



**KLASIFIKASI POLA KARIER ALUMNI BERDASARKAN
LABEL CLUSTER DATA TRACER STUDY MENGGUNAKAN
RANDOM FOREST, XGBOOST, DAN REGRESI LOGISTIK**

**TUGAS AKHIR
SKRIPSI**

MUHAMMAD ASADEL AYDIN

UNIVERSITAS 41522010023

MERCU BUANA

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2026**



**KLASIFIKASI POLA KARIER ALUMNI BERDASARKAN
LABEL CLUSTER DATA TRACER STUDY MENGGUNAKAN
RANDOM FOREST, XGBOOST, DAN REGRESI LOGISTIK**

**TUGAS AKHIR
SKRIPSI**

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana

MUHAMMAD ASADEL AYDIN

UNIV 41522010023 AS

MERCU BUANA

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2026**

HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Muhammad Asadel Aydin

NIM : 41522010023

Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika

Menyatakan dengan sebenar-benarnya bahwa Tugas Akhir berjudul:

“Klasifikasi pola karier alumni berdasarkan label cluster data tracer study menggunakan random forest, xgboost, dan regresi logistik” adalah hasil karya saya sendiri, tidak mengandung unsur plagiarisme, pelanggaran hak cipta, atau konten ilegal dalam bentuk apapun dan tidak melanggar hukum atau hak pihak manapun.

Apabila di kemudian hari ditemukan pelanggaran terhadap pernyataan ini, saya bersedia menanggung seluruh konsekuensi hukum dan membebaskan Universitas Mercu Buana dari segala bentuk tuntutan hukum dan saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

Demikian pernyataan ini saya buat dengan sebenar-benarnya untuk digunakan sebagaimana mestinya.

UNIVERSITAS
MERCU BUANA

Jakarta, 24 Juli 2026



Muhammad Asadel Aydin

PERNYATAAN SIMILARITY CHECK
/SIMILARITY CHECK STATEMENT

Saya yang bertanda tangan dibawah ini menyatakan, bahwa karya ilmiah yang ditulis oleh
/The undersigned, hereby declare that the scientific work written by

Nama /Name : MUHAMMAD ASADEL AYDIN
NIM /Student ID Number : 41522010023
Program Studi /Study Program : Teknik Informatika

Dengan Judul Tugas Akhir

/The title:

**“KLASIFIKASI POLA KARIER ALUMNI BERDASARKAN LABEL CLUSTER
DATA TRACER STUDY MENGGUNAKAN RANDOM FOREST, XGBOOST, DAN
REGRESI LOGISTIK”**

telah dilakukan pengecekan *similarity* dengan sistem Turnitin pada tanggal:

/Similarity checks have been carried out with the Turnitin system on the date:

24 Juli 2026

dengan nilai persentase sebesar :

/with a percentage value of:

24%

dinyatakan memenuhi standar sesuai dengan ketentuan berlaku di **Fakultas Ilmu Komputer**
Universitas Mercu Buana. */declared to meet standards in accordance with applicable
regulations at the Faculty of Computer Science, Universitas Mercu Buana*

File hasil cek similarity turnitin:

/Turnitin similarity report file

https://drive.google.com/file/d/1o_WYZnaFo4sG4_25IXxmNICbgfQUgoI/view?usp=drive_link



Jakarta, 24 Juli 2026
Admin Turnitin Fasilkom UMB



Agung Prawoto, S.Kom., B.Sc
NIK : 322970503

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama : MUHAMMAD ASADEL AYDIN
NIM : 41522010023
Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika
Judul Tugas Akhir : KLASIFIKASI POLA KARIER ALUMNI
BERDASARKAN LABEL CLUSTER DATA
TRACER STUDY MENGGUNAKAN RANDOM
FOREST, XGBOOST, DAN REGRESI LOGISTIK

Telah berhasil dipertahankan pada sidang tanggal 31 Juli 2026 dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

Disahkan oleh:
Pembimbing



Eliyani, Dr. Ir., M.Kom
NIDN/NUPTK: 0321026901

Jakarta, 27 Agustus 2026

Mengetahui,

UNIVERSITAS
MERCU BUANA
Dekan Fakultas Ilmu Komputer



Dr. Bambang Jokonowo, S.Si., MTI
NIDN/NUPTK: 0320037002

Ketua Program Studi
Teknik Informatika



Dr. Hadi Santoso, S.Kom., M.Kom
NIDN/NUPTK: 0225067701

KATA PENGANTAR

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, saya dapat menyelesaikan Tugas Akhir ini. Penulisan Tugas Akhir ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Teknik Informatika pada Fakultas Ilmu Komputer Universitas Mercu Buana. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan Tugas Akhir ini, sangatlah sulit bagi saya untuk menyelesaikan Tugas Akhir ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Andi Adriansyah, M.Eng. selaku Rektor Universitas Mercu Buana.
2. Bapak Dr. Bambang Jokonowo, S.Si., MTI selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Dr. Hadi Santoso, S.Kom., M.Kom. selaku Ketua Program Studi Teknik Informatika Universitas Mercu Buana.
4. Ibu Dr. Ir. Eliyani, M.Kom selaku dosen pembimbing tugas akhir yang telah memberikan pengarahan, motivasi, menyediakan waktu, tenaga, dan pikiran sehingga selama pembuatan tugas akhir ini terjadwal dengan baik.
5. Kedua Orang Tua saya yang selalu mendukung saya selama menjalani masa studi sebagai mahasiswa Universitas Mercu Buana.
6. Teman-teman kuliah Keluarga Topik yang selalu berbagi informasi dan memberikan dukungan dalam bentuk yang berbeda-beda.

Akhir kata, saya berharap Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga Tugas Akhir ini membawa manfaat bagi pengembangan ilmu.

Jakarta, 30 Agustus 2026



Muhammad Asadel Aydin

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR DI REPOSITORI UMB

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

Nama : MUHAMMAD ASADEL AYDIN
NIM : 41522010023
Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika
Judul Laporan Skripsi : KLASIFIKASI POLA KARIER ALUMNI
BERDASARKAN LABEL CLUSTER DATA TRACER
STUDY MENGGUNAKAN RANDOM FOREST,
XGBOOST, DAN REGRESI LOGISTIK

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 Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

UNIVERSITAS
MERCU BUANA

Jakarta, 24 Juli 2026

Yang menyatakan,



Muhammad Asadel Aydin

**KLASIFIKASI POLA KARIER ALUMNI BERDASARKAN LABEL
CLUSTER DATA TRACER STUDY MENGGUNAKAN RANDOM FOR-
EST, XGBOOST, DAN REGRESI LOGISTIK**

MUHAMMAD ASADEL AYDIN

ABSTRAK

Tracer study merupakan salah satu instrumen evaluasi yang digunakan perguruan tinggi untuk mengetahui gambaran pola karier alumni. Penelitian ini bertujuan menganalisis faktor-faktor yang memengaruhi pola karier alumni melalui metode klasifikasi berdasarkan label *cluster* hasil *clustering* data *tracer study* Universitas Mercu Buana. Tahapan penelitian meliputi prapemrosesan data, penyeimbangan kelas menggunakan *Synthetic Minority Over-sampling Technique (SMOTE)*, serta pembangunan model menggunakan Logistic Regression, Random Forest, dan XGBoost. Label klasifikasi disusun dalam skema dua *cluster* sebagai hasil utama dan empat cluster sebagai pembanding. Evaluasi model dilakukan menggunakan akurasi, *presisi*, *recall*, dan *F1-score*, sedangkan analisis faktor dilakukan melalui *feature importance*. Hasil penelitian menunjukkan bahwa Random Forest memberikan performa terbaik dengan akurasi 79,11% pada skema dua *cluster*. Faktor yang paling berpengaruh terhadap pola karier alumni adalah kerja sama tim, diikuti pengembangan diri, penggunaan teknologi informasi, komunikasi, kewirausahaan, kepemimpinan, jumlah lamaran, jumlah respons, dan jumlah wawancara.

Kata kunci: *Tracer Study*, pola karier alumni, klasifikasi, Random Forest, *feature importance*.

UNIVERSITAS
MERCU BUANA

CLASSIFICATION OF ALUMNI CAREER PATTERNS BASED ON CLUSTER LABELS IN DATA TRACER STUDY USING RANDOM FOREST, XGBOOST, AND LOGISTIC REGRESSION

MUHAMMAD ASADEL AYDIN

ABSTRACT

Tracer studies are widely used by higher education institutions to evaluate graduates' career outcomes. This study aims to analyze the factors influencing alumni career patterns using a classification approach based on cluster labels derived from tracer study data at Universitas Mercu Buana. The research process included data preprocessing, class balancing using the Synthetic Minority Over-sampling Technique (SMOTE), and model development using Logistic Regression, Random Forest, and XGBoost. The classification labels were organized into a two-cluster scheme as the primary analysis and a four-cluster scheme for comparison. Model performance was evaluated using accuracy, precision, recall, and F1-score, while influential factors were identified through feature importance analysis. The results indicate that Random Forest achieved the best performance, with an accuracy of 79.11% in the two-cluster scheme. Teamwork was identified as the most influential factor, followed by self-development, information technology skills, communication, entrepreneurship, leadership, number of job applications, number of responses, and number of interviews.

Keywords: *Tracer Study, alumni career patterns, classification, Random Forest, feature importance.*

UNIVERSITAS
MERCU BUANA

DAFTAR ISI

HALAMAN SAMPUL.....	i
HALAMAN JUDUL	i
HALAMAN PERNYATAAN KARYA SENDIRI	ii
HALAMAN SURAT KETERANGAN HASIL UJI TURNITIN	iii
HALAMAN PENGESAHAN.....	iv
KATA PENGANTAR.....	v
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR DI REPOSITORI UMB.....	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiv
DAFTAR LAMPIRAN.....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	3
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	4
1.5 Batasan Masalah	5
BAB II TINJAUAN PUSTAKA	7
2.1 Penelitian Terdahulu	7
2.2 Teori Pendukung.....	12
2.2.1 <i>Tracer Study</i>	12
2.2.2 <i>Data Mining</i>	14
2.2.3 Metodologi CRIPS-DM	15
2.2.4 <i>Clustering</i>	16
2.2.5 Klasifikasi.....	18
2.2.6 Algoritma Random Forest	19
2.2.7 Algoritma XGBoost	21
2.2.8 Algoritma Regresi Logistik	22
2.2.9 <i>Data Preprocessing</i>	23
2.2.10 Synthetic Minority Over-sampling Technique (SMOTE)	24

2.2.11	<i>Hyperparameter Tuning menggunakan RandomizedSearchCV</i>	25
2.2.12	Evaluasi Model Klasifikasi	27
2.2.13	<i>Feature Importance</i>	29
BAB III	METODE PENELITIAN	31
3.1.	Pendekatan Penelitian	31
3.2.	Desain Penelitian	31
3.3.	Subjek Penelitian	33
3.4.	Instrumen Penelitian	34
3.4.1.	Perangkat Keas	34
3.4.2.	Perangkat Lunak	35
3.5.	Teknik Pengumpulan Data	35
3.6.	Teknik Analisis Data	36
3.7.	Prosedur Penelitian	37
3.7.1.	<i>Business Understanding</i>	37
3.7.2.	<i>Data Understanding</i>	38
3.7.3.	<i>Data Preparation</i>	45
3.7.4.	<i>Modeling</i>	52
3.7.5.	<i>Evaluation</i>	55
BAB IV	PEMBAHASAN	60
4.1.	Persiapan Data	60
4.1.1.	Dataset	61
4.1.2.	<i>Exploratory Data Analysis (EDA)</i>	61
4.1.3.	<i>Data Preprocessing</i>	65
4.1.4.	Analisis Hasil Clustering	82
4.2.	Pemodelan	88
4.2.1.	Pemilihan Model	89
4.2.2.	<i>Hyperparameter Tuning Skema Dua Cluster</i>	90
4.2.3.	Pembangunan Model Skema Dua Cluster	92
4.2.4.	<i>Hyperparameter Tuning Skema Empat Cluster</i>	92
4.2.5.	Pembangunan Model Skema Empat Cluster	94
4.3.	Evaluasi Model	95
4.3.1.	Evaluasi Model Skema Dua Cluster	95
4.3.2.	Evaluasi Model Skema Empat Cluster	103
4.4.	Analisis Hasil	110
4.4.1.	Analisis Perbandingan Performa Algoritma pada Skema Dua Cluster dan Empat Cluster	111
4.4.2.	Analisis Faktor Dominan Berdasarkan <i>Feature Importance</i>	115
BAB V	KESIMPULAN DAN SARAN	121
5.1	Kesimpulan	121
5.2	Saran	122

DAFTAR PUSTAKA.....	124
LAMPIRAN.....	127



DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	7
Tabel 3. 1 Karakteristik Dataset Penelitian pada Skema Dua Cluster dan Empat Cluster	34
Tabel 3. 2 Perangkat Keras	35
Tabel 3. 3 Perangkat Lunak	35
Tabel 3. 4 Karakteristik Atribut yang Digunakan pada Penelitian	41
Tabel 3. 5 Atribut yang Digunakan pada Skema Dua Cluster	47
Tabel 3. 6 Atribut yang Digunakan pada Skema Empat Cluster	48
Tabel 3. 7 Proses Transformasi Data pada Skema Dua Cluster dan Empat Cluster	50
Tabel 4. 1 Karakteristik Awal Dataset	61
Tabel 4. 2 Daftar Atribut Skema Dua Cluster.....	66
Tabel 4. 3 Daftar Atribut Skema Empat Cluster.....	67
Tabel 4. 4 Hasil Penghapusan Data Duplikat Skema Dua Cluster	70
Tabel 4. 5 Hasil Penghapusan Data Duplikat Skema Empat Cluster.....	71
Tabel 4. 6 Hasil Transformasi Data Skema Dua Cluster	73
Tabel 4. 7 Hasil One-Hot Encoding Skema Empat Cluster.....	74
Tabel 4. 8 Hasil Pembagian Data Skema Dua Cluster.....	76
Tabel 4. 9 Hasil Pembagian Data Skema Empat Cluster	77
Tabel 4. 10 Distribusi Kelas Sebelum dan Sesudah SMOTE pada Skema Dua Cluster	80
Tabel 4. 11 Distribusi Kelas Sebelum dan Sesudah SMOTE pada Skema Empat Cluster	81
Tabel 4. 12 Dimensi Data Latih Sebelum dan Sesudah SMOTE pada Skema Empat Cluster.....	82
Tabel 4. 13 Distribusi Anggota Cluster pada Skema Dua Cluster	82
Tabel 4. 14 Karakteristik Cluster pada Skema Dua Cluster	83
Tabel 4. 15 Distribusi Anggota Cluster pada Skema Empat Cluster	85
Tabel 4. 16 Karakteristik Hasil Clustering Skema Empat Cluster Menggunakan K-Means Model	86
Tabel 4. 17 Parameter Hyperparameter Logistic Regression	90
Tabel 4. 18 Parameter Hyperparameter Random Forest.....	91
Tabel 4. 19 Parameter Hyperparameter XGBoost	91
Tabel 4. 20 Parameter Hyperparameter Logistic Regression	93
Tabel 4. 21 Parameter Hyperparameter Random Forest.....	93
Tabel 4. 22 Parameter Hyperparameter XGBoost	94
Tabel 4. 23 Hasil Evaluasi Model Skema Dua Cluster.....	96
Tabel 4. 24 Hasil Evaluasi Logistic Regression pada Skema Dua Cluster.....	97
Tabel 4. 25 Hasil Evaluasi Random Forest pada Skema Dua Cluster	99
Tabel 4. 26 Hasil Evaluasi XGBoost pada Skema Dua Cluster	101
Tabel 4. 27 Perbandingan Performa Model Sebelum dan Sesudah Penerapan SMOTE pada Skema Empat Cluster.....	104
Tabel 4. 28 Hasil Evaluasi Logistic Regression pada Skema Empat Cluster	105
Tabel 4. 29 Hasil Evaluasi Random Forest pada Skema Empat Cluster	107
Tabel 4. 30 Hasil Evaluasi XGBoost pada Skema Empat Cluster.....	109

Tabel 4. 31 Perbandingan Performa Algoritma pada Skema Dua Cluster dan Empat Cluster.....	111
---	-----



DAFTAR GAMBAR

Gambar 3. 1 Diagram Alur Penelitian Klasifikasi Berdasarkan Hasil Clustering	33
Gambar 3. 2 Alur Transformasi Dataset	46
Gambar 3. 3 Proses Penerapan SMOTE	52
Gambar 3. 4 Tahapan Modeling	53
Gambar 3. 5 Tahapan Evaluation.....	57
Gambar 4. 1 Distribusi Label Cluster pada Skema Dua Cluster.....	62
Gambar 4. 2 Distribusi Label Cluster pada Skema Empat Cluster.....	63
Gambar 4. 3 Visualisasi Missing Value pada Skema Dua Cluster	64
Gambar 4. 4 Visualisasi Missing Value pada Skema Empat Cluster	64
Gambar 4. 5 Boxplot Sebelum Penanganan Outlier pada Skema Empat Cluster.	75
Gambar 4. 6 Visualisasi Hasil Clustering Skema Empat Cluster Menggunakan PCA.....	85
Gambar 4. 7 Confusion Matrix Logistic Regression Skema Dua Cluster	98
Gambar 4. 8 Confusion Matrix Random Forest pada Skema Dua Cluster	100
Gambar 4. 9 Confusion Matrix XGBoost pada Skema Dua Cluster	102
Gambar 4. 10 Confusion Matrix Random Forest pada Skema Empat Cluster ...	108
Gambar 4. 11 Perbandingan Nilai Akurasi Algoritma pada Skema Dua Cluster dan Empat Cluster.....	112
Gambar 4. 12 Feature Importance Random Forest pada Skema Dua Cluster	116
Gambar 4. 13 Feature Importance XGBoost pada Skema Dua Cluster.....	118



DAFTAR LAMPIRAN

Lampiran 1. Kartu Asistensi	127
Lampiran 2 Bukti Surat Pernyataan HKI / Terbit	129
Lampiran 3 Curriculum Vitae	131
Lampiran 4 Sertifikat BNSP	132
Lampiran 5 Surat Ijin Riset Perusahaan.....	133

