

## DAFTAR ISI

HALAMAN PERNYATAAN KEASLIAN SKRIPSI .....	ii
HALAMAN PENGESAHAN .....	iii
KATA PENGANTAR .....	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI.....	iv
ABSTRAK.....	v
DAFTAR ISI.....	viii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	3
1.3. Tujuan Penelitian .....	3
1.4. Ruang Lingkup.....	4
1.5. Sistematika Penulisan .....	5
BAB II LANDASAN TEORI.....	7
2.1. Studi Literatur .....	7
2.2. Sistem Rekomendasi.....	10
2.3. TF-IDF ( <i>Term Frequency – Inverse Document Frequent</i> ).....	13
2.3.1. Parameter n-gram .....	16
2.3.2. Parameter L2 Norm .....	16
2.3.3. Parameter <i>Smooth</i> IDF.....	16
2.3.4. Parameter Sublinear <i>Term Frequency</i> .....	17
2.4. Arsitektur Transformer .....	18
2.5. IndoBERT ( <i>Indonesian Bidirectional Encoder Representations from Transformers</i> ).....	19
2.5.1. Representasi Input (Representasi Teks Awal).....	21
2.5.2. Proyeksi Linear untuk <i>Query</i> , <i>Key</i> , dan <i>Value</i> .....	22
2.5.3. Perhitungan <i>Scaled Dot-Product Attention</i> .....	22
2.5.4. <i>Multi-Head Self-Attention</i> (MHSA) .....	23
2.5.5. <i>Hidden State</i> (Hi).....	24
2.5.6. Pembentukan Representasi Teks Dokumen ( <i>Mean Pooling</i> ).....	24
2.5.7. Normalisasi L2 untuk Stabilitas <i>Cosine Similarity</i> .....	25
2.5.8. <i>Fine-Tuning</i> .....	25
2.6. <i>Content-Based Filtering</i> .....	26

2.7. <i>Cosine Similarity</i> .....	27
2.8. <i>Normalized Discounted Cumulative Gain (NDCG)</i> .....	28
BAB III METODOLOGI PENELITIAN .....	33
3.1. Gambaran Umum Penyelesaian Masalah .....	33
3.2. Pengumpulan Data .....	34
3.3. <i>Data Understanding</i> .....	35
3.4. <i>Data Preprocessing</i> .....	35
3.5. Representasi Teks .....	37
3.5.1. TF-IDF ( <i>Term Frequency - Inverse Document Frequency</i> ).....	38
3.5.2. IndoBERT tanpa <i>Fine-Tuning</i> .....	41
3.5.3. IndoBERT dengan <i>Fine-Tuning</i> .....	50
3.6. Rekomendasi dengan <i>Cosine Similarity</i> .....	57
3.7. Evaluasi dengan metrik <i>Normalized Discounted Cumulative Gain (NDCG)</i> .....	60
BAB IV HASIL DAN PEMBAHASAN .....	70
4.1. Lingkungan dan Perangkat yang Digunakan dalam Penelitian .....	70
4.2. Pengumpulan Data .....	70
4.3. <i>Data Understanding</i> .....	71
4.4. <i>Data Preprocessing</i> .....	73
4.4.1. <i>Data Preprocessing</i> pada <i>data frame</i> Data Tourism with ID.....	73
4.4.2. <i>Data Preprocessing</i> pada <i>data frame</i> Data Tourism Rating .....	74
4.4.3. <i>Data Preprocessing</i> pada <i>Content-Based Filtering</i> .....	76
4.5. Representasi Teks .....	77
4.5.1. Representasi Teks TF-IDF.....	77
4.5.2. Representasi menggunakan metode IndoBERT tanpa <i>Fine-Tuning</i> .....	83
4.5.3. Representasi menggunakan metode IndoBERT dengan <i>Fine-Tuning</i> .....	84
4.6. Rekomendasi dengan <i>Cosine Similarity</i> .....	91
4.6.1. Rekomendasi <i>Cosine Similarity</i> dengan Representatif TF-IDF .....	92
4.6.2. Rekomendasi <i>Cosine Similarity</i> dengan IndoBERT tanpa <i>Fine-Tuning</i> .....	97
4.6.3. Rekomendasi <i>Cosine Similarity</i> dengan Representatif IndoBERT <i>Fine-Tuning</i> ...	99
4.6.4. Perbandingan hasil <i>Cosine Similarity</i> representatif TF-IDF dan IndoBERT .....	104
4.6.4.1. Rekomendasi dengan Representasi TF-IDF .....	105
4.6.4.2. Rekomendasi dengan Representasi IndoBERT tanpa <i>Fine-Tuning</i> .....	110
4.6.4.3. Rekomendasi dengan Representasi IndoBERT dengan <i>Fine-Tuning</i> ....	112

4.7. Evaluasi dengan metrik <i>Normalized Discounted Cumulative Gain</i> (NDCG) .....	118
4.7.1. Evaluasi NDCG dengan Representasi TF-IDF.....	119
4.7.2. Evaluasi NDCG dengan Representasi IndoBERT tanpa <i>Fine-Tuning</i> .....	120
4.7.3. Evaluasi NDCG dengan Representasi IndoBERT dengan <i>Fine-Tuning</i> .....	121
4.7.4. Analisis Perbandingan Representasi TF-IDF, IndoBERT <i>Pretrained</i> , dan IndoBERT <i>Fine-Tuning</i> .....	122
BAB V KESIMPULAN DAN SARAN .....	125
5.1. Kesimpulan .....	125
5.2. Saran .....	126
DAFTAR PUSTAKA .....	127
LAMPIRAN.....	131