

Pengembangan Sistem Evaluasi Adaptif Berbasis LLM dalam Smart Tutor Menggunakan Bloom's Taxonomy

<http://dx.doi.org/10.28932/jste.v2i2.13159>

Received: 26 Agustus 2025 | Revised: 11 November 2025 | Accepted: 10 Desember 2025

Creative Commons License 4.0 (CC BY – NC)



Nicholas Christopher Rudy^{✉#1}, Hapnes Toba^{*2}

[#]Program Studi Teknik Informatika, Fakultas Teknologi dan Rekayasa Cerdas, Universitas Kristen Maranatha
Bandung, Jawa Barat 40164, Indonesia

¹2172020@maranatha.ac.id

^{*}Program Studi Teknik Informatika, Fakultas Teknologi dan Rekayasa Cerdas, Universitas Kristen Maranatha
Bandung, Jawa Barat 40164, Indonesia

²hapnestoba@it.maranatha.edu

✉ Corresponding author: 2172020@maranatha.ac.id

How to cite this article:

N. C. Rudy, H. Toba, "Development of an LLM-Based Adaptive Evaluation System in Smart Tutor Using Bloom's Taxonomy," *Journal of Smart Technology and Engineering*, vol. 2, no. 2, pp. 75-83, 2026. <https://doi.org/10.28932/jts.v2i2.13159>

Abstrak — Penelitian ini bertujuan untuk meningkatkan sistem pembelajaran adaptif dengan mengembangkan platform Smart Tutor yang dilengkapi dengan *Large Language Models* (LLMs). Sistem tradisional yang mengandalkan pencocokan kata kunci atau kesamaan kosinus seringkali gagal menilai respons siswa secara akurat karena kurangnya pemahaman semantik. Untuk mengatasi hal ini, kami mengusulkan sistem evaluasi berbasis semantik yang menggunakan *BERTScore* dan *TF-IDF* untuk menganalisis jawaban siswa secara lebih kontekstual.

Kata Kunci— *BERTScore*; *Evaluasi LLM*; *Pembelajaran adaptif*; *Smart Tutor*; *Taksonomi Bloom*.

Development of an LLM-Based Adaptive Evaluation System in Smart Tutor Using Bloom's Taxonomy

Abstract — *This research aims to improve adaptive learning systems by developing a Smart Tutor platform enhanced with Large Language Models (LLMs). Traditional systems that rely on keyword matching or cosine similarity often fail to accurately assess student responses due to their lack of semantic understanding. To address this, we propose a semantic-based evaluation system using BERTScore and TF-IDF to analyze student answers more contextually.*

Keywords — *Adaptive learning; BERTScore; Bloom's Taxonomy; LLM evaluation; Smart Tutor.*

I. PENDAHULUAN

Perkembangan teknologi kecerdasan buatan (AI) telah memberikan dampak signifikan dalam dunia pendidikan, khususnya dalam membangun sistem pembelajaran yang lebih interaktif dan otomatis. Salah satu tantangan utama dalam sistem pembelajaran digital adalah kemampuan untuk melakukan evaluasi jawaban mahasiswa secara akurat, efisien, dan adil. Evaluasi yang dilakukan secara manual membutuhkan waktu dan tenaga yang besar, sementara metode otomatis tradisional sering kali gagal menangkap makna yang terkandung dalam jawaban mahasiswa.

Salah satu sistem pembelajaran yang dikembangkan sebelumnya adalah *Smart Tutor*, yang berfungsi untuk menghasilkan pertanyaan dari dokumen PDF dan mengevaluasi jawaban mahasiswa secara otomatis. Namun, sistem awal ini masih menggunakan pendekatan *cosine similarity* sederhana yang sangat sensitif terhadap struktur kalimat dan variasi kata, sehingga tidak mampu menilai kesamaan makna secara semantik. Hal ini menyebabkan ketidaktepatan dalam evaluasi dan berkurangnya efektivitas sistem pembelajaran adaptif.

Untuk mengatasi keterbatasan tersebut, penelitian ini mengembangkan sistem Smart Tutor versi terbaru dengan mengintegrasikan metode evaluasi berbasis *semantic similarity* menggunakan *BERTScore* dan *TF-IDF*. Sistem ini juga mengadopsi empat level pertama dari *Bloom's Taxonomy*—yaitu *remembering*, *understanding*, *applying*, dan *analysing*—sebagai dasar dalam menyusun dan mengadaptasi pertanyaan sesuai tingkat pemahaman mahasiswa. Dengan demikian, sistem dapat memberikan pertanyaan yang semakin kompleks seiring meningkatnya performa mahasiswa secara bertahap.

Penelitian ini bertujuan untuk merancang dan membangun sistem evaluasi adaptif berbasis *LLM* yang tidak hanya menilai jawaban mahasiswa secara semantik, tetapi juga mampu menyesuaikan tantangan belajar secara real-time. Diharapkan, pendekatan ini dapat meningkatkan efektivitas pembelajaran mandiri berbasis *AI*, serta memberikan pengalaman belajar yang lebih personal, dinamis, dan relevan bagi setiap mahasiswa.

II. KAJIAN TEORI

A. Pembelajaran Adaptif dan Kecerdasan Buatan

Pemanfaatan kecerdasan buatan (*AI*) dalam dunia pendidikan semakin berkembang, khususnya dalam membangun sistem pembelajaran adaptif yang dapat menyesuaikan materi atau soal dengan kemampuan masing-masing siswa. Salah satu tantangan utama adalah bagaimana sistem dapat mengevaluasi jawaban mahasiswa secara akurat, terutama dalam menjawab soal esai yang memiliki variasi bahasa yang tinggi.

B. Bloom's Taxonomy dalam Penyesuaian Level Soal

Bloom's Taxonomy digunakan untuk mengelompokkan tingkat kognitif dalam pembelajaran. Dalam sistem ini, digunakan empat level awal, yaitu: (1) *Remembering*, mengenali atau mengingat informasi; (2) *Understanding*, menjelaskan konsep dengan kata sendiri; (3) *Applying*, menerapkan konsep dalam kasus baru; dan (4) *Analysing*, mengidentifikasi pola dan hubungan antar konsep. Penerapan *Bloom's Taxonomy* membantu sistem menyusun pertanyaan dengan tingkat kesulitan yang sesuai dengan perkembangan pemahaman mahasiswa.

C. Evaluasi Semantik dengan *BERTScore* dan *TF-IDF*

BERTScore adalah metode evaluasi berbasis representasi semantik yang memanfaatkan model *pre-trained* seperti *BERT* untuk mengukur kesamaan makna antara jawaban mahasiswa dan referensi. *TF-IDF* digunakan untuk mengidentifikasi kata kunci penting dari dokumen sumber dan jawaban mahasiswa. Kombinasi kedua metode ini menghasilkan evaluasi yang lebih adil dan akurat dibanding *cosine similarity*, yang terlalu bergantung pada kesamaan bentuk kata.

D. Prompt Engineering pada *LLM*

Prompt engineering adalah teknik yang digunakan untuk mengarahkan model bahasa besar (*LLM*) agar menghasilkan keluaran yang lebih sesuai dengan kebutuhan sistem. Dalam penelitian ini, digunakan

beberapa strategi seperti few-shot prompting, *chain-of-thought*, dan *self-consistency* untuk meningkatkan kualitas pertanyaan dan evaluasi. Teknik ini membuat sistem *Smart Tutor* mampu beradaptasi lebih baik terhadap konteks dan tingkat pemahaman pengguna.

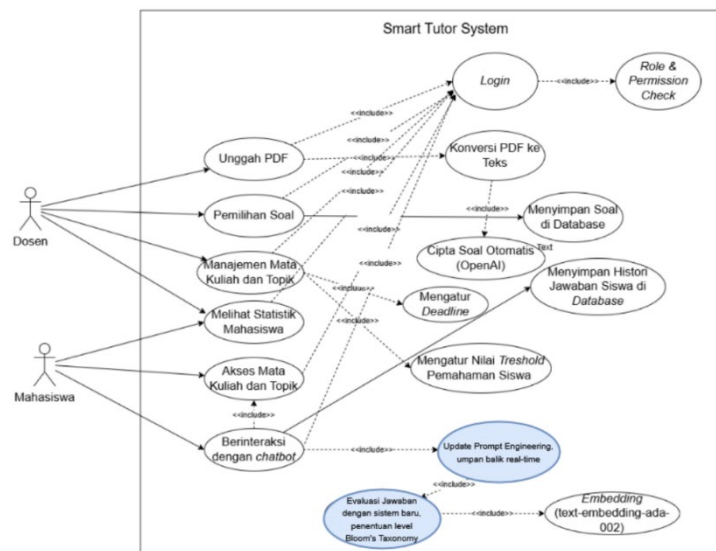
III. ANALISIS DAN RANCANGAN SISTEM

A. Gambaran Umum Sistem

Sistem *Smart Tutor* dikembangkan untuk mendukung pembelajaran adaptif berbasis kecerdasan buatan. Sistem ini memungkinkan dosen mengunggah dokumen materi dalam format PDF, yang kemudian diproses untuk menghasilkan pertanyaan secara otomatis. Mahasiswa dapat menjawab pertanyaan melalui chatbot, dan jawaban akan dievaluasi menggunakan pendekatan semantik berbasis *BERTScore* dan *TF-IDF*. Berbeda dengan sistem sebelumnya yang menggunakan cosine similarity, pendekatan baru ini lebih mampu menangkap makna semantik, sehingga menghasilkan evaluasi yang lebih akurat dan adil.

B. Use Case Diagram Sistem

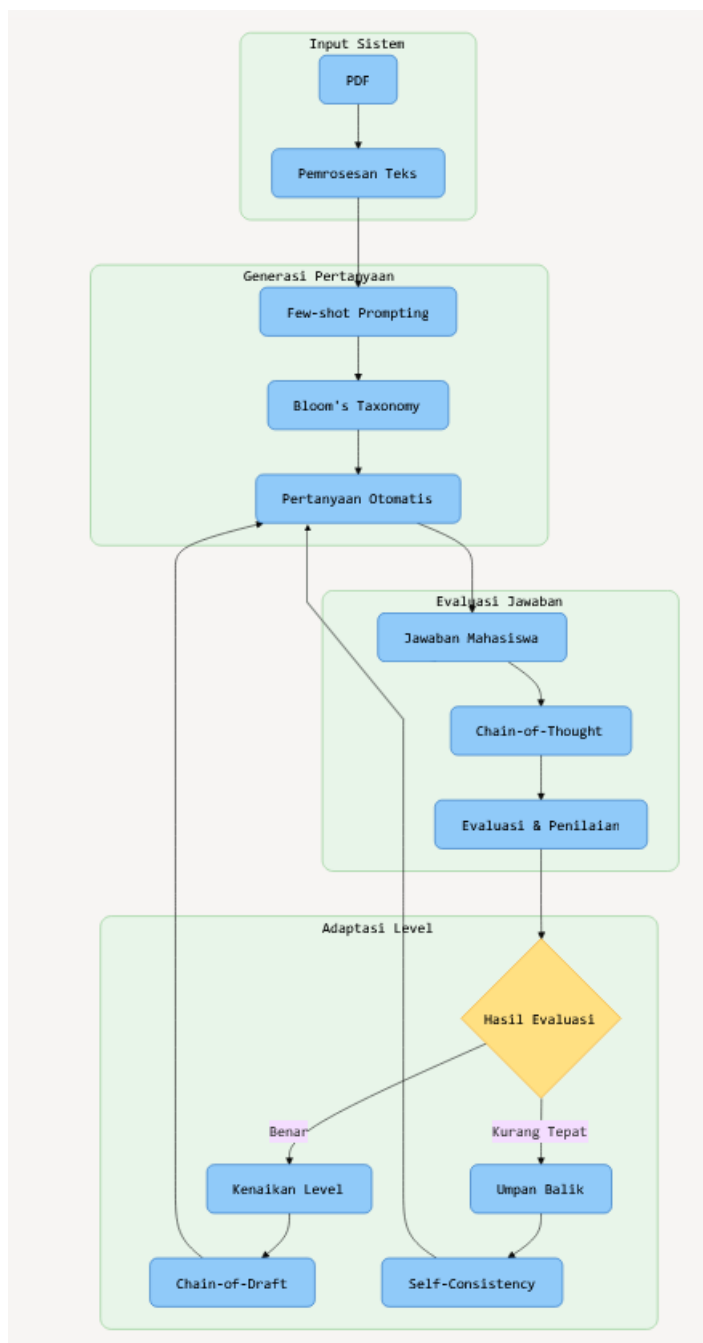
Smart Tutor melibatkan tiga aktor utama, yaitu dosen, mahasiswa, dan administrator. Dosen bertugas mengunggah materi dan meninjau soal, mahasiswa menjawab pertanyaan dan menerima umpan balik adaptif, sedangkan administrator bertugas mengelola sistem secara umum. Gambar 1 berikut menunjukkan *Use Case Diagram* sistem secara keseluruhan:



Gambar 1. Use Case Diagram Smart Tutor.

C. Alur Adaptasi Soal dan Evaluasi

Setelah dosen mengunggah dokumen, sistem akan mengekstrak isi materi dan melakukan pembangkitan pertanyaan berdasarkan *Bloom's Taxonomy*. Mahasiswa akan mendapatkan pertanyaan sesuai level awal (*Remembering*), dan apabila berhasil menjawab dengan baik dalam beberapa kali percobaan, sistem akan menaikkan tingkat kesulitan. Jika mahasiswa salah menjawab, sistem akan memberikan umpan balik dan tetap berada di level yang sama. Gambar 2 menggambarkan alur sistem dari generasi pertanyaan hingga evaluasi dan adaptasi level:



Gambar 2. Workflow Adaptasi Soal dan Evaluasi Smart Tutor.

D. Komponen Utama Sistem

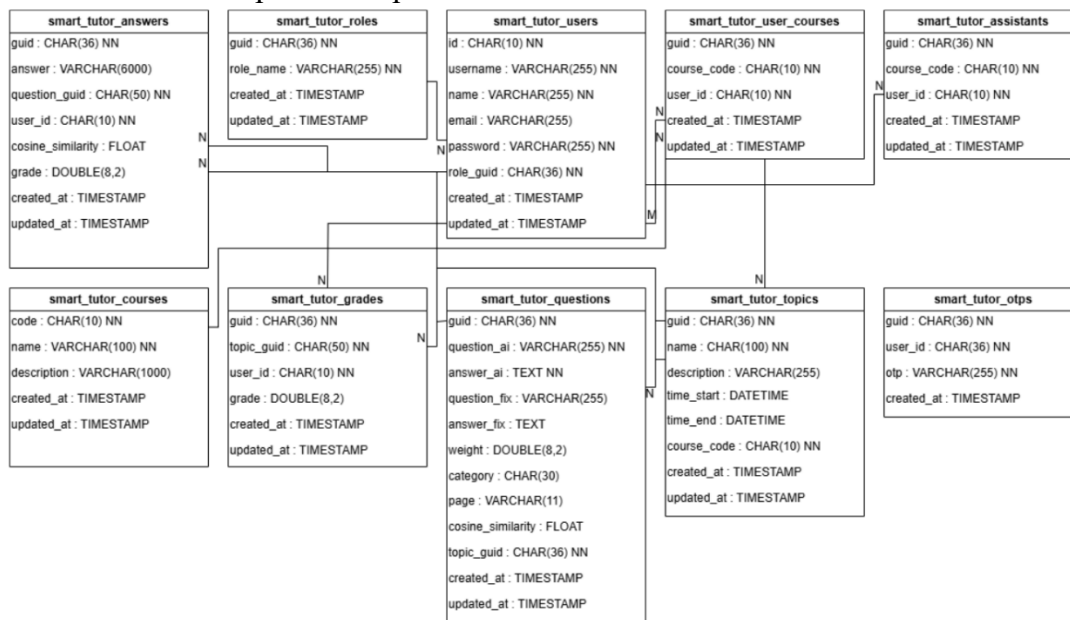
Sistem ini dibangun dengan arsitektur modular, di mana backend utama dikembangkan menggunakan *Laravel* untuk pengelolaan *database* dan API, sedangkan evaluasi semantik dijalankan melalui *Flask* dengan integrasi model *BERT* dan *TF-IDF*. Proses pembuatan pertanyaan dan evaluasi jawaban dilakukan menggunakan *OpenAI* melalui API GPT dengan berbagai teknik prompt engineering, seperti few-shot prompting untuk generasi soal, *chain-of-thought* prompting untuk evaluasi, dan *self-consistency* prompting untuk validasi umpan balik.

IV. IMPLEMENTASI

A. Arsitektur Teknologi dan Integrasi

Sistem *Smart Tutor* dibangun dengan arsitektur terpisah namun saling terhubung menggunakan *REST API*. *Laravel* digunakan sebagai *backend* dan *frontend* yang menangani manajemen pengguna, materi, Pengembangan Sistem Evaluasi Adaptif Berbasis LLM dalam Smart Tutor Menggunakan Bloom's Taxonomy
Nicholas Christopher Rudy, Hapnes Toba

database soal-jawaban, dan desain antarmuka secara keseluruhan. *Flask* digunakan untuk memproses evaluasi semantik terhadap jawaban mahasiswa dengan algoritma *BERTScore* dan *TF-IDF*. Sistem juga menggunakan *OpenAI GPT* melalui *HTTP* request untuk menghasilkan pertanyaan secara otomatis berdasarkan materi PDF dan strategi prompting yang telah dirancang. Komunikasi antara frontend dan backend dilakukan secara asinkron menggunakan *jQuery Ajax*. *Entity Relationship Diagram* untuk penyimpanan data sistem ini dapat dilihat pada Gambar 3.



Gambar 3. Entity Relationship Diagram.

B. Implementasi Evaluasi Jawaban

Evaluasi jawaban dalam sistem *Smart Tutor* dilakukan secara otomatis dengan pendekatan semantik yang menggabungkan dua teknik utama, yaitu *BERTScore* dan *TF-IDF*. *BERTScore* digunakan untuk menilai kesamaan makna antara jawaban mahasiswa dengan jawaban referensi yang dihasilkan oleh model bahasa besar (*LLM*). Teknik ini bekerja dengan cara mengubah setiap token dalam kalimat menjadi representasi vektor, kemudian menghitung kemiripan semantik antar token menggunakan *cosine similarity*, dan menghasilkan skor akhir dalam bentuk *F1-score*. Skor ini merepresentasikan seberapa baik makna jawaban mahasiswa sesuai dengan makna yang diharapkan.

