

BAB I

PENDAHULUAN

Bab pendahuluan ini membahas mengenai latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan yang digunakan dalam skripsi berjudul Pembuatan Umpan Balik Otomatis Berbasis Klasifikasi Multikelas pada Jawaban Mahasiswa dengan Pemodelan T5 (*Text-to-Text Transfer Transformer*).

1.1 Latar Belakang

Pendidikan memainkan peran sentral dalam membangun individu dan masyarakat yang berkualitas. Salah satu aspek penting dalam sistem pendidikan adalah proses evaluasi, yang bertujuan untuk mengukur sejauh mana peserta didik memahami dan mampu mengaplikasikan materi yang telah dipelajari. Dalam konteks ini, umpan balik dari pengajar menjadi elemen krusial yang membantu peserta didik mengidentifikasi kekurangan serta memperkuat pemahaman mereka terhadap materi. Umpan balik yang diberikan secara tepat dan konstruktif tidak hanya meningkatkan kualitas pembelajaran, tetapi juga mendorong keterlibatan dan motivasi belajar peserta didik (Winstone et al., 2019). Sayangnya, dalam praktiknya, keterbatasan waktu dan sumber daya sering kali menjadi hambatan dalam pemberian umpan balik secara menyeluruh, terutama di kelas-kelas dengan jumlah peserta yang besar.

Untuk mengatasi permasalahan tersebut, berkembanglah solusi berbasis kecerdasan buatan (AI) yang menawarkan efisiensi dalam penilaian dan pemberian umpan balik. Salah satu pendekatan yang semakin luas diterapkan adalah *Automated Essay Scoring* (AES), yang memanfaatkan *Natural Language Processing* (NLP) dan pembelajaran mesin untuk mengevaluasi jawaban secara otomatis (Misgna et al., 2024). Namun, sebagian besar sistem AES saat ini masih fokus pada pemberian skor saja, tanpa menyertakan umpan balik naratif yang informatif bagi mahasiswa. Padahal, umpan balik semacam inilah yang justru dibutuhkan untuk mendukung proses pembelajaran yang reflektif dan berkelanjutan.

Agar sistem dapat memberikan umpan balik yang lebih bermakna, dibutuhkan pemodelan yang mampu memahami isi jawaban mahasiswa secara lebih kontekstual, bukan sekadar mencocokkan kata kunci. Model berbasis Transformer seperti T5 (*Text-to-Text Transfer Transformer*) telah terbukti efektif dalam berbagai tugas NLP karena kemampuannya memahami konteks dan menghasilkan teks yang natural (Raffel et al., 2019). T5 menggunakan

pendekatan *transfer learning* yang fleksibel sehingga dapat diterapkan dalam tugas-tugas seperti klasifikasi teks maupun generasi umpan balik. Dalam penelitian sebelumnya, T5 juga menunjukkan potensi dalam menyampaikan umpan balik yang lebih akurat, termasuk dalam konteks pembelajaran bahasa (Behzad et al., 2022).

Dalam sistem berbasis T5 ini, tugas klasifikasi menjadi penting untuk menentukan kategori dari jawaban mahasiswa berdasarkan tingkat kesesuaian dengan jawaban referensi. Kategori seperti *correct*, *partial 2*, *partial 1*, dan *incorrect* digunakan untuk menyampaikan umpan balik yang lebih spesifik dan sesuai dengan kebutuhan pembelajaran mahasiswa. Oleh karena itu, pendekatan klasifikasi multikelas menjadi inti dari sistem ini, karena memungkinkan model untuk menyesuaikan bentuk umpan balik berdasarkan label klasifikasi yang diberikan.

Untuk mengukur kemiripan antara jawaban mahasiswa dan referensi, sistem ini mengintegrasikan berbagai metode pengukuran *similarity*, yaitu BLEU Score, Fuzzy *Similarity*, dan BERT *Similarity*. Masing-masing metode menangkap aspek yang berbeda, seperti BLEU untuk kemiripan n-gram, Fuzzy untuk *edit distance*, dan BERT untuk kesamaan semantik berbasis *embedding*. Dengan menggabungkan ketiganya menggunakan pendekatan *dynamic weight adjustment*, sistem diharapkan dapat mengenali makna jawaban mahasiswa yang menggunakan struktur kalimat yang berbeda tetapi bermakna sama (Kulshreshtha et al., 2022).

Evaluasi performa klasifikasi multikelas dilakukan menggunakan Macro F1-Score, karena metrik ini memberikan penilaian yang seimbang terhadap performa pada seluruh kelas, termasuk kelas minoritas. Macro F1 dinilai lebih tepat dibandingkan akurasi biasa dalam skenario klasifikasi multikelas dengan distribusi yang tidak merata (Grandini et al., 2020). Selain itu, untuk mengevaluasi kedekatan skor similarity yang diprediksi sistem terhadap skor penilai manusia, digunakan SMAPE (*Symmetric Mean Absolute Percentage Error*), yang secara efektif mengukur deviasi relatif dan dinilai lebih stabil dalam menangani skala nilai yang bervariasi (Bisht & Kumar, 2020).

Tidak hanya dari sisi metrik kuantitatif berbasis nilai prediksi, penelitian ini juga mempertimbangkan kecocokan antara keluaran sistem dan *ground truth* sebagai indikator kualitas sistem umpan balik otomatis. Oleh karena itu, digunakan pendekatan evaluasi berbasis *exact match accuracy*, yang mengukur persentase kecocokan umpan balik sistem dengan

umpan balik *ground truth* yang telah ditentukan. Metrik ini sangat penting karena setiap label klasifikasi memiliki bentuk umpan balik tetap yang harus dihasilkan secara presisi.

Dalam penelitian Chen et al. (2024) menyatakan bahwa evaluasi berbasis *exact match* telah menjadi standar dalam berbagai tugas *Natural Language Processing* yang menuntut ketepatan literal, termasuk dalam evaluasi sistem *code generation*, *machine translation*, dan *question answering*. Dalam penelitian tersebut, *exact match* digunakan untuk menilai kemampuan model dalam menghasilkan output yang sepenuhnya sesuai dengan referensi (*ground truth*), tanpa toleransi terhadap variasi redaksional. Pendekatan ini memungkinkan penilaian objektif terhadap sistem, terutama ketika keberhasilan model ditentukan oleh kemampuannya menghasilkan format atau isi yang telah distandarkan.

Dengan demikian, penelitian ini bertujuan untuk mengembangkan sistem umpan balik otomatis berbasis klasifikasi multikelas terhadap jawaban mahasiswa menggunakan pemodelan T5 dan pendekatan gabungan similarity. Sistem ini diharapkan dapat membantu dosen dalam memberikan umpan balik yang cepat, akurat, dan adaptif, serta membantu mahasiswa memahami kekurangan mereka secara lebih konstruktif untuk mendukung proses belajar mandiri yang lebih efektif.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, rumusan masalah yang diangkat dalam penelitian ini adalah bagaimana pembuatan umpan balik otomatis berbasis klasifikasi multikelas pada jawaban mahasiswa dengan pemodelan T5 (*Text-to-Text Transfer Transformer*). Adapun masalah khusus yang akan dibahas dalam penelitian ini meliputi.

1. Bagaimana kinerja metode *combined similarity* dengan *dynamic weight adjustment* dalam menghasilkan prediksi skor jawaban mahasiswa berdasarkan evaluasi menggunakan metrik SMAPE?
2. Bagaimana kinerja model dalam melakukan klasifikasi multikelas terhadap skor jawaban mahasiswa berdasarkan evaluasi menggunakan pendekatan Macro F1-Score?
3. Bagaimana pengaruh *hyperparameter tuning* (*learning rate* dan *batch size*) terhadap performa model T5 dalam menghasilkan skor jawaban mahasiswa berbasis klasifikasi multikelas?

4. Bagaimana hasil evaluasi sistem umpan balik otomatis berdasarkan evaluasi menggunakan metrik *exact match accuracy* untuk menilai kualitas umpan balik yang dihasilkan oleh model terhadap jawaban mahasiswa?

1.3 Tujuan dan Manfaat

Tujuan umum dari penelitian ini adalah membuat umpan balik otomatis berbasis klasifikasi multikelas pada jawaban mahasiswa dengan pemodelan T5 (*Text-to-Text Transfer Transformer*). Sistem ini dirancang untuk meningkatkan akurasi dan efisiensi dalam mengevaluasi jawaban mahasiswa secara otomatis. Adapun tujuan khusus dari penelitian ini adalah sebagai berikut.

1. Menganalisis kinerja metode *combined similarity* dengan *dynamic weight adjustment* dalam menghasilkan prediksi skor jawaban mahasiswa berdasarkan evaluasi menggunakan metrik SMAPE.
2. Mengevaluasi kinerja klasifikasi multikelas dari skor jawaban mahasiswa yang telah diprediksi, menggunakan pendekatan Macro F1-Score.
3. Mengetahui pengaruh *hyperparameter tuning* (*learning rate* dan *batch size*) terhadap performa model T5 dalam menghasilkan klasifikasi skor jawaban mahasiswa.
4. Mengevaluasi kualitas umpan balik yang dihasilkan oleh sistem evaluasi menggunakan metrik *exact match accuracy*.

Manfaat penelitian ini diharapkan dapat meningkatkan efisiensi evaluasi akademik berbasis kecerdasan buatan. Sistem umpan balik otomatis yang dikembangkan dapat membantu tenaga pendidik memberikan umpan balik lebih cepat, akurat, dan objektif, serta mengurangi beban penilaian manual. Selain itu, sistem ini memungkinkan mahasiswa menerima umpan balik yang lebih adaptif sesuai tingkat pemahaman mereka, mendukung pembelajaran mandiri.

1.4 Ruang Lingkup

Ruang lingkup dalam penelitian pembuatan umpan balik otomatis berbasis klasifikasi multikelas pada jawaban mahasiswa dengan pemodelan T5 (*Text-to-Text Transfer Transformer*) ini dijabarkan sebagai berikut.

1. Penelitian ini menerapkan pendekatan *dynamic weight adjustment* untuk memboboti tiga metode *similarity*, yaitu BLEU Score, Fuzzy Similarity, dan BERT Similarity.

2. Keluaran dari penelitian ini berupa klasifikasi jawaban mahasiswa ke dalam kategori *correct*, *partial 2*, *partial 1*, *incorrect*.
3. *Tools* yang digunakan dalam penelitian ini adalah Google Colaboratory dengan bahasa pemrograman Python.

1.5 Sistematika Penulisan

Sistematika penulisan digunakan untuk memberikan gambaran yang runtut dan jelas mengenai penyusunan laporan skripsi yang berjudul Pembuatan Umpan Balik Otomatis Berbasis Klasifikasi Multikelas Pada Jawaban Mahasiswa Dengan Pemodelan T5 (*Text-to-Text Transfer Transformer*). Sistematika penulisan laporan akan dijelaskan sebagai berikut.

BAB I PENDAHULUAN

Bab ini membahas mengenai beberapa aspek penting, termasuk latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup, dan sistematika penulisan laporan yang dibuat.

BAB II LANDASAN TEORI

Bab ini memuat tinjauan pustaka dan teori-teori yang digunakan dalam penelitian, meliputi *state of the art* sistem umpan balik otomatis, konsep klasifikasi multikelas, serta arsitektur model T5 yang digunakan sebagai dasar pemodelan. Selain itu, dibahas pula *sentence embedding*, prapemrosesan data, dan augmentasi data dengan BART. Penelitian ini juga mengintegrasikan tiga metode pengukuran *similarity*, yaitu BLEU, Fuzzy, dan BERT, yang digabungkan melalui pendekatan *combined similarity* dengan *dynamic weight adjustment*. Untuk mengoptimalkan performa model, digunakan strategi *hyperparameter tuning* berbasis Optuna, sedangkan evaluasi sistem dilakukan menggunakan metrik SMAPE, Macro F1-Score, dan *exact match accuracy*.

BAB III METODOLOGI PENELITIAN

Bab ini menyajikan alur serta tahapan-tahapan yang digunakan sebagai acuan dalam melaksanakan penelitian. Adapun hal-hal yang dibahas meliputi gambaran umum penelitian, pengumpulan data, augmentasi data, prapemrosesan data, *sentence embedding*, perhitungan *similarity*, penggabungan nilai *similarity* dengan pendekatan *dynamic weight adjustment*, evaluasi kinerja metode perhitungan, klasifikasi jawaban mahasiswa,

pembentukan dataset untuk model T5, pencarian *hyperparameter* terbaik menggunakan Optuna, pelatihan model akhir (*final training*), evaluasi model menggunakan SMAPE, Macro F1-Score, serta *exact match accuracy*.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil dan analisis dari setiap tahapan yang telah dilakukan dalam penelitian. Pembahasan dimulai dari pemaparan hasil pelaksanaan setiap tahapan penelitian yang mencakup pengumpulan data, augmentasi data, prapemrosesan data, *sentence embedding*, perhitungan *similarity*, evaluasi kinerja metode perhitungan, klasifikasi jawaban, pembentukan dataset untuk model T5, *hyperparameter tuning*, pelatihan akhir model, hingga *generate feedback*. Selanjutnya, dilakukan analisis terhadap eksperimen yang dilakukan, meliputi evaluasi kualitas prediksi *similarity* menggunakan SMAPE, evaluasi klasifikasi menggunakan Macro F1-Score, serta penilaian kualitas umpan balik dengan *exact match accuracy*.

BAB V KESIMPULAN DAN SARAN

Bab ini membahas mengenai kesimpulan dari penelitian berdasarkan hasil yang telah diperoleh serta saran untuk penelitian selanjutnya agar dapat dikembangkan lebih lanjut.