Salifort Motors Employee Turnover Project

Prepared under the guidance of the Senior Leadership Team

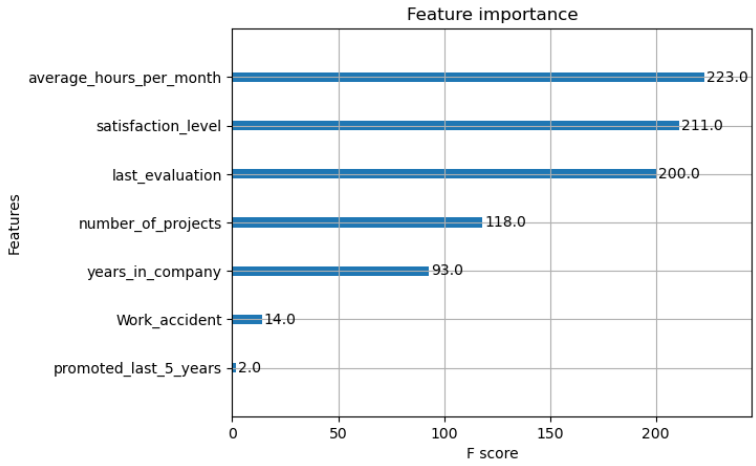
Project Overview

The data analytics team is developing a data analysis project to predict whether or not an employee will leave the company based on the combination of factors like average number of hours worked per month and overall satisfaction rating. Employee turnover means that either the employee has left the company via termination, laid off, and resignation. By building a machine learning model on whether the employee will leave due to various reasons, an ideal solution can be proposed to help improve retention rate and lower employee turnover.

Key Insights

- Three tree-based models have been built for this project: decision tree, random forest, and XGBoost.
- The dataset is divided into training, validation, and test sets, which helps decide which model is the most suitable one in predicting unseen data.
- All three models have performed effectively in fitting the model, owing to the metric scores being higher than 0.9.
- Out of all the three tree-based models built for the project, the decision tree is designated as the champion model in terms of the F1 score.
- The confusion matrices indicate that the combined total of false positives and false negatives are significantly lower than the combined total of true positives and true negatives, and that is how it contributes to the high scores for accuracy, recall, precision, and F1.
- Upon examination of the feature importance plot for the XGBoost model, the top three features that stands out are average_hours_per_monrth, satisfaction_level, and last_evaluation.
- Overall, tree-based models are more effective than logistic regression due to the fact that they require minimal preprocessing and data cleaning, and the models can handle outliers effectively without adversely affecting the results.

Details



Feature importance plot on employee turnover (extracted from the XGBoost model)

Next Steps

- The current data is sufficient enough to satisfy the problem domain and predict employee turnover based on the combination of factors. The tree-based machine learning models demonstrated the efficacy of making classifications, and a critical need for additional data to make more accurate predictions.
- For future work to consider, the top three features from the feature importance plot should be examined further to find out how they played an essential part in contributing to the likelihood of the employee leaving the company.