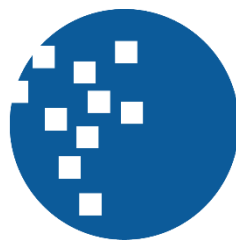


**ANALISIS SENTIMEN TERHADAP KOMENTAR  
PELECEHAN SEKSUAL DI PERGURUAN TINGGI PADA  
EMPAT MEDIA SOSIAL POPULER**



**UMN**

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

**SKRIPSI**

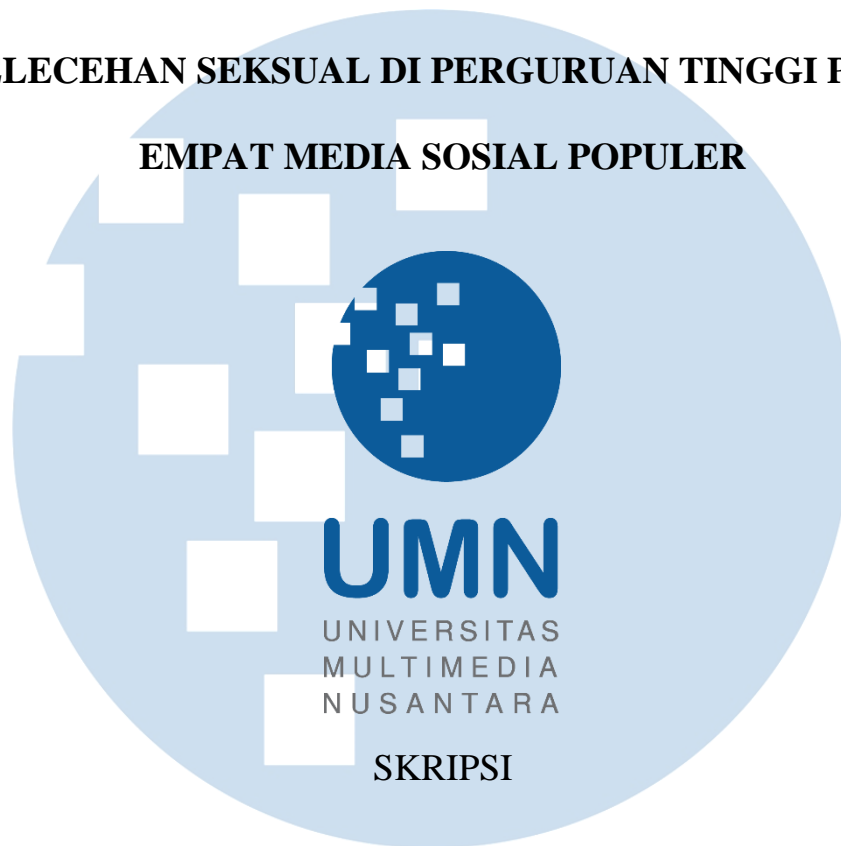
**Vinson Phoan**

**00000029963**

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

**2022**

**ANALISIS SENTIMEN TERHADAP KOMENTAR  
PELECEHAN SEKSUAL DI PERGURUAN TINGGI PADA  
EMPAT MEDIA SOSIAL POPULER**



Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer (S.Kom.)

**Vinson Phoan**

**0000029963**

**PROGRAM STUDI SISTEM INFORMASI**

**FAKULTAS TEKNIK DAN INFORMATIKA**

**UNIVERSITAS MULTIMEDIA NUSANTARA**

**TANGERANG**

**2022**

## HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Vinson Phoan

Nomor Induk Mahasiswa : 00000029963

Program studi : Sistem Informasi

Skripsi dengan judul:

Analisis Sentimen terhadap Komentar Pelecehan Seksual di Perguruan Tinggi Pada Empat Media Sosial Populer 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 2022



Vinson Phoan

UMN  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## HALAMAN PENGESAHAN

Skripsi dengan judul

“Analisis Sentimen terhadap Komentar Pelecehan Seksual di Perguruan Tinggi  
Pada Empat Media Sosial Populer”

Oleh

Nama : Vinson Phoan

NIM : 00000029963

Program Studi : Sistem Informasi

Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Jumat, 24 Juni 2022

Pukul 15.00 s.d 17.00 dan dinyatakan

LULUS

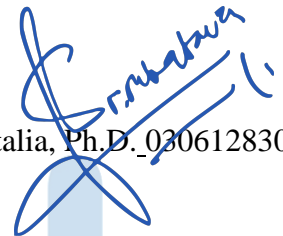
Dengan susunan pengujian sebagai berikut.

Ketua Sidang



Ririn Ikana Desanti, S.Kom.,  
M.Kom.\_0327106402

Penguji



Friska Natalia, Ph.D.\_0306128307

Pembimbing



Johan Setiawan, S.Kom., M.M.,  
M.B.A.\_0327106402

Ketua Program Studi Sistem Informasi

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA



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

## HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas academica Universitas Multimedia Nusantara, saya yang bertanda tangan di bawah ini:

Nama : Vinson Phohan  
NIM : 00000029963  
Program Studi : Sistem Informasi  
Fakultas : Fakultas Teknik dan Informatika  
JenisKarya : \*Tesis/Skripsi/Tugas Akhir (\*coret salah satu)

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.

“Analisis Sentimen terhadap Komentar Pelecehan Seksual di Perguruan Tinggi Pada Empat Media Sosial Populer” Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini, Universitas Multimedia Nusantara berhak menyimpan, mengalihmediakan/mengalihformatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan 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 2022

Yang menyatakan,



Vinson Phohan

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 Komentar Pelecehan Seksual di Perguruan Tinggi Pada Empat Media Sosial Populer” dilakukan untuk memenuhi salah satu syarat untuk memperoleh gelar sarjana komputer. 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 MA, 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. Johan Setiawan, S.Kom., M.M., M.B.A., sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Niknik Mediyawati, S.Pd., M.Hum., sebagai Ahli Bahasa yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
6. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.
7. Teman-teman yang menjadi relawan untuk membantu pelabelan data dan verifikasi data saya cepat dapat menyelesaikan laporan skripsi ini.

Semoga karya ilmiah ini dapat membantu untuk menambah pengetahuan dan berguna bagi pembaca agar mendapatkan inspirasi dan ide lebih baik.

Tangerang, 14 Juni 2022



Vinson Phohan



# ANALISIS SENTIMEN TERHADAP KOMENTAR PELECEHAN SEKSUAL DI PERGURUAN TINGGI PADA EMPAT MEDIA SOSIAL POPULER

Vinson Phohan

## ABSTRAK

Pelecehan seksual merupakan tindakan seksual yang berupa verbal dan nonverbal yang dilakukan secara sengaja dan adanya indikasi pemaksaan terhadap korban. Sering kali korban mengurungkan niatnya untuk tidak melapor atau bercerita karena korban takut terjadi hal yang tidak diinginkan kedepannya. Untuk mengetahui informasi yang berguna bagi penyintas dan penelitian pada penelitian ini, akan melakukan sentimen analisis terhadap komentar pelecehan seksual yang berkaitan dengan UMN.

Untuk menyelesaikan permasalahan tersebut, dilakukan analisis sentimen terhadap komentar pelecehan seksual di perguruan tinggi pada empat media sosial populer yang bertujuan untuk mengukur tanggapan netizen dari setiap sosial media terhadap kasus pelecehan seksual di sekitar lingkungan UMN serta mengukur performa dari penggunaan SVM dengan menggunakan metode *framework* CRISP-DM, FastText, dan algoritma SVM untuk mengidentifikasi sebuah komentar terhadap pelecehan seksual yang berkaitan dengan UMN.

Hasil Dari penelitian ini dengan menggunakan model yang berakurasi 55,14% dan data pelecehan seksual yang di kumpulkan dari tanggal 25 Juni 2021 hingga 7 Desember 2022 dengan jumlah 287 data yang didapatkan dari situs Twitter, Instagram, Medium, dan Line Today. Sentimen positif merupakan sentimen yang paling sedikit ditemukan dari keseluruhan data yang hanya sebesar 8,7% sedangkan sentimen negatif mendapatkan sebesar 36,6%. Dari ke-empat *platform* yang terbaik yaitu *platform* Twitter karena mendapatkan respon cukup baik dari segi sentimen positif dan netral ditimbang lainnya.

**Kata kunci:** *Natural Language Processing*, Pelecehan Seksual, Sentimen Analisis, Sosial Media, *Text Mining*

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 COMMENTS ON SEXUAL HARASSMENT IN COLLEGES ON FOUR POPULAR SOCIAL MEDIA**

Vinson Phoan

## ***ABSTRACT***

*Sexual harassment is a sexual act in the form of verbal and nonverbal that is carried out intentionally, and there are indications of coercion on the victim. Victims often give up their intention not to report or tell stories because the victim is afraid that something unwanted will happen in the future. To find helpful information for survivors and research in this study, we will conduct a sentiment analysis of comments on sexual harassment-related to UMN.*

*To solve this problem, sentiment analysis of comments about sexual harassment in universities on four popular social media aims to measure netizen responses from each social media to cases of sexual harassment around UMN and measure the performance of using SVM using the CRISP-framework method. DM, FastText, and SVM algorithm to identify a comment on sexual harassment-related to UMN.*

*Results From this study used a model with an accuracy of 55.14% and sexual harassment data collected from 25 June 2021 to 7 December 2022, with 287 data obtained from Twitter, Instagram, Medium, and Line Today sites. Positive sentiment is the least found sentiment from the overall data, only 8.7%, while negative sentiment is 36.6%. Of the four platforms, the best is the Twitter platform because it gets a pretty good response in terms of positive and neutral sentiments compared to others.*

***Keywords:*** *Natural Language Processing, Sexual Harrasment, Sentiment Analysis, Social Media, Text Mining*

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 PENGESAHAN.....	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS .....	iv
KATA PENGANTAR .....	v
ABSTRAK .....	vi
<i>ABSTRACT</i> .....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR .....	xii
DAFTAR RUMUS .....	xiv
DAFTAR LAMPIRAN.....	xv
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang Masalah .....	1
1.2 Rumusan Masalah .....	5
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian.....	6
1.5 Manfaat Penelitian.....	6
BAB II LANDASAN TEORI.....	8
2.1 Pelecehan Seksual .....	8
2.2 CRISP-DM .....	8
2.3 Data Mining.....	10
2.4 Text Mining.....	11
2.5 Sentimen Analisis.....	11
2.6 Natural Language Processing.....	11
2.6.1 Noise Removal.....	12
2.6.2 Case Folding .....	12
2.6.3 Tokenization.....	12
2.6.4 Stop Words.....	13
2.6.5 Stemming .....	13

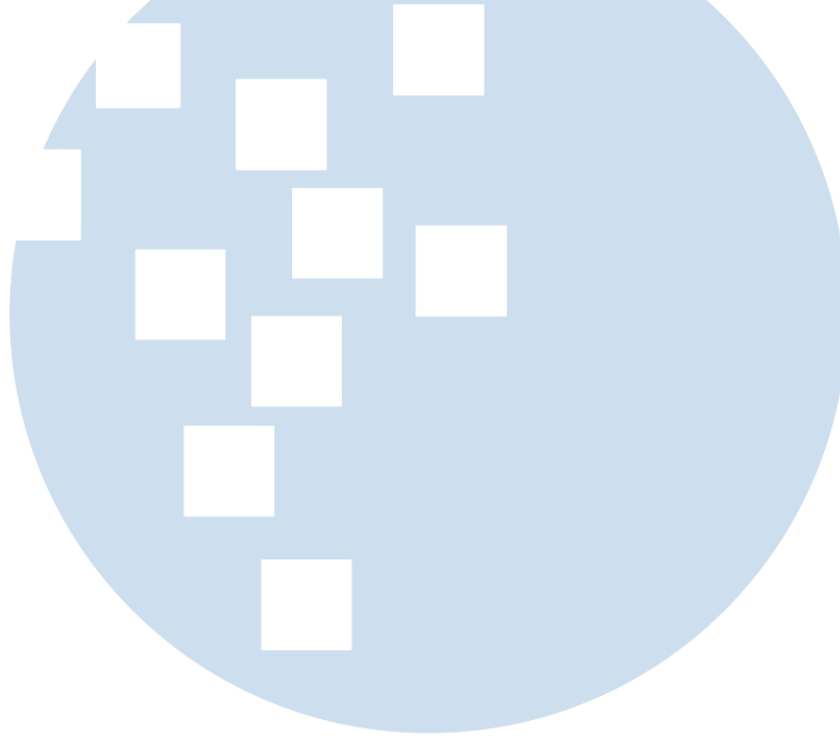
2.7	Polarity .....	14
2.8	Word Embedding .....	14
2.9	Fast Text .....	14
2.10	Uniform Manifold Approximation and Projection (UMAP).....	14
2.11	Machine learning .....	15
2.12	Support Vector Machine .....	15
2.13	Naïve Bayes .....	15
2.14	Confusion Matrix .....	16
2.14.1	Accuracy .....	17
2.14.2	Precision.....	17
2.14.3	Recall .....	17
2.14.4	F-Measure .....	18
2.15	Penelitian Terdahulu.....	19
<b>BAB III METODOLOGI PENELITIAN.....</b>		<b>29</b>
3.1	Gambaran Umum Objek Penelitian .....	29
3.2	Metode Penelitian.....	29
3.3	Alur Penelitian.....	30
3.3.1	Business Understanding.....	31
3.3.2	Data Understanding .....	31
1)	Pengumpulan Data Komentar Terkait Pelecehan Seksual .....	32
3.3.3	Data Preparation .....	35
3.3.4	Modelling.....	38
3.3.5	Evaluation .....	39
3.3.6	Deployment.....	40
3.4	Variabel Penelitian .....	40
3.4.1	Variabel Independen.....	40
3.4.2	Variabel Dependen .....	41
3.5	Teknik Pengumpulan Data.....	41
3.6	Teknik Pengambilan Sampel.....	42
3.7	Teknik Analisis Data .....	43
<b>BAB IV ANALISIS DAN HASIL PENELITIAN .....</b>		<b>44</b>
4.1	Spesifikasi Perangkat .....	44

4.2	Business Understanding .....	44
4.3	Data Understanding .....	45
4.3.1	Pengumpulan Data .....	45
4.3.2	Pelabelan Data .....	46
4.3.3	Analisis Data Eksplorasi .....	47
4.3.4	Data Cleaning .....	49
4.4	Data Preparation .....	50
4.4.1	Pemrosesan Teks .....	50
4.4.2	Wordcloud .....	51
4.4.3	Word Embedding .....	53
4.4.4	UMAP .....	54
4.4.5	Slicing Data .....	55
4.4.6	Splitting Data .....	56
4.5	Modelling .....	58
4.5.1	Training Model .....	58
4.5.2	Testing Model .....	59
4.6	Evaluation .....	60
4.7	Deployment .....	62
4.7.1	Visualisasi dan Analisa Sentimen .....	63
4.8	Hasil dan Analisis .....	70
BAB V SIMPULAN DAN SARAN .....		73
5.1	Simpulan .....	73
5.2	Saran .....	74
DAFTAR PUSTAKA .....		75
LAMPIRAN .....		80

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 TABEL

Tabel 2. 1 Penelitian Terdahulu .....	19
Tabel 3. 1 Perbandingan Frameworks Data Mining .....	29
Tabel 3.2 Tools Perbandingan.....	43



# UMMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## DAFTAR GAMBAR

Gambar 1. 1 Grafik Kekerasan Perempuan Berdasarkan Jenjang Pendidikan Tahun 2015-2020 [2].....	1
Gambar 1. 2 Grafik Pelaku Kekerasan Seksual Tahun 2015-2020 [2].....	2
Gambar 1. 3 Tindakan Kekerasan dalam Hubungan [3].....	2
Gambar 1. 4 Platform Media Sosial Indonesia Juli 2021 [6].....	3
Gambar 2. 1 Siklus CRISP-DM [9] .....	9
Gambar 2. 2 Confusion Matrix [29].....	16
Gambar 3. 1 Alur Penelitian Penerapan CRISP-DM [13]. .....	30
Gambar 3.2 Flowchart Business Understanding.....	31
Gambar 3.3 Flowchart Data Understanding.....	31
Gambar 3.4 Flowchart Labelling .....	33
Gambar 3.5 Flowchart Verifikasi Label.....	34
Gambar 3.6 Flowchart Data Preparation.....	36
Gambar 3.7 Flowchart Modelling .....	39
Gambar 3.8 Flowchart Evaluation .....	40
Gambar 3.9 Flowchart Deployment.....	40
Gambar 3.10 Flowchart Pengumpulan Data .....	41
Gambar 3.11 Flowchart Scraping .....	42
Gambar 4. 1 Hasil data pelecehan seksual .....	45
Gambar 4. 2 Hasil data terkait UMN .....	46
Gambar 4. 3 Hasil Pelabelan.....	46
Gambar 4. 4 Hasil Verifikasi Pelabelan.....	47
Gambar 4. 5 Dimensi Data.....	47
Gambar 4. 6 Total Sentimen .....	48
Gambar 4. 7 Persentase Sentimen.....	48
Gambar 4. 8 Informasi Data.....	49
Gambar 4. 9 Hasil Missing Values .....	49
Gambar 4. 10 Hasil Label Encoding.....	50
Gambar 4. 11 Hasil Pemorsesan Teks.....	51
Gambar 4. 12 Wordcloud Negative Tweets.....	52
Gambar 4. 13 Wordcloud Netral Tweets .....	52
Gambar 4. 14 Wordcloud Positive Tweets .....	53
Gambar 4. 15 Hasil Word Embedding.....	54
Gambar 4. 16 Hasil UMAP.....	54
Gambar 4. 17 Visualisasi Word Relation.....	55
Gambar 4. 18 Hasil Perbesar.....	55
Gambar 4. 19 Hasil Membersihkan Missing Values .....	56
Gambar 4. 20 Hasil Splitting Data .....	56
Gambar 4. 21 Training Data Sentimen .....	57
Gambar 4. 22 Testing Data Sentimen .....	58
Gambar 4. 23 Training Model.....	59
Gambar 4. 24 Hasil Prediksi Testing Data.....	59

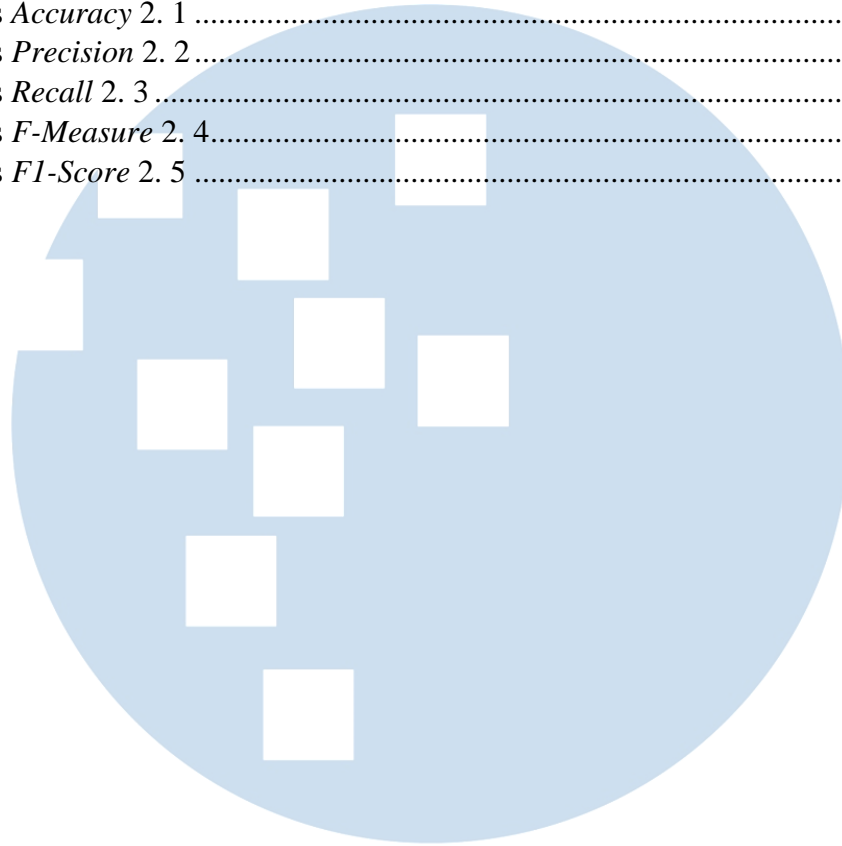
Gambar 4. 25 Confusion Matrix Training Data.....	60
Gambar 4. 26 Classification Report Training Data.....	61
Gambar 4. 27 Confusion Matrix Testing Data.....	61
Gambar 4. 28 Classification Report Testing Data .....	62
Gambar 4. 29 Hasil Prediksi .....	63
Gambar 4. 30 Persentase Sentimen.....	64
Gambar 4. 31 Total Sentimen .....	64
Gambar 4. 32 Total Persentase berdasarkan Sumber.....	65
Gambar 4. 33 Total Sentimen berdasarkan Sumber.....	66
Gambar 4. 34 Persentase Komentar berdasarkan Sumber dan Sentimen .....	67
Gambar 4. 35 Komentar Negatif.....	67
Gambar 4. 36 Komentar Positif .....	68
Gambar 4. 37 Komentar Netral.....	68
Gambar 4. 38 20 Frekuensi kata terbanyak pada sentimen netral.....	69
Gambar 4. 39 Total Sentimen Pada Tahun 2021 .....	70





