

Öneri Sistemlerinde Kişilik Özelliklerine Göre Performans ve Popülerlik Yanlılığı Analizi

Tuğba Türkoğlu Kaya^{1*}

¹Bilgisayar Mühendisliği Bölümü, Ardahan Üniversitesi, Ardahan, Türkiye (tugbaturkoglu@ardahan.edu.tr)
*corresponding author

Özet – Öneri sistemleri, kullanıcıların tercihlerine dayalı olarak kişisel ve memnuniyetin ön planda olduğu öneriler sunmayı amaçlamaktadır. Ancak her ne kadar amaç doğrultusunda hareket edilmeye çalışılsa da, önerilerde popüler yanlılığı problemi ile karşılaşmaktadır. Bu çalışma, öneri sistemlerinin kişilik tahmini alanındaki kullanımını ele almakta ve bireylerin kişilik özellikleri (Açıklık, Uyumluluk, Duygusalılık, Sorumluluk ve Dışadönüklük) temelinde popülerlik yanlılığını incelemeyi amaçlamaktadır. Çalışmada ayrıca, bu kişilik boyutlarına göre yapılan tahminlerin doğruluğunu değerlendirmek amacıyla geliştirilen bir modelin etkinliği analiz edilmiştir. Değerlendirmelerde sıralama tabanlı (nDCG, GAP) ve kapsama tabanlı (APLT, LTC) metrikler kullanılmıştır. Elde edilen sonuçlar, Açıklık boyutunda modelin sıralama başarısının en yüksek olduğunu gösterirken, GAP değerlerinin tüm boyutlar için negatif olması sıralama doğruluğunda önemli eksikliklere işaret etmektedir.

Anahtar Kelimeler – Öneri sistemleri, Popüler yanlılığı, Kullanıcı eğilimi

Analysis of Performance and Popularity Bias in Recommender Systems Based on Personality Traits

Abstract – Recommender systems aim to provide personal and satisfaction-oriented recommendations based on users' preferences. However, although efforts are made to act in line with the goal, the problem of popularity bias is encountered in recommendations. This study addresses the use of recommendation systems in the field of personality prediction and aims to examine the popularity bias based on the personality traits of individuals (Openness, Agreeableness, Emotionality, Conscientiousness and Extraversion). In addition, the effectiveness of a model developed to evaluate the accuracy of predictions made according to these personality dimensions was analyzed in the study. Ranking-based (nDCG, GAP) and coverage-based (APLT, LTC) metrics were used in the evaluations. The obtained results show that the model has the highest ranking success in the Openness dimension, while the negative GAP values for all dimensions indicate significant deficiencies in ranking accuracy.

Keywords – Recommender systems, Popularity bias, User tendency

I. GİRİŞ

İnternet ve teknolojinin giderek yaygınlaşması ile birlikte öneri sistemleri oldukça popüler hale gelen bir yapı olmuştur. Bu sistemler filmler, müzik, çevrim içi vb. gibi platformlarda yaygın olarak kullanılmaktadır. Bu sistemlerin amacı, birçok alternatif arasından kullanıcıların ilgilerini çekebilecek anlamlı tavsiyeleri üretmektir. Bu sayede, bu sistemler hem hizmet alan hem de hizmet sağlayanlar için oldukça faydalı olabilmektedir. Amazon' daki kitap önerileri veya Netflix' teki filmler, bu öneri sistemlerinin çevrimiçi hizmet sağlayıcıların nasıl bir işleyişe sahip olduğunun gerçek dünyadaki örnekleridir. Dolayısıyla bu sistemlerin varlığı, kullanıcılara zaman kaybı olmadan doğru hizmet alabilmelerine imkân tanımaktadır.

Öneri sistemleri, işbirlikçi filtreleme, içerik-tabanlı ve hibrit yöntemler olarak alt dallara ayrılmaktadır [1,2]. Ancak öneri amacıyla en çok kullanılan algoritma türlerinden biri işbirlikçi filtreleme (İF) yöntemleridir. Bu sistemler, günümüzde en yaygın ve etkili öneri yöntemlerinden biri olarak kabul edilmektedir. İF yaklaşımları, kullanıcının seçimlerini,

sistemdeki diğer kullanıcıların tercihlerine göre yönlendirmeyi amaçlar. Bu yöntemin temel varsayımı, geçmişte belirli öğeler hakkında benzer ya da zıt görüşler belirten kullanıcıların, gelecekte de benzer tercihlere sahip olacağıdır. Bu doğrultuda, İF algoritmaları önce hedef kullanıcıya düşünsel olarak en yakın kullanıcıları belirler, ardından bu kullanıcıların tercihlerini analiz ederek öneri oluşturur. Her ne kadar bu sistemler oldukça başarılı tavsiyeler üretip, kullanıcı memnuniyetini arttırsa da, bu sistemlerin önemli sınırlılıklarından biri popüler yanlılığı problemidir [3,4]. Bu durum, sistemlerin büyük ölçüde popüler öğelere odaklanarak daha az bilinen içeriklere yeterince yer vermemesiyle kendini gösterir. Oysa ki, popüler olmayan, yani "uzun kuyruk" olarak adlandırılan öğeler, kullanıcıların tercihlerine dair daha derin ve çeşitli içgörüler sunabilir. Özellikle, kullanıcı profillerini keşfetmeye yönelik etkin öğrenme (active learning) stratejileri uygulayan sistemler için bu tür öğelerin önerilmesi kritik öneme sahiptir. Çünkü genellikle kullanıcılar bu içerikleri henüz puanlamamıştır ve tercih yelpazeleri bu alanlarda daha çeşitlidir [5, 6]. Bununla birlikte, uzun kuyruk öğelerinin önerilmesi yalnızca bireysel kullanıcı deneyimi açısından

değil, aynı zamanda toplumsal fayda bakımından da önem taşır. Popülerlik odaklı sistemlerin hakim olduğu pazarlarda, daha az bilinen ürünler ya da sanatçılar görünürlük kazanmakta zorlanmaktadır. Bu da yenilikçiliğin ve çeşitliliğin önünü tıkayan, büyük markaların ve bilinen aktörlerin baskın olduğu daha tekdüze bir pazar yapısına yol açmaktadır [7]. Dolayısıyla bu çalışmada, bu problemi kullanıcı eğilimleri açısından ele alınması amaçlanmıştır. Her bir kullanıcı için ürün popülerliği belirlenirken, sistem üzerinden genel bir değerlendirmeden ziyade her bir kullanıcının eğilimini, sistem değerlendirmesini yaparken ki kişi üzerinde etkili olan faktörler dikkate alınmıştır.

II. ÖN ÇALIŞMALAR

Bu bölümde çalışmada kullanılan metotlara yer verilmiştir.

A. SHAP

SHAP (SHapley Additive exPlanations), makine öğrenmesi modellerinin kararlarını açıklamak amacıyla kullanılan güçlü bir açıklanabilirlik (explainability) yöntemidir [8,9]. SHAP, her bir özelliğin (feature) model tahmini üzerindeki katkısını nicel olarak belirlemeyi hedefler. Bu yöntem, işbirlikçi oyun teorisine dayanan Shapley değerleri kavramını temel alır ve her bir özelliğin modelin çıktısına yaptığı marjinal katkıyı hesaplar.

Shapley değerleri, çok oyunculu oyunlarda her oyuncunun oyuna olan katkısını adil biçimde dağıtmak için geliştirilmiş teorik bir yaklaşımdır. SHAP bu teoriyi makine öğrenmesi bağlamında yeniden yorumlayarak, her bir özelliği bir oyuncu, modelin tahminini ise oyunun toplam çıktısı olarak ele alır. Böylece, her özelliğin model kararına yaptığı katkı, diğer tüm olası özellik kombinasyonları üzerinden ortalama etkisi alınarak hesaplanır.

B. Pareto

Pareto İlkesi, İtalyan ekonomist Vilfredo Pareto tarafından ortaya konmuş ve genellikle “80/20 kuralı” olarak bilinen ekonomik bir ilkedir [10]. Bu ilkeye göre, birçok olayda sonuçların büyük bir bölümü, nedenlerin küçük bir kısmından kaynaklanır. Örneğin, bir ekonomide toplam gelirin %80’inin nüfusun %20’si tarafından elde edilmesi gibi.

Matematiksel olarak bu ilke, bir dağılımdaki yoğunluğun belirli bir azınlıkta yoğunlaştığını ifade eder. Öneri sistemleri gibi kullanıcı-tabanlı sistemlerde bu ilke genellikle şu şekilde gözlemlenir:

- Kullanıcıların %20’si, sistemdeki etkileşimlerin (tıklamalar, puanlamalar vb.) %80’ini gerçekleştirir.
- İçeriklerin %20’si, tüm önerilerin %80’ini oluşturur (örneğin en popüler ürünler, filmler veya müzikler).

III. DENEYLER

Bu bölümde çalışmanın ilerletilmesi aşamasındaki test metodolojisi, kullanılan veri seti ve performans ölçütlerine yer verilmiştir.

C. Test Metodolojisi

Bu çalışmada, ilk olarak kullanıcıların her birinin ‘Deneyime Açıklık, Uyumluluk, Duygusal Denge, Sorumluluk, Dışadönüklük’ kriterlerinin baskınlığına göre gruplandırılma işlemi yapılmıştır. Gruplandırma aşamasından sonra, her bir grup için leave one out tekniği kullanılarak veri setleri test ve

eğitim veri seti olarak ayrılmıştır. Sonrasında ise her bir kullanıcıyı kendi grubu içerisinde değerlendirilip, popüler ve popüler olmayan ürünler Pareto prensibine göre belirlenecektir. Son aşamada ise işbirlikçi filtreleme algoritması ile üretilen en iyi-n öneri listesinin popüler yanlılığının etkisi incelenmiştir.

D. Veri Seti

Bu çalışmada, kullanıcıların kişilik özelliklerini dikkate alan öneri sistemleri geliştirmek amacıyla Personality 2018 veri setinin bir alt kümesi kullanılmıştır [11]. Veri seti, 1.780 kullanıcı tarafından 7.228 filme verilmiş toplam 911.369 adet derecelendirmeyi içermekte olup, her kullanıcı için beş büyük kişilik özelliği (Big Five) kapsamında değerlendirilen sayısal kişilik puanları da içermektedir. Bu kişilik boyutları: deneyime açıklık (openness), uyumluluk (agreeableness), duygusalılık (emotional stability), sorumluluk (conscientiousness) ve dışadönüklük (extraversion) şeklindedir. Her bir özellik, [1, 1.5, 2, ..., 7] aralığında ayırık değerlerle temsil edilmiştir. Bu çalışmada, veri seti daha anlamlı analizler yapılabilmesi adına en az 20 kullanıcıdan değerlendirme almış ürünler ve en az 20 ürün değerlendirmesi yapmış kullanıcılar filtrelenerek indirgenmiştir. Bu işlem, hem veri yoğunluğunu artırmak hem de öneri sistemlerinin performansını daha doğru biçimde değerlendirebilmek için gerçekleştirilmiştir. Ayrıca, veri seti öneri sistemleri araştırmalarında sıkça karşılaşılan popülerite yanlılığı problemini barındırmaktadır. Ögelerin küçük bir kısmı büyük çoğunluk tarafından değerlendirilmiş olup, bu durum öneri listelerinde sadece popüler ürünlerin öne çıkmasına ve uzun kuyrukta yer alan içeriklerin göz ardı edilmesine neden olmaktadır. Bu dengesizlikler, öneri algoritmalarının çeşitlilik ve yaratıcılık düzeylerini olumsuz etkileyebilmektedir.

E. Performans Metrikleri

Normalleştirilmiş İndirimli Kümülatif Kazanç (Normalized Discounted Cumulative Gain-nDCG): Öneri sistemlerinde kullanıcıya sunulan ürünlerin sıralama kalitesini ölçen bir metriktir [12]. Kullanıcının ilgisine en uygun ürünlerin üst sıralarda yer alması hedeflenir. Bu metriğin hesaplanabilmesi için önce Cumulative Gain (CG) ve Discounted Cumulative Gain (DCG) hesaplanır. CG, önerilen ürünlerin alaka düzeylerinin toplamı iken, DCG bu alaka düzeylerini ürünlerin listede yer aldığı konuma göre indirger.

Grup Ortalama Popülerlik (Group Average Precision-GAP): Farklı kişilik temelli kullanıcı grupları için yayılan popülerlik yanlılığının düzeyini belirlemek amacıyla bu metrik önerilmiştir [13]. GAP değerinin yüksek çıkması, ilgili kullanıcı grubuna yapılan önerilerin, kullanıcı profillerindeki ürünlere kıyasla daha popüler ürünlere yöneldiğini göstermektedir. Bu durum, öneri sisteminin popülerlik yanlılığı taşıdığını ve kullanıcı tercihleriyle tam uyumlu olmayan, genel anlamda daha çok değerlendirilen ürünleri önerme eğiliminde olduğunu ortaya koyar. Buna karşılık, GAP değerinin düşük olması ise öneri sisteminin kullanıcı profillerine daha yakın, yani grubun geçmiş etkileşimlerine benzer popülerlik seviyesinde ürünler önerdiğini göstermektedir. Bu durum, daha dengeli bir öneri sunumu sağladığı şeklinde yorumlanabilir.

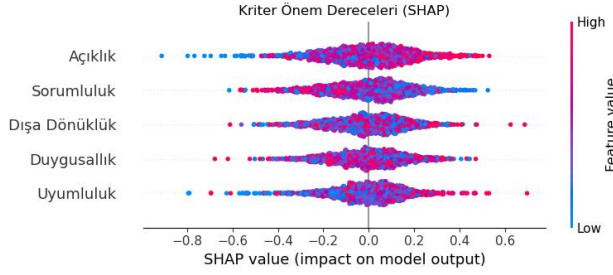
Kuyruk Ürünlerin Önerilerde Ortalama Yüzdesi (Average Percentage of Long Tail Items-APLT): Sıralamaya dayalı ürünleri değerlendirmek için kuyruk ürünlerin ortalama yüzdesini hesaplamaktadır [14]. Yüksek APLT algoritmanın

çeşitlilik açısından yüksek kaliteli önermeler yaptığını tam anlamıyla ifade edemeyebilir ancak düşük APLT öneri listelerinde çeşitlilik açısından eksiklik olduğunu kesin olarak ifade edebilir.

Kuyruk Ürünlerin Kapsamı (Long-Tail Coverage-LTC): Öneri listesinin kuyruk ürünleri ne kadar kapsadığını göstermektedir [15]. Yüksek LTC öneri listesinde daha az popüler ürünlerin olduğunu, daha çok kuyruk ürünlerin yer aldığını göstermektedir.

IV. DENEY SONUÇLARI

Çalışmada ilk olarak SHAP kullanılarak kullanıcıların kişilik özelliklerinin baskınlığına göre gruplama işlemi yapılmış ve sonrasında her bir gruptaki kullanıcılar için popülerlik yanlılığı incelenmiştir.

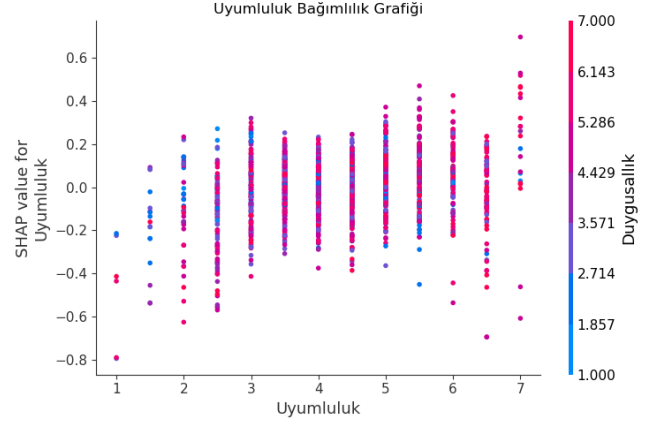


Şekil 1. Kriterlerin SHAP değerleri

Şekil 1, modelin çıktısını etkileyen kişilik özelliklerinin önem derecelerini SHAP değerleri aracılığıyla göstermektedir. Bu analiz, her bir özelliğin modelin karar mekanizmasındaki görece etkisini değerlendirmeye olanak tanımaktadır. Görselde her bir nokta bir kullanıcıyı temsil etmekte olup, renk skalası ilgili özelliğin değerini ifade etmektedir (kırmızı: yüksek değer, mavi: düşük değer). Yatay eksende gösterilen SHAP değeri, özelliğin model çıktısı üzerindeki etkisinin yönünü ve büyüklüğünü göstermektedir.

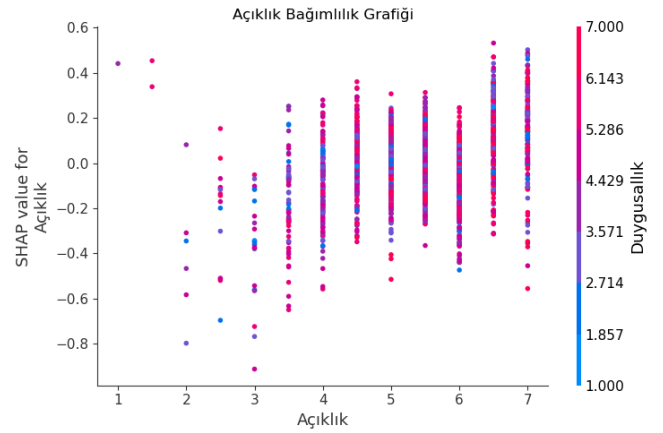
Elde edilen sonuçlara göre, Açıklık ve Dışa Dönüklük özellikleri model çıktısı üzerinde görece daha belirgin bir etkiye sahiptir. Özellikle bu özelliklerde yüksek değerlere sahip kullanıcıların model sonucuna pozitif katkıda bulunduğu, dolayısıyla bu kullanıcıların tercihleriyle modelin daha uyumlu çalıştığı gözlemlenmiştir. Sorumluluk özelliği de benzer bir şekilde pozitif etki yaratmakta; bu durum, daha sorumluluk sahibi bireylerin öneri sistemindeki davranışlarının model tarafından daha kolay öğrenilebilir olduğunu düşündürmektedir.

Buna karşın, Uyumluluk özelliğinin SHAP değerleri çoğunlukla sifira yakın olup, bu özelliğin model kararlarında anlamlı bir fark yaratmadığı görülmektedir. Duygusallık ise pozitif ve negatif yönlü etkileriyle daha değişken bir karakter sergilemektedir; bu da özelliğin model çıktısına olan etkisinin bağlama göre değişkenlik gösterebileceğini işaret etmektedir.



