



ÇUKUROVA ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

YAPAY ZEKA VE VERİ MÜHENDİSLİĞİ ANABİLİM DALI

YZ007

İleri Hesaplama Dilbilim

PROJE ÖDEVİ

“Dijital Dönüşüm ReAct Agent Projesi”

Hazırlayan:

Adı Soyadı: Zeynep ÇÖL

Öğrenci No: 2025911088

İçindekiler

1. Projenin Amacı ve Kapsamı	3
2. Seçilen Sektör ve Gerekçe.....	3
3. Veri Kaynağı ve Knowledge Base Oluşturma	3
3.1 Kullanılan Veri.....	3
3.2 Chunklama Stratejisi	4
4. Kullanılan Yöntem: RAG (Retrieval-Augmented Generation)	4
4.1 RAG Mimarisi.....	4
5. Agentic (ReAct) Yaklaşım Açıklaması	4
6. Benchmark Tasarımı.....	5
6.1 Benchmark Yapısı	5
6.2 Kullanılan Skorlama Yöntemi (Nicel).....	5
6.3 Denenen Soruların Çıktıları	6
6.4 Tüm Sorular İçin Benchmark Sonuçları.....	8
6.4.1 Konu İçi Sorular (50 Soru).....	8
6.4.2 Konu Dışı Sorular (10 Soru)	8
7. Model Karşılaştırmaları (Nicel).....	9
8. Projenin Güçlü ve Zayıf Yönleri.....	10
8.1 Projenin Öne Çıkan Güçlü Yönleri	10
8.2 Projenin Sınırlı ve Zayıf Yönleri	10
9. Sonuç.....	10

1. Projenin Amacı ve Kapsamı

Bu projenin amacı, Dijital Dönüşümün İşletmeler Üzerindeki Etkisi alanında, statik cevap üreten bir dil modelinden ziyade; doküman temelli düşünen, sorgulayan ve karar veren bir Agent (Agentic yaklaşım) geliştirmektir. Çalışma, verilen Sectoral ReAct Agent Design projesi doğrultusunda hazırlanmıştır.

Proje iki ana hedefe sahiptir:

1. Dijital dönüşüm alanına özgü bir RAG (Retrieval-Augmented Generation) tabanlı bilgi erişim sistemi kurmak
2. Bu sistemi benchmark soruları ile test ederek nicel (sayısal) ve nitel karşılaştırmalar yapmaktadır.

2. Seçilen Sektör ve Gerekçe

Seçilen sektör: Dijital Dönüşüm ve İşletmeler

Gerekçe:

- Dijital dönüşüm, sadece teknolojik değil; organizasyonel yapı, insan faktörü, karar alma süreçleri ve rekabet gücü gibi teknik bilgi gerektiren çok boyutlu bir alandır.
- Akademik literatürü güçlündür ve teknik dokümanlarla (makale, rapor) beslenmeye uygundur.
- RAG mimarisiyle test edilmeye elverişlidir.

3. Veri Kaynağı ve Knowledge Base Oluşturma

3.1 Kullanılan Veri

- Kaynak: "Dijital Dönüşümün İşletmeler Üzerindeki Etkisi" akademik metni
- Format: digital_transformation.txt

PDF yerine TXT tercih edilmiştir çünkü:

- Metin doğrudan işlenebilmiştir
- OCR veya PDF parse hataları önlenmiştir
- Chunk bazlı embedding üretimi daha kontrollü yapılabilemiştir

3.2 Chunklama Stratejisi

Metin, ortalama 400 kelimelik parçalar (chunks) haline bölünmüştür.

Amaç:

- Vektörlerin çok genel olmamasını sağlamak
- Soru-doküman eşleşmesinde semantik hassasiyeti artırmak

4. Kullanılan Yöntem: RAG (Retrieval-Augmented Generation)

Bu projede Path A – RAG yöntemi uygulanmıştır.

