

**ANALISIS SENTIMEN PADA APLIKASI WETV DENGAN
METODE NAÏVE BAYES, DECISION TREE DAN RANDOM
FOREST**



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

Yeremia Steven Putra Mareby

00000042528

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

2024

**ANALISIS SENTIMEN PADA APLIKASI WETV DENGAN
METODE NAÏVE BAYES, DECISION TREE DAN RANDOM**

FOREST



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer (S.Kom)

Yeremia Steven Putra Mareby

00000042528

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS TEKNIK DAN INFORMATIKA

UNIVERSITAS MULTIMEDIA NUSANTARA

TANGERANG

2024

i

HALAMAN PERNYATAAN TIDAK PLAGIAT

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Yeremia Steven Putra Mareby

Nomor Induk Mahasiswa : 00000042528

Program studi : Sistem Informasi

Skripsi dengan judul:

ANALISIS SENTIMEN PADA APLIKASI WETV DENGAN METODE NAÏVE BAYES, DECISION TREE DAN RANDOM FOREST

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, 30 Mei 2024



Yeremia Steven Putra Mareby

UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PENGESAHAN

HALAMAN PENGESAHAN

Skripsi dengan judul

ANALISIS SENTIMEN PADA APLIKASI WETV DENGAN METODE NAÏVE
BAYES, DECISION TREE DAN RANDOM FOREST

Oleh

Nama : Yeremia Steven Putra Mareby
NIM : 00000042528
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

Telah diujikan pada hari Kamis, 30 Mei 2024

Pukul 15.00 s.d 17.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut.

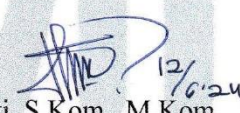
Ketua Sidang


Rudi Sutomo, S.Kom., M.Si., M.Kom.
0222057501

Penguji


Ir. Raymond Sunardi Oetama, M.CIS
328046803

Pembimbing


Ririn Ikana Desanti, S.Kom., M.Kom
313058001

Ketua Program Studi Sistem Informasi


Ririn Ikana Desanti, S.Kom., M.Kom
313058001

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini:

Nama : Yeremia Steven Putra Mareby

NIM : 00000042528

Program Studi : Sistem Informasi

Jenjang : S1

Judul Karya Ilmiah :

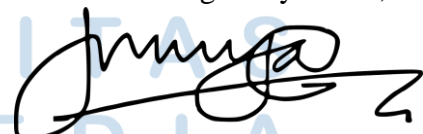
ANALISIS SENTIMEN PADA APLIKASI WETV DENGAN METODE NAÏVE
BAYES, DECISION TREE DAN RANDOM FOREST

Menyatakan dengan sesungguhnya bahwa saya bersedia:

- Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial. Saya tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.
- Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: Dalam proses pengajuan penerbitan ke dalam jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*)*.

Tangerang, 13 Juni 2024

Yang menyatakan,



Yeremia Steven Putra Mareby

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 peneliti panjatkan kepada Tuhan Yang Maha Esa karena berkat dan kasih karunia-Nya, peneliti dapat menyelesaikan skripsi dengan baik yang berjudul “Analisis Sentimen Pada Aplikasi WeTV Dengan Metode Naïve Bayes, Decision Tree dan Random Forest” dilakukan untuk memenuhi syarat mendapatkan gelar Sarjana Komputer pada jurusan Sistem Informasi, Teknik dan Informatika Universitas Multimedia Nusantara.

Peneliti juga menyadari bahwa adanya bantuan serta bimbingan dari berbagai pihak dari awal pembuatan skripsi ini sampai saat skripsi ini telah selesai dibuat. Oleh karena itu peneliti mengucapkan terima kasih kepada:

1. Bapak Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng Niki Prastomo, S.T., M.Sc, selaku Dekan Fakultas Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, S.Kom., M.Kom, selaku Ketua Program Studi Universitas Multimedia Nusantara dan sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
4. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.
5. Ivan Pratama, Amelia Wulandari, Garry Julius, Joseph Michael, Jessica, Shandra Isabelle dan Christin Helen yang telah memberikan dukungan dan semangat dalam menyelesaikan tugas akhir ini.
6. Teman-teman seperjuangan dari awal sampai penulisan tugas akhir Vincentius Juan, Rico Tanuwijaya, Marcellius Cahyadi, Gregorius Dinov, Henry Wirawan dan Ferrozio Wijaya, yang telah memberikan masukan serta bantuan dalam penulisan tugas akhir ini.

Akhir kata, dalam menyelesaikan penelitian ini, peneliti memiliki harapan yang kuat bahwa hasil penelitian yang telah dikembangkan dalam bentuk skripsi ini dapat memberikan kontribusi berarti dan bermanfaat kepada para pembaca

Tangerang, 13 Juni 2024



Yerima Steven Putra Mareby



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

ANALISIS SENTIMEN PADA APLIKASI WETV DENGAN METODE NAÏVE BAYES, DECISION TREE DAN RANDOM FOREST

Yeremia Steven Putra Mareby

ABSTRAK

Adanya peningkatan dari penggunaan internet di Indonesia khususnya pada pembayaran konten *digital* pada tahun 2023 lalu. Sejak saat itu penggunaan aplikasi khususnya *streaming* semakin meningkat karena sudah sebagai kebutuhan dari setiap masyarakat untuk mendapatkan hiburan seperti nonton atau *streaming* melalui gawai. Oleh sebab itu, tujuan dari penelitian ini adalah untuk mengidentifikasi mayoritas sentimen dari ulasan pengguna pada WeTV, sebuah aplikasi *streaming* terkenal di Indonesia yang masih baru belakangan ini hadir pada tahun 2018 yang mampu bersaing dengan aplikasi *streaming* besar di Indonesia dengan memuncaki top 6 khusus penggunaan aplikasi *streaming* terbanyak di Indonesia. Metode yang diterapkan meliputi dari penggunaan algoritma Naïve Bayes, Decision Tree, dan Random Forest. Berdasarkan ketiga algoritma tersebut, model dengan performa terbaik yaitu Naïve Bayes kemudian Random Forest dan Decision Tree. Hasil dari penelitian menunjukkan bahwa mayoritas pendapat pengguna terhadap aplikasi WeTV cenderung negatif. Dalam model digunakan dari ketiga algoritma yaitu Naïve Bayes yang merupakan algoritma terbaik dengan memiliki akurasi sebesar 79,1%. Setelah dilakukan akurasi model dari Naïve Bayes akan digunakan dalam *deployment* untuk *website* sederhana yang dapat dipakai oleh pengguna.

Kata kunci: Analisis sentimen, Decision Tree, Naïve Bayes, Random Forest, WeTV.

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

SENTIMENT ANALYSIS ON WETV APPLICATION WITH NAÏVE BAYES, DECISION TREE AND RANDOM FOREST

METHODS

Yeremia Steven Putra Mareby

ABSTRACT (English)

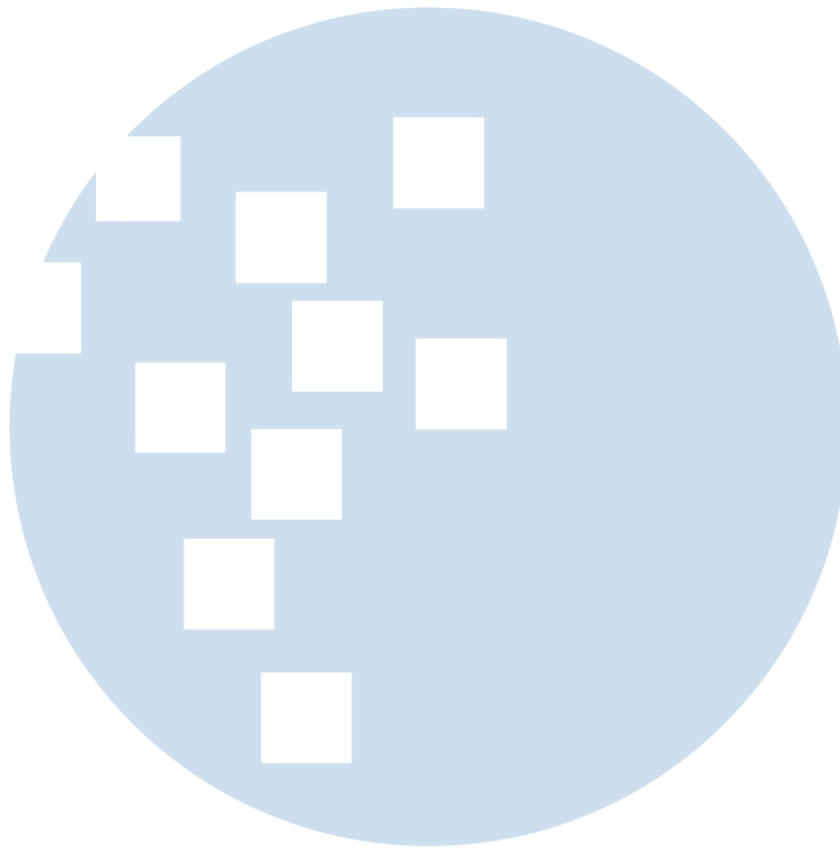
There is an increase in internet usage in Indonesia, especially in digital content payments in 2023. Since then, the use of applications, especially streaming, has increased because it has become a necessity for every community to get entertainment such as watching or streaming through devices. Therefore, the purpose of this research is to identify the majority of sentiments from user reviews on WeTV, a well-known streaming application in Indonesia that was recently present in 2018 which was able to compete with major streaming applications in Indonesia by topping the top 6 specifically the highest use of streaming applications in Indonesia. The methods applied include the use of Naïve Bayes, Decision Tree, and Random Forest algorithms. Based on the three algorithms, the model with the best performance is Naïve Bayes then Random Forest and Decision Tree. The results of the study show that the majority of user opinions of the WeTV application tend to be negative. In the model used from the three algorithms, Naïve Bayes is the best algorithm with an accuracy of 79.1%. After the accuracy of the model from Naïve Bayes will be used in deployment for a simple website that can be used by users

Keywords: *Sentiment analysis, Decision Tree, Naïve Bayes, Random Forest, WeTV.*

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS	iv
KATA PENGANTAR	v
ABSTRAK	vii
ABSTRACT (English)	viii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xv
DAFTAR RUMUS	xvi
BAB I PENDAHULUAN	1
1.1.1 Latar Belakang	1
1.1.1.1 Rumusan Masalah	5
1.2 Batasan Masalah	5
1.3 Tujuan dan Manfaat Penelitian	6
1.3.1 Tujuan Penelitian	6
1.3.2 Manfaat Penelitian	6
1.3.3 Sistematika Penulisan	7
BAB II LANDASAN TEORI	8
2.1 Penelitian Terdahulu	8
2.2 Teori Tentang Skripsi	16
2.2.1 Analisis Sentimen	16
2.2.2 Google Play Store	17
2.2.3 Text Mining	17
2.2.4 Confusion Matrix	18
2.2.5 Term Frequency-Inverse Document Frequency (TF-IDF)	20
2.3 Teori Tentang Algoritma	20
2.3.1 Naïve Bayes Classifier	20

2.3.2	Decision Tree	21
2.3.3	Random Forest	22
2.4	Teori Tentang Tools	23
2.4.1	Python	23
2.4.2	Anaconda Navigator	23
2.4.3	Jupyter Notebook	24
BAB III	METODOLOGI PENELITIAN	25
3.1	Gambaran Umum Objek Penelitian	25
3.2	Metode Penelitian	25
3.2.1	Alur Penelitian	25
3.2.2	Metode Data Mining	28
3.3	Teknik Pengumpulan Data	30
3.3.1	Populasi dan Sampel	30
3.3.2	Periode Pengambilan Data	31
3.4	Teknik Analisis Data	31
BAB IV	ANALISIS DAN HASIL PENELITIAN	32
4.1	<i>Business Understanding</i>	32
4.2	<i>Data Understanding</i>	33
4.3	<i>Data Preparation</i>	35
4.3.1	Labeling	35
4.3.2	Data Cleaning	36
4.3.3	Data Pre-Processing	36
4.4	<i>Modeling</i>	41
4.4.1.1	Data Split Training 70% dan Testing 30%	41
4.4.2	Data Split Training 80% dan Testing 20%	50
4.5	<i>Evaluation</i>	59
4.6	<i>Deployment</i>	62
4.6.1.1	Hasil dan Pembahasan	65
BAB V	SIMPULAN DAN SARAN	67
5.1	Simpulan	67
5.2	Saran	68
DAFTAR PUSTAKA	69



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	8
Tabel 2. 2 Tabel Confusion Matrix	18
Tabel 4. 1 Penjelasan Head Dataset	34
Tabel 4. 2 Perbandingan performa algoritma 70% dan 30%	60
Tabel 4. 3 Perbandingan performa algoritma 80% dan 20%	60
Tabel 4. 4 Perbandingan performa algoritma pada data validation 70% dan 30%	60
Tabel 4. 5 Perbandingan performa algoritma pada data validation 80% dan 20%	61
Tabel 4. 6 Perbandingan dari algoritma Naïve Bayes yang terbaik pada data validation terhadap data splitting	61
Tabel 4. 7 Perbandingan algoritma yang digunakan dengan penelitian lainnya...	65
Tabel 4. 8 Perbandingan akurasi Algoritma Naive Bayes pada penelitian sebelumnya.....	65

UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR GAMBAR

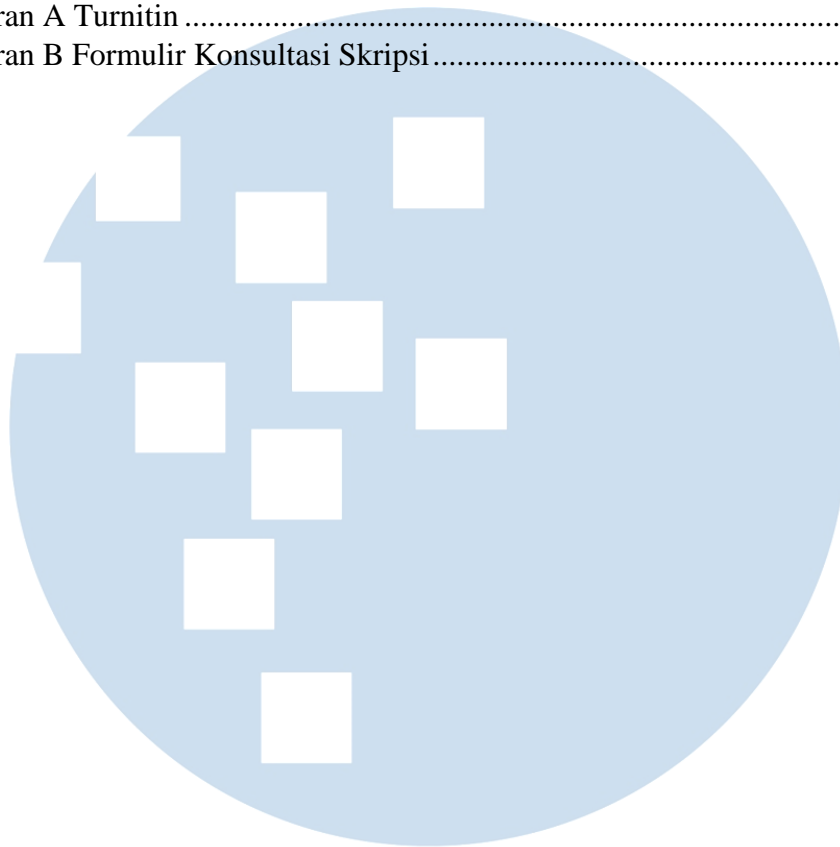
Gambar 1. 1. Daftar konten berbayar pada pengguna internet	1
Gambar 1. 4. Pengguna Aplikasi Video on Demand (VoD) Paling Banyak Digunakan di Indonesia.....	3
Gambar 2. 1 Sentimen Analisis.....	16
Gambar 2. 2 Decision Tree Classifier.....	21
Gambar 2. 3 Random Forest Classifier.....	22
Gambar 2. 4 Logo Python.....	23
Gambar 2. 5 Anaconda Navigator Logo.....	23
Gambar 2. 6 Jupyter Notebook Logo.....	24
Gambar 3. 1 Diagram Alur Penelitian.....	25
Gambar 4. 1 Proses pengambilan data dari Review Google Play Store	33
Gambar 4. 2 Tampilan Head dataset yang menampilkan 4 data teratas	34
Gambar 4. 3 Pemberian label pada data dari Ulasan dan pembersihan data.....	36
Gambar 4. 4 Proses Cleaning Data	36
Gambar 4. 5 Proses pengunduhan Package NLTK.....	37
Gambar 4. 6 Proses menjalankan fungsi Stop Words.....	37
Gambar 4. 7 Stemming pada Dataset.....	38
Gambar 4. 8 Fungsi info() untuk memberikan informasi struktur jumlah data	39
Gambar 4. 9 BarChart ulasan positif (atas) dan negatif (bawah).....	40
Gambar 4. 10 Grafik Distribusi Score (Kiri) dan Sentimen (Kanan)	40
Gambar 4. 11 Metrics modul.....	41
Gambar 4. 12 Proses pemisahan data 70% dan 30%	41
Gambar 4. 13 Fungsi dari value_counts 70% dan 30%	42
Gambar 4. 14 TF - IDF untuk data Training dan Validation	43
Gambar 4. 15 Algoritma Naive Bayes dan Akurasi 70% dan 30%	44
Gambar 4. 16 Hasil Classification_Report pada Alogritma Naive Bayes 70% dan 30%	44
Gambar 4. 17 Visualisasi Heatmap pada Confusion Matrix dari algoritma Naive Bayes 70% dan 30%.....	45
Gambar 4. 18 Algoritma Decision Tree dan Akurasi 70% dan 30%	46
Gambar 4. 19 Hasil Classification_Report pada Alogritma Decision Tree 70% dan 30%	46
Gambar 4. 20 Visualisasi Heatmap Confusion Matrix dari algoritma Decision Tree 70% dan 30%	47
Gambar 4. 21 Algoritma Random Forest dan Akurasi 70% dan 30%	48
Gambar 4. 22 Classification_Report pada Alogritma Random Forest 70% dan 30%	49
Gambar 4. 23 Visualisasi Heatmap Confusion Matrix dari algoritma Random Forest dan 30%	49
Gambar 4. 24 Proses pemisahan data 80% dan 20%	50
Gambar 4. 25 Fungsi dari value_counts 80% dan 20%	51

Gambar 4. 26 TF - IDF untuk data Training dan Validation 80% dan 20%	51
Gambar 4. 27 Code pada CountVectorizer	52
Gambar 4. 28 Algoritma Naive Bayes dan Akurasi 80% dan 20%	53
Gambar 4. 29 Hasil Classification_Report pada Alogritma Naive Bayes 80% dan 20%	54
Gambar 4. 30 Visualisasi Heatmap pada Confusion Matrix dari algoritma Naive Bayes 80% dan 20%	54
Gambar 4. 31 Algoritma Decision Tree dan Akurasi 80% dan 20%	55
Gambar 4. 32 Hasil Classification_Report pada Alogritma Decision Tree 80% dan 20%	56
Gambar 4. 33 Visualisasi Heatmap Confusion Matrix dari algoritma Decision Tree 80% dan 20%	57
Gambar 4. 34 Algoritma Random Forest dan Akurasi 80% dan 20%	58
Gambar 4. 35 Hasil Classification_Report pada Alogritma Random Forest 80% dan 20%	58
Gambar 4. 36 Visualisasi Heatmap Confusion Matrix dari algoritma Random Forest 80% dan 20%	59
Gambar 4. 37 Code untuk mengunduh file SAV dari model	62
Gambar 4. 38 Code untuk mengunduh file TF-IDF	62
Gambar 4. 39 Code untuk menjalankan Website Streamlit dengan localhost	63
Gambar 4. 40 Tampilan dari Website	63
Gambar 4. 41 Tampilan Website pada sentimen Positif	64
Gambar 4. 42 Tampilan Website pada sentimen Negatif	64



DAFTAR LAMPIRAN

Lampiran A Turnitin 75
Lampiran B Formulir Konsultasi Skripsi..... 87



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR RUMUS

Rumus 2. 1 Perhitungan terhadap Precision	19
Rumus 2. 2 Perhitungan terhadap Recall	19
Rumus 2. 3 Perhitungan terhadap F1 - Score.....	19
Rumus 2. 4 Perhitungan terhadap Accuracy	19
Rumus 2. 5 Perhitungan terhadap Algoritma Naive Bayes.....	20



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA