



SKRIPSI

**KOMPARASI KINERJA ALGORITMA MACHINE
LEARNING DENGAN PENDEKATAN
BERTOPIC UNTUK ANALISIS STANCE DAN
IDENTIFIKASI TOPIK PADA OPINI PUBLIK
TERKAIT IKN**

NABILA AHLISYA FATIR

NPM 22082010017

DOSEN PEMBIMBING

Eka Dyar Wahyuni, S.Kom, M.Kom.

Nambi Sembilu, S.Kom, M.Kom.

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



SKRIPSI

**KOMPARASI KINERJA ALGORITMA MACHINE
LEARNING DENGAN PENDEKATAN
BERTOPIC UNTUK ANALISIS STANCE DAN
IDENTIFIKASI TOPIK PADA OPINI PUBLIK
TERKAIT IKN**

NABILA AHLISYA FATIR

NPM 22082010017

DOSEN PROMOTOR

Eka Dyar Wahyuni, S.Kom, M.Kom.

Nambi Sembilu, S.Kom, M. Kom.

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

Halaman ini sengaja dikosongkan

LEMBAR PENGESAHAN

KOMPARASI KINERJA ALGORITMA MACHINE LEARNING DENGAN PENDEKATAN BERTOPIC UNTUK ANALISIS STANCE DAN IDENTIFIKASI TOPIK PADA OPINI PUBLIK TERKAIT IKN

Oleh :

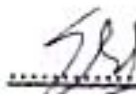
NABILA AHLISYA FATIR

NPM. 22082010017

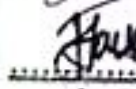
Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Sistem Informasi Fakultas Ilmu Komputer Universitas Pembangunan Veteran Jawa Timur Pada tanggal 8 Januari 2026.

Menyetujui,

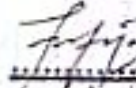
Eka Dyar Wahyuni, S.Kom, M.Kom.
NIP. 9841201 2021212 005

 (Pembimbing I)

Nambi Sembilu, S.Kom, M.Kom.
NIP. 19900516 2024061 003

 (Pembimbing II)

Seftin Fitri Ana Wati, S.Kom, M.Kom
NIP. 212199 10 320267

 (Ketua Penguji)

Abdul Rezha Efrat Najaf, S.Kom, M.Kom
NIP. 19940929 202203 1008

 (Penguji II)

Iqbal Ramadhani Mukhlis, S.Kom, M.Kom
NIP. 199303052024061002

 (Penguji III)

Menyetujui,

Dekan Fakultas Ilmu Komputer


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

Halaman ini sengaja dikosongkan

LEMBAR PERSETUJUAN

**KOMPARASI KINERJA ALGORITMA MACHINE LEARNING DENGAN
PENDEKATAN BERTOPIC UNTUK ANALISIS STANCE DAN
IDENTIFIKASI TOPIK PADA OPINI PUBLIK TERKAIT IKN**

Oleh :

NABILA AHLISYA FATIR

NPM. 22082010017

Menyetujui,

**Koordinator Program Studi Sistem Informasi
Fakultas Ilmu Komputer**



Agung Brastama Putra, S.Kom., M.Kom.

NIP. 198511242021211003

Halaman ini sengaja dikosongkan

SURAT PERNYATAAN BEBAS PLAGIASI

Yang bertanda tangan di bawah ini:

Nama : Nabila Ahlisya Fatir
NPM : 22082010017
Program : Sarjana (S1)
Program Studi : Sistem Informasi
Fakultas : Ilmu Komputer

Menyatakan bahwa dalam dokumen 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 dikemudian 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, 14 Januari 2026
Yang Membuat Pernyataan,



Nabila Ahlisya Fatir
NPM. 22082010017

Halaman ini sengaja dikosongkan

ABSTRAK

Nama Mahasiswa / NPM : Nabila Ahlisya Fatir / 22082010017
Judul Skripsi : Komparasi Kinerja Algoritma Machine Learning dengan Pendekatan BerTopic untuk Analisis Stance dan Identifikasi Topik pada Opini Publik Terkait IKN”
Dosen Pembimbing : 1. Eka Dyar Wahyuni, S.Kom, M.Kom.
2. Nambi Sembilu, S.Kom, M.Kom.

Pemindahan Ibu Kota Negara (IKN) merupakan kebijakan strategis yang memunculkan beragam sikap pro dan kontra di kalangan masyarakat, khususnya pada media sosial X yang bersifat terbuka dan berbasis teks singkat. Penelitian ini bertujuan untuk membandingkan kinerja algoritma *machine learning* klasik dalam mengklasifikasikan *stance* publik terhadap pemindahan IKN serta mengidentifikasi topik utama yang mendasari sikap tersebut. Data penelitian berupa *tweet* berbahasa Indonesia yang dikumpulkan dari media sosial X melalui teknik *scraping*. Lima algoritma klasifikasi yang diuji yaitu Support Vector Machine, Logistic Regression, Decision Tree, Random Forest, dan K-Nearest Neighbor. Evaluasi model dilakukan menggunakan *F1-score weighted* sebagai metrik utama untuk menangani ketidakseimbangan kelas. Hasil pengujian menunjukkan bahwa Random Forest merupakan model dengan performa terbaik, khususnya pada konfigurasi bigram dengan pembagian data 90:10, yang menghasilkan *F1-score weighted* sebesar 0,8011, serta nilai *accuracy*, *precision-weighted*, dan *recall-weighted* masing-masing sebesar 0,8012. Selanjutnya, penerapan metode BERTopic berhasil mengidentifikasi topik dominan seperti kebijakan dan isu sosial, investasi dan pembangunan ekonomi, serta infrastruktur. Penelitian ini menyimpulkan bahwa integrasi klasifikasi *stance* dan *topic modeling* mampu memberikan pemahaman yang komprehensif terhadap opini publik terkait pemindahan IKN.

Kata kunci: Data Mining, Stance Analysis, Machine Learning, BERTopic, IKN

Halaman ini sengaja dikosongkan

ABSTRACT

Student Name / NPM : Nabila Ahlisya Fatir / 22082010017
Thesis Title : Komparasi Kinerja Algoritma Machine Learning
dengan Pendekatan BerTopic untuk Analisis
Stance dan Identifikasi Topik pada Opini Publik
Terkait IKN”
Advisors : 1. Eka Dyar Wahyuni, S.Kom, M.Kom.
2. Nambi Sembilu, S.Kom, M.Kom.

The relocation of Indonesia’s capital city (IKN) is a strategic public policy that has generated diverse pro and contra stances among the public, particularly on social media platform X, which is open and dominated by short text-based content. This study aims to compare the performance of classical machine learning algorithms in classifying public stance toward the relocation of the IKN and to identify the main topics underlying these stances. The dataset consists of Indonesian-language tweets collected from social media X using a scraping technique. Five classification algorithms were evaluated, Support Vector Machine, Logistic Regression, Decision Tree, Random Forest, and K-Nearest Neighbor. Model evaluation was conducted using the weighted F1-score as the primary metric to address class imbalance. The experimental results indicate that Random Forest achieved the best performance, particularly with the bigram configuration and a 90:10 training–testing split, producing a weighted F1-score of 0.8011, along with accuracy, weighted precision, and weighted recall values of 0.8012. Furthermore, the application of the BERTopic method successfully identified dominant topics such as public policy and social issues, investment and economic development, and infrastructure. This study concludes that integrating stance classification with topic modeling provides a comprehensive understanding of public opinion regarding the relocation of the IKN.

Keywords: Data Mining, Stance Analysis, Machine Learning, BERTopic, IKN

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Allah SWT atas segala rahmat, karunia, dan kesempatan yang diberikan, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Komparasi Kinerja Algoritma Machine Learning dengan Pendekatan BerTopic untuk Analisis Stance dan Identifikasi Topik pada Opini Publik Terkait IKN”. Penyusunan skripsi ini bukanlah proses yang mudah, namun berkat bimbingan, dukungan, dan doa dari banyak pihak, penulis mampu melewati berbagai tantangan hingga tahap akhir. Dalam kesempatan ini, penulis ingin menyampaikan rasa terima kasih dan penghargaan kepada pihak-pihak yang telah mendukung selama proses penyusunan skripsi:

1. Mama dan Papa tercinta, yang selalu menjadi sumber kekuatan dan motivasi terbesar. Terima kasih atas doa, kasih sayang, kesabaran, serta pengorbanan yang tiada henti.
2. Ibu Eka Dyar Wahyuni, S.Kom, M.Kom dan Bapak Nambi Sembilu, S.Kom, M.Kom selaku dosen pembimbing I dan II, yang dengan sabar memberikan bimbingan, arahan, serta masukan yang sangat berharga. Terima kasih atas waktu dan perhatian yang telah diberikan di setiap tahap penyusunan skripsi ini.
3. Seluruh bapak dan ibu dosen Program Studi Sistem Informasi, yang telah menjadi bagian dari perjalanan akademik ini. Ilmu dan pengalaman yang telah diberikan akan selalu menjadi bekal berharga di masa depan.
4. Rido Izril Qonata, seseorang yang selalu membersamai, memberi semangat, motivasi, dan dukungan tanpa henti selama proses penyusunan skripsi. Terima kasih sudah selalu ada di setiap langkah, menemani di saat sulit, dan membuat perjuangan ini terasa lebih bermakna.
5. Alumni MIKTI Geng dan Pal, yang selalu menciptakan semangat positif dan inspirasi dalam meraih prestasi selama kuliah. Kehadiran kalian memberikan energi, motivasi, dan keceriaan yang tak ternilai bagi penulis.
6. Faiz Nur Fitrah Insani, saudara sekaligus dosen pembimbing pribadi yang selalu memberikan *support* moral maupun materil, serta tak pernah lelah memberikan arahan, motivasi, dan dorongan agar penulis tetap semangat menghadapi setiap tantangan dalam proses penyusunan skripsi ini.

7. Rekan-rekan seperjuangan Galaksi 2022, terima kasih karena telah menjadi keluarga pertama saat memasuki gerbang perkuliahan. Solidaritas, kebersamaan, dan dukungan kalian membuat setiap tantangan terasa lebih ringan, dan membuat penulis merasa bangga bisa menjadi bagian dari perjalanan kalian. Kehangatan serta semangat kalian akan selalu menjadi kenangan yang tak terlupakan sepanjang masa kuliah.
8. Diri sendiri, yang tak pernah lelah untuk berusaha menjadi versi terbaik setiap harinya. Terima kasih telah menapaki setiap langkah dengan keteguhan hati, kesabaran, dan tekad yang tak tergoyahkan. Perjalanan panjang menapaki jalan kuliah sambil bekerja, kini terbayar dengan haru dan rasa bangga.
9. Seluruh pihak yang tidak dapat disebutkan satu per satu, yang telah memberikan dukungan, bantuan, dan semangat dalam berbagai bentuk. Terima kasih telah menjadi bagian dari perjalanan ini.

Penulis menyadari sepenuhnya bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, saran, masukan, dan kritik yang membangun sangat penulis harapkan agar karya ini dapat terus diperbaiki di masa mendatang. Semoga skripsi ini tidak hanya bermanfaat bagi pembaca, tetapi juga menjadi referensi dan inspirasi bagi penelitian atau karya ilmiah berikutnya.

Surabaya, 26 Desember 2025

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	iii
LEMBAR PERSETUJUAN	v
SURAT PERNYATAAN ORISINALITAS	vii
ABSTRAK	ix
ABSTRACT	xi
KATA PENGANTAR.....	xiii
DAFTAR ISI.....	xv
DAFTAR GAMBAR.....	xix
DAFTAR TABEL	xxi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	6
1.6 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA.....	8
2.1 Penelitian Terdahulu	9
2.2 Landasan Teori.....	13
2.2.1 X (Twitter)	14
2.2.1 Text Mining	14
2.2.2 Pra-pemrosesan Teks	14
2.2.3 Stance Analysis.....	15
2.2.4 Binary Classification.....	15
2.2.5 TF-IDF	15
2.2.6 Confusion Matrix	16
2.2.7 Support Vector Machine (SVM)	16
2.2.8 Logistic Regression	17
2.2.9 Decision Tree	17
2.2.10 Random Forest.....	17

2.2.11	<i>K-Nearest Neighbor (KNN)</i>	18
2.2.12	Topic Modeling	18
2.2.13	BERTopic	18
2.2.14	<i>Cross-Validation</i>	19
2.2.15	<i>Hyperparameter Tuning</i>	19
BAB III METODOLOGI PENELITIAN.....		20
3.1	Studi Literatur.....	21
3.2	Analisis Kebutuhan	22
3.2.1	Kebutuhan Data	22
3.2.2	Kebutuhan <i>Software</i> dan <i>Hardware</i>	22
3.3	Pengumpulan Data.....	22
3.4	<i>Filtering</i> Data	23
3.5	Pelabelan Data	23
3.6	Eksplorasi Data.....	25
3.7	Pembagian Data.....	25
3.8	Praproses Data	25
3.8.1	<i>Cleaning</i>	26
3.8.2	<i>Case Folding</i>	26
3.8.3	<i>Normalization</i>	27
3.8.4	<i>Stopward Removal</i>	27
3.8.5	<i>Stemming</i>	27
3.8.6	<i>Tokenizing</i>	28
3.9	Perancangan Model	28
3.9.1	Uji <i>Stance Analysis</i>	29
3.9.2	Evaluasi Model	31
3.9.3	Perancangan <i>Topic Modeling</i>	32
3.9.4	Penyimpanan Model	32
3.10	Penerapan Model	33
3.11	Validasi Model	34
BAB IV HASIL DAN PEMBAHASAN.....		35
4.1	Hasil.....	35
4.1.1	Studi Literatur.....	35

