

BENGALI TRANSLITERATION

Technology

Google Colab, Streamlit, Hugging Face Hub

BENGALI TRANSLITERATION

Tasks:

- Task 1: Dataset Preparation
- Task 2: Data Preprocessing
- Task 3: Model Selection
- Task 4: Model Training and Saving

Task 1 - Dataset Preparation

- **Steps:**

- a. Install required libraries: datasets, transformers.**
- b. Load the dataset from Hugging Face: SKNahin/bengali-transliteration-data.**
- c. Split dataset into training (90%) and validation (10%) subsets.**

Task 2 - Data Preprocessing

- **Steps:**
 - a. **Load tokenizer for the google/mt5-small model.**
 - b. **Define a tokenization function to preprocess inputs (rm) and targets (bn).**
 - c. **Apply tokenization to the dataset.**

Task 3 - Model Selection

- **Model: mT5 (Multilingual Text-to-Text Transfer Transformer)**
- **Pretrained Model Used: google/mt5-small**

Task 4 - Model Training

- **Steps:**
 - a. **Define device (GPU/CPU).**
 - b. **Create a custom dataset class for DataLoader.**
 - c. **Train the model using AdamW optimizer.**
 - d. **Save the model and tokenizer.**

**THANK
YOU**