

BAB II

LANDASAN TEORI

2.1 Penelitian Terdahulu

Pada sub bab ini, dibahas penelitian-penelitian terdahulu yang relevan dengan topik penelitian ini, khususnya yang berkaitan dengan penggunaan *Small Language Models* dalam berbagai aplikasi seperti *text generation*, *question answering*, dan penerapan dalam konteks pendidikan. Penelitian-penelitian ini memberikan wawasan penting mengenai penerapan model seperti *GPT-2*, *ALBERT*, dan *ELECTRA-small* untuk menghasilkan teks berkualitas dan relevan, termasuk soal-soal *Higher-Order Thinking Skills* yang menuntut analisis, evaluasi, dan kreasi. Tabel berikut menyajikan ringkasan dari berbagai penelitian terdahulu yang mengkaji penggunaan model-model ini, serta hasil yang dicapai, yang memberikan dasar penting bagi penelitian ini dalam mengembangkan model untuk meningkatkan kualitas pembelajaran otomatis.



Tabel 2. 1 Penelitian Terdahulu

| No | Tahun | Judul | Peneliti | Model SLM | Hasil |
|----|-------|---------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------|----------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. | 2024 | <i>Investigating Data Efficient Methods for Enhancing Model Generalization using Dataset Cartography, Augmentation and Perturbation Techniques [27]</i> | Gupta & Rekha | ELECTRA-small (google/electra-small-discriminator) | Penelitian menilai dampak pelatihan data-efisien. Model baseline mencapai F1 in-domain 85,2 dan EM 77,4. Pelatihan pada subset variabilitas tinggi plus augmentasi (parafrase & adversarial) menaikkan F1 out-of-domain ke 27,9 dengan EM 18,4, melampaui baseline 18,0. |
| 2. | 2021 | <i>Fine-TOM matcher results for OAEI 2021 [28]</i> | Leon Knorr & Jan Portisch | ALBERT-base-v2 (fine-tuned) | Fine-TOM menggunakan model ALBERT-base-v2 yang difine-tune. Pada track Anatomy, sistem mencatat precision 0,916, recall 0,794 dan F-measure 0,851. Pada track Conference, precision 0,64, recall 0,53, F-measure 0,58. Pada track Knowledge Graph, precision 0,92, recall 0,75 dan F-measure 0,83. |
| 3. | 2025 | <i>ChatGPT vs. Human Journalists: Analyzing News Summaries Through BERTScore and Moderation Standards [29]</i> | Kim, H.-S.; Kang, J.-W.; Choi, S.-Y. | GPT 2 | BERTScore F1 > 0,87; kemiripan konteks tinggi; bahasa lebih netral |
| 4. | 2025 | <i>Chinese Story Generation Based on Style Control of Transformer Model and Content Evaluation Method [30]</i> | Lin, J.-W.; Su, T.-W.; Chang, C.-C. | Transformer (BERT, GPT) | Cycle-SG GAN mencapai skor tertinggi (rata-rata 4,13); keseimbangan keterbacaan 3,3–4,6; masih perlu peningkatan kelancaran |
| 5. | 2024 | <i>How does GPT-2 Predict Acronyms? Extracting and Understanding</i> | Jorge García-Carrasco, | GPT-2 Small | Penelitian mekanistik ini menemukan sirkuit yang terdiri dari 8 head atensi (~5 % dari total |

