

BAB II

TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1 Tinjauan Penelitian

2.1.1 Tinjauan Penelitian I

Penelitian ini merupakan penelitian yang membahas tentang pembelajaran secara daring atau *online*. Penelitian pertama yaitu penelitian yang dilakukan oleh Muhammad Nur Hidayat dan Rully Pramudita, Teknik Informatika dari Universitas Bina Insani dengan judul "Analisis Sentimen Terhadap Pembelajaran Secara Daring Pasca Pandemi Covid-19 Menggunakan Metode IndoBERT" tahun 2023. Berdasarkan hasil penelitian dilakukan terhadap 35 data, 10 data mempunyai sentimen positif, 23 data mempunyai sentimen negatif dan 2 data mempunyai sentimen netral. Dan sistem klasifikasi opini dengan metode IndoBERT dapat membantu untuk mengklasifikasikan opini secara akurat dan membaginya menjadi 3 kategori (positif, netral dan negatif) secara otomatis dengan nilai akurasi sebesar 0,87 atau 87%[3].

2.1.2 Tinjauan Penelitian II

Penelitian ini merupakan penelitian yang membahas tentang persepsi konsumen di aplikasi transportasi yaitu Gojek. Penelitian kedua yaitu penelitian yang dilakukan oleh Nofita Mahfudiyah dan Andry Alamsyah, Manajemen Bisnis Telekomunikasi & Informatika, Fakultas Ekonomi Bisnis di Universitas Telkom dengan judul "Analisis Persepsi Konsumen Terhadap Kualitas Layanan Gojek Menggunakan *Sentiment Analysis* Dan *Topic Modeling* Berdasarkan *Deep Learning* IndoBERT" tahun 2022. Hasil dari penelitian yang sudah dilakukan, terdapat tiga kesimpulan yang diperoleh. Kesimpulan pertama terkait pemodelan IndoBERT. Hasil pemodelan menggunakan IndoBERT memiliki akurasi 96% dengan rasio dataset 80:10:10. Akurasi *precision* dan *recall* masing-masing sebesar 93% dan 95%. IndoBERT mampu mengklasifikasikan dan memprediksi dengan baik karena semua akurasi di atas 90%. Sentimen

negatif mendominasi 56%, sedangkan sentimen positif 38% dan sentimen netral 6%. Konsumen Gojek cenderung memiliki persepsi yang kurang baik. *Topic modelling* menghasilkan *cluster* yang baik berdasarkan dimensi kualitas layanan. Frekuensi tertinggi pada sentimen positif adalah *mobile convenience* dan *reliability*, sedangkan pada sentimen negatif adalah *mobile customer service* dan *billing*. Sentimen netral memiliki frekuensi tertinggi pada *mobile security* dan *privacy* [4].

2.1.3 Tinjauan Penelitian III

Penelitian ini merupakan penelitian yang membahas tentang pendapat masyarakat terhadap konten *Deepfake*. Penelitian ketiga yaitu penelitian yang dilakukan oleh Shane Giorgio Alexander, Amadeus Terra Ananto, I Putu Adhitya Pratatama Mangku Purnama, Bayu Liano Leader Habibullah, Nur Aini Rakhmawati. Fakultas Ilmu Komputer di Insitut Teknologi Sepuluh Nopember dengan judul "Analisis Sentimen Opini Masyarakat Indonesia Terhadap Konten *Deepfake* Tokoh Publik" tahun 2023. Berdasarkan hasil penelitian tersebut menginvestigasi pandangan masyarakat terhadap konten *deepfake* yang melibatkan tokoh publik dan mengukur besarnya dampak yang ditimbulkan akibat dari penyebaran konten *deepfake* di media sosial terkhususnya Youtube. Penelitian ini menggunakan proses klasifikasi sentimen untuk mengevaluasi opini masyarakat Indonesia tentang konten *deepfake* di YouTube. Dataset terdiri dari 300 data yang dibagi menjadi data *training* dan data *testing*. Klasifikasi sentimen dilakukan dengan IndoBERT dan metode *semi-supervised learning*. Metode Naive Bayes Classifier juga digunakan untuk hasil prediksi sentimen. Hasilnya menunjukkan bahwa model memiliki akurasi 82%, *mean F1* 54.3%, *recall* 51.34%, dan *precision* 61%. Kata-kata seperti "teknologi," "oknum," "bohong," dan "kriminal" mencerminkan sentimen negatif, sementara kata seperti "iya," "sini," "bedain," dan "sulit" mencerminkan sentimen positif. Mayoritas masyarakat Indonesia cenderung netral dalam menyampaikan pendapat tentang penggunaan *deepfake* dengan tokoh publik.[5].

