

SKRIPSI

**IMPLEMENTASI *K-MEANS CLUSTERING* PADA HASIL
TOPIC MODELING MENGGUNAKAN *LATENT DIRICHLET*
ALLOCATION PADA PUBLIKASI ILMIAH DOSEN BIDANG
KOMPUTER UNIVERSITAS NEGERI JAKARTA**



MADA REKADARMA SEPTIANDA

1512621081

PROGRAM STUDI

PENDIDIKAN TEKNIK INFORMATIKA DAN KOMPUTER

FAKULTAS TEKNIK UNIVERSITAS NEGERI JAKARTA

2025

LEMBAR PENGESAHAN SKRIPSI

Judul : IMPLEMENTASI K-MEANS CLUSTERING PADA HASIL
TOPIC MODELING MENGGUNAKAN LATENT DIRICHLET
ALLOCATION PADA PUBLIKASI ILMIAH DOSEN BIDANG
KOMPUTER UNIVERSITAS NEGERI JAKARTA

Penyusun : Mada Rekadarma Septianda

NIM : 1512621081

Tanggal, 12 Mei 2025

Dosen Pembimbing I,



Dr. Widodo, M.Kom

NIP. 197203252005011002

Tanggal, 12 Mei 2025

Dosen Pembimbing II,



Murien Nugraheni, S.T, M.Cs

NIP. 198710112019032012

PENGESAHAN PANITIA UJIAN SKRIPSI

NAMA DOSEN

TANDA TANGAN

TANGGAL

Ressy Dwitias Sari, , S.T., M.T.I

Ketua Penguji



12/05/2025

Bambang Prasetya Adhi, S.Pd., M.Kom

Dosen Penguji I



12/05/2025

Diat Nurhidayat, , S.Pd., M.TI

Dosen Penguji II



12/05/2025

HALAMAN 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, 04 April 2025

Yang membuat pernyataan



Mada Rekadarma Septianda

NIM. 1512621081



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

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 : Mada Rekadarma Septianda

NIM : 1512621081

Fakultas/Prodi : Teknik/Pendidikan Teknik Informatika dan Komputer

Alamat email : mada.septianda@gmail.com

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

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

yang berjudul :

**IMPLEMENTASI *K-MEANS CLUSTERING* PADA HASIL *TOPIC MODELING*
MENGUNAKAN *LATENT DIRICHLET ALLOCATION* PADA PUBLIKASI ILMIAH
DOSEN BIDANG KOMPUTER UNIVERSITAS NEGERI JAKARTA**

Dengan Hak Bebas Royalti Non-Eksklusif ini UPT Perpustakaan 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, 19 Juli 2025

Penulis

Mada Rekadarma Septianda

KATA PENGANTAR

Puji dan syukur kehadirat Allah SWT atas berkah, rahmat, karunia, serta hidayah-Nya, sehingga penulis dapat menyelesaikan penelitian melalui skripsi ini dengan baik dan tepat waktu. Skripsi dengan judul “Implementasi *K-Means Clustering* pada Hasil *Topic Modeling* Menggunakan *Latent Dirichlet Allocation* pada Publikasi Ilmiah Dosen Bidang Komputer di Universitas Negeri Jakarta”, yang disusun sebagai syarat untuk memperoleh gelar Sarjana Pendidikan pada Program Studi Pendidikan Teknik Informatika dan Komputer, Fakultas Teknik, Universitas Negeri Jakarta.

Pada proses penyusunan skripsi dan penelitian, penulis diberkahi banyak bantuan, dukungan, serta doa dari berbagai pihak. Maka dari itu, dengan rasa hormat, penulis menyampaikan ucapan terima kasih kepada:

1. Marsudi, Endang, dan Marendra, serta Ka Hesti selaku Ayah, Ibu, dan Saudara penulis yang senantiasa memberikan dukungan berupa materi dan doa yang menyertai, serta memotivasi penulis untuk menyelesaikan skripsi tepat waktu.
2. Dr. Widodo, M.Kom., selaku Dosen Pembimbing 1 yang mampu menghargai kekurangan Mahasiswa dan selalu memberikan banyak bantuan berupa ilmu, arahan dan nasihat, serta obrolan seputar pendidikan selama proses penyusunan skripsi ini.
3. Murien Nugraheni, S.T., M.Cs., selaku Dosen Pembimbing 2 yang selalu memberikan nasihat dan dukungan penuh pada apa yang saya tuangkan di skripsi ini, serta menjadi Dosen Pembimbing seperti teman sendiri yang bisa memotivasi diri Penulis untuk terus maju.
4. Bapak dan Ibu Dosen Program Studi Pendidikan Teknik Informatika dan Komputer yang memberikan banyak hal berupa ilmu sederhana maupun kompleks untuk memberikan pandangan baru bahwa Dosen merupakan sosok yang bisa mendengarkan pendapat dan curahan hati dari Mahasiswa.

5. Luthvia Fariha selaku sosok yang memberikan banyak dukungan mental yang senantiasa menyemangati dan menemani dalam penyelesaian studi maupun skripsi yang telah dijalani.
6. Teman – teman terdekat penulis yakni Hisyam, Resky, Fathur, Bayu, Fadel, Cesar, Ido, Anan, Ical, Rizkyawan, Tegar, Tabah, Dheni, Denis yang menemani senang, dan sedih selama menjalani studi, serta Jiel Vayad Ramadhan yang senantiasa membantu penelitian dan penyusunan skripsi.
7. Teman – teman Forever Young Aldy, Alfin, Danan, Fathan yang senantiasa saling memberikan semangat untuk menyelesaikan studi perkuliahan bersama – sama.
8. Zhafran Panggomgomi, dan Fozan Bebe Moreno selaku programmer yang merealisasikan aplikasi SIMTA serta, SSO UNJ sebagai sarana penyelesaian Tugas Akhir Mahasiswa di UNJ

Penulis sadar akan kekurangan yang terdapat pada skripsi ini. Oleh karena itu, masukan dan saran diharapkan mampu membangun, serta menyempurnakan penelitian ini di masa mendatang. Akhir kata, penulis berharap bahwasannya skripsi ini dapat memberikan kontribusi, khususnya dalam pengembangan penelitian pada bidang studi Pendidikan Teknik Informatika dan Komputer.

Jakarta, 04 April 2025

Penulis,



Mada Rekadarma Septianda

NIM.1512621081

ABSTRAK

Mada Rekadarma Septianda, Implementasi *K-Means Clustering* pada Hasil *Topic Modeling* Menggunakan *Latent Dirichlet Allocation* pada Publikasi Ilmiah Dosen Bidang Komputer di Universitas Negeri Jakarta, Program Studi Pendidikan Teknik Informatika dan Komputer. Fakultas Teknik. Universitas Negeri Jakarta. 2025. Dosen Pembimbing: Dr. Widodo, M.Kom., dan Murien Nugraheni, S.T, M.Cs.

Publikasi Ilmiah Dosen memiliki peran penting dalam kontribusi di dunia pendidikan perguruan tinggi. Universitas Negeri Jakarta, sebagai salah satu lembaga pendidikan perguruan tinggi di Indonesia, berupaya meningkatkan kualitas penelitian dan publikasi ilmiah yang akan diterbitkan dalam kategori Internasional yakni *scopus*. Program Studi bidang komputer di Universitas Negeri Jakarta memiliki banyak publikasi ilmiah yang berkualitas dan sudah terindeks *scopus*. Namun, banyaknya publikasi ilmiah yang diterbitkan para dosen tidak teridentifikasi secara spesifik tema maupun subdisiplin yang sesuai dengan bidang program studinya memunculkan kompleksitas dalam mengidentifikasi hal tersebut. Oleh karena itu, tujuan penelitian ini adalah menentukan kecenderungan subdisiplin dari publikasi ilmiah dosen bidang komputer di Universitas Negeri Jakarta dengan mengidentifikasi secara menyeluruh dilakukan implementasi pembentukan klaster dengan *K-Means* dari hasil pengelompokan tema yang dijadikan *topic* menggunakan metode *Latent Dirichlet Allocation*. Pengelompokan dengan klaster menghasilkan pembagian klaster yang optimal dengan jumlah 10 klaster setiap program studinya, dengan nilai *Davies Bouldin Indeks* mendekati 0,5 dan *Silhouette Coefficient* melebihi 0,6. Hasil Evaluasi menggunakan kedua algoritma menunjukkan hasil yang konsisten dan baik sehingga, mampu memunculkan klaster yang berisikan kecenderungan *topic* dari masing – masing publikasi ilmiah dosen pada tiga program studi bidang komputer di Universitas Negeri Jakarta, berdasarkan *ACM Curricula* Tahun 2005, menunjukkan bahwa Program Studi Pendidikan Teknik Informatika dan Komputer memiliki kecenderungan publikasi ilmiah pada bidang *Information System* (IS), Program Studi Ilmu Komputer memiliki kecenderungan publikasi ilmiah pada bidang *Computer Science* (CS), dan Program Studi Sistem Teknologi dan Informasi memiliki kecenderungan publikasi ilmiah pada bidang *Computer Engineering* (CE). Kecenderungan tersebut bisa dikategorikan berdasarkan terms yang memiliki kecenderungan ke masing – masing bidang yang terdapat pada *ACM Curricula* Tahun 2005.

Kata Kunci: *K-Means*, *Latent Dirichlet Allocation*, Publikasi Ilmiah, *ACM Curricula* 2005.

ABSTRACT

Mada Rekadarma Septianda, Implementation of K-Means Clustering on Topic Modeling Results Using Latent Dirichlet Allocation in Scientific Publications of Lecturers in Computer Science at Jakarta State University, Informatics and Computer Engineering Education Study Program. Faculty of Engineering. State University of Jakarta. 2025. Supervisors: Dr. Widodo, M.Kom., and Murien Nugraheni, S.T, M.Cs.

Lecturer Scientific Publications have an important role in contributing to the world of higher education. Jakarta State University, as one of the higher education institutions in Indonesia, seeks to improve the quality of scientific research and publications that will be published in the International category, namely Scopus. The Computer Study Program at the State University of Jakarta has many quality scientific publications and has been indexed by Scopus. However, the large number of scientific publications published by lecturers who do not identify specific themes or subdisciplines that are in accordance with the field of their study program raises complexity in identifying them. Therefore, the purpose of this study is to determine the subdisciplinary tendency of scientific publications of lecturers in the field of computer at the State University of Jakarta by thoroughly identifying the implementation of cluster formation with K-Means from the results of the grouping of themes that are used as topics using the Latent Dirichlet Allocation method. Grouping by cluster results in an optimal cluster division with 10 clusters in each study program, with the Davies Bouldin Index value close to 0.5 and the Silhouette Coefficient exceeding 0.6. The results of the evaluation using both algorithms show consistent and good results so that they are able to produce clusters that contain topic trends from scientific publications in three computer science study programs at the State University of Jakarta, based on the ACM Curricula in 2005, show that the Informatics and Computer Engineering Education Study Program has a tendency to scientific publications in the fields of Information System (IS), the Computer Science Study Program has a tendency to scientific publications in the fields of Computer Science (CS), and the Technology and Information Systems Study Program has a tendency to publish scientific publications in the field of Computer Engineering (CE). These tendencies can be categorized based on terms that have a tendency to each field contained in the ACM Curricula in 2005.

Keywords: K-Means, Latent Dirichlet Allocation, Scientific Publications, ACM Curricula 2005

DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI	I
HALAMAN PERNYATAAN.....	II
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI.....	III
KATA PENGANTAR.....	IV
ABSTRAK	VI
ABSTRACT	VII
DAFTAR ISI.....	VIII
DAFTAR GAMBAR	XI
DAFTAR TABEL	XII
DAFTAR LAMPIRAN	XIII
BAB I PENDAHULUAN.....	14
1.1 Latar Belakang Masalah.....	14
1.2 Identifikasi Masalah	16
1.3 Pembatasan Masalah	17
1.4 Rumusan Masalah	17
1.5 Tujuan Penelitian.....	17
1.6 Manfaat Penelitian	17
BAB II TINJAUAN PUSTAKA.....	18
2.1 Kerangka Teoritik.....	18
2.1.1 Publikasi Ilmiah	18
2.1.2 Program Studi Bidang Komputer.....	19
2.1.3 <i>Association of Computer Machine (ACM)</i>	20
2.1.4 <i>Machine Learning</i>	21
2.1.5 <i>Supervised Learning</i>	22
2.1.6 <i>Unsupervised Learning</i>	23
2.1.7 <i>Topic Modeling</i>	24
2.1.8 <i>Latent Dirichlet Allocation</i>	26
2.1.9 <i>Clustering</i>	27
2.1.10 <i>K-Means Clustering</i>	28

2.1.11 <i>Text Preprocessing</i>	31
2.1.12 <i>Term Frequency (TF) – Inverse Document Frequency (IDF)</i>	33
2.1.13 <i>Davies Bouldin Index</i>	34
2.1.14 <i>Silhouette Coefficient</i>	37
2.2 Penelitian Relevan.....	39
2.3 Kerangka Berpikir.....	47
BAB III METODOLOGI PENELITIAN	50
3.1 Tempat dan Waktu Penelitian.....	50
3.2 Alat dan Bahan Penelitian	50
3.2.1 Alat Penelitian	50
3.2.2 Bahan Penelitian.....	51
3.3 Langkah – langkah Penelitian	52
3.4 Prosedur Penelitian.....	53
3.4.1 Pengumpulan Publikasi Ilmiah	53
3.4.2 <i>Text Preprocessing</i>	53
3.4.3 <i>Term Weighting TF-IDF</i>	55
3.4.4 <i>Latent Dirichlet Allocation</i>	56
3.4.5 <i>K-Means Clustering</i>	57
3.5 Teknik Pengumpulan Data	57
3.6 Metode Evaluasi Data	58
3.6.1 <i>Davies Bouldin Index</i>	58
3.6.2 <i>Silhouette Coefficient</i>	59
BAB IV HASIL PENELITIAN	60
4.1 Deskripsi Hasil Penelitian.....	60
4.1.1 <i>Text Preprocessing</i>	60
4.1.2 <i>Term Weighting TF-IDF</i>	70
4.1.3 <i>Latent Dirichlet Allocation</i>	74
4.1.4 <i>K-Means Clustering</i>	92
4.2 Analisis Data Penelitian	105
4.2.1 Evaluasi dengan <i>Davies Bouldin Index</i>	105
4.2.2 Evaluasi dengan <i>Silhouette Coefficient</i>	109
4.3 Pembahasan.....	113

4.4 Aplikasi Hasil Penelitian	115
BAB V KESIMPULAN DAN SARAN	117
5.1 Kesimpulan	117
5.2 Saran.....	118
DAFTAR PUSTAKA.....	119
LAMPIRAN.....	124
TENTANG PENULIS.....	142



DAFTAR GAMBAR

Gambar 2. 1 Diagram Blok <i>Supervised Learning</i>	23
Gambar 2. 2 Diagram Blok <i>Unsupervised Learning</i>	24
Gambar 2. 3 Teknis <i>topic modelling</i> (Chauhan & Shah, 2022)	25
Gambar 2. 4 Representasi model LDA (Blei et al., 2003)	26
Gambar 2. 5 Diagram Alir <i>K-Means</i>	29
Gambar 2. 6 Kerangka Berpikir	49
Gambar 3. 1 Diagram alir penelitian.....	52
Gambar 4. 1 Program <i>Text Preprocessing</i>	61
Gambar 4. 2 Deretan kata hasil <i>text preprocessing</i>	62
Gambar 4. 4 Cuplikan kode <i>Term Frequency</i>	71
Gambar 4. 5 Cuplikan kode <i>Inverse Document Frequency (IDF)</i>	71
Gambar 4. 6 Cuplikan kode <i>TF-IDF</i>	72
Gambar 4. 7 Kode untuk memproses <i>terms</i> menjadi <i>topic</i> menggunakan LDA ..	75
Gambar 4. 8 <i>Wordcloud Topic 1</i> prodi PTIK	81
Gambar 4. 9 <i>Wordcloud Topic 1</i> ILKOM.....	85
Gambar 4. 10 <i>Wordcloud Topic 1</i> prodi STI	91
Gambar 4. 11 Implementasi <i>K-Means Clustering</i>	92
Gambar 4. 12 menentukan <i>top topic</i>	92
Gambar 4. 13 <i>Heatmap</i> dari <i>Clustering</i> prodi PTIK.....	96
Gambar 4. 14 <i>Heatmap</i> dari <i>Cluster</i> prodi ILKOM.....	100
Gambar 4. 15 <i>Heatmap</i> dari <i>Clustering</i> prodi STI.....	104
Gambar 4. 16 Inisialisasi evaluasi.....	105
Gambar 4. 17 Grafik <i>Davies Bouldin Index</i> pada prodi PTIK.....	106
Gambar 4. 18 Grafik <i>Davies Bouldin Index</i> pada prodi ILKOM.....	107
Gambar 4. 19 Grafik <i>Davies Bouldin Index</i> pada prodi STI.....	108
Gambar 4. 20 Grafik <i>Silhouette Coefficient</i> prodi PTIK.....	110
Gambar 4. 21 Grafik <i>Silhouette Coefficient</i> prodi ILKOM	111
Gambar 4. 22 Grafik <i>Silhouette Coefficient</i> prodi STI	112

DAFTAR TABEL

Tabel 2. 1 Penerapan <i>Text Preprocessing</i>	32
Tabel 2. 2 Penelitian Relevan	42
Tabel 3. 1 Perangkat keras	50
Tabel 3. 2 Perangkat Lunak	51
Tabel 4. 1 Skor Hasil <i>Text Preprocessing</i> PTIK	63
Tabel 4. 2 Skor Hasil <i>Text Preprocessing</i> ILKOM	66
Tabel 4. 3 Skor Hasil <i>Text Preprocessing</i> STI	68
Tabel 4. 4 Skor <i>TF-IDF</i> PTIK	72
Tabel 4. 5 Skor <i>TF-IDF</i> ILKOM	73
Tabel 4. 6 Skor <i>TF-IDF</i> STI	74
Tabel 4. 7 Hasil pendistribusian <i>terms</i> menjadi topik pada PTIK	76
Tabel 4. 8 Hasil pendistribusian <i>terms</i> menjadi topik pada ILKOM	81
Tabel 4. 9 Hasil pendistribusian <i>terms</i> menjadi topik pada STI	86
Tabel 4. 10 Hasil klaster prodi PTIK	93
Tabel 4. 11 Hasil klaster prodi ILKOM	97
Tabel 4. 12 Hasil klaster prodi STI	101
Tabel 4. 13 Skor Penilaian <i>Davies Bouldin Index</i> prodi PTIK	105
Tabel 4. 14 Skor Penilaian <i>Davies Bouldin Index</i> prodi ILKOM	106
Tabel 4. 15 Skor Penilaian <i>Davies Bouldin Index</i> prodi STI	107
Tabel 4. 16 Skor Penilaian <i>Silhouette Coefficient</i> prodi PTIK	109
Tabel 4. 17 Skor Penilaian <i>Silhouette Coefficient</i> prodi ILKOM	110
Tabel 4. 18 Skor Penilaian <i>Silhouette Coefficient</i> prodi STI	111

DAFTAR LAMPIRAN

Lampiran 1. Publikasi Ilmiah Dosen terindeks <i>scopus</i>	124
Lampiran 2. Kode Program.....	133
Lampiran 3. Lembar Kelayakan Judul Skripsi.....	134
Lampiran 4. Surat Tugas Dosen Pembimbing	135
Lampiran 5. Lembar Konsultasi Dosen Pembimbing 1	136
Lampiran 6. Lembar Konsultasi Dosen Pembimbing 2	138
Lampiran 7. Surat Pernyataan Dosen Pembimbing 1	140
Lampiran 8. Surat Pernyataan Dosen Pembimbing 2	141

