

Assignment 2 (Contest) : Design of NER for Nepali language in `SpaCy` .

Objective:

Your task is to design and implement a Named Entity Recognition (NER) model for the Nepali language using the `SpaCy` library. This assignment will involve collecting, cleaning, annotating a dataset and training the NER Model. Your NER model should recognize and categorize entities into the following tags:

- **PERSON**: Names of individuals
- **DATE**: Dates and time expressions
- **ORGANIZATION**: Names of organizations, political parties, institutions, etc.
- **POST**: Titles or positions held by individuals (e.g., Prime Minister, Minister)
- **LOCATION**: Names of places (cities, countries, landmarks, etc.)
- **NUMBER**: Numeric expressions
- **MONEY**: Monetary amounts

Baseline:

The baseline model that achieves an accuracy of **60%** will be provided. Your goal is to **improve** upon this baseline by following the steps outlined in the provided python notebook file.

Most of the steps are explained in the notebook.

You can improve the code, add/reduce more steps as per your need.

Model Training:

- Use the `SpaCy` library to train an NER model on your annotated dataset.
- Fine-tune the model's parameters to improve its performance.

Model Evaluation:

- Evaluate the performance of your NER model on your own test set.
- Compare your model's accuracy, precision, recall, and F1-score with the provided baseline.

Evaluation Dataset:

- The evaluation dataset will be released after the deadline of the submission.

Deliverables:

- **Annotated Dataset**: The dataset you collected and annotated.
- **Python Notebook**: Containing your code and analysis.
- **Best Model**: The best model produced should be submitted in zip file.

Evaluation Criteria:

- **Accuracy Improvement**: How much you have improved over the baseline.
- **Data Quality**: The quality and consistency of the annotated dataset.
- **Model Design**: The effectiveness of your model's design and the techniques you used.

Declaring Winner:

- The model with highest accuracy will be considered as winner.
- The winner will be awarded with bonus marks.