



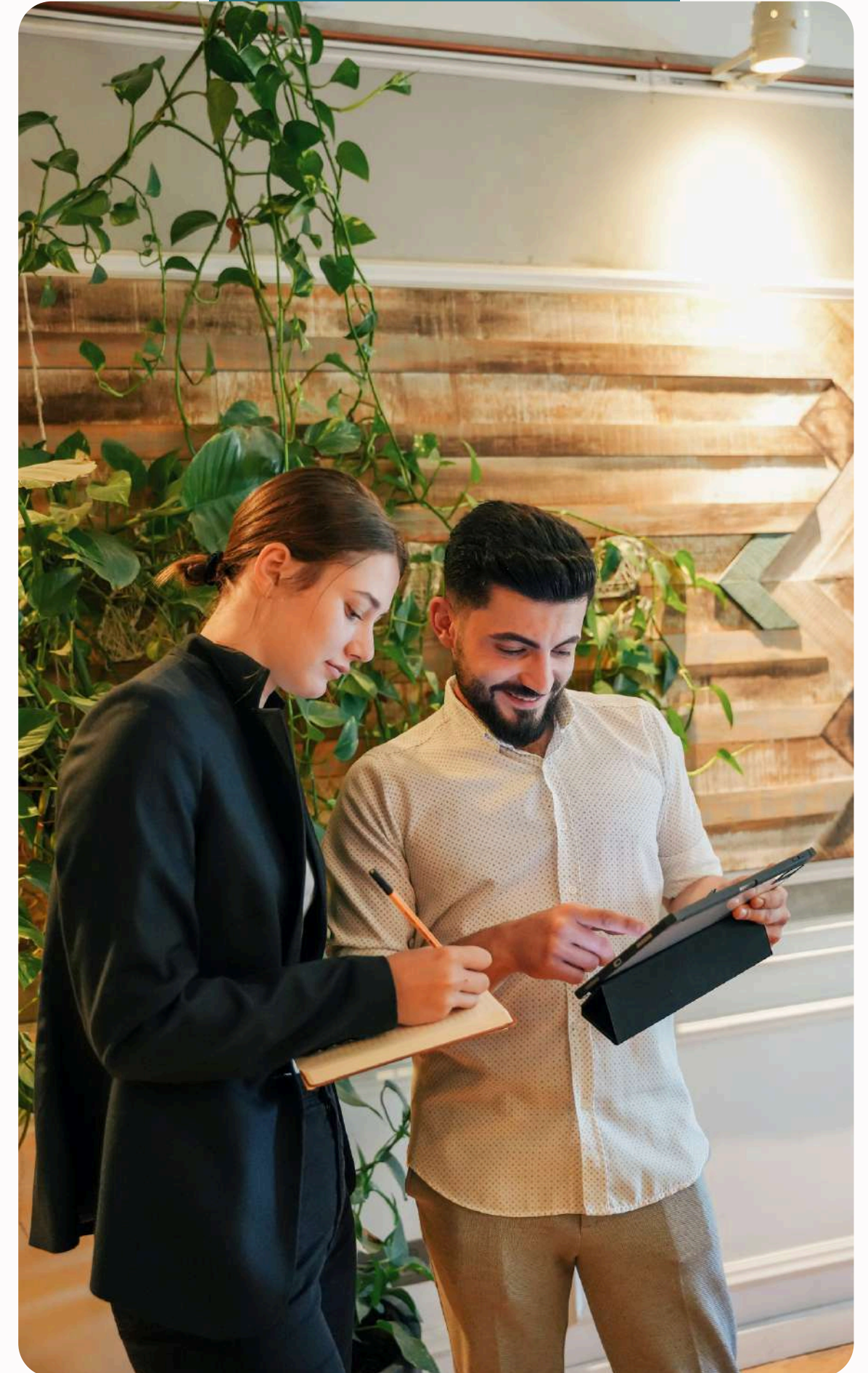
Tenera

Resign Risk Alert System

Dataset : HR Analytics: Job Change of Data Scientists

Mentor : Dedi Irwanto Samosir

Bernadetta Quinta | Santa Clara Krisfina Dohude | Phaedra Aldimo |
Yonatan Yusak Lestari | Nur Abdillah Al Fatih



OUR TEAM



Santa Clara
Krisfina D.

Project Manager



Bernadetta
Quinta

Data Analyst



Phaedra Aldimo

Data Engineer



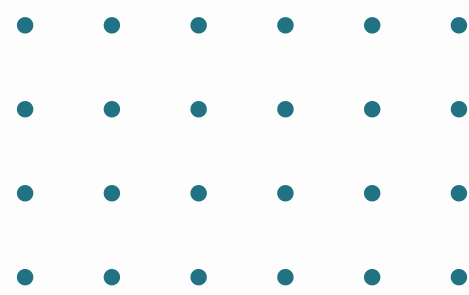
Nur Abdillah Al
Fatih

Data Scientist



Yonatan Yusak
Lestari

Data Scientist



AGENDA

01

Overview

02

EDA & Insight

03

Data Pre-Processing

04

Modelling

05

Recommendation and
deployment

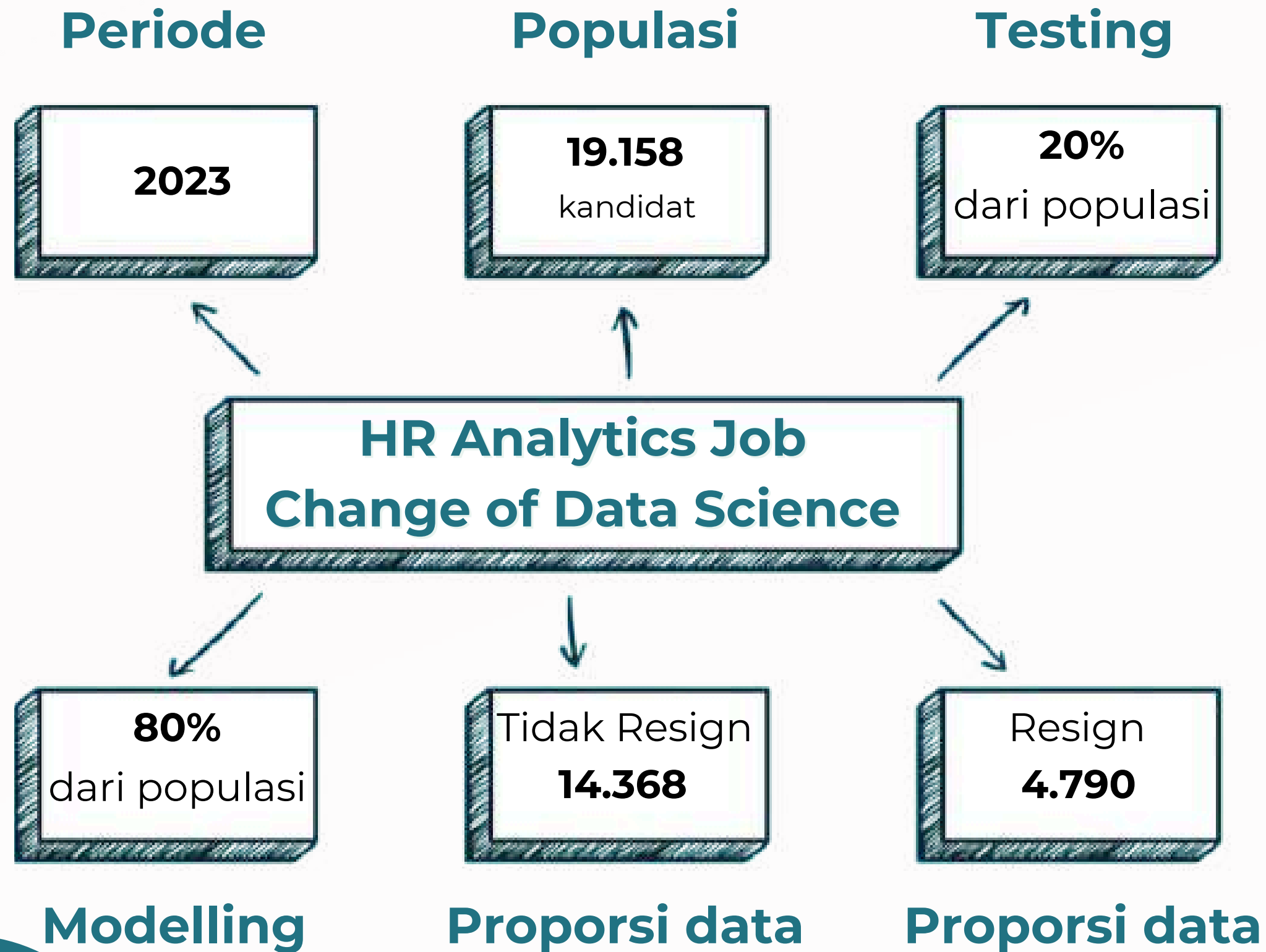


OVERVIEW

| 01



DATA OVERVIEW



LIST OF VARIABLE :

