

# Salifort Motors

## Employee Retention Project

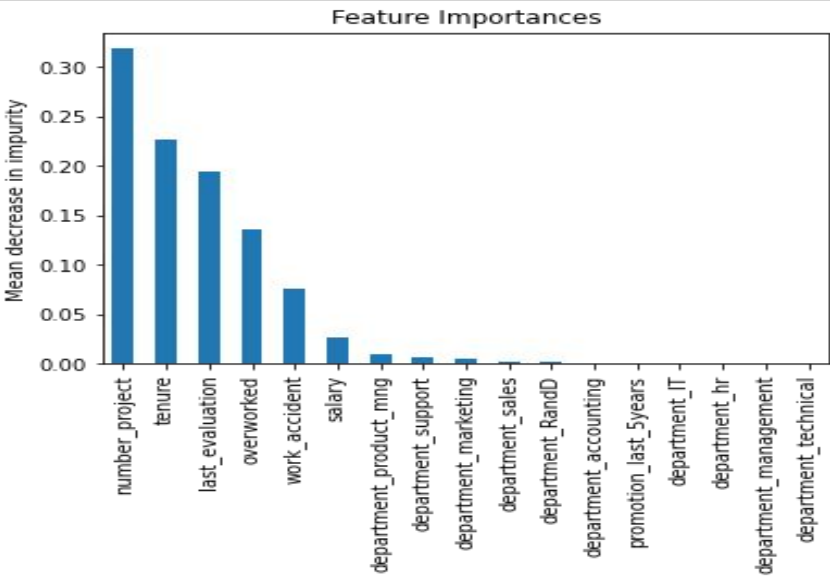
### Project Overview

Salifort Motors strives to create a corporate culture that supports employee success and professional development but the leadership team is concerned about the high rate of turnover among employees. 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.

### Key Insights

- 3-4 projects are the optimal amount that employees should work on.
- Consider promoting employees who have been with the company for at least 4 years, and investigate why 4-year tenured employees are so dissatisfied.
- Reward employees for working longer hours or don't require them to do so.
- Inform employees about the company's overtime pay policies and the expectations around workload and time off.
- High evaluation scores should not be reserved for employees who work 200+ hours per month. Consider a proportionate scale for rewarding employees who contribute more/put in more effort.
- Number of projects, tenure, evaluation scores and average monthly hours worked are the best predictors of employee retention.

### Details



In the XGBoost model above, `'number_project'`, `'tenure'`, `'last_evaluation'`, `'overworked'`, `'work_accident'`, and `'salary'`, and have the highest importance. These variables are most helpful in predicting the outcome variable, `'left'`.

### Next Steps

It may be justified to still have some concern about data leakage. It could be prudent to consider how predictions change when `'last_evaluation'` is removed from the data. It's possible that evaluations aren't performed very frequently, in which case it would be useful to be able to predict employee retention without this feature. It's also possible that the evaluation score determines whether an employee leaves or stays, in which case it could be useful to pivot and try to predict performance score. The same could be said for satisfaction score.

For another project, we could try building a K-means model on this data and analyzing the clusters because of the odd distribution for satisfaction level, evaluation scores, and average monthly hours worked.