

DAFTAR ISI

SKRIPSI.....	i
PERSETUJUAN DOSEN PEMBIMBING TUGAS AKHIR	iii
PERSETUJUAN TIM PENGUJI TUGAS AKHIR.....	iv
ABSTRAK	v
<i>ABSTRACT</i>	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL.....	xiii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian.....	3
1.4 Batasan Masalah.....	4
1.5 Manfaat Penelitian.....	4
1.5.1 Bagi Universitas.....	4
1.5.2 Bagi Penulis	5
1.6 Sistematika Penulisan.....	5
BAB II LANDASAN TEORI.....	7
2.1 TikTok.....	7
2.2 <i>Python</i>	7
2.3 <i>Machine Learning</i>	8
2.3.1 <i>Supervised Learning</i>	8
2.3.2 <i>Unsupervised learning</i>	9
2.3.3 <i>Reinforcement learning</i>	9
2.4 <i>Text mining</i>	10
2.4.1 <i>Text Preprocessing</i>	10

2.4.2	<i>TF – IDF</i>	14
2.5	<i>Confusion Matrix</i>	15
2.6	Klasifikasi.....	17
2.7	KNN	19
2.7.1	Jarak Euclidean	20
2.8	Universitas Pelita Harapan	21
2.9	Penelitian Terdahulu.....	22
BAB III METODOLOGI		23
3.1	Tahapan Penelitian	23
3.2	Prosedur Pengumpulan data	24
3.3	Seleksi Data.....	26
3.4	Pemberian Label.....	26
3.5	Text Preprocessing	27
3.5.1	Case Normalization.....	27
3.5.2	Pembersihan	28
3.5.3	Tokenization.....	28
3.5.4	Stopword Removal.....	29
3.5.5	<i>Lemmatization</i>	29
3.5.6	<i>Stemming</i>	30
3.6	Transformation (TF&IDF)	31
3.7	Pemisahan <i>dataset</i>	38
3.8	Metode analisis.....	38
3.9	Evaluasi model	39
3.10	Mengukur <i>accuracy</i> dengan KNN.....	39
BAB IV HASIL DAN PEMBAHASAN		43
4.1	Hasil Pengumpulan Data	43
4.1.1	Hasil data label.....	44
4.1.2	Hasil Text Preprocessing.....	44
4.2	Hasil Penelitian.....	48
4.3	Pembahasan	49

BAB V KESIMPULAN DAN SARAN.....	54
5.1 Kesimpulan.....	54
5.2 Saran.....	54
DAFTAR PUSTAKA	56
LAMPIRAN A : <i>SOURCE CODE PROGRAM</i>	1



DAFTAR GAMBAR

Gambar 2. 1 Contoh KNN - Data Baru Sebelum dan Sesudah.....	20
Gambar 2. 2 Rumus Jarak Euclidean	21
Gambar 3. 1 Prosedur Penelitian.....	23
Gambar 3. 2 Import Library	40
Gambar 3. 3 Fungsi untuk <i>preprocessing text</i>	41
Gambar 3. 4 <i>Hyperparameter</i> KNN.....	42
Gambar 4. 1 <i>Data Info</i> komentar <i>TikTok</i>	43
Gambar 4. 2 Total jumlah <i>positive</i> dan <i>negative</i>	44
Gambar 4. 3 Hasil <i>Case Folding</i> Sumber : Notebook peneliti (2023).....	45
Gambar 4. 4 Hasil Pembersihan.....	45
Gambar 4. 5 Hasil <i>Tokenization</i>	46
Gambar 4. 6 Hasil <i>Stopword Removal</i>	46
Gambar 4. 7 Hasil Lemmatization	47
Gambar 4. 8 Hasil <i>Stemming</i>	47
Gambar 4. 9 Hasil <i>cross validation</i>	48
Gambar 4. 10 <i>Confusion Matrix</i>	49
Gambar 4. 11 Hasil perhitungan <i>Confusion Matrix</i>	51
Gambar 4. 12 data validasi.....	52

DAFTAR TABEL

Tabel 2. 1 Sebelum dan sesudah <i>case folding</i>	11
Tabel 2. 2 Sebelum dan sesudah <i>normalization</i>	11
Table 2. 3 Sebelum dan sesudah pembersihan.....	12
Table 2. 4 Sebelum dan sesudah <i>tokenization</i>	12
Table 2. 5 Sebelum dan sesudah <i>stopword removal</i>	13
Table 2. 6 Sebelum dan sesudah <i>lemmatization</i>	13
Table 2. 7 Sebelum dan sesudah <i>stemming</i>	14
Table 2. 8 Rumus <i>confussion matrix</i>	16
Table 2. 9 Kelebihan dan kekurangan metode klasifikasi.....	18
Table 2. 10 Penelitian terdahulu.....	22
Tabel 3. 1 Contoh <i>dataset</i>	24
Tabel 3. 2 Tabel Seleksi Data	26
Tabel 3. 3 Tabel Pemberian Label	26
Tabel 3. 4 Tabel setelah <i>Case Normalization</i>	27
Tabel 3. 5 Tabel setelah Pembersihan.....	28
Tabel 3. 6 Tabel setelah <i>Tokenization</i>	28
Tabel 3. 7 Tabel setelah <i>Stopword Removal</i>	29
Tabel 3. 8 Tabel <i>Lemmatization</i>	30
Tabel 3. 9 Tabel <i>Stemming</i>	30
Tabel 3. 10 Tabel <i>TF - IDF</i>	31
Tabel 3. 11 Jumlah kata pada kalimat.....	31
Tabel 3. 12 Perhitungan Hasil TF	33
Tabel 3. 13 Hasil Perhitungan IDF	34
Tabel 3. 14 Hasil Perhitungan TF-IDF data <i>testing</i> 1	36
Tabel 3. 15 Hasil Perhitungan Data <i>Testing</i> 2	37
Tabel 3. 16 Hasil perhitungan Euclidean	38