| No | Tahun | Judul | Peneliti | Model SLM | Hasil |
|----|-------|-------------------------------------------------------------------------------------------------------------------------------------|--------------------------------|---------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | | <i>ing a Circuit via Mechanistic Interpretability</i> [31] | Alejandro Maté & Juan Trujillo | | head) pada GPT-2 Small yang bertanggung jawab atas prediksi akronim tiga huruf. Ablasi head lain tidak menurunkan kinerja—bahkan sedikit meningkat—menunjukkan bahwa sirkuit ini mengumpulkan fungsi prediksi akronim. Sirkuit memanfaatkan informasi posisi yang disediakan oleh mask kausal untuk menyalin huruf kapital secara tepat. |
| 6. | 2023 | <i>Unleash GPT-2 Power for Event Detection</i> [32] | Zhang et al. | GPT-2 teacher–student (GPTEDOT) | Arsitektur teacher–student GPTEDOT melampaui baseline pada beberapa dataset: pada ACE 2005, precision 82,3, recall 76,3 dan F1 79,2; pada CySecED, precision 65,9, recall 64,1, F1 65,0; pada RAMS, precision 55,5, recall 78,6, F1 65,1. |
| 7. | 2023 | <i>Comparison of Evaluation Metrics for Short Story Generation</i> [33] | Netisopakul & Taoto | N-gram, CBOW, GRU, GPT-2 | BLEU/ROUGE/BERTScore berkorelasi kuat; WMD berkorelasi negatif; Self-BLEU tidak berkorelasi; GPT-2 (pretrained/finetuned) unggul dalam keragaman & kemiripan |
| 8. | 2023 | <i>Application of Transformer-Based Deep Learning Models for Predicting the Suitability of Water for Agricultural Purposes</i> [22] | Khanday et al. | ALBERT Base v2 & ALBERT-WPD | Model baseline ALBERT Base v2 memperoleh akurasi 91 % dengan F1 0,91. Setelah dioptimasi, varian khusus ALBERT-WPD mencapai akurasi 95,88 % (\approx 96 %) dan meningkatkan presisi serta recall. |

| No | Tahun | Judul | Peneliti | Model SLM | Hasil |
|-----|-------|-----------------------------------------------------------------------------------------------------|-------------------------------------|--------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 9. | 2024 | <i>Using of Transformers Models for Text Classification to Mobile Educational Applications</i> [23] | Gizatdinova et al. | RoBERTa-base, BERT-base-cased/uncased, ALBERT-base-v2, DistilBERT-base-uncased | Tabel evaluasi menunjukkan bahwa pada epoch 4, roberta-base memperoleh akurasi 81 % dan F1-score 80 %. bert-base-cased mencapai akurasi 80 % (F1 77 %), bert-base-uncased 79 % (F1 77 %), albert-base-v2 78 % (F1 74 %) dan distilbert-base-uncased 76 % (F1 71 %). |
| 10. | 2025 | <i>Smart Building Recommendations with LLMs: A Semantic Comparison Approach</i> [25] | Papaioannou, Korkas & Kosmatopoulos | GPT-2, GPT-4, DeepSeek-R1-Distill Qwen | Rekomendasi LLM menurunkan konsumsi listrik ~10% dalam uji dunia nyata; model terbuka mampu menyamai/mengungguli GPT-4 dengan biaya lebih rendah |

Penelitian yang dilakukan oleh Rohling berjudul *Investigating Data Efficient Methods for Enhancing Model Generalization using Dataset Cartography, Augmentation and Perturbation Techniques* mengupas strategi hemat data untuk meningkatkan generalisasi model *ELECTRA-small-discriminator* dalam tugas *Question Answering*. Dengan baseline F1 dalam domain sebesar 85,2 dan EM 77,4 pada dataset *Stanford Question Answering Dataset (SQuAD)*, performa model menurun signifikan pada dataset *out-of-domain* seperti *AdversarialQA*. Melalui pendekatan Dataset *Cartography* yang diperluas dengan *Sampling with Perturbation for Uncertainty Quantification*, penelitian ini berhasil mengidentifikasi contoh pelatihan yang memiliki variabilitas tinggi dan melatih model menggunakan subset tersebut. Dengan tambahan augmentasi data berupa *paraphrase* dan *adversarial*, F1-score untuk *out-of-domain* meningkat menjadi 27,9 (EM 18,4), yang melampaui baseline 18,0. Hasil ini menguatkan ide bahwa dengan fokus pada contoh yang ambigu dan penerapan augmentasi data yang tepat, model kecil seperti *ELECTRA* dapat lebih tahan terhadap pergeseran distribusi data.

Keberhasilan pendekatan ini juga terlihat dalam beberapa studi lainnya, terutama yang memanfaatkan model *A Lite Bidirectional Encoder Representations from Transformers (ALBERT)*. Penelitian oleh Knorr & Portisch dalam *Fine-TOM Matcher Results for OAEI 2021* menunjukkan bagaimana *ALBERT-base-v2* yang di-fine-tune menggunakan hanya 20% data referensi dapat menghasilkan *F-measure* hingga 0,851 pada track *Anatomy*. Rejini et al. juga menunjukkan keberhasilan *ALBERT* dalam memprediksi potabilitas air dengan akurasi hingga 96% setelah optimasi, jauh melampaui baseline 91%. Ketiga penelitian ini menunjukkan bahwa *ALBERT*, melalui fine-tuning ringan dan optimasi arsitektur, mampu memberikan performa tinggi pada tugas yang sangat spesifik, namun tetap efisien dalam penggunaan sumber daya.

Di sisi lain, penggunaan *Generative Pre-trained Transformer 2 (GPT-2)* dalam generasi teks dan deteksi peristiwa juga telah terbukti efektif, seperti yang terlihat pada penelitian *Unleash GPT-2 Power for Event Detection*. Dalam penelitian ini, arsitektur *teacher-student berbasis GPT-2* berhasil melampaui

model *Bidirectional Encoder Representations from Transformers Event Detection (BERT-ED)* dengan F1 79,2 pada dataset *Automatic Content Extraction 2005 (ACE 2005)*. Sementara itu, García-Carrasco et al. mengungkapkan melalui *mechanistic interpretability* bahwa sirkuit atensi dalam *GPT-2* berperan penting dalam memprediksi akronim tiga huruf. Penelitian lain oleh Netisopakul & Taoto menunjukkan bahwa *GPT-2* unggul dalam generasi cerita pendek dengan korelasi tinggi antara *Bilingual Evaluation Understudy (BLEU)*, *Recall-Oriented Understudy for Gisting Evaluation (ROUGE)*, dan *BERTScore*, serta menyoroti pentingnya penggunaan metrik evaluasi yang tepat. Ketiga penelitian ini menegaskan bahwa *GPT-2* memiliki fleksibilitas yang luar biasa dalam berbagai aplikasi, baik untuk deteksi peristiwa maupun generasi teks, dan pentingnya memahami mekanisme internal model ini untuk meningkatkan hasil yang diinginkan.

Selanjutnya, penelitian oleh Papaioannou et al. tentang *Smart Building Recommendations with Large Language Models (LLMs): A Semantic Comparison Approach* menunjukkan bagaimana model *DeepSeek* dapat memberikan rekomendasi bangunan pintar dengan performa yang setara atau bahkan melampaui *GPT-4*, namun dengan biaya inferensi yang jauh lebih rendah. Hal ini sangat relevan untuk aplikasi dunia nyata, di mana efisiensi biaya dan skalabilitas menjadi faktor kunci, terutama dalam industri energi dan manajemen gedung. Studi ini menunjukkan bahwa model *open-access* seperti *DeepSeek* dapat mencapai kinerja yang lebih efisien dengan biaya yang lebih rendah, yang tentunya dapat menjadi acuan bagi penelitian lain, termasuk aplikasi dalam pengembangan model untuk edukasi berbasis *Artificial Intelligence (AI)*.

Lin et al., mengembangkan metode untuk generasi cerita berbasis *Transformer* dengan kontrol gaya dan evaluasi konten, yang meningkatkan kelancaran dan keterbacaan teks dengan menggunakan kecocokan *Part-of-Speech (PoS)* dan *fuzzy matching*. Pendekatan ini sangat relevan dalam konteks pembuatan soal HOTS berbasis cerita di Desa Pete, di mana konteks lokal dan kelancaran naratif menjadi penting dalam merancang soal yang tidak hanya menguji ingatan, tetapi juga keterampilan analitis dan kreatif siswa.

Terakhir, penelitian oleh Kim et al., membandingkan ringkasan berita yang dihasilkan ChatGPT dengan yang ditulis oleh jurnalis *The New York Times*. *ChatGPT* menunjukkan kemampuan untuk menghasilkan ringkasan dengan tingkat keselarasan semantik yang tinggi dan menghindari bias serta konten inflamasi, yang membuatnya menjadi alat yang efektif dan etis dalam otomatisasi pembuatan konten berita. Penelitian ini semakin memperkuat relevansi *GPT-2* dan model serupa dalam aplikasi praktis, termasuk pembuatan soal yang mendalam dan seimbang secara etis.

Skripsi ini berbeda dari penelitian-penelitian tersebut karena berfokus pada konteks lokal Desa Pete dan edukasi berbasis kearifan lokal. Dataset yang digunakan berasal dari tiga video tutorial budidaya angsa dengan total durasi sekitar 27 menit; transkripnya dibersihkan dan dipilih sepuluh kalimat kunci untuk dijadikan pasangan konteks–pertanyaan. Model yang dilatih merupakan *Small Language Model GPT-2 Small*, *ELECTRA-Small*, dan *ALBERT-base-v2* yang ditune-ulang dengan parameter kecil *5 epoch*, *batch size 2*, *learning rate 5e-5* supaya cocok dijalankan pada perangkat dengan sumber daya terbatas. Selain itu, sistem ini diintegrasikan ke website resmi Desa Pete melalui framework *Laravel* dan *FastAPI* sehingga pengguna dapat mengakses generator soal HOTS langsung dari tombol di situs desa. Data lokal seperti kegiatan budidaya angsa dan penduduk dimasukkan ke dalam basis data kecamatan. Integrasi ini menjadikan skripsi ini pionir dalam menerapkan pipeline CRISP-DM secara utuh untuk pembuatan soal HOTS berbasis kearifan lokal yang dapat berjalan di lingkungan *low-resource*, menawarkan kontribusi praktis bagi Desa Pete serta bukti teoretis bahwa SLM kecil cukup andal untuk konteks tersebut.

2.2 Teori yang berkaitan

2.2.1 Higher-Order Thinking Skills

Higher-Order Thinking Skills merujuk pada kemampuan berpikir tingkat tinggi yang mencakup tiga level utama dalam taksonomi berpikir: analisis, evaluasi, dan kreasi. Pada level analisis, siswa diharapkan dapat memecah informasi menjadi bagian-bagian yang lebih kecil dan memahami hubungan antar bagian tersebut. Evaluasi melibatkan

kemampuan untuk menilai dan memverifikasi informasi atau argumen berdasarkan kriteria tertentu, seperti validitas atau keandalan bukti. Sedangkan kreasi mengharuskan siswa untuk menghasilkan ide atau solusi baru yang orisinal, berdasarkan pemahaman yang mendalam. HOTS sangat penting dalam mempersiapkan siswa untuk menghadapi tantangan dunia nyata yang kompleks, di mana kemampuan berpikir kritis dan kreatif menjadi kunci dalam pemecahan masalah dan pengambilan keputusan [34].

Pengembangan HOTS tidak hanya mengandalkan pembelajaran pasif, melainkan melibatkan siswa sebagai aktor utama dalam proses belajar, dengan guru berperan sebagai fasilitator yang merancang pembelajaran yang menantang dan relevan. Salah satu cara untuk menumbuhkan HOTS adalah melalui soal-soal yang mendorong analisis, evaluasi, dan kreasi, meskipun tidak ada penandaan level Bloom secara eksplisit pada soal. Dengan pendekatan seperti ini, soal yang dihasilkan bertujuan untuk mengasah keterampilan berpikir kritis dan kreatif siswa dalam konteks lokal, seperti memahami konsep-konsep dalam budaya angsa dan aplikasi nyata lainnya [35]. Pendekatan pembelajaran aktif seperti pemecahan masalah berbasis kasus nyata dan proyek inkuiiri semakin memperkuat pengembangan HOTS melalui refleksi dan keterlibatan siswa dalam proses pembelajaran yang lebih mendalam. Temuan ini juga mendukung ide bahwa desain soal HOTS berbasis pembelajaran aktif dapat meningkatkan kualitas berpikir kritis siswa [36].

2.2.2 Desa Wisata

Banyak desa mengembangkan desa wisata sebagai cara meningkatkan kesejahteraan dan melestarikan budaya. Di Desa Nyatnyono, digitalisasi melalui situs dan media sosial digunakan untuk mempromosikan wisata religi dan kuliner [26], sedangkan penelitian di Desa Cupak menyoroti pentingnya partisipasi masyarakat, pemenuhan unsur wisata (atraksi, transportasi, akomodasi, pelayanan, infrastruktur)

dan pengelolaan tradisi lokal [24]. Pemetaan SDM di Desa Giri Gresik menegaskan bahwa keberhasilan desa wisata bergantung pada kualitas sumber daya manusia kecerdasan, spiritualitas, integritas yang mampu memberikan layanan ramah dan berkarakter [37], sehingga pengembangan desa wisata perlu mengintegrasikan teknologi, pemberdayaan komunitas, dan peningkatan kapasitas SDM.

