

**RINGKASAN TEKS OTOMATIS PADA DOKUMEN TUNGGAL  
BAHASA INDONESIA DENGAN METODE ABSTRAK  
SEQUENCE TO SEQUENCE**

**SKRIPSI**

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



disusun oleh

**ABI MAS HARDIANTO**

**19.11.3153**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2023**

**RINGKASAN TEKS OTOMATIS PADA DOKUMEN TUNGGAL  
BAHASA INDONESIA DENGAN METODE ABSTRAK  
SEQUENCE TO SEQUENCE**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**ABI MAS HARDIANTO**

**19.11.3153**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2023**

## HALAMAN PERSETUJUAN

HALAMAN PERSETUJUAN

SKRIPSI

RINGKASAN TEKS OTOMATIS PADA DOKUMEN TUNGGAL  
BAHASA INDONESIA DENGAN METODE ABSTRAK  
SEQUENCE TO SEQUENCE

yang disusun dan diajukan oleh

**Abi Mas Hardianto**

**19.11.3153**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 18 September 2023

Dosen Pembimbing,



**Arif Akbarul Huda, S.Si, M.Eng**  
**NIK. 190302287**

## HALAMAN PENGESAHAN

### HALAMAN PENGESAHAN

#### SKRIPSI

#### RINGKASAN TEKS OTOMATIS PADA DOKUMEN TUNGGAL BAHASA INDONESIA DENGAN METODE ABSTRAK SEQUENCE TO SEQUENCE

yang disusun dan diajukan oleh

**Abi Mas Hardianto**

**19.11.3153**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 18 September 2023

#### Susunan Dewan Penguji

**Nama Penguji**

**Tanda Tangan**

Arif Akbarul Huda, S.Si, M.Eng  
NIK. 190302287

Anggit Dwi Hartanto, M.Kom  
NIK. 190302163

Hendra Kurniawan, M.Kom  
NIK. 190302244



Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 18 September 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 : Abi Mas Hardianto  
NIM : 19.11.3153

Menyatakan bahwa Skripsi dengan judul berikut:

**Ringkasan Teks Otomatis pada Dokumen Tunggal Bahasa Indonesia dengan Metode Abstrak Sequence to Sequence**

Dosen Pembimbing : Arif Akbarul Huda, S.Si, M.Eng

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, 18 September 2023

Yang Menyatakan,

  
Abi Mas Hardianto

## HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur atas kesempatan yang telah Allah SWT berikan dan dengan telah diselesaikannya skripsi ini. Saya mempersembahkan skripsi ini kepada :

1. Ayah dan Ibu saya tercinta yang telah memberi dukungan moral, mendoakan, dan selalu memperhatikan kesehatan saya di setiap kondisi,
2. Bapak Arif Akbarul Huda, S.Si., M.Eng., selaku dosen pembimbing yang selalu membimbing dan mengarahkan jalannya proses skripsi ini,
3. Teman sekelas saya, anggota dari kelas IF 09, senantiasa hadir untuk menemani saya,
4. Serta semua pihak yang membantu dan mendukung saya yang tidak bisa saya sebutkan satu persatu.

## KATA PENGANTAR

Puji syukur kami panjatkan ke hadirat Allah SWT, karena atas rahmat dan karunia-Nya, penulis dapat menyelesaikan penulisan skripsi ini dengan judul "Ringkasan Teks Otomatis pada Dokumen Tunggal Bahasa Indonesia dengan Metode Abstrak Sequence to Sequence". Penelitian ini merupakan suatu langkah kecil dalam perjalanan akademis penulis yang penuh perjuangan, dukungan, dan bimbingan.

Penulis ingin mengucapkan terima kasih yang tak terhingga kepada Bapak Dosen Pembimbing, Arif Akbarul Huda, yang telah memberikan bimbingan, arahan, dan motivasi sehingga penelitian ini dapat terselesaikan dengan baik. Bapak tidak hanya menjadi pembimbing akademis, tetapi juga menjadi teladan dalam dedikasi dan ketekunan dalam dunia ilmu pengetahuan.

Selanjutnya, terima kasih kepada keluarga tercinta yang selalu memberikan doa dan dukungan moril sehingga penulis dapat menjalani proses perkuliahan dengan lancar. Terima kasih juga kepada teman-teman seperjuangan yang telah menjadi sumber inspirasi dan motivasi dalam meniti perjalanan ini.

Penulis menyadari bahwa penelitian ini jauh dari kesempurnaan, oleh karena itu, segala saran dan kritik membangun sangat penulis harapkan untuk perbaikan di masa yang akan datang. Semoga penelitian ini dapat memberikan manfaat dan kontribusi positif dalam bidang ilmu yang dipilih.

Akhir kata, penulis berharap bahwa skripsi ini dapat menjadi bahan referensi dan inspirasi bagi pembaca yang membutuhkan. Semoga hasil penelitian ini dapat membuka wawasan dan memberikan kontribusi kecil bagi perkembangan ilmu pengetahuan.

Demikianlah kata pengantar ini penulis sampaikan, dengan kerendahan hati dan harapan besar bahwa penelitian ini dapat memberikan nilai tambah bagi dunia ilmiah.

Yogyakarta, 1 September 2023

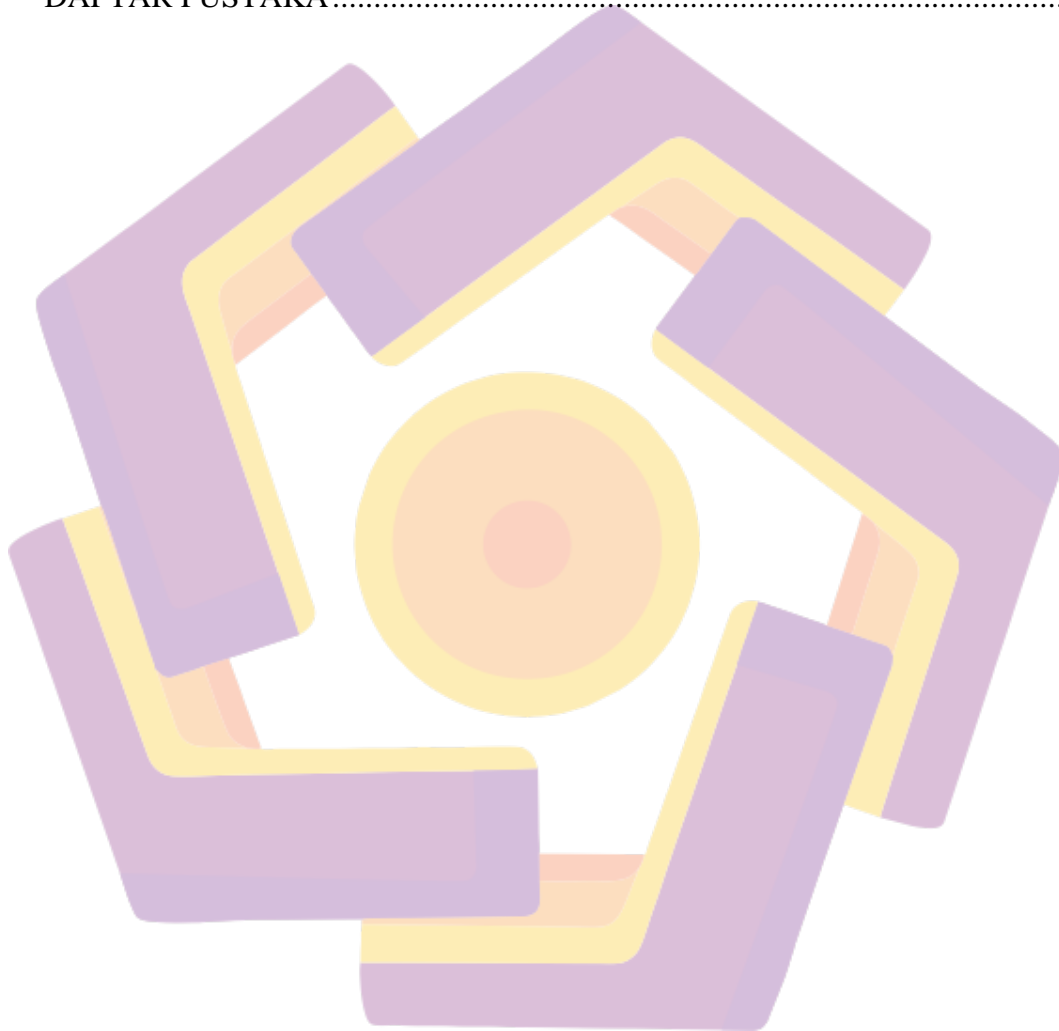
Penulis

## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN PERSETUJUAN .....	ii
HALAMAN PENGESAHAN .....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI .....	iv
HALAMAN PERSEMBAHAN .....	v
KATA PENGANTAR .....	vi
DAFTAR ISI .....	vii
DAFTAR TABEL .....	ix
DAFTAR GAMBAR .....	x
INTISARI .....	xi
ABSTRACT .....	xii
