

**PERBANDINGAN KINERJA LSTM PADA UNIT 8,16, DAN 32  
UNIT DENGAN WORD EMBEDDING TF-IDF, WORD2VEC,  
DAN FASTTEXT DALAM KASUS CYBERBULLYING DI  
MEDIA SOCIAL INDONESIA**

**SKRIPSI**

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



disusun oleh

**RICO BASYAR BARWAQAH**

**18.11.2556**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2024**

**PERBANDINGAN KINERJA LSTM PADA UNIT 8,16, DAN 32  
UNIT DENGAN WORD EMBEDDING TF-IDF, WORD2VEC,  
DAN FASTTEXT DALAM KASUS CYBERBULLYING DI  
MEDIA SOCIAL INDONESIA**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**RICO BASYAR BARWAQAH**

**18.11.2556**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2024**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**PERBANDINGAN KINERJA LSTM PADA UNIT 8,16, DAN 32 UNIT  
DENGAN WORD EMBEDDING TF-IDF, WORD2VEC, DAN FASTTEXT  
DALAM KASUS CYBERBULLYING DI MEDIA SOCIAL INDONESIA**

**Halaman Persetujuan**  
yang disusun dan diajukan oleh

**Rico Basyar Barwaqah**

**18.11.2556**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 19 Febuari 2024

**Dosen Pembimbing,**



**Arief Setyanto, S.Si., MT., Ph.D**

**NIK. 190302036**

**HALAMAN PENGESAHAN**

**SKRIPSI**



**PERBANDINGAN KINERJA LSTM PADA UNIT 8,16, DAN 32 UNIT DENGAN WORD  
EMBEDDING TF-IDF, WORD2VEC, DAN FASTTEXT DALAM KASUS  
CYBERBULLYING DI MEDIA SOCIAL INDONESIA**

yang disusun dan diajukan oleh

**Rico Basyar Barwaqah**

**18.11.2556**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 19 Febuari 2024

<b>Nama Penguji</b>	<b>Susunan Dewan Penguji</b>	<b>Tanda Tangan</b>
<b><u>Arif Akbarul Huda, S.Si, M.Eng</u></b> NIK. 190302287		
<b><u>Wahid Miftahul Ashari, S.Kom., M.T</u></b> NIK. 190302452		
<b><u>Arief Setyanto, S.Si., MT., Ph.D</u></b> NIK. 190302036		

Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 19 Febuari 2024

**DEKAN FAKULTAS ILMU KOMPUTER**



**Hanif Al Fatta, S.Kom., M.Kom., Ph.D.**  
NIK. 190302096

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

**Nama mahasiswa : Rico Basyar Barwaqah**  
**NIM : 18.11.2556**

Menyatakan bahwa Skripsi dengan judul berikut:

**Perbandingan Kinerja LSTM Pada Unit 8,16, Dan 32 Unit Dengan Word Embedding TF-IDF, Word2Vec, Dan FastText Dalam Kasus Cyberbullying Di Media Social Indonesia**

Dosen Pembimbing : Arief Setyanto, S.Si., MT., Ph.D

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, 19 Febuari 2024

Yang Menyatakan,



Rico Basyar Barwaqah

## KATA PENGANTAR

Puji syukur atas kehadiran Allah SWT yang telah melimpahkan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan sebaik-baiknya sebagai salah-satu syarat untuk dapat menempuh ujian sarjanan pada Fakultas Ilmu Komputer (FIK) Program Studi Informatika di Universitas Amikom Yogyakarta.

Dengan selesainya skripsi ini, maka penulis mengucapkan terima kasih kepada:

1. Allah SWT yang selalu memberikan nikmat, karunia serta kemudahan yang sangat berlimpah dalam proses penyusunan skripsi ini.
2. Prof. Dr. M.Suyanto, MM selaku rector Universitas Amikom Yogyakarta
3. Bapak Arief Setyanto, S.Si., MT., Ph.D selaku dosen pembimbing yang telah membimbing penulis dalam menyusun skripsi ini.

Penulis menyadari dalam penulisan skripsi ini masih ada kekurangan sehingga kritik dan saran yang membangun sangat penulis harapkan. Penulis berharap semoga skripsi ini bisa memberikan manfaat kepada para pembaca.

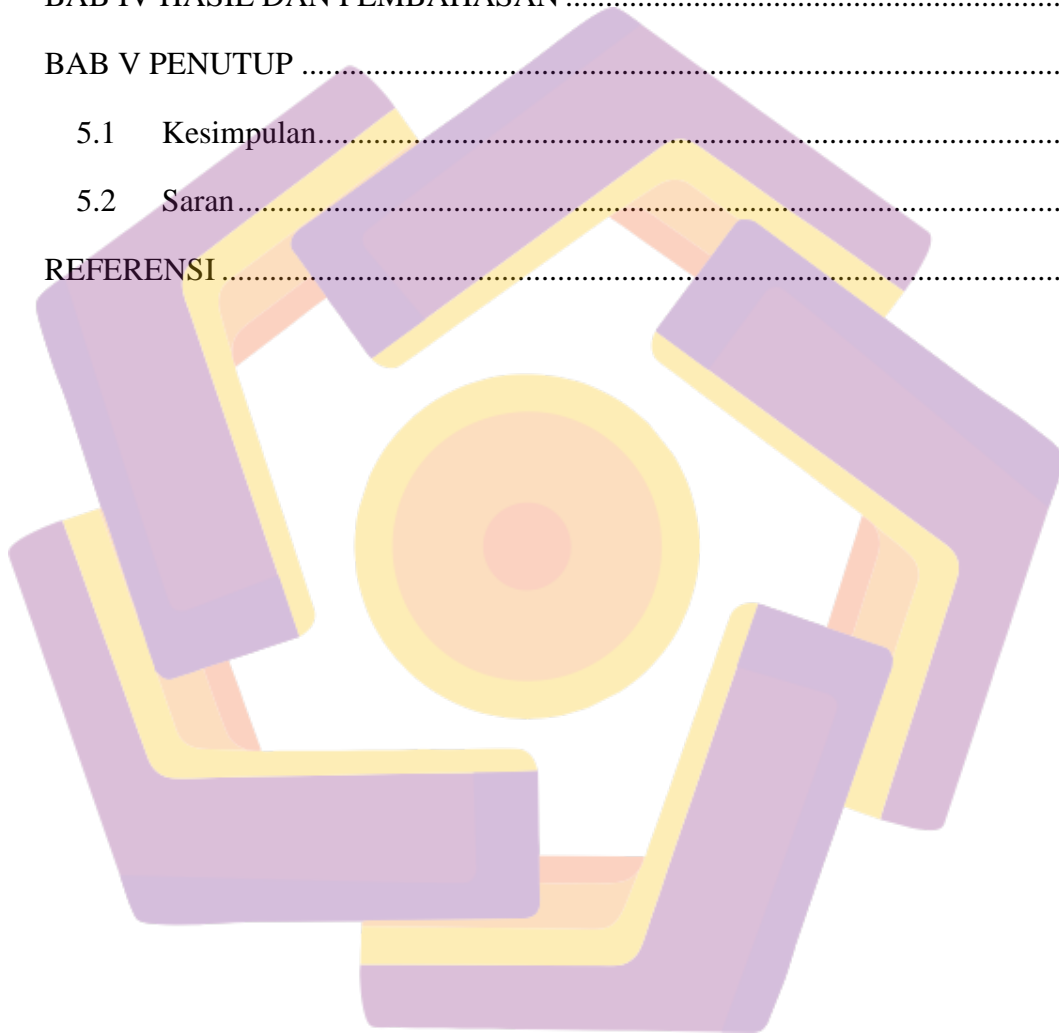
Yogyakarta, 19 Febuari 2024

Penulis

## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN .....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI .....	iv
KATA PENGANTAR .....	v
DAFTAR ISI.....	vi
DAFTAR TABEL.....	viii
DAFTAR GAMBAR .....	ix
DAFTAR LAMBANG DAN SINGKATAN .....	x
INTISARI .....	xi
<i>ABSTRACT</i> .....	xii
<b>BAB I PENDAHULUAN</b> .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan.....	5
<b>BAB II TINJAUAN PUSTAKA</b> .....	6
2.1 Studi Literatur .....	6
2.2 Dasar Teori.....	12
<b>BAB III METODE PENELITIAN</b> .....	20
3.1 Dataset.....	20

3.2	Alur Penelitian.....	22
3.2.1	Preprocessing.....	24
3.2.2	Preprocessing.....	25
3.2.3	Long Short-Term Memory (LSTM) .....	26
BAB IV HASIL DAN PEMBAHASAN .....		30
BAB V PENUTUP .....		41
5.1	Kesimpulan.....	41
5.2	Saran.....	41
REFERENSI.....		42





## DAFTAR TABEL

Tabel 2.1. Keaslian penelitian.....	8
Tabel 2.2. Confusion matrix.....	17
Tabel 3.1. Dataset.....	20
Tabel 3.2. Contoh data yang dieliminasi secara acak.....	21
Tabel 3.3. Contoh data yang tidak tereliminasi.....	21
Tabel 3.4. Text Preprocessing.....	24
Tabel 3.5. Tokenization.....	25
Tabel 3.6. Arsitektur LSTM 8 Unit.....	26
Tabel 3.7. Arsitektur LSTM 16 Unit.....	27
Tabel 3.8. Arsitektur LSTM 32 Unit.....	28
Tabel 4.1. Hasil Evaluasi Accuracy.....	34
Tabel 4.2. Hasil Evaluasi Precision.....	35
Tabel 4.3. Hasil Evaluasi Recall.....	36
Tabel 4.4. Hasil Evaluasi F1-Score.....	36
Tabel 4.5. Perbandingan penelitian sebelumnya dengan topic yang sama.....	40

## DAFTAR GAMBAR

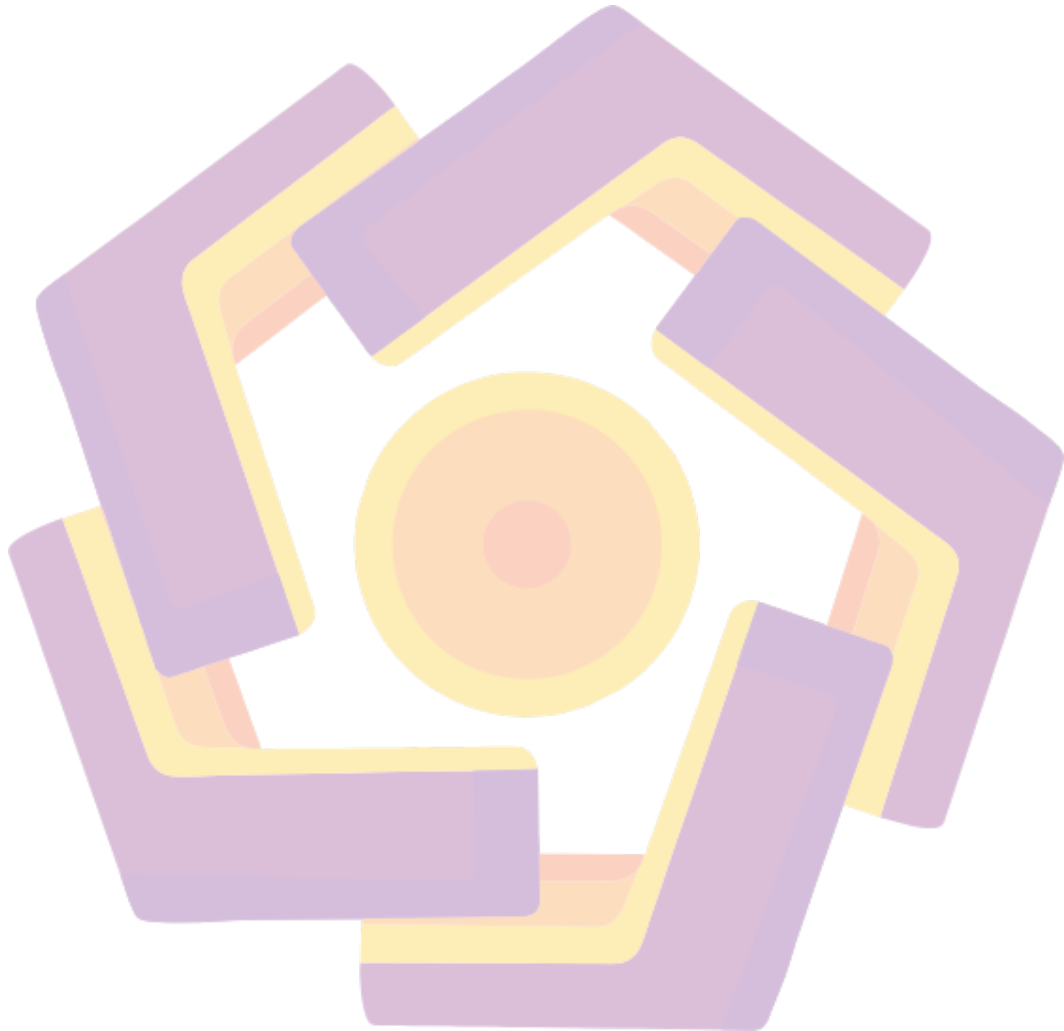
Gambar 3.1. kata yang sering muncul pada label non-cyberbullying.....	22
Gambar 3.2. kata yang sering muncul pada label cyberbullying.....	22
Gambar 3.3. Alur proses penelitian.....	23
Gambar 4.1. LSTM 8 unit dengan TF-IDF.....	30
Gambar 4.2. LSTM 8 unit dengan Word2Vec.....	31
Gambar 4.3. LSTM 8 unit dengan FastText.....	31
Gambar 4.4. LSTM 16 unit dengan TF-IDF.....	32
Gambar 4.5. LSTM 16 unit dengan Word2Vec.....	32
Gambar 4.6. LSTM 16 unit dengan FastText.....	32
Gambar 4.7. LSTM 32 unit dengan TF-IDF.....	33
Gambar 4.8. LSTM 32 unit dengan Word2Vec.....	33
Gambar 4.9. LSTM 32 unit dengan FastText.....	33
Gambar 4.10. Confusion Matrix TF-IDF dengan unit LSTM yang berbeda.....	37
Gambar 4.11. Confusion Matrix Word2Vec dengan unit LSTM yang berbeda...38	
Gambar 4.12. Confusion Matrix FastText dengan unit LSTM yang berbeda.....	39



