest precision at 10% threshold and below
2) Highest ROC
3) Second highest recall
4) Highest precision achieved by non-RF classifiers is 0.35 (AB)



Naïve Model vs Tuned Model

Metric	Random Guess	Naive Model	Tuned Model (10%)
Precision:	0.14	0.170	0.443
Recall:	0.5	0.180	0.306
Accuracy:	0.5	0.773	0.844

Comparison of a Naive Model to Tuned

Naive: Revenue Fall Last Year = Same This Year

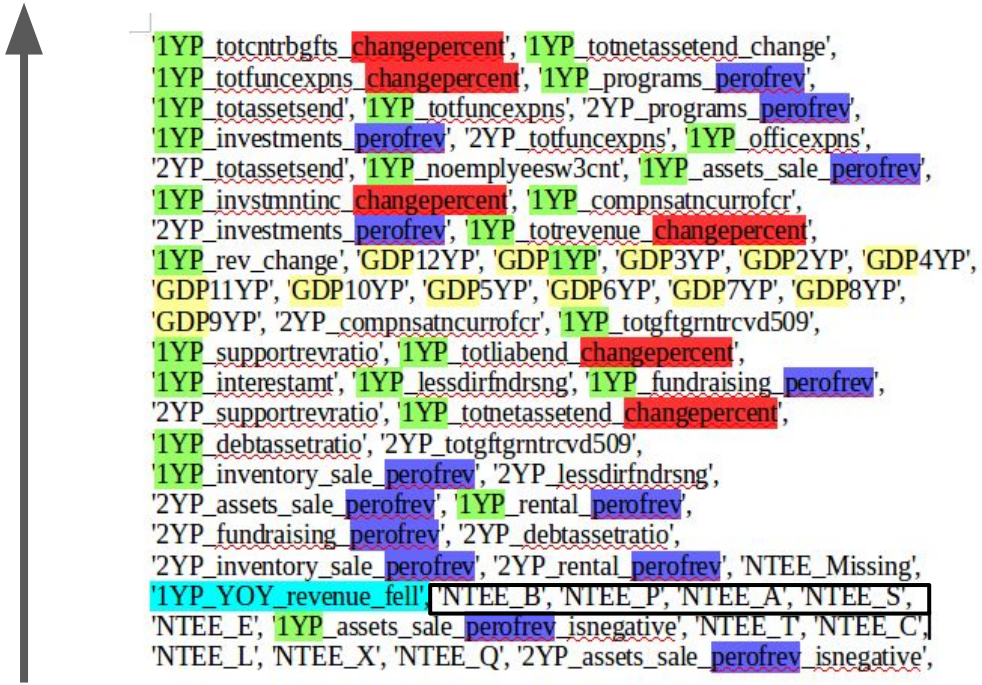
Tuned Model: Random Forest

(n_estimators=1000, max_depth=100)

Robustness Check

1) Dropped variables >40% missing (98) vs generate dummies (277)
2) Re-labeling: 20% to 40%
3) Include additional features (154->260) with 2013-2015 data

Final Model Performance: Variable Importance



Trend

Income Source

Most Recent Year

Econ & Fixed Effect

Type

Y_{t-1}

Validation on 2015

Metric	Performance
Precision (10%):	0.181
AUC (ROC):	0.541
Recall (10%):	0.127
Accuracy (10%):	0.793
10% Top Prob Threshold:	*0.563*

We greatly underperformed for 2015 compared to 2014

Likely, this is because the patterns that were predictive 2012/2013 into 2014 simply did not hold over time

Key Takeaway: We need more years to make the model robust to time period

Data Limitations

Noise in the IRS Data

Duplicated EINs, negative revenue values, missing EIN's in subsequent years

Restricted Extracts

Not all of the 990 data that nonprofits fill out are in the IRS extract, mostly financial, no governance

Missing Data

Dropped 56 features that were missing for more than 40% of the organizations; analysis shows they didn't help

Caveats & Future Work

Use Cases & Ethics

Discuss with subject experts the efficacy of a risk score based on revenue changes

Thresholds

Quantify relative costs of false positives and false negatives

More Data

Identify data sources which may be more predictive of future financial performance

Thanks!