

A SIOP Machine Learning Competition

Learning by Doing

Dan J. Putka

April 20, 2018



Acknowledgments

- Eli Lilly
- Meghan Lowery (Eli Lilly) & Alan Colquitt
- Alexander Schwall (Rhabit) and Ben Taylor (ZIFF)
- SIOP Special Sessions Committee
- HumRRO

Acknowledgments

- 12 teams representing 14 different institutions spanning academe and practice
 - Bowling Green State University
 - Capital One
 - Davidson College
 - DDI
 - Gartner
 - HumRRO
 - Korn Ferry
 - Old Dominion University
 - Rice University
 - Shaker Consulting
 - U.S. Customs & Border Protection
 - University of Georgia
 - University of Illinois
 - Walmart

Background & objectives

- Idea born out of conversations after SIOP's 2016 Big Data LEC
- Host a **Kaggle-like** machine learning competition focused on an **outcome and dataset of meaning to SIOP members**
- Provide an annual novel mechanism for **educating SIOP members** about advances in predictive modeling – *learning by doing*
- Provide the data sponsor “free” **crowd-sourced insight** into a prediction problem they face
- This year was a “by invite only” **pilot test** to clarify how the competition would work

The prediction problem and data

- **Predicting voluntary turnover at Eli Lilly**
- Data for **32,296** Eli Lilly employees active as of December 31, 2009
- Predict who voluntary left vs. who was still active as of December 31, 2014
- Criterion-related data
 - **Primary criterion** was **status** as of 12/31/14 (0 = Active, 1 = Voluntary Exit)
 - Base rate of turnover in the model training set was **18.7%**
 - Exit date and exit reason

162 predictor variables

- Unique ID, entry date, tenure as of 12/31/09
- **Demographics:** Age, gender, race/ethnicity
- **Time varying variables** captured yearly from 2004-2009
 - **Location:** City, country
 - **Job-Related:** Organizational function and sub-function, job type, pay grade, cross-functional experience
 - **Job performance**
 - Job performance ratings – overall and competency-level
 - Performance of incumbent employees' supervisors!
 - HiPo indicators
- Teams were also allowed to “engineer” new predictor variables

Ahh...the joys of real data

Longitudinal data ✓

Nesting & cross-classification ✓

Low n subgroups ✓

Missing data
(average of 37% per variable) ✓

the competition

Rules & process

- Teams submitted up to **5 entries per week** (early January – late February)
- Teams **developed** models based on a random portion of the full sample reserved for the “**training set**” ($n = 24,205$)
- Teams **applied** their models to random portion of the full sample reserved for the “**test set**” ($n = 8,091$)
 - **No criterion data** were provided to teams in the test set
- Teams emailed **entries** to the competition manager for processing
 - Simple .csv file with ID and predicted probability for each test set case

Evaluation of entries

- Entries evaluated on a version of the test set containing the turnover criterion
- **Evaluation metric:** Cross-validated area under the ROC curve (**AUC**) statistic
- **AUC** is commonly used to evaluate predicted probabilities when modeling a binary criterion

Leaderboards

- Updated **public leaderboards** were distributed on a weekly basis
- Two leaderboards were maintained for the competition
 - A **public leaderboard** based on ~50% of the test set ($n = 4,046$)
 - A **private leaderboard** based on ~50% of the test set ($n = 4,045$)
 - Public/private split designed to prevent teams from “learning” the test set
- Final standings were based on the **private leaderboard test set**

Final judging

- After the final **public leaderboard** was released, teams were asked to nominate up to 5 of their entries for “final judging”
- Final team rankings reflect the top nominated entry from each team based on its AUC in the **private leaderboard test set**

The top 4 teams will present their approach today!

overview of results

Quick summary

188

Total number of entries

15.6

Average number of entries
per team
(Range 2 to 34)

.79

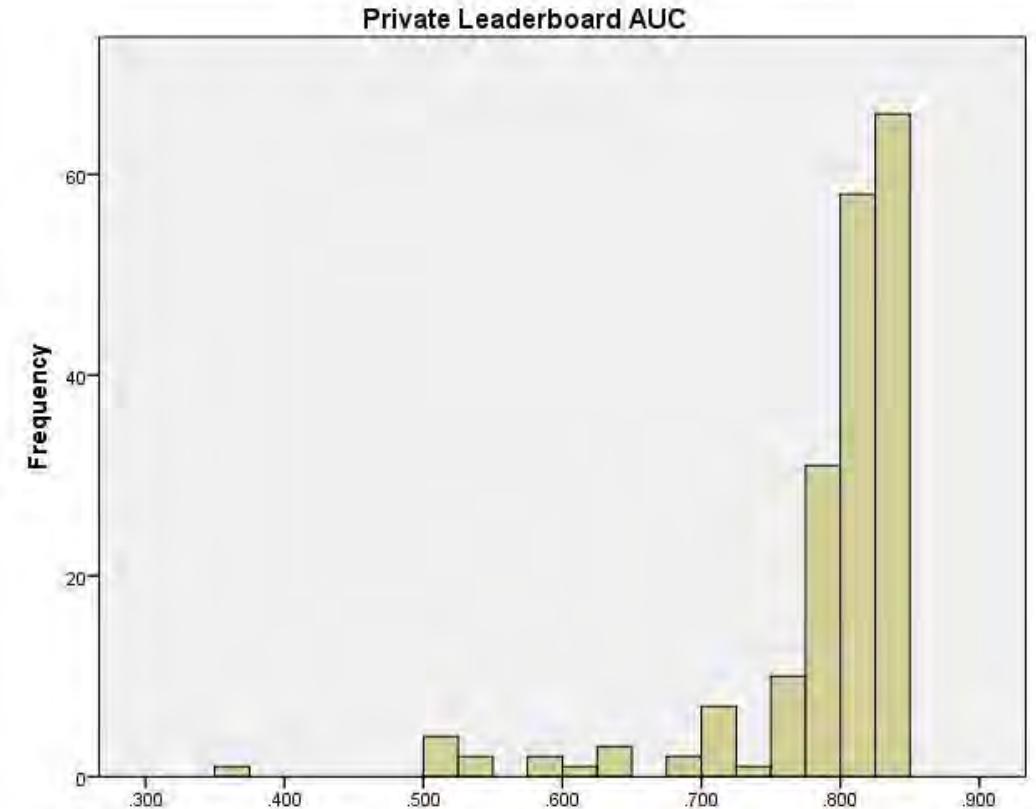
Average cross-validated
AUC across entries

24.5

Percentage of entries
reaching a cross-validated
AUC > .83

.96

Average correlation among
top 4 teams' winning entries



Final team standings

Ranking	Team	AUC
1	???	.839138
2	???	.836564
3	???	.834485
4	???	.834379
5	Team Blobfish	.830739
6	Team Procrastination	.828798
7	Crouching Tiger, Hidden Markov Model	.824695
8	Team TNTLAB	.820585
9	C3PO	.818404
10	Team Bulldogs	.816676
11	Valkyries of Shrinkage	.798661
12	Log Oddballs	.792855

- **Less than .005** separated the top 4 teams!
- **Less than .05** separated the first place team from the last place team!

drum roll please.....

#4 Byte Monsters

Private Test Set AUC = .834379

(only .003640 > #5)

#3 ROC You Like a Hurricane

Private Test Set AUC = .834485

(only .000105 > #4, sampling error smiles upon the ROC)

#2 Team DDI

Private Test Set AUC = .836564

(only .002079 > #3)

#1 An Enriching Meal

Private Test Set AUC = .839138

(only .002574 > #2)

Team presentations

- How did your team approach the prediction problem?
- Describe your winning solution.
- What were the most important predictors of turnover?
- What lessons did you learn?

**The top 4 teams' presentations and code are in
SIOP's *Document Library!***

team presentations