



**PERINGKASAN TEKS OTOMATIS ABSTRAK BAHASA
INDONESIA MENGGUNAKAN MODEL PEGASUS**

LAPORAN SKRIPSI

MOHAMAD IMAN SOLIHIN SUDRAJAT

41519110007

**UNIVERSITAS
MERCU BUANA**

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA JAKARTA

2023



**PERINGKASAN TEKS OTOMATIS ABSTRAK BAHASA
INDONESIA MENGGUNAKAN MODEL PEGASUS**

LAPORAN SKRIPSI

MOHAMAD IMAN SOLIHIN SUDRAJAT

41519110007

Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana

**UNIVERSITAS
MERCU BUANA**

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA JAKARTA

2023

HALAMAN PENYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Mohamad Iman Solihin Sudrajat
NIM : 41519110007
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Peringkasan Teks Otomatis Abstrak
Bahasa Indonesia Menggunakan Model PEGASUS

Menyatakan bahwa Laporan Skripsi ini adalah hasil karya saya sendiri dan bukan plagiat, serta semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Apabila ternyata ditemukan di dalam Laporan Skripsi saya terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

Jakarta, 19 Februari 2024

UNIVERSITAS
MERCU BUANA



Mohamad Iman Solihin Sudrajat

HALAMAN PENGESAHAN

Laporan Skripsi ini diajukan oleh:

Nama :Mohamad Iman Solihin Sudrajat

NIM :41519110007

Program Studi :Teknik Informatika

Judul Laporan Skripsi :Peringkasan Teks Otomatis Abstrak Bahasa Indonesia
Menggunakan Model PEGASUS

Telah berhasil dipertahankan pada sidang di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Strata 1 pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

Disahkan oleh:

Pembimbing : Dr. Ir. Eliyani

NIDN : 0321026901

Ketua Penguji : Dr. Bagus Priambodo, ST, M.TI

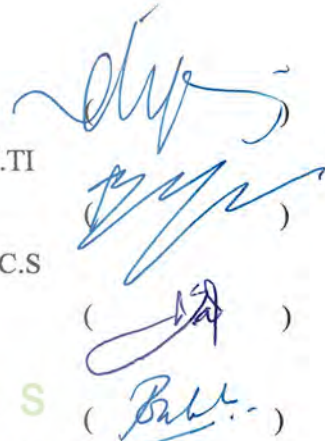
NIDN : 0313057905

Penguji 1 : Mohamad Yusuf, S.kom., M.C.S

NIDN : 0307097606

Penguji 2 : Roy Mubarak, ST., M.Kom

NIDN : 0310027402



(*[Signature]*)
(*[Signature]*)
(*[Signature]*)
(*[Signature]*)

UNIVERSITAS
MERCU BUANA

Jakarta, 19 Februari 2024

Mengetahui,

Dekan

Ketua Program Studi



Dr. Bambang Jekonowo, S.Si., M.T.I



Dr. Hadi Santoso, S.Kom., M.Kom

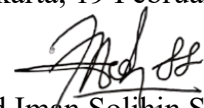
KATA PENGANTAR

Puji syukur kehadiran Allah SWT yang telah memberikan rahmat dan hidayah-Nya sehingga penulis mampu menyelesaikan penulisan skripsi ini dengan judul “Peringkasan Teks Otomatis Abstrak Bahasa Indonesia Menggunakan Model PEGASUS”. Penulis ingin mengucapkan terima kasih kepada pihak yang terlibat dan mendukung dalam penulisan skripsi ini. Oleh karena itu, saya mengucapkan terimakasih kepada:

1. Prof. Dr. Ir. Andi Adriansyah, M. Eng selaku Rektor Universitas Mercu Buana.
2. Dr. Bambang Jokonowo, S.Si., M.T.I. selaku Dekan Fakultas Ilmu Komputer.
3. Dr. Hadi Santoso, S.Kom., M.Kom selaku Ketua Program Studi Teknik Informatika.
4. Dr. Ir. Eliyani Selaku Dosen Pembimbing yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan saya dalam penyusunan skripsi ini;
5. Dr. Bagus Priambodo, ST., M.TI selaku Ketua Penguji Tugas Akhir atas koreksi dan arahan serta masukannya.
6. Mohamad Yusuf, S.kom., M.S.C selaku Penguji 1 Tugas Akhir atas koreksi dan arahan serta masukannya.
7. Roy Mubarak, ST., M.kom selaku Penguji 2 Tugas Akhir atas koreksi dan arahan serta masukannya.
8. Orang tua beserta adik dan teman-teman yang selalu memberikan doa, motivasi serta dukungan sehingga dapat mengerjakan tugas akhir ini.

Akhir kata, saya berharap Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga Laporan Skripsi ini membawa manfaat bagi pengembangan ilmu.

