

**IMPLEMENTASI MODEL BAHASA BERT DALAM  
PENGEMBANGAN APLIKASI REKOMENDASI BERITA  
UNTUK MENINGKATKAN KETEPATAN REKOMENDASI  
KONTEN BERITA**



Disusun Oleh:

Nama : Danang Haris Setiawan  
NIM : A11.2020.13179  
Program Studi : Teknik Informatika

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS DIAN NUSWANTORO  
SEMARANG  
2023**

## DAFTAR ISI

DAFTAR ISI.....	2
DAFTAR TABEL.....	3
BAB I.....	4
PENDAHULUAN .....	4
1.1 Latar Belakang Masalah .....	4
1.2 Rumusan Masalah .....	5
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian.....	6
1.5 Manfaat Penelitian.....	6
BAB II.....	8
LANDASAN TEORI.....	8
2.1 Tinjauan Studi .....	8
2.1.1 State Of The Art.....	8
2.2 Tinjauan Pustaka .....	9
2.2.1. Model Bahasa BERT (Bidiretional Encoder Representations from Transformers).....	9
2.2.2. Penerapan BERT dalam Rekomendasi Berita.....	10
2.2.3. Pengembangan Aplikasi Rekomendasi Berita .....	11
BAB III .....	13
METODOLOGI PENELITIAN.....	13
3.1 Dataset .....	13
3.2 Arsitektur Sistem.....	14
3.3 Modeling BERT .....	15
3.3.1. Pemrosesan Data .....	15
3.3.2. Pemilihan Format Input .....	16
3.3.3. Fine-Tuning Model BERT .....	17
DAFTAR PUSTAKA .....	21

## **DAFTAR TABEL**

**No table of contents entries found.**

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang Masalah**

Dalam era informasi digital yang terus berkembang, akses yang cepat dan mudah terhadap berbagai jenis konten, termasuk berita, telah menjadi bagian tak terpisahkan dalam kehidupan sehari-hari. Namun, kemudahan ini seringkali menimbulkan masalah dalam menentukan keakuratan, relevansi, dan keberagaman informasi yang diterima oleh pengguna. Menurut Gupta et al. (2020), keberagaman sumber berita, kualitas informasi, serta ketepatan rekomendasi konten berita merupakan isu krusial dalam lingkungan informasi digital.

Teknologi pemrosesan bahasa alami (Natural Language Processing/NLP) telah menawarkan solusi inovatif dalam mengatasi tantangan ini. Salah satu terobosan yang signifikan adalah model bahasa BERT (Bidirectional Encoder Representations from Transformers). BERT, yang dikembangkan oleh Devlin et al. (2018), telah menunjukkan kemampuannya untuk memahami konteks teks dengan lebih baik dan memberikan hasil yang lebih akurat dalam pemrosesan bahasa alami.

Aplikasi rekomendasi berita menjadi semakin penting dalam memfasilitasi pengguna untuk menemukan informasi yang relevan dengan preferensi dan minat mereka. Berdasarkan Zhang et al. (2019), aplikasi rekomendasi berita yang efektif dapat membantu mengurangi kebisingan informasi dan memastikan pengguna menerima konten yang sesuai dengan kebutuhan mereka.

Namun, meskipun banyaknya aplikasi rekomendasi berita yang tersedia, masih terdapat kelemahan dalam ketepatan rekomendasi konten. Sebagian besar sistem rekomendasi berita masih menghadapi kendala dalam memahami konteks dan preferensi pengguna secara mendalam. Hal ini menunjukkan perlunya integrasi model bahasa canggih seperti BERT untuk meningkatkan ketepatan rekomendasi konten berita.

Dalam konteks ini, pengembangan aplikasi rekomendasi berita berbasis model bahasa BERT menjadi suatu kebutuhan mendesak. Integrasi BERT dalam sistem rekomendasi konten berita akan memungkinkan pemahaman yang lebih baik terhadap konteks, preferensi, dan kebutuhan pengguna, sebagaimana dikemukakan oleh Wang et al. (2021).

