

Approach and Methodology for Smart Search Tool

1. Problem Statement

- Briefly describe the goal: Building a smart search system for Analytics Vidhya courses that efficiently matches user queries with course details.

2. Dataset Collection

- Mention that the dataset was manually curated from the [Analytics Vidhya Courses](#) platform.
- Include information about the dataset's structure (`title`, `description`, `curriculum` columns).

3. Embedding Model Selection

- Explain the choice of `all-MiniLM-L6-v2` from Sentence-Transformers:
 - Lightweight yet effective for generating contextual embeddings.
 - Well-suited for text similarity tasks with limited computational resources.

4. Vector Indexing with FAISS

- Describe how FAISS was used for efficient similarity search:
 - Indexed embedding vectors with L2 distance for fast query matching.

5. Methodology

- **Data Preprocessing:** Cleaned, concatenated, and embedded course data (`title`, `description`, `curriculum`).
- **Embedding Generation:** Used Sentence-Transformer to encode course data and user queries.
- **Search Functionality:** Utilized FAISS to find top-k similar embeddings for a user-provided query.

6. Tool Deployment

- The tool is deployed via Google Colab with Gradio. You can access it here: [Smart Search Tool Public Link](#).

7. Challenges and Future Enhancements

- Mention challenges like dataset size and manual curation.
- Propose enhancements like incorporating more advanced LLMs or a larger dataset for improved results.