

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

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

RONALD FERDINAND

18.11.1904

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2022

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

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

RONALD FERDINAND

18.11. 1904

Kepada

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA

2022

HALAMAN PERSETUJUAN

SKRIPSI

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

yang disusun dan diajukan oleh

Ronald Ferdinand

18.11.1904

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 27 Agustus 2022

Dosen Pembimbing,



Mardhiya Hayaty, S. T., M. Kom.

NIK. 190302108

HALAMAN PENGESAHAN

SKRIPSI

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

yang disusun dan diajukan oleh

Ronald Ferdinand

18.11.1904

Telah dipertahankan di depan Dewan Penguji
pada tanggal 24 Agustus 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Windha Mega Pradnya D, M.Kom

NIK. 190302185

Anna Balta, M.Kom

NIK. 190302290

Mardhlya Hayaty, S.T., M.Kom

NIK. 190302108

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 24 Agustus 2022

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, S.Kom., M.Kom.

NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Ronald Ferdinand

NIM : 18.11.1904

Menyatakan bahwa Skripsi dengan judul berikut:

Text Summarize Pada Dataset Berita Online Berbahasa Indonesia Dengan Pendekatan Abstraktif Menggunakan 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 2022

Yang Menyatakan,



Ronald Ferdinand

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 Abstraktif 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.

Penulis menyadari bahwa penulisan skripsi ini masih belum sempurna karena keterbatasan kemampuan dan pengetahuan yang penulis dapatkan, oleh karena itu penulis mohon maaf atas segala kekurangan.

Dalam penyusunan skripsi ini, penulis juga sadar banyak mendapat dukungan, bimbingan bantuan dan kemudahan dari berbagai pihak sehingga skripsi ini dapat diselesaikan. Oleh karena itu, penulis mengucapkan terima kasih sebesar-besarnya kepada semua pihak yang telah membantu dan mendorong terwujudnya skripsi ini.

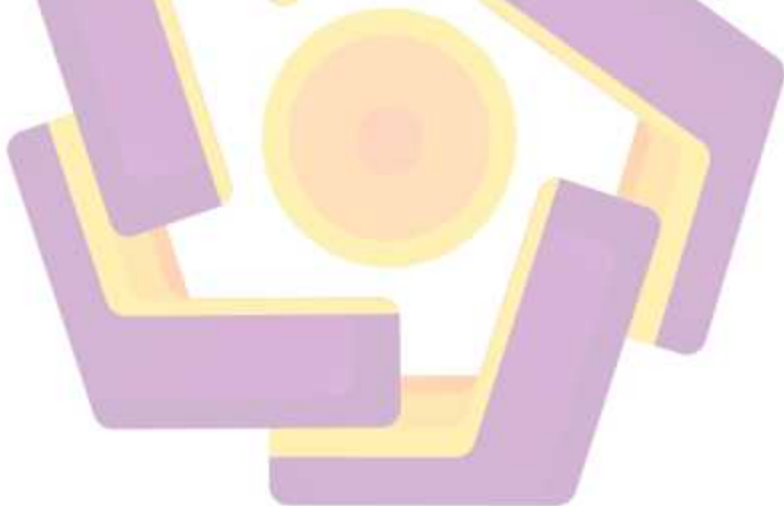
Dengan ketulusan hati, penulis menyampaikan terima kasih kepada :

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. Yupiansyah, Sarinda Lina, Laura Ladiesga dan Delanova Shakyta, selaku kedua orang tua dan adik-adik penulis yang telah memberikan kasih sayang, motivasi, dan doa untuk penulis.
5. Bapak dan Ibu dosen serta staff Universitas AMIKOM Yogyakarta yang telah memberikan ilmu selama saya berkuliah.

6. Aprilia Nurkasanah dan Hiero Panda, yang telah menemani, mendampingi, memberikan motivasi, serta selalu ada dan bisa bagi penulis.
7. Teman teman kelas 18-IF-02, yaitu Rizki, Galuh, Ikhwan, Koko, Andryansa, dan lainnya yang penulis tidak dapat sebutkan satu persatu.
8. Teman teman pengurus Himpunan Mahasiswa Informatika (HMIF) yang sudah saya anggap sebagai saudara, terutama kepengurusan periode 2020/2021 yaitu Alfian, Ayub, Danang, Anggi, Kurnia, dan lainnya yang penulis tidak dapat sebutkan satu persatu.
9. Dan untuk semuanya yang tidak bisa disebutkan oleh penulis satu per satu.

Yogyakarta, 24 Agustus 2022

Penulis



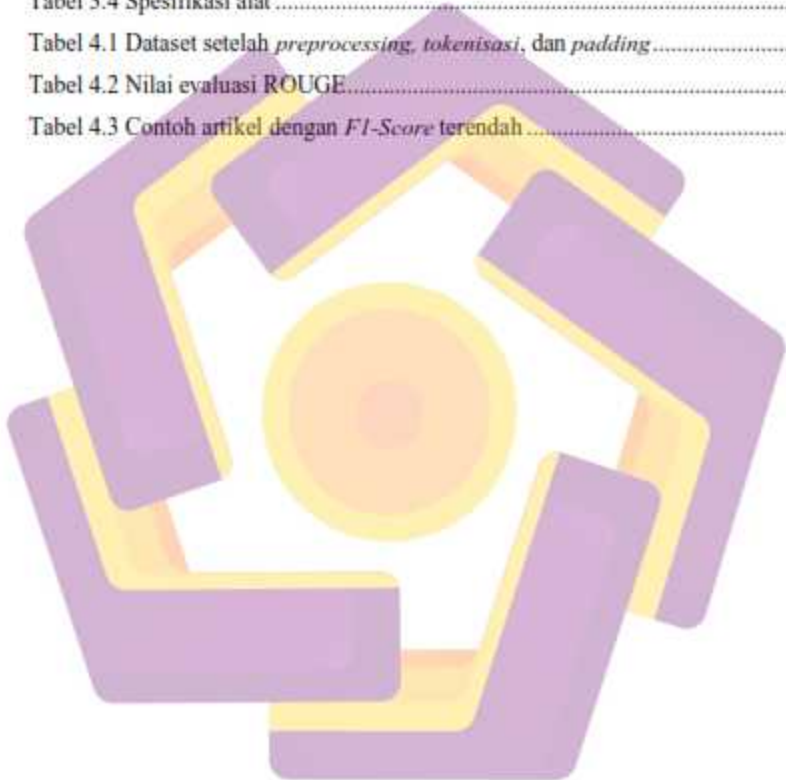
DAFTAR ISI

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	4
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Dasar Teori	9
2.2.1. <i>Bidirectional LSTM</i>	9
2.2.2. Mekanisme <i>Attention</i>	10
2.2.3. ROUGE.....	11
BAB III METODE PENELITIAN	13
3.1 <i>Dataset</i>	13
3.2 Alur Penelitian	14
3.2.1. <i>Preprocessing</i>	14
3.2.2. <i>Splitting Data</i>	15

3.2.3. Model <i>Bidirectional LSTM</i>	15
3.2.4. <i>Inference</i>	16
3.2.5. <i>Evaluation</i>	16
3.3 Eksperimen	16
BAB IV HASIL DAN PEMBAHASAN	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>Inference</i>	23
4.2. Hasil dan Diskusi	24
BAB V PENUTUP	27
5.1 Kesimpulan	27
5.2 Saran	27
REFERENSI	28

DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	7
Tabel 3.1 Detail jumlah kata artikel dan ringkasan	13
Tabel 3.2 Contoh artikel dan ringkasan asli.....	13
Tabel 3.3 Spesifikasi data yang di- <i>training</i>	17
Tabel 3.4 Spesifikasi alat	17
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 artikel dengan <i>F1-Score</i> terendah	26



DAFTAR GAMBAR

Gambar 2.1	Arsitektur sel LSTM	9
Gambar 2.2	Arsitektur <i>Bidirectional</i> LSTM.....	10
Gambar 2.3	Arsitektur mekanisme <i>attention</i>	11
Gambar 3.1	Alur Penelitian	14
Gambar 3.2	Model BLSTM.....	15
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	<i>Model summary</i>	22
Gambar 4.8	Fungsi <i>compile</i> model.....	23
Gambar 4.9	Proses pelatihan model.....	23
Gambar 4.10	Proses <i>inference</i> data.....	24
Gambar 4.11	Grafik <i>Training Loss</i>	24

DAFTAR ISTILAH

<i>Batch</i>	jumlah data yang diolah sebelum dihitung nilai loss.
<i>Deep learning</i>	salah satu bidang di ranah kecerdasaran 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.
<i>Ngram</i>	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 penggunaan *Bidirectional Long-Short Term Memory* untuk melakukan peringkasan teks secara abstraktif 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 dan ROGUE-L sebesar 3,63%.

Kata kunci: BLSTM, *Abstractive Text Summarize*, Mekanisme Attention, 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 abstractly 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 and ROGUE-L by 3.63%.

Keyword: *BLSTM, Abstractive Text Summarize, Attention Mechanism, Liputan6, Indonesian.*