Salifort Motors HR Machine Learning Project

Prepared by Bryan Koewler 08 Feb 2024

Overview

The human resources department collected and provided data to be analyzed for insight into employee satisfaction and turnover. They want to know the factors that cause an employee to leave. In addition to improving the satisfaction, culture and camaraderie of the employees, a reduction in hiring and training costs is anticipated.

Objective

To build a model that can predict whether an employee will leave or stay, to identify the factors that lead to employee turnover, and provide solutions to increase retention and decrease turnover costs.

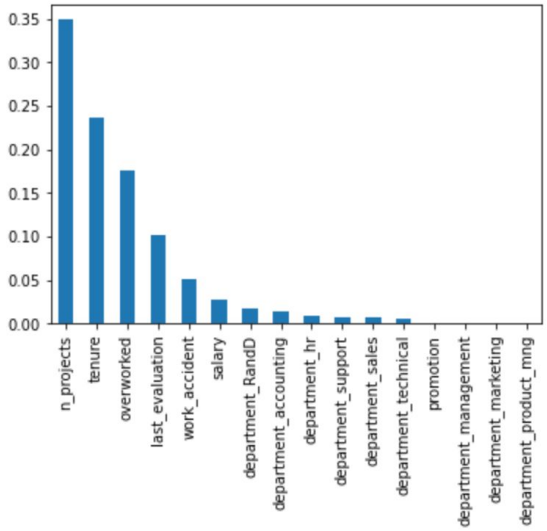

Fig 1. Feature importances of XGB model by total decrease in Gini impurity

Results

After testing multiple models including logistic regression, decision tree and random forest, the xgb technique produced the best results, listed in the table below. The model is very good at predicting those who would leave and those who would stay, with a ROC AUC of 0.94.

An employee is more likely to be dissatisfied or leave when :

- number of projects < 4 or > 5
- average monthly hours >= 175
- tenure year = 3, 4

	model name	precision	recall	accuracy	f1	auc
0	xgb champion model	0.909091	0.903614	0.968979	0.906344	0.942807

Fig. 2. Table of metrics from final xgb model on holdout data

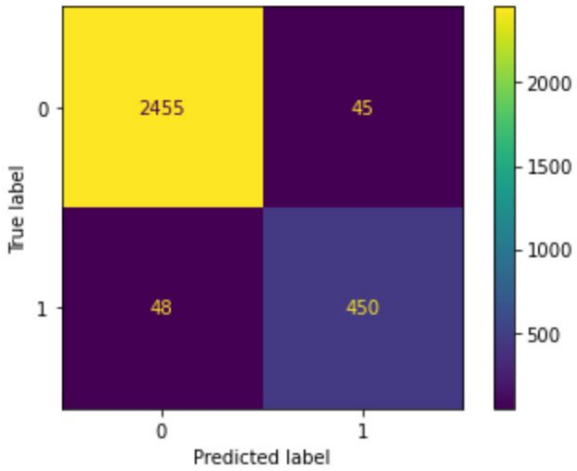

Fig. 3. Confusion matrix of xgb classification results on holdout data

Next Steps

Implementation of limits for monthly hours worked, number of projects, and gathering more information as to why tenure years 3-4 are the most likely for dissatisfaction. Implementation of regular surveys or discussion to keep tabs on employee satisfaction and reduce turnover. Further analysis may be completed by building a K-means model and clustering data. Finally, reworking the model to predict satisfaction level or evaluation scores may prove useful to further enhance ability to identify those who would leave.