SHL Assessment Recommendation System – Final Report

Objective:

To evaluate the performance of different embedding models used to build a recommendation engine for SHL assessments. Two approaches were compared: Google Gemini Embeddings and Sentence Transformers (BAAI/bge-small-en).

Data Collection:

I scraped the SHL official site using a custom notebook (SHL_Site_Scrape.ipynb) to obtain the assessment data. The scraped content includes assessment names, descriptions, durations, support types, job levels, and links. The final structured dataset was saved as shl_assessments_full_data.csv.

Approach:

Model 1: Google Gemini API (embedding-001)

- API-based embedding model from Google Generative AI
- Integration was done using embedContent() endpoint via HTTP requests
- Texts were wrapped with headers ("content": {"parts": ["..."]}) to match document embedding requirements

Challenges:

- API quota was limited, and the response was slower
- Embeddings were of lower quality when applied to SHL's catalogue
- Retrieval performance was inconsistent for practical queries

Results:

- Mean Recall@3: 0.57
- Mean MAP@3: 1.43

Model 2: Sentence Transformers (BAAI/bge-small-en)

- Open-source transformer model via Hugging Face's sentence-transformers
- Embeddings were generated locally in Colab
- Combined assessment title, description, type, duration, and level for each vector
- Indexed with FAISS using cosine similarity (inner product)

Advantages:

- Fast, local, reproducible
- No API quota or latency
- Performed better in job-specific matching tasks

Results:

- Mean Recall@3: 0.71
- Mean MAP@3: 1.57

Observations:

- Gemini worked well for generic embedding but lacked fine control and domain fit
- Sentence Transformers produced more consistent and accurate recommendations
- Sentence Transformers were chosen for final deployment due to:
  - Better semantic alignment
  - Higher accuracy
  - Simpler deployment (no API keys)

Challenges & Key Insights
- The dataset did not include technical assessments (e.g., Java, Python), which impacted some queries.
- Focused benchmark evaluation on customer service, industrial, and banking roles where data coverage exists.

Tools & Libraries Used
- NLP: sentence-transformers, Google Generative AI
- Search: faiss-cpu, sklearn, numpy, pandas
- Web/API: Gradio, FastAPI
- Deployment: Hugging Face Spaces (UI) and Render.com (API)

Deployment:
Frontend Demo (Hugging Face):
SHL Site Scrapping Link
Google Gemini Link
Sentence Transformer Link
API Endpoint (Render):
https://shl-assessment-engine-api.onrender.com/health **Test**: Link

Final Notes
- Built both Gemini-based and sentence-transformer-based models.
- Successfully deployed for public demo and evaluation.
- The system can scale with a larger dataset for more comprehensive recommendations.