IMPLEMENTING NEURAL MACHINE TRANSLATION USING BI-DIRECTIONAL LSTM ARCHITECTURE FOR BIDIRECTIONAL TRANSLATION OF INDONESIAN–NGAPAK BANYUMASAN

HERY SETIYAWAN JORDI

Program Studi Informatika, Fakultas Sains & Teknologi
Universitas Teknologi Yogyakarta
Jl. Ringroad Utara Jombor Sleman Yogyakarta
E-mail: herysetiyawan368@gmail.com

ABSTRACT

Manual translation between Indonesian and Ngapak Banyumasan still faces challenges in terms of consistency and accessibility, underscoring the need for an automatic translation system that delivers fast, accurate results. This study aims to design and implement a Neural Machine Translation (NMT) system based on an encoder–decoder architecture, utilizing stacked Bidirectional Long Short-Term Memory (Bi-LSTM) layers in the encoder and stacked LSTM layers in the decoder, enhanced with the Bahdanau Attention mechanism for bidirectional Indonesian–Ngapak translation. The dataset comprises 23,592 sentence pairs that have undergone curation and preprocessing, including tokenization using SentencePiece with the Unigram model. The model was trained with Early Stopping and evaluated using loss, accuracy, and BLEU scores. Test results indicate that the configuration using 80% of the data for training achieved the best performance, with a BLEU-4 score of 76% and validation accuracy of approximately 95%. Additionally, User Acceptance Testing (UAT) yielded an average score of 85%, considered very good. Therefore, the developed system is capable of producing accurate, well-structured translations and is suitable as an automatic translation solution, while also supporting the preservation of the Ngapak language through artificial intelligence.

Keywords: Neural Machine Translation, Bi-LSTM, Ngapak Banyumasan Language, SentencePiece.