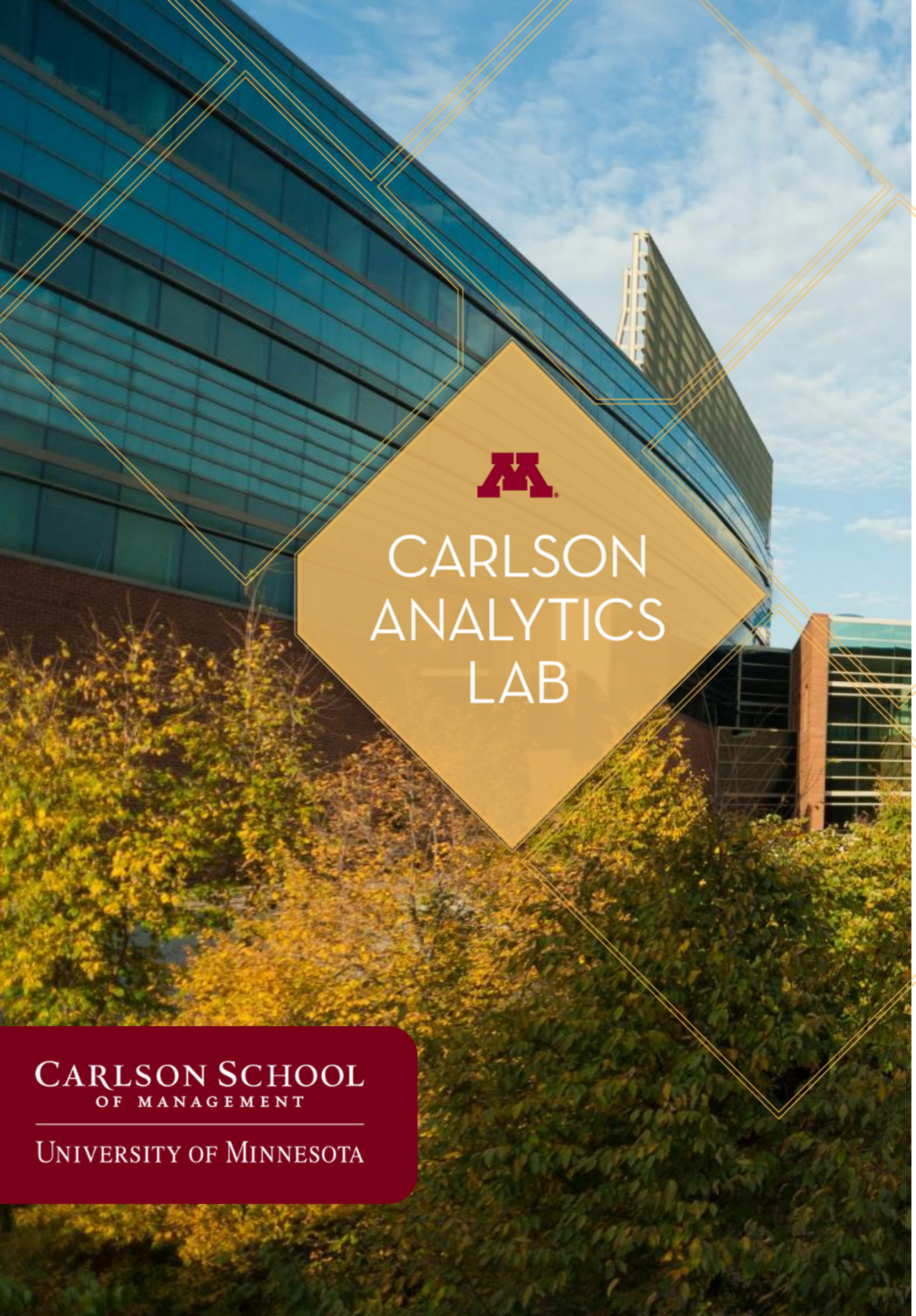


UMN OHR project

Predicting employee departures

June 23rd 2020

Kevin Grady, Wendy Lu, Anthony Meyers, Danny Moncada, Claire Ryan, & Jonathan Watkins

