

2020 SIOP Machine Learning Competition Overview

Summary

This year's competition is all about creating algorithms that predict important, competing outcomes while minimizing bias in a real-world use-case: hiring decisions. Using closed-ended pre-employment assessment data, can you make hiring decisions that not only drive key business outcomes but also ensure fairness? The top 4 teams with the most accurate and fairest predictions (on the Private Leaderboard) will be declared the winners and will be asked to present at the 2020 SIOP Special Session in April, 2020.

Introduction

All important decisions in life involve trade-offs. A potential mate may be stunningly attractive, but what if they have an incompatible? You might find the home of your dreams...but not in the neighborhood you want. Making great decisions requires balancing competing criteria and finding the optimal outcome. Hiring is no different. To hire effectively one must not only maximize outcomes for the business but also comply with legal requirements. This is often called the “diversity-validity trade-off.” This year's competition will require you to maximize business impact by developing an algorithm that simultaneously maximizes job performance and reduces turnover while minimizing bias.

2020 Competition Task

Your task will be to create the most accurate and fair method to select employees. You will use data from an actual assessment battery data from a large company and weight the various inputs to predict both turnover and job performance for an entry-level position. Your solutions will also be penalized based on the amount of unfairness (adverse impact) is created. You will train your algorithms on a large training set and submit your hiring decisions on two separate holdout samples. Your score will be calculated based on how well you determined who the true top performers were, whether your selected employees stayed with the company, the degree to which your selected employees meet BOTH those criteria and finally, on the fairness of your selection decisions. **Your selection ratio will be 50%**, meaning you will have to decide which half of dev and test set to hire and which half will be rejected. Think of this as similar to providing a “**Go**”/“**No Go**” decision on each candidate. Please see the evaluation tab for specifics on the calculation of your final score.

The Data

Data is provided to team leads and any team member emails you list when [registering on our Google form](#). Data will be shared via email after registering.

The data will be a slight variant of a true selection system. You will be provided with a training set that contains 7,890 respondents that have both performance (column = **High_Performer**) and turnover data (column = **Retained**) along with a protected status (column = **Protected_Group**). You will also be provided with an additional 36,212 respondents that have turnover data (Retained), but no performance data. Both the dev and test sets contain 2,250 candidates with all columns except the criterion. All sets were developed via stratified random sampling of the criterion and protected group status. For specifics on the data and examples of item types please view [this document](#).

Frequently Asked Questions: FAQs

Ranking of Submissions

The competition this year will use the following methodology to rank various submissions:

$$\text{Final_score} = \text{Overall_accuracy} - \text{Unfairness}$$

Overall_accuracy=

Percentage_of_true_top_performers_hired*25+

Percentage_of_true_retained_hired* 25+

Percentage_of_true_retained_top_perf_hired* 50

Unfairness =Absolute_value(1 -Adverse_impact_ratio)*100

Definition of Terms in Ranking Method

Final_score =The final score for a given set of predictions. This score will determine the rank on the leaderboard (both public and private) and is derived by the Overall_accuracy minus the Overall_unfairness.

Overall_accuracy = A score that is a weighting of what matters to the business. The hit rate of true top performers is given a weight of 25%. The hit rate of true retained hires is given a weight 25%. The hit rate of those selected that are both top performers and retained is given a weight of 50%.

Percentage_of_true_top_performers_hired = The ratio of top performers you hired (based on your predictions) out of the possible top hires you could have hired. Selected top performers divided by top performers in the data set. This constitutes 25% of the overall accuracy score.

Percentage_of_true_retained_hired = The ratio of employees you hired (based on your predictions) that stayed with the company out of the possible retained hires you could have made. Selected retained employees divided by retained employees possible. This constitutes 25% of the overall accuracy score.

Percentage_of_true_retained_top_performers_hired = The ratio of employees you hired (based on your predictions) that stayed with the company and were top performers out of the possible top performing retained hires. Selected top performing retained employees divided by top performing retained employees in the data set. This constitutes 50% of the overall accuracy score.

Unfairness = How much a selection decision deviates from fair (protected group selection ratio equal to the non-protected group selection ratio). Specifically, 1 minus the absolute value of the Adverse_impact_ratio multiplied by 100. The absolute value of this calculation so that a protected group over non-protected selection ratio of 1.20 is equal to that of 0.80 (both resulting in 0.20). Unfairness in either direction is treated equally. Lastly that score is multiplied by 100 so that it is on the same scale as the Overall_accuracy score (0.20 becomes 20 and is then subtracted from the Overall_accuracy).

Adverse_impact_ratio = The protected group selection ratio (protected hired over the protected in the applicant pool) divided by the non-protected selection ratio (non-protected hired over the non-protected in the applicant pool). Selection_rate_of_protected divided by Selection_rate_of_non_protected.

Selection_rate_of_protected = Protected hired over the protected in the applicant pool

Selection_rate_of_non_protected = Non-protected hired over the non-protected in the applicant pool

Example of Ranking

Here is an example of how the numbers work together to produce a Final_score. All numbers are imaginary.

Input numbers:

Test data set number of cases= 2250

Number selected as hired= 2250 * .50 (selection rate of .50)

True top performers in the hold out set = 650

From your selected hires, how many are top performers= 500

Percentage_of_true_top_performers_hired = 500/650

True retained employees in the hold out set = 1125

From your selected hires, how many were retained = 750

Percentage_of_true_retained_hired = 750/1125

True retained top performing employees in the hold out set = 500

From your selected hires, how many were retained top performers = 375

Percentage_of_true_retained_top_performers_hired= 375/500

Selection_rate_of_protected = 0.81

Selection_rate_of_non_protected = 0.91

Adverse_impact_ratio = 0.89

Overall_accuracy = ((500/650)*25)+ ((750/1125)*25)+ ((375/500)*50)

(.77 * 25) + (.67 *25) + (.75 *50)

19.2 + 16.7 + 37.5

73.4 out of a possible 100 points for Overall_accuracy

Unfairness = Absolute_value (1 – 0.81/.91) *100

Absolute_value (1 – 0.89) *100

(.11) * 100

11

Final_score = Overall_accuracy – Unfairness

73.4 – 11

62.4

The Final_score that would be used on the leaderboard is 62.4