

BAB II

TINJAUAN PUSTAKA

2.1. Penelitian Terkait

2.1.1. *IndoLEM IndoNLU: Benchmark and Resource for Evaluating Indonesian Natural Language Understanding* [2]

Meskipun Bahasa Indonesia digunakan oleh hampir 200 juta orang dan merupakan bahasa yang paling banyak digunakan ke-10 di dunia, namun bahasa tersebut kurang mendapatkan perhatian dalam penelitian Pemrosesan Bahasa Alami (NLP). Penelitian-penelitian sebelumnya tentang Bahasa Indonesia terhambat oleh kurangnya *dataset* yang dianotasi, keterbatasan sumber daya bahasa, dan kurangnya standarisasi sumber daya. Dalam penelitian ini, mereka merilis dataset IndoNLU yang terdiri dari 12 tugas untuk Bahasa Indonesia, meliputi sentimen analisis berbasis aspek, question-answering, part-of-speech tagging, dan banyak lainnya. Peneliti juga merilis INDOBERT, sebuah model bahasa pra-pelatihan baru untuk Bahasa Indonesia, dan mengevaluasinya dengan *dataset* indoLEM, serta membandingkannya dengan sumber daya yang sudah ada. Percobaan peneliti menunjukkan bahwa IndoBERT mencapai kinerja terbaik di sebagian besar tugas dalam IndoLEM.

Dikarenakan pada penelitian ini sentimen yang dianalisa berupa Bahasa Indonesia, maka model yang memiliki kemampuan Bahasa Indonesia secara komprehensif akan cocok digunakan pada penelitian ini. Pada penelitian ini disebutkan bahwa IndoBERT terdapat 4 miliar kata Bahasa Indonesia yang bersumber dari Wikipedia, Kompas, Liputan6, Tempo, OSCAR, dan korpus besar Bahasa Indonesia lainnya yang mengandung bahasa formal maupun non formal [2].

2.1.2. *ChatGPT vs Gemini LLaMA on Multilingual Sentiment Analysis* [3]

Dalam studi ini, peneliti membuat skenario-skenario yang nuansanya kompleks dan ambigu, peneliti menerjemahkannya ke dalam 10 bahasa, dan peneliti memprediksi sentimen yang terkait menggunakan LLM populer. Hasilnya divalidasi berdasarkan tanggapan manusia pasca-hoc. Skenario-skenario yang ambigu sering kali dapat diatasi dengan baik oleh ChatGPT dan Gemini, tetapi peneliti mengakui adanya bias signifikan dan kinerja yang tidak konsisten di antara model dan bahasa manusia yang dievaluasi. Penelitian ini menyediakan metodologi standar untuk evaluasi analisa sentimen otomatis dan mengajukan panggilan untuk lebih meningkatkan algoritma dan data yang mendasarinya, untuk meningkatkan kinerjanya, interpretabilitas, dan aplikabilitasnya. Penelitian ini menjadi acuan untuk mengetahui model terbaru yang cukup populer untuk *sentiment analysis*, oleh karena itu pada penelitian ini model lain yang dibandingkan adalah Gemini, GPT dan IndoBERT.

2.1.3. *Multi-label Text Classification of Indonesian Customer Review Using Bidirectional Encoder Representation from Transformer Language Model* [4]

Dalam penelitian ini, peneliti ingin mengetahui efektivitas dari dua strategi yang peneliti usulkan menggunakan representasi fitur dari model bahasa dengan enkoder bidireksional dari transformer (BERT) yang telah dilatih pada Bahasa Indonesia, yang disebut sebagai IndoBERT, untuk melakukan klasifikasi teks multi-label. Pertama, IndoBERT digunakan sebagai representasi fitur yang akan digabungkan dengan jaringan saraf konvulsional-ekstrim *gradient boosting* (CNN-XGBoost). Kedua, IndoBERT digunakan baik sebagai representasi fitur maupun klasifikasi untuk langsung menyelesaikan tugas klasifikasi. Analisis tambahan dilakukan untuk membandingkan hasil kami dengan menggunakan model BERT multibahasa. Menurut

hasil eksperimen peneliti, model pertama peneliti yang menggunakan IndoBERT sebagai representasi fitur menunjukkan kinerja yang signifikan dibandingkan dengan beberapa *baseline*. Model kedua peneliti yang menggunakan IndoBERT baik sebagai representasi fitur maupun klasifikasi dapat meningkatkan secara signifikan efektivitas model pertama peneliti. Secara ringkas, model yang kami usulkan dapat meningkatkan efektivitas *baseline* menggunakan Word2Vec-CNN-XGBoost sebesar 19,19% dan 6,17%, dalam hal akurasi dan skor F-1, secara berturut-turut.

Berbeda dengan kasus sentiment analysis yang hanya bersifat positive dan negative, dikarenakan struktur label penilaian untuk studi kasus penelitian TokoTeknoPro terdiri dari beberapa label maka studi literatur yang berbasis multi-label akan cocok untuk penelitian aspect based sentiment analysis.

2.1.4. An Experimental Study of Lexicon-based Sentiment Analysis on Bahasa Indonesia [5]

Penelitian ini mencoba melakukan *sentiment analysis* untuk Bahasa Indonesia menggunakan *lexicon-based method*, dan menjelaskan bahwa pendekatan *lexicon-based* cenderung lebih ringan dari sisi usaha yang diperlukan dibandingkan dengan pendekatan *supervised training* menggunakan data yang dilabel secara manual. Namun pendekatan *lexicon-based* sendiri cenderung memberikan hasil dengan akurasi yang lebih rendah dibandingkan pendekatan *supervised training*.

Selain itu, penelitian ini juga menyuarakan beberapa permasalahan yang dihadapi ketika menggunakan pendekatan *lexicon-based* yang kemudian diterjemahkan ke Bahasa Inggris, antara lain struktur kalimat yang menyebabkan "*The Phenomenon of Thwarted Expectation*", kata-kata yang kurang standar dalam Bahasa Indonesia namun sering digunakan terutama dalam komunikasi tekstual di internet, dan kesulitan dalam menerjemahkan kata-kata tertentu yang ambigu ke

Bahasa Inggris. Dengan mempelajari penelitian ini, saya memutuskan untuk mengambil pendekatan *supervised training*, dan menghindari proses penerjemahan untuk mengurangi dampak dari permasalahan yang dijelaskan di atas.

