

Text Generation untuk Profil Mata Kuliah pada Penilaian Outcome-Based Education Menggunakan Text-to-Text Transfer Transformers

Nurrohman^{1*}, Dian Sa'adillah Maylawati², Muhammad Insan Al-Amin³
 Teknik Informatika, UIN Sunan Gunung Djati Bandung, Bandung, Indonesia
 *e-mail Corresponding Author: 1207050095@student.uinsgd.ac.id

Abstract

The evaluation of Course Learning Outcomes (CPMK) in Outcome-Based Education (OBE) is still conducted manually, making it time-consuming and prone to errors. Additionally, the achievement profile of CPMK is often overlooked. This study aims to automate the generation of course profiles based on CPMK using Text Generation technology. The method employed is Transformers with the T5 (Text-to-Text Transfer Transformer) algorithm. Experiments were conducted using three variants of the T5 model: T5 Base, T5 Base with fine-tuning, and T5 XL, evaluated using the Bilingual Evaluation Understudy (BLEU) and Recall-Oriented Understudy for Gisting Evaluation (ROUGE) metrics. The results show that T5 XL achieved the best performance, with an average BLEU score of 0.592 and a ROUGE-L score of 0.721. T5 Base with fine-tuning recorded a BLEU score of 0.417 and a ROUGE-L score of 0.468, while T5 Base without fine-tuning had a BLEU score of 0.327 and a ROUGE-L score of 0.246. Additionally, more structured prompts yielded better evaluation results. This study demonstrates that T5 XL enhances the efficiency and accuracy of CPMK evaluation in OBE.

Keywords: Outcome Based Education; Text Generation; Text-To-Text Transfer Transformers; Penilaian

Abstrak

Evaluasi capaian pembelajaran mata kuliah (CPMK) dalam *Outcome-Based Education* (OBE) masih dilakukan secara manual, memakan waktu, dan rentan terhadap kesalahan. Selain itu, profil pencapaian CPMK sering diabaikan. Penelitian ini bertujuan mengotomasi pembuatan profil mata kuliah berbasis CPMK menggunakan teknologi *Text Generation*. Metode yang digunakan adalah *Transformers* dengan algoritma T5 (*Text-to-Text Transfer Transformers*). Eksperimen dilakukan dengan tiga varian model T5: T5 Base, T5 Base dengan *fine-tuning*, dan T5 XL, dievaluasi menggunakan metrik *Bilingual Evaluation Understudy* (BLEU) dan *Recall-Oriented Understudy for Gisting Evaluation* (ROUGE). Hasil menunjukkan T5 XL memiliki performa terbaik dengan BLEU rata-rata 0,592 dan ROUGE-L 0,721. T5 Base dengan *fine-tuning* mencatat BLEU 0,417 dan ROUGE-L 0,468, sedangkan T5 Base tanpa *fine-tuning* memiliki BLEU 0,327 dan ROUGE-L 0,246. Selain itu, prompt yang lebih terstruktur menghasilkan evaluasi lebih baik. Penelitian ini membuktikan bahwa T5 XL meningkatkan efisiensi dan akurasi evaluasi CPMK dalam OBE.

Kata kunci: Outcome Based Education; Text Generation; Text-To-Text Transfer Transformers; Penilaian

1. Pendahuluan

Salah satu pendekatan dalam pengajaran di abad 21 ini adalah *Outcome-based Education* atau disingkat dengan OBE. OBE dirancang untuk membantu peserta belajar untuk mencapai hasil yang telah ditentukan (*outcome targeted*) [1]. Pada sistem pembelajaran ini, capaian pembelajaran diidentifikasi terlebih dahulu kemudian merencanakan metode pembelajaran yang disesuaikan dengan hasil yang akan dicapai atau luaran [2]. Sistem pembelajaran bisa dilakukan dengan *blended learning*. Terdapat juga faktor-faktor pengaruh keberhasilan yaitu metode pembelajaran yang terdiri dari menghidupkan ruang-ruang diskusi mahasiswa diluar kelas, pembelajaran modern dan multidisiplin ilmu pengetahuan, faktor selanjutnya juga terdapat kebebasan akademik serta peran dosen dan mahasiswa yang sangat

menunjang keberhasilan implementasi kurikulum OBE [3]. Strategi dalam mengimplementasikan OBE menjadi sangat penting khususnya dalam mendesain metode pengukuran berdasarkan rubrik penilaian yang seragam sehingga dapat terdokumentasikan dengan baik. Capaian pembelajaran lulusan yang menjadi fokus utama kemudian membutuhkan evaluasi sehingga memerlukan pengukuran sesuai dengan kurikulum yang berjalan [4].

Capaian pembelajaran adalah suatu pernyataan yang menunjukkan tentang apa yang diinginkan, diketahui, dimengerti dan dapat dilaksanakan oleh peserta didik setelah menyelesaikan tahap belajar. Capaian pembelajaran adalah kompetensi yang diperoleh melalui proses penghayatan pengetahuan, sikap, keterampilan, kompetensi dan himpunan pengalaman belajar [5]. Faktanya, proses bisnis saat ini penilaian CPL dilakukan secara manual dengan dosen menginputkan data nilai pada *Microsoft excel*, kemudian dosen mengolah nilai secara manual menggunakan rumus *Excel*, dosen membuat laporan dari hasil pengolahan nilai. Kemudian data akan tersimpan pada *Microsoft Excel* dan terakhir dosen memberikan laporan kepada pihak departemen [6]. Selain pengolahan nilai yang masih menggunakan teknologi konvensional, evaluasi profil maha kuliah yang diselesaikan oleh mahasiswa pun belum dilakukan secara otomatis. Bahkan seringkali terlupakan. Perkembangan teknologi *Natural Language Processing* (NLP) dalam generating text dapat dimanfaatkan untuk menghasilkan profil mata kuliah secara efisien berbasis AI. *Text Generation* merupakan salah satu sub-bidang yang paling penting dalam ranah pemrosesan bahasa alami (NLP). *Text Generation* merupakan sistem yang dapat membangkitkan *text* berupa kalimat secara otomatis dari teks atau dokumen dengan menggunakan metode atau beberapa pola tertentu [7].

Penelitian sebelumnya memanfaatkan teknologi *text generation* ini pada berbagai studi kasus seperti: Alat bantu baca tunanetra berbasis teknologi *text to speech* dan *optical character recognition* [8]. Dimana pada penelitian yang ditulis oleh Li Gong, Joseph Crego dan Jean Senellart menyebutkan model neural baru-baru ini menunjukkan kemajuan yang signifikan dalam tugas-tugas pembuatan data-ke-teks di mana teks deskriptif dihasilkan berdasarkan catatan basis data. Dalam karya ini, kami menyajikan model pembangkitan data-ke-teks berbasis Transformers baru yang mempelajari pemilihan konten dan pembuatan ringkasan secara *end-to-end* [9]. Representasi dari *Transformers* (BERT) telah diusulkan dalam beberapa tahun terakhir untuk mengembangkan beberapa model canggih untuk berbagai tugas NLP, seperti menjawab pertanyaan (SQuAD v1.1) dan inferensi bahasa alami [10]. Optimasi diperlukan agar model dapat memberikan peningkatan dalam memprediksi suatu data, selain itu bertujuan melakukan peningkatan nilai akurasi pada model BART serta meningkatkan ketepatan dan kecepatan dalam melakukan klasifikasi [11]. Model-model inovatif ini, yang diwakili oleh ChatGPT, memanfaatkan algoritma Transformers yang mampu untuk memahami dan menghasilkan teks seperti bahasa manusia [12]. Dalam penelitian ini, dikembangkan model *Text Generation* menggunakan *Transformers* untuk mengotomasi evaluasi capaian pembelajaran mata kuliah setiap mahasiswa. *Transformers* adalah pendekatan populer saat ini karena strukturnya yang kompleks dan mekanisme perhatiannya yang memungkinkan pemahaman mendalam terhadap bahasa lisan dan tulisan. Contoh model *Transformers* yang dapat digunakan untuk teks berbahasa Indonesia adalah BERT dan T5 [13]. *Text-to-Text Transfer Transformers* (T5) adalah sebuah kerangka kerja yang dikembangkan di atas arsitektur populer seperti BERT dan GPT dengan memanfaatkan pembelajaran transfer berbasis teks-ke-teks. Model T5 merupakan model *Transformers* yang dilatih secara *end-to-end* dengan teks sebagai input dan teks yang telah dimodifikasi sebagai *output*. Model ini terbukti berhasil dan mampu melakukan tiga tugas utama. Yaitu, generasi teks, klasifikasi, dan regresi. Penelitian ini menunjukkan bahwa penggunaan teknologi kecerdasan buatan, khususnya T5, memiliki potensi besar dalam mendukung otomatisasi evaluasi capaian pembelajaran mahasiswa secara lebih efisien dan sistematis [14].

