Project Final Presentation

Identifying Key Variables to Explain Employee Performance

MGT 6203 | Fall 2024 | 11/29/2024

Ben Mascott (Analysis Co-lead), Erick Ordonez (Project Manager),
Jerry Fang (Data Visualization Lead), Isabella Reyna (Analysis Co-lead)

Georgia Tech

Introduction and Problem Statement

Business Problem

Our team collaborated closely with the Human Resources (HR) department at a firm to address challenges in improving employee performance but given limited time and resources. HR needed to find a data-driven approach to identify key factors that best explain the variability in employee performance and prioritize initiatives by developing a model to predict performance.

By utilizing linear and non-linear predictive models, we analyzed the employee data to uncover actionable insights that allowed HR to focus on key factors, while also reducing time and money spent on less significant factors. This approach has not only optimized its time and costs but also enabled HR to produce strategies to enhance employee productivity, which helped them to foster a more engaged, productive, and efficient workplace.

Project Objectives

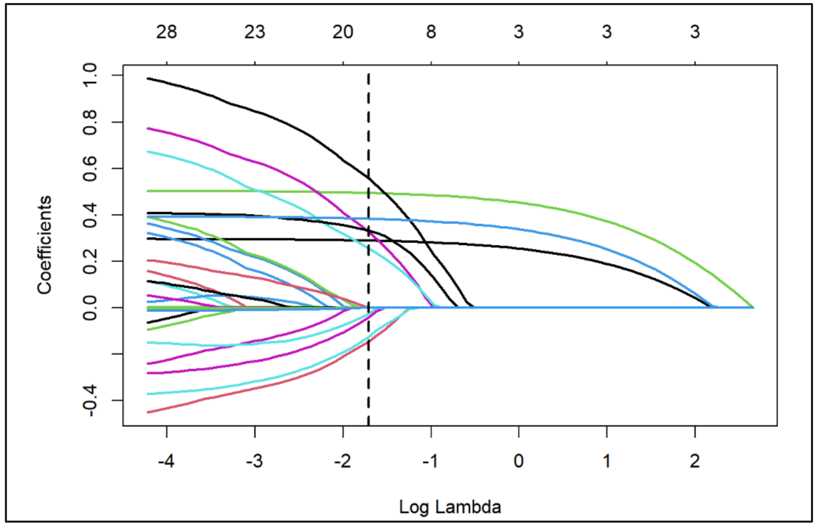
1. Identify significant factors, such as years of experience, job satisfaction, and education level, that significantly impact employee performance, allowing HR to design targeted support and intervention efforts aimed at low-performing employees.
2. Develop predictive models to help HR to allocate their resources efficiently by identifying employees who are at risk of being low-performers and provide proactive support to them.

Georgia Tech

Key Insights and Findings

An **80/20 train/test split** was used. **5 models** were trained on the same training set, and their respective performance was evaluated on the test set. The Mean Squared Prediction Error (**MSPE**) and Mean Absolute Error (**MAE**) for each model was calculated using the **test set**. The results are shown below.

Model	MSPE	MAE	# of Predictors (including dummies)
Full Model	101.201	7.940675	29
Stepwise Model	101.1876	7.932867	8
LASSO	101.094	7.936827	15
Decision Tree	124.1864	8.812438	29
Random Forest	105.2443	8.102373	29



LASSO Coefficient Paths

The Stepwise Model is the Selected Model.

Performance_Score ~ Experience + Work_Hours_Per_Week + Age + Education_Level + Annual_Bonus

- Forward, Backward, and Forward-Backward selection all selected the same subset of predictors (shown above). The model with these predictors will be referred to as the "stepwise model" for brevity for the rest of this presentation.
- The Stepwise Model has the **lowest MAE**. Although LASSO has lowest MSPE, an ANOVA test to compare the two models showed the two models are not statistically significantly different at a 0.05 significance level. Hence, the more **parsimonious model** (fewer predictors) was chosen.

Business Impact and Implications

- **Practical Implication:** Since only 5 of the 13 variables are in our dataset are in the selected model, HR can save time and money by only collecting data for 5 variables instead of 13.
- **Strategic Value:** By not having to devote resources to collecting data for unused predictors, HR can focus more resources on ensuring *data integrity* and collecting data *more frequently* to update their model. A more parsimonious model is also *easier to interpret*.
- **Statistical Evidence**: A partial ANOVA F-test comparing the full regression model to our selected model resulted in a p-value > 0.05. Thus, we could not reject the null hypothesis and concluded the full model was plausibly similar, in terms of explanatory power, to the stepwise model, which uses only a subset of predictors in the full model.

- **Practical Implication**: Experience is the most important predictor, but *existing* employees cannot increase their number of years of experience overnight.
- **Strategic Value:** Recommend HR investigate ways of retaining and hiring experienced employees to improve experience in the s*hort-term*. For a *longer-term* impact, create opportunities for more-experienced employees to share their knowledge with less-experienced employees.
- **Statistical Evidence**: In the stepwise model, Experience was the first variable added to the minimum (intercept-only) model since its marginal regression had the lowest AIC (39526). Experience was the first variable branched from in the decision tree.

- **Practical Implication:** Allocate resources to help less experienced, younger employees, who work less than average
- **Strategic Value:** Focusing on these employees may help the HR department improve their overall employee performance scores, since these are the three strongest predictors and all three variables exhibit a *positive linear association* with the response.
- **Statistical Evidence:** In the selected model (stepwise regression), Experience, Work hours Per Week, and Age are the only variables significant at the 0.01 significance level, in the presence of all other variables. The decision tree chose the same three variables as its top branches.

Georgia Tech

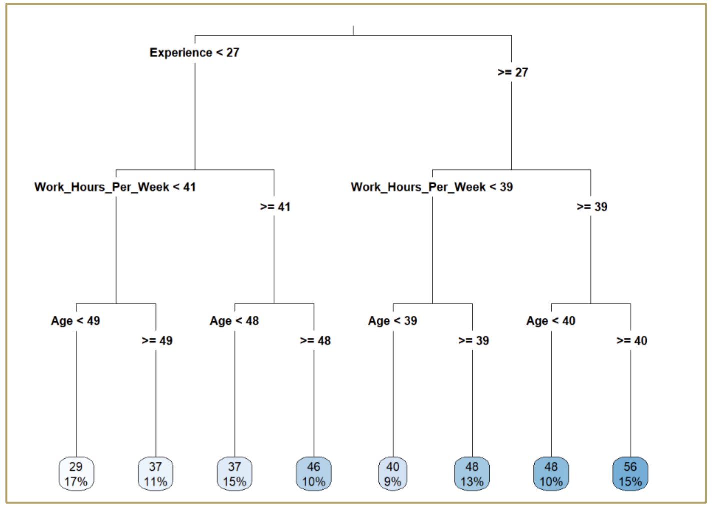
Recommendations

Actionable Recommendations:

- ***Recommendation #1***: Create a mentorship program that pairs experienced employees with early-career or less experienced staff. Goal: Enable less experienced employees to benefit from the knowledge and guidance of their more experienced counterparts. By asking questions, seeking advice, and learning best practices, less experienced employees can enhance their skills and improve their performance scores, even with fewer years of experience.
- ***Recommendation #2:*** Prioritize candidates with more years of experience when hiring new employees. Goal: Since the years of experience for current employees cannot be quickly increased, placing greater emphasis on experience in the hiring process can help improve overall employee performance in the short term.
- ***Recommendation #3:*** Relaxing overtime gates. Goal: incentivize overtime to increase the number of work hours per week
- ***Recommendation #4:*** Tuition Assistance Program. Goal: Incentivize pursuing advanced degrees with tuition assistance programs

Implementation Steps:

1. Propose our recommendations and implementation plan (next slide) to management.
2. Work with management to define a timeline for rolling out interventions and determining how much budget and labor will be allocated towards the project.
3. Select recommendations to pursue based on available time and resources. Recommendations above are given in ***priority order***, based on explanatory power of the predictor variables, as supported by our random forest, decision tree, LASSO, and stepwise models.
4. After selected recommendations are approved by management, carry out recommendations sequentially in priority order (if time allows). Re-fit the stepwise regression model after each recommendation (see next slide for details).



Decision Tree Nodes

Implementation Plan

- Buffer between phases to analyze data from completed phase and develop data collection strategy for the next phase.
- **Re-fit model** and **re-assess predictive performance** after each phase.
- At each phase, account for **selection bias** if programs introduced for the phase have an opt-in policy.
- **Multi-phase approach** minimizes risk and reduces confounding variables that may arise if all solutions implemented simultaneously. It also provides **flexibility** to allow for changes in budget and available resources.
- Modular phases allow for **updates to plan** depending on which recommendations management selects.

SCHEDULE:

Jan2025-July 2025	Aug 2025	Sep 2025-Dec 2025	Aug 2025	Feb 2026-Aug 2026	Sep 2026		Feb 2026
PHASE 1	DAC	PHASE 2	DAC	PHASE 3	DAC		1 Yr F/U

DAC = Data Analysis and Collection
1 Yr F/U = One-year follow-up with management to share findings and discuss path forward

Phase 1: Rollout Recommendation #1 and #2 (Establish Mentorship Programs, increase weight of experience when hiring)
- Variables addressed: Experience, Age
- Resources: Request support from experienced employees who have capacity to serve as mentors

Phase 2: Rollout Recommendation #3 (Relax Overtime gates to incentivize increasing work hours per week)
- Variables addressed: Work_Hours_Per_Week, Annual_Bonus
- Resources: Request managers in each department review the annual bonus process/criteria with employees

