

**ALOKASI DOSEN PEMBIMBING PERTAMA TUGAS AKHIR
MAHASISWA DENGAN PENDEKATAN *COSINE SIMILARITY*
DAN *GENETIC ALGORITHM***

SANDRA BERLIANA PUTRI



**PROGRAM STUDI STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**

@Hak cipta milik IPB University

IPB University



IPB University
Bogor Indonesia

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Perpustakaan IPB University



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Alokasi Dosen Pembimbing Pertama Tugas Akhir Mahasiswa dengan Pendekatan *Cosine Similarity* dan *Genetic Algorithm*” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juni 2026

Sandra Berliana Putri
G1401221021

ABSTRAK

SANDRA BERLIANA PUTRI. Alokasi Dosen Pembimbing Pertama Tugas Akhir Mahasiswa dengan Pendekatan *Cosine Similarity* dan *Genetic Algorithm*. Dibimbing oleh AGUS MOHAMAD SOLEH dan GERRY ALFA DITO.

Proses alokasi dosen pembimbing pertama tugas akhir mahasiswa di perguruan tinggi sering menghadapi kendala efisiensi karena perlu menyelaraskan kapasitas maksimum bimbingan dosen, dosen pilihan mahasiswa, dan kesesuaian topik penelitian berbasis teks terbuka. Penelitian ini bertujuan membangun mekanisme optimasi alokasi dosen pembimbing pertama menggunakan *genetic algorithm* (GA) yang dikombinasikan dengan *cosine similarity* untuk mengukur kemiripan topik. Permasalahan dimodelkan menggunakan metode penalti yang mengintegrasikan *hard constraint* (kapasitas maksimum bimbingan dosen) dan *soft constraint* (dosen pilihan mahasiswa dan topik penelitian). Data teks dibobotkan menggunakan *term frequency-inverse document frequency* (TF-IDF) dengan tiga skema n-gram (unigram, bigram, dan unibigram). GA kemudian dijalankan pada delapan skenario kombinasi jumlah generasi, mekanisme penalti kapasitas, dan skema pembobotan. Hasil penelitian menunjukkan bahwa hanya peningkatan jumlah generasi yang meningkatkan nilai *fitness* solusi dan waktu komputasi. S₆-unigram dipilih karena menghasilkan alokasi yang terbaik tanpa melanggar kapasitas maksimum bimbingan dosen, meminimumkan ketidaksesuaian dosen pilihan dan topik penelitian, serta efisien secara komputasi. Mekanisme terbaik ini kemudian diimplementasikan ke dalam *dashboard* interaktif berbasis R Shiny untuk mendukung pengambilan keputusan ke depannya.

Kata kunci: alokasi dosen pembimbing, *cosine similarity*, *genetic algorithm*, R Shiny, TF-IDF.



ABSTRACT

SANDRA BERLIANA PUTRI. Allocation of the First Supervisor for Students' Final Assignments Using Cosine Similarity and Genetic Algorithm Approaches. Supervised by AGUS MOHAMAD SOLEH and GERRY ALFA DITO.

The allocation process of the first supervisor for students' final assignments in higher education often faces efficiency challenges due to the need to align lecturer capacities, students' preferred lecturers, and research topic suitability based on open text. This study aims to build an optimization mechanism for allocating the first supervisor using a genetic algorithm (GA) combined with cosine similarity to measure topic similarity. The problem is modeled using a penalty method that integrates hard constraint (maximum advising capacity) and soft constraint (preferred lecturers and research topics). Text data are weighted using term frequency-inverse document frequency (TF-IDF) across three n-gram schemes (unigram, bigram, and unigram). GA is then run across eight scenarios combining numbers of generations, capacity penalty mechanisms, and weighting scheme. The results show that only the increase in the number of generations improves the solution's fitness value and computation time. S₆-unigram was selected because it produced the best allocation without violating lecturers' maximum advising capacities, minimizing mismatches in preferred lecturers and research topics, and remaining computationally efficient. This best mechanism was then implemented into an interactive dashboard based on R Shiny to support decision-making in the future.

Keywords: cosine similarity, genetic algorithm, R Shiny, supervisor allocation, TF-IDF.



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2026
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

ALOKASI DOSEN PEMBIMBING PERTAMA TUGAS AKHIR MAHASISWA DENGAN PENDEKATAN *COSINE SIMILARITY* DAN *GENETIC ALGORITHM*

SANDRA BERLIANA PUTRI

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



Judul Skripsi : Alokasi Dosen Pembimbing Pertama Tugas Akhir Mahasiswa
dengan Pendekatan *Cosine Similarity* dan *Genetic Algorithm*

Nama : Sandra Berliana Putri

NIM : G1401221021

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:

Dr. Agus Mohamad Soleh, S.Si., M.T.

Pembimbing 2:

Gerry Alfa Dito, S.Si., M.Si.

Diketahui oleh

Ketua Program Studi:

Dr. Bagus Sartono, S.Si., M.Si.

NIP 19780411 200501 1 002

Tanggal Ujian:
24 Juni 2026

Tanggal Lulus:

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah Subhanahu wa Ta'ala atas segala rahmat, karunia, dan hidayah-Nya sehingga penulis dapat menyelesaikan karya ilmiah yang berjudul “Alokasi Dosen Pembimbing Pertama Tugas Akhir Mahasiswa dengan Pendekatan *Cosine Similarity* dan *Genetic Algorithm*”. Penyusunan karya ilmiah ini tidak lepas dari bantuan, bimbingan, arahan, serta dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan apresiasi dan terima kasih yang tulus kepada:

1. Bapak Dr. Agus Mohamad Soleh, S.Si., M.T. dan Bapak Gerry Alfa Dito, S.Si., M.Si. selaku dosen pembimbing yang telah meluangkan waktu, memberikan bimbingan, arahan, dan ilmu selama proses penyusunan karya ilmiah ini;
2. Bapak Dr. Septian Rahardiantoro, S.Stat., M.Si. selaku penguji luar komisi, Ibu Dr. Yenni Angraini, S.Si., M.Si. selaku moderator seminar hasil, dan Ibu Aulia Rizki Firdawanti, S.Stat., M.Si. selaku moderator kolokium yang telah memberikan saran dan masukan dalam penulisan karya ilmiah ini;
3. Bapak Dr. Farit Mochamad Afendi, S.Si., M.Si. dan Teh Ika Purmatasari yang telah membantu dalam penyediaan data penelitian;
4. Seluruh dosen dan tenaga pendidik Program Studi Statistika dan Sains Data IPB University yang telah memberikan ilmu dan dukungan administratif selama masa perkuliahan;
5. Bapak Benny Oktavian, Ibu Hefni Gusnita, dan adik penulis, yaitu Glenn Marcellino Alenski, beserta seluruh keluarga besar yang selalu memberikan doa, semangat, kasih sayang, dukungan, dan nasihat dalam setiap langkah perjalanan penulis;
6. Fadlur Rizki, Kak Okta, Arbay, Oci, Rai, Yuli, Rifa, Caca, Delia, Shalma, Afi, Nafisa, Elke, Biki, Qonita, Ulfah, Husna, Bintang, Jodi, Haidar, Mahib, dan Kaysa yang telah banyak membantu dan menjadi bagian penting dalam perjalanan studi penulis;
7. DA Warriors, DA Alliance, Visdat Team 6, Fanpage Tubagoez, RBM Pak Be, dan Bodo Amat yang telah menemani dan memberi semangat dalam perjalanan studi penulis;
8. Marinestic Statistika 59 yang sudah kebersamai perkuliahan dan perjuangan penulisan karya ilmiah selama 4 tahun ini; serta
9. Seluruh pihak yang telah mendukung yang tidak dapat disebutkan satu per satu.

Semoga karya ilmiah ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan, khususnya di bidang statistika, sains data, dan optimasi.

Bogor, Juni 2026

Sandra Berliana Putri



DAFTAR ISI

DAFTAR TABEL	ix
DAFTAR GAMBAR	ix
DAFTAR LAMPIRAN	ix
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
1.4 Ruang Lingkup	3
TINJAUAN PUSTAKA	4
2.1 Tokenisasi N-gram	4
2.2 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	4
2.3 <i>Cosine Similarity</i>	5
2.4 Masalah Alokasi dan Metode Penalti	5
2.5 <i>Genetic Algorithm (GA)</i>	5
III METODE	10
3.1 Data	10
3.2 Prosedur Analisis Data	11
IV HASIL DAN PEMBAHASAN	16
4.1 <i>Preprocessing Data</i>	16
4.2 Tokenisasi N-gram	17
4.3 Pembobotan Token menggunakan TF-IDF	17
4.4 Perhitungan Kesesuaian Topik dengan <i>Cosine Similarity</i>	17
4.5 Evaluasi GA dengan 8 Skenario dan 3 Skema N-gram	18
4.6 Implementasi <i>Dashboard</i>	20
V SIMPULAN DAN SARAN	23
5.1 Simpulan	23
5.2 Saran	23
DAFTAR PUSTAKA	24
LAMPIRAN	26
RIWAYAT HIDUP	29

DAFTAR TABEL

1	Contoh pembentukan token pada unigram, bigram, dan unibigram	4
2	Istilah dalam GA	6
3	Peubah yang digunakan	10
4	Contoh data mahasiswa	10
5	Contoh data dosen	11
6	Skema pembobotan penalti	15
7	Skenario GA	15
8	Contoh tahapan <i>preprocessing</i> teks pada topik penelitian mahasiswa	16
9	Contoh tahapan <i>preprocessing</i> teks pada bidang penelitian dosen	16
10	Contoh hasil tokenisasi dengan unigram, bigram, dan unibigram	17
11	Contoh hasil pembobotan TF-IDF pada skema unibigram	17
12	Tiga dosen dengan nilai <i>cosine similarity</i> tertinggi untuk mahasiswa dengan topik penelitian “machine learning” pada setiap skema n-gram	18
13	Nilai maksimum jumlah mahasiswa bimbingan setiap dosen hasil alokasi GA	18
14	Nilai <i>fitness</i>	19
15	Perbandingan komponen penyusun nilai <i>fitness</i> pada S ₇ -unibigram dan S ₆ -unibigram	19
16	Waktu komputasi (detik)	20

DAFTAR GAMBAR

1	Ilustrasi populasi, kromosom, dan gen (Mayyani <i>et al.</i> 2023)	6
2	Tampilan unggah data mahasiswa ketika kapasitas tidak berubah	21
3	Tampilan unggah data kapasitas baru dan data mahasiswa	21
4	Tampilan proses alokasi dosen pembimbing pertama	22
5	Hasil akhir alokasi dosen pembimbing pertama	22

DAFTAR LAMPIRAN

1	Lampiran 1 Rekapitulasi jumlah judul penelitian setiap dosen dari Google Scholar tahun 2021—2026	27
2	Lampiran 2 Data bidang minat dan judul penelitian dosen dengan nilai <i>cosine similarity</i> tertinggi untuk mahasiswa dengan topik penelitian “machine learning”	28



@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.