

**IMPLEMENTASI ALGORITMA RANDOM FOREST DALAM ANALISIS  
SENTIMEN TERHADAP ULASAN PADA APLIKASI DIGITAL  
KORLANTAS POLRI**



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

**Afreza Danang Fathoni**

**0000034187**

UMN

UNIVERSITAS

MULTIMEDIA

NUSANTARA

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

**TANGERANG**

**2023**

**IMPLEMENTASI ALGORITMA RANDOM FOREST DALAM ANALISIS  
SENTIMEN TERHADAP ULASAN PADA APLIKASI DIGITAL  
KORLANTAS POLRI**



**SKRIPSI**

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

**Afreza Danang Fathoni**

**0000034187**

**UMN**

**UNIVERSITAS**

**MULTIMEDIA**

**NUSANTARA**

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

**TANGERANG**

**2023**

## HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Afreza Danang Fathoni  
Nomor Induk Mahasiswa : 00000034187  
Program Studi : Informatika

Skripsi dengan judul:

**Implementasi Algoritma Random Forest Dalam Analisis Sentimen Terhadap Ulasan Pada Aplikasi Digital Korlantas Polri**

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, 9 Oktober 2023



(Afreza Danang Fathoni)

## HALAMAN PENGESAHAN

Skripsi dengan judul

### **IMPLEMENTASI ALGORITMA RANDOM FOREST DALAM ANALISIS SENTIMEN TERHADAP ULASAN PADA APLIKASI DIGITAL KORLANTAS POLRI**

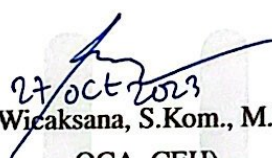
oleh

Nama : Afreza Danang Fathoni  
NIM : 00000034187  
Program Studi : Informatika  
Fakultas : Fakultas Teknik dan Informatika


Telah diujikan pada hari Selasa, 17 Oktober 2023  
Pukul 09.00 s/s 11.00 dan dinyatakan  
**LULUS**

Dengan susunan penguji sebagai berikut

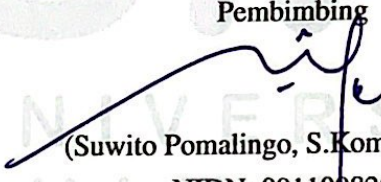
Ketua Sidang

  
(Arya Wicaksana, S.Kom., M.Eng.Sc.,  
OCA, CEH)  
NIDN: 0315109103

Penguji

  
(Dennis Gunawan, S.Kom., M.Sc.)  
NIDN: 0320059001

Pembimbing

  
(Suwito Pomalingo, S.Kom., M.Kom.)  
NIDN: 0911098201

Ketua Program Studi Informatika,

  
(Marlinda Vasty Overbeek, S.Kom., M.Kom.)  
NIDN: 0818038501

iii

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

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

Nama : Afreza Danang Fathoni  
NIM : 00000034187  
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:

**IMPLEMENTASI ALGORITMA RANDOM FOREST DALAM ANALISIS  
SENTIMEN TERHADAP ULASAN PADA APLIKASI DIGITAL  
KORLANTAS POLRI**

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.

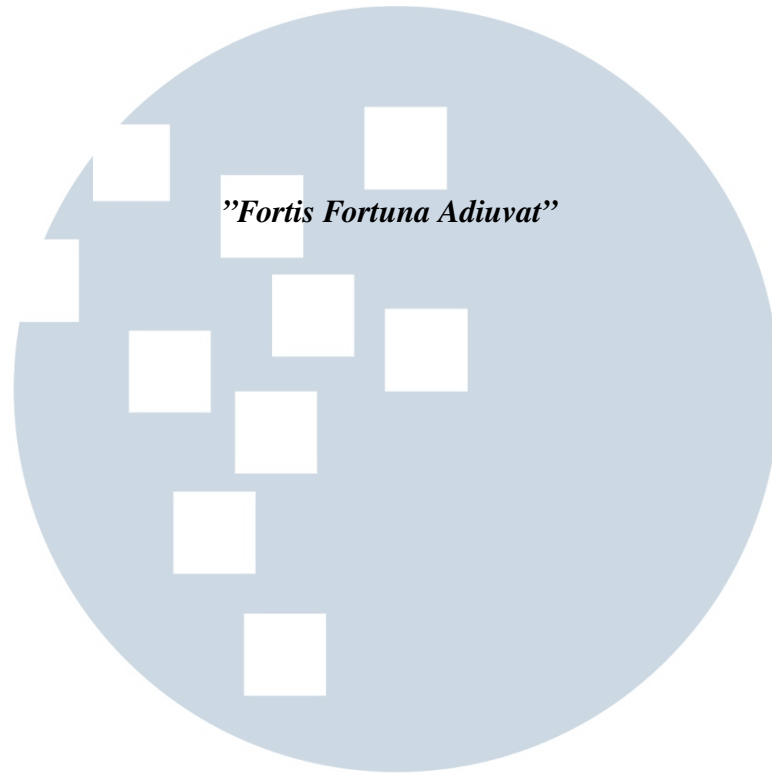
Tangerang, 9 Oktober 2023

Yang menyatakan



Afreza Danang Fathoni

**Halaman Persembahan / Motto**



Terence

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: Implementasi Algoritma Random Forest Dalam Analisis Sentimen Terhadap Ulasan Pada Aplikasi Digital Korlantas Polri 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. Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Marlinda Vasty Overbeek, S.Kom., M.Kom., selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Suwito Pomalingo, S.Kom., M.Kom., 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 skripsi ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 9 Oktober 2023



Afreza Danang Fathoni

**IMPLEMENTASI ALGORITMA RANDOM FOREST DALAM ANALISIS  
SENTIMEN TERHADAP ULASAN PADA APLIKASI DIGITAL  
KORLANTAS POLRI**

Afreza Danang Fathoni

**ABSTRAK**

Dalam era teknologi informasi dan komunikasi yang semakin canggih. Pelayanan publik khususnya dari POLRI, telah memanfaatkan aplikasi Digital Korlantas POLRI untuk menerima kritik dan saran dari masyarakat. Penelitian ini akan melakukan analisa sentimen terhadap ulasan pengguna terhadap aplikasi Digital Korlantas POLRI. Analisa sentimen akan menggunakan metode *Random Forest Classifier* dan juga menggunakan metode perbandingan dari peneliti terdahulu. Metode perbandingan yang digunakan yaitu *Support Vector Machine* dan *Multinomial Naive Bayes*. Sehingga mendapatkan metode terbaik dalam melakukan analisis sentimen khususnya pada data teks ulasan. Berdasarkan penelitian yang sudah dilakukan, *Random Forest Classifier* mendapatkan nilai akurasi 80,3%. Sedangkan pada metode perbandingannya yaitu *Support Vector Machine* mendapatkan nilai akurasi tertinggi yaitu 82,9% dan *Multinomial Naive Bayes* mendapatkan nilai akurasi 77,5%. Kemudian *model Random Forest Classifier* yang sudah dibuat dan dilatih sebelumnya, digunakan pada implementasi aplikasi *web* prediksi sentimen.

**Kata kunci:** Analisis Sentimen, Digital Korlantas POLRI, *Random Forest Classifier*, *Support Vector Machine*, *Multinomial Naive Bayes*, Ulasan Pengguna

U M I N  
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



# Implementation of Random Forest Algorithm in Sentiment Analysis of Reviews on Digital Korlantas Polri Application

Afreza Danang Fathoni

## ABSTRACT

In the era of increasingly sophisticated information and communication technology, public services, especially from the Indonesian National Police (POLRI), have utilized the Digital Korlantas POLRI application to receive criticisms and suggestions from the public. This research aims to conduct a sentiment analysis of user reviews of the Digital Korlantas POLRI application. The sentiment analysis will employ the Random Forest Classifier method and also use comparative methods from previous researchers. The comparative methods used are the Support Vector Machine and Multinomial Naive Bayes, aiming to find the best method for conducting sentiment analysis, especially on text review data. Based on the research conducted, the Random Forest Classifier achieved an accuracy score of 80.3%. In contrast, the comparative method, the Support Vector Machine, obtained the highest accuracy score of 82.9%, and the Multinomial Naive Bayes achieved an accuracy score of 77.5%. Subsequently, the previously created and trained Random Forest Classifier model was used in the implementation of a sentiment prediction web application.

**Keywords:** *Sentiment Analysis, Digital Korlantas POLRI, Random Forest Classifier, Support Vector Machine, Multinomial Naive Bayes, User Reviews*



## DAFTAR ISI

HALAMAN JUDUL . . . . .	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT . . . . .	ii
HALAMAN PENGESAHAN . . . . .	iii
HALAMAN PERSETUJUAN PUBLIKASI ILMIAH . . . . .	iv
HALAMAN PERSEMBAHAN/MOTO . . . . .	v
KATA PENGANTAR . . . . .	vi
ABSTRAK . . . . .	vii
ABSTRACT . . . . .	viii
DAFTAR ISI . . . . .	ix
DAFTAR GAMBAR . . . . .	xi
DAFTAR TABEL . . . . .	xiii
DAFTAR LAMPIRAN . . . . .	xiv
<b>BAB 1 PENDAHULUAN . . . . .</b>	<b>1</b>
1.1 Latar Belakang Masalah . . . . .	1
1.2 Rumusan Masalah . . . . .	3
1.3 Batasan Permasalahan . . . . .	3
1.4 Tujuan Penelitian . . . . .	3
1.5 Manfaat Penelitian . . . . .	4
1.6 Sistematika Penulisan . . . . .	4
<b>BAB 2 LANDASAN TEORI . . . . .</b>	<b>6</b>
2.1 Analisis Sentimen . . . . .	6
2.1.1 <i>Data Collection</i> . . . . .	6
2.1.2 <i>Text Preparation</i> . . . . .	6
2.1.3 <i>Sentiment Detection</i> . . . . .	7
2.1.4 <i>Sentiment Classification</i> . . . . .	7
2.2 <i>Decision Tree</i> . . . . .	7
2.3 <i>Random Forest</i> . . . . .	8
2.4 <i>Text Preprocessing</i> . . . . .	10
2.4.1 <i>Case Folding</i> . . . . .	10
2.4.2 <i>Tokenizing</i> . . . . .	11
2.4.3 <i>Filtering</i> . . . . .	11
2.4.4 <i>Stemming</i> . . . . .	12
2.4.5 <i>SMOTE</i> . . . . .	14
2.5 <i>TF-IDF</i> . . . . .	14
2.6 Evaluasi Klasifikasi . . . . .	15
2.6.1 <i>K-Fold Cross Validation</i> . . . . .	15
2.6.2 <i>Confusion Matrix</i> . . . . .	16
2.6.3 <i>Accuracy</i> . . . . .	17
2.6.4 <i>Precision</i> . . . . .	18
2.6.5 <i>Recall</i> . . . . .	18
2.6.6 <i>F1-Score</i> . . . . .	18
2.7 <i>Streamlit</i> . . . . .	18
<b>BAB 3 METODOLOGI PENELITIAN . . . . .</b>	<b>22</b>
3.1 Identifikasi Masalah . . . . .	23
3.2 Studi literatur . . . . .	23
3.3 Pengumpulan Data . . . . .	23
3.4 Text Pre-processing . . . . .	23
3.5 Klasifikasi . . . . .	23

3.6	Evaluasi . . . . .	24
3.7	Implementasi Web App . . . . .	24
3.8	Perancangan Sistem . . . . .	24
3.8.1	Aplikasi Training dan Evaluation . . . . .	25
3.8.2	Aplikasi Web Prediksi Sentimen . . . . .	33
3.9	Rancangan Tampilan Antarmuka . . . . .	34
BAB 4	HASIL DAN DISKUSI . . . . .	36
4.1	Kebutuhan Sistem . . . . .	36
4.2	Pengumpulan Data . . . . .	36
4.3	Pelabelan Data . . . . .	37
4.4	Text Pre-Processing . . . . .	38
4.4.1	Case Folding . . . . .	38
4.4.2	Tokenizing . . . . .	39
4.4.3	Filtering . . . . .	39
4.4.4	Stemming . . . . .	40
4.4.5	SMOTE . . . . .	41
4.5	TF-IDF . . . . .	42
4.6	Word Cloud . . . . .	42
4.7	Training dan Evaluation Model . . . . .	45
4.7.1	K-Fold Cross Validation . . . . .	46
4.7.2	Confusion Matrix . . . . .	46
4.7.3	Evaluation Matrix . . . . .	47
4.8	Hasil dan Diskusi . . . . .	48
BAB 5	SIMPULAN DAN SARAN . . . . .	49
5.1	Simpulan . . . . .	49
5.2	Saran . . . . .	49
DAFTAR PUSTAKA	. . . . .	51



## DAFTAR GAMBAR

Gambar 2.1	Visualisasi Algoritma Random Forest . . . . .	9
Gambar 3.1	Metodologi Penelitian . . . . .	22
Gambar 3.2	Flowchart Aplikasi Training dan Evaluation . . . . .	25
Gambar 3.3	Flowchart Pada Proses Scrap dan Label Data . . . . .	26
Gambar 3.4	Flowchart Pada Proses Text Preprocessing . . . . .	27
Gambar 3.5	Flowchart Pada Proses Punctuation Removal . . . . .	28
Gambar 3.6	Flowchart Pada Proses Tokenizing . . . . .	29
Gambar 3.7	Flowchart Pada Proses Filtering . . . . .	30
Gambar 3.8	Flowchart Pada Proses Stemming . . . . .	31
Gambar 3.9	Flowchart Pada Proses Training dan Evaluation Model . . . . .	32
Gambar 3.10	Flowchart Aplikasi Prediksi Sentimen . . . . .	33
Gambar 3.11	Rancangan Antar Muka Website Prediksi Sentimen . . . . .	34
Gambar 3.12	Rancangan Antar Muka Website Prediksi Sentimen . . . . .	35
Gambar 3.13	Rancangan Antar Muka Website Prediksi Sentimen . . . . .	35
Gambar 4.1	Distribusi Kelas Sentimen Sebelum Dan Setelah SMOTE . . . . .	41
Gambar 4.2	Hasil Visualisasi Word Cloud Dari Teks Yang Memiliki Sentimen Positif Dalam Data Frame . . . . .	43
Gambar 4.3	Hasil Visualisasi Word Cloud Dari Teks Yang Memiliki Sentimen Negatif Dalam Data Frame . . . . .	44
Gambar 4.4	Hasil Visualisasi Word Cloud Dari Teks Yang Memiliki Sentimen Netral Dalam Data Frame . . . . .	45
Gambar 5.1	Proses Scraping Data Ulasan Dari Play Store . . . . .	58
Gambar 5.2	Proses Filter Data Ulasan Berdasarkan Tanggal Dibuat . . . . .	58
Gambar 5.3	Proses Memisahkan Kolom Tanggal Bulan Dan Tahun . . . . .	58
Gambar 5.4	Proses Menghilangkan Kolom At . . . . .	59
Gambar 5.5	Proses Menyimpan Data Frame Ke Dalam Bentuk CSV . . . . .	59
Gambar 5.6	Proses Pelabelan Sentimen Data CSV . . . . .	59
Gambar 5.7	Data Ulasan Yang Disimpan Di Repositori GitHub . . . . .	60
Gambar 5.8	Potongan Code Import Dataset Dari Repositori GitHub . . . . .	60
Gambar 5.9	Potongan Code Import Library Yang Digunakan Untuk Text Pre-Processing . . . . .	60
Gambar 5.10	Potongan Code Pada Proses Pembersihan Teks . . . . .	61
Gambar 5.11	Potongan Code Pada Proses Tokenizing . . . . .	61
Gambar 5.12	Potongan Code Pada Proses Stop Word Removal . . . . .	61
Gambar 5.13	Potongan Code Pada Proses Stemming . . . . .	62
Gambar 5.14	Potongan Code Pada Proses Import Library Word Cloud . . . . .	62
Gambar 5.15	Hasil Confusion Matrix Dari Model Random Forest Classifier . . . . .	62
Gambar 5.16	Hasil Confusion Matrix Dari Model Support Vector Machine . . . . .	63
Gambar 5.17	Hasil Confusion Matrix Dari Model Multinomial Naive Bayes . . . . .	63
Gambar 5.18	Potongan Code Pada Proses Visualisasi Word Cloud Dari Teks Yang Memiliki Sentimen Positif Dalam Data Frame . . . . .	64
Gambar 5.19	Potongan Code Pada Proses Visualisasi Word Cloud Dari Teks Yang Memiliki Sentimen Negatif Dalam Data Frame . . . . .	64
Gambar 5.20	Potongan Code Pada Proses Visualisasi Word Cloud Dari Teks Yang Memiliki Sentimen Netral Dalam Data Frame . . . . .	64

Gambar 5.21	Potongan Code Pada Proses Import Library Dan Konfigurasi Pada Proses Training Dan Evaluasi Model . . .	64
Gambar 5.22	Potongan Code Pada Proses Vektorisasi TF-IDF Pada Data Dan Simpan Model TF-IDF . . . . .	64
Gambar 5.23	Potongan Code Pada Proses SMOTE . . . . .	65
Gambar 5.24	Potongan Command Terminal Pada Proses Instalasi Kerangka Kerja Streamlit . . . . .	65
Gambar 5.25	Potongan Code Import Library Dan Dependencies Untuk Web App . . . . .	65
Gambar 5.26	Potongan Code Pada Proses Pengambilan Model . . . . .	66
Gambar 5.27	Potongan Code Dari Fungsi Analisis Sentimen . . . . .	67
Gambar 5.28	Potongan Code Dari Fungsi Analisis Sentimen (2) . . . . .	68
Gambar 5.29	Tampilan Antar Muka Web App Dari Sentimen Positif . . .	69
Gambar 5.30	Tampilan Antar Muka Web App Dari Sentimen Netral . . .	69
Gambar 5.31	Tampilan Antar Muka Web App Dari Sentimen Negatif . . .	70
Gambar 5.32	Tampilan Antar Muka Web App Dari Notifikasi Alert . . .	70



