

## DAFTAR ISI

HALAMAN PENGESAHAN .....	ii
HALAMAN PENGESAHAN DEWAN PENGUJI.....	iii
PERNYATAAN BEBAS PLAGIASI.....	iv
BUKTI BEBAS PLAGIASI.....	v
DAFTAR ISI .....	vi
DAFTAR GAMBAR .....	ix
DAFTAR TABEL .....	xi
CATATAN REVISI DOKUMEN .....	xiii
INTISARI.....	xiv
ABSTRACT .....	xv
RINGKASAN EKSEKUTIF.....	xvi
BAB 1    PENGANTAR .....	1
BAB 2    DASAR TEORI PENDUKUNG .....	3
2.1 <i>Natural Language Processing</i> .....	3
2.2 <i>Data Preprocessing</i> .....	3
2.2.1 <i>Case Folding</i> .....	3
2.2.2 <i>Tokenizing</i> .....	4
2.2.3 <i>Filtering atau Stopwords</i> .....	4
2.2.4 <i>Stemming</i> .....	4
2.3 <i>Term Frequency Inverse Document Frequency (TF-IDF)</i> .....	4
2.4 <i>Bidirectional Encoder Representations from Transformers (BERT)</i> .....	5
2.4.1 <i>IndoBERT</i> .....	5
2.5 <i>Cosine Similarity</i> .....	6
2.6 <i>Kurva ROC Curve dan AUC</i> .....	6
BAB 3    ANALISIS STUDI PUSTAKA KUNCI DAN PEMILIHAN METODE .....	8
3.1 <i>Proses Stemming</i> .....	8
3.2 <i>Pembobotan Teks</i> .....	9
3.2.1 <i>Pembobotan Berbasis Term Frequency</i> .....	9
3.2.2 <i>Pembobotan Berbasis Preferensi</i> .....	10
3.3 <i>Perhitungan Matrik Kemiripan</i> .....	12
3.3.1 <i>Cosine Similarity</i> .....	12
3.3.2 <i>Euclidean Distance</i> .....	12

3.4	Pemilihan Metode .....	13
BAB 4	DETAIL IMPLEMENTASI .....	14
4.1	Luaran dan Spesifikasi Capstone Project .....	14
4.1.1	Luaran Capstone Project .....	14
4.1.2	Spesifikasi Model Capstone Project.....	17
4.2	Batasan Masalah.....	17
4.3	Implementasi dengan Model Pra-pelatihan BERT.....	17
4.3.1	Masukan Data ( <i>Data Input</i> ).....	19
4.3.2	<i>Pre-processing</i> Masukan Data .....	20
4.3.3	Pembobotan Teks dengan Model Pra-pelatihan IndoBERT .....	21
4.3.4	Perhitungan Kemiripan dengan <i>Cosine Similarity</i> .....	21
4.3.5	Keluaran Data ( <i>Data Output</i> ).....	22
4.4	Implementasi dengan Pembobotan TF-IDF .....	22
4.4.1	Masukan Data ( <i>Data Input</i> ).....	24
4.4.2	<i>Pre-processing</i> Masukan Data .....	25
4.4.3	Pembobotan Teks dengan TF-IDF .....	26
4.4.4	Perhitungan Kemiripan dengan <i>Cosine Similarity</i> .....	26
4.4.5	Keluaran Data ( <i>Data Output</i> ).....	26
4.5	Pengujian Model .....	26
BAB 5	PENGUJIAN DAN PEMBAHASAN .....	29
5.1	Pengujian Fungsionalitas Model Identifikasi Kemiripan Laporan .....	29
5.1.1	Pengujian Fungsionalitas dengan Model Pra-pelatihan IndoBERT Tanpa <i>Stemming</i> dan <i>Stopword</i> .....	29
5.1.2	Pengujian Fungsionalitas dengan Model Pra-pelatihan IndoBERT dengan <i>Stemming</i> dan <i>Stopword</i> .....	31
5.1.3	Pengujian Fungsionalitas dengan Pembobotan TF-IDF Tanpa <i>Stemming</i> dan <i>Stopword</i> .....	34
5.1.4	Pengujian Fungsionalitas dengan Pembobotan TF-IDF dengan <i>Stemming</i> dan <i>Stopword</i> .....	36
5.2	Pengujian <i>Threshold</i> dan Akurasi Model Identifikasi Kemiripan Laporan .....	38
5.3	Pengujian Waktu Eksekusi Model Identifikasi Kemiripan Laporan.....	42
BAB 6	ANALISIS MENGENAI PENGARUH SOLUSI <i>ENGINEERING DESIGN</i> .....	44
6.1	Mengurangi Redundansi Data Sistem UPIK.....	44
6.2	Memberikan <i>User Experience</i> Lebih Baik Untuk Pengguna UPIK.....	44

6.3	Pengembangan Fitur Lain .....	44
BAB 7	KESIMPULAN DAN SARAN .....	45
7.1	Kesimpulan.....	45
7.2	Saran.....	45
REFERENSI.....		47