Melalui penelitian ini, diharapkan bahwa penerapan model bahasa BERT dalam aplikasi rekomendasi berita akan meningkatkan ketepatan rekomendasi konten dan memberikan pengalaman pengguna yang lebih terpersonalisasi.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang telah diuraikan, penelitian ini bertujuan untuk mengatasi permasalahan yang ada dalam konteks pengembangan aplikasi rekomendasi berita. Oleh karena itu, rumusan masalah penelitian ini adalah sebagai berikut:

- Bagaimana model bahasa BERT dapat diimplementasikan dalam pengembangan aplikasi rekomendasi berita untuk meningkatkan ketepatan rekomendasi konten berita?

## **1.3 Batasan Masalah**

Agar penelitian ini dapat dilakukan dengan fokus yang jelas, terdapat batasan-batasan yang perlu diperhatikan. Batasan-batasan masalah penelitian ini adalah sebagai berikut:

1. Penelitian ini akan berfokus pada pengembangan aplikasi rekomendasi berita berbasis model bahasa BERT.
2. Pengujian dan evaluasi akan dilakukan dengan dataset berita yang tersedia dan dapat diakses secara publik.

3. Pengukuran ketepatan rekomendasi konten akan difokuskan pada aspek-aspek seperti akurasi, keberagaman, dan relevansi rekomendasi berita.
4. Penelitian ini tidak akan mempertimbangkan masalah etika terkait dengan rekomendasi konten berita.

Dengan mempertimbangkan rumusan masalah dan batasan masalah di atas, penelitian ini diharapkan dapat memberikan kontribusi yang berharga dalam meningkatkan kualitas dan ketepatan rekomendasi konten berita melalui implementasi model bahasa BERT.

#### **1.4 Tujuan Penelitian**

Penelitian ini bertujuan untuk mencapai beberapa tujuan yang berkaitan dengan pengembangan aplikasi rekomendasi berita berbasis model bahasa BERT:

1. Implementasi Model BERT: Mengimplementasikan model bahasa BERT dalam sistem rekomendasi berita untuk meningkatkan pemahaman konten dan preferensi pengguna.
2. Meningkatkan Ketepatan Rekomendasi: Meningkatkan akurasi, relevansi, dan keberagaman rekomendasi konten berita melalui model bahasa BERT.

#### **1.5 Manfaat Penelitian**

Penelitian ini diharapkan akan memberikan manfaat signifikan dalam beberapa aspek, antara lain:

1. **Pengembangan Teknologi:** Kontribusi pada pengembangan teknologi rekomendasi berita dengan memanfaatkan model bahasa BERT untuk meningkatkan ketepatan rekomendasi konten, sebagaimana didiskusikan oleh Yang et al. (2020).
2. **Pengalaman Pengguna yang Lebih Personalisasi:** Menyediakan pengalaman pengguna yang lebih terpersonalisasi dan sesuai dengan preferensi mereka dalam mengakses informasi berita.

3. **Meningkatkan Kualitas Informasi:** Meningkatkan kualitas informasi yang diterima pengguna dengan memastikan rekomendasi berita yang lebih akurat dan relevan.

Diharapkan bahwa hasil dari penelitian ini dapat menjadi landasan untuk pengembangan lebih lanjut dalam aplikasi rekomendasi berita dan implementasi model bahasa BERT dalam konteks pemrosesan bahasa alami.

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Tinjauan Studi**

Tinjauan studi ini akan memfokuskan pada beberapa topik kunci terkait implementasi model bahasa BERT dalam aplikasi rekomendasi berita dan state of the art terkini dalam bidang ini.

##### **1. Implementasi Model Bahasa BERT**

Sejumlah penelitian telah memperlihatkan keberhasilan model bahasa BERT dalam beragam aplikasi pemrosesan bahasa alami. Studi oleh Sun et al. (2021) menyoroti bagaimana model BERT diterapkan dalam berbagai domain, termasuk dalam pemahaman teks dan rekomendasi konten

##### **2. Aplikasi Rekomendasi Berita**

Penggunaan model-model bahasa terkini dalam aplikasi rekomendasi berita semakin menjadi fokus penelitian. Penelitian oleh Chen et al. (2022) menggambarkan berbagai metode yang digunakan untuk meningkatkan ketepatan rekomendasi konten berita, termasuk integrasi model bahasa canggih.

