# Interactive Visual Communication of Predictive Workforce Analytics

Shubir Kapoor*, Steven Rohall*, Aleksandra Mojsilovic*, Donna Gresh*, Moninder Singh*, Deepika Kakrania*

IBM T.J. Watson Research Center, Yorktown Heights, NY USA

## ABSTRACT

We describe a framework and methodology that combines enterprise data from human resource systems, with social engagement data from enterprise collaboration systems, then applies predictive modeling to identify employees at risk of voluntary attrition. We provide an interactive visual interface that allows line of business users and HR analysts to 1) explore the data and identify attrition patterns in the business, 2) identify population segments with elevated attrition risk and 3) design effective programs and actions to proactively combat attrition in the organization. By proactively identifying top talent at high risk of voluntarily leaving, an organization can take appropriate action in time to reduce employee departures.

**Keywords: Retention, attrition, visualization, predictive modeling, machine learning, workforce analytics.**

**Index Terms: Mathematics of computing~Decision diagrams, Applied computing~Decision analysis, Applied computing~Forecasting.**

## 1 INTRODUCTION

Developing predictive models and the algorithms that lead from data to decisions is a critical component of business analytic applications. However, equally important is the effective presentation of the data and the results of analyses to enable informed actions and ultimately guide decisions that could positively affect enterprise key performance indicators.

In our retention[1] analytics solution we use a data-driven approach to understanding attrition patterns within a business. We apply predictive modeling to rich enterprise data to identify high-attrition segments within the company and understand which employees are at risk of attrition and why. The interactive visual platform aids the business decision makers in identifying high attrition clusters within an organization, *e.g.* in a specific country, business unit, job category, or performance rating. The solution also "guides" the user in discovering the characteristics of the at-risk population and the key "influencers" of attrition. One of the actions frequently and successfully deployed by organizations is compensation. As a result, in addition to attrition prediction, the retention analytics solution also implements compensation analytics, which aids in identifying the most effective size and scope of compensation actions.

## 2 MODEL, RESULTS AND FINDINGS

The models and associated visualizations were applied to a proactive retention action for several thousand employees across several geographies and business units for a large Fortune 500 company. The models developed were used to choose employees with the best benefit/cost ratio for a targeted retention action based on the (i) expected probability of attrition in the next 12 months, (ii) cost of replacing them, including hiring/training costs and salary premium, and (iii) net benefit achieved by trying to retain them via compensation investments over the cost of replacing them if they left.

The four pronged framework for tackling voluntary employee attrition comprises the following steps:

1. *Attrition insights*: Based on a proprietary predictive model[2] using decision trees, the system helps to understand reasons for voluntary attrition in terms of actionable attributes so that appropriate retention actions can be formulated. Feature selection and feature creation were very important in determining the performance of the models. For example, salary by itself was not a strong indicator. However, by comparing salaries of individuals to peer groups, defined by various dimensions such as country, salary band, job role, and years of service, strong relationships could be observed.

2. *Attrition Costs*: This involved the development of techniques and/or gathering of data for determining salary premiums and hiring costs. We developed statistical models for learning salary premiums from historical data by looking at salaries of new hires as well as recent attriters. Hiring costs involve costs such as recruiting fees, sign-on bonuses, training, onboarding costs, etc., and are also important in deciding whether to invest in trying to retain a potential attriter.

3. *Optimizing compensation investments*: An analytics model was developed to determine the size of raises, as well as the population to target. This is achieved by modeling the effect of raises on attrition by mining historical data on prior off-cycle salary actions, as well as by interviewing HR experts. There is a point, for example, at which it becomes too costly to try to retain a certain employee population even considering lost work and replacement costs.

4. *Disbursement to employees*: We developed an optimization framework to optimally assign a fixed investment amount to a population of identified high-risk attriters so as to maximize the cumulative expected benefit of retaining them over some future time horizon.

## 3 INTERACTIVE VISUAL ANALYTICS

### 3.1 Current Prototype

Many analytical applications assume that users know precisely what they need before they've begun their analysis. However, frequently the users depend on the application to assist them in finding an answer with only a vague idea of where to start. To a business user not familiar with data exploration and analytics, such exploratory analysis can often feel like swimming upstream, and it is important to design the visual experience to facilitate the journey to allow improved decision-making by allowing consideration of outcomes and their implications.

* shubirk@us.ibm.com, steven_rohall@us.ibm.com, aleksand@us.ibm.com, donna@us.ibm.com, moninder@us.ibm.com, deepi@us.ibm.com

The attrition modeling dataset is high-dimensional. Employee geography, for example, is specified by market segment, region, and country. Employee business unit includes geographic segment, unit, and line. Data on employee job category and performance is included, as are an employee's attrition score, the size of the segment and key attrition influencers (e.g. compensation, performance, experience). Compensation information is more than just salary and includes bonuses, commissions, salary compared to peers, salary compared to market, time since last raise or bonus, etc. The challenge, therefore from a visualization perspective was to display this multi-dimensional dataset so that it communicates the analytic insights and allows line-of-business users to take the effective actions. Although the analytic models produce a probability of an employee's likelihood to leave, the probability alone would be of little use to the user. Instead, we wanted to emphasize the differences between populations—not that someone was likely to leave, but rather that they were more likely to leave than others in a similar segment.

To accomplish this, we designed three coordinated treemaps[3], each showing a particular categorical variable from the data: geography, business unit, and job information. The colors used were a non-linear red-green ramp with the switch from green to red at about 10%, the acceptable attrition rate for this data set. As a user drills in (e.g., into the "store manager" job role), all of the treemaps update. In this example, we can see that store managers in India are much more likely to leave relative to other geographies (Figure 1).

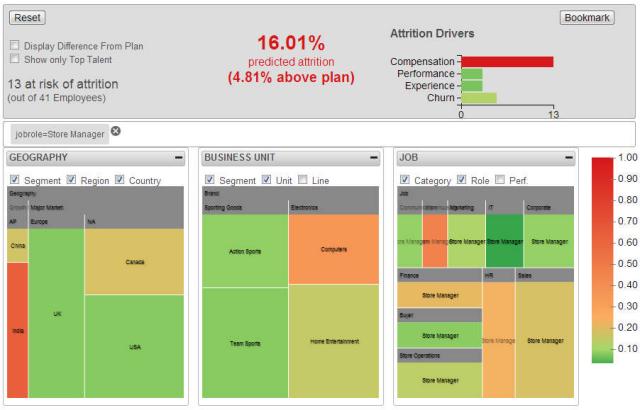


Figure 1: Coordinated treemaps show 3 different categorical dimensions of the data.

At the same time, summary statistics indicating the number of employees in the selected group, their aggregate predicted attrition, and the difference from the planned attrition rate are updated interactively in real time. This information is displayed at the top of the user interface. A simple bar chart (Figure 2) shows the 4 key attrition drivers allowing the user to understand the key influencers for attrition (compensation, in this example).

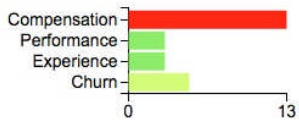


Figure 2: Compensation is the largest driver of attrition.

The bar graph shows that, for this population, compensation is the biggest driver for the high predicted attrition rate. Clicking on the compensation bar brings up the Compensation Adjuster (Figure 3) that guides the user on the investments required, cost savings incurred and projected net benefits achieved by reducing attrition to a desired state. This capability allows the user to ask "what if" questions and more optimally firm up the compensation budget.

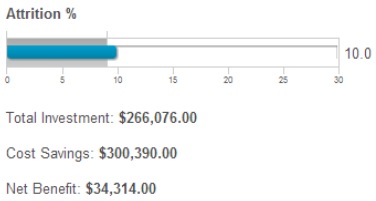


Figure 3: Compensation adjuster enables optimizing compensation investments

Bookmarks allow a user to capture and share their investigation with other users. In Figure 4, a bookmark of the store manager investigation is represented as a bullet chart showing summary data for that population. Predicted attrition is shown as the black

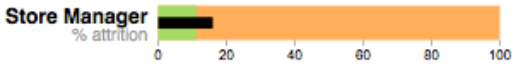


Figure 4: Bullet chart showing summary data for "store manager" role.

bar; the longer orange bar shows that this population is above the planned, acceptable attrition rate. Clicking on this bookmark returns the other visualizations in the interface to the correct state. Finally, at any time, the user is able to get a detailed view showing individual employees' predicted attrition rates as well as their individual attrition drivers.

### 3.2 Future Visualizations

One of the novel contributions of this work is the addition of corporate social networking data into the attrition model. It seems reasonable, for example, that if a person's friends and colleagues are leaving the company, that that person might also have a higher likelihood of leaving. However, this effect has not been quantified. We are currently experimenting with overlaying attrition prediction data on the corporate social network. This will aid with decisions to intervene based upon people who are key influencers in the corporate network.

## 4 Conclusion

Retaining their best employees is one of the most important challenges faced by an organization. In this work we have demonstrated an approach combining predictive modeling and interactive visualization to communicate the results intuitively to a line of business user. The visualization supports the lifecycle of decision making from insights to evidence and finally action, guiding the user at each step of the analysis. Evaluating the effectiveness of the actions as an additional visual guide for effective decisions is a planned future area of research.

## 5 References

[1] Holtom, B. C., Mitchell, T. R., Lee, T. W., & Eberly, M. B. 5 Turnover and Retention Research: A Glance at the Past, a Closer Review of the Present, and a Venture into the Future. The Academy of Management Annals, 2(1), 231-274, 2008

[2] Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.

[3] Shneiderman, B. Tree visualization with treemaps: a 2-d space-filling approach, ACM Transactions on Graphics, vol. 11, 1 (Jan. 1992) 92-99