

**PERBANDINGAN ALGORITMA SVM DAN NAIVE BAYES DALAM
ANALISIS SENTIMEN PEMINDAHAN IKN DI TWITTER**



SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

Bryant Christopher Hihola
00000051191

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2024**

**PERBANDINGAN ALGORITMA SVM DAN NAIVE BAYES DALAM
ANALISIS SENTIMEN PEMINDAHAN IKN DI TWITTER**



Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

Bryant Christopher Hihola

0000051191

UMMN

UNIVERSITAS

MULTIMEDIA

NUSANTARA

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA**

TANGERANG

2024

LEMBAR PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Bryant Christopher Hihola

Nomor Induk Mahasiswa : **0000051191**

Program Studi : Informatika

Jenjang : S1

Skripsi dengan judul:

**PERBANDINGAN ALGORITMA SVM DAN NAIVE BAYES DALAM
ANALISIS SENTIMEN PEMINDAHAN IKN DI TWITTER**

merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan magang maupun dalam penulisan laporan karya tulis , saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas Akhir yang telah saya tempuh.

Tangerang, 3 Juni 2024



Bryant Christopher Hihola

HALAMAN PENGESAHAN

Skripsi dengan judul
**PERBANDINGAN ALGORITMA SVM DAN NAIVE BAYES DALAM
ANALISIS SENTIMEN PEMINDAHAN IKN DI TWITTER**

oleh

Nama : Bryant Christopher Hihola
NIM : 00000051191
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Senin, 3 Juni 2024

Pukul 15.00 s/s 17.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut

Ketua Sidang



(David Agustriawan, S.Kom., M.Sc.,
Ph.D.)

NIDN: 0111223304

Penguji



(Dennis Gunawan, S.Kom., M.Sc.)

NIDN: 0320059001

Pembimbing I



(Dr. Ir. P. M. Winarno, M.Kom.)

NIDN: 330106002


Pembimbing II



(Suwito Pomalingo, S.Kom., M.Kom.)

NIDN: 0911098201

PJS Ketua Program Studi Informatika,



(Dr. Eng. Niki Prastomo, S.T., M.Sc)

NIDN: 0419128203

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH MAHASISWA

Yang bertanda tangan di bawah ini:

Nama : Bryant Christopher Hihola
NIM : 00000051191
Program Studi : Informatika
Jenjang : S1
Judul Karya Ilmiah : Perbandingan Algoritma SVM dan Naive Bayes dalam Analisis Sentimen Pemindahan IKN di Twitter

Menyatakan dengan sesungguhnya bahwa saya bersedia (**pilih salah satu**):

- Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial. Saya tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.
- Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: Dalam proses pengajuan penerbitan ke dalam jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*)*.

