

# Talent Flow Modeling: Using machine learning to predict workforce transitions

Steven T. Hunt, Ph.D. (with Jenngang Shih)

Click to follow my posts: [www.linkedin.com/in/steventhunt](http://www.linkedin.com/in/steventhunt)

The term “talent flow” refers to workforce changes associated with the job status of employees. It encompasses employee hires, promotions, job transfers, attendance and turnover. A characteristic of talent flow data is it is usually objective and discrete since it reflects employees’ formal contract status with their company or their physical presence at work. It is often relatively easy to collect large samples of talent flow data by tapping into organizations’ HCM systems. These things [make talent flow day well suited for modeling using artificial intelligence and machine learning applications](#).

SAP SuccessFactors is [undertaking a variety of project that use machine learning to model and predict talent flow](#). Most of these projects focus on employee turnover and retention. Other projects examine topics such as internal candidate applications, hiring, promotions, attendance, and job transitions. This is the first in a series of articles exploring insights and lessons we have gained from these various talent flow projects.

Table 1 provides an example of some of the insights we are gaining through this work. The table shows the results of using machine learning to model and predict retention across 10 different organizations. The data used for this analysis was drawn from different customer HCM systems. A more detailed explanation of the data and analysis the led to Table 1 can be found here (cite technical paper on machine learning work).

Three valuable insights can be drawn from the results in Table 1. First, machine learning applications reliably identified employee retention risk across all 10 companies. This tells us is there is meaningful variance in the talent flow data available from company HCM systems. This may seem like a trivial observation, but it is extremely important. Some people assume that all “big data” is inherently valuable. This is not true. Just because we have a lot of data does not mean the data has a lot of useful information. The value of data does not depend on the size of the dataset nearly as much as the quality of information it contains. What Table 1 shows is HCM datasets contain useful information for predicting talent flow outcomes such as turnover.

Table 1. Employee Characteristics Predicting Turnover in Different Organizations

Industry	# of Employees	Employee Characteristics Most Strongly Predictive of Turnover	Predictive Power	Prediction Confidence
Financial Services	5,000 to 10,000	Job Title, Organization Tenure, Position Title	Moderate	Medium
Asset Management	1000 to 5000	Last Performance Rating, Job Function, Organization Tenure	Strong	High
Semiconductors	5,000 to 10,000	Last Performance Rating, Job Function, Organization Tenure	Strong	High
Medical Equipment	20,000 to 30,000	Job Family, Performance Rating, Organization Tenure	Moderate	High
Banking	10,000 to 20,000	Organization Tenure, Job Family, Age	Moderate	High
Electric Utilities	5,000 to 10,000	Job Function, Age, Organization Tenure	Strong	High

<b>Oil &amp; Gas</b>	10,000 to 20,000	Business Function, Strategic Business Unit, Organization Tenure	Moderate	Medium
<b>Insurance</b>	30,000 to 40,000	Employment Status, Organization Tenure, Job Grade	Moderate	High
<b>Software technology</b>	10,000 to 20,000	Age, Employment Status, Job Function	Strong	High
<b>Manufacturing</b>	20,000 to 30,000	Position, Organization Tenure, Local Job Grade	Strong	High

The second valuable insight is the fact that predictors of retention vary depending on the organization. This suggests that factors that drive turnover change significantly from one organization to the next. Consequently, issues that drive turnover in one company may be of little importance in another. Interventions to increase retention that work in one company may not work elsewhere. This suggests that companies should ideally conduct turnover analysis using their own data. It is risky to assume that factors associated with turnover risk in one company will generalize.

The third major insight concerns the specific predictors listed in the third column. This column shows what sort of employee characteristics are associated with turnover. But it does not tell us why these characteristics predict turnover. We do not know if the relationships between these predictors and turnover are a result of things that can be directly influenced by companies such as employee engagement, or reflect socioeconomic realities such as mandatory retirement ages that are largely outside of company control. Table 1 tells us there is meaningful information to be found in HCM talent flow data, but more research is needed to understand how this information should guide HCM strategies and interventions.

In future articles, we will dive into exploring and discussing the data behind the results in Table 1 along with data from other SAP machine learning projects analyzing talent flow. This includes going beyond empirical summaries to look at potential theoretical reasons underlying the results. We will also discuss methodological issues and analytical lessons learned as we delve into this new area of human capital management analytics. Please let us know if there are any questions you would like us to address. The journey is just beginning!