



**ANALISIS SENTIMEN ULASAN APLIKASI LETTERBOXD
DI GOOGLE PLAY MENGGUNAKAN METODE
NAÏVE BAYES DAN INFORMATION GAIN**

SKRIPSI

**FERNALDI ANGGADHA
1810511069**

**PROGRAM STUDI S1 INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA
JAKARTA
2025**



**ANALISIS SENTIMEN ULASAN APLIKASI LETTERBOXD
DI GOOGLE PLAY MENGGUNAKAN METODE
NAÏVE BAYES DAN INFORMATION GAIN**

SKRIPSI

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

**FERNALDI ANGGADHA
1810511069**

**PROGRAM STUDI S1 INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA
JAKARTA
2025**

LEMBAR ORISINALITAS

PERNYATAAN ORISINALITAS

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

Nama : Fernaldi Anggadha

NIM : 1810511069

Tanggal : 20 Juni 2025

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

Jakarta, 20 Juni 2025

Yang menyatakan,

A blue and red rectangular stamp is visible, containing the text "METRAL" and "FERNALDI" along with a date "20 JUNI 2025". A handwritten signature in black ink is written over the stamp.

Fernaldi Anggadha

LEMBAR PUBLIKASI

PERNYATAAN PERSETUJUAN PUBLIKASI SKIRPSI UNTUK KEPERLUAN AKADEMIS

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

Nama : Fernaldi Anggadaha

NIM : 1810511069

Tanggal : 20 Juni 2025

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

ANALISIS SENTIMEN ULASAN APLIKASI LETTERBOXD DI GOOGLE PLAY MENGGUNAKAN METODE NAÏVE BAYES DAN INFORMATION GAIN

Beserta perangkat yang ada (jika diperlukan), dengan Hak Bebas Royalti di
Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih
media/formalitas, mengelola dalam bentuk pengkalan data (*database*), merawat dan
mempublikasi Skripsi saya selama tetap mencantumkan nama saya sebagai
penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian Pernyataan ini saya buat dengan sebenarnya.

Jakarta 20 Juni 2025,

Yang Menyatakan,



Fernaldi Anggadaha

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

Judul : Analisis Sentimen Ulasan Aplikasi Letterboxd Di Google Play
Menggunakan Metode Naive Bayes Dan Information Gain.
Nama : Fernaldi Anggadha
NIM : 1810511069
Program Studi : S1 Informatika

Disetujui oleh :

Penguji 1:

Jayanta, S.Kom., M.Si

Penguji 2:

Nurul Afifah Arifuddin, S.Pd., M.T.

Pembimbing 1:

Dr. Widya Cholil, M.I.T

Pembimbing 2:

Diketahui oleh:

Koordinator Program Studi:

Dr. Widya Cholil, M.I.T

NIP 221112080

Dekan Fakultas Ilmu Komputer

Prof. Dr. Ir. Supriyanto, S.T., M.Sc.

NIP 197605082003121002

Tanggal Ujian Skripsi:

4 Juli 2025

ANALISIS SENTIMEN ULASAN APLIKASI LETTERBOXD DI GOOGLE PLAY MENGGUNAKAN METODE NAÏVE BAYES DAN INFORMATION GAIN

Fernaldi Anggadha

Abstrak

Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi Letterboxd pada platform Google Play Store dengan menggunakan metode klasifikasi Naïve Bayes serta seleksi fitur Information Gain. Dalam era digital, ulasan pengguna menjadi indikator penting untuk menilai kualitas dan kinerja suatu aplikasi. Naïve Bayes digunakan karena efisien dalam mengklasifikasikan teks, sedangkan Information Gain dimanfaatkan untuk menyaring fitur-fitur yang paling relevan guna meningkatkan performa model. Penelitian ini menerapkan pendekatan kuantitatif dengan data ulasan yang dikumpulkan melalui API google-play-scraper dari Januari hingga Mei 2025. Proses praproses data mencakup case folding, pembersihan teks, tokenisasi, penghilangan stopword, dan lemmatization. Data kemudian dibagi dengan rasio 80:20 untuk pelatihan dan pengujian. Hasil evaluasi menunjukkan bahwa model Naïve Bayes tanpa seleksi IG (*Information Gain*) mampu mencapai akurasi sebesar 82,5% dengan nilai precision dan recall masing-masing sebesar 83% dan 84% serta model Naïve Bayes dengan seleksi IG mampu mencapai akurasi 83,5% dengan nilai precision dan recall masing-masing sebesar 84% dan 85%. Selain itu, dilakukan analisis lanjutan dengan memetakan hasil sentimen negatif ke dalam tujuh karakteristik kualitas perangkat lunak berdasarkan standar ISO/IEC 25010 menggunakan pendekatan rule-based. Temuan ini diharapkan dapat menjadi dasar bagi pengembang Letterboxd dalam meningkatkan aspek kualitas aplikasi berdasarkan persepsi dan pengalaman pengguna.

Kata kunci: analisis sentimen, Naïve Bayes, Information Gain, ISO 25010, Google Play, Letterboxd.

SENTIMENT ANALYSIS OF LETTERBOXD APPLICATION REVIEWS ON GOOGLE PLAY USING NAÏVE BAYES AND INFORMATION GAIN

Fernaldi Anggadha

Abstract

This research aims to analyze user sentiment from reviews of the Letterboxd application on the Google Play Store using the Naïve Bayes classification method and Information Gain feature selection. In the digital era, user reviews serve as a critical indicator for evaluating application quality and performance. Naïve Bayes was chosen for its efficiency in text classification, while Information Gain was applied to filter the most relevant features to enhance model performance. A quantitative approach was employed, with review data collected via the google-play-scraper API between January and May 2025. Text preprocessing steps included case folding, cleaning, tokenization, stopword removal, and lemmatization. The data was then split using an 80:20 ratio for training and testing. Evaluation results showed that the Naïve Bayes model achieved 82.5% accuracy with precision and recall scores of 83% and 84%, respectively, meanwhile, combined with Information Gain, Naïve Bayes achieved an accuracy of 83.5%, with a precision of 84% and a recall of 85%.. Furthermore, a follow-up analysis was conducted by mapping negative sentiment results into seven software quality characteristics based on the ISO/IEC 25010 standard using a rule-based classification approach. These findings are expected to assist Letterboxd developers in enhancing application quality based on user sentiment and experience.

Keywords: *sentiment analysis, Naïve Bayes, Information Gain, ISO 25010, Google Play, Letterboxd.*

KATA PENGANTAR

Puji dan Syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas segala rahmat, karunia, dan kuasa-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “Analisis Sentimen Ulasan Pengguna Aplikasi Letterboxd dalam Aplikasi Google Play Store Menggunakan Metode Naïve Bayes dan Seleksi Fitur Information Gain“. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Ilmu Komputer pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.

Penulis menyadari skripsi ini tidak lepas dari kekurangan sehingga berbagai masukan dan saran dapat menjadi sesuatu yang berharga untuk penyelesaian karya ini. Penulis juga menyadari bahwa tersusunnya skripsi ini tidak lepas dari bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih dan penghargaan yang sebesar- besarnya kepada:

1. Ayah dan Ibu, Sanggam Paindoan Siregar dan Rety Yuliana, atas dukungannya yang senantiasa menemani lewat perjuangan, doa, dan kasih sayang yang diberikan kepada saya tiada henti.
2. Ibu Dr. Widya Cholil, M.I.T, selaku Kepala Program Studi Informatika Universitas Pembangunan Nasional Veteran Jakarta dan selaku dosen pembimbing yang telah memberikan arahan, motivasi, dan masukan kepada saya selama bimbingan sehingga dapat menyelesaikan skripsi ini.
3. Prof. Dr. Ir. Supriyanto, ST, M.Sc., selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.
4. Adik saya, Natasya Miranda, yang senantiasa memberikan saya dukungan baik secara motivasi, saran, dan doa selama penyusunan skripsi ini.
5. Seluruh dosen dan staf di lingkungan Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta, yang telah memberikan ilmu dan fasilitas yaselama masa studi.
6. Rekan seangkatan saya yang telah memberikan dukungan moral dan material selama proses pengerjaan skripsi ini.

Akhir kata, penulis berharap semoga skripsi ini dapat bermanfaat untuk semua pihak, khususnya bagi pengembangan Ilmu Pengetahuan di bidang Ilmu Komputer.

Jakarta, 20 Juni 2025



Fernaldi Anggadha

DAFTAR ISI

LEMBAR ORISINALITAS	iii
LEMBAR PUBLIKASI	iv
LEMBAR PENGESAHAN	v
Abstrak	vi
<i>Abstract</i>	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR RUMUS	xiv
DAFTAR LAMPIRAN.....	xv
BAB 1. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2. Rumusan Masalah	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian	3
1.5. Manfaat Penelitian	3
1.6 Luaran yang Diharapkan	4
1.7. Sistematika Penulisan	4
BAB 2. LANDASAN TEORI.....	6
2.1. Film Review	6
2.2. Letterboxd	6
2.3. Google Play Store	7
2.4. Analisis Sentimen	7
2.5. <i>Machine Learning</i>	8
2.6. Text Mining.....	8
2.7. Praproses Data.....	9
2.7.1 <i>Case Folding</i>	9
2.7.2 <i>Cleaning</i>	9
2.7.3 <i>Normalization</i>	9
2.7.4 <i>Tokenizing</i>	9
2.7.5 <i>Stopword Removal</i>	10

2.7.6	<i>Lemmatization</i>	10
2.8.	TF-IDF	10
2.9.	Information Gain.....	11
2.7.	Naïve Bayes	12
2.8	Evaluasi	15
2.9.	Penelitian Terdahulu	16
BAB 3. METODOLOGI PENELITIAN.....		19
3.1.	Alur Penelitian	19
3.1.1.	Menentukan Topik	20
3.1.2.	Identifikasi Masalah	20
3.1.3.	Perumusan Masalah	20
3.1.4.	Studi Literatur	20
3.1.5.	Pengumpulan Data	21
3.1.6.	Pelabelan	21
3.1.7.	Praproses Data.....	22
3.1.7.1	<i>Case Folding</i>	22
3.1.7.2	<i>Normalization</i>	22
3.1.7.3	<i>Cleaning</i>	22
3.1.7.4	<i>Tokenizing</i>	23
3.1.7.5	<i>Stopword Removal</i>	23
3.1.7.6	<i>Lemmatization</i>	23
3.1.8.	Pembagian Data	24
3.1.9.	Pembobotan Kata dengan TF-IDF	24
3.1.10.	Information Gain.....	25
3.1.11.	Klasifikasi	26
3.1.12.	Evaluasi.....	26
3.2.	Alat Pendukung Penelitian.....	26
3.3.	Jadwal Penelitian.....	27
BAB 4. HASIL DAN PEMBAHASAN.....		29
4.1.	Pengumpulan Data	29
4.2.	Pelabelan	29
4.3.	Praproses Data.....	30
4.3.1.	Case Folding	30
4.3.2.	Normalization.....	32

4.3.3.	Cleaning	33
4.3.4.	Tokenizing.....	34
4.3.5.	Stopword Removal.....	35
4.3.6.	Lemmatization.....	36
4.4.	Pembagian Data	37
4.5.	TF-IDF	37
4.6.	Information Gain.....	40
4.6.1.	Perhitungan nilai Entropy Total	42
4.6.2.	Perhitungan nilai Entropy Fitur.....	42
4.6.3.	Perhitungan nilai Information Gain	43
4.6.4.	Pengambilan Fitur	45
4.7.	Klasifikasi	46
4.8.	Evaluasi	54
4.9.	Hasil dan Analisa Sentimen Ulasan Pengguna	56
4.9.1.	Ulasan Sentimen Positif.....	56
4.9.2.	Ulasan Sentimen Negatif.....	57
4.9.3.	Functional Suitability	58
4.9.4.	Performance Efficinecy.....	58
4.9.5.	Usability	59
4.9.6.	Reliability.....	60
4.9.7.	Security	61
4.9.8.	Maintainability	62
4.9.9.	Portability.....	63
BAB 5.	PENUTUP	64
5.1.	Kesimpulan	64
5.2.	Keterbatasan Penelitian.....	64
5.3.	Saran.....	65
DAFTAR PUSTAKA		66
DAFTAR RIWAYAT HIDUP		69
LAMPIRAN		70

DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian.....	19
Gambar 4. 1 Letterboxd Reviews English	29
Gambar 4. 2 Pelabelan Data.....	30
Gambar 4. 3 Fitur Frekuensi Sentimen Positif Terbanyak.....	57
Gambar 4. 4 WordCloud Sentimen Positif	57
Gambar 4. 5 Fitur Frekuensi Sentimen Negatif Terbanyak	57
Gambar 4. 6 WordCloud Sentimen Negatif.....	57
Gambar 4. 7 Fitur Frekuensi Functional Suitability.....	58
Gambar 4. 8 WordCloud Functional Suitability	58
Gambar 4. 9 Fitur Frekuensi Performance Efficiency	59
Gambar 4. 10 WordCloud Performance Efficiency	59
Gambar 4. 11 Fitur Frekuensi Usability.....	59
Gambar 4. 12 WordCloud Usability	59
Gambar 4. 13 Fitur Frekuensi Reliability	60
Gambar 4. 14 WordCloud Reliability	60
Gambar 4. 15 Fitur Frekuensi Security	61
Gambar 4. 16 WordCloud Security.....	61
Gambar 4. 17 Fitur Frekuensi Maintainability.....	62
Gambar 4. 18 WordCloud Maintainability	62
Gambar 4. 19 Fitur Frekuensi Portability	63
Gambar 4. 20 WordCloud Portability	63

DAFTAR TABEL

Tabel 2. 1 Confusion Matrix	15
Tabel 2. 2 Penelitian Terkait	17
Tabel 3. 1 Jadwal Penelitian.....	27
Tabel 4. 1 Case Folding	31
Tabel 4. 2 Case Folding	31
Tabel 4. 3 Normalization.....	32
Tabel 4. 4 Cleaning	33
Tabel 4. 5 Tokenizing	34
Tabel 4. 6 Stopword Removal.....	35
Tabel 4. 7 Lemmatization	36
Tabel 4. 8 Pembagian Data	37
Tabel 4. 9 Data Sampel	38
Tabel 4. 10 Pembobotan TF-IDF	38
Tabel 4. 11 Fitur Bobot TF-IDF.....	40
Tabel 4. 12 Fitur Information Gain.....	43
Tabel 4. 13 Urutan Skor Fitur IG	45
Tabel 4. 14 Data Input Naïve Bayes (TF-IDF)	46
Tabel 4. 15 Nilai Probabilitas Kelas	51
Tabel 4. 16 Sampel Data Uji.....	52
Tabel 4. 17 Sampel Data Uji Preproseccing	52
Tabel 4. 18 Bobot TF-IDF Data Uji.....	52
Tabel 4. 19 Confusion Matrix	54
Tabel 4. 20 Perbandingan Naïve Bayes dan Naïve Bayes + Information Gain	55

DAFTAR RUMUS

Rumus (1) TF-IDF	10
Rumus (2) IDF	11
Rumus_ (3) Perkalian TF dan IDF	11
Rumus_ (4) Entropy Total	12
Rumus (5) Entropy Atribut	12
Rumus (6) Information Gain	12
Rumus (7) Umum Teorema Bayes	13
Rumus (8) Teorema Bayes	14
Rumus (9) Likelihood	14
Rumus (10) Multinomial Naïve Bayes	15
Rumus (11) Skor Accuracy	16
Rumus (12) Skor Precision	16
Rumus (13) Skor Recall	16
Rumus (14) Prior Positif	48
Rumus (15) Prior Negatif	48
Rumus (16) Turunan Multinomial Naive Bayes	53
Rumus (17) Multinomial Naive Bayes log	53
Rumus (18) F-1 Score	55

DAFTAR LAMPIRAN

Lampiran 1 Hasil Cek Plagiarisme	70
--	----