

**PENDEKATAN EKSTRAKTIF PADA SUMMARIZE DOKUMEN  
MENGUNAKAN GRU**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika



disusun oleh

**WARAWANDANU**

**18.11.2532**

Kepada

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS AMIKOM YOGYAKARTA**

**YOGYAKARTA**

**2023**

**PENDEKATAN EKSTRAKTIF PADA SUMMARIZE DOKUMEN**

**MENGGUNAKAN GRU**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika



disusun oleh

**WARAWANDANU**

**18.11.2532**

Kepada

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS AMIKOM YOGYAKARTA**

**YOGYAKARTA**

**2023**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**PENDEKATAN EKSTRAKTIF PADA SUMMARIZE DOKUMEN**

**MENGGUNAKAN GRU**

yang disusun dan diajukan oleh

**Warawandanu**

**18.11.2532**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 21 Februari 2023

**Dosen Pembimbing,**



**Mardhiya Hayati M.Kom**

**NIK. 190302108**

**HALAMAN PENGESAHAN**

**SKRIPSI**

**PENDEKATAN EKSTRAKTIF PADA SUMMARIZE DOKUMEN**

**MENGGUNAKAN GRU**

yang disusun dan diajukan oleh

**Warawandanu**

**18.11.2532**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 21 Februari 2023

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

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

**Yoga Pristyanto S.Kom, M.Eng**  
**NIK. 190302412**

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



Two handwritten signatures in blue and black ink are present. The first signature is in blue ink and the second is in black ink. Both are written over horizontal lines.



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

Yang bertandatangan di bawah ini,

Nama mahasiswa : Warawandanu  
NIM : 18.11.2532

Menyatakan bahwa Skripsi dengan judul berikut:

**Pendekatan Ekstraktif Pada Summarize Dokumen Menggunakan GRU**

Dosen Pembimbing : Mardhiya Hayati, 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, 21 Februari 2023

Yang Menyatakan,



METERAI  
TEMPEL  
1000  
FBAX394391462

Warawandanu

## HALAMAN PERSEMBAHAN

Alhamdulillah segala puji syukur atas berkat rahmat dan karunia Allah SWT yang telah memberikan kemudahan dan kelancara bagi penulis dalam menyelesaikan skripsi ini dengan sebaik-baiknya. Tidak lepas dari bantuan dan dengan dukungan berbagai pihak. Dengan ini saya persembahkan skripsi ini kepada semua pihak yang turut mendukung penulis sehingga mampu menyelesaikan studi untuk meraih gelar sarjana, yaitu:

1. Bapak dan ibu saya tercinta yang tak hentinya memberikan do'a, cinta dan kasih sayangnya kepada penulis. Serta berkat pengorbanan dan perjuangan merekalah penulis mampu menyelesaikan skripsi ini.
2. Dosen pembimbing saya ibu Mardhiya Hayati, S.T., M.kom yang telah membimbing , mengarahkan dan memberikan pengetahuan yang berharga selama proses penulisan skripsi. Terima kasih atas kesabaran dan waktu yang diberikan untuk membantu saya mencapai hasil terbaik.
3. Teman SMA sekaligus sahabat penulis yang selalu kuat mendengar cerita dari penulis yaitu Bella, Fiulan, dan Viky.
4. Teman-teman Pondok Pesantren Ki Ageng Giring, Kak Aly, Mas Farhan, Mas Jaka, Kang Hisam yang selalu memberikan semangat dan pengalaman berharga bagi penulis.
5. Pengurus HMIF periode 2021/2022, yang selalu memberikan *support* dan memberikan energi *positif* kepada penulis
6. Teman Grup jalan-jalan , Reny, Kusuma, Aldin, Villa , dan Iman terima kasih sudah kebersamaan selama masa-masa skripsi dan memberikan coretan perjalanan yang sangat berkesan.
7. Teman asisten praktikum *Big Data* dan *Predictive Analytics* Lanjut Reny Fatarina yang telah memberikan energi positif dan keceriaan setiap harinya.
8. Teman kelas IF 11 yang selalu memberikan semangat dan bantuan dalam kelancaran skripsi yaitu Mas Bagus, Fajar, Wahyu.
9. Dan seluruh pihak yang tidak bisa penulis sebutkan satu-persatu.

## KATA PENGANTAR

Puji syukur atas kehadiran Allah SWT yang telah melimpahkan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “*Pendekatan Ekstraktif Pada Summarize Dokumen Menggunakan GRU*” dengan sebaik-baiknya sebagai salah-satu syarat untuk dapat menempuh ujian sarjana 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. Ibu Windha Mega Pradnya Duhita, M.Kom selaku Ketua Program Studi S1 Informatika Universitas Amikom Yogyakarta
4. Ibu Mardhiya Hayaty, S.T., M.Kom selaku dosen pembimbing yang telah memberikan pengalaman dan membimbing penulis dalam menyusun skripsi ini.

Penulis menyadari bahwa dalam penulisan skripsi ini masih banyak kekurangan sehingga saran dan kritik yang membangun sangat penulis harapkan. Penulis berharap semoga skripsi ini bias memberikan manfaat kepada para pembaca khususnya pembaca pribadi

Yogyakarta, 21 Februari 2023

Warawandanu

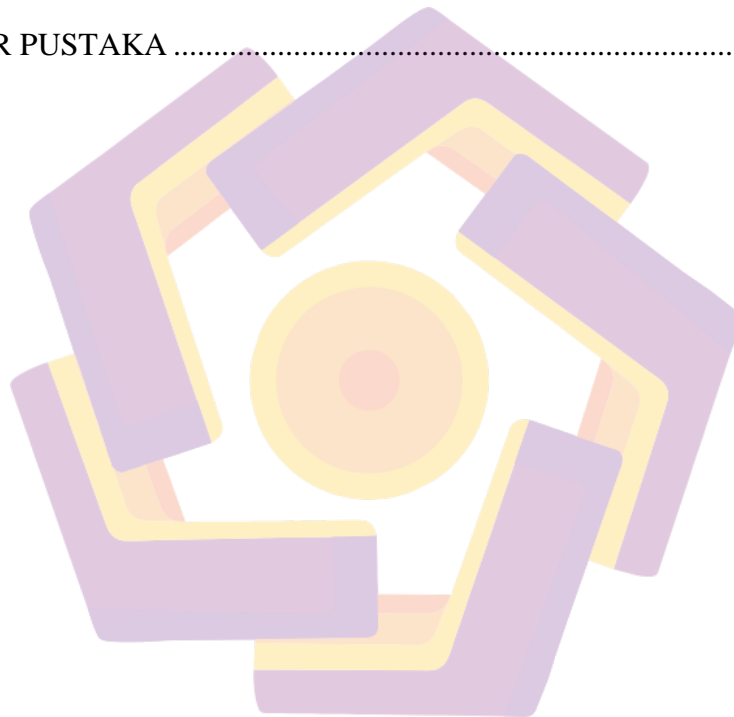
## DAFTAR ISI

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.....	x
DAFTAR GAMBAR .....	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 KAJIAN PUSTAKA.....	5
2.1 Studi Literatur .....	5
2.2 Dasar Teori .....	11
2.2.1 Peringkasan Teks Otomatis.....	11



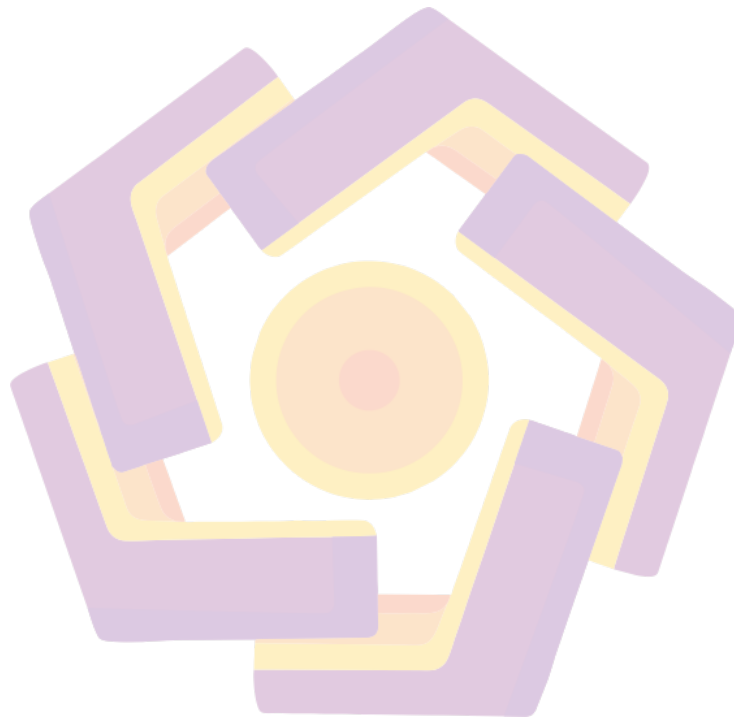
2.2.2 Artikel Berita.....	12
2.2.3 Algoritma Recurrent Neural Network ( RNN ) .....	13
2.2.4 Algoritma <i>Long Short Term Memory</i> ( LSTM ) .....	14
2.2.5 Gated Recurrent Unit ( GRU ) .....	14
2.2.6 Perbedaan RNN,LSTM dan GRU .....	16
2.2.7 ROUGE .....	17
<b>BAB III METODE PENELITIAN .....</b>	<b>21</b>
3.1 <i>Dataset</i> .....	21
3.2 Alur Penelitian .....	21
3.2.1 Preprocessing .....	22
3.2.2 <i>Splitting</i> Data.....	23
3.2.3 Model Gated Recurrent Unit ( GRU ).....	23
3.2.4 Evaluasi .....	23
3.3 Alat dan Bahan.....	23
3.3.1 Alat Penelitian.....	23
3.3.2 Bahan Penelitian .....	24
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>25</b>
4.1 Hasil.....	25
4.2 Pembahasan .....	25
4.2.2 Dataset.....	25
4.2.3 <i>Preprocessing</i> Data .....	26

4.2.4 Pembagian Data .....	29
4.2.5 Arsitektur Model .....	29
4.2.7 Evaluasi .....	31
<b>BAB V PENUTUP</b> .....	<b>34</b>
5.1 Kesimpulan .....	34
5.2 Saran .....	34
<b>DAFTAR PUSTAKA</b> .....	<b>35</b>



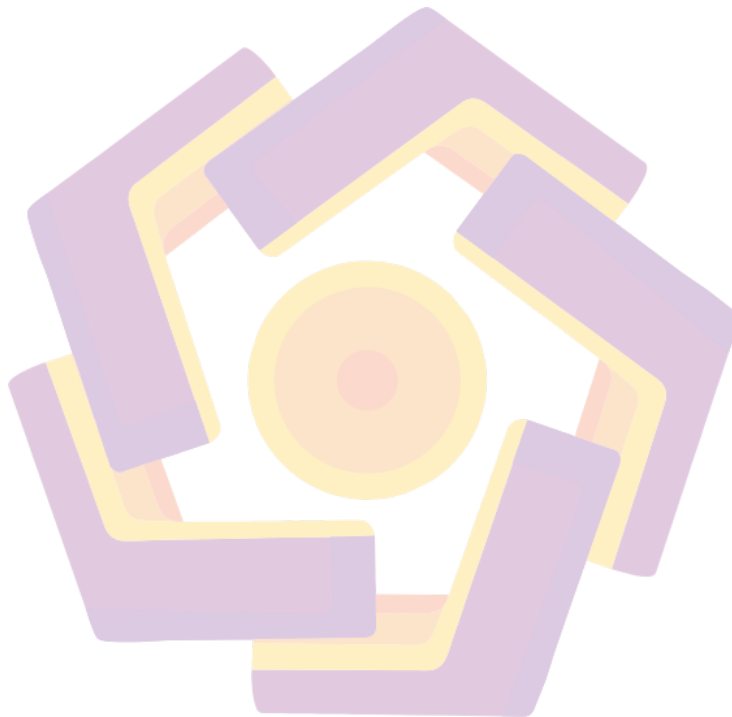
## DAFTAR TABEL

Tabel 2.1. Keaslian Penelitian.....	7
Tabel 3.1 Jumlah dataset liputan6.....	20
Tabel 4.1 Ukuran ROUGE pada daataset Liputan6.....	24
Tabel 4.2 Hasil <i>case folding</i> .....	25
Tabel 4.3 Hasil <i>stopword</i> .....	27
Tabel 4.4 Hasil <i>Tokenization</i> dan <i>Padding</i> .....	28
Tabel 4.5 Hasil Model GRU.....	29
Tabel 4.6 Hasil Ringkasan.....	30



## DAFTAR GAMBAR

Gambar 2.1. Skema GRU.....	14
Gambar 4.1. Isi <i>dataset</i> Liputan6.....	24
Gambar 4.2. <i>Dataset</i> Liputan6 setelah pembersihan .....	25
Gambar 4.3. <i>Splitting</i> Data.....	28
Gambar 4.4 Hasil permodelan dengan GRU.....	29



## INTISARI

Perubahan zaman yang begitu cepat, mengakibatkan teknologi yang berkembang juga melesat jauh perkembangannya. Salah-satu yang tampak dari perkembangan tersebut adalah Dengan adanya pertukaran informasi yang semakin maju dan juga cepat didapatkan. Salah-satu arus informasi yang berkembang saat ini adalah terkait artikel berita. Akan tetapi, artikel berita yang panjang dan banyak membuat para pembaca masih sangat kesulitan dalam menemukan intisari atau informasi penting dalam berita tersebut sehingga membutuhkan waktu yang lama dalam memahami artikel berita. Untuk mempermudah pembaca dalam menemukan intisari dari artikel berita, diperlukan sebuah Peringkasan Teks sehingga pembaca dapat mengefisienkan waktunya dalam memahami dan menangkap isi dari berita tersebut.

Dalam penelitian ini digunakan algoritma *Gated Recurrent Unit (GRU)* dengan pendekatan ekstraktif menggunakan *dataset* liputan6 untuk mengetahui seberapa besar akurasi yang didapatkan dalam peringkasan teks. Yang dimana algoritma ini bekerja dengan memproses data melalui *gated*. Yang dimana dari setiap *gated reset* dan *gated update* digunakan dalam proses input, *output* dan *forget* dalam pelatihan data. Sebelum data dilatih, data artikel berita yang akan digunakan dibersihkan dengan metode *preprocessing* yaitu *case folding*, *remove stopword*, *tokenization* dan *padding sequence*. Kemudian, didapat hasil data yang bersih dan dibagi menjadi data *training* dan data *testing*. Dari penelitian ini didapat Hasil evaluasi ringkasan artikel berita liputan6 menggunakan algoritma *Gated Recurrent Unit (GRU)* dengan membandingkan ringkasan mesin dan ringkasan manusia pada dataset Liputan6 sebanyak 193.883 menghasilkan nilai rata-rata *f1Score* pada ROUGE-L adalah 0,0375, ROUGE-1 adalah 0,0375, ROUGE-2 adalah 0. Dengan demikian kualitas ringkasan yang dihasilkan oleh algoritma *Gated Recurrent Unit (GRU)* pada artikel berita ini masih tergolong rendah

**Kata kunci:** GRU, Pendekatan Ekstraktif, *Gated Recurrent Unit*, Summarization, *text mining*.

## ABSTRACT

The changing times are so fast, resulting in the technology that develops also accelerates its development. One of the things that can be seen from this development is the exchange of information that is increasingly advanced and also quickly obtained. One of the information flows that is currently developing is related to news articles. However, long and many news articles make it very difficult for readers to find the essence or important information in the news so that it takes a long time to understand the news article. to make it easier for readers to find the essence of news articles, a Text Summarizer is needed so that readers can streamline their time in understanding and capturing the contents of the news.

In this study, the Gated Recurrent Unit (GRU) algorithm is used with an extractive approach using the liputan6 dataset to find out how much accuracy is obtained in text summarization. where this algorithm works by processing data through gated. Which of each gated reset and gated update is used in the input, output and forget process in data training. Before the data is trained, the news article data to be used is cleaned with preprocessing methods, namely case folding, remove stopwords, tokenization and sequence padding. Then, the results of clean data are obtained and divided into training data and testing data. From this study, the results of the evaluation of the summary of the liputan6 news article using the Gated Recurrent Unit (GRU) algorithm by comparing the machine summary and the human summary on the Liputan6 dataset of 193,883 resulted in an average value of f-1Score on ROUGE-L is 0,0375, ROUGE-1 is 0,0375, ROUGE-2 is 0. Thus the quality of the summary produced by the Gated Recurrent Unit (GRU) algorithm on this news article is still relatively low.

**Keyword:** GRU, Extractive Approach, Gated Recurrent Unit, Summarization, text mining.