2.1.5. *Challenges and Opportunities of Language Representation Model* [6]

Penelitian ini menjelaskan beberapa jenis representasi bahasa, menjelaskan karakteristik dari masing-masing representasi bahasa tersebut dan perbedaannya satu sama yang lain. Dari penelitian ini, tampak bahwa BERT yang mengambil pendekatan *bidirectional* dibandingkan dengan GPT yang kemudian berkembang menjadi GPT-2 yang tidak menggunakan *bidirectional transformer* dan hanya menggunakan *one-directional transformer*.

Penelitian ini sendiri hanya menjelaskan sedikit tentang GPT-3, karena penelitian ini dipublikasikan ketika GPT-3 baru dapat digunakan oleh publik. Melihat kondisi sekarang dimana GPT-3 sendiri sudah dapat digunakan publik, dan bahkan dapat dilatih (*fine-tune*) untuk dapat menyelesaikan tugas tertentu setelah dilatih, saya ingin menjadikan GPT-3 (terutama GPT-3.5) sebagai salah satu model yang dibandingkan dalam penelitian saya ini, mengingat GPT-3.5 sudah berkembang jauh lebih banyak dibandingkan model sebelumnya, dan OpenAI sudah membuka jalan untuk melakukan *fine-tuning* terhadap model GPT-3.5

2.1.6. *Impact of Training and Testing Data Splits on Accuracy of Time Series Forecasting in Machine Learning* [7]

Penelitian ini mencoba mengeksplorasi pengaruh dari perubahan jumlah dataset yang digunakan untuk training, dan pengaruhnya dengan tingkat akurasi dari prediksi model. Penelitian ini mencoba memvariasikan jumlah training data ke beberapa tingkat dan juga membandingkan jenis *output* dari modelnya untuk melihat dari beberapa sisi.

Penelitian ini memberikan wawasan menarik dengan melihat hasil yang menunjukkan tingkat akurasi dari model yang sudah dilatih. *Metric* yang digunakan untuk merepresentasikan tingkat akurasi adalah *Mean Average Error*, dan dari grafik yang ditunjukkan, tampak bahwa jumlah data yang terlalu sedikit umumnya berujung pada tingkat akurasi yang lebih rendah. Akan tetapi, dapat dilihat perilaku yang tidak konsisten terutama pada jumlah data yang lebih tinggi. Terkadang jumlah data *training* yang lebih tinggi malah menyebabkan tingkat akurasi yang lebih rendah untuk jenis *output* tertentu, tetapi untuk jenis *output* lain tingkat akurasi semakin meningkat untuk jumlah *training* data yang lebih tinggi. Peneliti juga menyimpulkan bahwa untuk jumlah data tertentu, tingkat akurasi dapat menurun tajam dan hal ini dapat dihindari bila jumlah data yang digunakan meningkat cukup besar. Akan tetapi, jumlah data yang besar akan berujung pada kebutuhan sistem yang lebih mumpuni, waktu yang lebih lama, dan cenderung biaya yang lebih besar, sehingga diperlukan pertimbangan dari berbagai aspek untuk menentukan jumlah data *training* dan lebih banyak belum tentu lebih baik.

2.1.7. *Large Language models for aspect-based sentiment analysis*[13]

Penelitian ini meninjau perkembangan Analisis Sentimen Berbasis Aspek (ABSA) menggunakan *Large Language Model*. Pendekatan awal ABSA mengandalkan teknik pembelajaran mesin tradisional seperti support vector machines (SVMs) dan *conditional random fields* (CRFs), yang memerlukan rekayasa fitur ekstensif dan sering kesulitan menangkap nuansa kontekstual dalam teks. Dengan kemajuan dalam pembelajaran mendalam, jaringan saraf berulang (RNN) dan long short-term memory (LSTM) mulai digunakan untuk menangkap ketergantungan kontekstual, meskipun masih menghadapi tantangan dalam menangani ketergantungan jarak jauh dan memerlukan data berlabel besar. Pengenalan mekanisme atensi, terutama dalam model Transformer, memungkinkan model memberi bobot dinamis pada kata-kata penting, dan model pra-pelatihan seperti BERT telah merevolusi

tugas-tugas NLP dengan kinerja unggul melalui pra-pelatihan skala besar diikuti dengan penyesuaian khusus tugas. LLMs seperti GPT-3.5 dan GPT-4 menawarkan kemampuan luar biasa dalam penyelesaian teks dan pemahaman yang mendalam, memungkinkan mereka unggul dalam pembelajaran zero-shot dan few-shot yang mengurangi kebutuhan dataset berlabel besar. Fine-tuning GPT-3.5 untuk tugas ABSA mencapai skor F1 83.8, menunjukkan peningkatan kinerja yang signifikan. Namun, tantangan tetap ada dalam hal biaya komputasi yang tinggi dan kebutuhan untuk meningkatkan interpretabilitas serta adaptasi domain. Penelitian masa depan harus fokus pada optimasi efisiensi LLMs, peningkatan interpretabilitas, dan memastikan adaptasi domain yang efektif. Karya ini menyoroti potensi transformasional LLMs dalam ABSA sambil mengakui tantangan dan arah penelitian yang perlu dieksplorasi lebih lanjut.

2.1.8. *Efficient Sentiment Analysis: A Resource-Aware Evaluation of Feature Extraction Techniques, Ensembling, and Deep Learning Models* [14]

Penelitian ini membahas pentingnya evaluasi yang memperhitungkan efisiensi sumber daya dalam analisis sentimen, selain hanya mengejar akurasi maksimal. Studi ini membandingkan berbagai teknik ekstraksi fitur, metode ensembling, model pembelajaran mendalam, LLM dalam konteks biaya sumber daya yang dibutuhkan untuk penerapan model. Hasil penelitian menunjukkan bahwa meskipun LLMs yang di-fine-tune mencapai akurasi terbaik, beberapa konfigurasi alternatif mampu memberikan penghematan sumber daya hingga 24.283 kali lipat dengan penurunan akurasi yang sangat kecil (<1%). Selain itu, pada dataset yang lebih kecil, perbedaan akurasi antara model berkurang sementara perbedaan konsumsi sumber daya semakin meningkat.

Penelitian ini juga menunjukkan bahwa LLMs, seperti GPT-3.5 dan GPT-4, mengungguli model tradisional seperti Support Vector Machines (SVM) dalam tugas analisis sentimen, terutama dalam konteks pengaturan zero-shot dan few-shot learning. LLMs mampu memberikan hasil yang lebih akurat dan konsisten dibandingkan SVM, yang membutuhkan rekayasa fitur ekstensif

dan data berlabel yang lebih besar untuk mencapai hasil yang serupa. Meskipun demikian, penting untuk dicatat bahwa keunggulan performa LLMs ini datang dengan biaya komputasi dan sumber daya yang jauh lebih tinggi dibandingkan SVM.

Penelitian ini menekankan bahwa dalam banyak kasus, algoritma yang lebih sederhana seperti SVM dapat mencapai performa yang hampir setara dengan LLMs namun dengan konsumsi daya dan sumber daya yang jauh lebih rendah, menjadikannya pilihan yang lebih efisien untuk aplikasi di dunia nyata. Dengan demikian, penelitian ini menggarisbawahi pentingnya mempertimbangkan efisiensi sumber daya dalam pengembangan dan pemilihan model pembelajaran mesin untuk analisis sentimen.

2.2. *Natural Language Processing*

Natural language processing (NLP), merupakan salah satu pendekatan terkomputerisasi untuk menganalisa teks berdasarkan aspek teori dan teknologi. NLP didefinisikan sebagai sebuah bidang teoritis mengenai suatu teknik komputasional yang digunakan untuk menganalisa dan merepresentasikan teks yang ditulis secara natural (bahasa manusia) pada satu atau lebih level analisis linguistik dengan tujuan untuk memperoleh *human-like language processing* yang dapat diimplementasikan dalam berbagai bidang [8].

NLP menyediakan teori dan implementasi untuk diaplikasikan dalam berbagai bidang. Namun faktanya segala bidang yang berhubungan dengan pengolahan teks merupakan kandidat untuk NLP. Beberapa bidang pengaplikasian yang memanfaatkan NLP diantaranya adalah *Information Retrieval, Information Extraction, Question Answering, Summarization, Machine Translation* dan *Dialogue System* [9].

Pustejovsky dan Stubbs, (2012) menjelaskan bahwa ada beberapa area utama penelitian pada field NLP, diantaranya:

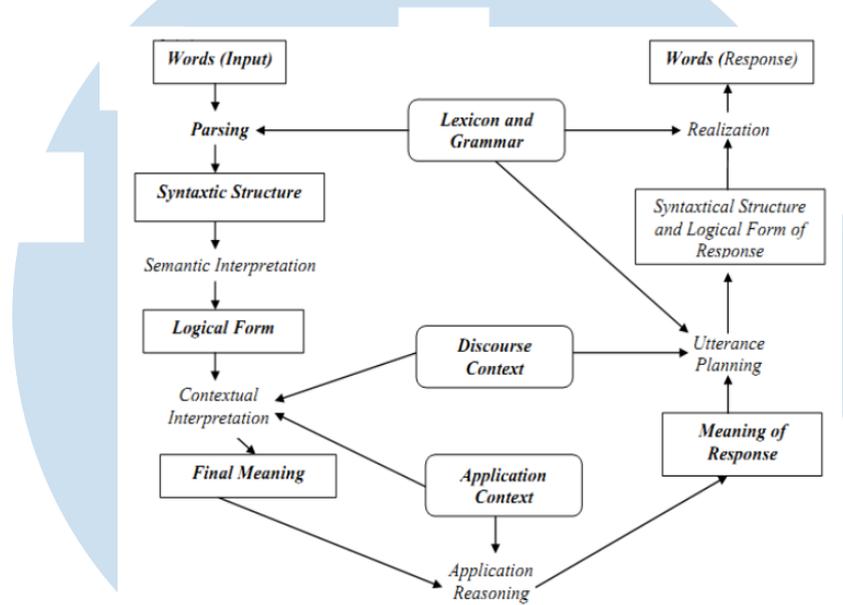
- *Question Answering Systems* (QAS). Kemampuan komputer untuk menjawab pertanyaan yang diberikan oleh pengguna. Daripada memasukkan *keyword* ke dalam mesin pencarian, dengan QAS,

pengguna bisa langsung bertanya dalam bahasa natural yang digunakannya, baik itu Inggris, Mandarin, ataupun Indonesia.

- *Summarization*. Pembuatan ringkasan dari sekumpulan konten dokumen atau email. Dengan menggunakan aplikasi ini, pengguna bisa dibantu untuk mengkonversikan dokumen teks yang besar ke dalam bentuk slide presentasi.
- *Machine Translation*. Produk yang dihasilkan adalah aplikasi yang dapat memahami bahasa manusia dan menerjemahkannya ke dalam bahasa lain. Termasuk di dalamnya adalah Google Translate yang apabila dicermati semakin membaik dalam penerjemahan bahasa. Contoh lain lagi adalah BabelFish yang menerjemahkan bahasa pada *real time*.
- *Speech Recognition*. Bidang ini merupakan cabang ilmu NLP yang cukup sulit. Proses pembangunan model untuk digunakan telepon/komputer dalam mengenali bahasa yang diucapkan sudah banyak dikerjakan. Bahasa yang sering digunakan adalah berupa pertanyaan dan perintah.
- *Document classification*. Aplikasi ini adalah merupakan area penelitian NLP yang paling sukses. Pekerjaan yang dilakukan aplikasi ini adalah menentukan dimana tempat terbaik dokumen yang baru dimasukkan ke dalam sistem. Hal ini sangat berguna pada aplikasi *spam filtering*, *news article classification*, dan *movie review*.

NLP tidak bertujuan untuk mentransformasikan bahasa yang diterima dalam bentuk teks atau suara menjadi data digital dan/atau sebaliknya pula; melainkan bertujuan untuk memahami arti dari kalimat yang diberikan dalam bahasa alami dan memberikan respon yang sesuai, misalnya dengan melakukan suatu aksi tertentu atau menampilkan data tertentu. Untuk mencapai tujuan ini dibutuhkan tiga tahap proses. Proses yang pertama ialah *parsing* atau analisa sintaksis yang memeriksa kebenaran struktur kalimat berdasarkan suatu *grammar* (tata bahasa) dan *lexicon* (kosakata) tertentu. Proses kedua ialah *semantic interpretation* atau interpretasi semantik yang bertujuan untuk merepresentasikan arti dari kalimat secara *context-independent* untuk keperluan lebih lanjut. Sedangkan proses ketiga ialah

contextual interpretation atau interpretasi kontekstual yang bertujuan untuk merepresentasikan arti secara *context dependent* dan menentukan maksud dari penggunaan kalimat. Gambaran organisasi dari sebuah sistem NLP yang lengkap ditunjukkan pada Gambar 2.1.



Gambar 2.1 Gambaran Organisasi Sistem NLP

Sumber : studi analisis metode-metode *parsing* dan interpretasi semantik pada *natural language processing*

2.3. *Natural Language Understanding*

Pemahaman bahasa alami (*Natural Language Understanding/NLU*) adalah cabang dari kecerdasan buatan (*Artificial Intelligence/AI*) yang menggunakan perangkat lunak komputer untuk memahami masukan dalam bentuk kalimat menggunakan teks atau ucapan. NLU memungkinkan interaksi manusia-komputer dengan menganalisis bahasa daripada hanya kata-kata.

NLU memungkinkan komputer untuk memahami sentimen yang diekspresikan dalam bahasa alami yang digunakan oleh manusia, seperti Bahasa Inggris, Prancis, atau Mandarin, tanpa sintaksis formal dari bahasa

komputer. NLU juga memungkinkan komputer untuk berkomunikasi kembali kepada manusia dalam bahasa mereka sendiri.

Sebuah bentuk dasar dari NLU disebut *parsing*, yang mengambil teks tertulis dan mengubahnya menjadi format terstruktur agar dapat dipahami oleh komputer. Alih-alih mengandalkan sintaksis bahasa komputer, NLU memungkinkan komputer untuk memahami dan merespons teks yang ditulis oleh manusia.

Salah satu tujuan utama dari NLU adalah untuk menciptakan bot yang dapat diaktifkan melalui obrolan (*chat*) dan suara yang dapat berinteraksi dengan orang tanpa pengawasan. Banyak *startup*, serta perusahaan IT besar, seperti Amazon, Apple, Google, dan Microsoft, baik memiliki atau sedang mengerjakan proyek-proyek NLU dan model bahasa.

NLU menganalisis data untuk menentukan maknanya dengan menggunakan algoritma untuk mengurangi ucapan manusia menjadi ontologi terstruktur - model data yang terdiri dari definisi semantik dan pragmatik. Dua konsep dasar dari NLU adalah pengenalan tujuan dan entitas:

1. **Intent recognition** adalah proses mengidentifikasi sentimen pengguna dalam teks masukan dan menentukan tujuan mereka. Ini adalah bagian pertama dan paling penting dari NLU, karena menetapkan makna teks.
2. **Entity recognition** adalah jenis khusus dari NLU yang fokus pada mengidentifikasi entitas dalam sebuah pesan dan kemudian mengekstrak informasi paling penting tentang entitas tersebut. Ada dua jenis entitas: entitas bernama dan entitas numerik. Entitas bernama dikelompokkan ke dalam kategori, seperti nama orang, nama bisnis, dan lokasi. Entitas numerik dikenali sebagai kuantitas, tanggal, mata uang, dan persentase.

Ketika diberikan masukan dalam bahasa alami, NLU membagi masukan tersebut menjadi kata-kata individual - disebut sebagai token - yang mencakup tanda baca dan simbol lainnya. Token-token tersebut dijalankan melalui kamus yang dapat mengidentifikasi sebuah kata dan bagian datanya. Token-token tersebut kemudian dianalisis untuk struktur tata-bahasa mereka, termasuk peran kata dan kemungkinan ambiguitas makna yang berbeda.

2.4. IndoBERT

BERT merupakan singkatan dari *Bidirectional Encoder Representations from Transformers*. BERT dirancang agar dapat melakukan pelatihan dua arah dari teks yang tidak berlabel dengan mengkondisikan bersama pada konteks kiri dan kanan di semua lapisan [2]. Arsitektur model BERT adalah *encoder Transformer* dua arah *multi-layer*. Terdapat dua tahapan yang dilakukan pada BERT, yaitu tahapan *pre-training* dan tahapan *fine-tuning*. Selama *pre-training*, model dilatih dengan menggunakan data yang tidak berlabel dan dilatih dengan tugas yang berbeda, sedangkan untuk tahapan *fine-tuning*, model BERT pertama kali diinisialisasi dengan parameter yang telah dilatih sebelumnya, dan semua parameter diatur dengan menggunakan data berlabel [2].

IndoBERT mempunyai arsitektur yang sama dengan BERT, namun yang membedakan IndoBERT dengan BERT adalah dataset yang digunakan pada tahapan *pre-training*. Dataset yang digunakan untuk melatih IndoBERT disebut sebagai Indo4B yang terdiri dari sekitar 4 miliar kata dengan sekitar 250 juta kalimat, dataset Indo4B mencakup kalimat bahasa Indonesia formal dan sehari-hari yang disusun dari 15 dataset, yang dua diantaranya mencakup Bahasa Indonesia sehari-hari, delapan mencakup Bahasa Indonesia formal, dan sisanya memiliki gaya campuran antara bahasa sehari-hari dan formal [2].

Arsitektur IndoBERT didasarkan pada arsitektur dasar *transformer*, arsitektur dasar ini mencakup beberapa komponen kunci [2]:

1. *Encoder Transformer*: Arsitektur *transformer* terdiri dari beberapa blok *encoder* yang bertugas untuk mengkonversi input teks menjadi representasi numerik yang dapat diproses lebih lanjut oleh model. Setiap blok *encoder* terdiri dari beberapa lapisan *self-attention* dan lapisan-lapisan *fully connected feedforward networks*.
2. *Multi-Head Self-Attention*: *Self-attention* adalah mekanisme yang memungkinkan model untuk "menggunakan" informasi dari setiap kata dalam kalimat untuk memahami konteksnya. *Multi-head self-attention* memungkinkan model untuk memperhatikan hubungan antara kata-kata dari beberapa sudut pandang yang berbeda.

3. *Layer Normalization*: Setiap lapisan *encoder* dilengkapi dengan normalisasi lapisan untuk mempercepat dan menjaga stabilitas pelatihan model.
4. *Feedforward Networks*: Setelah *self-attention*, representasi kata-kata dari input dikirim melalui jaringan *feedforward* untuk ekstraksi fitur lanjutan.
5. *Pre-Training* dan *Fine-Tuning*: Model IndoBERT, seperti BERT, menjalani dua tahap pelatihan: tahap *pre-training* dan tahap *fine-tuning*. Tahap *pre-training* dilakukan di atas data teks besar yang tidak diawasi, sementara tahap *fine-tuning* dilakukan pada tugas-tugas khusus, seperti pemahaman teks atau klasifikasi, dengan menggunakan data yang berlabel.

Arsitektur IndoBERT, khususnya, dioptimalkan untuk bahasa Indonesia dengan menyesuaikan parameter dan metode pelatihan pada data bahasa Indonesia. Hal ini memungkinkan model untuk memahami dan memproses teks dalam bahasa Indonesia dengan baik.

2.5. GPT by OpenAI

Kecerdasan buatan dan model bahasa generatif seperti ChatGPT memberikan potensi yang menarik. ChatGPT, yang merupakan singkatan dari *Chat Generative Pre-Trained Transformer*, merupakan contoh *chatbot*. *Chatbot* sendiri adalah program kecerdasan buatan yang terintegrasi dengan teknologi AI (*artificial intelligence*) yang mampu memberikan respon terhadap pertanyaan pengguna dengan cara yang mirip dengan interaksi manusia, tetapi dalam bentuk teks otomatis [8]. Adopsi teknologi ChatGPT memberikan peluang bagi pemanfaatan *chatbot* AI dalam dunia pendidikan di Indonesia, terutama dalam pengembangan keterampilan peserta didik yang relevan dengan tuntutan abad ke-21. Melalui kemajuan teknologi ini, sumber belajar tidak hanya terbatas pada pendekatan yang dikendalikan oleh pendidik, namun juga melibatkan pemanfaatan alat bantu yang dapat mempercepat akses terhadap berbagai sumber belajar dengan cakupan yang lebih luas (*broad-based learning*). Dalam konteks ini, teknologi menjadi suatu disiplin

ilmu yang menjadi kebutuhan esensial bagi pendidik dan peserta didik dalam menghadapi tantangan pembelajaran abad ke-21.

Deskripsi tentang ChatGPT yang peneliti berikan sesuai dengan fungsinya sebagai sistem kecerdasan buatan atau *artificial intelligence* (AI) yang beroperasi dalam format percakapan. Secara sederhana, ChatGPT dapat dianggap sebagai "guru virtual" yang memberikan jawaban otomatis kepada pertanyaan yang diajukan oleh pengguna, menirukan dinamika percakapan antara peserta didik dan guru dalam suatu kelas. Dengan demikian, ChatGPT berperan sebagai entitas pembantu yang menyediakan informasi dan jawaban dalam format interaktif. ChatGPT merupakan produk dari perusahaan AI non-profit bernama Open AI yang didirikan pada tahun 2015 yang di inovasikan oleh Elon Musk dan beberapa tokoh terkenal seperti Reid Hoffman dan Sam Altman. Dibalik canggihnya ChatGPT ini juga terdapat beberapa kelebihan dan kekurangan, contoh dari kelebihan ChatGPT ini adalah kecepatan dalam memberikan jawaban yang sifatnya sangat responsif dan bahasa yang mudah dipahami. Akan tetapi ChatGPT ini bukan tanpa kelemahan, ada beberapa kelemahan dalam ChatGPT ini, seperti

1. Ketidakakuratan informasi, kita ketahui ChatGPT ini merupakan model yang dibuat dari data dan pelatihan mesin pencari sehingga informasi yang diberikan oleh ChatGPT mungkin saja tidak sesuai atau akurat dengan fakta yang ada
2. Keterbatasan pemahaman, walaupun dapat memahami bahasa manusia akan tetapi pemahaman ini masih terbatas pada data yang ada seperti halnya perambah Google
3. Tidak memiliki emosi. ChatGPT tidak punya kemampuan emosi dan tidak dapat memberikan respon secara emosional
4. Dari sisi keamanan. dimana program algoritma pada ChatGPT dapat digunakan untuk tujuan yang tidak etis dan tidak bermoral atau malah melakukan kejahatan kriminal oleh krn itu perlu adanya pengawasan dan kendali dalam penggunaan ChatGPT
5. ChatGPT ini membutuhkan data dalam jumlah besar supaya dapat berfungsi dengan baik, dan karena itu juga perlu perhatian dan

perlindungan terhadap privasi dan perlindungan data ketika menggunakan teknologi ini

Arsitektur GPT terdiri dari beberapa komponen kunci, masing-masing memainkan peran penting dalam memahami dan menghasilkan teks. Berikut komponen-komponen ini secara detail:

- *Transformer architecture*

Pilar dari model-model GPT adalah arsitektur *transformer*. Arsitektur ini terdiri dari struktur *encoder-decoder*, tetapi dalam kasus GPT, hanya *decoder* yang digunakan. Arsitektur *transformer* terdiri dari beberapa lapisan, masing-masing mengandung mekanisme *self-attention* dan jaringan saraf *feedforward*.

- *Positional encoding*

Untuk menangani sifat berurutan dari bahasa, model-model GPT menggunakan pemetaan posisi. Pemetaan ini ditambahkan ke *embedding input* untuk memberikan informasi tentang posisi setiap *token* dalam urutan.

- *Token embeddings*

Model-model GPT menggunakan *embedding token* untuk merepresentasikan teks masukan. Setiap token di-*embed* ke dalam ruang vektor kontinu, yang memungkinkan model untuk memproses teks.

- *Layer stacking*

Model-model GPT umumnya menumpuk beberapa lapisan *transformer* satu di atas yang lain untuk menangkap pola-pola kompleks dalam teks. Jumlah lapisan adalah hiperparameter yang dapat disesuaikan berdasarkan tugas dan sumber daya yang tersedia.

2.6. *Gemini by Google*

Gemini adalah model dasar multimodal milik Google yang perusahaan ini sedang mengintegrasikan ke beberapa produknya. Gemini merupakan jawaban Google terhadap GPT-4 milik OpenAI, model bahasa besar multimodal (LLM) yang menggerakkan versi berbayar dari ChatGPT,

keberhasilan yang memulai perlombaan senjata AI generatif, karena beberapa perusahaan teknologi sejak itu berlomba-lomba untuk menghadirkan produk-produk terbaru dan terbaik ke pasar.

