İSTANBUL TİCARET ÜNİVERSİTESİ
FEN BİLİMLERİ DERGİSİ

İstanbul Commerce University Journal of Science

http://dergipark.org.tr/ticaretfbd

Araştırma Makalesi / Research Article

TRimCapS: MAKİNE ÖĞRENMESİ İLE TÜRKÇE DİLİNDEKİ GÖRÜNTÜ ALT YAZILARINI SINIFLANDIRMA SİSTEMİ

TRimCapS: A MACHINE LEARNING-BASED SYSTEM FOR CLASSIFYING IMAGE CAPTIONS IN THE TURKISH LANGUAGE

Merve PINAR[1] **Esra YILMAZ**[2] **Zeki ÇIPLAK**[3]
Ayşe Berna ALTINEL GİRGİN[4]

https://doi.org/10.55071/ticaretfbd.1635443

Sorumlu Yazar (Corresponding Author)
merve.pinar@marmara.edu.tr

Geliş Tarihi (Received)
08.02.2025

Revizyon Tarihi (Revised)
10.07.2025

Kabul Tarihi (Accepted)
11.07.2025

Öz

Dijital medyanın yaygınlaşmasıyla görüntü ve video içeriklerinin analizi önem kazanmıştır. Ancak, Türkçe alt yazı sınıflandırması, dilin yapısal zorlukları ve sınırlı veri kümeleri nedeniyle büyük bir araştırma sorunu oluşturmaktadır. Bu sorunu ele almak için TasvirEt, Flickr30k ve MS COCO veri kümeleri birleştirilerek 114.566 görüntü ve 588.867 Türkçe alt yazı içeren ImCapTR veri kümesi oluşturulmuştur. Önerilen TRimCapS sisteminde, alt yazılar TF-IDF, CountVectorizer ve GloVe ile vektörleştirilmiş, K-Means ve Latent Dirichlet Allocation kullanılarak kategorize edilmiştir. Özellik seçimi bilgi kazancı, ki-kare, Fisher skoru, karşılıklı bilgi ve temel bileşenler analizi yöntemleriyle gerçekleştirilmiştir. Çeşitli makine öğrenimi ve derin öğrenme modelleriyle yapılan sınıflandırma deneylerinde, CountVectorizer ve BERT kombinasyonu %98,84 doğruluk oranı ile en iyi sonucu vermiştir. Bilgi kazancı ve temel bileşenler analizi, diğer yöntemlere göre daha yüksek performans göstermiştir. Bu çalışma, Türkçe alt yazı sınıflandırması konusunda en kapsamlı deney sonuçlarını sunan ve oluşturulan veri kümesini araştırmacıların erişimine açan ilk çalışmadır.

Anahtar Kelimeler: Türkçe görüntü alt yazılandırma, alt yazı sınıflandırma, BERT, metin vektörleştirme, özellik seçimi.

Abstract

With the widespread adoption of digital media, the analysis of image and video content has gained significance. However, Turkish subtitle classification presents a major research challenge due to the structural complexities of the language and the limited availability of datasets. To address this issue, the TasvirEt, Flickr30k, and MS COCO datasets were combined to create the ImCapTR dataset, which contains 114.566 images and 588.867 Turkish subtitles. In the proposed TRimCapS system, subtitles were vectorized using TF-IDF, CountVectorizer, and GloVe, and categorized using K-Means and Latent Dirichlet Allocation. Feature selection was performed using information gain, chi-square, Fisher score, mutual information, and principal component analysis. Classification experiments utilizing various machine learning and deep learning models demonstrated that the combination of CountVectorizer and BERT achieved the highest accuracy of 98.84%. Information gain and principal component analysis outperformed other feature selection methods. This study is the first to provide the most comprehensive experimental results on Turkish subtitle classification while making the constructed dataset publicly accessible to researchers.

Keywords: Turkish image captioning, caption classification, BERT, text vectorization, feature selection.

[1] Marmara Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul, Türkiye.
merve.pinar@marmara.edu.tr.

[2] Marmara Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı, İstanbul, Türkiye.
esra@marmara.edu.tr.

[3] İstanbul Gedik Üniversitesi, Gedik Meslek Yüksekokulu, Bilgisayar Teknolojileri Bölümü, İstanbul, Türkiye.
zeki.ciplak@gedik.edu.tr.

[4] Marmara Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul, Türkiye.
berna.altinel@marmara.edu.tr.

1. GİRİŞ

Dijital medya içeriklerinin hızla artmasıyla birlikte, görüntü ve video veri tabanları giderek daha fazla popülerlik kazanmaktadır. Bu veri tabanlarındaki içeriğin işlenmesi, anlamsal olarak anlaşılması ve bilgi elde edilmesi önemli ve zorlu bir konu haline gelmektedir. İçeriklere hızlı ve etkili erişim sağlamak için bunların özelliklerine, kategorilerine ve anlam düzeylerine göre ayrıntılı şekilde sınıflandırılması gerekir. Bu süreç, kullanıcıların arama sorgularında daha hassas ve anlamlı sonuçlar elde etmelerine yardımcı olurken, aynı zamanda veri tabanlarının yönetimini ve içeriğin keşfedilmesini de kolaylaştırır.

Görüntü içeriği sınıflandırması, çeşitli teknikler ve metodolojiler kullanılarak gerçekleştirilebilir. Bu kapsamda, görüntü alt yazı sınıflandırması, nesne algılama algoritmaları ve görüntü sınıflandırma teknikleri gibi çeşitli yöntemler kullanılmaktadır. Bu yöntemler, görüntülerin içeriğini anlamak, nesneleri tanımlamak ve sahneleri sınıflandırmak için tasarlanmıştır. Bu sayede, büyük veri kümeleri üzerinde etkin ve doğru sınıflandırma yapmak mümkün olmaktadır. Szegedy ve ark. derin evrişimli sinir ağlarının kullanımını önererek, 2014 yılında düzenlenen ImageNet Büyük Ölçekli Görsel Tanıma Yarışması'nda tanıtılan Inception mimarisiyle görüntü sınıflandırma ve tespitinde önemli gelişmeler sağlamıştır (Szegedy ve ark., 2015). Gaspar ve Alexandre, görüntü içeriğine dayalı duygu sınıflandırması oluşturmak için hem metin hem de görüntü verilerini kullanarak sınıflandırmayı içeren görüntü duygu analizi için çok modelli bir yaklaşım sunmuştur (Gaspar & Alexandre, 2019). Bu daha kapsamlı görüntü içeriği sınıflandırması için çoklu yöntemlerin entegre edilme potansiyelini göstermektedir.

Alt yazı sınıflandırması, görüntülerin içeriğini doğru şekilde tanımlamayı ve kategorize etmeyi amaçlayan görüntü açıklamalarının analizini içerir. Bu süreç, görsellere ait metin tabanlı açıklamaların belirli kategorilere veya sınıflara ayrılmasını sağlar. Temelde, makine öğrenimi modellerinin görsellere eşlik eden metinsel açıklamaları analiz ederek bunları uygun sınıflara ataması hedeflenmektedir. Türkçe, zengin bir morfolojiye sahip olması ve anlam katmanlarının çeşitliliğiyle bilinen bir dildir. Bu dilsel özellikler, görüntü alt yazılarının doğru şekilde sınıflandırılmasını daha karmaşık hâle getirmektedir. Mevcut çalışmalar çoğunlukla İngilizce gibi kaynak zenginliği yüksek diller üzerine odaklanmış; Türkçe gibi kaynakların sınırlı olduğu dillerdeki potansiyel araştırma alanları yeterince incelenmemiştir. Bu makalede sunulan çalışma, görüntülerin Türkçe alt yazı sınıflandırması alanındaki boşluğu doldurmayı amaçlamaktadır. Literatürde yer alan üç farklı Türkçe alt yazı veri kümesinin birleştirilmesiyle oluşturulan ImCapTR veri kümesi, 114.566 görüntü içermekte olup, Türkçe alt yazı sınıflandırmasında kullanılabilecek kapsamlı ve zengin bir kaynak sunmaktadır. Bu veri kümesi üzerinde gerçekleştirilen analiz ve sınıflandırma çalışmaları kapsamında birçok makine öğrenimi ve derin öğrenme tekniği uygulanmıştır. Söz konusu teknikler, veri kümesinin ayrıntılı biçimde incelenmesini ve sınıflandırılmasını sağlamış; elde edilen sonuçlar, Türkçe doğal dil işleme ve görüntü sınıflandırma alanlarında önemli ve yol gösterici bulgular ortaya koymuştur.

Bu makale, öncelikle görüntülerin Türkçe alt yazılarının sınıflandırmasının önemini ve bu alandaki zorlukları tartışmakta; ardından ImCapTR veri kümesinin Şekil 1' de gösterildiği gibi oluşturulma sürecini ve bu veri kümesi üzerinde uygulanan

sınıflandırma metodolojilerini detaylandırmaktadır. Bu çalışmanın, Türkçe içerik analizi ve görüntü alt yazılarının sınıflandırılması konularında gelecek araştırmalara ilham vereceği ve bu alandaki teknik gelişmelere önemli katkılar sağlayacağı öngörülmektedir.

Bu çalışmada alt yazı metinlerinin içerikleri analiz edilerek belirli kategorilere ayrılması, bu sayede veri kümesindeki alt yazı metinlerinin anlamının daha iyi anlaşılması ve doğru bir şekilde sınıflandırılması amaçlanmıştır. Bu çalışmanın temel katkıları şunlardır:

- ImCapTR adlı Türkçe veri kümesi, mevcut çeşitli Türkçe veri kümelerinin birleştirilmesiyle oluşturulmuş olup, 114.566 görüntü ve 588.867 alt yazı içermektedir. Bildiğimiz kadarıyla, kamuya açık en geniş kapsamlı Türkçe veri kümesi olan ImCapTR, doğal dil işleme ve bilgisayarla görme alanlarında önemli bir referans noktasıdır. Çalışmada kullanılan veri kümesine GitHub' daki TRimCapS deposundan erişilebilir.

- ImCapTR veri kümesi, Türkçe dilinin çeşitliliğini yansıtacak şekilde hazırlanarak makine öğrenimi ve derin öğrenme modellerinin eğitimi için değerli bir kaynak oluşturmuştur. Bu sayede, Türkçe doğal dil işleme alanındaki çalışmaların daha kapsamlı ve güvenilir bir temele dayanması amaçlanmıştır.

- TRimCapS sistemi kapsamında, farklı vektörleştirme, kümeleme, özellik seçimi ve sınıflandırma algoritmaları kullanılarak Türkçe alt yazılar üzerinde analizler gerçekleştirilmiştir. Literatürdeki çalışmalar göz önüne alındığında, bu çalışma, Türkçe alt yazıları kullanarak gerçekleştirilen sınıflandırma süreçlerine dair kapsamlı bir inceleme sunan ilk çalışmalardan biridir.

