

DavaVirgioKertawijaya_000000 56848_mrsi

by --

Submission date: 28-Nov-2023 09:43AM (UTC-0500)

Submission ID: 2240838354

File name: DavaVirgioKertawijaya_00000056848_FINAL2_removed.pdf (444.5K)

Word count: 5646

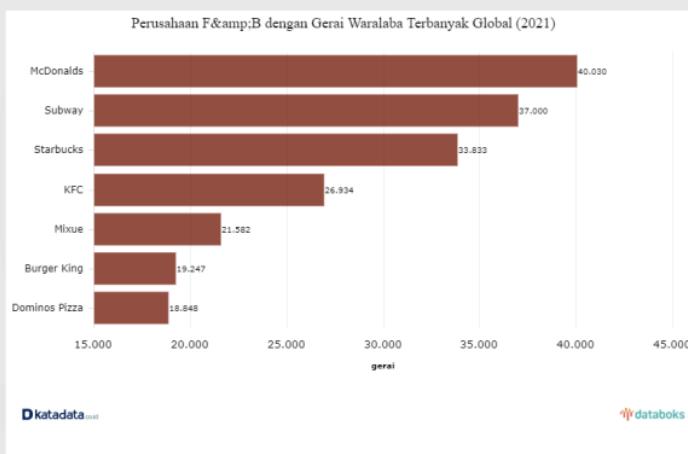
Character count: 38087

7 BAB I

PENDAHULUAN

1.1 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, terutama 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.

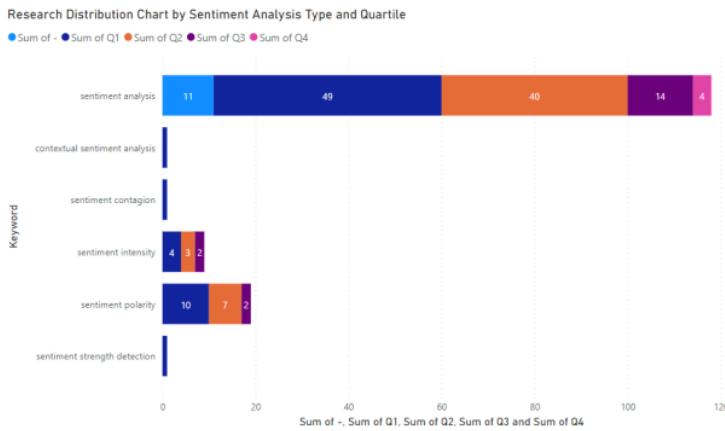


Gambar 1.1 Perusahaan F & B dengan Gerai Waralaba Terbanyak Global 2021[58]

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.



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 sentimen atau sentimen polarity 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 mengeksplorasi 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,¹¹ 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.

18

1.2 Rumusan Masalah

Berikut merupakan rumusan masalah dari penelitian ini:

1. Bagaimana menerapkan algoritma berbasis deep learning untuk deteksi kekuatan sentimen pada ulasan produk McDonald's?
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 (pR^2)?

11

1.3 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. Menggunakan bahasa pemrograman Python untuk melakukan analisis.

2 **1.4 Tujuan dan Manfaat Penelitian**

1.4.1 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 (pR^2) dari pendekatan CNN, LSTM, dan Bi-LSTM dengan leksikon SSS-Lex, dan Vader.

2 **1.4.2 Manfaat Penelitian**

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

6 **1.4.2.1 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.

6 **1.4.2.2 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.

- 33
- Hasil penelitian ini dapat memberikan dasar untuk penelitian lebih lanjut di bidang analisis sentimen, baik dalam konteks produk McDonald's maupun dalam domain lain.

1.5 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

10
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 Language 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

8
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 II²⁰

LANDASAN TEORI

2.1 Tinjauan Teori

2.1.1 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 ³ 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].

2.1.2 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.

2.1.3 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].

2.1.4 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 berkaitan 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].

2.1.4.1 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.

2.1.4.2 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.

2.1.4.3 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.

5

2.1.4.4 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].

2.1.4.5 Lemmatization

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

2.1.4.6 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.

3

2.1.4.7 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.

2.1.5 Metode-Metode Analisis Sentimen

27

2.1.5.1 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.

2.1.5.2 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.

2.1.5.3 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].

2.1.5.4 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.

2.1.5.5 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.

2.2 Framework

Berikut merupakan beberapa framework data mining yang dapat diterapkan pada penelitian deteksi kekuatan sentimen, serta algoritma yang akan digunakan dalam penelitian ini.⁵

16

2.2.1 CRISP-DM

CRISP-DM (Cross-Industry Standard Process for Data Mining)³⁶ 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.

14

2.2.2 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.

19

2.2.3 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.

2.3 Tools

35

2.3.1 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.

2.3.1.1 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.

2.3.1.2 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.

30

2.3.2 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.

31

2.3.2.1 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.

2.3.2.2 TextBlob

TextBlob adalah perpustakaan NLP Python yang menekankan kesederhanaan dan kemudahan penggunaan. TextBlob menyediakan antarmuka yang intuitif dan alat yang mudah digunakan, memungkinkan

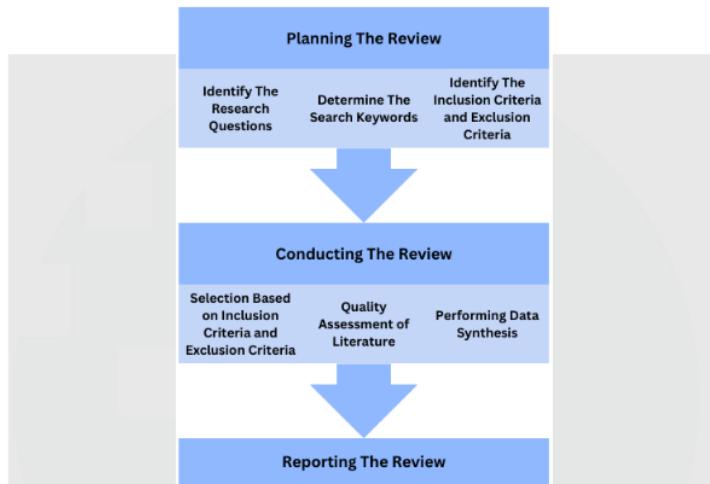
peneliti dengan berbagai tingkat keahlian untuk melakukan analisis teks [53].

2.3.2.3 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.

2.4 Penelitian Terdahulu

Sebuah Kajian Literatur Sistematik (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 Sistematic 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.



Gambar 2.1 Tahapan Systematic Literature Review [56]

2.4.1 Planning The Review

2.4.1.1 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 pendekstrian kekuatan sentimen menggunakan leksikon berbasis konteks dijalankan?
- b. RQ 2: Apa saja jurnal yang pernah menerbitkan artikel mengenai pendekstrian kekuatan sentimen dengan pendekatan leksikon berbasis konteks?

- c. RQ 3: Apa saja kata kunci yang digunakan dalam pendekripsi 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 pendekripsi kekuatan sentimen dengan leksikon berbasis konteks?
- f. RQ 6: Metode dan teknologi apa yang digunakan dalam penelitian pendekripsi kekuatan sentimen dengan leksikon berbasis konteks?
- g. RQ 7: Platform data apa yang digunakan dalam penelitian pendekripsi kekuatan sentimen dengan leksikon berbasis konteks?

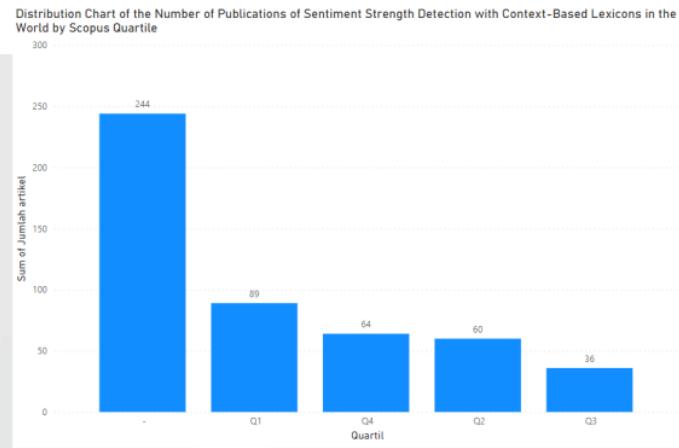
Dari rangkuman research question yang tercantum di atas, langkah berikutnya adalah menetapkan kata kunci pencarian untuk fase selanjutnya.

2.4.1.2 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") ³⁷ AND ("machine learning" OR "deep learning" OR "text mining" OR "data mining" ³² 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.



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

2.4.1.3 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 diterbitkan sebelum 2020
2	Jurnal penelitian yang ditulis dalam bahasa Inggris	Jurnal penelitian tidak dalam 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

2.4.2 Conducting The Review

2.4.2.1 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.

2.4.2.2 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:

- Artikel yang memiliki reputasi dan terindeks di Scopus.
- Hanya artikel jurnal yang dimasukkan dalam penelitian.
- Setiap langkah dalam sistematik literature review dilakukan sesuai dengan pedoman yang telah ditetapkan.

2.4.2.3 Penyintesan Data

Sintesis Proses penyintesan 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 penyintesan 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.

2.4.3 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.

BAB III

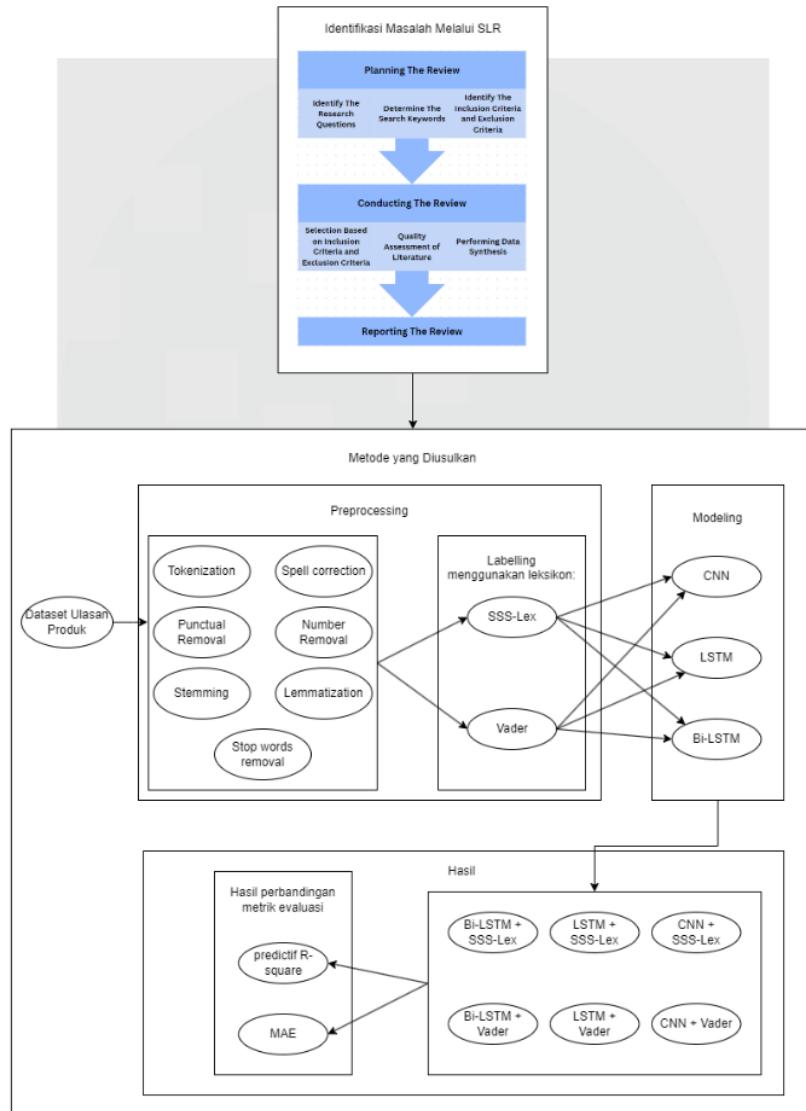
METODOLOGI PENELITIAN

3.1 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.

3.2 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 pR². 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

12 3.3 Variabel Penelitian

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

3.3.1 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

3.3.2 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.

3.4 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.

3.5 Teknik Pengambilan Sampel

Data ulasan produk McDonald's yang digunakan berasal dari Kaggle,²³ 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.

7

3.6 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.

29

3.6.1 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.

3.6.2 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.

3.6.3 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

33

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.

3.6.4 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.

13%

SIMILARITY INDEX

12%

INTERNET SOURCES

4%

PUBLICATIONS

3%

STUDENT PAPERS

PRIMARY SOURCES

- | | | |
|---|--|------|
| 1 | journals.upi-yai.ac.id
Internet Source | 3% |
| 2 | kc.umn.ac.id
Internet Source | 1 % |
| 3 | Anggreiny Rolangon, Axcel Weku, Green Arther Sandag. "Perbandingan Algoritma LSTM Untuk Analisis Sentimen Pengguna Twitter Terhadap Layanan Rumah Sakit Saat Pandemi Covid-19", TeIKA, 2023
Publication | 1 % |
| 4 | 123dok.com
Internet Source | <1 % |
| 5 | Dany Pratmanto, Fabriyan Fandi Dwi Imaniawan. "Analisis Sentimen Terhadap Aplikasi Canva Menggunakan Algoritma Naive Bayes Dan K-Nearest Neighbors", Computer Science (CO-SCIENCE), 2023
Publication | <1 % |
| 6 | repository.unhas.ac.id
Internet Source | <1 % |

7	docplayer.info Internet Source	<1 %
8	library.umn.ac.id Internet Source	<1 %
9	text-id.123dok.com Internet Source	<1 %
10	Submitted to Universitas Singaperbangsa Karawang Student Paper	<1 %
11	repository.binadarma.ac.id Internet Source	<1 %
12	dspace.uii.ac.id Internet Source	<1 %
13	Submitted to State Islamic University of Alauddin Makassar Student Paper	<1 %
14	ecampus.pelitabangsa.ac.id Internet Source	<1 %
15	Submitted to Konsorsium PTS Indonesia - Small Campus Student Paper	<1 %
16	download.garuda.kemdikbud.go.id Internet Source	<1 %
17	ejurnal.ung.ac.id Internet Source	<1 %

18	repository.unj.ac.id Internet Source	<1 %
19	Submitted to INTI International University Student Paper	<1 %
20	Submitted to iGroup Student Paper	<1 %
21	repository.its.ac.id Internet Source	<1 %
22	id.scribd.com Internet Source	<1 %
23	repository.unair.ac.id Internet Source	<1 %
24	widuri.raharja.info Internet Source	<1 %
25	www.gudangmakalah.com Internet Source	<1 %
26	Boby Pranata, Susanti. "Support Vector Machine untuk Sentiment Analysis Bakal Calon Presiden Republik Indonesia 2024", Indonesian Journal of Computer Science, 2023 Publication	<1 %
27	www.diva-portal.org Internet Source	<1 %
28	www.jawaraspeed.com Internet Source	<1 %

29	publikasi.dinus.ac.id Internet Source	<1 %
30	sunscrapers.com Internet Source	<1 %
31	www.boxer.or.id Internet Source	<1 %
32	Camila Corrêa Matias Pereira. "A comunicação nos ambientes virtuais e o comportamento suicida", Universidade de Sao Paulo, Agencia USP de Gestao da Informacao Academica (AGUIA), 2021 Publication	<1 %
33	doku.pub Internet Source	<1 %
34	jurnal.unai.edu Internet Source	<1 %
35	spectrum.library.concordia.ca Internet Source	<1 %
36	www.scribd.com Internet Source	<1 %
37	www.mdpi.com Internet Source	<1 %

Exclude quotes On

Exclude matches Off

Exclude bibliography On

DavaVirgioKertawijaya_00000058919_mrsi

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17

PAGE 18

PAGE 19

PAGE 20

PAGE 21

PAGE 22

PAGE 23

PAGE 24

PAGE 25

PAGE 26

PAGE 27
