

**PERBANDINGAN ALGORITMA MULTINOMIAL NAIVE BAYES DAN
LOGISTIC REGRESSION DALAM ANALISIS SENTIMEN TERHADAP
CHATGPT PADA MEDIA SOSIAL TWITTER**



SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

Muhammad Zhikri

00000033093

UMN

UNIVERSITAS

MULTIMEDIA

NUSANTARA

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA**

TANGERANG

2023

**PERBANDINGAN ALGORITMA MULTINOMIAL NAIVE BAYES DAN
LOGISTIC REGRESSION DALAM ANALISIS SENTIMEN TERHADAP
CHATGPT PADA MEDIA SOSIAL TWITTER**



SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

Muhammad Zhikri

0000033093

UMN

UNIVERSITAS

MULTIMEDIA

NUSANTARA

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA**

TANGERANG

2023

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Muhammad Zhikri
Nomor Induk Mahasiswa : 00000033093
Program Studi : Informatika

Skripsi dengan judul:

**Perbandingan Algoritma Multinomial Naive Bayes dan Logistic Regression
Dalam Analisis Sentimen Terhadap ChatGPT Pada Media Sosial Twitter**

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, 14 Juni 2023



(Muhammad Zhikri)

UMM
UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PENGESAHAN

Skripsi dengan judul

**PERBANDINGAN ALGORITMA MULTINOMIAL NAIVE BAYES DAN
LOGISTIC REGRESSION DALAM ANALISIS SENTIMEN TERHADAP
CHATGPT PADA MEDIA SOSIAL TWITTER**

oleh

Nama : Muhammad Zhikri
NIM : 00000033093
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Selasa, 27 Juni 2023

Pukul 15:00 s/d 17:00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut

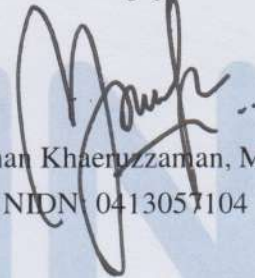
Ketua Sidang



(Moeljono Widjaja, B.Sc., M.Sc., Ph.D.)

NIDN: 0311106903

Penguji



(Yaman Khaeruzzaman, M.Sc)

NIDN: 0413057104

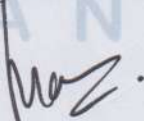
Pembimbing



(Wirawan Istiono, S.Kom., M.Kom.)

NIDN: 313048304

Ketua Program Studi Informatika,



(Marlinda Vasty Overbeek, S.Kom., M.Kom.)

NIDN: 0818038501

**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

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

Nama : Muhammad Zhikri
NIM : 00000033093
Program Studi : Informatika
Fakultas : Teknik dan Informatika
Jenis Karya : Skripsi

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

**PERBANDINGAN ALGORITMA MULTINOMIAL NAIVE BAYES DAN
LOGISTIC REGRESSION DALAM ANALISIS SENTIMEN TERHADAP
CHATGPT PADA MEDIA SOSIAL TWITTER**

