

BAB I

PENDAHULUAN

A. Latar Belakang

Ujian Nasional (UN) merupakan bagian penting dari sistem pendidikan di Indonesia yang bertujuan mengukur pencapaian belajar siswa. Setelah ditiadakan sejak 2021, pemerintah berencana memberlakukan kembali UN mulai tahun ajaran 2025/2026 untuk siswa SD, SMP, dan SMA. UN versi baru dirancang agar evaluasi capaian akademik peserta didik lebih objektif, sekaligus mengurangi beban mental dan finansial bagi siswa, tanpa dijadikan sebagai penentu kelulusan [1].

Rencana pemberlakuan kembali UN muncul dari aspirasi pemangku kebijakan pendidikan, termasuk kepala dinas pendidikan, organisasi profesi guru, perguruan tinggi, dan masyarakat. Salah satu pertimbangannya adalah kebutuhan menilai capaian akademik peserta didik secara individual, yang selama ini tidak sepenuhnya tercermin dari Asesmen Nasional (AN) berbasis sampling. Selain itu, standar prestasi melalui jalur prestasi perlu dinormalisasi agar evaluasi lebih adil [1].

Platform berbagai video seperti YouTube telah menjadi sarana penting bagi masyarakat untuk berdiskusi dan mengekspresikan pendapat terkait berbagai isu, termasuk kebijakan pendidikan. Masyarakat menyalurkan pendapatnya dengan meramaikan unggahan video melalui berbagai reaksi dan komentar. Opini yang disampaikan oleh pengguna YouTube dalam kolom komentar seringkali berjumlah sangat besar dan dapat dimanfaatkan untuk menilai suatu topik atau tujuan tertentu, yang dikenal dengan istilah analisis sentimen. Proses analisis sentimen pada data komentar biasanya meliputi tiga tahapan utama, yaitu klasifikasi, evaluasi, dan visualisasi hasil [2].

Seiring dengan perkembangan teknologi pemrosesan bahasa alami (*Natural Language Processing/NLP*), beberapa model seperti IndoBERT telah dikembangkan untuk memproses teks berbahasa Indonesia secara lebih efektif. IndoBERT merupakan model *deep learning* yang diadaptasi dari arsitektur BERT dan telah dilatih menggunakan korpus besar berbahasa Indonesia sehingga mampu memahami konteks dan makna kata dalam bahasa Indonesia dengan lebih baik. Model ini terbukti memiliki performa tinggi pada berbagai tugas NLP, termasuk analisis sentimen [3].

Selain model tunggal seperti IndoBERT, ada juga metode *hybrid* yang menggabungkan beberapa teknik atau model untuk meningkatkan akurasi analisis sentimen. Metode *hybrid* ini, yang sering kali menggabungkan teknik *machine learning* dan *deep learning*, bertujuan untuk meningkatkan kemampuan model dalam menangani kompleksitas bahasa yang lebih tinggi. Penelitian sebelumnya menunjukkan bahwa kombinasi model BERT dengan CNN atau LSTM dapat menghasilkan performa yang lebih baik dibandingkan model tunggal, di mana BERT-CNN mencapai akurasi tertinggi sebesar 0,9896 dibandingkan varian lain seperti BERT-LSTM-CNN dan BERT-LSTM [4].

Dengan meningkatnya jumlah komentar di YouTube, terdapat kebutuhan untuk menganalisis sentimen publik terkait isu seperti penerapan kembali UN. Penelitian ini bertujuan membandingkan kinerja IndoBERT dan metode *hybrid* dalam menganalisis komentar-komentar YouTube. Hasil analisis diharapkan memberikan wawasan mengenai respons masyarakat, mengidentifikasi kecenderungan opini publik, serta mengevaluasi efektivitas model IndoBERT dibandingkan metode *hybrid*. Dengan demikian, penelitian ini dapat berkontribusi pada pengembangan teknologi NLP sekaligus menjadi dasar bagi pemangku kebijakan dalam merancang strategi komunikasi dan kebijakan pendidikan yang sesuai dengan aspirasi masyarakat.

B. Perumusan Masalah

Berdasarkan uraian latar belakang yang telah dijelaskan di atas, adapun rumusan masalah dalam penelitian ini adalah: “Bagaimana perbandingan performa model IndoBERT dan *Hybrid* (IndoBERT + CNN) dalam analisis sentimen komentar YouTube tentang penerapan kembali Ujian Nasional menggunakan nilai *f1-score*?”

C. Batasan Penelitian

Untuk menjaga fokus penelitian dan menghindari cakupan yang terlalu luas, penelitian ini dibatasi pada beberapa aspek berikut :

1. Penelitian ini menggunakan sebanyak 4904 data komentar yang diambil dari video YouTube yang membahas pelaksanaan kembali Ujian Nasional.

2. Data yang digunakan akan dilabeli menggunakan *Inset lexicon*
3. Parameter pengklasifikasian dalam analisis sentimen ini terdiri dari kategori positif, netral, dan negatif.
4. Komentar video Youtube yang diambil berasal dari saluran berita terkenal di Indonesia seperti Kompas TV, Metro TV dan BeritaSatu
5. Kinerja masing-masing model akan dibandingkan menggunakan nilai *f1-score*. Perbandingan dilakukan dengan menggunakan *dataset* yang sama dan telah dilabeli secara manual untuk memastikan konsistensi dan keakuratan evaluasi.
6. Output yang dihasilkan meliputi:
 - a. Hasil klasifikasi sentimen: positif, negatif, dan netral.
 - b. Distribusi jumlah komentar berdasarkan kategori sentimen (jumlah komentar positif, netral, dan negatif).
 - c. Nilai akurasi model (*f1-Score*) yang didapatkan dari mengukur keseimbangan antara nilai *precision* dan *recall*.
 - d. Rangkuman setiap kelompok sentimen untuk memberikan gambaran informatif atas opini publik.
 - e. Topik atau Kata Kunci Populer per Sentimen untuk Mengungkap isu dominan di tiap kategori sentimen.

D. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah “Mengimplementasikan model IndoBERT dan *Hybrid* (IndoBERT + CNN) dalam analisis sentimen komentar YouTube tentang penerapan kembali Ujian Nasional untuk memperoleh model dengan performa terbaik berdasarkan perbandingan nilai *f1-score*.”

E. Manfaat Penelitian

Beberapa manfaat yang bisa didapatkan dari penelitian ini adalah sebagai berikut :

1. Untuk Universitas, mengetahui tolak ukur kemampuan mahasiswa dalam menguasai ilmu yang diperoleh selama kuliah.

2. Untuk Peneliti, hasil penelitian ini dapat berfungsi sebagai referensi dan dasar pengetahuan bagi peneliti lain di masa depan.
3. Untuk Pembaca, mendapatkan informasi mengenai sentimen masyarakat mengenai kebijakan pemerintah tentang penerapan kembali ujian nasional.
4. Untuk Pengguna Aplikasi, aplikasi ini memungkinkan analisis sentimen masyarakat mengenai ujian nasional dengan efisien, memberikan rangkuman sentimen, kata kunci, dan evaluasi akurasi secara akurat.

