1. Introduction:

Part-of-speech (POS) tagging is a fundamental task in natural language processing (NLP), aiming to assign grammatical categories (tags) to words in a sentence. In this report, we evaluate the POS tagging system using Viterbi algorithm.

---

2. Model Description:

The model is trained on a labeled dataset and employs techniques such as numerical stability through **logarithmic metrics** and handling unknown words using add-one smoothing.

---

3. Data Statistics:

- Vocabulary Size: 6606
- Number of Unique POS Tags: 46
- Number of Unique Tags After Start and End Tags: 48
- Size of **train_tag_transition** ($P(tag_i)$): 47
- Size of **train_tag_pair_counts** ($P(tag_j \mid tag_i)$): Variable based on training data
- Size of **train_tag_emission** ($P(tag_i)$): 46
- Size of **train_tag_word_counts** (P(word | tag)): Variable based on training data

---

4. Evaluation Metrics:

We evaluate the POS tagging system using standard metrics including Precision, Recall, F1 Score, and Accuracy from sklearn library. These metrics provide insights into the system's performance in terms of both correctness and completeness of tagging.

---

5. Inference Results on Training Data:

- F1 Score: 0.9808
- Precision: 0.9810
- Recall: 0.9807
- Accuracy: 0.9807

---

6. Inference Results on Test Data:

- F1 Score: 0.6973
- Precision: 0.7368
- Recall: 0.7153
- Accuracy: 0.7153

---

7. Assumptions:

- The tag sequence for each sentence is appended by <s> tag and ended by </s> tag to demarcate the sentence boundaries.
- Instead of multiplying probabilities across words, logarithms of probabilities are added for better numerical stability.
- In the case of test data, where words not present in the training emission probabilities are encountered, add-one smoothing is applied.
- In all other cases, word-pair doesn't exist, tag-tag pair doesn't exist, and probability score is assumed to be -infinity.
