

**ANALISIS SENTIMEN PROGRAM NATURALISASI PEMAIN TIMNAS
SEPAK BOLA INDONESIA MENGGUNAKAN ALGORITMA NAIVE
BAYES**



SKRIPSI

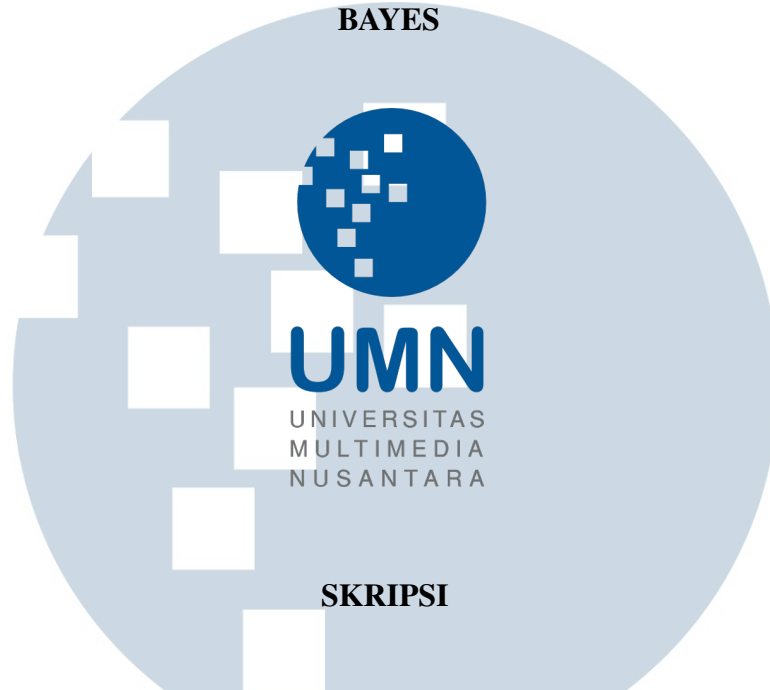
Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

Sabilillah Amory Reyhan Rusjdi
00000054377

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025**

**ANALISIS SENTIMEN PROGRAM NATURALISASI PEMAIN TIMNAS
SEPAK BOLA INDONESIA MENGGUNAKAN ALGORITMA NAIVE**

BAYES



SKRIPSI

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

Sabilillah Amory Reyhan Rusjdi

00000054377

UMN

**UNIVERSITAS
MULTIMEDIA
NUSANTARA**

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

TANGERANG

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Sabilillah Amory Reyhan Rusjdi

Nomor Induk Mahasiswa : 00000054377

Program Studi : Informatika

Skripsi dengan judul:

Analisis Sentimen Program Naturalisasi Pemain Timnas Sepak Bola Indonesia Menggunakan Algoritma Naive Bayes

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 Desember 2024



(Sabilillah Amory Reyhan Rusjdi)

HALAMAN PENGESAHAN

Skripsi dengan judul

**ANALISIS SENTIMEN PROGRAM NATURALISASI PEMAIN TIMNAS
SEPAK BOLA INDONESIA MENGGUNAKAN ALGORITMA NAIVE
BAYES**

oleh

Nama : Sabilillah Amory Reyhan Rusjdi
NIM : 00000054377
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

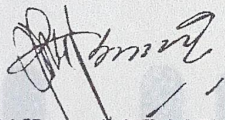
Telah diujikan pada hari Kamis, 9 Januari 2025

Pukul 08.00 s/s 10.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut

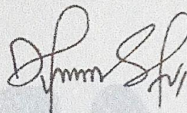
Ketua Sidang



(Adhi Kushadi, S.T, M.Si.)

NIDN: 0303037304

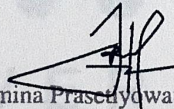
Penguji



(David Agustriawan, S.Kom., M.Sc.,
Ph.D.)

NIDN: 0525088601

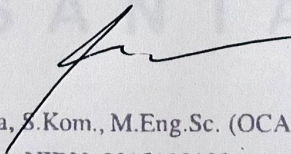
Pembimbing



(Dr. Maria Irminda Prasetyowati, S.Kom., M.T.)

NIDN: 0725057201

Ketua Program Studi Informatika,



(Arya Wicaksana, S.Kom., M.Eng.Sc. (OCA, CEH, CEI))

NIDN: 0315109103

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini:

Nama : Sabilillah Amory Reyhan Rusjdi
NIM : 00000054377
Program Studi : Informatika
Jenjang : S1
Judul Karya Ilmiah : Analisis Sentimen Program
Naturalisasi Pemain Timnas Sepak
Bola Indonesia Menggunakan
Algoritma Naive Bayes

Menyatakan dengan sesungguhnya bahwa saya bersedia (**pilih salah satu**):

- ☒ 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 bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
 - Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - Embargo publikasi karya ilmiah dalam kurun waktu tiga tahun.

Tangerang, 9 Desember 2024

Yang menyatakan

Sabilillah Amory Reyhan Rusjdi

**Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

Halaman Persembahan / Motto

"A good name is to be more desired than great wealth, Favor is better than silver and gold."

Proverbs 22:1 (NASB)



KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Analisis Sentimen Program Naturalisasi Pemain Timnas Sepak Bola Indonesia Menggunakan Algoritma Naive Bayes 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. Andrey Andoko, 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. Bapak Arya Wicaksana, S.Kom., M.Eng.Sc. (OCA, CEH, CEI), selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Ibu Dr. Maria Irminda Prasetyowati, S.Kom., M.T. , sebagai Pembimbing pertama yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya tesis ini.
5. Orang Tua, keluarga dan teman 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 Desember 2024



Sabilillah Amory Reyhan Rusjdi

ANALISIS SENTIMEN PROGRAM NATURALISASI PEMAIN TIMNAS SEPAK BOLA INDONESIA MENGGUNAKAN ALGORITMA NAIVE BAYES

Sabilillah Amory Reyhan Rusjdi

ABSTRAK

Naturalisasi adalah proses hukum yang dilakukan oleh seseorang dalam memperoleh kewarganegaraan suatu negara. Program naturalisasi pemain telah dilakukan oleh Persatuan Sepak Bola Seluruh Indonesia (PSSI) sejak 2010 namun hingga saat ini prestasi timnas sepak bola Indonesia kurang baik yang menyebabkan banyak menuai pro dan kontra. Oleh karena itu dilakukan analisa sentimen dengan tujuan mengetahui opini masyarakat terkait program naturalisasi. Untuk mencapai tingkat akurasi tertinggi, sentimen dibagi menjadi dua, yaitu sentimen positif dan sentimen negatif. Model yang digunakan dalam melakukan klasifikasi merupakan model dengan algoritma Naïve Bayes dan TF-IDF Vectorizer dalam mengukur relevansi kata. Penyeimbangan data juga dilakukan dengan tujuan data menjadi seimbang. Hasil yang didapat sebelum data seimbang yakni Complement Naive Bayes dengan akurasi tertinggi yaitu 84% dengan *precision* 65%, *recall* 94% dan *F1-score* 77% pada data positif, lalu data negatif didapat dengan *precision* 97%, *recall* 81% dan *F1-score* 88%. Setelah dilakukan penyeimbangan yakni kedua model Multinomial Naive Bayes dan Complement Naive Bayes dengan akurasi tertinggi yaitu 90% dengan *precision* 89%, *recall* 94% dan *F1-score* 84% pada data positif, lalu data negatif didapat dengan *precision* 98%, *recall* 89% dan *F1-score* 91%.

Kata kunci: Analisis sentimen, *Machine Learning*, *Naive Bayes*, Naturalisasi, Timnas Indonesia

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 THE INDONESIAN FOOTBALL NATIONAL
TEAM PLAYER NATURALIZATION PROGRAM USING NAIVE BAYES**

ALGORITHM

Sabilillah Amory Reyhan Rusjdi

ABSTRACT

Naturalization is a legal process carried out by a person in obtaining citizenship of a country. The player naturalization program has been carried out by the Indonesian Football Association (PSSI) since 2010, but until now the achievements of the Indonesian national football team are not good, which has caused many pros and cons. Therefore, sentiment analysis is carried out with the aim of knowing public opinion regarding the naturalization program. To achieve the highest level of accuracy, sentiment is divided into two, namely positive sentiment and negative sentiment. The model used in classification is a model with the Naïve Bayes algorithm and TF-IDF Vectorizer in measuring word relevance. Data balancing is also done with the aim of the data being balanced. The results obtained before the data is balanced are Complement Naive Bayes with the highest accuracy of 84% with precision 65%, recall 94% and F1-score 77% on positive data, then negative data is obtained with precision 97%, recall 81% and F1-score 88%. After balancing the two models Multinomial Naive Bayes and Complement Naive Bayes with the highest accuracy of 90% with precision 89%, recall 94% and F1-score 84% on positive data, then negative data is obtained with precision 98%, recall 89% and F1-score 91%.

Keywords: *Indonesian National Team, Machine Learning, Naive Bayes, Naturalization, Sentiment analysis*

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 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	xiv
DAFTAR KODE	xv
DAFTAR LAMPIRAN	xvi
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	5
1.3 Batasan Permasalahan	5
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	6
1.6 Sistematika Penulisan	6
BAB 2 LANDASAN TEORI	7
2.1 Naturalisasi/Pewarganegaraan	7
2.2 Analisa Sentimen	7
2.3 Text Preprocessing	7
2.4 Data Splitting	8
2.5 TF-IDF	8
2.6 ADASYN	9
2.7 Naive Bayes	9
2.8 Confusion Matrix	11
2.8.1 Accuracy	12
2.8.2 Precision	12
2.8.3 Recall	12
2.8.4 F1-Score	12
BAB 3 METODOLOGI PENELITIAN	13
3.1 Metodologi Penelitian	13
3.2 Rancangan Sistem	13
3.2.1 Text Crawling	14
3.2.2 Text Pre-processing	16
3.2.3 TF-IDF	18
3.2.4 Data Splitting	18
3.2.5 Data Balancing	19
3.2.6 Penerapan Model	19
3.2.7 Pengujian Model	21
3.2.8 Evaluasi	21
BAB 4 HASIL DAN DISKUSI	22
4.1 Spesifikasi Sistem	22
4.2 Implementasi Sistem	22
4.2.1 Text Crawling	22

4.2.2	Text Pre-Processing	23
4.2.3	TF-IDF	31
4.2.4	Training dan Testing Split Data	31
4.2.5	Data Balancing	32
4.2.6	Penerapan Model	33
4.3	Pengujian dan Evaluasi	33
4.3.1	Pelatihan	33
4.3.2	Pengujian	46
4.3.3	Evaluasi	75
BAB 5	SIMPULAN DAN SARAN	84
5.1	Simpulan	84
5.2	Saran	84
DAFTAR PUSTAKA		85

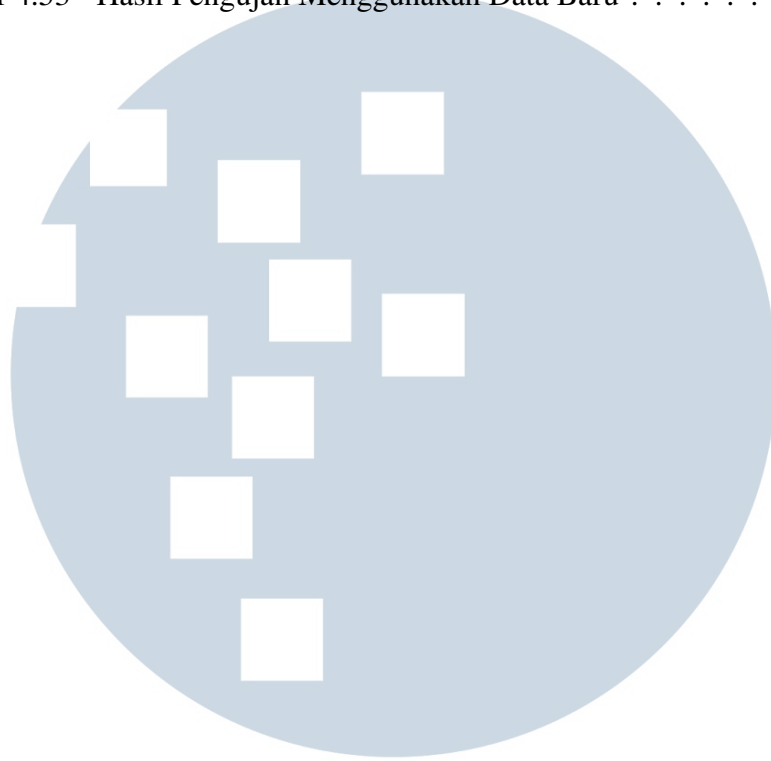


DAFTAR GAMBAR

Gambar 2.1	<i>Confusion Matrix</i> (sumber: Medium)	11
Gambar 3.1	<i>Flowchart</i> Rancangan Sistem	14
Gambar 3.2	<i>Flowchart Text Crawling</i>	15
Gambar 3.3	<i>Flowchart Text Pre-processing</i>	16
Gambar 3.4	<i>Flowchart Labelling</i>	17
Gambar 3.5	<i>Flowchart Model Naïve Bayes</i>	20
Gambar 4.1	Informasi data yang didapat	23
Gambar 4.2	Distribusi Kata	28
Gambar 4.3	Hasil pelabelan data	30
Gambar 4.4	Hasil pelabelan data	30
Gambar 4.5	Hasil Penyeimbangan Data <i>Train</i> label VADER	32
Gambar 4.6	Hasil Penyeimbangan Data <i>Train</i> label <i>pre-trained</i>	32
Gambar 4.7	Hasil Pelatihan <i>Complement Naive Bayes</i>	34
Gambar 4.8	<i>Confusion matrix Multinomial Naive Bayes</i> Tahap Pelatihan	34
Gambar 4.9	Hasil Pelatihan <i>Complement Naive Bayes</i> dengan <i>hyperparameter tuning</i>	36
Gambar 4.10	<i>Confusion matrix Multinomial Naive Bayes</i> Tahap Pelatihan	36
Gambar 4.11	Hasil Pelatihan <i>Multinomial Naive Bayes</i>	38
Gambar 4.12	<i>Confusion matrix Multinomial Naive Bayes</i> Model Pelatihan	38
Gambar 4.13	Hasil Pelatihan <i>Multinomial Naive Bayes</i> dengan <i>hyperparameter tuning</i>	39
Gambar 4.14	<i>Confusion matrix Multinomial Naive Bayes</i> Model Pelatihan	40
Gambar 4.15	Hasil Pelatihan <i>Complement Naive Bayes</i>	41
Gambar 4.16	<i>Confusion matrix Multinomial Naive Bayes</i> Tahap Pelatihan	42
Gambar 4.17	Hasil Pelatihan <i>Complement Naive Bayes</i> dengan <i>hyperparameter tuning</i>	43
Gambar 4.18	<i>Confusion matrix Multinomial Naive Bayes</i> Tahap Pelatihan	43
Gambar 4.19	Hasil Pelatihan <i>Multinomial Naive Bayes</i>	45
Gambar 4.20	<i>Confusion matrix Multinomial Naive Bayes</i> Model Pelatihan	45
Gambar 4.21	Hasil Pengujian <i>Complement Naive Bayes</i> Sebelum Penyeimbangan	47
Gambar 4.22	<i>Confusion matrix Complement Naive Bayes</i> Sebelum Penyeimbangan	47
Gambar 4.23	Hasil Pengujian <i>Complement Naive Bayes</i> Setelah Penyeimbangan	48
Gambar 4.24	<i>Confusion matrix Complement Naive Bayes</i> Setelah Penyeimbangan	49
Gambar 4.25	Hasil Pengujian <i>Complement Naive Bayes</i> dengan <i>hyperparameter tuning</i>	50
Gambar 4.26	<i>Confusion matrix Complement Naive Bayes</i> dengan <i>hyperparameter tuning</i>	50
Gambar 4.27	Hasil Pengujian <i>Complement Naive Bayes</i> Setelah Penyeimbangan dan <i>hyperparameter tuning</i>	52
Gambar 4.28	<i>Confusion matrix Complement Naive Bayes</i> Setelah Penyeimbangan dan <i>hyperparameter tuning</i>	52

Gambar 4.29	Hasil pengujian <i>Multinomial Naive Bayes</i> Sebelum Penyeimbangan	54
Gambar 4.30	<i>Confusion matrix Multinomial Naive Bayes</i> Sebelum Penyeimbangan	54
Gambar 4.31	Hasil pengujian <i>Multinomial Naive Bayes</i> Setelah Penyeimbangan	55
Gambar 4.32	<i>Confusion matrix Multinomial Naive Bayes</i> Setelah Penyeimbangan	56
Gambar 4.33	Hasil pengujian <i>Multinomial Naive Bayes</i> Setelah <i>hyperparameter tuning</i>	57
Gambar 4.34	<i>Confusion matrix Multinomial Naive Bayes</i> Setelah <i>hyperparameter tuning</i>	57
Gambar 4.35	Hasil pengujian <i>Multinomial Naive Bayes</i> Setelah Penyeimbangan dan <i>hyperparameter tuning</i>	59
Gambar 4.36	<i>Confusion matrix Multinomial Naive Bayes</i> Setelah Penyeimbangan dan <i>hyperparameter tuning</i>	59
Gambar 4.37	Hasil Pengujian <i>Complement Naive Bayes</i> Sebelum Penyeimbangan	61
Gambar 4.38	<i>Confusion matrix Complement Naive Bayes</i> Sebelum Penyeimbangan	61
Gambar 4.39	Hasil Pengujian <i>Complement Naive Bayes</i> Setelah Penyeimbangan	62
Gambar 4.40	<i>Confusion matrix Complement Naive Bayes</i> Setelah Penyeimbangan	63
Gambar 4.41	Hasil Pengujian <i>Complement Naive Bayes</i> Setelah <i>hyperparameter tuning</i>	64
Gambar 4.42	<i>Confusion matrix Complement Naive Bayes</i> Setelah <i>hyperparameter tuning</i>	64
Gambar 4.43	Hasil Pengujian <i>Complement Naive Bayes</i> Setelah Penyeimbangan dan <i>hyperparameter tuning</i>	66
Gambar 4.44	<i>Confusion matrix Complement Naive Bayes</i> Setelah Penyeimbangan dan <i>hyperparameter tuning</i>	66
Gambar 4.45	Hasil pengujian <i>Multinomial Naive Bayes</i> Sebelum Penyeimbangan	68
Gambar 4.46	<i>Confusion matrix Multinomial Naive Bayes</i> Sebelum Penyeimbangan	68
Gambar 4.47	Hasil pengujian <i>Multinomial Naive Bayes</i> Setelah Penyeimbangan	69
Gambar 4.48	<i>Confusion matrix Multinomial Naive Bayes</i> Setelah Penyeimbangan	70
Gambar 4.49	Hasil pengujian <i>Multinomial Naive Bayes</i> Setelah <i>hyperparameter tuning</i>	71
Gambar 4.50	<i>Confusion matrix Multinomial Naive Bayes</i> Setelah <i>hyperparameter tuning</i>	71
Gambar 4.51	Hasil pengujian <i>Multinomial Naive Bayes</i> Setelah Penyeimbangan dan <i>hyperparameter tuning</i>	73

Gambar 4.52	<i>Confusion matrix Multinomial Naive Bayes</i> Setelah Penyeimbangan dan <i>hyperparameter tuning</i>	73
Gambar 4.53	Hasil Pengujian Menggunakan Data Baru	74



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR TABEL

Tabel 4.1	Hasil <i>Pre-processing Cleaning</i> dan <i>Case Folding</i>	24
Tabel 4.2	Hasil <i>Pre-processing</i> Normalisasi	25
Tabel 4.3	Hasil <i>Pre-processing</i> Tokenisasi	26
Tabel 4.4	Hasil <i>Pre-processing Stopword Remove</i>	27
Tabel 4.5	Hasil <i>Pre-processing Lemmatized</i>	27
Tabel 4.6	Hasil <i>Pre-processing Translation</i>	29
Tabel 4.7	Hasil Evaluasi Performa Model Sebelum Penyeimbangan .	75
Tabel 4.8	Hasil Evaluasi Performa Model Setelah <i>hyperparameter tuning</i>	76
Tabel 4.9	Hasil Evaluasi Performa Model Setelah Penyeimbangan . .	77
Tabel 4.10	Hasil Evaluasi Performa Model Setelah Penyeimbangan . .	78
Tabel 4.11	Hasil Evaluasi Performa Model Sebelum Penyeimbangan .	79
Tabel 4.12	Hasil Evaluasi Performa Model Setelah <i>hyperparameter tuning</i>	80
Tabel 4.13	Hasil Evaluasi Performa Model Setelah Penyeimbangan . .	81
Tabel 4.14	Hasil Evaluasi Performa Model Setelah Penyeimbangan . .	82



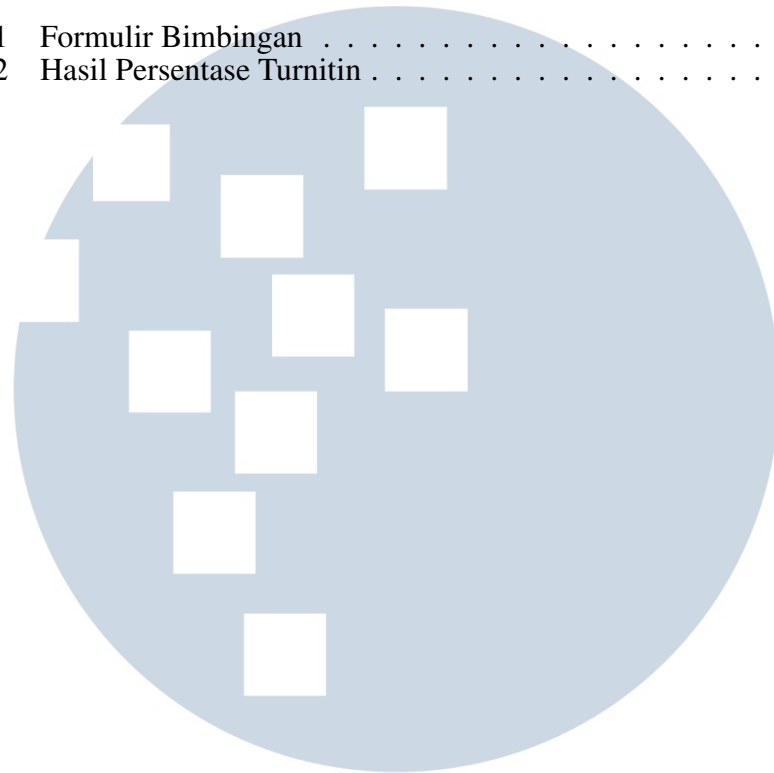
DAFTAR KODE

4.1	Potongan Kode <i>Text Crawling</i>	22
4.2	Potongan Kode <i>Cleaning dan Case Folding</i>	24
4.3	Potongan Kode Normalisasi	25
4.4	Potongan Kode Tokenisasi	25
4.5	Potongan Kode <i>Stopword Remove</i>	26
4.6	Potongan Kode <i>Lemmatize</i>	27
4.7	Potongan Kode <i>Translation</i>	28
4.8	Potongan Kode <i>Labelling</i>	29
4.9	Potongan Kode TF-IDF	31
4.10	Potongan Kode <i>Data Splitting</i>	31
4.11	Potongan Kode <i>Complement Naive Bayes</i>	33
4.12	Potongan Kode <i>Multinomial Naive Bayes</i>	33



DAFTAR LAMPIRAN

Lampiran 1	Formulir Bimbingan	88
Lampiran 2	Hasil Persentase Turnitin	89



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA