

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi berdampak signifikan dalam bidang pendidikan, salah satunya melalui penelitian sistem *Automatic Question Generation* (AQG), yaitu sistem pembangkit soal otomatis yang menghasilkan soal secara efisien dan mengatasi pembuatan soal manual yang memakan waktu [1]. Pendekatan awal pengembangan AQG dilakukan dengan metode *rule-based* (berbasis aturan) di mana soal yang dihasilkan memiliki struktur yang tetap tetapi kurang fleksibel dan sulit diterapkan pada teks yang beragam [2]. Kemudian berkembang ke metode *neural sequence-to-sequence* (seq2seq) yang lebih fleksibel tetapi kualitas pertanyaannya sangat bergantung pada data latih sehingga kurang konsisten dari segi makna [3]. Keterbatasan tersebut diatasi dengan munculnya arsitektur *transformer* yang meningkatkan pemahaman konteks dan kualitas soal yang dihasilkan [4]. Meskipun perkembangannya signifikan, sistem AQG umumnya masih menghasilkan soal tanpa mempertimbangkan kemampuan individu. Kondisi ini berpotensi mengurangi efektivitas di mana soal dengan tingkat kesulitan tidak sesuai dengan kemampuan pengguna sehingga menimbulkan frustrasi atau rasa tidak tertantang [5,6]. Oleh karena itu, AQG adaptif menjadi salah satu pendekatan untuk menghasilkan soal dengan menyesuaikan kemampuan individu [7].

Beberapa peneliti terdahulu telah mengembangkan AQG adaptif, seperti K. Suresh dkk. mengembangkan sistem ujian adaptif berbasis *transformer* dengan menggunakan model *text-to-text transfer transformer* (T5) untuk pembangkitan soal dan *Bidirectional Encoder Representations from Transformers* (BERT) untuk evaluasi jawaban esai berbasis kemiripan makna, hasilnya unggul dibandingkan pendekatan *machine learning* tradisional dan metode *rule-based* [5]. Selanjutnya, penelitian oleh Yuto Tomikawa dkk. mengintegrasikan *Item Response Theory* (IRT) dengan BERT dan T5 untuk menghasilkan pasangan pertanyaan-jawaban dengan tingkat kesulitan secara adaptif, serta meningkatkan akurasi estimasi kemampuan dengan nilai *mean absolute error* (MAE) sebesar 0,707 dibandingkan metode acak sebesar 0,744 [6]. Integrasi model *transformer* dengan pendekatan estimasi kemampuan seperti IRT yang dikombinasikan dengan *Elo Rating System* (ERS) untuk memperbaharui kemampuan user dapat digunakan dalam pengembangan sistem evaluasi adaptif. Selain itu, penggunaan model T5 pada kedua penelitian tersebut menunjukkan bahwa model ini mampu menghasilkan pertanyaan secara

efektif dalam sistem AQG adaptif. Hal ini didukung oleh penelitian yang menunjukkan bahwa T5 menghasilkan pertanyaan dengan kualitas lebih tinggi dibandingkan BART dan GPT-2 dalam bidang pendidikan [8], serta memerlukan waktu 33,3% hingga 68% lebih singkat dibanding model BERT dan GPT-3 [7].

Penelitian terdahulu telah berhasil mengembangkan AQG yang adaptif secara *real-time* tetapi masih menggunakan format jawaban esai. Oleh karena itu, evaluasi memerlukan proses penilaian berdasarkan kemiripan makna yang lebih kompleks dan menambah beban komputasi sistem [5]. Untuk mengatasi keterbatasan tersebut, model T5 yang digunakan oleh Rodriguez-Torrealba dkk. dapat dimanfaatkan untuk membangkitkan soal pilihan ganda secara langsung sehingga proses evaluasi dapat dilakukan dengan mencocokkan jawaban pengguna dengan kunci jawaban [8]. Pendekatan ini menyederhanakan proses evaluasi tanpa memerlukan model tambahan untuk menilai kemiripan jawaban. Dengan demikian, hasil evaluasi dapat diperoleh secara langsung setelah pengguna menjawab setiap soal sehingga mendukung proses evaluasi adaptif secara *real-time* [9].

Dalam sistem adaptif, estimasi kemampuan pengguna memiliki peran penting dalam menentukan tingkat kesulitan. Beberapa pendekatan telah digunakan, seperti *Bayesian Knowledge Tracing* (BKT) yang memantau perkembangan penguasaan materi berdasarkan riwayat jawaban pengguna, tetapi pendekatan ini belum mempertimbangkan tingkat kesulitan setiap soal secara langsung [10]. Pendekatan lain adalah *Classical Test Theory* (CTT) yang menentukan tingkat kesulitan soal berdasarkan hasil jawaban peserta. Akan tetapi, hasil pengukurannya masih dipengaruhi oleh karakteristik peserta yang digunakan dalam pengujian [11]. Berbeda dengan kedua pendekatan tersebut, IRT mempertimbangkan kemampuan pengguna dan tingkat kesulitan soal secara bersamaan [12]. Dalam praktiknya, IRT umumnya memperbarui estimasi kemampuan menggunakan seluruh riwayat jawaban yang telah diberikan pengguna. Sementara itu, mekanisme ERS hanya menggunakan nilai kemampuan sebelumnya dan respons terbaru untuk memperbaharui estimasi kemampuan sehingga proses perhitungannya menjadi lebih sederhana dan cepat [13].

Berdasarkan uraian di atas, maka dibangun sistem AQG adaptif dengan menggunakan T5 dalam pembangkitan soal, serta mekanisme IRT yang dikombinasikan dengan ERS untuk memperbarui estimasi kemampuan secara iteratif. Oleh karena itu, penelitian ini dilakukan dengan judul “*Pengembangan Sistem Pembangkit Soal Adaptif dengan Model Text-to-Text Transfer Transformer (T5) dan Item Response Theory (IRT)*”.

1.2 Rumusan Masalah

Rumusan masalah dalam tugas akhir ini adalah belum tersedianya sistem *Automatic Question Generation* (AQG) adaptif yang mampu menghasilkan soal pilihan ganda secara otomatis dan menyesuaikan tingkat kesulitan soal berdasarkan kemampuan pengguna.

1.3 Tujuan

Tujuan tugas akhir ini adalah mengembangkan sistem *Automatic Question Generation* (AQG) adaptif yang mampu menghasilkan soal pilihan ganda secara otomatis menggunakan model *Text-to-Text Transfer Transformer* (T5) serta menyesuaikan tingkat kesulitan soal berdasarkan estimasi kemampuan pengguna menggunakan pendekatan *Item Response Theory* (IRT) dengan kombinasi pembaruan kemampuan berbasis *Elo Rating System* (ERS).

1.4 Manfaat

Manfaat yang diperoleh dari penulisan tugas akhir ini:

1. Menjadi referensi ilmiah mengenai AQG dengan integrasi model generatif berbasis *transformer* dengan IRT sebagai mekanisme perhitungan adaptif.
2. Menghasilkan prototipe aplikasi yang otomatis pada siklus latihan soal berbasis web, di mana pembangkitan soal, kalkulasi tingkat kemampuan, dan rekomendasi soal berikutnya berjalan secara sinkron dalam satu platform.

1.5 Ruang Lingkup

Ruang lingkup pengembangan sistem pembangkit soal adaptif dalam tugas akhir ini meliputi:

1. Sistem yang dikembangkan berupa aplikasi latihan soal pilihan ganda adaptif berbasis web.
2. Evaluasi jawaban dilakukan secara *biner* berdasarkan kecocokan antara jawaban pengguna dan kunci jawaban.
3. Estimasi kemampuan pengguna dihitung menggunakan IRT dan diperbarui secara iteratif dengan *Elo Rating System* (ERS) setelah setiap respons.
4. Pemilihan soal berikutnya dilakukan berdasarkan estimasi kemampuan pengguna untuk menyesuaikan tingkat kesulitan soal secara adaptif.

5. *Input* sistem berupa file PDF tunggal dengan kapasitas maksimum 15 MB yang berisi materi pembelajaran bahasa Inggris.
6. Proses *fine-tuning* model dilakukan menggunakan *library* Hugging Face menggunakan *dataset* RACE (*Reading Comprehension Dataset*) [14].
7. Topik *dataset* RACE merupakan kumpulan materi bacaan dan pertanyaan ujian bahasa Inggris.
8. Hasil hanya memiliki dua tingkatan kesulitan "*middle*" dan "*high*".
9. Arsitektur sistem terdiri dari *frontend* menggunakan *React*, *backend* menggunakan *Express.js* dan *Flask*, serta database menggunakan *PostgreSQL*.
10. Pengujian sistem dilakukan menggunakan metode *blackbox testing* untuk memastikan fungsi-fungsi sistem berjalan sesuai dengan kebutuhan.
11. Pengujian kualitas pembangkitan soal dilakukan menggunakan metrik ROUGE.
12. Pengujian mekanisme adaptif dilakukan melalui simulasi perubahan estimasi kemampuan dan tingkat kesulitan soal yang dipilih sistem.
13. Tugas Akhir tidak membandingkan performa sistem dengan model atau metode lain.
14. Pengujian menggunakan simulasi respon pengguna untuk mengamati perubahan estimasi kemampuan serta penyesuaian tingkat kesulitan soal yang dipilih sistem secara adaptif.

