

PERBANDINGAN PERFORMA ALGORITMA *CLUSTERING* BERBASIS PARTISI DAN KEPADATAN PADA BERTOPIC ISU MAKAN BERGIZI GRATIS DI MEDIA SOSIAL X

RAIHAN SABILLAH ADISECHA



**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 “Perbandingan Performa Algoritma *Clustering* Berbasis Partisi dan Kepadatan pada BERTopic Isu Makan Bergizi Gratis di Media Sosial 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.

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

Bogor, Agustus 2026

Raihan Sabillah Adisecha
G1401221082



ABSTRAK

RAIHAN SABILLAH ADISECHA. Perbandingan Performa Algoritma *Clustering* Berbasis Partisi dan Kepadatan pada BERTopic Isu Makan Bergizi Gratis di Media Sosial X. Dibimbing oleh ANWAR FITRIANTO dan AKBAR RIZKI.

Perkembangan statistika dan sains data mendorong pengembangan *topic modelling* untuk mengidentifikasi tema utama dari data teks tidak terstruktur berskala besar, salah satunya BERTopic yang memanfaatkan *transformer embeddings* untuk menghasilkan topik yang lebih koheren. Penelitian ini menerapkan BERTopic pada 13.142 data opini publik terkait Program Makan Bergizi Gratis (MBG) di media sosial X periode 6 Januari 2025–6 Januari 2026, sekaligus membandingkan performa algoritma K-Means, K-Medoids, DBSCAN, dan HDBSCAN. Data diproses menggunakan *embedding* IndoSBERT-large, direduksi dengan UMAP, dikelompokkan dengan keempat algoritma yang dioptimasi menggunakan grid search, dan dievaluasi menggunakan *coherence score* sebagai metrik utama serta *diversity score* dan persentase *noise* sebagai metrik tambahan. Hasil menunjukkan algoritma berbasis kepadatan lebih unggul dibandingkan algoritma berbasis partisi, dengan DBSCAN sebagai yang terbaik (*coherence score* 0,711, *diversity score* 0,933, *noise* 0,495%), menegaskan efektivitasnya untuk BERTopic pada teks berbahasa Indonesia dalam mengungkap isu utama diskursus publik mengenai MBG.

Kata kunci: *BERTopic*, *clustering*, *coherence score*, DBSCAN, Makan Bergizi Gratis

ABSTRACT

RAIHAN SABILLAH ADISECHA. Performance Comparison of Partition and Density-Based *Clustering* Algorithms on BERTopic Issue of Free Nutritious Meals on Social Media X. Supervised by ANWAR FITRIANTO and AKBAR RIZKI.

The advancement of statistics and data science has driven the development of topic modeling, particularly BERTopic, which leverages *transformer embeddings* to produce more coherent topics than traditional methods. This study applies BERTopic to 13,142 public opinion posts on the Free Nutritious Meal (Makan Bergizi Gratis, MBG) program from platform X, collected between 6 January 2025 and 6 January 2026, and compares K-Means, K-Medoids, DBSCAN, and HDBSCAN *clustering* algorithms. Data were processed using IndoSBERT-large *embeddings*, reduced with UMAP, *clustered* using the four grid-search-optimized algorithms, and evaluated using *coherence score* as the primary metric, with diversity and noise percentage as supporting metrics. Results show density-based algorithms outperform partition-based ones, with DBSCAN achieving the best performance (coherence = 0.711, diversity = 0.933, *noise* = 0.495%), confirming its effectiveness for BERTopic on Indonesian-language text in uncovering major issues in public discourse on MBG.

Keywords: *BERTopic, clustering, coherence score, DBSCAN, Free Nutritious Meal Program.*



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.

PERBANDINGAN PERFORMA ALGORITMA *CLUSTERING* BERBASIS PARTISI DAN KEPADATAN PADA BERTOPIC ISU MAKAN BERGIZI GRATIS DI MEDIA SOSIAL X

RAIHAN SABILLAH ADISECHA

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

Tim Penguji pada Ujian Skripsi:
Gerry Alfa Dito, 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.



Judul Skripsi : Perbandingan Performa Algoritma *Clustering* Berbasis Partisi dan Kepadatan pada BERTopic Isu Makan Bergizi Gratis di Media Sosial X

Nama : Raihan Sabillah Adisecha

NIM : G1401221082

Disetujui oleh

Pembimbing 1:

Dr. Anwar Fitrianto, M.Sc

Pembimbing 2:

Akbar Rizki, M.Si

Diketahui oleh

Ketua Program Studi:

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

NIP 197804112005011002

Tanggal Ujian:
22 Juli 2026

Tanggal Lulus:



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 *Topic modelling*, dengan judul “Perbandingan Performa Algoritma *Clustering* Berbasis Partisi dan Kepadatan pada BERTopic Isu Makan Bergizi Gratis di Media Sosial X”.

Puji dan syukur penulis panjatkan ke hadirat Allah Swt. atas segala rahmat, karunia, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini. Dalam proses penyusunan skripsi ini, penulis memperoleh banyak bantuan, bimbingan, dukungan, serta doa dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih kepada:

1. Dr. Anwar Fitrianto, M.Sc. dan Akbar Rizki, M.Si. selaku dosen pembimbing yang telah memberikan bimbingan, arahan, masukan, serta saran selama proses penyusunan skripsi.
2. Dr. Anwar Fitrianto, M.Sc. selaku dosen pembimbing akademik yang telah memberikan bimbingan dan arahan selama penulis menempuh pendidikan di IPB University.
3. Moderator kolokium, seminar hasil, dan penguji luar komisi pembimbing yang telah memberikan kritik, saran, dan masukan yang membangun untuk penyempurnaan skripsi ini.
4. Ayah, ibu, dan seluruh keluarga atas segala doa, kasih sayang, dukungan, serta motivasi yang tiada henti selama penulis menempuh pendidikan hingga penyusunan skripsi ini.
5. Sania Nur Rahma Firdaus, selaku seseorang terdekat yang senantiasa memberikan doa, dukungan, semangat, motivasi, serta dengan sabar mendengarkan keluh kesah dan memberikan bantuan selama proses penyusunan skripsi ini.
6. Seluruh teman-teman perkuliahan yang telah menjadi rekan berdiskusi, berbagi pengalaman, serta saling memberikan bantuan selama masa studi.
7. Seluruh staf akademik yang telah membantu kelancaran berbagai urusan administrasi selama perkuliahan.
8. Semua pihak yang telah memberikan bantuan, dukungan, dan kontribusi, baik secara langsung maupun tidak langsung, yang tidak dapat penulis sebutkan satu per satu.

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

Bogor, Agustus 2026

Raihan Sabillah Adisecha

DAFTAR ISI

DAFTAR TABEL	x
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	x
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
1.4 Manfaat	3
II TINJAUAN PUSTAKA	4
2.1 Program Makan Bergizi Gratis	4
2.2 Metode <i>Topic Modelling</i> BERTopic	4
2.3 <i>Text Embedding</i> dengan Model IndoSBERT	5
2.4 <i>Uniform Manifold Approximation and Projection</i> (UMAP)	7
2.5 K-Means	9
2.6 K-Medoids	11
2.7 <i>Density-Based Spatial Clustering of Applications with Noise</i> (DBSCAN)	13
2.8 <i>Hierarchical Density-Based Spatial Clustering of Applications with Noise</i> (HDBSCAN)	14
2.9 <i>CountVectorizer</i>	16
2.10 <i>Class-based Term Frequency–Inverse Document Frequency</i> (c-TF-IDF)	16
2.11 Evaluasi Topik: <i>Coherence Score</i> dan <i>Diversity Score</i>	16
III METODE	19
3.1 Data	19
3.2 Prosedur Analisis	19
IV HASIL DAN PEMBAHASAN	22
4.1 Data <i>Preprocessing</i>	22
4.2 Eksplorasi Data	24
4.3 Hasil <i>Topic Modelling</i> BERTopic	26
4.3.1 Hasil BERTopic Menggunakan <i>Clustering</i> K-Means	28
4.3.2 Hasil BERTopic Menggunakan <i>Clustering</i> K-Medoids	29
4.3.3 Hasil BERTopic Menggunakan <i>Clustering</i> DBSCAN	30
4.3.4 Hasil BERTopic Menggunakan <i>Clustering</i> HDBSCAN	31
4.4 Perbandingan Hasil BERTopic Setiap Algoritma <i>Clustering</i>	33
4.5 Interpretasi Hasil BERTopic dengan Algoritma <i>Clustering</i> Terbaik	34
V SIMPULAN DAN SARAN	39
5.1 Simpulan	39
5.2 Saran	39
DAFTAR PUSTAKA	41
LAMPIRAN	44
RIWAYAT HIDUP	62



DAFTAR TABEL

1	Parameter algoritma <i>clustering</i> yang digunakan	20
2	Contoh data yang diseleksi	22
3	Penerapan data <i>preprocessing</i>	24
4	Statistik deskriptif karakteristik data unggahan	24
5	Hasil embedding dan reduksi dimensi UMAP untuk tiga dokumen contoh	26
6	Hasil pengukuran kemiripan semantik D1, D2, dan D3	27
7	Sepuluh hasil terbaik optimasi <i>hyperparameter</i> K-Means	28
8	Sepuluh hasil terbaik optimasi <i>hyperparameter</i> K-Medoids	30
9	Sepuluh hasil terbaik optimasi <i>hyperparameter</i> DBSCAN	31
10	Sepuluh hasil terbaik optimasi <i>hyperparameter</i> HDBSCAN	32
11	Perbandingan <i>hyperparameter</i> terbaik setiap algoritma	33
12	Interpretasi hasil <i>topic modelling</i> BERTopic menggunakan DBSCAN	34

DAFTAR GAMBAR

1	Diagram alir metode penelitian	21
2	<i>Word cloud</i> sebelum penghapusan <i>stop words</i>	25
3	<i>Word cloud</i> setelah penghapusan <i>stop words</i>	25
4	<i>Word cloud</i> kata representatif per topik	36

DAFTAR LAMPIRAN

1	Hasil pemodelan BERTopic dengan algoritma <i>clustering</i> K-Means	45
2	Hasil pemodelan BERTopic dengan algoritma <i>clustering</i> K-Medoids	47
3	Hasil pemodelan BERTopic dengan algoritma <i>clustering</i> DBSCAN	49
4	Hasil pemodelan BERTopic dengan algoritma <i>clustering</i> HDBSCAN	59