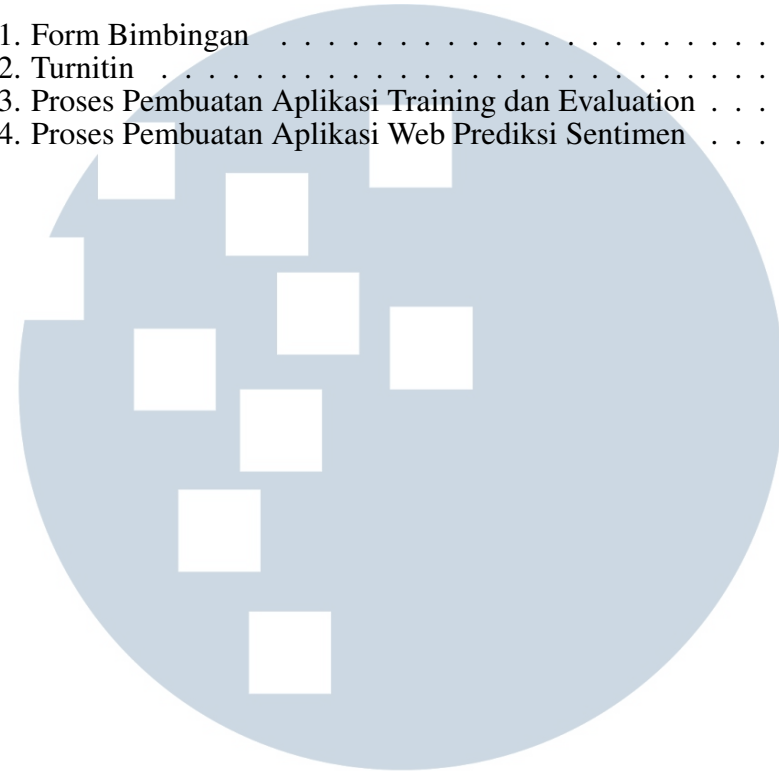
## DAFTAR TABEL

Tabel 2.1	Tabel proses case folding . . . . .	11
Tabel 2.2	Tabel proses tokenizing . . . . .	11
Tabel 2.3	Tabel proses Filtering . . . . .	12
Tabel 2.4	Tabel proses Stemming . . . . .	12
Tabel 2.5	Tabel Kombinasi Awalan Akhiran Yang Tidak Diijinkan . .	13
Tabel 2.6	Tabel 5 Fold Cross Validation . . . . .	16
Tabel 2.7	Tabel Confusion Matrix . . . . .	17
Tabel 2.8	Tabel Tinjauan Literatur . . . . .	19
Tabel 2.8	Tabel Tinjauan Literatur (lanjutan) . . . . .	20
Tabel 2.8	Tabel Tinjauan Literatur (lanjutan) . . . . .	21
Tabel 4.1	Tabel Sampel Data Ulasan Digital Korlantas Polri . . . . .	37
Tabel 4.2	Tabel Label Data Ulasan Digital Korlantas Polri . . . . .	38
Tabel 4.3	Tabel Sampel Data Saat Proses Tokenizing . . . . .	39
Tabel 4.4	Tabel Sampel Data Saat Proses Filtering . . . . .	40
Tabel 4.5	Tabel Sampel Data Saat Proses Stemming . . . . .	40
Tabel 4.6	Tabel Sampel Data Hasil TF-IDF . . . . .	42
Tabel 4.7	Tabel Confusion Matrix Dari Beberapa Model . . . . .	47
Tabel 4.8	Tabel Nilai Dari Precision, Recall, F1-Score, Accuracy . .	47



## DAFTAR LAMPIRAN

Lampiran 1. Form Bimbingan . . . . .	55
Lampiran 2. Turnitin . . . . .	57
Lampiran 3. Proses Pembuatan Aplikasi Training dan Evaluation . . . . .	58
Lampiran 4. Proses Pembuatan Aplikasi Web Prediksi Sentimen . . . . .	66



**UMMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA