

**RINGKASAN TEKS MENGGUNAKAN GENERATIVE PRE-
TRAINED TRANSFORMER 2**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1-Informatika



disusun oleh

AINI NUR KHASANA

19.11.2595

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2022

**RINGKASAN TEKS MENGGUNAKAN GENERATIVE PRE-
TRAINED TRANSFORMER 2**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1-Informatika



disusun oleh

AINI NUR KHASANAH

19.11.2595

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2022

HALAMAN PERSETUJUAN

SKRIPSI

**RINGKASAN TEKS MENGGUNAKAN GENERATIVE PRE-
TRAINED TRANSFORMER 2**

yang disusun dan diajukan oleh

Aini Nur Khasanah

19.11.2595

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 Februari 2023

Dosen Pembimbing,



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

HALAMAN PENGESAHAN
SKRIPSI
**RINGKASAN TEKS MENGGUNAKAN GENERATIVE PRE-
TRAINED TRANSFORMER 2**

yang disusun dan diajukan oleh

Aini Nur Khasanah

19.11.2595

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Februari 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Arif Dwi Laksito, M.Kom
NIK. 190302150

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

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



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 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 : Aini Nur Khasanah
NIM : 19.11.2595

Menyatakan bahwa Skripsi dengan judul berikut:

Ringkasan Teks Menggunakan Generative Pre-trained Transformer 2

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, 26 Januari 2023

Yang Menyatakan,



Aini Nur Khasanah

HALAMAN PERSEMBAHAN

Alhamdulillah segala puji bagi Allah SWT yang telah memberikan kelancaran dan keberhasilan bagi penulis dalam menyelesaikan penelitian pada skripsi ini. Skripsi ini ditujukan kepada semua orang yang telah membantu dalam mendukung perkuliahan dan menyelesaikan studi untuk mendapatkan gelar sarjana, yaitu :

1. Kedua orang tua dan adik penulis yang selalu mendukung dan mendoakan serta memberikan kasih sayangnya.
2. Keluarga besar penulis yang selalu memberikan doa.
3. Wahyuni Dwi Astuti, Pipin Ristriani, teman-teman sekelas, seangkatan, dan satu tim riset yang saling membantu dan sharing untuk berproses bersama.
4. Riefkyanov Surya Adia Pratama yang selalu membantu dan memberikan masukannya dalam pengerjaan penelitian pada skripsi ini.
5. Teman-teman dan sahabat penulis, Dina Sulistyowati, Anggun Lestari Ningtyastuti, Erika Yosita Putri, Dewi Setiyawati, Angga Reni Dwi Astuti, Nur Asikin yang selalu memberikan dukungan serta semangat kepada penulis.
6. Muhammad Muhajir Islam, Aryo Bimo Saputro, Deny Ali Saputra, Rachmat yang selalu berkembang bersama dalam menjalankan final project selama perkuliahan.
7. Dan untuk semua pihak yang berada disekitar penulis, terimakasih banyak atas dukungannya.

KATA PENGANTAR

Puji syukur kepada Allah SWT atas segala rahmat dan karunia-Nya, serta salam yang selalu ditujukan kepada Nabi Muhammad SAW sehingga penulis dapat menyelesaikan penelitian skripsi ini yang berjudul “Ringkasan Teks Menggunakan Model Generative Pre-trained Transformer 2” dengan baik.

Tanpa dukungan dan bantuan dari mereka, penulis tidak dapat menyelesaikan penelitian ini dengan baik. Dalam kesempatan ini, penulis ingin menyampaikan ucapan terima kasih kepada semua pihak yang telah membantu dalam penyusunan penelitian ini, dengan hormat :

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 dan Ibu Windha Mega Pradnya Duhita, M.Kom selaku Ketua Program Studi Informatika.
3. Ibu Mardhiya Hayaty S.T, M.Kom selaku dosen pembimbing yang telah memberikan arahan, ilmu dan kesabaran dalam membimbing penelitian dan penyusunan skripsi ini.
4. Bapak Arif Dwi Laksito, M.Kom dan Bapak Arif Akbarul Huda, S.Si, M.Eng selaku dosen penguji. Terimakasih atas saran yang telah diberikan selama pengujian guna memperbaiki sehingga menjadi lebih baik.

