



Generating headlines for Turkish news texts with transformer architecture based deep learning method

Abdulkadir Karaca*^{ID}, Özlem Aydın^{ID}

Department of Computer Engineering, Faculty Engineering, Trakya University, 22020, Edirne, Türkiye

Highlights:

- Transformer architecture is used for automatic headline generation
- The success of the transformer based model is compared with the designed LSTM and GRU based models
- It has been observed that with the transformer, more appropriate headlines are generated for the content of the news

Keywords:

- Turkish natural language processing
- Automatic headline generation
- Abstract text summarization
- Deep learning
- Transformers

Article Info:

Research Article
Received: 06.07.2021
Accepted: 16.03.2023

DOI:

10.17341/gazimmfd.963240

Correspondence:

Author: Abdulkadir Karaca
e-mail:
akadirkaraca@outlook.com
phone: +90 530 110 7093

Graphical/Tabular Abstract

Models were created by LSTM, GRU and Transformer architecture on news headline generation. The ROUGE-1, ROUGE-2 and ROUGE-L measurements of the experiments performed on the validation data at the end of the training are shown in Table A. In general, it has been observed that the transformer architecture draws a successful result.

Table A. General experimental results of the corpus

		LSTM	GRU	Transformer
Rouge-1	F-1 Score	0.2917	0.3217	0.3358
	Precision	0.4667	0.3143	0.4792
	Recall	0.2456	0.3333	0.3016
Rouge-2	F-1 Score	0.1000	0.1818	0.2667
	Precision	0.1000	0.1667	0.2857
	Recall	0.1000	0.2000	0.2500
Rouge-	F-1 Score	0.2500	0.3217	0.3358
	Precision	0.3667	0.3143	0.4792
	Recall	0.2193	0.3333	0.3016

Purpose:

It is an important need to be able to reach the right content covering the desired information in a short time from the intense data that we are exposed to today. The headlines of the content have a positive effect on meeting this need. However, if there are misleading headlines, this can turn into a disadvantage. In this study, it is aimed to produce more directive Turkish headlines in the best summaries of Turkish texts by using current natural language processing and deep learning technologies.

Theory and Methods:

A corpus of approximately 125,000 Turkish news texts and headlines was used in the study. The news texts and headlines in the corpus were preprocessed in order to train the model correctly. In order to obtain more human-like and original results, general and abstract summarization techniques were preferred as solution methods. Transformer architecture-based encoder-decoder deep learning architecture, which was published in 2017 and accepted as state-of-the-art in natural language processing, was used as the training model. This model was developed using deep learning framework TensorFlow and Keras. In addition, models were created with LSTM and GRU architectures to compare the success of the Transformer architecture. LSTM and GRU models were trained through Google Colab, which supports high system requirements in deep learning studies, and results were produced. The transformer architecture, on the other hand, is trained via a remote computer with higher system capacity. All models were trained for 25 epochs.

Results:

With the trained models, loss function values of 1.03, 0.55 and 2.49, respectively, were obtained at the end of 25 epochs. The data reserved for validation in the corpus were produced with the trained model. In this experiment, ROUGE-1, ROUGE-2 and ROUGE-L metrics, which are generally preferred in summary tasks, were preferred as success criteria. The metrics measured for all models are shown in Table A. It has been observed that the transformer architecture is generally successful compared to other models.

Conclusion:

It has been observed that the transformer architecture has achieved successful results in terms of understanding the context and evaluation criteria compared to studies with other deep learning models.



Transformatör mimarisi tabanlı derin öğrenme yöntemi ile Türkçe haber metinlerine başlık üretme

Abdulkadir Karaca*^{ID}, Özlem Aydın^{ID}

Trakya Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, 22020, Edirne, Türkiye

Ö N E Ç I K A N L A R

- Otomatik başlık üretmek için Transformatör mimarisi kullanılmıştır
- Transformatör tabanlı modelin başarısı, tasarlanan LSTM ve GRU tabanlı modeller ile karşılaştırılmıştır
- Transformatör ile haberin içeriğine daha uygun başlıklar üretildiği gözlemlenmiştir

Makale Bilgileri

Araştırma Makalesi

Geliş: 06.07.2021

Kabul: 16.03.2023

DOI:

10.17341/gazimmfd.963240

Anahtar Kelimeler:

Türkçe doğal dil işleme,
otomatik başlık üretme,
soyutlayıcı metin özetleme,
derin öğrenme,
transformatör

ÖZ

Günümüzde internet insanların kolay erişebildiği, kolaylıkla ve kontrolden bağımsız içerik üretebildiği bir yapıdır. Bu durumla paralel olarak, büyük veriyi oluşturan ham verilerden bilgiyi elde edebilme yeteneği daha karmaşıklaşmıştır. İçeriklerin başlıklarının kontrolsüz ve yanıltıcı olabilecek unsurlar barındırması doğru bilgiye ulaşmayı zorlaştırmaktadır. İnsanların sınırlı zamanlarında, istedikleri bilgiye ulaşabilmesinde içeriklerin başlıkları önemlidir. Bu çalışmada, haber metinlerine yanıltıcı olabilecek başlıklar yerine, içeriğe uygun başlıklar üretilmesi amaçlanmıştır. Bu amaçla derin öğrenme yöntemiyle Türkçe haber metinlerine başlık üreten bir uygulama geliştirilmiştir. Veri seti olarak SuDer haber derlemi kullanılmıştır. Modelin eğitimi için günümüzde doğal dil çalışmalarında sıklıkla tercih edilen Transformatör mimarisiyle soyut özetleme yöntemi kullanılarak, haber başlıklarının üretiminde daha insansı sonuçlar elde edilmesi amaçlanmıştır. Bu çalışmada Transformatör modelinin başarımının karşılaştırılması için Uzun Ömürlü Kısa Dönem Belleği (Long Short-Term Memory, LSTM) ve Geçitli Yineleme Birimi (Gated Recurrent Unit, GRU) mimarileriyle modeller hazırlanmış ve eğitimleri yapılmıştır. Derlem üzerinde LSTM, GRU ve Transformatör mimarileriyle 25 dönem yapılan eğitim sonunda kayıp değerleri sırasıyla 1,03, 0,55 ve 2,49 olarak ölçülmüştür. Doğrulama verileri üzerindeki deneylerde ROUGE-1, ROUGE-2 ve ROUGE-L metrikleri üzerinden yapılan ölçümler değerlendirildiğinde, Transformatör mimarisinin kısmen iyi olduğu gözlemlenmiştir. Ayrıca bu mimarilerle üretilen başlıklara bakıldığında, Transformatör mimarisiyle üretilen başlıkların diğer mimarilere göre haber içeriğine kısmen daha uygun başlıklar ürettiği gözlemlenmiştir.

Generating headlines for Turkish news texts with transformer architecture based deep learning method

H I G H L I G H T S

- Transformer architecture is used for automatic headline generation
- The success of the transformer based model is compared with the designed LSTM and GRU based models
- It has been observed that with the transformer, more appropriate headlines are generated for the content of the news

Article Info

Research Article

Received: 06.07.2021

Accepted: 16.03.2023

DOI:

10.17341/gazimmfd.963240

Keywords:

Turkish natural language
processing,
automatic headline
generation,
abstract text summarization,
deep learning,
transformer

ABSTRACT

Nowadays, the Internet is a structure that people can access easily and at the same time produce content easily and without control. In parallel with this situation, the ability of extract information from the raw data that makes up big data has become more complex. The fact that the headlines of the contents contain uncontrolled and misleading elements makes it difficult to reach the right information. The headlines of the contents are important for people to reach the information they want in their limited time. In this study, it is aimed to produce headlines suitable for the content instead of headlines that may be misleading for news. For this purpose, an application that produces headlines for Turkish news with deep learning method has been developed. SuDer news corpus is used as dataset. For the training of the model, it is aimed to obtain more humanoid results in the production of news headlines by using the Transformer architecture, which is frequently preferred in natural language studies today and the abstract summarization method. In this study, in order to compare the performance of the Transformer model, models are prepared and trained with Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures. At the end of 25 epochs of training with LSTM, GRU and Transformer architectures on the corpus, the values of loss are 1.03, 0.55 and 2.49 respectively. In the experiments performed on the validation data, measurements are made with ROUGE-1, ROUGE-2 and ROUGE-L metrics. As a result of the measurements, it is observed that the Transformer architecture is partially good, based on the metric values produced. In addition, when the headlines produced with these architectures are examined, it is observed that the headline obtained with the Transformer architecture produce headlines that are partially more suitable for the news content compared to other architectures.

1. Giriş (Introduction)

Günümüzde çevrimiçi içeriklerin üssel olarak artması, beraberinde servis edilen bilgilerin de aynı oranda yanıltıcı bir şekilde karşımıza çıkmasına neden olmaktadır. Bu durum hem gereksiz bilgiye maruz kalma ihtimalini ortaya çıkarmakta, hem de gün içinde kısıtlı olan zamanı etkili kullanamama problemlerine sebep olmaktadır. Oluşan bu zaman kaybına ve yanlış bilgilendirme problemlerine bir dereceye kadar engel olabilmek için, metinlerin içindeki ham veriden bağlamı doğru ifade edebilecek yeteneğe sahip başlıkları üretebilecek sistemlere ihtiyaç duyulmaktadır. Otomatik başlık üreten bu sistemler ile herhangi bir araştırmayla ilgili elde edilmek istenilen bilgi için hangi kaynakların kullanılıp kullanılmayacağına karar verme sorununun çözümü de kolaylaşacaktır.

Başlık üretmenin bir çeşit metin özetleme görevi olduğu düşünülebilir. Sonuç olarak da literatürdeki başlık üretme ile ilgili problemlerin genelinde de özetleme yöntemleri temel alınarak çözümler üretilmektedir. Özetleme yöntemleri ile çıkarıma dayalı ve soyut özetleme yapılabilmektedir. Haber metinlerindeki mevcut cümleleri belirli kurallara göre puanlayarak, bağlamı açıklama yeteneğine sahip olduğu tahmin edilen cümlelerin seçilmesi ile elde edilen sonuçlara çıkarıma dayalı özetleme yöntemleri ile ulaşılır. Soyut özetleme yöntemleri ile metin bağlamını ifade edebilecek ve metnin içinde birebir bulunmayan, aynı zamanda insan kullanım diline daha yakın ve özgün cümleler elde etmemize olanak sağlayacak özetler üretmek mümkündür.

Geçmişten günümüze metinlere başlık üretmek için çok farklı yöntemler önerilmiştir. Önerilen yöntemlerin tümü metin bağlamına bağlı, kapsamlı ve morfolojik açıdan okunabilir başlıkları üretmeyi hedeflemiştir. Üretilen başlık belirginlik, uzunluk, tutarlılık ve gramere uygun olmasının yanı sıra gereksiz sözcükleri de barındırmaması gibi kriterler göz önünde bulundurularak değerlendirilir. Başlık üretme problemlerinde genel çözüm olarak kullanılan metin özetleme çalışmalarında, İngilizce dili için 1950'lerden bugüne bilgisayar bilimleri alanında çok sayıda farklı yöntem önerilmiştir. Başlangıçta sunulan çözümlerde çıkarıma dayalı özetleme yöntemlerinin kullanıldığı gözlemlenmektedir. Metin özetleme çalışmalarının başladığı dönemlerde Edmunson [1] ve Luhn [2] gibi araştırmacılar, sözcük frekansı, cümlelerin gramer yapısı ve cümleleri puanlamakta kullanılan denklemler için fayda sağlayacak, önceden belirlenmiş sözlükler ile kurula bağımlı yöntemler geliştirmişlerdir. 2004 yılındaki başka bir çalışmada, aynı bağlam ile ilgili farklı dokümanlardan elde edilen farklı adetlerdeki cümlelerin, sözcük merkezli olacak şekilde aralarında benzerlik matrisi oluşturulmuştur. Seçilen cümlelerin düğüm olduğu, aralarındaki bağlantıların ise kenar olarak gösterildiği grafik tabanlı bir yapı ile çözüm sunulmuşlardır. Özeti oluşturacak cümleler, bu grafik tabanlı yapı içinden belirlenen kenar eşik değerine göre seçilerek elde edilmektedir [3]. 2014 yılına kadar doğal metin üretmedeki zorlukları nedeniyle soyut özetleme yöntemleri fazla popüler değildi. Ancak bilgisayar sistemlerinin karmaşık hesaplamaları yapabilecek kaynaklara sahip olmasıyla, makine öğrenmesi ve derin öğrenme yöntemleri de özetleme çalışmalarında kullanılmaya başlanmıştır. Böylelikle soyut özetleme çalışmaları ağırlık kazanmıştır. Özellikle kodlayıcı-kod çözücü modellerin özetleme çalışmalarında kullanımının başarılı sonuçlar verdiği görülmektedir. Başlık üretme üzerine yapılan çalışmalardan birinde haber metinlerine başlık üretmek için Uzun Ömürlü Kısa Dönem Belleği (Long Short-Term Memory, LSTM) ve dikkat mekanizması içeren bir kodlayıcı-kod çözücü Yinelemeli Sinir Ağı (Recurrent Neural Network, RNN) mimarisi oluşturulmuştur [4]. Takase vd. [5] tarafından geliştirilen bir uygulamada ise sözdizimsel ve anlamsal bilgiyi yapısal olarak içeren dikkat tabanlı bir kodlayıcı-kod çözücü model tasarlanmıştır. Önerilen

modelin diğer dikkat tabanlı kodlayıcı-kod çözücü modellere göre başlık üretmede daha başarılı sonuçlar elde edildiği gözlemlenmiştir. Nöral uçtan uca (sequence-2-sequence, seq2seq) model ile başlık üreten bir diğer çalışmada üretim birbirini takip eden iki aşamadan oluşmaktadır. İlk aşamada metin özetleme yöntemleri ile metnin önemli cümleleri seçilmekte, takip eden aşamada hiyerarşik dikkate dayalı çoklu cümle özetleyebilen bir modelden faydalanılmaktadır [6]. Çalışmanın deneysel sonuçlarına göre nöral metin özetleme yöntemleri başlık üretme görevinde oldukça başarılı sonuçlar vermiştir. Murao vd. [7] bir haber toplayıcının editöryal sürecine katkı sağlamak amacıyla haber başlıklarına kısa başlıklar üreten bir model sunmuşlardır. Önerilen model, birden fazla bilgi kaynağını ayrı ayrı kodlamak için birden çok kodlayıcıya sahip kodlayıcı-kod çözücü bir modeldir ve sundukları bu hibrit model ile özellikle yararlılık ölçüsü kullanılarak diğer temel çözümlere göre daha başarılı sonuçlar elde etmişlerdir.

Literatürde başlık üretmede özellikle son yıllarda Transformatör mimarisine dayanan modellerle başarılı sonuçlar elde eden bazı çalışmalar bulunmaktadır. Gavrilov vd. [8] Transformatör mimarisinin modifiye edilmiş bir modeli olan Universal Transformer mimarisini kullanarak, New York Times Annotated derlem üzerinde yaptıkları soyut özetlemeye dayalı başlık üretiminde başarılı sonuçlar elde etmişlerdir. Liao vd. [9] Standart Transformer (XL)-CC adını verdikleri, Transformer-XL [10] modelindeki uzun vadeli bağımlılık yakalama yeteneğini kullanarak İngilizce ve Çince haber derlemi üzerinde uyguladıkları modelde önemli değerlere ulaştıklarını belirtmişlerdir. Yine başka bir çalışmada önceden eğitilmiş Transformatör tabanlı modellerden mBart ve BerSumAbs'de başlık üretme görevinin etkinliğini, Rusça haberlerden oluşan RIA and Lenta derlemeleri için analiz etmişler ve bu dil için önemli sonuçlara ulaşılmıştır [11]. Gu vd. [12], haber metinlerine kısa ve bilgilendirici başlıklarla üretmeyi hedefleyen yaklaşımlarında, çok seviyeli ön eğitilmiş çatı (framework) yapısı kurmuşlardır. Bu yapıda, mevcut haber metinleri üzerinden uzaktan denetimle diğer farklı haber metinlerinden bilgi aktararak eğitim için daha büyük bir veri seti oluşturulmuştur. Ayrıca haber metinlerindeki önemli bilgiyi çıkarmak için kendi kendine oylamaya dayalı yeni bir dikkat katmanı modeli geliştirmişlerdir. Oluşturdukları bu çatı yapısı ve Transformatör mimarisini içeren modeli, kodlayıcı-kod çözücü bir model ile birleştirerek başarılı bir sonuç elde etmişlerdir.

Türkçe dili için derin öğrenme ile gerçekleştirilmiş haber başlığı üreten tek çalışma Karakoç ve Yılmaz'a [13] aittir. Kendi hazırladıkları bir derlem ile LSTM mimarisini kullanarak soyut, yoruma dayalı bir model oluşturmuşlardır. Kullandıkları derlem üzerinde bazı ön işlemler yaptıktan sonra vektörleştirme işlemi için Facebook AI Research tarafından eğitilmiş FastText kütüphanesinin gömme vektörlerini (word embedding) kullanmayı tercih etmişlerdir. Türkçe dilinde başlık üretme görevi için ihtiyacı karşılayabilecek yeterli içeriğe sahip bir derlemin mevcut olduğunu söylemek zordur. Yapılan çalışmanın ise bu koşullara rağmen iyi sonuçlar elde ettiği söylenebilir.

Yapılan tüm bu çalışmalar ile, derin öğrenme mimarilerinin günümüz bilgisayarların matematiksel ve istatistiksel hesaplama gücü dahil edilerek, metin özetleme ve başlık üretme problemleri için doğal dile yakın çözümler sunduğu görülmektedir. Bütün derin öğrenme mimarileri, birbirlerinin eksik yönlerini operasyonel açıdan veya performans açısından tamamlamak üzere tasarlanmışlardır [14]. Derin öğrenme mimarileriyle, Doğal Dil İşleme (DDİ) uygulamalarından olan görüntü alt yazısı üretme [15] ve finansal duygu analizi [16] gibi farklı alanlarda da modeller kurulabilmektedir. Yine bu mimarilerin, DDİ dışındaki alanlarda da başarılı sonuçlar verdiği yapılan farklı disiplinlerdeki çalışmalardan görülmektedir [17, 18].

Bu çalışmada Türkçe dilinde hazırlanmış haber metinlerine başlık üretmeyi hedefleyen bir model geliştirilmiştir. Geliştirilen modelde haberin içeriğine uygun, yanıltıcı bilgi taşımayan haber başlığı üretimi hedeflenmiştir. Literatürde özellikle son yıllarda derin öğrenme çalışmalarında başarılı sonuçlar elde ettiği gözlemlenmiş olan Transformatör mimarisi geliştirilen model tasarımında kullanılmıştır [21]. Bu çalışma, Türkçe dili için haber başlığı üretmede Transformatör mimarisini kullanan ilk çalışmadır ve bu yönüyle literatüre önemli bir katkı sağlamaktadır. Ayrıca çalışmada modelin başarımını kanıtlamak üzere LSTM ve Geçitli Yineleme Birimi (Gated Recurrent Unit, GRU) mimarileri ile de model eğitimleri yapılmış ve sonuçları karşılaştırılmıştır. SuDer haber derleminden elde edilen 125.000 haber metni kullanılarak geliştirilen LSTM, GRU ve Transformatör mimarileri ile 25 dönem (epoch) eğitim sonunda kayıp (loss) fonksiyonu değeri sırasıyla 1,03, 0,55 ve 2,49 olarak ölçülmüştür.

2. Teorik Metot (Theoretical Method)

Başlık üretimi gibi doğal dilin anlaşılması ve üretilmesi ile ilgili problemlerin çözümünde, anlaşılması gereken bağlam ve üretilmesi istenilen çıktı için ihtiyaç duyulan sözlüğün öncelikle oluşturulması gerekmektedir. Bu sözlük, derlem içindeki tekrar eden veya etmeyen her sözcüğü veya belirli sayıda sözcüğü eşsiz olarak temsil etmelidir. Bu çalışmada sözlük oluşturulurken Türkçe'nin morfolojik yapısını ifade etmekteki zorluklar ve semantik olarak anlaşılır çıktılar üretilebilmek için, kullanılan derlemdeki sözcükler üzerinde herhangi bir işlem (sözcük kökü bulma vb.) yapılmamıştır. Geliştirilen sistemin ürettiği başlıkların bağlamı en iyi şekilde ifade edebilmesi için başlık kavramını karşılayabilecek sonuçlar üretebilen özetleme teknikleri kullanılmıştır. Özetleme teknikleri genel olarak girdiye göre, bağlama göre ve çıktıya göre olmak üzere sınıflandırılabilir. Sistem ilgili haber ile tek kaynaktan beslendiği için girdiye göre tekli doküman özetleme; özetlenecek haberlerin içerikleri belirli bir konuya bağlı olmadığı ve sorguya bağlı özetler olmadığı için bağlama göre genel özetleme yöntemi; çıktıya göre ise daha insansı ve özgün sonuçlar elde edebilmek amacıyla soyut özetleme teknikleri birlikte kullanılarak çözüm sunulmaya çalışılmıştır [19].

Makine öğrenmesi ve derin öğrenme alanlarında en önemli parametre veridir. İlgili problemin çözümü için konuyla ilgili yeterli verinin edinilmesi gerekmektedir. Geliştirilen uygulamada Şen ve Yanıkoğlu [20] tarafından oluşturulan SuDer haber derlemi tercih edilmiştir. Bu derlem, üzerinde gerekli ayıklama ve ön işlemler yapılarak çözüm sunulmak istenen problem için daha uygun ve etkili bir alt derlem oluşturularak kullanılmıştır. Mevcut çalışmada elde edilen derlemin eğitimi için doğal dil gibi sıralı verilerin eğitilmesi için kullanılan derin öğrenme mimarilerinin (RNN, LSTM ve GRU) yerine, 2017 yılında "Attention is All You Need" başlıklı makale ile birlikte önerilen ve DDİ ile ilgili çalışmalarda en son teknoloji (state-of-the-art) olarak kabul edilen Transformatör mimarisi kullanılmıştır [21]. Bu tercihte literatür araştırması aşamasında karşılaşılan Google Çeviri'nin Transformatör mimarili hibrit bir modele geçmesi, GPT-2 (Generative Pre-trained Transformer 2) ve GPT-3 (Generative Pre-trained Transformer 3) gibi modellerin Transformatör tabanlı mimari kullanmaları en büyük etken olmuştur [22-24].

2.1. Veri Kümesi (Dataset)

Bu çalışmada derlem olarak Türkçe haberlerde belge sınıflandırması problemi için oluşturulmuş SuDer Türkçe haber derlemi tercih edilmiştir. Bu derlem 2010 ile 2017 yılları arasındaki iki ayrı gazeteden elde edilen haber metni, haber başlığı, haberin kategorisi ve haber tarihi bilgilerini içeren yaklaşık olarak 700.000 içerikten oluşmaktadır. SuDer derlemindeki bütün içerikler bir metnin kategorisini tahmin etme için uygun olsa da doğal dil anlayışı ve doğal

dil üretimi için aynı performansı göstermekte yeterli değildir. Bu nedenle derlemdeki içeriklerden, başlık üretme ve metin özetleme görevleri için uygun olabilecek metinler ve başlıklar ayıklanmıştır. Derlemi oluşturan gazetelerden biri problemin çözümü için genel olarak haber başlığı ve metni açısından yeterli olmadığından en başta elenmiştir. Bu yetersizliği oluşturan nedenler şunlardır: haber başlıkları ve gövdelerinin gereksiz bilgi barındırması, köşe yazılarının çok uzun haber gövdesi içermesi ve haber başlıklarının da çoğunlukla tek sözcükten oluşması veya haber bağlamından alakasız mecaz anlam taşıyan başlıklara sahip olmasıdır. Gerekli ayıklamalar neticesinde bu çalışmada kullanılan gazete haberlerinin kategorik olarak toplam doküman sayıları, toplam sözcük sayıları ve kategorik olarak haberlerin ortalama sözcük uzunlukları Tablo 1'de gösterilmiştir.

Tablo 1. Kullanılan gazetenin kategorik istatistikleri
(Categorical statistics of the newspaper used)

Kategori	Toplam doküman sayıları	Sözcük sayıları	
		Toplam	Ortalama
Gündem	143.842	35.749.880	248,54
Yaşam	123.086	22.878.732	180,86
Ekonomi	85.485	22.261.600	247,38
Yazarlar	681	16.335.364	239,87
Toplam	420.513	95.494.110	227,09

Derlem için seçilen bu gazetenin, başlığa sahip olup haber gövdesi olmayan resimli haber içerikleri ile aynı başlığa veya haber gövdesine sahip olan satırları derlemden atılmıştır. Yine haber gövdesinde sadece "Haberin Devamı..." ve "Ayrıntılar Geliyor..." ifadelerini barındıran haberler derlemden çıkartılmıştır. Böylece derlemin haber ve başlık metinlerinin toplam sayıları, ortalama sözcük uzunlukları, haber ve başlık uzunluklarının standart sapma, minimum, maksimum sayısal değerleri Tablo 2'de verilen genel sayısal istatistiksel sonuçlara sahip olmuştur. Elimizde böyle bir derlemi eğitebilecek bir makine olmadığı için genel derlemin üçte birini kapsayacak şekilde ilk 125.000 haber ve başlığı seçilmiştir. Eğitim için seçilen derleme ön işlemler uygulandıktan sonraki derlemin haber ve başlık metinlerinin toplam sayıları, ortalama sözcük uzunlukları, haber ve başlık uzunluklarının standart sapma, minimum, maksimum sayısal değerleri gibi elde edilen sayısal istatistikler ise Tablo 3'te gösterildiği gibidir. Ayrıca eğitime girecek derlem, eğitim ve %10 doğrulama verisi olacak şekilde ayrılmıştır.

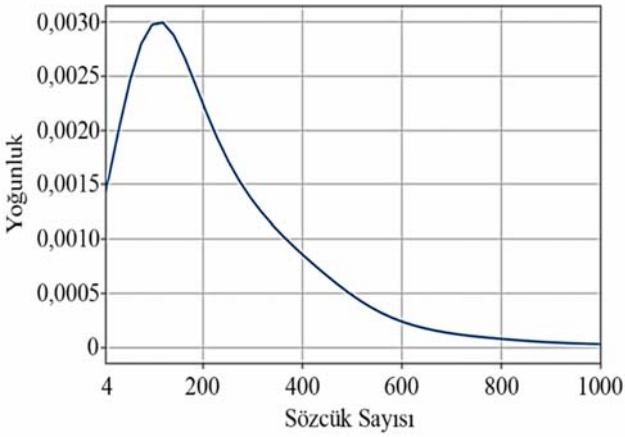
Tablo 2. Orijinal derlemin sayısal istatistikleri
(Numerical statistics of the original corpus)

	Haber	Başlık
İçerik sayısı	357.962	
Ortalama sözcük	225,19	4,86
Ortalama sözcük (%75)	297	6
Standart sapma	229,15	1,49
Minimum sözcük	4	3
Maksimum sözcük	10.969	23

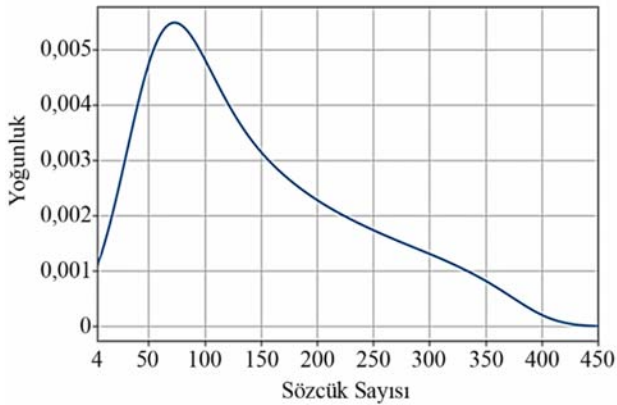
Tablo 3. Ayıklanmış derlemin sayısal istatistikleri
(Numerical statistics of the extracted corpus)

	Haber	Başlık
İçerik sayısı	125.000	
Ortalama sözcük	174,40	5,20
Ortalama sözcük (%75)	230	6
Standart sapma	169,65	1,59
Minimum sözcük	4	3
Maksimum sözcük	7.095	23

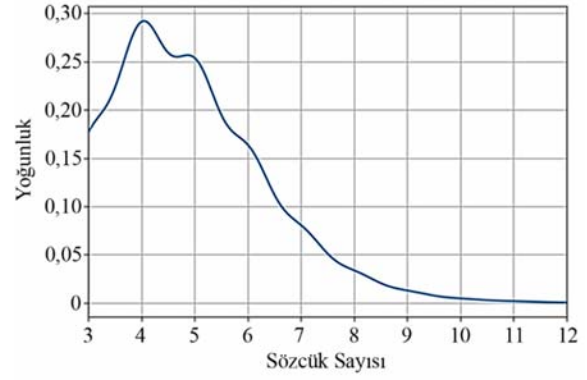
Haber gövdesinin uzunluğu için seçilen sayısal parametre, haber başlıklarının uzunluk parametresi ve eğitim veri kümesinin ayıklanması aşamasındaki diğer sayısal parametre değerleri belirlenirken gözlemsel, deneysel ve istatistiksel veriler değerlendirilerek seçilmiştir. Şekil 1’de genel derlemdeki haberlerin sözcük uzunluklarına göre toplam derlem içindeki yoğunluk dağılım grafiği gösterilmektedir. Eğitim için modele beslenecek olan haber gövdesinin ve başlıklarının uzunluklarının statik olarak belirlenmesi gerekmektedir. Bu amaçla, Tablo 3’teki ortalama sözcük sayısı ve standart sapma değerlerinin toplamı şeklinde seçilen giriş dizi uzunlukları belirlendikten sonra elde edilen, eğitime giren haberlerin sözcük uzunluklarına göre ön işleme yapılmış toplam derlem içindeki yoğunluk dağılım grafiği ise Şekil 2’deki gibi sunulmuştur. Tablo 3’teki istatistiksel parametreler ile belirlenen değerler ile derlemin normal dağılıma (Gauss dağılımı) yakın bir grafik çizmesi amaçlanarak eğitim için daha uygun bir derlem oluşturulmaya çalışılmıştır. Şekil 3’te ise genel derlemdeki başlıkların sözcük uzunluklarına göre toplam derlem içindeki yoğunluk dağılım grafiği gösterilmektedir. Yukarıda bahsedildiği üzere eğitime girecek olan başlıkların dizi uzunluklarının statik olarak belirlenmesi için ortalama sözcük sayısı, standart sapma değerleri ve başlıklar için başlangıç-bitiş jeton (token) sayısı toplamı kullanılmıştır. Böylece eğitime girecek olan başlıkların sözcük uzunluklarına göre ön işleme yapılmış toplam derlem içindeki yoğunluk dağılım grafiği Şekil 4’teki gibi elde edilmiştir.



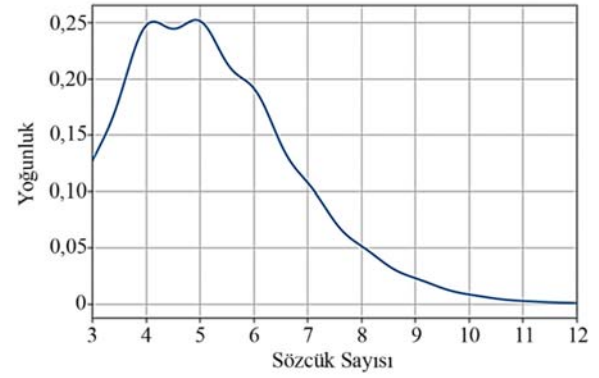
Şekil 1. Orijinal derlemin haber dağılım grafiği
(News distribution chart of the original corpus)



Şekil 2. Ayıklanan derlemin haber dağılım grafiği
(News distribution chart of the extracted corpus)



Şekil 3. Orijinal derlemin başlık dağılım grafiği
(Headline distribution chart of the original corpus)



Şekil 4. Ayıklanan derlemin başlık dağılım grafiği
(Headline distribution chart of the extracted corpus)

2.2. Ön İşlemler (Pre-Processings)

Bilgisayar sistemleri doğal dili insanların kullandığı şekliyle anlayamamaktadır. Bu sebeple de doğal dil bilgisayar sistemlerinin anlayabileceği deterministik şekillerde ifade edilmektedir. Doğal dilin bilgisayar sistemlerinde kullanılabilmesi için sayısal şekilde gösterimi gereklidir. Bu sayısal gösterimlerin derlemi ifade edebilmekte ne kadar güçlü olduğu aynı zamanda bilgisayar sistemlerindeki öğrenme modellerinin de doğal dili kullanmakta ve üretmekteki gücü ile doğru orantılıdır. Önerilen çözümün istenilen başarıyı elde edebilmesi için, derlemdeki ham verinin bilgiyi ne oranda ifade edebildiği önemli bir durumdur. Bu durumun en temel çözümü ise mevcut derlemin, bilgisayar biliminde “çöp girerse, çöp çıkar” prensibi ile uygun hale getirilmesidir. Bu sebeple modele besleyeceğimiz veriler, mevcut probleme uygun bir çözüm sağlayacak nitelikte bazı ön işlemlere ihtiyaç duymaktadır. Bu çalışmada kullanılan derlemde de haber gövdesi ve haber başlıkları için ayrı fonksiyonlar ile bu işlemler yapılmıştır. Bu fonksiyonlar sırasıyla şunlardır:

- Düzenli ifadeler (RegEx Kütüphanesi) kullanılarak noktalama işaretlerinin, arka arkaya olabilecek fazla boşluk veya iki sözcük arasında kalan tek karakterlerin filtrelenmesi.
- Ara başlıkların silinmesi.
- Haber metinlerindeki tarihlerin ve sayıların temizlenmesi, haber başlıklarındaki tarih ve sayıların ise “####” işareti ile temsil edilecek şekilde değiştirilmesi.
- Türkçe’deki frekansı en yüksek 10.000 sözcük içerisinde seçilen durak sözcüklerin (stop words)* metin içerisinde temizlenmesi. * Durak sözcüklerine <https://github.com/ahmetax/trstop> adresinden erişilebilir.

Bütün bu aşamalardan sonra, üç karakterden kısa sözcükler de metinlerden atılmış ve sonrasında tüm sözcükler küçük harfli olacak hale dönüştürülerek derlem eğitim için son halini almıştır. Orijinal haber gövdesi ve başlığı ile ön işlenmiş haber gövdesi ve başlığı için birer örnek sırasıyla Tablo 4 ve Tablo 5'te verilmiştir.

2.3. Transformatör Mimarisi (Architecture of Transformer)

Transformatör mimarisinin DDİ çalışmalarında kullanılan diğer mimarilerden en büyük farklarından biri verileri işlediği zaman adımındaki farklılıktır. Önceki RNN, LSTM ve GRU mimarilerinde sözcük dizileri, yani haber gövdesi, $t-1$ zaman adımından gelen gizli durum ile t zamanındaki girdi vektörünün işleme sokulması mantığı üzerine tasarlanmaktadır [25]. Transformatör mimarisinde ise bütün dizi aynı zaman adımında bir matris olarak modele girdi olarak alınır. Doğal dil üzerindeki çalışmalarda önceki mimariler ile kullanılan kodlayıcı-kod çözücü yapısı ise ortak olarak Transformatör mimarisinde de kullanılmaktadır. Yine RNN, LSTM ve GRU mimarilerindeki tekrarlayan sinir hücre yapılarının yerini, iki alt katmandan oluşan N adet Transformatör kodlayıcı ve kod çözücü blokları almaktadır. Transformatör mimarisi günümüzde, doğal dil ile alakalı problemlerin çözümünde RNN, LSTM ve GRU gibi mimarilere göre daha çok tercih edilir duruma gelmiştir. Bunun nedeni, çalışmalarda Transformatör mimarisi kullanılarak diğer mimarilere göre daha başarılı sonuçlar elde edilmesi ve mimari olarak hala geliştirilmeye devam edilmesidir. Transformatör mimarisinin bahsedilen önceki mimarilerden yapı olarak birçok farklılığı daha bulunmaktadır. Günümüzde mimari olarak farklı geliştirilmeler yapılmış, farklı Transformatör mimarileri de mevcuttur. Fakat bu çalışmada orijinal makalede de kullanılan ve gerekli yetenekleri sunabilen Şekil 5'teki temel Transformatör mimarisi tercih edilmiştir.

Transformatörlerde farklılık gösteren yapılardan operasyonel olarak ilk sırada yer alan konumsal kodlama metodudur [26]. Deterministik bilgisayar sistemleri doğal dili insanlar gibi metinsel veya fonetik olarak algılayamamaktadır. Bu sebeple mevcut verimizi bilgisayarın

anlayabileceği sayısal değerler şeklinde ifade etmemiz gerekmektedir. Bu ihtiyaçtan kaynaklı modelin eğitimi aşamasında, derlemin içeriğinden oluşturulan sözlüğün sözcük gösterimi oluşturulmalıdır. Sözlüğü oluşturan sözcüklerin her biri, gösterimden belirlenen boyutta, sayıda, kendisinin en yakınına konumlanan diğer sözcüklerle skaler değerler şeklinde ifade edilen temsil vektörleri oluşturulmaktadır. Bu temsil vektörleri, bilgisayar sistemlerinin doğal dili anlayacak şekilde sayısallaştırılmasında en önemli adımlardan biridir. Buraya kadar olan kısım diğer mimarilerde de ortak olarak bulunmaktadır. Transformatör mimarisi ise konumsal metod ile her sözcüğün cümledeki konum bilgisini ve sözcük temsillerinin konumlarını kullanarak, Eş. 1 ve Eş. 2'deki formül sayesinde konumsal değer matrisini oluşturur. Daha sonra mevcut girdi dizi matrisi ile konumsal değer matrisi toplanır. Bu sayede bağlama bağlı olarak gramer yapısının çözülmesi sağlanmış olur. Aynı zamanda eş sesli sözcüklerin cümledeki konumuna göre hangi anlamı ifade ettiği bilgisayar tarafından anlaşılabilir olmaktadır. Ayrıca Şekil 6'da, örnek olarak 32 sözcükten oluşan ve sözcük temsil boyutu 64 olan bir girdi dizi matrisi için cümledeki konumsal değer matrisi gösterilmiştir.

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d_{model}}) \quad (1)$$

$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d_{model}}) \quad (2)$$

Transformatör mimarisinin, önceki mimarilerden farklılaşmasının operasyonel olarak bir sonraki faktörü ise çok başlı öz dikkat mekanizmasıdır. Bu mekanizma kısmen LSTM ve GRU gibi mimarilerde ek katman olarak modele eklenen dikkat mekanizmasının bir türevidir. Fakat Transformatör mimarisi, kodlayıcı ve kod çözücü bloklarının alt katmanlarında çok başlı öz dikkat mekanizması kavramı ile daha kapsamlı hesaplama yeteneğine sahip olmaktadır. Bu adımda cümle içindeki her sözcüğün diğer sözcüklerle olan dikkat matrisi Şekil 7'deki gibi Eş. 3'teki hesaplamalar yapılarak ağırlıklar güncellenir. Bu hesaplamalarda sorgu (Q), anahtar (K) ve değer (V) matrisleri parametre olarak kullanılır. İlk başta bu üç parametre Q, K ve V aynı dizi matrisleriyken eğitim aşamasında farklılaşarak

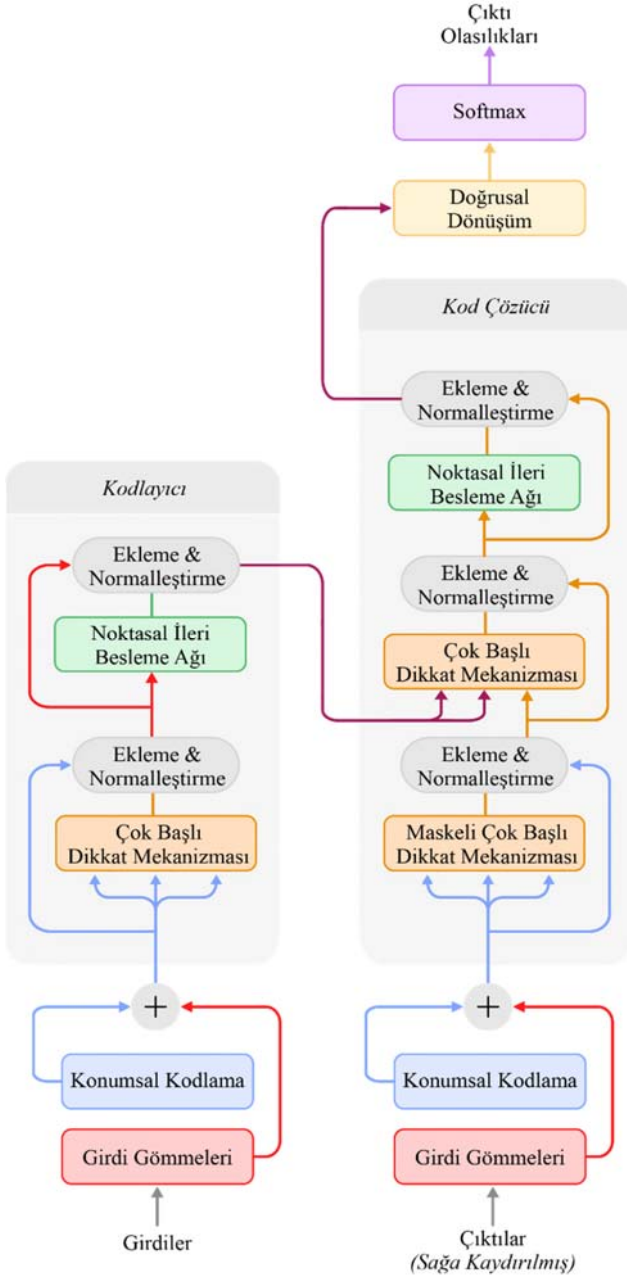
Tablo 4. Ön işlenmiş haber ve başlık örneği-1 (Pre-processed news and headline sample-1)

	Orijinal metin	Ön işlenmiş metin
Haber	Özel yetkili Ankara 12. Ağır Ceza Mahkemesi, 28 Şubat soruşturmasını yürüten savcıların itirazı üzerine, nöbetçi hâkimlikçe serbest bırakılan emekli Orgeneral Teoman Koman ile emekli Korgeneraller Engin Alan ve Kamuran Orhon hakkında "yakalama kararı" çıkarttı. Mahkeme kararın gerekçesini "zanlılara atılı suçun niteliği" olarak belirtti. MHP'den milletvekili seçilen Alan, Balyoz davasından tutuklu bulunuyor.	özel yetkili Ankara ağır ceza mahkemesi şubat soruşturmasını yürüten savcıların itirazı üzerine nöbetçi hâkimlikçe serbest bırakılan emekli Orgeneral Teoman koman emekli korgeneraller engin alan Kamuran Orhon hakkında yakalama kararı çıkarttı mahkeme kararın gerekçesini zanlılara atılı suçun niteliği belirtti MHP den milletvekili seçilen alan balyoz davasından tutuklu bulunuyor
Başlık	28 Şubat komutanlarına yakalama kararı çıktı	### şubat komutanlarına yakalama kararı çıktı

Tablo 5. Ön işlenmiş haber ve başlık örneği-2 (Pre-processed news and headline sample-2)

	Orijinal metin	Ön işlenmiş metin
Haber	Türk Hava Yolları'nın 2016'nın ilk iki ayındaki yolcu sayısı geçen yılın aynı dönemine göre yüzde 11.2 artışla 9.2 milyona ulaştı. Bu rakam 2015'te 8.3 milyon olmuştu. Yolcu sayısındaki artış iç hatlarda yüzde 12.6, dış hatlarda yüzde 10.2 oranında gerçekleşti. Dış Hatlar Business/Comfort Class yolcu sayısı ve Dıştan Dışa Transfer Yolcu sayılarında da Ocak-Şubat 2015 dönemine kıyasla sırasıyla yüzde 5.8 ve 20.9 artış sağlandı. Doluluk oranı 3 puanlık düşüş ile yüzde 73.3 oldu.	Türk hava yolları ilk ayındaki yolcu sayısı geçen yılın aynı dönemine yüzde artışla milyona ulaştı rakam milyon olmuştu yolcu sayısındaki artış hatlarda yüzde dış hatlarda yüzde oranında gerçekleşti dış hatlar business/comfort class yolcu sayısı dıştan dışa transfer yolcu sayılarında ocak şubat dönemine kıyasla sırasıyla yüzde artış sağlandı doluluk oranı puanlık düşüş yüzde oldu
Başlık	THY yolcu sayısını 9.2 milyona çıkardı	THY yolcu sayısını ### milyona çıkardı

güncellenir. Eş. 3'teki parantez içindeki hesaplamalar ile modele girdi olarak giren dizinin barındırdığı sözcüklerin, kendi aralarındaki ağırlık değerleri kare matris şeklinde hesaplanır. Daha sonra V matrisi ile çarpılarak, modelin kodlayıcı ve kod çözücü bloklarının boyutlarına uygun bir şekilde ve dikkat değerlerini de barındıran bir matris elde edilmiş olur. Dikkat mekanizmasındaki çok başlı dikkat kavramı ise Q, K ve V matrislerinin belirlenen baş sayısı kadar olacak şekilde boyutları küçültülerek dikkat hesapları yapılması sayesinde, cümlelerdeki bağlamın farklı açılardan da yakalanması sağlanmaktadır [26].



Şekil 5. Transformator mimarisi (Architecture of Transformer)

$$Dikkat(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3)$$

Transformator mimarisindeki diğer bir farklılık noktasal ileri besleme ağ katmanı yapısıdır. Buradaki noktasal kavramı buraya kadar dizi

matrisleriyle yapılan işlemler, bu aşamada matrisi oluşturan sözcük vektörleri üzerinde yapılarak ileri yayılımın gerçekleştirilmesidir. Bu katmanda işlemler birbirine bağlı, farklı aktivasyon fonksiyonlarına sahip iki ayrı yoğun katmanda (dense layer) yapılır. Bu sayede dikkat katmanlarında eğitilen ağırlıkların potansiyel olarak daha zengin temsil gücüne sahip olması sağlanır. Aynı zamanda modeldeki bir sonraki bloğun girişiyle aynı boyutta matrislerin ileriye beslenilmesi sağlanır [26].

Transformator mimarisinin bir başka farkı ise artık toplam (residual) yapısına sahip olmasıdır. Bu yapı Transformatorün kodlayıcı bloklarında sahip olduğu 2 alt katmanın ve kod çözücü bloklarında sahip olduğu 3 alt katmanın hepsinde katmanlara giren dizi matrisi ile katmanlara girip, katmanlarda yapılan işlemlerden sonra elde edilen matrisler Şekil 5'te gösterildiği gibi toplanmaktadır. Bu artık bağlantılar gradyanların* doğrudan ağırlar üzerinde akmasına izin vererek ağı eğitime yardımcı olur [26]. * Gradyan: Tensör işleminin türevidir. Türev kavramının fonksiyonlardan çok boyutlu girdilere genelleştirilmesidir.

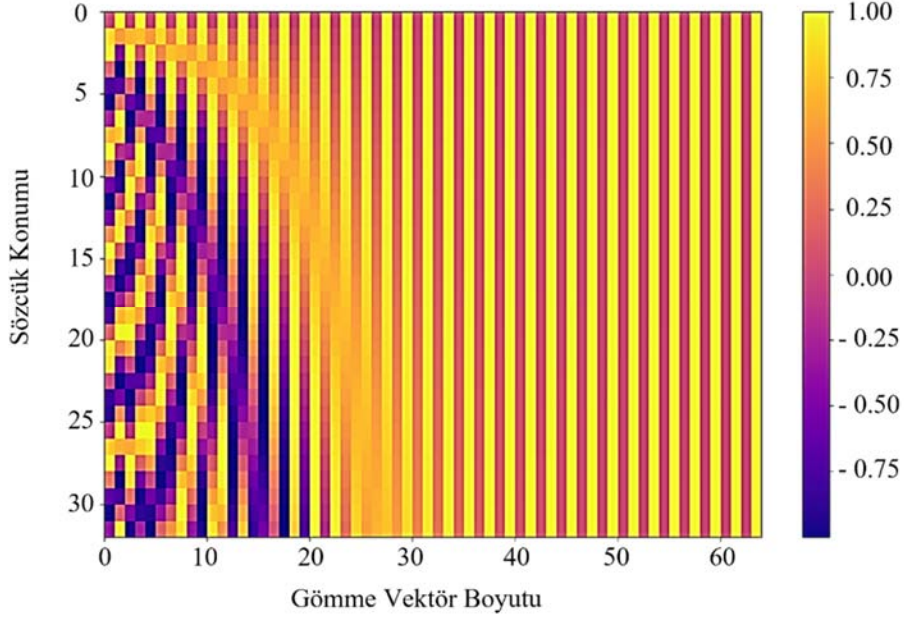
Transformator mimarisi, diğer mimarilerden operasyonel olarak birçok farklı yapı barındırmasının yanı sıra Tablo 6'da diğer modellerle birlikte işlem karmaşıklıklarının da belirtildiği üzere performans açısından da farklılaşmaktadır. Tablo 6'da n parametresi dizi boyutunu, uzunluğunu, d parametresi ise modelin hücre derinliğini yani gizli durum birim sayısını ifade etmektedir. Transformator t zaman adımında bütün bir diziyi tek seferde işleme aldığı, dizideki her sözcüğün öncesinde ve sonrasında da sözcükler olduğu için n^2 ile ifade edilmektedir. Diğer mimariler ise ayrı zaman adımlarında işlediği için n şeklinde maliyette yer alır. Diğer parametre olan d ise Transformator mimarisinde tek seferde hesaplanırken, diğer mimarilerde zaman adımından dolayı, her zaman adımında hesaplanır. Genel olarak n parametresi değer olarak 100 civarında sayısal değer olarak tercih edilir. Fakat d parametresi ise gelişmiş projelerde 2048, 4096 gibi büyük sayısal değerler alacak şekilde kullanılır [27].

Tablo 6. Transformator mimarisi hesaplama maliyeti (Transformer architecture calculation cost)

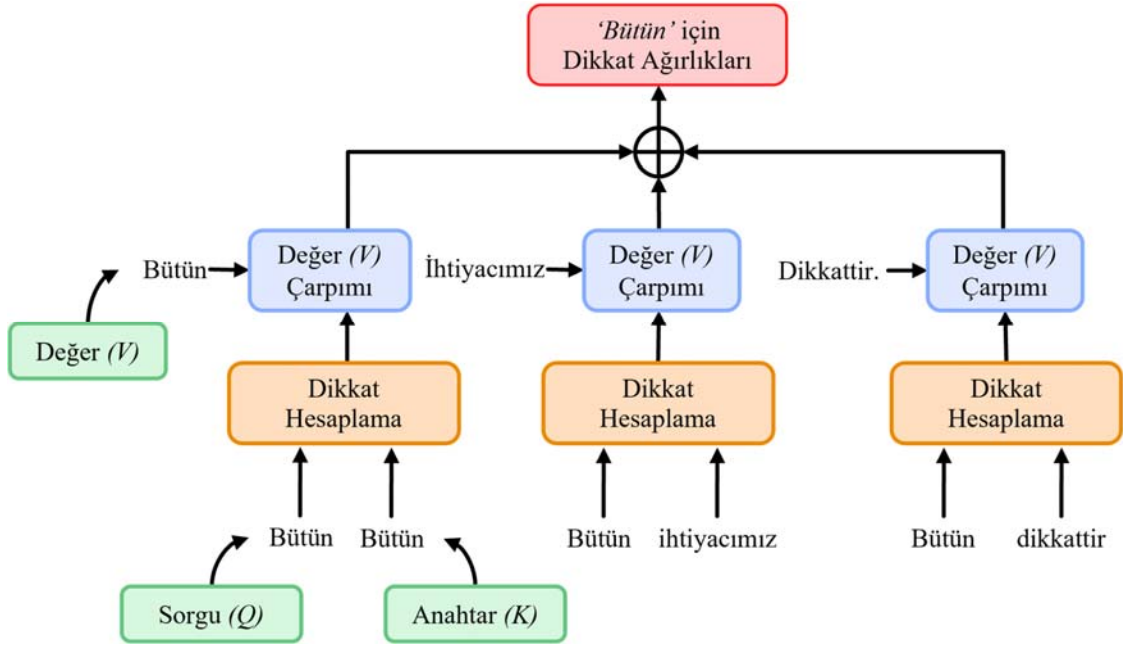
	İşlemler
Transformator	$n^2 \times d$
Yinelemeli Ağlar (RNN, LSTM, GRU)	$n \times d^2$
Konvolüsyonel Ağlar	$n \times d^2$

2.4. Eğitim Modeli (Training Model)

Bu çalışmada derlemin hazırlanmasından eğitime ve üretim aşaması da dahil olmak üzere TensorFlow ve Keras birlikte kullanılmıştır. Eğitim esnasında gerekli olan sistem gereksinimlerinin bulut hesaplama (cloud computing) üzerinden maliyeti göz önünde bulundurulduğunda, bunu en iyi şekilde karşılayabilecek ortamın Google Colab Pro platformu olduğu saptanmıştır [28]. TensorFlow, Google ürünü olması sebebiyle Google Colab tarafından sağlanan donanımsal özelliklerden en iyi şekilde faydalanabilmektedir, bu nedenle çalışmada tercih edilmiştir. Diğer taraftan Keras için TensorFlow'un arka uç (backend) çalıştırmasını sağlayan bir API'ye sahip olması da Keras'ın seçiminde etkili olmuştur [29]. Bu kütüphaneler kullanılarak haber metinlerindeki ve başlıklardaki cümleleri oluşturan her bir sözcük herhangi bir işleme tabi tutulmadan, yani morfolojik yapı üzerinde ayırıştırma veya sözcük kökleri bulunmadan, haber ve başlık sözlükleri ayrı ayrı oluşturularak sayısallaştırıldı. Toplamda eğitilmek üzere 450.000'in üzerinde jeton elde edildi. Haber metinleri ve başlıkları elde edilen sayısal sözlük kullanılarak sırasıyla 384 ve 12 uzunlukta dolgu (padding) kullanılarak sabit vektörler olarak ifade edilecek şekilde eğitim verisi,



Şekil 6. Konumsal değer matrisi (Positional value matrix)



Şekil 7. Öz dikkat mekanizması (Self-attention mechanism)

eğitim model mimarisine girdi olarak beslemek için uygun hale getirildi. Bu sayısal parametreler deneysel, gözlemsel ve Tablo 3'teki istatistiksel sayılar kullanılarak normal dağılım grafiği üzerinde ortalama ve standart sapma değerleri ile hesaplanan 68-95-99,7 kuralı göz önünde bulundurularak belirlenmiştir [30]. Aynı sözlükler olarak elde edilen girdi ve hedef sözlükleri yine Keras kütüphanesi kullanılarak, varsayılan sözcük gömme yöntemi kullanılarak her bir sözcüğü ifade eden 512 boyuta sahip vektörler şeklinde hazırlanmıştır. Oluşturulan vektörler derlemin oluşturduğu sözcük gösterimindeki vektörel uzayda her sözcüğün kendine en yakın 512 sözcüğü ifade edecek şekilde ve cümle olarak düşünüldüğünde ise sözcük vektörlerinden oluşan girdi dizisini bir matris şeklinde modele

beslenmektedir. Bu çalışmada iki farklı eğitim kaynağı kullanılmıştır. Bütün modeller 125.000 haber üzerinde 25 dönem boyunca eğitilmiştir. İlk eğitim kaynağı olarak; LSTM ve GRU modelleri, hesaplamaların eş zamanlı olarak akmasını sağlayan 3584 çekirdeğe ve 16 GB grafik belleğine sahip NVIDIA Tesla P100 grafik işlemcili Google Colab Pro bulut servisi üzerinde çalıştırılmıştır [31]. Eğitim süreleri GRU için 8 saat sürmüştür. LSTM için ise 6 saatte tamamlanmıştır. İkinci eğitim kaynağı olarak; Transformatör mimarisi modeli, 4608 Cuda çekirdeğine sahip 48 GB Quadro RTX 8000 grafik işlemcili ve 256 GB RAM belleğe sahip uzak bilgisayar üzerinde çalıştırılmıştır. 25 dönem boyunca eğitim yaklaşık olarak 7 saatte tamamlanmıştır.

3. Sonuçlar ve Tartışmalar (Results and Discussions)

Derlemin LSTM, GRU ve Transformatör mimarisi tabanlı kodlayıcı-kod çözücü modelle 25 dönem boyunca eğitilmesi sonucu, kayıp fonksiyonu değerleri Şekil 8’de gösterilmiştir. Kayıp fonksiyonu, derin öğrenme mimarilerinde eğitim esnasında hedef ile model tarafından tahmin edilen çıktı arasındaki hatanın hesaplanarak en aza indirilmesini amaçlayan araçtır [25]. Yani hedef ile tahmin arasındaki gradyan eğiminin hesaplanmasıdır. Bu çalışmada eğitim esnasındaki kayıp fonksiyon değerleri LSTM, GRU ve Transformatör mimarileri için sırasıyla 1,03, 0,55 ve 2,49 olarak ölçülmüştür.

Eğitilen model ile derlemedeki doğrulama için ayrılan haberler kullanılarak, bu haberlere başlık üretilmeye çalışılmıştır. Eğitime girmeyen doğrulama verileri ile sistem tarafından üretilen başlıklardan iki tanesi, orijinal haber ve başlık metinleriyle birlikte eğitilen modeller kullanılarak oluşturulan yeni başlıkların gösterildiği Tablo 7 ve Tablo 8’de örnek olarak verilmiştir. Çalışmanın

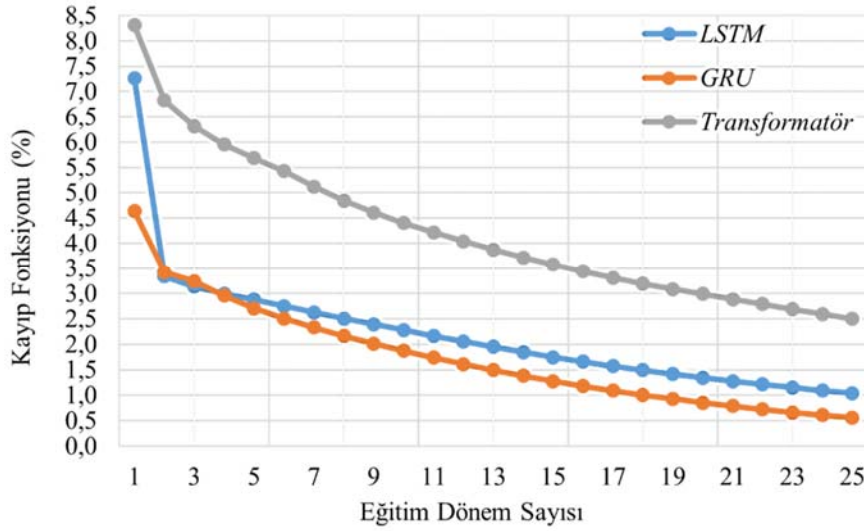
değerlendirmesinde özellikle metin özetleme problemlerinde kullanılan ROUGE ölçüm metriği kullanılmıştır. Eş. 4’teki Rouge Recall değeri BLEU metriği olarak, Eş. 5’teki Rouge Precision değeri ROUGE metriği olarak değerlendirilmektedir. Eş. 6’daki Rouge F1 Score değeri ise ROUGE ve BLEU değerlerinin harmonik ortalaması ile elde edilmektedir [32]. Bu metrikler hesaplanırken TorchMetric kütüphanesindeki ilgili modül kullanılmıştır [33].

$$\text{Rouge Recall} = \frac{\text{Çakışan Sözcük Sayısı}}{\text{Referans Başlıktaki Toplam Sözcük Sayısı}} \quad (4)$$

$$\text{Rouge Precision} = \frac{\text{Çakışan Sözcük Sayısı}}{\text{Sistemin Ürettiği Başlıktaki Toplam Sözcük Sayısı}} \quad (5)$$

$$\text{Rouge F1 Score} = 2 * \frac{\text{Rouge Recall} \times \text{Rouge Precision}}{\text{Rouge Recall} + \text{Rouge Precision}} \quad (6)$$

Derlemden doğrulama için ayrılan veriler kullanılarak, genel olarak özetleme görevlerinin başarımının ölçülmesi için kullanılan ROUGE-



Şekil 8. Modelin eğitim doğruluk oranı (Training accuracy rate of the model)

Tablo 7. Doğrulama verisi ile üretilen başlık-1 (Headline produced with validation data-1)

Haber	Erzurum’da iki siyasi parti taraftarları arasında çıkan olayda 3 kişi ile olaya müdahale eden 1 polis yaralandı. Dün akşam saatlerinde Yakutiye ilçesi Mahallebaşı semtinde ticari taksitle tur atarak bir partinin sembol işaretini yaptığı ileri sürülen taksici, karşı parti taraflarının saldırısına uğradı. Bunun üzerine kaçmaya çalışan taksici ve yanındakiler ile karşı taraf arasında taşlı sopalı kavga yaşandı. Olay yerine gelen polis, yaklaşık 50 kişinin karıştığı olaya müdahale etti. Soruşturma başlatılan olayda yaralanan 3 kişi ile polis, hastaneye kaldırıldı.
Başlık	Erzurum’da Tehlikeli Gerginlik
Üretilen başlıklar	
LSTM	pkk lılar mezarlığa bile patlatıcı döşemiş
GRU	Erzurum da ak partili başkan in acı haberi
Transformatör	ak parti seçim aracına saldırı ### yaralı

Tablo 8. Doğrulama verisi ile üretilen başlık-2 (Headline produced with validation data-2)

Haber	Adana’da cinsel ilişki isteğini reddeden eski çalışma arkadaşı 27 yaşındaki Ayça Tekin’i öldürdüğü iddiasıyla yargılan 20 yaşındaki Abdullah Koyuncu, ömür boyu hapis cezasına çarptırılırken, herhangi bir ceza indirimi uygulanmadı.
Başlık	Ayça’nın Katiline İndirimsiz Müebbet
Üretilen başlıklar	
LSTM	hamile kadına şiddete karşı sağır ya da vardır
GRU	hastanede taciz iddiasına soruşturma
Transformatör	cinsel saldırıya ### yıl hapis cezası

1, ROUGE-2 ve ROUGE-L metrikleri ile elde edilmiştir. Her birinin F-1 Score, Precision ve Recall değerlerini içermektedir. Genel deney sonuçları belirtilen metrikler üzerinden, 25 dönem eğitilen LSTM, GRU ve Transformatör modelleri için ayrı ayrı Tablo 9'daki gibi ölçülmüştür. Sonuç olarak ise ölçümler sonucunda elde edilen metrik değerler üzerinden hareketle Transformatör mimarisinin kısmen iyi olduğu gözlemlenmiştir.

Tablo 9. Derlemin genel deney sonuçları
(General experimental results of the corpus)

		LSTM	GRU	Transformatör
Rouge-1	F-1 Score	0,2917	0,3217	0,3358
	Precision	0,4667	0,3143	0,4792
	Recall	0,2456	0,3333	0,3016
Rouge-2	F-1 Score	0,1000	0,1818	0,2667
	Precision	0,1000	0,1667	0,2857
	Recall	0,1000	0,2000	0,2500
Rouge-L	F-1 Score	0,2500	0,3217	0,3358
	Precision	0,3667	0,3143	0,4792
	Recall	0,2193	0,3333	0,3016

Türkçe dili için haber başlığı üretme üzerine, derin öğrenme tabanlı ve yoruma dayalı çalışma yapmış olan Karakoç ve Yılmaz'ın [13] doğrulama verileri üzerinde yapmış oldukları deneyde ROUGE-1 değeri üzerinden 0,20'lik bir başarı elde ettikleri görülmüştür. Bu metrik değer için değerlendirme yapıldığında Transformatör mimarisinin 0,4792 bir değerle belirtilen çalışmadan daha başarılı olduğu görülmektedir.

4. Sonuçlar (Conclusions)

Bu çalışma Transformatör mimarisini temel alan bir model ile Türkçe haber metinlerine başlık üretme görevini yerine getirmek üzerinedir. Model oluşturulurken soyut özetleme yaklaşımı kullanılarak bağlamı doğru ifade eden özgün başlıklar üretilmesi hedeflenmiştir. Transformatör mimarisi ile oluşturulan modelin başarımını göstermek için LSTM ve GRU mimarileri ile de modeller oluşturulmuş ve eğitimleri gerçekleştirilmiştir. Modellerin test aşamasında ise Greedy algoritması kullanılmıştır. Sonuçlar ROUGE-1, ROUGE-2 ve ROUGE-L metrikleri üzerinden değerlendirilmiştir. Elde edilen sonuçlara göre Transformatör mimarisi, LSTM ve GRU mimarilerinden daha başarılı sonuçlar vermiştir. Bunun yanı sıra Transformatör mimarisi ile elde edilen başlıkların diğer mimarilere göre haber içeriğine kısmen daha uygun başlıklar ürettiği gözlemlenmiştir. Gelecek çalışmalarda Transformer-XL [10] veya Google Pegasus [34] gibi modellerle bağlamı daha iyi ifade edebilen, derlem bağımsız modeller geliştirilebilir. Bağlamdan bağımsız görevler için Türkçe'nin morfolojik yapısına uygun jetonlaştırma (tokenizer) ve Çift Yönlü Transformatör Kodlayıcı Temsilleriyle (Bidirectional Encoder Representation from Transformer, BERT) veya Google tarafından (Google BERT) Wikipedia verileriyle eğitilen ön eğitilmiş derlem sözlüğü ve dil modeli kullanılabilir. Böylelikle kaynaktan eğitilen, fakat eğitildiği kaynağa bağlı kalmadan başarılı içerik üretebilen genel dil modelleri oluşturulabilir. Aynı zamanda üretim aşamasında kod çözücü bloğunda varsayılan olarak kullanılan en yüksek oranlı tahmini kullanarak ilerleyen Greedy algoritması yerine, sezgisel ışın arama (Heuristic Beam Search, HBS) algoritması ile n farklı en yüksek oranlı tahminleri de hesaba katan bir üretim ağaç yapısıyla daha başarılı sonuçlar elde edilebilir.

Teşekkür (Acknowledgement)

Bu çalışmada, transformatör mimarisinin eğitimi aşamasında uzak bilgisayar üzerinde modelimizin eğitiminde bize yardımcı olan Dr. Öğretim Üyesi Emir ÖZTÜRK'e teşekkürü borç biliriz.

Kaynaklar (References)

- Edmunson, H.P., New Methods in Automatic Extracting, Journal of the Association for Computing Machinery (J.ACM), 16 (2), 264-285, 1969.
- Luhn, H.P., The Automatic Creation of Literature Abstracts, IBM Journal of research and development (IBM J. Res. Dev), 2 (2), 159-165, 1958.
- Erkan, G., Radev, D.R., LexRank: Graph-based Lexical Centrality as Saliency in Text Summarization, Journal of Artificial Intelligence Research (JAIR), 22, 457-479, 2004.
- Lopyrev, K., Generating News Headlines with Recurrent Neural Networks, Journal of Controlled Release (JCR), 230, 73-78, 2015.
- Takase, S., Suzuki, J., Okazaki, N., Hirao, T., Nagata, M., Neural Headline Generation on Abstract Meaning Representation, In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP), Austin-Texas-A.B.D, 1054-1059, 1-5 November, 2016.
- Tan, J., Wan, X., Xiao, J., From Neural Sentence Summarization to Headline Generation: A Coarse-to-Fine Approach, In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI), Melbourne- Australia, 4109-4115, 19-25 August, 2017.
- Murao, K., Kobayashi, K., Kobayashi, H., Yatsuka, T., Masuyama, T., Higurashi, T., Tabuchi, Y., A Case Study on Neural Headline Generation for Editing Support, In Proceedings of the 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT), Minneapolis-Minnesota-A.B.D, 73-82, 2-7 June, 2019.
- Gavrilov, D., Kalaidin, P., Malykh, V., Self-attentive Model for Headline Generation, In: Azzopardi L., Stein B., Fuhr N., Mayr P., Hauff C., Hiemstra D. (eds) Advances in Information Retrieval. ECIR 2019. Lecture Notes in Computer Science (Lect. Notes Comput. Sci.), 11438, 87-93. 2019.
- Liao, Y., Meng, K., Zhang, J., Liu, G., Unleashing the potential of attention model for news headline generation, In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow-United Kingdom, 1-8, 19-24 July, 2020.
- Dai, Z., Yang, Z., Yang, Y., Cohen, W.W., Carbonell, J., Le, Q.V., Salakhutdinov, R., Transformer-xl: Attentive language models beyond a fixed-length context, Proceedings of the 57th Conference of the Association for Computational Linguistics (ACL 2019), Florence-Italy, 2978-2988, 2019.
- Bukhtiyarov, A., Gusev, I., Advances of Transformer-Based Models for News Headline Generation, Artificial Intelligence and Natural Language 9th Conference (AINL 2020), Helsinki-Finland, 54-61, 7-9 October, 2020.
- Gu, X., Mao, Y., Han, J., Liu, J., Yu, H., Wu, Y., Yu, C., Finnie, D., Zhai, J., Zukoski, N., Generating Representative Headlines for News Stories, In Proceedings of the The Web Conference 2020, Taipei-Taiwan, 1773-1784, 20-24 April, 2020.
- Karakoç, E., Yılmaz, B., Deep Learning Based Abstractive Turkish News Summarization, In Proceedings of the 27th Signal Processing and Communications Applications Conference (SIU), Sivas-Türkiye, 1-4, 24-26 Nisan, 2019.
- Chollet, F., Uygulamalı Derin Öğrenme, Python ile Derin Öğrenme, Editör: Kuyumcu, B., Buzdağı Yayınevi, Ankara, Türkiye, 123-360, 2017.
- Kilimci Z.H., Financial sentiment analysis with Deep Ensemble Models (DEMs) for stock market prediction, Journal of the Faculty of Engineering and Architecture of Gazi University, 35 (2), 635-650, 2019.
- Battini Sönmez E., Yıldız T., Yılmaz B.D., Demir A.E., Image captioning in Turkish language: Database and model, Journal of the Faculty of Engineering and Architecture of Gazi University, 35 (4), 2089-2100, 2020.
- Güzel Turhan, C., Bilge, H.Ş., Scalable image generation and super resolution using generative adversarial networks, Journal of the Faculty of Engineering and Architecture of Gazi University, 35 (2), 953-966, 2019.
- Metlek, S., Kayaalp, K., Detection of bee diseases with a hybrid deep learning method, Journal of the Faculty of Engineering and Architecture of Gazi University, 36 (3), 1715-1732, 2021.
- Nekic, M. Automatic Text Summarization. <https://www.youtube.com/watch?v=d0OXm0dRZ4>. Yayın tarihi Haziran 27, 2019. Erişim Tarihi Mayıs 15, 2020.

20. Şen, M.U., Yanıkoğlu, B., Document classification of SuDer Turkish News Corpora, In Proceedings of the 26th Signal Processing and Communications Applications Conference (SIU), İzmir-Türkiye, 1-4, 2-5 Mayıs, 2018.
21. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., Attention is All You Need, In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS), Long Beach California-A.B.D., 6000-6010, 4-9 December, 2017.
22. Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D., Language Models are Few-Shot Learners, Advances in Neural Information Processing Systems 33 (NeurIPS 2020), Online Conference, 1877-1901, 6-12 December, 2020.
23. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., Language Models are Unsupervised Multitask Learners. https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf. Yayın tarihi Şubat 14, 2019. Erişim tarihi Temmuz 11, 2022.
24. Google AI Blog. Recent Advances in Google Translate. <https://ai.googleblog.com/2020/06/recent-advances-in-google-translate.html>. Yayın tarihi Haziran 8, 2020. Erişim tarihi Temmuz 11, 2022.
25. Goodfellow, I., Bengio, Y., Courville, A., Derin Ağlar: Modern Pratikler, Derin Öğrenme, Editör: Aksoy, B., Kurt, F., Çelik, V., Buzdağı Yayınevi, Ankara, Türkiye, 163-488, 2018.
26. Zhang, A., Lipton, Z.C., Li, M., Smola, A.J. Dive into Deep Learning. <https://d21.ai/d21-en.pdf>. Yayın tarihi Kasım 6, 2020. Erişim tarihi Ocak 10, 2021.
27. Kaiser, L. Attention is all you need; Attentional Neural Network Models | Lukasz Kaiser | Masterclass. <https://www.youtube.com/watch?v=rBCqOTefxvg>. Yayın tarihi Ekim 4, 2017. Erişim tarihi Şubat 10, 2021.
28. Google Colab. Making the Most of your Colab Subscription. <https://colab.research.google.com/?hl=en>. Erişim tarihi Temmuz 11, 2022.
29. Chollet, F., Fundamentals of Deep Learning, Deep Learning with Python, Manning Publications Company, Shelter Island, USA, 1-116, 2018.
30. Wheelan, C., Merkezi Limit Teoremi, Çıplak İstatistik, Editör: Kaya, Ö., Buzdağı Yayınevi, Ankara, Türkiye, 201-223, 2020.
31. NVIDIA. Data Sheet: Tesla P100. <https://images.nvidia.com/content/tesla/pdf/nvidia-tesla-p100-PCIe-datasheet.pdf>. Yayın tarihi Ekim 4, 2016. Erişim tarihi Mart 1, 2021.
32. Özkan, C., İnternet Tabanlı Türkçe Metinler İçin Otomatik Özetleme Tekniği, Yüksek Lisans Tezi, Maltepe Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul, Türkiye, 2019.
33. TorchMetric. ROUGE SCORE. https://torchmetrics.readthedocs.io/en/stable/text/rouge_score. Erişim tarihi Temmuz 11, 2022.
34. Zhang, J., Zhao, Y., Saleh, M., Liu, P.J., PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization, In Proceedings of the 37th International Conference on Machine Learning, Online Conference (ICML'20), 11328-11339, 13-18 July, 2020.

