

Evaluasi Kinerja Model Transformer: Mengukur Kesesuaian Capaian Pembelajaran Mata Kuliah

Sumijan¹, Husni Thamrin*², Galih Pramuja Inngam Fanani³

^{1,2}Prodi Magister Informatika, Universitas Muhammadiyah Surakarta

³Prodi Sistem dan Teknologi Informasi, Universitas 'Aisyiyah Surakarta

Email: ¹l208230002@student.ums.ac.id, ²husni.thamrin@ums.ac.id, ³galihfanani@aiska-university.ac.id

Abstrak

Kurikulum berorientasi pada hasil (Outcome-Based Education/OBE) menjadi pendekatan penting dalam pendidikan tinggi untuk memastikan kesesuaian antara Capaian Pembelajaran Lulusan (CPL) dan kebutuhan dunia kerja. Namun, tantangan dalam implementasinya masih muncul, khususnya dalam memastikan konsistensi antara CPL dan Capaian Pembelajaran Mata Kuliah (CPMK). Penelitian ini mengevaluasi efektivitas dua model transformer berbasis deep learning, BERT dan SBERT, dalam mengotomatisasi proses evaluasi kesesuaian CPL–CPMK. Eksperimen dilakukan pada data kurikulum dari Universitas 'Aisyiyah Surakarta. Hasil menunjukkan bahwa BERT unggul dalam tugas klasifikasi teks dengan nilai akurasi 0,61, precision 0,72, dan F1 score 0,73. Sebaliknya, SBERT menunjukkan recall lebih tinggi (0,75), yang berguna dalam konteks minimalisasi kesalahan tipe dua. Temuan ini menjadikan BERT sebagai model yang lebih cocok digunakan untuk mengukur keselarasan antara CPL dengan CPMK.

Kata kunci: BERT, CPL, CPMK, Evaluasi Kurikulum, SBERT, Transformer.

Evaluating the Performance of Transformer Models: Measuring the Alignment of Course Learning Outcomes

Abstract

Outcome-Based Education (OBE) has become an essential approach in higher education to ensure alignment between Graduate Learning Outcomes (Capaian Pembelajaran Lulusan – CPL) and the needs of the job market. However, challenges in its implementation persist, particularly in maintaining consistency between CPL and Course Learning Outcomes (Capaian Pembelajaran Mata Kuliah – CPMK). This study evaluates the effectiveness of two transformer-based deep learning models, BERT and SBERT, in automating the evaluation process of CPL–CPMK alignment. Experiments were conducted using curriculum data from Universitas 'Aisyiyah Surakarta. The results show that BERT outperforms in text classification tasks, achieving an accuracy of 0.61, precision of 0.72, and F1 score of 0.73. In contrast, SBERT demonstrated a higher recall (0.75), which is beneficial in minimizing type II errors. These findings suggest that BERT is the more suitable model for assessing the alignment between CPL and CPMK.

Keywords: BERT, CPL, CPMK, Curriculum Evaluation, SBERT, Transformer.

1. PENDAHULUAN

Kurikulum merupakan fondasi utama dalam penyelenggaraan pendidikan tinggi yang mencakup perencanaan dan pengaturan mengenai Capaian Pembelajaran Lulusan (CPL), materi ajar, proses pembelajaran, dan sistem penilaian [1]. Kurikulum berfungsi sebagai pedoman dalam pelaksanaan program studi, baik di tingkat fakultas maupun universitas [2]. Salah satu elemen kritikal dalam kurikulum adalah keterkaitan antara capaian pembelajaran dengan kebutuhan nyata di dunia kerja dan masyarakat [3]. Namun, realitas saat ini menunjukkan adanya kesenjangan yang signifikan antara kompetensi lulusan perguruan tinggi dengan tuntutan dunia kerja. Lulusan kerap dianggap belum siap secara praktis karena tidak memiliki keterampilan yang relevan, baik dari sisi teknis maupun non-teknis [4]. Padahal, di era globalisasi dan disruptif digital, kompetensi lulusan dituntut semakin tinggi, tidak hanya untuk menjadi pekerja, tetapi juga inovator, wirausahawan, dan agen perubahan sosial [5].

Berbagai upaya telah dilakukan untuk menyelaraskan pendidikan tinggi dengan kebutuhan industri, salah satunya melalui konsep *link and match*. Namun, pendekatan ini saja tidak cukup. Dinamika sosial, ekonomi, dan teknologi yang sangat cepat menuntut sistem pendidikan tinggi untuk lebih adaptif, akuntabel, dan berorientasi

pada luaran (*outcome-based*), bukan sekadar proses pengajaran formal [6]. Demi menjawab tantangan tersebut, diterapkanlah pendekatan *Outcome-Based Education* (OBE), yang menekankan pada pencapaian hasil belajar sebagai pusat dari seluruh proses pendidikan. Dalam OBE, CPL menjadi titik tolak perancangan kurikulum, yang kemudian diturunkan menjadi Capaian Pembelajaran Mata Kuliah (CPMK) melalui dokumen Rencana Pembelajaran Semester (RPS). CPMK harus terformulasi secara spesifik dan terukur, sehingga memungkinkan untuk dievaluasi secara obyektif [7].