city_development_index

last_new_job

relevent_experience

experience

education_level

gender

city

major_discipline

enrolled_university

target

DATASET EXPLORATION

No	Variabel	Definisi
1	Enrollee_id	ID unik untuk setiap kandidat
2	City	Kota tempat kandidat tinggal
3	City_development_index	Alat ukur yang digunakan untuk menilai kemajuan dan kinerja pembangunan di suatu kota. CDI mencakup berbagai indikator, termasuk infrastruktur, ekonomi, sosial, dan lingkungan, untuk memberikan gambaran menyeluruh tentang kualitas hidup dan pembangunan kota.
4	Gender	Jenis kelamin
5	Relevant_experience	Apakah mempunyai pengalaman yang relevan dengan job yang baru
6	Enrolled_university	Status aktif di Universitas
7	Education_level	Jenjang pendidikan
8	Major_discipline	Jurusan atau bidang studi kandidat
9	Experience	Lama pengalaman bekerja (tahun)
10	Last_new_job	Berapa lama bertahan di pekerjaan terakhir sebelum pindah / berapa kali dimutasi ke pekerjaan lain sebelumnya
11	Training_hours	Berapa lama pelatihan yang dijalani (jam)
12	Target	Apakah kandidat berpindah kerja (1 = Ya, 0 = Tidak)

PROBLEM

‘Turnover’ is when employees leave a company, voluntarily or otherwise. **‘Turnover rate’**, meaning the percentage of employees who leave a company during a specific period of time.

Most leaders agree that a **turnover rate of about 10% is ideal.**
(www.wellhub.com)

PROBLEM

- Jumlah keluar: 4.790
- Karyawan awal : 19.158
- Karyawan akhir: 14.368

$$\text{Turnover Rate} = \frac{\text{Jumlah karyawan keluar}}{(\text{Awal tahun} + \text{Akhir tahun})/2} \times 100\%$$

Rata-rata jumlah karyawan:

$$= \frac{19.158 + 14.368}{2} = 16.763$$

Turnover Rate:

$$= \frac{4.790}{16.763} \times 100\% = 28,5\%$$

Jumlah angka turnover yang sangat tinggi yaitu sebesar 28.5% di atas angka turnover ideal yang berkisar antara 5 - 10 %

- Sumber: Rumus Turnover rate

GOALS & OBJECTIVES

Goals

Adanya profiling resiko probability karyawan resign atau tidak untuk mengoptimalkan pencegahan dan alokasi training & development cost

- Membangun model machine learning untuk memprediksi karyawan resign atau tidak
- Melakukan segmentasi resiko karyawan resign atau tidak untuk memberikan treatment yang tepat sesuai dengan tingkat resiko

Objectives

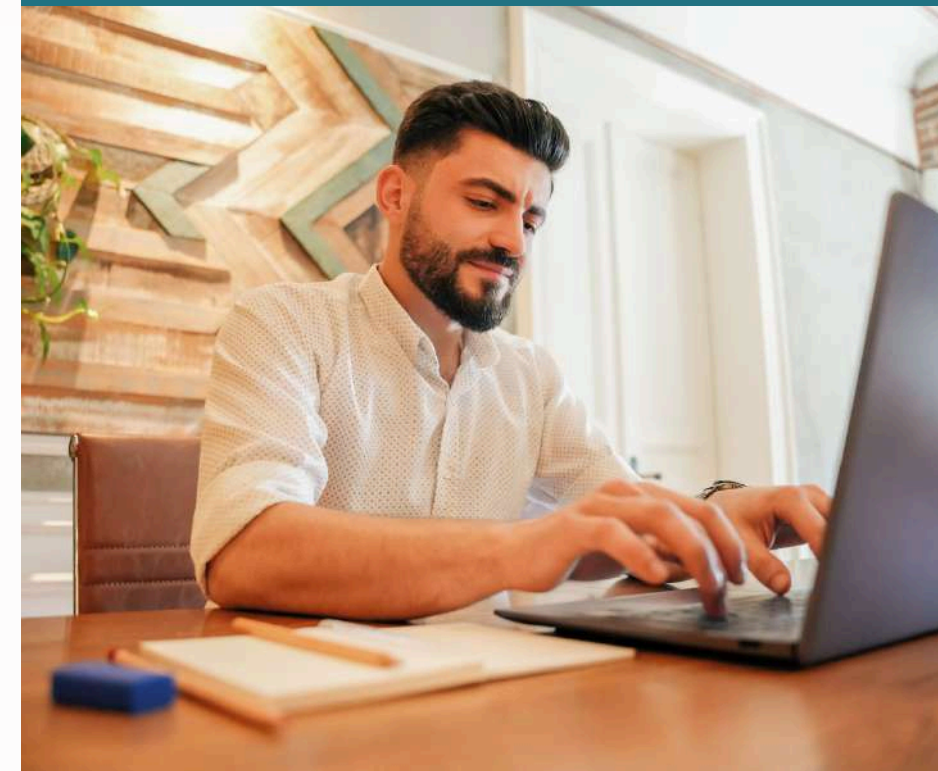
SUCCESS METRICS

TURN OVER RATE



Diharapkan mengurangi Turn Over karyawan di angka 5-10%. Source: *HR & Deloitte Reports & SHRM (Society for Human Resource Management)*

COST

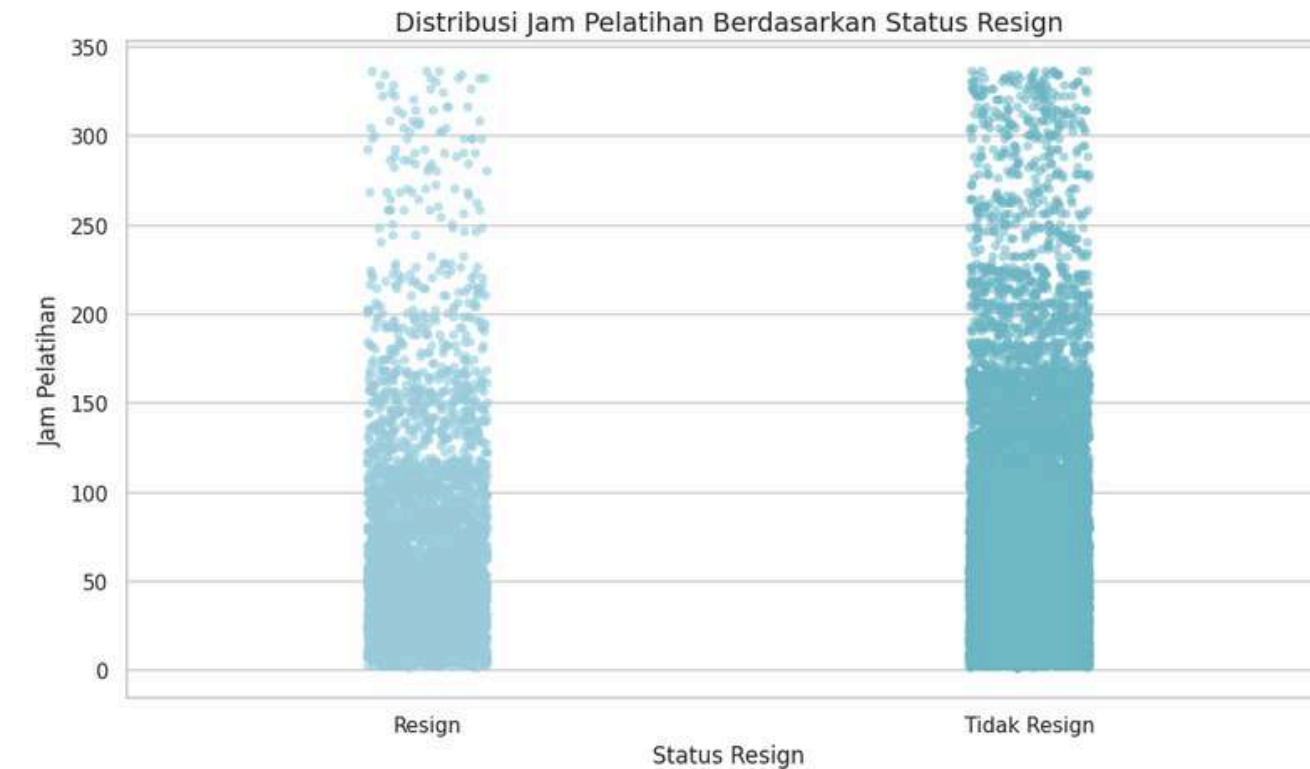
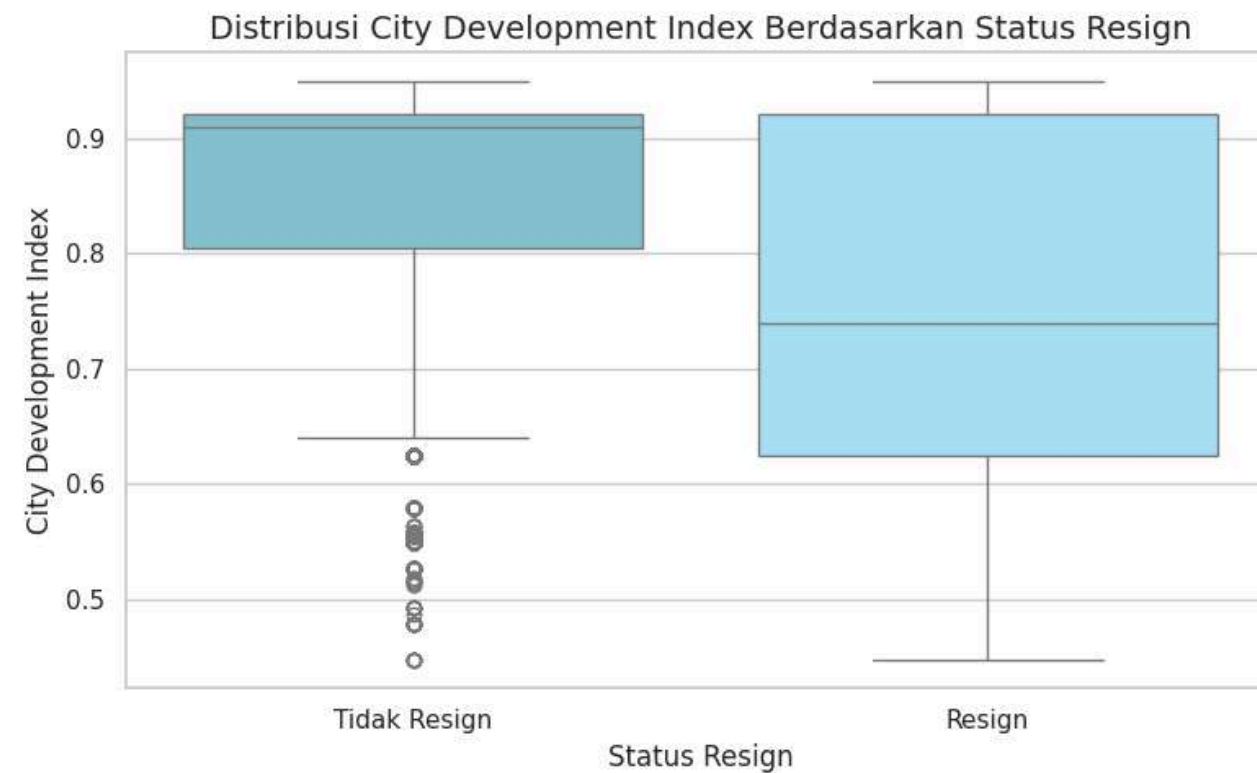


Alokasi Training & Development Cost tepat sasaran (50%: high, 35% medium, 15% low) Source: *Workramp LSM Report*

DATA UNDERSTANDING



EXPLORATORY DATA ANALYSIS



1. City Development Index

- Karyawan yang tidak resign cenderung berasal dari kota dengan City Development Index yang lebih tinggi (median mendekati 0.9).
- Sebaliknya, karyawan yang resign memiliki distribusi yang lebih luas dan median lebih rendah (sekitar 0.75).
- Tingkat pembangunan kota berpengaruh terhadap keputusan resign, semakin berkembang kotanya, karyawan cenderung bertahan.

2. Training Hours

- karyawan dengan jam pelatihan lebih tinggi cenderung tidak resign.

MULTIVARIATE ANALYSIS

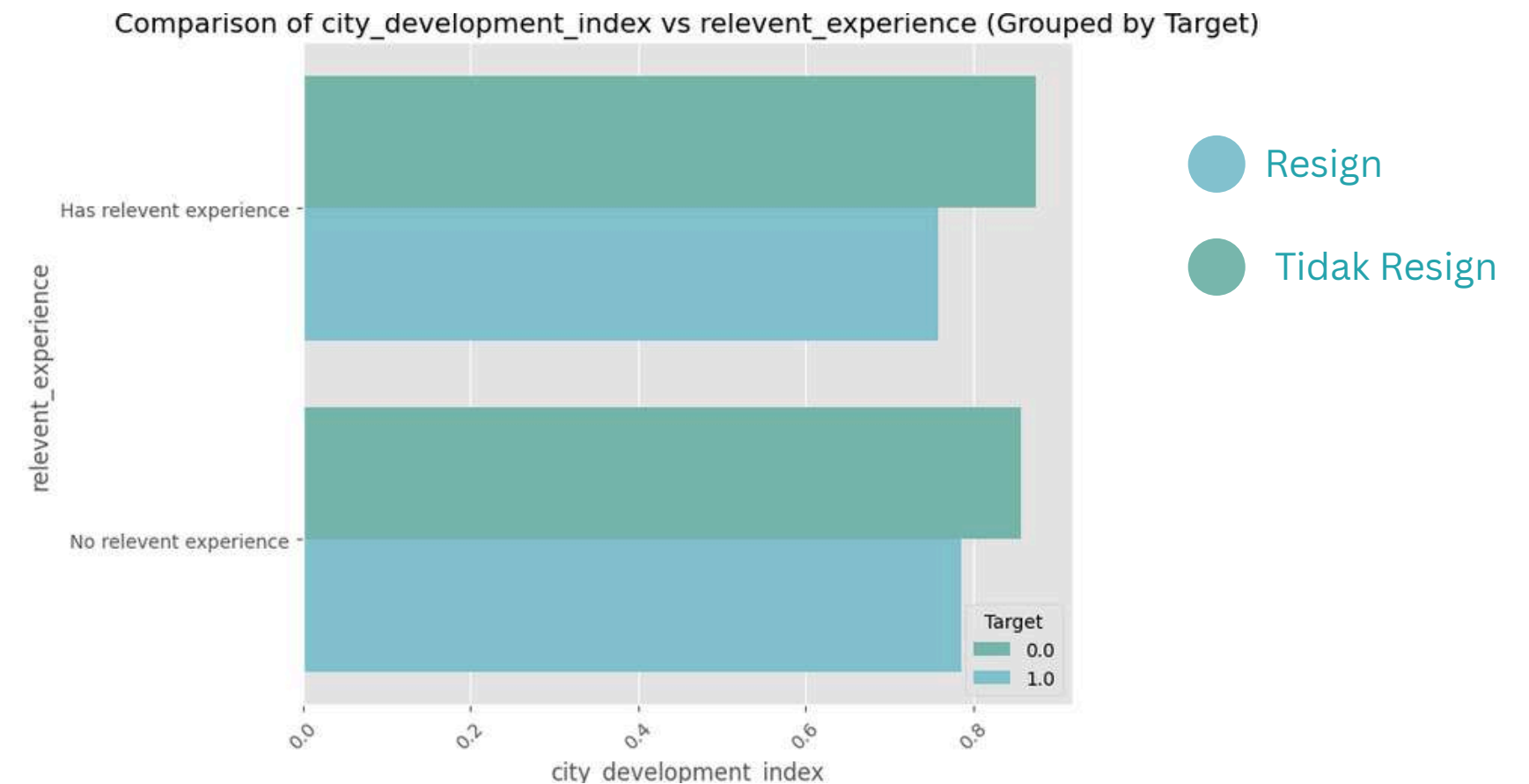
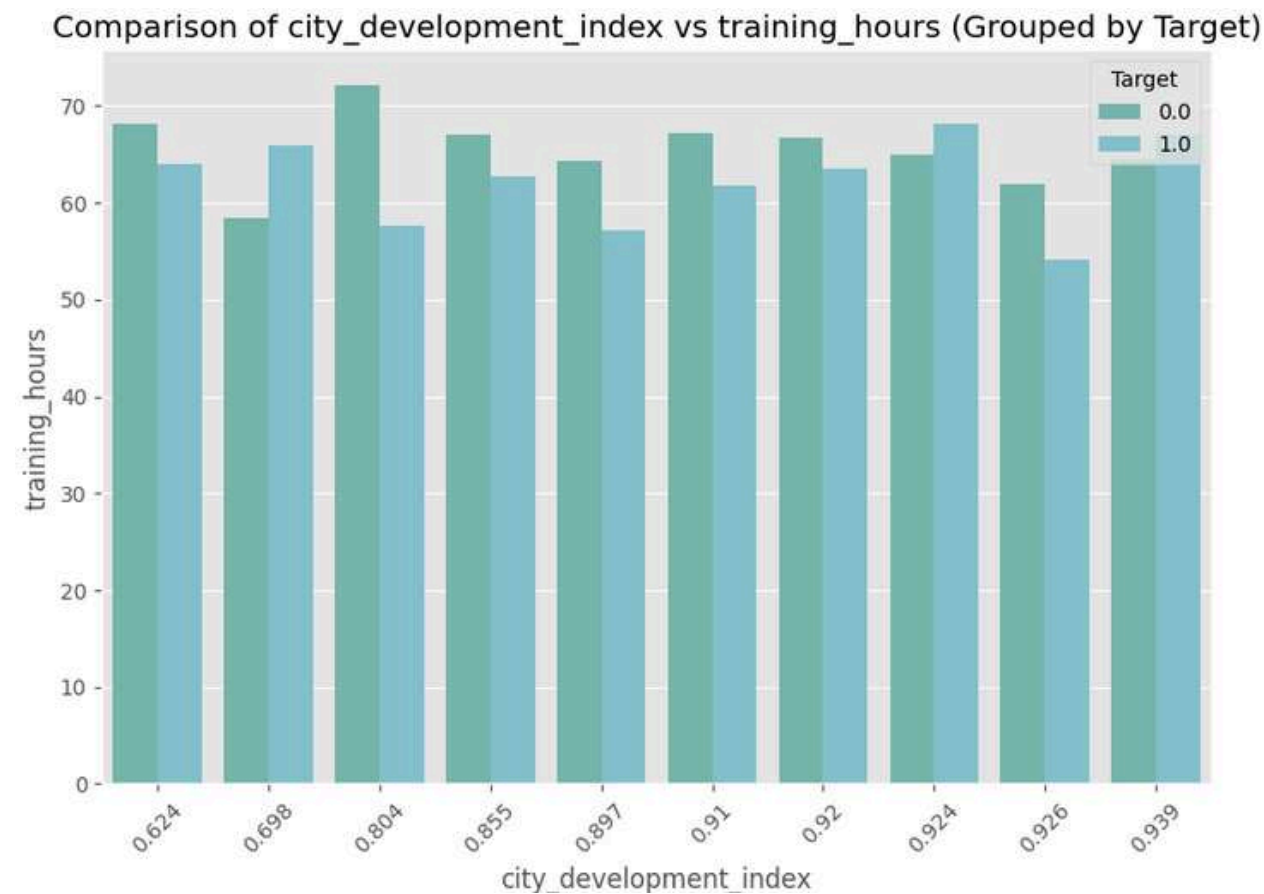
HEATMAP CORRELATION



Insight :

1. City Development Index adalah fitur paling informatif dalam memprediksi kemungkinan resign (berkorelasi negatif dengan target).
2. Fitur lain seperti training_hours, experience, dan relevent_experience memiliki pengaruh sangat kecil terhadap keputusan resign dari sudut pandang korelasi linier

MULTIVARIATE ANALYSIS



Comparison of City Development Index vs Training Hours (Grouped by Target)

- Pada kota-kota dengan indeks pembangunan yang lebih tinggi, jumlah training hours cenderung lebih tinggi. Namun, distribusi antara individu yang memutuskan resign (1) dan yang tetap bertahan (0) cukup seimbang. Artinya, indeks pembangunan kota mungkin berpengaruh terhadap jumlah pelatihan yang diikuti, tetapi tidak secara langsung menentukan apakah seseorang akan resign atau tidak.

Comparison of City Development Index vs Relevant Experience (Grouped by Target)

- Individu dengan pengalaman relevan cenderung berasal dari kota dengan indeks pembangunan yang lebih tinggi. Ini menunjukkan bahwa di kota-kota yang lebih maju, orang lebih mungkin memiliki pengalaman relevan. Namun, perbedaan antara mereka yang memutuskan resign (1) dan yang tetap bertahan (0) cukup kecil, yang menunjukkan bahwa memiliki pengalaman relevan tidak secara langsung mempengaruhi seseorang untuk resign atau tidak.

DATA PRE-PROCESSING



PRE-PROCESSING

Data Cleaning

- Handle missing values pada data kategorikal dengan menggunakan **modus**.
- Handle outliers pada 'city_development_index' dengan metode **IQR** karena terdapat data yang skewed.

Feature Encoding

- Mengubah **data kategorikal** menjadi **data numerikal** terutama inconsistent entries pada kolom **last_new_job**

Feature Selection

- Drop **company size**, **company type** dan
- Memilih fitur relevan: **city_development_index**, **training_hours**, **target**, **relevent_experience**, **experience**

Feature Extraction

- Menambah variable baru yaitu **estimated_salary** dan dipetakan berdasarkan tingkat pendidikan dan disesuaikan menurut jumlah tahun pengalaman kerja.
- **Gaji menjadi salah satu faktor utama** dalam keputusan resign:
 1. Ketidakpuasan terhadap gaji (journal.yrpiiku.com)
 2. Tawaran gaji lebih tinggi dari perusahaan lain (qmfincial.com)

Standardization Data

- Menggunakan **StandardScaler** dari **scikit-learn** untuk menstandarkan fitur

Handle Imbalance

- Menggunakan **SMOTE** dari **imbalanced-learn** dan **class_weight** untuk **Logistic Regression** agar menyesuaikan imbalanced class weights.

FEATURE EXTRACTION

Mapping Gaji Berdasarkan Pendidikan dan Pengalaman Kerja (variabel estimated_salary)

Tingkat Pendidikan	Gaji Dasar (Rp)	0-1 Tahun	2-5 Tahun (+10%)	6-10 Tahun (+20%)	>10 Tahun (+30%)
Primary School	3.000.000	3.000.000	3.300.000	3.600.000	3.900.000
High School	5.000.000	5.000.000	5.500.000	6.000.000	6.500.000
Graduate (S1)	8.000.000	8.000.000	8.800.000	9.600.000	10.400.000
Masters (S2)	12.000.000	12.000.000	13.200.000	14.400.000	15.600.000
PhD (S3)	16.000.000	16.000.000	17.600.000	19.200.000	20.800.000

Sumber: [Digital Skola](#), [Katadata](#)

MODELLING



MODEL

Model	Train Recall	Test Recall	Train Accuracy	Test Accuracy	Train ROC-AUC	Test ROC-AUC
Decision Tree	0.88	0.50	0.89	0.69	0.97	0.63
Random Forest	0.86	0.53	0.87	0.72	0.94	0.72
Logistic Regression	0.64	0.61	0.69	0.69	0.74	0.71
XGBoost	0.71	0.59	0.76	0.74	0.84	0.75

Logistic Regression adalah model paling tepat untuk prediksi karyawan resign ini, karena :

- Test Recall tertinggi (0.61) → lebih banyak karyawan yang resign bisa terdeteksi.
- Tidak overfitting → gap antara train dan test relatif kecil.

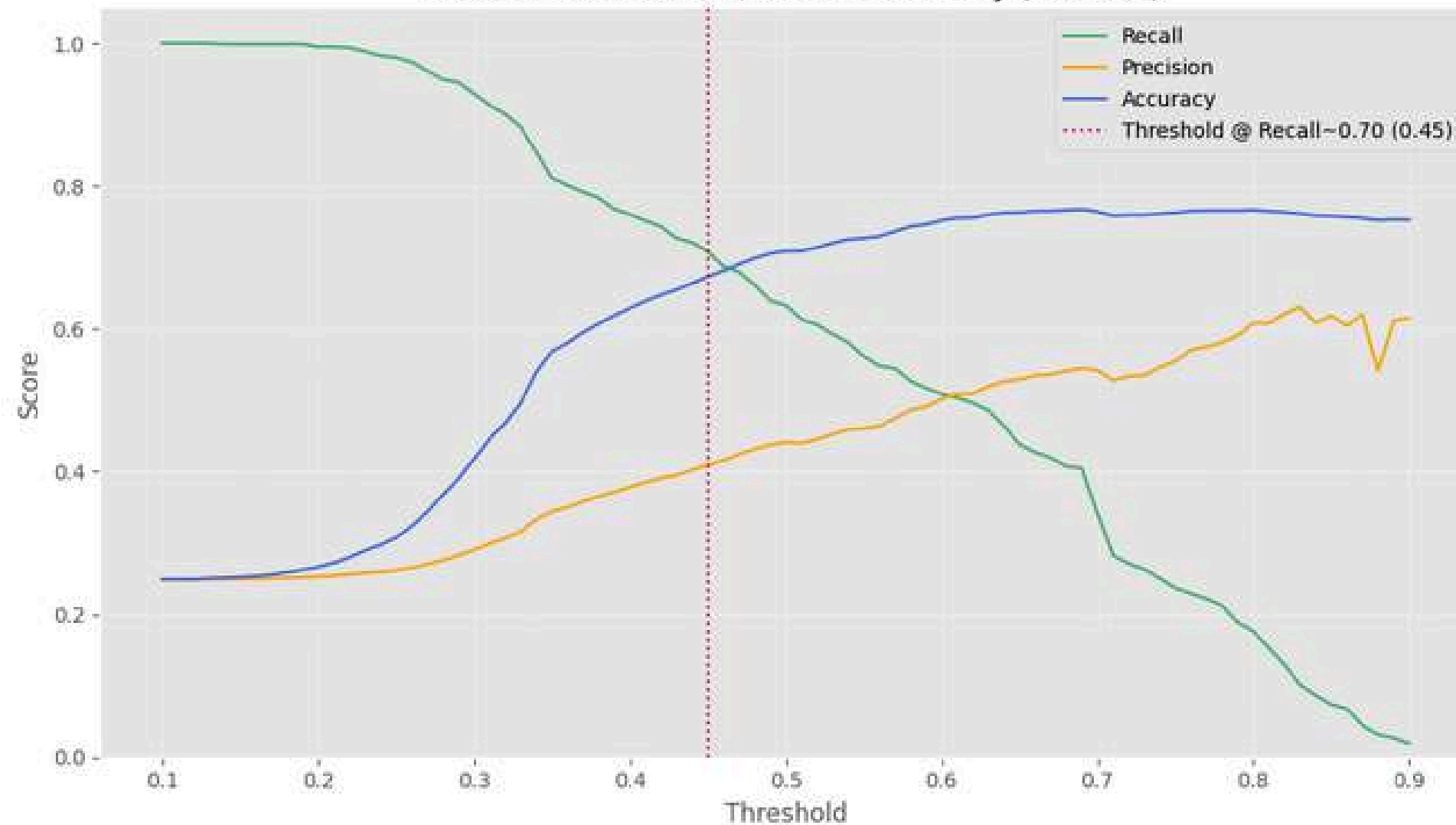
Evaluasi Model Lain :

- Random Forest & Decision Tree: Meskipun akurasi tinggi, tapi sangat overfitting dan recall rendah.
- XGBoost: Accuracy & ROC-AUC tinggi, tapi recall rendah, sehingga banyak orang resign tidak terdeteksi

OPTIMASI MODEL FINAL

Threshold : 0.45
Recall : 0.7086
Precision : 0.4894
Accuracy : 0.6738

Threshold vs Recall, Precision, Accuracy (Test Set)

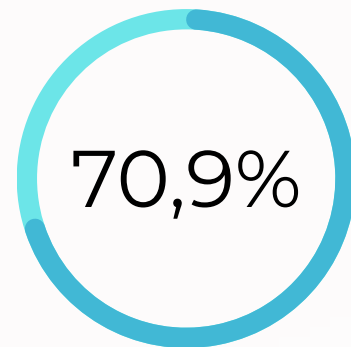


- Default model biasanya menganggap seseorang akan resign jika probabilitasnya ≥ 0.5 .
- Dengan threshold 0.45, maka:
- Recall meningkat, karena lebih banyak kasus resign yang terdeteksi
- Precision menurun, karena semakin banyak prediksi resign yang ternyata tidak benar.
- Nilai threshold ini akan di gunakan untuk mengsegmentasi resiko kedalam High Risk, Medium dan Low Risk

METRIKS MODEL OPTIMUM

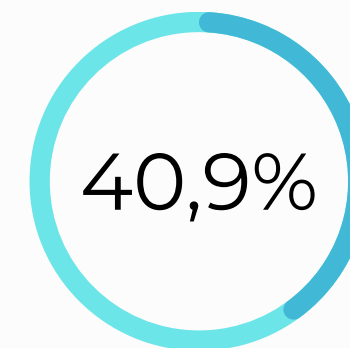
- Dengan threshold 0,45 dipilih karena memberikan trade-off terbaik antara recall yang tinggi dan akurasi serta precision yang masih layak.
- **Kesimpulan:** Model prediksi ini sudah cukup efektif dalam mengidentifikasi potensi resign, dengan recall yang baik (0.71). Meskipun precision masih tergolong rendah.

Recall



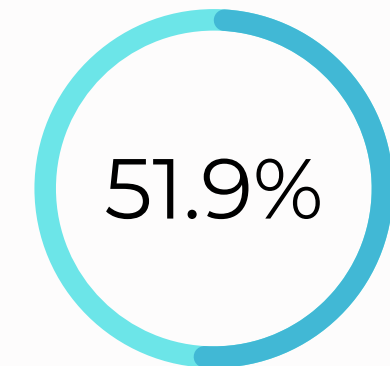
Jika terdapat 100 karyawan yang benar-benar resign, maka model dapat menangkap sekitar 71 di antaranya.

Precision



Jika dari hasil prediksi terdapat 100 karyawan yang diprediksi resign, maka hanya 41 di antaranya yang benar-benar resign.

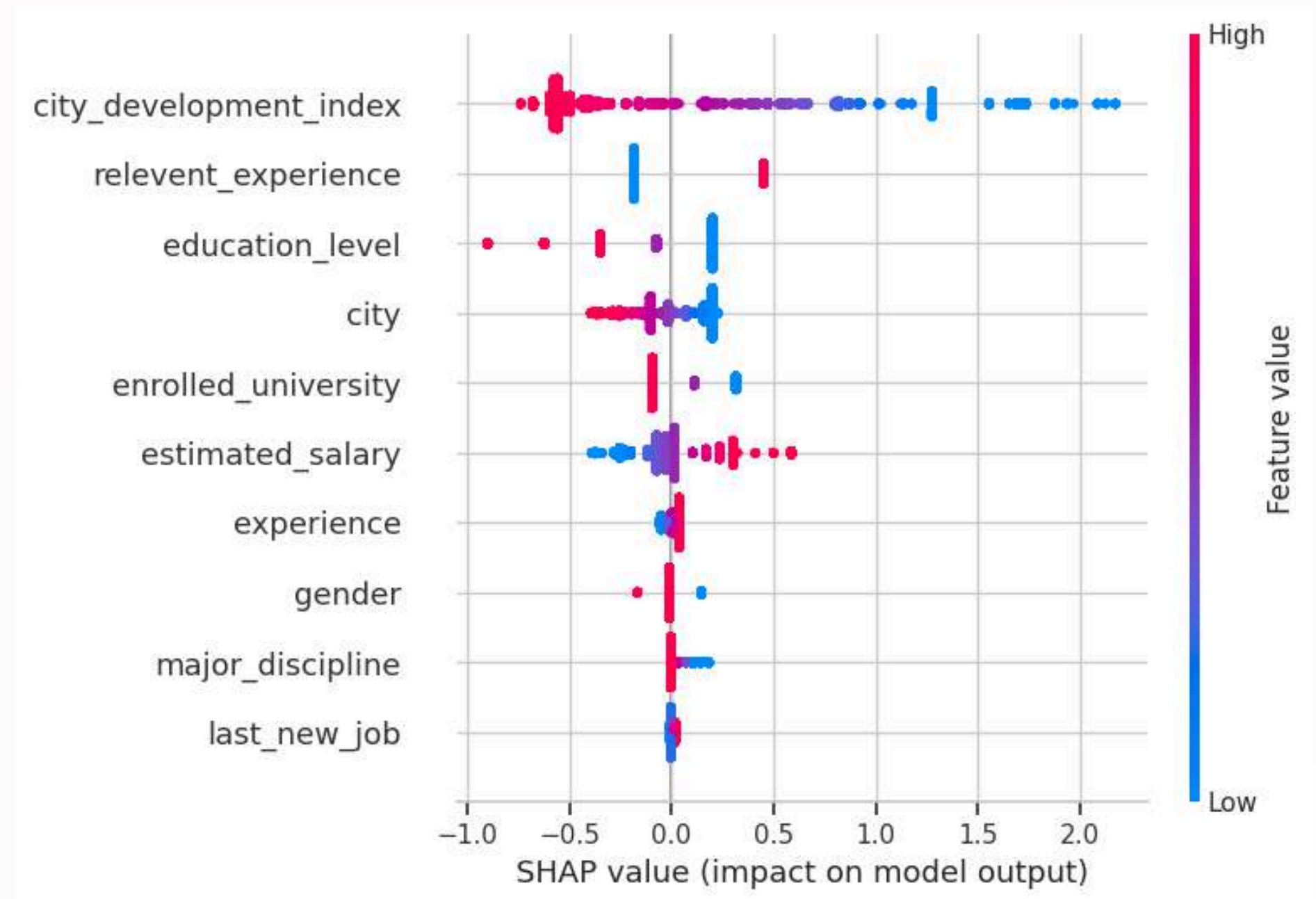
F1 Score



Jika mempertimbangkan keseimbangan antara recall dan precision, model memiliki performa sedang dalam mendeteksi karyawan yang resign.

FEATURE IMPORTANCE

City Development Index menjadi faktor paling berpengaruh terhadap prediksi resign. Nilai indeks yang lebih rendah cenderung meningkatkan risiko resign karyawan.



SEGMENTASI

Segmen Risiko	Threshold
High-Risk	0.70
Medium-Risk	0.45 – 0.70
Low-Risk	<0.45

- High-Risk (Probabilitas ≥ 0.70):
 - Sangat berisiko untuk resign, prioritas utama untuk intervensi.
- Medium-Risk (Probabilitas antara 0.45 – 0.70):
 - Cukup berisiko, perlu pemantauan dan perhatian ekstra.
- Low-Risk (Probabilitas < 0.45):
 - Risiko resign rendah, namun tetap perlu dipertahankan motivasinya.

BUSINESS IMPACT

🔍 ANALISIS TURNOVER

- Total Karyawan: 19.158
- Estimasi Karyawan Resign (28.5% attrition): 4790 orang
- Recall Model (71%): Berhasil mengidentifikasi 3401 dari 4790 orang yang benar-benar akan resign.

✅ ANALISIS PREVENTED RESIGN

- Rasio Averted vs Expected Resign:

$$\frac{3401}{4790} = 0,71 \Rightarrow 71,0\%$$

- Interpretasi:
Sekitar 71% dari karyawan yang diprediksi akan resign berhasil dipertahankan melalui intervensi.

Keterangan	Nilai Sebelum	Nilai Setelah
Estimasi Resign	4790 orang	1389 orang
Turnover Rate	28,5%	7,5%

BUSINESS IMPACT

TOTAL T&D RP. 1 000 000 000 / YEARS

Segmen Risiko	Sebelum (Equal 33,3%)	Sesudah (Target)	Δ (%)	Δ (Rp)
High-Risk	33,3% → Rp 333.333.333	50% → Rp 500.000.000	+16,7 pp	+Rp 166.666.667
Medium-Risk	33,3% → Rp 333.333.333	35% → Rp 350.000.000	+1,7 pp	+Rp 16.666.667
Low-Risk	33,3% → Rp 333.333.333	15% → Rp 150.000.000	-18,3 pp	-Rp 183.333.333
Total	100% → Rp 1.000.000.000	100% → Rp 1.000.000.000	0 pp	Rp 0

🔍 Insight:

- High-Risk mendapat tambahan alokasi signifikan (+16,7 pp) untuk memfokuskan upaya preventif.
- Medium-Risk sedikit bertambah (+1,7 pp) agar tetap dijaga momentum pengembangannya.
- Low-Risk dipangkas (-18,3 pp) karena turnover mereka relatif rendah dan kebutuhan training reguler sudah mencukupi.
- Dengan redistribusi ini, proporsi anggaran T&D kini lebih mencerminkan prioritas retensi memaksimalkan dukungan pada karyawan yang paling berisiko resign.
- Segmentasi alokasi anggaran berdasarkan pendekatan risk-based prioritization (Workramp)

BUSINESS IMPACT



SEGMENTASI RISIKO

Segmentasi Prediksi Resign berdasarkan Threshold
Menjembatani trade-off antara recall 70% vs precision 40%, agar:

- Risiko resign tetap bisa dideteksi secara luas (via recall)
- Tapi tindakan HR lebih akurat (via segmentasi scoring)

★ DAMPAK BISNIS (SEBELUM SEGMENTASI)

Aspek	Dampak
Precision 40%	60% tindakan HR salah sasaran
Biaya retensi boros	Banyak program diterapkan ke karyawan yang tidak resign
Overload tim HR	Terlalu banyak intervensi, potensi burnout tim
Potensi demotivasi	Karyawan yang tidak resign merasa dicurigai

✅ DAMPAK BISNIS (SETELAH SEGMENTASI)

Aspek	Dampak
Biaya lebih terkendali	Program difokuskan ke misalnya top 100 orang → lebih efisien
Fokus strategi retensi	Bisa membedakan prioritas tindakan
HR lebih strategis	Tidak reaktif, tetapi berbasis data dan level risiko

RECOMMENDATION AND DEPLOYMENT



RECOMMENDATION

BERDASAR SEGMENTASI RISIKO (Threshold)

HIGH RISK

TINDAKAN: Retensi Aktif

1.Budgeting Cost:

Menggunakan Alokasi Budgeting Cost 50%

2.Komunikasi 1-on-1 dengan Atasan:

Jika Counseling dirasa kurang cukup. Lakukan Interaksi langsung dengan manajer atau atasan guna memperkuat hubungan interpersonal. Karyawan lebih cenderung bertahan jika merasa didengar dan dianggap penting.

3.Bonus & Insentif:

Memberikan penghargaan finansial seperti bonus atau insentif berbasis performa dapat meningkatkan motivasi dan loyalitas karyawan, terutama mereka yang merasa kurang diapresiasi.

MEDIUM RISK

TINDAKAN: Engagement

1.Budgeting Cost:

Menggunakan Alokasi Budgeting Cost 35%

2.Survei Kepuasan & Engagement Berkala:

Gunakan model survei yang terpercaya seperti Gallup Q12. Pastikan hasil survei ditindaklanjuti dengan tindakan nyata agar karyawan merasa didengar dan dihargai.

3.Reminder untuk Manajer:

Sekitar 70% variasi engagement tim dipengaruhi oleh peran manajer, sehingga keterlibatan manajer sangat penting. Oleh karena itu, manajer wajib melakukan sesi coaching ringan setiap minggu guna memastikan anggota tim tahu ekspektasi kerja yang jelas, merasa didukung dalam pengembangan karier, dan merasa bahwa pekerjaannya bermakna serta memberikan kontribusi.

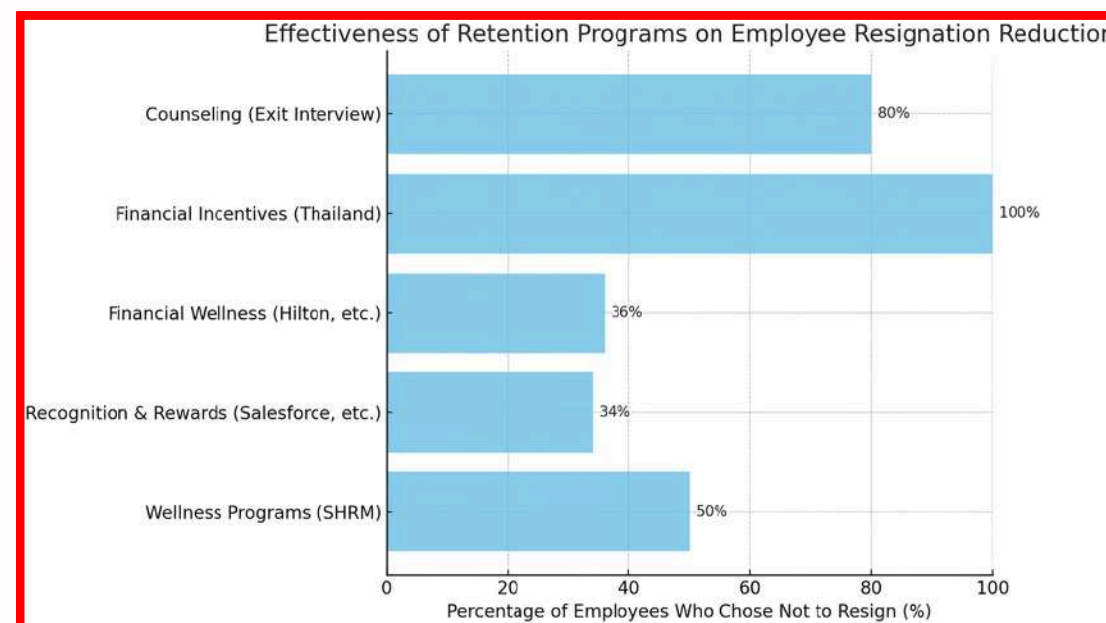
LOW RISK

TIDAK PERLU TINDAKAN KHUSUS

Tidak diperlukan tindakan khusus. Namun, disarankan untuk melakukan pemantauan secara berkala guna memastikan situasi tetap stabil dan mencegah potensi perkembangan kelompok ini menjadi ke tahap medium ataupun tinggi.

1.Budgeting Cost:

Menggunakan Alokasi Budgeting Cost 15%



Threshold : > 0.70

What Is Employee Engagement and How Do You Improve It?

01. Employee Engagement Definition
02. Why Is Employee Engagement Important?
03. Whose Job Is Employee Engagement?
04. What Are the Drivers of Employee Engagement?
05. Improving Employee Engagement Strategies
06. Measuring Employee Engagement: Gallup's Questions
07. The Employee Engagement Model
08. Employee Engagement Examples: The 3 Types of Employees You Have
09. How to Improve Employee Engagement: Team Engagement Ideas
10. Improving Employee Engagement Begins Here

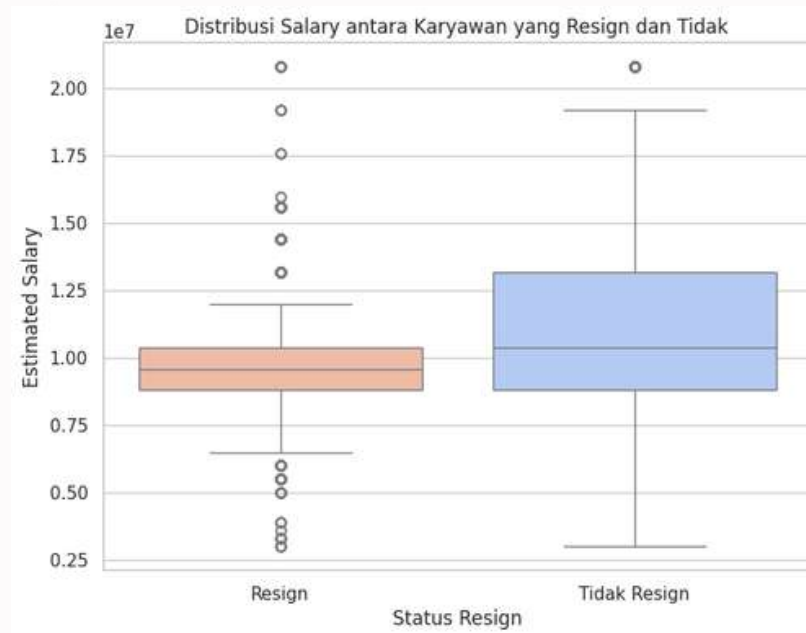
Source : <https://www.gallup.com/workplace/285674/improve-employee-engagement-workplace.aspx>

Threshold : 0.45 - 0.70

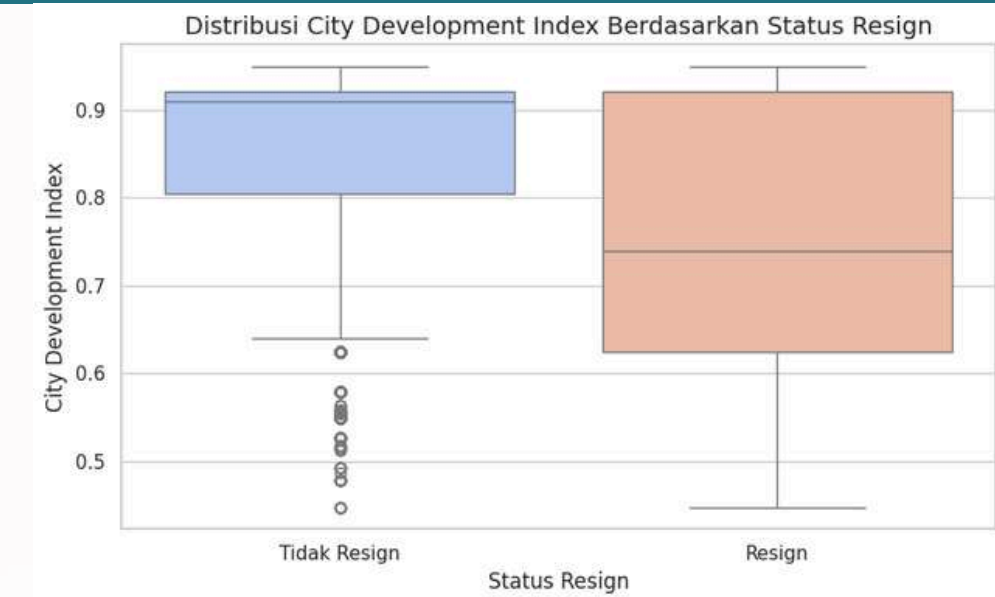


Threshold : < 0.45

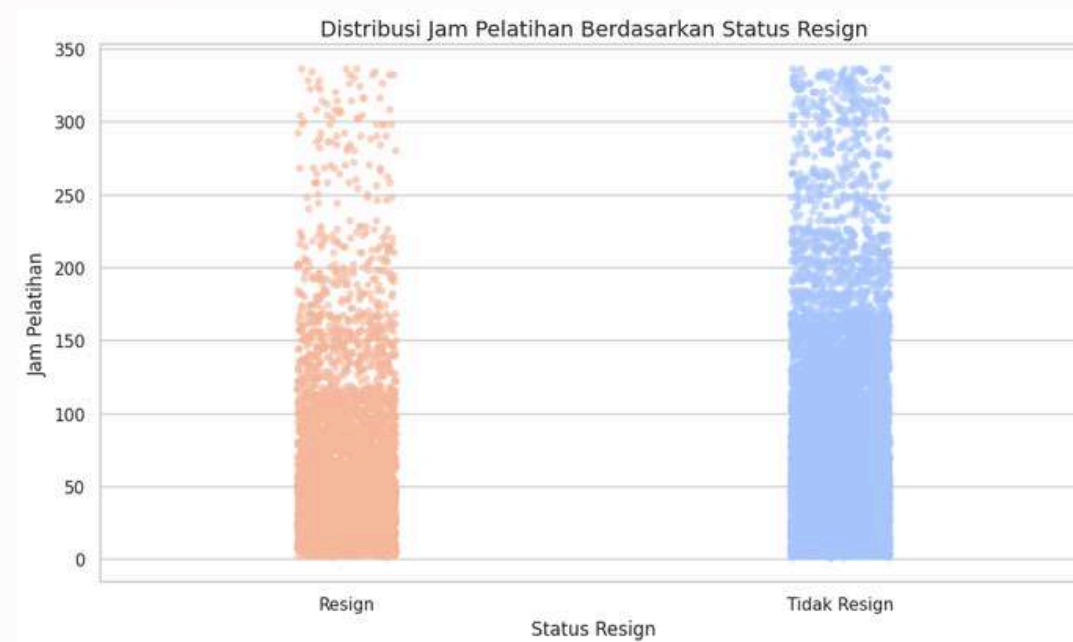
RECOMMENDATION



Perusahaan **disarankan meninjau ulang terhadap gaji**, karena **gaji rendah berkorelasi dengan risiko resign yang lebih tinggi**.



Pemebrian Benefit lebih bagi karyawan di kota **dengan indeks pembangunan yang rendah** untuk mengurangi risiko resign.



Perusahaan disarankan untuk **meningkatkan akses pelatihan bagi karyawan**, karena data menunjukkan bahwa **karyawan dengan jam pelatihan lebih tinggi cenderung tidak resign**

RECOMMENDATION

ENHANCEMENT MODEL (ACCURACY)

Melakukan penambahan variabel seperti :



Level Jabatan:

Mengetahui yang banyak resign itu jabatannya apa, sehingga bisa lebih spesifik memberikan solusi sesuai dengan jabatan



Alasan Resign:

Menanyakan alasan resign untuk mendapatkan insight resign secara detail



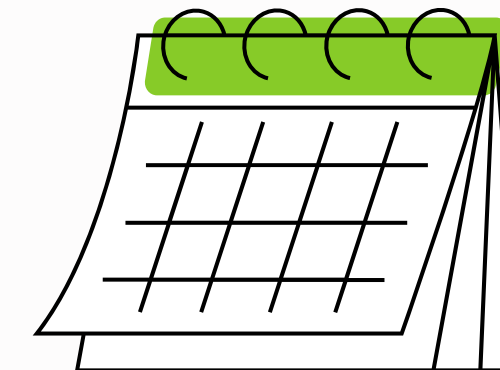
Jarak Kantor:

Mengetahui apakah jarak ke kantor yang terlalu jauh menjadi alasan resign



Gaji:

Mengetahui apakah gaji rendah menjadi alasan karyawan resign



Tanggal Resign:

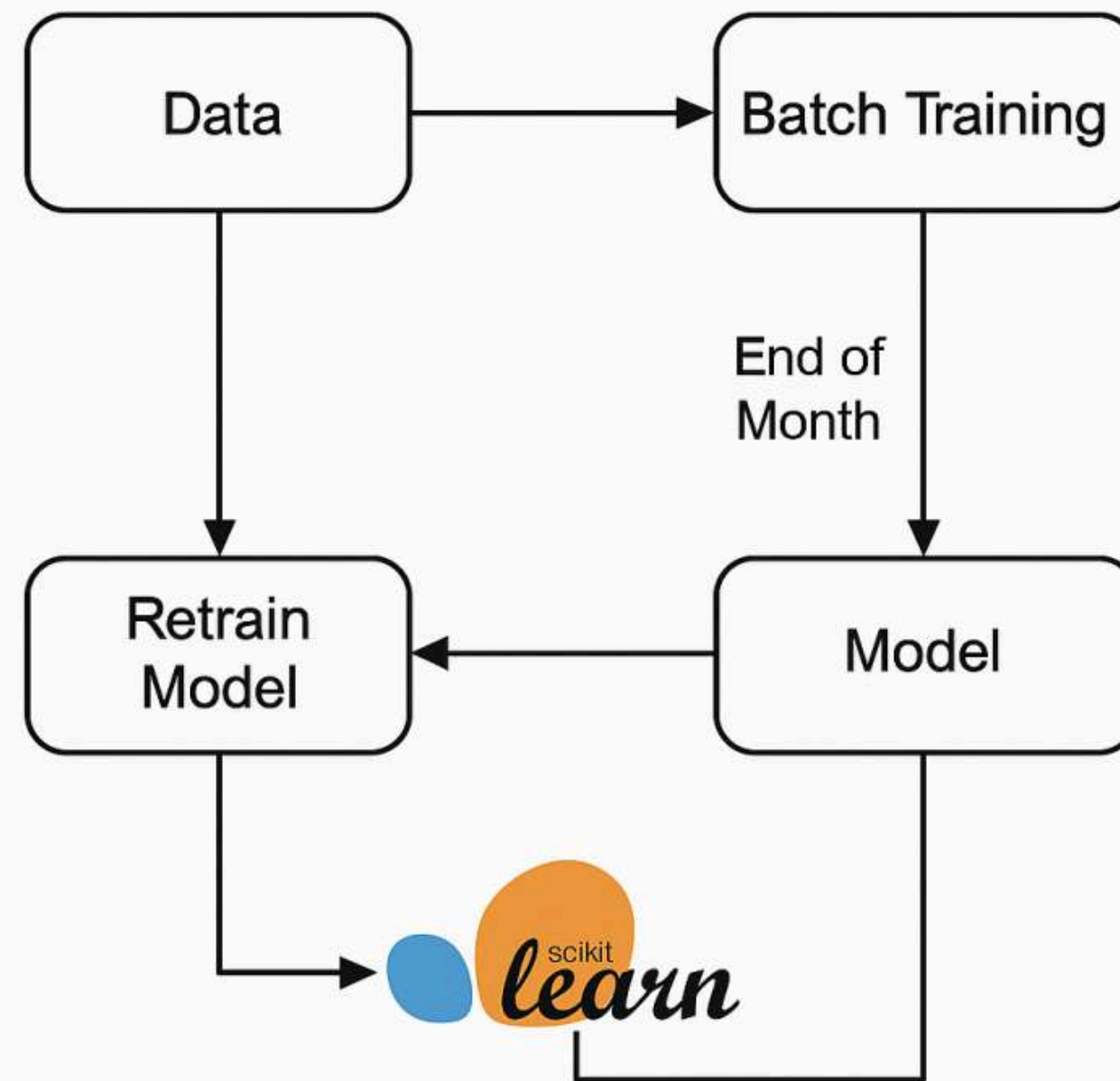
Bisa untuk melihat pola/prediksi di periode mana saja resign sering terjadi

DEPLOYMENT

Menggunakan model terbaik untuk deployment


Menggunakan Streamlit Cloud dengan integrasi Github

Retraining Flow



DEPLOYMENT

 **Prediksi Kemungkinan Karyawan Resign**

 **Input Data Karyawan** 

Kota

city_117

▼

City Development Index (CDI)

0.698

Jenis Kelamin

Male

▼

Tingkat Pendidikan

Primary School

▼

Jurusan

STEM

▼

Status Universitas

no_enrollment

▼

Pengalaman Kerja (tahun)

<1

▼

Pengalaman Relevan

Has relevent experience

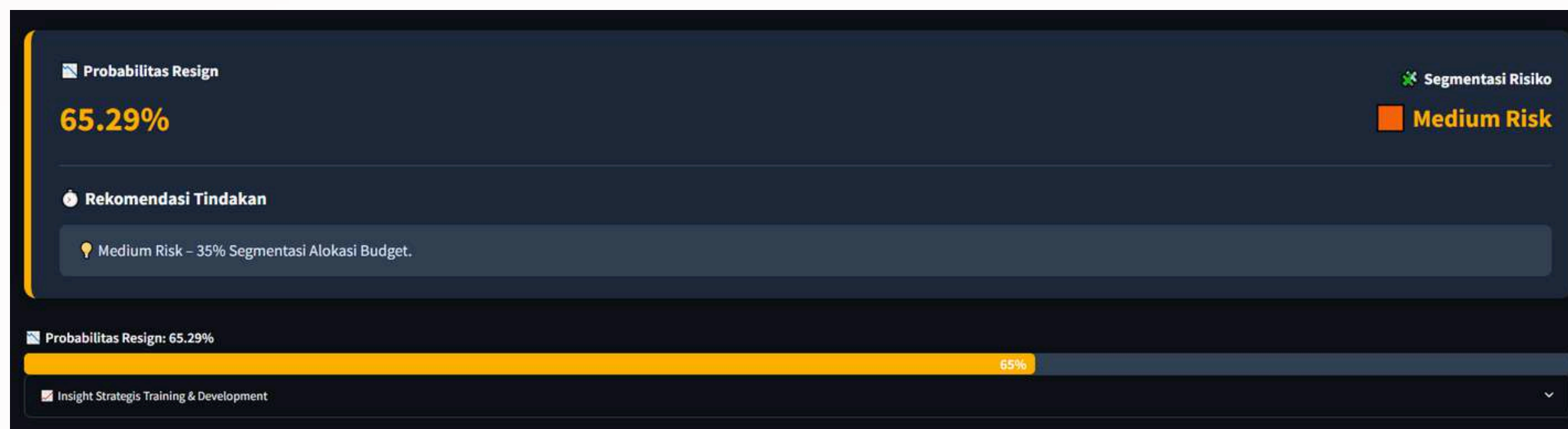
▼

Terakhir Ganti Pekerjaan

0

▼

[LINK
DEPLOYMENT](#)





THANK YOU