Context

- Payroll costs are the University of Minnesota's largest single expense.
- There is a general push to reduce administrative costs/overall costs across the board.
- Turnover is expensive: some estimates put the cost of losing one employee at 1-2 times the employee's salary.
- HR is decentralized across the university, leading to disparate analytics capabilities and different budget and retention challenges. Currently, there is no standardized process for reducing turnover.
- OHR has been building up their analytics efforts over the past two years.
- Given this, OHR sees an opportunity to leverage predictive modeling to identify employees who are likely to leave so they can intervene and potentially reduce turnover, leading to cost savings.

Defining the Problem

- OHR has tasked us with building a model that predicts which employees are likely to leave their jobs, with a lead time of *at least* two months, but preferably longer.
- Scope is limited to:
  - Full-time, salaried employees
  - Job Code Groups: Civil Service, Professional & Administrative and Faculty job groups. Labor Represented group potentially included in this group - waiting for OHR confirmation*
  - Pay Groups: P12, P10, and P09
  - Predicting departures from primary jobs only (not secondary jobs)*
  - Retirements: OHR has indicated they would like to predict retirements as well as resignations.*
- End goal: predictions at the employee level that will enable OHR/unit HR departments/other stakeholders to intervene in some way before the employee leaves.

*Needs further discussion with OHR

Modeling Approach

Proposed Development Steps

Minimum Viable Product

No.	Description	Priority
1	Produce a working model predicting resignations from U of M positions	MUST
2	Add granularity by breaking out forecast to an individual employee level	MUST
3	Identify features that are highly predictive (explainability)	SHOULD
4	Incorporate retirements into the main predictive model or secondary model	SHOULD
5	Add non-public data to the model: age and other demographic info	FUTURE*

Notes

*OHR will incorporate age and other private data after model delivery.

Logistics & Timelines

Next Steps:

- Next Week:
  - Finalize our population of interest
  - Discuss variable selection/consolidation
  - Continue pre-processing to generate (1) a clean flattened file and (2) a cleaned dataset formatted for survival analysis
- Following Week:
  - Begin model building

Progress To Date

To date we have:

- Met several times to discuss and establish our core problem statement
- Narrowed down our population of interest and drafted questions to further clarify scope
- Parsed the data and data dictionary to understand the available data
- Performed an initial descriptive analysis to establish churn rates over time and among various subpopulations
- Discussed the pros and cons of various modeling approaches (tree-based versus black-box); discussed the suitability of a survival analysis/ proportional hazard approach
- Begun transforming and encoding our data for further analysis/modelling

June 23rd, 2020 Team 1 - Client Meeting

1. Problem Statement - Clarification
   a. With the amount of additional variables involved in an employee's decision to retire, our proposed approach is to build a separate model to predict employees who retire (secondary model), if there is time remaining after we have completed the resignation prediction model (primary model). Are they any concerns with this approach?
   b. As a starting point for our model, we are only including "Resignations" to determine if an employee churned. However, there are several other forms of termination (i.e. "Termination for Cause"); should we be including these in our model selection/building process? One concern is that this might dilute the model, leading to lower performance.

1. What is the existing process for handling resignations? How is it conducted, and what events take place before, during, and following an employee's resignation? (related to Need for Change)

1. How do you envision/in what ways will your key stakeholders use our model? Will it be used differently by department managers versus the dean of a school, for example? (related to Solution Delivery)

June 23rd, 2020 Team 1 - Client Meeting

4. What will be the project team's measure of success? What metric(s) can we use to show that we have delivered value as a result of this project? How can the team demonstrate that the model is making a difference, even without the benefit of seeing the final implementation in place?
   a. Is **explainability** something that we as a team should be exploring further? How important is it to know the factors that lead to an employee churning? (This question applies both overall and for individual predictions)

5. We see that there is about an 80%/20% split between churned (resignations) and active employees. Therefore "accuracy" may not be an appropriate measure of success because there are many more employees who do not resign than employees who do.
   a. Is it more important to, with greater *precision*, correctly identify employees that will <u>truly</u> churn?
   b. Or is it more important to, with greater *recall*, correctly identify a greater <u>population</u> of employees who will churn? (explanation on next slide)

June 23rd, 2020 Team 1 - Client Meeting

Precision/Recall/Accuracy - Background Information

- During our initial discussions, one of the requirements for successful delivery is a highly “accurate” model. After further exploration, we see that only around 20% of employees resign (“churn”).

