

**U-TAPIS: DETEKSI HURUF MIRING DAN TEGAK PADA
TEKS BERITA MENGGUNAKAN RULE-BASED DAN
CONDITIONAL RANDOM FIELD**



LAPORAN RESEARCH & TECHNOLOGY PROGRAM

**LUIS GANTAR
00000073190**

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

**U-TAPIS: DETEKSI HURUF MIRING DAN TEGAK PADA
TEKS BERITA MENGGUNAKAN RULE-BASED DAN
CONDITIONAL RANDOM FIELD**



LAPORAN RESEARCH & TECHNOLOGY PROGRAM

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

**LUIS GANTAR
00000073190**

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

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Luis Gantar
Nomor Induk Mahasiswa : 00000073190
Program Studi : Informatika

Skripsi dengan judul:

U-TAPIS: Deteksi Huruf Miring dan Tegak pada Teks Berita Menggunakan Rule-Based dan Conditional Random Field

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, 23 Desember 2025



(Luis Gantar)

UMM
UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PERNYATAAN PENGGUNAAN BANTUAN KECERDASAN ARTIFISIAL (AI)

Saya yang bertanda tangan di bawah ini:

Nama Lengkap : Luis Gantar
NIM : 00000073190
Program Studi : Informatika
Judul Laporan : U-TAPIS: Deteksi Huruf Miring dan Tegak pada
Teks Berita Menggunakan Rule-Based dan Conditional
Random Field

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




(Luis Gantar)

HALAMAN PERNYATAAN KEABSAHAN PERUSAHAAN

Nama Lengkap : Luis Gantar
NIM : 00000073190
Program Studi : Informatika
Fakultas : Teknik dan Informatika

Dengan ini menyatakan bahwa saya melaksanakan kegiatan di:

1. Nama Perusahaan/Organisasi : LPPM Universitas Multimedia Nusantara
2. Alamat : Jalan Scientia Boulevard Gading, Curug
Sangereng, Serpong, Kabupaten Tangerang,
Banten 15810
3. Email Perusahaan/Organisasi : lppm@umn.ac.id

- (1) Perusahaan / Organisasi tempat saya melakukan kegiatan dapat divalidasi keberadaannya.
- (2) Jika di kemudian hari terbukti ditemukan kekurangan / 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.

Mengetahui,



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

Tangerang, 7 Januari 2026

Menyatakan,



(Luis Gantar)

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

Yang bertanda tangan di bawah ini:

Nama : Luis Gantar
NIM : 00000073190
Program Studi : Informatika
Jenjang : S1
Judul Karya Ilmiah : U-TAPIS: Deteksi Huruf Miring dan
Tegak pada Teks Berita Menggunakan
Rule-Based dan Conditional Random
Field

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, 23 Desember 2025

Yang menyatakan


Luis Gantar

**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

"You can improve on everything. You're never perfect"

Max Verstappen



KATA PENGANTAR

Puji dan syukur dipanjatkan kepada Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya sehingga laporan penelitian ini yang berjudul “U-TAPIS: Deteksi Huruf Miring Dan Tegak Pada Teks Berita Menggunakan Rule-Based Dan Conditional Random Field” dapat terselesaikan tepat waktu. Laporan ini disusun sebagai salah satu bentuk pertanggungjawaban akademik dalam program *PRO-STEP Research & Technology Program* yang dilaksanakan di Universitas Multimedia Nusantara.

Mengucapkan terima kasih

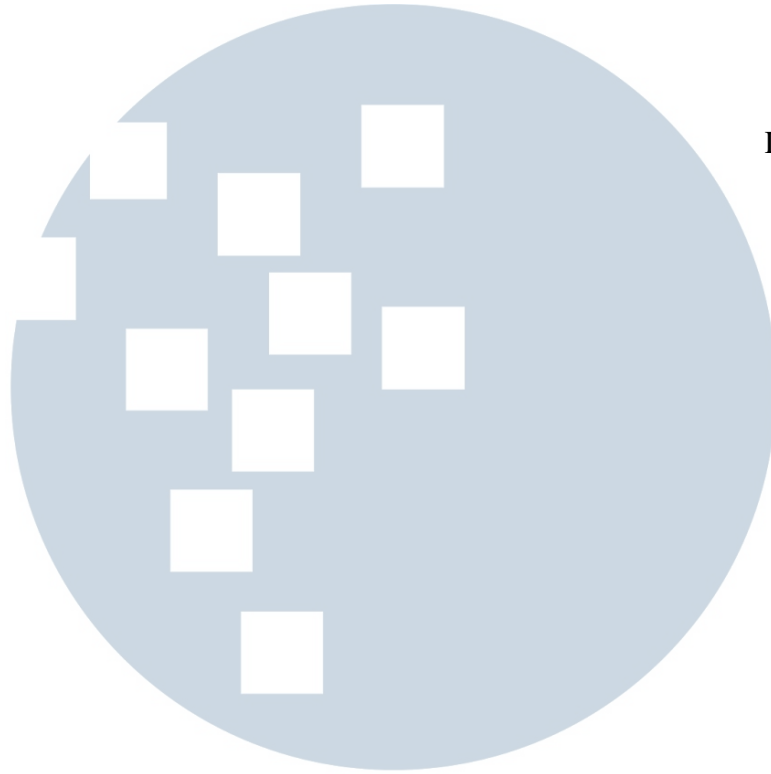
1. Bapak Dr. Ir. Andrey Andoko, M.Sc., selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak Vincentius Kurniawan, S.Kom., M.Eng.Sc., sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya laporan penelitian ini.
5. Ibu Marlinda Vasty Overbeek, S.Kom, M.Kom, sebagai Pembimbing lapangan yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya laporan penelitian ini.
6. Ibu Dr. Niknik M. Kuntarto, M.Si. sebagai pakar bahasa dalam penelitian ini.
7. Bapak Rudi Sutomo, S.Kom., M.Si., M.Kom. sebagai pakar pengembang.
8. Pihak Keluarga yang telah memberikan bantuan dukungan material dan moral, sehingga penelitian ini dapat terselesaikan.

Semoga laporan magang ini dapat memberikan manfaat dan menjadi referensi bagi pihak-pihak yang membutuhkannya.

Tangerang, 23 Desember 2025



Luis Gantar



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

U-TAPIS: DETEKSI HURUF MIRING DAN TEGAK PADA TEKS BERITA MENGUNAKAN RULE-BASED DAN CONDITIONAL RANDOM FIELD

Luis Gantar

ABSTRAK

Di era digital, tuntutan kecepatan publikasi berita sering kali mengakibatkan kesalahan linguistik, khususnya pada penggunaan huruf miring dan tegak yang tidak sesuai dengan Pedoman Umum Ejaan Bahasa Indonesia (PUEBI). Platform U-Tapis yang dikembangkan untuk membantu jurnalis saat ini belum memiliki fitur spesifik tersebut. Penelitian ini bertujuan untuk mengatasi hal tersebut dengan mengembangkan model deteksi otomatis *menggunakan pendekatan hibrida, yaitu penggabungan metode Rule-Based dan Conditional Random Field (CRF)*. Penelitian ini menggunakan dataset berupa 10.000 artikel berita daring serta data sintesis yang dilabeli dengan skema IOB. Proses deteksi melibatkan ekstraksi fitur kontekstual, leksikal, dan kata pemicu (*trigger word*) untuk mengenali 12 kategori huruf miring dan lima kategori huruf tegak. Hasil evaluasi menunjukkan bahwa model memiliki performa yang sangat tinggi dengan akurasi global (*Global Accuracy*) sebesar 99,84% dan rata-rata *F1-Score* sebesar 0,993.

Kata kunci: *Conditional Random Field, Deteksi Huruf Miring, Natural Language Processing, PUEBI, U-Tapis.*



***U-TAPIS: DETECTION OF ITALIC AND UPRIGHT TEXT IN NEWS
ARTICLES USING RULE-BASED AND CONDITIONAL RANDOM FIELD***

Luis Gantar

ABSTRACT

In the digital era, the demand for speed in news publication often leads to linguistic errors, particularly regarding the use of italics and upright text that do not comply with the General Guidelines for Indonesian Spelling (PUEBI). The U-Tapis platform, developed to assist journalists, currently lacks a specific feature to address this issue. This study aims to fill this gap by developing an automatic detection model using a hybrid approach that combines Rule-Based methods and Conditional Random Fields (CRF). This study utilizes a dataset consisting of 10,000 online news articles and synthetic data labeled using the IOB scheme. The detection process involves the extraction of contextual, lexical, and trigger word features to recognize 12 categories of italics and five categories of upright text. Evaluation results demonstrate that the model achieves very high performance, with a Global Accuracy of 99.84% and an average F1-Score of 0.993

Keywords: Conditional Random Field, Italic Detection, Natural Language Processing, PUEBI, U-Tapis



DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
PERNYATAAN PENGGUNAAN BANTUAN AI	iii
PERNYATAAN KEABSAHAN	iv
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	v
HALAMAN PERSEMBAHAN/MOTO	vi
KATA PENGANTAR	vii
ABSTRAK	ix
ABSTRACT	x
DAFTAR ISI	xi
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR KODE	xiv
DAFTAR RUMUS	xv
DAFTAR LAMPIRAN	xvi
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Batasan Masalah	3
1.5 Urgensi Penelitian	4
1.6 Luaran Penelitian	4
1.7 Manfaat Penelitian	4
BAB 2 LANDASAN TEORI	5
2.1 Aturan Penggunaan Huruf Miring	5
2.2 <i>Natural Language Processing</i>	6
2.2.1 Pelabelan Sekuensial (<i>Sequence Labeling</i>) sebagai Tugas Fundamental	7
2.3 IOB (<i>Inside, Outside, Beginning</i>) Labeling	8
2.4 <i>Conditional Random Field</i> (CRF)	8
2.5 <i>Confusion Matrix</i>	10
BAB 3 METODOLOGI PENELITIAN	13
3.1 Metode Penelitian	13
3.2 Tahapan Penelitian	13
3.3 Pengumpulan dan Persiapan Dataset	14
3.4 Pra-pemrosesan Data	16
3.5 Rekayasa Fitur dan Pembagian Data	18
3.6 Pelatihan dan Evaluasi Model	21
BAB 4 HASIL DAN DISKUSI	23
4.1 Implementasi Sistem Deteksi Penggunaan Huruf Miring	23
4.1.1 Persiapan Dataset	23
4.1.2 Implementasi Model CRF	28
4.1.3 Evaluasi Model CRF	29
4.1.4 Implementasi Hasil Model Menjadi API	30
BAB 5 SIMPULAN DAN SARAN	33
5.1 Simpulan	33
5.2 Saran	33
DAFTAR PUSTAKA	34

DAFTAR TABEL

Tabel 2.1	Tabel <i>Confusion Matrix</i>	11
Tabel 3.1	Distribusi data entitas kategori isi	15
Tabel 3.1	Lanjutan 3.1	16
Tabel 3.2	Distribusi Data Entitas Kategori Wadah	16
Tabel 3.3	Daftar fitur pemicu, posisi, dan kontekstual model CRF	18
Tabel 3.4	Lanjutan Tabel 3.3 (fitur pemicu dan posisil)	19
Tabel 3.5	Daftar fitur lokal dan leksikal model CRF	19
Tabel 3.5	Lanjutan Tabel 3.5 (Fitur tanda baca dan leksikal)	20
Tabel 4.1	Contoh data latih korpus berita	23
Tabel 4.2	Contoh data yang memiliki label	26
Tabel 4.3	Contoh data dengan format BIO	27
Tabel 4.4	Spesifikasi parameter request API	31
Tabel 4.5	Spesifikasi Parameter <i>Response</i> Utama (Root)	31
Tabel 4.6	Spesifikasi Atribut Objek Entitas	32



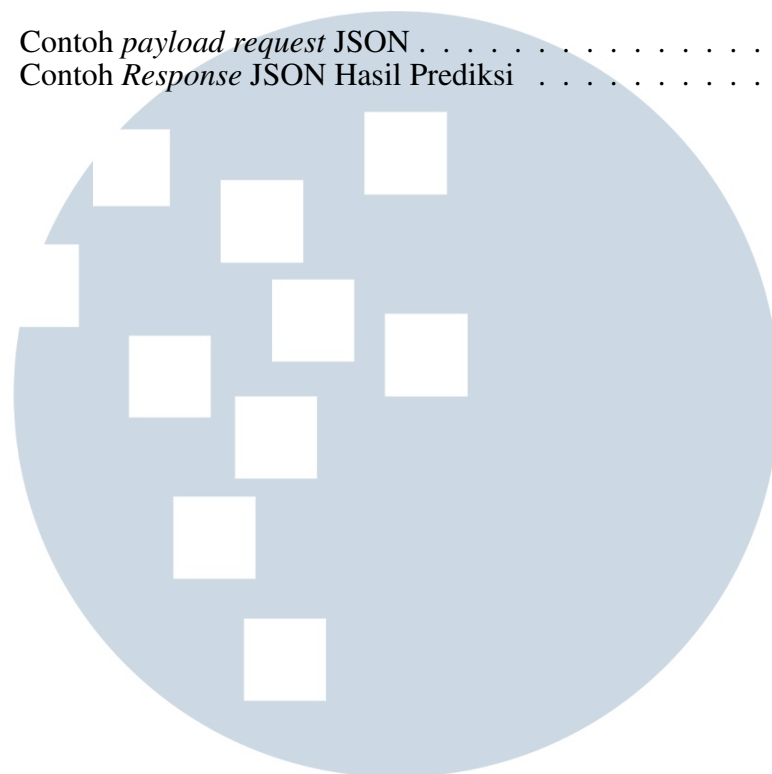
DAFTAR GAMBAR

Gambar 3.1	<i>Pipeline</i> penelitian	14
Gambar 3.2	<i>Flowchart</i> persiapan dan ekstraksi dataset	15
Gambar 3.3	Flowchart tahapan pra-pemrosesan dan pelabelan token	17
Gambar 3.4	<i>Flowchart</i> rekayasa fitur	20
Gambar 3.5	<i>Flowchart</i> algoritma pelatihan model CRF	22
Gambar 4.1	<i>Log</i> proses pembagian data dan pelatihan model	28
Gambar 4.2	Contoh hasil ekstraksi fitur pada satu token	29
Gambar 4.3	Laporan hasil evaluasi kinerja model per kategori	30



DAFTAR KODE

Kode 4.1	Contoh <i>payload request</i> JSON	31
Kode 4.2	Contoh <i>Response</i> JSON Hasil Prediksi	32



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR RUMUS

2.1	Rumus Model <i>Conditional Random Fields</i> (CRF)	9
2.2	Rumus <i>Log-Likelihood</i>	9
2.3	Rumus <i>Accuracy Confusion Matrix</i>	11
2.4	Rumus <i>Precision Confusion Matrix</i>	11
2.5	Rumus <i>Recall Confusion Matrix</i>	11
2.6	Rumus <i>F1-Score Confusion Matrix</i>	11



DAFTAR LAMPIRAN

Lampiran 1	PRO-STEP-01 Cover Letter	38
Lampiran 2	PRO-STEP-02 Research and Technology Program Card	39
Lampiran 3	PRO-STEP-03 Daily Task	40
Lampiran 4	PRO-STEP-04 Verification Form	55
Lampiran 5	Formulir Bimbingan	55
Lampiran 6	Hasil Persentase Turnitin	57
Lampiran 7	Draft Artikel Nasional	65
Lampiran 8	Formulir Penggunaan Perangkat Kecerdasan Artifisial (AI)	69
Lampiran 9	Surat Kesepahaman UMN dan TRIBUN DIGITAL ONLINE	71
Lampiran 10	Surat Penerimaan PRO-STEP	76

