

**PENGEMBANGAN MODUL U-TAPIS UNTUK MENDETEKSI  
KESALAHAN PENGGUNAAN TANDA BACA MENGGUNAKAN  
RECURRENT NEURAL NETWORK**



**LAPORAN MBKM PENELITIAN**

**Andrea Zoe Putri Sukonco  
00000059793**

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

**PENGEMBANGAN MODUL U-TAPIS UNTUK MENDETEKSI  
KESALAHAN PENGGUNAAN TANDA BACA MENGGUNAKAN  
RECURRENT NEURAL NETWORK**



**LAPORAN MBKM PENELITIAN**

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

**Andrea Zoe Putri Sukonco**

**00000059793**

**UMN**

**UNIVERSITAS**

**MULTIMEDIA**

**NUSANTARA**

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

**TANGERANG**

**2025**

## HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Andrea Zoe Putri Sukonco  
NIM : 00000059793  
Program Studi : Informatika

Menyatakan dengan sesungguhnya bahwa Tesis/Skripsi/Tugas Akhir/Laporan Magang/MBKM saya yang berjudul:

**Pengembangan Modul U-Tapis Untuk Mendeteksi Kesalahan Penggunaan Tanda Baca Menggunakan Recurrent Neural Network**

merupakan hasil karya saya sendiri, bukan merupakan hasil plagiat, dan tidak pula dituliskan oleh orang lain; Semua sumber, baik yang dikutip maupun dirujuk, telah saya cantumkan dan nyatakan dengan benar pada bagian Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan karya ilmiah, saya bersedia menerima konsekuensi untuk dinyatakan TIDAK LULUS. Saya juga bersedia menanggung segala konsekuensi hukum yang berkaitan dengan tindak plagiarisme ini sebagai kesalahan saya pribadi dan bukan tanggung jawab Universitas Multimedia Nusantara.

Tangerang, 3 Januari 2025



(Andrea Zoe Putri Sukonco)

**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  
Jenis Karya : Laporan MBKM Penelitian

Menyatakan dengan sesungguhnya bahwa:

- Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya di repositori Knowledge Center, sehingga dapat diakses oleh Civitas Akademika/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial dan saya juga tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.
- Saya tidak bersedia karena dalam proses pengajuan untuk diterbitkan ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*)\*\*.

Tangerang, 3 Januari 2025

Yang menyatakan



Andrea Zoe Putri Sukonco

U M M N  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

\*\* Jika tidak bisa membuktikan LoA jurnal/HKI selama enam bulan ke depan, saya bersedia mengizinkan penuh karya ilmiah saya untuk diunggah ke KC UMN dan menjadi hak institusi UMN.

**Halaman Persembahan / Motto**

”Karena semua butuh usaha”

Unknown



**UMMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA



## KATA PENGANTAR

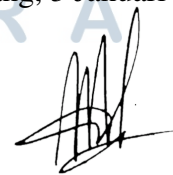
Puji syukur saya panjatkan ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga Tugas Akhir dengan judul "Pengembangan Modul U-Tapis Untuk Mendeteksi Kesalahan Penggunaan Tanda Baca Menggunakan *Recurrent Neural Network*" dapat diselesaikan dengan baik. Penulisan laporan MBKM Penelitian ini bertujuan untuk mengembangkan modul yang mampu mendeteksi kesalahan tanda baca pada teks menggunakan pendekatan *Recurrent Neural Network* (RNN).

Saya mengucapkan terima kasih kepada pihak-pihak yang telah memberikan dukungan dan bantuan selama proses penyelesaian tugas akhir ini:

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 Assoc. Prof. Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Ibu Eunike Endariahna Surbakti, S.Kom., M.T.I., sebagai Pembimbing pertama yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya tesis ini.
5. Ibu Marlinda Varsty Overbeek, S.Kom., M.Kom., sebagai Dosen Penelitian yang telah banyak membantu dan memberikan bimbingan atas terselesainya Penelitian ini.
6. Orang Tua, keluarga, teman-teman, dan hewan peliharaan saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan penelitian ini.

Semoga penelitian ini dapat menjadi manfaat bagi kita semua.

Tangerang, 3 Januari 2025



Andrea Zoe Putri Sukonco

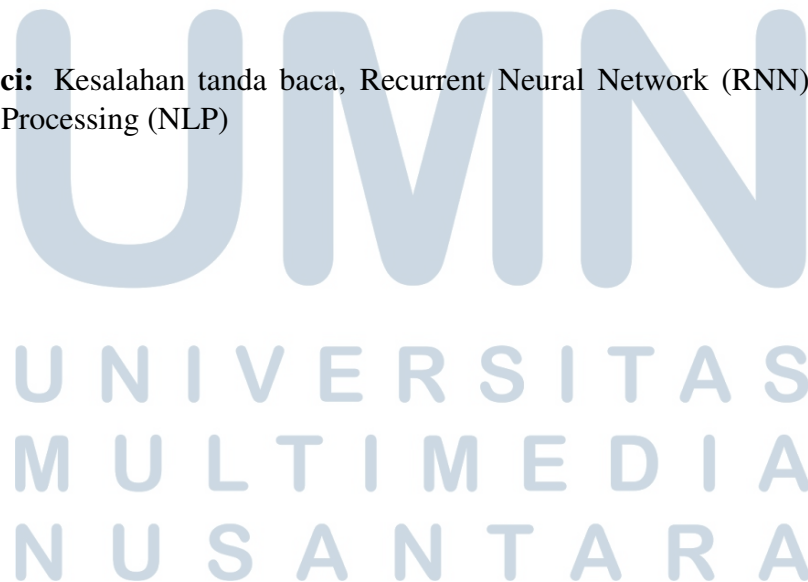
# PENGEMBANGAN MODUL U-TAPIS UNTUK MENDETEKSI KESALAHAN PENGGUNAAN TANDA BACA MENGGUNAKAN RECURRENT NEURAL NETWORK

Andrea Zoe Putri Sukonco

## ABSTRAK

Penelitian ini mengeksplorasi pengembangan U-Tapis, sebuah alat yang dirancang untuk secara otomatis mendeteksi dan memperbaiki kesalahan tanda baca berdasarkan edisi kelima Ejaan yang Disempurnakan (EYD V). Dengan memanfaatkan *Recurrent Neural Network* (RNN), sebuah model pembelajaran mendalam yang cocok untuk data sekuensial, U-Tapis secara efisien mengidentifikasi kesalahan tanda baca dalam teks Bahasa Indonesia. Penelitian ini menggunakan dataset yang terdiri dari 1.035 artikel dari Tribunnews.com, mencakup 9.930 kalimat, untuk pelatihan dan evaluasi. Teknik pra-pemrosesan data seperti tokenisasi digunakan untuk mempersiapkan teks sebelum analisis. Model RNN yang digunakan menunjukkan akurasi sebesar 69%, yang menunjukkan potensi untuk meningkatkan efisiensi dalam deteksi kesalahan tanda baca, meskipun masih memerlukan penyempurnaan lebih lanjut untuk mencapai akurasi yang lebih tinggi. Hasil penelitian menunjukkan bagaimana *Deep Learning* dan *Natural Language Processing* (NLP) dapat menyederhanakan proses editorial dengan mengotomatiskan deteksi kesalahan, memastikan konsistensi bahasa, dan mengurangi upaya manual dalam jurnalisme dan bidang lain yang bergantung pada teks.

**Kata kunci:** Kesalahan tanda baca, Recurrent Neural Network (RNN), Natural Language Processing (NLP)



***Development of the U-Tapis Module for Detecting Punctuation Misuse Using  
Recurrent Neural Network***

Andrea Zoe Putri Sukonco

***ABSTRACT***

*This study explores the development of U-Tapis, a tool designed to automatically detect and correct punctuation errors based on the fifth edition of Ejaan yang Disempurnakan (EYD V). By utilizing a Recurrent Neural Network (RNN), a deep learning model suitable for sequential data, U-Tapis efficiently identifies punctuation errors in Indonesian text. The study employs a dataset comprising 1,035 articles from Tribunnews.com, covering 9,930 sentences, for training and evaluation. Data preprocessing techniques, such as tokenization, were applied to prepare the text for analysis. The RNN model achieved an accuracy of 69%, indicating its potential to enhance efficiency in punctuation error detection, although further refinement is needed to achieve higher accuracy. The findings demonstrate how Deep Learning and Natural Language Processing (NLP) can simplify editorial processes by automating error detection, ensuring language consistency, and reducing manual effort in journalism and other text-dependent fields.*

***Keywords:*** *Punctuation Errors, Recurrent Neural Network (RNN), Natural Language Processing (NLP)*





## DAFTAR ISI

HALAMAN JUDUL . . . . .	i
HALAMAN PERNYATAAN TIDAK PLAGIAT . . . . .	ii
HALAMAN PERSETUJUAN PUBLIKASI ILMIAH . . . . .	iii
HALAMAN PERSEMBAHAN/MOTO . . . . .	iv
KATA PENGANTAR . . . . .	v
ABSTRAK . . . . .	vi
ABSTRACT . . . . .	vii
DAFTAR ISI . . . . .	viii
DAFTAR GAMBAR . . . . .	x
DAFTAR TABEL . . . . .	xi
DAFTAR KODE . . . . .	xii
DAFTAR LAMPIRAN . . . . .	xiii
BAB 1 PENDAHULUAN . . . . .	1
1.1 Latar Belakang Masalah . . . . .	1
1.2 Rumusan Masalah . . . . .	2
1.3 Tujuan Penelitian . . . . .	3
1.4 Urgensi Penelitian . . . . .	3
1.5 Luaran Penelitian . . . . .	3
1.6 Manfaat Penelitian . . . . .	3
BAB 2 TINJAUAN PUSTAKA . . . . .	4
2.1 Tanda Baca . . . . .	4
2.1.1 Tanda Titik (.) . . . . .	5
2.1.2 Tanda Koma (,) . . . . .	5
2.1.3 Tanda Tanya (?) . . . . .	7
2.1.4 Tanda Seru (!) . . . . .	7
2.2 Deep Learning . . . . .	7
2.3 Natural Language Processing (NLP) . . . . .	8
2.4 Recurrent Neural Network (RNN) . . . . .	8
2.5 Permutasi . . . . .	10
BAB 3 METODOLOGI PENELITIAN . . . . .	11
3.1 <i>Data Collection</i> . . . . .	11
3.1.1 <i>Dataset</i> Utama . . . . .	11
3.1.2 <i>Dataset</i> Pendukung . . . . .	11
3.2 <i>Data Preprocessing</i> . . . . .	12
3.3 <i>Data Preprocessing</i> . . . . .	12
3.3.1 <i>Data Cleaning</i> . . . . .	12
3.3.2 <i>Variate Dataset</i> . . . . .	13
3.4 <i>Data Variation</i> . . . . .	13
3.4.1 <i>Random Order Dataset</i> . . . . .	13
3.4.2 <i>Split Dataset</i> . . . . .	13
3.4.3 <i>Permute Dataset</i> . . . . .	14
3.4.4 <i>Select Data</i> . . . . .	14
3.4.5 <i>Drop Columns</i> . . . . .	14
3.5 <i>Feature Engineering</i> . . . . .	15
3.5.1 <i>Rule-Based</i> . . . . .	15
3.5.2 <i>Extract Rule-Based</i> . . . . .	16
3.5.3 <i>Generate Error Description</i> . . . . .	17
3.5.4 <i>Generate Error Label</i> . . . . .	17

3.6	<i>Fold Dataset</i>	17
3.6.1	<i>Split Dataset</i>	18
3.6.2	<i>Tokenization</i>	18
3.6.3	<i>Pad Sequence</i>	18
3.6.4	<i>Balance Dataset</i>	18
3.7	<i>Modelling</i>	19
3.8	<i>Evaluation</i>	20
BAB 4	HASIL DAN PEMBAHASAN	22
BAB 5	SIMPULAN DAN SARAN	51
5.1	Simpulan	51
5.2	Saran	51
DAFTAR PUSTAKA		52



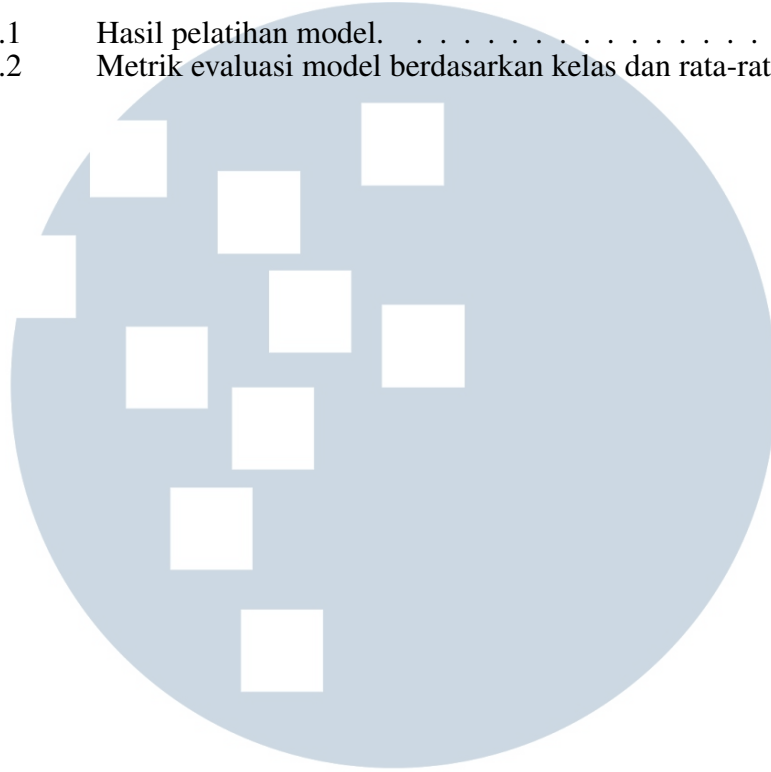
## DAFTAR GAMBAR

Gambar 2.1	Neuron buatan pada <i>deep learning</i> . . . . .	7
Gambar 2.2	Proses <i>recurrent neural network</i> . . . . .	8
Gambar 3.1	Metodologi Penelitian . . . . .	11
Gambar 3.2	Urutan <i>Data Preprocessing</i> . . . . .	12
Gambar 3.3	Urutan Variasi Data . . . . .	13
Gambar 3.4	Urutan <i>Feature Engineering</i> . . . . .	15
Gambar 3.5	Urutan <i>Feature Engineering</i> . . . . .	17
Gambar 3.6	Urutan <i>Modelling</i> . . . . .	19
Gambar 4.1	Urutan Variasi Data . . . . .	40
Gambar 4.2	Urutan Variasi Data . . . . .	43



## DAFTAR TABEL

Tabel 4.1	Hasil pelatihan model. . . . .	45
Tabel 4.2	Metrik evaluasi model berdasarkan kelas dan rata-rata. . .	47



**UMMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## DAFTAR KODE

4.1	Konversi ke .csv . . . . .	23
4.2	Memisahkan teks ke dalam kalimat . . . . .	24
4.3	Normalisasi teks pada dataset . . . . .	25
4.4	Membuat dataset kombinasi acak dan menggabungkan kolom . . . . .	25
4.5	Membuat kolom kombinasi dan memilih data secara acak . . . . .	26
4.6	Memuat dataset dan mengidentifikasi pola-pola penting . . . . .	27
4.7	Ekstraksi fitur berbasis tanda baca . . . . .	29
4.8	Ekstraksi fitur berbasis penomoran . . . . .	31
4.9	Ekstraksi fitur berbasis koma dalam kalimat . . . . .	33
4.10	Deteksi dan pemberian label kesalahan . . . . .	37
4.11	Visualisasi distribusi data . . . . .	39
4.12	Tokenisasi dan padding teks untuk model . . . . .	41
4.13	Split Dataset . . . . .	41
4.14	Implementasi SMOTETomek . . . . .	42
4.15	Visualisasi distribusi kelas "Errors" dan "No Errors" . . . . .	43
4.16	Implementasi model SimpleRNN . . . . .	43
4.17	Implementasi <i>Early Stopping</i> . . . . .	45
4.18	Threshold optimal . . . . .	45
4.19	Classification report . . . . .	46
4.20	Save model h5 . . . . .	47
4.21	Flask API untuk model . . . . .	47
4.22	Kode Client untuk Mengirim File ke Endpoint Flask . . . . .	50





## DAFTAR LAMPIRAN

Lampiran 1	MBKM-01 Cover Letter MBKM Internship Track 1 . . . . .	55
Lampiran 2	MBKM-02 MBKM Internship Track 1 Card . . . . .	56
Lampiran 3	MBKM-03 Daily Task - Internship Track 1 . . . . .	57
Lampiran 4	MBKM-04 Verification Form of Internship Report MBKM Internship Track 1 . . . . .	69
Lampiran 5	Form Bimbingan . . . . .	70
Lampiran 6	Memorandum of Understanding U-Tapis: Universitas Multimedia Nusantara dan Tribunnews.com . . . . .	71
Lampiran 7	Draft Paper SINTA . . . . .	74
Lampiran 8	Turnitin . . . . .	82