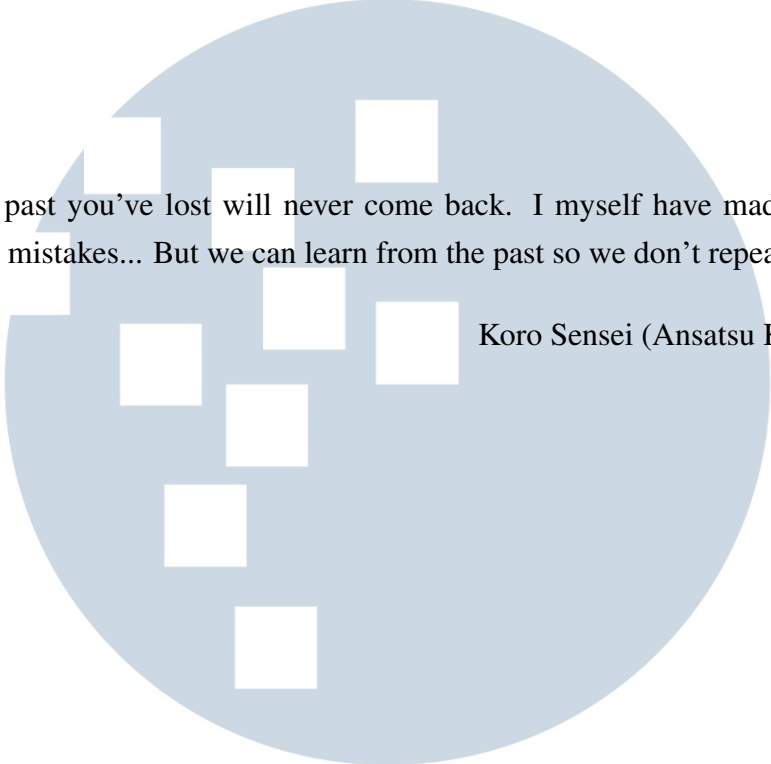
Tangerang, 3 Juni 2024



(Bryant Christopher Hihola)

* Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

Halaman Persembahan / Motto



"The past you've lost will never come back. I myself have made so many mistakes... But we can learn from the past so we don't repeat it."

Koro Sensei (Ansatsu Kyōshitsu)

UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

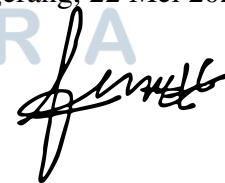
KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Perbandingan Algoritma SVM dan Naive Bayes Dalam Analisis Sentimen Pemindahan IKN di Twitter dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc, selaku PJS Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak Dr. Ir. P. M. Winarno, M.Kom., sebagai Pembimbing pertama yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya skripsi ini.
5. Bapak Suwito Pomalingo, S.Kom., M.Kom., sebagai Pembimbing kedua yang telah banyak membantu dan memberikan bimbingan atas terselesainya Skripsi ini.
6. Orang Tua dan keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan skripsi ini.

Semoga skripsi ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 22 Mei 2024



Bryant Christopher Hihola

PERBANDINGAN ALGORITMA SVM DAN NAIVE BAYES DALAM ANALISIS SENTIMEN PEMINDAHAN IKN DI TWITTER

Bryant Christopher Hihola

ABSTRAK

Pemindahan pusat pemerintahan Indonesia dari Jakarta ke lokasi baru atau wilayah lain telah menjadi topik perbincangan sejak masa kepemimpinan Presiden Soekarno sampai Susilo Bambang Yudhoyono. Meskipun demikian, hingga saat ini rencana tersebut belum terealisasi. Kehadiran pemindahan IKN Indonesia tentu memunculkan berbagai respons, terutama dari masyarakat Indonesia. Mengingat bahwa perubahan ibu kota negara Indonesia merupakan isu yang sangat sensitif, banyak pembahasannya terjadi di media sosial, termasuk di platform seperti *twitter*. Oleh karena itu, dilakukan penelitian yang bertujuan untuk mengklasifikasikan sentimen dari opini masyarakat pengguna *twitter* terhadap pemindahan ibukota negara. Pada penelitian ini terdapat model yang dibuat untuk klasifikasi menggunakan dua algoritma *Naive Bayes Classifier* dan *Support Vector Machine* untuk membandingkan tingkat akurasi kedua algoritma tersebut. Hasil penelitian ini adalah akurasi tertinggi pada skenario *balance* data didapat pada rasio 90:10 oleh *Naive Bayes* sebesar 94,78% dibanding dengan *Support Vector Machine* sebesar 92,87%. Sedangkan pada skenario *imbalance* data, akurasi tertinggi didapat pada rasio 90:10 oleh *Support Vector Machine* sebesar 84,88% dibanding dengan *Naive Bayes* sebesar 81,10%. Mayoritas sentimen pada pemindahan IKN adalah positif. Oleh karena itu, dapat disimpulkan bahwa sentimen pengguna *twitter* terhadap pemindahan IKN adalah positif.

