



**PERANCANGAN MACHINE LEARNING UNTUK
PERPUTARAN KARYAWAN MENGGUNAKAN LOGISTIC
REGRESSION DAN RANDOM FOREST**

LAPORAN TUGAS AKHIR

ABANG IZZAN RIFKI CHAIRULLAH

41821010032

**UNIVERSITAS
MERCU BUANA**

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2025



**PERANCANGAN MACHINE LEARNING UNTUK
PERPUTARAN KARYAWAN MENGGUNAKAN LOGISTIC
REGRESSION DAN RANDOM FOREST**

LAPORAN TUGAS AKHIR

ABANG IZZAN RIFKI CHAIRULLAH

41821010032

Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana

**UNIVERSITAS
MERCU BUANA**

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2025

HALAMAN PENYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama	: Abang Izzan Rifki Chairullah
NIM	: 41821010032
Program Studi	: Sistem Informasi
Judul Laporan Skripsi	: Perancangan Machine Learning Untuk Perputaran Karyawan Menggunakan Logistic Regression Dan Random forest

Menyatakan bahwa Laporan Aplikatif/Tugas Akhir/Jurnal/Media Ilmiah ini adalah hasil karya saya sendiri dan bukan plagiat, serta semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Apabila ternyata ditemukan di dalam Laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

28 Juli 2025
Jakarta,.....

UNIVERSITAS
MERCU BUANA



Abang Izzan Rifki Chairullah

LEMBAR PENGESAHAN

Nama Mahasiswa : Abang Izzan Rifki Chairullah
NIM (41821010032)
Judul Tugas Akhir : Perancangan Machine Learning Untuk Perputaran Karyawan Menggunakan Logistic Regression Dan Random Forest

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 28 Juli 2025

Menyetujui,

Pembimbing : Abdi Wahab, S.Kom, MT ()

NIDN : 0305068502

Ketua Penguji : Wawan Gunawan, S.Kom, MT., M.Kom. ()

NIDN : 0424108104

Penguji 1 : Misni, S.Kom, M.Kom ()

NIDN : 0413046802

Penguji 2 : Suhendra, S.Kom, M.Kom ()

NIDN : 0308019002

Mengetahui,



Dr. Bambang Jokonowo, S.Si., M.T.I.
Dekan Fakultas Ilmu Komputer



Wawan Gunawan, S.Kom, MT., M.Kom.
Ka.Prodi Sistem Informasi

KATA PENGANTAR

Puji syukur kita panjatkan kehadirat Allah Subhanahu Wa Ta'ala, yang telah melimpahkan berkat dan karunia-Nya sehingga penulis dapat menyelesaikan Laporan Tugas Akhir dengan baik. Dalam kesempatan ini, saya ingin menyampaikan laporan yang telah saya lakukan selama beberapa waktu dengan judul “Penerapan Machine Learning Untuk Perputaran Karyawan Menggunakan Logistic Regression Dan Random Forest. Laporan ini disusun dengan tujuan sebagai syarat kelulusan dalam fakultas ilmu komputer universitas Mercu Buana.

Saya ingin menyampaikan terimakasih yang sebesar besarnya kepada :

1. .Prof. Dr. Ir. Andi Adriansyah, M.Eng, selaku Rektor Universitas Mercu Buana.
2. Dr. Bambang Jokonowo, S. Si, MTI, selaku Dekan Fakultas Ilmu Komputer.
3. Pak Wawan Gunawan, S.Kom, MT., M.Kom, selaku Ketua Program Studi Sistem Informasi Universitas Mercu Buana.
4. Pak Abdi Wahab, S.Kom, MT Selaku Dosen Pembimbing yang telah menyediakan waktu tenaga dan pikiran untuk mengarahkan saya dalam menyusun skripsi ini.
5. Ibu Lusianah, selaku Ibu kandung saya yang selalu memberikan kasih dan sayang tiada tara dalam perjalanan saya selama hidup hingga menempuh hidup perkuliahan.
6. Bapak Abang Akhmad Fauzi, selaku Ayah kandung saya, beliau lah yang selalu mengajarkan saya arti menjadi seorang mahasiswa yang bertanggung jawab sehingga saya bisa sampai berada di titik ini,
7. Keluarga Besar Alm Bpk Muhammad inin, yang selalu memberikan doa dan semangat untuk penulis dalam menyusun Tugas Akhir ini.
8. Sahabat-Sahabat Sistem Informasi, yang selalu memberikan dorongan Motivasi dan Semangat bagi penulis.
9. Mahasiswa dengan Nim 41821010064, yang selalu berada di sisi penulis dalam senang maupun sedih, terima kasih atas cinta dan dukungan nya.

10. Sahabat-Sahabat Metro, yang selalu membuat saya ceria dikala saya mengalami kehambatan dalam menyusun Tugas Akhir ini.

11. Club Manchester United, yang telah memberikan saya motivasi dan pertandingan nya menemani saya dalam menyusun Tugas Akhir ini, meskipun dalam musim ini sedang dalam masa terpuruk seperti halnya diri saya dalam menyusun Tugas Akhir ini.

Penulis juga menyadari penyusunan ini masi jauh dari sempurna, oleh karena itu, penulis mengharapkan kritik dan saran demi terciptanya hasil yang lebih baik di masa depan. Semoga laporan Tugas Akhir ini ini dapat memberikan manfaat dan menjadi bahan informasi yang bermanfaat bagi pembaca yang ingin mengetahui lebih dalam mengenai machine learning.

Jakarta, 6 Juni 2025



Abang Izzan Rifki Chairullah



UNIVERSITAS
MERCU BUANA

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama	: Abang Izzan Rifki Chairullah
NIM	: 41821010032
Program Studi	: Sistem Informasi
Judul Laporan Skripsi	: Perancangan Machine Learning Untuk Perputaran Karyawan Menggunakan Logistic Regression Dan Random Forest

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Laporan Magang/Skripsi/Tesis/Disertasi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 28 Juli 2025

Yang menyatakan,



(Abang Izzan Rifki Chairullah)

ABSTRAK

Nama : Abang Izzan Rifki Chairullah
NIM : 41821010032
Program Studi : Sistem Informasi
Judul Laporan Skripsi : Perancangan Machine Learning Untuk Perputaran Karyawan Menggunakan Logistic Regression Dan Random Forest
Pembimbing : Abdi Wahab, S.Kom, MT

Penelitian ini berfokus pada penerapan model prediksi untuk mengidentifikasi tingkat perputaran karyawan dengan memanfaatkan algoritma Logistic Regression dan Random Forest. Selain itu, teknik oversampling sintetis SMOTE digunakan untuk mengatasi ketidakseimbangan data. Model ini bertujuan mengungkap faktor-faktor utama yang memengaruhi keputusan karyawan untuk keluar dari perusahaan, seperti pendidikan, usia, jenis kelamin, tingkat pembayaran, pengalaman kerja, dan lokasi. Dataset penelitian berasal dari Kaggle, yang mencakup informasi karyawan dalam format historis. Dengan membandingkan performa Logistic Regression dan Random Forest, baik dengan maupun tanpa SMOTE, penelitian ini menemukan bahwa Random Forest yang dikombinasikan dengan SMOTE menghasilkan akurasi yang lebih tinggi dan mampu menangani distribusi data yang tidak merata secara lebih efektif. Temuan ini memberikan kontribusi signifikan bagi perusahaan, khususnya dalam mengurangi tingkat turnover, merancang kebijakan retensi yang lebih baik, dan mengoptimalkan strategi rekrutmen karyawan.

Kata Kunci : *Logistic Regression, Random Forest, SMOTE, Perputaran Karyawan, Machine Learning.*

ABSTRACT

Name : Abang Izzan Rifki Chairullah
NIM : 41821010032
Study Program : Sistem Informasi
Title Thesis : Perancangan Machine Learning Untuk Perputaran
Karyawan Menggunakan Logistic Regression Dan
Random Forest
Counsellor : Abdi Wahab, S.Kom, MT

This study focuses on implementation a predictive model to identify employee turnover rates by utilizing Logistic Regression and Random Forest algorithms. Additionally, the Synthetic Minority Oversampling Technique (SMOTE) is employed to address data imbalance issues. The model aims to uncover key factors influencing employees' decisions to leave a company, such as education, age, gender, payment tier, work experience, and location. The research dataset, obtained from Kaggle, contains historical employee information. By comparing the performance of Logistic Regression and Random Forest, both with and without SMOTE, the study found that Random Forest combined with SMOTE delivers higher accuracy and effectively manages uneven data distribution. These findings provide significant contributions to companies, particularly in reducing turnover rates, designing better retention policies, and optimizing employee recruitment strategies.

Keywords : Logistic Regression, Random Forest, SMOTE, Employee Turnover, Machine Learning

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERNYATAAN KARYA SENDIRI	ii
HALAMAN PENGESAHAN	iii
KATA PENGANTAR.....	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS.....	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI.....	ix
DAFTAR TABLE	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Permasalahan.....	2
1.3 Tujuan Penelitian	2
1.4 Batasan Masalah.....	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA.....	6
2.1 Teori/Konsep Terkait	6
2.1.1 Logistic Regression.....	6
2.1.2 Random Forest	7
2.1.3 Oversampling SMOTE	8

2.1.4	Perputaran Karyawan	9
2.1.5	Confusion Matrix	10
2.2	Penelitian Terdahulu	13
2.3	Analisis Literature Riview	22
BAB III METODE PENELITIAN		24
3.1	Deskripsi Sumber Data	24
3.2	Teknik Pengumpulan Data	24
3.3	Pemilihan Parameter Model	25
3.4	Diagram Alir Penelitian	26
3.5	Jadwal Penelitian	27
BAB IV HASIL PENELITIAN		30
4.1	Pengelolaan Data	30
4.2	Data Preparation	30
4.3	Pre Prosesing	33
4.4	Modeling Machine Learning	36
4.4.1	Modeling logistic Regression	36
4.4.2	Modeling Logistic Regression Oversampling Dengan Teknik SMOTE	40
4.4.3	Modeling Random Forest	45
4.4.4	Modeling Random Forest Oversampling dengan Teknik SMOTE ...	51
4.5	Visualisasi Data	59
4.5.1	Visualisasi Awal	59
4.5.2	Visualisasi 1 : Distribusi Usia Karyawan	62
4.5.3	Visualisasi 2 : Distribusi Gender Karyawan	63
4.5.4	Visualisasi 3 : Visualisasi Yang Pernah Dibenched Berdasarkan Kota 64	
4.5.5	Visualisasi 4 : Tingkat Pembayaran Vs Tinggalkan Pekerjaan	65

4.5.6	Visualisasi 5 : Pengalaman Di Bidang Saat Ini vs Tinggalkan Pekerjaan	66
4.5.7	Visualisasi 6 : Pendidikan vs Tinggalkan Pekerjaan	67
4.5.8	Visualisasi 8 : Usia vs Tinggalkan Pekerjaan	68
4.5.9	Visualisasi 8 : Kota vs Tinggalkan Pekerjaan.....	69
4.5.10	Visulisasi 9 : Karyawan Yang Pernah Dibinched vs Tinggalkan Pekerjaan71	
4.5.11	Tahun Lamanya Masuk (Joining Year) vs Tinggalkan pekerjaan	72
4.6	Visualisasi Inti.....	73
4.7	Penjelasan Hasil	75
BAB V KESIMPULAN DAN SARAN		77
5.1	Kesimpulan	77
5.2	Saran.....	77
DAFTAR PUSTAKA		79
LAMPIRAN.....		82



