Employee Turnover Project | ML Model Results

ISSUE / PROBLEM

Salifort Motors leadership has asked the data team to build a ML model that will accurately predict future employee turnover. A powerful model can help managers better understand why employees are leaving and take proactive measures to increase employee retention. The model is based on data collected by Human Resources of Saliford Motors.

RESPONSE

To obtain a model with the highest predictive power, we developed two different models to compare results: the logistic regression model and the random forest model.

Prior to model construction, we emphasized data quality by cleaning, addressing missing values, outliers, and inconsistencies. Categorical variables were transformed for model compatibility. We split the dataset into training and test sets to validate performance and ensure generalizability.

IMPACT

- The Random Forest model fits the data better than the Logistic Regression model.

- The champion model achieved 98% accuracy, 96% precision, 93% recall and 94% F1-score. Based on the performance metrics. We conclude that the model is effective for predicting employee turnover
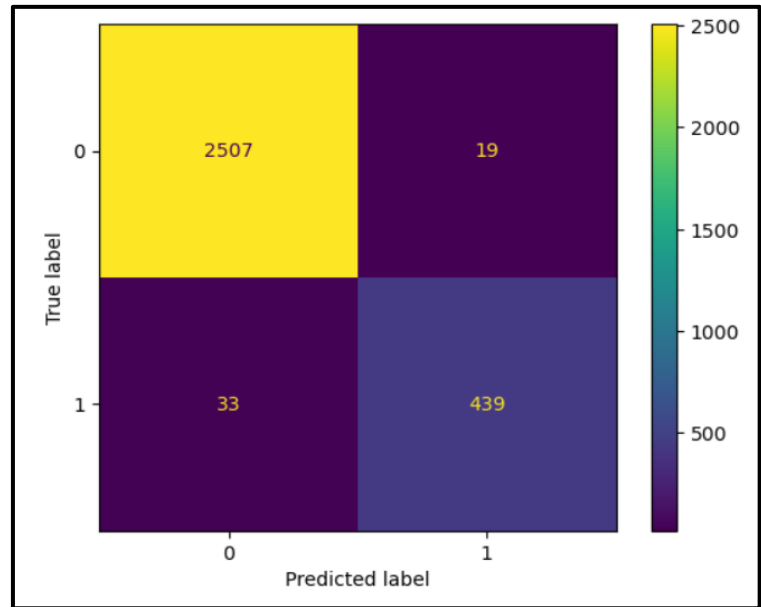
Confusion matrix of Random Forest Model
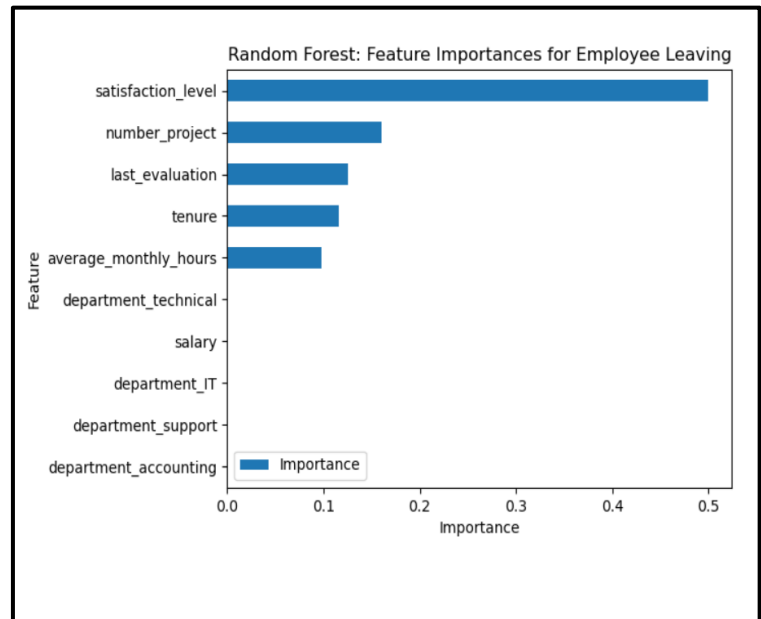
Feature Importance of Random Forest Model

KEY INSIGHTS

The results of the model prediction:
- 2507 correctly predicted as not leaving the company.
- 439 people accurately predicted as leaving.
- 19 people who did not leave and the model inaccurately predicted as leaving.
- 33 people who left and the model inaccurately predicted did not leave.

The top 5 important features for the model prediction are:
- Satisfaction level
- Number of projects
- Last evaluation
- Tenure
- Average monthly hours