Killian McKee

3/10/2019

Reducing Employee Attrition: Implementing AI Retention Systems

**Overview**

Employee turnover presents an increasingly pressing and costly problem for employers. A strong job market, changing workforce, and technological advances are all compounding talent retention difficulties. According to a 2017 *Fortune* [study](http://fortune.com/2016/12/28/employers-2017-employee-retention-unemployment/), 87% of employers rated improving retention as a critical priority. This is unsurprising given the high cost of replacing staff-[estimates](https://cnmsocal.org/featured/true-cost-of-employee-turnover/) range from about half of the annual salary for those with medium skillsets to more than double for highly skilled employees. These numbers are often significantly larger in highly technical or cleared settings due to protracted hiring and onboarding times. Fortunately for employers, many are unknowingly generating goldmines of employee data that can be cheaply, anonymously, and securely tapped with artificial intelligence (AI) to reduce employee churn and save significantly. This paper enumerates some of the prominent challenges to retain workers in both unclassified and classified settings, how predictive AI could offer retention improvements, benefits to addressing these obstacles, AI adoption challenges, and an example implementation using Python.

**Retention Challenges**

Becoming familiar with the major agents causing high churn rates provides a foundation upon which to build an AI solution. There are three overarching challenges facing most employers:

* **Strong Job Market:** The U.S. economy is strong. With unemployment hovering around [4%](https://ig.ft.com/sites/numbers/economies/us/) and many industries facing more new openings than hires, power is shifting from the hands of employers to employees.
* **Shifting Workforce:** Millennials are now the [largest cohort](http://www.pewresearch.org/fact-tank/2018/04/11/millennials-largest-generation-us-labor-force/) in the U.S. labor market (35%) and Post-Millennials are rapidly on the rise. Both groups are young, willing to uproot, harder to engage at work, and tend to be less loyal to single organizations than older employees. A recent Gallup [study](http://www.gallup.com/reports/189830/millennials-work-live.aspx?utm_source=gbj&utm_medium=copy&utm_campaign=20160512-gbj) estimates Millennial turnover costs the U.S. about $30.5 billion annually.
* **Technological Advances:** Employees have never had more access to jobs than now. Online job boards like Indeed and Glassdoor make it easy for anyone to quickly find and apply to a job. Similarly, sites like LinkedIn make it discreet and inexpensive for recruiters and employees to engage one another.

Additionally, there is a subset of problems specifically pertinent to highly technical/cleared staff:

* **Security Clearance Process:** The security clearance process consumes time and money. This represents an increased cost to onboarding employees and makes them more difficult to replace.
* **Cleared/Skilled Employee Shortage:** There is a [severe backlog](https://news.clearancejobs.com/2018/12/01/finding-a-workaround-to-the-defense-talent-shortage/) of highly technical/cleared employees (about 700,000 jobs) due to the length of the clearance process and increasing demand for these attributes. Cleared employees tend to be recruited more aggressively by competing firms (e.g. Amazon) and command more leverage over their employers.

**A Partial Solution: AI Attrition Prediction**

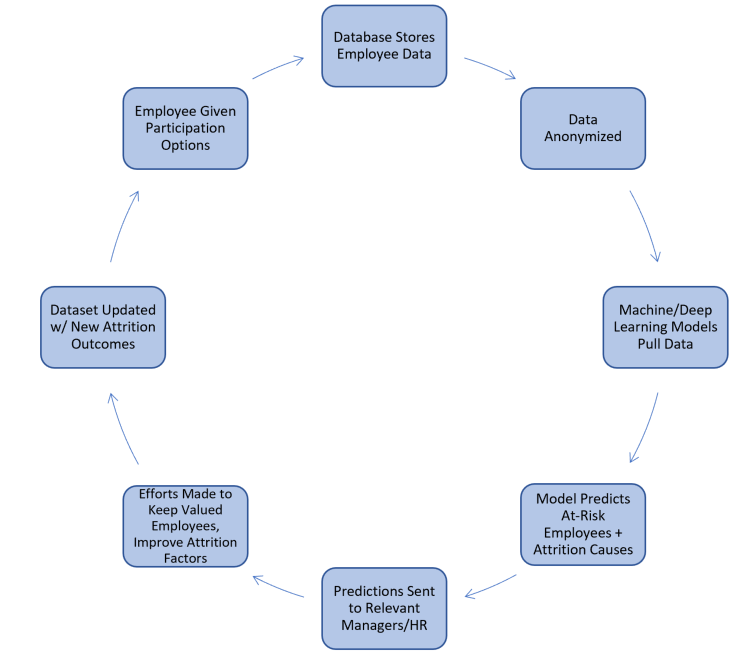
Routinely collected company data can be anonymized and securely utilized to predict employee attrition down to an individual level. More specifically, information like job title, salary, years with the company, years since promotion, boss, etc. can be used as features in a dataset where attrition is treated as a classification problem. In a classification setting, all past and current employees are treated as either, ‘someone who left the company’ or, ‘someone who didn’t leave the company’. Based on this past data, one can predict whether current employees (even when data is anonymized) are at risk of leaving given their similarity to the profiles of other employees and what factors could be driving their exodus. These similarities can be determined by machine learning and deep learning algorithms with a high degree of accuracy. For example, the walkthrough provided in this paper’s appendix achieves an accuracy of 89.4% at predicting if an employee has left on a small, imperfect dataset. Results would likely improve in a full scale, production setting. Decision leaders and human resources staff can use AI generated attrition predictions to reach out and work with employees to address issues, reduce costs, and improve employee satisfaction before new attrition occurs.

Figure 1: An example workflow for an AI Retention System

**Benefits of Addressing Retention Shortcomings with AI**

The benefits to addressing high employee churn rates via AI are straightforward and significant:

* **Cost Savings:** The cost of losing an employee is composed of numerous factors like hiring, onboarding, lost productivity, training, etc. Even modest reductions in attrition rates can lead to dramatic savings due to the high cost of replacing staff.
* **Ease of Implementation:** Utilizing algorithms like those provided in the walkthrough can be done easily and inexpensively. All necessary software is available open source and the data needed to generate predictions is routinely collected by human resources departments.
* **Happier Employees:** Finding and addressing potential sources of attrition can produce happier, more engaged employees
* **Identify Attrition Sources:** using AI tools like principal component analysis can elucidate what consistently drives employees to leave.For example, using an AI algorithm one might find ‘pay’ and ‘work life balance’ are the two most important predictors to determine if an employee is going to leave their company.
* **Consistent Customer Contact:** Reducing attrition helps preserve strongly performing teams and the relationships they’ve built with external contacts.

**Obstacles to Adopting Artificially Intelligent Practices**

Understanding AI challenges prepares stakeholders to assess where it belongs in their organization. Here are some of the common concerns facing artificially intelligent attrition tools and how they can be assuaged:

* **Fears About Losing Employee Privacy:** Most companies could perform an AI based attrition analysis using the information they are already collecting about their employees with information like the employee’s manager, current pay, tenure, etc.
* **Distrust of AI:** A lack of understanding about how AI works often breeds distrust. Making it clear that these practices are used to better understand employees and meet their needs can help with adoption.
* **Resistance to Change:** Employees are often somewhat neophobic because they’ve become accustomed to working in a certain way. Clearly defining the mutual benefits of AI (like an increased mutual understanding between management and employees or faster problem solving) can help ease adoption.
* **Overestimating Implementation Difficulty:** Many companies are unaware attrition prediction systems don’t require excessive time, money, or expertise. Thanks to open source tools like Github, sci-kit learn, and Tensorflow, a small team familiar with Python or R could implement a functioning version of an attrition predictor in short order (depending on availability of data, desired complexity of models, and familiarity with data science).
* **A Lack of Data:** Large sets of data (hundreds or thousands of employees) are needed to generate accurate predictions. Upsampling (generating artificial data based on the data one already possesses) can help address this problem, although it is inferior to real data.
* **Poor Data Governance:** Even if companies have enough data, it is often mismanaged. Consolidating disparate data sources into larger databases using tools like SQL and Python can help improve access to high quality data.

**Conclusion**

External factors are driving employees to change jobs more frequently. This transient workforce presses employers to spend ballooning amounts on hiring, onboarding, training, and retention, especially with technical or cleared staff.

Despite strong market forces, employers can use machine learning and AI to leverage the data they already collect to help predict and prevent attrition. The code walkthrough in the appendix section highlights the powerful potential of using artificial intelligence in human resources settings. Ultimately, more organizations should consider artificially intelligent retention systems because they are cheap to implement, offer significant cost savings, and can create happier, more engaged employees by revealing recurring churn themes.

**Appendix: An AI Implementation Example**

This section is intended for a more technical audience interested in building a retention model. The code steps through several machine learning algorithms and a deep neural net to elucidate the benefits of using AI in an employee retention setting. Using past data on employees, we can treat retention as a classification problem i.e. is an employee at risk of leaving or not? Managers can then use these predictions to preserve their talent accordingly.

This walkthrough uses IBM’s attrition [dataset.](https://www.ibm.com/communities/analytics/watson-analytics-blog/hr-employee-attrition/) It steps through 5 powerful, widely used classification algorithms: extreme gradient boosting, logistic regression, a random forest, support vector machines, and a basic deep neural net. To keep this section relatively compact, a pre-encoded version of the data is used for the first 4 algorithms, but the full data preparation pipeline is included in the neural net section. Ultimately, a peak accuracy of 89.4% is achieved using logistic regression, which is acceptable considering the small dataset size and the serious class imbalance (83% of the data falls into the ’did not commit attrition’ class). In a large-scale implementation, more data would likely be available to improve results. Additionally, there is room for improvement in hyperparameter tuning (using methods like GridSearchCV) and training time. The purpose of this guide is not to maximize accuracy, but to demonstrate a general workflow and the straightforwardness of using machine learning to derive improved attrition insights from an imperfect dataset.

**All Model Accuracies**

1. Logistic Regression: 89.4%
2. Extreme Gradient Boosting: 88.5%
3. Support Vector Machine: 88.0%
4. Deep Neural Net: 87.4%
5. Random Forest: 87.2%

**Complete Python Code**

