



**Pengembangan Sistem Tinjauan Makalah dengan Large
Language Models**

UJIAN KUALIFIKASI

Utami Lestari

99223141

**PROGRAM DOKTOR TEKNOLOGI INFORMASI
UNIVERSITAS GUNADARMA
2024**

Daftar Isi

Daftar Isi.....	ii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Batasan dan Rumusan Masalah.....	3
1.2.1 Batasan Masalah.....	3
1.2.2 Rumusan Masalah	4
1.3 Tujuan Penelitian.....	4
1.4 Kontribusi dan Manfaat Penelitian.....	4
BAB II TELAAH PUASTAKA	5
2.1 Pemrosesan Bahasa Alami	5
2.2 Deep Learning.....	6
2.3 Transformers	8
2.4 Model Bahasa Besar (LLM)	9
2.4.1 Prompt Learning.....	12
2.4.2 Architecture	13
2.4.3 Generative Pre-trained Transformer (GPT)	15
2.4.4 Bidirectional Encoder Representations from Transformers (BERT) ...	15
2.5 Penelitian Terdahulu.....	16
2.6 Peer Review	21
BAB III METODOLOGI PENELITIAN.....	22
3.1 Gambaran Umum.....	22
3.1.1 Pengumpulan data	23
3.1.2 Preprocessing data.....	23
3.1.3 Pembuatan Model LLM	24
3.1.4 Evaluasi Model LLM	25
3.1.5 Validasi Ahli	25
3.2 Jadwal Penelitian.....	25
DAFTAR PUSTAKA	27

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi telah menjadi pendorong utama transformasi digital yang telah mengubah pola kehidupan manusia secara fundamental. Teknologi informasi telah menyebar ke berbagai sektor serta memfasilitasi akses ke berbagai informasi, mempercepat pertukaran data, dan mengubah cara manusia berinteraksi. Saat ini teknologi bukan hanya sekedar kebutuhan sekunder namun dapat dikatakan sebagai kebutuhan primer. Perkembangan teknologi mempengaruhi berbagai aspek kehidupan seperti sosial, politik dan budaya. Salah satu bentuk dari perkembangan teknologi adalah kecerdasan buatan atau *artificial intelligence* (AI).

Kecerdasan buatan atau AI mengacu pada aplikasi algoritma dan teknik perangkat lunak yang memungkinkan komputer dan mesin untuk mensimulasikan persepsi manusia dan proses pengambilan keputusan untuk menyelesaikan tugas dengan sukses (Murphy, 2019). Teknologi kecerdasan buatan (AI) berkembang mengiringi perkembangan teknologi informasi, hal ini membawa perubahan diberbagai bidang seperti pendidikan, kesehatan, ekonomi, industry, dan transportasi. Teknologi AI menawarkan berbagai kemudahan, efisiensi, analisis data yang mendalam dan pengenalan pola. Perkembangan kecerdasan buatan (AI) telah memainkan peran krusial dalam menghadirkan kemajuan signifikan diberbagai aplikasi. Melalui teknik-teknik seperti deep learning dan machine learning, AI mampu menganalisis data yang kompleks dan mengidentifikasi pola-pola yang tersembunyi di dalamnya. Pada pengenalan gambar misalnya, AI telah mencapai tingkat ketepatan yang mengesankan dalam mengenali objek, wajah, atau bahkan pola-pola mikroskopis dalam citra medis. pada bidang teks, AI dapat mengenali pola-pola kompleks dalam teks seperti struktur gramatikal, entitas, dan makna kata. AI juga dapat mengidentifikasi pola-pola seperti opini atau sentimen dalam teks, topik pembicaraan, atau bahkan pemahaman konteks dari suatu kalimat.

Pertumbuhan pesat penelitian ilmiah dan produksi artikel ilmiah dalam berbagai disiplin ilmu menghasilkan tantangan baru dalam proses tinjauan artikel ilmiah. Volume besar publikasi membuat tugas penelaahan sejawat (*peer-review*) semakin kompleks dan memakan waktu yang lebih banyak. Jumlah naskah yang diajukan ke jurnal untuk proses *peer-review* mengalami pertumbuhan tahunan sebesar 6,1% (Checco et al., 2021). Sementara itu, kebutuhan untuk memastikan kualitas dan keakuratan penilaian tetap menjadi prioritas. Adanya pertumbuhan pengajuan naskah untuk di tinjau menimbulkan potensi bias dan konsistensi penelaahan yang kurang baik. Sehingga terdapat peluang untuk mengembangkan suatu model yang mampu melakukan tinjauan secara mandiri dengan bantuan kecerdasan buatan.

Pada saat yang sama, perkembangan kecerdasan buatan dalam bidang pemrosesan teks atau yang dikenal dengan *Natural Language Processing (NLP)* telah membuka peluang baru untuk otomatisasi beberapa aspek dari proses tinjauan artikel ilmiah. Kemampuan algoritma tersebut dalam memahami bahasa alami, analisis konten, dan penyajian informasi dapat dimanfaatkan untuk mempermudah dan mempercepat proses tinjauan. Pemrosesan bahasa alami (NLP) adalah subbidang kecerdasan buatan dan linguistik komputasi. Bidang ini berfokus pada kemampuan komputer untuk memahami, menafsirkan, dan menghasilkan bahasa manusia dengan cara yang bermakna dan berguna (Amaratunga, 2023). Salah satu pendekatan dalam pemrosesan bahasa alami yang memungkinkan untuk melakukan tugas mendalam dengan data besar adalah model bahasa besar atau *Large Language Models (LLM)*.

Model Bahasa besar adalah model bahasa yang telah dilatih sebelumnya dengan ukuran parameter jauh lebih besar dibandingkan dengan pendekatan lainnya (Liu et al., 2024). Model bahasa besar adalah hasil dari kombinasi pemrosesan bahasa alami, konsep pembelajaran mendalam, dan model kecerdasan buatan generative. Akhir-akhir ini model bahasa besar telah memberikan terobosan yang signifikan terutama yang berkaitan dengan transformer. Hal ini mencakup peningkatan komputasi dan ketersediaan pelatihan data dengan skala yang besar.

Perkembangan tersebut telah menghasilkan transformasi revolusioner dengan memberikan kemungkinan pembuatan LLM yang dapat mendekati kinerja manusia pada berbagai tugas (Naveed et al., 2023).

Beberapa penelitian sebelumnya membahas mengenai penelaahan sejawat dengan kecerdasan buatan seperti yang ditulis oleh Alessandro, Lorenzo, Pierpaolo, Stephen & Giuseppe pada tahun 2021 menunjukkan bahwa kecerdasan buatan mampu melakukan penelaahan sejawat dan memprediksi sesuai dengan hasil ahli (Checco et al., 2021). Penelitian lainnya dilakukan oleh Iddo Drori & Dov Te'eni tahun 2024 menunjukkan kecerdasan buatan khususnya model Bahasa besar mampu melakukan tugas telaah sejawat dengan cukup baik (Drori & Te'eni, 2024).

Penelitian ini bertujuan untuk mengembangkan model kecerdasan buatan untuk tinjauan artikel berbasis large language models dengan fokus pada peningkatan efisiensi dan kecepatan proses telaah sejawat, tanpa mengorbankan kualitas dan keakuratan penilaian ilmiah. Model ini diharapkan dapat membantu dalam identifikasi aspek-aspek kritis, serta memberikan analisis otomatis yang dapat digunakan sebagai dasar untuk penilaian lebih lanjut. Pengembangan model-tools ini diharapkan dapat mempercepat alur kerja tinjauan artikel, meminimalkan potensi kesalahan manusia, dan pada akhirnya dapat meningkatkan efisiensi penelitian ilmiah secara keseluruhan.

1.2 Batasan dan Rumusan Masalah

1.2.1 Batasan Masalah

Untuk mencegah meluasnya permasalahan dalam domain yang diteliti penulis membuat Batasan agar fokus pada penyelesaian masalah dapat tercapai. Adapun Batasan masalah dari penelitian ini adalah sebagai berikut :

1. Penelitian berfokus pada pengembangan model platform tinjauan artikel ilmiah dengan menggunakan LLM
2. Model LLM yang digunakan adalah GPT-4
3. Model difokuskan pada artikel ilmiah dalam disiplin ilmu komputer

1.2.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah disampaikan diperoleh beberapa permasalahan yang harus di selesaikan. Permasalahan tersebut dirumuskan sebagai berikut :

1. Bagaimana membangun model platform tinjauan artikel ilmiah dengan LLM?
2. Bagaimana algoritma untuk klasifikasi artikel ilmiah untuk publikasi jurnal?

1.3 Tujuan Penelitian

Secara umum penelitian ini bertujuan untuk menciptakan platform tinjauan artikel ilmiah berbasis kecerdasan buatan dengan menggunakan model bahasa besar. Tujuan penelitian secara khusus adalah sebagai berikut :

1. Membangun model platform tinjauan artikel ilmiah dengan LLM
2. Mengembangkan algoritma untuk klasifikasi artikel ilmiah untuk publikasi jurnal

1.4 Kontribusi dan Manfaat Penelitian

Kontribusi dalam bidang akademik adalah tersedianya model platform tinjauan artikel ilmiah dengan LLM dan algoritma untuk klasifikasi artikel ilmiah untuk publikasi jurnal. Kontribusi penelitian pada bidang teknologi adalah tersedianya platform tinjauan artikel ilmiah yang berbasis teknologi kecerdasan buatan.

Manfaat dari hasil penelitian ini:

1. Memudahkan para penelaah sejawat dalam melakukan tinjauan artikel.
2. Menciptakan standarisasi dalam penelaahan sejawat.
3. Memberikan efisiensi waktu dalam penelaahan sejawat.

BAB II

TELAAH PUSTAKA

2.1 Pemrosesan Bahasa Alami

Pemrosesan Bahasa Alami atau *Natural Language Processing (NLP)* adalah cabang dari ilmu komputer dan kecerdasan buatan yang bertujuan untuk mengajarkan mesin agar dapat memahami dan memproses bahasa manusia secara efektif dan akurat (Vivi P Ratung, 2023). Teknik NLP digunakan dalam berbagai konteks, seperti pengenalan suara, penerjemahan bahasa, analisis sentimen, pengembangan chatbot, serta pembuatan ringkasan teks, dan sebagainya. NLP menggabungkan prinsip linguistik, statistik, dan pembelajaran mesin untuk membantu komputer dalam memahami bahasa manusia dan memberikan respons yang sesuai dengan input yang diberikan. Secara tradisional, pekerjaan dalam pemrosesan bahasa alami cenderung melihat proses analisis bahasa sebagai suatu yang dapat diuraikan menjadi beberapa tahap yang mencerminkan perbedaan linguistik teoritis yang ditarik antara sintaksis, semantik dan pragmatik.

Sistem pemrosesan bahasa alami sering dijuluki sebagai "pipeline" karena umumnya melibatkan beberapa tahapan pemrosesan. Bahasa alami mengalir dari satu titik ke titik lainnya dalam sistem ini. Bahasa alami berisi informasi atau instruksi yang dapat diekstraksi tetapi tidak langsung diterjemahkan menjadi serangkaian operasi matematika. Informasi dan instruksi ini dapat disimpan, diindeks, dicari, dan dijalankan untuk merespons pertanyaan. Ini adalah salah satu fungsi utama dari pemrosesan bahasa alami (Hannes Hapke et al., 2019).

Menjembatani antara manusia dengan computer maka terdapat beberapa tugas-tugas dari NLP yaitu:

1. Klasifikasi teks : Menetapkan label atau kategori pada sebuah teks. Misalnya, mengklasifikasikan email sebagai spam atau bukan spam, analisis sentimen (mengidentifikasi sentimen sebagai positif, negatif, atau netral), kategorisasi topik, dll.
2. Terjemahan mesin: Menerjemahkan teks secara otomatis dari satu bahasa ke bahasa lain.

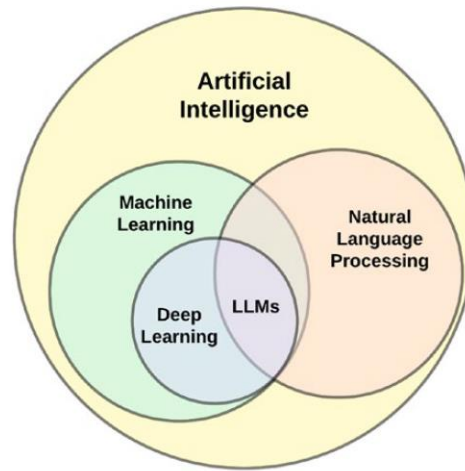
3. Pembuatan teks: Membuat teks seperti manusia, yang dapat berupa chatbot, konten yang dibuat secara otomatis, atau ringkasan teks.
4. Pengenalan ucapan: Mengubah bahasa lisan menjadi teks tertulis.
5. Peringkasan teks: Secara otomatis menghasilkan ringkasan yang ringkas dan koheren dari teks yang lebih panjang.
6. Pemodelan bahasa: Memprediksi kemungkinan munculnya urutan kata tertentu dalam suatu bahasa.

Konsep dasar dari NLP untuk mencapai tugas-tugas yang disebutkan sebelumnya, NLP menggunakan serangkaian konsep utama. Ini adalah beberapa yang paling umum:

- a. Tokenisasi : proses memecah kalimat menjadi unit-unit yang lebih kecil. biasanya berupa kata atau subkata. Unit-unit yang lebih kecil ini disebut token, dan tokenisasi adalah langkah preprocessing yang penting dalam sebagian besar tugas NLP. Misalnya kata “budi sedang bermain bola” yang akan di pecah menjadi [“Budi”, ”sedang”, ”bermain”, ”bola”]
- b. Penghapusan *Stopword* : Stopword adalah kata-kata umum yang sering muncul dalam teks tetapi memiliki makna semantik yang kecil. Menghapus stopwords dapat membantu mengurangi noise dan meningkatkan efisiensi komputasi.
- c. *Stemming dan Lemmatization* : Stemming dan lemmatization adalah teknik yang digunakan untuk mengurangi kata menjadi bentuk dasar.

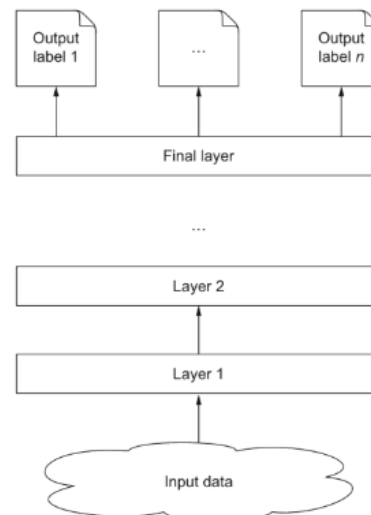
2.2 Deep Learning

Deep learning merupakan bidang penelitian baru dalam pembelajaran mesin yang bertujuan untuk meniru kemampuan otak manusia dalam mengolah dan mempelajari data masukan yang kompleks serta menyelesaikan tugas-tugas rumit dengan baik. Penggunaan deep learning telah berhasil diterapkan dalam berbagai bidang, seperti pengolahan gambar, suara, teks, dan gerak. Kemajuan teknik yang berasal dari penelitian deep learning telah memberikan dampak signifikan pada pengembangan Natural Language Processing (NLP), yang merupakan proses pemrosesan bahasa alami (Du & Shanker, n.d.).



Gambar 2. 1 Cabang Ilmu Artificial Intelligence

Pada gambar 2.1 menunjukkan cabang ilmu kecerdasan buatan yang mencakup pembelajaran mesin, pemrosesan Bahasa alami, deep learning dan model Bahasa besar.



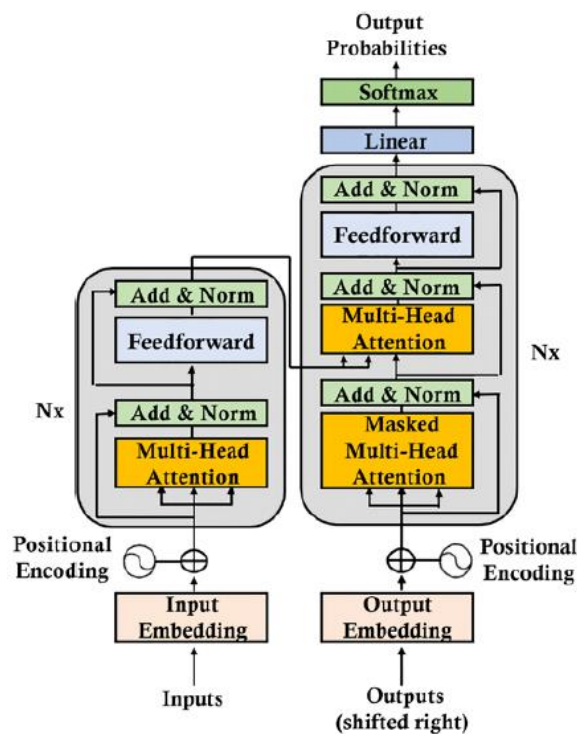
Gambar 2. 2 Arsitektur Umum Deep Learning

Lapisan keluaran, akhirnya, menghasilkan sebuah hasil: label yang diberikan model pada input τ_0 ke τ_α . Biasanya, jaringan menghasilkan kemampuan probabilitas untuk satu set hasil yang mungkin. Hasil dengan probabilitas tertinggi kemudian menjadi label keluaran akhir. Semua lapisan kecuali lapisan input dan output adalah

lapisan tersembunyi, karena mereka tidak dapat diamati dengan mudah. Seperti yang telah disebutkan, lapisan tersembunyi dalam neural net bekerja menguraikan lapisan data input yang tidak dapat dipisahkan secara linier, langkah demi langkah (Stephan Raaijmakers, 2022).

2.3 Transformers

Transformer adalah model pembelajaran yang dirancang untuk komputasi paralel pada superkomputer dengan homogenisasi dan dihomogenisasi. Melalui proses homogenisasi, satu model transformer dapat menangani berbagai tugas tanpa memerlukan penyesuaian khusus. Dengan kemampuan ini, Transformer dapat melakukan pembelajaran sendiri pada data mentah yang tidak berlabel dalam jumlah besar, menggunakan miliaran parameter (Rothman & Gulli, n.d.).



Gambar 2. 3 *Arsitektur Transformer*

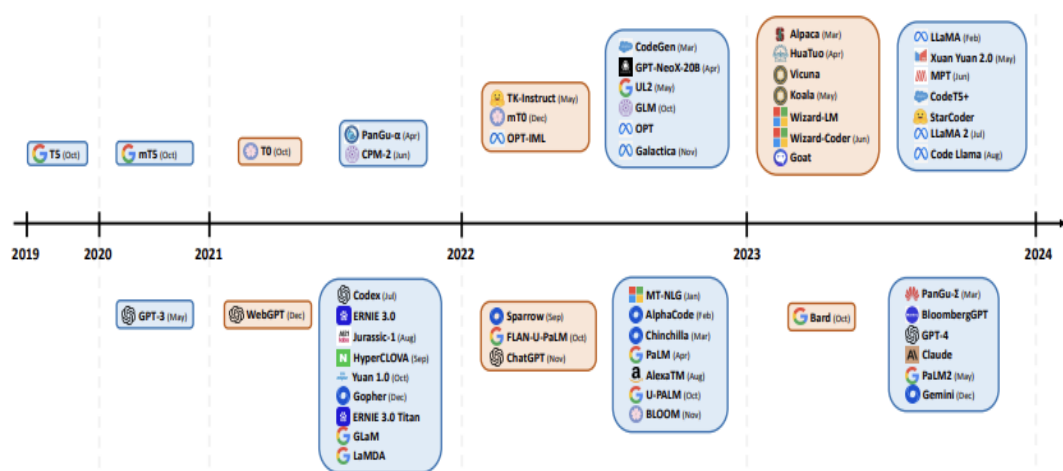
Pada gambar 2.3 di sebelah kiri, input masuk ke sisi *encoder* Transformer melalui *attention sublayer* dan *feedforward sublayer*. Di sebelah kanan, output target masuk ke sisi dekoder Transformer melalui dua *attention sublayer* dan *feedforward*

sublaye. Pada arsitektur ini tidak ada lagi RNN, LSTM, atau CNN. *Recurrence* telah ditinggalkan dalam arsitektur ini (Rothman & Gulli, n.d.).

Attention telah menggantikan fungsi recurrence yang membutuhkan parameter yang meningkat seiring bertambahnya jarak antara dua kata meningkat. Mekanisme attention adalah operasi "kata ke kata". Ini sebenarnya adalah operasi token-ke-token. Mekanisme attention akan menemukan bagaimana setiap kata berhubungan dengan semua kata lain dalam sebuah urutan, termasuk kata yang sedang dianalisis itu sendiri.

2.4 Model Bahasa Besar (LLM)

Large language models adalah model-model bahasa yang memiliki ukuran dan kompleksitas yang besar, terdiri dari jutaan atau bahkan miliaran parameter. Model-model ini dirancang untuk memahami dan menghasilkan bahasa manusia dengan tingkat keterampilan yang tinggi. LLM dibangun dengan menggunakan teknik-teknik deep learning, terutama menggunakan arsitektur seperti Transformer, yang memungkinkan LLM untuk menangani sejumlah besar data dan menangkap pola-pola kompleks dalam bahasa manusia. Large language models telah menjadi fokus utama dalam penelitian NLP karena kemampuan mereka untuk melakukan berbagai tugas, seperti penerjemahan bahasa, generasi teks, analisis sentiment dan sebagainya.



Gambar 2. 4 Model Bahasa Besar(LLM)

Pelatihan model bahasa besar (LLM) pada dataset yang sangat besar telah mengubah secara signifikan pemrosesan bahasa alami, memungkinkan LLM untuk meniru interaksi manusia dan berperan sebagai asisten serbaguna dalam berbagai tugas. Ini termasuk memberikan jawaban yang luas terhadap pertanyaan, membantu dalam penulisan, memberikan pengajaran, dan beragam tugas lainnya (Wolf et al., 2023).

Model bahasa yang besar dapat dikategorikan berdasarkan berbagai faktor seperti arsitektur, tujuan pelatihan, tipe data, dan aplikasi (Amaratunga, 2023). Berdasarkan arsitekturnya LLM dibedakan berdasarkan beberapa kategori, yaitu:

- Transformer
- Recurrent Neural Network
- Convolutional Neural Network

Adapun komponen-komponen penting dalam LLM :

1. Tokenization (tokenisasi)

Tokenization adalah langkah pra-pemrosesan penting dalam pelatihan model bahasa besar (LLM) yang memecah teks menjadi unit-unit tak terpisahkan yang disebut token (Naveed et al., 2023). Token dapat berupa karakter, subkata, simbol, atau kata, tergantung pada proses tokenisasi yang digunakan. Beberapa skema tokenisasi yang umum digunakan dalam LLM termasuk wordpiece, byte pair encoding (BPE), dan unigramLM (Webster & Kit, n.d.).

2. Encoding Positions

Transformer memproses urutan input secara paralel dan independen, tanpa mencatat informasi posisi. Oleh karena itu, diperkenalkan positional encoding di transformer, di mana vektor positional encoding ditambahkan ke token embedding. Varian positional embedding meliputi positional encoding absolute, relative, or learned positional encodings. Dalam relative encoding terdapat 2 positional embedding yang sering digunakan dalam LLMs yaitu Alibi dan RoPE.

- a. Alibi : bertugas mengurangi bias skalar dari attention score yang meningkat dengan Jarak antara posisi token.
- b. RoPE : teknik ini membantu model untuk lebih efektif memproses dan memahami teks dengan memperhatikan hubungan relatif antar token yang lebih relevan.

3. Attention

Attention bertugas memberikan bobot kepada token-token input berdasarkan kepentingannya sehingga model memberikan penekanan lebih pada token-token yang relevan (Vaswani et al., n.d.). Attention dalam transformer menghitung pemetaan query, key, dan value untuk urutan input, di mana skor perhatian diperoleh dengan mengalikan query dan key, dan kemudian digunakan untuk memberi bobot pada nilai-nilai. Beberapa strategi attention yang digunakan dalam LLM yaitu :

- a. Self-Attention
- b. Cross Attention
- c. Sparse Attention
- d. Flash Attention

4. Activation Functions

Fungsi aktivasi memiliki peran penting dalam kemampuan kurva jaringan saraf. Fungsi aktivasi adalah komponen kunci dalam LLM dan jaringan saraf lainnya yang memungkinkan model untuk belajar pola kompleks dan non-linear dalam data. Pilihan fungsi aktivasi yang tepat dapat secara signifikan mempengaruhi kinerja dan efisiensi model. Beberapa fungsi aktivasi yang digunakan dalam LLM yaitu :

- a. ReLU
- b. GeLU
- c. GLU variants

5. Encoder

Modul encoder dari model Transformer terdiri dari beberapa lapisan identik, masing-masing mencakup mekanisme multi-head attention dan jaringan saraf feed-

forward(Liu et al., 2024). Dalam mekanisme multi-head attention, setiap posisi dalam urutan input dihitung untuk perhatian dengan posisi lainnya guna menangkap ketergantungan antar posisi. Jaringan saraf feed-forward digunakan untuk memproses dan mengekstrak fitur dari output mekanisme perhatian. Modul encoder secara bertahap mengekstraksi fitur urutan input melalui penumpukan beberapa lapisan tersebut dan mengirimkan hasil encoding akhir ke modul decoder untuk decoding. Desain modul encoder memungkinkan penanganan ketergantungan jarak jauh dalam urutan input dan meningkatkan kinerja dalam berbagai tugas NLP.

6. Decoder

Modul decoder dari model Transformer terdiri dari beberapa lapisan identik, masing-masing mencakup mekanisme multi-head attention dan jaringan saraf feed-forward(Liu et al., 2024). Berbeda dengan encoder, decoder juga mencakup mekanisme attention tambahan antara encoder dan decoder, digunakan untuk menghitung perhatian pada urutan input selama proses decoding. Pada setiap posisi, decoder hanya dapat melakukan perhitungan self-attention dengan posisi sebelumnya untuk memastikan urutan yang dihasilkan tidak melanggar aturan tata bahasa. Mask memainkan peran penting dalam decoder, memastikan bahwa hanya informasi sebelum langkah waktu saat ini yang diperhatikan saat menghasilkan urutan output, dan tidak membocorkan informasi dari langkah waktu mendatang. Mekanisme self-attention pada decoder menggunakan mask untuk mencegah model mengakses informasi masa depan saat menghasilkan prediksi pada setiap langkah waktu, menjaga kausalitas model. Hal ini memastikan bahwa output yang dihasilkan oleh model bergantung pada informasi pada langkah waktu saat ini dan sebelumnya, tanpa dipengaruhi oleh informasi masa depan.

2.4.1 Prompt Learning

Prompt learning adalah pendekatan machine learning yang digunakan secara luas, terutama di bidang NLP. Metodologi ini melibatkan pembuatan pernyataan prompt yang hati-hati untuk mengarahkan model menghasilkan perilaku atau output tertentu. Pendekatan ini sering digunakan untuk fine-tuning dan mengarahkan LLM yang sudah dilatih sebelumnya untuk menjalankan tugas tertentu atau menghasilkan

hasil yang diinginkan(Liu et al., 2024). Desain pernyataan prompt dapat mengarahkan model pre-trained untuk melakukan berbagai tugas seperti menjawab pertanyaan, menghasilkan teks, dan memahami semantik. Kekuatan pendekatan ini terletak pada kemampuannya untuk beradaptasi dengan berbagai tugas melalui modifikasi sederhana pada pernyataan prompt, tanpa perlu melatih ulang seluruh model. Untuk LLM seperti seri GPT dan model pre-trained lainnya, prompt learning menyediakan cara yang mudah dan kuat untuk fine-tuning model. Dengan memberikan prompt yang sesuai, peneliti dan praktisi dapat menyesuaikan perilaku model agar lebih cocok untuk domain atau kebutuhan tugas tertentu. Singkatnya, prompt learning adalah pendekatan machine learning yang membangun model bahasa yang telah dilatih sebelumnya dan mengarahkannya untuk melakukan berbagai tugas melalui desain pernyataan prompt, menawarkan fleksibilitas yang meningkat untuk menyesuaikan aplikasi model.

2.4.2 Architecture

Saat ini semua Large Language Models (LLMs) dibangun dengan menggunakan arsitektur Transformer. Arsitektur ini memungkinkan model-model ini untuk memiliki skala hingga beberapa miliar atau bahkan triliun parameter. Secara umum, arsitektur Pretrained language model (PLM) dapat dikelompokkan menjadi tiga kategori: Encoder-only, Encoder-decoder, dan Decoder-only. Arsitektur Encoder-only tidak lagi digunakan dalam LLMs terbaru dan tidak akan dibahas lebih lanjut di sini. Sebagai gantinya, fokus bagian ini adalah untuk memperkenalkan arsitektur Encoder-decoder dan Decoder-only.

1. Encoder-Decoder

Arsitektur Encoder-decoder pada LLMs didasarkan pada arsitektur Transformer tradisional. Terdiri dari dua komponen utama: Encoder dan Decoder, di mana Encoder menyandikan urutan input melalui beberapa lapisan Multi-Head Self-Attention, sedangkan Decoder menggunakan cross-attention pada representasi output dari Encoder untuk menghasilkan urutan target secara autoregresif. Arsitektur ini menjadi dasar bagi LLM terkenal seperti T5, flan-T5, dan BART.

2. Decoder only

LLMs dengan arsitektur Decoder-only memanfaatkan komponen decoder dari arsitektur Transformer tradisional. Berbeda dengan arsitektur Encoder-Decoder yang menggabungkan encoder dan decoder, arsitektur Decoder-only hanya fokus pada proses dekoding. Model ini secara berurutan menghasilkan token-token dengan memperhatikan token-token sebelumnya dalam urutan. Arsitektur ini telah diterapkan dalam berbagai tugas generasi bahasa, menunjukkan efektivitasnya dalam menghasilkan teks tanpa memerlukan fase encoding eksplisit. Arsitektur Decoder-only dapat dibedakan lagi menjadi dua kategori: arsitektur Causal Decoder dan arsitektur Prefix Decoder.

- a. **Arsitektur Causal Decoder**, setiap token dalam urutan input model hanya dapat memperhatikan token-token input yang terjadi sebelumnya dan dirinya sendiri selama proses dekoding. Ini mencapai perhatian unidireksional pada urutan input dengan menggunakan masker khusus. Arsitektur ini dikonfigurasi dengan matriks masker yang berbeda-beda untuk mengimplementasikan berbagai arsitektur. Arsitektur Causal Decoder adalah dasar bagi serangkaian LLM terkenal seperti seri GPT, yang dikenal karena kinerja superior mereka dan banyak diterapkan dalam LLM lain seperti BLOOM, OPT, Gopher, dan LLaMA.
- b. **Arsitektur Prefix Decoder**, menggabungkan keunggulan dari arsitektur Encoder-decoder dan Causal Decoder. Dengan konfigurasi masker unik seperti yang diilustrasikan pada Gambar 1, arsitektur ini memungkinkan perhatian dua arah (bidireksional) untuk token-token dalam awalan (prefix), sementara tetap mempertahankan perhatian unidireksional untuk menghasilkan token-token berikutnya. Desain ini memungkinkan generasi autoregresif dari urutan output dengan fleksibilitas untuk memperhatikan secara bidireksional token-token awalan. Contoh LLM yang menerapkan arsitektur Prefix Decoder termasuk PaLM dan GLM.

2.4.3 Generative Pre-trained Transformer (GPT)

Generative Pre-trained Transformer (GPT) adalah model yang mempopulerkan LLM kepada masyarakat umum. GPT adalah keluarga LLM yang dirilis oleh OpenAI, sebuah laboratorium penelitian kecerdasan buatan Amerika yang terdiri dari organisasi nirlaba OpenAI Inc. Model GPT OpenAI menggunakan pendekatan semi-supervisi, yang merupakan pertama kalinya pendekatan semacam itu digunakan dengan model transformator. Pendekatan ini melibatkan dua tahap:

1. Tahap pra-pelatihan generatif tanpa pengawasan di mana tujuan pemodelan bahasa digunakan untuk menetapkan parameter awal
2. Tahap "fine-tuning" yang diawasi di mana parameter ini disesuaikan dengan tugas target

2.4.4 Bidirectional Encoder Representations from Transformers (BERT)

BERT adalah model transformator khusus encoder. Inovasi BERT terletak pada kemampuannya untuk menangkap konteks dari arah maju dan mundur dalam sebuah urutan, sehingga memungkinkannya untuk membuat representasi kata yang sangat kontekstual. Tidak seperti model bahasa tradisional sebelumnya yang bersifat searah (memprediksi kata berikutnya berdasarkan kata-kata sebelumnya), BERT memprediksi kata-kata yang hilang dalam sebuah kalimat dengan mempertimbangkan konteks kiri dan kanan, yang memungkinkannya untuk menangkap nuansa kontekstual secara lebih efektif.

2.5 Penelitian Terdahulu

Application of LLM Agents in Recruitment: A Novel Framework for Resume Screening yang ditulis oleh Chengguang Gan, Qinghao Zhang, Tatsunori Mori pada tahun 2024 (Gan et al., 2024) membahas mengenai kerangka kerja agen berbasis Large Language Models (LLMs) untuk proses penyaringan resume otomatis dalam rekrutmen. Tujuan utamanya adalah meningkatkan efisiensi dan manajemen waktu dalam proses rekrutmen. Penelitian ini secara keseluruhan menunjukkan potensi besar dalam penggunaan agen LLM untuk otomatisasi penyaringan resume, meskipun ada beberapa tantangan yang harus diatasi untuk penerapan yang lebih luas dan aman. Dalam penelitian tersebut dijelaskan bahwa penggunaan LLM dapat mengurangi waktu dan tenaga yang dibutuhkan untuk menyaring resume secara manual. LLM juga memberikan peningkatan signifikan dalam klasifikasi alimats resume dan kemampuan dalam mengidentifikasi informasi penting. Pada penelitian ini menggunakan model open-source seperti LLaMA2 memungkinkan eksekusi lokal yang lebih aman untuk data pribadi. LLM yang dapat digunakan dalam skenario rekrutmen nyata untuk meningkatkan efektivitas.

Paper Review: AI-Assisted Peer Review yang ditulis oleh Alessandro Checco, Lorenzo Bracciale, Pierpaolo Loreti, Stephen Pinfield, Giuseppe Bianchi pada tahun 2021 (Checco et al., 2021) membahas mengenai investigasi kemampuan AI untuk mendekati atau membantu keputusan manusia dalam proses penilaian kualitas dan peer review dari output penelitian. Para peneliti merancang alat AI dan melatihnya menggunakan 3300 makalah dari tiga konferensi beserta evaluasi review-nya. Tujuannya adalah untuk menguji kemampuan AI dalam memprediksi skor review dari manuskrip baru yang belum diamati, hanya dengan menggunakan konten tekstualnya. Penelitian ini memberikan gambaran bahwa AI dapat memprediksi hasil peer review yang dicapai berdasarkan rekomendasi reviewer manusia, meskipun hanya menggunakan metrik yang cukup sederhana seperti distribusi kata, skor keterbacaan, dan format dokumen. Korelasi yang kuat ditemukan antara ukuran kualitas proxy sederhana dan keputusan akhir penerimaan/penolakan, menunjukkan bahwa beberapa komponen dari proses

penilaian kualitas dan peer review dapat dibantu atau digantikan oleh alat berbasis AI. pada penelitian ini dijelaskan beberapa kelebihan mengenai penggunaan AI yaitu dapat mengurangi waktu yang diperlukan reviewer untuk menilai makalah dengan menangani bagian-bagian yang lebih membosankan dari proses review, seperti pemeriksaan keterbacaan dan format. Proses metodologi yang dijelaskan secara rinci memungkinkan replikasi setup eksperimental ini, yang meningkatkan kredibilitas hasil penelitian. AI dapat membantu mengungkap bias dalam proses review, yang berpotensi untuk dikurangi atau dihilangkan dengan pengembangan lebih lanjut. Penelitian ini memberikan pandangan mendalam tentang potensi dan batasan penggunaan AI dalam mendukung proses peer review, sekaligus menggarisbawahi pentingnya peran manusia dalam aspek-aspek yang lebih kompleks dan kritis dari penilaian penelitian.

Human-in-the-Loop AI Reviewing: Feasibility, Opportunities, and Risks yang ditulis oleh Iddo Drori dan Dov Te'eni pada tahun 2024 (Drori & Te'eni, 2024) membahas mengenai kelayakan, peluang, dan risiko penggunaan model bahasa besar (LLM) seperti GPT-4 dalam proses penelaahan akademik, dengan tetap mempertahankan manusia dalam loop. Penulis melakukan eksperimen dengan GPT-4 untuk menilai dan membandingkan ulasan yang dihasilkan oleh LLM dengan ulasan manusia. Penelitian ini menggunakan dataset dari International Conference on Learning Representations (ICLR) 2023, yang terdiri dari 2,040 makalah dengan total 7,698 ulasan. Lima versi ulasan GPT-4 per makalah dibuat, masing-masing dengan peningkatan jumlah dokumen kontekstual. Ulasan dibandingkan berdasarkan skor dan komentar bebas. Penilaian kelayakan dilakukan dengan membandingkan distribusi skor dan komentar antara ulasan manusia dan LLM. Hasil dari penelitian yang telah dilakukan adalah LLM dapat menghasilkan ulasan yang cukup akurat dan membantu mengurangi beban penelaahan, meskipun tidak sepenuhnya dan tidak untuk semua kasus. LLM menunjukkan bias positif sekitar 23% pada skor rekomendasi dibandingkan ulasan manusia. Bias ini dapat diminimalkan dengan menambahkan statistik tahun sebelumnya sebagai konteks. Kualitas komentar LLM dianggap sebanding dengan ulasan manusia dalam hal penjelasan skor, panduan perbaikan, dan kekhususan konten. Menggunakan

LLM dapat mengurangi beban penelaahan dan mempercepat proses penelaahan. Selain itu AI dapat membantu memastikan ulasan yang tidak bias dan adil sesuai dengan misi dan kebijakan jurnal. Penelitian ini menunjukkan bahwa AI-augmented reviewing memiliki potensi besar untuk meningkatkan efisiensi dan kualitas proses penelaahan akademik.

No	Judul	Tahun	Penulis	Metode	Kelebihan	Kekurangan
1	AI-assisted peer review	2021	Alessandro Checco,Lorenzo Bracciale,Pierpaolo Loreti,Stephen Pinfield & Giuseppe Bianchi	Regression Performance & Naïve Regressor	Dapat memprediksi sesuai dengan hasil ahli	hanya menggunakan beberapa aspek peer-review saja
2	Human-in-the-Loop AI Reviewing: Feasibility, Opportunities, and Risk	2024	Iddo Drori & Dov Te'eni	LLM (GPT-4)	Memberikan bukti empiris yang cukup kuat mengenai kelayakan penggunaan AI untuk membantu proses tunjauan artikel	Hanya menggunakan 1 model LLM yaitu GPT-4

No	Judul	Tahun	Penulis	Metode	Kelebihan	Kekurangan
3	Application of LLM Agents in Recruitment: A Novel Framework for Resume Screening	2024	Chengguang Gan, Qinghao Zhang, Tatsunori Mori	LLM (LLaMA2 dan GPT-3.5)	Mampu mempercepat dan mengotomatisasi proses screening resume	Memerlukan dataset pelatihan yang besar untuk penilaian dan ringkasan resume
4	Automated Paper Screening for Clinical Reviews Using Large Language Models: Data Analysis Study	2024	Eddie Guo, Mehul Gupta, Jiawen Deng, Ye-Jean Park, Michael Paget, Christopher Naugler	Large language Models	Mampu melakukan penalaran sehingga dapat menjelaskan keputusan dan memperbaiki kesalahan	Sensitivitas rendah sehingga ada kemungkinan melewatkan artikel yang sesuai
5	Zero-shot Generative Large Language Models for Systematic Review Screening Automation	2024	Shuai Wang, Harris Scells, Shengyao Zhuang, Martin Potthast, Bevan Koopman, Guido Zuccon	Generative Large Language Models	Adanya evaluasi dari berbagai arsitektur LLM tanpa memerlukan finetuning	Tidak adanya perbandingan dengan model lain seperti GPT

Berdasarkan penelitian telaah pustaka yang telah dilakukan dapat disimpulkan bahwa proses telaah sejawat dapat dilakukan secara otomatis dengan bantuan teknologi kecerdasan buatan. beberapa penelitian sebelumnya telah membahas mengenai penggunaan kecerdasan buatan terutama menggunakan model bahasa besar. Berdasarkan pemaparan tersebut maka terdapat peluang untuk dilakukan pengembangan platform tinjauan artikel ilmiah dalam bidang ilmu computer dengan menggunakan model bahasa besar.

2.6 Peer Review

Peer Review Process adalah proses di mana jurnal menilai kualitas naskah sebelum diterbitkan, ditinjau oleh para ahli yang relevan di bidangnya untuk mereview dan mengomentari naskah yang diterima. Proses ini bertujuan untuk membantu editor menentukan apakah naskah harus diterbitkan dalam jurnal.

Poin penting dalam Peer Review Process :

1. Naskah yang dikirimkan ke jurnal terlebih dahulu melalui penyaringan awal oleh tim editorial.
2. Naskah yang lolos pemeriksaan akan dikirim pada minimal dua peer reviewer untuk ditinjau.
3. Peer reviewer secara independen membuat rekomendasi kepada editor jurnal, apakah naskah harus ditolak atau diterima (dengan atau tanpa revisi).
4. Editor jurnal mempertimbangkan semua umpan balik dari peer reviewer dan membuat keputusan untuk menerima atau menolak naskah.
5. Peer Review Process untuk publikasi jurnal pada dasarnya adalah mekanisme kendali mutu, dimana para ahli mengevaluasi naskah yang bertujuan untuk memastikan kualitas dari naskah yang diterbitkan.

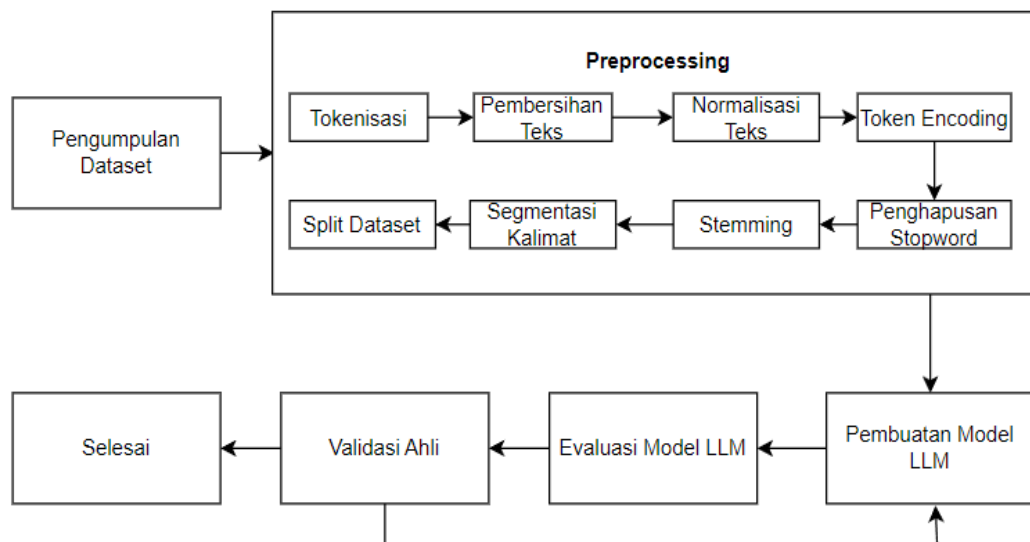
BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum

Penelitian ini bertujuan untuk mengembangkan aplikasi berbasis Large Language Model (LLM) dengan arsitektur GPT-4 yang mampu melakukan telaah sejawat(peer review) secara otomatis pada artikel ilmiah dari jurnal komputer. Data utama yang digunakan adalah artikel ilmiah berbahasa Indonesia dalam bidang ilmu komputer dari berbagai jurnal akademik. Sebelum digunakan, data akan diperiksa untuk menghilangkan informasi pribadi yang dapat mengidentifikasi penulis atau reviewer. Aplikasi ini diharapkan dapat membantu para peneliti dan editor jurnal dalam menganalisis dan memperoleh wawasan dari artikel yang seringkali bersifat kompleks dan teknis.

Untuk melakukan penelitian ini perlu dilakukan beberapa tahapan hingga penelitian selesai, tahapan yang dilakukan mulai dari pengumpulan data, preprocessing data, melakukan pemodelan untuk telaah sejawat, mengevaluasi model dan validasi ahli. Untuk tahapan penelitian dapat dilihat pada gambar 3.1.



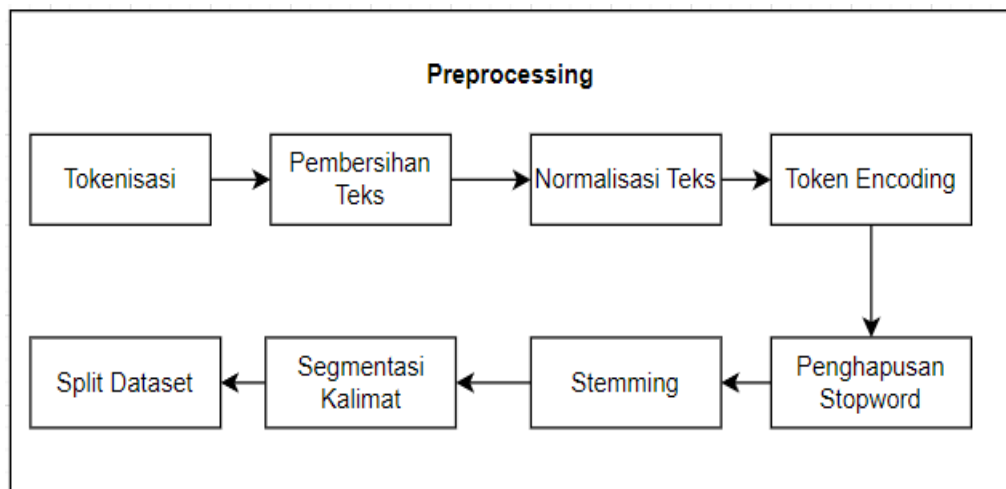
Gambar 3. 1Tahapan Penelitian

3.1.1 Pengumpulan data

Proses pengumpulan data dilakukan dengan cara mengumpulkan artikel ilmiah dari berbagai sumber terbuka dengan topik artikel ilmu computer. Pengumpulan data menggunakan teknik webscraping, artikel yang telah dikumpulkan akan diproses melalui tahap preprocessing.

3.1.2 Preprocessing data

Proses preprocessing data merupakan langkah yang sangat penting dalam persiapan data untuk pemodelan LLM. Proses ini melibatkan beberapa tahap penting yang bertujuan untuk membersihkan dan menyiapkan data teks agar sesuai dengan kebutuhan model serta meningkatkan kualitas dan konsistensi representasi teks. Proses preprocessing dilakukan melalui beberapa tahap seperti tokenisasi, pembersihan teks, normalisasi, token encoding, penghapusan stopwords, stemming, segmentasi kalimat dan pemisahan dataset.



Gambar 3. 2 Tahapan Preprocessing

Tahap pertama adalah tokenisasi, di mana teks dipecah menjadi unit-unit yang lebih kecil yang dikenal sebagai token, memungkinkan model untuk menganalisis teks pada tingkat yang lebih granular. Selanjutnya, dilakukan pembersihan teks untuk menghilangkan karakter atau simbol yang tidak diinginkan seperti tanda baca, angka, dan karakter khusus lainnya, serta penghapusan spasi berlebih dan karakter yang tidak relevan. Normalisasi juga dilakukan untuk mengubah teks

menjadi bentuk standar, termasuk mengubah semua huruf menjadi huruf kecil, menghapus aksent dari huruf, dan menangani variasi penulisan yang berbeda untuk kata yang sama. Setelah itu, token yang dihasilkan dari tokenisasi perlu diubah menjadi representasi numerik melalui token encoding, menggunakan teknik embeddings dari model transformer. Penghapusan stopword, yaitu kata-kata umum yang sering muncul dalam teks tetapi tidak memiliki makna khusus yang penting untuk analisis, juga dilakukan untuk mengurangi dimensi data dan fokus pada kata-kata yang lebih bermakna. Proses selanjutnya adalah stemming dan lemmatisasi yang bertujuan untuk mengurangi kata-kata ke bentuk dasar atau akar katanya, dengan stemming memotong akhiran kata dan lemmatisasi menggunakan kamus bahasa untuk mengembalikan kata ke bentuk dasar yang benar secara gramatikal. Selanjutnya segmentasi kalimat dilakukan untuk memisahkan teks menjadi kalimat-kalimat individu yang bisa dianalisis lebih lanjut secara terpisah. Tahap terakhir dalam preprocessing adalah pemisahan dataset menjadi bagian-bagian yang berbeda, seperti data latih, data validasi, dan data uji, yang penting untuk mengevaluasi kinerja model secara adil dan menghindari overfitting. Melalui proses preprocessing yang cermat dan terstruktur, data teks menjadi lebih bersih, terorganisir, dan siap digunakan dalam pemodelan, sehingga tidak hanya meningkatkan efisiensi pemrosesan data tetapi juga memungkinkan model untuk belajar dan melakukan prediksi dengan lebih akurat.

3.1.3 Pembuatan Model LLM

Setelah dataset yang di kumpulkan dan melalui proses preprocessing maka dilanjutkan tahap pemodelan dengan menggunakan LLM. Pada tahap ini dilakukan pemodelan dengan arsitektur GPT-4 untuk platform tinjauan artikel ilmiah. Proses pemodelan dimulai dengan fine-tuning GPT-4 menggunakan dataset yang telah dipreprocessing sebelumnya. Fine-tuning dilakukan untuk menyesuaikan model dengan gaya penulisan dan terminologi spesifik yang digunakan dalam artikel ilmiah. Selama fase pelatihan, model dievaluasi secara berkala untuk memastikan kinerjanya sesuai dengan harapan, dan parameter model dioptimalkan untuk meningkatkan kualitas output. Penggunaan GPT-4 untuk platform tinjauan artikel ilmiah dapat menyediakan analisis yang mendalam dan komprehensif, membantu

reviewer untuk lebih cepat dan efisien dalam menilai kualitas dan kontribusi sebuah artikel. Hal ini tidak hanya meningkatkan produktivitas tetapi juga memastikan bahwa artikel yang dipublikasikan memenuhi standar ilmiah yang tinggi.

3.1.4 Evaluasi Model LLM

Evaluasi model merupakan langkah yang penting dalam pengembangan sistem kecerdasan buatan, karena memungkinkan untuk menilai kinerja dan efektivitas model dalam menyelesaikan tugas tertentu. Proses evaluasi membantu mengidentifikasi kelemahan dan kekuatan model, serta memberikan wawasan tentang seberapa baik model dapat digunakan. Tanpa evaluasi yang tepat, model yang dikembangkan dapat menghasilkan prediksi yang tidak akurat atau tidak dapat diandalkan, yang berpotensi menyebabkan kinerja sistem yang buruk secara keseluruhan. Pada penelitian ini dilakukan evaluasi model dengan melihat nilai akurasi, presisi, recall dan F1-Score.

1. Akurasi memberikan gambaran umum tentang seberapa baik model klasifikasi melakukan prediksi secara keseluruhan.
2. Presisi memberikan informasi tentang seberapa banyak prediksi positif yang sebenarnya benar dari semua prediksi positif yang dilakukan oleh model.
3. Recall memberikan informasi tentang seberapa banyak instance positif yang berhasil diidentifikasi oleh model dari semua instance positif yang
4. sebenarnya dalam dataset.
5. F1-Score berguna ketika kelas target tidak seimbang dalam dataset, karena mencakup baik presisi maupun recall dalam perhitungannya.

3.1.5 Validasi Ahli

Proses validasi ahli ini memastikan bahwa model GPT-4 yang digunakan untuk telaah sejawat mampu memberikan evaluasi yang akurat, relevan, dan sesuai dengan standar akademik, dengan masukan berharga dari para ahli di bidangnya.

3.2 Jadwal Penelitian

Jadwal penelitian bertujuan untuk mengatasi target waktu penelitian, memastikan bahwa penelitian ini dapat diselesaikan sesuai dengan batas waktu

DAFTAR PUSTAKA

- Amaratunga, T. (2023). Understanding Large Language Models. In *Understanding Large Language Models*. Apress. <https://doi.org/10.1007/979-8-8688-0017-7>
- Checco, A., Bracciale, L., Loreti, P., Pinfield, S., & Bianchi, G. (2021). AI-assisted peer review. *Humanities and Social Sciences Communications*, 8(1). <https://doi.org/10.1057/s41599-020-00703-8>
- Drori, I., & Te'eni, D. (2024). Human-in-the-Loop AI Reviewing: Feasibility, Opportunities, and Risks. In *Journal of the Association for Information Systems* (Vol. 25, Issue 1, pp. 98–109). Association for Information Systems. <https://doi.org/10.17705/1jais.00867>
- Du, T., & Shanker, V. K. (n.d.). *Deep Learning for Natural Language Processing*.
- Gan, C., Zhang, Q., & Mori, T. (2024). *Application of LLM Agents in Recruitment: A Novel Framework for Resume Screening*. <http://arxiv.org/abs/2401.08315>
- Hannes Hapke, Cole Howard, & Hobson Lane. (2019). *Natural Language Processing in Action: Understanding, analyzing, and generating text with Python*. Simon and Schuster.
- Liu, Y., He, H., Han, T., Zhang, X., Liu, M., Tian, J., Zhang, Y., Wang, J., Gao, X., Zhong, T., Pan, Y., Xu, S., Wu, Z., Liu, Z., Zhang, X., Zhang, S., Hu, X., Zhang, T., Qiang, N., ... Ge, B. (2024). *Understanding LLMs: A Comprehensive Overview from Training to Inference*. <http://arxiv.org/abs/2401.02038>
- Murphy, R. F. (2019). *RAND Corporation Artificial Intelligence Applications to Support K-12 Teachers and Teaching: A Review of Promising Applications, Opportunities, and Challenges*.
- Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2023). *A Comprehensive Overview of Large Language Models*. <http://arxiv.org/abs/2307.06435>
- Rothman, D., & Gulli, A. (n.d.). *Transformers for natural language processing : build, train, and fine-tuning deep neural network architectures for NLP with Python, PyTorch, TensorFlow, BERT, and GPT-3*.
- Stephan Raaijmakers. (2022). *Deep Learning for Natural Language Processing*. Simon and Schuster.

Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (n.d.). *Attention Is All You Need*.

VIVI P RATUNG, S. T. , M. (2023). *TEKNIK-TEKNIK PEMROSESSAN BAHASA ALAMI(NLP)*. Lakeisha.

Webster, J. J., & Kit, C. (n.d.). *TOKENIZATION AS THE INITIAL PHASE IN NLP*.

Wolf, Y., Wies, N., Avnery, O., Levine, Y., & Shashua, A. (2023). *Fundamental Limitations of Alignment in Large Language Models*.
<http://arxiv.org/abs/2304.11082>