

DAFTAR ISI

JUDUL	i
LEMBAR PENGESAHAN	ii
PERNYATAAN BEBAS PLAGIASI	iii
KATA PENGANTAR	iv
DAFTAR ISI.....	v
DAFTAR GAMBAR	viii
DAFTAR TABEL.....	x
DAFTAR SIMBOL	xi
INTISARI	xii
ABSTRACT.....	xiii
BAB I PENDAHULUAN	1
1.1. Latar belakang	1
1.2. Rumusan masalah.....	3
1.3. Tujuan dan Manfaat Proyek Akhir.....	3
1.4. Batasan penelitian.....	5
1.5. Sistematika penulisan	5
BAB II KAJIAN PUSTAKA	7
2.1. Studi Pustaka	7
2.2. Dasar Teori	15
2.2.1. Database.....	15
2.2.2. Big Data.....	15
2.2.3. Data Warehouse.....	16

2.2.4.	Hadoop	16
2.2.5.	Hadoop <i>Distributed File System</i> (HDFS).....	17
2.2.6.	<i>Artificial Intelligence</i>	18
2.2.7.	Natural Language Processing	20
2.2.8.	Transformer	20
2.2.9.	BERT	21
2.2.10.	DistilBERT	22
2.2.11.	MiniLM	23
2.2.12.	ELECTRA	24
2.2.13.	FP16 Mixed Precision	26
2.2.14.	Metrik Evaluasi	27
2.2.15.	Metriks Kinerja Komputasi	29
2.3.	Hipotesis	30
BAB III	METODE PENELITIAN	31
3.1.	Alat dan Bahan	31
3.1.1	Dataset	31
3.1.2	Perangkat Keras.....	32
3.1.3	Perangkat Lunak.....	32
3.2.	Tahapan proyek	33
3.2.1	Identifikasi Masalah	33
3.2.2	Studi Literatur.....	34
3.2.3	Identifikasi Kebutuhan Sistem	34
3.2.4	Perancangan Sistem.....	34
3.2.5	Implementasi Sistem	34
3.3.	Rancang Bangun Sistem Web	45

BAB IV	HASIL DAN PEMBAHASAN	49
4.1	Analisis Efisiensi Penyimpanan dan Query Data.....	49
4.1	Evaluasi Kinerja Model.....	50
4.2	Analisis Pemanfaatan Sumber Daya	54
4.3	Pengujian Sistem Web	57
BAB V	PENUTUP	72
5.1	Kesimpulan.....	72
5.2	Saran.....	72
DAFTAR PUSTAKA	74
LAMPIRAN.....		78

DAFTAR GAMBAR

Gambar 2. 1 Arsitektur HDFS	17
Gambar 2. 2 Cakupan AI	18
Gambar 2. 3 Arsitektur Transformer [24]	21
Gambar 2. 4 Mekanisme MLM.....	22
Gambar 2. 5 Mekanisme NSP.....	22
Gambar 2. 6 Komparasi parameter antara <i>pretrained models</i> [26].....	23
Gambar 2. 7 Proses Distilasi <i>Deep Self Attention</i> [27]	24
Gambar 2. 8 Skema <i>replaced token detection</i> [28]	25
Gambar 2. 9 Implementasi <i>Mixed Precision training</i> [29]	27
Gambar 2. 10 <i>Confusion matrix</i>	28
Gambar 3. 1 Diagram Alir Tahapan Penelitian.....	33
Gambar 3. 2 Kode instalasi dan <i>import library</i>	35
Gambar 3. 3 Alur kombinasi data	35
Gambar 3. 4 Data mentah <i>Event ID</i>	36
Gambar 3. 5 Data mentah <i>Event Template</i>	36
Gambar 3. 6 Proses <i>mapping</i> secara keseluruhan	38
Gambar 3. 7 Proses menggabungkan data	39
Gambar 3. 8 Proses <i>train-test splitting</i>	39
Gambar 3. 9 Proses <i>encode label</i>	39
Gambar 3. 10 Model <i>Loading</i> dan tokenisasi data.....	40
Gambar 3. 11 Alur logika pembuatan <i>class</i>	41
Gambar 3. 12 Konfigurasi Dataset dan DataLoader	41
Gambar 3. 13 Konfigurasi model dan implementasi FP-16	42
Gambar 3. 14 Fungsi pelatihan	43
Gambar 3. 15 Fungsi evaluasi.....	44
Gambar 3. 16 Proses iterasi pelatihan dan evaluasi	45
Gambar 3. 17 Konfigurasi dan integrasi model	45
Gambar 3. 18 Desain sistem web.....	46
Gambar 3. 19 Kode model <i>loading</i> dan input data dari user	47
Gambar 3. 20 Kode memilih antara ketiga opsi model.....	47

Gambar 3. 21 Diagram alir proses inferensi	48
Gambar 4. 1 Grafik <i>train loss</i> ketiga model.....	50
Gambar 4. 2 Grafik <i>val loss</i> ketiga model	51
Gambar 4. 3 Grafik <i>train accuracy</i> ketiga model	52
Gambar 4. 4 Grafik <i>val accuracy</i> ketiga model	52
Gambar 4. 5 Grafik <i>CPU Utilization</i> ketiga model.....	54
Gambar 4. 6 Grafik <i>Disk utilization</i> ketiga model	55
Gambar 4. 7 Grafik penggunaan GPU ketiga model	57
Gambar 4. 8 DistilBERT dengan input format <i>Event Id</i>	58
Gambar 4. 9 DistilBERT dengan input format <i>log message</i>	58
Gambar 4. 10 MiniLM dengan input format <i>Event Id</i>	59
Gambar 4. 11 MiniLM dengan input format <i>log message</i>	59
Gambar 4. 12 ELECTRA dengan input format <i>Event Id</i>	60
Gambar 4. 13 ELECTRA dengan input format <i>log message</i>	60

DAFTAR TABEL

Tabel 2.1 Perbandingan penelitian.....	10
Tabel 3. 1 Spesifikasi laptop yang digunakan.....	32
Tabel 3. 2 Spesifikasi Kaggle <i>runtime</i>	32
Tabel 3. 3 Pustaka yang digunakan.....	33
Tabel 4. 1 Perbandingan Waktu Eksekusi Query Pandas dan SQLite.....	49
Tabel 4. 2 Hasil metrik evaluasi ketiga model.....	53
Tabel 4. 3 Hasil prediksi dan nilai <i>ground truth</i>	61
Tabel 4. 4 Pengujian kinerja waktu saat integrasi web pada ketiga model .	71

DAFTAR SIMBOL

Q	=	Vektor <i>Query</i>
K^T	=	<i>Key Transpose</i>
$\sqrt{d_k}$	=	Dimensi Vektor
V	=	Vektor <i>Value</i>
θ_D	=	Parameter <i>Discriminator</i>
θ_G	=	Parameter <i>Generator</i>
y_{ij}	=	Label sebenarnya untuk sampel ke-i pada kelas ke-j (<i>one-hot encoded</i>)
\hat{y}_{ij}	=	Probabilitas prediksi model untuk sampel ke-i pada kelas ke-j