



SKRIPSI

SISTEM GovBERTic UNTUK ANALISIS OPINI PUBLIK TERHADAP KINERJA SATU TAHUN PERTAMA PRABOWO-GIBRAN (STUDI KASUS: MEDIA SOSIAL X)

NAURA ULAYYA NARISWARI
NPM 22083010034

DOSEN PEMBIMBING
Amri Muhaimin, S.Stat., M.Stat., M.S.
Shindi Shella May Wara, M. Stat

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2026**

LEMBAR PENGESAHAN

SISTEM GovBERTic UNTUK ANALISIS OPINI PUBLIK TERHADAP KINERJA SATU TAHUN PERTAMA PRABOWO-GIBRAN (STUDI KASUS: MEDIA SOSIAL X)

Oleh:
NAURA ULAYYA NARISWARI
NPM. 22083010034

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Sidang Skripsi Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur pada Tanggal 24 Juni 2026:

Menyetujui,

Amri Muhaimin, S.Stat., M.Stat., M.S.
NIP. 19950723 202406 1 002



(Pembimbing I)

Shindi Shella May Wara, M.Stat.
NIP. 19960518 202406 2 003



(Pembimbing II)

Wahyu Syaifullah J. S, S.Kom., M.Kom.
NIP. 19860825 202121 1 003



(Ketua Penguji)

Kartika Maulida Hindrayani, S.Kom., M.Kom.
NIP. 19920909 202203 2 009



(Penguji I)

Mengetahui,

Dekan Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

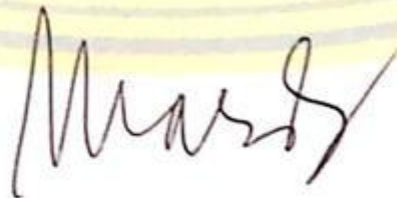
SISTEM GovBERTic UNTUK ANALISIS OPINI PUBLIK TERHADAP KINERJA SATU TAHUN PERTAMA PRABOWO-GIBRAN (STUDI KASUS: MEDIA SOSIAL X)

Oleh:
NAURA ULAYYA NARISWARI
NPM. 22083010034

Telah disetujui untuk mengikuti Ujian Skripsi

Menyetujui,

Plt Koordinator Program Studi Sains Data
Fakultas Ilmu Komputer



Dr. Ir. I Gede Susrama Mas Divasa S.T. M.T., IPU
NIP. 19700619 2021211 009

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama Mahasiswa : Naura Ulayya Nariswari
NPM : 22083010034
Program : Sarjana (S1)
Program Studi : Sains Data
Fakultas : Fakultas Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila di kemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.



Surabaya, Juni 2026
Yang Membuat Pernyataan,



NAURA ULAYYA NARISWARI
NPM. 22083010034

ABSTRAK

Nama Mahasiswa / NPM : Naura Ulayya Nariswari / 22083010034
Judul Skripsi : Sistem GovBERTic untuk Analisis Opini Publik terhadap Kinerja Satu Tahun Pertama Prabowo-Gibran (Studi Kasus: Media Sosial *X*)
Dosen Pembimbing : 1. Amri Muhaimin, S.Stat., M.Stat., M.S.
2. Shindi Shella May Wara, M. Stat.

Dalam satu tahun pertama masa pemerintahan Presiden dan Wakil Presiden yang baru, media sosial menjadi ruang utama bagi masyarakat untuk menyampaikan opini terhadap kebijakan dan kinerja pemerintah. Namun, narasi yang berkembang di Platform *X* sering kali didominasi oleh klaim yang tidak didukung metodologi analisis yang transparan sehingga berpotensi membentuk opini publik yang bias. Penelitian ini bertujuan membangun sistem GovBERTic yang mengintegrasikan pemodelan topik dan analisis sentimen untuk mengidentifikasi isu-isu utama serta kecenderungan sentimen masyarakat terhadap kinerja pemerintah berdasarkan data dari Platform *X*. Tahap pemodelan topik dilakukan menggunakan BERTopic dengan representasi semantik berbasis GovBERTic, reduksi dimensi menggunakan UMAP, klasterisasi menggunakan HDBSCAN, serta ekstraksi kata kunci melalui c-TF-IDF dan KeyBERT. Analisis sentimen dilakukan dengan membandingkan tiga model berbasis *transformer*, yaitu IndoBERT, IndoBERTweet, dan RoBERTa, menggunakan metode Stratified 5-Fold Cross Validation dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Dataset penelitian terdiri atas 5.787 *tweet* dengan 5.538 *tweet* bersih untuk dianalisis. Hasil pemodelan topik menunjukkan bahwa BERTopic berhasil menghasilkan delapan topik utama dengan *Topic Diversity* sebesar 0,9375 dan nilai *Coherence Score* (Cv) sebesar 0,4326 yang menunjukkan kualitas topik yang baik pada data media sosial. Pada tahap analisis sentimen, RoBERTa memberikan performa terbaik dibandingkan IndoBERT dan IndoBERTweet dengan rata-rata *accuracy* sebesar $74,44\% \pm 9,69\%$ dan *F1-score* makro sebesar $73,75\% \pm 10,57\%$, sehingga dipilih sebagai model klasifikasi sentimen. Integrasi hasil pemodelan topik dan analisis sentimen menunjukkan bahwa dari delapan topik yang terbentuk terdapat empat topik dengan sentimen

dominan netral, dua topik dengan sentimen dominan negatif, dan dua topik dengan sentimen dominan positif. Hasil penelitian menunjukkan bahwa GovBERTic mampu mengidentifikasi isu-isu utama beserta kecenderungan sentimen masyarakat secara objektif dan berbasis data, sehingga dapat menjadi instrumen analisis yang mendukung evaluasi persepsi publik terhadap kinerja pemerintahan di media sosial.

Kata Kunci : *analisis sentimen, BERTopic, GovBERTic, media sosial X, pemerintahan Prabowo–Gibran*

ABSTRACT

Student Name / NPM : Naura Ulayya Nariswari / 22083010034
Undergraduate thesis title : GovBERTic System for Public Opinion Analysis of Prabowo-Gibran's First-Year Performance (Case Study: Social Media X)
Advisors : 1. Amri Muhaimin, S.Stat., M.Stat., M.S.
2. Shindi Shella May Wara, M. Stat.

In the first year of the new President and Vice President's administration, social media became the main space for the public to express their opinions on government policies and performance. However, the narrative that develops on Platform X is often dominated by claims that are not supported by transparent analysis methodologies that have the potential to shape biased public opinion. This research aims to build a GovBERTic system that integrates topic modeling and sentiment analysis to identify key issues as well as the tendency of public sentiment towards government performance based on data from Platform X. The topic modeling stage was carried out using BERTopic with GovBERTic-based semantic representation, dimension reduction using UMAP, clustering using HDBSCAN, and keyword extraction through c-TF-IDF and KeyBERT. Sentiment analysis was carried out by comparing three transformer-based models, namely IndoBERT, IndoBERTtweet, and RoBERTa, using the Stratified 5-Fold Cross Validation method with accuracy, precision, recall, and F1-score metrics. The research dataset consisted of 5,787 tweets with 5,538 clean tweets to analyze. The results of the topic modeling showed that BERTopic managed to produce eight main topics with a Topic Diversity of 0.9375 and a Coherence Score (Cv) value of 0.4326 which shows good topic quality on social media data. At the sentiment analysis stage, RoBERTa performed better than IndoBERT and IndoBERTtweet with an average accuracy of $74.44\% \pm 9.69\%$ and a macro F1-score of $73.75\% \pm 10.57\%$, so it was chosen as the sentiment classification model. The integration of topic modeling and sentiment analysis results showed that of the eight topics formed, there were four topics with neutral dominant sentiment, two topics with negative dominant

sentiment, and two topics with positive dominant sentiment. The results of the study show that GovBERTic is able to identify the main issues and trends of public sentiment objectively and based on data, so that it can be an analytical instrument that supports the evaluation of public perception of government performance on social media.

Keywords: *sentiment analysis, BERTopic, GovBERTic, social media X, Prabowo–Gibran government.*

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi dengan judul **“Sistem GovBERTic untuk Analisis Opini Publik terhadap Kinerja Satu Tahun Pertama Prabowo-Gibran (Studi Kasus: Media Sosial X)”** dapat terselesaikan dengan baik.

Penulis mengucapkan terima kasih kepada Bapak Amri Muhaimin, S.Stat., M.Stat., M.S. selaku Dosen Pembimbing I dan Bapak Dr.Eng.Ir.Dwi Arman Prasetya, S.T., M.T., IPU. selaku Dosen Pembimbing II yang telah meluangkan waktunya untuk memberikan bimbingan, nasehat serta motivasi kepada penulis. Selain itu, selama penyusunan skripsi penulis juga banyak menerima bantuan dari berbagai pihak. Untuk itu penulis mengucapkan terima kasih kepada:

1. Ibu/Bapak [...] selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Ibu/Bapak [...] selaku Koordinator Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur.
3. Bapak/Ibu dosen Program Studi Sains Data yang telah memberikan ilmu dan bimbingan selama masa perkuliahan.

Penulis menyadari bahwa di dalam penyusunan skripsi ini banyak terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhirnya, dengan segala keterbatasan yang penulis miliki semoga laporan ini dapat bermanfaat bagi semua pihak umumnya dan penulis pada khususnya.

Surabaya, Juni 2026

Penulis

DAFTAR ISI

HALAMAN SAMPUL	i
LEMBAR PENGESAHAN.....	ii
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI	iv
ABSTRAK.....	v
<i>ABSTRACT</i>.....	vii
KATA PENGANTAR	x
DAFTAR ISI	xi
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvii
DAFTAR LAMPIRAN	xviii
DAFTAR NOTASI	xx
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah	7
1.3. Batasan Masalah	7
1.4. Tujuan Penelitian.....	8
1.5. Manfaat Penelitian.....	8
BAB II TINJAUAN PUSTAKA	10
2.1. Penelitian Terdahulu.....	10
2.2. Kerangka Teori	21
2.2.1. Media Sosial	21
2.2.2. Media Sosial <i>X</i>	21
2.2.3. Opini Publik dalam Sosial Media	22
2.2.4. <i>Natural Language Processing</i>	22
2.2.5. <i>Text Mining</i>	23
2.2.6. <i>Text Preprocessing</i>	25

2.2.7.	<i>Sentence Embedding</i>	27
2.2.8.	<i>Topic modeling</i>	29
2.2.9.	BERTopic.....	30
2.2.10.	Evaluasi Topic Modeling	36
2.2.11.	Analisis Sentimen	38
2.2.12.	<i>Transformer Architecture</i>	39
2.2.13.	IndoBERT	41
2.2.14.	IndoBERTweet.....	44
2.2.15.	RoBERTa	45
2.2.13.	Stratified K-Fold Cross Validation	48
2.2.14.	<i>Majority Voting</i>	49
2.2.15.	Evaluasi Klasifikasi.....	50
BAB III METODOLOGI PENELITIAN		52
3.1.	Variabel Penelitian dan Sumber Data	52
3.2.	Langkah Analisis.....	53
3.3.	Desain Sistem.....	64
BAB IV HASIL DAN PEMBAHASAN		67
4.1.	Hasil Penelitian	67
4.1.1.	Persiapan Data.....	67
4.1.2.	<i>Sentence Embedding</i>	73
4.1.3.	BERTopic.....	74
4.1.4.	<i>Model Selector</i> Sentimen	91
4.1.5.	Prediksi Sentimen per Topik BERTopic dengan Model Terpilih	100
4.1.6.	Output Final GovBERTic	106
4.2.	Tampilan Antar Muka (<i>Graphical User Interface</i>).....	108
4.2.1.	Halaman Dashboard Analisis	108
4.2.2.	Halaman <i>Model Selector</i> Sentimen.....	110
4.2.3.	Halaman <i>Insight Extraction</i>	111
BAB V PENUTUP.....		113

5.1.	Kesimpulan.....	113
5.2.	Saran Pengembangan.....	114
DAFTAR PUSTAKA		116
LAMPIRAN		119

DAFTAR GAMBAR

Gambar 2.1. Arsitektur BERTopic [24]	31
Gambar 2.2. Model <i>Sentence -BERT</i>	31
Gambar 2.3. Reduksi Dimensi.....	33
Gambar 3.1. Prediksi Sentimen per Topik	63
Gambar 4.1. Informasi Dataset.....	68
Gambar 4.2. Visualisasi Perbandingan Model Sentimen	99
Gambar 4.3. Distribusi <i>Tweet</i> per Topik dan Sentimen Dominan.....	104
Gambar 4.4. Distribusi <i>Tweet</i> per Topik dan Sentimen Dominan.....	105
Gambar 4.5. Halaman Dashboard Analisis GovBERTic	108
Gambar 4.6. Halaman Model <i>Selector</i> Sentimen	110
Gambar 4.7. Halaman <i>Insight Extraction</i>	111

DAFTAR TABEL

Tabel 2.1. Penelitian Terdahulu.....	10
Tabel 3.1. Contoh <i>tweet</i> hasil <i>scraping</i>	52
Tabel 2.2. Sample data yang akan digunakan	55
Tabel 3.4. Contoh interpretasi topik pada Model BERTopic	60
Tabel 4.1. Analisis Jumlah Topik dan Outlier.....	85
Tabel 4.2. Hasil Pembentukan Topik BERTopic	86
Tabel 4.4. Hasil Prediksi Sentimen Per Topik	102

DAFTAR LAMPIRAN

Lampiran 1. <i>Letter of Acceptance</i> (LoA)	119
Lampiran 2. <i>Kode Script</i>	120

DAFTAR NOTASI

E	:	Vektor embedding kalimat
n_R	:	Jumlah token dalam kalimat
e	:	Vektor embedding token ke- i
$Sim(A,B)$	:	Nilai cosine similarity antara vektor A dan B
$W(t,c)$	:	Bobot c -TF-IDF untuk istilah t dalam klaster c
dst	:	Frekuensi istilah t pada klaster c
Jumlah klaster/topik	:	Gambar input
N	:	Jumlah klaster yang mengandung istilah t
$DF(t)$	:	Rata-rata jumlah kata per kelas dalam korpus
A	:	Frame
$P(y_i)$	:	Probabilitas sentimen kelas ke- i
z_i	:	Nilai logit dari lapisan linear untuk kelas ke- i
TD	:	Jumlah kata unik dari seluruh topik
$f(x)m$	:	Jumlah seluruh kata kunci pada semua topik
$d(x', x)$	:	Jarak Euclidean antara titik X dan y
$K(x, y)$	:	Frekuensi kemunculan kata t dalam dokumen d
$G(x)$	:	Nilai coherence score rata-rata
σ	:	Standard deviasi dari distribusi
$G(x, y)$	:	Nilai keterkaitan semantik kata ke- i