

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Revolusi digital telah membawa perubahan dalam dunia pendidikan, khususnya pada proses penilaian dan pemberian umpan balik. Perubahan tersebut menggeser pendekatan konvensional menuju proses yang lebih adaptif dan berpusat pada peserta didik [1]. Dalam konteks pembelajaran modern, umpan balik (*feedback*) menjadi elemen kunci yang memengaruhi perkembangan belajar siswa. Penelitian menunjukkan bahwa umpan balik yang diberikan secara cepat memiliki korelasi kuat dengan peningkatan motivasi serta pemahaman konsep, sedangkan umpan balik yang diberikan terlambat cenderung kehilangan efektivitas karena siswa telah beralih ke materi berikutnya [2]. Namun, penyediaan umpan balik yang ideal sering kali sulit diterapkan dalam praktik. Pengajar menghadapi beban kerja yang tinggi, khususnya dalam memberikan umpan balik tertulis yang menuntut penilaian mendalam. Kondisi ini memunculkan siklus negatif, yaitu tingginya beban kerja menyebabkan umpan balik menjadi tidak spesifik atau terlambat, siswa menganggap umpan balik tersebut tidak membantu atau tidak dimanfaatkan, dan pengajar semakin enggan atau mengalami kesulitan dalam menyediakan umpan balik formatif yang berkualitas pada masa mendatang [3]. Situasi ini menegaskan perlunya solusi yang mampu mengotomatisasi pemberian umpan balik secara *real-time* dengan tingkat akurasi yang dapat diandalkan, guna memutus lambatnya siklus evaluasi tanpa menambah beban pengajar.

Pembelajaran sejarah abad ke-21 tidak lagi berfokus pada hafalan fakta, melainkan pada pengembangan kemampuan berpikir kritis melalui analisis narasi, evaluasi sumber, dan interpretasi peristiwa. Namun, kondisi di lapangan menunjukkan adanya kesenjangan yang signifikan. Penelitian di SMA Negeri 7 Kota Serang mengungkap bahwa kemampuan berpikir kritis siswa pada materi Demokrasi Liberal berada pada kategori “kurang sekali” [4]. Banyak siswa hanya mengulang isi buku tanpa mampu merumuskan kesimpulan maupun pendapat kritis.

Kebutuhan tersebut mendorong munculnya solusi teknologi pendidikan berbasis *Natural Language Processing* (NLP), salah satunya *Automated Essay Scoring* (AES). AES mampu mempercepat proses penilaian dan menghasilkan skor yang lebih konsisten dibandingkan penilai manusia [5]. Sistem awal, seperti

Project Essay Grade (PEG) dan *e-rater*, menjadi tonggak awal pengembangan teknologi ini [6]. Namun, sistem AES klasik memiliki kelemahan mendasar, yaitu berperan sebagai penilai, bukan sebagai pendidik. Sistem ini umumnya hanya menghasilkan skor numerik yang bersifat sumatif dan tidak menyediakan informasi diagnostik yang membantu siswa memperbaiki kualitas tulisan [7]. Selain itu, banyak sistem AES bergantung pada fitur linguistik dangkal, seperti panjang teks [6]. Pendekatan ini menyebabkan AES rentan dimanipulasi, misalnya dengan menambah panjang esai, serta gagal mengevaluasi aspek berpikir tingkat tinggi, seperti kualitas argumentasi [8].

Berbeda dengan model tradisional, *Transformer* menggunakan mekanisme *self-attention* yang mampu menangkap dependensi kata jarak jauh secara bersamaan sehingga memungkinkan pemahaman konteks yang lebih kaya dan akurat [9]. Terobosan ini melahirkan berbagai model bahasa pra-terlatih yang membawa perubahan paradigma dalam analisis dan generasi teks [9].

Untuk mewujudkan sistem penilaian esai yang komprehensif, yaitu mampu mengevaluasi jawaban sekaligus membangkitkan umpan balik, penelitian ini mengintegrasikan dua arsitektur *Transformer* yang dikembangkan untuk Bahasa Indonesia. Pemahaman teks (*Natural Language Understanding/NLU*) dilakukan menggunakan IndoBERT, sedangkan pembangkitan teks (*Natural Language Generation/NLG*) dilakukan menggunakan IndoT5 (idT5). Sebagai model berbasis arsitektur *encoder*, IndoBERT unggul dalam menghasilkan representasi teks yang bermakna dan telah dilatih pada korpus besar Bahasa Indonesia [10]. Penelitian sebelumnya menunjukkan bahwa IndoBERT mencapai skor *Quadratic Weighted Kappa* (QWK) sebesar 0,9 pada penilaian esai berbahasa Indonesia, yang menunjukkan tingkat kesepakatan sangat tinggi dengan penilai manusia [11]. Namun, penelitian tersebut hanya berfokus pada prediksi skor holistik dan umpan balik (*feedback*) masih bersifat statis berdasarkan nilai. Keterbatasan utama dari pendekatan berbasis angka tunggal ini adalah hilangnya dimensi evaluasi kognitif lainnya. Skor numerik tidak dapat menjelaskan secara mandiri apakah nilai yang rendah disebabkan oleh ketidakakuratan fakta atau rendahnya tingkat penalaran siswa. Tanpa pemetaan profil diagnostik yang multidimensi, model generatif pada tahap selanjutnya tidak memiliki acuan (*constraint*) yang memadai untuk merangkai teks umpan balik yang adaptif dan tepat sasaran.

Untuk mengatasi keterbatasan tersebut, penelitian ini mengusulkan pengembangan arsitektur *multi-head* pada IndoBERT yang memungkinkannya menjalankan *multi-task learning* (MTL). Penggunaan satu *encoder* bahasa yang

membagikan representasi semantik ke beberapa lapisan *output* terbukti mampu meningkatkan kemampuan generalisasi model. Pendekatan ini memanfaatkan efek regularisasi yang mencegah *overfitting* pada satu tugas spesifik sehingga model dapat mempelajari representasi yang lebih universal [12]. Lebih lanjut, berbeda dengan pendekatan AES konvensional yang umumnya hanya memprediksi skor holistik tunggal, implementasi MTL pada domain AES terbukti efektif untuk memprediksi berbagai dimensi kualitas tulisan secara bersamaan melalui representasi bersama (*shared representation*) [13]. Oleh karena itu, metode ini dirancang untuk memproses satu masukan jawaban dan menghasilkan tiga luaran evaluasi secara bersamaan, yaitu prediksi skor numerik, klasifikasi label kebenaran untuk validasi fakta, serta identifikasi *Level of Reasoning* (LOR) guna mengukur kedalaman kognitif siswa. Pendekatan multidimensi ini dipilih agar sistem tidak hanya menghasilkan penilaian yang akurat secara kuantitatif, tetapi juga kaya secara diagnostik sehingga dapat mendukung penyusunan umpan balik yang lebih spesifik.

Sementara IndoBERT berfokus pada proses diagnostik, penyediaan umpan balik deskriptif memerlukan pendekatan generatif yang berbeda. Oleh karena itu, penelitian ini memanfaatkan idT5, yaitu varian arsitektur *Text-to-Text Transfer Transformer* yang dikembangkan khusus untuk Bahasa Indonesia [14]. Berbeda dengan IndoBERT, idT5 menggunakan struktur *encoder-decoder* yang memungkinkannya tidak hanya memahami konteks masukan, tetapi juga merangkai teks keluaran. Pemilihan arsitektur ini didasarkan pada studi sebelumnya yang menunjukkan bahwa model berbasis T5 efektif dalam menghasilkan *rationale*, yaitu penjelasan penilaian yang menyerupai umpan balik manusia [15]. Dengan demikian, integrasi idT5 diharapkan mampu mengisi kebutuhan fitur formatif yang belum dapat diakomodasi oleh model penilai.

Berdasarkan potensi tersebut, penelitian ini merumuskan sistem AES dua tahap yang mengintegrasikan keunggulan kedua model. Tahap pertama adalah tahap diagnosis, yaitu IndoBERT mengevaluasi kualitas esai secara akurat berdasarkan ketepatan konseptual dan penalaran. Tahap kedua adalah tahap remediasi, yaitu idT5 memanfaatkan hasil diagnosis tersebut untuk membangkitkan umpan balik korektif yang bersifat formatif, spesifik, dan membangun bagi siswa.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana merancang dan mengimplementasikan model *Natural Language Understanding* (NLU) berbasis IndoBERT dengan arsitektur *multi-task learning* untuk mengevaluasi esai sejarah secara otomatis?
2. Bagaimana mengintegrasikan hasil evaluasi NLU ke dalam model generatif *Natural Language Generation* (NLG) berbasis idT5 untuk membangkitkan teks umpan balik (*feedback*) yang adaptif serta relevan secara kontekstual terhadap kesalahan siswa?
3. Bagaimana tingkat akurasi dan kinerja model kecerdasan buatan yang dikembangkan dalam melakukan evaluasi esai serta membangkitkan umpan balik berdasarkan metrik pengujian standar pada tugas NLU dan NLG?

1.3 Batasan Permasalahan

Untuk menjaga agar penelitian ini tetap terfokus dan mendalam, ditetapkan batasan masalah sebagai berikut:

1. Penelitian ini berfokus pada esai mata pelajaran sejarah dengan materi Demokrasi Liberal untuk kelas XII sekolah menengah atas. Sistem tidak dikembangkan atau diuji untuk mata pelajaran lain.
2. Sistem ini dirancang untuk memproses dan menganalisis esai yang ditulis dalam Bahasa Indonesia.
3. Pengembangan sistem menggunakan arsitektur dua tahap yang terdiri atas IndoBERT untuk tahap diagnosis kesalahan konseptual dan idT5 untuk tahap pembangkitan umpan balik korektif.
4. Sistem diimplementasikan melalui proses *fine-tuning* menggunakan dataset esai sejarah berbahasa Indonesia yang disusun dengan bantuan OpenAI dan diverifikasi oleh ahli.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut:

1. Mengembangkan dan mengimplementasikan model *Natural Language Understanding* (NLU) berbasis IndoBERT dengan arsitektur *multi-task learning* untuk mengevaluasi esai sejarah secara otomatis.
2. Mengembangkan model *Natural Language Generation* (NLG) berbasis idT5 yang terintegrasi dengan hasil evaluasi NLU untuk menghasilkan umpan balik korektif yang adaptif.
3. Mengevaluasi kinerja sistem AES yang dibangun menggunakan metrik evaluasi NLU, seperti *Accuracy*, F1-Score, MAE, dan QWK, serta metrik evaluasi NLG, seperti BLEU dan ROUGE, guna memastikan kelayakan sistem.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat praktis bagi siswa dan guru, serta manfaat teoritis bagi pengembangan teknologi pendidikan.

1.5.1 Manfaat Praktis

1. Menyediakan umpan balik korektif yang cepat, spesifik, dan mudah dipahami untuk mendukung perbaikan mandiri siswa.
2. Meringankan beban kerja guru dalam koreksi esai serta menyediakan alat evaluasi yang lebih objektif dan konsisten, sehingga waktu dapat dialihkan untuk aktivitas pengajaran yang lebih strategis.

1.5.2 Manfaat Teoritis

1. Mendorong pergeseran paradigma AES dari alat penilaian sumatif menjadi instrumen pembelajaran formatif yang berorientasi pada pengembangan kemampuan berpikir kritis.
2. Memberikan kontribusi terhadap penelitian NLP berbahasa Indonesia melalui penerapan model *Transformer* dalam konteks pendidikan dan tugas generatif.

1.6 Sistematika Penulisan

Bagian ini berisi uraian singkat mengenai struktur isi laporan penelitian, dimulai dari pendahuluan hingga simpulan dan saran. Sistematika penulisan laporan adalah sebagai berikut:

- **Bab 1 PENDAHULUAN**
Bab ini berisi pendahuluan yang mencakup latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan laporan.
- **Bab 2 LANDASAN TEORI**
Bab ini membahas landasan teori dan tinjauan pustaka yang relevan dengan penelitian, meliputi konsep AES, NLP, IndoBERT, idT5, serta penelitian terdahulu yang menjadi dasar pengembangan sistem.
- **Bab 3 METODOLOGI PENELITIAN**
Bab ini menjelaskan metodologi penelitian yang digunakan, meliputi perancangan sistem, arsitektur model, pembangunan dan pelatihan model IndoBERT serta idT5, dan dataset yang digunakan.
- **Bab 4 HASIL DAN DISKUSI**
Bab ini menyajikan hasil implementasi sistem dan pembahasan, analisis performa penilaian dan umpan balik, serta evaluasi sistem AES yang dikembangkan.
- **Bab 5 KESIMPULAN DAN SARAN**
Bab ini berisi simpulan dari hasil penelitian yang telah dilakukan serta saran untuk pengembangan sistem dan penelitian selanjutnya.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A