

**Analisis Sentimen *Review* Produk Menggunakan Algoritma
Bidirectional Long Short-Term Memory (Bi-LSTM) dan
*Word Embedding***

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika



Oleh:

Nadia Rizky Hairunnisa

NIM: 09021281823039

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2022**

LEMBAR PENGESAHAN SKRIPSI

ANALISIS SENTIMEN *REVIEW* PRODUK MENGGUNAKAN ALGORITMA *BIDIRECTIONAL LONG SHORT-TERM MEMORY (Bi-LSTM)* DAN WORD EMBEDDING

Oleh:

Nadia Rizky Hairunnisa
NIM : 09021281823039

Palembang, 16 Januari 2023

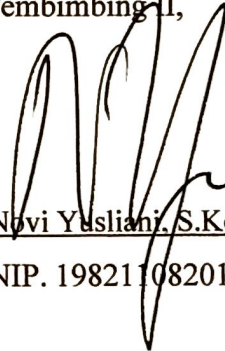
Pembimbing I



Dr. Abdiansah, S.Kom., M.Cs

NIP. 198410012009121005

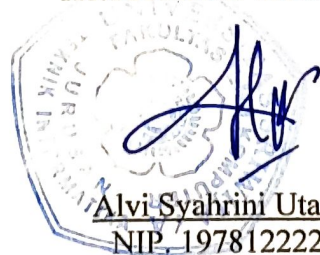
Pembimbing II,



Novi Yusliani, S.Kom., M.T

NIP. 198211082012122001

Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M. Kom.

NIP. 197812222006042003

TANDA LULUS UJIAN KOMPREHENSIF

Pada hari Selasa tanggal 6 Januari 2023 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Nadia Rizky Hairunnisa

NIM : 09021281823039

Judul : Analisis Sentimen *Review* Produk Menggunakan Algoritma *Bidirectional Long Short-Term Memory* dan *Word Embedding*

dan dinyatakan **LULUS**.

1. Ketua

Dr. M. Fachrurrozi, M.T
NIP. 198005222008121002



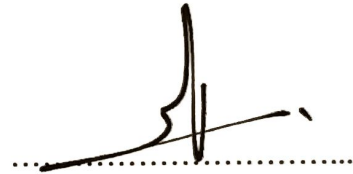
2. Penguji I

Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003



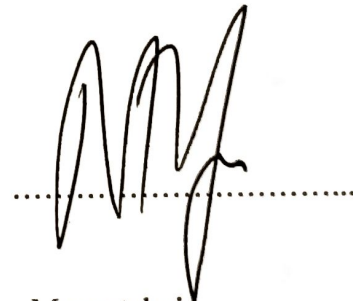
3. Pembimbing I

Dr. Abdiansah, S.Kom, M.Cs
NIP. 198410012009121005

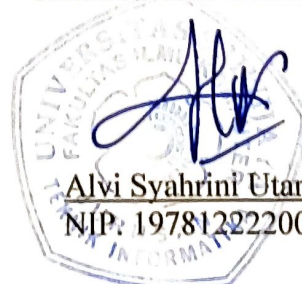


4. Pembimbing II

Novi Yusliani, M.T
NIP. 198211082012122001



Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M. Kom.
NIP. 197812222006042003

PERNYATAAN

Yang bertanda tangan dibawah ini:

Nama : Nadia Rizky Hairunnisa

NIM : 09021281823039

Program Studi : Teknik Informatika

Judul Skripsi : Analisis Sentimen *Review* Produk Menggunakan Algoritma
Bidirectional Long Short-Term Memory dan *Word Embedding*

Hasil Pengecekan Software iThenticate/Turnitin : 10%

Menyatakan bahwa Laporan Proyek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Inderalaya, 23 Januari 2023



Nadia Rizky Hairunnisa
NIM. 09021281823039

MOTTO DAN PERSEMBAHAN

Motto:

- *Authenticity, progress, and compassion*
- *Whatever you choose to water will grow*
- *Cogito, ergo sum* –Descartes
- *Veritas numquam perit* –Seneca
- Terbentur, terbentur, terbentur, terbentuk –Tan Malaka
- *Verily, along with every hardship there is relief [Quran 94:5]*

Kupersembahkan karya tulis ini kepada:

- Allah Subhanahu Wa Ta'ala dan
Rasulullah Shalallahu Alaihi Wasallam
- Mama, Papa, dan Adik Tercinta
- Keluarga Besar
- Teman-teman Seperjuangan
- Dosen Pembimbing
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

**Analisis Sentimen *Review* Produk Menggunakan Algoritma
Bidirectional Long Short-Term Memory (Bi-LSTM) dan
*Word Embedding***

Oleh:

Nadia Rizky Hairunnisa (09021281823039)

Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Sriwijaya

Email: nadiarizkyhairunnisa@gmail.com

ABSTRACT

Sentiment analysis is a computational study of human opinions, sentiments, emotions, and behavior toward entities or attributes expressed through written text. Sentiment analysis plays a significant role for companies and organizations because public opinion about their products and services is valuable for business strategy and evaluation. This research developed a system to classify product review sentiment using Bidirectional Long Short-Term Memory (Bi-LSTM) algorithm and Word Embedding, Word2Vec as the embedding layer. The built system uses two models, the base model, which has the same parameter configuration as the CNN model used in previous research, and the tuned model, whose parameter configuration is based on the results of hyperparameter tuning. The results showed that the second model has the best performance with an accuracy of 90.33%, precision of 99.41%, recall of 90.29%, and F1-Score of 94.61%.

Keywords: Analysis Sentiment, Product Review, Bidirectional Long Short Term-Memory, Word Embedding, Word2Vec

Analisis Sentimen *Review* Produk Menggunakan Algoritma *Bidirectional Long Short-Term Memory (Bi-LSTM)* dan *Word Embedding*

Oleh:

Nadia Rizky Hairunnisa (09021281823039)

Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Sriwijaya

Email: nadiarizkyhairunnisa@gmail.com

ABSTRAK

Analisis sentimen adalah bidang komputasional yang mempelajari tentang opini, sentimen, emosi, dan perilaku manusia terhadap entitas dan atribut yang diekspresikan melalui teks tertulis. Analisis sentimen berperan besar bagi perusahaan atau organisasi karena pendapat publik mengenai produk dan layanan mereka sangat berharga untuk dijadikan evaluasi dan strategi bisnis. Pada penelitian ini, dikembangkan sistem yang dapat mengklasifikasikan sentimen ulasan produk menggunakan metode *Bidirectional Long Short-Term Memory (Bi-LSTM)* dan *Word Embedding*, yaitu *Word2Vec* sebagai *embedding layer*. Pengembangan sistem menggunakan dua model, yaitu *base model* yang konfigurasi parameternya sama dengan model CNN yang digunakan pada penelitian sebelumnya, dan *tuned model* yang konfigurasi parameternya berdasarkan hasil *hyperparameter tuning*. Hasil penelitian menunjukkan model kedua memiliki performa terbaik pada dengan nilai akurasi 90.33%, *precision* 99.41%, *recall* 90.29%, dan *F1-Score* 94.61%.

Kata Kunci: Sentimen Analisis, Ulasan Produk, *Bidirectional Long Short Term-Memory*, *Word Embedding*, *Word2Vec*

KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Allah SWT, Tuhan Semesta Alam atas berkat, rahmat, dan karunia-Nya yang telah diberikan kepada penulis sehingga Tugas Akhir berjudul “Analisis Sentimen *Review* Produk Menggunakan Algoritma *Bidirectional Long Short-Term Memory* (Bi-LSTM) dan *Word Embedding*” dapat disusun dengan baik sebagai syarat dalam menyelesaikan studi Strata-1 program studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Sriwijaya.

Pada kesempatan ini penulis ingin mengucapkan ucapan terima kasih yang luar biasa besar kepada semua pihak yang memberikan dukungan, motivasi dan bimbingan selama penyusunan Tugas Akhir dan penelitian ini berlangsung. Secara khusus ucapan terima kasih ini ditujukan kepada:

1. Allah Subhanahu wa Ta’ala yang telah memberikan hamba kesehatan, kemudahan, rahmat, dan karunia-Nya.
2. Kedua orang tua tercinta, Mama dan Papa yang tak henti memberikan semangat dan dukungan baik dalam bentuk doa, nasihat, bimbingan, solusi, dan banyak hal lainnya yang tak dapat penulis sebutkan satu per satu. Kedua adik tercinta juga keluarga besar yang selalu memberi motivasi untuk tetap belajar dan berproses dalam hidup.
3. Bapak Dr. Abdiansah, S. Kom., M.Cs selaku dosen pembimbing pertama dan Ibu Novi Yusliani, S.Kom, M.T selaku dosen pembimbing kedua yang senantiasa memberikan ilmu, bimbingan, saran, dan dukungan dalam menyelesaikan Tugas Akhir ini.

4. Ibu Alvi Syahrini Utami, M.Kom selaku dosen penguji ujian komprehensif dan Ketua Jurusan Teknik Informatika yang telah memberikan masukan dan saran yang membangun dalam Tugas Akhir ini.
5. Seluruh dosen program studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya yang telah memberikan banyak ilmu, pengalaman, serta arahan selama masa perkuliahan
6. Kak Ricy Firnando selaku admin Jurusan Teknik Informatika dan seluruh staf tata usaha yang telah membantu kelancaran proses administrasi perkuliahan.
7. Suna Alkayuni Aresta, Muhammad Tiansyah Pratama, Muhammad Sholeh, Kurnia Oktaviani, dan Peter Wilburn, sahabat yang telah menemani, mendukung, membantu, dan berbagi suka maupun duka dengan penulis sebelum dan selama masa perkuliahan, serta proses pengerjaan Tugas Akhir.
8. Adi Kurniawan, Febyk Alek Satria, Della Octa Amelia, *ciwi-ciwi* dan keluarga besar TI RegCeh, dan teman-teman seperjuangan Teknik Informatika 2018 yang telah mengulurkan bantuan dan memberikan dukungan dalam masa perkuliahan dan proses pengerjaan Tugas Akhir.
9. Elizabeth Fillips, Ali Abdaal, James Scholz, Heleen Study Vibes, Sabda PS, Cania Citta Irlaine, Gita Wirjawan, Melissa Xyreads, Zahid Azmi Ibrahim, Fellexandro Ruby, Linh Truong, Gazi Jarin, Alffy Rev & Linka Angelia, dan Paola Merrill, yang konten kurasinya telah mencerahkan, mengedukasi, menginspirasi, dan menghibur penulis semasa masa perkuliahan.
10. *Myself*, terima kasih untuk tetap bangun dan berjuang, terima kasih untuk segalanya.

Penulis juga mengucapkan terima kasih kepada semua pihak yang telah membantu penyusunan dalam menyelesaikan Tugas Akhir ini yang tidak dapat penulis sebutkan satu per satu. Penulis harap semoga tulisan ini dapat menjadi bahan pembelajaran di masa yang akan datang. Penulis menyadari bahwa Tugas Akhir ini masih memiliki kekurangan, oleh karena itu penulis memohon maaf apabila terdapat kesalahan, kekhilafan, dan keterbatasan pengetahuan penulis dalam menyusun Tugas Akhir ini. Penulis sangat mengharapkan kritik dan saran yang membangun.

DAFTAR ISI

ABSTRACT.....	vi
ABSTRAK.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL.....	xv
BAB I PENDAHULUAN.....	I-1
1.1 Pendahuluan.....	I-1
1.2 Latar Belakang.....	I-1
1.3 Rumusan Masalah.....	I-4
1.4 Tujuan Penelitian.....	I-4
1.5 Manfaat Penelitian.....	I-5
1.6 Batasan Masalah.....	I-5
1.7 Sistematika Penulisan.....	I-5
1.8 Kesimpulan.....	I-6
BAB II KAJIAN LITERATUR.....	II-1
2.1. Pendahuluan.....	II-1
2.2. Landasan Teori.....	II-1
2.2.1 Analisis Sentimen.....	II-1
2.2.2 Pra-Pengolahan Teks.....	II-4
2.2.3 Word Embedding.....	II-5
2.2.4 Recurrent Neural Network.....	II-8
2.2.5 Long Short-Term Memory.....	II-13
2.2.6 Bidirectional Long Short-Term Memory (Bi-LSTM).....	II-21
2.2.7 Confusion Matrix.....	II-24
2.2.8 Agile Software Development Methodology.....	II-26
2.3. Penelitian Lain yang Relevan.....	II-29
2.4 Kesimpulan.....	II-34
BAB III METODE PENELITIAN.....	III-1
3.1 Pendahuluan.....	III-1
3.2 Pengumpulan Data.....	III-1

3.2.1	Jenis dan Sumber Data	III-1
3.2.2	Metode Pengumpulan Data	III-1
3.3	Tahapan Penelitian	III-2
3.3.1	Mengumpulkan Data	III-3
3.3.2	Mengembangkan Perangkat Lunak untuk Model Bi-LSTM ...	III-3
3.3.3	Menentukan Kerangka Kerja Penelitian	III-5
3.3.4	Menentukan Kriteria dan Format Data Pengujian	III-7
3.3.5	Melakukan Evaluasi Hasil Pengujian	III-8
3.3.6	Membuat Kesimpulan	III-9
3.3.7	Membuat Laporan Penelitian	III-9
3.4	Metode Pengembangan Perangkat Lunak	III-9
3.5	Kesimpulan.....	III-10
BAB IV PENGEMBANGAN PERANGKAT LUNAK		IV-1
4.1	Pendahuluan	IV-1
4.2	Pengembangan Perangkat Lunak	IV-1
4.2.1	Analisis Kebutuhan (<i>Requirement Analysis</i>)	IV-1
4.2.2	Perancangan Perangkat Lunak (<i>Software Design</i>).....	IV-2
4.2.3	Implementasi Perangkat Lunak (<i>Software Implementation</i>)..	IV-20
4.2.4	Pengujian Perangkat Lunak.....	IV-30
4.3	Kesimpulan.....	IV-31
BAB V HASIL DAN ANALISIS PENELITIAN		V-1
5.1	Pendahuluan	V-1
5.2	Data Hasil Penelitian	V-1
5.2.1	Konfigurasi Pengujian.....	V-1
5.2.2	Hasil Penelitian	V-3
5.3	Analisis Hasil Penelitian	V-20
5.4	Kesimpulan.....	V-20
BAB VI KESIMPULAN DAN SARAN		VI-1
6.1	Pendahuluan	VI-1
6.2	Kesimpulan.....	VI-1
6.3	Saran.....	VI-2
DAFTAR PUSTAKA		xi
LAMPIRAN.....		xx

DAFTAR GAMBAR

Gambar II-1. Arsitektur CBOW dan Skip-gram (Landthaler et al., 2017)	II-6
Gambar II-2. Visualisasi Glove embedding (Pennington et al., 2014)	II-8
Gambar II-3. Diagram RNN berdasarkan input pada tiap time step.....	II-9
Gambar II-4. Ilustrasi kemampuan RNN menangkap short-term dependency .	II-12
Gambar II-5. Ilustrasi kegagalan RNN menangkap long-term dependency	II-12
Gambar II-6. Diagram arsitektur LSTM (Yan, 2015).....	II-14
Gambar II-7. Forget gate pada arsitektur LSTM (Yan, 2015)	II-15
Gambar II-8. Input gate pada arsitektur LSTM (Yan, 2015)	II-16
Gambar II-9. Cell state pada arsitektur LSTM (Yan, 2015)	II-19
Gambar II-10. Output gate pada arsitektur LSTM (Yan, 2015)	II-20
Gambar II-11. Ilustrasi kalimat yang membutuhkan konteks	II-21
Gambar II-12. Arsitektur Unidirectional LSTM dan Bidirectional LSTM	II-22
Gambar II-13. Arsitektur Bidirectional LSTM (He & Han, 2020)	II-24
Gambar II-14. Ilustrasi Confusion Matrix (Mohajon, 2020)	II-24
Gambar II-15. Diagram Agile SDLC (Anurina, 2021)	II-26
Gambar III-1. Diagram Tahapan Penelitian.....	III-2
Gambar III-2. Arsitektur Sistem Penelitian	III-5
Gambar IV-1 <i>Use Case</i> Diagram Sistem Analisis Sentimen.....	IV-3
Gambar IV-2 Activity Diagram Melihat Halaman Informasi Dataset.....	IV-11
Gambar IV-3 Activity Diagram Melihat Halaman Informasi Model	IV-12
Gambar IV-4 Activity Diagram Melakukan Prediksi Sentimen	IV-13
Gambar IV-5 Activity Diagram Melakukan Prapengolahan Teks.....	IV-14
Gambar IV-6 Activity Diagram Melakukan Pelatihan dan Pengujian Model BiLSTM-Word2Vec	IV-15
Gambar IV-7 Sequence Diagram Melihat Halaman Informasi Dataset.....	IV-16
Gambar IV-8 Sequence Diagram Melihat Halaman Informasi Pelatihan dan Pengujian Model	IV-17
Gambar IV-9 Sequence Diagram Melakukan Prediksi Sentimen.....	IV-18

Gambar IV-10 <i>Class Diagram</i> Sistem	IV-19
Gambar IV-11 Implementasi <i>Interface</i> Halaman Utama	IV-23
Gambar IV-12 Implementasi <i>Interface</i> Halaman Demo	IV-24
Gambar IV-13 Implementasi <i>Interface</i> Halaman Dataset (Deskripsi).....	IV-24
Gambar IV-14 Implementasi <i>Interface</i> Halaman Dataset (Sampel Data)	IV-25
Gambar IV-15 Implementasi <i>Interface</i> Halaman Model (Deskripsi Model)...	IV-26
Gambar IV-16 Implementasi <i>Interface</i> Halaman Model (Konfigurasi Penelitian)	IV-27
Gambar IV-17 Implementasi <i>Interface</i> Halaman Model (Kurva Hasil Pelatihan dan Pengujian Model)	IV-28
Gambar IV-18 Implementasi <i>Interface</i> Halaman Model (Tabel Hasil Evaluasi Model).....	IV-29
Gambar V-1 Grafik Metrik Evaluasi <i>Base Model</i> pada Dataset 1	V-8
Gambar V-2 Grafik <i>Loss Base Model</i> pada Dataset 1	V-9
Gambar V-3 Grafik Metrik Evaluasi <i>Base Model</i> pada Dataset 2	V-12
Gambar V-4 Grafik <i>Loss Base Model</i> pada Dataset 2	V-12
Gambar V-5 Grafik Metrik Evaluasi <i>Tuned Model</i> pada Dataset 1	V-15
Gambar V-6 Grafik <i>Loss Tuned Model</i> pada Dataset 1	V-16
Gambar V-7 Grafik Metrik Evaluasi <i>Tuned Model</i> pada Dataset 2.....	V-19
Gambar V-8 Grafik <i>Loss Tuned Model</i> pada Dataset 2	V-19

DAFTAR TABEL

Tabel III-1. Format Data Pengujian	III-8
Tabel III-2. Evaluasi Hasil Pengujian	III-8
Tabel IV-1 Kebutuhan Fungsional.....	IV-2
Tabel IV-2 Kebutuhan Non-Fungsional	IV-2
Tabel IV-3 Definisi Aktor Use Case	IV-4
Tabel IV-4 Definisi Use Case	IV-4
Tabel IV-5 Skenario Use Case Melihat Halaman Informasi Dataset	IV-5
Tabel IV-6 Skenario Use Case Melihat Halaman Informasi Model	IV-6
Tabel IV-7 Skenario Use Case Melakukan Prediksi Sentimen.....	IV-7
Tabel IV-8 Skenario <i>Use Case</i> Melihat Tabel Hasil Pelatihan Model	IV-9
Tabel IV-9 Skenario Use Case Melihat Tabel Hasil Pengujian Model	IV-10
Tabel IV-11 Daftar Implementasi Kelas	IV-20
Tabel IV-12 Daftar Implementasi Endpoint REST API	IV-22
Tabel IV-13 Rencana Pengujian Use Case Melihat Halaman Informasi Dataset.....	IV-30
Tabel IV-14 Rencana Pengujian <i>Use Case</i> Melihat Halaman Informasi Pelatihan dan Pengujian Model.....	IV-30
Tabel IV-15 Rencana Pengujian Use Case Melakukan Prediksi Sentimen	IV-30
Tabel IV-16 Hasil Pengujian <i>Use Case</i> Melihat Halaman Informasi Dataset.....	IV-28
Tabel IV-17 Hasil Pengujian <i>Use Case</i> Melihat Halaman Informasi Pelatihan dan Pengujian Model	IV-28
Tabel IV-18 Hasil Pengujian <i>Use Case</i> Melakukan Prediksi Sentimen	IV-29
Tabel V-1 <i>Confusion Matrix</i> Pelatihan <i>Base Model</i> pada Dataset 1.....	V-6
Tabel V-2 Hasil Evaluasi Pelatihan <i>Base Model</i> pada Dataset 1	V-6
Tabel V-3 <i>Confusion Matrix</i> Pengujian <i>Base Model</i> pada Dataset 1	V-7
Tabel V-4 Hasil Evaluasi Pengujian <i>Base Model</i> pada Dataset 1.....	V-7
Tabel V- 5 <i>Confusion Matrix</i> Pelatihan <i>Base Model</i> pada Dataset 2.....	V-9
Tabel V-6 Hasil Evaluasi Pelatihan <i>Base Model</i> pada Dataset 2.....	V-10

Tabel V-7 <i>Confusion Matrix</i> Pengujian <i>Base Model</i> pada Dataset 2	V-10
Tabel V-8 Hasil Evaluasi Pengujian <i>Base Model</i> pada Dataset 2.....	V-11
Tabel V-9 <i>Confusion Matrix</i> Pelatihan <i>Tuned Model</i> pada Dataset 1	V-13
Tabel V-10 Hasil Evaluasi Pelatihan <i>Tuned Model</i> pada Dataset 1.....	V-13
Tabel V-11 <i>Confusion Matrix</i> Pengujian <i>Tuned Model</i> pada Dataset 1	V-14
Tabel V-12 Hasil Evaluasi Pengujian <i>Tuned Model</i> pada Dataset 1	V-14
Tabel V-13 <i>Confusion Matrix</i> Pelatihan <i>Tuned Model</i> pada Dataset 2	V-16
Tabel V-14 Hasil Evaluasi Pelatihan <i>Tuned Model</i> pada Dataset 2.....	V-17
Tabel V-15 <i>Confusion Matrix</i> Pengujian <i>Tuned Model</i> pada Dataset 2	V-17
Tabel V-16 Hasil Evaluasi Pengujian <i>Tuned Model</i> pada Dataset 2	V-18
Tabel V-17 Hasil Rata-Rata Evaluasi Pengujian Sistem	V-20
Tabel V-18 Hasil Rata-Rata Evaluasi Pengujian Sistem Berdasarkan Model	V-20
Tabel V-19 Hasil Rata-Rata Evaluasi Pengujian Sistem Berdasarkan Dataset.....	V-20

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini akan dibahas berkenaan dengan garis besar pokok-pokok pikiran dalam penelitian ini. Pokok pikiran yang akan dibahas antara lain latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, dan manfaat penelitian. Pokok-pokok pikiran yang diuraikan akan dijadikan acuan dalam kajian penelitian ini.

1.2 Latar Belakang

Menurut Kestenbaum (2017), *online marketplace* adalah sebuah website atau aplikasi yang memfasilitasi aktivitas perbelanjaan dari berbagai jenis sumber tanpa harus memiliki inventori. Konsumen dapat mengakses inventori *supplier* secara *online* karena informasi produk yang tersedia bersifat *real-time* dan *marketplace* berperan sebagai fasilitator dalam setiap transaksi yang terjadi.

Dilansir dari laporan Google¹, pertumbuhan *marketplace* di Indonesia merupakan yang tertinggi di Asia Tenggara dengan nilai *Gross Merchandise Value* (GVM) sebesar 44 Miliar USD dan diprediksi akan meningkat menjadi 124 Miliar USD pada tahun 2025. Faktor utama yang mempengaruhi pesatnya perkembangan industri *e-commerce* di Indonesia adalah pandemi Covid-19, yang memicu faktor-

¹ https://storage.googleapis.com/gweb-economy-sea.appspot.com/assets/pdf/eEconomy_SEA_2020_Report.pdf

faktor lainnya seperti meningkatnya penetrasi internet, meningkatnya kebutuhan dasar seperti pangan dan kesehatan, serta perkembangan infrastruktur pembayaran non-tunai (Negara & Soesilowati, 2021). Dengan laju pertumbuhan *marketplace* yang kian memuncak, memahami opini dan preferensi konsumen sangatlah penting untuk sebuah bisnis.

Analisis sentimen adalah bidang komputasional yang mempelajari tentang opini, sentimen, emosi, dan perilaku manusia terhadap entitas dan atribut yang diekspresikan melalui teks tertulis (Liu, 2015). Menurut Bannister (2018), analisis sentimen memiliki peran yang penting untuk bisnis maupun organisasi karena pendapat publik mengenai produk dan layanan mereka sangat berharga untuk dijadikan evaluasi.

Penelitian terkait analisis sentimen telah dilakukan sebelumnya menggunakan berbagai jenis algoritma *machine learning* maupun *deep learning*. Algoritma *machine learning* seperti Naïve Bayes Classifier (NBC), *Support Vector Machine* (SVM), *Decision Tree* (DT) *Classifier*, dan *Maximum Entropy* (ME) *Classifier* memiliki akurasi yang lebih rendah dibandingkan algoritma *deep learning* (Jain & Kaushal, 2018).

Penelitian terbaru yang membandingkan metode-metode *deep learning* pada analisis sentimen menunjukkan bahwa optimisasi LSTM, yaitu *Bidirectional LSTM* (Bi-LSTM) memiliki akurasi terbaik di antara algoritma *deep learning* lainnya seperti LSTM, RNN, CNN (Rachman & Santoso, 2021). Penggunaan metode Bi-LSTM untuk analisis sentimen berbahasa Indonesia masih jarang dilakukan, dan peneliti belum menemukan penelitian terkait yang

mengimplementasikan algoritma tersebut pada data ulasan produk. Hal tersebut menjadi alasan mengapa peneliti memutuskan untuk menggunakan Bi-LSTM sebagai metode utama dalam tugas akhir.

Pemakaian metode *word embedding* pada analisis sentimen terbukti dapat meningkatkan akurasi analisis sentimen, terlepas apapun model utama yang dipakai (Deho et al., 2018; Kurniasari & Setyanto, 2020). *Word embedding* adalah teknik yang dipakai untuk merepresentasikan kata menjadi vektor yang berupa bilangan riil kontinu, sehingga vektor-vektor yang jaraknya berdekatan dapat merepresentasikan kata-kata yang mirip satu sama lain. Ide dasar *word embedding* berasal dari *distributional hypothesis* yang dikembangkan oleh Zellig S. Harris dan John Firth (Lagrari et al 2021).

Metode *word embedding* sebelumnya telah dilakukan pada analisis sentimen Twitter, hasil vektor kata digunakan untuk melatih model *Random Forest* dan terbukti memiliki akurasi yang lebih baik dibandingkan model *Random Forest* tanpa *word embedding* (Deho et al., 2018). Terdapat juga penelitian analisis sentimen data ulasan jasa Traveloka menggunakan algoritma RNN sebagai metode utama, hasilnya juga menunjukkan bahwa RNN yang dikombinasikan dan Word Embedding memiliki akurasi yang lebih baik dibandingkan RNN tanpa *word embedding* (Kurniasari & Setyanto, 2020). Hal tersebut menjadi alasan mengapa peneliti memutuskan untuk menggunakan metode *word embedding* sebagai metode tambahan dalam tugas akhir.

Terdapat beberapa jenis *word embedding* yang biasanya digunakan untuk analisis sentimen. Menurut Imaduddin et al. (2019) yang membandingkan beberapa

metode *word embedding* pada analisis sentimen ulasan hotel, Glove *embedding* memiliki akurasi yang lebih baik dibandingkan metode *word embedding* lainnya seperti Word2Vec dan Doc2Vec. Meskipun memiliki akurasi yang lebih rendah, peneliti memutuskan untuk menggunakan Word2Vec sebagai metode tambahan pada penelitian ini, karena tugas akhir ini menggunakan penelitian Daffa (2022) sebagai referensi utama. Penelitian tersebut menggunakan algoritma Convolutional Neural Network (CNN) yang dikombinasikan dengan Word2Vec, sehingga peneliti akan membandingkan performa metode utama, yaitu Bidirectional Long Short-Term Memory, dengan penelitian sebelumnya yang dilakukan Daffa (2022).

1.3 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana cara membangun sistem analisis sentimen *review* produk menggunakan algoritma Bi-LSTM dan metode *word embedding*?
2. Bagaimana kinerja algoritma Bi-LSTM dan metode *word embedding* untuk sistem analisis sentimen pada *review* produk berdasarkan akurasi, presisi, *recall*, dan F1 *score*?

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menghasilkan sistem analisis sentimen *review* produk menggunakan algoritma Bi-LSTM dan metode *word embedding*.
2. Mengetahui kinerja algoritma Bi-LSTM dan metode *word embedding* untuk sistem analisis sentimen pada *review* produk berdasarkan akurasi, presisi, *recall*, dan F1 *score*.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Dapat mengetahui tingkat akurasi dari sistem analisis sentimen *review* produk menggunakan algoritma Bi-LSTM dan metode *word embedding*.
2. Sistem yang dihasilkan dapat digunakan sebagai bahan evaluasi untuk organisasi atau perusahaan yang bergerak di bidang bisnis.
3. Hasil penelitian dapat dijadikan sebagai rujukan untuk penelitian terkait.

1.6 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data sentimen menggunakan Bahasa Indonesia dengan kategori positif dan negatif.
2. Data sentimen dikumpulkan dari *review* produk pada situs *marketplace* Lazada.
3. Hanya menganalisis *review* produk dalam format karakter A-Z.

1.7 Sistematika Penulisan

Sistematika penulisan dalam penelitian ini adalah sebagai berikut:

BAB I. PENDAHULUAN

Pada bab ini menjelaskan tentang latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian dan batasan masalah. Pokok-pokok pikiran ini akan menjadi dasar dan acuan pengembangan penelitian pada bab selanjutnya.

BAB II. KAJIAN LITERATUR

Pada bab ini dibahas landasan teori yang digunakan di dalam penelitian, termasuk di dalamnya mengenai *marketplace*, analisis sentimen, *machine learning*,

deep learning, text mining, scrapping, word embedding, algoritma Bidirectional Long-Short Term Memory, dan penelitian terkait yang relevan.

BAB III. METODOLOGI PENELITIAN

Pada bab ini dibahas proses pengumpulan data dan tahapan-tahapan di dalam penelitian. Tahapan penelitian dibahas lebih rinci berdasarkan kerangka kerja tertentu. Di bagian akhir bab ini akan dimuat rancangan manajemen proyek penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini menjelaskan analisa dan proses pengembangan sistem, seperti analisis kebutuhan sistem dan konstruksi sistem. Pada akhir bab dilakukan analisa pengujian.

BAB V. HASIL DAN ANALISA PENELITIAN

Pada bab ini berisi hasil yang dilakukan pada sistem serta keakuratan sistem. Pada akhir bab dijelaskan analisis dari penelitian.

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini menjelaskan kesimpulan dari penelitian dan saran ke depannya dari penelitian ini.

1.8 Kesimpulan

Pada Bab ini telah dibahas mengenai latar belakang penelitian serta acuan penting dalam penelitian seperti latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan sistematika penulisan.

DAFTAR PUSTAKA

- Abduljabbar, R. L., Dia, H., & Tsai, P.-W. (2021). Unidirectional and Bidirectional LSTM Models for Short-Term Traffic Prediction. In J. Tang (Ed.), *Journal of Advanced Transportation* (Vol. 2021, pp. 1–16). Hindawi Limited.
- Almeida, F., & Xexéo, G. (2019). Word Embeddings: A Survey.
- Al-Saqqa, S., & Awajan, A. (2019). The Use of Word2vec Model in Sentiment Analysis: A Survey. *Proceedings of the 2019 International Conference on Artificial Intelligence, Robotics and Control*, 39–43. Presented at the Cairo, Egypt.
- Anurina, Olha. (2021). Agile SDLC: Skyrocketing Your Project with Agile Principles. Diakses pada 21 Desember 2021, dari <https://mlsdev.com/blog/agile-sdlc>
- Baktha, K., & Tripathy, B. K. (2017). Investigation of recurrent neural networks in the field of sentiment analysis. *2017 International Conference on Communication and Signal Processing (ICCSP)*, 2047–2050.
- Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*. 5. 157-66.
- Cui, Z., Ke, Ruimin., Pu, Ziyuan & Wang, Y. (2020). Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values. *Transportation Research Part C: Emerging Technologies*. 118. 102674.

- Daffa, R. M. (2022). Analisis Sentimen Ulasan Produk Berbahasa Indonesia Menggunakan *Deep Learning Convolutional Neural Network* dan Word2Vec.
- Divakar, Varun. (2019). Understanding The Chain Rule. Diakses pada 5 November 2021, dari <https://blog.quantinsti.com/understanding-chain-rule/>
- Firth, J. R. (1957). A synopsis of linguistic theory, 1930-55. In Palmer, F.R. (ed.) (1968). *Selected papers of J.R. Firth 1952-59*. London and Harlow: Longman.
- Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to Forget: Continual Prediction with LSTM. *Neural Computation*, 12(10), 2451–2471.
- Ghiffarie, A., Salsabila, K., Baistama, R., Variadi, M., & Rhajendra, M. (2019). Analisis Sentimen Terhadap Produk The Body Shop Tea Tree Oil. *Jurnal Teknologi dan Manajemen Informatika* 5.
- Godin, F., Degraeve, J., Dambre, J., & De Neve, W. (2018). Dual Rectified Linear Units (DReLU): A replacement for tanh activation functions in Quasi-Recurrent Neural Networks. *Pattern Recognition Letters*, 116, 8–14.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. <https://www.deeplearningbook.org/>
- Google, Temasek, Bain & Company (2020). The e-Conomy SEA 2020. Diakses pada 25 Agustus 2021, dari https://storage.googleapis.com/gweb-economy-sea.appspot.com/assets/pdf/e-Conomy_SEA_2020_Report.pdf
- Graves, A., Fernández, S., & Schmidhuber, J. (2005). Bidirectional LSTM Networks for Improved Phoneme Classification and Recognition. *Proceedings of the 15th International Conference on Artificial*

Neural Networks: Formal Models and Their Applications - Volume Part II,
799–804.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10), 2222–2232.

Gunawan, B., Sastypratiwi, H., Pratama, E. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*. 4. 113.

Harris, Z. S. (1954). Distributional structure. *Word*, 10, 146–162.

Haryanto, D., Muflikhah, L., & Fauzi, M. (2018). Analisis Sentimen Review Barang Berbahasa Indonesia Dengan Metode Support Vector Machine Dan Query Expansion. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(9), 2909-2916.

He, Shirong & Han, Dezhi. (2020). An Effective Dense Co-Attention Networks for Visual Question Answering. *Sensors*. 20. 4897.

Hochreiter, Sepp & Schmidhuber, Jürgen. (1996). LSTM can solve hard long time lag problems. *Advances in Neural Information Processing Systems*. 473-479.

Hochreiter, Sepp & Schmidhuber, Jürgen. (1997). Long Short-term Memory. *Neural computation*. 9. 1735-80.

Humairah, H., Darmawan, I., & Pratiwi, O.N. (2020). Analisis Sentimen Ulasan Produk Toko Online Rubylicious Untuk Peningkatan Layanan Menggunakan Algoritma Naive Bayes.

Ibrahim, N. (2012). An Overview of Agile Software Development Methodology and Its Relevance to Software Engineering.

- IBM Cloud Education. (2020). What is Text Mining? Diakses pada 20 September 2021, dari <https://www.ibm.com/cloud/learn/text-mining>
- Irfani, F.F., Triyanto, M., Hartanto, A.D., & Kusnawi, K. (2020). Analisis Sentimen Review Aplikasi RuangGuru Menggunakan Algoritma Support Vector Machine.
- Isnain, A., Sihabuddin, A., & Suyanto, Y. (2020). Bidirectional Long Short Term Memory Method and Word2vec Extraction Approach for Hate Speech Detection. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 14, 169.
- Jain, P. K., Saravanan, V., & Pamula, R. (2021). A Hybrid CNN-LSTM: A Deep Learning Approach for Consumer Sentiment Analysis Using Qualitative User-Generated Contents. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 20(5).
- Jiao, Q., & Zhang, S. (2021). A Brief Survey of Word Embedding and Its Recent Development. *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 5, 1697–1701.
- Kadhim, Ammar. (2018). An Evaluation of Preprocessing Techniques for Text Classification. *International Journal of Computer Science and Information Security*. 16. 22-32.
- Kestenbaum, Richard. (2017). What Are Online Marketplaces and What Is Their Future? Diakses pada 25 Agustus 2021, dari <https://www.forbes.com/sites/richardkestenbaum/2017/04/26/what-are-online-marketplaces-and-what-is-their-future>
- King, Timothy. (2019). 80 Percent of Your Data Will Be Unstructured in Five Years. Diakses pada 20 September 2021, dari <https://solutionsreview.com/>

data-management/80-percent-of-your-data-will-be-unstructured-in-five-years/

Lagrari, F.-E., & Elkettani, Y. (2021). Traditional and Deep Learning Approaches for Sentiment Analysis: A Survey. *Advances in Science, Technology and Engineering Systems Journal*, 6(5), 1–7.

Landthaler, J., Waltl, B., Huth, D., Braun, D., Matthes, F., Stocker, C., & Geiger, T. (2017). Extending Thesauri Using Word Embeddings and the Intersection Method.

Li, X., Xianyu, H., Tian, J., Chen, W., Fanhang, M., Xu, M. & Cai, L. (2016). A deep bidirectional long short-term memory based multi-scale approach for music dynamic emotion prediction. 544-548.

Liu, B. (2015). *Sentiment analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge: Cambridge University Press.

Liu, Y., Lu, J., Yang, J., & Mao, F. (2020). Sentiment analysis for e-commerce product reviews by deep learning model of Bert-BiGRU-Softmax. *Mathematical Biosciences and Engineering*, 17, 7819–7837.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013a). Efficient Estimation of Word Representations in Vector Space.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013b). Distributed Representations of Words and Phrases and their Compositionality.

Mohajon, Joydwip. (2020). Confusion Matrix for Your Multi-Class Machine Learning Model. Diakses pada 4 Desember 2021, dari <https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826>

- Mutabaruka, E. (2021). Agile Methodology Software Development Adaptability Challenges in Corporate Organization. Student
- Muktafin, E. H., Kusriani, K., & Luthfi, E. T. (2020). Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing. *Eksplora Informatika*, 10(1), 32–42.
- Negara, S. D. dkk. (2021). E-Commerce in Indonesia: Impressive Growth but Facing Serious Challenges. Diakses pada 25 Agustus 2021, dari https://www.iseas.edu.sg/wp-content/uploads/2021/07/ISEAS_Perspective_2021_102.pdf
- Olah, Christopher. (2015). Understanding LSTM Networks. Diakses pada 2 November 2021, dari <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- Deho, O. B., Agangiba, W. A., Aryeh, F. L., & Ansah, J. A. (2018). Sentiment Analysis with Word Embedding. *2018 IEEE 7th International Conference on Adaptive Science & Technology (ICAST)*, 1-4.
- Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. EMNLP. 14.
- Prabha, M. I., & Umarani Srikanth, G. (2019). Survey of Sentiment Analysis Using Deep Learning Techniques. *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*, 1–9.
- Pujari, C., Bhat, A., & Shetty, N. (2016). Comparison of Classification Techniques for Feature Oriented Sentiment Analysis of Product Review Data.
- Pykes, Kurtis. (2020). The Vanishing/Exploding Gradient Problem in Deep Neural Network: Understanding the Obstacles That Faces Us When Building Deep Neural Networks. Diakses pada 5 November 2021, dari

<https://towardsdatascience.com/the-vanishing-exploding-gradient-problem-in-deep-neural-networks-191358470c11>

Rachman, F., & Santoso, H. (2021). Perbandingan Model Deep Learning untuk Klasifikasi Sentiment Analysis dengan Teknik Natural Language Processing. *Jurnal Teknologi dan Manajemen Informatika*, 7(2), 103-112.

Rahman, M. M., & Siddiqui, F. H. (2019). An Optimized Abstractive Text Summarization Model Using Peephole Convolutional LSTM. *Symmetry*, 11(10).

Rokhman, K. A., Berlilana, B., & Arsi, P. (2021). Perbandingan Metode Support Vector Machine dan Decision Tree untuk Analisis Sentimen Review Komentar pada Aplikasi Transportasi Online. *Journal of Information System Management (JOISM)*, 2(2), 1-7.

Rupapara, V., Rustam, F., Amaar, A., Washington, P. B., Lee, E., & Ashraf, I. (2021). Deepfake tweets classification using stacked Bi-LSTM and words embedding. *PeerJ. Computer science*, 7, e745.

Sari, P. K., Alamsyah, A., & Wibowo, S. (2018). Measuring e-Commerce service quality from online customer review using sentiment analysis. *Journal of Physics: Conference Series*. 971. 012053.

Schuster, M., & Paliwal, K. (1997). Bidirectional recurrent neural networks. *Signal Processing, IEEE Transactions on*. 45. 2673 - 2681.

Srujan, K. S., Nikhil, S. S., Rao, R., Karthik, K., Harish, B. S., & Kumar, H. M. K., (2018) Classification of Amazon Book Reviews Based on Sentiment Analysis. In: Bhateja V., Nguyen B., Nguyen N., Satapathy S., Le DN.

- (eds) Information Systems Design and Intelligent Applications. Advances in Intelligent Systems and Computing, vol 672. Springer, Singapore.
- Surohman, S., Aji, S., Rousyati, R., & Wati, F. (2020). Analisa Sentimen Terhadap Review Fintech Dengan Metode Naive Bayes Classifier Dan K- Nearest Neighbor. *EVOLUSI: Jurnal Sains dan Manajemen*. 8.
- Elli, M. S. & Yi-Fan, W. (2016). Amazon Reviews, business analytics with sentiment analysis.
- Tan, T. G., Teo, J., & Anthony, P. (2011). A comparative investigation of non-linear activation functions in neural controllers for search-based game AI engineering. *Artificial Intelligence Review*, 41, 1–25.
- Ting K.M. (2011). Confusion Matrix. In: Sammut C., Webb G.I. (eds) *Encyclopedia of Machine Learning*. Springer, Boston, MA.
- Tripathy, A., & Rath, S. K. (2017). Classification of Sentiment of Reviews using Supervised Machine Learning Techniques. *International Journal of Rough Sets and Data Analysis (IJRSDA)*, 4(1), 56-74.
- Wang, S., Zhou, W., & Jiang, C. (2020). A survey of word embeddings based on deep learning.
- Wint, Z. Z., Manabe, Y., & Aritsugi, M. (2018). Deep Learning Based Sentiment Classification in Social Network Services Datasets. *2018 IEEE International Conference on Big Data, Cloud Computing, Data Science Engineering (BCD)*, 91–96.
- Yan, Shi. (2015). Understanding LSTM and Its Diagram. Diakses pada 11 November 2021, dari <https://blog.mlreview.com/understanding-lstm-and-its-diagrams-37e2f46f1714>

Yin, W., Kann, K., Yu, M., & Schütze, H. (2017). Comparative Study of CNN and RNN for Natural Language Processing. *ArXiv, abs/1702.01923*.

Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2021). Dive into Deep Learning.