

Natural Language Processing Lab

Exercise 8:

Application: Machine Translation

Dataset: Download the dataset from

[Tamil-English-Dataset/Dataset at master · Ishikahooda/Tamil-English-Dataset · GitHub](#)

[English to tamil data \(kaggle.com\)](#)

Task:

- Develop a model that can translate a Tamil input sentence to English.
- Analyze the impact of various embedding techniques in the translation task.
- Use sequence model to translate the Tamil input sentence and analyze the performance of various techniques.
- Translate the input sentences using Transformer model and understand its functioning.
- Translate the following Tamil Phrase and evaluate your model based on the expected output.

Input 1: நான் மிகவும் சந்தோஷமாக இருக்கிறேன்

Expected Output: Im so happy

Input 2: அது அவசியமில்லை

Expected Output: It wasnt necessary

Input 3: தயவுசெய்து அதை மீண்டும் செய்யவும்

Expected Output: Please do that again

Input 4: அது ஒரு நல்ல யோசனை

Expected Output: That is a good idea

Input 5: அவர்கள் ஒன்றாக வேலை செய்ய ஒப்புக்கொண்டனர்

Expected Output: They agreed to work together