

**U-TAPIS: DETEKSI KESALAHAN PENULISAN PARTIKEL
BAHASA INDONESIA MENGGUNAKAN XGBOOST DAN
RULE-BASED SYSTEM**



LAPORAN RESEARCH TECHNOLOGY PROGRAM

**FRILIYANDRE
00000068633**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2026**

**U-TAPIS: DETEKSI KESALAHAN PENULISAN PARTIKEL
BAHASA INDONESIA MENGGUNAKAN XGBOOST DAN
RULE-BASED SYSTEM**



LAPORAN RESEARCH TECHNOLOGY PROGRAM

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

**FRILIYANDRE
00000068633**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2026**

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Friliyandre
Nomor Induk Mahasiswa : 00000068633
Program Studi : Informatika

Skripsi dengan judul:

U-Tapis: Deteksi Kesalahan Penulisan Partikel Bahasa Indonesia Menggunakan XGBoost dan Rule-Based System

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, 02 Januari 2026



(Friliyandre)

UMM
UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PERNYATAAN PENGGUNAAN BANTUAN KECERDASAN ARTIFISIAL (AI)

Saya yang bertanda tangan di bawah ini:

Nama Lengkap : Friliyandre
NIM : 00000068633
Program Studi : Informatika
Judul Laporan : U-Tapis: Deteksi Kesalahan Penulisan Partikel Bahasa Indonesia Menggunakan XGBoost dan Rule-Based System

Dengan ini saya menyatakan secara jujur menggunakan bantuan Kecerdasan Artifisial (AI) dalam pengerjaan Laporan sebagai berikut:

- ☐ Menggunakan AI sebagaimana diizinkan untuk membantu dalam menghasilkan ide-ide utama saja.
- ☐ Menggunakan AI sebagaimana diizinkan untuk membantu menghasilkan teks pertama saja.
- ☒ Menggunakan AI untuk menyempurnakan sintaksis dan tata bahasa untuk pengumpulan tugas.
- ☐ Karena tidak diizinkan: Tidak menggunakan bantuan AI dengan cara apa pun dalam pembuatan tugas

Saya juga menyatakan bahwa:

- (1) Menyerahkan secara lengkap dan jujur penggunaan perangkat AI yang diperlukan dalam tugas melalui Formulir Penggunaan Perangkat Kecerdasan Artifisial (AI).
- (2) Mengakui telah menggunakan bantuan AI dalam tugas saya baik dalam bentuk kata, parafrase, penyertaan ide, atau fakta penting yang disarankan oleh AI dan telah dicantumkan dalam sitasi serta referensi.
- (3) Terlepas dari pernyataan di atas, tugas ini sepenuhnya merupakan karya saya sendiri.

Tangerang, 7 Januari 2026



(Friliyandre)

HALAMAN PERNYATAAN KEABSAHAN PERUSAHAAN

Nama : Friliyandre
NIM : 00000068633
Program Studi : Informatika
Fakultas : Teknik & Informatika

Dengan ini menyatakan bahwa saya melaksanakan kegiatan di:

Nama Perusahaan/Organisasi : LPPM Universitas Multimedia Nusantara

Alamat : Jalan Scientia Boulevard
Gading, Curug Sangereng,
Serpong, Kabupaten
Tangerang, Banten 15810

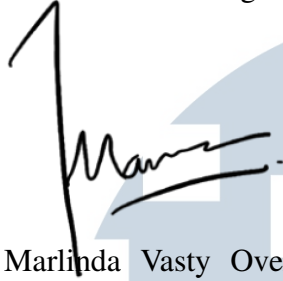
Email Perusahaan/Organisasi : lppm@umn.ac.id

1. Perusahaan/Organisasi tempat saya melakukan kegiatan dapat di validasi keberadaannya.
2. Jika dikemudian hari, terbukti ditemukan kecurangan/penyimpangan data yang tidak valid di perusahaan/organisasi tempat saya melakukan kegiatan, maka:
 - (a) Saya bersedia menerima konsekuensi dinyatakan **TIDAK LULUS** untuk mata kuliah yang telah saya tempuh.
 - (b) Saya bersedia menerima semua sanksi yang berlaku sebagaimana ditetapkan dalam peraturan yang berlaku di Universitas Multimedia Nusantara.

Pernyataan ini saya buat dengan sebenar-benarnya dan digunakan sebagaimana mestinya.

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

Mengetahui,
Dosen Pembimbing



Marinda Vasty Overbeek, S.Kom,
M.Kom

Tangerang, 02 Januari 2026

Yang menyatakan,



Friliyandre



**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Yang bertanda tangan di bawah ini:

Nama : Friliyandre
NIM : 00000068633
Program Studi : Informatika
Jenjang : S1
Judul Karya Ilmiah : U-Tapis: Deteksi Kesalahan
Penulisan Partikel Bahasa Indonesia
Menggunakan XGBoost dan Rule-
Based System

Menyatakan dengan sesungguhnya bahwa saya bersedia:

- ☒ 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, 02 Januari 2026

Yang menyatakan



Friliyandre

****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

"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 ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga Tugas Akhir berjudul "U-TAPIS: Deteksi Kesalahan Penulisan Partikel Bahasa Indonesia Menggunakan Algoritma XGBoost dan Rule-Based System" dapat diselesaikan. Penelitian ini disusun sebagai salah satu syarat kelulusan Program Studi Informatika serta sebagai upaya pengembangan pemahaman dan kemampuan penelitian di bidang *Natural Language Processing* (NLP).

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. Bapak Dr. Ivransa Zuhdi Pane, M.Eng., B.CS., sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Ibu Marlinda Vasty Overbeek, S.Kom, M.Kom, sebagai Pembimbing kedua yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
6. Ibu Dr. Niknik M. Kuntarto, M. Si. dan Bapak Samiaji Bintang Nusantara S.T., M.A., sebagai Pakar bahasa yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
7. Bapak Rudi Sutomo, S.Kom., M.Si., M.Kom., sebagai Pengembang *website* yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
8. Bapak Moeljono Widjaja, B.Sc., M.Sc., Ph.D., sebagai Pakar data analisis yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.

9. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Semoga laporan penelitian ini dapat bermanfaat, baik bagi sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 02 Januari 2026



Friliyandre



U-TAPIS: DETEKSI KESALAHAN PENULISAN PARTIKEL BAHASA INDONESIA MENGGUNAKAN XGBOOST DAN RULE-BASED SYSTEM

Friliyandre

ABSTRAK

Kesalahan penulisan partikel Bahasa Indonesia, seperti *pun*, *per*, *-lah*, dan *-kah*, masih ditemukan pada teks jurnalistik meskipun telah diatur dalam Pedoman Umum Ejaan Bahasa Indonesia (PUEBI). Penelitian ini mengusulkan sistem U-TAPIS untuk mendeteksi kesalahan penulisan partikel menggunakan kombinasi algoritma XGBoost dan *rule-based system*. Dataset dibangun dari korpus berita daring berbahasa Indonesia dengan total 1.060.341 kalimat yang mencakup sembilan kategori kesalahan dan satu kategori penulisan benar. Teks diproses melalui tahapan pra-pemrosesan dan rekayasa fitur linguistik manual. Model dilatih menggunakan GridSearchCV dengan metrik F1-Macro dan menghasilkan akurasi 99,33% pada data uji. Model kemudian diintegrasikan ke dalam API berbasis Flask untuk mendukung deteksi kesalahan partikel secara otomatis.

Kata kunci: deteksi kesalahan gramatikal, partikel Bahasa Indonesia, XGBoost, Natural Language Processing



***U-TAPIS: DETECTION OF INDONESIAN PARTICLE WRITING ERRORS
USING THE XGBOOST AND RULE-BASED SYSTEM***

Friliyandre

ABSTRACT

Errors in the use of Indonesian particles, such as pun, per, -lah, and -kah, are still found in journalistic texts despite being clearly regulated in the Pedoman Umum Ejaan Bahasa Indonesia (PUEBI). This study proposes U-TAPIS, a system for detecting particle usage errors by combining the XGBoost algorithm with a rule-based system. The dataset was constructed from an Indonesian online news corpus consisting of 1,060,341 sentences, covering nine categories of particle errors and one category of correct usage. The text data were processed through preprocessing and manual linguistic feature engineering. The model was trained using GridSearchCV with the F1-Macro metric and achieved an accuracy of 99.33% on the test set. The trained model was subsequently integrated into a Flask-based API to support automatic detection of Indonesian particle usage errors.

Keywords: grammatical error detection, Indonesian particles, Natural Language Processing, XGBoost



DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
PERNYATAAN PENGGUNAAN BANTUAN AI	iii
HALAMAN PERNYATAAN KEABSAHAN PERUSAHAAN	iv
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	vi
HALAMAN PERSEMBAHAN/MOTO	viii
KATA PENGANTAR	ix
ABSTRAK	xi
ABSTRACT	xii
DAFTAR ISI	xiii
DAFTAR TABEL	xv
DAFTAR GAMBAR	xvi
DAFTAR RUMUS	xvii
DAFTAR LAMPIRAN	xviii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian	2
1.4 Urgensi Penelitian	3
1.5 Luaran Penelitian	3
1.6 Manfaat Penelitian	3
BAB 2 LANDASAN TEORI	5
2.1 Deteksi Kesalahan Gramatikal	5
2.2 Partikel dalam Bahasa Indonesia	5
2.3 Natural Language Preprocessing (NLP)	6
2.4 Extreme Gradient Boosting (XGBoost)	6
2.5 Rule-Based System	10
2.6 Evaluasi Model	11
BAB 3 METODOLOGI PENELITIAN	13
3.1 Tahapan Penelitian	13
3.2 Studi Literatur	13
3.3 Pengumpulan Data	13
3.3.1 Penandaan Bagian Kalimat yang Salah	14
3.3.2 Penggabungan dan Pembersihan Dataset Kesalahan	15
3.3.3 Penyusunan Dataset Partikel Benar	15
3.3.4 Pembersihan Dataset Benar	16
3.3.5 Dataset Akhir	16
3.3.6 Flowchart Pembuatan Dataset	17
3.4 Pra-Pemrosesan Data	18
3.5 Rekayasa Fitur	19
3.5.1 Fitur Leksikal	21
3.5.2 Fitur Pola Penulisan	21
3.5.3 Fitur Morfologis dan POS	22
3.5.4 Fitur Relasi Linguistik Partikel	22
3.5.5 Pembentukan Dataset Fitur Akhir	23
3.6 Pelatihan Model (Training)	24
3.7 Integrasi Rule-Based System	26
3.8 Evaluasi Model	26

3.9	Pembuatan API	26
3.10	Spesifikasi Sistem	27
BAB 4	HASIL DAN DISKUSI	28
4.1	Hasil Implementasi Pengumpulan Dataset U-Tapis	28
4.2	Hasil Rekayasa Fitur	29
4.3	Pelatihan Model	30
4.4	Implementasi Model dengan Integrasi API	31
BAB 5	SIMPULAN DAN SARAN	36
5.1	Simpulan	36
5.2	Saran	36
DAFTAR PUSTAKA		37



DAFTAR TABEL

Tabel 3.1	Distribusi Data untuk Sembilan Kategori Kesalahan	15
Tabel 3.2	Distribusi Label pada Dataset Akhir	17
Tabel 3.3	Fitur Leksikal untuk Dua Kalimat	23
Tabel 3.4	Fitur Pola Penulisan untuk Dua Kalimat	23
Tabel 3.5	Fitur Morfologis dan POS untuk Dua Kalimat	24
Tabel 3.6	Fitur Relasi Linguistik untuk Dua Kalimat	24
Tabel 3.7	Spesifikasi sistem perangkat keras dan perangkat lunak penelitian	27



DAFTAR GAMBAR

Gambar 2.1	XGBoost	7
Gambar 2.2	Features of XGBoost Algorithm	7
Gambar 3.1	Pipeline Penelitian	13
Gambar 3.2	Flowchart Pembuatan Dataset	17
Gambar 3.3	Flowchart Pra-Pemrosesan	18
Gambar 3.4	Flowchart Rekayasa Fitur	20
Gambar 3.5	Contoh Proses XGBoost	25
Gambar 3.6	API Request	27
Gambar 4.1	Dataset Kotor Kandidat Partikel	28
Gambar 4.2	Dataset Bersih Partikel Label 1	29
Gambar 4.3	Dataset Fitur Ekstraksi	29
Gambar 4.4	Sebagian Hasil Eksperimen Grid Search pada XGBoost	30
Gambar 4.5	Hasil Pelatihan Model XGBoost Menggunakan Parameter Terbaik	31
Gambar 4.6	Contoh Respons <i>Rule</i> ke-1	32
Gambar 4.7	Contoh Respons <i>Rule</i> ke-2 dan ke-3	32
Gambar 4.8	Contoh Respons <i>Rule</i> ke-4 dan ke-5 (<i>lah</i>)	33
Gambar 4.9	Contoh Respons <i>Rule</i> ke-5 (<i>kah</i>)	34
Gambar 4.10	Contoh Respons <i>Rule</i> ke-6 dan ke-7	34
Gambar 4.11	Contoh Respons <i>Rule</i> ke-8 dan ke-9	35



DAFTAR RUMUS

Rumus 2.1	Fungsi Objektif XGBoost	9
Rumus 2.2	Rumus Gain pada XGBoost	9
Rumus 2.3	<i>Accuracy</i>	11
Rumus 2.4	<i>Precision</i>	11
Rumus 2.5	<i>Recall</i>	12
Rumus 2.6	<i>F1-Score</i>	12



DAFTAR LAMPIRAN

Lampiran 1	PRO-STEP-01 Cover Letter	42
Lampiran 2	PRO-STEP-02 Program Card	43
Lampiran 3	PRO-STEP-03 Daily Task	44
Lampiran 4	PRO-STEP-04 Verification Form	56
Lampiran 5	Form Bimbingan	57
Lampiran 6	Turnitin	58
Lampiran 7	Draft Artikel	65
Lampiran 8	Surat Kerjasama	72
Lampiran 9	Surat Penerimaan PRO-STEP	76

