

**ANALISIS SENTIMEN ULASAN PRODUK BERBAHASA
INDONESIA MENGGUNAKAN *DEEP LEARNING*
CONVOLUTIONAL NEURAL NETWORK DAN *WORD2VEC***

*Diajukan sebagai syarat untuk menyelesaikan
Pendidikan Program Strata-I
pada Jurusan Teknik Informatika*



Oleh :

Revan Muhammad Dafa

NIM : 09021181722081

Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2023

LEMBAR PENGESAHAN SKRIPSI

ANALISIS SENTIMEN ULASAN PRODUK BERBAHASA INDONESIA MENGGUNAKAN *DEEP LEARNING* *CONVOLUTIONAL NEURAL NETWORK* DAN *WORD2VEC*

Oleh :

Revan Muhammad Dafa
NIM : 09021181722081

Indralaya, 14 Agustus 2023

Pembimbing I



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

Pembimbing II


Desty Rodiah, S.Kom., M.T.
NIP. 198912212020122011

Mengetahui,
Ketua Jurusan Teknik Informatika




Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

TANDA LULUS UJIAN SIDANG SKRIPSI

Pada hari Jum'at, 26 Agustus 2022 telah dilaksanakan ujian sidang skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Revan Muhammad Dafa

NIM 09021181722081

Judul : Analisis Sentimen Ulasan Produk Berbahasa Indonesia
menggunakan *Deep Learning Convolutional Neural Network* dan *Word2Vec*

Dengan ini dinyatakan bahwa telah lulus ujian sidang skripsi.

1. Ketua Penguji

Dr. M. Fachrurrozi, M.T.

NIP. 198005222008121002



2. Penguji I

Novi Yusliani, M.T.

NIP. 198211082012122001



3. Penguji II

Annisa Darmawahyuni, M.Kom

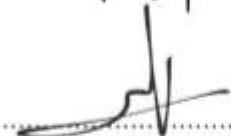
NIP. 1671147006900002



4. Pembimbing I

Dr. Abdiansah, M.CS

NIP. 198410012009121005



5. Pembimbing II

Desty Rodiah, M.T.

NIP. 198912212020122011



Mengetahui

Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.

NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini :

Nama : Revan Muhammad Dafa
NIM : 09021181722081
Program Studi : Teknik Informatika
Judul : Analisis Sentimen Ulasan Produk Berbahasa Indonesia
menggunakan *Deep Learning Convolutional Neural
Network* dan *Word2Vec*


Hasil pengecekan software ithenticate/Turnitin : 7%

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

Demikian pernyataan ini saya buat dengan sebenar-benarnya dan tidak ada paksaan dari pihak manapun.



Palembang, Agustus 2023


Revan Muhammad Dafa
NIM. 09021181722081

MOTO DAN PERSEMBAHAN

“Setia dengan yang baik”

Kupersembahkan karya tulis ini kepada:

- Kedua orang tuaku
- Saudaraku
- Sahabatku
- Almamaterku

ANALISIS SENTIMEN ULASAN PRODUK BERBAHASA INDONESIA MENGGUNAKAN *DEEP LEARNING* *CONVOLUTIONAL NEURAL NETWORK* DAN *WORD2VEC*

Oleh:

Revan Muhammad Dafa (09032282733082)

Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas
Sriwijaya

ABSTRAK

Fenomena perdagangan elektronik timbul akibat dari pesatnya perkembangan teknologi dan media komunikasi. Ulasan produk menjadi informasi yang penting bagi pengguna media perdagangan elektronik, namun sering sekali ulasan dituliskan dengan cara yang tidak benar. Kami menggunakan metode *Convolutional Neural Network* dengan fitur *Word2Vec* untuk melakukan analisis sentimen ulasan yang berjumlah 5.080.101. *Probability of Similarity* akan digunakan sebagai tambahan proses pra pengolahan untuk mengetahui pengaruh kata yang diperbaiki terhadap model yang dilatih. Berdasarkan hasil pengujian, model menunjukkan performansi yang sangat baik dengan akurasi 0.91119, presisi 0.91123, recall 0.91232, dan F1-score 0.91178. Performansi model dengan tahapan normalisasi memiliki hasil yang sangat baik dibandingkan dengan tahapan *Stemming* dan *Stopword Removal*. Tahapan perbaikan kata dengan metode *Probability of Similarity* meningkatkan 4 dari 5 skema data, namun dengan data latih yang cukup besar proses ini dirasa cukup memakan waktu dan sumber daya.

Kata kunci: *Convolutional Neural Network*, *Probability of Similarity*, Analisis Sentimen

SENTIMENT ANALYSIS OF INDONESIAN PRODUCT REVIEWS USING DEEP LEARNING CONVOLUTIONAL NEURAL NETWORK AND WORD2VEC

By:

Revan Muhammad Dafa (09032282733082)

Informatics Engineering, Faculty of Computer Science, Sriwijaya

University Sriwijaya

ABSTRACT

The phenomenon of electronic commerce arises as a result of the rapid development of technology and communication media. Product reviews are important information for users of electronic commerce media, but reviews are often written incorrectly. The Deep Learning method is a method that is developing very rapidly at this time. We used the Convolutional Neural Network method with the Word2Vec feature to carry out a sentiment analysis of 5,080,101 reviews. Probability of Similarity will be used as an additional pre-processing process to determine the effect of the corrected word on the model being trained. Based on the test results, the model shows excellent performance with an accuracy of 0.91119, a precision of 0.91123, a recall of 0.91232, and an F1-score of 0.91178. The performance of the model with the normalization stage which has quite good performance compared to the stemming and stopword removal stages. The stages of word improvement using the Probability of Similarity method improve 4 out of 5 data schemes, but with large enough training data this process is considered to be quite time consuming and resource consuming.

Keywords : Convolutional Neural Network, Probability of Similarity, Sentiment Analysis

KATA PENGANTAR

Pujian syukur kepada Allah SWT atas berkat dan rahmat-Nya yang telah diberikan kepada Penulis sehingga dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas akhir ini disusun untuk memenuhi salah satu syarat guna menyelesaikan pendidikan program Strata-1 pada Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Dalam menyelesaikan Tugas Akhir ini banyak pihak yang telah memberikan bantuan dan dukungan baik secara langsung maupun secara tidak langsung. Penulis ingin menyampaikan rasa terima kasih kepada:

1. Kedua orang tua saya yang telah setia dan sabar menunggu saya dalam menyelesaikan studi ini
2. Ibu Alvi Syahrini Utami, M.Kom. selaku ketua jurusan Teknik Informatika yang telah memberi kesempatan saya untuk mendapatkan tambahan waktu dalam mengerjakan penelitian ini.
3. Bapak Dr. Abdiansah, M.Cs. dan Ibu Desty Rodiah, M.Kom. yang telah membimbing saya dalam mengerjakan penelitian ini.
4. Seluruh Bapak dan Ibu Dosen Teknik Informatika Universitas Sriwijaya.
5. Bapak Cipta Atsahlantusay yang sudah memberikan kesempatan saya untuk mengasah ilmu *programming* saya di bagian Mitra Bisnis Layanan TI PT Pupuk Sriwijaya.
6. Kak Handika, Kak Denny, dan Kak Sukirman yang telah menjadi mentor saya selama magang dan bekerja di PT Pupuk Sriwijaya.
7. Teman teman MBLTI yang telah berbagi cerita drama dan ilmunya selama

magang dan bekerja di PT Pupuk Sriwijaya.

8. Teman-teman kelas dan angkatan yang selalu berjuang bersama sama dalam menuntut ilmu di perkuliahan

Penulis menyadari dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya. Akhir kata semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Palembang, Agustus 2023

Revan Muhammad Dafa

DAFTAR ISI

	Halaman
JUDUL	i
LEMBAR PENGESAHAN SKRIPSI.....	ii
TANDA LULUS UJIAN SIDANG SKRIPSI	iii
HALAMAN PERNYATAAN	iv
ABSTRAK	vi
ABSTRACT.....	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR	xvi
DAFTAR LAMPIRAN	xvi
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-4
1.6 Batasan Masalah.....	I-5
1.7 Sistematika Penulisan.....	I-6
1.8 Kesimpulan.....	I-7
BAB II KAJIAN LITERATUR	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori.....	II-1
2.2.1 Lazada.....	II-1
2.2.2 Analisis Sentimen	II-4
2.2.3 <i>Document Preprocessing</i>	II-5
2.2.4 Klasifikasi Teks	II-6
2.2.5 <i>Convolutional Neural Network</i>	II-7
2.2.6 <i>Word2Vec</i>	II-12

2.2.7	<i>Probability of Similarity</i>	II-16
2.2.8	<i>Rational Unified Process</i>	II-16
2.2.9	Pengukuran Hasil Klasifikasi.....	II-20
2.3	Penelitian Lain yang Relevan.....	II-22
2.4	Kesimpulan.....	II-26
BAB III METODOLOGI PENELITIAN.....		III-1
3.1	Pendahuluan	III-1
3.2	Pengumpulan Data	III-1
3.3.1	Jenis dan Sumber Data.....	III-1
3.3.2	Metode Pengumpulan Data.....	III-4
3.3	Tahapan Penelitian	III-6
3.3.1	Mengumpulkan Data Penelitian.....	III-7
3.3.2	Kerangka Kerja Penelitian	III-8
3.3.3	Kriteria Pengujian	III-12
3.3.4	Format Data Pengujian	III-12
3.3.5	Alat yang digunakan pada Penelitian.....	III-13
3.3.6	Pengujian Penelitian	III-14
3.3.7	Analisis Hasil Pengujian Penelitian dan Kesimpulan.....	III-15
3.4	Metode Pengembangan Perangkat Lunak.....	III-16
3.4.1	Fase Insepsi.....	III-16
3.4.2	Fase Elaborasi	III-17
3.4.3	Fase Konstruksi.....	III-18
3.4.4	Fase Transisi	III-18
3.5	Rencana Jadwal Penelitian	III-19
3.6	Kesimpulan.....	III-21
BAB IV PENGEMBANGAN PERANGKAT LUNAK.....		IV-22
4.1	Pendahuluan	IV-22
4.2	Fase Insepsi	IV-22
4.2.1	Pemodelan Bisnis	IV-22
4.2.1.1	Kebutuhan Bisnis.....	IV-22
4.2.1.2	Kebutuhan Teknis.....	IV-2
4.2.1.2.1	Modul <i>Crawl</i>	IV-2
4.2.1.2.2	Modul Prapengolahan.....	IV-2
4.2.1.2.3	Modul Pembobotan.....	IV-2

