

**IMPLEMENTASI ALGORITMA BIDIRECTIONAL LONG SHORT-TERM
MEMORY (BI-LSTM) UNTUK ANALISIS SENTIMEN REVIEW
(FEEDBACK) PELANGGAN**



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

Casey Tjiptadjaja
00000045957

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2024**

**IMPLEMENTASI ALGORITMA BIDIRECTIONAL LONG SHORT-TERM
MEMORY (BI-LSTM) UNTUK ANALISIS SENTIMEN REVIEW
(FEEDBACK) PELANGGAN**



Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

Casey Tjiptadjaja

00000045957

UMN

**UNIVERSITAS
MULTIMEDIA
NUSANTARA**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA**

TANGERANG

2024

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Casey Tjiptadjaja
NIM : 00000045957
Program Studi : Informatika

Menyatakan dengan sesungguhnya bahwa Skripsi saya yang berjudul:
**Implementasi Algoritma Bidirectional Long Short-Term Memory (Bi-LSTM)
untuk Analisis Sentimen Review (Feedback) Pelanggan**

merupakan hasil karya saya sendiri, bukan merupakan hasil plagiat, dan tidak pula dituliskan oleh orang lain; Semua sumber, baik yang dikutip maupun dirujuk, telah saya cantumkan dan nyatakan dengan benar pada bagian Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan karya ilmiah, saya bersedia menerima konsekuensi untuk dinyatakan TIDAK LULUS. Saya juga bersedia menanggung segala konsekuensi hukum yang berkaitan dengan tindak plagiarisme ini sebagai kesalahan saya pribadi dan bukan tanggung jawab Universitas Multimedia Nusantara.

Tangerang, 16 Mei 2024



(Casey Tjiptadjaja)

HALAMAN PENGESAHAN

Skripsi dengan judul

IMPLEMENTASI ALGORITMA BIDIRECTIONAL LONG SHORT-TERM MEMORY (BI-LSTM) UNTUK ANALISIS SENTIMEN REVIEW (FEEDBACK) PELANGGAN

oleh

Nama : Casey Tjiptadjaja
NIM : 00000045957
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Kamis, 30 Mei 2024
Pukul 15.00 s/s 17.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut

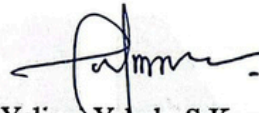
Ketua Sidang



(Alexander Waworuntu, S.Kom., M.T.I.) (Sy Yuliani Yakub, S.Kom., M.T. Ph.D.)

NIDN: 309068503

Penguji



NIDN: 0411037904

Pembimbing,



(Moeljono Widjaja, B.Sc., M.Sc., Ph.D.)

NIDN: 0311106903

Pjs. Ketua Program Studi Informatika,



(Dr. Eng. Niki Prastomo, S.T., M.Sc.)

NIDN: 0419128203

iii

**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Yang bertanda tangan di bawah ini:

Nama : Casey Tjiptadjaja
NIM : 00000045957
Program Studi : Informatika
Jenjang : S1
Jenis Karya : Skripsi

Menyatakan dengan sesungguhnya bahwa:

- Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya di repositori Knowledge Center, sehingga dapat diakses oleh Civitas Akademika/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial dan saya juga tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.
- Saya tidak bersedia karena dalam proses pengajuan untuk diterbitkan ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*)**.

Tangerang, 16 Mei 2024

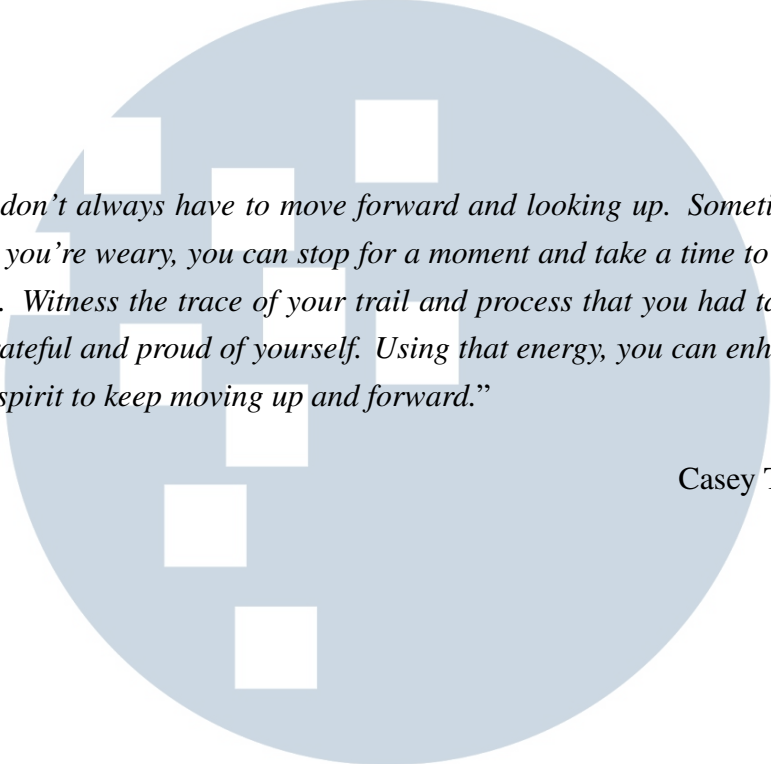
Yang menyatakan



Casey Tjiptadjaja

** Jika tidak bisa membuktikan LoA jurnal/HKI selama enam bulan ke depan, saya bersedia mengizinkan penuh karya ilmiah saya untuk diunggah ke KC UMN dan menjadi hak institusi UMN.

Halaman Persembahan / Motto



”You don’t always have to move forward and looking up. Sometimes, when you’re weary, you can stop for a moment and take a time to look down. Witness the trace of your trail and process that you had taken. Be grateful and proud of yourself. Using that energy, you can enhance your spirit to keep moving up and forward.”

Casey Tjiptadjaja

UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Implementasi Algoritma Bidirectional Long Short-Term Memory (Bi-LSTM) untuk Analisis Sentimen Review (Feedback) Pelanggan dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika dan Pjs. Ketua Program Studi Informatika Universitas Multimedia Nusantara. Informatika Universitas Multimedia Nusantara.
3. Bapak Moeljono Widjaja, B.Sc., M.Sc., Ph.D., sebagai Pembimbing pertama yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya tesis ini.
4. Orang Tua, teman-teman dan keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tesis ini.

Semoga skripsi ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 16 Mei 2024



Casey Tjiptadjaja

IMPLEMENTASI ALGORITMA BIDIRECTIONAL LONG SHORT-TERM MEMORY (BI-LSTM) UNTUK ANALISIS SENTIMEN REVIEW (FEEDBACK) PELANGGAN

Casey Tjiptadjaja

ABSTRAK

Pada generasi sekarang banyak perusahaan besar maupun kecil sedang bersaing dengan ketat menciptakan hal yang lebih baik dari yang terdapat di pasaran, salah satu cara mereka dapat mewujudkan hal tersebut adalah dengan cara penjaminan kualitas jasa atau barang yang lebih baik dari yang lainnya. Untuk meningkatkan kualitas layanan dari barang atau jasa yang ditawarkan, pihak penyedia harus melakukan penelitian mengenai *feedback* dari pengguna atau pelanggan mereka. Kebanyakan perusahaan menengah dan kecil, seperti UMKM, toko *online*, dan sebagainya, melakukan riset mengenai *feedback customer* secara manual, dengan melihat satu per satu *feedback* dari pelanggan, yang dimana sangat tidak efektif dan efisien jika *feedback customer* yang didapatkan sangat banyak. Maka dari itu, penelitian ini dibuat dengan maksud dan tujuan ingin membantu para perusahaan-perusahaan menengah dan kecil dalam analisis sentimen pelanggan mereka, serta tren dalam periode tertentu. Pada penelitian ini akan dilakukan penerapan algoritma *Bidirectional Long Short-Term Memory* (Bi-LSTM) untuk melakukan analisis sentimen pada *feedback* pelanggan. Penelitian ini juga melakukan perbandingan dengan metode *deep learning* lainnya sebagai pembanding dengan metode yang diusulkan, yaitu algoritma Uni-LSTM, GRU, CNN, dan *Simple-RNN*. Setelah dilakukan pengujian, hasil akurasi algoritma Uni-LSTM, Bi-LSTM, GRU, CNN, dan *Simple-RNN* secara berurutan adalah 52.2%, 92.4%, 52.2%, 90.9%, dan 49.7%. Lalu didapatkan bahwa algoritma Bi-LSTM dengan unit 64, *epoch training* sebanyak lima (5), dan rasio *training* dan *testing* sebesar 80:20, serta model yang mengimplementasikan aktivasi 'relu' (*Rectified Linear Unit*) mendapatkan hasil yang paling maksimal, yaitu mendapatkan akurasi 92.4%, yang dimana merupakan akurasi paling unggul jika dibandingkan dengan akurasi yang menggunakan model dan algoritma lainnya.

Kata kunci: Analisis Sentimen, Bi-LSTM, *Deep Learning*, *Review* Pelanggan

UNIVERSITAS
MULTIMEDIA
NUSANTARA

**IMPLEMENTATION OF BIDIRECTIONAL LONG SHORT-TERM MEMORY
(BI-LSTM) ALGORITHM FOR SENTIMENT ANALYSIS OF CUSTOMER
REVIEWS (FEEDBACK)**

Casey Tjiptadjaja

ABSTRACT

In the current generation, many large and small companies are competing fiercely to create better things than those on the market, one way they can realize this is by guaranteeing the quality of services or goods that are better than others. To improve the service quality of the goods or services offered, the provider must conduct research on feedback from their users or customers. Most medium and small companies, such as MSMEs, online stores, and so on, conduct research on customer feedback manually, by looking at one by one feedback from customers, which is very ineffective and efficient if a lot of customer feedback is obtained. Therefore, this research is made with the intent and purpose of wanting to help medium and small companies in analyzing their customer sentiment, as well as trends in a certain period. This research will apply the Bidirectional Long Short-Term Memory (Bi-LSTM) algorithm to perform sentiment analysis on customer feedback. This research also compares with other deep learning methods as a comparison with the proposed method, namely the Uni-LSTM, GRU, CNN, and Simple-RNN algorithms. After testing, the accuracy results of the Uni-LSTM, Bi-LSTM, GRU, CNN, and Simple-RNN algorithms are 52.2%, 92.4%, 52.2%, 90.9%, and 49.7%, respectively. Then it was found that the Bi-LSTM algorithm with 64 units, five (5) training epochs, and a training and testing ratio of 80:20, as well as a model that implements 'relu' activation (Rectified Linear Unit) obtained the maximum results, namely getting 92.4% accuracy, which is the most superior accuracy when compared to accuracy using other models and algorithms.

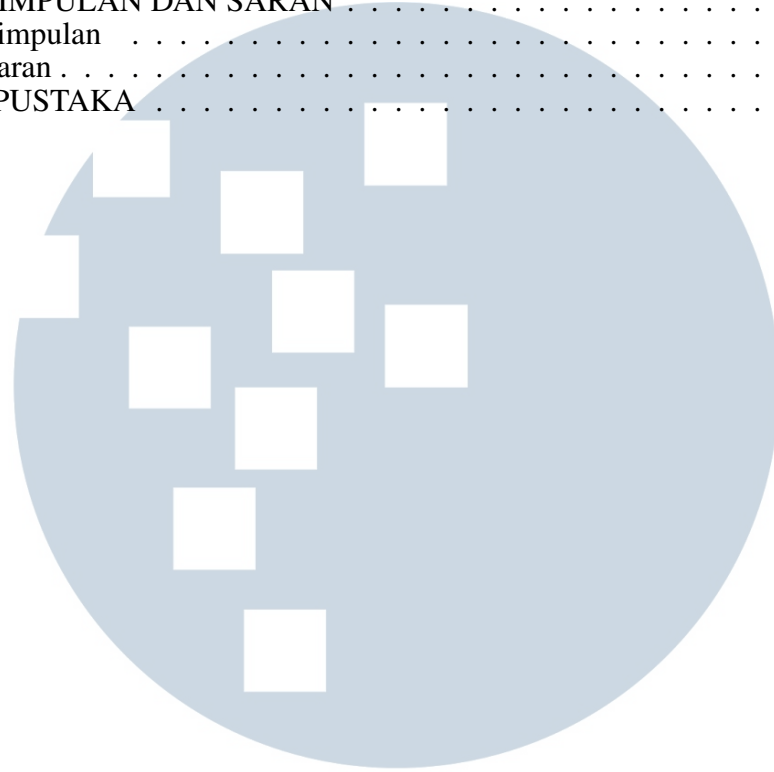
Keywords: *Bi-LSTM, Customer Review, Deep Learning, Sentiment Analysis*

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI ILMIAH	iv
HALAMAN PERSEMBAHAN/MOTO	v
KATA PENGANTAR	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xiii
DAFTAR KODE	xiv
DAFTAR LAMPIRAN	xv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Batasan Permasalahan	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	5
BAB 2 LANDASAN TEORI	7
2.1 Deep learning	7
2.2 Analisis Sentimen	7
2.3 Recurrent Neural Network	8
2.4 Long Short-Term Memory	9
2.5 Natural Language Processing	11
BAB 3 METODOLOGI PENELITIAN	13
3.1 Pengumpulan Data	13
3.2 Data Cleaning and Preprocessing	15
3.3 Encoding Labels dan Pembagian Dataset Train-Test	23
3.3.1 Tokenizing	25
3.3.2 Sequence Padding	26
3.4 Pembuatan Model	26
3.5 Model Training and Evaluation	27
3.6 Model and Algorithm Comparison	27
3.7 Pengujian Analisis Sentimen	27
3.8 Spesifikasi Sistem	28
3.8.1 Laptop	28
3.8.2 Jupyter Notebook	28
BAB 4 HASIL DAN DISKUSI	29
4.1 Hasil data sebelum dan setelah dilakukan data cleaning dan pre-processing	29
4.2 Implementasi model	31
4.3 Proses training dan testing	33
4.3.1 Proses training	34
4.3.2 Proses testing	35
4.4 Tahap Pengujian	38
4.4.1 Input kalimat secara manual	38

4.4.2	File tanpa tanggal	39
4.4.3	File dengan tanggal	44
BAB 5	SIMPULAN DAN SARAN	50
5.1	Simpulan	50
5.2	Saran	50
DAFTAR PUSTAKA	52



UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR GAMBAR

Gambar 2.1	Arsitektur RNN	9
Gambar 2.2	Arsitektur Unidirectional LSTM	10
Gambar 2.3	Arsitektur Bidirectional LSTM	11
Gambar 3.1	Tahapan-tahapan yang dilakukan pada penelitian	13
Gambar 3.2	Data yang digunakan	15
Gambar 3.3	Tahapan-tahapan yang dilakukan pada <i>data cleaning</i> dan <i>preprocessing</i>	15
Gambar 3.4	<i>Column</i> data yang akan digunakan	16
Gambar 3.5	Penamaan <i>column</i> yang akan digunakan	16
Gambar 3.6	Penghapusan data yang kosong	17
Gambar 3.7	Data yang digunakan	19
Gambar 3.8	Hasil penghapusan kata henti (<i>stopwords</i>) Bahasa Indonesia	19
Gambar 3.9	Hasil dari penghapusan data kosong setelah dilakukan penghapusan <i>stopwords</i>	20
Gambar 3.10	<i>Top 10</i> kata yang sering disebutkan di semua label	21
Gambar 3.11	<i>Top 10</i> kata yang sering disebutkan di label positif	21
Gambar 3.12	<i>Top 10</i> kata yang sering disebutkan di label negatif	22
Gambar 3.13	Hasil dari dilakukannya <i>stemming</i>	23
Gambar 3.14	Tahapan-tahapan yang dilakukan pada <i>encoding labels</i> dan pembagian <i>dataset train-test</i>	23
Gambar 4.1	Data asli	29
Gambar 4.2	Histogram data asli	29
Gambar 4.3	Data hasil <i>data cleaning</i> dan <i>pre-processing</i>	30
Gambar 4.4	Histogram data hasil <i>textitdata cleaning</i> dan <i>pre-processing</i>	30
Gambar 4.5	Hasil dari <i>training</i> model	33
Gambar 4.6	<i>Plot</i> dari hasil akurasi <i>training</i> dan <i>validation</i>	34
Gambar 4.7	<i>Plot</i> dari hasil nilai <i>loss training</i> dan <i>validation</i>	35
Gambar 4.8	Hasil akurasi data <i>testing</i>	35
Gambar 4.9	Hasil dari klasifikasi <i>confusion metrics</i>	36
Gambar 4.10	Hasil dari <i>confusion metrics</i>	37
Gambar 4.11	Contoh isi <i>array 'sentences'</i> pada kode pengujian	39
Gambar 4.12	Hasil dari pengujian dengan cara <i>input</i> kalimat secara manual	39
Gambar 4.13	Hasil dari upload data	40
Gambar 4.14	Hasil dari penghapusan kolom yang tidak digunakan	40
Gambar 4.15	Hasil dari pengubahan nama kolom	41
Gambar 4.16	Hasil dari data yang diproses	42
Gambar 4.17	Sebelum dimasukkan ke dalam model	42
Gambar 4.18	Setelah dimasukkan ke dalam model	42
Gambar 4.19	<i>Top 10</i> kata-kata pada label 'Positive'	43
Gambar 4.20	<i>Top 10</i> kata-kata pada label 'Negative'	43
Gambar 4.21	Hasil dari upload data	44
Gambar 4.22	<i>Cell</i> kode untuk menghilangkan kolom yang tidak digunakan	45
Gambar 4.23	<i>Cell</i> kode untuk mengubah nama kolom	45
Gambar 4.24	Hasil dari data yang diproses	46
Gambar 4.25	Hasil dari tren sentimen tiap bulannya	47
Gambar 4.26	Sebelum dimasukkan ke dalam model	47
Gambar 4.27	Setelah dimasukkan ke dalam model	48

Gambar 4.28 *Top* 10 kata-kata pada label 'Positive' 48
Gambar 4.29 *Top* 10 kata-kata pada label 'Negative' 49



UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR TABEL

Tabel 3.1	Hasil penghapusan <i>tag-tag</i> HTML, URL, dan karakter yang bukan alfabet	18
Tabel 3.2	Hasil perubahan semua kata menjadi huruf kecil (<i>lowercase</i>)	18
Tabel 3.3	Hasil penghapusan kata henti (<i>stopwords</i>)	19
Tabel 3.4	Hasil pemecahan kalimat menjadi per kata	20
Tabel 3.5	Hasil perubahan semua kata menjadi huruf kecil (<i>lowercase</i>)	22
Tabel 3.6	Proses <i>Tokenizing</i>	26
Tabel 3.7	Proses <i>Sequence Padding</i>	26
Tabel 4.1	Hasil perbandingan algoritma-algoritma	31
Tabel 4.2	Hasil perbandingan LSTM Unit	31
Tabel 4.3	Proses <i>embedding layer</i> 100 dimensi	32
Tabel 4.4	Hasil perbandingan penggunaan aktivasi 'relu'	32
Tabel 4.5	Hasil perbandingan penggunaan rasio <i>training</i> dan <i>testing</i>	33
Tabel 4.6	Hasil perbandingan <i>epoch</i> yang berbeda	34



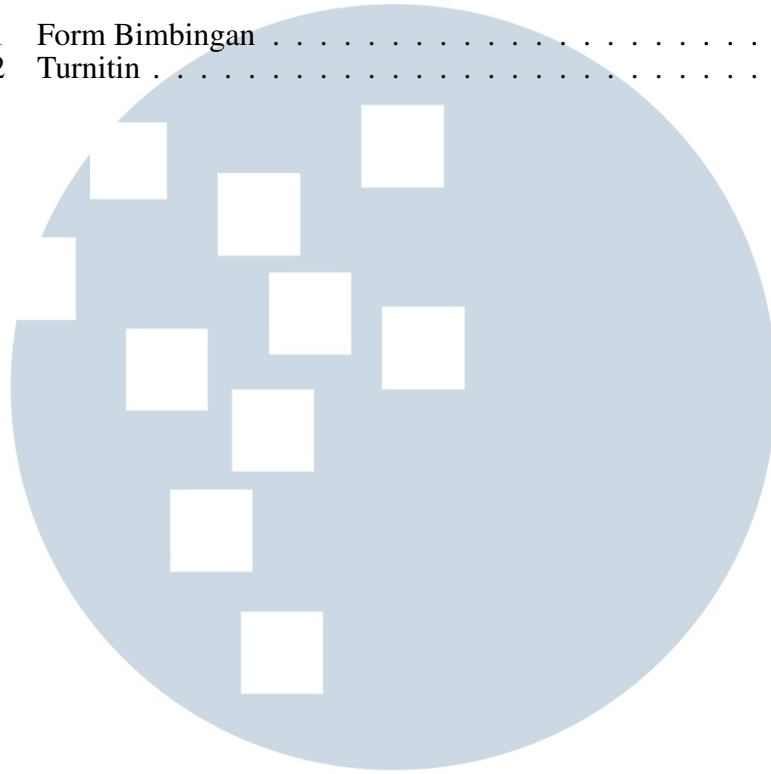
DAFTAR KODE

3.1	<i>Packages</i> dan <i>libraries</i> yang digunakan	14
3.2	Penghapusan <i>tag</i> HTML dan URL serta karakter non-alfabet	17
3.3	Melakukan <i>encoding</i> kategorikal label	23
3.4	Pembagian data	24
3.5	Persiapan data sebelum dimasukkan ke model	24
4.1	Model Bi-LSTM yang digunakan	31



DAFTAR LAMPIRAN

Lampiran 1	Form Bimbingan	55
Lampiran 2	Turnitin	57



UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA