

YILDIZ TEKNİK ÜNİVERSİTESİ
Elektrik-Elektronik Fakültesi
Bilgisayar Mühendisliği Bölümü



BLM3510 Yapay Zeka

Prof. Dr. Mehmet Fatih Amasyalı

Öğrenci-1

Bilal Müftüoğlu

Öğrenci-2

Doğukan Baş

İsim:

E-mail:

Öğrenci No:

bilal.muftuoglu1@std.yildiz.edu.tr

20011007

dogukan.bas@std.yildiz.edu.tr

21011003

İçindekiler

1. GPT Model

1.1- GPT Top Accuracy

1.2- GPT t-SNE

2. BERT Model

2.1- BERT Top Accuracy

2.2- BERT t-SNE

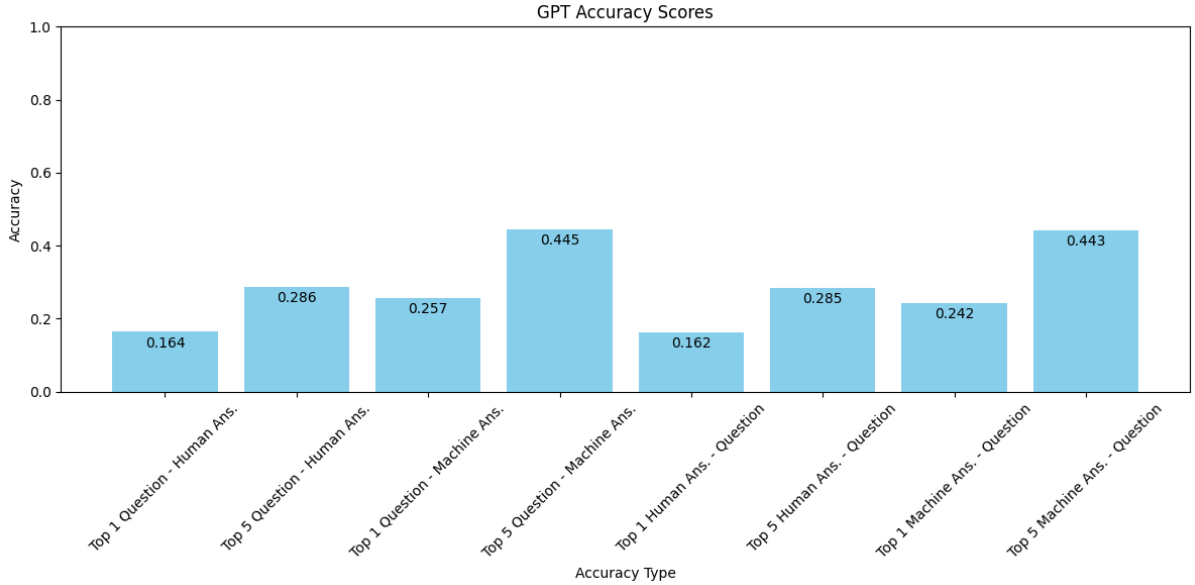
3. Genel Çıkarım ve Karşılaştırma

1- GPT Model

Kod içerisinde “ytu-ce-cosmos/turkish-gpt2” modelini kullandık.

1.1- GPT Top Accuracy Graphic

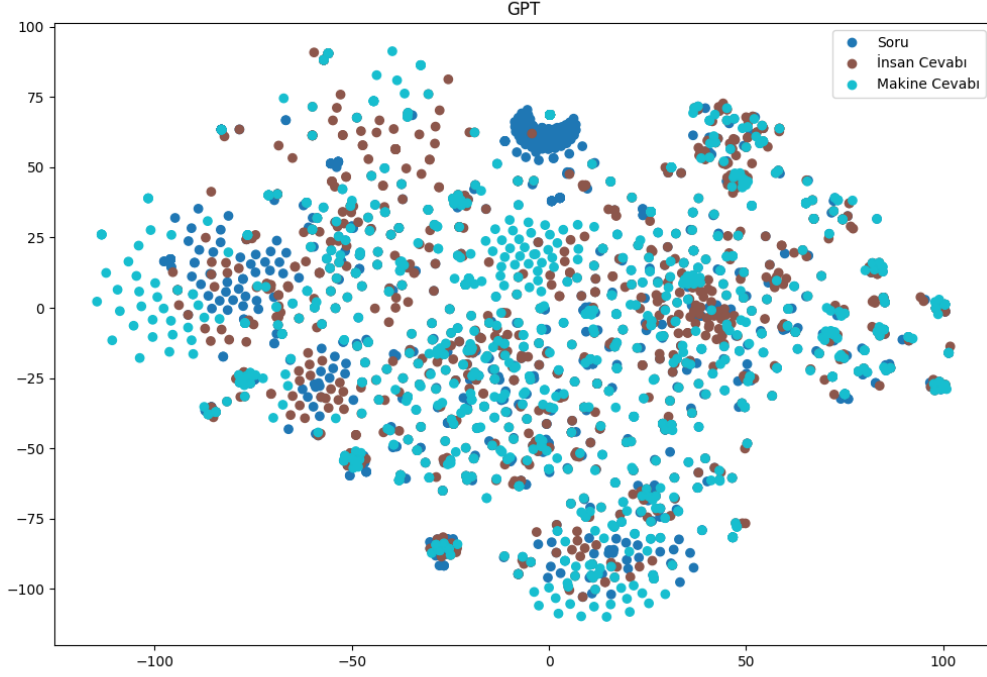
GPT modelinin tokenizer’ını kullanarak oluşturduğumuz vektörleri ve kosinüs benzerlik algoritmasını kullanarak elde ettiğimiz benzerlik oranları ile gerçek değerleri karşılaştırdığımızda, rastgele bir çalıştırmada, Şekil 1’de görülen sonuçları aldık. Farklı çalıştırmalarda elde ettiğimiz değerlerde Top 1 doğruluk skorları için ortalama 0.15-0.25 arası sonuçlar elde ederken, Top 5 doğruluk skorları için ortalama 0.25-0.45 arası sonuçlar elde ettik. Makine cevapları ile sorular arasındaki oranlar insan cevapları ile sorular arasındaki oranlara göre yüksek çıkmıştır. Bunun nedeni yapay zekanın genel olarak soruları tekrar ederek cevaplamaı olabilir.



Şekil 1 - GPT Top Accuracies

1.2- GPT t-SNE Graphic

Verisetinde bulunan soru, insan cevabı ve makine cevabı kategorilerinin temsillerini GPT modelinin tokenizer’ını kullanarak elde ettikten sonra, t-SNE uygulayarak 2 boyutlu bir grafik üzerinde görselleştirilmiş halleri rastgele bir çalıştırma için Şekil 2’de verilmiştir. Soru kategorisindeki bazı elemanların belli bir noktada toplandıkları gözükmemektedir. Bunun nedeni sorulan soruların kelime olarak veya cümle yapısı olarak birbirlerine çok benzemesi olabilir. Aynı şekilde bazı noktalarda insan ve makine cevapları toplanmıştır. Bunun nedeni benzer türdeki sorular için verilen cevaplar olmaları olabilir.



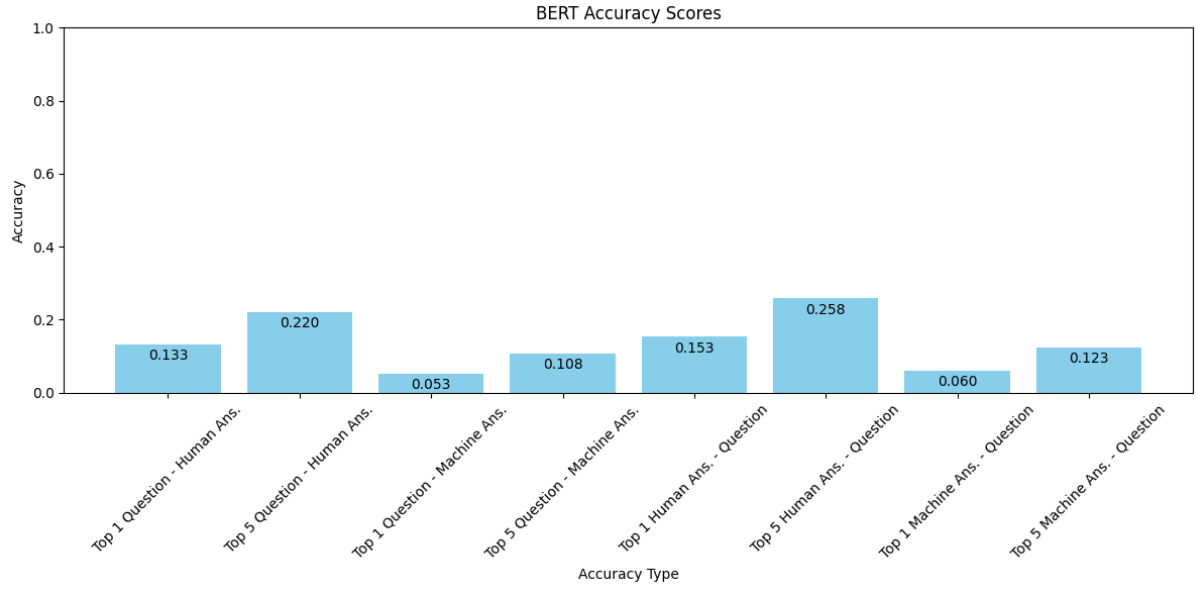
Şekil 2 - GPT t-SNE

2- BERT Model

Kod içerisinde “ytu-ce-cosmos/turkish-medium-bert-uncased” modelini kullandık.

2.1- BERT Top Accuracy Graphic

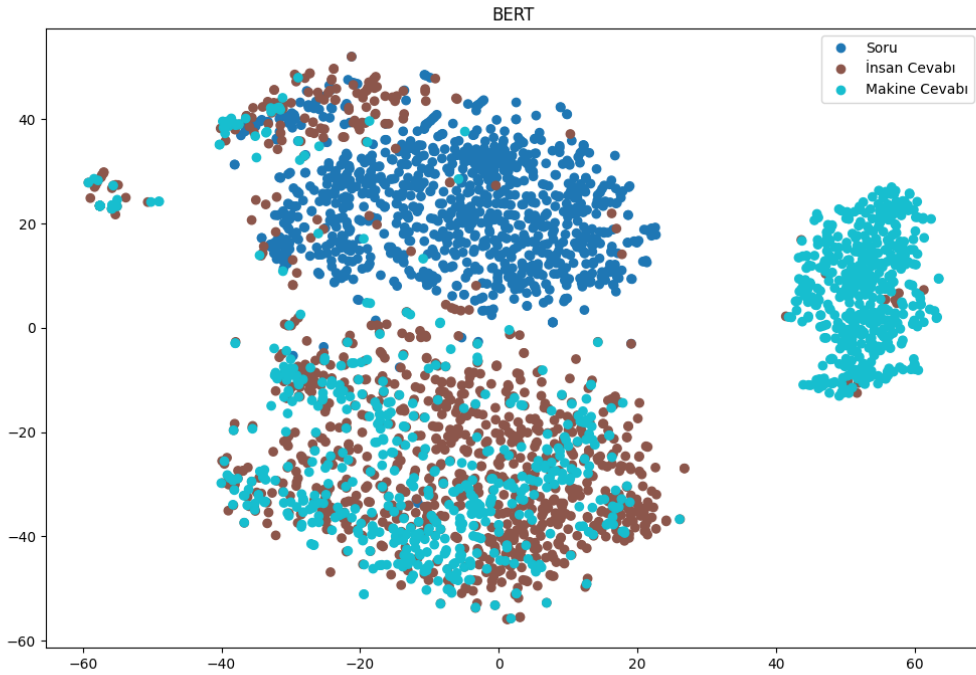
BERT modelinin tokenizer'ını kullanarak oluşturduğumuz vektörleri ve kosinüs benzerlik algoritmasını kullanarak elde ettiğimiz benzerlik oranları ile gerçek değerleri karşılaştırdığımızda, rastgele bir çalıştırmada, Şekil 3'te görülen sonuçları aldık. Farklı çalıştırmalarda elde ettiğimiz değerlerde Top 1 doğruluk skorları için ortalama 0.05-0.15 arası sonuçlar elde ederken, Top 5 doğruluk skorları için ortalama 0.10-0.25 arası sonuçlar elde ettik.



Şekil 3 - BERT Top Accuracies

2.2- Bert t-SNE Graphic

Verisetinde bulunan soru, insan cevabı ve makine cevabı kategorilerinin temsillerini BERT modelinin tokenizer'ını kullanarak elde ettikten sonra, t-SNE uygulayarak 2 boyutlu bir grafik üzerinde görselleştirilmiş halleri rastgele bir çalıştırma için Şekil 4'te verilmiştir. Genel olarak soru ve cevap kategorilerini birbirinden ayırmış gözükmetedir. Makine cevaplarının büyük bir kısmı uzak bir noktada konumlanmıştır. Şekil 3'te gördüğümüz makine cevabı ile soru arasındaki değerlerin, insan cevabı ile soru arasındaki değerlere göre düşük kalmasının nedeni bu olabilir.



Şekil 4 - BERT t-SNE

3- Genel Çıkarım ve Karşılaştırma

GPT ve BERT modellerinin “Top Accuracy” değerlerini kıyasladığımızda GPT modelinin daha başarılı sonuçlar elde ettiğini görüyoruz. Bunun nedeni Şekil 4’te gördüğümüz gibi BERT modelinin soru ve cevapları daha başarılı bir şekilde kümeleştirme yaparak ayırması ve bunun sonucu olarak soru ve cevaplar arasındaki benzerlik oranlarının daha düşük ve yanıltıcı olmasından kaynaklanmış olabilir.

GPT modelinin sorular ile makine cevapları arasında kurduğu ilişki yüksekken, BERT modelinin sorular ile insan cevapları arasında kurduğu ilişki daha yüksek gözükmemektedir. Bunun nedeni modellerin kullandıkları temsil vektörlerinin uzunlukları olabilir. GPT’nin vektör uzunluğu 768 iken BERT’in vektör uzunluğu 512’dir. Makinelerin verdiği cevaplar genel olarak daha uzunken, insan cevapları daha kısadır. Model vektör uzunlukları da metin uzunlukları ile sınırlı olduğundan dolayı GPT modeli makine cevaplarında, BERT modeli ise insan cevaplarında başarı göstermiş olabilir.