

**RESUME ARTIKEL ILMIAH**

**ALBERT: A Lite BERT for Self-supervised Learning of Language  
Representations**



**Rafif Ilafi Wahyu Gunawan**

**21081010093**

**MATA KULIAH PENGOLAHAN CITRA DIGITAL A081**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL**

**“VETERAN” JAWA TIMUR**

**2024**

## DETAIL PUBLIKASI

<b>Judul Artikel</b>	: ALBERT: A Lite BERT for Self-supervised Learning of Language Representations
<b>Nama Jurnal</b>	: -
<b>Volume</b>	: Vol. 57, DOI 10.1080, 25 April 2024, Halaman 1 - 11
<b>Penulis</b>	: Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, Radu Soricut
<b>Tahun Terbit</b>	: 2019
<b>Link Artikel</b>	: <a href="https://arxiv.org/abs/1909.11942">10.48550/arXiv.1909.11942</a>

## LATAR BELAKANG

Peningkatan ukuran model dalam pretraining representasi bahasa alami (NLP) sering menghasilkan performa yang lebih baik dalam tugas-tugas hilir. Namun, pertumbuhan model menghadapi tantangan seperti keterbatasan memori GPU/TPU, waktu pelatihan yang lama, dan degradasi performa. Untuk mengatasi masalah ini, ALBERT diperkenalkan sebagai versi ringan dari BERT dengan teknik pengurangan parameter yang mengurangi konsumsi memori sekaligus meningkatkan efisiensi pelatihan. Model ini juga memperkenalkan *sentence-order prediction* (SOP) sebagai pengganti metode prediksi urutan kalimat BERT, yang terbukti lebih efektif dalam meningkatkan tugas-tugas NLP multikalimat.

## ISI ARTIKEL

ALBERT memperkenalkan dua teknik utama untuk meningkatkan efisiensi parameter BERT:

1. Parameterisasi Embedding yang Difaktorkan: Memisahkan embedding kata dari dimensi layer tersembunyi untuk mengurangi jumlah parameter yang tidak relevan.
2. Berbagi Parameter Antar-Layer: Membagi parameter pada semua lapisan model, yang tidak hanya mengurangi jumlah parameter secara keseluruhan tetapi juga bertindak sebagai bentuk regulasi.

Selain itu, ALBERT menggantikan next-sentence prediction (NSP) dengan SOP yang lebih fokus pada koherensi antar-kalimat. Hasilnya, ALBERT menunjukkan peningkatan signifikan dalam benchmark NLP seperti GLUE, SQuAD, dan RACE.

## **DATASET**

Penelitian pada artikel ini menggunakan dataset pretraining umum yang terdiri dari:

- BOOKCORPUS: Dataset besar yang berisi koleksi buku.
- Wikipedia Inggris: Artikel-artikel terstruktur dari Wikipedia.

Dataset ini menyediakan sekitar 16GB teks tanpa kompresi. Tokenisasi dilakukan menggunakan SentencePiece dengan ukuran kosa kata 30.000.

## **METODELOGI**

Penelitian melibatkan beberapa langkah sebagai berikut:

1. Arsitektur Model: Menggunakan encoder Transformer dengan fungsi aktivasi GELU. Model ALBERT menggunakan dua modifikasi utama BERT:
  - a. Pengurangan parameter embedding melalui dekomposisi matriks embedding besar.
  - b. Berbagi parameter lintas layer.
2. Loss Fungsi: Mengganti NSP dengan SOP, yang lebih efektif dalam menangkap koherensi antar-kalimat.
3. Pelatihan: Dilakukan di TPU Cloud dengan optimizer LAMB dan batch size 4096, selama 125.000 langkah pelatihan.

## HASIL DAN PEMBAHASAN

ALBERT dirancang untuk mengatasi kendala dalam model BERT yang terkait dengan ukuran parameter besar dan kebutuhan memori yang tinggi. Eksperimen menunjukkan bahwa ALBERT berhasil memberikan hasil lebih baik atau setara dengan BERT pada berbagai tugas NLP, dengan keunggulan berikut:

- **Pengurangan Parameter:**
  - ALBERT menggunakan parameterisasi embedding yang difaktorkan dan berbagi parameter antar-layer. Dengan pendekatan ini, model ALBERT-large memiliki 18x lebih sedikit parameter dibandingkan dengan BERT-large (18 juta vs. 334 juta parameter).
  - Model ini juga menunjukkan efisiensi waktu pelatihan, dapat dilatih hingga 1,7x lebih cepat dibandingkan BERT.
- **Efisiensi Model:**
  - Meskipun memiliki parameter yang lebih sedikit, ALBERT-xxlarge menghasilkan skor **+1,5% lebih tinggi rata-rata** dibandingkan BERT-large pada tugas hilir, menunjukkan efisiensi parameter tanpa degradasi performa.

ALBERT dievaluasi pada tiga benchmark utama NLP:

- GLUE (General Language Understanding Evaluation): Mengukur kemampuan pemahaman bahasa alami secara umum.
- SQuAD (Stanford Question Answering Dataset): Dataset untuk tugas tanya-jawab (QA).
- RACE (ReAding Comprehension from Examinations): Tes pemahaman membaca yang dirancang untuk siswa menengah dan tinggi.

Hasil Evaluasi:

- GLUE Benchmark:
  - ALBERT-xxlarge mencetak 89,4 pada GLUE, meningkat dibandingkan BERT-large (86,6). Model ini menunjukkan keunggulan dalam tugas-tugas seperti pengenalan hubungan teks (MNLI, RTE) dan pengelompokan semantik (STS-B).
- SQuAD v1.1 dan v2.0:

- Pada SQuAD v2.0, ALBERT mencapai skor F1 sebesar 92,2, lebih baik dari model-model sebelumnya seperti XLNet (90,1) dan RoBERTa (89,8). Skor F1 ini menunjukkan kemampuan ALBERT untuk memahami konteks dan menghasilkan jawaban yang lebih presisi.
- RACE Benchmark:
  - ALBERT-xxlarge mencapai akurasi tertinggi 89,4%, meningkat signifikan dibandingkan dengan model BERT-large (72%) dan RoBERTa (83,2%). Ini menunjukkan kemampuan ALBERT dalam memahami teks multikalimat dengan struktur kompleks.

### **KESIMPULAN**

ALBERT menawarkan solusi inovatif untuk mengatasi keterbatasan BERT, seperti konsumsi memori tinggi dan waktu pelatihan yang panjang, melalui teknik pengurangan parameter. Model ini mempertahankan atau bahkan meningkatkan performa pada berbagai benchmark NLP, sekaligus menjadi lebih efisien. SOP sebagai komponen baru juga terbukti lebih berguna dibanding NSP. Hasil ini menunjukkan potensi besar ALBERT dalam meningkatkan efisiensi model pra-latihan di NLP.