

BAB II

TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 Media Sosial

Media sosial adalah platform digital yang memungkinkan pengguna untuk berinteraksi, berbagi informasi, dan membangun jaringan dengan orang lain secara *online*. Berbagai bentuk media sosial termasuk Facebook, Twitter, Instagram, dan YouTube telah menjadi alat yang sangat penting bagi masyarakat modern dalam berkomunikasi dan menyebarkan informasi. Media sosial menawarkan ruang di mana individu, komunitas, dan organisasi dapat berbagi opini, ide, dan konten (A. Tiwari & Singh, 2019).

Media sosial memainkan peran sentral dalam membentuk opini publik dan memungkinkan ekspresi pribadi secara luas. Platform ini memungkinkan penyebaran informasi secara cepat dan interaktif, serta memfasilitasi diskusi tentang peristiwa terkini dan isu-isu sosial. Media sosial juga dapat memperbesar polarisasi dalam masyarakat karena pengguna cenderung mencari informasi yang sesuai dengan pandangan pribadi mereka, sebuah fenomena yang dikenal sebagai "*echo chambers*" (F. Defersha & Tune, 2021).

Dalam konteks pemerintah, media sosial digunakan sebagai sarana untuk berkomunikasi dengan publik serta mengukur persepsi masyarakat terhadap kebijakan yang diterapkan. Khususnya di Indonesia, Dewan Perwakilan Rakyat Republik Indonesia (DPR RI) dapat memantau opini publik untuk memahami bagaimana masyarakat menanggapi kebijakan yang dibuat (H. Sadat, 2022).

2.1.2 Analisis Sentimen

Sentimen analisis atau *Opinion Mining* adalah studi komputasi untuk mengenali dan mengekspresikan opini, sentimen, evaluasi, sikap, emosi, subjektivitas, penilaian atau pandangan yang terdapat dalam suatu teks. Teks yang dimaksud dapat berupa artikel, komentar di *social media* ataupun dari sumber lainnya. (Satya Marga dkk., 2021).

Analisis sentimen digunakan untuk mengelompokkan kalimat atau teks ke dalam kalimat atau dokumen kemudian menentukan bagaimana pendapat yang dikemukakan dalam kalimat yang dianalisis apakah bersifat positif, negatif ataupun netral. Secara umum analisis sentimen dilakukan pada 3 level:

a. *Document Level*

Pada level ini analisis bertugas untuk mengklasifikasikan apakah seluruh pendapat dokumen mengekspresikan sentimen positif atau negatif.

b. *Sentence Level*

Pada level ini, analisis bertugas untuk menentukan apakah setiap kalimat menyatakan pendapat positif, negatif, atau netral. Kalimat netral berarti bahwa tidak ada pendapat yang dinyatakan dalam kalimat tersebut.

c. *Entity and Aspect Level*

Pada level ini, analisis sampai pada tahap menentukan sentimen positif atau negatif berdasarkan aspek (target) yang terdapat dalam suatu kalimat. Level yang digunakan dalam proyek ini adalah *Aspect Level*.

2.1.3 *Data Preprocessing*

Data *preprocessing* merupakan proses pengolahan data asli yang sebelumnya tidak terstruktur menjadi terstruktur dengan tujuan agar data yang digunakan dalam proses analisis sentimen bersih dari *noise*. *Preprocessing* merupakan langkah awal dalam analisis sentimen yang dapat meningkatkan performa model *classifier*. Dalam tahap *preprocessing* ini, dilakukan penghapusan *URL*, *user mention*, *hashtag*, tanda baca, dan lain sebagainya.

Tahapan *Preprocessing* :

1. *Case Folding*

Pada tahap ini, seluruh kata dalam sebuah kalimat diubah menjadi huruf kecil. Proses ini diperlukan agar semua kata setara dan tidak terjadi kesalahan dalam pemberian bobot pada suatu kata (Filemon dkk., 2022).

2. *N-grams*

N-Gram adalah penggabungan kata yang sering muncul bersamaan dengan tujuan menunjukkan suatu sentimen. Terdapat beberapa jenis token dalam *N-gram*, di antaranya *Unigram*, *Bigram*, dan *Trigram*. *Unigram* merupakan token yang hanya terdiri dari satu kata, *Bigram* terdiri dari dua kata, dan *Trigram* terdiri dari tiga kata. *N-gram* digunakan karena dalam bahasa Indonesia banyak frasa yang terdiri lebih dari satu kata (Satya Marga dkk., 2021).

