

**PENGARUH PENERAPAN 4 TEKNIK *PREPROCESSING* PADA
KINERJA *AUTOMATIC TEXT SUMMARIZATION***

SKRIPSI

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



disusun oleh

MUHAMMAD FARHAN JUNA

19.11.3091

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

**PENGARUH PENERAPAN 4 TEKNIK *PREPROCESSING* PADA
KINERJA *AUTOMATIC TEXT SUMMARIZATION***

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

MUHAMMAD FARHAN JUNA

19.11.3091

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA**

YOGYAKARTA

2023

HALAMAN PERSETUJUAN

SKRIPSI

**PENGARUH PENERAPAN 4 TEKNIK *PREPROCESSING* PADA
KINERJA *AUTOMATIC TEXT SUMMARIZATION***

yang disusun dan diajukan oleh

Muhammad Farhan Juna

19.11.3091

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 Februari 2022

Dosen Pembimbing,



Mardhiya Hayaty, ST, M.Kom
NIK. 190302108

HALAMAN PENGESAHAN

SKRIPSI

PENGARUH PENERAPAN 4 TEKNIK *PREPROCESSING* PADA
KINERJA *AUTOMATIC TEXT SUMMARIZATION*

yang disusun dan diajukan oleh

Muhammad Farhan Juna

19.11.3091

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Februari 2022

Susunan Dewan Penguji

Nama Penguji

Arifivanto Hadinegoro, S.Kom, MT
NIK. 190302289

Tanda Tangan



Bayu Setiaji, M.Kom
NIK. 190302216



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



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Februari 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 : **Muhammad Farhan Juna**
NIM : **19.11.3091**

Menyatakan bahwa Skripsi dengan judul berikut:

Pengaruh Penerapan 4 Teknik *Preprocessing* Pada Kinerja *Automatic Text Summarization*

Dosen Pembimbing : **Mardhiya Hayaty, ST, 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, 20 Februari 2022

Yang Menyatakan,



Muhammad Farhan Juna

HALAMAN PERSEMBAHAN

Assalamu'alaikum Warahmatullahi Wabarakatuh

Puji syukur ke hadirat Tuhan yang Maha Esa, yang telah melimpahkan segala rahmat dan karuniaNya, sehingga penulis dapat menyelesaikan tugas akhir. Shalawat serta salam, semoga senantiasa tercurah kepada junjungan kita Nabi Besar Muhammad SAW, yang membawa dan menerangi hati nurani kita, menjadi cahaya bagi segala perbuatan mulia. Dan Insya Allah kita semua termasuk umat Nabi Muhammad SAW hingga akhir zaman.

Melalui kesempatan yang sangat berharga ini, persembahkan tugas akhir dan rasa terimakasih saya ucapkan untuk:

1. Untuk kedua orang tua tercinta, termakasih bapak dan ibu yang telah membantu dan membimbing anaknya mengerjakan tugas akhir. Tanpa doa dukungan dari orang tua mungkin saya tidak akan mampu menyelesaikan tugas akhir ini.
2. Untuk dosen pembimbing saya, bu Mardhiya Hayaty, ST, M.Kom atas ilmu dan bimbingan yang telah diberikan selama proses bimbingan.
3. Untuk teman-teman kelas IF-08 yang telah mengukir cerita bersama selama perkuliahan dan teman-teman himpunan yang telah memberikan banyak ilmu dan pengalaman diluar perkuliahan.
4. Untuk teman riset kolab yang selalu mendukung satu salam lain untuk menyelesaikan tugas akhir bersama.

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah subhanahu wa ta'ala yang telah melimpahkan nikmat dan rahmat-Nya yang melimpah, sehingga penulis dapat menyelesaikan Skripsi yang berjudul “Pengaruh Teknik Preprocessing Pada Kinerja Automatic Text Summarization” dengan tepat waktu. Penulis juga ingin menyampaikan rasa terima kasih yang sedalam-dalamnya kepada semua pihak yang telah membantu dalam pembuatan skripsi ini, khususnya kepada:

1. Bapak Prof, Dr. M. Suyanto, MM. selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta, M.Kom selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Ibu Windha Mega Pradnya D, M.Kom. selaku ketua Program Studi S1 Informatika Universitas Amikom Yogyakarta.
4. Bapak Rizqi Sukma Kharisma, S.Kom., M.Kom sebagai Dosen Wali yang selalu memberikan nasihat.
5. Ibu Mardhiya Hayaty, ST, M.Kom selaku dosen pembimbing yang telah membimbing penulis dalam proses pembuatan Skripsi
6. Seluruh Dosen Universitas AMIKOM Yogyakarta yang telah memberikan ilmu yang bermanfaat.
7. Teman-teman IF-08 angkatan 2019 yang telah menjadi keluarga selama proses akademik.
8. Semua pihak yang telah banyak membantu dalam penyusunan skripsi ini yang tidak dapat disebutkan satu persatu

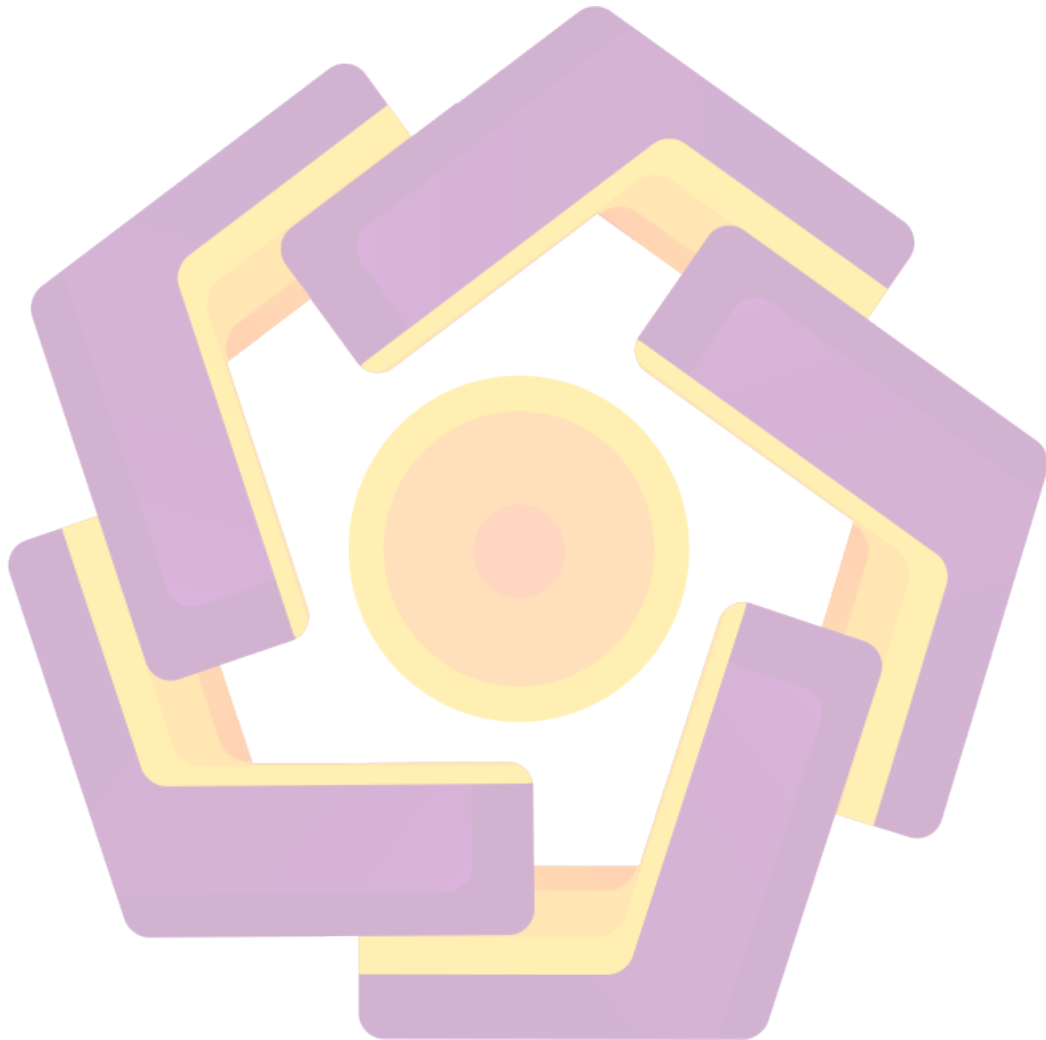
Yogyakarta, 20 Februari 2023

Penulis

DAFTAR ISI

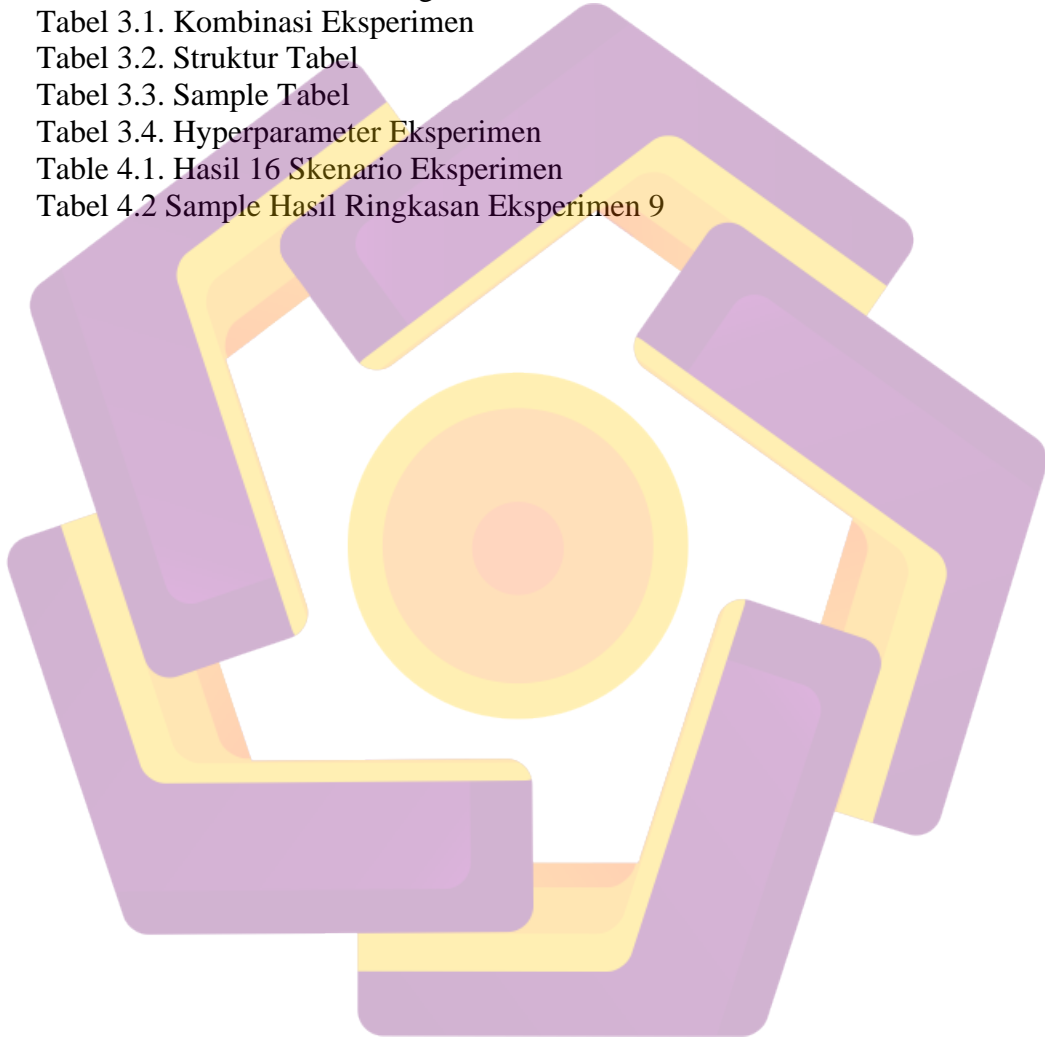
HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN.....	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH	xiii
INTISARI	xiv
ABSTRACT.....	xv
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	11
BAB III METODE PENELITIAN	20
3.1 Objek Penelitian.....	20
3.2 Alur Penelitian	20
3.3 Alat dan Bahan.....	22
BAB IV HASIL DAN PEMBAHASAN	26
4.1 Hasil Eksperimen.....	26
4.2 Hasil Ringkasan Model.....	28
BAB V PENUTUP	30

