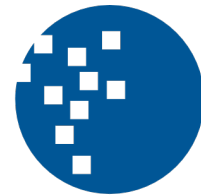


**PENGEMBANGAN MODUL U-TAPIS UNTUK
DETEKSI KESALAHAN PENGGUNAAN ANGKA DAN
BILANGAN MENGGUNAKAN CONDITIONAL
RANDOM FIELDS**



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

LAPORAN RESEARCH TECHNOLOGY PROGRAM

VINSON
00000070870

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

**PENGEMBANGAN MODUL U-TAPIS UNTUK DETEKSI
KESALAHAN PENGGUNAAN ANGKA DAN BILANGAN
MENGUNAKAN CONDITIONAL RANDOM FIELDS**



LAPORAN RESEARCH TECHNOLOGY PROGRAM

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

**VINSON
00000070870**

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

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Vinson
Nomor Induk Mahasiswa : 00000070870
Program Studi : Informatika

Menyatakan dengan sesungguhnya bahwa Laporan Research Technology Program saya yang berjudul:

Pengembangan Modul U-Tapis untuk Deteksi Kesalahan Penggunaan Angka dan Bilangan Menggunakan Conditional Random Fields

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, 30 Desember 2025



(Vinson)

HALAMAN PERNYATAAN PENGGUNAAN BANTUAN KECERDASAN ARTIFISIAL (AI)

Saya yang bertanda tangan di bawah ini:

Nama : Vinson
NIM : 00000070870
Program Studi : Informatika
Judul Laporan : Pengembangan Modul U-Tapis untuk Deteksi Kesalahan
Penggunaan Angka dan Bilangan Menggunakan
Conditional Random Fields

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

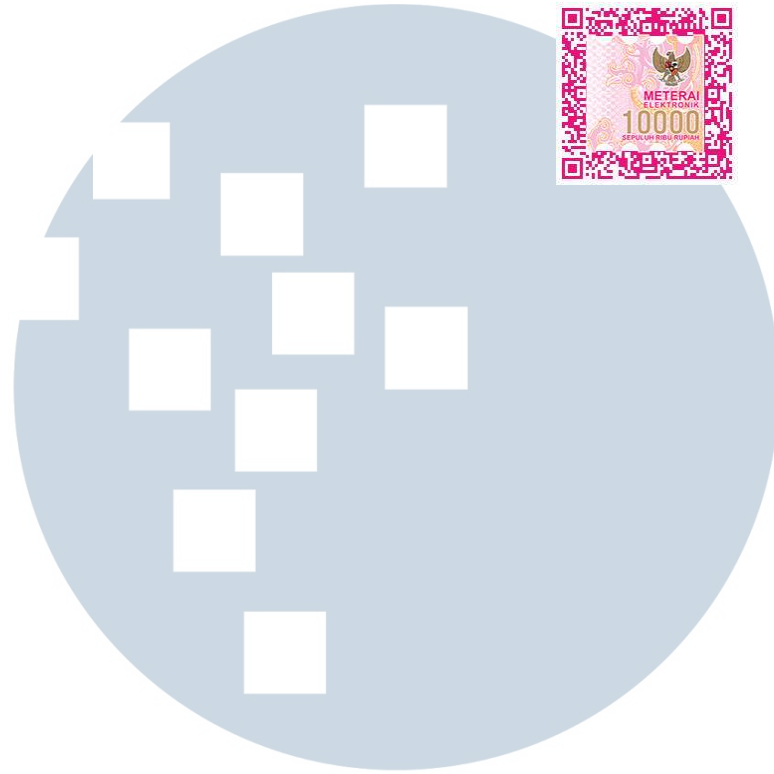
Saya juga menyatakan bahwa:

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, paraphrase, penyertaan ide atau fakta penting yang disarankan oleh AI dan saya telah menyantumkan dalam sitasi serta referensi
3. Terlepas dari pernyataan di atas, tugas ini sepenuhnya merupakan karya saya sendiri

Tangerang, 30 Desember 2025



(Vinson)



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

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

Yang bertanda tangan di bawah ini:

Nama : Vinson
NIM : 00000070870
Program Studi : Informatika
Jenjang : S1
Judul Karya Ilmiah : Pengembangan Modul U-Tapis untuk
Deteksi Kesalahan Penggunaan
Angka dan Bilangan Menggunakan
Conditional Random Fields

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*) **.

Tangerang, 30 Desember 2025

Yang menyatakan

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



Vinson

**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

”Selesaikan permainan ini dengan heroik supaya kita bisa memaknai kehidupan ini dengan lebih baik lagi.”

RD Bonifasius Lumintang



KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas berkat dan rahmat-Nya sehingga penulis dapat menyelesaikan laporan penelitian yang berjudul “Pengembangan Modul U-Tapis untuk Deteksi Kesalahan Penggunaan Angka dan Bilangan Menggunakan *Conditional Random Fields*” Laporan ini disusun sebagai salah satu syarat untuk memenuhi Pro-Step *Research & Technology Program* di Program Studi Informatika, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara. Oleh karena itu, saya mengucapkan terima kasih kepada:

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 Yustinus Widya Wiratama, S.Kom., M.Sc., sebagai Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Ibu Marlinda Vasty Overbeek, S.Kom, M.Kom, sebagai Pembimbing Lapangan yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
6. Ibu Dr. Niknik M. Kuntarto dan Samiaji Bintang Nusantara S.T., M.A. sebagai pakar bahasa dalam penelitian ini.
7. Bapak Rudi Sutomo, S.Kom., M.Si. sebagai pakar pengembang *website* untuk U-Tapis dan Bapak Moeljono Widjaja, B.Sc., M.Sc., PH.D. sebagai pakar *data analysis*.
8. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan penelitian ini.

Semoga karya ilmiah ini berguna dan menjadi manfaat bagi para pembaca.

Tangerang, 30 Desember 2025



Vinson



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

**PENGEMBANGAN MODUL U-TAPIS UNTUK DETEKSI KESALAHAN
PENGUNAAN ANGKA DAN BILANGAN MENGGUNAKAN
CONDITIONAL RANDOM FIELDS**

Vinson

ABSTRAK

Penelitian ini mengembangkan model machine learning berbasis Conditional Random Fields (CRF) untuk mendeteksi kesalahan penggunaan angka dan bilangan yang tidak sesuai dengan Ejaan Bahasa Indonesia yang Disempurnakan (EYD) pada teks berita daring. Proses penelitian mencakup pengumpulan 738.548 artikel berita Indonesia, ekstraksi kalimat, identifikasi kesalahan menggunakan pola berbasis aturan (regex), pelabelan data menggunakan format IOB, serta rekayasa fitur token-level untuk menangkap konteks sekuensial. Data kemudian dibagi menjadi data latih dan uji menggunakan skema 80:20. Model CRF dilatih untuk mengenali sebelas kategori kesalahan numerik, termasuk bilangan awal kalimat, angka lawan huruf, bilangan besar, alamat, penomoran buku, bilangan tingkat, akhiran -an, serta nama geografi. Evaluasi menunjukkan performa tinggi dengan nilai rata-rata F1-score, precision, dan recall sebesar 0,94. Analisis confusion matrix memperlihatkan kemampuan model membedakan kategori secara konsisten dengan tingkat kekeliruan yang rendah. Model kemudian diintegrasikan ke dalam backend berbasis Flask API sebagai layanan inferensi untuk sistem U-Tapis. Hasil ini menegaskan bahwa CRF efektif untuk tugas pelabelan sekuensial yang berkaitan dengan kesalahan angka dalam bahasa Indonesia dan berpotensi meningkatkan akurasi penyuntingan berita bagi jurnalis.

Kata kunci: *API Flask, Conditional Random Fields, Deteksi kesalahan angka dan bilangan, Ejaan Bahasa Indonesia yang Disempurnakan, Inside Outside Beginning, Natural Language Processing*

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 THE U-TAPIS MODULE FOR DETECTING ERRORS
IN THE USE OF NUMBERS AND NUMERALS USING CONDITIONAL
RANDOM FIELDS**

Vinson

ABSTRACT

This study develops a machine learning model based on Conditional Random Fields (CRF) to detect numerical expression errors in Indonesian online news articles according to the Enhanced Spelling System (EYD). The research workflow involves collecting 738,548 Indonesian news articles, extracting sentences, identifying numerical errors using rule-based regular expressions, applying IOB labeling, and performing token-level feature engineering to capture sequential context. The dataset was split into training and testing sets using an 80:20 ratio. The CRF model was trained to recognize eleven categories of numerical errors, including sentence-initial numbers, number-to-word inconsistencies, large-number formatting, address numbers, book section numbering, ordinal expressions, the “-an” suffix, and geographic names. Evaluation results show strong performance, yielding average F1-score, precision, and recall of 0.94. The confusion matrix further demonstrates consistent category separation with minimal misclassification. The trained model was integrated into a Flask-based API to support inference services in the U-Tapis system. These findings confirm that CRF is effective for sequence labeling tasks involving numerical expression validation in Indonesian text and can enhance editorial accuracy for journalists.

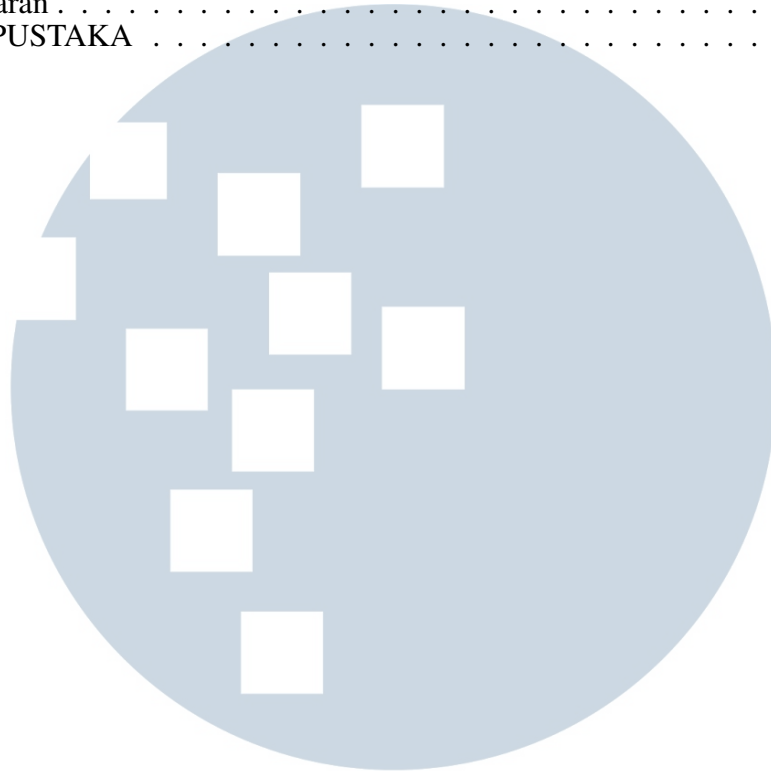
Keywords: *Conditional Random Fields, Enhanced spelling system, Flask API, Inside Outside Beginning, Natural Language Processing, Numerical error detection*

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 PERNYATAAN PENGGUNAAN BANTUAN KECERDASAN ARTIFISIAL (AI)	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	v
HALAMAN PERSEMBAHAN/MOTO	vi
KATA PENGANTAR	vii
ABSTRAK	ix
ABSTRACT	x
DAFTAR ISI	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR	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	4
1.4 Urgensi Penelitian	4
1.5 Luaran Penelitian	4
1.6 Manfaat Penelitian	4
BAB 2 LANDASAN TEORI	6
2.1 Aturan Penggunaan Angka dan Bilangan	6
2.2 <i>Natural Language Processing</i>	6
2.3 <i>IOB (Inside, Outside, Beginning) Labelling</i>	7
2.4 <i>Conditional Random Fields</i>	8
2.5 <i>Confusion Matrix</i>	9
BAB 3 METODOLOGI PENELITIAN	10
3.1 Metode Penelitian	10
3.2 Tahapan Penelitian	10
3.3 Pengumpulan dan Persiapan Data	11
3.4 Pelabelan IOB	14
3.5 Rekayasa Fitur & Pembagian data	15
3.6 Pelatihan Model	19
3.7 Evaluasi Model	21
3.8 Inferensi Model	22
3.9 Implementasi <i>API</i>	23
BAB 4 HASIL DAN DISKUSI	24
4.1 Hasil Implementasi <i>Backend</i>	24
4.1.1 <i>Request API</i>	24
4.1.2 <i>Response API</i>	24
4.2 Implementasi Sistem Deteksi Kesalahan Penggunaan Angka dan Bilangan	25
4.2.1 Pengumpulan Dataset	25
4.2.2 Implementasi Algoritma <i>Conditional Random Fields</i>	27
4.2.3 Evaluasi Model	28
4.2.4 Uji Coba Deteksi Kalimat dengan Model	30
4.2.5 Implementasi Model Menggunakan <i>Flask</i>	37

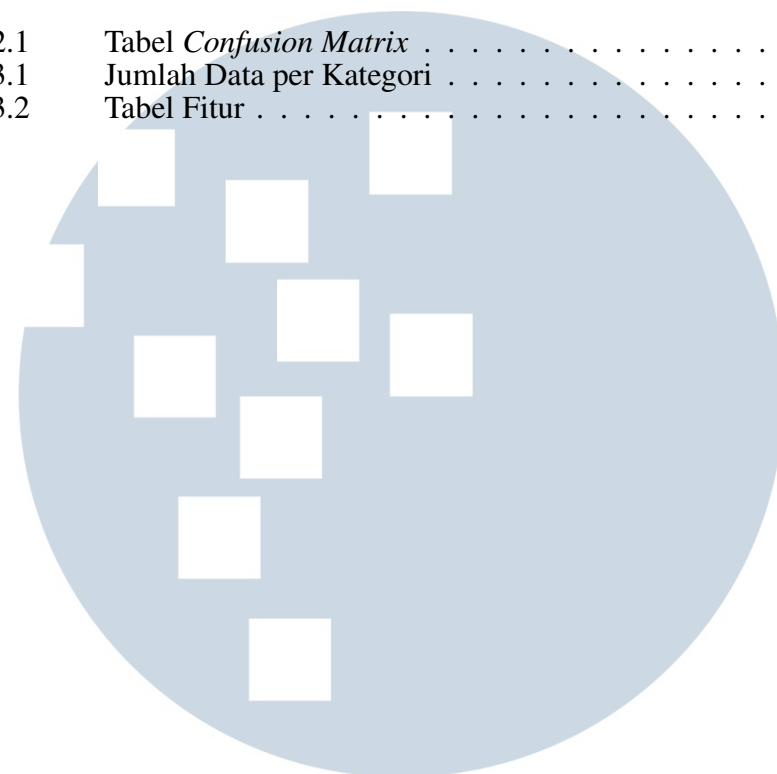
BAB 5	SIMPULAN DAN SARAN	38
5.1	Simpulan	38
5.2	Saran	38
DAFTAR PUSTAKA		40



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR TABEL

Tabel 2.1	Tabel <i>Confusion Matrix</i>	9
Tabel 3.1	Jumlah Data per Kategori	14
Tabel 3.2	Tabel Fitur	16



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR GAMBAR

Gambar 3.1	<i>Pipeline Penelitian</i>	11
Gambar 3.2	<i>Flowchart Ekstraksi Kalimat</i>	12
Gambar 3.3	<i>Flowchart IOB Labelling</i>	15
Gambar 3.4	<i>Flowchart Rekayasa Fitur</i>	16
Gambar 3.5	<i>Flowchart pelatihan model CRF</i>	20
Gambar 3.6	<i>Flowchart inferensi model CRF</i>	22
Gambar 4.1	Contoh <i>Request API</i>	24
Gambar 4.2	Contoh <i>Response API</i>	25
Gambar 4.3	Contoh Data Artikel yang Telah Dipecah per Kalimat	26
Gambar 4.4	Contoh Data yang Telah Difilter	26
Gambar 4.5	Contoh Hasil Feature Engineering	27
Gambar 4.6	<i>F1-score, Precision, dan Recall</i> per Kategori	28
Gambar 4.7	<i>Confussion Matrix</i>	29
Gambar 4.8	Hasil Deteksi Kalimat 1	30
Gambar 4.9	Hasil Deteksi Kalimat 2	31
Gambar 4.10	Hasil Deteksi Kalimat 3	32
Gambar 4.11	Hasil Deteksi Kalimat 4	32
Gambar 4.12	Hasil Deteksi Kalimat 5	33
Gambar 4.13	Hasil Deteksi Kalimat 6	34
Gambar 4.14	Hasil Deteksi Kalimat 7	34
Gambar 4.15	Hasil Deteksi Kalimat 8	35
Gambar 4.16	Hasil Deteksi Kalimat 9	36
Gambar 4.17	Hasil Deteksi Kalimat 10	36
Gambar 4.18	Hasil Deteksi Kalimat 11	37



DAFTAR RUMUS

Rumus 2.1	CRF	8
Rumus 2.2	<i>Accuracy</i>	9
Rumus 2.3	<i>Precision</i>	9
Rumus 2.4	<i>Recall</i>	9
Rumus 2.5	<i>F1 Score</i>	9



DAFTAR LAMPIRAN

Lampiran 1	PRO-STEP-01 Cover Letter	43
Lampiran 2	PRO-STEP-02 PRO-STEP Card	44
Lampiran 3	PRO-STEP-03 Daily Task	45
Lampiran 4	PRO-STEP-04 Verification Form	56
Lampiran 5	Form Bimbingan	57
Lampiran 6	Turnitin	58
Lampiran 7	Formulir Penggunaan Perangkat Kecerdasan Artifisial (AI) . .	63
Lampiran 8	Draft Artikel Nasional	65
Lampiran 9	Bukti Kerja Sama (MoU) Dengan TribunNews	70

