Explanation of the Task

Data Preparation:

- Started with the data preparation, explored the data by counting the rows, considering noise, assuming multiple entries in a single field by row and column, handling Bengali & English tags, sentence and word identification etc.
- Padded the sentences, POS tags, and NER tags to a fixed length (max_len).

Mapping and Conversion:

- Converted tokens, POS tags, and NER tags to indices for model input.
- Created special mappings for unknown words and padding tokens.

Data Splitting:

- Split the data into training, validation, and test sets.

Model Definition:

- Defined a bidirectional LSTM model to perform both POS tagging and NER tasks simultaneously. The model is defined using Keras, where the model predicts both POS tags and NER tags simultaneously for each token in a sentence. The architecture includes an embedding layer, a bidirectional LSTM, and two TimeDistributed dense layers for POS and NER outputs.

  The primary goal was to use pre-trained transformer-based models. However, being a first-timer in handling the Bangla language in my journey, I had to end up with this architecture learning from online sources.

Model Training:

- Trained the model using the training data and validated it using the validation data.

Model Evaluation:

- Evaluated the model on the test set and printed the results.

Visualization:

- Plotted the accuracy and loss values during training and validation to visualize the model's performance.