2.2.3 Artificial Intelligence

AI merujuk pada bidang dan teori yang berupaya membangun mesin dengan kemampuan belajar, memecahkan masalah dan mengenali pola layaknya manusia. Telaah literatur menjelaskan bahwa teknik seperti jaringan saraf, *fuzzy logic* dan sistem hibrida serta domain seperti expert systems, pemrosesan bahasa alami dan robotika menopang kemampuan ini; AI bahkan disebut lebih tahan lama dan dapat diandalkan dibanding kecerdasan manusia [38]. Dalam pemasaran digital, AI memanfaatkan teknik tersebut untuk menganalisis data besar, mempersonalisasi konten dan menargetkan iklan demi meningkatkan keterlibatan pelanggan [39]. Lembaga pendidikan dan organisasi lain dapat mengadopsi AI untuk menghadirkan pembelajaran adaptif, sementara AI sendiri mendapat manfaat dari praktik pengelolaan data institusional [40].

Tinjauan sistematis menunjukkan bahwa AI merevolusi layanan kesehatan dengan mengotomatiskan tugas administratif, mempercepat diagnosis dini dan meningkatkan kualitas perawatan [41]. Namun, pengaplikasian ini menimbulkan isu etika dan sosial, terutama perlindungan privasi serta keamanan data pasien. Untuk itu, kerangka kerja yang diusulkan oleh Villegas-Ch dkk. menekankan perlunya perlindungan data, penguatan keamanan sistem dan penyesuaian dengan regulasi guna menangkal manipulasi dan penyalahgunaan data. Secara keseluruhan, literatur mengingatkan bahwa manfaat AI harus diimbangi dengan strategi keamanan, privasi, dan etika yang kuat [42].

2.2.4 Fine Tuning

Fine-tuning adalah proses penyesuaian model *pre-trained* (model yang telah dilatih dengan dataset besar sebelumnya) untuk tugas spesifik

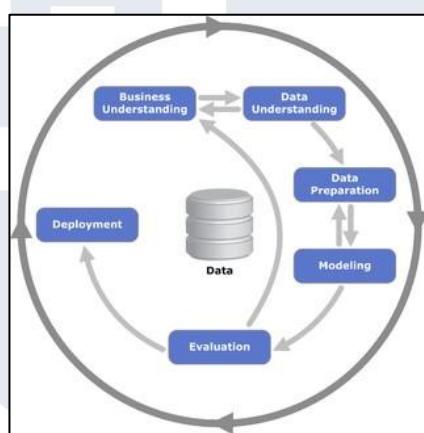
yang lebih sempit, menggunakan dataset yang lebih kecil dan lebih relevan. Proses ini mengadaptasi model umum yang telah mempelajari pola umum dalam data besar, agar lebih tepat dan optimal untuk aplikasi atau domain tertentu, seperti pemrosesan bahasa atau pengenalan gambar. Sebagai contoh, model seperti BERT atau GPT yang sudah dilatih pada korpus besar, dapat disesuaikan untuk kebutuhan spesifik dengan fine-tuning agar dapat menghasilkan output yang lebih akurat dalam konteks yang lebih terbatas, seperti soal HOTS pada topik pendidikan atau aplikasi medis [43].

1. *Pre-trained Model* menjadi langkah awal dalam fine-tuning karena model yang sudah dilatih sebelumnya memiliki pemahaman dasar yang kuat terhadap pola bahasa atau citra yang umum [44]. Dengan memanfaatkan model ini, kita tidak perlu memulai pelatihan dari awal, yang tentu lebih memakan waktu dan biaya komputasi. Fine-tuning memungkinkan kita untuk menyesuaikan model tersebut dengan data yang lebih kecil namun lebih relevan, sehingga menghasilkan model yang lebih efektif tanpa harus membangun dari nol.
2. *Data Efficient Fine-Tuning* berfokus pada pengoptimalan penggunaan data yang lebih sedikit dengan memilih sampel yang paling relevan dan representative [43]. Teknik ini bertujuan untuk mengurangi biaya pelatihan dengan memilih data yang memiliki dampak besar pada kinerja model, sehingga fine-tuning menjadi lebih cepat dan efisien. Hal ini memungkinkan pengembangan model yang lebih baik meskipun dengan sumber daya data yang terbatas, serta mengurangi waktu pelatihan tanpa mengurangi kualitas model yang dihasilkan.

2.3 Framework/Algoritma yang digunakan

2.3.1 Crisp DM

Metode *Cross-Industry Standard Process for Data Mining* telah banyak digunakan dalam berbagai aplikasi untuk menangani alur kerja dalam proses data mining. Sebagai contoh, penelitian yang dilakukan oleh Abdurrazzaq et al. (2025) menerapkan CRISP-DM untuk meningkatkan sistem pencarian informasi medis, yang mengintegrasikan semantic similarity dengan SLMs dalam mendeteksi penyakit [45]. Mereka mengadopsi enam tahap dalam CRISP-DM, mulai dari pemahaman bisnis untuk memahami konteks dan tujuan aplikasi, hingga penerapan model dalam deployment.



Gambar 2. 1 Crisp DM [46]

Penjelasan gambar ini menunjukkan siklus CRISP-DM, mulai dari pemahaman bisnis hingga deployment, yang dapat diterapkan dalam penelitian ini. Dengan menggunakan pendekatan ini, pastinya dapat mengolah transkrip untuk menghasilkan soal yang sesuai dengan kurikulum dan tujuan pembelajaran, tanpa bergantung pada sistem berbasis LLM eksternal. Di sisi lain, penelitian yang dilakukan oleh Cevallos (2024) tentang pengklasifikasian berita digital menggunakan teknik *machine learning* juga menggunakan CRISP-DM sebagai metodologi untuk mengatur alur proses dari awal hingga akhir [46]. Dimulai dengan *data understanding* melalui pengumpulan berita dari sumber-sumber yang berbeda, sistem ini menyimpan data dalam data lake dan melakukan *pre-processing*, *modeling*, serta evaluasi data. Tahapan data preparation dan modeling dalam CRISP-DM akan menjadi

dasar dalam mempersiapkan data transkrip dan menerapkan model untuk menghasilkan soal yang sesuai dengan tingkat kesulitan dan tujuan pembelajaran siswa.

2.3.2 Small Language Model

Small Language Models muncul sebagai solusi untuk mengatasi masalah komputasi dan privasi yang dihadapi oleh *Large Language Models*. Mereka menggunakan teknik kompresi seperti pruning, distilasi pengetahuan, dan kuantisasi untuk mengurangi ukuran model tanpa mengorbankan performa secara signifikan [47]. Dalam aplikasi dunia nyata, SLMs memiliki keunggulan dalam kecepatan inferensi dan biaya yang lebih rendah, menjadikannya ideal untuk tugas yang memerlukan penyesuaian cepat dan efisien, terutama pada perangkat mobile atau aplikasi berbasis edge [48]. Dengan demikian, meskipun SLMs tidak memiliki kemampuan emergent yang dimiliki oleh LLMs yang lebih besar, mereka menawarkan keunggulan dalam hal penggunaan sumber daya dan privasi data, yang semakin penting dalam konteks aplikasi spesifik seperti kesehatan dan keuangan.

Secara umum, metode yang digunakan untuk mengoptimalkan SLM termasuk teknik-teknik kompresi yang bertujuan untuk mengurangi ukuran model tanpa mengorbankan kemampuan generalisasi. Persamaan matematika yang digunakan dalam pruning dan distilasi pengetahuan, misalnya, dapat dijelaskan sebagai berikut:

$$L_{pruned} = L_{full} - \sum_{i \in P} \Delta W_i$$

Rumus 2. 1 Rumus Pruning [49]

Dimana L_{pruned} adalah kehilangan (loss) setelah pruning, L_{full} adalah kehilangan pada model penuh, dan ΔW_i adalah perubahan bobot pada parameter yang dihapus selama proses pruning.

$$L_{distilled} = \lambda L_{teacher} + (1 - \lambda) L_{student}$$

Rumus 2. 2 Rumus Distilled [49]

Dimana $L_{distilled}$ adalah kehilangan gabungan, $L_{teacher}$ adalah kehilangan model pengajar, dan $L_{student}$ adalah kehilangan model pelajar, dengan λ sebagai parameter yang mengontrol kontribusi dari kedua model.

