

**IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK
KLASIFIKASI DATA FILM DAN TV SERIES DI
NETFLIX**

SKRIPSI

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



disusun oleh

Reny Aprillia Safarty

19.11.2896

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2023**

**IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK
KLASIFIKASI DATA FILM DAN TV SERIES DI
NETFLIX**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika



disusun oleh:

Reny Aprillia Safarty

19.11.2896

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

HALAMAN PERSETUJUAN

SKRIPSI

IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK KLASIFIKASI DATA FILM DAN TV SERIES DI NETFLIX

yang disusun dan diajukan oleh

Reny Aprillia Safarty

19.11.2896

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 27 Maret 2023

Dosen Pembimbing,



Erni Seniwati, S.Kom, M.Cs

NIK. 190302231

HALAMAN PENGESAHAN

SKRIPSI

IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK KLASIFIKASI DATA FILM DAN TV SERIES DI NETFLIX

yang disusun dan diajukan oleh

Reny Aprillia Safarty

19.11.2896

Telah dipertahankan di depan Dewan Penguji
pada tanggal 27 Maret 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

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



Wiwi Widayani, M.Kom
NIK. 190302272



Windha Mega PD, M.Kom.
NIK. 190302185



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 27 Maret 2023

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Reny Aprillia Safarty
NIM : 19.11.2896

Menyatakan bahwa Skripsi dengan judul berikut:

IMPLEMENTASI ALGORITMA NAÏVE BAYES UNTUK KLASIFIKASI DATA FILM DAN TV SERIES DI NETFLIX

Dosen Pembimbing : Erni Seniwati, S.Kom, M.Cs

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 penelitian 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 disebutkan 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 pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 27 Maret 2023

Yang Menyatakan,



Reny Aprillia Safarty

HALAMAN PERSEMBAHAN

Yang Utama Dari Segalanya Puji syukur kepada Allah SWT. Taburan cinta dan kasih sayang-Mu telah memberikan kekuatan, membekaliku dengan ilmu serta memperkenalkanku dengan cinta. Atas karunia serta kemudahan yang Engkau berikan akhirnya skripsi yang sederhana ini dapat terselesaikan. Sholawat dan salam selalu terlimpahkan keharibaan Rasullah Muhammad SAW.

Penulis juga mengucapkan terimakasih kepada para pihak yang terlibat baik secara langsung maupun tidak langsung dalam pembuatan skripsi ini sampai selesai.

1. Kedua orang tua saya, Maryunis dan Jamilah yang telah memberikan kasih sayang, segala dukungan, dan cinta kasih maupun materi serta doa yang tiada henti. Semoga ini menjadi langkah awal untuk membuat kalian senang bahagia karena diri ini bisa memenuhi harapan untuk mencapai target waktu kelulusan.
2. Adikku, Rheina Aulia Rahmanda, yang senantiasa memberikan semangat walaupun sering bertengkar tapi hal itu selalu menjadi warna yang tak akan bisa tergantikan.
3. Dosen pembimbing, Erni Seniwati, S.Kom, M.Cs yang telah membimbing serta memberikan masukan terhadap penyusunan skripsi ini sehingga penulis dapat menyelesaikan.
4. Teman kelas Informatika-05 yang selalu memberi masukan dan saran ketika saya kebingungan dalam penyusunan skripsi ini, teman-teman yang selalu makan di burjo panembongan yang selalu menjadi saksi keseruan atau permasalahan yang mengesalkan, temen psu yang sampai sekarang masih akrab seperti pertama kali kenal. Terimakasih untuk kebersamaannya selama menempuh kuliah. Tanpa kalian jogja tidak mungkin bisa seasik ini.
5. Para dosen Universitas Amikom Yogyakarta yang telah membagikan ilmunya selama penulis menempuh pendidikan.

KATA PENGANTAR

Puji dan syukur saya panjatkan kepada Allah SWT atas ridhonya saya dapat menyelesaikan penyusunan skripsi ini. Adapun judul skripsi yang saya ajukan adalah “Implementasi Algoritma Naïve Bayes untuk Klasifikasi Data Film dan Tv Series di Netflix, sebagai salah satu syarat untuk menyelesaikan Program Sarjana (S1) Jurusan Informatika Fakultas Komputer Universitas Amikom Yogyakarta.

Penulis menyadari bahwa skripsi ini tidak mungkin terselesaikan tanpa adanya usaha yang keras, dukungan, bantuan, bimbingan dan nasehat dari berbagai pihak selama penyusunan skripsi ini. Pada kesempatan ini penulis menyampaikan terima kasih setulus-tulusnya kepada:

1. Allah SWT, Tuhan yang Maha Pengasih lagi Maha Penyayang yang telah memberikan rahmat yang tak kunjung henti kepada penulis serta diberikan kelancaran dalam menyelesaikan skripsi.
2. Prof. Dr. M. Suyanto, M.M selaku Rektor Universitas Amikom Yogyakarta.
3. Hanif Al Fatta, S.Kom., M.Kom selaku Dekan Fakultas Ilmu Komputer.
4. Erni Seniwati, S.Kom, M.Cs selaku Dosen Pembimbing.
5. Orang tua dan keluarga yang selalu mendoakan dan mendukung serta memotivasi penulis.
6. Semua pihak yang telah membantu dan tidak dapat disebutkan satu persatu.

Semoga segala kebaikan dan pertolongan semuanya mendapat berkah dari Allah SWT dan akhirnya saya menyadari bahwa skripsi ini masih jauh dari kata sempurna, karena keterbatasan ilmu yang saya miliki. Untuk itu saya dengan kerendahan hati mengharapkan saran dan kritik yang sifatnya membangun dari semua pihak demi membangun dan menyempurnakan penulisan skripsi ini.

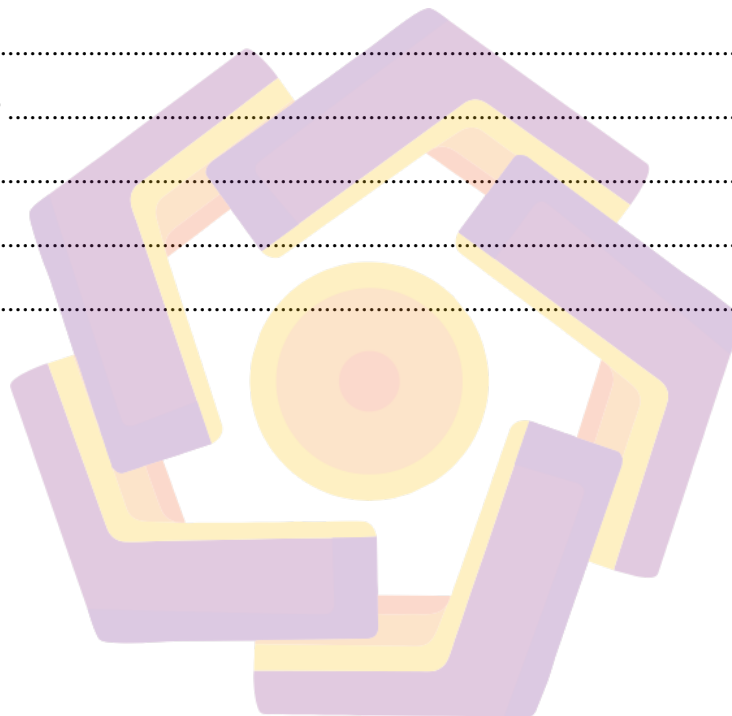
Yogyakarta, 31 Mei 2023

Penulis

DAFTAR ISI

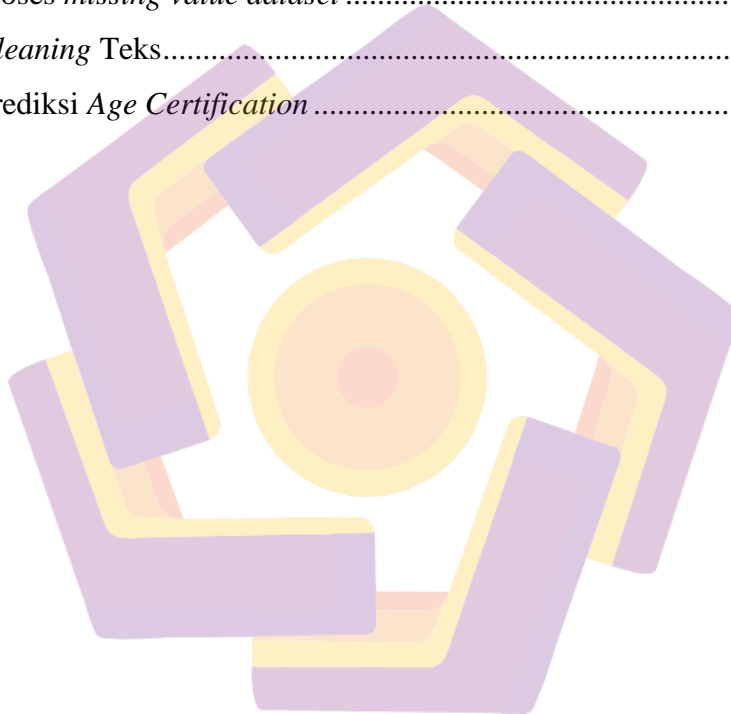
HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR	x
DAFTAR LAMBANG DAN SINGKATAN	xi
DAFTAR ISTILAH.....	xii
INTISARI	xiii
ABSTRACT.....	xiv
BAB 1 PENDAHULUAN.....	1
1.1 Latar belakang.....	1
1.2 Rumusan masalah	2
1.3 Batasan masalah.....	2
1.4 Tujuan Penelitian	3
1.5 Manfaat penelitian.....	3
1.6 Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA.....	5
2.1 Literatur Review	5
2.2 Netflix	15
2.3 <i>Age Rating</i>	16
2.4 Data mining.....	16

2.5 <i>Naïve Bayes</i>	20
2.5 Metode Evaluasi.....	21
BAB III METODE PENELITIAN	23
3.1 Alur penelitian	23
3.2 <i>Dataset</i>	24
3.3 <i>Preprocessing</i>	24
3.8 Spesifikasi Alat	29
BAB IV	30
HASIL DAN PEMBAHASAN	30
4.1 Pembahasan.....	30
BAB V PENUTUP	41
5.1 Kesimpulan	41
5.2 Saran	41
REFERENSI	42



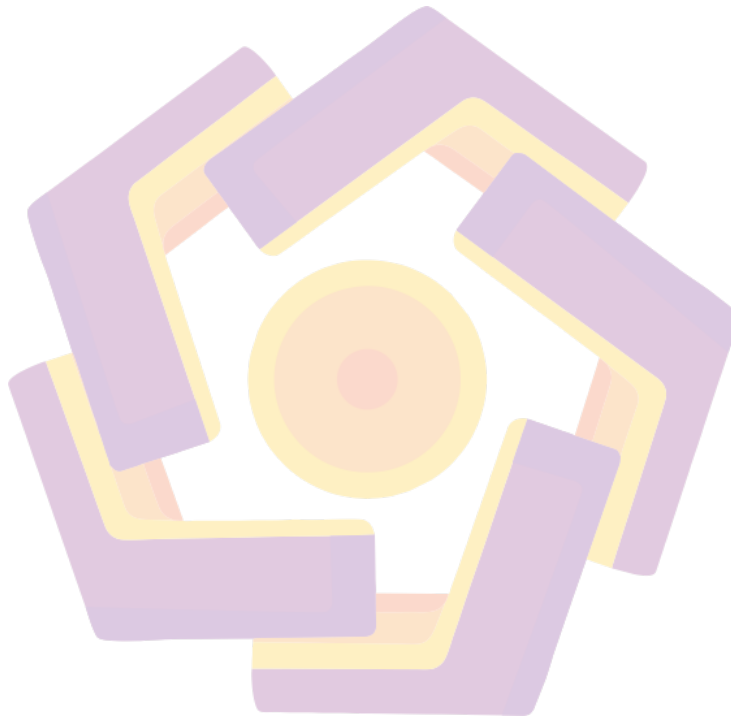
DAFTAR TABEL

Tabel 2.1 Literatur Review.....	15
Tabel 2.2 Evaluasi.....	29
Tabel 3.1 <i>Case Folding</i>	33
Tabel 3.2 <i>Tokenizing</i>	34
Tabel 3.3 <i>Filtering</i>	34
Tabel 3.4 <i>Stemming</i>	35
Tabel 4.1 Isi <i>dataset</i> Netflix Tv Show and Movies	38
Tabel 4.2 Hasil proses <i>missing value dataset</i>	39
Tabel 4.3 Hasil <i>Cleaning</i> Teks.....	43
Tabel 4.4 Hasil Prediksi <i>Age Certification</i>	4



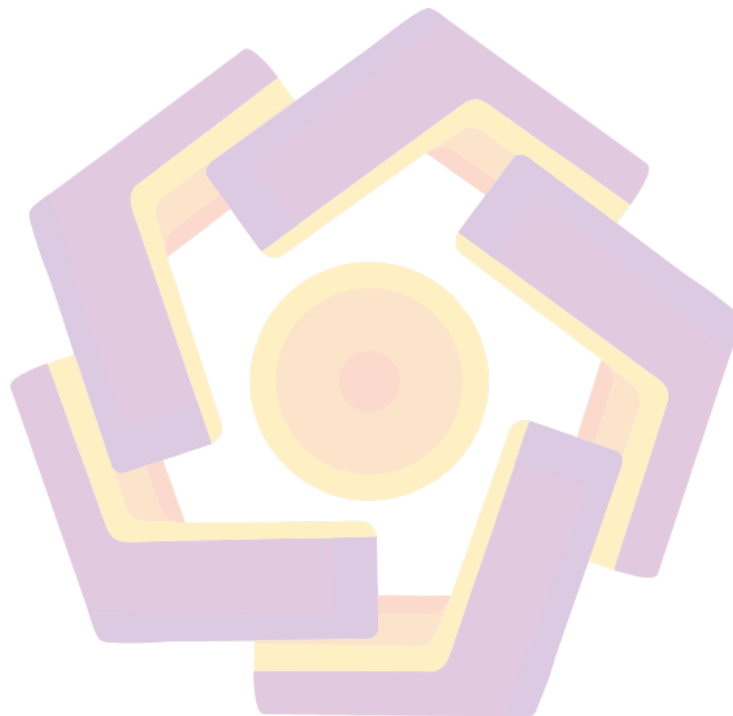
DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	31
Gambar 3.2 <i>Dataset</i> Netflix TV Shows and Movies	32
Gambar 4.1 NLTK	41
Gambar 4.2 Proses <i>Cleaning</i> Teks	42
Gambar 4.3 <i>Build Model</i>	44
Gambar 4.4 Hasil Evaluasi	45
Gambar 4.5 Rating Usia	46



DAFTAR LAMBANG DAN SINGKATAN

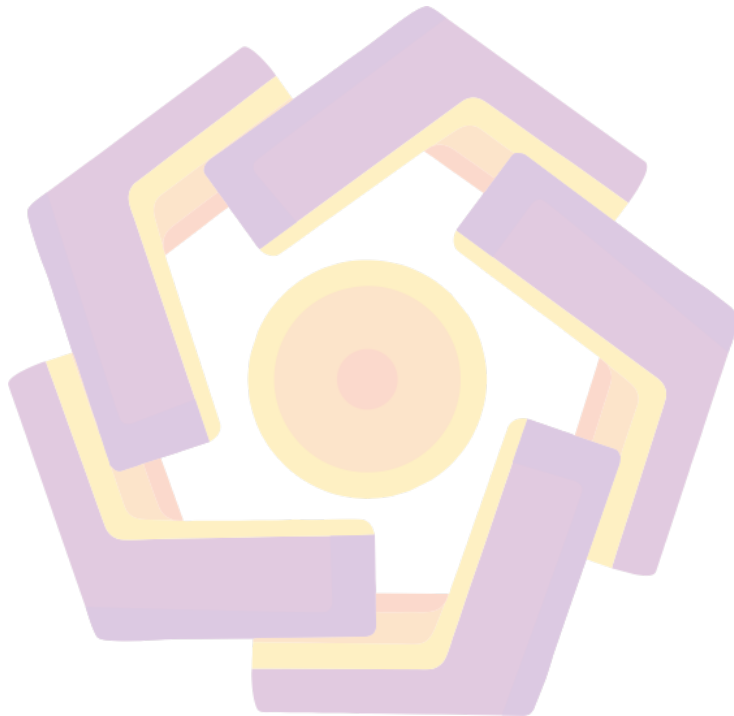
SVoD	Subscription Video on Demand
KDD	Knowledge Discovery in Database
SVM	Support Vector Machines



DAFTAR ISTILAH

Data Mining
Noise

proses ekstraksi informasi dari data yang besar
data yang tidak relevan atau tidak akurat



INTISARI

Netflix adalah platform streaming yang menawarkan berbagai macam film dan tv series hingga dokumenter. Dengan perkembangan teknologi yang sangat pesat kita bisa menonton di berbagai perangkat termasuk *smart TV*, *smartphone*, *tablet*, dan komputer. Ini memungkinkan penonton untuk menikmati layanan di mana saja dan kapan saja dengan berbagai genre seperti komedi, horor, *action*, *thriller*, romantis dan masih banyak lagi. Netflix menggunakan algoritma yang canggih untuk merekomendasi film kepada pengguna berdasarkan *history* yang pernah ditonton sebelumnya. Hal ini membuat mudah pengguna untuk menemukan konten-konten baru dan menjelajah genre yang mungkin belum pernah mereka tonton sebelumnya. Untuk memudahkan penonton dalam memilih film yang sesuai dengan selera, maka perlunya sistem rating usia untuk memberikan informasi bahwa film tersebut layak ditonton atau tidak dan apakah sesuai dengan batasan umur. Proses klasifikasi ini menggunakan metode *Naïve Bayes* dan mengukur hasil atau nilai dari akurasi *naïve bayes*. Terdapat 5850 data film yang digunakan dalam penelitian ini. Sebelum data dilatih, data yang akan digunakan akan dibersihkan dengan metode *preprocessing* diantaranya yaitu *case folding*, *tokenizing*, *stopword dan stemming*. Kemudian data tersebut mendapatkan hasil yang bersih dan dibagi menjadi data *training* dan data *testing*. Hasil dari penelitian ini menunjukkan bahwa metode *Naïve Bayes* berhasil mengklasifikasikan data mengenai film dan TV series di netflix dengan nilai rata-rata akurasi sebesar 59%. Meski skor akurasinya tidak cukup tinggi, Hasil penelitian ini bisa menjadi titik awal untuk mengembangkan metode yang lebih baik untuk mengklasifikasikan konten di Netflix.

Kata kunci: Klasifikasi, *Naïve Bayes*, Netflix, *Classification Report*

ABSTRACT

Netflix is a streaming platform that offers a wide variety of movies and TV series to documentaries. With the rapid development of technology, we can watch on various devices including smart TVs, smartphones, tablets and computers. This allows viewers to enjoy services anywhere and anytime with various genres such as comedy, horror, action, thriller, romance and many more. Netflix uses a sophisticated algorithm to recommend films to users based on history they have watched before. This makes it easy for users to discover new content and explore genres they may not have watched before. To make it easier for viewers to choose films according to their tastes, an age rating system is needed to provide information whether the film is worth watching or not and whether it is within the age limit. This classification process uses the Naïve Bayes method and measures the result or value of naïve Bayes accuracy. There are 5850 film data used in this study. Before the data is trained, the data to be used will be cleaned with preprocessing methods including case folding, tokenizing, stopwords and stemming. Then the data gets clean results and is divided into training data and testing data. The results of this study indicate that the Naïve Bayes method succeeded in classifying data regarding films and TV series on Netflix with an average accuracy value of 59%. Although the accuracy score is not high enough, the results of this research can be a starting point for developing better methods for classifying content on Netflix.

Keyword: *Classification, Naïve Bayes, Netflix, Classification Report.*