DAFTAR TABLE

Table 2. 1 Literature Riview	13
Table 3. 1 Table Parameter Model.....	25
Table 3. 2 Jadwal Penelitian.....	28
Table 4. 1 Table akurasi berdasarkan algoritma	76



DAFTAR GAMBAR

Gambar 2. 1 Logistic Regression	7
Gambar 2. 2 Random Forest	8
Gambar 2. 3 Oversampling SMOTE.....	9
Gambar 3. 1 Dataset Employee.....	24
Gambar 3. 2 Diagram Alir Penelitian	26
Gambar 4. 1 Informasi kolom data set	30
Gambar 4. 2 Pemeriksaan nilai yang hilang.....	31
Gambar 4. 3 Pemeriksaan nilai hilang pada fitur numerik.....	31
Gambar 4. 4 Pemeriksaan nilai hilang pada fitur kategorikal	32
Gambar 4. 5 Visualisasi distribusi target	33
Gambar 4. 6 Encoding data kategorikal	35
Gambar 4. 7 Menilai distribusi kelas target	36
Gambar 4. 8 Mengidentifikasi fitur dan label	36
Gambar 4. 9 Membagi data set	37
Gambar 4. 10 Membangun Model Logistic Regression	38
Gambar 4. 11 Pelatihan model	38
Gambar 4. 12 Evaluasi model	38
Gambar 4. 13 Hasil Modeling Logistic Regression	39
Gambar 4. 14 Menyeimbangkan data dengan Oversampling	40
Gambar 4. 15 Modelling ulang setelah oversampling	42
Gambar 4. 16 Nilai hasil Logistic regression oversampling	42
Gambar 4. 17 Implementasi confusion matrix	43
Gambar 4. 18 Visualisasi confusion matrix logistic regressin dengan oversampling (SMOTE)	44
Gambar 4. 19 Pembagian data pelatihan dan pengujian	45
Gambar 4. 20 Tranformasi variable kategorikal menjadi label encoding	46
Gambar 4. 21 Transformasi data kategorikal untuk random forest	47
Gambar 4. 22 pelatihan dan model evaluasi random forest	48
Gambar 4. 23 Visualisasi confusion matrix random forest.....	50
Gambar 4. 24 Random Oversampling.....	52

Gambar 4. 25 Mengevaluasi hasil penyeimbangan data.....	54
Gambar 4. 26 Membangun model Random forest dengan oversampling.....	56
Gambar 4. 27 Confusion matrix Random Forest Oversampling dengan smote ...	57
Gambar 4. 28 Visualisasi histogram	59
Gambar 4. 29 Visualisasi box plot dan scatter plot.....	60
Gambar 4. 30 Visualisasi distribusi karyawan.....	62
Gambar 4. 31 Visualisasi distribusi gender.....	63
Gambar 4. 32 Visualisasi yang pernah dibenched berdasar kota.....	64
Gambar 4. 33 Visualisasi Tingkat pembayaran vs tinggalkan pekerjaan	65
Gambar 4. 34 Visualisasi pengalaman di bidang saat ini vs tinggalkan pekerjaan	66
Gambar 4. 35 Visualisasi Pendidikan vs tinggalkan pekerjaan	67
Gambar 4. 36 Visualisasi Usia vs tinggalkan pekerjaan	68
Gambar 4. 37 Visualisasi kota vs tinggalkan pekerjaan	70
Gambar 4. 38 Visualisasi karyawan yang pernah dibinced vs tinggalkan pekerjaan	71
Gambar 4. 39 Visualisasi Tahun lamanya Masuk vs tinggalkan pekerjaan.....	72
Gambar 4. 40 Visualisaasi inti	74

DAFTAR LAMPIRAN

Lampiran 1 Kartu Bimbingan Asistensi.....	82
Lampiran 2 Curriculum Vitae	83
Lampiran 3 Surat Keterangan BNSP	84
Lampiran 4 Lembar Pernyataan Hasil Cek Uji Plagiasi	85

