



SKRIPSI

**PERBANDINGAN ALGORITMA DECISION TREE
DAN NAÏVE BAYES CLASSIFIER PADA
ANALISIS SENTIMEN TWITTER MENGENAI
KEBIJAKAN PENGHAPUSAN KEWAJIBAN
SKRIPSI**

ILHAM AINUR IDHANA

NPM 19081010099

DOSEN PEMBIMBING

Dr. Basuki Rahmat, S.Si., MT.

Eva Yulia Puspaningrum, S.Kom., M.Kom.

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025**



SKRIPSI

PERBANDINGAN ALGORITMA DECISION TREE DAN NAÏVE BAYES CLASSIFIER PADA ANALISIS SENTIMEN TWITTER MENGENAI KEBIJAKAN PENGHAPUSAN KEWAJIBAN SKRIPSI

ILHAM AINUR IDHANA
NPM 19081010099

DOSEN PEMBIMBING
Dr. Basuki Rahmat, S.Si., MT.
Eva Yulia Puspaningrum, S.Kom., M.Kom.

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025**

Halaman ini sengaja dikosongkan

LEMBAR PENGESAHAN

PERBANDINGAN ALGORITMA DECISION TREE DAN NAÏVE BAYES CLASSIFIER PADA ANALISIS SENTIMEN TWITTER MENGENAI KEBIJAKAN PENGHAPUSAN KEWAJIBAN SKRIPSI

Oleh :

ILHAM AINUR IDHANA

NPM. 19081010099

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 11 Maret 2025

Menyetujui

Dr. Basuki Rahmat, S.Si. MT.

NIP. 19690723 2021211 002

(Pembimbing I)

Eva Yulia Puspaningrum, S.Kom., M.Kom.

NIP. 19890705 2021212 002

(Pembimbing II)

Fetty Tri Anggraeny, S.Kom. M.Kom

NIP. 19820211 2021212 005

(Ketua Penguji)

Muhammad Muharrom A., S.Kom., M.Kom.

NIP. 19950601 202203 1 006

(Anggota Penguji II)

Mengetahui

Dekan Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT

NIP. 19681126 199403 2 001

Halaman ini sengaja dikosongkan

LEMBAR PERSETUJUAN

PERBANDINGAN ALGORITMA DECISION TREE DAN NAÏVE BAYES CLASSIFIER PADA ANALISIS SENTIMEN TWITTER MENGENAI KEBIJAKAN PENGHAPUSAN KEWAJIBAN SKRIPSI

Oleh :

ILHAM AINUR IDHANA

NPM. 19081010099



Menyetujui

**Koordinator Program Studi Informatika
Fakultas Ilmu Komputer**

Fetty Tri Anggraeny, S.Kom., M.Kom.

NIP. 19820211 202121 2 005

Halaman ini sengaja dikosongkan

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan dibawah ini:

Nama : Ilham Ainur Idhana
NPM : 19081010099
Program : Sarjana(S1)
Program Studi : Informatika
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 dikemudian hari ditemulan 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, 11 Maret 2025
Yang Membuat Pernyataan,



Ilham Ainur Idhana
NPM. 19081010099

Halaman ini sengaja dikosongkan

ABSTRAK

Nama Mahasiswa / NPM : Ilham Ainur Idhana / 19081010099
Judul Skripsi : Perbandingan Algoritma Decision Tree dan Naïve Bayes Classifier pada Analisis Sentimen Twitter Mengenai Kebijakan Penghapusan Skripsi
Dosen Pembimbing : 1. Dr. Basuki Rahmat, S.Si. MT.
2. Eva Yulia Puspaningrum, S.Kom., M.Kom.

Kebijakan penghapusan kewajiban skripsi sebagai syarat kelulusan mahasiswa S1/D4 yang diatur dalam Permendikbudristek Nomor 53 Tahun 2023 telah memunculkan berbagai reaksi di kalangan masyarakat, terutama di media sosial seperti Twitter. Oleh karena itu, algoritma Decision Tree (DT) Naïve Bayes Classifier (NBC) dan diterapkan untuk mengklasifikasikan sentimen masyarakat terhadap kebijakan penghapusan kewajiban skripsi di Indonesia. Tujuan utama dari penelitian ini adalah untuk mengevaluasi kinerja algoritma DT dan NBC dalam mengklasifikasikan data teks ke dalam tiga kategori, yaitu positif, netral, dan negatif. Proses tuning hyperparameter akan dilakukan pada model DT dan NBC untuk memperoleh parameter terbaik, kemudian menggunakan akurasi dan rata-rata skor AUC-ROC sebagai indikator untuk menentukan model yang paling optimal. Dataset yang digunakan berupa ribuan tweet yang telah melalui tahap preprocessing dan pelabelan menggunakan pendekatan lexicon-based serta ekstraksi fitur menggunakan *TF-IDF*. Hasil penelitian menunjukkan bahwa sentimen negatif lebih dominan dibandingkan sentimen positif dan netral. Evaluasi performa model dilakukan menggunakan metrik akurasi, di mana hasil terbaik diperoleh pada model Naïve Bayes dengan akurasi 77.13% dan nilai rata-rata AUC-ROC sebesar 56%, sementara Decision Tree mencapai akurasi 74.21% dan nilai rata-rata AUC-ROC 53%. Hasil ini menunjukkan bahwa Naïve Bayes lebih unggul dalam klasifikasi teks berbasis opini masyarakat di media sosial dibandingkan dengan Decision Tree.

Kata Kunci: Analisis Sentimen, Decision Tree, Naïve Bayes Classifier, Twitter.

Halaman ini sengaja dikosongkan

ABSTRACT

Student Name / NPM : Ilham Ainur Idhana / 19081010099
Thesis Title : Perbandingan Algoritma Decision Tree dan Naïve Bayes Classifier pada Analisis Sentimen Twitter Mengenai Kebijakan Penghapusan Skripsi
Advisors : 1. Dr. Basuki Rahmat, S.Si. MT.
2. Eva Yulia Puspaningrum, S.Kom., M.Kom.

The policy of removing the thesis obligation as a graduation requirement for S1 / D4 students stipulated in Permendikbudristek Number 53 of 2023 has raised various reactions among the public, especially on social media such as Twitter. Therefore, the Decision Tree (DT) Naïve Bayes Classifier (NBC) algorithm is applied to classify public sentiment towards the policy of removing the thesis obligation in Indonesia. The main objective of this research is to evaluate the performance of DT and NBC algorithms in classifying text data into three categories, namely positive, neutral, and negative. The hyperparameter tuning process will be carried out on the DT and NBC models to obtain the best parameters, then use the accuracy and average AUC-ROC score as indicators to determine the most optimal model. The dataset used is thousands of tweets that have gone through preprocessing and labeling stages using a lexicon-based approach and feature extraction using TF-IDF. The results show that negative sentiment is more dominant than positive and neutral sentiment. Model performance evaluation was conducted using accuracy metrics, where the best results were obtained in the Naïve Bayes model with an accuracy of 77.13% and an average AUC-ROC value of 56%, while Decision Tree achieved an accuracy of 74.21% and an average AUC-ROC value of 53%. These results show that Naïve Bayes is superior in the classification of public opinion-based text on social media compared to Decision Tree.

Keywords: Sentiment Analysis, Decision Tree, Naïve Bayes Classifier, Twitter.

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Puji s yukur kehadiran Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi dengan judul **“Perbandingan Algoritma Decision Tree dan Naïve Bayes Classifier pada Analisis Sentimen Twitter Mengenai Kebijakan Penghapusan Skripsi”** dapat terselesaikan dengan baik.

Penulis mengucapkan terima kasih kepada kepada Bapak Dr. Basuki Rahmat, S.Si., MT. selaku Dosen Pembimbing I dan Ibu Eva Yulia Puspaningrum, S.Kom., M.Kom selaku Dosen Pembimbing II yang telah meluangkan waktunya untuk memberikan bimbingan, nasehat serta motivasi kepada penulis. Dan penulis juga banyak menerima bantuan dari berbagai pihak, baik itu berupa moril, spiritual maupun materiil. Untuk itu penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Ibu Dr. Novirina Hendrasarie, S.T, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Fetty Tri Anggraeny S.Kom., M.Kom. selaku Koordinator Program Studi Informatika Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur.
4. Bapak Andreas Nugroho Sihananto, S.Kom., M.Kom. selaku Koordinator Skripsi Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
5. Bapak Agung Mustika Rizki, S.Kom, M.Kom selaku dosen wali yang membantu dalam perwalian dari awal sampai akhir perkuliahan.
6. Bapak Dr. Basuki Rahmat, S.Si., MT. selaku dosen pembimbing 1 yang telah memberikan arahan selama proses penyelesaian skripsi.
7. Ibu Eva Yulia Puspaningrum, S.Kom., M.Kom. selaku dosen pembimbing 2 yang telah memberikan arahan selama proses penyelesaian skripsi.
8. Seluruh Dosen dan Staff Tata Usaha Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
9. Orang Tua, Adik, dan Keluarga yang telah memberikan dukungan secara materi dan non-materi serta memberikan doa sehingga penulis dapat menyelesaikan perkuliahan dari awal hingga akhir dengan baik.

10. ‘Grup Ambyar’ yang beranggotakan Abiyan Naufal Hilmi, Eriyansyah Yusuf Suwandana, Muhammad Arif Saifudin dan Thomas Andrew Imanzaghi yang telah membantu dan memberikan masukan serta ide-idenya sehingga skripsi ini dapat terselesaikan.
11. Teman-teman seperjuangan Informatika angkatan 2019, kakak dan adik tingkat, serta seluruh pihak yang telah membantu dan mendukung penulis dalam menyelesaikan skripsi dan perkuliahan ini.

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, 11 Maret 2025

Penulis



Ilham Amur Idhana
19081010099

DAFTAR ISI

LEMBAR JUDUL SKRIPSI.....	i
LEMBAR PENGESAHAN	iii
LEMBAR PERSETUJUAN	v
SURAT PERNYATAAN BEBAS PLAGIASI.....	vii
ABSTRAK	ix
ABSTRACT	xi
KATA PENGANTAR.....	xiii
DAFTAR ISI	xv
DAFTAR TABEL.....	xviii
DAFTAR GAMBAR.....	xiv
DAFTAR KODE PROGRAM	xxi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Tujuan.....	3
1.4. Manfaat.....	3
1.5. Batasan Masalah	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Penelitian Terdahulu.....	5
2.2 Data Mining.....	8
2.3 Analisis Sentimen	9
2.4 Twitter	10
2.5 Text Mining	10
2.6 Preprocessing.....	11

2.6.1	Casefolding	11
2.6.2	Tokenizing.....	11
2.6.3	Stopword	11
2.6.4	Stemming	11
2.7	Decision Tree.....	12
2.7.1	Kelebihan Decision Tree	13
2.7.2	Kekurangan Decision Tree.....	13
2.9	Ekstraksi Fitur.....	16
2.10	Visualisasi Data	18
2.11	Python.....	18
2.12	K-Fold Cross Validation.....	19
2.13	Confusion Matrix.....	20
2.14	Kurva ROC (Receiver Operation Characteristic)	21
BAB III METODE PENELITIAN		23
3.1	Studi Literatur.....	24
3.2	Identifikasi Masalah	24
3.3	Kebutuhan Data	24
3.4	Pengumpulan Data.....	24
3.5	Pengolahan Data	26
3.6	Perancangan Model Klasifikasi	30
3.7	Pengujian	32
BAB IV HASIL DAN PEMBAHASAN.....		37
4.1.	Pengumpulan Data.....	37
4.2	Preprocessing Data	42
4.2.1	Case Folding.....	42

4.2.2 Tokenizing	44
4.2.3. Spelling Normalization	45
4.2.4. Filtering	46
4.2.5. Stemming	47
4.3. Labeling	48
4.4. Visualisasi Dataset.....	50
4.5. Pembagian Data	52
4.6. Feature Extraction.....	53
4.7. Klasifikasi Data	54
4.7.1. Klasifikasi Decision Tree	54
4.7.2. Klasifikasi Naïve Bayes.....	57
4.8. Pengujian Model.....	61
4.8.1. Hasil Pengujian Tuning Hyperparameter <i>Decision Tree</i>	69
4.8.2. Hasil Pengujian Tuning Hyperparameter <i>NBC</i>	73
4.8.3 Hasil Pengujian K-Fold Cross Validation	75
4.9. Interpretasi Model Terbaik	76
BAB V PENUTUP	79
5.1 Kesimpulan.....	79
5.2 Saran	79
DAFTAR PUSTAKA	81

DAFTAR TABEL

Tabel 2.1	Struktur Document-term Matrix	17
Tabel 3.1	Struktur dataset sebelum dilakukan preprocessing	25
Tabel 3.2	Struktur dataset setelah tahap ekstraksi fitur	26
Tabel 3.3	Contoh dokumen untuk TF-IDF	29
Tabel 3.4	Contoh hasil TF-IDF	29
Tabel 3.5	Contoh hasil Ekstraksi Fitur	30
Tabel 3.6	Struktur tabel skenario pengujian Naïve Bayes Classifier.....	34
Tabel 3.7	Struktur tabel skenario pengujian Decision Tree.....	35
Tabel 4.1	Akumulasi Jumlah Data Tiap Label	50
Tabel 4.2	Hasil Uji Skenario Hyperparameter DT	71
Tabel 4.3	Hasil Uji Skenario Hyperparameter NBC	73

DAFTAR GAMBAR

Gambar 2.1	Data Processing Pipeline	8
Gambar 2.2	Representasi Confusion Matrix.....	21
Gambar 3.1	Langkah-langkah Penelitian.....	23
Gambar 3.2	Tahapan Pengolahan Data.....	27
Gambar 3.3	Diagram Alir Metode Klasifikasi.....	30
Gambar 3.4	Struktur Decision Tree	32
Gambar 3.5	Diagram Alir Skema Pengujian	33
Gambar 4.1	Dataframe output crawling data	42
Gambar 4.2	Dataframe casefolding	43
Gambar 4.3	Dataframe hasil tokenizing	43
Gambar 4.4	Dataframe hasil normalisasi teks tweet.....	45
Gambar 4.5	Dataframe hasil filtering tweet.....	47
Gambar 4.6	Dataframe hasil stemming tweet.....	48
Gambar 4.7	Dataframe hasil labeling	49
Gambar 4.8	Grafik label sentimen	50
Gambar 4.9	Wordcloud Frekuensi Fitur Dataset	51
Gambar 4.10	Dataframe Hasil Ekstraksi Fitur.....	54
Gambar 4.11	Struktur Pohon Keputusan ID3 Decision Tree.....	56
Gambar 4.12	Datatest Baru.....	67
Gambar 4.13	Dataframe Hasil Pengujian Hyperparameter DT	70
Gambar 4.14	Dataframe Hasil Pengujian Hyperparameter DT (2)	70
Gambar 4.15	Nilai Confusion Matrix DT	72
Gambar 4.16	Dataframe Hasil Pengujian Hyperparameter NBC	73
Gambar 4.17	Nilai Confusion Matrix NBC	74

Gambar 4.18	Hasil <i>k-fold cross validation</i> pada DT	75
Gambar 4.19	Hasil <i>k-fold cross validation</i> pada NBC.....	75
Gambar 4.20	AUC-ROC DT	77
Gambar 4.21	AUC-ROC NBC.....	77

DAFTAR KODE PROGRAM

Kode Program 4.1	Akses Token Twitter	37
Kode Program 4.2	Install Library yang digunakan.....	38
Kode Program 4.3	Crawling Data Tweet.....	38
Kode Program 4.4	Crawling Data Tweet (2).....	39
Kode Program 4.5	Crawling Data Tweet (3).....	40
Kode Program 4.6	Crawling Data Tweet (4).....	40
Kode Program 4.7	Baca File CSV & Menampilkan Dataframe.....	41
Kode Program 4.8	Fungsi Casefolding.....	43
Kode Program 4.9	Library untuk Tokenizing.....	44
Kode Program 4.10	Fungsi Tokenize data Tweet	44
Kode Program 4.11	Fungsi Spelling Normalization Data Tweet	45
Kode Program 4.12	Install Library untuk Filtering	46
Kode Program 4.13	Penghapusan Stopwords.....	47
Kode Program 4.14	Install Library untuk Stemming	47
Kode Program 4.15	Stemming Tweet	48
Kode Program 4.16	Proses Sebelum Labeling	49
Kode Program 4.17	Labeling Setiap Data Tweet	49
Kode Program 4.18	Visualisasi Distribusi Sentimen	50
Kode Program 4.19	Wordcloud Sentimen Fitur Dataset	51
Kode Program 4.20	Splitting Dataset	52
Kode Program 4.21	Membuat DTM dengan Fungsi Vectorizer	53
Kode Program 4.22	Membuat model prediksi Data Test	61
Kode Program 4.23	Inisialisasi Nilai Parameter Pengujian DT	62

Kode Program 4.24	Inisialisasi Nilai Parameter Pengujian NBC	62
Kode Program 4.25	K-Fold Cross Validation	67
Kode Program 4.26	Confusion Matrix	69
Kode Program 4.27	Classification Report	69