KOMPARASI DETEKSI KEKUATAN SENTIMEN PADA ULASAN PRODUK MCDONALD’S: PENDEKATAN CNN, LSTM, DAN BI-LSTM DENGAN LEKSIKON SSS-LEX DAN VADER

****

SKRIPSI

**Dava Virgio Kertawijaya**

**00000056848**

**PROGRAM STUDI SISTEM INFORMASI**

**FAKULTAS TEKNIK DAN INFORMATIKA**

**UNIVERSITAS MULTIMEDIA NUSANTARA**

**TANGERANG**

**2023**

KOMPARASI DETEKSI KEKUATAN SENTIMEN PADA ULASAN PRODUK MCDONALD’S: PENDEKATAN CNN, LSTM, DAN BI-LSTM DENGAN LEKSIKON SSS-LEX DAN VADER

****

SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Sistem Informasi

**Dava Virgio Kertawijaya**

**00000056848**

**PROGRAM STUDI SISTEM INFORMASI**

**FAKULTAS TEKNIK DAN INFORMATIKA**

**UNIVERSITAS MULTIMEDIA NUSANTARA**

**TANGERANG**

**2023**

# **HALAMAN PERNYATAAN TIDAK PLAGIAT**

Dengan ini saya,

Nama : Dava Virgio Kertawijaya

Nomor Induk Mahasiswa : 00000056848

Program studi : Sistem Informasi

Skripsi dengan judul:

“KOMPARASI DETEKSI KEKUATAN SENTIMEN PADA ULASAN PRODUK MCDONALD’S: PENDEKATAN CNN, LSTM, DAN BI-LSTM DENGAN LEKSIKON SSS-LEX DAN VADER”

merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas Akhir yang telah saya tempuh.

Tangerang, 29 November 2023

\*materai Rp 10.000,00

Dava Virgio Kertawijaya

# **HALAMAN PERSETUJUAN**

Skripsi dengan judul

“KOMPARASI DETEKSI KEKUATAN SENTIMEN PADA ULASAN PRODUK MCDONALD’S: PENDEKATAN CNN, LSTM, DAN BI-LSTM DENGAN LEKSIKON SSS-LEX DAN VADER”

Oleh

Nama : Dava Virgio Kertawijaya

NIM : 00000056848

Program Studi : Sistem Informasi

Fakultas : Teknik dan Informatika

Telah disetujui untuk diajukan pada

Sidang Ujian Skripsi Universitas Multimedia Nusantara

Tangerang, 29 November 2023

Pembimbing

A black background with a black square

Description automatically generated with medium confidence

Dinar Ajeng Kristiyanti, S.Kom., M.Kom. 0330128801/079159

Dosen MRSI Kelas B

A black background with a black square

Description automatically generated with medium confidence

Ir. Reymond Sunardi Oetama, M.CIS.

0328046803

# **HALAMAN PENGESAHAN**

Skripsi dengan judul

“KOMPARASI DETEKSI KEKUATAN SENTIMEN PADA ULASAN PRODUK MCDONALD’S: PENDEKATAN CNN, LSTM, DAN BI-LSTM DENGAN LEKSIKON SSS-LEX DAN VADER”

Oleh

Nama : Dava Virgio Kertawijaya

NIM : 00000056848

Program Studi : Sistem Informasi

Fakultas : Teknik dan Informatika

Telah diujikan pada hari Rabu, 29 November 2023

Pukul 09.00 s.d 15.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut.

Ketua Sidang

<<Nama Lengkap & gelar Dosen>> <<NIDN/NIK>>

Penguji

Dr. Irmawati, S.Kom., M.M.S.I. 0805097703/081431

Pembimbing

Dinar Ajeng Kristiyanti, S.Kom., M.Kom. 0330128801/079159

Ketua Program Studi Sistem Informasi

Ririn Ikana Desanti, S.Kom., M.Kom.

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

Sebagai civitas academica Universitas Multimedia Nusantara, saya yang bertanda tangan di bawah ini:

Nama : Dava Virgio Kertawijaya

NIM : 00000056848

Program Studi : Sistem Informasi

Fakultas : Teknik dan Informatika

JenisKarya : \*Tesis/~~Skripsi~~/Tugas Akhir (\*coret salah satu)

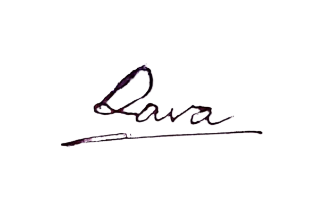
Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Multimedia Nusantara Hak Bebas Royalti Nonekslusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul.

“KOMPARASI DETEKSI KEKUATAN SENTIMEN PADA ULASAN PRODUK MCDONALD’S: PENDEKATAN CNN, LSTM, DAN BI-LSTM DENGAN LEKSIKON SSS-LEX DAN VADER”

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini, Universitas Multimedia Nusantara berhak menyimpan, mengalihmediakan/mengalihformatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 29 November 2023

Yang menyatakan,



Dava Virgio Kertawijaya

# **KATA PENGANTAR**

Segala puji dan syukur kami panjatkan ke hadirat Allah SWT atas segala rahmat, hidayah, serta karunia-Nya yang senantiasa melimpah pada kita semua. Shalawat serta salam semoga selalu tercurahkan kepada junjungan kita, Nabi Muhammad SAW, yang telah menjadi suri tauladan bagi umat manusia.

Pada kesempatan ini, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah turut serta dalam penelitian ini. Penulisan tugas akhir ini tidak mungkin terlaksana tanpa dukungan, bimbingan, dan motivasi dari berbagai pihak yang telah berperan penting.

Mengucapkan terima kasih

1. Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Universitas Multimedia Nusantara.
3. Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Universitas Multimedia Nusantara.
4. Dinar Ajeng Kristiyanti, S.Kom., M.Kom., sebagai Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.
6. Semua pihak yang telah membantu dan mendukung penelitian ini, baik secara langsung maupun tidak langsung.

Semoga karya ilmiah ini dapat memberikan sumbangan berharga dalam bidang Sentiment Analysis, khususnya dalam aspek deteksi kekuatan sentimen. Diharapkan karya ilmiah ini dapat memberikan manfaat yang signifikan bagi komunitas akademik serta praktisi di masa depan.

Tangerang, 29 November 2023

A close up of a logo

Description automatically generated

Dava Virgio Kertawijaya

KOMPARASI DETEKSI KEKUATAN SENTIMEN PADA ULASAN PRODUK MCDONALD’S: PENDEKATAN CNN, LSTM, DAN BI-LSTM DENGAN LEKSIKON SSS-LEX DAN VADER

Dava Virgio Kertawijaya

# **ABSTRAK**

Dalam era teknologi yang terus berkembang, pemahaman yang mendalam terhadap sentimen konsumen menjadi krusial, terutama dalam industri makanan cepat saji. Penelitian ini bertujuan untuk membandingkan pendekatan Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), dan Bidirectional LSTM (Bi-LSTM) dengan penerapan leksikon SSS-LEX dan Vader dalam deteksi kekuatan sentimen pada ulasan produk McDonald's. Menggunakan dataset ulasan produk McDonald's dari Kaggle periode Januari 2023 hingga Juni 2023, penelitian ini memusatkan perhatian pada analisis kekuatan sentimen untuk memahami respons konsumen secara lebih rinci daripada hanya melihat polaritas umum. Pendekatan berbasis leksikon, seperti SSS-LEX dan Vader, dikombinasikan dengan model CNN, LSTM, dan Bi-LSTM untuk mendapatkan pemahaman yang holistik mengenai sentimen konsumen. Melalui penerapan deteksi kekuatan sentimen pada domain ulasan produk McDonald's, penelitian ini bertujuan untuk memberikan wawasan mendalam dan kontekstual tentang cara konsumen merespons produk tersebut. Kontribusi utama penelitian ini mencakup penggabungan pendekatan leksikon dengan model neural network yang canggih, serta perbandingan kinerja model yang berbeda. Diharapkan hasil penelitian ini tidak hanya memberikan pemahaman lebih mendalam terkait sentimen konsumen pada produk McDonald's, tetapi juga dapat memberikan dasar bagi pengembangan metode analisis sentimen yang lebih canggih dalam konteks ulasan produk makanan cepat saji. Wawasan yang diperoleh dari penelitian ini diharapkan dapat mendukung pengambilan keputusan yang lebih baik dalam pengembangan produk dan strategi pemasaran di industri makanan cepat saji yang kompetitif.

Kata kunci: Deteksi kekuatan sentimen, Analisis sentimen, *CNN,* LSTM, Bi-LSTM, SSS-Lex, Vader.

KOMPARASI DETEKSI KEKUATAN SENTIMEN PADA ULASAN PRODUK MCDONALD’S: PENDEKATAN CNN, LSTM, DAN BI-LSTM DENGAN LEKSIKON SSS-LEX DAN VADER

Dava Virgio Kertawijaya

# ***ABSTRACT (English)***

*In an era of ever-evolving technology, a deep understanding of consumer sentiment is crucial, especially in the fast-food industry. This study aims to compare Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Bidirectional LSTM (Bi-LSTM) approaches with the application of SSS-LEX and Vader lexicons in sentiment strength detection on McDonald's product reviews. Using the McDonald's product review dataset from Kaggle from January 2023 to June 2023, this study focuses on sentiment strength analysis to understand consumer responses in more detail rather than just looking at general polarity. Lexicon-based approaches, such as SSS-LEX and Vader, are combined with CNN, LSTM, and Bi-LSTM models to gain a holistic understanding of consumer sentiment. Through the application of sentiment strength detection to the McDonald's product review domain, this research aims to provide deep and contextual insights into how consumers respond to the product. The main contributions of this research include combining the lexicon approach with advanced neural network models, as well as comparing the performance of different models. It is expected that the results of this study will not only provide a deeper understanding of consumer sentiment on McDonald's products but can also provide a basis for the development of more sophisticated sentiment analysis methods in the context of fast food product reviews. The insights gained from this research are expected to support better decision-making in product development and marketing strategies in the competitive fast-food industry.*

***Keywords****: Sentiment strength detection, Sentiment analysis, CNN,**LSTM,**Bi-LSTM,**SSS-Lex,**Vader.*

# **DAFTAR ISI**

[**HALAMAN PERNYATAAN TIDAK PLAGIAT** ii](#_Toc152100236)

[**HALAMAN PERSETUJUAN** iii](#_Toc152100237)

[**HALAMAN PENGESAHAN** iv](#_Toc152100238)

[**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS** v](#_Toc152100239)

[**KATA PENGANTAR** vi](#_Toc152100240)

[**ABSTRAK** viii](#_Toc152100241)

[***ABSTRACT (English)*** ix](#_Toc152100242)

[**DAFTAR ISI** x](#_Toc152100243)

[**DAFTAR TABEL** xiii](#_Toc152100244)

[**DAFTAR GAMBAR** xiv](#_Toc152100245)

[**DAFTAR LAMPIRAN** xv](#_Toc152100246)

[**BAB I PENDAHULUAN** 1](#_Toc152100247)

[**1.1** **Latar Belakang** 1](#_Toc152100248)

[**1.2** **Rumusan Masalah** 5](#_Toc152100249)

[**1.3** **Batasan Masalah** 5](#_Toc152100250)

[**1.4** **Tujuan dan Manfaat Penelitian** 6](#_Toc152100251)

[**1.4.1** **Tujuan Penelitian** 6](#_Toc152100252)

[**1.4.2** **Manfaat Penelitian** 6](#_Toc152100253)

