BAB 2
LANDASAN TEORI

2.1 Artificial Intelligence (AI)

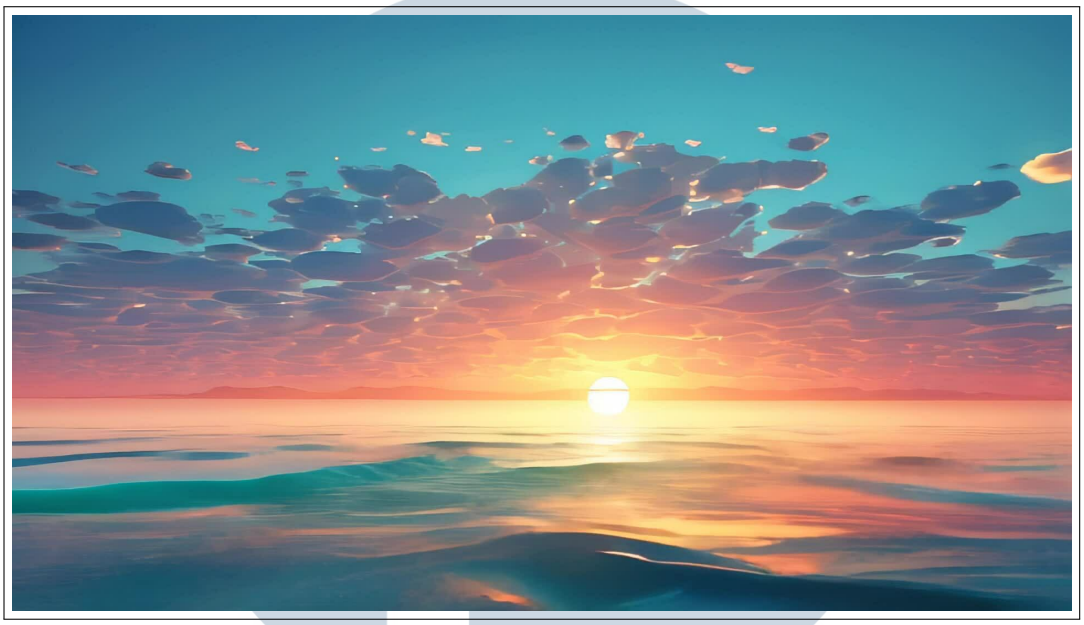
Kecerdasan buatan atau *Artificial Intelligence (AI)* merupakan sistem *Machine-Based* yang mampu membuat sebuah rekomendasi, prediksi, dan diagnosa yang sama seperti manusia[19]. Beberapa tahun belakangan, dampak dari kemajuan yang terjadi pada *AI* telah memasuki banyak bidang seperti bisnis, agrikultur, kesehatan, manufaktur, pendidikan, dan lainnya[19][5]. Meskipun begitu, *AI* dianggap dapat mengancam penyaluran tenaga kerja dikarenakan kompleksitas pekerjaan yang mulai bisa diambil alih oleh kecerdasan buatan demi efisiensi pekerjaan[19][6].

2.2 Twitter

Dalam beberapa tahun terakhir, penggunaan kecerdasan buatan dalam proses pembuatan karya seni digital telah berkembang pesat, dengan salah satu platform media sosial utama yang digunakan untuk mengekspresikan hasil kreasi para pengguna merupakan Twitter[20]. Banyak pengguna Twitter yang menghargai kreativitas dan hasil yang telah dibentuk dari kecerdasan buatan, namun masih ada kekhawatiran mengenai karya yang dibuat oleh mesin dengan berbagai macam alasan, seperti nilai artistik dan orisinalitas dari karya-karya yang telah dibuat[20]. Penelitian yang telah dilakukan oleh Kusumaningrum dan Susanto menunjukkan bahwa analisis sentimen di media sosial dapat dilakukan untuk menganalisis data dari berbagai platform sosial, termasuk Twitter, dan menemukan bahwa reaksi pengguna dapat dibedakan menjadi positif, negatif, dan netral[20].

Dengan menggunakan API Twitter, data dapat diperoleh dan dikumpulkan untuk melakukan berbagai macam analisis, seperti analisis sentimen dengan menggunakan *Natural Language Processing (NLP)*.[20]. NLP dapat digunakan untuk mengelompokkan komentar dan interaksi pengguna Twitter untuk mendapatkan data yang relevan dengan efisien dan dinamis[20]. Untuk mendapatkan pemahaman lebih lanjut mengenai analisis sentimen dari topik yang diinginkan, perlu ditekankan pentingnya penggunaan analisis yang tepat untuk meningkatkan akurasi dalam proses pengukuran sentimen[20].

2.3 AI-generated artwork



(a) Sunset over the Ocean[21]

(b) Spongebob Squarepants[22]

(c) End of Humanity[23]

Gambar 2.1. Contoh hasil generasi karya digital buatan AI.

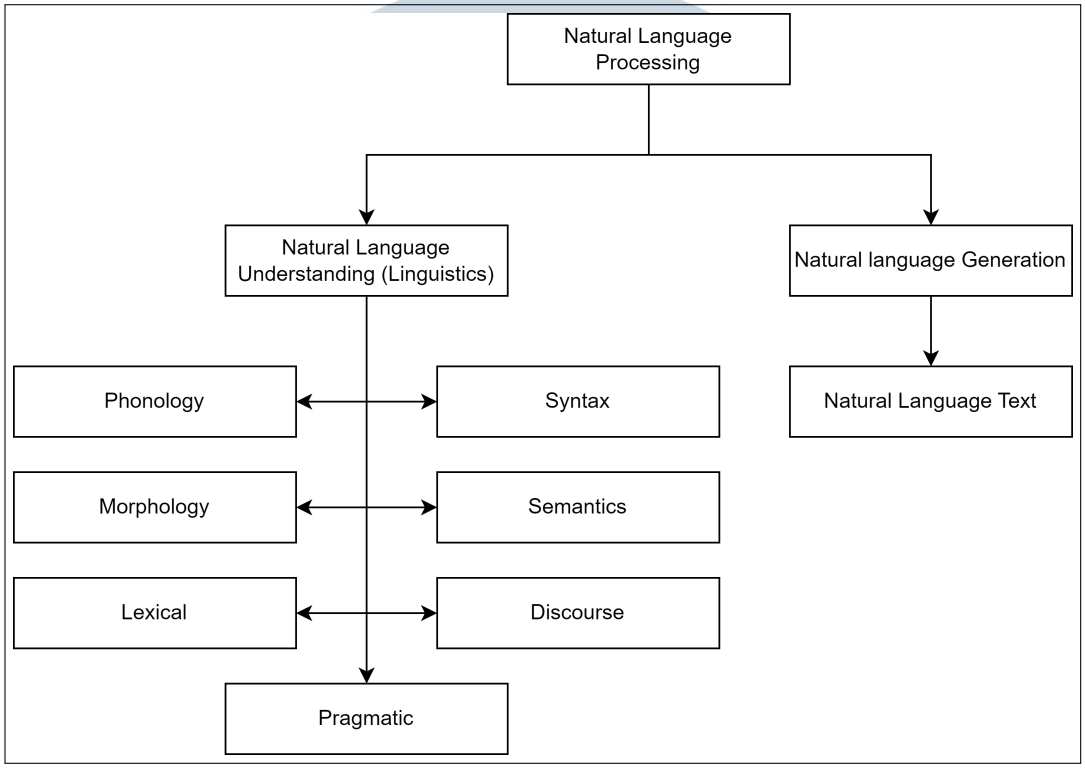
AI-generated artwork merupakan setiap bentuk seni digital yang dibuat melalui penggunaan kecerdasan buatan menggunakan algoritma. Penggunaan *AI* dalam pembuatan karya seni membawa banyak dampak positif dan negatif terhadap perkembangan teknologi, dengan plagiarisme dan pemanfaatan *AI* yang berkesan

tidak bermoral atau beretika, serta melanggar hak cipta dari seniman yang karyanya digunakan tanpa izin dalam proses pembuatan karya digital[7]. Penggunaan kecerdasan buatan dalam proses pembuatan karya seni merupakan topik yang penuh dengan pro dan kontra.

Gambar 2.2. Website Midjourney[1]

Salah satu contoh dari program aplikasi yang digunakan untuk membuat karya digital dengan bantuan proses dari kecerdasan buatan adalah *Midjourney*. *Midjourney* merupakan program kecerdasan buatan generatif yang digunakan melalui ketikan berdeskriptif[8]. Berdasarkan data statistik yang diberikan oleh *Midjourney*, mereka telah mendapatkan 14 juta pengguna yang telah mendaftar akun untuk menggunakan *Midjourney*[9].

2.4 Natural Language Processing (*NLP*)



Gambar 2.3. Terminologi NLP[2]

Natural Language Processing (NLP) merupakan cabang ilmu *Artificial Intelligence* yang digunakan untuk memahami dan mereplikasi bahasa manusia melalui teks dan/atau kata yang terucapkan layaknya seorang manusia [15] [16]. NLP biasa dikaitkan dengan berbagai teori dan teknik dalam menghadapi masalah dari berkomunikasi dengan komputer menggunakan bahasa manusia, dan terdiri dari dua komponen, yaitu *Natural Language Understanding* (*Linguistic*) (NLU), dan *Natural Language Generator* (*NLG*) [15].

1. *Phonology* *Phonology* merupakan bagian dari linguistik yang mengacu kepada rangkaian suara secara sistematis. Kata *Phonology* berasal dari Yunani kuno dengan *phono* berarti suara, dan suffix *-logy* teracukan terhadap kata atau kalimat. Pada tahun 1993, Nikolai Trubetzkoy (Linguistik dan Sejarahwan Russia) menyatakan bahwa *Phonology* merupakan ”Ilmu pembelajaran suara yang berkaitan dengan sistem perbahasaan” [15].

   Contoh dari penggunaan *Phonology* adalah ketika memanggil nama seseorang dengan nada biasa dan naga tegas.

2. *Morphology* Bagian yang berbeda dari suatu kata merepresentasikan unit makna terkecil yang dikenal sebagai *Morphemes*. *Morphology* terbentuk dari dasar dari perkataan, yang dimulai dari *morphemes*. Contoh *Morpheme* adalah kata *precancellation* yang secara *morphologically* bisa disederhanakan menjadi tiga *morphemes* berbeda; *Prefix pre*, *root cancella*, dan *suffix -tion*. Interpretasi dari *morphemes* tetap bermakna sama di semua kata. Layaknya menambahkan *verb -ed* untuk sebuah verb (*Work = Worked, Study = Studied*) yang menandakan suatu aksi dari *verb* itu yang sudah berada di masa lalu atau sudah terjadi [15].

3. *Lexical* Dalam *Lexical*, manusia, termasuk dengan sistem NLP, menginterpretasi makna dari kata-kata secara individu. Dalam proses ini, kata yang bisa menjadi lebih dari satu bagian dari suatu rangkaian kalimat yang diberikan berdasarkan konteks yang ada. Secara lebih sederhana, *Lexical* merupakan suatu upaya untuk memahami makna dari suatu kata atau kalimat berdasarkan hubungan satu sama lain dan konteks mereka [15].

   Contoh dari penggunaan *Lexical* adalah kata ’kunci’, yang bisa berarti solusi dari suatu masalah atau alat yang digunakan untuk membuka pintu, pagar, dan sebagainya.

4. *Syntax* *Syntax* merupakan proses rangkaian dari kata-kata dalam sebuah kalimat untuk membentuk suatu konteks yang masuk akal, karena dengan sekedar mengganti urutan suatu kata, konteks dari kalimat tersebut bisa saja berubah. Contohnya, ”*Shyam beats Ram in a competition*” dan ”*Ram beats Shyam in a competition*” memiliki dan makna konteks yang berbeda [15].

5. *Semantic* Dalam *Semantic*, *Semantic* adalah proses analisa dari format grammar suatu kalimat, termasuk rangkaian kalimat, frasa, dan klausa untuk

mengdeterminasikan hubungan dari suatu konteks. *Semantic* bertujuan untuk mencari makna yang sesuai dari suatu kalimat. Manusia bisa memahami pengetahuan dari logika dan konsep yang terdapat dari kalimat tersebut, namun mesin tidak dapat mengandalkan teknik-teknik ini. [15].

Contoh dari penggunaan *Semantic* adalah kata 'cahaya hidup' dari kalimat "Dia adalah cahaya hidupku". 'Cahaya hidup' dapat dianggap sebagai metafor tentang sesuatu yang sangat berarti.

