



**FILMFEELS: ANALISIS SENTIMEN REVIEW FILM
MENGUNAKAN ALGORITMA NAÏVE BAYES, DECISION
TREE, LOGISTIC REGRESSION, DAN ENSEMBLE
LEARNING**

LAPORAN TUGAS AKHIR

**UNIVERSITAS
MERCU BUANA**

DHAIFULLAH FADHIL HASAN

41520120040

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA 2025**



**FILMFEELS: ANALISIS SENTIMEN REVIEW FILM
MENGUNAKAN ALGORITMA NAÏVE BAYES,
DECISION TREE, LOGISTIC REGRESSION, DAN
ENSEMBLE LEARNING**

LAPORAN TUGAS AKHIR

DHAIFULLAH FADHIL HASAN
UNIVERSITAS
41520120040
MERCU BUANA

Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA 2025**

HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Dhaifullah Fadhil Hasan
NIM : 41520120040
Program Studi : Teknik Informatika
Judul Laporan Skripsi : FILMFEELS: ANALISIS SENTIMEN REVIEW FILM
MENGGUNAKAN ALGORITMA NAÏVE BAYES,
DECISION TREE, DAN LOGISTIC REGRESSION,
DAN ENSEMBLE LEARNING

Menyatakan bahwa Laporan Tugas Akhir ini adalah hasil karya saya sendiri dan bukan plagiat, serta semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Apabila ternyata ditemukan di dalam Laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

Jakarta, 15 Januari 2025

A handwritten signature in black ink is written over a yellow and red 10,000 Rupiah stamp. The stamp features the Garuda Pancasila emblem and the text 'REPUBLIK INDONESIA', '10000', and 'METERAI TEMPEL'. A unique serial number 'BDF43AJX697435620' is visible at the bottom of the stamp.

Dhaifullah Fadhil Hasan

HALAMAN PENGASAHAN

Laporan Skripsi diajukan oleh:

Nama : Dhaifullah Fadhil Hasan
NIM : 41520120040
Program Studi : Teknik Informatika
Judul Laporan Skripsi : FILMFEELS: ANALISIS SENTIMEN REVIEW FILM
MENGGUNAKAN ALGORITMA NAÏVE BAYES,
DECISION TREE, DAN LOGISTIC REGRESSION

Telah berhasil dipertahankan pada sidang di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Strata I pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Disahkan Oleh:

Pembimbing : Wawan Gunawan, S.Kom., MT
NIDN : 0424108104
Ketua Penguji : Dr. Afiyati, S.Si., M.T
NIDN : 0316106908
Penguji 1 : Prastika Indriyani, S.Kom, MCS
NIDN : 0312089401
Penguji 2 : Muhaimin Hasanudin, S.T, M.Kom
NIDN : 0420027508

Jakarta, Januari 2025
Mengetahui,

Dekan
Program Studi Teknik Informatika

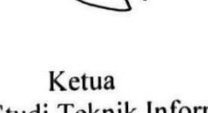


Dr. Bambang Jokonowo, S.Si., MTI
NIDN : 0320037002

Ketua
Program Studi Teknik Informatika



Dr. Hadi Santoso, S.Kom, M.Kom
NIDN : 0225067701

()
()
()
()

KATA PENGANTAR

Puji syukur Penulis panjatkan ke hadirat Allah swt. karena atas segala limpahan nikmat dan karunia-Nya, Penulis dapat menyelesaikan Laporan Tugas Akhir ini dengan lancar dan tepat waktu.

Penyelesaian Laporan Tugas Akhir ini tentunya tidak lepas dari bantuan dan dorongan berbagai pihak yang telah bersedia meluangkan waktu untuk memberikan semangat, saran dan masukan, serta bimbingan kepada Penulis. Oleh karena itu, Penulis ingin menyampaikan terima kasih yang tidak terhingga kepada:

1. Bapak Prof. Dr. Ir. Andi Adriansyah, M.Eng selaku Rektor Universitas Mercu Buana.
2. Bapak selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Dr. Hadi Santoso, S.Kom., M.Kom selaku Ketua Program Studi Informatika.
4. Bapak Wawan Gunawan S.Kom., M.T.I selaku Dosen Pembimbing yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan saya dalam penyusunan skripsi ini.
5. Ibu Dr. Afyati, S.Si., M.T selaku Ketua Tim Penguji Tugas Akhir atas koreksi dan arahan serta masukannya.
6. Ibu Prastika Indriyani, S.Kom, MCS selaku Dosen Penguji 1 Tugas Akhir atas koreksi dan arahan serta masukannya.
7. Orang tua Penulis, Alm. Bapak M. Danih dan Ibu Eva Latifah.
8. Rekan-rekan seperjuangan dan pihak-pihak lain yang tidak dapat Penulis sebutkan satu-persatu.

Akhir kata, Penulis menyadari bahwa Laporan Tugas Akhir ini masih jauh dari kata sempurna. Oleh karena itu, Penulis mengharapkan kritik dan saran yang membangun dari pembaca untuk menjadikan Proposal Penelitian ini lebih baik.

Jakarta, 15 Januari 2025


Dhaifulah Fadhil Hasan

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKEDEMIS

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama	: Dhaifullah Fadhil Hasan
NIM	: 41520120040
Program Studi	: Teknik Informatika
Judul Laporan Skripsi	: FILMFEELS: ANALISIS SENTIMEN REVIEW FILM MENGGUNAKAN ALGORITMA NAÏVE BAYES, DECISION TREE, LOGISTIC REGRESSION, DAN ENSEMBLE LEARNING

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana Hak Bebas Royalti Non-Eksklusif (Non-exclusive Royalty-Free Right) atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Laporan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 15 Januari 2025

A handwritten signature in black ink is written over a red and yellow 10000 Rupiah stamp. The stamp includes the text '10000', 'METERAI TEMPEL', and a serial number '9C2AAJX697415644'.

Dhaifullah Fadhil Hasan

ABSTRAK

Nama : Dhaifullah Fadhil Hasan
NIM : 41520120040
Program Studi : S1 Teknik Informatika
Judul Laporan Skripsi : FILMFEELS: ANALISIS SENTIMEN REVIEW
FILM MENGGUNAKAN ALGORITMA
NAÏVE BAYES, DECISION TREE, LOGISTIC
REGRESSION, AND ENSEMBLE LEARNING
Pembimbing : Wawan Gunawan S.Kom., M.T.I

Penelitian ini mengembangkan sistem analisis sentimen review film menggunakan Ensemble Learning dengan algoritma Naïve Bayes, Decision Tree, dan Logistic Regression. Metode pengembangan sistem menggunakan pendekatan incremental dengan dua tahap utama, yaitu preprocessing untuk persiapan data dan klasifikasi sentimen pada 397.146 data review film. Hasil evaluasi menunjukkan bahwa model AdaBoost dengan Decision Tree sebagai estimator dasar memperoleh akurasinya sebesar 76.8%, dengan performa sangat baik pada kelas Positif (recall 96% dan F1-score 0.86). Namun, model ini kesulitan dalam mengidentifikasi kelas Negatif, dengan precision 0.67 dan recall 0.22. Selanjutnya, model Voting Classifier yang menggabungkan DecisionTreeClassifier, LogisticRegression, dan MultinomialNB dengan pendekatan soft voting dan pembobotan menghasilkan akurasinya 82.23%, dengan performa terbaik pada kelas Positif (presisi 0.83, recall 0.96, F1-score 0.89), namun kinerjanya kurang optimal pada kelas Negatif (presisi 0.79, recall 0.44, F1-score 0.56).

Kata Kunci: *Sentimen Review Film, Ensemble Learning, Naïve Bayes, Decision Tree, Logistic Regression*

ABSTRACT

Name : Dhaifullah Fadhil Hasan
NIM : 41520120040
Study Program : S1 Teknik Informatika
Title Thesis Report : FILMFEELS: ANALISIS SENTIMEN REVIEW
FILM MENGGUNAKAN ALGORITMA
NAÏVE BAYES, DECISION TREE, LOGISTIC
REGRESSION
Advisor : Wawan Gunawan S.Kom., M.T.I

This study develops a sentiment analysis system for movie reviews using Ensemble Learning with Naïve Bayes, Decision Tree, and Logistic Regression algorithms. The system development method employs an incremental approach with two main stages: preprocessing for data preparation and sentiment classification of movie reviews. Evaluation results show that the AdaBoost model with Decision Tree as the base estimator achieved an accuracy of 76.8%, with excellent performance on the Positive class (recall 96% and F1-score 0.86). However, the model struggled to identify the Negative class, with precision 0.67 and recall 0.22. Furthermore, the Voting Classifier model, combining DecisionTreeClassifier, LogisticRegression, and MultinomialNB with a soft voting approach and weighting, achieved an accuracy of 82.23%, with the best performance on the Positive class (precision 0.83, recall 0.96, F1-score 0.89). However, its performance on the Negative class was less optimal (precision 0.79, recall 0.44, F1-score 0.56). The system was tested on a movie review dataset consisting of 397,146 entries and successfully provided ease of use, prediction accuracy, fast access, and improved information visualization. The parameters analyzed include accuracy, precision, recall, and F1-score for each algorithm.

Keywords: Sentimen Review Film, Ensemble Learning, Naïve Bayes, Decision Tree, Logistic Regression

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERNYATAAN KARYA SENDIRI	ii
HALAMAN PENGSAHAN	iii
KATA PENGANTAR.....	iv
HALAMAN PERNYATAAN	v
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI.....	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xii
BAB 1 PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian	3
1.5 Batasan Penelitian.....	3
1.1 Sistematika Penelitian.....	4
BAB 2 TINJAUAN PUSTAKA	5
2.1 Penelitian Terdahulu	5
2.2 Teori Utama	6
2.3 Teori Pendukung	8
2.2.1 Natural Language Processing (NLP)	8
2.2.2 Teknik Pra-pemrosesan Teks.....	9
2.2.3 Algoritma Klasifikasi.....	9
2.2.4 Ensemble Learning	12
2.2.5 Metode Evaluasi Kinerja Model	12
2.4 Metode CRISP DM.....	13
BAB 3 METODE PENELITIAN	15
3.1 Pendekatan Penelitian	15

3.2	Desain Penelitian	15
3.3	Subjek Penelitian	15
3.4	Instrumen Penelitian	15
3.5	Teknik Pengumpulan Data	18
3.6	Analisis Data.....	19
3.7	Prosedur Penelitian	20
3.8	Evaluasi Hasil Penelitian	22
BAB 4 PEMBAHASAN		23
4.1	Business Understanding.....	23
4.2	Data Understanding	23
4.3	Data Preparation	29
4.5	Implementasi Sistem.....	42
4.6	Evaluasi	61
BAB 5 KESIMPULAN DAN SARAN.....		64
5.1	Kesimpulan.....	64
5.2	Saran.....	64
DAFTAR PUSTAKA		66
LAMPIRAN		69



DAFTAR TABEL

Tabel 1Penghapusan Stopword	32
Tabel 2 Pembersihan Kata	33
Tabel 3 Tokenisasi dan Stemming	34
Tabel 4 Lematisasi	35
Tabel 5 Spell Checker	36



DAFTAR GAMBAR

Gambar 1 Informasi Dataset Awal	27
Gambar 2 Tampilan Dataset	27
Gambar 3 Informasi Data Sebelum Penanganan Nilai Kosong.....	29
Gambar 4 Informasi Data Setelah Penanganan Dataset	30
Gambar 5 Informasi Data Setelah Penanganan Duplikasi Data	30
Gambar 6 Informasi Data Setelah Penanganan Outlier	37
Gambar 7 Text Clean Bernilai Null Setelah Outlier	38
Gambar 8 Ulasan Film di Rotten Tomatoes dari Tahun ke Tahun.....	40
Gambar 9 Distribusi Sentimen pada Dataset.....	41
Gambar 10 Gambar Distribusi Sampel Sentimen.....	43
Gambar 11. Penerapan ADASYN	45
Gambar 12. Penerapan Random Oversampling.....	46
Gambar 13. Penerapan Tomek Link	46
Gambar 14 Evaluasi Kinerja Algoritma Naive Bayes	48
Gambar 15 Evaluasi Kinerja Naive Bayes dengan SMOTE	50
Gambar 16 Evaluasi Kinerja Logistic Regression.....	52
Gambar 17 Evaluasi Kinerja Algoritma Decision Tree.....	54

DAFTAR LAMPIRAN

Lampiran 1. Lembar Persetujuan	69
Lampiran 2. Kartu Bimbingan	70
Lampiran 3. Curriculum Vitae	71
Lampiran 4. Surat Pernyataan Haki	72
Lampiran 5. Form Revisi Penguji.	73
Lampiran 6. Sertifikat BNSP.	75
Lampiran 7. Turnitin	76