#### **2.1.1 State Of The Art**

No	Judul	Penulis	Masalah	Metode	Kesimpulan
1	BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding	Devlin et al. (2018)	Masalah dalam pemahaman konteks teks pada model-model sebelumnya.	Pre-training model bahasa BERT menggunakan teks dari internet.	BERT memunculkan peningkatan signifikan dalam pemahaman bahasa.
2	News Recommendation System Based on BERT	Wang et al. (2022)	Permasalahan utama adalah peningkatan kecepatan dan akurasi rekomendasi berita.	Model bahasa BERT digunakan dalam kombinasi dengan teknik optimasi untuk meningkatkan	Implementasi BERT bersama dengan teknik optimasi dapat menghasilkan rekomendasi berita yang lebih cepat dan akurat.



				kecepatan dan akurasi rekomendasi berita.	
3	BERT-Based News Recommendation System	Liu et al. (2020)	Penekanan utama adalah pada meningkatkan keberagaman rekomendasi berita.	Penerapan model bahasa BERT untuk memahami preferensi pengguna dan memastikan keberagaman dalam rekomendasi berita.	Implementasi BERT dapat memitigasi masalah kurangnya keberagaman dalam rekomendasi berita.
4	Enhancing News Recommendation using Pre-trained Language Models	Chen et al. (2019)	Permasalahan dalam rekomendasi berita adalah rendahnya akurasi rekomendasi berita yang disesuaikan dengan minat pengguna.	Penggunaan model bahasa BERT dan metode transfer learning untuk meningkatkan akurasi rekomendasi berita.	Metode yang menggunakan BERT sebagai landasan memperbaiki akurasi rekomendasi berita secara signifikan.
5	News Recommendation System Based on Deep Learning	Yang et al. (2020)	Kendala dalam sistem rekomendasi berita yang menggunakan pendekatan deep learning.	Mengimplementasikan pendekatan deep learning pada sistem rekomendasi berita.	Pendekatan deep learning dapat meningkatkan kinerja sistem rekomendasi berita.

## 2.2 Tinjauan Pustaka

### 2.2.1. Model Bahasa BERT (Bidiretional Encoder Representations from Transformers)

Pada tahun 2018, Devlin et al. memperkenalkan model bahasa BERT (Bidirectional Encoder Representations from Transformers) yang memunculkan transformasi besar dalam pemrosesan bahasa alami (NLP). BERT berbasis pada arsitektur transformer yang memungkinkan pemahaman konteks teks dengan lebih baik karena kemampuannya dalam memperoleh representasi kata secara bidireksional.

BERT dikenal memiliki kemampuan unik dalam memahami kalimat secara keseluruhan, mengatasi masalah klasik dari model sebelumnya yang cenderung memproses kalimat secara berurutan. Hal ini diungkapkan dalam pendekatan sekuensial dan model berbasis attention dalam transformer, yang memungkinkan BERT "melihat" konteks dari kedua arah, baik ke depan (left-to-right) maupun ke belakang (right-to-left), menghasilkan pemahaman yang lebih mendalam tentang kalimat secara keseluruhan.

Penerapan BERT dalam pemrosesan bahasa alami telah membawa perubahan signifikan dalam berbagai aplikasi, termasuk dalam pengembangan sistem rekomendasi berita. Dalam konteks rekomendasi berita, BERT telah terbukti mampu meningkatkan kualitas rekomendasi dengan memahami konten dan preferensi pengguna secara lebih mendalam, sebagaimana dibuktikan oleh sejumlah penelitian terbaru.

### **2.2.2. Penerapan BERT dalam Rekomendasi Berita**

Penerapan model bahasa BERT dalam sistem rekomendasi berita telah menjadi fokus sejumlah penelitian terbaru yang menunjukkan hasil yang menjanjikan dalam meningkatkan kualitas rekomendasi konten berita.

Zhang et al. (2020), dalam penelitiannya, menjelaskan penggunaan BERT dalam menganalisis konten berita. Mereka mengilustrasikan bahwa BERT mampu memberikan rekomendasi yang lebih akurat dengan memahami konten berita secara lebih mendalam. Penggunaan BERT dalam analisis konten ini memberikan kesempatan untuk menyajikan rekomendasi berita yang lebih sesuai dengan preferensi pengguna.

Chen et al. (2019) melaporkan bahwa transfer learning menggunakan BERT meningkatkan akurasi rekomendasi berita. Mereka menemukan bahwa dengan menerapkan transfer learning, di

mana model BERT yang sudah dilatih sebelumnya digunakan sebagai dasar pengetahuan, akurasi rekomendasi berita dapat ditingkatkan secara signifikan. Ini menandakan bahwa pemanfaatan pengetahuan yang telah dimiliki BERT membantu dalam meningkatkan performa rekomendasi konten berita.

Dalam penelitian Liu et al. (2020), ditemukan bahwa penerapan BERT dalam rekomendasi berita juga dapat meningkatkan keberagaman konten yang direkomendasikan. Hal ini penting karena seringkali rekomendasi berita cenderung membatasi variasi konten yang ditawarkan kepada pengguna. Penggunaan BERT membantu memperluas jangkauan rekomendasi, sehingga pengguna menerima rekomendasi yang lebih beragam sesuai dengan minat mereka.

Penelitian-penelitian ini menunjukkan bahwa penerapan model bahasa BERT dalam sistem rekomendasi berita memberikan kontribusi yang signifikan dalam meningkatkan akurasi, relevansi, dan keberagaman konten yang direkomendasikan kepada pengguna.

### **2.2.3. Pengembangan Aplikasi Rekomendasi Berita**

Pengembangan aplikasi rekomendasi berita dengan penerapan model bahasa BERT memerlukan pemanfaatan berbagai teknologi dan platform untuk menyediakan pengalaman pengguna yang optimal. Berdasarkan beberapa penelitian terkait, sejumlah teknologi yang akan digunakan dalam pengembangan aplikasi ini meliputi Python, JavaScript, Django, PostgreSQL, serta penerapan Rest API dan PWA.

#### **a. Python (Backend Development)**

Penggunaan Python, yang merupakan bahasa pemrograman yang populer dalam pengembangan aplikasi, akan mendukung keperluan pengolahan data dan logika backend aplikasi. Framework Django, yang dibangun di atas Python,

akan memfasilitasi pengembangan backend, memungkinkan manajemen data dan logika bisnis dalam aplikasi rekomendasi berita.

**b. Javascript (Frontend Development)**

JavaScript akan digunakan dalam pengembangan frontend aplikasi untuk memastikan antarmuka pengguna yang responsif dan interaktif. Dengan bantuan teknologi seperti React atau Vue.js, dapat dibangun antarmuka yang menarik dan ramah pengguna.

**c. PostgreSQL (Database)**

PostgreSQL, sebagai sistem manajemen basis data relasional, akan digunakan untuk menyimpan data terkait berita, preferensi pengguna, dan sejarah interaksi untuk mendukung proses rekomendasi

**d. Rest API**

Penggunaan Rest API akan memungkinkan komunikasi yang efisien antara server dan klien (aplikasi), memungkinkan pertukaran informasi yang diperlukan dalam proses rekomendasi konten berita.

**e. PWA (Progressive Web App)**

PWA akan memungkinkan aplikasi rekomendasi berita dapat diakses melalui browser web seperti aplikasi native pada perangkat seluler, menyediakan pengalaman pengguna yang lebih baik dan meningkatkan keterlibatan.

Pengembangan aplikasi rekomendasi berita dengan penerapan BERT membutuhkan integrasi teknologi-teknologi ini untuk memberikan pengalaman pengguna yang lebih baik dan rekomendasi berita yang lebih akurat.