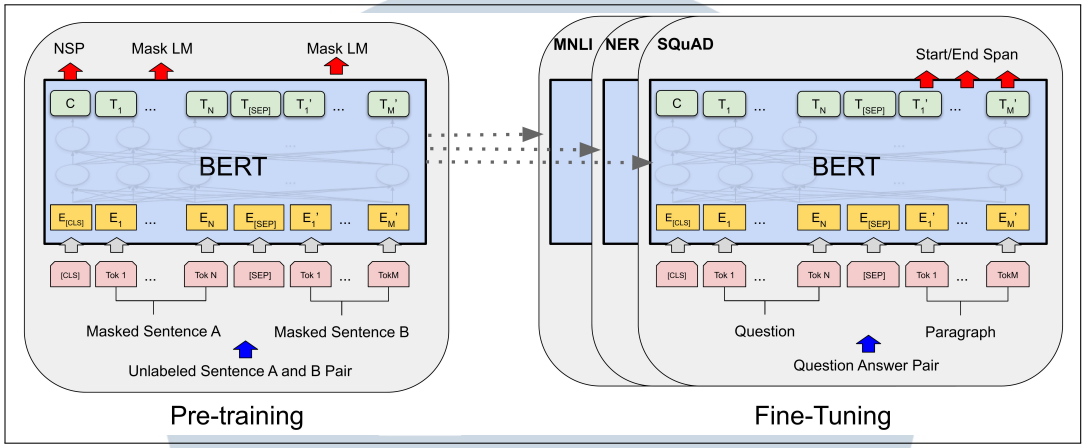
6. *Discourse* Ketika *syntax* dan *semantic* mengatasi unit kalimat biasa, *discourse* merupakan suatu bahasa yang terdiri dari beberapa kalimat. *Discourse* mengatasi analisis struktur logika dengan membuat hubungan antar kata dan kalimat untuk memastikan koherensi mereka. *Discourse* mencoba untuk melihat pesan dari setiap kalimat dan mencoba untuk mencari interpretasi logikal antar kalimat untuk memastikan linguistik kata yang akurat [15].

7. *Pragmatic Pragmatic* merupakan pembelajaran dari aspek perilaku manusia dan pikiran atau pembelajaran dari penggunaan tanda linguistik, kata, dan kalimat di situasi nyata. *Pragmatic* fokus dalam pengetahuan atau isi yang datang dari luar isi dokumen. *Pragmatic* menganalisis konteks demi mendapatkan interpretasi yang bermakna. Keambiguan *pragmatic* akan muncul ketika kalimat yang dianalisis tidak spesifik dan konteks yang disediakan tidak cukup. Sementara *semantic* berusaha mencari makna suatu kalimat secara literal, *pragmatic* justru mencari makna tersirat dari suatu kalimat. Contohnya adalah "Apakah anda tahu jam berapa ini?" akan di-interpretasikan oleh *semantic* sebagai "Mempertanyakan jam sekarang", sementara *pragmatic* akan mendapatkan konklusi "Menyatakan kekesalan terhadap seseorang yang melewati jam (telat)", sehingga analisis *pragmatic* bisa dinyatakan sebagai pembelajaran konteks yang mendorong pemahaman kita mengenai ekspresi linguistik [15].

2.5 Bidirectional Encoder Representations from Transformers (*BERT*)

Bidirectional Encoder Representations from Transformers (*BERT*) adalah model representasi bahasa yang menggunakan arsitektur Transformer secara mendalam dan bidireksional[24]. *BERT* mempelajari konteks teks dari kedua arah (kiri dan kanan) dengan menggunakan teknik *pre-training Masked Language Model* (*MLM*) dan *Next Sentence Prediction* (*NSP*)[24]. Pendekatan ini memungkinkan

BERT memahami hubungan kompleks antara kata-kata dan kalimat dalam teks, menghasilkan representasi yang lebih baik untuk berbagai tugas NLP[24].



Gambar 2.4. Pre-Training dan Fine-Tuning BERT[3]

Merujuk pada gambar 2.5, terdapat dua tahapan proses untuk klasifikasi teks, yaitu *pre-training* dan *fine-tuning*. Pada tahapan pre-training, model belajar dari data yang belum terlabel, sedangkan pada tahapan fine-tuning, model BERT akan melakukan proses klasifikasi teks dengan menggunakan parameter yang sudah disiapkan dan diadaptasikan dengan data yang sudah berlabel[25].

