

**PENGELOMPOKAN TOPIK ARTIKEL JURNAL SI/TI
MENGUNAKAN METODE LDA DAN *K-MEANS CLUSTERING***

**KELLIN
NIM H1101211031**

SKRIPSI



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2025**

**PENGELOMPOKAN TOPIK ARTIKEL JURNAL SI/TI
MENGUNAKAN METODE LDA DAN *K-MEANS CLUSTERING***

**KELLIN
NIM H1101211031**

Skripsi
Sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer pada Program Studi Sistem Informasi



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2025**

LEMBAR PENGESAHAN

Judul Tugas Akhir : "Pengelompokan Topik Artikel Jurnal SI/TI Menggunakan Metode LDA dan *K-Means Clustering*"
Nama Mahasiswa : Kellin
NIM : H1101211031
Jurusan/Program Studi : Sistem Informasi
Tanggal Lulus : 2 Juni 2025
SK Pembimbing : 3862/UN22.8/TD.06/2024/21 November 2024
SK Penguji : 1612/UN22.8/TD.06/2025/28 Mei 2025

Pembimbing I Dosen Pembimbing



Dian Prawira, S.T., M.Eng
NIP. 198411132015041001

Pembimbing II

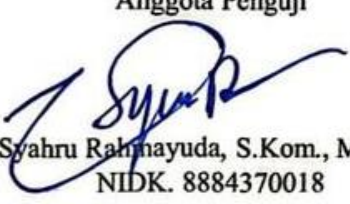

Ilhamsyah, S.Si., M.Cs
NIP. 198405102012121001

Dosen Penguji

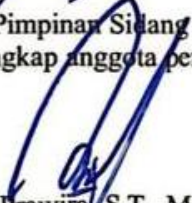
Ketua Penguji


Ferdy Febriyanto, S.Kom., M.Kom
NIP. 198902012019031008


Anggota Penguji


Syahru Rahmayuda, S.Kom., M.Kom
NIDK. 8884370018

Pimpinan Sidang
(merangkap anggota penguji)


Dian Prawira, S.T., M.Eng
NIP. 198411132015041001

Sekretaris Sidang
(merangkap anggota penguji)


Ilhamsyah, S.Si., M.Cs
NIP. 198405102012121001

Mengesahkan
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Tanjungpura


Prof. Dr. Gusrizal, S.Si., M.Si
NIP. 197108022000031001

PERNYATAAN INTEGRITAS

Saya yang bertanda tangan di bawah ini:

Nama : Kellin

NIM : H1101211031

Program Studi/ Jurusan : Sistem Informasi

Fakultas : Matematika dan Ilmu Pengetahuan Alam

dengan ini menyatakan bahwa dokumen ilmiah Tugas Akhir yang disajikan ini tidak mengandung unsur pelanggaran integritas akademik sesuai Peraturan Menteri Pendidikan, Kebudayaan, Riset, Dan Teknologi Republik Indonesia Nomor 39 Tahun 2021. Apabila di kemudian hari dokumen ilmiah Tugas Akhir ini mengandung unsur pelanggaran integritas akademik sesuai ketentuan perundangan tersebut, maka saya bersedia menerima sanksi akademik dan/atau sanksi hukum yang berlaku.

Demikian pernyataan ini untuk digunakan sebagaimana mestinya.

Pontianak, 2 Juni 2025



Kellin

H1101211031

Abstrak

Transformasi digital telah meningkatkan jumlah publikasi artikel ilmiah di bidang Sistem Informasi dan Teknologi Informasi (SI/TI), yang membuat akademisi kesulitan dalam mengidentifikasi tren topik terkini. Penelitian ini bertujuan untuk memetakan dan mengelompokkan topik artikel jurnal SI/TI berbahasa Indonesia yang dipublikasikan pada tahun 2024 dan terindeks SINTA. Data yang digunakan berupa judul dan abstrak artikel dari jurnal SINTA 1-6. Metode yang digunakan adalah *Latent Dirichlet Allocation* (LDA) untuk pemodelan topik dan *K-Means Clustering* untuk pengelompokan dokumen. Setelah melalui tahap *preprocessing* dan representasi teks menggunakan *Bag of Words*, diperoleh *coherence* tertinggi sebesar 0.5494, sehingga model LDA optimal dengan 11 topik ($\alpha = 0,11$; $\beta = 0,81$). Hasil visualisasi menunjukkan bahwa topik dominan berkaitan dengan pengembangan sistem dan aplikasi. Klasterisasi dengan *K-Means* menghasilkan dua kelompok utama, yaitu tema pengolahan data dan tema pengembangan sistem. Evaluasi dengan *silhouette score* sebesar 0.4168 menunjukkan bahwa pemisahan antar-klaster masih lemah, karena berada di antara rentang 0.25-0.5. Penelitian ini memberikan kontribusi dalam memahami distribusi topik artikel SI/TI, yang dapat dijadikan acuan dalam penentuan arah penelitian ke depan.

Kata Kunci: *Latent Dirichlet Allocation* (LDA), *K-Means Clustering*, Pemodelan Topik, Pengelompokan Dokumen, SINTA

CLUSTERING OF SI/TI JOURNAL ARTICLE TOPICS USING LDA AND K-MEANS CLUSTERING METHODS

Abstract

Digital transformation has increased the number of publications of scientific articles in the field of Information Systems and Information Technology (IS/IT), which makes it difficult for academics to identify the latest topic trends. This research aims to map and categorize the topics of Indonesian IS/IT journal articles published in 2024 and indexed by SINTA. The data used are titles and abstracts of articles from SINTA journals 1-6. The methods used are Latent Dirichlet Allocation (LDA) for topic modeling and K-Means for document clustering. After going through the preprocessing stage and text representation using Bag of Words, the highest coherence of 0.5494 is obtained, so the LDA model is optimal with 11 topics ($\alpha = 0.11$; $\beta = 0.81$). The visualization results show that the dominant topics are related to system and application development. Clustering with K-Means produces two main groups, namely data processing themes and system development themes. Evaluation with a silhouette score of 0.4168 shows that the separation between clusters is still weak, because it is between the range of 0.25-0.5. This research contributes to understanding the distribution of SI/TI article topics, which can be used as a reference in determining future research directions.

Keywords: *Latent Dirichlet Allocation (LDA), K-Means Clustering, Topic Modeling, Document Clustering, SINTA*

PRAKATA

Segala puji dan syukur penulis panjatkan ke hadirat Tuhan Yesus Kristus dan Bunda Maria atas berkat, rahmat, dan penyertaan-Nya, sebagai sumber kekuatan, pengharapan, dan ketenangan jiwa dalam setiap proses yang telah dijalani penulis sampai saat ini, sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul “Pengelompokan Topik Artikel Jurnal SI/TI Menggunakan Metode LDA dan *K-Means Clustering*”. Tugas akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer (S.Kom) pada Program Studi Sistem Informasi, Universitas Tanjungpura.

Penyusunan tugas akhir ini tidak terlepas dari bimbingan, dukungan, dan kontribusi dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Keluarga tercinta, khususnya kedua orang tua penulis yaitu Ibu Agustina Yuli dan Bapak Sugianto, atas segala doa, kasih sayang, semangat, serta dukungan moral maupun materil yang diberikan tiada henti kepada penulis. Gelar ini penulis persembahkan sebagai wujud nyata dari salah satu harapan besar kalian, yaitu melihat kedua anak kalian berhasil menyelesaikan pendidikan hingga ke jenjang sarjana. Terima kasih atas segala pengorbanan dan kesabaran yang telah menjadi fondasi utama dalam setiap langkah penulis.
2. Bapak Dian Prawira, S.T., M.Eng., selaku dosen pembimbing akademik sekaligus dosen pembimbing I, atas waktu, bimbingan, serta arahan yang telah diberikan selama proses penyusunan tugas akhir maupun bimbingan akademik.
3. Bapak Ilhamsyah, S.Si., M.Cs., selaku dosen pembimbing II, atas waktu, bimbingan, serta arahan yang telah diberikan selama proses penyusunan tugas akhir.
4. Bapak Ferdy Febriyanto, S.Kom., M.Kom., selaku dosen penguji I, atas saran dan masukan yang telah diberikan selama proses penyusunan tugas akhir.
5. Bapak Syahru Rahmayuda, S.Kom., M.Kom., selaku dosen penguji II, atas saran dan masukan yang telah diberikan selama proses penyusunan tugas akhir.

6. Bapak, Ibu, dan staf Program Studi Sistem Informasi yang telah memberikan ilmu dan pengetahuan, serta dukungan selama masa perkuliahan.
7. Teman-teman penulis yaitu Tia, Farhan, Rangga, Tiyara, Luis, Ghina, dan Mella yang telah memberikan semangat dan bantuan, serta menjadi tempat berbagi cerita selama proses perkuliahan maupun dalam penyusunan tugas akhir.
8. Teman-teman Program Studi Sistem Informasi Angkatan 2021, atas kebersamaan, bantuan, dan semangat saling mendukung selama masa perkuliahan hingga penyusunan tugas akhir.

Penulis menyadari bahwa penulisan Tugas Akhir ini masih memiliki kekurangan yang disebabkan oleh keterbatasan penulis. Oleh karena itu, apabila terdapat kekeliruan atau hal-hal yang kurang berkenan dalam penulisan ini, penulis memohon maaf yang sebesar-besarnya. Penulis juga mengharapkan kritik dan saran yang bersifat membangun demi perbaikan di masa mendatang. Semoga Tugas Akhir ini dapat memberikan manfaat, baik bagi penulis maupun bagi pihak-pihak yang berkepentingan.

Pontianak, 2 Juni 2025



Kellin
H1101211031

DAFTAR ISI

	Halaman
LEMBAR PENGESAHAN	ii
PERNYATAAN INTEGRITAS AKADEMIK.....	iii
Abstrak.....	iv
Abstract.....	v
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL	x
DAFTAR GAMBAR.....	xi
DAFTAR LISTING CODE	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian	2
1.4 Batasan Masalah.....	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II LANDASAN TEORI	6
2.1 Dasar Teori.....	6
2.1.1 Artikel Ilmiah	6
2.1.2 <i>Text Mining</i>	6
2.1.3 <i>Machine Learning</i>	7
2.1.4 <i>Topic Modeling</i>	7
2.1.5 <i>Text Preprocessing</i>	8
2.1.6 <i>Bag of Words (BoW)</i>	8
2.1.7 <i>Latent Dirichlet Allocation (LDA)</i>	9
2.1.8 <i>K-Means Clustering</i>	10
2.1.9 <i>Coherence Score</i>	11
2.1.10 <i>Elbow</i>	11
2.1.11 <i>Silhouette Coefficient</i>	12
2.2 Kajian Penelitian	13
BAB III METODOLOGI PENELITIAN	16
3.1 Bahan Penelitian.....	16
3.2 Alat Penelitian.....	16
3.3 Kerangka Kerja Penelitian	16
3.4 Data Penelitian	20
3.4.1 Objek Penelitian	20
3.4.2 Populasi dan Sampel	20
3.4.3 Jenis dan Sumber Data	20
BAB IV HASIL DAN PEMBAHASAN	21
4.1 Pengumpulan Data	21
4.2 Text Preprocessing	22
4.2.1 <i>Preprocessing Manual</i>	22
4.2.2 <i>Case Folding</i>	23
4.2.3 <i>Cleaning</i>	25

4.2.4	<i>Tokenizing</i>	28
4.2.5	<i>Stopword Removal</i>	30
4.2.6	<i>Stemming</i>	31
4.3	Bag of Words	33
4.4	Pemodelan Topik	35
4.4.1	<i>Coherence Score</i>	36
4.4.2	Pemodelan LDA	40
4.4.3	Visualisasi <i>PyLDAVis</i> dan <i>Word Cloud</i>	44
4.5	<i>Clustering</i>	64
4.5.1	<i>Elbow</i>	64
4.5.2	<i>K-Means Clustering</i>	67
4.5.3	Visualisasi <i>Cluster</i>	70
4.6	<i>Silhouette Coefficient</i>	74
BAB V PENUTUP		77
5.1	Kesimpulan	77
5.2	Saran	78
DAFTAR PUSTAKA		80

DAFTAR TABEL

	Halaman
Tabel 2. 1 Ukuran <i>Silhouette Score</i>	13
Tabel 2. 2 Penelitian terdahulu.....	13
Tabel 4. 1 Cuplikan hasil pengumpulan dataset.....	22
Tabel 4. 2 Perbandingan hasil <i>preprocessing</i> manual.....	23
Tabel 4. 3 Perbandingan hasil <i>case folding</i>	24
Tabel 4. 4 Perbandingan hasil <i>cleaning</i>	28
Tabel 4. 5 Perbandingan hasil <i>tokenizing</i>	29
Tabel 4. 6 Perbandingan hasil <i>stopword removal</i>	30
Tabel 4. 7 Perbandingan hasil <i>stemming</i>	32
Tabel 4. 8 Hasil <i>tuning Coherence Score</i>	37
Tabel 4. 9 Urutan hasil <i>tuning Coherence Score</i>	39
Tabel 4. 10 Sebaran kata per topik.....	43
Tabel 4. 11 Model LDA topik ke-0.....	45
Tabel 4. 12 Model LDA topik ke-1.....	46
Tabel 4. 13 Model LDA topik ke-2.....	48
Tabel 4. 14 Model LDA topik ke-3.....	49
Tabel 4. 15 Model LDA topik ke-4.....	51
Tabel 4. 16 Model LDA topik ke-5.....	52
Tabel 4. 17 Model LDA topik ke-6.....	54
Tabel 4. 18 Model LDA topik ke-7.....	55
Tabel 4. 19 Model LDA topik ke-8.....	57
Tabel 4. 20 Model LDA topik ke-9.....	58
Tabel 4. 21 Model LDA topik ke-10.....	60
Tabel 4. 22 Hasil interpretasi 11 topik	61
Tabel 4. 23 Hasil SSE dan selisih	66
Tabel 4. 24 Dasar interpretasi <i>cluster</i> ke-0	72
Tabel 4. 25 Dasar interpretasi <i>cluster</i> ke-1	73

DAFTAR GAMBAR

	Halaman
Gambar 2. 1 Model LDA	9
Gambar 3. 1 Kerangka kerja <i>IS Research</i>	17
Gambar 3. 2 Alur penelitian.....	19
Gambar 4. 1 Artikel jurnal bidang SI/TI.....	21
Gambar 4. 2 Representasi <i>Bag of Words</i>	34
Gambar 4. 3 Visualisasi <i>Bag of Words</i> dokumen 1	35
Gambar 4. 4 Grafik <i>Coherence Score</i>	39
Gambar 4. 5 Hasil distribusi kata per topik	42
Gambar 4. 6 Hasil distribusi topik per dokumen	43
Gambar 4. 7 Distribusi topik per dokumen dalam file CSV	44
Gambar 4. 8 Visualisasi <i>PyLDAVis</i> topik ke-0	45
Gambar 4. 9 Visualisasi <i>Word Cloud</i> topik ke-0	46
Gambar 4. 10 Visualisasi <i>PyLDAVis</i> topik ke-1	47
Gambar 4. 11 Visualisasi <i>Word Cloud</i> topik ke-1	47
Gambar 4. 12 Visualisasi <i>PyLDAVis</i> topik ke-2	48
Gambar 4. 13 Visualisasi <i>Word Cloud</i> topik ke-2	49
Gambar 4. 14 Visualisasi <i>PyLDAVis</i> topik ke-3	50
Gambar 4. 15 Visualisasi <i>Word Cloud</i> topik ke-3	50
Gambar 4. 16 Visualisasi <i>PyLDAVis</i> topik ke-4	51
Gambar 4. 17 Visualisasi <i>Word Cloud</i> topik ke-4	52
Gambar 4. 18 Visualisasi <i>PyLDAVis</i> topik ke-5	53
Gambar 4. 19 Visualisasi <i>Word Cloud</i> topik ke-5	53
Gambar 4. 20 Visualisasi <i>PyLDAVis</i> topik ke-6	54
Gambar 4. 21 Visualisasi <i>Word Cloud</i> topik ke-6	55
Gambar 4. 22 Visualisasi <i>PyLDAVis</i> topik ke-7	56
Gambar 4. 23 Visualisasi <i>Word Cloud</i> topik ke-7	56
Gambar 4. 24 Visualisasi <i>PyLDAVis</i> topik ke-8	57
Gambar 4. 25 Visualisasi <i>Word Cloud</i> topik ke-8	58
Gambar 4. 26 Visualisasi <i>PyLDAVis</i> topik ke-9	59
Gambar 4. 27 Visualisasi <i>Word Cloud</i> topik ke-9	59
Gambar 4. 28 Visualisasi <i>PyLDAVis</i> topik ke-10	60
Gambar 4. 29 Visualisasi <i>Word Cloud</i> topik ke-10	61
Gambar 4. 30 Visualisasi <i>PyLDAVis</i> keseluruhan	63
Gambar 4. 31 Visualisasi grafik <i>Elbow</i>	67
Gambar 4. 32 Cuplikan hasil klusterisasi	69
Gambar 4. 33 Jumlah dokumen di setiap <i>cluster</i>	69
Gambar 4. 34 Visualisasi <i>scatter plot</i>	70
Gambar 4. 35 Visualisasi <i>line chart cluster</i> ke-0	71
Gambar 4. 36 Visualisasi <i>line chart cluster</i> ke-1	73
Gambar 4. 37 Hasil <i>Silhouette Score</i>	75

DAFTAR LISTING CODE

	Halaman
<i>Listing Code 4. 1 Proses case folding</i>	24
<i>Listing Code 4. 2 Proses cleaning</i>	25
<i>Listing Code 4. 3 Proses tokenizing</i>	28
<i>Listing Code 4. 4 Proses stopword removal</i>	30
<i>Listing Code 4. 5 Proses stemming</i>	32
<i>Listing Code 4. 6 Proses Bag of Words</i>	33
<i>Listing Code 4. 7 Proses Coherence Score</i>	36
<i>Listing Code 4. 8 Proses visualisasi grafik Coherence Score</i>	38
<i>Listing Code 4. 9 Proses pemodelan LDA</i>	40
<i>Listing Code 4. 10 Proses Elbow</i>	64
<i>Listing Code 4. 11 Implementasi K-Means Clustering</i>	68
<i>Listing Code 4. 12 Perhitungan Silhouette Score</i>	74

BAB I

PENDAHULUAN

1.1 Latar Belakang

Transformasi digital yang pesat, menjadikan Sistem Informasi dan Teknologi Informasi (SI/TI) sebagai pilar utama dalam berbagai sektor kehidupan (N. Rahmawati & Mardiani, 2024). Perkembangan pesat ini melahirkan tren dan inovasi baru di bidang SI/TI, sehingga akademisi ikut berkontribusi melalui publikasi ilmiah. Publikasi ilmiah berupa jurnal yang digunakan sebagai media untuk menyebarluaskan artikel ilmiah. Melalui artikel ilmiah, temuan-temuan inovatif dapat divalidasi oleh komunitas ilmiah, menginspirasi penelitian selanjutnya, serta berkontribusi dalam membangun pengetahuan yang terpercaya.

Jumlah artikel ilmiah di bidang SI/TI terus meningkat setiap tahunnya, sehingga menjadi tantangan bagi akademisi dalam mengidentifikasi arah penelitian terkini. Hal tersebut dikarenakan beragamnya topik dan tingkat kompleksitas yang tinggi. Oleh karena itu, digunakan metode *Latent Dirichlet Allocation* (LDA) yang dapat membantu dalam mengeksplorasi dan memetakan topik dari artikel ilmiah tersebut, sehingga dapat memudahkan akademisi dalam melihat tren penelitian terkini.

Latent Dirichlet Allocation (LDA) mampu memperkirakan probabilitas kemunculan topik dalam kumpulan teks. LDA mengasumsikan bahwa setiap kumpulan teks terdiri dari beberapa topik laten, dan setiap topik di karakterisasi oleh distribusi kata (Kamil dkk., 2023). Sebelum diterapkan, data teks diubah ke dalam bentuk numerik menggunakan *Bag of Words* (BoW) agar dapat diproses oleh LDA (Ariansyah & Indahyanti, 2024).

Untuk memberikan gambaran umum tema penelitian berdasarkan distribusi topik yang dimilikinya, dibutuhkan proses lanjutan berupa *clustering*. *Clustering* digunakan untuk mengelompokkan data berdasarkan kemiripan karakteristiknya (Putri dkk., 2022). Penelitian ini menggunakan *K-Means Clustering*, yang bekerja dengan cara menentukan sejumlah *cluster* secara acak untuk dijadikan sebagai

centroid. Kemudian, dilakukan perhitungan jarak antara setiap data dan *centroid* hingga posisinya stabil atau tidak mengalami perubahan (Hasibuan & Puspasari, 2024).

Metode *Latent Dirichlet Allocation* (LDA) digunakan untuk menemukan topik-topik tersembunyi pada kumpulan artikel ilmiah. *Output* LDA berupa distribusi topik per dokumen digunakan sebagai *input* untuk *K-Means Clustering*, guna mengelompokkan dokumen berdasarkan kemiripan distribusi topiknya. Meskipun kombinasi kedua metode ini telah digunakan dalam penelitian sebelumnya, tetapi pemanfaatan distribusi topik dari LDA sebagai *input* untuk *K-Means Clustering* masih belum pernah dilakukan. Kombinasi ini diharapkan dapat memperlihatkan kecenderungan tema dan tren penelitian yang sedang berkembang.

1.2 Rumusan Masalah

Berdasarkan permasalahan dalam latar belakang, diperoleh rumusan masalah yaitu:

1. Bagaimana memperoleh probabilitas kemunculan topik-topik tersembunyi untuk melihat tren penelitian di bidang Sistem Informasi dan teknologi Informasi (SI/TI)?
2. Bagaimana memperoleh gambaran umum tema penelitian berdasarkan kemiripan distribusi topik yang terkandung dalam setiap artikel ilmiah?
3. Bagaimana mengevaluasi *cluster* yang dihasilkan dari metode yang digunakan dalam pengelompokan artikel jurnal Sistem Informasi dan Teknologi Informasi (SI/TI)?

1.3 Tujuan Penelitian

Tujuan yang diharapkan dapat dicapai melalui penelitian ini yaitu:

1. Melakukan pemodelan topik artikel ilmiah dari jurnal bidang SI/TI menggunakan metode *Latent Dirichlet Allocation* (LDA).
2. Melakukan pengelompokan artikel ilmiah dari jurnal bidang SI/TI menggunakan metode *K-Means Clustering*, berdasarkan *output* dari metode *Latent Dirichlet Allocation* (LDA).

3. Mengadopsi *Silhouette Coefficient* untuk mengukur kualitas pemisahan antar anggota dalam setiap *cluster* yang dihasilkan dari metode *K-Means Clustering*.

1.4 Batasan Masalah

Adapun batasan masalah yang terdapat dalam penelitian ini yaitu:

1. Fokus pada artikel ilmiah dari jurnal bidang Sistem Informasi dan Teknologi Informasi (SI/TI) yang terindeks SINTA.
2. Fokus pada artikel ilmiah yang diterbitkan pada tahun 2024.
3. Bagian yang akan diolah dari artikel ilmiah adalah judul dan abstrak.
4. Fokus pada artikel yang judul dan abstraknya menggunakan bahasa Indonesia.
5. Menggunakan metode *Latent Dirichlet Allocation (LDA)* untuk melakukan pemodelan topik.
6. Menggunakan *K-Means Clustering* untuk melakukan pengelompokan artikel.
7. Menggunakan *Bag of Words (BoW)* untuk mengubah data ke dalam bentuk numerik.
8. Menggunakan *Coherence Score* untuk mengevaluasi model LDA dan menentukan parameter optimal.
9. Menggunakan *Elbow* untuk menentukan jumlah *cluster* optimal.
10. Menggunakan *Silhouette Coefficient* untuk mengevaluasi hasil *cluster*.
11. Menggunakan *Google Colaboratory* untuk menjalankan kode program.
12. Bahasa pemrograman yang digunakan adalah *Python*.

1.5 Manfaat Penelitian

Penelitian ini diharapkan mampu memberikan manfaat bagi peneliti, bidang ilmu sistem informasi, dan akademisi, dengan manfaat antara lain:

1. Bagi Peneliti

Manfaat yang diharapkan dari penelitian ini bagi peneliti yaitu dapat memahami secara mendalam tentang konsep pemodelan topik menggunakan *Latent Dirichlet Allocation* dan pengelompokan data atau *clustering* menggunakan *K-Means Clustering*.

2. Bagi Bidang Ilmu Sistem Informasi

Manfaat yang diharapkan dari penelitian ini bagi bidang ilmu sistem informasi yaitu dapat menjadi bahan pustaka atau referensi untuk penelitian selanjutnya yang juga mengambil topik terkait pemodelan topik dan pengelompokan data atau *clustering*.

3. Bagi Pembaca

Manfaat yang diharapkan dari penelitian ini bagi pembaca, khususnya akademisi yaitu dapat memberikan referensi atau inspirasi dalam merancang penelitian yang relevan dengan perkembangan terkini di bidang SI/TI.

1.6 Sistematika Penulisan

Sistematika penulisan laporan tugas akhir disusun berdasarkan pedoman penulisan ilmiah pada Program Studi Sistem Informasi. Laporan tugas akhir terdiri dari 5 bab yang saling berkaitan dan tersusun secara sistematis. Adapun sistematika penulisan laporan tugas akhir sebagai berikut:

BAB I PENDAHULUAN

Bab ini memuat uraian mengenai latar belakang penelitian, rumusan masalah, tujuan yang ingin dicapai dari penelitian, batasan ruang lingkup tugas akhir, manfaat penelitian, serta penjelasan mengenai sistematika penulisan laporan tugas akhir secara keseluruhan.

BAB II LANDASAN TEORI

Bab ini berisi pembahasan mengenai teori-teori yang mendasari penelitian, mencakup kajian pustaka, referensi ilmiah, serta penelitian terdahulu yang relevan sebagai acuan dalam mendukung proses analisis dalam penelitian ini.

BAB III METODOLOGI PENELITIAN

Bab ini menguraikan pendekatan dan tahapan penelitian secara sistematis, serta alat dan bahan yang digunakan dalam melakukan pemodelan topik menggunakan *Latent Dirichlet Allocation* (LDA) dan pengelompokan menggunakan *K-Means Clustering*.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil pengelompokan artikel jurnal bidang SI/TI menggunakan metode *Latent Dirichlet Allocation* (LDA) dan *K-Means Clustering*. Pembahasan meliputi tahapan pengumpulan data, pembersihan data (*preprocessing*), membuat korpus dengan *Bag of Words* (BoW), penerapan *hyperparameter tuning* menggunakan *coherence score* untuk memperoleh parameter optimal, pemodelan topik menggunakan LDA, menginterpretasikan topik berdasarkan visualisasi menggunakan *word cloud* dan *PyLDAVis*, melakukan klasterisasi menggunakan *K-Means Clustering* dengan bantuan *elbow* untuk penentuan jumlah *cluster* optimal, menginterpretasi *cluster* berdasarkan visualisasi menggunakan *scatter plot* dan *line chart*, serta mengevaluasi hasil *cluster* menggunakan *silhouette coefficient*.

BAB V PENUTUP

Bab terakhir ini memuat kesimpulan dari hasil penelitian yang telah dilakukan, serta saran yang dapat diberikan berdasarkan temuan yang diperoleh. Selain itu, disampaikan pula rekomendasi untuk arah penelitian selanjutnya.