

## INTISARI

Pendidikan tinggi sebagai pemasok tenaga kerja di dunia industri dapat memanfaatkan informasi *job requirement* melalui *job portal online* sebagai bahan pertimbangan dalam peninjauan dan perumusan kembali kurikulum. Ekstraksi informasi pada *job requirement* dapat digunakan untuk mencari atau mengenali entitas keahlian kerja. Metode ekstraksi yang dapat digunakan untuk mengenali entitas pada *job requirement* adalah *named entity recognition* (NER).

Beberapa tahun terakhir, penelitian di bidang *Natural Language Processing* (NLP) telah banyak mengadopsi *Transformer* untuk berbagai macam penugasan termasuk pada pemodelan NER. Mekanisme *self-attention* merupakan salah satu fitur utama yang membuat arsitektur *Transformer* unggul dalam penugasan NLP. Untuk mengenali urutan dari setiap masukan *sequence*, arsitektur *Transformer* menyematkan posisi absolut dengan menggunakan fungsi sinusoidal agar model memahami jarak antar kata pada saat perhitungan nilai *attention*. Akan tetapi dalam pemodelan NER dibutuhkan informasi jarak dan informasi arah sehingga hasil identifikasi entitas dapat lebih akurat. Selain itu pada penelitian terdahulu dikatakan bahwa nilai *attention* yang tidak dinormalisasi membuat identifikasi pada model NER menjadi lebih baik. Pemanfaatan *Pretrained Model Language* (PML) berbasis arsitektur *Transformer*, seperti *Bidirectional Encoder Representations from Transformers* (BERT) sangat membantu dalam mengatasi keterbatasan dataset untuk pemodelan NER. Oleh karena itu penelitian untuk modifikasi *Transformer* diimplementasikan pada BERT, selanjutnya disebut sebagai MASTER. Sebagai *baseline* dibuat PML IndoBERTSkill, BERTSkill-Shaw dan BERTSkill-Huang. Pemodelan NER identifikasi keahlian kerja dilakukan dengan melakukan *fine-tuning* pada PML.

Hasil evaluasi PML dengan menggunakan PLL memperlihatkan PML IndoBERTSkill memiliki performansi yang lebih baik dibandingkan BERTSkill-Shaw, BERTSkill-Huang, dan MASTER.

Hasil evaluasi performansi model NER pada tingkat token dan tingkat entitas memperlihatkan model yang dilatih dari *fine-tuning* PML MASTER (MAS-TENER) mempunyai nilai *precision*, *recall* dan *F1-Score* tertinggi dibandingkan model NER yang dilatih dari *fine-tuning* IndoBERTSkill, BERT-Shaw dan BERT-Huang. Nilai *precision*, *recall* dan *F1-Score* MAS-TENER pada evaluasi tingkat token adalah 83.49%; 83.52% dan 83.51%. Sedangkan pada evaluasi tingkat entitas nilai *precision*, *recall* dan *F1-Score* MAS-TENER adalah 93.55%; 93.13% dan 93.32%. Performansi MAS-TENER menggunakan nilai *precision*, *recall* dan *F1-Score* meningkat sebesar 1%—3% pada tingkat *token* dan 0,34% — 3% pada tingkat entitas.

Kata kunci: MASTER, MAS-TENER, BERTSkill, NERSkill, *Pretrained Model Language*

## ABSTRACT

*Higher education, as a supplier of the workforce in the industrial world, can utilize job requirement information from online job portals as a consideration in reviewing and reformulating the curriculum. Information extraction from job requirements can be employed to search for or recognize job skill entities. One of the methods for extracting entities from job requirements is named entity recognition (NER).*

*In recent years, research in Natural Language Processing (NLP) has widely adopted the Transformer, including in NER modeling. The self-attention mechanism is a crucial feature that makes the Transformer architecture excel in NLP tasks. To identify the sequence input, the Transformer architecture embeds absolute positions using a sinusoidal function to help the model understand the distances between words when calculating attention scores. However, NER modeling requires information about distances and directions to improve entity identification accuracy. Previous research has also suggested that unnormalized attention scores improve NER model identification. Utilizing Pretrained Model Language (PML) based on the Transformer architecture, such as Bidirectional Encoder Representations from Transformers (BERT), has been very helpful in overcoming dataset limitations for NER modeling. Therefore, the study aimed to modify the Transformer implemented in BERT referred to as MASTER. As a baseline, PML models were created, including IndoBERTSkill, BERTSkill-Shaw, and BERTSkill-Huang. Modeling NER for skill identification was performed by fine-tuning these PMLs.*

*PML evaluation results using PLL show that PML IndoBERTSkill has better performance than BERTSkill-Shaw, BERTSkill-Huang, and MASTER.*

*The performance evaluation results of the NER model at the token level and entity level show that the model trained from fine-tuning PML MASTER (MAS-TENER) has the highest precision, recall and F1-Score values compared to the NER model trained from fine-tuning IndoBERTSkill, BERT-Shaw and BERT-Huang. The precision, recall and F1-Score MAS-TENER values at the token level evaluation are 83.49%; 83.52% and 83.51%. Meanwhile, at the entity level evaluation, the precision, recall and F1-Score MAS-TENER values were 93.55%; 93.13% and 93.32%. MAS-TENER performance using precision, recall and F1-Score values increased by 1%—3% at the token level and 0.34%—3% at the entity level.*

**Keywords:** MASTER, MAS-TENER, BERTSkill, NERSkill, Pretrained Model Language