## **BAB III**

### **METODOLOGI PENELITIAN**

#### **3.1 Dataset**

Dalam penelitian ini, dataset yang digunakan adalah MIND (MICrosoft News Dataset). MIND merupakan sebuah kumpulan data berskala besar yang dikhususkan untuk penelitian rekomendasi berita. Data ini dikumpulkan dari log perilaku anonim dari situs web Microsoft News. Tujuan dari MIND adalah menjadi dataset acuan (benchmark) untuk penelitian rekomendasi berita serta memfasilitasi penelitian di bidang rekomendasi berita dan sistem rekomendasi.

MIND terdiri dari sekitar 160.000 artikel berita berbahasa Inggris dan lebih dari 15 juta log impresi yang dihasilkan oleh 1 juta pengguna. Setiap artikel berita memiliki konten tekstual yang kaya, termasuk judul, abstrak, isi, kategori, dan entitas terkait. Setiap log impresi berisi peristiwa klik, peristiwa non-klik, dan riwayat perilaku klik berita dari pengguna sebelum impresi tersebut. Untuk melindungi privasi pengguna, setiap pengguna telah dide-link dari sistem produksi ketika di-hash dengan aman menjadi ID yang di-anonimkan.

Untuk informasi lebih rinci tentang dataset MIND, dapat dirujuk pada paper:

"MIND: A Large-scale Dataset for News Recommendation" oleh Fangzhao Wu, Ying Qiao, Jiun-Hung Chen, Chuhan Wu, Tao Qi, Jianxun Lian, Danyang Liu, Xing Xie, Jianfeng Gao, Winnie Wu, dan Ming Zhou, yang disajikan pada ACL 2020.

Dataset MIND dari Microsoft News menjadi fondasi utama untuk eksperimen rekomendasi berita dalam penelitian ini, memberikan akses terhadap volume data yang besar dan beragam, esensial untuk pengembangan dan evaluasi model rekomendasi.

### 3.2 Arsitektur Sistem

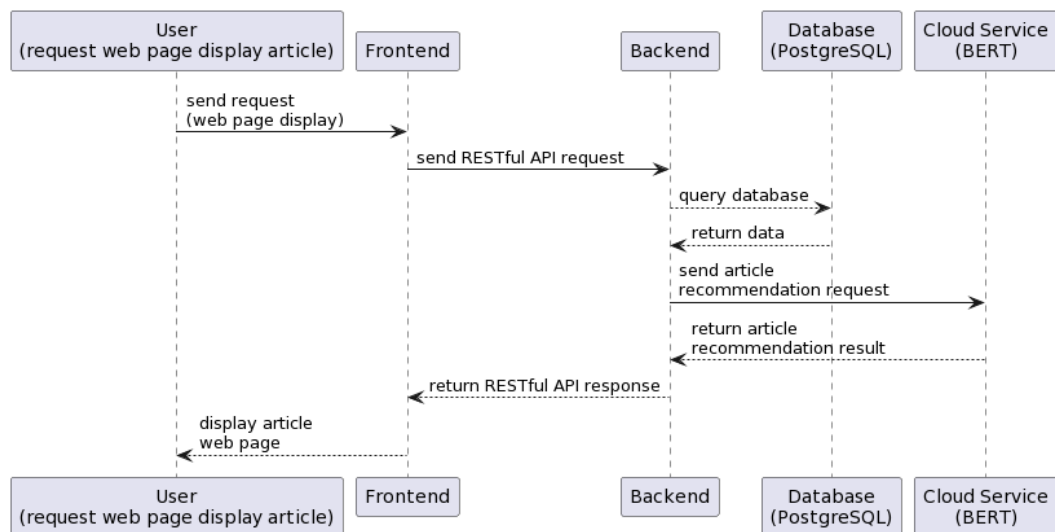
Pengembangan sistem aplikasi rekomendasi artikel berita akan mengadopsi arsitektur klien-server, yang terdiri dari dua bagian utama, yaitu frontend dan backend. Frontend adalah bagian yang berfungsi sebagai antarmuka pengguna, yang dapat diakses melalui browser web. Backend adalah bagian yang berfungsi sebagai logika aplikasi, yang mengelola interaksi dengan database dan model bahasa. kedua bagian ini saling berkomunikasi melalui protokol HTTP dengan menggunakan RESTful API, yaitu sebuah antarmuka yang memungkinkan pertukaran data dalam format JSON.

Frontend dikembangkan dengan menggunakan bahasa pemrograman Javascript. Pemilihan tersebut didasarkan pada keunggulan yang dimilikinya dalam hal kinerja, skalabilitas, dan efisiensi pengembangan. Tugas frontend mencakup tampilah halaman web dengan data dan hasil text generation, serta pengiriman permintaan RESTful API ke backend.

Backend, sebaliknya, dibangun dengan menggunakan Django, sebuah kerangka kerja web yang menggunakan bahasa pemrograman Python. pemilihan django dilandaskan pada keunggulan yang dimilikinya dalam hal fleksibilitas, keamanan, dan dukungan yang kuat dari komunitas pengembang. Backend memiliki tanggung jawab utama dalam mengelola logika aplikasi, berinteraksi dengan sistem basis data PostgreSQL, serta berkomunikasi dengan model bahasa BERT melalui API.

Model bahasa BERT merupakan sebuah model kecerdasan buatan yang digunakan untuk memahami konteks teks seara mendalam. Dalam konteks rekomendasi berita, BERT dapat digunakan untuk memahami konten berita dan preferensi pengguna dengan lebih baik, sehingga memberikan rekomendasi yang lebih akurat dan relevan. Kemudian model BERT dapat diakses melalui API yang disediakan oleh penyedia layanan cloud, seperti Azure atau Hugging Face, atau dapat dijalankan secara lokal dengan dukungan GPU.

Diagram arsitektur sistemnya sebagai berikut:



### 3.3 Modeling BERT

#### 3.3.1. Pemrosesan Data

##### 1) Tokenisasi Teks Berita:

Tokenisasi merupakan langkah kunci dalam pemrosesan data untuk menggunakan model BERT. Proses ini melibatkan mengonversi teks artikel berita menjadi token-token yang dipahami oleh model. Tokenisasi dilakukan menggunakan tokenizer BERT yang mengubah teks menjadi representasi token BERT yang spesifik. Ini memungkinkan model untuk memahami konteks teks dengan lebih baik.

##### 2) Pembentukan Sequences dan Padding

Setelah tokenisasi, data teks yang sudah diubah menjadi token-token BERT akan dibentuk menjadi sequences yang sesuai dengan format input yang diharapkan oleh model. Sejumlah token akan membentuk sequences yang sesuai dengan panjang maksimum yang diatur. Jika sequences memiliki panjang yang lebih pendek, dilakukan padding dengan menambahkan token khusus agar panjang sequences menjadi seragam.

### 3.3.2. Pemilihan Format Input

Pemilihan format input merupakan tahap krusial dalam menggunakan model BERT untuk tugas rekomendasi artikel berita. Tahapan ini melibatkan penentuan format input yang sesuai dengan arsitektur model BERT yang digunakan. Format input yang umum digunakan dalam konteks rekomendasi artikel berita melibatkan penambahan token-token khusus dan pengaturan struktur sequences. Berikut penjelasan lebih lanjut:

#### 1) Penambahan Token Khusus:

Dalam tahap ini, token-token khusus seperti token awal (seperti [CLS]), token akhir (seperti [SEP]), dan token pemisah (jika diperlukan) ditambahkan ke sequences. Token awal digunakan untuk menandai awal dari input, token akhir menandai akhir dari input, dan token pemisah digunakan jika ada lebih dari satu teks yang diinputkan (misalnya, kalimat pertama dan kedua). Ini membantu model BERT dalam memahami struktur teks dan mengenali elemen-elemen kunci dalam input.

#### 2) Penyesuaian Panjang Sequences:

Dalam tahap ini, token-token khusus seperti token awal (seperti [CLS]), token akhir (seperti [SEP]), dan token pemisah (jika diperlukan) ditambahkan ke sequences. Token awal digunakan untuk menandai awal dari input, token akhir menandai akhir dari input, dan token pemisah digunakan jika ada lebih dari satu teks yang diinputkan (misalnya, kalimat pertama dan kedua). Ini membantu model BERT dalam memahami struktur teks dan mengenali elemen-elemen kunci dalam input.

Pemilihan format input ini memastikan bahwa model BERT dapat mengonversi input teks ke dalam representasi vektor yang sesuai dengan arsitektur model. Dengan memasukkan token-token khusus dan mengatur struktur sequences, model BERT akan lebih efektif dalam



memahami konteks teks dan menghasilkan rekomendasi artikel berita yang akurat.

### **3.3.3. Fine-Tuning Model BERT**

Fine-tune model BERT merupakan Langkah kunci dalam pemanfaatan model yang sudah dilatih sebelumnya agar sesuai dengan data spesifik dari domain rekomendasi artikel berita. Proses fine-tuning melibatkan penyesuaian parameter model yang telah dipelajari secara umum dengan data khusus dari domain rekomendasi artikel berita. Berikut penjelasan rini tentang tahapan ini:

#### **1) Pemilihan Layer dan Parameter**

Pada tahapan ini, layer dan parameter pada model BERT yang akan diadaptasi terhadap dataset rekomendasi artikel berita ditentukan. Biasanya pada penelitian sebelumnya layer-layer terakhir atau Sebagian layer tertentu dari model BERT yang dipilih untuk disesuaikan dengan kata spesifik dari domain rekomendasi.

#### **2) Penyesuaian Hyperparameter**

Proses penyesuaian hyperparameter adalah langkah penting yang harus dilakukan dalam proses pelatihan. Hyperparameter adalah parameter yang digunakan untuk mengatur proses pelatihan model, mencakup variabel seperti ukuran batch, tingkat pembelajaran (learning rate), jumlah iterasi (epoch), jumlah lapisan (layer), jumlah unit per lapisan, fungsi aktivasi, fungsi kerugian, metode optimasi, dan variabel lain yang relevan. Parameter hipermodel memiliki peran kritis dalam mengarahkan proses pelatihan model deep learning menuju tingkat kinerja yang optimal. Terdapat dua pendekatan dalam penyetelan parameter hipermodel. Pendekatan pertama adalah penyetelan manual, sedangkan pendekatan kedua adalah mengkomparasikan proses penyetelan dengan tujuan menemukan konfigurasi parameter hipermodel terbaik. Pada tahap

awal eksperimen, parameter hipermodel diatur secara manual dengan maksud untuk mengamati interaksi antara setiap parameter dengan model yang bersangkutan. Setelah mendapatkan seperangkat parameter hipermodel yang sesuai, langkah selanjutnya melibatkan penyetelan lanjutan dengan menggunakan metode Grid Search.

### 3) Training, Validation dan Testing

Data awal yang digunakan untuk fine-tuning pertama-tama dibagi menjadi tiga subset: data training, data validation dan data testing. Skema yang akan digunakan adalah 80 untuk training, 10 untuk validation dan 10 digunakan untuk testing. Subset data training (80%) digunakan untuk melatih kembali model BERT. Model ini disesuaikan dengan data Microsoft News untuk rekomendasi artikel berita. Selama proses pelatihan, model akan belajar menyesuaikan representasinya dengan karakteristik data. Subset data validation (10%) digunakan untuk mengukur kinerja model selama proses fine-tuning. Metrik evaluasi seperti Precision, Recall, F1 Score, atau metrik rekomendasi lainnya digunakan untuk menilai performa model. Subset data testing (10%) akan digunakan untuk menguji performa model yang sudah di-fine-tuning. Hasil evaluasi pada data testing memberikan gambaran akurat tentang sejauhmana model mampu memberikan rekomendasi artikel berita yang akurat dan relevan.

### 4) Evaluasi Model

Evaluasi model BERT setelah proses fine-tuning adalah tahapan kuni dalam memastikan performa model yang dihasilkan sesuai dengan tujuan rekomendasi artikel berita. Dalam penelitian ini, evaluasi model BERT akan memanfaatkan beberapa metrik evaluasi yang inklusif dan sesuai dengan tujuan rekomendasi artikel berita.

