

***TEXT SUMMARIZE PADA DATASET BERITA ONLINE***  
**BERBAHASA INDONESIA DENGAN PENDEKATAN**  
**EKSTRAKTIF MENGGUNAKAN BLSTM**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**DANANG WAHYU SURYONO**

**18.11.2386**

Kepada

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

**2023**

***TEXT SUMMARIZE PADA DATASET BERITA ONLINE***  
**BERBAHASA INDONESIA DENGAN PENDEKATAN**  
**EKSTRAKTIF MENGGUNAKAN BLSTM**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**DANANG WAHYU SURYONO**

**18.11. 2386**

Kepada

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

**2023**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

***TEXT SUMMARIZE PADA DATASET BERITA ONLINE BERBAHASA  
INDONESIA DENGAN PENDEKATAN EKSTRAKTIF  
MENGUNAKAN BLSTM***

yang disusun dan diajukan oleh

**Danang Wahyu Suryono**

**18.11.2386**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal Februari 2023

Dosen Pembimbing,



**Mardhiya Hayaty, S. T., M. Kom.**  
**NIK. 190302108**

**HALAMAN PENGESAHAN**

**SKRIPSI**

**TEXT SUMMARIZE PADA DATASET BERITA ONLINE BERBAHASA  
INDONESIA DENGAN PENDEKATAN EKSTRAKTIF  
MENGUNAKAN BLSTM**

yang disusun dan diajukan oleh

**Danang Wahyu Suryono**

**18.11.2386**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 21 Februari 2023

**Susunan Dewan Penguji**

**Nama Penguji**

**Anna Baita, M.Kom**  
**NIK. 190302290**

**Arifiyanto Hadinegoro S.Kom, MT**  
**NIK. 190302289**

**Mardhiya Hayaty, S.T., M.Kom**  
**NIK. 190302108**

**Tanda Tangan**



Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 21 Februari 2023

**DEKAN FAKULTAS ILMU KOMPUTER**



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

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

### HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Danang Wahyu Suryono  
NIM : 18.11.2386

Menyatakan bahwa Skripsi dengan judul berikut:

**TEXT SUMMARIZE PADA DATASET BERITA ONLINE BERBAHASA  
INDONESIA DENGAN PENDEKATAN EKSTRAKTIF  
MENGUNAKAN BLSTM**

Dosen Pembimbing : Mardhiya Hayaty, S.T., 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, 24 Agustus 2023

Yang Menyatakan,



METRAI  
TEMPEL  
BALAI 0226254837

Danang Wahyu Suryono

## KATA PENGANTAR

Puji dan syukur penulis ucapkan kehadirat Tuhan Yang Maha Esa karena telah memberikan Rahmat dan Karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “***Text Summarize Pada Dataset Berita Online Berbahasa Indonesia Dengan Pendekatan Ekstraktif Menggunakan BLSTM***”. Penyusunan skripsi ini dilakukan sebagai salah satu syarat kelulusan yang harus dipenuhi untuk dapat mencapai Gelar Sarjana pada Fakultas Ilmu Komputer Program Studi Informatika Universitas AMIKOM Yogyakarta.

Dalam penulisan skripsi ini, penulis mengucapkan rasa terimakasih atas motivasi, bimbingan, dukungan, saran dan masukan dari berbagai pihak secara moral maupun spiritual. Pada kesempatan kali ini penulis mengucapkan terimakasih kepada :

1. Bapak Prof. Dr. M. Suyanto, M.M, selaku Rektor Universitas AMIKOM Yogyakarta,
2. Bapak Hanif Al Fatta, S. Kom., M. Kom, selaku Dekan Fakultas Ilmu Universitas AMIKOM Yogyakarta,
3. Ibu Mardhiya Hayaty, S.T., M.Kom, selaku dosen pembimbing yang telah memberikan dedikasi penuh dalam membantu penulis untuk penyelesaian skripsi ini,
4. Bapak Suparno, Ibu Sukiyem, kedua adik saya serta Alm. Nenek, selaku orang tua dan anggota keluarga saya yang selalu memberikan motivasi, restu, dukungan dan doa untuk penulis,
5. Teman-teman pengurus Himpunan Mahasiswa Informatika (HMIF) terutama Ronald, Ayub, Anggi dan Silvi yang sudah banyak membantu baik ketika dikepengurusan maupun ketika proses penulisan skripsi ini,
6. Gunawan Prasetyo yang telah berbaik hati meminjamkan laptop dalam proses pengerjaan skripsi ini,
7. Sai, Tomo, Yazid, Rizka dan Dilla selaku teman MSIB Batch 2 yang memotivasi saya untuk segera menyelesaikan skripsi ini,

8. Teman-teman seperjuangan dari kelas 18-IF-09 yang telah menemani saya dari semester awal hingga akhir, terutama yang sering saya repotkan dalam mengerjakan tugas kuliah,
9. Sahabat, teman, dan kerabat penulis yang tidak bisa disebutkan satu per satu, dan
10. Diri Sendiri yang pantang menyerah hingga skripsi ini berhasil di selesaikan.

Penulis menyadari bahwa naskah skripsi ini masih jauh dari kata sempurna dan masih terdapat banyak kekurangan, untuk itu penulis mengharapkan kritik dan saran dari pembaca untuk melakukan perbaikan. Semoga skripsi ini dapat bermanfaat bagi pembaca.

Yogyakarta, 21 Februari 2023

Penulis

## DAFTAR ISI

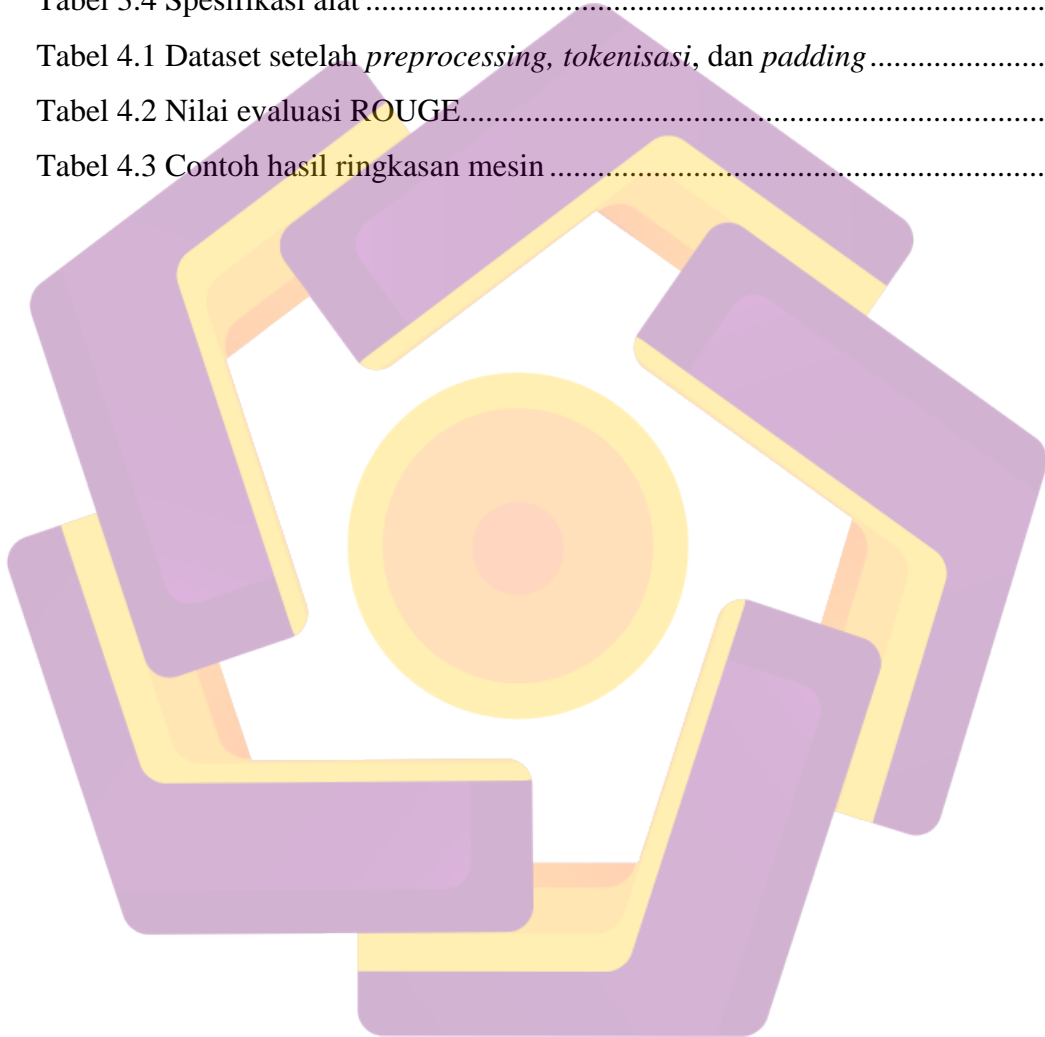
HALAMAN JUDUL .....	1
HALAMAN JUDUL .....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN .....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI .....	iv
KATA PENGANTAR .....	v
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
DAFTAR ISTILAH.....	xi
INTISARI .....	xii
ABSTRACT.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
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
2.2 Dasar Teori .....	9
2.2.1. ROUGE.....	9
2.2.2. <i>Bidirectional LSTM</i> .....	10
BAB III METODE PENELITIAN .....	12
3.1 <i>Dataset</i> .....	12
3.2 Alur Penelitian .....	13
3.2.1. <i>Preprocessing</i> .....	13



3.2.2.	<i>Splitting Data</i> .....	14
3.2.3.	Model <i>Bidirectional LSTM</i> .....	14
3.2.2.	<i>Extractive Summarize</i> .....	15
3.2.3.	<i>Evaluation</i> .....	15
3.3	Eksperimen .....	15
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....		19
4.1.	Implementasi .....	19
4.1.1.	Pemanggilan <i>Library</i> .....	19
4.1.2.	Pemanggilan <i>Dataset</i> .....	19
4.1.3.	<i>Preprocessing Data</i> .....	20
4.1.4.	Pemisahan Data .....	21
4.1.5.	Pembuatan Model .....	21
4.1.6.	Pelatihan Model .....	23
4.1.7.	Proses <i>Extractive Summarize</i> .....	23
4.2.	Hasil dan Diskusi .....	24
<b>BAB V PENUTUP</b> .....		27
5.1	Kesimpulan .....	27
5.2	Saran .....	27
<b>DAFTAR PUSTAKA</b> .....		28

## DAFTAR TABEL

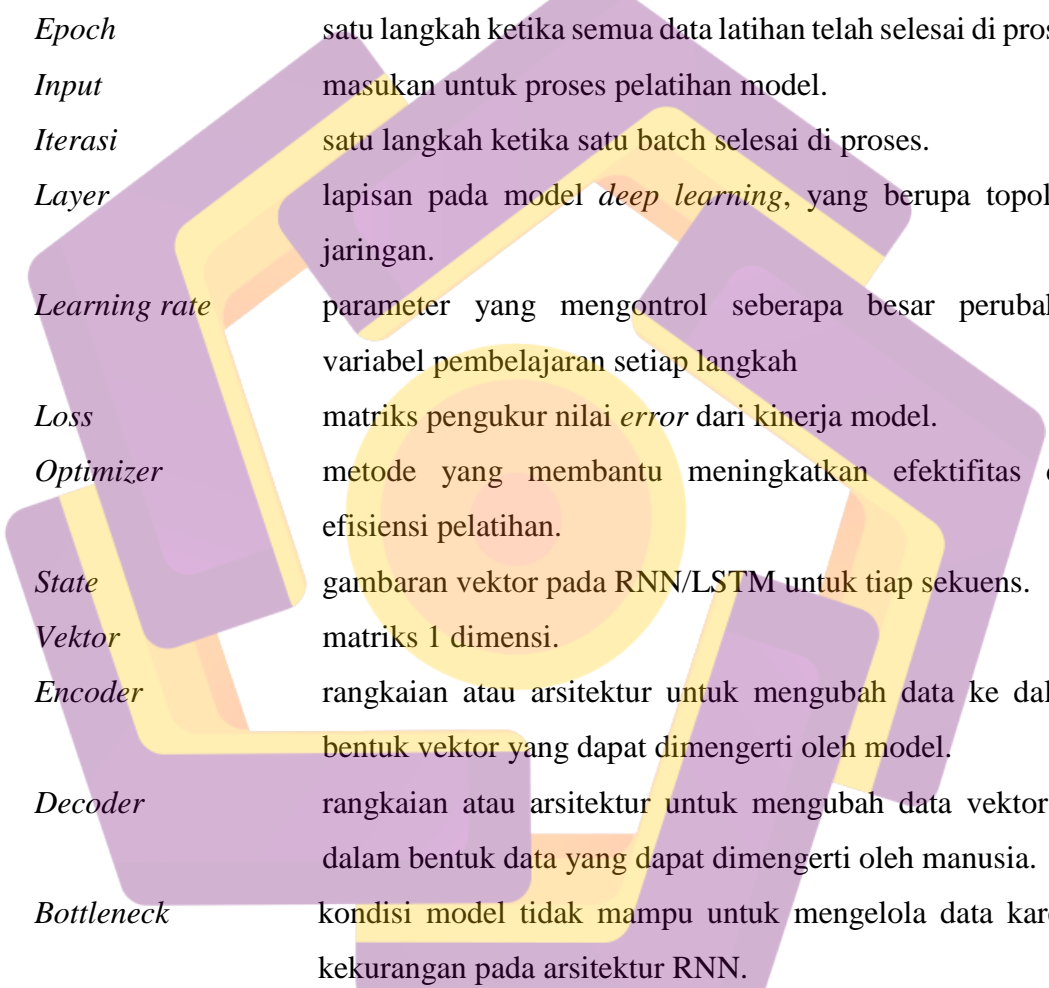
Tabel 2.1 Keaslian Penelitian .....	7
Tabel 3.1 Detail jumlah kata artikel dan ringkasan .....	12
Tabel 3.2 Contoh artikel dan ringkasan asli.....	12
Tabel 3.3 Spesifikasi data yang di- <i>training</i> .....	16
Tabel 3.4 Spesifikasi alat .....	16
Tabel 4.1 Dataset setelah <i>preprocessing</i> , <i>tokenisasi</i> , dan <i>padding</i> .....	20
Tabel 4.2 Nilai evaluasi ROUGE.....	25
Tabel 4.3 Contoh hasil ringkasan mesin.....	25



## DAFTAR GAMBAR

Gambar 2.1 Arsitektur sel LSTM .....	10
Gambar 2.2 Arsitektur <i>Bidirectional LSTM</i> .....	11
Gambar 3.1 Alur Penelitian .....	13
Gambar 3.2 Model Encoder BLSTM, Decoder LSTM .....	14
Gambar 4.1 Pemanggilan <i>library</i> .....	19
Gambar 4.2 Pemanggilan <i>dataset</i> .....	19
Gambar 4.3 <i>Preprocessing dataset</i> .....	20
Gambar 4.4 Pemisahan <i>dataset</i> .....	21
Gambar 4.5 Pembuatan <i>encoder layer</i> .....	21
Gambar 4.6 Pembuatan <i>decoder layer</i> .....	22
Gambar 4.7 Model <i>summary</i> .....	22
Gambar 4.8 Fungsi <i>compile model</i> .....	23
Gambar 4.9 Pelatihan model.....	226
Gambar 4.10 Proses <i>extractive summarize</i> .....	227
Gambar 4.11 Grafik <i>Training Loss</i> .....	237

