

## INTISARI

### NEURAL MACHINE TRANSLATION DENGAN STRATEGI PENANGANAN LOW-RESOURCE UNTUK BAHASA INDONESIA- WOLIO

LA ODE ABD. EL HAFIZH HIDAYAT  
21/484528/PPA/06191

Neural Machine Translation (NMT) mengandalkan pemodelan dengan arsitektur sequence-to-sequence (seq2seq) untuk mencapai performa optimal dalam membangun mesin penerjemah. Model seq2seq seperti Transformer terbukti meningkatkan kualitas penerjemahan dengan memetakan secara efektif rangkaian kata dari bahasa asal ke bahasa tujuan. Namun model NMT berpotensi performanya menurun ketika menghadapi bahasa dengan *data training* yang terbatas. Penyebabnya adalah NMT membutuhkan puluhan hingga ratusan juta kalimat paralel, yang hanya tersedia untuk bahasa dengan sumber daya melimpah, sedangkan sebagian besar bahasa, termasuk bahasa Wolio, kekurangan dataset sebesar itu.

Penelitian ini bertujuan mengatasi kerentanan NMT terhadap skenario *low-resource*, dengan tidak bergantung kepada jumlah dataset yang besar. Penelitian diawali dengan memodelkan NMT berbasis Transformer untuk mencapai performa yang paling baik pada dataset autentik. Basis model ini digunakan untuk menerjemahkan monolingual dataset guna menghasilkan dataset sintesis berbahasa Wolio. Langkah berikutnya adalah menerapkan augmentasi data dengan memodelkan ulang sistem NMT melalui skema Back-translation dan BPE dropout. Back-translation memanfaatkan data sintesis yang dihasilkan, sementara BPE-dropout menambah data dengan memanfaatkan variasi morfologis dari dataset autentik.

Adapun pendekatan yang berhasil mengatasi situasi *low-resource* di bahasa Indonesia-Wolio adalah dengan cara: pertama, mengurangi jumlah operasi BPE; kedua, mengurangi kedalaman parameter Transformer; dan ketiga, menerapkan augmentasi data berbasis BPE-dropout. Kombinasi strategi ini mencapai skor evaluasi tertinggi dengan 17,98 BLEU dan 40,60 chrF, serta secara akumulatif meningkatkan performa MT sebesar 7,51 BLEU dan 12,93 chrF. Sebaliknya, augmentasi data menggunakan Back-translation kurang berdampak dan malah menurunkan kinerja model.

**Kata kunci:** NMT, *low-resource*, Back-Translation, Transformer, BPE-Dropout

## ABSTRACT

### NEURAL MACHINE TRANSLATION WITH LOW-RESOURCE HANDLING STRATEGIES FOR INDONESIAN-WOLIO

LA ODE ABD. EL HAFIZH HIDAYAT  
21/484528/PPA/06191

Neural Machine Translation (NMT) relies on sequence-to-sequence (seq2seq) architecture to achieve optimal performance in machine translation. Seq2seq models, such as the Transformer, have demonstrated improvements in translation quality by effectively mapping word sequences from the source language to the target language. However, NMT models may exhibit diminished performance when applied to languages with limited training data. This is due to the requirement of tens to hundreds of millions of parallel sentences, which are predominantly available for high-resource languages. Consequently, most languages, including Wolio, lack access to such extensive datasets.

This study seeks to address the vulnerability of NMT in low-resource scenarios without depending on large number of datasets. It begins by modeling Transformer-based NMT to achieve adequate performance on authentic datasets. This model serves as a foundation for translating monolingual datasets to produce synthetic Wolio datasets. Subsequently, the study employs data augmentation by re-modeling the NMT system using Back-translation and BPE dropout schemes. Back-translation leverages the generated synthetic data, while BPE dropout enhances data by incorporating morphological variations of the authentic dataset.

The strategies that effectively address the low-resource challenge in Indonesian-Wolio translation include: first, reducing the number of BPE operations; second, decreasing the depth and width of Transformer parameters; and third, implementing BPE-dropout-based data augmentation. This combination of methods resulted in the highest evaluation scores, achieving a BLEU score of 17.98 and a chrF score of 40.60, with cumulative improvements in MT performance of 7.51 BLEU and 12.93 chrF. Conversely, data augmentation through Back-translation proved less effective and led to a decline in model performance.

**Keywords:** NMT, *low-resource*, Back-Translation, Transformer, BPE-Dropout