ANALISIS PENGARUH DATA NORMALISASI DAN NON NORMALISASI PADA DETEKSI *CYBERBULLYING* BERBAHASA INDONESIA DENGAN METODE INDOBERT

TUGAS AKHIR

Oleh:

MITA YANI NURMA PRATIWI

NIM. 1910817320003

PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS TEKNIK

UNIVERSITAS LAMBUNG MANGKURAT

BANJARMASIN

2023

ANALISIS PENGARUH DATA NORMALISASI DAN NON NORMALISASI PADA DETEKSI *CYBERBULLYING* BERBAHASA INDONESIA DENGAN METODE INDOBERT

TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat

Sarjana Strata-1 Teknologi Informasi

Oleh:

MITA YANI NURMA PRATIWI

NIM. 1910817320003

PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS TEKNIK

UNIVERSITAS LAMBUNG MANGKURAT

BANJARMASIN

2023

LEMBAR PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama	: Mita Yani Nurma Pratiwi
NIM	: 1910817320003
Fakultas	: Teknik
Program Studi	: Teknologi Informasi
Judul Tugas Akhir	: Analisis Pengaruh Data Normalisasi dan Non Normalisasi pada Deteksi *Cyberbullying* Berbahasa Indonesia dengan Metode IndoBERT
Pembimbing Utama	: Dr. Ir. Yuslena Sari, S.Kom., M.Kom.

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar akademik di suatu perguruan tinggi, dan sepanjang pengetahuan saya, juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar rujukan.

Banjarmasin, 28 Juli 2023

Mita Yani Nurma Pratiwi

LEMBAR PENGESAHAN

SKRIPSI PROGRAM STUDI S-1 TEKNOLOGI INFORMASI

Analisis Pengaruh Data Normalisasi dan Non Normalisasi pada Deteksi *Cyberbullying* Berbahasa Indonesia dengan Metode IndoBERT

oleh

Mita Yani Nurma Pratiwi (1910817320003)

Telah dipertahankan di depan Tim Penguji pada 3 Juli 2023 dan dinyatakan

L U L U S

Komite Penguji :

Ketua : **Andreyan Rizky Baskara, S.Kom., M.Kom**
NIP 199307032019031011

Anggota 1 : **Eka Setya Wijaya, S.T., M.Kom**
NIP 198205082008011010

Anggota 2 : **Nurul Fathanah Mustamin, S.Pd., M.T**
NIP 199110252019032018

Pembimbing Utama : **Dr. Ir. Yuslena Sari, S.Kom., M.Kom**
NIP 198411202015042002

Banjarbaru, 01 SEP 2023
diketahui dan disahkan oleh:

Wakil Dekan Bidang Akademik

[illegible], S.T., M.T.
NIP 197401071998021001

Koordinator Program Studi S-1 Teknologi Informasi,

Dr. Ir. Yuslena Sari, S.Kom., M.Kom
NIP 198411202015042002

PERSETUJUAN TUGAS AKHIR

ANALISIS PENGARUH DATA NORMALISASI DAN NON NORMALISASI PADA DETEKSI CYBERBULLYING BERBAHASA INDONESIA DENGAN METODE INDOBERT

OLEH
MITA YANI NURMA PRATIWI
NIM. 19108171320003

Telah diperiksa dan terpenuhi semua persyaratan akademik, administrasi, dan disetujui untuk dipertahankan di hadapan dewan penguji

Banjarmasin, 27 Juni 2023

Pembimbing Utama,

Dr. Ir. Yuslena Sari S.Kom., M.Kom.
NIP. 198411202015042002

ABSTRAK

Kasus *cyberbullying* yang marak terjadi di media sosial merupakan salah satu dampak negatif dari kebebasan dan kemudahan yang ditawarkan oleh media sosial. Berdasarkan 2.777 hasil *polling* pada remaja Indonesia berusia 14-24 tahun, 45% dari mereka melaporkan pernah mengalami *cyberbullying*. Hal ini menunjukkan bahwa deteksi *cyberbullying* sangat diperlukan sebagai salah satu bentuk pencegahan guna mengurangi kasus *cyberbullying*. NLP (*Natural Language Processing*) merupakan salah satu cara yang dapat digunakan untuk membantu dalam mendeteksi *cyberbullying*. IndoBERT dan IndoBERTweet merupakan *pre-trained* model yang dapat digunakan untuk deteksi *cyberbullying*. Data yang digunakan untuk melatih model NLP umumnya melewati tahap normalisasi, yaitu tahap mengubah kalimat non baku ke dalam kalimat baku sesuai dengan kaidah bahasa. Pengguna media sosial di Indonesia sendiri jarang menggunakan bahasa baku dalam melakukan komunikasi. Mereka cenderung menggunakan bahasa non baku dengan berbagai istilah dan kosakata baru. Oleh karena itu, penelitian ini dilakukan untuk mengetahui pengaruh data yang melewati proses normalisasi dan tidak melewati proses normalisasi pada model IndoBERT dan IndoBERTweet dalam mendeteksi *cyberbullying*. Data yang digunakan berasal dari Twitter dengan jumlah data 13.446. Model IndoBERT dan IndoBERTweet dengan data yang tidak melewati proses normalisasi memiliki hasil performansi lebih baik dibandingkan dengan data yang melewati proses normalisasi. Model IndoBERT tersebut memiliki nilai *accuracy, f1-score, recall*, dan *precision* sebesar 0.8520, 0.8520, 0.8520, dan 0.8521. Sementara itu, model IndoBERTweet memiliki memiliki nilai *accuracy, f1-score, recall*, dan *precision* sebesar 0.8602, 0.8602, 0.8602, dan 0.8609. Secara keseluruhan, model dengan hasil performansi terbaik adalah model IndoBERTweet dengan data yang tidak dinormalisasi.

Kata kunci: *Cyberbullying*, IndoBERT, IndoBERTweet, NLP, *Text Classification*

ABSTRACT

Cyberbullying cases that have been prevalent on social media are one of the negative consequences of the freedom and convenience offered by these platforms. Based on a survey of 2,777 Indonesian teenagers aged 14-24 years, 45% of them reported experiencing cyberbullying. This underscores the importance of cyberbullying detection as a preventive measure to reduce such incidents. Natural Language Processing (NLP) is one of the methods that can be used to aid in cyberbullying detection. IndoBERT and IndoBERTweet are pre-trained models that can be employed for this purpose. Data used to train NLP models typically undergoes normalization, which involves converting non-standard sentences into standardized sentences according to language norms. However, social media users in Indonesia rarely use standardized language in their communications. They tend to utilize non-standard language with various new terms and vocabulary. Therefore, this study aimed to investigate the influence of data that undergoes normalization and data that does not undergo normalization on the performance of IndoBERT and IndoBERTweet models in detecting cyberbullying. The data used in this research was obtained from Twitter, consisting of 13,446 samples. The results show that IndoBERT and IndoBERTweet models trained on data that did not undergo normalization achieved better performance compared to those trained on normalized data. The IndoBERT model exhibited accuracy, f1-score, recall, and precision values of 0.8520, 0.8520, 0.8520, and 0.8521, respectively. On the other hand, the IndoBERTweet model demonstrated accuracy, f1-score, recall, and precision values of 0.8602, 0.8602, 0.8602, and 0.8609, respectively. Overall, the model with the best performance was the IndoBERTweet model with non-normalized data.

Keywords: Cyberbullying, NLP, IndoBERT, IndoBERTweet, Text Classification

LEMBAR PERSEMBAHAN

Penulis mempersembahkan Tugas Akhir ini kepada:

1. Ayah, Ibu, Alm. Nenek, Adik serta keluarga tercinta yang telah memberikan doa, motivasi, dan dukungan saat menghadapi suatu masalah dalam keberlangsungan penyelesaian Tugas Akhir ini.
2. Ibu Dr. Ir. Yuslena Sari, S.Kom., M.Kom. selaku Koordinator Program Studi Teknologi Informasi dan Dosen Pembimbing Utama yang selalu menyempatkan waktunya untuk memberikan bimbingan dan arahan kepada mahasiswa teknologi informasi dan penulis untuk segera menyelesaikan Tugas Akhir ini**.**
3. Bapak Andry Fajar Zulkarnain, S.ST., M.T selaku Dosen Pembimbing Akademik yang selalu menyempatkan waktu untuk memberikan bimbingan, arahan, dan dukungan kepada penulis selama mengikuti proses perkuliahan.
4. Bapak Andreyan Rizky Baskara, S.Kom,. M.Kom, Bapak Eka Setya Wijaya, S.T., M.Kom., dan Ibu Nurul Fathanah Mustamin, S.Pd., M.T. selaku Dosen Penguji yang telah memberikan banyak arahan, masukan, dan solusi untuk penelitian ini.
5. Seluruh Dosen beserta Staf Program Studi Teknologi Informasi yang turut membantu dan mengarahkan dalam penyelesaian Tugas Akhir ini.
6. Seluruh teman Program Studi Teknologi Informasi Angkatan 2019 (SAKTI) beserta adik dan kakak tingkat yang telah banyak membantu serta memberikan dukungan dari perkuliahan hingga penyelesaian Tugas Akhir ini.
7. Teman-teman yang telah membantu dalam melakukan pemrosesan data untuk penelitian ini.
8. Eugynia Jessica Virgynia Rahman, Nurul Hikmah, Vania Laili Rahmah, Rahmadani, Nur Azlina, Nina Hafizah, Resya Cahaya Maharani, dan Rif'at Maulana yang selalu memberikan semangat dan dukungan, menghibur, mendengarkan keluh kesah, dan membantu selama proses studi berjalan hingga proses penyelesaian studi dilakukan.

9. Teman seperjuangan saya sejak SMA hingga detik ini, Mariati, Nurul Firdaus, Talitha Az Zahra Rosadi, Siti Rosita Sari, Irma Maulida, Deftri Sekar Ningrum, dan Nur Jannah yang selalu memberikan semangat dan dukungan, menghibur, mendengarkan keluh kesah, dan membantu selama SMA hingga proses penyelesaian studi dilakukan.
10. Diri sendiri yang tidak menyerah untuk terus belajar hingga detik ini.

KATA PENGANTAR

Puji dan syukur penulis ucapkan kehadirat Allah Subhanahu wa ta'ala yang telah melimpahkan rahmat dan hidayah-Nya. Sehingga penulis dapat menyelesaikan Tugas Akhir dengan Judul "*Analisis Pengaruh Data Normalisasi dan Non Normalisasi pada Deteksi Cyberbullying Berbahasa Indonesia dengan Metode IndoBERT*". Penulis menyadari penyelesaian laporan ini tidak akan terwujud tanpa adanya bimbingan dan bantuan dari berbagai pihak. Oleh karena itu, dalam kesempatan ini penulis menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Rektor Universitas Lambung Mangkurat, Bapak Prof. Dr. Ahmad, S.E., M.Si., yang memimpin dan memanajemen jalannya seluruh perkuliahan yang ada di Universitas Lambung Mangkurat.
2. Dekan Fakultas Teknik, Bapak Prof. Dr. Ir. Irphan Fitrian Radam, S.T., M.T., IPU, yang memberikan layanan terbaik dalam perkuliahan, terkhusus pada pelaksanaan Tugas Akhir di lingkungan Fakultas Teknik.
3. Koordinator Program Studi Teknologi Informasi serta Pembimbing Utama, Ibu Dr. Ir. Yuslena Sari, S.Kom. M.Kom. yang telah memberikan arahan, bimbingan, serta solusi dalam penyelesaian Tugas Akhir.
4. Dosen-dosen beserta staff di Program Studi Teknologi Informasi yang telah mengarahkan dan teman-teman yang membantu dalam proses penyelesaian Tugas Akhir.

Akhir kata, penulis menyampaikan terimakasih kepada semua pihak yang turut serta membantu dalam penyelesaian laporan Tugas Akhir ini. Penulis mengharap saran dan kritik yang membangun demi perbaikan dan penyempurnaan laporan ini. Semoga laporan ini dapat bermanfaat baik untuk pribadi, pembaca dan semua pihak yang membutuhkan.

Banjarmasin, Agustus 2023

Penulis

Mita Yani Nurma Pratiwi

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN