

BAB I PENDAHULUAN

1.1 Latar Belakang Masalah

Analisis opini publik menjadi komponen penting untuk memahami reaksi masyarakat terhadap peristiwa sosial dan kebijakan publik, terutama sejak ketersediaan data digital dan kemajuan teknik komputasi memungkinkan pemantauan skala besar secara *real-time*. Media sosial seperti X (Twitter) menyediakan interaksi cepat, jangkauan luas, dan jejak digital terstruktur sehingga banyak dimanfaatkan sebagai sumber data untuk studi opini publik dalam ranah olahraga [1]. Platform ini merupakan kanal bagi penggemar, jurnalis, dan aktor olahraga untuk bertukar informasi dan menilai keputusan atau kebijakan taktis [1].

Dalam konteks olahraga nasional, khususnya sepak bola, penunjukan Patrick Kluivert sebagai pelatih Timnas Indonesia meningkatkan aktivitas diskusi publik di Twitter dan menghasilkan beragam reaksi dari publik pada periode awal [2]. Kumpulan pembicaraan yang muncul selama masa kepelatihan tersebut berpotensi menyimpan opini yang lebih matang dan relevan untuk dianalisis, bukan sekadar reaksi awal. Penerapan *Aspect Based Sentiment Analysis* (ABSA) pada ulasan penggemar Timnas Indonesia menunjukkan bahwa pemisahan opini menurut aspek dapat mengurai opini ke dalam dimensi yang lebih spesifik [3]. ABSA juga direkomendasikan oleh kajian lain pada kajian Timnas Indonesia untuk analisis opini lebih mendalam sehingga dapat memberikan wawasan yang lebih terperinci bagi pemangku kepentingan olahraga [4].

Berdasarkan temuan dan rekomendasi literatur tersebut, pemilihan metode analisis menjadi penting. Berbagai metode telah diterapkan untuk analisis sentimen dan deteksi aspek, dari model linier hingga arsitektur *transformer* modern. Model *transformer* pralatih (*pre-trained*) seperti IndoBERT memberikan performa lebih baik dalam klasifikasi sentimen dibandingkan beberapa model klasik [5], model ini juga menunjukkan performa yang baik dalam ABSA untuk teks *Twitter* ketika di *fine-tune* [6]. Di sisi lain, Logistic Regression cenderung lebih unggul dalam hal

klasifikasi sentimen teks Twitter dibanding *supervised machine learning* lainnya [7]. Studi lain yang menerapkan ABSA pada domain aplikasi melaporkan bahwa Logistic Regression tampil kompetitif pada banyak aspek dari beberapa *classifier* tunggal [8]. Pendekatan *ensemble* memungkinkan penggabungan probabilitas dari beberapa model untuk dianalisis efektivitasnya dalam meningkatkan performa klasifikasi dibandingkan model tunggal [8].

Namun, hingga kini masih terbatas bukti empiris yang secara eksplisit mengevaluasi penggabungan model linear klasik dan model *transformer* prelatih melalui mekanisme *ensemble weighted soft-voting* pada tugas deteksi aspek multi-label dan klasifikasi sentimen per aspek pada cuitan Twitter berbahasa Indonesia. Penggabungan model dengan karakteristik representasi yang berbeda belum tentu menghasilkan peningkatan kinerja secara konsisten, sehingga efektivitas pendekatan tersebut perlu diuji secara objektif melalui eksperimen terkontrol. Oleh karena itu, penelitian ini diarahkan untuk mengevaluasi secara empiris apakah kombinasi Logistic Regression dan IndoBERT memberikan kontribusi yang bermakna dibandingkan model tunggal, serta untuk mengidentifikasi kondisi di mana pendekatan *ensemble* tersebut efektif maupun tidak efektif.

1.2 Rumusan Masalah

Dengan dasar latar belakang yang telah dipaparkan, masalah penelitian ini dirumuskan sebagai berikut:

1. Bagaimana merancang dan mengimplementasikan *ensemble weighted soft-voting* antara Logistic Regression dan IndoBERT untuk tugas deteksi aspek multi-label dan klasifikasi sentimen per aspek pada data Twitter tentang Timnas Indonesia era Patrick Kluivert?
2. Apakah *ensemble weighted soft-voting* antara Logistic Regression dan IndoBERT meningkatkan performa deteksi aspek multi-label dan klasifikasi sentimen per aspek dibandingkan model tunggal?

1.3 Batasan Masalah

Untuk menegaskan ruang lingkup penelitian serta memastikan arah pembahasan terfokus pada tujuan yang akan dicapai, peneliti menetapkan batasan masalah sebagai berikut:

1. Data dibatasi pada cuitan teks dari platform Twitter berbahasa Indonesia yang membahas Timnas Indonesia.
2. Rentang waktu pengambilan data dibatasi pada beberapa pertandingan Timnas Indonesia periode kepelatihan Patrick Kluivert dengan pengambilan data satu hari sebelum dan dua hari setelah setiap pertandingan.
3. Pengambilan cuitan disaring menggunakan kata kunci topikal bertimestamp yang relevan dengan Timnas Indonesia.
4. Data berupa cuitan teks, termasuk cuitan asli, balasan, dan *retweet*, sedangkan konten multimedia tidak diproses.
5. Metode dibatasi pada Logistic Regression dan IndoBERT, metode lain tidak menjadi fokus penelitian.
6. Penerapan *ensemble* keluaran kedua model dilakukan melalui teknik *weighted soft-voting*, teknik *ensemble* lain tidak dieksplorasi.

1.4 Tujuan Penelitian

Penelitian ini bertujuan menerapkan *ensemble weighted soft-voting* untuk menggabungkan Logistic Regression dan IndoBERT pada korpus cuitan Twitter berbahasa Indonesia tentang Timnas Indonesia pada periode kepelatihan Patrick Kluivert serta mengukur apakah *ensemble* tersebut meningkatkan kinerja deteksi aspek multi-label dan klasifikasi sentimen per aspek.

1.5 Manfaat Penelitian

Hasil dari penelitian ini diharapkan memberikan gambaran mengenai kinerja penggabungan Logistic Regression dan IndoBERT menggunakan teknik *weighted soft-voting* pada tugas deteksi aspek multi-label dan klasifikasi sentimen per aspek. Hasil evaluasi yang diperoleh dapat menjadi referensi dalam

pengembangan dan pengujian pendekatan kombinasi model dengan karakteristik representasi berbeda pada data berbahasa Indonesia, khususnya dalam konteks analisis sentimen multi-aspek.

Selain itu, hasil analisis sentimen per aspek yang dihasilkan dari penelitian ini dapat memberikan ilustrasi mengenai pola opini publik terhadap Timnas Indonesia di media sosial serta dapat menjadi dasar pertimbangan bagi penelitian selanjutnya yang mengkaji penggabungan model dengan karakteristik representasi berbeda pada domain serupa.

1.6 Sistematika Penulisan

Dalam upaya memudahkan pembaca dalam mengikuti uraian dan pembahasan, struktur penelitian ini disusun sebagai berikut:

BAB I PENDAHULUAN, berisi latar belakang masalah yang mendasari penelitian, perumusan masalah dalam bentuk pertanyaan, batasan masalah yang membatasi cakupan studi, tujuan dilakukannya penelitian, manfaat dari hasil penelitian, serta panduan alur baca berupa sistematika penulisan.

BAB II TINJAUAN PUSTAKA, mencakup literatur dari studi terdahulu yang relevan, celah riset yang ditemukan, serta dasar-dasar teori yang mendukung penelitian.

BAB III METODE PENELITIAN, mengulas tinjauan umum objek dan sumber data, kriteria pemilihan data, alur penelitian dari pengumpulan data hingga evaluasi, serta rincian alat dan bahan.

BAB IV HASIL DAN PEMBAHASAN, menyajikan hasil implementasi metode, evaluasi efektivitas model, analisis perbandingan, serta pembahasan temuan utama.

BAB V PENUTUP, memaparkan kesimpulan dari hasil penelitian, kontribusi penelitian, serta rekomendasi pengembangan selanjutnya.