Turner Sale

Initial Project Proposal

02/16/20

At my current place of work, our HR department has indicated that our turnover rates are higher than desired, especially in certain departments. This of course has a knock-on effect to other aspects of the business, such as training expenses, on-time work completion, and potentially on-time shipments to customers. As such, it would be useful to predict who are the employee’s at highest risk of turnover, and if possible, when they may be likely to leave.

Thankfully, we have several data sets at our disposal that could be of use. The first is of course employee data from our HRM (Human Resource Management system). This data would include demographic data, salary information, location data, department and supervisor data, etc. This high-level data itself may hold many insights (for example, do employees that live far from work have a higher turnover rate?), but I think that adding additional sources could improve on this.

The first set that I would combine would be time recording entries. Raw entries would be nearly useless, but perhaps transforming them to be focused on a previous period could assist. The time recording entries are the records created to indicate employee time spent on certain ‘itemclasses’ (our term for work types). This could include things like training, project management, engineering work, etc. Focusing on the past month or year and summarizing the time recording entries per employee might give additional relationships (e.g. a worker that spent a lot of time in administrative tasks might have higher turnover due to burnout). Additionally, it might be possible to find certain itemclasses that have higher turnover rates themselves (things like boredom, repetitiveness, or simplicity might be at play).

Another set that will have to be developed will be comparative pay scales by job type. This may be difficult to gather for a local level, but there are sources for national level pay scales, and comparing current salary to a national level may show an interesting correlation, or at the very least offer a solution to keeping an employee (if you want them, maybe paying them more will keep them).

It would also be possible to include the change records for each employee (such as change in position, change in salary, etc.) that might provide a better timeline aspect to the data set. In effect, adding a few features that describe whether the employee is seeing advancement, if their pay is changing, or similar alterations might help to better round out the ‘situation’ they find themselves in. An example could be an employee that has not had a position change in 5 years might not be as likely to stay as one who has seen advancement (or at least change) in that timespan. Or perhaps the opposite is true, and those with long tenure in a position are unlikely to leave.

This time sensitive concept would also be useful for applying a predicted turnover time. If an employee is likely to leave, it would be useful to have an idea as to when, particularly in the factory where turnover is highest. If we can predict that employees will leave in a certain time period, it may give HR some foresight into the impacts on our business.

Another addition would be looking at the effects of turnover on operations, which would not directly predict the turnover or assist in this effort but would show us where we may have impacts. This ties into the concept of predicting the time at which an at-rick employee may leave. If we see that turnover directly impacts on-time shipment, and we predict to have a high turnover in June, then perhaps we find ways to retain those people or hire new ones before we reach that time. I have a suspicion that although patterns will certainly appear, it will be difficult to predict when an employee will turnover, so I will leave this as a secondary task if the prediction goes well. But in either case, showing the impacts of turnover gives an immediately actionable set of information. The example of turnover effecting on-time shipments is a prime example. If we know this to be the case, then we could prepare some reporting to notify HR that we may see an issue if we predict high turnover. For example, if we reach June and predict high turnover for the summer, and we know that it will impact on-time shipments, we can set up notifications to consider hiring new workers, training other employees, or providing incentives during that period, or even working higher backwards and altering production schedules to include more lead time. This is of course a lofty goal but knowing how quickly production can turn south might motivate managers.

In order of importance, the following goals will be the target of this analysis:

1. Predict employees with a risk of turnover (binary)
2. Predict the likelihood of turnover (probability)
3. Predict when the turnover will occur
4. Find relationships between turnover and operational aspects
   1. Turnover to training costs
   2. Turnover to on-time shipments
   3. Turnover to hiring costs (if possible to find General Ledger Entries)
5. Offer suggestions as to what the largest factors in turnover are so they can be addressed by management
   1. Things like salary, location, training, etc.
6. Predict operation challenges that may arise
   1. If possible to predict turnover date