Adobe India Hackathon 2025 – Connecting the Dots

# Round 1B – Persona-Based Section Retrieval

This round was more advanced — we were given multiple PDFs and a user persona with a specific task. The goal was to intelligently extract the most relevant sections from across the documents and return them as a ranked list in JSON format.

To implement this, we turned to semantic search. We used pre-trained embeddings from Sentence Transformers (MiniLM model) to understand the context of both the user’s task and the document content.

Steps followed:

* • Chunked the content of each PDF into meaningful sections
* • Embedded both task query and content using Sentence Transformers
* • Calculated cosine similarity scores
* • Selected top-matching results and returned them in ranked order

The entire system again ran inside a Docker container (CPU-only), producing results in under 60 seconds.

Here is a sample of the terminal output:

A screenshot of a computer program

AI-generated content may be incorrect.