

**KLASIFIKASI MULTILABEL TERHADAP CYBERBULLYING
MENGUNAKAN BIDIRECTIONAL LSTM DAN INDOBERT
DENGAN RANDOM SEARCH HYPERPARAMETER
OPTIMIZATION**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

RIFAN AZIS ARDIANSYAH

21.11.4525

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

**KLASIFIKASI MULTILABEL TERHADAP CYBERBULLYING
MENGUNAKAN BIDIRECTIONAL LSTM DAN INDOBERT
DENGAN RANDOM SEARCH HYPERPARAMETER
OPTIMIZATION**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

RIFAN AZIS ARDIANSYAH

21.11.4525

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**KLASIFIKASI MULTILABEL TERHADAP CYBERBULLYING
MENGUNAKAN BIDIRECTIONAL LSTM DAN INDOBERT
DENGAN RANDOM SEARCH HYPERPARAMETER
OPTIMIZATION**

yang disusun dan diajukan oleh

Rifan Azis Ardiansyah

21.11.4525

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 30 Juni 2025

Dosen Pembimbing,



Windha Mega Pradnya Dhuhita, M.Kom.
NIK. 190302185

HALAMAN PENGESAHAN

SKRIPSI

**KLASIFIKASI MULTILABEL TERHADAP CYBERBULLYING
MENGUNAKAN BIDIRECTIONAL LSTM DAN INDOBERT
DENGAN RANDOM SEARCH HYPERPARAMETER
OPTIMIZATION**

yang disusun dan diajukan oleh

Rifan Azis Ardiansyah

21.11.4525

Telah dipertahankan di depan Dewan Penguji
pada tanggal 30 Juni 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Ike Verawati, M.Kom.

NIK. 190302237

Nafiatun Sholihah, S.Kom., M.Cs

NIK. 190302524

Windha Mega Pradnya Dhuhiha, M.Kom.

NIK. 190302185

Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 30 Juni 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.

NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Rifan Azis Ardiansyah
NIM : 21.11.4525

Menyatakan bahwa Skripsi dengan judul berikut:

Klasifikasi Multilabel Terhadap Cyberbullying Menggunakan Bidirectional LSTM dan IndoBERT Dengan Random Search Hyperparameter Optimization

Dosen Pembimbing : Windha Mega Pradnya Dhuhita, M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 30 Juni 2025

Yang Menyatakan,



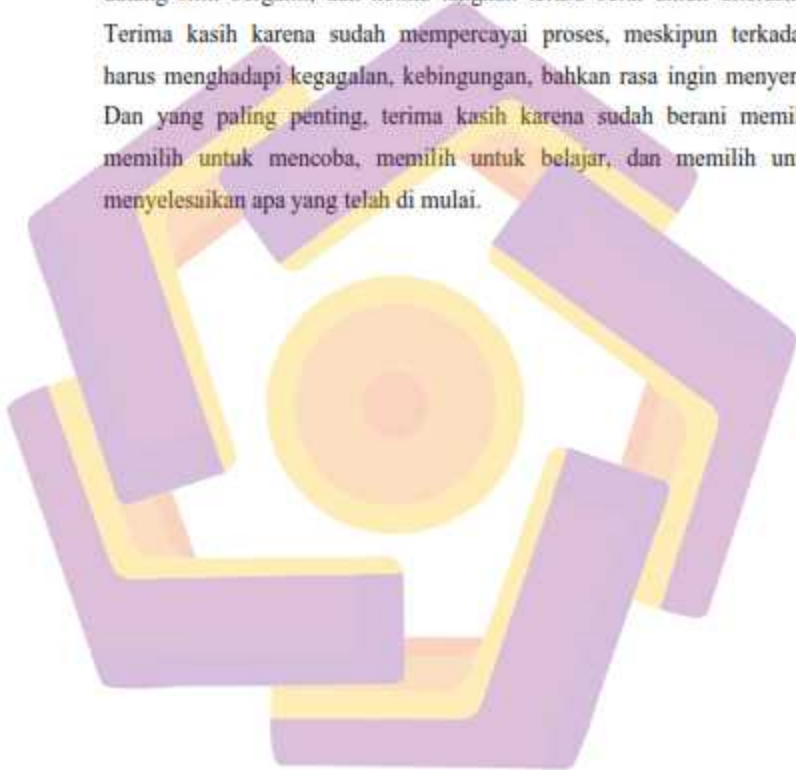
Rifan Azis Ardiansyah

HALAMAN PERSEMBAHAN

Alhamdulillah, segala puji syukur kehadiran Allah SWT atas segala nikmat berupa kesehatan, kekuatan, dan inspirasi yang sangat banyak dalam proses penyelesaian skripsi ini. Shalawat serta salam senantiasa tercurahkan kepada Nabi Muhammad SAW. Dengan penuh rasa syukur dan ketulusan hati, penulis menyadari bahwa perjalanan hingga titik ini bukanlah sesuatu yang mudah. Proses penyusunan skripsi ini telah dilalui dengan tekad, doa, serta dukungan dari orang-orang tersayang. Skripsi sederhana ini penulis persembahkan dengan penuh cinta dan rasa hormat kepada:

1. Kedua Orang Tua, Ayah Sarno dan Ibu Asih Sayekti, gelar sarjana ini penulis persembahkan untuk kedua orang tua tercinta. Terima kasih atas doa yang tidak ada putusnya, dan selalu memberikan dukungan berupa moril maupun materiil yang tak terhingga kepada penulis sehingga penulis mampu menyelesaikan studi hingga selesai. Semoga keberkahan dari Allah SWT senantiasa menyertai kehidupan beliau, serta diberikan kesehatan dan panjang umur.
2. Kakak Tercinta, Irna Kurniawati, yang selalu menjadi panutan dalam setiap langkah. Terima kasih atas segala motivasi, dukungan, dan kasih sayang. Terima kasih sudah menjadi kakak paling hebat bagi penulis.
3. Teman-Teman Seperjuangan Informatika 2021, Universitas Amikom Yogyakarta yang tidak bisa disebutkan satu persatu. Terima kasih atas kebersamaan, kerja keras, dan semangat yang tidak pernah padam dalam menghadapi segala tantangan. Semoga persahabatan dan solidaritas yang terjalin selama ini terus berkembang dan membawa manfaat bagi kita semua.
4. Almamater Tercinta, Universitas Amikom Yogyakarta, yang telah menjadi tempat menimba ilmu, mengembangkan potensi, dan membentuk karakter selama masa studi. Semoga karya ini dapat menjadi kontribusi kecil yang mencerminkan dedikasi dan rasa terima kasih penulis kepada institusi yang telah memberikan landasan kuat untuk melangkah ke masa depan.

5. Terakhir, penulis ingin mengucapkan terima kasih kepada satu sosok yang selama ini diam-diam berjuang tanpa henti, seorang laki-laki sederhana dengan impian yang tinggi. Terima kasih kepada diri sendiri, Rifan Azis Ardiansyah, terima kasih karena telah bertahan sejauh ini. Terima kasih karena tidak menyerah ketika jalan di depan terasa gelap, ketika keraguan datang silih berganti, dan ketika langkah terasa berat untuk diteruskan. Terima kasih karena sudah mempercayai proses, meskipun terkadang harus menghadapi kegagalan, kebingungan, bahkan rasa ingin menyerah. Dan yang paling penting, terima kasih karena sudah berani memilih, memilih untuk mencoba, memilih untuk belajar, dan memilih untuk menyelesaikan apa yang telah di mulai.



KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadirat Allah SWT atas limpahan rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul **“Klasifikasi Multilabel Terhadap Cyberbullying Menggunakan Bidirectional LSTM dan IndoBERT Dengan Random Search Hyperparameter Optimizattion”**. Penulisan skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta. Penulis menyadari bahwa tanpa bimbingan dan bantuan dari berbagai pihak, maka penulisan skripsi ini tidak akan berjalan lancar. Oleh karena itu, melalui kata pengantar ini, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta.
2. Ibu Prof. Dr. Kusriani, M.Kom., selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Ibu Eli Pujastuti, M.Kom., selaku Ketua Program Studi Informatika Universitas Amikom Yogyakarta.
4. Ibu Windha Mega Pradnya Dhuhiya, M.Kom., selaku Dosen Pembimbing Skripsi yang telah meluangkan waktu, tenaga, dan pikiran untuk mengarahkan penulis dalam penyusunan skripsi ini. Ucapan terima kasih disampaikan atas segala ilmu dan wawasan yang telah diberikan.
5. Ibu Ike Verawati, M.Kom. dan Ibu Nafiatun Sholihah, S.Kom., M.Cs., selaku Dosen Penguji yang telah memberikan masukan dan saran dalam penulisan skripsi ini.
6. Ibu Rumini, M.Kom., selaku Dosen Wali yang telah mendukung dan memberikan arahan kepada penulis selama proses studi.
7. Seluruh Staf Dosen dan Pegawai Fakultas Ilmu Komputer Universitas Amikom Yogyakarta yang telah membantu kelancaran penulis dalam menyelesaikan studi dan pelayanan terbaiknya selama masa perkuliahan.

8. Keluarga penulis, Ayah, Ibu, Kakak, dan keluarga besar tercinta yang senantiasa memberikan doa, dukungan, dan semangat kepada penulis dalam penyusunan skripsi ini.

Akhir kata, penulis dapat menyadari tanpa ridho dan pertolongan dari Allah SWT, serta bantuan, dukungan, motivasi dari segala pihak skripsi ini tidak dapat diselesaikan. Kepada semua pihak yang telah memberikan bantuan dalam penulisan ini, penulis ucapkan banyak terima kasih dan semoga Allah SWT membalas segera kebaikan kalian. *Aamiin Yarabbal'alam*.

Yogyakarta, 30 Juni 2025

Penulis

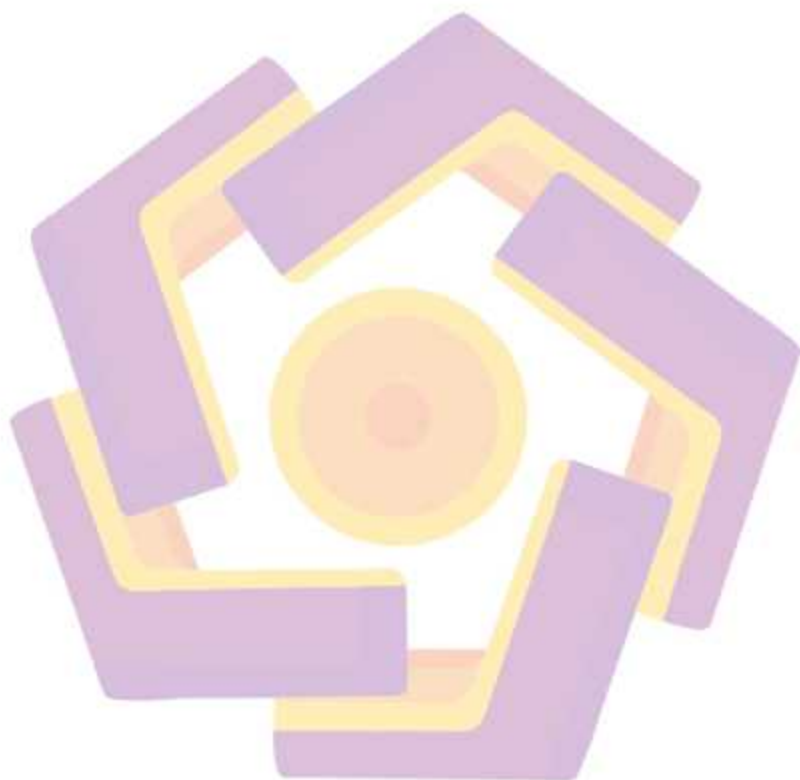


DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN PERSEMBAHAN.....	v
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xiv
DAFTAR LAMPIRAN.....	xv
DAFTAR LAMBANG DAN SINGKATAN.....	xvi
DAFTAR ISTILAH.....	xvii
INTISARI.....	xviii
ABSTRACT.....	xix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Sistematika Penulisan.....	4

BAB II TINJAUAN PUSTAKA.....	5
2.1 Studi Literatur.....	5
2.1 Dasar Teori.....	14
2.2.1. Cyberbullying.....	14
2.2.2. Deep Learning.....	14
2.2.3. Crawling data.....	15
2.2.4. Preprocessing data.....	15
2.2.5. Long Short Term Memory (LSTM).....	16
2.2.6. Bidirectional Long Short Term Memory (BiLSTM).....	18
2.2.7. BERT.....	19
2.2.8. IndoBERT.....	20
2.2.9. Random search.....	21
2.2.10. Confusion Matrix.....	22
BAB III METODE PENELITIAN.....	24
3.1 Objek Penelitian.....	24
3.2 Alur Penelitian.....	24
3.2.1 Crawling.....	25
3.2.2 Labeling.....	25
3.2.3 Preprocessing Data.....	25
3.2.4 Data Splitting.....	26
3.2.5 Model BiLSTM.....	26
3.2.6 Evaluasi Model.....	27

3.2.7	Optimasi Random Search.....	27
3.2.8	Perbandingan Sebelum dan Sesudah Optimasi.....	27
3.3	Alat dan Bahan	27
3.3.1	Data Penelitian	27
3.3.2	Alat/Instrumen.....	28
BAB IV HASIL DAN PEMBAHASAN		30
4.1	Crawling.....	30
4.2	Labeling.....	31
4.3	Preprocessing.....	32
4.1	Case folding	33
4.2	Cleaning.....	34
4.3	Normalisasi	35
4.4	Augmentasi	36
4.5	Tokenisasi	37
4.4	Splitting Data.....	38
4.5	Pembangunan dan Pelatihan Model BiLSTM.....	39
4.6	Optimasi Random Search	46
4.7	Perbandingan Sebelum dan Sesudah Optimasi	52
BAB V PENUTUP.....		55
5.1	Kesimpulan.....	55
5.2	Saran	56
REFERENSI		57



DAFTAR TABEL

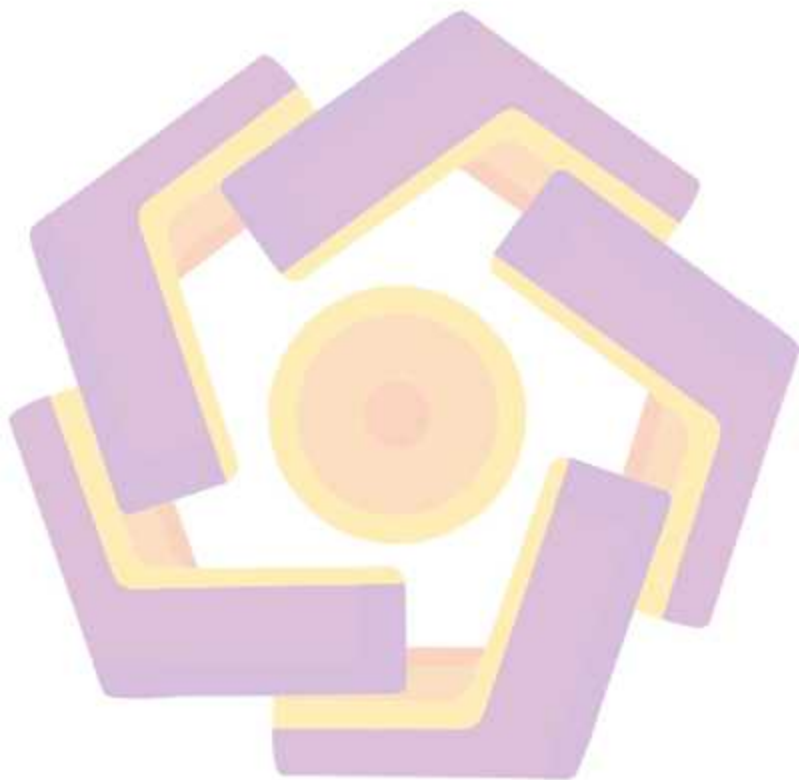
Tabel 2. 1 Keaslian Penelitian.....	8
Tabel 4. 1 Hasil Crawling dan Penggabungan Dataset.....	30
Tabel 4. 2 Hasil Case Folding.....	33
Tabel 4. 3 Hasil Cleaning Data	34
Tabel 4. 4 Hasil Normalisasi Data	35
Tabel 4. 5 Hasil Tokenisasi IndoBERT	37
Tabel 4. 6 Splitting Data	38
Tabel 4. 7 Arsitektur Model BiLSTM.....	39
Tabel 4. 8 Hasil Evaluasi Baseline Model	40
Tabel 4. 9 Classification Report Setiap Label.....	42
Tabel 4. 10 Hasil Kombinasi Hyperparameter Terbaik.....	46
Tabel 4. 11 Hasil Evaluasi Setelah Random Search	47
Tabel 4. 12 Classification Report Setelah Optimasi	48
Tabel 4. 13 Perbandingan Metrik Evaluasi Sebelum dan Setelah Optimasi.....	52

DAFTAR GAMBAR

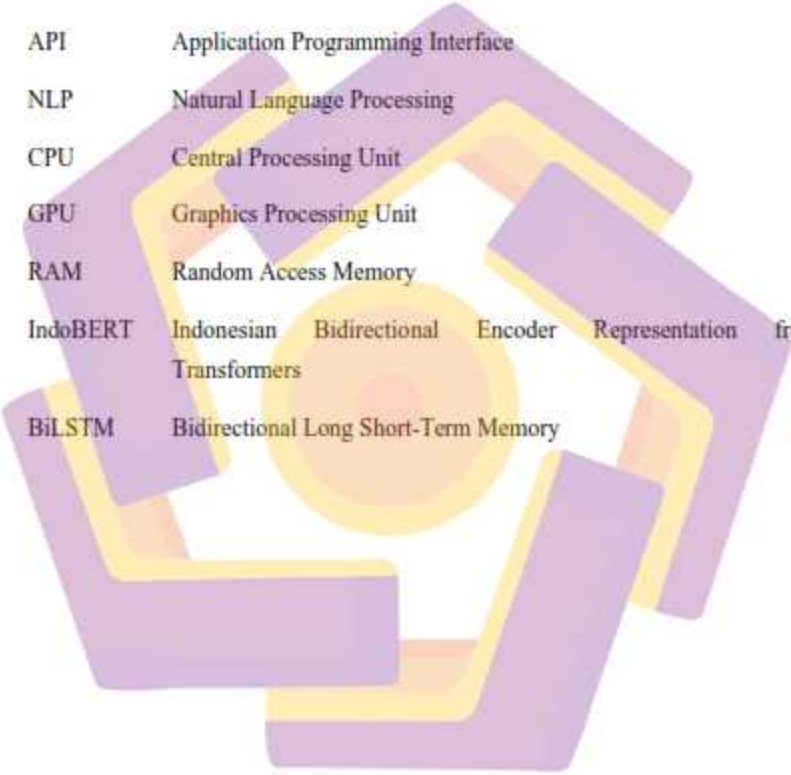
Gambar 2. 1 Arsitektur Long Short Term Memory (LSTM).....	17
Gambar 2. 2 Arsitektur Bidirectional Long Short Term Memory (BiLSTM)	18
Gambar 2. 3 Encoder dan Decoder	20
Gambar 2. 4 Pre-training dan Fine-Tuning BERT.....	21
Gambar 2. 5 Ilustrasi Pengambilan Nilai Algoritma Random Search	22
Gambar 2. 6 Confusion Matrix	23
Gambar 3. 1 Alur Penelitian.....	24
Gambar 4. 1 Hasil Labeling Data Manual	31
Gambar 4. 2 Distribusi Label dalam Dataset	32
Gambar 4. 3 Distribusi Label Sebelum dan Sesudah Augmentasi.....	36
Gambar 4. 4 Grafik Pelatihan Model	41
Gambar 4. 5 Confusion Matrix Setiap Label	45
Gambar 4. 6 Grafik Training Loss dan Training Accuracy Final Model.....	48
Gambar 4. 7 Confusion Matrix Setelah Optimasi.....	51
Gambar 4. 8 Perbandingan Metrik Sebelum dan Sesudah Optimasi	54
Gambar 4. 9 Visualisasi Co-occurrence Matrix Sebelum dan Sesudah Optimasi	54

DAFTAR LAMPIRAN

Lampiran 1. Link Dataset.....	62
Lampiran 2. Link Program Crawling dan Klasifikasi.....	62



DAFTAR LAMBANG DAN SINGKATAN



w	Bobot
b	Bias
RNN	Reccurent Neural Network
API	Application Programming Interface
NLP	Natural Language Processing
CPU	Central Processing Unit
GPU	Graphics Processing Unit
RAM	Random Access Memory
IndoBERT	Indonesian Bidirectional Encoder Representation from Transformers
BiLSTM	Bidirectional Long Short-Term Memory

DAFTAR ISTILAH



Tweet	Pesan singkat yang diposting di platform media sosial X.
Cyberbullying	Tindakan perundungan atau kekerasan verbal yang dilakukan melalui media digital.
Klasifikasi	Proses mengelompokkan data ke dalam kategori.
Multilabel	Jenis klasifikasi di mana satu data dapat memiliki lebih dari satu label secara bersamaan.
Embedding	Representasi vektor dari kata atau token dalam dimensi kontinu.
Augmentasi	Proses menambah atau memodifikasi data.
Accuracy	Proporsi prediksi yang benar dibandingkan total prediksi.
Precision	Proporsi prediksi positif yang benar dari seluruh prediksi positif.
Recall	Proporsi prediksi positif yang benar dari seluruh data yang sebenarnya positif.
F1-Score	Rata-rata harmonik dari precision dan recall.
Hamming Loss	Proporsi label yang salah diklasifikasikan terhadap total label.

INTISARI

Cyberbullying merupakan bentuk kekerasan verbal di dunia digital yang dapat menimbulkan dampak negatif terhadap kondisi psikologis dan sosial korbannya. Permasalahan ini semakin kompleks dengan banyaknya jenis cyberbullying yang terjadi di media sosial, sementara upaya deteksi terhadap berbagai bentuk cyberbullying masih tergolong minim. Permasalahan ini menunjukkan pentingnya pengembangan sistem klasifikasi yang mampu mengenali berbagai jenis ujaran kekerasan dalam teks, khususnya yang bersifat multilabel.

Penelitian ini bertujuan untuk membangun model klasifikasi multilabel cyberbullying dalam teks berbahasa Indonesia menggunakan pendekatan deep learning dengan memanfaatkan arsitektur BiLSTM dan embedding dari IndoBERT. Data yang digunakan terdiri dari teks berbahasa Indonesia dengan enam kategori label, yaitu penghinaan, pelecehan, ancaman, ujaran kebencian, body shaming, dan bukan cyberbullying. Tahapan penelitian meliputi preprocessing data, augmentasi, tokenisasi, dan pelatihan model. Selain itu, dilakukan optimasi hyperparameter menggunakan metode Random Search untuk memperoleh performa model terbaik. Evaluasi dilakukan menggunakan metrik akurasi, precision, recall, F1-score, dan Hamming Loss.

Model hasil optimasi menunjukkan peningkatan performa yang signifikan, dengan akurasi mencapai 72,47%, F1-score mikro sebesar 80,10%, dan Hamming Loss sebesar 0,0819. Hasil ini menunjukkan bahwa BiLSTM dan IndoBERT yang dioptimasi menggunakan metode Random Search mampu memberikan kinerja yang lebih efektif dalam klasifikasi multilabel teks cyberbullying berbahasa Indonesia. Temuan ini diharapkan dapat memberikan kontribusi bagi pengembangan sistem deteksi cyberbullying yang lebih efektif, peneliti di bidang NLP, serta lembaga yang bergerak dalam perlindungan pengguna digital.

Kata kunci: Klasifikasi Multilabel, Cyberbullying, BiLSTM, IndoBERT, Random Search.

ABSTRACT

Cyberbullying is a form of verbal violence in the digital world that can have a negative impact on the psychological and social conditions of its victims. This problem is increasingly complex with the many types of cyberbullying that occur on social media, while detection efforts for various forms of cyberbullying are still relatively minimal. This problem shows the importance of developing a classification system that is able to recognize various types of violent speech in text, especially those that are multilabel.

This research aims to build a multilabel cyberbullying classification model in Indonesian text using a deep learning approach by utilizing the BiLSTM architecture and embedding from IndoBERT. The data used consists of Indonesian texts with six label categories, namely insults, harassment, threats, hate speech, body shaming, and not cyberbullying. The research stages include data preprocessing, augmentation, tokenization, and model training. In addition, hyperparameter optimization is carried out using the Random Search method to obtain the best model performance. The evaluation used the accuracy, precision, recall, F1-score, and Hamming loss metrics.

The optimized model shows significant performance improvement, with accuracy reaching 72.47%, micro F1-score of 80.10%, and Hamming Loss of 0.0819. These results show that BiLSTM and IndoBERT optimized using Random Search method can provide more effective performance in multilabel classification of Indonesian cyberbullying text. This finding is expected to contribute to the development of a more effective cyberbullying detection system, researchers in the field of NLP, as well as institutions engaged in digital user protection.

Keyword: Multilabel Classification, Cyberbullying, BiLSTM, IndoBERT, Random Search.