



**PERBANDINGAN METODE ALGORITMA *NAÏVE BAYES* DAN
DECISION TREE PADA ANALISIS SENTIMEN MASYARAKAT
TERHADAP CALON PRESIDEN INDONESIA TAHUN 2024**

SKRIPSI

FARHAN YUSUF

2010512082

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI S1 SISTEM INFORMASI

2024



**PERBANDINGAN METODE ALGORITMA *NAÏVE BAYES* DAN
DECISION TREE PADA ANALISIS SENTIMEN MASYARAKAT
TERHADAP CALON PRESIDEN INDONESIA TAHUN 2024**

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana
Komputer**

FARHAN YUSUF

2010512082

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI S1 SISTEM INFORMASI

2024

PERNYATAAN ORISINALITAS

PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Farhan Yusuf

NIM : 2010512082

Tanggal : 26 Juli 2024

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan berlaku.

Jakarta, 26 Juli 2024

Yang Menyatakan,



(Farhan Yusuf)

**PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK
KEPENTINGAN AKADEMIS**

**PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI
UNTUK KEPENTINGAN AKADEMIS**

Sebagai civitas akademik Universitas Pembangunan Nasional “Veteran”
Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Farhan Yusuf

NIM : 2010512082

Fakultas : Ilmu Komputer

Program Studi : S1 – Sistem Informasi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional “Veteran” Jakarta Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

**Perbandingan Metode Algoritma *Naïve Bayes* dan *Decision Tree* Pada
Analisis Sentimen Masyarakat Terhadap Calon Presiden Indonesia Tahun
2024**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional “Veteran” Jakarta berhak menyimpan, mengalih data/formatkan, mengelola dalam bentuk pangkalan data (database), merawat dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenar-benarnya.

Dibuat di : Jakarta

Pada tanggal : 26 Juli 2024

Yang Menyatakan,



Farhan Yusuf

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

Skripsi ini diajukan oleh

Nama : Farhan Yusuf

NIM : 2010512082

Program Studi : S1 Sistem Informasi

Judul Tugas Akhir : Perbandingan Metode Algoritma Naïve Bayes dan Decision Tree
Pada Analisis Sentimen Masyarakat Terhadap Pemilihan Umum
Calon Presiden 2024

Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian dari persyaratan yang diperlukan untuk memperoleh gelar Sarjana Ilmu Komputer pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jakarta.



Ati Zaidiah, S.Kom, MTL.

Penguji 1



Zatin Niqotaini, S.Tr.Kom., M.Kom.

Penguji 2



Nurhafifah Matondang, S.Kom., MM.,

M.T.I

Pembimbing 1



Intan Hesti Indriana, SE., MM

Pembimbing 2



Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM

Dekan



Anita Muliawati, S.Kom., MTL.

Ketua Program Studi S1 Sistem Informasi

Ditetapkan di : Jakarta

Tanggal Ujian : 2 Juli 2024

PERBANDINGAN METODE ALGORITMA NAÏVE BAYES DAN DECISION TREE PADA ANALISIS SENTIMEN MASYARAKAT TERHADAP CALON PRESIDEN INDONESIA TAHUN 2024

ABSTRAK

Pemilihan presiden Indonesia tahun 2024 banyak menarik minat dan diskusi publik yang signifikan di platform media sosial seperti X. Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap para calon presiden menggunakan algoritma Naïve Bayes dan Decision Tree. Data dikumpulkan dari X selama tiga periode: pendaftaran calon pada Oktober 2023, sebelum pemilihan pada Februari 2024, dan setelah pemilihan pada pertengahan Februari 2024. Pra-pemrosesan data meliputi *case folding*, *data cleaning*, *text normalization*, *stemming*, *stop-word removal*, and *tokenization*. Penelitian ini mengklasifikasikan sentimen menjadi kategori positif dan negatif serta mengevaluasi kinerja algoritma berdasarkan metrik akurasi, presisi, *recall*, dan *F1-score*. Hasilnya menunjukkan bahwa kedua algoritma secara efektif mengklasifikasikan sentimen, dengan *F1-score* untuk Naïve Bayes: Anies Baswedan sebesar 79%, Prabowo Subianto sebesar 81%, dan Ganjar Pranowo sebesar 89%, sedangkan Decision Tree mencapai: Anies Baswedan sebesar 80%, Prabowo Subianto sebesar 76%, dan Ganjar Pranowo sebesar 87%. Setiap algoritma memiliki kelebihan dan kekurangannya sendiri untuk menangkap sentimen pemilih yang beragam. Penelitian ini berkontribusi pada pemahaman dinamika opini publik dalam pemilihan umum calon presiden di Indonesia dan memberi wawasan untuk studi analisis sentimen di masa depan, terutama dalam strategi kampanye politik dan pemantauan pemilu secara real-time.

Kata Kunci: Analisis Sentimen, *Naïve Bayes*, *Decision Tree*, Pemilihan Presiden Indonesia 2024, X

**COMPARISON OF NAÏVE BAYES AND DECISION TREE ALGORITHMS
IN SENTIMENT ANALYSIS OF PUBLIC OPINION TOWARDS THE 2024
INDONESIAN PRESIDENTIAL CANDIDATES**

ABSTRACT

The 2024 Indonesian presidential election has generated significant public interest and discussion on social media platforms like X. This research aims to analyze public sentiment towards the presidential candidates using Naïve Bayes and Decision Tree algorithms. Data were gathered from X during three periods: candidate registration in October 2023, pre-election in February 2024, and post-election in mid-February 2024. Data preprocessing involved case folding, data cleaning, text normalization, stemming, stop-word removal, and tokenization. This study classifies sentiments into positive and negative categories and evaluates the algorithms' performance based on accuracy, precision, recall, and F1-score metrics. Results indicate that both algorithms effectively classify sentiments, with F1-scores for Naïve Bayes: Anies Baswedan at 79%, Prabowo Subianto at 81%, and Ganjar Pranowo at 89%, while Decision Tree achieved: Anies Baswedan at 80%, Prabowo Subianto at 76%, and Ganjar Pranowo at 87%. Each algorithm has its strengths and weaknesses in capturing diverse voter sentiments. This research contributes to understanding the dynamics of public opinion in Indonesia's presidential election and provides insights for future sentiment analysis studies, particularly in political campaign strategies and real-time election monitoring.

Keywords: *Sentiment Analysis, Naïve Bayes, Decision Tree, Indonesia Presidential Election 2024, X*

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas rahmat dan hidayah-Nya sehingga peneliti dapat menyelesaikan Tugas Akhir dengan baik.

Dalam menyelesaikan Tugas Akhir ini tentunya peneliti tidak lepas dari banyak bantuan dan dukungan dari berbagai pihak. Oleh karena itu peneliti ingin menyampaikan terima kasih kepada:

1. Allah SWT. Atas segala rahmat dan karunia-Nya.
2. Kedua orang tua dan keluarga atas doa dan dukungannya.
3. Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM, selaku dekan Fakultas Ilmu Komputer.
4. Ibu Anita Muliawati, S.Kom., MTI. selaku Ketua Program Studi Sarjana Jurusan Sistem Informasi.
5. Ibu Nur Hafifah Matondang, S.Kom., M.M., M.T.I., selaku Dosen Pembimbing 1.
6. Dra. Intan Hesti Indriana, M.M., selaku Dosen Pembimbing 2.
7. Teman-teman, kerabat, dan saudara yang telah mendukung dan membantu satu sama lain.
8. Seluruh pihak yang telah membantu yang tidak mungkin untuk penulis sebutkan satu persatu.

Peneliti menyadari bahwa penyusunan Tugas Akhir ini masih jauh dari kata sempurna, oleh karena itu peneliti mengharapkan kritik dan saran yang membangun dari semua pihak.

Jakarta, 20 Juni 2024

Peneliti

DAFTAR ISI

PERNYATAAN ORISINALITAS	ii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS	iii
LEMBAR PENGESAHAN	iv
ABSTRAK	v
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR RUMUS	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiv
BAB I	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	5
1.6 Luaran yang Diharapkan	5
1.7 Sistematika Penulisan	6
BAB II	7
TINJAUAN PUSTAKA	7
2.1 Pemilihan Umum	7
2.2 Analisis Sentimen	7
2.3 <i>Text Mining</i>	7
2.4 <i>Web Scraping</i>	8
2.5 <i>Text Preprocessing</i>	8
2.5.1 <i>Case Folding</i>	8
2.5.2 <i>Data Cleaning</i>	9
2.5.3 <i>Text Normalization</i>	9
2.5.4 <i>Stemming</i>	9
2.5.5 <i>Stop Words Removal</i>	9
2.5.6 <i>Tokenization</i>	9
2.6 Pembobotan TF-IDF	10
2.7 Algoritma Naive Bayes	10
2.8 Algoritma <i>Decision Tree</i>	11
2.9 Fleiss Kappa	12
2.10 Penelitian Terdahulu	13
BAB III	19
METODOLOGI PENELITIAN	19
3.1 Kerangka Pikir	19
3.1.1 Identifikasi Masalah	20
3.1.2 Studi Literatur	20
3.1.3 <i>Crawling Data</i>	20

3.1.4	Pelabelan Data.....	20
3.1.5	<i>Text Preprocessing</i>	21
3.1.6	Pembobotan TF-IDF	22
3.1.7	Klasifikasi Dengan Multinomial Naive Bayes.....	22
3.1.8	Klasifikasi Dengan <i>Decision Tree</i>	23
3.1.9	Evaluasi.....	24
3.1.10	Hasil Analisis	25
3.2	Perangkat Penelitian	25
3.2.1	Perangkat Keras	25
3.2.2	Perangkat Lunak.....	26
3.3	Jadwal Penelitian	26
BAB IV		27
HASIL DAN PEMBAHASAN.....		27
4.1	Pengumpulan Data.....	27
4.2	Pelabelan Data	29
4.3	<i>Text Preprocessing</i>	33
4.3.1	<i>Case Folding</i>	33
4.3.2	<i>Data Cleaning</i>	33
4.3.3	<i>Text Normalization</i>	34
4.3.4	<i>Stemming</i>	35
4.3.5	<i>Stop Words Removal</i>	35
4.3.6	<i>Tokenization</i>	36
4.4	Pembobotan TF-IDF.....	36
4.5	Pembagian Data.....	37
4.6	<i>Resampling</i>	38
4.7	Klasifikasi Model Dengan Multinomial Naive Bayes	39
4.8	Klasifikasi Model Dengan <i>Decision Tree</i>	39
4.9	Pengujian dan Evaluasi Model	40
4.9.1	Hasil Evaluasi Model Anies Baswedan	40
4.9.2	Hasil Evaluasi Model Prabowo Subianto.....	43
4.9.3	Hasil Evaluasi Model Ganjar Pranowo	46
4.9.4	Perbandingan Model Berdasarkan Algoritma.....	49
4.10	Analisis Hasil Klasifikasi.....	50
4.10.1	<i>Wordcloud</i> Anies Baswedan	51
4.10.2	<i>Wordcloud</i> Prabowo Subianto.....	52
4.10.3	<i>Wordcloud</i> Ganjar Pranowo.....	53
BAB V.....		55
PENUTUP.....		55
5.1	Simpulan.....	55
5.2	Saran.....	56
DAFTAR PUSTAKA		57
LAMPIRAN.....		xv

DAFTAR RUMUS

No.	Rumus	Keterangan	Halaman
(2.1)	$TF_{t,d} = F_{t,d}$	Rumus dalam menghitung TF	10
(2.2)	$IDF_t = \log \frac{N}{DF_t}$	Rumus dalam menghitung IDF	10
(2.3)	$W_{t,d} = TF_{t,d} \times IDF_t$	Rumus dalam menghitung TF-IDF	10
(2.4)	$P(Y X) = \frac{p(x y)p(Y)}{p(X)}$	Rumus Teorema Naïve Bayes	11
(2.5)	$kappa = \frac{p_0 - p_e}{1 - p_e}$	Rumus Persamaan <i>Kappa Value</i>	12
(2.6)	$P_0 = \frac{1}{n} \sum_{i=1}^n q_i$	Rumus perhitungan persentase jumlah pengukuran antar <i>rater</i>	12
(2.7)	$P_e = \sum_{j=1}^k q_j^2$	Rumus perhitungan persentase jumlah perubahan antar <i>rater</i>	12
(2.8)	$q_i = \frac{1}{m(m-1)} \left(\sum_{j=1}^k m_{ij}^2 - m_{ij} \right)$	Rumus mencari nilai q_i untuk mendapatkan P_0	12
(2.9)	$q_j = \frac{x.j}{nm} = \frac{1}{nm} \sum_{i=1}^n x_{ij}$	Rumus mencari nilai q_j untuk mendapatkan P_e	13
(3.1)	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$	Rumus menghitung <i>accuracy</i> klasifikasi Naïve Bayes	25
(3.2)	$Precision = \frac{TP}{TP + FP}$	Rumus menghitung <i>precision</i> klasifikasi Naïve Bayes	25

(3.3)	$Recall = \frac{TP}{TP+FN}$	Rumus menghitung <i>recall</i> klasifikasi Naïve Bayes	25
(3.4)	$F1-Score = 2 \times \frac{Precision \times Recall}{Precision+Recall}$	Rumus menghitung <i>F1-Score</i> klasifikasi Naïve Bayes	25

DAFTAR TABEL

Tabel 2.1 Interpretasi Fleiss Kappa.....	13
Tabel 2.2 Penelitian Terdahulu (Peneliti, 2024)	13
Tabel 3.1 Jadwal Penelitian (Peneliti, 2024).....	26
Tabel 4.1 Menghapus data <i>tweet</i> yang tidak memiliki sentimen (Peneliti, 2024) 28	
Tabel 4.2 Ulasan dan <i>Labelling</i> dari Anonotator pada Data <i>Tweet</i> (Peneliti, 2024)	29
Tabel 4.3 Perhitungan <i>Fleiss Kappa</i> Anies Baswedan (Peneliti, 2024)	30
Tabel 4.4 <i>Case Folding</i> (Peneliti, 2024)	33
Tabel 4.5 <i>Data Cleaning</i> (Peneliti, 2024).....	34
Tabel 4.6 <i>Text Normalization</i> (Peneliti, 2024).....	34
Tabel 4.7 <i>Stemming</i> (Peneliti, 2024).....	35
Tabel 4.8 <i>Stop Words Removal</i> (Peneliti, 2024)	35
Tabel 4.9 <i>Tokenization</i> (Peneliti, 2024).....	36
Tabel 4.10 Pembagian Data untuk Calon Presiden Ganjar Pranowo (Peneliti, 2024)	37
Tabel 4.11 Pembagian Data untuk Calon Presiden Anies Baswedan (Peneliti, 2024)	37
Tabel 4.12 Pembagian Data untuk Calon Presiden Prabowo Subianto (Peneliti, 2024)	38
Tabel 4.13 Hasil <i>Confusion Matrix Model</i> Anies Baswedan dengan Algoritma Multinomial Naive Bayes (Peneliti, 2024).....	41
Tabel 4.14 Hasil <i>Confusion Matrix Model</i> Anies Baswedan dengan Algoritma <i>Decision Tree</i> (Peneliti, 2024)	42
Tabel 4.15 Hasil <i>Confusion Matrix Model</i> Prabowo Subianto dengan Algoritma Multinomial Naive Bayes (Peneliti, 2024).....	43
Tabel 4.16 Hasil <i>Confusion Matrix Model</i> Prabowo Subianto dengan Algoritma <i>Decision Tree</i> (Peneliti, 2024)	45
Tabel 4.17 Hasil <i>Confusion Matrix Model</i> Ganjar Pranowo dengan Algoritma Multinomial Naive Bayes	46
Tabel 4.18 Hasil <i>Confusion Matrix Model</i> Ganjar Pranowo dengan Algoritma <i>Decision Tree</i> (Peneliti, 2024)	47
Tabel 4.19 Hasil <i>Confusion Matrix Model</i> Menggunakan Algoritma Multinomial Naive Bayes (Peneliti, 2024).....	49
Tabel 4.20 Hasil <i>Confusion Matrix Model</i> Menggunakan Algoritma <i>Decision Tree</i> (Peneliti, 2024).....	50
Tabel 4.21 Hasil <i>Worldcloud</i> dari Model Anies Baswedan.....	51
Tabel 4.22 Hasil <i>Wordcloud</i> dari Model Prabowo Subianto	52
Tabel 4.23 Hasil <i>Wordcloud</i> dari Model Ganjar Pranowo.....	53

DAFTAR GAMBAR

Gambar 1.1 <i>Ranking</i> pengguna media sosial X per Januari 2023 (Statista, 2023) .	2
Gambar 3.1 Kerangka Pikir (Peneliti, 2024).....	19
Gambar 3.2 <i>The Basic Structure of a Confusion Matrix</i> (Ahmed, N. A., 2023) ..	24
Gambar 4.1 <i>Crawling</i> data menggunakan Tweet Harvest (Peneliti, 2024)	27
Gambar 4.2 Hasil <i>crawling</i> data dari media sosial X (Peneliti, 2024).....	28
Gambar 4.3 Pembobotan TF-IDF pada Data <i>Tweet</i> (Peneliti, 2024).....	37
Gambar 4.4 <i>Resampling</i> Kelas Minoritas Pada Dataset (Peneliti, 2024).....	38
Gambar 4.5 Proses <i>modelling</i> dengan algoritma Multinomial Naive Bayes (Peneliti, 2024)	39
Gambar 4.6 Proses <i>modelling</i> dengan algoritma Decision Tree (Peneliti, 2024) .	40

DAFTAR LAMPIRAN

Lampiran 1. Lembar Hasil Turnitin	xv
Lampiran 2. <i>Source Code</i>	xxvii
Lampiran 3. Daftar Riwayat Hidup.....	xxxiv