2.3.3 GPT 2

GPT-2 adalah model bahasa berbasis *Transformer* yang menunjukkan kemampuannya dalam berbagai tugas pemrosesan bahasa alami, termasuk prediksi teks dan bahkan operasi matematika dasar seperti perbandingan *greater-than* meskipun tidak dilatih secara eksplisit untuk itu [50]. Penelitian menunjukkan bahwa *GPT-2* dapat memprediksi hubungan antara angka dengan menggunakan struktur internal seperti multi-layer perceptrons dan kepala perhatian [31]. Kemampuan ini dapat diterapkan dalam pembuatan soal HOTS untuk anak SD, khususnya dalam soal logika dan matematika sederhana yang menguji pemahaman konsep seperti perbandingan angka. Dengan demikian, *GPT-2* dapat digunakan untuk menghasilkan soal yang mendalam keterampilan berpikir kritis dan analitis anak-anak, sambil memudahkan pembuatan soal yang sesuai dengan tingkat pemahaman mereka.

2.3.4 Electra-small

ELECTRA-small adalah model bahasa *pre-training* yang didesain untuk efisiensi dalam memanfaatkan data pelatihan. Dibandingkan dengan model *pre-training* lainnya seperti *BERT*, yang menggunakan metode *masked language modeling*, *ELECTRA-small* mengusung pendekatan *replaced token detection*, di mana sebagian token dalam input diganti dengan token yang salah dan model dilatih untuk membedakan token asli dan yang diganti. Pendekatan ini memungkinkan model untuk lebih cepat mempelajari representasi kata, karena token yang diganti secara sengaja memberikan informasi lebih kaya daripada hanya memprediksi token yang dimasked [27]. Dalam konteks pembuatan soal untuk edukasi, *ELECTRA-small* dapat digunakan untuk menghasilkan soal *Higher Order Thinking Skills* yang membutuhkan analisis dan pemahaman mendalam. Model ini mampu memahami

konteks dan menghasilkan pertanyaan yang menguji kemampuan berpikir kritis siswa, seperti dalam soal esai atau analisis yang melibatkan pemahaman konsep-konsep abstrak [51]. Keefektifan *ELECTRA-small* dalam menangani berbagai tugas *NLP*, termasuk pengklasifikasian teks dan analisis konteks, menjadikannya alat yang sangat berguna dalam pengembangan sistem pembuatan soal otomatis, memberikan efisiensi baik dari segi waktu pelatihan maupun komputasi.

2.3.5 Albert-Base-V2

ALBERT-base-v2 adalah model *transformer* yang dirancang untuk efisiensi komputasi dengan mengurangi jumlah parameter tanpa mengorbankan kinerja. Model ini terbukti efektif dalam berbagai tugas *NLP*, termasuk deteksi kualitas air untuk pertanian dengan akurasi tinggi hingga 96% [52]. Keunggulannya yang efisien dalam memproses teks memungkinkan *ALBERT-base-v2* digunakan untuk menghasilkan soal HOTS untuk anak SD. Model ini dapat memproduksi soal yang menguji pemahaman dan analisis, seperti pertanyaan esai atau pilihan ganda, dengan bahasa yang mudah dipahami dan sesuai dengan tingkat kemampuan kognitif anak-anak, sehingga memfasilitasi pembelajaran yang lebih mendalam. Selain itu, dengan kemampuannya dalam memahami konteks secara mendalam, *ALBERT-base-v2* mampu menghasilkan pertanyaan yang bukan hanya menguji ingatan, tetapi juga mendorong siswa untuk berpikir kritis dan kreatif, yang sangat penting untuk perkembangan keterampilan berpikir tingkat tinggi pada usia dini [22].

2.4 Tools/software yang digunakan

2.4.1 Jupyter Notebook

Jupyter Notebook merupakan platform komputasi interaktif yang menggabungkan kode, visualisasi, serta penjelasan teks dalam satu dokumen sehingga sangat bermanfaat untuk pendidikan dan penelitian. Menurut Bascuña et al. (2023), *Jupyter Notebook* mampu meningkatkan proses pembelajaran di bidang teknik kimia karena memberikan pengalaman belajar berbasis pemodelan, visualisasi, dan perhitungan

secara langsung, membuat mahasiswa lebih mudah memahami konsep yang abstrak [53]. Sementara itu, Owusu et al. (2022) melalui pengembangan *PyGEE-SWToolbox* menunjukkan bahwa *Jupyter Notebook* juga efektif digunakan untuk pemetaan dan analisis sumber daya air berbasis *Google Earth Engine*, sekaligus memudahkan pengguna non-programmer dalam mengakses dan mengolah data satelit [54]. Dengan fleksibilitas tersebut, *Jupyter Notebook* relevan digunakan dalam penelitian ini sebagai sarana untuk mengolah transkrip budidaya hewan melalui tahapan CRISP-DM hingga menghasilkan soal HOTS yang sesuai dengan konteks pembelajaran di Desa Pete.

2.4.2 Python

Python adalah bahasa pemrograman tingkat tinggi yang sangat populer di kalangan ilmuwan data dan pengembang perangkat lunak berkat sintaksisnya yang jelas dan ekosistem yang kaya. Correa de Barros dan Martins (2024) menunjukkan bahwa *Python* menjadi pilihan utama dalam pengembangan perangkat lunak karena kemudahan penggunaannya dalam mengembangkan aplikasi berbasis AI, khususnya dalam kode otomatis dengan menggunakan SLM untuk meningkatkan produktivitas pengembangan perangkat lunak [55]. Selain itu, Öztürk (2024) menekankan bagaimana *Python* dapat digunakan dalam toolkit seperti *XCompress* yang menggabungkan kompresi teks berbasis SLM untuk memilih algoritma kompresi yang paling efisien berdasarkan jenis data, menunjukkan fleksibilitas *Python* dalam aplikasi kompresi data yang lebih cepat dan efisien [56]. Dengan keberagaman pustaka yang tersedia, seperti NumPy untuk komputasi numerik dan *SciPy* untuk pemrograman ilmiah, *Python* memungkinkan pengembang untuk memanfaatkan berbagai metode analisis dan kompresi, sesuai dengan kebutuhan spesifik dalam berbagai konteks, termasuk dalam penelitian dan pendidikan [57]. Keunggulan *Python* ini sangat relevan dengan pengembangan alat dan sistem seperti generator soal HOTS berbasis transkrip di penelitian saya, yang memanfaatkan kemudahan integrasi pustaka dan algoritma dalam *Python* untuk analisis data dan pemodelan.

2.4.3 Visual Studio Code

Visual Studio Code merupakan editor kode sumber terbuka yang ringan, dilengkapi *IntelliSense*, *debugging* terintegrasi, dan ribuan ekstensi, sehingga menjadi pilihan utama dalam pengembangan aplikasi Laravel berbasis MVC [58]. VS Code juga mendukung pendekatan *model-driven architecture* melalui pengeditan metamodel *Ecore* serta aturan transformasi ATL untuk menghasilkan kode Laravel secara otomatis dengan tetap mematuhi standar MVC2 [59]. Dalam skripsi ini, VS Code digunakan sebagai environment utama untuk mengembangkan *backend* *FastAPI*, mengelola model *GPT-2 Small*, serta mengintegrasikan API dengan antarmuka Streamlit, sekaligus menjadi fondasi migrasi ke Laravel pada pengembangan selanjutnya agar aplikasi “Belajar Seru Bersama AI” lebih mudah dilanjutkan oleh mahasiswa Sistem Informasi.

2.4.4 Laragon

Laragon merupakan perangkat lunak pengembangan web lokal berbasis Windows yang berfungsi sebagai server lengkap untuk menjalankan aplikasi *PHP*, *MySQL*, dan *Apache/Nginx* secara terintegrasi, memungkinkan developer membangun dan menguji sistem web seperti pengumpulan data anggota lembaga tanpa bergantung pada server eksternal selama tahap awal [60]. Dengan fitur seperti *auto-virtualhost*, *pretty URLs*, dan dukungan SSL otomatis, *Laragon* mempercepat proses pengembangan *end-to-end* dari perencanaan hingga evaluasi, serta memfasilitasi migrasi ke hosting produksi dengan konfigurasi sederhana dan portabilitas tinggi, sehingga ideal untuk proyek seperti sistem pengumpulan data berbasis web yang memerlukan keamanan dan efisiensi dalam pengelolaan basis data *MySQL/MariaDB*.