

**IMPLEMENTASI MACHINE LEARNING UNTUK TOPIC
MODELING PADA APLIKASI DONGENG NUSANTARA
PANJI KEDIRI**

SKRIPSI

Diajukan Guna Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer (S. Kom.)
Pada Program Studi Teknik Informatika



Oleh :

Yosan Okta Odhianto
NPM : 2113020137

**FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS NUSANTARA PGRI KEDIRI
2025**

Skripsi oleh:

Yosan Okta Odhianto
NPM : 2113020137

Judul :

**IMPLEMENTASI MACHINE LEARNING UNTUK TOPIC MODELING
PADA APLIKASI DONGENG NUSANTARA PANJI KEDIRI**

Telah Disetujui Untuk Diajukan Kepada Panitia Ujian/Sidang Skripsi
Program Studi Teknik Informatika Fakultas Teknik dan Ilmu Komputer
Universitas Nusantara PGRI Kediri

Tanggal : 16 Juli 2025

Pembimbing I



Daniel Swanjaya, M.Kom.
NIDN. 0723098303

Pembimbing II



Julian Sahertian, S.Pd. M.T.
NIDN. 0707079001

Skripsi oleh:

Yosan Okta Odhianto
NPM : 2113020137

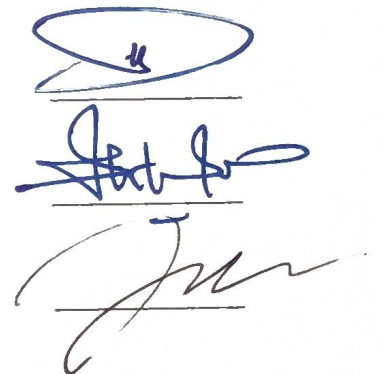
Judul :

**IMPLEMENTASI MACHINE LEARNING UNTUK TOPIC MODELING
PADA APLIKASI DONGENG NUSANTARA PANJI KEDIRI**

Telah dipertahankan di depan Panitia Ujian/Sidang Skripsi
Program Studi Teknik Informatika Fakultas Teknik dan Ilmu Komputer
Universitas Nusantara PGRI Kediri
Pada tanggal : 16 Juli 2025
Dan Dinyatakan telah Memenuhi Syarat

Panitia Penguji :

1. Ketua : Daniel Swanjaya, M.Kom
2. Penguji I : Intan Nur Farida, M.Kom
3. Penguji II : Julian Sahertian, S.Pd, M.T



Mengetahui,
Dekan FTIK



Dr. Sulistiono M.Si
NIDN. 0007076801

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini saya,

Nama : Yosan Okta Odhianto
Jenis Kelamin : Laki-laki
Tempat/Tgl Lahir : Kediri / 20 Oktober 2002
NPM : 2113020137
Fakultas/Prodi : Teknik dan Ilmu Komputer/ Teknik Informatika

Menyatakan dengan sebenarnya, bahwa dalam Skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi, dan sepanjang pengetahuan saya tidak terdapat karya tulis atau pendapat yang pernah diterbitkan oleh orang lain, kecuali yang secara sengaja dan tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Kediri, 16 Juli 2025
Yang Menyatakan



Yosan Okta Odhianto

NPM : 2113020137

HALAMAN PERSEMBAHAN

Penulisan skripsi ini dengan tulus saya dedikasikan kepada:

1. Kedua orang tua saya, yang dengan penuh kesabaran senantiasa mendoakan, memberikan dukungan terbaik, serta menjadi sumber motivasi tak henti-hentinya dalam perjalanan menyelesaikan skripsi ini.
2. Adik saya, yang selalu memberikan semangat dan dukungan moral, sehingga saya dapat menyelesaikan proses ini dengan baik.
3. Seluruh dosen Universitas Nusantara PGRI Kediri, yang telah memberikan bimbingan, ilmu, dan pelajaran berharga, baik dalam ranah akademik maupun kehidupan sehari-hari.
4. Teman-teman seperjuangan di kampus, yang menjadi tempat berbagi suka dan duka, serta saling menyemangati selama menjalani masa perkuliahan hingga tahap akhir ini.
5. Almamater tercinta, Universitas Nusantara PGRI Kediri, yang telah menjadi tempat saya tumbuh, belajar, dan mengembangkan diri selama masa perkuliahan.
6. Semua pihak yang tidak dapat saya sebutkan satu per satu, yang telah memberikan kontribusi dalam berbagai bentuk dukungan selama proses penyusunan skripsi ini.

Semoga dedikasi ini dapat menjadi penghormatan atas semua dukungan, doa, dan kebersamaan yang telah diberikan.

RINGKASAN

Yosan Okta Odhianto Implementasi *Machine learning* Untuk *Topic Modeling* Pada Aplikasi Dongeng Nusantara Panji Kediri, Skripsi, Program Studi Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri, 2025.

Kata Kunci : *Latent Dirichlet Allocation* (LDA), *Agglomerative Hierarchical Clustering* (AHC), *Topic Modeling*, Pemrosesan Bahasa Alami, Dongeng Nusantara.

Pengelompokan dongeng secara manual berdasarkan tema memerlukan waktu dan tenaga yang besar, serta menyulitkan pengguna dalam mencari cerita sesuai minat, sehingga dibutuhkan sistem otomatis yang mampu mengelompokkan dongeng berdasarkan kesamaan tema. Penelitian ini bertujuan untuk mengembangkan sistem pengelompokan dongeng Nusantara secara otomatis menggunakan metode *Latent Dirichlet Allocation* (LDA) untuk ekstraksi topik dan *Agglomerative Hierarchical Clustering* (AHC) untuk klasterisasi berdasarkan distribusi topik. Hasilnya menunjukkan bahwa model LDA dengan 6 topik memberikan nilai *Coherence Score* terbaik sebesar 0,4644, sedangkan klasterisasi dengan AHC optimal pada 9 klaster dengan nilai *Silhouette Score* sebesar 0,8787 dan *Davies-Bouldin Index* sebesar 0,4942. Sistem ini diimplementasikan dalam aplikasi web yang memudahkan pencarian cerita berdasarkan tema dan berkontribusi pada pelestarian budaya melalui teknologi NLP.

PRAKATA

Puji Syukur dipanjatkan kepada Tuhan Yang Maha Esa, karena atas ridha dan karunianya peneliti dapat menyelesaikan penyusunan laporan penelitian ini. Penulisan ini juga tak lepas dari dukungan pihak yang selalu membantu dalam penulisan penelitian ini. Oleh karenanya peneliti ingin mengucapkan terima kasih kepada:

1. Dr. Zainal Afandi, M.Pd. selaku Rektor Universitas Nusantara PGRI Kediri.
2. Dr. Sulistiono, M.Si selaku Dekan Fakultas Teknik dan Ilmu Komputer Universitas Nusantara PGRI Kediri.
3. Risa Helilintar, M.Kom selaku Ketua Program Studi Teknik Informatika Universitas Nusantara PGRI Kediri.
4. Daniel Swanjaya, M.Kom dan Julian Sahertian, M.Kom selaku Dosen Pembimbing Skripsi yang telah dan mengarahkan saya selama mengerjakan skripsi.
5. Kedua Orang Tua saya dan Keluarga atas doa dan dukungannya.
6. Ucapan terimakasih juga disampaikan kepada pihak-pihak lain yang tidak dapat disebutkan satu persatu, yang telah banyak membantu menyelesaikan penulisan penelitian ini.

Disadari penelitian ini masih banyak kekurangan, maka diharapkan kritik dan saran dari berbagai pihak sangat diharapkan. Semoga penelitian ini bermanfaat bagi semua pihak.

Kediri, 24 Juni 2025



Yosan Okta Odhianto

NPM : 2113020137

Daftar Isi

HALAMAN PERNYATAAN	iv
RINGKASAN	vi
PRAKATA.....	vii
Daftar Isi.....	viii
Daftar Tabel.....	x
Daftar Gambar.....	xi
Daftar Lampiran	xii
BAB I Pendahuluan	13
A. Latar Belakang	13
B. Identifikasi Masalah	14
C. Rumusan Masalah	15
D. Batasan Masalah.....	15
E. Tujuan Penelitian.....	16
F. Manfaat Penelitian	16
BAB II Landasan Teori	17
A. Landasan Teori	17
1. Dongeng	17
2. Tema Dongeng	17
3. Natural Language Processing (NLP)	18
4. Topic Modeling.....	19
5. Latent Dirichlet Allocation (LDA).....	19
6. Agglomerative Hierarchical Clustering (AHC)	20
B. Kajian Pustaka.....	21
C. Kerangka Berpikir.....	24
BAB III Metode Penelitian	26
A. Desain Penelitian.....	26
1. Jenis Penelitian.....	26
2. Variabel Penelitian	26
B. Instrumen Penelitian.....	30

1. Perangkat Keras (Hardware)	30
2. Perangkat Lunak (Software)	30
3. Dataset.....	31
4. Analisis Hasil	31
C. Tempat dan Jadwal Penelitian	32
1. Tempat Penelitian.....	32
2. Jadwal Penelitian.....	32
D. Objek dan Subjek Penelitian	33
1. Analisis Kebutuhan Sistem	33
2. Objek Penelitian	34
3. Subjek Penelitian.....	34
E. Prosedur Penelitian.....	34
F. Teknik Analisis Data	35
1. Desain Sistem.....	35
2. Simulasi Pemecahan Masalah	40
BAB IV Hasil dan Pembahasan	46
A. Hasil Penelitian	46
1. Implementasi desain sistem.....	46
2. Pengujian fungsional	54
3. Pengujian non-fungsional.....	58
B. Pembahasan.....	60
1. Analisis model LDA dan visualisasi	60
2. Analisis model AHC dan visualisasi.....	65
3. Analisis hasil kluster	66
4. Keunggulan sistem	76
5. Kekurangan sistem dan tantangan.....	77
BAB V Penutup.....	79
A. Simpulan	79
B. Saran.....	79
Daftar Pustaka	80
Lampiran	82

Daftar Tabel

Tabel 3. 1 Jadwal penelitian.....	32
Tabel 3. 2 Preprocessing text	40
Tabel 3. 3 Corpus	43
Tabel 3. 4 Topik	43
Tabel 3. 5 Topic Distribution	44
Tabel 4. 1 Pengujian fungsional halaman dashboard.....	55
Tabel 4. 2 Pengujian fungsional menu sidebar	55
Tabel 4. 3 Pengujian fungsional halaman clustering	56
Tabel 4. 4 Pengujian fungsional halaman daftar data dongeng	56
Tabel 4. 5 Pengujian fungsional menu tambah data dongeng.....	57
Tabel 4. 6 Pengujian fungsional halaman hasil klaster	58
Tabel 4. 7 Tabel Coherence score	58
Tabel 4. 8 Tabel Silhouette Score dan DBI	60
Tabel 4. 9 klaster pertama	67
Tabel 4. 10 Klaster kedua.....	68
Tabel 4. 11 Klaster ketiga	69
Tabel 4. 12 Klaster keempat.....	70
Tabel 4. 13 Klaster kelima	71
Tabel 4. 14 klaster keenam.....	72
Tabel 4. 15 Klaster ketujuh	73
Tabel 4. 16 Klaster kedelapan	74
Tabel 4. 17 Klaster kesembilan.....	75

Daftar Gambar

Gambar 2. 1 Alur kerja LDA (Nawang Sari & Dwi Purnomo, 2022)	19
Gambar 2. 2 Alur kerja AHC (Eric U. & Michael O., 2024).....	20
Gambar 2. 3 Kerangka berpikir.....	24
Gambar 3. 3 Prosedur penelitian.....	34
Gambar 3. 4 Flowchart sistem	35
Gambar 3. 5 Diagram konteks	36
Gambar 3. 6 DFD level 1	37
Gambar 3. 7 DFD Level 2 proses 1.1	38
Gambar 3. 8 ERD Sistem Klasifikasi.....	39
Gambar 3. 9 Alur Pengolahan Data	42
Gambar 4. 1 Halaman Dashboard	46
Gambar 4. 2 Sidebar.....	47
Gambar 4. 3 Halaman Clustering.....	47
Gambar 4. 4 Halaman List Dongeng.....	48
Gambar 4. 5 Input Dongeng.....	49
Gambar 4. 6 Halaman Hasil Cluster	49
Gambar 4. 7 Tampilan Kata Per-klaster.....	50
Gambar 4. 8 Topik pertama	62
Gambar 4. 9 Topik kedua.....	62
Gambar 4. 10 Topik ketiga.....	63
Gambar 4. 11 Topik keempat.....	63
Gambar 4. 12 Topik kelima	64
Gambar 4. 13 Topik keenam.....	64
Gambar 4. 14 Dendrogram.....	66

Daftar Lampiran

Lampiran 1. Output DBI dan Shioulette Score	82
Lampiran 2. Lembar Bimbingan Skripsi 1	87
Lampiran 3. Lembar Bimbingan Skripsi 2.....	88

BAB I

Pendahuluan

A. Latar Belakang

Aplikasi Dongeng Panji Nusantara adalah sebuah *platform* digital berbasis *website* yang menyediakan layanan baca dongeng secara online, dirancang untuk memperkenalkan dan melestarikan budaya nusantara melalui cerita – cerita dongeng yang kaya akan nilai moral, tradisi, dan kearifan lokal. Aplikasi ini ditujukan untuk semua kalangan, dari anak-anak hingga dewasa, yang ingin mengeksplorasi kisah – kisah menarik dari berbagai daerah di Indonesia. Untuk meningkatkan pengalaman pengguna, implementasi *machine learning* pada *topic modeling* diterapkan guna mengelompokkan tema cerita secara otomatis. Berdasarkan penelitian terdahulu tentang implementasi klasifikasi artikel berita Indonesia menggunakan metode *Latent Dirichlet Allocation* (LDA), Kusumaningrum et al. (2016) berhasil melakukan klasifikasi terhadap dokumen teks berita berbahasa Indonesia dengan akurasi rata – rata sebesar 70% untuk 10 arsip artikel secara keseluruhan sebagai data uji, dan pada arsip ke-5, akurasi mencapai 100%(Sakti Wardhana & Kesumawati, 2023).

Latent Dirichlet Allocation (LDA) adalah salah satu topik modeling yang populer digunakan, *Latent Dirichlet Allocation* (LDA) adalah bentuk *Topic Modeling* yang digunakan untuk mengekstrak fitur dari data teks (Hasan dkk., 2020). Dari hasil ekstrak fitur data teks oleh model *Latent Dirichlet Allocation* (LDA) diperoleh beberapa fitur baru antara lain (1) distribusi topik dalam dokumen : probabilitas dari topik – topik yang ada dalam sebuah dokumen, setiap dokumen diberi bobot untuk setiap topik, yang menunjukkan seberapa relevan dokumen tersebut dengan topik tertentu (2) topik : topik didefinisikan oleh sekumpulan kata – kata yang memiliki bobot probabilitas tinggi dalam sebuah dokumen. selanjutnya akan diimplementasikan *Agglomerative Hierarchical Clustering* (AHC) untuk mengelompokkan dokumen-dokumen berdasarkan kesamaan topik yang telah teridentifikasi.

Untuk mengelompokkan dokumen berdasarkan kemiripan distribusi topik, *Agglomerative Hierarchical Clustering* (AHC) menjadi salah satu metode *clustering* yang dapat digunakan dengan baik. *Agglomerative Hierarchical Clustering* (AHC) adalah jenis pengelompokan hierarkis yang paling populer yang digunakan untuk mengelompokkan objek dalam kelompok berdasarkan kesamaannya. Metode ini menggunakan pendekatan *bottom-up* dan mulai mengelompokkan dengan memperlakukan titik data individu sebagai satu kluster, kemudian digabungkan terus menerus berdasarkan kesamaan hingga membentuk satu kluster besar yang berisi semua objek (Eric U. & Michael O., 2024). AHC efektif dalam mengelompokkan dokumen berdasarkan distribusi topik, sehingga cocok untuk digunakan bersama LDA dalam pengelompokan tema.

Meskipun LDA dan AHC telah banyak digunakan dalam pemrosesan teks, kombinasi kedua metode ini jarang diterapkan untuk pengklasifikasian tema dalam dongeng. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan sistem yang mampu mengidentifikasi dan mengelompokkan tema dalam dongeng secara otomatis menggunakan kombinasi LDA dan AHC.

Oleh karena itu, penelitian ini bertujuan untuk mengembangkan sebuah sistem yang mampu mengidentifikasi dan mengelompokkan tema-tema dalam dongeng secara otomatis dengan menggunakan kombinasi metode LDA dan AHC. Melalui pendekatan ini, diharapkan proses pengklasifikasian tema dapat dilakukan lebih cepat dan akurat, sehingga memberikan kontribusi signifikan dalam pengembangan teknologi pengolahan teks, khususnya dalam ranah sastra dan naratif.

B. Identifikasi Masalah

Masalah utama yang dihadapi pengelola aplikasi adalah kurangnya efisiensi dalam mengelompokkan dongeng berdasarkan tema, yang berdampak pada kesulitan pengguna dalam menemukan cerita yang sesuai dengan minat mereka.

C. Rumusan Masalah

Bagaimana algoritma *Latent Dirichlet Allocation*, yang dikombinasikan dengan *Agglomerative Hierarchical Clustering* untuk menghasilkan pengelompokan yang lebih akurat dan relevan?

D. Batasan Masalah

Agar penulisan proposal ini lebih fokus tentang penelitian yang akan dibahas peneliti telah menetapkan batasan masalah sebagai berikut:

1. Penelitian hanya berfokus pada Aplikasi Dongeng Nusantara Panji Kediri.
2. Praktisi pada penelitian ini adalah Encil Puspitoningrum, M.Pd., sekaligus penanggungjawab Aplikasi Dongeng Nusantara Panji Kediri.
3. Algoritma *Latent Dirichlet Allocation* (LDA) digunakan untuk *Topic Modeling* sehingga didapat N Topik.
4. Algoritma *Agglomerative Hierarchical Clustering* digunakan untuk menyederhanakan Topik yang dihasilkan oleh LDA, sehingga didapat Tema dan Sub Temanya.
5. Proses pelabelan Tema dilakukan oleh pengelola Aplikasi Dongeng Nusantara Panji Kediri secara manual, dengan mempertimbangkan komposisi kata dalam tiap topik.
6. Metrik Evaluasi yang digunakan adalah *Coherence* dan *Silhouette Score*
7. Aplikasi berbasis Web
8. Bahasa Pemrograman yang digunakan adalah *python*
9. *Database* Server yang digunakan adalah *mysql*

E. Tujuan Penelitian

Adapun tujuan Penelitian ini dilakukan adalah :

1. Mengembangkan model *Topic Modeling* menggunakan algoritma LDA untuk mengidentifikasi tema-tema laten dalam kumpulan dongeng Panji Kediri.
2. Menerapkan algoritma *Agglomerative Hierarchical Clustering* untuk menyederhanakan hierarki topik yang dihasilkan oleh LDA.

Mengevaluasi kinerja model dalam mengklasifikasikan dongeng berdasarkan tema menggunakan metrik evaluasi *coherence*.

F. Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah :

1. Membantu pengelola aplikasi dalam mengklasifikasikan dongeng secara otomatis, sehingga menghemat waktu dan tenaga.
2. Memungkinkan pengelola aplikasi untuk mengorganisir koleksi dongeng dengan lebih baik.
3. Pengguna dapat dengan mudah menemukan dongeng yang sesuai dengan minat mereka.

Daftar Pustaka

- Amien, M. (2023). *Sejarah dan Perkembangan Teknik Natural Language Processing (NLP) Bahasa Indonesia: Tinjauan tentang sejarah, perkembangan teknologi, dan aplikasi NLP dalam bahasa Indonesia*. <http://arxiv.org/abs/2304.02746>
- Eric U., O., & Michael O., O. (2024). Overview of Agglomerative Hierarchical Clustering Methods. *British Journal of Computer, Networking and Information Technology*, 7(2), 14–23. <https://doi.org/10.52589/bjcnit-cv9poogw>
- Furqan, M., Sriani, S., & Shidqi, M. N. (2023). Chatbot Telegram Menggunakan Natural Language Processing. *Walisongo Journal of Information Technology*, 5(1), 15–26. <https://doi.org/10.21580/wjit.2023.5.1.14793>
- Hasan, M., Rahman, A., Razaul Karim, M., Saikat Islam Khan, M., & Jahidul Islam, M. (2020). *Normalized Approach to Find Optimal Number of Topics in Latent Dirichlet Allocation (LDA)*. https://link.springer.com/chapter/10.1007/978-981-33-4673-4_27
- Iyan Yulianti, D., Iman Hermanto, T., & Defriani, M. (2023). RESOLUSI: Rekayasa Teknik Informatika dan Informasi Analisis Clustering Donor Darah dengan Metode Agglomerative Hierarchical Clustering. *RESOLUSI: Rekayasa Teknik Informatika dan Informasi*, 3(6), 308. <https://djournals.com/resolusi>
- Maulidiya, D. (2023). Topic Modelling using Latent Dirichlet Allocation (LDA) to Investigate the Latent Topics of Mathematical Creative Thinking Research in Indonesia. *Journal of Intelligent Computing & Health Informatics*, 3(2), 35. <https://doi.org/10.26714/jichi.v3i2.11428>
- Nawang Sari, W. A., & Dwi Purnomo, H. (2022). TOPIC MODELING USING THE LATENT DIRICHLET ALLOCATION METHOD ON WIKIPEDIA

- PANDEMIC COVID-19 DATA IN INDONESIA. *Jurnal Teknik Informatika (Jutif)*, 3(5), 1223–1230. <https://doi.org/10.20884/1.jutif.2022.3.5.321>
- Nurani, R. Z., Nugraha, F., & Arga, H. S. P. (2021). Penggunaan Metode Pembelajaran Guided Discovery untuk Meningkatkan Kemampuan Mengidentifikasi Unsur-Unsur Intrinsik Dongeng di Sekolah Dasar. *Jurnal Basicedu*, 5(2), 627–635. <https://doi.org/10.31004/basicedu.v5i2.761>
- Putu Candra Prastya Dewi, N., Bagus Putrayasa, I., & Sudiana, I. N. (2021). *MEMBENTUK KARAKTER ANAK MELALUI HABITUASI DONGENG PADA PEMBELAJARAN DI SEKOLAH DASAR*.
- Sakti Wardhana, Y., & Kesumawati, A. (2023). Implementasi Klasifikasi Naïve Bayes dan Pemodelan Topik dengan Latent Dirichlet Allocation untuk Data Ulasan Video Game Lokal Pada Platform Steam. *Emerging Statistics and Data Science Journal*, 1(3).
- Septianingsih, A., Pertama, S. A., Kependudukan, D., Sipil, P., & Tangerang, K. (2022). *PEMETAAN KABUPATEN KOTA DI PROVINSI JAWA TIMUR BERDASARKAN TINGKAT KASUS PENYAKIT MENGGUNAKAN PENDEKATAN AGGLOMERATIF HIERARCHICAL CLUSTERING*. 3(2). <https://doi.org/10.46306/lb.v3i2>