2019 SIOP Machine Learning Competition

Natural Selection

This isthat place

The Team

- Josh Allen, Walmart Selection & Assessment
- Matthew Arsenault, Walmart Selection & Assessment
- Blaize Berry, Walmart Labs
- David Futrell, Walmart Selection & Assessment

Key Support (Content Experts)

- Fred Macoukji, Walmart Selection & Assessment
- Josh Rogers, Walmart Global Talent Management
- Meline Schaffer, Walmart Global Talent Management

Approach

- **Feature engineering**
  - Constructed features based on our knowledge of the *content*, creating word lists that captured the behaviors related to the personality variables
  - Created full set of text-based statistical features typically used in NLP
- **Multiple modeling strategies for generating predictors**
  - Simple linear combinations of statistics from Word Lists
  - Machine Learning
  - Deep Learning
- **Trial, error & tracking**
  - Train/Test Splits
  - K-fold Cross-Validation
  - Many submissions to the development set
  - Monitored potential overfitting/shrinkage
- **Averaged predictor scores from best models**

Feature Engineering (Word Lists)

- Team of Content Experts read a subsection of the train responses
- Created lists of words that should be related to trait, separated by “high” and “low”
- Developed a variable with counts of low and high words

Low Agreeableness	High Agreeableness	Low	High
I would find out how much time and money my colleague has invested so far in her vacation plans. If it is a lot I would demand that I get the week, unless she is willing to compensate me for my trouble.	I would schedule my vacation for another week. My colleague may have important plans for that week. So, I would allow them to take that week off and schedule my vacation for another time.	demand	schedule for another week, another time
Change my dates and then talk smack about them when they leave. Fill up in rage there that week while overdosing on black coffee. When they come back just stare at them at there cube.	I would go ahead and switch my vacation. I never make concrete plans before getting approved from work. So it would be no problem unless it is something like a family reunion or something. I would then try to compromise with the other employee.	Rage, stare	would switch, compromise

Item-Level Analysis

- Coding lead to a list of 250+ words/short phrases
- Created a variable for each word in each item
- Explored correlation for each word by trait

	O_1	O_2	O_3	O_4	O_5
not interested	-0.039		-0.023		-0.102
risk	-0.038	-0.021	-0.010	0.043	-0.102
depend	0.007	-0.041	0.018	-0.003	-0.075
difficult	-0.044	0.020	0.022	0.020	-0.066
concerned	-0.036		-0.053	0.042	-0.063

- Created 4 “features” – High 1 item, low 1 item, high across items, low across items
  - Added 2 additional features – “Go” and “Not Go”

Word List Composite Creation

- Submitted the 6 features separately to the dev data
- Retained the best predictors and created composites
- Added in additional elements (e.g., word counts, average length of word) and created final composites

Word List Test Submission

- Small drop from dev to test

	O	C	A	E	N	Mean
Dev	0.273	0.121	0.280	0.288	0.193	0.231
Test	0.146	0.164	0.237	0.279	0.222	0.209
Difference	-0.127	0.043	-0.043	-0.009	0.029	-0.022

- Moderate correlation with other submissions

	O	C	A	E	N
Words vs. DL	0.386	0.317	0.232	0.374	0.319

Selected Statistical Features for Machine-Learning Models

FEATURE	1	2	3	4	5	1-5 Concatenated
Number of words	X	X	X	X	X	X
Number misspelled	X	X	X	X	X	X
Percent mispelled	X	X	X	X	X	X
Number of characters	X	X	X	X	X	X
Number of characters	X	X	X	X	X	X
Number of syllables	X	X	X	X	X	X
Sentence Difficultly						X
Average syllables/word						X
Average sentence length						X
Average Word Length	X	X	X	X	X	X
Flesch Reading Ease Score						X
Flesch grade level						X
Linsear write score						X
Dale Chall score						X
SMOG score						X
Coleman-Liau score						X
Number of grammar errors						X

Best Machine-Learning Models

- **Conscientiousness:** XGBoost
- **Neuroticism:** Auto Machine Learning (TPOT): LassoLarsCV
- **Extraversion:** Auto Machine Learning (TPOT): ExtraTreesRegressor
- **Openness:** Auto Machine Learning (TPOT): ElasticNetCV
- **Agreeableness:** Auto Machine Learning (TPOT): ElasticNetCV

Deep Learning

- ELMo: Deep, conceptualized, character-based word representations
- Can be used instead of GloVe, Word2Vec or other pre-trained vector representations
- This generated our best Agreeableness predictor

Getting to the Final Submission

Five Submissions

1. Word List
2. Machine Learning
3. Deep Learning
4. Average predictor scores of the two best predictors for each scale
5. Best of first four submissions (Openness was Word-List Only)

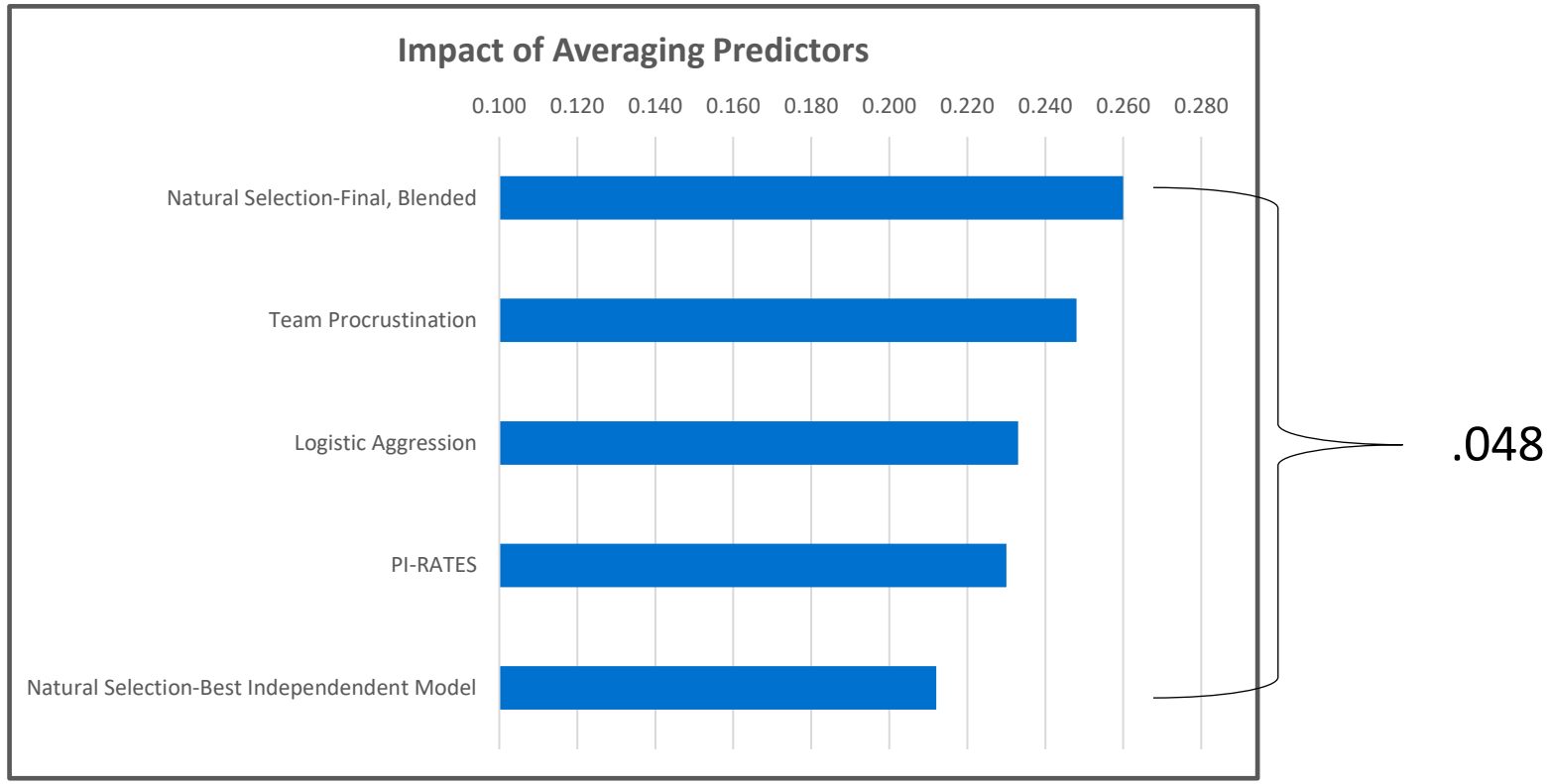
Submission	Openness	Conscientiousness	Agreeableness	Extraversion	Neuroticism	Mean (Submission)
1. Word List (WL)	**0.146**	**0.164**	0.237	**0.279**	**0.222**	0.209
2. Machine Learning (ML)	0.110	0.095	**0.303**	0.228	0.196	0.186
3. Deep Learning (DL)	0.126	**0.128**	**0.327**	**0.279**	**0.201**	0.212
4. Combining Best 2 of first 3	0.126	0.179	0.379	0.337	0.260	0.255

Combined Predictors:	WL	Mean(WL, DL)	Mean(ML, DL)	Mean(WL, DL)	Mean(WL, DL)	Mean (Submission)
5. Final	0.146	0.179	0.379	0.337	0.260	0.260

Our best trick:

The boost we obtained from averaging the predictor scores across models was greater than the spread across the top 4 teams.

Our best “independent” model would have finished 4th



How to Win

Collaborate

- You're not smart enough to do this by yourself
- Even if you are, you don't have time

Pay attention to the content

- You are I/O psychologists and this is a psychological measurement problem: This should give you an advantage over pure data scientists

Data Science

- Pipelining, iterative approaches that test multiple models with grid-searches of parameters have a much better chance of finding a good solution than traditional methods

Grind

- Prepare to spend a lot of time collaborating, brainstorming, running models, and poring over results