Jakarta, 19 Februari 2024



Mohamad Iman Solihin Sudrajat

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama : MOHAMAD IMAN SOLIHIN SUDRAJAT
NIM : 41519110007
Program Studi : TEKNIK INFORMATIKA
Judul Laporan Skripsi : PERINGKASAN TEKS OTOMATIS ABSTRAK BAHASA INDONESIA MENGGUNAKAN MODEL PEGASUS

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana Hak Bebas Royalti Non-Eksklusif (Non-exclusive Royalty-Free Right) atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Laporan Magang/Skripsi/Tesis/Disertasi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta,

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 19 Februari 2024
Yang Menyatakan,



Mohamad Iman Solihin Sudrajat

ABSTRAK

Nama : Mohamad Iman Solihin Sudrajat
NIM : 41519110007
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Peringkasan Teks Otomatis Abstrak Bahasa Indonesia
menggunakan Model PEGASUS
Pembimbing : Dr.Ir. Eliyani

Dokumen teks besar sulit bisa dipahami dan membutuhkan waktu untuk mengekstrak informasi penting. Salah satu cara mendapatkan ringkasan teks dengan cepat yaitu dengan peringkasan teks otomatis abstrak. Penelitian ini menggunakan dataset indosum yang berisi kumpulan teks berita. Dengan data yang digunakan 2000 sampel dengan ukuran size dokumen yang berkisar 1 paragraf – 22 paragraf. Model algoritma yang digunakan adalah PEGASUS `tunner007/pegasus_summarize` yang disediakan oleh library `huggingface`. Skenario yang dilakukan adalah penggunaan pre-processing dan split data. Skenario 1 implementasikan tanpa stemming dan stopwords removal dengan split data 70:30, lalu skenario 2 implementasikan stemming tanpa stopwords dengan split data 80:20 dan skenario 3 implementasikan stemming sama stopwords removal dengan dataset 90:10. Dari hasil penelitian menunjukkan bahwa skenario 1 memberikan hasil terbaik dengan nilai matriks ROUGE-1 precision 0.581760, recall 0.627699, f1-score 0.602227. lalu ROUGE-2 precision 0.461695, recall 0.498631, f1-score 0.478117 dan ROUGEL-L precision 0.545045, recall 0.588313, f1-score 0.564325.

Kata Kunci : Peringkasan Teks Otomatis, Abstrak, Bahasa Indonesia, PEGASUS, ROUGE

ABSTRACT

Name : Mohamad Iman Solihin Sudrajat
NIM : 41519110007
Study Program : Teknik Informatika
Title Thesis : Automatic Text Summarization of Indonesian Abstracts
Using PEGASUS Model
Counsellor : Dr.Ir. Eliyani

Large text documents are difficult to understand and take time to extract important information. One way to quickly summarize text is with abstract automatic text summarization. This research uses the indosum dataset which contains a collection of news texts. With the data used 2000 samples with document size ranging from 1 paragraph - 22 paragraphs. The algorithm model used is PEGASUS `tunner007/pegasus_summarize` provided by the huggingface library. The scenarios performed are the use of pre-processing and split data. Scenario 1 implement without stemming and stopwords removal with 70:30 split data, then scenario 2 implement stemming without stopwords with 80:20 split data and scenario 3 implement stemming and stopwords removal with 90:10 dataset. The results show that scenario 1 gives the best results with ROUGE-1 precision 0.581760, recall 0.627699, f1-score 0.602227. then ROUGE-2 precision 0.461695, recall 0.498631, f1-score 0.478117 and ROUGEL-L precision 0.545045, recall 0.588313, f1-score 0.564325.

Keywords: Automatic Text Summarization, Abstract, Bahasa Indonesia, PEGASUS, ROUGE

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENYATAAN KARYA SENDIRI	ii
HALAMAN PENGESAHAN.....	iii
KATA PENGANTAR.....	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS.....	v
ABSTRAK	vi
ABSTRACT.....	vii
DAFTAR ISI	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah.....	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Penelitian	4
BAB II TINJAUAN PUSTAKA	5
2.1 Penelitian Terdahulu.....	5
2.2 Teori Pendukung.....	17
2.2.1 Peringkasan Teks Otomatis.....	17
2.2.2 Peringkasan Teks Abstrak.....	18
2.2.3 Text Mining.....	19
2.2.4 Natural Language Processing	20
2.2.5 Transformers	21
2.2.6 Transformers : Self-Attention	22
2.2.7 Fine Tuning.....	23
2.2.8 Masked Language Model (MLM)	23
BAB III METODE PENELITIAN.....	25
3.1 Jenis Penelitian	25
3.2 Tahapan Penelitian.....	25
3.2.1 Pemilihan Dataset	26

3.2.2 Teks Preprocessing.....	26
3.2.3 Model Development	27
3.2.4 Training Model	28
3.2.5 Pengujian pada data test.....	28
3.2.6 Evaluasi.....	29
3.3 Arsitektur Model.....	30
BAB IV HASIL DAN PEMBAHASAN.....	32
4.1 Dataset	32
4.2 Pre-Processing	32
4.3 Pembuatan Model.....	43
4.3.1 Training Preparation	45
4.3.2 Training model.....	46
4.4 Visualisasi Data	49
4.4.1 Visualisasi Skenario Pertama.....	49
4.4.2 Visualisasi Skenario Kedua	50
4.4.3 Visualisasi Skenario Tiga.....	50
4.5 Pengujian	51
4.5.1 Evaluasi ROUGE	52
4.6 Analisis Hasil.....	53
4.6.1 Perbandingan Hasil Pengujian	59
4.6.2 Perbandingan penelitian Terdahulu.....	60
4.7 Kontribusi Penelitian.....	61
BAB V KESIMPULAN DAN SARAN.....	63
5.1 Kesimpulan.....	63
5.2 Saran	63
DAFTAR PUSTAKA	64
LAMPIRAN.....	67
Lampiran 1. Bimbingan.....	67
Lampiran 2. Halaman Persetujuan	68
Lampiran 3. Bukti Submit	69
Lampiran 4. Naskah Artikel Jurnal.....	70
Lampiran 5. Plagiarisme Check	78
Lampiran 6. Curriculum Vitae.....	79
Lampiran 7. Surat Pernyataan HKI (jika belum published HKI).....	81
Lampiran 8. Surat Pengalihan Hak Cipta	82

Lampiran 9. Sertifikat BNSP.....	83
Lampiran 10. Halaman Pernyataan Luaran Tugas Akhir	85



UNIVERSITAS
MERCU BUANA

DAFTAR TABEL

Tabel 2. 1 Penelitian terdahulu.....	5
Tabel 3. 1 Data deskriptif datasets indosum.....	26
Tabel 3. 2 Konfigurasi Hyperparameter.....	28
Tabel 4. 1 Hasil pre-processing.....	34
Tabel 4. 2 Contoh hasil proses tokenizing.....	42
Tabel 4. 3 Skenario Pembagian Data.....	45
Tabel 4. 4 Skenario Jumlah Epoch.....	47
Tabel 4. 5 Hasil evaluasi ROUGE skenario 1.....	53
Tabel 4. 6 Hasil peringkasan skenario 1.....	54
Tabel 4. 7 Hasil evaluasi ROUGE skenario 2.....	55
Tabel 4. 8 Hasil peringkasan skenario 2.....	56
Tabel 4. 9 Hasil evaluasi ROUGE skenario 3.....	57
Tabel 4. 10 Hasil peringkasan skenario 3.....	57
Tabel 4. 11 Hasil perbandingan dari 3 skenario.....	59
Tabel 4. 12 Perbandingan Penelitian Terdahulu.....	60



UNIVERSITAS
MERCU BUANA

DAFTAR GAMBAR

Gambar 2. 1 Arsitektur transformers.....	21
Gambar 3. 1 Tahapan penelitian.....	25
Gambar 3. 2 Model Pegasus bekerja.....	30
Gambar 3. 3 Contoh a self-supervised pada Pegasus.....	30
Gambar 4. 1 Sampel dataset indosum.....	32
Gambar 4. 2 Install modul.....	32
Gambar 4. 3 Import modul.....	33
Gambar 4. 4 Source code class pre-processing.....	33
Gambar 4. 5 Lanjutan source code proses case folding dan cleaning data.....	33
Gambar 4. 6 Lanjutan source code proses stemming.....	34
Gambar 4. 7 Lanjutan source code proses stopwords.....	34
Gambar 4. 8 Load pretrained tokenizer pegasus.....	42
Gambar 4. 9 Load pre-trained model pegasus.....	43
Gambar 4. 10 Source code class datasets.....	43
Gambar 4. 11 Source code lanjutan fungsi collate_fn.....	44
Gambar 4. 12 Source code persiapan training.....	46
Gambar 4. 13 Lanjutan source code training model.....	48
Gambar 4. 14 Grafik hasil skenario 1.....	50
Gambar 4. 15 Grafik hasil skenario 2.....	50
Gambar 4. 16 Grafik hasil skenario 3.....	50
Gambar 4. 17 Source code pengujian model pada data test.....	51
Gambar 4. 18 Lanjutan source code pengujian model pada data test.....	52
Gambar 4. 19 Lanjutan source code fungsi calcs_metrics.....	52
Gambar 4. 20 Lanjutan source code mencetak hasil rouge.....	52

DAFTAR LAMPIRAN

Lampiran 1 Bimbingan	67
Lampiran 2 Halaman Persetujuan	68
Lampiran 3 Bukti Submit.....	69
Lampiran 4 Naskah Artikel Jurnal	70
Lampiran 5 Plagiarisme Check	78
Lampiran 6 Curriculum Vitae	80
Lampiran 7 Surat Pernyataan HKI (jika belum published HKI)	81
Lampiran 8 Surat Pengalihan Hak Cipta.....	82
Lampiran 9 Sertifikat BNSP	83
Lampiran 10 Halaman Pernyataan Luaran Tugas Akhir	85



UNIVERSITAS
MERCU BUANA