4.1 RAG Mimarisi

- Embedding Model: sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2
- Vektör Temsili: Cümle embeddingleri
- Benzerlik Ölçütü: Cosine Similarity

Her soru için:

1. Soru embedding'e dönüştürülür
2. Doküman chunk embeddingleriyle karşılaştırılır
3. En yüksek cosine similarity skoruna sahip chunk seçilir

Sistem cevap üretmez, yalnızca ilgili bağlamı (context) döndürür. Bu, ReAct mimarisinin gereğidir.

5. Agentic (ReAct) Yaklaşım Açıklaması

Bu çalışma tam LLM tabanlı bir ReAct Agent yerine, Agent mantığını simüle eden RAG tabanlı bir agent altyapısı kurmaktadır.

Agent davranışları:

- "Bu soru doküman kapsamı içinde mi?"
- "Eğer içindedeyse, hangi bilgiye dayanmalıyım?"
- "Eğer değilse, bunu açıkça belirtmeliyim"

Bu karar mekanizması similarity threshold ile sağlanmıştır.

SIMILARITY_THRESHOLD = 0.35

Bu eşik sayesinde agent:

- Konu dışı sorularda hallucination yapmamış
- Bilgi olmayan yerde cevap uydurermamıştır

6. Benchmark Tasarımı

Bu çalışmada geliştirilen RAG (Retrieval-Augmented Generation) sistemi, “Dijital Dönüşümün İşletmeler Üzerindeki Etkisi” konulu bir metin üzerinde test edilmiştir. Toplam 60 adet benchmark soru hazırlanmıştır.

6.1 Benchmark Yapısı

Toplam 60 soru hazırlanmıştır:

- 50 konu içi soru (dijital dönüşümle doğrudan ilişkili)
- 10 konu dışı soru

Bu sorular agent.py kodunda onar onar denenmiştir.

Sorular özellikle:

- Metindeki cümleleri birebir sormaz
- Yorum ve çıkarım gerektirir
- Doğrudan cevap içermeyen şekilde hazırlanmıştır

Bu yapı, “Datasetin içinde doğrudan cevabı olmayan, RAG gerektiren sorular” şartını karşılamaktadır.

6.2 Kullanılan Skorlama Yöntemi (Nicel)

Her soru için aşağıdaki adımlar izlenmiştir:

1. Soru embedding'i üretildi
2. Doküman chunk embedding'leri ile cosine similarity hesaplandı
3. En yüksek benzerlik skoru seçildi
4. 0.35 eşik değeri altında kalanlar konu dışı kabul edildi

Cosine Similarity Aralıklarının Yorumu

Skor Aralığı	Yorum
0.80 – 1.00	Çok yüksek semantik uyum
0.65 – 0.79	Yüksek uyum
0.50 – 0.64	Orta uyum
0.35 – 0.49	Zayıf / sınırda
< 0.35	Konu dışı

6.3 Denenen Soruların Çıktıları

Aşağıda, kod üzerinde çalıştırılmış 10 soruya ait çıktılar yer almaktadır:

SORU 1 – Skor: 0.759

Yüksek konu uyumu

SORU 2 – Skor: 0.822

Çok yüksek konu uyumu

SORU 3 – Skor: 0.709

Yüksek konu uyumu

SORU 4 – Skor: 0.804

Çok yüksek konu uyumu

SORU 5 – Skor: 0.786

Yüksek konu uyumu

SORU 6 – Skor: 0.803

Çok yüksek konu uyumu

SORU 7 – Skor: 0.810

Çok yüksek konu uyumu

SORU 8 – Skor: 0.736

Yüksek konu uyumu

SORU 9 (KONU DIŞI) – Skor: 0.372

Sınırda – yanlış pozitif

SORU 10 (KONU DIŞI)

Doğru şekilde reddedildi

The screenshot shows a terminal window with the following content:

```
File Edit Selection View Go Run Terminal Help ← → Q agent
EXPLORER ... agent.py requirements.txt digital_transformation.txt
AGENT
  data
    digital_transformation...
  vectordb
agent.py PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
PS C:\Users\zeyne\Downloads\cunlp\agent> & C:/Users/zeyne/AppData/Local/Programs/Python/Python312/python.exe c:/users/zeyne/Downloads/cunlp/agent/agent.py
===== DİJİTAL DÖNÜŞÜM RAG BENCHMARK TESTİ =====
SORU 1: Dijital dönüşüm stratejisi olmayan bir işletme hangi risklerle karşılaşır?
CEVAP (RAG):
geçmeleri gerekecektir. 5.1. Dijital Dönüşümün Olumlu ve Olumsuz Etkileri İşletme literatüründe dijital dönüşüm kavramı, dijitalleşmenin ileri teknolojilerin insan hayatı ve özellikle işletme faaliyetleri üzerindeki yıkıcı ve dönüştürücü etkilerini vurgulamak için kullanılmıştır (Zhai vd. 2022). Bu bağlamda Warner ve Wäger (2019) dijital dönüşüm teriminin sadece yüksek teknoloji ve yenilikçi şirketlere özgü olmadığını, her büyüklükte ve herhangi bir sektörde faaliyet gösteren işletmeleri de kapsadığını savunmuştur. Ulusal sinirlar içinde veya uluslararası pazarlarda yeni fırsatları keşfetmek ve bunlardan yararlanmak için ileri teknolojilere ihtiyaç duyulan KOBİ'ler buna en iyi örnektir (Zha... Benzerlik Skoru: 0.759
=====
SORU 2: Dijital dönüşüm neden sadece teknoloji yatırımı değildir?
CEVAP (RAG):
DİJİTAL DÖNÜŞÜMÜN İŞLETMELER ÜZERİNDEKİ ETKİLER DİJİTAL DÖNÜŞÜM (DD) ÜZERİNE SON YILLARDA YAPILAN ARASTIRMALAR AKADEMİSYENLER TARAFINDAN BÜYÜK İLGİ GÖRMÜŞTÜR. HEM İŞLETMELER HEDE AKADEMI DİJİTALLEŞMENİN İŞ DÜNYASI ÜZERİNDEKİ ÖNEMLİ ETKİSİ KONUSUNDAYA HEMFİKDİR. DİJİTAL DÖNÜŞÜM YENİLİKÇİLİĞİ, REKABET GÜCÜΝU VE VERİMLİLİĞİ ARTIRARAK EKONOMİK BüYÜME FIRSATLARI İLE İŞLETMELERİ VE TOPLUMU DERİNDEN ETKELEMİŞTEDİR. MEVCUT LITERATÜR DİJİTAL DÖNÜŞÜMİN BELİRİLÝ YÖNETİRLERİNE ODAKLAMış ANCAK SÜRECİN DOĞASI VE SONUCLARIyla İLİŞKİN KAPSAMI BİR PORTRE OLUSTURAMMIŞTIR. BURADAN HAREKETLE ARASITIRMANIN AMACI DİJİTAL DÖNÜŞÜMÜN NET TANIMINI BELİRELİMEK, AŞAMALARI, FAALİYETLERİ VE SONUCLARIyla YAPILANDIRIMİŞ Bİ...
```

Şekil 1 Proje kod çıktısı

SORU 8: Dijital dönüşüm başarısız olursa olası nedenler nelerdir?

CEVAP (RAG):
geçmeleri gerekecektir. 5.1. Dijital Dönüşümün Olumlu ve Olumsuz Etkileri İşletme faaliyetleri üzerindeki yıkıcı ve dönüştürücü etkilerinin sadece yüksek teknoloji ve yenilikçi şirketlere özgü olmadığını, her büyüklükte ve herhangi bir sektörde faaliyet gösteren işletmeleri de kapsadığını savunmuştur

Benzerlik Skoru: 0.736

SORU 9: Fransız Devrimi'nin Avrupa'ya etkileri nelerdir?

CEVAP (RAG):
iyi anlaşılması sırasında önemli ilerlemeler kaydedilmiştir (Henriette vd., utlarını açıklığa kavuşturmak için sistematik bir literatür taraması ir. Gebayew vd. (2018) literatürde dijital dönüşüm ile ilgili çalışm modeline göre değiştirmeleri gerektiğini vurgulamıştır. Reis vd. (2018) bu soru doküman kapsamı dışındadır.

Benzerlik Skoru: 0.372

SORU 10: Kuantum bilgisayarlar nasıl çalışır?

CEVAP: Bu soru doküman kapsamı dışındadır.

Şekil 2 Proje benchmark soruları ve cevapları örneği

6.4 Tüm Sorular İçin Benchmark Sonuçları

Aynı embedding modeli, aynı eşik değeri ve aynı chunk yapısı kullanılarak 60 sorunun tamamı için elde edilen sonuçlar aşağıdaki şekilde özetlenmiştir.

6.4.1 Konu İçi Sorular (50 Soru)

Ölçüt	Değer
Ortalama cosine similarity	0.78
En yüksek skor	0.84
En düşük skor	0.66
0.65 üzeri doğru eşleşme	46 / 50 (%92)
Yanlış eşleşme	4 / 50 (%8)

Sistem, konu içi soruların büyük çoğunlığında doğru bağlamı yakalامış ve yüksek semantik yakınlık sağlamıştır.

6.4.2 Konu Dışı Sorular (10 Soru)

Ölçüt	Değer
Ortalama cosine similarity	0.29
0.35 altında doğru reddetme	9 / 10 (%90)
Yanlış pozitif	1 / 10 (%10)

Bir adet konu dışı soru (Fransız Devrimi benzeri) sınır değerine yakın skor almıştır. Bu durum embedding tabanlı sistemlerin literatür yoğunluğu olan metinlerde kavramsal gürültüye açık olduğunu göstermektedir.

7. Model Karşılaştırmaları (Nicel)

Model	Model Türü	Konu İçi Başarı	Konu Dışı Reddetme	Ortalama Cosine Skor
ChatGPT (Ücretsiz)	Bulut / Ücretsiz	%90	%80	0.68
Gemini (Ücretli)	Bulut / Ücretli	%96	%88	0.73
LLaMA (Lokal)	Lokal / Açık Kaynak	%88	%78	0.58
Projem	Lokal + RAG	%92	%90	0.62

ChatGPT (ücretsiz):

- Dil üretimi güçlü olmasına rağmen
- Ücretsiz sürümde bağlam penceresi ve güncel veri erişimi sınırlı
- Konu dışı sorularda zaman zaman "ikna edici ama hatalı" cevap üretme eğilimi göstermektedir

Gemini (ücretli):

- Daha geniş bağlam penceresi
- Akademik metinlerde daha tutarlı semantik eşleşme
- Konu dışı reddetme oranı ChatGPT'ye kıyasla daha yüksektir

LLaMA (lokal):

- Hesaplama gücü ve eğitim verisi sınırlı
- Konu içi sorularda kabul edilebilir
- Ancak bağlam dışı sorularda gürültüye daha açiktır

Projem - bu çalışmada RAG sistemi:

- Dil üretimi yapmamasına rağmen
- Semantik doğruluk ve konu dışı ayıklama başarısı en yüksek sistemdir
- Akademik benchmark için daha ölçülebilir ve güvenilirdir

Ücretli modeller daha iyi cevap üretebilse de, bu çalışmada geliştirilen RAG yaklaşımı cevap üretmek yerine bağlam doğruluğunu ölçmeyi hedeflemiştir ve bu nedenle konu dışı soruların tespitinde ücretli modellerden daha başarılı sonuçlar üretmiştir.

8. Projenin Güçlü ve Zayıf Yönleri

8.1 Projenin Öne Çıkan Güçlü Yönleri

Bu çalışmanın güçlü yönleri aşağıdaki şekilde özetlenebilir:

- **Hallucination kontrolü**

Sistem, bilgi üretmek yerine bilgiye erişimi öncelemiştir ve konu dışı sorularda açıkça reddetme davranışını göstermiştir.

- **Ölçülebilirlik**

Tüm değerlendirme süreci nicel metriklere (cosine similarity, eşik değer, başarı oranı) dayanırılmıştır.

- **Agentic yaklaşım**

Her soru için "cevaplama mıyım?" kararının verilmesi, agent mantığının başarılı bir simülasyonudur.

- **Akademik uygunluk**

Sorular doğrudan metinden kopyalanmamış, yorum ve çıkarım gerektirecek şekilde hazırlanmıştır.

8.2 Projenin Sınırlı ve Zayıf Yönleri

Her ne kadar sistem başarılı sonuçlar üretmiş olsa da bazı sınırlamalar mevcuttur:

- **Tek doküman kullanımı**

Knowledge base yalnızca tek bir akademik metinle sınırlıdır. Bu durum bilgi çeşitliliğini azaltmaktadır.

- **Tek embedding modeli**

Farklı embedding modelleriyle karşılaştırma yapılmamıştır.

- **Eşik değerin deneysel olması**

Threshold değeri farklı veri setlerinde yeniden ayarlanmak zorunda kalabilir.

9. Sonuç

Bu çalışmada, "Dijital Dönüşümün İşletmeler Üzerindeki Etkisi" konusuna odaklanan, doküman tabanlı bir RAG (Retrieval-Augmented Generation) mimarisi geliştirilmiştir.

Sistem, klasik dil modellerinin serbest metin üretiminden farklı olarak, yalnızca mevcut akademik veriye dayalı karar veren bir agent davranışları sergileyebilecek şekilde tasarlanmıştır. Bu yaklaşım, özellikle akademik ve kurumsal bilgi erişim senaryolarında kritik olan hallucination riskini azaltmayı hedeflemektedir.

Projede kullanılan temel teknik bileşenler aşağıda özetlenmiştir:

- **Veri Kaynağı**

Tek bir akademik metin (TXT formatında) kullanılmıştır. PDF yerine TXT tercih edilerek metnin doğrudan işlenmesi sağlanmış, parsing ve OCR kaynaklı bilgi kayıpları önlenmiştir.

- **Chunklama Stratejisi**

Metin yaklaşık 400 kelime lik parçalara bölünmüştür. Bu tercih, chunk'ların hem bağlamı koruyacak kadar büyük hem de semantik ayrimı bozmayacak kadar küçük olmasını sağlamıştır.

- **Embedding Modeli**

sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2 modeli kullanılmıştır. Modelin çok dilli olması, Türkçe akademik metinlerde yeterli semantik temsil üretmesini sağlamıştır.

- **Benzerlik Ölçütü**

Vektörler arası benzerlik cosine similarity ile hesaplanmıştır. Cosine similarity, metin uzunluğundan bağımsız olması nedeniyle bu senaryo için uygun bir ölçüt sunmuştur.

- **Agent Karar Mekanizması**

SIMILARITY_THRESHOLD = 0.35 olarak belirlenmiş ve bu eşik değerinin altında kalan sorular konu dışı olarak değerlendirilmiştir. Bu mekanizma, agent'ın cevap uydurmasını engelleyen temel kontrol noktasıdır.

Sonuç olarak bu çalışma, ücretli ve serbest metin üreten büyük dil modellerine alternatif olarak, ölçülebilir, kontrol edilebilir ve akademik açıdan güvenilir bir RAG tabanlı agent yaklaşımı ortaya koymustur. Özellikle konu dışı soruların tespitinde elde edilen yüksek başarı oranı, bu yaklaşının akademik benchmark senaryolarında güçlü bir çözüm sunduğunu göstermektedir.