**METU NLP BENCHMARK**

**MEVCUT TÜRKÇE VERİ KÜMELERİNİN KALİTE DEĞERLENDİRMESİ**

**v3 (13 Ağustos 2024)**

*Hazırlayanlar:*

*Abdullah Topraksoy (İstanbul Üniversitesi Dilbilimi Bölümü)*

*Ahmet Kaan Sever (Bilkent Üniversitesi Bilgisayar Mühendisliği Bölümü)*

*Ayşe Aysu Cengiz (ODTÜ Bilgisayar Mühendisliği Bölümü)*

*Münire Büşra Tufan (Hacettepe Üniversitesi Sosyoloji Bölümü)*

*Çağrı Toraman (ODTÜ Bilgisayar Mühendisliği Bölümü)*

*Esra Darıcı (ODTÜ Türk Dili Bölümü)*

Bu doküman, ODTÜ Türkçe LLM Leaderboard için belirlediğimiz mevcut veri kümelerinin kalitesinin etiketleyenler tarafından değerlendirilmesi sırasında yol göstermesi amacıyla hazırlanmıştır.

**GÖREV SEÇİMİ**

Aşağıdaki mevcut 9 Türkçe veri kümesi için kalite değerlendirmesi yapılacaktır.

1. MMLU-pro-TR
2. XQuAD
3. XNLI
4. XL-Sum
5. XCOPA
6. WikiANN
7. XTREME (POS)
8. STS-B
9. Belebele

Her veri kümesini değerlendirmek için veri kümesinden rastgele seçilen 100 satırla bir excel tablosu oluşturuldu. Değerlendirmenin nasıl yapılacağının detaylı açıklaması “**Kriterler**” bölümünde, her bir görevin/veri kümesinin detaylı açıklaması ise “**Veri Kümeleri**” bölümünde verilmiştir.

**KRİTERLER**

Tüm değerlendirmeler iki tabanında yapılacak olup kriter sağlanıyorsa 1, sağlanmıyorsa 0 olarak etiketleme bekleniyor.

Kriterin sağlanması için o kriter altındaki tüm alt soruların evet olarak cevaplanması gerekir. En az bir soruya hayır diyorsanız o kriter sağlanmamıştır.

Bu kriterlerin veri kümelerine özel açıklamaları **“Veri Kümeleri”** başlığı altında yer almaktadır.

1. **Doğruluk:** Sorunun cevabı doğru mu? Doğru olarak etiketlenen cevap ile sorulan sorunun gerçek cevabı aynı mı? Bu kriter bazı sorular için araştırma gerektirebilir. Soru doğrudan yanlış olabilir veya cevabın doğruluğu çeviri sırasında kaybolmuş olabilir.
2. **Dil bilgisi:** Metin Türkçe dil bilgisine uygun mu? Yazım kurallarına ve noktalama işaretlerine uyulmuş mu?

Özetleme metinlerinde bulunan yazım kuralları hatalarına örnekler:

* “Filistin topraklarından fırlatılan bir füzenin İsrail'in güneyine isabet etmesinin ardından İsrail, Cumartesi gecesi Gazze Şeridi'ne 12 defa hava saldırısı düzenledi.”
  + Belirli bir tarih bildirmeyen ay ve gün adları küçük harfle yazılır.
  + **Doğrusu:** …İsrail, cumartesi gecesi…
* “Venezuela Devlet Başkanı Nicolas Maduro, yeni anayasayı yazacak Kurucu Meclis'in üyelerini belirleyen seçimler sonrası zaferini ilan etti.”
  + Kurum, kuruluş, kurul, birleşim, oturum ve iş yeri adlarına gelen ekler kesmeyle ayrılmaz.
  + **Doğrusu:** …Kurucu Meclisin üyelerini…

Detaylı bilgi için ["Yazım Kuralları"](https://drive.google.com/file/d/1NJ3-SzXrQIRiFoLRw8oHc1qrzoeqH8HN/view?usp=drive_link) dokümanına bakabilirsiniz.

1. **Bağdaşıklık ve Tutarlılık (Cohesion and Coherence):** Metni oluşturan sözcük ve cümleler arasındaki dil bilgisi kuralları uyumlu mu? Cümle ve paragrafların anlam bütünlüğü mantık düzenine göre yapılandırılmış mı? Terimler ve dil yapıları çeviri boyunca tutarlı bir şekilde kullanılıyor mu?

* Bağdaşıklık (Cohesion): “Bir metni oluşturan çeşitli bölümlerin bir dilsel bütünlük sağlayacak biçimde birbirlerine bağlanması” (Onursal, 2003:8) Bağdaşıklık “bir yazının metin olmasını sağlayan, metin içi ilişkileri kuran, dille ilgili özelliklerin tümünü belirten bir kavram” (Günay, 2007: 71) olarak tanımlanabilir.

Bağdaşıklık (doğru örnek):

"Ben dün okula gittim. Okulda arkadaşlarımla görüştüm. Sonra eve geldim."

Bu örnekte, cümleler arasında "ben" zamiri, "dün" ve "sonra" gibi bağlaçlar ve zaman zarfları kullanarak bir dilbilgisel bağlantı kurulmuştur.

Bağdaşıklık (yanlış örnek):

"Ben dün okula gittim. Sonra kitap okumayı seviyorum. Ardından arkadaşlarımla görüştüm."

Bu örnekte, cümleler arasında dilbilgisel bağlantı kurulmaya çalışılmış, ancak "kitap okumayı seviyorum" cümlesi, metnin geri kalanıyla dilbilgisel olarak bağlantılı değildir. Bu cümle, bir genel ifade olarak dururken, diğer cümleler bir hikaye anlatmaktadır.

* Tutarlılık (Coherence): Tutarlılık, “metindeki mantıksal bağlantı” (Subaşı Uzun, 1995: 110); bir metni oluştururken sözcükler arasındaki anlam ve mantık bağıntısı” (Aksan, 1998: 35) olarak tanımlanabilir.

Tutarlılık (doğru örnek):

"Ben dün okula gittim. Okulda arkadaşlarımla görüştüm. Sonra eve geldim ve yemek yedim."

Bu örnekte, cümleler arasında mantıksal bir bağlantı kurulmuştur. Okula gitmek, arkadaşlarla görüşmek ve eve gelmek, ardından yemek yemek, birbiriyle ilgili ve mantıklı bir sıralamadır.

Tutarlılık (yanlış örnek):

"Ben dün okula gittim. Okulda arkadaşlarımla görüştüm. Sonra aniden uzaya gittim ve bir uzaylıyla tanıştım."

Bu örnekte, cümleler arasında mantıksal bir bağlantı kurulmaya çalışılmış, ancak "aniden uzaya gittim ve bir uzaylıyla tanıştım" cümlesi, metnin geri kalanıyla mantıksal olarak bağlantılı değildir. Okulda arkadaşlarla görüştükten sonra uzaya gitmek ve bir uzaylıyla tanışmak, mantıklı bir sıralama değildir.

1. **Anlaşılabilirlik, Akıcılık ve Anlatım Bozukluğu:** Metin doğal ve akıcı bir şekilde okunabiliyor mu? Ana dili Türkçe olan biri bu cümleleri kurar mıydı? Metinde anlatım bozukluğu var mı?

Anlatım Bozukluğuna özetleme metinlerinden örnekler:

* “75 Nobel ödülü sahibi açık mektup yazdı 2017 yılında 75 Nobel ödülü sahibi İranlı yetkililere hitaben kaleme aldıkları bir açık mektupta Ahmadreza Celali'nin derhal serbest bırakılması çağrısı yaptılar.”
  + Cümlede hem noktalama işaretleri eksikliği hem de özne-yüklem uyumsuzluğundan kaynaklanan anlatım bozukluğu var.
  + **Doğrusu:** 75 Nobel ödülü sahibi açık mektup yazdı: 2017 yılında 75 Nobel ödülü sahibi, İranlı yetkililere hitaben kaleme aldıkları bir açık mektupta Ahmadreza Celali'nin derhal serbest bırakılması çağrısı yaptı.
* “Polisin silahsız siyah bir genci vurması nedeniyle 11 gündür protesto ve şiddet olaylarının yaşandığı ABD'nin St Louis bölgesinde, polisin ikinci siyah erkeği öldürmesiyle gerginliğin daha da artmasından korkuluyor.”
* Cümlede tamlama hatasından kaynaklanan anlatım bozukluğu var.
* **Doğrusu:** Polisin silahsız siyahi bir genci vurması nedeniyle 11 gündür protesto ve şiddet olaylarının yaşandığı ABD'nin St Louis bölgesinde, polisin ikinci kez siyahi bir erkeği öldürmesiyle gerginliğin daha da artmasından korkuluyor.

Detaylı bilgi için ["Anlatım Bozuklukları"](https://drive.google.com/file/d/1NHQyZXApRtmwbtvALU4ARt8rvO13amE2/view?usp=drive_link) dokümanına bakabilirsiniz.

1. **Terim Kullanımı:** Teknik ve özel terimler doğru ve yerinde kullanılıyor mu?

Örnekler:

* MMLU-pro-TR veri kümesinde Jovian Planets ibaresi Jüpiter Gezegenleri olarak çevrilmiş fakat Jüpiter Gezegenleri diye bir gezegen sınıfı yok. Jovian Planets Dev gezegenler demek.
* Prime facie duty ibaresi birincil görev diye çevrilmiş fakat felsefede prime facie ilk bakışta ödevler olarak çevrilir.

1. **Türkiye’deki Ortak Kültürel Sağduyu Bilgisine Uygunluk:** Aşağıdaki kriterlerden her ikisi de (**a** ve **b**) doğru ise 1, değilse 0 veriniz.
   1. Bu kriter doğrudan, salt, değişmez, ülkenin konumu, kuralları, yasaları, siyasi işleyiş biçimi ile ilgili verilen bilgileri kapsamaktadır. Belirtilen veri Türkiye’nin sosyal, ekonomik, kültürel, coğrafi kuralları ile ilgili mi?

Örnekler:

* “İngiliz hukukunda geleneğin hukuka dahil edilmesinin önkoşulu nedir? ….” başka bir ülkenin kurallar bütünü ile ilgilidir.
* "4 feet uzunluğunda bir tel parçası kullandı. Tel, 6 inç uzunluğunda parçalara kesildi…Yerel bir benzin istasyonu galon başına 3 dolar fiyatla benzin satıyor" örneklerinde feet, inç, galon metrikleri ülkemizde kullanılmamaktadır.  
  1. Aşağıda belirtilen kriterlerle ilgili yargılar içeriyorsa aşağıdaki kurallara göre yanıtlayınız. Bu kriterler belirlenirken Nguyen vd. (2023), Anacleto vd. (2006), Acharya vd. (2020) kaynaklarından yararlanılmıştır. Bunlar elbette sınıflara, statülere, inançlara, eğitim düzeyine, toplumsal cinsiyete, ırk ve etnik kökenlere göre değişiklik göstermektedir. Bu yüzden amaç kesin bir bilimsel anket sunmak değil, makul varsayımlara ulaşmak olacaktır.
* Yemek ve Yiyecekler

