**ALBERT (A Lite BERT) Architecture**

**Опис**

ALBERT (A Lite BERT) є спрощеною версією BERT (Bidirectional Encoder Representations from Transformers), яка зменшує параметри та покращує швидкість тренування і продуктивність моделі. ALBERT була представлена компанією Google в роботі "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations" у 2019 році.

**Архітектура**

ALBERT використовує декілька ключових оптимізацій в порівнянні з оригінальним BERT:  
1. Factorized Embedding Parameterization: ALBERT розділяє параметри векторів слів (word embedding) і параметри векторів прихованих шарів (hidden layers). Це дозволяє значно зменшити кількість параметрів у моделі.  
2. Cross-layer Parameter Sharing: ALBERT застосовує спільне використання параметрів між шарами трансформера, що зменшує кількість параметрів і дозволяє моделі бути більш легкою.  
3. Inter-sentence Coherence Loss: ALBERT використовує новий метод для навчання відносин між реченнями, що покращує здатність моделі розуміти текст на більш високому рівні.

**Особливості**

- Меньша кількість параметрів: Завдяки факторизації параметрів та спільному використанню параметрів, ALBERT має значно меншу кількість параметрів у порівнянні з BERT, що дозволяє швидше навчати модель та зменшує обчислювальні витрати.  
- Швидше тренування: Модель тренується швидше за рахунок зменшення кількості параметрів і оптимізацій в архітектурі.  
- Висока продуктивність: Незважаючи на меншу кількість параметрів, ALBERT досягає результатів, які конкурентоспроможні з оригінальним BERT на різних завданнях NLP.

**Переваги над іншими моделями**

- Оптимізація ресурсів: ALBERT потребує менше обчислювальних ресурсів та часу для тренування, що робить її більш практичною для використання в реальних додатках.  
- Збереження точності: ALBERT зберігає високий рівень точності на різних задачах NLP, навіть з меншою кількістю параметрів.  
- Гнучкість: Завдяки своїй архітектурі, ALBERT може бути адаптована для різних завдань обробки природної мови без значного зростання обчислювальних витрат.

**Порівняння з іншими моделями**

Порівнюючи ALBERT з іншими моделями, такими як BERT і RoBERTa:  
- BERT: Оригінальна модель з великою кількістю параметрів, потребує більше обчислювальних ресурсів і часу на тренування. Висока точність на багатьох завданнях NLP.  
- RoBERTa: Поліпшена версія BERT з кращими результатами на багатьох завданнях, але також вимагає більше ресурсів на тренування.  
- ALBERT: Оптимізована версія BERT з меншою кількістю параметрів і швидшим тренуванням, зберігаючи високу точність.

**Блок-схема алгоритму**

