Using Predictive Analytics in HR in Eight Steps

About the Author

Tej Mehta

Tej Mehta is an entrepreneur, consultant and student of human sciences.

He has founded OWEN Analytics with the objective of uncovering deep insights about organizations and their employees using data. Tej and his team utilize latest concepts in AI, machine learning and ONA (Organization Network Analysis) to develop predictive as well as prescriptive solutions for people analytics.

Previously, Tej was a Vice President at Seabury Group, the largest aviation advisory in the world. He worked with C-level clients to assist them with restructuring, strategy and operations issues. He has lived in India, US and Europe, and worked in over 20 countries around the world.

Tej has an MBA degree from University of Southern California.

Article Contributed by S. Ramesh

Using Predictive Analytics in HR in Eight Steps

Analytics typically goes on a 3 stage maturity model.

- Descriptive
- Predictive
- Prescriptive

Descriptive analysis is what we do on an everyday basis. All our measures are descriptive in nature. How many people do we employ, what is our gender distribution, how many offers we have made in the last quarter all either describe an action that has happened or the status of an organization or the efficiency of a function. Within this of course, the measures can indicate either an activity or an outcome.

We have interviewed 250 people is an activity measure. While the fact that 125 people have joined us is an outcome measure. Within this, we calculate ratios on the basis of historic performance. We create measures like Join ratios and build a recruitment pyramid to arrive at ratios. This really works on the basis that if on an average 70% of all offered employees have joined us; the number is likely to be the same going forward. If we need an additional headcount of 200 people in 3 months' time, we should make 200/ 0.7 = 285 offers.

This type of analysis is typically used for projecting into the future and prescribing what should happen. So, why is this not predictive analytics?

The big assumption we are making when using the ratios is that past performance is a likely indicator for future performance. In general, tolerances within HR accept the variations arising from extrapolation. You may not get it right, but you won't be too wrong either.

However, Subtly Reframing the Question, Changes our Perspective.

1. What if we really have the resources only to hire within a +/- 5% range? If we need 200 people we need to make sure it happens by making only 220 offers?
2. What if we need to track the number of offers dynamically as the market conditions are changing rapidly? Our final objective could be a range between 150 to 200 that we do not even know now?

So, a static, extrapolation based prediction works in most cases. However it is not very efficient or dynamic. We would need a more dynamic way of prediction to manage in such situations.

Based on experiences with predictive retention modeling, the following steps need to be undertaken for a predictive model in HR with high degree of accuracy. The example relates to the context of retention, but the principles hold good across the board.

1. **Visualize the business case**

   A proper exercise in predictive analysis needs rigor as well as accurate data. This needs investment of time from not only the specialists but also business. There is any number of scenarios for predicting outcomes. However, the following have gained traction because of their business impact.

   - Retention modeling
   - Offer to join modeling
   - Competency to performance modeling.

It is not enough if we just say we want to increase retention. One of the organizations employed 2500 people in frontline sales. This function had high attrition to the tune of 35%. In terms of lost opportunities, lower productivity and cost of training this cost the company around Rs 100 crores per annum.

A reduction of attrition from 35% levels to 25% levels would at the least result in savings of Rs 30 crores. Obviously then the scope of retention modeling selects itself! Prediction exercises should be carefully selected for business impact.

2. **Develop the Hypothesis**

Companies have different approaches when employees leave. A few have extensive models for capturing the information on the exiting employee while it has become a ritual in many places. Having decided to go forward with modeling retention, it is important for us to select key hypothesis. It is not sufficient to stick to generalizations like

- Employees are leaving in search of a better package after the increments.
- Employees leave managers not companies
- Employees do not leave for compensation, they leave for growth.

It is not unusual for HR research studies that will offer one or more of the above insights. And these have more than a whiff of truth. However, unless your company has participated in the study, the extrapolation cannot happen as a rule. Yes, there would be bad managers and some would leave for growth. However we need to capture all our assumptions by doing a workshop with business and HR leaders.

A connected step is to see whether we have measures and data to examine the hypothesis. Our summary table could look like

Hypothesis	Key variables/metrics	Availability of data
• Specific demography of employees have higher churn rate	• Gender • Age • Level of education	• Yes
• Market specific factors affect attrition rate	• Location • Historic location performance	• Yes
• Executives with low or declining sales performance are expected to exit faster	• Sales performance of the executive • Decline in sales performance	• Yes
• Manager's attitude and influence is a key factor contributing to attrition		• No

Now not only know what our hypotheses are but also the practicability of each one of them based on dataset available.

3. **Exploratory Analysis**

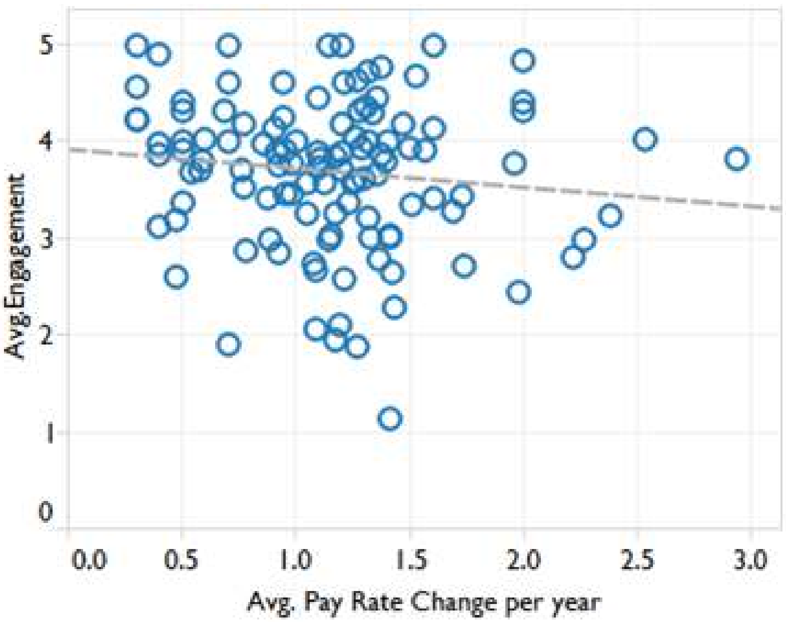
We need to examine each hypothesis to not only validate them but also see whether any patterns have been missed out.

For example let us say one of our hypotheses is that the number of pay rate changes an employee has impacts their engagement.

Then a plot like the following would help us explore the impact of pay rate changes on employee engagement.

This is a scatter plot for a group of employees. For employees in front line the pay levels often change due to different commissions being paid based on the performance. In this visualization we compare the frequency of such changes against their employment levels.

The analysis does support our original assumption that changes in pay-rates do have an impact on the engagement levels. While the engagement stays high as long as the average is around 1.5, with increasing changes the engagement levels actually drop. Our hypothesis could be that a high variation in pay rates during a year reduces engagement and leads to attrition.



By such explorations, we create a shorter set of hypothesis that is meaningful.

4. Completing the Data Set with Subjective Analysis where Feasible.

It would be very important for a predictive model to be able to map against the individual engagement level. On the other hand, our traditional surveys insist on confidentiality. We would know what demographic of employees have low engagement. However we have no way of knowing who exactly is unhappy.

Once we have completed the hypothesis short listing, our data sufficiency would look like this. Where there are gaps, perception data is used to supplement the existing dataset.

S.No	Factors	Historic Data	Survey Data
1	Tenure	Y	
2	Education Background	Y	
3	Source of Hiring	Y	
4	Gender	Y	
5	Marital Status	Y	
6	Location	Y	
7	Hometown	Y	
8	Supervisor	Y	
9	Individual Performance	Y	
10	Absenteeism	N	
11	Manager Infulance	N	Y
12	Motivation Levels	N	Y
13	Level of Collaboration	N	Y
14	Access to Resources	N	Y
15	Industry Factors	N	
16	Compensation w.r.t Competition	N	

This is where a technique like “Organization Network Analysis” becomes relevant. Organizations are not what their formal structure denotes. Let us look at a social network metaphor. In social networks, the stories can emanate from anywhere and gain likes. Influencers in social networks are not necessarily the biggest stars but ones with a very original point of view. It helps to be a celebrity.

Similarly, work and goals cascade down a formal organization structure. Increasingly, organizations resemble a network and the successful ones are far more cohesive than those which are not. The informal, uncharted organization wields as much influence or more, than the formal.

Mapping information flows helps us in recreating the informal organization. Correlating the cohesion of the informal organization with the percentage of in-house/ lateral talent would establish the case to groom from within or otherwise.

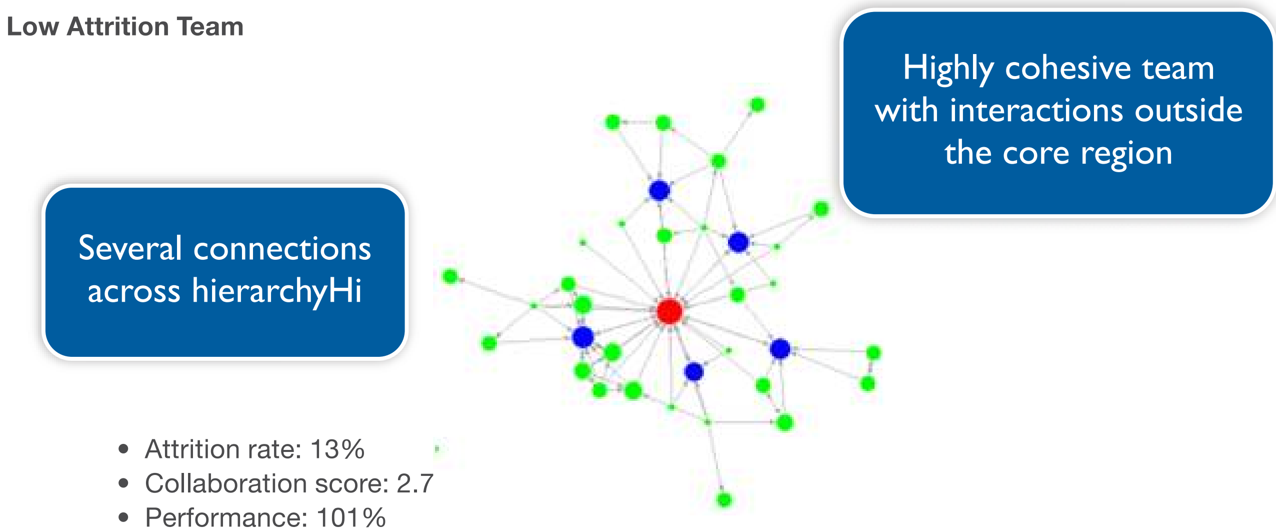
Mapping the informal organization will need us to actually run simple surveys and build visualizations based on the response.

