IMPLEMENTASI BLOOM FILTER UNTUK PENDETEKSI HATE SPEECH BILINGUAL BAHASA INDONESIA DAN INGGRIS

SKRIPSI

George Marcellino Jo
00000045841

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2024

IMPLEMENTASI BLOOM FILTER UNTUK PENDETEKSI HATE SPEECH BILINGUAL BAHASA INDONESIA DAN INGGRIS

SKRIPSI

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

George Marcellino Jo
00000045841

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2024

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama	:	George Marcellino Jo
NIM	:	00000045841
Program Studi	:	Informatika

Menyatakan dengan sesungguhnya bahwa Skripsi saya yang berjudul:

IMPLEMENTASI BLOOM FILTER UNTUK PENDETEKSI HATE SPEECH BILINGUAL BAHASA INDONESIA DAN INGGRIS

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, 18 Juni 2024

(George Marcellino Jo)

HALAMAN PENGESAHAN

Skripsi dengan judul

IMPLEMENTASI BLOOM FILTER UNTUK PENDETEKSI HATE SPEECH BILINGUAL BAHASA INDONESIA DAN INGGRIS

oleh

Nama	:	George Marcellino Jo
NIM	:	00000045841
Program Studi	:	Informatika
Fakultas	:	Fakultas Teknik dan Informatika

Telah diujikan pada hari Selasa, 4 Juni 2024
Pukul 13.00 s/s 15.00 dan dinyatakan
LULUS
Dengan susunan penguji sebagai berikut

Ketua Sidang

(Eunike Endariahna Surbakti, S.Kom., M.T.I.)
NIDN: 0322099401

Penguji

(Alexander Waworuntu, S.Kom., M.T.I.)
NIDN: 0309068503

Pembimbing

14 Juni 2024

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

Plt. Ketua Program Studi Informatika,

(Dr. Eng. Niki Prastomo, S.T., M.Sc.)
NIDN: **0419128203**

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini:

Nama	:	George Marcellino Jo
NIM	:	00000045841
Program Studi	:	Informatika
Jenjang	:	S1
Jenis Karya	:	Skripsi

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, 18 Juni 2024

Yang menyatakan

George Marcellino Jo

** 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

”The more you fear something to happen, the more likely it is to occur”

No.1 Murphy Law

KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: IMPLEMENTASI BLOOM FILTER UNTUK PENDETEKSI HATE SPEECH BILINGUAL BAHASA INDONESIA DAN INGGRIS dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada: Mengucapkan terima kasih

1. Bapak Dr. Ninok Leksono, 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 Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Pjs. Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak Arya Wicaksana, S.Kom., M.Eng.Sc. (OCA, CEH, CEI, sebagai Pembimbing pertama yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya tesis ini.
5. Orang Tua, dan keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tesis ini.

Semoga skripsi ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 18 Juni 2024

George Marcellino Jo

IMPLEMENTASI BLOOM FILTER UNTUK PENDETEKSI HATE SPEECH BILINGUAL BAHASA INDONESIA DAN INGGRIS

George Marcellino Jo

ABSTRAK

Hate speech adalah ungkapan atau pidato yang menghina individu atau kelompok berdasarkan (dugaan) keanggotaan dalam suatu kelompok sosial yang dikenali melalui ciri-ciri seperti ras, etnis, jenis kelamin, orientasi seksual, agama, usia, disabilitas fisik atau mental, dan faktor-faktor lainnya. Dikarenakan pertumbuhan pesat internet yang memungkinkan untuk terjadinya komunikasi dari berbagai belahan dunia dengan mudah tanpa ada batasan jarak, interferensi bahasa lain ke dalam bahasa Indonesia pun tidak dapat dihindari, dan membuat terciptanya penggunaan bahasa bilingual di kehidupan sehari-hari. BLOOM-560M akan digunakan untuk mendeteksi apakah sebuah teks bilingual Indonesia-Inggris merupakan teks *hate speech* atau non-*hate speech*. Hasil dari penelitian mendeteksi *hate speech* atau non-*hate speech* pada teks bilingual dengan menggunakan BLOOM Filter pada jumlah *epoch* 6 dengan menggunakan nilai *learning rate* sebesar 2e-5 dan nilai *epsilon* 1e-8, didapatkan *accuracy* 94,28%, *precision* 88,63%, *recall* 92,85%, dan *F1-score* 90,69%. Model juga mendapatkan hasil yang cukup memuaskan dalam mendeteksi teks manual yang di-*forward pass* ke dalam model dengan baik, 5 dari 6 teks yang dimasukkan dapat diprediksi oleh model dengan benar.

Kata kunci: Bilingual, BLOOM Filter, *Code-mixed*, *Hate Speech*, *Machine Learning*

BLOOM FILTER IMPLEMENTATION FOR DETECTING BILINGUAL HATE SPEECH OF INDONESIA AND ENGLISH

George Marcellino Jo

ABSTRACT

Hate speech is an expression or speech that insults individuals or groups based on (alleged) membership in a social group identified by characteristics such as race, ethnicity, gender, sexual orientation, religion, age, physical or mental disability, and other factors. Due to the rapid growth of the internet, which enables communication from various parts of the world easily without distance limitations, language interference into Indonesian cannot be avoided, leading to the use of bilingual language in daily life. BLOOM-560M will be used to detect whether a bilingual Indonesian-English text is hate speech or non-hate speech. The results of the research on detecting hate speech or non-hate speech in bilingual texts using the BLOOM Filter, with 6 epochs, a learning rate of 2e-5, and an epsilon value of 1e-8, showed an accuracy of 94.28%, precision of 88.63%, recall of 92.85%, and an F1-score of 90.69%. The model achieved satisfactory results and was able to effectively detect manually input texts during the forward pass. The model correctly predicted 5 out of the 6 texts that were inputted.

Keywords: *Bilingual, BLOOM Filter, Code-mixed, Hate Speech, Machine Learning*

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR KODE

DAFTAR LAMPIRAN