Penelitian pada skripsi ini disusun dengan tujuan untuk memberikan informasi yang bermanfaat bagi pembaca, sehingga berharap dapat memberikan kontribusi yang positif dalam bidang yang diangkat dalam penelitian ini dan dapat dijadikan sebagai referensi yang bermanfaat bagi pembaca, serta juga mengharapkan saran dan kritik yang membangun demi kesempurnaan penelitian ini.

Yogyakarta, 26 Januari 2023

Penulis

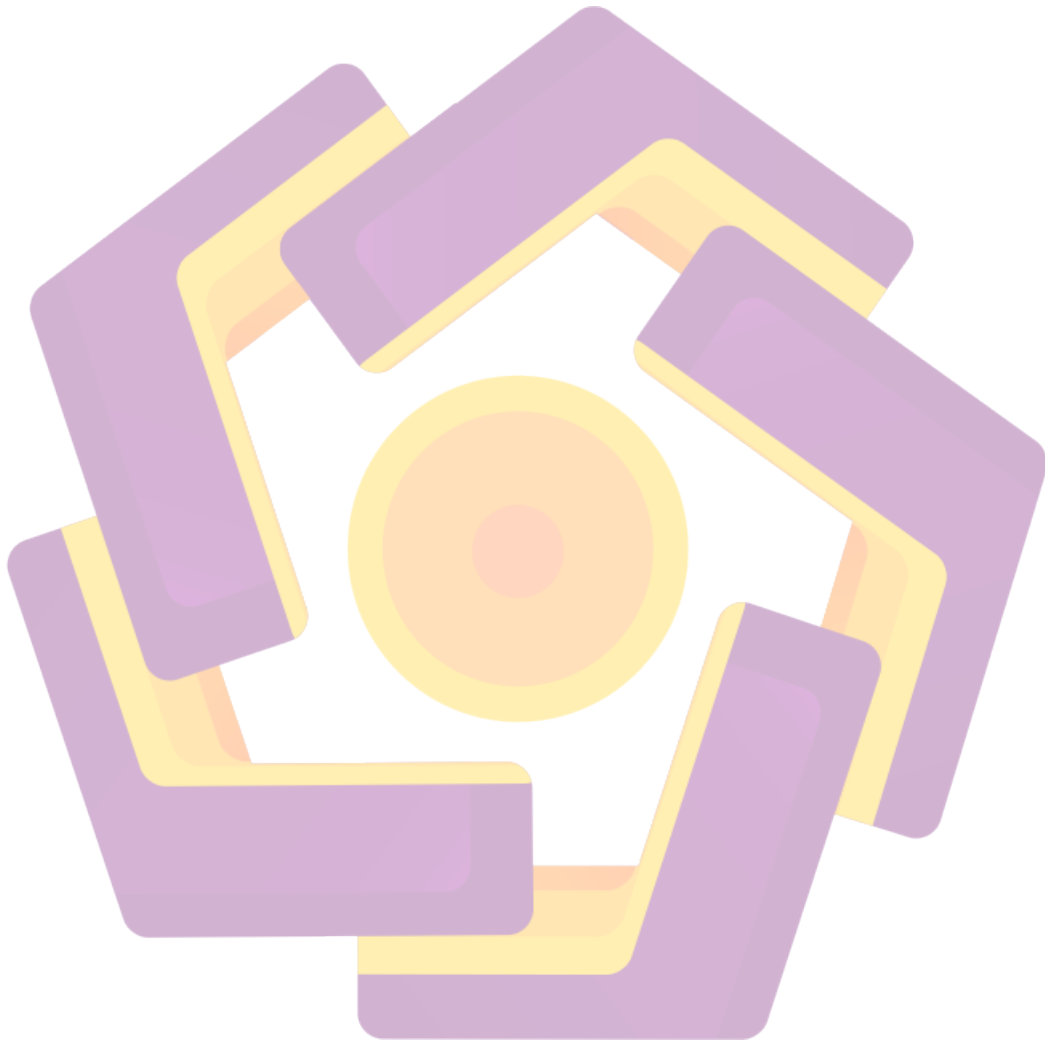
DAFTAR ISI

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 LAMBANG DAN SINGKATAN	xi
DAFTAR ISTILAH	xii
INTISARI	xiii
ABSTRACT	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	2
1.5 Manfaat Penelitian	2
1.6 Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA	4
2.1 Studi Literatur	4
2.2 Dasar Teori	8
2.2.1 Machine Learning dan Deep Learning	8
2.2.2 Automatic Text Summarization	8
2.2.3 Transformer	9
2.2.4 Model Generative Pre-trained Transformer 2 (GPT-2)	11
2.2.5 Evaluasi ROUGE	13

BAB III METODE PENELITIAN	14
3.1 Objek Penelitian	14
3.2 Alur Penelitian	14
3.2.1 Preparing Data	15
3.2.2 Preprocessing Data	15
3.2.3 Modeling Data	15
3.2.4 Evaluation Data	16
3.3 Alat dan Bahan Penelitian	16
3.3.1 Alat Penelitian	16
3.3.2 Bahan Penelitian	17
BAB IV HASIL DAN PEMBAHASAN	18
4.1. Hasil Penelitian	18
4.2. Pembahasan	20
BAB V PENUTUP	22
5.1 Kesimpulan	22
5.2 Saran	22
REFERENSI	24

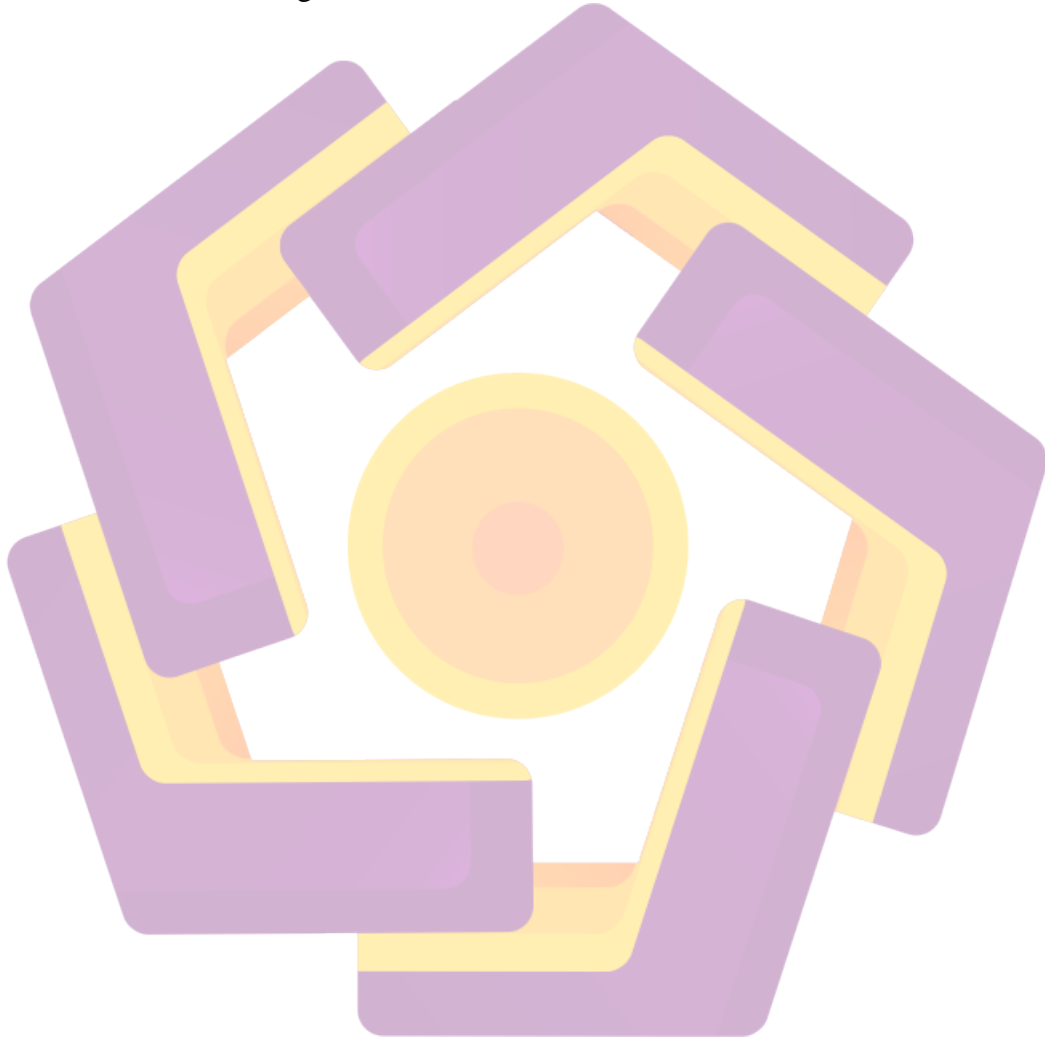
DAFTAR TABEL

Tabel 2.1. Keaslian Penelitian	6
Tabel 3.1. Hardware	16
Tabel 4.1. Hasil rata-rata evaluasi ROUGE	19
Table 4.2. Hasil peringkasan teks dengan skor rouge	19



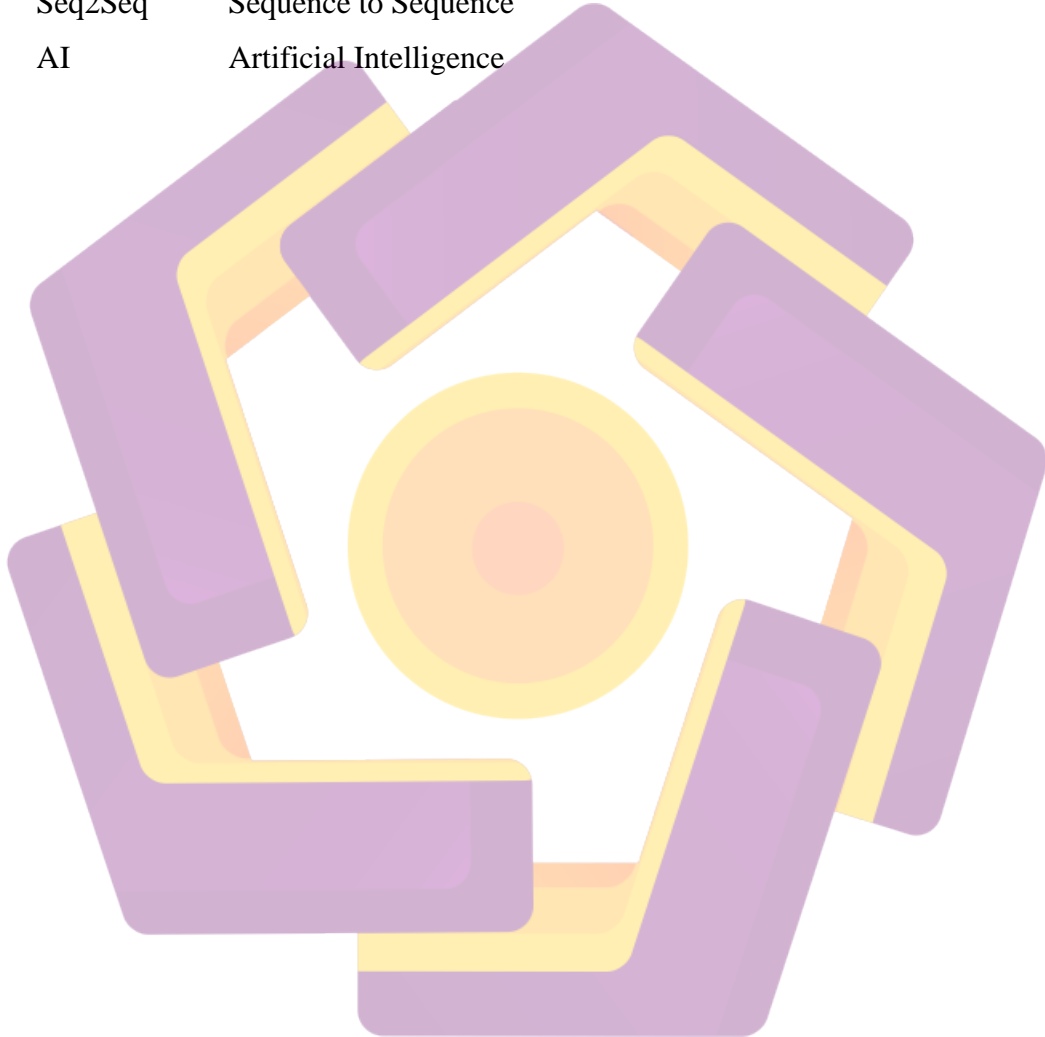
DAFTAR GAMBAR

Gambar 2.1. Model Arsitektur Transformer	10
Gambar 2.2. Transformer-Decoder Block GPT-2 dengan Masked Self Attention	11
Gambar 2.3. Masked Self Attention	12
Gambar 3.1. Alur Penelitian	14
Gambar 4.1. Training dan Validation Loss	18



DAFTAR LAMBANG DAN SINGKATAN

GPT	Generative Pre-trained Transformer
ROUGE	Recall-Oriented Understudy for Gisting Evaluation
NLP	Natural Language Processing
Seq2Seq	Sequence to Sequence
AI	Artificial Intelligence



DAFTAR ISTILAH

Hidden state	menyimpan informasi tentang input sebelumnya
Token	unit dasar dari teks dalam Pemrosesan Bahasa Alami (NLP)



INTISARI

Peringkasan teks otomatis adalah salah satu masalah yang dihadapi dalam bidang Pemrosesan Bahasa Alami (NLP), yang bertujuan untuk mengidentifikasi dan menyajikan informasi penting dari teks yang panjang secara cepat dan akurat. Ada dua pendekatan utama yang digunakan dalam proses peringkasan teks, yaitu peringkasan abstraktif dan peringkasan ekstraktif. Peringkasan abstraktif lebih sulit dilakukan karena tingkat kesulitan yang lebih tinggi, namun hasilnya lebih alami dan dapat menghasilkan kata-kata baru yang tidak ada dalam teks asli.

Dalam penelitian sebelumnya, RNN termasuk model *Seq2Seq* yang paling populer dalam peringkasan teks. Namun, metode ini masih memiliki kelemahan dalam menyimpan *memory* dan gradien yang hilang pada kalimat yang panjang, sehingga menyebabkan penurunan kualitas peringkasan teks yang panjang. Oleh karena itu, dalam penelitian ini mengusulkan untuk menggunakan model *Transformer* dengan mekanisme *Attention* yang dapat mengambil informasi penting dan dapat melakukan peringkasan pada teks yang panjang.

Model *Transformer* yang diusulkan adalah model GPT-2 (*Generative Pre-Trained Transformer 2*) yang menggunakan *decoder* sehingga dapat memprediksi kata berikutnya dengan menggunakan model pre-trained dari *w11wo/indo-gpt2-small* yang diimplementasikan pada dataset *Indosum* yang berisikan teks dan ringkasan berita berbahasa Indonesia *Indosum*. Kinerja model dievaluasi menggunakan evaluasi ROUGE. Hasil penelitian dengan model pre-trained GPT-2 yang diusulkan berhasil mencapai rata-rata *Recall* untuk R-1, R-2, dan R-L masing-masing 0.56, 0.46, dan 0.52.

Kata kunci: ringkasan teks, transformer, gpt-2.

ABSTRACT

Automatic text summarization is one of the problems faced in the field of Natural Language Processing (NLP), which aims to quickly and accurately identify and present important information from long texts. There are two main approaches used in the process of summarizing texts, namely abstractive summarization and extractive summarization. Abstractive summaries are more difficult due to the higher level of difficulty, but the results are more natural and can generate new words that were not in the original text.

In previous research, RNN is one of the most popular Seq2Seq models in text summarization. However, this method still has the disadvantage of memory saving and missing gradients in long sentences, causing a decrease in the quality of long text summaries. Therefore, this study proposes to use the Transformer model with an Attention mechanism which can retrieve important information and can summarize long texts.

The proposed Transformer model is the GPT-2 (Generative Pre-Trained 2) model which uses a decoder so that it can predict the next word using the pre-trained model of `w11wo/indo-gpt2-small` which is implemented Indosum dataset that contained text and summary about news in Indonesian. Model performance was evaluated using the ROUGE evaluation. The results of the research with the proposed GPT-2 pre-trained model succeeded in achieving an average Recall for R-1, R-2, and R-L respectively 0.56, 0.46, dan 0.52.

Keyword: text summarization, transformers, gpt-2