

BAB I

PENDAHULUAN

A. Latar Belakang

Perkembangan teknologi informasi dan komunikasi di Indonesia mengalami pertumbuhan yang cukup pesat dalam beberapa tahun terakhir. Berdasarkan data dari Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) [1], jumlah pengguna internet di Indonesia pada tahun 2024 mencapai 221.563.479 jiwa dari total populasi 278.696.200 jiwa, dengan tingkat penetrasi sebesar 79,5%. Angka ini menunjukkan bahwa hampir mayoritas orang Indonesia sudah menggunakan internet secara aktif.

Seiring dengan meningkatnya pengguna internet di Indonesia, kebutuhan akan informasi yang cepat dan akurat semakin penting di era digital saat ini. Salah satu bentuk informasi yang sering diakses oleh masyarakat adalah berita yang tersebar di berbagai media massa. Saat ini semua pengguna internet dapat mengakses berita dengan cepat dengan melakukan pencarian di peramban dan perangkat yang memiliki jaringan daring akan menyediakan informasi yang diinginkan konsumen [2]. Dalam sebuah berita, terdapat elemen-elemen penting seperti judul dan isi.

Menurut Dewi et al. [3] judul sering kali dijadikan dasar penilaian oleh pembaca tanpa membaca isi keseluruhan berita. Hal ini dapat menyebabkan kesalahan persepsi dari suatu berita. Kesalahan persepsi terhadap informasi atau berita dapat berdampak terhadap pembentukan opini publik serta menurunkan tingkat kepercayaan publik terhadap institusi [4]. Hal ini bisa berisiko memicu konflik sosial. Pada pemilu Amerika Serikat tahun 2016, penyebaran berita palsu dan judul berita yang menyesatkan berdampak pada manipulasi opini pemilih, polarisasi politik, serta menurunkan kepercayaan publik terhadap media. Konflik sosial lain terjadi pada masa pandemi COVID-19 di India, dimana misinformasi yang menghubungkan penyebaran virus dengan kelompok tertentu memicu kekerasan fisik dan perlakuan diskriminatif terhadap anggota komunitas di beberapa daerah di India [5]. Kedua kasus tersebut adalah bukti bahwa kesalahan persepsi terhadap informasi atau berita dapat berdampak langsung pada konflik

sosial. Di sisi lain, judul berita dinilai memiliki potensi untuk menggiring opini pembaca dan mencerminkan ideologi media terhadap isu yang disampaikan.

Media massa adalah media penyebaran informasi yang berhubungan langsung dengan masyarakat luas misalnya radio, televisi, surat kabar, dan internet. Salah satu media massa yang paling banyak dikonsumsi publik adalah media berita online, karena bisa diakses dimana saja serta kapan saja selama terkoneksi dengan internet [6]. Di Indonesia terdapat beberapa media massa terpopuler antara lain, Detik, Kompas, TribunNews, Tempo, dan CNN Indonesia. Selain media berita *online*, media sosial juga menjadi media massa yang sering digunakan oleh masyarakat Indonesia sebagai tempat mencari berita. Beberapa media sosial yang sering digunakan oleh masyarakat Indonesia antara lain, Instagram, TikTok, Facebook, dan YouTube.

Salah satu berita yang sedang menjadi perbincangan masyarakat Indonesia adalah program makan bergizi gratis (MBG) yang baru diterapkan di tahun 2025 ini. Makan bergizi gratis adalah program yang dibentuk oleh Presiden Prabowo Subianto yang berfokus pada pemenuhan kebutuhan gizi masyarakat, khususnya anak-anak dan kelompok rentan [7]. Program ini resmi dimulai pada 6 Januari 2025 di 190 Satuan Pelayanan Pemenuhan Gizi (SPPG) pada 26 provinsi di Indonesia, yang menargetkan balita, santri, siswa PAUD, TK, SD, SMP, SMA, ibu hamil, dan ibu menyusui [8].

Program ini memiliki tujuan dan manfaat yang baik, namun keberadaannya menimbulkan beragam kontroversi di tengah masyarakat. Sebagian pihak menganggap program ini sebagai upaya yang membantu mengatasi persoalan gizi buruk. Namun, ada juga pihak yang melihat program makan bergizi gratis ini tidak optimal atau tidak merata distribusinya [9]. Adanya pro dan kontra terkait program makan bergizi gratis ini membuat publik kesulitan menilai apakah program ini memiliki sentimen positif, negatif, atau netral. Oleh karena itu, penelitian ini akan berfokus pada analisis sentimen judul berita terkait program makan bergizi gratis atau MBG.

Menurut Rifqi dan Yunianto [10] analisis sentimen adalah proses menganalisis teks untuk mengidentifikasi apakah maknanya memiliki sentimen positif, negatif, atau netral. Dalam beberapa tahun terakhir, berbagai algoritma

Machine Learning maupun *Deep Learning* telah diimplementasikan untuk analisis sentimen dengan tingkat keberhasilan yang bervariasi. Setiadi et al. [11] menganalisis sentimen multi aspek dari review hotel berbahasa Inggris dengan memanfaatkan model *RoBERTa Twitter*. Aspek yang dianalisis adalah kamar, pelayanan, lokasi, fasilitas, harga, dan sentimen. Hasilnya model *RoBERTa Twitter* mendapat akurasi aspek sentimen sebesar 88%. Untuk aspek fasilitas sebesar 75%, aspek pelayanan 78%, aspek lokasi 84%, aspek kamar 81%, dan aspek harga 84%. Selanjutnya, Pookduang et al. [12] mengevaluasi efektivitas model RoBERTa, dalam analisis sentimen ulasan buku di Amazon dan membandingkannya dengan model tradisional (*Naive Bayes*, KNN, CART) serta model LSTM. Hasil pengujian menunjukkan bahwa RoBERTa unggul secara signifikan, dengan akurasi mencapai 96,30% jauh melebihi model lain termasuk LSTM dengan akurasi 91,20%.

Implementasi model berbasis Transformer, khususnya RoBERTa telah memajukan bidang *Natural Language Processing* (NLP) secara signifikan. Model berbasis Transformer seperti RoBERTa, unggul dalam menangkap makna dalam kalimat dan hubungan kontekstual [12]. Hal ini membuat RoBERTa dipilih dalam penelitian ini untuk diimplementasikan dalam analisis sentimen judul berita.

Berdasarkan uraian yang telah dijelaskan sebelumnya, RoBERTa menunjukkan hasil yang lebih baik dalam analisis sentimen dibandingkan dengan algoritma tradisional. Perbedaan penelitian ini dengan penelitian terdahulu adalah penelitian ini berfokus pada model RoBERTa untuk klasifikasi sentimen dengan data judul berita terkait topik makan bergizi gratis atau MBG berbahasa Indonesia. Sumber data judul berita adalah dari tiga media massa yang dipilih yaitu Detik, Kompas, dan YouTube. Pemilihan Detik, Kompas, dan YouTube sebagai sumber data didasarkan pada jangkauannya sebagai media massa *online* paling banyak diakses oleh masyarakat Indonesia di awal tahun 2025.

Menurut data *Digital News Report* dari *Reuters Institute* [13], Detik merupakan media berita *online* paling banyak diakses di Indonesia dengan proporsi sebanyak 46% responden, diikuti Kompas sebesar 37%, TvOne News dan Tribunnews sama yaitu sebesar 26%, CNN Indonesia dan Metro TV News sama yaitu sebesar 24%, Liputan 6 sebesar 22%, Tempo sebesar 16%, Kumparan sebesar

15%, dan RCTI News sebesar 13%. Pemilihan Detik dan Kompas mewakili mayoritas media berita *online* yang paling banyak digunakan di Indonesia. Sementara itu, berdasarkan sumber yang sama media sosial paling banyak digunakan sebagai sumber berita bagi masyarakat Indonesia adalah Whatsapp dengan proporsi sebesar 43% responden, diikuti YouTube sebesar 41%, Facebook sebesar 39%, TikTok sebesar 34%, Instagram sebesar 31%, dan X sebesar 13%. YouTube dipilih untuk mewakili media sosial paling banyak digunakan sebagai sumber berita bagi masyarakat Indonesia. Hal ini dikarenakan Whatsapp yang menempati urutan pertama lebih bersifat pribadi dan tertutup. Gabungan dari Detik, Kompas, dan YouTube diharapkan memberikan spektrum sentimen yang lebih luas. Oleh karena itu, penulis melakukan penelitian yang berjudul “Analisis Sentimen Judul Berita pada Media Massa Menggunakan RoBERTa”.

B. Perumusan Masalah

Berdasarkan latar belakang di atas maka rumusan masalahnya adalah “Bagaimana Kinerja Algoritma RoBERTa dalam Analisis Sentimen Judul Berita pada Media Massa”.

C. Batasan Penelitian

Adapun batasan masalah yang diberlakukan dalam penelitian ini adalah sebagai berikut:

1. Analisis sentimen judul berita difokuskan pada ruang lingkup makan bergizi gratis atau MBG berbahasa Indonesia.
2. Dataset judul berita didapat melalui proses *scraping* dari tiga media massa yaitu Detik, Kompas, dan YouTube.
3. Proses *scraping* menggunakan *library* Python BeautifulSoup dan Youtube Data V3 API.
4. Data yang digunakan dalam rentang 1 Januari 2025 – 31 Maret 2025 sebanyak 2093 data.
5. Pelabelan data menggunakan model IndoBERT *taufiqdp/indonesian-sentiment*.

6. Klasifikasi sentimen dikategorikan menjadi tiga yaitu negatif, netral, dan positif.
7. Kinerja yang diukur adalah berdasarkan metrik *accuracy*, *precision*, *recall*, *f1-score*, waktu pelatihan model, dan sumber daya komputasi yang diperlukan.

D. Tujuan Penelitian

Penelitian ini bertujuan untuk mengukur kinerja algoritma RoBERTa dalam mengklasifikasikan sentimen judul berita berbahasa Indonesia dari media massa Detik, Kompas, dan YouTube terkait topik makan bergizi gratis.

E. Manfaat Penelitian

Terdapat beberapa manfaat yang bisa didapat dari penelitian ini. Manfaat dari penelitian ini meliputi:

1. Menambah wawasan terkait bidang *Natural Language Processing* (NLP) melalui implementasi model Transformer RoBERTa untuk analisis sentimen judul berita sebagai referensi perkembangan ilmu pengetahuan.
2. Mendukung peningkatan kualitas penelitian prodi dan universitas serta membuka peluang penelitian lanjutan terkait bidang NLP.