

Pemodelan Topik dalam Judul Berita Pemilu 2019 Indonesia Menggunakan Metode Latent Dirichlet Allocation (LDA)

Oleh:

Muhammad Lazuardi Imani,

Arif Senja Fitrani

Progam Studi Informatika

Universitas Muhammadiyah Sidoarjo

Juli 2025



www.umsida.ac.id



[umsida1912](#)



[umsida1912](#)



[universitas
muhammadiyah
sidoarjo](#)



[umsida1912](#)

Pendahuluan

Pemilu 2019 di Indonesia merupakan peristiwa politik besar yang melibatkan berbagai pemilihan tingkat nasional dan daerah. Penelitian ini bertujuan menganalisis judul berita online tentang Pemilu 2019 menggunakan metode Latent Dirichlet Allocation (LDA) untuk mengidentifikasi topik-topik utama dalam pemberitaan. LDA digunakan karena kemampuannya mengelompokkan kata-kata yang sering muncul bersama ke dalam topik tertentu. Penelitian ini juga memanfaatkan coherence score untuk memilih jumlah topik optimal dan visualisasi interaktif melalui pyLDAvis. Beberapa penelitian sebelumnya juga menjadi acuan, terutama dalam konteks analisis media dan persepsi publik.

Pertanyaan Penelitian (Rumusan Masalah)

Berdasarkan fenomena yang diuraikan pada pendahuluan , maka rumusan masalah penelitian ini adalah sebagai berikut.

1. Bagaimana memodelkan topik dalam judul berita tentang pemilu 2019 dengan metode Latent Dirichlet Allocation (LDA)?
2. Bagaimana hasil pengelompokkan topik-topik dalam judul berita pemilu 2019 berdasarkan model LDA yang dibangun?

Metode

Penelitian ini menggunakan pendekatan analisis topik dengan metode Latent Dirichlet Allocation (LDA) untuk mengidentifikasi tema-tema utama dari judul berita Pemilu 2019. Berikut tahapan utamanya:

1. Pengambilan Dataset

- Data dikumpulkan melalui proses **web scraping** dari portal berita online.
- Dataset kemudian disimpan dalam format Excel untuk memudahkan pengolahan.



A	B	C	D	E	F
1	Title				
2	Kampanye Pilkada Bangkalan Mulai Tidak Sehat				
3	Benarkah Polsek Tanjung Bumi Tak Gubris Laporan Tim Farid Alfauzi?				
4	Bila Pendaftaran Dibuka, Farid Alfauzi Sudah Bisa Mendaftar Pilkada Bangkalan				
5	Ratusan Baleho Bakal Calon Bupati Bangkalan Dirusak				
6	Begitu Cara Nyaleg Lewat Hanura Bangkalan				
7	Serikat Buruh Sepakat Deklarasi Damai Pemilu 2019				
8	Antisipasi Rawan Pelanggaran, Bawaslu Minta Masyarakat Aktif				
9	survey lembaga adiwangsa Jokowi unggut Prabowo di Malang				
10	Dana Reses Capai Rp 12,1 Miliar, Dewan Anggap-Wajar				
11	Poster Paslon Jokowi Maruf Dipasang di Tengah Jalan Berlubang di Kota Malang				
12	ASN Pose Paslon Capres, Bawaslu Kota Malang: Dia Ngakunya Terbawa Suasana				
13	Dispendukcapil Kota Batu Bakar Ratusan E-KTP Rusak				
14	Polisi Wartawan Kompak Jaga Kota Batu Tetap Sejuk di Tahun Politik				
15	raih penghargaan kualifikasi b di ajang ppid award 2018 kota malang masih kalah dengan kota madiun				
16	polisi grebek penjual miras berkedok toko jamu di bandulan				
17	puluhan ribu warga kota malang belum perekaman ktp elektronik				
18	terkait suban kota malang pertanyaan keputusan KPU RI				
19	seluruh elemen sepakat jaga kedamaian kota malang di pemilu 2019				
20	pkb optimis rebut 3 kursi DPR RI di dapil malang raya				
21	ribuan warga malang deklarasi mendukung Jokowi				
22	ribuan warga malang ikuti senam sehat deklarasi dukungan Jokowi Maruf Amin				
23	romahurmuzily ajak masyarakat cerdas dalam memilih pemimpin				
24	polisi tanggap aduan mci terkait postingan herman jenggot harland				
25	ratusan santri di malang deklarasi dukung Jokowi Maruf Amin				
26	seribu siswa di kota malang dapat ktp				
27	ridwan hisjam prof dr hc kh maruf amin patut jadi teladan				
28	presiden Jokowi sapa puluhan ribu muslimat nu				

Metode

2. Pra-pemrosesan Teks

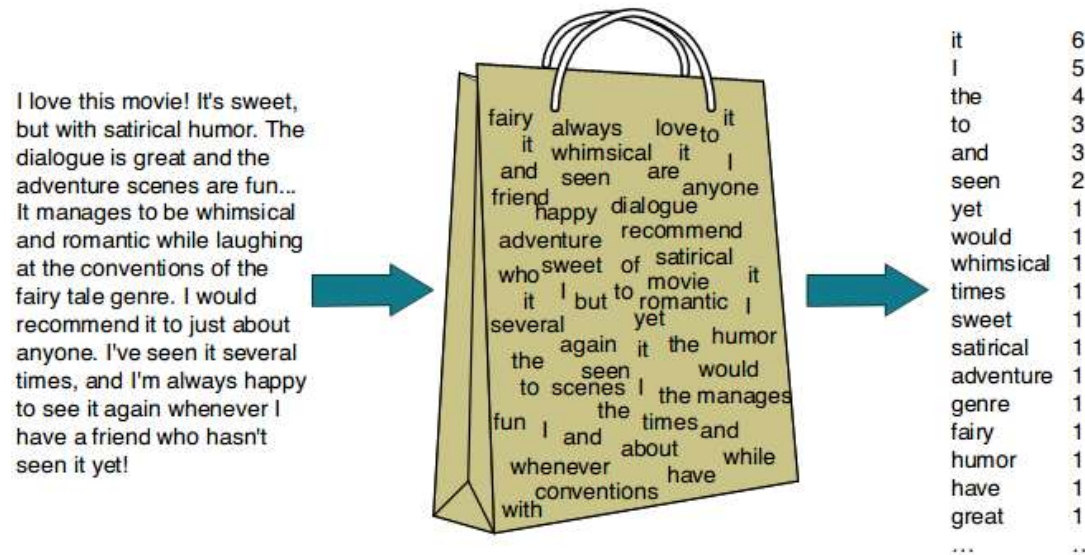
Beberapa tahap pengolahan teks dilakukan agar data siap dianalisis:

- **Lowercase folding:** Mengubah seluruh teks menjadi huruf kecil.
- **Tokenisasi:** Memecah kalimat menjadi kata-kata (token).
- **Penghapusan karakter khusus:** Menghapus emoji, tanda baca, tagar, dan URL.
- **Stopwords removal:** Menghapus kata-kata umum yang tidak bermakna signifikan.
- **Normalisasi teks:** Mengurangi duplikasi dan bentuk kata yang tidak baku.
- **Stemming:** Mengubah kata ke bentuk dasar (misalnya “mendaftar” jadi “daftar”) menggunakan library **Sastrawi**.

Metode

3. Pembuatan Bag of Words (BoW)

- Teks yang telah diproses dikonversi menjadi **representasi numerik** melalui model **BoW** untuk menghitung frekuensi kata dalam dokumen.



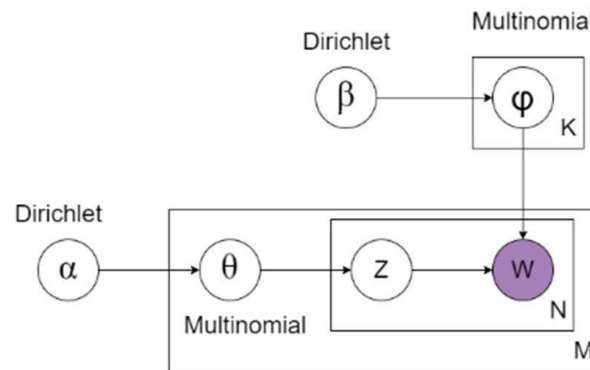
Metode

4. Penerapan Model LDA

Model **LDA** digunakan untuk memetakan kumpulan kata menjadi beberapa topik tersembunyi dalam dataset.

Parameter LDA yang digunakan mencakup:

- Jumlah dokumen (M)
- Jumlah topik (K)
- Iterasi (i)
- Koefisien α (alpha) dan β (beta)
- Nilai **coherence score** dihitung untuk menentukan jumlah topik terbaik (optimum).