Sayangnya, dalam praktik implementasi OBE, sering kali ditemukan ketidaksesuaian antara CPMK dan CPL, yang menandakan bahwa proses perancangan kurikulum belum sepenuhnya selaras dan konsisten [8]. Ketidakseselarasan ini dapat menyebabkan tujuan pembelajaran tidak tercapai secara optimal, sehingga mempengaruhi mutu lulusan dan menurunkan tingkat relevansi mereka di pasar kerja [9].

Penelitian sebelumnya yang berjudul "Mengenal Model BERT dan Implementasinya untuk Analisis Sentimen Ulasan Game" menunjukkan bahwa model BERT secara konsisten memberikan performa yang lebih baik dibandingkan model-model tradisional, dengan peningkatan akurasi rata-rata sebesar 3–7% [10]. Studi lain berjudul "Japanese Short Answer Grading for Japanese Language Learners Using the Contextual Representation of BERT" mengindikasikan bahwa BERT memiliki keunggulan dibandingkan SBERT, dengan peningkatan akurasi dari 59,3% menjadi 61,5% [11]. Penelitian dengan judul "*Performance of Machine Learning Algorithms on Automatic Summarization of Indonesian Language Texts*" menyatakan bahwa Pendekatan hibrida LexRank+BERT menunjukkan kinerja superior dibandingkan metode ekstraktif konvensional, dengan pencapaian skor evaluasi ROUGE yang lebih tinggi pada aspek presisi, recall, dan F1 [12]. Penelitian dengan judul Efektivitas Algoritma Semantik dengan Keterkaitan Kata dalam Mengukur Kemiripan Teks Bahasa Indonesia menghasilkan jumlah kata korelasi 0,7128, cosine similarity menghasilkan korelasi 0,7408, dan semantik berbasis hubungan kata memberikan performa terbaik dengan korelasi 0,7508 [13].

Penelitian lain dengan judul *Classification of SWOT Statements Employing BERT PreTrained Model Embedding* menyimpulkan bahwa model SBERT yang digunakan untuk mengubah kalimat menjadi vektor, jika digabung dengan algoritma klasifikasi Support Vector Machine (SVM), menghasilkan kinerja terbaik [14]. Sementara itu, penelitian ketiga yang berjudul "A Review and Comparative Study of Cancer Detection Using Machine Learning: SBERT and SimCSE Application" menunjukkan bahwa SBERT mampu merepresentasikan sekuen DNA secara efektif untuk keperluan deteksi kanker, dengan tingkat akurasi berkisar antara 70–75% [15]. SBERT terbukti efektif dan andal untuk mendeteksi kemiripan semantik dalam deduplikasi judul-judul penelitian, dan hasilnya mendukung asumsi bahwa duplikasi semantik dalam dataset tersebut rendah [16].

Sebagai respons terhadap hal tersebut, Universitas 'Aisyiyah Surakarta mulai menerapkan kurikulum berbasis OBE pada semester ganjil tahun akademik 2023/2024, yang melibatkan tiga fakultas dan sepuluh program studi. Dalam konteks ini, evaluasi sistematis dan akurat terhadap kesesuaian CPL dan CPMK menjadi kebutuhan mendesak. Pendekatan konvensional yang mengandalkan evaluasi manual cenderung tidak efisien, rawan bias, dan sulit diterapkan secara skala besar [17]. Seiring dengan perkembangan teknologi kecerdasan buatan, khususnya model pembelajaran mesin berbasis *transformer*, muncul peluang baru untuk mengotomatisasi dan menyempurnakan proses evaluasi tersebut [18]. Model seperti BERT dan SBERT telah menunjukkan performa yang menjanjikan dalam memahami makna semantik teks dan relasi antar dokumen. Oleh karena itu, penelitian ini dilakukan untuk mengevaluasi dan membandingkan kinerja dua model transformer, BERT dan SBERT, dalam mengukur kesesuaian antara CPL dan CPMK [19].

Tulisan ini mendeskripsikan hasil penelitian terkait evaluasi kesesuaian antara capaian pembelajaran matakuliah dan capaian pembelajaran lulusan. Evaluasi tersebut menjadi bagian awal dari desain strategi pembelajaran sebuah matakuliah yang merupakan bagian integral sebuah kurikulum OBE. Kontribusi tulisan ini terletak pada pembuktian bahwa model transformer mampu membantu dalam menilai keselarasan capaian pembelajaran mata kuliah dan capaian pembelajaran lulusan.

2. METODE PENELITIAN

Penelitian bertujuan untuk mengevaluasi kinerja model *transformer*, khususnya algoritma BERT (*Bidirectional Encoder Representations from Transformers*) dan SBERT (*Sentence-BERT*), dalam mengukur kesesuaian antara CPMK dengan CPL. Model BERT digunakan untuk menganalisis kesesuaian berbasis kesamaan kata, sementara SBERT digunakan untuk mengukur kesesuaian pada level kalimat secara semantik [20]. Secara umum, metode penelitian ini terbagi ke dalam beberapa tahapan sebagai berikut.

2.1. Alat Dan Bahan

Penelitian memanfaatkan kombinasi perangkat keras (*hardware*) dan perangkat lunak (*software*) untuk mendukung pelaksanaan eksperimen secara optimal. Pemilihan alat dan bahan dilakukan berdasarkan kebutuhan fungsional, efisiensi, serta kemampuan dalam mendukung proses pengumpulan dan analisis data. Table 1. Merupakan beberapa software yang dibutuhkan untuk mendukung penelitian ini. Table 2. Merupakan beberapa perangkat yang digunakan untuk membangun

Tabel 1. *Software* untuk membuat aplikasi

<i>Software</i>	Keterangan
Windows 10	Sebagai Sistem Operasi
Visual Studio Code	Tools Coding
Bahasa Pemrograman	Sebagai analisis BERT
Pyton	dan SBERT otomatis
<i>Microsoft Excel</i>	Tools Analisis manual

Tabel 2. *Hardware* untuk server dan analisis

<i>Hardware</i>	Spesifikasi
Laptop 1	Intel Core i3, RAM DDR4 8 GB, SSD 512 GB
Server	Ubuntu Server 20.04.06 LTS
Komputer 1	Intel Core i5, RAM DDR4 32 GB, SSD 2 TB

2.2. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan metode eksperimen sistem untuk mengevaluasi kinerja model berbasis *transformer* dalam konteks pendidikan tinggi. Evaluasi dilakukan dengan membandingkan hasil kesesuaian yang diukur oleh sistem berbasis BERT dan SBERT dengan hasil penilaian manual menggunakan pendekatan taksonomi Bloom melalui Kata Kerja Operasional (KKO) [21][22].

2.3. Tahapan Pengumpulan Data

Pengumpulan data dilakukan dengan cara sebagai berikut:

- Mengumpulkan data CPL dari 10 Program Studi (Prodi) yang berasal dari 3 fakultas berbeda.
- Mengumpulkan data CPMK dari masing-masing mata kuliah, dengan total 210 mata kuliah dan 799 CPMK.
- Proses perhitungan kesesuaian dilakukan secara otomatis menggunakan model BERT dan SBERT.
- Sebagai pembanding, dilakukan penilaian manual oleh manusia menggunakan pendekatan KKO Bloom.
- Analisis dan evaluasi *accuracy*, *precision*, *recall* dan *F1 score*. Evaluasi dilakukan dengan cara membandingkan hasil prediksi model terhadap data referensi atau label sebenarnya. Analisis ini bertujuan untuk menentukan model mana yang memiliki kinerja paling baik dalam mengukur kesesuaian antara CPMK dan CPL secara akurat dan efektif. Gambar 1. Merupakan alur diagram blok dari tahapan awal sampai akhir.



Gambar 1. Alur Blok Diagram

2.4. Tahapan Perhitungan Data

Data hasil pengolahan dari sistem kemudian dianalisis menggunakan perangkat lunak *Microsoft Excel*. Analisis ini dilakukan dalam dua tahap utama:

1. Penentuan ambang batas (*threshold*) untuk menentukan klasifikasi kesesuaian antara CPMK dan CPL pada hasil skor dari BERT dan SBERT.
2. Perhitungan metrik evaluasi untuk menilai kinerja model, yaitu:
 - a) *Accuracy*
 - b) *Precision*
 - c) *Recall*
 - d) *F1 Score*

Perhitungan ini dilakukan dengan membandingkan hasil prediksi sistem dengan hasil evaluasi manual sebagai data *ground truth* [23].

2.5. Rumus Evaluasi Transformer

Berikut adalah penjelasan dari masing-masing istilah yang sering digunakan dalam evaluasi kinerja model klasifikasi berdasarkan metrik evaluasi [24][25] :

a) *Accuracy*

Accuracy (Akurasi) adalah proporsi prediksi yang benar dari seluruh prediksi yang dilakukan.

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+FP+TN} \quad (1)$$

b) *Precision*

Precision (Presisi) adalah seberapa akurat proporsi prediksi positif yang benar dari semua prediksi positif yang dilakukan oleh model.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

c) Recall

Recall (Sensitivitas atau *True Positive Rate*) adalah seberapa banyak proporsi data positif yang berhasil diprediksi dengan benar oleh model.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

d) F1 Score

F1 Score adalah rata-rata harmonis antara *Precision* dan *Recall*. Digunakan ketika kita ingin menyeimbangkan kedua metrik tersebut.

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

Keterangan :

- TP (*True Positive*): Prediksi positif yang benar
- TN (*True Negative*): Prediksi negatif yang benar
- FP (*False Positive*): Prediksi positif yang salah
- FN (*False Negative*): Prediksi negatif yang salah

3. HASIL DAN PEMBAHASAN

Hasil evaluasi kinerja model BERT dan SBERT dalam mengukur kesesuaian antara CPMK dan CPL. Model BERT yang digunakan adalah *cahya/bert-base-indonesia-1.5G* yaitu model berbahasa indonesia sebesar 1,5 GB sedangkan model SBERT menggunakan *naufalihsan/indonesia-sbert-large* yaitu model berbahasa indonesia sebesar 1,5 GB. Hasil secara otomatis yang dilakukan sistem computer kemudian dibandingkan dengan perhitungan yang dilakukan oleh manusia secara manual berdasarkan taksonomi KKO Bloom. Evaluasi kinerja model dilakukan dengan menghitung metrik klasifikasi yaitu *accuracy*, *precision*, *recall*, dan *F1-score*.

3.1. Tahapan Pengumpulan Data

Setelah data berhasil dikumpulkan ke dalam sistem aplikasi RPS yang telah disiapkan, langkah selanjutnya adalah melakukan proses perhitungan secara otomatis menggunakan algoritma BERT dan SBERT. Proses perhitungan ini diterapkan secara terpisah pada masing-masing mata kuliah yang diteliti. Setiap mata kuliah memiliki jumlah dan karakteristik CPMK yang berbeda, sehingga hasil perhitungannya pun bervariasi. Gambar 2. Menyajikan visualisasi hasil perhitungan otomatis yang dihasilkan oleh sistem berdasarkan algoritma yang digunakan.

Data Capaian Pembelajaran Mata Kuliah (CPMK) -- Manajemen Dalam Keperawatan						
No	No CPL	CPL	No CPMK	CPMK	BERT15	SBERT15
1	7	Mampu melakukan asuhan keperawatan dalam berbagai tatanan pelayanan keperawatan (KKU4)--P2	2	Merumuskan <i>Planning, Organizing, Staffing, Actuating</i> dan <i>Controlling</i> (POSAC) dalam manajemen keperawatan dan mendemonstrasikan metode penugasan keperawatan sesuai MPKP.	0.8	0.54
2	8	Mampu menerapkan dakwah Islami dan menjadi kader Muhammadiyah - 'Aisyiyah (KKU5) -- P1	1	Menjelaskan konsep manajemen secara umum, manajemen kesehatan dan manajemen keperawatan dengan memperhatikan prinsip dan etik keperawatan dan nilai-nilai islami.	0.8	0.47
3	14	Mampu melaksanakan prosedur manajemen bencana sesuai standar dan kewenangannya(KKK6) -- P2	3	Mahasiswa mampu menjelaskan dan mendemonstrasikan strategi dalam manajemen konflik.	0.75	0.33

Gambar 2. Hasil perhitungan secara otomatis

Setelah perhitungan otomatis dilakukan, hasil tersebut divalidasi melalui proses peninjauan manual oleh evaluator manusia. Validasi ini menggunakan pendekatan taksonomi KKO Bloom. Gambar 3. memvisualisasikan proses pemilihan dan penyelarasan antara CPL dan CPMK. Dalam tahap ini, CPL dikategorikan ke dalam salah satu dari tiga domain utama, yaitu kognitif, afektif, dan psikomotorik, sebagaimana merujuk pada taksonomi pembelajaran. Sementara itu, CPMK dipilih dari enam level berdasarkan KKO, yang mencerminkan tingkat kompleksitas pembelajaran yang diharapkan. Proses penilaian kesesuaian dilakukan dengan membandingkan tingkat CPL terhadap tingkat CPMK. Kesesuaian dianggap tercapai apabila level CPMK berada pada tingkat yang sama atau lebih rendah dibandingkan dengan CPL. Sebaliknya, apabila CPMK memiliki tingkat yang lebih tinggi dari CPL, maka hasilnya dikategorikan sebagai tidak sesuai. Setelah seluruh data dianalisis dan kesesuaian antar capaian dievaluasi, proses dilanjutkan dengan mengunduh hasil evaluasi dalam format yang telah ditentukan untuk keperluan dokumentasi dan pelaporan.

No	No CPL	CPL	No CPMK	CPMK	Hasil Person CPL	Hasil Person CPMK	Hasil Akhir	Komentar																						
1	7	Mampu melakukan asuhan keperawatan dalam berbagai tataran pelayanan keperawatan (KKU4)–P2	2	Merumuskan <i>Planning, Organizing, Staffing, Actuating and Controlling</i> (POSAC) dalam manajemen keperawatan dan mendemonstrasikan metode penugasan keperawatan sesuai MPKP.	<table border="1"> <thead> <tr> <th>Kognitif</th> <th>Afektif</th> <th>Psikomotor</th> </tr> </thead> <tbody> <tr> <td><input type="radio"/> C1</td> <td><input type="radio"/> A1</td> <td><input type="radio"/> P1</td> </tr> <tr> <td><input type="radio"/> C2</td> <td><input type="radio"/> A2</td> <td><input checked="" type="radio"/> P2</td> </tr> <tr> <td><input type="radio"/> C3</td> <td><input type="radio"/> A3</td> <td><input type="radio"/> P3</td> </tr> <tr> <td><input type="radio"/> C4</td> <td><input type="radio"/> A4</td> <td><input type="radio"/> P4</td> </tr> <tr> <td><input type="radio"/> C5</td> <td><input type="radio"/> A5</td> <td><input type="radio"/> P5</td> </tr> <tr> <td><input type="radio"/> C6</td> <td><input type="radio"/> A6</td> <td><input type="radio"/> P6</td> </tr> </tbody> </table>	Kognitif	Afektif	Psikomotor	<input type="radio"/> C1	<input type="radio"/> A1	<input type="radio"/> P1	<input type="radio"/> C2	<input type="radio"/> A2	<input checked="" type="radio"/> P2	<input type="radio"/> C3	<input type="radio"/> A3	<input type="radio"/> P3	<input type="radio"/> C4	<input type="radio"/> A4	<input type="radio"/> P4	<input type="radio"/> C5	<input type="radio"/> A5	<input type="radio"/> P5	<input type="radio"/> C6	<input type="radio"/> A6	<input type="radio"/> P6	<input type="radio"/> 0 <input checked="" type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input checked="" type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6	<input checked="" type="radio"/> Tidak Sesuai <input type="radio"/> Sesuai		
Kognitif	Afektif	Psikomotor																												
<input type="radio"/> C1	<input type="radio"/> A1	<input type="radio"/> P1																												
<input type="radio"/> C2	<input type="radio"/> A2	<input checked="" type="radio"/> P2																												
<input type="radio"/> C3	<input type="radio"/> A3	<input type="radio"/> P3																												
<input type="radio"/> C4	<input type="radio"/> A4	<input type="radio"/> P4																												
<input type="radio"/> C5	<input type="radio"/> A5	<input type="radio"/> P5																												
<input type="radio"/> C6	<input type="radio"/> A6	<input type="radio"/> P6																												
2	8	Mampu menerapkan dakwah Islami dan menjadi kader Muhammadiyah – 'Aisyiyah (KKU5) – P1	1	Menjelaskan konsep manajemen secara umum, manajemen kesehatan dan manajemen keperawatan dengan memperhatikan prinsip dan etik keperawatan dan nilai-nilai islami.	<table border="1"> <thead> <tr> <th>Kognitif</th> <th>Afektif</th> <th>Psikomotor</th> </tr> </thead> <tbody> <tr> <td><input type="radio"/> C1</td> <td><input type="radio"/> A1</td> <td><input checked="" type="radio"/> P1</td> </tr> <tr> <td><input type="radio"/> C2</td> <td><input type="radio"/> A2</td> <td><input type="radio"/> P2</td> </tr> <tr> <td><input type="radio"/> C3</td> <td><input type="radio"/> A3</td> <td><input type="radio"/> P3</td> </tr> <tr> <td><input type="radio"/> C4</td> <td><input type="radio"/> A4</td> <td><input type="radio"/> P4</td> </tr> <tr> <td><input type="radio"/> C5</td> <td><input type="radio"/> A5</td> <td><input type="radio"/> P5</td> </tr> <tr> <td><input type="radio"/> C6</td> <td><input type="radio"/> A6</td> <td><input type="radio"/> P6</td> </tr> </tbody> </table>	Kognitif	Afektif	Psikomotor	<input type="radio"/> C1	<input type="radio"/> A1	<input checked="" type="radio"/> P1	<input type="radio"/> C2	<input type="radio"/> A2	<input type="radio"/> P2	<input type="radio"/> C3	<input type="radio"/> A3	<input type="radio"/> P3	<input type="radio"/> C4	<input type="radio"/> A4	<input type="radio"/> P4	<input type="radio"/> C5	<input type="radio"/> A5	<input type="radio"/> P5	<input type="radio"/> C6	<input type="radio"/> A6	<input type="radio"/> P6	<input type="radio"/> 0 <input checked="" type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6	<input type="radio"/> Tidak Sesuai <input checked="" type="radio"/> Sesuai		
Kognitif	Afektif	Psikomotor																												
<input type="radio"/> C1	<input type="radio"/> A1	<input checked="" type="radio"/> P1																												
<input type="radio"/> C2	<input type="radio"/> A2	<input type="radio"/> P2																												
<input type="radio"/> C3	<input type="radio"/> A3	<input type="radio"/> P3																												
<input type="radio"/> C4	<input type="radio"/> A4	<input type="radio"/> P4																												
<input type="radio"/> C5	<input type="radio"/> A5	<input type="radio"/> P5																												
<input type="radio"/> C6	<input type="radio"/> A6	<input type="radio"/> P6																												

Gambar 3. Hasil perhitungan secara manual

3.2. Tahapan Perhitungan dan Analisis Data

Tahapan perhitungan dimulai dengan menentukan nilai ambang batas. Nilai ambang ini akan menjadi tolak ukur dalam proses pengambilan keputusan. Tabel 3. Merupakah hasil perhitungan nilai BERT dan SBERT pada pengukuran CPL dan CPMK.

Tabel 3. Nilai perhitungan BERT dan SBERT

Keterangan	Nilai BERT	Nilai SBERT
Maksimal	0,95	0,97
Minimal	0,62	0,09
Batas	0,79	0,53

Berdasarkan Tabel 3. Nilai tertinggi pada model BERT adalah 0,95 dan yang terendah adalah 0,62. Menentukan ambang batas yang sesuai, digunakan rata-rata dari kedua nilai tersebut 0,79. Nilai ini digunakan sebagai batas untuk membedakan apakah pasangan teks dianggap relevan atau tidak. Pada model SBERT, nilai maksimum adalah 0,97 dan nilai minimum 0,09. Memakai metode yang sama, diperoleh ambang batas sebesar 0,53. Penentuan ambang batas ini bertujuan untuk menyaring hasil prediksi model agar lebih akurat dalam mengidentifikasi hubungan semantik antar teks.

Tabel 4. Menunjukkan hasil evaluasi model BERT berdasarkan empat kategori klasifikasi: *True positive* (TP), *False negative* (FN), *False positive* (FP), dan *True negative* (TN). TP terjadi ketika model BERT dan penilaian manusia sama-sama menyatakan *sesuai*, dengan jumlah 414. FN terjadi ketika model menyatakan sesuai tetapi manusia menyatakan tidak sesuai, sebanyak 150 data. FP adalah ketika model menyatakan tidak sesuai namun manusia menyatakan sesuai, sebanyak 160 data. Sementara itu, TN terjadi saat kedua pihak menyatakan tidak sesuai, dengan jumlah 75. Nilai-nilai ini digunakan untuk menghitung performa model, seperti akurasi, presisi, *recall*, dan *F1-score*.

Tabel 4. *Confusion Matrix BERT*

		Manusia	
		Positif	Negatif
BERT	Positif	414	150
	Negatif	160	75

Tabel 5. Menyajikan hasil klasifikasi antara sistem berbasis SBERT dan penilaian manusia terhadap kesesuaian suatu entitas. Berdasarkan data tersebut, diperoleh nilai TP sebanyak 252, yang menunjukkan bahwa sistem SBERT berhasil mengidentifikasi kesesuaian dengan benar sesuai dengan penilaian manusia. Selanjutnya, terdapat FN sebanyak 83, yaitu kasus ketika SBERT menilai suatu entitas sesuai, namun penilaian manusia menyatakan tidak sesuai. Sebaliknya, terdapat FP sebanyak 322, di mana SBERT menilai entitas tidak sesuai padahal menurut penilaian manusia sebenarnya sesuai. Terakhir, nilai TN sebesar 142 menunjukkan bahwa SBERT dan manusia sama-sama menilai entitas sebagai tidak sesuai. Data ini mencerminkan performa sistem SBERT dalam mengklasifikasikan kesesuaian entitas dan menjadi dasar evaluasi metrik performa seperti akurasi, presisi, *recall*, dan F1-score.

Tabel 5. *Confusion Matrix SBERT*

		Manusia	
		Positif	Negatif
SBERT	Positif	252	83
	Negatif	322	152

Table 6 merangkum hasil evaluasi yang menunjukkan bahwa model BERT secara umum memiliki kinerja yang lebih baik dibandingkan SBERT dalam konteks tugas klasifikasi teks yang diujikan. Hal ini terlihat dari nilai Akurasi, *Precision*, dan F1 Score yang lebih tinggi pada BERT. Akurasi sebesar 0,61 pada BERT menunjukkan kemampuan model dalam mengklasifikasikan data secara keseluruhan dengan lebih tepat dibandingkan SBERT yang hanya mencapai 0,49. *Precision* BERT yang mencapai 0,72 juga menunjukkan bahwa model ini lebih handal dalam menghasilkan prediksi positif yang benar, yang penting dalam konteks di mana kesalahan positif palsu harus diminimalkan. F1 Score BERT yang lebih tinggi 0,73 juga memperkuat bahwa model ini memiliki keseimbangan yang baik antara *Precision* dan *Recall*. Namun demikian, SBERT menunjukkan keunggulan pada metrik *Recall* 0,75, mengindikasikan kemampuannya dalam mendeteksi lebih banyak instance relevan atau positif, meskipun dengan tingkat *Precision* yang lebih rendah. Ini menunjukkan bahwa SBERT dapat menjadi pilihan yang lebih sesuai pada skenario yang lebih menekankan pada minimisasi *false negative*, seperti dalam deteksi anomalai atau sistem peringatan dini.

Tabel 6. Hasil perbandingan metrik antara nilai BERT dan SBERT.

Metrik	Nilai BERT	Nilai SBERT
Akurasi	0,61	0,49
Precision	0,72	0,44
Recall	0,73	0,75
F1 Score	0,73	0,55

Perbandingan metrik kinerja model BERT dan SBERT yang diperlihatkan pada tabel 6 menunjukkan perbedaan yang cukup signifikan kinerja kedua model dalam menilai kesesuaian antara capaian pembelajaran lulusan dan capaian pembelajaran mata kuliah. Perbedaan diduga disebabkan oleh arsitektur dan tujuan desain oleh masing-masing model. BERT digunakan untuk optimalisasi pengukuran kesamaan semantik antar kata sedangkan SBERT digunakan untuk optimalisasi pengukuran kesamaan semantik antar kalimat. Kesesuaian antara capaian pembelajaran lebih dititikberatkan pada penggunaan kata kerja yang tercantum dalam pernyataan capaian pembelajaran. Sebagai contoh terdapat capaian pembelajaran lulusan “CPL1. mampu melakukan skrining penyakit secara dini” dan capaian pembelajaran matakuliah “CPMK1. mampu mengembangkan prosedur skrining penyakit”. Secara kalimat, kedua pernyataan (CPL1 dan CPMK1) mengandung banyak kata yang sama namun karena kata kerja yang digunakan berada pada level kognitif yang berbeda maka kedua pernyataan itu tidak selaras. Sedangkan pernyataan “CPMK2. Mampu melaksanakan prosedur skrining penyakit jantung bawaan pada bayi” selaras dengan CPL1 karena menggunakan kata kerja yang setara. Titik berat kesesuaian capaian pembelajaran

diduga bukan karena makna semantik dari kalimat sehingga penggunaan model SBERT dalam pengujian keselarasan capaian pembelajaran tidak lebih baik dibanding model BERT.

4. KESIMPULAN

Hasil evaluasi, dapat disimpulkan bahwa model BERT memiliki kinerja yang lebih unggul dibandingkan SBERT dalam tugas klasifikasi teks pada konteks evaluasi pembelajaran berbasis OBE, dengan nilai Akurasi, *Precision*, dan F1 Score yang lebih tinggi, masing-masing sebesar 0,61, 0,72, dan 0,73. Hal ini menunjukkan bahwa BERT lebih efektif dalam menghasilkan prediksi yang akurat dan seimbang, sehingga lebih sesuai untuk mengidentifikasi kesesuaian antara capaian pembelajaran matakuliah dan capaian pembelajaran lulusan. Meskipun demikian, SBERT menunjukkan keunggulan pada metrik *Recall* (0,75), yang menjadikannya lebih sensitif dalam mendekripsi capaian pembelajaran yang relevan, terutama dalam skenario yang menuntut minimisasi kesalahan tipe dua. Temuan ini menjadikan BERT sebagai model yang lebih cocok digunakan untuk mengukur keselarasan antara CPL dengan CPMK.

DAFTAR PUSTAKA

- [1] I. arum tri Rahayu, L. Hidayati, I. Nahari, E. E. Saputra, and N. A. P. Reztanty, “Analisa Ketercapaian Program Learning Outcomes Berbasis Obe Pada Program Studi S-1 Pendidikan Tata Busana Fakultas Teknik Universitas Negeri Surabaya,” *J. Vocat. Tech. Educ.*, vol. 4, no. 2, pp. 33–43, 2022, doi: 10.26740/jvte.v4n2.p33-43.
- [2] M. I. Muzakir and Susanto, “Implementasi Kurikulum Outcome Based Education (Obe) Dalam Sistem Pendidikan Tinggi Di Era Revolusi Industri 4.0 ((Implementation of the Outcome Based Education (OBE) Curriculum in the Higher Education System in the Era of Industrial Revolution 4.0),” *Edukasiana J. Islam. Educ.*, vol. 2, no. 1, pp. 118–139, 2023.
- [3] M. V. Panichkina, I. A. Sinyavskaya, and E. V. Shestova, “Challenges of Professional Adaptation of University Graduates in Response to the Economics’ Digital Transformation,” in *2018 XVII Russian Scientific and Practical Conference on Planning and Teaching Engineering Staff for the Industrial and Economic Complex of the Region (PTES)*, IEEE, Nov. 2018, pp. 44–46. doi: 10.1109/PTES.2018.8604207.
- [4] S. Setiono, S. Windyariani, and A. Juhanda, “Implementasi Sistem Penilaian Berbasis Oucome Based Education di Perguruan Tinggi,” *J. Pendidik.*, vol. 11, no. 1, pp. 1–9, 2023, doi: 10.36232/pendidikan.v11i1.2617.
- [5] A. H. A. Rasyid, B. Yunitasari, I. W. Susila, D. Dewanto, Y. Yunus, and D. I. Santoso, “Pengembangan Model Evaluasi Pembelajaran Berbasis Obe,” *J. Pendidik. (Teori dan Prakt.)*, vol. 7, no. 1, pp. 8–17, 2022, doi: 10.26740/jp.v7n1.p8-17.
- [6] U. M. Ishaq, M. F. Wicaksono, and S. Nurhayati, “Aplikasi Probe Untuk Penilaian Capaian Pembelajaran Mahasiswa Pada Kurikulum OBE (Outcame-Based Education),” *Komputika J. Sist. Komput.*, vol. 12, no. 2, pp. 67–74, 2023, doi: 10.34010/komputika.v12i2.9763.
- [7] S. Masruroh and S. Shofia Hilabi, “Implementasi Pendidikan Agama Islam Berbasis Outcame Based Education (OBE) di Perguruan Tinggi Umum,” *J. Dirosah Islam.*, vol. 4, no. 3, pp. 374–381, 2022, doi: 10.47467/jdi.v4i3.1263.
- [8] M. Dzulkifli, A. Samsudin, and A. S. Budi, “Analisis Tingkat Kompetensi Mahasiswa Terhadap Capaian Pembelajaran Mata Kuliah Pada Dacum Chart Program Studi Teknologi Industri Pangan,” *J. Jendela Pendidik.*, vol. 2, no. 03, pp. 347–357, 2022, doi: 10.57008/jjp.v2i03.253.
- [9] A. L. Ramdani, D. H. Widyatgoro, and R. Munir, “Optimalisasi Rekomendasi Rute Pada Perencanaan Perjalanan Wisata: Studi Pustaka,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 515–525, 2024, doi: 10.57152/malcom.v4i2.1213.
- [10] M. F. Ramadhan, B. Siswoyo, and S. I. Cirebon, “Mengenal Model BERT dan Implementasinya untuk Analisis Sentimen Ulasan Game,” pp. 395–398.
- [11] D. L. Luhurkinanti, P. D. Purnamasari, T. Tsunakawa, and A. A. P. Ratna, “Japanese Short Answer Grading for Japanese Language Learners Using the Contextual Representation of BERT,” *IEEE Access*, vol. 13, no. January, pp. 17195–17207, 2025, doi: 10.1109/ACCESS.2025.3532659.
- [12] G. Wiratmoko, H. Thamrin, and W. E. Pamungkas, “Performance of Machine Learning Algorithms on Automatic Summarization of Indonesian Language Texts,” *J. Online Inform.*, vol. 10, no. 1, pp. 196–204, 2025, doi: 10.15575/join.v10i1.1506.

- [13] H. Thamrin and A. Sabardilla, “Efektivitas Algoritma Semantik dengan Keterkaitan Kata dalam Mengukur Kemiripan Teks Bahasa Indonesia,” *Khazanah Inform.*, vol. 1, no. 1, pp. 7–11, 2021.
- [14] H. Thamrin, D. Oktafiani, I. I. Rasyid, and I. M. Fauzi, “Classification of SWOT Statements Employing BERT Pre-Trained Model Embedding,” *J. Sist. Inf. Bisnis*, vol. 14, no. 2, pp. 143–152, 2024, doi: 10.21456/vol14iss2pp143-152.
- [15] M. Mokoatle, V. Marivate, D. Mapiye, R. Bornman, and V. M. Hayes, “A review and comparative study of cancer detection using machine learning: SBERT and SimCSE application,” *BMC Bioinformatics*, vol. 24, no. 1, pp. 1–25, 2023, doi: 10.1186/s12859-023-05235-x.
- [16] P. Lombaers, J. de Bruin, and R. van de Schoot, “Reproducibility and Data Storage for Active Learning-Aided Systematic Reviews,” *Appl. Sci.*, vol. 14, no. 9, 2024, doi: 10.3390/app14093842.
- [17] J. J. Teijema *et al.*, “Active learning-based systematic reviewing using switching classification models: the case of the onset, maintenance, and relapse of depressive disorders,” *Front. Res. Metrics Anal.*, vol. 8, 2023, doi: 10.3389/frma.2023.1178181.
- [18] H. H. Mohammed, E. Dogdu, A. K. Gorur, and R. Choupani, “Multi-Label Classification of Text Documents Using Deep Learning,” in *2020 IEEE International Conference on Big Data (Big Data)*, 2020, pp. 4681–4689. doi: 10.1109/BigData50022.2020.9378266.
- [19] A. S. Harahap, M. Irfan, and K. Usman, “Outcome-Based Education pada Mata Kuliah Kewirausahaan Olahraga Program Studi Pendidikan Jasmani Kesehatan dan Rekreasi,” *J. Phys. Heal. Recreat.*, vol. 4, no. 1, pp. 15–24, 2023, [Online]. Available: <http://www.jurnal.stokbinaguna.ac.id/index.php/JPHR/article/view/1652>
- [20] H. Halabieh *et al.*, “The Future of Higher Education: Identifying Current Educational Problems and Proposed Solutions,” *Educ. Sci.*, vol. 12, no. 12, 2022, doi: 10.3390/educsci12120888.
- [21] Ulfah and O. Arifudin, “Analisis Teori Taksonomi Bloom pada Pendidikan di Indonesia,” *J. Al-Amar*, vol. 4, no. 1, pp. 13–22, 2023.
- [22] N. Khalishah and N. Iklilah, “Taksonomi Bloom (Revisi): Tujuan Pendidikan dan Implementasinya dalam Pembelajaran Matematika,” *Pros. Semin. Nas. Tadris Mat.*, pp. 248–266, 2021.
- [23] A. Sujatha Ravindran and J. Contreras-Vidal, “An empirical comparison of deep learning explainability approaches for EEG using simulated ground truth,” *Sci. Rep.*, vol. 13, no. 1, pp. 1–20, 2023, doi: 10.1038/s41598-023-43871-8.
- [24] A. A. Salih and A. M. Abdulazeez, “Evaluation of Classification Algorithms for Intrusion Detection System: A Review,” *J. Soft Comput. Data Min.*, vol. 2, no. 1, pp. 31–40, 2021, doi: 10.30880/jscdm.2021.02.01.004.
- [25] M. Owusu-Adjei, J. Ben Hayfron-Acquah, T. Frimpong, and G. Abdul-Salaam, “Imbalanced class distribution and performance evaluation metrics: A systematic review of prediction accuracy for determining model performance in healthcare systems,” *PLOS Digit. Heal.*, vol. 2, no. 11 November, pp. 1–19, 2023, doi: 10.1371/journal.pdig.0000290.