

DAFTAR ISI

HALAMAN PENGESAHAN.....	ii
PERNYATAAN BEBAS PLAGIASI	iii
MOTTO DAN PERSEMBAHAN	iv
KATA PENGANTAR.....	v
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xii
DAFTAR LAMPIRAN	xiii
INTISARI	xiv
ABSTRACT.....	xv
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
1.6 Tinjauan Pustaka	7
1.7 Metodologi Penelitian	10
1.8 Sistematika Penulisan	10
BAB II LANDASAN TEORI.....	13
2.1 Ringkasan Numerik.....	13
2.2 Probabilitas	20
2.2.1 Eksperimen Acak dan Ruang Sampel	20
2.2.2 Event dan Operasi Himpunan	20
2.2.3 Definisi Aksiomatik Probabilitas	21
2.2.4 Sifat-sifat dan Teorema Dasar Probabilitas	22
2.2.5 Probabilitas Bersyarat dan Independensi.....	22
2.3 Variabel <i>Random</i>	23
2.4 <i>Semantic similarity</i>	25
2.5 <i>Machine learning</i>	26
2.6 <i>Deep learning</i>	28

2.7	<i>Artificial Neural Network</i>	29
2.8	<i>Natural Language Processing</i>	32
2.9	<i>Text preprocessing</i>	33
2.10	<i>Text Classification</i>	35
2.11	<i>Multi-label Retrieval</i>	37
2.12	<i>Hardware Accelerator</i>	39
BAB III ANALISIS KOMPARATIF MODEL TRANSFORMER UNTUK PEMETAAN OTOMATIS SCOPE OF SUPPLY PADA PENGADAAN JASA		41
3.1	<i>Procurement</i>	41
3.1.1	<i>Procurement Cycle</i>	43
3.1.2	<i>Scope of Supply</i>	45
3.1.3	Klasifikasi Baku Lapangan Usaha Indonesia (KBLI)	46
3.2	Pengukuran Kemiripan Teks Berbasis Vektor	48
3.2.1	Konsep Dasar Vektor dalam Ruang Euclid	51
3.2.2	<i>Cosine similarity</i>	53
3.3	<i>Model Transformer</i>	54
3.3.1	<i>Encoder</i>	56
3.3.2	<i>Decoder</i>	57
3.4	BERT	59
3.4.1	Arsitektur Model	60
3.4.2	<i>Pre-training Task Model BERT</i>	63
3.4.3	IndoBERT	70
3.5	DistilBERT	72
3.5.1	Arsitektur Model DistilBERT	73
3.6	XLNet	76
3.7	Metrik Evaluasi Model	84
3.8	Alur Analisis Eksperimen	90
BAB IV STUDI KASUS		93
4.1	Deskripsi Data	94
4.2	Eksplorasi dan Visualisasi Data	109
4.2.1	Pemeriksaan <i>Missing value</i> dan Data Duplikat	110
4.2.2	Visualisasi Distribusi Jumlah ISC per <i>Service Code</i>	111
4.2.3	Visualisasi Distribusi Panjang Teks	113
4.2.4	Karakteristik <i>Multi-label</i> pada <i>Ground truth</i>	115
4.3	<i>Text preprocessing</i>	116

4.3.1	<i>Case Folding</i>	117
4.3.2	<i>Cleaning Text</i>	117
4.3.3	Normalisasi dan Formalisasi Bahasa	118
4.3.4	<i>Stopword Removal</i>	119
4.3.5	<i>Stemming</i>	120
4.3.6	Pemeriksaan Data Setelah <i>Text Preprocessing</i>	121
4.4	Pengaturan Model	122
4.5	<i>Output Mapping</i> Model	123
4.5.1	<i>Output Mapping</i> IndoBERT	123
4.5.2	<i>Output Mapping</i> DistilBERT	124
4.5.3	<i>Output Mapping</i> XLNet	126
4.6	Perbandingan Performa dan Kinerja Model	129
BAB V PENUTUP		132
5.1	Kesimpulan	132
5.2	Saran	134
DAFTAR PUSTAKA		135
LAMPIRAN		139