3. Tokenisasi

Tokenisasi merupakan proses memisahkan setiap kata dalam sebuah kalimat berdasarkan spasi di antara kata-kata dalam kalimat tersebut. Tokenisasi memilah

kalimat menjadi kata-kata dan membuat daftar, yaitu setiap kalimat menjadi daftar kata. Pemisahan kata dilakukan berdasarkan spasi di antara kata-kata (Darwis dkk., 2020). Hasil dari proses tokenisasi akan digunakan sebagai masukan dalam tahap transformasi teks ke bentuk angka.

4. *Lematisasi*

Lematisasi merupakan teknik yang digunakan untuk mengubah kata berimbuhan menjadi kata dasar dengan memperhatikan makna yang dimaksud serta disesuaikan dengan Kamus Besar Bahasa Indonesia (Aby Vonega dkk., 2022).

5. *Modeling*

Modeling adalah proses pengembangan atau pembuatan model dari suatu sistem. Dalam analisis sentimen, *modeling* adalah pembuatan atau penggunaan algoritma yang dapat melakukan analisis teks untuk menentukan sentimen yang terkandung dalam teks. Proses membangun model dilakukan berdasarkan dataset yang telah dibersihkan (M. I. Fikri dkk., 2020).

2.1.4 Komputasi Hijau

Komputasi hijau adalah konsep yang berkaitan dengan desain, pembuatan, dan penggunaan perangkat keras serta perangkat lunak komputasi secara efisien untuk meminimalkan dampak lingkungan. Tujuan utama komputasi hijau adalah untuk mengurangi konsumsi energi, emisi karbon, dan limbah elektronik (Poongodi, Sharma, dkk., 2020) .

Prinsip Komputasi Hijau Komputasi hijau menurut (Kumar, Gupta, & Chauhan, 2023) menekankan pada:

1. Efisiensi Energi

Penggunaan perangkat keras dan perangkat lunak yang meminimalkan penggunaan listrik dan sumber daya lainnya.

2. Pengurangan Emisi Karbon

Menggunakan sumber daya komputasi yang ramah lingkungan dengan memanfaatkan teknologi hemat energi seperti *CPU* dan perangkat penyimpanan yang hemat energi.

3. Pengurangan Dampak Lingkungan

Menerapkan praktik desain dan pengembangan yang mempertimbangkan dampak ekologis selama siklus hidup perangkat teknologi.

Dalam konteks pembelajaran mesin (*machine learning*), penerapan komputasi hijau mencakup penggunaan algoritma yang lebih efisien dan teknik optimasi untuk mengurangi penggunaan sumber daya komputasi (Zubar & Balamurugan, 2020).

Beberapa contoh praktis penerapan komputasi hijau dalam pemodelan *machine learning* termasuk:

1. Pemilihan Algoritma yang Efisien

Algoritma seperti *Random Forest* dan *Extra Trees* memiliki keunggulan dalam efisiensi komputasi karena mereka mendukung pemrosesan *paralel*. Pemrosesan *paralel* memungkinkan model untuk mengoptimalkan penggunaan sumber daya secara lebih efisien, yang pada akhirnya mengurangi waktu komputasi dan

konsumsi energi (M. Gupta dkk., 2023). Dibandingkan dengan algoritma yang lebih kompleks seperti *deep learning*, algoritma *ensemble* ini memerlukan daya komputasi yang lebih rendah, yang berkontribusi pada upaya pengurangan energi.

2. Penggunaan Infrastruktur *Cloud* Hemat Energi

Komputasi hijau juga dapat diimplementasikan melalui penggunaan infrastruktur *cloud* yang hemat energi. Layanan *cloud* seperti *Google Cloud* dan *Amazon Web Services* (AWS) telah mengadopsi teknologi ramah lingkungan dengan menggunakan pusat data yang efisien dalam hal penggunaan daya. Dengan memanfaatkan layanan ini, organisasi dapat menekan konsumsi energi yang dibutuhkan untuk pelatihan dan implementasi model *machine learning* (Poongodi, Rad, dkk., 2020). Infrastruktur *cloud* ini tidak hanya memberikan skalabilitas tetapi juga memastikan bahwa sumber daya digunakan secara efisien dan sesuai dengan prinsip keberlanjutan.

