ISPEC Fen Bilimleri Enstitüsü Dergisi, 4(1):1-11, 2025
ISPEC Journal of Science Institute, 4(1):1-11, 2025
DOI: https://doi.org/10.5281/zenodo.15775633

Yapay Zekâ ve Derin Öğrenme Yöntemleriyle Türkçe Metin Özetleme: Önceden Eğitilmiş Büyük Dil Modellerinin Vektörel Kosinüs Benzerlik Tabanlı İçerik Seçimi ile Türkçe Özelinde Performanslarının İyileştirilmesi

Bilal KOBANOĞLU [1*], Fazlı YILDIRIM [1]

[1] İstanbul Topkapı Üniversitesi, Yönetim Bilişim Sistemleri Bölümü, İstanbul, Türkiye
Sorumlu Yazar (Corresponding author): bilalkobanoglu@gmail.com

Özet

Bu makale, Türkçe metin özetleme sürecinde yapay zekâ ve derin öğrenme tabanlı büyük dil modellerinin kullanılmasına odaklanarak mevcut önceden eğitilmiş dil modelleri ile yenilikçi bir çözüm sunmayı amaçlamaktadır. Bu bağlamda, makalenin temel amacı, önceden eğitilmiş modellerin Türkçe metin özetleme görevine uyarlanarak, eğitim maliyetlerinden kaçınan ve aynı zamanda verimlilik ile performans odaklı bir yaklaşım sunan bir yöntem geliştirmektir. Bu araştırmanın kuramsal dayanağı dilin dağılımsal düzenliliklerinin sayısal temsillerle ifadesini kullanarak, bu düzenliliklerin benzerlik oranına göre üretken dil modellerinin daha düzenli çıktı üretmesi üzerinedir. Bu bağlamda her kelimenin bir vektör olarak ifade edildiği uzayda, cümlelerin oluşturduğu bağlamsal vektörler arasındaki benzerlik (düzenlilik) kullanılarak, büyük dil modellerinin daha düzenli üretken çıktı üretmesi beklenmektedir. Bu diktonomi yaklaşımı, giriş vektörler arasındaki benzerliklerin başka bir süreçte hesaplanarak giriş vektöründeki düzenliliklerin artırımı ile üretken çıktının daha düzenli olmasını hedeflemektedir. Bu çalışmada, Türkçe metin özetleme için büyük dil modellerinde cümle vektörlerinin kosinüs benzerliğine dayalı içerik seçimi stratejisi önerilmektedir. Amaç, en anlamlı cümleleri seçerek modele daha hedefli bir giriş sağlamak ve böylece hem özetleme performansını artırmak hem de hesaplama maliyetlerini düşürmektir. T5 modelinin, en yüksek vektörel benzerliğe sahip cümlelerle beslenerek performansının iyileştirilmesi incelenmiş ve ROUGE metrikleriyle değerlendirilmiştir. Sonuçlar, bu yöntemin özetleme kalitesini artırdığını ve t-testi ile yapılan analizlerin bazı örneklerde anlamlı iyileşmeler gösterdiğini ortaya koymuştur.

Araştırma Makalesi

Makale Tarihçesi
Geliş Tarihi: 09.02.2025
Kabul Tarihi: 28.04.2025

Anahtar Kelimeler
Yapay Zekâ,
Derin Öğrenme,
Önceden Eğitilmiş Dil Modelleri,
ROUGE Skoru

Turkish Text Summarization with Artificial Intelligence and Deep Learning Methods: Improving the Performance of Pretrained Large Language Models for Turkish through Vectorial Cosine Similarity-Based Content Selection

Abstract

This paper aims to present an innovative solution for the Turkish text summarization process by leveraging artificial intelligence and deep learning-based large language models (LLMs), specifically focusing on the use of pre-trained language models. In this context, the primary objective of this study is to adapt pre-trained models for the task of Turkish text summarization, thereby proposing a cost-efficient and performance-oriented approach that eliminates the need for extensive training. The theoretical foundation of this research is based on the numerical representation of the distributional regularities of language, positing that leveraging the similarity ratios of these regularities can enhance the structured output generation of generative language models. In this regard, it is expected that large language models will generate more structured and coherent outputs by utilizing the contextual similarity between sentence vectors in a vector space where each word is represented as a vector. This dichotomic approach aims to improve the consistency of generative outputs by increasing the regularity of input vectors through a separate process that computes the similarities between them. In this study, a cosine similarity-based content selection strategy for sentence vectors in large language models is proposed for Turkish text summarization. The objective is to select the most meaningful sentences, providing the model with more targeted and summarization-friendly input, thereby enhancing summarization performance while reducing computational costs. The study investigates the improvement of the T5 model's performance by feeding it with sentences that exhibit the highest vectorial similarity. The evaluation is conducted using ROUGE-1, ROUGE-2, and ROUGE-L metrics. The results indicate that selecting sentences with high vectorial cosine similarity contributes to a significant improvement in summarization quality. Furthermore, statistical analyses using t-tests demonstrate that this method enhances summarization performance in certain cases.

Research Article

Article History
Received : 09.02.2025
Accepted: 28.04.2025

Keywords
Artificial Intelligence,
Deep Learning,
Pretrained Language Models,
ROUGE Score

1.Giriş

Bu araştırmanın genel amacı, mühendislik prensiplerini ve ayrık düşünme yöntemlerini kullanarak, mevcut araçlarla problemlere daha verimli ve etkili çözümler bulmak ve aynı zamanda Yeşil Mutabakata (Avrupa Komisyonu, 2019) uyum sağlamaktır.

Doğal Dil İşleme (Natural Language Programming, NLP), bilgisayarların insan dillerinde yazılmış ifadeleri veya kelimeleri anlamalarını sağlamaya odaklanan Yapay Zekâ ve Dilbilim alanında bir disiplindir. Bu, kullanıcının işini kolaylaştırmak ve bilgisayarla doğal dilde iletişim kurma isteğini karşılamak amacıyla ortaya çıkmakta ve Doğal Dil Anlama veya Dilbilim ile Doğal Dil Üretimi olmak üzere iki ana kategoriye ayrılmaktadır (Khurana ve ark., 2023).

Türkçe metin özetleme, doğal dil işleme (NLP) alanında önemli bir araştırma konusu olup, metinlerin kısa, öz ve anlamlı bir şekilde özetlenmesi işlemidir. Ancak, Türkçe'nin yapısal farklılıkları ve dil özellikleri, özetleme süreçlerini daha karmaşık hale getirebilmektedir. Bu nedenle, Türkçe metin özetleme görevlerinde daha etkili sonuçlar elde edebilmek için, son yıllarda yapay zekâ ve derin öğrenme yöntemleri üzerine yoğunlaşan araştırmalar artmıştır. Özellikle, önceden eğitilmiş dil modelleri metinlerin daha doğru özetlenmesi için önemli araçlar haline gelmiştir.

Uygulamamızın stratejisi, önceden eğitilmiş T5 büyük dil modelinin performansını iyileştirmek amacıyla, Türkçe metin özetleme sürecinde özetlenecek metin içindeki cümleler arasındaki vektörel kosinüs benzerliği en yüksek olan cümleleri seçerek modele daha hedeflenmiş bir giriş sağlamaktır. Bu yaklaşım, modelin daha anlamlı ve doğru özetler üretmesini amaçlamaktadır. Kosinüs benzerliği, cümlelerin anlamlarını karşılaştıran bir teknik olup, seçilen cümlelerin metnin genel anlamını daha iyi yansıtmasına yardımcı olabilir.

Bu araştırma, Türkçe metin özetleme alanında mevcut büyük dil modelleriyle yapılan çalışmalara katkı sağlamak ve bu modellerin Türkçe'yi daha verimli bir şekilde işleyebilmesini mümkün kılacak yenilikçi bir çözüm önermeyi amaçlamaktadır. Ayrıca, büyük dil modellerinde özetleme performansını artırmak için, modele verilen giriş verilerinin benzerlik oranı yüksek olanlarının seçilmesine dayanan bir strateji sunulmakta ve bu yaklaşım, literatürde daha önce ele alınmamış bir yöntemi ortaya koyarak, Türkçe metin özetleme sürecine dair yeni bir perspektif kazandırmayı hedeflemektedir.

2. Literatürdeki Çalışmalar

Literatürde, büyük dil modellerinin Türkçe verileriyle eğitilmesi ve ince ayarlanması üzerine pek çok çalışma bulunmaktadır. Bu çalışmalar, Türkçe'nin kendine has dil bilgisel ve yapısal özelliklerinin dikkate alınarak büyük dil modellerinin performansının iyileştirilmesine yönelik çeşitli metodolojiler geliştirmiştir. Türkçe metinlere özgü dil özelliklerinin doğru bir şekilde modellenebilmesi için, önceden eğitilmiş büyük dil modellerinin Türkçe veri setleriyle ince ayar yapılarak, dilin zengin yapısının daha etkin bir şekilde öğrenilmesi sağlanmıştır. Bu bağlamda, dilin morfolojik, sentaktik ve semantik özelliklerini doğru bir şekilde yakalamayı hedefleyen metotlar, modelin Türkçe verileriyle uyumlu hale gelmesini sağlamıştır.

Türkçe dilindeki performansı artırmaya yönelik diğer bir yaklaşım ise, modelin Türkçe veri setlerinde daha iyi genelleme yapabilmesi için ince ayar sırasında kullanılan tekniklerde yapılan özelleştirmelerdir. Özellikle, büyük dil modellerinin Türkçe verilerine adapte edilmesi sürecinde yeni model eğitme, transfer öğrenme, veri artırma ve dil özelliklerine duyarlı parametre optimizasyonu gibi stratejiler, modelin doğruluğunu ve verimliliğini önemli ölçüde iyileştirmiştir. Bu çalışmalar, büyük dil modellerinin Türkçe metin işleme alanındaki performansını artırmaya yönelik somut çözümler sunmakta ve daha güçlü, Türkçe'ye özgü metin işleme sistemlerinin geliştirilmesine olanak tanımaktadır.

Kartal ve Kutlu (2020) kendi modellerini eğitirken cümle seçiminde, Bakan ve Yakut (2024) graf tabanlı özetlemede, Karakoç ve Yılmaz (2019) kendi modellerini eğitirken kosinüs benzerlik tabanlı içerik seçimi yapmaktadırlar.

Şakar ve Emekci (2025) farklı veri kümelerinde, vektör veri tabanlarında, RAG yöntemlerinde, büyük dil modellerinde (LLM'ler), gömme modelleri ve gömme filtre puanları üzerinde bir grid search optimizasyonu üzerine çalışmış ve RAG performansı, her bir veri kümesi ve soru-cevap çiftinden elde edilen gömülü LLM yanıtı ile referans yanıtı arasındaki fark kosinüs benzerliğini ölçerek değerlendirmiştir.

Gopalakrishnan ve ark. (2025) GPT-4 modelinin tıbbi yönergelerden neden-sonuç ilişkilerinin çıkarılması üzerine yaptıkları çalışmada modelin doğruluğunu değerlendirmek için kosinüs benzerliği gibi alternatif yöntem kullanmıştır.

Literatürde yer alan büyük dil modellerinin Türkçe verileriyle eğitilmesi ve ince ayarlanması üzerine yapılan çalışmalar genellikle yüksek işlem gücü gereksinimi ve bu süreçlerin çevresel etkileri açısından bazı sınırlamalara sahiptir. Bu modellerin eğitilmesi ve optimize edilmesi, büyük miktarda verinin işlenmesi ve model parametrelerinin ayarlanması sürecinde önemli bilgi işlem gücü ve hesaplama kaynakları talep etmektedir. Bu durum, özellikle derin öğrenme tabanlı modellerin büyük veri setlerinde eğitim süreçlerinde hesaplama sürelerinin uzamasına ve bu süreçlerde büyük miktarda enerji tüketilmesine yol açmaktadır.

Ayrıca, bu yüksek işlem gücü gereksinimi, çevresel etkiler ve karbon salınımı açısından da önemli bir problem teşkil etmektedir. Büyük dil modellerinin eğitilmesi, büyük veri merkezlerinde yoğun hesaplama kaynaklarının kullanılmasına ve bu süreçlerin sonucunda ciddi bir enerji tüketimi ve karbon salımı meydana gelmesine neden olmaktadır.

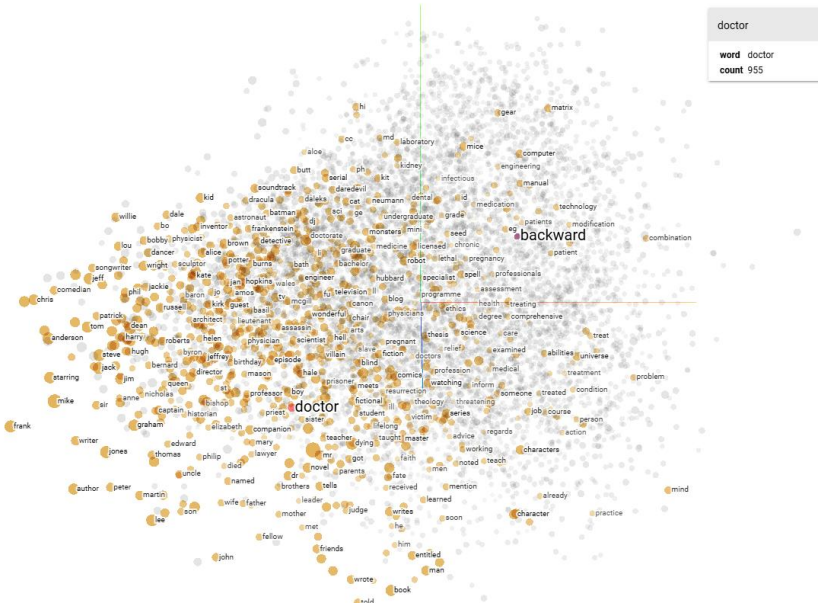
3. Önerilen Yöntem

Harris'e (1954) göre, dilin yapısının dağılımsal düzenlilikler temelinde anlaşılabileceğini savunur. Dil öğeleri rastgele dağılmamaktadır; her bir unsur, belirli bir sırayla ve sınıflardan seçilen üyelerle birleştirilir. Dilin unsurlarının sınıfları, belirli kısıtlamalara tabi olup, bu kısıtlamalar keyfi olarak ihlal edilemez. Dilin her bir unsuru, diğer unsurlarla olan ilişkileri belirli ölçütlere göre tanımlanabilir ve bu sayede dilin yapısına dair daha derin bir anlayış elde edilebilir. Ayrıca, her bir unsurun dağılımı, ilişkili ifadeler arasındaki kısıtlamalarla tanımlanabilir ve bu da dildeki unsurların yapısal düzeninin matematiksel bir şekilde ifade edilmesine olanak tanır.

Harris (1954) bu ilişkilerin dağılımsal düzenlilikler açısından incelenebileceğini ifade ederek dağıtık kelime temsilinin temellerini atmıştır. Kelime temsilleri, kelime gömme vektörleriyle kelimeler vektörel ifadelerle tanımlanıp, aralarındaki ilişkiler tensörlerle ifade edilebilmektedir.

McDonald ve Ramscar' a (2001) göre dağılım teorisi, benzer veya ilişkili anlamlara sahip kelimelerin aynı bağlamda meydana geldiğini öne sürmektedir. Sonuç olarak, sözdizimsel veya anlamsal olarak bağlantılı kelimler için gömme temsilleri, ilişkili olmayan terimlere kıyasla vektör alanında daha yakın yer almaktadır. Bu ilişki, bu gömme temsillerinin üretildiği metin verilerine veya külliyata tamamen bağımlı olmaktadır. İlişkili olan vektör temsillerine örnek Figure 1 de verilmiştir. Word embedding, kelimelerin matematiksel temsillerini oluştururken, kelimeler arasındaki bağlamsal ilişkileri dikkate almaktadır. Örneğin, "doktor" kelimesi, sağlık, tedavi, hastalık gibi terimlerle güçlü bağlara sahiptir. Şekil 1 de de görüleceği üzere word embedding algoritmaları, bu kelimelerin semantik olarak birbirine yakın olduğunu ve benzer bağlamlarda sıklıkla kullanıldıklarını gösteren vektörler üretir. Bu sayede, "doktor" kelimesi ve onunla ilişkilendirilebilecek kelimeler (örneğin, "hemşire", "hastane", "muayene") benzer vektörlere sahip olur, çünkü bu kelimeler genellikle aynı temada ve anlamda kullanılır. Bu tür

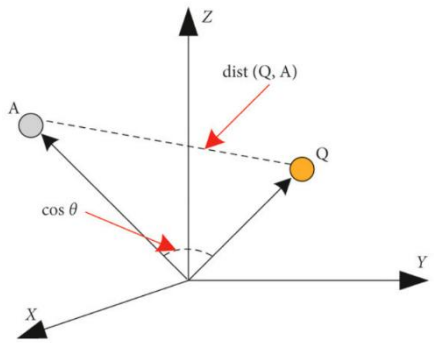
bir temsil, makinelerin dilin derinliklerine inmelerine ve dilin anlamını daha iyi kavrayarak görevleri daha etkili bir şekilde yerine getirmelerine olanak tanımaktadır.



Şekil 1. Kelime Vektörleri Projeksiyonu (Anonim, 2025a).

Johnson ve ark. (2024)'a göre Makinelerin metni işleyebilmesi için, kelime dağarcığındaki serbest biçimli metin terimleri sayısal değerlere (vektörlere) dönüştürülmelidir. Farklı makine öğrenimi teknikleri, metin verilerini otomatikleştirilmiş süreçler, tahmine dayalı modelleme, bilgi keşfi ve anlama amacıyla kullanmak için çeşitli işletmeler tarafından yaygın olarak kullanılmaktadır. Önemli sorunlardan biri, serbest biçimli metni makine öğrenimi algoritmalarının verimli bir şekilde kullanabileceği bir temsile dönüştürmektir

Kosinüs benzerliği, iki vektörün nokta çarpımının her bir vektörün büyüklüğü ile bölünerek hesaplanır. Genellikle, iki öğe arasındaki kosinüs benzerliği değeri 1 olduğunda, bu öğeler yüksek derecede benzer kabul edilir. Kosinüs benzerliği, iki öğe arasındaki karşılaştırılabilirliği nicelendirmenin bir yoludur. İki öğe olan A ve Q' yi ele alınacak olursa, bu öğeler arasındaki kosinüs benzerliği, vektörlerin iç çarpımının, her iki vektörün normlarının çarpımına bölünmesiyle elde edilir. Kosinüs benzerliği genellikle yönelimdeki benzerliği kontrol eder, büyüklükteki benzerliği değil (Mana ve Sasipraba, 2021). Vektörler ve arasındaki kosinüs ilişkisi Şekil 2 de gösterilmiştir.



Şekil 2. Kosinüs benzerliği ve Öklid mesafesinin şematik diyagramları (Anonim, 2025b).

Kosinüs benzerliği, bilgi getirme ve ilgili çalışmalarda yaygın olarak kullanılan bir metriktir. Bu metrik, bir metin belgesini terim vektörleri olarak modellemektedir. Bu model sayesinde, iki belge arasındaki benzerlik, belgelerin terim vektörleri arasındaki kosinüs değerinin hesaplanmasıyla elde edilebilir (Rahutomo, Kitasuka ve Aritsugi, 2012). Bu metriğin uygulanması, herhangi iki metin (cümle, paragraf veya tüm belge) için geçerli olabilir. Arama motoru örneğinde, kullanıcı sorgusu ile belgeler arasındaki benzerlik değeri, en yüksekten en düşüğe doğru sıralanır. Belgenin terim vektörü ile sorgunun terim vektörü arasındaki daha yüksek benzerlik skoru, belgenin sorgu ile daha alakalı olduğunu gösterir (Salton ve Buckley, 1988).

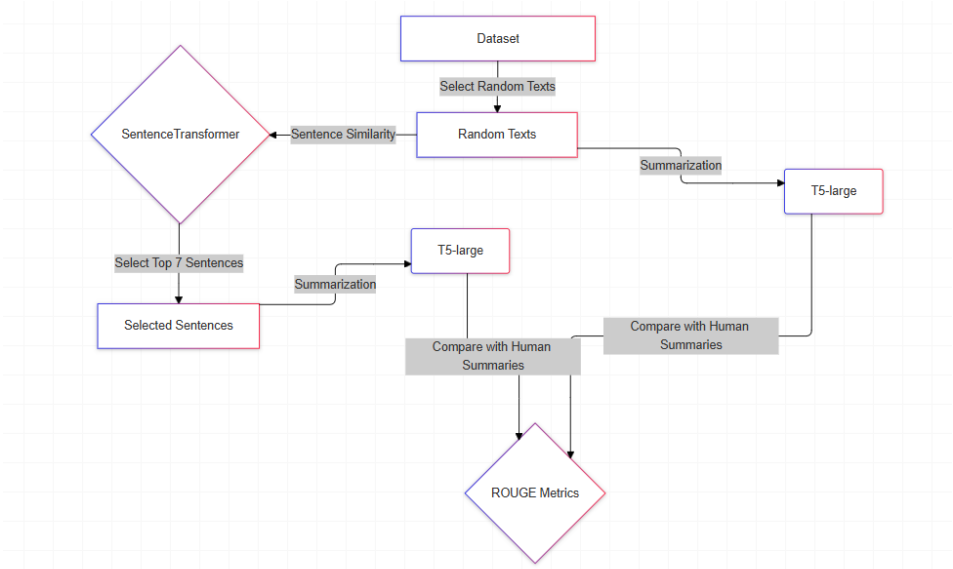
Ancak, kosinüs benzerliği, metnin anlamını tam olarak ele alamaz. Kosinüs benzerliğinin uygulanması, iki terim vektörü arasındaki sözdizimsel eşleşmeye dayanır ve bu bazen güvenilir olmayan sonuçlar verebilir. Sözdizimsel eşleşme, anlam farklarını yeterince iyi yakalayamayabilir. Bu durum, örneğin bilgi getirme sistemlerinde yanlış sonuçlar üretmesine ve performans düşüşüne yol açabilir (Salton ve Buckley, 1988).

Benzerlik, sınıflandırma ve kümeleme görevlerinde temel bir kavramdır. Benzerlik kavramı, mesafe kavramı ile yakından ilişkilidir, ancak tam olarak aynı şey değillerdir. Örneğin, benzerlik, iki örnek arasındaki ortak özellikleri ölçerken, mesafe, bunlar arasındaki farkları gösterir. Yine de, benzerlik ve mesafe arasında güçlü bir bağ vardır. İlk olarak iki örnek arasındaki mesafe hesaplanabilir ve ardından benzer olup olmadıklarını belirlemek için uygun bir eşik belirlenebilir. Mesafe azaldıkça, iki örnek birbirine daha benzer hale gelir (Xia ve ark., 2015).

Öklid mesafesi, basitliği nedeniyle en sık kullanılan metriklerden biridir. Öklid uzayında, iki nokta arasındaki mesafe, onları birleştiren doğru parçasının uzunluğu ile ölçülür. Ne yazık ki, Öklid mesafesi büyüklüklere karşı yüksek bir hassasiyet sergiler. Alternatif olarak, kosinüs benzerliği, iki vektör arasındaki açı ile benzerliği ölçen başka bir yaygın kullanılan metriktir. Herhangi iki desen için, bu desenler arasındaki Öklid mesafesi arttıkça daha az benzer kabul edilirken, kosinüs benzerlikleri arttıkça daha benzer kabul edilir (Xia ve ark., 2015).

Bu araştırmanın kuramsal dayanağı dilin dağılımsal düzenliliklerinin sayısal temsillerle ifadesini kullanarak, bu düzenliliklerin benzerlik oranına göre üretken dil modellerinin daha düzenli çıktı üretmesi üzerinedir. Bu bağlamda her kelimenin bir vektör olarak ifade edildiği uzayda, cümlelerin oluşturduğu bağlamsal vektörler arasındaki benzerlik (düzenlilik) kullanılarak, büyük dil modellerinin daha düzenli üretken çıktı üretmesi beklenmektedir. Bu diktonomi yaklaşımı, giriş vektörler arasındaki benzerliklerin başka bir süreçte hesaplanarak giriş vektöründeki düzenliliklerin artırımı ile üretken çıktının daha düzenli olmasını hedeflemektedir.

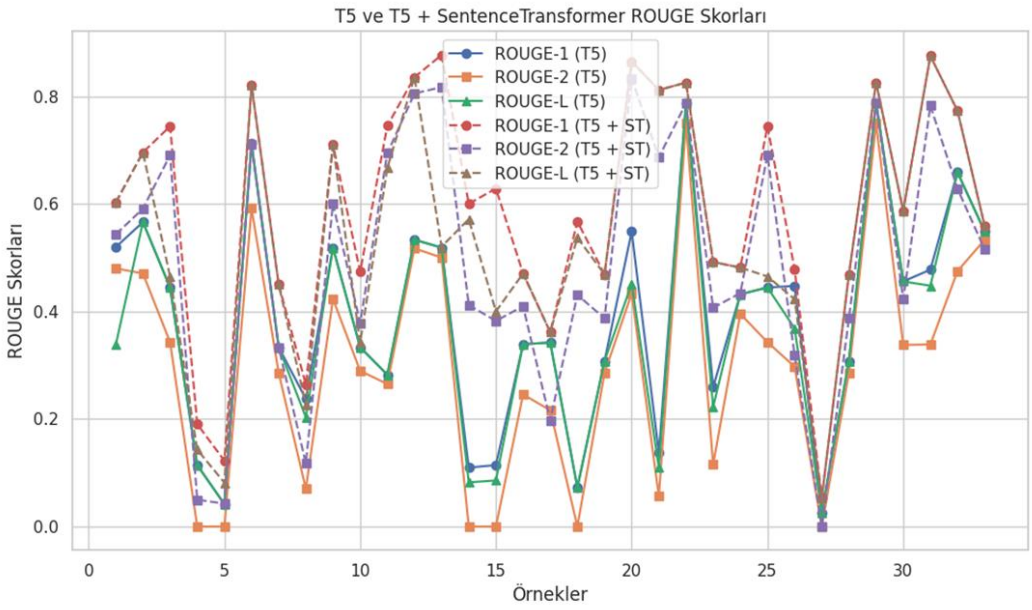
Bu yöntem Google T5-large (Raffel ve ark., 2020) modelinde uygulanmaktadır. Wikipedia Türkçe veri setinden (Güngör, 2023) rastgele seçilmiş 262 metnin özeti önce T5-large modeline gönderilerek metinlerin özeti elde edilmiştir. Aynı 262 metin içindeki cümlelerin kosinüs benzerlik hesaplamaları için SenteceTransformer modeli olan “paraphrase-MiniLM-L6-v2” e (Reimers, 2019) gönderilerek kosinüs benzerliği en yüksek olan yedi cümle seçilerek T5-large modeline özetlenmesi için gönderilmiştir. Seçilen cümle sayısı kaynak verimliği göz önünde bulundurularak belirlenmiştir. Elde edilen özet metinler beşeri özetle Rouge metrikleriyle karşılaştırılarak sonuca ulaşılmıştır. Bu akış Şekil 3 de ayrıntılı olarak gösterilmiştir.



Şekil 3. Önerilen Yöntem İş Akış Şeması

4. Bulgular

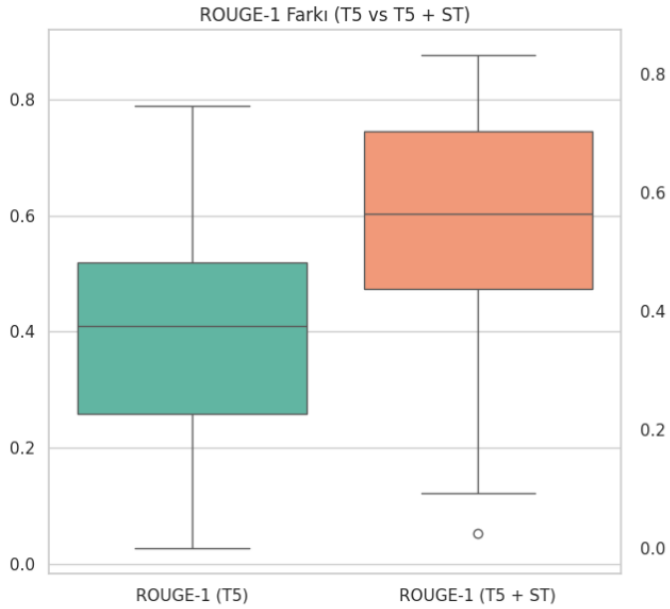
ROUGE-1, ROUGE-2 ve ROUGE-L, metin özetleme ve metin karşılaştırma çalışmalarında kullanılan yaygın değerlendirme metrikleridir. Bu metrikler, özetlerin doğruluğunu, kapsamını ve tutarlılığını ölçmek için kullanılır. ROUGE-1, unigram (tek kelime) benzerliğini ölçen bir metriktir. Bu metrik, referans özet ile model tarafından üretilen özet arasındaki ortak tek kelimeleri sayarak bir örtüşme oranı hesaplarken, ROUGE-2, bigram (iki kelime) benzerliğini ölçen bir metriktir. Burada amaç, referans özet ile modelin özetindeki iki ardışık kelimenin (bigram) örtüşmesini değerlendirmektir. ROUGE-L, uzunluk bazlı en uzun ortak alt diziyi (Longest Common Subsequence - LCS) ölçen bir metriktir. Bu metrik, özetlerin içerdiği kelimelerin sırasına ve yapısına daha fazla odaklanır. ROUGE-L, özetin anlamını ve sırasını ne kadar doğru şekilde koruduğunu ölçer.



Şekil 4. MT5-Large ve SentenceTransformer+T5-Large Modelleri Performans Sonuçları

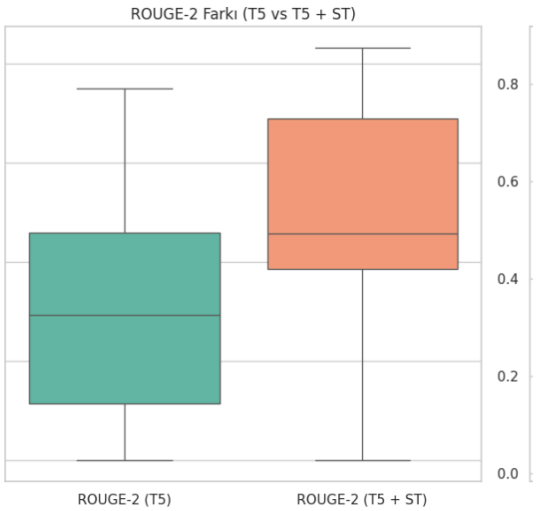
Figure 2 incelendiğinde, SentenceTransformer'ın (ST) kullanımı, özetleme performansını bazı durumlarda artırdığı görülmektedir. ROUGE-1, ROUGE-2 ve ROUGE-L metriklerinde,

SentenceTransformer'ın etkisi bazı örneklerde olumlu olmuştur. Örneğin, ROUGE-1 (T5) skoru 0.52'den 0.60'a yükselmiş, aynı şekilde ROUGE-2 (T5) skoru da 0.48'den 0.54'e çıkmıştır. ROUGE-L skorları da benzer şekilde iyileşmiştir. Özellikle, bazı örneklerde ROUGE-1, ROUGE-2 ve ROUGE-L değerlerinde 0.712, 0.591, 0.712 gibi yüksek sonuçlar elde edilmiştir. Bu sonuçlar, SentenceTransformer'ın bazı metin içerisindeki önemli cümleleri daha iyi tespit edebildiğini ve bu sayede T5 modelinin özetleme başarısını artırdığını göstermektedir. Şekil 4 de görülen bazı örneklerde ise SentenceTransformer'ın kullanılması ile ROUGE skorları oldukça düşük kalmıştır. Örneğin, ROUGE-1 (T5) skoru 0.04 gibi çok düşük bir değere sahip örnekler bulunmaktadır. Bu tür düşük skorlar, özetlemenin kalitesiz veya yetersiz olduğu, ya da metnin karmaşıklığından kaynaklanan zorluklar olduğunu gösterebilir. ROUGE-2 ve ROUGE-L metriklerinde de benzer şekilde düşük değerler gözlemlenmiştir. Bazı örneklerde, SentenceTransformer'ın etkisi sınırlı olmuş ve ROUGE-1 skoru, sadece 0.05'e yükselmiştir. Bu, SentenceTransformer'ın yalnızca sınırlı ölçüde fayda sağladığı ya da seçilen cümlelerin modelin özetleme performansını önemli ölçüde iyileştiremediği durumları işaret edebilir. Bazı örneklerde, SentenceTransformer ile yapılan işlem sonucunda özetlemenin kalitesinde belirgin bir iyileşme gözlemlenmiştir. Örneğin, ROUGE-1 (T5) skoru 0.44'ten 0.74'e, ROUGE-2 (T5) skoru 0.34'ten 0.69'a, ROUGE-L (T5) skoru ise 0.44'ten 0.47'ye çıkmıştır. Bu tür artışlar, SentenceTransformer'ın belirli metinlerde anlamlı cümleleri seçerek, özetin kalitesini artırmaya yardımcı olduğunu ve modelin daha doğru özetler üretmesini sağladığını göstermektedir. Bazı örneklerde ROUGE skorlarında dalgalanma görülmektedir. Örneğin, ROUGE-1 (T5) skoru 0.71 gibi yüksek bir değerden 0.11'e kadar düşerken, ROUGE-2 ve ROUGE-L metriklerinde de benzer bir düşüş yaşanmıştır. Bu tür dalgalanmalar, bazı metinlerin özetlenmesinin özellikle zor olduğuna işaret edebilir. Bu tür metinler genellikle karmaşık yapılar, T5 modelinin kelime temsilinin SentenceTransformer kelime temsili ile uyumsuzluğunu ve çok sayıda teknik terim veya anlamı zor cümleler içerebilir, bu da modelin başarısını etkileyebilmektedir. Şekil 4 de özellikle dikkat çeken yüksek performanslı örnekler, ROUGE-1 ve ROUGE-2 metriklerinde sırasıyla 0.79 ve 0.75 gibi yüksek değerler elde edilmiştir. Bu örnekler, SentenceTransformer ve T5 modelinin birleşimiyle elde edilen yüksek kaliteli özetleri temsil etmektedir. Bu, SentenceTransformer'ın anlamlı ve önemli cümleleri doğru bir şekilde özetlere dâhil edebilmesinin bir göstergesidir.



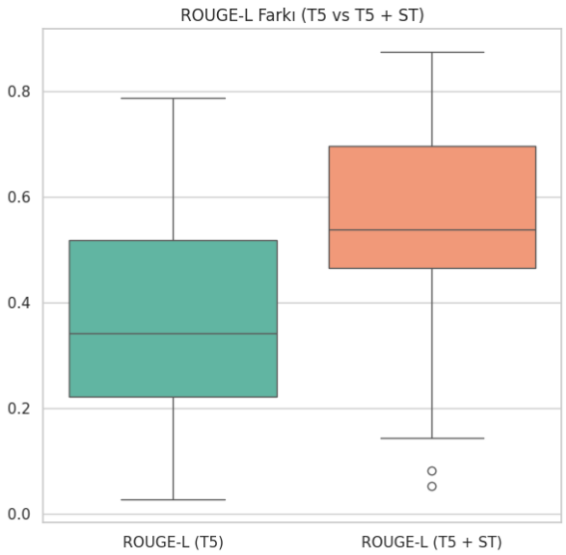
Şekil 5. İyi Performans Gösteren Örneklerin ROUGE-1 (T5) ile ROUGE-1 (T5 + ST) Karşılaştırması

Şekil 5 den anlaşılacağı gibi kosinüs benzerliklerinin kullanımda iyi performans gösteren örneklerin T5-Large'a göre ortalamada belirgin iyileşme gözlemleniyor. SentenceTransformer, T5 modeline kıyasla ROUGE-1 skorlarında ortalama olarak kayda değer bir artış sağlamaktadır. Bu, SentenceTransformer'ın metnin genel anlamını ve önemli cümlelerini daha iyi yakalayarak özetleme kalitesini artırdığını göstermektedir. SentenceTransformer, özellikle kelime ve cümle düzeyindeki ilişkileri daha etkili bir şekilde modelleyerek, özetin anlam bütünlüğünü iyileştirmekte ve dolayısıyla ROUGE-1 skorlarında önemli bir artışa yol açmaktadır.



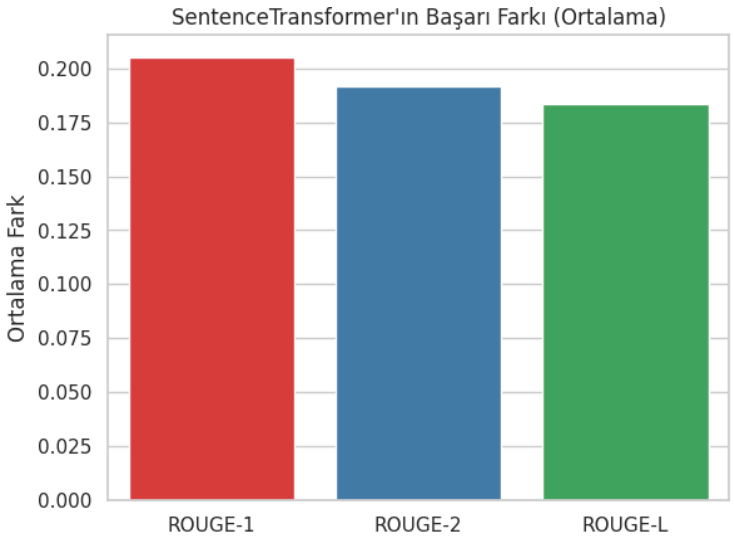
Şekil 6. İyi Performans Gösteren Örneklerin ROUGE-2 (T5) ile ROUGE-1 (T5 + ST) Karşılaştırması

ROUGE-2 skorlarında ise (Şekil 6), kosinüs benzerliklerinin kullanımda iyi performans gösteren örneklere etkisi daha ılımlı olmuştur. Bu skor, genellikle iki kelime arasındaki n-gram benzerliğini ölçtüğü için, anlam düzeyindeki iyileşmeler kadar belirgin değildir. Ancak, yine de SentenceTransformer'ın etkisi görülmektedir, çünkü bazı örneklerde büyük artışlar kaydedilmiştir. Bu durum, SentenceTransformer'ın, özellikle daha uzun ve anlamlı n-gram'lar için daha fazla bilgi sunarak, modelin kelime ve ifade ilişkilerini daha etkili bir şekilde öğrenmesine olanak tanıdığına işaret etmektedir.



Şekil 7. İyi Performans Gösteren Örneklerin ROUGE-L (T5) ile ROUGE-L (T5 + ST) Karşılaştırması

ROUGE-L (Longest Common Subsequence) skoru, cümleler arasındaki en uzun ortak alt diziyi ölçer ve metnin genel anlamının yanı sıra yapısal bütünlüğünü de dikkate alır. Burada gözlemlenen orta düzeydeki iyileşmeler (Şekil 7), kosinüs benzerliklerinin kullanımda iyi performans gösteren örneklerde metnin uzun vadeli yapısal ve anlamsal ilişkilerini anlamada daha az etkili olduğunu düşündürmektedir. Bununla birlikte, SentenceTransformer hala özetin bağlamını geliştirmekte ve modelin sonuçlarını iyileştirmektedir, ancak bu etki ROUGE-1 ve ROUGE-2'ye kıyasla daha sınırlıdır.



Şekil 8. İyi Performans Gösteren Örneklerin SentenceTransformer'ın Başarı Farkı (Ortalama)

İyi performans gösteren örneklerin genel olarak (Şekil 8), SentenceTransformer'ın T5 modelinin özetleme performansını iyileştirmede olumlu bir etkisi olduğu söylenebilir. Özellikle ROUGE-1'deki belirgin artış, SentenceTransformer'ın özetleme sürecine katkılarının anlamın doğruluğunu artırmaya yönelik olduğunu gösteriyor. ROUGE-2'deki daha küçük iyileşmeler, bu katkının daha çok kelime düzeyindeki n-gram benzerliklerine dayalı olduğunu, ROUGE-L'deki daha düşük iyileşmeler ise metnin yapısal bütünlüğü üzerinde yapılan iyileştirmelerin sınırlı olduğunu göstermektedir.

Bu bulgular, kullanılan metodun özellikle anlam düzeyindeki iyileştirmeler konusunda güçlü olduğunu ve metnin özünü daha iyi kavrayarak özetlemeyi daha etkili hale getirdiğini, ancak yapısal ve n-gram bazlı iyileştirmelerde etkisinin daha sınırlı kaldığını ortaya koymaktadır.

T-test sonuçları, T5 ve T5 + Kosinüs Benzerlik Tabanlı İçerik Seçimi yöntemleri arasındaki farkın istatistiksel olarak anlamlı olduğunu göstermektedir. Her üç ROUGE metriği için yapılan t-testlerinde, p-değerlerinin son derece düşük olması, farkın rastlantısal olmadığını ve Kosinüs Benzerlik Tabanlı İçerik Seçimi kullanımının özetleme performansını bazı örneklerde belirgin şekilde iyileştirdiğini doğrulamaktadır. Bu istatistiksel bulgular, kosinüs benzerliğiyle seçilen cümlelerin metnin anlamını doğru şekilde yansıtarak, özetlerin n-gram yapısını ve anlam bütünlüğünü koruyarak kalitesini artırdığını ortaya koymaktadır.

5. Sonuç

Bu bulgular, kosinüs benzerlik tabanlı içerik seçimi ile özetleme görevlerinde T5 modeline bütünleştirilerek bazı örneklere performans artışı sağladığını ve bu artışın istatistiksel olarak anlamlı olduğunu göstermektedir. Bu iyileştirmeler, Türkçe özetleme gibi dil işleme görevlerinde daha doğru ve anlamlı metinler üretmek için güçlü bir adım atıldığını ortaya koymaktadır.

Özellikle bazı örneklerde, SentenceTransformer'ın kullanımı ile özetlerin ROUGE skorları önemli ölçüde artmıştır. ROUGE-1 ve ROUGE-2 skorlarında sırasıyla 0.52'den 0.60'a, 0.48'den 0.54'e kadar iyileşmeler gözlemlenmiştir. Daha yüksek performanslı örneklerde, ROUGE-1 skoru 0.71'e, ROUGE-2 skoru ise 0.75'e kadar yükselmiştir. Bu iyileşmeler, SentenceTransformer'ın metin içindeki anlamlı ve önemli cümleleri daha iyi tespit ederek, T5 modelinin özetleme başarısını artırdığına işaret etmektedir.

Ancak, tüm örneklerde benzer iyileşmeler gözlemlenmemiştir. Bazı metinlerde, kosinüs benzerlik tabanlı içerik seçimi ile özetleme etkisi sınırlı kalmış ve ROUGE skorlarında düşüşler yaşanmıştır. Örneğin, bazı örneklerde ROUGE-1 skoru sadece 0.04'e kadar düşerken, ROUGE-2 ve ROUGE-L metriklerinde de düşük sonuçlar alınmıştır. Bu, özetlemenin kalitesiz veya yetersiz olduğu, iki modelin kelime vektör düzeyindeki uyumsuzluğu, ya da metnin karmaşıklığından kaynaklanan zorlukların etkisiyle olduğu varsayılabilir. Ayrıca, metinlerdeki teknik terimler veya anlam karmaşıklıkları, modelin özetleme performansını olumsuz etkilemesi nedenleri arasında gösterilebilir.

Bu araştırmanın gelecekteki çalışmalara katkı sağlayabilecek birkaç önerisi bulunmaktadır:

- Farklı dil modellerinin ve diktonomi yöntemiyle performans karşılaştırılması: Çalışmamızda kullanılan model ve yöntemin yanı sıra, diğer dil modelleri ve benzer tekniklerinin de benzer görevler üzerindeki performansları karşılaştırılarak daha geniş bir perspektif elde edilebilir.

- Karmaşık veri kümeleri üzerinde denemeler: Özellikle Türkçe'nin zengin dil yapısı göz önüne alındığında, farklı veri kümeleri ve daha karmaşık metinler üzerinde yapılan denemelerle modelin performans değerlendirmesi yapılabilir.

Açıklama

Bu makale, 2024-2025 yıllarında yürütülen "Mevcut Derin Doğal Dil İşleme Modellerinin Türkçe Özelinde Performanslarının Artırılması" başlıklı ilk yazarın doktora tezinden üretilmiştir.

Referanslar

Anonim, 2024a. Tensorflow Kelime Projeksiyonları.

Anonim, 2020b. Schematic Diagram of Cosine Similarity and Euclidean Distance.

Avrupa Komisyonu, 2019. Avrupa Yeşil Mutabakatı. Avrupa Birliği Yayın Ofisi. (https://eur-lex.europa.eu/legal-content/TR/TXT/?uri=CELEX%3A52019DC0640), (Erişim tarihi: 29.02.2025).

Bakan, C.T., Yakut, S., 2024. Graf teorisi ve malatya merkezilik algoritmasına dayalı haber metinlerinin özetlemesi. *Bilişim Teknolojileri Dergisi,* 17(3): 189-198.

Gopalakrishnan, S., Garbayo, L., Zadrozny, W., 2025. Causality extraction from medical text using large language models (LLMs). *Information*, 16(1): 13.

Güngör, M., 2023. Wikipedia TR Summarization Dataset [Veri kümesi]. Hugging Face.

Harris, Z.S., 1954. Distributional Structure. *WORD,* 10(2–3): 146–162.

Johnson, S.J., Murty, M.R., Navakanth, I., 2024. A detailed review on word embedding techniques with emphasis on word2vec. *Multimedia Tools and Applications,* 83(13): 37979-38007.

Karakoç, E., Yılmaz, B., 2019. Deep learning based abstractive Turkish news summarization. In *2019 27th Signal Processing and Communications Applications Conference (SIU)*, pp. 1-4, IEEE.

Kartal, Y.S., Kutlu, M., 2020. Türkçe haber metinleri için makine öğrenmesi temelli özetleme. In *2020 28th Signal Processing and Communications Applications Conference, SIU 2020-Proceedings*. Institute of Electrical and Electronics Engineers Inc.

Khurana, D., Koli, A., Khatter, K., Singh, S., 2023. Natural language processing: State of the art, current trends and challenges. *Multimedia Tools and Applications,* 82(3): 3713-3744.

Mana, S.C., Sasipraba, T., 2021. Research on cosine similarity and Pearson correlation based recommendation models. In *Journal of Physics: Conference Series* 1: 012014.

McDonald, S., Ramscar, M., 2001. Testing the distributional hypothesis: The influence of context on judgements of semantic similarity. *Proceedings of the Annual Meeting of the Cognitive Science Society,* 23: 23.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ...Liu, P.J., 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research,* 21(140): 1-67.

Rahutomo, F., Kitasuka, T., Aritsugi, M., 2012. Semantic cosine similarity. *The 7th International Student Conference on Advanced Science and Technology ICAST*, p. 1, South Korea: University of Seoul.

Reimers, N., 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *arXiv preprint arXiv:1908.10084.*

Şakar, T., Emekci, H., 2025. Maximizing RAG efficiency: A comparative analysis of RAG methods. *Natural Language Processing,* 31(1): 1-25.

Salton, G., Buckley, C., 1988. Term-weighting approaches in automatic text retrieval. *Information Processing and Management,* 24(5): 513−523.

TeraGron, 2021. Wikisum: A dataset for summarization of Wikipedia articles. Hugging Face. (https://huggingface.co/datasets/teragron/wikisum), (Erişim Tarihi: 25.02.2024)

Xia, P., Zhang, L., Li, F., 2015. Learning similarity with cosine similarity ensemble. *Information Sciences,* 307: 39-52.