University of Minnesota Office of Human Resources

Employee Resignation Predictive Modeling

August 11, 2020

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

Our team

Kevin Grady

Loyalty Analytics Consultant

20+ years leading marketing and analytics teams, transforming customer database insights into results.

His passions include consumer psychology, loyalty marketing, strategic planning, and data visualization.

Wendy Lu

Reporting Analyst
US Bank

Wendy has been in the banking industry for 5 years and is working on ETL implementation, reporting solutions and dashboards for senior leaders.

Her focus is mainly on strategic planning, marketing analysis and insights, and data visualization.

Anthony Meyers

Managing Director
Aon

10 years experience leading analytics teams to support clients understanding of natural catastrophe risk.

Also focuses on strategy and execution of modern technology and analytics to provide clients and prospects best in class analytics solutions.

Danny Moncada

Data Analyst
Workforce Data Mgmt.

6+ years experience as a report developer helping business leaders make informed data driven decisions.

Works in the Office of Human Resources as primary front end developer for HR Analytics and site administrator for the *HR Executive Dashboard*

Claire Ryan

Senior Analyst
DaVita Kidney Care

5 years experience in healthcare analytics, with expertise in data visualization and predictive modeling.

Currently a Senior Analyst at DaVita, supporting medical directors in designing data-driven patient care programs.

Jonathan Watkins

Data & Planning Analyst
Hennepin County

7+ years of research experience in academic and governmental settings.

Jonathan holds a Bachelor of Science in Physics from Howard University and a Master of Public Policy from the University of Minnesota.

Disclaimer

Any explicit example highlighting an employee or employees is purely illustrative and does not represent any factual information.

Functional architecture, methods utilized, and model performance are factual.

Correlation is a necessary, but insufficient condition for causation.

Predicting resignations to avoid losing talented employees

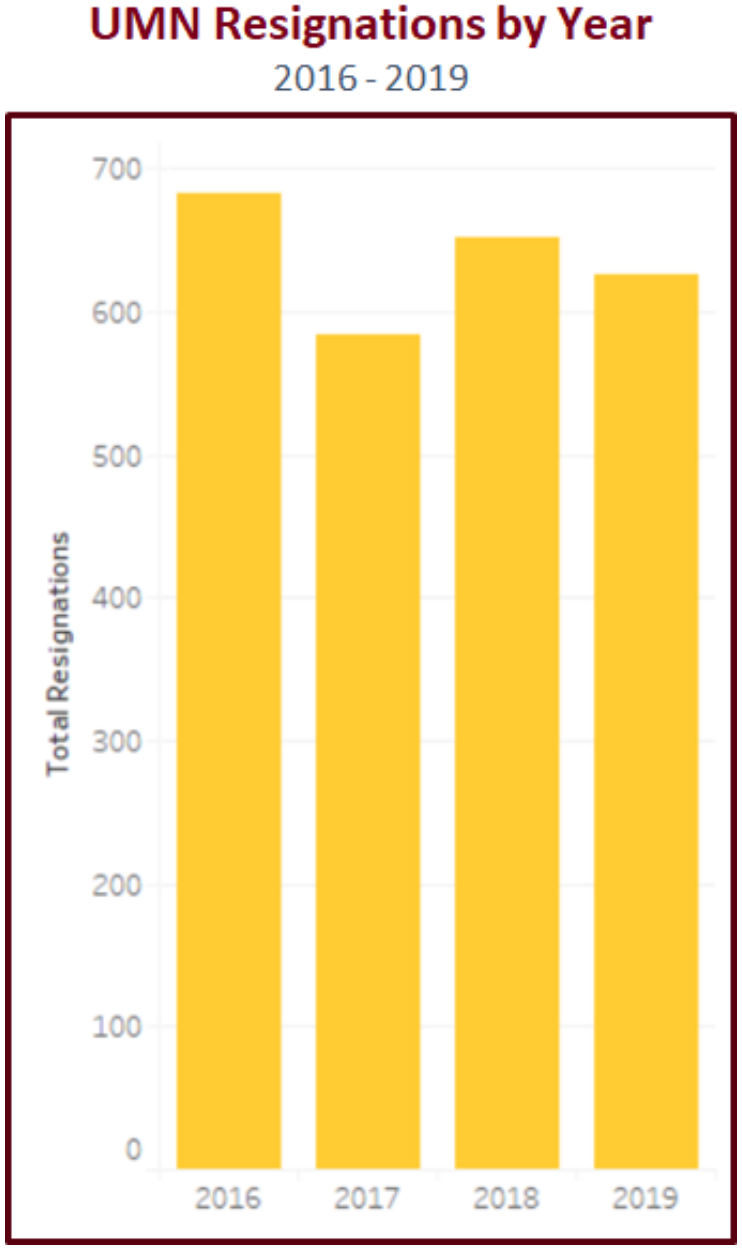
To aid the Office of Human Resources in reducing turnover, our team has been tasked with building a model to identify employees most likely to resign.

A recent Deloitte study showed that the cost of losing one employee ranged from 1 -2 times their annual salary.

Over 2,500 University staff resigned from 2016 - 2019

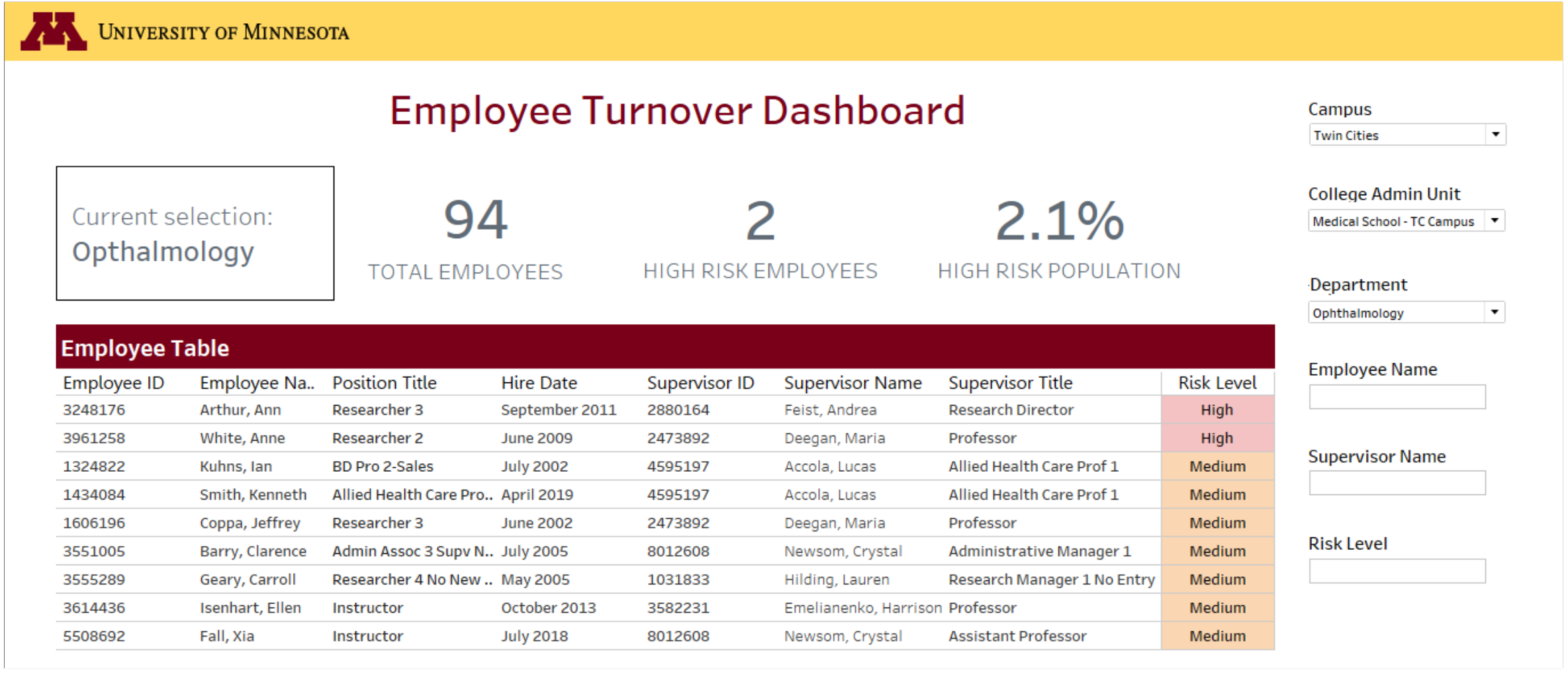
- The goal: use predictive modeling to **identify full -time faculty and professional staff who are likely to leave UMN**

- Final five year population consisted of 22,514 faculty and professional staff, of which 3,261 had resigned (15%)



Our model can be viewed through a flexible, simplified dashboard

This enables HR leaders to have an impact on resignations by predicting who is most likely to resign



Employee Table

Employee ID	Employee Na..	Position Title	Hire Date	Supervisor ID	Supervisor Name	Supervisor Title	Risk Level
3248176	Arthur, Ann	Researcher 3	September 2011	2880164	Feist, Andrea	Research Director	High
3961258	White, Anne	Researcher 2	June 2009	2473892	Deegan, Maria	Professor	High
1324822	Kuhns, Ian	BD Pro 2-Sales	July 2002	4595197	Accola, Lucas	Allied Health Care Prof 1	Medium
1434084	Smith, Kenneth	Allied Health Care Pro..	April 2019	4595197	Accola, Lucas	Allied Health Care Prof 1	Medium
1606196	Coppa, Jeffrey	Researcher 3	June 2002	2473892	Deegan, Maria	Professor	Medium
3551005	Barry, Clarence	Admin Assoc 3 Supv N..	July 2005	8012608	Newsom, Crystal	Administrative Manager 1	Medium
3555289	Geary, Carroll	Researcher 4 No New ..	May 2005	1031833	Hilding, Lauren	Research Manager 1 No Entry	Medium
3614436	Isenhart, Ellen	Instructor	October 2013	3582231	Emelianenko, Harrison	Professor	Medium
5508692	Fall, Xia	Instructor	July 2018	8012608	Newsom, Crystal	Assistant Professor	Medium

We deliver a holistic analytics framework that is successful in meeting the project goals

Architecture

We deliver a **cost-effective,** open-source machine learning infrastructure built on frameworks already in use at the University.

Data Science

We deliver an analytics framework utilizing standard practices for **Data engineering** and model building.

Performance

We deliver models with **high** performance

88%+ **Precision** identifying faculty or staff that might resign

84% Time-to-Event **order accuracy**

500% Potential **increase in effectiveness**

Our model will help you (**the experts**) make more informed decisions

Predictive Analytics

Based on historical trends and patterns, **what is likely to happen in the future at the University?**

Prescriptive Analytics

What should the University do?

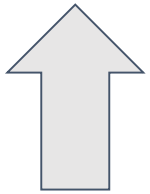
Our model will help you (**the experts**) make more informed decisions

Predictive Analytics

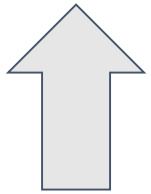
Based on historical trends and patterns, **what is likely to happen in the future at the University?**

Prescriptive Analytics

What should the University do?



Our team's predictive model



Your expertise

We think there are **4 primary reasons** related to staff and faculty resignations

HOW LONG a staff or faculty member has worked somewhere

WHO a staff or faculty member works with

HOW MUCH a staff or faculty member gets paid

HOW RECENT was their last **SALARY CHANGE**

4 primary reasons related to staff and faculty resignations - key factors used in our models

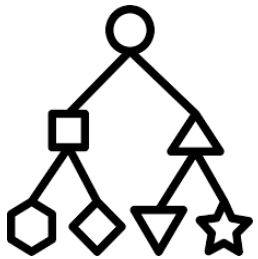
- How long have they worked at UMN?
- How long at their job code?
- How long in their current job position?
- How long in their current department?

- How many different supervisors have they had?
- How many direct reports does their supervisor have?
- How long since last their supervisor change?
- How big is their department?
- What's their department turnover rate?

- What is their salary?
- How many raises have they received?
- Is their salary above or below the position median salary?

- When did they last have a salary increase?
- Whether they hold multiple jobs
- Whether in a tenure track position (faculty)

Our multi-model approach enables department leaders to make more informed decisions.



Classification Model

Describes **WHO and WHAT FACTORS**

- Professional Staff Model
- Faculty Model

Time-to-Event Model

Describes **WHEN**

- Professional Staff Model

Results from the models highlight **who** might resign, **what factors** may predict resignations, and **when** resignations might occur.

Enables department leaders to leverage their **expertise** to to react appropriately.

Multiple classification models driven by behavior differences between professional staff and faculty.

Our framework leverages **free**, open source software

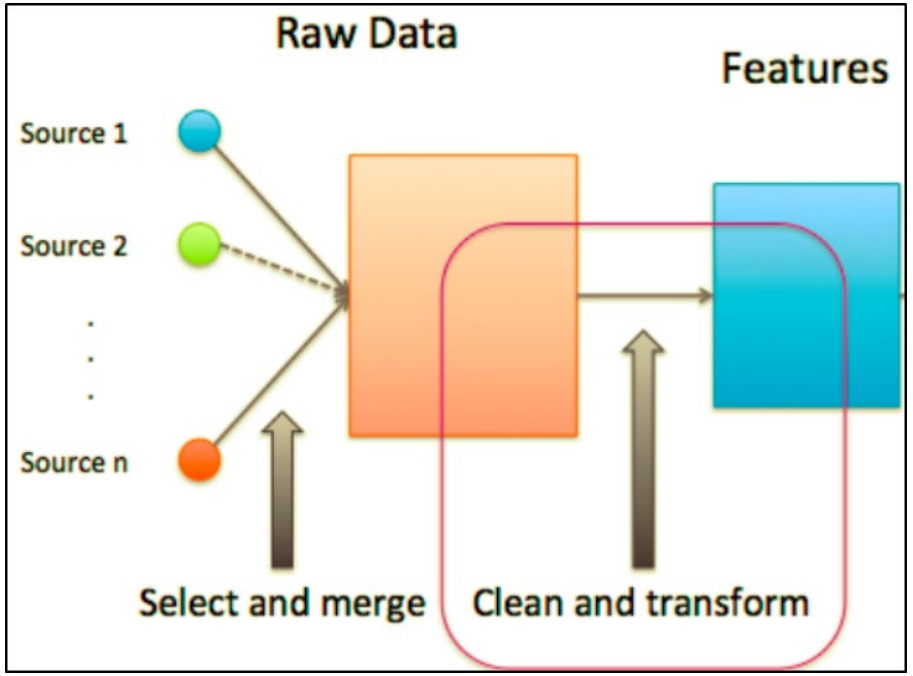
- Carlson Analytics lab provided Virtual Machine with Anaconda environment, a free and open source distribution of Python programming language for scientific computing
- All models were developed using **existing** open-source machine learning libraries maintained by data science community

- Additional software utilized:
  - **R** (data exploration)
  - **Jupyter** Notebooks (code development)
  - **GitHub** (repository for storing code)
  - **Tableau** Desktop (data visualization)

```
Here are the environment details...

C:\Python\envs\ohr_2020\python.exe
3.7.7 (default, May  6 2020, 11:45:54) [MSC v.1916 64 bit (AMD64)]
sys.version_info(major=3, minor=7, micro=7, releaselevel='final', serial=0)

This notebook is using Pandas version: 0.25.3.
This notebook is using Numpy version: 1.18.5.
This notebook is using Matplotlib version: 3.2.2.

Here are the machine learning libraries:
This notebook is using Scikit-learn version: 0.23.1.
The following algorithms are a part of the sklearn package: Logistic Regression, k-Nearest Neighbor, Decision Tree, Random Forest, Support Vector Machine.
This notebook is using XGBoost version: 0.90.
This notebook is using LightGBM version: 2.3.1.
This notebook is using Scikit-plot version: 0.3.7.
This notebook is using Lifelines version: 0.24.16.
This notebook is using PySurvial version: 0.2.1.
```

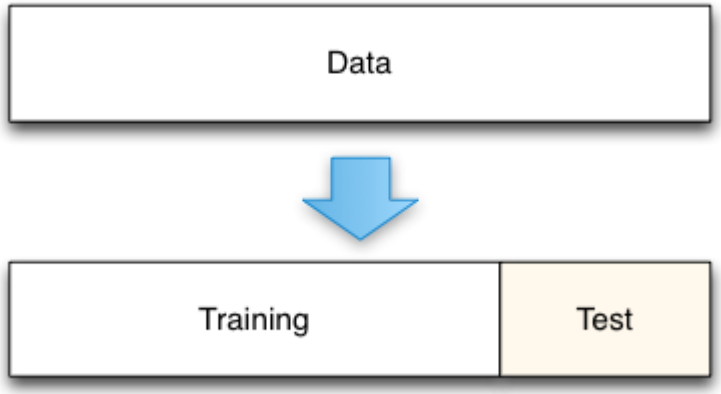
Our framework provides **value** through our data engineering process

- Our framework is beneficial because we didn't use raw, unstructured data to plug directly into the models
- We created new, more **informative** data points from the raw data in a process known as *feature engineering*
- We selected data points that were most relevant to our four primary reasons and generated **clean**, easy to interpret factors



Our framework provides the ability to evaluate our model outcomes

- We calibrated and "fit" *both* our classification and time-to-event models on a **training** dataset, a subset of employees that are used to "learn" latent patterns between our factors and whether an employees is active or resigned
- We set aside a **testing** set to evaluate how well the model performs on **unobserved** data points and to provide an estimate of how well the model will do on future data



Professional Staff

- Training examples: ~10,500 employees
- Testing examples: ~5,200 employees

Faculty

- Training examples: ~3,200 employees
- Testing examples: ~1,600 employees

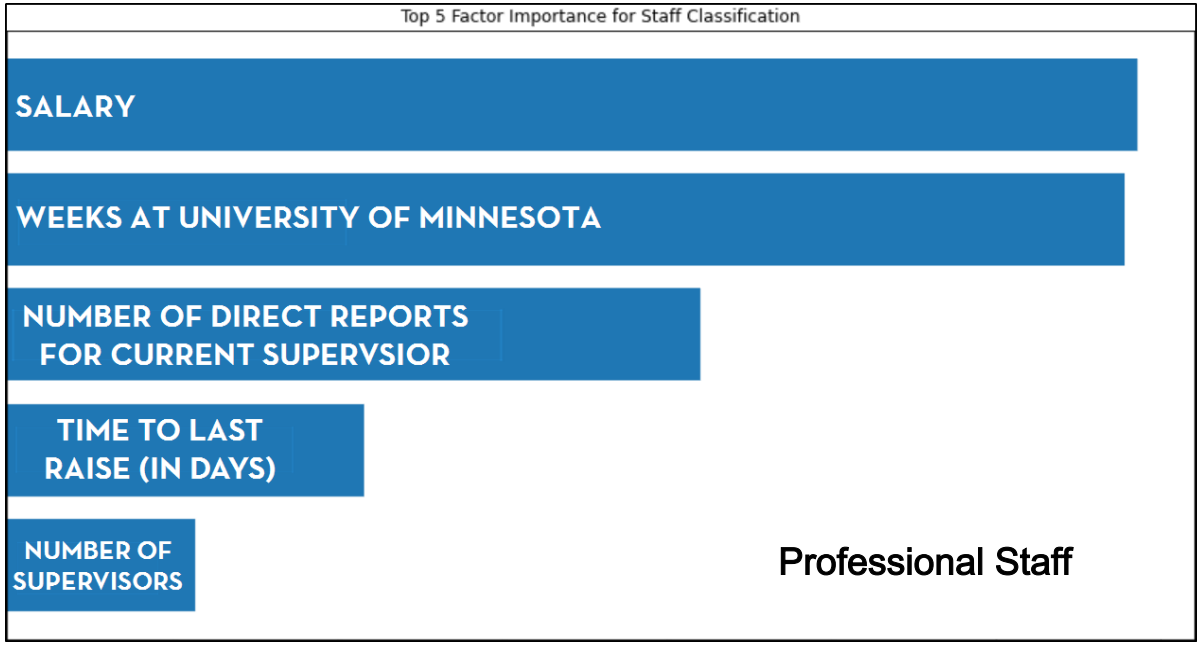
Reviewing classification model performance

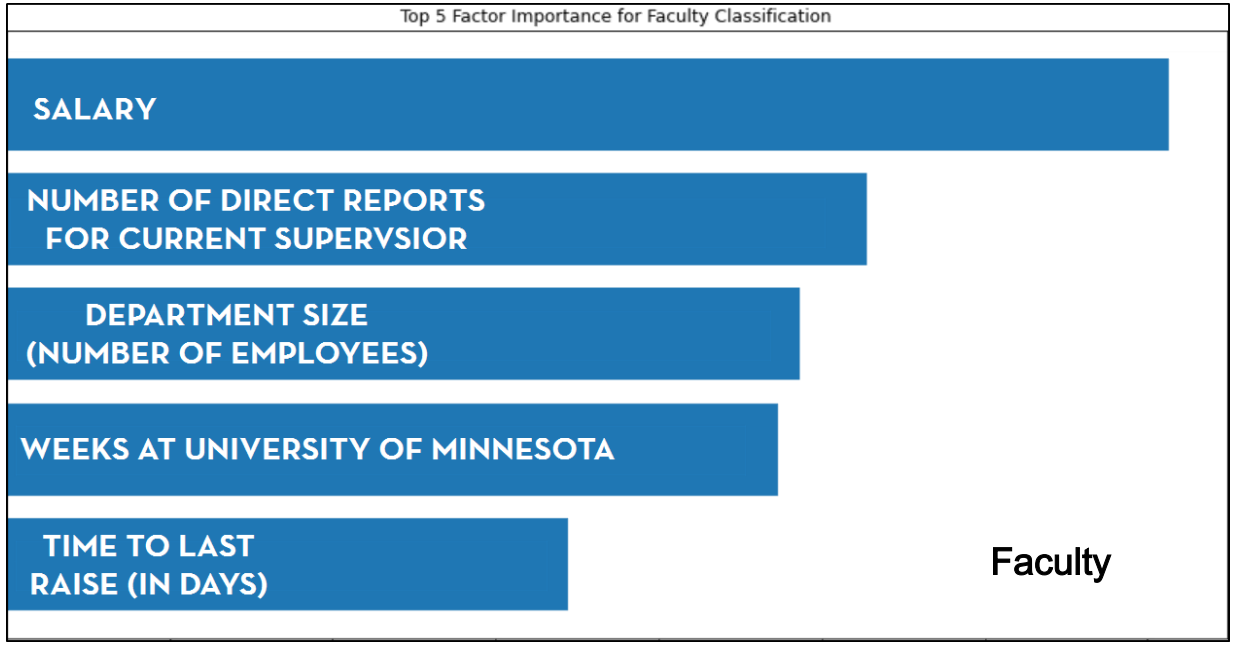
STAFF	Our model predicted **867** professional staff would resign (out of 5,200 active employees in test data)	**766** predicted correctly by our model	90 % precision
FACULTY	Our model predicted **184** faculty would resign (out of 1,600 active employees in test data)	**162** predicted correctly by our model	88 % precision

- These precision scores come from using our fitted models to predict the outcomes for employees in our "**testing** "dataset
- Our two final classification models were both constructed using the **LightGBM** algorithm

We can use factor importance to gain insights into model predictions

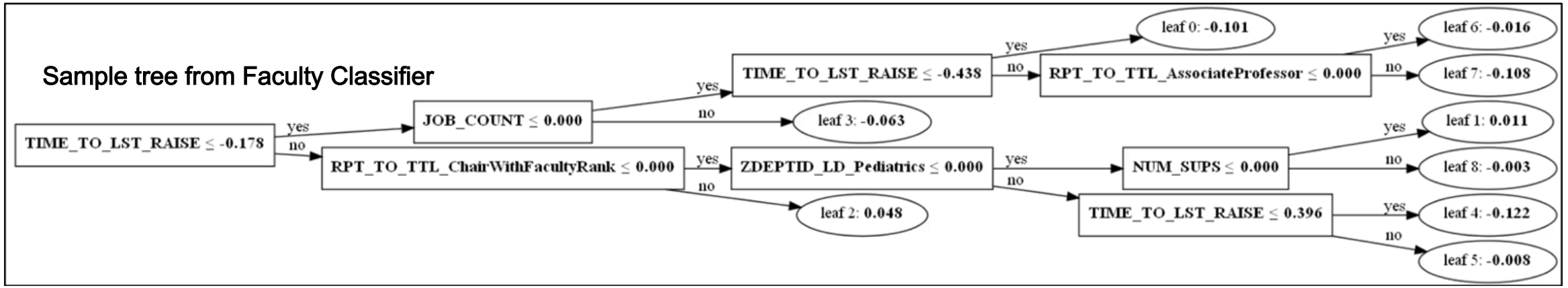
- Classification models can help us uncover latent or undiscovered factors for identifying employees who have resigned
- Factor importance provides insight into the **data & model predictions**
- **Note:** there may be additional factors influencing employee resignations that are outside the scope of the data set used and our model outcomes





Unpacking the model decision -making process

- **LightGBM** is a gradient boosting framework that uses decision trees to predict the value of a target variable (i.e. *active* vs. *resigned* employee) based on several input variables (i.e. our **factors**)
- LightGBM will "fit" trees and determine which employees it did not classify with great enough **precision** , giving those employees higher weights during the next training cycle; it will iterate through this process ***thousands*** of times to get the best possible separation of active vs. resigned employees
- Advantages:
  - Simple to understand and interpret
  - Performs well with large datasets
  - **Mirrors human decision making more closely than other approaches**



Examining Time-to-Event Metrics

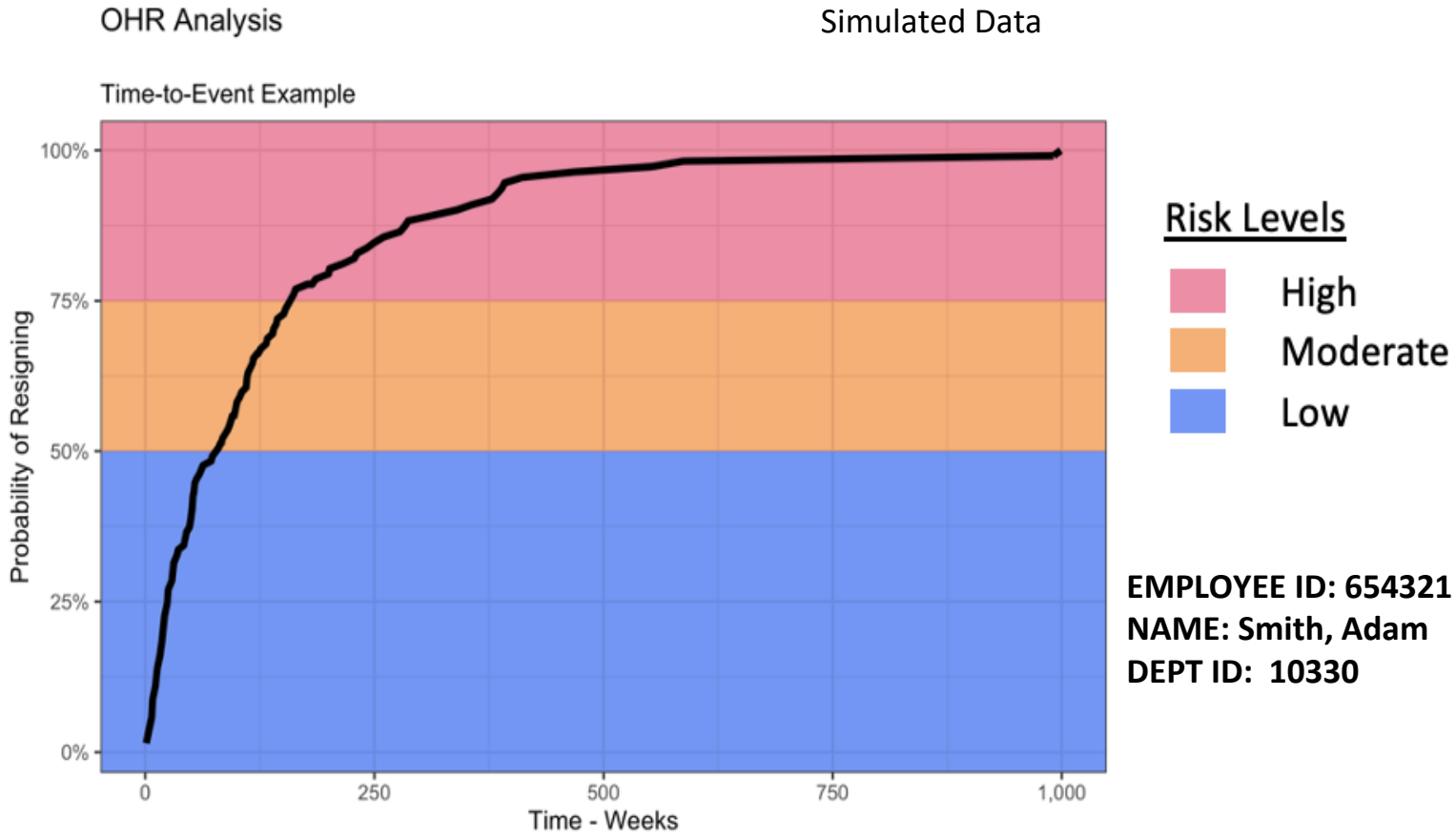
In the context of the problem, the Time-to-event model is showing **high predictiveness** around the *order* of who might quit, given the volatility around *precisely* when someone might quit

Concordance Score: **84%**

- Concordance gives us a sense of model performance
- High-quality models typically achieve 55 - 75% concordance
- Values range between 0 and 100
  - 100 is perfect
  - 50 is no better than random chance
  - 0 is perfectly wrong
- Concordance is concerned with the order of our predictions, rather than exact values

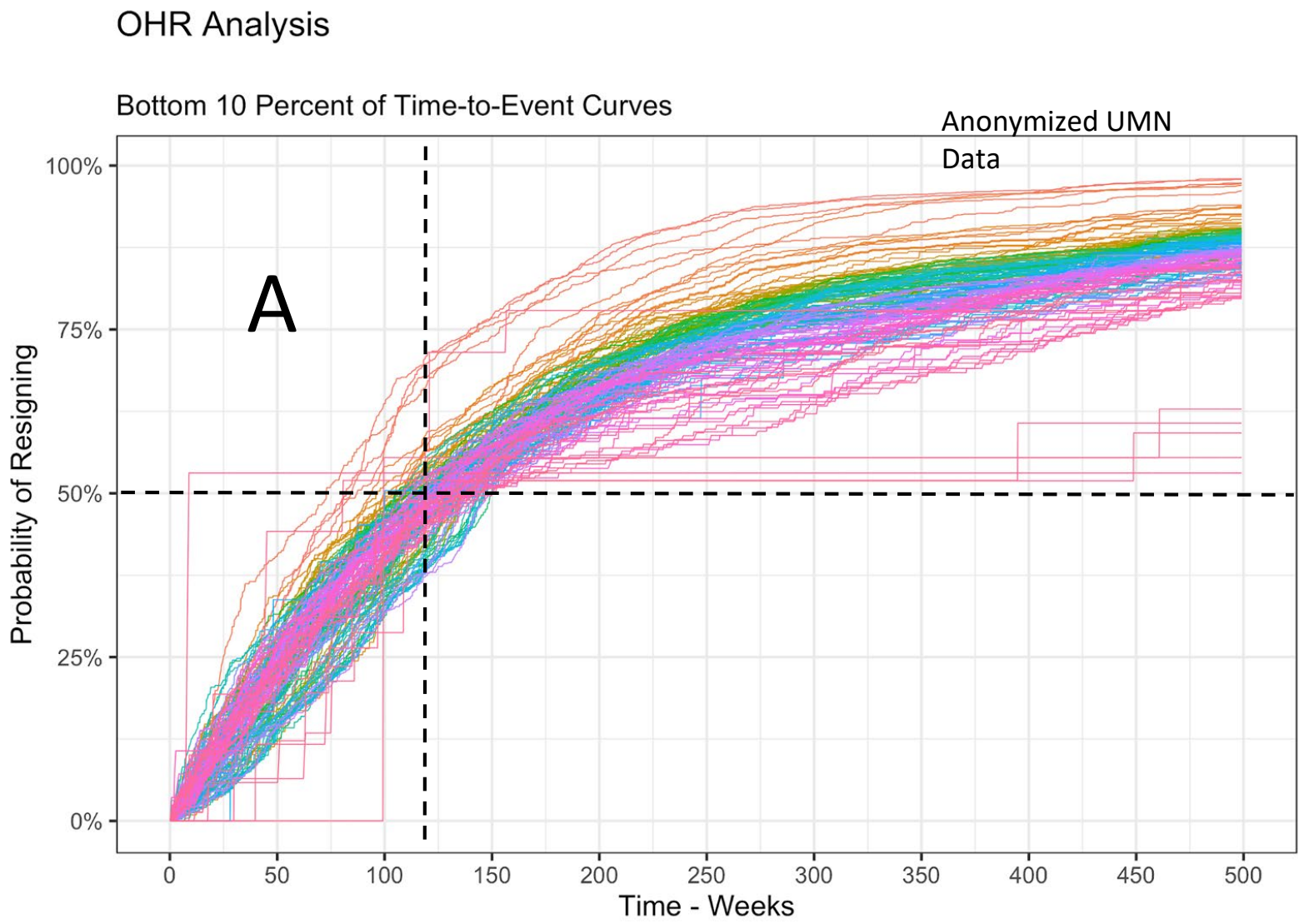
Understanding Time -to-Event Output

- The model allows us to predict the time-to-resignation for every employee
- The trajectory of each curve can help us identify intervention opportunities
- As we move forward in time and acquire more data we can continue to update and refine our predictions



Examining OHR Time-to-Event Data

- Algorithm can predict the probability of resignation over time for staff and faculty
- Higher risk individuals appear above and to the left of lower risk individuals
- In this example, we might consider everyone appearing in quadrant A high risk



Model output: what we see

The raw model output looks like this:

EMPLID	EVENT_RESIGN	WEEKS_UMN	JOB_COUNT	LOCATION	NUM_RAISES	TIME_TO_LST_RAISE	EMP_CLSS_CD	JOBCD_GRP_CD	WKFC_ACTN_RSN_LD	DEPTID	JOB_TTL	CLLG_ADM_UNT_LD
724371	0	718.1429	1	TCEASTBANK	3	155.1	ACA	AA	Retirement	10237	U of MN Foundation VP	UNIVERSITY RELATION
3583275	0	280.1429	1	TCEASTBANK	0	280.1	ACP	AP	End of Appointment	10197	Assistant Coach	INTERCOLLEGIATE ATH
5150754	0	269.1429	1	CHANHASSEN	0	42.1	CVL	CS	Rehire after 30 Days	11050	CO Pro 1-Bookstore Svcs	FOOD, AGRI/NAT RSRC
4868112	0	195.0000	1	TCEASTBANK	0	41.1	CVL	CS	Position Data Update	10194	Athl Pro 1-Ticket Sls/Svc	INTERCOLLEGIATE ATH
4978498	0	209.7143	1	TCEASTBANK	0	41.1	CVL	CS	Position Data Update	10138	Rec Pro 2-Fit/Wellness	STUDENT AFFAIRS, VIC

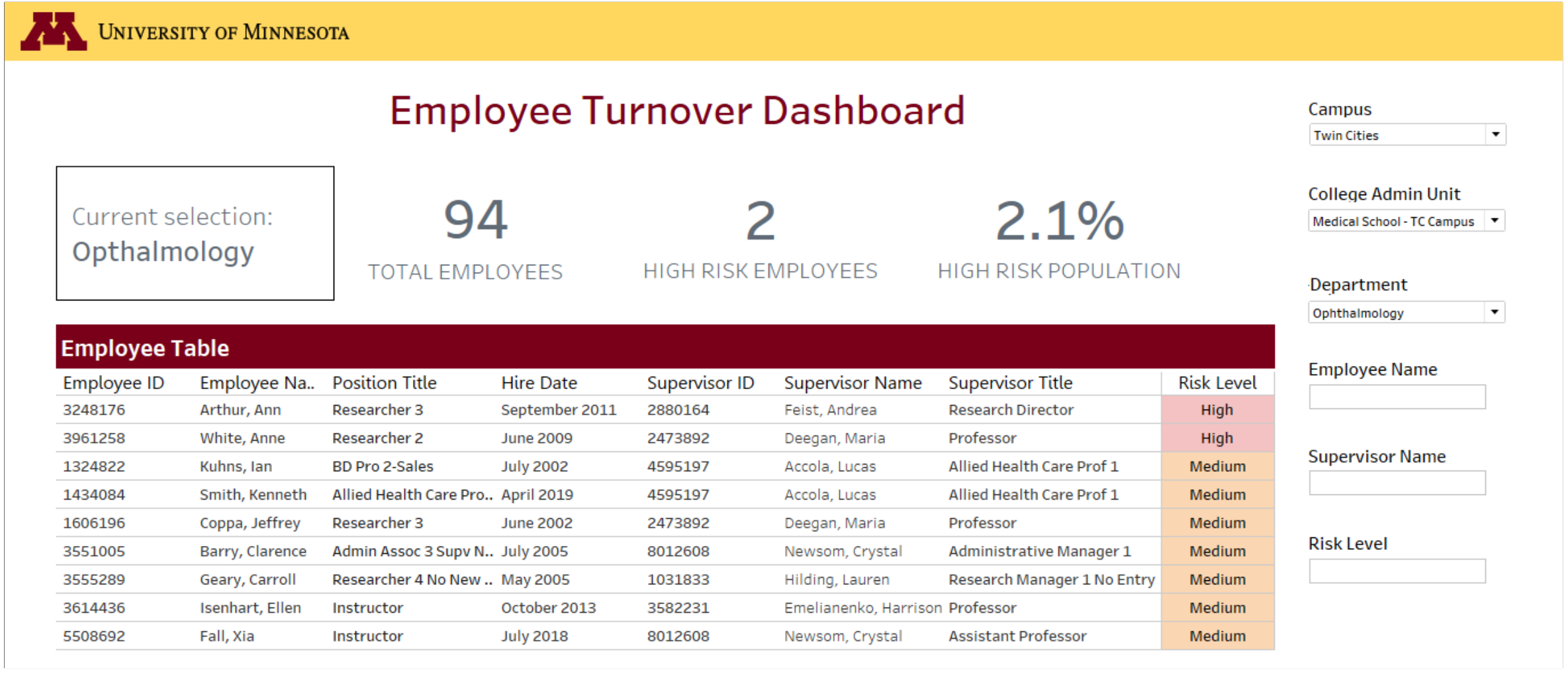
ATGY_DESC_Other	WKFC_CATGY_DESC_Police.Security	WKFC_CATGY_DESC_Service.and.Maintenance	WKFC_CATGY_DESC_Skilled.Generalists	WKFC_CATGY_DESC_Skilled.Trades	WKFC_CATGY_DESC_Student.Services	ZDEPT_MEDIAN_BOOL_	X0.5	partial_hazards
	0	0	0	0	0	1	8.857143	0.4535576
	0	0	0	0	0	0	73.000000	8.6076045
	0	0	1	0	0	0	80.714286	3.7579784
	0	0	0	0	0	0	82.142857	9.6389566
	0	0	0	0	0	0	84.142857	9.6389566

The last step in our process is to turn it into something valuable and user - friendly for the end users (**you** !)

HR Analytics will also have access to the raw model output to generate additional useful content.

What you see: flexible, easy to use dashboard

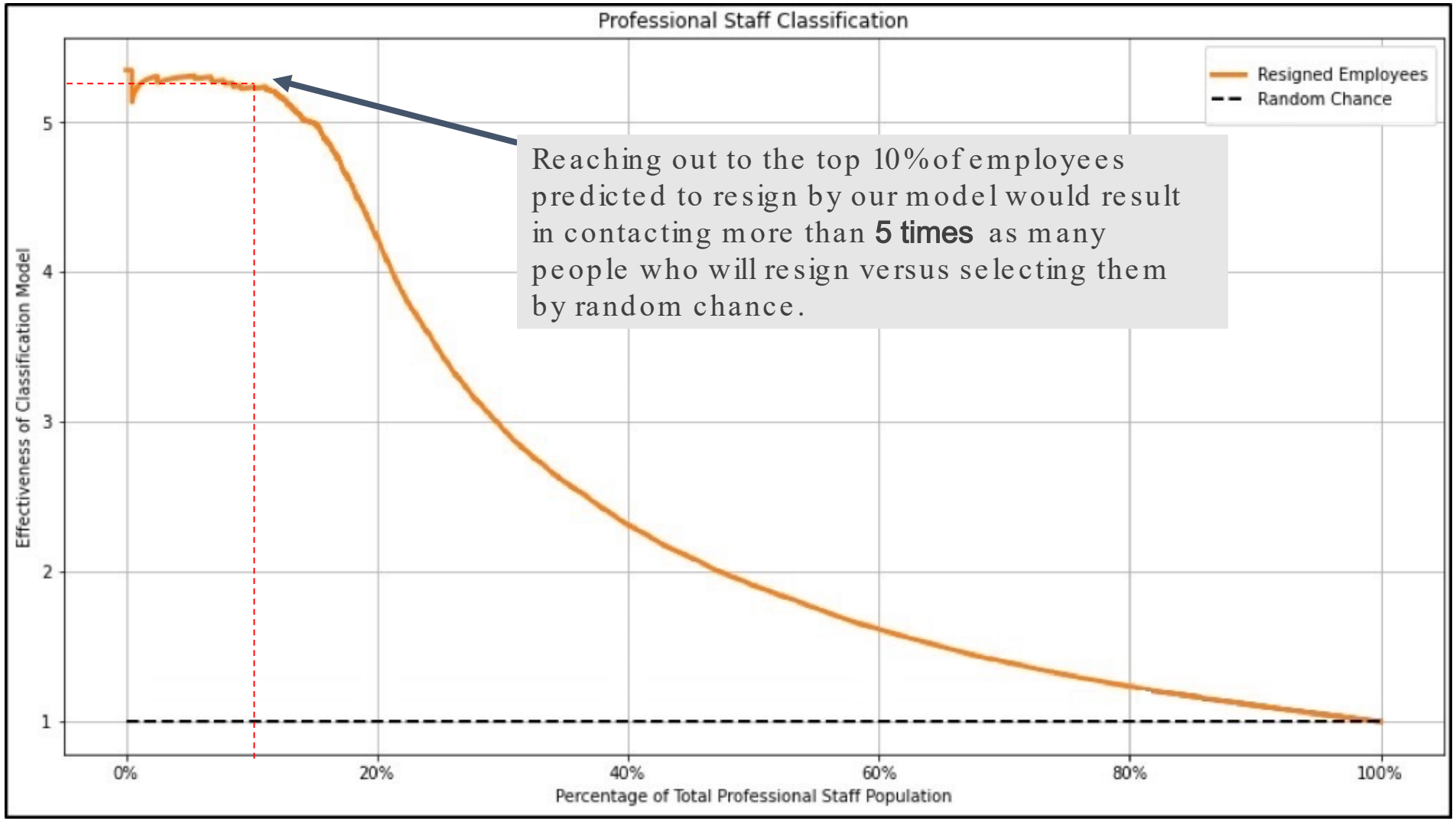
This enables HR leaders to have an impact on resignations by predicting who is most likely to resign



Employee Table

Employee ID	Employee Na..	Position Title	Hire Date	Supervisor ID	Supervisor Name	Supervisor Title	Risk Level
3248176	Arthur, Ann	Researcher 3	September 2011	2880164	Feist, Andrea	Research Director	High
3961258	White, Anne	Researcher 2	June 2009	2473892	Deegan, Maria	Professor	High
1324822	Kuhns, Ian	BD Pro 2-Sales	July 2002	4595197	Accola, Lucas	Allied Health Care Prof 1	Medium
1434084	Smith, Kenneth	Allied Health Care Pro..	April 2019	4595197	Accola, Lucas	Allied Health Care Prof 1	Medium
1606196	Coppa, Jeffrey	Researcher 3	June 2002	2473892	Deegan, Maria	Professor	Medium
3551005	Barry, Clarence	Admin Assoc 3 Supv N..	July 2005	8012608	Newsom, Crystal	Administrative Manager 1	Medium
3555289	Geary, Carroll	Researcher 4 No New ..	May 2005	1031833	Hilding, Lauren	Research Manager 1 No Entry	Medium
3614436	Isenhart, Ellen	Instructor	October 2013	3582231	Emelianenko, Harrison	Professor	Medium
5508692	Fall, Xia	Instructor	July 2018	8012608	Newsom, Crystal	Assistant Professor	Medium

Making you more effective at picking the right people



We deliver a holistic analytics framework that is successful in meeting the project goals

Architecture

We deliver a **cost-effective,** open-source machine learning infrastructure built on frameworks already in use at the University.

Data Science

We deliver an analytics framework utilizing standard practices for **Data Engineering** and model building.

Performance

We deliver models with **high** performance

88%+ **Precision** identifying faculty or staff that might resign

84% Time-to-Event **order accuracy**

500% Potential **increase in effectiveness**

Limitations

There are many additional reasons that may drive employees to quit, that we can't observe and therefore our model can't capture. (For example, a spouse accepting a job out of state)

Predictive models rely solely and completely on historical record. Future shifts in behavior will require model retraining.

Any model is a best guess and should be used as an additional data point rather than the final word.

Future Enhancements

Due to privacy constraints, many key data points were not available to our team to use in the models. We believe that adding these to the models will boost performance.

- Age & other demographic information
- Education level and field of study
- Employee zip code (daily commute distance)
- Previous employment history

More sophisticated data mining and machine learning techniques can be applied to the data set given increased computational resources.

CARLSON SCHOOL
OF MANAGEMENT
UNIVERSITY OF MINNESOTA

CARLSON
ANALYTICS
LAB

Thank You!