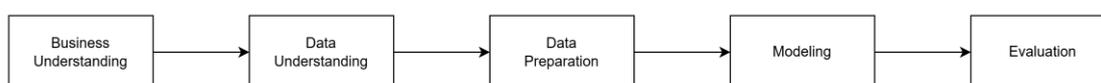
Text generation digunakan untuk mengotomasi evaluasi capaian pembelajaran mata kuliah dengan menggunakan *Transformers*, tepatnya menggunakan *Text-To-Text Transfer Transformers* (T5). Dimana model ini cocok digunakan dalam penelitian ini terlihat dari jumlah parameter yang lebih besar dan arsitektur kompleks yang memungkinkan T5 menangkap konteks dengan lebih baik dibandingkan model lainnya [15]. Selain itu, struktur *prompt* juga memainkan peran penting dalam hasil evaluasi. Hal ini mengindikasikan bahwa pola atau struktur tertentu lebih mudah dipahami oleh model [16]. *Prompt* yang lebih terstruktur

memberikan konteks yang jelas, sehingga model dapat menghasilkan teks yang lebih relevan [17].

Penelitian ini bertujuan untuk mendeskripsikan profil kemampuan mahasiswa dalam menyelesaikan masalah pada mata kuliah. Dengan memanfaatkan teknologi kecerdasan buatan, penelitian ini mengeksplorasi penggunaan *Text Generation* berbasis *Transformers* untuk mengotomatisasi proses tersebut. Hasil penelitian ini diharapkan dapat berkontribusi pada pengembangan asesmen OBE serta evaluasi capaian kompetensi mahasiswa.

2. Metodologi

Metode *Cross-Industry Standard Process for Data Mining* (CRISP-DM) diadaptasi untuk sebagai aktivitas penelitian. Metode CRISP-DM cocok digunakan dalam penelitian ini karena memiliki kerangka kerja yang sistematis dan terstruktur untuk penelitian ini, mengingat sifatnya yang berbasis pada analisis data dan pengembangan model berbasis kecerdasan buatan [18]. CRISP-DM memiliki proses yang saling berhubungan dan bersifat iteratif, dimulai dengan pemahaman bisnis hingga evaluasi pada model yang dihasilkan. Aktivitas CRISP-DM dimulai dari *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, dan *Evaluation*. Sedangkan aktivitas *Deployment* tidak dilakukan pada penelitian ini. Gambar 1 menunjukkan aktivitas penelitian yang mengadaptasi CRISP-DM.



Gambar 1. Adaptasi Metode CRISP-DM

2.1. Business Understanding

Penelitian ini bertujuan untuk mengotomasi pembuatan profil kemampuan mahasiswa dalam menyelesaikan masalah pada mata kuliah berbasis *Outcome Based Education* (OBE) dengan memanfaatkan teknologi kecerdasan buatan. Melalui *Text Generation* menggunakan *Transformers*, penelitian ini diharapkan dapat menghasilkan deskripsi yang lebih informatif dan koheren mengenai capaian kompetensi mahasiswa serta diharapkan dapat memberi kontribusi dalam peningkatan asesmen OBE dan evaluasi capaian mahasiswa secara efisien dan akurat.

Table 1. Contoh Rubrik Penilaian

SKOR	RUBRIK PENILAIAN					
	CPMK0101	CPMK0102	CPMK0103	CPMK0104	CPMK0105	CPMK0106
0 - 20	<p>Tidak Kompeten: Mahasiswa tidak menunjukkan pemahaman yang memadai terhadap materi yang telah diajarkan. Mereka tidak mampu menjelaskan konsep-konsep dengan jelas dan tidak dapat memberikan contoh yang tepat untuk mendukung penjelasan tersebut. Ketidakmampuan ini juga tercermin dalam ketidakmampuan mereka untuk mengemukakan ide atau argumen yang masuk akal, yang mengindikasikan kurangnya pengetahuan yang mendasar. Tidak ada pemahaman atau pengetahuan yang dapat diidentifikasi dari penjelasan yang diberikan, sehingga menciptakan kesan bahwa mahasiswa belum menguasai materi dan memerlukan perhatian lebih lanjut dalam proses pembelajaran.</p>					
21 - 40	<p>Kurang Kompeten: Mahasiswa menunjukkan kesulitan dalam memahami materi yang telah diajarkan. Penjelasan yang diberikan sering kali tidak jelas atau bahkan salah, menciptakan kebingungan dalam menyampaikan informasi. Contoh yang disertakan kurang relevan atau tidak tepat, yang semakin memperburuk pemahaman yang seharusnya dicapai. Selain itu, gagasan atau argumen yang dikemukakan tidak logis atau terfragmentasi, menunjukkan ketidakmampuan untuk menyusun pemikiran secara sistematis. Dengan demikian, pemahaman mahasiswa terhadap materi ini tampak sangat terbatas dan sering kali salah, mengindikasikan perlunya pendekatan yang lebih mendalam dalam proses pembelajaran untuk membantu mereka mencapai pemahaman yang lebih baik.</p>					
41 - 60	<p>Cukup Kompeten: Mahasiswa memiliki pemahaman dasar terhadap materi yang diajarkan, tetapi penjelasan yang diberikan menunjukkan kejelasan yang terbatas. Mereka sering kali memberikan contoh yang kurang relevan, yang tidak sepenuhnya mendukung konsep yang dijelaskan. Selain itu, ide atau argumen yang dikemukakan cenderung kurang terstruktur, membuatnya sulit untuk diikuti dan dipahami dengan baik. Meskipun pemahaman mereka mungkin benar dalam beberapa aspek, namun tetap saja tidak lengkap, mengindikasikan perlunya pengembangan lebih lanjut agar dapat menyampaikan informasi dengan lebih efektif dan terorganisir.</p>					