2.2 Landasan Teori

2.2.1 Analisis Sentimen

Analisis sentimen menggunakan analisis teks untuk memperoleh berbagai sumber data dari internet dan berbagai platform media sosial. Tujuannya adalah untuk mendapatkan opini dari pengguna platform. Analisis sentimen adalah bidang pemrosesan bahasa alami (NLP) yang membangun sistem untuk mengenali dan mengekstrak opini dalam format teks. Informasi tekstual kini banyak tersedia di Internet dalam bentuk forum, blog, media sosial, dan situs ulasan. Dengan bantuan analisis sentimen, informasi yang sebelumnya tidak terstruktur dapat diubah menjadi data yang lebih terstruktur[6].

2.2.2 Persepsi Dan Opini

Secara singkat, persepsi adalah cara seseorang memahami atau menginterpretasikan informasi, opini adalah pandangan atau keyakinan seseorang tentang suatu topik atau masalah[7]. Persepsi yaitu anggapan berdasarkan pengalaman atau pengetahuan yang kadang berbeda antara satu orang dengan orang lain atau kadang berbeda dengan kondisi yang sebenarnya. Opini adalah suatu pemikiran atau gagasan yang menjelaskan suatu kebutuhan atau keinginan terhadap pemikiran dan gagasan yang tidak bermakna.

2.2.3 Analisis Media Sosial

Analisis media sosial melibatkan pengumpulan data berharga dari saluran media sosial yang Anda gunakan dan kemudian mengidentifikasi dan mengubahnya menjadi kesimpulan yang dapat ditindaklanjuti. Platform media sosial sangat bermanfaat untuk mengidentifikasi ketertarikan audiens, mempelajari kompetitor dan menjadi yang terbaik di industri[8].

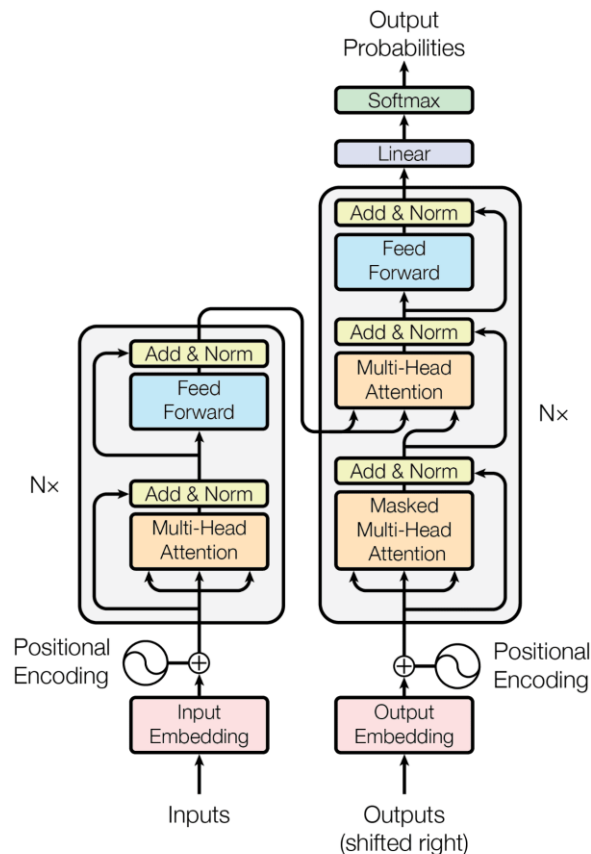
2.2.4 Skizofrenia

Skizofrenia adalah gangguan mental serius dimana pasien mengalami kesulitan membedakan antara khayalan dan kenyataan. Karena kondisi ini umumnya ditandai dengan perilaku abnormal seperti delusi dan halusinasi, pasien sering dianggap “gila”. Skizofrenia merupakan masalah kesehatan jangka panjang yang memerlukan pengobatan berkelanjutan. Artinya, penderita

skizofrenia harus menerima pengobatan seumur hidup untuk mengendalikan gejala, mencegah komplikasi, dan membantu mereka berfungsi dalam kehidupan sehari-hari[9].

