

ABSTRAK

Perkembangan teknologi di era digital mengubah paradigma penyiaran televisi dari analog ke digital. Penelitian ini mengeksplorasi analisis sentimen masyarakat terhadap migrasi televisi menggunakan model *Bidirectional Long Short-Term Memory* (BiLSTM). Dataset berisi 9.597 data tweet dari *Twitter*, diperoleh dengan *library SNScrape* melalui proses *scraping*. Model BiLSTM digunakan untuk mengekstraksi sentimen positif, negatif, dan netral dari data tweet. Dataset ini diperoleh melalui proses sampling *Slovin* dengan tingkat error 2%, untuk memastikan representativitas sampel yang optimal dengan total 3.098 data tweet, penelitian ini menerapkan rasio split data 80:20, dengan 2.478 data untuk pelatihan model dan 620 data untuk pengujian. Teknik validasi silang (*cross-validation*) 10 kali memberikan reliabilitas hasil yang baik. Hasil eksperimen mencapai rata-rata akurasi 91.36%, menunjukkan kinerja yang impresif. Analisis sentimen dilakukan pada sejumlah besar data dengan distribusi yang signifikan, terdiri dari 1.698 sentimen positif, 5.208 sentimen negatif, dan 2.691 sentimen netral, memberikan wawasan yang tentang variasi sentimen dalam dataset. Selain itu, distribusi hasil prediksi sentimen teridentifikasi kata-kata kunci yang mempengaruhi sentimen negatif, seperti "stb," "analog," dan "beli," mencerminkan ketidakpuasan terhadap kebijakan migrasi. Di sisi lain, kata-kata positif seperti "digital" dan "siar" mencerminkan pandangan positif terhadap teknologi digital. Hasil eksperimen menunjukkan bahwa pendekatan ini mampu melakukan klasifikasi sentimen pada data teks dengan akurasi tinggi.

Kata kunci : Analisis Sentimen, *Twitter*, Perubahan Televisi Analog ke Digital, *Bidirectional Long Short Term Memory*

ABSTRACT

Technological developments in the digital era changed the paradigm of television broadcasting from analog to digital. This research explores the analysis of public sentiment towards television migration using a model Bidirectional Long Short-Term Memory (BiLSTM). The dataset contains 9,597 tweet data from Twitter, obtained with library SNScrape through the process scraping. The BiLSTM model is used to extract positive, negative and neutral sentiments from tweet data. This dataset was obtained through a sampling process Slovin with an error rate of 2%, to ensure optimal sample representativeness with a total of 3,098 tweet data, this study applied a data split ratio of 80:20, with 2,478 data for model training and 620 data for testing. Cross validation technique (cross-validation) 10 times provides good reliability of results. The experimental results achieved an average accuracy of 91.36%, showing impressive performance. Sentiment analysis was performed on a large amount of data with a significant distribution, consisting of 1,698 positive sentiments, 5,208 negative sentiments, and 2,691 neutral sentiments, providing insight into the variations in sentiment within the dataset. In addition, the distribution of sentiment prediction results identified key words that influence negative sentiment, such as "stb," "analog," and "buy," reflecting dissatisfaction with migration policies. On the other hand, positive words such as "digital" and "broadcast" reflect a positive view of digital technology. Experimental results show that this approach is able to perform sentiment classification on text data with high accuracy.

Keywords: *Sentiment Analysis, Twitter, Change from Analog to Digital Television, Bidirectional Long Short Term Memory*

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERNYATAAN ORISINALITAS.....	iii
ABSTRAK.....	iv
<i>ABSTRACT</i>	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR	x
DAFTAR TABEL	xii
DAFTAR KODE PROGRAM.....	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.2.1 Permasalahan	5
1.2.2 Pertanyaan Penelitian.....	5
1.3 Tujuan dan Manfaat.....	5
1.3.1 Tujuan Penelitian	5
1.3.2 Manfaat Penelitian	6
1.4 Batasan Masalah.....	6
1.5 Sistematika Penulisan.....	7
BAB II TEORI PENUNJANG	8
2.1 Analisis Sentimen.....	8
2.2 <i>Text mining</i>	9
2.3 <i>Twitter</i>	9
2.4 <i>Crawling Data</i>	9
2.5 <i>Preprocessing Data</i>	10
2.6 Esktrasi Fitur	11
2.7 <i>Tokenizer</i>	11
2.8 <i>Pad Sequences</i>	12
2.9 <i>Long Short Term Memory</i>	13

2.10	<i>Bidirectional Long Short Term Memory</i>	14
2.11	<i>Confusion matrix</i>	16
2.10	<i>K-Fold Cross Validation</i>	18
2.11	Penelitian Terkait.....	19
BAB III METODE PENELITIAN.....		25
3.1	Metode Penelitian CRISP-DM.....	25
3.1.1	<i>Business Understanding</i>	26
3.1.2	<i>Data Understanding</i>	27
3.1.3	<i>Data Preparation</i>	28
3.1.4	<i>Modeling</i>	35
3.1.5	<i>Evaluation</i>	39
3.1.6	<i>Deployment</i>	40
BAB IV HASIL DAN PEMBAHASAN		41
4.1	<i>Business Understanding</i>	41
4.2	<i>Data Understanding</i>	41
4.1.1	Pengumpulan Data	42
4.1.2	Filterisasi Data	45
4.1.3	Pelabelan Data.....	47
4.1.4	<i>Sampling Slovin</i>	49
4.3	<i>Data Preparation</i>	51
4.2.1	<i>Preprocessing Data</i>	51
4.2.2	Visualisasi Data.....	63
4.2.3	Ekstrasi Fitur	68
4.2.4	<i>Split Data</i>	72
4.4	<i>Modelling</i>	74
4.3.1	Klasifikasi <i>Bidirectional Long Short Term Memory (BiLSTM)</i>	75
4.5	<i>Evaluation</i>	77
4.4.1	<i>Confusion Matrix</i>	77
4.6	Hasil Skenario Pengujian	83
4.5.1	Skenario Pengujian Pertama	83
4.5.2	Skenario Pengujian Kedua	93
4.7	<i>Deployment</i>	101

BAB V PENUTUP.....	109
5.1 Kesimpulan.....	109
5.2 Saran.....	110
REFERENSI	111
LAMPIRAN.....	115
DAFTAR RIWAYAT HIDUP.....	119

DAFTAR GAMBAR

Gambar 1. 1 Presentase Penduduk yang Menonton Televisi di Indonesia	2
Gambar 2. 1 <i>Long Short Term Memory Model</i>	13
Gambar 2. 2 <i>Bidirectional Long Short Term Memory Model</i> [21]	14
Gambar 2. 3 Ilustrasi Proses <i>10-fold Cross Validation</i> [25]	19
Gambar 3. 1 Alur Metode CRISP-DM	25
Gambar 3. 2 <i>Flowchart</i> Alur Penelitian	26
Gambar 3. 3 <i>Flowchart</i> Ekstrasi Fitur Tokenisasi dan <i>Padding</i>	33
Gambar 3. 4 <i>Flowchart</i> Klasifikasi BiLSTM	36
Gambar 3. 5 Diagram Alir Skenario Penelitian	39
Gambar 4. 1 Proses Filterisasi Data	46
Gambar 4. 2 Proses Filterisasi Data	46
Gambar 4. 3 Proses Filterisasi Data	47
Gambar 4. 4 Grafik Perbandingan Kelas Sentimen	49
Gambar 4. 5 Proses Split Data	50
Gambar 4. 6 Proses Split Data	50
Gambar 4. 7 Proses Split Data	50
Gambar 4. 8 <i>Word Cloud</i> Semua Sentimen	65
Gambar 4. 9 <i>Word Cloud</i> Sentimen Positif	66
Gambar 4. 10 <i>Word Cloud</i> Sentimen Negatif	66
Gambar 4. 11 <i>Word Cloud</i> Sentimen Netral	67
Gambar 4. 12 Implementasi Ekstrasi Fitur	70
Gambar 4. 13 <i>Output Confusion Matrix 1</i>	78
Gambar 4. 14 <i>Output Confusion Matrix 2</i>	78
Gambar 4. 15 <i>Output Confusion Matrix 3</i>	79
Gambar 4. 16 <i>Output Confusion Matrix 4</i>	79
Gambar 4. 17 <i>Output Confusion Matrix 5</i>	80
Gambar 4. 18 <i>Output Confusion Matrix 6</i>	80
Gambar 4. 19 <i>Output Confusion Matrix 7</i>	81
Gambar 4. 20 <i>Output Confusion Matrix 8</i>	81
Gambar 4. 21 <i>Output Confusion Matrix 9</i>	82
Gambar 4. 22 <i>Output Confusion Matrix 10</i>	82

Gambar 4. 23 <i>Model Sequential BiLSTM</i>	84
Gambar 4. 24 Diagram Batang Jumlah Kelas Sentimen.....	99
Gambar 4. 25 Diagram Alir <i>Deployment Streamlit</i>	101
Gambar 4. 26 Hasil <i>Deployment</i> Prediksi Kalimat Baru	103
Gambar 4. 27 Hasil Prediksi Negatif <i>Deployment</i> Streamlit	107
Gambar 4. 28 Hasil Prediksi Positif <i>Deployment</i> Streamlit.....	108
Gambar 4. 29 Hasil Prediksi Netral <i>Deployment</i> Streamlit	108

DAFTAR TABEL

Tabel 2. 1 Tabel <i>Confusion matrix</i> [23]	16
Tabel 2. 2 Penelitian Terkait	22
Tabel 4. 1 Jumlah Dataset Setiap Kata Kunci	44
Tabel 4. 2 Contoh Data Tweet	45
Tabel 4. 3 Label Data <i>Tweet</i>	47
Tabel 4. 4 Hasil <i>Preprocessing Cleaning</i>	53
Tabel 4. 5 Hasil <i>Preprocessing Case Folding</i>	54
Tabel 4. 6 Hasil <i>Preprocessing Tokenization</i>	56
Tabel 4. 7 Hasil <i>Preprocessing Slangword</i>	58
Tabel 4. 8 Hasil <i>Preprocessing Stopword Removal</i>	60
Tabel 4. 9 Hasil <i>Preprocessing Stemming</i>	62
Tabel 4. 10 <i>Preprocessing Final</i>	63
Tabel 4. 11 Hasil <i>Output</i> Pembagian Data	74
Tabel 4. 12 Hasil Pelatihan Iterasi <i>K-fold Cross Validation</i> 1	84
Tabel 4. 13 Hasil Pelatihan Iterasi <i>K-fold Cross Validation</i> 2	85
Tabel 4. 14 Hasil Pelatihan Iterasi <i>K-fold Cross Validation</i> 3	85
Tabel 4. 15 Hasil Pelatihan Iterasi <i>K-fold Cross Validation</i> 4	85
Tabel 4. 16 Hasil Pelatihan Iterasi <i>K-fold Cross Validation</i> 5	86
Tabel 4. 17 Hasil Pelatihan Iterasi <i>K-fold Cross Validation</i> 6	86
Tabel 4. 18 Hasil Pelatihan Iterasi <i>K-fold Cross Validation</i> 7	87
Tabel 4. 19 Hasil Pelatihan Iterasi <i>K-fold Cross Validation</i> 8	87
Tabel 4. 20 Hasil Pelatihan Iterasi <i>K-fold Cross Validation</i> 9	88
Tabel 4. 21 Hasil Pelatihan Iterasi <i>K-fold Cross Validation</i> 10	88
Tabel 4. 22 Hasil Evaluasi <i>Confusion Matrix</i> 1	89
Tabel 4. 23 Hasil Evaluasi <i>Confusion Matrix</i> 2	89
Tabel 4. 24 Hasil Evaluasi <i>Confusion Matrix</i> 3	90
Tabel 4. 25 Hasil Evaluasi <i>Confusion Matrix</i> 4	90
Tabel 4. 26 Hasil Evaluasi <i>Confusion Matrix</i> 5	91
Tabel 4. 27 Hasil Evaluasi <i>Confusion Matrix</i> 6	91
Tabel 4. 28 Hasil Evaluasi <i>Confusion Matrix</i> 7	91
Tabel 4. 29 Hasil Evaluasi <i>Confusion Matrix</i> 8	92

Tabel 4. 30 Hasil Evaluasi <i>Confusion Matrix</i> 9	92
Tabel 4. 31 Hasil Evaluasi <i>Confusion Matrix</i> 10	92
Tabel 4. 32 Hasil Prediksi Model BiLSTM	95
Tabel 4. 33 5 Kata Teratas Kelas Negatif	100
Tabel 4. 34 5 Kata Teratas Kelas Positif.....	100

DAFTAR KODE PROGRAM

Kode Program 4. 1 Implementasi program pengumpulan data.....	43
Kode Program 4. 2 Implementasi Program <i>Cleaning</i>	52
Kode Program 4. 3 Implementasi Program <i>Case Folding</i>	54
Kode Program 4. 4 Implementasi Program <i>Tokenization</i>	55
Kode Program 4. 5 Implementasi Program <i>Slangword</i>	57
Kode Program 4. 6 Implementasi Program <i>Stopword Removal</i>	59
Kode Program 4. 7 Implementasi Program <i>Stemming</i>	61
Kode Program 4. 8 <i>World Cloud Sentiment</i>	64
Kode Program 4. 9 Implementasi Ekstrasi Fitur.....	69
Kode Program 4. 10 Ekstrasi Fitur.....	71
Kode Program 4. 11 Split Data	73
Kode Program 4. 12 Klasifikasi Model BiLSTM	76
Kode Program 4. 13 <i>Confusion Matrix</i>	77
Kode Program 4. 14 Prediksi Sentimen	94
Kode Program 4. 15 Pengambilan Kata Sering Muncul Setiap Kelas Sentimen..	99
Kode Program 4. 16 <i>Deployment</i> Model BiLSTM	102
Kode Program 4. 17 <i>Deploy Streamlit</i>	106