

Title: Improving Bangla Linguistics: Advanced LSTM, Bi-LSTM, and Seq2Seq Models for Translating Sylheti to Modern Bangla

Author: Sourav Kumar Das; Md. Julkar Naeen; Md. Jahidul Islam; Md. Anisul Haque Sajeeb; Narayan Ranjan Chakraborty; Mayen Uddin Mojumdar

Abstract: Bangla or Bengali is the national language of Bangladesh, people from different regions don't talk in proper Bangla. Every division of Bangladesh has its own local language like Sylheti, Chittagong etc. In recent years some papers were published on Bangla language like sentiment analysis, fake news detection and classifications, but a few of them were on Bangla languages. This research is for the local language and this particular paper is on Sylheti language. It presented a comprehensive system using Natural Language Processing or NLP techniques for translating Pure or Modern Bangla to locally spoken Sylheti Bangla language. Total 1200 data used for training 3 models LSTM, Bi-LSTM and Seq2Seq and LSTM scored the best in performance with 89.3% accuracy. The findings of this research may contribute to the growth of Bangla NLP researchers for future more advanced innovations.

Keywords: Bangla local language, Sylheti language, LSTM, Bi-LSTM, Language translation, Natural Language, Processing Techniques

DOI: <https://doi.org/10.1109/ICCCNT61001.2024.10723886>

<https://ieeexplore.ieee.org/document/10723886>