



PEMANFAATAN BERTOPIC UNTUK *TOPIC MODELING* DAN GPT UNTUK ANALISIS SENTIMEN PADA KOMENTAR YOUTUBE TERKAIT ISU MOBIL LISTRIK DI INDONESIA

YULIANTI NURDZANAH



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

- 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 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 DAN SUMBER INFORMASI, PENGGUNAAN AI, DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Pemanfaatan BERTopic untuk *Topic Modeling* dan GPT untuk Analisis Sentimen pada Komentar YouTube terkait Isu Mobil Listrik di Indonesia” 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, Juli 2026

Yulianti Nurdzanah
G140122011



ABSTRAK

YULIANTI NURDZANAH. Pemanfaatan BERTopic untuk *Topic Modeling* dan GPT untuk Analisis Sentimen pada Komentar YouTube terkait Isu Mobil Listrik di Indonesia. Dibimbing oleh CICI SUHAENI dan GERRY ALFA DITO.

Pertumbuhan pasar mobil listrik di Indonesia diikuti oleh meningkatnya aktivitas diskusi pada platform YouTube. Komentar yang dihasilkan memuat informasi mengenai isu yang dibahas serta kecenderungan sentimen pengguna, namun jumlahnya yang besar dan tidak terstruktur menyulitkan analisis secara manual. Penelitian ini bertujuan untuk memanfaatkan BERTopic untuk pemodelan topik dan GPT untuk analisis sentimen. Data yang digunakan terdiri atas 28.054 komentar YouTube yang dipublikasikan selama periode 2022–2025. BERTopic mengelompokkan komentar ke dalam enam topik utama dengan nilai *topic coherence* sebesar 0,558. Analisis sentimen dilakukan menggunakan pendekatan *zero-shot prompting* dengan GPT-5.4-mini dan memperoleh *balanced accuracy* sebesar 0,81 serta *macro F1-score* sebesar 0,84. Hasil klasifikasi menunjukkan bahwa 61,5% komentar tergolong ke dalam kategori sentimen negatif. Integrasi hasil menunjukkan bahwa topik kebijakan subsidi dan transisi energi serta aspek teknis dan infrastruktur memiliki proporsi sentimen negatif tertinggi. Hasil eksplorasi temporal menunjukkan pergeseran fokus pembahasan dari kebijakan dan transisi energi pada 2022–2023 menuju aspek teknis, industri, ekonomi, dan minat pembelian pada 2024–2025.

Kata kunci: Analisis sentimen, BERTopic, GPT-5.4-mini, mobil listrik, YouTube



ABSTRACT

YULIANTI NURDZANAH. Utilization of BERTopic for Topic Modeling and GPT for Sentiment Analysis on YouTube Comments regarding Electric Vehicle Issues in Indonesia. Supervised by CICI SUHAENI and GERRY ALFA DITO.

The growth of the electric vehicle market in Indonesia has been accompanied by increasing discussion activity on YouTube. The comments generated contain information about the issues being discussed as well as trends in user sentiment. However, their sheer volume and unstructured nature make manual analysis difficult. This study aimed to utilize BERTopic for topic modeling and GPT for sentiment analysis. The data used consisted of 28,054 YouTube comments published during the 2022–2025 period. BERTopic grouped the comments into six main topics with a topic coherence value of 0.558. Sentiment analysis was performed using a zero-shot prompting approach with GPT-5.4-mini, achieving a balanced accuracy of 0.81 and a macro F1-score of 0.84. The classification results showed that 61.5% of comments were categorized as negative sentiment. The integration of both results showed that the subsidy policy and energy transition topic, as well as the technical and infrastructure topic, had the highest proportion of negative sentiment. Temporal exploration showed a shift in discussion focus from policy and energy transition in 2022–2023 toward technical aspects, industry, economy, and purchase intention in 2024–2025.

Keywords: BERTopic, electric vehicles, GPT-5.4-mini, sentiment analysis, YouTube



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.



PEMANFAATAN BERTOPIC UNTUK *TOPIC MODELING* DAN GPT UNTUK ANALISIS SENTIMEN PADA KOMENTAR YOUTUBE TERKAIT ISU MOBIL LISTRIK DI INDONESIA

YULIANTI NURDZANAH

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:
Eka Putri Nur Utami, S.Stat., M.Si



Judul Skripsi : Pemanfaatan BERTopic untuk *Topic Modeling* dan GPT untuk Analisis Sentimen pada Komentar YouTube terkait Isu Mobil Listrik di Indonesia

Nama : Yulianti Nurdzanah
NIM : G1401221011

Disetujui oleh

Pembimbing 1:
Cici Suhaeni, S.Si., M.Si., Ph.D

Pembimbing 2:
Gerry Alfa Dito, S.Si., M.Si

Diketahui oleh

Ketua Program Studi:
Dr. Bagus Sartono, S.Si., M.Si
NIP 197804112005011002

Tanggal Ujian:
29 Juli 2026

Tanggal Lulus:



PRAKATA

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat dan karunia-Nya sehingga karya ilmiah ini dapat diselesaikan dengan baik. Penelitian ini disusun dengan judul "Pemanfaatan BERTopic untuk *Topic Modeling* dan GPT untuk Analisis Sentimen pada Komentar YouTube terkait Isu Mobil Listrik di Indonesia". Penulisan skripsi ini tidak lepas dari dukungan berbagai pihak yang telah membantu penulis sejak masa studi hingga selesainya karya ilmiah ini. Oleh karena itu, penulis menyampaikan ucapan terima kasih kepada:

1. Kedua orangtua penulis, Ibu Newi Karnewi dan Bapak Sunanto, kakak kandung penulis, Mutia Latifah Febriany, keponakan penulis, Azhar Abdul Hakam Al'Alawy, serta seluruh keluarga besar yang telah memberikan doa, dukungan, dan semangat terbesar bagi penulis.
2. Ibu Cici Suhaeni, S.Si., M.Si., Ph.D dan Bapak Gerry Alfa Dito, S.Si, M.Si selaku komisi pembimbing yang telah meluangkan waktu untuk memberikan bimbingan, arahan, ilmu, serta masukan yang sangat berarti selama penyusunan karya ilmiah ini.
3. Ibu Aulia Rizki Firdawanti, S.Stat., M.Si selaku moderator kolokium, Ibu Laily Nissa Atul Mualifah, S.Si., M.Si selaku moderator seminar, serta Ibu Eka Putri Nur Utami, S.Stat., M.Si selaku dosen penguji skripsi yang telah memberikan kritik dan saran yang membangun demi penyempurnaan karya ilmiah ini.
4. Seluruh Bapak dan Ibu Dosen serta tenaga kependidikan Program Studi Statistika dan Sains Data IPB University yang telah membekali penulis dengan ilmu pengetahuan dan membantu selama masa perkuliahan hingga penyelesaian studi.
5. Teman-teman seperjuangan Statistika Angkatan 59, khususnya Adinda Pratiwi, Deswita Nur Alpharofi, Cindy Indriyani, Siti Arbaynah, Raihana Asma Amani, Delia Fitri Audina, Salsa Rifda Erira dan Salsabila Fayiza yang selalu menemani, memberikan semangat, serta menjadi rekan diskusi selama masa perkuliahan.
6. Verawati Ayu Lestari sebagai sahabat sekaligus rekan bertumbuh yang senantiasa menemani penulis selama empat tahun terakhir, serta selalu memberikan dukungan, semangat, dan kebersamaan selama menjalani kehidupan di bangku perkuliahan.
7. Semua pihak yang tidak dapat penulis sebutkan satu per satu, yang telah membantu baik secara langsung maupun tidak langsung dalam penyelesaian karya ilmiah ini.

Penulis menyadari bahwa karya ilmiah ini masih belum mencapai tingkat kesempurnaan yang diharapkan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun demi penyempurnaan karya ilmiah ini. Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan serta dapat memberikan kontribusi bagi pengembangan ilmu pengetahuan.

Bogor, Juli 2026

Yulianti Nurdzanah



DAFTAR ISI

DAFTAR TABEL	x
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	x
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	3
II TINJAUAN PUSTAKA	4
2.1 Mobil Listrik	4
2.2 YouTube Data API v3	4
2.3 BERTopic	4
2.4 Evaluasi Pemodelan Topik	8
2.5 Analisis Sentimen Berbasis GPT	10
2.6 Metrik Evaluasi Analisis Sentimen	12
III METODE	13
3.1 Data	13
3.2 Prosedur dan Analisis Data	13
IV HASIL DAN PEMBAHASAN	18
4.1 Pengumpulan dan Pra-Pemrosesan Data	18
4.2 Eksplorasi Data	20
4.3 Pemodelan Topik	23
4.4 Analisis Sentimen	27
4.5 Eksplorasi integratif topik, sentimen, dan dinamika waktu	31
V SIMPULAN DAN SARAN	35
5.1 Simpulan	35
5.2 Saran	35
DAFTAR PUSTAKA	37
LAMPIRAN	41
RIWAYAT HIDUP	50



DAFTAR TABEL

1	Interpretasi rentang nilai Kappa (Landis dan Koch 1977)	16
2	Contoh video yang tidak memenuhi kriteria penelitian	18
3	Contoh komentar yang tidak memenuhi kriteria penelitian	19
4	Contoh tahapan pra-pemrosesan data	20
5	statistik deskriptif panjang komentar berdasarkan jumlah token	20
6	Representasi numerik teks	23
7	Hasil evaluasi pemodelan topik	24
8	Ringkasan hasil pemodelan topik	25
9	Cohen's Kappa antarpasangan <i>anotator</i>	28
10	Fleiss' Kappa	29

DAFTAR GAMBAR

11	Representasi <i>input</i> pada BERT (Devlin <i>et al.</i> 2019)	5
12	Ilustrasi reduksi dimensi dengan UMAP (Oide dan Sugita 2022)	6
13	Arsitektur GPT (Yang <i>et al.</i> 2023)	11
14	Ringkasan jumlah data	19
15	Distribusi (a) video dan (b) komentar berdasarkan kanal	21
16	Lima video dengan komentar terbanyak	22
17	Distribusi komentar berdasarkan tahun	22
18	<i>Wordcloud</i> komentar	23
19	<i>Inter-topic distance map</i> (a) $K = 5$ (b) $K = 6$ (c) $K = 8$	25
20	Proporsi hasil pelabelan sentimen setiap anotator	28
21	Proporsi kategori sentimen berdasarkan pelabelan manual	29
22	<i>Classification report</i>	30
23	Proporsi jumlah komentar berdasarkan sentimen	31
24	Visualisasi <i>wordcloud</i> berdasarkan kategori sentimen (a) positif, (b) netral, dan (c) negatif	31
25	Distribusi sentimen berdasarkan topik	32
26	Distribusi topik berdasarkan tahun	33
27	Distribusi sentimen berdasarkan tahun	34

DAFTAR LAMPIRAN

28	Lampiran 1 Contoh kamus <i>custom normalization</i>	42
29	Lampiran 2 <i>prompt</i> pelabelan topik menggunakan ChatGPT 5.5	43
30	Lampiran 3 <i>Workflow</i> n8n analisis sentimen dengan GPT-5.4-mini	44
31	Lampiran 4 <i>Prompt</i> analisis sentimen dengan GPT-5.4-mini	45
32	Lampiran 5 Contoh komentar yang merepresentasikan setiap topik	46
33	Lampiran 6 Contoh hasil pelabelan manual	49