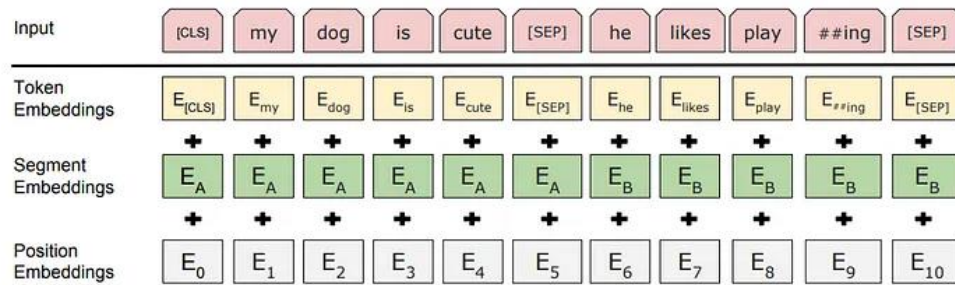
2.2.5 *Bidirectional Encoder Representations from Transformers* (BERT)

Bidirectional Encoder Representations from Transformers (BERT) adalah model representasi bahasa terlatih yang pertama kali dipopulerkan oleh Google pada 11 Oktober 2018. Teknik pra-pelatihan berbasis jaringan saraf untuk pemrosesan bahasa alami (*Natural Language Processing*) yang dikenal dengan nama *Bidirectional Encoder Representations from Transformers* (BERT). Dengan cara ini model BERT dapat *fine-tuning* dengan tambahan satu layer saja. Dilatih secara tak terawasi pada data teks besar untuk mempelajari representasi bahasa umum. Ini dilakukan dengan memprediksi kata-kata yang hilang (*masked words*) dalam kalimat dan memprediksi hubungan antara dua kalimat yang berdekatan.



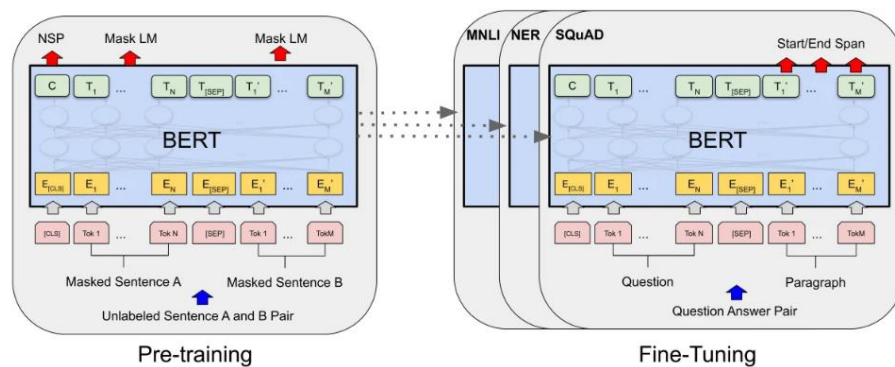
Gambar 2. 1 Arsitektur Transformers

Arsitektur *Transformer* merupakan arsitektur yang revolusioner dalam NLP. Kemampuannya dalam menangani tugas urutan-ke-urutan telah membuka berbagai peluang baru dalam berbagai bidang. *Transformer* menggunakan mekanisme *self-attention* untuk memahami representasi input dan output. Transformer memiliki 2 mekanisme terpisah, yaitu *encoder* untuk membaca teks input dan *decoder* yang menentukan prediksi. Tiap *layer encoder* memiliki dua *sub layer*, yang pertama adalah *multi-head self-attention mechanism*, dan yang kedua *fully connected feed forward network*. Pada *encoder*, setiap input akan melewati *self-attention layer*. Layer ini membantu *encoder* untuk melihat kata lain yang ada pada kalimat. Output dari *attention layer* kemudian digunakan sebagai input ke *feed-forward neural network*[10].



Gambar 2. 2 Representasi Input BERT

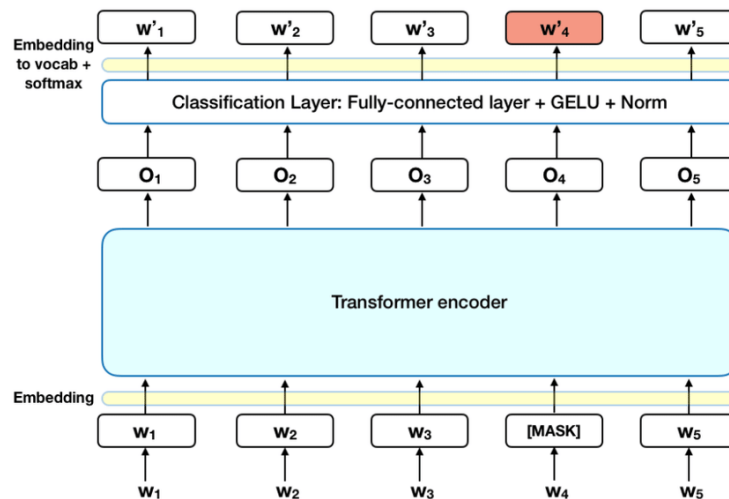
Awalnya, kata dalam NLP direpresentasikan sebagai *one-hot vector*, yaitu vektor dengan banyaknya elemen sama dengan jumlah kata dalam kosakata, dimana hanya elemen yang 对应 (*duì yì - corresponding*) dengan kata tersebut bernilai 1 dan sisanya bernilai 0. *Word embedding* kemudian memperkenalkan representasi vektor yang lebih padat dan bermakna, dimana kata-kata yang memiliki makna serupa akan memiliki representasi vektor yang saling dekat. Teknik ini membantu model NLP memahami hubungan antar kata dan meningkatkan performa[11].



Gambar 2. 3 Ilustrasi Tahapan Pre-training dan Fine-tuning

Framework BERT terdiri dari dua tahap utama: tahap pra-latih (*pre-training*) dan tahap *fine-tuning* (penyesuaian halus). Tahap *pre-training* merupakan tahap pada saat model mempelajari bahasa dan konteksnya dengan melakukan *Mask Language Modelling* (MLM) dan *Next Sentence Prediction* (NSP) secara bersamaan. MLM memungkinkan representasi untuk memadukan konteks kiri dan konteks kanan, yang memungkinkan kita untuk melatih

transformator dua arah yang dalam dan NSP yang menggabungkan representasi pasangan *pre-train* teks. Tahap *pre-training* dan *fine-tuning* merupakan kerja sama yang membuat BERT begitu kuat. Pra-latih memberikan representasi input yang kaya akan makna dan konteks, sedangkan *fine-tuning* menyesuaikan model untuk tugas NLP spesifik yang diinginkan[12].



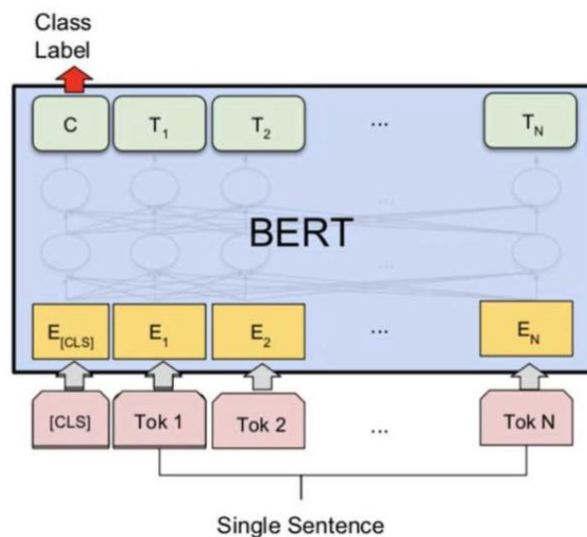
Gambar 2. 4 Proses Masked Language Modelling (MLM)

Masked Language Modeling (MLM) adalah teknik pelatihan untuk model bahasa di mana beberapa token dalam teks input diacak dan digantikan dengan token [MASK]. Model kemudian diajarkan untuk memprediksi nilai token yang digantikan berdasarkan konteks sekitarnya. Tujuannya adalah untuk membantu model memahami hubungan antara kata-kata dalam sebuah kalimat dengan mempertimbangkan konteks yang ada[13].

Next Sentence Prediction (NSP) adalah teknik pelatihan untuk model bahasa yang bertujuan untuk membantu model memahami hubungan antara dua kalimat berurutan dalam teks. Dengan NSP, model belajar untuk memahami hubungan antara kalimat-kalimat dalam teks, yang dapat membantu dalam pemahaman konteks yang lebih luas daripada hanya memahami kalimat secara individu[14].

2.2.6 Fine-tuning Bidirectional Encoder Representations from Transformers (BERT)

Pertama, model BERT dilatih pada dataset besar menggunakan teknik pelatihan MLM (*Masked Language Modeling*) dan NSP (*Next Sentence Prediction*). Proses ini membantu BERT memahami representasi kata-kata dan kalimat dalam konteks yang lebih luas.



Gambar 2. 5 Ilustrasi fine-tuning BERT untuk Klasifikasi

Setelah BERT dilatih secara umum, langkah selanjutnya adalah *fine-tuning*. Pada tahap ini, model BERT disesuaikan dengan tugas spesifik dengan memperbarui parameter-parameter model pada dataset tugas tersebut. Misalnya, jika tujuannya adalah untuk klasifikasi teks, seperti sentimen analisis atau klasifikasi topik, BERT akan disesuaikan dengan dataset yang sesuai dan kemudian dilatih kembali pada tugas tersebut. Setelah proses *fine-tuning* selesai, model yang dihasilkan dievaluasi pada dataset uji untuk mengukur kinerjanya dalam menyelesaikan tugas spesifik tersebut. *Fine-tuning* BERT memungkinkan model untuk belajar representasi yang lebih khusus untuk tugas-tugas tertentu, sehingga meningkatkan kinerja model dalam menyelesaikan tugas-tugas tersebut[15].

2.2.7 Confusion Matrix

Sebuah tabel yang digunakan dalam *machine learning* untuk mengevaluasi performa dari suatu model klasifikasi. *Confusion matrix* menggambarkan hasil prediksi dari model pada data uji, dibandingkan dengan nilai sebenarnya dari data tersebut.

| | | PREDICTED classification | | | | |
|-----------------------|---|--------------------------|----|----|----|---|
| | | Classes | a | b | c | d |
| ACTUAL classification | a | TN | FP | TN | TN | |
| | b | FN | TP | FN | FN | |
| | c | TN | FP | TN | TN | |
| | d | TN | FP | TN | TN | |

Gambar 2. 6 Confusion Matrix

True Positives (TP): Kasus di mana model memprediksi kelas positif dengan benar, dan kelas aktualnya juga positif. *True Negatives* (TN): Kasus di mana model memprediksi kelas negatif dengan benar, dan kelas aktualnya juga negatif. *False Positives* (FP): Kasus di mana model memprediksi kelas positif, tetapi kelas aktualnya negatif. Juga dikenal sebagai *Type I error*. *False Negatives* (FN): Kasus di mana model memprediksi kelas negatif, tetapi kelas aktualnya positif. Juga dikenal sebagai *Type II error*.

Dengan menggunakan nilai-nilai ini, kita dapat menghitung berbagai metrik evaluasi, seperti akurasi, presisi, *recall* (sensitivitas), spesifisitas, dan *F1-score*, yang membantu kita memahami performa keseluruhan dari model klasifikasi. *Confusion matrix* sangat penting karena memberikan pandangan yang lebih rinci tentang bagaimana model kita melakukan prediksi,

khususnya dalam mengidentifikasi jenis kesalahan yang dilakukan oleh model tersebut[16].

2.2.8 IndoBERT

IndoBERT adalah kepanjangan dari Indonesia *Bidirectional Encoder Representations from Transformers*. Ini adalah model bahasa natural terbaru yang dikembangkan oleh komunitas pengembang AI Indonesia, menggunakan teknologi NLP (*Natural Language Processing*). IndoBERT menggunakan pendekatan *deep learning* dimana model belajar secara mandiri dari data teks yang disediakan. Model ini dapat mempelajari hubungan antara kata dan frasa dalam sebuah kalimat, sehingga menghasilkan terjemahan yang lebih akurat dan terlihat alami. IndoBERT adalah model pemrosesan bahasa alami (NLP) yang dibangun menggunakan kerangka pembelajaran mendalam PyTorch. PyTorch adalah kerangka pembelajaran mendalam populer yang banyak digunakan oleh para peneliti untuk membangun model pembelajaran mendalam di berbagai bidang termasuk NLP[17].

2.2.9 X

Twitter atau X merupakan media sosial yang sangat populer di kalangan pengguna internet. Platform ini memungkinkan pengguna untuk berbagi *tweet* dan *tweet* dalam bentuk teks pendek, gambar, dan video. Seperti media sosial lainnya, X memiliki beragam istilah yang berkaitan dengan fitur dan karakter penggunaannya[18].