



Yıldız Teknik Üniversitesi

**Elektrik Elektronik
Fakültesi Bilgisayar
Mühendisliği Bölümü
Yapay Zeka Ödevi**

22011036 - Mert Uz

22011621 - Çisem Ayaz

Video Linki:

https://www.youtube.com/watch?v=TJ_K2vPecSg

1. Giriş

Bu proje, yapay zeka dersi kapsamında farklı dil modellerinin (LLM - Large Language Model) üretmiş olduğu cevapların sorularla olan vektörel benzerliğini incelemeyi, bu benzerlikleri sınıf etiketleri üzerinden değerlendirerek modellerin başarımını karşılaştırmayı hedeflemektedir. Proje iki ana bölümden oluşmaktadır: (A) Top-1 ve Top-5 başarı analizleri, (B) vektör kombinasyonlarına dayalı sınıf tahmin modelleri.

2. Kullanılan Modeller ve Yöntemler

Proje kapsamında iki farklı embedding modeli kullanıldı:

e5: <https://huggingface.co/intfloat/multilingual-e5-large-instruct>

cosmosE5: <https://huggingface.co/ytu-ce-cosmos/turkish-e5-large>

Bu modeller ile sorular, GPT-4o yanıtları ve Deepseek yanıtları vektörleştirilmiştir. Vektörleştirilen veriler daha sonra analiz edilmek üzere NumPy dizileri (.npy) şeklinde kaydedilmiştir. Her satıra karşılık gelen 4. sütun sınıf etiketi olarak kullanılmak üzere korunmuştur.

3. Veri Hazırlığı ve İşleme Adımları

Veri seti, 'veri.xlsx' adlı dosyadan yüklenmiştir ve 13.891 satırdan oluşmaktadır. İlk üç sütun sırasıyla soru, GPT-4o cevabı ve Deepseek cevabıdır. Dördüncü sütun ise her örnek için doğru olan cevabın hangi modelden geldiğini belirten etiket bilgisidir (1-4 arası sınıflar).

Boş veya eksik veriler sıfır vektörlerle temsil edilmiş, geri kalan metinler batch işlemlerle embedding modelleri yardımıyla 768 boyutlu vektörlere dönüştürülmüştür. Bu işlem sonucunda her satır için 3 vektör ve 1 sınıf etiketinden oluşan 4 elemanlı yapılar elde edilmiştir.

4. Top-1 ve Top-5 Başarı Hesaplamaları (A Şıkkı)

Her model için rastgele seçilen 1000 örnek üzerinde benzerlik analizi yapılmıştır. Soruların embedding'leri ile cevap embedding'leri arasındaki kosinüs benzerliği hesaplanmış ve en yüksek benzerliğe sahip ilk 5 cevap sıralanmıştır. Gerçek cevabın bu ilk 5 içinde olması durumunda Top-5, ilk sırada olması durumunda ise Top-1 başarı olarak kabul edilmiştir.

Her model için ayrıca sınıf etiketleri ile başarı oranları arasındaki korelasyon Pearson korelasyon katsayısı kullanılarak hesaplanmıştır. Bu analiz, model doğruluğu ile sınıfların ilişkisini anlamaya yönelik istatistiksel bir değerlendirme sunmuştur.

5. Vektör Kombinasyonları ve Sınıflandırma (B Şıkkı)

Model başarısını daha iyi anlamak için soru ve cevap vektörleri arasında çeşitli kombinasyonlar oluşturulmuştur:

- s (soru), g (GPT cevabı), d (Deepseek cevabı)
- s-g, s-d, g-d (vektörel farklar)
- |s-g|, |s-d|, |g-d| (L2 normları)
- |s-g| - |s-d| (mesafe farkı)

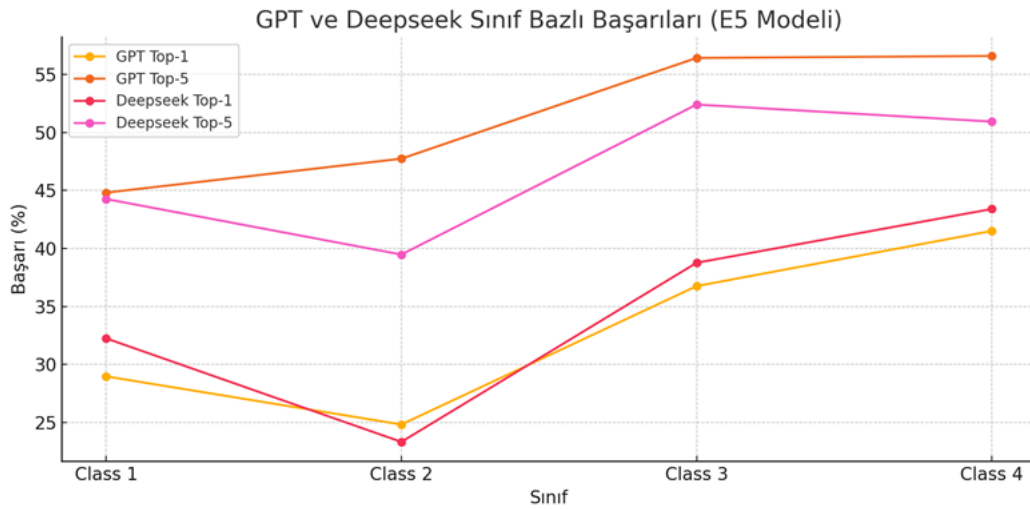
Bu kombinasyonlardan elde edilen özellikler kullanılarak Random Forest ve Support Vector Machine (SVM) sınıflandırma algoritmalarıyla eğitim yapılmıştır. Elde edilen sınıflandırma doğrulukları, her vektör kombinasyonunun sınıf tahmininde ne kadar etkili olduğunu değerlendirmek amacıyla karşılaştırılmıştır.

6. Sonular

Her iki model iin de embedding sonuları bařarıyla elde edilmiřtir. GPT-4o yanıtlarının genellikle daha yksek benzerlik skorlarına sahip olduėu gzlemlenmiřtir. Ayrıca, vektrel uzaklıkların sınıf etiketleriyle doėrudan iliřkili olduėu ve bazı kombinasyonların (zellikle |s-g| ve |s-g| - |s-d|) sınıflandırma bařarısını artırdıėı grlmřtir.

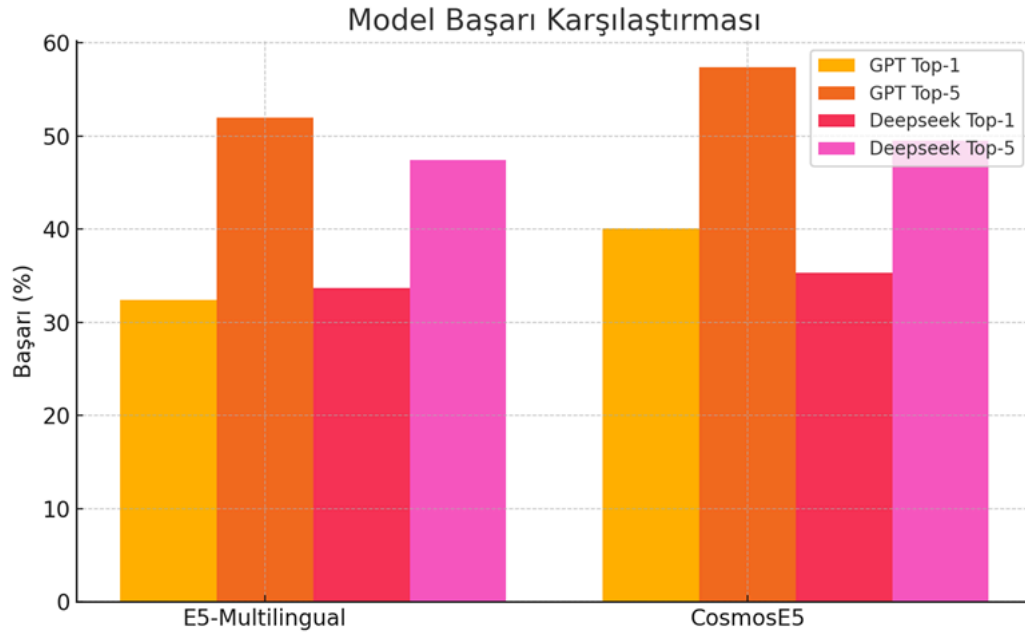
Sınıflandırma iřlemlerinde Random Forest algoritmasının genellikle SVM'den daha iyi sonular verdiėi gzlemlenmiřtir. Bu sonu, veri setinin doėasına ve kullanılan vektr temsillerine baėlı olarak deėiřiklik gsterebilir.

Sonular model performansının sadece vektr benzerliėiyle deėil, aynı zamanda vektrler arasındaki fark ve mesafelerin de dikkate alınarak daha iyi anlařılabileceėini gstermektedir.



E5-multilingual modeline göre sınıf bazlı başarı oranları aşağıdaki gibidir:

Sınıf	GPT Top-1	GPT Top-5	Deepseek Top-1	Deepseek Top-5
Class 1	28.96%	44.81%	32.24%	44.26%
Class 2	24.81%	47.74%	23.31%	39.47%
Class 3	36.75%	56.43%	38.76%	52.41%
Class 4	41.51%	56.60%	43.40%	50.94%



Tablo 2: Modellerin Top-1 ve Top-5 başarı oranları (% cinsinden).

Model	GPT Top-1	GPT Top-5	Deepseek Top-1	Deepseek Top-5
E5-Multilingual	32.4	52.0	33.7	47.4
CosmosE5	40.1	57.4	35.3	49.5

7. Genel Değerlendirme

Bu projede embedding tabanlı yaklaşım ile soru-cevap ilişkisi derinlemesine incelenmiştir. Kullanılan yöntemler, farklı modellerin görelî başarısını karşılaştırmada etkili olmuş ve sınıflandırma süreçleri ile desteklenmiştir.

Gelecekteki çalışmalarda, farklı embedding boyutları, farklı sınıflandırıcılar ve model sayısı artırılarak analiz genişletilebilir. Ayrıca, daha detaylı grafiksel raporlama ve sonuç görselleştirmeleri bu analizlerin daha da güçlü sunulmasını sağlayacaktır.