

SKRIPSI

**PENGKLUSTERAN ARTIKEL PADA JURNAL PINTER
MENGUNAKAN ALGORITMA DBSCAN**



Disusun oleh:

Genta Abdi Fathir

1512622026

PENDIDIKAN TEKNIK INFORMATIKA DAN KOMPUTER

FAKULTAS TEKNIK

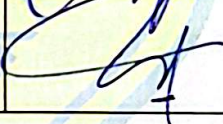
UNIVERSITAS NEGERI JAKARTA

2026

LEMBAR PENGESAHAN TUGAS AKHIR
BENTUK : SKRIPSI

Nama Mahasiswa : Genta Abdi Fathir
NIM : 1512622026
Program Studi/Fakultas : Pendidikan Teknik Informatika dan Komputer /
Teknik
Judul Penelitian : Pengklusteran Artikel Pada Jurnal PINTER
Menggunakan Algoritma DBSCAN

ini telah dipertahankan dihadapan dewan penguji dan dinyatakan **Lulus** / ~~Tidak Lulus~~ * Ujian Tugas Akhir oleh Tim Penguji pada Tanggal 9 Juli 2026
Tim Penguji,

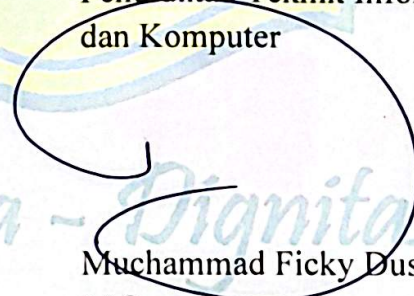
Ketua Penguji	Bambang Prasetya Adhi, S.Pd., M.Kom NIDN. 0025028303	
Anggota Penguji (Penguji Ahli 1)	Hamidillah Ajie, M.T NIDN. 0024087402	
Anggota Penguji (Penguji Ahli 2)	Wiranti Kusuma Hapsari, S.Kom., M.Cs NIDN. 0416079403	
Anggota Penguji (Pembimbing 1)	Dr. Widodo, S. Kom., M. Kom. NIDN. 0025037206	
Anggota Penguji (Pembimbing 2)	Neng Ayu Herawati, S. Pd., M. T. NUPTK. 125277667723015	

Mengetahui,
Dean Fakultas Teknik



Prof. Dr. Neneng Siti Silfi Ambarwati,
S.Sr., Apt., M.Si
NIDN. 0029027204

Koordinator Program Studi
Pendidikan Teknik Informatika
dan Komputer



Muchammad Ficky Duskarnaen,
M.Sc.
NIDN. 0024097304

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa:

1. Skripsi ini merupakan karya asli dan belum pernah diajukan untuk mendapatkan gelar akademik sarjana, baik di Universitas Negeri Jakarta maupun di Perguruan Tinggi lain.
2. Skripsi ini belum dipublikasikan kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
3. Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Universitas Negeri Jakarta.

Jakarta, 1 Juli 2026



Intelligentia - Dignitas

Genta Abdi Fathir

No. Reg 1512622026



KEMENTERIAN PENDIDIKAN TINGGI, SAINS DAN TEKNOLOGI
UNIVERSITAS NEGERI JAKARTA
PERPUSTAKAAN DAN KEARSIPAN

Jalan Rawamangun Muka Jakarta 13220

Telepon/Faksimili: 021-4894221

Laman: lib.unj.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Negeri Jakarta, yang bertanda tangan di bawah ini, saya:

Nama : Genta Abdi Fathir
NIM : 1512622026
Fakultas/Prodi : Teknik/Pendidikan Teknik Informatika dan Komputer
Alamat email : gentaabdi@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan dan Kearsipan Universitas Negeri Jakarta, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

☒ Skripsi ☐ Tesis ☐ Disertasi ☐ Lain-lain (.....)

yang berjudul :

Pengklusteran Artikel Pada Jurnal PINTER Menggunakan Algoritma DBSCAN

Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan dan Kearsipan Universitas Negeri Jakarta berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain secara **fulltext** untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Negeri Jakarta, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 8 Agustus 2026

Penulis

(Genta Abdi Fathir)

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT atas segala rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan proposal skripsi yang berjudul “Pengkusteran Artikel pada Jurnal PINTER menggunakan Algoritma DBSCAN” dengan baik dan tepat waktu. Proposal skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Pendidikan pada Program Studi Pendidikan Teknik Informatika dan Komputer, Fakultas Teknik, Universitas Negeri Jakarta.

Penulis menyadari bahwa terselesaikannya proposal skripsi ini tidak lepas dari bantuan, bimbingan, dukungan, dan doa dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Kedua orang tua tercinta, Bapak Mardaih dan Ibu Susana, serta adik tersayang, Thirza, terima kasih atas doa, kasih sayang, dukungan moral maupun materiil, serta semangat yang tiada henti diberikan kepada penulis sepanjang perjalanan akademik ini.
2. Bapak Muhammad Ficky Duskarnaen, S.T., M.Sc., selaku Koordinator Program Studi Pendidikan Teknik Informatika dan Komputer. Terima kasih atas arahan dan perhatian yang telah Bapak berikan kepada seluruh mahasiswanya untuk menyelesaikan pendidikan tepat waktu.
3. Bapak Dr. Widodo, S.Kom., M.Kom., selaku Dosen Pembimbing I yang dengan penuh keikhlasan, ketelitian, dan kesabaran telah meluangkan waktu, tenaga, dan pikiran dalam memberikan bimbingan, arahan, serta masukan yang sangat berharga bagi penulis selama proses penyusunan proposal skripsi ini.
4. Ibu Neng Ayu Herawati, S.Pd., M.T., selaku Dosen Pembimbing II yang dengan penuh perhatian, ketulusan, dan dedikasi telah meluangkan waktu, tenaga, dan pikiran dalam memberikan bimbingan, arahan, serta motivasi yang sangat berarti bagi penulis selama proses penyusunan proposal skripsi ini.

5. Seluruh Dosen dan Staf Program Studi Pendidikan Teknik Informatika dan Komputer, atas ilmu pengetahuan dan bantuan yang telah diberikan selama masa perkuliahan.
6. Teman-teman seperjuangan di Program Studi Pendidikan Teknik Informatika dan Komputer angkatan 2022, yang telah memberikan lingkungan yang suportif, semangat, bantuan, dan hiburan selama masa studi.
7. Semua pihak yang tidak dapat penulis sebutkan satu per satu, yang telah turut berkontribusi dalam penyelesaian proposal skripsi ini.

Penulis menyadari bahwa proposal skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun dari semua pihak demi perbaikan dan penyempurnaan penelitian ini. Semoga proposal skripsi ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan, khususnya di bidang text mining, pengklusteran dokumen, dan pengolahan publikasi ilmiah berbahasa Indonesia.

Jakarta, 1 Juli 2026



Genta Abdi Fathir
No. Reg 1512622026

Intelligentia - Dignitas

ABSTRAK

Genta Abdi Fathir, Pengklusteran Artikel pada Jurnal PINTER menggunakan Algoritma DBSCAN. Dosen Pembimbing: Dr. Widodo, S.Kom., M.Kom., Neng Ayu Herawati, S.Pd., M.T. Program Studi Pendidikan Teknik Informatika dan Komputer. Fakultas Teknik. Universitas Negeri Jakarta. 2026.

Penelitian ini bertujuan menganalisis performa pengklusteran artikel pada Jurnal Pendidikan Teknik Informatika dan Komputer (PINTER) berdasarkan kemiripan konten menggunakan algoritma Density-Based Spatial Clustering of Applications with Noise (DBSCAN). Dataset yang digunakan terdiri atas 160 artikel dari Jurnal PINTER Volume 1 Nomor 1 hingga Volume 8 Nomor 2, dengan fitur judul, abstrak, dan isi artikel. Representasi teks dilakukan menggunakan Word2Vec, dilanjutkan dengan reduksi dimensi menggunakan *Principal Component Analysis* (PCA) untuk menjaga kebermaknaan jarak antar dokumen. Parameter DBSCAN (eps dan min_samples) ditentukan melalui *grid search* yang dievaluasi menggunakan *Silhouette Coefficient*, *Davies Bouldin Index* (DBI), dan *Coverage*, dengan pengujian dilakukan pada tiga skenario, yaitu pengklusteran terhadap seluruh artikel (Skenario 1), per volume (Skenario 2), dan per nomor penerbitan (Skenario 3). Hasil pengujian menunjukkan Skenario 1 membentuk tujuh kluster dengan *Silhouette Coefficient* sebesar 0,7223 dan DBI sebesar 0,3847, meskipun menyisakan 73 artikel (45,6%) sebagai noise. Skenario 2 dan Skenario 3 memperoleh rata-rata *Silhouette Coefficient* yang lebih tinggi, masing-masing 0,7470 dan 0,7264, namun menghasilkan kluster yang jauh lebih kecil dan banyak (27 dan 35 kluster) sehingga sulit diinterpretasikan sebagai gambaran menyeluruh isi jurnal. Berdasarkan pertimbangan tersebut, Skenario 1 ditetapkan sebagai konfigurasi utama karena menghasilkan kluster yang lebih besar dan stabil, dengan enam kecenderungan isi yang konsisten muncul pada ketiga skenario, yaitu model pembelajaran di kelas, media pembelajaran berbasis video/multimedia, jaringan komputer, web service/sistem informasi, analisis/klasifikasi teks, dan arsitektur enterprise pada lingkungan SMK. Karakteristik media pembelajaran berbasis video merupakan yang paling dominan, membentuk kluster terbesar dengan 20 artikel.

Kata Kunci: Clustering, DBSCAN, Jurnal PINTER, *Silhouette Coefficient*, Word2Vec.

ABSTRACT

Genta Abdi Fathir, *Article Clustering on Jurnal PINTER Using the DBSCAN Algorithm*. Advisors: Dr. Widodo, S.Kom., M.Kom., Neng Ayu Herawati, S.Pd., M.T. Informatics and Computer Engineering Education Study Program. Faculty of Engineering. Universitas Negeri Jakarta. 2026.

This study aims to analyze the clustering performance of articles from Jurnal Pendidikan Teknik Informatika dan Komputer (PINTER) based on content similarity using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm. The dataset consists of 160 articles from Jurnal PINTER Volume 1 Number 1 to Volume 8 Number 2, using the title, abstract, and body of each article as features. Text representation was performed using Word2Vec, followed by dimensionality reduction using Principal Component Analysis (PCA) to preserve meaningful distances between documents. DBSCAN parameters (eps and min_samples) were determined through grid search evaluated with the Silhouette Coefficient, Davies Bouldin Index (DBI), and Coverage, tested across three scenarios: clustering the entire corpus (Scenario 1), per volume (Scenario 2), and per issue (Scenario 3). The results show that Scenario 1 formed seven clusters with a Silhouette Coefficient of 0.7223 and a DBI of 0.3847, although 73 articles (45.6%) were classified as noise. Scenarios 2 and 3 achieved higher average Silhouette Coefficients of 0.7470 and 0.7264 respectively, but produced much smaller and more numerous clusters (27 and 35 clusters), making them harder to interpret as an overall picture of the journal's content. Based on these considerations, Scenario 1 was selected as the main configuration for producing larger, more stable, and interpretable clusters, with six content tendencies consistently appearing across all three scenarios: classroom learning models, video/multimedia-based learning media, computer networks, web services/information systems, text analysis/classification, and enterprise architecture in vocational school (SMK) settings. Video-based learning media was the most dominant characteristic, forming the largest cluster with 20 articles.

Keywords: Clustering, DBSCAN, Jurnal PINTER, Silhouette Coefficient, Word2Vec.

DAFTAR ISI

LEMBAR PENGESAHAN PROPOSAL SKRIPSI	ii
LEMBAR PERNYATAAN	iii
KATA PENGANTAR.....	iv
ABSTRAK	vi
ABSTRACT.....	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL.....	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Identifikasi Masalah.....	3
1.3 Pembatasan Masalah.....	3
1.4 Rumusan Masalah.....	4
1.5 Tujuan Penelitian	4
1.6 Manfaat Penelitian	4
1.6.1 Manfaat Teoritis:.....	4
1.6.2 Manfaat Praktis:.....	4
BAB II TINJAUAN PUSTAKA	6
2.1 Kerangka Teoritik	6
2.1.1 Jurnal Ilmiah	6
2.1.2 Jurnal PINTER.....	7
2.1.3 <i>Machine Learning</i>	7
2.1.4 <i>Clustering</i>	8
2.1.5 Algoritma DBSCAN.....	9
2.1.6 Text Mining.....	11
2.1.7 <i>Text Preprocessing</i>	11
2.1.8 Representasi Teks dengan Word Embedding (Word2Vec)	14
2.1.9 <i>Principal Component Analysis (PCA)</i>	16
2.1.10 <i>Silhouette Coefficient</i>	17
2.1.11 <i>Davies Bouldin Index</i>	17

2.1.12 Knowledge Discovery in Databases (KDD).....	18
2.2 Penelitian Relevan	20
2.3 Kerangka Berpikir	28
BAB III METODOLOGI PENELITIAN.....	31
3.1 Tempat dan Waktu Penelitian	31
3.2 Alat dan Bahan Penelitian	31
3.2.1 Alat Penelitian	31
3.2.2 Bahan Penelitian	32
3.3 Langkah-Langkah Penelitian.....	33
3.4 Prosedur Penelitian.....	34
3.4.1 Pengumpulan Data.....	35
3.4.2 Text Preprocessing	35
3.4.3 Transformation	36
3.4.4 Data Mining.....	38
3.4.5 Interpretation/Evaluation	40
3.5 Teknik Pengumpulan Data.....	42
BAB IV HASIL PENELITIAN	43
4.1 Hasil Pengumpulan Data	43
4.2 Preprocessing dan Representasi Vektor.....	43
4.2.1 Text Preprocessing.....	43
4.2.2 Representasi Vektor dengan Word2Vec	46
4.2.3 Reduksi Dimensi dengan PCA dan Normalisasi	47
4.3 Grid search dan Pemilihan Parameter DBSCAN.....	48
4.3.1 Pencarian Parameter	48
4.3.2 Metrik Evaluasi.....	49
4.3.3 Logika Pemilihan Parameter Terbaik.....	50
4.4 Hasil Clustering	51
4.4.1 Skenario 1 Gabungan.....	51
4.4.2 Skenario 2 Per Volume.....	56
4.4.3 Skenario 3 Per nomor	65
4.5 Hasil Evaluasi dan Interpretasi	79
4.5.1 Perbandingan <i>Silhouette Coefficient</i> dan DBI Antar Skenario 79	
4.5.2 Penentuan Konfigurasi Terbaik	81

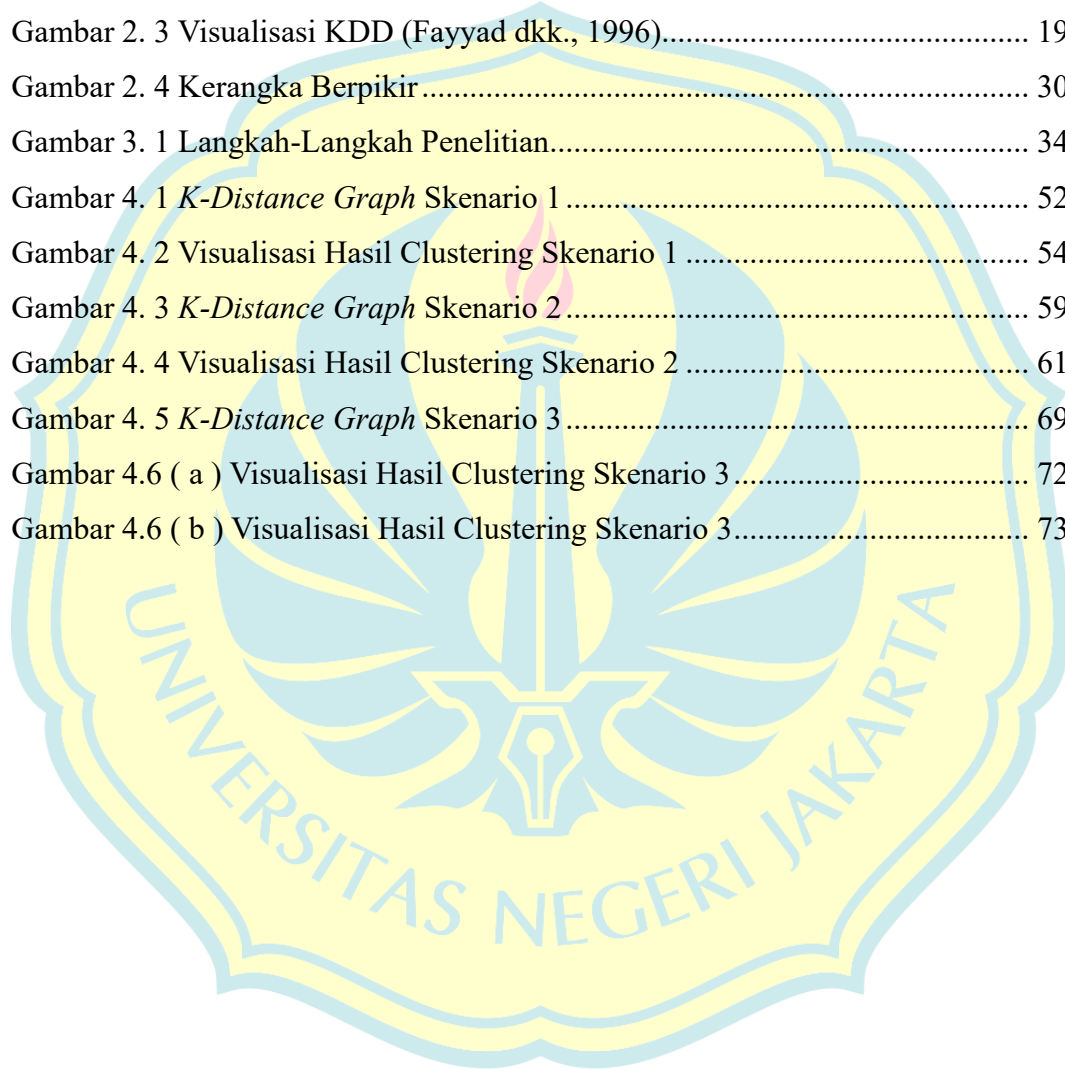
4.5.3 Interpretasi Cluster Berdasarkan Kata Dominan	82
4.6 Pembahasan	84
BAB V . KESIMPULAN DAN SARAN.....	87
5.1 Kesimpulan	87
5.2 Saran	88
DAFTAR PUSTAKA	90
LAMPIRAN.....	94



Intelligentia - Dignitas

DAFTAR GAMBAR

Gambar 2. 1 Hubungan antara core dan <i>border point</i> (Ester dkk., 1996).....	9
Gambar 2. 2 Arsitektur CBOW dan Skip-gram pada Word2Vec (Mikolov dkk., 2013)	15
Gambar 2. 3 Visualisasi KDD (Fayyad dkk., 1996).....	19
Gambar 2. 4 Kerangka Berpikir	30
Gambar 3. 1 Langkah-Langkah Penelitian.....	34
Gambar 4. 1 <i>K-Distance Graph</i> Skenario 1	52
Gambar 4. 2 Visualisasi Hasil Clustering Skenario 1	54
Gambar 4. 3 <i>K-Distance Graph</i> Skenario 2	59
Gambar 4. 4 Visualisasi Hasil Clustering Skenario 2	61
Gambar 4. 5 <i>K-Distance Graph</i> Skenario 3	69
Gambar 4.6 (a) Visualisasi Hasil Clustering Skenario 3	72
Gambar 4.6 (b) Visualisasi Hasil Clustering Skenario 3.....	73



Intelligentia - Dignitas

DAFTAR TABEL

Tabel 2.1 Alur Transformasi Teks pada Setiap Tahapan <i>Preprocessing</i>	13
Tabel 2.2 Penelitian Relevan.....	23
Tabel 3.1 Perangkat Keras.....	31
Tabel 3.2 Perangkat Lunak.....	32
Tabel 3.3 Kriteria nilai <i>Silhouette Coefficient</i> (Atira & Sari, 2023).	41
Tabel 4. 1 <i>Preprocessing</i>	45
Tabel 4. 2 Parameter Pelatihan Word2vec.....	46
Tabel 4. 3 Rentang nilai yang diuji pada setiap skenario	49
Tabel 4. 4 Parameter terbaik pada Skenario 1	51
Tabel 4. 5 Distribusi artikel Skenario 1	53
Tabel 4. 6 Kata dominan tiap klaster skenario 1	54
Tabel 4. 7 Parameter terbaik pada Skenario 2 (per volume)	57
Tabel 4. 8 Distribusi Artikel Skenario 2	60
Tabel 4. 9 Kata dominan tiap klaster skenario 2	62
Tabel 4. 10 Parameter terbaik pada Skenario 3 (per nomor).....	66
Tabel 4. 11 Distribusi Artikel Skenario 3	70
Tabel 4. 12 Kata dominan tiap klaster skenario 3	74
Tabel 4. 13 Perbandingan Rata-rata Metrik Evaluasi Antar Skenario	79
Tabel 4. 14 Karakteristik Ukuran dan Stabilitas Klaster Antar Skenario.....	80
Tabel 4. 15 Konsistensi Kemunculan Kecenderungan Isi antar Skenario	82

Intelligentia - Dignitas

DAFTAR LAMPIRAN

Lampiran 1 Lembar Pernyataan Kelayakan judul	94
Lampiran 2 Surat tugas dosen pembimbing.....	95
Lampiran 3 Lembar konsultasi	96
Lampiran 4 kartu Kehadiran sempro.....	100
Lampiran 5 Biodata peserta ujian skripsi.....	101
Lampiran 6 Surat permohonan Ujian skripsi	105
Lampiran 7 surat persetujuan dosen pembimbing skripsi.....	106
Lampiran 8 Surat persetujuan dosen pembimbing akademik	108
Lampiran 9 pra-transkrip akademik	109
Lampiran 10 Surat Pernyataan Menyelesaikan Mata Kuliah Wajib	112
Lampiran 11 Hasil <i>Preprocessing</i>	113



Intelligentia - Dignitas