**Automated Response Generation for Customer Support**

**Assignment Title:** Automated Response Generation for Customer Support

**Objective:**

This project aims to develop a model capable of generating automated responses to customer queries within the customer support domain. The model will learn from existing customer support interactions, allowing it to provide efficient and accurate responses to user inquiries.

**Dataset:**

The dataset used for training the model is the "Customer-Support-Responses" dataset, publicly available on Hugging Face Datasets: <https://huggingface.co/datasets/Kaludi/Customer-Support-Responses>

**Tasks:**

This project encompasses several key steps:

1. **Data Exploration and Preprocessing:** We will begin by exploring the characteristics of the dataset, including analyzing the vocabulary, sentence lengths, and overall structure of the conversations. Following exploration, the data will undergo preprocessing steps to ensure it's suitable for model training. This might involve tasks like cleaning text, handling typos, and potentially tokenizing the data.
2. **Model Selection and Training:** Our approach will involve training a transformer-based model like T5 model.
3. **Model Fine-tuning:** After initial training, we will employ fine-tuning techniques to enhance the model's performance. This may involve focusing on specific aspects like response coherence and relevance to the user's query. Fine-tuning can involve techniques like curriculum learning or specifically tailoring the training data to emphasize desired qualities in the generated responses.
4. **Evaluation:** Once the model is trained and fine-tuned, a comprehensive evaluation process will be undertaken. We will employ metrics like BLEU score and ROUGE score to assess the quality and similarity of the generated responses to the human-written references. Additionally, human evaluation can be performed to gauge the appropriateness and naturalness of the model's output.

**Deliverables:**

* **Code:** The code used for data exploration, preprocessing, model training, fine-tuning, and evaluation will be provided, ensuring reproducibility of the results.
* **Documentation:** This document serves as a detailed explanation of the project's purpose, steps involved, and the selected dataset. Additionally, it can be further expanded to include insights from the data exploration, results of the evaluation process, and a discussion of potential improvements.