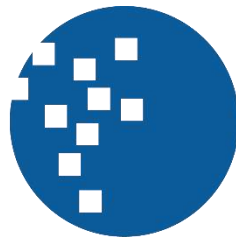


**PENINGKATAN ALGORITMA KNN DAN NAÏVE BAYES  
DENGAN TEKNIK FEATURE SELECTION DALAM  
MEMPREDIKSI RATING FILM INDONESIA**



**UMN**

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

**SKRIPSI**

**Frentzen**

**0000025725**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS TEKNIK DAN INFORMATIKA  
UNIVERSITAS MULTIMEDIA NUSANTARA  
TANGERANG**

**2022**

**PENINGKATAN ALGORITMA KNN DAN NAÏVE BAYES  
DENGAN TEKNIK FEATURE SELECTION DALAM  
MEMPREDIKSI RATING FILM INDONESIA**



**SKRIPSI**

Diajukan sebagai salah satu syarat untuk memperoleh

Gelar Sarjana Komputer (S. Kom)

**Frentzen**

**0000025725**

**UMMN**

**PROGRAM STUDI SISTEM INFORMASI  
TEKNIK DAN INFORMATIKA**

**UNIVERSITAS MULTIMEDIA NUSANTARA  
TANGERANG**

**2022**

## HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Frentzen  
Nomor Induk Mahasiswa : 00000025725  
Program studi : Sistem Informasi

Skripsi dengan judul:

### **PENINGKATAN ALGORITMA KNN DAN NAÏVE BAYES DENGAN TEKNIK FEATURE SELECTION DALAM MEMPREDIKSI RATING FILM INDONESIA**

merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/ penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas akhir yang telah saya tempuh.

Tangerang, 18 Januari 2022



(Frentzen)

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A

## HALAMAN PERSETUJUAN

Skripsi dengan judul

“PENINGKATAN ALGORITMA NAIVE BAYES DAN KNN DENGAN TEKNIK FEATURE SELECTION DALAM MEMPREDIKSI RATING FILM INDONESIA”

Oleh

Nama : Frentzen  
NIM : 00000025725  
Program Studi : Sistem Informasi  
Fakultas : Teknik dan Informatika

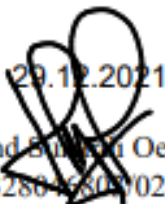
Telah disetujui untuk diajukan pada

Sidang Ujian Skripsi Universitas Multimedia Nusantara

Tangerang, 3 Januari 2022

Pembimbing

Ketua Program Studi

  
29.12.2021

Ir. Raymond Sulaiman Oetama, M.C.I.S.  
328055809023897

  
29122021

Ririn Ikana Desanti, S.Kom., M.Kom.

## HALAMAN PENGESAHAN

Skripsi dengan judul

“PENINGKATAN ALGORITMA KNN DAN NAÏVE BAYES DENGAN  
TEKNIK FEATURE SELECTION DALAM MEMPREDIKSI RATING FILM  
INDONESIA”

Oleh

Nama : Frentzen  
NIM : 00000025725  
Program Studi : Sistem Informasi  
Fakultas : Teknik dan Informatika

Telah diujikan pada hari Selasa, 18 Januari 2022


Pukul 10.00 s/d 12.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut.

Ketua Sidang

Penguji

  
Digitally signed by Suryasari  
DN: cn=Suryasari  
gn=Suryasari c=Indonesia  
h=D=Universitas Multimedia  
Nusantara ou=Information  
System  
e=suryasari@umm.ac.id  
Reason: I am approving this  
document  
Date: 2022.02.09  
Ketua Sidang  
Suryasari, S.Kom., M.T.  
0323088301/067345

  
08/02/2022  
Penguji  
Monika Evelin Johan, S.Kom., M.MSI.  
0327059501/071281

Pembimbing

  
Ir. Raymond Suardi Oetama, MCIS.  
328046803/023897

Ketua Prodi Sistem Informasi

  
09022021  
Ketua Prodi Sistem Informasi  
Ririn Ikana Desanti, S.Kom., M.Kom.

## HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Multimedia Nusantara, saya yang bertandatangan di bawah ini:

Nama : Frentzen  
NIM : 000000025725  
Program Studi : Sistem Informasi  
Fakultas : Teknik dan Informatika  
Jenis Karya : ~~Tesis/Skripsi/Laporan Magang~~

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Multimedia Nusantara Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul:

“PENINGKATAN ALGORITMA KNN DAN NAÏVE BAYES DENGAN TEKNIK FEATURE SELECTION DALAM MEMPREDIKSI RATING FILM INDONESIA”

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non eksklusif ini Universitas Multimedia Nusantara berhak menyimpan, mengalih media / format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis / pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 3 Januari 2022

Yang menyatakan,



(Frentzen)

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A

## KATA PENGANTAR

Puji Syukur atas selesainya penulisan Skripsi ini dengan judul: “PENINGKATAN ALGORITMA KNN DAN NAÏVE BAYES DENGAN TEKNIK FEATURE SELECTION DALAM MEMPREDIKSI RATING FILM INDONESIA” dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Jurusan Sistem Informasi Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan tugas akhir ini, sangatlah sulit bagi saya untuk menyelesaikan tugas akhir ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc. , selaku Dekan Fakultas Universitas Multimedia Nusantara.
3. Ririn Ikana Desanti, S.Kom., M.Kom, selaku Ketua Program Studi Universitas Multimedia Nusantara.
4. Ir. Raymond Sunardi Oetama, M.C.I.S., sebagai Pembimbing yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya tesis ini.
5. Orang Tua dan keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tesis ini.

Semoga karya ilmiah ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 3 Januari 2022



(Frentzen)

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A

**PENINGKATAN ALGORITMA KNN DAN NAÏVE BAYES  
DENGAN TEKNIK FEATURE SELECTION DALAM  
MEMPREDIKSI RATING FILM INDONESIA**

(Frentzen)

**ABSTRAK**

Berdasarkan survey yang dilakukan oleh IDN Times dengan 411 responden, menunjukkan penonton di Indonesia memilih film yang akan ditonton berdasarkan beberapa faktor. Orang-orang memilih film untuk ditonton di bioskop berdasarkan *genre* ada sebesar 16,4 persen, pemain atau aktor sebesar 14,5 persen, dan sutradara sebesar 12,7 persen.

Saat ini minat peneliti-peneliti tentang rating film Indonesia sangatlah sedikit. Sebagian besar dari mereka lebih banyak membahas tentang film di luar negeri dibandingkan dengan film-film Indonesia. Sehingga dibutuhkan penelitian yang membahas tentang rating film Indonesia. Penelitian ini akan menggunakan tools *data mining* yaitu RapidMiner dengan metode kerangka kerja CRISP-DM. Penelitian dilakukan dengan melakukan perbandingan antara 2 algoritma yaitu Naïve Bayes dan KNN untuk melakukan prediksi pada rating film Indonesia dengan menggunakan teknik *feature selection*.

Penelitian ini menghasilkan hasil yaitu algoritma KNN memiliki performa yang lebih baik daripada Naïve Bayes dengan teknik *feature selection* maupun tanpa *feature selection*. Dari 3 model yang dibuat untuk KNN yaitu dengan Forward Selection, Backward Elimination, dan tanpa *feature selection* semuanya menghasilkan nilai akurasi yang sama yaitu 86,65% sedangkan Naïve Bayes mendapat nilai akurasi sebesar 85,13% dengan forward selection, 84,91% dengan backward elimination, dan 71,32% tanpa teknik *feature selection*.

**Kata kunci:** CRISP-DM, *feature selection*, KNN, Naïve Bayes, *rating*

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A



**ENHANCEMENT OF KNN AND NAÏVE BAYES ALGORITHM  
WITH FEATURE SELECTION TECHNIQUE IN PREDICTING  
INDONESIAN FILM RATINGS**

(Frentzen)

**ABSTRACT (English)**

*Based on a survey conducted by IDNTimes with 411 respondents, it shows that viewers in Indonesia choose films to watch based on several factors. People choose films to watch in theaters based on genre by 16.4 percent, players or actors by 14.5 percent, and directors by 12.7 percent.*

*Currently, there is very little interest among researchers regarding Indonesian film ratings. Most of them talk more about films abroad than Indonesian films. So it takes research that discusses the rating of Indonesian films. This research will use data mining tools, namely RapidMiner with the CRISP-DM framework method. The research was conducted by comparing 2 algorithms, namely Naïve Bayes and KNN to make predictions on Indonesian film ratings using feature selection techniques.*

*This study resulted in the result that the KNN algorithm has a better performance than Naïve Bayes with feature selection techniques or without feature selection. Of the 3 models made for KNN, namely with Forward Selection, Backward Elimination, and without feature selection, all of them produce the same accuracy value of 86.65% while Naïve Bayes gets an accuracy value of 85.13% with forward selection, 84.91% with backward elimination, and 71.32% without feature selection technique.*

**Keywords:** CRISP-DM, feature selection, KNN, Naïve Bayes, rating

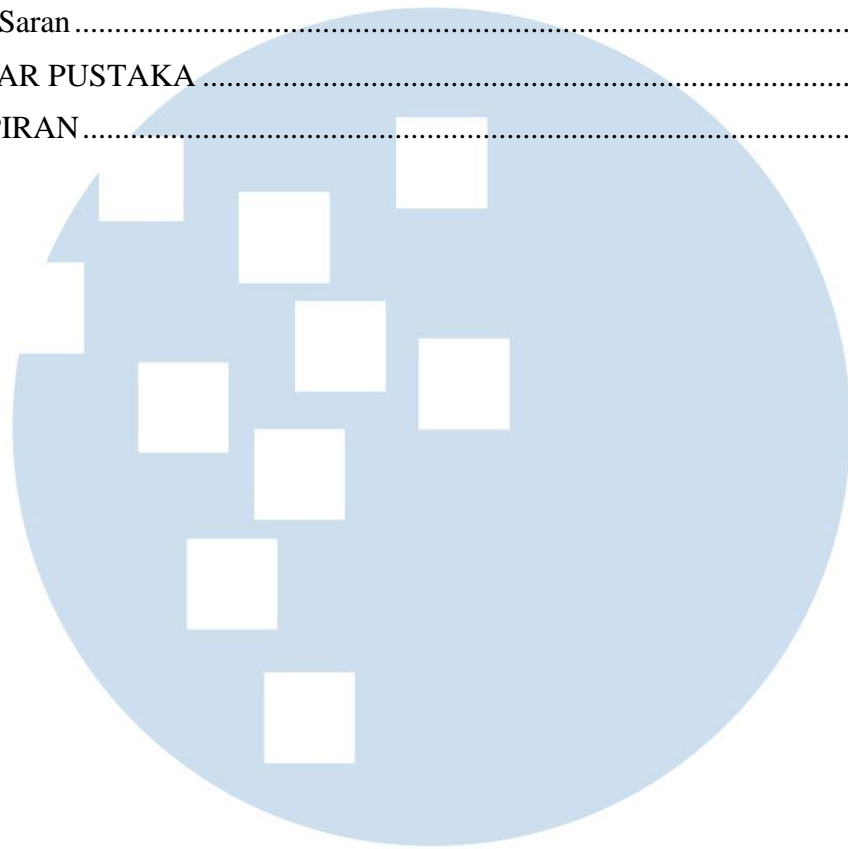
U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A

## DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT .....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS .....	v
KATA PENGANTAR .....	vi
ABSTRAK.....	vii
<i>ABSTRACT (English)</i> .....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR RUMUS .....	xv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	3
1.3. Batasan Masalah.....	3
1.4. Tujuan dan Manfaat Penelitian.....	4
1.4.1. Tujuan Penelitian .....	4
1.4.2. Manfaat Penelitian .....	4
BAB II LANDASAN TEORI.....	5
2.1 Film .....	5
2.2 <i>Rating</i> Film.....	6
2.3 CRISP-DM .....	6
2.4 Naïve Bayes.....	8
2.5 KNN (K-Nearest Neighbor) .....	9
2.6 <i>Feature selection</i> .....	9
2.7 <i>Forward selection</i> .....	9
2.8 <i>Backward elimination</i> .....	10
2.9 Tools.....	10
2.10 Penelitian Terdahulu.....	11

BAB III METODOLOGI PENELITIAN .....	13
3.1. Gambaran Umum Objek Penelitian .....	13
3.1 Metode Penelitian.....	13
3.1.1 Data Collection.....	13
3.1.2 Variabel Independen .....	14
3.1.3 Variabel Dependen.....	15
3.2 Alur Penelitian.....	15
3.2.1 Business Understanding.....	16
3.2.2 Data Understanding.....	16
3.2.3 Data Preparation.....	17
3.2.4 Modeling .....	17
3.2.5 Evaluation .....	20
3.2.6 Deployment.....	20
BAB IV ANALISIS DAN HASIL PENELITIAN .....	21
4.1 Business Understanding .....	21
4.2 Data Understanding .....	22
4.3 Data Preparation .....	22
4.4 Modeling .....	24
4.4.1. Retrieve Dataset .....	26
4.4.2. Select Attributes .....	27
4.4.3. Filter Tahun.....	28
4.4.4. Menentukan Label.....	29
4.4.5. Split Data.....	29
4.4.6. Multiply.....	30
4.4.7. Modeling <i>Forward selection</i> KNN.....	31
4.4.8. Modeling <i>Forward selection</i> Naïve Bayes .....	34
4.4.9. Modeling <i>Backward elimination</i> KNN .....	36
4.4.10. Modeling <i>Backward elimination</i> Naïve Bayes.....	39
4.4.11. Modeling KNN tanpa <i>Feature selection</i> .....	41
4.4.12. Modeling Naïve Bayes Tanpa <i>Feature selection</i> .....	43
4.5 Evaluation.....	44
BAB V KESIMPULAN SARAN .....	47

5.1. Kesimpulan.....	47
5.2. Saran.....	48
DAFTAR PUSTAKA.....	49
LAMPIRAN.....	51



UMMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu .....	11
Tabel 3. 1 Tipe Data.....	14
Tabel 3. 2 Perbandingan Tools .....	19
Tabel 4. 1 Pembagian Class .....	23
Tabel 4. 2 Variabel Terpilih.....	45
Tabel 5. 1 Perbandingan Akurasi Penelitian .....	47



# UMMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

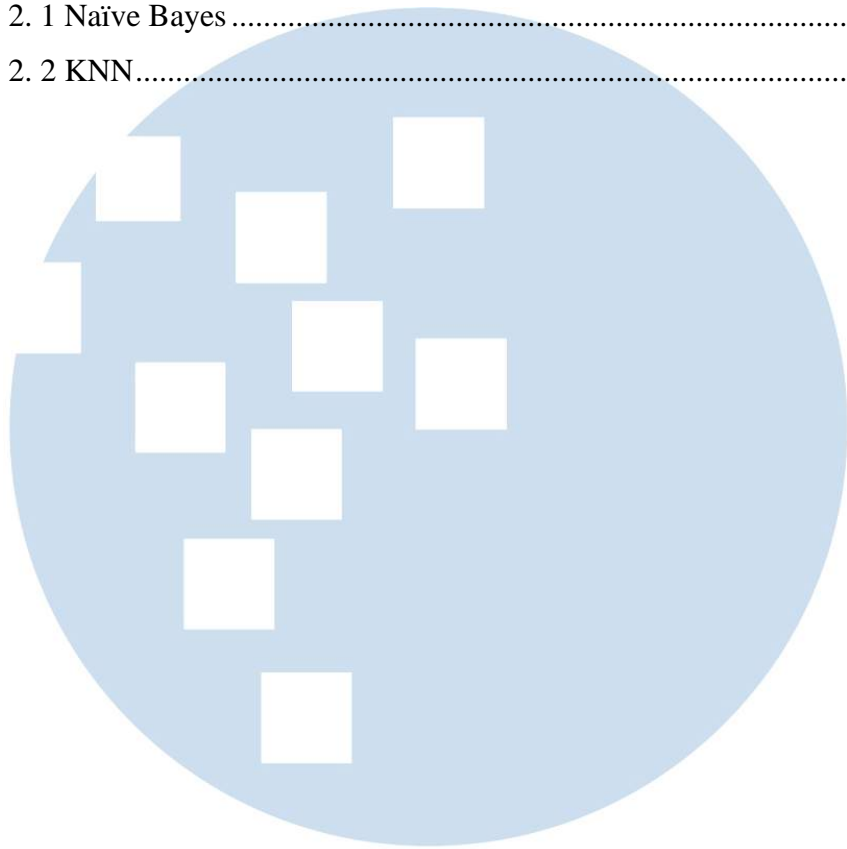
## DAFTAR GAMBAR

Gambar 2. 1 CRISP-DM .....	7
Gambar 3. 1 CRISP-DM .....	16
Gambar 3. 2 Flowchart Modeling .....	18
Gambar 4. 1 Grafik Perbandingan Penonton dan <i>Vote</i> .....	21
Gambar 4. 2 Syntax Missing Value .....	23
Gambar 4. 3 Hasil Syntax Missing Value .....	24
Gambar 4. 4 Hasil Pencarian K .....	25
Gambar 4. 5 Flowchart Modeling .....	25
Gambar 4. 6 Hasil RapidMiner .....	26
Gambar 4. 7 Operator Retrieve .....	26
Gambar 4. 9 Menu Select Attributes .....	27
Gambar 4. 8 Operator Select Attributes .....	27
Gambar 4. 11 Menu Filter Examples .....	28
Gambar 4. 10 Operator Filter Examples .....	28
Gambar 4. 12 Operator Set Role .....	29
Gambar 4. 13 Menu Set Role .....	29
Gambar 4. 15 Menu Split Data .....	30
Gambar 4. 14 Operator Split Data .....	30
Gambar 4. 16 Operator Multiply .....	31
Gambar 4. 17 Operator Forward Selection KNN .....	31
Gambar 4. 18 Operator Cross Validation .....	32
Gambar 4. 19 Isi Operator Cross Validation .....	32
Gambar 4. 20 Nilai K Operator KNN .....	33
Gambar 4. 21 Hasil Forward Selection KNN .....	33
Gambar 4. 22 Akurasi Forward Selection KNN .....	33
Gambar 4. 23 Performance Vector Forward Selection KNN .....	34
Gambar 4. 24 Operator Forward Selection NB .....	34
Gambar 4. 25 Operator Cross Validation NB .....	34
Gambar 4. 26 Isi Operator Cross Validation .....	35
Gambar 4. 27 Hasil Forward Selection NB .....	35

Gambar 4. 28 Akurasi Forward Selection NB .....	35
Gambar 4. 29 Performance Vector Forward Selection NB.....	36
Gambar 4. 30 Operator Backward Elimination KNN.....	36
Gambar 4. 31 Operator Cross Validation.....	37
Gambar 4. 32 Isi Operator Cross Validation.....	37
Gambar 4. 33 Nilai K Pada Operator KNN .....	37
Gambar 4. 34 Hasil Backward Elimination KNN.....	38
Gambar 4. 35 Hasil Backward Elimination KNN.....	38
Gambar 4. 36 Performance Vector Backward Elimination KNN.....	38
Gambar 4. 37 Operator Backward Elimination NB .....	39
Gambar 4. 38 Operator Cross Validation.....	39
Gambar 4. 39 Isi Cross Validation.....	40
Gambar 4. 40 Hasil Backward Elimination NB.....	40
Gambar 4. 41 Akurasi Backward Elimination NB.....	40
Gambar 4. 42 Performance Vector Backward Elimination NB.....	41
Gambar 4. 43 Operator Cross Validation.....	41
Gambar 4. 44 Isi Cross Validation.....	42
Gambar 4. 45 Nilai K Operator KNN .....	42
Gambar 4. 46 Hasil KNN Tanpa <i>Feature selection</i> .....	42
Gambar 4. 47 Performance Vector KNN.....	43
Gambar 4. 48 Operator Cross Validation.....	43
Gambar 4. 49 Hasil NB.....	44
Gambar 4. 50 Isi Cross Validation.....	44
Gambar 4. 51 Performance Vector NB .....	44
Gambar 4. 52 Grafik Perbandingan Performance .....	45

DAFTAR RUMUS

Rumus 2. 1 Naïve Bayes ..... 8  
Rumus 2. 2 KNN..... 9



UMMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA