Employee Turnover Project | ML Model Results

Prepared for: Salifort Motors Leadership Team

ISSUE / PROBLEM

Currently, there is a high rate of turnover among Salifort employees. Salifort's senior leadership team is concerned about how many employees are leaving the company. If Salifort could predict whether an employee will leave the company, and discover the reasons behind their departure, they could better understand the problem and develop a solution. The leadership team requested a model that predicts whether an employee will leave the company based on their department, number of projects, average monthly hours, and any other relevant data points. **This report offers details and key insights from Milestones 8 to 10, which could impact the future development of the project, should further work be undertaken.**

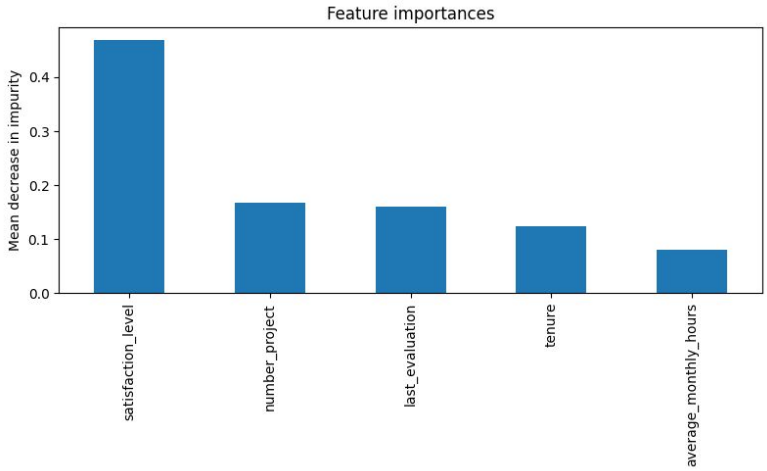
IMPACT

- ➔ The company should utilize the random forest model developed to predict if an employee will leave the company or not.
- ➔ To retain the employees, the company should focus on the following features, in the order of importance,
  - ◆ Each employee's satisfaction level (satisfaction_level),
  - ◆ The number of projects each employee contributed to (number_project),
  - ◆ Each employee's last performance review score (last_evaluation),
  - ◆ How many years each employee has been with the company (tenure), and
  - ◆ Each employee's average number of work hours per month (average_monthly_hours).

RESPONSE

- **To obtain a model with the highest predictive power, the Salifort Motors data team used four model architectures to create nine models to cross-compare results: Logistic Regression, Decision Tree, Random Forest, and XGBoost.**
- To prepare for this work, the data was split into training, validation, and test sets. Splitting the data three ways means that there is less data available to train the model than splitting just two ways. However, **performing model selection on a separate validation set enables testing of the champion model by itself on the test set, which gives a better estimate of future performance than splitting the data two ways and selecting a champion model by performance on the test data.**

KEY INSIGHTS



- From comparing the models' evaluation metric results, **the random forest model** performed the best on the data.
- For the test dataset:
  - The champion model correctly identified 92.7% of the employees who actually left the company, in other words, the model's type II error rate is 7.3%.
  - The champion model correctly identified 99.8% of the employees who actually stayed with the company, in other words, the model's type I error rate is 0.2%.
- **There are four features that are not important for the random forest model**, they are department, salary, work_accident, and promotion_last_5years, and are not included in the training of the champion model.
- The champion model is sensical because both type II and type I error rates are relatively small. Whether these error rates are acceptable depends on the costs of the measures taken to prevent an employee from leaving versus the value of retaining them.