Aalto-yliopisto
Aalto-universitetet
Aalto University

15.3.2023

Employee Attrition at AmazingIT

TU-E5030 – Creating Value with Analytics D

Tuomo Salo

Executive summary

Highest turnover in AmazingIT is with junior and next tier designers and software developers. The turnover might be related to worse benefits and career building motivation of younger professionals. Disproportionate workload compared to their superiors is not the case here. Logistic regression model says that most influential variables are personal performance rating and evaluation of their boss, overtime amount and travel days. The predicting accuracy of this model was underwhelming and required manipulating training data. Finally, the model could predict the number of leavers somewhat accurately but further testing and adjusting is required.

1. Why employees leave the company?

There are multiple reasons why employees leave them company and most turnover happens in low tier designers and software developers. Figure 1 illustrates this observation.

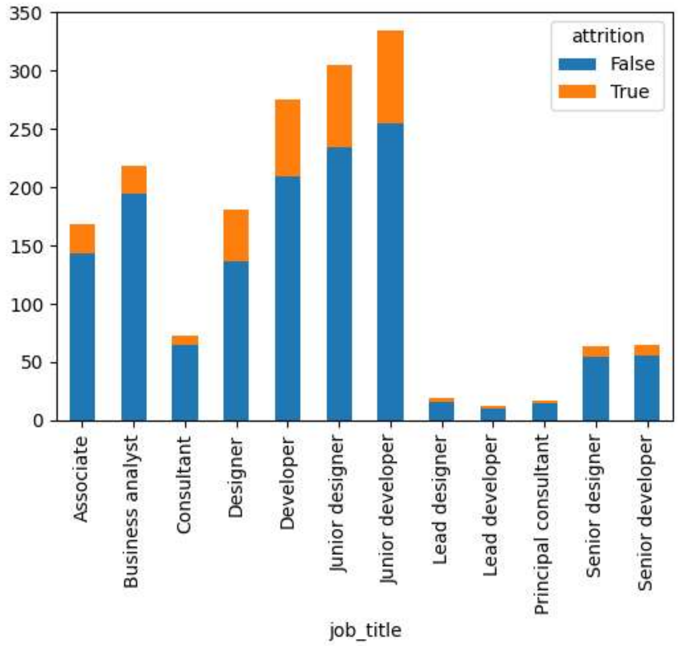


Figure 1. Attrition amounts by job title.

After constructing a logistic regression model from the employee training data the most probable causes for attribution were low employee performance rating and low bosses performance rating. Next most significant variable was the amount of travel days per year. Other variables, such as the education level and gender had impact on the logistic regression, but their p values were over 0.05 so they were removed to reduce overfitting of the model.

The company has the most people in the lowest hierarchy category, and least in the highest category with the rest in between. Performance rating distribution does not seem to discriminate lower tier workers, and rather high tier employees tend to be rated more critically (Appendix 1.1).

Travel days distribute evenly when scaled and compared with all hierarchy levels which indicates that younger workers are not disproportionately pressured to make travel trips (Appendix 1.2).

To summarize, mostly lowest and second-to-lowest tier designer and developers are the categories with highest turnover. It can be related to career building where switching jobs often nets more salary and better benefits. AmazingIT could be worse than its competitors to its employees which cannot be evaluated from provided data. It is not related to disproportionately high work load compared to higher tier employees inside the company.

2. Identifying corrective actions

Easiest and most effective way to reduce attrition rate is to reduce overtime hours worked, particularly with junior designers and developers, and reduce their travel days per year.

The constructed model identifies several easy options to reduce attrition rate. Table 1 depicts the most important variables in the logistic regression model and useful column to determine the efficiency of changing the variables. The last column tells how much and in what direction the probability of employee resigning this year moves, when the variable is increased by one unit. For example, one hour of overtime per week more increases the chance of resigning by 11.3 %. And on the contrary, increasing the training spending by 1000 € would lower the chance of leaving by 7.18 %.

Table 1. Most meaningful variables, their coefficients and chance to resign per unit change.

Variable	Coefficient	Chance to resign per unit increase
training_spend	-0.000718	-0.000718
performance_rating	-0.401	-0.330039
boss_rating_avg	-0.339	-0.287
overtime_per_week	0.107	0.113
traveldays_per_year	0.0262	0.0266
lastpromotion_months	0.052	0.0536

The chances would not linearly affect the probabilities since this is an exponential model, and infinitely increasing training spending would not prevent an employee from ever leaving the company. However, the model is a good starting point to estimate the efficiency of different measures.

Decreasing travel days per year is an efficient way to reduce the attrition rate and should be aimed for. Unnecessary or excessive travel may wear off the employee and reduce happiness of the employee. Even small decreases would help the attrition rate. Average and median travel days for an employee in AmazingIT is around 20 days, and eliminating all of these would decrease the chance of resigning by 80 % [1].

