UNIVERSITAS
GADJAH MADA
**Analisis Performa Model Hybrid Extractiveâ€“Abstractive Summarization (TextRank + IndoT5) pada Teks
Berita Bahasa Indonesia**
Ahmad Siddiq Priaji, Faizah, S.Kom., M.Kom.; Aina Musdholifah, S.Kom., M.Kom. Ph.D
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR ISI

HALAMAN PENGESAHAN....iii
PERNYATAAN....iv
KATA PENGANTAR....v
DAFTAR ISI....vi
DAFTAR GAMBAR....viii
DAFTAR TABEL....ix
INTISARI....x
ABSTRACT....xi
BAB I PENDAHULUAN....1
1.1. Latar Belakang....1
1.2. Rumusan Masalah....4
1.3. Batasan Masalah....5
1.4. Tujuan Penelitian....5
1.5. Manfaat Penelitian....6
BAB II TINJAUAN PUSTAKA....7
2.1. Peringkas Teks Otomatis....7
2.2. Pendekatan Ekstraktif....7
2.3. Pendekatan Abstraktif....8
2.4. Pendekatan *Hybrid*....8
2.5. TextRank....9
2.6. *Text-to-Text Transfer Transformer*....12
2.7. Dataset Liputan6....15
2.8. Metrik Evaluasi Model....16
2.9. Studi Peringkas Teks Otomatis dengan Pendekatan *Hybrid*....18
BAB III METODOLOGI PENELITIAN....22
3.1. Deskripsi Umum....22
3.2. Studi Literatur....23
3.3. Analisis Permasalahan....24
3.4. Akuisisi Data....25
3.5. Persiapan dan Pembersihan Data....26
3.6. Arsitektur Model *Hybrid*....27
3.6.1. Pra-Pemrosesan Artikel....28
3.6.2. Tahap Ekstraktif: TextRank....30
3.6.3. Tahap Abstraktif: IndoT5....33
3.7. Rancangan Pengujian dan Analisis....37
3.7.1. Metrik Evaluasi Kinerja Model....37
3.7.2. Skenario Pengujian....40
BAB IV IMPLEMENTASI....42
4.1. Lingkungan Pengembangan....42

Analisis Performa Model Hybrid Extractiveâ€“Abstractive Summarization (TextRank + IndoT5) pada Teks
Berita Bahasa Indonesia
Ahmad Siddiq Priaji, Faizah, S.Kom., M.Kom.; Aina Musdholifah, S.Kom., M.Kom. Ph.D
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

4.2. Persiapan Data 42
4.2.1. *Data Cleaning* 42
4.2.2. Rekonstruksi Teks 43
4.2.3. Konversi Format JSON ke CSV 44
4.3. Implementasi IndoT5 45
4.3.1. Pemuatan Data dan Pengambilan Sampel 46
4.3.2. Tokenisasi dan Pra-Pemrosesan *Input* 46
4.3.3. Konfigurasi *Low-Rank Adaptation* (LoRA) 47
4.3.4. Penyesuaian Dinamis (Data Collator) 48
4.3.5. Pengaturan Pelatihan 48
4.3.6. Eksekusi, Penyimpanan, dan Pengemasan 49
4.4. Implementasi TextRank 50
4.4.1. Pra-Pemrosesan Teks Khusus TextRank 50
4.4.2. Implementasi TextRank dengan TF-IDF 51
4.5. Implementasi Model *Hybrid* 54
4.5.1. Inisialisasi Model Bahasa 54
4.5.2. Alur Kerja (*Pipeline*) Ekstraktif ke Abstraktif 56
4.5.3. Otomatisasi Skenario Pengujian 57
4.6. Implementasi Evaluasi 59
4.6.1. Paralelisme Data pada Lingkungan Terbatas 59
4.6.2. Agregasi Data dan Pembersihan (*Data Merging*) 60
4.6.3. Perhitungan Metrik ROUGE dan Waktu Komputasi 61
BAB V HASIL DAN PEMBAHASAN 63
5.1. Distribusi Dataset 63
5.2. Hasil *Fine-Tuning* IndoT5 66
5.3. Hasil Evaluasi 68
5.3.1. Perbandingan Performa Model Utama 68
5.3.2. Perbandingan Pendekatan Ekstraktif Murni 72
5.3.3. Pengaruh Tingkat Kompresi pada *Model Hybrid* 73
5.3.4. Analisis Efisiensi Waktu Komputasi 75
5.4. Analisis *Ablation Study* 77
5.5. Pembahasan 79
5.5.1. Efektivitas Pendekatan *Hybrid* TextRank + IndoT5 79
5.5.2. Perbandingan TextRank dan *Lead* sebagai Modul Ekstraktif 81
5.5.3. *Trade-off* antara Kualitas dan Efisiensi Waktu 81
5.5.4. Analisis Kesenjangan dengan *Hybrid Oracle* 82
BAB VI SIMPULAN DAN SARAN 84
6.1. Simpulan 84
6.2. Saran 85
DAFTAR PUSTAKA 86
LAMPIRAN 89