A Predictive Approach to Understand How to Get a Promotion

PURDUE UNIVERSITY | Krannert School of Management

Victoria Bergamini, Lauren Ostrowski, Justin Soeldner, Luis Gil, Hana Abdel-Dayem, Matthew A. Lanham
Purdue University, Krannert School of Management
vbergami@purdue.edu; lostrow@purdue.edu; jsoeldne@purdue.edu; gil0@purdue.edu; habdelda@purdue.edu; lanhamm@purdue.edu

ABSTRACT

We develop both interpretable and non-interpretable predictive models to predict how to get a promotion. The motivation for our study is that promoting driven and capable employees is a necessary step that every company must take to ensure long-term success. Knowing who a company expects to promote early can streamline company growth. Our interpretable model provides a way for employees and policy-makers to understand what attributes historically have made an employee more promotable than others. We posit our research may provide employees a novel way to strategically identify controllable characteristics (e.g., education, KPIs met, performance evals, etc.) they might work on to improve their odds of being promotable. Likewise, we believe our model provides policy-makers a way to re-examine how promotions have been made so they are made more fairly in the future. Our non-interpretable model shows how we were able to increase predictive accuracy at the expense of explainability.

INTRODUCTION

Nowadays, employees are highly motivated by incentives, such as promotions, raises in wages, etc. To ensure a low rate of employee turnover & higher satisfaction, companies must allow for room to grow and recognition of skills. In return, employers seek capable employees with high leadership and delivery skills. High motivation, capability, and reliability are just a few of the many important traits that employers look for in future leaders for the company.

We want to determine the best model to predict who is more likely to get a promotion based on their features. Below, the graph determines that turnover trends demonstrated an 8.3% increase over 2018 and 88% increase since 2010. (The Work Institute 2020 Retention Report,2020)

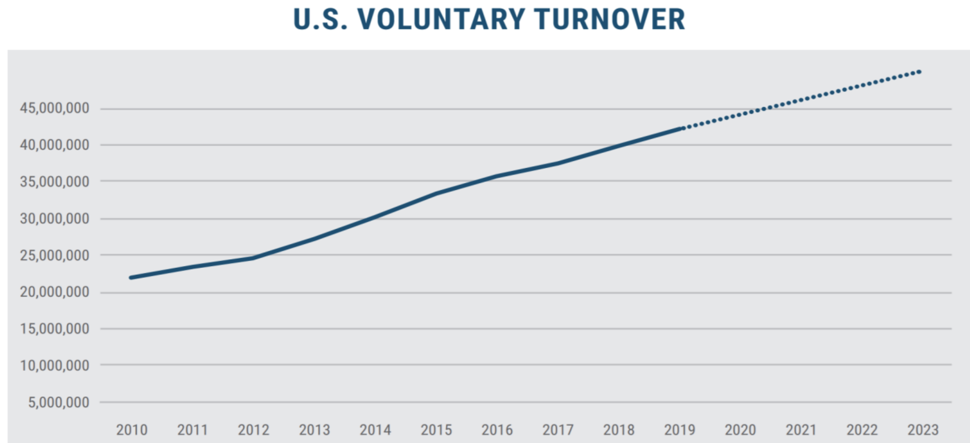

Fig 1. U.S. Voluntary Turnover

Our Research Question:

- What features make an employee more likely to get a promotion in a company?

LITERATURE REVIEW

In the following research studies, each performed statistical modeling using a linear model and regression analysis. In our study we combined the use of linear modeling with boosting to get a more accurate prediction. Our study is unique because we compare the previous methods used and combine them with an ensemble model.

