

# Salifort Motors Turnover Analysis

## Data Analysis and Machine Learning Model

### Overview

Salifort Motors senior managerial staff has been faced with the notable issue of high turnover in the company. There is a lot of investment in the onboarding and training of new staff. They aim to have a professional and supportive corporate culture, which is built to promote growth and success. However, through employees leaving the company and being laid off, Salifort is taking sizeable damages. The questions are where are the issues that seem to be the root of this endemic problem being faced? And is there a way to predict when an employee is in the group moving towards leaving the company via turnover?

### Objective

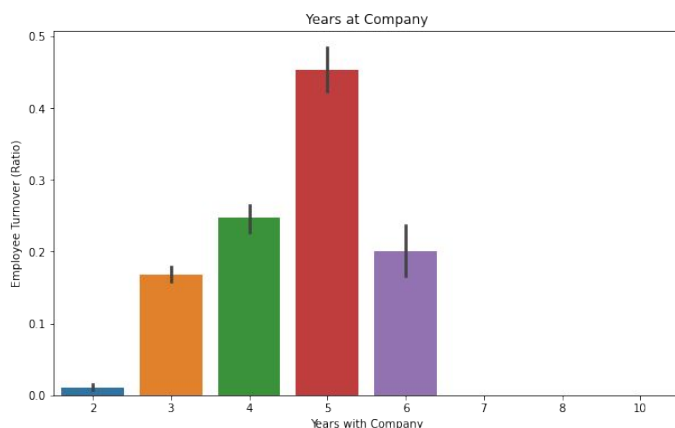
There is a technical goal and holistic goal. The technical is the review, clean, then model the employee data to predict which employees are likely to be 'turnover'. The holistic goal is to analyze the data and what features are most impacting the prediction of turnover and hopefully better understand where inside the company there are factors influencing employees in such a way that they either are laid off or choose to leave the company. Our deliverables are: recommendations and insights as to why turnover is high throughout Salifort Motors and a machine learning model that will predict what employees are likely to be 'turnover'.

### Results

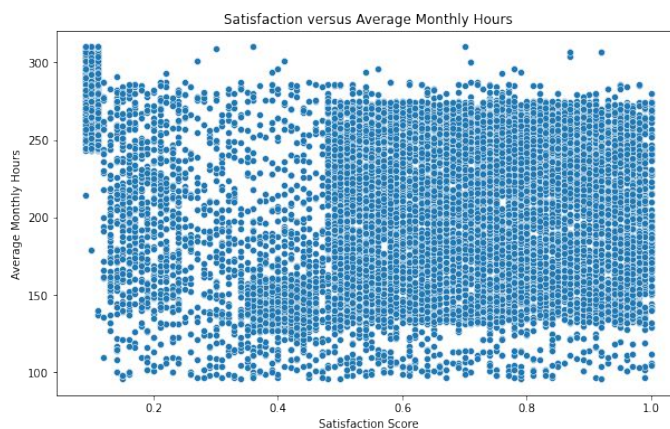
- Random Forest model on test dataset results: Accuracy and the weighted average of Recall, Precision, and F1-score are all 98%. While the results for 'left' specifically were Recall of 90%, Precision of 98%, and F1-score of 94%. Original test dataset results ended with a F1-score of 93%, but once the importance of the features was reviewed and the unimportant features were dropped, it increased by 1%. With these scores being produced by the model we can with reasonable confidence use the model to predict 'turnover' employees.
- Statistical review results: a two tailed t test was run on employees with high and low satisfaction. The null hypothesis was that the mean of the target variable 'left' of each group was the same. The p-value found was 3.136e-89%, which is essentially zero and less than the 5% significance level, and therefore, statistically significant. We reject the null hypothesis and must conclude that there is a difference between the mean of the target variable 'left' of each group.
- From analyzing the data and its statistics, there were a couple noticeable trends.
  - Highest turnover is employees in their fifth year with the company.
  - The average employee works 201 hours per month. 40 hours greater than a standard 4 week month (160 hours).
  - Satisfied and high scoring managerial reviewed employees are leaving the company.
  - Employees working at the contribution rate of 1-2 projects per year are most likely to be turnover.

### Next Steps

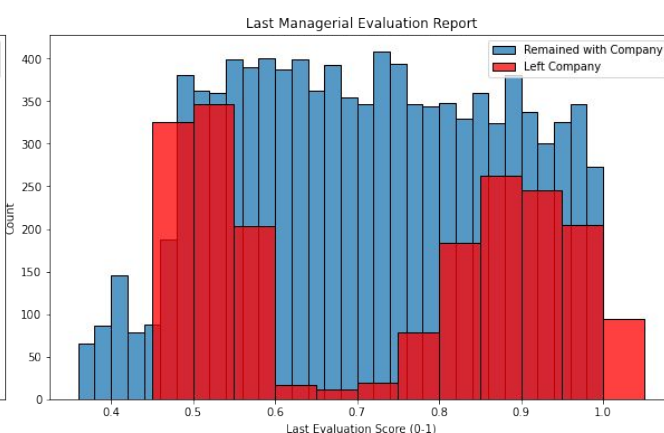
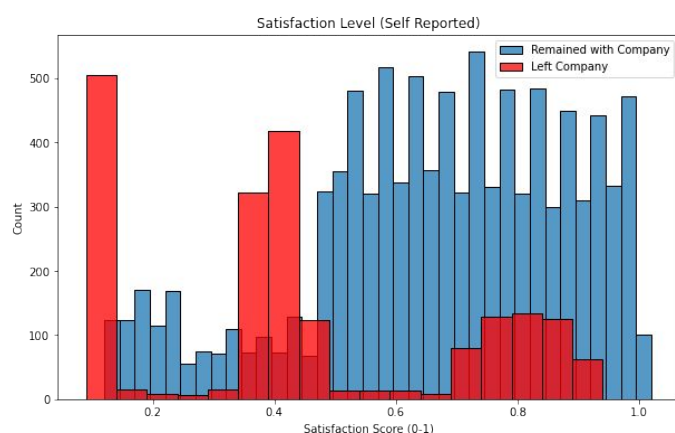
- I recommend implementing the model, so the management can address and better retain employees that may be heading towards turnover through dissatisfaction and overwork.
- The predictions of the model include a moderate amount of False Negatives. To improve the model we should build better employee review systems, project points for instance, for a better indicator of employee engagement and progress. Additionally, we can approach building the model differently, with new hyperparameter tuning and feature transformations and selection.
- Effective immediately, the management must implement better stewardship of employees. This may include: better hours and time management, improve long term incentives, and increase moral and team building exercises.



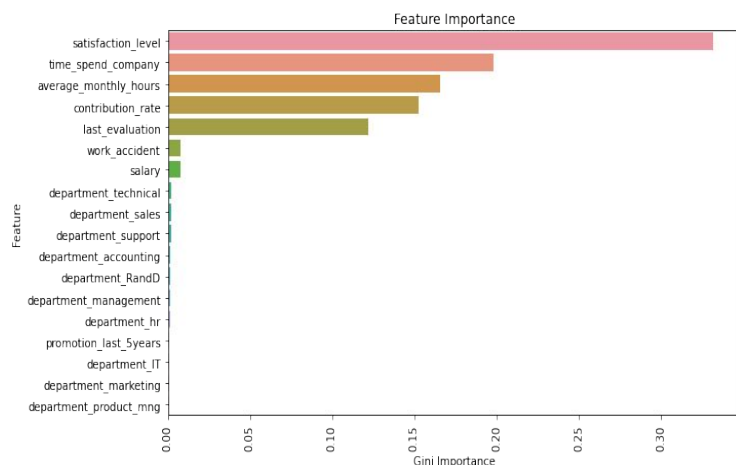
We observe a near exponential increase in turnover approaching 5 years of employment, and then a sudden and near complete slow down after year five.



There is a distinct grouping of employees with low satisfaction, which are working approximately 240-310 hours per month. It can be seen that more than half the company is working above the normal time for a work week.



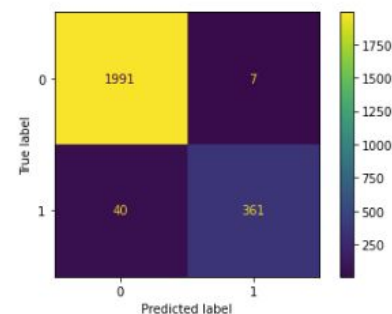
For both the satisfaction level and the evaluation report score, we observe a phenomena of turnover occurring on either end of the spectrum. There are clearly endemic issues of the company pushing good, loyal, satisfied employees out, alongside issues demotivating and disengaging another group of employees.



Our top five features include employee satisfaction, time with the company, average hours worked per month, projects per year, and manager evaluation score. These features are dominantly used by the model to predict which employees will turnover.

#### Random Forest (On Test Set)

<Figure size 1440x1440 with 0 Axes>



	precision	recall	f1-score	support
stay	0.98	1.00	0.99	1998
left	0.98	0.90	0.94	401
accuracy			0.98	2399
macro avg	0.98	0.95	0.96	2399
weighted avg	0.98	0.98	0.98	2399

After dropping department and promotion data, and retraining the Random Forest Model, a slight improvement was observed. The F1-score increased by 1% to 94%. However, the model is still experiencing a more False Negatives than desired, lower left quadrant of the confusion matrix. We aim to reduce or eliminate them with further testing and development.