Yemek Saatleri

* + - Sabah: 07:00-09:00; Öğlen 12:00-13:30; Akşam 17:00-21:00: bu yargılar görüldüğünde (1)
    - Bu saatler dışında saatlere rastlandığında (0)
* Öğünlerde Yenilen Yiyecekler
  + Sabah: Yumurta, sosis, sucuk, salam, peynir, zeytin, ekmek, tost, reçel; Öğlen ve Akşam: et, tavuk, balık yemekleri, sulu yemekler görüldüğünde (1)
  + Sabah: Bacon, kruvasan, mantar, fasulye, marmelat; Öğlen ve Akşam club sandviç, fish and chips, tofu vs. görüldüğünde (0)
* Özel Gün Yemekleri
* Bayramlarda, kutlamalarda, düğünlerde yenilen yemeklere dikkat edilebilir. Örneğin yeni yıl kutlamalarında hindi yenilmesi bize daha sonradan gelen geleneklerden bir tanesidir.
* Yemekler
  + Domuz eti gibi ürünler görüldüğünde (0)
* İçecekler
  + Türk kahvesi, rakı, bira, şarap, şalgam, ayran (1)
  + Uzo, noel içecekleri vs. (0)
* Kıyafetler
  + İskoçların giydiği Kilt gibi geleneksel kıyafetler (0)
  + Başörtüsü, şalvar, kaftan, takke, bindallı, yemeni, mayo, etek, elbise, pantolon (1)
* Ritüeller
  + Cenaze, doğum günü, yılbaşı, doğum, sünnet, evlilik, bayramlar gibi ritüellere dikkat edilmeli. Bizim için düğünlerin kalabalık olması, bayramlarda el öpme, diş hediği, kırkı çıkartmak vs. gibi alışkanlıklar ortak kültürel sağduyu bilgisine uygundur (1)
  + Bu gibi ritüellerde şampanya patlatmak, düğünlerde resepsiyon vermek, bayramlarda kiliseye gitmek vs. bizim ortak kültürel sağduyu bilgimize uygun değildir (0)
  + Bu kriteri etiketlerken lütfen bahsedilen ritüellerdeki davranışları dikkatlice düşünerek etiketleyiniz.
* Gelenekler
  + El öpme, harçlık alma vs. (1)
  + Harakiri, ötenazi, diş doldurma, bebek atma vs. (0)
* Davranışlar
  + Sokağa tükürme, trafik kuralları, topluluk içinde gaz çıkartma vs. (0)
  + Büyüklerle saygılı konuşma, bacak bacak üstüne atmama, bahşiş bırakma vs. (1)
* Sporlar
  + Amerikan Futbolu vs. (0)
  + Futbol, Yüzme, Güreş vs. (1)

1. **Yorum:** Değerlendirme sütunları arasında yukarıdaki altı kritere ek olarak bir de “Yorum” sütunu yer almaktadır. Bu sütunun her satırını doldurmanız beklenmemektedir. Neden bir kritere 0 verdiğiniz ya da soruda kriterlerin kapsamadığı bir problemin varlığı gibi durumları not almak ve yorumlarınız varsa iletmek adına bu sütunu kullanabilirsiniz.

**VERİ KÜMELERİ**

1. **MMLU-pro-TR**
   1. **Açıklama:** MMLU-pro-TR, çoktan seçmeli ve bilgi gerektiren sorulardan oluşan bir veri kümesidir. veri kümesinde farklı alanlarda bilgi gerektiren sorular ve her soru için on cevap seçeneği bulunmaktadır. MMLU-Pro adlı İngilizce veri kümesinden Türkçe’ye çevrilmiştir.
   2. **Veri kümesi sütunları:**
      1. **question\_id:** Soruya özel numara
      2. **question:** Soru metni
      3. **options:** On adet cevap şıkkı
      4. **answer:** Doğru cevabın İngilizce alfabede karşılık geldiği harf. (örn: 1. şık → A, 4. şık → D)
      5. **answer\_index:** Doğru cevabın listedeki indisi (indisler 0’dan başlıyor)
      6. **category:** Sorunun gerektirdiği bilginin alanı
      7. **src:** Kaynak
   3. **Değerlendirme sütunları:**
      1. **Doğruluk:** Aşağıdaki iki soruya da cevabınız “evet” ise kutuya 1 yazın, birine bile cevabınız hayır ise 0 yazın.
         1. Doğru cevap şıklarda var mı?
         2. Soru için verilen cevap şıkkı doğru şık mı?
      2. **Dil bilgisi:** Aşağıdaki iki sorunun cevabı da “hayır” ise kutuya 1 yazın, aksi takdirde 0 yazın.
         1. Yazım hatası var mı?
         2. Noktalama hatası var mı?
      3. **Bağdaşıklık ve Tutarlılık:** Aşağıdaki üç sorunun cevabı da “evet” ise kutuya 1 yazın, aksi takdirde 0 yazın.
         1. Metni oluşturan sözcük ve cümleler arasındaki dil bilgisi kuralları uyumlu mu?
         2. Cümle ve paragrafların anlam bütünlüğü mantık düzenine göre yapılandırılmış mı?
         3. Terimler ve dil yapıları çeviri boyunca tutarlı bir şekilde kullanılıyor mu?
      4. **Anlaşılabilirlik, Akıcılık ve Anlatım Bozukluğu:** Aşağıdaki üç sorunun cevabı da “evet” ise kutuya 1 yazın, aksi takdirde 0 yazın.
         1. Metin doğal ve akıcı bir şekilde okunabiliyor mu?
         2. Ana dili Türkçe olan biri bu cümleleri kurar mıydı?
         3. Metin anlatım bozukluğu içermeyen cümlelerden mi oluşuyor?
      5. **Terim kullanımı:** Teknik ve özel terimler doğru ve yerinde kullanılıyor ise 1, kullanılmıyorsa 0 verin. Örnek:

* Jovian Planets ibaresi Jüpiter Gezegenleri olarak çevrilmiş fakat Jüpiter Gezegenleri diye bir gezegen sınıfı yok. Jovian Planets Dev gezegenler demek.
* Prime facie duty ibaresi birincil görev diye çevrilmiş fakat felsefede prime facie ilk bakışta ödevler olarak çevrilir.
  + 1. **Türkiye’deki Ortak Kültürel Sağduyu Bilgisine Uygunluk:** Belirtilen veri, Türkiye’nin sosyal, ekonomik, kültürel, coğrafi kuralları ile ilgili mi? Kriterler kısmındaki koşulları içeriyorsa belirlenen kurallar çerçevesinde mi içeriyor? Cevabınız her ikisine de evet ise 1, hayır ise 0 verin.
    2. **Yorum**

1. **XQuAD** 
   1. **Açıklama:** XQuAD, paragraf anlama üzerine bir veri kümesidir. veri kümesinde Wikipedia’dan alınan bir paragrafla ilgili birkaç soru sorularak cevabın paragrafın içinden bulunması bekleniyor. Paragrafların aslı İngilizce Wikipedia’dan olup hepsi profesyonel çevirmenler tarafından çevrilmiştir. 100 örnek sorunun hepsi farklı paragraflardan alınmıştır.
   2. **Veri kümesi sütunları:**
      1. **id:** Soruya özel numara
      2. **context:** Soru için gerekli olan bilgiyi içeren paragraf(lar)
      3. **question:** “Context” sütununda verilen paragrafla ilgili sorulan soru
      4. **answer\_text:** “question” sütununda yazan sorunun cevabı (bu cevap paragrafın içinde doğrudan yer alıyor olmalı)
      5. **answer\_start:** “question” sütununda yazan sorunun cevabının “context” sütunundaki paragrafın kaçıncı karakterden itibaren başladığını gösteren sayı
      6. **title:** Paragrafın bulunduğu Wikipedia sayfasının başlığı
   3. **Değerlendirme sütunları:**
      1. **Doğruluk:** Aşağıdaki üç sorunun cevabı da “evet” ise kutuya 1 yazın, aksi takdirde 0 yazın.
         1. Sorunun cevabı paragrafta yer alıyor mu?
         2. “answer\_text” sütunundaki cevap doğru ve eksiksiz mi?
         3. Cevap paragraftan tek bir doğru olarak çıkarılabiliyor mu?
      2. **Dil bilgisi:** Aşağıdaki iki sorunun cevabı da “hayır” ise kutuya 1 yazın, aksi takdirde 0 yazın.
         1. Yazım hatası var mı?
         2. Noktalama hatası var mı?
      3. **Bağdaşıklık ve Tutarlılık:** Aşağıdaki üç sorunun cevabı da “evet” ise kutuya 1 yazın, aksi takdirde 0 yazın.
         1. Metni oluşturan sözcük ve cümleler arasındaki dil bilgisi kuralları uyumlu mu?
         2. Cümle ve paragrafların anlam bütünlüğü mantık düzenine göre yapılandırılmış mı?
         3. Terimler ve dil yapıları çeviri boyunca tutarlı bir şekilde kullanılıyor mu?
      4. **Anlaşılabilirlik, Akıcılık ve Anlatım Bozukluğu:** Aşağıdaki beş sorunun cevabı da “evet” ise kutuya 1 yazın, aksi takdirde 0 yazın.
         1. Metin doğal ve akıcı bir şekilde okunabiliyor mu?
         2. Ana dili Türkçe olan biri bu cümleleri kurar mıydı?
         3. Metin anlatım bozukluğu içermeyen cümlelerden mi oluşuyor?
         4. Soru cümlesini okuduğunuzda neyin sorulduğunu anlayabiliyor musunuz?
         5. Paragrafı okuduğunuzda ne anlattığını anlayabiliyor musunuz?
      5. **Terim kullanımı:** Paragraf ve soruda teknik ve özel terimlerin çevirisi doğru yapılmış mı? Bu soruya cevabınız evet ise 1, hayır ise 0 verin. Örnek için bu dokümanın “Kriterler” bölümüne bakabilirsiniz.
      6. **Türkiye’deki Ortak Kültürel Sağduyu Bilgisine Uygunluk:** Belirtilen veri, Türkiye’nin sosyal, ekonomik, kültürel, coğrafi kuralları ile ilgili mi? Kriterler kısmındaki koşulları içeriyorsa belirlenen kurallar çerçevesinde mi içeriyor? Cevabınız her ikisine de evet ise 1, hayır ise 0 verin.
      7. **Yorum**
2. **XNLI** 
   1. **Açıklama:** XNLI, doğal dil çıkarımı üzerine bir veri kümesidir. MNLI veri kümesinin bir kısmının Türkçeye çevrilmesi ile oluşturulmuştur. Bu veri kümesinin her sorusunda birer önerme ve hipotez verilir ve önermeye bakarak hipotezin alakası belirlenir. Bu alaka üç şekilde olabilir:

0 (Entailment): Gerektirme, önermenin doğru olması hipotezin doğru olmasını gerektiriyor.

1 (Neutral): Karar verilemez, önerme ile hipotez arasında bir bağ yok.

2 (Contradiction): Çelişki, önerme ile hipotezin aynı anda doğru olması bir çelişki oluşturuyor.

* 1. **Veri kümesi sütunları:** 
     1. **premise:** Önerme
     2. **hypothesis:** Önermeye bağlı olarak gözden geçirilecek hipotez
     3. **label:** Hipotezin önerme ile alakası
  2. **Değerlendirme sütunları:**
     1. **Doğruluk:** Önerme ve hipotez arasındaki ilişki doğru belirlenmiş ise 1, yanlış belirlenmiş ise 0 verin.
     2. **Dil bilgisi:** Aşağıdaki iki sorunun cevabı da “hayır” ise kutuya 1 yazın, aksi takdirde 0 yazın.
        1. Yazım hatası var mı?
        2. Noktalama hatası var mı?
     3. **Bağdaşıklık ve Tutarlılık:** Aşağıdaki üç sorunun cevabı da “evet” ise kutuya 1 yazın, aksi takdirde 0 yazın.
        1. Metni oluşturan sözcük ve cümleler arasındaki dil bilgisi kuralları uyumlu mu?
        2. Cümle ve paragrafların anlam bütünlüğü mantık düzenine göre yapılandırılmış mı?
        3. Terimler ve dil yapıları çeviri boyunca tutarlı bir şekilde kullanılıyor mu?
     4. **Anlaşılabilirlik, Akıcılık ve Anlatım Bozukluğu:** Aşağıdaki üç sorunun cevabı da “evet” ise kutuya 1 yazın, aksi takdirde 0 yazın.
        1. Metin doğal ve akıcı bir şekilde okunabiliyor mu?
        2. Ana dili Türkçe olan biri bu cümleleri kurar mıydı?
        3. Metin anlatım bozukluğu içermeyen cümlelerden mi oluşuyor?
     5. **Terim kullanımı:** Önerme ve hipotezde teknik ve özel terimlerin çevirisi doğru yapılmış mı? Bu soruya cevabınız evet ise 1, hayır ise 0 verin. Örnek için bu dokümanın “Kriterler” bölümüne bakabilirsiniz.
     6. **Türkiye’deki Ortak Kültürel Sağduyu Bilgisine Uygunluk:** Belirtilen veri, Türkiye’nin sosyal, ekonomik, kültürel, coğrafi kuralları ile ilgili mi? Kriterler kısmındaki koşulları içeriyorsa belirlenen kurallar çerçevesinde mi içeriyor? Cevabınız her ikisine de evet ise 1, hayır ise 0 verin.
     7. **Yorum**

1. **XL-Sum**
   1. **Açıklama:** XL-Sum, özetleme üzerine bir veri kümesidir. veri kümesinin Türkçe alt kümesi, BBC Türkiye ve Dünya’dan alınan haberlerden oluşuyor. Bu haberlerin internet sitesinde kalın punto ile yazılmış ilk birkaç cümlesi ana metinden ayrılıp özet olarak kabul edilmiş. Veri kümesinin bazı alt dil kümelerinin aksine Türkçe alt kümesi insan kontrolünden geçmemiş.
   2. **Veri kümesi sütunları:**
      1. **id:** Soruya özel numara
      2. **text:** Soru için gerekli olan haberi içeren paragraf(lar)
      3. **summary:** “text” sütununda verilen paragrafın özeti
      4. **title:** “text” sütununda yazan haberin başlığı
      5. **url: “**text” sütununda verilen haberin linki
   3. **Değerlendirme sütunları:**
      1. **Doğruluk:** Aşağıdaki ilk soruya cevabınız “evet”, diğer iki soruya cevabınız “hayır” ise 1 verin. Verdiğiniz cevap bir soru için bile farklıysa 0 verin. (Sorular XL-Sum makalesinden alınmıştır.)
         1. “summary” sütununda verilen özet haberin içeriğinin ne hakkında olduğunu doğru aktarıyor mu?
         2. Özet, haberle çelişen bir bilgi içeriyor mu?
         3. Özet, haberden çıkarılamayacak bir bilgi içeriyor mu?
      2. **Dil bilgisi:** Aşağıdaki iki sorunun cevabı da “hayır” ise kutuya 1 yazın, aksi takdirde 0 yazın.
         1. Yazım hatası var mı?
         2. Noktalama hatası var mı?
      3. **Bağdaşıklık ve Tutarlılık:** Aşağıdaki üç sorunun cevabı da “evet” ise kutuya 1 yazın, aksi takdirde 0 yazın.
         1. Metni oluşturan sözcük ve cümleler arasındaki dil bilgisi kuralları uyumlu mu?
         2. Cümle ve paragrafların anlam bütünlüğü mantık düzenine göre yapılandırılmış mı?
         3. Terimler ve dil yapıları tutarlı bir şekilde kullanılıyor mu?
      4. **Anlaşılabilirlik, Akıcılık ve Anlatım Bozukluğu:** Aşağıdaki beş sorunun cevabı da “evet” ise kutuya 1 yazın, aksi takdirde 0 yazın.
         1. Metin doğal ve akıcı bir şekilde okunabiliyor mu?
         2. Ana dili Türkçe olan biri bu cümleleri kurar mıydı?
         3. Metin anlatım bozukluğu içermeyen cümlelerden mi oluşuyor?
         4. Haber metnini okuduğunuzda ne anlatıldığını anlayabiliyor musunuz?
         5. Haber metni “Videoyu izlemek için tıklayın.” gibi internet sitesine ait ibarelerden arındırılmış mı? (“Haberin sonu” ibaresinin metinlerin yaklaşık %50’sinde bulunması sebebiyle bu soruyu cevaplarken bu ibareyi görmezden geliniz)
      5. **Terim kullanımı:** Paragraf ve soruda teknik ve özel terimler doğru kullanılmış mı? Bu soruya cevabınız evet ise 1, hayır ise 0 verin. Örnek için bu belgenin “Kriterler” bölümüne bakabilirsiniz.
      6. **Türkiye’deki Ortak Kültürel Sağduyu Bilgisine Uygunluk:** Belirtilen veri, Türkiye’nin sosyal, ekonomik, kültürel, coğrafi kuralları ile ilgili mi? Kriterler kısmındaki koşulları içeriyorsa belirlenen kurallar çerçevesinde mi içeriyor? Cevabınız her ikisine de evet ise 1, hayır ise 0 verin.
      7. **Yorum**
2. **XCOPA**
   1. **Açıklama:** XCOPA, sağduyuya dayalı akıl yürütme üzerine bir veri kümesidir. Bu veri kümesinde bir cümle verilir ve sorunun etiketine göre devamına en uygun gelecek olan cümleyi verilen iki şık arasından seçilmesi beklenir. İki çeşit etiket var:
      1. Effect (Etki): Bu etikete sahip sorularda verilen cümlenin sonucunda oluşabilecek durumu veren şık doğru cevap olarak kabul ediliyor. Cevapların doğruluğunu ölçerken şu şekilde cümleleri birleştirebilirsiniz: (verilen cümle) Bu yüzden (doğru şık)
      2. Cause (Sebep): Bu etikete sahip sorularda iseverilen cümlede yazan durumun oluşma sebebini veren şık doğru cevap olarak kabul ediliyor. Cevapların doğruluğunu ölçerken şu şekilde cümleleri birleştirebilirsiniz: (verilen cümle) Çünkü (doğru şık)
   2. **Veri kümesi sütunları:**
      1. **idx:** Soruya özel numara
      2. **premise:** Soruda yaşanan olayı belirten cümle.
      3. **question type:** Sorunun tipini belirten etiket. “effect” ya da “cause” olabiliyor.
      4. **choice1/choice2:** Cevap şıkları
      5. **label:** 0 ise cevap “choice1”, 1 ise cevap “choice2” anlamına geliyor.
      6. **changed:** Sorunun daha önce değiştirilip değiştirilmediğini belirtiyor.
   3. **Değerlendirme sütunları:**
      1. **Doğruluk:** Aşağıdaki iki soruya da cevabınız “evet” ise 1 verin, aksi takdirde 0 verin.
         1. Verilen etikete göre (cause/effect) cevap olarak gösterilen şık uyumlu mu? Yani aşağıdaki şekilde cümleleri birleştirdiğinizde mantıklı bir anlam akışı oluşuyor mu?
            1. Effect: (premise) Bu yüzden (cevap)
            2. Cause: (premise) Çünkü (doğru cevap)
         2. İki şık arasından doğru bir cevap seçilebiliyor mu?
      2. **Dil bilgisi:** Aşağıdaki iki sorunun cevabı da “hayır” ise kutuya 1 yazın, aksi takdirde 0 yazın.
         1. Yazım hatası var mı?
         2. Noktalama hatası var mı?
      3. **Bağdaşıklık ve Tutarlılık:** Aşağıdaki üç sorunun cevabı da “evet” ise kutuya 1 yazın, aksi takdirde 0 yazın.
         1. Metni oluşturan sözcük ve cümleler arasındaki dil bilgisi kuralları uyumlu mu?
         2. Cümle ve paragrafların anlam bütünlüğü mantık düzenine göre yapılandırılmış mı?
         3. Terimler ve dil yapıları çeviri boyunca tutarlı bir şekilde kullanılıyor mu?
      4. **Anlaşılabilirlik, Akıcılık ve Anlatım Bozukluğu:** Aşağıdaki dört sorunun cevabı da “evet” ise kutuya 1 yazın, aksi takdirde 0 yazın.
         1. Metin doğal ve akıcı bir şekilde okunabiliyor mu?
         2. Ana dili Türkçe olan biri bu cümleleri kurar mıydı?
         3. Metin anlatım bozukluğu içermeyen cümlelerden mi oluşuyor?
         4. “Premise”, “choice1” ve “choice2” sütunlarındaki cümlelerde anlatılan olayların ne olduğu anlaşılıyor mu?
      5. **Terim kullanımı:** Paragraf ve soruda teknik ve özel terimlerin çevirisi doğru yapılmış mı? Bu soruya cevabınız evet ise 1, hayır ise 0 verin. Örnek için bu dokümanın “Kriterler” bölümüne bakabilirsiniz.
      6. **Türkiye’deki Ortak Kültürel Sağduyu Bilgisine Uygunluk:** Belirtilen veri, Türkiye’nin sosyal, ekonomik, kültürel, coğrafi kuralları ile ilgili mi? Kriterler kısmındaki koşulları içeriyorsa belirlenen kurallar çerçevesinde mi içeriyor? Cevabınız her ikisine de evet ise 1, hayır ise 0 verin.
      7. **Yorum**
3. **WikiANN** 
   1. **Açıklama:**  WikiANN (ya da PAN-X), varlık ismi tanıma üzerine bir veri kümesidir. veri kümesinin Türkçe alt kümesi, Türkçe Wikipedia’dan alınan çeşitli cümleler ya da ibarelerden oluşuyor. Kelime ve noktalama işaretleri bir listenin elemanları olarak ayrılıyor ve bu listenin hangi elemanlarının bir varlığın ismi olup olmadığının, eğer bir varlığın ismi ise ne çeşit bir varlık olduğunun bulunması bekleniyor. Bu veri kümesinde üç çeşit varlık ele alınıyor:
      1. PER: Person (kişi),
      2. ORG: Organization (kuruluş, organizasyon),
      3. LOC: Location (mekan).
   2. **Veri kümesi sütunları:**
      1. **tokens:** Kelime ve noktalama işareti olarak ayrılmış ibareleri içerir.
      2. **ner\_tags:** “tokens” listesindeki sıralamaya karşılık gelecek şekilde aşağıdaki NER Etiketlerine uygun verilen sayılardan oluşan liste. NER Etiketlemeleri:

0: Aşağıdakilerden hiçbiri değil

1: İnsan isminin ilk kelimesi

2: İnsan isminin devamı

3: Kuruluş/organizasyon isminin ilk kelimesi

4: Kuruluş/organizasyon isminin devamı

5: Mekan isminin ilk kelimesi

6: Mekan isminin devamı

* + 1. **langs:** “tokens” listesindeki sıralamaya karşılık gelecek şekilde listedeki kelimelerin dili. Yalnızca Türkçe veri kümesi ele alındığı için bu listelerdeki tüm elemanlar ‘tr’den ibaret.
    2. **spans:** “ner\_tags”de verilen sayılara göre “tokens” listesinde verilen varlığın tam adı ve türü.
  1. **Değerlendirme sütunları:**
     1. **Doğruluk:** Aşağıdaki dört soruya da cevabınız evet ise 1 verin, aksi takdirde 0 verin.
        1. Cümlede bulunan her etiketlenebilir varlık belirtilmiş mi?
        2. Etiketlenebilir varlıklar isimlerinde eksik ya da fazla olmadan “spans”da belirtilmiş mi?
        3. Belirtilen varlıkların etiketlemesi (PER, ORG, LOC) doğru mu?
        4. PER, ORG ya da LOC ile etiketlenemeyecek varlıklar/kelimeler “ner\_tags”de 0 (sıfır) ile gösterilmiş mi (yani etiketleme dışarısında tutulmuş mu)?
     2. **Dil bilgisi:** Aşağıdaki iki sorunun cevabı da “hayır” ise kutuya 1 yazın, aksi takdirde 0 yazın.
        1. Yazım hatası var mı?
        2. Noktalama hatası var mı?
     3. **Bağdaşıklık ve Tutarlılık:** Aşağıdaki üç sorunun cevabı da “evet” ise kutuya 1 yazın, aksi takdirde 0 yazın.
        1. Metni oluşturan sözcük ve cümleler arasındaki dil bilgisi kuralları uyumlu mu?
        2. Cümle ve paragrafların anlam bütünlüğü mantık düzenine göre yapılandırılmış mı?
        3. Terimler ve dil yapıları çeviri boyunca tutarlı bir şekilde kullanılıyor mu?
     4. **Anlaşılabilirlik, Akıcılık ve Anlatım Bozukluğu:** Aşağıdaki üç sorunun cevabı da “evet” ise kutuya 1 yazın, aksi takdirde 0 yazın.
        1. Metin doğal ve akıcı bir şekilde okunabiliyor mu?
        2. Ana dili Türkçe olan biri bu cümleleri kurar mıydı?
        3. Metin anlatım bozukluğu içermeyen cümlelerden mi oluşuyor?
     5. **Terim kullanımı:** Paragraf ve soruda teknik ve özel terimler doğru kullanılmış mı? Bu soruya cevabınız evet ise 1, hayır ise 0 verin. Örnek için bu belgenin “Kriterler” bölümüne bakabilirsiniz.
     6. **Türkiye’deki Ortak Kültürel Sağduyu Bilgisine Uygunluk:** Belirtilen veri, Türkiye’nin sosyal, ekonomik, kültürel, coğrafi kuralları ile ilgili mi? Kriterler kısmındaki koşulları içeriyorsa belirlenen kurallar çerçevesinde mi içeriyor? Cevabınız her ikisine de evet ise 1, hayır ise 0 verin.
     7. **Yorum**

1. **XTREME (POS)**
   1. **Açıklama:** Kelime ve noktalama işaretleri bir listenin elemanları olarak ayrılıyor ve bu listenin her elemanı cümlenin ögeleri olarak sınıflandırılıyor.
   2. **veri kümesi sütunları:**
      1. **tokens:** Kelime ve noktalama işareti olarak ayrılmış ibareleri içerir.
      2. **pos\_tags:** “tokens” listesindeki sıralamaya karşılık gelecek şekilde sayılarla sınıflandırılmış POS etiketleri.

0: sıfat

1: önleç

2: zarf

3: yardımcı fiil

4: eşgüdümlü bağlaç

5: belirteç

6: ünlem

7: ad/isim

8: sayı

9: edat/ilgeç

10: zamir

11: özel isim

12: noktalama

13: yantümce bağlacı

14: sembol

15: fiil

16: diğerleri

daha fazla bilgi için: https://universaldependencies.org/u/pos/

* 1. **Değerlendirme sütunları:**
     1. **Doğruluk:** Cümlenin ögeleri doğru sırada ve doğru sınıflarda gösterilmiş mi? Bir yanlış bile varsa 0, yoksa 1 verin.
     2. **Dil bilgisi:** Aşağıdaki iki sorunun cevabı da “hayır” ise kutuya 1 yazın, aksi takdirde 0 yazın.
        1. Yazım hatası var mı?
        2. Noktalama hatası var mı?
     3. **Bağdaşıklık ve Tutarlılık:** Aşağıdaki üç sorunun cevabı da “evet” ise kutuya 1 yazın, aksi takdirde 0 yazın.
        1. Metni oluşturan sözcük ve cümleler arasındaki dil bilgisi kuralları uyumlu mu?
        2. Cümle ve paragrafların anlam bütünlüğü mantık düzenine göre yapılandırılmış mı?
        3. Terimler ve dil yapıları çeviri boyunca tutarlı bir şekilde kullanılıyor mu?
     4. **Anlaşılabilirlik, Akıcılık ve Anlatım Bozukluğu:** Aşağıdaki üç sorunun cevabı da “evet” ise kutuya 1 yazın, aksi takdirde 0 yazın.
        1. Metin doğal ve akıcı bir şekilde okunabiliyor mu?
        2. Ana dili Türkçe olan biri bu cümleleri kurar mıydı?
        3. Metin anlatım bozukluğu içermeyen cümlelerden mi oluşuyor?
     5. **Terim kullanımı:** Cümlede teknik ve özel terimlerin çevirisi doğru yapılmış mı? Bu soruya cevabınız evet ise 1, hayır ise 0 verin. Örnek için bu dokümandaki “Kriterler” bölümüne bakabilirsiniz.
     6. **Türkiye’deki Ortak Kültürel Sağduyu Bilgisine Uygunluk:** Belirtilen veri, Türkiye’nin sosyal, ekonomik, kültürel, coğrafi kuralları ile ilgili mi? Kriterler kısmındaki koşulları içeriyorsa belirlenen kurallar çerçevesinde mi içeriyor? Cevabınız her ikisine de evet ise 1, hayır ise 0 verin.
     7. **Yorum**

1. **STS-B Turkish**
   1. **Açıklama:** Verilen iki verinin birbirlerine ne kadar yakın oldukları sorusu. İngilizce STS-B veri kümesinin makine çevirisi ile Türkçeye çevrilmiş hali.
   2. **Veri kümesi sütunları:**
      1. **sentence1:** Verilen ilk cümle
      2. **sentence2:** Verilen ikinci cümle
      3. **score:** Verilen iki cümlenin yakınlığına 5.0 ile 0.0 arası verilen bir skor. Tam sayı olmak zorunda değil.
   3. **Değerlendirme sütunları:**
      1. **Doğruluk:** İki cümlenin birbirine yakınlığı anlamlı bir puan almış mı? Puan mantıklı ise 1, değilse 0 verin.
      2. **Dil bilgisi:** Aşağıdaki iki sorunun cevabı da “hayır” ise kutuya 1 yazın, aksi takdirde 0 yazın.
         1. Yazım hatası var mı?
         2. Noktalama hatası var mı?
      3. **Bağdaşıklık ve Tutarlılık:** Aşağıdaki üç sorunun cevabı da “evet” ise kutuya 1 yazın, aksi takdirde 0 yazın.
         1. Metni oluşturan sözcük ve cümleler arasındaki dil bilgisi kuralları uyumlu mu?
         2. Cümle ve paragrafların anlam bütünlüğü mantık düzenine göre yapılandırılmış mı?
         3. Terimler ve dil yapıları çeviri boyunca tutarlı bir şekilde kullanılıyor mu?
      4. **Anlaşılabilirlik, Akıcılık ve Anlatım Bozukluğu:** Aşağıdaki üç sorunun cevabı da “evet” ise kutuya 1 yazın, aksi takdirde 0 yazın.
         1. Metin doğal ve akıcı bir şekilde okunabiliyor mu?
         2. Ana dili Türkçe olan biri bu cümleleri kurar mıydı?
         3. Metin anlatım bozukluğu içermeyen cümlelerden mi oluşuyor?
      5. **Terim kullanımı:** Cümlelerde teknik ve özel terimlerin çevirisi doğru yapılmış mı? Bu soruya cevabınız evet ise 1, hayır ise 0 verin. Örnek için bu dokümandaki “Kriterler” bölümüne bakabilirsiniz.
      6. **Türkiye’deki Ortak Kültürel Sağduyu Bilgisine Uygunluk:** Belirtilen veri, Türkiye’nin sosyal, ekonomik, kültürel, coğrafi kuralları ile ilgili mi? Kriterler kısmındaki koşulları içeriyorsa belirlenen kurallar çerçevesinde mi içeriyor? Cevabınız her ikisine de evet ise 1, hayır ise 0 verin.
      7. **Yorum**
2. **Belebele**
   1. **Açıklama:** Belebele, paragraf anlama üzerine çoktan seçmeli sorulardan oluşan bir veri kümesidir. veri kümesinde Wikinews, Wikijunior ve WikiVoyage’dan alınan bir metinle ilgili bir ya da birkaç soru sorularak doğru cevabın şıklar arasından seçilmesi bekleniyor. Paragrafların ve soruların aslı İngilizce hazırlanmış olup hepsi akıcı İngilizce ve Türkçe konuşabilen çevirmenler tarafından çevrilmiştir. Değerlendirme için 100 örnek soru alınırken hepsinin farklı metinlere ait olmasına dikkat edilmiştir.
   2. **Veri kümesi sütunları:**
      1. **link:** İngilizce metnin alındığı link
      2. **question\_number:** Srounun metnin soruları arasından kaçıncı olduğunu gösteren sayı
      3. **flores\_passage:** Metnin kendisi
      4. **question:** Metinle ilgili sorulan soru
      5. **mc\_answerX:** X. cevap şıkkı ( X: 1, 2, 3 ya da 4 olabilir)
      6. **correct\_answer\_num:** Doğru cevabın kaç numaralı şık sütununa denk geldiğini belirten sayı, indisler 1’den başlamaktadır (örn: bu sütundaki değer 4 ise, doğru cevap “mc\_answer4” sütunundaki cevap)
      7. **dialect:** Sorunun hangi dilde olduğunu ve hangi alfabenin kullanıldığını gösteren sütun. Türkçe alt kümesi için bu sütundaki tüm değerler “tur\_Latn”
      8. **ds:** tarih damgası
   3. **Değerlendirme sütunları:**
      1. **Doğruluk:** Aşağıdaki üç sorunun cevabı da “evet” ise kutuya 1 yazın, aksi takdirde 0 yazın.
         1. Sorunun cevabına paragraftan ulaşmak mümkün mü?
         2. “correct\_answer\_num” sütunundaki sayıya karşılık gelen cevap doğru ve eksiksiz mi?
         3. Şıkların sadece biri mi doğru cevabı veriyor?
      2. **Dil bilgisi:** Aşağıdaki iki sorunun cevabı da “hayır” ise kutuya 1 yazın, aksi takdirde 0 yazın.
         1. Yazım hatası var mı?
         2. Noktalama hatası var mı?
      3. **Bağdaşıklık ve Tutarlılık:** Aşağıdaki üç sorunun cevabı da “evet” ise kutuya 1 yazın, aksi takdirde 0 yazın.
         1. Metni oluşturan sözcük ve cümleler arasındaki dil bilgisi kuralları uyumlu mu?
         2. Cümle ve paragrafların anlam bütünlüğü mantık düzenine göre yapılandırılmış mı?
         3. Terimler ve dil yapıları çeviri boyunca tutarlı bir şekilde kullanılıyor mu?
      4. **Anlaşılabilirlik, Akıcılık ve Anlatım Bozukluğu:** Aşağıdaki beş sorunun cevabı da “evet” ise kutuya 1 yazın, aksi takdirde 0 yazın.
         1. Metin doğal ve akıcı bir şekilde okunabiliyor mu?
         2. Ana dili Türkçe olan biri bu cümleleri kurar mıydı?
         3. Metin anlatım bozukluğu içermeyen cümlelerden mi oluşuyor?
         4. Soru cümlesini okuduğunuzda neyin sorulduğunu anlayabiliyor musunuz?
         5. Paragrafı okuduğunuzda ne anlattığını anlayabiliyor musunuz?
      5. **Terim kullanımı:** Paragraf ve soruda teknik ve özel terimlerin çevirisi doğru yapılmış mı? Bu soruya cevabınız evet ise 1, hayır ise 0 verin. Örnek için bu dokümanın “Kriterler” bölümüne bakabilirsiniz.
      6. **Türkiye’deki Ortak Kültürel Sağduyu Bilgisine Uygunluk:** Belirtilen veri, Türkiye’nin sosyal, ekonomik, kültürel, coğrafi kuralları ile ilgili mi? Kriterler kısmındaki koşulları içeriyorsa belirlenen kurallar çerçevesinde mi içeriyor? Cevabınız her ikisine de evet ise 1, hayır ise 0 verin.
      7. **Yorum**

**REFERANSLAR**

Onursal, İrem (2003). Türkçe Metinlerde Bağdaşıklık ve Tutarlılık. Günümüz Dilbilim Çalışmaları. Multilingual Yayınları.

Günay, Doğan (2007). Metin Bilgisi. Multilingual Yayınları.

Acharya, A., Talamadupula, K., & Finlayson, M. A. (2020). Towards an atlas of cultural commonsense for machine reasoning. *arXiv preprint arXiv:2009.05664*.

Anacleto, J., Liebennan, H., Tsutsumi, M., Neris, Y, Can'alho, A. et. al., 2006, in IFIP International Federation for Information Processing, Volume 217, Artificial Intelligence in Theory and Practice, ed. M. Bramer, (Boston: Springer), pp. 1-10.

Nguyen, T. P., Razniewski, S., Varde, A., & Weikum, G. (2023, April). Extracting cultural commonsense knowledge at scale. In *Proceedings of the ACM Web Conference 2023* (pp. 1907-1917).

["Yazım Kuralları"](https://drive.google.com/file/d/1NJ3-SzXrQIRiFoLRw8oHc1qrzoeqH8HN/view?usp=drive_link) dokümanı.

["Anlatım Bozuklukları"](https://drive.google.com/file/d/1NHQyZXApRtmwbtvALU4ARt8rvO13amE2/view?usp=drive_link) dokümanı.