Case Study: Employee Retention & Loyalty Benefits

Pushkar Chougule

Overview

Just like the quote from *Vincent Van Gogh "Great things are done by a series of small things brought together*" – all the Great organizations are what they are with all the people – the human talent they possess and nurture. So retaining employees is core to protecting any organization. One of the five ideological tenets set forth by a founder of HP is "We achieve our common objective through teamwork". Employees take on complementary roles and bring complementary skills to the table, learning to work and grow together in the process. So, when good employees quit the company, it causes multiple pain points – Loss of Talent and Knowledge, Time and Cost of identifying equally good candidates, training them and getting them to be as much or more productive than the predecessors.

As a part of this case study, we would be looking to utilize the power of predictive analytics techniques in identifying the employees who are most likely to leave the company, which can be very helpful in order to devise retention efforts plans for them. We have heard the idiom – "a stich, in time, saves nine", holds so true here. It's like identifying the leaks in the a boat's hull in order to patch them up and keep the sheep afloat. HP, though came in as the 27^{th} largest employer of 2011, with revenues to the tune of $127 billions, is still not worry free in terms of employee turnover – some of the workgroups reported overs in the range of 20%, which is very high and concerning.

As a part of the employee retention efforts, the dynamic duo of Gitali Halder and Anindya Dey teamed up to come up with predictive modeling system to identify all employees who have higher chance of quitting their jobs – phenomenon also known as employees with higher “Flight Risk”. The employee flight risk prediction system promises a $300 million in estimated potential savings with respect to new replacement employee hires (by focus on retaining them) and also by preventing loss of work productivity globally.

Business Understanding

Some of the work departments in HP face higher attrition rates to the tune of 20%, which can cause disruptions in the functioning of those departments and eventually the smoother functioning of the company. So, as a part of the employee retention program, Halder and Dey combined to develop a predictive system to identify the employees with higher possibility of quitting the company. Then the respective managers take these predictions into account, at their own discretion, about the corrective measures to be taken. This process is built more as a decision support mechanism rather than an automated system for decision making. So, human actions aspects for the final decision making are still in play.

HP tags more than 330,000 of its employees with a so called Flight Risk score. This is a simple number that assists in knowing the likelihood whether an employee will resign from his or her current job. The predictive analytics empowers the organizations by providing potential insights, although somewhat sensitive, in some cases. These predictions are all derived from the readily available data. With such powerful predictions, HP can focus its efforts on retaining the employees with high flight risk, thereby avoiding the overhead of hiring and training new employees.

Data Understanding and Preparation

The dataset consisted of around 330,000 employees with HP at the time of the employee retention / flight risk prediction initiative. Halder and Dey pulled together two years of employee data. Some of the important features to consider in the data understanding and preparation are the employee salaries, raises, job ratings and job rotations. Then they added on the most important feature – whether the employee had quit or not. Since most of the data was up to date, very little effort was needed in terms of cleanup and formatting. This data served as training set for employee Flight Risk detection system to identify and understand various factors contributing to this problem. As we know “Garbage In, Garbage Out” holds so true for Machine Learning / Predictive Analytics projects - Data quality check was performed to ensure that all the relevant fields were adhering to the standard guidelines like job ratings encoding to indicate higher to lower rating values appropriately, employee salaries are numeric with two decimal place, promotion and raise information is recorded correctly etc. Employee departure information should be appropriately encoded to indicate the correct cases of yes / no type of answers. Other factors like Job role, number of years of experience, gender, time in the current role etc. were also collected and used. For statistical data modeling analysis, conversion / encoding of Categorical and text data to appropriate numeric data is essential and hence it was performed to get numeric values.

Modeling / Evaluation

The data will first be split into training and testing data set. The training set will be used to train / build the decision tree model. Training of the data was performed through decision tree analysis using Chi-Squared Automatic Interaction Detector (CHAID) model. The train set produces large sets of decision trees with randomly selected number of variables for model building / analysis. It is important to have large number of decision making variables to be selected since these models typically use square root of the number of variables selected as number of random variables to be used in the decision trees created. We will use 80% of the dataset as training set and the remainder 20% set for testing purposes so as to ensure the selected metrics, precision and accuracy are well within expected ranges. We can evaluate how many of the predicted outcomes result in accurate classification as employees quitting the company vs. continuing with the company for longer duration. We can calculate the positive and negative prediction results ratios for correct (true) and incorrect (false) result counts and ensure the accuracy and precision level of the model is at least above 70%, since this is the first such predictive analytics endeavor of this kind for the company. We can use the k-fold (e.g. 10-fold, 11-fold etc.) cross validation technique to try to fit the model. Once the model fitting is performed, we can evaluate the model against the test data set and validate whether the model predicted the employee flight risk in terms of predicted outcomes vs. actual cases. If the results are within the expected accuracy levels, then we will continue to use the model or train / evaluate / fine-tune the model iteratively until the predicted results meet the expectations of accuracy over 70% levels. The criteria for efficient model building is with help of evaluation metrics like precision, accuracy, lift, recall etc. If criteria were met on these various parameters, model can be deployed to production environment and monitored for performance and continued improvement opportunities.

Deployment

Predictions of the Employee flights risks does come with its own challenges. Identifying which employees will leave and which ones will continue to stay on correctly is a big challenge in itself and depends on variety of considerations like current compensation, ratings / promotions, salary increments, current job role and time spent, the overall experience, and outside job market conditions. What's the accuracy level of the Flight risk score – what if we fail to calculate this correctly and tag the employees incorrectly as "at risk"? What are the consequences if the employees come to know about this? Will the awareness of such mechanism make employees to quit at higher rate, being concerned about their privacy? This is the reason why carefully deploying the predictive analytics CHAID model is so important in the employee flight risk process. It helps identify which are highly likely to quit, so that company can prepare in advance to extend additional offers / perks to retain the employees. This will greatly benefit the company by avoiding the overhead cost and time in hiring of newer staff members, training them and also helps maintain higher productivity levels.

Conclusion

Flight Risk model identified $300 million worth of estimated savings considering the factors like staff replacements and productivity loss across all global regions of HP operations. Also these model correctly identified 40 percent of HP employees with highest Flight Risk scores included 75 percent of the people actually leaving the company (lift of 1.9). The system highlighted just the higher number of promotions could not sufficiently prevent the attrition unless it was accompanied with higher salary raises. It also underscored that periodic changes in the nature of job duties and giving the interesting work also kept employees engaged and thus with the company for longer. Flight Risk reports helped shape management decisions in a productive direction. It empowered the company and management with knowing the right factors which could lead to attrition thereby allowing an opportunity to take corrective measures in right direction. Ultimately, this helps in developing more robust strategies to retain the staff which results in reduction in the costs of staff replacements / training and maintain the higher level of productivity along with better business continuity.

References

1) Abbott, D. (2014). *Applied Predictive Analytics – Principles and Techniques for the Professional Data Analyst.* John Wiley & Sons.
2) Siegel, E. (2016). *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die.* John Wiley & Sons.
3) Siegel, Eric (2013). Predictive Analytics: The privacy pickle (Hewlett-Packard's prediction of employee behavior) https://pubsonline.informs.org/do/10.1287/LYTX.2013.06.02/full/
4) Nair, H. (2019). *A predictive analytics model to determine the Flight Risk Score of employees* https://www.linkedin.com/pulse/predictive-analytics-model-determine-flight-risk-score-dr-hari-nair