

PERBANDINGAN KINERJA LDA DAN KEYNMF DALAM PEMODELAN TOPIK DISKURSUS PUBLIK TERKAIT ISU LPDP PADA MEDIA SOSIAL

ADINDA PRATIWI



**PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PERNYATAAN MENGENAI SKRIPSI, SUMBER INFORMASI, PENGGUNAAN AI, DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Perbandingan Kinerja LDA dan KeyNMF dalam Pemodelan Topik Diskursus Publik Terkait Isu LPDP pada Media Sosial” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dalam penyusunan karya ini, penulis menggunakan bantuan kecerdasan buatan berupa ChatGPT, Claude, dan Gemini untuk identifikasi dan perbaikan kesalahan kode program, penyusunan kerangka penulisan, serta penyuntingan bahasa. Setelah menggunakan alat/layanan tersebut, penulis meninjau dan menyunting konten sesuai kebutuhan serta bertanggung jawab penuh atas isi karya tugas akhir ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Agustus 2026

Adinda Pratiwi
G1401221041



ABSTRAK

ADINDA PRATIWI. Perbandingan Kinerja LDA dan KeyNMF dalam Pemodelan Topik Diskursus Publik Terkait Isu LPDP pada Media Sosial. Dibimbing oleh CICI SUHAENI dan AKBAR RIZKI.

Sebagai program pemerintah, beasiswa Lembaga Pengelola Dana Pendidikan (LPDP) menjadi salah satu program yang banyak diperbincangkan di media sosial, terutama setelah munculnya kontroversi alumni LPDP pada Februari 2026. Diskursus tersebut menghasilkan data teks dalam jumlah besar yang memerlukan metode *topic modeling* untuk mengidentifikasi topik-topik dominan secara efektif. Penelitian ini bertujuan membandingkan kinerja metode LDA dan KeyNMF dalam pemodelan topik pada data media sosial berbahasa Indonesia terkait LPDP serta mengidentifikasi dan menginterpretasikan topik-topik dominan menggunakan metode terbaik. Penelitian menggunakan 1709 *tweet* dari platform X dan 3800 komentar YouTube yang dikumpulkan melalui *scraping* pada periode 15 Februari hingga 28 Februari 2026. Tahapan penelitian meliputi prapemrosesan teks, pemodelan topik, serta evaluasi berdasarkan *topic coherence*, *topic diversity*, *document coverage*, dan interpretabilitas topik. Hasil penelitian menunjukkan bahwa LDA memberikan performa terbaik pada data platform X dengan nilai *topic diversity* sebesar 1 dan *document coverage* sebesar 0.809, sedangkan KeyNMF memberikan performa terbaik pada data YouTube dengan nilai *topic coherence* sebesar 0.506 dan interpretabilitas topik yang lebih baik. Jumlah topik optimal pada kedua data sama-sama berjumlah 7 topik yang secara garis besar berfokus pada 5 isu utama terkait kewarganegaraan penerima, pengelolaan dana publik, seleksi dan sasaran penerima, tata kelola dan kebijakan, serta komitmen alumni dan manfaat program.

Kata kunci: diskursus, KeyNMF, LDA, LPDP, pemodelan topik



ABSTRACT

ADINDA PRATIWI. Performance Comparison of LDA and KeyNMF in Topic Modeling of Public Discourse on LPDP Issues in Social Media. Supervised by CICI SUHAENI and AKBAR RIZKI.

As a government program, the Education Fund Management Agency (LPDP) scholarship has become one of the most talked-about topics on social media, particularly following the controversy involving LPDP alumni in February 2026. This discourse has generate large volume of text data that requires topic modeling methods to effectively identify dominant topics. This study aims to compare the performance of LDA and KeyNMF for topic modeling on Indonesian-language social media data related to LPDP and to identify and interpret the dominant topics using the best method. The dataset consists of 1709 tweets from X and 3800 YouTube comments collected via scraping between February 15 and February 28, 2026. The research stages included text preprocessing, topic modeling, and evaluation based on topic coherence, topic diversity, document coverage, and topic interpretability. The results showed that LDA performed best on the X platform data, with a topic diversity score of 1 and a document coverage score of 0.809, while KeyNMF performed best on the YouTube data, with a topic coherence score of 0.506 and better topic interpretability. The optimal number of topics for both datasets is 7, broadly focusing on 5 issues related to recipients' citizenship, public fund management, recipient selection and targeting, governance and policy, and alumni engagement and program benefits.

Keywords: discourse, KeyNMF, LDA, LPDP, topic modeling.



