

Kısa Konuşma Cümlelerinin Dönüştürücü Yöntemleriyle Otomatik Etiketlenmesi

Auto-tagging of Short Conversational Sentences using Transformer Methods

D. Emre Taşar*, Şükrü Ozan†, Umut Özdi†, M. Fatih Akca‡, Oğuzhan Ölmez§, Semih Gülüm¶, Seçilay Kutal¶, Ceren Belhan||

* emretasa@garantibbva.com.tr

† sukruozan@adresgezgini.com, umutozdil@adresgezgini.com

‡ mehmet.akca1@ogr.sakarya.edu.tr

§ 172803041@ogr.cbu.edu.tr

¶ semihgulum@marun.edu.tr, secilaykutsal@marun.edu.tr

|| ceren.belhan@std.ieu.edu.tr

Özetçe —Kısa konuşma cümlelerinin anlamsal özelliklerine göre yüksek doğruluk ile kategorilere ayrıştırılması problemi, doğal dil işleme alanında üzerinde çalışılan bir konudur. Bu çalışmada 46 farklı kategoride sınıflandırılan örnekler ile oluşturulan bir veri seti kullanılmıştır. Örnekler, bir firmanın müşteri temsilcileriyle, firmanın internet sitesi ziyaretçileri arasında gerçekleşen yazılı sohbet (chat) görüşmelerinden alınmış cümlelerden oluşmaktadır. Ana amaç, sorulan sorulara anlamlı yanıtlar üretebilen bir sohbet uygulamasında kullanmaya yönelik olarak internet sitesi ziyaretçilerden gelen soruları ve talepleri önceden belirlenen 46 kategori için en doğru şekilde otomatik olarak etiketlemektir. Bunun için Türkçe dilinde ön eğitime tabi tutulmuş birbirinden farklı BERT modelleri ve bir adet GPT-2 modeli tercih edilmiştir. İlgili modellerin sınıflandırma başarımları detaylı bir şekilde incelenerek raporlanmıştır.

Anahtar Kelimeler—otomatik etiketleme, doğal dil işleme, BERT, GPT-2, çoklu sınıf.

Abstract—The problem of categorizing short speech sentences according to their semantic features with high accuracy is a subject studied in natural language processing. In this study, a data set created with samples classified in 46 different categories was used. Examples consist of sentences taken from chat conversations between a company's customer representatives and the company's website visitors. The primary purpose is to automatically tag questions and requests from visitors in the most accurate way for 46 predetermined categories for use in a chat application to generate meaningful answers to the questions asked by the website visitors. For this, different BERT models and one GPT-2 model, pre-trained in Turkish, were preferred. The classification performances of the relevant models were analyzed in detail and reported accordingly.

Keywords—auto-tagging, natural language processing, BERT, GPT-2, multi-class.

I. GİRİŞ

Bu çalışmada doğal dil işleme modelleri yardımı ile sorulara otomatik olarak yanıtlar üretmesi hedeflenen bir chatbot için, gelen soruları önceden belirlenmiş 46 temel kategoriye

mümkün olduğunca doğru bir şekilde ayırabilmek hedeflenmektedir. Günümüzde yoğun olarak çalışılmakta olan bu konu ile ilgili olarak literatürde birçok güncel çalışmaya rastlamak mümkündür.

Metin sınıflandırma problemi için Jain vd. tarafından hibrit kümeleme ve sınıflandırma modeli (HCC) [1] önerilmiştir. Bu modelde, metinlerin kümelenmesi için k-ortalama metodu işleme alınırken, çok etiketli bir sınıflandırma yapılabilmesi için ise derin öğrenme algoritmaları kullanılmıştır. Çalışma sonunda sınıflandırma problemi için Evrişimli Sinir Ağları (CNN) başarılı bir sonuç verirken, çok etiketli sınıflandırma için Uzun-Kısa Süreli Bellek (LSTM) ya da Derin CNN'in daha etkili olabileceği belirtilmiştir.

BERT modeli doğal dil işlemede metnin sınıflandırması için kullanılan açık kaynaklı bir modeldir. Devlin vd. tarafından geliştirilen transformatörlerden çift yönlü kodlayıcı gösterimlere sahip olan BERT modeli [2] ile birlikte, etiketlenmemiş metinlerin derin çift yönlü temsillerinin önceden eğitimi sırasında tüm katmanlarda sağ ve sol bağlam bilgilerinin dahil edilmesi sağlanmaktadır. Model, sonrasında ince ayar yapılarak göreve özgü şekilde eğitilebilmektedir. Bu görevler soru cevaplama, duygu analizi, metin sınıflandırma ve adlandırılmış varlık tanıma gibi farklı çeşitlerde olabilir [3].

İnce ayar yapılabilmesi için önceden eğitilmiş çeşitli BERT modelleri mevcuttur. Bu modeller dil açısından, dile özgü ve çok dilli olarak ikiye ayrılmaktadır. Türkçe dili özelinde literatürde yeteri kadar iyi eğitilmiş çok fazla dil modeli olmadığı için çalışmada Türkçe diline özel olarak eğitilmiş olanların yanı sıra çok dilli modeller de kullanılmıştır. Ronnqvist vd. çalışmalarında [4] kullandıkları çok dilli model ile 6 dil üzerinde çeşitli görevlerde test gerçekleştirmiş ve bu modelin başarısının diller arasındaki değişimini gözlemlemişlerdir. Çalışmada Türkçe ile aynı dil ailesine üye olan Fince dilinde model %93.2 doğruluk değerine ulaşmıştır.

Özçift vd. çalışmasında [3] morfolojik olarak zor bir dil olan Türkçe ile 6 farklı görevde geleneksel makine öğrenmesi yöntemlerinin yanında BERT modelinin perfor-

mansını karşılaştırmıştır. Spam etiketleme görevi dışında, BERT modeli diğer modellere nazaran daha yüksek doğruluk oranı vermiştir.

Tanaka vd. çalışmasında [5] BERT modelini sınıflandırıcı olarak kullanmıştır. Dokümanların sınıflandırılması için ince ayar yapılan BERT modelinde, girdi olarak doküman verilmiş ve BERT'in metin sınıflandırmadaki başarısı ortaya konulmuştur.

Ozan vd. çalışmasında [6] kategorik sınıflandırma probleminde yaygın olarak kullanılan Doc2Vec, LSTM ve BERT gibi modellerin performanslarını karşılaştırmıştır. Çalışmada BERT modelinin %93 doğruluk seviyesiyle diğer modellere göre yüksek başarımlar gösterdiği tespit edilmiştir.

Bahsedilen çalışmalar doğrultusunda seçilen BERT modelleri, Bölüm II'de detayları verilen veri seti ile, sınıflandırma problemine yönelik olarak ince ayar yapılarak eğitilmiştir.

BERT modelleri dışında çalışmada yer verilen bir diğer model olan GPT-2, Radford vd. tarafından geliştirilmiş üretken yapıya sahip olan bir kod çözücü transformatördür [7]. Çalışmada bu model, kendi içinde özel olarak sınıflandırma yapan bir kütüphanenin yardımı ile metin sınıflandırma amacıyla kullanılmıştır.

Whitfield, sınıflandırma problemlerinde model performansını arttırabilmek adına GPT-2 modelini kullanarak ürettiği sentetik verilerle, veri setini büyüterek başarımlarını metriklerini karşılaştırmayı hedeflemiştir [8]. Çalışmadan ulaşılan sonuçlara göre GPT-2, kuruluşların ve işletmelerin, büyük ölçekli veri toplamaya ilişkin yüksek maliyetlere katlanmadan yüksek performansa sahip sınıflandırma modelleri oluşturmasına izin vermektedir.

II. KULLANILAN VERİ SETİ

Veri seti için firmanın internet sitesinde kullanılan sohbet arayüzü aracılığıyla web sitesi ziyaretçileri tarafından yöneltilen sorular ele alınmıştır. Elde edilen 73551 adet ve kelime sayısı 2 ile 30 arasında değişen seçili cümle, 46 kategori için manuel olarak etiketlenmiştir. Örnek yedi kategori için seçilmiş bazı örnek cümleler Tablo I'de verilmiştir. Veri setinde bulunan cümle-cümle gruplarının uzunluğunun geniş bir aralıkta değişkenlik göstermesi (Şekil 1), kategorilere göre veri dağılımının düzensiz oluşu (Şekil 2) ve benzer kalıplarda verilerin olması hazırlanan veri setinin kullanılmasının zorlukları olarak sayılabilir. Ön işleme aşamasında veri setindeki cümlelerde yer alan sayılar ve noktalama işaretleri atılmış, bütün harfler küçük harflere çevrilmiştir. Veri seti %70'i eğitim, %30'u da test için olacak şekilde iki gruba ayrılmıştır. Eğitim aşamasında kullanılan yöntemlere ait detaylar Bölüm III'de verilecektir.

III. YÖNTEMLER

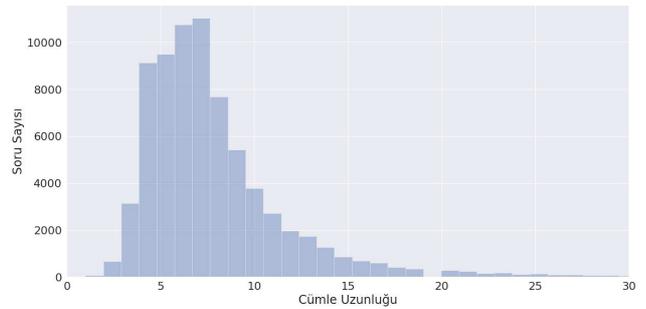
A. BERT Yöntemi

BERT modeli ön eğitim ve ince ayar olarak 2 aşamadan meydana gelmektedir. Ön eğitim, birbirinden farklı ve etiketlenmemiş veriler üzerinden bir konuşma diline özel yada birden fazla konuşma dili için eğitilir. Ön eğitimden geçmiş bir BERT modeline etiketlenmiş veriler ile ince ayar yapılır. İnce ayar sayesinde BERT'in dikkat mekanizması neye dikkat

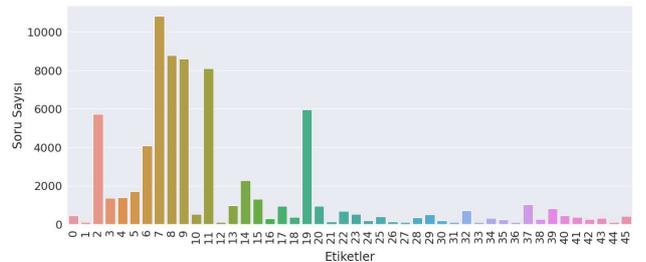
TABLO I: Örnek Cümleler ve Eşleştirilmiş Etiketler

Facebook reklamları	Facebook reklamı için görmek istiyorum
	Facebook reklam paketleri, ücretleri ve kampanyalarınız ile ilgili bilgi verir misiniz?
	sitemi Facebook reklam yapmak istiyorum
Instagram reklamları	Instagram şirket reklamı vermek istiyordum bu konu ile yardımcı olabilir misiniz
	Instagram nasıl reklam verebilirim
	Instagram reklamlarıyla ilgili bilgi almak istemişim
Mobil uygulama	mobil uygulamalarının web entegrasyonunu yapabiliyor musunuz
	Merhaba, mobil uygulama için fiyat teklifi almak istiyordum detayları hangi mail adresi ile paylaşabiliriz?
	mobil uygulama yaptırmak istiyorum ne yapmalıyım
Ofisiniz nerede	Adres bilgisi alabilir miyim
	hangi ilden hizmet veriyorsunuz
	Ankara'da yüzyüze görüşüp ilgi alabileceğimiz bir ekibiniz mevcut mudur
İş başvurusu	firmanızda çalışma imkanımız var mı
	Bu yaz için staj imkanlarınızı öğrenmek istiyorum. Bilgisayar Mühendisliği 3.sınıf öğrencisiyim
	acaba iş başvuru yapabilmesi için mail adresiniz var mıdır
Web sitesine canlı destek ekleme	Canlı destek hizmeti sunuyormusunuz
	müşteri telefonlarına yetişemiyoruz canlı destek hizmetinizden yararlanmak isteriz ücreti ne kadar
	Birde bu canlı destek sisteminizin ücreti nedir ?
Web sitesi yaptırmak	Siz web tasarımı yapıyor musunuz peki
	ben web sitesi yaptırmayı düşünüyorum
	ya peki web adresi kurabilir miyiz

Not: Toplam kategori sayısı 46 olmasına karşın tabloda 7 örnek kategori paylaşılmıştır.



Şekil 1: Veri Setindeki Cümlelerin Uzunluklarına Göre Dağılım Grafiği



Şekil 2: Veri Setinin Etiketlere Göre Dağılım Grafiği

etmesi gerektiğini anlayabilir. Çalışmada kullanılan BERT-base modeli, 12 transformatör bloğu, 12 öz-dikkat başlığı ve 768 gizli katmana sahip bir kodlayıcı içerir ve çift yönlü öz-dikkat kullanır [9]. Kullanılan distil-BERT modeli ise BERT-base modeline göre daha küçük bir yapıya sahip olduğundan %60 daha hızlı çalışırken BERT performansını da büyük ölçüde koruyabilmektedir [10]. BERT modelinin eğitimi büyük harflere duyarlılık açısından, "cased" (büyük küçük harf kullanımına karşı duyarlı) ve "uncased" (büyük küçük harf kullanımına karşı duyarlısız) olarak iki farklı model tipi ile gerçekleştirilebilmektedir.

B. GPT-2 Yöntemi

Mevcut sistemler, genel veri setleri ile çalışmak yerine daha dar konularda çalışmaya uygundur. “Üretken Eğitim Öncesi Transformatör” olarak adlandırılan ve anlamlı metin oluşturmada oldukça başarılı olan GPT-2, mevcut sistemlerin aksine çoklu görevlerde çalışabilecek ve özel görevlere göre de konfigüre edilebilecek şekilde tasarlanmıştır.

Bir kod çözücü transformatörü olan GPT-2’nin eğitimi esasında girdi dizisinin son belirteci, tahminde ihtiyaç duyulan tüm bilgileri içerdiğinden, girdiyi takip eden bir sonraki token hakkında tahmin yapılması amacıyla kullanılmaktadır. Girdi dizisinin son belirtecinin taşıdığı bilgi, üretim problemi yerine bir sınıflandırma probleminde tahmin yapmak amacıyla da kullanılabilir. GPT-2 ile tahmin yapılırken, BERT modelinin sınıflandırma probleminde uygulandığı gibi ilk token yerleştirmek yerine son token yerleştirme uygulanmaktadır [11].

Bu kapsamda gerçekleştirilmeye çalışılan sınıflandırma problemi için Bölüm I’de bahsedilen yöntemler ile birlikte BERT ve GPT-2 modellerinin başarımı test edilmiştir. Daha önceden firma bünyesinde yapılmış olan çalışma [6] genişletilerek toplamda 10 adet BERT modeli ve bir adet GPT-2 modeli kullanılmıştır. GPT-2’nin sınıflandırıcı olarak kullanıldığı çalışmada, ilgili modeller söz konusu veri seti ile ince ayar yapılarak eğitilmiştir. Model mimarilerine ait bazı temel parametrelere ait değerler Tablo II’de gösterilmiştir.

C. F1 Skoru ve Başarım Kriterleri

Eğitilen modellerin başarısı doğruluk değeri, F1 skoru, kayıp değerleri ve hata düzeyi (confusion matrix) baz alınarak değerlendirilmiştir. Bu başarı ölçütleri; True Positive (TP), False Positive (FP), True Negative (TN) ve False Negative (FN) değerlerine bakılarak hesaplanır. Modelin çıktısı ikili sınıflandırma olarak göz önüne alınırsa; TP, modelin tahmini bir durum için olumlu sonuç vermesi ve asıl durumda da sonucun olumlu olduğunu gösterirken TN, algoritma çıktısının olumsuz iken asıl durumun da olumsuz olduğunu temsil eder. İki tip hata vardır, bunlar FP ve FN kavramlarıdır. FP algoritmanın tahminini olumluyken asıl değer olumsuz olması durumudur. FN ise algoritmanın olumsuz dediği bir sonuç karşısında asıl durumun olumlu olmasıdır. Çalışmada kullanılan veri setinde örnek bir etiket için elde edilen sonuçlarla oluşturulmuş olan Tablo III, TP, FP, TN ve FN kavramlarının daha iyi anlaşılabilmesi için incelenebilir.

Doğruluk değeri en temel başarım kriteri olarak ifade edilebilir. Bu metrikte modelin ürettiği doğru tahmin sonuçları üretilen tüm tahminlere oranlanmaktadır (Denklem 1). Ancak bu yöntem kategorik sınıflandırma problemlerinde sınıf bazlı bir sonuç veremediğinden tercih edilmemektedir. Bu durumlarda F1 skoru adı verilen ve kesinlik-doğruluk değerlerine dayanan ayrı bir metrik kullanılmaktadır.

Kesinlik (precision) değeri modelin tahmin ettiği TP sayısının modelin ürettiği tüm olumlu sonuçlara oranı ile bulunmaktadır (Denklem 2). Duyarlılık (recall) değeri ise modelin tahmin ettiği TP sayısının modelin tahmin etmesi gereken tüm olumlu sonuçlara oranı ile bulunmaktadır (Denklem 3). Bu iki kavram modelin ürettiği sonuçların farklı iki durumu üzerinden hesaplandığından her iki durumun da etkin bir

şekilde yorumlandığı F1 skoru, Denklem 4’te ifade edildiği gibi kesinlik ve duyarlılık değerlerinin harmonik ortalaması alınarak hesaplanmaktadır [21].

$$\text{Doğruluk} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Kesinlik} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Duyarlılık} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{F1skoru} = 2 \left(\frac{\text{Kesinlik} \times \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \right) \quad (4)$$

Kategorilere göre veri sayısının düzenli dağılmaması, çalışmada kullanılan modellerin genel başarımının tespiti için farklı bir metriğin kullanımını gerektirmektedir. Kategori bazında hesaplanmış F1 skorunun kategori bazlı veri dağılımına göre ağırlıklı ortamasının alınarak hesaplanmasıyla dengesiz veri dağılımında ortaya çıkan bu problem belirli bir ölçüde çözülebilmektedir.

IV. UYGULAMA SONUÇLARI

Çalışmada sınıflandırma problemi, Bölüm III-A ve Bölüm III-B’de anlatıldığı gibi farklı BERT-base, distilBERT-base ve GPT-2 modellerinin eğitimleri ile çözülmeye çalışılmıştır. Bu bölümde ise eğitilen modellerin sonuçları Bölüm III-C’de verilen kriterler göz önüne alınarak incelenecektir. Her modelin eğitimi üç dönem (epoch) gerçekleştirilerek parametreler hesaplanmıştır. Şekil 4’de modellerin doğruluk/dönem grafikleri çizdirilmiştir.

GPT-2 metin üretmede göstermiş olduğu başarısını metin sınıflandırma probleminde gösterememiş ve doğruluk değeri ancak %84 değerine kadar çıkabilmiştir. Ancak çalışma için daha önemli bir karşılaştırma kriteri olan F1 skoru %73 olarak gözlemlenmiştir.

BERT modeli için önceden eğitilmiş “cased” ve “uncased” modeller ile gerçekleştirilen ince ayar eğitimlerinin ardından başarım oranları incelendiğinde, iki farklı model tipinin sonuç metrikleri arasında, ilgili problem ve veri seti düşünüldüğünde, çok büyük bir fark gözlemlenmemiştir.

Başarım oranları incelendiğinde en başarılı modelin %89 doğruluk değeri ve %82 F1 skoruyla loodos/bert-base-turkish-cased [17] olduğu gözlemlenmiştir. Bu eğitimin çoklu sınıf bazlı başarısı hata düzeyi üzerinden Şekil 5’de gösterilmiştir. Aynı modelin eğitimi dönem sayısının etkisinin gözlemlenmesi amacıyla 10 dönem ile tekrarlanmıştır. Bulgular sonucunda 3 dönemden sonra modelin aşırı öğrenmeye gittiği gözlemlenmiştir. Modelin dönemlere göre başarısı doğruluk/dönem grafiğine göre Şekil 3’den incelenebilir.

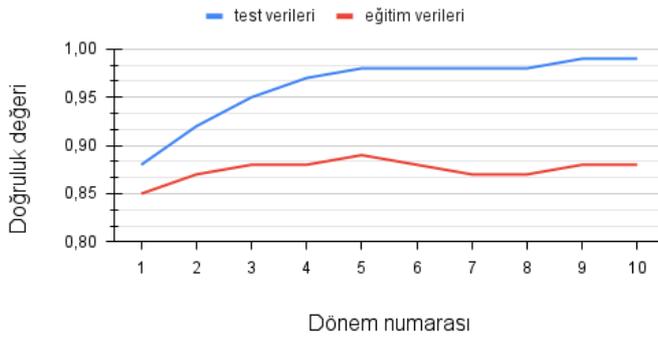
TABLO II: Kullanılan Ön Eğitimli BERT Modelleri ve GPT-2 Modelinin Parametreleri

Model	Gizli Katman Boyutu	Maksimum Sekans Uzunluğu	Dikkat Ana Başlığı Sayısı	Gizli Katman Sayısı	Mimari	Dil	Sözlük Boyutu
bert-base-multilingual-cased [12]	768	512	12	12	BertForMaskedLM	Çok dilli	119547
bert-base-multilingual-uncased [12]						Çok dilli	105879
dbmdz/bert-base-turkish-128k-cased [13]						Türkçe	128000
dbmdz/bert-base-turkish-128k-uncased [14]						Türkçe	128000
dbmdz/bert-base-turkish-cased [15]						Türkçe	32000
dbmdz/bert-base-turkish-uncased [14]						Türkçe	32000
Geotrend/bert-base-tr-cased [16]						Türkçe	19099
loodos/bert-base-turkish-cased [17]						Türkçe	32000
loodos/bert-base-turkish-uncased [18]						Türkçe	32000
dbmdz/distilbert-base-turkish-cased [19]						6	DistilBertForMaskedLM
adresgezgini/turkish-gpt-2 [20]	768	1024	12	12	GPT2LMHeadModel	Türkçe	50000

TABLO III: Karmaşıklık Matrisi

Tahmin edilen sınıf	İş başvurusu (1)	Ait olduğu sınıf	
		İş başvurusu (1)	İş başvurusu değil (0)
İş başvurusu değil (0)	Şirketinizde çalışmak için nereyi doldurmam gerekir TP	İşyerim için reklam vermek istiyorum FP	Sosyal medya reklamları nasıl oluşturuluyor TN
	İşe alımlar devam ediyor mu FN		

Doğruluk/Dönem Grafiği



Şekil 3: loodos/bert-base-turkish-cased [17] Modeli 10 Dönem Eğitim Sonuçları

V. SONUÇLAR

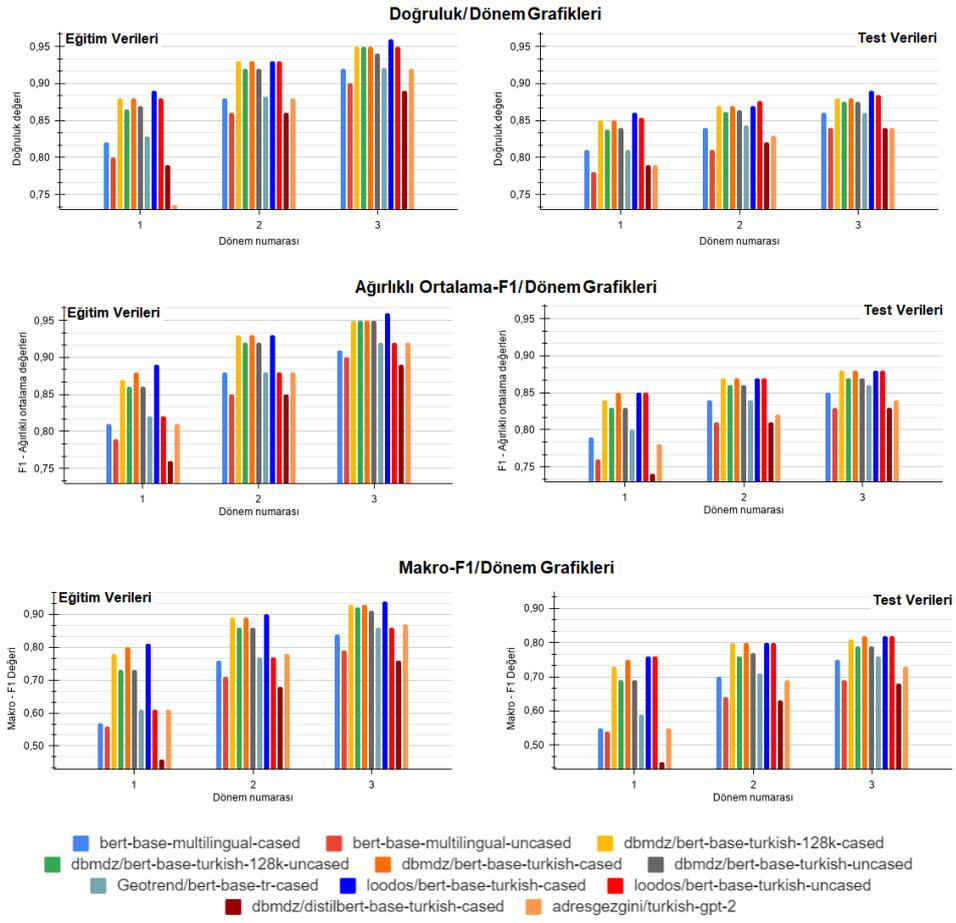
Bu çalışmada, önceden eğitilmiş BERT modellerinin sınıflandırma problemi üzerindeki performansları karşılaştırılarak en yüksek başarı seviyesine sahip model belirlenmeye çalışılmış ve bu modeller aynı zamanda kendi eğittiğimiz GPT-2 modelinin aynı problem özelindeki başarısı ile karşılaştırılmıştır. GPT-2 ile BERT modelinde elde edilen başarı seviyelerine ulaşamamıştır. Gerçekleştirilen eğitimler doğrultusunda BERT modelleri arasından Loodos ekibine ait online bloglardan, e-kitaplardan, gazetelerden, Twitter'dan, makalelerden ve Wikipedia sitesi üzerinden toplanan 200GB boyutundaki külliyat ile önceden eğitilmiş, 32000 adet kelime bulunan sözlüğü ve 110 milyon parametreye sahip BERT modelinin en iyi başarımlarını verdiği gözlemlenmiştir. İlerleyen çalışmalarda, doğruluk değeri %90'ın altında kalan sınıflar için GPT-2 gibi üretken bir özelliğe sahip model yardımıyla veri artırma (data augmentation) yoluna gidilerek sonuçların iyileştirilebileceği öngörülmektedir.

BİLGİLENDİRME

Bu çalışma TÜBİTAK TEYDEB 1501 programı kapsamında desteklenmiş olan 3190585 numaralı "Makine Öğrenmesi ile Anlamlı Diyalog Üretebilecek Genel Amaçlı Chatbot Uygulaması" isimli proje kapsamında gerçekleştirilmiştir.

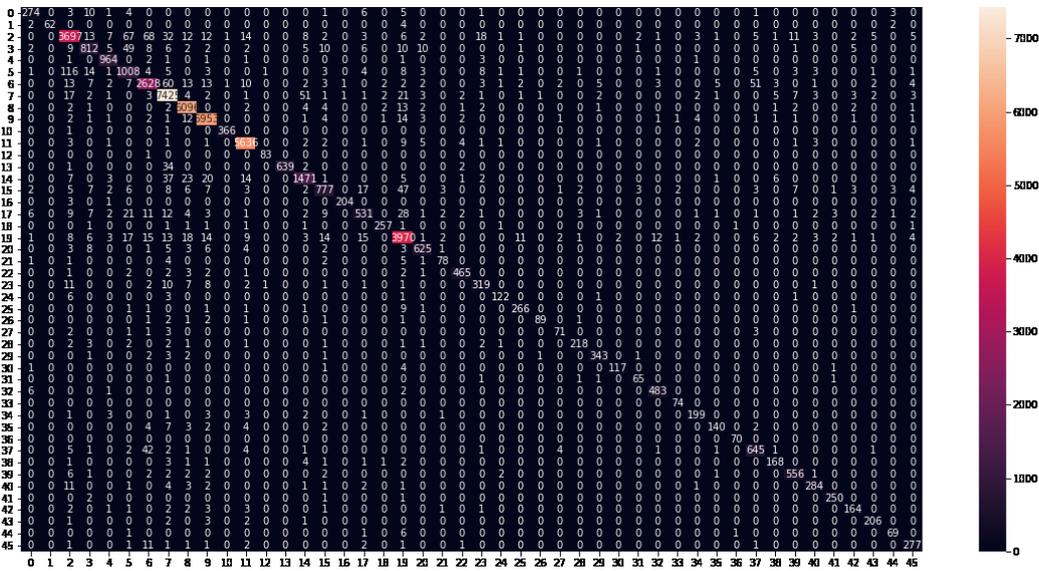
KAYNAKÇA

- Jain, S., Jain, A. K., and Singh, S. P., "Building a machine learning model for unstructured text classification: Towards hybrid approach," in *Rising Threats in Expert Applications and Solutions*, pp. 447–454. Springer, 2021.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K., "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
- Özçift, A., Akarsu, K., Yumuk, F., and Söylemez, C., "Advancing natural language processing (nlp) applications of morphologically rich languages with bidirectional encoder representations from transformers (bert): an empirical case study for turkish," *Automatika*, pp. 1–13, 2021.
- Rönqvist, S., Kanerva, J., Salakoski, T., and Ginter, F., "Is multilingual bert fluent in language generation?," *arXiv preprint arXiv:1910.03806*, 2019.
- Tanaka, H., Shinnou, H., Cao, R., Bai, J., and Ma, W., "Document classification by word embeddings of bert," in *International Conference of the Pacific Association for Computational Linguistics*. Springer, 2019, pp. 145–154.
- Şükrü Ozan ve Emre Taşar,, "Kısa konuşma cümlelerinin doğal dil İşleme yöntemlerini kullanarak otomatik etiketlenmesi," 29. IEEE Sinyal İşleme ve İletişim Uygulamaları Kurultayı, SIU 2021, Çevrimiçi, Basım aşamasında bildiri., 2021.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., and Sutskever, I., "Language models are unsupervised multitask learners," 2019.
- Whitfield, D., "Using gpt-2 to create synthetic data to improve the prediction performance of nlp machine learning classification models," *arXiv preprint arXiv:2104.10658*, 2021.
- Sun, C., Qiu, X., Xu, Y., and Huang, X., "How to fine-tune bert for text classification?," *ArXiv*, vol. abs/1905.05583, 2019.
- Sanh, V., Debut, L., Chaumond, J., and Wolf, T., "Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter," *arXiv preprint arXiv:1910.01108*, 2019.
- Mihaila, G., "Gpt2 for text classification using hugging face transformers," https://gmihaila.github.io/tutorial_notebooks/gpt2_finetune_classification/, Mar 2021.
- Devlin, J., Chang, M., Lee, K., and Toutanova, K., "BERT: pre-training of deep bidirectional transformers for language understanding," *CoRR*, vol. abs/1810.04805, 2018.
- dbmdz,, "dbmdz/bert-base-turkish-128k-cased · hugging face," <https://huggingface.co/dbmdz/bert-base-turkish-128k-cased>, Mar 2020.



Şekil 4: BERT ve GPT-2 Modelleri Eğitim Sonuçları

- [14] dbmdz,, “dbmdz/bert-base-turkish-128k-uncased · hugging face,” <https://huggingface.co/dbmdz/bert-base-turkish-128k-uncased>, Mar 2020.
- [15] dbmdz,, “dbmdz/bert-base-turkish-uncased · hugging face,” <https://huggingface.co/dbmdz/bert-base-turkish-uncased>, Mar 2020.
- [16] Abdaoui, A., Pradel, C., and Sigel, G., “Load what you need: Smaller versions of multilingual bert,” in *SustaiNLP / EMNLP*, 2020.
- [17] Loodos,, “loodos/bert-base-turkish-cased · hugging face,” <https://github.com/Loodos/turkish-language-models>, Dec 2020.
- [18] Loodos,, “loodos/bert-base-turkish-uncased · hugging face,” <https://github.com/Loodos/turkish-language-models>, Aug 2020.
- [19] dbmdz,, “dbmdz/distilbert-base-turkish-uncased · hugging face,” <https://huggingface.co/dbmdz/distilbert-base-turkish-cased>, Mar 2020.
- [20] AdresGezgini, I., “/adresgezgini/turkish-gpt-2 · hugging face,” <https://api-inference.huggingface.co/models/adresgezgini/turkish-gpt-2>, Jan 2021.
- [21] Powers, D. M., “Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation,” *arXiv preprint arXiv:2010.16061*, 2020.



Şekil 5: lodos/bert-base-turkish-cased [17] Modeli 3 Dönem Eğitim Sonuçları