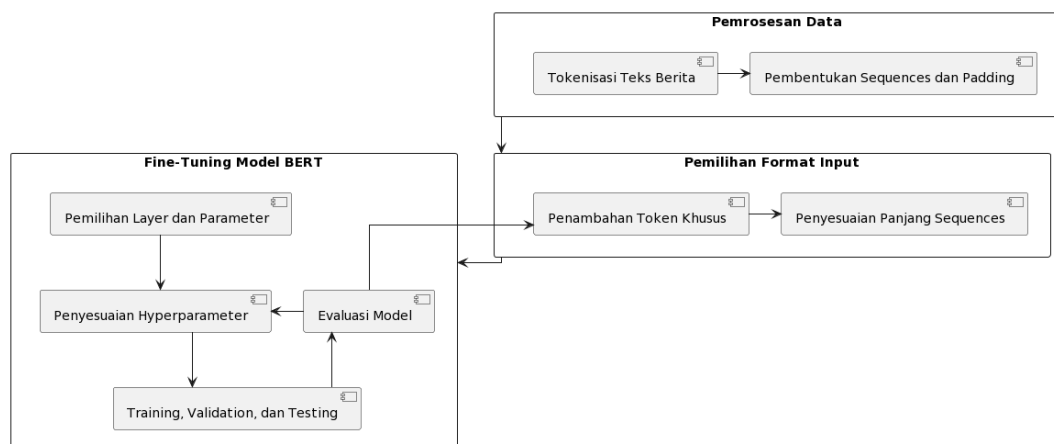
**Precision:** mengukur sejauh mana artikel yang direkomendasikan relevan dengan preferensi pengguna.

**Recall:** untuk mengukur sejauh mana artikel relevan yang direkomendasikan dibandingkan dengan jumlah total artikel yang seharusnya direkomendasikan.

**F1 Score:** penggabungan dari Precision dan Recall yang memberikan gambaran holistic tentang performa model.

**NDCG (Normalized Discounted Cumulative Gain):** metrik evaluasi yang mempertimbangkan peringkat artikel yang direkomendasikan, memberikan bobot yang lebih tinggi pada artikel yang relevan yang diurutkan lebih tinggi.

**AUC-ROC (Area Under the Receiver Operating Characteristic Curve):** metrik yang mengukur kemampuan model dalam memisahkan artikel yang relevan dan tidak relevan berdasarkan nilai threshold yang berbeda.





## DAFTAR PUSTAKA

Gupta, R., Jain, S., & Jain, R. (2020). News Recommendation System Using Collaborative Filtering. *International Journal of Computer Applications*, 170(10), 38-42.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*.

Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep Learning for News Recommendation Based on Long- and Short-term User Representations. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*.

Wang, X., Wang, X., Wu, J., Li, C., & Wang, H. (2021). A Survey on News Recommendation. *IEEE Access*, 9, 18938-18956.

Yang, F., Li, J., Wang, H., & Zhang, C. (2020). News Recommendation System Based on Deep Learning. *2020 4th International Conference on Data Science and Information Technology (ICDSIT)*.

Sun, Z., Wang, Y., & Zhang, J. (2021). Applications of BERT Model in Text Mining. *2021 International Conference on Artificial Intelligence and Big Data (ICAIBD)*.

Chen, Q., Liu, H., & Zhang, P. (2022). A Survey on News Recommendation Technology. *2022 IEEE 5th International Conference on Computer and Communication Engineering Technology (CCET)*.

Chen, W., Chen, J., & Zhao, J. (2019). Enhancing News Recommendation using Pre-trained Language Models. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval.

Wang, X., Wang, X., Wu, J., Li, C., & Wang, H. (2022). News Recommendation System Based on BERT. IEEE Access, 10, 15382-15392.

Liu, S., Li, J., & Wu, Q. (2020). BERT-Based News Recommendation System. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval.

Fangzhao Wu, Ying Qiao, Jiun-Hung Chen, Chuhan Wu, Tao Qi, Jianxun Lian, Danyang Liu, Xing Xie, Jianfeng Gao, Winnie Wu, dan Ming Zhou. "MIND: A Large-scale Dataset for News Recommendation." ACL 2020.