

DAFTAR ISI

ABSTRAK	i
ABSTRACT	ii
DAFTAR ISI	iii
DAFTAR GAMBAR	v
DAFTAR TABEL	vii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	5
1.4.1 Manfaat Bagi Peneliti	5
1.4.2 Manfaat Bagi Masyarakat	5
1.4.3 Manfaat Bagi Akademik	6
1.5 Batasan Masalah	6
1.6 Kerangka Pemikiran	6
1.7 Sistematika Penulisan	9
BAB II LANDASAN TEORI	11
2.1 Penelitian Terdahulu	11
2.2 Dasar Teori	14
2.2.1 Text Summarization	14
2.2.2 Transformer Model	15
2.2.3 Extractive and Abstractive Summarization	15
2.2.4 Encoder dan Encoder-Decoder	16
2.2.5 Bidirectional Encoder Representations from Transformers (BERT) 17	17
2.2.6 Bidirectional and Auto-Regressive Transformers (BART)	17
2.2.7 Multilingual BART (MBART)	17
2.2.8 Dataset untuk Summarization Bahasa Indonesia	17
2.2.9 Fine-Tuning Transformers	17
2.2.10 Recall-Oriented Understudy of Gisting Evaluation (ROUGE)	18

2.2.11	BERTSCORE.....	18
2.2.12	Cross-Industry Standard Process for Data Mining (CRISP-DM)...	19
BAB III METODOLOGI		22
3.1	Metode yang digunakan	22
3.1.1	<i>Business Understanding</i>	23
3.1.2	<i>Data Understanding</i>	24
3.1.3	<i>Data Preparation</i>	24
3.1.4	<i>Modelling</i>	25
3.1.5	<i>Evaluation</i>	25
3.2	Sumber daya yang diperlukan	26
3.2.1	Dataset.....	26
3.2.2	Literatur.....	27
3.2.3	Peralatan.....	27
BAB IV HASIL DAN PEMBAHASAN.....		28
4.1	Hasil dan pembahasan	28
4.1.1	<i>Business Understanding</i>	28
4.1.2	<i>Data Understanding</i>	35
4.1.3	<i>Data Preparation</i>	37
4.1.4	<i>Modelling</i>	40
4.1.5	<i>Evaluation</i>	45
BAB V PENUTUP		58
5.1	Kesimpulan.....	58
5.2	Saran	60
DAFTAR PUSTAKA.....		61

DAFTAR GAMBAR

Gambar 1.1 Kerangka Pemikiran.....	7
Gambar 2.1 Model Proses CRISP-DM.....	20
Gambar 3.1 Metodologi Penelitian	22
Gambar 4.1 Latar Belakang Responden Survei	31
Gambar 4.2 Persentase Responden yang Memiliki Pengalaman Text Summarization	31
Gambar 4. 3 Model Transformer yang Diketahui atau Digunakan Responden	32
Gambar 4. 4 Faktor yang Dipertimbangkan dalam Memilih Model untuk Produksi	32
Gambar 4.5 Responden yang Mengalami Kesulitan Menentukan Model	33
Gambar 4. 6 Pandangan Responden tentang Pentingnya Penelitian Komparatif .	33
Gambar 4.7 Proses Web Scrapping untuk Dataset.....	36
Gambar 4.8 Pengecekan Dataset.....	36
Gambar 4.9 Struktur Folder Dataset	37
Gambar 4.10 Kode Memanggil Dataset.....	38
Gambar 4.11 Kode Pengambilan Kolom yang Diperlukan.....	38
Gambar 4.12 Kode Proses Data Cleaning.....	38
Gambar 4.13 Diagram Batang Distribusi Split Data.....	39
Gambar 4.14 Kode Text Chunking	39
Gambar 4.15 Kode Library Tokenisasi Model.....	40
Gambar 4.16 Kode Tokenisasi Otomatis BERT.....	41
Gambar 4.17 Kode Tokenisasi Otomatis BART	41
Gambar 4.18 Kode Tokenisasi Otomatis mBART	41
Gambar 4.19 Kode Pemilihan Model BERT.....	43
Gambar 4.20 Kode Pemilihan Model BART	43
Gambar 4.21 Kode Pemilihan Model mBART	43
Gambar 4.22 Inferensi Model BART	44
Gambar 4.23 Inferensi Model BERT	44
Gambar 4.24 Inferensi Model mBART	45

Gambar 4. 25 Kurva Konvergensi Rata-Rata Metrik pada 200 Data Uji (Pita = ± 1 Galat Baku)	53
Gambar 4. 26 Perbandingan Evaluasi Model: Diagram Batang ROUGE, Diagram Batang BERTScore, dan Heatmap Seluruh Metrik.....	54



DAFTAR TABEL

Tabel 2.1 The State of The Art	11
Tabel 3.1 Peralatan	27
Tabel 4. 1 Pemetaan Pertanyaan Survei terhadap Rumusan Masalah dan Urgensi	29
Tabel 4. 2 Rincian Dataset per Sumber Berita	35
Tabel 4. 3 Konfigurasi Model yang Digunakan	42
Tabel 4. 4 Perhitungan ROUGE pada Sampel 1	47
Tabel 4. 5 Perhitungan ROUGE pada Sampel 2	47
Tabel 4. 6 Perhitungan ROUGE pada Sampel 3	48
Tabel 4. 7 Hasil BERTScore pada Sampel 1, 2, dan 3	49
Tabel 4. 8 Nilai Evaluasi Lima Sampel Pertama Model BERT	49
Tabel 4. 9 Nilai Evaluasi Lima Sampel Pertama Model BART	50
Tabel 4. 10 Nilai Evaluasi Lima Sampel Pertama Model mBART	50
Tabel 4. 11 Jumlah Nilai 200 Sampel dan Rata-Rata Akhir	50
Tabel 4. 12 Rata-Rata Kumulatif, Galat Baku, dan Interval Kepercayaan 95% ...	52
Tabel 4. 13 Hasil ROUGE dan BERTScore Tiap Model (Rata-Rata 200 Data Uji)	54
Tabel 4. 14 Selisih BART terhadap BERT per Sampel dan Interval Kepercayaan 95%	55
Tabel 4. 15 Jumlah Sampel Uji dengan Skor Tertinggi pada Setiap Metrik	55