

DAFTAR ISI

| | |
|--|-----|
| HALAMAN PENGESAHAN..... | ii |
| PERNYATAAN BEBAS PLAGIASI..... | iii |
| PRAKATA..... | iv |
| DAFTAR ISI..... | v |
| DAFTAR GAMBAR..... | ix |
| DAFTAR TABEL..... | x |
| INTISARI..... | xi |
| ABSTRACT..... | xii |
| BAB I PENDAHULUAN..... | 1 |
| 1.1 Latar Belakang..... | 1 |
| 1.2 Rumusan Masalah..... | 3 |
| 1.3 Batasan Masalah..... | 4 |
| 1.4 Tujuan Penelitian..... | 4 |
| 1.5 Manfaat Penelitian..... | 4 |
| BAB II TINJAUAN PUSTAKA..... | 5 |
| BAB III LANDASAN TEORI..... | 12 |
| 3.1 Peringkasan Teks Otomatis..... | 12 |
| 3.1.1 Abstraktif..... | 12 |
| 3.1.2 Ekstraktif..... | 13 |
| 3.2 Dokumen Tunggal..... | 13 |
| 3.3 Pra-pemrosesan..... | 13 |
| 3.3.1 <i>Detokenization</i> | 13 |
| 3.3.2 <i>Remove Punctuations</i> | 13 |
| 3.3.3 <i>Case Folding</i> | 14 |
| 3.3.4 <i>Stemming</i> | 14 |
| 3.4 <i>Cosine Similarity</i> | 15 |
| 3.5 Ekstraksi Fitur..... | 16 |
| 3.5.1 <i>Title Score</i> | 16 |

| | |
|---|----|
| 3.5.2 <i>Position Score</i> | 17 |
| 3.5.3 <i>Similarity Sentence to Title (SRST)</i> | 18 |
| 3.5.4 <i>Similarity Sentence to Sentence Cluster (SRSC)</i> | 19 |
| 3.5.5 <i>Sentence Score (SS)</i> | 20 |
| 3.6 Representasi Vektor Pengukuran Kemiripan Kalimat | 22 |
| 3.6.1 Doc2Vec | 22 |
| 3.6.2 Sentence-BERT | 24 |
| 3.7 <i>Light Gradient Boosted Machine (LightGBM)</i> | 26 |
| 3.8 Evaluasi Hasil Ringkasan..... | 27 |
| 3.8.1 Intrinsik..... | 27 |
| 3.8.2 Ekstrinsik | 27 |
| 3.8.3 ROUGE..... | 28 |
| BAB IV RANCANGAN PENELITIAN | 30 |
| 4.1 Deskripsi Umum Penelitian..... | 30 |
| 4.2 Tahapan Penelitian | 30 |
| 4.3 Rancangan Pengumpulan Data..... | 31 |
| 4.4 Rancangan Prapemrosesan Data..... | 33 |
| 4.5 Rancangan Model..... | 34 |
| 4.5.1 Pengukuran Kemiripan Kalimat dengan Doc2Vec..... | 36 |
| 4.5.2 Pengukuran Kemiripan Kalimat dengan Sentence-BERT..... | 37 |
| 4.6 Rancangan Pengujian | 37 |
| 4.7 Rancangan <i>Summary Extraction</i> | 38 |
| 4.8 Rancangan Evaluasi..... | 39 |
| BAB V IMPLEMENTASI..... | 41 |
| 5.1 Lingkungan Implementasi Model..... | 41 |
| 5.2 Implementasi Pengumpulan Data..... | 42 |
| 5.3 Implementasi Pengembangan Model | 42 |
| 5.4 Implementasi Prapemrosesan | 43 |
| 5.5 Implementasi Ekstraksi Fitur <i>Title Score</i> | 44 |
| 5.6 Implementasi Ekstraksi Fitur <i>Sentence Position</i> | 44 |
| 5.7 Implementasi Pembuatan Model Doc2Vec | 45 |

| | |
|---|-----------|
| 5.8 Implementasi Pembuatan Model Sentence-BERT | 46 |
| 5.9 Implementasi Ekstraksi Fitur <i>Similarity Sentence to Title</i> (SRST) | 46 |
| 5.10 Implementasi Ekstraksi Fitur <i>Similarity Sentence to Sentence Cluster</i> (SRSC) | 47 |
| 5.11 Implementasi Ekstraksi Fitur <i>Sentence Score</i> | 49 |
| 5.12 Implementasi LightGBM | 50 |
| 5.13 Implementasi Pembuatan Ringkasan | 51 |
| 5.14 Implementasi Evaluasi Kualitas Hasil Ringkasan | 52 |
| BAB VI HASIL DAN PEMBAHASAN | 53 |
| 6.1 Hasil Implementasi Prapemrosesan | 53 |
| 6.1.1 Judul Berita | 53 |
| 6.1.2 Detokenisasi Kalimat Berita | 53 |
| 6.1.3 Pemisahan dan Pembersihan Kalimat dari Tanda Baca | 55 |
| 6.1.4 <i>Stemming</i> Judul Berita dan Kalimat Berita | 55 |
| 6.2 Hasil Implementasi | 56 |
| 6.2.1 Hasil Implementasi Doc2Vec | 58 |
| 6.2.3 Hasil Implementasi Sentence-BERT | 58 |
| 6.2.5 Performa Model LightGBM | 59 |
| 6.3 Pembahasan | 60 |
| 6.3.1 Perhitungan Kemiripan Kalimat Menggunakan Doc2Vec | 60 |
| 6.3.2 Perhitungan Kemiripan Kalimat Menggunakan Sentence-BERT | 61 |
| 6.3.3 Pengaruh Prapemrosesan <i>Stemming</i> | 64 |
| 6.3.4 Pengaruh Parameter pada LightGBM | 64 |
| 6.3.4 Kualitas Ringkasan berdasarkan Nilai ROUGE | 65 |
| BAB VII KESIMPULAN DAN SARAN | 70 |
| 7.1 Kesimpulan | 70 |
| 7.2 Saran | 70 |
| DAFTAR PUSTAKA | 71 |
| LAMPIRAN | 74 |
| Lampiran A: Nilai Fitur pada Doc2Vec tanpa <i>stemming</i> | 74 |
| Lampiran B: Nilai Fitur pada Sentence-BERT tanpa <i>stemming</i> | 75 |

| | |
|---|----|
| Lampiran C: Nilai ROUGE-1 pada Doc2Vec dengan LightGBM-Parameter1 tanpa <i>stemming</i> | 76 |
| Lampiran D: Nilai ROUGE-1 pada Doc2Vec dengan LightGBM-Parameter2 tanpa <i>stemming</i> | 77 |
| Lampiran E: Nilai ROUGE-1 pada Sentence-BERT dengan LightGBM-Parameter1 tanpa <i>stemming</i> | 78 |
| Lampiran F: Nilai ROUGE-1 pada Sentence-BERT dengan LightGBM-Parameter2 tanpa <i>stemming</i> | 79 |

DAFTAR GAMBAR

| | |
|---|----|
| Gambar 3.1 Proses peringkasan teks otomatis (Gunawan, dkk., 2017) | 12 |
| Gambar 3.2 Model PV-DM (Le dan Mikolov, 2014) | 23 |
| Gambar 3.3 Model PV-DBOW (Le dan Mikolov, 2014) | 24 |
| Gambar 3.4 Ilustrasi BERT base network (Towards Data Science, 2021) | 25 |
| Gambar 3.5 Ilustrasi Sentence-BERT | 26 |
| Gambar 4.1 Contoh teks berita dari dataset IndoSum | 33 |
| Gambar 4.2 Rancangan prapemrosesan data | 33 |
| Gambar 4.3 Rancangan pengembangan model | 35 |
| Gambar 5.1 Implementasi prapemrosesan untuk mendapatkan judul berita | 43 |
| Gambar 5.2 Implementasi detokenisasi untuk mendapatkan kalimat | 43 |
| Gambar 5.3 Implementasi <i>text cleaning</i> untuk membersihkan kalimat dari tanda baca..... | 44 |
| Gambar 5.4 Implementasi ekstraksi fitur title score | 44 |
| Gambar 5.5 Implementasi ekstraksi fitur sentence position | 45 |
| Gambar 5.6 Implementasi pembuatan model doc2vec | 45 |
| Gambar 5.7 Implementasi pembuatan model Sentence-BERT | 46 |
| Gambar 5.8 Implementasi ekstraksi fitur SRST menggunakan doc2vec | 47 |
| Gambar 5.9 Implementasi ekstraksi fitur SRST dengan Sentence-BERT | 47 |
| Gambar 5.10 Implementasi ekstraksi fitur SRSC dengan doc2vec | 48 |
| Gambar 5.11 Implementasi ekstraksi fitur SRSC dengan Sentence-BERT | 48 |
| Gambar 5.12 Implementasi ekstraksi fitur sentence score | 49 |
| Gambar 5.13 Implementasi LightGBM | 50 |
| Gambar 5.14 Hasil fine tuning LightGBM | 51 |
| Gambar 5.15 Implementasi pembuatan ringkasan | 51 |
| Gambar 5.16 Implementasi evaluasi hasil ringkasan menggunakan ROUGE | 52 |

DAFTAR TABEL

| | |
|--|----|
| Tabel 2.1 Perbandingan penelitian | 10 |
| Tabel 3.1 Probabilitas distribusi dari posisi kalimat (Gunawan, dkk., 2017) | 18 |
| Tabel 4.1 Contoh kalimat yang akan dicari kemiripannya | 36 |
| Tabel 4.2 Hasil pengukuran antar kalimat menggunakan Doc2Vec | 36 |
| Tabel 4.3 Parameter pada Doc2Vec | 36 |
| Tabel 4.4 Hasil pengukuran kemiripan antar kalimat menggunakan Sentence- BERT | 37 |
| Tabel 4.5 Parameter pada Sentence-BERT | 37 |
| Tabel 4.6 Parameter pada LightGBM | 38 |
| Tabel 4.7 Hasil ringkasan ideal (gold_label = True pada dataset) | 39 |
| Tabel 4.8 Hasil ringkasan oleh model | 39 |
| Tabel 4.9 Rancangan skenario implementasi | 40 |
| Tabel 6.1 Cuplikan judul berita | 53 |
| Tabel 6.2 Cuplikan kalimat berita | 54 |
| Tabel 6.3 Cuplikan hasil prapemrosesan | 55 |
| Tabel 6.4 Cuplikan judul berita setelah stemming | 55 |
| Tabel 6.5 Cuplikan kalimat berita setelah stemming | 56 |
| Tabel 6.6 Performa model LightGBM pada skenario 1 Doc2Vec | 58 |
| Tabel 6.7 Kualitas ringkasan (ROUGE-1) yang dihasilkan pada skenario 1 | 58 |
| Tabel 6.8 Performa model LightGBM pada skenario 2 Sentence-BERT | 59 |
| Tabel 6.9 Kualitas ringkasan yang dihasilkan pada skenario 2 | 59 |
| Tabel 6.10 Performa model LightGBM | 60 |
| Tabel 6.11 Waktu yang dibutuhkan pada Doc2Vec tanpa stemming | 60 |
| Tabel 6.12 Waktu yang dibutuhkan pada Doc2Vec dengan stemming | 61 |
| Tabel 6.13 Waktu yang dibutuhkan pada Sentence-BERT tanpa stemming | 62 |
| Tabel 6.14 Waktu yang dibutuhkan pada Sentence-BERT dengan stemming | 63 |
| Tabel 6.15 Hasil evaluasi ROUGE 1 | 68 |
| Tabel 6.16 Hasil evaluasi ROUGE 2 | 68 |
| Tabel 6.17 Hasil evaluasi ROUGE L | 69 |