

NLP Project Part 2

Q5: Custom Feature Extractor 1

1. **Feature Description:**

The feature counts the number of unique parts of speech (POS) tags present in the input text.

2. **Purpose and Motivation:**

This feature measures the lexical diversity of the text, providing insight into how varied the user's language is. A richer variety of POS tags may indicate more sophisticated input, allowing the chatbot to adjust its responses accordingly.

3. **Sample Inputs and Outputs:**

- **Input 1:** "The cat sat on the mat."
 - **Output:** 4 (Unique POS tags: DT, NN, VBD, IN)
- **Input 2:** "I love programming in Python because it is fun."
 - **Output:** 7 (Unique POS tags: PRP, VBP, VBG, IN, NNP, RB, JJ)

Q6: Custom Feature Extractor 2

1. **Feature Description:**

This feature counts the number of coordinating conjunctions, such as "and" and "or", in the input text.

2. **Purpose and Motivation:**

It helps measure sentence complexity. The use of multiple coordinating conjunctions generally indicates the presence of multiple clauses, which may correlate with more complex language. The chatbot can use this information to provide simpler or more complex responses based on the user's writing style.

3. **Sample Inputs and Outputs:**

- **Input 1:** "I wanted to go out, but it started raining."
 - **Output:** 1 (One occurrence: "but")
- **Input 2:** "She likes coffee, and he likes tea or juice."
 - **Output:** 2 (Occurrences: "and", "or")