

**KLASIFIKASI GENRE FILM BERDASARKAN SINOPSIS
MENGUNAKAN METODE NAIVE BAYES**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh
DAMAR JATI BAYUAJI
19.11.1979

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026**

**KLASIFIKASI GENRE FILM BERDASARKAN SINOPSIS
MENGUNAKAN METODE NAIVE BAYES**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

DAMAR JATI BAYUAJI

19.11.2979

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

KLASIFIKASI GENRE FILM BERDASARKAN SINOPSIS
MENGUNAKAN METODE NAIVE BAYES

yang disusun dan diajukan oleh

DAMAR JATI BAYU AJI

19.11.2979

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 Oktober 2025

Dosen Pembimbing



Yoga Pristiwanto, S.Kom, M.Eng

NIK. 190302412

HALAMAN PENGESAHAN

SKRIPSI

KLASIFIKASI GENRE FILM BERDASARKAN SINOPSIS
MENGUNAKAN METODE NAIVE BAYES

yang disusun dan diajukan oleh

DAMAR JATI BAYUAJI

19.11.2979

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Oktober 2025

Nama Penguji
Susunan Dewan Penguji

Tanda Tangan

Eti Pujiastuti, S.kom., M.kom.
NIK. 190302227

Dwi Nurani, S.kom., M.kom.
NIK. 190302236

Yoga Priyanto, S.Kom, M.Eng
NIK. 190302412

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Oktober 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : DAMAR JATI BAYU AJI
NIM : 19.11.2979

Menyatakan bahwa Skripsi dengan judul berikut,

KLASIFIKASI GENRE FILM BERDASARKAN SINDRIPS MENGGUNAKAN METODE NAIVE BAYES

Dosen Pembimbing : Yoga Priyanto, S.Kom, M.Eng

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penemuan SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan penaubatan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 20 November 2025

Yang Menyatakan,



Damar Jati Bayu Aji

HALAMAN PERSEMBAHAN

Puji syukur penulis panjatkan ke hadirat Allah S.W.T atas segala rahmat, nikmat, hidayah dan kesempatan untuk dapat menimba ilmu, sehingga penulis dapat menyelesaikan laporan ini dengan lancar. Dalam menyusun laporan ini penulis banyak menerima bantuan, bimbingan, dan dukungan oleh berbagai pihak. Sebagai wujud rasa syukur, laporan ini penulis persembahkan kepada:

1. Kepada kedua orang tua, Almarhum Bapak Sahlan Rosidi dan Ibu Supadmi atas Segala doa, dukungan tanpa henti, semangat, serta pengorbanan yang telah Ibu dan Almarhum Bapak berikan menjadi fondasi dan penyemangat bagi penulis dalam menuntut ilmu. Jasanya akan selalu dikenang dan doanya senantiasa dipanjatkan.
2. Kepada Bapak Yoga Pristyanto, S.Kom, M.Eng selaku dosen pembimbing saya yang telah mengarahkan dan membantu dalam penyusunan skripsi ini. Saya sangat berterimakasih atas bimbingannya selama ini yang telah memberikan masukan, kritik dan saran yang membangun agar menjadi lebih baik lagi untuk kedepannya.
3. Kepada sahabat dan teman-teman saya yang telah memberikan saya dukungan dan semangat selama masa perkuliahan.
4. Dan semua pihak yang mendukung saya secara langsung ataupun tidak langsung.

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah Subhanahu wa Ta'ala atas rahmat, hidayah, dan karunia-Nya yang melimpah, sehingga penulis dapat menyelesaikan laporan Skripsi yang berjudul **“Klasifikasi Genre Film Berdasarkan Sinopsis Menggunakan Metode Naive Bayes”**. Tujuan dari penyusunan skripsi ini ialah untuk memenuhi syarat memperoleh gelar Sarjana pada program studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta. Didalam pengerjaan skripsi ini telah melibatkan banyak pihak yang sangat membantu dalam banyak hal. Oleh sebab itu disini penulis sampaikan rasa terimakasih sedalam-dalamnya kepada:

1. Allah SWT atas Rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi ini dengan baik dan semoga dapat memberikan manfaat dikemudian hari.
2. Bapak Prof. Dr .M. Suyanto, M.M. selaku Rektor Universitas Amikom Yogyakarta.
3. Ibu Prof. Dr. Kusriani, M.Kom. selaku Deakan Fakultas Ilmu Komputer.
4. Ibu Eli Pujiastuti, M.Kom. selaku Ketua Program Studi S1 Teknik Komputer Universitas Amikom Yogyakarta.
5. Bapak Yoga Pristiyanto, S.Kom, M.Eng selaku dosen pembimbing saya yang telah mengarahkan dan membantu dalam penyusunan skripsi ini.
6. Kepada segenap Dosen, Staff, Karyawan Universitas Amikom Yogyakarta yang telah memberikan ilmu kepada penulis dibangku kuliah dan juga membantu penulis dalam kelancaran administrasi sampai terselesaikannya skripsi ini.
7. Kedua orang tua tercinta, Almarhum Bapak Sahlan Rosidi dan Ibunda Supadmi, yang telah memberikan doa, dukungan, semangat, dan pengorbanan yang tiada henti. Perjuangan dan kasih sayang kalian adalah pondasi dari setiap langkah penulis.
8. Saudara-saudara beserta keluarga yang selalu mendoakan, memberi semangat dan dukungan penuh kepada penulis.

9. Serta kepada pihak yang telah membantu dalam penyusunan skripsi ini yang tidak dapat penulis sebutkan satu per satu.

Penulis berharap semoga skripsi ini dapat bermanfaat bagi semua pihak yang terkait dengan penulisan ini. Dalam penulisan skripsi ini, penulis menyadari masih banyak kekurangan karena perbatasnya pengetahuan dan pengalaman penulis. Oleh karena itu, dengan lapang hati penulis mengharapkan kritik dan saran yang membangun guna menyempurnakan skripsi ini. Akhirnya dengan kerendahan hati penulis mengucapkan terima kasih dan semoga skripsi ini dapat bermanfaat bagi penulis maupun pembaca.

Yogyakarta, 10 Oktober 2025

Penulis



DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR SINGKATAN	xiii
DAFTAR ISTILAH	xiv
INTISARI	xvi
ABSTRACT	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	2
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA	5
2.1 Tinjauan Pustaka	5
2.2 Dasar Teori	9
2.2.1 Machine Learning	9
2.2.2 Python	10
2.2.3 The Movie Database (TMDB)	10
2.2.4 API (Application Programming Interface)	10
2.2.5 Klasifikasi Teks	11
2.2.6 Preprocessing Teks	11

2.2.7 TF-IDF (Term Frequency–Inverse Document Frequency).....	13
2.2.8 Naive Bayes Classifier	14
2.2.9 SMOTE (Synthetic Minority Oversampling Technique).....	16
2.2.10 Evaluasi Model	17
2.2.11 Cross Validation.....	18
BAB III METODE PENELITIAN	19
3.1 Objek Penelitian.....	19
3.2 Alur Penelitian	20
3.2.1 Pengumpulan Data	20
3.2.2 Preprocessing	21
3.2.3 Pengolahan Data	22
3.2.4 Pelatihan Model Naive Bayes	23
3.2.5 Evaluasi dan Hasil Analisa	26
3.3 Alat dan Bahan.....	26
3.3.1 Data Penelitian	26
3.3.2 Alat/Instrumen Penelitian	27
3.3.2.1 Perangkat Lunak (Software)	27
3.3.2.2 Perangkat Keras (Hardware).....	27
BAB IV HASIL DAN PEMBAHASAN	28
4.1 Hasil Penelitian.....	28
4.1.1 Pengumpulan Data.....	28
4.1.2 Preprocessing	29
4.1.2.1 Menghapus Nilai Null	29
4.1.2.2 Melakukan Filter Pada Data Genre.....	30
4.1.2.3 Text Preprocessing.....	32
4.1.3 Pengolahan Data	33
4.1.3.1 Pembobotan TF-IDF	33
4.1.3.2 Penanganan Imbalance Data.....	34
4.1.3.3 Pembagian Data Training dan Testing	35
4.1.4 Pelatihan Model Complement Naive Bayes	35
4.1.5 Evaluasi Model dan Analisis Hasil.....	36
4.1.5.1 Confusion Matrix	36

4.1.5.2 Cross Validation.....	39
4.1.6 Perhitungan Manual Complement Naive Bayes (CNB)	40
4.1.7 Perbandingan Dengan Metode Lain	46
4.1.7.1 Multinomial Naive Bayes (MNB).....	47
4.1.7.2 Support Vector Machine (SVM).....	48
4.1.8 Perbandingan Sebelum dan Sesudah SMOTE-Tomek	48
4.2 Pembahasan	49
4.2.1 Klasifikasi Genre Film Berdasarkan Teks Sinopsis	49
4.2.2 Analisis Performa Complement Naive Bayes	49
4.2.3 Faktor-Faktor yang Mempengaruhi Akurasi	50
4.2.4 Analisis Perbandingan CNB, MNB, dan SVM.....	50
4.2.5 Keterbatasan Umum	52
4.2.6 Perbandingan dengan Penelitian Terdahulu	52
BAB V PENUTUP	54
5.1 Kesimpulan	54
5.2 Saran	55
REFERENSI	57

DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	8
Tabel 3.1 Software penelitian	27
Tabel 3.2 Hardware penelitian	27
Tabel 4.1 dataset film.....	28
Tabel 4.2 informasi dataset	29
Tabel 4.3 informasi dataset setelah menghapus nilai null	30
Tabel 4.4 data genre primer	30
Tabel 4.5 konfigurasi TF-IDF	33
Tabel 4.6 Hasil Penanganan <i>Imbalance data</i>	34
Tabel 4.7 Hasil <i>Complement NB</i>	36
Tabel 4.8 Performa Klasifikasi per Genre	37
Tabel 4.9 Cross Validation	39
Tabel 4.10 Sampel 3 dataset	41
Tabel 4.11 Bobot tiap fitur.....	41
Tabel 4.12 Komplemen untuk kelas Action.....	42
Tabel 4.13 Komplemen untuk kelas Animation	43
Tabel 4.14 Perbandingan Hasil sebelum Balancing Data.....	49
Tabel 4.15 Perbandingan CNB, MNB, SVM	51
Tabel 4.16 perbandingan penelitian terdahulu	53

DAFTAR GAMBAR

Gambar 2.1. confusion matrix.....	18
Gambar 3.1. <i>flowchart</i> penelitian.....	19
Gambar 3.2. flowchar pengumpulan data	20
Gambar 3.3. <i>flowchart</i> preprocessing	21
Gambar 3.4. pengolahan data.....	23
Gambar 3.5. <i>flowchart</i> pengolahan data	25
Gambar 4.1. Distribusi Genre Primer setelah Filter	31
Gambar 4.2. perbandingan data sebelum dan sesudah <i>preprocessing</i>	33
Gambar 4.3. confusion matrik.....	38
Gambar 4.4. Hasil Cross Validation	40

DAFTAR SINGKATAN



API	Application Programming Interface
TMDB	The Movie Database
NLP	Natural Language Processing
TF-IDF	Term Frequency-Inverse Document Frequency
NB	Naive Bayes
NBML	Naive Bayesian Multi-label
SMOTE	Synthetic Minority Over-sampling Technique
SMOTE-Tomek	SMOTE + Tomek Links
K-NN	K-Nearest Neighbors
SVM	Support Vector Machines
LSTM	Long Short-Term Memory
CNN	Convolutional Neural Network
JSON	JavaScript Object Notation
CSV	Comma-Separated Values
HTTP	Hypertext Transfer Protocol
REST	Representational State Transfer

DAFTAR ISTILAH

Genre Film	Kategori film berdasarkan tema, gaya, atau konten cerita.
Sinopsis	Ringkasan cerita film yang digunakan sebagai data teks untuk klasifikasi.
Preprocessing Teks	Proses pembersihan dan persiapan teks sebelum analisis.
Case Folding	Mengubah semua huruf dalam teks menjadi huruf kecil.
Tokenization	Memecah teks menjadi unit-unit kecil (token) seperti kata.
Stemming	Mereduksi kata ke bentuk dasarnya (akar kata).
Stopword Removal	Menghapus kata umum yang tidak bermakna signifikan.
TF-IDF Vectorization	Mengubah teks menjadi vektor numerik berdasarkan bobot TF-IDF.
Imbalanced Data	Kondisi jumlah sampel antar kelas tidak seimbang.
Cross-Validation	Teknik evaluasi model dengan membagi data berulang kali.
Confusion Matrix	Tabel perbandingan hasil prediksi dengan nilai sebenarnya.
Accuracy	Proporsi prediksi benar dari total prediksi.
Precision	Proporsi prediksi positif yang benar.

Recall	Kemampuan model menemukan semua sampel positif.
F1-Score	Rata-rata harmonik dari precision dan recall.
Laplace Smoothing	Teknik menghindari probabilitas nol dengan menambah nilai alpha.
Script Teks	Naskah film berisi dialog dan arahan cerita.
Micro Average F1-Measure	Metrik evaluasi yang menghitung F1-score secara global.
Stratified Sampling	Pembagian data dengan menjaga proporsi kelas.
Feature Selection	Proses memilih fitur paling relevan untuk model.
Overfitting	Kondisi model terlalu kompleks dan menghafal data latih.
Underfitting	Kondisi model terlalu sederhana untuk menangkap pola data.

INTISARI

Pertumbuhan industri film yang pesat, ditandai dengan peningkatan jumlah film nasional dari 36 (2021) menjadi 129 (2023), menimbulkan tantangan dalam mengorganisir dan mengkategorikan film berdasarkan genre secara efisien. Klasifikasi manual yang selama ini dilakukan rentan terhadap subjektivitas dan tidak konsisten, sehingga diperlukan solusi otomatis yang akurat. Penelitian ini bertujuan mengembangkan model klasifikasi genre film berdasarkan sinopsisnya menggunakan algoritma Complement Naive Bayes untuk mengatasi permasalahan tersebut. Metode yang diterapkan meliputi pengumpulan 1000 data film dari TMDb API, dilanjutkan dengan tahapan preprocessing teks (*case folding, cleaning, tokenization, stopword removal, stemming*), ekstraksi fitur TF-IDF, penanganan ketidakseimbangan data dengan SMOTE-Tomek, pelatihan model, dan evaluasi menggunakan confusion matrix serta *5-fold cross-validation*. Hasil penelitian menunjukkan model yang dibangun mencapai akurasi 76,87% dan Macro F1-Score, 76,27% dengan performa terbaik pada genre Horror (F1-Score 0,85) dan terendah pada Action (F1-Score 0,64). Validasi silang mengkonfirmasi konsistensi model dengan akurasi rata-rata 76,21% (± 0.0123). Hasil ini membuktikan bahwa sinopsis film mengandung pola leksikal yang dapat dimanfaatkan untuk klasifikasi genre secara komputasional. Penelitian ini diharapkan berkontribusi dalam pengembangan sistem rekomendasi film, pengkategorian database otomatis, serta memberikan insight bagi pengembang platform streaming dan peneliti dalam bidang *Natural Language Processing* (NLP). Untuk pengembangan selanjutnya, disarankan menggunakan pendekatan *multi-label classification*, dataset yang lebih besar dan seimbang, serta mengeksplorasi model deep learning.

Kata kunci: Klasifikasi Teks, Genre Film, Naive Bayes, Sinopsis, TF-IDF

ABSTRACT

The rapid growth of the film industry, marked by an increase in national film production from 36 (2021) to 129 (2023), presents a challenge in efficiently organizing and categorizing films by genre. Manual classification, as has been done so far, is prone to subjectivity and inconsistency, thus requiring an accurate automated solution. This research aims to develop a model for classifying movie genres based on their synopses using the Complement Naive Bayes algorithm to address this problem. The applied method includes collecting 1000 movie data points from the TMDB API, followed by text preprocessing stages (case folding, cleaning, tokenization, stopword removal, stemming), TF-IDF feature extraction, handling data imbalance with SMOTE-Tomek, model training, and evaluation using a confusion matrix and 5-fold cross-validation. The results show that the built model achieved an accuracy of 76.87% and a Macro F1-Score of 76.27%, with the best performance on the Horror genre (F1-Score 0.85) and the lowest on Action (F1-Score 0.64). Cross-validation confirmed the model's consistency with an average accuracy of 76.21% (± 0.0123). These results prove that movie synopses contain lexical patterns that can be utilized for computational genre classification. This research is expected to contribute to the development of film recommendation systems, automatic database categorization, and provide insights for streaming platform developers and researchers in the field of Natural Language Processing (NLP). For further development, it is recommended to use a multi-label classification approach, a larger and more balanced dataset, and to explore deep learning models.

Keyword: Text Classification, Movie Genre, Naive Bayes, Synopsis, TF-IDF