3. Optimasi Model melalui Reduksi Dimensi dan *Pruning*

Reduksi dimensi dengan teknik seperti *Principal Component Analysis* (PCA) dapat digunakan untuk mengurangi kompleksitas model dengan mengeliminasi fitur-fitur yang tidak relevan. Dengan menurunkan dimensi data, algoritma dapat beroperasi dengan lebih sedikit sumber daya komputasi tanpa mengurangi performa prediksi secara signifikan (Zhou dkk., 2021). Selain itu, teknik *model pruning* dapat diterapkan pada model seperti Random Forest dan Extra Trees untuk menghilangkan elemen-elemen yang tidak diperlukan, sehingga ukuran model lebih kecil dan konsumsi daya lebih rendah selama fase inferensi.

4. Praktik *Quantization* untuk Penghematan Energi

Quantization adalah teknik yang digunakan untuk mengurangi presisi numerik dalam representasi data dan model tanpa mengorbankan akurasi prediksi. Penggunaan *quantization* dapat memperkecil ukuran model, yang berarti lebih sedikit penggunaan memori dan daya selama fase komputasi (Zhou dkk., 2021). Dengan menerapkan teknik ini, algoritma seperti *Extra Trees* dan *Random Forest* dapat menjadi lebih hemat energi dan lebih efisien dalam hal performa operasional.

5. Algoritma Optimasi dan Pengurangan Penggunaan Energi pada Model Ensemble

Algoritma *ensemble* seperti *Random Forest* dan *Extra Trees* tidak hanya efisien dalam segi akurasi, tetapi juga lebih efektif dalam mengurangi waktu komputasi karena pembagian tugas secara *paralel*. Selain itu, penelitian terbaru menunjukkan bahwa *Random Forest* yang dioptimalkan dengan metode pemangkasan pohon (*tree pruning*) serta penggunaan metode berbasis aturan dapat menurunkan konsumsi energi hingga 30% tanpa kehilangan akurasi yang signifikan (M. Gupta dkk., 2023).

Melalui implementasi prinsip-prinsip komputasi hijau ini, analisis sentimen berbasis machine learning dapat dilakukan dengan meminimalkan jejak lingkungan, sambil tetap mempertahankan efisiensi dalam hal akurasi dan kecepatan pemrosesan (M. Gupta dkk., 2023). Ini menunjukkan bagaimana penggabungan inovasi teknologi dengan prinsip keberlanjutan dapat mendukung tujuan komputasi yang lebih hijau dan ramah lingkungan.

2.2 Penelitian Terkait

2.2.1 *State of The Art*

Penelitian terdahulu dalam bidang analisis sentimen telah mengeksplorasi berbagai algoritma seperti *Naïve Bayes*, *Support Vector Machine (SVM)*, *Decision Tree*, hingga *Random Forest*. Fokus utama dari sebagian besar studi ini adalah mengevaluasi tingkat akurasi dan efektivitas model dalam mengklasifikasikan opini publik di media sosial. Misalnya, Adrian et al. (2021) menemukan bahwa *Random Forest* memiliki keunggulan akurasi dibandingkan SVM dalam analisis sentimen terkait kebijakan PSBB. Penelitian lain oleh Fikri et al. (2020) juga memperkuat dominasi pendekatan konvensional dengan membandingkan SVM dan *Naïve Bayes* di *platform* Twitter.

Meskipun studi-studi tersebut memberikan kontribusi penting dalam mengidentifikasi metode klasifikasi yang andal, sebagian besar masih bersifat teknis dan belum secara eksplisit mempertimbangkan aspek efisiensi komputasi. Parameter seperti konsumsi energi, waktu proses, dan dampak ekologis model masih jarang digunakan sebagai bagian dari evaluasi. Ini menunjukkan adanya celah penting dalam literatur, terutama dalam konteks penerapan model analisis sentimen pada data berskala besar seperti media sosial.

