

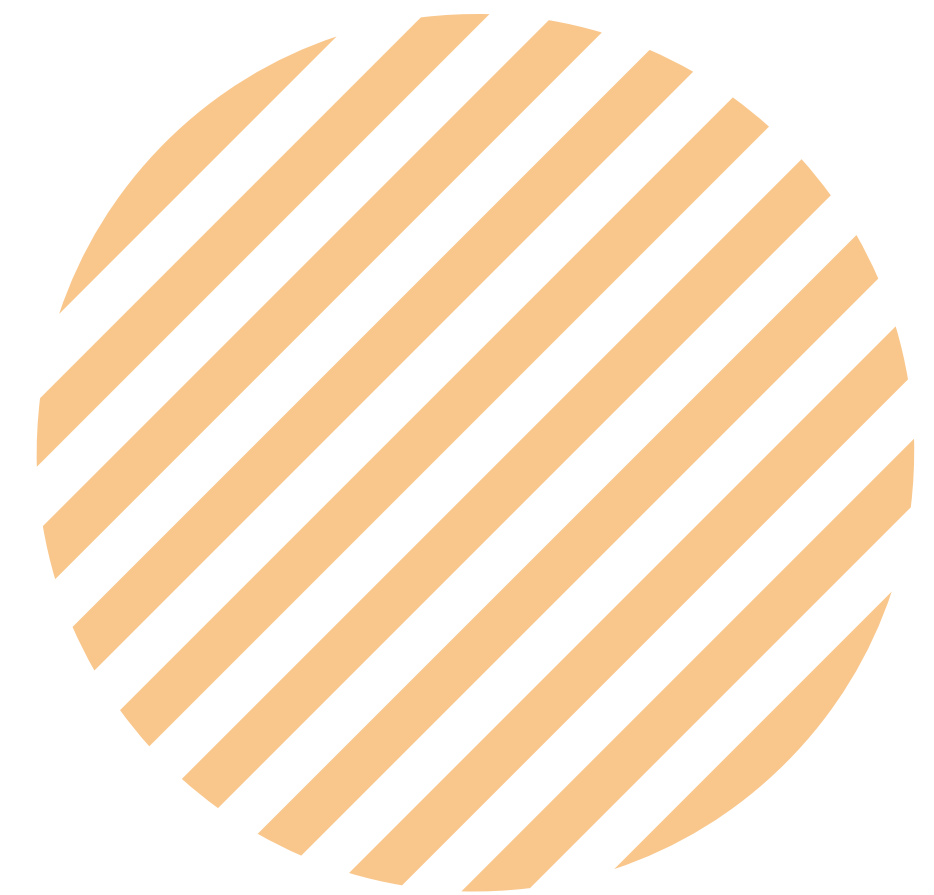


Human-AI Interaction | 2020 Fall

Exploring Employee Attrition for Talent Retention

Final Presentation

Janet Dai, Yi Wang, Dianne Kim

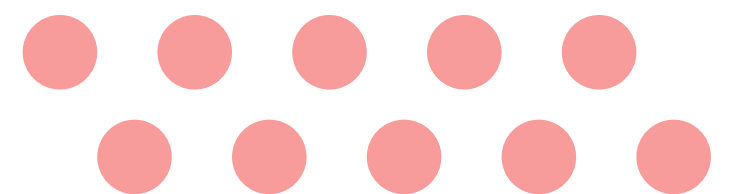




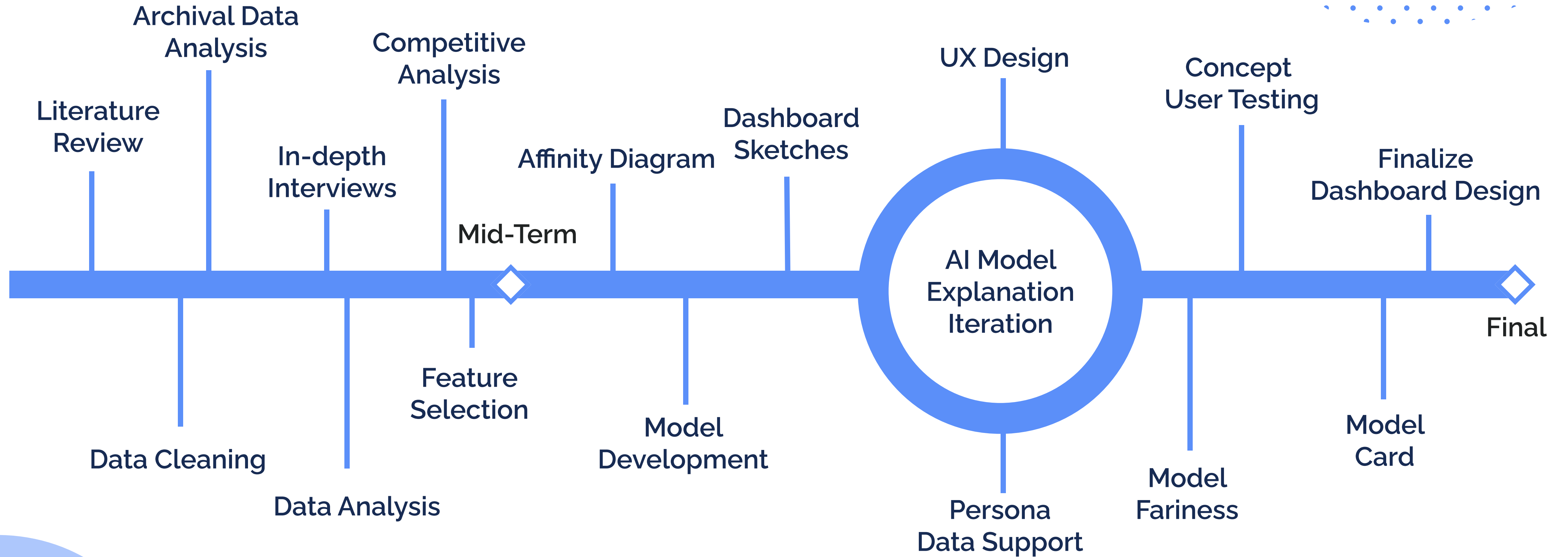
Abstract/Overview

Attrition is a widespread and expensive problem for companies
Costs of not only recruiting, interviewing, onboarding, training but also loss of productivity and output when an employee leaves unexpectedly
Cost of replacement = 20% salary of the lost employee

How might we create a solution that could help companies predict when employees are likely to leave...with the goal of providing a strategy to retain them?



Project Timeline



Literature Review

AI applications have been developed to try to tackle workers who are planning to leave their jobs, with companies like IBM. Armed with this knowledge, managers can take strategic actions to encourage that employee to stay. By focusing specifically on high performers with in-demand skills, IBM says it has saved \$300 million in retention costs.

In-depth Interview

8 Participants (3 managers, 1 HR , 4 Employees)

- The choice to leave a job can be very complex and subjective, but the relationship with managers and co-workers is the most common one based on the interviews
- Trust and explanation is super important for this kind of tool
- Attrition prediction tools are ideal for large companies where HR are not familiar to all the employees

The image is a composite of three elements. At the top right, a video call window shows two participants: Eric Nordquist (top) and Dianne Kim (bottom). The main central area displays a questionnaire with the following text:

managers/ HR Professionals

agreed to participate in this study! As a quick reminder, we are working on a project for our Human-AI team that is specifically look at the topic of worker attrition, and the companies are using to manage their workforce and we wanted to get your thoughts on a potential solution we're considering that could help to address this problem area.

Do you have any questions for us before we begin?

Do we have your permission to record this interview, just for note-taking purposes?

Section 1: Demographic

1. What is your age? Gender?
2. What is the industry in which you work? (Finance, Education, Healthcare, etc.) Selfie
3. What is your highest level of education? (High school, associate's degree, bachelor's degree, masters, PhD) - Master's
4. How many years of work experience do you have, in total?

Section 2: Job Role and Organization

5. What is your job title (as it is written on your business card or email signature?)
6. What is the size of the organization (number of employees)? Your best estimate is fine.
7. How long have you been in your current role?

At the bottom, a large whiteboard is covered with numerous colorful sticky notes (blue, pink, yellow, green) arranged in a complex mind map structure, representing data collected from the interviews.

Exploratory Data Analysis

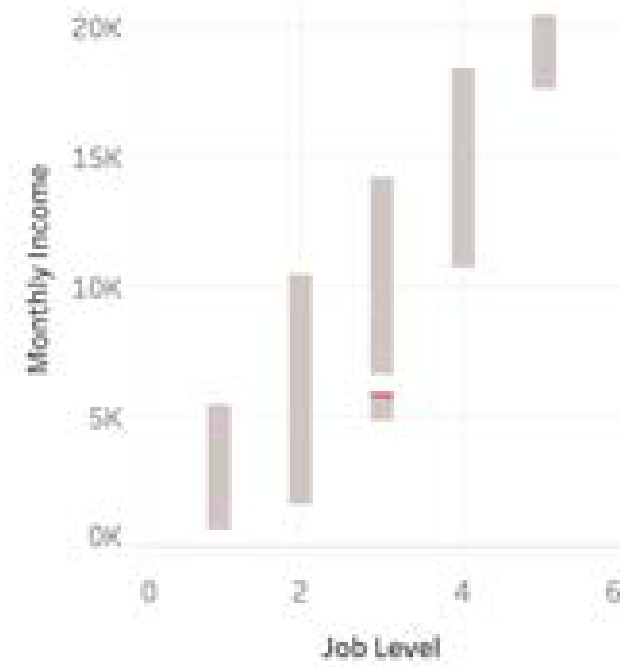
IBM Kaggle dataset: 35 features, included 1427 employees

IBM Employee Attrition Analysis

16.1% Attrition Rate



Lower job level (with lower income), higher attrition



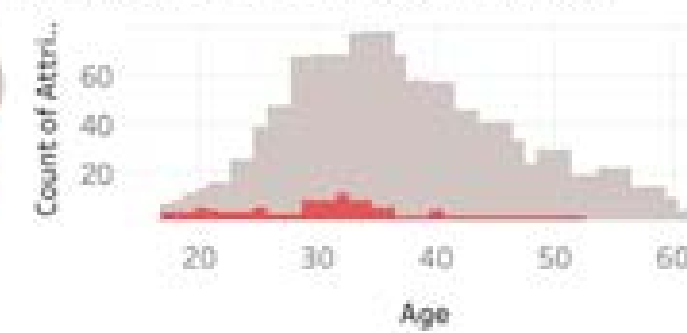
Sales Representative and Lab Researchers have higher attrition



More frequently business travel, more attrition



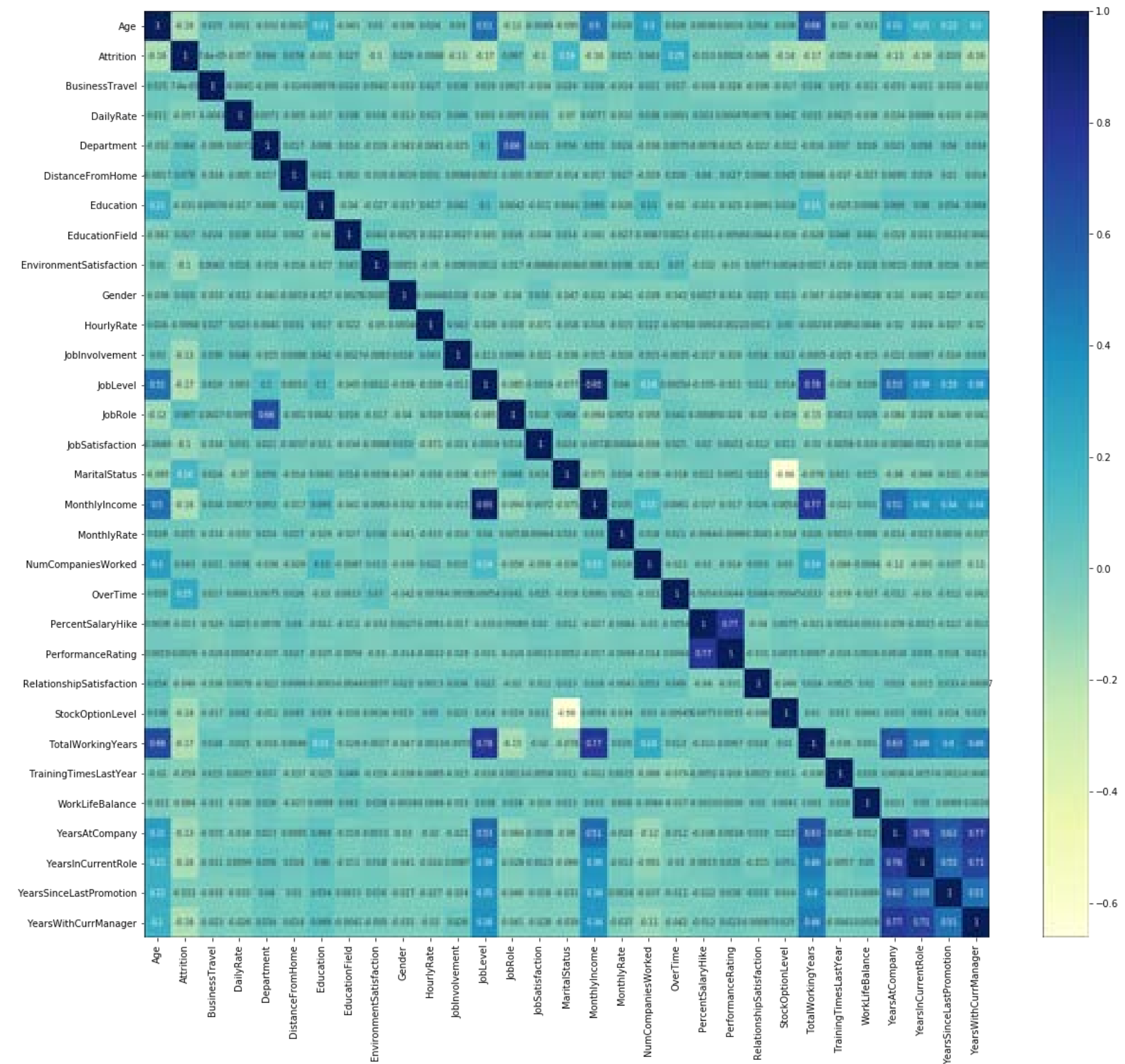
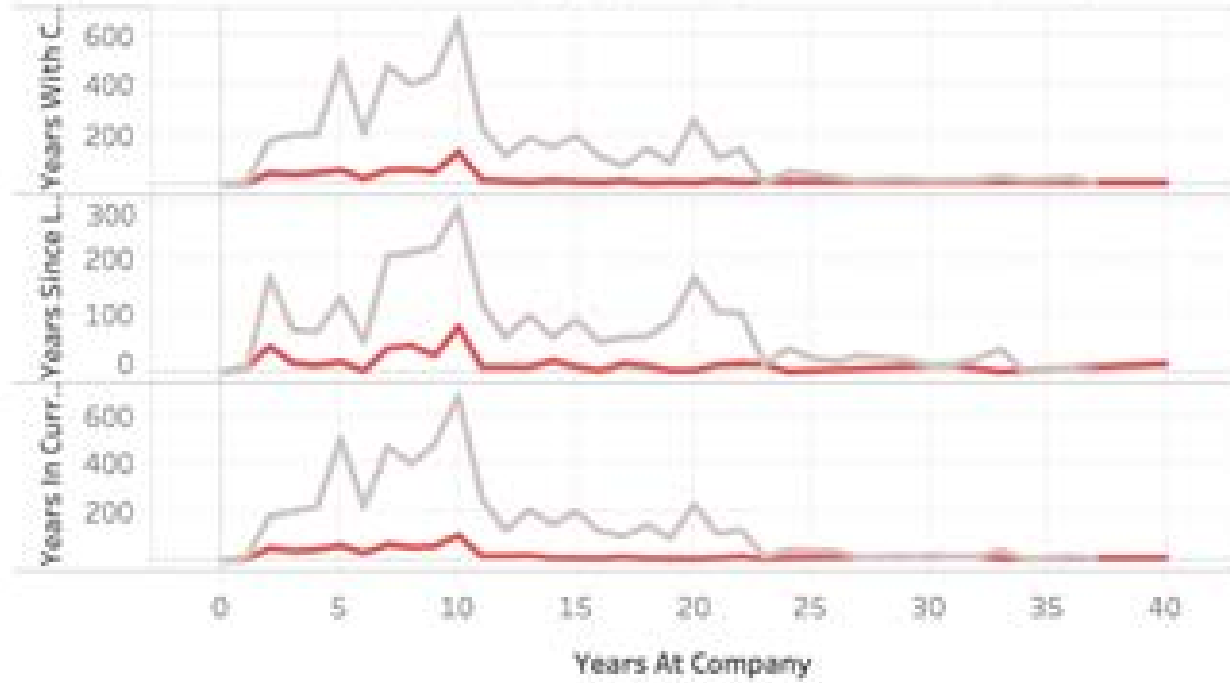
Younger than 40 has higher attrition



A small number of working years in company leads to a lower salary, which leads to a high attrition rate

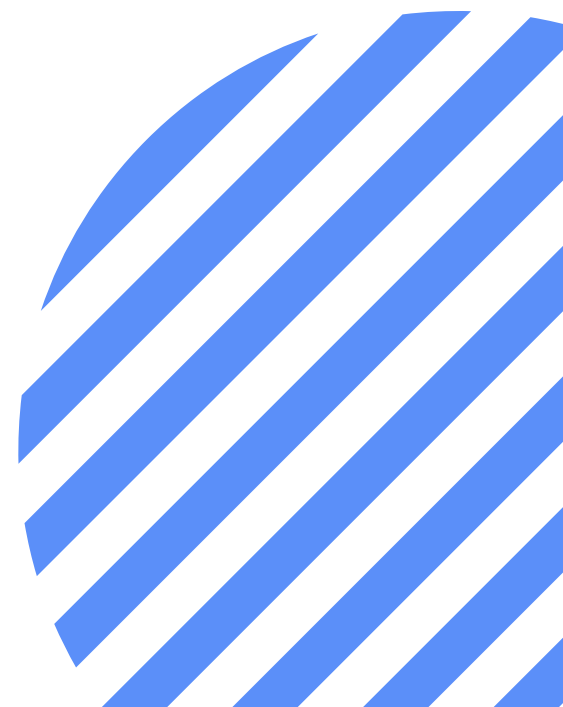


Relation between Years in Company with current manager, last promotion and current position (10-year might be a cut-off point)



Model Card

Logistics Regression has the best accuracy score is **87.9%** with the best hyperparameter



Employee Attrition Prediction Model

The model predicts the risk of employee attrition based on their demographic and working information and returns a probability score to assist Human Resource Managers to evaluate their employees. In this model card, you can understand about how the model performs on 28 features and learn about how to provide meaningful recommendations to HR based on the feature importances calculate from each input.



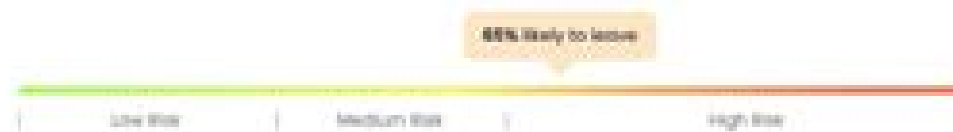
MODEL DESCRIPTION

Input: Employee's basic working information (IBM dataset features)

OverTime, JobLevel, MonthlyIncome, StockOptionLevel, TrainingTimesLastYear, BusinessTravel, DailyRate, EnvironmentSatisfaction, JobInvolvement, JobSatisfaction, WorkLifeBalance, Department, DistanceFromHome, PerformanceRating, EducationField, JobRole, NumCompaniesWorked, PercentSalaryHike, RelationshipSatisfaction, TotalWorkingYears, YearsAtCompany, YearsInCurrentRole, YearsSinceLastPromotion, YearsWithCurrManager

Output: The evaluation for the employee attrition decision. For each persona, the model will provide a leaving probability score.

- Attrition final results
- Probability
- Feature Importance
- Confidence score



LIMITATIONS

This model is established on the IBM data set, which limits the input of HR to be within the range of features included in the dataset. Although these features are splendid, it cannot contain all relevant factors. Different enterprises are hindered by different factors, resulting in unstable forecast results.

Meanwhile, the limitation of specific dataset will lead users to provide corresponding suggestions based on the conclusion of the importance of features, which requires us to further collect relevant data to make up the corresponding thresholds for each feature.

TRADE-OFF

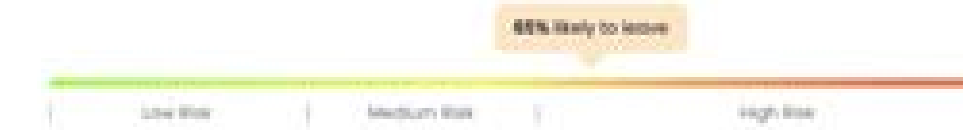
There are some personal demographic information in the dataset, including **Age, Marital Status and Gender**.

Input: Employee's basic working information (IBM dataset features)

OverTime, JobLevel, MonthlyIncome, StockOptionLevel, TrainingTimesLastYear, BusinessTravel, DailyRate, EnvironmentSatisfaction, JobInvolvement, JobSatisfaction, WorkLifeBalance, Department, DistanceFromHome, PerformanceRating, EducationField, JobRole, NumCompaniesWorked, PercentSalaryHike, RelationshipSatisfaction, TotalWorkingYears, YearsAtCompany, YearsInCurrentRole, YearsSinceLastPromotion, YearsWithCurrManager

Output: The evaluation for the employee attrition decision. For each persona, the model will provide a leaving probability score.

- Attrition final results
- Probability
- Feature Importance
- Confidence score



LIMITATIONS

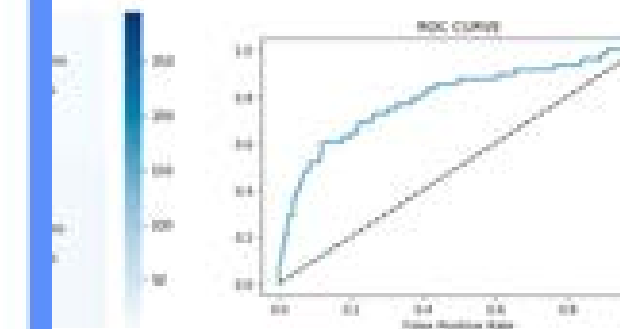
This model is established on the IBM data set, which limits the input of HR to be within the range of features included in the dataset. Although these features are splendid, it cannot contain all relevant factors. Different enterprises are hindered by different factors, resulting in unstable forecast results.

Meanwhile, the limitation of specific dataset will lead users to provide corresponding suggestions based on the conclusion of the importance of features, which requires us to further collect relevant data to make up the corresponding thresholds for each feature.

There are some personal demographic information in the dataset, including **Age, Marital Status and Gender**. Due to the diversity of Gender and the imbalance of the data, we first analyzed Gender by AI Fairness. It was found that Gender did not lead to bias. However, Age and Marital Status are closely related and both are sensitive factors. Even if they have high correlation to Attrition, in order to guarantee the fairness of the model, we excluded both.

The model's performance on the testing set by utilizing Logistics Regression Model is nearly **88%**.

The model's Precision, which is **0.98**. Disparities in recall are relatively small. The area under ROC curve is a great proven metric for our current model.



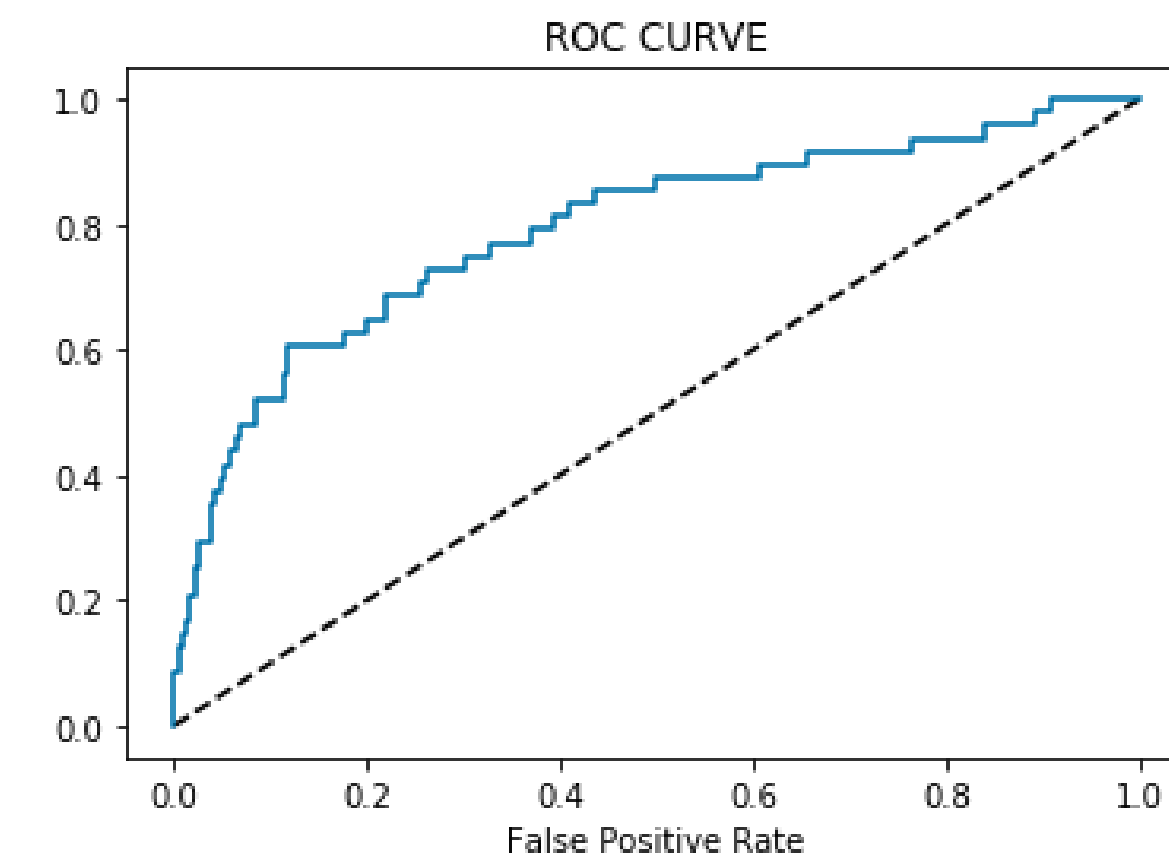
CONCLUSION

Based on the results of the dataset, if you can bring company's employee dataset with you, through our recommendation system, we would not only make your results more accurate, but also avoid the bias of the model to the sensitive data of the AI model (after we debug whether the sensitive data is biased and balance the data).

