

LAPORAN AKHIR
FINAL PROJECT
Resign Risk Alert System

Diajukan untuk memenuhi persyaratan kelulusan
Bootcamp Data Science Rakamin

Disusun oleh :
Bernadetta Quinta
Santa Clara Krisfina Dohude
Phaedra Aldimo
Yonatan Yusak Lestari
Nur Abdillah Al Fatih



Tenera
TAHUN 2025

Abstraksi

Proyek ini bertujuan untuk mengembangkan sistem prediksi risiko resign karyawan berbasis machine learning guna membantu perusahaan mengurangi angka turnover yang tinggi, yakni sebesar 28.5%, jauh di atas batas ideal 5–10%. Studi ini menggunakan dataset HR Analytics: Job Change of Data Scientists untuk menganalisis faktor-faktor yang mempengaruhi keputusan karyawan dalam berpindah kerja.

Proses analisis dimulai dengan eksplorasi data, penanganan missing values dan outlier, serta pembuatan fitur baru seperti `estimated_salary`. Data kemudian diproses melalui encoding, standardisasi, dan penyeimbangan kelas menggunakan SMOTE. Beberapa model diterapkan, dan Logistic Regression dipilih sebagai model terbaik karena memiliki recall tertinggi sebesar 61,3%, sejalan dengan tujuan utama yaitu mendeteksi sebanyak mungkin karyawan yang berpotensi resign (minimasi false negative). Lalu dimaksimalkan dengan optimasi penurunan threshold menjadi 0.45 yang mengakibatkan recall menjadi 70.9%.

Selain prediksi, dilakukan segmentasi berdasarkan threshold yang telah dibuat untuk merancang strategi retensi yang tepat. Dengan adanya segmentasi maka memaksimalkan intervensi difokuskan pada karyawan berisiko sesuai prioritasnya dibandingkan dengan strategi tanpa segmentasi yang kurang efektif dalam mencegah turnover rate.

Rekomendasi meliputi pemberian pelatihan yang disesuaikan, evaluasi gaji minimum berdasarkan prediksi, serta relokasi ke kota dengan indeks pembangunan tinggi. Model di-deploy menggunakan Streamlit Cloud dan GitHub sebagai integrasi teknologi.

Kata Pengantar

Puji syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya, sehingga Final Project Bootcamp Data Science Rakamin Academy dengan judul “Resign Risk Alert System” dapat diselesaikan dengan baik dan tepat waktu.

Proyek ini merupakan implementasi dari materi yang telah kami pelajari, mulai dari EDA, pemodelan machine learning, hingga deployment. Penyusunan proyek ini tidak lepas dari dukungan berbagai pihak, oleh karena itu kami sampaikan terima kasih kepada:

1. Rakamin Academy, yang telah menyediakan program pembelajaran intensif dan terstruktur, serta memberikan kesempatan kepada kami untuk mengasah kemampuan di bidang data science melalui proyek nyata.
2. Bapak Dedi Irwanto Samosir, selaku mentor, atas bimbingan, arahan, dan masukan yang sangat berharga selama proses pengerjaan proyek ini.
3. Seluruh rekan satu tim, yaitu Bernadetta Quinta, Santa Clara Krisfina Dohude, Phaedra Aldimo, Yonatan Yusak Lestari, dan Nur Abdillah Al Fatih, atas kerja sama, semangat kolaboratif, dan dedikasi yang luar biasa selama menyelesaikan tugas ini bersama.
4. Instruktur dan tim pengajar Rakamin lainnya, yang telah membagikan ilmu, pengalaman, dan wawasan yang sangat mendalam selama program berlangsung.

Kami menyadari bahwa proyek ini masih memiliki keterbatasan, oleh karena itu kami terbuka terhadap saran dan masukan yang membangun. Besar harapan kami agar hasil proyek ini dapat memberikan manfaat, baik dalam pembelajaran maupun sebagai kontribusi nyata di dunia kerja profesional.

Jakarta, Mei 2025

Hormat kami,

Tim Final Project – Tenera

Daftar Isi

I.	Pendahuluan	1
1.1	Latar belakang.....	1
1.2	Lingkup	1
1.3	Tujuan	2
II.	Kelompok Final Project Bootcamp Rakamin	4
2.1	Struktur Organisasi	4
2.2	Lingkup Pekerjaan	4
2.3	Deskripsi Pekerjaan.....	6
2.4	Jadwal Kerja.....	7
III.	Prediksi Resign Karyawan Menggunakan Machine Learning.....	9
3.1	Deskripsi Permasalahan Proyek.....	9
3.2	Proses Pelaksanaan Proyek	9
3.3	Pencapaian Hasil Proyek.....	13
IV.	Penutup.....	22
4.1	Kesimpulan	22
4.2	Saran.....	22
V.	Referensi	24
VI.	Lampiran A. PPT Final Project Presentation	25
VII.	Lampiran B. Notulensi Mentoring	26
VIII.	Lampiran C. Dokumen Teknik	29

DAFTAR GAMBAR

Gambar 2. 1 Struktur Organisasi	4
Gambar 2. 2 Timeline Proyek.....	7
Gambar 3. 1 Analisis Univariat & Multivariat	10
Gambar 3. 2 Heatmap Korelasi	11
Gambar 3. 3 Alur Pre-Processing	12
Gambar 3. 4 Model Optimum.....	14
Gambar 3. 5 Shap Value.....	16
Gambar 3. 6 Tampilan Streamlit Deployment.....	20
Gambar 3. 7 Alur Deployment	21

DAFTAR TABEL

Tabel 2. 1 Peran, Tugas, dan Kontribusi Anggota Tim Proyek	6
Tabel 3. 1 Mapping Fitur Estimasi_Salary.....	13
Tabel 3. 2 Perbandingan Hasil Evaluasi Antar Model	14
Tabel 3. 3 Kategori Segmentasi Risiko Resign.....	16
Tabel 3. 4 Estimasi Turnover Rate Sebelum dan Setelah Intervensi	17
Tabel 3. 5 Redistribusi Anggaran T&D Berdasarkan Risiko Resign.....	18
Tabel 3. 6 Perbandingan Dampak Sebelum dan Setelah Segmentasi	18

I. Pendahuluan

1.1 Latar belakang

Tingkat *turnover* karyawan merupakan salah satu indikator penting dalam menilai stabilitas dan efektivitas suatu organisasi, terutama pada bidang yang berkembang pesat seperti data science. Karyawan yang keluar dari perusahaan, baik secara sukarela maupun tidak, dapat menyebabkan terganggunya alur kerja, meningkatnya beban kerja pada tim yang tersisa, serta bertambahnya biaya rekrutmen dan pelatihan. Oleh karena itu, mengukur dan mengelola *turnover rate* menjadi aspek krusial dalam manajemen sumber daya manusia.

Pada data tahun 2023, diperoleh bahwa dari total 19.158 kandidat, sebanyak 4.790 di antaranya memilih untuk resign. Dengan menghitung rata-rata jumlah karyawan antara awal dan akhir periode (yaitu sebesar 16.763), maka diperoleh tingkat *turnover* sebesar 28,5%. Nilai ini dihitung menggunakan rumus standar yaitu:

$$Turnover\ rate = \left(\frac{Jumlah\ Karyawan\ Keluar}{(Karyawan\ awal + Karyawan\ Akhir)/2} \right) 100\% \quad (1.1)$$

Angka tersebut jauh melebihi ambang batas ideal *turnover rate* yang direkomendasikan oleh para pakar, yaitu antara 5% hingga 10% [1]. Tingginya angka *turnover* ini menunjukkan adanya masalah dalam retensi karyawan yang perlu segera diidentifikasi lebih lanjut. Oleh sebab itu, pendekatan berbasis data sangat diperlukan untuk memahami faktor-faktor utama yang mempengaruhi keputusan resign karyawan dan merumuskan strategi retensi yang lebih efektif dan tepat sasaran.

1.2 Lingkup

Lingkup proyek ini mencakup seluruh proses analisis prediktif untuk mengidentifikasi faktor-faktor yang mempengaruhi keputusan resign karyawan, serta mengklasifikasikan risiko resign berdasarkan data kandidat di bidang data science. Tahapan yang dilakukan dalam proyek ini meliputi:

- Eksplorasi dan pembersihan data, untuk memahami struktur data dan menangani data yang tidak lengkap atau tidak konsisten.
- Pembuatan fitur (feature engineering), guna mengoptimalkan input yang digunakan dalam proses pemodelan.

- Pemodelan prediksi resign menggunakan algoritma machine learning, dengan tujuan mengklasifikasikan apakah seorang kandidat berpotensi melakukan *job change*.
- Segmentasi risiko resign ke dalam kategori *High*, *Medium*, dan *Low*, berdasarkan probabilitas hasil prediksi.
- Evaluasi performa model menggunakan metrik yang sesuai, serta perumusan rekomendasi kebijakan berbasis hasil prediksi.
- Penyajian hasil melalui aplikasi interaktif menggunakan Streamlit, agar hasil analisis lebih mudah dipahami dan diakses oleh pemangku kepentingan.

Melalui tahapan-tahapan tersebut, proyek ini tidak hanya menghasilkan model prediksi, tetapi juga memberikan dasar bagi pengambilan keputusan strategis yang bersifat preventif terhadap potensi resign karyawan.

1.3 Tujuan

Tujuan dari proyek ini adalah membangun sistem analitik berbasis machine learning untuk memprediksi kemungkinan karyawan melakukan resign dan memberikan rekomendasi strategi retensi yang tepat. Dengan adanya sistem ini, perusahaan dapat mengambil keputusan berbasis data dalam mengelola risiko resign serta mengoptimalkan alokasi anggaran pelatihan dan pengembangan karyawan. Secara lebih rinci, tujuan ini dapat dijabarkan ke dalam tiga bagian utama:

a) Goals

Menyediakan *profiling* risiko resign karyawan berdasarkan probabilitas, guna mendukung upaya pencegahan resign dan optimalisasi *training & development cost*.

b) Objectives

1. Membangun model machine learning untuk memprediksi apakah seorang karyawan berpotensi resign atau tidak.

2. Melakukan segmentasi risiko resign ke dalam tiga kategori (*High*, *Medium*, dan *Low*) agar perusahaan dapat memberikan *treatment* atau intervensi yang tepat sesuai tingkat risiko masing-masing karyawan.

c) Success Metrics

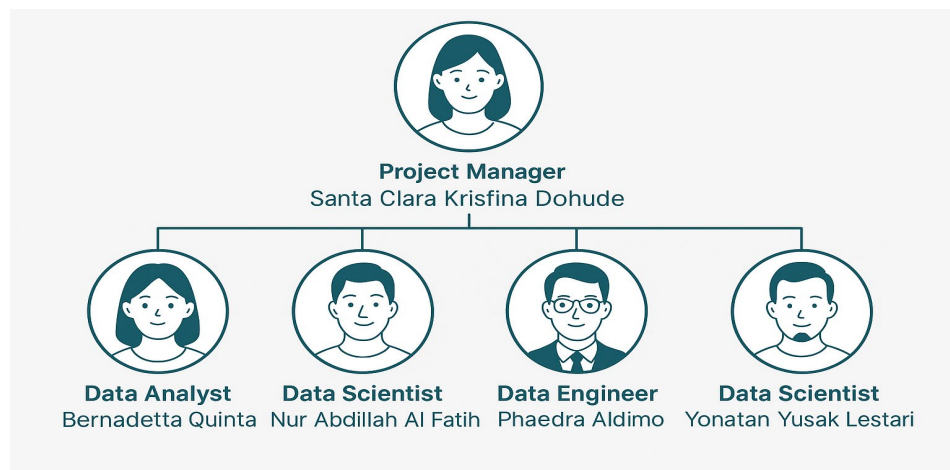
1. Turnover Rate: Keberhasilan proyek ini diharapkan dapat membantu perusahaan menurunkan angka turnover menuju batas ideal 5–10%, sesuai standar dari SHRM dan Deloitte[3].
2. Training & Development Cost Optimization: Efisiensi pengalokasian biaya pelatihan, dengan distribusi yang disarankan yaitu 50% untuk high risk, 35% untuk medium, dan 15% untuk low risk[4].

Dengan tercapainya tujuan-tujuan tersebut, perusahaan dapat mengelola risiko resign secara proaktif dan mengoptimalkan investasi dalam pengembangan sumber daya manusia.

II. Kelompok Final Project Bootcamp Rakamin

2.1 Struktur Organisasi

Struktur organisasi kelompok Final Project Bootcamp Rakamin disusun untuk mendukung kelancaran pelaksanaan proyek serta memastikan pembagian peran yang jelas antar anggota tim. Dalam struktur ini, seluruh anggota tergabung dalam satu divisi proyek yang secara langsung terlibat dalam kegiatan analisis data, pembuatan model prediktif, dan implementasi solusi berbasis machine learning. Divisi proyek ini ditandai secara khusus dalam gambar struktur organisasi berikut untuk membedakannya dari unit atau divisi lain. Dengan struktur sebagai berikut.



Gambar 2. 1 Struktur Organisasi

2.2 Lingkup Pekerjaan

1. Project Manager – Santa Clara Krisfina Dohude

- Mengawasi jadwal proyek, memastikan tugas-tugas selesai tepat waktu, dan mengelola komunikasi dalam tim
- Memastikan bahwa semua dokumentasi proyek terorganisir dengan baik.
- Jika grup memilih untuk menerapkan model (misalnya, menggunakan Flask atau Streamlit), tangani aspek teknis penerapan.
- Memastikan kualitas proyek secara keseluruhan, memeriksa konsistensi antara tujuan bisnis, kinerja model, dan pelaporan.

2. Data Engineer – Phaedra Aldimo

- Menangani semua aspek pembersihan, transformasi, dan integrasi data dari berbagai sumber dan memastikan kumpulan data terstruktur dengan baik untuk analisis
 - Lakukan EDA untuk menemukan tren, hubungan, dan potensi masalah dalam data.
 - Gunakan visualisasi untuk mengkomunikasikan temuan awal kepada tim.
 - Menyiapkan pipeline data untuk mendukung analisis prediktif turnover dan variabel terkait seperti jam pelatihan, gaji, masa kerja, dan status resign.
3. Data Scientist – Nur Abdillah Al Fatih & Yonatan Yusak Lestari
- Mengembangkan model pembelajaran mesin atau solusi analitis tingkat lanjut berdasarkan kebutuhan proyek.
 - Bereksperimen dengan model yang berbeda (misalnya, Logistic Regression, Random Forest, XGBoost) untuk mengidentifikasi resiko resign.
 - Menggunakan data uji untuk memvalidasi kinerja model dan mengoptimalkan akurasi menggunakan validasi silang dan metrik yang tepat.
 - Berkolaborasi dengan Data Engineer untuk memastikan bahwa fitur yang digunakan dalam model meningkatkan kinerja.
 - Membuat model di-deploy menggunakan Streamlit Cloud dan GitHub sebagai integrasi teknologi.
4. Data Analyst – Bernadetta Quinta
- Menentukan masalah bisnis dan tujuan analisis serta mengidentifikasi tren dan kebutuhan industri.
 - Menerjemahkan wawasan data menjadi rekomendasi bisnis yang dapat ditindaklanjuti. Memastikan bahwa hasil yang dihasilkan sesuai dengan tren industri atau kebutuhan pemangku kepentingan.
 - Menentukan seperti apa kesuksesan proyek (pengukuran proyek) pada penurunan turnover dan efisiensi cost training.
 - Membuat visualisasi yang menyajikan wawasan dan hasil dengan cara yang dapat diakses oleh pemangku kepentingan non-teknis

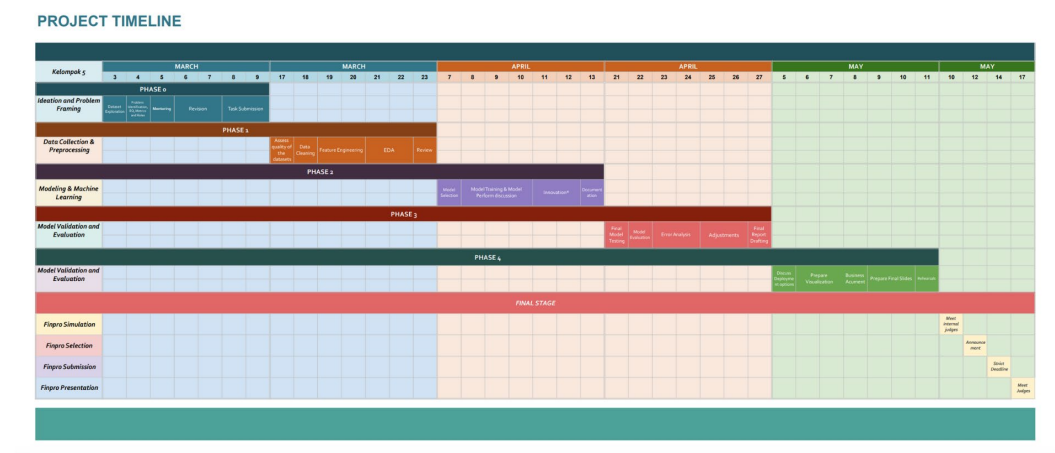
2.3 Deskripsi Pekerjaan

Tabel 2. 1 Peran, Tugas, dan Kontribusi Anggota Tim Proyek

<i>Role</i>	<i>Task Description</i>	<i>Contribution</i>
<i>Project Manager</i>	Mengawasi jadwal proyek, memastikan tugas-tugas selesai tepat waktu, dan mengelola komunikasi dalam tim	Memastikan semua dokumentasi terorganisir, mengelola komunikasi dengan tim, serta memastikan kualitas dan konsistensi proyek.
	Memastikan bahwa semua dokumentasi proyek terorganisir dengan baik.	Memastikan insight data diterjemahkan menjadi kebijakan nyata dalam pengembangan karyawan.
	Jika grup memilih untuk menerapkan model, tangani aspek teknis penerapan (misalnya, menggunakan Flask atau Streamlit).	Mengawasi penerapan model dan memastikan seluruh proses berjalan dengan baik hingga pengujian akhir model.
	Memastikan kualitas proyek secara keseluruhan, memeriksa konsistensi antara tujuan bisnis, kinerja model, dan pelaporan.	Menjaga kualitas laporan dan tujuan bisnis tetap sejalan dengan hasil analisis serta presentasi dari anggota tim.
<i>Data Engineer</i>	Menangani semua aspek pembersihan, transformasi, dan integrasi data dari berbagai sumber.	Menyiapkan dan membersihkan data dari berbagai sumber yang relevan, termasuk mempersiapkan pipeline untuk analisis prediktif turnover.
	Lakukan EDA untuk menemukan tren, hubungan, dan potensi masalah dalam data.	Menyediakan wawasan awal kepada tim terkait potensi masalah dalam data dan menemukan pola-pola yang dapat digunakan untuk model prediktif.
	Gunakan visualisasi untuk mengkomunikasikan temuan awal kepada tim.	Membantu tim dengan visualisasi data awal yang mudah dipahami untuk perencanaan dan keputusan lebih lanjut.
<i>Data Scientist</i>	Mengembangkan model Machine Learning berdasarkan kebutuhan project.	Membangun dan mengembangkan model Machine Learning
	Bereksperimen dengan model yang berbeda (misalnya, Logistic Regression, Random Forest, XGBoost) untuk mengidentifikasi resiko resign.	Menguji model berbeda untuk mengidentifikasi risiko resign dan memilih model terbaik.

	Menggunakan data uji untuk memvalidasi kinerja model dan mengoptimalkan akurasi menggunakan validasi silang dan metrik yang tepat.	Meningkatkan akurasi model dan memastikan kinerja model memadai sebelum implementasi lebih lanjut.
	Berkolaborasi dengan Data Engineer untuk memastikan bahwa fitur yang digunakan dalam model meningkatkan kinerja.	Memastikan data yang digunakan dalam model sudah bersih dan dapat memaksimalkan prediksi terkait turnover karyawan.
Data Analyst	Menentukan masalah bisnis dan tujuan analisis serta mengidentifikasi tren dan kebutuhan industri.	Menganalisis data untuk menentukan faktor-faktor yang mempengaruhi turnover karyawan dan bagaimana program pelatihan bisa membantu.
	Menerjemahkan insight menjadi rekomendasi bisnis yang dapat ditindaklanjuti. Memastikan bahwa hasil yang dihasilkan sesuai dengan tren industri.	Memberikan rekomendasi berbasis data kepada stakeholder terkait kebijakan pengurangan turnover dan pengembangan pelatihan.
	Menentukan seperti apa kesuksesan proyek (pengukuran proyek) pada penurunan turnover dan efisiensi cost training.	Menyusun metrik kesuksesan proyek dan memastikan laporan berfokus pada efektivitas pengurangan turnover dan cost efisiensi pelatihan.
	Membuat visualisasi yang menyajikan wawasan dan hasil dengan cara yang dapat diakses oleh pemangku kepentingan non-teknis.	Membuat laporan visual yang mudah dipahami oleh pihak-pihak non-teknis, seperti HR dan manajemen.