Selain itu, sebagian besar penelitian terdahulu juga berfokus pada domain komersial (*e-commerce*, aplikasi konsumen) atau isu sosial umum. Studi yang secara khusus mengkaji opini publik terhadap lembaga pemerintahan, seperti DPR RI, masih sangat terbatas. Padahal, pemahaman terhadap persepsi masyarakat

terhadap institusi negara sangat penting dalam mendukung pengambilan kebijakan berbasis data.

Oleh karena itu, penelitian ini hadir dengan kontribusi utama berupa pendekatan yang mengintegrasikan performa akurasi model dengan prinsip efisiensi energi dalam konteks institusi publik. Kombinasi algoritma *ensemble* seperti *Extra Trees* dan *Random Forest* diuji tidak hanya berdasarkan akurasi, tetapi juga berdasarkan efisiensi energi, untuk menjawab tantangan keberlanjutan dalam analisis data besar. Rangkuman posisi dan kontribusi dari beberapa penelitian terdahulu dapat dilihat pada Tabel 2.1.

Tabel 2. 1 State Of the Art

No	Peneliti	Model	Fokus Penelitian	Gap Penelitian
1.	(Adrian dkk., 2021)	Random Forest, SVM	Analisis sentimen PSBB menggunakan Random Forest dan SVM	Belum mempertimbangkan efisiensi energi dan belum membandingkan Extra Trees
2.	(Amaliah dkk., 2022)	Lexicon Based, Naïve Bayes	Analisis sentimen aplikasi investasi di Twitter	Belum menggunakan ensemble method dan tidak mengkaji konsumsi energi
3.	(Asshiddiqi & Lhaksmana, 2020)	Decision Tree, SVM	Analisis sentimen kinerja PSSI di Instagram	Belum mempertimbangkan green computing
4.	(Rasyidin Muqsith Rizqi Prasetyo & Musthafa, 2023)	Naïve Bayes, SVM	Analisis sentimen relokasi ibu kota	Tidak menggunakan ensemble learning dan belum mempertimbangkan efisiensi energi
5.	(Kotelnikova, 2020)	Deep Learning, Rule-Based	Perbandingan metode DL dan Rule-Based dalam tugas analisis sentimen	Tidak membahas efisiensi energi dan konsumsi sumber daya
6.	(Dwi Ahmad Dzulhijjah dkk., 2023)	Random Forest, KNN	Analisis sentimen di Twitter	Belum membandingkan konsumsi energi antara model ensemble secara spesifik
7.	(Swati Agarwal dkk., 2022)	Random Forest, XGBoost, SVM	Analisis sentimen berbasis green computing menggunakan ensemble classifier	Belum membahas aspek interpretabilitas dan analisis pada sektor publik
8.	(Mahesh Kumar & Neha Tyagi, 2023)	Umum (ML di Edge)	Sentiment analysis hemat energi di lingkungan <i>edge computing</i>	Tidak membandingkan performa energi antar model secara eksplisit
9.	(Jingjing Xu dkk., 2021)	CNN, RNN, LSTM, BERT	Survei <i>green deep learning</i> dan teknik efisiensi model	Fokus hanya pada DL; tidak membahas model ringan seperti RF atau ETC
10.	(Lacoste dkk., 2019)	Umum (Machine Learning)	Kalkulator emisi karbon dari pelatihan model ML	Belum meninjau pengaruh metode ensemble terhadap efisiensi energi
11.	(Lannelongue dkk., 2020)	Umum	Framework umum untuk menghitung jejak karbon	Belum dikaitkan dengan performa model NLP/ML spesifik

No	Peneliti	Model	Fokus Penelitian	Gap Penelitian
12.	(Strubell dkk., 2019)	Transformer, BERT, NAS	Estimasi emisi karbon dalam NLP	Fokus hanya pada NLP dan tidak bahas alternatif model hemat energi
13.	(Thakur dkk., 2024)	LSTM	Analisis sentimen dan topik seputar Pemilu Indonesia 2024 pada komentar YouTube	Belum mempertimbangkan efisiensi energi atau pembanding model ringan
14.	(Geni dkk., 2023)	IndoBERT, SVM	Analisis sentimen terhadap Pemilu 2024 menggunakan model BERT Indonesia	Tidak membahas konsumsi daya dan tidak membandingkan model hemat energi
15.	(Sembiring & Dewa, 2025)	Naïve Bayes, SVM, LSTM	Perbandingan model sentimen terhadap Pemilu 2024 di Twitter	Tidak mempertimbangkan efisiensi energi dan tidak meninjau interpretabilitas model
16.	(Ma’aly dkk., 2024)	CNN, Bi-LSTM	Analisis emosi multi-label pada debat calon presiden 2024	Belum mempertimbangkan konsumsi energi dan tidak membandingkan efisiensi antar model
17.	Penelitian ini	Random Forest, Extra Trees	DPR RI, akurasi dan efisiensi energi	Menambahkan aspek komputasi hijau serta penggunaan ensemble learning

2.2.2 Matriks Penelitian

Tabel 2. 2 Matriks Penelitian

No.	Penulis	Model/Algoritma								Metrik Perbandingan							
		Random Forest	Support Vector Machine	Naïve Bayes	Lexicon Based	Decision Tree	Deep Learning	Rule Base	KNN	Extra Trees	Akurasi	Presisi	Recall	F1-Score	Macro Average	Weighted Average	Green Computing
1.	(Adrian dkk., 2021)	✓	✓								✓	✓	✓	✓	✓	✓	
2.	(Amaliah dkk., 2022)			✓	✓						✓	✓	✓	✓	✓	✓	

No.	Penulis	Model/Algoritma									Metrik Perbandingan						
		Random Forest	Support Vector Machine	Naïve Bayes	Lexicon Based	Decision Tree	Deep Learning	Rule Base	KNN	Extra Trees	Akurasi	Presisi	Recall	F1-Score	Macro Average	Weighted Average	Green Computing
3.	(Asshiddiqi & Lhaksana, 2020)		√			√					√	√	√	√			
4.	(Rasyidin & Musthafa, 2023)		√	√							√						
6.	(Kotelnikova, 2020)						√	√			√			√			
7.	(Dwi Ahmad Dzulhijjah dkk., 2023)	√							√		√	√	√	√			
8.	(Swati Agarwal dkk., 2022)	√		√							√	√	√	√			√
9.	(Mahesh Kumar & Neha Tyagi, 2023)	√	√								√	√	√	√			√
10.	(Jingjing Xu dkk., 2021)						√				√	√	√	√			√
11.	(Lacoste dkk., 2019)																√
12.	(Lannelongue dkk., 2020)																√
13.	(Strubell dkk., 2019)						√				√						√
14.	(Thakur dkk., 2024)						√				√	√	√	√			
15.	(Geni dkk., 2023)						√				√			√			
16.	(Sembiring & Dewa, 2025)		√	√			√				√						
17.	(Ma'aly dkk., 2024)						√				√	√	√	√			
18.	Penelitian ini	√								√	√	√	√	√	√	√	√

Berdasarkan tinjauan terhadap studi-studi terdahulu yang tercantum dalam Tabel 2.2, dapat disimpulkan bahwa meskipun telah banyak dilakukan perbandingan algoritma klasifikasi untuk analisis sentimen, fokus utama sebagian besar penelitian masih terbatas pada aspek akurasi. Belum ditemukan studi yang secara eksplisit membandingkan performa algoritma *ensemble* seperti *Extra Trees* dan *Random Forest* dengan mempertimbangkan efisiensi energi sebagai bagian dari pendekatan komputasi hijau. Selain itu, konteks penelitian terkait institusi publik, khususnya DPR RI, masih sangat terbatas dalam literatur. Oleh karena itu, penelitian ini menempati posisi yang strategis karena menghadirkan pendekatan baru yang mengintegrasikan evaluasi performa model dengan prinsip efisiensi komputasi dalam domain analisis kebijakan publik, yang menjadi kontribusi utama penelitian ini.

2.3 Sintesis Tinjauan Pustaka

Berdasarkan tinjauan literatur, sebagian besar penelitian analisis sentimen masih berfokus pada aspek performa akurasi dan pengolahan teks. Sangat sedikit studi yang mempertimbangkan aspek efisiensi energi dan berkelanjutan dalam pemrosesan data skala besar. Oleh karena itu, penelitian ini berupaya untuk mengisi kekosongan tersebut dengan :

1. Mengintegrasikan evaluasi akurasi dan efisiensi energi
2. Mengaplikasikan prinsip komputasi hijau pada algoritma ensemble
3. Memfokuskan objek kajian pada opini publik terhadap DPR RI