A more pronounced effect would be gained by reducing the overtime hours per week. Median overtime hours are 4 per week and eliminating all of them would decrease the chance to resign by 75%. However, eliminating overtime should be targeted to the specific job titles that have high resign rates and are shown in figure 1. Taking action to reduce overtime hours of junior developers, developers and junior designers should be the main priority as they are most likely to change company.

Finally, performance ratings of the employee and the quality of their boss play a key role in keeping people in the company. It is complicated to suggest fit-for-all solution to improve these metrics, but human resources should identify underperforming bosses and either train or move them. When employees like their boss, they are more likely to continue in the job. This is one metric which might contribute to the higher-than-average attribution rate.

[1] 1/(0.0266 * 20 days) = 0.188

3.

The training data was used to create logistic regression model, but the result was quite underwhelming. Figure 2 shows the confusion matrix of the predicted result compared to actual test data. The model predicted that almost all employees would stay in the company and only eight would leave. In reality, 66 employees left the company and the model did not predict that at all.

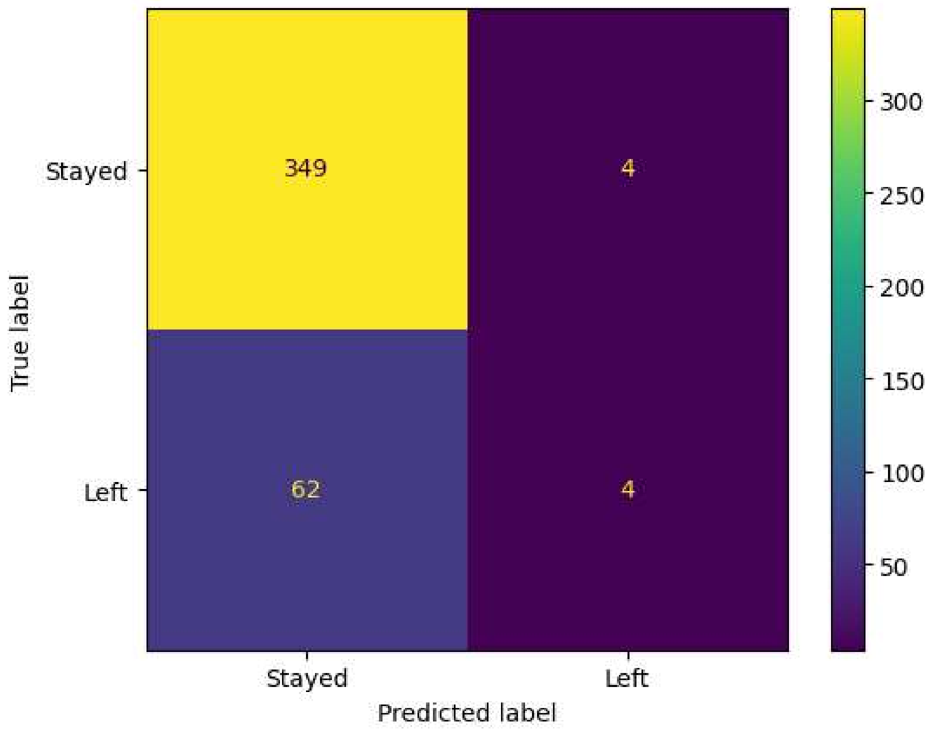


Figure 2. Confusion matrix of logistic regression prediction model.

Scaling the variables to be standard length or limiting the amount of variables did not meaningfully change the model predictions. Further iterating of the model should be conducted to make it predict that not everybody will stay.

The model has an accuracy score of 84 % which is misleading, since it is equal to the 84 % of people who stayed in the company.

However, I filtered 50 % of the negative attribution label -rows out of the training data to obtain less biased model that has better results. Notably, imbalanced data can lead regression models to always predict the to be safe, and biasing the data can make the model more accurate (He & Garcia, 2009). Figure 3 depicts the improved models confusion matrix.