- Consider the accuracy metric:
  - Accuracy is simply the number of *correct predictions* over *total predictions.*
  - Our model could achieve 80% accuracy by simply predicting that that the employee will not resign, every single time the model makes a prediction. Since we see in the data that only 20% of employees resign, this prediction will be accurate 80% of the time.
  - However, this model is completely useless! In this case, accuracy is not the best metric to strive for.

- Two other well known / widely used metrics are Precision & Recall.
  - Precision: how many employees selected from our model will actually churn?
  - Recall: how many employees who could potentially resign did our model actually identify?

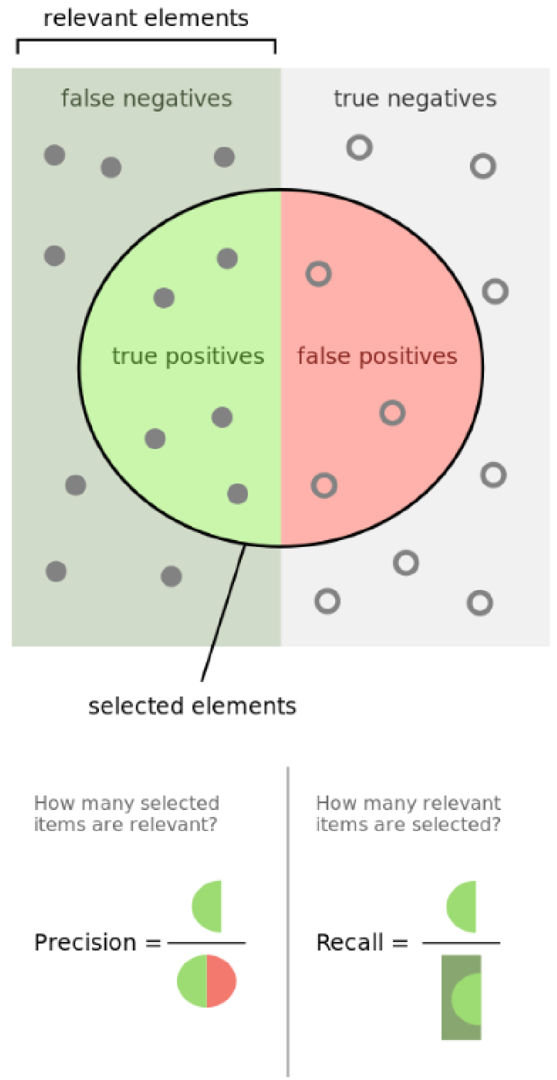
June 23rd, 2020 Team 1 - Client Meeting

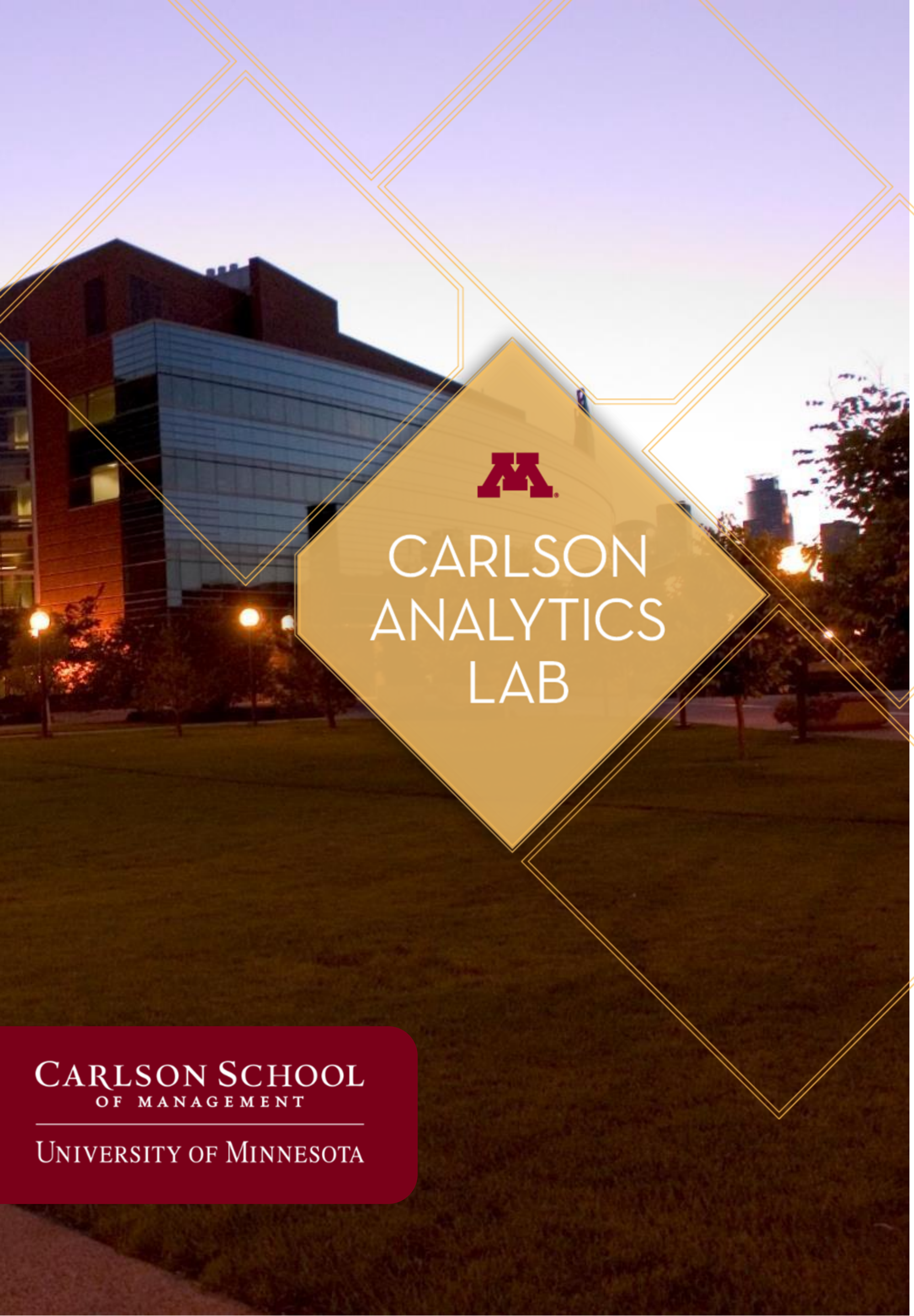
Precision/Recall/Accuracy - Background Information

- Example of *high precision:*
  - The model predicts that 10 employees will resign. Amazingly, every single employee that the model predicted to resign, does resign or was planning to resign! However, there were 100 employees who also resigned that the model did not flag, and therefore there was no opportunity to intervene and potentially keep these employees.
  - High precision means that a prediction of an employee resigning is highly trustworthy (they are very likely to resign), but the model will miss out on many other employees that will resign.

- Example of *high recall:*
  - The model predicts that 500 employees will resign, and staff work very hard and intervene with all 500. Through the course of the conversations, the staff find out that only 100 of these employees were actually planning to resign – the other 400 employees had no intentions of leaving, leading to a wasted effort. However, they were able to have a conversation with all 100 employees who were planning on leaving and prevent some of them from quitting.
  - High recall means that the model will flag all or most people who are likely to resign, but the model will also incorrectly flag many people who have no plans to leave.

- How to decide which metric is more desirable? *Think about the costs/drawbacks of intervening when someone is not planning to leave, versus not intervening if someone is planning to leave.*




CARLSON ANALYTICS LAB
CARLSON SCHOOL OF MANAGEMENT
UNIVERSITY OF MINNESOTA

Thank You