Phase 3: Rollout Recommendation #4 (Tuition Assistance Program)
- Variable addressed: Education_Level
- Resources: Request Finance allocate resources to incentivize continuing education programs

Challenges and Considerations

Challenges	Mitigation Strategies
Inability to make causal claims. Since the dataset consists of observational data and not data from a controlled randomized experiment, we cannot make causal claims. We can only report association between the predictors and response.	Implement interventions using a phased-approach. This will allow HR to better assess if key predictors may have a causal relationship with the response, since each phase will only focus on one or two predictors.
Risk of Age-Based Discrimination. While age has a positive relationship with performance score, we don't want to take away resources from older employees to improve the performance of younger ones.	Early-Career programs will not have age limitations. Rather, they will be targeted at employees new to the company or employees who underwent a recent career change.
Time is needed to see potential changes in performance. Different predictors may have a different timeline before there are noticeable changes in performance.	For employees who participate in tuition assistance programs, we recommend tracking their performance over time (collect data at the end of each semester) to see if there is a positive trend in the individual's performance score.

Georgia Tech

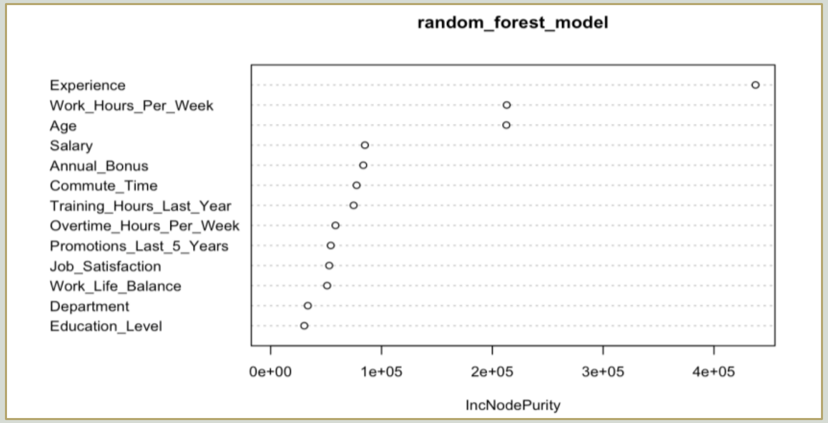
Future Directions

Further Analysis

- **Collect More Specific Data:** For example, for "Experience", not just "Years of experience", but also years of relevant experience, and years with same company.
- **Cost of required features:** LASSO had a better MSPE than the stepwise model but requires more features. Explore if the cost of data collection for more features is worth it for a slight increase in metric performance.
- **Reformat Response Variable:** Set a performance score threshold and reformat performance score to "High Performer" and "Low Performer". This would help if budget limitations means the goal changes to increasing the number of high performers.

Long Term Improvements

- **Cost associated with improving employee performance:** Understand at what point does cost of investing in employees break even with the added value of improved performance.
- **Employee Performance over time:** Explore how employee performance varies over time. This can help measure the effectiveness of interventions.
- **Adding More Features:** Since we did consider Non-Linear models (Random Forest), if we decide to add more features, we will also be able to handle non-linear relationships.



Conclusion and Summary

Recap of Key Points:
We propose four recommendations and a multi-phase plan for implementation (see slides 5 and 6)
Explaining Performance: Linear regression, regularized regression, decision tree, and random forest models consistently identify experience, weekly work hours, and age as the most influential factors in explaining employee performance, even when accounting for all other variables. We recommend that HR prioritize these variables when hiring new employees and designing programs to support current staff. This focus aligns with the goal of improving overall employee performance.
Predicting Performance: We recommend using the stepwise model for predicting performance, as it has the lowest Mean Absolute Error (MAE) among the tested models. Additionally, the stepwise model is significantly more streamlined than the next best option, the LASSO model. This efficiency reduces the amount of data required for model re-fitting, saving both time and resources.

Concluding Thoughts:
Recommendation #1 focuses on mentorship for existing employees and recommendation #2 is concerned with the hiring processes for new employees, thus providing both short-term and long-term solutions for improving experience.
Finally, Recommendations #3 and #4 target the remaining predictors in the selected model.

- If there is not enough resources to implement all recommendations, then after implementing #1 and #2, we recommend HR pursue #3 if the company more focused on performance improvements in the short-term , and #4 if the company's strategy is geared towards long-term improvements.

In summary, we have provided HR a list of key variables used to explain employee performance, a stepwise regression model for predicting performance, specific recommendations with expected short-term and long-term impacts, and a multi-phase plan for rolling out recommendations. We believe with these tools, HR can strategically allocate its resources, saving time and money, to improve employee performance across the company.

Georgia Tech