4.2.1.2.4	Modul Perbaikan Kata.....	IV-3
4.2.1.2.5	Modul Analisis Sentimen	IV-3
4.2.2	Analisis dan Desain	IV-3
4.2.2.1	Analisis Kebutuhan Perangkat Lunak	IV-3
4.2.2.2	Desain	IV-4
4.2.2.2.1	Diagram <i>Use-case</i>	IV-4
4.2.2.2.1.1	Definisi Aktor.....	IV-5
4.2.2.2.1.2	Definisi <i>Use-case</i>	IV-5
4.2.2.2.1.3	Skenario <i>Use-case</i>	IV-6
4.3	Fase Elaborasi	IV-11
4.3.1	Antarmuka Perangkat Lunak	IV-11
4.3.2	Kebutuhan Sistem.....	IV-14
4.3.3	Pemodelan Modul.....	IV-15
4.4	Fase Konstruksi	IV-22
4.4.1	Implementasi.....	IV-23
4.5	Fase Transisi.....	IV-25
4.5.1	Pengujian <i>Use-case</i> Mengambil Data Ulasan.....	IV-25
4.5.2	Pengujian <i>Use-case</i> Melakukan Prapengolahan Teks	IV-25
4.5.3	Pengujian <i>Use-case</i> Melakukan Pembobotan.....	IV-26
4.5.4	Pengujian <i>Use-case</i> Melakukan Perbaikan Kata	IV-26
4.5.5	Pengujian <i>Use-case</i> Melakukan Sentimen Analisis.....	IV-27
4.5.6	Pengujian <i>Use-case</i> Mengambil Data Ulasan.....	IV-27
4.5.7	Pengujian <i>Use-case</i> Melakukan Prapengolahan Teks	IV-28
4.5.8	Pengujian <i>Use-case</i> Melakukan Pembobotan.....	IV-29
4.5.9	Pengujian <i>Use-case</i> Melakukan Perbaikan Kata	IV-30
4.5.10	Pengujian <i>Use-case</i> Melakukan Sentimen Analisis	IV-31
BAB V HASIL DAN PEMBAHASAN.....		V-1
5.1	Pendahuluan	V-1
5.2	Hasil Pengujian Program.....	V-1
5.3	Hasil Pengujian	V-4
5.3.1	Hasil Pengujian Pada Data D1	V-4
5.3.2	Hasil Pengujian Pada Data D2.....	V-5
5.3.3	Hasil Pengujian Pada Data D3.....	V-6
5.3.4	Hasil Pengujian Pada Data D4.....	V-7

5.3.5 Hasil Pengujian Pada Data D5.....	V-9
5.3.6 Hasil Pengujian Pada Data D6.....	V-10
5.3.7 Hasil Pengujian Pada Data D7.....	V-11
5.3.8 Hasil Pengujian Pada Data D8.....	V-13
5.3.9 Hasil Pengujian Pada Data D9.....	V-14
5.3.10 Hasil Pengujian Pada Data D10	V-15
5.3 Hasil Penelitian	V-17
5.3.1 Konfigurasi Pengujian	V-17
5.3.2 Hasil <i>Convolutional Neural Network</i>	V-17
5.3.3 Hasil Prapengolahan Terhadap Model Analisis Sentimen ...	V-19
5.3.4 Hasil Perbaikan Kata Terhadap Tahapan Prapengolahan.....	V-21
BAB VI KESIMPULAN DAN SARAN	VI-1
6.1 Kesimpulan.....	VI-1
6.2 Saran.....	VI-1
DAFTAR PUSTAKA	xvii

DAFTAR TABEL

Tabel II-1. <i>Confusion Matrix</i>	II-21
Tabel III-1 Contoh Ulasan Produk pada Lazada.....	III-2
Tabel III-2 Jumlah Rating pada data Ulasan Produk	III-3
Tabel III-3 Fenomena Penulisan Kata pada Data Ulasan	III-3
Tabel III-4 <i>Confusion Matrix</i> Analisis Sentimen.....	III-12
Tabel III-5 Kinerja Model Analisis Sentimen.....	III-13
Tabel III-6 Hasil Pengujian Model	III-16
Tabel III-7 Kedekatan Kata Sebelum Perbaikan Kata	III-16
Tabel III- 8 Rencana Jadwal Penelitian	III-19
Tabel IV-1 Definisi Aktor.....	IV-5
Tabel IV-2 Definisi <i>Use-case</i>	IV-5
Tabel IV-3 Skenario Mengambil Data Ulasan.....	IV-6
Tabel IV-4 Skenario Prapengolahan Teks	IV-7
Tabel IV-5 Skenario Pembobotan Teks	IV-8
Tabel IV-6 Skenario Perbaikan Kata	IV-9
Tabel IV-7 Skenario Analisis Sentimen.....	IV-10
Tabel IV-8 Keterangan Implementasi Kelas.....	IV-23
Tabel IV-9 Rencana Pengujian <i>Use-case</i> Mengambil Data Ulasan.....	IV-25
Tabel IV-10 Rencana Pengujian <i>Use-case</i> Melakukan Prapengolahan Teks ..	IV-25
Tabel IV-11 Rencana Pengujian <i>Use-case</i> Melakukan Pembobotan.....	IV-26
Tabel IV-12 Rencana Pengujian <i>Use-case</i> Melakukan Perbaikan Kata	IV-26
Tabel IV-13 Rencana Pengujian <i>Use-case</i> Melakukan Sentimen Analisis	IV-23
Tabel IV-14 Rencana Pengujian <i>Use-case</i> Mengambil Data Ulasan.....	IV-27
Tabel IV-15 Rencana Pengujian <i>Use-case</i> Melakukan Prapengolahan Teks ..	IV-28
Tabel IV-16 Rencana Pengujian <i>Use-case</i> Melakukan Pembobotan.....	IV-29
Tabel IV-17 Rencana Pengujian <i>Use-case</i> Melakukan Perbaikan Kata	IV-30
Tabel IV-18 Rencana Pengujian <i>Use-case</i> Melakukan Sentimen Analisis	IV-28
Tabel V-1 Dataset	V-2
Tabel V-2 Model Word2Vec	V-2
Tabel V-3 Confusion Matrix Model D1.....	V-4
Tabel V-4 Hasil Klasifikasi Sentimen Model D1	V-4
Tabel V-5 Confusion Matrix Model 2	V-5
Tabel V-6 Hasil Klasifikasi Sentimen Model 2	V-6
Tabel V-7 Confusion Matrix Model 3	V-7
Tabel V-8 Hasil Klasifikasi Sentimen Model 3	V-7
Tabel V- 9. Confusion Matrix Model 4	V-8
Tabel V-10. Hasil Klasifikasi Sentimen Model 4	V-8
Tabel V-11. Confusion Matrix Model D1.....	V-9
Tabel V-12. Hasil Klasifikasi Sentimen Model D1	V-9
Tabel V-13. Confusion Matrix Model D6.....	V-11
Tabel V-14. Hasil Klasifikasi Sentimen Model D6	V-11

Tabel V-15. Confusion Matrix Model 7	V-12
Tabel V-16. Hasil Klasifikasi Sentimen Model 7	V-12
Tabel V-17 Confusion Matrix Model 8	V-13
Tabel V-18 Hasil Klasifikasi Sentimen Model 8	V-13
Tabel V-19 Confusion Matrix Model 9	V-14
Tabel V-20 Hasil Klasifikasi Sentimen Model 9	V-15
Tabel V-21. Confusion Matrix Model 10	V-16
Tabel V-22. Hasil Klasifikasi Sentimen Model 10	V-16
Tabel V-23 Hasil Pelatihan Seluruh Model	V-18
Tabel V-24 Perbandingan Hasil Model M1 dengan M6	V-21
Tabel V-25 Perbandingan Hasil Model M1 dengan M6	V-22
Tabel V-26 Perbandingan Hasil Model M3 dan M8	V-22
Tabel V-27 Perbandingan Hasil Model M4 dengan M9	V-22
Tabel V-28 Perbandingan Hasil Model M5 dengan M10	V-22

DAFTAR GAMBAR

Gambar II-1. Operasi <i>Pooling</i> pada CNN.....	II-10
Gambar II-2. <i>Fully Connected Layer</i> pada CNN.....	II-10
Gambar II-3. Arsitektur CBOW (kiri) dan Arsitektur <i>Skip-Gram</i> (kanan).....	II-13
Gambar II-4. <i>Rational Unified Process</i>	II-17
Gambar III-1. Halaman Ulasan Produk <i>website</i> Lazada.....	III-2
Gambar III-2. JSON Data Ulasan Produk pada API Ulasan Lazada.....	III-6
Gambar III-3. Tahapan Penelitian.....	III-7
Gambar III-4. Kerangka Kerja Penelitian	III-8
Gambar IV-1. Diagram <i>Use-case</i> Perangkat Lunak	IV-4
Gambar IV-2. Halaman Informasi Penelitian	IV-11
Gambar IV-3. Halaman Informasi Dataset	IV-12
Gambar IV-4. Halaman Informasi Model.....	IV-13
Gambar IV-5. Halaman Informasi Hasil Pelatihan	IV-13
Gambar IV-6. Halaman Demo Analisis Sentimen.....	IV-14
Gambar IV-7. Diagram Aktivitas Melakukan <i>Crawling Dataset</i> API Lazada.....	IV-15
Gambar IV-8. Diagram Aktivitas Prapengolahan Teks	IV-16
Gambar IV-9. Diagram Aktivitas <i>Word Embedding</i>	IV-17
Gambar IV-10. Diagram Aktivitas Perbaikan Kata	IV-18
Gambar IV-11. Diagram Aktivitas Membuka Halaman Dokumentasi.....	IV-19
Gambar IV-12. Diagram Aktivitas Analisis Sentimen	IV-20
Gambar IV-13. Diagram Aktivitas Membuka Halaman Dataset	IV-21
Gambar IV-14. Diagram Aktivitas Melakukan Prediksi Sentimen	IV-22
Gambar V-1. Grafik Akurasi Model 1	V-4
Gambar V-2. Grafik <i>Loss</i> Model 1	V-4
Gambar V-3. Grafik Akurasi Model 2	V-5
Gambar V-4. Grafik <i>Loss</i> Model 2	V-5
Gambar V-5. Grafik Akurasi Model 3	V-6
Gambar V-6. Grafik <i>Loss</i> Model 3	V-6
Gambar V-7. Grafik Akurasi Model 4	V-7
Gambar V-8. Grafik <i>Loss</i> Model 4	V-8
Gambar V-9. Grafik Akurasi Model 5	V-9
Gambar V-10. Grafik <i>Loss</i> Model 5	V-9
Gambar V-11. Grafik Akurasi Model 6	V-10
Gambar V-12. Grafik <i>Loss</i> Model 6	V-10
Gambar V-13. Grafik Akurasi Model 7	V-11
Gambar V-14. Grafik <i>Loss</i> Model 7	V-12
Gambar V-15. Grafik Akurasi Model 8	V-13
Gambar V-16. Grafik <i>Loss</i> Model 8	V-13
Gambar V-17. Grafik Akurasi Model 9	V-14
Gambar V-18. Grafik <i>Loss</i> Model 9	V-14
Gambar V-19. Grafik Akurasi Model 10	V-15

Gambar V-20. Grafik <i>Loss</i> Model 10	V-16
Gambar V-21. Grafik Akurasi Berdasarkan Prapengolahan.....	V-19
Gambar V-22. Grafik Akurasi Berdasarkan Presisi	V-19
Gambar V- 23. Grafik Presisi Berdasarkan <i>Recall</i>	V-20
Gambar V- 24. Grafik <i>Recall</i> Berdasarkan Proses Prapengolahan.....	V-20

BAB I

PENDAHULUAN

1.1 Pendahuluan

Berisi gagasan utama yang menjadi dasar penulisan skripsi. Gagasan utama meliputi latar belakang dari penelitian, rumusan masalah, tujuan dan manfaat penelitian, batasan-batasan masalah, kerangka penulisan skripsi, dan kesimpulan.

1.2 Latar Belakang

Fenomena perdagangan elektronik atau lebih dikenal dengan nama *e-commerce* adalah salah satu dampak pesatnya perkembangan teknologi dan media komunikasi. Variabel kepercayaan, kemudahan, dan kualitas informasi menjadi pengaruh yang besar terkait pemilihan lazada sebagai situs *e-commerce* (Mulyadi et al., 2018). Ulasan produk menjadi informasi yang penting bagi pengguna situs *e-commerce*. Konsumen pada umumnya melihat ulasan sebelum melakukan transaksi pembelian. Konsumen dapat melihat ulasan dan rating konsumen lain untuk mengetahui kualitas produk yang ingin mereka beli. Walaupun begitu penting, informasi pada ulasan sering sekali dituliskan menggunakan bahasa yang tidak baku.

Analisis sentimen adalah studi komputasi tentang mengenali dan mengekspresikan opini, sentimen, evaluasi, sikap, emosi, subjektivitas, penilaian atau pandangan yang terkandung dalam teks (Liu, 2012). Analisis sentimen ulasan produk pada situs *e-commerce* adalah hal yang sulit. Hal ini dikarenakan terdapat berbagai fenomena linguistik yang ditemukan pada ulasan di situs *e-commerce*, antara lain : kesalahan tipografi, singkatan, kata slang, penggunaan bahasa daerah.

Selain itu, ulasan produk di situs *e-commerce* juga memiliki panjang teks yang pendek.

Metode *machine learning* sering digunakan pada analisis sentimen berbahasa Indonesia. *Support Vector Machine* adalah metode *machine learning* dengan performa terbaik dibandingkan metode *machine learning* lainnya. Sutoyo (2020) mengevaluasi kinerja metode *Naïve Bayes*, *Logistic Regression*, dan *Support Vector Machine* dalam analisis sentimen berbahasa Indonesia. Hasil dari evaluasi kinerja menunjukkan bahwa *Support Vector Machine* memiliki kinerja terbaik dengan *accuracy*, *recall*, *precision*, *F-measure* adalah 97.72%, 96.01%, 99.18%, dan 97.57% (Sutoyo & Almaarif, 2020).

Selain menggunakan teknik *machine learning*, *Deep learning* juga bisa digunakan untuk melakukan analisis sentimen. Kim (2014) menunjukkan bahwa metode *Convolutional Neural Network* dengan fitur *Word2Vec* dapat mengungguli beberapa metode dari *machine learning* populer, antara lain: *Support Vector Machine*, *Logistic Regression*, *Multinomial Naïve Bayes*, *Naïve Bayes* (Kim, 2014). *Word2Vec* merupakan bahan utama dari analisis sentimen menggunakan *Convolutional Neural Network* (Kim, 2014). Model *Word2Vec* dengan arsitektur *Skip-Gram* dan ukuran maksimum vektor sebesar 100 memperoleh performa terbaik pada analisis sentimen ulasan produk (Nawang Sari et al., 2019).

Data ulasan produk berbahasa Indonesia yang diperoleh dari situs *e-commerce* memiliki berbagai fenomena linguistik. Peneliti pada umumnya menggunakan tahapan pengolahan khusus untuk mengatasi fenomena yang ada data mereka. Haryanto, Muflikhah dan Fauzi (2018) menggunakan tahapan *Feature*

Expansion untuk mengatasi masalah pada data ulasan dengan panjang teks yang pendek. Tahapan Feature Expansion dengan *API Kateglo* mampu meningkatkan akurasi sebesar 2% (Haryanto et al., 2018). Namun tahapan *FeatureExpansion* tidak signifikan mempengaruhi hasil akurasi apabila menggunakan data berukuran besar (Fauzi, Firmansyah, & Afirianto, 2018). Febrianti, Indriati dan Widodo menggunakan tahapan perbaikan kata untuk mengatasi masalah pada data ulasan yang memiliki kesalahan tipografi kata. Hasil dari analisis sentimen menggunakan koreksi kata lebih unggul dari pada analisis sentimen tanpa koreksi kata (Febrianti et al., 2018).

Tahapan perbaikan kata berfokus untuk menangani kata salah eja yang dibentuk oleh kesalahan tipografi secara umum (Soleh & Purwarianti, 2011). Metode *Probability of Similarity* adalah metode perbaikan kata dengan mengkalkulasi probabilitas kedekatan dari kata yang memiliki kesalahan tipografi dengan kata pada korpus. Metode *Probability of Similarity* punya performa lebih baik dibandingkan dengan metode *edit distance* dengan algoritma *Forward Reversed Dictionary* (Aqsath Rasyid et al., 2011). Metode ini mampu mengukur kedekatan antara kata kandidat dengan kata yang memiliki kesalahan tipografi (Soleh & Purwarianti, 2011).

Berdasarkan referensi dari penelitian sebelumnya, algoritma *Convolutional Neural Network* dan *Word2Vec* akan menjadi metode yang digunakan dalam penelitian analisis sentimen ulasan produk berbahasa Indonesia pada situs *e-commerce*. Metode *Probability of Similarity* akan digunakan dalam menangani kesalahan tipografi kata yang banyak ditemukan pada data ulasan produk yang ada

di situs *e-commerce*. Hasil dari penelitian ini diharapkan bisa digunakan sebagai data referensi dalam pemilihan metode dan tahapan pra-pengolahan data pada analisis sentimen ulasan produk dengan berbagai fenomena linguistik.

1.3 Rumusan Masalah

Berdasarkan latar belakang tersebut maka di dapatlah rumusan masalah sebagai berikut:

1. Bagaimana kinerja metode *Convolutional Neural Network* dengan fitur *Word2Vec* untuk analisis sentimen ulasan produk berbahasa Indonesia?
2. Bagaimana pengaruh perbaikan kata dengan metode *Probability of Similarity* terhadap kinerja analisis sentimen menggunakan metode *Convolutional Neural Network* dengan fitur *Word2Vec*?

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah :

1. Mengetahui kinerja dari metode *Convolutional Neural Network* dengan fitur *Word2Vec* pada analisis sentimen ulasan produk berbahasa Indonesia yang memiliki berbagai fenomena linguistik.
2. Mengetahui pengaruh dari tahapan perbaikan kata dengan metode *Probability of Similarity* terhadap kinerja dari analisis sentimen ulasan produk berbahasa Indonesia menggunakan metode *Convolutional Neural Network* dengan fitur *Word2Vec*.

1.5 Manfaat Penelitian

Manfaat yang bisa didapat dari penelitian ini adalah :

1. Didapatlah informasi terkait kinerja metode *Convolutional Neural Network* untuk analisis sentimen pada ulasan produk berbahasa Indonesia. Informasi ini diharapkan dapat menjadi acuan pembaca untuk memilih metode yang tepat untuk analisis sentimen pada ulasan produk berbahasa Indonesia.
2. Memperoleh informasi mengenai pengaruh dari tahapan perbaikan kata terhadap analisis sentimen menggunakan *Convolutional Neural Network* dengan fitur *Word2Vec*. Informasi ini diharapkan dapat menjadi acuan pembaca untuk memilih tahapan pra-pengolahan data yang tepat pada analisis sentimen ulasan produk.

1.6 Batasan Masalah

Batasan masalah dari penelitian ini adalah :

1. Data yang akan dipakai diperoleh pada *website lazada.co.id* dengan kategori produk : olahraga dan *outdoor*, otomotif, peralatan elektronik, aksesoris elektronik, tas dan aksesoris, kesehatan dan kecantikan, *fashion* wanita, *fashion* pria, bayi dan mainan, TV dan elektronik rumah, kebutuhan rumah tangga, keperluan rumah dan gaya hidup.
2. Data penelitian yang digunakan berjumlah 400.000 dengan masing-masing label berjumlah 200.000.
3. Korpus yang digunakan untuk proses perbaikan kata berjumlah 13000 kata.
4. Parameter ukuran filter yang digunakan pada layer konvolusi CNN bernilai 3,4,5.

5. Parameter maksimum fitur yang digunakan pada layer konvolusi CNN bernilai 100.
6. Parameter *dropout* yang digunakan pada layer dropout CNN bernilai 0,5.
7. Fungsi aktivasi yang digunakan pada layer konvolusi adalah *Rectifier Lienar Unit* (ReLU).
8. Fungsi *optimizer* yang digunakan pada pelatihan CNN adalah SGD dengan *learning rate* 0.001.
9. Arsitektur *Word2Vec* yang digunakan adalah Skip-Gram.
10. Panjang vektor dari *Word2Vec* adalah 100.

1.7 Sistematika Penulisan

Penyusunan skripsi ini disusun dengan sistematika penulisan sebagai berikut:

BAB I. PENDAHULUAN

Menguraikan pokok-pokok pikiran yang mendasari penelitian, seperti latar belakang dari penelitian, rumusan masalah yang diperoleh, tujuan dan manfaat dari penelitian, batasan-batasan masalah yang dibuat serta kerangka penulisan.

BAB II. KAJIAN LITERATUR

Menjelaskan dasar dari teori yang akan dipakai di dalam penelitian ini, meliputi definisi analisis sentimen, tahapan pra-pengolahan data, algoritma *Convolutional Neural Network*, algoritma *Word2Vec*, algoritma *Probability of Similarity*, serta kajian literatur

dari beberapa penelitian lain yang terkait dengan penelitian ini.

BAB III. METODE PENELITIAN

Menjelaskan proses dalam penelitian. Meliputi pengumpulan dan analisis data, mekanisme dari pengujian dan perancangan sistem. Setiap langkah dari rencana studi akan dijabarkan dengan rinci dengan mengacu pada kerangka kerja.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pembahasan tentang analisis dan desain dari perangkat lunak. Pengembangan dan perancangan tersebut meliputi pengumpulan dan analisis kebutuhan, desain dan konstruksi, lalu pengujian untuk melihat kesesuaian hasil dengan perancangan.

BAB V. HASIL DAN ANALISA PENELITIAN

Menguraikan hasil dari pengujian. Hasil dari pengujian dan hasil analisisnya akan disajikan sebagai landasan kesimpulan dalam penelitian.

BAB VI. KESIMPULAN DAN SARAN

Berisi kesimpulan dari penelitian dan juga saran yang diharapkan mampu berguna dalam penelitian lain terkait algoritma *Convolutional Neural Network* dan *Word2Vec* untuk analisis sentimen ulasan produk berbahasa Indonesia.

1.8 Kesimpulan

Telah dijabarkan secara umum gagasan pokok yang mendasari penelitian ini, meliputi latar belakang dari penelitian, rumusan masalah yang diperoleh, tujuan

dan manfaat dari penelitian, batasan-batasan masalah yang dibuat serta kerangka penulisan.

DAFTAR PUSTAKA

- Amanah, D. & Harahap, D.A. 2020. Competitive Advantages of Lazada Indonesia. *European Journal of Business and Management Research*, 5(1).
- Aqsath Rasyid, N., Kamayani, M., Reinanda, R., Simbolon, S., Soleh, M.Y. & Purwarianti, A. 2011. Application of document spelling checker for Bahasa Indonesia. *ICACISIS 2011 - 2011 International Conference on Advanced Computer Science and Information Systems, Proceedings*.
- Dave, K., Lawrence, S. & Pennock, D.M. 2003. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. *Proceedings of the 12th International Conference on World Wide Web, WWW 2003*.
- Edeki, C. 2013. Agile Unified Process. *International Journal of Computer Science and Mobile Applications, IJCSMA*.
- Fauzi, M.A. 2018. Indonesian Review Rating Prediction Using Support Vector Machine and N-Gram. *Journal of Southwest Jiaotong*, 53(5).
- Fauzi, M.A. 2019. *Word2Vec* model for sentiment analysis of product reviews in Indonesian language. *International Journal of Electrical and Computer Engineering (IJECE)*.
- Fauzi, M.A., Firmansyah, R.F.N. & Afirianto, T. 2018. Improving Sentiment Analysis of Short Informal Indonesian Product Reviews using Synonym Based Feature Expansion. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, (Vol 16, No 3: June 2018): 1345–1350. (<http://journal.uad.ac.id/index.php/TELKOMNIKA/article/view/7751>).
- Febrianti, Y.M., Indriati, I. & Widodo, A.W. 2018. Analisis Sentimen Pada Ulasan

- “Lazada ” Berbahasa Indonesia Menggunakan K-Nearest Neighbor (K-NN) Dengan Perbaikan Kata Menggunakan Jaro Winkler Distance. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer.
- Fitri Niasita, A., Adikara, P.P. & Adinugroho, S. 2019. Analisis Sentimen Pembangunan Infrastruktur di Indonesia dengan Automated Lexicon *Word2Vec* dan Naive-Bayes. J-Ptiik.
- Han, J., Kamber, M. & Pei, J. 2012. Data Mining: Concepts and Techniques. Data Mining: Concepts and Techniques.
- Haryanto, D.J., Muflikhah, L. & Fauzi, M.A. 2018. Analisis Sentimen Review Barang Berbahasa Indonesia Dengan Metode Support Vector Machine Dan Query Expansion. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya.
- Ipmawati, J., Kusriani & Taufiq Luthfi, E. 2017. Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen. Indonesian Journal on Networking and Security.
- Kim, Y. 2014. Convolutional neural networks for sentence classification. *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*.
- Liu, B. 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1): 1–184.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. & Dean, J. 2013. Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*.

- Mubarok, F., Harliana, H. & Hadijah, I. 2015. Perbandingan Antara Metode RUP dan Prototype Dalam Aplikasi Penerimaan Siswa Baru Berbasis Web. *Creative Information Technology Journal*.
- Mulyadi, A., Eka, D. & Nailis, W. 2018. Pengaruh Kepercayaan, Kemudahan, dan Kualitas Informasi terhadap Keputusan Pembelian di Toko Online Lazada. *Jembatan*.
- Nawangsari, R.P., Kusumaningrum, R. & Wibowo, A. 2019. *Word2Vec* for Indonesian sentiment analysis towards hotel reviews: An evaluation study. *Procedia Computer Science*.
- Patel, B. & Shah, D. 2013. Significance of stop word elimination in meta search engine. *2013 International Conference on Intelligent Systems and Signal Processing, ISSP 2013*.
- Putra, I.W.S.E., Wijaya, A.Y. & Soelaiman, R. 2016. Klasifikasi Citra Menggunakan Convolutional Neural Network (Cnn) Pada Caltech 101 Image Classification Using Convolution Neural Network (Cnn) on Caltech 101. Institut Teknologi Sepuluh November.
- Rational 1998. Rational Unified Process: Best Practices for Software Development Teams. Rational Software white Paper TP026B.
- Ridok, A. & Latifah, R. 2015. Klasifikasi Teks Bahasa Indonesia Pada Corpus Tak Seimbang Menggunakan NWKNN. Konferensi Nasional Sistem dan Informatika 2015.
- Schneider, K.M. 2005. Techniques for improving the performance of naive bayes for text classification. *Lecture Notes in Computer Science*.

- Seller Center Lazada. Rating dan Ulasan Penjualan. Diakses dari <https://sellercenter.lazada.co.id/seller/helpcenter/tingkatkan-rating-dan-ulasan-untuk-meningkatkan-penjualan-anda!.html>
- Soleh, M.Y. & Purwarianti, A. 2011. A non word error spell checker for Indonesian using morphologically analyzer and HMM. *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics, ICEEI 2011*.
- Sutoyo, E. & Almaarif, A. 2020. Twitter sentiment analysis of the relocation of Indonesia's capital city. *Bulletin of Electrical Engineering and Informatics*.
- Zhang, Y. & Wallace, B.C. 2017. A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification. ArXiv, abs/1510.0.