

KLASIFIKASI SENTIMEN DENGAN INDOBERTWEET DAN PEMODELAN BERTOPIC ISU PROGRAM MAGANG BERDAMPAK PADA PLATFORM-X

AURALLY BUDI ARINI



PROGRAM STUDI SARJANA 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.



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Klasifikasi Sentimen Dengan Indobertweet Dan Pemodelan Bertopic Isu Program Magang Berdampak Pada Platform-X” 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.

Dalam penyusunan karya ini, penulis menggunakan bantuan kecerdasan buatan ChatGPT untuk mencari ide serta membantu penyusunan dan pengembangan kode program yang digunakan dalam proses analisis data. Setelah menggunakan alat/layanan tersebut, penulis meninjau dan menyunting konten sesuai kebutuhan serta bertanggung jawab penuh atas isi karya tugas akhir ini.

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

Bogor, Agustus 2026

Aurally Budi Arini
G1401221088



ABSTRAK

AURALLY BUDI ARINI. Klasifikasi Sentimen Dengan Indobertweet Dan Pemodelan Bertopic Isu Program Magang Berdampak Pada Platform-X. Dibimbing oleh KUSMAN SADIK dan CICI SUHAENI.

Program Magang Berdampak 2025 menimbulkan berbagai diskusi di media sosial, khususnya pada platform X, tempat mahasiswa aktif membagikan opini dan pengalaman mereka terkait program tersebut. Penelitian ini bertujuan untuk mengklasifikasikan sentimen publik serta mengidentifikasi topik-topik dominan yang muncul dalam pembahasan mengenai Program Magang Berdampak 2025 menggunakan pendekatan klasifikasi sentimen dan pemodelan topik. Klasifikasi sentimen dilakukan menggunakan IndoBERTweet, yaitu model transformer prelatih yang dirancang khusus untuk data Twitter berbahasa Indonesia. Sementara itu, pemodelan topik dilakukan menggunakan *Improved BERTopic* yang mengintegrasikan *embedding* IndoBERTweet, reduksi dimensi menggunakan *Uniform Manifold Approximation and Projection* (UMAP), *clustering* menggunakan *Balanced Iterative Reducing and Clustering using Hierarchies* (BIRCH), serta representasi topik menggunakan pembobotan BM25. Hasil klasifikasi sentimen menunjukkan bahwa model IndoBERTweet memperoleh nilai *balanced accuracy* sebesar 0,77 yang mengindikasikan performa yang baik dalam mengklasifikasikan sentimen. Model *Improved BERTopic* berhasil menghasilkan 9 topik. Hasil penelitian menunjukkan bahwa mahasiswa memiliki ketertarikan yang tinggi terhadap Program Magang Berdampak 2025. Namun, diskusi masih didominasi oleh permasalahan terkait sistem pendaftaran, lowongan, aksesibilitas informasi, serta munculnya jasa joki CV. Temuan ini diharapkan dapat mendukung evaluasi program dan menjadi bahan pertimbangan dalam pengembangan program magang pendidikan di masa mendatang.

Kata Kunci: BERTopic, IndoBERTweet, Klasifikasi Sentimen, Pemodelan Topik, Program Magang Berdampak



ABSTRACT

AURALLY BUDI ARINI. Sentiment Classification Using IndoBERTweet and BERTopic Modeling of the Magang Berdampak Program Issues on X Platform. Supervised by KUSMAN SADIK and CICI SUHAENI.

The Magang Berdampak 2025 program has generated extensive discussions on social media, particularly on the X platform, where students actively share their opinions and experiences regarding the program. This study aims to classify public sentiment and identify the dominant topics emerging from discussions on the Magang Berdampak 2025 program using sentiment classification and topic modeling approaches. Sentiment classification was performed using IndoBERTweet, a pre-trained transformer model specifically designed for Indonesian Twitter data. Meanwhile, topic modeling was conducted using an Improved BERTopic framework that integrates IndoBERTweet embeddings, dimensionality reduction through Uniform Manifold Approximation and Projection (UMAP), clustering with Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH), and topic representation based on BM25 weighting. The sentiment classification results show that the IndoBERTweet model achieved a balanced accuracy of 0.77, indicating good performance in classifying sentiment. The Improved BERTopic model successfully identified nine distinct topics. The findings reveal that students demonstrate a high level of interest in the Magang Berdampak 2025 program. However, the discussions are still dominated by issues related to the registration system, internship vacancies, information accessibility, and the emergence of CV-writing services. These findings are expected to support program evaluation and provide valuable insights for the future development of educational internship programs.

Keywords: BERTopic, IndoBERTweet, Magang Berdampak Program, Sentiment Classification, Topic Modeling



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 2026¹
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.

KLASIFIKASI SENTIMEN DENGAN INDOBERTWEET DAN PEMODELAN BERTOPIC ISU PROGRAM MAGANG BERDAMPAK PADA PLATFORM-X

AURALLY BUDI ARINI

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

**PROGRAM STUDI SARJANA 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 —

Perpustakaan IPB University

Tim Penguji pada Ujian Skripsi:

1. Laily Nissa Atul Mualifah, S.Si., M.Si

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 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



Judul Skripsi : Klasifikasi Sentimen Dengan Indobertweet Dan Pemodelan
Bertopic Isu Program Magang Berdampak Pada Platform-X

Nama : Aurally Budi Arini
NIM : G1401221088

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:

Dr. Kusman Sadik, S.Si., M.Si.

Pembimbing 2:

Cici Suhaeni, S.Si., M.Si., Ph.D.

Diketahui oleh

Ketua Program Studi:

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

NIP 197804112005011002

Tanggal Ujian:
3 Agustus 2026

Tanggal Lulus:



@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



PRAKATA

Puji dan syukur penulis panjatkan kepada Allah *subhanaahu wa ta'ala* atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan Januari 2026 sampai bulan Juli 2026 ini ialah analisis teks, dengan judul “Klasifikasi Sentimen Dengan Indobertweet Dan Pemodelan Bertopic Isu Program Magang Berdampak Pada Platform-X”.

Penulisan karya ilmiah ini tidak lepas dari dukungan dan bimbingan berbagai pihak sehingga karya ilmiah dapat terselesaikan dengan baik. Oleh karenanya, penulis mengucapkan terima kasih kepada:

1. Ibu Ranie, Bapak Arianto, Ghia, Audi, Tante Ria, Om Jovie, Giani, dan Aurora selaku keluarga penulis yang sudah kebersamai dan mendukung penulis;
2. Dr. Kusman Sadik, S.Si., M.Si. dan Ibu Cici Suhaeni, S.Si., M.Si., Ph.D. selaku komisi pembimbing yang telah membimbing penulis selama penyusunan karya ilmiah ini;
3. Ibu Laily Nissa Atul Mualifah, S.Si., M.Si selaku dosen penguji luar pada ujian akhir studi;
4. Seluruh dosen dan tenaga pendidik Program Studi Statistika dan Sains Data yang telah memberi ilmu bermanfaat dan menunjang segala kebutuhan penulis selama perkuliahan dan penyusunan karya ilmiah ini;
5. Keyzha, Naiya, Jasmita, Yumna, dan Avrel yang sudah kebersamai penulis dalam menempuh program studi Statistika dan Sains Data;
6. Eci, Rakha, Clarizsa, Jesa, Laras, Nita, Olivia, Aisha, Nathania, dan Sanaya selaku teman penulis yang sudah memberikan dukungan;
7. Marinestic Statistika 59 yang sudah kebersamai perkuliahan dan perjuangan penulisan karya ilmiah selama hampir 4 tahun ini; dan
8. Seluruh pihak yang telah mendukung yang tidak dapat disebutkan satu persatu.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Agustus 2026

Aurally Budi Arini



@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.



DAFTAR ISI

DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiii
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	2
TINJAUAN PUSTAKA	3
2.1 Program Magang Berdampak	3
2.2 Klasifikasi Sentimen dengan IndoBERTweet	3
2.3 Pemodelan Topik dengan BERTopic	5
III METODE	10
3.1 Data	10
3.2 Tahapan Analisis Data	10
IV HASIL DAN PEMBAHASAN	15
4.1 Hasil Praproses dan Pelabelan Data	15
4.2 Eksplorasi Data	18
4.3 Klasifikasi Sentimen	21
4.4 Pemodelan Topik dengan <i>Improved</i> BERTopic	24
4.5 Integrasi IndoBERTweet dan <i>Improved</i> BERTopic	30
V SIMPULAN DAN SARAN	32
5.1 Simpulan	32
5.2 Saran	32
DAFTAR PUSTAKA	33
LAMPIRAN	35
RIWAYAT HIDUP	41

DAFTAR TABEL

1	Kriteria pelabelan data	10
2	<i>Confusion Matrix</i>	12
3	<i>Dataset</i> yang digunakan	15
4	Contoh cuitan tidak relevan	15
5	Hasil <i>cleaning</i> dan normalisasi teks	17
6	Kandidat <i>hyperparameter</i>	22
7	Hasil evaluasi model akhir pada data uji	23
8	Hasil klasifikasi model akhir	23
9	Hasil optimasi nilai <i>threshold</i>	26
10	Penentuan nilai <i>threshold</i> dengan metode <i>Rank Aggregation</i>	26
11	Hasil pemodelan topik dengan <i>Improved BERTopic</i>	27

DAFTAR GAMBAR

1	Ilustrasi <i>pre-training</i> dan <i>fine-tuning</i> BERT (Devlin <i>et al.</i> 2019)	3
2	Diagram alir tahapan analisis	14
3	Distribusi hasil pelabelan sentimen secara manual	18
4	Visualisasi word cloud hasil pelabelan (a) positif, (b) netral, dan (c) negatif	19
5	Top 5 kata unigram atau bigram pada sentimen positif dan negatif	19
6	Grafik tren bulanan jumlah cuitan beserta sentimennya	20
7	Grafik proporsi sentimen setiap bulan	20
8	Distribusi jumlah kata	21
9	Sebaran sentimen data latih dan data uji	21
10	<i>Confusion matrix</i> hasil klasifikasi dengan IndoBERTweet	24
11	Visualisasi <i>similarity matrix</i>	29
12	Distribusi sentimen setiap topik oleh label manual (a) dan prediksi (b)	30

DAFTAR LAMPIRAN

1	Lampiran 1 Kamus normalisasi tambahan	36
2	Lampiran 2 Distribusi sentimen label manual (<i>ground truth</i>) dan prediksi pada setiap topik	40



@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.