2.4 Jadwal Kerja



Gambar 2. 2 Timeline Proyek

Keterangan Kegiatan Mingguan:

1. Hari Senin – Diskusi Kelompok

- Seluruh anggota melakukan diskusi kelompok (group discussion) untuk menyepakati pembagian tugas mingguan.
- Setiap anggota memahami scope pekerjaan berdasarkan phase yang sedang dijalankan.

2. Hari Selasa – Pengerjaan Individu

- Anggota kelompok mengerjakan tugas masing-masing sesuai peran yang telah disepakati
- Tugas dilakukan secara asinkron melalui platform kolaboratif internal (misalnya Google Collab/Gdrive/Github).

3. Hari Rabu – Penyusunan Laporan dan Presentasi

- Kelompok mengumpulkan hasil pekerjaan masing-masing dan mengintegrasikannya dalam bentuk laporan dan presentasi mingguan.
- Visualisasi, insight bisnis, dan progress model dimasukkan dalam deck presentasi yang akan dilaporkan ke mentor.

4. Hari Rabu atau Kamis – Sesi Mentoring

- Project Manager bertugas berkoordinasi dengan mentor untuk menjadwalkan sesi mentoring.
- Sesi ini digunakan untuk memperoleh masukan terhadap hasil sementara, klarifikasi metode, dan arah kerja tim.

Melalui proyek ini, Kelompok Tenera berupaya memberikan kontribusi nyata dalam pengembangan solusi data-driven untuk masalah yang sering dihadapi oleh perusahaan, yaitu tingginya angka turnover karyawan. Kelompok Tenera berharap hasil project ini dapat memberikan dampak positif dalam proses pengambilan keputusan di bidang HR.

III. Prediksi Resign Karyawan Menggunakan Machine Learning

3.1 Deskripsi Permasalahan Proyek

Dalam dunia kerja modern, khususnya pada sektor teknologi dan data science, tingkat mobilitas karyawan sangat tinggi. Fenomena resign atau job change menjadi salah satu tantangan serius yang dihadapi oleh perusahaan karena dapat berdampak langsung terhadap stabilitas tim, efisiensi operasional, dan biaya sumber daya manusia.

Berdasarkan data yang dianalisis, ditemukan bahwa tingkat turnover karyawan mencapai 28,5%, jauh di atas batas ideal 5–10% sebagaimana direkomendasikan oleh para pakar manajemen sumber daya manusia (Wellhub, 2023). Angka ini dihitung berdasarkan proporsi karyawan yang keluar dibandingkan dengan rata-rata jumlah karyawan selama periode tertentu.

Tingginya angka turnover ini menunjukkan bahwa banyak perusahaan belum mampu secara efektif memetakan risiko resign dan melakukan intervensi preventif yang tepat. Salah satu penyebabnya adalah belum adanya sistem berbasis data yang dapat memprediksi siapa saja karyawan yang berisiko tinggi untuk resign, serta bagaimana strategi pencegahan yang optimal dapat dirancang.

Oleh karena itu, proyek ini hadir untuk menjawab persoalan tersebut dengan pendekatan berbasis machine learning, yaitu membangun model prediktif dan sistem segmentasi risiko resign, yang pada akhirnya diharapkan dapat membantu pengambilan keputusan strategis dalam pengelolaan SDM perusahaan.

3.2 Proses Pelaksanaan Proyek

Proyek dilaksanakan secara kolaboratif dengan pembagian peran yang jelas dalam tim:

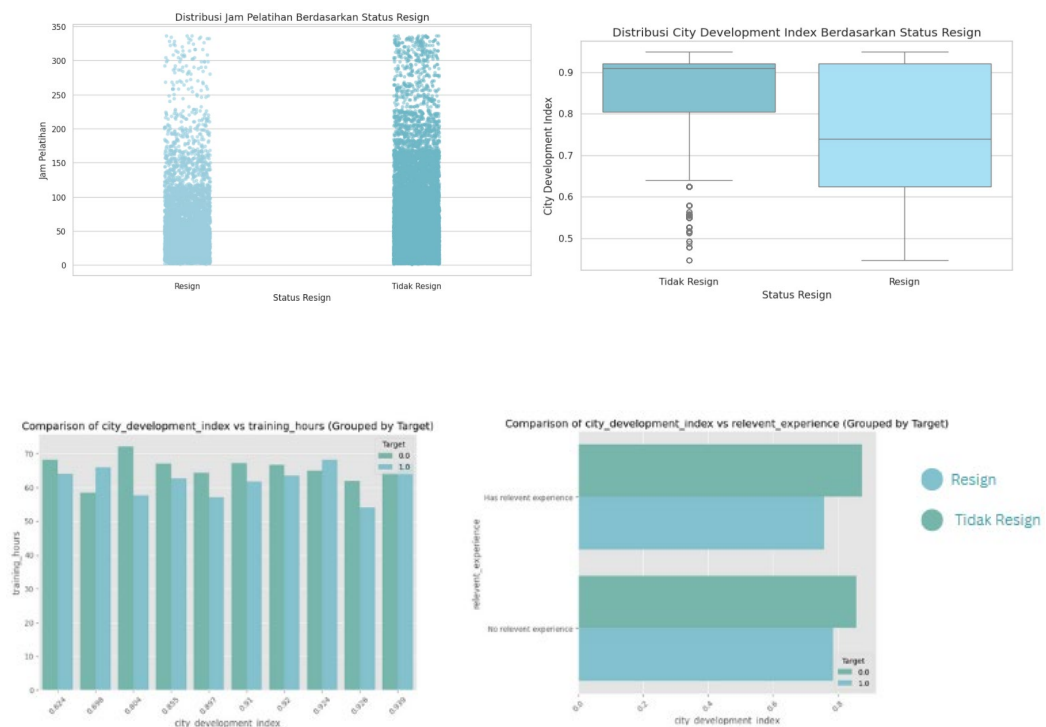
- Senin: Diskusi mingguan dan pembagian tugas.
- Selasa: Pengerjaan individu menggunakan tools seperti Google Drive, Colab, GitHub.
- Rabu: Integrasi laporan dan pembuatan presentasi.
- Kamis: Sesi mentoring untuk evaluasi dan arahan.

Proyek ini juga dilakukan melalui beberapa tahapan yang sistematis, mulai dari eksplorasi data hingga tahap pre-processing, dengan tujuan menghasilkan data yang siap untuk dipakai dalam pemodelan machine learning. Proses ini dilakukan dengan bantuan berbagai pustaka Python seperti *pandas*, *seaborn*, *scikit-learn*, dan *imbalanced-learn*. Dengan tahapan sebagai berikut.

1. Eksplorasi Data (Exploratory Data Analysis)

Langkah pertama dalam proses analisis adalah melakukan eksplorasi data untuk memahami struktur serta pola awal yang mungkin berhubungan dengan status resign karyawan. Hasil eksplorasi menunjukkan bahwa karyawan yang tidak resign umumnya berasal dari kota dengan City Development Index yang lebih tinggi, dengan nilai median mendekati 0,9. Sebaliknya, karyawan yang resign memiliki sebaran indeks kota yang lebih luas dengan median lebih rendah, yaitu sekitar 0,75.

Selain itu, variabel *training_hours* juga menunjukkan pengaruh terhadap keputusan resign. Karyawan dengan jam pelatihan yang lebih tinggi cenderung memiliki kemungkinan lebih kecil untuk mengundurkan diri.



Gambar 3. 1 Analisis Univariat & Multivariat

2. Korelasi Antar Fitur

Untuk mengetahui seberapa besar hubungan antar fitur dengan target resign, digunakan heatmap correlation.



Gambar 3. 2 Heatmap Korelasi

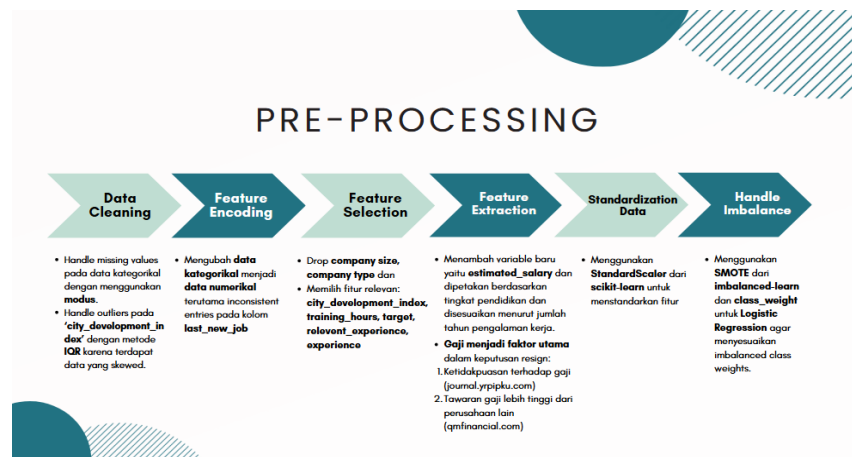
Hasilnya menunjukkan bahwa city_development_index memiliki korelasi negatif terhadap target resign (bernilai -0.34), menjadikannya salah satu fitur paling informatif. Sementara fitur lain seperti training_hours, experience, dan relevent_experience memiliki pengaruh sangat kecil terhadap target. Tidak ditemukan korelasi tinggi antar fitur, sehingga tidak ada indikasi multikolinearitas yang signifikan.

3. Pembersihan dan Pra-pemrosesan Data (*Pre-Processing*)

Tahapan ini mencakup beberapa proses penting:

- Data Cleaning: Missing value diisi dengan modus, dan outlier pada city_development_index ditangani dengan metode IQR.
- Feature Encoding: Data kategorikal seperti last_new_job dikonversi ke bentuk numerik.

- Feature Selection: Fitur yang kurang relevan seperti `company_size` dan `company_type` di-drop, sedangkan fitur yang dinilai relevan seperti `training_hours`, `experience`, dan `city_development_index` tetap dipertahankan.
- Feature Extraction: Ditambahkan fitur `estimated_salary` yang dihitung berdasarkan tingkat pendidikan dan pengalaman kerja.
- Standardisasi Data: Fitur numerik dinormalisasi menggunakan `StandardScaler` dari `scikit-learn`.
- Penanganan Data Imbalance: Menggunakan metode `SMOTE` dari `imbalanced-learn` serta `class_weight` di model `Logistic Regression`.



Gambar 3.3 Alur *Pre-Processing*

4. Pembuatan Fitur Tambahan (Feature Engineering)

Salah satu fitur baru yang dikembangkan dalam proyek ini adalah `estimated_salary`, yaitu estimasi gaji berdasarkan kombinasi antara tingkat pendidikan dan pengalaman kerja. Penambahan fitur ini dilakukan karena gaji merupakan salah satu faktor utama yang memengaruhi keputusan seorang karyawan untuk mengundurkan diri dari perusahaan.

Ketidakpuasan terhadap gaji yang dirasa tidak sesuai dengan beban kerja atau tidak kompetitif dibandingkan standar industri dapat mendorong karyawan untuk mencari pekerjaan baru[2]. Selain itu, adanya tawaran gaji yang lebih tinggi dari perusahaan lain juga menjadi alasan umum dalam keputusan resign [5].

Estimasi gaji ditentukan berdasarkan pemetaan realistis untuk posisi data scientist di Indonesia, dengan mengacu pada data dari Digital Skola[6] dan Katadata[7]. Estimasi ini kemudian disesuaikan dengan lama pengalaman kerja, yang diasumsikan berpengaruh terhadap potensi resign. Nilai estimasi tersebut ditampilkan dalam tabel berikut.

Tabel 3. 1 Mapping Fitur *Estimasi_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

3.3 Pencapaian Hasil Proyek

1. Model Prediktif yang Efektif

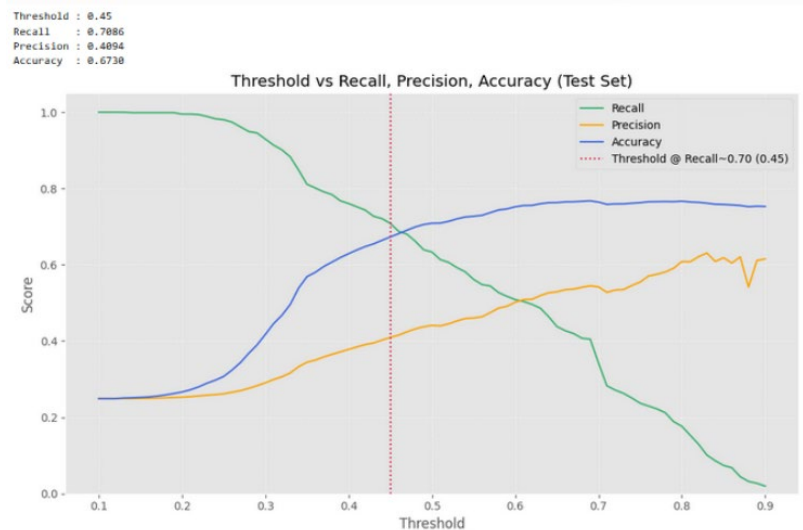
Beberapa algoritma telah diuji, antara lain: Decision Tree, Random Forest, Logistic Regression, dan XGBoost. Berdasarkan hasil evaluasi performa, Logistic Regression dipilih sebagai model terbaik karena memiliki keseimbangan antara akurasi, recall, dan overfitting yang rendah.

Dibandingkan model lain, Logistic Regression menunjukkan performa recall tertinggi pada data test (0,61), yang berarti model mampu menangkap lebih banyak karyawan yang berpotensi resign. Selain itu, perbedaan performa antara data train dan test juga cukup kecil, sehingga tidak terjadi overfitting yang signifikan. Dengan ringkasan terdapat pada table sebagai berikut.

Tabel 3. 2 Perbandingan Hasil Evaluasi Antar 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

Selanjutnya, dilakukan optimasi threshold agar model lebih sensitif terhadap prediksi resign.



Gambar 3. 4 Model Optimum

Hasil dari penggunaan threshold 0,45 adalah sebagai berikut:

- Recall: 70,9%

Dari 100 karyawan yang benar-benar resign, model dapat mengidentifikasi sekitar 71 orang.

- Precision: 40,9%

Dari 100 karyawan yang diprediksi akan resign oleh model, hanya 41 yang benar-benar resign.

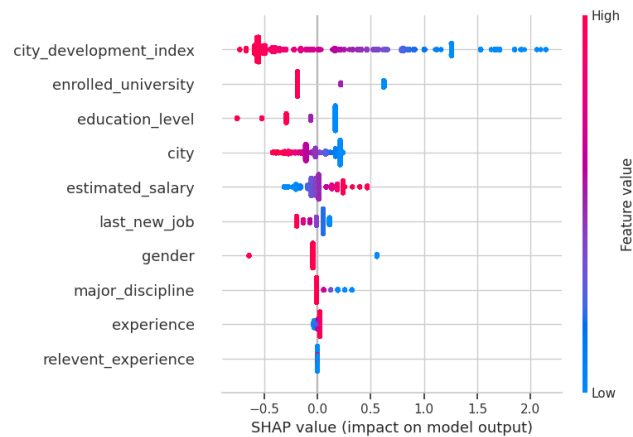
- F1-Score: 51,9%

Menggambarkan keseimbangan antara recall dan precision. Nilai ini menunjukkan bahwa model masih memiliki performa sedang, namun cukup layak untuk digunakan dalam tahap screening awal.

Untuk mengetahui fitur mana yang paling berpengaruh terhadap prediksi resign, digunakan metode interpretabilitas model yaitu SHAP (*SHapley Additive exPlanations*). Hasil visualisasi SHAP menunjukkan bahwa:

- City Development Index adalah fitur paling berpengaruh dalam mempengaruhi keputusan model. Nilai indeks yang rendah (berwarna biru) cenderung meningkatkan prediksi resign (tergeser ke arah kanan/grafik positif).
- Fitur lain yang juga cukup berkontribusi, meskipun tidak sebesar City Development Index, antara lain:
 - Relevant Experience: karyawan dengan pengalaman yang relevan memiliki risiko resign yang lebih tinggi.
 - Estimated Salary: karyawan dengan gaji yang lebih rendah dari ekspektasi menunjukkan kecenderungan resign lebih besar.

- Education Level dan Enrolled University juga memberikan pengaruh, meskipun lebih kecil skalanya.



Gambar 3. 5 *Shap Value*

Dengan memahami fitur-fitur yang paling mempengaruhi keputusan model, tim HR dapat menyusun strategi intervensi yang lebih terarah, seperti mengembangkan program retensi untuk karyawan dari kota dengan indeks pembangunan rendah atau mereka yang bergaji di bawah rata-rata.

2. Segmentasi

Setelah model Logistic Regression dipilih sebagai model terbaik, dilakukan proses tuning threshold untuk menyesuaikan strategi klasifikasi dengan kebutuhan bisnis. Default threshold 0,50 dinilai kurang sensitif dalam mendeteksi karyawan yang berisiko resign. Oleh karena itu, dilakukan penyesuaian menjadi threshold = 0,45, karena nilai ini memberikan trade-off terbaik antara recall dan precision . Berdasarkan threshold ini, dilakukan segmentasi risiko terhadap seluruh populasi karyawan ke dalam tiga kategori:

Tabel 3. 3 Kategori Segmentasi Risiko Resign

Kategori Risiko	Probabilitas Prediksi
High Risk	$\geq 0,70$
Medium Risk	0,45 – 0,69

Low Risk	< 0,45
----------	--------

Segmentasi ini memungkinkan perusahaan mengidentifikasi kelompok karyawan yang perlu perhatian lebih besar, baik dalam bentuk retensi, evaluasi ulang job satisfaction, maupun alokasi pelatihan yang tepat.

Dampak Bisnis (Business Impact)

Model prediksi resign tidak hanya memberikan insight terhadap risiko resign, tetapi juga menciptakan dampak nyata bagi strategi manajemen SDM perusahaan. Berikut beberapa dampak utama yang dihasilkan:

3. Penurunan Turnover Rate Secara Signifikan

Berdasarkan hasil prediksi:

- Total Karyawan: 19.158
- Estimasi yang resign (tanpa intervensi): 4.790 orang (28,5%)
- Recall Model (71%): mampu mengidentifikasi 3.401 orang dari 4.790 yang benar-benar akan resign
- Bila 3.401 karyawan berhasil dicegah untuk resign, maka turnover turun menjadi 7,5%, mendekati standar ideal 5–10% [1].

Tabel 3. 4 Estimasi Turnover Rate Sebelum dan Setelah Intervensi

Keterangan	Nilai Sebelum	Nilai Setelah
Estimasi Resign	4.790 orang	1.389 orang
Turnover Rate	28,5%	7,5%

4. Efisiensi Alokasi Training & Development (T&D)

Setelah dilakukan segmentasi risiko, perusahaan bisa mengalokasikan anggaran T&D sebesar Rp1 miliar lebih efisien:

Tabel 3. 5 Redistribusi Anggaran T&D Berdasarkan Risiko Resign

Segmen Risiko	Sebelum (Equal 33,3%)	Sesudah (Target)	Δ (%)	Δ (Rp)
High-Risk	33,3% → Rp333.333.333	50% → Rp500.000.000	+16,7 pp	+Rp166.666.667
Medium-Risk	33,3% → Rp333.333.333	35% → Rp350.000.000	+1,7 pp	+Rp16.666.667
Low-Risk	33,3% → Rp333.333.333	15% → Rp150.000.000	-18,3 pp	-Rp183.333.333
Total	100%	100%	0 pp	Rp0

Insight:

- Alokasi difokuskan ke karyawan risiko tinggi.
- Karyawan risiko rendah tetap mendapat pelatihan sesuai kebutuhan dasar.

5. Strategi Retensi Lebih Akurat melalui Segmentasi Risiko

Segmentasi prediksi resign berdasarkan threshold 0,45 dan 0,70 menjembatani recall 70% dan precision 40%, menghasilkan tindakan HR yang lebih terfokus.

Tabel 3. 6 Perbandingan Dampak Sebelum dan Setelah Segmentasi

Aspek	Dampak Sebelum Segmentasi	Dampak Setelah Segmentasi
Precision	40% → banyak salah sasaran	Program difokuskan ke misalnya top 100 → lebih efisien
Biaya Retensi	Boros karena menasar karyawan yang tidak resign	Biaya lebih terkendali
Beban Tim HR	Overload dan burnout	HR lebih strategis, kerja berdasarkan prioritas risiko

Dampak pada Karyawan Tetap	Potensi demotivasi jika tidak resign tapi disasar	Tidak reaktif, tetapi berbasis data dan risk scoring
----------------------------	---	--

6. Rekomendasi

Berdasarkan keseluruhan hasil yang telah diperoleh dari proses pemodelan dan analisis, berikut adalah rekomendasi yang dapat diberikan untuk membantu perusahaan dalam menurunkan angka turnover dan mengoptimalkan retensi karyawan:

- a. 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: Mengetahui alasan resign membantu membedakan resign sukarela atau tidak.
 - 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.
- b. Rekomendasi Berdasarkan Segmentasi Risiko (Threshold):
 - High Risk (Threshold > 0.70):
 - Budgeting Cost: Menggunakan alokasi 50%.
 - Tindakan Retensi Aktif: Komunikasi 1-on-1 dengan atasan dan pemberian bonus/insentif.
 - Medium Risk (0.45 – 0.70):
 - Budgeting Cost: Menggunakan alokasi 35%.
 - Tindakan Engagement: Survei kepuasan, feedback, dan komunikasi dua arah.
 - Low Risk (< 0.45):
 - Budgeting Cost: Menggunakan alokasi 15%.
 - Tidak diperlukan tindakan khusus, cukup pemantauan rutin.

c. Rekomendasi Berdasarkan Fitur-Fitur Penting:

- Gaji: Perusahaan disarankan meninjau ulang terhadap gaji, karena gaji rendah berkorelasi dengan risiko resign yang lebih tinggi.
- City Development Index: Penempatan/Pemindahan karyawan di kota dengan indeks pembangunan yang lebih tinggi mengurangi risiko resign. (Rentang 0.8 – 1.0)
- Training Hours: Perusahaan disarankan untuk meningkatkan akses pelatihan bagi karyawan, karena data menunjukkan bahwa karyawan dengan jam pelatihan lebih tinggi cenderung tidak resign.

d. Deployment

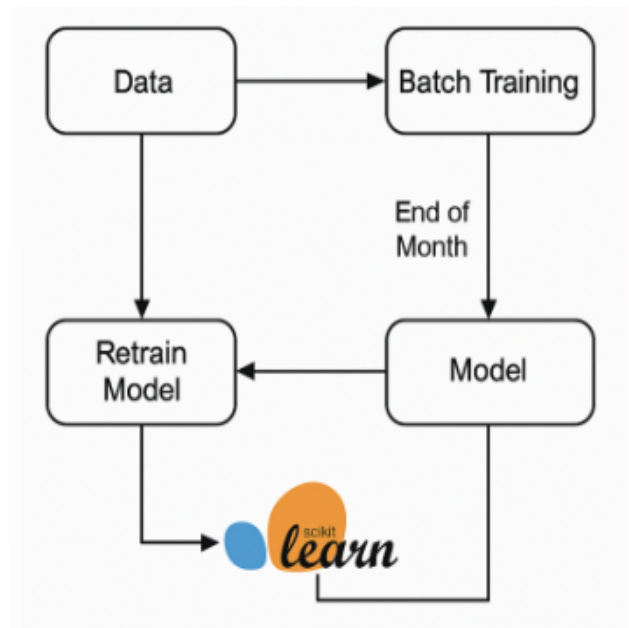
Model terbaik yang telah dipilih dari hasil evaluasi (Logistic Regression dengan threshold optimal 0,45) kemudian di-*deploy* menggunakan platform **Streamlit Cloud** agar dapat digunakan oleh pihak HR secara praktis. Aplikasi ini memungkinkan pengguna untuk memasukkan data karyawan dan mendapatkan prediksi kemungkinan resign, lengkap dengan segmentasi risiko dan rekomendasi tindak lanjut.



The screenshot displays a web application interface for predicting employee resignation. The title is 'Prediksi Kemungkinan Karyawan Resign'. Below the title is a section labeled 'Input Data Karyawan'. The form contains several input fields and dropdown menus arranged in two columns. The left column includes: 'Kota' (City) with a dropdown showing 'city_117', 'City Development Index (CDI)' with a text input '0.850', 'Jenis Kelamin' (Gender) with a dropdown showing 'Male', 'Tingkat Pendidikan' (Education Level) with a dropdown showing 'Primary School', and 'Jumlah' (Count) with a dropdown showing '1000'. The right column includes: 'Status Pernikahan' (Marital Status) with a dropdown showing 'no_married', 'Pengalaman Kerja (tahun)' (Work Experience in years) with a dropdown showing '<1', 'Pengalaman Relawan' (Volunteer Experience) with a dropdown showing 'Has relevant experience', and 'Tingkat Gaji Penerimaan' (Salary Level) with a dropdown showing '3'. There are also radio buttons for 'Status Pernikahan' and 'Pengalaman Kerja (tahun)'.

Gambar 3. 6 Tampilan *Streamlit Deployment*

Alur deployment dilakukan melalui integrasi GitHub dengan Streamlit, sehingga memungkinkan pengelolaan versi model yang lebih terstruktur. Selain itu, disusun juga flow retraining yang dilakukan secara berkala (batch training setiap akhir bulan) untuk memastikan model tetap akurat seiring dengan bertambahnya data baru.



Gambar 3. 7 Alur *Deployment*

Dengan adanya deployment ini, perusahaan dapat:

- Melakukan prediksi resign secara real-time.
- Mengidentifikasi risiko resign per individu.
- Memberikan rekomendasi tindakan berdasarkan skor risiko.

Tampilan antarmuka aplikasi menampilkan probabilitas resign, segmentasi risiko (High, Medium, Low), serta saran alokasi strategi retensi berdasarkan hasil prediksi tersebut.

IV. Penutup

4.1 Kesimpulan

Final Project Bootcamp ini berhasil mengembangkan sebuah sistem prediksi risiko resign karyawan berbasis machine learning yang mampu berfungsi sebagai sistem peringatan dini. Model terbaik yang digunakan adalah Logistic Regression dengan recall sebesar 70,9%, sehingga mampu mendeteksi sekitar 71% dari karyawan yang benar-benar akan resign. Model ini telah melalui proses evaluasi, tuning threshold, serta deployment melalui Streamlit Cloud agar dapat diakses oleh pihak HR secara real-time.

Dari sisi bisnis, proyek ini menunjukkan dampak yang signifikan:

- Dengan model prediksi dan segmentasi risiko yang telah dilakukan, perusahaan dapat mengurangi estimasi turnover rate dari 28,5% menjadi 7,5%.
- Strategi ini juga berpotensi memaksimalkan biaya T&D perusahaan melalui alokasi ulang anggaran pelatihan berdasarkan tingkat risiko resign (risk-based budgeting).
- Segmentasi risiko juga membuat strategi HR lebih fokus dan efisien, karena tindakan retensi dapat diarahkan hanya pada kelompok karyawan berisiko tinggi.

Keberhasilan proyek ini tidak lepas dari kolaborasi yang baik antar anggota tim, pembagian peran yang jelas, serta dukungan mentor selama proses pengerjaan. Proyek ini telah menyelesaikan seluruh pipeline analisis: mulai dari eksplorasi data, preprocessing, pemodelan, evaluasi, interpretasi, hingga deployment.

4.2 Saran

1. Implementasi Nyata:

- Perusahaan disarankan untuk mulai menerapkan hasil prediksi sebagai dasar kebijakan retensi karyawan.
- Gunakan segmentasi risiko untuk menyesuaikan alokasi pelatihan berdasarkan tingkat risiko resign agar intervensi lebih tepat sasaran.

2. Pengembangan Model:

- Tambahkan variabel penting yang saat ini belum tersedia seperti kepuasan kerja, gaji relatif terhadap standar industri, dan evaluasi kinerja.
- Uji performa model pada data real-time atau lintas sektor industri untuk menguji generalisasi.

3. Penguatan Teknologi:

- Tingkatkan tampilan dan kemudahan penggunaan dari aplikasi Streamlit agar lebih ramah bagi tim HR non-teknis.
- Pertimbangkan integrasi langsung dengan sistem HR perusahaan agar solusi ini dapat digunakan secara berkelanjutan dan otomatis.

Dengan penerapan lanjutan dan peningkatan ini, sistem prediksi risiko resign ini diharapkan dapat membantu perusahaan dalam mengurangi turnover dan meningkatkan efektivitas strategi retensi karyawan.

V. Referensi

- [1] Wellhub, “What is employee turnover rate and why it matters,” 2023. [Online]. Available: <https://wellhub.com/en-us/blog/wellbeing-culture/employee-turnover-rate/>
- [2] A. A. P. Andini and S. Aslamiyah, “The Effect of Salary Satisfaction, Workload and Career Development on Employee Turnover Intention of PT. XXX,” *Manajemen Studies and Entrepreneurship Journal*, vol. 5, no. 2, pp. 7720–7730, 2024. [Online]. Available: <http://journal.yrpiiku.com/index.php/msej>
- [3] SHRM and Deloitte, “Managing for Employee Retention & 2023 Global Human Capital Trends,” 2023. [Online]. Available: <https://www.shrm.org> and <https://www2.deloitte.com>
- [4] Workramp, Internal Learning Strategy Model Report, 2023.
- [5] QM Financial, “Resign Karyawan: Wajar Gak Sih?” Nov. 2023. [Online]. Available: https://qmfinancial.com/2023/11/resign-karyawan/?utm_source=chatgpt.com
- [6] Digital Skola, “Berapa Gaji Data Scientist di Indonesia?”, Digital Skola Blog, 2023. [Online]. Available: <https://digitalskola.com/blog/data-science/gaji-data-science>
- [7] Katadata, “Menilik Faktor yang Mempengaruhi Gaji Data Scientist di Indonesia,” *Katadata.co.id*, May 22, 2023. [Online]. Available: <https://katadata.co.id/lifestyle/varia/646b1d557a39f/menilik-faktor-yang-mempengaruhi-gaji-data-scientist-di-indonesia>

VI. Lampiran A. PPT Final Project Presentation



Link PPT : [Kelompok Tenera](#)

VII. Lampiran B. Notulensi Mentoring

Pada Bagian ini berisi log activity atau notulensi mentoring dengan format sebagai berikut.

Tanggal Mentoring	Materi Diskusi	Hasil Diskusi dan Keputusan	Tindak Lanjut
05 Maret 2025	<ul style="list-style-type: none">- Problem Understanding- Dataset HR Analytics- Penentuan goal & KPI	<ul style="list-style-type: none">- Turnover rate aktual diasumsikan 25% (melebihi ideal 10%)- Fokus pada prediksi probabilitas resign dan segmentasi risiko (tinggi/sedang/rendah)- training_hours akan digunakan untuk efisiensi cost, bukan sebagai input model.	<ul style="list-style-type: none">- Menyusun problem, goals, objective, KPI, dan output- Menyusun timeline project sesuai alur ML
19 Maret 2025	<ul style="list-style-type: none">- Data Understanding- Data Cleaning- Feature Engineering- EDA	<ul style="list-style-type: none">- company_size dan company_type di-drop- training_hours tidak dianggap outlier- Feature extraction berupa estimated_salary ditambahkan- Visualisasi EDA fokus pada fitur relevan: experience, relevant_experience, dll.	<ul style="list-style-type: none">- Tambah visualisasi histogram missing value- Siapkan mapping estimated_salary dari education_level dan experience
16 April 2025	<ul style="list-style-type: none">- Modeling (LogReg, XGBoost, RF, DT)- API & Deployment- Evaluasi Model	<ul style="list-style-type: none">- Logistic Regression dipilih karena recall tertinggi- Mulai eksplorasi deployment dengan Streamlit- Segmentasi dilakukan dengan threshold probabilitas atau	<ul style="list-style-type: none">- Terapkan SMOTE pada LogReg- Pelajari dan siapkan Streamlit untuk deployment- Dokumentasikan di GitHub

		pendekatan statistik kuartil	
01 Mei 2025	<ul style="list-style-type: none"> - SHAP Value & Fairness - Error & Business Impact - Rekomendasi Strategis 	<ul style="list-style-type: none"> - Tidak perlu menghapus fitur dengan kontribusi kecil - Fairness analysis hanya digunakan sebagai bahan pertimbangan - Error analysis berfokus pada impact bisnis dan retensi - Rekomendasi dibagi: teknikal & bisnis 	<ul style="list-style-type: none"> - Buat visualisasi SHAP & distribusi fitur - Tambah grafik dampak retensi, counseling, insentif terhadap penurunan risiko resign
16 Mei 2025	<ul style="list-style-type: none"> - Penyusunan slide presentasi final - Validasi narasi dan alur tim - Goals, Objective, KPI - Modeling, insight & impact bisnis 	<ul style="list-style-type: none"> - Struktur dan alur slide perlu disederhanakan dan dirapikan - Narasi tim harus menggambarkan simulasi tim data di perusahaan - Tambah grafik kondisi turnover vs ideal (dengan sumber) - Fokus pada insight yang berdampak langsung pada bisnis - Gunakan sumber kredibel seperti SHRM dan Workramp 	<ul style="list-style-type: none"> - Revisi slide sesuai urutan dan saran - Buat slide pembuka dan "Goals & Objectives" - Tambah grafik perbandingan turnover rate - Update analisis cost & prevented resign- Gunakan kembali slide rekomendasi dari stage 3 - Tambahkan flow retraining dan tampilkan output utama deployment

VIII. Lampiran C. Dokumen Teknik

- Link Collab : [Collab_Kelompok Tenera](#)
- Link Streamlit : [Steamlit_Kelompok Tenera](#)
- Link Drive : [Drive_Kelompok Tenera](#)