Bilgisayar Bilimleri ve Mühendisliği Dergisi

Cilt-19
Sayı-1
2026

Türkçe ve İngilizce Anahtar Sözcük Üretimi için Özgün Veri Küme Oluşturulması ve Büyük Dil Modellerine İnce Ayar Uygulanması

Construction of a Novel Dataset for Turkish and English Keyword Generation and Fine-Tuning of Large Language Models

Şeyda Nur ARSLAN
Sivas Cumhuriyet Üniversitesi
Fen Bilimleri Enstitüsü, Yapay Zekâ ve Veri Bilimi
Sivas/Türkiye
20239258022@cumhuriyet.edu.tr
ORCID: 0009-0005-1343-0258

Yasin GÖRMEZ
Sivas Cumhuriyet Üniversitesi
İktisadi ve İdari Bilimler Fakültesi, Yönetim Bilişim Sistemleri
Sivas/Türkiye
ysngrmz@hotmail.com
ORCID:0000-0001-8276-2030

Halil ARSLAN
Sivas Cumhuriyet Üniversitesi
Mühendislik Fakültesi, Bilgisayar Mühendisliği
Sivas, Türkiye
harslan@cumhuriyet.edu.tr
ORCID: 0000-0003-3286-5159

Öz

Bu çalışmada, Türkçe ve İngilizce metinler için özgün anahtar sözcük üretiminde kullanılmak üzere TR Dizin veri tabanından geniş kapsamlı iki veri kümesi oluşturulmuş ve bu veri kümeleri T5 ile FlanT5 büyük dil modellerine ince ayar yapılarak değerlendirilmiştir. Hiper parametrelerin belirlenmesinde Bayesian optimizasyon yaklaşımı kullanılmış ve bu süreç modellerin öğrenme verimliliğini artırmıştır. Deneysel analizler, özellikle Türkçe modellerin hem n gram tabanlı ölçütlerde hem de anlamsal benzerlik değerlendirmelerinde yüksek başarı sağladığını göstermektedir. Elde edilen sonuçlar, büyük dil modellerinin otomatik anahtar sözcük üretimi için etkili ve ölçeklenebilir bir çözüm sunduğunu ortaya koymaktadır.

Anahtar sözcükler: Anahtar sözcük çıkarımı, Büyük dil modelleri, Derin Öğrenme, Türkçe veri küme, Doğal dil işleme.

Makale Bilgileri	**Article Info**
Türü: Araştırma	Type: Research
Geliş tarihi: 08.12.2025	Received date: 08.12.2025
Düzeltme tarihi: 29.12.2025	Revision date: 29.12.2025
Kabul tarihi: 14.01.2026	Accepted date: 14.01.2026

Atıf/ to Cite (IEEE): Ş. N. ARSLAN, Y. GÖRMEZ, H. ARSLAN, Türkçe ve İngilizce Anahtar Sözcük Üretimi için Özgün Veri Küme Oluşturulması ve Büyük Dil Modellerine İnce Ayar Uygulanması: Bilgisayar Bilimleri ve Mühendisliği Dergisi, Cilt 19 2026 Sayı-19-1. Sf: 107-116
DOI: 10.54525/bbmd.1838196

Abstract

This study presents two comprehensive datasets constructed from the TR Dizin database for use in generating original keywords in Turkish and English texts. These datasets were employed to fine-tune the T5 and FlanT5 large language models, which were subsequently evaluated for their effectiveness. Bayesian optimization was utilized to determine the optimal hyperparameters, enhancing the learning efficiency of the models. Experimental analyses revealed that the Turkish models achieved high performance in both n-gram–based metrics and semantic similarity evaluations. The findings indicate that large language models offer an effective and scalable solution for automatic keyword generation.

Keywords: Keyword extraction, Large language models, Deep learning, Turkish dataset, Natural language processing.

1. Giriş

Sayısallaşmanın ivme kazanmasıyla birlikte, veri toplama, depolama ve erişim süreçleri kayda değer ölçüde kolaylaşmıştır. Ancak, yapılan araştırmalar veri hacmindeki artışla kaliteli bilgi oranının azaldığını göstermektedir [1], [2]. Bu bağlamda, sayısal platformlardaki veri miktarının artmasına rağmen, bireylerin doğru ve güvenilir bilgiye ulaşımı giderek daha karmaşık bir hal almaktadır. Akademik makalelerde, arama motorlarında, veri tabanlarında, sayısal

kitaplıklarda, sosyal medya etiketlerinde ve arama motoru optimizasyonunda sıklıkla kullanılan anahtar sözcükler, bir metnin içeriğini en iyi şekilde özetleyen ve bilgiye erişimi kolaylaştıran temel sözcükler olarak tanımlanabilmektedir. Belgeleri temsil eden anahtar sözcüklerin doğru seçilmesi araştırma süreçlerinde doğru bilgiye ulaşmayı kolaylaştırmaktadır. Doğru seçilmiş anahtar sözcüklerin hem kullanıcıların hem de sistemlerin belgeyi daha iyi tanımasını sağladığı görülmüştür [3]. Yapılan bir çalışmada ise anahtar sözcüklerle yapılan dinamik filtreleme ve görüntüleme teknikleri sayesinde kullanıcıların yarı yarıya daha az ilgisiz dokümanla karşılaştığı gösterilmiştir [4]. Tüm bu durumlar dikkate alındığında dokümanlar için doğru anahtar sözcüklerin seçilmesinin büyük önem arz ettiği değerlendirmesi yapılabilmektedir.

Bir belge için hem manuel yollarla hem de otomatik olarak anahtar sözcük belirlemek mümkündür. Manuel yöntem, konu uzmanı ya da yazar tarafından metne özgü özgün anahtar sözcükler seçilerek gerçekleştirilmektedir. Bu yaklaşım yüksek öznellik, zaman maliyeti ve bir kişiye bağlı tutarlılık getirdiği için son zamanlarda otomatik anahtar sözcük belirlemenin önemi giderek artmaktadır. Otomatik yöntemle belirlenen anahtar sözcüklerin tutarlılık açısından bir uzman tarafından belirlenen anahtar sözcükler ile uyuştuğu görülmüştür. Bu durumun yanı sıra otomatik yollarla anahtar sözcük belirlemenin özellikle büyük belge koleksiyonlarında ölçeklenebilirlik açısından avantaja sahip olduğu ve anahtar sözcük belirlemeyi çok daha hızlı yapabildiği değerlendirmeleri yapılmaktadır [5]. Otomatik yollarla yapılan anahtar sözcük tahmininde ise metinde geçen en önemli sözcüklerin tahmin edildiği çıkarımsal ve metinde açıkça yer almayan ama içeriği özetleyen yeni kavramlar üreten yorumlayıcı olmak üzere iki farklı yaklaşım kullanılmaktadır. Yapay zeka alanındaki çalışmalar, büyük dil modelleri kullanıldığında yorumlayıcı yöntemlerin hem özgünlük hem de anlam yakalama açısından daha üstün başarım sağlayabileceğini göstermektedir [6].

Otomatik anahtar sözcük üretmenin zorluğu dilden dile de değişiklik göstermektedir. İngilizce gibi yüksek kaynağa sahip ve terminolojik standardizasyonu yüksek olan dillerde anahtar sözcük tahmini yapmak daha kolay olmaktadır. Bu durumun aksine Türkçe gibi morfolojik açından zengin ve daha az kaynağa sahip dillerde ise anahtar sözcük tahmini yapmak daha zor bir hale gelmektedir. Makine öğrenmesi modellerinin başarımını önemli ölçüde etkileyen veri sayısı anahtar sözcük çıkarımı da oldukça etkilemektedir. Yapılan bir çalışmada anahtar tahmini için geliştirilmiş olan denetimli yöntemlerin az veri nedeniyle düşük başarım gösterdiği değerlendirilmiştir [7]. Veri yetersizliğinin yanı sıra karmaşık lokmalama (lokmaizasyon) süreçleri ve metin dışı anahtar sözcüklerin yüksek oranı gibi sebeplerden ötürü Türkçe gibi dillerde geliştirilen anahtar sözcük tahmin modellerinin doğruluğu düşük olmaktadır.