- Türkçe alt yazı sınıflandırması üzerine yapılan önceki çalışmalarla karşılaştırıldığında, ImCapTR veri kümesi hem geniş kapsamlı içeriği hem de uygulanan yöntemlerin çeşitliliği bakımından literatüre önemli bir katkı sağlamaktadır. Bildiğimiz kadarıyla, bu kadar fazla sayıda vektörleştirme, kümeleme, özellik seçimi ve sınıflandırma yönteminin aynı çalışma kapsamında bir araya getirildiği ve karşılaştırmalı bir analizinin yapıldığı ilk araştırmadır.

Şekil 1. ImCapTR Veri Kümesinden Örnek Görüntü ve Alt Yazılar

Bu makale, görüntü alt yazılarının sınıflandırılması üzerine kapsamlı bir çalışma sunmayı amaçlamaktadır. İlk olarak, giriş bölümünde, çalışmanın önemi vurgulanmakta, araştırmanın amacı ve kapsamı açıklanmaktadır. İkinci bölümde, konu ile ilgili geniş bir literatür taraması yapılarak mevcut bilgiler derlenmiş, benzer çalışmalar ayrıntılı bir şekilde açıklanmış ve tablolar halinde sunulmuştur. Ardından Bölüm 3' te ImCapTR veri kümesinin oluşturulması, işlenmesi, çeşitli vektörleştirme, özellik seçimi ve konu modelleme yöntemlerinin teknik detayları verilmiştir. Bölüm 4' te uygulanan yöntemler kıyaslanmış ve bunların karşılaştırmalı analizleri yapılmıştır. Son olarak, bölüm 5' te karşılaşılan mevcut zorluklarla birlikte gelecekteki araştırmalar için yönlendirmeler ve çalışmanın değerlendirme sonuçları ele alınmıştır.

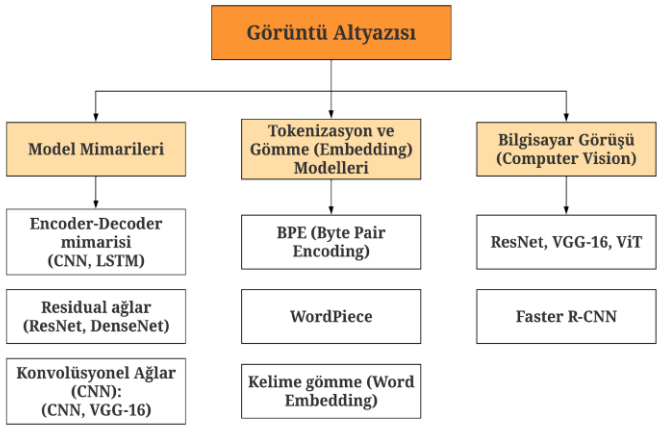
2. İLGİLİ ÇALIŞMALAR

Görüntü alt yazısı, video içeriklerde izleyicilere görsel deneyimi zenginleştirmek, anlamı derinleştirmek ve dil bariyerlerini aşmak amacıyla önemli bir özelliktir. Bu özellik yalnızca video içerikleriyle sınırlı kalmayıp, aynı zamanda sosyal medya paylaşımları, eğitim materyalleri, reklam ve pazarlama malzemeleri, müze açıklamaları, web siteleri ve bloglar gibi çeşitli platformlarda resim tabanlı içeriklerde de geniş bir kullanım alanına sahiptir. Görüntü alt yazısıyla ilgili yapılan literatür taraması, bu alandaki çalışmaların devam ettiğini göstermektedir. Yapılan çalışmaların modellerini, Şekil 2' de görüldüğü gibi gruplandırmak için öncelikle model mimarileri kategorisi incelenebilir. Bu kategori encoder-decoder mimarisi gibi yapıları içerir ve Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM) gibi modelleri kapsar. Diğer bir grup ise, tokenizasyon ve gömme modelleridir. Bu kısımda tokenizasyon için Byte Pair Encoding (BPE) ve WordPiece gibi yöntemlerle kelime

gömme için kullanılan teknikler bulunur. Üçüncü olarak, öğrenme algoritmaları ve kayıp fonksiyonlarını bir araya getiren modeller incelenebilir. Bu kategoride No Language Left Behind (NLLB) ve NLLB Deep gibi kayıp fonksiyonları ile oluşturma ve getirim tabanlı modeller yer almaktadır.

Ünal ve ark. 2016 yılında "TasvirEt" ismiyle Türkçe açıklamalar içeren bir veri kümesi oluşturmuşlardır (Unal ve ark., 2016). Flickr8K veri kümesindeki 8091 görüntü için bir veya iki adet alt yazı olmak üzere toplamda 12.222 Türkçe alt yazı oluşturmuşlardır. Daha sonra bu veri kümesi üzerinde iki farklı alt yazılama modeli önerilmiştir: İlk modelde görüntüye en yakın görsel içeriğin alt yazısı transfer edilmiştir. İkinci modelde ise aday görüntü koleksiyonundaki Türkçe alt yazılar köküne ayrılmıştır. BLEU metriği kullanan deneysel sonuçlarda, Türkçe alt yazılama modeli (0,26), en yakın görüntü tabanlı alt yazılama modeline (0,21) göre daha yüksek başarı elde etmiştir.

Samet ve ark. 2017 yılında yeni bir Türkçe veri kümesi oluşturup, görüntü alt yazılama modeli önermişlerdir (Samet ve ark., 2017). Flickr8k, Flickr30k, MS COCO ve TasvirEt veri kümelerini kullanmışlardır. TasvirEt veri kümesinde her görüntüde bir veya iki Türkçe açıklama, diğerlerinde ise her görüntüde beş İngilizce açıklama vardır, İngilizce alt yazıları Google Translate ile Türkçeye çevirmişlerdir. Yaklaşık 120.000 görüntüyü eğitim, 42.000 görüntüyü test ve doğrulama için kullanarak CNN ve LSTM algoritmalarını içeren bir model oluşturmuşlardır. İngilizce açıklamalar önce Türkçeye çevrilerek ve ayrıca doğrudan İngilizce ile eğitilip çıktıların Türkçeye çevrilmesiyle model eğitimi yapılmış, ardından bu iki yöntemin sonuçları karşılaştırılmıştır.



Şekil 2. Görüntü Alt Yazısı ile İlgili Yapılan Çalışmalarda Oluşturulan Modeller

Kuyu ve ark. 2018 yılında, görüntü alt yazılama alanında Türkçe diline özgün bir yöntem geliştirerek alt sözcük öğelerini kullanmış ve MS COCO 2014 ile Flickr30k veri kümelerini Google Translate ile Türkçeye çevirip, BPE kullanarak alt yazıları alt sözcüklere ayrıştırmış ve yinelemeli derin ağ modeliyle beslenen LSTM tabanlı bir dil modeli kullanmışlardır (Kuyu ve ark., 2018). Ayrıca, tercüme sonrası bazı görüntü açıklamalarının gürültü içerdiği gözlemlendiği için TasvirEt eğitim kümesi kullanılarak

ince ayar yapılmıştır. Deney sonuçları, BLEU, METEOR, Rouge-L, CIDEr ve WMD başarı metrikleriyle karşılaştırmalı olarak rapor edilmiştir.

Sönmez ve ark. otomatik görüntü Türkçe alt yazısı oluşturmak için Vinyals ve ark. ait çalışmadan esinlenmiş ve MS COCO 2014 veri kümesini kullanmışlardır (Sonmez ve ark., 2019, Vinyals ve ark., 2015). Görüntülere ait alt yazılar Yandex API ile Türkçeye çevrilmiş ve insanların değerlendirdiği çevirilerin doğruluğu test edilmiştir. Her görüntü için beş farklı tercüme girilebilen bir web sitesi geliştirilmiş ve CNN ile Recurrent Neural Network (RNN) algoritmalarını kullanan modeller eğitilmiştir. Modellerin başarısı BLEU, METEOR, ROUGE ve CIDEr ölçütleriyle değerlendirilmiş ve sırasıyla 0,297, 0,129, 0,272 ve 0,528 başarı oranları elde edilmiştir.

Golech ve ark. 2022 yılında görüntülerin Türkçe alt yazısını oluşturma üzerinde çalışmış ve Flickr8K, Flickr30K ile MS COCO 2014 veri kümelerini kullanmışlardır (Golech ve ark., 2022). MS COCO 2014 veri kümesi Google API ile Türkçeye çevrilerek 123.287 görüntü ve 616.767 yorum içeren yeni bir veri kümesi oluşturulmuş, İstanbul Bilgi Üniversitesi lisans öğrencileri tarafından gözden geçirilerek doğruluğu sağlanmış ve GitHub' da paylaşılmıştır. Çalışmada, encoder, decoder ve meshed memory transfer (MMT) kullanarak beş model oluşturulmuş ve karşılaştırılmıştır. Önerilen MMT modeli, BLEU skoru 0,72 ile diğer dört Türkçe modelden daha başarılı olmuştur.

Yıldız ve ark. 2023 yılında derin makine çevirisinin Türkçe otomatik görüntü açıklama üzerindeki etkisini araştırmışlardır (Yıldız ve ark., 2023). MS COCO veri tabanındaki İngilizce görüntü açıklamaları, NLLB modeli kullanılarak Türkçeye çevrilmiş ve büyük veri kümeleri eksikliği giderilmiştir. ResNet, ResNext ve Swin gibi derin öğrenme yapıları kullanılarak, LSTM tabanlı bir model görüntülerin Türkçe açıklanması için değerlendirilmiştir. Performans karşılaştırmaları, çeşitli BLEU, METEOR, ROUGE-L ve CIDEr değerleriyle yapılmıştır.

Yıldız ve ark. 2023 yılında otomatik Türkçe görüntü açıklamaları için TRCaptionNet adında yeni bir derin öğrenme tabanlı model tanıtmıştır (Yildiz ve ark., 2023). Model, bir görüntü kodlayıcı, görüntü dönüştürücü tabanlı bir özellik yansıtma modülü ve bir metin decoderinden oluşmaktadır. Giriş görüntüleri, CLIP görüntü kodlayıcısı ile işlenip nihai özellikler elde edilirken, Bidirectional Encoder Representations from Transformers (BERT) tabanlı derin metin decoderi ile Türkçe açıklamalar üretilmiştir. Performans değerlendirmeleri, yaygın görüntü açıklama ölçütleriyle MS COCO ve Flickr30K veri kümelerinde başarılı sonuçlar elde edildiğini göstermiştir.

Ersoy ve ark. 2023 yılında ORTPiece adlı bir çözüm geliştirerek, görüntülerin Türkçe açıklamalarını oluşturmak amacıyla encoder-decoder mimarisini kullanmışlardır (Ersoy ve ark., 2023). Giriş görüntüsü önce bir nesne tespitçisinden geçirilip, görünüm ve geometri tabanlı özellikler elde edilmiş ve bu özellikler encoder mimarisine giriş olarak verilmiştir. Kelime parçacığı yerleştirmesi ve encoder çıktısı, decoder bloğuna giriş olarak verilmiş ve decoder çıktısı Türkçe metin tabanlı bir görüntü açıklaması olmuştur. Çalışmada, önceden eğitilmiş bir Faster R-CNN kullanılmış ve sadece encoder ve decoder kısımları eğitilmiştir. Encoder, modifiye edilmiş bir transformer modeline dayanmakta olup, decoder bloğu önerilen bir mimariye dayanmaktadır.