5.1 Kesimpulan	30
5.2 Saran	30
REFERENSI	31
LAMPIRAN.....	34



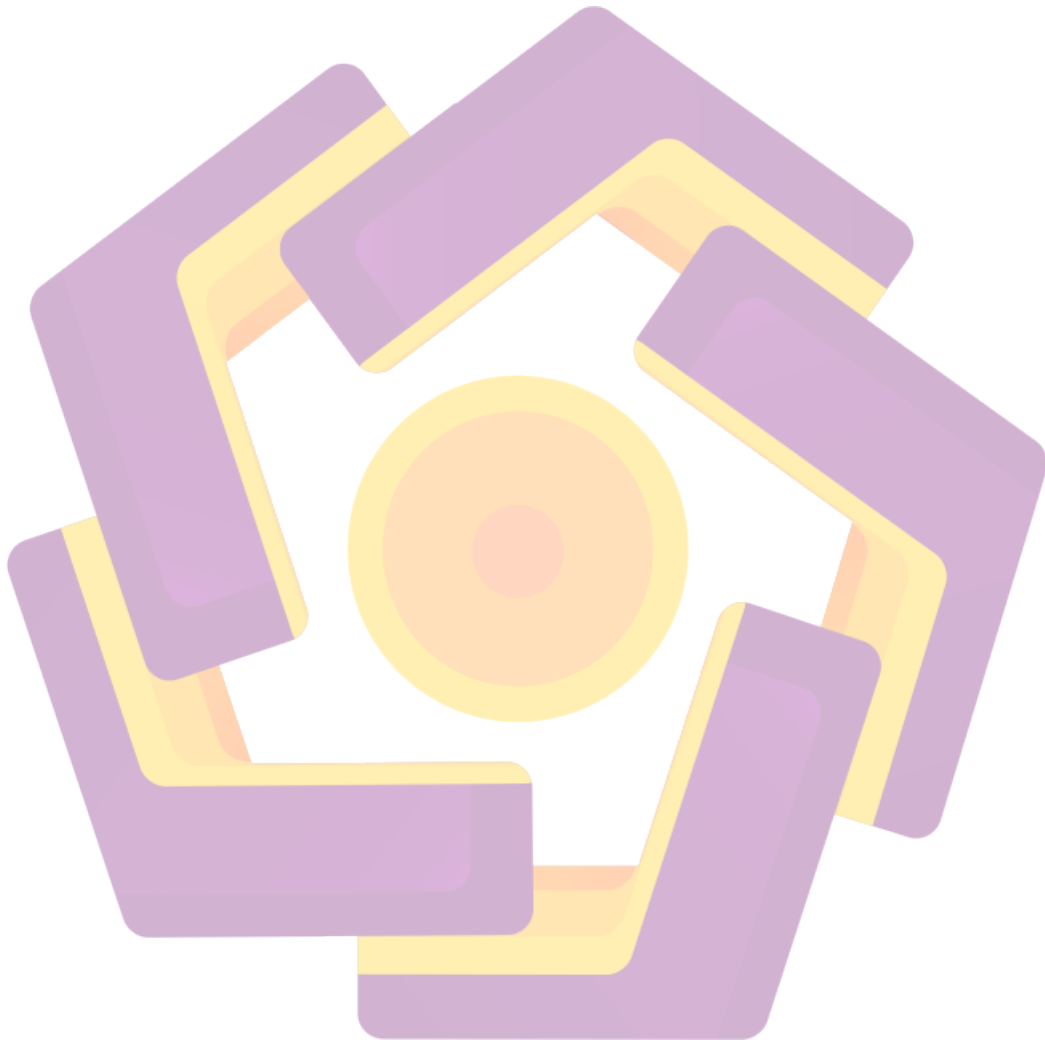
DAFTAR TABEL

Tabel 2.1. Keaslian Penelitian	8
Tabel 2.2. Proses Data Cleaning	16
Tabel 2.3. Proses Stemming	17
Tabel 2.4. Proses Stopwords	18
Tabel 2.5. Proses Case Folding	18
Tabel 3.1. Kombinasi Eksperimen	21
Tabel 3.2. Struktur Tabel	23
Tabel 3.3. Sample Tabel	23
Tabel 3.4. Hyperparameter Eksperimen	24
Table 4.1. Hasil 16 Skenario Eksperimen	26
Tabel 4.2 Sample Hasil Ringkasan Eksperimen 9	28

