

Graf Teorisi ve Malatya Merkezilik Algoritmasına Dayalı Haber Metinlerinin Özetlemesi

Araştırma Makalesi/Research Article

 Cevher Tayyib BAKAN^{1*},  Selman YAKUT^{1*}

¹Yazılım Mühendisliği, İnönü Üniversitesi, Malatya, Türkiye
cevherbakan@gmail.com, selman.yakut@inonu.edu.tr
(Geliş/Received:01.04.2024; Kabul/Accepted:16.05.2024)
DOI: 10.17671/gazibtd.1463107

Özet— Günümüzde internetin yaygın kullanımıyla, bilgi kaynaklarındaki doğru bilgiye erişimi önemli kılmaktadır. Bilgi kaynaklarının artmasıyla birlikte özgün içeriğe sahip bilgiye erişim güçleşmektedir. Bu nedenle metin özetleme yöntemlerinin önemi giderek artmaktadır. Haber metinleri gibi önemli temel bilgi kaynaklarının etkili bir şekilde özetlenmesi günümüzde bir gereklilik haline gelmiştir. Bu çalışmada haber metinlerinin etkili bir şekilde özetlenmesi için Malatya merkezilik algoritmasını temel alan bir özetleme yaklaşımı önerildi. Önerilen yaklaşımda orijinal metin tanımlayıcıların çıkarılması, kelime köklerinin elde edilmesi gibi çeşitli ön işlemlerden geçirilerek graf yapısına dönüştürülür. Graf'a dönüştürülen metin için Malatya merkezilik algoritması kullanılarak graftaki düğümlerin Malatya merkezilik değerleri hesaplanır. Bu değerler dikkate alınarak metin özetini oluşturan özetler seçilir. Seçilen özetler graftan çıkarılır. Oluşan yeni graf yapısı için merkezilik değeri hesaplanarak seçim işlemleri devam ettirilir. Graf Teorisi ve Malatya merkezilik algoritmasının birlikte kullanımı, haber metinlerinin özetlenmesinde verimliliği artırdığı gösterildi. Bununla birlikte haber içeriklerinin anlamlı bir şekilde özetlenmesi sağlandı. Bu yaklaşımın başarısını değerlendirmek amacıyla BBC veri seti üzerinde toplamda 2224 İngilizce haber metniyle kapsamlı bir şekilde test edildi. Çalışmada haber metinleri etkili bir şekilde özetlendiği yapılan testlerle ve alınan etkili rouge değerleriyle gösterildi. Graf teorisi ve Malatya merkezilik algoritması, bilgiye erişimi kolaylaştırmak ve anlam düzeyinde etkileşimi artırmak adına önemli bir potansiyele sahip olduğu gösterildi. Elde edilen uygulama sonuçları, haber metinlerini daha anlamlı bir şekilde sunabileceğini ve etkili özetler üretilebileceğini göstermektedir.

Anahtar Kelimeler— haber metinleri özetleme, graf tabanlı özetleme, malatya merkezilik algoritması, çıkarım tabanlı özetleme yöntemleri, metin işleme

Summarization of News Texts Based on Graph Theory and Malatya Centrality Algorithm

Abstract— The widespread use of the internet today highlights the importance of accessing accurate information from reliable sources. With the proliferation of information sources, accessing original content has become increasingly challenging. Therefore, the importance of text summarization methods is steadily increasing. Effectively summarizing essential information sources such as news articles has become a necessity. In this study, an approach based on the Malatya centrality algorithm is proposed for effectively summarizing news articles. In the proposed approach, various preprocessing steps such as extracting descriptive terms and obtaining word roots are applied to transform the original text into a graph structure. The Malatya centrality algorithm is then utilized to calculate the centrality values of nodes in the graph representing the text. Based on these values, summaries constituting the text summary are selected. The selected summaries are removed from the graph. Centrality values are then calculated for the resulting new graph structure, and the selection process continues. The combined use of Graph Theory and the Malatya centrality algorithm is shown to enhance efficiency in summarizing news articles. Additionally, meaningful summarization of news content is achieved. To evaluate the success of this approach, it was comprehensively tested on a total of 2224 English news articles from the BBC dataset. The study demonstrates effective summarization of news articles through conducted tests and achieved effective ROUGE scores. The utilization of Graph Theory and the Malatya centrality algorithm is shown to have significant potential in facilitating information access and increasing interaction at the semantic level. The obtained application results indicate the ability to present news articles in a more meaningful manner and produce effective summaries.

Keywords— news texts summarization, graph based summarization, malatya centrality algorithm, inference based summarization methods, text processing

1. GİRİŞ (INTRODUCTION)

Günümüzde, bilgi kaynaklarının ve internetin gelişimiyle birlikte birçok bilgi kaynağına erişilmekte ancak bu kaynaklar içinde doğru ve anlamlı içeriğe ulaşmak giderek zorlaşmaktadır. Metin özetleme, bu bilgi yoğunluğu altında önemli bir araç haline geldi. Özellikle haber metinleri gibi uzun ve detaylı içeriklerin özünü kavramak, anlamlı bir şekilde özetlemek, günümüzün bilgi çağında vazgeçilmez bir gerekliliktir. Bu bağlamda, geleneksel metin özetleme yöntemleri zaman zaman sınırlamalarla ve hatalarla karşılaşır. İnternet teknolojisinin gelişmesiyle birlikte, kullanıcılar geniş bir bilgi yelpazesine erişebilmekte ancak doğru, özgün ve öne çıkan içeriği belirlemek oldukça zor hale gelmektedir. Metinlerde tekrar eden ifadeler, gereksiz detaylar, yoğunluk asıl mesajı bulmayı ve önemli bilgileri belirlemeyi zorlaştırır.

Metin özetlemede çıkarıma dayalı ve yoruma dayalı olmak üzere temel olarak iki yöntem kullanılır. Çıkarıma dayalı özetleme, temelde metindeki önemli ifadelerin tespitini amaçlar. Bu süreçte, kelimelerin yinelenmesi ve benzerleri göz önünde bulundurularak ana fikirler belirlenir. Öte yandan, yoruma dayalı özetleme ise derin öğrenme ve yapay sinir ağları gibi yöntemlerle, verilen metinden yeni ve anlamlı cümlelerle özet çıkarır. Her iki yönteminde olumlu ve olumsuz yanları bulunur. Yoruma dayalı özetlemede, yeni cümleler oluştururken anlam bozuklukları ortaya çıkabilir. Yoruma dayalı özetleme, çıkarıma dayalı özetlemeye göre çok daha uzun sürebilmektedir. Çıkarıma dayalı özetlemelerde, yeni cümle oluşturamazken metindeki en anlamlı cümleleri bir araya getirerek özet oluşturur. Bu yöntem çok daha hızlı ve anlam bozukluğu olmadan özetleri oluşturabilir.

Minimum tepe örtme problemi, bir graftaki bütün kenarların kapsandığı düğümlerinin minimum alt kümesini seçmeyi amaçlar. Bu seçilen düğümler, grafa ait tüm kenarları en az bir kez kapsamalıdır. Yani, seçilen düğümler sayesinde grafin tüm kenarlarıyla temas edilir. Bu problem, çeşitli uygulamalarda, özellikle ağ tasarımı ve optimizasyonunda kullanılır. Örneğin, bir iletişim ağındaki düğümleri temsil eden bir graf düşünülürse bu düğümleri bir alt küme olarak seçilip herhangi iki düğüm arasında iletişim kuracak bir bağlantı oluşturulabilir. Ancak bu bağlantıların oluşturulması maliyetlidir. En az maliyetle tüm düğümler arasında iletişim kurulmalıdır. Minimum tepe örtme problemi, bu durumda bize en az sayıda düğüm seçerek bu bağlantıların kurulmasını sağlayacak bir çözüm sunar. Bu problem, polinom zamanında çözülemeyen bir problem olarak bilinir. Ancak pratikte, çeşitli yaklaşımlar ve algoritmalarla çözümüne yaklaşmak mümkündür. Bu nedenle, Minimum tepe örtme problemi, bilgisayar biliminde ve birçok alanda önemli bir optimizasyon sorunudur. Yakut ve diğerleri, Minimum tepe örtme problemi için, çalışmasında Malatya vertex-cover algoritmasını sundular. [1] Bu algoritmanın metin özetleme üzerinde de başarılı olabileceği düşünüldü. Haber metinlerini özetlerken bu algoritma kullanılarak algoritmanın metin özetlemedeki başarısı da ortaya çıkarıldı.

Bu çalışma, BBC veri seti üzerinde gerçekleştirilen 2224 haber metni ile Malatya merkezilik algoritması ve graf teorisi gibi tekniklerin performansını değerlendirerek haber metinlerinin özetlenmesinde yeni bir yaklaşım sunar. Önerilen yaklaşımda haber metinleri tanımlayıcıların

atılması, kelime köklerinin belirlenmesi gibi belirli önışlemlerden geçirilerek graf yapısına dönüştürülür. Daha sonra oluşturulan bu graf yapısına Malatya merkezilik algoritması uygulanarak graftaki düğümlerin Malatya merkezilik değerleri hesaplanır. Bu algoritma, metin içindeki temel yapıları vurgulayarak özünü koruyan, anlaşılır bir özetleme süreci sunabilmektedir. Ayrıca çalışma, daha önceki araştırmalarda belirtilen graf temelli çıkarımsal ve yoruma dayalı yöntemleri de göz önüne alır. Bilgi çağının gerekliliklerine uygun olarak daha etkili ve anlaşılır metin özetleme yöntemlerinin geliştirilmesine katkı sağlamayı hedefler. BBC veri seti üzerinde gerçekleştirilen 2224 haber metni üzerinde yapılan deneyler ve üretilen Rouge değerleri önerilen yöntemin haber kaynakları gibi farklı metinleri içerebilen metinlerde etkili metin özetleme yapıldığını gösterir.

Bu makale çalışmasının devamında şu bölümler yer almaktadır: bölüm 2’de önerilen çalışma ile ilgili literatüre yer verildi. Bölüm 3’te önerilen algoritma yer almaktadır. Bölüm 4’te çalışmanın deneysel sonuçlarına yer verildi. Bölüm 5’te ise sonuçlar yer almaktadır.

2. LİTERATÜRDE İLGİLİ ÇALIŞMALAR (LITERATURE RELATED STUDIES)

Metin özetleme, akademik literatürde farklı şekillerde ele alınan geniş bir konu başlığıdır. Bu alandaki çalışmalar, graf tabanlı çıkarma yöntemlerini kullananlar olduğu gibi, makine öğrenimi yöntemleri ile yorumlama odaklı yaklaşımları benimseyenleri de barındırır. Literatürdeki bu yaklaşımları ve uygulama alanlarını farklı şekillerde sınıflandırmak ve ele almak mümkündür. Bununla beraber bu makalede daha çok literatürdeki çıkarsama odaklı yöntemlere ilişkin örnek çalışmalar incelendi.

Bakan ve Yakut, metin özetleme işlemi için graf teorisi ve Malatya merkezilik algoritması kullanmıştır. [2] Bu çalışmada örnek bir metin graf yapısına dönüştürülerek bu metinde özet çıkarma işlemi gerçekleştirilmiştir. Önerilen algoritmanın, metin özetleme işleminin hızlı ve tam olarak gerçekleştirdiği örnek uygulamayla gösterilmiştir. Önerilen algoritma Malatya merkezilik algoritması kullanılarak metinler için etkili özetler üretildiğini göstermektedir.

Yakut ve diğerleri, minimum tepe örtme problemi için Malatya vertex-cover algoritması önermişlerdir. Bu algoritma, grafin düğümleri için Malatya merkezilik değerleri kullanılarak oluşturulur. İlk adımda grafin düğümlerinin Malatya merkezilik değerleri, Malatya merkezilik algoritması kullanılarak hesaplanır. İkinci adımda minimum tepe örtme çözümü için düğümler, graf içindeki maksimum Malatya merkezilik değerine sahip düğüm seçilerek ve çözüm kümesine eklenerek seçilir. Daha sonra bu düğüm ve bu düğüme bağlı olan kenarlar graftan çıkarılır. Kalan düğümlerden oluşan graf için Malatya merkezilik değerleri tekrar hesaplanır ve seçim işlemi devam eder. İşlem, grafin tüm kenarları kaplandığında sona erer. [3]

Tülek, Türkçe metinlerin özetlerini oluşturmak için kök çıkarma algoritmaları ve kök-ek analizini kullanılmış ve ardından özetlenecek cümleler belirlenmiş. [4] Khushboo ve diğerleri ise graf sıralama algoritmalarını en kısa yol algoritmalarıyla birleştirerek otomatik cümle çıkarma tekniklerini geliştirdi. [5] Erkan ve Dragomi ise LexRank yaklaşımını benimseyerek cümlelerin önemini, cümleler

arası kosinüs benzerliğine dayalı olarak belirlendi.[6] Moawad ve Aref, orijinal belgeyi özetlemek için zengin bir anlamsal graf kullandı ve bu grafik üzerinden daha kısa bir belge oluşturdu. [7] Ferreira ve diğerleri, metin işleme uygulamaları için dört ayrı boyutta temellendirilmiş bir graf modeli sundu. [8]

Mallick ve diğerleri, her bir web sayfasına özgü olarak düzenlenmiş bir TextRank metodu kullanarak belgenin özünü yakalayan bir graf-tabanlı metin özetleme yöntemi geliştirdi.[9] Sankarasubramaniam vd. Vikipedi'yi grafik sıralama ile birleştirerek orijinal metne başlık cümleleri ekleyerek yeni bir yaklaşım geliştirdi.[10] Alguliev ve Aliguliyev ise belgesiz bir özetleme yöntemi üzerinde çalıştı, orijinal belgeden cümlelerin kümelmesi ve seçilerek özet oluşturulması konusunda yeni kriter fonksiyonları sundular.[11] Nagwani ve Verma, Java programlama dilinde sık kullanılan terimlere dayalı olarak bir metin özetleme algoritması tasarladı ve uyguladı.[12] Mihalcea'nın çalışmasında, otomatik cümle çıkarma için graf-tabanlı sıralama algoritmaları kullanılan bir yöntem sunulmaktadır.[13] Akter ve diğerleri, tek veya çoklu Bengalce belgelerinden önemli cümleleri çıkaran bir metin özetleme yöntemi önermişlerdir. Bu yaklaşımda, giriş belgeleri tokenizasyon ve kök çıkarma gibi ön işlemlerden geçmekte; daha sonra Terim Frekans-Ters Belge Frekansı (TF/IDF) kelime puanlarını hesaplamakta ve bu kelimelerin puanları cümle puanlarını belirlemek için toplanmaktadır. Ayrıca, çalışma K-ortalama kümeleme algoritmasını kullanmaktadır.[14]

Literatürdeki diğer çalışmalar, duygusal temelli yaklaşımların önemli ilgi gördüğünü göstermektedir. Babar ve Patil'in çalışması, metin özetleme için Fuzzy Mantık Çıkarımı ve Latent Semantic Analysis gibi anlamsal yaklaşımlara odaklanmaktadır.[15] Ayrıca Ozsoy ve diğerleri, Latent Semantic Analysis (LSA) tabanlı özetleme algoritmalarını tanıtmış ve bu öneriler de makalenin yazarlarına aittir. [16]

Makine öğrenimi ve derin öğrenme gibi teknikleri kullanan bazı çalışmalar da incelendi. Erhandı, derin öğrenme yaklaşımını kullanarak Türkçe ve İngilizce metinler üzerinde bir özetleme çalışması yürütmüş.[17] Neto ve diğerleri, orijinal metinden doğrudan çıkarılan özellikler kullanarak eğitilebilir makine öğrenimi algoritmaları kullanarak metin özetleme sürecini sunmuşlar.[18] Silla vd. metin özetleme sorununu bir sınıflandırma problemi olarak ele alarak otomatik özetleme yöntemi geliştirmişlerdir. Çalışmanın amacı, sınıflandırma algoritmalarının performansını artırmak için Genetik Algoritma tabanlı özellik seçiminin etkinliğini araştırmaktır.[19]

Genetik algoritmaları kullanan birçok çalışmada incelenmiştir. Kaynar'ın çalışmasında, metin tabanlı özetleme için genetik algoritma kullanılmıştır. Bu genetik algoritma, veri setleri üzerinde sistemi eğitmek için kullanılmıştır.[20] Al-Abdallah ve Al-Taani, Arapça belge özetleme için Parçacık Sürü Optimizasyonu algoritmasını önermişler. Önerilen yaklaşım, Evrimsel Algoritmalar ve Harmony Search yöntemleriyle karşılaştırılmıştır.[21]

Karçioğlu ve Yaşa, otomatik metin özetleme için genetik algoritmaların başarı oranlarının yüksek olduğu ve bu alanda daha iyi bireyleri seçerek problemi çözebileceği gösterilmiştir. Gelecekte, makine öğrenimi algoritmalarıyla genetik algoritmanın uyum fonksiyonunu geliştirmeyi ve

farklı yapay zeka yöntemlerini bir araya getirerek sistem performansını artırmayı hedeflemektedirler.[22]

Karakoç ve Yılmaz, derin öğrenme yöntemleriyle Türkçe haberlerin özetlenmesi üzerine odaklanılmıştır. Makalede, haber başlıklarının tahmin edilmesinde kodlayıcı-kod çözücü modeli kullanılarak yoruma dayalı metin özetleme yöntemi uygulanmıştır. Yapılan deneysel çalışmalar, haber metninin tamamının eğitimiyle elde edilen modelin, özetleme performansı açısından diğer modellere göre daha başarılı olduğunu göstermiştir.[23]

Torun ve İner, internet haberciliğinin geleneksel medya üzerindeki etkilerini inceleyerek, TextRank algoritmasıyla Türkçe haber metinlerini özetleyerek haber vektörleri oluşturmuş ve benzerlik tespiti için TF-IDF ve Doc2Vec kullanmıştır. Uzmanlarca değerlendirilen sistem, detaylı bir başarı analizi sunmuştur.[24]

Kartal ve Kutlu, haber metinlerini en iyi temsil eden cümleleri seçerek otomatik bir özetleme sistemi öneriyor. Bu model, cümle konumu, anlatım ifadeleri, varlık isimleri, kelime sıklığı ve başlık benzerliği gibi öznitelikleri kullanarak geliştirilmiş ve LSA tabanlı metotlardan daha başarılı sonuçlar elde etmiştir.[25]

Hark ve diğerleri, entropi tabanlı metin özetleme yöntemi üzerine yapılan çalışma inceleniyor. Bu metodun, metinlerin yapısını ve içerdiği bilgi miktarını değerlendirerek ROUGE metrikleriyle başarılı bir performans gösterebildiği belirtiliyor. Entropi tabanlı bu yöntem, diğer metotlara göre umut verici sonuçlar sunarak araştırmacıların dikkatini çekiyor.[26]

Wang vd. biyomedikal literatür ve elektronik sağlık kayıtları üzerindeki özetleme tekniklerini ve değerlendirme yöntemlerini inceledi. 2013 ile 2021 arasındaki 7235 makaleden 58'i incelenerek, mevcut sistemlerin çoğunlukla literatür özetlemeye odaklandığı, hibrit yöntemlerin başarılı olduğu ve elektronik sağlık kayıtlarının özetlenmesine dair çalışmaların arttığı belirlendi.[27]

Kumar vd. farklı dillerdeki metin özetleme yöntemlerini inceleyerek Hindistan ve yabancı dillerde gerçekleştirilen çalışmaları karşılaştırır. Genellikle makine ve derin öğrenme temelli sınıflandırma yöntemlerinin başarılı sonuçlar verdiğini belirtirken, gelecekteki araştırma alanları olarak özellik çıkarma, sınıflandırma yöntemlerinin geliştirilmesi ve farklı dil veri tabanlarının bulunabilirliği gibi konulara işaret eder.[28]

Uçkan vd. çoklu belgelerden özet çıkarmak için CatSumm adında yeni bir yöntem tanıtıyor. Bu yöntem, metinlerin temsili, spektral grafik bölütleme ve cümlelerin puanlanması üzere üç aşamadan oluşuyor. Yapılan deneyler, önerilen CatSumm metodunun DUC-2002 ve DUC-2004 veri setlerinde %44.073 ile %56.513 arasında başarılı özetleme sonuçları verdiğini gösteriyor 2021.[29]

Aydın ve Uçkan, metin özetleme için farklı benzerlik ölçütleri ve çizge temsilleri kullanarak çıkarımsal metin özetleme yöntemlerini incelemiştir. Kosinüs Benzerliği, Jaccard Benzerliği, Levenshtein Benzerliği ve Pearson Korelasyon Katsayısı gibi ölçütlerle çizgeler oluşturulmuş, ardından Arasındalık Merkeziliği, Yakınlık Merkeziliği, Derece Merkeziliği ve Özvektör Merkeziliği ölçütleriyle en değerli cümleler tespit edilerek 200 ve 400 kelimelik özetler üretilmiştir. DUC-2002 veri seti üzerinde yapılan deneysel

çalışmalar sonucunda, Jaccard Benzerliği ve Yakınlık Merkeziliği ile 200 kelimelik, Kosinüs Benzerliği ve Özvektör Merkeziliği ile 400 kelimelik özetlerin en yüksek performansı sergilediği belirlenmiştir.[30]

Kuş ve Acı, Türkçe tıbbi metinlerin özetlenmesi üzerine odaklanıyor. COVID-19'a ilişkin 84 makale üzerinde çalışarak, Term Frekansı ve LexRank gibi algoritmalarla genişletilmiş özetler elde etti. Yapılan değerlendirmeler, özetlerin orijinal metinlere benzediğini gösteriyor. Gelecekte daha derin öğrenme metodlarıyla çalışmayı hedefliyor.[31]

Literatürde metin özetleme için farklı şekillerde sınıflandırılabilir birçok yaklaşım önerilmekle beraber bu yaklaşımları genel olarak yorumlayıcı ve çıkarıcı metin özetleme yaklaşımları olarak ele almak mümkündür. Yorumlayıcı metin özetleme yaklaşımları metinde anlamı dikkate alarak özetleme yaparken çıkarımsal yaklaşımlar metindeki ifadeleri özet belirlemede kullanılmaktadır. Bu yaklaşımların kullanım amacı, kullanılan dilin türü, başarı ölçütü gibi parametreler dikkate alınarak değerlendirilebilir. Benzer algoritmaların farklı dil türlerinde veya farklı içeriğe ve amaca sahip veri setlerinde metin özetleme gerçekleştirilebilmektedir. Özellikle haber kaynakları gibi farklı alanlara hitap edebilen yaklaşımların kullanılması etkili metin özetleme açısından önemlidir.

3. ÖNERİLEN YÖNTEM (RECOMMENDED METHOD)

Bu çalışmada haber metinlerinin özetlenmesinde etkili bir algoritma önerildi. Malatya merkezilik algoritması, metinlerin anlamını temsil eden etkili özetlerin elde edilmesi için önemli kelimeleri belirlenmesi ve metin özetlerinin oluşturulması amacıyla kullanıldı. Bu süreç, birkaç adımdan oluşur. İlk olarak, gelen haber metinleri ön işleme adımları ile işlenir. Bu adımlar arasında metinlerin ayrıştırılması, tokenlara bölünmesi, gereksiz kelimelerin (stopwords) çıkarılması ve kelimelerin köklerinin belirlenmesi yer alır. Ardından, Malatya merkezilik algoritması, bu ön işlemde geçmiş metinleri kelime düğümlerinden oluşan bir graf olarak temsil eder.

Malatya merkezilik algoritması, kelime düğümleri arasındaki bağlantıları ve kelimelerin önemini belirlemek için kullanılır. Graf yapısındaki her kelimenin bağlantıları, metindeki diğer kelimelerle olan ilişkileri gösterir. Bu ilişkilerin incelenmesi ve analiziyle Malatya merkezilik skorları hesaplanır.

Skorlar, bir kelimenin metindeki konumu, sıklığı ve diğer kelimelerle olan ilişkileri temel alınarak belirlenir. Malatya merkezilik skorlarına göre en yüksek skorlara sahip kelimeler seçilir. Bu kelimeler, metnin ana fikrini ve en önemli noktalarını yansıtan kelimelerdir. Bu seçilen kelimelerin içerdiği cümleler birleştirilerek özet metin oluşturulur. Özet metin, özgün metnin temel noktalarını içerir ve orijinal metnin anlamını korurken daha kısa ve öz bir yapı sunar. Bu yöntem, Malatya merkezilik algoritmasının kullanımıyla haber metinlerinin özetlenmesini sağlar. Graf tabanlı bu yaklaşım, metnin içeriğini temsil eden önemli kelimeleri belirler. Bu önemli kelimeler üzerinden metnin özeti oluşturulur.

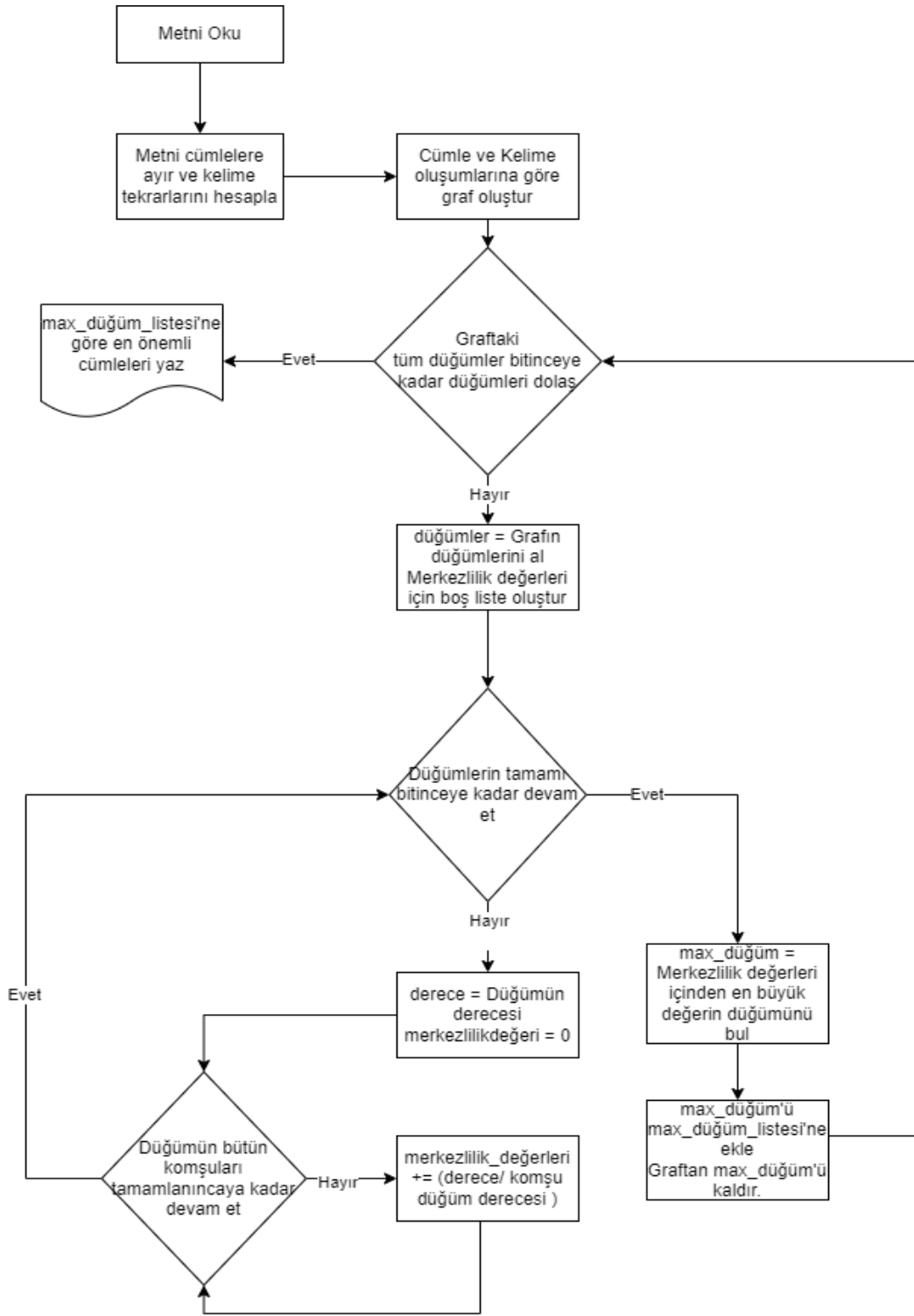
Şekil-1'de, metinde yer alacak kelimelerin belirlenmesi sürecini gösteren algoritmanın akış şeması bulunmaktadır.

Bu şemada, öncelikle durak kelimelerinin belirlenmesi ve bunların graftan çıkarılması yer alır. Graf oluşturulduktan sonra her cümle graf üzerine yerleştirilir. Malatya merkezilik algoritması uygulanarak en etkili düğüm seçilir. Her aşamada, en etkili düğüm graftan çıkarılır ve yeni graftaki en etkili düğüm bulunur. Bu işlem düğümler bitinceye kadar tekrarlanır. Ardından, bu belirlenen en etkili düğümler birleştirilerek özetlenmiş metin oluşturulur. Şekil-1 de gösterilen şema, yeni yöntemin ayrıntılı adımlarını gösterir. Ayrıca metin özetlemede Malatya merkezilik algoritmasının kullanımını açıklar.

Önerilen algoritmaya ait sözde kodlar verildi. Bu kodda, başlangıçta orijinal metinden tanımlayıcı unsurların çıkarılması ve kelime köklerinin belirlenmesi gibi çeşitli ön işlemler uygulanır. Bu ön işlemler sonucunda metin, bir graf yapısına dönüştürülür. Daha sonra, bu graf yapısındaki düğümlerin Malatya merkezilik değerleri, Malatya merkezilik algoritması kullanılarak hesaplanır. Bu değerler doğrultusunda, metin özetini oluşturacak özetler belirlenir. Seçilen özetler graf yapısından çıkarılır. Ardından, oluşan yeni graf yapısı için merkezilik değeri hesaplanarak seçim işlemleri sürdürülür. Metin özeti elde edildiğinde merkezilik değerlerinin hesaplanması ve seçim işlemleri tamamlanır.

Önerilen algoritmanın sözde kodu

```
Fonksiyon: MerkezilikHesapla(graf)
düğümler = GrafDüğümünüAl(graf)
merkezilik_değerleri = []
Her bir düğüm için:
    derece = GrafDerecesiniAl(graf, düğüm)
    merkezilik_değeri = 0
    Her bir komşu için:
        merkezilik_değeri = merkezilik_değeri +
(derece/GrafDerecesiniAl(graf, komşu))
    merkezilik_değeri merkezilik_değerleri'ne ekle
    merkezilik_değerlerini döndür
Fonksiyon: MaksMerkezilikDeğeriBul(graf)
merkezilik_değerleri = MerkezilikHesapla(graf)
maks_indeks = MaksIndexiBul(merkezilik_değerleri)
maks_düğüm = GrafDüğümünüIndexleAl(graf, maks_indeks)
GrafDüğümüKaldır(graf, maks_düğüm)
maks_düğümü döndür
Fonksiyon: MaxDüğümListesiHesapla(graf)
Graftaki tüm düğümler bitinceye kadar:
    maks_düğüm = MaksMerkezilikDeğeriBul(graf)
    max_düğüm'ü listeye ekle
    maks_düğüm_listesini sırala ve döndür
Fonksiyon: GrafOluştur(cümleler, kelime_oluşumları)
durak_kelimeler = DurakKelimeleriniAl()
graf = BoşGrafOluştur()
Her bir cümle için:
    GrafDüğümüEkle(graf, cümle)
Her bir kelime, oluşum_sayısı içinde:
    Eğer KelimeCümledeVar(cümle, kelime) ve
KelimeDurakKelimesiDeğil(kelime):
    GrafKenarıEkle(graf, cümle, kelime, ağırlık=oluşum_sayısı)
graf'ı döndür
Fonksiyon: RougeHesapla(referans, özet)
RougeMetrikleriniHesapla ve döndür
Fonksiyon: Main()
metin = MetniOku()
cümleler = CümleleriAyr(metin)
kelime_oluşumları = KelimeOluşumlarınıSay(metin)
graf = GrafOluştur(cümleler, kelime_oluşumları)
maks_düğüm_listesi = MaxDüğümListesiHesapla(graf)
etkili_cümleler = CümleleriIndexle(cümleler,
min_max_düğüm_listesi)
referans_metin = "Örnek Referans Metin"
skorlar = RougeHesapla(referans_metin, etkili_cümleler)
```



Şekil 1. Önerilen Yöntemin Akış Şeması

4. DENEYSEL SONUÇLAR (EXPERIMENTAL RESULTS)

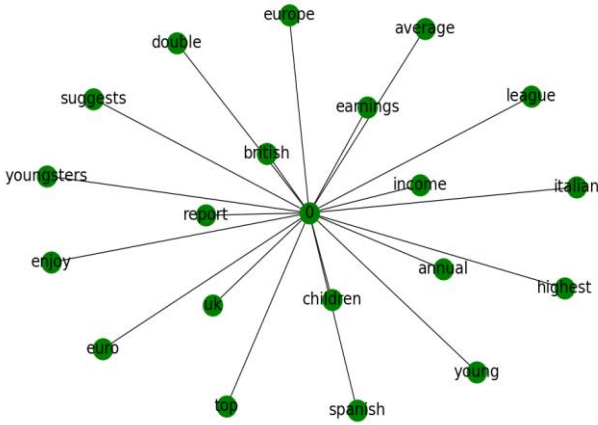
Bu çalışmada, metin özetleme sürecinde Malatya merkezilik algoritmasının etkinliği değerlendirildi. Bu amaçla, toplamda 2224 haber metni, Textrank, Lexrank, Pagerank ve önerilen yöntem aracılığıyla özetlendi. Elde edilen sonuçlar, rouge değerleri üzerinden objektif bir değerlendirmeye tabi tutuldu.

Malatya merkezilik algoritması, bu çalışmanın odak noktasını oluşturan temel algoritma olarak önerilen yöntem kullanıldı. Malatya merkezilik algoritmasının özetleme sürecindeki başarı oranları, metinlerin en etkili noktalarını belirleme konusundaki yeteneğiyle ön plana çıkmaktadır.

Rouge değerleri incelendiğinde, Malatya merkezilik algoritmasının Textrank, Lexrank ve Pagerank gibi diğer yaygın kullanılan algoritmalarla kıyaslandığında genel olarak daha yüksek performans sergilediği belirlendi. Bu durum, Malatya merkezilik algoritmasının haber metinlerini özetleme sürecinde sağladığı avantajları vurgular. Textrank, Lexrank ve Pagerank algoritmaları da önerilen yöntemdeki gibi graf tabanlı algoritmalar olduğu için bu algoritmalar ile karşılaştırıldı. Bu algoritmalar yaygın olarak birçok çalışmada kullanılmış, diğer yöntemler ile karşılaştırılmıştır.

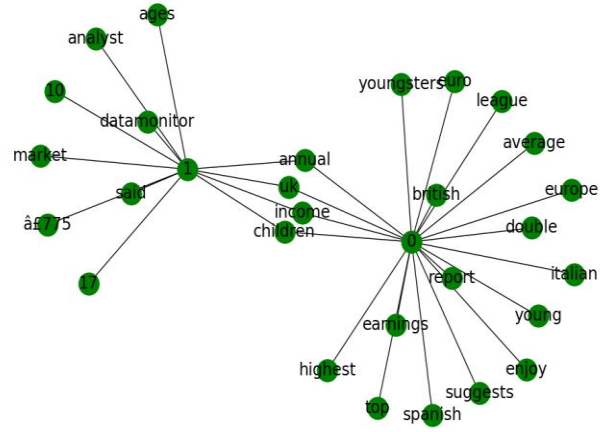
Önerilen yöntem, Malatya merkezilik algoritmasını içerdiği için diğer yöntemlere göre daha başarılı bir özetleme sağlandı. Araştırmanın bulguları, metin özetleme alanında Malatya merkezilik algoritmasının etkili bir çözüm sunabileceği yönünde önemli bir perspektif sunmaktadır.

Bu çalışma, Malatya merkezilik algoritmasının metin özetleme sürecindeki etkinliğini vurgular ve bu algoritmanın gelecekteki çalışmalarda daha fazla kullanılmasını teşvik eder.



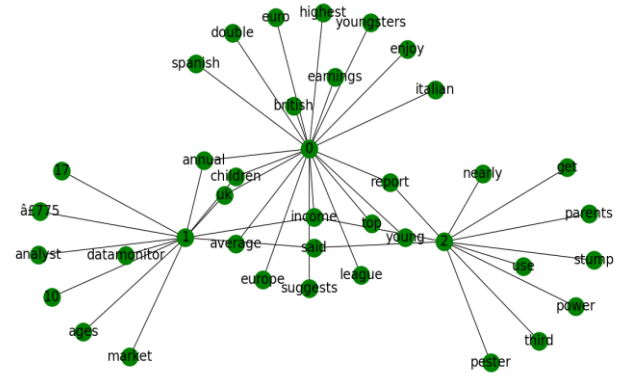
Şekil 2. İlk Cümledeki Graf Yapısı

Şekil-2 İlk cümledeki her kelimenin 1.düğüme eklendiğini gösteren bir örnek.

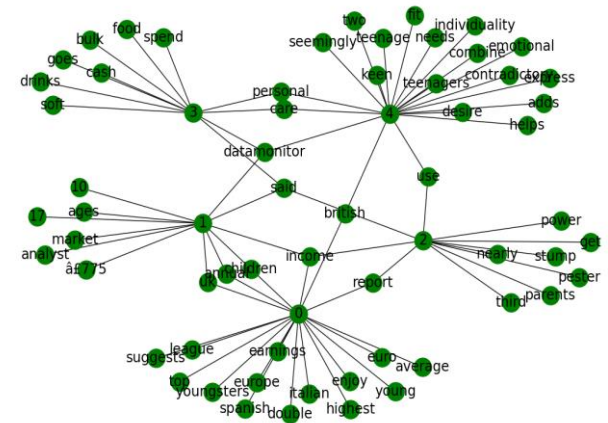


Şekil 3. İkinci cümle eklendikten sonraki graf yapısını gösterir.

Şekil 3'de ikinci cümledeki paylaşımın graf'a eklenmesiyle ilk cümledeki paylaşılan düğümler belirlenerek dahil edilmektedir. Bu süreç cümleler arasındaki ilişkiyi ortaya çıkarır.



Şekil 4. Üçüncü Cümle Eklendikten Sonra Graf Yapısı.



Şekil 5. Beşinci Cümle Eklendikten Sonra Graf Yapısı.

Şekil-5 'te grafikteki 5. cümleye ulaştıkça cümleler arasındaki ilişkilerin daha da arttığı gözlemlenebilir. İlerledikçe daha sık kullanılan kelimeler ortaya çıkacak ve bu da metnin içinden en etkili cümlelerin seçilmesini sağlayacaktır.

Aynı örnek metin, metin özetlemede kullanılan diğer algoritmalarla test edildi. Önerilen yöntemle karşılaştırılan algoritmalar TextRank, LexRank ve PageRank algoritmalarıdır. Bu algoritmalarından elde edilen sonuçların rouge puanları, önerilen yöntemden elde edilen sonuçlarla karşılaştırıldı. Karşılaştırmalar Tablo-1'de Rouge-1 metriği, Tablo-2'de Rouge-2 metriği ve Tablo-3'te Rouge-1 metriği temel alınarak yapıldı.

Tablo 1. ROUGE-1 Metriğine Dayalı Metin Özetleme için Önerilen Yöntemin TextRank, LexRank ve PageRank Algoritmaları ile Karşılaştırılması

Özetleme Yöntemleri	Rouge-1		
	Recall	Precision	F-Score
TextRank	0,3286	0,7946	0,4526
LexRank	0,276	0,6919	0,3846
PageRank	0,3514	0,7606	0,4694
Önerilen Yöntem	0,6088	0,4265	0,4869

Tablo-1'de ROUGE-1 metriklerine dayalı karşılaştırmaya göre TextRank, LexRank ve PageRank algoritmaları tarafından uygulanan özetleme yöntemleri, bir önerilen yöntemle karşılaştırıldı. Rouge-1, özet metin ile referans metin arasındaki tek kelime uzunluğundaki parçacıkların benzerliğini değerlendirir. Bu puanlar kelime düzeyindeki benzerliği ölçer.

Elde edilen sonuçlara göre, her bir özetleme yönteminin Recall (hatırlama) değeri, özetin gerçek metindeki önemli bilgilerin yüzdesini ifade eder. TextRank, LexRank ve PageRank sırasıyla %32,86, %27,6 ve %35,14 Recall değerine sahipken, önerilen yöntem bu değeri %60,88 olarak göstermektedir. Bu, önerilen yöntemin daha fazla önemli bilgiyi özetle koruduğunu ortaya çıkarır.

Precision, özetin içinde bulunan bilgilerin ne kadarının gerçekten önemli olduğunu ölçer. Özetin ne kadarının gerçekten gerekli ve doğru bilgilerle doldurulduğunu gösterir. Yüksek bir Precision değeri, özetin genellikle

gereksiz bilgiler içermediğini ve sadece gerekli olanların bulunduğunu gösterir. Yapılan çalışmada Rouge-1 metriğine göre Precision değerleri; TextRank, LexRank ve PageRank sırasıyla %79,46, %69,19 ve %76,06 Precision değerine sahipken, önerilen yöntem bu değeri %42,65 olarak göstermektedir. Rouge-1 metriğine göre Precision değeri karşılaştırılan diğer algoritmalara göre biraz daha düşük çıktığı için özette gereksiz kelimelerin de var olduğu sonucunu çıkarmaktadır.

F-Score değeri, Recall ve Precision'ın harmonik ortalamasıdır. Bu değer, özetlemenin genel başarısını yansıtır. TextRank, LexRank ve PageRank sırasıyla %45,26, %38,46 ve %46,94 F-Score'a sahipken, önerilen yöntem bu değeri %48,69 olarak göstermektedir. Bulgular, önerilen yöntemin diğerlerine kıyasla daha yüksek bir Recall ancak daha düşük bir Precision ile birlikte daha dengeli bir F-Score sağladığını ortaya koymaktadır.

Tablo 2. ROUGE-2 Metriğine Göre Metin Özetlemede Önerilen Yöntemin TextRank, LexRank ve PageRank Algoritmaları ile Karşılaştırılması

Özetleme Yöntemleri	Rouge-2		
	Recall	Precision	F-Score
TextRank	0,2514	0,6914	0,3566
LexRank	0,1835	0,551	0,2667
PageRank	0,2626	0,6561	0,3639
Önerilen Yöntem	0,469	0,296	0,3513

Tablo-2'de ROUGE-2 metriklerine dayalı bir karşılaştırma gerçekleştirilmiştir. TextRank, LexRank ve PageRank algoritmaları ile uygulanan özetleme yöntemleri, önerilen yöntemle kıyaslandı.

Rouge-2, iki kelime uzunluğundaki parçacıkların benzerliğini ölçer. Rouge-2 metriğine göre Recall (hatırlama) değeri, metnin orijinalinde bulunan önemli iki kelimelik öbeklerin yüzdesini ifade eder. TextRank, LexRank ve PageRank sırasıyla %25,14, %18,35 ve %26,26 Recall değerine sahipken, önerilen yöntem bu değeri %46,9 olarak göstermektedir. Bu sonuç Rouge-2 metriğine göre önemli bilgilerin özet içinde daha fazla bulunduğunu gösterir.

Precision (kesinlik) değeri ise özetin doğruluğunu ölçer. TextRank, LexRank ve PageRank sırasıyla %69,14, %55,1 ve %65,61 Precision değerine sahipken, önerilen yöntem bu değeri %29,6 olarak göstermektedir. Bu sonuç özette gerekli olmayan bilgilerin olabileceğini gösterir.

F-Score değeri, Recall ve Precision'ın harmonik ortalamasıdır ve bu değer, özetlemenin genel başarısını yansıtır. TextRank, LexRank ve PageRank sırasıyla %35,66, %26,67 ve %36,39 F-Score'a sahipken, önerilen yöntem bu değeri %35,13 olarak göstermektedir. Bu sonuca göre de önerilen yöntemin Rouge-2 metriğinin ortalama olarak iyi olduğu sonucu ortaya çıkar.

Tablo 3. ROUGE-L Metriğine Göre Metin Özetlemede Önerilen Yöntemin TextRank, LexRank ve PageRank Algoritmaları ile Karşılaştırılması

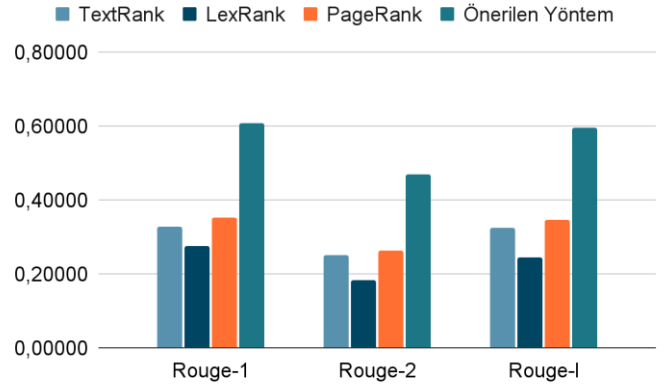
Özetleme Yöntemleri	Rouge-l		
	Recall	Precision	F-Score
TextRank	0,323	0,7811	0,4448
LexRank	0,2445	0,6638	0,3687
PageRank	0,3446	0,7462	0,4605
Önerilen Yöntem	0,5953	0,4141	0,4744

Tablo-3'de ROUGE-L metriklerine dayalı bir karşılaştırma gerçekleştirildi. TextRank, LexRank ve PageRank algoritmaları tarafından uygulanan özetleme yöntemleri, bir önerilen yöntemle karşılaştırılmıştır. Bulgulara göre, her bir özetleme yönteminin ROUGE-L metrikleri altında değerlendirilmiştir. Rouge-1, tüm özet ve referans metinleri arasındaki en uzun paylaşılan alt dizinin benzerliğini ölçen En Uzun Ortak Alt Dizi (LCS) tabanlı benzerlik ölçüsünü kullanır.

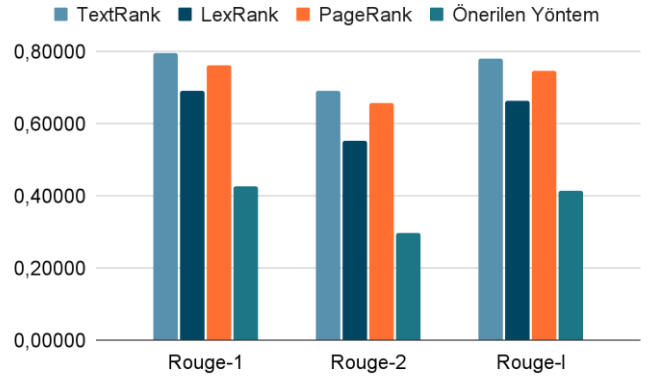
Recall (hatırlama) değeri, TextRank, LexRank ve PageRank sırasıyla %32,3, %24,45 ve %34,46 Recall değerine sahipken, önerilen yöntem bu değeri %59,53 olarak göstermektedir. Bu, önerilen yöntemin daha fazla önemli bilgiyi özetle koruduğunu gösterir.

Precision (kesinlik) değeri, TextRank, LexRank ve PageRank sırasıyla %78,11, %66,38 ve %74,62 Precision değerine sahiptirken, önerilen yöntem bu değeri %41,41 olarak göstermektedir. Rouge-1 metriğine göre Precision değeri karşılaştırılan diğer algoritmalara göre biraz daha düşük çıktığı için özetle gereksiz kelimelerin bulunabileceği sonucu ortaya çıkar.

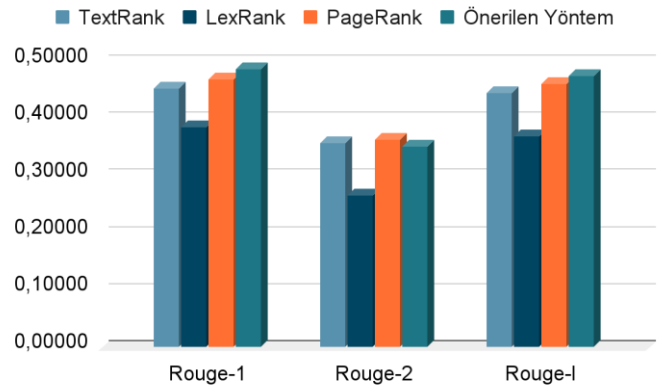
F-Score değeri, TextRank, LexRank ve PageRank sırasıyla %44,48, %36,87 ve %46,05 F-Score'a sahipken, önerilen yöntem bu değeri %47,44 olarak göstermektedir. F-Score değeri hesaplanırken hem Recall hem de Precision kullanıldığı için bu sonuca göre Rouge-1 metriğinin diğer yöntemlere kıyasla daha iyi olduğu sonucuna varılır.



Şekil 6. ROUGE Metriklerine Göre Recall Değerlerinin Karşılaştırılması



Şekil 7. ROUGE Metriklerine Göre Precision Değerlerinin Karşılaştırılması



Şekil 8. ROUGE Metriklerine Göre F-Skor Değerlerinin Karşılaştırılması

5. SONUÇLAR (CONCLUSIONS)

Bu çalışmada, Malatya merkezilik algoritmasının graf tabanlı bir yaklaşım kullanarak İngilizce haber metinleri ile etkili bir şekilde özetleme işlemi yapılabildiği gösterildi. Algoritma, metnin içeriğini temsil eden önemli kelimeleri belirler. Bu kelimeler üzerinden metnin özetini oluşturur. Bu graf tabanlı yaklaşım, metin özetleme alanında önemli bir ilerleme sağlamaktadır.

Malatya merkezilik algoritmasının etkinliğini değerlendirmek amacıyla bu çalışma, geniş bir veri kümesi olan bbcnews veri seti üzerinde gerçekleştirilmiştir. Toplamda 2224 haber metni kullanılarak yapılan testlerle algoritmanın performansı incelendi. Bu veri seti, farklı konulardaki zengin ve çeşitli haber içeriklerini içermektedir. Bu da Malatya merkezilik algoritmasının genel uygulanabilirliğini değerlendirmek için uygun bir ortam sağlar. Algoritma, kelime düğümleri arasındaki bağlantıları ve kelimelerin önemini belirlemek üzere Malatya merkezilik skorlarını kullanır. Bu skorlar, metin içindeki kelimenin belirginliğini gösterir. Yüksek skorlara sahip kelimeler, metnin ana fikrini ve kritik noktalarını yansıtan önemli kelimeler olarak seçilir. Seçilen kelimelerin içerdiği cümleler birleştirilerek özet metin oluşturulur.

Sonuçlar değerlendirildiğinde, Rouge-1 metriğinde Recall ve f-skore değerlerinde iyi olduğu fakat Precision değeri karşılaştırılan diğer algoritmalara göre biraz daha düşük çıktığı için özet de gereksiz kelimelerin de var olduğunu gösteriyor. Algoritma Rouge-1 de genellikle başarılı olsa da bu konuda biraz daha geliştirilmesi gerekir. Rouge-2 metriğine göre Precision ve f-skore değerlerinde diğer algoritmalara kıyasla yeteri kadar başarılı olamadığı görüldü. Bu metrik, özetlenen metinle referans metin arasındaki aynı iki kelimenin yan yana gelme oranının diğer algoritmalara göre biraz daha düşük olduğunu gösterir. Rouge-1 metriğine göre algoritmanın genel olarak başarılı olduğu görüldü. Rouge-1 de en uzun ortak alt-dizi benzerliği diğer algoritmalara göre daha başarılı sonuç verdiği anlaşıldı.

