Methodology

Data Collection: Collected data from **Ajker Bissho**, **MediGK**, **Current Affairs**, etc.

Preprocessing:
- Tokenization
- Stop-word removal
- Data cleaning for improving dataset quality.

Model Selection: Fine-tune pre-trained models like **mt5-small**, **BanglaGPT**, **Dialo-GPT or GPT-2.**

Training Process: Train with the dataset using transfer learning techniques.

Evaluation Metrics: Use metrics like F1-score, BLEU score, and accuracy to measure model performance.

Daffodil *International* **University**

B.Sc. Title Defense

Expected Outcomes

Our model will achieve accuracy in this new Bengali dataset, enhancing accessibility. The system will be useful for educational platforms, customer support, and general inquiries.

cont..