



YAPAY ZEKA 2. ÖDEV MODELLER

Mert Tuna Kurnaz
20011031

mert.tuna.kurnaz@std.yildiz.edu.tr

Büşra Medine Gural
20011038

medine.gural@std.yildiz.edu.tr

1. GİRİŞ

Bu raporda, metinler arası benzerlik ölçümlerini inceleyen bir analiz sunulmaktadır. Verilen excel dosyası; soru, insan cevabı ve makine cevabından oluşan verileri içermektedir. Veri setinden rastgele seçilen 1000 örnekte, her bir soru için en benzer insan ve makine cevaplarını bulmak amacıyla analiz yapılmış ve TOP1 ile TOP5 başarıları hesaplanmıştır. Benzer şekilde, rastgele 1000 insan cevabı ve 1000 makine cevabı üzerinde, her biri için en benzer soruları bulmak ve gerçek soruya göre TOP1 ve TOP5 başarılarını belirlemek için analizler gerçekleştirilmiştir. Analiz sırasında, metin temsilleri için BERT ve GPT modelleri kullanılarak, bu modellerin benzerlik ölçümleri üzerindeki etkisi değerlendirilmektedir. Son olarak, bu temsiller üzerinde t-SNE uygulanarak metinlerin iki boyutlu bir görselleştirilmesi sunulmaktadır.

2. SONUÇLAR VE YORUMLAR

Aşağıdaki çalışma sürelerini içeren sonuçlar, verilerin embedding hesaplamaları haricindeki toplam çalışma zamanını içermektedir. Embedding işlemi dahil edildiğinde GPT2'nin çalışma süresi, BERT'in yaklaşık 3 katı kadardır. Ek olarak, makine cevaplarının embedding hesaplama süresi, insan cevapları ve soruların embedding hesaplama süresinin yaklaşık 2 katı kadardır.

2.1. Rastgele 1000 soru için seçilen insan cevaplarının başarı durumu:

	TOP1	TOP5	ÇALIŞMA SÜRESİ
BERT	358	512	303.6631
GPT2	458	662	304.8876

2.2. Rastgele 1000 soru için seçilen makine cevaplarının başarı durumu:

	TOP1	TOP5	ÇALIŞMA SÜRESİ
BERT	358	547	302.9873
GPT2	395	640	306.3967

2.3. Rastgele 1000 insan cevabı için seçilen soruların başarı durumu:

	TOP1	TOP5	ÇALIŞMA SÜRESİ
BERT	278	432	302.8761
GPT2	463	683	304.1093

2.4. Rastgele 1000 makine cevabı için seçilen soruların başarı durumu:

	TOP1	TOP5	ÇALIŞMA SÜRESİ
BERT	195	279	302.9018
GPT2	483	740	304.8047

Tablodaki sonuçlardan yola çıkarak GPT2'nin TOP1 ve TOP5 başarı oranlarının BERT'e göre daha yüksek olduğu görülmektedir. Bu, GPT2'nin doğru cevapları daha sık bir şekilde ve ilk denemede bulduğu anlamlarına gelmektedir. Her iki model için çalışma süresi neredeyse aynıdır. Eğer doğruluk önemli bir etken ise, GPT daha iyi bir seçenek olacaktır. Eğer süre önemli bir etken ise her iki model de benzer çalışma süresine sahip olduğundan doğruluğun daha başarılı olduğu GPT2 tercih edilebilir.

Örnek – GPT2: Makine cevabından sorunun tahmin edilmesi.

TOP1 = 1

TOP5 = 1

```
Response75: İstanbul'dan Ankara'ya gitmek için tren, otobüs ve uçak seçenekleri bulunmaktadır. Bu seçeneklerden en hızlısı uçaktır.
True Question: İstanbul'dan Ankara'ya en hızlı nasıl gidebilirim?
1 : İstanbul'dan Ankara'ya en hızlı nasıl gidebilirim?
2 : İstanbul ile Londra arası kaç km'dir?
3 : A ve B kentleri arası 720 km'dir. A kentinden B kentine 9 saatte giden aracın hızı saatte kaç km'dir?
4 : Bir otobüs durağına günde 5 otobüs gelmektedir. 1 haftada kaç otobüs gelir?
5 : Ankara'da toplam kaç tane teknopark vardır?
```

Örnek– GPT2: İnsan cevabından sorunun tahmin edilmesi.

TOP1 = 1

TOP5 = 1

```
Response93: İnsan kaslarının bazı ilaçlarla güçlendirilebileceği, ancak bu ilaçların dikkatlice kullanılması ve bir doktorun gözetiminde olması gerektiği bilinmektedir.
True Question: İnsan kasları bazı ilaçlar ile güçlendirilebilir mi
1 : İnsan kasları bazı ilaçlar ile güçlendirilebilir mi
2 : İnsan vücudunun temel parçaları nelerdir?
3 : İnsan davranışlarını etkileyen faktörler nelerdir?
4 : Yediğimiz bir yemeğin sindirilme sürecinde gerçekleşen olaylar nelerdir?
5 : İnsanların başparmağı olmasa ne gibi sorunlar olurdu
```

Örnek– GPT2: Sorudan insan cevabının tahmin edilmesi.

TOP1 = 0

TOP5 = 0

```
Question67: Bir ağaçta 7 kuş var ve bir avcı bir kuş vurursa, ağaçta kaç kuş kalır?
True Response: 6
1 : Bir elimizde 5 parmak vardır ama bizim 2 elimiz vardır. Dolayısıyla 12 kişide 120 parmak vardır.
2 : Yılanların bacağı olmadığı için toplam 0 bacak vardır.
3 : Evet yapraksız çiçek bulunmaktadır. Yapraksız çiçekler fotosentez yapamazlar. Örneğin erguvan çiçeği yapraksızdır.
4 : Çam ağaçları gibi bitkiler kışları da yeşil kalabilmektedir.
5 : Evet, yarasalar uçan memelilerdir.
```

Örnek – BERT: Sorudan insan cevabının tahmin edilmesi.

TOP1 = 0

TOP5 = 1

```
Question45: The Elder Scrolls V: Skyrim oyununda karakterimiz hangi liseye gitmektedir ?
True Response: Bahsedilen oyunda lise vb. eğitim kurumları yoktur. Oyun daha çok keşif üzerinedir.
1 : Bir araba ve motosiklet için ortak parçalar şunlar olabilir: Motor,Direksiyon,Hız Ekranı,Egzoz,Tekerlek...
2 : Turp (Transüretral Prostat Rezeksiyonu) en sık kullanılan cerrahi operasyonlardan biridir. İyi huylu prostat büyümelerinin neredeyse tamamı TURP uygulamaları kapsamında tedavi edilmektedir.
3 : Bahsedilen oyunda lise vb. eğitim kurumları yoktur. Oyun daha çok keşif üzerinedir.
4 : Türkiye'de son referandum, 2017 yılında gerçekleştirilmiştir. 16 Nisan 2017 tarihinde yapılan referandumda, Türkiye'de anayasa değişikliği oylaması yapılmış ve seçmenler, yeni anayasa taslağını kabul etme veya reddetme hakkına sahip olmuştur.
5 : Günde 2-3 litre su içmeliyiz
```

Örnek – BERT: Sorudan makine cevabının tahmin edilmesi.

TOP1 = 1

TOP5 = 1

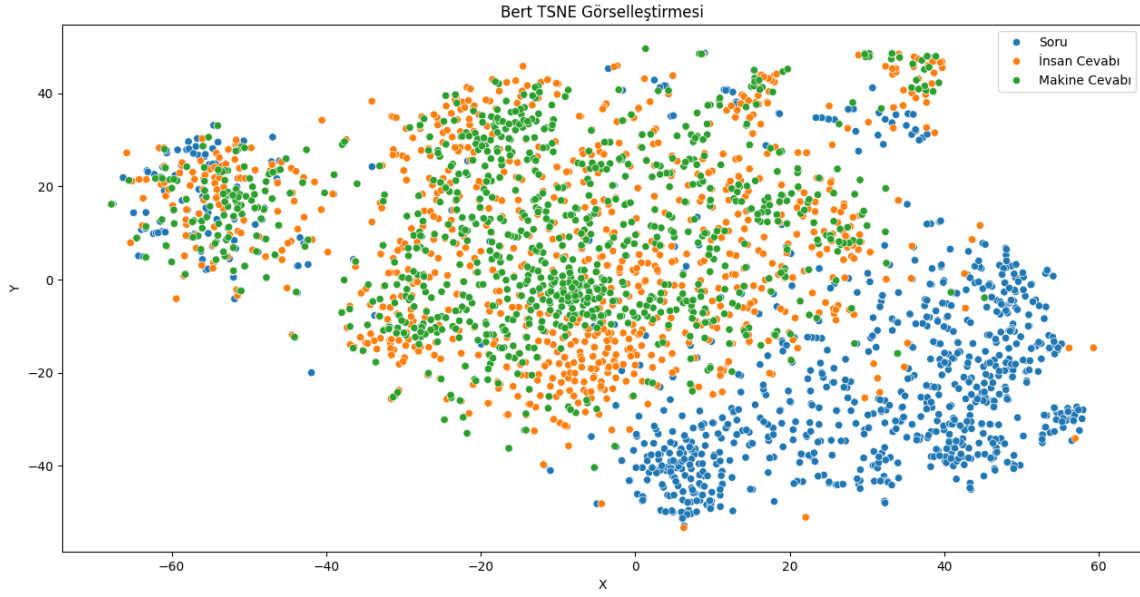
```
Question89: Zaman içerisinde zeytinin rengi neden değişir?
True Response: Zeytinler olgunlaştıkça renkleri yeşil, açık kahverengi ve sonunda koyu siyah veya bordo arasında değişir. Bu renk değişimi, antioksidanlar ve pigmentler olan karoten ve klorofil gibi bileşiklerin oluşumundan kaynaklanır. Olgunlaşma süreci ilerledikçe, klorofil parçalanır ve karoten daha baskın hale gelir. Sonuç olarak, zeytinler olgunlaştıkça yeşil renklerini kaybeder ve daha koyu renklere dönüşür.
1 : Zeytinler olgunlaştıkça renkleri yeşil, açık kahverengi ve sonunda koyu siyah veya bordo arasında değişir. Bu renk değişimi, antioksidanlar ve pigmentler olan karoten ve klorofil gibi bileşiklerin oluşumundan kaynaklanır. Olgunlaşma süreci ilerledikçe, klorofil parçalanır ve karoten daha baskın hale gelir. Sonuç olarak, zeytinler olgunlaştıkça yeşil renklerini kaybeder ve daha koyu renklere dönüşür.
2 : Türkiye'de en çok çay tüketilen iller arasında Rize önemli bir yer tutar.
3 : En etkili leke çıkarma yöntemleri arasında lekenin tipine uygun kimyasallar kullanma ve leke oluşumuna hızlı müdahale bulunur, kalıcı lekeleri önler.
4 : Bir haftada 5 otobüsün geldiği duraktaki otobüs sayısı: 5 otobüs/gün x 7 gün = 35 otobüs.
5 : Kışın yeşil kalan bitki türleri arasında çam ağaçları, çalılar ve bazı yosunlar bulunabilir.
```

Örnek – BERT: İnsan cevabından sorunun tahmin edilmesi.

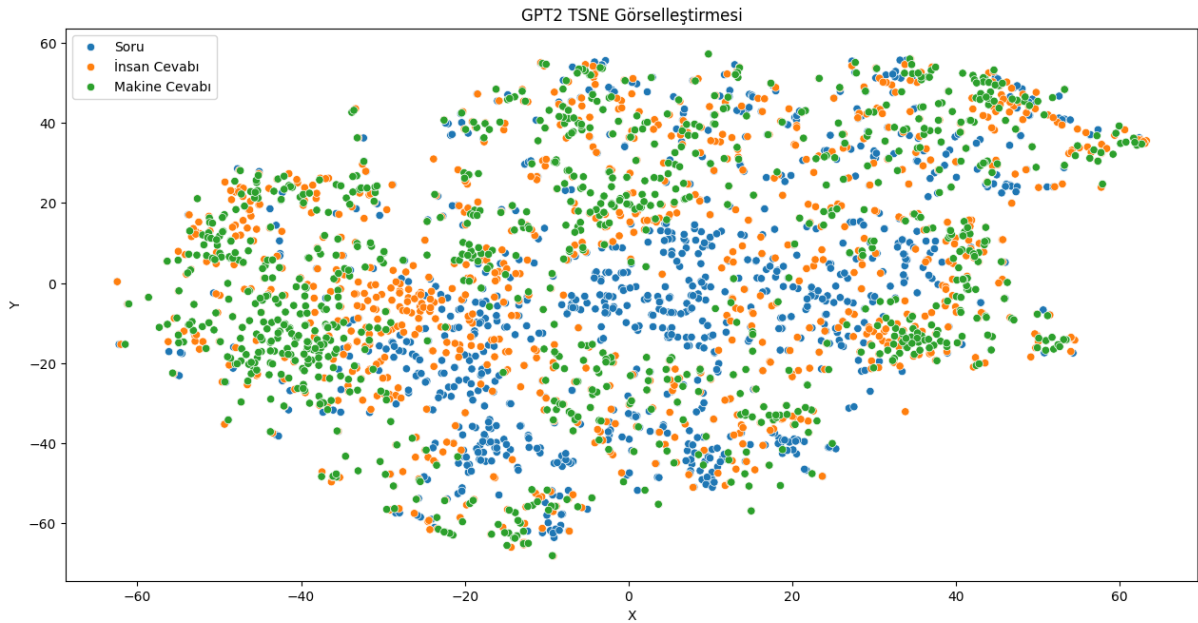
TOP1 = 0

TOP5 = 1

```
Response37: Dünya ile Ay arasındaki mesafe yaklaşık olarak 300.000 km civarındadır. Işığın aydan dünyaya ulaşması yaklaşık 1 saniye sürmektedir.
True Question: Dünya ile Aya arasındaki mesafe ne kadardır?
1 : A ve B kentleri arası 720 km'dir. A kentinden B kentine 9 saatte giden aracın hızı saatte kaç km'dir?
2 : Dünya ile Aya arasındaki mesafe ne kadardır?
3 : Bir su ısıtıcısında suyun kaynama noktası 95 derece olarak ölçülüyor. Bu durumda su ısıtıcısı deniz seviyesine göre nerededir? Nedenini açıklayın.
4 : Bir otobüs durağına günde 5 otobüs gelmektedir. 1 haftada kaç otobüs gelir?
5 : Bir ormandaki tavşanların sayısının iki katı ile yine bu tavşan sayısının 10 fazlasının toplamı 160 dan küçüktür. Bu ormanda en fazla kaç tavşan olabilir?
```



Soruların sağ alta toplanması, bu veri noktasının taşıdığı kelime dağarcığı veya anlam yapısı gibi özelliklerle diğerlerinden farklı bir kümeyi temsil ettiğini göstermektedir. Sorular, genellikle bir tür belirgin yapıya sahiptir veya cevaplardan farklı özelliklere sahiptir. Makine ve insan cevaplarının ortada iç içe geçmiş görünmesi, bu iki tür veri arasında önemli bir benzerlik olduğunu göstermektedir. Bu durum, makine ve insan cevaplarının birbirine benzediğini veya benzer bir dağılıma sahip olduğuna işaret etmektedir. Diğer bir deyişle, makinelerin insan benzeri cevaplar verme yeteneğinin yüksek olduğunu temsil etmektedir.



Grafikten görüleceği üzere, üç farklı veri noktaları arasında kelime dağarcığı açısından önemli benzerlikler vardır. Bu, soruların belirli bir şekilde yanıtlandığını veya hem insan hem de makine cevaplarının benzer dil kalıplarını kullandığını göstermektedir. Buradan makine cevaplarının insan cevaplarıyla yüksek benzerlik gösterdiği çıkarılabilir. BERT ile karşılaştırıldığında daha iyi sonuç verdiği grafikteki kümelenmemiş dağılımda görülmektedir.