## DAFTAR RUMUS

Rumus <i>Accuracy</i> 2. 1 .....	17
Rumus <i>Precision</i> 2. 2 .....	17
Rumus <i>Recall</i> 2. 3 .....	18
Rumus <i>F-Measure</i> 2. 4 .....	18
Rumus <i>F1-Score</i> 2. 5 .....	18

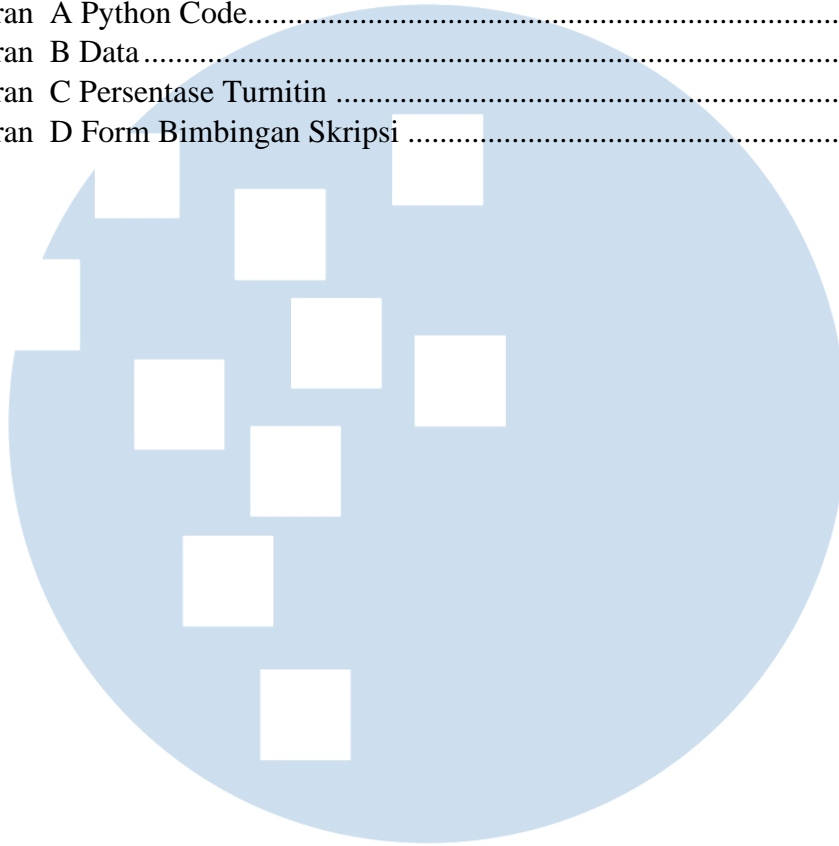


# UMMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## DAFTAR LAMPIRAN

Lampiran A Python Code.....	80
Lampiran B Data.....	91
Lampiran C Persentase Turnitin .....	176
Lampiran D Form Bimbingan Skripsi .....	182



# UMMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA