

**ANALISIS SENTIMEN PUBLIK PADA PEMBATALAN TUAN
RUMAH INDONESIA DI PIALA DUNIA U-20
MENGUNAKAN METODE FASTTEXT DAN ALGORITMA
RECURRENT NEURAL NETWORK (RNN)**

SKRIPSI



Oleh :

AAN EVIAN NANDA

NPM 19081010040

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR**

2024

**ANALISIS SENTIMEN PUBLIK PADA PEMBATALAN TUAN
RUMAH INDONESIA DI PIALA DUNIA U-20
MENGUNAKAN METODE FASTTEXT DAN ALGORITMA
RECURRENT NEURAL NETWORK (RNN)**

SKRIPSI

Diajukan Untuk Memenuhi Sebagai Persyaratan Dalam Menempuh Gelar Sarjana
Komputer Program Studi Informatika



Oleh :

AAN EVIAN NANDA

NPM. 19081010040

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
2024**

LEMBAR PENGESAHAN SKRIPSI

Judul : ANALISIS SENTIMEN PUBLIK PADA PEMBATALAN TUAN RUMAH
INDONESIA DI PIALA DUNIA U-20 MENGGUNAKAN METODE
FASTTEXT DAN ALGORITMA RECURRENT NEURAL NETWORK (RNN)

Oleh : AAN EVIAN NANDA

NPM : 19081010040

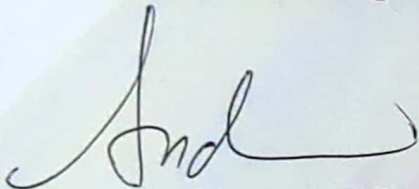
Telah Diseminarkan Dalam Ujian Skripsi Pada :

Hari Rabu, Tanggal 3 April 2024

Mengetahui

Dosen Pembimbing


1.



Andreas Nugroho Sihananto, S.Kom., M.Kom.

NPT. 211199 00 412271

2.



Agung Mustika Rizki, S.Kom., M.Kom

NIP. 19930725 202203 1008

Dosen Penguji

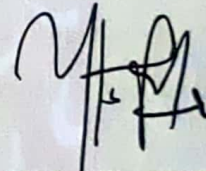
1.



Chrystia Afi Putra, S.Kom., M.T.

NIP. 19861008 2021211 001

2.



Yisti Vita Via, S.ST., M.Kom.

NIP. 19860425 2021212 001

Menyetujui

Dekan
Fakultas Ilmu Komputer

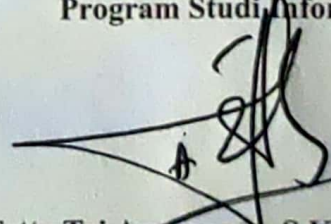


Prof. Dr. Ir. Novirina Hendrasarie. MT.

NIP. 19681126 199403 2 001

Koordinator

Program Studi Informatika



Fetty Tri Anggraeni, S.Kom., M.Kom.

NIP. 19820211 202121 2 005

SURAT PERNYATAAN ANTI PLAGIAT

Saya mahasiswa Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur, yang bertanda tangan di bawah ini:

Nama : Aan Evian Nanda

NPM : 19081010040

Dengan ini menyatakan bahwa judul skripsi atau tugas akhir yang saya ajukan dan kerjakan, yang berjudul:

**“ANALISIS SENTIMEN PUBLIK PADA PEMBATALAN TUAN RUMAH
INDONESIA DI PIALA DUNIA U-20 MENGGUNAKAN METODE
FASTTEXT DAN ALGORITMA RECURRENT NEURAL NETWORK
(RNN)”**

Bukan merupakan plagiat dari skripsi atau tugas akhir maupun penelitian orang lain dan juga bukan merupakan produk atau *software* yang saya beli dari pihak lain. Saya juga menyatakan bahwa skripsi ini adalah pekerjaan Saya sendiri, kecuali yang dinyatakan dalam daftar pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional “Veteran” Jawa Timur maupun institusi pendidikan lainnya.

Jika ternyata kemudian hari pernyataan ini terbukti tidak benar, maka Saya siap menerima segala konsekuensinya.

Surabaya, 03 April 2024

Hormat Saya,



Aan Evian Nanda
NPM. 19081010040

ANALISIS SENTIMEN PUBLIK PADA PEMBATALAN TUAN RUMAH INDONESIA DI PIALA DUNIA U-20 MENGGUNAKAN METODE FASTTEXT DAN ALGORITMA RECURRENT NEURAL NETWORK (RNN)

Nama Mahasiswa : Aan Evian Nanda

NPM : 19081010040

Program Studi : Informatika

Dosen Pembimbing : Andreas Nugroho Sihananto, S.Kom., M.Kom.

Agung Mustika Rizki, S.Kom., M.Kom.

ABSTRAK

Peluang emas Indonesia untuk mengikuti kompetisi bola kelas dunia pada kompetisi Piala Dunia U-20 musnah, karena FIFA memberi keputusan pencabutan status Indonesia sebagai tuan rumah Piala Dunia U-20. Netizen Indonesia yang merasa kecewa mengungkapkan opini mereka dan menjadi *trending* di media sosial Twitter.

Penelitian ini berfokus pada analisis sentimen *tweet* menggunakan kombinasi metode FastText *embeddings* untuk vektorisasi kata dan menggunakan algoritma RNN jenis LSTM untuk klasifikasi sentimen. Dataset yang digunakan berjumlah total 9.645 data yang terdiri dari 4.141 data positif dan 5.504 data negatif yang diambil pada rentang waktu 29 maret 2023 s.d 05 April 2023. Hasil pengujian pada model LSTM memberikan performa terbaik dengan nilai *accuracy* sebesar 74.92%, *precision* 74.73%, *recall* 74.92%, dan *f1-score* 74.77%.

Kesimpulan dari penelitian ini adalah mayoritas dari *dataset* memiliki sentimen negatif yang berarti masyarakat lebih cenderung memberikan opini negatif daripada memberikan dukungan kepada sepak bola Indonesia yang sedang mengalami masalah. Diharapkan dengan adanya kesimpulan tersebut kedepannya masyarakat lebih mengontrol opini mereka dan memberikan opini yang positif saat Indonesia sedang mengalami masalah.

Kata Kunci : *Twitter, Piala Dunia U-20, Analisis Sentimen, FastText, RNN, LSTM.*

KATA PENGANTAR

Dengan penuh rasa syukur, penulis ingin mengucapkan terima kasih kepada Allah SWT atas berkat dan rahmat-Nya yang telah memungkinkan penyelesaian skripsi berjudul "Analisis Sentimen Publik pada Pembatalan Tuan Rumah Indonesia di Piala Dunia U-20 menggunakan Metode FastText dan Algoritma *Recurrent Neural Network* (RNN)."

Laporan skripsi ini disusun dalam rangka memenuhi persyaratan mata kuliah skripsi dan sebagai salah satu langkah penting dalam menyelesaikan pendidikan tingkat Sarjana (S1) di Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur.

Penulis sadar akan adanya keterbatasan dalam penulisan skripsi ini, dan dengan rendah hati menerima saran serta kritik yang membangun. Semoga skripsi ini dapat memberikan manfaat yang signifikan bagi semua pihak, baik bagi pembaca maupun untuk penelitian mendatang.

Surabaya, 03 April 2024

Penulis

UCAPAN TERIMA KASIH

Penulisan laporan skripsi ini tidak dapat terwujud tanpa bantuan, motivasi, dan dukungan dari berbagai pihak. Oleh karena itu, dengan tulus, penulis ingin mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT., selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Ibu Dr. Ir. Novirina Hendrasarie, M.T., selaku Dekan Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Fetty Tri Anggraeny, S.Kom., M.Kom., selaku Koordinator Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur dan Dosen Wali penulis.
4. Bapak Andreas Nugroho Sihananto, S.Kom., M.Kom., selaku Dosen Pembimbing I yang meluangkan waktu, tenaga, serta pikiran untuk membimbing dan mengarahkan penulis selama proses penyelesaian skripsi.
5. Bapak Agung Mustika Rizki, S.Kom., M.Kom., selaku Dosen Pembimbing II yang telah memberikan arahan, dukungan, serta saran kepada penulis sehingga penulis dapat menyelesaikan penyusunan skripsi.
6. Seluruh Dosen Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur yang telah memberikan ilmunya kepada penulis selama perkuliahan.
7. Kedua orang tua penulis, Bapak Sukadi dan Ibu Muryani yang selalu memberikan dukungan dan motivasi kepada penulis selama perkuliahan.
8. Alfinas Agung Mujiono, Muhammad Alfyando, Arif Widiasan Subagio, dan seluruh teman-teman dari program studi Informatika angkatan 2019 yang telah memberikan bantuan dan dukungan kepada penulis dalam menempuh perkuliahan dan menyelesaikan skripsi.

