

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Studi Literatur

Ada beberapa penelitian terdahulu yang dijadikan referensi utama pada Tabel 1.

*Tabel 1 Penelitian Terdahulu*

<b>Penulis</b>	<b>Judul(Tahun)</b>	<b>Hasil</b>	<b>Perbandingan</b>
Muhammad Fadhil Mubaraq	<i>Sentiment Analysis on Twitter Social Media towards Climate Change on Indonesia Using IndoBERT Model(2022)</i>	Pada penelitian tersebut, peneliti berhasil menyempurnakan model IndoBERT untuk menerapkan analisis sentimen dengan topik perubahan iklim di Indonesia pada media Twiter. Hasil rata-rata terbaik F1-Score 95,6% dengan <i>learning rate</i> $2e-5$ dengan 16 batch[6].	Perbandingan dari penelitian ini yaitu ada pada jumlah data, pada penelitian Muhammad Fadhil Mubaraq data yang digunakan yaitu berjumlah 1533 tweet. Sedangkan jumlah data yang digunakan pada penelitian ini data yang digunakan yaitu 7448 data. Karena jumlah data dapat mempengaruhi model dalam melakukan pelatihan. Semakin banyak data yang digunakan semakin besar kemungkinan untuk mendapatkan hasil akurasi yang baik.
Priyan Fadhil Supriyadi	<i>Xiaomi Smartphone Sentiment Analysis on Twitter Social Media Using IndoBERT(2023)</i>	Didasarkan pada hasil pengujian serta analisis yang dilakukan, pertama, hasil pengujian yang dibandingkan antara model	Pada penelitian Priyan Fadhil Supriyadi telah dilakukan perbandingan melakukan pelatihan dataset berbahasa

		<p>BERT dan IndoBERT menunjukkan bahwa BERT membuat lebih banyak kesalahan prediksi sentimen daripada IndoBERT. Skor rata-rata akurasi sebesar 98% pada jumlah <i>epoch</i> 4 [7].</p>	<p>Indonesia menggunakan BERT dan IndoBERT. Pada penggunaan BERT didapatkan hasil rata-rata akurasi 87% dan IndoBERT 98%. Hal tersebut mencari referensi bahwa penggunaan IndoBERT pada klasifikasi teks berbahasa Indonesia dapat menghasilkan akurasi pelatihan model lebih baik karena IndoBERT memiliki token yang dilatih dari lebih dari 220 juta kata bahasa Indonesia, sedangkan BERT melakukan pelatihan bahasa Indonesia menggunakan <i>multilingual</i> yang dapat menerjemahkan semua bahasa.</p>
<p>D Muhammad Sholeh Abdurohim</p>	<p>Analisis Sentimen Peraturan Pengeras Suara di Masjid pada Sosial Media Twitter Menggunakan Algoritma IndoBERT(2023)</p>	<p>Pada penelitian ini algoritma IndoBERT berhasil mendapatkan f1-score 73% dari 420 data tweet[8].</p>	<p>Pada penelitian D Muhammad Sholeh Abdurohim data yang digunakan yaitu berjumlah 420 data sehingga model hasil yang didapatkan kurang akurat dikarenakan model terlalu sederhana dalam</p>

			menangkap pola yang ada dalam data. Oleh karena itu, kekurangan tersebut menjadi acuan pada penelitian ini.
Lenggo Geni	<i>Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia using IndoBERT Language Model(2023)</i>	Peneliti menganalisis sentimen pada opini publik Indonesia untuk pemilu 2024 mendatang menggunakan data Twitter. Peneliti menggunakan model IndoBERT dengan akurasi sebesar 83,5% [9].	Pada penelitian Lenggo Geni, data sentimen yang digunakan tidak seimbang yaitu 837 data netral, 82 data negatif dan 81 data positif dengan total 1000 data. Hal tersebut menjadi acuan penelitian ini untuk melakukan <i>data balancing</i> untuk mendapatkan hasil yang lebih akurat.

## 2.2 Komparasi Metode

Sub Bab ini menjelaskan perbandingan penggunaan metode IndoBERT dengan model lain.

### 1. Naïve Bayes

Naïve Bayes adalah model klasifikasi yang sederhana yang mengasumsikan independensi antara fitur-fiturnya, artinya model ini menganggap setiap kata dalam ulasan sebagai fitur yang independen satu sama lain. Ketika naïve bayes diberikan ulasan, model akan melihat frekuensi kemunculan kata-kata tertentu dalam kumpulan data pelatihan untuk menentukan probabilitas ulasan tersebut termasuk dalam kelas sentimen yang berbeda (positif, negatif atau netral). Naïve Bayes tidak dapat menangkap konteks kalimat dan hubungan antar kata-kata dengan baik seperti yang dilakukan model IndoBERT. Penelitian Lenggo Geni pada tabel 1 dengan judul *Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia using IndoBERT Language Model(2023)* melakukan perbandingan metode IndoBERT dengan naïve bayes, hasil rata-rata yang didapatkan dari penggunaan naïve bayes yaitu 74% sedangkan IndoBERT 83,4% yang dapat disimpulkan bahwa

model IndoBERT dapat memahami hubungan antar kata-kata dengan lebih baik sehingga didapatkan hasil yang lebih akurat.

## 2. Support Vector Machine

*Support Vector Machine(SVM)* adalah algoritma klasifikasi yang bekerja dengan menemukan *hyperplane* terbaik yang memisahkan *instance* dari kelas yang berbeda dalam ruang fitur. Ketika SVM diberikan data teks, model ini mencoba untuk menemukan *hyperplane* terbaik yang memisahkan ulasan positif dari ulasan negatif berdasarkan fitur-fitur yang diekstraksi dari teks. SVM tidak memiliki pemahaman tentang konteks atau hubungan antar kata-kata dalam kalimat. Ini hanya bekerja dengan fitur yang diekstraksi dari data teks. Misalnya, jika sebuah ulasan mengandung kata-kata positif dan negatif, SVM akan mencoba untuk menemukan *hyperplane* yang memisahkan ulasan tersebut berdasarkan kemunculan kata-kata tersebut, tanpa memperhatikan konteks atau nuansa kalimat. Penelitian yang membandingkan metode Support Vector Machine dan BERT dilakukan oleh Husnul Mawaddah dengan judul “Analisis dan Perbandingan Klasifikasi *Hate Speech* Menggunakan Metode Fitur Manual SVM dan BERT” didapatkan hasil rata-rata yaitu 76% menggunakan SVM dan 98% menggunakan BERT. Artinya model BERT melakukan klasifikasi dengan lebih akurat[10].

Perbandingan dari 2 metode di atas menunjukkan keunggulan metode IndoBERT dibandingkan dengan Naïve Bayes dan Support Vector Machine. Dari perbandingan di atas dapat disimpulkan bahwa IndoBERT dapat melakukan klasifikasi teks lebih baik karena IndoBERT dapat memahami konteks dan hubungan antar kata-kata sehingga mendapatkan hasil yang lebih baik dan lebih akurat.

### 2.3 Analisis Sentimen

Analisis sentimen ialah proses menganalisis opini masyarakat pada media online untuk menentukan emosional dari penulisnya. Analisis ini penting untuk dapat mengetahui emosi atau sentimen masyarakat terhadap sebuah topik, brand dan sebagainya bernada positif, negatif atau netral. Dalam konteks penelitian ini, sentimen berguna untuk mengetahui kepercayaan dan kepuasan masyarakat terhadap instansi. Sentimen positif, negatif dan netral dapat membantu instansi untuk evaluasi layanan berdasarkan keluhan masyarakat pada opini tersebut.

Analisis sentimen juga memanfaatkan berbagai teknik dalam pemrosesan bahasa alami(NLP) untuk memahami dan mengklasifikasi teks berdasarkan ekspresi emosi yang terkandung di dalamnya. Selain itu, penggunaan model pembelajaran mesin dapat membantu dalam prediksi sentimen dengan akurasi yang tinggi.

## 2.4 Penggunaan IndoBERT dalam Analisis Sentimen

BERT adalah model bahasa yang mampu memahami konteks dan nuansa dalam teks dengan tingkat akurasi yang tinggi. Dengan menggunakan BERT, kita dapat mengurai pesan teks pada sosial media X yang sering kali singkat, adanya bahasa informal dan memahami perasaan dan pandangan yang terkandung dalamnya dengan lebih baik[4]. Untuk melakukan tugas pemrosesan bahasa alami berbahasa Indonesia, juga terdapat IndoBERT yang merupakan model BERT yang dilatih untuk memahami bahasa Indonesia. IndoBERT dapat mengenali sentimen atau perasaan yang terkandung dalam teks. IndoBERT dilatih dengan lebih dari 220 juta kata bahasa Indonesia, diantaranya 74 juta dari Wikipedia Indonesia, 55 juta dari artikel berita Kompas, tempo dan Liputan6 serta 90 juta dari Web Corpus Indonesia.

BERT menggunakan pendekatan transformer yang memungkinkan model untuk memahami konteks dari setiap kata dalam teks dengan lebih baik. Ini berarti BERT dapat menghasilkan representasi yang lebih akurat untuk setiap kata, berdasarkan konteks di sekitarnya. BERT dapat melihat konteks dari kedua arah, baik dari kiri ke kanan maupun sebaliknya. Hal ini memungkinkan BERT untuk memahami hubungan antara kata-kata yang jauh di dalam kalimat, yang sering kali penting untuk pemahaman yang akurat. BERT telah berhasil diterapkan pada berbagai tugas pemrosesan bahasa alami, termasuk analisis sentimen, pemahaman teks, penerjemahan mesin dan lainnya. BERT memiliki keunggulan dalam pemahaman konteks yang kuat, kemampuan transfer learning yang efisien, dan ketersediaan sebagai open-source. Ini menjadikan BERT sebagai salah satu model bahasa yang populer dan efektif dalam pemrosesan bahasa alami saat ini.