## DAFTAR LAMBANG DAN SINGKATAN

TF-IDF      Term Frequency-Inverse Document Frequency

LSTM      Long Short-Term Memory



## INTISARI

Media sosial sebagai platform yang dimana individu dapat berinteraksi satu sama lain. Media sosial tidak hanya memungkinkan pengguna untuk membagikan konten yang mereka buat tetapi juga dapat menghasilkan banyak opini dari penggunanya. Dari sekian banyak sudut pandang beberapa mengarah ke perbedaan pendapat dan, dari perbedaan pendapat tersebut sayangnya beberapa individu melakukan tindakan cyberbullying di platform ini. Tujuan dari penelitian ini adalah mengetahui perbedaan performa klasifier teks pada arsitektur LSTM dengan unit yang berbeda untuk klasifikasi pada cyberbully dengan menggunakan teknik word embedding TF-IDF, Word2vec, dan FastText. Dataset yang digunakan adalah kombinasi komentar Instagram yang diambil dari Kaggle dan tweet yang bersumber dari penelitian sebelumnya. Data yang diperoleh untuk penelitian ini terdiri dari total 1570 entri, data tersebut seimbang dengan 785 diberi label sebagai '0' untuk yang bukan cyberbullying dan 785 sebagai '1' untuk cyberbullying. Ukuran pengujian data adalah 30% dari total entri dan pengujian data yang dilakukan sebanding antara label '0' dan '1'. Hasil tes menunjukkan dari teknik word embedding yang berbeda TF-IDF memiliki nilai accuracy yang paling tinggi, tetapi menurut gambar diagram garis jika unit LSTM ditambahkan maka TF-IDF memiliki potensi overfitting sedangkan Word2Vec memiliki kinerja yang optimal dan FastText tidak bisa melakukan pembelajaran pada kasus ini.

**Kata kunci:** Cyberbullying, LSTM, TF-IDF, Word2vec, FastText, klasifikasi teks, representasi teks.

## **ABSTRACT**

*Social media serves as a platform where individuals interact with each other. Social media not only allows users to share their created content but also generates a wide range of opinions. From these many viewpoints, some may lead to disagreements and, unfortunately, certain individuals may resort to cyberbullying on these platforms. The aim of this research is to determine the performance differences of text classifiers on LSTM architectures with different units for the classification of cyberbullying using the TF-IDF, Word2Vec, and FastText word embedding techniques. The dataset utilized is a combination of Instagram comments obtained from Kaggle and tweets from previous research. The data for this study comprises a total of 1570 entries, with a balanced distribution of 785 labeled as '0' for non-cyberbullying and 785 as '1' for cyberbullying. The testing data size is 30% of the total entries, with an equal distribution of '0' and '1' labels for testing. The test results indicate that among the different word embedding techniques, TF-IDF has the highest accuracy. However, according to the line graph, if LSTM units are increased, TF-IDF may potentially lead to overfitting, while Word2Vec exhibits optimal performance. FastText, on the other hand, is unable to learn effectively in this case.*

**Keyword:** Cyberbullying, LSTM, TF-IDF, Word2vec, FastText, text classification, text representation.