Bu çalışma kapsamında, anahtar sözcük üretimi için gerekli olan özgün İngilizce ve Türkçe veri kümeleri, Türkiye ulusal atıf dizini olan TR Dizin veri tabanındaki makalelerden yararlanılarak oluşturulmuştur. Oluşturulan bu veri kümeleri kullanılarak T5 ve FlanT5 modellerine hiper-parametre optimizasyonu (Bayesian yaklaşımı kullanılarak) uygulanmış ve ardından ince ayar yapılmıştır. Özellikle kaynaklarda daha az kaynağa sahip olması ve eklemeli yapısının dil işleme zorluklarını artırması nedeniyle Türkçe için büyük dil modellerine ince ayar uygulanması, çalışmanın temel yenilikçi yönünü oluşturmaktadır. Ayrıca hem Türkçe hem de İngilizce için yeni bir özgün veri kümenin oluşturulması da çalışmanın literatüre olan katkısını pekiştirmektedir. Takip eden bölümlerde, öncelikle ilgili literatür taraması sunulmuş, ardından çalışmada kullanılan materyal ve metodoloji detaylıca açıklanmıştır. Son olarak, elde edilen deney sonuçları paylaşılmış ve makale, sonuçların tartışıldığı bir bölümle tamamlanmıştır.

2. Kaynak Taraması

Anahtar sözcük tahmini için kaynaklarda birçok çalışma yapılmıştır. Aday anahtar sözcüklerin içerik ve anlam tabanlı son işleme tutulduğu yöntemin önerildiği çalışmada 17 farklı veri kümendeki F1 değeri 5 farklı yöntemde ortalama %25,8 iyileştirilmiştir [8]. Anahtar sözcük tahmini için KeyBERT yaklaşımı; Text Rank, Rake, Gensim, Yake, ve TF-IDF yöntemleri ile karşılaştırılmıştır. JUCS veri kümende KeyBERT yönteminin diğer yöntemlerden daha iyi olduğu kanaatine varılmıştır. Aynı çalışmada KeyBERT yöntemi kullanılarak geliştirilen model tarafından üretilen anahtar sözcükler, yazarlar tarafından oluşturulan anahtar sözcükler ile %51 benzerlik göstermiştir [9]. Ön eğitilmiş BERT modelinin örnek, prosedür, patolojik tanı anahtar sözcüklerini çıkarmak için kullanıldığı çalışmada 6093 rapor ile ince ayar yapılmış ve model başarımları 678 rapor üzerinde hesaplanmıştır. Yapılan hesaplamalar sonucunda örnek, prosedür ve patolojik tanı için anahtar sözcük üretiminde sırasıyla %99.51, %99.85 ve %99.61 hassasiyet değeri edilmiş ayrıca BERT modelinin uzun kısa vadeli bellek, evrişimsel sinir ağı, Bayes, Kea ve WINGNUS yöntemlerine göre daha iyi olduğu sonucuna ulaşılmıştır [10]. Anahtar sözcük tahmin başarısını artırmak için hedef merkez tabanlı ve dikkat mekanizması ile güçlendirilmiş uzun kısa vadeli bellek modelinin önerildiği çalışmada %80.35 hassasiyet, %88.11 duyarlılık ve %84.05 F1 değeri elde edilmiştir [11]. Karmaşık bir ağ olarak modellenmiş ve düğüm özelliklerinden elde edilen özniteliklerle sınıflandırıcının önerilmiş olduğu çalışmada farklı dil ve alanlarda kaynaklardan daha iyi sonuç elde edildiği raporlanmıştır [12]. Yapılan bir çalışmada farklı boyut, alan ve okunabilirlikteki belge grupları için lojistik regresyon ve LASSO düzenlileştirme temelli, etiket varlığına duyarlı, genel kullanıma uygun bir anahtar sözcük çıkarma modeli önerilmektedir. Model, terim frekanslarına bağımlılığı azaltarak ayırt edici ve temsil edici anahtar sözcükler çıkarmada klasik yöntemleri geride bırakmakta ve çeşitli belge gruplarında sağlam başarım göstermektedir [13]. Belgelerden anahtar sözcükleri çıkarmak için hem küresel hem de yerel bağlam bilgilerini kullanan yeni bir yöntemin önerildiği çalışmada destek vektör makinaları tabanlı yöntem, deneysel sonuçlara göre temel yöntemlerden önemli ölçüde daha iyi başarım göstererek belge sınıflandırma doğruluğunu artırmaktadır [14]. Tekil belgelerden anahtar

sözcük çıkarmak için TextRank, RAKE ve TAKE yöntemlerini birleştiren bir topluluk yönteminin tasarlanarak analiz edildiği çalışmada filtreleme sezgiseli ve dinamik eşik fonksiyonları kullanılarak oluşturulan yöntem, kullanılan veri kümende bireysel anahtar sözcük çıkarıcılardan daha iyi başarım göstermiştir [15]. Çok dilli belgelerden anahtar sözcükleri çıkarmak için makine öğrenimi yöntemlerine dayanmayan ve tek bir belgeyle etkili anahtar sözcükler üreten yeni bir algoritma öneren çalışmada Japonca ve İngilizce olmak üzere iki dilli bir doküman ve Reuter's dokümanının bir kısmında sınama edilen algoritmanın belgeleri benzersiz bir şekilde tanımlayan anahtar sözcükler çıkardığını gösterilmiştir [16]. Biyomedikal metinlerde hastalık, semptom, ilaç gibi varlık isimlerini tanımlayan biyomedikal adlandırılmış varlık tanıma yöntemlerinin incelendiği çalışmada bu yöntemler sözlük tabanlı, kural tabanlı, makine öğrenimi ve derin öğrenme olarak dört kategoriye ayrılmıştır. Aynı çalışmada MedMention veri küme üzerinde yapılan deneysel çalışma, BioBERT yönteminin hastalık ve semptom çıkarımında 0,72 F1 değeri ile en başarılı sonuçları verdiğini gösterilmiştir [17]. Biyomedikal makalelerde anahtar sözcük çıkarımını bir dizi etiketleme görevi olarak ele alan çalışmada sözcükleri derin bağlamsal gömülü vektörlerle temsil eden ve XLNET ile BERT tabanlı modelleri ince ayar yaparak anahtar sözcük etiketlerini tahmin eden bir yöntem önerilmiştir. Biyomedikal anahtar sözcük çıkarımı için kıyaslama veri kümende yapılan başarım değerlendirmesi, alan spesifik bağlamsal gömülü vektörlerin (BioBERT, SciBERT) genel alan gömülü vektörlerine (BERT, RoBERTa, XLNET) ve denetimsiz yöntemlere kıyasla en iyi sonuçları elde ettiğini gösterilmiştir [18].

Son yıllarda büyük dil modelleri farklı alanlarda kullanılmaya başlanmış ve bu alanda farklı akademik çalışmalar yapılmıştır. Büyük dil modelinin tıbbi alanda kullanım sorunlarına değinilen çalışmada Med-PaLM modeline ince ayar yapılarak üretilen Med-PaLM2 ile %86,5 skor elde edilmiştir. Bu başarı oranı ile Med-PaLM2 modeli, Med-PaLM modeline göre %19 daha iyi sonuç elde etmiş ve gerçek dünya tıbbi sorularında yapılan pilot çalışmada %65 oranında Med-PaLM2 tarafından verilen cevaplar hekim cevaplarına tercih edilmiştir [19]. Biyomedikal doğal dil işleme görevlerinde dört büyük dil modelinin sıfır öğrenme, az öğrenme ve ince ayar olmak üzere 12 kıyaslama üzerinde başarımlarının karşılaştırıldığı çalışmada geleneksel ince ayarın çoğu görevde daha iyi başarım gösterdiği kanaatine varılmıştır. Aynı çalışmada GPT-4 gibi kapalı kaynak modellerin tıbbi soru cevaplama gibi akıl yürütme görevlerinde üstünlük sağladığı ancak açık kaynak büyük dil modellerinin başarım açıklarını kapatmak için ince ayara ihtiyaç duyduğu çıkarımı yapılmıştır [20]. Biyomedikal adlandırılmış varlık tanıma görevini bir dizi etiketleme görevinden üretim görevine dönüştüren bir talimat tabanlı öğrenme paradigması sunan çalışmada LLaMA-7B tabanlı BioNER-LLaMA modeli geliştirilmiştir. BioNER-LLaMA hastalık, kimyasal ve gen varlıklarını içeren üç biyomedikal adlandırılmış varlık tanıma veri kümende GPT-4'ün az öğrenme yeteneklerine kıyasla %5-30 daha yüksek F1 Değerleri elde etmiştir [21]. Büyük dil modellerinin bilimsel literatürü entegre ederek nörobilim deney sonuçlarını insan uzmanlardan daha iyi tahmin edebildiğini göstermek için BrainBench adı verilen bir ileri görüşlü kıyaslama sunan çalışmada nörobilim literatürüne ince ayar yapılmış BrainGPT modelinin daha yüksek başarım sergilediği ortaya koyulmuştur. Aynı çalışmada büyük dil modellerinin yüksek güvenle yaptığı tahminlerin daha doğru olduğu ve bu modellerin keşif süreçlerinde insanlara yardımcı olabileceği belirtilirken, yaklaşımın nörobilimle sınırlı olmadığı ve diğer bilgi yoğun alanlara aktarılabileceği vurgulanmıştır [22].GPT-4, GPT-3.5, PaLM ve LLaMA2 modellerinin otomatik kompozisyon değerlendirmede kullanıldığı çalışmada insan değerlendiricilerle karşılıklı olarak değerlendirmeler yapılmış ve GPT-4 modelinin en iyi başarımı gösterdiği gözlemlenmiştir [23]. Nefret söylemi tespiti için GPT-2 modelinin kullanıldığı çalışmada büyük dil modellerinin sınıf dengesizliği, çok dilli içerik ve kavramsal kaymalar gibi sorunlarla rahatlıkla başa çıkabileceğine değinilmiştir [24].

Doğal dil işleme literatüründe anahtar sözcük üretimi, metin sınıflandırma ve metin özetleme görevleri ile yakından ilişkili olmakla birlikte, amaç ve çıktı uzayı bakımından bu görevlerden ayrışmaktadır. Metin sınıflandırma, belgeleri önceden tanımlanmış, sabit ve sınırlı sayıda kategori etiketine (örneğin; spor, ekonomi, sağlık) atamayı hedeflerken [25], [26]; anahtar sözcük üretimi, metni temsil edecek etiketleri sınırsız bir sözcük dağarcığından seçmekte veya üretmektedir. Öte yandan metin özetleme, kaynak metni daha kısa ancak dilbilgisel ve anlamsal akışa sahip bir anlatıya dönüştürmeyi amaçlar [27], [28].Anahtar sözcük üretimi ise metin özetlemedeki gibi bilgiyi sıkıştırma prensibine dayanmakla birlikte, sonuç olarak dilbilgisel bir cümle yapısı yerine, belgenin özünü temsil eden ayrık ve yoğun anlamlı terimler dizisi sunmaktadır. Bu çalışma, büyük dil modellerinin özetleme yeteneklerinden faydalanarak, metin sınıflandırmanın kesinliği ile özetlemenin anlamsal derinliğini birleştiren üretken bir anahtar sözcük belirleme yaklaşımı sunmasıyla kaynaklardaki diğer görevlerden farklılaşmaktadır.

Literatür taraması incelendiğinde, anahtar sözcük çıkarımında makine öğrenmesi temelli yaklaşımların yaygın olarak kullanıldığı ve özellikle derin öğrenme tabanlı yöntemlerin, insanlar tarafından belirlenen anahtar sözcüklere oldukça benzer sonuçlar üretebildiği görülmüştür. Farklı dillerde anahtar sözcük çıkarımına yönelik çok sayıda çalışma bulunmasına rağmen, Türkçe diline odaklanan araştırmaların sınırlı olduğu dikkat çekmektedir. Ayrıca, büyük dil modellerinin çeşitli alanlardaki farklı problemleri çözmede etkili olduğu bilinmekle birlikte, bu modellerle gerçekleştirilen anahtar sözcük tahmin çalışmalarının oldukça kısıtlı olduğu sonucuna ulaşılmıştır.

3. Yöntem

Çalışmada, anahtar sözcük üretimi amacıyla Türkçe ve İngilizce dillerine yönelik iki ayrı özgün veri küme oluşturulmuş ve bu veri kümeleri kullanılarak büyük dil modellerine ince ayar uygulanmıştır. İnce ayar sürecine geçilmeden önce, modeller için en uygun hiper-

parametrelerin belirlenmesinde Bayesian optimizasyon yaklaşımından yararlanılmıştır.

3.1 Veri Kümesi Oluşturma

Çalışma kapsamında, Türkçe ve İngilizce anahtar sözcük üretim modellerinde kullanılmak üzere özgün bir veri küme oluşturulmuştur. Veri küme hazırlanırken makalelerin özetleri ile özetlerin sonunda yer alan anahtar sözcükler kullanılmıştır. Bu doğrultuda, TRDizin veri tabanındaki makalelerin Türkçe ve İngilizce özetleri ile anahtar sözcüklerine dayalı iki ayrı veri küme geliştirilmiştir. Özet ve anahtar sözcük bilgilerine erişim için TRDizin tarafından sağlanan resmi uygulama programlama arayüzü (API) kullanılmıştır [29].

TRDizin-API aracılığıyla bir arama terimi verildiğinde, ilgili makalelerin listelenmesi ve izin verilen veri alanlarına erişim mümkün olmaktadır. Bu kapsamda ilk aşamada Fen, Mühendislik, Sağlık, Edebiyat, Tarih gibi farklı disiplinlerden 777 arama terimi manuel olarak belirlenmiştir. Ancak bu 777 terim aracılığıyla hedeflenen sayıda makaleye ulaşılamadığı için, ilk sorgulamalar sonucunda elde edilen makalelerin anahtar sözcükleri de arama terimlerine eklenmiştir. Böylece toplam 23.255 terimden oluşan yeni bir arama listesi oluşturulmuştur. Bu terimlerin bazıları tek sözcükten, bazıları ise birden fazla sözcükten oluşmaktadır.

Anahtar sözcük tahmin modellerinin genelleme yeteneğinin artırılması ve karşılaştırmaların adil bir şekilde yapılabilmesi için veri kümenin farklı alanlardan metinler içermesi özellikle gözetilmiştir. Bu doğrultuda arama terimlerinin disiplinler arası çeşitliliğine özen gösterilmiştir. Oluşturulan 23.000 terim içeren liste temel alınarak oluşturulan sözcük bulutu Şekil 1'de sunulmaktadır.

Şekil-1: Veri küme oluşturma aşamasında kullanılan arama terimlerinden üretilmiş sözcük bulutu

Şekilde sunulan sözcük bulutunda görüldüğü üzere, veri kümenin genelleştirilmesini sağlamak amacıyla farklı disiplin alanlarına ait terimler kullanılmıştır. Belirlenen arama terimlerinin her biri ile TRDizin veri tabanında ayrı ayrı tarama yapılarak ilgili makaleler listelenmiştir. Listelenen makalelerin başlıkları, özetleri ve anahtar sözcükleri hem İngilizce hem de Türkçe olarak kayıt altına alınmıştır. Tekrarları önlemek adına, TRDizin tarafından her makaleye tahsis edilen benzersiz kimlik numarası (ID) bir listeye eklenmiştir. Bu sayede, önceki aramalarda tespit edilen bir makalenin veri kümene tekrar dahil edilmesi engellenmiştir. Veri toplama aşamasının sonunda oluşturulan veri kümelerine ilişkin detaylı bilgiler Çizelge 1'de sunulmaktadır.

Çizelge-1: Oluşturulan Veri Kümelerine Ait İçerik Bilgisi

Veri Küme	Toplam Özet Sayısı	Makale Başına Ortalama Sözcük Sayısı	Toplam Anahtar Sözcük Sayısı	Makale Başına Ortalama Anahtar Sözcük Sayısı
Türkçe	102.625	180,14	510.709	4,97
İngilizce	95.324	221,10	470.305	4,93

Derin öğrenme modellerinin, özellikle büyük dil modellerinin eğitimi veya bu modellere ince ayar yapılması için veri kümelerinde yer alan örnek sayısının yüksek olması büyük önem taşımaktadır. Çizelge 1'de görüldüğü üzere hem Türkçe hem de İngilizce veri kümelerinde 95.000'in üzerinde özet metnin bulunması, derin öğrenme modellerinin ince ayar süreçlerine anlamlı bir katkı sunacaktır. Türkçe ve İngilizce özet sayılarındaki farklılığın temel nedeni, bazı dergilerde İngilizce özet bulundurma zorunluluğunun olmamasıdır. Makale başına düşen ortalama anahtar sözcük sayıları incelendiğinde, TRDizin veri tabanında yer alan çalışmalarda genel olarak beş civarında anahtar sözcük talep edildiği sonucuna ulaşılmaktadır. Ayrıca, çizelgede görülen bir diğer dikkat çekici bulgu, İngilizce özetlerin ortalama sözcük sayısının Türkçe özetlerden daha yüksek olmasıdır. Bu durumun, Türkçenin eklemeli bir dil yapısına sahip olmasından kaynaklandığı değerlendirilmektedir. Türkçede eklerin ayrı bir sözcük olarak sayılmaması, ancak İngilizcede eklerin bağımsız sözcük olarak değerlendirilmesi söz konusu farkın temel nedenini oluşturmaktadır. Oluşturulan veri kümene HuggingFace platformundan ilgili link kullanılarak erişmek mümkündür [30].

3.2 Büyük Dil Modelleri

Çalışmada anahtar sözcük üretimi için büyük dil modellerinden faydalanılmış ve büyük dil modellerine ince ayar yapılmıştır. Büyük dil modelleri, çok büyük ölçekli metin veri kümeleri üzerinde eğitilmiş, milyarlarca (hatta trilyonlarca) parametre içeren derin öğrenme tabanlı dil modelleridir. Bu modeller, dil anlama, dil üretimi, özetleme, çeviri, soru-cevap, anahtar sözcük tahmini gibi çok çeşitli doğal dil işleme görevlerinde yüksek başarım göstermektedir. Dönüştürücü mimarisine dayanarak bağlamı dikkate alır ve metinler arası ilişkileri modelleyebilir. Modellerin ölçeği büyüdükçe genelleme gücü, transfer öğrenme potansiyeli ve uyarlanabilirlikleri artmaktadır. Çalışmamızda anahtar sözcük üretimi için T5 ve FlanT5 büyük dil modellerine ince ayar işlemi yapılmıştır.

Raffel ve ark. tarafından geliştirilen T5 (Text-to-Text Transfer Dönüştürücü) modeli, tüm dil görevlerini "girdi metninin çıktı metnine dönüştürülmesi" yaklaşımıyla birleştirmekte ve büyük ölçekli C4 veri küme üzerinde eğitilmesi sayesinde aktarım öğrenmesi için güçlü bir temel sunmaktadır [31]. T5 modelinin bir uzantısı olan Flan-T5 ise Chung ve çalışma arkadaşları tarafından talimat-temelli ince ayar yaklaşımıyla yeniden yapılandırılmıştır [32]. Bu model, farklı görev türlerini içeren geniş veri kümeleriyle eğitilerek özellikle sıfır atışlı öğrenme (zero-shot) ve az atışlı öğrenme (few-shot)

senaryolarda daha yüksek genelleme başarımı göstermektedir. Dolayısıyla T5 ve Flan-T5 modelleri hem ince ayara uygunlukları hem de dilsel görevlerdeki esneklikleri nedeniyle büyük dil modeli tabanlı anahtar sözcük üretimi çalışmalarında öne çıkan iki önemli mimaridir.

3.3 Bayesian Optimizasyon

Bayesian optimizasyon, özellikle değerlendirilmesi maliyetli ve kapalı formda tanımlanmayan hedef fonksiyonlar için kullanılan olasılık temelli bir hiper-parametre optimizasyon yöntemidir. Bu yaklaşımda hedef fonksiyon doğrudan modellenmek yerine, Gauss süreçleri veya ağaç tabanlı ardıl modeller gibi sezgisel tahminleyiciler kullanılarak olası parametre uzayı üzerinde bir ön ihtimal dağılımı oluşturulur. Ardından, bu dağılıma dayanarak belirlenen edinim fonksiyonu aracılığıyla yeni örnekleme noktaları seçilir ve böylece arama süreci rastgele arama ya da grid arama gibi kör yöntemlere kıyasla çok daha az deneme ile daha iyi sonuçlara ulaşır. Snoek, Larochelle ve Adams, derin öğrenme hiper-parametre ayarı için Bayesian optimizasyonun klasik yöntemlere göre daha hızlı yakınsadığını ve daha düşük maliyetle daha iyi sonuçlar verdiğini göstermiştir [33]. Benzer şekilde Bergstra, Bardenet, Bengio ve Kégl, grid arama ve rastgele aramanın yüksek boyutlu parametre uzaylarında verimsiz kaldığını, Bayesian yaklaşımların ise önceki gözlemlerden bilgi edinerek arama sürecini yönlendirdiğini ortaya koymuştur [34]. Bu nedenlerle Bayesian optimizasyon, özellikle derin öğrenme ve büyük dil modeli ince ayarı gibi hesaplama açısından yoğun alanlarda hem örnek verimliliği hem de yakınsama hızının yüksekliği bakımından üstün bir seçenek olarak kabul edilmektedir.

4. Deney Sonuçları

Çalışmanın deneysel aşamasında, veri kümeleri öncelikle eğitim, sınama ve doğrulama olmak üzere üç ana kümeye ayrılmıştır. Bu kapsamda, toplam veri kümenden rastgele örnekleme yoluyla sırasıyla %20'lik ve %10'luk dilimler seçilerek sınama ve doğrulama veri kümeleri oluşturulmuştur. Geriye kalan örnekler ise modelin eğitimi için kullanılmak üzere eğitim veri kümeni teşkil etmiştir. İngilizce ve Türkçe veri kümeleri için eğitim, sınama ve doğrulama kümelerindeki kesin örnek sayıları Çizelge 2'de detaylı olarak sunulmaktadır.

Çizelge-2: Eğitim, Sınama ve Doğrulama Kümleri Örnek Sayıları

Veri Küme	Alt Veri Küme	Örnek Sayısı
Türkçe	Eğitim	71.837
	Sınama	20.525
	Doğrulama	10.263
İngilizce	Eğitim	66.726
	Sınama	19.065
	Doğrulama	9.533

Veri küme bölütleme aşamasının tamamlanmasının ardından, çalışmada kullanılan T5 ve FlanT5 modelleri, Hugging Face platformu üzerinden sağlanan dönüştürücü kitaplığı aracılığıyla geliştirilmiştir [35], [36]. Modelin girdi ve çıktı sınırlamaları, maksimum girdi uzunluğu 512 lokma ve maksimum etiket (hedef) uzunluğu 64 lokma olarak belirlenmiştir. Ayrıca, girdi metinlerinin uzunluk sınırını aşmaması amacıyla kesme parametresi True olarak ayarlanmıştır.

Geliştirilen büyük dil modellerinin anahtar sözcük üretimi görevinde uzmanlaşmasını sağlamak amacıyla, istem özelleştirmesi yapılmış ve girdiler sisteme bu yapıya uygun olarak iletilmiştir. Bu özelleştirme kapsamında, model girdisi, öncelikle yönlendirici bir talimat olan "anahtar sözcük üret:" terimiyle başlatılmış, hemen ardından ilgili makale özeti eklenmiştir. Modelin öğrenme hedefi (etiket metni) olarak ise, istenen anahtar sözcükler virgüllerle ayrılmış şekilde hazırlanarak hedef metin veri küme oluşturulmuştur. Son olarak, oluşturulan bu metin veri kümeleri, modele girdi olarak verilebilmesi için sayısallaştırma işlemine tabi tutulmuştur.

Model geliştirme sürecini takiben, ince ayar işleminin başarımını maksimize etmek üzere hiper-parametre optimizasyonu gerçekleştirilmiştir. Bu amaçla, öğrenim oranı (learning_rates), epok sayısı (num_train_epochs), ağırlık azaltma (weight_decays) ve ısınma oranı (warmup_ratios) hiper-parametreleri seçilmiş ve eğitim ile Doğrulama kümeleri kullanılarak Bayesian optimizasyon yöntemi ile optimize edilmiştir. Her bir model için optimize edilen hiper-parametrelerin detayları Çizelge-3'te sunulmaktadır.

Çizelge-3: İnce Ayar Aşamasında Kullanılan Modellere Ait Hiper-Parametre Optimizasyon Ayrıntıları

Hiper-Parametre Adı	Hiper-Parametre Türü	Hiper Parametre Uzayı	Model Adı	Optimum Değer
Öğrenim oranı	Reel Sayı	$[10^{-7}, 10^{-2}]$	T5 Türkçe	0.00156
			T5 İngilizce	0.00487
			FlanT5 Türkçe	0.00518
			FlanT5 İngilizce	0.00194
Epok sayısı	Tam Sayı	[20, 200]	T5 Türkçe	48
			T5 İngilizce	95
			FlanT5 Türkçe	139
			FlanT5 İngilizce	115
Ağırlık azaltma	Reel Sayı	$[10^{-2}, 0]$	T5 Türkçe	0.00058
			T5 İngilizce	0.00726
			FlanT5 Türkçe	0.00022
			FlanT5 İngilizce	0.00215
Isınma Oranı	Reel Sayı	$[2 \times 10^{-1}, 0]$	T5 Türkçe	0.17323
			T5 İngilizce	0.00874
			FlanT5 Türkçe	0.05501
			FlanT5 İngilizce	0.13942

Hiper-parametre optimizasyonu aşamasında, Python programlama dilindeki scikit-optimize (skopt) kitaplığı kullanılarak Bayesian optimizasyon yöntemi uygulanmıştır [37]. Bu yöntem, optimize edilecek hiper-parametreler için belirlenen minimum ve maksimum aralık değerleri arasında en uygun değeri Gauss Süreçleri yardımıyla tespit etmektedir. Optimizasyon sürecinde, skopt kitaplığının gp_minimize fonksiyonu kullanılmıştır. Bu fonksiyonun edinim fonksiyonu Beklenen İyileşme (Expected Improvement - EI) olarak seçilmiş ve optimizasyonun çağrı sayısı (n_calls) 10 olarak belirlenmiştir. Son olarak, optimizasyon metriği (hedef başarım ölçütü) olarak ROUGE-1 metriği kullanılmıştır.

Hiper-parametre optimizasyonu aşamasından sonra optimum hiper-parametreleri kullanan ve eğitim veri küme kullanılarak eğitilen modelin başarım Değerleri sınama veri küme üzerinde hesaplanmıştır. Bu kapsamda ROUGE-1, ROUGE-2 ve ROUGE-L metrikleri hesaplanmıştır. Anahtar sözcük üretimi görevinde elde edilen sonuçların değerlendirilmesinde ROUGE-1, ROUGE-2 ve ROUGE-L metriklerinin tercih edilmesi, sözcük düzeyindeki örtüşmeleri ve kısmi sıralama bilgisini etkin bir şekilde yakalamaları açısından gerekçelendirilmiştir. Bu metrikler sırasıyla tekli sözcük, ikili sözcük dizisi ve en uzun ortak alt dizi temelli çakışmayı ölçmektedir. Anahtar sözcük ve anahtar ifade değerlendirmesi genellikle kısa ve lokma tabanlı bir biçimde gerçekleştiği için, n-gram temelli bu karşılaştırmalar, model başarımına ilişkin doğrudan ve açıklanabilir bir ölçüt sunmaktadır [38]. Çizelge 4'te her bir model için hesaplanmış olan başarım Değerleri gösterilmektedir.

Çizelge-4: İnce Ayar Uygulanmış Modellerin Sınama Veri Küme Üzerindeki Başarım Değerleri

Model	ROUGE-1	ROUGE-2	ROUGE-L
T5 Türkçe	0.4089	0.2041	0.4054
T5 İngilizce	0.3996	0.1675	0.3886
FlanT5 Türkçe	0.4106	0.2046	0.4070
FlanT5 İngilizce	0.3965	0.1654	0.3854

Çizelge 4'te sunulan sonuçlar incelendiğinde, anahtar sözcük tahmini görevinde değerlendirilen dört modelin ROUGE-1, ROUGE-2 ve ROUGE-L metrikleri açısından genel olarak tatmin edici bir başarım sergilediği görülmektedir. Türkçe modellerin (T5 ve FlanT5), hem ROUGE-1 hem de ROUGE-L değerlerinde İngilizce modellerden daha yüksek değerlere ulaşması, modellerin veri kümesinin dilsel özellikleriyle daha uyumlu bir öğrenme süreci gerçekleştirdiğini göstermektedir. Özellikle FlanT5 Türkçe modelinin ROUGE-1'de 0.4106 ve ROUGE-L'de 0.4070 seviyelerine ulaşması, kaynaklarda yer alan benzer çalışmalarla karşılaştırıldığında rekabetçi bir başarı düzeyine işaret etmektedir; nitekim Türkçe anahtar sözcük çıkarımı üzerine yürütülen bir çalışmada BERT tabanlı modelin ROUGE-1 F1 değerinin 0.399 olarak raporlanmış olması, elde edilen sonuçların alanda kabul gören başarım aralığıyla uyumlu olduğunu göstermektedir [39]. Bununla birlikte ROUGE-2 Değerlerinin nispeten düşük kalması, modellerin çok sözcükli anahtar ifade üretiminde sınırlı başarı gösterdiğine işaret etmekte olup, kaynaklarda ROUGE tabanlı değerlendirmelerin bigram ve çok sözcüklü örtüşmeleri daha düşük skorlarla yansıtmasının doğal bir sonucu olarak değerlendirilebilir. Genel olarak ele alındığında, elde edilen ROUGE Değerleri anahtar sözcük tahmini görevindeki başarımın yeterli olduğunu ortaya koymakta; bunun yanında, gelecekte semantik benzerlik odaklı ek metriklerin kullanılmasıyla değerlendirme sürecinin daha bütüncül şekilde desteklenebileceği öngörülmektedir.

Her ne kadar ROUGE-1, ROUGE-2 ve ROUGE-L metrikleri anahtar sözcük tahmini modellerinin başarımını değerlendirmede yaygın olarak kullanılmakta olsa da büyük dil modelleri kullanılarak geliştirilen sistemlerin değerlendirilmesinde bu metrikler çoğu zaman yetersiz kalabilmektedir. Bunun temel nedeni, büyük dil modellerinin aynı kavramı farklı sözcüklerle ifade edebilmesi ve bu nedenle yüzeysel n-gram eşleşmesine dayalı metriklerin gerçek başarımı tam olarak yansıtamamasıdır. Örneğin, "tahmin modeli" ve "öngörü modeli" ifadeleri farklı sözcüklerden oluşsa da anlamsal olarak eşdeğerdir ve model çıktısı açısından doğru kabul edilmelidir. Bu durumu dikkate alarak çalışmamızda, modellerin anlamsal düzeydeki başarısı da ayrıca analiz edilmiştir. Bu kapsamda, sentence_ transformers [40] kitaplığı kullanılarak hem üretilen hem de gerçek anahtar sözcüklerin anlamsal gömme (embedding) vektörleri elde edilmiş ve iki sözcük küme arasındaki benzerlik, cosine similarity ölçütü ile hesaplanmıştır. Benzerlik hesaplamalarında HuggingFace platformunda bulunan modellerden yararlanılmıştır [41]. Türkçe anahtar sözcük benzerlik Değerleri hesaplanırken "dbmdz/bert-base-turkish-cased", "savasy/bert-base-turkish-sentiment-cased" ve "BAAI/bge-m3" modelleri; İngilizce anahtar sözcük benzerlik Değerleri için ise "all-distilroberta-v1", "sentence-transformers/stsb-roberta-large" ve "sentence- transformers /stsb-bert-base" modelleri kullanılmıştır. Hesaplanan anlamsal benzerlik sonuçları Çizelge 5'te sunulmaktadır.

Çizelge-5: İnce Ayar Uygulanmış Modellerin Sınama Veri Küme Üzerindeki Anlamsal Başarım Analizi

Anahtar Sözcük Tahmin Modeli	Gömme Çıkarımı Modeli	Benzerlik Değeri
T5 Türkçe	dbmdz/bert-base-turkish-cased	0.8494
	savasy/bert-base-turkish-sentiment-cased	0.7520
	BAAI/bge-m3	0.7028
FlanT5 Türkçe	dbmdz/bert-base-turkish-cased	0.8485
	savasy/bert-base-turkish-sentiment-cased	0.7524
	BAAI/bge-m3	0.7043
T5 İngilizce	all-distilroberta-v1	0.6884
	sentence-transformers/stsb-roberta-large	0.7332
	sentence-transformers/stsb-bert-base	0.7632
FlanT5 İngilizce	all-distilroberta-v1	0.6264
	sentence-transformers/stsb-roberta-large	0.6549
	sentence-transformers/stsb-bert-base	0.5989

Çizelge 5'te, ince ayar uygulanmış modellerin anlamsal düzeydeki başarımları, farklı gömme modelleri kullanılarak hesaplanan cosine benzerlik Değerleri üzerinden karşılaştırmalı olarak sunulmaktadır. Elde edilen sonuçlar, Türkçe modellerinin hem T5 hem de FlanT5 yapılarında, özellikle "dbmdz/bert-base-turkish-cased" modeli ile hesaplanan benzerlik Değerlerinde 0,84'ün üzerinde değerlere ulaşarak yüksek bir tutarlılık sergilediğini göstermektedir. Bu durum, Türkçe modellerin üretmiş olduğu anahtar sözcüklerin, referans anahtar sözcüklerle anlam bakımından büyük ölçüde örtüştüğüne işaret etmektedir. "savasy/bert-base-turkish-sentiment-cased" ve "BAAI/bge-m3" modelleriyle hesaplanan Değerlerin nispeten daha düşük olmasına rağmen, her iki T5 tabanlı Türkçe modelinin sonuçları birbirine yakın olup, model davranışının gömme tekniğinden bağımsız biçimde istikrarlı olduğunu göstermektedir. İngilizce modellerde ise Değerlerin gömme modeline daha duyarlı olduğu görülmektedir. T5 İngilizce modelinin özellikle "sentence-transformers/stsb-bert-base" ile 0.7632 değerine ulaşması, model çıktılarının anlamsal düzeyde belirgin bir doğruluk taşıdığını ortaya koyarken, FlanT5 İngilizce modelinin tüm gömme modellerinde daha düşük benzerlik Değerleri üretmesi dikkat çekicidir. Genel olarak değerlendirildiğinde, Çizelge 5'te sunulan bulgular, Türkçe modellerin anlamsal benzerlik açısından daha güçlü başarım sergilediğini, İngilizce modellerde ise gömme modeline bağlı değişkenliklerin daha belirgin olduğunu göstermekte; bu durum, ROUGE tabanlı yüzeysel ölçümlerin ötesine geçilerek anlamsal doğruluğun değerlendirilmesinin model başarımını daha bütüncül şekilde yansıttığını ortaya koymaktadır.

Model tarafından üretilen anahtar sözcüklerin başarısını daha iyi görebilmek için çalışmaya kalitatif analizde eklenmiştir. Bu kapsamda T5 modeli için Türkçe ve İngilizce dillerinde en yüksek ve en düşük anlamsal benzerlik değerine sahip girdiler Çizelg- 6'da paylaşılmıştır.

Çizelge-6: T5 Modeli için Türkçe ve İngilizce Metinlerde En Yüksek ve En Düşük Anlamsal Benzerlik Değerlerine Sahip Anahtar Sözcük Tahminleri

Anahtar Sözcük Tahmin Modeli	Özet	Gerçek Anahtar Sözcükler	Üretilen Anahtar Sözcükler	Benzerlik Değeri
T5 Türkçe	İnsan beslenmesinde stratejik açıdan en önemli bitkilerin başında yer alan buğday, büyüme gelişme ve verimini etkileyen çevresel stres faktörleri içerisinde en önemlilerinin başında yer alan kuraklığa tolerensının yüksek olması mutlak gerekli bir bitkidir. 2025 yılına kadar yaklaşık 1.8 milyar insanın mutlak su kıtlığı ile karşı karşıya kalacağı ve dünya nüfusunun % 65'inin su sıkıntısı çeken ortamlarda yaşayacağı tahmin edildiği göz önüne alındığında, kuraklık stresi önemli bir sorun olarak her zaman önemsenmesi gereken bir durum olacaktır. Bu çalışmada; 12 adet makarnalık buğday genotipi, Bahri Dağdaş Uluslararası Tarımsal Araştırma Enstitüsü Konya Merkez lokasyonuna ait, sulu ve yağışa dayalı deneme alanlarında, kuraklık hassasiyet indeksi bakımından karakterize edilmiştir. Ayrıca genotiplere ait tane verimleri ile indeks arasında ilişkiler belirlenmiştir. Kuraklık hassasiyet indeksine göre çalışmalarımızda hem sulanan alanlar için, hem de yağışa dayalı alanlar için genotiplerin seçilmesine karar verilmiştir. Çalışmada, Türköz çeşidinin kuraklık hassasiyetinin diğer genotiplere göre daha iyi olduğu belirlenmiştir.	verim, genotip, Makarnalk buday, kuraklk hassasiyet indeksi	verim, genotip, Makarnalk buday, kuraklk hassasiyet indeksi	1
	Son yıllarda endüstrileşme ve artan sanayi faaliyetleri nedeniyle denize kıyısı olan şehirlerde görülen deniz kirliliği, Türkiye'nin üçüncü en kalabalık kenti olan İzmir'de de görülmekteydi. Ege Denizinin en verimli alanlarından birisi olan "İzmir Körfezi'nde başlatılan büyük kanal projesi ve yıkılan Ragıp Paşa Dalyanı sayesinde körfez akıntısında düzelme ve kirlilikte azalma olmasına rağmen; dönem dönem gözlenen alg patlamaları, denizdeki canlıları olduğu kadar insan sağlığını da tehdit etmektedir. Bu çalışmada 1990-2016 yılları arasında İzmir Büyük Şehir Belediyesi, Dokuz Eylül Üniversitesi Deniz Bilimleri ve Teknolojisi Enstitüsü ve Ege Üniversitesi Su Ürünleri Fakültesinin ortaklaşa yürüttüğü izleme projeleri sırasında elde edilen	DSP	zmir, körfezi, deniz kirlilii, salk politikas, gözlenen zararllar	0.32

	360 kantitatif ve 1080 kalitatif örnek incelenerek; hem körfezde geçen süre boyunca gözlenen zararlı alg patlamaları hem de baskın 2 büyük sınıfı olan Dinophyceae ve Bacillariophyceae türlerinin dağılımları incelenmiştir. DSP, ASP ve PSP gibi toksinler üreten türler yanında, herhangi bir toksin içermeyen fakat aşırı üreyerek ortamda olumsuz koşulları oluşturan türler incelenmiştir.			
T5 İngilizce	Microorganisms are not homogeneously distributed in environments, soil systems are heterogeneous. Soil can be an important evidence value in forensic investigations. It is among the important evidences that contribute to the solution of forensic events in forensic sciences. Bacteria contained in the soil are microbiological evidences. Not all bacteria can be cultured by conventional methods and the amount of cultured bacteria remains limited. Metagenomic studies have been carried out for non-culturable Bacteria. The aim of this study is to perform DNA isolation from soil samples taken from Yeşil Lake (swamp), Faculty of Arts and Sciences garden, agricultural land, Sıklık (forest area) regions of Çorum Province in Türkiye and to determine bacterial diversity by metagenomic analysis of DNA isolated from soil samples. Density and differences of isolates according to habitats were determined. It is thought that the result of this study can shed light on previous crime scene studies in the determined habitats and will contribute to possible future crime scene studies and forensic science that may occur later.	DNA isolation, Habitat, Bacterial diversity, Criminalistics	DNA isolation, Habitat, Bacterial diversity, Criminalistics	1
	Anacamptis coriophora (Orchidaceae) is a highly endangered orchid in Türkiye due to its excessive collection and the continuing deterioration of its habitat. In this study, the cultivation conditions of A. coriophora were determined. A sterile soil mixture was filled into jars and the fungal isolate (previously isolated from A. coriophora roots), Ceratobasidiaceae MG762693 was inoculated in separate glass jars, producing fungal compost when hyphae were developed. This fungal compost was then filled into pots where A. coriophora seed packs (0.001 g) were placed and subsequently moistened with sterile liquid nutrient medium. After 45 days of germination, fifty seedlings of approximately equal size were transferred directly to a natural environment and after 6 months of development the measuring of the tubers was done. The phenological process was then monitored until flowering. After 45 days, germination and developmental stages rates were determined from the seed packs in the pots inoculated with the Ceratobasidiaceae MG762693 fungal isolate and 64.3% germination and 11.75% leaf-rooted seedlings (stage 4) occurred. Plants flowered in June the following year, and the seeds ripened in July. The largest tuber in adult individuals was about 3 times the weight of first-year tubers. Each individual formed 2 or 3 tubers, thus increasing the number of tubers approximately 2.5 times in 2 years. In this study, ex vitro symbiotic seedlings were planted in the natural environment and a small population was formed in a 2-year period. The results revealed that orchids can be grown on a large scale with this method, both economically and for conservation and reintroduction.	Orchidaceae	Orchidaceae, Seed germination, Symbiotic, Orchid cultivation	0.28

Çizelge 6'da T5 modelinin Türkçe ve İngilizce metinler üzerindeki anahtar sözcük üretim başarımını nitel olarak değerlendirmek amacıyla, en yüksek ve en düşük anlamsal benzerlik Değerlerine sahip örnekler sunulmuştur. Yüksek benzerlik değerine sahip örneklerde, modelin gerçek anahtar sözcükleri birebir ya da büyük ölçüde doğru biçimde tahmin ettiği görülmektedir. Düşük benzerlik değerine sahip örneklerde ise, çoğu durumda gerçek anahtar sözcüklerin eksik olduğu ve model tarafından üretilen anahtar sözcüklerin anlamsal açıdan daha uygun olduğu açıkça

gözlemlenmektedir. Bununla birlikte, düşük skorlu örneklerde yeniden yazıma bağlı olarak bazı sözcük düzeyi hataların ortaya çıktığı da görülmektedir. Örneğin, Türkçe düşük başarımlı örnekte yer alan "zmir, körfezi, deniz kirlilii, salk politikas, gözlenen zararllar" ifadeleri biçimsel olarak hatalıdır. Bu durum, modelin öğrenme sürecindeki sınırlılıklardan kaynaklanabileceği gibi, daha gelişmiş modellerin veya ek iyileştirme stratejilerinin kullanılmasının gerekliliğine de işaret etmektedir.

5. Tartışma ve Sonuç

Bu çalışmada, Türkçe ve İngilizce metinler için özgün anahtar sözcük üretimi amacıyla iki büyük ölçekli veri küme oluşturulmuş, ardından T5 ve FlanT5 modellerine Bayesian optimizasyon destekli ince ayar uygulanmıştır. Elde edilen bulgular, özellikle Türkçe modellerin hem ROUGE tabanlı hem de anlamsal benzerlik ölçütlerinde İngilizce modellere kıyasla daha yüksek başarım sergilediğini göstermektedir. Bu sonuç, Türkçe veri kümenin hacmi, morfolojik özelliklerin model tarafından başarılı şekilde öğrenilmesi ve veri çeşitliliğinin genelleme kapasitesine katkısıyla açıklanabilir. İngilizce modellerde ise özellikle anlamsal benzerlik değerlerindeki dalgalanma, gömme modellerinin varyansına duyarlılığı ortaya koymaktadır.

Modellerin ROUGE-2 değerlerinin görece düşük kalması, çok sözcükli anahtar ifadelerin üretilmesindeki zorluklara işaret etmekte olup, bu durum özellikle n gram tabanlı metriklerin sınırlılıklarıyla ilişkilidir. Bununla birlikte, anlamsal benzerlik sonuçları, büyük dil modellerinin yüzeysel eşleşme yerine kavramsal örtüşme üretme eğilimini doğrulamakta ve otomatik anahtar sözcük üretimi süreçlerinde semantik ölçümlerin gerekliliğini vurgulamaktadır. Geniş ve disiplinler arası veri küme sayesinde modellerin farklı alanlara uygulanabilirliği artmış, özellikle Türkçe gibi kaynakları sınırlı dillerde büyük dil modellerine ince ayarın önemli bir katkı sunduğu ortaya konmuştur.

Öte yandan, kaynaklarda anahtar sözcük belirleme amacıyla sıklıkla başvurulan TF-IDF, TextRank ve KeyBERT gibi denetimsiz veya istatistiksel yöntemler, temel olarak 'çıkarımsal' (extractive) bir yaklaşım sergilemektedir. Bu yöntemler, belge içerisinde halihazırda var olan sözcükleri belirli skorlama algoritmalarına göre seçerek aday anahtar sözcük olarak sunmaktadır. Ancak bu çalışmada kullanılan T5 ve FlanT5 gibi büyük dil modelleri, 'üretken/yorumlayıcı' (abstractive) bir yeteneğe sahiptir. Bu sayede modeller, metinde açıkça geçmese dahi, belgenin anlamsal bütünlüğünü temsil eden, eş anlamlı veya kaynaklarda kabul görmüş terminolojik kavramları anahtar sözcük olarak üretebilmektedir. Dolayısıyla, çıkarımsal yöntemlerin başarısını ölçen metriklerle, üretken modellerin anlamsal zenginliğini karşılaştırmak metodolojik bir farklılık içermektedir. Bu çalışmada elde edilen yüksek anlamsal benzerlik Değerleri (Çizelge 5), üretken modellerin sadece sözcük eşleşmesi değil, kavramsal temsil yeteneği açısından da çıkarımsal yöntemlerin ötesine geçme potansiyeli taşıdığını göstermektedir.

Gelecek çalışmalar kapsamında, daha gelişmiş değerlendirme metrikleri (BERTScore, MoverScore ve semantik kümeleme tabanlı ölçütler gibi) kullanılacaktır. Ayrıca, çok sözcükli anahtar ifade üretiminin geliştirilmesi amacıyla dizilim odaklı modeller, talimat tabanlı öğrenme stratejileri ve pekiştirmeli öğrenme yaklaşımları uygulanacaktır. Veri küme genişletilecek, alan bazlı ince ayar süreçleri yürütülecek ve farklı büyük dil modelleri (LLaMA, Mistral, Gemma vb.) üzerinde karşılaştırmalı analizler yapılacaktır. Ek olarak, üretilen anahtar sözcüklerin arama motoru odaklı kullanılabilirliğini incelemek için bilgi erişim başarımına dayalı ölçütler sisteme entegre edilecektir. Bu çalışmalar, modelin kapsamını ve gerçek dünya uygulamalarındaki etkinliğini daha da artıracaktır.

Bu çalışma hem özgün veri küme üretimi hem de büyük dil modellerinin Türkçe ve İngilizce anahtar sözcük üretimi bağlamında incelenmesi açısından literatüre önemli katkılar sunmakta ve özellikle Türkçe doğal dil işleme çalışmalarında büyük ölçekli derin öğrenme yöntemlerinin etkinliğini ortaya koymaktadır.

Teşekkür

Bu çalışmada elde edilen sonuçların bir kısmı Doç. Dr. Yasin GÖRMEZ tarafından danışmanlığı yapılan, Şeyda Nur ARSLAN tarafından hazırlanacak olan "Türkçe Metinlerde Anahtar Sözcük Tahmini: Özgün Veri Küme Oluşturulması ve Derin Öğrenme Tabanlı Model Geliştirme" başlıklı tezde kullanılacaktır. Bu çalışma aynı zamanda "Detay Danışmanlık Bilgisayar Hizmetleri Sanayi ve Dış Ticaret A.Ş" Ar-Ge merkezinde yürütülen çalışmaların bir ürünüdür, destekleri için teşekkür ederiz. Bu araştırmada yer alan kısmi nümerik hesaplamalar TÜBİTAK ULAKBİM, Yüksek Başarım ve Grid Hesaplama Merkezi'nde (TRUBA kaynaklarında) gerçekleştirilmiştir.

Kaynakça

[1] Waldherr, A. Maier, D. Miltner, P. ve Günther, E. "Big Data, Big Noise: The Challenge of Finding Issue Networks on the Web", *Social Science Computer Review*, c. 35, sy 4, ss. 427-443, Ağu. 2017, doi: 10.1177/0894439316643050.

[2] Cai L. ve Zhu, Y. "The Challenges of Data Quality and Data Quality Assessment in the Big Data Era", *Data Science Journal*, c. 14, sy 0, May. 2015, doi: 10.5334/dsj-2015-002.

[3] Grant, M. J. "Key words and their role in information retrieval", *Health Info Libr J*, c. 27, sy 3, ss. 173-175, Eyl. 2010, doi: 10.1111/j.1471-1842.2010.00904.x.

[4] Berenci, E. Carpineto, C. Giannini, V. ve Mizzaro, S. "Effectiveness of Keyword-Based Display and Selection of Retrieval Results for Interactive Searches", içinde *Research and Advanced Technology for Digital Libraries*, S. Abiteboul ve A.-M. Vercoustre, Ed., Berlin, Heidelberg: Springer, 1999, ss. 106-125. doi: 10.1007/3-540-48155-9_9.

[5] Rose, S. Engel, D. Cramer, N. ve Cowley, W. "Automatic Keyword Extraction from Individual Documents", içinde *Text Mining*, John Wiley & Sons, Ltd, 2010, ss. 1-20. doi: 10.1002/9780470689646.ch1.

[6] Maragheh R. Y. ve ark., "LLM-TAKE: Theme Aware Keyword Extraction Using Large Language Models", 01 Aralık 2023, *arXiv*: arXiv:2312.00909. doi: 10.48550/arXiv.2312.00909.

[7] Koloski, B. Pollak, S. Škrlj, B. ve Martinc, M. "Out of Thin Air: Is Zero-Shot Cross-Lingual Keyword Detection Better Than Unsupervised?", 14 Şubat 2022, *arXiv*: arXiv:2202.06650. doi: 10.48550/arXiv.2202.06650.

[8] Altuncu, E. Nurse, J. R. C. Xu, Y. Guo, J. ve Li, S. "Improving Performance of Automatic Keyword Extraction (AKE) Methods Using PoS Tagging and Enhanced Semantic-Awareness", *Information*, c. 16, sy 7, Art. sy 7, Tem. 2025, doi: 10.3390/info16070601.

[9] Khan M. Q. ve ark, "Impact analysis of keyword extraction using contextual word embedding", *PeerJ Comput. Sci.*, c. 8, s. e967, May. 2022, doi: 10.7717/peerj-cs.967.

[10] Kim Y. ve ark., "Validation of deep learning natural language processing algorithm for keyword extraction from pathology reports in electronic health records", *Sci Rep*, c. 10, sy 1, s. 20265, Kas. 2020, doi: 10.1038/s41598-020-77258-w.

[11] Zhang, Y. Tuo, M. Yin, Q. Qi, L. Wang, X. ve Liu, T. "Keywords extraction with deep neural network model", *Neurocomputing*, c. 383, ss. 113-121, Mar. 2020, doi: 10.1016/j.neucom.2019.11.083.

[12] Duari S. ve Bhatnagar, V. "Complex Network based Supervised Keyword Extractor", *Expert Systems with Applications*, c. 140, s. 112876, Şub. 2020, doi: 10.1016/j.eswa.2019.112876.

[13] Shin, H. Lee, H. J. ve Cho, S. "General-use unsupervised keyword extraction model for keyword analysis", *Expert Systems with Applications*, c. 233, s. 120889, Ara. 2023, doi: 10.1016/j.eswa.2023.120889.

[14] Zhang, K. Xu, H. Tang, J. ve Li, J. "Keyword Extraction Using Support Vector Machine", içinde *Advances in Web-Age Information Management*, J. X. Yu, M. Kitsuregawa, ve H. V. Leong, Ed., Berlin, Heidelberg: Springer, 2006, ss. 85-96. doi: 10.1007/11775300_8.

[15] Pay T. ve Lucci, S. "Automatic keyword extraction: An ensemble method", içinde *2017 IEEE International Conference on Big Data (Big Data)*, Ara. 2017, ss. 4816-4818. doi: 10.1109/BigData.2017.8258552.

[16] Bracewell, D. B. Ren, F. ve Kuriowa, S. "Multilingual single document keyword extraction for information retrieval", içinde *2005 International Conference on Natural Language Processing and Knowledge Engineering*, Eki. 2005, ss. 517-522. doi: 10.1109/NLPKE.2005.1598792.

[17] Çelikten, A. Onan, A. ve Bulut, H. "Investigation of Biomedical Named Entity Recognition Methods", içinde *4th International Conference on Artificial Intelligence and Applied Mathematics in Engineering*,

[18 Hemanth,]D. J. Yigit, T. Kose, U. ve Guvenc, U. Ed., Cham: Springer International Publishing, 2023, ss. 218-229. doi: 10.1007/978-3-031-31956-3_18.

[19] Çelikten, A. Uğur, A. ve Bulut, H. "Keyword Extraction from Biomedical Documents Using Deep Contextualized Embeddings", içinde *2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, Ağu. 2021, ss. 1-5. doi: 10.1109/INISTA52262.2021.9548470.

[20] Singhal K. ve ark., "Toward expert-level medical question answering with large language models", *Nat Med*, c. 31, sy 3, ss. 943-950, Mar. 2025, doi: 10.1038/s41591-024-03423-7.

[21] Chen Q. ve ark., "Benchmarking large language models for biomedical natural language processing applications and recommendations", *Nat Commun*, c. 16, sy 1, s. 3280, Nis. 2025, doi: 10.1038/s41467-025-56989-2.

[22] Keloth V. K. *ve ark.*, "Advancing entity recognition in biomedicine via instruction tuning of large language models", *Bioinformatics*, c. 40, sy 4, s. btae163, Nis. 2024, doi: 10.1093/bioinformatics/btae163.

[23] Luo X. *ve ark.*, "Large language models surpass human experts in predicting neuroscience results", *Nat Hum Behav*, c. 9, sy 2, ss. 305-315, Şub. 2025, doi: 10.1038/s41562-024-02046-9.

[24] Liew P. Y. ve T. Tan, I. K. "On Automated Essay Grading using Large Language Models", içinde *Proceedings of the 2024 8th International Conference on Computer Science and Artificial Intelligence*, içinde CSAI '24. New York, NY, USA: Association for Computing Machinery, Şub. 2025, ss. 204-211. doi: 10.1145/3709026.3709030.

[25] Choudhary, M. Agarwal, B. ve Goyal, V. "Hate Speech Detection: Leveraging LLM-GPT2 with Fine-Tuning and Multi-Shot Techniques", *Procedia Computer Science*, c. 258, ss. 2817-2825, Oca. 2025, doi: 10.1016/j.procs.2025.04.542.

[26] Wan, Z. "Text Classification: A Perspective of Deep Learning Methods", 24 Eylül 2023, *arXiv*: arXiv:2309.13761. doi: 10.48550/arXiv.2309.13761.

[27] Li Q. *ve ark.*, "A Survey on Text Classification: From Traditional to Deep Learning", *ACM Trans. Intell. Syst. Technol.*, c. 13, sy 2, s. 31:1-31:41, Nis. 2022, doi: 10.1145/3495162.

[28] Zhang, T. Ladhak, F. Durmus, E. Liang, P. McKeown, K. ve Hashimoto, T. B. "Benchmarking Large Language Models for News Summarization", *Transactions of the Association for Computational Linguistics*, c. 12, ss. 39-57, Oca. 2024, doi: 10.1162/tacl_a_00632.

[29] Basyal L. ve Sanghvi, M. "Text Summarization Using Large Language Models: A Comparative Study of MPT-7b-instruct, Falcon-7b-instruct, and OpenAI Chat-GPT Models", 17 Ekim 2023, *arXiv*: arXiv:2310.10449. doi: 10.48550/arXiv.2310.10449.

[30] "TR Dizin Entegrasyonu Geliştirme Dokümanı - TR Dizin Yardım". Erişim: 19 Ekim 2025. [Çevrimiçi]. Erişim adresi: https://development.trdizin.gov.tr/

[31] "ysngrmz/trdizin_keywords · Datasets at Hugging Face". Erişim: 27 Aralık 2025. [Çevrimiçi]. Erişim adresi: https://huggingface.co/datasets/ysngrmz/trdizin_keywords

[32] Raffel C. *ve ark.*, "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Dönüştürücü", 19 Eylül 2023, *arXiv*: arXiv:1910.10683. doi: 10.48550/arXiv.1910.10683.

[33] Chung H. W. *ve ark.*, "Scaling Instruction-Finetuned Language Models", 06 Aralık 2022, *arXiv*: arXiv:2210.11416. doi: 10.48550/arXiv.2210.11416.

[34] Snoek, J. Larochelle, H. ve Adams, R. P. "Practical Bayesian Optimization of Machine Learning Algorithms", 29 Ağustos 2012, *arXiv*: arXiv:1206.2944. doi: 10.48550/arXiv.1206.2944.

[35] Bergstra, J. Bardenet, R. Bengio, Y. ve Kégl, B. "Algorithms for Hyper-Parameter Optimization", içinde *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2011. Erişim: 19 Ekim 2025. [Çevrimiçi]. Erişim adresi: https://proceedings.neurips.cc/paper/2011/hash/86e8f7ab32cfd12577bc2619bc635690-Abstract.html

[36] “FLAN-T5”. Erişim: 19 Ekim 2025. [Çevrimiçi]. Erişim adresi: https://huggingface.co/docs/Transformers/model_doc/flan-t5

[37] “T5”. Erişim: 19 Ekim 2025. [Çevrimiçi]. Erişim adresi: https://huggingface.co/docs/Transformers/model_doc/t5

[38] *scikit-optimize: Sequential model-based optimization toolbox.* (2025). Python. Erişim: 19 Ekim 2025. [MacOS, Microsoft :: Windows, POSIX, Unix]. Erişim adresi: https://scikit-optimize.readthedocs.io/en/lasınama/contents.html

[39] Wu, D. Cheng, P. ve Zheng, Y. “Multistage Mixed-Attention Unsupervised Keyword Extraction for Summary Generation”, *Applied Sciences*, c. 14, sy 6, s. 2435, Oca. 2024, doi: 10.3390/app14062435.

[40] Aydın Ö. ve Kantarcı, H. “Türkçe Anahtar Sözcük Çıkarımında LSTM ve BERT Tabanlı Modellerin Karşılaştırılması”, *bbmd*, c. 17, sy 1, ss. 9-18, Haz. 2024, doi: 10.54525/bbmd.1454220.

[41] *sentence-Transformers: Embeddings, Retrieval, and Reranking*. Python. Erişim: 06 Aralık 2025. [Çevrimiçi]. Erişim adresi: https://www.SBERT.net

[42] “Hugging Face – The AI community building the future.” Erişim: 06 Aralık 2025. [Çevrimiçi]. Erişim adresi: https://huggingface.co/