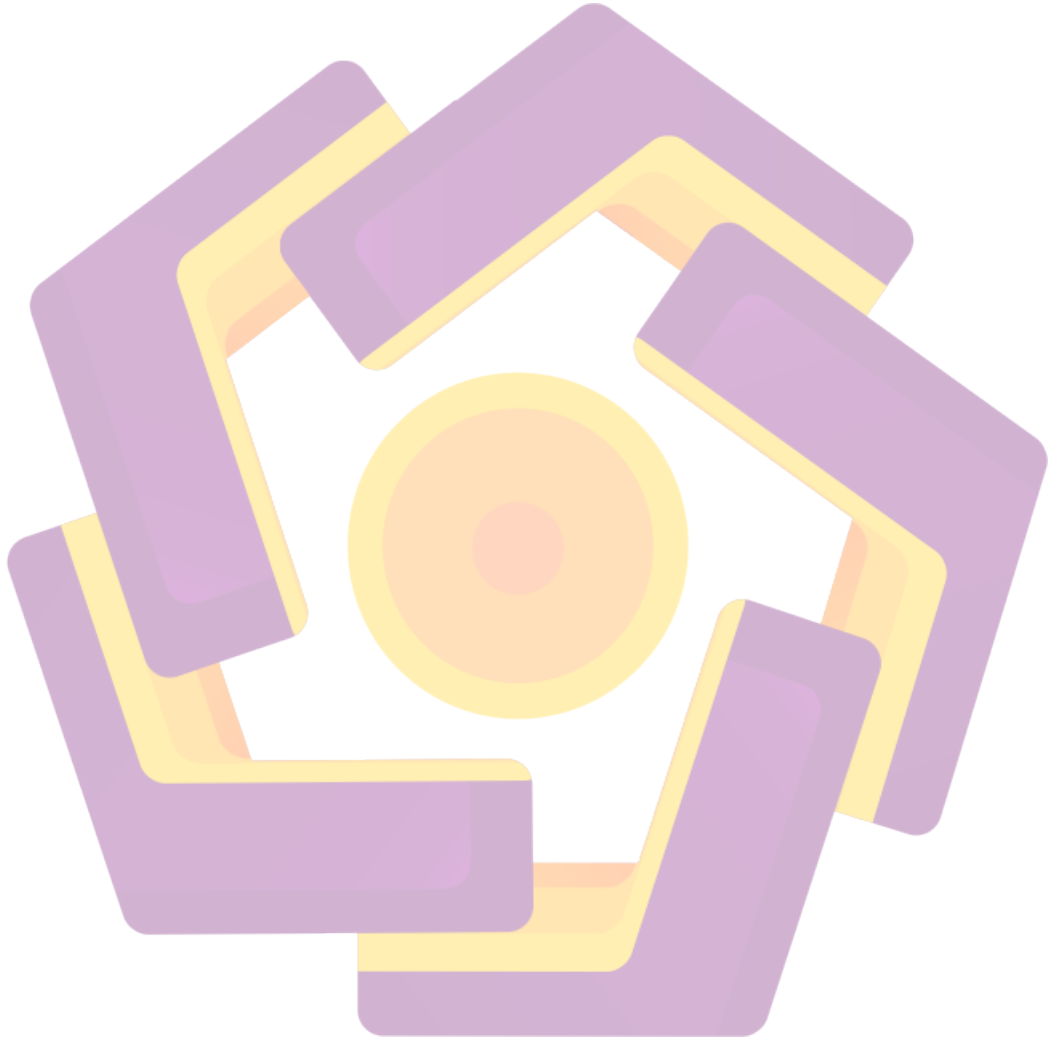


DAFTAR GAMBAR

Gambar 2.1. Arsitektur Transformers	14
Gambar 2.2. BERT Pre-training dan Fine-Tuning – Arsitektur Model	15
Gambar 3.1. Alur Penelitian	20
Gambar 4.1. Hasil R-Avg Score	27

