

**Analisis Sentimen Terhadap Ulasan *IDN News* Menggunakan  
Komparasi Random Forest, Decision Tree dan K-Nearest  
Neighbor**



**UMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

**SKRIPSI**

**Rico Tanuwijaya Saputra**

**0000047553**

**PROGRAM STUDI SISTEM INFORMASI  
TEKNIK DAN INFORMATIKA  
UNIVERSITAS MULTIMEDIA NUSANTARA  
TANGERANG  
2024**

**Analisis Sentimen Terhadap Ulasan *IDN News* Menggunakan  
Komparasi Random Forest, Decision Tree dan K-Nearest**

**Neighbor**



**UMN**

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

**SKRIPSI**

Diajukan sebagai Salah Satu Syarat untuk Memperoleh  
Gelar Sarjana Komputer (S.Kom)

**Rico Tanuwijaya Saputra**

**00000047553**

**PROGRAM STUDI SISTEM INFORMASI**

**TEKNIK DAN INFORMATIKA**

**UNIVERSITAS MULTIMEDIA NUSANTARA**

**TANGERANG**

**2024**

## HALAMAN PERNYATAAN TIDAK PLAGIAT

### HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Rico Tanuwijaya Saputra

Nomor Induk Mahasiswa : 00000047553

Program studi : Sistem Informasi

Skripsi dengan judul:

**Analisis Sentimen Terhadap Ulasan IDN News Menggunakan Komparasi Random Forest, Decision Tree dan K-Nearest Neighbor**

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, 21 Mei 2024



Rico Tanuwijaya Saputra

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 PENGESAHAN

### HALAMAN PENGESAHAN

Skripsi dengan judul

Analisis Sentimen Terhadap Ulasan IDN News Menggunakan Komparasi  
Random Forest, Decision Tree dan K-Nearest Neighbor

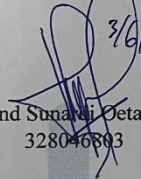
Oleh

Nama : Rico Tanuwijaya Saputra  
NIM : 00000047553  
Program Studi : Sistem Informasi  
Fakultas : Teknik dan Infomatika

Telah diujikan pada hari Selasa, 21 Mei 2024  
Pukul 13.00 s.d 15.00 dan dinyatakan  
LULUS

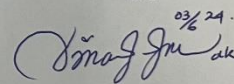
Dengan susunan penguji sebagai berikut.

Ketua Sidang

  
3/6/2024

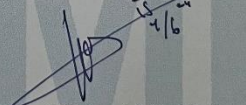
Ir. Raymond Sunardi Detama, M.CIS  
328046803

Penguji

  
03/6/24


Dinar Ajeng Kristiyanti, S.Kom, M.Kom.  
0330128801

Pembimbing

  
4/6/24

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

Ketua Program Studi Sistem Informasi

  
4/6/24

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

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 PUBLIKASI KARYA ILMIAH MAHASISWA

Yang bertanda tangan di bawah ini:

Nama : Rico Tanuwijaya Saputra

NIM : 00000047553

Program Studi : Sistem Infomasi

Jenjang : S1

Judul Karya Ilmiah :

**Analisis Sentimen Terhadap Ulasan IDN News Menggunakan  
Komparasi Random Forest, Decision Tree dan K-Nearest Neighbor**

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, 4 Juni 2024



Rico Tanuwijaya Saputra

U M 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

## KATA PENGANTAR

Puji Syukur atas selesainya penulisan Laporan skripsi ini dengan judul: “Analisis Sentimen Terhadap Ulasan IDN News Menggunakan Komparasi Random Forest, Decision Tree dan K-Nearest Neighbor” dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Jurusan Sistem Informasi Pada 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, selaku Dekan Fakultas Teknik Dan Informatika Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Bapak Rudi Sutomo, S.Kom., Msi., M.Kom. sebagai Pembimbing yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya skripsi ini.
5. Keluarga yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan laporan Skripsi.
6. Marcellius Cahyadi, Vincentius Juan Antonio, Yeremia Steven Putra Mareby, Gregorius Dinov, Henry Wirawan Tamrin teman – teman yang telah membantu selama kuliah dan penelitian.

Semoga laporan ini dapat bermanfaat untuk dijadikan referensi atau informasi bagi pembaca

Tangerang, 21 Mei 2024



Rico Tanuwijaya Saputra



# Analisis Sentimen Terhadap Ulasan IDN News Menggunakan Komparasi Random Forest, Decision Tree dan K-Nearest

**Neighbor**

Rico Tanuwijaya Saputra

## ABSTRAK

Aplikasi IDN news merupakan aplikasi yang dimiliki oleh IDN media. Aplikasi ini telah dirilis sejak 8 Juni 2020. IDN news telah menjadi aplikasi berita paling populer pada kategori “berita & majalah”. Aplikasi IDN telah diunduh lebih dari 1 juta kali dengan rating 4.7. Dalam sebuah ulasan pengguna terdapat ulasan yang positif yang berisi kepuasan pengguna dan ulasan negatif yang merupakan ketidakpuasan pengguna. Penelitian ini dilakukan untuk mengetahui sentimen pengguna aplikasi IDN news cenderung ke arah positif atau negatif. Proses klasifikasi akan menggunakan tiga algoritma berbeda yaitu Random Forest, Decision Tree, dan KNN. Penelitian ini akan menggunakan tiga skenario pembagian train dan test data yaitu 80:20, 70:30, dan 60:40. Penggunaan algoritma yang berbeda dilakukan untuk mengetahui performa algoritma yang terbaik dalam melakukan klasifikasi. Pengukuran algoritma didasarkan pada hasil *confusion matrix*. Hasil pengujian yang telah dilakukan mendapatkan hasil bahwa sentimen positif sebanyak 559 ulasan sedangkan sentimen negatif sebanyak 396 ulasan. Berdasarkan hasil pelabelan berdasarkan score/rating. Berdasarkan hasil visualisasi wordcloud ulasan/kata yang paling sering disebut oleh pengguna yaitu live, bagus, aplikasi. Perbandingan algoritma Random Forest, Decision Tree, KNN telah mendapatkan hasil bahwa algoritma Random Forest merupakan algoritma yang terbaik dengan menggunakan dataset review aplikasi IDN. Dengan hasil akurasi 82%, *Precision* 82%, *Recall* 89%, *F1\_Score* 86%.

Kata kunci: *decision tree, k-nearest neighbor, random forest.*

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 of IDN News Reviews Using Random Forest, Decision Tree and K-Nearest Neighbor Comparisons

Rico Tanuwijaya Saputra

## **ABSTRACT (English)**

*IDN news application is an application owned by IDN media. This application has been released since June 8, 2020. IDN news has become the most popular news app in the “news & magazines” category. The IDN app has been downloaded more than 1 million times with a rating of 4.7. in a user review there are positive reviews that contain user satisfaction and negative reviews which are user dissatisfaction. This research was conducted to find out the sentiment of IDN news application users tending towards positive or negative. This research will perform labeling to IDN News application user reviews. Labeling is based on the rating that has been given. Ratings 1-3 will be given a negative label while labels 4-5 will be given a positive label. The classification process will use three different algorithms namely Random Forest, Decision Tree, and KNN. This research will use three scenarios of train and test data distribution, namely 80:20, 70:30, and 60:40. The use of different algorithms is done to find out the best algorithm performance in classification. Algorithm measurement is based on confusion matrix results. The test results that have been carried out get the results that positive sentiment is 559 reviews while negative sentiment is 396 reviews. Based on the labeling results based on score/rating. Based on the results of wordcloud visualization, the reviews/words most often mentioned by users are live, good, application. Comparison of Random Forest, Decision Tree, KNN algorithms has found that the Random Forest algorithm is the best algorithm using the IDN application review dataset. With the results of 82% accuracy, Precision 82%, Recall 89%, F1\_Score 86%.*

**Keywords:** *decision Tree, k-nearest neighbors, random forest.*

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

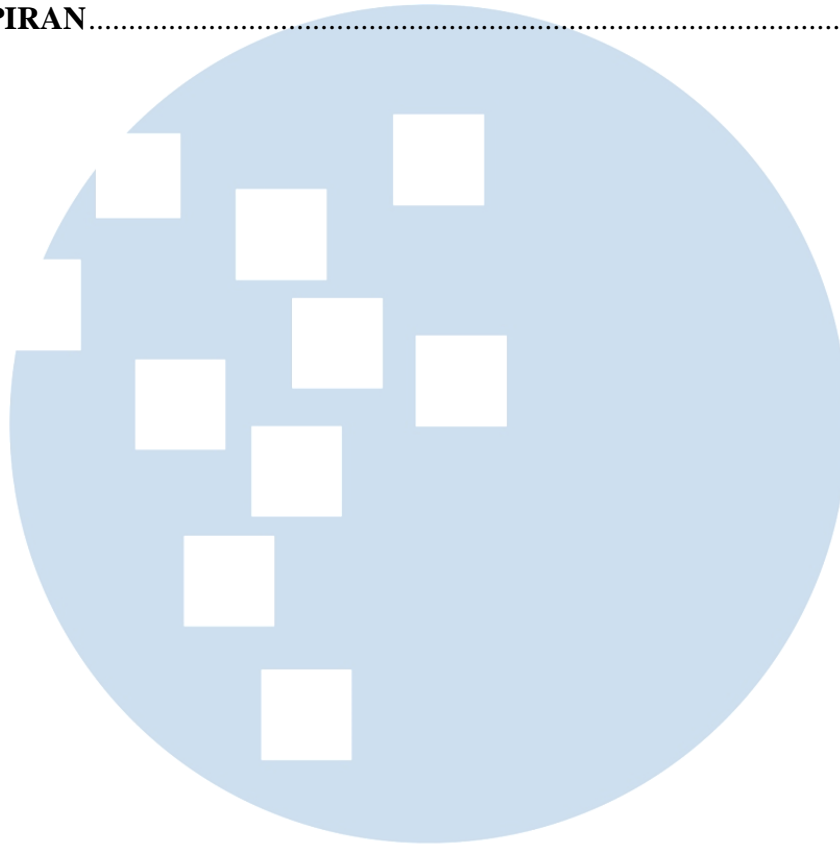


## DAFTAR ISI

<b>HALAMAN PERNYATAAN TIDAK PLAGIAT</b> .....	ii
<b>HALAMAN PENGESAHAN</b> .....	iii
<b>HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH MAHASISWA</b> .....	iv
<b>KATA PENGANTAR</b> .....	v
<b>ABSTRAK</b> .....	vi
<b>ABSTRACT (English)</b> .....	vii
<b>DAFTAR ISI</b> .....	viii
<b>DAFTAR TABEL</b> .....	xi
<b>DAFTAR GAMBAR</b> .....	xii
<b>DAFTAR RUMUS</b> .....	xiv
<b>DAFTAR LAMPIRAN</b> .....	xv
<b>BAB I PENDAHULUAN</b> .....	1
<b>1.1 Latar Belakang</b> .....	1
<b>1.2 Rumusan Masalah</b> .....	5
<b>1.3 Batasan Masalah</b> .....	5
<b>1.4 Tujuan dan Manfaat Penelitian</b> .....	5
<b>1.4.1 Tujuan Penelitian</b> .....	5
<b>1.4.2 Manfaat Penelitian</b> .....	6
<b>1.5 Sistematika Penulisan</b> .....	6
<b>BAB II LANDASAN TEORI</b> .....	8
<b>2.1 Penelitian Terdahulu</b> .....	8
<b>2.2 Analisis Sentimen</b> .....	13
<b>2.3 Google Play store</b> .....	14
<b>2.4 Random Forest</b> .....	15
<b>2.5 Decision Tree</b> .....	17
<b>2.6 K Nearest Neighbors</b> .....	18
<b>2.7 Jupyter Notebook</b> .....	18
<b>2.8 Anaconda Navigator</b> .....	19
<b>2.9 Visual Studio</b> .....	19

2.10	<i>Scraping</i> .....	19
2.11	<i>Data Preprocessing</i> .....	20
2.12	<i>Term Frequency - Inverse Document Frequency (TF-IDF)</i> .....	21
2.13	<i>Python</i> .....	21
2.14	<i>Confusion Matrix</i> .....	22
2.15	<i>Chi-Square</i> .....	24
2.16	<i>Select K best</i> .....	24
2.17	<i>N-Gram</i> .....	24
2.18	<i>Streamlit</i> .....	25
2.19	<i>CRISP-DM</i> .....	25
<b>BAB III METODOLOGI PENELITIAN</b> .....		27
3.1	<b>Gambaran Umum Objek Penelitian</b> .....	27
3.2	<b>Alur Penelitian</b> .....	27
3.3	<b>Metode Penelitian</b> .....	29
3.4	<b>Teknik Pengumpulan data</b> .....	30
3.5	<b>Teknik Analisis Data</b> .....	31
<b>BAB IV ANALISIS DAN HASIL PENELITIAN</b> .....		33
4.1	<b>Busines Understanding</b> .....	33
4.2	<b>Pengumpulan Data</b> .....	33
4.3	<b>Data Preparation</b> .....	37
4.3.1	<b>Cleaning data</b> .....	37
4.3.2	<b>Data Processing</b> .....	46
4.3.3	<b>Pelabelan</b> .....	49
4.4	<b>Modeling</b> .....	50
4.4.1	<b>Uji Skenario Train dan Test Data 80:20</b> .....	51
4.4.2	<b>Uji Skenario Train dan Test Data 70:30</b> .....	60
4.4.3	<b>Uji Skenario Train dan Test Data 60:40</b> .....	69
4.5	<b>Evaluation</b> .....	77
4.6	<b>Deployment</b> .....	81
<b>BAB V SIMPULAN DAN SARAN</b> .....		83
5.1	<b>Simpulan</b> .....	83
5.2	<b>Saran</b> .....	84

<b>DAFTAR PUSTAKA</b> .....	85
<b>LAMPIRAN</b> .....	90



**UMMN**

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu .....	8
Tabel 2. 2 TP, FP, FN, TN .....	22
Tabel 2. 3 Rumus Confusion Matrix.....	23
Tabel 3. 1 Kelebihan dan Kekurangan Algoritma .....	30
Tabel 4. 1 Hasil Perubahan Huruf kapital.....	39
Tabel 4. 2 Hasil Penghapusan URL .....	39
Tabel 4. 3 Hasil Penghapusan angka.....	39
Tabel 4. 4 Hasil Penghapusan Simbol.....	40
Tabel 4. 5 Hasil Menghapus Tanda Baca.....	40
Tabel 4. 6 Hasil Menghapus Spasi Kosong .....	40
Tabel 4. 7 Hasil Normalization .....	41
Tabel 4. 8 Hasil Stopwords .....	42
Tabel 4. 9 Hasil Stemming.....	43
Tabel 4. 10 Hasil Penelitian Skenario 80:20.....	78
Tabel 4. 11 Hasil Penelitian Skenario 70:30.....	78
Tabel 4. 12 Hasil penelitian skenario 60:40.....	79
Tabel 4. 13 Hasil perbandingan penelitian terdahulu.....	80



## DAFTAR GAMBAR

Gambar 1. 1 Aplikasi IDN pada google playstore. ....	1
Gambar 1. 2 Aplikasi Berita paling populer .....	2
Gambar 2. 1 Tampilan Googleplaystore Pada Website .....	15
Gambar 2. 2 Cara kerja algoritma Random Forest .....	15
Gambar 3. 1 Alur Penelitian.....	27
Gambar 3. 2 CRISP-DM .....	31
Gambar 4. 1 Instal dan Import library scraping. ....	33
Gambar 4. 2 Scraping data. ....	34
Gambar 4. 3 Data frame .....	34
Gambar 4. 4 Jumlah data yang didapatkan .....	34
Gambar 4. 5 data hasil Scraping. ....	35
Gambar 4. 6 Menghilangkan kolom.....	36
Gambar 4. 7 Simpan data dengan format CSV .....	36
Gambar 4. 8 Tampilan data dengan format csv .....	37
Gambar 4. 9 Import Library .....	38
Gambar 4. 10 Proses Casefolding .....	38
Gambar 4. 11 Proses Normalization .....	41
Gambar 4. 12 Proses Stopwords .....	42
Gambar 4. 13 Proses Stemming .....	43
Gambar 4. 14 Menggabungkan seluruh proses cleaning. ....	43
Gambar 4. 15 Cleaning seluruh ulasan pada data .....	44
Gambar 4. 16 Info data setelah clean .....	44
Gambar 4. 17 Wordcloud sentimen positif .....	45
Gambar 4. 18 Wordcloud Sentimen Negatif.....	45
Gambar 4. 19 Define x dan y .....	46
Gambar 4. 20 Proses TF-IDF .....	46
Gambar 4. 21 Merubah hasil TF-IDF menjadi tabular .....	47
Gambar 4. 22 Hasil model tabular .....	47
Gambar 4. 23 Proses Filterisasi.....	47
Gambar 4. 24 Hasil filterisasi .....	48
Gambar 4. 25 Menyipan proses .....	48
Gambar 4. 26 Pelabelan .....	49
Gambar 4. 27 Hasil pelabelan .....	49
Gambar 4. 28 Visualisasi Sentimen Positif dan Negatif .....	50
Gambar 4. 29 Import Library .....	51
Gambar 4. 30 Pembagian data train and test.....	51
Gambar 4. 31 Model Random Forest .....	52
Gambar 4. 32 Confusion Matrix Random Forest.....	52
Gambar 4. 33 Heat map confusion matrix .....	53
Gambar 4. 34 Model Decision Tree .....	54
Gambar 4. 35 Confusion Matrix Decision Tree.....	54

Gambar 4. 36 Heat Hasil Confusion Matrix Decision Tree.....	55
Gambar 4. 37 Model KNN K=5.....	56
Gambar 4. 38 Hasil Model KNN K=5 .....	57
Gambar 4. 39 Model KNN K=7.....	57
Gambar 4. 40 Confusion Matrix KNN K=7 .....	58
Gambar 4. 41 Heatmap Confusion Matrix KNN K=7 .....	59
Gambar 4. 42 Pembagian Data Train dan Test 70:30 .....	60
Gambar 4. 43 Model Random Forest 70:30.....	60
Gambar 4. 44 Hasil Confusion Matrix Random Forest 70:30 .....	61
Gambar 4. 45 Heatmap Random Forest 70:30.....	61
Gambar 4. 46 Model Decision Tree 70:30.....	62
Gambar 4. 47 Hasil Confusion Matrix Decision Tree 70:30 .....	63
Gambar 4. 48 Heatmap Confusion Matrix Decision Tree 70:30 .....	64
Gambar 4. 49 Model KNN 70 :30 K=5 .....	65
Gambar 4. 50 Hasil KNN 70:30 K=5 .....	66
Gambar 4. 51 Model KNN 70:30 K=7 .....	67
Gambar 4. 52 Confusion Matrix KNN 70:30 K=7 .....	67
Gambar 4. 53 Tabel Confusion Matrix KNN 70:30 K=7 .....	68
Gambar 4. 54 Split Train dan Test 60:40.....	69
Gambar 4. 55 Model Random Forest 60:40.....	69
Gambar 4. 56 Confusion Matrix Random Forest 60:40.....	70
Gambar 4. 57 Heatmap Random Forest 60:40.....	71
Gambar 4. 58 Pemodelan Decision Tree 60:40 .....	72
Gambar 4. 59 Confusion Matrix Decision Tree 60:40.....	72
Gambar 4. 60 Heatmap Decision Tree 60:40.....	73
Gambar 4. 61 Model KNN 60:40 K=5 .....	74
Gambar 4. 62 Hasil KNN 60:40 K=5 .....	75
Gambar 4. 63 Model KNN 60:40 K=7 .....	75
Gambar 4. 64 Confusion Matrix KNN 60:40 K=7 .....	76
Gambar 4. 65 Heatmap KNN 60:40 K=7 .....	76
Gambar 4. 66 Uji coba prediksi sentimen.....	81
Gambar 4. 67 Deployment website hasil positif.....	82
Gambar 4. 68 Deployment website hasil negatif.....	82