BAB I PENDAHULUAN .....	13
1.1 Latar Belakang .....	13
1.2 Rumusan Masalah .....	14
1.3 Batasan Masalah .....	15
1.4 Tujuan Penelitian .....	15
1.5 Manfaat Penelitian .....	15
1.6 Sistematika Penulisan .....	16
BAB II TINJAUAN PUSTAKA .....	17
2.1 Studi Literatur .....	17
2.2 Dasar Teori .....	21
BAB III METODE PENELITIAN .....	31
3.1 Alat dan Bahan Penelitian .....	31
3.2 Alur Penelitian .....	32
3.3 Metode Pengolahan Data .....	34
BAB IV HASIL DAN PEMBAHASAN .....	42
4.1 Pengumpulan Dataset Artikel Berita .....	42
4.2 Text Preprocessing .....	42
4.3 Pembuatan Model dan Model Trainer .....	49

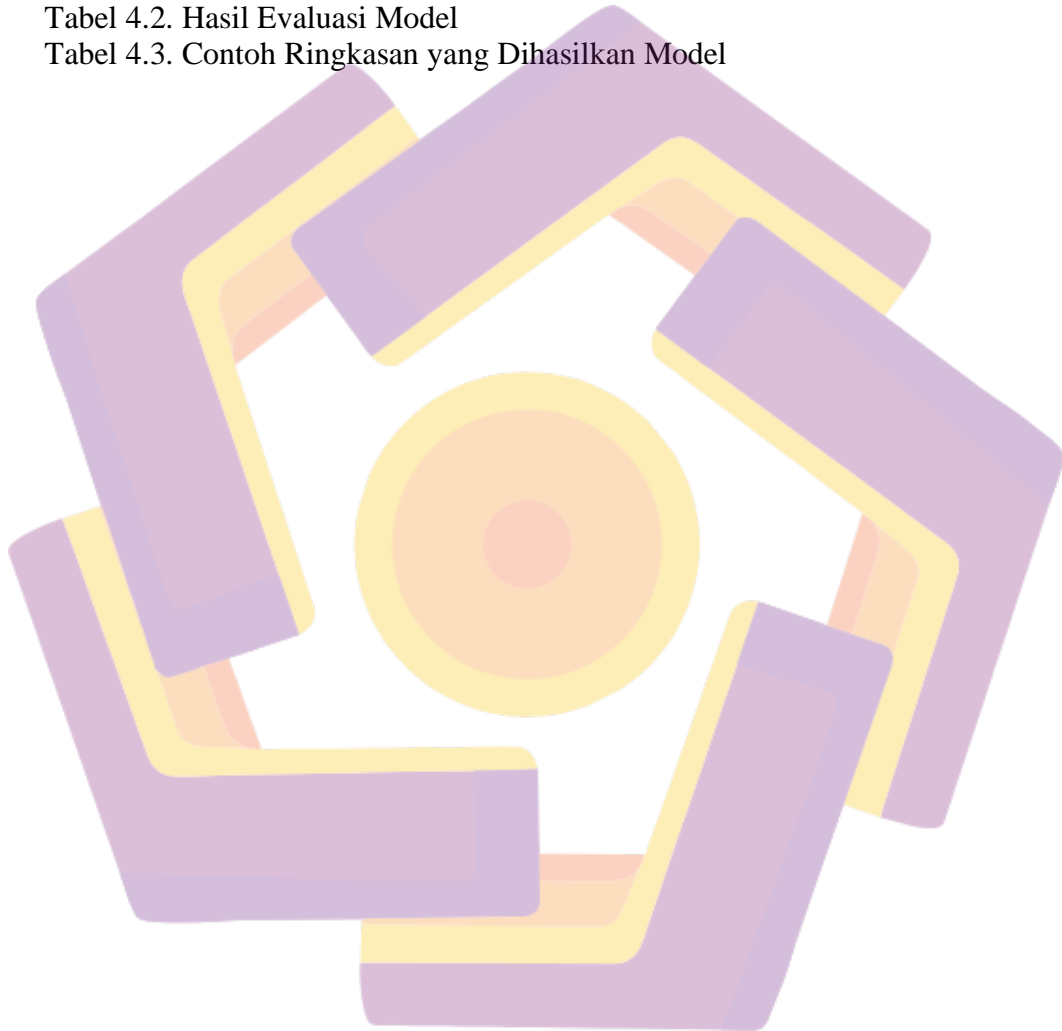


4.4 Model Training .....	50
4.5 Evaluasi dan <i>Testing</i> .....	52
BAB V PENUTUP .....	61
5.1 Kesimpulan.....	61
5.2 Saran.....	61
DAFTAR PUSTAKA.....	63



## DAFTAR TABEL

Tabel 2.1. Keaslian Penelitian	18
Tabel 2.2. Contoh N-gram	29
Tabel 3.1. Alat Penelitian	31
Tabel 3.2. Prefiks Untuk Berbagai Tugas NLP	37
Tabel 4.1. Contoh Data Hasil Tokenization	46
Tabel 4.2. Hasil Evaluasi Model	53
Tabel 4.3. Contoh Ringkasan yang Dihasilkan Model	53



## DAFTAR GAMBAR

Gambar 2.1. Arsitektur Ringkasan Teks Otomatis	22
Gambar 2.2. Arsitektur Sequence to Sequence	24
Gambar 2.3. Arsitektur Transformer	26
Gambar 3.1. Data Mentah Dataset Indosum	32
Gambar 3.2. Alur Penelitian	32
Gambar 3.3. Alur Flattening Data	34
Gambar 3.4. Alur Case Folding	35
Gambar 3.5. Alur Filtering	36
Gambar 3.6. Alur Proses Menambahkan Prefiks	37
Gambar 3.7. Alur Tokenization	38
Gambar 3.8. Alur Model Training	39
Gambar 3.9. Alur Evaluasi Testing Data	41
Gambar 4.1. Kode Program Akses Database MongoDB	42
Gambar 4.2. Alur Text Preprocessing	42
Gambar 4.3. Kode Program Flatten Function	43
Gambar 4.4. Data Mentah Sebelum Dilakukan Flattening Data	43
Gambar 4.5. Data Setelah Dilakukan Flattening Data	43
Gambar 4.6. Hasil Case Folding	44
Gambar 4.7. Daftar Stopwords	44
Gambar 4.8. Hasil Filtering	44
Gambar 4.9. Hasil Penambahan Prefiks	45
Gambar 4.10. Kode Program Tokenization	45
Gambar 4.11. Bentuk Dataset Setelah Proses Tokenization	46
Gambar 4.12. Kode Program Pembuatan Model	49
Gambar 4.13. Kode Program Pembuatan Model Trainer	50
Gambar 4.14. Kode Program Variabel Args	50
Gambar 4.15. Model Training	51
Gambar 4.16. Grafik Training Loss dan Validation Loss	52

## INTISARI

Dengan munculnya internet mengakibatkan jumlah dokumen yang tersedia dalam bentuk daring meningkat drastis. Dampaknya untuk mendapatkan informasi yang relevan bagi pengguna internet menjadi semakin sulit karena jumlah informasi yang ada sangat masif. Meringkas teks menjadi salah satu cara yang efektif untuk mendapatkan informasi penting dalam dokumen. Namun terdapat keterbatasan waktu dan tenaga untuk meringkas seluruh teks secara manual.

Ringkasan teks otomatis dapat menjadi solusi. Tujuan ringkasan teks otomatis adalah menghasilkan teks ringkas yang tetap mempertahankan informasi utama. Salah satu metode yang dapat diterapkan untuk mendapat ringkasan teks abstraktif adalah dengan *sequence-to-sequence* (Seq2Seq). Seq2Seq berbasis *fine-tuned model* transformer menunjukkan hasil yang lebih baik dibandingkan Seq2Seq berbasis CNN atau RNN. *Text to Text Transfer Transformer* (T5) merupakan salah satu contoh *fine-tuned model* transformer.

Uji coba Seq2Seq berbasis T5 transformer untuk tugas ringkasan teks, dengan 1.877 dokumen tunggal berita bahasa Indonesia, menghasilkan nilai ROUGE-1 0,742, ROUGE-2 0,677, ROUGE-L 0,713, dan ROUGE-Lsum 0,734.

**Kata kunci:** ringkasan teks otomatis, ringkasan teks abstraktif, Seq2Seq, Transformer, T5

## ABSTRACT

*Internet makes the number of available documents in digital form has increased. The impact is obtaining relevant information for internet users has become increasingly difficult due to the massive amount of information available. Summarizing text is one effective way to extract important information from documents. However, there are time and effort limitations for manually summarizing entire texts.*

*Automatic text summarization can provide a solution. The goal of automatic text summarization is to generate concise text while retaining the main information. One method that can be applied for abstractive text summarization is the sequence-to-sequence (Seq2Seq) approach. Fine-tuned transformer-based Seq2Seq models, such as the Text to Text Transfer Transformer (T5), outperform Seq2Seq models based on CNN or RNN.*

*Experimenting with a T5 transformer-based Seq2Seq for text summarization tasks using 1,877 individual Indonesian news documents yielded ROUGE-1 score of 0.742, ROUGE-2 score of 0.677, ROUGE-L score of 0.713, and ROUGE-Lsum score of 0.734.*

**Keyword:** *Automatic text summarization, abstractive text summarization, Seq2Seq, transformer, T5*