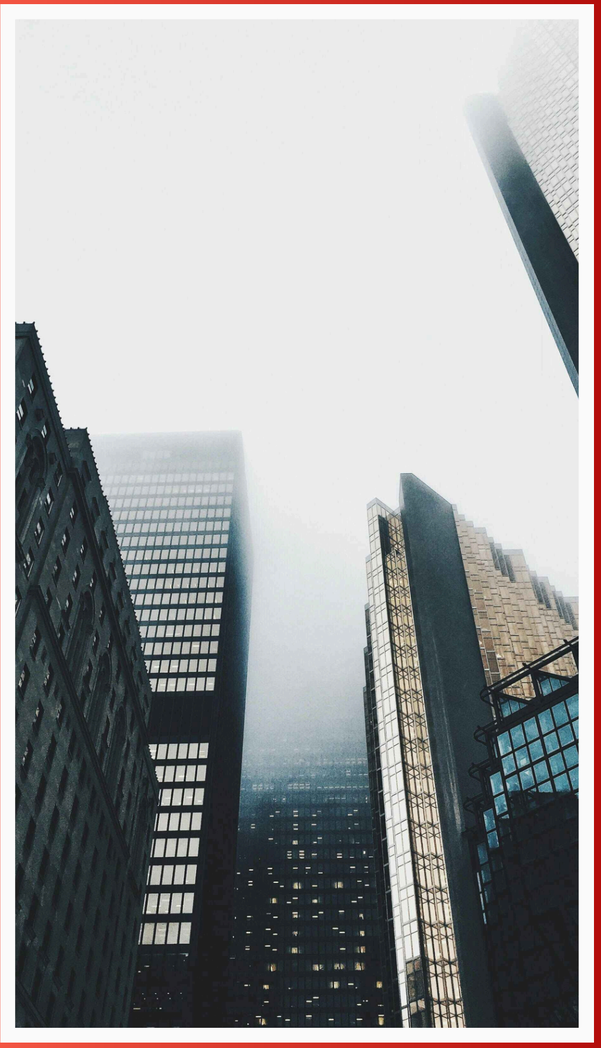
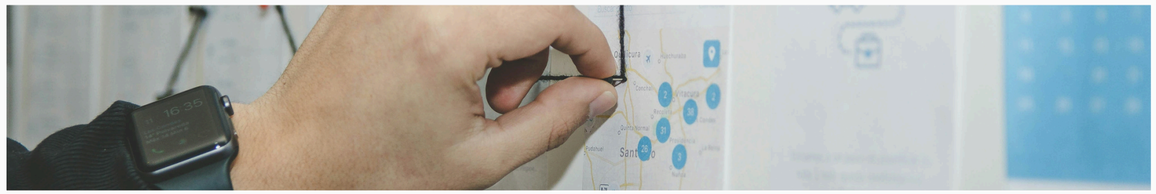
Improving Employee Retention by Predicting Employee Attrition

By: **Az-Zukhrufu Fi Silmi Suwondo**



Problem & Objectives



Problem Statement

The organization struggles with **high employee turnover**, leading to **costly recruitment** and **lost productivity**. Attrition also **harms team morale and cohesion**. A lack of data-driven insight into why employees leave makes retention efforts mostly reactive.

Goal

To **reduce the overall employee turnover rate by 15%** within the **next 18 months** by developing and fully implementing a **predictive, data-driven retention framework** that proactively identifies at-risk employees and enables targeted intervention strategies

Objectives

- Analyze **key variables** to identify **top predictors of attrition**
- Build a **machine learning model** to **assign attrition risk scores** to employees
- Translate insights into **actionable HR recommendations** to improve culture, growth, and satisfaction

Success Metrics

-15%

Attrition Rate

Compared to the previous period (resignations)

+15%

Employee Engagement Score

Target increase from previous period

Dataset Overview

Rakamin Academy

Dataset Summary

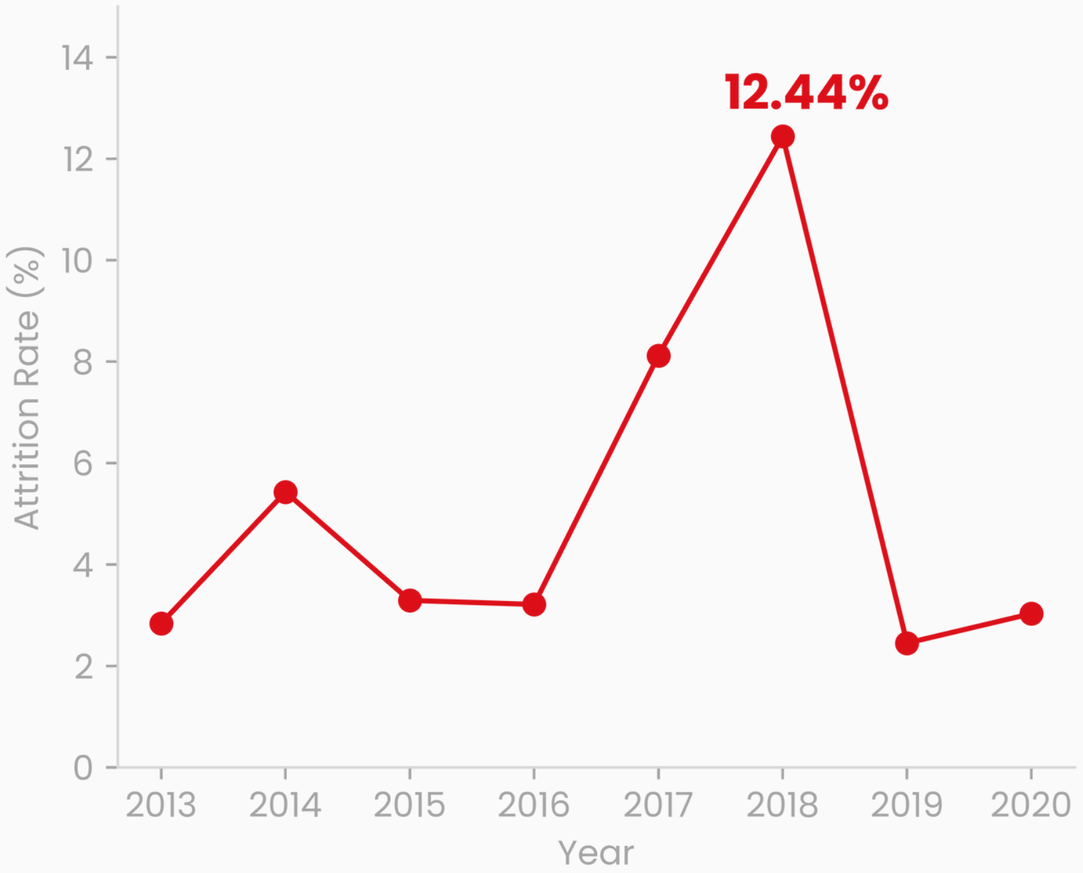
- Total Employees: 287
- Resigned Employees: 89
- Hiring Period: Jan 2006 – Jul 2018
- Resignation Period: May 2013 – Sep 2020

Features

- Demographics: Gender, Marital Status, Region
- Employment Info: Job Role, Career Track, Hiring Platform
- Scores: Engagement Score, Satisfaction Score, Project Count
- Behavioral Data: Absences, Late Arrivals, Resignation Reason

Source Data

The 2018 Crisis



Product Manager

Severe attrition in a strategic role signals leadership instability

Long Working Hours

Most cited reason for leaving indicates unsustainable workload

Website

Highest attrition source shows gap between employer branding and reality

Rakamin Academy

April Attrition Spike

8.2%

High-performing employees resigned during this period

14.3%

Employees with low engagement were more likely to resign

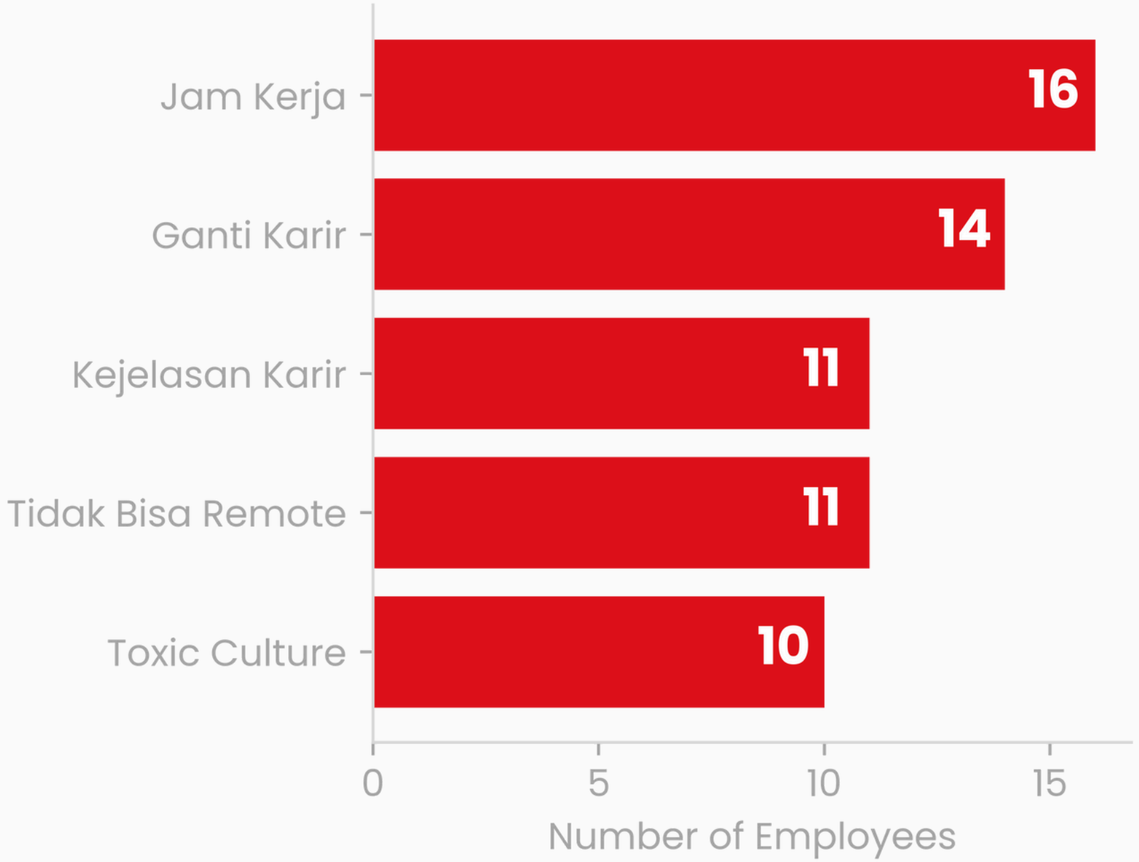
12.5%

The Product Design team experienced a significant attrition rate

13.6%

Employees hired through CareerBuilder showed a higher attrition rate

Top 5 Resignation Reasons



Front-End Engineers

8 out of 11 resigned due to lack of **remote work** options

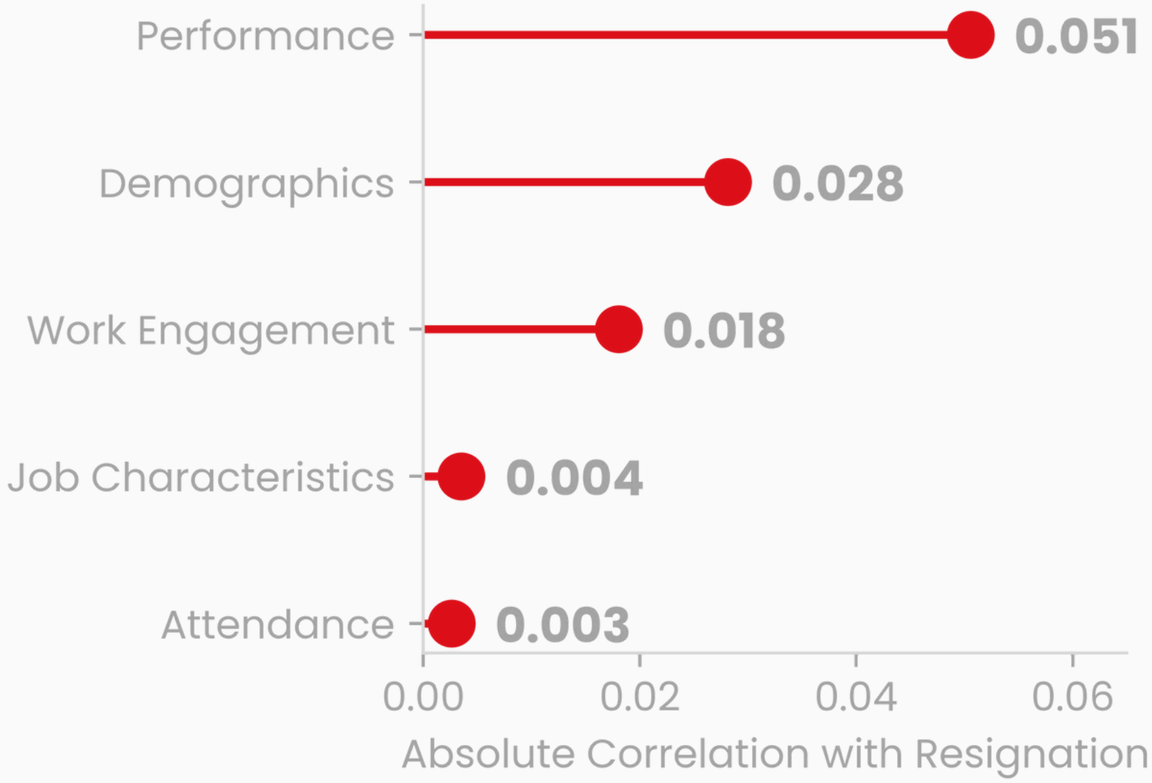
Data Analysts

6 employees cited **toxic culture** as their main reason for leaving

Top Performers

7 high-performing employees left because **remote work wasn't allowed**

No Significant Predictors Found



Key Finding

A comprehensive analysis of **287 employees** across **five feature categories** revealed **extremely weak correlations**, all below 0.06. The highest was just 0.051, indicating no factor group shows predictive power.

Implications

- Turnover is **complex** and **personal**
- **HR metrics alone can't** predict resignations
- **Culture** and **personal** factors matter more

Preprocessing

Data Splitting

- Features (X) and target (y) separated
- Stratified train-test split (80/20)

Missing Value & Outlier

- Numerical columns filled with median
- Categorical column filled with mode
- Outliers imputed with median

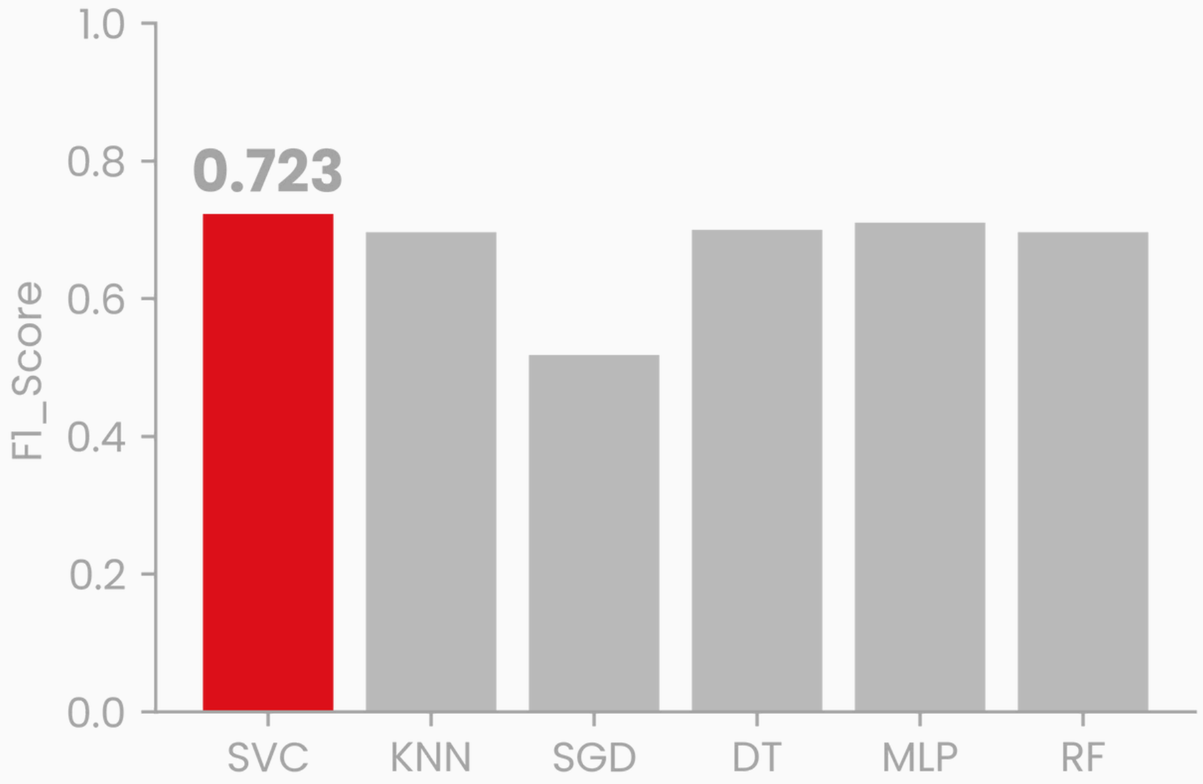
Feature Scaling & Encoding

- Numerical features scaled using RobustScaler
- Ordinal encoding for ordinal variables
- Frequency encoding for high-cardinality columns

Feature Engineering Highlights

- Time-Based Features
- Performance & Engagement
- Attendance Patterns
- Career Signals
- Demographics
- Advanced Features
- Clustering & Risk Scores

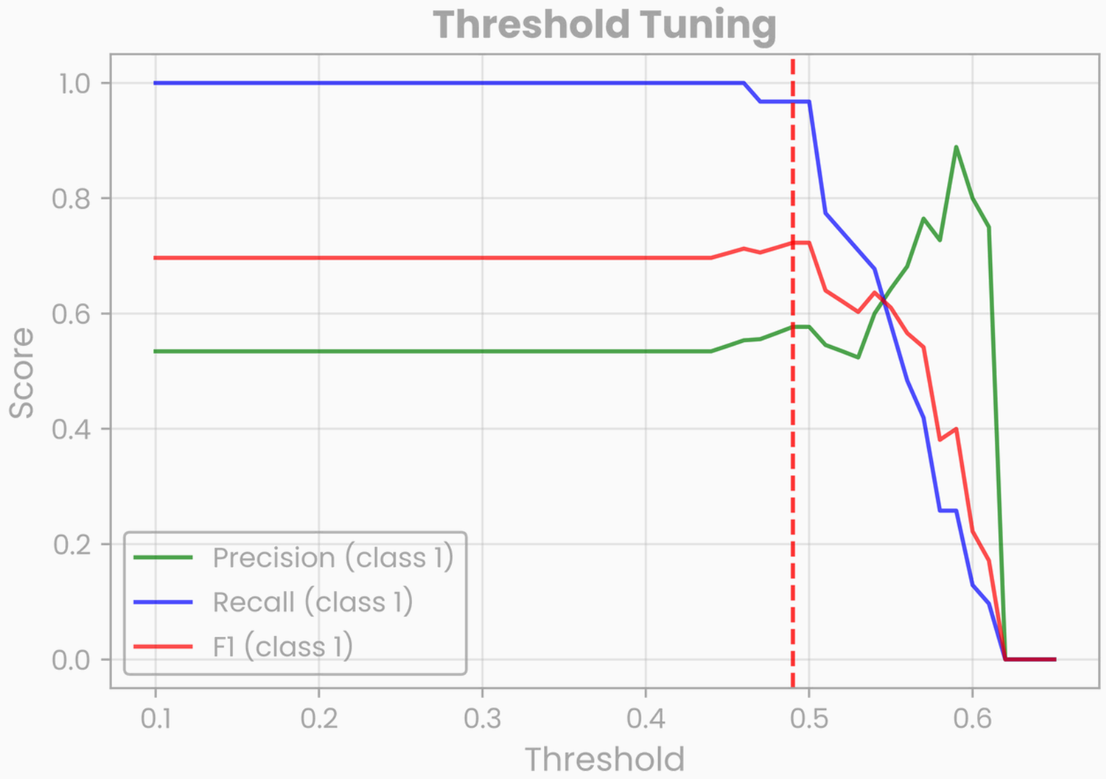
Model Selection



The **SVC model outperformed** other candidates, showing consistent metrics. Its **highest F1 score** indicates strong class separation ability.

Model	F1-Score	Precision	Recall
SVC	**0.72**	**0.58**	**0.97**
MLP	0.71	0.6	0.87
SGD	0.7	0.57	0.9
KNN	0.7	0.53	1
RF	0.7	0.53	1

Best Model Optimization



Hyperparameter Tuning

- Method: **Optuna** (30 trials, 5-fold CV, scoring = ROC AUC)
- Best Parameters:
  - C=**3.069**
  - kernel=**linear**
  - gamma=**auto**

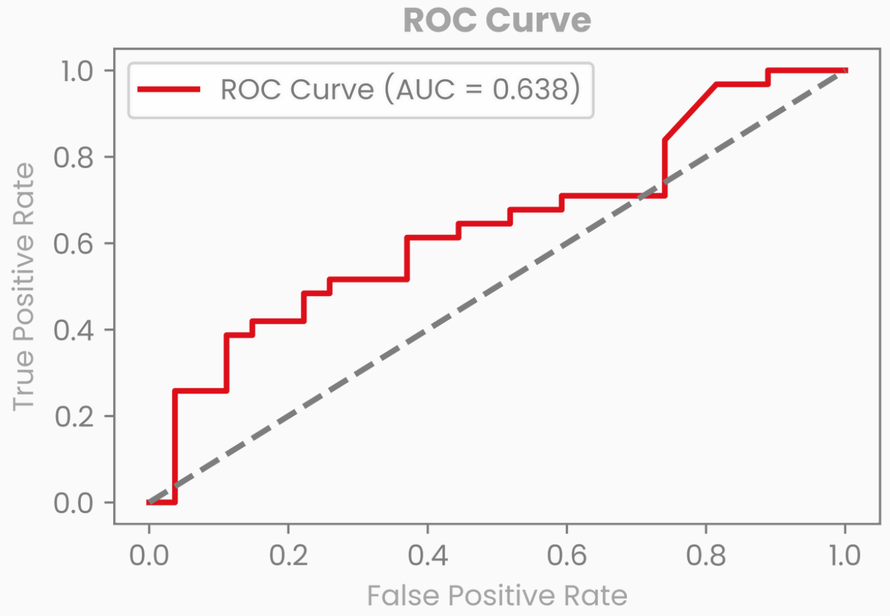
Threshold Tuning

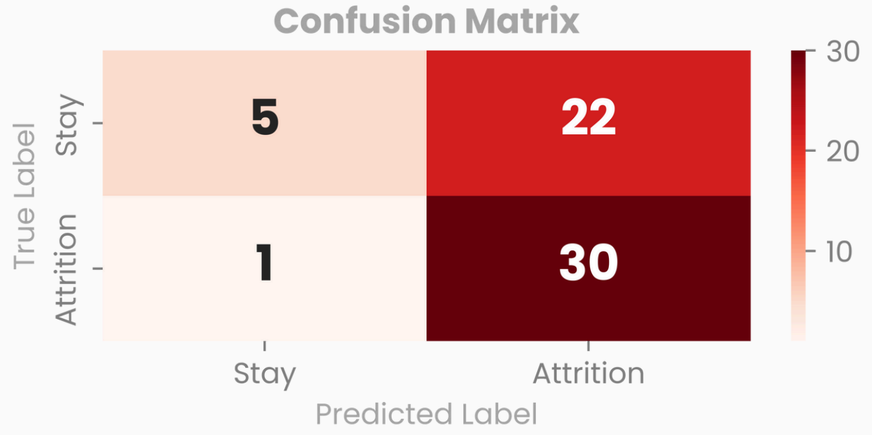
- Method: Threshold search (0.1–0.9) with **Precision–Recall–F1 trade-off** evaluation
- Best Threshold = **0.49**

Best Model Evaluation

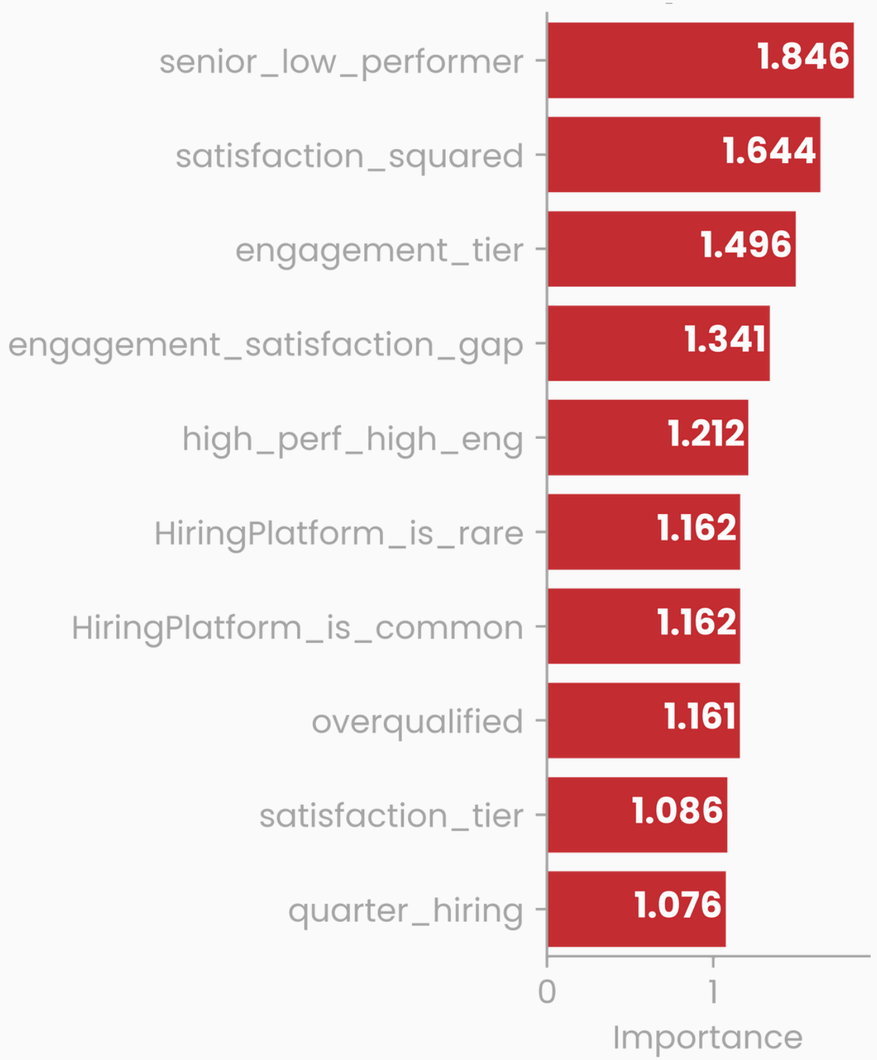
Recall: 0.97
Precision: 0.58
F1-Score: 0.72

The model demonstrates strong predictive capability with an AUC of **0.638** and F1-score of **0.72**, achieving exceptional sensitivity (**97% recall**) in identifying at-risk cases while maintaining acceptable overall performance metrics.





Rakamin Academy



Feature Importance

Performance & Engagement

The model shows senior employees with low performance have the highest attrition risk, followed by satisfaction and engagement factors

Talent Mismatch

Qualification misalignment (overqualified) and engagement-satisfaction gaps are strong turnover predictors

Hiring & Onboarding

Recruitment platforms and hiring timing (quarter) significantly impact employee retention

Strategic Recommendations

Revamp Flexibility & Working Hours

- Implement hybrid/remote work for technical roles
- Conduct workload audit, especially for roles with frequent overtime

Strengthen Work Culture & Team Environment

- Launch cultural intervention for Data team: anonymous surveys, FGDs
- Create safe, anonymous feedback channels

Create Career Clarity & Growth Opportunities

- Develop & communicate clear career paths for each role
- Launch mentorship & skill development programs

Use Predictive Model for Proactive Intervention

Rakamin Academy

Thank You

Source Code

LinkedIn

afsilmis@gmail.com