

LEARN MAP: Kerangka Kecerdasan Buatan untuk Klasifikasi Aktivitas Belajar Mahasiswa dan Pemetaan Potensi Akademik

Ita Margaretta Br Tarigan^{1*}, Siti Jamilah Br Tarigan², Raheliya Br Ginting³

^{1*,3} Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Institut Teknologi dan Bisnis Indonesia, Deli Serdang, Indonesia

² Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Institut Teknologi dan Bisnis Indonesia, Deli Serdang, Indonesia

^{1*}itamargaretta1997@gmail.com, ²jamilah.tarigan89@gmail.com, ³raheliyabrginting@gmail.com

^{*)} itamargaretta1997@gmail.com

Abstrak-Transformasi digital dalam pendidikan tinggi menghasilkan data aktivitas mahasiswa dalam jumlah besar yang sebagian besar tersimpan dalam bentuk teks tidak terstruktur. Data tersebut berpotensi dimanfaatkan untuk memahami pola belajar, kompetensi, dan potensi akademik mahasiswa, namun pengelolaannya masih menghadapi kendala karena metode konvensional umumnya hanya mengukur kemiripan kata tanpa memahami konteks aktivitas. Penelitian ini bertujuan mengembangkan LEARN MAP, yaitu kerangka kecerdasan buatan berbasis *Natural Language Processing* (NLP) dan *Bidirectional Encoder Representations from Transformers* (BERT) untuk mengklasifikasikan aktivitas belajar mahasiswa serta memetakan potensi akademik secara otomatis. Penelitian menggunakan dataset sebanyak 8.954 aktivitas mahasiswa yang berasal dari kegiatan pembelajaran dan MBKM. Tahapan penelitian meliputi pengumpulan data, preprocessing teks, fine-tuning model BERT, klasifikasi aktivitas, pemetaan soft skill, dan analisis potensi akademik. Hasil penelitian menunjukkan bahwa model BERT mampu mengklasifikasikan aktivitas mahasiswa dengan nilai *accuracy* 92,31%, *precision* 91,27%, *recall* 90,84%, dan *F1-score* 91,05%. Aktivitas mahasiswa didominasi oleh kategori komunikasi, pengabdian masyarakat, kolaborasi, dan penelitian. Hasil pemetaan menunjukkan bahwa kemampuan komunikasi dan kemandirian merupakan soft skill yang paling berkembang, sementara mayoritas mahasiswa berada pada kategori potensi akademik sedang hingga tinggi. Kebaruan penelitian terletak pada integrasi klasifikasi aktivitas berbasis BERT dengan pemetaan soft skill dan potensi akademik dalam satu kerangka analitik. LEARN MAP berpotensi mendukung pengambilan keputusan akademik berbasis data serta transformasi digital pendidikan tinggi.

Kata Kunci: BERT, Natural Language Processing, Student Activity Classification, Academic Potential Mapping, Learning Analytics.

Abstract-The digital transformation of higher education has generated large volumes of student activity data, most of which are stored as unstructured text. These data have significant potential to provide insights into learning patterns, competencies, and academic potential; however, conventional approaches often fail to capture contextual meaning and semantic relationships within student activity records. This study proposes LEARN MAP (Learning Activity and Academic Potential Mapping), an artificial intelligence framework based on Natural Language Processing (NLP) and Bidirectional Encoder Representations from Transformers (BERT) to automatically classify student learning activities and map academic potential. The study utilized a dataset of 8,954 student activity records collected from learning and Independent Learning–Independent Campus (MBKM) programs. The research process consisted of data collection, text preprocessing, BERT fine-tuning, activity classification, soft-skill mapping, and academic potential analysis. The results demonstrate that the BERT model achieved an accuracy of 92.31%, precision of 91.27%, recall of 90.84%, and F1-score of 91.05% in classifying student activities. The findings indicate that student activities were predominantly associated with communication, community engagement, collaboration, and research. Furthermore, the soft-skill mapping revealed that communication and self-directed learning were the most developed competencies, while most students were categorized as having moderate to high academic potential. The novelty of this study lies in the integration of BERT-based activity classification with soft-skill and academic potential mapping within a unified analytical framework. LEARN MAP has the potential to support data-driven academic decision-making and advance digital transformation initiatives in higher education.

Keywords: BERT, Natural Language Processing, Student Activity Classification, Academic Potential Mapping, Learning Analytics.

1. PENDAHULUAN

Transformasi digital pada pendidikan tinggi telah mendorong perubahan signifikan dalam cara mahasiswa belajar dan berinteraksi dengan lingkungan akademik. Aktivitas belajar mahasiswa tidak lagi terbatas pada ruang kelas formal, melainkan berlangsung secara luas melalui platform digital seperti *Learning Management System* (LMS), forum diskusi daring, laporan aktivitas, refleksi pembelajaran, dan berbagai bentuk interaksi akademik di luar kelas [1], [2]. Kegiatan pembelajaran di luar kelas seperti magang, PMM, asistensi mengajar, proyek kemanusiaan, wirausaha, studi independen, dan KKNT dapat berjalan dengan baik untuk peningkatan SDM [3]. Kondisi ini menghasilkan data pembelajaran dalam jumlah besar, terutama dalam bentuk teks tidak terstruktur, yang sejatinya menyimpan informasi penting mengenai pola belajar, keterlibatan, dan potensi akademik mahasiswa [4].

Namun, sebagian besar perguruan tinggi belum mampu memanfaatkan data aktivitas belajar tersebut secara optimal [5]. Pemantauan akademik umumnya masih berfokus pada indikator konvensional; akibatnya, intervensi akademik sering terlambat dan tidak tepat sasaran sehingga berdampak pada mutu capaian pendidikan tinggi secara keseluruhan. Permasalahan ini menjadi semakin krusial dalam konteks pembangunan nasional jangka panjang. Asta Cita Presiden dalam RPJMN 2025–2029, khususnya poin keempat, menegaskan pentingnya pembangunan Sumber Daya Manusia (SDM), sains, teknologi, dan pendidikan sebagai fondasi daya saing bangsa.

Untuk mewujudkan SDM unggul dan adaptif, perguruan tinggi dituntut memiliki sistem yang mampu memetakan potensi akademik mahasiswa secara akurat dan berkelanjutan. Hal ini juga selaras dengan RIRN 2027–2045 yang menempatkan pengembangan kecerdasan buatan dan riset pendidikan adaptif sebagai topik strategis nasional. Penelitian sebelumnya dalam bidang monitoring aktivitas belajar mahasiswa umumnya menggunakan pendekatan statistik tradisional atau metode text mining dasar seperti TF-IDF, *Cosine Similarity*, dan Word2Vec [6],[7]. Pendekatan tersebut hanya mengukur kemiripan kata secara permukaan tanpa memahami konteks, relasi semantik, dan makna aktivitas secara utuh. Akibatnya, tingkat akurasi klasifikasi aktivitas relatif rendah; banyak aktivitas mahasiswa yang bersifat kompleks dan kontekstual tidak teridentifikasi dengan tepat, serta pemetaan potensi akademik menjadi kurang akurat dan sulit dijadikan dasar pengambilan keputusan akademik.

Perkembangan kecerdasan buatan, khususnya dalam bidang *Natural Language Processing* (NLP) berbasis deep learning, membuka peluang untuk mengatasi keterbatasan tersebut [8],[9]. Model berbasis Transformer seperti BERT (*Bidirectional Encoder Representations from Transformers*) memiliki kemampuan unggul dalam memahami konteks bahasa alami secara dua arah, sehingga mampu menangkap makna, intensi, dan pola aktivitas belajar mahasiswa secara lebih mendalam [10],[11]. Pemanfaatan BERT dalam analisis aktivitas belajar memungkinkan klasifikasi yang lebih presisi serta pemetaan potensi akademik yang komprehensif. Sejumlah studi terbaru membuktikan bahwa BERT unggul dalam klasifikasi teks pendidikan, analisis keterlibatan mahasiswa, dan prediksi performa akademik [12], [13], [14]. Dengan demikian, BERT dapat menangkap dan mengkategorikan sebuah kata dalam konteks dan tingkat tertentu sesuai dengan struktur modelnya [15]. Tokenisasi merupakan pembagian kalimat menjadi bagian yang lebih kecil pada BERT yang disebut "sub-word" atau "token-pieces" [16], [17], [18]. Tokenisasi di BERT mencakup token awal dan token yang dibuat khusus, seperti [CLS] (Klasifikasi) dan [SEP] (Pemisah), dan lain-lain [19], [20]. Prosedur ini dilakukan pada tingkat kata dan karakter individual, memungkinkan model untuk memahami keadaan di sekitarnya dan hubungan antar keadaan tersebut [21]. Metode ini membantu BERT mengatasi kata-kata yang sulit atau morfologi yang berbeda [22]. Metode BERT sudah digunakan di dalam ekstraksi jaringan saraf kerangka manusia [23], media sosial [24], [25], [26], [27], layanan akademik-nonakademik [28]. Berdasarkan permasalahan di atas, peneliti mengusulkan LEARN MAP, yaitu kerangka kecerdasan buatan berbasis NLP dan BERT untuk memahami pola aktivitas belajar mahasiswa secara kontekstual. LEARN MAP dirancang untuk mengklasifikasi aktivitas belajar mahasiswa secara akurat dan memetakan potensi akademik berbasis representasi semantik, sehingga perguruan tinggi dapat melakukan prediksi risiko, asesmen adaptif, dan rekomendasi pembelajaran secara tepat waktu.

Dengan demikian, penelitian ini tidak hanya berkontribusi pada pengembangan keilmuan di bidang kecerdasan buatan dan pendidikan, tetapi juga mendukung kebijakan nasional dalam penguatan SDM unggul, transformasi digital pendidikan tinggi, dan pemanfaatan data pembelajaran berbasis AI secara berkelanjutan. Penelitian ini bertujuan untuk menjawab kebutuhan strategis pendidikan tinggi dalam memanfaatkan AI guna memahami aktivitas belajar mahasiswa secara kontekstual. LEARN MAP mendukung Asta Cita dan RIRN melalui peningkatan mutu pembelajaran, intervensi akademik tepat sasaran, serta penguatan kualitas SDM dan pembangunan nasional berbasis data. Novelty penelitian ini menghadirkan kebaruan melalui pengembangan LEARN MAP, kerangka kecerdasan buatan berbasis BERT dan NLP yang mampu memahami konteks, relasi semantik, dan makna aktivitas belajar mahasiswa secara komprehensif. Berbeda dengan pendekatan TF-IDF dan Word2Vec yang bersifat leksikal, LEARN MAP menganalisis laporan aktivitas belajar di luar kelas secara kontekstual untuk mengklasifikasikan aktivitas kompleks dan memetakan potensi akademik secara presisi.

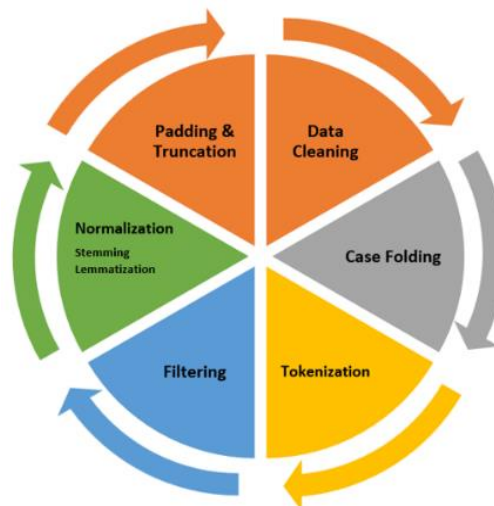
Kebaruan juga terletak pada integrasi klasifikasi aktivitas, analisis representasi pembelajaran, dan rekomendasi intervensi akademik adaptif berbasis data pendidikan.

2. METODOLOGI PENELITIAN

Metode penelitian ini menggunakan pendekatan machine learning berbasis *Natural Language Processing* (NLP) dengan model BERT (*Bidirectional Encoder Representations from Transformers*) sebagai inti kerangka LEARN MAP. Penelitian diawali dengan pengumpulan data aktivitas belajar mahasiswa, khususnya laporan dan narasi aktivitas belajar di luar kelas. Data kemudian melalui tahap praproses teks dengan NLP meliputi pembersihan teks, tokenisasi, normalisasi, dan penghapusan kata tidak bermakna untuk memastikan kualitas dan konsistensi data [29]. Selanjutnya, dilakukan fine-tuning model BERT untuk mengklasifikasi aktivitas belajar secara kontekstual. Hasil representasi semantik dari BERT dimanfaatkan untuk pemetaan potensi akademik mahasiswa, yang menggambarkan kecenderungan belajar, tingkat keterlibatan, dan risiko akademik mahasiswa. Kinerja sistem dapat dijamin secara objektif dengan mengevaluasi kinerja model menggunakan ukuran skor F1, yang meliputi akurasi, presisi, recall, dan waktu eksekusi. Berikut ini beberapa tahap penelitian yang dilakukan secara lebih terperinci sebagai berikut:

2.1 *Natural Language Processing* (NLP)

Gambar 1 merupakan tahap preprocessing data yang merupakan langkah krusial dalam penelitian berbasis *Natural Language Processing* (NLP), karena kualitas data teks sangat menentukan kinerja model kecerdasan buatan yang digunakan [30], [31]. Dalam konteks penelitian LEARN MAP, preprocessing dilakukan untuk memastikan bahwa data aktivitas belajar mahasiswa yang berasal dari laporan kegiatan, refleksi pembelajaran, forum diskusi, dan log aktivitas digital dapat diproses secara optimal oleh model berbasis BERT dengan tahapan sebagai berikut:



Gambar 1. Tahapan Preprocessing Data

2.2 Ekstraksi fitur dengan BERT (*Bidirectional Encoder Representations from Transformers*)

Klasifikasi aktivitas belajar secara kontekstual dilakukan dengan menerapkan model BERT (*Bidirectional Encoder Representations from Transformers*), yaitu model bahasa berbasis transformer yang dikembangkan oleh Google. BERT digunakan untuk berbagai tugas menemukan kembali informasi (*information retrieval*), khususnya dalam hal pencarian informasi untuk memahami konteks kata dalam kalimat dengan cara yang lebih mendalam. Ekstraksi fitur dengan BERT adalah sebagai berikut [32], [33]:

- Tokenisasi (WordPiece Tokenization)
Teks aktivitas belajar diubah menjadi token:
 $X = [x_1, x_2, \dots, x_n]$
Setiap token dipetakan ke indeks kosakata:
 $t_i = V(X_i)$
- Tahap Encoding (Self-Attention Transformer)



Masukan berupa embedding kata, posisi, dan segmentasi:

$$E = X + P + S$$

Self-Attention Mechanism (Mekanisme menghitung antar kata didalam memahami konteks kalimat secara penuh)

$$Q = EWQ, \quad K = EWK, \quad V = EWV$$

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{Q \cdot K^T}{\sqrt{d_k}} \right) V$$

c. Tahap *Fine-tuning* untuk Klasifikasi Aktivitas Belajars

Fine-tuning dilakukan dengan melatih ulang parameter BERT menggunakan data aktivitas belajar mahasiswa yang telah dilabeli. Proses ini memungkinkan model memahami konteks spesifik pendidikan tinggi, sehingga klasifikasi aktivitas belajar dan pemetaan potensi akademik menjadi lebih akurat dan adaptif.

