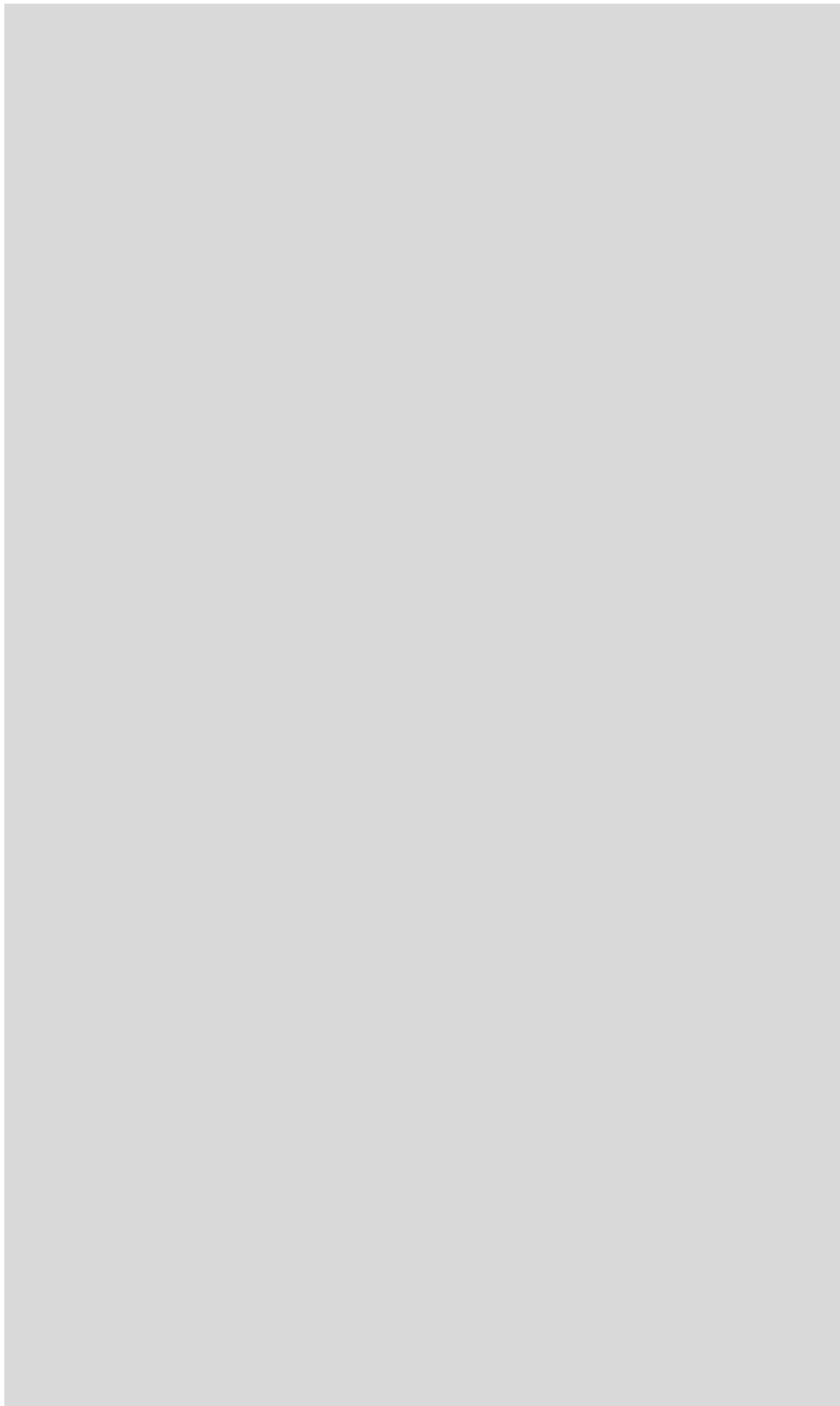
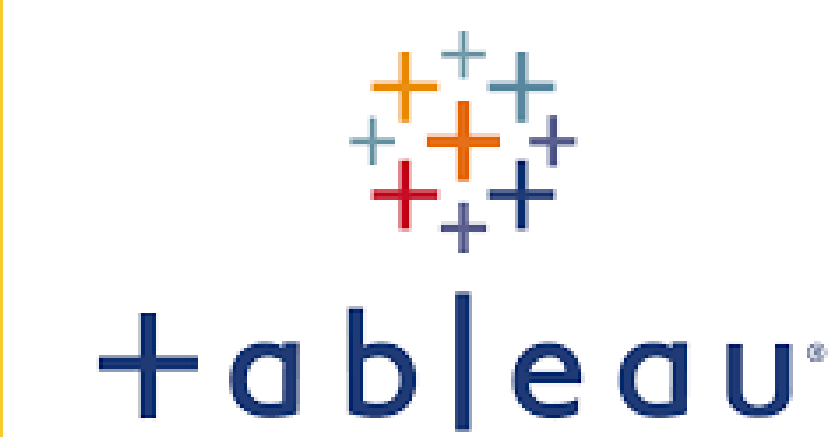


XGBoost Unleashed: A Data-Driven Approach to Predicting Employee Churn

James Down Graduation: December 2023

Faculty Advisor: Dr. Paul Johnson



Employee turnover is a costly endeavor for companies.

- Wide range between 30% and 500% of employee salary.

- Includes posting, interviewing, hiring and training costs.

Identifying the key reasons for employee departures within an organization can equip them with the knowledge needed to address voluntary turnover.

- Reduce turnover costs if they can retain employees

- Take preventative steps for identified at risk employees

- Explore influential factors and reduce their negative effect on employee turnover

Utilizing classification models can uncover the essential variables driving voluntary employee turnover and offer predictions for future turnover.

Random Forest

- Collection of Decision Trees, that are each slightly different, that make individual decisions which are then averaged to make a final decision.

The data exploration in **Figure 1** reveals an imbalance between the number of employees who left the company and those who are still employed. Instead of attempting rebalancing or dataset sampling, two models were tested to determine which one could better manage this imbalance.

During the data exploration shown in **Figure 2**, a distinct group of employees stood out, constituting over 50% of all those who left – specifically, those who worked overtime. This factor's significance will be further highlighted in the variable importance analysis.

Random Forest:

Accuracy : 83.28% True Positive Rate: 95.12%

Kappa : 20.89% True Negative Rate: 21.28%

AUC: 80% (**Figure 4**)

The Random. Factors with higher gains had a more pronounced on the model. Moreover, the higher the cover, the greater the number of observations impacted by the factor.

The scaled importance of features in **Figure 6** shows how influential OverTime is compared to all other features. The rest of the features exponentially decay.