We welcome your valuable comments on our recommendation system and look forward to your suggestions on improving the functions of our recommendation system. We not only need you to enrich our data, but also need your valuable suggestions on improving the model.

AI Model

- Convert Binary Classification into Probability score in order to present the risk of employee attrition
- Confusion Matrix analysis - Precision, Recall & ROC
- SMOTE Method for Imbalanced Data
- Limitations: The dataset itself has some limitations. (Feature dissatisfaction, lack of thresholds for recommendation, unclear features categories)

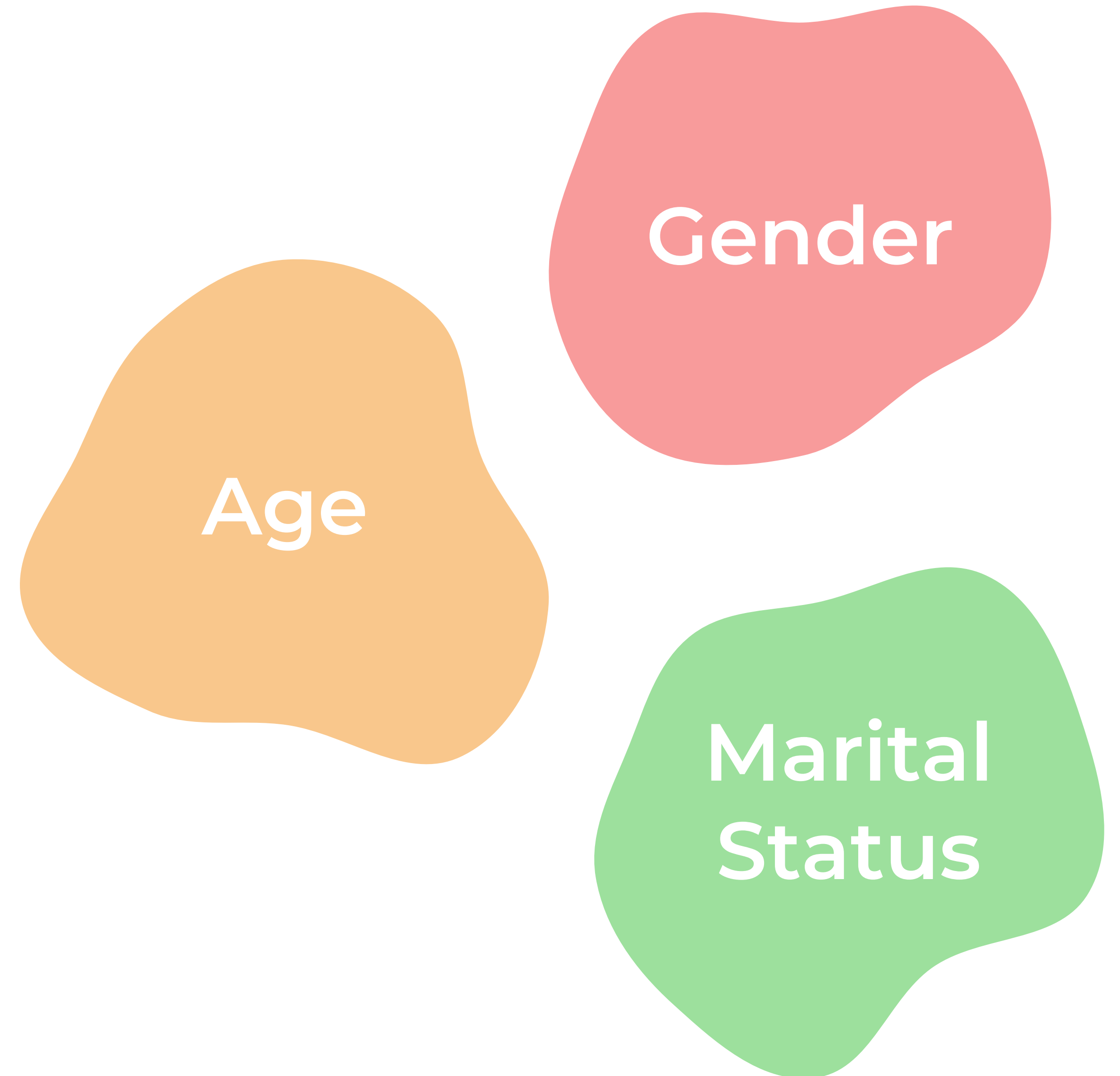


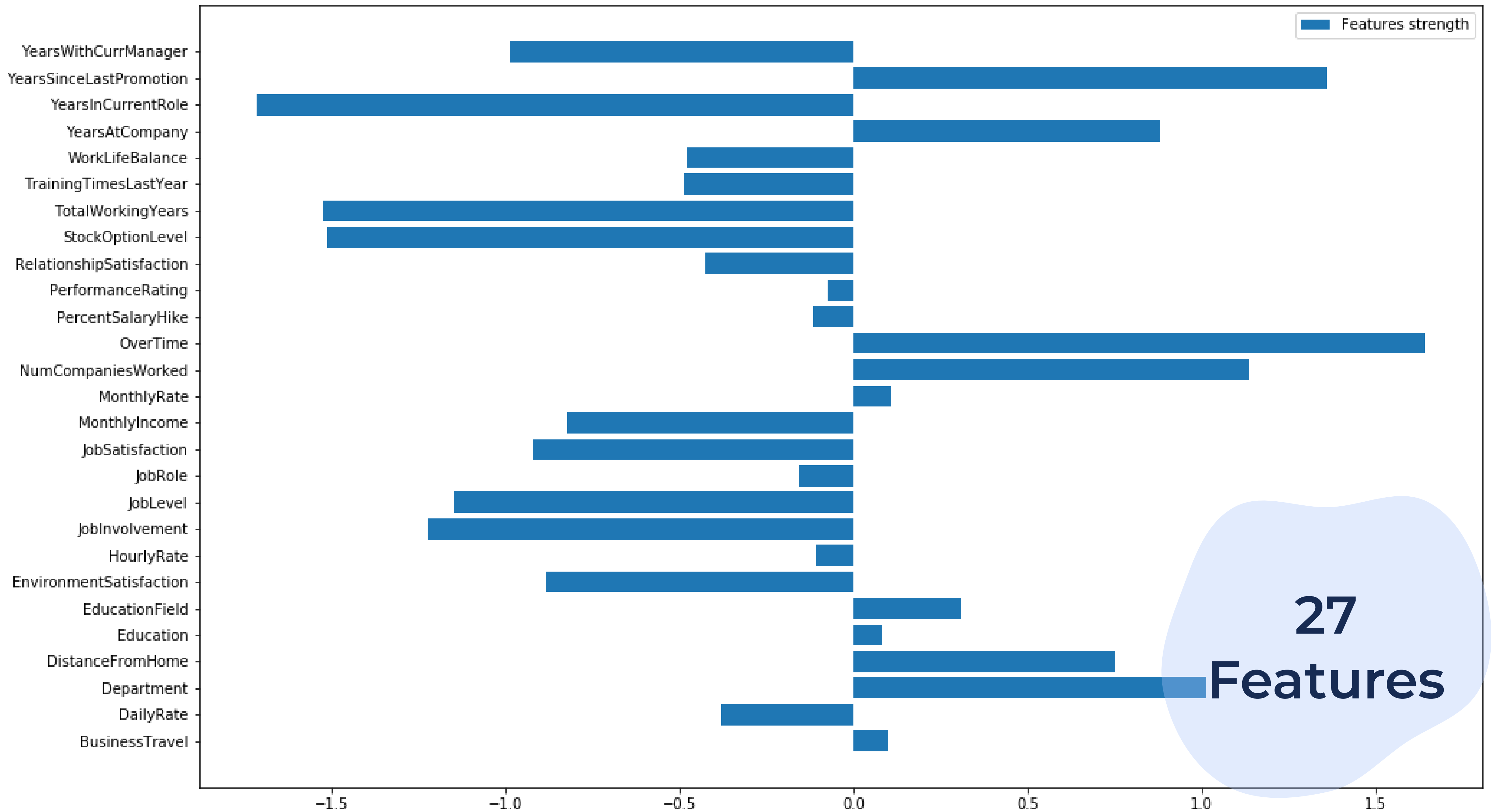
Fairness/Bias

Excluded employee sensitive demographic features as they are illegal to collect at workspaces

AI fairness metrics

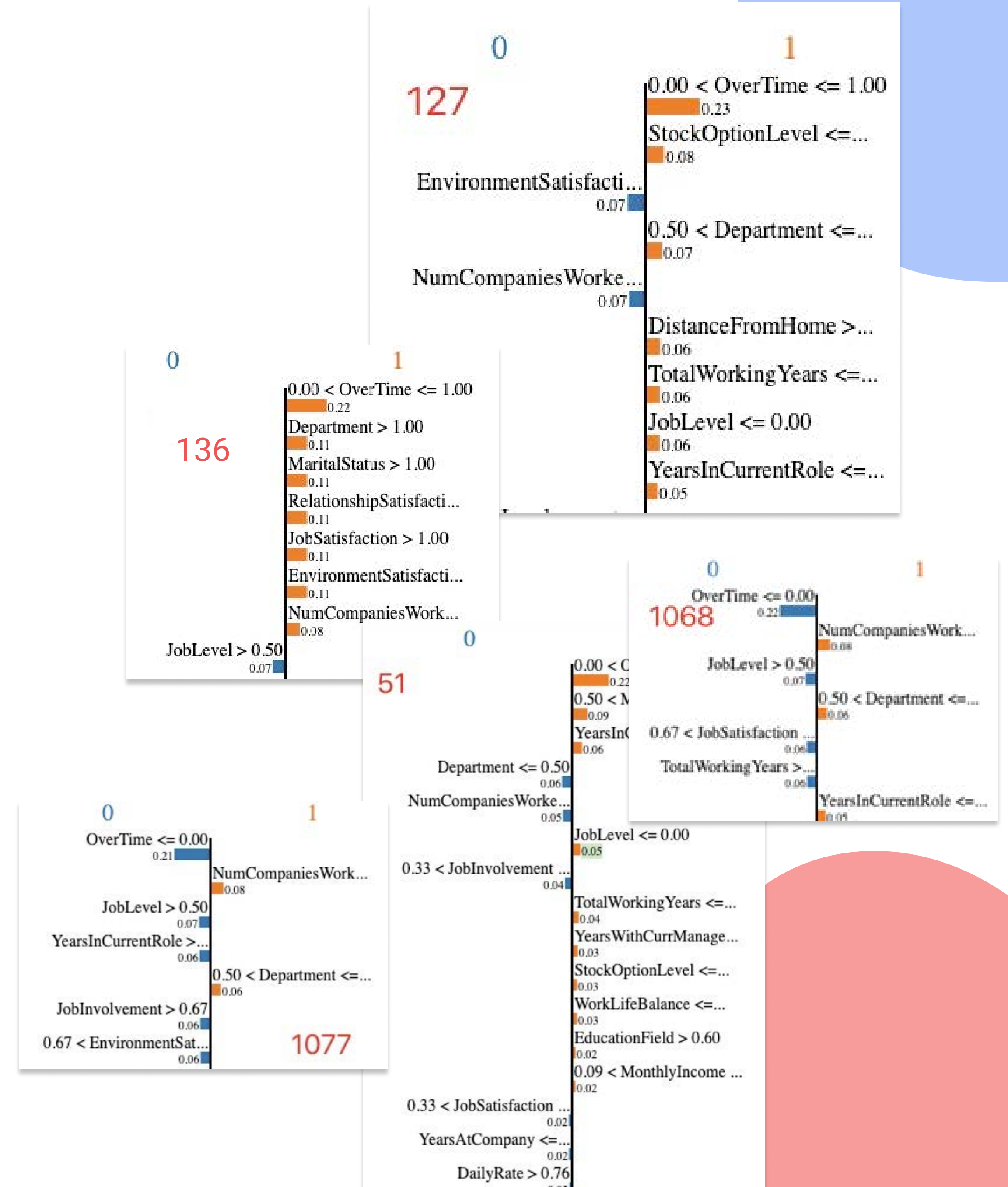
- Equalized odds
- Test Fairness
- Counterfactual Fairness



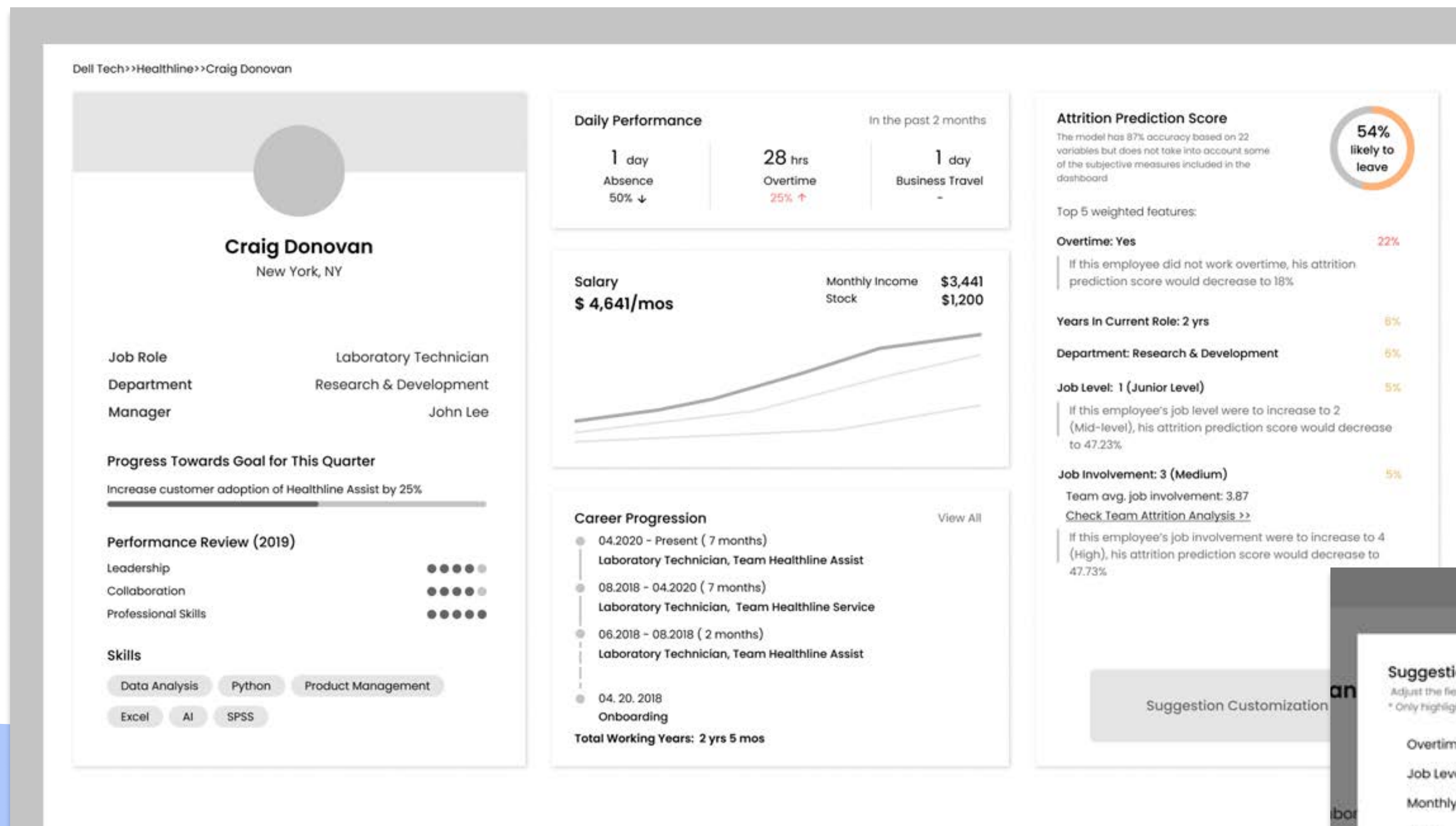


Persona

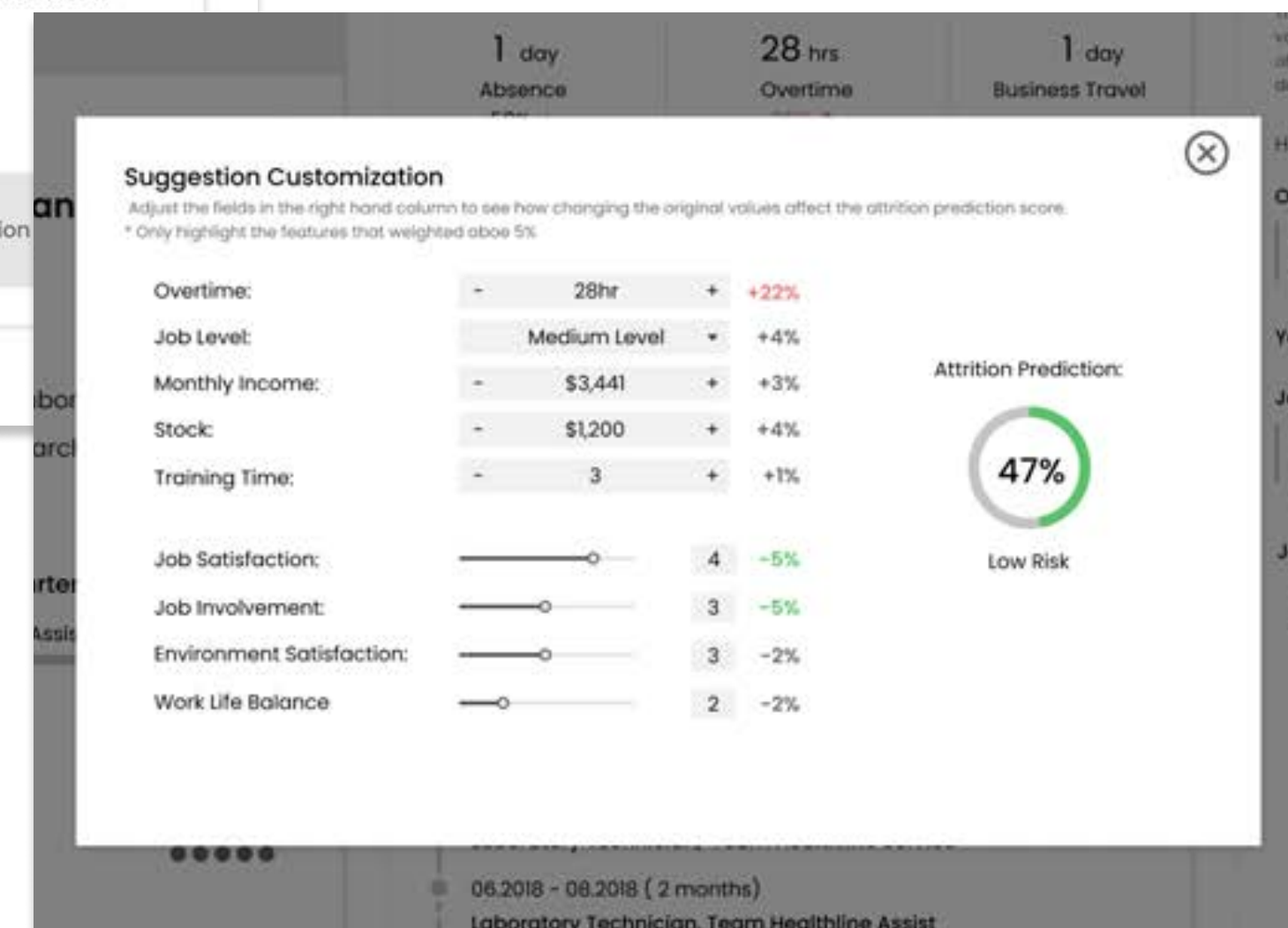
- Overtime is always the top reason for attrition
- Salary ranks much lower than expected, and has a low impact on attrition
- Easy to have stereotypes on unactionable features
- Most features weigh lower than 5%



Dashboard Design



- Performance vs. Attrition
- Attrition Explanation
 - Percentage Risk
 - Highly Weighted Features
 - Understandable Language
 - Suggestion Customization



User Testing

3 users: 1 manager, 1 HR, 1 Employee

- Like the performance metric
- Outcome explanations themselves were clear but not understandable
- Customization was too granular and overwhelming to be useful
- Employee had a negative reaction

The image shows a screenshot of a dashboard for 'Attrition Prediction Score'. The main heading is 'Attrition Prediction Score' with a sub-note: 'The model has 87% accuracy based on 22 variables but does not take into account some of the subjective measures included in the dashboard'. A circular gauge indicates '54% likely to leave'. Below this, 'Top 5 weighted features' are listed: 'Overtime: Yes' (22%), 'Years In Current Role: 2 yrs' (8%), 'Department: Research & Development' (6%), 'Job Level: 1 (Junior Level)' (5%), and 'Job Involvement: 3 (Medium)' (5%). Each feature has a brief explanation of how it affects the score. A 'Suggestion Customization' panel is open, showing a table of adjustable variables and their impact on the score. A 'Trust 3/5' overlay is present in the top right, and a '47% Low Risk' gauge is in the bottom right.

Attrition Prediction Score
The model has 87% accuracy based on 22 variables but does not take into account some of the subjective measures included in the dashboard

54% likely to leave

Top 5 weighted features:

- Overtime: Yes** 22%
If this employee did not work overtime, his attrition prediction score would decrease to 18%
- Years In Current Role: 2 yrs** 8%
- Department: Research & Development** 6%
- Job Level: 1 (Junior Level)** 5%
If this employee's job level were to increase to 2 (Mid-level), his attrition prediction score would decrease to 47.23%
- Job Involvement: 3 (Medium)** 5%
Team avg. job involvement: 3.87
[Check Team Attrition Analysis >>](#)
If this employee's job involvement were to increase to 4 (High), his attrition prediction score would decrease to 47.73%

Suggestion Customization
Adjust the fields in the right hand column to see how changing the original values affect the attrition prediction score.
* Only highlight the features that weighted above 5%

Overtime:	- 28hr +	+22%
Job Level:	Medium Level ▾	+4%
Monthly Income:	- \$3,441 +	+3%
Stock:	- \$1,200 +	+4%
Training Time:	- 3 +	+1%
Job Satisfaction:	—○— 4	-5%
Job Involvement:	—○— 3	-5%
Environment Satisfaction:	—○— 3	-2%
Work Life Balance	—○— 2	-2%

Attrition Prediction:
47%
Low Risk

Trust 3/5

Personal Dashboard

Healthline Product Team >> Craig Donovan

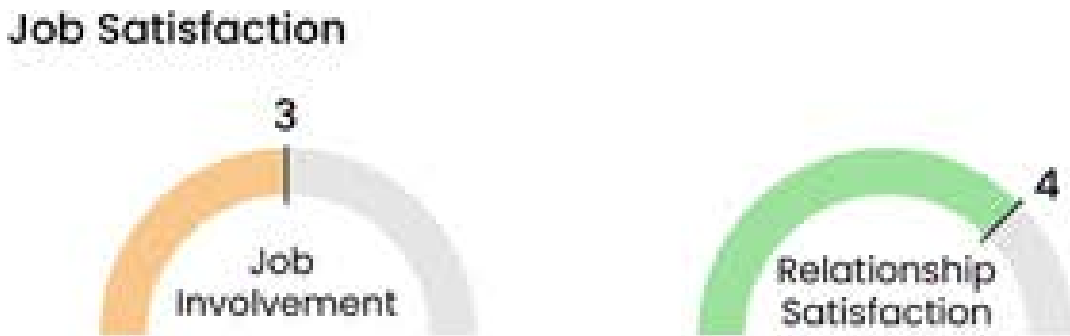


Craig Donovan
New York, NY

Department: Sales
Job Title: Sales Representative
Manager: [John Lee](#)
Current Team: [Healthline Product Team](#)

Total Working Yrs: 1 yrs 8mos
Yrs in Current Team: 8 mos

- Career Progression
- 04.2020 - 11. 2020 (7 months)
Sales Trainer , Healthline Assist Team
 - 12.2019 - 04.2020 (5 months)
Sales Trainer , Healthline Product Team



[View Full Report >>](#)

AI Attrition Prediction

*The model has 87% accuracy based on 22 quantitative variables but the attrition rate may also be related to other subjective reasons.

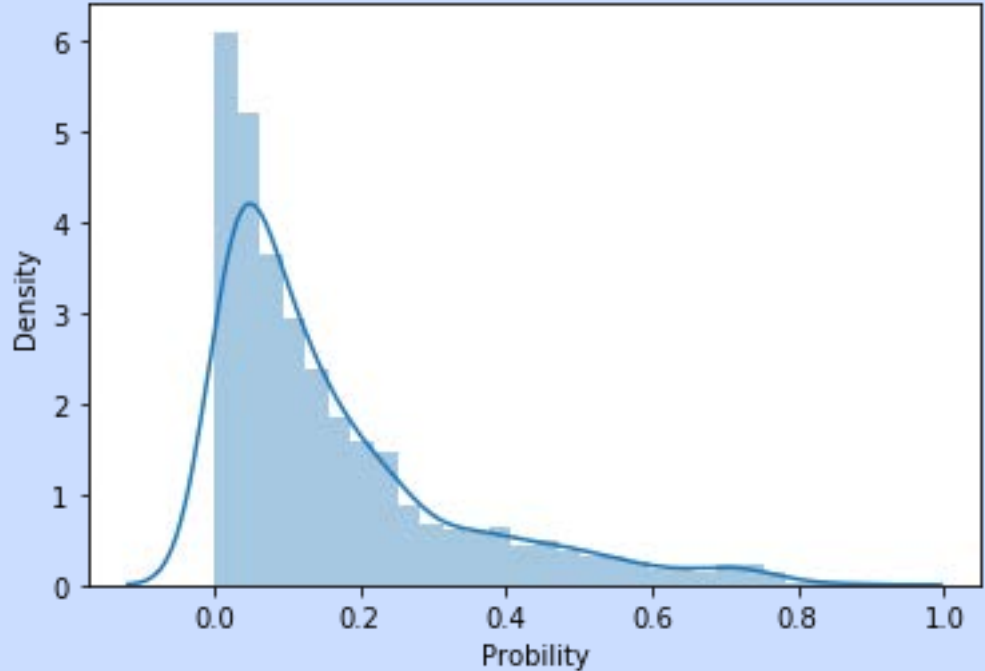


[View AI Model Explanation >>](#)

Personal Dashboard

"I don't even know how to make sense of the 65% because everybody in the company has a 50/50 chance of leaving."

