

**KLASIFIKASI OPINI PENONTON FILM PADA
PLATFORM STREAMING VIDEO DENGAN
ALGORITMA NAÏVE BAYES**

SKRIPSI



Oleh

Alita Tiffana

211510017

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2025**

**KLASIFIKASI OPINI PENONTON FILM PADA
PLATFORM STREAMING VIDEO DENGAN
ALGORITMA NAÏVE BAYES**

SKRIPSI

**Untuk memenuhi salah satu syarat
memperoleh gelar Sarjana**



Oleh

Alita Tiffana

211510017

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2025**

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini saya:

Nama : Alita Tiffana

NPM : 211510017

Fakultas : Teknik dan Komputer

Program Studi : Sistem Informasi

Menyatakan bahwa “**Skripsi**” yang saya buat dengan judul:

KLASIFIKASI OPINI PENONTON FILM PADA PLATFORM STREAMING VIDEO DENGAN ALGORITMA NAÏVE BAYES

Adalah hasil karya sendiri dan bukan “duplikasi” dari karya orang lain. Sepengetahuan saya, didalam naskah Skripsi. ini tidak terdapat karya ilmiah atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip didalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata di dalam naskah Skripsi. ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia naskah Skripsi ini digugurkan dan gelar akademik yang saya peroleh dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya tanpa ada paksaan dari siapapun

Batam, 19 Juli 2025



Alita Tiffana

211510017

**KLASIFIKASI OPINI PENONTON FILM PADA
PLATFORM STREAMING VIDEO DENGAN
ALGORITMA NAÏVE BAYES**

SKRIPSI

**Untuk memenuhi salah satu syarat
memperoleh gelar Sarjana**

Oleh

Alita Tiffana

211510017

**Telah disetujui oleh Pembimbing pada tanggal
seperti tertera di bawah ini**

Batam, 18 Juli 2025

A handwritten signature in blue ink, appearing to read 'Erlin Elisa', with a horizontal line underneath.

Erlin Elisa, S.Kom., M.Kom.

Pembimbing

ABSTRAK

Platform streaming video telah berkembang menjadi komponen penting dalam masyarakat kontemporer karena menawarkan akses hiburan yang fleksibel. Masyarakat umum masih memiliki opini yang beragam mengenai *platform* ini, meskipun ada beberapa yang lebih suportif daripada yang lain. Tiga kategori sentiment positive, negative, dan netral akan digunakan dalam studi ini untuk mengklasifikasikan sentimen penonton terhadap layanan *streaming video*, khususnya *Netflix* dan *Disney+*. Menggunakan Algoritma Naïve Bayes, dan data dari komentar pengguna di *Google Play Store* diperoleh menggunakan *web scraping*. Prapemrosesan, pelabelan data, klasifikasi, serta penilaian model yakni metrik akurasi, presisi, *recall*, skor-F1 adalah semua langkah dalam proses analisis. Berdasarkan hasil penelitian, algoritma *Gaussian Naïve Bayes* mencapai akurasi 43,52% untuk *Disney+* dan 41,99% untuk *Netflix*. Studi ini menunjukkan kemungkinan awal dalam penerapan analisis opini publik otomatisasi, meskipun tingkat akurasinya saat ini tergolong rendah. Secara teknis, *Disney+* berkinerja lebih baik dalam klasifikasi, sedangkan evaluasi pengguna berdasarkan rating menghasilkan ulasan positif yang lebih baik di *Netflix*. Diharapkan bahwa studi ini bisa bermanfaat sebagai landasan untuk pengembang sistem klasifikasi opini yang lebih tepat pada masa yang akan datang.

Kata Kunci: Analisis Sentimen, *Data Mining*, *Naïve Bayes*, *Streaming Video*, *Netflix*, *Disney+*,

ABSTRACT

Video streaming platforms have become an essential component of contemporary society since they offer flexible access to entertainment. However, opinions about these programs are still divided, with some members of the public being more supportive than others. This study aims to categorize viewer attitudes toward video streaming services, specifically Netflix and Disney+, into three groups: good, negative, and neutral. The strategy used is the Naïve Bayes Algorithm, and web scraping techniques are used to collect user comment data from the Google Play Store. Preprocessing, data labeling, classification, and model evaluation using metrics like accuracy, precision, recall, and F1 score are all part of the analytical process. Because they provide flexible access to entertainment, video streaming platforms have become an indispensable part of modern society. There is still disagreement among the public on these programs, though, with some people being more in favor of them than others. The purpose of this study is to divide viewer opinions on Netflix and Disney+ in particular into three categories: neutral, negative, and positive. The Naïve Bayes Algorithm is the approach employed, and data from user comments on the Google Play Store is gathered using web scraping techniques. The analytical process includes preprocessing, data labeling, classification, and model evaluation utilizing metrics such as accuracy, precision, recall, and F1 score.

Keywords: Sentiment Analysis, Data Mining, Naïve Bayes, Video Streaming, Netflix, Disney+

KATA PENGANTAR

Segala puji dan syukur kepada Allah SWT., yang telah melimpahkan segala rahmat dan karuniaNya, sehingga penulis dapat menyelesaikan laporan tugas akhir yang merupakan salah satu persyaratan untuk menyelesaikan program studi strata satu (S1) pada Program Studi Sistem Informasi. Universitas Putera Batam.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Karena itu, kritik dan saran akan senantiasa penulis terima dengan senang hati. Dengan segala keterbatasan, penulis menyadari pula bahwa skripsi ini takkan terwujud tanpa bantuan, bimbingan, dan dorongan dari berbagai pihak. Untuk itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih kepada:

1. Rektor Universitas Putera Batam, Ibu Dr. Nur Elfi Husda, S.Kom., M.SI.
2. Dekan Fakultas, Bapak Welly Sugianto, S.T., M.M.
3. Ketua Program Studi, Bapak Muhammad Rasid Ridho, S.Kom., M.SI.
4. Ibu Erlin Elisa, S.Kom., M.Kom., selaku pembimbing Skripsi pada Program Studi Sistem Informasi Universitas Putera Batam
5. Dosen dan Staff Universitas Putera Batam
6. Orang Tua, Keluarga dan teman-teman penulis
7. Rekan-rekan mahasiswa di Program Studi Sistem Informasi Universitas Putera Batam

Semoga Allah SWT. membalas kebaikan dan selalu mencurahkan hidayah serta taufik-Nya, Amin.

Batam, 19 Juli 2025



Alita Tiffana

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL	ii
SURAT PERNYATAAN	iii
HALAMAN PENGESAHAN	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Identifikasi Masalah	5
1.3. Batasan Masalah	6
1.4. Rumusan Masalah	6
1.5. Tujuan Penelitian	7
1.6. Manfaat Penelitian	7
1.6.1. Manfaat Teoritis	7
1.6.2. Manfaat Praktis	8
BAB II TINJAUAN PUSTAKA	9
2.1. Teori Dasar	9
2.1.1. Platform Streaming	9
2.1.2. Analisis Sentimen	10
2.1.3. Knowledge Discovery In Database	10
2.1.4. Data mining	12
2.1.5. Klasifikasi	13
2.1.6. Algoritma Naïve Bayes	13
2.1.7. Google Colab	16
2.1.8. Python	17
2.2. Penelitian Terdahulu	18
2.3. Kerangka Pemikiran	23
2.4. Hipotesis	24
BAB III METODE PENELITIAN	25
3.1. Desain Penelitian	25
3.2. Variabel	28
3.3. Populasi Dan Sampel	29
3.3.1. Populasi	29
3.3.2. Sampel	29
3.4. Teknik Pengumpulan Data	29
3.5. Metode Analisis Data	31
3.6. Lokasi Dan Jadwal Penelitian	31
3.6.1. Lokasi	31

3.6.2 Jadwal Penelitian	32
BAB IV HASIL DAN PEMBAHASAN	33
4.1. Hasil	33
4.1.1. Scraping Data	33
4.1.2. Preprocessing Data	34
4.1.3. Pelabelan Data	50
4.1.4. Data Split	53
4.1.5. Naïve Bayes	54
4.1.6. Rating	58
4.2. Pembahasan	60
BAB V SIMPULAN DAN SARAN	63
5.1. Simpulan	63
5.2. Saran	64
DAFTAR PUSTAKA	65
LAMPIRAN	
Lampiran 1. Pendukung Penelitian	
Lampiran 2. Daftar Riwayat Hidup	
Lampiran 3. Surat Keterangan Penelitian	
Lampiran 4. Jurnal Artikel	
Lampiran 5. Turnitin Skripsi	

DAFTAR GAMBAR

Gambar 1.1 Lanskap Media Digital 2024	2
Gambar 2.1 KDD.....	11
Gambar 2.2 Kerangka Pemikiran	24
Gambar 3.1 Desain Penelitian	25
Gambar 3.2 Metode Analisis Data	31
Gambar 4.1 Scraping Data Netflix	34
Gambar 4.2 Scraping Data Disney+	34
Gambar 4.3 Proses Hapus Data Duplikat Netflix.....	35
Gambar 4.4 Proses Hapus Data Duplikat Disney+	35
Gambar 4.5 Stopwords Sebelum Preprocessing Netflix	36
Gambar 4.6 Frekuensi kata Sebelum Preprocessing Netflix	37
Gambar 4.7 Stopwords Sebelum Preprocessing Disney+	38
Gambar 4.8 Frekuensi kata Sebelum Preprocessing Disney+.....	38
Gambar 4.9 Cleaning Netflix	39
Gambar 4.10 Cleaning Disney+	39
Gambar 4.11 Case Folding Netflix.....	40
Gambar 4.12 Case Folding Disney+	40
Gambar 4.13 Normalisasi Netflix.....	41
Gambar 4.14 Normalisasi Disney+	42
Gambar 4.15 Tokenize Netflix	42
Gambar 4.16 Tokenize Disney+	42
Gambar 4.17 Stopword Removal Netflix	43
Gambar 4.18 Stopword Removal Disney+.....	44
Gambar 4.19 Steming Netflix	45
Gambar 4.20 Steming Disney+	46
Gambar 4.21 NaN Netflix	46
Gambar 4.22 NaN Disney+	47
Gambar 4.23 Stopword Setelah Preprocessing Netflix	48
Gambar 4.24 Frekuensi Kata Setelah Preprocessing Netflix	48
Gambar 4.25 Stopword Setelah Preprocessing Disney+.....	49
Gambar 4.26 Frekuensi Kata Setelah Preprocessing Disney+	49
Gambar 4.27 Pelabelan Data Netflix.....	50
Gambar 4.28 Pelabelan Data Netflix.....	51
Gambar 4.29 Pelabelan Data Disney+.....	52
Gambar 4.30 Pelabelan Data Disney+.....	52
Gambar 4.31 Data Split Netflix	53
Gambar 4.32 Data Split Disney+.....	54
Gambar 4.33 GaussianNB Netflix.....	55
Gambar 4.34 Confusion Matrix GaussianNB Netflix	55
Gambar 4.35 GaussianNB Netflix.....	56
Gambar 4.36 GaussianNB Disney+	57
Gambar 4.37 Confusion Matrix GaussianNB Disney+.....	57
Gambar 4.38 GaussianNB Disney+	58

Gambar 4.39 Rating Netflix	59
Gambar 4.40 Rating Disney+	60

DAFTAR TABEL

Tabel 3.1 Jadwal Penelitian.....	32
---	----