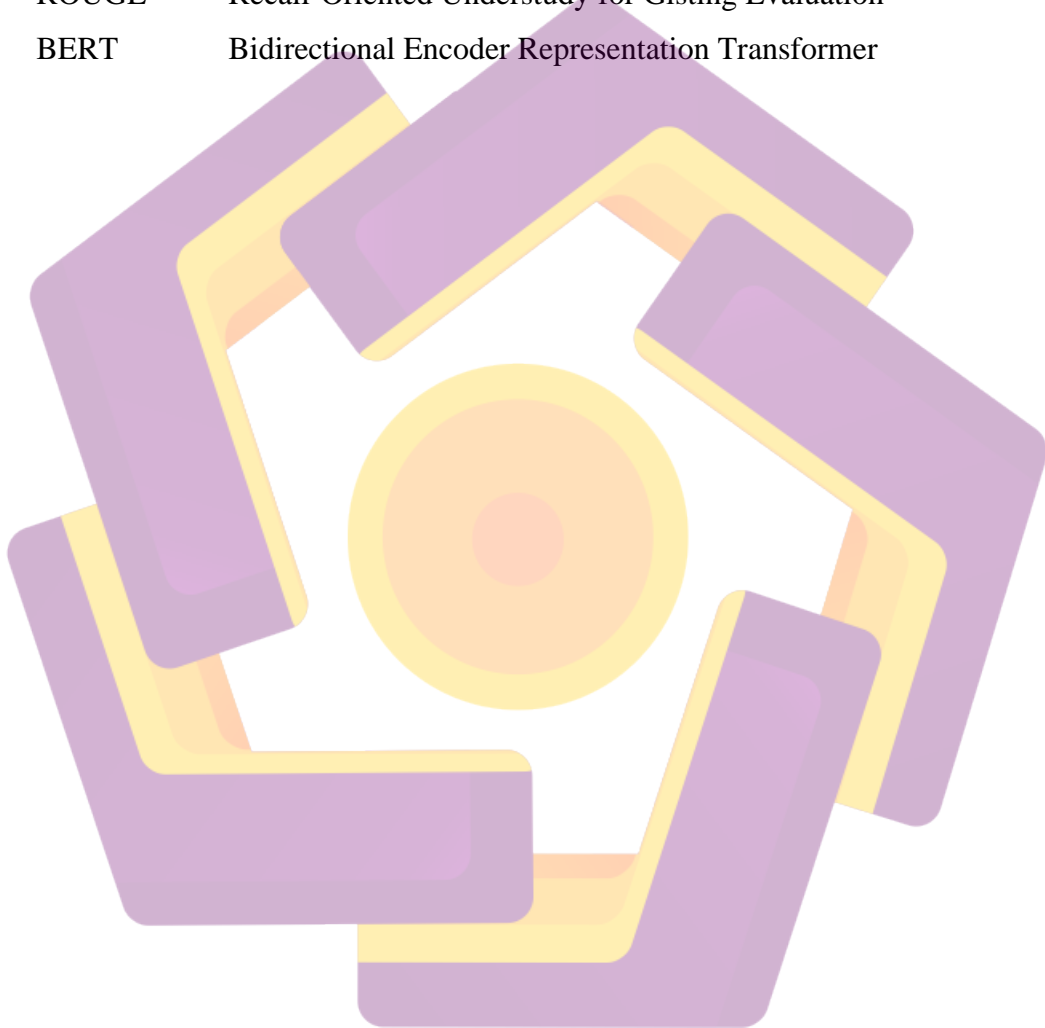


DAFTAR LAMPIRAN



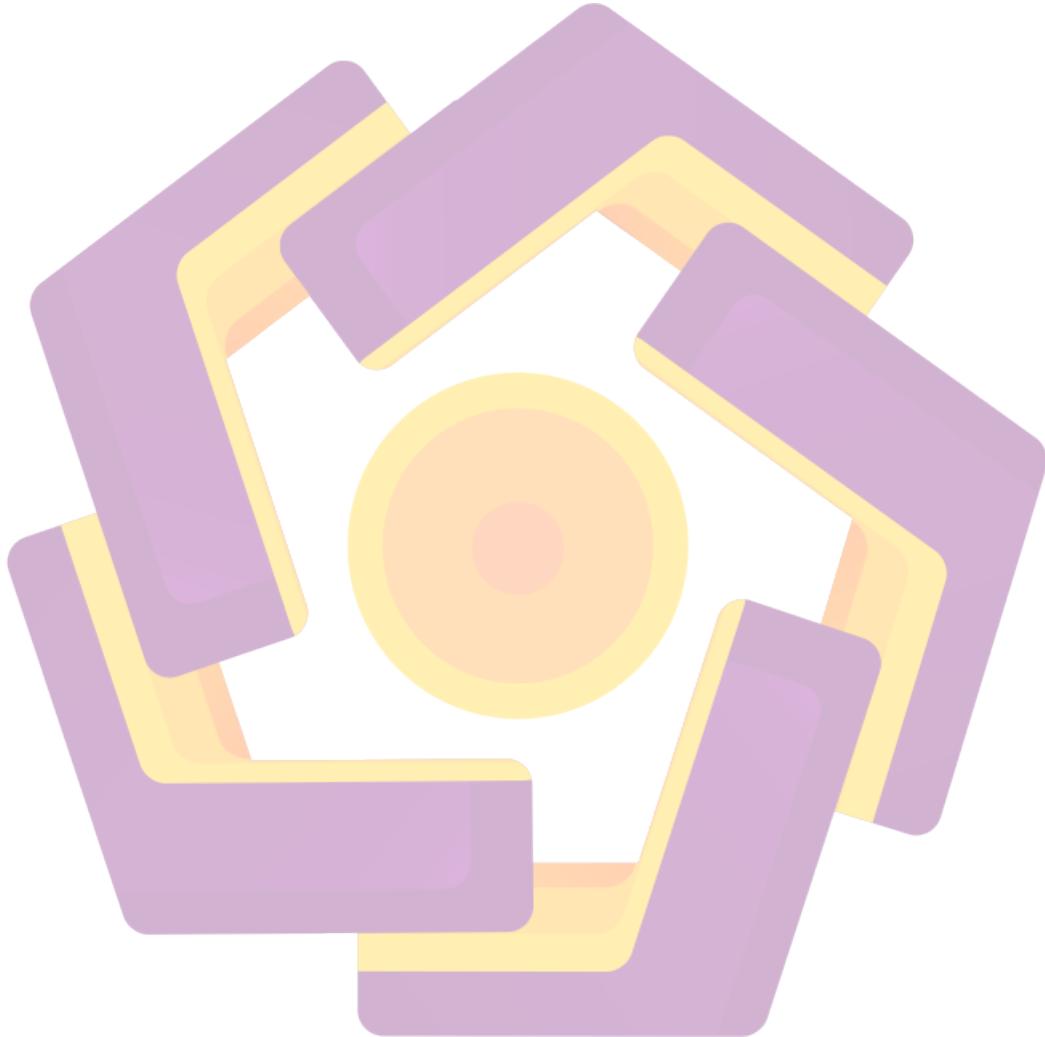
DAFTAR LAMBANG DAN SINGKATAN

AI	Artificial Intelligence
NLP	Natural Language
ATS	Automatic Text Summarization
ROUGE	Recall-Oriented Understudy for Gisting Evaluation
BERT	Bidirectional Encoder Representation Transformer



DAFTAR ISTILAH

Vektor besaran yang mempunyai arah



INTISARI

Informasi digital saat ini meningkat dengan sangat cepat dan manusia harus mengatasi ratusan hingga ribuan informasi di internet saat mencari informasi. Namun, ketika sebuah informasi terlalu panjang membuat sulit bagi pembacanya untuk memahami informasi secara cepat. Salah satu cara untuk mengatasi masalah ini adalah menggunakan Automatic text summarization (ATS). Permasalahan muncul ketika data yang digunakan dalam peringkasan teks masih bersifat kotor seperti tanda baca, URLs, karakter kosong, dll. Preprocessing merupakan tahap penting dalam ATS untuk membersihkan data kotor sehingga, penelitian ini bertujuan untuk mengetahui seberapa besar pengaruh preprocessing terhadap hasil kualitas ringkasan dalam peringkasan teks otomatis. Oleh karena itu, penelitian ini mengusulkan 16 skenario eksperimen yang akan menggunakan pre-trained model dari IndoBERT. Kajian ini akan secara khusus membahas teknik preprocessing dengan melakukan serangkaian percobaan dengan kombinasi preprocessing seperti data cleaning, stopwords, stemming, dan case folding. Kemudian, hasil penelitian akan diukur dengan menggunakan evaluasi ROUGE. Hasil penelitian ini menunjukkan bahwa kinerja terbaik diperoleh dengan menggunakan kombinasi pembersihan data dan pelipatan kasus, dengan skor ROUGE-1 (0,78), ROUGE-2 (0,60), dan ROUGE-L (0,68).

Kata kunci: automatic text summarization, preprocessing, rouge, data cleaning, peringkasan teks.

ABSTRACT

Today's digital information is increasing very quickly and humans have to navigate hundreds to thousands of pieces of information on the internet when looking for information. However, when a piece of information is too long it makes it difficult for the reader to understand the information quickly. One way to overcome this problem is to use Automatic text summarization (ATS). The problem arises when the data used in the text summary is still dirty, such as punctuation marks, URLs, empty characters, etc. Preprocessing is an important step in ATS to clean up dirty data, so this study aims to determine how much influence preprocessing has on the results of summary quality in automatic text summarization. Therefore, this study proposes 16 experimental scenarios that will use pre-trained models from IndoBERT. This study will specifically discuss preprocessing techniques by conducting a series of experiments with preprocessing combinations such as data cleaning, stopwords, stemming, and case folding. Then, the research results will be measured using the ROUGE evaluation. The results of this study indicate that the best performance is obtained by using a combination of data cleaning and case folding, with ROUGE-1 (0.78), ROUGE-2 (0.60), and ROUGE-L (0.68) scores.

Keyword: automatic text summarization, preprocessing, rouge, data cleaning, text summarizes.