Assessment of Salifort Motors' employee retention problem

Charly Moreno - Data Professional at Salifort Motors

ISSUE / PROBLEM

The main objective of the project is to generate a statistical / ML model that predicts whether an employee will leave the company or not, to identify relevant features that drive the turnout rate of employees and **propose an strategy to increase employee retention**.

The main research question is: **what's likely to make the employee leave the company?**

RESPONSE

Using the provided dataset, an EDA was conducted to identify main variables that could explain the turnout rate of the company and adequate the data to further modeling activities.

The data was splitted into three buckets:Training (60%), Validation (20%) and Testing (20%). This minimized the bias when comparing models performance. Three models were evaluated: Logistic Regression, Random Forest and XGBoost, **being selected the XGBoost model as the winner**.

The XGBoost model had more than 95% in each evaluation metric and a Recall of 96.4% (True positives against all positives), which makes it a production model.

KEY INSIGHTS

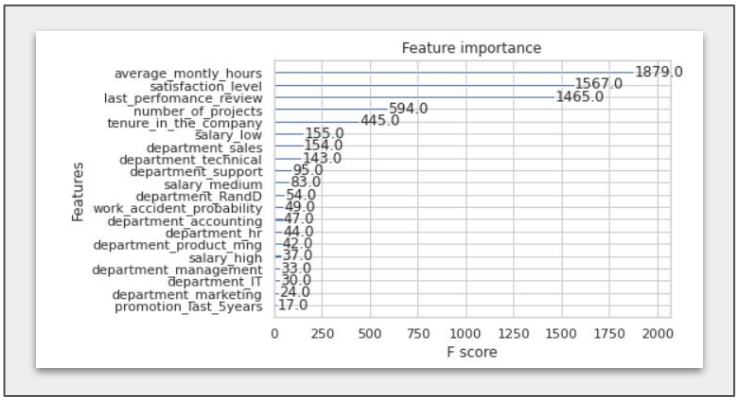
Out of the 20 variables used to fit the winner model (XGBoost), three of them were the **major predictor variables: average monthly hours, satisfaction level and last performance review**.

From the EDA, we identified that the mean average monthly hours of turnout employees is 225 hours, 10% higher than the median of stayed employees (200 monthly hours).

Other relevant variables are: number of projects and tenure in the company. Less relevant variables are salary, department and promotion in last five years.

	model	precision	recall	F1	accuracy
0	XG Boost CV (Training)	0.979572	0.956138	0.967680	0.984776
1	XG Boost CV (Validation)	0.985612	0.959384	0.972321	0.987000
2	XG Boost CV (Testing)	0.982882	0.964986	0.973852	0.987667

Evaluation metrics of XGBoost models (validation model was the winner)



Feature Importance plot of XGBoost Testing model

IMPACT

A retention strategy can be set using the results gathered from the XGBoost model selected. For example:

- **Reducing the labor charge for the segment of workers that are considering to leave** the company will help improving the retention, since that is the most valued action considered by the workers of the company.
- The **segment of workers** that should receive this benefit are workers that **had a very good performance review** last year, and also that **manifest low engagement** with the company when asked through the satisfaction level survey.
- Based on our research, the **laboral hours of these workers should be diminished from 225 to around 200 monthly hours**. This is the best strategy to tackle the current HR problem.