



YouTube trend büyük veri kümelerinden ülkeler arası kalıcı etiketlerin keşfi

Cross-country persistent tags discovery from YouTube trending video big dataset

Yeşim Dokuz^{1,*} 

¹ Niğde Ömer Halisdemir Üniversitesi, Bilgisayar Mühendisliği Bölümü, 51240, Niğde Türkiye

Öz

YouTube kolay kullanılan arayüzü ve büyük miktarda kullanıcı sayısı ile video paylaşım sosyal medya platformları arasında birinci video paylaşım platformudur. YouTube video veri kümelerinin büyük veri doğasından dolayı bu veri kümelerinin analizi ve bilgi çıkarımı, araştırmacılar ve kurum yöneticilerine YouTube kullanıcılarının sosyal eğilimleri hakkında fikir vermektedir. Ancak, YouTube büyük verilerinin analizi, görüntü ve ses işleme uygulamalarının zorluğu, semantik analiz metotlarını düzensiz YouTube içeriklerine uygulamanın zorluğu ve YouTube video veri kümelerinin büyük veri özelliği nedeniyle zordur. Literatürdeki çalışmalar video tavsiye sistemleri, YouTube yorumlarından semantik analizler ve trend video analizleri üzerine odaklanmaktadır. Bu çalışmada, üç ülkeye ait YouTube trend video büyük verisi (Amerika Birleşik Devletleri, Kanada ve İngiltere) kullanılarak ülkeler arası kalıcı etiketlerin keşfi için yeni bir metot ve algoritma önerilmiştir. Keşfedilen ülkeler arası kalıcı etiketler, bazı YouTube video etiketlerinin küresel olarak kullanıldığı, ancak bazı etiketlerin ise yalnız bir ülkede kullanıldığını göstermektedir.

Anahtar kelimeler: YouTube trend video analizi, Ülkeler arası kalıcı etiketler keşfi, Video etiket analizi, YouTube büyük veri analizi

1 Giriş

Bilgi teknolojileri ve web uygulamalarındaki son gelişmelerle birlikte, sosyal medya platformları kullanıcılar arasında büyük önem kazanmıştır. Bu platformlar arasında video içerik platformları önemli bir paya sahiptir. YouTube ise kolay kullanılan arayüzü ve büyük miktarda kullanıcı sayısı ile video paylaşım sosyal medya platformları arasında birinci video platformudur [1].

YouTube video veri kümelerinin analizi ve bilgi çıkarımı, araştırmacılar ve kurum yöneticilerine YouTube kullanıcılarının sosyal eğilimleri hakkında fikir vermektedir [2-4]. Özellikle, video içerikleri ve etiketleri videoyu yükleyen kullanıcıların düşünceleri ve fikirleri hakkında önemli bilgiler sağlamaktadır. Kullanıcılara yeni video tavsiye sistemleri de YouTube video verilerinin analiziyle mümkün olmaktadır [5-8]. Nefret ve şiddet, kriminal içerikler ve diğer başlıklar YouTube video içeriklerinin analiziyle keşfedilebilmektedir [9-11].

Abstract

YouTube is the primary video content platform among video sharing social media platforms with its easy-to-use interface and huge number of users. Due to the big data nature of YouTube video datasets, analyzing and extracting knowledge from these datasets would provide insights into researchers and government directors on social orientation and tendency YouTube users. However, analyzing YouTube big datasets is challenging due to the difficulty of image and speech processing applications, the hardness of utilizing semantic analysis methods on irregular YouTube contents, and big data nature of YouTube video datasets. Literature studies focus on video recommendation systems, semantic analysis on YouTube comments and trending video analysis. In this study, a new method and an algorithm are proposed to discover cross-country persistent tags over YouTube trending video big dataset for three countries (United States of America, Canada, and Great Britain). The discovered cross-country persistent tags show that some YouTube video tags are globally utilized on videos, while some certain tags are utilized for only one country.

Keywords: YouTube trend video analysis, Cross-country persistent tags discovery, Video tag analysis, YouTube big data analytics

Ancak, YouTube videolarının analizi ve işlenmesinin çeşitli zorlukları bulunmaktadır. Birincisi, YouTube video veri kümeleri kullanılarak bilgi keşfi, görüntü ve ses işleme uygulamalarının zorlukları nedeniyle zordur. İkincisi, YouTube videolarındaki metinsel içeriklerden semantik analiz metotlarıyla bilgi keşfi düzensiz içerikler bulunması nedeniyle zordur. Üçüncüsü, YouTube video veri kümeleri doğası gereği büyük veri kaynaklarıdır ve yeni analiz metotları geliştirilmesi gerekmektedir.

Trend video analizi, YouTube trend videolarının analizini hedefleyen YouTube büyük veri analitiğinin önemli bir alanıdır. YouTube, her ülkedeki kullanıcılarının video izleme tercihlerine göre ülke tabanlı trend video listeleri üretmektedir. YouTube trend video büyük verisinin analizi, ülke bazında YouTube kullanıcılarının eğilimlerini ortaya çıkarmakta ve ülke bazında önemli videoların analizi, küresel önemli başlıkların keşfi ve ülkeler bazında trend etiketlerin keşfi başta olmak üzere çok çeşitli konularda fayda sağlamaktadır.

* Sorumlu yazar / Corresponding author, e-posta / e-mail: ytorun@ohu.edu.tr (Y. Dokuz)

Geliş / Received: 20.07.2023 Kabul / Accepted: 27.09.2023 Yayınlanma / Published: 15.10.2023

doi: 10.28948/ngumuh.1330386

Literatürde, YouTube trend video büyük veri kümelerinin analizini hedefleyen çeşitli çalışmalar bulunmaktadır. Figueiredo vd. [12] YouTube videolarının yüklendikleri andan itibaren popülerite gelişimlerini karakterize etmek için gelişim örüntülerini analiz etmişlerdir. Chelaru vd. [13] beğenme, beğenmeme ve yorum sayısı gibi sosyal özelliklere dayalı olarak YouTube videolarının sıralanmasını yapmışlardır. Brodersen vd. [14] YouTube videolarını coğrafi olarak analiz ederek videoların popüleritesiyle lokasyonları arasında bir bağlantı olup olmadığını araştırmışlardır. Krishna vd. [15] YouTube videolarındaki trendler, mevsimsellik ve tahminleri keşfetmek için duygu analizi metotlarını kullanarak YouTube videolarındaki yorumları analiz etmişlerdir. Hoiles vd. [16] meta özelliklere ve sosyal dinamiklere dayalı olarak YouTube videolarının popüleritesini tahmin etmek için makine öğrenmesi algoritmalarını kullanmışlardır. Chen ve Chang [17] YouTube videolarının popüleritesinin erken tahmini için gözetimli öğrenme ve topluluk modeli tabanlı iki aşamalı bir çerçeve önermişlerdir. Gajanayake ve Sandanayake [18] oyun videolarındaki yorumları duygu analizi metotları ile analiz ederek videoları daha popüler hale getirmek için tavsiye sistemi geliştirmişlerdir. Dokuz [19] YouTube trend videolarına ait bilgileri kullanarak popüler ve kalıcı etiketlerin keşfi için bir yöntem ve algoritma önermiştir.

Bu çalışmada, YouTube trend video büyük verisi kullanılarak birden fazla ülke üzerinde ülkeler arası kalıcı etiketler keşfedilmiştir. Aynı dili kullanmaları ve ülkeler arası kalıcı etiketlerin keşfi bu sayede mümkün olacağı için Amerika Birleşik Devletleri, Kanada ve İngiltere olmak üzere üç ülke seçilmiştir. PPTagD algoritması her bir ülke üzerinde uygulanarak ülkeler için popüler ve kalıcı etiketlerin keşfi yapılmıştır. Daha sonra yeni bir yöntem ve algoritma kullanılarak ülkeler arası kalıcı etiketler keşfedilmiştir. Bu yöntemde, üç adet eşik değeri önerilmiş ve bu eşik değerleri kullanarak ülkeler arası kalıcı etiketleri keşfeden bir yöntem geliştirilmiştir. Keşfedilen ülkeler arası kalıcı etiketler bazı YouTube video etiketlerinin küresel olarak kullanıldığı ancak bazı etiketlerin ise yalnız bir ülkede kullanıldığı görülmüştür.

Bu çalışmanın devam eden kısımları şu şekilde organize edilmiştir. Bölüm 2’de YouTube trend video büyük veri kümesi, temel tanımlar ve önerilen algoritma sunulmuştur. Bölüm 3’te önerilen algoritmanın sonuçları sunulmuş ve tartışılmıştır. Bölüm 4’te ise sonuçlar ve gelecekteki çalışmalar sunulmuştur.

2 Materyal ve metot

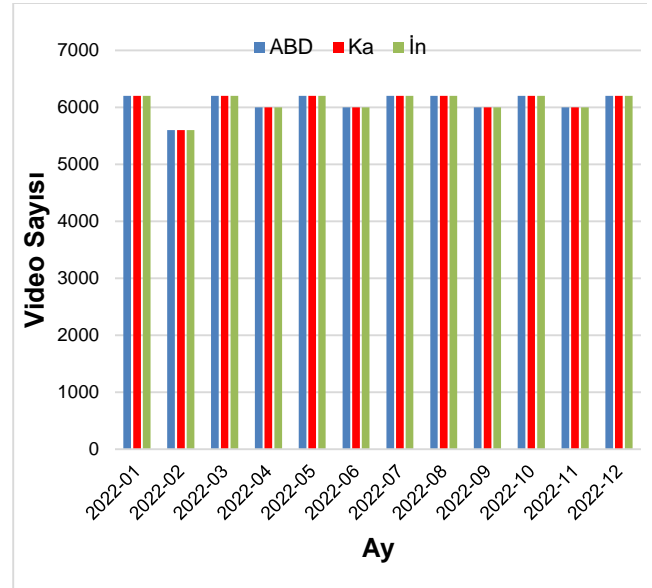
Bu bölümde öncelikle YouTube trend video büyük veri kümesi tanımlanmıştır. Daha sonra, temel tanımlar ve önerilen algoritma sunulmuştur.

2.1 YouTube trend video büyük veri kümesi

Bu çalışmada, 12 Ağustos 2020 tarihinden itibaren günlük trend videoları içeren YouTube trend video büyük veri kümesi kullanılmıştır [20]. Veri kümesinde Amerika Birleşik Devletleri, Kanada ve İngiltere de olmak üzere 11 ülkeye ait trend videolar bulunmaktadır. Veri kümesinde her bir videoya ait video başlığı, kategori numarası, etiketler,

beğenme sayısı, beğenmeme sayısı ve yorum sayısı gibi 16 parametre bulunmaktadır. Her gün için 200’e kadar günlük trend video verisi bulunmaktadır.

Bu çalışmada, ülkeler arası kalıcı etiketlerin keşfi için Amerika Birleşik Devletleri, Kanada ve İngiltere (ABD, Ka, İn) trend video veri kümeleri ve 2022 yılına bir yıllık trend video kayıtları seçilmiştir. 2022 yılına ait Amerika Birleşik Devletleri, Kanada ve İngiltere veri kümelerindeki toplam video sayıları sırasıyla 72.999, 72.999 ve 72.998’dir. Veri kümelerindeki video sayılarının aylara göre gösterimi Şekil 1’de sunulmuştur.



Şekil 1. Seçilen ülke veri kümeleri için aylar bazında video sayıları

2.2 Önerilen metot

Bu bölümde, önerilen ülkeler arası kalıcı etiketlerin keşfi metodu ve algoritması sunulmuştur. İlk olarak temel tanımlar yapılmış, daha sonra önerilen algoritma sunulmuştur. Bu çalışmaya ait akış diyagramı Şekil 2’de sunulmuştur.

Şekil 2’de görüldüğü üzere, ilk olarak Ülkeler YouTube trend video veri kümesi ön işlemden geçirilmekte ve her ülke için PPTagD algoritması çalıştırılmaktadır. Daha sonra, her bir benzersiz etiket için ülkeler arası etiket kalıcılığı değeri hesaplanmakta ve eğer etiketler verilmiş olan bir minimum eşik değerini sağlıyorsa ülkeler arası kalıcı etiketler listesine eklenmektedir. Son olarak ülkeler arası kalıcı etiketler listesi çıktı olarak verilmektedir.

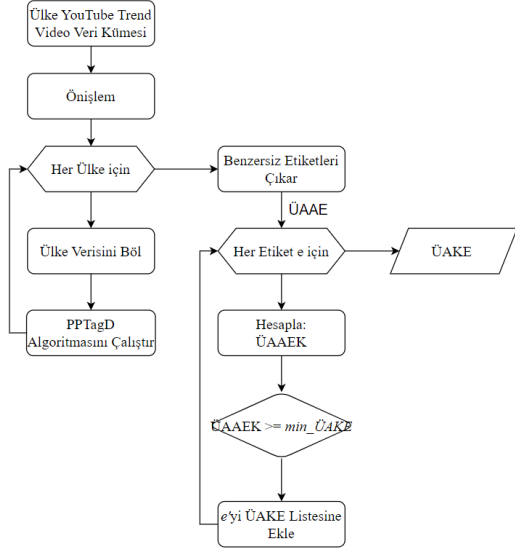
2.2.1 Temel tanımlar

Bu bölümde, ülkeler arası kalıcı etiketlerin keşfi ile ilgili temel tanımlar sunulmuştur.

Tanım 1. Ülke YouTube Trend Video Büyük Veri Kümesi (ÜYTVK) birden fazla ülkenin YouTube üzerindeki günlük trend videolarına ait meta bilgilerini içeren veri kümesidir. Veri kümesine ait detaylar Bölüm 2.1’de sunulmuştur.

Tanım 2. Ülke Günlük Trend Video Etiket (ÜGTVE) ÜYTVK içerisinde bulunan bir ülkedeki günlük trend video ile ilişkili etikettir.

Tanım 3. Ülke Kalıcı Etiket (ÜKE) PPTagD algoritması [19] sonucu kalıcı etiket olarak keşfedilmiş olan etikettir.



Şekil 2. Bu çalışmaya ait akış diyagramı

Tanım 4. Ülke Kalıcı Etiketler Listesi (ÜKEL) ÜYTVK içerisinde bulunan her ülke için keşfedilmiş olan ÜKE listesidir.

Tanım 5. Ülkeler Arası Aday Etiketler (ÜAAE) ÜYTVK içerisindeki her bir ülke için keşfedilmiş olan benzersiz ÜKEL'dir.

Tanım 6. Ülkeler Arası Aday Etiket Kalıcılığı (ÜAAEK) verilen bir t etiketinin her bir ülkedeki etiket kalıcılığı değerlerinin ortalamasıdır.

$$\text{ÜAAEK}_t = \frac{\sum \text{ülke_etiket_kalıcılığı}_t}{\text{Veri kümesindeki ülke sayısı}} \quad (1)$$

Tanım 7. Ülkeler Arası Kalıcı Etiket (ÜAKE) verilen minimum ülkeler arası etiket kalıcılığı eşik değeri min_ÜAKE 'yi sağlayan aday etikettir.

$$\text{ÜAAEK}_t \geq \text{min_ÜAKE} \quad (2)$$

2.2.2 Önerilen ÜA-KE keşfi algoritması

Bu bölümde, ülkeler arası kalıcı etiketlerin keşfi algoritması olarak hazırlanan Ülkeler Arası Kalıcı Etiket Keşfi (ÜA-KE Keşfi) algoritması sunulmuştur. Önerilen algoritma, sadece bir ülkeye ait popüler ve kalıcı etiketlerin keşfini yapmak için daha önce önerilmiş olan PPTagD algoritmasının [19] geliştirilmiş bir versiyonudur. Önerilen ÜA-KE Keşfi algoritması birden fazla ülkeye ait YouTube trend veri kümesini kullanarak kalıcı etiketlerin keşfini yapmaktadır. Önerilen ÜA-Ke Keşfi algoritmasına ait sözde kod Algoritma 1'de sunulmuştur.

Algoritma 1'de, 1. adımda ÜAKE ve Tum_Etiketler dizileri oluşturulmuştur. 2. adımda, Ülke YouTube Trend Video Büyük Veri Kümesi önişlemeden geçirilmektedir. 3. ve 8. adımlar arasında her bir ülke ayrı ayrı analiz edilmekte ve ülkelere ait kalıcı etiketler PPTagD algoritması [19] kullanılarak keşfedilmektedir. 5. adımda İşlenmiş_ÜYTVK veri kümesi kullanılarak ülke günlük veri kümesi çıkarılmaktadır. 6. adımda, PPTagD algoritması çalıştırılmakta ve ülkelere ait kalıcı etiketler listesi keşfedilmektedir. 7. adımda ülke kalıcı etiketler listesi ÜKEL Tum_Etiketler genel dizisine eklenmektedir. 9. adımda, her ülkeye ait kalıcı etiketlerin listelerini içeren Tum_Etiketler genel dizisindeki benzersiz etiketler çıkarılmaktadır. 10. ve 15. adımlar arasında, ülkeler arası aday etiketler analiz edilmekte ve min_ÜAKE 'yi sağlayıp sağlamadıkları değerlendirilmektedir. 11. adımda, ülkeler arası aday etiket kalıcılığı (ÜAAEK) hesaplanmaktadır. 12. adımda, üaae 'ye ait ÜAAEK kontrol edilmekte ve eğer min_ÜAKE 'ye eşit veya büyük ise, 13. adımda üaae ÜAKE dizisine eklenmektedir. Son olarak 16. adımda ÜAKE dizisi algoritmanın çıktısı olarak döndürülmektedir.

Algoritma 1. ÜA-KE Keşfi algoritması

Girdiler:

ÜYTVK: Ülke YouTube Trend Video Veri Kümesi
 min_popülerite : Minimum etiket popüleritesi eşik değeri
 min_kalıcılık : Minimum etiket kalıcılığı eşik değeri
 min_ÜAKE : Minimum ülkeler arası etiket kalıcılığı eşik değeri

Çıktı:

ÜAKE: Ülkeler arası kalıcı etiketler

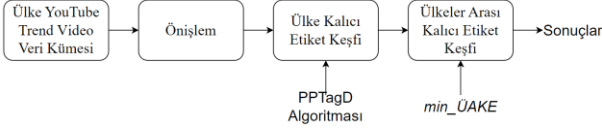
Algoritma:

1. ÜAKE = [], Tum_Etiketler = []
2. İşlenmiş_ÜYTVK = önişlem(ÜYTVK)
3. **for each** ülke $ü$ **in** ÜYTVK
4. ÜKEL = []
5. ülke_günlük_veri_kumesi = verikümesi_bol ($ü$, İşlenmiş_ÜYTVK)
6. ÜKEL = PPTagD(ülke_günlük_veri_kumesi, min_popülerite , min_kalıcılık)
7. Tum_Etiketler \leftarrow ÜKEL
8. **end for**
9. ÜAAE = benzersiz_popüler_etiketleri_çıkar (Tum_Etiketler)
10. **for each** ülkeler arası aday etiket üaae **in** ÜAAE
11. ÜAAEK = kalıcılık_hesapla (üaae , Tum_Etiketler)
12. **if** ÜAAEK \geq min_ÜAKE
13. ÜAKE \leftarrow üaae
14. **end if**
15. **end for**
16. **return** ÜAKE

3 Bulgular ve tartışma

Bu bölümde önerilen ÜA-KE Keşfi algoritmasına ait sonuçlar ve değerlendirmesi sunulmuştur. Algoritmanın işlem zamanı performansı ve min_ÜAKE eşik değerinin algoritmaya etkisi değerlendirilmiştir. Ayrıca, ABD, Kanada

ve İngiltere'ye ait ülkeler arası kalıcı etiketler sunulmuştur. Çalışmaya ait deneysel sistem Şekil 3'te sunulmuştur.



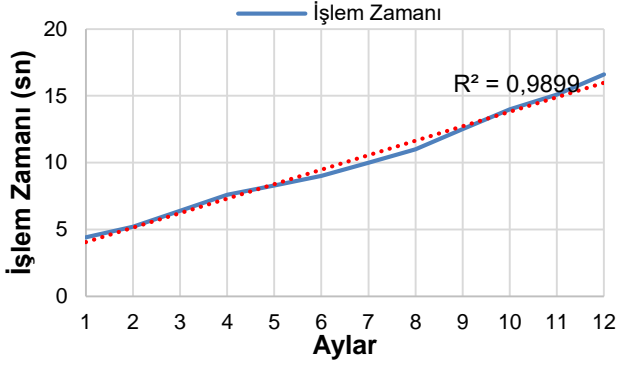
Şekil 3. Çalışmaya ait deneysel sistem

3.1 Deneysel sonuçları

Bu bölümde, önerilen algoritmaya ait deneysel sonuçlar sunulmuştur. İşlem zamanı performansı, $min_ÜAKE$ eşik değerinin etkisi ve keşfedilen sonuçları değerlendirmesi olmak üzere üç deney sunulmuştur.

3.1.1 İşlem zamanı performansı

Bu deneyde ÜA-KE Keşfi algoritmasının işlem zamanı performansı değerlendirilmiştir. Veri kümesindeki 12 ayın tamamı kullanılmış ve $min_popülarite$, $min_kalıcılık$ ve $min_ÜAKE$ eşik değerleri sırasıyla 0.025, 0.2 ve 0.2 olarak seçilmiştir. ÜA-KE Keşfi algoritmasının farklı aylara göre işlem zamanı performansı Şekil 4'te sunulmuştur.



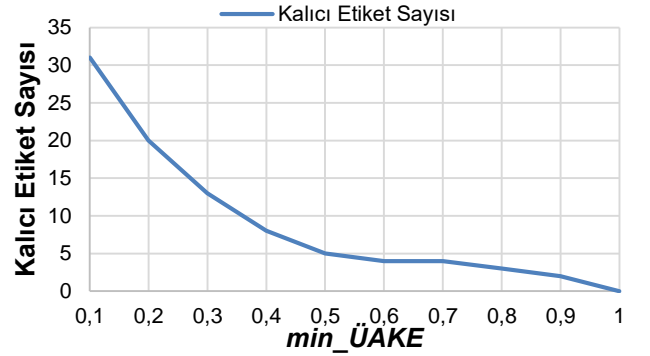
Şekil 4. ÜA-KE Keşfi algoritmasının işlem zamanı

Şekil 4'te görüldüğü üzere, ÜA-KE Keşfi algoritmasının işle zamanı, veri kümesindeki ayların sayısı arttıkça artmaktadır. Ayrıca, algoritmanın işlem zamanının 0.9899 R^2 değeriyle veri kümesinin artışına bağlı olarak lineer şekilde artmasıyla ise ÜA-KE Keşfi algoritmasının

ölçeklenebilir bir algoritma olduğunu göstermektedir. 12 aylık veri kümesinden 17 saniye içerisinde sonuçların elde edilebilmesi ise algoritmanın kullanılabilirliğini göstermektedir.

3.1.2 $min_ÜAKE$ Eşik değerinin etkisi

Bu deneyde, $min_ÜAKE$ eşik değerinin etkisi incelenmiştir. Veri kümesindeki 12 ayın tamamı kullanılmış ve $min_popülarite$ ve $min_kalıcılık$ eşik değerleri sırasıyla 0.025 ve 0.2 olarak seçilmiştir. $min_ÜAKE$ eşik değeri 0.1'den 1'e kadar 0.1'er artırılmıştır. $min_ÜAKE$ eşik değerinin etkisi Şekil 5'te sunulmuştur.



Şekil 5. $min_ÜAKE$ eşik değerinin etkisi

Şekil 5'te görüldüğü üzere, $min_ÜAKE$ eşik değerinin artışıyla birlikte ülkeler arası kalıcı etiketlerin sayısı azalmaktadır. $min_ÜAKE$ eşik değerinin 0.1 ile 0.5 değerleri arasında azalma eğilimi daha yüksek iken, $min_ÜAKE$ eşik değerinin daha yüksek değerleri için azalma oranı düşmektedir. $min_ÜAKE$ eşik değeri 1 olduğu zaman hiçbir etiket ülkeler arası kalıcı etiket olarak keşfedilememektedir.

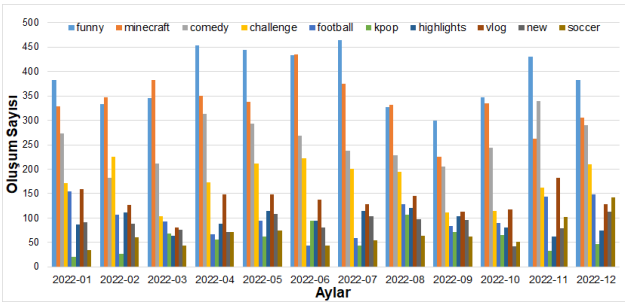
3.1.3 Algoritma sonuçlarının değerlendirilmesi

Bu deneyde, ÜA-KE Keşfi algoritmasından elde edilen sonuçlar değerlendirilmiştir. Veri kümesindeki 12 ayın tamamı kullanılmış ve $min_popülarite$, $min_kalıcılık$ ve $min_ÜAKE$ eşik değerleri sırasıyla 0.025, 0.2 ve 0.2 olarak seçilmiştir. Tablo 1'de ilk 10 ülkeler arası kalıcı etiket ÜAAEK değerleri ve her bir ülkedeki kalıcılık değerleriyle birlikte sunulmuştur.

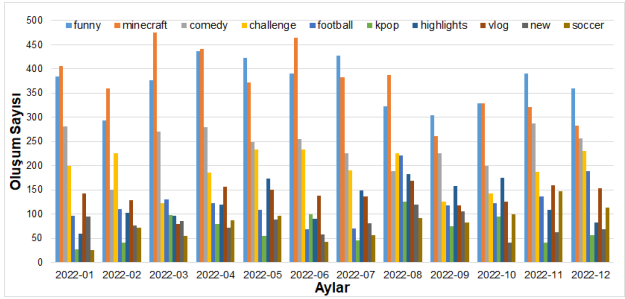
Tablo 1. İlk 10 ülkeler arası kalıcı etiket ve ülkelerdeki kalıcılık değerleri

Sıra	Etiket	ÜAAEK	ABD	Kanada	İngiltere
1	funny	0.995421	1	0.991758	0.994505
2	minecraft	0.945055	0.953297	0.961538	0.92033
3	comedy	0.893773	0.906593	0.892857	0.881868
4	challenge	0.748168	0.678571	0.777472	0.788461
5	football	0.554945	0.315934	0.447802	0.901099
6	kpop	0.437729	0.445055	0.436813	0.431319
7	highlights	0.434066	0.324176	0.425824	0.552198
8	vlog	0.401099	0.381868	0.406593	0.414835
9	new	0.386447	0.381868	0.35989	0.417582
10	soccer	0.340659	0.203297	0.247253	0.571429

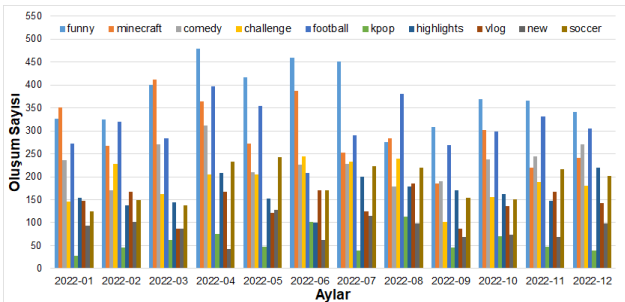
Tablo 1’de görüldüğü üzere, ilk dört etiket bütün ülkeler için oldukça yüksek etiket kalıcılık değerine sahiptir ve bu nedenle bu etiketlerin ÜAAEK değerleri daha yüksektir. Ancak, diğer etiketlerin ÜAAEK değerleri nispeten daha düşüktür. Ayrıca, ülkeler arası kalıcı etiketlerin sıralaması, her bir ülkedeki kalıcı etiketlerin sıralamasına uymamaktadır. Örneğin, *football* etiketi ABD ve Kanada’da daha düşük etiket kalıcılık değerine sahip iken İngiltere için oldukça yüksek etiket kalıcılık değerine sahiptir ve bu nedenle ülkeler arası kalıcı etiketler listesinde beşinci sırada bulunmaktadır. Benzer durum, *highlights* ve *soccer* etiketlerinde de bulunmaktadır. Keşfedilen etiketler değerlendirildiği zaman, elde edilen sonuçların [19]’daki 2021 yılına ait veri kümesindeki ABD sonuçlarıyla uyumlu olduğu görülmektedir.



Şekil 6. ABD için kalıcı etiketlerin aylara göre dağılımı



Şekil 7. Kanada için kalıcı etiketlerin aylara göre dağılımı



Şekil 8. İngiltere için kalıcı etiketlerin aylara göre dağılımı

Şekiller 6-8’de ülkeler arası kalıcı etiketlerin her bir ülkedeki aylara göre dağılımı sunulmuştur. Şekillerde görüldüğü üzere, tüm etiketlerin ülkelere göre farklı bir dağılımı olduğu gözlemlenmiştir. Örneğin, ABD ve İngiltere’de *funny* etiketi diğer etiketlere karşı önemli bir

üstünlüğü bulunmakta iken Kanada’da *funny* ve *minecraft* etiketleri farklı aylarda diğer etiketlere karşı üstünlük göstermektedir. *football* ve *soccer* etiketleri İngiltere’de belirli aylarda önemli bir paya sahip iken ABD ve Kanada’da bu etiketlerde anlamlı bir artış gözlemlenmemektedir. Bu sonuç, *football* ve *soccer* etiketlerinin İngiltere’de ABD ve Kanada’ya göre daha yaygın kullanıldığını göstermektedir. Şekillerden gözlemlenen bir diğer durum ise, İngiltere’de etiketlerin oluşum sayıları ABD ve Kanada’ya göre daha düşük olmasıdır. Bu durum etiketlerin İngiltere’de, ABD ve Kanada’ya göre daha düzgün dağıldığını ve ilk 10 dışındaki etiketlerin de belirli bir miktar paya sahip olduğunu göstermektedir.

Tablo 2. Tek bir ülkede keşfedilen bağımsız etiketler

ABD	Kanada	İngiltere
hip hop	shorts	sidemen
news	fortnite	premier league
comics		ksi
records		sport
simon cowell		action
		formula 1
		formula one
		pop
		grand prix
		sky sports
		miniminter
		music video
		auto racing

Tablo 2’de sadece bir ülkede keşfedilen bağımsız etiketler sunulmuştur. Tablo analiz edildiği zaman İngiltere’nin ABD ve Kanada’ya göre çok daha fazla bağımsız etikete sahip olduğu görülmüştür. Ayrıca, bağımsız etiketlerin içerikleri incelendiği zaman, İngiltere’deki videoların YouTube etiketlerini kullanmada farklı konulara, futbol, formula 1 ve oto yarışları gibi, odaklandığı görülmektedir. Bununla birlikte, ABD ve Kanada’nın diğer ülkelerden farklılaştığı bir konu kesin olarak keşfedilmemiştir. Burada ABD ve Kanada’nın sınır komşusu olmaları ve benzer YouTube kullanımına sahip olmalarının etkisi olabileceği düşünülmektedir.

4 Sonuçlar

YouTube, kolay kullanımı ve çok fazla kullanıcı sayısı ile video içerik platformları arasında birinci platform olarak sosyal medya kullanıcıları arasında kabul görmüş bir platformdur. YouTube büyük verilerinin analizi için yeni tekniklerin kullanılması gitgide yaygınlaşmakta ve YouTube kullanıcılarının eğilimleri hakkında önemli bilgiler sağlamaktadır. Özellikle video içerikleri ve etiketleri video sahipleri ve onların düşünceleri hakkında önemli bilgiler vermektedir.

Bu çalışmada, birden fazla ülke arasından, ülkeler arası kalıcı etiketlerin keşfi için YouTube trend video büyük veri

kümesi kullanılmıştır. Aynı dili kullandıkları için ülkeler arası kalıcı etiketlerin keşfi mümkün olabileceğinden dolayı ABD, Kanada ve İngiltere seçilmiştir. Daha önce önerilmiş olan PPTagD algoritması her bir ülke için uygulanmış ve her bir ülke için popüler ve kalıcı etiketler keşfedilmiştir. Daha sonra, yeni bir yöntem ve algoritma önerilerek ülkeler arası kalıcı etiketlerin keşfi yapılmıştır. Keşfedilen ülkeler arası kalıcı etiketler, bazı etiketlerin videolarda küresel olarak kullanıldığını, ama bununla birlikte bazı etiketlerin sadece bir ülkede kullanıldığını göstermektedir.

Gelecekteki çalışmalarda, video içeriklerini de dikkate alan bir yöntem ile ülkeler arası kalıcı etiketlerin keşfi yapılması planlanmaktadır. Ayrıca İngilizce dışındaki diğer dillerdeki veri kümeleri üzerinde de önerilen algoritma kullanılarak kalıcı etiketlerin keşfi yapılabilir.

Teşekkür

Bu araştırma Niğde Ömer Halisdemir Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimi Koordinatörlüğü'nce desteklenmiştir. Proje No: MMT2023/1-BAGEP

Çıkar çatışması

Yazarlar çıkar çatışması olmadığını beyan etmektedir.

Benzerlik oranı (iThenticate): %10

Kaynaklar

- [1] Oberlo, Oberlo YouTube statistics. <https://www.oberlo.com/blog/youtube-statistics>, Accessed 2023.
- [2] R. Novendri, A. S. Callista, D. N. Pratama, and C. E Puspita, Sentiment analysis of YouTube movie trailer comments using Naïve Bayes. *Bulletin of Computer Science and Electrical Engineering*, 1, 1, 26–32, 2020. <https://doi.org/10.25008/bcsee.v1i1.5>
- [3] M. Alkaff, A. R. Baskara and Y. H. Wicaksono, Sentiment analysis of Indonesian movie trailer on YouTube using Delta TF-IDF and SVM. 2020 5th International Conference on Informatics and Computing, pp. 1–5, Gorontalo, Indonesia, 2020. <https://doi.org/10.1109/ICIC50835.2020.9288579>
- [4] S. Singh and G. Sikka, YouTube sentiment analysis on US elections 2020. ICSCCC 2021 -International Conference on Secure Cyber Computing and Communications, pp. 250–254, Jalandhar, India, 2021. <https://doi.org/10.1109/ICSCCC51823.2021.9478128>
- [5] M. Yan, J. Sang and C. Xu, Unified YouTube video recommendation via cross-network collaboration. ICMR 2015- Proceedings of the 2015 ACM International Conference on Multimedia Retrieval, pp. 19–26, Shanghai, China, 2015. <https://doi.org/10.1145/2671188.2749344>
- [6] P. Covington, J. Adams and E. Sargin, Deep neural networks for YouTube recommendations. Proceedings of the 10th ACM Conference on Recommender Systems, pp. 191–198, Boston, Massachusetts, USA, 2016. <https://doi.org/http://dx.doi.org/10.1145/2959100.2959190>
- [7] R. Zhou, D. Xia, J. Wan and S. Zhang, An intelligent video tag recommendation method for improving video popularity in mobile computing environment. *IEEE Access*, 8, 6954–6967, 2020. <https://doi.org/10.1109/ACCESS.2019.2961392>
- [8] Q. Liu, R. Xie, L. Chen, S. Liu, K. Tu, P. Cui, B. Zhang and L. Lin, Graph neural network for tag ranking in Tag-enhanced video recommendation. *International Conference on Information and Knowledge Management, Proceedings*, 1, pp. 2613–2620, Ireland, 2020. <https://doi.org/10.1145/3340531.3416021>
- [9] S. Agarwal and A. Sureka, A focused crawler for mining hate and extremism promoting videos on YouTube. HT 2014- Proceedings of the 25th ACM Conference on Hypertext and Social Media, pp. 294–296, Santiago, Chile, 2014. <https://doi.org/10.1145/2631775.2631776>
- [10] A. Matamoros-Fernández, Platformed racism: the mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube. *Information Communication and Society*, 20, 6, 930–946, 2017. <https://doi.org/10.1080/1369118X.2017.1293130>
- [11] R. Ottoni, E. Cunha, G. Magno, P. Bernardina, W. Meira Jr. And V. Almeida, Analyzing Right-wing YouTube channels: Hate, violence and discrimination. Proceedings of the 10th ACM Conference on Web Science, pp. 323–332, Amsterdam, Netherlands, 2018. <https://doi.org/10.1145/3201064.3201081>
- [12] F. Figueiredo, F. Benevenuto and J. M. Almeida, The tube over time: Characterizing popularity growth of YouTube videos. Proceedings of the 4th ACM International Conference on Web Search and Data Mining, pp. 745–754, Hong Kong, China, 2011. <https://doi.org/10.1145/1935826.1935925>
- [13] S. V. Chelaru, C. Orellana-Rodriguez and I. S. Altingovde, Can social features help learning to rank YouTube videos? Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7651 LNCS, pp. 552–566, 2012. https://doi.org/10.1007/978-3-642-35063-4_40
- [14] A. Brodersen, S. Scellato and M. Wattenhofer, YouTube around the world: Geographic popularity of videos. WWW'12- Proceedings of the 21st Annual Conference on World Wide Web, pp. 241–250, Lyon France, 2012. <https://doi.org/10.1145/2187836.2187870>
- [15] A. Krishna, J. Zambreno and S. Krishnan, Polarity trend analysis of public sentiment on YouTube. The 19th International Conference on Management of Data (COMAD), pp. 125–128, Ahmedabad, India, 2013.
- [16] W. Hoiles, A. Aprem and V. Krishnamurthy, Engagement and popularity dynamics of YouTube videos and sensitivity to Meta-Data. *IEEE Transactions on Knowledge and Data Engineering*, 29, 7, 1426–1437, 2017. <https://doi.org/10.1109/TKDE.2017.2682858>

- [17] Y. L. Chen and C. L. Chang, Early prediction of the future popularity of uploaded videos. *Expert Systems with Applications*, 133, 59–74, 2019. <https://doi.org/10.1016/j.eswa.2019.05.015>
- [18] G. M. H. C. Gajanayake and T. C. Sandanayake, Trending pattern identification of youtube gaming channels using sentiment analysis. 20th International Conference on Advances in ICT for Emerging Regions, ICTer 2020- Proceedings, ICTer, pp. 149–154, Colombo, Sri Lanka, 2020. <https://doi.org/10.1109/ICTer51097.2020.9325476>
- [19] Y. Dokuz, Discovering popular and persistent tags from YouTube trending video big dataset. *Multimedia Tools and Applications*, 1-19, 2023. <https://link.springer.com/article/10.1007/s11042-023-16019-z>
- [20] R. Sharma, YouTube Trending Video Dataset. <https://www.kaggle.com/rsrishav/youtube-trending-video-dataset>, Accessed 2022

