

Türkçe Duygu Sınıflandırma İçin Transformers Tabanlı Mimarilerin Karşılaştırılmalı Analizi

Comparative Analysis of Transformers Based Architectures for Turkish Sentiment Classification

Mehmet ARZU^{*1} , Murat AYDOĞAN² 

¹Bilgisayar Mühendisliği Bölümü, Malatya Turgut Özal Üniversitesi, Malatya, Türkiye

²Yazılım Mühendisliği Bölümü, Fırat Üniversitesi, Elâzığ, Türkiye

(mehmet.arzu@ozal.edu.tr, m.aydogan@firat.edu.tr)

Received:Aug.26, 2023

Accepted: Sep.7-8,2023

Published:Oct.18,2023

Özetçe— Transformers tabanlı ile duygu sınıflandırma, son zamanlarda doğal dil işleme ve makine öğrenmesi alanında yaygın olarak çalışılan bir konudur. Metinler içerisinde karşılaşılan duygusal ifadelerin anlamlandırılması ve sınıflandırılması, sosyal medya analizi, piyasa araştırması, kullanıcı deneyimleri vb. gibi kullanılabileceği birçok alan mevcuttur. Bu sebeple, bu çalışmada Transformers tabanlı mimariler kullanılarak duygu sınıflandırmasının gerçekleştirilmesi hedeflenmiştir. Bu çalışmada, 150000 veriden oluşan TRSAv1 veriseti üzerinde, 8 farklı BERTurk ve 2 farklı ELECTRA varyasyonu üzerinde duygu sınıflandırma işlemi için kullanılmıştır. Bu modeller, Türkçe metinler üzerinde duygu sınıflandırılması çalışmalarında kullanılmak için önceden eğitilmiş modellerdir. Veri seti üzerinde 3 farklı metot kullanılarak modeller eğitilmiş ve sonuçlar karşılaştırmalı olarak değerlendirilmiştir. Yapılan deneyler sonucunda, modellerin duygu sınıflandırma performansları doğruluk ve F1-skor metrikleri kullanılarak ölçülmüştür. Deney sonuçları, Transformers modellerinin duygu sınıflandırması konusundaki etkinliğini ve kullanılan modellerin performans değerlendirmelerini ortaya koymuştur.

Anahtar Kelimeler : *Transformers, duygu sınıflandırma, BERTurk, doğal dil işleme*

Abstract— Sentiment classification based on Transformers is a topic that has recently been widely studied in natural language processing and machine learning. There are many areas where it can be used, such as the interpretation and classification of emotional expressions in texts, social media analysis, market research, user experiences, etc. For this reason, this study aims to realize sentiment classification using Transformers based architectures. In this study, 8 different BERTurk and 2 different ELECTRA variations were used for sentiment classification on the TRSAv1 dataset consisting of 150000 data. These models are pre-trained models to be used in sentiment classification studies on Turkish texts. The models were trained on the dataset using 3 different methods and the results were evaluated comparatively. As a result of the experiments, the sentiment classification performance of the models was measured using accuracy and F1-score metrics. The results of the experiments revealed the effectiveness of Transformers models in sentiment classification and the performance evaluations of the models used.

Keywords : *Transformers, sentiment classification, BERTurk, natural language processing*

1. Giriş

Teknolojinin ve internetin yaygınlaşmasıyla birlikte sosyal medyada, e-ticaret sitelerinde vb. platformlarda her gün milyarlarca metin verisi üretilmektedir. Üretilen milyarlarca metin verisi markalar ve işletmeler için değerli bir hale gelmiştir. Bu veriler, markalara kullanıcıların düşüncelerini anlamalarına, hedef kitleleriyle iletişim kurmalarına ve pazarlama stratejilerini daha iyi hale getirmelerine yardımcı olur.

Duygu analizi, metin verilerindeki duygusal ifadeleri tanımlamak ve sınıflandırmak için kullanılan bir yöntemdir. Metinlerin pozitif, negatif veya nötr duyguları yansıtan etiketlerle sınıflandırılması, markaların kullanıcıların gerçek hislerini anlamasını ve hedeflerine daha uygun bir şekilde yanıt vermesini sağlar. Duygu

analizi çalışmalarında, metinlerin pozitif, negatif veya nötr içeriğe sahip olup olmadığı sorgulanarak analiz edilir. Bu analizin sonucuna göre, kişilerin veya belirli bir topluluğun çalışmayla ilgili konu hakkındaki görüşü belirlenmiş olur. Bu bağlamda, duygu analizi şirketler için piyasaya sürülecek bir ürün için ön pazar araştırması yapmada, bir topluluk için alınacak bir kararın pozitif veya negatif nasıl tepki alacağını tahmin etmede yönlendirme sağlayabilir (Kaynar, 2016). Ancak büyük miktardaki metin veri setlerini el ile incelemek zorlayıcı ve uzun zaman alıcı bir süreçtir. Transformers bu gibi süreçlerde devreye giren bir yapay zekâ modelidir.

