Employee Attrition Prediction Project

Nil Virgili
Ironhack
July 2025
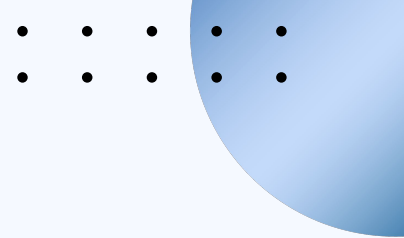
Table of contents
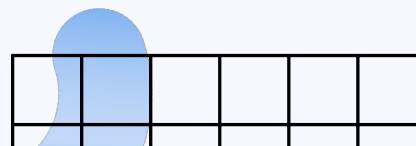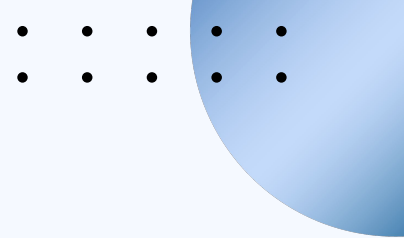
1. Introduction

Context and Business Objective

2. Data Overview
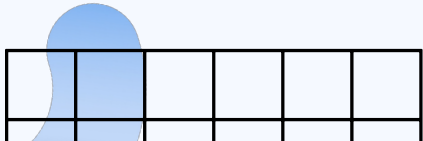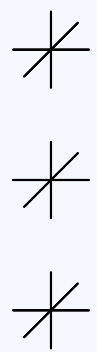
Dataset, Description and Fields

74k rows

24 features:

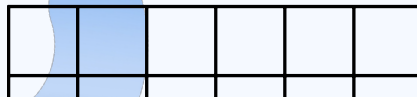5 numerical

11 categorical/ordinal
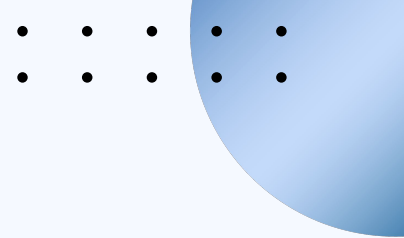
5 binary

Measures:

47% attrition rate (Target)

38.5 Avg. age

7300$ Avg. monthly salary

54% Males
46% Females

3. Data preparation
Cleaning, EDA and Feature Engineering

Find inconsistent features

Cleaning

Understand data characteristics / relations

EDA

Create new features

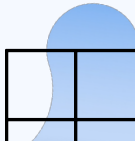## Feature Engineering
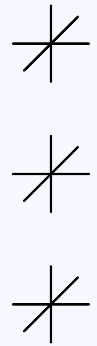
Removed features:
company_tenure
years_at_company

Higher attr. Rates:
Age: Young (19–25)
Gender: Female
Job_level: Entry level
Marital status: Single
WLB: poor, fair
Performance: low, b.avg.
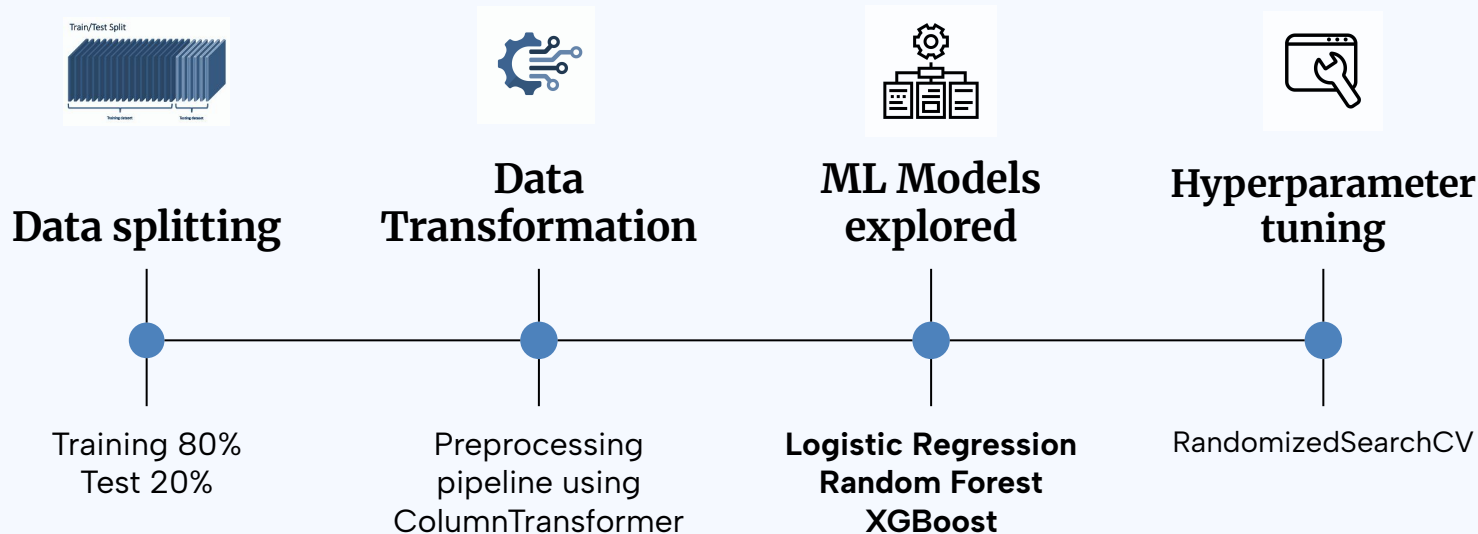Overtime: yes
Nr. Dependants: <=3
Nr. promotions: <=2
Comp.reput: poor, fair

New features:
Age_group
Promotions_group
has_many_dependants

4. Model

Development and optimization

Data splitting

Training 80%
Test 20%

Data Transformation

Preprocessing pipeline using ColumnTransformer

ML Models explored

**Logistic Regression
Random Forest
XGBoost**
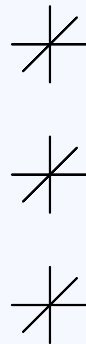
Hyperparameter tuning

RandomizedSearchCV

5. Key Findings

EDA highlights

Explore Tableau

5. Key Findings

Model performance

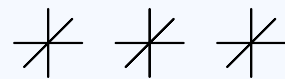XGBoost (Broad/Fine Tuned) provide the best results

Model Comparison Across Metrics

Model
- Logistic Regression
- Random Forest
- XGBoost (Untuned)
- XGBoost (Broad Tuned)
- XGBoost (Fine Tuned)
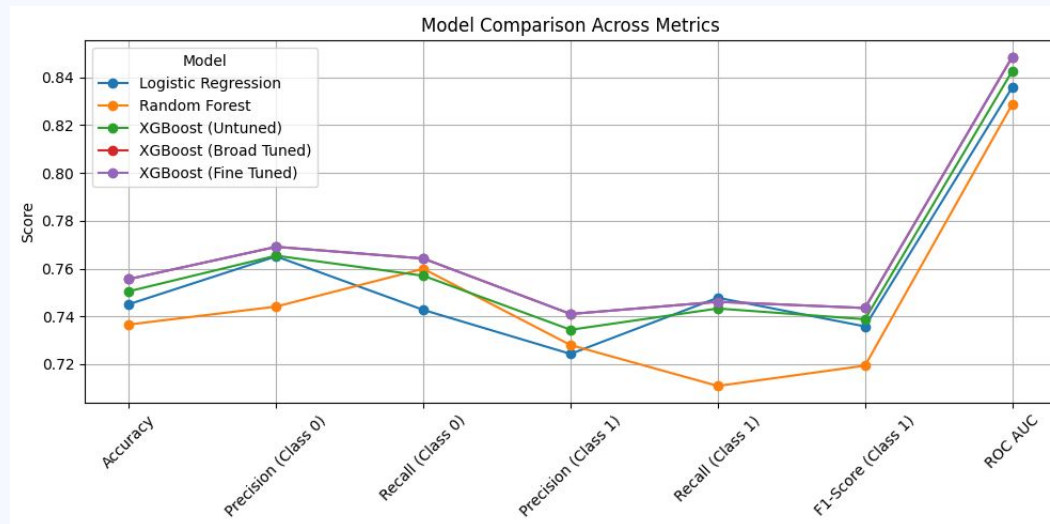
5. Key Findings

Feature Importance

Top 10 Features - XGBoost (Fine Tuned)
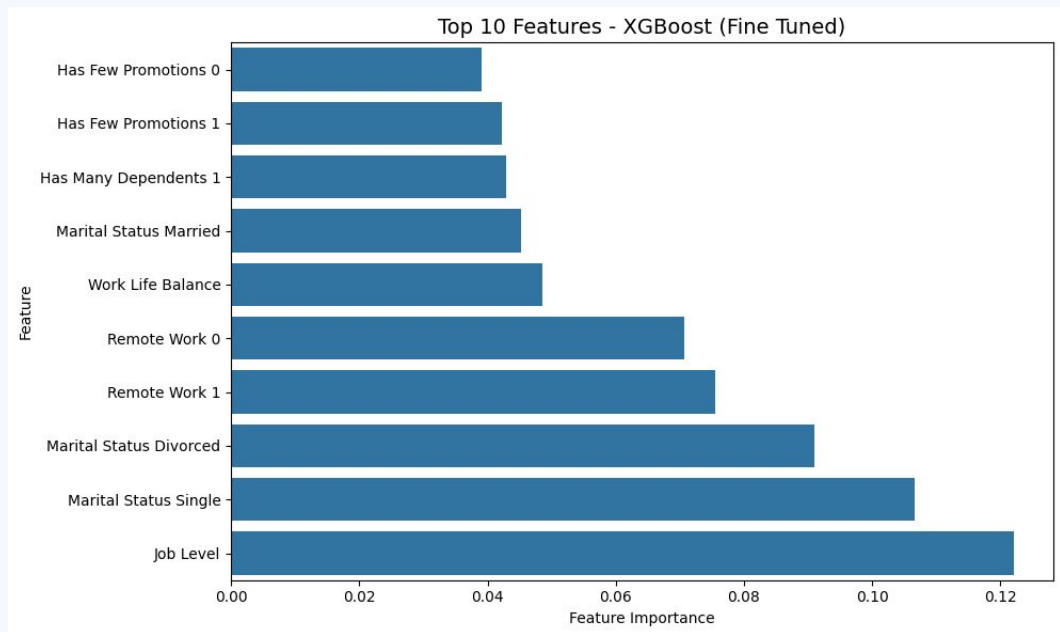
6. Recommendations

Adopt model 1.

XGBoost (Broad Tuned)

Investigate 3.

Nuances

Feature engineering 5.

Explore additional features
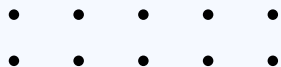
Proposals

2. **Retention**

Key drivers

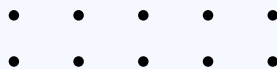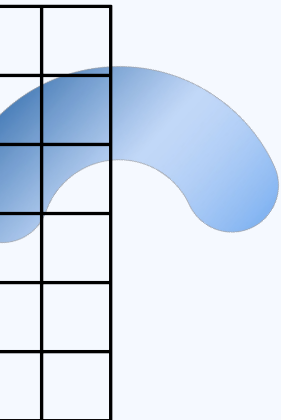4. **Threshold tuning**

Define optimal classification threshold

6. **CI & monitoring**

Model monitoring and training

Thanks