

DAFTAR ISI

LEMBAR JUDUL	
LEMBAR PERSETUJUAN.....	i
LEMBAR PENGESAHAN	ii
ABSTRAK	iii
<i>ABSTRACT</i>	iv
KATA PENGANTAR.....	v
DAFTAR ISI	vii
DAFTAR GAMBAR	x
DAFTAR TABEL.....	xii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Tujuan.....	4
1.4. Manfaat	4
1.5. Batasan Masalah.....	5
1.6. Kerangka Pemikiran.....	5
1.7. Sistematika Penulisan	6
BAB II KAJIAN LITERATUR.....	8
2.1. Tinjauan Pustaka	8
2.1.1. <i>State of The Art</i>	8
2.2. Landasan Teori	13
2.2.1. Analisis Sentimen	13
2.2.2. Pernikahan.....	13
2.2.3. Aplikasi X (Twitter)	15
2.2.4. <i>Natural Language Processing (NLP)</i>	15
2.2.5. <i>Text Preprocessing</i>	16
2.2.6. <i>Tokenization</i>	16
2.2.7. <i>Transformers</i>	17
2.2.8. <i>Bidirectional Encoder Representations from Transformers (BERT)</i> 18	
2.2.9. <i>Confusion Matrix</i>	20

2.2.10. <i>Cross Industry Standard Process for Data Mining (CRISP-DM)</i>	21
BAB III METODOLOGI PENELITIAN.....	24
3.1. <i>Business Understanding</i>	25
3.1.1. Menganalisis Pemahaman Tujuan Bisnis Penelitian.....	25
3.1.2. Megidentifikasi Kebutuhan dan Tantangan Penelitian.....	25
3.1.3. Menyusun Rencana Proyek Penelitian.....	26
3.2. <i>Data Understanding</i>	26
3.2.1. Pengumpulan Data Penelitian	26
3.2.2. Menelaah Data	28
3.3. <i>Data Preparation</i>	28
3.3.1. Seleksi <i>Dataset</i> Penelitian.....	28
3.3.2. Pembersihan <i>Dataset</i>	28
3.3.3. Pembagian <i>Dataset</i>	33
3.3.4. Pelabelan <i>Dataset</i>	34
3.3.5. Penanganan Data Tidak Seimbang.....	35
3.3.6. Pembagian Data Untuk Pelatihan	36
3.3.7. Tokenisasi Data	36
3.3.8. Transformasi Data Sebelum Pelatihan	38
3.4. <i>Modeling</i>	39
3.4.1. <i>Fine-tuning</i> Model Pada Pelatihan.....	39
3.5. <i>Evaluation</i>	41
3.5.1. Evaluasi Hasil Kinerja Model.....	42
3.5.2. Analisis Sentimen dan Perbandingan Hasil Analisis	42
BAB IV HASIL DAN PEMBAHASAN.....	44
4. 1. Hasil Serta Pembahasan Rumusan Masalah Penelitian 1	44
4. 2. 1. Hasil Proses <i>Business Understanding</i>	44
4. 2. 2. Hasil <i>Data Understanding</i>	47
4. 2. 3. Hasil Pengolahan <i>Data Preparation</i>	50
4. 2. 4. <i>Modeling</i>	59
4. 2. Hasil Serta Pembahasan Rumusan Masalah Penelitian 2	68
4. 2. 1. <i>Evaluation</i>	68
4. 3. Hasil Serta Pembahasan Rumusan Masalah Penelitian 3	82

4.3.1. Hasil Analisis Sentimen Publik pada Media Sosial X Mengenai “ <i>Marriage is Scary</i> ”	82
BAB VPENUTUP	91
5.1. Simpulan	91
5.2. Saran.....	92
DAFTAR PUSTAKA	93



DAFTAR GAMBAR

Gambar 1.1 Kerangka Pemikiran.....	6
Gambar 2.1 Arsitektur <i>Transformers</i> [38].....	17
Gambar 2.2 <i>Pre-Training</i> dan <i>Fine-Tuning</i> BERT[15].....	19
Gambar 2.3 Representasi <i>Input</i> dari BERT [15]	19
Gambar 2.4 Fase CRISP-DM[44].....	21
Gambar 3.1 Diagram Alur Tahapan Penelitian CRISP-DM	24
Gambar 3.2 Alur Pengambilan Data	27
Gambar 3.3 Alur Pembersihan Data.....	29
Gambar 4.1 Proses <i>Scraping Dataset</i>	47
Gambar 4.2 Hasil <i>Scraping</i> Data Indonesia.....	49
Gambar 4.3 Hasil <i>Scraping</i> Data Amerika Serikat	49
Gambar 4.4 Seleksi Data Berdasarkan Kata Kunci	50
Gambar 4.5 Seleksi Data Berdasarkan Lokasi.....	50
Gambar 4.6 Pembersihan data	51
Gambar 4.7 Hasil Pembersihan <i>Dataset</i> Indonesia.....	52
Gambar 4.8 Hasil Pembersihan <i>Dataset</i> Amerika Serikat	52
Gambar 4.9 Pemilihan Kolom yang Relevan.....	53
Gambar 4.10 Pembagian <i>Dataset</i> untuk Pelabelan.....	53
Gambar 4.11 Pelabelan Data.....	55
Gambar 4.12 Hasil Pelabelan <i>Dataset</i> Indonesia.....	55
Gambar 4.13 Hasil Pelabelan <i>Dataset</i> Amerika Serikat	55
Gambar 4.14 Pembagian <i>Dataset</i> untuk Pelatihan.....	57
Gambar 4.15 Hasil Auto Tokenisasi Data	58
Gambar 4.16 Konversi Label Menjadi Numerik.....	59
Gambar 4.17 Hasil Instalasi Model.....	60
Gambar 4.18 Hasil <i>Training</i> Eksperimen Pertama Indonesia.....	61
Gambar 4.19 Grafik <i>Loss</i> Eksperimen Pertama Indonesia	61
Gambar 4.20 Hasil <i>Training</i> Data Eksperimen Kedua Indonesia.....	62
Gambar 4.21 Grafik <i>Loss</i> Percobaan Kedua Indonesia	62
Gambar 4.22 Hasil <i>Training</i> Data Eksperimen Ketiga Indonesia.....	63
Gambar 4.23 Grafik <i>Loss</i> Percobaan Ketiga Indonesia	64

Gambar 4.24 Hasil Pelatihan Pertama Data Amerika Serikat.....	64
Gambar 4.25 Grafik <i>Loss</i> Pelatihan Pertama Amerika Serikat.....	65
Gambar 4.26 Hasil Pelatihan Kedua Amerika Serikat.....	65
Gambar 4.27 Grafik <i>Loss</i> Pelatihan Kedua Amerika Serikat.....	66
Gambar 4.28 Hasil Pelatihan Ketiga Amerika Serikat.....	66
Gambar 4.29 Grafik <i>Loss</i> Pelatihan Ketiga Amerika Serikat	67
Gambar 4.30 Hasil Pelatihan Keempat Amerika Serikat	67
Gambar 4.31 Grafik <i>Loss</i> Pelatihan Keempat Amerika Serikat.....	68
Gambar 4.32 <i>Confusion Matrix</i> Pelatihan Pertama Indonesia	69
Gambar 4.33 <i>Confusion Matrix</i> Percobaan Kedua Indonesia.....	70
Gambar 4.34 <i>Confusion Matrix</i> Pelatihan Ketiga Indonesia	72
Gambar 4.35 <i>Confusion Matrix</i> Pelatihan Pertama Amerika Serikat	73
Gambar 4.36 <i>Confusion Matrix</i> Pelatihan Kedua Amerika Serikat	75
Gambar 4.37 <i>Confusion Matrix</i> Percobaan Ketiga Amrika Serikat.....	76
Gambar 4.38 <i>Confusion Matrix</i> Pelatihan Keempat Amerika Serikat	77
Gambar 4.39 Distribusi Panjang Teks Data Indonesia.....	81
Gambar 4.40 Distribusi Panjang Teks Data Amerika Serikat	81
Gambar 4.41 Distribusi Sentimen Data Indonesia.....	83
Gambar 4.42 <i>WordCloud</i> Sentimen Negatif Indonesia	84
Gambar 4.43 <i>WordCloud</i> Sentimen Positif Indonesia.....	85
Gambar 4.44 Distribusi Sentimen Amerika Serikat.....	87
Gambar 4.45 <i>WordCloud</i> Sentimen Negatif Amerika Serikat	88
Gambar 4.46 <i>WordCloud</i> Sentimen Positif Amerika Serikat	89

DAFTAR TABEL

Tabel 2. 1 <i>State of The Art</i>	8
Tabel 2. 2 Tokenisasi Teks	16
Tabel 3. 1 Proses <i>Lowercase</i>	30
Tabel 3. 2 Proses Menghapus URL.....	30
Tabel 3. 3 Proses Menghapus <i>Mention</i>	31
Tabel 3. 4 Proses Menghapus <i>Hashtag</i>	31
Tabel 3. 5 Proses Hapus Simbol.....	32
Tabel 3. 6 Proses Menghapus Spasi Berlebih	32
Tabel 3. 7 Proses Menghapus Angka	33
Tabel 3. 8 Proses Menghapus Tanda Baca	33
Tabel 3. 9 Proses Tokenisasi	36
Tabel 3. 10 Proses Penambahan Spesial Token.....	37
Tabel 3. 11 Proses <i>Truncation</i> dan <i>Padding</i>	37
Tabel 3. 12 Proses <i>Attention Mask</i>	38
Tabel 3. 13 Eksperimen Pelatihan Data Indonesia.....	41
Tabel 3. 14 Eksperimen Pelatihan Data Amerika Serikat	41
Tabel 4. 1 Rancangan Proyek Penelitian.....	46
Tabel 4. 2 Jumlah <i>Dataset</i> yang Terkumpul.....	48
Tabel 4. 3 Jumlah <i>Dataset</i> Setelah Seleksi.....	51
Tabel 4. 4 Jumlah Data Setelah Pembersihan	52
Tabel 4. 5 Jumlah Data <i>Subset</i> Penelitian	54
Tabel 4. 6 Jumlah Sentimen <i>Dataset</i> Indonesia	56
Tabel 4. 7 Jumlah Sentimen <i>Dataset</i> Amerika Serikat.....	56
Tabel 4. 8 Jumlah Data Indonesia Setelah Augmentasi	57
Tabel 4. 9 Jumlah Data <i>Train, Val, Test</i> di Dataset Indonesia	57
Tabel 4. 10 Jumlah Data <i>Train, Val, Test</i> di Dataset Amerika Serikat.....	58
Tabel 4. 11 <i>Classification Report</i> Pelatihan Pertama Indonesia	69
Tabel 4. 12 <i>Classification Report</i> Pelatihan Kedua Indonesia.....	70
Tabel 4. 13 <i>Classification Report</i> Pelatihan Ketiga Indonesia.....	71
Tabel 4. 14 Hasil Eksperimen Data Indonesia	72
Tabel 4. 15 <i>Classification Report</i> Pelatihan Pertama Amerika Serikat.....	73

Tabel 4. 16 <i>Classification Report</i> Pelatihan Kedua Amerika Serikat	74
Tabel 4. 17 <i>Classification Report</i> Pelatihan Ketiga Amerika Serikat	75
Tabel 4. 18 <i>Classification Report</i> Pelatihan Keempat Amerika Serikat	76
Tabel 4. 19 Hasil Eksperimen Data Amerika Serikat.....	77
Tabel 4. 20 Perbedaan Karakteristik Penulisan Indonesia dan Amerika Serikat ..	79
Tabel 4. 21 Kalimat Sentimen Negatif Data Indonesia.....	84
Tabel 4. 22 Kalimat Sentimen Positif Data Indonesia	85
Tabel 4. 23 Kalimat Sentimen Negatif Data Amerika Serikat	88
Tabel 4. 24 Kalimat Sentimen Positif Data Amerika Serikat.....	89