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, 14 Juni 2023

Yang menyatakan

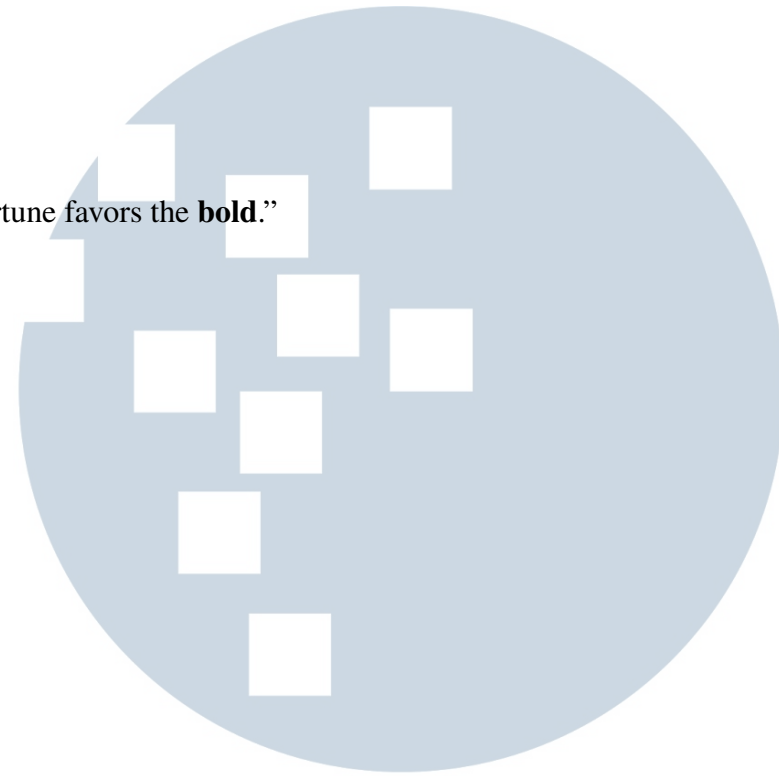


Muhammad Zhikri

UNIVERSITAS
MULTIMEDIA
NUSANTARA

Halaman Persembahan / Motto

"Fortune favors the **bold.**"



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

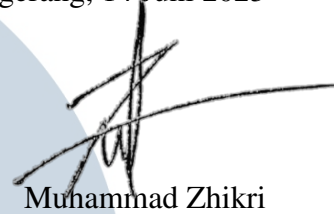
KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Perbandingan Algoritma Multinomial Naive Bayes dan Logistic Regression Dalam Analisis Sentimen Terhadap ChatGPT Pada Media Sosial Twitter dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

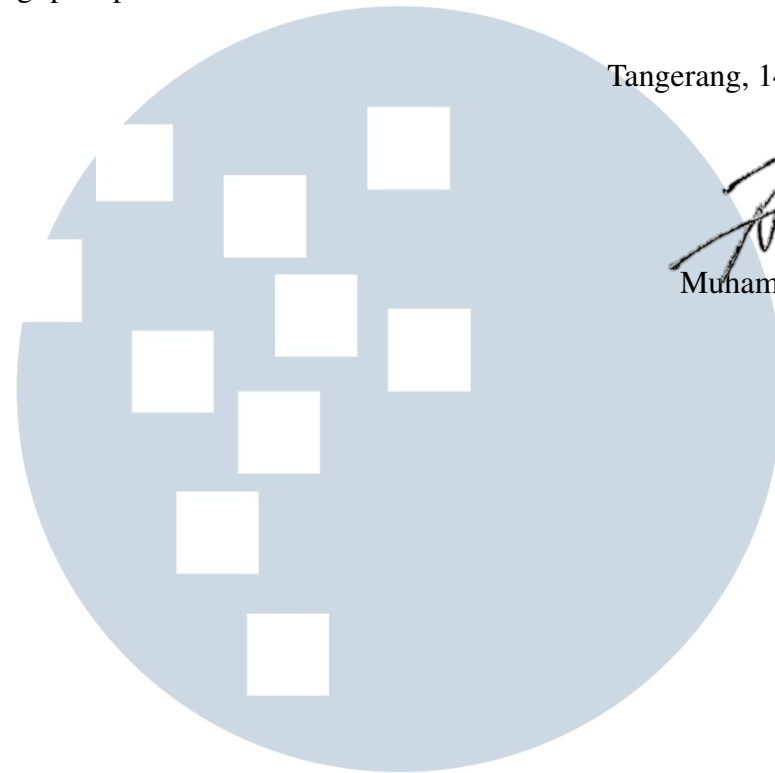
1. Bapak Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Marlinda Vasty Overbeek, S.Kom., M.Kom., selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak Wirawan Istiono, S.Kom., M.Kom., sebagai Pembimbing pertama yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya skripsi ini.
5. Diri saya sendiri Muhammad Zhikri yang telah berjuang menyelesaikan skripsi ini.
6. Orang Tua saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan skripsi ini.
7. Adik saya Novia Farika yang telah membantu menjumpai pembimbing dan kepala prodi di UMN untuk penandatanganan halaman persetujuan skripsi.
8. Teman-teman saya di UMN, Raisa, Aqsa, Dinda, Fathur, Danke, Riyan. Terima kasih atas kebersamaan dan kekompakkannya dalam menjalani hari-hari perkuliahan di UMN. Semoga kalian semua sukses selalu.
9. Teman-teman mabar saya, Cerni Amelia, Ravin, Sardi, Pamu, Chandra dan yang lainnya. Terima kasih atas dukungan, motivasi dan waktu luang kalian yang sering digunakan untuk menemani saya mabar.

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

Tangerang, 14 Juni 2023



Muhammad Zhikri



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

PERBANDINGAN ALGORITMA MULTINOMIAL NAIVE BAYES DAN LOGISTIC REGRESSION DALAM ANALISIS SENTIMEN TERHADAP CHATGPT PADA MEDIA SOSIAL TWITTER

Muhammad Zhikri

ABSTRAK

ChatGPT merupakan produk kecerdasan buatan yang dikembangkan dan diluncurkan oleh OpenAI. Sejak awal peluncurannya, ChatGPT terus menjadi perbincangan warga dunia karena robot kecerdasan buatan ini mampu memahami pertanyaan, memberikan jawaban yang relevan, dan berinteraksi dengan pengguna dengan tanggapan yang menyerupai percakapan manusia. Namun, karena kemampuan dan kepopulerannya ini lah membuat banyak orang mengekspresikan opini mereka di media sosial termasuk media sosial Twitter. Terdapat ratusan bahkan ribuan opini pengguna Twitter yang berkaitan dengan ChatGPT yang dapat dimanfaatkan untuk melakukan analisis sentimen pada opini pengguna Twitter di Indonesia. Penelitian ini akan membandingkan algoritma *multinomial naive bayes* dan *logistic regression* dalam sentimen analisis yang berkaitan dengan ChatGPT. Sebelum klasifikasi dilakukan, TF-IDF (*Term Frequency-Inverse Document Frequency*) dan SMOTE (*Synthetic Minority Over-sampling Technique*) akan diimplementasikan. Hasil pengujian menunjukkan hasil bahwa performa *logistic regression* lebih baik dibandingkan *multinomial naive bayes*, dengan skor akurasi sebesar 86%, presisi 84%, *recall* 100%, dan *f1-score* 91% dengan perbandingan *data train* 80% dan *data test* 20%. Pengujian dengan SMOTE juga menunjukkan bahwa *Logistic Regression* memiliki performa yang lebih baik dengan tingkat akurasi mencapai 91% dibandingkan *Multinomial Naive Bayes* yang memiliki akurasi 88%

Kata kunci: *Analisis Sentimen, Multinomial Naive Bayes, Logistic Regression, ChatGPT, TF-IDF*

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

Comparison of Multinomial Naive Bayes and Logistic Regression Algorithms in Sentiment Analysis of ChatGPT on Twitter Social Media

Muhammad Zhikri

ABSTRACT

ChatGPT is an artificial intelligence product developed and launched by OpenAI. Since its initial release, ChatGPT has been widely discussed worldwide due to its ability to understand questions, provide relevant answers, and interact with users in a conversational manner. However, due to its capabilities and popularity, many people have expressed their opinions on social media, including Twitter. There are hundreds, if not thousands, of Twitter users' opinions related to ChatGPT that can be utilized for sentiment analysis in Indonesian. This research aims to compare the performance of the multinomial naive Bayes and logistic regression algorithms in sentiment analysis related to ChatGPT. Before classification, TF-IDF (Term Frequency-Inverse Document Frequency) and SMOTE (Synthetic Minority Over-sampling Technique) will be implemented. The test results indicate that logistic regression outperforms multinomial naive Bayes with an accuracy score of 86%, precision of 84%, recall of 100%, and an F1-score of 91%, with an 80% training data and 20% testing data split. The testing with SMOTE also shows that Logistic Regression performs better with an accuracy rate of 91% compared to Multinomial Naive Bayes, which has an accuracy of 88%.

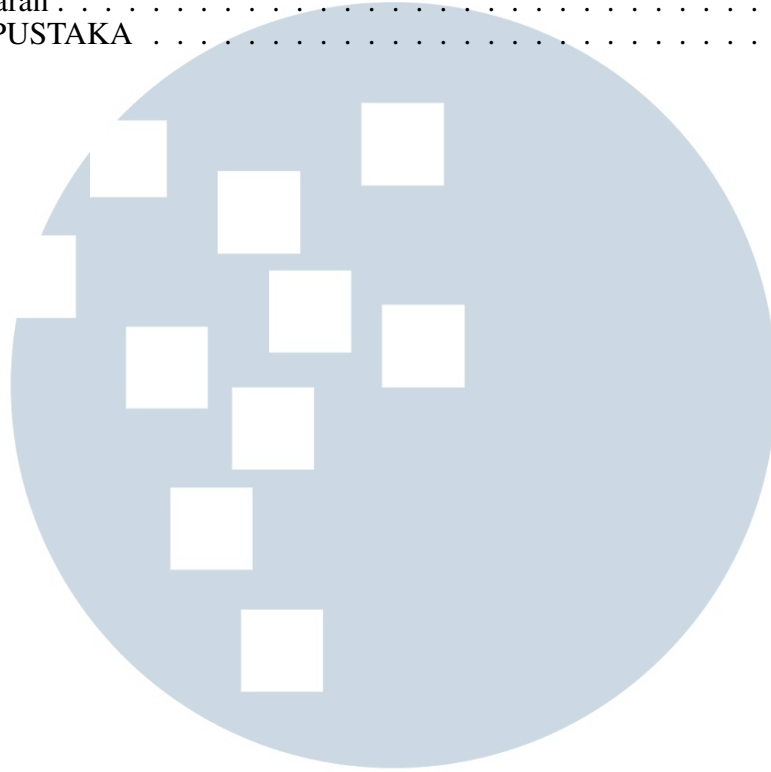
Keywords: *Sentiment Analysis, Multinomial Naive Bayes, Logistic Regression*
ChatGPT, TF-IDF



DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERSETUJUAN PUBLIKASI ILMIAH	v
HALAMAN PERSEMBAHAN/MOTO	vi
KATA PENGANTAR	vii
ABSTRAK	ix
ABSTRACT	x
DAFTAR ISI	xi
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv
DAFTAR KODE	xv
DAFTAR LAMPIRAN	xvi
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Permasalahan	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3
BAB 2 LANDASAN TEORI	5
2.1 Analisis Sentimen	5
2.2 Multinomial Naive Bayes	5
2.3 Logistic Regression	6
2.4 Synthetic Minority Oversampling Technique (SMOTE)	7
2.5 Labelling	8
2.6 Text Processing	8
2.6.1 Case Folding	8
2.6.2 Data Cleaning	8
2.6.3 Tokenization	8
2.6.4 Stopword Removal	9
2.6.5 Stemming	9
2.7 Term Frequency - Inverse Document Frequency	9
2.8 Matrix Confussion	10
BAB 3 METODOLOGI PENELITIAN	13
3.1 Gambaran Umum Objek Penelitian	13
3.2 Metode Penelitian	15
3.3 Perancangan Sistem	15
BAB 4 HASIL DAN DISKUSI	21
4.1 Spesifikasi Sistem	21
4.2 Import Library	21
4.3 Metode Pengumpulan Data	22
4.4 Labelling	23
4.5 Text Preprocessing	25
4.5.1 Penerapan TF-IDF Vectorizer	29
4.6 Uji Coba	31
4.6.1 Skenario Uji Coba	31

BAB 5	SIMPULAN DAN SARAN	38
5.1	Simpulan	38
5.2	Saran	38
DAFTAR PUSTAKA		39



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR GAMBAR

Gambar 3.1	Gambaran umum penelitian	14
Gambar 3.2	Proses <i>Text Preprocessing</i>	17
Gambar 3.3	Proses penerapan <i>TF-IDF</i>	18
Gambar 3.4	Proses penerapan <i>multinomial naive bayes</i>	19
Gambar 3.5	Proses penerapan <i>logistic regression</i>	20
Gambar 4.1	Hasil pemberian label	23
Gambar 4.2	Jumlah masing-masing label	24
Gambar 4.3	Potongan kode <i>import dataset</i> dan outputnya	25
Gambar 4.4	Hasil <i>case folding</i>	25
Gambar 4.5	Hasil <i>data cleansing</i>	26
Gambar 4.6	Hasil <i>tokenization</i>	27
Gambar 4.7	Hasil <i>stopwords removal</i>	27
Gambar 4.8	Hasil <i>stemming</i>	28
Gambar 4.9	Proses pembersihan data duplikat	29
Gambar 4.10	Hasil penghitungan TF-IDF	30
Gambar 4.11	Wordcloud visualiasi kemunculan kata	31
Gambar 4.12	Hasil pengujian Multinomial Naive Bayes 80:20 dengan SMOTE	36
Gambar 4.13	Hasil pengujian Multinomial Naive Bayes 70:30 dengan SMOTE	36
Gambar 4.14	Hasil pengujian Logistic Regression 80:20 dengan SMOTE	37
Gambar 4.15	Hasil pengujian Logistic Regression 70:30 dengan SMOTE	37



DAFTAR TABEL

Tabel 2.1	Confusion Matrix	11
Tabel 4.1	Hasil penghitungan TF-IDF kumulatif keseluruhan dokumen	30
Tabel 4.2	Hasil perbandingan rasio 70:30 dan 80:20 dengan <i>multinomial naive bayes</i>	32
Tabel 4.3	<i>Confussion matrix</i> MNB rasio perbandingan 70:30	32
Tabel 4.4	<i>Confussion matrix</i> MNB rasio perbandingan 80:20	32
Tabel 4.5	Hasil perbandingan rasio 70:30 dan 80:20 dengan <i>logistic</i> <i>regression</i>	34
Tabel 4.6	<i>Confussion matrix</i> LR rasio perbandingan 70:30	34
Tabel 4.7	<i>Confussion matrix</i> LR rasio perbandingan 80:20	34



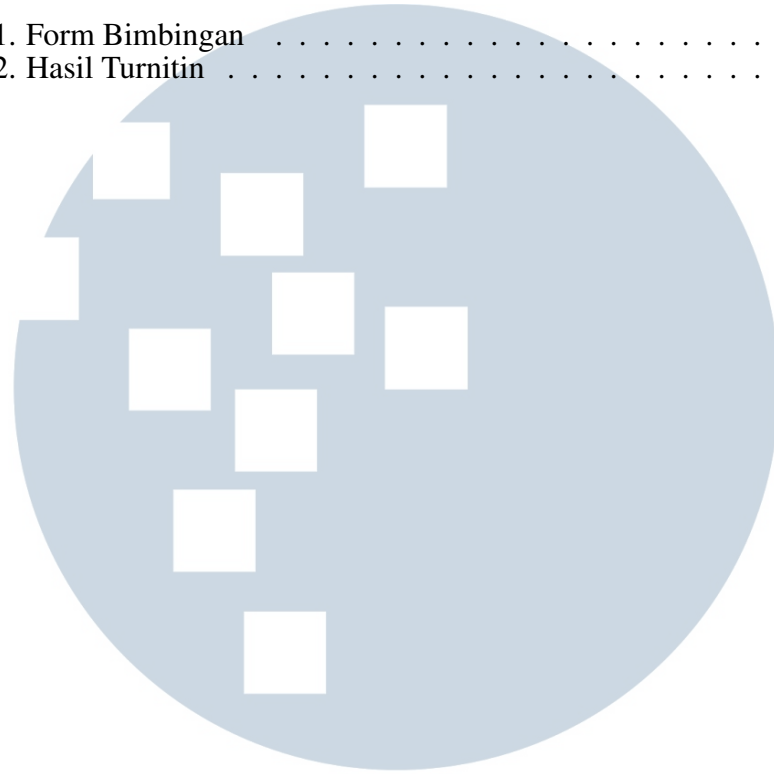
DAFTAR KODE

4.1	Potongan kode <i>import library</i>	21
4.2	Potongan kode <i>tweet crawling</i>	22
4.3	Potongan kode <i>slangword replacement</i>	23
4.4	Potongan kode untuk menghapus data yang berlabel netral	24
4.5	Potongan kode <i>case folding</i>	25
4.6	Potongan kode <i>data cleansing</i>	26
4.7	Potongan kode <i>tokenization</i>	27
4.8	Potongan kode <i>stopword removal</i>	27
4.9	Potongan kode <i>stemming</i>	28
4.10	Potongan kode implementasi TF-IDF	29



DAFTAR LAMPIRAN

Lampiran 1. Form Bimbingan	40
Lampiran 2. Hasil Turnitin	42



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA