

DAFTAR ISI

Lembar Pengesahan.....	i
PERNYATAAN BEBAS PLAGIASI.....	ii
INTISARI.....	iii
ABSTRACT.....	iv
DAFTAR ISI.....	v
DAFTAR GAMBAR.....	viii
DAFTAR TABEL.....	ix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	4
1.4 Batasan Masalah.....	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA.....	6
BAB III LANDASAN TEORI.....	15
3.1 <i>Clickbait</i>	15
3.2 <i>Few-shot Learning</i>	15
3.3 BERT	17
3.4 Sentence-BERT.....	19
3.5 IndoBERT	20
3.6 SetFit	21
3.7 Matriks Evaluasi.....	23
3.7.1 <i>Confusion matrix</i>	24
3.7.2 Skor Akurasi	25
3.7.3 <i>Precision</i>	25
3.7.4 <i>Recall</i>	25
3.7.5 F-score.....	26
BAB IV METODOLOGI PENELITIAN.....	27
4.1 Alat dan Bahan.....	27
4.1.1 Alat.....	27
4.1.2 Bahan	27
4.2 Prosedur Penelitian.....	27
4.2.1 Deskripsi Penelitian	28

4.2.2	Akuisisi Data.....	29
4.2.3	Arsitektur model	30
4.2.3.1	ST <i>Fine tuning</i>	33
4.2.3.2	Arsitektur IndoBERT.....	37
4.2.3.3	<i>Pooling Layer</i> dan <i>Cosine-Similarity</i>	40
4.2.3.4	<i>Classification head</i>	41
4.2.4	Pra-pemrosesan Data.....	42
4.2.4.1	<i>Case-folding</i>	42
4.2.4.2	<i>Punctuation Removal</i>	43
4.2.4.3	<i>Lemmatization</i>	43
4.2.4.4	Menghilangkan <i>Stopword</i>	44
4.2.4.5	Mengubah Struktur <i>Dataset</i>	44
4.2.4.6	Pemisahan data.....	44
4.2.4.7	<i>Sampling</i> data.....	44
4.2.5	Inisialisasi Model	46
4.2.6	Melatih Model dengan <i>Training</i> set.....	46
4.2.7	Prosedur Evaluasi.....	47
4.2.8	Evaluasi Model	47
4.2.9	Komparasi Model.....	47
BAB V	IMPLEMENTASI.....	49
5.1	<i>Environment</i> Implementasi	50
5.2	Akuisisi Data.....	50
5.3	Pra-Pemrosesan Data	50
5.3.1	Pembuatan <i>Dataframe</i>	50
5.3.2	<i>Case-folding</i>	51
5.3.3	<i>Punctuation Removal</i>	52
5.3.4	<i>Stopword Removal</i>	52
5.3.5	<i>Lemmatization</i>	53
5.3.6	<i>Splitting</i> Data.....	54
5.3.7	<i>Data restructure</i> untuk SetFit	55
5.3.8	<i>Sampling</i> Data <i>Training</i>	55
5.3.9	<i>Tokenisasi</i> Data	56
5.3.10	<i>Data restructure</i> untuk <i>Fine-tuning</i> IndoBERT	57
5.4	Inisialisasi Model	59
5.4.1	Inisialisasi Model SetFit.....	59
5.4.2	Inisialisasi Model IndoBERT.....	60

5.5	Pelatihan Model	62
5.5.1	Pelatihan Model SetFit.....	62
5.5.2	Pelatihan Model IndoBERT.....	65
5.6	Evaluasi dan Komparasi Model	67
5.6.1	Evaluasi SetFit	67
5.6.2	Evaluasi IndoBERT	69
BAB VI HASIL DAN PEMBAHASAN.....		72
6.1	Pengumpulan Data	72
6.2	Pra-pemrosesan Data.....	72
6.2.1	<i>Case-folding</i>	72
6.2.2	<i>Punctuation Removal</i>	73
6.2.3	<i>Stopword Removal</i>	73
6.2.4	<i>Lemmatization</i>	74
6.2.5	<i>Splitting Data</i>	75
6.2.6	<i>Data restructure</i> untuk SetFit	75
6.2.7	<i>Sampling Training Data</i>	75
6.2.8	Pra-pemrosesan data untuk <i>fine-tuning</i> IndoBERT	77
6.3	Hasil <i>Hyperparameter-Tuning</i> SetFit model	77
6.4	Inferensi pada data Testing	78
6.5	Komparasi dengan Fine-tuned IndoBERT	81
6.6	Komparasi durasi <i>training</i> SetFit dan IndoBERT.....	83
BAB VII PENUTUP.....		85
7.1	Kesimpulan	85
7.2	Saran.....	85
DAFTAR PUSTAKA		87

DAFTAR GAMBAR

Gambar 3.1 Diagram Cara Kerja <i>Few-shot Learning</i> (Rohit, 2022)	16
Gambar 3.2 Representasi <i>Input</i> BERT (Devlin dkk. (2018))	19
Gambar 3.3 Diagram Alur Kerja SetFit (Moshe, 2022)	22
Gambar 3.4 Contoh <i>Confusion matrix</i>	24
Gambar 4.1 - Alur Penelitian	30
Gambar 4.2 - Sampel <i>Dataset</i> CLICK-ID	31
Gambar 4.3 - Diagram Blok <i>Fine-tuning</i> Dan <i>Training</i> SetFit	32
Gambar 4.4 - Arsitektur Model SetFit	33
Gambar 4.5 - <i>Pseudo Code</i> Pembentukan <i>Triplet</i> Pasangan Kalimat	35
Gambar 4.6 - Diagram Ilustrasi Pembentukan <i>Triplet</i>	38
Gambar 4.7 - Arsitektur IndoBERT	38
Gambar 4.8 - Diagram <i>Training</i> dan Inferensi <i>Classification head</i> SetFit	42
Gambar 5.1 - Pembuatan <i>Dataframe</i>	51
Gambar 5.2 - Proses <i>Case-folding</i>	51
Gambar 5.3 - Proses <i>Punctuation Removal</i>	52
Gambar 5.4 - Proses <i>Stopword Removal</i>	53
Gambar 5.5 - Proses <i>Lemmatization</i>	54
Gambar 5.6 - Proses Data <i>Splitting</i>	54
Gambar 5.7 - Proses <i>Data restructure</i> untuk SetFit	55
Gambar 5.8 - Proses <i>Sampling</i> Data <i>Training</i> SetFit	56
Gambar 5.9 - Proses <i>Tokenisasi</i> Data <i>Training</i> IndoBERT	57
Gambar 5.10 - Proses <i>Restructure</i> Data <i>Training</i> IndoBERT	58
Gambar 5.11 - Proses Inisialisasi Model SetFit	60
Gambar 5.12 - Proses Insialisasi Model IndoBERT	61
Gambar 5.13 - Proses <i>Training</i> Model SetFit	64
Gambar 5.14 - Proses <i>Training</i> Model IndoBERT	66
Gambar 5.15 - Proses Evaluasi SetFit	69
Gambar 5.16 - Proses Evaluasi IndoBERT	71
Gambar 6.1 - <i>Dataframe</i> yang Akan Digunakan	72
Gambar 6.2 - Hasil <i>Splitting</i> Data	75
Gambar 6.3 - Struktur Data untuk SetFit	75
Gambar 6.4 - Perbandingan Performa Model SetFit Pada Tiap Nilai Sampel	80
Gambar 6.5 - Grafik Perbandingan Performa SetFit dan IndoBERT	83
Gambar 6.6 - Grafik Perbandingan Durasi <i>Training</i> SetFit dan IndoBERT	84

DAFTAR TABEL

Tabel 2.1 - Perbandingan Metode <i>Few-shot Learning</i> Untuk <i>Task NLP</i>	12
Tabel 4.1 - Spesifikasi <i>Dataset</i> yang Digunakan	30
Tabel 6.1 - Tabel Perbandingan Sebelum dan Setelah Dilakukan <i>Case-folding</i>	73
Tabel 6.2 - Tabel Perbandingan Sebelum dan Setelah Dilakukan <i>Punctuation Removal</i> . ..	73
Tabel 6.3 - Tabel Perbandingan Sebelum dan Setelah Dilakukan <i>Stopword Removal</i>	74
Tabel 6.4 - Tabel Perbandingan Sebelum dan Setelah Dilakukan <i>Lemmatization</i>	74
Tabel 6.5 - Tabel Hasil <i>Sampling</i> 8 Data <i>Training</i>	76
Tabel 6.6 - Tabel Perbandingan Hasil Pra-Pemrosesan Data Untuk IndoBERT	77
Tabel 6.7 - Tabel Hasil <i>Hyperparameter-Tuning</i> SetFit Model.....	78
Tabel 6.8 - Tabel Hasil Inferensi SetFit pada Data Testing.....	79
Tabel 6.9 – Tabel Standar Deviasi Inferensi SetFit pada Data Testing	79
Tabel 6.10 - Tabel Hasil <i>Hyperparameter-Tuning</i> IndoBERT Model	81
Tabel 6.11 - Tabel Perbandingan Hasil Inferensi SetFit dan IndoBERT pada Data Testing ..	82
Tabel 6.12 - Tabel Standar Deviasi Hasil Inferensi IndoBERT	82