

## DAFTAR GAMBAR

Gambar 2.1 Grafik fungsi <i>Sigmoid</i> .....	15
Gambar 2.2 <i>Hyperplane</i> .....	17
Gambar 2.3 <i>Hyperplane Optimal</i> .....	17
Gambar 3.1 <i>Flowchart</i> langkah-langkah pengerjaan.....	23
Gambar 3.2 Kode untuk mengekstrak data dari <i>file</i> HTML situs <i>turnbackhoax.id</i> ..	24
Gambar 3.3 Kode untuk membaca <i>file</i> <i>.csv</i> .....	25
Gambar 3.4 Informasi <i>df_turnbackhoax</i> .....	26
Gambar 3.5 Informasi <i>df_detik</i> .....	27
Gambar 3.6 Informasi <i>df_kompas</i> .....	27
Gambar 3.7 Kode yang menampilkan jumlah data kosong pada <i>dataframe</i> <i>turnbackhoax</i> .....	30
Gambar 3.8 Kode untuk menghapus baris dengan data hilang pada kolom “narasi” pada <i>dataframe</i> <i>turnbackhoax</i> .....	30
Gambar 3.9 Kode untuk menghapus baris duplikat pada <i>dataframe</i> <i>turnbackhoax</i> ..	30
Gambar 3.10 Kode untuk menghapus kolom-kolom yang tidak dibutuhkan .....	30
Gambar 3.11 Kode untuk mencari teks dengan awalan tertentu dan mengganti teks tersebut dengan teks setelah kata “narasi” .....	30
Gambar 3.12 Kode untuk melakukan seleksi data.....	32
Gambar 3.13 Kode untuk melakukan <i>case folding</i> .....	35
Gambar 3.14 Kode untuk melakukan tokenisasi .....	36
Gambar 3.15 Kode untuk filtering <i>URL removal</i> .....	37
Gambar 3.16 Kode untuk melakukan <i>filtering stopwords</i> .....	37

Gambar 3.17 Kode untuk melakukan <i>filtering punctuational removal</i> .....	38
Gambar 3.18 Fungsi untuk <i>filtering number removal</i> .....	39
Gambar 3.19 Kode untuk melakukan <i>filtering number removal</i> .....	40
Gambar 3.20 Kode untuk melakukan <i>stemming</i> .....	41
Gambar 3.21 Kode untuk melakukan pembagian data .....	42
Gambar 3.22 Kode untuk melakukan transformasi TF-IDF .....	42
Gambar 3.23 Kode untuk membuat dan melatih model <i>Logistic Regression</i> .....	48
Gambar 3.24 Kode untuk memprediksi data dengan model yang telah dibuat .....	48
Gambar 3.25 Kode untuk membuat dan melatih model SVM .....	51
Gambar 4.1 Bar chart perbandingan jumlah data berita hoaks dengan berita faktual .....	55
Gambar 4.2 Deskripsi <i>dataset</i> faktual .....	56
Gambar 4.3 Deskripsi <i>dataset</i> hoaks .....	56
Gambar 4.4 <i>Bar chart</i> distribusi frekuensi kata berita hoaks .....	57
Gambar 4.5 <i>Bar chart</i> distribusi frekuensi kata berita faktual .....	57
Gambar 4.6 <i>Word cloud</i> distribusi 20 kata-kata teratas pada berita hoaks .....	58
Gambar 4.7 <i>Word cloud</i> distribusi 20 kata-kata teratas pada berita faktual .....	58
Gambar 4.8 Media-media sosial yang banyak digunakan masyarakat Indonesia .....	59
Gambar 4.9 Kode untuk mengecek sumber berita hoaks dari <i>dataset</i> turnbackhoax. ....	59
Gambar 4.10 Persentase penyebaran berita hoaks pada media sosial dari <i>dataset</i> turnbackhoax. ....	60
Gambar 4.14 Data uji .....	63
Gambar 4.15 <i>Confusion matrix</i> hasil prediksi pada data uji .....	64

Gambar 4.16 Metrik evaluasi hasil prediksi pada data uji ..... 64

Gambar 4.17 *Line chart* perbandingan metrik evaluasi pada *random\_state* yang berbeda-beda..... 67

Gambar 4.18 *Line chart* perbandingan metrik evaluasi pada rasio pembagian yang berbeda-beda..... 69



## DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i> .....	19
Tabel 2.2 Penelitian terdahulu .....	21
Tabel 3.1 <i>DataFrame</i> df_turnbackhoax.....	25
Tabel 3.2 <i>DataFrame</i> df_detik .....	26
Tabel 3.3 <i>DataFrame</i> df_kompas.....	26
Tabel 3.4 Contoh data hilang .....	29
Tabel 3.5 Contoh konten salah .....	29
Tabel 3.6 <i>Dataset</i> hasil penggabungan judul dan narasi menjadi "judul_dan_narasi" pada <i>dataframe</i> turnbackhoax .....	31
Tabel 3.7 <i>Dataset</i> berita faktual hasil seleksi data .....	33
Tabel 3.8 <i>Dataset</i> setelah pemberian label.....	34
Tabel 3.9 <i>Dataset</i> setelah tahap <i>case folding</i> .....	35
Tabel 3.10 <i>Dataset</i> setelah tahap tokenisasi.....	36
Tabel 3.11 <i>Dataset</i> setelah tahap <i>filtering stopwords</i> .....	38
Tabel 3.12 Data setelah proses <i>filtering punctuational removal</i> .....	39
Tabel 3.13 Data setelah tahap <i>filtering number removal</i> .....	40
Tabel 3.14 Data setelah proses <i>stemming</i> .....	41
Tabel 3.15 Tiga dokumen yang digunakan untuk perhitungan pada <i>Microsoft Excel</i> ..	43
Tabel 3.16 Menghitung frekuensi <i>term</i> pada setiap dokumen .....	43
Tabel 3.17 Menghitung nilai TF untuk setiap <i>term</i> .....	44
Tabel 3.18 Menghitung nilai IDF untuk setiap <i>term</i> .....	45
Tabel 3.19 Menghitung nilai TF-IDF untuk setiap <i>term</i> pada setiap dokumen .....	46

Tabel 3.20 Tabel TF-IDF setelah dilakukan normalisasi.....	47
Tabel 3.21 Nilai <i>intercept</i> dan koefisien model <i>Logistic Regression</i> .....	49
Tabel 3.22 Tabel TF-IDF ditambah kolom "probabilitas klasifikasi awal" .....	50
Tabel 3.23 Nilai <i>bias</i> dan <i>weight</i> model SVM .....	51
Tabel 3.24 Hasil klasifikasi dengan metode SVM .....	53
Tabel 4.1 <i>Dataset</i> setelah tahap <i>text preprocessing</i> .....	55
Tabel 4.2 <i>Confusion matrix</i> hasil pengujian dengan <i>random_state</i> yang berbeda-beda	65
Tabel 4.3 Metrik evaluasi hasil pengujian dengan <i>random_state</i> yang berbeda-beda ..	65
Tabel 4.4 <i>Confusion matrix</i> hasil pengujian dengan skenario pembagian data yang berbeda-beda.....	66
Tabel 4.5 Metrik evaluasi hasil pengujian dengan skenario pembagian data yang berbeda-beda.....	66

## DAFTAR ISI

<b>HALAMAN JUDUL</b>	
<b>PERNYATAAN DAN PERSETUJUAN UNGGAH TUGAS AKHIR</b>	
<b>PERSETUJUAN DOSEN PEMBIMBING TUGAS AKHIR</b>	
<b>PERSETUJUAN TIM PENGUJI TUGAS AKHIR</b>	
<b>ABSTRAK.....</b>	<b>v</b>
<b>ABSTRACT .....</b>	<b>vi</b>
<b>KATA PENGANTAR.....</b>	<b>vii</b>
<b>DAFTAR GAMBAR.....</b>	<b>ix</b>
<b>DAFTAR TABEL .....</b>	<b>xii</b>
<b>DAFTAR ISI.....</b>	<b>xiv</b>
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang Masalah .....	1
1.2 Rumusan Masalah .....	2
1.3 Tujuan Penelitian.....	2
1.4 Batasan Masalah.....	2
1.5 Manfaat Penelitian.....	3
1.5.1 Manfaat bagi Akademis.....	3
1.5.2 Manfaat bagi Masyarakat .....	3
1.6 Sistematika Penulisan .....	4
<b>BAB II LANDASAN TEORI.....</b>	<b>6</b>
2.1 Berita Hoaks.....	6
2.2 <i>Machine Learning</i> .....	8
2.3 <i>Text Mining</i> .....	10
2.4 <i>Supervised Learning</i> .....	14
2.5 <i>Logistic Regression</i> .....	14
2.6 <i>Support Vector Machine (SVM)</i> .....	16
2.7 Evaluasi.....	19
2.8 Penelitian Terdahulu.....	21
<b>BAB III METODOLOGI .....</b>	<b>23</b>
3.1 Pengumpulan Data .....	24

3.2	Pembersihan Data.....	27
3.3	Seleksi Data .....	31
3.4	Pelabelan data .....	33
3.5	<i>Text Preprocessing</i> .....	34
3.5.1	<i>Case folding</i> .....	34
3.5.2	Tokenisasi ( <i>Tokenizing</i> ) .....	35
3.5.3	<i>Filtering</i> .....	36
3.5.4	<i>Stemming</i> .....	40
3.5.5	Pembagian <i>Dataset</i> .....	41
3.5.6	<i>Transformation</i> (TF-IDF).....	42
3.5.7	Metode <i>Logistic Regression</i> .....	48
3.5.8	Metode SVM .....	51
<b>BAB IV</b>	<b>HASIL DAN PEMBAHASAN.....</b>	<b>54</b>
4.1	Analisis Data Eksploratif (EDA) .....	54
4.2	Hasil.....	63
4.3	Pembahasan.....	66
4.3.1	Pengujian Model .....	67
4.3.2	Pengujian Model dengan <i>random_state</i> yang Berbeda-beda .....	67
4.3.3	Pengujian Model dengan Skenario Pembagian Data yang Berbeda-beda.....	69
<b>BAB V</b>	<b>KESIMPULAN DAN SARAN .....</b>	<b>72</b>
5.1	Kesimpulan .....	72
5.2	Saran .....	72
	<b>DAFTAR PUSTAKA.....</b>	<b>73</b>
	<b>LAMPIRAN A .....</b>	<b>A-1</b>