Terdapat kemiripan arsitektur pada kedua model BERT dengan perbedaan pada output layer. Pada proses fine-tuning, model dimulai dengan parameter yang terdapat seperti pada pre-training, namun disempurnakan dengan menggunakan data yang sudah berlabel untuk tugas dan fungsi yang spesifik, seperti klasifikasi teks, analisis sentimen, atau entailment [24]. Token unik seperti [CLS] dan [SEP] dimasukkan ke dalam input untuk membantu BERT dalam proses fine-tuning[24].

Pada tahapan pre-training yang dapat dilihat pada gambar kiri, model BERT dilatih menggunakan teknik Masked Language Model (MLM) dan Next Sentence Prediction (NSP). MLM akan menyembunyikan atau menggantikan kata yang terdapat pada suatu kalimat dengan [MASK]. NSP akan melatih model dengan mencari hubungan logis antar kalimat yang terdapat pada teks tersebut.

Pada tahapan fine-tuning, model BERT akan menyesuaikan bobot hasil pre-training terhadap tugas spesifik yang ingin diselesaikan, seperti klasifikasi sentimen, analisis emosi, atau pertanyaan-jawaban. Proses ini dilakukan dengan menggunakan dataset yang telah diberi label sesuai dengan target klasifikasi. Parameter awal yang diperoleh dari pre-training tidak diinisialisasi ulang, melainkan disempurnakan melalui proses pembelajaran tambahan terhadap data

berlabel tersebut. Selama fine-tuning, struktur arsitektur utama dari BERT tetap dipertahankan, namun lapisan akhir (output layer) dimodifikasi sesuai jenis tugas. Misalnya, untuk tugas klasifikasi, lapisan akhir dapat berupa fully connected layer dengan fungsi aktivasi softmax yang mengklasifikasikan keluaran dari token [CLS] ke dalam kategori yang diinginkan. Melalui pendekatan ini, BERT dapat secara efektif memanfaatkan pengetahuan umum dari pre-training dan mengadaptasinya untuk performa yang optimal pada tugas spesifik.

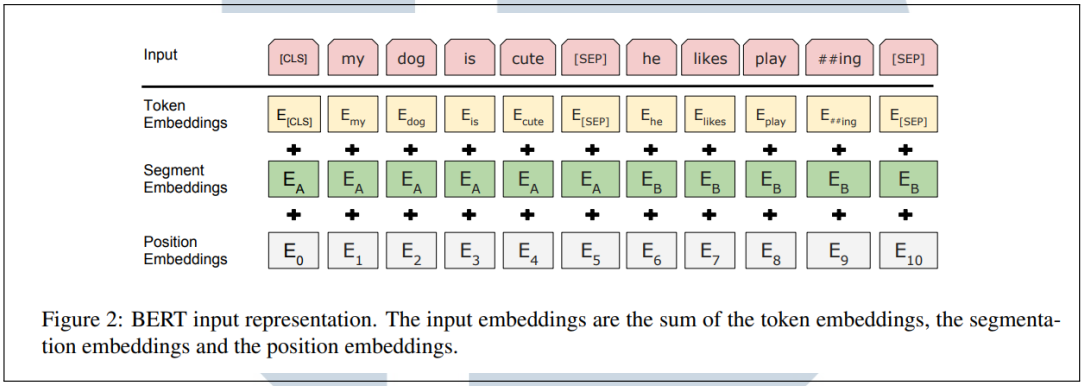


Figure 2: BERT input representation. The input embeddings are the sum of the token embeddings, the segmentation embeddings and the position embeddings.

Gambar 2.5. Input-Output Embedding pada BERT[3]

BERT dapat disesuaikan atau *fine-tuning* untuk berbagai tugas spesifik seperti analisis sentimen atau klasifikasi teks tanpa modifikasi besar terhadap sistem atau model yang digunakan. Hal ini membuat *BERT* sangat fleksibel dan mudah untuk digunakan sekaligus memberikan hasil yang lebih akurat setelah proses Fine-Tuning[24].

BERT dilatih menggunakan dua jenis tugas, yaitu Masked Language Modeling (MLM) dan Next Sentence Prediction (NSP). Total fungsi kerugian selama proses pretraining merupakan penjumlahan dari kedua tugas tersebut, yang dirumuskan sebagai:

$$\mathscr{L}_{BERT} = \mathscr{L}_{MLM} + \mathscr{L}_{NSP}$$

Dengan meminimalkan nilai dari $\mathscr{L}_{BERT}$, model dapat secara simultan belajar memahami makna kata dalam konteks kalimat (MLM) dan hubungan antar kalimat (NSP).

2.5.1 Masked Language Model (*MLM*)

MLM adalah salah satu teknik pre-training yang digunakan oleh *BERT*. Dalam proses *BERT*, *MLM* akan memblokir secara acak (*masking*) beberapa kata dalam kalimat dan model diminta untuk memprediksi kata yang hilang berdasarkan konteks kiri dan kanan[24]. Teknik ini memungkinkan model untuk membangun representasi kontekstual yang mendalam dari setiap kata[24]. Contoh dari proses MLM dapat dilihat pada tabel 2.1.

Tabel 2.1. Contoh Proses MLM

No	Sebelum	Sesudah	Prediksi
1	Dia pergi ke perpustakaan untuk membaca buku.	Dia pergi ke **[MASK]** untuk membaca buku.	perpustakaan
2	Cuaca hari ini sangat cerah dan menyenangkan.	Cuaca hari ini sangat **[MASK]** dan menyenangkan.	cerah
3	Saya sangat senang bisa bertemu dengan teman-teman lama.	Saya sangat [**MASK**] bisa bertemu dengan teman-teman lama.	senang
4	Dia membeli sepatu baru untuk olahraga.	Dia membeli [**MASK**] baru untuk olahraga.	sepatu
5	Anjing itu berlari ke taman untuk bermain.	Anjing itu berlari ke [**MASK**] untuk bermain.	taman

Fungsi kerugian (loss) pada MLM didefinisikan sebagai:

$$\mathscr{L}_{MLM} = -\sum_{i \in M} \log P(w_i | \hat{w})$$

- M adalah himpunan indeks token yang di-mask
- w_i adalah token asli pada posisi ke-i
- $\hat{w}$ adalah token-token input dengan masking

Fungsi loss ini menghitung seberapa besar kesalahan prediksi model dalam menebak token yang dihilangkan. Semakin kecil nilai loss, semakin baik model dalam memahami konteks kalimat dan memprediksi kata yang hilang.

2.5.2 Next Sentence Prediction (*NSP*)

NSP adalah tugas pre-training kedua pada BERT yang dirancang untuk mempelajari hubungan antar-kalimat[24]. Dalam NSP, model diberi dua kalimat dan diminta menentukan apakah kalimat kedua adalah kelanjutan logis dari kalimat pertama (IsNext) atau tidak (NotNext)[24].

Tabel 2.2. Contoh Proses NSP

No	Sebelum	Sesudah	Label
1	Saya pergi ke pasar untuk membeli buah.	Setelah itu, saya pergi ke toko buku.	*IsNext*
2	Saya suka sekali menonton film action.	Di toko ini, mereka menjual berbagai macam peralatan dapur.	*NotNext*
3	Pagi ini saya berjalan ke taman untuk berolahraga.	Sesampainya di sana, saya bertemu dengan teman lama.	*IsNext*
4	Kucing saya suka tidur di atas sofa.	Namun, hari ini dia tidur di tempat tidur.	*IsNext*
5	Dia pergi ke toko untuk membeli pakaian baru.	Toko itu memiliki banyak pilihan sepatu.	*NotNext*

Fungsi loss NSP ditulis sebagai:

$$\mathscr{L}_{NSP} = -[y \log P(A|B) + (1-y) \log(1 - P(A|B))]$$

- A dan B adalah dua kalimat masukan
- $y = 1$ jika B merupakan kelanjutan dari A, dan 0 jika tidak

Loss NSP digunakan untuk melatih kemampuan model dalam memahami hubungan antar kalimat. Nilai loss yang rendah menunjukkan bahwa model mampu mengenali urutan kalimat dengan baik.

2.6 IndoNLU

IndoNLU Benchmark merupakan suatu kumpulan sumber daya untuk melatih, mengevaluasi, dan menganalisis sistem pemahaman *Natural Language Understanding* atau bahasa alami untuk bahasa Indonesia[17]. *IndoNLU* menangani dua belas tugas atau *tasks* seperti sentimen analisis, klasifikasi teks, *entailment*, dan fungsi-fungsi lainnya. *IndoNLU* menyediakan berbagai benchmark dan model dengan menggunakan *IndoBERT*[17]).

2.6.1 IndoBERT-base

IndoBERT-base merupakan Pre-trained BERT yang digunakan untuk *benchmark IndoNLU*[17]. IndoBERT merupakan model BERT yang khusus dilatih dalam bahasa Indonesia untuk memahami dan menganalisis teks berbahasa Indonesia. IndoBERT-base memiliki parameter sebesar 124,5M, 12 layer, 12 head, 768 embedding size, 768 hidden size, monolingualitas dalam bahasa Indonesia, dan tipe pre-train dengan metode kontekstual. Dataset yang digunakan untuk Pre-trained BERT mencapai 4 milyar kata dari teks bahasa Indonesia yang sudah mengikuti tahapan pre-processing (dengan ukuran sekitar 23 gigabyte data atau sekitar 124,5 juta parameter yang dihasilkan) sebagai dataset standar dari IndoBERT[17].

2.6.2 Model SmSA

SmSA merupakan salah satu model IndoNLU yang dapat digunakan untuk melakukan analisis sentimen dengan label positif, negatif, dan netral. SmSA merupakan salah satu downstream task yang menggunakan IndoBERT. Model ini menggunakan dataset dalam bentuk ulasan dan komentar dalam bahasa Indonesia yang telah dikumpulkan dari beberapa platform online[17].

2.7 Split Dataset

Splitting dataset merupakan proses pembagian dari total data menjadi tiga secara merata, yaitu training, valid, dan testing. Data training akan digunakan untuk melatih model dengan dataset yang sudah berlabel positif, negatif, dan netral menggunakan model fine-tuned terbaik yang telah dipilih dalam *skenario testing*. Data valid akan digunakan bersama dengan data training untuk mengevaluasikan

performa model secara berkala[26]. Data testing merupakan dataset utama yang akan dievaluasikan menggunakan model yang sudah terlatih dengan dataset training dan valid yang sudah dipersiapkan sebelumnya[26].

2.7.1 Data Training

Training set adalah bagian dari dataset yang digunakan untuk melatih model dengan menyesuaikan parameter internalnya berdasarkan data yang tersedia. Model menggunakan data dalam training set untuk belajar mengenali pola dan hubungan antara fitur serta label yang telah disediakan[26]. Selama proses pelatihan, model akan mengoptimalkan bobot dan parameter yang ada dengan tujuan untuk meminimalkan kesalahan, sehingga dapat menghasilkan prediksi yang lebih akurat[26]. *Data valid* akan membantu mengoptimisasi performa dari *data training* untuk memastikan tidak terjadinya *overfitting* terhadap dataset yang akan dianalisis[27].

Overfitting merupakan kesalahan pembelajaran yang terjadi pada suatu model machine learning ketika model terlalu menyesuaikan diri dengan data pelatihan (training dataset), sehingga performanya untuk menganalisis data baru menjadi berkurang[26]. Hal ini biasanya ditandai dengan perbedaan signifikan antara akurasi pada training dataset yang sangat tinggi dan akurasi pada testing dataset yang jauh lebih rendah[26].

Sebaliknya, underfitting merupakan kesalahan pembelajaran yang terjadi ketika model tidak mampu menangkap pola yang cukup dari data pelatihan, sehingga menghasilkan performa yang buruk untuk training dataset maupun testing dataset[26]. Hal ini biasanya disebabkan oleh model yang terlalu sederhana untuk data yang diberikan atau jumlah data pelatihan yang tidak mencukupi, sehingga model tidak bisa memahami struktur mendasar dari data yang digunakan[26].

2.7.2 Data Valid

Validation set adalah subset dari data yang digunakan untuk menyesuaikan hyperparameter model dan mengevaluasi performa model selama proses pelatihan tanpa mempengaruhi parameter utama (atau dataset yang akan di-testing) yang dipelajari[26]. Validation set membantu dalam proses scenario testing atau model selection, yaitu memilih konfigurasi model yang memberikan kinerja terbaik sebelum diterapkan pada data baru[26]. Jika model menunjukkan kinerja yang

sangat baik pada training set tetapi buruk pada validation set, kemungkinan model mengalami overfitting. Oleh karena itu, validation set digunakan untuk memantau dan mencegah terjadinya overfitting pada model. Dalam praktiknya, validation set sering kali diambil sekitar 10-20 persen dari total dataset, tergantung pada jumlah data yang tersedia[26]. *Data valid* akan membantu mengoptimisasi performa dari *data training* untuk memastikan tidak terjadinya *overfitting* terhadap dataset yang akan dianalisis[27].