## DAFTAR ISTILAH



<i>Batch</i>	jumlah data yang diolah sebelum dihitung nilai loss.
<i>Deep learning</i>	salah satu bidang di ranah kecerdasan buatan yang mencakup pengembangan program jaringan saraf tiruan dengan parameter pembelajaran yang besar.
<i>Epoch</i>	satu langkah ketika semua data latihan telah selesai di proses.
<i>Input</i>	masukan untuk proses pelatihan model.
<i>Iterasi</i>	satu langkah ketika satu batch selesai di proses.
<i>Layer</i>	lapisan pada model <i>deep learning</i> , yang berupa topologi jaringan.
<i>Learning rate</i>	parameter yang mengontrol seberapa besar perubahan variabel pembelajaran setiap langkah
<i>Loss</i>	matriks pengukur nilai <i>error</i> dari kinerja model.
<i>Optimizer</i>	metode yang membantu meningkatkan efektifitas dan efisiensi pelatihan.
<i>State</i>	gambaran vektor pada RNN/LSTM untuk tiap sekuens.
<i>Vektor</i>	matriks 1 dimensi.
<i>Encoder</i>	rangkaian atau arsitektur untuk mengubah data ke dalam bentuk vektor yang dapat dimengerti oleh model.
<i>Decoder</i>	rangkaian atau arsitektur untuk mengubah data vektor ke dalam bentuk data yang dapat dimengerti oleh manusia.
<i>Bottleneck</i>	kondisi model tidak mampu untuk mengelola data karena kekurangan pada arsitektur RNN.
Ngram	potongan n-karakter yang diambil dari suatu string.

## INTISARI

Arus informasi pada saat ini bergerak sangat cepat melalui media daring, sehingga artikel berita dalam bentuk digital sudah sangat wajar dan sering kita temui melalui berbagai media sosial. Banyaknya artikel berita yang beredar di media sosial menyebabkan kita perlu untuk menghabiskan lebih banyak waktu untuk mengikuti perkembangan informasi tersebut. Hal tersebut mendorong NLP untuk melakukan peringkasan pada teks artikel dengan tujuan menghasilkan teks yang lebih singkat tanpa menghilangkan informasi utama pada artikel berita tersebut. Penelitian ini mengusulkan menggunakan *Bidirectional Long-Short Term Memory* untuk melakukan peringkasan teks secara ekstraktif terhadap data berbahasa Indonesia. *Bidirectional LSTM* dipilih karena memiliki proses pembelajaran yang lebih baik dari LSTM karena memiliki dua lapisan paralel untuk memperoleh informasi dari data. Hasil penelitian ini menunjukkan bahwa *Bidirectional LSTM* belum dapat menghasilkan ringkasan teks yang baik untuk data berbahasa Indonesia dengan menunjukkan hasil evaluasi ROGUE-1 23,9% dan ROGUE-L sebesar 19,3%.

**Kata kunci:** BLSTM, *Extractive Text Summarize*, Liputan6, Bahasa Indonesia.



## ABSTRACT

Nowadays, the flow of information moves very quickly through online media, so news articles in digital form are widespread and often encountered through various social media. The number of news articles circulating on social media causes us to spend more time following the development of the information. It encourages NLP to summarize the article to produce a shorter text without losing the main idea of the news article. This study proposes using Bidirectional Long-Short Term Memory (BiLSTM) in extractive summarizing the text into Indonesian news articles from Liputan6 media. BiLSTM is an LSTM structure that has been developed to cover the shortcomings of LSTM. Long-Short Term Memory (LSTM) is an RNN structure with an Attention mechanism and the ability to store memory which was developed to cover the shortcomings of RNNs that cannot store memory. BiLSTM is chosen because it has a better learning process than LSTM. The reason is that BiLSTM has two parallel layers to obtain information from the data. The results of this study indicate that BiLSTM has optimal loss and val\_loss values by reaching values of 0.0028 and 0.064. However, the model developed in this study has not been able to produce a good summary of the text for Indonesian data by showing the evaluation results of ROGUE-1 23,9 and ROGUE-L by 19.3%.

**Keyword:** BLSTM, Ekstraktif Text Summarize, Liputan6, Indonesian