Ayrıca, WordPiece tabanlı bir belirteçleme modeli kullanılarak decoder' a beslenen açıklamalar ön işlenmiştir.

Ertuğrul ve ark. 2023 yılında TasvirEt veri kümesini kullanarak VGG-16 ve DenseNet ile görüntülerin özellik çıkarımını, LSTM ile de açıklama oluşturma görevini gerçekleştiren bir çalışma sunmuştur (Ertuğrul & Omurca, 2023). Derin sinir ağları, önceki sonuçları iyileştirerek daha doğru ve anlamlı açıklamalar oluşturmuştur. Unal ve diğerlerinin çalışmasına göre, aynı veri kümesinde BLEU-1 için 0,260 ve BLEU-2 için 0,102 olmak üzere daha yüksek skorlar elde edilmiştir.

Görüntülerin Türkçe alt yazısı üzerine yapılan araştırmaların sınırlı olduğu göz önüne alındığında, daha fazla çalışma ve gelişme için bu alandaki potansiyelin artırılması gerekmektedir. Tablo 1, görüntü Türkçe alt yazısıyla ilgili yapılan bazı çalışmaları içermektedir.

Tablo 1. Türkçe Dilinde Görüntü Alt Yazısı ile İlgili Çalışmalar

Yazar	Araştırma odağı	Veri kümesi	Boyut	Bölme oranı	Model	En Başarılı Model	Blue	Meteor	Rouge	CIDer
Yıldız ve ark., 2023	Türkçe görüntü alt yazılamada derin makine çeviri etkisinin analizi	MS COCO	123.287	92:4:4	LSTM, NLLB	NLLB	0,31	0,11	0,26	0,04
Yıldız ve ark., 2023	TRCaptionNet: Otomatik Türkçe görüntü açıklama için modeli	MS COCO, Flickr 30K	142.287	92:4:4	NLLB deep linguistic model, ViT, ResNet	Önerilen model	0,577	0,254	0,465	0,655
Ersoy ve ark., 2023	ORTPiece: Transformatör kullanan Türkçe tabanlı bir görüntü alt yazılama algoritması	Özel veri seti	5.000	80:10:10	Faster R-CNN, encoder, decoder, Word Piece	Önerilen model	0,334	0,174	0,033	0,553
Ertuğrul & Omurca, 2023	Türkçe açıklamalar oluşturmak	TasvirEt	8.000	80:20	VGG-16, LSTM Dense Net, Word Embedding	VGG-16, Dense Net	0,260			
Golech ve ark., 2022	Türkçe dilinde otomatik görüntü alt yazılama	MS COCO 2014, Flickr8K, Flickr 30K	123.287	70:30	encoder, decoder, meshed memory transfer	Önerilen model	0,72	-	-	-
Sönmez ve ark., 2019	Türkçe dilinde otomatik görüntü alt yazılama	MS COCO 2014	123.287	50:25:25	CNN, RNN	Önerilen model	0,297	0,129	0,272	0,528

Kuyu ve ark., 2018	Türkçe dilinde, altsözcük öğelerini kullanarak görüntü alt yazılama	MS COCO 2014, Flickr30, TasvirEt	123.287	70:30	BPE, LSTM	Önerilen model	0,308	0,147	0,302	0,058
Samet ve ark., 2017	Türkçe açıklamalar içeren yeni bir veri kümesi	MS COCO2014, Flickr8K, TasvirEt, Flickr30k	123.287	70:30	CNN, LSTM	Önerilen model	0,342	0,157	0,322	0,461
Unal ve ark., 2016	TasvirEt: Türkçe açıklamalar içeren yeni bir veri kümesi	Flickr8K	8.091	90:10	Oluşturma ve getirim tabanlı modeller	Önerilen model	0,26	-	-	-

Görüntü alt yazısının kullanım alanları ve bu alandaki çalışmaların incelenmesinin ardından, bu çalışmada daha özgün bir araştırma alanı ele alınmıştır. Özellikle, alt yazıların sınıflandırılması konusu bu bağlamda önemli bir noktayı oluşturmaktadır. Literatür taraması, araştırmaların görüntü alt yazısının sınıflandırılması problemine odaklandığını göstermektedir. Ancak, Türkçe alt yazıların sınıflandırılmasıyla ilgili özgün bir çalışmaya rastlanmamıştır. Bu bağlamda, alt yazı sınıflandırılmasıyla ilgili literatür taraması yapıldığında çok fazla çalışmanın olmadığı görülmüştür. Yıl bağımsız Google Scholar üzerinde "caption classification" anahtar kelimesi ile yapılan taramada, 105 sonuç elde edilmiş ve bu çalışmalar incelendiğinde görüntü Türkçe alt yazı sınıflandırması olmadığı görülmüştür. "caption classification" konusuna yoğunlaşan bazı çalışmalar Tablo 2' de verilmiştir. Bulunan çalışmalar incelendiğinde, örneğin Anjoletto Ferreira ve ark., Ferreira ve ark., Shekhar ve ark. gibi çalışmalar, FOIL-COCO veri kümesini kullanarak, alt yazıları doğru-yanlış olarak iki sınıflı olacak şekilde sınıflandırmışlardır (Anjoletto Ferreira ve ark., 2020; Ferreira ve ark., 2022; Shekhar ve ark., 2017). Bu çalışmaların amacı görüntü ve alt yazı arasındaki uyumsuzluğu bulmaktır, bu nedenle alt yazı sınıflandırması kapsamında başka alana dahil olmaktadır. Konuyla ilgili bulunan diğer çalışmalar ise genelde iki aşamadan oluşmuştur; görüntüden alt yazı oluşturma ve alt yazıları sınıflandırma. Bu çalışmada, Türkçe alt yazı sınıflandırması gerçekleştirilerek, bu alanda Türkçe dilinde literatürdeki boşluğun giderilmesine katkı sağlanmıştır.

Tablo 2. Görüntüden Alt Yazı Oluşturma ve Sınıflandırma Çalışmaları

Yazar	**Araştırma odağı**	**Frame Work**	**Veri kümesi**	**Boyut**	**Öğrenme Türü**	**Model**	**En Başarılı Model**	**Acc**	**Recall**	**F1-Skor**
Bharne & Bhaladhare 2024	Görüntü alt yazı çetimi ve görüntü alt yazı sınıflandırma Sahte profil tespiti.	Python, Pycharm, Windows 10	scamdigger.com datingnmore.com	12.000+	Makine öğrenmesi Derin öğrenme	Otomatik alt yazı için: ResNet, CNN, BLIP Alt yazı sınıflandırma: SVM, Lineat SVM, RF, XTree	XTree	0,91	83,73	85,87
Cao ve ark 2024	Alt yazı sınıflandırması Görsel kültürel farkındalığı sınıflandırmak	4-core Linux system, OpenAI	MaRVL benchmark dataset	429	-	GPT-4V	GPT-4V	-	-	67,4

Yu ve ark. 2021	Görsel açıklama ve sınıflandırma ile sosyal medya görsellerini anlamlandırma ve kategorilendirme.	Tensor Flow	Özel Instagram	400	Derin öğrenme	BERT, Inception V3	BERT	Train: 98 Val: 75	-	-
Kaur & Kiesel, 2020	Resim ve alt yazıları sınıflandırma Alt yazı analizi ile grafik türü sınıflandırması konusu.	Python	1: Science Direct API 2: Özel	96.837	Makine öğrenmesi Derin öğrenme	RF, SVM, CNN	RF	-	-	92,2
Rafi ve ark., 2020	Alt yazıları sınıflandırma Instagram hashtagleri ile LSTM kullanarak alt yazı sınıflandırması ile tepkileri otomatikleştirmek	Tensor Flow	1: Instagram posts 2: Movie Review	11.974	Derin öğrenme	Attention-based LSTM networks	LSTM	1: 73,6 2: 87,7	-	-
Andrearczyk & Müller 2018	Inorganik malzeme bilimi için grafiksel veri analizine dayalı resim ve alt yazı sınıflandırma sistemi geliştirmek	Amazon Mechanical Turk3; Amazon EC2, Tensor; PyTorch7	Özel	16.668	Derin öğrenme	CNN, Mat-CNN	Tüm çok modlu bilgilerle sınıflandırma	-	-	96,1
Kafka, 2017	Resimden alt yazı oluşturma ve alt yazıları sınıflandırma İnsan fiillerinin sınıflandırmak	Python, Tensor Flow	MS COCO içinden seçilen 2 farklı alt küme	1: 533 2: 366	Makine öğrenmesi Derin öğrenme	Encoding (BoW or FLS), Classifier (SVM or CNN), Otomatik alt yazı için: S&T, Alt yazıları sınıflandırma: GT	GT	GT: 94,5 S&T: 59,1	-	-

3. MATERYAL VE METOT

ImCapTR veri kümesi, Türkçe alt yazılardan elde edilmiş olup, literatürdeki veri kümelerine kıyasla daha geniş kapsamlı bir yapıya sahiptir. Bu durum, sınıflandırma modellerinin performansını artırma potansiyeli taşımaktadır. Bu bölümde, TRimCapS sisteminin bileşenleri ve kullanılan yöntemler detaylı biçimde sunulmuştur. Görüntülerin Türkçe alt yazılarını sınıflandırmak amacıyla TasvirEt, Flickr30k ve MS COCO veri kümeleri birleştirilerek ImCapTR adında yeni ve zengin içerikli bir veri kümesi oluşturulmuştur. Şekil 3' te, TRimCapS sisteminin mimarisini açıklayan akış diyagramı yer almaktadır.

3.1. Veri Önişleme

Toplam 114.566 görüntüye ait 588.867 alt yazı kullanılmıştır. Bu alt yazılar temel ön işleme aşamalarından geçirilmiştir. Bu temel aşamalar; tüm harflerin küçültülmesi, şapkalı harflerin normal formlarına çekilmesi (â karakterinin a karakterine dönüştürülmesi gibi) ve kelimeler arasında bulunan birden fazla boşluğun tek boşluğa indirgenmesidir. Daha sonra harfler dışında metinsel olarak herhangi bir anlam ifade etmeyen tüm karakterlerin de çıkarılması işlemi gerçekleştirilmiştir. Bu anlamda, belirli harflerin (aâbcçdefgğhiîjklmnoöprsştuüûvyzwxq) dışındaki tüm karakterler, alt yazılardan kaldırılmıştır. Ardından bir görüntüye ait birden fazla alt yazı, boşlukla art arda eklenerek birleştirilmiştir. Böylece 588.867 adet olan alt yazı sayısı, görüntü sayısı ile eşit hale gelmiştir. Tablo 3' te, ön işleme sürecinden geçirilen veri kümesine ait temel istatistiksel bilgiler sunulmuştur.

3.1.1. Gereksiz ifadelerin tespit edilmesi ve ayıklanması

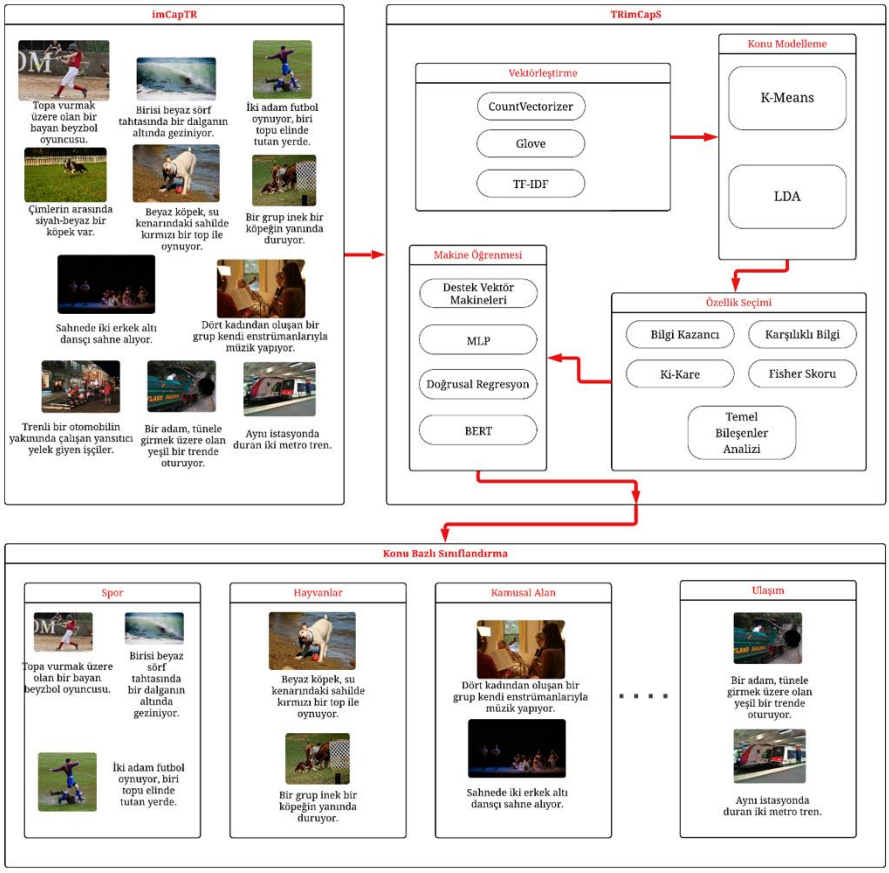
Bir başka önemli adım ise İngilizce 'stopwords' olarak adlandırılan kelimelerin alt yazı metinlerinden çıkarılması işlemidir. Doğal dil işleme alanındaki literatürde sıkça kullanılan NLTK (Natural Language Toolkit) kütüphanesinin Türkçe için sunduğu stopwords listesinin yetersiz olduğu anlaşılmış, bu nedenle el ile yeni bir stopwords listesi hazırlanması gereği ortaya çıkmıştır (Loper & Bird, 2002). Bu liste 'Gereksiz Kelimeler (GK) Listesi' olarak isimlendirilmiştir. Bu doğrultuda, çeşitli ön işlemeler sonrası yeni elde edilen veri kümesindeki her bir kelime ve kelime öbekleri sayılmıştır. Ardından sıklıklarına göre sıralandığı bir liste hazırlanmıştır. Böylece yeterli sayıda GK tespit edilmiştir. GK' lerin tespiti manuel olarak gerçekleştirilmiş ve sonucunda iki farklı liste oluşturulmuştur. İlk liste, toplamda 1226 adet tek kelimeli ifadeler içerirken, ikinci liste ise 686 adet iki kelimeli ifadeler içermektedir. Bu listeler, ileride Türkçe tabanlı doğal dil işleme projelerinde kullanılmak üzere kaydedilmiştir. Daha sonra, bu listeler kullanılarak tüm GK' lerin metinlerden çıkarılmıştır.

Tablo 3. Ön İşleme Sonrası Veri Kümesine Ait İstatistikler

Toplam satır sayısı	114.566
Toplam kelime sayısı	2.811.640
Alt yazılardaki ortalama kelime sayısı	24,54
En az kelime sayısı	7
En çok kelime sayısı	77
Benzersiz kelime sayısı	18.388
Ngram-1	18.367

3.1.2. Zemberek kütüphanesi ile kelime köklerinin oluşturulması

Türkçe, sondan eklemeli ve çekimli bir dil olduğundan, aynı anlama gelebilecek birçok kelime farklı bir görünümde bulunabilmektedir. Örneğin, "kitap" kelimesi ile "kitabı, kitabını, kitabıyla" gibi kelimeler aslında aynı anlamdadır. Bu kelimelere getirilen eklerin çıkarılması ve hepsinin "kitap" kelimesine indirgenmesi gerekir. Alt yazılar içerisinde bunun gibi çokça örneğe rastlamak mümkündür. Bu işlemin yapılarak doğru kelime köklerine inilmesi, kurulacak olan modellerin daha iyi sonuçlar üretmesini sağlayacağından, önem arz etmektedir. Bu amaçla, Zemberek kütüphanesinin ilgili metotları kullanılarak, veri kümesinde bulunan alt yazılardaki her bir kelime, kelime köküyle değiştirilmiştir (Akın & Akın, 2007). Bu aşamadan sonra alt yazılar, sadece kategori ifade eden ve kelime köklerinden oluşan metinler haline gelmiştir.



Şekil 3. TRimCapS Sistem Mimarisi

3.2. Vektörleştirme

3.2.1. CountVectorizer

Metin verilerini sayısal bir forma dönüştürmek için kullanılan bir araçtır. Her belgeyi bir vektör olarak temsil eder ve her kelimenin belgedeki sıklığını içerir. Metin verileri CountVectorizer kullanılarak sayısal matrislere dönüştürülür, böylece makine öğrenimi algoritmalarıyla daha etkili bir şekilde işlenebilir. Örneğin, duygu analizi, metin sınıflandırma ve kümeleme gibi görevler için kullanılabilir (Turki & Roy, 2022).

3.2.2. TF-IDF

TF-IDF (Term Frequency-Inverse Document Frequency), bilgi erişim ve metin madenciliğinde yaygın olarak kullanılan bir ağırlıklandırma yöntemidir. Bir kelimenin bir belge içindeki önemini değerlendirmeye yarar. TF-IDF' nin TF kısmı bir terimin bir belgedeki sıklığını temsil ederken, IDF kısmı bir terimin tüm belge içinde ne kadar

benzersiz veya nadir olduğunu gösterir. TF-IDF, bu iki bileşeni birleştirerek hem bir belgede sık görülen hem de belge genelinde ayırt edici olan kelimeleri etkili bir şekilde vurgulayabilir, böylece metin sınıflandırma ve bilgi alma gibi görevlere yardımcı olur (Robertson, 2004).

3.2.3. GloVe

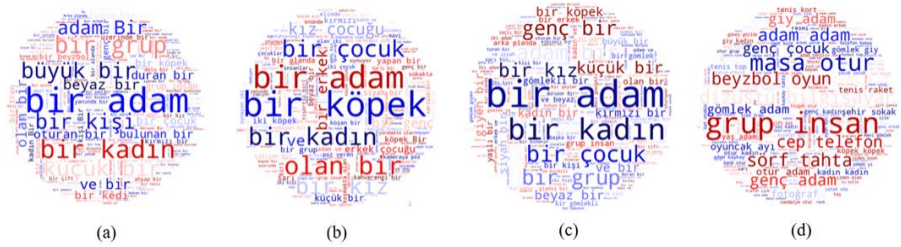
GloVe (Global Vectors for Word Representation), kelime temsillerini öğrenmek için kullanılan bir başka önemli doğal dil işleme yöntemidir. GloVe, kelime vektörlerini oluştururken kelimenin bağlamını ve küresel metin istatistiklerini göz önünde bulundurur, bu da daha tutarlı ve genel geçerli kelime temsilleri sağlar. Bu yöntem, kelime ilişkilerini daha iyi yakalamak için kelime vektörlerini bir araya getirir, böylece anlamsal benzerlikleri ve farklılıkları daha doğru bir şekilde yansıtır (Pennington, Socher ve ark., 2014).

3.2.4. BERT

BERT, çift yönlü bağlamı etkin bir şekilde yakalayan bir yöntemdir. Bu model, dikkat mekanizmalarına dayanan Transformer mimarisi üzerine inşa edilmiştir, bu da tekrarlama ve konvolüsyon ihtiyacını ortadan kaldırır. BERT, aynı anda hem sol hem de sağ bağlamlardan bilgiyi yakalayabilmesini sağlar ve dil anlayışını gerektiren görevlerde etkinliğini artırır. Geleneksel yaklaşımlara kıyasla daha derin ve daha geniş bir dil anlama yeteneği sunar, bu da metin tabanlı uygulamalarda önemli bir ilerleme sağlar (Devlin ve ark., 2018).

3.3. Veri Kümeleri

Bu çalışmada, literatür taraması sonucunda, Türkçe alt yazılara sahip üç veri kümesi kullanılmıştır. Bu veri kümelerindeki görüntü ve alt yazılar birleştirilerek "ImCapTR" adında yeni bir Türkçe alt yazı veri kümesi oluşturulmuştur. Şekil 4' te literatürde kullanılan üç veri kümesi ve yeni oluşturulan ImCapTR veri kümelerine ait kelime bulutları verilmiştir.



Şekil 4. Veri Kümelerine Ait Kelime Bulutu Görüntüleri
(a) MS COCO 2014 (b) TasvirEt (c) Flickr30k (d) ImCapTR

3.3.1. TasvirEt

"TasvirEt" veri kümesi, Flickr8K veri kümesindeki görüntülerin Türkçe alt yazılarının toplanması amacıyla oluşturulmuştur (Unal ve ark., 2016). Bu amaçla, bir sistem geliştirilmiş ve toplanacak olan veriler bu sistem aracılığıyla kitlesel katılım ile toplanmıştır. Veri girişi, herkesin katkıda bulunabileceği bir şekilde halka açık olarak tasarlanmıştır. Kullanıcılar, belirli bir hesaptan giriş yaparak Flickr8K görüntü kümesine ait görüntülere açıklamalar eklemiştir. Veri kümesi 8.000 görüntü ve 16.037 alt yazıdan oluşmaktadır.

3.3.2. Flickr30k

Google çeviri aracının son zamanlarda başarılı sonuçlar vermesi ile görüntü ve İngilizce alt yazı içeren Flickr30k veri kümesindeki alt yazılar Türkçe diline çevrilmiştir (Samet ve ark., 2017). Çeviri için Google Translate API kullanılmıştır. Sonuç olarak 31.783 görüntü ve her bir görüntü için beş alt yazı bilgisi olacak şekilde toplamda 158.915 alt yazı elde edilmiştir.

3.3.3. MS COCO 2014

Microsoft Common Objects in Context (MS COCO) veri kümesi, görüntü alt yazısı oluşturma için yaygın olarak kullanılan bir referans veri kümesidir (Lin ve ark., 2014). MS COCO veri kümesindeki görüntü alt yazıları Yandex Çeviri Uygulama Program Arayüzü (Yandex API) üzerinden Türkçe' ye çevrilmiştir (Sönmez ve ark., 2019). Böylece her bir görüntü için beş alt yazı olacak şekilde 82.783 görüntü için 413.915 alt yazı elde edilmiştir.

3.3.4. ImCapTR

Literatürde Türkçe alt yazı çalışmalarında yaygın olarak kullanılan üç veri kümesi olan TasvirEt, Flickr30k ve MS COCO veri kümeleri birleştirilerek "ImCapTR" adında yeni bir veri kümesi oluşturulmuştur. Veri kümesi birleştirme sürecinde TasvirEt veri kümesindeki görüntülerin Flickr30k veri kümesinde de bulunduğu tespit edilmiş ve tekrar eden, birbirinin aynısı olan görüntüler temizlenerek, her biri yalnızca bir defa yer alacak şekilde düzenlenmiştir. Fakat TasvirEt veri kümesi manuel olarak kullanıcı tarafından oluşturulmuş özgün bir veri kümesi olduğundan Flickr30k veri kümesinden farklı alt yazılara sahiptir. Bu nedenle her bir görüntü için beş alt yazıya sahip olan Flickr30k ile TasvirEt veri kümelerinde ortak olan görüntülerin alt yazıları birleştirilerek 8.000 görüntü için yedi veya sekiz adet alt yazı elde edilmiştir. Veri kümelerindeki alt yazıların birleştirilmesi sonucu 106.566 görüntünün beş, 7963 görüntünün yedi ve 37 görüntünün ise sekiz adet alt yazısı bulunmaktadır. Böylece TrimCapS veri kümesi toplamda 114.566 görüntü ve 588.867 alt yazı içermektedir. Tablo 4' te veri kümelerine ait görüntü ve alt yazı sayıları verilmiştir.

Şekil 5' te veri kümesinde bulunan örnek dört görüntü ve bu görüntülere ait alt yazılar verilmiştir. Veri kümelerinin birleştirilmesi ile elde edilen yeni veri kümesinde her bir görüntü minimum beş farklı alt yazı ile temsil edilmektedir. Ancak veri kümesinin kullanımını daha anlaşılır hale getirmek ve bilgi karmaşasını önlemek amacıyla, bir görüntüye ait olan birden fazla alt yazı, her bir görsel için tek alt yazı olacak şekilde

birleştirilmiştir. Yeni düzenleme ile her görsel için sadece bir alt yazı satırı oluşturulmuştur. Böylece veri kümesi hem görsel hem de metinsel analizlerde daha işlevsel hale getirilmiştir.

Tablo 4. Veri Kümeleri Görüntü ve Alt Yazı Dağılımları

Veri Kümesi Adı	Görüntü Sayısı	Alt yazı Sayısı
TasvirEt	8000	16.037
Flickr30k	31.783	158.915
MS COCO 2014	82.783	413.915
ImCapTR	114.566	588.867

Şekil 5. ImCapTR Veri Kümesi Örnek Görüntü ve Alt Yazılar

3.4. Konu Modelleme

3.4.1. K-Means

K-Means, kümeleme algoritmalarından biridir ve verileri gruplamak için kullanılır. Bu algoritma, önceden belirlenmiş bir sayıda küme (K sayısı) kullanılarak çalışır, veri noktalarını belirli bir uzaklık metriğine göre kümelere ayırmaya çalışır ve her bir kümenin merkezini bulmaya çalışır.

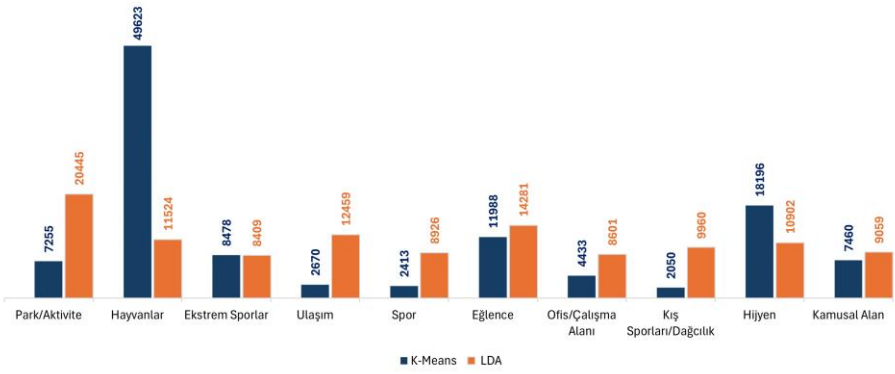
Başlangıçta, K adet küme merkezi rastgele seçilir. Daha sonra her veri noktası, bulunduğu konumdan en yakın küme merkezine atanır. Her kümenin merkezi, o kümeye ait veri noktalarının ortalaması alınarak tekrar güncellenir. Veri noktaları en yakın küme merkezine tekrar atanır ve merkezler tekrar hesaplanır. Durdurma kriterine göre küme merkezi belirleme işlemleri tekrarlanır. Belirli bir duruma ulaşıldığında, algoritma tamamlanır ve veri noktaları kümeleme işlemine göre gruplanmış olur. Veri noktaları ile küme merkezleri arasındaki uzaklığı belirlemek için genellikle Öklidyen metriği kullanılır. Ek olarak K-Means algoritmasının k değerini belirlemek ve

başlangıçta rastgele seçilen küme merkezlerine bağlı olarak farklı sonuçlar elde edilmesi gibi dezavantajları bulunur (Hu ve ark., 2023, Mussabayev ve ark., 2023).

3.4.2. Gizli dirichlet ayrımı

Gizli Dirichlet Ayrımı, belgelerdeki gizli temaları ve bu temaların içerdikleri kelimeler arasındaki ilişkiyi modelleyen bir olasılık modelleme yöntemidir. Bu model, bir belge koleksiyonundaki belgelerin içeriklerini temsil eden gizli temaları keşfetmeye çalışır. Her belge, bu gizli temaların bir karışımını içerir ve her tema belirli bir olasılıkla belgede bulunan kelimeleri üretir. LDA, metin korpusları için üretken bir olasılıksal model olarak kabul edilir ve üç seviyeli hiyerarşik bir Bayes modelidir. Parametre tahmini için varyasyonel yöntemler ve bir Beklenti-Maksimizasyon (Expectation Maximization- EM) algoritması kullanır. LDA, metin madenciliği, konu modelleme ve bilgi çıkarma gibi birçok uygulama alanında kullanılmaktadır. Temel amacı, büyük belge koleksiyonlarından anlam çıkarmak ve bu belgelerdeki gizli yapıları anlamak olan LDA, özellikle konu analizi, belge sınıflandırma ve öneri sistemleri gibi birçok alanda etkili bir araç olarak kabul edilmektedir (Blei ve ark., 2003).

Küme sayısını belirlemek için Elbow yöntemi ve çeşitli testler uygulanarak küme yani kategori sayısı 10 olarak belirlenmiştir (Syakur ve ark., 2018). K-Means ve LDA ile kategorilere ayrılan veri kümesinde her bir kategoriye ait görüntü sayıları Şekil 6' da verilmiştir.



Şekil 6. K-Means ve LDA Sonucu Elde Edilen Kategorilere Ait Görüntü Sayıları

3.5. Özellik Seçimi

Özellik seçimi, veri kümesindeki özelliklerin sayısını azaltma veya en önemli özellikleri seçme sürecidir. Bu, modelin performansını artırır, eğitim süresini kısaltır ve veri kümesini daha iyi anlamayı sağlar. Özellikler arasındaki korelasyon tespit edilerek, gürültülü veya gereksiz özellikler çıkarılır ve modelin genelleme yeteneği artırılır. Metin sınıflandırma işlemlerinde özelliklerin sayısı çok yüksek olduğundan, filtre tabanlı yaklaşımlar yaygın olarak kullanılır. Bilgi kazancı ve ki-kare istatistiği gibi yöntemler, yüksek boyutlu metin verilerinin karmaşıklığını azaltarak hesaplama maliyetini düşürür ve genelleme yeteneğini artırır. Özellik seçme süreci, özellik alt

kümesi oluşturma, değerlendirme, durdurma koşullarının sağlanması ve sınıflandırıcı ile doğrulama adımlarını içerir. Filtreleme, sarım, gömülü ve hibrit olmak üzere dört ana özellik seçme yöntemi bulunmaktadır (Deng ve ark., 2019).

3.5.1. Bilgi kazancı

Bilgi kazancı, filtreleme yöntemlerindendir ve etiketli veriye ihtiyaç duyar. Özelliklerin seçilip seçilmemesi için entropi bilgisini kullanır ve her özniteliği ayrı ayrı değerlendirir. Entropi, veri kümesinin homojen veya heterojen olup olmadığını gösterir; düşük entropi homojenliği, yüksek entropi ise heterojenliği belirtir. Entropi değeri yüksek olan özellikler daha fazla bilgi içerir ve hedef değişkeni daha iyi açıklar, bu yüzden bilgi kazancı yönteminde bu tür özellikler tercih edilir. Ayrıca, bilgi kazancı pozitif ve negatif korelasyonlu özelliklere eşit yaklaşır. Ancak, bu yöntem veri kümesine karşı aşırı uyum gösterme eğiliminde olabilir (Budak, 2018, Forman, 2003). Matematiksel olarak, bir veri kümesinin entropisi (1) formülü ile hesaplanır:

$$Entropi = -\sum_{i=1}^{n} p_i \cdot log_2(p_i) \quad (1)$$

3.5.2. Ki-kare testi

Ki-kare testi, etiketli veri gerektiren ve iki kategorik değişken arasındaki ilişkinin anlamlılığını belirleyen bir filtreleme yöntemidir. Metin belgelerinde, kelime frekanslarının sınıflandırma modeline etkilerini değerlendirmek ve en anlamlı özellikleri seçmek için kullanılır (Peters & Van Voorhis, 1940). Ayrıca, gereksiz özellikleri filtreleyerek aşırı uyumu azaltır ve modelin genelleştirile bilirliğini artırır. Bu yöntem, metin madenciliği ve doğal dil işleme uygulamalarında yaygındır. Matematiksel olarak, ki-kare testinin formülü aşağıdaki (2) gibidir.

$$X^2 = \Sigma\ ((O_i - E_i)^2 / E_i) \quad (2)$$

3.5.3. Karşılıklı bilgi

Karşılıklı bilgi, etiketli veri gerektiren bir filtreleme yöntemidir ve iki rastgele değişken arasındaki ilişkinin gücünü, bir değişkenin değerinin diğerinin belirsizliğini ne kadar azalttığını ölçerek belirler (Kraskov ve ark., 2004). İki değişken X ve Y arasındaki Karşılıklı Bilgi, aşağıdaki formülle (3) ifade edilir:

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y)\ log(p(x,y) / p(x)\ \cdot\ p(y)) \quad (3)$$

Bu formülde I (X; Y) Karşılıklı Bilgiyi temsil eder. p(x,y) X ve Y' nin ortak olasılığını ifade eder. p(x) ve p(y) sırasıyla X ve Y' nin olasılıklarını ifade eder. Bu formül, değişkenler arasındaki bağımlılığı ölçer. Eğer I (X; Y) değeri yüksekse, X ve Y arasındaki bağımlılık güçlüdür; düşükse, bağımlılık zayıftır veya yoktur.

3.5.4. Fisher skoru

Fisher skoru, etiketli veri gerektiren bir filtreleme yöntemidir ve özelliklerin sınıflar arasındaki farklarını göz önüne alarak en önemli özellikleri belirler. İyi bir özellik, sınıflar arasındaki ayrımı artırırken, sınıflar içindeki değişkenliği azaltmalıdır. Metin

sınıflandırmasında, Fisher skoru gürültülü özellikleri belirleyip önemli olmayan kelimeleri eleyerek modelin performansını artırabilir (Gu ve ark., 2012). Fisher Skorun formülü (4) aşağıdaki gibi ifade edilebilir:

$$Fisher\ Score = (Mean\ of\ Between - Class\ Variability)^2\ /\ (Mean\ of\ Within\ -Class\ Variability) \quad (4)$$

3.5.5. Temel bileşenler analizi

Bir veri kümesindeki değişkenler arasındaki ilişkileri anlamak ve veri kümesini daha az sayıda değişkenle temsil etmek için kullanılır. Amacı, değişkenler arasındaki korelasyonu azaltarak veri kümesini yeni bir koordinat sistemi içinde, varyansı maksimize eden bileşenlere dönüştürmektir. Bu yeni değişkenlere "temel bileşenler" denir ve her bileşen, veri kümesinin varyansını azalan sırayla açıklar (Bro & Smilde, 2014). Korelasyon matrisinin ve özdeğer ve özvektörlerin formülü (5) aşağıdaki gibidir:

$$\Sigma\ =\ 1\ /\ (n-1)\ \times\ (X\ -\ \underline{X})^T\ \times\ (X\ -\ \underline{X}) \quad (5)$$

Burada, X veri kümesini, n gözlem sayısını, $\bar{X}$ ise veri kümesinin sütun ortalamasını temsil eder.

$$\Sigma \times \upsilon = \lambda \times \upsilon \quad (6)$$

Bu formülde (6) λ, özdeğerleri ve υ, özvektörleri temsil eder. Özdeğerler büyükten küçüğe sıralandığında, en büyük özdeğerlere karşılık gelen özvektörler temel bileşenleri oluşturur.

4. DENEYSEL SONUÇLAR VE ANALİZ

4.1. Deneysel Ortam

Veri kümelerinin eğitim ve test süreçlerinde modellerin oluşturulması için Intel i7 12700K 3.60 GHz hızında bir işlemciye sahip bilgisayar kullanılmıştır. Hesaplamalar için NVIDIA RTX 4070 Ti model 12 GB kapasiteli bir GPU desteği sağlanmıştır. Kullanılan sistemin RAM desteği de 32 GB ve 6000 MHz hızındadır. Çalışmada her tür depolama ihtiyacını karşılamak amacıyla 2 TB Samsung 990 Pro SSD ve işletim sistemi olarak Windows 11 Pro kullanılmıştır. Tüm kodlamalar Python programlama diliyle, Pycharm 2023.1 editöründe ve Python' ın 3.8 versiyonu ile gerçekleştirilmiştir.

4.2. Deneysel Sonuçlar

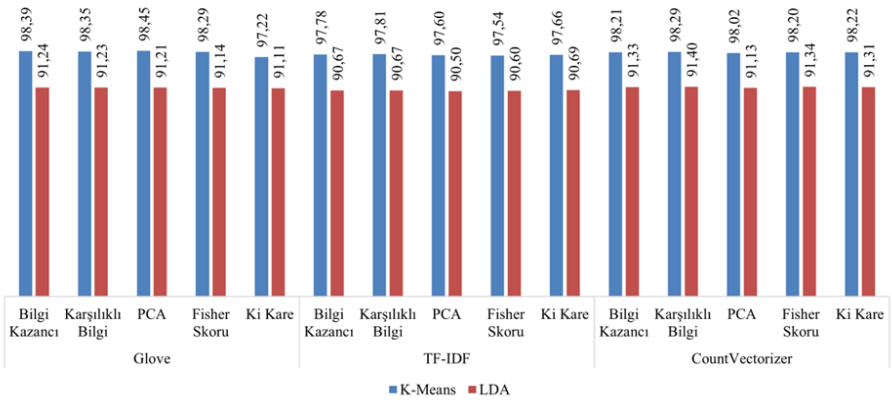
ImCapTR veri kümesi çeşitli vektörleştiriciler kullanılarak vektörleştirilmiş ve K-Means ile LDA yöntemleri kullanılarak kategorilere ayrılmıştır. SVM, LR, MLP ve BERT gibi çeşitli makine öğrenimi modelleriyle birlikte GloVe, TF-IDF ve CountVectorizer gibi çeşitli vektörleştiricilerin ve özellik seçim tekniklerinin performansı incelenmiştir. Sonuçlar, doğruluk yüzdeleri cinsinden Tablo 5' te sunulmuştur. Çalışmada, farklı vektörleştiriciler ve özellik seçim teknikleri (PCA, Bilgi Kazancı, Karşılıklı Bilgi, Fisher Skoru, ki-kare) kullanılarak çeşitli makine öğrenimi

modellerinin performansı incelenmiştir. Sonuçlar, vektörleştirici ve özellik seçim yönteminin model doğruluğunu önemli ölçüde etkilediğini ortaya koymuştur.

Tablo 5. Farklı Vektörleştirici, Özellik Seçim Teknikleri ve Makine Öğrenmesi Modellerinin Doğruluk Değerlerinin Performans Karşılaştırması

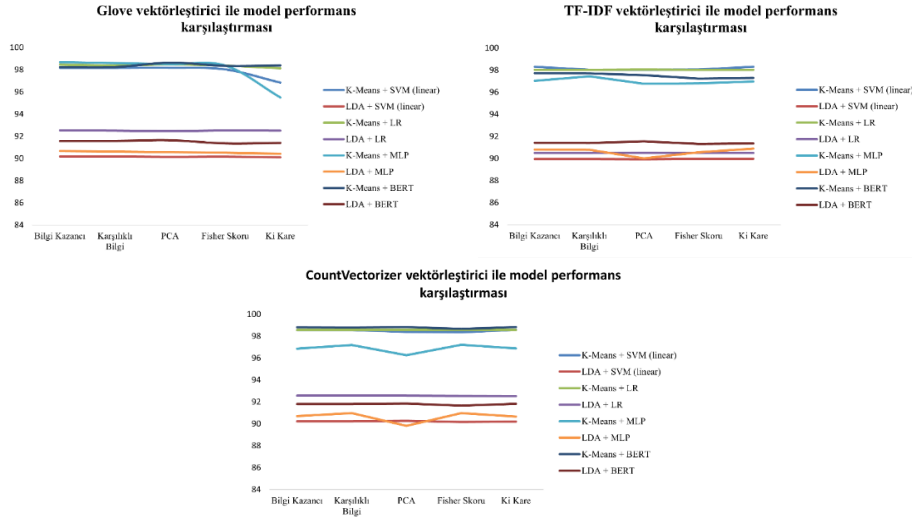
Vektörleştirici	Özellik Seçimi	K-Means + SVM	LDA + SVM	K-Means + LR	LDA + LR	K-Means + MLP	LDA + MLP	K-Means + BERT	LDA + BERT
GloVe	Bilgi Kazancı	98,16	90,19	98,45	92,53	**98,69**	90,67	98,25	91,57
GloVe	Karşılıklı Bilgi	98,16	90,19	98,38	92,51	**98,61**	90,62	98,25	91,58
GloVe	PCA	98,17	90,15	98,47	92,48	98,51	90,58	**98,64**	91,64
GloVe	Fisher Skoru	98,03	90,17	98,37	92,52	**98,39**	90,52	98,36	91,36
GloVe	Ki-kare	96,84	90,1	98,13	92,51	95,5	90,44	**98,4**	91,4
TF-IDF	Bilgi Kazancı	**98,30**	89,95	98,05	90,51	97,03	90,81	97,73	91,42
TF-IDF	Karşılıklı Bilgi	98,03	89,95	**98,05**	90,51	97,44	90,8	97,71	91,4
TF-IDF	PCA	98,03	89,94	**98,06**	90,52	96,77	90,01	97,54	91,54
TF-IDF	Fisher Skoru	**98,06**	89,98	98,05	90,5	96,8	90,57	97,23	91,33
TF-IDF	Ki-kare	**98,31**	89,97	98,05	90,5	96,97	90,9	97,3	91,37
Count Vectorizer	Bilgi Kazancı	98,57	90,23	98,57	92,57	96,87	90,7	**98,81**	91,81
Count Vectorizer	Karşılıklı Bilgi	98,57	90,24	98,57	92,57	97,2	90,97	**98,8**	91,8
Count Vectorizer	PCA	98,4	90,28	98,57	92,57	96,27	89,82	**98,84**	91,84
Count Vectorizer	Fisher Skoru	98,36	90,18	98,57	92,56	97,22	90,97	**98,66**	91,66
Count Vectorizer	Ki-kare	98,59	90,22	98,57	92,54	96,89	90,66	**98,83**	91,83

GloVe vektörleştirici kullanılarak çeşitli özellik seçimi yöntemleriyle elde edilen doğruluk oranları incelendiğinde, en yüksek değerler bilgi kazancı yöntemiyle K-Means + MLP (%98,69) ve PCA yöntemiyle K-Means + BERT (%98,64) modellerinde elde edilmiştir. GloVe için ki-kare yöntemi genellikle daha düşük performans sergilerken, PCA diğer yöntemlere göre yüksek performans gösteren özellik seçimi yöntemi olmuştur. TF-IDF vektörleştirici kullanılarak farklı özellik seçimi yöntemleriyle elde edilen doğruluk oranları incelendiğinde, en yüksek doğruluk oranlarının bilgi kazancı (%98,3) ve ki-kare (%98,31) yöntemleriyle elde edildiği görülmüştür. K-Means + LR ve K-Means + SVM kombinasyonu, özellik seçimi yöntemleriyle genellikle yüksek doğruluk oranları sağlamıştır.



Şekil 7. K-Means ve LDA Kategorileştirme Yöntemleri ile Elde Edilen Doğruluk Oranlarının Karşılaştırılması

CountVectorizer vektörleştirici ile çeşitli özellik seçimi yöntemleri kullanılarak yüksek ve tutarlı doğruluk oranları elde edilmiştir. En yüksek doğruluk oranı, K-Means + BERT (%98,84) modeli ile PCA kullanılarak elde edilmiştir. CountVectorizer için tüm özellik seçim yöntemleri genel olarak benzer performans sergilemiştir, bu da CountVectorizer' ın özellik seçim yönteminden bağımsız olarak tutarlı bir performans sunduğunu göstermektedir. Şekil 7' de K-Means ve LDA ile kategorize edilmiş verilerle elde edilen modellerin başarı oranlarının ortalamasını içeren grafik sunulmuştur. Her iki şekil de incelendiğinde K-Means yöntemi ile daha yüksek doğruluk değerlerinin elde edildiği görülmüştür. Şekil 8' de ise kullanılan vektörleştirme yöntemleri kullanılarak oluşturulan modellerin doğruluk oranlarının karşılaştırıldığı grafikler verilmiştir.



Şekil 8. GloVe, TF-IDF ve Countvectorizer ile Model Performanslarının Karşılaştırılması

Boosting, topluluğun genel kaybını en aza indirmek amacıyla eklemeli modellerin ardışık olarak eğitilmesini içerir. Bu yöntem, önceki modellerin hatalarını düzelterek giriş verilerinden karmaşık kalıpları öğrenmeyi sağlar (Chen & Guestrin, 2016). Gözlemlerin ağırlıklarını yinelemeli olarak ayarlayarak ve doğru sınıflandırılması zor olan alanlara odaklanarak modellerin genel performansını iyileştirmeyi amaçlamaktadır. Extreme Gradient Boosting (XGBoost) ise, ikinci dereceden Taylor açılımı kullanarak kayıp fonksiyonunun doğruluğunu artırıp geleneksel Gradyan Artırma Karar Ağacı (GAKA) modelinin performansını geliştiren güçlü bir boosting algoritmasıdır. ImCapTR veri kümesi üzerinde bu yöntemler test edilmiş ve Gradient Boosting için %98,90, XGBoost için %98,92 doğruluk değerleri elde edilmiştir.

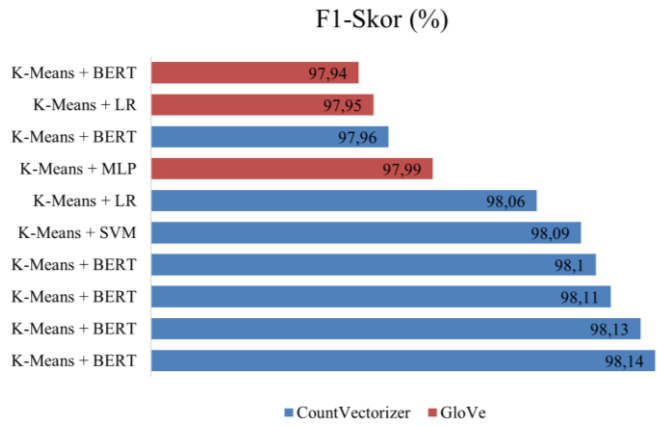
Tablo 6. Farklı Vektörleştirici, Özellik Seçim Teknikleri ve Makine Öğrenmesi Modellerinin F1-Skor Değerlerinin Performans Karşılaştırması

Vektörleştirici	Özellik Seçimi	K-Means + SVM	LDA + SVM	K-Means + LR	LDA + LR	K-Means + MLP	LDA + MLP	K-Means + BERT	LDA + BERT
GloVe	Bilgi Kazancı	97,46	89,49	97,95	92,03	**97,99**	89,97	97,55	90,87
GloVe	Karşılıklı Bilgi	97,45	89,49	97,68	92,01	**97,91**	89,92	97,55	90,88
GloVe	PCA	97,47	89,49	97,89	91,82	97,89	90,09	**97,94**	90,76
GloVe	Fisher Skoru	97,46	89,47	**97,91**	92,02	97,7	90,02	97,86	90,86
GloVe	Ki-kare	97,61	89,45	97,75	91,84	97,84	89,76	**97,9**	90,92
TF-IDF	Bilgi Kazancı	**97,6**	89,45	97,36	89,81	96,53	90,01	97,03	90,92
TF-IDF	Karşılıklı Bilgi	**97,6**	89,3	97,37	90,02	96,74	90,34	97,01	90,7

TF-IDF	PCA	**97,81**	89,54	97,56	89,82	96,07	89,51	96,81	90,9
TF-IDF	Fisher Skoru	**97,8**	89,49	97,35	89,87	96,4	89,87	96,73	90,63
TF-IDF	Ki-kare	**97,81**	89,52	97,36	89,86	96,19	90,16	97,7	90,92
Count Vectorizer	Bilgi Kazancı	97,68	89,75	97,87	91,84	96,5	90,27	**98,11**	91,11
Count Vectorizer	Karşılıklı Bilgi	97,67	89,74	97,87	92,07	96,94	90,22	**98,1**	91,1
Count Vectorizer	PCA	97,7	89,58	97,87	92,07	95,57	89,12	**98,14**	91,14
Count Vectorizer	Fisher Skoru	97,66	89,73	**98,06**	92,06	96,72	90,27	97,96	91,1
Count Vectorizer	Ki-kare	98,09	89,52	97,87	91,84	96,19	90,16	**98,13**	91,13

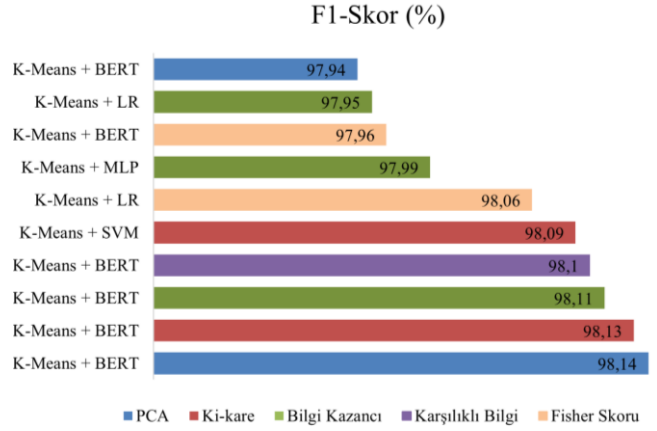
Sınıflandırma problemlerinde model başarısının değerlendirilmesinde doğruluk metriği yaygın olarak kullanılmakla birlikte, bu ölçüt sınıf dağılımının dengesiz olduğu veri kümelerinde modelin gerçek performansını yansıtmakta yetersiz kalabilmektedir. Doğruluk, modelin doğru tahmin ettiği örneklerin toplam örnek sayısına oranı olarak tanımlansa da özellikle azınlık sınıfların göz ardı edilmesi durumunda yüksek doğruluk değerleri elde etmek mümkündür. Bu bağlamda, yalnızca doğruluk metriğine dayanarak yapılan değerlendirmeler, modelin sınıflar arası ayırt edici gücünü ve hataya karşı duyarlılığını yeterince ortaya koyamamaktadır. Bu çalışmada, model performansını daha bütüncül bir biçimde değerlendirebilmek amacıyla doğruluk metriğine ek olarak F1-skoru da hesaplanmıştır. F1-skoru, kesinlik ve duyarlılık ölçütlerinin harmonik ortalaması olarak tanımlanmakta olup, modelin hem doğru pozitif tahmin üretme kabiliyetini hem de pozitif sınıfları eksiksiz yakalayabilme yetisini birlikte değerlendirmektedir. Bu özelliği sayesinde F1-skoru, sınıflar arası dengesizlik içeren veri kümelerinde daha hassas ve güvenilir bir performans ölçütü olarak öne çıkmaktadır.

Tablo 5 ve Tablo 6 incelendiğinde, bazı modellerin yüksek doğruluk değerleri elde ettiği, ancak buna rağmen F1-skorlarının daha düşük seviyelerde kaldığı gözlemlenmektedir. Örneğin, TF-IDF vektörleştirici ve PCA özellik seçimi ile K-Means + LR modeli %98,06 doğruluk göstermiş, karşılık gelen F1-skoru %97,56 olarak hesaplanmıştır. Benzer şekilde, GloVe vektörleştirici ve Ki-kare yöntemi ile elde edilen K-Means + LR modeli %98,13 doğruluğa sahipken, F1-skoru %97,75' te kalmıştır. Bu durum, bazı modellerin ağırlıklı olarak çoğunluk sınıflara odaklandığını, azınlık sınıflarda ise başarısız tahminler üretebildiğini ortaya koymaktadır. Buna karşılık, CountVectorizer kullanılarak geliştirilen ve K-Means + BERT modeliyle eğitilen yapı, hem yüksek doğruluk (%98,84) hem de yüksek F1-skoru (%98,14) elde ederek daha dengeli bir performans sergilemiştir.



Şekil 9. En İyi 10 F1-Skor Değerine Göre Vektörleştiricilerin Karşılaştırılması

Şekil 9 ve Şekil 10 incelendiğinde, en yüksek F1-skoruna sahip ilk 10 modelin performansları vektörleştirici türleri (Şekil 9) ve özellik seçimi yöntemleri (Şekil 10) açısından karşılaştırılmıştır. Şekil 9' da görüldüğü üzere, CountVectorizer kullanılarak elde edilen modeller genellikle daha yüksek F1-skorları ile öne çıkmıştır. Özellikle "K-Means + BERT" modeli, farklı özellik seçim teknikleriyle birlikte her iki vektörleştiriciyle test edilmiş; ancak CountVectorizer ile 98,14 F1-skoruna ulaşarak en yüksek başarıyı göstermiştir. GloVe vektörleştiricisi ise "K-Means + MLP" ve "K-Means + LR" gibi modellerde rekabetçi sonuçlar üretmişse de genel sıralamada geride kalmıştır. Şekil 10' da özellik seçimi yöntemlerinin etkisi değerlendirilmiş olup PCA, Ki-kare, Bilgi Kazancı, Karşılıklı Bilgi ve Fisher Skoru yöntemleri karşılaştırılmıştır. En yüksek başarıyı sağlayan özellik seçim tekniği PCA olurken, bu yöntemi sırasıyla Ki-kare ve Bilgi Kazancı takip etmiştir. Bununla birlikte, farklı modellerin aynı vektörleştirici altında çeşitli özellik seçimleri ile test edilmesi, bu bileşenlerin nihai başarı üzerindeki etkisini açıkça ortaya koymaktadır.



Şekil 10. En İyi 10 F1-Skor Değerine Göre Özellik Seçimi Yönteminin Karşılaştırılması

Çalışma kapsamında test edilen modeller arasında, genel olarak CountVectorizer tabanlı temsillerin, özellikle PCA ve Ki-kare gibi özellik seçimi yöntemleriyle birleştirildiğinde, en yüksek doğruluk ve F1-skoru değerlerine ulaştığı gözlemlenmiştir. Bu kombinasyonlar içerisinde K-Means + BERT modeli, %98' in üzerindeki başarı oranlarıyla öne çıkmıştır. GloVe ve TF-IDF gibi gömme temelli yöntemler ise bazı durumlarda yüksek doğruluk sağlasa da F1-skorlarında daha dalgalı bir performans sergilemiştir. Modeller düzeyinde ise, K-Means + BERT, K-Means + MLP ve K-Means + SVM yapıları genel olarak üstün performans göstermiştir. LDA tabanlı modeller diğerlerine göre daha düşük başarı elde etmiştir. Bu sonuçlar, klasik vektörleştirme ve özellik seçimi yöntemlerinin, derin öğrenme tabanlı sınıflayıcılarla etkili biçimde entegre edilebildiğini göstermektedir.

5. SONUÇLAR VE TARTIŞMA

Teknolojinin hızla ilerlemesiyle sosyal platformlar aracılığıyla video ve resim tabanlı iletişim büyük ölçüde artmıştır. Bu iletişimde görüntü altyazıları önemli bir rol oynamakta ve çeşitli kullanım amaçlarıyla dikkat çekmektedir. Ancak, Türkçe dilindeki görüntü alt yazı veri kümelerinin yetersizliği, bu alandaki çalışmaların önündeki büyük bir engeldir. MS COCO ve Flickr gibi uluslararası veri kümeleri, geniş bir dil yelpazesine sahip olmalarına rağmen, Türkçe' nin semantik yapısının diğer dillere göre farklılıkları, mevcut veri kümelerinin Türkçe problemler üzerinde etkili bir şekilde kullanılmasını zorlaştırmaktadır. Bu nedenle, Türkçe alt yazı veri kümelerinin oluşturulması, yapay zekâ modellerinin Türkçe dilinde daha etkili ve duyarlı olmasını sağlamak adına büyük öneme sahiptir. Ayrıca, görüntü alt yazılarının sınıflandırılması, bu alanda yapılan çalışmalara katkı sağlamaktadır. Ancak literatür taraması sonucu, az miktarda Türkçe olmayan alt yazı sınıflandırma çalışması olmasına rağmen, görüntü Türkçe alt yazı sınıflandırılması üzerine yapılmış bir çalışmaya rastlanmamıştır.

Bu doğrultuda gerçekleştirilen bu çalışmada, TasvirEt, Flickr30k ve MS COCO veri kümeleri birleştirilerek, ImCapTR adında bir veri kümesi oluşturulmuştur. Bu veri kümesi, 114.566 görüntü ve 588.867 Türkçe alt yazı içermektedir. TF-IDF, CountVectorizer ve Glove vektörleştiriciler kullanılarak metinler sayısal hale getirilmiştir. 1-gram değeriyle elde edilen vektörleştirme boyutunun 18.367 olduğu tespit edilmiştir. Ayrıca, bilgi kazancı, karşılıklı bilgi, Fisher skor, PCA ve ki-kare olmak üzere 5 farklı özellik seçim yöntemi de kullanılarak sonuçlar zenginleştirilmiştir. LDA ve K-Means algoritmalarıyla kümeleme yapılarak elde edilen sonuçlar çeşitli makine öğrenme yöntemleriyle değerlendirilmiştir. SVM (linear), LR, MLP ve BERT yöntemleri kullanılarak elde edilen sonuçlar incelenmiştir. Bu analizler sonucunda, CountVectorizer + K-Means + BERT kombinasyonunun özellik seçiminden bağımsız olarak yüksek performans metrikleri elde ettiği görülmüştür. Genel olarak PCA ve bilgi kazancı yöntemleri ile daha yüksek doğruluğa ulaşılmıştır. Bu çalışmada, daha önce raporlanmamış birleştirilmiş veri kümesi üzerinde çeşitli algoritmalar uygulanmış ve elde edilen deneysel sonuçlar paylaşılmıştır. Türkçedeki çalışmalara ek olarak, bu kapsamlı veri kümesi üzerinde gerçekleştirilen analizler ile literatüre katkı sağlanmıştır.

Gelecek çalışmalarda, görüntülerin Türkçe alt yazıları ile ilgili daha büyük bir veri kümesi oluşturmak hedeflenmektedir. Ayrıca, sınıflandırma aşamasında daha karmaşık derin öğrenme modellerini kullanarak başarı oranlarının artırılması planlanmaktadır. Bellek tüketimi ve eğitim süreleri gibi önemli faktörleri de göz önünde bulundurarak,

hangi modellerin gerçekleştirilebilir olduğunun kapsamlı bir şekilde incelenmesi hedeflenmektedir. Bu sayede, görüntü alt yazı sınıflandırması alanında daha etkili ve verimli çözümler geliştirmek amaçlanmaktadır.

Yazarların Katkısı
Yazarlar, çalışmaya eşit şekilde katkı sunmuşlardır.

Çıkar Çatışması Beyanı
Yazarlar arasında herhangi bir çıkar çatışması bulunmamaktadır.

Araştırma ve Yayın Etiği Beyanı
Bu çalışma Marmara Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimi tarafından 11558 nolu proje ve 11635 nolu proje kapsamında desteklenmiştir.

KAYNAKÇA

Akın, A. A., & Akın, M. D. (2007). Zemberek, an open source NLP framework for Turkic languages. *Structure*, *10*(2007), 1–5.

Andrearczyk, V., & Müller, H. (2018). Deep multimodal classification of image types in biomedical journal figures. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction: 9th International Conference of the CLEF Association, CLEF 2018* (3-14). Springer.

Anjoletto Ferreira, L., De Rizzo Meneghetti, D., & Santos, P. E. (2020). CAPTION: Correction by analyses, POS-tagging and interpretation of objects using only nouns. *arXiv preprint arXiv:2010.00839*.

Bharne, S., & Bhaladhare, P. (2024). Enhancing user profile authenticity through automatic image caption generation using a bootstrapping language–image pre-training model. *Engineering Proceedings*, *59*(1), 182.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.

Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Analytical Methods*, *6*(9), 2812–2831.

Budak, H. (2018). Özellik seçim yöntemleri ve yeni bir yaklaşım. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, *22*, 21–31.

Cao, Y., Li, W., Li, J., Yuan, Y., & Hershcovich, D. (2024). Exploring visual culture awareness in GPT-4V: A comprehensive probing. *arXiv preprint arXiv:2402.06015*.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (785–794).

Deng, X., Li, Y., Weng, J., & Zhang, J. (2019). Feature selection for text classification: A review. *Multimedia Tools and Applications*, *78*(3), 3797–3816.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Ersoy, A., Yıldız, O. T., & Özer, S. (2023). ORTPiece: An ORT-based Turkish image captioning network based on transformers and WordPiece. In *2023 31st Signal Processing and Communications Applications Conference (SIU)* (1-4).

Ertuğrul, M. A. C., & Omurca, S. İ. (2023). Generating image captions using deep neural networks. In *2023 8th International Conference on Computer Science and Engineering (UBMK)* (271–275).

Ferreira, L. A., De Rizzo Meneghetti, D., Lopes, M., & Santos, P. E. (2022). CAPTION: Caption analysis with proposed terms, image of objects, and natural language processing. *SN Computer Science*, *3*(5), 390.

Forman, G. (2003). An extensive empirical study of feature selection metrics for text classification. *Journal of Machine Learning Research*, *3*(Mar), 1289–1305.

Gaspar, A., & Alexandre, L. A. (2019). A multimodal approach to image sentiment analysis. In *Intelligent Data Engineering and Automated Learning – IDEAL 2019: 20th International Conference* (302–309). Springer.

Golech, S. B., Karacan, S. B., Sönmez, E. B., & Ayral, H. (2022). A complete human verified Turkish caption dataset for MS COCO and performance evaluation with well-known image caption models trained against it. In *2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)* (1–6).

Gu, Q., Li, Z., & Han, J. (2012). Generalized Fisher score for feature selection. *arXiv preprint arXiv:1202.3725*.

Hu, H., Liu, J., Zhang, X., & Fang, M. (2023). An effective and adaptable K-means algorithm for big data cluster analysis. *Pattern Recognition*, *139*, 109404.

Kafka, A. (2017). Caption aided action recognition using single images. *Lehigh University*

Kaur, P., & Kiesel, D. (2020). Combining image and caption analysis for classifying charts in biodiversity texts. In *VISIGRAPP (3: IVAPP)* (157–168).

Kraskov, A., Stögbauer, H., & Grassberger, P. (2004). Estimating mutual information. *Physical Review E*, *69*(6), 066138.

Kuyu, M., Erdem, A., & Erdem, E. (2018). Image captioning in Turkish with subword units. In *2018 26th Signal Processing and Communications Applications Conference (SIU)* (1–4).

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference* (740–755). Springer.

Loper, E., & Bird, S. (2002). NLTK: The natural language toolkit. *arXiv preprint cs/0205028*.

Mussabayev, R., Mladenovic, N., Jarboui, B., & Mussabayev, R. (2023). How to use K-means for big data clustering? *Pattern Recognition*, *137*, 109269.

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (1532–1543).

Peters, C. C., & Van Voorhis, W. R. (1940). Chi square.

Rafi, A. M., Rana, S., Kaur, R., Wu, Q. J., & Zadeh, P. M. (2020). Understanding global reaction to the recent outbreaks of COVID-19: Insights from Instagram data analysis. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (3413–3420).

Robertson, S. (2004). Understanding inverse document frequency: On theoretical arguments for IDF. *Journal of Documentation*, *60*(5), 503–520.

Samet, N., Hiçsönmez, S., Duygulu, P., & Akbaş, E. (2017). Görüntü altyazılama için otomatik tercümeyle eğitim kümesi oluşturulabilir mi? In *25th Signal Processing and Communications Applications Conference (SIU)*.

Shekhar, R., Pezzelle, S., Klimovich, Y., Herbelot, A., Nabi, M., Sangineto, E., & Bernardi, R. (2017). FOIL it! Find one mismatch between image and language caption. *arXiv preprint arXiv:1705.01359*.

Sönmez, E. B., Yıldız, T., Yılmaz, B. D., & Demir, A. E. (2019). Türkçe dilinde görüntü altyazısı: Veritabanı ve model. *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, *35*(4), 2089–2100.

Syakur, M., Khotimah, B. K., Rochman, E., & Satoto, B. D. (2018). Integration K-means clustering method and elbow method for identification of the best customer profile cluster. In *IOP Conference Series: Materials Science and Engineering* (Vol. 336, 012017).

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (1–9).

Turki, T., & Roy, S. S. (2022). Novel hate speech detection using word cloud visualization and ensemble learning coupled with count vectorizer. *Applied Sciences*, *12*(13), 6611.

Ünal, M. E., Çıtamak, B., Yağcıoğlu, S., Erdem, A., Erdem, E., Cinbis, N. I., & Çakıcı, R. (2016). Tasviret: A benchmark dataset for automatic Turkish description generation from images. In *2016 24th Signal Processing and Communication Application Conference (SIU)* (1977–1980).

Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2015). Show and tell: A neural image caption generator. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (3156–3164).

Yıldız, S., Memiş, A., & Çarlı, S. (2023). Automatic Turkish image captioning: The impact of deep machine translation. In *2023 8th International Conference on Computer Science and Engineering (UBMK)* (414–419).

Yıldız, S., Memiş, A., & Varlı, S. (2023). TRCaptionNet: A novel and accurate deep Turkish image captioning model with vision transformer based image encoders and deep linguistic text decoders. *Turkish Journal of Electrical Engineering and Computer Sciences*, *31*(6), 1079–1098.

Yu, X., Ahn, Y., & Jeong, J. (2021). High-level image classification by synergizing image captioning with BERT. In *2021 International Conference on Information and Communication Technology Convergence (ICTC)* (1686–1690).