

BAB II

TINJAUAN PUSTAKA

Bab II digunakan untuk menjelaskan penelitian terdahulu dan teori-teori yang berkaitan dan mendukung penelitian.

2.1 Studi Sebelumnya

Steam sebagai platform distribusi *game* digital terbesar di dunia telah menjadi objek penelitian dalam berbagai studi akademik. Studi yang dilakukan oleh Zhang [20] menunjukkan bahwa ulasan pengguna di Steam dapat menjadi indikator kuat terhadap keberhasilan *game*. Mereka menemukan bahwa ulasan yang berisi kritik teknis cenderung memiliki dampak lebih besar terhadap skor akhir *game* dibandingkan dengan ulasan yang membahas elemen naratif atau *gameplay*.

Fadhlurrahman dkk. [21] dalam penelitiannya melakukan klasifikasi sentimen ulasan pengguna di platform Steam menggunakan pendekatan BERT, BiLSTM, dan CRF. Penelitian ini menunjukkan bahwa ulasan negatif cenderung lebih panjang dan lebih mendalam dibandingkan dengan ulasan positif. Selain itu, mereka menemukan bahwa BERT memberikan akurasi yang lebih tinggi dibandingkan CRF, sedangkan LSTM mampu menangkap konteks kalimat lebih baik dalam analisis sentimen berbasis teks panjang dikarenakan LSTM lebih kompleks dibandingkan yang lainnya.

Britto dkk. [22] dalam penelitiannya membahas bagaimana perubahan mekanisme permainan dalam sebuah *game* dapat memengaruhi reaksi komunitas pemain. Studi ini membandingkan sentimen ulasan sebelum dan sesudah adanya update besar dalam sebuah *game*. Hasil penelitian menunjukkan bahwa perubahan *gameplay* sering kali menimbulkan reaksi beragam, terutama di antara penggemar lama yang sudah terbiasa dengan mekanisme sebelumnya.

Ramadhan dkk. [23] dalam studinya mengatakan bahwa analisis sentimen merupakan alat yang sangat relevan untuk digunakan. Metode ini memungkinkan klasifikasi opini pengguna berdasarkan kategori sentimen positif, negatif, atau netral, sehingga dapat memberikan pemahaman yang lebih komprehensif mengenai pengalaman dan pandangan pengguna terhadap layanan yang disediakan.

Dalam penelitian di luar industri *game*, Ruziqiana [24] membandingkan efektivitas metode Naïve Bayes, SVM, dan Random Forest dalam klasifikasi sentimen. Hasil penelitian menunjukkan bahwa Naïve Bayes unggul dalam analisis teks pendek, sementara SVM lebih unggul dalam menangani teks yang lebih kompleks.

Studi oleh Rasyidi[25] dalam penelitiannya meneliti fenomena "*trash talking*" dalam konten *game* online, menunjukkan bagaimana analisis konten dapat mengungkap pola komunikasi yang spesifik dalam komunitas *game*.

Penelitian oleh Nurdy dkk.[26] mengaplikasikan analisis sentimen pada ulasan *game* di platform digital. Penelitian ini membuktikan bahwa analisis konten dapat digunakan untuk mengukur dan memahami sentimen pemain terhadap suatu *game* berdasarkan ulasan yang mereka berikan.

Sari dkk.[27] dalam penelitiannya menganalisis bagaimana Bank Syariah Indonesia memanfaatkan konten Instagram sebagai strategi pemasaran digital untuk menarik minat calon nasabah. Penelitian ini menggunakan metode deskriptif kualitatif dengan teknik pengumpulan data melalui dokumentasi postingan Instagram Bank Syariah Indonesia. Hasil penelitian menunjukkan bahwa konten-konten yang disajikan merupakan strategi pemasaran yang efektif dalam menarik minat calon nasabah di era digital.

No	Penulis	Tahun	Metode	Tujuan	Temuan
1	Zhang [20]	2022	MLR, LASSO, RIDGE, SVR, MLPR	Menganalisis pengaruh ulasan pengguna terhadap skor <i>game</i> di Steam.	Ulasan kritik teknis lebih berdampak pada skor <i>game</i> di Steam.
2	Fadhlurrahman dkk. [21]	2023	BERT, BiLSTM, CRF	Mengklasifikasikan sentimen ulasan <i>game</i> di Steam menggunakan model deep learning.	Ulasan negatif lebih panjang; BERT lebih akurat dari CRF; LSTM lebih baik dalam konteks kalimat Panjang, dikarenakan LSTM lebih kompleks
3	Britto dkk. [22]	2020	Random Forest, SVM, LR	Mengevaluasi penerimaan <i>game</i> berdasarkan analisis sentimen ulasan.	Perubahan <i>gameplay</i> menimbulkan reaksi beragam, terutama dari penggemar lama.
4	Ramadhan dkk.[23]	2024	BERT	Mengimplementasikan model BERT untuk analisis sentimen ulasan <i>game</i> .	Analisis sentimen relevan untuk memahami pengalaman pengguna.
5	Ruziqiana[24]	2020	Naïve Bayes, SVM, Random Forest	Membandingkan efektivitas algoritma klasifikasi sentimen pada komentar Instagram.	Naïve Bayes unggul dalam teks pendek, SVM untuk teks kompleks.
6	Rasyidi [25]	2024	Analisis Konten	Menganalisis konten "trash talking" dalam <i>game</i> online <i>Mobile Legend</i> di YouTube.	Analisis konten mengungkap pola komunikasi spesifik dalam komunitas <i>game</i> .
7	Nurdy dkk. [26]	2024	Naïve Bayes	Menganalisis sentimen ulasan <i>game</i> Stumble Guys di Playstore.	Analisis konten mengukur sentimen pemain berdasarkan ulasan.
8	Sari dkk. [27]	2023	Analisis Konten	Menganalisis konten Instagram sebagai strategi digital marketing Bank Syariah Indonesia	Hasil penelitian menunjukkan bahwa konten-konten yang disajikan merupakan strategi pemasaran yang

					efektif dalam menarik minat calon nasabah di era digital.
--	--	--	--	--	---

Tabel 2.1 Penelitian Terdahulu



2.2 Dasar Teori

2.2.1 Natural Language Processing (NLP)

Natural language processing (NLP), merupakan salah satu pendekatan terkomputerisasi untuk menganalisa teks berdasarkan aspek teori dan teknologi. Menurut Liddy, NLP didefinisikan sebagai sebuah bidang teoritis mengenai suatu teknik komputasional yang digunakan untuk menganalisa dan merepresentasikan teks yang ditulis secara natural (bahasa manusia) pada satu atau lebih level analisis linguistik dengan tujuan untuk memperoleh *human-like language processing* yang dapat diimplementasikan dalam berbagai bidang[28].

2.2.2 Analisis Sentimen

Analisis sentimen adalah teknik dalam pemrosesan bahasa alami (*Natural Language Processing/NLP*) yang bertujuan untuk mengidentifikasi dan mengklasifikasikan opini atau emosi dalam suatu teks [29]. Teknologi ini digunakan dalam berbagai bidang, seperti pemasaran digital, layanan pelanggan, dan analisis opini publik di media sosial.

Analisis sentimen sering kali disebut sebagai *opinion mining* karena fokusnya dalam mengekstrak informasi subjektif dari data tekstual[30]. Proses ini memungkinkan organisasi untuk memahami bagaimana perasaan pelanggan terhadap produk, layanan, atau merek mereka, serta membantu dalam pengambilan keputusan berbasis data[31]. Secara umum, analisis sentimen dapat dikelompokkan ke dalam tiga kategori utama:

1. **Analisis Sentimen Polaritas** – Menentukan apakah sentimen suatu teks bersifat positif, negatif, atau netral.
2. **Analisis Sentimen Berbasis Emosi** – Mengidentifikasi jenis emosi tertentu dalam teks, seperti marah, bahagia, sedih, atau takut [32].
3. **Analisis Sentimen Aspek-Sentris** – Menganalisis sentimen yang berhubungan dengan aspek tertentu dari suatu entitas [33].

Teknik dalam analisis sentimen dapat dibagi menjadi beberapa pendekatan utama:

1. Pendekatan Berbasis Kamus (*Lexicon-Based Approach*)

Pendekatan ini menggunakan daftar kata yang telah dikategorikan berdasarkan sentimen (positif atau negatif). Beberapa contoh kamus yang sering digunakan dalam analisis sentimen adalah SentiWordNet dan AFINN [34],[35].

Kelebihan:

- Tidak memerlukan data pelatihan yang besar.
- Mudah diimplementasikan untuk berbagai bahasa.

Kekurangan:

- Tidak dapat memahami konteks atau makna kata dalam kalimat.
- Kesulitan dalam menangani ironi, sarkasme, atau kata-kata yang memiliki makna ganda.

2. Pendekatan Berbasis Pembelajaran Mesin (*Machine Learning-Based Approach*)

Metode ini menggunakan algoritma pembelajaran mesin yang dilatih pada dataset berlabel untuk mengklasifikasikan sentimen [36]. Algoritma yang umum digunakan dalam analisis sentimen berbasis pembelajaran mesin meliputi:

- ***Naïve Bayes (NB)*** – Algoritma probabilistik yang sering digunakan untuk klasifikasi teks[37].
- ***Support Vector Machine (SVM)*** – Model berbasis vektor yang mencari hiperbidang terbaik untuk memisahkan kelas sentimen [38].
- ***Random Forest (RF)*** – Algoritma berbasis pohon keputusan yang dapat digunakan untuk klasifikasi sentimen[39].

Kelebihan:

- Lebih akurat dibandingkan pendekatan berbasis kamus.
- Dapat menangkap pola kompleks dalam teks.

Kekurangan:

- Membutuhkan dataset pelatihan yang cukup besar.
- Performanya dapat dipengaruhi oleh kualitas fitur yang diekstraksi dari teks.

3. Pendekatan Berbasis Deep Learning (*Deep Learning-Based Approach*)

Dengan kemajuan dalam jaringan saraf tiruan (*Artificial Neural Networks*), metode berbasis deep learning semakin populer dalam analisis sentimen[40]. Beberapa model deep learning yang sering digunakan adalah:

- ***Long Short-Term Memory (LSTM)*** – Versi RNN yang lebih baik dalam menangani dependensi jangka panjang dalam teks[41].

- ***Bidirectional Encoder Representations from Transformers (BERT)*** – Model berbasis *Transformer* yang dapat memahami konteks kata dalam kalimat lebih baik dibandingkan model lain [42].

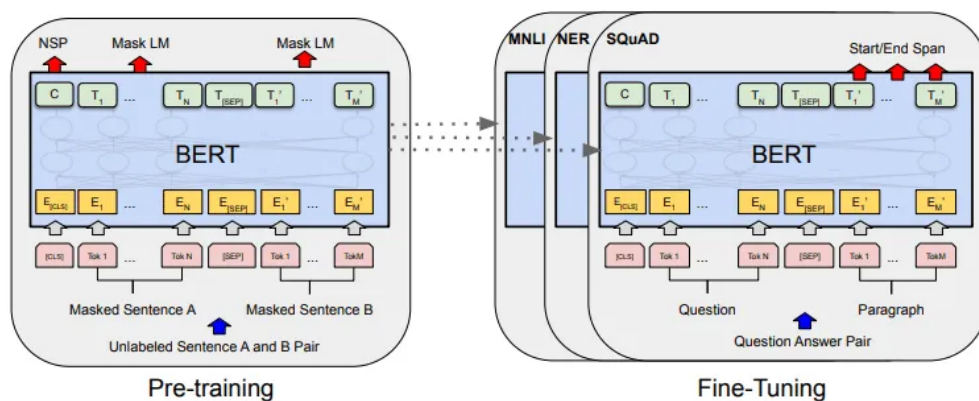
2.2.3 BERT (Bidirectional Encoder Representations from Transformers)

BERT (*Bidirectional Encoder Representations from Transformers*) adalah model bahasa berbasis deep learning yang diperkenalkan oleh Google AI pada tahun 2018 [43]. Model ini menggunakan arsitektur *Transformer* dan dirancang untuk memahami konteks dalam teks secara *bidirectional* (dua arah) [43]. BERT telah merevolusi pemrosesan bahasa alami (*Natural Language Processing/NLP*) dengan kemampuannya untuk menangkap makna dari suatu kata berdasarkan konteks keseluruhan dalam suatu kalimat [43].

BERT dibangun di atas arsitektur *Transformer* [44]. *Transformer* menggunakan mekanisme *self-attention*, yang memungkinkan model memproses setiap kata dalam teks secara paralel, tidak seperti pendekatan berbasis RNN atau LSTM yang bersifat sekuensial[45]. Komponen utama *Transformer* yang digunakan dalam BERT:

- ***Self-Attention***: Menghitung hubungan antar kata dalam sebuah kalimat, terlepas dari urutan kata[46].
- ***Feed-Forward Layers***: Lapisan fully connected untuk memperkuat fitur yang diekstrak[46].
- ***Positional Encoding***: Menambahkan informasi posisi kata dalam kalimat karena *Transformer* tidak memiliki pemrosesan sekuensial seperti RNN [46].

2.2.3.1 Pre- Training dan Fine-Tuning



Gambar 2.1 Arsitektur Pre-Training dan Fine-Tuning dalam BERT

Tahap *pre-training* BERT melibatkan proses pelatihan model menggunakan kumpulan teks yang sangat besar[47]. Pada tahap ini, model belajar memahami struktur bahasa dengan cara memprediksi kata-kata yang disembunyikan dalam sebuah kalimat (*Masked Language Model*) serta menentukan hubungan antara dua kalimat (*Next Sentence Prediction*)[48]. Proses ini memungkinkan BERT untuk membangun representasi bahasa yang kontekstual dan bidirectional[49].

Setelah tahap *pre-training*, model BERT yang telah dilatih dapat di-*fine-tune* untuk tugas-tugas spesifik. Selama tahap *fine-tuning*, beberapa lapisan atas pada model BERT disesuaikan dengan dataset tugas tertentu, seperti klasifikasi teks, *Named Entity Recognition* (NER), atau *Question Answering* (QA)[46]. Dengan *fine-tuning*, BERT dapat menyesuaikan diri dengan kebutuhan spesifik suatu tugas dan menghasilkan hasil yang lebih akurat dan relevan [50].

2.2.3.2 DistilBERT

DistilBERT merupakan versi ringan dari model BERT (Bidirectional Encoder Representations from Transformers) yang dikembangkan melalui teknik *knowledge distillation*. Tujuan dari pengembangan DistilBERT adalah untuk menghasilkan model dengan performa mendekati BERT namun dengan ukuran yang lebih kecil, kecepatan lebih tinggi, dan efisiensi komputasi yang lebih baik[51].

DistilBERT memiliki sekitar 40% lebih sedikit parameter dibandingkan BERT-base, berjalan sekitar 60% lebih cepat, namun tetap mempertahankan lebih dari 95% akurasi pada berbagai tugas Natural Language Processing (NLP). Hal ini dicapai dengan mentransfer pengetahuan dari model BERT yang lebih besar ke dalam model DistilBERT melalui proses *distillation*, di mana model siswa (DistilBERT) dilatih untuk meniru keluaran model guru (BERT)[51].

DistilBERT tetap menggunakan arsitektur Transformer encoder seperti BERT, tetapi dengan beberapa penyederhanaan, seperti penghapusan token-type embeddings dan pengurangan jumlah lapisan dari 12 menjadi 6. Meskipun demikian, DistilBERT tetap dapat menangani tugas-tugas NLP seperti analisis sentimen, klasifikasi teks, dan *question answering* secara efektif.

Dalam konteks analisis sentimen, penggunaan DistilBERT sangat relevan karena kemampuannya dalam memahami konteks kalimat secara bidirectional dan efisiensinya dalam hal sumber daya komputasi. Oleh karena itu, model ini sangat cocok

digunakan untuk penelitian yang memerlukan klasifikasi teks dalam jumlah besar namun dengan keterbatasan perangkat keras[52].

2.3 Tabel Perbandingan

Tabel 2.2 Tabel Perbandingan antara SVM, LSTM, dan BERT

Aspek	<i>Naïve Bayes</i>	<i>SVM (Support Vector Machine)</i>	<i>LSTM (Long Short-Term Memory)</i>	BERT <i>(Bidirectional Encoder Representations from Transformers)</i>
Metode	Probabilistik (berbasis Teorema Bayes)	Pemisahan data berdasarkan hyperplane optimal	Jaringan saraf berulang dengan memori jangka panjang	Transformer-based model dengan pemahaman kontekstual bidirectional
Pemahaman Konteks	Rendah (tidak mempertimbangkan urutan kata)	Sedang (bergantung pada representasi fitur)	Tinggi (memahami urutan kata)	Sangat Tinggi (memahami hubungan antar kata dalam dua arah)
Kecepatan Pelatihan	Cepat	Sedang	Lambat	Sangat lambat
Kecepatan Inferensi	Cepat	Cepat	Sedang	Lambat
Akurasi pada <i>Review Steam</i>	Rendah (karena tidak menangkap hubungan kata dengan baik)	Sedang (tergantung pemrosesan fitur)	Tinggi (menangkap konteks <i>review</i> lebih baik)	Sangat tinggi (memahami makna ulasan secara mendalam)
Kemampuan Generalisasi	Rendah	Baik	Baik	Sangat baik
Kompleksitas Implementasi	Mudah	Sedang	Tinggi	Sangat tinggi
Kebutuhan Data	Kecil hingga sedang	Sedang	Besar	Sangat besar