Kata kunci: *Analisis Sentimen, Confusion Matrix, Naive Bayes Classifier, Support Vector Machine*

U M N
U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

Comparison of SVM and Naive Bayes Algorithms in Sentiment Analysis of IKN Transfers on Twitter

Bryant Christopher Hihola

ABSTRACT

Moving the center of Indonesian government from Jakarta to a new location or region others have been a topic of conversation since President Soekarno's leadership to Susilo Bambang Yudhoyono. Nevertheless, until now this plan has not been realized. The presence of the transfer of Indonesian IKN is certain elicited various responses, especially from the Indonesian people. Remember that changing the capital city of Indonesia is a very sensitive issue, much of the discussion occurs on social media, including on platforms such as Twitter. Therefore, research was carried out which aimed to classify sentiment from the opinion of the Twitter user community regarding the relocation of the capital country. In this research, a model was created for classification using two algorithms Naive Bayes Classifier and Support Vector Machine to compare the level of accuracy of the two algorithms. The results of this research show that in the imbalance data scenario, the best accuracy was achieved on 90:10 ratio by the Support Vector Machine model with 84.88% compared to Naive Bayes model with 81.10%, while for the balanced data scenario, the highest accuracy was achieved by the Naive Bayes model with 94.78% compared to Support Vector Machine model with 92.87%. The majority of sentiments for the IKN transfers is positive. Therefore it can be concluded that sentiment twitter users regarding IKN transfers are positive.

Keywords: *Naive Bayes Classifier, Confusion Matrix, Sentiment Analysis, Support Vector Machine*

U M M N
U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI ILMIAH	iv
HALAMAN PERSEMBAHAN/MOTO	v
KATA PENGANTAR	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
DAFTAR KODE	xiii
DAFTAR LAMPIRAN	xiv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Permasalahan	2
1.4 Tujuan Penelitian	2
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3
BAB 2 LANDASAN TEORI	5
2.1 Penelitian Terdahulu	5
2.2 Media Sosial Twitter	7
2.3 Text Mining	8
2.4 Text Preprocessing	8
2.5 Ekstraksi Fitur (TF-IDF)	10
2.6 SMOTE	11
2.7 K-Fold Cross-Validation	11
2.8 Analisis Sentimen	12
2.9 <i>VADER (Valence-Aware Dictionary and sEntiment Reasoner)</i>	12
2.10 <i>Naive Bayes Classifier</i>	13
2.11 <i>Support Vector Machine</i>	14
2.12 <i>Confusion Matrix</i>	14
2.13 <i>ROC - AUC Score</i>	16
BAB 3 METODOLOGI PENELITIAN	17
3.1 Metode Penelitian	17
3.2 Alur Penelitian	17
3.2.1 Studi Literatur	18
3.2.2 Pengumpulan Data	19
3.2.3 Proses Pengolahan Teks (<i>Text Preprocessing</i>)	19
3.2.4 Klasifikasi Sentimen dan Implementasi Algoritma	22
BAB 4 HASIL DAN DISKUSI	25
4.1 Spesifikasi Sistem	25
4.2 Pengumpulan Data	25
4.3 Proses Pengolahan Teks (<i>Text Preprocessing</i>)	29
4.3.1 <i>Data Cleaning</i>	29
4.3.2 <i>Translate</i>	31
4.3.3 <i>Case Folding</i>	33

4.3.4	Tokenizing	34
4.3.5	Filtering	36
4.3.6	Lemmatization	37
4.4	Klasifikasi Sentimen dan Implementasi Algoritma	39
4.4.1	Data Labelling	40
4.4.2	TF-IDF	42
4.4.3	Data Balancing	43
4.4.4	Grid Search	45
4.4.5	Validation	47
4.4.6	Modelling dan Evaluasi Model	47
BAB 5	SIMPULAN DAN SARAN	68
5.1	Simpulan	68
5.2	Saran	68
	DAFTAR PUSTAKA	69



DAFTAR GAMBAR

Gambar 3.1	Diagram alir metodologi penelitian.	18
Gambar 3.2	Diagram alir penerapan <i>Text Preprocessing</i>	20
Gambar 3.3	<i>Flow</i> terjemahan data menggunakan <i>google.trans</i>	20
Gambar 3.4	<i>Flowchart</i> klasifikasi sentimen dan implementasi kedua algoritma.	22
Gambar 4.1	Tampilan <i>tweet-harvest</i> ketika <i>crawling data</i>	27
Gambar 4.2	Tampilan <i>tweet-harvest</i> ketika selesai <i>crawling data</i>	27
Gambar 4.3	Tampilan data <i>imbalance</i> sebelum <i>oversampling</i>	44
Gambar 4.4	Tampilan data <i>balance</i> setelah <i>oversampling</i>	45
Gambar 4.5	<i>Confusion matrix</i> untuk model <i>Naive Bayes</i>	52
Gambar 4.6	<i>Confusion matrix</i> untuk <i>balanced data</i> model <i>Naive Bayes</i>	54
Gambar 4.7	<i>Confusion matrix</i> untuk model <i>Support Vector Machine</i>	56
Gambar 4.8	<i>Confusion matrix</i> untuk <i>balanced data</i> model <i>Support Vector Machine</i>	58
Gambar 4.9	<i>ROC - AUC Score</i> untuk <i>balanced data</i> model <i>Naive Bayes</i>	60
Gambar 4.10	<i>ROC - AUC Score</i> untuk <i>balanced data</i> model <i>Support Vector Machine</i>	61
Gambar 4.11	<i>Learning curve</i> untuk <i>overfitting test</i> model <i>Naive Bayes</i>	62
Gambar 4.12	<i>Learning curve</i> untuk <i>overfitting test</i> model <i>Support Vector Machine</i>	63
Gambar 4.13	<i>Word Cloud</i> Sentimen Positive model <i>Naive Bayes</i>	64
Gambar 4.14	<i>Word Cloud</i> Sentimen netral model <i>Naive Bayes</i>	65
Gambar 4.15	<i>Word Cloud</i> Sentimen negatif model <i>Naive Bayes</i>	65
Gambar 4.16	<i>Word Cloud</i> Sentimen positif model <i>Support Vector Machine</i>	66
Gambar 4.17	<i>Word Cloud</i> Sentimen netral model <i>Support Vector Machine</i>	66
Gambar 4.18	<i>Word Cloud</i> Sentimen negatif model <i>Support Vector Machine</i>	67



DAFTAR TABEL

Tabel 2.1	Riset Gap	5
Tabel 2.2	Ilustrasi Pembagian Data dalam <i>K-Fold Cross Validation</i> .	11
Tabel 2.3	<i>Confusion Matrix</i> yang terdiri dari prediksi dan nilai aktual positif dan negatif.	15
Tabel 4.1	Hasil <i>data crawling</i> menggunakan <i>tweet-harvest</i>	28
Tabel 4.2	Hasil <i>data cleaning</i>	30
Tabel 4.3	Hasil <i>translate</i>	32
Tabel 4.4	Hasil <i>case fold</i>	33
Tabel 4.5	Hasil <i>tokenize</i>	35
Tabel 4.6	Hasil <i>filtering</i>	36
Tabel 4.7	Hasil <i>Lemmatization</i>	38
Tabel 4.8	Hasil <i>labelling</i>	41
Tabel 4.9	Total Data Klasifikasi Sentimen	42
Tabel 4.10	Kombinasi parameter yang digunakan dalam <i>Grid Search</i> .	47
Tabel 4.11	Hasil uji perbandingan <i>Split Data</i> model <i>Naive Bayes</i> dengan <i>balance data</i>	49
Tabel 4.12	Hasil uji perbandingan <i>split data</i> model <i>Support Vector Machine</i> dengan <i>balance data</i>	50
Tabel 4.13	Hasil uji perbandingan <i>split data</i> model <i>Naive Bayes</i> dengan <i>imbalance data</i>	50
Tabel 4.14	Hasil uji perbandingan <i>split data</i> model <i>Support Vector Machine</i> dengan <i>imbalance data</i>	51
Tabel 4.15	Hasil akhir perbandingan akurasi model <i>Naive Bayes</i> dan <i>SVM</i> dengan <i>Balanced</i> dan <i>Imbalanced Data</i>	51
Tabel 4.16	Tabel <i>Confusion Matrix</i> sentimen model <i>NBC</i>	52
Tabel 4.17	Tabel <i>Confusion Matrix</i> sentimen model <i>NBC</i> setelah <i>Oversampling</i>	54
Tabel 4.18	Tabel <i>Confusion Matrix</i> sentimen model <i>SVM</i>	56
Tabel 4.19	Tabel <i>Confusion Matrix</i> sentimen model <i>SVM</i> setelah <i>Oversampling</i>	58

UNIVERSITAS
MULTIMEDIA
NUSANTARA

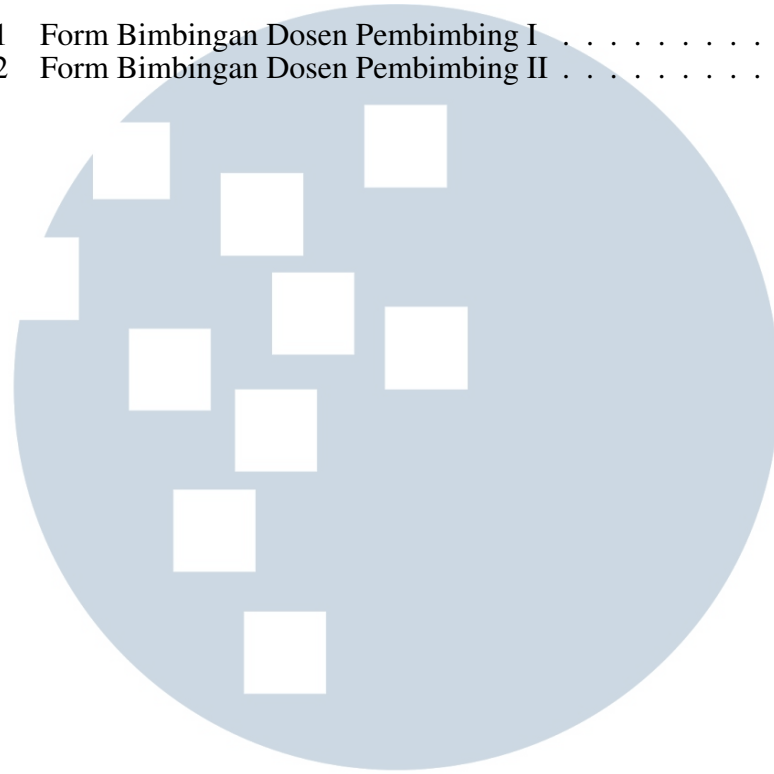
DAFTAR KODE

4.1	<i>Script python</i> pengumpulan data dengan <i>tweet-harvest</i>	25
4.2	<i>Script python</i> penggabungan dataset dan penghapusan duplikat	29
4.3	<i>Script python</i> pembersihan data	29
4.4	<i>Script python</i> terjemahan data	31
4.5	<i>Script python</i> untuk <i>case folding</i>	33
4.6	<i>Script python</i> untuk <i>tokenizing</i> data	34
4.7	<i>Script python</i> untuk <i>filterin stopwords removal</i> data	36
4.8	<i>Script python</i> untuk <i>Lemmatization</i> data	37
4.9	<i>Script python</i> untuk <i>labelling</i> data	40
4.10	<i>Script python</i> untuk pembobotan <i>TF-IDF</i>	43
4.11	<i>Script python</i> untuk <i>balancing data</i>	44
4.12	<i>Script python</i> untuk <i>Grid Search</i>	45
4.13	<i>Script python</i> untuk inialisasi <i>Stratified K-Fold</i>	47
4.14	<i>Script python</i> untuk training, testing, dan evaluasi model	48



DAFTAR LAMPIRAN

Lampiran 1	Form Bimbingan Dosen Pembimbing I	72
Lampiran 2	Form Bimbingan Dosen Pembimbing II	74



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA