Çukurova Üniversitesi
Mühendislik Fakültesi Dergisi
Çukurova University
Journal of the Faculty of Engineering
ISSN: 2757-9255 CİLT/VOLUME: 40 SAYI/ISSUE: 3 EYLÜL/SEPTEMBER 2025

Kullanıcı Eğilimlerinin Göreceli Modellenmesinin Popülerlik Yanlılığına Etkisi

Nilüfer BALLI [1,a], **Tuğba TÜRKOĞLU KAYA** [2,b]

[1]*Ardahan Üniversitesi, Lisansüstü Eğitim Enstitüsü, İleri Teknolojiler Bölümü, Ardahan, Türkiye*
[2]*Ardahan Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Ardahan, Türkiye*

[a]***ORCID**: 0009-0004-2826-1189;* [b]***ORCID**: 0000-0003-2698-8961*

Makale Bilgileri
Geliş : 23.07.2025
Kabul : 16.09.2025
DOI: 10.21605/cukurovaumfd.1748924

Sorumlu Yazar
Tuğba Türkoğlu Kaya
tugbaturkoglu@ardahan.edu.tr

Anahtar Kelimeler
Çeşitlilik
Çok ölçütlü öneri sistemleri
Göreceli skor
Popülerlik yanlılığı

***Atıf şekli:** BALLI, N., KAYA, T.T., (2025). Kullanıcı Eğilimlerinin Göreceli Modellenmesinin Popülerlik Yanlılığına Etkisi. Çukurova Üniversitesi, Mühendislik Fakültesi Dergisi, 40(3), 671-685.*

ÖZ

Dijitalleşmenin hızlanmasıyla kullanıcılar çok sayıda ürün ve hizmet seçeneğiyle karşı karşıya kalmakta, bu da kişiselleştirilmiş içeriklere erişimi zorlaştırmaktadır. Öneri sistemleri bu soruna çözüm sunmakla birlikte, geleneksel yaklaşımlar genellikle tek boyutlu puanlara dayanmakta ve popüler içerikleri öne çıkarma eğilimindedir. Bu durum önerilerin çeşitlilik ve adaletini sınırlandırmaktadır. Bu çalışmada, çok boyutlu veriler üzerinde kullanıcıların puanlama davranışlarını bağlamsal olarak analiz eden yeni bir yöntem olan RelPop önerilmektedir. RelPop, aynı puanın farklı kullanıcılar için göreli anlamını dikkate alarak içerikleri yeniden sıralamakta, böylece önerilerin özgünlük ve adalet düzeyi artmaktadır. Ayrıca, öneri listelerindeki popülerlik yanlılığını ölçmek için ADPI metriği geliştirilmiştir. İki çok ölçütlü veri seti üzerinde yapılan deneyler, RelPop'un önerilerin yenilik ve özgünlüğünü artırdığını, ADPI'nın ise popülerlik yanlılığını daha hassas biçimde ortaya koyduğunu göstermektedir.

The Effect of Relative Modeling of User Trends on Popularity Bias

Article Info
Received : 23.07.2025
Accepted : 16.09.2025
DOI: 10.21605/cukurovaumfd.1748924

Corresponding Author
Tuğba Türkoğlu Kaya
tugbaturkoglu@ardahan.edu.tr

Keywords
Diversity
Multi-criteria recommender system
Preference score
Popularity bias

***How to cite:** BALLI, N., KAYA, T.T., (2025). The Effect of Relative Modeling of User Trends on Popularity Bias. Cukurova University, Journal of the Faculty of Engineering, 40(3), 671-685.*

ABSTRACT

Users are faced with an overwhelming number of products and services, making personalized content harder to access. Recommender systems address this challenge, but traditional approaches often rely on one-dimensional ratings and tend to prioritize popular items. This limits diversity and fairness in recommendations. To overcome these issues, this study proposes RelPop, a novel method that contextually analyzes users' rating behaviors on multi-dimensional data. RelPop accounts for the relative meaning of the same score across different users, re-ranking items according to individual evaluation habits. This leads to more original, fair, and user-specific recommendations. Furthermore, a new performance metric, ADPI, is introduced to objectively measure popularity bias by evaluating how much recommended items deviate from the popularity center in a non-directional way. Experimental results on two multi-criteria datasets demonstrate that RelPop enhances novelty and uniqueness, while ADPI provides a more precise assessment of popularity bias.

1. GİRİŞ

Dijitalleşmenin ve çevrimiçi platformların yaygınlaşmasıyla birlikte, kullanıcıların karşılaştığı ürün ve hizmet seçeneklerinin çeşitliliği hızla artmaktadır [1]. Örneğin, Netflix' te 5000' den fazla film ve dizi, Amazon' da milyonlarca ürün, Spotify'da 70 milyondan fazla şarkı bulunmaktadır [2]. Bu artış, kullanıcıların kendi ilgi ve ihtiyaçlarına uygun içeriklere ulaşmasını zorlaştırmakta ve karar verme süreçlerini karmaşıklaştırmaktadır [3]. Kullanıcılar, bu geniş içerik havuzunda kendilerine en uygun seçenekleri bulmakta zorlanmakta ve bu durum "seçim paradoksu" olarak adlandırılan bir soruna yol açmaktadır. Örneğin, bir kullanıcı Netflix' te film seçerken, binlerce seçenek arasından kendisine en uygun olanı bulmakta zorlanabilir ve bu durum karar verme sürecini olumsuz etkileyebilmektedir [4]. Dolayısıyla e-ticaret, dijital medya ve çevrimiçi hizmet platformlarında, kullanıcıların ihtiyaçlarına ve zevklerine uygun içeriklere ulaşabilmesi için öneri sistemleri vazgeçilmez bir araç haline gelmiştir [5]. Bu sistemler, kullanıcıların geçmiş davranışlarını, tercihlerini ve etkileşimlerini analiz ederek kişiselleştirilmiş öneriler sunmaktadır [6]. Ancak, geleneksel öneri sistemlerinin çoğu, kullanıcıların değerlendirmelerini tek bir genel puan üzerinden modellemekte ve ürünlerin çok boyutlu doğasını göz ardı etmektedir [7,8]. Örneğin, bir film değerlendirmesinde kullanıcılar oyunculuk, senaryo, yönetmenlik, görsel efektler gibi farklı kriterleri dikkate alırken, sistem bu değerlendirmeleri tek bir genel puan üzerinden ele almaktadır [7,9]. Bu yaklaşım, kullanıcıların her bir kriter için farklı beklentileri ve öncelikleri olduğu gerçeğini göz ardı etmektedir [10,11]. Bu çok yönlü değerlendirme ihtiyacını karşılamak üzere geliştirilen çok ölçütlü öneri sistemleri, her bir ürün veya hizmetin çeşitli yönlerini ayrı ayrı analiz ederek daha rafine ve kullanıcı odaklı öneriler sunmaktadır [11,12]. Bu sistemler, kullanıcıların farklı kriterlere verdikleri önem derecelerini dikkate alarak, daha doğru ve kişiselleştirilmiş öneriler oluşturabilmektedir [13]. Örneğin, bir film öneri sisteminde, bir kullanıcı oyunculuk performansına daha fazla önem verirken, başka bir kullanıcı senaryo kalitesine daha fazla önem verebilir. Çok ölçütlü öneri sistemleri olarak adlandırılan bu yapılar, bu farklılıkları dikkate alarak her kullanıcı için daha uygun öneriler sunabilmektedir. Ancak, her ne kadar oldukça başarılı tavsiyeler üretip kullanıcı memnuniyetini arttırsa da bu sistemler popülerlik yanlılığı olarak ifade edilen problemle karşı karşıya kalabilmektedir. Bu problem, sistemlerin çoğunlukla yüksek oylama almış ya da geniş kitlelerce bilinen içerikleri öne çıkarması; buna karşılık daha az bilinen, ancak kullanıcıya potansiyel olarak daha uygun içerikleri arka planda bırakması şeklinde ortaya çıkmaktadır [14,15]. Literatürde bu sorunun etkilerini azaltmaya yönelik çeşitli algoritmalar ve kullanıcı eğilimi modelleme yaklaşımları geliştirilmiş olmakla birlikte, çalışmaların büyük çoğunluğu ya tek ölçütlü sistemlere odaklanmakta ya da kullanıcıların kriterler arası tercih farklılıklarını yeterince dikkate almamaktadır [16], [17]. Örneğin, Klimashevskaia ve arkadaşları [18], işbirlikçi filtreleme tabanlı algoritmaların popüler ürünleri aşırı öne çıkardığını ve bunun kişiselleştirme kalitesini olumsuz etkilediğini göstermiştir. Kowald ve Lacic [19] ise popülerlik yanlılığının hem ürün hem de kullanıcı düzeyinde farklı yansımalarını ele alarak, popüler içeriklere ilgi duymayan kullanıcılar için öneri doğruluğunun anlamlı şekilde düştüğünü ortaya koymuştur. Ayrıca, Abdollahpouri ve arkadaşları [20], bu yanlılığın adalet ve kalibrasyon üzerindeki etkilerini incelemiş, popüler olmayan ürünlerin bu ürünlere ilgi duyan kullanıcılara dahi yeterince önerilmediğini vurgulamıştır. Bu kapsamda, öneri sistemlerinde popülerlik etkisini daha doğru şekilde değerlendirmek üzere Steck [21] “popularity-stratified recall” metriğini önermiş, Wei ve arkadaşları [22] ise nedensel çıkarım ve karşıt-gerçeksel analiz teknikleriyle popülerlik etkisini ayrıştırmaya çalışmıştır. Dolayısıyla çalışmaların her birinde çok ölçütlü yapılar üzerinde bu problemin etkisinin göz ardı edildiği görülmektedir.

Bu çalışmada, öneri sistemlerinde sıklıkla karşılaşılan popülerlik yanlılığının çok ölçütlü sistemlerde etkisini azaltmaya yönelik olarak, kullanıcıların geçmiş puanlama davranışlarını dikkate alan yenilikçi bir kullanıcı popülerlik eğilimi hesaplama yöntemi olan Göreceli Popülerlik Skoru (RelPop) önerilmektedir. RelPop yöntemi, her bir kullanıcının puanlama ölçeğini bağlamsal olarak analiz ederek, benzer puanların farklı kullanıcılar nezdinde farklı anlamlar taşıyabileceği gerçeğini temel almaktadır. Bu bağlamda yöntem, mutlak puan değerlerinden ziyade, puanların kullanıcının kendi puanlama eğilimleri içindeki göreli konumunu dikkate alarak, bireysel memnuniyet düzeyini daha hassas biçimde modellemektedir. Örneğin, genel olarak düşük puanlar veren bir kullanıcının bir ürüne verdiği 4 ya da 5 gibi yüksek bir puan, bu ürünün söz konusu kullanıcı için gerçekten öne çıkan ve tatmin edici bir deneyim sunduğunu işaret ederken; sıkça yüksek puanlar veren bir kullanıcı için benzer bir puan daha sıradan bir değerlendirme olarak yorumlanabilir. RelPop yöntemi bu farklılıkları sistematik biçimde analiz ederek, kullanıcıların göreli memnuniyet düzeylerini nesnel şekilde hesaplamayı ve bu sayede öneri listelerinde daha çeşitli, daha az popüler ama potansiyel olarak daha ilgi çekici içeriklerin yer almasını sağlamayı amaçlamaktadır. Bu

yaklaşım, öneri sistemlerinin bireysel bağlamları daha etkin biçimde dikkate almasını sağlayarak hem içerik çeşitliliğini hem de kullanıcı adaletini artırma potansiyeline sahiptir.

Çalışmanın başlıca katkıları aşağıda özetlenmektedir:

- Kullanıcıların göreli puanlama eğilimlerine dayalı olarak popülerlik eğilimini daha hassas biçimde ölçebilen yeni bir hesaplama yöntemi geliştirilmiştir. Bu yöntem, kullanıcıların genel değerlendirme davranışlarını dikkate alarak sistematik bir popülerlik analizi yapma olanağı sunmaktadır.
- Bu yöntem, çok ölçütlü öneri sistemlerine uyarlanarak farklı algoritmalarla kapsamlı karşılaştırmalı deneyler yapılmıştır.
- Önerilen yöntem, çok ölçütlü öneri sistemlerine uyarlanarak farklı algoritmalarla kapsamlı karşılaştırmalı deneyler gerçekleştirilmiştir. Bu kapsamda, öneri performansı ve popülerlik eğilimi açısından çeşitli sistemlerin duyarlılıkları karşılaştırmalı olarak değerlendirilmiştir.
- Popülerlikten sapma düzeyini yönsüz biçimde ölçen yeni bir değerlendirme metriği olan Mutlak Popülerlik Eğilimindeki Sapma (Absolute Deviation in Popularity Inclination-ADPI) önerilmiştir. Bu metrik, mevcut literatürdeki yönlü ölçütlerin tamamlayıcısı olarak sistemlerin popülerlikten sapma derecelerini daha bütüncül biçimde analiz etme olanağı sunmaktadır.
- Ayrıca, çok ölçütlü öneri sistemleri alanında popülerlik yanlılığına yönelik çalışmaların sınırlı sayıda olması göz önünde bulundurularak, bu çalışma söz konusu boşluğu doldurmayı ve alana yeni bir bakış açısı kazandırmayı hedeflemektedir

Bu çalışma altı bölümden oluşmaktadır. Çalışma boyunca kullanılan yöntemler ve metotlar Bölüm 2' de yer alırken; Bölüm 4 ve 5' te sırasıyla Deney metodolojisi ve Deneysel sonuçlar yer almaktadır. Son bölüm ise sonuç ve önerilerden oluşmaktadır.

2. ÖN ÇALIŞMALAR

Bu bölümde çalışma boyunca kullanılan yöntemlere ve metotlara yer verilmiştir.

2.1. Kullanıcı Popülerlik Eğilimi Belirleme Yöntemleri

Öneri sistemlerinde popülerlik yanlılığını azamaya yönelik geliştirilen pek çok yöntem, genellikle öneri listelerini sonradan işleme teknikleriyle yeniden sıralamaya odaklanır; bu süreçte popüler ürünler geri plana itilirken, daha az bilinen ürünler öne çıkarılır. Ancak bu yaklaşımlar, çoğu zaman kullanıcıların ürünlerin popülerliğiyle ilgili gerçek tercihlerini dikkate almaz. Oysa pratikte, bazı kullanıcılar daha çok popüler içerikleri tercih ederken, bazıları ise daha az bilinen seçeneklere yönelmektedir. Bu nedenle, kullanıcılara daha kişiselleştirilmiş ve tatmin edici öneriler sunabilmek için, onların popülerlik eğilimlerinin doğru şekilde tespit edilmesi ve öneri algoritmalarına entegre edilmesi gerekmektedir. Bu bölümde, literatürde kullanıcıların popülerlik eğilimini ölçmek amacıyla yaygın olarak kullanılan yöntemler özetlenmektedir.

2.1.1. Popüler Ürünlerin Oranı Yaklaşımı

Popüler ürünlerin oranı yaklaşımı (RPI), kullanıcıların öneri sistemlerinde yaygın olarak tercih edilen içeriklere ne kadar ilgi gösterdiğini anlamak için kullanılan bir metriktir. Bu yaklaşımda, sistemdeki tüm öğeler, aldıkları toplam oy veya puan sayısına göre sıralanır. En çok oy alan ve sistemdeki toplam oyların yaklaşık beşte birini oluşturan öğeler "popüler" olarak tanımlanırken, geri kalanlar "az popüler" olarak kabul edilir [20]. Bu doğrultuda bu yöntem, bir kullanıcının etkileşimde bulunduğu içeriklerin ne kadarının bu popüler gruba girdiğini hesaplar. Örneğin, bir film öneri platformunda, kullanıcı A'nın izlediği 10 filmin 7'si en çok izlenenler listesindeyse, bu kullanıcının RPI değeri 0,7 olur. Böylece, öneri sistemi kullanıcıların popüler içeriklere mi yoksa daha niş seçeneklere mi yöneldiğini daha iyi analiz edebilir.

$$RPIu = \frac{|\{i \in Iu \mid i \in H\}|}{|Iu|} \quad (1)$$

Burada, I_u kullanıcının değerlendirdiği tüm öğeleri, H ise sistemde popüler olarak tanımlanan öğeleri temsil etmektedir.

2.1.2. Derecelendirilen Ürünlerin Ortalama Popülerliği

Derecelendirilen ürünlerin ortalama popülerliği (APRI), kullanıcıların popüler içeriklere olan eğilimini ölçmek için geliştirilmiş bir metriktir. Bu metrik, her bir öğenin popülerliğini, o öğeyi değerlendiren kullanıcı sayısının sistemdeki toplam kullanıcı sayısına oranı olarak tanımlamaktadır [23].

$$APRI_u = \frac{1}{|Iu|} \sum_{\in Iu} \frac{|\{u' \in U \mid i \in Iu'\}|}{|U|} \tag{2}$$

Burada, U tüm kullanıcı kümesini, I_u herhangi bir kullanıcının daha önce değerlendirme yaptığı ürünler kümesini ifade etmektedir.

2.1.3. Ortalamadan Daha İyi Yaklaşımı

Bu yöntemde kullanıcıların derecelendirdiği tüm ürünler yerine yalnızca beğendikleri ürünler dikkate alınmaktadır. Önceki bölümlerde açıklanan RPI veAPRI yöntemlerinin bu mantığa (BTA_{RPI}, BTA_{APRI}) uyarlanmasıdır [24]. BTA_{RPI} değeri, beğenilen öğelerin popüler kategorideki oranını ölçerek, kullanıcının popüler öğelere olan eğilimini ortaya koyarken; BTA_{APRI} metriği, kullanıcının beğendiği öğelerin popülerlik değerlerinin ağırlıklı ortalamasını alarak popüler içerik eğilimini ölçmektedir. Her iki yaklaşımda da yalnızca beğeni eşiğini aşan öğeler dikkate alınarak, kullanıcının gerçek tercihleri üzerinden popülerlik eğilimi analiz edilir.

2.1.4. Kullanıcı Popülerlik Eğilimini Tespit Etmek İçin Önerilen Yeni Yaklaşım: Göreceli Popülerlik Skoru

Kullanıcıların puanlama davranışlarının daha derinlemesine ve niteliksel olarak analiz edilebilmesi için, Lee ve arkadaşları [25] geliştirdiği tercih modeli bu çalışmada referans alınmıştır. Söz konusu model, kullanıcının bir ürüne verdiği puanı, o kullanıcının genel puanlama alışkanlıklarıyla birlikte ele alarak değerlendirir. Bu sayede, aynı puan değeri, farklı kullanıcılar için farklı anlamlar taşıyabilmektedir. Örneğin, çoğunlukla yüksek puanlar veren bir kullanıcı ile nadiren yüksek puan veren bir kullanıcının aynı ürüne verdiği yüksek puan, tercih düzeyi açısından aynı şekilde yorumlanmamalıdır. Tercih Skoru (Preference Score) olarak adlandırılan bu yöntemde, her bir kullanıcının değerlendirdiği ürünler arasında, belirli bir ürüne verdiği puanın göreceli konumu dikkate alınarak bir tercih skoru bulunmaktadır (Eşitlik 3).

$$Pref_{(r_{u,i})} = \alpha \frac{pref > (r_{u,i})}{R_u^+} + \beta \frac{pref = (r_{u,i})}{R_u^+} \tag{3}$$

Eşitlik 3'te:

- $pref > (r_{u,i})$: Kullanıcının i ürününe verdiği puandan daha düşük puan verdiği ürünlerin sayısı,
- $pref = (r_{u,i})$): Kullanıcının i ürününe verdiği puanla aynı puan verdiği ürünlerin sayısı,
- R_u^+: Kullanıcının değerlendirdiği toplam ürün sayısı,
- α ve β: Ağırlık parametreleri olup, literatürde genellikle sırasıyla 1 ve 0,5 olarak alınmaktadır

Önerilen bu yaklaşımda (Göreceli Popülerlik Skoru, RelPop), yukarıda tanımlanan tercih puanları, her kullanıcı için değerlendirdiği tüm ürünler üzerinde ortalama alınarak, kullanıcının genel popülerlik eğilimi skoru aşağıdaki denklem kullanılarak elde edilmiştir.

$$RelPop_u = \frac{1}{|I_u|} \sum_{i \in I_u} Pref_{(r_{u,i})} \tag{4}$$

Eşitlik 4'de I_u, u kullanıcısının değerlendirdiği ürünlerin kümesini ifade etmektedir. Bu yöntem sayesinde, kullanıcıların puanlama davranışlarındaki bireysel farklılıklar dikkate alınarak, daha adil ve kişiselleştirilmiş bir popülerlik eğilimi ölçümü sağlanmıştır.

RelPop yöntemi ile geleneksel işbirlikçi filtreleme algoritmalarının sıklıkla göz ardı ettiği önemli bir soruna, yani kullanıcıların puanlama davranışlarındaki bağlamsal ve göreli farklılıkların dikkate alınmaması problemine doğrudan çözüm sunması amaçlanmıştır. Klasik yöntemler, tüm kullanıcıların aynı puanı benzer anlamlarla verdiği varsayımıyla hareket ederek, bireysel değerlendirme ölçeklerindeki farklılıkları göz ardı etmekte ve bu durum önerilerin doğruluğunu, çeşitliliğini ve kullanıcıya olan uygunluğunu sınırlamaktadır. Bu çalışmada önerilen bu yaklaşım, kullanıcıların geçmiş puanlama dağılımlarını analiz ederek her bir puanın o kullanıcı özelindeki anlamını modellemekte ve böylece daha gerçekçi bir tercih temsili sunmaktadır. Bu yöntem sayesinde, önerilen içeriklerin yalnızca popülerliğe dayalı olarak değil, kullanıcıların göreli memnuniyet düzeyleri doğrultusunda kişiselleştirilmiş biçimde sunulması mümkün hale getirmesi amaçlanmıştır. Sonuç olarak, sistem hem popülerlik yanlılığını azaltacak hem de kullanıcı memnuniyetini ve öneri kalitesini arttıracaktır. Özellikle çok ölçütlü karar ortamlarında bu bağlamsal analiz, öneri sistemlerinin adil, dengeli ve kullanıcı odaklı hale gelmesine önemli katkılar sağlayacaktır.

2.2. Popülerlik Yanlılığını Azaltmaya Yönelik Algoritmalar

Öneri sistemlerinde karşılaşılan temel sorunlardan biri olan popülerlik önyargısı, sistemin etkinliğini ve kullanıcı memnuniyetini önemli ölçüde düşürmektedir [20]. Bu sorunun çözümüne yönelik olarak, literatürde: artırılmış fayda üretimi (Aug), çarpımsal fayda öğrenme (Mul) ve popülerlik duyarlı öğe ağırlıklandırma (Paiw) gibi yaklaşımlar sunulmuş ve her biri bu bölümde incelenmiştir.

2.2.1. Artırılmış Fayda Üretimi Algoritması

Artırılmış fayda üretimi algoritması (Aug), öneri sistemlerinde içerik dağılımını dengelemek için geliştirilmiş yenilikçi bir yaklaşımdır [26]. Bu yöntem, makine öğrenmesi modelinin tahmin ettiği kullanıcı ilgisini, içeriğin popülerlik seviyesine göre belirlenen bir dengeleme katsayısıyla birleştirerek daha dengeli öneriler sunmayı amaçlar. Bu sayede, çok popüler içeriklerin ağırlığı azaltılırken, daha az bilinen içerikler öne çıkarılır.

2.2.2. Çarpımsal Fayda Öğrenme Algoritması

Çarpımsal fayda öğrenme algoritması (Mul), öneri sistemlerinde popüler içeriklerin etkisini azaltmak için geliştirilmiş yenilikçi bir sıralama tekniğidir. Her içeriğin tahmini puanı, popülerlik seviyesiyle ters orantılı bir katsayıyla çarpılarak, popüler içeriklerin puanları belirgin şekilde düşürülür. Bu sayede, ana akım içeriklerin öneri listelerinde baskın olması engellenir ve kullanıcıya daha çeşitli içerikler sunulur [27].

2.2.3. Popülerlik Duyarlı Öğe Ağırlıklandırma Algoritması

Popülerlik duyarlı öğe ağırlıklandırma (Paiw), öneri sistemlerinde içerik çeşitliliğini artırmak için geliştirilmiş bir yaklaşımdır. Yöntem, içeriklerin popülerlik seviyelerine göre logaritmik bir dönüşümle ağırlıklar hesaplar; popüler içerikler düşük, az bilinenler ise yüksek ağırlık alır. Bu ağırlıklar, tahmini öneri puanlarıyla λ parametresiyle birleştirilerek, hem kullanıcı tercihlerine uygun hem de çeşitli öneriler sunulur. Paiw, öneri kalitesini koruyarak çeşitliliği artırır [28].

3. DENEY METODOLOJİSİ

Bu çalışmada, önerilen RelPop yönteminin etkinliğini değerlendirmek için kapsamlı bir deneysel çerçeve tasarlanmıştır. Değerlendirme süreci, yaygın olarak kullanılan tek-çıkışlı çapraz (leave-one out) doğrulama stratejisi temel alınarak yapılandırılmıştır. Her bir değerlendirme iterasyonunda, veri setinden rastgele bir kullanıcı test kullanıcısı olarak seçilirken, kalan kullanıcılar eğitim setini oluşturmak için kullanılmıştır. Eğitim verileri kullanılarak, test kullanıcısının ürünleri için tahmini puanları oluşturmak üzere bir işbirlikçi filtreleme algoritması uygulanmış ve bu tahmini değerlerden en yüksek 10 ürün kullanıcıya öneri olarak verilmiştir. Bu kapsamda, iki farklı analiz senaryosu tasarlanmıştır. İlk senaryoda, kullanıcıların film ve otellere verdikleri genel puanlar dikkate alınarak analiz gerçekleştirilmiştir. Bu yaklaşımda her bir kullanıcının tek bir bütünsel değerlendirmesi üzerinden tercih eğilimleri modellenmiş ve öneri sistemi

performansı, bu genel puanlar üzerinden test edilmiştir. Böylece, sistemin genel beğeni düzeyini ne ölçüde yakalayabildiği ortaya konulmuştur. İkinci senaryoda ise, kullanıcıların her bir kriter için (örneğin filmler için hikâye, oyunculuk, yönetmenlik, görsel efektler; oteller için konum, temizlik, hizmet, değer ve odalar) ayrı ayrı verdikleri puanlar kullanılarak kriter bazlı analiz uygulanmıştır. Bu yöntemde, kullanıcıların tercihlerinin farklı boyutlarını yansıtan çok ölçütlü verilerden yararlanılmış ve öneri sistemi performansının bu ayrıntılı tercihler altında nasıl değiştiği incelenmiştir. Bu iki senaryonun karşılaştırmalı değerlendirilmesi sayesinde, önerilen yöntemin hem genel değerlendirmeler temelinde hem de ayrıntılı kriter bazında kullanıcı tercihlerini ne ölçüde yansıttığı ortaya konulmuştur. Böylece, öneri sisteminin farklı veri senaryolarında gösterdiği başarı ve esneklik hakkında daha kapsamlı bir analiz yapılmıştır.

Çalışmamızın ilk aşaması olan kullanıcı popülerlik analizinde, test setindeki her bir kullanıcı geleneksel yöntemlere (*APRI, RPI, BTA*$_{APRI}$, *BTA*$_{RPI}$) ek olarak önerilen RelPop yöntemine göre eğilimler tespit edilip, bunların karşılaştırmalı analiz sonuçları verilirken; ikinci aşamada önerilen bu yöntemin performansı, Mul, Paiw ve Aug gibi popüler yeniden sıralama yöntemleriyle karşılaştırılmıştır.

3.1. Veri Seti

Bu çalışmada, öneri sistemleri alanında yaygın olarak kullanılan çok ölçütlü Yahoo! Movies (YM) ve TripAdvisor (TA) veri setleri kullanılmıştır. YM20 veri seti, 429 kullanıcı ve 491 filmden oluşmakta, toplam 18.504 değerlendirme içermektedir. Her film, hikâye, oyunculuk, yönetmenlik ve görsel efektler olmak üzere dört kriter üzerinden değerlendirilmiştir [29]. TA veri seti ise 1.354 kullanıcı ve 1.291 otelden oluşmakta, toplam 38.512 değerlendirme içermektedir [30]. %99,72 seyreklik oranı ile oldukça seyrek bir yapıdadır ve kullanıcı başına ortalama 28,44 değerlendirme ile düşük etkileşim göstermektedir. TA, konum, temizlik, hizmet, değer ve odalar olmak üzere beş kriter içermektedir. Her iki veri seti de 1-5 arası değerlendirme ölçeği kullanmakta ve seyreklik oranları ile kullanıcı başına değerlendirme sayıları bakımından farklılık göstermektedir. Bu farklılıklar, öneri sistemleri için çeşitli zorluk seviyeleri oluşturmakta ve kullanıcı davranış profillerinin analizine olanak sağlamaktadır. Veri setlerine ilişkin özet bilgiler Çizelge 1'de sunulmaktadır.

Çizelge 1. YM ve TA veri seti ile ilgili bilgiler

Veri seti	#Kullanıcı	#Ürün	#Oylama
YM20	429	491	18.504
TA	1346	1289	38.512

3.2. Değerlendirme Metrikleri

Öneri sistemlerinde popülerlik önyargısının değerlendirilmesinde çeşitli performans ölçütleri kullanılmaktadır. Bu çalışmada, ilk olarak sistemin kullanıcıya özel öneriler sunma başarısını ve popüler içeriklere olan eğilimini analiz etmek amacıyla kesinlik metriği temel alınmıştır.

3.2.1. Kesinlik

Kesinlik metriği, öneri sistemlerinde hem alaka düzeyinin hem de yanlılık etkisinin ölçülmesinde literatürde yaygın olarak tercih edilmekte ve bu metrikte aktif kullanıcı için ilginç ürünlerin önerilme oranı belirlenmektedir [31]. Aşağıdaki eşitlik kullanılarak hesaplama yapılabilmektedir:

$$Kesinlik@K(u) = \frac{|Rel_u \cap Rec_u^K|}{|Rec_u^K|} \tag{5}$$

Eşitlik 5'te, Rel_u kullanıcı u için ilgili öğeler kümesini, Rec_u^K ise kullanıcı u için önerilen en iyi K öğe kümesini ifade etmektedir.

3.2.2. Geri Çağırma

Öneri sistemlerinin başarısını değerlendirmek için yaygın olarak kullanılan temel ölçütlerden biridir. Bu metrik, sistemin kullanıcıya ait tüm ilgili öğeler arasından ne kadarını öneri listesine dahil edebildiğini gösterir [32] ve matematiksel formülasyonu aşağıda yer almaktadır:

$$Geri\ çağırma@K(u) = \frac{|Rel_u \cap Rec_u^K|}{|Rec_u|} \quad (6)$$

Kullanıcı u için ilgili öğelerin tamamı Rel_u kümesinde yer alırken, sistemin kullanıcıya sunduğu en iyi K adet öneri ise Rec_u^K kümesinde bulunmaktadır (Eşitlik 6).

3.2.3. F1-Ölçütü

Kesinlik ve geri çağırma değerlerinin harmonik ortalaması olan F1-ölçütü, bu iki performans metriğini dengeli bir şekilde birleştirerek sistemin genel başarısını değerlendirmeye olanak tanır. Bu ölçütün matematiksel ifadesi şu şekildedir (Eşitlik 7):

$$F1@K(u) = 2 \times \frac{Precision@K(u) \times Recall@K(u)}{Precision@K(u) + Recall@K(u)} \quad (7)$$

3.2.4. Yenilik

Yenilik metriği, öneri sistemlerinin kullanıcılara tanıdık olmayan ancak ilgi alanlarına uygun yeni içerikler sunma yeteneğini ölçer. Bu metrik, sistemin kullanıcıların daha önce karşılaşmadığı fakat beğenme olasılığı yüksek olan öğeleri önerme başarısını değerlendirir. Yüksek bir yenilik değeri, sistemin kullanıcıları yeni ve çeşitli içeriklerle tanıştırabildiğini, böylece kullanıcı deneyimini zenginleştirdiğini gösterir [33].

$$Yenilik = \frac{1}{|R|} \sum_{i \epsilon R} -\log_2(p(i)) \quad (8)$$

Öneri listesindeki toplam öğe sayısı R ile gösterilirken, $p(i)$ değişkeni, i. öğenin herhangi bir kullanıcı tarafından değerlendirilme olasılığını ifade eder (Eşitlik 8).

3.2.5. Ortalama Uzun Kuyruk Popülerliği

Ortalama uzun kuyruk popülerliği (APLT), öneri sistemlerinin kullanıcılara sunduğu öneri listesinde yer alan az popüler veya niş içeriklerin dağılımını değerlendiren bir metriktir [33]. Herhangi bir K kullanıcısı için, sistemin oluşturduğu N_K öneri listesindeki uzun kuyruk öğelerinin oranı, aşağıdaki denklem kullanılarak hesaplanabilir

$$APLT_K = \frac{|T \cap N_K|}{N} \quad (9)$$

Burada, uzun kuyruk öğelerinin kümesi T ile gösterilirken, APLT değerinin düşük olması, sistemin öneri listesinde yeterli çeşitlilik sağlayamadığını ve genellikle popüler içeriklere ağırlık verdiğini ifade etmektedir. Buna karşılık, yüksek APLT değerleri, sistemin uzun kuyruk içeriklerini başarıyla öneri listesine dahil edebildiğini işaret eder. Ancak, yüksek APLT değeri tek başına öneri kalitesinin iyi olduğunu garanti etmez, çünkü sistem az popüler içerikleri önerirken kullanıcı tercihlerini göz ardı edebilir eder [34].

3.2.6. Uzun Kuyruk Kapsamı

Uzun Kuyruk Kapsamı (LTC) metriği, öneri sisteminin tüm kullanıcılar için oluşturduğu öneri listelerinde yer alan az popüler içeriklerin genel dağılımını değerlendirilmektedir (Eşitlik 10):

$$LTC = \frac{|I_N \cap T|}{|T|} \quad (10)$$

LTC metriğinin yüksek değerler göstermesi, öneri sisteminin az popüler veya niş içerikleri daha kapsamlı bir şekilde öneri listelerine dahil edebildiğini işaret eder. Bu durum, sistemin sadece popüler içeriklere odaklanmak yerine, daha geniş bir içerik yelpazesini kullanıcılara sunabildiğini gösterir. Yüksek LTC

değerleri, öneri sisteminin çeşitlilik açısından daha başarılı olduğunu ve popülerlik yanlılığından daha az etkilendiğini ortaya koyar. Bu da kullanıcıların daha çeşitli ve kişiselleştirilmiş öneriler almasını sağlar [35].

3.2.7. Gini İndeksi Ölçütü

Gini İndeksi ölçütü, veri setindeki sınıfların dağılımını analiz ederek, bir düğümdeki verilerin ne kadar heterojen olduğunu ortaya koymaktadır.

$$Gini(D) = 1 - \sum {p_i}^2 \tag{11}$$

Gini değeri hesaplanırken Eşitlik 11'de, *D* ile gösterilen veri setindeki her bir sınıfın *(i)* dağılım oranı p_i değişkeni ile ifade edilir. Gini değerinin sıfır olması, incelenen düğümdeki tüm verilerin aynı sınıfa ait olduğu, yani tam bir homojenlik sağlandığını açıklayabilir [36].

3.2.8. Popülerlik Eğilimindeki Sapma

Popülerlik Eğilimindeki Sapma (Deviation in Popularity Inclination-DPI): Kullanıcıların mevcut yöntemlerle hesaplanan popülerlik eğilimleriyle yeni önerilen teknikler sonucunda elde edilen popülerlik eğilimleri arasındaki farkı normalize ederek ölçmektedir:

$$DPI_u = \frac{P_u - \varepsilon_u}{\varepsilon_u} \tag{12}$$

Eşitlik 12'de, ε_u mevcut yöntemlerle hesaplanan kullanıcı popülerlik eğilimi iken, P_u yeni yöntemle hesaplanan kullanıcı popülerlik eğilimini temsil etmektedir. DPI değerinin negatif çıkması, önerilen yöntemin kullanıcılara daha az popüler, dolayısıyla daha özelleştirilmiş ya da farklı içerikler sunduğunu göstermektedir [24].

3.2.9. Mutlak Popülerlik Eğilimindeki Sapma

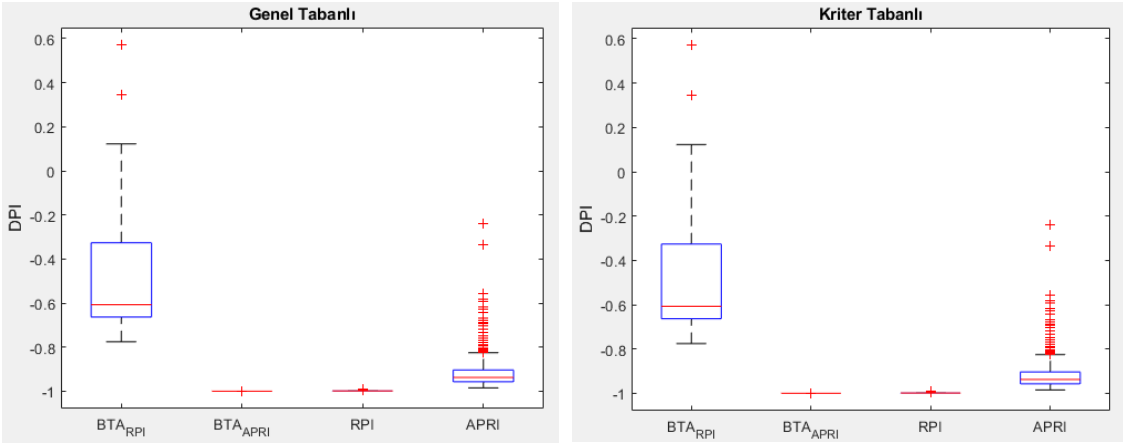
Mutlak Popülerlik Eğilimindeki Sapma (Absolute Deviation in Popularity Inclination-ADPI): Kullanıcıların mevcut yöntemlerle hesaplanan popülerlik eğilimleriyle yeni önerilen teknikler sonucunda elde edilen popülerlik eğilimleri arasındaki mutlak farkı normalize ederek ölçmektedir. Önerilen bu yöntemde popülerlik eğilim farkının artık pozitif/negatif yönü değil, sadece büyüklüğü dikkate alınır.

$$ADPI_u = \left|\frac{P_u - \varepsilon_u}{\varepsilon_u}\right| \tag{13}$$

Mutlak DPI değeri, klasik yöntemle önerilen içeriklerle yeni yöntemin önerileri arasındaki popülerlik eğilimi farkının büyüklüğünü göstermektedir. Bu ölçüt sayesinde sistemin öneri davranışındaki sapmaların yönünden bağımsız olarak, öneri çeşitliliği ve popülerlikten uzaklaşma düzeyi değerlendirilebilmektedir (Eşitlik 13).

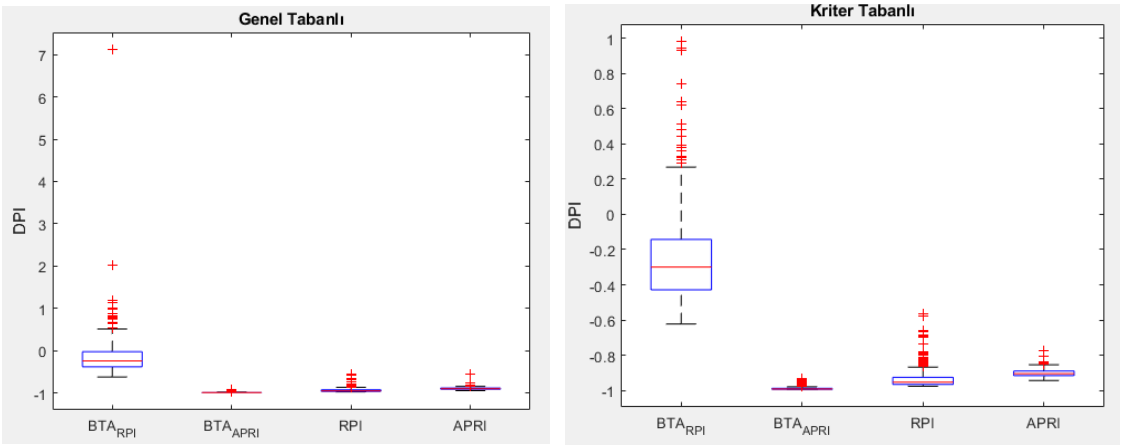
4. DENEYSEL SONUÇLAR

Çalışmanın temel amacı, öneri sistemlerinde kullanıcıların değerlendirme davranışlarındaki bağlamsal farklılıkları dikkate alarak popülerlik yanlılığını azaltmaya yönelik daha etkili ve kullanıcı odaklı bir yaklaşım geliştirmektir. Bu kapsamda, önerilen *RelPop* yöntemi ile literatürde yaygın olarak kullanılan BTA_{APRI}, BTA_{RPI}, *APRI* ve *RPI* yöntemleri, iki farklı çok ölçütlü veri kümesi (YM20 ve TA) üzerinde karşılaştırmalı olarak incelenmiştir. Analizlerde, geleneksel performans metriklerine ek olarak popülerlik yanlılığını daha hassas ölçen yeni bir ADPI metriği de kullanılmıştır. Kullanıcı popülerlik eğilimlerinin tespitine yönelik karşılaştırmalı analiz sonuçları Şekil 1-4'te sunulmuştur.



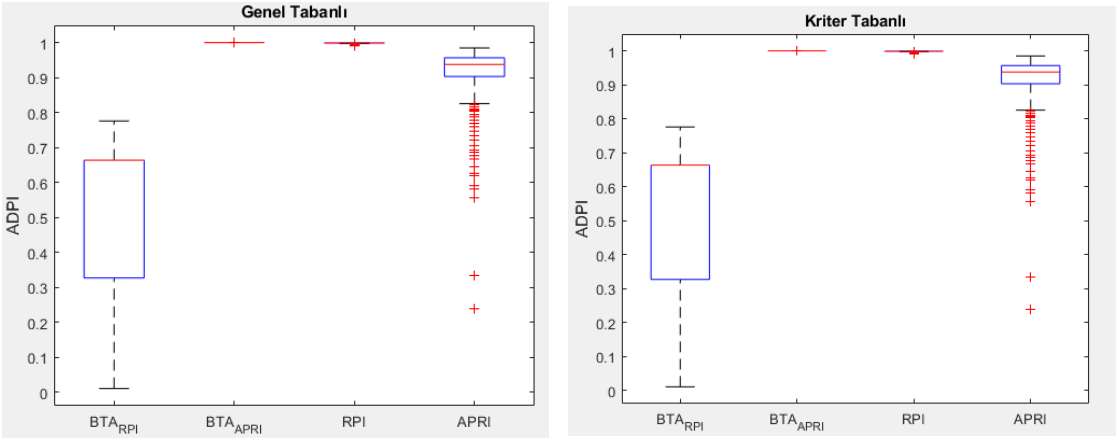
Şekil 1. TA veri seti üzerinde DPI analiz sonuçları

Şekil 1'de TA veri kümesi üzerinde gerçekleştirilen DPI analizleri sunulmaktadır. Genel tabanlı değerlendirmede, BTARPI yöntemi, negatif medyan değeri ve geniş dağılım aralığıyla dikkat çekmekte; bu durum yöntemin, kullanıcıya sunulan içerikleri daha az popüler hale getirme potansiyeline sahip olduğunu göstermektedir. Diğer taraftan, APRI yöntemi de negatif değerlere sahip olmakla birlikte, daha dar bir dağılım sergilemekte ve önerilerin popülerlikten sınırlı biçimde uzaklaştığını ortaya koymaktadır. RPI ve BTA_{APRI} yöntemleri ise DPI açısından neredeyse sıfıra yakın değerlere sahip olup, kullanıcılara sunulan içeriklerin mevcut popüler eğilimlerden anlamlı biçimde sapmadığını göstermektedir. Kriter tabanlı değerlendirmede de benzer bir eğilim gözlemlenmiş; BTA_{RPI} yöntemi en düşük medyan değeri ile öne çıkarken, APRI yöntemi negatif değerlere sahip olmakla birlikte daha sınırlı sapma göstermiştir. RPI ve BTA_{APRI} yöntemlerinin ise çeşitliliği sınırlı tutan daha sabit sonuçlar ürettiği görülmektedir.



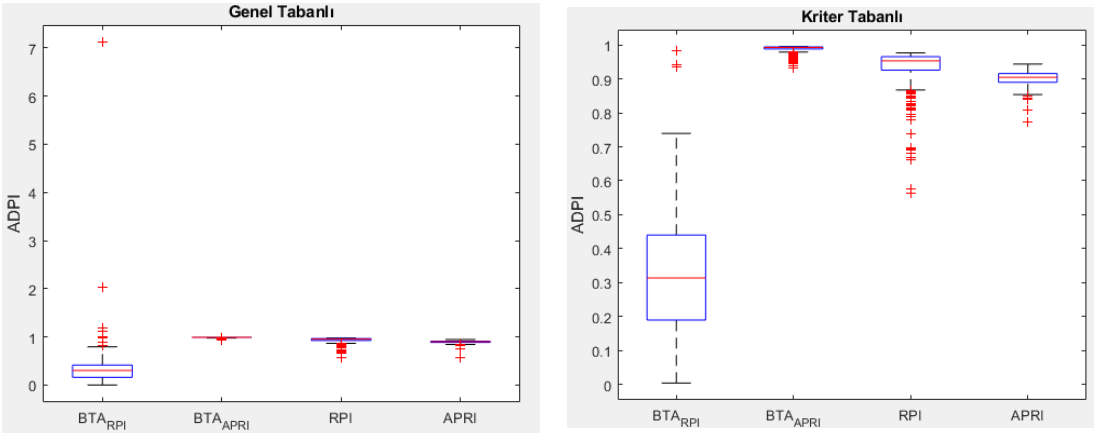
Şekil 2. YM20 veri seti üzerinde DPI analiz sonuçları

Şekil 2' de ise bir diğer veri seti olan YM20 üzerindeki sonuçlara yer verilmektedir. Genel tabanlı analizde, BTA_{RPI} yöntemi uç değerlerin etkisiyle geniş bir varyans aralığına sahiptir; bu durum, yöntemin bazı kullanıcılar için önerilen içeriklerin popülerlikten güçlü biçimde sapabildiğini göstermektedir. Buna karşın, RPI ve BTA_{APRI} yöntemleri, düşük varyans ve sıfıra yakın DPI değerleriyle daha tutarlı ancak popülerliğe yakın önerilerde bulunmaktadır. APRI yöntemi ise negatif değerler üretmesine rağmen öneri çeşitliliği açısından oldukça sınırlı bir etki göstermektedir. Kriter tabanlı değerlendirmede ise BTA_{RPI} yöntemi, yine popülerlik yanlılığından uzak öneriler sunarken, diğer yöntemler –özellikle BTA_{APRI} ve RPI– istikrarlı ancak oldukça benzer dağılımlar göstermiştir. Bu durum, klasik yöntemlerin kullanıcıya önerdiği içeriklerin popülerliğe olan bağımlılığının sürdüğünü ortaya koymaktadır.



Şekil 3. TA veri seti üzerinde ADPI analiz sonuçları

Şekil 3 incelendiğinde, mutlak sapma bağlamında BTA_{RPI} yöntemi, medyan açısından en düşük ADPI değerine sahip olup önerilerin, kullanıcıların genel eğilimlerinden sapma derecesinin kontrollü olduğunu göstermektedir. RPI ve BTA_{APRI} yöntemleri neredeyse sabit ve yüksek ADPI değerleriyle dikkat çekmektedir; bu durum, önerilerin kullanıcıların mevcut popülerlik eğilimlerinden anlamlı düzeyde sapmadığını göstermektedir. APRI yöntemi ise genel tabanlı analizde en yüksek ADPI değerine sahiptir; bu durum, önerilerin farklılaşma açısından göreceli olarak daha uzak olduğunu ancak bu farklılaşmanın rastlantısal biçimde oluşabileceğini düşündürmektedir. Kriter tabanlı sonuçlarda da benzer örüntü korunmuş, BTA_{RPI} göreli olarak daha düşük ADPI değerleri üretirken, diğer yöntemler sabit ve yüksek değerlere sahip olmuştur.

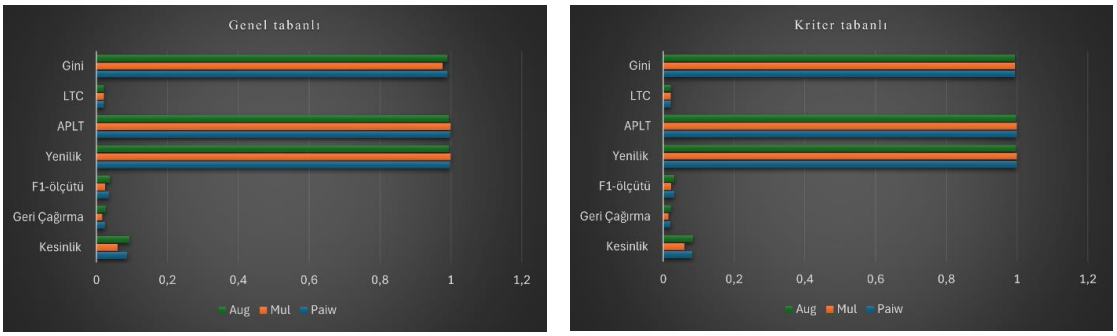


Şekil 4. YM20 veri seti üzerinde ADPI analiz sonuçları

Aynı performans metriğinin diğer veri setindeki sonuçları Şekil 4'te gösterilmektedir. Genel tabanlı analizde BTA_{RPI} yöntemi, ADPI değerlerinde düşük medyan ve geniş dağılım aralığı ile bazı kullanıcılar için önerileri ciddi biçimde özelleştirirken, bazıları için daha dengeli içerikler sunmaktadır. *RPI* ve BTA_{APRI} yöntemleri ise sabit ve yüksek ADPI değerleriyle, sistemin popülerlik merkezli çalıştığını göstermektedir. *APRI* ise yüksek medyan ve dar varyansla, önerilerin kullanıcıların popülerlik eğilimlerinden çok sapmadığını ortaya koymaktadır. Kriter tabanlı analizde de benzer bir görünüm sergilenmiş, BTA_{RPI} görece düşük sapma ile öne çıkarken, diğer yöntemler yüksek ve homojen sapma düzeylerinde kalmıştır.

Genel olarak, DPI ve ADPI analizleri öneri sistemlerinin popülerlik yanlılığı açısından farklı performanslar sergilediğini göstermektedir. BTA_{RPI}, her iki veri kümesinde de popülerlik eğiliminden saparak daha özelleştirilmiş öneriler sunma potansiyeli taşımaktadır. *APRI*, önerilerin popülerlikten uzaklaştığını gösterse de bu sapma genellikle sınırlı kalmıştır. *RPI* ve BTA_{APRI} ise popülerlik odaklı öneriler sunmaya devam etmekte, bu da çeşitliliğin artırılmasında yetersiz kalındığını göstermektedir. Sonuçlar, öneri sistemlerinde adil ve çeşitli öneriler için DPI ve ADPI metriklerinin birlikte değerlendirilmesinin önemini vurgulamaktadır. Özellikle BTA_{RPI} yöntemi, farklı kullanıcı grupları için popülerlik yanlılığını azaltmada etkili bir yaklaşım sunmaktadır.

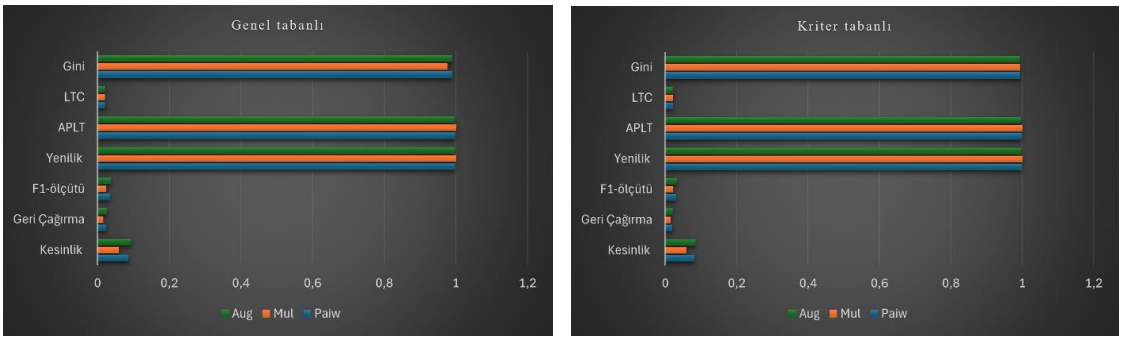
Çalışmanın ikinci aşaması olan, kullanıcıların popülerlik eğilimlerinde meydana gelen değişimlerin kapsamlı analizi Şekil 5-10 arasında gösterilmektedir.



Şekil 5. YM20 veri seti üzerinde ADPI analiz sonuçları

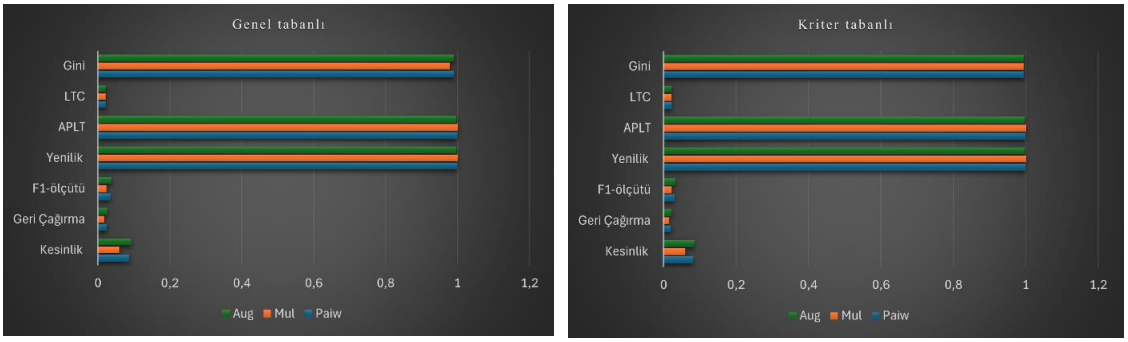
Şekil 5'te, YM20 veri seti üzerinde BTA_{RPI} yöntemiyle yapılan genel ve kriter tabanlı analizlerin sonuçları sunulmaktadır. Aug, Mul ve Paiw algoritmalarının performansları yedi farklı metrik üzerinden karşılaştırılmıştır. Her iki analizde de Gini, APLT ve yenilik metriklerinde tüm algoritmalar yüksek ve benzer değerlere ulaşmıştır; bu da öneri sistemlerinin çeşitli ve yenilikçi içerikler sunmada başarılı olduğunu göstermektedir. Yüksek Gini indeksi, öneri listelerinin kullanıcılar arasında dengeli ve adil dağıldığını, popülerlik yanlılığının azaldığını göstermektedir. Buna karşılık, LTC metriği tüm algoritmalar için düşük kalmış, sistemlerin uzun kuyruktaki niş içerikleri önerme oranının sınırlı olduğunu ortaya koymuştur.

Doğruluk metrikleri olan kesinlik, geri çağırma ve F1-ölçütü ise tüm algoritmalarda düşük çıkmıştır; bu da sistemlerin isabetli ve ilgili öneriler sunmada yetersiz kaldığını göstermektedir. Genel ve kriter tabanlı analizler karşılaştırıldığında, metriklerin dağılımı ve algoritmaların sıralaması büyük ölçüde benzerdir. Sonuç olarak, Aug, Mul ve Paiw algoritmaları çeşitlilik ve yenilik açısından başarılı olsa da, doğruluk ve uzun kuyruk kapsamı açısından geliştirmeye ihtiyaç duymaktadır. Bu analizler, öneri sistemlerinin yalnızca doğruluk değil, aynı zamanda yeni ve çeşitli içerik sunma kapasitesiyle de değerlendirilmesi gerektiğini göstermektedir.



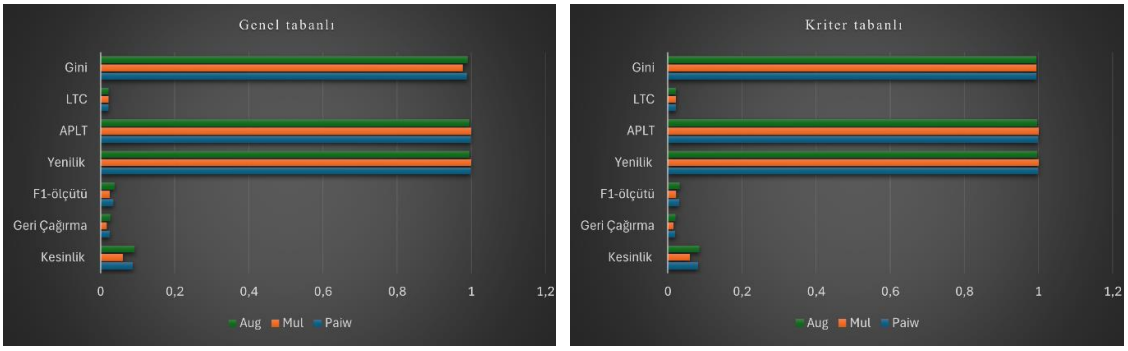
Şekil 6. YM20 veri seti üzerinde BTA_{APRI} analiz sonuçları

Şekil 6'da, aynı veri setinin BTA_{APRI} yöntemiyle yapılan analiz sonuçları sunulmaktadır. Grafikler, Aug, Mul ve Paiw algoritmalarının Gini, APLT ve yenilik metriklerinde 0,95'in üzerinde, yüksek değerlere ulaştığını göstermektedir. Özellikle yenilik değerinin 0,98'e yaklaşması, öneri listelerinin büyük ölçüde yeni ve kullanıcı için keşfedilmemiş ürünlerden oluştuğunu ortaya koymaktadır. Buna karşılık, LTC metriğinin 0,03'ün altında kalması, öneri listelerinde niş içeriklerin sınırlı olduğunu göstermektedir. Doğruluk metriklerinde ise kesinlik yaklaşık 0,04, geri çağırma ve F1-ölçütü ise 0,01'in üzerindedir; bu da sistemlerin isabetli ve kişiselleştirilmiş öneriler sunmada yetersiz kaldığını göstermektedir. Genel ve kriter tabanlı analizler benzer sonuçlar vermiş, yalnızca kriter tabanlı analizde APLT ve yenilikte küçük artışlar gözlenmiştir. Sonuç olarak, Aug, Mul ve Paiw algoritmaları çeşitlilik ve yenilikte başarılı olsa da, doğruluk ve uzun kuyruk kapsamı açısından geliştirmeye ihtiyaç duymaktadır. Özellikle LTC ve doğruluk metriklerindeki düşük değerler, sistemlerin niş içeriklere ve kullanıcıya özel önerilere daha fazla odaklanması gerektiğini göstermektedir.



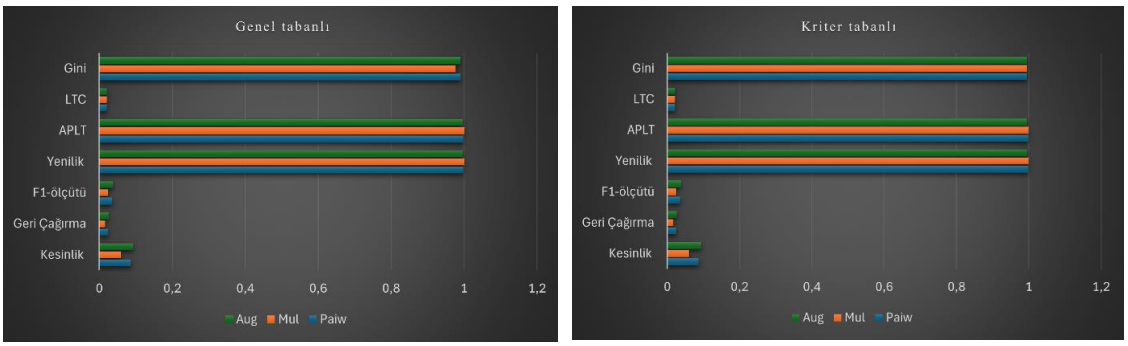
Şekil 7. YM20 veri seti üzerinde *RPI* analiz sonuçları

Şekil 7'de, Aug, Mul ve Paiw algoritmalarının Gini, APLT ve yenilik metriklerinde 0,95'in üzerinde ve birbirine yakın değerlere ulaştığı görülmektedir; bu da sistemlerin yenilikçi ve çeşitli içerikler sunmada başarılı olduğunu göstermektedir. LTC metriğinde ise Mul algoritması, diğerlerine göre daha yüksek bir değere ulaşarak uzun kuyrukta yer alan içerikleri önerme konusunda kısmi bir avantaj sağlamıştır. Doğruluk metrikleri genel olarak düşük kalırken, Aug algoritması ise kesinlik ve F1-ölçütlerinde diğerlerine göre daha iyi performans göstermiştir. Kriter tabanlı analizde ise APLT ve yenilikte küçük artışlar gözlenmiş, bu da kriter bazlı yaklaşımın çeşitlilik ve yenilik üretimini bir miktar artırdığını göstermektedir. Sonuç olarak, RPI yöntemi altında algoritmalar yenilik ve çeşitlilikte güçlü performans sergilerken, Mul uzun kuyruk kapsamı, Aug ise doğruluk metriklerinde öne çıkmıştır.



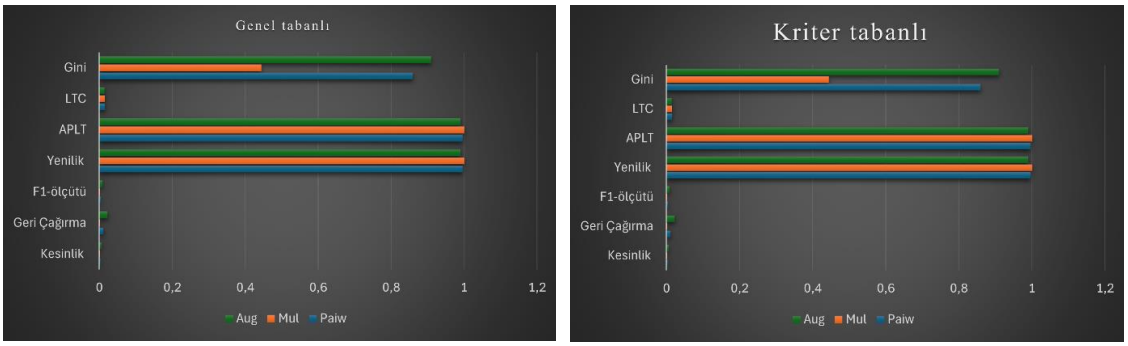
Şekil 8. YM20 veri seti üzerinde *APRI* analiz sonuçları

Şekil 8'de, YM20 veri setinde APRI yöntemiyle yapılan analizlerde Aug, Mul ve Paiw algoritmaları Gini, APLT ve yenilik metriklerinde yüksek değerler elde ederek çeşitlilik ve yenilikte başarılı olmuştur. Ancak, LTC ve doğruluk metrikleri tüm algoritmalarda düşük kalmış, özellikle kesinlik değerlerinin 0,1'in altında olması sistemlerin isabetli önerilerde yetersiz olduğunu göstermiştir. Kriter tabanlı ve genel analizler arasında anlamlı bir fark görülmemiş, APRI yöntemi tutarlı bir performans sergilemiştir. Sonuç olarak, algoritmalar yenilik ve çeşitlilikte güçlü, doğruluk ve uzun kuyruk kapsamı açısından ise geliştirmeye açıktır.



Şekil 9. YM20 veri seti üzerinde *RelPop* analiz sonuçları

Şekil 9'da, RelPop yöntemiyle YM20 veri setinde algoritmaların genel ve kriter tabanlı performansları karşılaştırılmıştır. Tüm algoritmalar Gini, APLT ve yenilik metriklerinde üst sınıra yakın değerlere ulaşarak, RelPop'un hem çeşitliliği hem de öneri özgünlüğünü koruduğunu göstermiştir. Gini değerlerinin yüksek ve birbirine yakın olması, önerilerin kullanıcılar arasında dengeli dağıldığını işaret etmektedir. LTC metriği ise orta düzeyde kalarak, RelPop'un niş içerikleri dengeli şekilde önerdiğini göstermektedir. Doğruluk metriklerinde Aug algoritması, kesinlik ve F1'de diğerlerine göre daha tutarlı sonuçlar vermiştir. Kriter tabanlı analizde ise yenilik ve APLT'deki küçük artışlar, kullanıcı tercih ağırlıklarının öneri kalitesine olumlu katkı sağladığını göstermektedir. Sonuç olarak, RelPop yöntemiyle algoritmalar, çeşitlilik ve kullanıcı odaklılıkta dengeli ve tatmin edici bir performans sergilemiş, önerilerde denge odaklı bir yaklaşım sunmuştur.



Şekil 10. TA veri seti üzerinde *BTA_{PRI}, BTA_{APRI}, RPI, APRI ve RelPop* analiz sonuçları

Şekil 10'da, TA veri seti üzerinde farklı yöntemlerle yapılan genel ve kriter tabanlı analizlerin sonuçları sunulmuştur. Tüm algoritmalar Gini, APLT ve yenilik metriklerinde yüksek skorlar elde ederek, kullanıcıya çeşitli ve özgün içerikler sunmada başarılı olmuştur. Ancak, LTC metriği tüm yöntemlerde düşük kalmış, sistemlerin uzun kuyruktaki niş içerikleri önermede yetersiz olduğunu göstermiştir. Doğruluk metrikleri de genel olarak düşük çıkmış, özellikle kesinlik ve geri çağırma oranlarının zayıf olması, sistemlerin isabetli öneriler sunmakta zorlandığını ortaya koymuştur.

Genel ve kriter tabanlı analizlerde benzer sonuçlar alınması ise, TA veri setinin seyrek yapısından kaynaklanmaktadır. TA veri seti, çalışmadaki en seyrek veri seti olduğundan, kullanıcıya kişisel öneriler sunabilmek için kullanılan oylama değerleri oldukça sınırlıdır. Bu durum, yöntemlerin tüm metriklerde aynı değerleri üretmesine yol açmış ve algoritmalar arası farkların belirginleşmesini engellemiştir. Sonuç olarak, yöntemler çeşitlilik ve yenilikte başarılı, ancak doğruluk ve uzun kuyruk kapsamı açısından sınırlı kalmıştır; veri setinin seyrekliği ise performans farklarının ortaya çıkmasını zorlaştırmıştır.

5. SONUÇ VE ÖNERİLER

Bu çalışmada, öneri sistemlerinde popülerlik yanlılığını azaltarak kullanıcıya daha çeşitli, adil ve bağlama duyarlı içerikler sunmayı amaçlayan RelPop yöntemi ile bu amaca yönelik geliştirilen yeni bir değerlendirme metriği olan ADPI önerilmiştir. RelPop, kullanıcıların mutlak puanlamalarından ziyade kendi puanlama geçmişleri içindeki göreli eğilimlerine odaklanarak, kişisel anlamı daha etkili bir şekilde modellemektedir. Böylece öneri listelerinde hem çeşitliliğin hem de kişiselleştirme kalitesinin artırılması hedeflenmiştir. YM10 ve TA veri setleri üzerinde yürütülen deneysel çalışmalar sonucunda, RelPop yönteminin özellikle yenilik, APLT ve Gini endeksi gibi çeşitlilik ve adalet odaklı metriklerde yüksek performans sergilediği gözlemlenmiştir. Ayrıca, RelPop'un Aug algoritmasıyla birlikte kullanıldığında, az bilinen ve potansiyel olarak kullanıcıya uygun içeriklerin öneri listelerine daha dengeli bir şekilde dahil edildiği görülmüştür. Bu durum, RelPop'un popüler içeriklerin baskınlığını azaltarak daha kapsayıcı ve kullanıcı odaklı öneriler sunduğunu göstermektedir. Özellikle düşük etkileşimli kullanıcılar veya farklı kriterlere göre değerlendirme yapan kullanıcılar açısından da sistemin etkili sonuçlar ürettiği belirlenmiştir. Ancak, kesinlik, geri çağırma ve F1-ölçütü gibi klasik doğruluk metriklerinde RelPop'un geleneksel yöntemlere benzer sınırlılıkları bulunmaktadır. Ayrıca, TA veri setinin seyrek yapısı ve YM20 veri setindeki kullanıcı sayısının azlığı, yöntemin potansiyelinin tam olarak ortaya konmasını kısıtlamıştır.

Performans değerlendirmesi kapsamında kullanılan DPI ve önerilen ADPI metrikleri, öneri sistemlerinin popülerlik yanlılığından ne ölçüde saptığını değerlendirmede birbirini tamamlayıcı niteliktedir. DPI,

sapmanın yönünü ve büyüklüğünü gösterirken; ADPI, sapmanın yönünden bağımsız olarak büyüklüğünü ölçmektedir. DPI metriklerinin kullanıcı odaklı çeşitlilik analizlerinde; ADPI'nın ise farklı yöntemler arasındaki genel performans farklarını karşılaştırmada daha uygun olduğu sonucuna ulaşılmıştır.

Bu kapsamda elde edilen bulgular doğrultusunda aşağıdaki öneriler sunulabilir:

- Özellikle seyrek veri setlerinde kullanıcı-öğe etkileşimlerinin artırılması, RelPop'un bağlamsal analiz gücünü daha da geliştirecektir.
- RelPop'un çeşitlilik ve adalet açısından sunduğu avantajlar, doğruluk odaklı algoritmalarla birleştirilerek hibrit öneri sistemleri tasarlanabilir.
- Gelecek çalışmalarda farklı veri setleriyle benzer analizler yapılarak yöntemin genellenebilirliği ve sınırlılıkları daha kapsamlı biçimde değerlendirilebilir.

Bununla birlikte, yöntemin genellenebilirliğini artırmak için yalnızca film ve otel gibi belirli alanlarla sınırlı kalınmaması, farklı türde veri setlerinde (ör. müzik, e-ticaret, sosyal medya) test edilmesi önemlidir. Veri seti sınırlılıklarının azaltılması amacıyla sentetik veri üretme veya transfer learning tabanlı yaklaşımlar gelecekte değerlendirilebilir. Ayrıca RelPop'un, çok ölçütlü sistemlerde her bir kriter için ayrı ayrı veya kriterlerin ağırlıklandırılması yoluyla hesaplanması, daha rafine kişiselleştirme imkânı sağlayarak yöntemin çok ölçütlü yapılara özgü avantajını ortaya koyacaktır. Bu tür genişletmeler, RelPop'un yalnızca popülerlik yanlılığını azaltmakla kalmayıp, aynı zamanda kullanıcıların bireysel kriter önceliklerini de dikkate alan daha esnek ve kapsamlı öneri sistemlerinin geliştirilmesine katkı sunacaktır.

6. KAYNAKLAR

1. Zhang, S., Yao, L., Sun, A. & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys, 52*(1), 1-38.
2. Gomez-Uribe, C. A. & Hunt, N. (2016). The Netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems, 6*(4), 1-19.
3. Ricci, F., Rokach, L. & Shapira, B. (2022). Recommender systems: Techniques, applications, and challenges. *In F. Ricci, L. Rokach & B. Shapira (Eds.), Recommender Systems Handbook* (1-35). Cham: Springer.
4. Schwartz, B. (2016). *The paradox of choice: Why more is less* (Rev. ed.). New York: Harper Perennial.
5. Jannach, D. & Adomavicius, G. (2017). Price and profit awareness in recommender systems. *In Proceedings of the 11th ACM Conference on Recommender Systems* (192-200).
6. Koren, Y. & Bell, R. (2015). Advances in collaborative filtering. *In F. Ricci, L. Rokach & B. Shapira (Eds.), Recommender Systems Handbook* (77-118). Cham: Springer.
7. Adomavicius, G. & Kwon, Y. (2015). Multi-criteria recommender systems. *In F. Ricci, L. Rokach & B. Shapira (Eds.), Recommender Systems Handbook* (847-880). Cham: Springer.
8. Baltrunas, L. & Ricci, F. (2014). Experimental evaluation of context-dependent collaborative filtering using item splitting. *User Modeling and User-Adapted Interaction, 24*(1-2), 7-34.
9. Kalkan, İ.E. ve Şahin, C. (2023). Çapraz satışı destekleyebilecek transformer ile geliştirilmiş bir öneri sistemi. *Çukurova Üniversitesi Mühendislik Fakültesi Dergisi, 38*(2), 571-584.
10. Liu, H., Wang, Y. & Chen, X. (2021). Multi-criteria recommendation with user preference learning. *In Proceedings of the 15th ACM Conference on Recommender Systems* (456-465).
11. Jannach, D. & Adomavicius, G. (2016). Recommendations with a purpose. *In Proceedings of the 10th ACM Conference on Recommender Systems* (7-10).
12. Koşar, O. ve Atak, M. (2023). Bulut bilişim sanal sunucu ürün seçiminde çok kriterli bir karar destek modeli. *Çukurova Üniversitesi Mühendislik Fakültesi Dergisi, 38(*4), 939-953.
13. Baltrunas, L., Ludwig, B. & Ricci, F. (2011). Matrix factorization techniques for context-aware recommendation. *In Proceedings of the 5th ACM Conference on Recommender Systems* (301-304).
14. Wang, X., He, X., Wang, M., Feng, F. & Chua, T. S. (2019). Neural graph collaborative filtering. *In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (165-174).
15. Steck, H. (2018). Calibrated recommendations. *In Proceedings of the 12th ACM Conference on Recommender Systems* (154-162).

16. Ekstrand, M.D., Azpiazu, I.M., Anuyah, O., McNeill, D., Ekstrand, J.D. & Pera, M.S. (2018). All the cool kids, how do they fit in?: Popularity and demographic biases in recommender evaluation and effectiveness. *In Proceedings of the Conference on Fairness, Accountability and Transparency* (172-186).
17. Abdollahpouri, H., Mansoury, M., Burke, R. & Mobasher, B. (2020). The connection between popularity bias, calibration, and fairness in recommendation. *In Proceedings of the 14th ACM Conference on Recommender Systems* (726-731).
18. Klimashevskaia, A., Abdollahpouri, H. & Mobasher, B. (2022). Popularity bias in collaborative filtering: A survey. *In Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization* (1–11).
19. Kowald, D. & Lacic, E. (2022). Popularity bias in different domains: A multi-domain analysis. *In Proceedings of the 16th ACM Conference on Recommender Systems* (1-11).
20. Abdollahpouri, H., Burke, R. & Mobasher, B. (2019). The unfairness of popularity bias in recommendation. *In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (148-154).
21. Steck, H. (2011). Item popularity and recommendation accuracy. *In Proceedings of the 5th ACM Conference on Recommender Systems* (125-132).
22. Wei, D., Wang, X., Li, Y. & He, X. (2021). Causal inference for recommender systems. *In Proceedings of the 30th ACM International Conference on Information and Knowledge Management* (3043-3049).
23. Mansoury, M., Abdolllahpouri, H., Smith, J.,Dehpanah, A., Pechenizkiy, M., Mobasher, B. (2019). Investigating potential factors associated with gender discrimination in collaborative recommender systems. *Information Processing & Management, 57*(2), 102371.
24. Taçlı, Y., Yalçın, E. & Bilge, A. (2022). Novel approaches to measuring the popularity inclination of users for the popularity bias problem. *In Proceedings of the 6th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)* (555-560).
25. Lee, J., Lee, D., Lee, Y., Hwang, W. & Kim, S. (2016). Improving the accuracy of top-n recommendation using a preference model. *Information Sciences, 348*, 290-304.
26. Yalcin, E. & Bilge, A. (2021). Investigating and counteracting popularity bias in group recommendations. *Information Processing & Management, 58*(5), 102608.
27. Zhang, Y., Chen, X., Ai, Q., Croft, W. B. & Guo, J. (2017). Towards conversational search and recommendation: System ask, user respond. *In Proceedings of the 27th ACM International Conference on Information and Knowledge Management* (177-186).
28. Abdollahpouri, H., Burke, R. & Mobasher, B. (2017). Controlling popularity bias in learning-to-rank recommendation. *In Proceedings of the 11th ACM Conference on Recommender Systems (RecSys)* (42-46).
29. Lakiotaki, K., Matsatsinis, N.F. & Tsoukias, A. (2011). Multicriteria user modeling in recommender systems. *IEEE Intelligent Systems, 26*(2), 64-76.
30. Wang, H., Lu, Y. & Zhai, C. (2011). Latent aspect rating analysis without aspect keyword supervision. *In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (618-626).
31. Herlocker, J.L., Konstan, J.A., Terveen, L.G. & Riedl, J.T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS), 22*(1), 5-53.
32. Bellogín, A., Castells, P. & Cantador, I. (2011). Precision-oriented evaluation of recommender systems: An algorithmic comparison. *In Proceedings of the 5th ACM Conference on Recommender Systems* (333-336).
33. Vargas, S. & Castells, P. (2011). Rank and relevance in novelty and diversity metrics for recommender systems. *In Proceedings of the 5th ACM Conference on Recommender Systems* (109-116).
34. Park, Y.J. & Tuzhilin, A. (2008). The long tail of recommender systems and how to leverage it. *In Proceedings of the 2008 ACM Conference on Recommender Systems* (11-18).
35. Burke, R. (2002). Hybrid recommender systems: Survey and experiments. User *Modeling and User-Adapted Interaction, 12*(4), 331-370.
36. Breiman, L., Friedman, J., Stone, C.J. & Olshen, R.A. (1984). *Classification and regression trees.* Chapman & Hall/CRC.