Sementara itu, *TF-IDF* digunakan sebagai pelengkap untuk mengevaluasi keberadaan kata-kata kunci penting yang berkaitan langsung dengan konteks pertanyaan dan materi pembelajaran. Dengan menganalisis bobot kata berdasarkan frekuensinya dalam dokumen dan konteks keseluruhan, *TF-IDF* memberikan skor yang mencerminkan kekuatan representasi topik dalam jawaban mahasiswa.

Kedua skor tersebut dikombinasikan menggunakan rumus terintegrasi dengan pembobotan: 60% dari *BERTScore* dan 40% dari *TF-IDF*. Rumus ini dirancang untuk menyeimbangkan antara kemiripan semantik dan keberadaan kata kunci yang penting. Jika hasil akhir dari kombinasi tersebut melebihi nilai ambang batas tertentu, maka jawaban dianggap benar. Sistem kemudian akan melanjutkan ke pertanyaan selanjutnya dengan tingkat kesulitan yang lebih tinggi. Namun, apabila skor evaluasi berada di bawah ambang batas, sistem akan memberikan umpan balik berupa koreksi, penjelasan singkat, serta pertanyaan lanjutan pada level yang sama guna memperkuat pemahaman mahasiswa terhadap konsep yang sedang dipelajari.

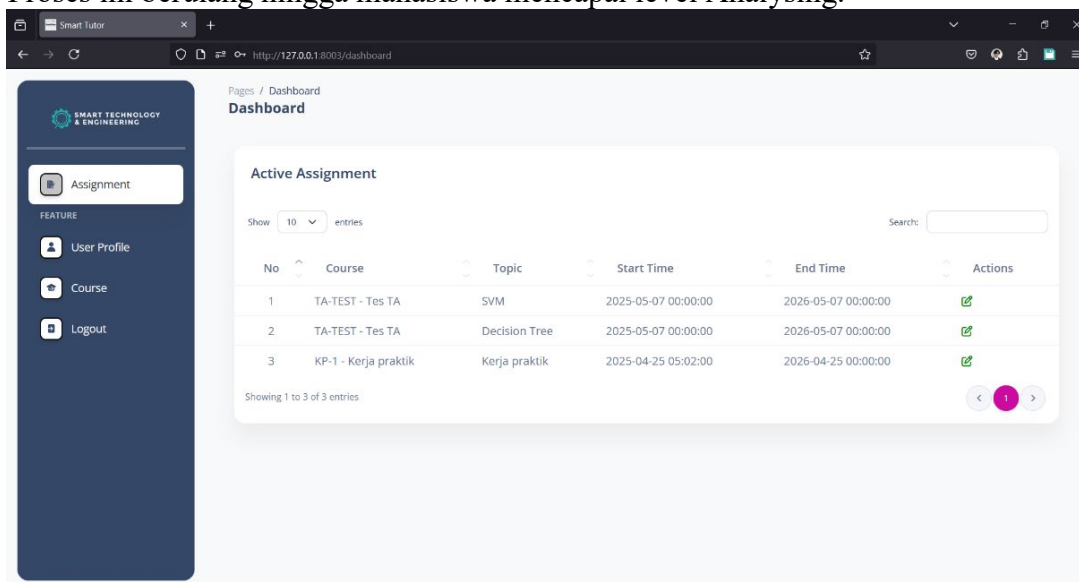
C. Prompt Engineering dan Adaptasi Soal

Proses pembuatan dan penyesuaian soal dilakukan melalui teknik *prompt engineering*. Sistem menggunakan *Few-shot prompting* untuk memberi contoh sebelum memicu model membuat soal baru; *Chain-of-thought prompting*, untuk mengevaluasi alasan atau alur berpikir mahasiswa; *Self-consistency prompting*, untuk memvalidasi koreksi jawaban secara berulang. Logika adaptasi sistem dirancang agar

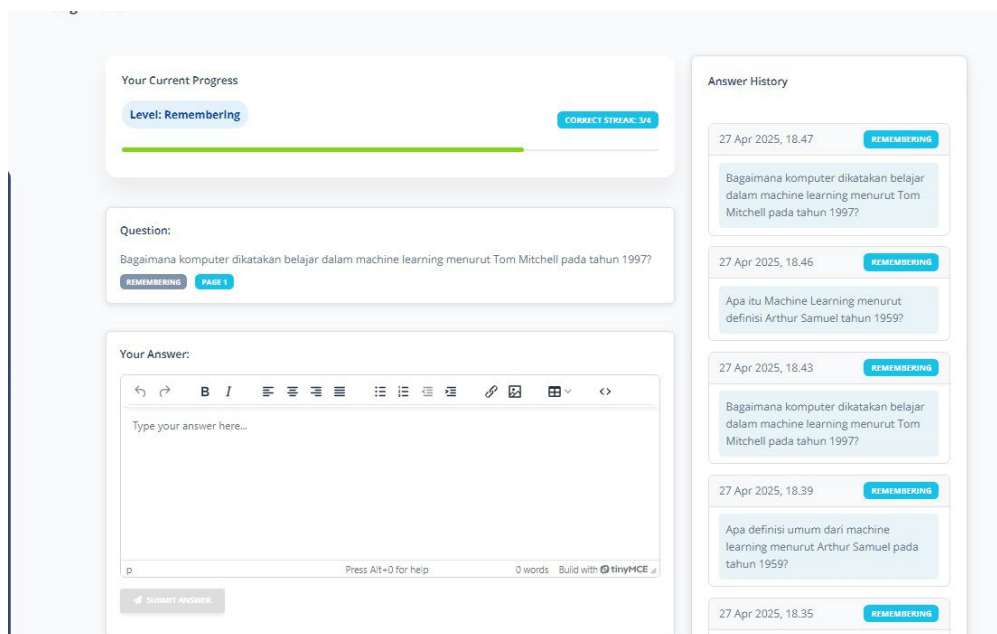
mahasiswa naik level setelah menjawab benar 4 kali berturut-turut. Jika salah, sistem mengulang soal dengan konteks yang sama namun lebih sederhana, disertai penjelasan tambahan. Proses ini menciptakan pembelajaran yang responsif dan personal.

D. Tampilan Antarmuka Sistem

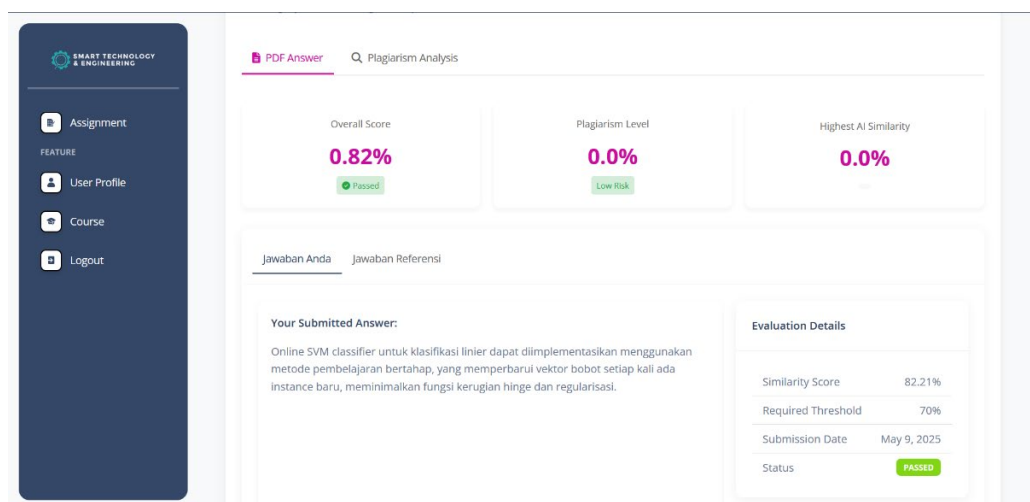
Antarmuka utama berupa chatbot interaktif yang digunakan oleh mahasiswa. Mahasiswa menjawab langsung di dalam sesi chatbot, dan sistem memberikan umpan balik secara real-time. Jika jawaban salah, sistem menjelaskan kesalahan dan memberi soal serupa. Jika benar, sistem memberikan soal pada level berikutnya. Proses ini berulang hingga mahasiswa mencapai level Analysing.



Gambar 4. Antarmuka Dashboard Mahasiswa.



Gambar 5. Antarmuka Halaman Pengerjaan.



Gambar 6. Antarmuka Ringkasan Evaluasi.

V. PENGUJIAN

A. Metode Pengujian Sistem

Pengujian sistem dilakukan dengan pendekatan *black-box* untuk memastikan seluruh fitur utama berjalan sesuai fungsinya. Setiap modul sistem diuji secara mandiri, termasuk proses unggah dokumen, pembangkitan pertanyaan, penilaian jawaban mahasiswa, pemberian umpan balik, serta logika adaptasi level pertanyaan. Pengujian dilakukan dengan menggunakan *Postman* untuk menguji *API*, serta secara langsung melalui antarmuka *chatbot* oleh mahasiswa.

B. Evaluasi Respons dan Adaptasi Sistem

Pengujian difokuskan pada respons sistem terhadap berbagai kondisi jawaban mahasiswa. Jawaban yang benar secara semantik akan meningkatkan skor dan memicu sistem untuk memberikan pertanyaan pada level berikutnya. Sebaliknya, jawaban yang kurang tepat akan menghasilkan umpan balik berupa koreksi dan pertanyaan baru dalam level yang sama. Sistem berhasil mempertahankan alur adaptif berdasarkan performa mahasiswa, dan seluruh transisi level berjalan sesuai logika yang dirancang.

C. Pengujian TAM

Sebagai pelengkap, dilakukan pengujian berdasarkan pendekatan *Technology Acceptance Model (TAM)* dengan melibatkan mahasiswa sebagai pengguna sistem. Evaluasi ini bertujuan untuk mengukur sejauh mana teknologi *Smart Tutor* diterima dan dirasakan manfaatnya dalam proses pembelajaran. Pengumpulan data dilakukan melalui kuesioner digital, dengan fokus pada persepsi kemudahan penggunaan, kegunaan sistem, dan kenyamanan antarmuka. Hasil menunjukkan bahwa sebagian besar responden merasa sistem memberikan pengalaman belajar yang menarik dan responsif.

Pengujian berbasis *Technology Acceptance Model (TAM)* dilakukan untuk mengukur tingkat penerimaan dan kenyamanan pengguna dalam menggunakan sistem *Smart Tutor*. Survei dilakukan kepada 31 mahasiswa pengguna sistem, yang mengisi kuesioner berbasis skala Likert (1–5) terhadap 14 pernyataan, dan 2 pertanyaan tambahan untuk kritik dan saran. Hasil rata-rata menunjukkan bahwa sebagian besar aspek sistem dinilai positif. Respons tertinggi muncul pada pernyataan mengenai kemudahan penggunaan dan interaksi tanpa hambatan (rata-rata > 4,0). Evaluasi terhadap validitas metode *BERTScore* dan *TF-IDF* sebagai evaluator jawaban mahasiswa juga memperoleh respons yang cukup positif (rata-rata 4,0)

Meskipun demikian, minat pengguna untuk menggunakan sistem dalam jangka panjang, seperti integrasi dengan LMS, mendapat skor lebih rendah (rata-rata 3,8), yang mengindikasikan perlunya pengembangan fitur tambahan dan peningkatan pengalaman pengguna. Secara umum, hasil pengujian TAM menunjukkan bahwa sistem diterima dengan baik, dan mendukung tujuan utama pengembangan sistem, yakni menyediakan pengalaman belajar adaptif yang nyaman, akurat, dan efisien.

D. Hasil Pengujian

Berdasarkan hasil uji coba dan observasi langsung, sistem Smart Tutor mampu menjalankan fungsinya secara optimal. Proses evaluasi berbasis semantik dapat membedakan makna jawaban mahasiswa dengan cukup akurat, dan sistem mampu mengadaptasi tingkat kesulitan soal secara real-time. Selain itu, fitur umpan balik dan perulangan soal terbukti mampu membantu mahasiswa dalam memperbaiki pemahaman terhadap materi. Pengujian ini mengonfirmasi bahwa sistem telah memenuhi tujuan utama pengembangan, yaitu menciptakan pengalaman belajar adaptif dan interaktif berbasis AI.

TABEL I
 HASIL SURVEY SYSTEM USABILITY SCALE

No	Pertanyaan Singkat	Skor Rata-rata
1	Soal mudah dipahami tanpa kebingungan.	4.5
2	Bahasa soal mudah dimengerti.	4.5
3	Sistem mudah digunakan untuk pemula.	4.1
4	Navigasi sistem intuitif.	4.2
5	Tidak butuh banyak instruksi untuk menggunakan.	4.0
6	Evaluasi sistem relevan dan adil.	4.2
7	Evaluasi dengan TF-IDF dan BERTscore valid.	4.1
8	Sistem bantu belajar lebih efektif.	4.1
9	Sistem tingkatkan pemahaman materi.	4.2
10	Akan dipertimbangkan jika ada di LMS.	3.8
11	Mudah digunakan tanpa pelatihan tambahan.	3.7
12	Tidak butuh banyak usaha/waktu.	4.1
13	Fokus pada soal tanpa gangguan sistem.	4.3
14	Tertarik menggunakan kembali.	4.0

VI. SIMPULAN DAN SARAN

A. Kesimpulan

Penelitian ini berhasil mengembangkan sistem Smart Tutor yang mampu mengevaluasi jawaban mahasiswa secara otomatis dan adaptif dengan pendekatan berbasis model bahasa besar (LLM). Dengan menggabungkan BERTScore dan TF-IDF, sistem mampu melakukan penilaian semantik terhadap jawaban mahasiswa secara lebih akurat dibandingkan metode konvensional seperti cosine similarity.

Penerapan Bloom's Taxonomy sebagai dasar penyusunan dan eskalasi tingkat pertanyaan memungkinkan sistem memberikan pengalaman belajar yang lebih terstruktur dan bertahap. Dengan logika adaptasi yang mempertimbangkan performa mahasiswa, sistem mampu menyesuaikan tingkat kesulitan pertanyaan secara dinamis, sehingga mendukung pembelajaran yang lebih personal dan responsif.

Hasil pengujian menunjukkan bahwa sistem berfungsi dengan baik, baik dari sisi teknis maupun pengalaman pengguna. Evaluasi melalui kuesioner TAM menunjukkan bahwa pengguna merasa terbantu dalam memahami materi dan menilai sistem sebagai alat pembelajaran yang efektif dan mudah digunakan.

B. Saran

Untuk pengembangan ke depan, disarankan agar sistem Smart Tutor dilengkapi dengan fitur tambahan seperti pilihan jenis soal (misalnya soal pilihan ganda), visualisasi progress pembelajaran, serta umpan balik yang lebih terperinci. Selain itu, integrasi sistem ke dalam Learning Management System (LMS) seperti Moodle atau Google Classroom dapat memperluas jangkauan dan mempermudah akses mahasiswa dalam pembelajaran mandiri. Pengujian lebih lanjut dengan jumlah pengguna yang lebih besar juga perlu dilakukan untuk mengukur skalabilitas, efektivitas sistem dalam konteks nyata, serta memperkuat validitas model evaluasi semantik yang digunakan.

DAFTAR PUSTAKA

- [1] Khairunnisa, & Wulan, Nur. Perancangan Intelligent Tutoring System sebagai Upaya Inovatif pada Pembelajaran Algoritma dan Struktur Data. Universitas Harapan Medan, 2020. [Online]. Available: https://www.researchgate.net/publication/349092065_Perancangan_Intelligent_Tutoring_System_Sebagai_Upaya_Inovatif_Pada_Pembelajaran_Algoritma_dan_Struktur_Data
- [2] Bulathwela, S., Muse, H., & Yilmaz, E. "Scalable Educational Question Generation with Pre-trained Language Models," arXiv preprint arXiv:2305.07871, 2023. [Online]. Available: <https://arxiv.org/abs/2305.07871>
- [3] Bhowmick, A. K., et al., "Automating Question Generation From Educational Text," dalam Artificial Intelligence XL. SGAI 2023. Lecture Notes in Computer Science, vol. 14381, Bramer, M., & Stahl, F. (eds), Springer, Cham, 2023, doi: 10.1007/978-3-031-47994-6_38. Available: <https://arxiv.org/abs/2309.15004>
- [4] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv preprint arXiv:1810.04805, 2019. [Online]. Available: <https://arxiv.org/abs/1810.04805>
- [5] Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. "GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding," arXiv preprint arXiv:1804.07461, 2019. [Online]. Available: <https://arxiv.org/abs/1804.07461>
- [6] Anderson, L. W., & Krathwohl, D. R. A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. Longman, 2001. [Online]. Available: https://quincycollege.edu/wp-content/uploads/Anderson-and-Krathwohl_Revised-Blooms-Taxonomy.pdf
- [7] Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. "BERTscore: Evaluating Text Generation with BERT," arXiv preprint arXiv:1904.09675, 2020. [Online]. Available: <https://arxiv.org/abs/1904.09675>
- [8] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. "Language Models are Few-Shot Learners," arXiv preprint arXiv:2005.14165, 2019. [Online]. Available: <https://arxiv.org/abs/2005.14165>
- [9] Hussein, M. A., Hassan, H., & Nassef, M. "Automated Language Essay Scoring Systems: A Literature Review," PeerJ Computer Science, vol. 5, e208, 2019. [Online]. Available: <https://peerj.com/articles/cs-208/>
- [10] Kurdi, M. H., Al-Madi, N., & Al-Khatib, W. "A Systematic Review of Automated Question Generation for Educational Purposes," International Journal of Artificial Intelligence in Education, vol. 30, pp. 121–158, 2020. [Online]. Available: <https://link.springer.com/article/10.1007/s40593-019-00186-y>
- [11] Tamkin, A., Brundage, M., Ganguli, D., & Clark, J. "Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models," arXiv preprint arXiv:2102.02503, 2021. [Online]. Available: <https://arxiv.org/abs/2102.02503>
- [12] Hwang, G.-J., & Chang, S.-C. "A review of opportunities and challenges of chatbots in education," Smart Learning Environments, vol. 7, no. 1, pp. 1–14, 2020. doi: 10.1186/s41039-019-0098-z. Available: <https://www.tandfonline.com/doi/full/10.1080/10494820.2021.1952615>