Machine Learning II:

Employee Attrition Prediction

Group B

Table of Contents

01

Executive Summary

Executive Summary

Goal:
Predict the attrition of employees

Feature Engineering:
Label encoding and Creating New Variables

Model Deployment:
Iterative sampling, Model Selection

Final Model:
Random Forest Classifier and XGBoost

Business Insights and Strategies:
Benefits and Employee Retention Strategies

02

Feature Engineering

Feature Engineering

01

Dropping Unnecessary Features

Features with one single value in all rows

02

Creating New Features

Aggregating existing features

03

Dealing with Numerical Features

Not removing outliers nor applying scaling

04

Dealing with Categorical Features

Label encoding

05

Deciding Final Features

Dropping variables that have been aggregated and highly correlated ones

Feature Engineering: Creating New Variables

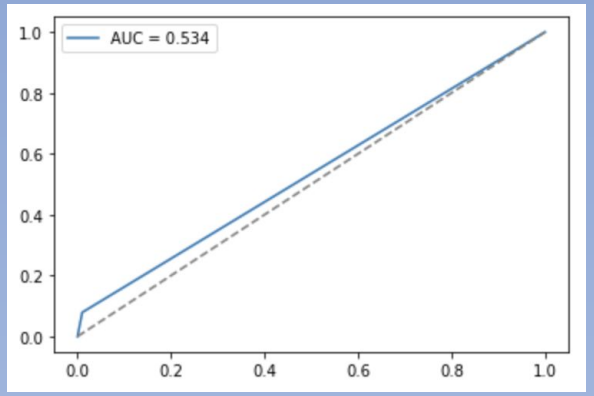
New Variable Name	Description
Age_group	categorizing 'Age' into 5 different groups ('18-30', '31-40', '41-50', '51-60', '61-70')
Distance_group	categorizing 'DistanceFromHome' into 3 different groups ('Short', 'Medium', 'Long')
Satisfaction_level	'EnvironmentSatisfaction' + 'RelationshipSatisfaction'
Job_Title	'JobRole' + ' - ' + 'Department'
Experience_Factor	'TotalWorkingYears' * 'JobLevel'

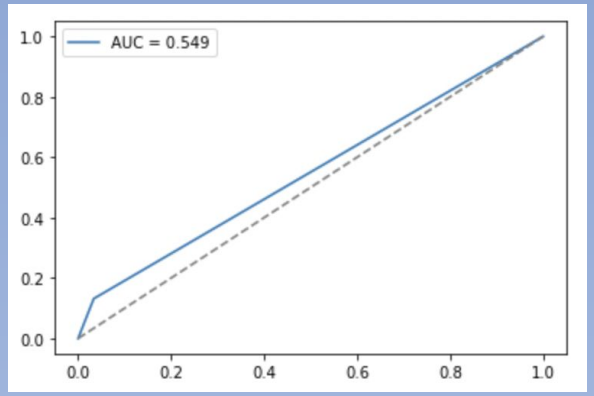
03

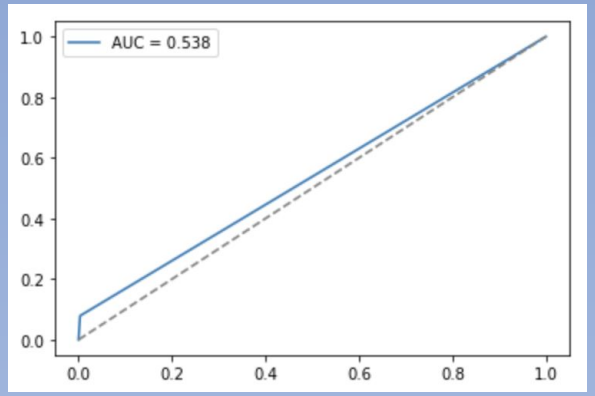
Model Deployment

Models Deployment

Trial: Basic Random Forest	Trial: XGboost Classifier	Trial: Logistic Regression





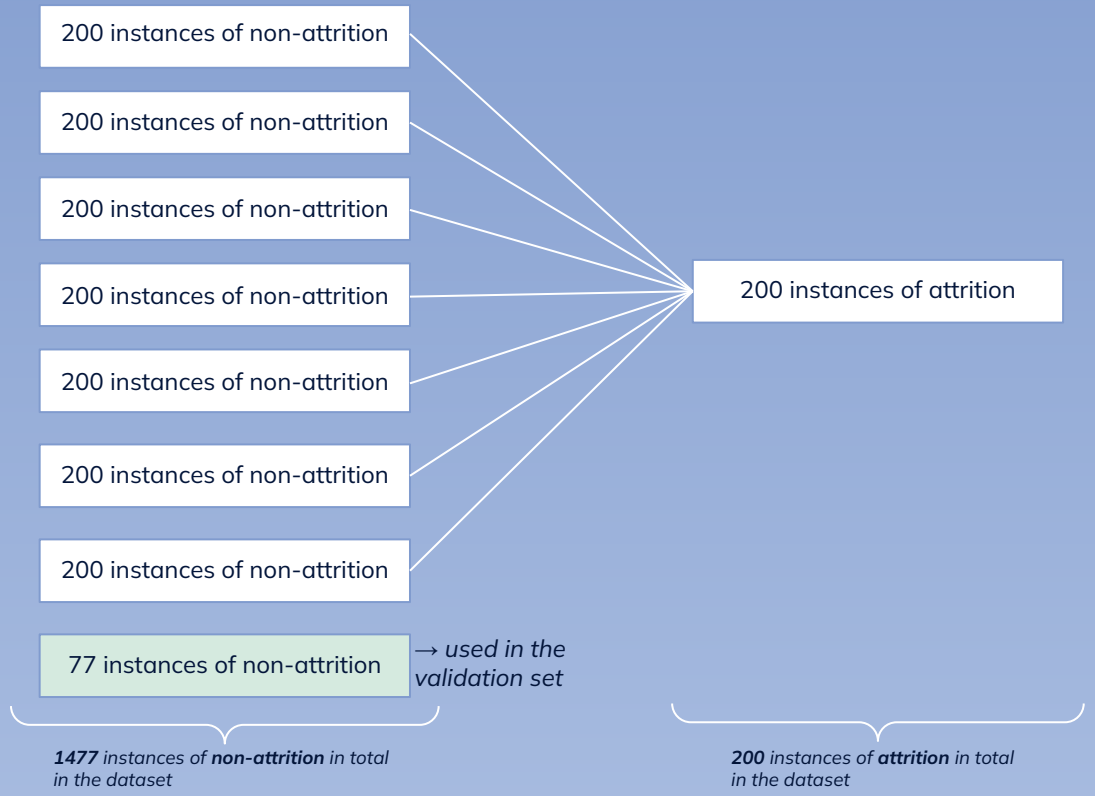


Problem: Hugely imbalanced dataset (attrition : non-attrition ≈ 1 : 10)

Models Deployment: Handling Imbalanced Data

Iterative Sampling:
*similar with **Bagging**, but instead of running in parallel, its run iteratively, storing all predictions made instead of discarding the worst ones.*

The benefit is that it generates synthetic data points that are similar to the minority class, reducing bias and improving the model's ability to generalize



04

Final Predictive Model

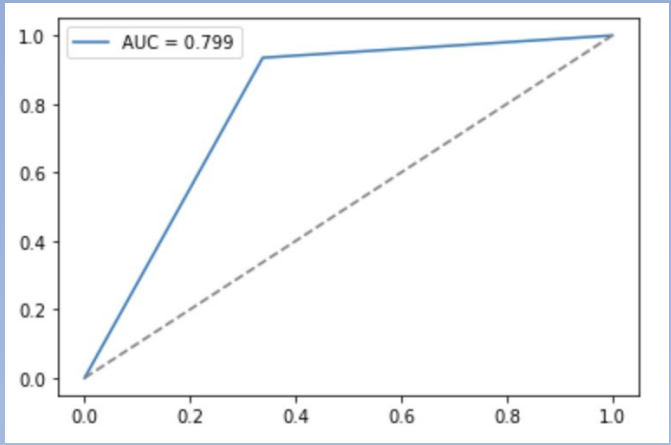
Final Model Choice

Benefits	XGBoost	Random Forest	Combination
Better prediction accuracy	✔	✔	Increased accuracy
Improved generalization	✔	✔	Reduces the risk of overfitting and improves generalization.
Enhanced feature importance	✔	✔	Better understanding of important features
Faster training time	✔	✔	Faster training, without sacrificing performance.
Better interpretability	✘	✔	More interpretable model
Robustness to outliers	✘	✔	Robust model that is less affected by outliers.

Combination of XGboost & Random Forest

How:
Calculate the mean of each iteration of XGboost & Random Forest and round it to a binary value

Result:
AUC score on Kaggle: **0.76 for Private and .82 for Public.**



05

Benefits of Model's Predictions

Benefits of Model's Prediction

According to the U.S. Bureau of Labor,

the average employee turnover rate in 2021 was 47.2%.

The turnover rate includes employees who voluntarily quit, layoffs, retirements and discharges.

Benefits of Model's Prediction

Managers can take preventative measures with employees who are predicted to be at risk of leaving

- Ex: Meetings to talk about job satisfaction, goal setting, and general feedback

Better understanding of employee behavior and the factors that contribute to retention

- Ability to make informed decisions to enhance performance and gain a competitive advantage

Benefits of Model's Prediction

Less energy and time spent by the human resources team in the recruitment process

- No need to constantly replace lost talent
- More resources to be used for career development

Decreased onboarding and training costs

- More time for managers and senior employees to focus on projects

06

Employee Retention Strategies

Employee Retention Strategies

Sending quarterly surveys to all employees, adding emphasis to those who are predicted as leaving.

- Gather feedback for what factors are causing dissatisfaction
- Create an environment of trust

Sending invitations for focus groups to those who responded to the survey

- Get more in depth information about their current job satisfaction and also ideas for how to improve the employee experience.

Employee Retention Strategies

Providing benefits or salary increases to those employees who are predicted as leaving

- Have the highest performance rating
- Employees who have not been promoted in 4 years or more.

Implementing career development programs for entry-level employees

- Keep employees engaged and increase performance
- Keep employees in track for promotions/salary increases

07

Conclusion