

**PENGEMBANGAN MODUL U-TAPIS UNTUK
MENGKOREKSI KESALAHAN PENGGUNAAN TANDA
BACA MENGGUNAKAN *LONG SHORT-TERM MEMORY***



SKRIPSI

**ANDREA ZOE PUTRI SUKONCO
00000059793**

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

**PENGEMBANGAN MODUL U-TAPIS UNTUK
MENGKOREKSI KESALAHAN PENGGUNAAN TANDA
BACA MENGGUNAKAN *LONG SHORT-TERM MEMORY***



SKRIPSI

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

**ANDREA ZOE PUTRI SUKONCO
00000059793**

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

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Andrea Zoe Putri Sukonco
Nomor Induk Mahasiswa : 00000059793
Program Studi : Informatika

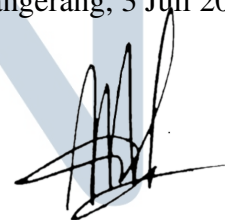
Skripsi dengan judul:

Pengembangan Modul U-Tapis untuk Mengkoreksi Kesalahan Penggunaan Tanda Baca Menggunakan *Long Short-Term Memory*

merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis 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 maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 3 Juli 2025



(Andrea Zoe Putri Sukonco)

UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PENGESAHAN

Skripsi dengan judul

PENGEMBANGAN MODUL U-TAPIS UNTUK MENGKOREKSI KESALAHAN PENGGUNAAN TANDA BACA MENGGUNAKAN *LONG SHORT-TERM MEMORY*

oleh

Nama : Andrea Zoe Putri Sukonco
NIM : 00000059793
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Jumat, 18 Juli 2025

Pukul 13.00 s/s 15.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut

Ketua Sidang

(Wirawan Istiono, S.Kom., M.Kom)

NIDN: 0313048304

Penguji

(SY.Yuliani., S.Kom., MT., PhD)

NIDN: 0411037904

Pembimbing

(Marlinda Vasty Overbeek, S.Kom, M.Kom)

NIDN: 0818038501

Ketua Program Studi Informatika,

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

NIDN: 0315109103

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini:

Nama : Andrea Zoe Putri Sukonco
NIM : 00000059793
Program Studi : Informatika
Jenjang : S1
Judul Karya Ilmiah : Pengembangan Modul U-Tapis untuk
Mengkoreksi Kesalahan Penggunaan
Tanda Baca Menggunakan *Long
Short-Term Memory*

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, 3 Juli 2025

Yang menyatakan



Andrea Zoe Putri Sukonco

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



UMN

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 PERSEMBAHAN / MOTTO

"We lived well today, so let's live well tomorrow. Today is not our end, there is tomorrow. I hope we can welcome that tomorrow."

Park Boyoung



KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa karena atas limpahan rahmat-Nya, penulis dapat menyelesaikan Tugas Akhir dengan judul "Pengembangan Modul U-Tapis untuk Mengkoreksi Kesalahan Penggunaan Tanda Baca Menggunakan Long Short-Term Memory" ini dengan baik dan lancar.

Tugas Akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Informatika. Penelitian ini dilatarbelakangi oleh pentingnya penggunaan tanda baca yang benar dalam komunikasi formal, akademik, maupun profesional. Kesalahan tanda baca, meskipun tampak sepele, dapat mengganggu pemahaman dan makna sebuah kalimat. Oleh karena itu, penelitian ini bertujuan untuk merancang sistem berbasis *deep learning* yang mampu mendeteksi dan mengoreksi kesalahan penggunaan tanda baca secara otomatis, sehingga dapat membantu penulis, editor, dan pengguna bahasa lainnya dalam menghasilkan teks dengan penggunaan tanda baca yang lebih baik.

Mengucapkan terima kasih Mengucapkan terima kasih

1. Bapak Dr. Ir. Andrey Andoko, M.Sc., 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, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Ibu Marlinda Vasty Overbeek, S.Kom, M.Kom, sebagai Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya skripsi ini.
5. Ibu Dr. Niknik Mediyawati, S.Pd., M.Hum. dan Bapak Samiaji Bintang Nusantara, S.T., M.A. yang telah menerima saya dalam tim U-Tapis serta memberikan kontribusi dalam validasi dataset dan aplikasi.
6. Keluarga saya yang telah memberikan dukungan tanpa henti, baik secara moral maupun material, selama proses pengerjaan skripsi ini.
7. Teman-teman saya yang telah memberikan semangat, kebersamaan, serta bantuan teknis maupun emosional dalam menyelesaikan tugas akhir ini.

8. Tim So G.oo.D NewKids; Kim Jiyong, Wu Hao, Jang Kyuhyun, Lee Junseo, dan Lee Taegeon yang telah merilis lagu So G.oo.D di saat saya merasa menyerah untuk menyelesaikan skripsi ini.

Semoga skripsi ini dapat bermanfaat bagi semua pihak.

Tangerang, 3 Juli 2025



Andrea Zoe Putri Sukonco



**PENGEMBANGAN MODUL U-TAPIS UNTUK MENGKOREKSI
KESALAHAN PENGGUNAAN TANDA BACA MENGGUNAKAN *LONG
SHORT-TERM MEMORY***

Andrea Zoe Putri Sukonco

ABSTRAK

Penelitian ini mengembangkan U-Tapis, sebuah sistem otomatis yang dirancang untuk mendeteksi dan mengoreksi kesalahan tanda baca berdasarkan kaidah Ejaan Yang Disempurnakan (EYD) edisi kelima. Sistem ini terdiri dari dua tahapan utama: tahap deteksi kesalahan tanda baca yang dibangun menggunakan arsitektur *Recurrent Neural Network* (RNN), serta tahap koreksi kesalahan yang mengandalkan arsitektur *Long Short-Term Memory* (LSTM). Dataset yang digunakan dalam penelitian ini terdiri dari 10.000 artikel dari situs Tribunnews.com dengan total 96.674 kalimat, yang telah melalui tahap prapemrosesan seperti tokenisasi. Model LSTM dirancang sebagai model klasifikasi biner dan dilatih menggunakan teknik optimasi berbasis pencarian acak untuk mendapatkan konfigurasi parameter terbaik. Model RNN terbukti efektif dalam mengidentifikasi posisi kesalahan dengan tingkat akurasi mencapai 92%. Selanjutnya, kesalahan yang terdeteksi dikoreksi secara otomatis oleh model LSTM, yang mampu menangkap ketergantungan jangka panjang dalam data sekuensial dan menghasilkan prediksi koreksi yang lebih akurat, dengan akurasi koreksi mencapai 86%. Temuan ini menunjukkan bahwa pendekatan berbasis *Deep Learning* dan *Natural Language Processing* (NLP) pada U-Tapis memiliki potensi besar dalam mendukung proses editorial, meningkatkan konsistensi penggunaan tanda baca, serta mengurangi intervensi manual dalam penyuntingan teks secara signifikan.

Kata kunci: Long Short-Term Memory, Natural Language Processing, Recurrent Neural Network, Tanda Baca, U-Tapis

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

DEVELOPMENT OF U-TAPIS' MODULE FOR CORRECTING PUNCTUATION ERRORS USING LONG SHORT-TERM MEMORY

Andrea Zoe Putri Sukonco

ABSTRACT

This study developed U-Tapis, an automated system designed to detect and correct punctuation errors based on the fifth edition of the Enhanced Spelling Guidelines (EYD). The system consists of two main stages: the punctuation error detection phase, built using a Recurrent Neural Network (RNN) architecture, and the error correction phase, which utilizes a Long Short-Term Memory (LSTM) architecture. The dataset used in this study consists of 10,000 articles from the Tribunnews.com website, totaling 96,674 sentences that have undergone preprocessing steps such as tokenization. The LSTM model was designed as a binary classification model and trained using a random search optimization technique to obtain the best parameter configuration. The RNN model has proven effective in identifying the location of errors with an accuracy of up to 92%. Subsequently, the detected errors are automatically corrected by the LSTM model, which is capable of capturing long-term dependencies in sequential data and producing more accurate correction predictions, achieving an accuracy of 86%. These findings demonstrate that a Deep Learning and Natural Language Processing (NLP)-based approach in U-Tapis holds great potential in supporting editorial processes, improving punctuation consistency, and significantly reducing manual intervention in text editing.

Keywords: Long Short-Term Memory, Natural Language Processing, Punctuation, Recurrent Neural Network, U-Tapis

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

DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
HALAMAN PERSEMBAHAN/MOTO	vi
KATA PENGANTAR	vii
ABSTRAK	ix
ABSTRACT	x
DAFTAR ISI	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR KODE	xv
DAFTAR RUMUS	xvii
DAFTAR LAMPIRAN	xviii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Permasalahan	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB 2 LANDASAN TEORI	5
2.1 Tanda Baca	5
2.1.1 Tanda Titik (.)	6
2.1.2 Tanda Koma (,)	7
2.1.3 Tanda Tanya (?)	8
2.1.4 Tanda Seru (!)	8
2.2 Deep Learning	9
2.3 Natural Language Processing (NLP)	10
2.4 Recurrent Neural Network (RNN)	10
2.5 Long Short-Term Memory (LSTM)	12
2.6 Long Short-Term Memory (LSTM)	12
2.7 Dot Product	15
2.8 Random Search	17
2.9 Rectified Linear Unit (ReLU)	17
2.10 Sigmoid	17
2.11 Receiver Operating Characteristic (ROC) Curve	18
2.11.1 Luas Area di Bawah Kurva (AUC)	18
2.11.2 Keunggulan ROC Curve Menurut Penelitian Sebelumnya	19
2.11.3 Implementasi pada Penelitian Ini	19
2.12 Classification Report	19
BAB 3 METODOLOGI PENELITIAN	21
3.1 Tahapan Penelitian	21
3.2 Deteksi	21
3.2.1 Deteksi: Pengumpulan data	22
3.2.2 Deteksi: Praproses Data	23
3.2.3 Pembersihan data	24
3.2.4 Deteksi: Pembangunan Fitur	26

3.2.5	<i>Deteksi: Fold Dataset</i>	28
3.2.6	<i>Deteksi: Pemodelan</i>	30
3.3	Koreksi	33
3.3.1	Koreksi: Pembangunan Fitur	34
3.3.2	Koreksi: Fold Dataset	35
3.3.3	Koreksi: Pemodelan	36
3.3.4	Koreksi: Evaluasi Model	38
BAB 4	HASIL DAN DISKUSI	40
4.1	Deteksi	40
4.1.1	Deteksi: Pengumpulan Data	40
4.1.2	Deteksi: Praproses Data	41
4.1.3	Deteksi: Pembangunan Fitur	52
4.1.4	Deteksi: <i>Fold Dataset</i>	72
4.1.5	Deteksi: Pemodelan	74
4.1.6	Deteksi: Evaluasi Model	81
4.2	Koreksi	85
4.2.1	Koreksi: Pembangunan Fitur	85
4.2.2	Koreksi: <i>Fold Dataset</i>	99
4.2.3	Koreksi: Pemodelan	102
4.2.4	Koreksi: Evaluasi Model	106
4.3	Integrasi API	109
BAB 5	SIMPULAN DAN SARAN	113
5.1	Simpulan	113
5.2	Saran	114
SARAN		114
DAFTAR PUSTAKA		115



DAFTAR TABEL

Tabel 4.1	Distribusi Jenis Kesalahan Tanda Baca	68
Tabel 4.2	Aturan Penggunaan Tanda Baca dan Jumlah Data	70
Tabel 4.3	Hasil Random Search Hyperparameter untuk VanillaRNN	77
Tabel 4.4	Hasil Pelatihan Model (58Epoch)	78
Tabel 4.5	Ambang Batas (Threshold) Terbaik per Jenis Kesalahan	82
Tabel 4.6	Hasil Evaluasi Model Deteksi Kesalahan Tanda Baca	83
Tabel 4.7	Hasil Randomsearch hyperparameter model LSTM	104
Tabel 4.8	Cuplikan nilai threshold beserta metrik TPR, FPR, dan selisih TPR–FPR Best threshold = 0.86641157	108
Tabel 4.9	Laporan klasifikasi deteksi kesalahan tanda baca	108
Tabel 4.10	Deteksi Kesalahan Berdasarkan Rule-Based	111



DAFTAR GAMBAR

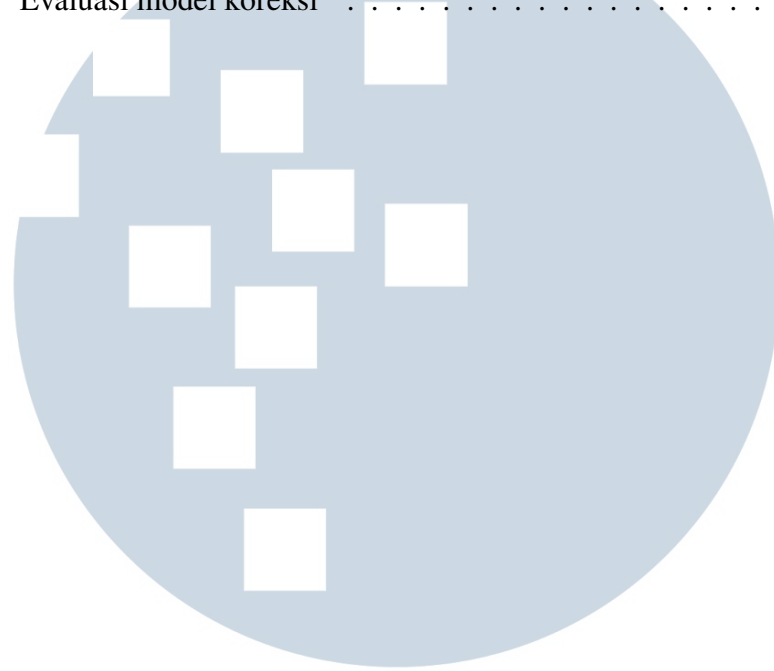
Gambar 2.1	Ilustrasi konsep Deep Learning	9
Gambar 2.2	Arsitektur Recurrent Neural Network	10
Gambar 2.3	Ilustrasi arsitektur Long Short-Term Memory (LSTM)	12
Gambar 3.1	Metodologi Penelitian	21
Gambar 3.2	Deteksi	22
Gambar 3.3	Urutan <i>Praproses Data</i>	23
Gambar 3.4	Urutan Variasi Data	24
Gambar 3.5	Urutan <i>Feature Engineering</i>	26
Gambar 3.6	Urutan <i>Fold dataset</i>	29
Gambar 3.7	<i>Data Splitting</i>	29
Gambar 3.8	Urutan <i>Modelling</i>	30
Gambar 3.9	Urutan <i>Stack Model</i>	31
Gambar 3.10	Urutan <i>Parameter Tuning</i>	31
Gambar 3.11	<i>Model Fitting</i>	31
Gambar 3.12	Alur metodologi koreksi tanda baca	33
Gambar 3.13	<i>Feature Engineering</i> Koreksi	34
Gambar 3.14	<i>Fold Dataset</i> Koreksi	35
Gambar 3.15	<i>Modelling</i> Koreksi	36
Gambar 3.16	<i>Stack Modelling</i> Koreksi	36
Gambar 3.17	Urutan <i>Parameter Tuning</i>	37
Gambar 3.18	<i>Model Fitting untuk Koreksi Tanda Baca</i>	37
Gambar 4.1	Grafik <i>Best Fitting</i>	81
Gambar 4.2	Distribusi kelas koreksi	101
Gambar 4.3	Distribusi kelas koreksi setelah SMOTEENN	102
Gambar 4.4	Grafik <i>Modelling</i> SMOTEENN	106



DAFTAR KODE

Kode 4.1	Menggabungkan dataset ke satu <i>file</i> *.csv	40
Kode 4.2	Membersihkan dataset	41
Kode 4.3	Memecah dan memvariasikan dataset	43
Kode 4.4	Memecah dataset berdasarkan urutan	46
Kode 4.5	Membuat <i>product</i> dataset dan memilih acak	49
Kode 4.6	Impor dataset	52
Kode 4.7	Inisialisasi pola jalan	53
Kode 4.8	Inisialisasi pola gelar	53
Kode 4.9	Inisialisasi kata baku	54
Kode 4.10	Inisialisasi pola penggunaan tanda baca	54
Kode 4.11	Inisialisasi pola pertanyaan	56
Kode 4.12	Inisialisasi pola dan/atau dan kalau/karena/agar	57
Kode 4.13	Inisialisasi pola kalimat langsung	58
Kode 4.14	Inisialisasi pola tetapi/melainkan/sedangkan	58
Kode 4.15	Inisialisasi pola oleh karena itu/jadi/meskipun demikian	59
Kode 4.16	Inisialisasi pola dalam/atas	59
Kode 4.17	Ekstrak fitur tanda titik, tanya, dan seru	60
Kode 4.18	Ekstrak fitur tanda koma	62
Kode 4.19	Membuat label untuk masing-masing kelas	64
Kode 4.20	Menerapkan label pada dataset	67
Kode 4.21	Menghitung jumlah dataset setiap label	67
Kode 4.22	Menggabungkan dataset utama dengan data sintesis	69
Kode 4.23	Membuat label vektor	71
Kode 4.24	<i>Fold</i> Dataset	72
Kode 4.25	Pembagian dataset	73
Kode 4.26	Ukuran dari masing-masing dataset	73
Kode 4.27	<i>Hyperparameter tuning</i> menggunakan <i>Random Search</i>	74
Kode 4.28	<i>Best model fitting</i>	77
Kode 4.29	<i>Plotting</i> hasil pelatihan model	80
Kode 4.30	Evaluasi model deteksi	81
Kode 4.31	Evaluasi model deteksi	84
Kode 4.32	Impor dataset utama dan data pendukung untuk proses koreksi	86
Kode 4.33	Memperbaiki tanda baca di akhir kalimat	86
Kode 4.34	Memperbaiki tanda tanya pada kalimat tanya	87
Kode 4.35	Memperbaiki penggunaan huruf kecil setelah tanda baca	88
Kode 4.36	Memperbaiki huruf kapital yang sebelumnya tidak ada tanda baca	88
Kode 4.37	Memperbaiki tanda seru lebih dari satu	90
Kode 4.38	Memperbaiki koma dan/atau	90
Kode 4.39	Memperbaiki koma karena/kalau/agar	91
Kode 4.40	Memperbaiki koma pada kalimat langsung	92
Kode 4.41	Memperbaiki koma pada tetapi/melainkan/sedangkan	93
Kode 4.42	Memperbaiki koma pada oleh karena itu/jadi/meskipun demikian	93
Kode 4.43	Memperbaiki koma pada nama jalan	94
Kode 4.44	Memperbaiki penggunaan koma pada gelar	96
Kode 4.45	Memperbaiki koma pada dalam/atas	96
Kode 4.46	Perbaikan tulisan	97
Kode 4.47	<i>Fold dataset</i> koreksi	99
Kode 4.48	Membagi dataset modeling koreksi	100

Kode 4.49	<i>Plotting</i> dataset	100
Kode 4.50	SMOTEENN	101
Kode 4.51	<i>Hyperparameter tuning</i> model koreksi	103
Kode 4.52	<i>Best model fitting dan plotting</i>	105
Kode 4.53	Prediksi hasil tes	107
Kode 4.54	Evaluasi model koreksi	108



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

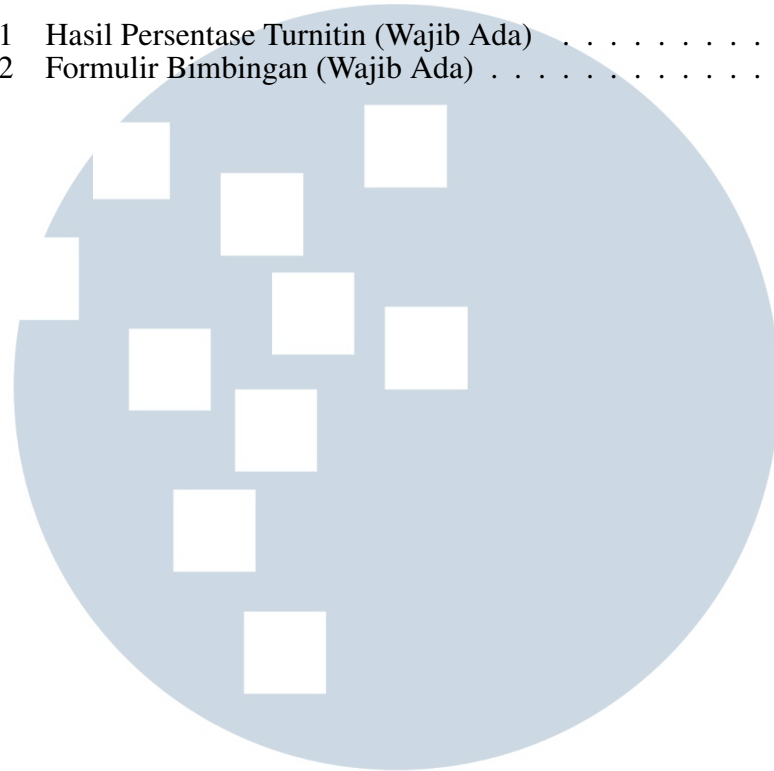
DAFTAR RUMUS

Rumus 2.1	Persamaan umum RNN	11
Rumus 2.7	Logika <i>cell state</i>	13
Rumus 2.9	Persamaan umum <i>Dot Product</i>	15
Rumus 2.10	Persamaan <i>Dot Product</i> antar klaster	16
Rumus 2.11	Contoh perhitungan <i>Dot Product</i> antar klaster	16
Rumus 2.12	Persamaan umum ReLU	17
Rumus 2.13	Persamaan umum Sigmoid	17
Rumus 2.15	Rumus TPR dan FPR pada ROC Curve	18
Rumus 2.16	Persamaan umum Precision	19
Rumus 2.17	Persamaan umum Recall	19
Rumus 2.18	Persamaan umum F1-Score	19



DAFTAR LAMPIRAN

Lampiran 1	Hasil Persentase Turnitin (Wajib Ada)	119
Lampiran 2	Formulir Bimbingan (Wajib Ada)	133



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA