XLM-ROBERTA-BASED DETECTION OF HATE SPEECH IN INDONESIAN-ENGLISH CODE-MIXED TEXT

PUBLICATION THESIS

FARREL DINARTA
00000055702

INFORMATICS STUDY PROGRAM
FACULTY OF ENGINEERING AND INFORMATICS
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025

XLM-ROBERTA-BASED DETECTION OF HATE SPEECH IN INDONESIAN-ENGLISH CODE-MIXED TEXT

PUBLICATION THESIS

Submitted as one of the requirements to obtain the degree of Bachelor of Computer Science (S.Kom.)

FARREL DINARTA
00000055702

INFORMATICS STUDY PROGRAM
FACULTY OF ENGINEERING AND INFORMATICS
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025

DECLARATION OF NON-PLAGIARISM

I, the undersigned,

Name	:	Farrel Dinarta
Student ID Number	:	00000055702
Study Program	:	Informatics

Hereby declare that the thesis entitled:
XLM-RoBERTa-Based Detection of Hate Speech in Indonesian-English Code-Mixed Texts

is entirely my own work and not a result of plagiarism from any other scientific writing. All sources, whether quoted or referenced, have been properly acknowledged and listed in the References section.

Should any form of academic misconduct or irregularity in the execution or writing of this scientific report be discovered in the future, I am willing to accept the consequences, including being declared FAIL in the respective course.

Tangerang, 20 May 2025

(Farrel Dinarta)

ENDORSEMENT PAGE

The publication titled:

XLM-ROBERTA-BASED DETECTION OF HATE SPEECH IN INDONESIAN-ENGLISH CODE-MIXED TEXT

By

Full Name	:	Farrel Dinarta
Student ID	:	00000055702
Study Program	:	Informatics
Faculty	:	Engineering and Informatics

Has been tested on Wednesday, July 9, 2025
from 13.00 s/s 14.00 and was stated
PASSED

Chair Person of the Session

(Alethea Suryadibrata, S.Kom., M.Eng.)
NIDN: 0322099201

Supervisor / Mentor

(Arya Wicaksana, S.Kom., M.Eng.Sc., OCA)
NIDN: 0315109103

Head of Informatics

(Arya Wicaksana, S.Kom., M.Eng.Sc., OCA)
NIDN: 0315109103

SCIENTIFIC WORK PUBLICATION APPROVAL PAGE FOR ACADEMIC PURPOSES

The undersigned below:

Name	:	Farrel Dinarta
Student ID	:	00000055702
Study Program	:	Informatics
Degree Level	:	Bachelor
Title of Scientific Work	:	XLM-RoBERTa-Based Detection of Hate Speech in Indonesian-English Code-Mixed Texts

Hereby sincerely declare my willingness (**please select one option**):

☐ I fully grant permission to Universitas Multimedia Nusantara to publish my scientific work in the Knowledge Center repository, so that it may be accessed by the UMN Academic Community/Public. I declare that the scientific work I created does not contain any confidential data.

☑ I do not grant permission to publish this scientific work in the Knowledge Center repository, because it is currently under submission for publication in a national/international journal or conference (evidenced by a *letter of acceptance*)**.

☐ Other, please select one:

- May only be accessed internally by Universitas Multimedia Nusantara
- Publication of the scientific work is under embargo for a period of three years

Tangerang, 20 May 2025

Declarant,

Farrel Dinarta

**If I am unable to provide proof of the journal/patent letter of acceptance, I hereby agree to fully authorize the publication of my scientific work to the UMN Knowledge Center and acknowledge it as the property of Universitas Multimedia Nusantara.

PREFACE

All praise and gratitude to God Almighty for His blessings and grace, through which this Publication report entitled: XLM-RoBERTa-Based Detection of Hate Speech in Indonesian-English Code-Mixed Texts has been successfully completed. This report was prepared as one of the requirements to obtain a Bachelor of Computer Science degree in the Informatics Study Program at the Engineering and Informatics Faculty of Universitas Multimedia Nusantara.

Without the help and guidance from various parties throughout my studies and during the preparation of this Publication it would have been difficult for me to complete this report. Therefore, I would like to express my sincere gratitude to:

1. Mr. Dr. Ir. Andrey Andoko, M.Sc., Rector of Universitas Multimedia Nusantara.
2. Mr. Dr. Eng. Niki Prastomo, S.T., M.Sc., Dean of the Faculty of Engineering and Informatics, Universitas Multimedia Nusantara.
3. Mr. Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, Head of the Informatics Study Program at Universitas Multimedia Nusantara, as well as my Supervisor, for the valuable guidance, direction, and motivation throughout the completion of this final project.
4. Mr. Dian Susanto and Mrs. Lisda Nirmala, my parents, for their material and moral support during the preparation of this Publication.
5. My fellow students and friends throughout my academic journey, who have provided continuous support, motivation, and encouragement during my studies and the writing of this final project.

I sincerely hope that this publication report will be useful both as a source of information and inspiration for future readers.

Tangerang, 20 May 2025

Farrel Dinarta

XLM-ROBERTA-BASED DETECTION OF HATE SPEECH IN INDONESIAN-ENGLISH CODE-MIXED TEXT

Farrel Dinarta

ABSTRACT

The prevalence of hate speech on digital platforms presents significant challenges, particularly in multilingual communities where code-switching complicates detection. This study explores the use of XLM-RoBERTa, a transformer-based model with robust multilingual capabilities, to detect hate speech within mixed-language texts, focusing on Indonesian-English code-switching. Traditional hate speech detection models rely on single-language datasets, limiting their effectiveness in such environments. We employ a dataset consisting of Indonesian, English, and code-mixed Indonesian-English language to evaluate XLM-RoBERTa's performance, comprised 24.844 training samples, 2.760 test samples, and 100 supplementary samples additionally. Key hyperparameters included batch size of 16 and 32, with learning rate spanning from 1e-5 to 5e-5. The model achieved near-perfect accuracy (99.6%) on the primary test set and strong generalization across realistic supplementary data with an F1-score of 90.94%. These findings underscore the model's potential for application in complex linguistic contexts, contributing to the development of effective multilingual hate speech detection.

Keywords: *code-switching, hate speech detection, mixed-language texts, XLM-RoBERTa*

DETEKSI UJARAN KEBENCIAN DALAM TEKS CAMPURAN BAHASA INDONESIA-INGGRIS MENGGUNAKAN XLM-ROBERTA

Farrel Dinarta

ABSTRAK

Prevalensi ujaran kebencian di platform digital menimbulkan tantangan besar, terutama di komunitas multibahasa di mana praktik alih kode (*code-switching*) mempersulit proses deteksi. Studi ini mengeksplorasi penggunaan XLM-RoBERTa, sebuah model berbasis *transformer* dengan kemampuan multibahasa yang kuat, untuk mendeteksi ujaran kebencian dalam teks campuran bahasa, dengan fokus pada konteks alih kode bahasa Indonesia-Inggris. Model deteksi ujaran kebencian tradisional umumnya bergantung pada *dataset* satu bahasa, sehingga efektivitasnya terbatas dalam konteks multibahasa. Kami menggunakan *dataset* yang terdiri dari bahasa Indonesia, bahasa Inggris, dan campuran Indonesia-Inggris untuk mengevaluasi performa XLM-RoBERTa, yang mencakup 24.844 sampel pelatihan, 2.760 sampel pengujian, dan 100 sampel tambahan. Beberapa hiperparameter utama meliputi ukuran *batch* 16 dan 32, serta laju pembelajaran antara 1e-5 hingga 5e-5. Model ini mencapai akurasi hampir sempurna (99,6%) pada *set* pengujian utama dan menunjukkan kemampuan generalisasi yang kuat pada data tambahan realistis dengan skor F1 sebesar 90,94%. Temuan ini menegaskan potensi model dalam konteks linguistik yang kompleks, serta kontribusinya terhadap pengembangan deteksi ujaran kebencian multibahasa yang efektif.

Kata Kunci: alih kode, deteksi ujaran kebencian, teks campuran bahasa, XLM-RoBERTa

LIST OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

List of Appendices