Vektor final token [CLS] digunakan: $h_{CLS} = \text{BERT}(X)$

Kemudian melewati layer klasifikasi: $y = \text{softmax}(Wh_{CLS} + b)$

2.3. Pemetaan Potensi Akademik

Tahap pemetaan potensi akademik dilakukan dengan memanfaatkan representasi fitur hasil ekstraksi model BERT dari data aktivitas belajar mahasiswa. Representasi ini menggambarkan pola keterlibatan, konsistensi belajar, kompleksitas aktivitas, dan konteks akademik mahasiswa di luar kelas. Selanjutnya, dilakukan analisis agregasi dan pengelompokan untuk mengidentifikasi profil potensi akademik mahasiswa, seperti potensi unggul, berkembang, atau berisiko, yang dapat digunakan sebagai dasar intervensi akademik dan rekomendasi pembelajaran adaptif.

Potensi mahasiswa dihitung dari representasi semantik aktivitas:

$$Z = f(h_{CLS})$$

Contoh formula pemetaan (mis. skor potensi akademik komposit):

$$S = \alpha C + \beta A + \gamma E$$

2.4. Tahap Evaluasi Model

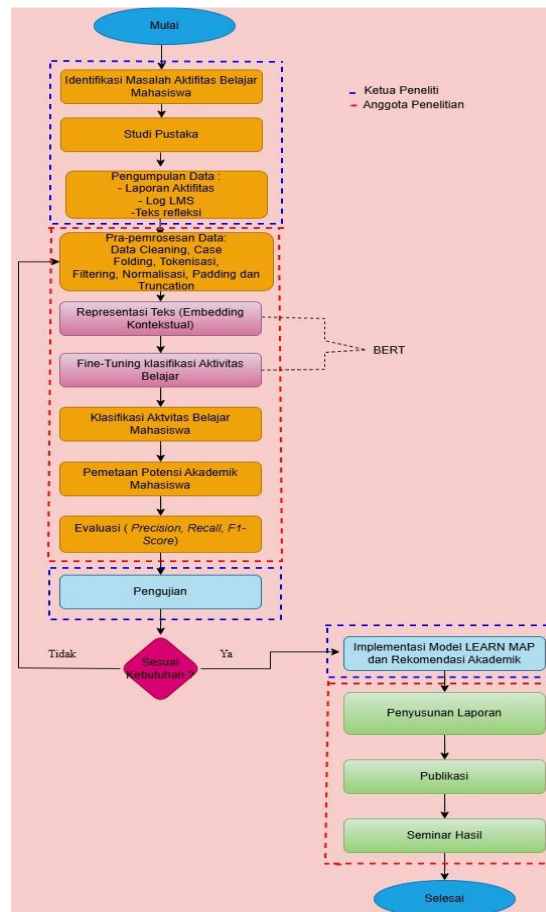
Tahap evaluasi bertujuan untuk mengukur kinerja dan keandalan model LEARN MAP. Evaluasi dilakukan menggunakan metrik standar machine learning, yaitu accuracy, precision, recall, dan F1-score pada proses klasifikasi aktivitas belajar. Selain itu, dilakukan evaluasi validitas pemetaan potensi akademik dengan membandingkan hasil model terhadap data capaian akademik aktual dan umpan balik dosen pembimbing. Bertujuan memastikan model yang dikembangkan bersifat akurat, adaptif, dan layak diimplementasikan dalam mendukung peningkatan mutu pendidikan tinggi.

Evaluasi model LEARN MAP dilakukan dengan matriks berikut:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad \text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}, \quad \text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

2.5 . Diagram Alir Penelitian

Diagram alir penelitian yang dilaksanakan dapat dilihat pada Gambar 2.



Gambar 2. Diagram Alir Penelitian

3. HASIL DAN PEMBAHASAN

3.1 Deskripsi Dataset Aktivitas Mahasiswa

Dataset yang digunakan dalam penelitian ini merupakan data aktivitas mahasiswa yang berasal dari implementasi Praktik Kerja Lapangan (PKL) di ITB Indonesia. Data dikumpulkan dari laporan aktivitas harian mahasiswa selama mengikuti kegiatan pembelajaran di luar kampus yang terdokumentasi pada sistem pelaporan akademik. Dataset ini dipilih karena mampu merepresentasikan aktivitas nyata mahasiswa dalam berbagai konteks pembelajaran, baik yang berkaitan dengan pengembangan kompetensi akademik maupun kompetensi nonakademik.

Secara keseluruhan, dataset terdiri atas 8.954 aktivitas mahasiswa yang dihasilkan oleh 284 mahasiswa dan tersebar pada 122 lokasi pelaksanaan PKL. Setiap aktivitas disimpan dalam bentuk teks naratif yang ditulis langsung oleh mahasiswa untuk menjelaskan kegiatan yang telah dilakukan. Karakteristik data tersebut menyebabkan dataset termasuk dalam kategori *unstructured text* data atau data teks tidak terstruktur, sehingga memerlukan pendekatan *Natural Language Processing* (NLP) untuk proses analisis.

Aktivitas yang tercatat dalam dataset mencakup berbagai bidang kegiatan, seperti pengajaran, pendataan masyarakat, observasi lapangan, penelitian, sosialisasi program, pelayanan administrasi, pengembangan masyarakat, kegiatan organisasi, hingga aktivitas pendukung lainnya. Variasi aktivitas yang cukup beragam memungkinkan penelitian ini memperoleh gambaran yang lebih komprehensif mengenai pola keterlibatan mahasiswa selama mengikuti program PKL.

Sebelum digunakan dalam proses pelatihan model, dilakukan analisis deskriptif terhadap dataset untuk mengetahui karakteristik data yang akan diolah. Analisis menunjukkan bahwa panjang dokumen aktivitas bervariasi antara satu hingga beberapa kalimat dengan rata-rata 10–25 kata per aktivitas. Perbedaan panjang teks tersebut mencerminkan variasi tingkat detail pelaporan mahasiswa dalam mendokumentasikan aktivitas yang dilakukan.

3.2 Hasil Preprocessing Data

Praproses diperlukan sebelum model BERT dapat diterapkan pada data aktivitas mahasiswa yang diperoleh, yang masih dalam format teks tidak terstruktur. Dengan membersihkan data dari informasi yang tidak relevan, menstandarisasi format teks, dan menciptakan representasi data yang lebih baik untuk klasifikasi, tahap praproses berupaya meningkatkan kualitas data. Karena korelasi yang tinggi antara kualitas data masukan dan kinerja model keluaran, langkah ini sangat penting.

Karena saat ini masih dalam format teks tidak terstruktur, data program magang tentang aktivitas mahasiswa tidak dapat langsung digunakan untuk tujuan klasifikasi. Efektivitas model klasifikasi dapat dipengaruhi oleh berbagai properti data, yang meliputi kapitalisasi yang tidak teratur, tanda baca, akronim, istilah umum (stopwords), dan variasi ejaan. Oleh karena itu, sebelum digunakan dalam model BERT, tahap praproses dilakukan untuk meningkatkan kualitas data dan menghasilkan representasi teks yang lebih akurat.

Pembersihan data, pengubahan huruf besar/kecil, tokenisasi, penghapusan stopwords, dan normalisasi adalah langkah-langkah praproses dalam penelitian ini. Untuk membuat data lebih mudah dipahami oleh model NLP, langkah-langkah ini bertujuan untuk menghilangkan informasi yang tidak relevan, menstandarisasi format teks, dan mempersiapkan data.

3.2.1 Data Cleaning

Tahap data *cleaning* dilakukan untuk menghilangkan spasi, tanda baca, simbol khusus, dan angka yang tidak diperlukan dari teks. Proses ini bertujuan mengurangi noise pada data sehingga informasi yang digunakan dalam proses klasifikasi menjadi lebih fokus pada aktivitas yang dilaporkan mahasiswa. Hasil contoh data *cleaning* dapat dilihat pada tabel 1 berikut.

Tabel 1. Hasil Data *Cleaning*

Data Asli	Hasil <i>Cleaning</i>
Menginput data calon pengantin!!!	Menginput data calon pengantin
Pendataan warga RT 03/04.	Pendataan warga RT

3.2.2 Case Folding

Tahap *case folding* dilakukan dengan mengubah seluruh huruf menjadi huruf kecil (*lowercase*). Tujuan proses ini adalah menghindari perbedaan representasi kata yang sebenarnya memiliki makna yang sama.

Contoh:

Melakukan Pendataan Warga Desa

menjadi

melakukan pendataan warga desa

Dengan proses ini, kata "Pendataan" dan "pendataan" akan dianggap sebagai kata yang sama oleh sistem.

3.2.3 Tokenization

Tahap tokenisasi bertujuan memecah kalimat menjadi unit-unit kata (tokens) sehingga dapat diproses lebih lanjut oleh model BERT. Setiap kata yang terkandung dalam kalimat dipisahkan menjadi token individual.

Contoh:

melakukan pendataan warga desa

Hasil tokenisasi:

["melakukan", "pendataan", "warga", "desa"]

Token yang dihasilkan akan digunakan untuk membentuk representasi teks pada tahap berikutnya.

3.2.4 Stopword Removal

Tahap *stopword removal* dilakukan untuk mengecualikan kata-kata yang sering digunakan tetapi tidak secara signifikan membantu dalam prosedur kategorisasi. Kata seperti dan, yang, di, ke, dari, untuk, dan dengan termasuk dalam daftar *stopwords* bahasa Indonesia.

Contoh:

melakukan pendataan warga yang berada di desa

menjadi

pendataan warga berada desa

Penghapusan stopwords membantu sistem lebih fokus pada kata-kata yang memiliki informasi penting terkait aktivitas mahasiswa.

3.2.5 Normalization

Tahap normalisasi dilakukan untuk menyeragamkan bentuk kata yang memiliki makna yang sama namun ditulis dengan format berbeda. Selain itu, proses ini juga digunakan untuk memperbaiki singkatan atau istilah tidak baku yang sering ditemukan pada laporan aktivitas mahasiswa. Berikut ini contoh hasil *Normalization* dapat dilihat pada tabel 2.

Tabel 2. Contoh Hasil *Normalization*

Sebelum	Sesudah
mhs	mahasiswa
keg	kegiatan
observasi lap	observasi lapangan

Normalisasi membantu meningkatkan konsistensi data sehingga hubungan semantik antarkata dapat dipahami lebih baik oleh model.

Setelah seluruh tahapan dilakukan, diperoleh data hasil akhir *processing* yang lebih bersih dan terstruktur. Contoh transformasi data pada setiap tahapan ditunjukkan pada Tabel 3 berikut.

Tabel 3. Hasil Transformasi *Preprocessing*

Tahapan	Hasil
<i>Data Asli</i>	Melakukan pendataan warga yang berada di Desa Suka Maju
<i>Cleaning</i>	Melakukan pendataan warga yang berada di Desa Suka Maju
<i>Case Folding</i>	melakukan pendataan warga yang berada di desa suka maju
<i>Tokenization</i>	melakukan, pendataan, warga, yang, berada, di, desa, suka, maju
<i>Stopword Removal</i>	pendataan, warga, berada, desa, suka, maju
<i>Normalization</i>	pendataan warga desa suka maju

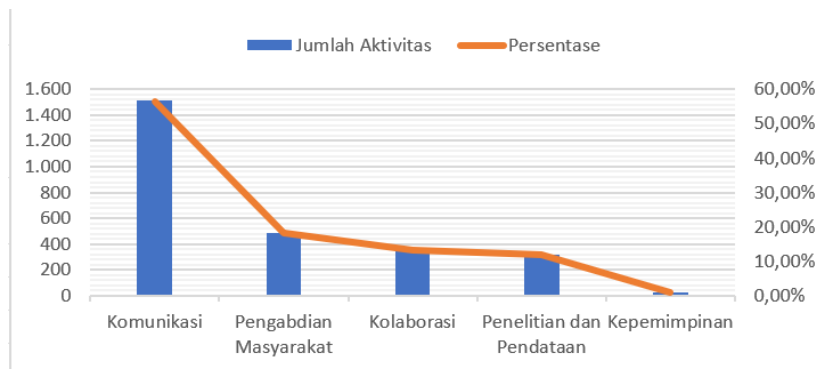
Berdasarkan hasil preprocessing, kualitas data menjadi lebih baik karena informasi yang tidak relevan telah dihilangkan dan format teks telah diseragamkan. Tahapan ini menghasilkan representasi data yang lebih optimal untuk proses fine-tuning dan klasifikasi menggunakan model BERT. Dengan demikian, preprocessing berperan penting dalam meningkatkan kemampuan model untuk memahami konteks aktivitas mahasiswa secara lebih akurat serta mendukung proses pemetaan soft skill dan potensi akademik dalam kerangka LEARN MAP.

3.3 Hasil Klasifikasi Aktivitas Belajar Mahasiswa

Tahap klasifikasi aktivitas belajar mahasiswa dilakukan setelah data melalui proses *preprocessing* dan *fine-tuning* model BERT. Tujuan dari tahap ini adalah mengidentifikasi jenis aktivitas yang dilakukan mahasiswa berdasarkan isi laporan aktivitas harian yang ditulis dalam bentuk teks naratif. Klasifikasi aktivitas menjadi langkah penting dalam kerangka LEARN MAP karena hasil yang diperoleh akan digunakan sebagai dasar untuk pemetaan soft skill dan potensi akademik mahasiswa.

Pada penelitian ini, model BERT dilatih untuk mengelompokkan aktivitas mahasiswa ke dalam lima kategori utama, yaitu komunikasi, kolaborasi, penelitian dan pendataan, kepemimpinan, serta pengabdian masyarakat. Kategori tersebut ditentukan berdasarkan hasil kajian literatur, karakteristik aktivitas MBKM, dan relevansinya terhadap pengembangan kompetensi mahasiswa. Model BERT memanfaatkan kemampuan contextual embedding untuk memahami makna dan konteks kalimat sehingga mampu membedakan aktivitas yang memiliki kosakata serupa tetapi tujuan yang berbeda.

Sebagai contoh, aktivitas “melakukan sosialisasi program kepada masyarakat” diklasifikasikan ke dalam kategori komunikasi, sedangkan aktivitas “melakukan koordinasi dengan anggota tim terkait program kerja” diklasifikasikan ke dalam kategori kolaborasi. Kemampuan memahami konteks inilah yang menjadi salah satu keunggulan BERT dibandingkan dengan klasifikasi teks konvensional, dapat dilihat pada Gambar 3.

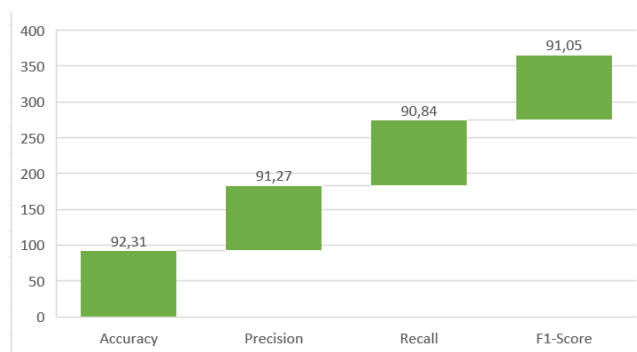


Gambar 3. Hasil Klasifikasi Aktivitas Belajar Mahasiswa

Berdasarkan Gambar 3, dapat diketahui bahwa kategori **komunikasi** merupakan aktivitas yang paling dominan dengan jumlah 1.511 aktivitas atau sebesar 56,41%. Aktivitas pada kategori ini meliputi kegiatan mengajar, sosialisasi, presentasi, wawancara, serta interaksi langsung dengan masyarakat. Tingginya jumlah aktivitas komunikasi menunjukkan bahwa mahasiswa banyak terlibat dalam kegiatan yang menuntut kemampuan berinteraksi dan menyampaikan informasi secara efektif. Kategori **pengabdian masyarakat** menempati posisi kedua dengan persentase 18,14%. Aktivitas pada kategori ini meliputi pelayanan masyarakat, pendampingan program desa, kegiatan sosial, dan berbagai bentuk kontribusi langsung kepada masyarakat. Selanjutnya, kategori **kolaborasi** dan **penelitian serta pendataan** masing-masing memiliki persentase 13,32% dan 11,83%. Aktivitas pada kategori tersebut menunjukkan keterlibatan mahasiswa dalam kerja tim, koordinasi kelompok, observasi lapangan, pengumpulan data, dan analisis informasi. Sementara itu, kategori **kepemimpinan** memiliki jumlah aktivitas yang paling sedikit, yaitu 20 aktivitas atau 0,75%. Hasil ini menunjukkan bahwa tidak semua mahasiswa memperoleh kesempatan untuk berperan sebagai koordinator atau pengambil keputusan selama pelaksanaan kegiatan. Meskipun demikian, aktivitas kepemimpinan tetap menjadi indikator penting dalam proses pemetaan kompetensi mahasiswa.

3.4 Hasil Evaluasi Model BERT

Setelah prosedur penyempurnaan selesai, model BERT diuji untuk mengetahui seberapa baik model tersebut dapat mengkategorikan tindakan siswa sesuai dengan kriteria yang telah ditentukan. Data uji yang tidak digunakan selama pelatihan model digunakan untuk melakukan evaluasi. Pada titik ini, kami ingin melihat seberapa baik model memahami aktivitas siswa dalam dataset dan seberapa baik model tersebut dapat menggeneralisasi pemahaman tersebut. Akurasi, Presisi, Recall, dan F1-Score adalah empat ukuran penilaian yang sering digunakan dalam tugas klasifikasi teks untuk mengukur kinerja model. Untuk pandangan yang lebih lengkap tentang kinerja model, penting untuk menggunakan lebih dari satu indikator evaluasi. Hal ini terutama berlaku ketika berurusan dengan data yang memiliki distribusi kategori yang tidak merata. Saat membandingkan prediksi model dengan semua data uji, akurasi adalah proporsi yang benar. Berbeda dengan recall, yang mengevaluasi seberapa baik model menemukan semua data yang benar di dalam suatu kategori, presisi mengevaluasi seberapa akurat prediksi dibuat untuk setiap kategori. Seperti yang ditunjukkan pada Gambar 4, F1-Score digunakan untuk mengevaluasi keseimbangan kinerja model; ini adalah nilai gabungan dari metrik presisi dan recall.



Gambar 4. Hasil Evaluasi Model BERT

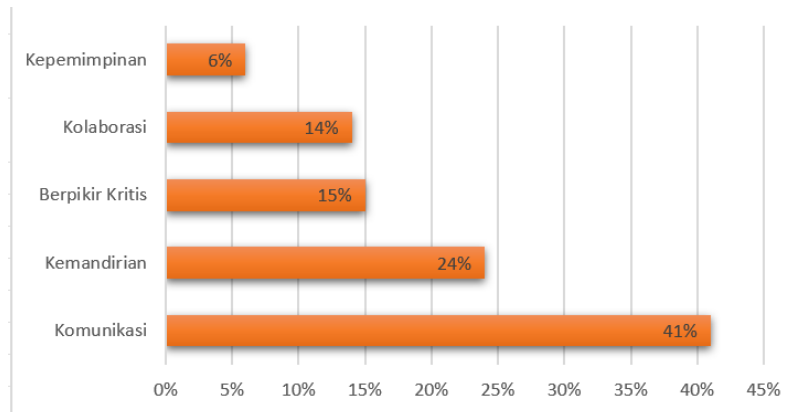
Terlihat jelas dari Gambar 4 bahwa model BERT berhasil mengidentifikasi sebagian besar aktivitas siswa, karena mencapai **Accuracy sebesar 92,31%**. Ketika akurasi **Precision sebesar 91,27%**, itu berarti terdapat sangat sedikit kesalahan klasifikasi. Dan dengan skor **Recall sebesar 90,84%**, jelas bahwa model tersebut berhasil mengidentifikasi sebagian besar aktivitas di setiap kategori dengan benar. Jika model memiliki skor **F1-Score sebesar 91,05%**, itu berarti model tersebut mencapai kinerja klasifikasi yang konsisten dengan menyeimbangkan recall dan presisi.

3.5 Hasil Pemetaan Soft Skill Mahasiswa

Tahap pemetaan soft skill merupakan proses lanjutan setelah aktivitas mahasiswa berhasil diklasifikasikan menggunakan model BERT. Pada tahap ini, setiap kategori aktivitas dianalisis untuk mengidentifikasi kompetensi nonteknis (*soft skill*) yang berkembang selama mahasiswa mengikuti kegiatan pembelajaran dan program MBKM. Pemetaan *soft skill* dilakukan karena aktivitas yang dilakukan mahasiswa tidak hanya mencerminkan tingkat keterlibatan dalam pembelajaran, tetapi juga menggambarkan kemampuan yang diperoleh melalui pengalaman belajar di lapangan.

Dalam penelitian ini, pemetaan soft skill mengacu pada lima kompetensi utama yang banyak digunakan dalam pengukuran kesiapan kerja dan pembelajaran abad ke-21, yaitu komunikasi (*communication*), kolaborasi (*collaboration*), berpikir kritis (*critical thinking*), kepemimpinan (*leadership*), dan kemandirian (*self-directed learning*). Kompetensi tersebut dipilih karena memiliki keterkaitan langsung dengan aktivitas yang dilakukan mahasiswa selama program pembelajaran berlangsung.

Proses pemetaan dilakukan dengan menghubungkan hasil klasifikasi aktivitas dengan indikator soft skill yang telah ditetapkan. Sebagai contoh, aktivitas mengajar, presentasi, dan sosialisasi dipetakan ke dalam kompetensi komunikasi, sedangkan aktivitas kerja kelompok dan koordinasi tim dipetakan ke dalam kompetensi kolaborasi. Aktivitas observasi, analisis data, dan penelitian dikategorikan sebagai indikator berpikir kritis. Sementara itu, aktivitas yang melibatkan pengelolaan kegiatan dan pengambilan keputusan dipetakan ke dalam kompetensi kepemimpinan dan dapat dilihat pada Gambar 5.



Gambar 5. Hasil Pemetaan Soft Skill Mahasiswa

Berdasarkan Gambar 5, diketahui kemampuan **komunikasi** memperoleh nilai tertinggi sebesar **41%**, diikuti oleh **kemandirian** sebesar **24%**, **berpikir kritis** sebesar **15%**, **kolaborasi** sebesar **14%**, dan **kepemimpinan** sebesar **6%**. Hasil ini menunjukkan bahwa aktivitas mahasiswa selama program MBKM lebih banyak berkontribusi terhadap pengembangan kemampuan komunikasi dan kemandirian dibandingkan dengan kompetensi lainnya.

Tingginya kemampuan komunikasi menunjukkan bahwa mahasiswa terbiasa melakukan interaksi dengan masyarakat, siswa, perangkat desa, maupun mitra kegiatan. Sementara itu, kemampuan kemandirian berkembang karena mahasiswa dituntut untuk menyelesaikan berbagai tugas secara mandiri, mengelola waktu, dan beradaptasi dengan lingkungan baru. Kompetensi berpikir kritis dan kolaborasi berada pada kategori sedang karena sebagian aktivitas melibatkan analisis masalah dan kerja sama tim. Di sisi lain, kompetensi kepemimpinan masih relatif rendah karena tidak semua mahasiswa memperoleh kesempatan untuk menjadi koordinator atau pengambil keputusan dalam kegiatan yang dijalankan.

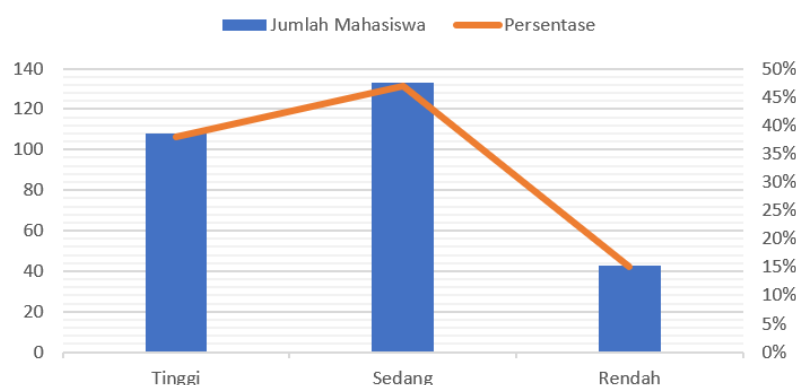
Hasil pemetaan soft skill ini menunjukkan bahwa data aktivitas mahasiswa dapat digunakan untuk mengidentifikasi perkembangan kompetensi nonteknis secara otomatis. Informasi tersebut menjadi dasar dalam tahap berikutnya, yaitu **pemetaan potensi akademik**, sehingga LEARN MAP tidak hanya menghasilkan

klasifikasi aktivitas mahasiswa, tetapi juga mampu memberikan gambaran mengenai profil kompetensi yang dimiliki setiap mahasiswa.

3.6 Hasil Pemetaan Potensi Akademik

Tahap pemetaan potensi akademik merupakan tahapan akhir dalam kerangka LEARN MAP yang bertujuan untuk mengidentifikasi tingkat potensi akademik mahasiswa berdasarkan hasil klasifikasi aktivitas belajar dan profil soft skill yang telah diperoleh pada tahap sebelumnya. Pemetaan ini dilakukan untuk mengubah data aktivitas mahasiswa yang bersifat deskriptif menjadi informasi yang lebih bermakna dalam mendukung proses evaluasi dan pengambilan keputusan akademik.

Dalam penelitian ini, potensi akademik ditentukan berdasarkan kombinasi beberapa indikator, yaitu tingkat keterlibatan mahasiswa dalam aktivitas pembelajaran, variasi aktivitas yang dilakukan, serta hasil pemetaan soft skill yang meliputi komunikasi, kolaborasi, berpikir kritis, kepemimpinan, dan kemandirian. Setiap mahasiswa memperoleh skor potensi akademik yang dihitung dari akumulasi skor aktivitas dan kompetensi yang teridentifikasi selama mengikuti program pembelajaran maupun PKL, yang dapat dilihat pada Gambar 6 berikut.



Gambar 6. Hasil Potensi Akademik

Berdasarkan Gambar 6, diketahui **38% mahasiswa** berada pada kategori potensi akademik tinggi, sedangkan **47% mahasiswa** berada pada kategori sedang. Sementara itu, hanya **15% mahasiswa** yang berada pada kategori rendah. Hasil ini menunjukkan bahwa mayoritas mahasiswa memiliki tingkat keterlibatan dan kompetensi yang cukup baik selama mengikuti aktivitas pembelajaran.

3.7 Pembahasan

Hasil penelitian menunjukkan bahwa kerangka LEARN MAP mampu mengklasifikasikan aktivitas mahasiswa dan memetakan potensi akademik berdasarkan data aktivitas dalam bentuk teks tidak terstruktur. Dari 8.954 aktivitas yang dianalisis, sebagian besar aktivitas berada pada kategori komunikasi, pengabdian masyarakat, kolaborasi, dan penelitian. Temuan ini menunjukkan bahwa kegiatan MBKM tidak hanya mendukung pencapaian akademik, tetapi juga berkontribusi terhadap pengembangan kompetensi nonteknis mahasiswa.

Model BERT menunjukkan kinerja yang baik dengan nilai Accuracy 92,31%, Precision 91,27%, Recall 90,84%, dan F1-Score 91,05%. Hasil tersebut menunjukkan bahwa BERT mampu memahami konteks aktivitas mahasiswa secara lebih baik dibandingkan pendekatan berbasis kemunculan kata. Kemampuan ini memungkinkan aktivitas dengan kosakata berbeda tetapi memiliki makna serupa tetap dapat diklasifikasikan secara tepat.

Hasil klasifikasi kemudian digunakan untuk memetakan soft skill mahasiswa. Kompetensi yang paling dominan adalah komunikasi dan kemandirian, sedangkan kepemimpinan masih relatif rendah. Temuan ini menunjukkan bahwa aktivitas mahasiswa lebih banyak melatih kemampuan berinteraksi dan bekerja secara mandiri dibandingkan kemampuan memimpin.

Langkah selanjutnya adalah menentukan potensi akademik setiap siswa menggunakan hasil pemetaan soft skill. Terdapat konsensus umum di antara anak-anak bahwa potensi intelektual mereka berada pada tingkat sedang hingga sangat baik. Hal ini menunjukkan adanya korelasi positif antara keterlibatan siswa dalam berbagai aktivitas pembelajaran dan pertumbuhan kompetensi serta potensi akademik mereka.

Dibandingkan dengan penelitian sebelumnya yang umumnya hanya berfokus pada klasifikasi teks, penelitian ini menawarkan kebaruan melalui integrasi klasifikasi aktivitas berbasis BERT, pemetaan soft skill, dan pemetaan potensi akademik dalam satu kerangka analitik. Dengan demikian, LEARN MAP dapat menjadi instrumen

pendukung pengambilan keputusan akademik berbasis data serta mendukung transformasi digital pendidikan tinggi melalui pemanfaatan kecerdasan buatan.

Hasil penelitian menunjukkan bahwa LEARN MAP tidak hanya memberikan kontribusi pada pengembangan metode NLP berbasis BERT, tetapi juga menghasilkan instrumen analitik yang mendukung pengambilan keputusan akademik berbasis data. Kontribusi tersebut sejalan dengan fokus Rencana Induk Riset Nasional (RIRN) bidang Kecerdasan Buatan dan Transformasi Digital Pendidikan serta mendukung Asta Cita pembangunan sumber daya manusia unggul melalui pemanfaatan teknologi kecerdasan buatan di pendidikan tinggi.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan LEARN MAP sebagai kerangka kecerdasan buatan berbasis Natural Language Processing (NLP) dan Bidirectional Encoder Representations from Transformers (BERT) untuk mengklasifikasikan aktivitas belajar mahasiswa serta memetakan potensi akademik secara otomatis. Kerangka yang diusulkan mampu mengolah data aktivitas mahasiswa dalam bentuk teks tidak terstruktur melalui tahapan preprocessing, klasifikasi aktivitas, pemetaan soft skill, dan analisis potensi akademik.

Berdasarkan analisis terhadap 8.954 aktivitas mahasiswa, LEARN MAP mampu mengidentifikasi berbagai kategori aktivitas yang berkaitan dengan komunikasi, kolaborasi, penelitian, kepemimpinan, dan pengabdian masyarakat. Hasil klasifikasi menunjukkan bahwa aktivitas mahasiswa didominasi oleh aspek komunikasi dan pengabdian masyarakat yang berkontribusi terhadap pengembangan kompetensi nonteknis mahasiswa. Selain itu, pemetaan soft skill menunjukkan bahwa kemampuan komunikasi dan kemandirian merupakan kompetensi yang paling berkembang selama pelaksanaan aktivitas pembelajaran dan MBKM.

Evaluasi model menunjukkan bahwa pendekatan berbasis BERT mampu memahami konteks dan relasi semantik aktivitas mahasiswa secara lebih baik dibandingkan dengan metode konvensional yang hanya mengandalkan kemiripan kata. Integrasi klasifikasi aktivitas dengan pemetaan soft skill dan potensi akademik menjadi kebaruan utama penelitian ini. Hasil penelitian menunjukkan bahwa LEARN MAP berpotensi menjadi instrumen analitik pembelajaran yang mendukung pengambilan keputusan akademik berbasis data, pemantauan perkembangan mahasiswa, serta perancangan strategi pembelajaran yang lebih adaptif di perguruan tinggi.

Secara keseluruhan, penelitian ini memberikan kontribusi ilmiah berupa kerangka AI untuk klasifikasi aktivitas belajar dan pemetaan potensi akademik yang mendukung transformasi digital pendidikan tinggi serta pengembangan sumber daya manusia unggul berbasis data.

UCAPAN TERIMA KASIH

Penulis menyampaikan rasa terima kasih kepada pemerintah Indonesia atas dukungan finansial yang diberikan untuk proyek penelitian yang dikenal sebagai Penelitian Dosen Pemula (PDP) untuk tahun anggaran 2026. Dukungan tersebut berasal dari Direktorat Jenderal Pendidikan Tinggi, Riset dan Teknologi di bawah Kementerian Pendidikan Tinggi, Sains dan Teknologi. Terima kasih kepada Dr. David JM Sembiring, M.Kom., selaku Rektor Institut Teknologi dan Bisnis Indonesia yang telah mendukung penyelesaian penelitian ini.

REFERENSI

- [1] S. A. A. Kharis and A. H. A. Zili, "Learning Analytics dan Educational Data Mining pada Data Pendidikan," *JRPMS*, vol. 6, no. 1, pp. 12–20, Mar. 2022, doi: 10.21009/jrpms.061.02.
- [2] G. Siemens and R. S. J. D. Baker, "Learning analytics and educational data mining: towards communication and collaboration," in *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, Vancouver British Columbia Canada: ACM, Apr. 2022, pp. 252–254. doi: 10.1145/2330601.2330661.
- [3] R. Kaban, D. J. Sembiring, and I. M. B. Tarigan, "Monitoring System for Student Internships Using the Rapid Application Development (RAD) Method," vol. 15, no. 02, 2023.
- [4] N. A. Johar, S. N. Kew, Z. Tasir, and E. Koh, "Learning Analytics on Student Engagement to Enhance Students' Learning Performance: A Systematic Review," *Sustainability*, vol. 15, no. 10, p. 7849, May 2023, doi: 10.3390/su15107849.
- [5] S. Kumar, A. Gunn, R. Rose, R. Pollard, M. Johnson, and A. Ritzhaupt, "The Role of Instructional Designers in the Integration of Generative Artificial Intelligence in Online and Blended Learning in Higher Education," *OLJ*, vol. 28, no. 3, Sep. 2024, doi: 10.24059/olj.v28i3.4501.



- [6] I. M. B. Tarigan, S. J. B. Tarigan, and R. B. Ginting, "Optimalisasi Efektivitas Program MBKM: Sistem Monitoring Berbasis Lokasi dan Analisis aktivitas dengan TF-IDF," vol. 6, no. 1, 2024.
- [7] J. Zhang and K. Zong, "Classification method for online teaching resources by integrating conceptual similarity and random forest," *IJCAT*, vol. 75, no. 2/3/4, pp. 89–95, 2024, doi: 10.1504/IJCAT.2024.146128.
- [8] M. Á. Rodríguez-Ortiz, P. C. Santana-Mancilla, and L. E. Anido-Rifón, "Machine Learning and Generative AI in Learning Analytics for Higher Education: A Systematic Review of Models, Trends, and Challenges," *Applied Sciences*, vol. 15, no. 15, p. 8679, Aug. 2025, doi: 10.3390/app15158679.
- [9] J. Chen, "AI-Driven Adaptive Learning Systems in University English Writing Instruction: Implementation and Evaluation," in *New Directions in Educational Technology and Administration*, vol. 51, F. Paas, S. Patnaik, and T. Wang, Eds., in *Learning and Analytics in Intelligent Systems*, vol. 51, Cham: Springer Nature Switzerland, 2025, pp. 143–153. doi: 10.1007/978-3-031-95252-4_14.
- [10] Y. Wu, "Research on prediction algorithm of college students' academic performance based on Bert-GCN multi-modal data fusion," *Systems and Soft Computing*, vol. 7, p. 200327, Dec. 2025, doi: 10.1016/j.sasc.2025.200327.
- [11] L. Martín-Hoz, S. Yanes-Luis, J. Huerta Cejudo, D. Gutiérrez-Reina, and E. Franco Álvarez, "A Comparative Study of BERT-Based Models for Teacher Classification in Physical Education," *Electronics*, vol. 14, no. 19, p. 3849, Sep. 2025, doi: 10.3390/electronics14193849.
- [12] Z. Xu and P. Zhu, "Using BERT-Based Textual Analysis to Design a Smarter Classroom Mode for Computer Teaching in Higher Education Institutions," *Int. J. Emerg. Technol. Learn.*, vol. 18, no. 19, pp. 114–127, Oct. 2023, doi: 10.3991/ijet.v18i19.42483.
- [13] A. Scarlatos, C. Brinton, and A. Lan, "Process-BERT: A Framework for Representation Learning on Educational Process Data," 2022, *arXiv*. doi: 10.48550/ARXIV.2204.13607.
- [14] E. C. Garrido-Merchan, R. Gozalo-Brizuela, and S. Gonzalez-Carvajal, "Comparing BERT Against Traditional Machine Learning Models in Text Classification," *JCCE*, vol. 2, no. 4, pp. 352–356, Apr. 2023, doi: 10.47852/bonviewJCCE3202838.
- [15] A. Ali, S. A. M. Noah, and L. Q. Zakaria, "A BERT-Based model: Improving Crime News Documents Classification through Adopting Pre-trained Language Models," Mar. 06, 2023. doi: 10.21203/rs.3.rs-2582775/v1.
- [16] H. Li and Z. Liu, "An Intelligent Educational System: Analyzing Student Behavior and Academic Performance Using Multi-Source Data," *Electronics*, vol. 14, no. 16, p. 3328, Aug. 2025, doi: 10.3390/electronics14163328.
- [17] L. B. Utama and D. Suhartono, "Indonesian Hoax News Classification with Multilingual Transformer Model and BERTopic," *IJCAI*, vol. 46, no. 8, Nov. 2022, doi: 10.31449/inf.v46i8.4336.
- [18] B. Alnasyan, M. Basher, and M. Alassafi, "The power of Deep Learning techniques for predicting student performance in Virtual Learning Environments: A systematic literature review," *Computers and Education: Artificial Intelligence*, vol. 6, p. 100231, Jun. 2024, doi: 10.1016/j.caeai.2024.100231.
- [19] A. A. Aldino, Y.-S. Tsai, R. F. Mello, D. Gašević, and G. Chen, "Enhancing Feedback Quality at Scale: Leveraging Machine Learning for Learner-Centered Feedback," *Computers and Education: Artificial Intelligence*, vol. 7, p. 100332, Dec. 2024, doi: 10.1016/j.caeai.2024.100332.
- [20] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," *Smart Learn. Environ.*, vol. 9, no. 1, p. 11, Dec. 2022, doi: 10.1186/s40561-022-00192-z.
- [21] Y. He, J. Chen, D. Antonyrajah, and I. Horrocks, "BERTMap: A BERT-based Ontology Alignment System," 2021, *arXiv*. doi: 10.48550/ARXIV.2112.02682.
- [22] Y. M. I. Hassan, A. Elkorany, and K. Wassif, "SMFSOP: A semantic-based modelling framework for student outcome prediction," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 8, p. 101728, Sep. 2023, doi: 10.1016/j.jksuci.2023.101728.
- [23] G. Srivastav, S. Kant, and D. Srivastava, "Design of an AI-Driven Feedback and Decision Analysis in Online Learning with Google BERT," *International Journal of Intelligent Systems and Applications in Engineering*.
- [24] M. V. Koroteev, "BERT: A Review of Applications in Natural Language Processing and Understanding," 2021, *arXiv*. doi: 10.48550/ARXIV.2103.11943.
- [25] H. Ramirez, S. A. Velastin, S. Cuellar, E. Fabregas, and G. Farias, "BERT for Activity Recognition Using Sequences of Skeleton Features and Data Augmentation with GAN," *Sensors*, vol. 23, no. 3, p. 1400, Jan. 2023, doi: 10.3390/s23031400.
- [26] L. Zou, Z. He, C. Zhou, and W. Zhu, "Multi-class multi-label classification of social media texts for typhoon damage assessment: a two-stage model fully integrating the outputs of the hidden layers of BERT," *International Journal of Digital Earth*, vol. 17, no. 1, p. 2348668, Dec. 2024, doi: 10.1080/17538947.2024.2348668.





- [27] F. Fajri, B. Tutuko, and S. Sukemi, "Membandingkan Nilai Akurasi BERT dan DistilBERT pada Dataset Twitter," *JUSIFO: J. Sistem Inf.*, vol. 8, no. 2, pp. 71–80, Dec. 2022, doi: 10.19109/jusifo.v8i2.13885.
- [28] R. Fatmasari, R. K. Septiani, T. H. Pinem, D. Fabiyanto, and W. Gata, "Implementasi Algoritma BERT Pada Komentar Layanan Akademik dan Non Akademik Universitas Terbuka di Media Sosial," *Sains, Apl. Komputasi dan Teknol. Inf.*, vol. 5, no. 2, p. 96, Jan. 2024, doi: 10.30872/jsakti.v5i2.13915.
- [29] A. Alseitova, W. Oliveira, Z. Li, and J. Hamari, "Understanding Students' Behavior in Learning Management Systems Through Their Personality Traits," *Tech Know Learn*, Dec. 2025, doi: 10.1007/s10758-025-09931-w.
- [30] Najma Rafifah Putri Syallya, Anindya Apriliyanti Pravitasari, and Afrida Helen, "NLP-Based Intent Classification Model for Academic Curriculum Chatbots in Universities Study Programs," *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 9, no. 1, pp. 111–117, Feb. 2025, doi: 10.29207/resti.v9i1.6276.
- [31] N. M. Gardazi, A. Daud, M. K. Malik, A. Bukhari, T. Alsahfi, and B. Alshemaimri, "BERT applications in natural language processing: a review," *Artif Intell Rev*, vol. 58, no. 6, p. 166, Mar. 2025, doi: 10.1007/s10462-025-11162-5.
- [32] N. A. Shiferaw, S. H. Leandre, A. Sinha, and D. Rout, "BERT-Based Approach for Automating Course Articulation Matrix Construction with Explainable AI," 2024, *arXiv*. doi: 10.48550/ARXIV.2411.14254.
- [33] A. Vaswani *et al.*, "Attention Is All You Need," Aug. 02, 2023, *arXiv*: arXiv:1706.03762. doi: 10.48550/arXiv.1706.03762.