

Employee Retention Project | XGBoost Modeling Results

Prepared for: Salifort HR & Leadership Team




Google Advanced Data Analyst Capstone Project by Israel Hernandez

OVERVIEW

HR requested a data analyst review the results from an employee satisfaction survey to develop a model that can predict employees will leave the company. For this task, we used XGBoost binary classification as an excellent choice in turnover prediction for improved employee retention and reduced cost of re-training new hires.

Problem/Issue

XGBoost - Binomial Classification

-  **Target Goal:** Use survey data to build and analyze a binary logistic model.
-  **Methods:**
 - Use GridSearchCV to find the best estimator
 - Split data into 60/20/20 Training/Validation/Test
 - Model XGBoost against the training data.
 - Test against the validation data
 - Choose the champion model and run it against the test data.
-  **Impact:** Without having to deal with outliers or collinearity, decision tree modeling is flexible.

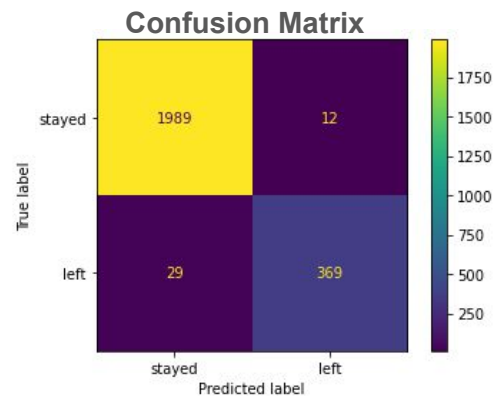
NEXT STEPS

We recommend using the XGBoost model which had great scores, as shown below.

model	precision	recall	F1	accuracy
XGB test	0.968504	0.927136	0.947368	0.98291

- Find ways to reduce or limit workweeks to 45 hours per week or less, while ensuring employees with shorter workweeks be given more work time.
- Look into ways to alleviate the load on those with 6 or more projects--possibly dividing the work with those who tend to have much less than 160 monthly hours.
- Ascertain why employees leave after 5 years of tenure.
- For employees with excellent evaluations, discuss what incentives would help keep these talented individuals.

KEY INSIGHTS

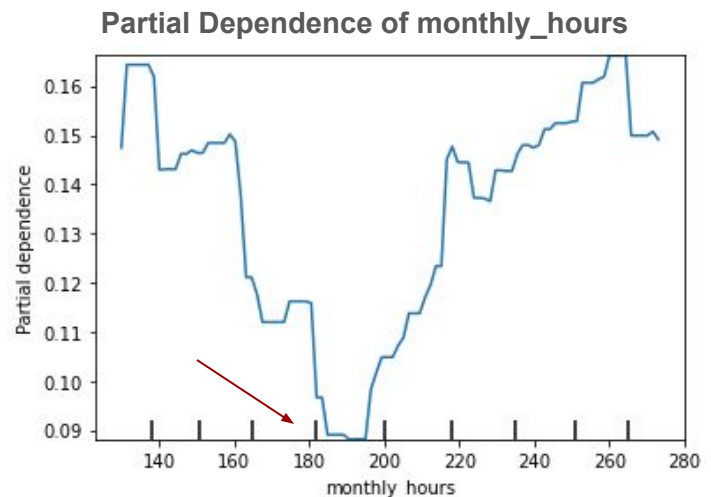


Per the Confusion Matrix, the model predicts that a small number of employees stayed (29) but actually left.

Feature Importance Top 5: (Refer to the Appendix)

1. satisfaction_level
2. average_monthly_hours
3. last_evaluation
4. number_project
5. Tenure

Partial dependence of monthly_hours shows higher churn when hours are extremely **below OR above** the **176-hour** typical monthly hours, based on a 40-hour work week.



Appendix

Feature Importance