4.1.2	Analisis Kebutuhan	35
4.1.3	Pengumpulan Data	36
4.1.4	Filtering Data	38
4.1.5	Pelabelan Data.....	39
4.1.6	Eksplorasi Data	42
4.1.7	Pembagian Data	44
4.1.8	Praproses Data.....	45
4.1.9	Perancangan Model.....	49
4.1.10	Penerapan Model	66
4.1.11	Validasi Model.....	70
4.2	Pembahasan.....	73
BAB V PENUTUP		78
5.1	Kesimpulan	78
5.2	Saran.....	79
DAFTAR PUSTAKA		81

Halaman ini sengaja dikosongkan

DAFTAR GAMBAR

Gambar 3.1 Diagram Alir Penelitian	19
Gambar 3.2 Tahap Pengumpulan Data	21
Gambar 3.3 Tahap Pelabelan Data	22
Gambar 3.4 Alur Praproses Data	24
Gambar 3.5 Alur Perancangan Model.....	26
Gambar 3.6 Topic Modeling	30
Gambar 4.1 Potongan Kode Pengumpulan Data.....	35
Gambar 4.2 Contoh Hasil Scraping Data	36
Gambar 4.3 Potongan Kode Tahap Filtering	37
Gambar 4.4 Kode Perhitungan Fleiss' Kappa.....	39
Gambar 4.5 Hasil Perhitungan Fleiss' Kappa	39
Gambar 4.6 Kode Metode Majority Vote	40
Gambar 4.7 Grafik Distribusi Kategori.....	41
Gambar 4.8 Wordcloud Setiap Kategori.....	42
Gambar 4.9 Grafik Tren Per Bulan	43
Gambar 4.10 Potongan Kode Pembagian Data.....	44
Gambar 4.11 Potongan Kode Fungsi Praproses Data	45
Gambar 4.12 Potongan Kode Praproses Data	46
Gambar 4.13 Wordcloud Sebelum Praproses	47
Gambar 4.14 Wordcloud Sesudah Praproses	48
Gambar 4.15 Potongan Kode Variabel Input dan Label	48
Gambar 4.16 Potongan Kode Konfigurasi Model dan Grid Hyperparameter	49
Gambar 4.17 Kode Pembagian Data	49
Gambar 4.18 Kode Ekstraksi Fitur Teks.....	50
Gambar 4.19 Potongan Kode Evaluasi Coherence dan Stability.....	58
Gambar 4.20 Hasil Perhitungan Coherence dan Stability.....	59
Gambar 4.21 Hasil Perhitungan Coherence dan Stability.....	59
Gambar 4.22 Jumlah Topik Ideal.....	60
Gambar 4.23 Informasi Topik.....	61
Gambar 4.24 Intertopic Distance Map	61
Gambar 4.25 Informasi Topik.....	62
Gambar 4.26 Kata Kunci Representatif Setiap Topik.....	62
Gambar 4.27 Kata Kunci Representatif Setiap Topik	63
Gambar 4.28 Potongan Kode Simpan Model Stance.....	66
Gambar 4.29 Potongan Kode Simpan Model Topik.....	66
Gambar 4.30 Perancangan Wireframe	67
Gambar 4.31 Integrasi Flask	67
Gambar 4.32 Fitur Prediksi Berbasis Teks	68

Gambar 4.33 Fitur Prediksi Berbasis File Input	69
Gambar 4.34 Contoh Hasil Prediksi Berbasis Teks	70
Gambar 4.35 Tampilan Awal Unggah File Input	70
Gambar 4.36 Contoh Hasil Prediksi Berbasis File Input	71
Gambar 4.37 Contoh Hasil Visualisasi Bar Chart.....	71
Gambar 4.38 Contoh Hasil Visualisasi Wordcloud.....	72
Gambar 4.39 Confusion Matrix Model Terbaik.....	74
Gambar 4.40 Classification Report Model Terbaik.....	75
Gambar 4.41 Distribusi Topik Teratas	75

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	8
Tabel 4.1 Contoh Hasil Pelabelan Data	38
Tabel 4.2 Contoh Hasil Praproses	46
Tabel 4.3 Output Skenario Pengujian Model	51
Tabel 4.4 Ringkasan Model Terbaik	56
Tabel 4.5 Interpretasi Topik	64

Halaman ini sengaja dikosongkan