



## DAFTAR ISI

HALAMAN PENGESAHAN.....	i
PERNYATAAN.....	ii
KATA PENGANTAR .....	iii
DAFTAR ISI.....	v
DAFTAR TABEL.....	viii
DAFTAR GAMBAR .....	ix
INTISARI.....	xii
<i>ABSTRACT</i> .....	xiii
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah .....	3
1.4 Tujuan Penelitian .....	3
1.5 Manfaat Penelitian .....	3
1.6 Metodologi Penelitian.....	4
1.7 Sistematika Penulisan .....	5
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>7</b>
2.1 Evolusi Metode Analisis Kebijakan Privasi .....	7
2.2 Penggunaan Dataset OPP-115 untuk Tugas Klasifikasi Kebijakan Privasi.....	8
2.3 Pendekatan Metode <i>Hybrid</i> dalam Klasifikasi Kebijakan Privasi.....	9
<b>BAB III LANDASAN TEORI.....</b>	<b>28</b>
3.1 Kebijakan Privasi .....	28
3.2 Online Privacy Policies Corpus (OPP-115).....	29
3.3 Machine Learning .....	30
3.3.1 Supervised Learning.....	31
3.3.2 Unsupervised Learning .....	32



3.3.3 Reinforcement Learning.....	32
3.4 Support Vector Machine .....	32
3.5 Term Frequency-Inverse Document Frequency .....	34
3.6 Transformer.....	35
3.6.1 Mekanisme Self Attention.....	37
3.6.2 Positional Encoding.....	38
3.6.3 Struktur Encoder-Decoder.....	38
3.6.4 Jaringan Feed-Forward.....	39
3.6.5 Training dan Optimization .....	39
3.7 BERT .....	40
3.7.1 Masked Language Modelling.....	41
3.7.2 Next Sentence Prediction .....	42
3.8 Pengujian.....	42
<b>BAB IV METODE PENELITIAN .....</b>	<b>44</b>
4.1 Deskripsi Umum Penelitian .....	44
4.2 Tahapan Penelitian.....	44
4.3 Perancangan Sistem .....	44
4.3.1 Dataset.....	45
4.3.2 Preprocessing .....	47
4.3.3 Representasi Teks.....	47
4.3.4 Pembagian Data.....	49
4.3.5 <i>Training</i> .....	49
4.3.6 Pengujian dan Evaluasi Model.....	51
4.3.7 Analisis Performa Model .....	51
<b>BAB V IMPLEMENTASI.....</b>	<b>52</b>
5.1 Pemahaman Data .....	52
5.2 Pemrosesan Data .....	55
5.3 TF-IDF .....	60
5.4 BERT .....	61
5.5 SVM.....	67



5.6 Pengujian dan Evaluasi .....	72
BAB VI HASIL DAN PEMBAHASAN .....	74
6.1 Hasil Pemahaman dan Pra-pemrosesan Data.....	74
6.2 Hasil Model <i>Baseline</i> TF-IDF-SVM .....	76
6.3 Hasil Model <i>Hybrid</i> BERT-SVM .....	79
6.4 Pembahasan.....	86
BAB VII SIMPULAN DAN SARAN .....	91
DAFTAR PUSTAKA .....	93



## DAFTAR TABEL

Tabel 2. 1 Perbandingan Penelitian Tugas Klasifikasi dengan Model <i>Hybrid</i> pada Beberapa Domain.....	10
Tabel 2. 2 Penelitian Terdahulu .....	11
Tabel 3.1 Jenis Tipe Machine Learning .....	31
Tabel 3.2 Confusion Matrix 2-kelas.....	43
Tabel 4.1 Pembagian Data Anotasi berdasarkan <i>Threshold</i> .....	46
Tabel 6.1 Tabel Distribusi Anotasi pada Kebijakan Privasi .....	75
Tabel 6.2 Hasil Pengujian Hyperparameter pada Setiap Kernel SVM .....	82
Tabel 6.4 Hasil Precision (P), Recall (R), F1, dan Support (S) pada Data dengan Threshold Anotasi 0.5 .....	86
Tabel 6.5 Hasil Precision (P), Recall (R), F1, dan Support (S) pada Data dengan Threshold Anotasi 0.75 .....	87
Tabel 6.6 Hasil Precision (P), Recall (R), F1, dan Support (S) pada Data dengan Threshold Anotasi 1.0 .....	87
Tabel 6.7 Nilai <i>Overall Accuracy</i> Tiap Model yang Diuji.....	88
Tabel 6.8 Komparasi Performa Model dengan Model Terdahulu .....	89



## DAFTAR GAMBAR

Gambar 3.1 Berbagai kategori Machine Learning dan contoh algoritmanya (Jain & Kumar, 2022) .....	30
Gambar 3.2 Kemungkinan-kemungkinan hyperplane (Gandhi, 2018).....	33
Gambar 3.3 Kernel SVM (Gandhi, 2018).....	33
Gambar 3.4 Arsitektur Transformer (Vaswani <i>et al.</i> , 2017).....	36
Gambar 3.5 Mekanisme <i>Attention</i> pada Transformer (Vaswani <i>et al.</i> , 2017) .....	36
Gambar 4.1 Diagram Alir Tahapan Penelitian.....	44
Gambar 4.2 Perancangan Sistem.....	45
Gambar 4. 3 Kategori Data Practices .....	46
Gambar 4. 4 Contoh data kebijakan dari Atlantic Monthly Group.....	47
Gambar 4. 5 Tahapan <i>Preprocessing</i> .....	47
Gambar 4.6 Representasi teks menggunakan BERT Embeddings .....	49
Gambar 4. 7 Penerapan Linear Support Vector Machine (SVM).....	50
Gambar 5.1 File Kebijakan .....	51
Gambar 5.2 Fungsi untuk Memuat Kebijakan Privasi.....	52
Gambar 5.3 File Anotasi .....	52
Gambar 5.4 Konten File Anotasi .....	52
Gambar 5.5 Kode Program untuk Mendefinisikan Kolom Anotasi.....	52
Gambar 5.6 Fungsi untuk Memuat Anotasi Kebijakan.....	53
Gambar 5.7 Visualisasi Distribusi Label Dataset OPP-115.....	53
Gambar 5.8 Fungsi untuk Mengkombinasikan Data Kebijakan dan Anotasi.....	54
Gambar 5.9 Fungsi untuk Menghapus Tag HTML.....	55
Gambar 5.10 Fungsi untuk Tokenisasi Kalimat.....	56
Gambar 5.11 Fungsi untuk Menghilangkan <i>Stopwords</i> .....	57
Gambar 5.12 Fungsi untuk <i>Stemming</i> Kata dalam Kalimat.....	57
Gambar 5.13 Fungsi Utama untuk Memproses Data ( <i>Preprocessing</i> ) .....	58
Gambar 5.14 <i>Pipeline Tuning Hyperparameter</i> TF-IDF dan SVM.....	59
Gambar 5.15 Vektorisasi TF-IDF dengan <i>Hyperparameter</i> Terbaik .....	60
Gambar 5.16 Label Mapping .....	61



Gambar 5.17 <i>Data Splitting</i> .....	61
Gambar 5.18 Impor Library BERT .....	61
Gambar 5.19 Kelas TextDataset .....	62
Gambar 5.20 Proses <i>Finetuning</i> .....	64
Gambar 5.21 Visualisasi <i>Loss</i> dari <i>Finetuning</i> .....	64
Gambar 5.22 Proses Menghasilkan Embeddings .....	65
Gambar 5.23 Training SVM untuk Model <i>Baseline</i> .....	66
Gambar 5.24 Grid Parameter untuk <i>Hyperparameter Tuning</i> Model Hybrid .....	67
Gambar 5.25 Inisialisasi Variabel sebelum Cross Validation .....	67
Gambar 5.26 Fungsi Cross Validation untuk Model Hybrid .....	69
Gambar 5.27 Tuning Hyperparameter SVM untuk Model Hybrid .....	70
Gambar 5.28 <i>Re-training</i> SVM dengan Data Gabungan Train dan Val .....	70
Gambar 5.29 Evaluasi Model menggunakan <i>classification_report</i> .....	71
Gambar 5.30 Metrics Micro-Averaged .....	71
Gambar 5.31 Confusion Matrix .....	72
Gambar 6.1 Jumlah Anotasi dan Kebijakan .....	74
Gambar 6.2 Jumlah Anotasi dengan <i>Threshold</i> Konsolidasi yang Berbeda .....	74
Gambar 6.3 Plot Distribusi Label .....	74
Gambar 6.4 Kolom Dataframe sebelum dilakukan <i>Preprocessing</i> .....	75
Gambar 6.5 Kolom Dataframe setelah dilakukan <i>Preprocessing</i> .....	76
Gambar 6.6 Alur Klasifikasi dengan Model <i>Baseline</i> .....	76
Gambar 6.7 Hasil <i>Data Splitting</i> .....	76
Gambar 6.8 Pipeline GridSearchCV untuk TF-IDF dan SVM .....	76
Gambar 6.9 Hasil Vektorisasi TF-IDF .....	77
Gambar 6.10 Klasifikasi menggunakan SVM dan <i>Hyperparameter</i> Terbaik .....	77
Gambar 6.11 Hasil Klasifikasi menggunakan Model <i>Baseline</i> .....	77
Gambar 6.12 Nilai Micro-Average Model <i>Baseline</i> .....	77
Gambar 6.13 Confusion Matrix Model <i>Baseline</i> .....	79
Gambar 6.14 Alur Klasifikasi dengan Model <i>Hybrid</i> .....	80
Gambar 6.15 Hasil <i>Label Mapping</i> .....	80
Gambar 6.16 Data Split untuk <i>Finetuning</i> .....	80



Gambar 6.17 Visualisasi <i>Loss</i> per Step.....	80
Gambar 6.18 Visualisasi <i>Loss</i> per Epoch.....	81
Gambar 6.19 Struktur Model BERT sudah di Finetune.....	81
Gambar 6.20 Hasil Hyperparameter Tuning Model <i>Hybrid</i> .....	84
Gambar 6.21 Hasil Klasifikasi menggunakan Model <i>Hybrid</i> .....	84
Gambar 6.22 Nilai Micro-Average Model <i>Hybrid</i> .....	84
Gambar 6.23 Confusion Matrix Model <i>Hybrid</i> .....	85