[**1.4.2.1** **Manfaat Teoritis** 6](#_Toc152100254)

[**1.4.2.2** **Manfaat Praktis** 6](#_Toc152100255)

[**1.5** **Sistematika Penulisan** 7](#_Toc152100256)

[**BAB II LANDASAN TEORI** 8](#_Toc152100257)

[**2.1** **Tinjauan Teori** 8](#_Toc152100258)

[**2.1.1** **Analisis Sentimen** 8](#_Toc152100259)

[**2.1.2** **Deteksi Kekuatan Sentimen** 8](#_Toc152100260)

[**2.1.3** **Ulasan Produk McDonald’s** 8](#_Toc152100261)

[**2.1.4** **Data Pre-processing** 9](#_Toc152100262)

[**2.1.4.1** **Tokenization** 9](#_Toc152100268)

[**2.1.4.2** **Punctual Removal** 9](#_Toc152100275)

[**2.1.4.3** **Number Removal** 9](#_Toc152100283)

[**2.1.4.4** **Stemming** 10](#_Toc152100292)

[**2.1.4.5** **Lemmatization** 10](#_Toc152100302)

[**2.1.4.6** **Stop words removal** 10](#_Toc152100313)

[**2.1.4.7** **Spell correction** 10](#_Toc152100325)

[**2.1.5** **Metode-Metode Analisis Sentimen** 10](#_Toc152100326)

[**2.1.5.1** **Convolutional Neural Networks (CNN)** 10](#_Toc152100327)

[**2.1.5.2** **Long Short-Term Memory (LSTM)** 11](#_Toc152100328)

[**2.1.5.3** **Bidirectional Long Short-Term Memory (Bi-LSTM)** 11](#_Toc152100329)

[**2.1.5.4** **Kontekstual Leksikon SSS-Lex** 12](#_Toc152100330)

[**2.1.5.5** **Vader** 12](#_Toc152100331)

[**2.2** **Framework** 12](#_Toc152100332)

[**2.2.1** **CRISP-DM** 13](#_Toc152100333)

[**2.2.2** **Knowledge Discovery in Databases (KDD)** 13](#_Toc152100334)

[**2.2.3** **Sample, Explore, Modify, Model, Assess (SEMMA)** 13](#_Toc152100335)

[**2.3** **Tools** 14](#_Toc152100336)

[**2.3.1** **Deep Learning Frameworks** 14](#_Toc152100337)

[**2.3.1.1** **TensorFlow** 14](#_Toc152100338)

[**2.3.1.2** **Keras** 15](#_Toc152100339)

[**2.3.2** **Natural Language Processing (NLP) Libraries** 15](#_Toc152100340)

[**2.3.2.1** **Natural Language Toolkit (NLTK)** 15](#_Toc152100341)

[**2.3.2.2** **TextBlob** 16](#_Toc152100342)

[**2.3.2.3** **Gensim** 16](#_Toc152100343)

[**2.4** **Penelitian Terdahulu** 16](#_Toc152100344)

[**2.4.1** **Planning The Review** 17](#_Toc152100345)

[**2.4.1.1** **Menentukan Research Question** 17](#_Toc152100351)

[**2.4.1.2** **Menetapkan Istilah Pencarian** 18](#_Toc152100352)

[**2.4.1.3** **Identifikasi Kriteria Inklusi dan Eksklusi** 20](#_Toc152100353)

[**2.4.2** **Conducting The Review** 21](#_Toc152100354)

[**2.4.2.1** **Pemilihan Jurnal dari Kriteria Inklusi dan Eksklusi** 21](#_Toc152100360)

[**2.4.2.2** **Penilaian Kualitas Literatur** 21](#_Toc152100361)

[**2.4.2.3** **Penyintesisan Data** 21](#_Toc152100362)

[**2.4.3** **Reporting The Review** 22](#_Toc152100363)

[**BAB III METODOLOGI PENELITIAN** 30](#_Toc152100364)

[**3.1** **Gambaran Umum Objek Penelitian** 30](#_Toc152100368)

[**3.2** **Metode Penelitian** 30](#_Toc152100369)

[**3.3** **Variabel Penelitian** 31](#_Toc152100370)

[**3.3.1** **Variabel Independen** 32](#_Toc152100374)

[**3.3.2** **Variabel Dependen** 32](#_Toc152100375)

[**3.4** **Teknik Pengumpulan Data** 32](#_Toc152100376)

[**3.5** **Teknik Pengambilan Sampel** 32](#_Toc152100377)

[**3.6** **Teknik Analisis Data** 33](#_Toc152100378)

[**3.6.1** **Data Pre-Processing** 33](#_Toc152100379)

[**3.6.2** **Modeling CNN dengan Leksikon SSS-Lex dan Vader** 33](#_Toc152100380)

[**3.6.3** **Modeling LSTM dengan Leksikon SSS-Lex dan Vader** 33](#_Toc152100381)

[**3.6.4** **Modeling Bi-LSTM dengan Leksikon SSS-Lex dan Vader** 34](#_Toc152100382)

[**DAFTAR PUSTAKA** 35](#_Toc152100383)

[**LAMPIRAN** 43](#_Toc152100384)

# **DAFTAR TABEL**

[Tabel 3.1 Tabel Kriteria Inklusi & Eksklusi 20](#_Toc73803138)

# **DAFTAR GAMBAR**

[Gambar 1.1 Perusahaan F & B dengan Gerai Waralaba Terbanyak Global 2021 1](#_Toc152078897)

[Gambar 1.2 Grafik Sebaran Penelitian Berdasarkan Jenis Analisis Sentimen dan Kuartil 3](#_Toc152079203)

[Gambar 3.2 Grafik Distribusi Jumlah Publikasi Deteksi Kekuatan Sentimen dengan Leksikon Berbasis Konteks di Dunia berdasarkan Kuartil Scopus 20](#_Toc152080013)

[Gambar 3.3 Alur Penelitian 31](#_Toc152080341)

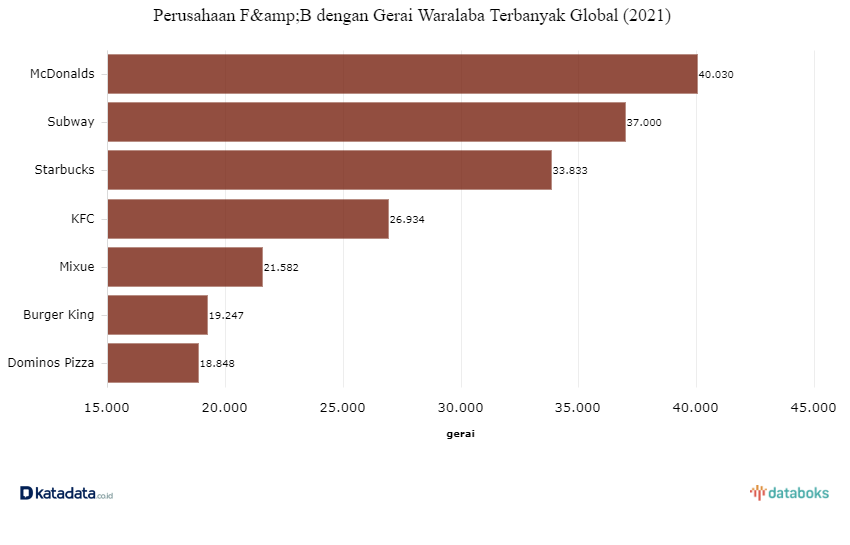
# **DAFTAR LAMPIRAN**

[Lampiran A Hasil Turnitin 43](#_Toc71293273)

# **BAB I PENDAHULUAN**

## **Latar Belakang**

Perkembangan bisnis di era teknologi digerakkan oleh perubahan preferensi konsumen. Saat ini, teknologi begitu membantu dalam mempercepat proses bisnis di semua industri. Kejadian ini menciptakan tren dimana orang mulai berharap dapat melakukan aktifitas mereka dengan cepat, mudah, dan praktis. Alhasil, model bisnis berkembang mengikuti arus kemauan konsumen di berbagai sektor industri, tarutama industri yang menjual kebutuhan primer, seperti makanan. Industri makanan telah berkembang sampai pada tahap dimana ada istilah makanan cepat saji. Istilah ini muncul dalam bisnis makanan sebagai respon atas kebutuhan konsumen yang menginginkan pelayanan cepat mudah, dan praktis. Salah satu merk brand terkenal yang menjadi pionir di sektor makanan cepat saji ini, yaitu McDonald’s dengan menjadi pemilik gerai terbanyak di dunia dengan grafik sebaran yang terlihat pada Gambar 1.1.



Gambar 1.1 Perusahaan F & B dengan Gerai Waralaba Terbanyak Global 2021[58]

Kunci keberhasilan bisnis ini bergantung pada produk dan layanan yang memenuhi kepuasan konsumen. Namun, keberhasilan McDonald's tidak hanya terletak pada efisiensi operasionalnya, melainkan juga pada kemampuannya memahami dan merespons sentimen konsumen.

Sentimen konsumen yang termanifestasi dalam ulasan produk dapat dilihat ketika pelanggan dimintai pendapatnya tentang produk atau pengalaman, mereka akan merasa dihargai dan terhubung dengan organisasi [1]. Dalam industri makanan, pelanggan sering kali melihat ulasan restoran sebelum melakukan pemesanan. Hal ini menjadikan sentimen konsumen yang termanifestasi dalam ulasan produk memiliki dampak signifikan pada citra merek dan keputusan pembelian. Analisis sentimen menjadi semakin penting dalam memahami kompleksitas respons konsumen terhadap produk tertentu. McDonald's, sebagai salah satu merek terkemuka di industri makanan cepat saji, terus menghadapi tantangan dalam mengelola dan memahami sentimen konsumen yang berkembang di platform daring.

Pendekatan konvensional dalam analisis sentimen sering kali terbatas pada penentuan polaritas umum, yaitu positif, negatif, atau netral dari suatu teks [2]. Namun, hanya sedikit penelitian yang mempertimbangkan kekuatan sentimen pengguna, khususnya dalam domain bisnis tentang ulasan pelanggan, padahal kekuatan sentimen pengguna merupakan sebuah aspek yang sangat penting untuk berbagai aplikasi [3]. Berikut ini grafik jenis penelitian sentimen analisis yang didapat melalui sistematic literature review.

A graph with numbers and a bar

Description automatically generated with medium confidence

Gambar 1.2 Grafik Sebaran Penelitian Berdasarkan Jenis Analisis Sentimen dan Kuartil

Berdasarkan grafik di atas, terlihat bahwa penelitian terkait sentimen analisis masih umum dilakukan. Sedangkan, penelitian terkait intensitas atau polaritas sentimen yang digunakan dalam deteksi kekuatan sentimen masih sedikit dilakukan. Dalam domain bisnis, kekuatan sentimen pengguna perlu untuk menganalisis kekuatan sentimen ulasan untuk menentukan peringkat produk dan pedagang. Ini karena ulasan yang berbeda yang berbeda untuk produk yang sama dapat memiliki kekuatan sentimen yang sangat berbeda, meskipun mereka mengekspresikan polaritas sentimen yang sama [4].

Penelitian telah cukup aktif di bidang analisis sentimen [5], dan sejumlah kecil pendekatan untuk memprediksi sentimen dari suatu teks telah diusulkan. Dua pendekatan utama untuk deteksi kekuatan sentimen adalah pendekatan berbasis leksikon dan pendekatan berbasis pembelajaran mesin [6]. Pendekatan berbasis leksikon menggunakan leksikon sentimen seperti Opinion Lexicon, SSS-Lex, Vader, dan SentiWordNet [7]. Terdapat penelitian yang mengusulkan untuk memprediksi sentimen komentar dengan menghitung jumlah kata sentimen positif dan negatif [8]. Ada juga penelitian yang mengusulkan pendekatan berbasis leksikon holistik untuk menganalisis sentimen dengan mengeksploitasi bukti eksternal dan konvensi linguistik dari ekspresi bahasa alami [9]. Di samping itu, ada yang mengintegrasikan sinyal emosional ke dalam metode berbasis leksikon tradisional untuk menganalisis sentimen ekspresi di media sosial [10]. Namun, mayoritas pendekatan berbasis leksikon mengabaikan informasi kontekstual, karena mereka mengikuti asumsi kantong kata, di mana sebuah dokumen direpresentasikan hanya sebagai sekumpulan kata [11]. Tidak seperti pendekatan berbasis leksikon yang memprediksi sentimen dengan dataset yang tidak berlabel, pendekatan berbasis pembelajaran mesin pertama-tama dilatih pada banyak dataset berlabel dan kemudian digunakan untuk memprediksi kekuatan sentimen dari sebuah dokumen atau kalimat [12]. Namun, pendekatan berbasis pembelajaran mesin belum mempertimbangkan informasi di balik kata-kata sentimen, yang sangat penting untuk mendeteksi kekuatan sentimen suatu dokumen.

Untuk mengintegrasikan keuntungan dari leksikon dan metode supervised learning, ada yang menggunakan petunjuk subjektivitas dan penandaan bagian dari ucapan untuk mengekstrak kata-kata dan konsep teratas, dan kemudian memasukkannya ke dalam jaringan saraf tiruan (CNN) untuk identifikasi subjektivitas [13]. Ada juga yang menggunakan sentimen positif dan negatif, serta kepercayaan konsumen terhadap produk dari pesan Twitter dalam pengklasifikasi pohon keputusan untuk menggambarkan pergerakan harga harian saham [14]. Di samping itu, ada juga yang mengintegrasikan leksikon ke dalam tugas klasifikasi sentimen [15]. Meskipun telah ada penelitian-penelitian sebelumnya yang membahas deteksi sentimen dan pengembangan leksikon, sebagian besar dari mereka cenderung fokus pada pendekatan tunggal dan jarang memperbandingkan kinerja beberapa metode. Oleh karena itu, penelitian ini bertujuan untuk melakukan komparasi detail antara CNN, LSTM, dan Bi-LSTM, dengan mempertimbangkan penggunaan leksikon SSS-LEX dan Vader, dalam deteksi kekuatan sentimen pada ulasan produk McDonald's.

Kontribusi pada penelitian ini adalah 1) mengetahui penerapan deteksi kekuatan sentimen pada ulasan produk McDonald’s, 2) data yang digunakan dalam penelitian adalah data ulasan produk McDonald’s bersumber dari Kaggle pada periode waktu January 2023 – Juni 2023, 3) menerapkan pendekatan CNN, LSTM, dan Bi-LSTM dengan leksikon SSS-Lex dan Vader pada deteksi kekuatan sentimen khususnya pada domain ulasan produk McDonald’s, 4) membandingkan model CNN-SSS-Lex, LSTM-SSS-Lex, Bi-LSTM-SSS-Lex, CNN-Vader, LSTM-Vader, dan Bi-LSTM-Vader untuk mencari pendekatan dan leksikon terbaik dalam deteksi kekuatan sentimen pada ulasan produk McDonald’s.

## **Rumusan Masalah**

Berikut merupakan rumusan masalah dari penelitian ini:

1. Bagaimana menerapkan algoritma berbasis deep learning untuk deteksi kekuatan sentimen pada ulasan produk McDonalds?
2. Bagaimana hasil perbandingan dari pendekatan CNN, LSTM, dan Bi-LSTM dengan leksikon SSS-Lex, dan Vader berdasarkan metrik evaluasi MAE?
3. Bagaimana hasil perbandingan dari pendekatan CNN, LSTM, dan Bi-LSTM dengan leksikon SSS-Lex, dan Vader berdasarkan metrik evaluasi prediktif R-square ()?

## **Batasan Masalah**

Berikut merupakan batasan masalah dari penelitian ini:

1. Objek analisis yang diteliti dibatasi pada ulasan produk McDonald's.
2. Data yang digunakan berupa ulasan pelanggan terkait produk McDonald’s pada periode Januari – Juni 2023 yang berasal dari data open source Kaggle.
3. Faktor-faktor eksternal, seperti tren sosial atau peristiwa tertentu yang dapat mempengaruhi sentimen masyarakat terhadap McDonald's, tidak dimasukkan dalam penelitian ini.
4. Pendekatan dan leksikon yang digunakan dibatasi dalam penerapan deteksi kekuatan sentimen pada ulasan produk.
5. Analisis yang dilakukan, yaitu membandingkan pendekatan CNN, LSTM, dan Bi-LSTM dengan leksikon SSS-Lex, dan Vader dalam menerapkan deteksi kekuatan sentimen.
6. Menerapkan framework CRISP-DM, namun tanpa melakukan tahap deployment.

## **Tujuan dan Manfaat Penelitian**

### **Tujuan Penelitian**

Berikut merupakan tujuan dari penelitian ini, yakni:

1. Menerapkan pendekatan berbasis deep learning, yaitu CNN, LSTM, dan Bi-LSTM untuk deteksi kekuatan sentimen pada ulasan produk McDonalds.
2. Mengetahui hasil perbandingan berdasarkan metrik evaluasi MAE dari pendekatan CNN, LSTM, dan Bi-LSTM dengan leksikon SSS-Lex, dan Vader.
3. Mengetahui hasil perbandingan berdasarkan metrik evaluasi prediktif R-square () dari pendekatan CNN, LSTM, dan Bi-LSTM dengan leksikon SSS-Lex, dan Vader.

### **Manfaat Penelitian**

Berikut merupakan manfaat yang dibagi menjadi manfaat praktis dan manfaat teoritis dari penelitian ini, diantaranya:

### **Manfaat Teoritis**

Berikut merupakan manfaat praktis dalam penelitian ini.

* + - 1. Penelitian ini memberikan kontribusi signifikan terhadap literatur analisis sentimen dengan membandingkan kinerja tiga pendekatan utama (CNN, LSTM, Bi-LSTM) dalam deteksi kekuatan sentimen pada ulasan produk.
      2. Melalui penggunaan leksikon SSS-LEX dan Vader dalam konteks analisis sentimen, penelitian ini dapat memperluas pemahaman kita tentang penggunaan leksikon kontekstual.

### **Manfaat Praktis**

Berikut merupakan manfaat praktis dalam penelitian ini.

* + - 1. Hasil penelitian dapat digunakan untuk mengembangkan alat analisis sentimen yang lebih efektif, khususnya dalam menilai kekuatan sentimen dalam ulasan produk McDonald's.
      2. Hasil penelitian ini dapat memberikan dasar untuk penelitian lebih lanjut di bidang analisis sentimen, baik dalam konteks produk McDonald's maupun dalam domain lain.

## **Sistematika Penulisan**

Skripsi ini menggunakan template yang terstandarisasi untuk mahasiswa Sistem Informasi di Universitas Multimedia Nusantara. Untuk mempermudah melihat dan mengetahui pembahasan secara menyeluruh, maka sistematika penulisan perlu dibuat sebagai kerangka sekaligus pedoman dalam penulisan. Berikut sistematika penulisan untuk karya ilmiah ini.

BAB I PENDAHULUAN

Pada pendahuluan bab 1 berisikan latar belakang penelitian, rumusan masalah yang mendasari penelitian ini, batasan masalah, tujuan dan manfaat penelitian, serta sistematika penulisan.

BAB II LANDASAN TEORI

Pada landasan teori bab 2 berisikan teori-teori terkait penelitian yang terdiri dari Text Mining, Sentiment Analysis, Context-Dependent Lexicon, dan Natural Langguage Processing (NLP). Selain teori-teori yang mendasari penelitian, bab 2 juga berisikan Study Literature Review (SLR), yaitu penelitian terdahulu berkaitan topik penelitian ini.

BAB III METODOLOGI PENELITIAN

Pada metodologi penelitian bab 3 berisikan teori-teori terkait penelitian yang terdiri dari gambaran umum objek penelitian, tinjauan literatur, metode penelitian, variabel penelitian, teknik pengumpulan data, teknik pengambilan sampel, dan teknik analisis data.

# **BAB II LANDASAN TEORI**

## **Tinjauan Teori**

### **Analisis Sentimen**

Analisis Sentimen adalah teknik yang sangat penting dalam pemrosesan bahasa alami (*Natural Language Processing*, NLP) [16]. Ini memberikan pemahaman mendalam tentang opini, sentimen, dan emosi yang terkandung dalam teks. Dalam era di mana kita dibanjiri oleh data teks dari berbagai sumber seperti media sosial, ulasan produk, atau berita, analisis sentimen menjadi kunci untuk memahami pandangan pengguna, mengevaluasi respons pasar, atau bahkan mendeteksi perubahan sikap masyarakat terhadap isu-isu tertentu. Dengan metode ini, kita dapat mengklasifikasikan apakah suatu teks mengandung sentimen positif, negatif, netral, atau bahkan dengan tingkat kekuatan tertentu, yang sangat penting dalam pengambilan keputusan [17].

### **Deteksi Kekuatan Sentimen**

Deteksi kekuatan sentimen merupakan perkembangan penting dalam analisis sentimen, karena tidak semua sentimen dalam teks memiliki kekuatan yang sama [18]. Dalam beberapa kasus, sentimen mungkin sangat kuat dan ekspresif, sementara dalam kasus lain, sentimen hanya sebatas sedikit ekspresi ringan. Dengan deteksi kekuatan sentimen, pemahaman tingkat intensitas sentimen yang diungkapkan dalam teks dapat dimengerti [19]. Misalnya, dalam analisis ulasan produk, deteksi kekuatan sentimen dapat membantu dalam memahami apakah pelanggan sangat puas atau hanya sedikit puas dengan produk tersebut. Hal ini memberikan wawasan yang lebih kaya dan nuansa dalam analisis sentimen.

### **Ulasan Produk McDonald’s**

McDonald's sebagai merek global dalam industri makanan cepat saji memiliki jutaan ulasan produk dari konsumen di berbagai platform. Analisis sentimen pada ulasan produk McDonald's memberikan wawasan berharga terkait persepsi konsumen, memungkinkan perusahaan untuk mengambil tindakan yang relevan [20].

### **Data Pre-processing**

*Pre-processing* dilakukan untuk membersihkan dan mengubah data untuk ekstraksi fitur yang relevan [21]. Hal ini melibatkan pembersihan dataset dari *noise*. *Noise* di sini pada dasarnya berkonotasi dengan ketidakteraturan bahasa yang sering muncul dalam teks *mikro-blog*. Karena data yang berisik dan tidak terstruktur mempengaruhi kualitas tugas klasifikasi sentimen, data tersebut diubah menjadi format input yang terstruktur [22].



### **Tokenization**

Tokenisasi adalah prosedur penting untuk memecah teks menjadi unit-unit bermakna yang disebut token [23]. Token bisa berupa kata atau frasa. Tujuannya adalah untuk membuat representasi yang lebih terstruktur dan mempermudah analisis teks lebih lanjut. Proses tokenisasi dapat melibatkan pemisahan berdasarkan spasi atau karakter tertentu.



### **Punctual Removal**

*Punctual Removal* merupakan proses menghapus tanda baca atau karakter yang dianggap tidak memberikan kontribusi signifikan pada analisis sentimen atau pemahaman teks. Misalnya, tanda tanya, tanda seru, tanda koma, tanda titik, dan sebagainya [24]. Penghapusan tanda baca membantu menghindari *noise* yang tidak perlu dalam data teks.



### **Number Removal**

*Number removal* adalah langkah untuk menghilangkan angka dari teks [25]. Angka seringkali dianggap tidak relevan dalam analisis sentimen dan dapat dihapus agar fokus tetap pada makna kata dan struktur kalimat.



### **Stemming**

*Stemming* adalah proses menghilangkan afiks (imbuhan) dari kata untuk mendapatkan akar kata atau bentuk dasarnya [26]. Hal ini dilakukan untuk mengurangi variasi kata ke bentuk dasarnya sehingga kata-kata yang sebenarnya memiliki makna yang sama dapat diidentifikasi dengan lebih baik [27].



### **Lemmatization**

*Lemmatization* adalah proses mengubah kata-kata ke bentuk dasarnya atau *lemma* [28]. Dibandingkan dengan *stemming*, lemmatization lebih canggih karena melibatkan pemahaman struktur gramatikal dan makna kata, sehingga menghasilkan representasi yang lebih akurat [29].



### **Stop words removal**

*Stop words removal* adalah langkah di mana kata-kata umum yang tidak memberikan makna signifikan, seperti "dan," "atau," "yang," dihapus dari teks [30]. Hal ini dilakukan agar fokus analisis tetap pada kata-kata yang lebih informatif.



### **Spell correction**

*Spell correction* adalah proses memperbaiki kesalahan pengejaan dalam teks. Tujuan dari langkah ini adalah untuk mengurangi ambiguitas dan memastikan kata-kata diekspresikan dengan benar, sehingga meningkatkan akurasi analisis teks [31]. *Spell correction* dapat melibatkan penggunaan kamus atau algoritma pemrosesan bahasa alami.

### **Metode-Metode Analisis Sentimen**

### **Convolutional Neural Networks (CNN)**

*Convolutional Neural Networks* (CNN) awalnya dikembangkan untuk mengolah data gambar, namun dengan sukses telah diadopsi dalam analisis teks [32]. CNN memanfaatkan lapisan konvolusi untuk mengidentifikasi pola dan fitur dalam data [33]. Dalam konteks analisis sentimen, CNN digunakan untuk mengekstraksi komponen penting dari teks, seperti frasa-frasa atau pola kata yang mencerminkan sentimen tertentu [34]. Keunggulan utama CNN adalah kemampuannya memahami struktur hierarkis dalam teks [35]. Dengan kata lain, CNN dapat mengenali bagaimana kata-kata, frasa, dan kalimat saling berhubungan dalam suatu dokumen, memungkinkan hasil analisis sentimen yang lebih canggih dan akurat [36]. Karena kemampuannya yang sangat adaptif, CNN membuka peluang untuk analisis yang lebih mendalam dan berlaku luas, tidak hanya dalam bahasa Inggris tetapi juga dalam berbagai bahasa lainnya [37]. Dengan demikian, penggunaan CNN dalam analisis sentimen telah membantu menghadirkan solusi yang lebih kuat dan efektif dalam memahami sentimen manusia yang terkandung dalam teks.

### **Long Short-Term Memory (LSTM)**

LSTM adalah jenis arsitektur jaringan saraf yang memegang peranan krusial dalam memodelkan urutan data, seperti teks [38]. Keunggulan LSTM terletak pada kemampuannya untuk mengatasi masalah vanishing gradient, memungkinkan perhitungan yang lebih efektif pada data sekuensial [39]. LSTM diperkenalkan oleh Hochreiter & Schmidhuber pada tahun 1997 sebagai solusi untuk memperbaiki masalah yang dihadapi oleh RNN.

### **Bidirectional Long Short-Term Memory (Bi-LSTM)**

Bidirectional Long Short-Term Memory (Bi-LSTM) merupakan pengembangan dari arsitektur LSTM yang memiliki kemampuan untuk menggabungkan informasi dari kedua arah, yaitu sebelum dan sesudah suatu titik waktu dalam suatu urutan data [40]. Dengan memperkenalkan arah maju (*forward*) dan arah mundur (*backward*), Bi-LSTM dapat meningkatkan pemahaman konteks dan memperluas jangkauan informasi yang dapat diakses oleh model [41].

### **Kontekstual Leksikon SSS-Lex**

Leksikon SSS-LEX (*Sentiment Strength-Specific Lexicon*) adalah sumber daya linguistik yang dikembangkan khusus untuk analisis sentimen [37]. Leksikon ini dirancang untuk memberikan informasi tentang kekuatan sentimen yang terkandung dalam kata-kata, memungkinkan analisis yang lebih nuansatif dan mendalam terhadap tingkat intensitas sentimen.

### **Vader**

VADER (*Valence Aware Dictionary and Sentiment Reasoner*) adalah leksikon sentimen yang memanfaatkan pendekatan valensi untuk menganalisis sentimen teks [42]. Dengan mempertimbangkan nilai sentimen positif, negatif, atau netral dari setiap kata, VADER menyediakan skala kontinum untuk mengukur sentimen dalam teks, mulai dari sangat negatif hingga sangat positif [43]. Leksikon ini mencakup analisis bigram dan trigram, memungkinkan penangkapan nuansa emosional dalam konteks kata-kata yang digabungkan. Selain itu, VADER menyesuaikan sentimen untuk kata-kata amplifier dan negasi, serta memberikan bobot sentimen relatif untuk mengukur pengaruh kata-kata terhadap sentimen keseluruhan teks [44]. Kemampuan VADER untuk menyediakan analisis sentimen yang lebih kaya dan kontekstual membuatnya menjadi alat yang efektif dalam berbagai aplikasi, termasuk analisis ulasan produk, pemantauan media sosial, dan evaluasi sentimen terhadap berita.

## **Framework**

Berikut merupakan beberapa framework data mining yang dapat diterapkan pada penelitian deteksi kekuatan sentimen, serta algoritma yang akan digunakan dalam penelitian ini.

### **CRISP-DM**

CRISP-DM (*Cross-Industry Standard Process for Data Mining*) adalah salah satu kerangka kerja yang paling umum digunakan dalam data mining dan analisis data [45]. Ini terdiri dari enam tahap, yaitu pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan distribusi. Tahap ini secara terstruktur memandu peneliti dalam mengelola seluruh proyek analisis data, termasuk analisis sentimen. Kerangka kerja ini memberikan struktur yang kuat untuk mengelola proyek analisis data, membantu dalam pemahaman bisnis yang kuat, dan memungkinkan fleksibilitas dalam menggunakan berbagai teknik pemodelan. Akan tetapi, mungkin terlalu kompleks untuk proyek kecil dan sederhana. Selain itu, CRISP-DM cenderung terfokus pada tugas-tugas teknis daripada interpretasi hasil.

### **Knowledge Discovery in Databases (KDD)**

*Knowledge Discovery in Databases* (KDD) adalah pendekatan sistematis untuk mengidentifikasi pola dan pengetahuan dari data yang kompleks dan besar [46]. Ini terdiri dari beberapa tahap, yaitu pemahaman masalah, pemahaman data, ekstraksi, transformasi, pemodelan, evaluasi, dan penyajian hasil. KDD menekankan pentingnya pemahaman masalah dan data yang mendalam sebelum analisis dimulai. Ini membantu mengidentifikasi sumber daya data yang paling relevan dan fokus pada penemuan pengetahuan yang bermanfaat. Namun, proses KDD cenderung memakan waktu dan bisa rumit. Memerlukan pemahaman yang mendalam tentang domain data yang sedang dianalisis.

### **Sample, Explore, Modify, Model, Assess (SEMMA)**

*Sample, Explore, Modify, Model, Assess* (SEMMA) adalah pendekatan data mining yang dikembangkan oleh *SAS Institute* [47]. Ini terdiri dari tahap-tahap berurutan: sampel data, eksplorasi data, modifikasi data, pemodelan, dan penilaian. SEMMA menekankan pentingnya langkah-langkah persiapan data yang kuat sebelum pemodelan. SEMMA memprioritaskan pemahaman yang mendalam tentang data, membantu mengidentifikasi perbaikan yang diperlukan dalam data sebelum pemodelan, dan fokus pada evaluasi hasil dengan cermat. Akan tetapi, mungkin memerlukan lebih banyak waktu untuk persiapan data dibandingkan dengan beberapa kerangka kerja lainnya. Selain itu, persiapan data yang ketat dapat menambah kompleksitas proyek.

## **Tools**

### **Deep Learning Frameworks**

*Deep Learning Frameworks* adalah kerangka kerja yang mendasari dalam pengembangan dan implementasi model jaringan saraf dalam penelitian dan aplikasi berbasis *deep learning* [48]. Mereka menyediakan alat dan sumber daya penting yang mempermudah pelatihan dan pengujian model deep learning. Dalam penelitian Anda, menggunakan kerangka kerja deep learning merupakan pilihan yang sangat relevan karena penggunaan jaringan saraf yang dalam untuk tugas analisis teks.

### **TensorFlow**

TensorFlow adalah salah satu kerangka kerja deep learning yang sangat populer. Dikembangkan oleh Google, TensorFlow menawarkan fleksibilitas dan dukungan komprehensif untuk pembuatan dan pelatihan model deep learning. Ia juga mendukung berbagai jenis arsitektur jaringan saraf dan dapat digunakan dalam berbagai tugas analisis teks seperti klasifikasi teks, pemrosesan bahasa alami, dan penerjemahan mesin [49]. TensorFlow memiliki komunitas yang besar dan aktif, serta sumber daya pelatihan yang melimpah, yang membuatnya sangat relevan dalam penelitian Anda.

### **Keras**

Keras, di sisi lain, adalah kerangka kerja yang sangat fokus pada kegunaan dan kesederhanaan. Keras adalah antarmuka tingkat tinggi yang berjalan di atas TensorFlow (dan juga di atas Theano) [50]. Ini memudahkan peneliti untuk membangun dan menguji model *deep learning* dengan cepat. Keras sering digunakan dalam penelitian dan pengembangan prototipe karena kecepatan dan kemudahan penggunaannya. Anda dapat menggunakan Keras untuk membangun dan menguji model deep learning dengan cepat, terutama jika Anda memiliki pengalaman yang lebih terbatas dalam *deep learning*.

### **Natural Language Processing (NLP) Libraries**

*Natural Language Processing (NLP) Libraries* adalah komponen penting dalam penelitian dan pengembangan dalam bidang analisis teks dan pemrosesan bahasa alami [51]. Mereka menyediakan alat dan fungsi penting untuk memproses, menganalisis, dan memahami teks manusia. Tiga perpustakaan NLP yang relevan yang bisa menjadi pilihan Anda dalam penelitian adalah NLTK (*Natural Language Toolkit*), TextBlob, dan Gensim.

### **Natural Language Toolkit (NLTK)**

NLTK (*Natural Language Toolkit*) adalah salah satu perpustakaan NLP yang paling banyak digunakan dalam bahasa pemrograman Python. NLTK menawarkan berbagai modul dan sumber daya, seperti tokenisasi, stemming, analisis sentimen, analisis tata bahasa, dan banyak lagi [52]. Keunggulan NLTK adalah fleksibilitasnya, yang memungkinkan peneliti untuk mengakses alat NLP yang beragam sesuai dengan kebutuhan penelitian. NLTK juga sering digunakan oleh para peneliti untuk keperluan pendidikan dan eksperimen dalam NLP.

### **TextBlob**

TextBlob adalah perpustakaan NLP Python yang menekankan kesederhanaan dan kemudahan penggunaan. TextBlob menyediakan antarmuka yang intuitif dan alat yang mudah digunakan, memungkinkan peneliti dengan berbagai tingkat keahlian untuk melakukan analisis teks [53].

### **Gensim**

Gensim adalah perpustakaan yang fokus pada pemodelan topik dan pemahaman kata dalam teks. Gensim sering digunakan dalam penelitian yang melibatkan teks besar atau koleksi dokumen. Gensim memungkinkan peneliti untuk membangun model pemahaman kata (*word embedding*) yang dapat digunakan dalam berbagai tugas pemrosesan bahasa alami [54]. Misalnya, model pemahaman kata yang dihasilkan oleh Gensim dapat membantu dalam klasifikasi dokumen, pemodelan topik, atau pencarian dokumen berdasarkan konten.

## **Penelitian Terdahulu**

Sebuah *Systematic Literature Review* (SLR) dilaksanakan untuk mengeksplorasi dan menyajikan wawasan dari berbagai artikel penelitian yang fokus pada deteksi kekuatan sentimen melalui pendekatan leksikon berbasis konteks. Artikel-artikel yang dikaji mencakup periode tahun 2020 hingga 2023, sesuai dengan rentang waktu ketika metode leksikon berbasis konteks mulai mendapatkan popularitas dalam analisis sentimen. Tujuan dari SLR ini adalah untuk mendapatkan pemahaman yang mendalam tentang variasi literatur yang ada yang membahas aspek deteksi kekuatan sentimen. Lebih lanjut, SLR ini juga bertujuan untuk menyelidiki berbagai teknologi yang diterapkan dalam domain analisis sentimen yang memanfaatkan leksikon berbasis konteks. Penelitian ini mengadopsi langkah-langkah yang telah dijelaskan dalam panduan *Systematic Literature Review* yang diajukan oleh Kitchenham dan Charters[55]. Proses SLR ini terstruktur dan diilustrasikan dengan rinci dalam Gambar 2.1, dengan rincian model diperjelas pada subseksi berikutnya.

A diagram of a review

Description automatically generated

Gambar 2.1 Tahapan Systematic Literature Review [56]

## **Planning The Review**



## **Menentukan Research Question**

Langkah awal dalam mengidentifikasi permasalahan penelitian adalah merumuskan pertanyaan penelitian. Proses Tinjauan Pustaka yang sistematis sangat penting dalam mengidentifikasi permasalahan penelitian[57]. Tindakan ini dilakukan untuk menetapkan batasan dalam konteks diskusi, sehingga memandu penelitian agar lebih terfokus. Dalam upaya untuk mempermudah dan menyelaraskan penelitian, *Research Question* (RQ) dirinci sebagai berikut.:

a. RQ 1: Pada tahun berapakah penelitian yang terkait dengan pendeteksian kekuatan sentimen menggunakan leksikon berbasis konteks dijalankan?

b. RQ 2: Apa saja jurnal yang pernah menerbitkan artikel mengenai pendeteksian kekuatan sentimen dengan pendekatan leksikon berbasis konteks?

c. RQ 3: Apa saja kata kunci yang digunakan dalam pendeteksian kekuatan sentimen dengan leksikon berbasis konteks?

d. RQ 4: Metode analisis sentimen apa yang diterapkan dalam penelitian deteksi kekuatan sentimen dengan leksikon berbasis konteks?

e. RQ 5: Algoritma atau teknik optimasi apa yang digunakan dalam penelitian pendeteksian kekuatan sentimen dengan leksikon berbasis konteks?

f. RQ 6: Metode dan teknologi apa yang digunakan dalam penelitian pendeteksian kekuatan sentimen dengan leksikon berbasis konteks?

g. RQ 7: Platform data apa yang digunakan dalam penelitian pendeteksian kekuatan sentimen dengan leksikon berbasis konteks?

Dari rangkuman research question yang tercantum di atas, langkah berikutnya adalah menetapkan kata kunci pencarian untuk fase selanjutnya.

## **Menetapkan Istilah Pencarian**

Peran penting kata kunci dalam penelitian tidak bisa diabaikan. Pemilihan kata kunci yang tepat memiliki dampak signifikan dalam mendapatkan artikel yang relevan dengan topik penelitian. Untuk memastikan hasil pencarian yang menyeluruh, penelitian ini mengidentifikasi istilah-istilah dari pertanyaan penelitian, menggunakan istilah dasar dan afiks yang sesuai dengan fokus penelitian, dan menerapkan operator *AND* dan *OR* sesuai kebutuhan. Penelitian ini menelaah artikel-artikel yang terindeks di Scopus, dan semua dokumen tersebut disusun dalam bahasa Inggris. Oleh karena itu, perumusan kata kunci juga dilakukan dalam bahasa Inggris. Berdasarkan pengembangan kata kunci ini, sebuah rangkaian string pencarian berhasil disusun.

*(("sentiment") AND ("strength" OR "power" OR "intensity" OR "detection") AND ("context" OR "lexicon" OR "sentistrength" OR "vader" OR "sss-lex" OR "affin" OR "wordnet") AND ("machine learning" OR "deep learning" OR "text mining" OR “data mining” OR "natural language processing" OR "NLP" OR "feature extraction" OR "feature selection" OR "dimensionality reduction" OR "cross-validation" OR "hyperparameter tuning" OR "ensemble methods" OR "transfer learning" OR "sequence-to-sequence" OR "attention mechanisms" OR "word embeddings" OR "sentiment lexicons" OR "feature engineering" OR "time series analysis" OR "Regression Analysis"))*

String pencarian di atas berfungsi sebagai kata kunci untuk mencari artikel yang akan digunakan dalam penelitian ini. Proses pencarian artikel dilakukan melalui aplikasi Harzing's Publish or Perish, dengan menyaring hanya artikel yang terdaftar di Scopus. Dengan menggunakan beberapa kombinasi string dalam pencarian, ditemukan total 498 artikel yang paling relevan dengan topik penelitian, dengan rentang Quartile 1 hingga Quartile 4. Dari jumlah tersebut, 89 artikel terkategori sebagai Kuartil 1, 60 artikel di Kuartil 2, 36 artikel di Kuartil 3, dan 64 artikel di Kuartil 4. Distribusi jumlah artikel berdasarkan Quartil Ranking Scopus tergambar dalam Grafik 3.2.

A graph with blue squares

Description automatically generated

Gambar 2.2 Diagram Distribusi Jumlah Publikasi untuk Deteksi Sentimen Intensitas dengan Pendekatan Leksikon Berbasis Konteks di Seluruh Dunia Berdasarkan Peringkat Kuartil Scopus.

## **Identifikasi Kriteria Inklusi dan Eksklusi**

Dalam memperoleh literatur yang berkaitan dengan topik penelitian, sangat penting untuk menetapkan kriteria khusus, yang terdiri dari kriteria inklusi dan eksklusi. Kriteria inklusi berfungsi sebagai parameter yang memandu pemilihan literatur, menggambarkan karakteristik yang penting untuk dimasukkan ke dalam penelitian. Sebaliknya, kriteria eksklusi adalah faktor-faktor yang tidak dipertimbangkan dalam proses pemilihan literatur. Kriteria ini secara kolektif berkontribusi pada pendekatan yang disempurnakan dan metodis, memastikan pengambilan artikel yang paling relevan dan relevan secara kontekstual untuk upaya penelitian. Berikut ini Tabel 3.1.

Tabel 3.1 Identifikasi Kriteria Inklusi dan Eksklusi

|  |  |  |
| --- | --- | --- |
| No. | Kriteria Inklusi | Kriteria Eksklusi |
| 1 | Jurnal penelitian yang terbit dari 2020 hingga 2023 | Jurnal penelitian yang diteritkan sebelum 2020 |
| 2 | Jurnal penelitian yang ditulis dalam bahasa Inggris | Jurnal penelitian tidakdalam bahasa inggris |
| 3 | Jurnal penelitian yang terindeks di Scopus. | Jurnal penelitian tidak terindex di scopus |
| 4 | Jurnal penelitian yang membahas deteksi kekuatan sentimen dengan menggunakan pendekatan leksikon berbasis konteks. | Jurnal penelitian yang membahas deteksi kekuatan sentimen dengan menggunakan pendekatan kecuali leksikon berbasis konteks |
| 5 | Jurnal penelitian yang menyediakan informasi lengkap. | Jurnal penelitian yang memiliki duplikat |

## **Conducting The Review**



## **Pemilihan Jurnal dari Kriteria Inklusi dan Eksklusi**

Pada tahap perencanaan tinjauan, identifikasi kriteria inklusi dan eksklusi dilakukan, menghasilkan 5 kriteria inklusi dan 5 kriteria eksklusi. Selanjutnya, kriteria inklusi dan eksklusi yang telah diidentifikasi digunakan untuk menyaring artikel. Berdasarkan hasil seleksi menggunakan kriteria inklusi dan eksklusi dari 493 artikel awal, diperoleh 165 artikel yang paling relevan.

## **Penilaian Kualitas Literatur**

Untuk memastikan kualitas dalam proses sistematik literature review dan mencapai tinjauan yang kuat, penilaian kualitas artikel penelitian dilakukan berdasarkan kriteria tertentu. Parameter-parameter yang digunakan dalam penelitian ini melibatkan:

a. Artikel yang memiliki reputasi dan terindeks di Scopus.

b. Hanya artikel jurnal yang dimasukkan dalam penelitian.

c. Setiap langkah dalam sistematik literature review dilakukan sesuai dengan pedoman yang telah ditetapkan.

## **Penyintesisan Data**

Sintesis Proses penyintesisan data dilakukan untuk mengumpulkan bukti dari literatur yang telah dipilih. Selama proses ini, makalah dianalisis, disintesis, dibandingkan, dan dirangkum untuk menghasilkan informasi yang paling relevan dalam menjawab pertanyaan penelitian. Hasil dari penyintesisan data disusun dalam format tabel. Selain itu, data divisualisasikan melalui diagram garis, diagram lingkaran, dan diagram batang untuk memudahkan presentasi informasi mengenai distribusi artikel yang terkait dengan deteksi kekuatan sentimen menggunakan pendekatan leksikon berbasis konteks.

## **Reporting The Review**

Penutup dari tahap tinjauan pustaka sistematis ini adalah penyajian hasil tinjauan. Dalam konteks ini, disajikan penelitian yang paling relevan terkait dengan deteksi kekuatan sentimen menggunakan leksikon berbasis konteks.

Tabel 2.4 Penelitian Terdahulu

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Authors | Title | Journal | Methods | Dataset | Result |
| Minghui Huang, Haoran Xie, Yanghui Rao, Jingrong Feng, Fu Lee Wang | Sentiment strength detection with a context-dependent lexicon-based convolutional neural network | Information Science | Menggunakan CNN berdasarkan leksikon yang bergantung pada konteks | BBC, Digg, MySpace, Runners World, Twitter, Youtube | Model yang diusulkan dapat memprediksi kekuatan sentimen dokumen secara lebih efektif dibandingkan dengan baseline. Selain itu, leksikon khusus kekuatan sentimen memiliki kualitas yang tinggi. |
| Liang-Chih Yu, Jin Wang, K. Robert Lai, Xuejie Zhang | Refining Word Embeddings Using Intensity Scores for Sentiment Analysis | IEEE/ACM Transactions on Audio Speech and Language Processing | Menggunakan leksikon intensitas sentimen; menggunakan CNN untuk SemEval dan untuk SST menggunakan CNN, DAN, Bi-LSTM, Tree-LSTM. Metrik evaluasi SemEval, yaitu metrik evaluasi resmi, Macro − F1=(F1pos + F1neg )/2. Sementara, SST menggunakan akurasi | SemEval dan SST | Metode yang diusulkan dapat meningkatkan embedding kata konvensional maupun embedding sentimen untuk klasifikasi sentimen biner, ternary, dan berbutir halus (fine grained). Selain itu, kinerja berbagai model jaringan syaraf tiruan juga ditingkatkan. |
| Zhao Fang, Qiang Zhang, Xiaoan Tang, Anning Wang, Claude Baron | An implicit opinion analysis model based on feature-based implicit opinion patterns | Artificial Intelligence Review | FBIOP mining, sentiment intensity dan polarity calculation, dilanjutkan dengan FBIOP core selection, kemudian dikomparasi dengan model SVM dan CNN | Data dari PCauto.com.cn; IMOPCs dan IMOPRs dataset yang merupakan review (R) dan clausa (C) tanpa opini pada leksikon sentimen | Model yang diusulkan mengungguli 2 model lainnya (SVM dan CNN) |
| Xiaoping Yang, Zhongxia Zhang, Zhongqiu Zhang, Yuting Mo, Lianbei Li, Li Yu, Peican Zhu | Automatic Construction and Global Optimization of a Multisentiment Lexicon | Computational Intelligence and Neuroscience | Generate leksikon SentiRuc kemudian dibandingkan dengan leksikon lain seperti NTUSD, HowNet, dan DUT Ontology. Dilanjutkan dengan eksperimen sentiment disambiguation dan evaluasi sinonim dan antonim keempat leksikon. Terakhir, dilakukan evaluasi model leksikon NTUSD, HowNet, DUT, dan SentiRuc (before dan after disambiguation) | Data dari HIT Tongyicicilin, sinonim dan antonim leksikon dor pupils dan dataset NLPCC 2013 competition, dan NLPCC 2014 Competition | SentiRuc perform dengan bagus pada dataset general, namun pada tes labeling SentiRuc outperform peringkat kedua sebesar 21% yang membuktikan bahwa pemodelan statistical languange berkinerja luar biasa dalam representasi semantik dari sentimen |
| Hardik Meisheri, Rupsa Saha, Priyanka Sinha, Lipika Dey | Textmining at EmoInt-2017: A deep learning approach to sentiment intensity scoring of English tweets | EMNLP 2017 - 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, WASSA 2017 - Proceedings of the Workshop | CNN, LSTM, dan Custom model, yaitu model penelitian ini yang mengintegrasi CNN dan LSTM; Penambahan feature untuk dibandingkan | Amazon reviews dataset | Hasil penelitian menunjukkan bahwa fitur membantu seperti yang ditunjukkan oleh korelasi yang lebih tinggi. Selain itu, model campuran berkinerja lebih baik pada set pengembangan sementara pada set uji coba model LSTM terbukti lebih akurat. |
| Ali Mert Ertugrul, Itir Onal, Cengiz Acarturk | Does the strength of sentiment matter? A regression based approach on turkish social media | Lecture Notes in Computer Science | SVM dan SVR | Turkish tweets | Hasil menunjukkan bahwa menggunakan regresi pada confidence skor sedikit meningkatkan akurasi, dibandingkan dengan mengklasifikasikan tweet berdasarkan label kelas diskritnya. Selan itu, ditemukan juga bahwa menggunakan penyematan kata alih-alih Fitur N-gram menghasilkan representasi tweet yang berdimensi rendah dan lebih baik untuk analisis sentimen |
| Maganti Syamala, N. J. Nalini | A speech-based sentiment analysis using combined deep learning and language model on real-time product review | International Journal of Engineering Trends and Technology | CNN + Bi-RNN model, spectogram acoustic feature, 2 gram language model. | Youtube review dataset untuk implementasi proposed model, Librispeech (english accent speech dataset) untuk training modelnya. | Hasil menunjukkan bahwa diperoleh bukti peningkatan WER dan CER (sebagai metrik evaluasi) yang lebih baik jika dibandingkan dengan model ASR tradisional. Analisis sentimen dilakukan dengan menggunakan teknik NLP pada naskah teks yang telah ditranskrip dan mencapai akurasi 90% ketika  akurasi 90% ketika divalidasi menggunakan algoritma Regresi Logistik machine learning. |
| Jianeng Yang, Aimin Yang, Yongmei Zhou | A method of building Chinese sentiment lexicon based on semantics | Applied Mechanics and Materials 596 | Penelitian ini menggunakan SVM text classifier dan leksikon proposed CSLS yang dibandingkan dengan NTUSD dan TUPNL | Chinese microblog sentiment analysis evaluation of NLP&CC2012 | Eksperimen pada microblog sentimen analisis memverifikasi efektifitas dari lexicon yang dibangun. |
| Petr Hajek, Aliaksandr Barushka, Michal Munk | Opinion mining of consumer reviews using deep neural networks with word-sentiment associations | IFIP Advances in Information and Communication Technology 583 | Menggunakan DNN Model dengan asosiasi word sentiment (DNN-WSA) dan dibandingkan dengan model yang lain, seperti CNN, LSTM, dll. | Amazon dataset dari Kaggle | Hasil dari eksperimen menunjukkan bahwa asosiasi kata-sentimen mungkin lebih efektif daripada kata representasi berdasarkan sematan kata saja. Mengintegrasikan asosiasi kata-sentimen dengan representasi n-gram akan memberikan peningkatan lebih lanjut. |
| Aobo Yue, Chao Mao, Linyan Chen, Zebang Liu, Chaojun Zhang, Zhiqiang Li | Detecting Changes in Perceptions towards Smart City on Chinese Social Media: A Text Mining and Sentiment Analysis | Buildings | Model dibuat berdasarkan CNN dan LSTM, dan menggunakan LDA topic model dan CNN-BiLSTM model untuk construct topic mining model dari smart city | Weibo tentang smart city related topics | Hasil penelitian menunjukkan bahwa topik yang diminati publik sangat beragam, termasuk desain tingkat tinggi, pengembangan ekonomi, dan aplikasi teknologi. |

# **BAB III METODOLOGI PENELITIAN**

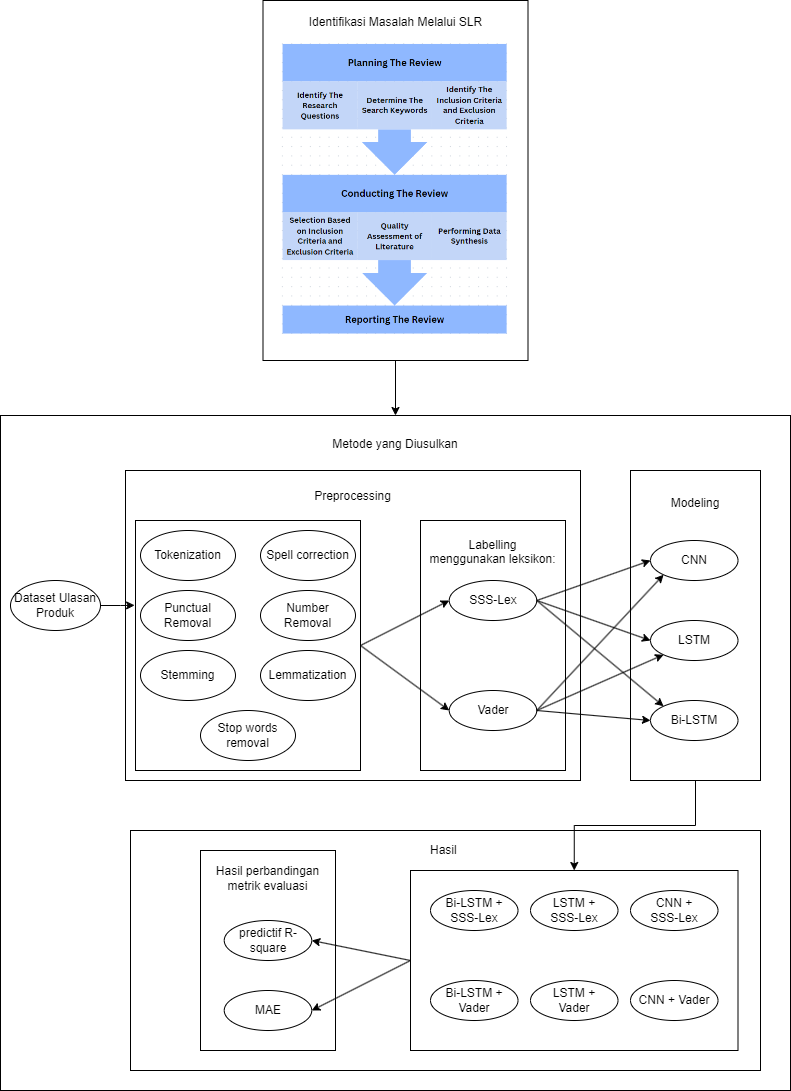


## **Gambaran Umum Objek Penelitian**

Objek penelitian yang digunakan dalam penelitian ini, yaitu data open source ulasan pelanggan McDonald’s dari Kaggle pada periode Januari – Juni 2023. McDonald’s merupakan waralaba makanan cepat saji global yang telah membangun reputasi sebagai salah satu merek paling ikonik di dunia kuliner. Dikenal dengan menu beragamnya, termasuk burger, kentang goreng, dan minuman ringan, restoran ini telah menjadi tempat favorit banyak orang dari berbagai kalangan. Ulasan pelanggan terhadap McDonald's sering kali mencerminkan berbagai aspek, mulai dari rasa makanan hingga pengalaman layanan di restoran tersebut.

## **Metode Penelitian**

Penelitian ini menerapkan deteksi kekuatan sentimen yang dapat mengukur skor intensitas sentimen dalam kalimat ulasan produk McDonald’s. Penelitian ini menggunakan framework CRISP-DM. Tak hanya itu, analisis yang diterapkan dalam penelitian ini, yaitu membandingkan hasil dari pendekatan CNN, LSTM, dan Bi-LSTM dengan leksikon SSS-Lex, dan Vader dalam menerapkan deteksi kekuatan sentimen berdasarkan metriks evaluasi MAE dan . Data preprocessing yang dilakukan dalam penelitian ini, yaitu menggunakan *tokenization, punctuation removal, number removal, stemming, lemmatization, stop words removal, dan spell correction*. Kemudian, dilanjutkan dengan pemodelan dengan pendekatan CNN, LSTM, dan Bi-LSTM dengan leksikon SSS-Lex dan Vader. Grafik alur penelitian dapat dilihat pada Gambar 3.3.



Gambar 3.3 Alur Penelitian

## **Variabel Penelitian**

Variabel penelitian yang terdapat pada penelitian ini, yaitu variabel independen dan variabel dependen.



## **Variabel Independen**

Data ulasan produk McDonald’s merupakan variabel independen dalam penelitian ini, sebab ulasan pelanggan biasanya memuat sentimen. Sehingga, dari ulasan pelanggan, dapat diketahui sentimen apa yang dirasakan pada produk/layanan McDonalds

## **Variabel Dependen**

Sentimen dari ulasan pelanggan pada produk McDonald’s merupakan variabel dependen dalam penelitian ini. Polaritas dari deteksi kekuatan sentimen terdapat sentimen positif, negatif, dan netral dengan skor sentimen yang ditentukan melalui leksikon. Sehingga, sentimen yang didapatkan dari deteksi kekuatan sentimen dapat dimanfaatkan dengan baik bagi McDonald’s dalam mengambil keputusan.

## **Teknik Pengumpulan Data**

Dataset utama yang digunakan berasal dari Kaggle, berfokus pada ulasan produk McDonald's dari Januari – Juni 2023 untuk menganalisis sentimen konsumen. Proses dimulai dengan mengidentifikasi dan mengunduh dataset yang sesuai dari Kaggle, mencakup teks ulasan, label sentimen, dan atribut pendukung lainnya. Pemilihan dataset didasarkan pada kelengkapan informasi yang memungkinkan analisis mendalam terhadap kekuatan sentimen dalam ulasan. Dataset ini telah mematuhi ketentuan lisensi dan kebijakan privasi Kaggle, serta memperhatikan etika penelitian dan hak privasi pengguna. Informasi lebih lanjut tentang dataset, termasuk lisensi dan batasan penggunaan, dapat ditemukan pada dokumentasi dataset yang terkait.

## **Teknik Pengambilan Sampel**

Data ulasan produk McDonald's yang digunakan berasal dari Kaggle, meliputi periode Januari – Juni 2023. Pengambilan sampel dilakukan dengan menggunakan teknik pengambilan sampel acak (*random sampling*) untuk mengurangi bias dan meningkatkan generalisabilitas hasil analisis sentimen. Teknik ini diarahkan untuk memastikan dataset mencakup variasi pendapat konsumen selama Januari-Juni 2023, serta memberikan gambaran seimbang dan representatif terhadap sentimen konsumen terhadap produk McDonald's dalam periode yang ditentukan.

## **Teknik Analisis Data**

Teknik analisis data dalam karya ilmiah ini menggunakan teknik kuantitatif. Teknik kuantitatif dilakukan pada deteksi kekuatan sentimen dengan pengujian beberapa model yang dilihat dari perbandingan berdasarkan statistik.

## **Data Pre-Processing**

Pada tahap ini, terdapat 7 langkah pre-processing yang dilakukan, yaitu *tokenization, punctuation removal, number removal, stemming, lemmatization, stop words removal, dan spell correction*. Semua tahapan *pre-processing* tersebut dilakukan dengan menggunakan *library* dari Python, seperti NLTK, *string*, dan menggunakan *function*.

## **Modeling CNN dengan Leksikon SSS-Lex dan Vader**

Convolutional Neural Network (CNN) diterapkan sebagai model untuk mengekstraksi fitur dari teks ulasan. CNN dikenal efektif dalam menangkap pola spasial dan hierarki fitur dalam data teks. Pemodelan CNN yang dilakukan berjumlah 2, yaitu dengan leksikon SSS-LEX dan Vader. SSS-Lex memberikan pendekatan leksikal yang mempertimbangkan kekuatan sentimen khusus, sementara Vader menggunakan pendekatan yang memperhitungkan konteks sosial dan tata bahasa. Integrasi kedua leksikon ini diharapkan dapat memberikan nuansa dan kedalaman analisis sentimen pada ulasan produk McDonald's, sehingga menciptakan landasan yang kuat untuk perbandingan dan evaluasi kinerja model.

## **Modeling LSTM dengan Leksikon SSS-Lex dan Vader**

*Long Short-Term Memory* (LSTM) diimplementasikan sebagai algoritma modeling kedua. LSTM, sebagai jenis dari *Recurrent Neural Network* (RNN), dirancang untuk mengatasi tantangan *vanishing gradient* yang sering muncul dalam RNN, sehingga dapat menangkap dependensi jangka panjang dalam urutan teks. Pemodelan LSTM yang dilakukan berjumlah 2, yaitu dengan leksikon SSS-LEX dan Vader. SSS-Lex memberikan pendekatan leksikal yang mempertimbangkan kekuatan sentimen khusus, sementara Vader menggunakan pendekatan yang memperhitungkan konteks sosial dan tata bahasa. Integrasi kedua leksikon ini diharapkan dapat memberikan nuansa dan kedalaman analisis sentimen pada ulasan produk McDonald's, sehingga menciptakan landasan yang kuat untuk perbandingan dan evaluasi kinerja model.

## **Modeling Bi-LSTM dengan Leksikon SSS-Lex dan Vader**

*Bidirectional* LSTM (Bi-LSTM) digunakan untuk memperluas kemampuan pemahaman konteks algoritma LSTM. Dengan mampu menangkap informasi dari kedua arah urutan, Bi-LSTM dapat lebih efektif dalam menangani konteks dan hubungan kompleks dalam teks ulasan. Pemodelan Bi-LSTM yang dilakukan berjumlah 2, yaitu dengan leksikon SSS-LEX dan Vader. SSS-Lex memberikan pendekatan leksikal yang mempertimbangkan kekuatan sentimen khusus, sementara Vader menggunakan pendekatan yang memperhitungkan konteks sosial dan tata bahasa. Integrasi kedua leksikon ini diharapkan dapat memberikan nuansa dan kedalaman analisis sentimen pada ulasan produk McDonald's, sehingga menciptakan landasan yang kuat untuk perbandingan dan evaluasi kinerja model.

# **DAFTAR PUSTAKA**

[1] D. Suhartanto, M. H. Ali, K. Tan, F. Sjahroeddin, and L. Kusdibyo, “Loyalty toward online food delivery service: the role of e-service quality and food quality,” *J. Foodserv. Bus. Res.*, vol. 22, pp. 1–17, Nov. 2018, doi: 10.1080/15378020.2018.1546076.

[2] B. Bansal and S. Srivastava, “Hybrid attribute based sentiment classification of online reviews for consumer intelligence,” *Appl. Intell.*, vol. 49, no. 1, pp. 137–149, Jan. 2019, doi: 10.1007/s10489-018-1299-7.

[3] H. Chen, H. Xie, X. Li, F. L. Wang, Y. Rao, and T. L. Wong, *Sentiment strength prediction using auxiliary features*. 2017. doi: 10.1145/3041021.3054149.

[4] M. McGlohon, N. Glance, and Z. Reiter, *Star quality: Aggregating reviews to rank products and merchants*. 2010. doi: 10.1609/icwsm.v4i1.14019.

[5] B. Liu, “Sentiment analysis and opinion mining,” in *Synthesis Lectures on Human Language Technologies*, 2012, pp. 1–184. doi: 10.2200/S00416ED1V01Y201204HLT016.

[6] I. Habernal, T. Ptáček, and J. Steinberger, “Reprint of ‘supervised sentiment analysis in Czech social media,’” *Inf. Process. Manag.*, vol. 51, no. 4, pp. 532–546, Jul. 2015, doi: 10.1016/j.ipm.2015.05.006.

[7] Y. Lu, X. Kong, X. Quan, W. Liu, and Y. Xu, *Exploring the sentiment strength of user reviews*, vol. 6184 LNCS. 2010. doi: 10.1007/978-3-642-14246-8\_46.

[8] M. Hu and B. Liu, “Mining and summarizing customer reviews,” *KDD-2004 - Proc. Tenth ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 168–177, 2004, doi: 10.1145/1014052.1014073.

[9] X. Ding, B. Liu, and P. S. Yu, *A holistic lexicon-based approach to opinion mining*. 2008. doi: 10.1145/1341531.1341561.

[10] X. Hu, J. Tang, H. Gao, and H. Liu, *Unsupervised sentiment analysis with emotional signals*. 2013. doi: 10.1145/2488388.2488442.

[11] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, “Distributed Representations of Words and Phrases and their Compositionality. arXiv e-prints, page,” *arXiv Prepr. arXiv1310.4546*, vol. 26, pp. 1–9, Oct. 2013.

[12] J. Wang, B. Peng, and X. Zhang, “Using a stacked residual LSTM model for sentiment intensity prediction,” *Neurocomputing*, vol. 322, pp. 93–101, Sep. 2018, doi: 10.1016/j.neucom.2018.09.049.

[13] I. Chaturvedi, E. Cambria, and D. Vilares, *Lyapunov filtering of objectivity for Spanish Sentiment Model*, vol. 2016-Octob. 2016. doi: 10.1109/IJCNN.2016.7727785.

[14] Tien Thanh Vu, Shu Chang, Quang Thuy Ha, and Nigel Collier, *An Experiment in Integrating Sentiment Features for Tech Stock Prediction in Twitter*. 2012.

[15] B. Shin, T. Lee, and J. D. Choi, *Lexicon integrated CNN models with attention for sentiment analysis*. 2017. doi: 10.18653/v1/w17-5220.

[16] P. Shruthi and K. M. Anil Kumar, “Novel approach for generating hybrid features set to effectively identify hate speech,” *Intel. Artif.*, vol. 23, no. 66, pp. 97–111, 2020, doi: 10.4114/intartif.vol23iss66pp97-111.

[17] J. Hu and J. Li, “Sentiment evaluation based hierarchical text representation method for sentiment classification,” *Jisuanji Gongcheng/Computer Eng.*, vol. 46, no. 3, 2020, doi: 10.19678/j.issn.1000-3428.0054521.

[18] R. I. Mukhamediev *et al.*, “Classification of negative information on socially significant topics in mass media,” *Symmetry (Basel).*, vol. 12, no. 12, pp. 1–23, 2020, doi: 10.3390/sym12121945.

[19] M. R. R. Rana *et al.*, “Aspect-Based Sentiment Analysis for Social Multimedia: A Hybrid Computational Framework,” *Comput. Syst. Sci. Eng.*, vol. 46, no. 2, pp. 2415–2428, 2023, doi: 10.32604/csse.2023.035149.

[20] Y. S. Mehanna and M. Mahmuddin, “The Effect of Pre-processing Techniques on the Accuracy of Sentiment Analysis Using Bag-of-Concepts Text Representation,” *SN Comput. Sci.*, vol. 2, no. 4, 2021, doi: 10.1007/s42979-021-00453-7.

[21] A. Kumar, V. Dabas, and P. Hooda, “Text classification algorithms for mining unstructured data: a SWOT analysis,” *Int. J. Inf. Technol.*, vol. 12, Feb. 2018, doi: 10.1007/s41870-017-0072-1.

[22] J. Zhao and X. Gui, “Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis,” *IEEE Access*, vol. PP, p. 1, Feb. 2017, doi: 10.1109/ACCESS.2017.2672677.

[23] T. Parlar, S. Özel, and F. Song, “Analysis of data pre-processing methods for the sentiment analysis of reviews,” *Comput. Sci.*, vol. 20, p. 123, Jan. 2019, doi: 10.7494/csci.2019.20.1.3097.

[24] P. D. Turney, “Thumbs up or thumbs down?,” *Comput. Res. Repos. - CORR*, p. 417, Dec. 2001, doi: 10.3115/1073083.1073153.

[25] S. V. Balshetwar, R. S. Abilash, and R. Dani Jermisha, “Fake news detection in social media based on sentiment analysis using classifier techniques,” *Multimed. Tools Appl.*, vol. 82, no. 23, pp. 35781–35811, 2023, doi: 10.1007/s11042-023-14883-3.

[26] D. Šandor and M. Bagić Babac, “Sarcasm detection in online comments using machine learning,” *Inf. Discov. Deliv.*, 2023, doi: 10.1108/IDD-01-2023-0002.

[27] M. Andre, R. F. Mello, A. Nascimento, R. D. Lins, and D. Gasevic, “Toward Automatic Classification of Online Discussion Messages for Social Presence,” *IEEE Trans. Learn. Technol.*, vol. 14, no. 6, pp. 802–816, 2021, doi: 10.1109/TLT.2022.3150663.

[28] N. Rai, D. Kumar, N. Kaushik, C. Raj, and A. Ali, “Fake News Classification using transformer based enhanced LSTM and BERT,” *Int. J. Cogn. Comput. Eng.*, vol. 3, pp. 98–105, 2022, doi: 10.1016/j.ijcce.2022.03.003.

[29] M. M. Agüero-Torales, A. G. López-Herrera, and D. Vilares, “Multidimensional Affective Analysis for Low-Resource Languages: A Use Case with Guarani-Spanish Code-Switching Language,” *Cognit. Comput.*, vol. 15, no. 4, pp. 1391–1406, 2023, doi: 10.1007/s12559-023-10165-0.

[30] A. Mewada and R. K. Dewang, “SA-ASBA: a hybrid model for aspect-based sentiment analysis using synthetic attention in pre-trained language BERT model with extreme gradient boosting,” *J. Supercomput.*, vol. 79, no. 5, pp. 5516–5551, 2023, doi: 10.1007/s11227-022-04881-x.

[31] D. Kim, “Research On Text Classification Based On Deep Neural Network,” *Int. J. Commun. Networks Inf. Secur.*, vol. 14, no. 1s, pp. 100–113, 2022, doi: 10.17762/IJCNIS.V14I1S.5618.

[32] A. R. Abas, I. Elhenawy, M. Zidan, and M. Othman, “Bert-cnn: A deep learning model for detecting emotions from text,” *Comput. Mater. Contin.*, vol. 71, no. 2, pp. 2943–2961, 2022, doi: 10.32604/cmc.2022.021671.

[33] M. E. Basiri, S. Nemati, M. Abdar, E. Cambria, and U. R. Acharya, “ABCDM: An Attention-based Bidirectional CNN-RNN Deep Model for sentiment analysis,” *Futur. Gener. Comput. Syst.*, vol. 115, pp. 279–294, 2021, doi: 10.1016/j.future.2020.08.005.

[34] M. Wankhade, C. S. R. Annavarapu, and A. Abraham, “CBMAFM: CNN-BiLSTM Multi-Attention Fusion Mechanism for sentiment classification,” *Multimed. Tools Appl.*, 2023, doi: 10.1007/s11042-023-17437-9.

[35] S. A. Waheeb, “Discharge Summaries Based Sentiment Detection Using Multi-Head Attention and CNN-BiGRU,” *Comput. Syst. Sci. Eng.*, vol. 46, no. 1, pp. 981–998, 2023, doi: 10.32604/csse.2023.035753.

[36] Y. Peng and J. Wang, “Rumor detection based on attention cnn and time series of context information,” *Futur. Internet*, vol. 13, no. 11, 2021, doi: 10.3390/fi13110267.

[37] M. Huang, H. Xie, Y. Rao, J. Feng, and F. L. Wang, “Sentiment strength detection with a context-dependent lexicon-based convolutional neural network,” *Inf. Sci. (Ny).*, vol. 520, pp. 389–399, 2020, doi: 10.1016/j.ins.2020.02.026.

[38] K. L. Tan, C. P. Lee, K. S. M. Anbananthen, and K. M. Lim, “RoBERTa-LSTM: A Hybrid Model for Sentiment Analysis With Transformer and Recurrent Neural Network,” *IEEE Access*, vol. 10, pp. 21517–21525, 2022, doi: 10.1109/ACCESS.2022.3152828.

[39] K. Veena, V. Sasirekha, and S. Devi, “Detection of Sarcastic Sentiment Analysis in Tweets Using Lstm With Improved Attention Based Feature Extraction (Iaten),” *J. Theor. Appl. Inf. Technol.*, vol. 101, no. 18, pp. 7369–7380, 2023, [Online]. Available: https://api.elsevier.com/content/abstract/scopus\_id/85175494200

[40] D. Jain, A. Kumar, and G. Garg, “Sarcasm detection in mash-up language using soft-attention based bi-directional LSTM and feature-rich CNN,” *Appl. Soft Comput. J.*, vol. 91, 2020, doi: 10.1016/j.asoc.2020.106198.

[41] A. Verma *et al.*, “Identification of Hate Speech on Social Media using LSTM,” *GMSARN Int. J.*, vol. 17, no. 4, pp. 468–474, 2023, [Online]. Available: https://api.elsevier.com/content/abstract/scopus\_id/85150831135

[42] C. J. Hutto and E. Gilbert, “VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text,” in *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*, 2015.

[43] C. K. C. Yu, “Automated Analysis of Dream Sentiment—The Royal Road to Dream Dynamics?,” *Dreaming*, vol. 32, no. 1, pp. 33–51, 2022, doi: 10.1037/drm0000189.

[44] L. M. G. Salazar, O. Loyola-González, and M. A. Medina-Pérez, “An explainable approach based on emotion and sentiment features for detecting people with mental disorders on social networks,” *Appl. Sci.*, vol. 11, no. 22, 2021, doi: 10.3390/app112210932.

[45] M. M. Krishna, Midhunchakkaravarthy, and J. Vankara, “Detection of Sarcasm Using Bi-Directional RNN Based Deep Learning Model in Sentiment Analysis,” *J. Adv. Res. Appl. Sci. Eng. Technol.*, vol. 31, no. 2, pp. 352–362, 2023, doi: 10.37934/araset.31.2.352362.

[46] C. Zheng, M. Bouazizi, and T. Ohtsuki, “An Evaluation on Information Composition in Dementia Detection Based on Speech,” *IEEE Access*, 2022, doi: 10.1109/ACCESS.2022.3203068.

[47] H. Chouikhi, M. Alsuhaibani, and F. Jarray, “BERT-Based Joint Model for Aspect Term Extraction and Aspect Polarity Detection in Arabic Text,” *Electron.*, vol. 12, no. 3, 2023, doi: 10.3390/electronics12030515.

[48] S. A. Waheeb, N. A. Khan, and X. Shang, “an Efficient Sentiment Analysis Based Deep Learning Classification Model To Evaluate Treatment Quality,” *Malaysian J. Comput. Sci.*, vol. 35, no. 1, pp. 1–20, 2022, doi: 10.22452/mjcs.vol35no1.1.

[49] M. Jbene, S. Tigani, S. Rachid, and A. Chehri, “Deep neural network and boosting based hybrid quality ranking for e-commerce product search,” *Big Data Cogn. Comput.*, vol. 5, no. 3, 2021, doi: 10.3390/bdcc5030035.

[50] S. Giri, S. Das, S. B. Das, and S. Banerjee, “SMS Spam Classification-Simple Deep Learning Models With Higher Accuracy Using BUNOW And GloVe Word Embedding,” *J. Appl. Sci. Eng.*, vol. 26, no. 10, pp. 1501–1511, 2023, doi: 10.6180/jase.202310\_26(10).0015.

[51] Y. Li *et al.*, “PecidRL: Petition expectation correction and identification based on deep reinforcement learning,” *Inf. Process. Manag.*, vol. 60, no. 3, 2023, doi: 10.1016/j.ipm.2023.103285.

[52] Z. Ren, Q. Shen, X. Diao, and H. Xu, “A sentiment-aware deep learning approach for personality detection from text,” *Inf. Process. Manag.*, vol. 58, no. 3, 2021, doi: 10.1016/j.ipm.2021.102532.

[53] A. Alzu’bi, L. B. Younis, A. Abuarqoub, and M. Hammoudeh, “Multimodal Deep Learning with Discriminant Descriptors for Offensive Memes Detection,” *J. Data Inf. Qual.*, vol. 15, no. 3, 2023, doi: 10.1145/3597308.

[54] R. Akila and S. Revathi, “Fine Grained Analysis of Intention for Social Media Reviews Using Distance Measure and Deep Learning Technique,” *J. Internet Serv. Inf. Secur.*, vol. 13, no. 2, pp. 48–64, 2023, doi: 10.58346/JISIS.2023.I2.003.

[55] B. Kitchenham and S. Charters, “Guidelines for performing Systematic Literature Reviews in Software Engineering,” vol. 2, Jan. 2007.

[56] S. Hardani and D. Kristiyanti, “Systematic Literature Review: Analisa Sentimen Penerimaan Masyarakat Terhadap Jenis Vaksin Covid-19 Di Dunia,” *ikraith-informatika*, vol. 6, Nov. 2022, doi: 10.37817/ikraith-informatika.v6i3.2204.

[57] B. Kitchenham, “Procedures for Performing Systematic Reviews,” *Keele, UK, Keele Univ.*, vol. 33, Aug. 2004.

[58] Annur, Cindy. (2022, December 29). Perusahaan F&B dengan Gerai Waralaba Terbanyak Global (2021). Databoks. [https://databoks.katadata.co.id/datapublish/ 2022/12/29/mixue-punya-gerai-terbanyak-ke-5-di-dunia-burger-king-kalah](https://databoks.katadata.co.id/datapublish/%202022/12/29/mixue-punya-gerai-terbanyak-ke-5-di-dunia-burger-king-kalah)

# **LAMPIRAN**

Lampiran A Hasil Turnitin

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a web page

Description automatically generated

**FORMULIR KONSULTASI SKRIPSI – FAKULTAS TEKNIK & INFORMATIKA**

**Dosen Pembimbing : Dinar Ajeng Kristiyanti, S.Kom., M.Kom.**

**Jurusan : Sistem Informasi**

**Semester : 5**

**Nama : Dava Virgio Kertawijaya**

**NIM : 00000056848**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tanggal Konsultasi** | **Agenda/Pokok Bahasan** | **Saran Perbaikan** | **Paraf Dosen Pembimbing** |
| **22 September 2023** | **Bimbingan Perdana** | **Mempelajari integrasi tools SLR: Publish or Perish, Mendeley & Vos Viewer** |  |
| **29 September 2023** | **Pengajuan Topik 1** | **Lakukan SLR untuk mencari keterbaruan penelitian** |  |
| **6 Oktober 2023** | **Revisi SLR 1** | **Kata kunci yang digunakan tidak spesifik membahas Sentiment Strength Detection, harap lebih rinci lagi kata kuncinya** |  |
| **20 Oktober 2023** | **Laporan Proposal MRSI** | **Dilengkapi lagi penulisan yang masih kurang** |  |
| **10 November 2023** | **Revisi SLR 2** | **Kata kunci yang digunakan terlalu spesifik membahas Sentiment Strength Detection, harap tidak terlalu rinci dan tidak terlalu umum kata kuncinya** |  |
| **24 November 2023** | **Menentukan Topik** | **Mencari objek penelitian dalam pencarian dataset** |  |
| **27 November 2023** | **Revisi Bab 1, Bab 2, Bab 3 Proposal MRSI, PPT, Serta Artikel SLR** | **Diperhatikan jumlah referensi artikel SLR dan proposal MRSI minimal 50. Jelaskan apa yang mau dibuat pada Bab 3, tambahkan metode dan hasil SLR.** |  |
| **28 November 2023** | **Finalisai Proposal MRSI, PPT, Serta Artikel SLR** |  |  |

Catatan : Form ini wajib dibawa pada saat konsultasi & dilampirkan didalam skripsi (**Minimal 8 kali Konsultasi**)

Tangerang, 28 November 2023

Dinar Ajeng Kristiyanti, S.Kom., M.Kom.

Dosen Pembimbing