Study	Linear Regression	Logistic Regression	Ensembled Model
Our Study, 2020		✓	✓
Bialmi Z., Asroalah L., 2018	✓		
Drageni,L., In Sue,Q., Vankatesk, P., Tesluk, P.,2011		✓	
Ghani. N., Yonus, N., Bakry,N, 2016	✓		
Frenkel,S., Bednall, T.,2016		✓	
Bonet, R.,Cappelli, P.,Hamori, M.,2018	✓		

METHODOLOGY

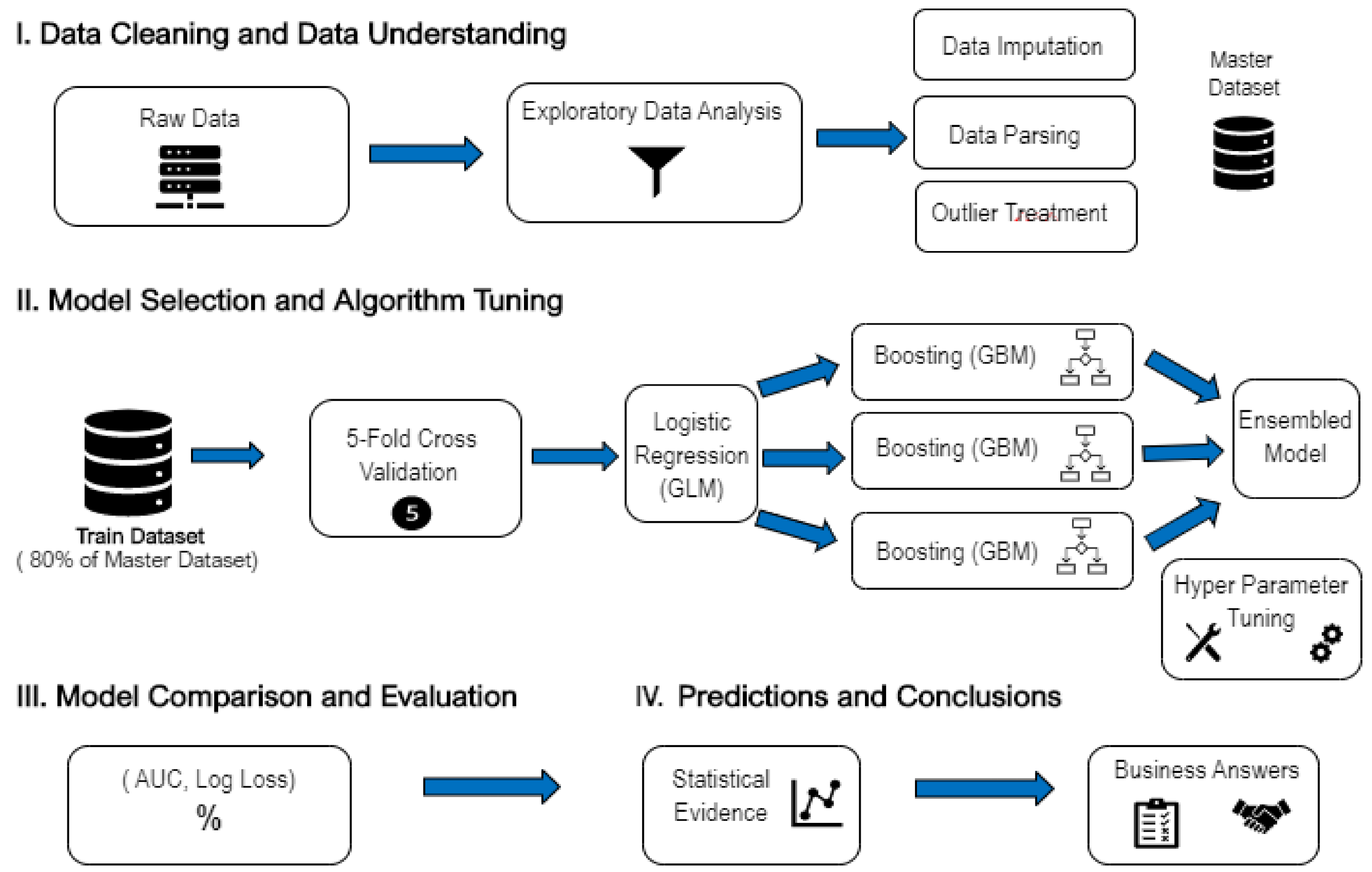

Fig 2. Study Design

Classification Problem:

The caveats we encountered in our study include the following:

- Imbalanced data set
- Too many features (13 variables)

To address these concerns, we rebalanced the training data, and used the R h2o library for variable selection.

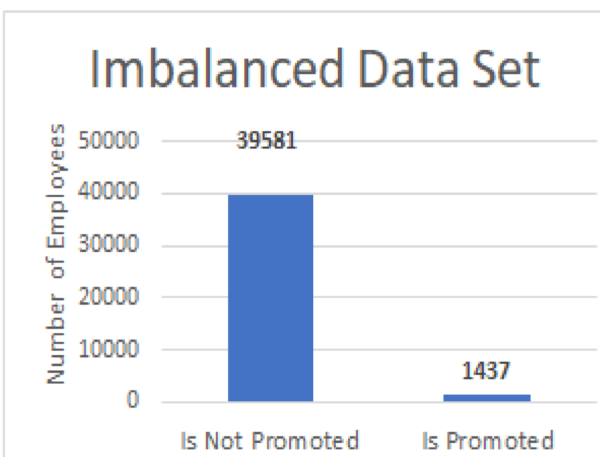

Fig 3. Y Variable

STATISTICAL RESULTS

In our study we performed four different models including three individual gradient boosting models and a stacked ensemble of those three. We found based on an AUC of 0.902 and a Log Loss of 0.165 that the Stacked Ensemble performed the best. Therefore, our results are based on this model. We referenced the confusion matrix for our test set to determine how well our model can predict those who should be promoted and those who are not promoted.

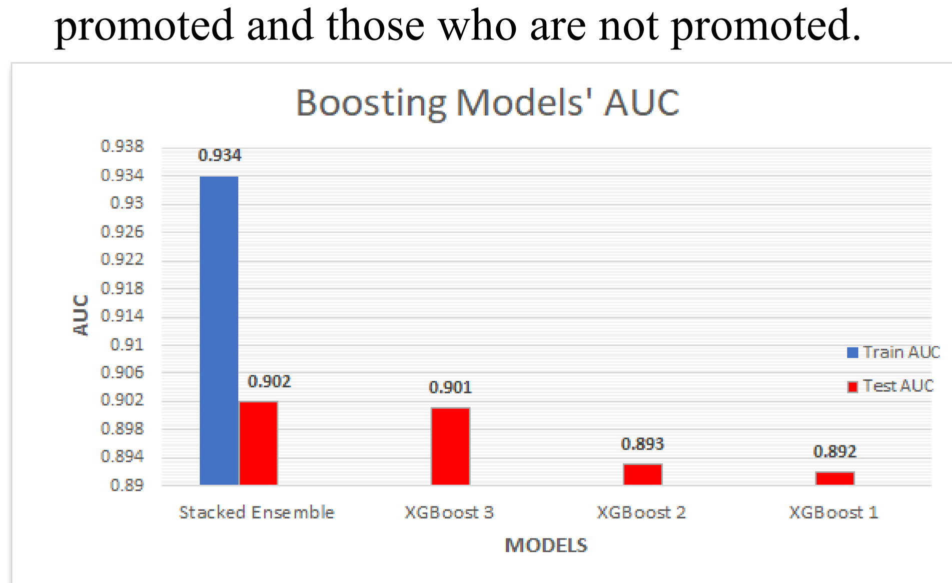

Fig 4. Model Comparison

Promotion	Is Not Promoted	Is Promoted
Is Not Promoted	39,581	493
Is Promoted	2,334	1,437

Fig 5. Confusion Matrix

EXPECTED BUSINESS IMPACT

Identifying promotable candidates early has many advantages to firms as well as the candidates themselves. The highest performing employees based on training score, on average, come from the Marketing and Operations. Departments. Focusing on employees with these attributes at an early point in time could shorten the promotion candidate list to focus on candidates that are predicted to be promoted.

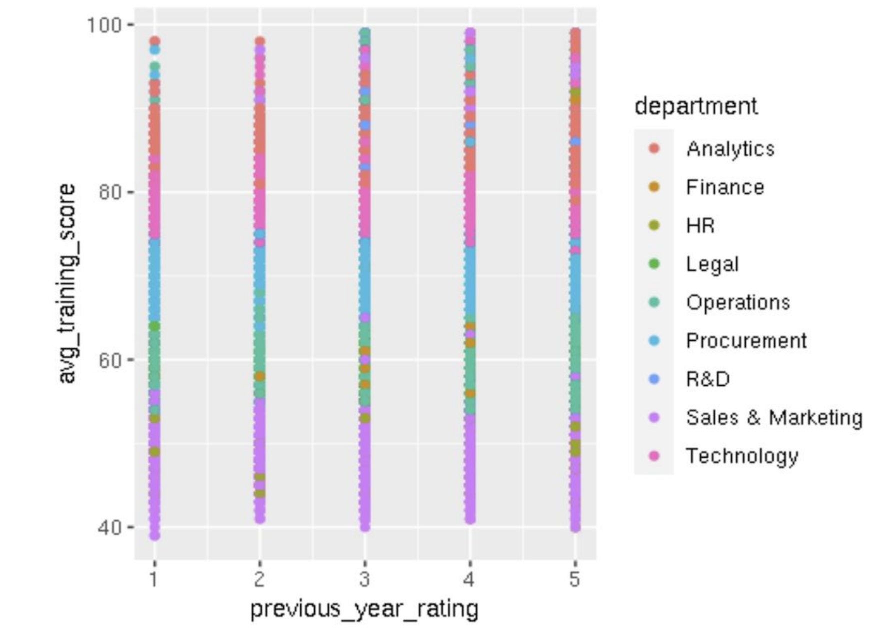

Fig 6. Training Score by Department

	Actual						
	Cost: 30%	0	1	Failure Rate	Success Rate	Expected Benefit	
Predicted	0	39581	493	1.23%	98.77%	29.26%	
	1	2334	1437	61.89%	38.11%	-7.14%	22.10%

Fig 7. Predicted Savings

The US Department of Labor estimates that hiring the wrong candidate for a job costs the company about 30% of that employee's salary on the first year as a result of his or her underperformance. The same metric will be used to evaluate the benefit of promoting the right employee (gain from a true positive). Using this predictive model alone for promotional decisions, firms are expected to gain 29.26% as a result of identifying true negatives and lose 7.14% as a result of false positive predictions leading to net savings of 22.1% of the respective employee's salary on the first year. Saving time on interviewing mostly candidates predicted to earn a promotion leads to increased cost and time efficiency.

CONCLUSIONS

- The three most important features for getting a promotion include:
  (1) Average training score, (2) Marketing department, and (3) Operations department

Using several parameters and multiple models, we have found that **early promotion** is extremely beneficial in retaining employees for longer periods of time. Many companies struggle to retain their top employees in today's world leading to a struggle to always be on the lookout for new employees. **Promoting early and often helps prevent turnover** along with understanding your industry and implementing strategies based on this. **Using a multi-layered predictive model to consider promotion candidates, firms can save time and money by shortening their candidate list and finding the right employee to for the job**.

Our predictive model does include some limitations, however. First, while a GLM model does accurately predict the binary classification problem, its flexibility does cause more variance within the bias-variance tradeoff. As mentioned previously, we ran into two problems during our project one was having too many features and the other was an imbalanced data set. We addressed the data dimensionality problem by using h2o which allowed us to focus on a fewer number of variables. In order to address the imbalanced data set, we rebalanced our training data prior to training and evaluating our models.

ACKNOWLEDGEMENTS

We would like to thank Professor Matthew Lanham, Xinyu Wang, Theo Ginting, and our graduate student mentors Tianrong Li and Penghao Deng for their guidance and support on this project.