

HR Employee Retention Survey: Data Analysis | ML Model Results

Prepared for: Salifort Motors Leadership Team

ISSUE / PROBLEM

Salifort Motors leadership are looking to address an unusually high turnover rate as of late. As a first step in the process, they have tasked the HR department with surveying employees to collect data on their satisfaction levels.

The goal of this project is to use HR data to train a machine learning model to predict whether an employee will leave the company, while garnering insights into what is driving turnover.

This report offers key insights from the development of the machine learning model, in order to help the company increase retention and job satisfaction for current employees, and save money and time training new employees.

IMPACT

Recommendations:

Establish initiatives to align employees' workload with those who are less likely to leave.

- **Employees leave less frequently when working an average of 165-220 hours per month.**
- **Employees leave less frequently when contributing to 3-5 projects.**

Conduct exit interviews to gain insights into why some employees left dissatisfied despite high performance review scores, and identify ways to improve support for those with low scores.

Consider capping monthly hours worked at 275. Employee satisfaction tends to be the lowest by far among those who work more than 275 hours monthly on average.

Focus retention efforts on employees who are approaching 3 years tenure. Most employees who left had a tenure of 3 to 6 years.

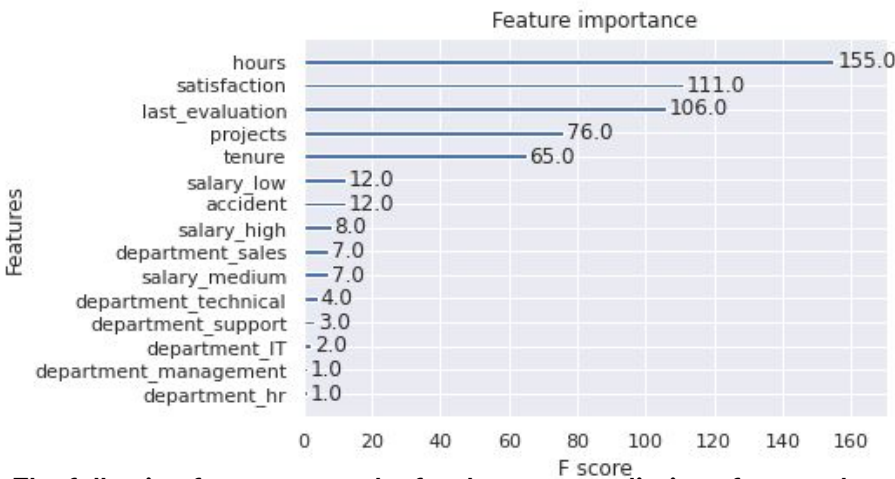
RESPONSE

The machine learning model chosen for this task was XGBoost, as it is less sensitive to class imbalance and unscaled data, allowing the use of all predictive features available with little preprocessing. The data was split into a training and testing set, allowing the fit of the model to be evaluated so that it may be tuned to fit the data as well as possible. **Utilizing cross-validation with GridSearchCV, the hyperparameters of the model were tuned to prioritize the recall evaluation metric.**

The final model performs excellently at predicting whether an employee will leave the company, but will need to be tested with new data and re-evaluated to ensure that the performance continues to be strong. Feature engineering, the addition of more granular data, and further hyperparameter tuning could all improve performance.

The final model yielded the following evaluation scores upon testing:
Accuracy: 0.985 | Precision: 0.985 | Recall: 0.924 | F1: 0.953

KEY INSIGHTS



The following features were by far the most predictive of an employee leaving the company, in order of importance:

Hours: (Average number of hours employee worked per month)

Employees who left very rarely worked in the range of 165-220 average hours per month.

Satisfaction: (The employee's self-reported satisfaction level [0-1])

Employees with a reported satisfaction level below 0.2 or around 0.4 tend to leave most frequently. Satisfaction tends to be highest among those who work 130-260 hours per month on average.

Last evaluation: (Score of employee's last performance review [0-1])

Employees who left rarely had scores in the range of 0.6-0.75, with the distribution of scores outside that range being relatively uniform. Some employees who left had scores approaching 0.8 despite low satisfaction.

Projects: (Number of projects employee contributes to)

Those who left very rarely contributed to just 3 projects. Average satisfaction is consistently high for employees contributing to a range of 3-5 projects. Employees who left have a wider distribution of project contribution count.

Tenure: (How many years the employee has been with the company)

Nearly all employees who left had a tenure of 3-6 years. Most had 3 years tenure, with turnover rate decreasing as tenure increases.