

Proje Özeti: Bilgi Tabanı (Embedding + FAISS Index) Süreci (Büşra)

1. Yapılanlar

- **PDF'den bilgi çıkarma:**
Projemize eklenen PDF dosyasının (`all_pdf_bootcamp.pdf`) tamamındaki metinler otomatik olarak okunup dijitalleştirildi.
 - **Metinleri parçalara ayırma (chunk'lama):**
Okunan metinler, anlam bütünlüğü ve verimli arama için 500 karakterlik çakışmalı parçalara ayrıldı. 983 adet chunk oluşturuldu.
 - **Embedding (vektörleştirme) işlemi:**
Her bir metin parçası, Hugging Face'in açık kaynak `all-MiniLM-L6-v2` modeliyle çok boyutlu vektörlere (embedding) dönüştürüldü.
 - **FAISS index oluşturma:**
Elde edilen embedding vektörleri FAISS kütüphanesiyle bir vektör veri tabanında indekslendi. Böylece daha sonra hızlı ve etkili arama yapılabilecek altyapı hazırlandı.
 - **Çıktıların kaydedilmesi:**
Oluşan vektör veritabanı (`faiss_index.index`) ve parça metinler (`faiss_index_chunks.pkl`) dosya olarak projeye kaydedildi.
-

2. Elde Edilen Çıktılar

- **`faiss_index.index`**
→ Embedding vektörlerini barındıran FAISS vektör veritabanı dosyası
 - **`faiss_index_chunks.pkl`**
→ Tüm chunk (parça) metinlerini içeren pickle dosyası
-

3. Özetle

- Projemiz için **veri arama ve soru-cevap altyapısının** en kritik kısmı olan bilgi tabanı başarıyla oluşturuldu.
- Kodlar ve çıktılar tekrar kullanılabilir; yeni PDF eklenirse süreç kolayca tekrarlanabilir.
- Sonraki adımda, bu bilgi tabanından akıllı arama veya chatbot uygulamaları geliştirilebilir.