The brief survey can have questions across areas like mentorship, innovation, career etc. In the survey, specific questions to employees to enlist, who in their peer group is

- Someone who they interact with for helping out with their daily issues
- Someone who they look upto as mentors
- Someone who is their role-model and whose inputs they value for career growth.

That someone can be their manager, but also a peer or a manager in a different group. These lead to us not only data points to supplement the model but also unique insights into the functioning of different teams and their strong and weak links.

As the illustration shows, there are big differences between the team with high attrition and with low attrition.

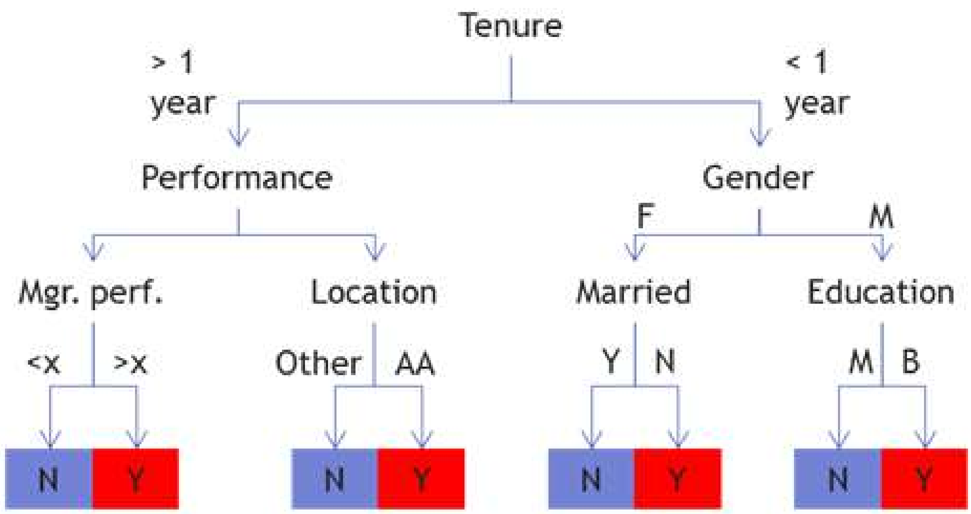


5. Building the Predictive Model

Now HR completes the work with business and passes it onto the analytics experts. There are enough tools and techniques that help with analyzing the data. For instance,

- Logistic regression
- Random forest
- Decision trees

Are all techniques that can be used to build predictive models. A decision tree could look like:



6. **Model Validation**

The validation of model is done using the truth table. Essentially an algorithm is generated using the existing variables. Then it is run on existing data to see its predictive abilities.

1. How many cases where the model predicts retention and the employee stays?
2. How many cases does the model predict attrition but the employee quits?
3. Where does the model predict retention but employee quits?
4. Where model predicts attrition but employee stays?

For the ideal model, 1+2 should be 100% and 3+4 should be 0%. While that is the ideal state, even a model that predicts 80% of the time is a huge improvement on what we currently have.

Predicted Class

Actual Class		No Attrition	Attrition	Class accuracy
	No Attrition	562	57	90%
	Attrition	78	117	60%
	Overall accuracy: 83%			

7. **Fine Tuning the Model**

The model is a dynamic entity. While one could run a version of the finalized model, continuous learning can take place by adding more variables and seeing whether the accuracy improves. This lies at the heart of artificial intelligence.

8. **Working with the Outcomes**

It is tempting to use the model as a novelty. However, the real benefit comes from actually going back to the assumptions. We estimated that reducing attrition from 35% to 25% will save us 30 crores at least. So, a risk profile is created for each employee and the nature of risk is flagged. (Manager relationship, comparative salary, last relocation etc.) Then necessary actions are taken to minimize attrition.

For example, if we have been able to create a risk profile for each employee. We need to take this forward from the risk. The objective is to address the high risk employees and take actions within the framework. Simplistically, it could look like

EMPLOYEE NO	RISK PERCENTAGE	WHAT COULD TRIGGER ATTRITION	ACTION
2001	35%	Safe in the near future	Continue communicating
3001	60%	Next project allocation in the same technology	Have a 1-0-1 to identify what she wants after the present assignment. Get her inputs and try to arrive at a joint plan
4001	90%	Is disengaged. Work quality is diminishing	Have a conversation. Start making back-up plans in case the person leaves.