Transformer modelleri, doğal dil işleme alanında önemli bir gelişim sağlayan son derece etkili yapay sinir ağı modelleridir. Bu modeller, büyük veri setlerinde gözetimsiz öğrenme ile eğitilir. Metinler üzerinde çıkarımlar yapmak, metin oluşturmak, makine çevirisi ve sınıflandırma gibi görevlerde kullanılır. Duygu analizi alanında da Transformers tabanlı modeller, metinlerdeki duyguları anlamak ve bu duyguları belirli etiketlere göre sınıflandırmak için kullanılan önemli bir araçtır.

Yapılan çalışmada, transformers tabanlı mimariler ile duygu sınıflandırma konusu incelenmektedir. Transformerlar metinlerin içerisinde bulunan anlamsal ilişkileri daha iyi kavrayabilen ve dil anlama görevlerinde yüksek seviyede başarılı sonuçlar elde edebilen yapıya sahiptir. Bu modeller büyük miktarda veri üzerinde eğitildiği için dil anlama becerileri, genelleme ve duygu analizi için daha uygundur.

2. İlgili Çalışmalar

Köksal, 2021 yılında yaptığı çalışmada, 6 farklı BERT modelini farklı hiperparametreler kullanarak duygu analizi yapmış ve sonuçları kıyaslamıştır. Bu deneyleri gerçekleştirirken kitap ve film yorumları içeren 2 farklı veri seti kullanmıştır. Deneyler sonucunda en yüksek başarı oranını Türkçe duygu analizi için geliştirilmiş ve Türkçe duygu analizi veri setleri ile ince ayarlanmış (fine-tuning) BERT modeli olan “sabasy/bert-turkish-sentiment-cased” modelinde %97 doğruluk oranı ile elde etmiştir. Buna ek olarak veri seti üzerinde hiçbir ön işleme yapmamasına rağmen, literatürde aynı veri seti kullanılarak derin öğrenme ve makine öğrenmesi algoritmaları ile yapılan çalışmaların tamamını geride bıraktığını vurgulamıştır (Köksal, 2021).

Adoma ve arkadaşlarının 2020 yılında yayınladığı çalışmada, 4 Transformer modelinin (BERT, RoBERTa, DistilBERT ve XLNet) metinlerden duyguları tanımadaki etkinliğini incelemiş ve her bir modelin çıktısını diğerleriyle karşılaştırmıştır. Veri seti olarak 7 farklı duygu etiketi ile sınıflandırılmış, 37 ülkede anket çalışmalarıyla oluşturulmuş ISEAR veri kümesini kullanmışlardır. Deneyler sonucunda RoBERTa için 0,7431, XLNet için 0,7299, BERT için 0,7009 ve DistilBERT için 0,6693 doğruluk oranı elde etmişlerdir (Adoma ve ark., 2020).

2021 yılında yayınlanan bir çalışmada, Güven, Twitter üzerindeki tweetler üzerinde BERT modelleri ve makine öğrenmesi yöntemleri kullanarak duygu analizi gerçekleştirmiş ve elde edilen sonuçları karşılaştırmıştır. Veri seti sinirli, mutlu, üzgün, şaşkın ve korkunç etiketlerinden oluşan 5 farklı duygusal sınıf içermektedir. Bert modellerinden Bert-T modeli ile %98.75 doğruluk değeri elde edilirken makine yöntemlerinden LR yönteminde %98.4'lük doğruluk değeri elde edilmiştir. BERT modellerinin duygu analizine olumlu etkisini çalışmada göstermiştir (Güven, 2021).

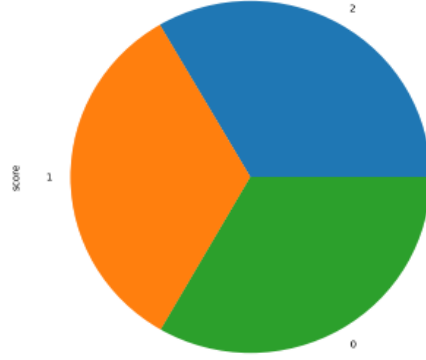
Açıkalin ve arkadaşları, pozitif ve negatif etiketli film ve otel yorumlarından oluşan veri seti üzerinde BERT modellerini kullanarak Türkçe duygu analizi yapmışlardır. Film veri setinde mevcut çalışmalardan daha yüksek doğruluk elde edildiğini gösterirken otel veri seti, film yorumları ile eğitildiğinde mevcut çalışmalardan daha düşük doğruluk oranına sahip olduğu belirtilmiştir (Açıkalin ve ark., 2021).

3. Materyal ve Metot

3.1 Veri Seti

"TRSAv1" (Aydoğan, 2022) veri seti, Aydoğan tarafından oluşturulmuş ve toplamda 150000 veriden oluşan bir veri setidir. Veri seti, e-ticaret sitelerindeki kullanıcı ürün yorumlarından elde edilmiştir. Bu veri seti, pozitif, nötr ve negatif olmak üzere 3 sınıfa ayrılmıştır ve her sınıfta 50000 veri bulunmaktadır. Veri setindeki etiketlenmiş veri sayıları eşit tutularak dengeli bir veri seti oluşturulması amaçlanmıştır. Şekil 1'de veri setindeki etiket dağılımı gösterilmektedir.

Veri etiketleri 0,1 ve 2'den oluşmaktadır. 0 “nötr”ü, 1 “negatif”i, 2 ise “pozitif”i temsil etmektedir.



Şekil 1. Veri setindeki etiket dağılımı

Veri seti 3 sütundan oluşmaktadır. Veri setinde yer alan "id" sütunu, her bir verinin benzersiz bir numarası olduğunu gösterir. Her bir numara bir veri örneğinin diğer örneklerden ayırt edilebilmesi için kullanılır. "Score" sütunu, veri örneklerinin hangi sınıfa ait olduğunu belirtir. Burada yer alan "0", "1" ve "2" etiketleri kullanıcı yorumlarının sırasıyla "nötr", "negatif" ve "pozitif" sınıflarına ait olduklarını belirtmek için kullanılır. "Review" sütunu, kişilerin duygu, düşünce veya deneyimlerini aktardıkları alanı temsil etmektedir. Şekil 2 veri seti örneğini göstermektedir.

id	score	review
59770	0	aldim ama hep kullanmıyorum, uzun süre faydası...
21362	2	Saçı yumuşacık yapıyo ayrıca uzamasında da etk...
127324	0	Kaliteli çok güzel bir ürün öneriyorum kuru s...
140509	0	Kokusu inanılmaz güzel
144297	1	Neden tırtıklı olduğunu anlayamadım. Heyerim ç...
96370	1	Resimdeki kadar iyi durmuyo iade ettim
49742	1	Çok sert asla tavsiye etmem defalarca üstünde...
47389	2	kokusu güzel.nemlendirici ligi de

Şekil 2. Veri seti örneği

Tablo 1, veri setinin eğitim, doğrulama ve test olarak üç bölüme ayrıldığını ve her bir bölümdeki veri ve sınıf dağılımını göstermektedir. 150000 veriden oluşan veri setinin eğitim bölümünde toplamda 75000 veri bulunmaktadır. Bu verilerin 25128'i pozitif, 24907'si nötr ve 24965'i negatif olarak etiketlenmiştir. Doğrulama bölümünde 37500 veri bulunurken, bunların 12420'si pozitif, 12559'u nötr ve 12521'i negatif olarak etiketlenmiştir. Test bölümünde de 37500 veri yer alır ve bu verilerin 12452'si pozitif, 12534'ü nötr ve 12514'ü negatif olarak etiketlenmiştir. Bu dağılım, veri setinin her bir bölümde dengeli bir sınıf dağılımına sahip olduğunu göstermektedir.

Tablo 1. Veri setinin eğitim, doğrulama ve test dağılımı

Bölüm	Veri Sayısı	Pozitif Sayısı	Nötr Sayısı	Negatif Sayısı
Eğitim	75000	25128	24907	24965
Doğrulama	37500	12420	12559	12521
Test	37500	12452	12534	12514

3.2 Metot

Bu bölümde, verinin ön işleme adımları, analizi ve kullanılan transformers tabanlı mimariler sunulmaktadır. Yapılan çalışmada model değerlendirme ölçütü olarak doğruluk ve F1 Skor metrikleri kullanılmıştır.

Deneylerin gerçekleştirilmesinde, programlama dili olarak Python kullanılmış, kodlama ortamı olarak da "Colab" tercih edilmiştir.

3.2.1 Veri Ön İşleme

E-ticaret sitelerindeki kullanıcı yorumlarından elde edilen veri seti üzerinde ön işleme adımları uygulanmıştır. Bu adımlar arasında durak kelimelerin (stopwords) kaldırılması, noktalama işaretlerinin ve özel karakterlerin kaldırılması, gereksiz boşlukların temizlenmesi gibi işlemler bulunmaktadır. Bunlara ek olarak güçlü-etkili kelimelerin çıkarılması (removing strong words) işlemi uygulanmıştır.

Durak kelimeler (Stopwords): Herhangi bir dilde sıklıkla kullanılan ve metin içerisinde bir anlam ifade etmeyen kelimelere durak kelimeler (stopwords) denir (Çoban, 2021).

Veri setinde yaygın olarak geçen durak kelimelerin örnekleri şunlardır: bir, ve, veya, gibi, için, ile. Bu kelimeler veri setinden çıkarılmıştır.

Etkili-yoğun Kelimeler (Strong words): "Strong words" terimi, bir kelimenin etkisini ve duygusal yoğunluğunu tanımlamak için kullanılır. Bu kelimeler genel olarak tanımlayıcı, güçlü veya ilgi çekici bir anlam taşırlar. İnsanların duygusal tepkilerini tetikleyebilir, inandırıcılığı artırabilir veya bir mesajın etkisini artırabilirler. İnsanlar üzerinde güvenilirlik oluşturabilir, mesajın veya duyguların etkisini arttırabilirler.

Şekil 3'te, kullanıcı yorumlarındaki güçlü duygusal ifadeleri temsil eden kelimeler listelenmiştir.

['çok', 'güzel', 'beğenmedim', 'iyi', 'gibi', 'muhteşem', 'hiç', 'gereksiz', 'ortalama']

Şekil 3. Etkili-yoğun Kelimeler

Yapılan çalışmada "strong words" listesinde geçen "çok", "güzel", "beğenmedim", "iyi", "gibi", "muhteşem", "hiç", "gereksiz" ve "ortalama" gibi kelimeler çıkarılmıştır.

3.2.2. Kullanılan Transformers Tabanlı Mimariler

Transformer kavramı ilk olarak 2017 yılında Toronto Üniversitesi ve Google'daki araştırmacılar tarafından geliştirildi ve "Attention is all you need" makalesi (Vaswani ve arkadaşları, 2017) ile tanıtıldı. Bu model, bir derin öğrenme modeli olup yaygın olarak doğal dil işleme alanında soru cevaplama, metin özetleme, sınıflandırma gibi görevlerde kullanılmaktadır.

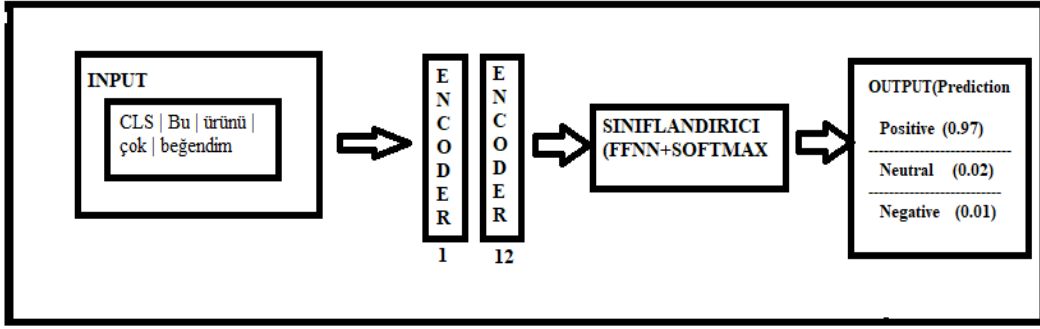
Bu bölümde, duygu sınıflandırması için kullanılan Transformers tabanlı mimariler anlatılacaktır. Bu modeller arasında BERTurk (cased, 32k), BERTurk (uncased, 32k), BERTurk (cased, 128k), BERTurk (uncased, 128k), DistilBERTurk (cased), ConvBERTurk (cased), ConvBERTurk mC4 (cased), ConvBERTurk mC4 (uncased), ELECTRA Base mC4 (cased), ELECTRA Small (cased) yer almaktadır.

BERTTurk: BERTurk, Türkçe metinler üzerinde doğal dil işleme işlevleri için eğitilmiş olan oldukça kapsamlı bir dil modelidir. "Cased", büyük/küçük harf duyarlı, "uncased" ise büyük/küçük harf duyarsız BERTurk versiyonlarıdır. "Cased" ve "uncased" versiyonlarının 32000 ve 128000 kelime parçacığı boyutlarında eğitilmiş 2 farklı eğitim modeli vardır.

DistilBERT: DistilBERT modeli BERT modelinden çıkan öğrenmeyi sıkıştırarak daha küçük bir boyuta getirir. Bu sayede daha az bellek kullanımı ve daha hızlı tahmin süreleri gibi avantajlar sağlamaktayken, daha az veri ile çalışmasından dolayı BERT'e göre daha düşük bir performans sunar (Huggingface, 2016).

ConvBERTurk: Öz-dikkat bloğu tabanlı yapısı nedeniyle, BERTurk hafıza ve maliyet açısından dezavantajlıdır. Yerel bağımlılıkları daha iyi modellemek için önceden eğitilmiş ConvBERTurk önerilmiştir (Savcı ve arkadaşları, 2023). ConvBERTurk modelinin varyasyonlarından olan ConvBERTurk mC4, eğitim için Maksimum Çıkarım Kümesi (Maximum Credible Corpus) olarak adlandırılan büyük bir veri setinin kullanılması ile oluşturulan bir modeldir.

ELECTRA: ELECTRA da BERT gibi doğal dil işleme alanında kullanılan bir modeldir. Bu modeli BERT'ten ayıran temel fark eğitim aşamasındaki yaklaşımıdır. BERT önceden eğitilmiş bir dil modeli olarak büyük bir metin veri kümesi üzerinde maskeleyip yaparken, ELECTRA metni bozma ve geri kazanma metodu ile çalışır. ELECTRA modelinin varyasyonlarından olan ELECTRA based mC4 daha büyük veri kümesi ile eğitilmişken ELECTRA small daha küçük veri kümesi ile eğitilmiştir.



Şekil 4. Bert model sınıflandırma

Şekil 4 basit bir sınıflandırma sürecini göstermektedir. Encoder yığın işlemi, cümle kelimelerini paralel olarak işler. Bir encoder, self-attention bloğunu ve feed-forward sinir ağını içerir.

4.Bulgular ve Tartışma

Bu çalışmada 150000 adet ürün yorumu içeren veri seti, yukarıda belirtilen modeller üzerinde farklı metotlar kullanılarak sınanmış ve sonuçların karşılaştırılması yapılmıştır.

Yapılan tüm deneyler sonucunda elde edilen bulgular Tablo 2’de gösterilmiştir. İlk olarak veri seti üzerinde hiçbir ön işlem yapılmadan, belirtilen modeller üzerinde deneyler gerçekleştirilmiştir. Bu deney sonuçlarına göre ön işlemez gerçekleştirilen deneylerde en yüksek doğruluk oranı 82.89(%) ile ConvBERTurk mC4(uncased) modelinde, en düşük değer ise 81.46 ile DistilBERTurk (cased) modelinde gözlemlenmiştir. F1 score metriğine göre ise en yüksek değer yine ConvBERTurk mC4(uncased) modelinde görülmüştür.

Diğer bir yöntemde, ön işleme adımlarından noktalama işaretlerinin kaldırılması, durak kelimelerin kaldırılması, özel karakterlerin kaldırılması işlemleri uygulanmıştır. Veri seti ön işlemeli adımdan geçirildikten sonra en yüksek başarımlı oranı 83.05(%) ile ConvBERTurk mC4(uncased) modeli tarafından sağlandığı gözlemlenmiştir.

Uygulanan son yöntemde ise etkili-yoğun kelimelerin kaldırılması işlemi gerçekleştirilmiştir. Bu yöntemde en yüksek başarımlı oranı 81.32(%) doğruluk oranıyla ConvBERTurk mC4(cased) modeli tarafından sağlandığı tespit edilmiştir.

Tablo 2. Deney sonuçları

Model	Ön işlemez (%)		Ön İşlemeli (%)		Etkili-Yoğun Kelimelerin Kaldırılması (%)	
	Doğruluk	F1	Doğruluk	F1	Doğruluk	F1
BERTurk(cased, 32k)	82.30	82.14	82.87	82.91	80.70	80.75
BERTurk (uncased, 32k)	82.40	82.33	82.94	82.93	80.84	80.94
BERTurk (cased, 128k)	82.62	82.57	82.54	82.55	80.80	80.84
BERTurk(uncased, 128k)	82.43	82.31	82.70	82.71	80.59	80.61
DistilBERTurk (cased)	81.46	81.47	81.40	81.46	79.58	79.74
ConvBERTurk (cased)	82.70	82.64	83.02	83.06	80.90	80.96
ConvBERTurk mC4 (cased)	82.70	82.64	83.04	83.08	81.32	81.41
ConvBERTurk mC4 (uncased)	82.89	82.79	83.05	83.08	81.11	81.20
ELECTRA Base mC4 (cased)	82.42	82.32	82.69	82.73	80.81	80.86
ELECTRA Small (cased)	80.98	80.92	80.91	80.93	78.82	78.85

Tablo 2’de gösterilen tüm sonuçlar üzerinde genel bir değerlendirme yapıldığında; veri seti üzerinde ön işlem uygulandığında ön işlemez yöntemde göre daha yüksek doğruluk oranına sahip olduğu görülmektedir. Diğer modellere kıyasla ConvBERTurk modelleri, bu veri seti üzerinde genel olarak daha iyi performans göstermektedir.

Tüm modeller için etkili-yoğun kelimelerin kaldırılması yönteminin duygu sınıflandırma performansı üzerinde negatif bir etkisi olduğu görülmüştür. En düşük başarımları ise DistilBERTTurk (cased) modelinde görülmüştür.

5. Sonuçlar

Gerçekleştirilen bu çalışmada, e-ticaret sitelerindeki kullanıcıların Türkçe yorumlarından oluşan veri seti üzerinde transformer tabanlı mimariler kullanılarak duygu sınıflandırması yapılmıştır. Çalışmanın sonuçlarına göre modeller arasında karşılaştırma yapılmış ve duygu sınıflandırma performansı doğruluk oranı ve F1 skor metrikleri kullanılarak değerlendirilmiştir. Ayrıca bu çalışma, Türkçe dilinde duygu sınıflandırması için transformers tabanlı mimarilerin kullanılabilirliğini de ortaya koymaktadır.

İleride yapılacak çalışmalarda, veri sayısı artırılarak transformers tabanlı mimariler ile geleneksel makine öğrenme tekniklerinin performanslarının karşılaştırılması hedeflenmektedir.

6. Kaynaklar

Kaynar, O., Görmez, Y., Yıldız, M., & Albayrak, A. (2016, September). Makine öğrenmesi yöntemleri ile Duygu Analizi. In International Artificial Intelligence and Data Processing Symposium (IDAP'16) (Vol. 17, No. 18, pp. 17-18).

Köksal, Ö. (2021, June). Enhancing Turkish sentiment analysis using pre-trained language models. In 2021 29th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.

Adoma, A. F., Henry, N. M., & Chen, W. (2020, December). Comparative analyses of bert, roberta, distilbert, and xlnet for text-based emotion recognition. In 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP) (pp. 117-121). IEEE.

Güven, Z. A. (2021, September). Comparison of BERT models and machine learning methods for sentiment analysis on Turkish tweets. In 2021 6th International Conference on Computer Science and Engineering (UBMK) (pp. 98-101). IEEE.

Acikalın, U. U., Bardak, B., & Kutlu, M. (2021). BERT modeli ile türkçe duygu analizi.

Aydoğan, (2022, February). TRSAv1: A new benchmark dataset for classifying user reviews on Turkish e-commerce websites. <https://journals.sagepub.com/doi/abs/10.1177/01655515221074328> . Erişim Tarihi: 10 Haziran 2023

Çoban, Ö., Özyer, B., & Özyer, G. T. (2015, May). Sentiment analysis for Turkish Twitter feeds. In 2015 23rd Signal Processing and Communications Applications Conference (SIU) (pp. 2388-2391). IEEE.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

DistilBERT . https://huggingface.co/docs/transformers/model_doc/distilbert . Erişim Tarihi: 17 Temmuz 2023

Savcı, P., & Das, B. (2023). Comparison of pre-trained language models in terms of carbon emissions, time and accuracy in multi-label text classification using AutoML. Heliyon, 9(5).