Salifort Motors: Reducing Employee Attrition

Executive Summary of Machine Learning Construction

Prepared by: Rebecca Iglesias-Flores

ISSUE / PROBLEM

Salifort Motors recently conducted its annual employee satisfaction survey, with executive leadership particularly focused on this year's results due to a recent divestiture, organizational restructuring, and a surge in employee attrition. The key question they seek to answer is: "**What factors contribute to employee turnover?**"

RESPONSE

The Data Team has been tasked with building a Machine Learning Model to predict when an employee will leave the company. This is part of a greater effort to keep talent and increase employee retention at the company.

A deep dive into the dataset revealed that none of the features exhibited a linear relationship with the target variable or with each other.

Logistic Regression was initially selected as the best model for fitting the data, but various model were explored for efficacy.

IMPACT

This model will uncover key insights into what's driving employee turnover, helping us understand the real reasons behind their decisions to leave. With this knowledge, the company can take meaningful action to address concerns, improve the employee experience, and ultimately create an environment where people feel valued and want to stay.

Breaking Down the Results

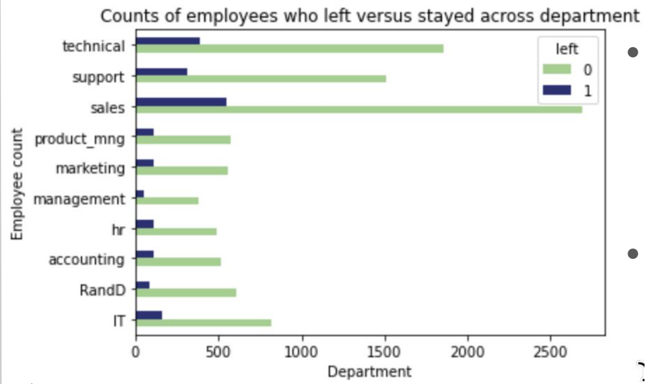

Fig 1. Employee Attrition per Department

- The departments with the highest proportion of employee attrition is **HR** at 18.8%, followed by **accounting** at 17.55%, and **engineering (technical)** at 17.38%
- Based on numbers alone, the department having the most employees leave was **Sales.**

Round 1

- Logistic Regression achieved 80% precision, 82% recall, and an F1-score of 80%.
- Decision Tree Classifier performed significantly better, with 93% precision, 92% recall, an F2-score of 92%, 97% accuracy, and an AUC of 0.97.

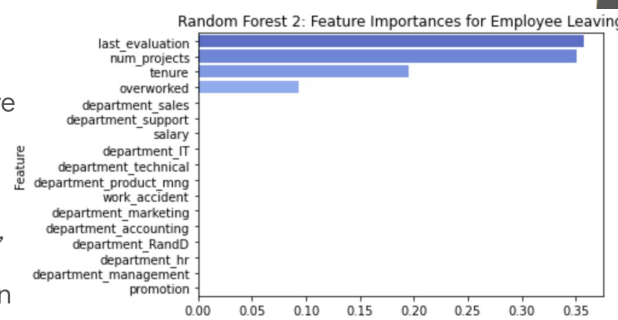

Fig 2. Feature Importance SOTA model

Round 2

- Random Forest Classifier slightly outperformed the Decision Tree Classifier.
- XGBoost far exceeded decision tree and random classifier, but at the cost of high computation time during cross-validation.
- Feature engineering did not improve model performance across any of the tested models.

Fig. 3. Model experiments and their evaluation metrics

model	precision	recall	F1	accuracy	auc
XGBoost CV	0.977977	0.916510	0.946200	0.982694	NaN
RandomForest CV	0.949339	0.914625	0.931569	0.977690	0.979395
decision tree CV	0.926282	0.917764	0.921857	0.974145	0.971091
XGBoost 2 CV	0.907583	0.878846	0.892868	0.964971	NaN
RandomForest 2 CV	0.863791	0.902698	0.882784	0.960175	0.964051
decision tree 2 CV	0.824575	0.902711	0.861388	0.951731	0.959327

- Our initial hypothesis that data leakage may be present in avg_mo_hours appears to be correct, as the second round of models showed improved generalization.

KEY INSIGHTS

Overwork is a strong predictor of employee attrition, as indicated by the correlation between number of projects, workload (overworked), and employee departures.

The most predictive features were number of projects, performance evaluation, overwork, and tenure. Workload-related factors (num_projects and overworked) directly impact retention, one of which was feature-engineered by our data team.

To improve employee retention and morale, the company should consider the following strategies:

- Cap the number of projects employees can be assigned.
- Address tenure-related dissatisfaction
- Reassess performance evaluation criteria
- Ensure high evaluations are not disproportionately tied to working 200+ hours per month.
- Implement a proportionate reward system that fairly recognizes contributions and effort.
- Encourage open discussions about workplace culture
- Hold company-wide and team-specific discussions to identify and address concerns related to work-life balance.
- Implementing these strategies can reduce attrition, enhance employee satisfaction, and improve overall workplace morale.