

**PENGEMBANGAN MODUL U-TAPIS BERBASIS ALGORITMA  
RANDOM FOREST UNTUK DETEKSI PENGGUNAAN HURUF KAPITAL**



**LAPORAN MBKM PENELITIAN**

**Yustira Nhisya Shabilla  
00000063592**

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

**PENGEMBANGAN MODUL U-TAPIS BERBASIS ALGORITMA  
RANDOM FOREST UNTUK DETEKSI PENGGUNAAN HURUF KAPITAL**



**LAPORAN MBKM PENELITIAN**

**Yustira Nhisya Shabilla**

**00000063592**

**UMN**

**UNIVERSITAS**

**MULTIMEDIA**

**NUSANTARA**

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

**TANGERANG**

**2025**

## HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Yustira Nhisya Shabilla  
NIM : 00000063592  
Program Studi : Informatika

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

**PENGEMBANGAN MODUL U-TAPIS BERBASIS ALGORITMA  
RANDOM FOREST UNTUK DETEKSI PENGGUNAAN HURUF KAPITAL**

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, Tgl. Pengumpulan



(Yustira Nhisya Shabilla)

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

Yang bertanda tangan di bawah ini:

Nama : Yustira Nhisya Shabilla  
NIM : 00000063592  
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

UMMN  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA



Yustira Nhisya Shabilla

\*\* 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**

”Jauhi orang-orang negatif, mereka punya masalah untuk setiap solusi.”

- Albert Einstein



## 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 Berbasis Algoritma Random Forest untuk Deteksi Penggunaan Huruf Kapital”\* dapat diselesaikan dengan baik. Laporan MBKM Penelitian ini bertujuan untuk mengembangkan modul yang mampu mendeteksi kesalahan penggunaan huruf kapital dalam teks menggunakan pendekatan berbasis algoritma *Random Forest*.

Dengan ini, 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 Assoc. Prof. Arya Wicaksana, S.Kom., M.Eng.Sc., OCA., selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak Dr. Ir. Winarno, M.Kom., selaku Pembimbing MBKM pertama yang telah memberikan bimbingan dan arahan.
5. Ibu Marlinda Vasty Overbeek, S.Kom., M.Kom., selaku pembimbing penelitian yang memberikan banyak bimbingan dan dukungan.
6. Orang tua, keluarga, dan teman-teman saya (Areta, Dea, Arvin, Kelsha, Kesya, Bella, Nidaa, Andara, Hava) yang telah memberikan dukungan moral dan material.

Semoga laporan penelitian ini bermanfaat bagi kita semua.

Tangerang, 3 Januari 2025

Yustira Nhisya Shabilla

# **PENGEMBANGAN MODUL U-TAPIS BERBASIS ALGORITMA RANDOM FOREST UNTUK DETEKSI PENGGUNAAN HURUF KAPITAL**

Yustira Nhisya Shabilla

## **ABSTRAK**

Pengolahan teks otomatis menjadi kebutuhan penting dalam media dan jurnalistik untuk mendeteksi kesalahan kapitalisasi huruf yang memengaruhi profesionalisme tulisan dan kredibilitas berita. Pedoman Ejaan yang Disempurnakan (EYD) menyediakan aturan jelas, namun penerapannya secara konsisten masih menjadi tantangan, terutama pada teks yang diproduksi dalam jumlah besar. Teknologi Pemrosesan Bahasa Alami (NLP) menawarkan solusi efisien, namun perangkat lunak serupa untuk bahasa Indonesia masih terbatas. U-TAPIS dikembangkan berbasis aturan EYD, memanfaatkan algoritma Random Forest yang efektif untuk klasifikasi teks, menangani data berdimensi tinggi, dan robust terhadap overfitting. Modul ini dirancang untuk meningkatkan akurasi, konsistensi, dan kualitas teks jurnalistik, sekaligus mendukung kompetensi bahasa jurnalis. Dengan fokus pada efisiensi dan spesifisitas kebutuhan jurnalistik, U-TAPIS menjadi alat yang berharga dalam pengembangan teknologi bahasa di Indonesia.

**Kata Kunci:** Ejaan Yang Disempurnakan (EYD), kesalahan kapitalisasi, Pemrosesan Bahasa Alami (NLP), Random Forest, U-TAPIS.





**DEVELOPMENT OF THE U-TAPIS MODULE BASED ON THE RANDOM FOREST ALGORITHM FOR DETECTING CAPITALIZATION USAGE**

Yustira Nhisya Shabilla

**ABSTRACT**

*Automated text processing has become an essential need in media and journalism to detect capitalization errors that affect the professionalism of writing and the credibility of news. The Enhanced Spelling Guidelines (EYD) provide clear rules, but consistent application remains a challenge, especially for texts produced in large volumes. Natural Language Processing (NLP) technology offers efficient solutions, but similar software for the Indonesian language remains limited. U-TAPIS is developed based on EYD rules, leveraging the Random Forest algorithm, which is effective for text classification, handles high-dimensional data, and is robust against overfitting. This module is designed to improve the accuracy, consistency, and quality of journalistic texts while supporting journalists' language competence. With a focus on efficiency and the specific needs of journalism, U-TAPIS becomes a valuable tool in advancing language technology in Indonesia.*

**Keywords:** *capitalization errors, Enhanced Spelling Guidelines (EYD), Natural Language Processing (NLP), Random Forest, U-TAPIS.*



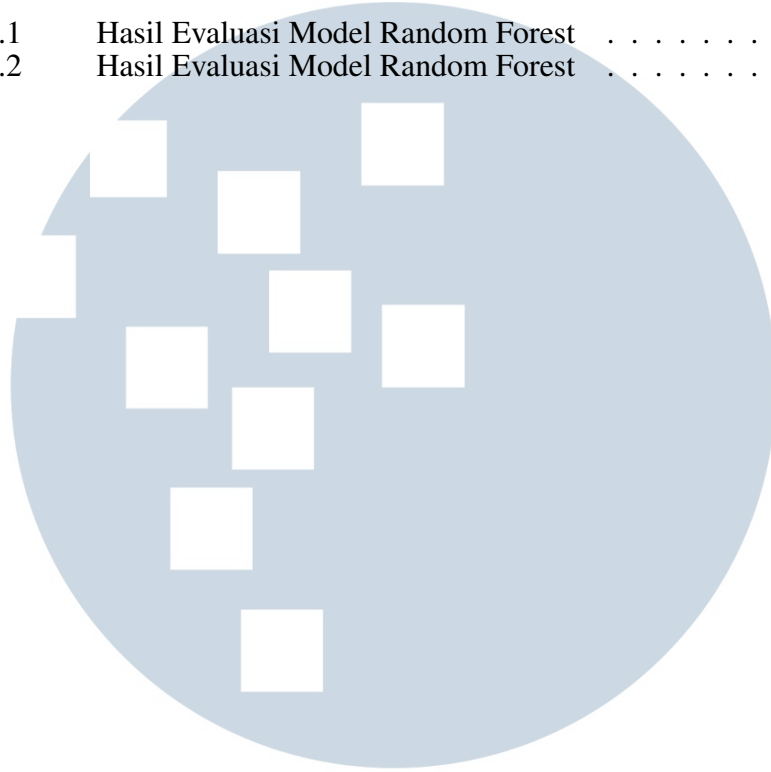


## DAFTAR ISI

HALAMAN JUDUL . . . . .	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT . . . . .	ii
HALAMAN PERSETUJUAN PUBLIKASI ILMIAH . . . . .	iii
HALAMAN PERSEMBAHAN/MOTO . . . . .	iv
KATA PENGANTAR . . . . .	v
ABSTRAK . . . . .	vi
ABSTRACT . . . . .	vii
DAFTAR ISI . . . . .	viii
DAFTAR TABEL . . . . .	ix
DAFTAR GAMBAR . . . . .	x
DAFTAR KODE . . . . .	xi
DAFTAR LAMPIRAN . . . . .	xii
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 . . . . .	4
BAB 2 LANDASAN TEORI . . . . .	5
2.1 Ejaan Yang Disempurnakan (EYD) . . . . .	5
2.2 Decision Tree . . . . .	7
2.3 Random Forests . . . . .	9
2.4 Ensemble Learning . . . . .	10
2.5 Confusion Matrix . . . . .	11
BAB 3 METODOLOGI PENELITIAN . . . . .	13
3.1 Tinjauan Pustaka . . . . .	13
3.2 Pengumpulan Data . . . . .	13
3.3 Pre-processing . . . . .	14
3.4 Ekstraksi Fitur . . . . .	15
3.5 Pembagian Dataset . . . . .	15
3.6 Pembangunan Model . . . . .	17
3.7 Evaluasi Model . . . . .	18
BAB 4 HASIL DAN DISKUSI . . . . .	19
4.1 Implementasi Sistem . . . . .	19
4.2 Import Dataset . . . . .	19
4.3 Preprocessing Data . . . . .	20
4.4 Ekstrasi Fitur . . . . .	22
4.5 Split Dataset . . . . .	24
4.6 Pelatihan Model . . . . .	24
4.7 Hasil dan Evaluasi Model . . . . .	25
BAB 5 SIMPULAN DAN SARAN . . . . .	30
5.1 Simpulan . . . . .	30
5.2 Saran . . . . .	30
DAFTAR PUSTAKA . . . . .	32

## DAFTAR TABEL

Tabel 4.1	Hasil Evaluasi Model Random Forest . . . . .	25
Tabel 4.2	Hasil Evaluasi Model Random Forest . . . . .	27



**UMMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## DAFTAR GAMBAR

Gambar 2.1	Arsitektur Decision Tree [1] . . . . .	7
Gambar 3.1	Preprocessing Flowchart . . . . .	14
Gambar 3.2	Ekstrasi Fitur Flowchart . . . . .	16
Gambar 3.3	Pembangunan Model Flowchart . . . . .	17
Gambar 4.1	Confussion Matrix . . . . .	26



## DAFTAR KODE

Kode 4.1	Import Dataset . . . . .	19
Kode 4.2	Data Preprocessing and Transformation . . . . .	20
Kode 4.3	Fungsi untuk Analisis Kapitalisasi . . . . .	22
Kode 4.4	Membagi Dataset untuk Pelatihan dan Pengujian . . . . .	24
Kode 4.5	Pelatihan Model Random Forest . . . . .	24



**UMMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## DAFTAR LAMPIRAN

Lampiran 1	MBKM-01 Cover Letter MBKM Research Track 1 . . . . .	38
Lampiran 2	MBKM-02 MBKM Research Track 1 Card . . . . .	39
Lampiran 3	MBKM-03 Daily Task - Research Track 1 . . . . .	40
Lampiran 4	MBKM-04 Verification Form of Research Report MBKM Research Track 1 . . . . .	53
Lampiran 5	Form Bimbingan . . . . .	54
Lampiran 6	Turnitin . . . . .	55
Lampiran 7	MoU Berita Tribun . . . . .	57
Lampiran 8	Draft Artikel Nasional . . . . .	60