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2026¹
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

PERBANDINGAN KINERJA LDA DAN KEYNMF DALAM PEMODELAN TOPIK DISKURSUS PUBLIK TERKAIT ISU LPDP PADA MEDIA SOSIAL

ADINDA PRATIWI

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Tim Penguji pada Ujian Skripsi:

1. Aulia Rizki Firdawanti, S.Stat., M.Si.



Judul Skripsi : Perbandingan Kinerja LDA dan KeyNMF dalam Pemodelan
Topik Diskursus Publik Terkait Isu LPDP pada Media Sosial

Nama : Adinda Pratiwi

NIM : G1401221041

Disetujui oleh

Pembimbing 1:

Cici Suhaeni, S.Si., M.Si., Ph.D.

Pembimbing 2:

Akbar Rizki, S.Stat., M.Si.

Diketahui oleh

Ketua Program Studi:

Dr. Bagus Sartono, S.Si., M.Si.

NIP 197804112005011002

Tanggal Ujian:
27 Juli 2026

Tanggal Lulus:



PRAKATA

Puji dan syukur penulis panjatkan kepada Allah SWT atas segala rahmat dan karunia-Nya, sehingga karya ilmiah ini dapat diselesaikan dengan baik. Penelitian ini disusun dengan judul “Perbandingan Kinerja LDA dan KeyNMF dalam Pemodelan Topik Diskursus Publik Terkait Isu LPDP pada Media Sosial”. Penulisan ini tidak lepas dari dukungan berbagai pihak yang telah membantu penulis sejak masa studi hingga selesainya karya ilmiah ini. Oleh karena itu, penulis menyampaikan terima kasih yang tulus kepada:

1. Ibu Cici Suhaeni, S.Si., M.Si., Ph.D. dan Ibu Akbar Rizki, S.Stat., M.Si. selaku komisi pembimbing yang telah meluangkan waktunya untuk memberikan bimbingan, arahan, dan ilmu selama penyusunan karya ilmiah ini.
2. Ibu Eka Putri Nur Utami, S.Stat., M.Si. selaku moderator kolokium, Bapak Dr. Agus Mohamad Soleh, S.Si., M.T. selaku moderator seminar, dan Ibu Aulia Rizki Firdawanti, S.Stat., M.Si. selaku dosen penguji yang telah memberikan kritik serta saran membangun demi kesempurnaan penelitian ini.
3. Seluruh Bapak dan Ibu Dosen serta tenaga kependidikan Program Studi Statistika dan Sains Data IPB University yang telah membekali penulis dengan ilmu pengetahuan dan bantuan administratif selama masa perkuliahan.
4. Kedua orang tua penulis dan seluruh keluarga besar yang senantiasa memberikan doa, dukungan moril, serta kasih sayang yang menjadi kekuatan terbesar penulis.
5. Teman-teman seperjuangan Statistika Angkatan 59, khususnya Deswita, Cindy, Yulianti, Arbaynah, Ria, Biki, Azanti, Virgie, Haikal, dan Laras, yang selalu menemani, memberikan semangat, serta menjadi rekan diskusi yang luar biasa selama masa studi.
6. Elmanda, Nifa, Rei, Putri, Dinda, Nita, Yiska, dan Fuji sebagai sahabat semasa kuliah yang telah menemani dan mendukung penulis hingga saat ini.
7. Dinda, Inayah, Angel, serta Alfija sebagai sahabat yang telah memberikan dukungan dan semangat bagi penulis sejak masa sekolah hingga saat ini.
8. Rekan-rekan dari organisasi, terutama GSB dan Panti Goceng yang telah memberikan pengalaman berharga dan semangat dalam melewati masa-masa kuliah yang menantang.
9. Semua pihak yang tidak dapat penulis sebutkan satu per satu, yang telah membantu baik secara langsung maupun tidak langsung dalam penyelesaian skripsi ini.

Penulis menyadari bahwa karya ilmiah ini masih jauh dari sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun dari pembaca. Semoga penelitian ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan, khususnya di bidang statistika dan sains data.

Bogor, Agustus 2026

Adinda Pratiwi

DAFTAR ISI

DAFTAR TABEL	xi
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xi
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	3
II TINJAUAN PUSTAKA	4
2.1 <i>Topic Modeling</i>	4
2.2 <i>Latent Dirichlet Allocation (LDA)</i>	4
2.3 <i>Variational Inference</i>	5
2.4 Tokenisasi N-gram	6
2.5 <i>Bag-of-Words (BoW)</i>	7
2.6 KeyNMF	8
2.7 <i>Sentence Embedding</i> Berbasis <i>Multilingual MPNet</i>	9
2.8 Metrik Evaluasi	10
2.8.1 <i>Topic Coherence</i>	10
2.8.2 <i>Topic Diversity</i>	11
2.8.3 <i>Document Coverage</i>	11
III METODE	12
3.1 Data	12
3.2 Prosedur Analisis Data	12
IV HASIL DAN PEMBAHASAN	16
4.1 Praproses Data	16
4.2 Eksplorasi Data	18
4.2.1 Eksplorasi Data Platform X	18
4.2.2 Eksplorasi Data Platform YouTube	20
4.3 Pemodelan Topik Latent Dirichlet Allocation (LDA)	23
4.3.1 Tokenisasi <i>N-gram</i>	23
4.3.2 <i>Bag-of-Word (BoW)</i>	24
4.3.3 Pemodelan LDA	24
a) LDA pada Data Platform X	25
b) LDA pada Data Platform YouTube	28
4.4 Pemodelan Topik KeyNMF	29
4.4.1 Hasil <i>Sentence Embedding</i>	29
4.4.2 Pemodelan KeyNMF	30
a) KeyNMF pada Data Platform X	31
b) KeyNMF pada Data Platform YouTube	32
4.5 Pemilihan Model Terbaik dan Interpretasi	34
4.5.1 Pemilihan Model Terbaik pada Data Platform X	34
4.5.2 Interpretasi Topik Model Terbaik pada Data Platform X	35
4.5.3 Pemilihan Model Terbaik pada Data Platform YouTube	37
4.5.4 Interpretasi Topik Model Terbaik pada Data Platform YouTube	38



4.5.5	Ringkasan Temuan Topik pada Platform X dan YouTube	40
-------	--	----

V	SIMPULAN DAN SARAN	43
---	--------------------	----

5.1	Simpulan	43
-----	----------	----

5.2	Saran	43
-----	-------	----

	DAFTAR PUSTAKA	44
--	----------------	----

	LAMPIRAN	47
--	----------	----

	RIWAYAT HIDUP	53
--	---------------	----



DAFTAR TABEL

1	Jumlah data sebelum dan sesudah <i>data selection</i>	16
2	Contoh <i>tweet</i> sebelum dan sesudah praproses data	17
3	Hasil tokenisasi <i>unigram-bigram</i>	23
4	Hasil <i>bag-of-words</i>	24
5	Hasil evaluasi model LDA pada data platform X	25
6	Hasil evaluasi model LDA pada data platform YouTube	28
7	Contoh vektor <i>contextual embedding</i>	30
8	Hasil evaluasi model KeyNMF pada data platform X	31
9	Hasil evaluasi model KeyNMF pada data platform YouTube	33
10	Perbandingan performa model terbaik LDA dan KeyNMF pada data platform X	34
11	Interpretasi topik model LDA pada data Platform X	35
12	Perbandingan performa model terbaik LDA dan KeyNMF pada data platform YouTube	37
13	Interpretasi topik model KeyNMF pada Data Platform YouTube	38
14	Hasil pengelompokan topik menjadi isu utama pada platform X dan platform YouTube	41

DAFTAR GAMBAR

1	Representasi model grafis LDA	4
2	<i>Top 10</i> akun dengan <i>tweet</i> terbanyak	18
3	<i>Word cloud</i> kata dominan pada data platform X	19
4	<i>Top 10</i> kata dengan frekuensi tertinggi pada data platform X	20
5	Distribusi komentar berdasarkan <i>channel</i>	20
6	<i>Top 10</i> video dengan jumlah komentar terbanyak	21
7	<i>Word cloud</i> kata dominan pada data platform YouTube	22
8	<i>Top 10</i> kata dengan frekuensi tertinggi pada data platform YouTube	23
9	<i>Intertopic distance map</i> untuk (a) 5 topik, (b) 7 topik, (c) 8 topik, (d) 9 topik, dan (e) 10 topik	26
10	<i>Intertopic distance map</i> untuk (a) 5 topik, (b) 6 topik, (c) 7 topik, dan (d) 9 topik	28

DAFTAR LAMPIRAN

1	Daftar kata kunci dominan model LDA pada data platform X	48
2	Daftar kata kunci dominan model LDA pada data platform YouTube	49
3	Daftar kata kunci dominan model KeyNMF pada data platform X	50
4	Daftar kata kunci dominan model KeyNMF pada data platform YouTube	51



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.