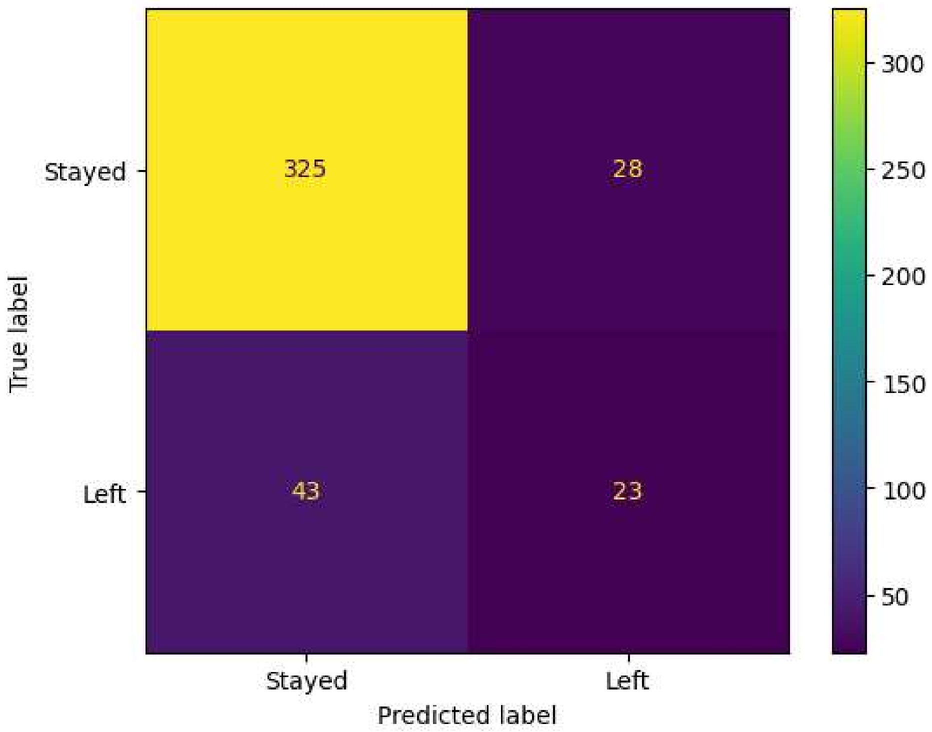


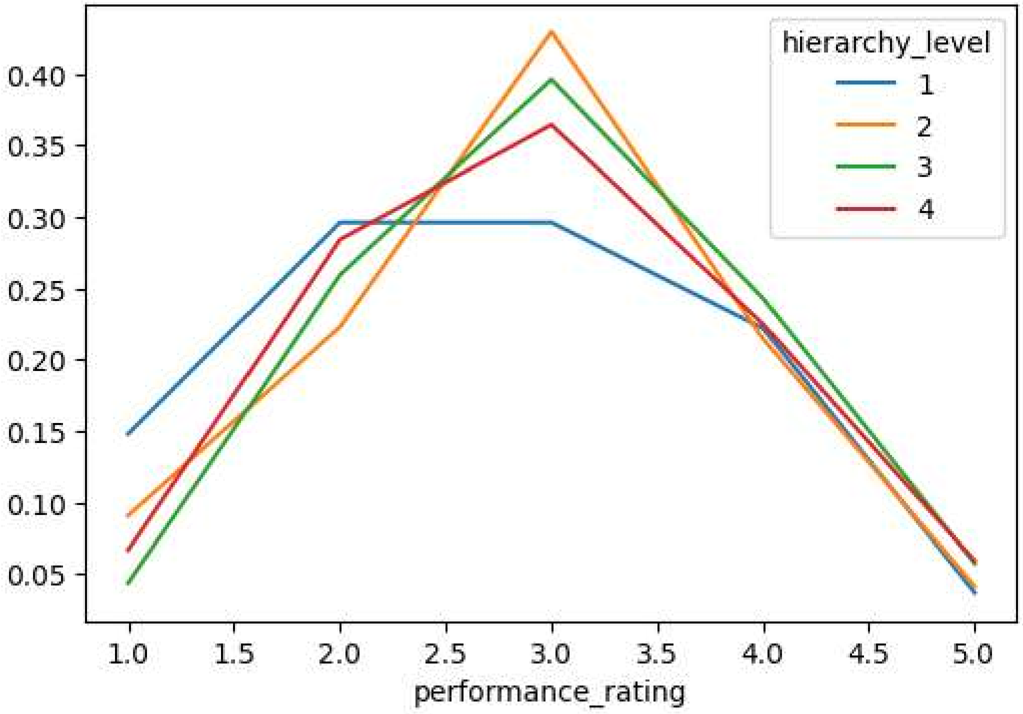
Figure 3. Confusion matrix of improved model trained with less biased data.

This result is better but not perfect. This time the model predicts the number of quitters to be 51 which is somewhat near to the real number. The model did not correctly specify which of them were quitting and was right of the specific employees under 50 % of the time. Further adjusting is still needed and the variables could be given different weights to mimic reality more precisely.

References

He, H. and Garcia, E.A., 2009. Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9), pp.1263-1284.

Appendix 1.1. Performance rating by hierarchy level, scaled.



Appendix 1.2. Travel days by hierarchy level, scaled.