Minimize the impact of attrition prediction
More as a support for the performance



Most of the employees' attrition risk are lower than 20%

AI Attrition Prediction

*The model has 87% accuracy based on 22 quantitative variables but the attrition rate may also be related to other subjective reasons.

65% likely to leave

Low Medium High

[View AI Model Explanation >>](#)



Attrition Explanation

< Back

AI Attrition Prediction Explanation

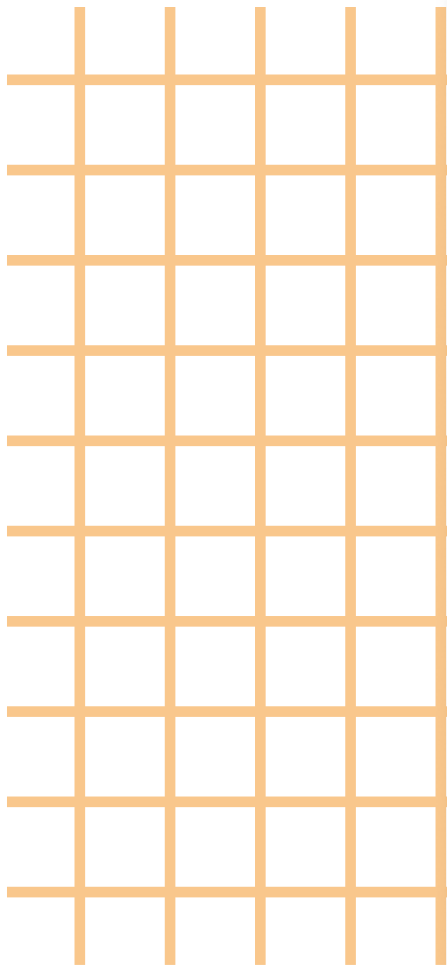
The model has 87% accuracy based on 22 quantitative variables but the attrition rate may also be related to other subjective reasons

- High Impact
- Medium Impact
- Low Impact
- Negative Impact

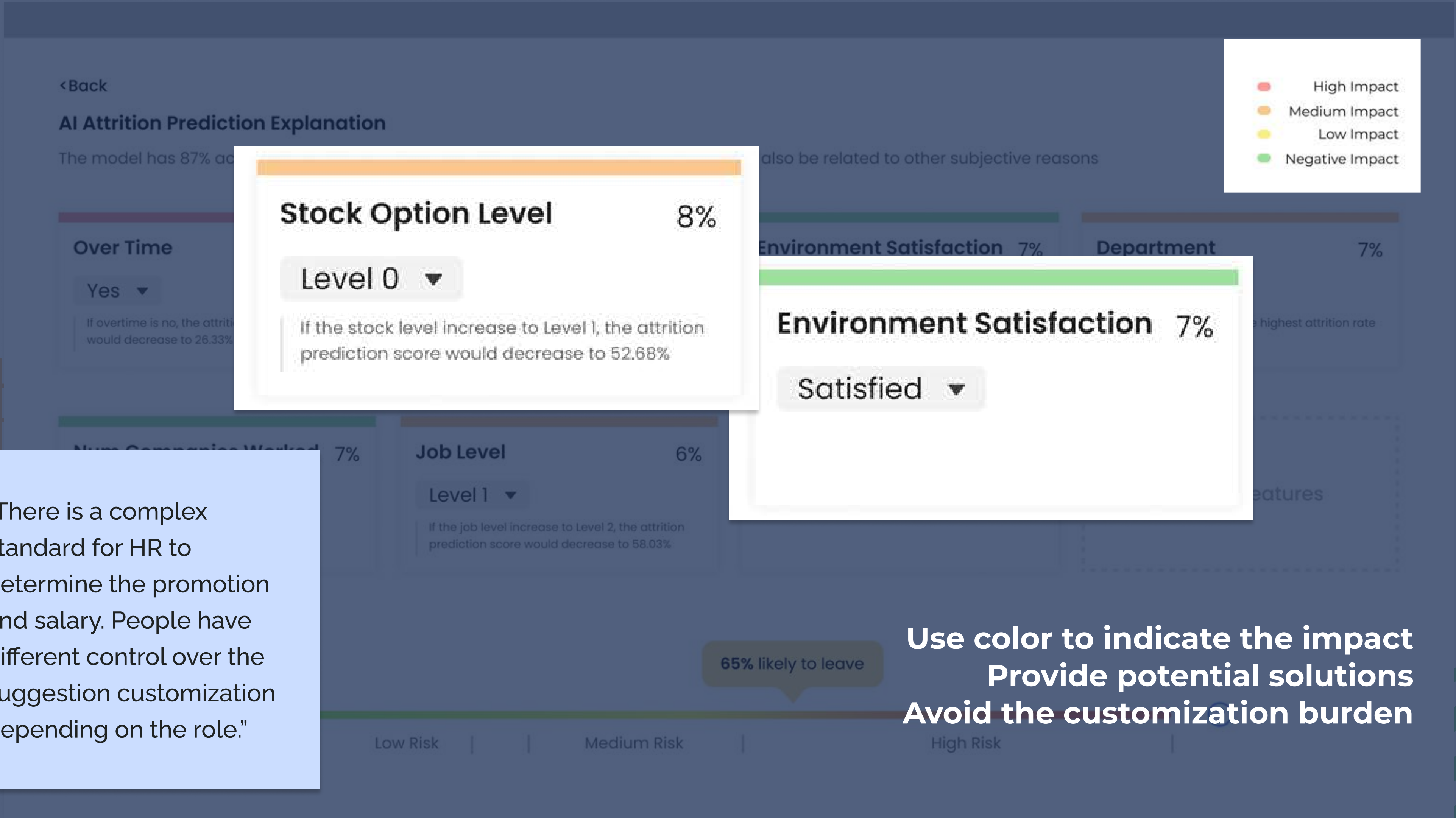
Over Time 23% Yes ▾ If overtime is no, the attrition prediction score would decrease to 26.33%	Stock Option Level 8% Level 0 ▾ If the stock level increase to Level 1, the attrition prediction score would decrease to 52.68%	Environment Satisfaction 7% Satisfied ▾	Department 7% Sales Department Sales has the highest attrition rate in the company
Num Companies Worked 7% 1	Job Level 6% Level 1 ▾ If the job level increase to Level 2, the attrition prediction score would decrease to 58.03%	Yrs in Current Role 6% 0	+ Other Features

65% likely to leave

Low Risk | Medium Risk | High Risk



Attrition Explanation



“There is a complex standard for HR to determine the promotion and salary. People have different control over the suggestion customization depending on the role.”

Use color to indicate the impact
Provide potential solutions
Avoid the customization burden

Attrition Explanation

< Back

AI Attrition Prediction Explanation

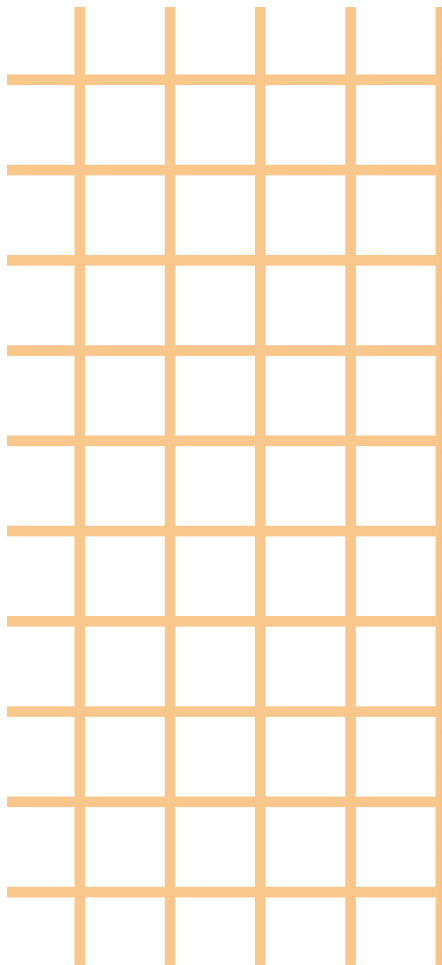
The model has 87% accuracy based on 22 quantitative variables but the attrition rate may also be related to other subjective reasons

- High Impact
- Medium Impact
- Low Impact
- Negative Impact

Feature	Impact	Value	Change Description
Over Time	High Impact	23%	Yes If overtime is no, the attrition prediction score would decrease to 18.23%
Stock Option Level	Medium Impact	8%	Level 0 Level 1 Level 2 Level 3 Increase to Level 1, the attrition could decrease to 47.23%
Environment Satisfaction	Negative Impact	7%	Satisfied
Department	Medium Impact	7%	Sales Department Sales has the highest attrition rate in the company
Num Companies Worked	Negative Impact	7%	1
Job Level	Medium Impact	6%	Level 2
Yrs in Current Role	Medium Impact	6%	0
+ Other Features			

48% likely to leave

Low Risk | Medium Risk | High Risk



Team Dashboard

Healthline Product Team

This team is responsible for connecting doctors, hospitals, researchers and patients on a shared technology platform to improve the patient experience

Current Employees

32

Attritioned Employees 2020

4 12%↓

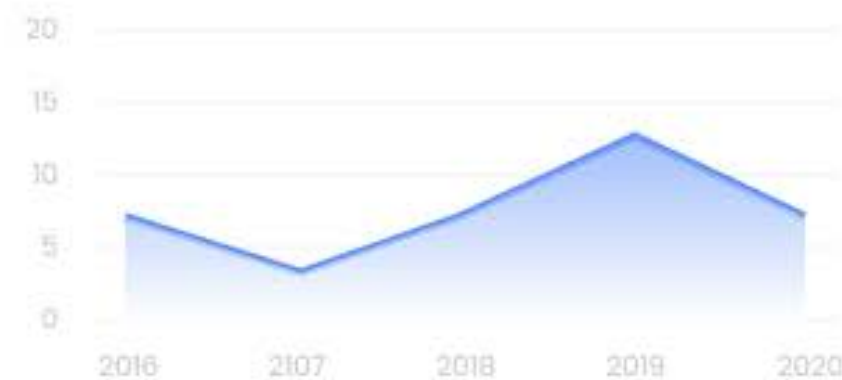
Performance Rate 2019-2020

86 25%↑

Avg. Yrs in the Team

4.78

Attrition History 2016-2020



Top Reasons for Attrition

1. Monthly Income(28)
2. Years with Current Managers(21)
3. Job Level(18)
4. Total Years(17)
5. Job Satisfaction(15)

Top Jobs Roles for Attrition

1. Sales Repreresitives(8)
2. Software Developer(7)
3. Product Manager(4)
4. Sales Manager(3)
5. Senior Product Manager(3)

Satisfaction Survey Rate 2019-2020

Job Involvement 12% ↗



Relationship Satisfaction 8% ↗



Work Life Balance 7% ↘

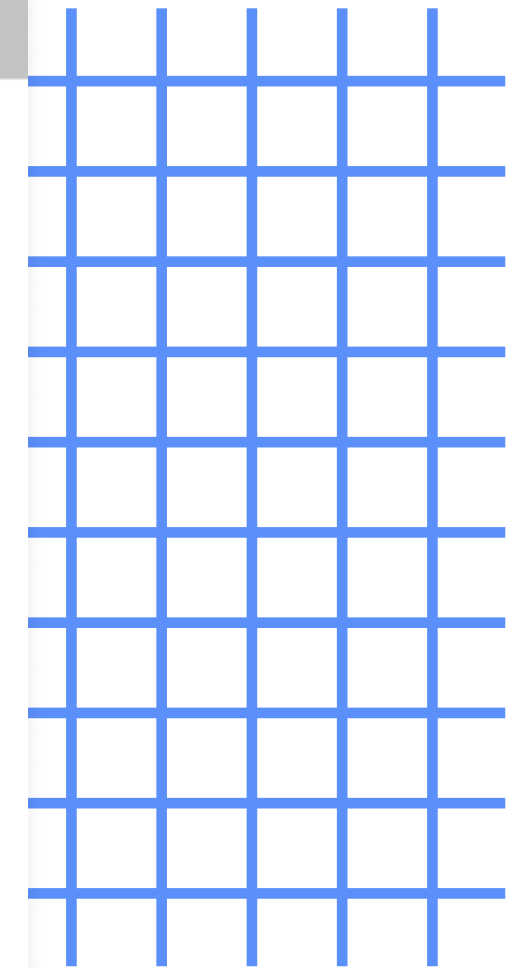


Current Employees

Attritioned Employees

Filter

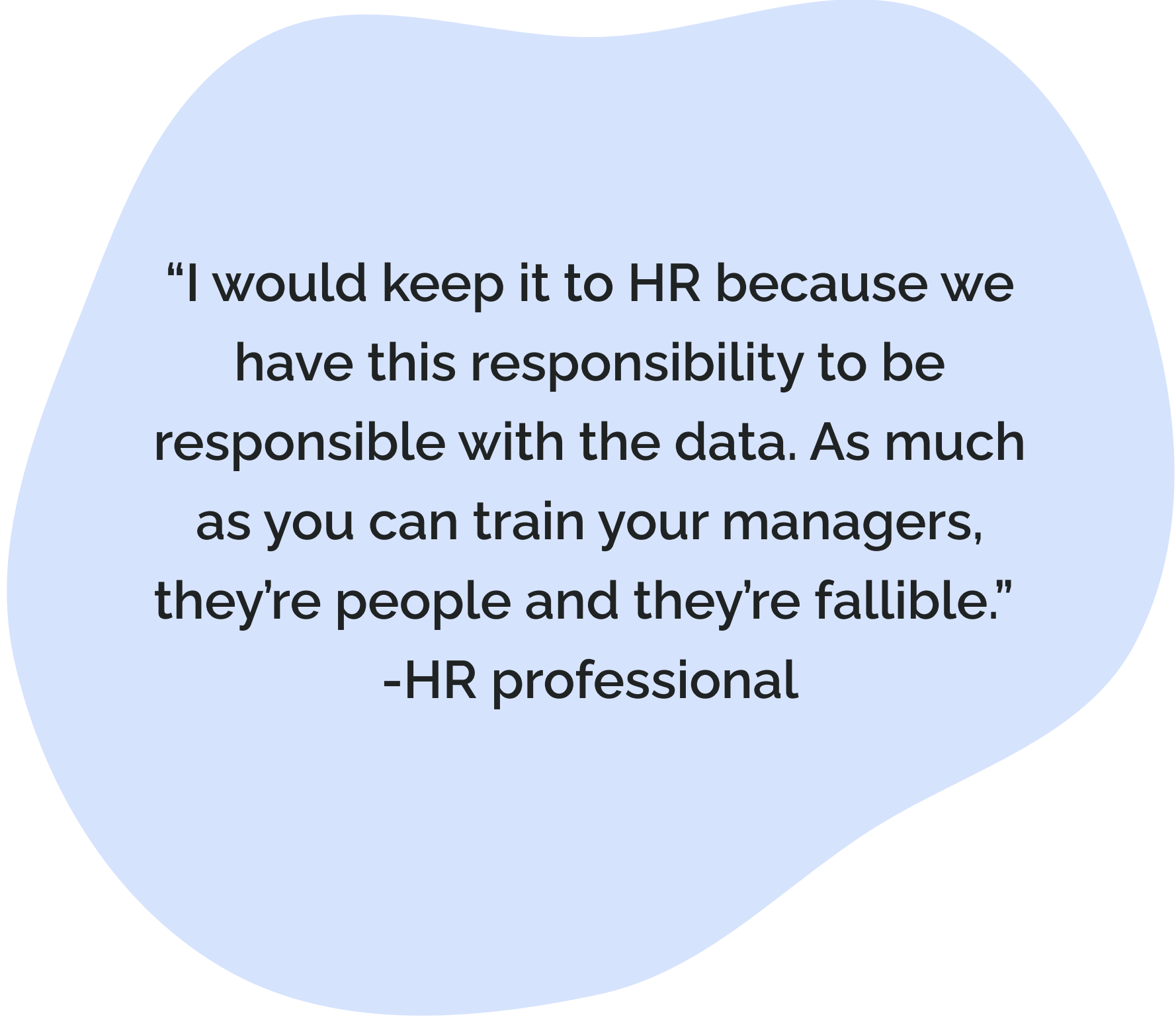
Name	Job Role	Yrs in the Team	Performance Rate	Risk
Robert Fox	President of Sales	5 Yrs 2 mos	87	Low Risk
Cameron Williamson	Marketing Coordinator	5 Yrs 2 mos	78	Medium Risk
Brooklyn Simmons	Marketing Coordinator	4 Yrs 8 mos	83	Low Risk
Darlene Robertson	Software Development Manager	3 Yrs 2 mos	90	Low Risk
Craig Donovan	Sales Representative	1 Yrs 2 mos	87	High Risk
Maria Mufson	Marketing Coordinator	1 Yrs 2 mos	88	Medium Risk



Guidelines for Future Use

Guidelines for HR

- Ideal for large companies where HR is a “black box” and HR does not have a 1:1 relationship with every employee or see dynamics within various teams
- Team-level view of attrition may be more valuable than individual-level
- Best for flagging flight risk for senior-level or above
- Neither employees nor managers would not have access to the platform (but do employees have a right to know?)



“I would keep it to HR because we have this responsibility to be responsible with the data. As much as you can train your managers, they’re people and they’re fallible.”
-HR professional

Implications for Future Use

False positive/negative scenarios

“How do you explain the idea of an AI thing that's judging you behind the books without showing them so they don't freak out? Also, you show it to them and they'll know what they can do to game it. They'll know the measures and then it becomes irrelevant.”

-Manager



Challenges & Lessons Learned

- Limitations with the dataset
- Communication with technical/non-technical
- How to design for data



Thank you!

Questions?

Human-AI Interaction | 2020 Fall

Janet Dai, Yi Wang, Dianne Kim