Penulis hanya bisa berharap, semoga Tuhan Yang Maha Esa senantiasa memberi perlindungan dan membalas semua kebaikan yang telah diberikan.

DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI	i
SURAT PERNYATAAN ANTI PLAGIAT	ii
ABSTRAK.....	iii
KATA PENGANTAR	iv
UCAPAN TERIMA KASIH.....	v
DAFTAR ISI.....	vi
DAFTAR TABEL.....	ix
DAFTAR GAMBAR	xi
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah.....	4
1.3. Tujuan Penelitian	5
1.4. Manfaat Penelitian	5
1.5. Batasan Masalah	6
BAB II TINJAUAN PUSTAKA.....	7
2.1. Penelitian Terdahulu	7
2.2. Piala Dunia U-20.....	9
2.3. <i>Machine Learning</i>	10
2.4. Twitter.....	10
2.5. Analisis Sentimen	11
2.6. Snsrape	11
2.7. VADER.....	12
2.8. <i>Text Mining</i>	13
2.9. <i>Text Preprocessing</i>	13
2.9.1. <i>Cleansing</i>	14
2.9.2. <i>Case-folding</i>	14
2.9.3. <i>Tokenizing</i>	15
2.9.4. <i>Normalization</i>	15
2.9.5. <i>Stopwords Removal</i>	16
2.9.6. <i>Stemming</i>	16

2.10. <i>Word Embedding</i>	17
2.10.1. <i>FastText</i>	17
2.11. <i>Recurrent Neural Network</i>	18
2.11.1. <i>Long Short Term Memory</i>	20
2.12. <i>Confusion Matrix</i>	23
BAB III METODOLOGI PENELITIAN	26
3.1. Tahapan Penelitian.....	26
3.2. Studi Literatur	27
3.3. Analisa dan Desain	27
3.4. Akuisisi Data.....	28
3.5. <i>Text Preprocessing</i>	30
3.5.1. <i>Cleansing</i>	31
3.5.2. <i>Case-folding</i>	32
3.5.3. <i>Tokenizing</i>	33
3.5.4. <i>Normalization</i>	34
3.5.5. <i>Stopwords Removal</i>	35
3.5.6. <i>Stemming</i>	36
3.6. Pelabelan dan Sortir data	37
3.7. <i>FastText Word Embedding</i>	41
3.8. Klasifikasi Data.....	43
3.8.1. Pembuatan Model LSTM.....	43
3.9. Skenario Uji Coba.....	46
BAB IV HASIL DAN PEMBAHASAN	47
4.1. Lingkungan Penelitian	47
4.2. Spesifikasi Perangkat Keras.....	47
4.3. Spesifikasi Perangkat Lunak.....	47
4.4. Hasil Akuisisi Data	48
4.5. Hasil <i>Text Preprocessing</i>	50
4.5.1. <i>Cleansing</i>	51
4.5.2. <i>Case-Folding</i>	53
4.5.3. <i>Tokenizing</i>	54
4.5.4. <i>Normalization</i>	55

4.5.5. <i>Stopwords Removal</i>	57
4.5.6. <i>Stemming</i>	58
4.6. Hasil Pelabelan Dan Sortir Data	59
4.7. Hasil FastText <i>Word Embedding</i>	65
4.8. Hasil Model Klasifikasi LSTM.....	68
4.9. Hasil Skenario Uji Coba	70
4.9.1. <i>Split Data</i>	71
4.9.2. <i>Learning Rate</i>	77
4.9.3. <i>Epoch</i>	85
4.9.4. <i>Optimizer</i>	93
4.9.5. <i>Unit</i>	98
4.10. Hasil dan Analisis Uji	102
4.11. Hasil Skenario Lanjutan.....	104
BAB V PENUTUP.....	105
5.1. Kesimpulan	105
5.2. Saran	105
DAFTAR PUSTAKA	107
Lampiran I.....	111
Biodata Penulis	118

DAFTAR TABEL

Tabel 2.1 Contoh Penerapan <i>Cleansing</i>	14
Tabel 2.2 Contoh Penerapan <i>Case-folding</i>	15
Tabel 2.3 Contoh Penerapan <i>Tokenizing</i>	15
Tabel 2.4 Contoh Penerapan <i>Normalization</i>	16
Tabel 2.5 Contoh Penerapan <i>Stopwords Removal</i>	16
Tabel 2.6 Contoh Penerapan <i>Stemming</i>	17
Tabel 2.7 Multi Class <i>Confusion Matrix</i>	24
Tabel 3.1 Proses <i>Cleansing</i>	31
Tabel 3.2 Proses <i>Case-folding</i>	32
Tabel 3.3 Proses <i>Tokenizing</i>	33
Tabel 3.4 Proses <i>Normalization</i>	34
Tabel 3.5 Proses <i>Stopwords Removal</i>	35
Tabel 3.6 Proses <i>Stemming</i>	37
Tabel 3.7 Contoh Dataset Positif	40
Tabel 3.8 Contoh Dataset Negatif.....	40
Tabel 3.9 <i>Tuning Hyperparameter Unit LSTM</i>	44
Tabel 3.10 <i>Tuning Hyperparameter Learning Rate</i>	45
Tabel 3.11 Contoh <i>Tuning Hyperparameter Pelatihan Model</i>	45
Tabel 3.12 <i>Tuning Hyperparameter Penelitian</i>	46
Tabel 4.1 Detail Spesifikasi Perangkat Keras	47
Tabel 4.2 Detail Spesifikasi Perangkat Lunak	48
Tabel 4.3 Hasil <i>Scraping</i> Data Pembatalan Piala Dunia.....	49
Tabel 4.4 Sentimen Sebelum Diproses	51
Tabel 4.5 Hasil Penerapan <i>Cleansing</i>	53
Tabel 4.6 Hasil Penerapan <i>Case-folding</i>	54
Tabel 4.7 Hasil Penerapan <i>Tokenizing</i>	55
Tabel 4.8 Hasil Penerapan <i>Normalization</i>	56
Tabel 4.9 Hasil Penerapan <i>Stopwords Removal</i>	58
Tabel 4.10 Hasil Penerapan <i>Stemming</i>	59

Tabel 4.11 Informasi <i>Dataset</i> Hasil Pelabelan VADER.....	62
Tabel 4.12 Informasi <i>Dataset</i> Final	65
Tabel 4.13 <i>Summary</i> Arsitektur Model LSTM	69
Tabel 4.14 Hasil Pengujian Terbaik Pada Perbandingan <i>Split Data</i>	76
Tabel 4.15 Hasil Pengujian Terbaik Pada <i>Learning Rate</i>	84
Tabel 4.16 Hasil Pengujian Terbaik Pada <i>Epoch</i>	93
Tabel 4.17 Hasil Pengujian Terbaik Pada <i>Optimizer</i>	97
Tabel 4.18 Hasil Pengujian Terbaik Pada Jumlah Unit	102
Tabel 4.19 Akurasi <i>Tuning Hyperparameter</i> Terbaik.....	102
Tabel 4.20 <i>Classification Report</i> Hasil <i>Tuning Hyperparameter</i> Terbaik.....	103

DAFTAR GAMBAR

Gambar 2.1	Arsitektur N-Grams Pada FastText	18
Gambar 2.2	Arsitektur <i>Recurrent Neural Network</i> (Firmansyah dkk., 2020)	19
Gambar 2.3	Arsitektur <i>Long Short Term Memory</i> (Priyanka dkk., 2023)	20
Gambar 2.4	Letak <i>Forget Gate</i> LSTM	21
Gambar 2.5	Letak <i>Input Gate</i> LSTM.....	22
Gambar 2.6	Letak <i>Output Gate</i> LSTM	23
Gambar 3.1	Tahapan Penelitian	26
Gambar 3.2	<i>Flowchart</i> Alur Sistem	27
Gambar 3.3	<i>Flowchart</i> Akuisisi Data	28
Gambar 3.4	<i>Scraping Tweet</i>	29
Gambar 3.5	<i>Flowchart</i> <i>Preprocessing</i> Data.....	30
Gambar 3.6	<i>Flowchart</i> <i>Cleansing</i>	31
Gambar 3.7	<i>Flowchart</i> <i>Case-folding</i>	32
Gambar 3.8	<i>Flowchart</i> <i>Tokenizing</i>	33
Gambar 3.9	<i>Flowchart</i> <i>Normalization</i>	34
Gambar 3.10	<i>Flowchart</i> <i>Stopwords Removal</i>	35
Gambar 3.11	<i>Flowchart</i> <i>Stemming</i>	36
Gambar 3.12	<i>Flowchart</i> Pelabelan Data	37
Gambar 3.13	Contoh Pelabelan VADER.....	38
Gambar 3.14	Mencari Nilai Sentimen Setiap Kata.....	38
Gambar 3.15	<i>Flowchart</i> Sortir Data	39
Gambar 3.16	<i>Flowchart</i> FastText <i>Word Embedding</i>	41
Gambar 3.17	Contoh N-gram Pada FastText.....	42
Gambar 3.18	Contoh <i>Embeddings</i> Vektor Kalimat Pada FastText.....	42
Gambar 3.19	<i>Flowchart</i> Klasifikasi Data	43
Gambar 3.20	<i>Flowchart</i> Pembuatan Model LSTM	43
Gambar 4.1	Potongan <i>Dataset</i> Hasil <i>Cleansing</i>	53
Gambar 4.2	Potongan <i>Dataset</i> Hasil <i>Case-folding</i>	54
Gambar 4.3	Potongan <i>Dataset</i> Hasil <i>Tokenizing</i>	55
Gambar 4.4	Potongan <i>Dataset</i> Hasil <i>Normalization</i>	57

Gambar 4.5 Potongan <i>Dataset</i> Hasil <i>Stopword Removal</i>	58
Gambar 4.6 Potongan <i>Dataset</i> Hasil <i>Stemming</i>	59
Gambar 4.7 <i>Dataset</i> Hasil Pelabelan Manual	60
Gambar 4.8 <i>Dataset</i> Hasil <i>Translate</i> Ke Bahasa Inggris	61
Gambar 4.9 <i>Dataset</i> Hasil Pelabelan	62
Gambar 4.10 Informasi <i>Dataset</i> Sortir <i>Tweet</i> Duplikat	63
Gambar 4.11 Informasi <i>Dataset</i> Sortir <i>Tweet</i> < 5 kata	64
Gambar 4.12 Informasi <i>Dataset</i> Sortir <i>Tweet</i> Netral	64
Gambar 4.13 Tampilan Hasil <i>Dataset</i> Final	65
Gambar 4.14 Hasil Vektor Kata Batal	66
Gambar 4.15 Hasil Vektor Kalimat Pertama	67
Gambar 4.16 <i>Output</i> Pelatihan Model LSTM.....	70
Gambar 4.17 Hasil Grafik Skenario <i>Split Data</i> 70:30	71
Gambar 4.18 Hasil Akurasi Skenario <i>Split Data</i> 70:30	72
Gambar 4.19 Hasil <i>Confusion Matrix</i> Skenario <i>Split Data</i> 70:30	72
Gambar 4.20 Hasil Grafik Skenario <i>Split Data</i> 80:20	73
Gambar 4.21 Hasil Akurasi Skenario <i>Split Data</i> 80:20	73
Gambar 4.22 Hasil <i>Confusion Matrix</i> Skenario <i>Split Data</i> 80:20	74
Gambar 4.23 Hasil Grafik Skenario <i>Split Data</i> 90:10	75
Gambar 4.24 Hasil Akurasi Skenario <i>Split Data</i> 90:10	75
Gambar 4.25 Hasil <i>Confusion Matrix</i> Skenario <i>Split Data</i> 90:10	76
Gambar 4.26 Hasil Grafik Skenario <i>Learning Rate</i> 0.001	77
Gambar 4.27 Hasil Akurasi Skenario <i>Learning Rate</i> 0.001.....	78
Gambar 4.28 Hasil <i>Confusion Matrix</i> Skenario <i>Learning Rate</i> 0.001	78
Gambar 4.29 Hasil Grafik Skenario <i>Learning Rate</i> 0.002.....	79
Gambar 4.30 Hasil Akurasi Skenario <i>Learning Rate</i> 0.002.....	79
Gambar 4.31 Hasil <i>Confusion Matrix</i> Skenario <i>Learning Rate</i> 0.002.....	80
Gambar 4.32 Hasil Grafik Skenario <i>Learning Rate</i> 0.003.....	81
Gambar 4.33 Hasil Akurasi Skenario <i>Learning Rate</i> 0.003.....	81
Gambar 4.34 Hasil <i>Confusion Matrix</i> Skenario <i>Learning Rate</i> 0.003.....	82
Gambar 4.35 Hasil Grafik Skenario <i>Learning Rate</i> 0.004.....	83
Gambar 4.36 Hasil Akurasi Skenario <i>Learning Rate</i> 0.004.....	83

Gambar 4.37 Hasil <i>Confusion Matrix</i> Skenario <i>Learning Rate</i> 0.004.....	84
Gambar 4.38 Hasil Grafik Skenario <i>Epoch</i> 15.....	85
Gambar 4.39 Hasil Akurasi Skenario <i>Epoch</i> 15	86
Gambar 4.40 Hasil <i>Confusion Matrix</i> Skenario <i>Epoch</i> 15.....	86
Gambar 4.41 Hasil Grafik Skenario <i>Epoch</i> 20.....	87
Gambar 4.42 Hasil Akurasi Skenario <i>Epoch</i> 20	88
Gambar 4.43 Hasil <i>Confusion Matrix</i> Skenario <i>Epoch</i> 20.....	88
Gambar 4.44 Hasil Grafik Skenario <i>Epoch</i> 25.....	89
Gambar 4.45 Hasil Akurasi Skenario <i>Epoch</i> 25	89
Gambar 4.46 Hasil <i>Confusion Matrix</i> Skenario <i>Epoch</i> 25.....	90
Gambar 4.47 Hasil Grafik Skenario <i>Epoch</i> 30.....	91
Gambar 4.48 Hasil Akurasi Skenario <i>Epoch</i> 30	91
Gambar 4.49 Hasil <i>Confusion Matrix</i> Skenario <i>Epoch</i> 30.....	92
Gambar 4.50 Hasil Grafik Skenario <i>Optimizer Adam</i>	94
Gambar 4.51 Hasil Akurasi Skenario <i>Optimizer Adam</i>	94
Gambar 4.52 Hasil <i>Confusion Matrix</i> Skenario <i>Optimizer Adam</i>	95
Gambar 4.53 Hasil Grafik Skenario <i>Optimizer RMSprop</i>	96
Gambar 4.54 Hasil Akurasi Skenario <i>Optimizer RMSprop</i>	96
Gambar 4.55 Hasil <i>Confusion Matrix</i> Skenario <i>Optimizer RMSprop</i>	97
Gambar 4.56 Hasil Grafik Skenario Unit 128.....	98
Gambar 4.57 Hasil Akurasi Skenario Unit 128	99
Gambar 4.58 Hasil <i>Confusion Matrix</i> Skenario Unit 128.....	99
Gambar 4.59 Hasil Grafik Skenario Unit 256.....	100
Gambar 4.60 Hasil Akurasi Skenario Unit 256	101
Gambar 4.61 Hasil <i>Confusion Matrix</i> Skenario Unit 256.....	101
Gambar 4.62 Hasil <i>Confusion Matrix Tuning Hyperparameter Terbaik</i>	103