SKOR	RUBRIK PENILAIAN					
	CPMK0101	CPMK0102	CPMK0103	CPMK0104	CPMK0105	CPMK0106
61 - 80	Kompeten: Mahasiswa menunjukkan pemahaman yang baik terhadap materi yang diajarkan. Mereka dapat menjelaskan konsep dengan cukup jelas, memberikan contoh yang relevan dan mendukung penjelasan tersebut. Ide atau argumen yang dikemukakan disusun dengan struktur yang masuk akal, sehingga mudah diikuti dan dipahami. Meskipun terdapat beberapa kesalahan minor, pemahaman mahasiswa secara umum adalah akurat dan mencerminkan kemampuan mereka untuk mengaitkan informasi dengan cara yang efektif. Hal ini menunjukkan bahwa mereka memiliki landasan yang kuat untuk terus mengembangkan pemahaman yang lebih mendalam terhadap materi yang dipelajari.					
81 – 100	Sangat Kompeten: Mahasiswa menunjukkan pemahaman mendalam tentang materi yang diajarkan. Mereka dapat menjelaskan konsep dengan jelas dan tepat, sehingga informasi yang disampaikan mudah dipahami. Contoh yang diberikan sangat relevan dan efektif dalam mendukung penjelasan, sementara ide atau argumen yang dikemukakan disusun dengan logis dan kohesif. Pemahaman yang ditunjukkan tidak hanya akurat, tetapi juga bersifat kritis dan reflektif, menunjukkan kemampuan mahasiswa untuk menganalisis dan mengevaluasi informasi secara mendalam. Hal ini mencerminkan bahwa mahasiswa tidak hanya menguasai materi, tetapi juga mampu mengaitkannya dengan konteks yang lebih luas dan relevan.					

2.2. Data Understanding

Data yang digunakan dalam penelitian ini mengacu pada rubrik penilaian yang disusun berdasarkan kurikulum *Outcome Based Education* (OBE). Rubrik ini mencakup berbagai komponen untuk mendeskripsikan capaian kompetensi mahasiswa, yaitu CPMK (Capaian Pembelajaran Mata Kuliah), nilai yang di peroleh mahasiswa, dan narasi deskriptif yang menjelaskan pencapaian kompetensi. CPMK menggambarkan kompetensi yang harus di dicapai oleh mahasiswa dalam setiap mata kuliah sesuai dengan standar OBE. Nilai yang tercatat mencerminkan Tingkat pencapaian mahasiswa terhadap kompetensi yang diharapkan, sementara narasi deskriptif memberikan penjelasan mengenai proses dan hasil belajar mahasiswa. Data ini menjadi dasar untuk menghasilkan profil kemampuan mahasiswa yang lebih objektif dan terstruktur. Contoh rubrik penilaian CPMK terdapat pada Gambar 1.

2.3. Data Preparation

Dataset yang digunakan terdiri dari 155 data yang berisi CPMK, Skor, dan Narasi sebagai referensi keluaran. Dataset dibagi menjadi dua jenis, yaitu data *train* dan data *test* dengan metode pembagian yang tersaji pada Tabel 2.

Tabel 2. Data Train dan Data Test

Kategori	Jumlah	Persentase
<i>Train</i>	108	70.00%
<i>Test</i>	47	30.00%

Data kemudian di tokenisasi menggunakan T5 tokenizer dengan panjang maksimum 512 token, kemudian dilakukan padding dan truncation agar konsisten. Tokenisasi menghasilkan *input IDs*, *Attention Mask*, dan *Labels* dalam format numerik. Dataset diformat menjadi tipe torch.Tensor agar dapat digunakan untuk fine-tuning model T5.

2.4. Modeling

Setelah tahapan pengolahan awal data selesai, langkah selanjutnya adalah mengembangkan model menggunakan T5 (*Text-to-Text Transfer Transformers*). T5, yang dikembangkan oleh Google, merupakan model berbasis arsitektur *Transformers* yang dirancang untuk memecahkan berbagai masalah pemrosesan bahasa alami (NLP) dengan pendekatan serbaguna, yaitu merubah semua tugas menjadi tugas *text-to-text*. Salah satu fitur utama dari arsitektur ini adalah penggunaan *encoder-decoder*, yang memungkinkan model untuk memproses dan menghasilkan teks dengan mengandalkan mekanisme *self-attention* untuk mempelajari hubungan antara kata-kata dalam suatu kalimat.

Pada penelitian ini, T5 diterapkan dengan dua varian model, T5 Base dan T5 XL, yang masing-masing memiliki jumlah parameter dan kapasitas yang berbeda. T5 Base dan T5 XL memiliki perbedaan utama dalam ukuran model, kebutuhan komputasi, serta kualitas hasil yang dihasilkan. T5 Base, dengan 220 juta parameter, lebih ringan dan efisien, sehingga cocok untuk

tugas-tugas dengan dataset yang lebih kecil dan dapat dijalankan pada perangkat dengan daya komputasi terbatas. Sementara itu, T5 XL memiliki 3 miliar parameter, memungkinkan model untuk menangani tugas yang lebih kompleks dengan pemahaman yang lebih dalam terhadap pola linguistik, namun membutuhkan sumber daya komputasi yang lebih besar. Dari segi performa, T5 Base lebih cepat dalam proses inferensi tetapi kurang mampu menangkap konteks dalam data, sedangkan T5 XL mampu menghasilkan teks dengan kualitas lebih tinggi, terutama dalam skenario yang membutuhkan pemrosesan informasi secara lebih mendalam. Dalam konteks penelitian ini, T5 Base dapat digunakan untuk eksperimen awal atau skenario dengan keterbatasan komputasi, sedangkan T5 XL lebih unggul dalam menghasilkan narasi deskripsi CPMK yang lebih akurat dan kaya akan informasi. Arsitektur ini memungkinkan model untuk menghasilkan narasi atau ringkasan berdasarkan input yang diberikan, seperti dalam hal ini, deskripsi CPMK dan skor yang digunakan untuk menghasilkan narasi yang relevan.

T5 diadaptasi untuk menghasilkan narasi berbasis input berbentuk teks yang mencakup deskripsi CPMK dan skor. Tabel 3 menguraikan konfigurasi lapisan dan parameter yang digunakan untuk membangun model T5 yang diterapkan dalam penelitian ini, serta perbedaan antara T5 Base dan T5 XL dalam hal jumlah parameter, kapasitas model, dan kemampuan generalisasi untuk tugas-tugas *text-to-text*

Tabel 3. Arsitektur T5

Input Size	Operator	t	c	n	s
Variable (Text)	Embedding Layer	-	512	-	-
512	Encoder Layer (Self-Attention)	-	512	12	-
512	Encoder Layer (Feed-Forward)	-	512	12	-
512	Decoder Layer (Self-Attention)	-	512	12	-
512	Decoder Layer (Cross-Attention)	-	512	12	-
512	Decoder Layer (Feed-Forward)	-	512	12	-
512	Output Layer (Softmax)	-	Vocabulary size	-	-

Dalam model ini, input awal berupa CPMK dan skor digunakan untuk menghasilkan narasi yang relevan. Sebagai model berbasis *Transformers*, T5 memanfaatkan arsitektur *encoder-decoder*, yang dirancang untuk memproses input teks dan menghasilkan *output* yang sesuai dengan konteks yang diberikan. *Input* teks diolah terlebih dahulu menggunakan T5 *tokenizer*, yang mengubah teks menjadi representasi numerik agar dapat diproses oleh model.

Pada tahap awal, teks dimasukkan ke dalam *encoder*, yang mengubahnya menjadi representasi tersembunyi (*hidden representation*). Representasi ini kemudian diteruskan ke *decoder*, yang bertugas menyusun narasi akhir. Arsitektur T5 mengadopsi *self-attention mechanism*, yang memungkinkan model memahami hubungan antar kata dalam konteks yang lebih luas, tidak hanya terbatas pada kata-kata yang berdekatan. Selain itu, T5 menggunakan pendekatan *text-to-text*, di mana baik input maupun output diperlakukan sebagai teks, sehingga lebih fleksibel dalam berbagai tugas NLP.

Dalam penerapannya, arsitektur T5 memungkinkan model untuk menangani berbagai tugas generatif, termasuk *abstractive summarization*, *translation*, dan *text generation*. Dengan menggunakan mekanisme *cross-attention*, *decoder* dapat berfokus pada bagian-bagian penting dari *input* yang telah diproses oleh *encoder*, sehingga narasi yang dihasilkan lebih relevan dan koheren. Dalam konteks penelitian ini, pendekatan ini digunakan untuk menyusun deskripsi CPMK berdasarkan skor, sehingga model dapat menghasilkan teks yang sesuai dengan standar OBE.

2.5. Evaluation

Dilakukan pengujian dan evaluasi dengan dataset validasi yang telah disiapkan untuk mengukur kinerja model *text generation* yang telah dibangun. Kinerja model dalam tugas ini diukur berdasarkan kemampuan model untuk menghasilkan narasi yang sesuai dengan konteks yang diberikan. Evaluasi dilakukan menggunakan metrik seperti *Bilingual Evaluation Understudy* (BLEU) dan *Recall-Oriented Understudy for Gisting Evaluation* (ROUGE) untuk menilai kesesuaian, keberagaman, dan koherensi teks yang dihasilkan. BLEU digunakan untuk mengukur kesamaan antara n-gram dalam output model dan narasi referensi yang diberikan. Metrik BLEU yang tersaji pada rumus (1) memberikan gambaran tentang seberapa banyak kata atau frasa yang dihasilkan oleh model cocok dengan kata atau frasa dalam narasi referensi [19]. Semakin tinggi skor BLEU, semakin baik model dalam menghasilkan teks yang sesuai.

$$BLEU = BP \cdot \exp \left(\sum_{n=1}^N w_n \cdot \log p_n \right) \quad (1)$$

Keterangan:

1. Brevity Penalty (BP): Dikenakan penalty jika output lebih pendek dari referensi.
2. w : bobot semua n-gram.
3. p : nilai precision.

ROUGE adalah metrik evaluasi yang digunakan untuk mengukur kualitas teks yang dihasilkan oleh model berdasarkan teks referensi [20]. ROUGE berfokus untuk mengukur kesamaan teks dengan narasi referensi, baik dari sisi struktur maupun informasi penting yang terkandung. Dalam konteks penelitian ini, evaluasi ROUGE mencakup tiga metrik utama: ROUGE-1, ROUGE-2, dan ROUGE-L, yang masing-masing mengukur aspek berbeda dari teks yang dihasilkan:

1. ROUGE-1

Mengukur kesesuaian antara teks yang dihasilkan dengan referensi menggunakan unigram (kata Tunggal). Pengukuran *precision*, *recall*, dan *f1-score* pada ROUGE-1 terdapat pada rumus (2)-(4).

$$Precision = \frac{\text{Jumlah unigram yang cocok}}{\text{Jumlah total unigram dalam teks yang dihasilkan}} \quad (2)$$

$$Recall = \frac{\text{Jumlah total unigram dalam referensi}}{\text{Jumlah unigram yang cocok}} \quad (3)$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (4)$$

2. ROUGE-2

Mengukur kesesuaian menggunakan bigram (dua kata berurutan). Pengukuran *precision*, *recall*, dan *f1-score* pada ROUGE-2 terdapat pada rumus (5)-(7).

$$Precision = \frac{\text{Jumlah bigram yang cocok}}{\text{Jumlah total bigram dalam teks yang dihasilkan}} \quad (5)$$

$$Recall = \frac{\text{Jumlah bigram yang cocok}}{\text{Jumlah total bigram dalam referensi}} \quad (6)$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (7)$$

3. ROUGE-L

Mengukur kesesuaian berdasarkan Longest Common Subsequence (LCS), yaitu urutan kata terpanjang yang sama antara teks yang dihasilkan dan referensi. Pengukuran *precision*, *recall*, dan *f1-score* pada ROUGE-L terdapat pada rumus (8)-(10).

$$Precision = \frac{\text{Panjang LCS}}{\text{Jumlah kata dalam teks yang dihasilkan}} \quad (8)$$

$$Recall = \frac{\text{Panjang LCS}}{\text{Jumlah kata dalam referensi}} \quad (9)$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (10)$$

3. Hasil dan Pembahasan

Pada penelitian kali ini peneliti melakukan pada model berbasis arsitektur T5, yaitu T5 base, T5 XL, dan T5 base yang telah di *fine-tuning* dengan menggunakan dua Prompt yaitu, Prompt Mata Kuliah dan Prompt CPMK. Dimana pada masing-masing prompt berisi 3 variasi Prompt yang berisikan tiga perintah, pada Prompt-1 perintah berisi "Berdasarkan mata kuliah yang diberikan, buatlah narasi baru yang relevan dengan informasi di atas.", Prompt-2 "Gunakan narasi referensi untuk membuat narasi baru yang sesuai dengan konteks mata kuliah dan nilai yang diberikan.", dan Prompt-3 "Buatlah narasi baru yang sesuai dengan konteks mata kuliah, cpmk, dan skor yang diberikan.". Hasil penelitian menunjukkan perbandingan performa tiga model yang telah digunakan dalam menghasilkan narasi berdasarkan deskripsi CPMK dan skor yang diberikan. Setelah dilakukan percobaan menggunakan tiga model menggunakan arsitektur T5, yaitu T5 base, T5 XL, dan T5 base yang di *fine-tuning*, menghasilkan evaluasi performa dalam matrik evaluasi.

3.1. Pengujian Mata Kuliah

Tabel 4 menyajikan hasil pengujian berdasarkan Mata Kuliah. Data dalam tabel ini mencakup perbandingan performa model T5 dalam menghasilkan profil mata kuliah berbasis CPMK. Evaluasi dilakukan menggunakan metrik BLEU dan ROUGE untuk mengukur akurasi dan kesesuaian teks yang dihasilkan. Hasil ini memberikan gambaran tentang efektivitas masing-masing varian model dalam pendekatan *Outcome-Based Education* (OBE).

Tabel 4. Hasil Pengujian Mata Kuliah

Pengujian Mata Kuliah													
Model	Nilai	Hasil Evaluasi											
		Prompt-1				Prompt-2				Prompt-3			
		BLEU	Rouge-1	Rouge-2	Rouge-L	BLEU	Rouge-1	Rouge-2	Rouge-L	BLEU	Rouge-1	Rouge-2	Rouge-L
T5 XL	0-20	0,40	0,469	0,376	0,508	0,56	0,629	0,520	0,639	0,37	0,457	0,332	0,453
	21-40	0,52	0,651	0,462	0,655	0,49	0,592	0,459	0,602	0,31	0,433	0,282	0,462
	41-60	0,47	0,605	0,418	0,577	0,50	0,572	0,455	0,549	0,44	0,457	0,354	0,482
	61-80	0,53	0,642	0,599	0,649	0,47	0,576	0,464	0,560	0,31	0,372	0,305	0,404
	81-100	0,55	0,684	0,537	0,695	0,57	0,634	0,527	0,644	0,49	0,573	0,424	0,614
	T5 Base	0-20	0,39	0,309	0,239	0,293	0,35	0,313	0,225	0,284	0,35	0,294	0,190
21-40	0,37	0,292	0,272	0,272	0,35	0,320	0,286	0,297	0,36	0,275	0,221	0,277	
41-60	0,37	0,248	0,257	0,222	0,33	0,221	0,207	0,218	0,39	0,287	0,364	0,265	
61-80	0,28	0,288	0,189	0,302	0,36	0,309	0,202	0,295	0,34	0,322	0,233	0,301	
81-100	0,38	0,310	0,223	0,269	0,32	0,261	0,290	0,257	0,33	0,326	0,409	0,339	
T5 Base FT	0-20	0,45	0,534	0,500	0,524	0,51	0,639	0,609	0,635	0,47	0,581	0,543	0,574
21-40	0,37	0,422	0,393	0,414	0,34	0,555	0,533	0,553	0,34	0,461	0,431	0,453	
41-60	0,34	0,490	0,466	0,481	0,43	0,541	0,524	0,541	0,37	0,515	0,493	0,511	
61-80	0,41	0,460	0,411	0,452	0,47	0,569	0,551	0,568	0,46	0,471	0,439	0,467	
81-100	0,33	0,471	0,440	0,463	0,33	0,532	0,514	0,529	0,47	0,523	0,487	0,520	

Dalam hasil percobaan menggunakan prompt mata kuliah dengan tiga variasi prompt menunjukkan bahwa model T5 XL, T5 Base yang di *fine-tuning* (T5 Base FT) memiliki performa yang bervariasi dalam menghasilkan narasi berbasis nilai CPMK. Pada model T5 XL secara konsisten menghasilkan skor evaluasi tertinggi di semua metrik dan rentan nilai. Pada variasi Prompt-1, Model mencapai skor BLEU 0,552 dan ROUGE-L 0,695 untuk rentan nilai 81-100, sementara pada variasi Prompt-2 dan Prompt-3, skor BLEU masing-masing mencapai 0,573 dan 0,491, dengan ROUGE-L tertinggi sebesar 0,644 dan 0,614. Hal ini menunjukkan kapasitas model T5 XL dalam menangkap konteks dan menghasilkan narasi yang relevan.

Model T5 Base *fine-tuning* menunjukkan peningkatan yang signifikan disbanding T5 Base standar. *Fine-tuning* memungkinkan T5 Base *fine-tuning* menghasilkan skor BLEU dan ROUGE-L yang lebih tinggi, terutama pada rentang nilai 0-20 dan 61-80. Sebagai contoh, pada Prompt-1 untuk rentang nilai 0-20, BLEU mencapai 0,452 dan ROUGE-L 0,524, sementara pada Prompt-2 untuk rentang nilai 61-80, BLEU tercatat sebesar 0,468 dengan ROUGE-L 0,568. Namun, meskipun T5 Base FT menunjukkan perbaikan, performanya masih lebih rendah dibandingkan T5 XL, khususnya pada rentang nilai tinggi.

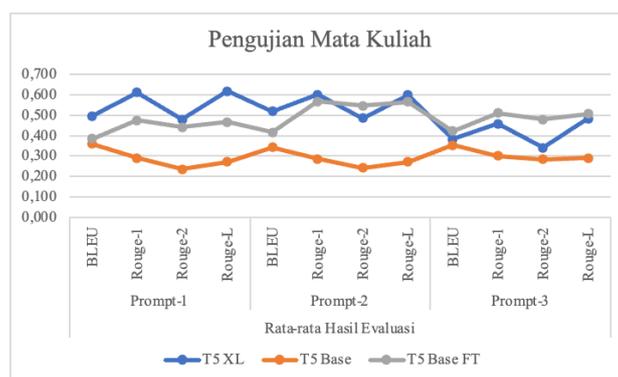
Di sisi lain, model T5 Base tanpa *fine-tuning* memiliki performa terendah di semua metrik evaluasi. Skor BLEU dan ROUGE-L untuk model ini cenderung stagnan pada nilai rendah, seperti pada Prompt-1 dengan BLEU 0,394 dan ROUGE-L 0,293 untuk rentang nilai 0-20, serta pada Prompt-2 dengan BLEU 0,361 dan ROUGE-L 0,295 untuk rentang nilai 61-80. Hal ini menunjukkan keterbatasan model T5 Base dalam memahami konteks tanpa adaptasi terhadap data spesifik.

Secara umum, performa semua model cenderung meningkat pada rentang nilai 81-100, yang menunjukkan bahwa narasi dengan nilai tinggi lebih mudah diprediksi. Selain itu, Prompt-2 memberikan skor evaluasi yang lebih tinggi dibandingkan Prompt-1 dan Prompt-3, menunjukkan bahwa pola atau struktur pada Prompt-2 lebih mudah dipahami oleh model. Dari hasil ini, dapat disimpulkan bahwa T5 XL adalah model terbaik untuk menghasilkan narasi berbasis CPMK, sementara *fine-tuning* memberikan dampak positif pada performa T5 Base, meskipun tetap lebih rendah dari T5 XL.

Berikutnya pada Tabel 5 dan Gambar 2 menjelaskan bahwa rata-rata skor yang di dapat pada hasil pengujian tiga model yaitu, T5 XL, T5 Base, dan T5 *Base fine-tuning* pada prompt Mata Kuliah.

Tabel 5. Rata-Rata Hasil Pengujian Mata Kuliah

Model	Pengujian Mata Kuliah											
	Rata-rata Hasil Evaluasi											
	Prompt-1				Prompt-2				Prompt-3			
	BLEU	Rouge-1	Rouge-2	Rouge-L	BLEU	Rouge-1	Rouge-2	Rouge-L	BLEU	Rouge-1	Rouge-2	Rouge-L
T5 XL	0,496	0,610	0,478	0,617	0,518	0,601	0,485	0,599	0,383	0,459	0,340	0,483
T5 Base	0,359	0,290	0,236	0,272	0,341	0,285	0,242	0,270	0,354	0,301	0,283	0,290
T5 Base FT	0,385	0,476	0,442	0,467	0,415	0,567	0,546	0,565	0,422	0,510	0,479	0,505



Gambar 2. Grafik Pengujian Berdasarkan Mata Kuliah

Dimana pada model T5 XL rata rata skor evaluasi yang di dapatkan lebih tinggi pada variasi Prompt-1 dan Prompt-2, sedangkan pada model T5 Base fine-tuning rata rata skor evaluasi yang dihasilkan lebih rendah dari model T5 XL meskipun, pada variasi Prompt-3 skor hasil evaluasi masih di ungguli oleh model T5 Base fine-tuning. Sedangkan pada model T5 Base rata-rata hasil skor evaluasi yang didapatkan masih lebih rendah dibandingkan dengan model T5 XL dan T5 fine-tuning, mengindikasikan kemungkinan struktur atau pola tertentu yang lebih mudah dipelajari oleh model. Dengan demikian, T5 XL tetap menjadi pilihan terbaik untuk menghasilkan narasi berbasis CPMK, sementara *fine-tuning* memberikan dampak positif pada T5 Base untuk meningkatkan performanya. Hasil rata-rata juga didapatkan pada tiga model berbeda yaitu T5 XL, T5 Base, dan T5 Base *fine-tuning* dengan menggunakan prompt CPMK.

3.2. Pengujian Capaian Pembelajaran Mata Kuliah (CPMK)

Tabel 6. Tabel ini berisi perbandingan kinerja berbagai varian model T5 dalam pengujian CPMK. Evaluasi dilakukan menggunakan metrik BLEU dan ROUGE untuk mengukur akurasi dan kesesuaian teks yang dihasilkan. Hasil ini memberikan gambaran tentang efektivitas masing-masing varian model dalam pendekatan *Outcome-Based Education (OBE)*.

Tabel 6. Hasil Pengujian CPMK

Model	Nilai	Pengujian CPMK											
		Hasil Evaluasi											
		Prompt-1				Prompt-2				Prompt-3			
	BLEU	Rouge-1	Rouge-2	Rouge-L	BLEU	Rouge-1	Rouge-2	Rouge-L	BLEU	Rouge-1	Rouge-2	Rouge-L	
T5 XL	0-20	0,63	0,794	0,722	0,789	0,61	0,800	0,702	0,794	0,51	0,626	0,555	0,621
	21-40	0,68	0,826	0,758	0,823	0,67	0,807	0,734	0,802	0,68	0,753	0,704	0,750
	41-60	0,68	0,817	0,753	0,811	0,62	0,753	0,688	0,731	0,65	0,779	0,713	0,775
	61-80	0,73	0,847	0,794	0,846	0,73	0,864	0,795	0,863	0,61	0,694	0,637	0,690
	81-100	0,72	0,861	0,798	0,858	0,73	0,863	0,800	0,861	0,61	0,729	0,662	0,749
T5 Base	0-20	0,27	0,220	0,205	0,194	0,32	0,211	0,246	0,180	0,30	0,208	0,257	0,226
	21-40	0,38	0,261	0,215	0,222	0,40	0,225	0,237	0,195	0,30	0,307	0,212	0,316
	41-60	0,23	0,175	0,278	0,183	0,31	0,259	0,247	0,227	0,35	0,212	0,196	0,231
	61-80	0,30	0,265	0,230	0,304	0,32	0,261	0,239	0,233	0,23	0,201	0,201	0,195
	81-100	0,30	0,203	0,253	0,195	0,34	0,177	0,306	0,153	0,30	0,275	0,289	0,275
T5 Base FT	0-20	0,50	0,558	0,534	0,558	0,49	0,544	0,522	0,544	0,46	0,517	0,494	0,517
	21-40	0,46	0,415	0,393	0,413	0,45	0,494	0,474	0,492	0,43	0,445	0,404	0,438
	41-60	0,42	0,477	0,448	0,474	0,34	0,491	0,473	0,491	0,43	0,521	0,349	0,471
	61-80	0,46	0,458	0,441	0,458	0,46	0,478	0,460	0,478	0,46	0,489	0,447	0,489
	81-100	0,40	0,447	0,430	0,447	0,37	0,515	0,499	0,515	0,40	0,438	0,420	0,438

Pada pengujian menggunakan *prompt* CPMK hasilnya performa masing-masing model T5 XL, T5 Base, dan T5 Base *fine-tuning*. Model T5 XL secara konsisten menunjukkan performa terbaik dibandingkan dengan model lainnya di semua rentang nilai dan variasi *prompt*. Hal ini terlihat dari skor BLEU dan ROUGE yang selalu lebih tinggi dibandingkan T5 Base dan T5 Base *fine-tuning*. Pada Prompt-1, misalnya, untuk rentang nilai 61-80, T5 XL memperoleh skor BLEU 0,732 dan ROUGE-L 0,846, yang menunjukkan keunggulan model ini dalam menghasilkan narasi yang mendekati data referensi. Skor tertinggi lainnya dapat diamati pada rentang nilai 81-100, di mana BLEU mencapai 0,731 pada Prompt-2 dan 0,608 pada Prompt-3, mengindikasikan akurasi model yang konsisten pada narasi dengan nilai tinggi.

Sebaliknya, model T5 Base memiliki performa terendah dibandingkan dengan dua model lainnya. Skor evaluasinya cenderung rendah pada semua variasi *prompt* dan rentang nilai, terutama pada rentang nilai rendah seperti 0-20 dan 21-40. Sebagai contoh, pada variasi Prompt-2 untuk rentang nilai 0-20, T5 Base hanya mencapai BLEU 0,323 dan ROUGE-L 0,180. Hal ini menunjukkan bahwa model T5 Base tanpa *fine-tuning* memiliki keterbatasan dalam memahami data dan menghasilkan narasi yang sesuai dengan kebutuhan evaluasi CPMK.

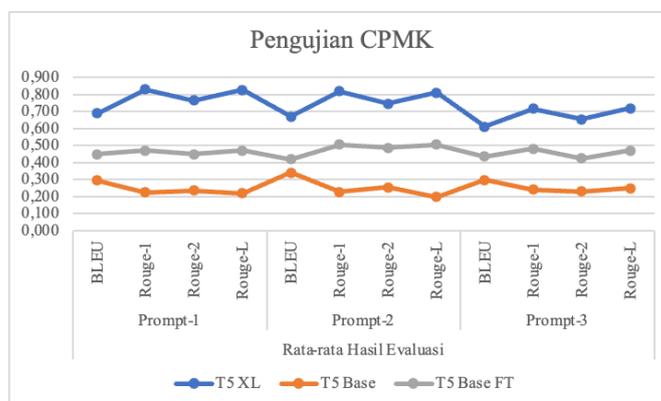
Model T5 Base *fine-tuning*, menunjukkan peningkatan signifikan dibandingkan dengan T5 Base standar. Pada variasi Prompt-1 untuk rentang nilai 0-20, BLEU model ini mencapai 0,502 dan ROUGE-L 0,558, yang jauh lebih tinggi dibandingkan dengan skor T5 Base pada variasi *prompt* yang sama. Meskipun performanya meningkat setelah *fine-tuning*, T5 Base yang telah di *fine-tuning* masih tidak dapat menyaingi performa T5 XL, terutama pada rentang nilai tinggi seperti 81-100, di mana skor BLEU hanya mencapai 0,396 di variasi Prompt-1 dan 0,368 di variasi Prompt-2.

Secara umum, performa semua model cenderung lebih baik pada rentang nilai tinggi, khususnya 81-100, di mana model lebih mudah memprediksi pola narasi yang lebih jelas. Selain itu, variasi Prompt-1 dan variasi Prompt-2 secara umum memberikan hasil evaluasi yang lebih tinggi dibandingkan variasi Prompt-3, mengindikasikan kemungkinan struktur atau pola tertentu yang lebih mudah dipelajari oleh model. Dengan demikian, T5 XL tetap menjadi pilihan terbaik untuk menghasilkan narasi berbasis CPMK, sementara *fine-tuning* memberikan dampak positif pada T5 Base untuk meningkatkan performanya.

Tabel 7 dan Gambar 3 menjelaskan hasil berbeda didapat pada model yang menggunakan *prompt* CPMK bahwa rata-rata skor yang di dapat pada hasil pengujian tiga model yaitu, T5 XI, T5 Base, dan T5 Base *fine-tuning*.

Tabel 7. Rata-Rata Hasil Pengujian CPMK

Model	Pengujian CPMK											
	Rata-rata Hasil Evaluasi											
	Prompt-1				Prompt-2				Prompt-3			
	BLEU	Rouge-1	Rouge-2	Rouge-L	BLEU	Rouge-1	Rouge-2	Rouge-L	BLEU	Rouge-1	Rouge-2	Rouge-L
T5 XL	0,688	0,229	0,765	0,826	0,670	0,818	0,744	0,810	0,610	0,716	0,654	0,717
T5 Base	0,295	0,225	0,236	0,220	0,340	0,227	0,255	0,198	0,296	0,240	0,231	0,249
T5 Base FT	0,448	0,471	0,449	0,470	0,420	0,504	0,486	0,504	0,435	0,482	0,423	0,471



Gambar 3. Grafik Hasil Pengujian Berdasarkan CPMK

Dimana pada model T5 XL rata rata skor evaluasi yang di dapatkan lebih tinggi pada semua variasi *prompt*, sedangkan pada model T5 Base *fine-tuning* rata rata skor evaluasi yang dihasilkan lebih rendah dari model T5 XL meskipun, skor hasil evaluasi masih lebih tinggi jika

dibandingkan dengan model T5 Base tanpa *fine-tuning*, Dimana pada model T5 Base rata-rata hasil skor evaluasi yang didapatkan masih lebih rendah dibandingkan dengan model T5 XL dan T5 *fine-tuning*, ini mengindikasikan bahwa bukan hanya struktur atau pola tertentu yang lebih mudah dipelajari oleh model namun kualitas model juga cukup penting. Dengan demikian, T5 XL tetap menjadi pilihan terbaik untuk menghasilkan narasi berbasis CPMK, sementara *fine-tuning* memberikan dampak positif pada T5 Base untuk meningkatkan performanya.

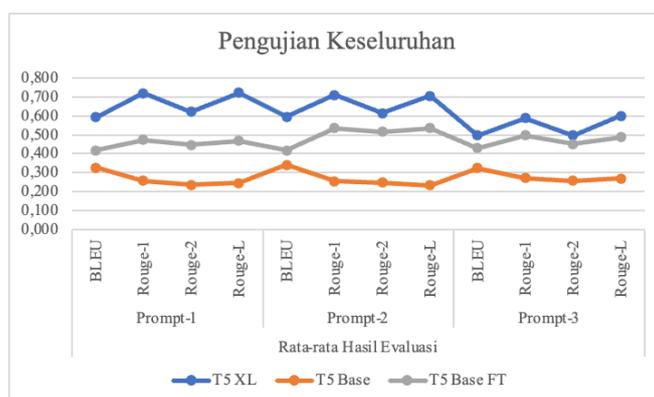
3.3. Rata-Rata Hasil Pengujian

Dari hasil percobaan yang sudah dilakukan pada masing-masing model T5 pada 2 prompt dimana setiap prompt-nya masing-masing memiliki 3 variasi prompt, diperoleh rata rata skor evaluasi dari seluruh pengujian.

Tabel 8 dan Gambar 4 menunjukkan hasil evaluasi dari tiga model T5 yang diuji pada 2 prompt dengan masing-masing 3 variasi prompt di dalamnya menunjukkan perbedaan kinerja yang signifikan.

Tabel 8. Rata-Rata Hasil Seluruh Pengujian

Model	Pengujian CPMK											
	Rata-rata Hasil Evaluasi											
	Prompt-1				Prompt-2				Prompt-3			
	BLEU	Rouge-1	Rouge-2	Rouge-L	BLEU	Rouge-1	Rouge-2	Rouge-L	BLEU	Rouge-1	Rouge-2	Rouge-L
T5 XL	0,592	0,720	0,622	0,721	0,594	0,709	0,614	0,705	0,496	0,587	0,497	0,600
T5 Base	0,327	0,257	0,236	0,246	0,341	0,256	0,248	0,234	0,325	0,270	0,257	0,269
T5 Base FT	0,417	0,473	0,446	0,468	0,417	0,536	0,516	0,535	0,428	0,496	0,451	0,488



Gambar 4. Grafik Rata-Rata Hasil Seluruh Pengujian

Model T5 XL tampil sebagai model dengan performa terbaik, menghasilkan nilai BLEU dan ROUGE yang lebih tinggi dibandingkan kedua model lainnya di seluruh prompt yang diuji. Ini mengindikasikan bahwa T5 XL lebih mampu menghasilkan teks yang sangat mirip dengan referensi yang diinginkan, dengan nilai BLEU rata-rata 0,592 dan ROUGE-L mencapai 0,721 pada Prompt-1, serta hasil serupa pada prompt lainnya. Sebaliknya, T5 Base menunjukkan kinerja yang lebih rendah, dengan nilai BLEU dan ROUGE yang lebih rendah, yang menunjukkan bahwa model ini kurang efektif dalam menghasilkan teks yang sesuai dengan referensi. Nilai BLEU untuk T5 Base rata-rata hanya 0,327, dan nilai-nilai ROUGE yang lebih rendah mencerminkan bahwa hasil teks yang dihasilkan masih jauh dari yang diharapkan.

Namun, T5 Base *Fine-Tuning* (FT) mengalami peningkatan signifikan setelah melalui proses *fine-tuning*, menghasilkan kinerja yang lebih baik dibandingkan dengan T5 Base. Nilai BLEU rata-rata pada T5 Base FT mencapai 0,417, dan nilai-nilai ROUGE juga menunjukkan peningkatan yang jelas, meskipun masih berada di bawah T5 XL. Secara keseluruhan, hasil ini menunjukkan bahwa T5 XL memberikan kinerja terbaik, diikuti oleh T5 Base FT yang lebih baik setelah *fine-tuning*, sementara T5 Base tetap berada di posisi paling bawah dalam hal kualitas teks yang dihasilkan. Hasil penelitian ini menunjukkan keselarasan dengan beberapa penelitian sebelumnya yang membuktikan bahwa T5 XL memiliki hasil yang lebih baik dibandingkan T5 Base [21]. Begitupun penelitian yang membuktikan bahwa T5 Base FT yang merupakan pengembangan dari model T5 Base memiliki hasil yang lebih baik dibandingkan dengan T5 base [22].

5. Simpulan

Penelitian ini bertujuan untuk mengotomasi pembuatan profil kemampuan mahasiswa berdasarkan capaian pembelajaran mata kuliah (CPMK) dalam sistem *Outcome-Based Education* (OBE) menggunakan teknologi *Text Generation* berbasis model *Transformers*, khususnya T5 (*Text-to-Text Transfer Transformers*). Hasil penelitian menunjukkan bahwa model T5 XL memiliki performa terbaik dibandingkan dengan T5 Base dan T5 Base yang telah di *fine-tuning*. T5 XL secara konsisten menghasilkan skor evaluasi BLEU dan ROUGE yang lebih tinggi, menunjukkan kemampuannya dalam memahami konteks narasi dan menghasilkan teks yang lebih relevan dan koheren. Proses *fine-tuning* pada T5 Base memberikan peningkatan performa signifikan, namun tetap tidak melebihi hasil yang dicapai oleh T5 XL, terutama pada narasi dengan nilai tinggi. Selain itu, variasi prompt yang lebih terstruktur, seperti Prompt-1 dan Prompt-2, terbukti lebih efektif dalam menghasilkan evaluasi yang lebih baik dibandingkan Prompt-3. Evaluasi menunjukkan bahwa model cenderung menghasilkan narasi yang lebih akurat pada nilai tinggi (81-100), dengan metrik BLEU dan ROUGE yang memberikan pengukuran kualitas teks secara efektif. Penelitian ini memberikan kontribusi pada pengembangan sistem asesmen berbasis OBE, dengan T5 XL sebagai model unggulan.

Untuk pengembangan ke depan, penelitian dapat diarahkan pada integrasi model ke dalam sistem yang lebih luas, peningkatan kemampuan generalisasi terhadap data yang lebih beragam, serta optimasi *fine-tuning* untuk efisiensi sumber daya. Penelitian ini menggunakan model yang masih termasuk dengan model yang efisien, pengembangan dapat difokuskan pada optimasi *fine-tuning* untuk meningkatkan efisiensi dan akurasi. Selain itu, eksperimen dengan variasi *dataset* yang lebih beragam dapat membantu meningkatkan generalisasi model. Untuk penelitian lebih lanjut, disarankan menggunakan model dengan kualitas dan arsitektur yang baik agar kualitas hasil yang dihasilkan juga lebih baik.

Daftar Referensi

- [1] Z. D. Rahmawati and S. Wahyuni, "Pengembangan Kurikulum Pendidikan Islam Multikultural Berbasis Outcome Based Education (OBE)," *TA'LIM: Jurnal Studi Pendidikan Islam*, vol. 7, no. 2, pp. 218-236, Jul.2024. doi: 10.52166/talim.v7i2.6895.
- [2] A. Aminuddin, R. Salambue, Y. Andriyani, E. Mahdiyah, P. Studi Sistem Informasi, and U. Riau, "Aplikasi E-OBE untuk Integrasi Komponen KurikulumOBE (Outcome Based Education)," *JSI: Jurnal Sistem Informasi (E-Journal)*, vol. 13, no. 1, pp. 2168-2182, Aug. 2021, doi: 10.18495/jsi.v13i1.34.
- [3] M. I. Muzakir and Susanto, "Implementasi Kurikulum Outcome Based Education (Obe) Dalam Sistem Pendidikan Tinggi Di Era Revolusi Industri 4.0", *edukasiana*, vol. 2, no. 1, pp. 118-139, May 2023. doi: 10.61159/edukasiana.v2i1.86.
- [4] M. I. Rarmli, M. A. T. ., and M. W. T. ., "Pelatihan Metode Pengukuran Capaian Pembelajaran Kurikulum Prodi Teknik Sipil Berbasis Outcome Based Education (OBE) pada Anggota BMPTTSSI", *Jurnal_Tepat*, vol. 5, no. 1, pp. 118-126, Jun. 2022. doi: 10.25042/jurnal_tepat.v5i1.226.
- [5] K. Henra, N. Q. Tayibu, and I. N. Masliah, "Pengaruh Pembelajaran Daring Asynchronous Terhadap Tingkat Pemenuhan CPMK Statistika," *JIPM (Jurnal Ilmiah Pendidikan Matematika)*, vol. 10, no. 1, p. 100-110, Aug. 2021, doi: 10.25273/jipm.v10i1.8537.
- [6] M. H. Zikry, A. B. Prasetijo, and R. Septiana, "Implementasi Front-end Sistem Penilaian Capaian Pembelajaran Lulusan (CPL) dan Capaian Pembelajaran Mata Kuliah (CPMK) (Studi Kasus pada Teknik Komputer Undip) Front-end Implementation of Program Learning Outcome (PLO) and Course Learning Outcome (CLO) (Case Study at Computer Engineering Undip)," *Jurnal Teknik Komputer*, vol. 3, no. 1, pp. 38-47, doi: 10.14710/jtk.v3i1.44282.
- [7] R. Putra, R. Ilyas, and F. Kasyidi, "Pembangkitan Kalimat Ilmiah Menggunakan Recurrent Neural Network", *SisInfo*, vol. 3, no. 1, pp. 11-20, Feb. 2021. doi: 10.37278/sisinfo.
- [8] Loviera Azahra, Qoniah Milladunka Nurhayati, Nuriya Parsa, Albany Indra Hafidz, and Ahmad Fu'adin, "Alat Bantu Baca Tunanetra Berbasis Teknologi Text-To-Speech Dan Optical Character Recognition", *Kohesi*, vol. 3, no. 2, pp. 1-10, Apr. 2024. doi: 10.3785/kohesi.v3i2.2899.
- [9] L. Gong, J. Crego, and J. Senellart, "Enhanced Transformers Model for Data-to-Text Generation." In *Proceedings of the 3rd Workshop on Neural Generation and Translation*, pp. 148-156, Nov.2019 doi: 10.18653/v1/D19-5615.

- [10] C. Alifia Putri and S. Al Faraby, "Classification of Sentiment Analysis on English Film Reviews with Analisis Sentimen Review Film Berbahasa Inggris Dengan Pendekatan Bidirectional Encoder Representations from Transformers approval. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, vol. 6, no. 2, pp. 181–193, Jan.2020, doi: 10.35957/jatisi.v6i2.206.
- [11] A. R. Subagyo and T. B. Sasongko, "Implementasi Algoritma Transformers BART dan Penggunaan Metode Optimasi Adam Untuk Klasifikasi Judul Berita Palsu," *Jurnal Media Informatika Budidarma*, vol. 8, no. 3, p. 1768-1777, Jul. 2024, doi: 10.30865/mib.v8i3.7852.
- [12] S. Bature, A. Olorunleke, O. Ibrahim, A. Kayode, and J. Bolakale, "Assessing Agricultural Literacy Among Senior Secondary School Students in Kwara State, Nigeria: Implications for Educational Interventions", *IJCETS*, vol. 11, no. 1, pp. 1-8, Apr. 2023. doi: 10.15294/ijcets.v11i1.54877.
- [13] I Nyoman Purnama and Ni Nengah Widya Utami, "Implementasi Peringkat Dokumen Berbahasa Indonesia Menggunakan Metode Text To Text Transfer Transformers (T5)", *JUTIK*, vol. 9, no. 4, pp. 381-391, Aug. 2023. doi: 10.36002/jutik.v9i4.2531.
- [14] Q. A. Itsnaini, M. Hayaty, A. D. Putra, and N. A. M. Jabari, "Abstractive Text Summarization using Pre-Trained Language Model 'Text-to-Text Transfer Transformers (T5)'," *ILKOM Jurnal Ilmiah*, vol. 15, no. 1, pp. 124–131, Apr. 2023, doi: 10.33096/ilkom.v15i1.1532.124-131.
- [15] J. Kaplan *et al.*, "Scaling Laws for Neural Language Models" Jan. 2020, doi: 10.48550/arXiv.2001.08361.
- [16] I. Borhan and A. Bajaj, "The Effect of Prompt Types on Text Summarization Performance With Large Language Models," *Journal of Database Management*, vol. 35, no. 1, pp. 1-23, Jan.2024, doi: 10.4018/JDM.358475.
- [17] Z. Y. Zhang, A. Verma, F. Doshi-Velez, and B. K. H. Low, "Understanding the Relationship between Prompts and Response Uncertainty in Large Language Models" Jul. 2024, doi: 10.48550/arXiv.2407.14845.
- [18] S. S. Hartinah and Sugiyono, "Pemodelan Data Mining Transaksi Penjualan Menggunakan Algoritma Apriori (Studi Kasus: Kedai Ngodeng & Smoothies)", *jimik*, vol. 5, no. 3, pp. 3080-3098, Sep. 2024. doi: 10.35870/jimik.v5i3.992
- [19] D. H. Putra and N. A. A. N. Syam, "Automasi Pembangkit Lirik Lagu Dalam Bahasa Indonesia: Implementasi GPT-NEO Dalam Pemrosesan Bahasa Kreatif", *Syntax: Journal of Software Engineering, Computer Science and Information Technology*, vol. 5, no. 2, pp. 612-616, Des.2024. doi: 10.46576/syntax.v5i2.5542.
- [20] Y. Yuniati, K. Milani Fitria, S. Purwiyanti, E. Nasrullah, M. A. Muhammad, and P. Korespondensi "Analisis Performa Ekstraksi Konten GPT-3 Dengan Matrik Bertscore Dan Rouge", *JTIK*, vol. 11, no. 6, pp. 1273–1280, Dec. 2024, doi: 10.25126/jtiik.1168088.
- [21] M. Guevara *et al.*, "Large language models to identify social determinants of health in electronic health records," *NPJ Digit Med*, vol. 7, no. 1, p. 1-37 Dec. 2024, doi: 10.1038/s41746-023-00970-0.
- [22] C. Raffel *et al.*, "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformers", *The Journal of Machine Learning Research*, vol. 21, No.140, p. 5485 - 5551, Jan. 2020, doi: 10.48550/arXiv.1910.10683.