## DAFTAR RUMUS

Rumus 2. 1 Random Forest.....	16
Rumus 2. 2 Decision Tree.....	17
Rumus 2. 3 KNN.....	18
Rumus 2. 4 Akurasi Confusion Matrix.....	23
Rumus 2. 5 Recall Confusion Matrix.....	23
Rumus 2. 6 Precision Confusion Matrix.....	23
Rumus 2. 7 F1_Score Confusion Matrix.....	24

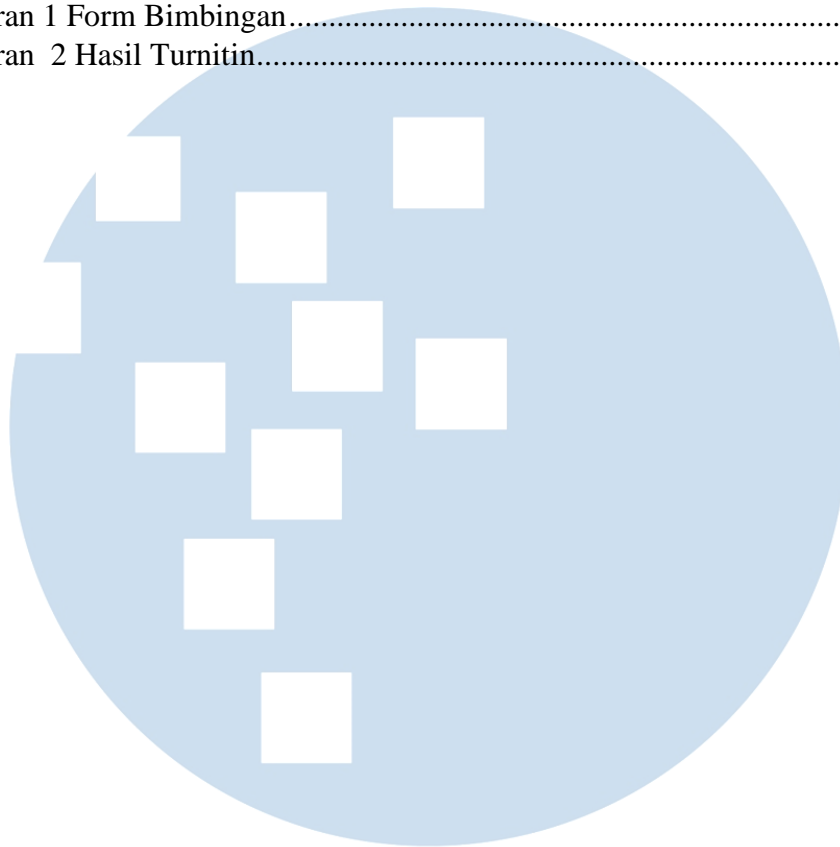


# UMMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

**DAFTAR LAMPIRAN**

Lampiran 1 Form Bimbingan..... 90  
Lampiran 2 Hasil Turnitin..... 91



**UMMN**

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA