

***KNOWLEDGE DISCOVERY* TERHADAP ULASAN APLIKASI
MENGUNAKAN PERBANDINGAN *WORD EMBEDDINGS* DENGAN
MODEL CNN-LSTM PADA SENTIMEN ANALISIS**

SKRIPSI

Program Studi Sistem Informasi

Jenjang Sarjana



Oleh :

Vanessa Novalia

09031182126008

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER

UNIVERSITAS SRIWIJAYA

2024

LEMBAR PENGESAHAN

SKRIPSI

***KNOWLEDGE DISCOVERY* TERHADAP ULASAN APLIKASI
MENGUNAKAN PERBANDINGAN *WORD EMBEDDINGS* DENGAN
MODEL CNN-LSTM PADA SENTIMEN ANALISIS**

Sebagai salah satu syarat untuk penyelesaian
studi di Program Studi Sistem Informasi S1

Oleh:

Vanessa Novalia 09031182126008

**Mengetahui,
Ketua Jurusan Sistem Informasi,**



**Ahmad Rifai, S.T., M.T.
NIP. 19791020201021003**

**Palembang, 12 November 2024
Pembimbing,**

A handwritten signature in blue ink, consisting of several overlapping loops and lines.

**Ken Ditha Tania, M.Kom., PhD.
NIP. 198507182012122003**

HALAMAN PERNYATAAN BEBAS PLAGIAT

Yang bertanda tangan di bawah ini:

Nama : Vanessa Novalia
NIM : 09031182126008
Program Studi : Sistem Informasi Regular
Judul Skripsi : *Knowledge Discovery* terhadap Ulasan Aplikasi Menggunakan Perbandingan *Word Embeddings* dengan Model CNN-LSTM pada Sentimen Analisis

Hasil Pengecekan iThenticate/Turnitin: 3%

Menyatakan bahwa laporan skripsi saya merupakan hasil karya saya sendiri dan bukan penjiplakan atau plagiat. Apabila ditemukan unsur penjiplakan atau plagiat dalam laporan skripsi 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.

Palembang, 12 November 2024



Vanessa Novalia
NIM. 09031182126008

HALAMAN PERSETUJUAN

Telah diterima jurnal di International Conference on Electrical Engineering And
Computer Science 2024 pada:

Hari : Kamis
Tanggal : 15 Agustus 2024
Nama : Vanessa Novalia
NIM : 09031182126008
Judul : *Knowledge Discovery of Application Review Using Word
Embedding's Comparison with CNN-LSTM Model on Sentiment
Analysis*

Tim Pembimbing :

1. Pembimbing : Ken Ditha Tania, M.Kom., PhD.



Mengetahui,

Ketua Jurusan Sistem Informasi



Ahmad Rifai, S.T., M.T.

NIP. 19791020201021003

HALAMAN PERSEMBAHAN DAN MOTTO

“Nothing’s Imposibble! Work Hard, Dream Big, Never Give Up! Just Believe in Yourself!”

“Do Your Best and Let God Do the Rest!”

“Dreams Don’t Come True by Wishing – They Come True by Working!”

Skripsi ini dipersembahkan kepada:

- ❖ Tuhan Yang Maha Esa
- ❖ Papa, Mama, Adik, dan Keluarga Penulis Tercinta
- ❖ Dosen Pembimbing, Bu Ken Ditha Tania
- ❖ Dosen Pembimbing Akademik, Bu Putri Eka Sevtiyuni
- ❖ Sahabat dan Teman Seperjuangan
- ❖ Almamaterku, Universitas Sriwijaya
- ❖ Diri Saya Sendiri, Vanessa Novalia

KATA PENGANTAR

Segala puji syukur atas kehadiran Tuhan Yang Maha Esa, atas segala limpahan rahmat, karunia dan petunjuk-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul “***Knowledge Discovery* terhadap Ulasan Aplikasi Menggunakan Perbandingan *Word Embeddings* dengan Model CNN-LSTM pada Sentimen Analisis**” ini dengan baik dan tepat waktu. Tugas Akhir ini disusun sebagai salah satu syarat kelulusan untuk menyelesaikan program studi S1 di Jurusan Sistem Informasi, Fakultas Ilmu Komputer, Universitas Sriwijaya.

Dalam proses penyusunan Tugas Akhir ini, penulis menyadari bahwa terdapat beberapa tantangan yang tidak selalu mudah diatasi. Namun, berkat bimbingan, dukungan, motivasi, serta doa dari banyak pihak, akhirnya penulisan Tugas Akhir ini dapat diselesaikan dengan baik. Oleh karena itu, dengan segala rasa hormat, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Kepada Tuhan Yang Maha Esa, yang telah memberikan petunjuk dan rahmat-Nya kepada penulis, sehingga penulis dapat memiliki kesempatan untuk mencapai titik ini dan menyelesaikan pendidikan ini dengan baik.
2. Kepada kedua orang tua tercinta, Mama dan Papa, serta saudara penulis, Melissa, yang telah menjadi sumber inspirasi dan motivasi utama dalam perjalanan ini. Terima kasih atas segala doa, semangat, dan dukungan tulus yang tak pernah berhenti mengalir. Setiap langkah dan pencapaian ini tak lepas dari pengorbanan yang telah kalian berikan, baik dalam bentuk kasih sayang, waktu, maupun tenaga. Tanpa kalian, semua ini tak akan mungkin terwujud.
3. Bapak Prof. Dr. Erwin, S.Si., M.Si. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.

4. Bapak Ahmad Rifai, S.T., M.T. selaku Ketua Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Ibu Ken Ditha Tania, M.Kom., PhD. selaku Dosen Pembimbing Tugas Akhir penulis yang telah memberikan bimbingan, arahan dan dukungan penuh selama proses penyusunan tugas akhir ini. Penulis sangat menghargai ilmu, ide-ide, dan nasihat yang telah diberikan, yang sangat membantu dalam menyelesaikan tugas akhir ini dengan baik.
6. Alm. Dr. Jaidan Jauhari, M.T. selaku Dosen Pembimbing Akademik penulis yang telah membimbing penulis sejak awal semester hingga semester empat.
7. Ibu Putri Eka Sevtiyuni, S.SI., M.T selaku Dosen Pembimbing Akademik penulis yang telah membimbing penulis sejak semester lima hingga akhir perkuliahan.
8. Mba Ayu, selaku Admin Jurusan Sistem Informasi yang telah dengan sabar dan sigap membantu penulis dalam mengurus semua administrasi dan pemberkasan penulis di akhir semester ini.
9. Seluruh dosen, admin jurusan, dan staf pegawai yang telah membantu, memberikan banyak ilmu, dan arahan selama penulis menempuh pendidikan di Fakultas Ilmu Komputer Universitas Sriwijaya.
10. Sahabat dekat dan seperjuanganku ABDH (Hesti, Nanda, Fanny, Aida, Jess) yang selalu menjadi pendukung setia, teman berbagi cerita, dan seperjuangan dalam setiap langkah. Perjalanan ini terasa lebih ringan berkat kebersamaan dan segala momen yang kita lewati bersama. Terima kasih telah menjadi sumber inspirasi, motivasi, dan keceriaan dalam setiap langkah yang kita ambil. Tanpa kalian, perjalanan ini tidak akan terasa sama.

11. Sahabat kecilku, Christi, yang sudah seperti saudara tak sedarah, yang telah menemani setiap langkah penulis sejak kecil, memberikan dukungan, keceriaan dan kebersamaan yang tak ternilai. Meskipun jarak memisahkan kita, semoga kita selalu bisa tetap saling *support each other and create beautiful memories together. Thank you for being an inseparable part of my life journey up until now!*
12. Sahabat SMA penulis, MSF (Christi, Carol, Siti), yang hingga saat ini masih menjadi bagian dari cerita hidup penulis. Terima kasih telah menjadi sumber semangat, inspirasi, dan keceriaan dalam setiap langkah. Kalian adalah sosok yang penuh ambisi dan selalu mendorong penulis untuk menjadi lebih baik. *Your presence in my life means a lot, and I am grateful that you guys can still continue to support each other despite the passing of time.*
13. Teman seperjuangan Studi Independen dan magang penulis, Najwa, yang telah menemani perjuangan dalam pembuatan project Bangkit Academy sebagai Kerja Praktik, hingga akhirnya dapat meraih *best project* di *company track* dan mendapatkan *offering* magang di Braincore. Tanpamu semua ini pasti akan terasa lebih sulit. Terima kasih telah menjadi rekan kerja yang luar biasa, menyenangkan bisa bekerja denganmu dan berkolaborasi dengan sangat baik.
14. Teman yang menemani masa semester akhir penulis, Danti, yang telah menjadi bagian dari perjalanan di akhir semester ini terutama telah menjadi sosok yang mendorong penulis untuk berjuang dalam tes sulit. Terima kasih atas dukungan, kebersamaan, dan waktu yang menyenangkan selama ini.
15. Teman-teman SI Reguler B 2021 yang telah bersama-sama berjuang dan mendukung selama perkuliahan.

16. Kepada diri sendiri yang senantiasa kuat dan ingin terus berusaha, meski dihadapi berbagai tantangan. *It wasn't easy, but you made it through. Keep moving forward!*

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan dan memerlukan pengembangan lebih lanjut. Oleh karena itu, penulis sangat terbuka untuk menerima kritik dan saran yang membangun demi perbaikan di masa mendatang. Penulis juga berharap skripsi ini dapat memberikan manfaat dalam menambah ilmu pengetahuan serta menjadi referensi untuk penelitian-penelitian serupa di masa yang akan datang.

Palembang, 12 November 2024

Penulis,



Vanessa Novalia

NIM. 09031182126008

***KNOWLEDGE DISCOVERY* TERHADAP ULASAN APLIKASI
MENGUNAKAN PERBANDINGAN *WORD EMBEDDINGS* DENGAN
MODEL CNN-LSTM PADA SENTIMEN ANALISIS**

Oleh:

Vanessa Novalia

09031182126008

ABSTRAK

Pertumbuhan *e-commerce* yang pesat di Indonesia mencerminkan akselerasi ekonomi negara ini dan memberikan berbagai manfaat. Salah satunya adalah TikTok Shop, yang telah menjadi platform populer untuk kegiatan ekonomi digital yang menawarkan peluang bagi para pelaku bisnis dan konsumen. Untuk memaksimalkan peluang tersebut, perusahaan harus secara efektif memahami dan mengelola sentimen pengguna untuk meningkatkan layanan dan pengalaman pengguna. Analisis sentimen yang akurat memainkan peran penting dalam mencapai tujuan ini. Penelitian ini berfokus pada perbandingan efektivitas berbagai metode *word embedding* dalam meningkatkan analisis sentimen ulasan pengguna TikTok Shop. Penelitian ini menerapkan pendekatan gabungan CNN-LSTM, membandingkan *classic pre-trained* seperti GloVe dengan *custom-trained embeddings* menggunakan Word2Vec dan FastText, yang dilatih menggunakan data spesifik domain. Tujuan dari penelitian ini adalah untuk mengidentifikasi metode penyematan yang paling efektif untuk menangkap sentimen pengguna dan memberikan wawasan yang lebih akurat. Hasil temuan menunjukkan bahwa teknik yang dilatih secara khusus yaitu Word2Vec, memberikan performa terbaik, dengan mencapai akurasi, presisi, F1-score, dan nilai AUC-ROC tertinggi. Sementara FastText berada di peringkat kedua dengan kinerja yang hampir sama, hanya berbeda sedikit. Sebaliknya, kemampuan *pre-trained* GloVe untuk menangkap sentimen pengguna tidak sekuat Word2Vec dan FastText yang telah dilatih secara khusus, meskipun secara keseluruhan tetap cukup efektif.

Kata Kunci: *Sentiment Analysis, Tiktok Shop, Word Embedding, CNN-LSTM*

**KNOWLEDGE DISCOVERY OF APPLICATION REVIEW USING WORD
EMBEDDINGS COMPARISON WITH CNN-LSTM MODEL ON
SENTIMENT ANALYSIS**

By:

Vanessa Novalia

09031182126008

ABSTRACT

The rapid growth of e-commerce in Indonesia reflects the country's economic acceleration and provides a variety of benefits. One of them is TikTok Shop, which has become a popular platform for digital economic activities that offers opportunities for businesses and consumers. To maximize these opportunities, companies must effectively understand and manage user sentiment to improve services and user experience. Accurate sentiment analysis plays a crucial role in achieving this goal. This research focuses on comparing the effectiveness of various word embedding methods in enhancing sentiment analysis of TikTok Shop user reviews. The study applies a combined CNN-LSTM approach, comparing classic pre-trained embeddings like GloVe with custom-trained embeddings using Word2Vec and FastText, which are trained on domain-specific data. The objective of this research is to identify the most effective embedding method for capturing user sentiment and providing more accurate insights. The findings show that the custom-trained technique, Word2Vec, delivered the best performance, achieving the highest accuracy, precision, F1-score, and AUC-ROC values. FastText ranked second with nearly the same performance, differing only slightly. In contrast, GloVe pre-trained ability to capture user sentiment is not as strong as that of custom-trained Word2Vec and FastText, although it remains quite effective overall.

Keywords: *Sentiment Analysis, Tiktok Shop, Word Embedding, CNN-LSTM*

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PENGESAHAN	ii
HALAMAN PERNYATAAN BEBAS PLAGIAT.....	iii
HALAMAN PERSETUJUAN	iv
HALAMAN PERSEMBAHAN DAN MOTTO.....	v
KATA PENGANTAR.....	vi
ABSTRAK	x
ABSTRACT	xi
DAFTAR ISI.....	xii
DAFTAR GAMBAR.....	xiv
DAFTAR TABEL	xv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian.....	5
1.4 Manfaat Penelitian.....	5
1.5 Batasan Masalah.....	6
BAB II TINJAUAN PUSTAKA.....	7
2.1 Landasan Teori	7
2.1.1 <i>Knowledge Discovery</i>	7
2.1.2 <i>Text Mining</i>	7
2.1.3 Analisis Sentimen	8
2.1.4 Aplikasi Tiktok Shop Seller Center	9
2.1.5 <i>Word Embedding</i>	10
2.1.6 <i>Deep learning</i>	11
2.1.7 <i>Confusion Matrix</i>	12
2.2 Penelitian Terdahulu	15
BAB III METODOLOGI PENELITIAN	18
3.1 Objek Penelitian	18
3.2 Alat dan Bahan	18
3.3 Tahapan Penelitian.....	19
3.3.1 <i>Data Scrapping</i>	19

3.3.2	<i>Data Labelling</i>	19
3.3.3	<i>Data Preprocessing</i>	20
3.3.4	<i>Data Splitting</i>	21
3.3.5	<i>Model Building</i>	22
3.3.6	<i>Model Evaluation</i>	24
BAB IV HASIL DAN PEMBAHASAN		26
4.1	Hasil Proses Pengumpulan Data	26
4.2	Hasil <i>Pre-Processing</i> Data dan <i>Text Analytics</i>	27
4.2.1	<i>Case Folding</i>	27
4.2.2	<i>Cleansing Text Using RegEx</i>	28
4.2.3	<i>Slang Word</i>	28
4.2.4	<i>Tokenizing</i>	29
4.2.5	<i>Stemming</i>	29
4.3	Hasil Implementasi <i>Word Embedding</i> Terhadap Model	32
4.3.1	Hasil <i>Word Embedding</i> Word2Vec	32
4.3.2	Hasil <i>Word Embedding</i> GloVe	33
4.3.3	Hasil <i>Word Embedding</i> FastText	34
4.4	Hasil Evaluasi	34
4.4.1	Hasil Evaluasi Word2Vec	35
4.4.2	Hasil Evaluasi GloVe	36
4.4.3	Hasil Evaluasi FastText	37
BAB V KESIMPULAN DAN SARAN		41
5.1	Kesimpulan	41
5.2	Saran	42
DAFTAR PUSTAKA		43

DAFTAR GAMBAR

Gambar 2.1 Logo Tiktok Shop	9
Gambar 3.1 Metodologi Penelitian.....	19
Gambar 3.2 <i>Model Architecture</i>	22
Gambar 3.3 Kode <i>Model Building</i>	23
Gambar 3.4 Kode <i>Embedding Word2Vec</i>	24
Gambar 3.5 Kode <i>Embedding GloVe</i>	24
Gambar 3.6 Kode <i>Embedding FastText</i>	24
Gambar 3.7 Kode Evaluasi Model	25
Gambar 3.8 Kode ROC-AUC.....	25
Gambar 4.1 Hasil <i>Scrapping dan Labelling Data</i>	26
Gambar 4.2 Distribusi Data	27
Gambar 4.3 <i>WordCloud</i> Sentimen Positif.....	30
Gambar 4.4 <i>WordCloud</i> Sentimen Negatif	31
Gambar 4.5 Distribusi Sentimen dari Waktu ke Waktu	31
Gambar 4.6 Hasil Akurasi Model <i>Word Embedding Word2Vec</i>	32
Gambar 4.7 Hasil Akurasi Model <i>Word Embedding GloVe</i>	33
Gambar 4.8 Hasil Akurasi Model <i>Word Embedding FastText</i>	34
Gambar 4.9 Hasil Evaluasi <i>Confusion Matrix</i>	34
Gambar 4.10 Hasil <i>Classification Report</i> Word2Vec	35
Gambar 4.11 Hasil <i>Classification Report</i> GloVe.....	36
Gambar 4.12 Hasil <i>Classification Report</i> FastText	37
Gambar 4.13 Perbandingan Kurva ROC Ketiga Model	38
Gambar 4.14 Perbandingan Performa Metriks Ketiga Model	39

DAFTAR TABEL

Table 2.1 <i>Confusion Matrix</i>	13
Table 2.2 Penelitian Sebelumnya.....	15
Table 3.1 Pembagian Data	22
Table 4.1 Hasil Implementasi <i>Case Folding</i>	27
Table 4.2 Hasil Implementasi <i>Cleaning Data</i> Menggunakan <i>RegEx</i>	28
Table 4.3 Hasil Implementasi <i>Slang Word</i>	28
Table 4.4 Hasil Implementasi <i>Tokenizing</i>	29
Table 4.5 Hasil Akhir Setelah Tahap <i>Pre-processing</i>	29

DAFTAR LAMPIRAN

Lampiran 1 Surat Kesiediaan Membimbing.....	A-1
Lampiran 2 Form Pengajuan Topik Skripsi	B-1
Lampiran 3 Surat Keputusan Pembimbing Tugas Akhir.....	C-1
Lampiran 4 Form Desk Evaluation	D-1
Lampiran 5 Hasil Pengecekan Similarity	E-1
Lampiran 6 Surat Pernyataan Similarity	F-1
Lampiran 7 Sampel Dataset Awal Penelitian	G-1
Lampiran 8 Sampel Dataset Hasil <i>Pre-processing Data</i>	H-1
Lampiran 9 Kartu Konsultasi	I-1
Lampiran 10 Log Book Dosen Pembimbing Skripsi	J-1
Lampiran 11 <i>Letter of Acceptance (LoA)</i>	K-1
Lampiran 12 Korespondensi	L-1

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pertumbuhan *e-commerce* yang kuat di Indonesia mencerminkan percepatan ekonomi negara yang memberikan berbagai keuntungan. *E-commerce* mengacu pada kegiatan transaksi ekonomi digital yang memungkinkan pelanggan dan penjual melakukan transaksi secara *online* dengan memanfaatkan perangkat teknologi tertentu melalui internet (Yin et al., 2022)(Ahmad et al., 2023). Banyaknya data yang dihasilkan dari interaksi pengguna dan transaksi *online* menyoroti peran penting *knowledge discovery*, karena data tersebut mengandung pengetahuan tersembunyi yang berharga yang dapat bermanfaat bagi peningkatan ekonomi dan pengelolaan sistem yang kompleks baik di tingkat individu maupun organisasi (Bagheri et al., 2023).

Text mining adalah penerapan dari berbagai teknik pemrosesan analitik untuk mengekstrak pola-pola signifikan dan menarik dari data yang tidak terstruktur, yang memungkinkan eksplorasi pengetahuan tersembunyi dalam sumber data tekstual (Razzaq & Basak, 2022), biasanya melalui tugas-tugas seperti *text classification* , *text clustering*, ekstraksi informasi dan pengambilan informasi (Yan et al., 2022). Dalam konteks ini, klasifikasi teks seperti analisis sentimen, banyak digunakan untuk menganalisis ulasan produk karena memberikan wawasan mendalam tentang umpan balik pelanggan (Huang et al., 2023).

Berbagai platform *e-commerce* menyediakan kesempatan bagi pengguna untuk menyampaikan tanggapan mengenai produk dan layanan mereka, sehingga dapat memperoleh informasi berharga seperti spesifikasi produk dari perspektif

pengguna, baik yang positif maupun negatif, yang penting untuk memahami sentimen pengguna (Daza et al., 2024). Menganalisis opini pengguna ini berperan penting dalam pertumbuhan global perusahaan, menjadikan pengumpulan umpan balik sebagai strategi kuat untuk mengevaluasi dan memahami ulasan pelanggan (Alzahrani et al., 2022).

Salah satu platform media sosial yang juga berfungsi sebagai platform *e-commerce* populer saat ini yaitu TikTok Shop, yang diperkenalkan pada 17 April 2021 sebagai fitur tambahan dari aplikasi Tiktok (Rahmania et al., 2024). Fitur ini sangat menarik bagi para ekonom yang melakukan aktivitas ekonomi digital di platform tersebut (Sudiantini et al., 2024). Dengan basis pengguna terbesar secara global (Putri et al., 2024), yang mencapai sekitar 143 juta pengguna aktif bulanan di Indonesia per Februari 2024 melalui kemitraan strategisnya dengan Tokopedia (Adi Ahdiat, 2024), Tiktok Shop menghasilkan data ulasan yang kaya akan informasi sentimen. Analisis sentimen adalah bagian dari pengolahan data berbasis teks yang mengevaluasi opini, perilaku, dan emosi seseorang untuk memperoleh informasi tentang sentimen yang bernilai positif atau negatif (Anjani & Irmanda, 2024). Namun, dalam pendekatan analisis sentimen ini terdapat tantangan signifikan dalam mengidentifikasi sentimen sebuah kalimat yang akurat (Geetha & Renuka, 2021).

Salah satu pendekatan dalam pemrosesan bahasa alami (NLP) yang telah terbukti dapat meningkatkan kinerja analisis sentimen adalah representasi berbasis *word embedding*, yang merepresentasikan teks ke dalam ruang vektor berdimensi tetap (Onan, 2021). *Word embedding* digunakan sebagai teknik untuk mengubah kata-kata menjadi representasi vektor numerik multidimensi yang terdistribusi

untuk menangkap makna kata-kata (Asudani et al., 2023) , agar dapat diproses oleh model *machine learning*. Beberapa metode *word embedding* yang populer, seperti Word2Vec, Glove, dan FastText, telah banyak digunakan dalam berbagai studi. Namun, efektivitas dari masing-masing teknik ini dapat bervariasi tergantung pada konteks dan jenis data yang digunakan.

Dalam penelitian sebelumnya yang berkaitan dengan *word embedding*, (Santosh Kumar et al., 2021) menekankan pada perbandingan metode *word embedding*, baik yang sudah dilatih (*pre-trained*) sebelumnya maupun yang belum dilatih (*untrained*), dalam model *deep learning*. Para peneliti menyoroti bahwa *embedding* yang sudah dilatih sebelumnya umumnya menghasilkan hasil yang lebih baik dibandingkan dengan *embedding* yang belum dilatih (*default Keras*), yang memperkuat pentingnya memanfaatkan pengetahuan yang telah tertanam dalam model yang telah dilatih sebelumnya (*pre-trained*). Namun, studi tersebut tidak meneliti potensi manfaat dari pelatihan *embedding* pada data khusus domain (*custom-trained*), sehingga meninggalkan celah dalam pemahaman tentang bagaimana *embedding* yang dilatih khusus mungkin berkinerja dibandingkan dengan model yang sudah dilatih sebelumnya.

Sebuah studi oleh (Sabbeh & Fasihuddin, 2023) membandingkan kinerja teknik *word embedding* klasik (GloVe, Word2Vec, FastText) dan kontekstual (*ARBERT*) menggunakan arsitektur BiLSTM dan CNN. Hasil penelitian menunjukkan bahwa *embedding* kontekstual memiliki kinerja terbaik, dan *embedding* yang dilatih khusus umumnya lebih unggul daripada versi yang sudah dilatih sebelumnya. Juga ditemukan bahwa BiLSTM mengungguli CNN pada dataset besar, sementara CNN lebih efektif pada dataset yang lebih kecil. Namun,

para peneliti tidak secara mendalam mengeksplorasi efektivitas *embedding* klasik yang dilatih khusus *versus* yang sudah dilatih sebelumnya dalam arsitektur *neural* gabungan antara CNN dan LSTM.

Berdasarkan temuan tersebut, penelitian ini bertujuan untuk mengeksplorasi lebih lanjut terkait efektivitas teknik penyematan kata dengan menerapkan pendekatan gabungan CNN-LSTM pada analisis sentimen dengan menggunakan data pada aplikasi *e-commerce* TikTok Shop Seller Center. Penelitian ini berfokus pada perbandingan efektivitas metode *word embedding* klasik dengan membandingkan model GloVe yang telah dilatih sebelumnya (*pre-trained*) dengan penyematan yang dilatih khusus (*custom-trained*) pada data spesifik domain menggunakan Word2Vec dan FastText. Hipotesis dari penelitian ini adalah bahwa penyematan yang dilatih secara khusus akan mengungguli model yang telah dilatih sebelumnya dalam menangkap nuansa sentimen, sehingga memberikan wawasan yang lebih dalam dan lebih akurat ke dalam ulasan pengguna.

Berdasarkan penjelasan yang telah diuraikan di atas, penulis ingin melakukan penelitian lebih lanjut tentang penerapan *word embedding* menggunakan model *deep learning* dengan judul **”*Knowledge Discovery* terhadap Ulasan Aplikasi Menggunakan Perbandingan *Word Embeddings* dengan Model CNN-LSTM pada Sentimen Analisis”**

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang di atas, maka rumusan masalah dalam penelitian ini diperoleh sebagai berikut:

1. Bagaimana mendapatkan model CNN-LSTM yang optimal dengan menggunakan teknik *word embedding* untuk mengklasifikasikan sentimen ulasan pada aplikasi Tiktok Shop?
2. Bagaimana hasil analisis sentimen terhadap kepuasan pengguna Tiktok Shop pada Google Playstore?
3. Penerapan metode *word embedding* manakah yang paling baik dalam penerapan metode CNN-LSTM pada analisis sentimen terhadap kepuasan pengguna Tiktok Shop?

1.3 Tujuan Penelitian

Adapun tujuan yang ingin diperoleh dari penelitian ini adalah sebagai berikut:

1. Memperoleh model CNN-LSTM yang optimal dengan penerapan *word embedding* yang tepat pada ulasan aplikasi Tiktok Shop.
2. Mengetahui hasil kinerja dari metode CNN-LSTM terhadap sentimen pengguna Tiktok Shop pada Google Play Store.
3. Membandingkan kinerja teknik *word embedding* dengan metode *deep learning* pada aplikasi Tiktok Shop.

1.4 Manfaat Penelitian

1. Penelitian ini dapat memperkaya literatur akademik bagi peneliti lain yang tertarik dalam mengembangkan model *deep learning*, khususnya CNN-LSTM, untuk analisis sentimen, dengan pemahaman yang lebih baik tentang metode *word embedding* yang optimal.
2. Meningkatkan akurasi klasifikasi sentimen ulasan TikTok untuk pemahaman pengguna yang lebih baik.

3. Dengan menemukan kombinasi terbaik antara model CNN-LSTM dan teknik *word embedding*, penelitian ini dapat membantu meningkatkan akurasi analisis sentimen, yang sangat bermanfaat bagi pengembang aplikasi dan pengambil keputusan dalam memahami kepuasan pengguna.

1.5 Batasan Masalah

Batasan masalah dalam penelitian ini ditetapkan agar pembahasan tetap terfokus pada lingkup permasalahan yang telah ditentukan, yaitu sebagai berikut:

1. Data penelitian diambil dari ulasan aplikasi Tiktok Shop pada Google Playstore menggunakan teknik *Web Scrapping* untuk periode Januari 2023 sampai dengan Juli 2024.
2. Data yang diolah hanya data teks yang berbahasa Indonesia.
3. Sentimen yang dianalisis terbatas pada dua kategori kelas yaitu sentimen positif dan negatif
4. Parameter yang diuji dalam menentukan model terbaik adalah metode *word embedding* yang digunakan.

DAFTAR PUSTAKA

- Adi Ahdiat. (2024, March 19). *TikTok Shop Gabung Tokopedia, Ini Jumlah Penggunanya*. Databooks. <https://databoks.katadata.co.id/datapublish/2024/03/19/tiktok-shop-gabung-tokopedia-ini-jumlah-penggunanya>
- Aghzal, M., & Mourhir, A. (2021). Distributional word representations for code-mixed text in Moroccan Darija. *Procedia Computer Science*, 189, 266–273.
- Ahmad, A., Digdowiseiso, K., & Lantana, D. A. (2023). The Determinants of E-Commerce Development in Indonesia: a Bibliometrical Analysis. *INTERNATIONAL JOURNAL OF ECONOMICS, MANAGEMENT, BUSINESS, AND SOCIAL SCIENCE (IJEMBIS)*, 3(3), 984–993.
- Al-Ghifari, M. L., & Tania, K. D. (2023). Sentiment Analysis Performance Value Optimization Using Hyperparamater Tunning With Grid Search On Shopee App Reviews. *Indonesian Journal of Computer Science*, 12(5).
- Alzahrani, M. E., Aldhyani, T. H. H., Alsubari, S. N., Althobaiti, M. M., & Fahad, A. (2022). Developing an Intelligent System with Deep Learning Algorithms for Sentiment Analysis of E-Commerce Product Reviews. *Computational Intelligence and Neuroscience*, 2022(1), 3840071.
- Angraini, N. A., & Lhaksana, K. M. (2024). Sentiment Analysis About Legislative Elections using Deep Learning with LSTM and CNN Models. *Building of Informatics, Technology and Science (BITS)*, 6(1), 289–299.
- Anjani, M., & Irmanda, H. N. (2024). COMPARISON PERFORMANCE OF WORD2VEC, GLOVE, FASTTEXT USING SUPPORT VECTOR

MACHINE METHOD FOR SENTIMENT ANALYSIS. *Jurnal Teknik Informatika (Jutif)*, 5(3), 669–674.

Asshiddiqi, M. F., & Lhaksana, K. M. (2020). Perbandingan Metode Decision Tree dan Support Vector Machine untuk Analisis Sentimen pada Instagram Mengenai Kinerja PSSI. *EProceedings of Engineering*, 7(3).

Asudani, D. S., Nagwani, N. K., & Singh, P. (2023). Impact of word embedding models on text analytics in deep learning environment: a review. *Artificial Intelligence Review*, 56(9), 10345–10425.

Bagheri, A., Taghvaeian, S., & Delen, D. (2023). A text analytics model for agricultural knowledge discovery and sustainable food production: A case study from Oklahoma Panhandle. *Decision Analytics Journal*, 9, 100350. <https://doi.org/https://doi.org/10.1016/j.dajour.2023.100350>

Bimantara, M. D., & Zufria, I. (2024). Text Mining Sentiment Analysis on Mobile Banking Application Reviews using TF-IDF Method with Natural Language Processing Approach. *JINAV: Journal of Information and Visualization*, 5(1), 115–123.

Chandra, S., Gourisaria, M. K., Harshvardhan, G. M., Rautaray, S. S., Pandey, M., & Mohanty, S. N. (2021). Semantic Analysis of Sentiments through Web-Mined Twitter Corpus. *ISIC*, 122–135.

Chen, Y., & Zhang, D. (2022). Integration of knowledge and data in machine learning. *ArXiv Preprint ArXiv:2202.10337*.

Daza, A., Rueda, N. D. G., Sánchez, M. S. A., Espíritu, W. F. R., & Quiñones, M. E. C. (2024). Sentiment Analysis on E-Commerce Product Reviews Using Machine Learning and Deep Learning Algorithms: A Bibliometric

- Analysis and Systematic Literature Review, Challenges and Future Works. *International Journal of Information Management Data Insights*, 4(2), 100267.
- Foody, G. M. (2023). Challenges in the real world use of classification accuracy metrics: From recall and precision to the Matthews correlation coefficient. *Plos One*, 18(10), e0291908.
- Frank, E., Oluwaseyi, J., & Olaoye, G. (2024). *Data preprocessing techniques for NLP in BI*.
- Geetha, M. P., & Renuka, D. K. (2021). Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model. *International Journal of Intelligent Networks*, 2, 64–69.
- Hasanah, K. (2024). Comparison of Sentiment Analysis Model for Shopee Comments on Google Play Store. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 13(1), 21–30.
- Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. (2022). Text preprocessing for text mining in organizational research: Review and recommendations. *Organizational Research Methods*, 25(1), 114–146.
- Hindocha, E., Yazhiny, V., Arunkumar, A., & Boobalan, P. (2019). Short-text Semantic Similarity using GloVe word embedding. *International Research Journal of Engineering and Technology*, 6(4).
- Hossny, A. H., Mitchell, L., Lothian, N., & Osborne, G. (2020). Feature selection methods for event detection in Twitter: a text mining approach. *Social Network Analysis and Mining*, 10, 1–15.

- Huang, H., Asemi, A., & Mustafa, M. B. (2023). Sentiment analysis in e-commerce platforms: A review of current techniques and future directions. *IEEE Access*.
- Ibias, A., Antona, H., Ramirez-Miranda, G., & Guinovart, E. (2024). Knowledge Discovery using Unsupervised Cognition. *ArXiv Preprint ArXiv:2409.20064*.
- Islam, M. S., Kabir, M. N., Ghani, N. A., Zamli, K. Z., Zulkifli, N. S. A., Rahman, M. M., & Moni, M. A. (2024). Challenges and future in deep learning for sentiment analysis: a comprehensive review and a proposed novel hybrid approach. *Artificial Intelligence Review*, 57(3), 62.
- Jeffrey Pennington, R. S. C. D. M. (2014). *GloVe: Global Vectors for Word Representation*. <https://nlp.stanford.edu/projects/glove/>
- Kathuria, A., Gupta, A., & Singla, R. K. (2021). A review of tools and techniques for preprocessing of textual data. *Computational Methods and Data Engineering: Proceedings of ICMDE 2020, Volume 1*, 407–422.
- Kumar, S., Singh, A. K., Bhushan, S., & Singh, V. K. (2024). A cross CNN-LSTM model for sarcasm identification in sentiment analysis. *Artificial Intelligence, Blockchain, Computing and Security Volume 1*, 322–328.
- Mewada, A., & Dewang, R. K. (2023). SA-ASBA: a hybrid model for aspect-based sentiment analysis using synthetic attention in pre-trained language BERT model with extreme gradient boosting. *The Journal of Supercomputing*, 79(5), 5516–5551.
- Mittal, D., & Agrawal, S. R. (2022). Determining banking service attributes from online reviews: text mining and sentiment analysis. *International Journal of Bank Marketing*, 40(3), 558–577.

- Murfi, H., Gowandi, T., Ardaneswari, G., & Nurrohmah, S. (2024). BERT-based combination of convolutional and recurrent neural network for Indonesian sentiment analysis. *Applied Soft Computing, 151*, 111112.
- Nguyen, H., Veluchamy, A., Diop, M., & Iqbal, R. (2018). Comparative study of sentiment analysis with product reviews using machine learning and lexicon-based approaches. *SMU Data Science Review, 1(4)*, 7.
- Onan, A. (2021). Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks. *Concurrency and Computation: Practice and Experience, 33(23)*, e5909.
- Pambudi, A., & Suprpto, S. (2021). Effect of Sentence Length in Sentiment Analysis Using Support Vector Machine and Convolutional Neural Network Method. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems), 15(1)*, 21–30.
- Putri, N., Prasetya, Y., Handayani, P. W., & Fitriani, H. (2024). TikTok Shop: How trust and privacy influence generation Z's purchasing behaviors. *Cogent Social Sciences, 10(1)*, 2292759.
- Rahmania, K. N., Ernawati, S., & Badar, M. (2024). Pengaruh Sikap dan Motivasi Pada Minat Beli di Tiktok Shop. *ECo-Fin, 6(2)*, 331–340.
- Razzaq, M. A., & Basak, T. (2022). Text mining in unstructured text: techniques, methods and analysis. *World Scientific News, 174*, 68–84.
- Rehman, A. U., Malik, A. K., Raza, B., & Ali, W. (2019). A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis. *Multimedia Tools and Applications, 78*, 26597–26613.

- Sabbeh, S. F., & Fasihuddin, H. A. (2023). A Comparative analysis of word embedding and deep learning for Arabic sentiment classification. *Electronics*, *12*(6), 1425.
- Santosh Kumar, P., Yadav, R. B., & Dhavale, S. V. (2021). A comparison of pre-trained word embeddings for sentiment analysis using deep learning. *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2020, Volume 1*, 525–537.
- Solikah, M., & Kusumaningtyas, D. (2022). Tik Tok shop: Quality system and marketing mix on consumer satisfaction of online shopping. *International Conference of Business and Social Sciences*, 877–887.
- Sudiantini, D., Yunita, E., Fauzia, N., Ramadhania, R., & Yuliani, Z. (2024). ANALYSIS OF THE COLLABORATIVE MARKETING STRATEGY OF THE TOKOPEDIA AND TIKTOK SHOP PLATFORMS IN INCREASING SALES. *SENTRI: Jurnal Riset Ilmiah*, *3*(6), 2804–2812.
- Suraj, D., Dinesh, S., Balaji, R., Deepika, P., & Ajila, F. (2023). Deciphering Product Review Sentiments Using BERT and TensorFlow. *FMDB Transactions on Sustainable Computing Systems*, *1*(2), 77–88.
- Vakili, M., Ghamsari, M., & Rezaei, M. (2020). Performance analysis and comparison of machine and deep learning algorithms for IoT data classification. *ArXiv Preprint ArXiv:2001.09636*.
- Wang, C., Nulty, P., & Lillis, D. (2020). A comparative study on word embeddings in deep learning for text classification. *Proceedings of the 4th International Conference on Natural Language Processing and Information Retrieval*, 37–46.

- Yan, H., Ma, M., Wu, Y., Fan, H., & Dong, C. (2022). *Overview and analysis of the text mining applications in the construction industry. Heliyon, 8 (12), e12088.*
- Yani, D. D. A., Pratiwi, H. S., & Muhandi, H. (2019). Implementasi web scraping untuk pengambilan data pada situs marketplace. *JUSTIN (Jurnal Sistem Dan Teknologi Informasi), 7(4), 257–262.*
- Yin, J. Y. B., Saad, N. H. M., & Yaacob, Z. (2022). Exploring Sentiment Analysis on E-Commerce Business: Lazada and Shopee. *Tem Journal, 11(4), 1508–1519.*