# Project 1: Q&A Chatbot with Vector Store

**Project Overview:**

In this project, you will be tasked with creating a Q&A chatbot that utilizes a vector store and Language Model (LLM) to provide accurate responses to user queries. You will be provided with the necessary resources to get started, including bootstrap code, input files, and a demo video showcasing the working of the chatbot. Your main objectives will be to create a vector store from provided Ubuntu documentation in markdown format and integrate it with the chatbot system.

**Tasks:**

Part 1: Create Vector Store

Your first task is to create a vector store using the provided Ubuntu documentation in markdown format. The vector store will serve as the knowledge base for the chatbot.

Part 2: Create Chatbot using Vector Store and LLM

In the second part of the project, you will build a chatbot that utilizes the vector store created in Part 1 and a Language Model (LLM) to perform Q&A with users. The chatbot should be able to understand user queries, search for relevant answers from the vector store, and provide appropriate responses.

Requirements:

- Utilize the provided bootstrap code, input files, and demo video to understand the workings of the chatbot.

- Integrate the vector store and LLM into the chatbot codebase.

- Implement a chatbot that can interpret user queries.

- Utilize the vector store to search for relevant answers.

- Generate appropriate responses based on the user's query.

- Implement proper error and exception handling.

**General Requirements:**

- Utilize best practices in coding and documentation.

- Follow a modular and organized approach in your code implementation.

- Ensure your code is well-structured, readable, and maintainable.

- Utilize logging for different levels.

# Project 2: Text Classification Model for Ubuntu Customer Centre Inquiries

**Project Overview:**

The aim of this project is to develop a text classification model that can accurately categorize the Ubuntu Customer Centre inquiries into distinct topics or categories. The provided dataset contains approximately 150,000 inquiries from customers, and the candidate may choose to work with a subset of this data. The end goal is to build a robust model that achieves decent accuracy and identifies at least 5 distinct topics.

**Deliverables:**

At the completion of this project, you should provide the following deliverables:

Codebase: Share your well-structured and well-documented code for data preprocessing, feature engineering, model training, and evaluation. Use appropriate Python libraries and frameworks for efficient implementation.

**Note:**

- You are allowed to use any relevant open-source libraries or frameworks.

- Consider utilizing techniques to address challenges such as class imbalance, overfitting, or handling long or noisy texts.

- Feel free to explore and experiment with different approaches or improvements beyond the initial project requirements.

**Ubuntu customer inquiries dataset**