Diluncurkan pada Desember 2023, Gemini adalah model terbesar dan paling canggih milik Google hingga saat ini, menurut perusahaan. Model ini dikembangkan oleh laboratorium riset kecerdasan buatan Google, DeepMind, dan Google Research, dan merupakan hasil dari hampir satu dekade kerja.

Model ini tersedia dalam tiga versi yang berbeda, yang bervariasi dalam ukuran dan kompleksitas:

1. Gemini Ultra

Gemini ultra adalah model "terbesar" dan "paling canggih" untuk melakukan tugas-tugas yang sangat kompleks, menurut Google. Perusahaan mengatakan bahwa model ini melebihi kinerja GPT-4 pada sebagian besar *benchmark* akademis yang paling banyak digunakan dalam penelitian dan pengembangan LLM, serta berbagai tugas *multimodal*. Model ini sedang diintegrasikan ke beberapa produk paling populer milik Google, termasuk Gmail, Docs, Slides, dan Meet. Pengguna dapat mengakses Gemini Ultra melalui layanan Gemini Advanced dengan biaya \$19.99 per bulan.

2. Gemini Pro

Gemini Pro adalah model tingkat menengah, dirancang untuk memahami pertanyaan-pertanyaan kompleks dan merespon dengan cepat, menjadikannya model terbaik untuk "skalabilitas melintasi berbagai tugas," seperti yang dijelaskan oleh Google. Versi khusus yang terlatih dari Pro saat ini menggerakkan chatbot AI Gemini dan tersedia melalui API Gemini di Google AI Studio dan Google Cloud Vertex AI.

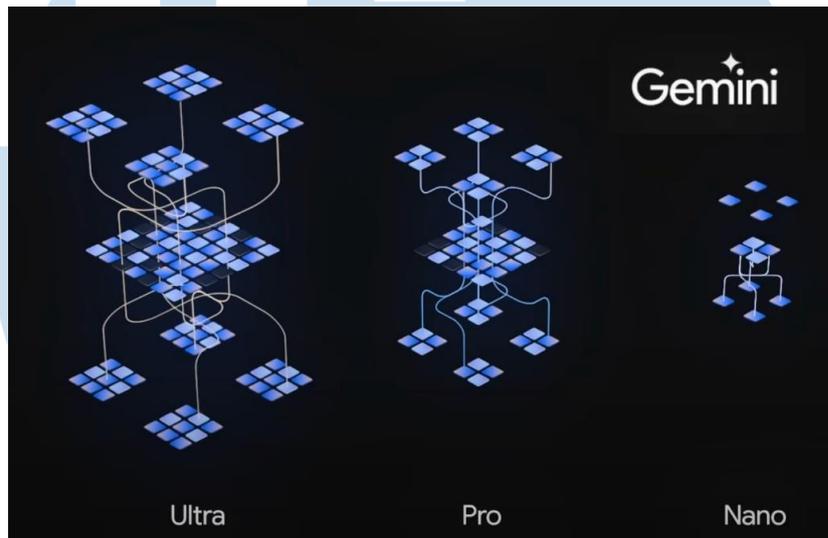
3. Gemini Nano

Sebuah versi yang jauh lebih kecil dari model Pro dan Ultra, Gemini Nano dirancang untuk menjadi cukup efisien untuk melakukan tugas-tugas secara langsung pada perangkat pintar, daripada harus terhubung ke server eksternal. Nano saat ini menggerakkan fitur-fitur pada Pixel 8 Pro seperti

Ringkasan di aplikasi Recorder dan Balasan Cerdas di aplikasi *keyboard virtual* Gboard.

Berikut arsitektur model pada Gemini sebagai berikut:

1. *A Suite of Models for any Application*



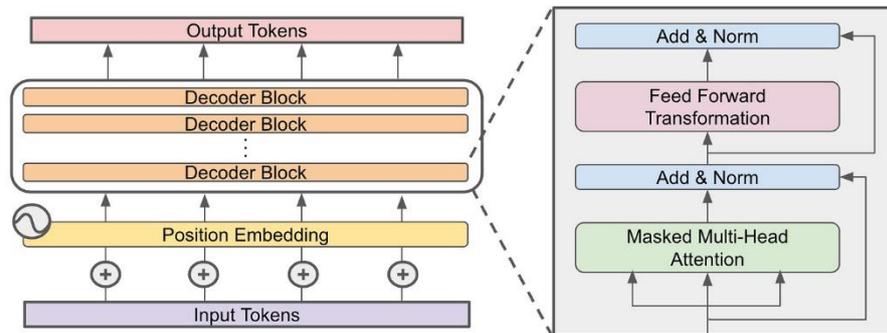
Gambar 2. 2 Rangkaian Model Gemini

Gemini bukanlah satu model tunggal, tetapi merupakan rangkaian beberapa LLM (*Large Language Models*). Setiap LLM ini memiliki ukuran yang berbeda, memberikan masing-masing dari mereka kompromi yang bervariasi antara efisiensi dan kapasitas pemecahan masalah. Tujuan dari Gemini adalah untuk merilis serangkaian model yang mampu dan cocok untuk berbagai aplikasi, baik itu di perangkat maupun memecahkan tugas-tugas penalaran kompleks dengan akses ke sumber daya komputasi yang besar (misalnya, pusat data). Berbagai ukuran model. Rilis awal Gemini mencakup tiga varian model yang berbeda:

- Gemini Ultra: model terbesar dan paling canggih yang mencapai kinerja terbaik pada banyak tugas yang sangat kompleks.
- Gemini Pro: Sebuah model yang dioptimalkan kinerjanya (dalam hal kualitas dan latensi) yang berkinerja baik dalam banyak tugas dan dapat diterapkan secara besar-besaran.

- Gemini Nano: sekelompok model yang dirancang untuk berjalan dengan efisien pada perangkat edge meskipun mencapai kinerja yang mengesankan.

2. Decoder-only Architecture



Gambar 2.3 Decoder-only transformer architecture

Seperti banyak LLM generatif, Gemini didasarkan pada *decoder-only transformer architecture*; lihat di atas untuk gambarannya. Detail-detail tentang arsitektur model tidak dijelaskan secara eksplisit dalam laporan teknis, tetapi kita diberikan cukup informasi untuk (secara kasar) menyimpulkan beberapa detail yang relevan tentang model tersebut.

3. Multimodal Language Models

Model-model Gemini bersifat multimodal, yang berarti bahwa mereka dapat menerima input dari dan menghasilkan output dengan beberapa modalitas data. Upaya sebelumnya dalam menciptakan model *multimodal* sangat menekankan dua modalitas - teks dan gambar - dan cenderung menggunakan pendekatan pasca-hoc yaitu i) melatih model-model terpisah pada setiap modalitas dan ii) "menjahit" model-model ini bersama dengan penyetulan lanjutan.

4. Model Training and Dataset

Campuran data *pre-training* yang tepat yang digunakan untuk Gemini dikecualikan dari laporan teknis. Namun, campuran data yang digunakan untuk melatih LLM telah ditunjukkan oleh karya sebelumnya (misalnya, Falcon, MPT, BioMedLM, dll.) sebagai faktor besar yang mempengaruhi kualitas model. Di [1], penulis menyebutkan beberapa pertimbangan yang

memberikan wawasan yang berguna ke dalam proses pelatihan dan dataset yang digunakan untuk Gemini.

2.7. *Google Vertex AI*

Vertex AI adalah platform machine learning (ML) yang memungkinkan Anda melatih dan men-*deploy* model ML dan aplikasi AI, serta menyesuaikan model bahasa besar (LLM) untuk digunakan dalam aplikasi yang didukung teknologi AI. Vertex AI menggabungkan *data engineering*, *data science*, dan alur kerja *engineering* ML sehingga tim Anda dapat berkolaborasi menggunakan kumpulan alat yang umum dan menskalakan aplikasi menggunakan manfaat Google Cloud.

Vertex AI menyediakan beberapa opsi untuk pelatihan dan *deployment* model:

- AutoML memungkinkan Anda melatih data tabulasi, gambar, teks, atau video tanpa perlu menulis kode atau menyiapkan pemisahan data.
- Pelatihan kustom memberi Anda kontrol penuh atas proses pelatihan, termasuk menggunakan *framework* ML pilihan Anda, menulis kode pelatihan Anda sendiri, dan memilih opsi penyesuaian *hyperparameter*.
- Dengan Model Garden, Anda dapat menemukan, menguji, menyesuaikan, dan men-*deploy* Vertex AI serta memilih model dan aset *open-source* (OSS).
- AI Generatif memberi Anda akses ke model AI generatif besar dari Google untuk berbagai modalitas (teks, kode, gambar, ucapan). Anda dapat menyesuaikan LLM Google untuk memenuhi kebutuhan, lalu men-*deploy*-nya untuk digunakan dalam aplikasi yang didukung teknologi AI.

2.8. *Multi-Aspect Based Sentiment Analysis*

Analisa sentimen berbasis multi aspek (*Multi-Aspect based Sentiment Analysis*) merupakan perkembangan dari analisa sentimen yang hanya mengacu pada lebih dari satu label. Analisa sentimen berbasis multi aspek dari sebuah kalimat merupakan sebuah opini yang mengacu kepada entitas yang spesifik dan aspek yang dibahasnya. Analisa sentimen berbasis aspek

bertujuan untuk mendeteksi polaritas teks tertulis berdasarkan dengan aspek tertentu.. Berdasarkan kebutuhan TokoTeknoPro, beberapa penilaian yang akan dijadikan sebagai penilaian antara lain:

2.8.1. Kualitas Produk

Analisa sentimen berdasarkan aspek kualitas produk adalah proses untuk mengevaluasi pandangan dan perasaan konsumen terhadap kualitas produk tertentu. Ini melibatkan pengumpulan, pemrosesan, dan analisis ulasan, pendapat, atau umpan balik yang diberikan oleh konsumen tentang produk tersebut.

Dalam analisa sentimen berdasarkan kualitas produk, tujuannya adalah untuk menentukan apakah pandangan konsumen terhadap kualitas produk tersebut negatif. Hal ini dapat dilakukan dengan menggunakan teknik-teknik pemrosesan bahasa alami untuk mengidentifikasi kata-kata atau frasa-frasa yang mengindikasikan sentimen positif atau negatif dalam ulasan atau pendapat konsumen.

2.8.2. Kecepatan Pengiriman

Analisa sentimen berdasarkan aspek kecepatan pengiriman adalah proses untuk mengevaluasi opini dan perasaan pelanggan terhadap kecepatan atau waktu pengiriman suatu produk atau layanan. Ini melibatkan pengumpulan, pengolahan, dan analisis umpan balik, ulasan, atau pendapat yang diberikan oleh pelanggan tentang pengalaman mereka dalam menerima produk atau layanan dalam waktu yang diharapkan.

Dalam analisa sentimen berdasarkan kecepatan pengiriman, tujuannya adalah untuk menentukan apakah pandangan pelanggan terhadap kecepatan pengiriman tersebut negatif. Hal ini dapat dilakukan dengan menggunakan teknik-teknik pemrosesan bahasa alami untuk mengidentifikasi kata-kata atau frasa-frasa yang menunjukkan kepuasan atau ketidakpuasan terhadap kecepatan pengiriman.

2.8.3. Layanan Penjual

Analisa sentimen berdasarkan aspek layanan penjualan adalah proses evaluasi terhadap pendapat dan perasaan pelanggan terhadap pelayanan yang diberikan selama proses penjualan suatu produk atau layanan. Ini melibatkan pengumpulan, pemrosesan, dan analisis umpan balik, ulasan, atau pendapat yang diberikan oleh pelanggan tentang interaksi mereka dengan staf penjualan, proses pembelian, dan layanan yang diterima.

Dalam analisa sentimen berdasarkan layanan penjualan, tujuannya adalah untuk menentukan apakah pandangan pelanggan terhadap layanan penjualan tersebut negatif. Hal ini dapat dilakukan dengan menggunakan teknik-teknik pemrosesan bahasa alami untuk mengidentifikasi kata-kata atau frasa-frasa yang menggambarkan kepuasan atau ketidakpuasan terhadap layanan penjualan.

2.8.4. Pengemasan

Analisa sentimen berdasarkan pengemasan adalah proses evaluasi terhadap pendapat dan perasaan pelanggan terhadap kualitas dan desain kemasan suatu produk. Ini melibatkan pengumpulan, pemrosesan, dan analisis umpan balik, ulasan, atau pendapat yang diberikan oleh pelanggan tentang pengalaman mereka dengan kemasan produk.

Dalam analisa sentimen berdasarkan pengemasan, tujuannya adalah untuk menentukan apakah pandangan pelanggan terhadap kemasan produk tersebut negatif. Hal ini dapat dilakukan dengan menggunakan teknik-teknik pemrosesan bahasa alami untuk mengidentifikasi kata-kata atau frasa-frasa yang menggambarkan kepuasan atau ketidakpuasan terhadap kualitas, desain, atau fungsionalitas kemasan.

2.8.5. Harga

Analisa sentimen berdasarkan harga adalah proses evaluasi terhadap pandangan dan perasaan pelanggan terhadap harga suatu produk atau layanan. Ini melibatkan pengumpulan, pemrosesan, dan analisis umpan balik, ulasan, atau pendapat yang diberikan oleh pelanggan tentang harga produk atau layanan yang mereka beli.

Dalam analisa sentimen berdasarkan harga, tujuannya adalah untuk menentukan apakah pandangan pelanggan terhadap harga tersebut negatif. Hal ini dapat dilakukan dengan menggunakan teknik-teknik pemrosesan bahasa alami untuk mengidentifikasi kata-kata atau frasa-frasa yang menggambarkan kepuasan atau ketidakpuasan terhadap harga.

2.9. *Precision, Recall, F1-Score, Accuracy*

Pengujian *Precision, Recall, Accuracy* dan *F1-Score* adalah metode evaluasi yang umum digunakan dalam pemrosesan bahasa alami dan bidang *machine learning* lainnya untuk mengukur kinerja dari model klasifikasi atau model yang melakukan tugas-tugas serupa. Metode ini biasanya digunakan dalam konteks klasifikasi biner, di mana model diminta untuk mengklasifikasikan *instance-instance* dalam dua kelas yang berbeda, seperti positif dan negatif, atau benar dan salah. Berikut adalah penjelasan singkat dari masing-masing pengujian:

1. *Precision*: *Precision* (presisi) mengukur seberapa banyak *instance* yang diklasifikasikan sebagai positif oleh model yang sebenarnya merupakan positif. Ini diperoleh dengan membagi jumlah *instance* yang diklasifikasikan sebagai positif secara benar oleh jumlah total *instance* yang diklasifikasikan sebagai positif (baik yang benar maupun yang salah).

$$\text{Presisi} = \frac{\text{TruePositives}}{\text{TruePositives} + \text{FalsePositives}}$$

2. *Recall*: *Recall* (*recall*) mengukur seberapa banyak *instance* positif yang berhasil diidentifikasi oleh model dari semua *instance* positif yang sebenarnya ada. Ini diperoleh dengan membagi jumlah *instance* yang diklasifikasikan sebagai positif secara benar oleh jumlah total *instance* positif yang sebenarnya ada.

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives}$$

3. *F1-Score*: *F1-Score* adalah metrik gabungan yang mengukur keseimbangan antara presisi dan *recall*. Ini adalah rata-rata harmonis dari presisi dan *recall*, dan memberikan gambaran yang lebih baik tentang kinerja model ketika ada ketidakseimbangan antara jumlah *instance* positif dan negatif dalam dataset.

$$F1Score = 2x \frac{Precision \times Recall}{Precision + Recall}$$

4. *Accuracy*: *accuracy* adalah metrik gabungan yang mengukur kemampuan model untuk dengan tepat mengidentifikasi sentimen dari teks yang diberikan.

$$Accuracy = \frac{CorrectPrediction}{Totaldata}$$

Dalam pengujian *Precision*, *Recall*, *F1-Score*, dan *Accuracy*, terdapat beberapa aspek yang perlu dipertimbangkan untuk mendapatkan evaluasi yang komprehensif terhadap kinerja model klasifikasi. Beberapa aspek tersebut meliputi:

3. *True Positives* (TP): Jumlah *instance* yang diklasifikasikan dengan benar sebagai positif oleh model.
4. *True Negatives* (TN): Jumlah *instance* yang diklasifikasikan dengan benar sebagai negatif oleh model.
5. *False Positives* (FP): Jumlah *instance* yang seharusnya diklasifikasikan sebagai negatif, tetapi salah diklasifikasikan sebagai positif oleh model.
6. *False Negatives* (FN): Jumlah *instance* yang seharusnya diklasifikasikan sebagai positif, tetapi salah diklasifikasikan sebagai negatif oleh model.
7. *Precision*: Proporsi dari *instance* positif yang diklasifikasikan dengan benar oleh model terhadap total *instance* yang diklasifikasikan sebagai positif oleh model.

8. *Recall*: Proporsi dari total *instance* positif yang berhasil diidentifikasi oleh model terhadap total *instance* positif yang sebenarnya ada dalam *dataset*.
9. *F1-Score*: Rata-rata harmonis dari *precision* dan *recall*, memberikan gambaran yang lebih baik tentang kinerja model ketika ada ketidakseimbangan antara jumlah *instance* positif dan negatif dalam *dataset*.
10. Akurasi (*Accuracy*): Proporsi dari total *instance* yang diklasifikasikan dengan benar oleh model terhadap total *instance* dalam *dataset*.
11. Kompromi antara *Precision* dan *Recall*: Dalam beberapa kasus, terdapat *trade-off* antara *precision* dan *recall*, di mana peningkatan *precision* dapat mengurangi *recall*, dan sebaliknya. Pemilihan nilai *threshold* atau pengaturan model tertentu dapat mempengaruhi kompromi ini.
12. Kurva *Precision-Recall*: Kurva yang menggambarkan hubungan antara *precision* dan *recall* pada berbagai *threshold* yang berbeda, membantu dalam memilih *threshold* yang optimal untuk tujuan klasifikasi tertentu.

Dengan mempertimbangkan aspek-aspek ini, evaluasi kinerja model klasifikasi dapat menjadi lebih komprehensif dan informatif.

2.10. Temperatur

Temperatur atau yang biasa dikenal juga dengan *temp* merupakan salah satu hyperparameter yang sering digunakan untuk menentukan pola acak dan variasi dari sebuah LLM. Variabel ini dapat menyesuaikan model untuk berbagai jenis tugas, dari yang membutuhkan konsistensi tinggi hingga kreativitas dan variasi dan model yang diinginkan.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A