Bu çalışma ile geliştirilmesi öngörülen kısımlar geliştirilerek önerilen yöntemin daha iyi sonuçlar verebileceği ortaya koyuldu. Malatya merkezilik algoritmasının bu çalışmadaki başarıları, metin özetleme tekniklerinin geliştirilmesine önemli bir katkı sunacaktır. Bu yöntem, gelecekteki araştırmalar için bir temel oluşturabilir ve haber metinlerinin daha etkili bir şekilde özetlenmesi için yeni stratejilere yol açabilir.

KAYNAKLAR (REFERENCES)

- [1] Yakut, S., Öztemiz, F., & Karci, A. "A New Approach Based on Centrality Value in Solving the Minimum Vertex Cover Problem: Malatya Centrality Algorithm." *Computer Science*. Volume Vol:7, Issue Issue:2, 81 - 88. (07.12.2022)
- [2] Bakan, C. & Yakut, S. "Development of Text Summarization Method based on Graph Theory and Malatya Centrality Algorithm." *Computer Science*. Volume: IDAP-2023. (18.10.2023)
- [3] Yakut, S., Öztemiz, F., & Karci, A. "A New Approach Based on Centrality Value in Solving the Maximum Independent Set Problem: Malatya Centrality Algorithm." *Computer Science*, Vol:8(Issue:1), 16-23. (2023)
- [4] Tülek, M. "Türkçe için Metin Özetleme." (2007)
- [5] Khushboo S. Thakkar, R.V. Dharaskar, & M.B. Chandak. "Graph-Based Algorithms for Text Summarization." *IEEE*. 10.1109/ICETET.2010.104. (2010)
- [6] Güneş Erkan, & Dragomir R. Radev. "LexRank: Graph-based Lexical Centrality as Saliency in Text Summarization." (2004)
- [7] Ibrahim F. Moawad, & Mostafa Aref. "Semantic Graph Reduction Approach for Abstractive Text Summarization." *IEEE*. 10.1109/ICCES.2012.6408498. (2013)
- [8] Rafael Ferreira, Frederico Freitas, Luciano de Souza Cabral, Rafael Dueire Lins, Rinaldo Lima, Gabriel França, Steven J. Simske, & Luciano Favaro. "A Four Dimension Graph Model for Automatic Text Summarization." *IEEE*. 10.1109/WI-IAT.2013.55. (2013)
- [9] Chirantana Mallick, Ajit Kumar Das, Madhurima Dutta, Asit Kumar Das, & Apurba Sarkar. "Graph-Based Text Summarization Using Modified TextRank." *Springer*. (2018)
- [10] Yogesh Sankarasubramaniam, Krishnan Ramanathan, & Subhankar Ghosh. "Text Summarization using Wikipedia." *ScienceDirect*. (2014)
- [11] Rasim Alguliev, & Ramiz Aliguliyev. "Evolutionary Algorithm for Extractive Text Summarization." *Scientific Research*. (2009)
- [12] Naresh Kumar Nagwani, & Dr. Shrish Verma. "A Frequent Term and Semantic Similarity based Single Document Text Summarization Algorithm." *ResearchGate*. (2011)
- [13] Rada Mihalcea. "Graph-based Ranking Algorithms for Sentence Extraction, Applied to Text Summarization." (2004)
- [14] Sumya Akter, Aysa Siddika Asa, Md. Palash Uddin, Md. Delowar Hossain, Shikhor Kumer Roy, & Masud Ibn Afjal. "An Extractive Text Summarization Technique for Bengali Document(s) Using K-means Clustering Algorithm." *IEEE*. 10.1109/ICIVPR.2017.7890883. (2017)
- [15] S.A. Babar, & Pallavi D. Patil. "Improving Performance of Text Summarization." *ScienceDirect*. (2015)
- [16] Makbule Gulcin Ozsoy, & Ferda Nur Alpaslan. "Text Summarization using Latent Semantic Analysis." 10.1177/0165551511408848. (2011)
- [17] Erhandi, B. "Derin Öğrenme ile Metin Özetleme." (2020)
- [18] Joel Larocca Neto, Alex A. Freitas, & Celso A. A. Kaestner. "Automatic Text Summarization Using a Machine Learning Approach." *Springer*. (2003)
- [19] Carlos N. Silla Jr., Gisele L. Pappa, Alex A. Freitas, & Celso A. A. Kaestner. "Automatic Text Summarization with Genetic Algorithm-Based Attribute Selection." *Springer*. (2004)
- [20] O. Kaynar, Y. E. Işık, Y. Görmez, ve F. Demirkoparan, "Otomatik Metin Özetleme için Genetik Algoritma Tabanlı Cümle Çıkarımı", *Yönetim Bilişim Sistemleri Dergisi*, c. 3, sy. 2, ss. 62–75, 2017.
- [21] Raed Z. Al-Abdallah, & Ahmad T. Al-Taani. "Arabic Single-Document Text Summarization Using Particle Swarm Optimization Algorithm." *ScienceDirect*. (2017)

- [22] Abdullah Ammar Karcioğlu & Ahmet Cahit Yaşa. "Automatic Summary Extraction in Texts Using Genetic Algorithms." IEEE. (2021)
- [23] Enise Karakoç, Burcu Yılmaz. "Deep Learning Based Abstractive Turkish News Summarization." IEEE. (2019)
- [24] Hakan Torun, A. Burak İner. "A Method for Similarity Detection in Vector Space by Summarizing News Articles." IEEE. (2022)
- [25] Yavuz Selim Kartal, Mucahid Kutlu. "Machine Learning Based Text Summarization for Turkish News." IEEE. (2021)
- [26] Hark, C., Taner Uçkan, T., Seyyarer, E., & Karcı, A. "Metin Özetlemesi için Düğüm Merkezliklerine Dayalı Denetimsiz Bir Yaklaşım." *dergipark*, 8(3). (30.09.2019)
- [27] Mengqian Wang, Manhua Wang, Fei Yu, Yue Yang, Jennifer Walker, Javed Mostafa. "A Systematic Review of Automatic Text Summarization for Biomedical Literature and EHRs." *Journal of the American Medical Informatics Association*, Volume 28, Issue 10, October. (2021)
- [28] Kumar, Y., Kaur, K. & Kaur, S. "Study of Automatic Text Summarization Approaches in Different Languages." *Artif Intell Rev* 54, 5897–5929. (2021)
- [29] Taner Uçkan, Cengiz Hark & Ali Karcı. "SSC: Clustering of Turkish Texts by Spectral Graph Partitioning." *Politeknik Dergisi*. (2021)
- [30] Aydın, A., & Uçkan, T. "Çizgeler Üzerinde Farklı Ağırlıklandırma Yöntemleri Ve Merkezilik Ölçütleri İle Çıkarımsal Metin Özetleme." *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 35(1), 71-79. (2023)
- [31] Kuş, A., & Acı, Ç. İ. "An Extractive Text Summarization Model for Generating Extended Abstracts of Medical Papers in Turkish." *Bilgisayar Bilimleri Ve Teknolojileri Dergisi*, 4(1), 19-26. (2023)