Metode

5. Evaluasi dan Visualisasi

- Evaluasi dilakukan menggunakan **model coherence-gensim** untuk mengukur keakuratan topik.
- Hasil visualisasi topik divisualisasikan menggunakan **pyLDAvis** agar hubungan antar topik dan kata kunci lebih mudah dianalisis.

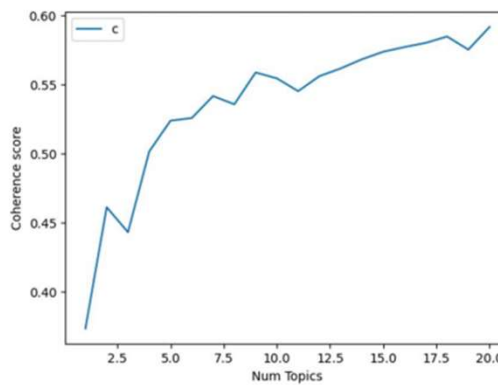
Hasil

1. Pembuatan Bag of Word (BOW)

Kata	Frequency Kemunculan
milu	1
warga	1
kpu	1
ajak	1
nyoblos	1
run	1

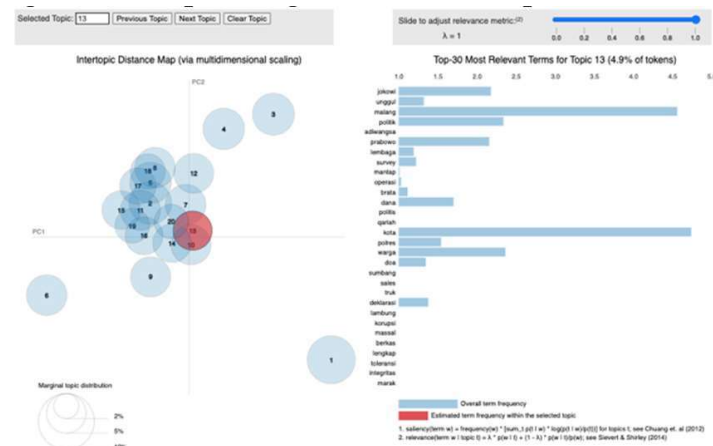
2. Latent Dirichlet Allocation (LDA)

- Coherence Score



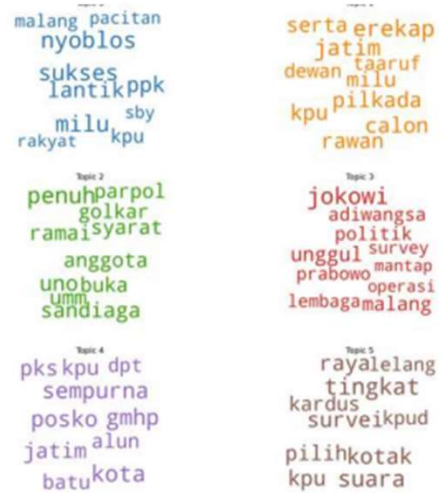
Hasil

3. Visualisasi pyLDAvis



Hasil

4. Wordcloud



Kesimpulan

Setelah menganalisis data dan mengolahnya melalui berbagai tahap pemrosesan teks dan pemodelan LDA, penelitian ini telah berhasil mengidentifikasi topik-topik utama yang muncul dalam pemberitaan terkait Pemilu 2019 di Indonesia. Topik-topik yang teridentifikasi mencakup dinamika pemilihan umum, fokus pada figur politik spesifik, serta aspek-aspek logistik dan administratif yang berkaitan dengan pemilihan. Dengan menggunakan model LDA dan alat visualisasi seperti pyLDAvis, penelitian ini memberikan wawasan mendalam tentang cara topik-topik tersebut ditangani dan dibahas dalam media. Hasil ini menawarkan pandangan yang berharga tentang bagaimana berbagai aspek dari pemilu dipersepsikan dan disajikan kepada publik, yang tidak hanya penting untuk memahami konteks pemilu tersebut tetapi juga untuk merancang strategi komunikasi politik yang lebih efektif di masa depan.

Referensi

- [1] D. Z. T. Kannitha, M. Mustafid, and P. Kartikasari, "Pemodelan Topik Pada Keluhan Pelanggan Menggunakan Algoritma Latent Dirichlet Allocation Dalam Media Sosial Twitter," *J. Gaussian*, vol. 11, no. 2, pp. 266–277, 2022, doi: 10.14710/j.gauss.v11i2.35474.
- [2] M. Y. Febrianta, S. Widiyanesti, and S. R. Ramadhan, "Analisis Ulasan Indie Video Game Lokal pada Steam Menggunakan Analisis Sentimen dan Pemodelan Topik Berbasis Latent Dirichlet Allocation," *J. Animaf. Games Stud.*, vol. 7, no. 2, pp. 117–144, 2021, doi: 10.24821/jags.v7i2.5162.
- [3] A. Syaifuddin, R. A. Harianto, and J. Santoso, "Analisis Trending Topik untuk Percakapan Media Sosial dengan Menggunakan Topic Modelling Berbasis Algoritme LDA," *J. Intell. Syst. Comput.*, vol. 2, no. 1, pp. 12–19, 2021, doi: 10.52985/insyst.v2i1.150.
- [4] J. W. Fernanda, "Pemodelan Persepsi Pembelajaran Online Menggunakan Latent Dirichlet Allocation," *J. Stat. Univ. Muhammadiyah Semarang*, vol. 9, no. 2, p. 79, 2021, doi: 10.26714/jsunimus.9.2.2021.79-85.
- [5] M. Istianda and A. Zastrawati, "Evaluasi Penyelenggaraan Pemilu Serentak 2019 Kota Makassar," *Sebatik*, vol. 25, no. 1, pp. 92–101, 2021, doi: 10.46984/sebatik.v25i1.1203.
- [6] L. H. Husin et al., "Malpraktik Pemilu dan Korupsi: Analisis terhadap Proses Penghitungan dan Rekapitulasi pada Pemilu Indonesia 2019," *Integritas J. Antikorupsi*, vol. 7, no. 1, pp. 57–78, 2021, doi: 10.32697/integritas.v7i1.720.
- [7] U. Malihatin S. Y. Findawati, and U. Indahyanti, "Topic Modeling in Covid-19 Vaccination Refusal Cases Using Latent Dirichlet Allocation and Latent Semantic Analysis," *J. Tek. Inform.*, vol. 4, no. 5, pp. 1063–1074, 2023, doi: 10.52436/1.jutif.2023.4.5.951.
- [8] Y. Matira and I. Setiawan, "Pemodelan Topik pada Judul Berita Online Detikcom Menggunakan Latent Dirichlet Allocation," *Estimasi J. Stat. Its Appl.*, vol. 4, no. 1, pp. 2721–379, 2023, doi: 10.20956/ejsa.vi.24843.
- [9] S. Mifrah, "Topic Modeling Coherence: A Comparative Study between LDA and NMF Models using COVID'19 Corpus," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 4, pp. 5756–5761, 2020, doi: 10.30534/ijatcse/2020/231942020.
- [10] N. A. K. Febriyani and R. B. Hadiprakoso, "Rancang Bangun Aplikasi Naskah Dinas Elektronik Berbasis Web Menggunakan WDLC," *Ultim. InfoSys J. Ilmu Sist. Inf.*, vol. 12, no. 1, pp. 43–51, 2021, doi: 10.31937/si.v12i1.1747.
- [11] D. Qiu, H. Jiang, and S. Chen, "Fuzzy information retrieval based on continuous bag-of-words model," *Symmetry (Basel)*, vol. 12, no. 2, 2020, doi: 10.3390/sym12020225.
- [12] S. Sunanto, R. Firdaus, and Makmur Setiawan Siregar, "Implementasi Logika Fuzzy Mamdani Pada Kendali Suhu dan Kelembaban Ruang Server," *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 2, no. 2, pp. 128–136, 2021, doi: 10.37859/coscitech.v2i2.3362.
- [13] J. Akbar, T. A. M., Y. Tolla, A. E. Ahmad, A. Yaqin, and E. Utami, "Pemodelan Topik Menggunakan Latent Dirichlet Allocation pada Ulasan Aplikasi PeduliLindungi," *InComTech J. Telekomun. dan Komput.*, vol. 13, no. 1, p. 40, 2023, doi: 10.22441/incomtech.v13i1.15572.
- [14] R. Solihah, "Peluang dan tantangan pemilu serentak 2019 dalam perspektif politik," *J. Ilm. Ilmu Pemerintah.*, vol. 3, no. 1, p. 73, 2018, doi: 10.14710/jiip.v3i1.3234.
- [15] D. L. C. Pardede and M. A. I. Waskita, "Analisis Pemodelan Topik Untuk Ulasan Tentang Peduli Lindungi," *J. Ilm. Inform. Komput.*, vol. 28, no. 1, pp. 17–26, 2023, doi: 10.35760/ik.2023.v28i1.7925.

