

Salifort Motors

Employee Retention Project

ISSUE / PROBLEM

Human Resources (HR) at Salifort Motors have created a dataset with working information of their employees. They want to understand what makes the employees leave the company. With this, they intend to increase the employee retention.

RESPONSE

Since the objective is to predict a categorical variable (whether an employee leaves or stays), we face a binary classification problem.

The team evaluated three classifiers (Logistic regression, Decision tree and Random forest) using a train / validation / test split.

The Random forest yielded the best performance, closely followed by the Decision tree.

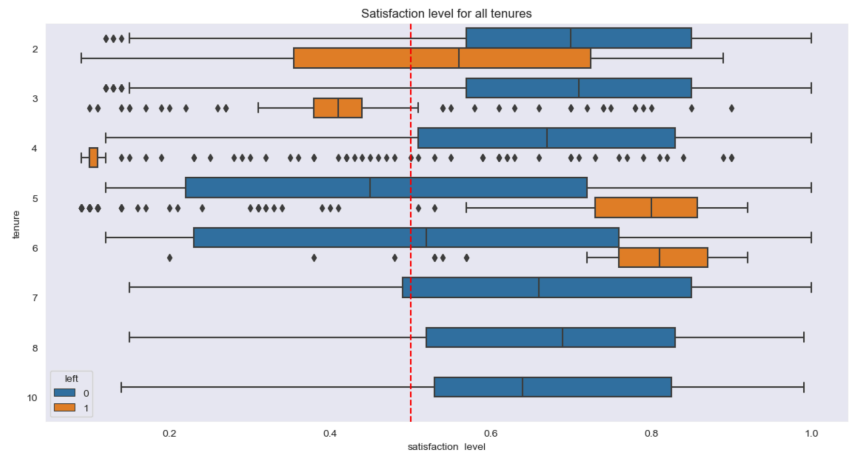
IMPACT

Being able to predict which employees will leave and which are the more influential variable may help the HR department to increase the employee retention, saving time and money.

INSIGHTS/NEXT STEPS

To increase employee retention, here are some suggestions derived from the analysis of the dataset:

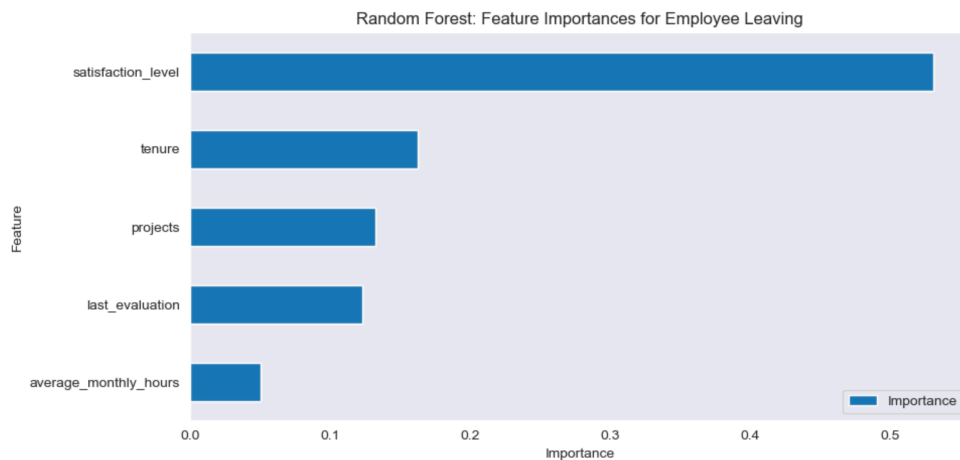
- Overworking decreases employee satisfaction: do not assign employees to many projects or long hours.
- Fourth-year employees may feel they are not rewarded and feel dissatisfied. Promote them.
- Positive evaluations should not be exclusive for overworked employees. Create standard scales.
- When joining, inform employees about the company's philosophy so they do not feel frustrated later on.



After working in the company for more than seven years, employees tend to stay. Fourth-year employees experience extremely low satisfaction levels.

	accuracy	roc_auc	f1	precision	recall
LogisticRegression	0.807968	0.155844	0.031915	0.052980	0.498466
DecisionTreeClassifier	0.978962	0.948229	0.925532	0.936743	0.957653
RandomForestClassifier	0.981647	0.958904	0.930851	0.944669	0.961389

The Random forest classifier yielded better results during the grid-search cross-validation performed using the training set. This was also the case for the validation stage.



The most important features (those which mostly reduced the Gini entropy of the Random forest) were satisfaction, tenure, number of projects, last evaluation and monthly hours.