

DAFTAR ISI

KATA PENGANTAR.....	i
ABSTRAK	iv
ABSTRACT	v
DAFTAR ISI	vi
DAFTAR GAMBAR	ix
DAFTAR TABEL	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	5
1.5.1 Bagi Peneliti.....	6
1.5.2 Bagi Universitas	6
1.6 Kerangka Pemikiran	7
1.7 Sistematika Penulisan.....	7
BAB II KAJIAN LITERATUR	9
2.1 Tinjauan Pustaka	9
2.2 Landasan Teori.....	18
2.2.1 Isu <i>Overclaim Skincare</i>	18
2.2.2 <i>Natural Language Processing (NLP)</i>	19
2.2.3 BERT	21
2.2.4 <i>Confusion matrix</i>	21
2.2.5 <i>Cross-Industry Standard Process for Data Mining (CRISP-DM)</i>	23
2.2.6 Media Sosial X (Twitter).....	25
BAB III METODOLOGI	28
3.1 Metode yang Digunakan.....	28
3.2 <i>Bussines Understanding</i>	28

3.2.1	Mengidentifikasi Masalah.....	28
3.2.2	Penilaian Situasi dan Tantangan.....	29
3.2.3	Rencana Penelitian	30
3.3	<i>Data Understanding</i>	31
3.3.1	<i>Scraping Data</i>	31
3.3.2	Analisis Data	32
3.4	<i>Data preparation</i>	33
3.4.1	Validasi Data	34
3.4.2	Pembersihan Data.....	35
3.4.3	Normalisasi Data	38
3.4.4	<i>Split Data</i>	39
3.4.5	<i>Labeling Data</i>	40
3.4.6	<i>Oversampling Data</i>	41
3.5	<i>Modeling</i>	42
3.5.1	Inisialisasi IndoBERT	42
3.5.2	Tokenisasi Data	43
3.5.3	<i>Fine Tuning Model</i>	44
3.6	<i>Evaluation</i>	45
3.6.1	Evaluasi Model.....	45
3.6.2	Analisis Sentimen Isu <i>Overclaim Skincare</i>	46
3.7	Bahan dan Peralatan yang Digunakan.....	47
BAB IV HASIL DAN PEMBAHASAN		21
4.1	Hasil dan Pembahasan Rumusan Masalah 1.....	21
4.1.1	Hasil <i>Business Understanding</i>	21
A.	Hasil Identifikasi Masalah	49
B.	Penilaian Situasi dan Tantangan.....	49
C.	Rencana Proyek.....	50
4.1.2	Hasil <i>Data Understanding</i>	52
A.	<i>Scraping Data</i>	52
B.	Analisis Data.....	53

4.1.3 Hasil <i>Data Preparation</i>	54
A. Validasi Data	55
B. Pembersihan Data.....	56
C. Normalisasi Data	58
D. <i>Split Data</i>	60
E. <i>Labeling Data</i>	61
F. <i>Oversampling Data</i>	64
4.1.4 <i>Modeling</i>	65
A. Inisialisasi IndoBERT.....	65
B. Tokenisasi Data	66
C. <i>Fine Tuning Model</i>	67
4.2 Hasil dan Pembahasan Rumusan Masalah 2.....	69
4.2.1 <i>Evaluation</i>	69
A. Hasil Evaluasi Model	69
B. Kesimpulan Hasil Evaluasi Model.....	86
4.2.2 Hasil Analisis Sentimen Isu <i>Overclaim Skincare</i>	88
A. Hasil Perbandingan Analisis Sentimen.....	88
BAB V PENUTUP	100
4.1 Kesimpulan	100
4.2 Saran.....	101
DAFTAR PUSTAKA.....	104

DAFTAR GAMBAR

Gambar 1 Kerangka pemikiran.....	7
Gambar 2 <i>Confusion matrix</i> [29].....	22
Gambar 3 Tahapan CRISP-DM [27].....	23
Gambar 4 Tahapan tahapan CRISP-DM	28
Gambar 5. Alur <i>Scraping</i> data.....	32
Gambar 6. Alur <i>Data preparation</i>	33
Gambar 7 Alur <i>Modeling</i>	42
Gambar 8 proses <i>scraping</i> data	52
Gambar 9 Hasil <i>scraping</i> data	53
Gambar 10 Pembersihan data karakter	56
Gambar 11 Pembersihan data emoji	57
Gambar 12 hasil pembersihan data <i>verified</i>	57
Gambar 13 hasil pembersihan data <i>verified</i>	58
Gambar 14 Kode hapus <i>stopword</i>	58
Gambar 15 Kode <i>Stemming</i>	59
Gambar 16 Hasil <i>stopword</i> dan <i>stemming</i> data <i>non verified</i>	59
Gambar 17 Hasil <i>stopword</i> dan <i>stemming</i> data <i>non verified</i>	59
Gambar 18 Kode <i>split</i> data.....	60
Gambar 19 Kode <i>labeling</i> data.....	61
Gambar 20 presentase <i>labeling</i> data <i>verified</i>	62
Gambar 21 presentase <i>labeling</i> data <i>verified</i>	63
Gambar 22 Hasil <i>labeling</i> data <i>verified</i>	63
Gambar 23 Hasil <i>labeling</i> data <i>non verified</i>	63
Gambar 24 Kode <i>Oversampling</i> data.....	64
Gambar 25 Distribusi <i>labeling</i> data <i>non verified</i>	65
Gambar 26 Distribusi <i>labeling</i> data <i>verified</i>	65
Gambar 27 Inisialisasi <i>tokenizer</i> IndoBERT	66

Gambar 28 Konfigurasi kelas label	66
Gambar 29 Inisialisasi model IndoBERT	66
Gambar 30. Pengaturan <i>TrainingArguments</i> untuk pelatihan model IndoBERT	68
Gambar 31. Inisialisasi dan pelatihan model menggunakan <i>Trainer</i>	68
Gambar 32. <i>Confusion matrix</i> konfigurasi LR 1e-5 epoch 1	70
Gambar 33. Grafik <i>Eval</i> dan <i>Train loss</i> LR 1e-5 epoch 1.....	71
Gambar 34. <i>Confusion matrix</i> konfigurasi LR 1e-5 epoch 2	72
Gambar 35. Grafik <i>Eval</i> dan <i>Train loss</i> LR 1e-5 epoch 2.....	73
Gambar 36. <i>Confusion matrix</i> konfigurasi LR 1e-5 epoch 3	75
Gambar 37. Grafik <i>Eval</i> dan <i>Train loss</i> LR 1e-5 epoch 3.....	75
Gambar 38. <i>Confusion matrix</i> konfigurasi LR 1e-5 epoch 5	77
Gambar 39. Grafik <i>Eval</i> dan <i>Train loss</i> LR 1e-5 epoch 5.....	77
Gambar 40. <i>Confusion matrix</i> konfigurasi LR 2e-5 epoch 1	79
Gambar 41 Grafik <i>Eval</i> dan <i>Train loss</i> LR 2e-5 epoch 1.....	79
Gambar 42. <i>Confusion matrix</i> konfigurasi LR 2e-5 epoch 2	81
Gambar 43. Grafik <i>Eval</i> dan <i>Train loss</i> LR 2e-5 epoch 2.....	82
Gambar 44. <i>Confusion matrix</i> konfigurasi LR 2e-5 epoch 3	83
Gambar 45. Grafik <i>Eval</i> dan <i>Train loss</i> LR 2e-5 epoch 3.....	84
Gambar 46. <i>Confusion matrix</i> konfigurasi LR 2e-5 epoch 5	85
Gambar 47. Grafik <i>Eval</i> dan <i>Train loss</i> LR 2e-5 epoch 5.....	86
Gambar 48 Hasil distribusi sentimen positif data <i>verified</i>	89
Gambar 49 Hasil presentase sentimen data <i>verified</i>	90
Gambar 50 Hasil <i>wordcloud</i> sentimen positif data <i>verified</i>	91
Gambar 51 Hasil <i>wordcloud</i> sentimen negatif data <i>verified</i>	92
Gambar 52 Hasil distribusi sentimen positif data <i>non verified</i>	94
Gambar 53 Hasil presentase sentimen data <i>non verified</i>	94
Gambar 54 Hasil <i>wordcloud</i> sentimen positif data <i>non verified</i>	95
Gambar 55 Hasil <i>wordcloud</i> sentimen negatif data <i>non verified</i>	96

DAFTAR TABEL

Tabel 2. 1 <i>State of The Art</i>	9
Tabel 3. 1 pembersihan data karakter	36
Tabel 3. 2 pembersihan karakter emoji dan <i>Unicode</i>	37
Tabel 3. 3 pemberdihan karakter tunggal.....	38
Tabel 3. 4 standarisasi huruf dan menghapus <i>newline</i>	38
Tabel 3. 5 Aturan klasifikasi sentimen	40
Tabel 3. 6. Contoh Tokenisasi data	44
Tabel 4. 1 Rencana proyek	51
Tabel 4. 2 Jumlah data setelah dipisah.....	55
Tabel 4. 3 Jumlah distribusi label	62
Tabel 4. 4 Contoh hasil tokenisasi	67
Tabel 4. 5 <i>Classification report</i> LR 1e-5 epoch 1.....	70
Tabel 4. 6 <i>Classification report</i> LR 1e-5 epoch 2.....	72
Tabel 4. 7 <i>Classification report</i> LR 1e-5 epoch 3.....	74
Tabel 4. 8 <i>Classification report</i> LR 1e-5 epoch 5.....	76
Tabel 4. 9 <i>Classification report</i> LR 2e-5 epoch 1.....	78
Tabel 4. 10 <i>Classification report</i> LR 2e-5 epoch 2.....	80
Tabel 4. 11 <i>Classification report</i> LR 2e-5 epoch 3.....	83
Tabel 4. 12 <i>Classification report</i> LR 2e-5 epoch 5.....	85
Tabel 4. 13 perbandingan percobaan performa model.....	87
Tabel 4. 14 Hasil kalimat dari sentimen positif isu <i>overclaim</i> produk kecantikan.....	91
Tabel 4. 15 Hasil kalimat dari sentimen negatif isu <i>overclaim</i> produk kecantikan	93
Tabel 4. 16 Hasil kalimat dari sentimen positif isu <i>overclaim</i> produk kecantikan.....	96
Tabel 4. 17 Hasil kalimat dari sentimen negatif isu <i>overclaim</i> produk kecantikan	97