BŞEÜ Fen Bilimleri Dergisi
8(1), 279-291, 2021

BSEU Journal of Science
https://doi.org/10.35193/bseufbd.884634

e-ISSN: 2458-7575 (https://dergipark.org.tr/tr/pub/bseufbd)

Araştırma Makalesi - Research Article

İşbirlikçi Filtreleme Algoritmalarının Çok-Beğenilen Ürünlere Yönelik Yanlılığı

Collaborative Filtering Algorithms' Bias Towards Highly-liked Items

Emre Yalçın[1*]

Geliş / Received: 22/02/2021 ***Revize / Revised: 20/03/2021*** ***Kabul / Accepted: 26/03/2021***

ÖZ

Öneri sistemleri, bireysel kullanıcılara herhangi bir kişisel çaba gerektirmeden geçmişteki tercihlerine ve özelliklerine göre uygun ürünleri/hizmetleri öneren otomatikleştirilmiş araçlardır. Bu sistemlerde, işbirlikçi filtreleme algoritmaları, ürünler için bireysel tahminler veya kullanıcılar için tercih edilir ürünlerin sıralı bir listesini üretmek için en çok kullanılan yaklaşımlardır. Bu tür algoritmaların verimliliği genellikle sağlanan önerilerin doğruluğu ile değerlendirilse de, ürün kataloğu kapsamı gibi doğruluk-üstü değerlendirmeler de nitelikli önerilerde kritik faktörler olarak kabul edilir. Ancak, son zamanlarda yapılan birçok çalışma, bu algoritmaların, belirli özellikleri (örn. popülerlik) nedeniyle bazı ürünleri üretilen sıralı listelerde diğerlerinden daha çok öne çıkarma eğiliminde olduğunu göstermiştir. Bu çalışmada, ürün profillerini farklı bir bakış açısıyla, beğenilme dereceleriyle irdeliyor ve işbirlikçi filtreleme algoritmalarının çok beğenilen ürünlere yönelik bir yanlılığının olup olmadığını araştırıyoruz. Bu amaçla, üç farklı kategoriden dokuz önemli işbirlikçi filtreleme algoritmasını kullanıyoruz ve iki gerçek-dünya veri kümesi üzerinde çeşitli deneyler gerçekleştiriyoruz. Deneysel sonuçlar, hemen hemen tüm algoritmaların çok beğenilen ürünlere yönelik güçlü bir yanlılığının olduğunu ve SVD ile SVD++ gibi matris çarpanlarına ayırma tabanlı algoritmaların yüksek kalitede öneriler üretmede diğerlerinden daha başarılı olduğunu göstermiştir.

Anahtar Kelimeler- Öneri Sistemleri, İşbirlikçi Filtreleme, Algoritmik Yanlılık, Ürün Profili, Katalog Kapsamı

ABSTRACT

Recommender systems are automated tools that suggest appropriate products/services to individual users based on their preferences in the past and characteristics without requiring any personal effort. In these systems, collaborative filtering algorithms are the most utilized approaches to produce individual predictions or a ranked list of preferable items for users. Although such algorithms' efficiency is generally assessed with the accuracy of provided recommendations, beyond-accuracy evaluations such as item catalog coverage are also considered critical factors in qualified recommendations. However, many recent studies demonstrate that these algorithms tend to feature certain items than others in the produced ranked lists because of their specific properties (e.g., popularity). In this study, we scrutinize item profiles with a different point of view, the degrees of being liked, and investigate whether there is any bias of collaborative filtering algorithms towards highly-liked items or not. To this end, we adopt nine prominent collaborative filtering algorithms in three different categories and perform various experiments on two real-world datasets. The experimental results demonstrate that almost all algorithms

[1]*Sorumlu yazar iletişim: eyalcin@cumhuriyet.edu.tr (https://orcid.org/0000-0003-3818-6712)
Bilgisayar Mühendisliği Bölümü, Sivas Cumhuriyet Üniversitesi, Mühendislik Fakültesi, Sivas, Türkiye

BSEU Journal of Science
https://doi.org/10.35193/bseufbd.884634

e-ISSN: 2458-7575 (https://dergipark.org.tr/tr/pub/bseufbd)

are strongly biased towards highly-liked items, and matrix factorization based algorithms such as SVD and SVD++ are more successful than others in producing high-quality recommendations.

Keywords- Recommender Systems, Collaborative Filtering, Algorithmic Bias, Item Profile, Catalog Coverage

I. GİRİŞ

Günümüzde, hızla artan İnternet kullanımına bağlı olarak, bireylerin alışveriş yapmak, film izlemek, müzik dinlemek, haber takip etmek, konaklama rezervasyonu yapmak gibi çeşitli günlük aktiviteyi gerçekleştirebilecekleri birçok faydalı çevrimiçi sistem geliştirilmiştir. İnternet gelişimine paralel şekilde sürekli geliştirilen ve çeşitlendirilen bu sistemler kullanıcılarına çeşitli kolaylıklar sağlasa da bazı zorlukların ortaya çıkmasına neden olmaktadır. Bu zorluklardan en önemlisi, hızlı artan İnternet kullanımına bağlı olarak işlenmesi ve depolanması gereken verilerin miktarının dramatik bir şekilde artmasıdır. Veri miktarındaki bu artış, aşırı bilgi yükleme sorunu olarak bilinen soruna neden olarak, bireylerin ihtiyaç duyduğu hizmetlere/ürünlere erişiminin zorlaşmasına ve karar verme süreçlerinin daha karmaşık bir hale gelmesine neden olur [1]. Öneri sistemleri, kullanıcıların İnternet ortamında erişmeye çalıştıkları ilgili ve değerli bilgileri keşfedip onların karar verme süreçlerini destekleyerek aşırı bilgi yükleme sorununun üstesinden gelinmesine yardımcı olan yazılım araçlarıdır [2]. Bu sistemler, günümüzde birçok çevrimiçi sistemde e-ticaret (eBay, Amazon vb.), film (Netflix, Amazon Prime Video vb.), müzik (Spotify, iTunes vb.), video (YouTube, Instagram vb.), haber (Google News vb.) ve konaklama (Booking vb.) gibi farklı amaçlar doğrultusunda etkin bir şekilde kullanılmaktadır.

Geleneksel öneri sistemlerinde temel amaç, kullanıcıların geçmişte çeşitli hizmetler/ürünler için yapmış olduğu tercihlerden yola çıkarak onların ilgisini çekebilecek ancak henüz deneyimlemedikleri hizmetleri/ürünleri kapsayan kişiselleştirilmiş tavsiyeler üretebilmektedir [3]. Bu amaç doğrultusunda birçok çevrimiçi sistem, İşbirlikçi Filtreleme (İF) olarak adlandırılan ve bireysel tavsiye üretme sürecinde oldukça başarılı sonuçlar elde edilmesini sağlayan algoritmaları kullanmaktadır [4]. Bu algoritmalar temel olarak, geçmişte benzer tercihler yapan kullanıcıların gelecekte de benzer tercihler yapma eğiliminde olduğu varsayımına dayanmaktadır ve kullandığı yaklaşımlar temelinde hafıza-tabanlı, model-tabanlı ve bunların birlikte kullanıldığı melez yaklaşımlar olarak üç kategoride değerlendirilmektedir [5]. Tipik bir İF algoritması, m adet kullanıcının n adet ürün için sağladığı derecelendirmeleri içeren $m \times n$ boyutunda, genellikle oldukça büyük ve boşluklu bir yapıda olan *kullanıcı* × *ürün* matrisi üzerinde işlem yaparak kullanıcıların henüz deneyimlemedikleri ürünler için tahmin değerleri hesaplamaya veya ilgisini çekebilecek ürünleri içeren sıralı öneri listeleri üretmeye odaklanmaktadır [6].

İF algoritmaları *kullanıcı* × *ürün* derecelendirme matrisi üzerinde eğitildikleri için, içerisinde bulunan tercihlerin karakteristik özelliklerinden etkilenerek ürünler açısından adil olmayan önerilerin üretilmesine neden olabilmektedir [7]. Bu durumun temel nedeni, bu matrislerin doğası gereği ürünler arasında eşit olmayan bir şekilde dağılmış kullanıcı tercihlerini içermesidir. Örneğin, bazı popüler ürünler yüksek sayıda kullanıcı tarafından değerlendirilirken, diğer ürünler yalnızca birkaç derecelendirme almaktadır. Derecelendirme dağılımındaki bu tür bir eşitsizlik, İF algoritmalarının ürettiği önerilerde, popüler ürünlerin diğerlerine (uzun-kuyruktaki ürünlere) kıyasla daha fazla öne çıkmasına neden olmaktadır [8-10]. Böylece, popüler olan ürünler üretilen öneri listelerinde daha fazla görünürlük kazanmakta ve bu da ilgili ürünlerin tüketilme oranlarını daha da arttırmaktadır. Öte yandan, uzun-kuyrukta yer alan ürünler, özellikle sisteme yakın zamanda eklenenler, ilgi çekici olsalar bile üretilen önerilerde hak ettikleri ilgiyi görememektir. Böyle bir yanlılığın farkında olunması, ayrıca, sağlayıcıların kendi ürünlerini öne çıkarmak amacıyla kullanıcılara derecelendirme sağlamaları için çeşitli rüşvetler vermelerine bile neden olabilmektedir [11]. İF algoritmalarının bu tür içsel yanlılıkları literatürde popülerlik-yanlılığı olarak adlandırılan soruna neden olarak "zengin daha zenginleşir" etkisine yol açmaktadır [12].

Her ne kadar İF algoritmalarının başarısı genellikle ürettikleri önerilerin doğruluğu ile analiz edilse de, son birkaç yılda, önerilen ürünlerin yeni olup olmadığı ve ürün kataloğunu iyi kapsayıp kapsamadığı gibi doğruluk-ötesi diğer unsurların, genel öneri kalitesi üzerinde olumlu bir etkisinin olduğu gösterilmiştir [13]. Dolayısıyla, ürünlerin popülerlikleri dışında, derecelendirme profillerindeki özellikler daha ayrıntılı bir biçimde ele alınarak, öneri sonuçlarında neden olabilecekleri muhtemel ayrımcılıklar üzerine daha fazla araştırma yapılması gerekmektedir. Bazı ürünlerin diğer ürünlere kıyasla kullanıcılar tarafından daha çok beğenildiği gerçeği göz önünde bulundurularak, bu çalışmada, bir ürünün kullanıcılar tarafından aldığı derecelendirmelerin ortalamaları, yani kullanıcılar tarafından beğenilme oranlarının, önerilen sıralı ürün listelerinde bir yanlılığa neden olup olmadığı analiz edilmektedir. Diğer bir deyişle, bu çalışmanın temel motivasyonu, öneri algoritmalarının ürettiği

BSEU Journal of Science
https://doi.org/10.35193/bseufbd.884634

e-ISSN: 2458-7575 (https://dergipark.org.tr/tr/pub/bseufbd)

sıralı ürün listelerinde çok beğenilen ürünler lehinde herhangi bir yanlılığın olup olmadığının araştırılmasıdır. Bu çalışmanın literatüre sağladığı temel katkılar aşağıda özetlenmiştir.

1. Öneri sistemlerinde ortaya çıkan farklı yanlılık türlerinin derin bir analizi sunulmaktadır.
2. Öneri sistemleri çalışmalarında sıklıkla kullanılan film alanındaki iki popüler veri kümesinde ürünlerin beğenilme oranları bağlamında bir eşitsizliğin olup olmadığı keşfedilmektedir.
3. Literatürde öne çıkan dokuz popüler İF algoritması ile üretilen sıralı öneri listelerinde, yüksek beğeniye sahip ürünlere karşı bir yanlılığın olup olmadığı analiz edilmektedir. Ayrıca, bu algoritmaların performansları, çeşitli doğruluk metrikleri ile değerlendirilerek analiz edilmektedir.

Bu çalışmanın geri kalanı şu şekilde organize edilmiştir: Sonraki bölümde, bireysel öneri üretme sürecinde ortaya çıkan yanlılık türleri açıklanarak literatürde öne çıkan çalışmaları özetlenmektedir. Bölüm 3'te bu çalışmanın kapsamını ortaya koyan motivasyon örneği ve çalışmanın yöntemi anlatılmaktadır. Bölüm 4'de ise gerçekleştirilen deneysel çalışmalar ayrıntılı şekilde anlatılmıştır. Son olarak Bölüm 5'te elde edilen bulgular tartışılarak gelecekte yapılacak çalışmalar açıklanmıştır.

II. İLGİLİ ÇALIŞMALAR

Bu bölüm, öneri sürecinde ortaya çıkan yanlılık türlerinin daha iyi anlaşılabilmesi için iki alt-bölüme ayrılmıştır. İlk alt-bölümde bir öneri mekanizmasının kullanıcılara öneri üretme sürecinde gerçekleştirdiği temel işlem adımlarını içeren geri bildirim döngüsü açıklanmıştır. İkinci alt-bölümde ise, öneri üretme sürecinde ortaya çıkan yanlılıklar ayrıntılı bir şekilde açıklanmış ve literatürde bu yanlılıkları ele alan çalışmalara yer verilmiştir.

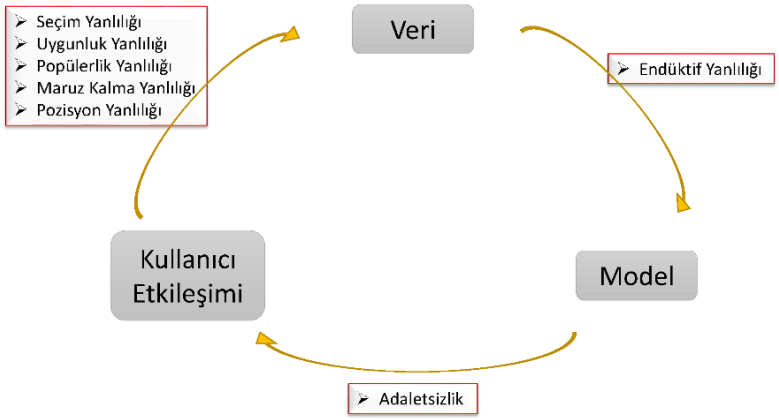
A. Önerilerde Geri Bildirim Döngüsü

Bir öneri sisteminde, kullanıcılar için tavsiyelerin üretilme süreci, üç temel bileşen, *Kullanıcı*, *Veri* ve *Model*, arasında bir geri bildirim döngüsü olarak düşünülebilir [10]. Şekil 1'de gösterildiği gibi, bu geri bildirim döngüsü üç temel aşamadan oluşur:

- *Kullanıcı* → *Veri*: Bu aşama, kullanıcı-ürün etkileşimleri ve diğer yan bilgiler (örneğin, kullanıcı profili ve ürün özellikleri) dâhil olmak üzere kullanıcılardan veri toplama adımıdır. Kullanıcı ve ürün kümeleri arasında toplanan bu etkileşimler genellikle bir kullanıcı × ürün matrisi aracılığı ile saklanmaktadır. Toplanan veriler genellikle iki kategoride değerlendirilir: kullanıcıların ürünler için doğrudan sağladığı sayısal derecelendirme değerleri (yani, açık geribildirimler) ve kullanıcıların ürünler ile etkileşimlerinden çıkarılan satın alma, tıklama, görüntüleme gibi diğer bilgiler (yani, örtük geribildirimler).
- *Veri* → *Model*: Toplanan verilere dayalı olarak öneri modellerinin öğrenilmesi aşamasıdır. Bu adım, kullanıcı tercihlerini geçmiş etkileşimlerden türetmek ve bir kullanıcının bir hedef ürünü beğenme olasılığını tahmin etme aşamasıdır. Bu amaçla, geçtiğimiz yıllar boyunca birçok araştırma yapılmış ve çeşitli öneri algoritmaları geliştirilmiştir.
- *Model* → *Kullanıcı:* Kullanıcıların bilgi ihtiyacını karşılamak için öneri sonuçlarının onlara döndürülmesi adımıdır. Bu aşama, kullanıcıların gelecekteki davranışlarını ve kararlarını önemli ölçüde etkilemektedir.

BSEU Journal of Science
https://doi.org/10.35193/bseufbd.884634

e-ISSN: 2458-7575 (https://dergipark.org.tr/tr/pub/bseufbd)



Şekil 1. Farklı aşamalarında çeşitli yanlılıkların meydana geldiği öneri geri bildirim döngüsü

B. Önerilerde Yanlılık

Öneri sistemleri, çeşitli alanlardaki birçok uygulamada büyük etkiler oluştursa da, üstesinden gelinmesi zor olan ve öneri etkinliğini bozabilecek birçok yanlılık sorunuyla karşı karşıyadır. Son yıllarda, öneri algoritmalarının sebep olduğu bu tür yanlılıkların nedenleri, öneri kalitesi üzerindeki olumsuz etkileri ve bu etkilerin hafifletilmesi konuları giderek önem kazanmış ve bu alanda yapılan birçok araştırmanın ilgi odağı olmuştur [7]. Şekil 1'de gösterildiği gibi, geri bildirim döngüsünün farklı aşamalarında ortaya çıkabilecek bazı yanlılık türleri tanımlanmıştır. Giriş bölümünde belirtildiği üzere, bu çalışmanın temel motivasyonu, ürünlere sağlanan derecelendirme değerlerinin ortalaması göz önünde bulundurularak, çok beğenilen ürünlere yönelik bir yanlılığın olup olmadığının araştırılmasıdır. Böyle bir yanlılık esasında derecelendirme verilerinden kaynaklanacağı için, bu bölümde, verideki yanlılık türleri analiz edilerek literatürde öne çıkan çalışmalar özetlenmektedir.

Kullanıcı etkileşimlerinin verileri deneysel olmaktan çok gözlemsel olduğundan, bu verilerde çeşitli eşitsizlikler/yanlılıklar görülmesi olasıdır. Bu yanlılıklar, tipik olarak farklı veri alt gruplarından kaynaklanmaktadır ve öneri modellerinin sistematik bir ırkçılık yapmasına veya niteliksiz öneriler üretmesine neden olurlar [11]. Verideki yanlılıklar genel olarak iki kategoride değerlendirilmektedir. Bunlar, açık geribildirimde *Seçim*, *Uygunluk* ve *Popülerlik* yanlılığı, örtük geribildirimde ise *Maruz Kalma* ve *Pozisyon* yanlılığıdır [10].

Seçim Yanlılığı, kullanıcıların hangi ürünleri derecelendireceklerini seçmekte özgür olması nedeniyle, toplanan derecelendirmelerin tüm derecelendirmelerin temsili bir örneği olmadığı için gerçekleşir. [14]'de araştırmacılar, kullanıcı derecelendirmelerini rastgele seçilen bazı ürünler açısından ele alarak kullanıcı-tabanlı bir inceleme yapmışlar ve iki temel bulguya ulaşmışlardır: (i) kullanıcılar beğendikleri ürünleri seçme ve derecelendirme eğilimindedir ve (ii) kullanıcıların özellikle kötü veya iyi ürünleri derecelendirme olasılığı daha yüksektir. Bu sonuç, gözlemlenen derecelendirme verilerinin dağılımının tüm derecelendirmelerin dağılımından farklı olduğunu göstermiştir [15, 16]. Ayrıca, bu sorunun öneri kalitesi üzerindeki etkisini hafifletmek amacıyla, son yıllarda eğilim puanı [17], veri isnat etme [18, 19] ve yarı öğrenme [20] gibi yöntemler geliştirilmiştir.

Uygunluk Yanlılığı, bir gruptaki kullanıcıların kendi yargılarına aykırı olsa bile gruptaki diğer bireylere benzer şekilde derecelendirme eğilimi içerisinde olması olarak tanımlanır. Bu durum, toplanan derecelendirme değerlerinin her zaman kullanıcının gerçek tercihini yansıtmamasına neden olur. Örneğin, bir ürünle ilgili kamuya açık yorumların yüksek derecelendirmesinden etkilenen bir kullanıcının, muhalif olmaktan kaçınarak düşük derecelendirmesini değiştirme olasılığı yüksektir [21, 22]. Bu tür bir uygunluk olgusu yaygındır ve kullanıcı derecelendirmelerinde yanlılığa sebep olur. [23]'de araştırmacılar, kullanıcıların derecelendirme dağılımlarının kamuoyunun görüşüne maruz kalmadan önceki ve sonraki hallerinde önemli ölçüde farklılık olduğunu göstermiştir. Ayrıca, uygunluk yanlılığı, sosyal faktörlerden dolayı kullanıcıların arkadaşlarıyla benzer şekilde davranma eğiliminde olduğu durumlarda da sıklıkla ortaya çıkabilmektedir [24, 25]. Bu sorun genellikle kullanıcıların derecelendirme değerlerini, kullanıcı tercihi ve sosyal etkilerin sentetik sonuçları göz önünde bulundurularak ele alınması ile giderilmektedir [25, 26].

BSEU Journal of Science
https://doi.org/10.35193/bseufbd.884634

e-ISSN: 2458-7575 (https://dergipark.org.tr/tr/pub/bseufbd)

Popülerlik Yanlılığı, mevcut ürünlerin küçük bir kısmının hak etmeseler bile diğer birçok ürüne kıyasla gerektiğinden fazla derecelendirmeye sahip olmasıdır. Uzun-kuyruk fenomeni olarak da bilinen bu sorun, öneri sistemleri literatüründe en yaygın bilinen yanlılık türüdür [9, 27]. Esasında, toplanan derecelendirme verisinden kaynaklanan bu eşitsizlik durumu, modelin eğitimi sırasında popüler ürünlere popüler olmayanlara kıyasla ideal değerlerinden daha yüksek derecelendirmeler üretilmesine yol açar ve daha çok tüketilmesine neden olur [8]. Bu durumun öneri kalitesi üzerindeki etkileri genellikle kullanıcı verilerinin değiştirilmesi [28], model adımında kullanılan algoritmanın modifiye edilmesi [12] ve üretilen öneri listelerinin yeniden sıralanması [27] gibi yöntemler kullanılarak giderilmiştir.

Kullanıcıların satın alma, görüntüleme ve tıklama gibi doğan davranışlarını yansıtan geri bildirimler açıkça bildirilen derecelendirmelere ek olarak öneri üretme sürecinde yaygın olarak kullanılır. Sayısal derecelendirmeler sunan açık geribildirimlerden farklı olarak örtük geribildirimler kullanıcıların tercihleri bağlamında kısmi bir bilgi sunar; bu nedenle, *Maruz Kalma* ve *Konum* gibi yanlılıklar bu tür tek-sınıf verilerden kaynaklanmaktadır. *Maruz Kalma* Yanlılığı, kullanıcıların yalnızca belirli ürünlerin bir kısmına maruz kalması nedeniyle oluşur, dolayısıyla gözlemlenmeyen etkileşimler her zaman olumsuz bir tercih olduğu manasına gelmez. Bir kullanıcı ile bir ürün arasındaki gözlemlenmemiş bir etkileşim, iki olası nedene bağlanabilir: *(i)* ürün, kullanıcının ilgisiyle eşleşmemektedir ve *(ii)* kullanıcı, ürünün farkında değildir [29]. Bu nedenle, gözlemlenmemiş etkileşimlerin yorumlanmasında belirsizlik ortaya çıkar. Gerçek olumsuz etkileşimleri (örneğin, maruz kalan ancak ilgisiz) potansiyel olarak olumlu olanlardan (örneğin, maruz kalmayan) ayırt edememek ciddi yanlılıklara neden olmaktadır. Son olarak *Pozisyon Yanlılığı*, kullanıcıların gerçek alakalı düzeyine bakmaksızın öneri listesinin daha üst konumunda bulunan ürünlerle etkileşim kurma eğilimi olduğunda gerçekleşmektedir. Dolayısıyla, kullanıcıların etkileşim sağladığı ürünler onların gerçekten ilgisini çeken ürünler olmayabilir [30, 31].

Bu bölümde sunulan literatür özeti göz önünde bulundurulduğunda, veri toplama aşamasında ortaya çıkabilecek ve öneri kalitesini doğrudan etkileyebilecek yanlılık türleri son yıllarda yapılan birçok araştırmacının ilgisini çekmiştir. Açık geribildirim olarak toplanan derecelendirmelerin farklı birçok karakteristik özelliği olduğu düşünüldüğünde, ürünlerin beğenilme oranlarındaki dağılımında da popülerlikleri gibi eşit olmayacağı öngörülebilir. Dolayısıyla, ürünlere sağlanan derecelendirmelerin bu tür özellikleri analiz edilerek, derecelendirme verileri üzerinde eğitilen algoritmalar tarafından üretilen önerilerde bir yanlılığa neden olup olmadığı hususunda daha kapsamlı araştırmaların yapılması gerekmektedir.

III. MOTİVASYON ÖRNEĞİ VE YÖNTEM

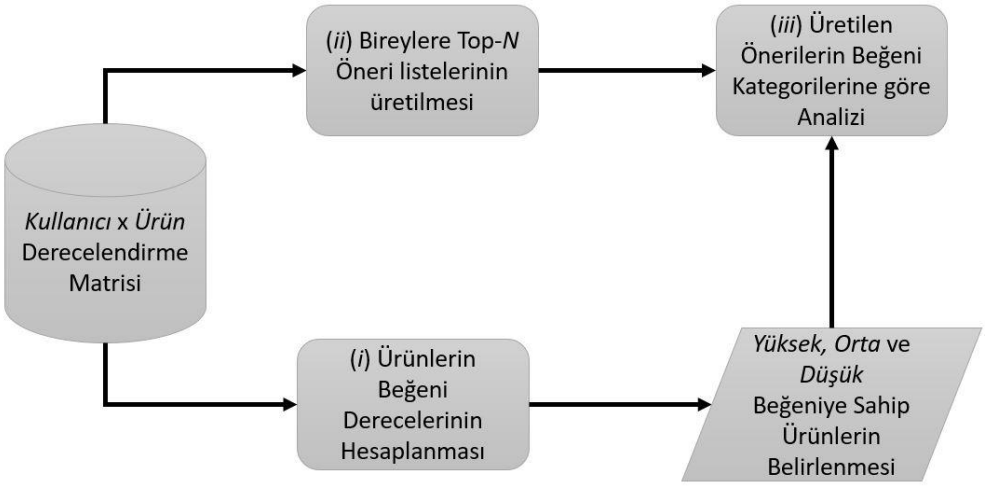
Öneri algoritmaları, daha önce Giriş Bölümü'nde belirtildiği gibi kullanıcıların ürünler için sağladığı derecelendirmeleri içeren derecelendirme matrisi üzerinde eğitilerek kullanıcılar için tavsiyeler üretirler. Derecelendirmelerin dağılımlarındaki bazı karakteristik özellikler bu algoritmaların performansını etkileyerek öneri çıktılarında adil olmayan sonuçların oluşmasına neden olabilirler. Bu durumun literatürde en yaygın örneği, bazı ürünlerin diğerlerinden daha popüler olması (kullanıcılar tarafından daha çok derecelendirilmesi) ve üretilen önerilerde bu ürünlerin lehine bir yanlılığın oluşmasıdır. Ürünlerin kullanıcılar tarafından farklı oranlarda beğenileceği gerçeğine dayanarak bu çalışmada ürünlere sağlanan derecelendirmeler farklı bir bakış açısıyla ele alınmış ve beğenilme oranları açısından bir eşitsizliğin olup olmadığı araştırılmıştır. Bununla birlikte, gerçekleştirilen deneysel çalışmalar ile beğeni derecesinin üretilen önerilerde bir yanlılığa sebep olup olmadığı kapsamlı bir şekilde analiz edilmiştir.

Bu amaçla kullanılan yöntemin işlem adımları Şekil 2'de gösterilmektedir. İlk olarak, (*i*) işlem adımında, ürünler, aşağıda ayrıntılı olarak açıklanan stratejiye dayanarak yüksek, orta ve düşük beğeniye sahip ürünler olmak üzere üç farklı kategoriye ayrılmıştır. Bu işlem adımında ayrıca literatürde popüler olarak kullanılan ve ayrıntılı bilgileri IV. Bölüm'de sunulan iki farklı veri kümesinde beğeni kategorilerine göre ürünlerin dağılımları analiz edilmiştir. Ardından, (*ii*) işlem adımında IV. Bölüm'ün başında ayrıntılı olarak açıklanan öneri algoritmaları yardımıyla kullanıcılar için Top-*N* sıralı ürün listeleri üretilmiştir. Son olarak, (*iii*) işlem adımında, üretilen önerilerdeki ürünlerin, beğeni kategorilerine göre kapsamlı bir analizi gerçekleştirilerek beğeni derecesinin neden olabileceği potansiyel yanlılık sorunları gözlemlenmiştir.

BŞEÜ Fen Bilimleri Dergisi
8(1), 279-291, 2021

BSEU Journal of Science
https://doi.org/10.35193/bseufbd.884634

e-ISSN: 2458-7575 (https://dergipark.org.tr/tr/pub/bseufbd)



Şekil 2. Kullanılan yöntemin blok şeması

Bir öneri sisteminde kullanıcı kümesinin $\{K = k_1, k_2, \dots, k_m\}$ ile, ürün kümesinin $\{I = i_1, i_2, \dots, i_n\}$ ile temsil edildiği durumda, tipik bir kullanıcı × ürün matrisi, kullanıcıların ürünler için yapmış oldukları derecelendirmeleri d_{ki} $\{k \in K, i \in I\}$ içermektedir. Bir i ürünün kullanıcılar tarafından beğenilme oranı b_i ise, Denklem 1'de gösterildiği gibi aldığı derecelendirme değerlerinin ortalaması alınarak hesaplanabilir.

$$b_i = \frac{\sum_{k \in K'} d_{ki}}{|K'|} \tag{1}$$

Burada, K', i ürünü için derecelendirme sağlayan kullanıcıların alt kümesini, $|K'|$ ise bu alt kümede bulunan kullanıcı sayısını ifade etmektedir.

Bununla birlikte, tipik bir [1,5] derecelendirme ölçeğine sahip öneri sisteminde, ürünler, onlar için hesaplanan b değerlerine göre aşağıdaki gibi sınıflandırılabilir [13]:

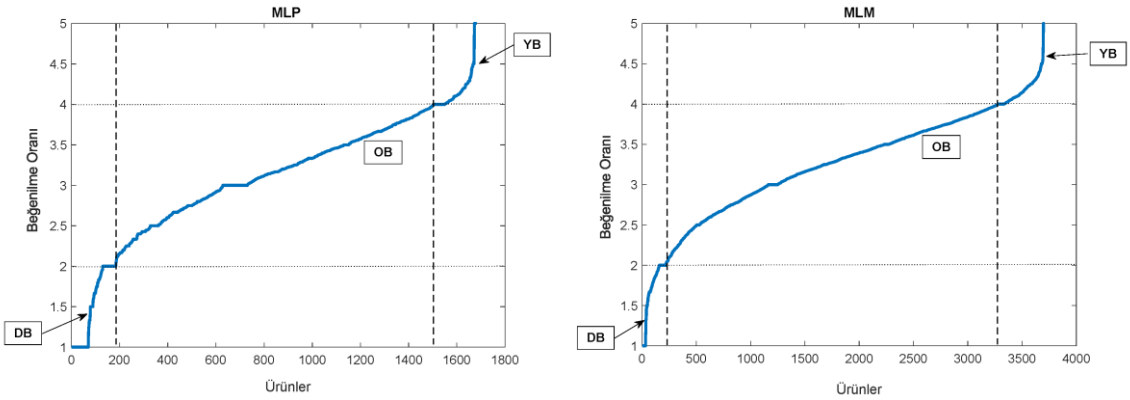
- Eğer $b_i \geq 4$ ise, i ürünü Yüksek Beğeniye (YB),
- Eğer $2 < b_i < 4$ ise, i ürünü Orta Beğeniye (OB),
- Eğer $b_i \leq 2$ ise, i ürünü Düşük Beğeniye (DB) sahiptir.

Bu çalışmada ilk olarak, öneri sistemi çalışmalarında sıklıkla kullanılan ve kullanıcıların çeşitli filmler için sağladığı [1,5] değerlendirme ölçeğindeki derecelendirmelerini içeren ve ayrıntıları özellikleri Tablo 1'de sunulan popüler MLP ve MLM veri kümelerinde her bir ürünün beğenilme oranı hesaplanarak hangi beğeni kategorisinde olduğu belirlenmiştir. Ardından, her iki veri kümesinde de ürünler beğeni oranlarına göre artan şekilde sıralanarak Şekil 3'de gösterildiği dağılımları analiz edilmiştir. Verilen şekilde, ürün dağılımları, beğeni kategorileri bağlamında üç farklı bölgeye ayrılmıştır. Sunulan şekilden açıkça görülebileceği gibi her iki kümesinde de, ürünlerin beğenilme oranlarına göre dağılımı benzerdir ve OB kategorisindeki ürünlerin sayısı hem YB hem de DB kategorisindeki ürünlere göre önemli ölçüde fazladır; OB kategorisindeki ürünlerin sayısının tüm ürünlerin sayısına oranı MLP veri kümesinde yaklaşık %79 iken MLM veri kümesinde %82'dir. Bu bulgu, ürünlerin beğeni oranlarının dağılımları bağlamında her iki veri kümesinde de önemli bir eşitsizlik olduğunu göstermektedir. Öte yandan, YB ve DB kategorisindeki ürünlerin sayısının tüm ürünlerin sayısına oranı her iki veri kümesi için de birbirine benzer olduğu ve OB ürünlerine kıyasla oldukça düşük olduğu söylenebilir. Spesifik olarak DB ve YB ürünlerin MLP veri kümesindeki oranı sırasıyla %11 ve %10 iken, MLM veri kümesindeki oranı %6 ve %12'dir.

BŞEÜ Fen Bilimleri Dergisi
8(1), 279-291, 2021

BSEU Journal of Science
https://doi.org/10.35193/bseufbd.884634

e-ISSN: 2458-7575 (https://dergipark.org.tr/tr/pub/bseufbd)



Şekil 3. MLP ve MLM veri kümelerindeki ürünlerin beğenilme oranları (X ekseninde ürünler beğenilme oranlarına göre artan bir şekilde sıralanarak temsil edilmektedir.)

IV. DENEYSEL ÇALIŞMALAR

Bu bölümde, öneri sistemi literatüründe öne çıkan çeşitli öneri algoritmaları ile bireyler için üretilen sıralı öneri listelerinde önceki bölümde ayrıntılı olarak açıklanan farklı beğeni kategorileri bağlamında bir eşitsizlik/yanlılık oluşup oluşmadığı ve bu listelerin doğrulukları analiz edilmektedir. Bu kapsamda ilk olarak deneylerde kullanılan dokuz popüler işbirlikçi filtreleme algoritması, veri kümeleri, değerlendirme ölçütleri ve deney metodolojisi açıklanmıştır. Ardından, deney sonuçları sunularak elde edilen bulgular tartışılmıştır.

A. Kullanılan İşbirlikçi Filtreleme Algoritmaları

Deneysel çalışmalarda, üç farklı kategoride öne çıkan, etkinliği ve verimliliği ile bilinen dokuz farklı işbirlikçi filtreleme (İF) algoritması kullanılmıştır. Bu algoritmalar ve parametreleri aşağıda açıklanmaktadır.

Komşuluk-tabanlı algoritmalar [32]:

- *User-kNN:* Popüler hafıza tabanlı İF algoritmalarından birisi olan bu algoritma aktif kullanıcı ile diğer kullanıcılar arasındaki benzerlikleri dikkate alarak en benzer *k* kullanıcıdan oluşan komşular oluşturur. Belirlenen komşuların hedef ürün için sağladığı derecelendirmeler üzerinden aktif kullanıcı için bireysel tahmin değerlerinin üretilmesini sağlamaktadır.
- *Item-kNN:* Bu algoritmada komşulukla, hedef ürüne en benzer ürünler göz önünde bulundurularak oluşturulur. Bireysel tahmin üretme süreci bu komşuluklar üzerinden hesaplanmaktadır.
- *User-kNNBase:* Bazı kullanıcılar sistematik olarak diğerlerinden daha yüksek derecelendirme sağlama eğiliminde olurlar. Bu algoritmanın *User-kNN* algoritmasından farkı, kullanıcıların bu tür eğilimlerinin üretilen bireysel tahmin değerleri içerisinde kapsüllenmesidir.
- *Item-kNNBase*: Benzer şekilde bazı ürünler genel olarak diğer ürünlerden daha çok beğenilir ve daha yüksek derecelendirmeler alırlar. Bu yöntemde, ürünlerin derecelendirmelerindeki bu tür özellikler üretilen bireysel tahmin değerlerinde kapsüllenir. Not olarak, açıklanan tüm *kNN* temelli algoritmalarda kullanıcılar/ürünler benzerlikler, en başarılı benzerlik metriklerinden birisi olan Pearson Korelasyonu Katsayısı ile hesaplanmıştır. Ayrıca, bu algoritmalar uygulanırken maksimum komşuluk değeri 40 olarak seçilmiştir; çünkü beş farklı komşuluk değeri (yani, 10, 20, 40, 60 ve 80) ile gerçekleştirilen deneysel çalışmalarda, en yüksek doğruluğa sahip sıralı öneri listeleri bu ayarlama ile elde edilmiştir. Bununla birlikte, bu ayarlamanın, *kNN* temelli algoritmalar açısından doğruluğu yüksek tahminlerin üretilmesinde ideal bir konfigürasyon olduğu önceki çalışmalarda gösterilmiştir [33].
- *SlopeOne*: *Item-kNN* ailesinden gelen bu algoritma, ürünlerin derecelendirme vektörleri arasındaki ilişkileri daha basit regresyon biçiminde tanımlayarak aşırı yüklenme sorununun olumsuz etkilerini hafifletmeyi sağlayan bir algoritmadır [34].

Matris Çarpanlara ayırma tabanlı algoritmalar [35]:

- *SVD*: Kullanıcılar ve ürünler arasındaki gizli ilişkileri sağlanan açık geribildirimleri kullanarak tespit ederek genellikle büyük ve oldukça boşluklu yapıda olan derecelendirme matrisini daha küçük boyuta indirgeyerek ölçeklenebilirlik sorununun üstesinden gelmeye yardımcı olan bir model-tabanlı algoritmadır.
- *SVD++*: *SVD* algoritmasının gelişmiş bir versiyonu olan bu yöntemde kullanıcılar ve ürünler arasındaki etkileşimler yalnızca mevcut açık geribildirimlerden değil aynı zamanda beğeniler, satın almalar, tıklanma sayıları gibi çeşitli örtük geribildirimler kullanılarak belirlenir.
- *NMF*: Negatif olmayan matris ayrıştırması tabanlı bu algoritmada indirgenmiş matrisin oluşturma sürecinde, açık geribildirimlerdeki negatif derecelendirmeler dikkate alınmaz. Diğer bir deyişle yalnızca pozitif derecelendirmeler dikkate alınarak kullanıcılar/ürünler arasındaki etkileşimler belirlenmektedir.

Kümeleme tabanlı algoritmalar:

- *Co-clustering:* Bu algoritma, derecelendirme matrisi üzerinde kümeleme yöntemleri kullanarak çeşitli özet istatistikler çıkarımı yapar. Elde edilen bu istatistikler daha sonra bireysel tahmin üretme sürecinde kullanılır [36].

B. Veri Kümeleri

Önceki bölümde bilgileri verilen algoritmaları, ürettikleri öneri listelerinde ürünlerin beğenilerine göre herhangi bir yanlılığa neden olup olmadığını analiz etmek için öneri sistemleri literatüründe yaygın olarak kullanılan iki farklı veri kümesi kullanılmıştır; MovieLens-100K (MLP) ve MovieLens-1M (MLM) [4]. Her iki veri kümesinde de kullanıcıların çeşitli filmler için sağladığı derecelendirmeler, ayrık tipte ve [1,5] derecelendirme yelpazesindedir. Bu veri kümelerinin ayrıntılı özellikleri Tablo 1'de sunulmaktadır. Bununla birlikte, bu veri kümelerinin, öneri sistemleri literatüründeki potansiyel yanlılık sorunları üzerine yapılan çalışmalarda sıklıkla kullanılması nedeniyle bu çalışmanın amacına uygun olduğu söylenebilir [8, 12, 37].

Tablo 1. Veri kümelerinin özellikleri

Veri Kümesi	Kullanıcı Sayısı	Ürün Sayısı	Derecelendirme Sayısı	Yoğunluk Oranı
MLP	943	1.682	100.000	%6,3
MLM	6.040	3.952	1.000.000	%4,25

C. Değerlendirme Ölçütleri ve Deney Metodolojisi

Bireyler için üretilen top-*N* öneri listelerinin doğruluğunu analiz etmek için literatürde yaygın biçimde kullanılan ve formülleri sırasıyla Denklem 2, 3 ve 4'te sunulan Kesinlik (Precision), Duyarlılık (Recall) ve F1-Skoru (F1-Score) değerlendirme ölçütleri kullanılmıştır [4]. Spesifik olarak, Precision, önerilen uygun ürünlerin sayısının önerilen tüm ürünlerin sayısına oranı olarak hesaplanmaktadır. Diğer yandan, Recall, önerilen uygun ürünlerin sayısının kullanıcının profilindeki bütün uygun ürünlerin sayısına oranı olarak hesaplanır. Son olarak F1-Score, hesaplanan Precision ve Recall değerlerinin harmonik ortalaması hesaplanarak belirlenir. Not olarak, bu ölçümler hesaplanırken bir ürünün uygun olup olmadığını belirlemek için eşik değeri 3.5 olarak seçilmiştir; çünkü [1, 5] değerlendirme ölçeğinde pozitif derecelendirmeler 4 ve 5'e karşılık gelmektedir [13].

$$Precision@N = \frac{\sum_{i \in N} \mathbb{1}(i \in p_k)}{|N|} \tag{2}$$

$$Recall@N = \frac{\sum_{i \in N} \mathbb{1}(i \in p_k)}{|p_k|} \tag{3}$$

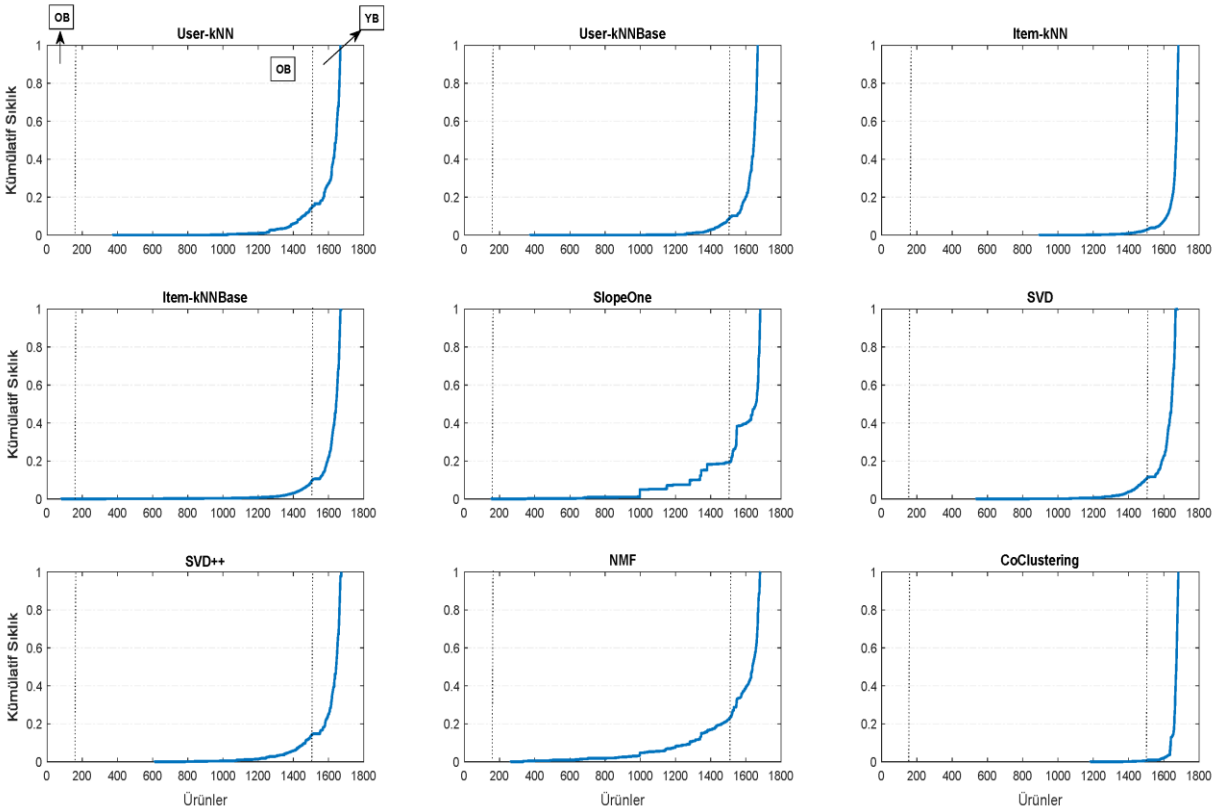
$$F1@N = 2 \times \frac{Precision@N \times Recall@N}{Precision@N + Recall@N} \quad (4)$$

Burada p_k, k kullanıcısının profilindeki uygun ürünlerin kümesini belirtir ve $\mathbb{1}(.)$, bağımsız değişkeni doğruysa 1, aksi halde 0 döndüren gösterge fonksiyonunu temsil etmektedir.

Gerçekleştirilen deneylerde bireysel öneri listeleri tek-çıkışlı çapraz doğrulama (leave-one-out cross validation) deney metodolojisi kullanılarak üretilmiştir. Buna göre, bir aktif kullanıcı test kullanıcısı olarak seçilmiş ve ürünler için sağladığı derecelendirmeler boş kabul edilmiştir. Geriye kalan tüm kullanıcılar eğitim kümesi olarak kabul edilip, bu kümedeki veriler üzerinde önceki bölümde açıklanan algoritmalardan biri kullanılarak aktif kullanıcıya tüm ürünler için bireysel tahmin değerleri üretilmiştir. Bu işlem veri kümesindeki tüm kullanıcılar için gerçekleştirilmiştir. Ardından, kullanıcılara ürünler için üretilen bireysel tahmin değerleri azalan sırada sıralanarak, her bir kullanıcı için en yüksek tahmin değerine sahip N ürünü içeren top-*N* öneri listeleri üretilmiştir. Son olarak, tüm kullanıcılar için üretilen tüm top-*N* listelerinde DB, OB ve YB kategorisindeki ürünlerin görülme sıklıkları analiz edilmiştir. Ayrıca, her kullanıcı için üretilen top-*N* ürün listesinin doğruluğu Precision, Recall ve F1-Score ile hesaplanarak ortalaması alınmıştır; böylece, ilgili algoritmanın genel doğruluk performansı hesaplanmıştır. Not olarak, gerçekleştirilen deneylerde top-20 öneri listeleri dikkate alınmıştır.

D. Deneysel Sonuçlar ve Tartışma

Bu bölümde ilk olarak öneri algoritmalarının ürettikleri top-*N* listelerinde bulunan ürünlerin Bölüm 3'de açıklanan DB, OB ve YB kategorileri bağlamında dağılımını analiz etmek için gerçekleştirilen deneylerin sonuçları sunulmaktadır. Şekil 4 ve 5 sırasıyla MLP ve MLM veri kümesinde üretilen top-20 listelerindeki ürünlerin beğeni kategorilerine göre görülme durumlarını, kümülatif sıklık grafikleri aracılığı ile göstermektedir.

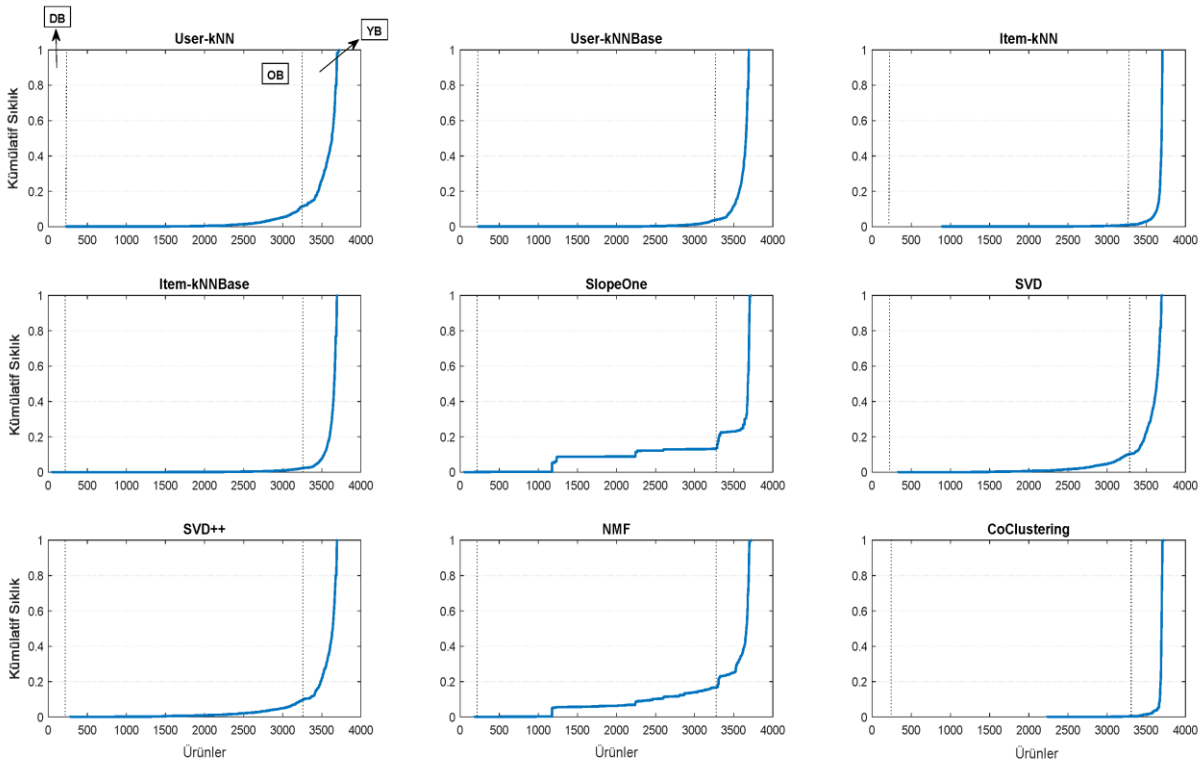


Şekil 4. MLP veri kümesinde algoritmaların ürettikleri top-20 öneri listelerinde DB, OB ve YB kategorilerindeki ürünlerin görülme sıklıkları (X ekseninde ürünler beğenilme oranlarına göre artan bir şekilde sıralanarak temsil edilmektedir.)

Daha önce Bölüm 3'de hem MLP hem de MLM veri kümesinde OB kategorisindeki ürünlerin sayısının YB ve DB kategorisindeki ürünlerin sayısına kıyasla oldukça fazla olduğu gösterilmişti. Buna rağmen, Şekil 4 ve 5'ten açıkça görüldüğü üzere her iki veri kümesi için de tüm algoritmalarla üretilen bireysel öneri listelerinde YB

kategorisindeki ürünlerin görülme sıklıklarının OB ve DB kategorisindeki ürünlere kıyasla oldukça fazla olduğu söylenebilir. Hatta MLP veri kümesinde seçilen algoritmanın *User-kNNBase*, *Item-kNN*, *Item-kNNBase*, *SVD* veya *Co-Clustering* olduğu durumda, üretilen öneri listelerindeki ürünlerin yaklaşık %90'ı YB kategorisindedir. Bu gözlem, MLM veri kümesi için gerçekleştirilen deneylerde de benzerdir. Diğer bir deyişle, üretilen öneri listelerinde OB ve DB kategorisindeki ürünler neredeyse hiç bulunmamaktadır. Bu bulgu, bu algoritmaların YB kategorisindeki ürünlerin lehine ciddi oranda bir ayrımcılık yaptığı, OB ve DB kategorisindeki ürünlere önerdikleri ürün listelerinde neredeyse hiç şans tanımadığını göstermektedir.

Seçilen algoritmanın *User-kNN*, *SlopeOne*, *SVD++* veya *NMF* olduğu durumda ise, YB kategorisindeki ürünlerin önerilen ürün listelerindeki oranı her iki veri kümesinde de yaklaşık %80'dir. Bu sonuç, matris çarpanlara ayırma tabanlı algoritmaların az da olsa diğer kategorideki algoritmalara kıyasla ürün çeşitliliği açısından daha başarılı öneriler üretebildiğini göstermiştir. Ancak yine de YB kategorisindeki ürünler bu algoritmalarla üretilen önerileri listelerinde OB ve DB kategorisindeki ürünlere kıyasla önemli ölçüde fazladır. Gerçekleştirilen deneylerin sonuçları, hem MLP hem de MLM veri kümesinde tüm algoritmaların YB kategorisindeki ürünlere yönelik önemli bir yanlılığının olduğunu ve ürettikleri öneri listelerinde bu ürünlerin diğerlerine kıyasla oldukça daha sık gözükmesi nedeniyle ürün kataloğu kapsamı açısından niteliksiz önerilere yol açtığını göstermiştir.



Şekil 5. MLM veri kümesinde algoritmaların ürettikleri top-20 öneri listelerinde DB, OB ve YB kategorilerindeki ürünlerin görülme sıklıkları (X ekseninde ürünler beğenilme oranlarına göre artan bir şekilde sıralanarak temsil edilmektedir.)

Bu algoritmaların ürettikleri top-20 öneri listelerinin kalitesi ayrıca Tablo 2'de sonuçlarının gösterildiği gibi üç farklı doğruluk ölçütü ile analiz edilmiştir. Gerçekleştirilen deneylerin sonuçlarına göre, hemen hemen tüm algoritmaların Precision değerleri MLM veri kümesinde daha yüksek iken, Recall değerleri MLP veri kümesinde daha yüksektir. Bu sonucun temel nedeni, MLM veri kümesinde bir kullanıcının profilindeki ortalama derecelendirme sayısının MLP veri kümesine kıyasla daha fazla olmasıdır. Bu durum, Denklem 3'te sunulan formül göz önünde bulundurulduğunda payda değerinin artmasına ve dolayısıyla hesaplanan Recall değerinin düşmesine neden olur. Öte yandan, kullanıcı profili başına düşen derecelendirme sayısının artması, üretilen öneri listelerinde uygun ürünlerin yakalanma ihtimalini arttırarak Precision değerlerinin artmasına yol açmıştır.

BSEU Journal of Science
https://doi.org/10.35193/bseufbd.884634

e-ISSN: 2458-7575 (https://dergipark.org.tr/tr/pub/bseufbd)

Elde edilen deney sonuçlarına göre, en kötü doğruluk sonuçlarının *SlopeOne* algoritmasının seçildiği durumda gözlemlendiği söylenebilir. Ancak, *SlopeOne* önceki deneylerde gösterildiği üzere ürün kataloğu açısından en başarılı algoritmalardan birisidir. Bununla birlikte, matris çarpanlara ayırma tabanlı algoritmalar, özellikle *SVD* ve *SVD++*, diğer algoritmalara kıyasla doğruluk açısından biraz daha iyi performans gösterdiği gözlemlenmektedir. Önceki deneylerin sonuçları da göz önünde bulundurulduğunda, bu iki algoritmaların ürün kataloğu açısından diğer algoritmalara kıyasla hem daha nitelikli hem de doğruluk açısından genellikle makul bireysel öneri listelerinin üretilmesini sağladığı söylenebilir.

Tablo 2. MLP ve MLM veri kümesi için doğruluk sonuçları

	MLP			**MLM**		
	Precision	**Recall**	**F1**	**Precision**	**Recall**	**F1**
UserkNN	0.363	0.159	0.193	0.358	0.090	0.195
UserkNN-Base	0.360	0.154	0.186	0.322	0.087	0.189
ItemkNN	0.127	0.038	0.053	0.101	0.023	0.033
ItemkNN-Base	0.326	0.113	0.147	0.272	0.071	0.097
SlopeOne	0.020	0.006	0.008	0.033	0.007	0.011
SVD	0.351	0.162	0.194	0.303	0.095	0.193
SVD++	0.386	0.182	0.211	0.348	0.106	0.203
NMF	0.070	0.026	0.032	0.059	0.015	0.021
Co-Clustering	0.066	0.024	0.031	0.083	0.023	0.031

V. SONUÇLAR

Öneri sistemleri, bireylerin herhangi bir kişisel çabası olmadan onların henüz deneyimlemedikleri ancak ilgilerini çeken ürünleri/hizmetleri tavsiye eden sistemler olarak tanımlanır. İnternet'in yaygın kullanımına paralel olarak birçok çevrimiçi platform tarafından etkin bir şekilde kullanılmaktadır. Her ne kadar bu sistemlerin başarısı genellikle ürettikleri önerilerin doğruluğu ile analiz edilse de son yıllarda üretilen öneri listelerinin ürün kataloğunu hangi oranda kapsadığı ve içerdiği ürünlerin çeşitliliği servis sağlayıcıları açısından giderek önem kazanmıştır.

Son yıllarda araştırmacılar ürün profillerini analiz ederek üretilen öneri listelerinde bazı özellikteki ürünlere karşı kullanılan algoritmaların bazı pozitif ayrımcılıklar yaptığını göstermişlerdir. Örneğin, öneri algoritmalarının popüler ürünlere yönelik bir yanlılığının olduğu ve ürettikleri öneri listelerinde popüler ürünlerin diğerlerine kıyasla çok daha fazla gözüktüğü bilinen bir olgudur. Bu durum, diğer ürünlerin hak ettikleri ilgili görseler bile yeteri kadar tüketilmemesine neden olur. Bu çalışmada, ürün profilleri farklı bir bakış açısıyla ele alınarak, beğenilme oranları açısından analiz edilmektedir ve üretilen öneri listelerinde ürün kataloğu açısından bir adaletsizliğe yol açıp açmadığı keşfedilmektedir. Bu amaçla ilk olarak, üç farklı beğeni kategorisi tanımlanarak iki popüler veri kümesinde ürünlerin bu kategorilere göre dağılımları analiz edilmiştir. Ardından, öneri sistemi literatüründe öne çıkan dokuz farklı işbirlikçi filtreleme algoritmasının ürettikleri öneri listelerinde bulunan ürünler, tanımlanan beğeni kategorilerine göre ayrıntılı bir şekilde irdelenmiştir. Ayrıca, bu algoritmaların performansları üç farklı doğruluk ölçütü ile değerlendirilmiştir. Gerçekleştirilen deneylerin sonuçları, hemen hemen bütün algoritmaların yüksek beğeniye (yani, yüksek derecelendirme ortalamasına) sahip ürünlere karşı bir yanlılığının olduğunu ve ürettikleri öneri listelerinde bu ürünlerin diğerlerine kıyasla daha sık görüldüğünü göstermektedir. Bu bulgu, algoritmaların doğruluk açısından makul öneriler üretebilmesine rağmen ürün kataloğu açısından niteliksiz önerilere neden olabildiğini göstermektedir. Bunula birlikte, SVD ve SVD++ gibi matris çarpanlarına ayırma tabanlı algoritmaların hem makul doğrulukta hem de yüksek beğeniye olan yanlılık bağlamında diğer algoritmalara kıyasla daha nitelikli öneriler ürettiği gözlemlenmiştir.

Gelecekte, algoritmaların yüksek beğeniye sahip ürünlere olan yanlılığı gidermek ve üretilen önerilerdeki etkisini hafifletmek amacıyla ürünleri yeniden sıralayan yeni yöntemlerin geliştirilmesi hedeflenmektedir. Ayrıca, ürünlere sağlanan derecelendirmeler farklı özellikler (sapması, entropisi vb.) açısından daha ayrıntılı bir şekilde irdelenerek, üretilen önerilerde bir adaletsizliğe yol açıp açmadığının araştırılması planlanmaktadır.

KAYNAKLAR

[1] Lu, J., Wu, D., Mao, M., Wang, W. & Zhang, G. (2015). Recommender system application developments: a survey. *Decision Support Systems*, 74, 12-32.
[2] Afoudi, Y., Lazaar, M. & Al Achhab, M. (2018). Collaborative filtering recommender system. International Conference on Advanced Intelligent Systems for Sustainable Development, 332-345.
[3] Batmaz, Z., Yurekli, A., Bilge, A. & Kaleli, C. (2019). A review on deep learning for recommender systems: challenges and remedies. *Artificial Intelligence Review*, 52(1), 1-37.
[4] Su, X. & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in artificial intelligence*, 2009, 1-19.
[5] Kaleli, C. (2014). An entropy-based neighbor selection approach for collaborative filtering. *Knowledge-Based Systems*, 56, 273-280.
[6] Yalçın, E., İsmailoğlu, F. & Bilge, A. (2021). An entropy empowered hybridized aggregation technique for group recommender systems. *Expert Systems with Applications*, 166, 114111.
[7] Abdollahpouri, H., Mansoury, M., Burke, R. & Mobasher, B. (2020). The Connection Between Popularity Bias, Calibration, and Fairness in Recommendation. Fourteenth ACM Conference on Recommender Systems, 726-731.
[8] Abdollahpouri, H., Mansoury, M., Burke, R. & Mobasher, B. (2019). The unfairness of popularity bias in recommendation. *arXiv preprint arXiv*, 1907, 13286.
[9] Boratto, L., Fenu, G. & Marras, M. (2019). The effect of algorithmic bias on recommender systems for massive open online courses. European Conference on Information Retrieval, 457-472.
[10] Chen, J., Dong, H., Wang, X., Feng, F., Wang, M. & He, X. (2020). Bias and Debias in Recommender System: A Survey and Future Directions. *arXiv preprint arXiv*, 2010, 03240.
[11] Ramos, G., Boratto, L. & Caleiro, C. (2020). On the negative impact of social influence in recommender systems: A study of bribery in collaborative hybrid algorithms. *Information Processing & Management*, 57(2), 102058.
[12] Boratto, L., Fenu, G. & Marras, M. (2021). Connecting user and item perspectives in popularity debiasing for collaborative recommendation. *Information Processing & Management*, 58(1), 102387.
[13] Bobadilla, J., Ortega, F., Hernando, A. & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-based systems*, 46, 109-132.
[14] Marlin, B., Zemel, R. S., Roweis, S. & Slaney, M. (2012). Collaborative filtering and the missing at random assumption. *arXiv preprint arXiv*, 1206. 5267.
[15] Steck, H. (2013). Evaluation of recommendations: rating-prediction and ranking. Proceedings of the 7th ACM conference on Recommender systems, 213-220.
[16] Hernández-Lobato, J. M., Houlsby, N. & Ghahramani, Z. (2014). Probabilistic matrix factorization with non-random missing data. International Conference on Machine Learning, 1512-1520.
[17] Ge, Y., Zhao, S., Zhou, H., Pei, C., Sun, F., Ou, W. & Zhang, Y. (2020). Understanding echo chambers in e-commerce recommender systems. Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2261-2270.
[18] Steck, H. (2010). Training and testing of recommender systems on data missing not at random. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, 713-722.
[19] Chen, J., Wang, C., Ester, M., Shi, Q., Feng, Y. & Chen, C. (2018). Social recommendation with missing not at random data. 2018 IEEE International Conference on Data Mining (ICDM), 29-38.
[20] Saito, Y. (2020). Asymmetric Tri-training for Debiasing Missing-Not-At-Random Explicit Feedback. Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 309-318.
[21] Wang, T. & Wang, D. (2014). Why Amazon's ratings might mislead you: The story of herding effects. *Big data*, 2(4), 196-204.
[22] Liu, Y., Cao, X. & Yu, Y. (2016). Are You Influenced by Others When Rating? Improve Rating Prediction by Conformity Modeling. Proceedings of the 10th ACM Conference on Recommender Systems, 269-272.
[23] Krishnan, S., Patel, J., Franklin, M. J. & Goldberg, K. (2014). A methodology for learning, analyzing, and mitigating social influence bias in recommender systems. Proceedings of the 8th ACM Conference on Recommender systems, 137-144.

[24] Chaney, A. J., Blei, D. M. & Eliassi-Rad, T. (2015). A probabilistic model for using social networks in personalized item recommendation. Proceedings of the 9th ACM Conference on Recommender Systems, 43-50.
[25] Wang, X., Hoi, S. C., Ester, M., Bu, J. & Chen, C. (2017). Learning personalized preference of strong and weak ties for social recommendation. Proceedings of the 26th International Conference on World Wide Web, 1601-1610.
[26] Tang, J., Gao, H. & Liu, H. (2012). mTrust: Discerning multi-faceted trust in a connected world. Proceedings of the fifth ACM international conference on Web search and data mining, 93-102.
[27] Abdollahpouri, H., Burke, R. & Mobasher, B. (2017). Controlling popularity bias in learning-to-rank recommendation. Proceedings of the Eleventh ACM Conference on Recommender Systems, 42-46.
[28] Jannach, D., Kamehkhosh, I. & Bonnin, G. (2016). Biases in automated music playlist generation: A comparison of next-track recommending techniques. Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization, 281-285.
[29] Liu, D., Cheng, P., Dong, Z., He, X., Pan, W. & Ming, Z. (2020). A general knowledge distillation framework for counterfactual recommendation via uniform data. Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 831-840.
[30] Joachims, T., Granka, L., Pan, B., Hembrooke, H. & Gay, G. (2017). Accurately interpreting clickthrough data as implicit feedback. *ACM SIGIR Forum*, 4-11.
[31] Collins, A., Tkaczyk, D., Aizawa, A. & Beel, J. (2018). A study of position bias in digital library recommender systems. *arXiv preprint arXiv*, 1802.06565.
[32] Koren, Y. (2010). Factor in the neighbors: Scalable and accurate collaborative filtering. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 4(1), 1-24.
[33] Herlocker, J., Konstan, J.A. & Riedl, J. (2002). An Empirical Analysis of Design Choices in Neighborhood-Based Collaborative Filtering Algorithms. *Information Retrieval*, 5, 287–310.
[34] Lemire, D. & Maclachlan, A. (2005). Slope one predictors for online rating-based collaborative filtering. Proceedings of the 2005 SIAM International Conference on Data Mining, 471-475.
[35] Bokde, D., Girase, S. & Mukhopadhyay, D. (2015). Matrix factorization model in collaborative filtering algorithms: A survey. *Procedia Computer Science*, 49, 136-146.
[36] George, T. & Merugu, S. (2005). A scalable collaborative filtering framework based on co-clustering. Fifth IEEE International Conference on Data Mining (ICDM'05), 4.
[37] Khenissi, S. & Nasraoui, O. (2020). Modeling and counteracting exposure bias in recommender systems. *arXiv preprint arXiv*, 2001.04832.