2.7.3 Data Testing

Test set adalah bagian dari dataset yang digunakan untuk mengukur kinerja akhir model setelah proses pelatihan dan validasi selesai[26]. Data dalam test set tidak pernah digunakan dalam pelatihan atau pemilihan hyperparameter, sehingga memberikan estimasi objektif, atau penilaian kinerja model yang tidak dipengaruhi oleh bias dari data pelatihan atau parameter yang telah disesuaikan selama proses training, terhadap kemampuan model pada data baru yang belum pernah dilihat sebelumnya[26]. Test set sangat penting karena memastikan bahwa model tidak hanya menghafal pola dalam training set (overfitting) tetapi benar-benar mampu membuat prediksi yang akurat terhadap data dunia nyata[26].

2.8 Evaluasi Kinerja Model

Dalam tugas klasifikasi seperti analisis sentimen, evaluasi kinerja model sangat penting untuk memahami seberapa baik model mampu mengenali pola dalam data dan menghasilkan prediksi yang akurat. Beberapa metrik yang umum digunakan adalah sebagai berikut:

2.8.1 Confusion Matrix

Confusion Matrix adalah tabel yang digunakan untuk menggambarkan kinerja model klasifikasi dengan membandingkan antara label aktual dengan label prediksi. Matriks ini terdiri dari empat komponen utama:

- True Positive (TP): Prediksi positif dan benar.
- True Negative (TN): Prediksi negatif dan benar.
- False Positive (FP): Prediksi positif tapi salah.

- False Negative (FN): Prediksi negatif tapi salah.

2.8.2 Accuracy

Akurasi mengukur proporsi prediksi yang benar dari keseluruhan data. Dirumuskan sebagai:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

2.8.3 Precision

Precision mengukur seberapa banyak prediksi positif yang benar-benar relevan. Dirumuskan sebagai:

$$\text{Precision} = \frac{TP}{TP + FP}$$

2.8.4 Recall

Recall mengukur seberapa banyak kasus positif yang berhasil ditangkap oleh model. Dirumuskan sebagai:

$$\text{Recall} = \frac{TP}{TP + FN}$$

2.8.5 F1-Score

F1-Score adalah harmonic mean dari precision dan recall, berguna saat ada ketidakseimbangan antara jumlah data di setiap kelas. Rumusnya adalah:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Metrik-metrik ini secara teknis tetap dihitung oleh model untuk memberikan estimasi performa selama proses pelatihan dan pengujian. Perlu dicatat bahwa dalam penelitian ini tidak terdapat proses pelabelan manual oleh annotator manusia. Oleh karena itu, metrik evaluasi seperti akurasi, precision, recall, F1-score, dan confusion matrix hanya digunakan sebagai estimasi performa internal yang dihasilkan oleh model selama proses fine-tuning dan pengujian otomatis.

2.9 Penelitian Terkait

Beberapa penelitian sebelumnya telah dilakukan untuk menganalisis opini publik terhadap kecerdasan buatan menggunakan pendekatan analisis sentimen. Penelitian oleh Beneng dan Zubair (2024) [28] membahas sentimen terhadap *AI-generated artwork* di Twitter menggunakan algoritma *Naive Bayes* dan *Neural Network*. Meskipun topiknya serupa, pendekatan yang digunakan berbeda dengan penelitian ini yang memanfaatkan *IndoBERT*, sebuah model berbasis *Transformer* yang telah dilatih khusus untuk bahasa Indonesia.

Di sisi lain, studi seperti yang dilakukan oleh Wijanarko et al. (2024) [29] dan Syafira (2023) [30] lebih berfokus pada analisis sentimen terhadap AI secara umum, tanpa membahas aspek seni digital yang dihasilkan oleh kecerdasan buatan.

Hingga saat ini, belum ditemukan penelitian sebelumnya dalam konteks Indonesia yang secara spesifik menganalisis *AI-generated artwork* menggunakan pendekatan *IndoBERT*. Oleh karena itu, penelitian ini diharapkan dapat memberikan kontribusi baru dengan mengisi kekosongan pada ranah penelitian sentimen terhadap seni digital berbasis AI di Indonesia, sekaligus memperluas pemanfaatan model *IndoBERT* dalam domain sosial dan budaya.