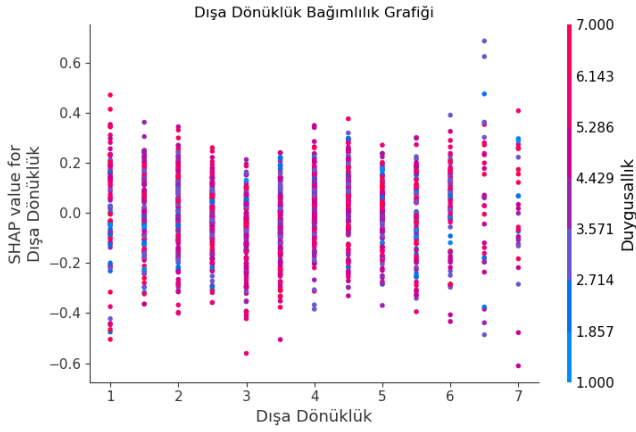
Şekil 2. Uyumluluk bağımlılık grafiği

Şekil 2, Uyumluluk özelliğinin model çıktısına etkisini gösteren SHAP bağımlılık grafiğidir. Grafikte, uyumluluk düzeyi arttıkça SHAP değerlerinin genellikle pozitif yöne kaydığı görülmektedir. Bu durum, yüksek uyumluluk düzeyine sahip kullanıcıların model tarafından daha olumlu değerlendirildiğini, yani model çıktısına daha fazla katkı sağladığını göstermektedir. Ayrıca renk skalası üzerinden görüldüğü üzere, duygusallık değişkeni ile de etkileşim söz konusudur. Duygusallık seviyesi arttıkça (pembe ve kırmızı tonlar), SHAP değerlerinin daha geniş bir aralığa yayıldığı gözlemlenmektedir; bu da modelin bu iki özelliğin birlikte etkisinden etkilendiğini düşündürmektedir.



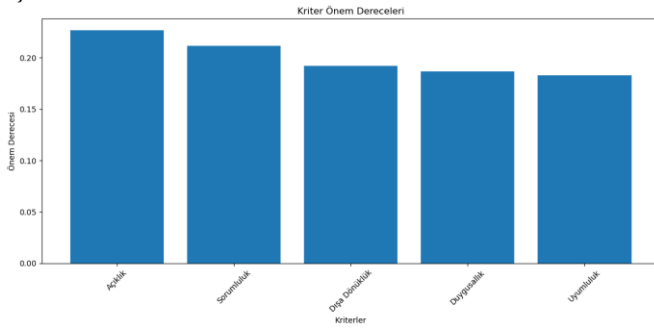
Şekil 3. Kriterlerin SHAP değerleri

Şekil 2, Açıklık değişkeninin SHAP değerlerine olan etkisini göstermektedir. Açıklık düzeyi arttıkça modelin çıktısına katkısının pozitif yönde arttığı gözlemlenmiştir. Duygusallık değişkeni renk skalasıyla temsil edilmesine rağmen, Açıklık ve SHAP değeri arasındaki ilişki üzerinde belirgin bir etkisi bulunmamaktadır. Özellikle düşük Açıklık seviyelerinde daha geniş bir Duygusallık dağılımı dikkat çekmektedir.



Şekil 4. Dışa Dönüklük bağımlılık grafiği

Şekil 4, Dışa Dönüklük değişkeninin SHAP değerleri ile olan ilişkisini göstermektedir. Dışa Dönüklük düzeyleri arttıkça SHAP değerlerinde anlamlı ve sistematik bir değişim gözlemlenmemektedir. Bu durum, modelin Dışa Dönüklük değişkenine karşı nötr bir tutum sergilediğini düşündürmektedir. Renk skalasıyla gösterilen Duygusalılık değişkeni ise SHAP değerleri üzerinde belirgin bir ayrışma oluşturmamaktadır.



Şekil 5. Kriterlerin önem dereceleri

Şekil 5, her bir kriterin model üzerindeki göreceli önem derecelerini (SHAP değerleri ile) göstermektedir. Bu tür grafiklerde, SHAP değerinin büyüklüğü, ilgili kriterin tahmin edilen sonuca (örneğin bir öneri puanına veya ürün sıralamasına) ne ölçüde katkıda bulunduğunu göstermektedir. Açıklık kriteri, en yüksek önem derecesine (~0.22) sahiptir. Bu, modelin öneri üretirken en çok bu kriteri dikkate aldığı ve kullanıcı tercihleri üzerinde en güçlü etkiye sahip olduğunu gösterir. Sorumluluk kriteri (~0.21) ikinci sırada gelmekte ve Açıklık'a oldukça yakın bir öneme sahiptir. Bu da kullanıcıların Sorumluluk kriterine verdikleri önemin model tarafından iyi yakalandığını gösterir. Dışa Dönüklük (~0.19), Dayanıklılık (~0.18) ve Uyumluluk (~0.18) kriterleri ise daha düşük ama birbirine yakın önem derecelerine sahiptir. Bu, modelin bu kriterleri daha dengeli şekilde dikkate aldığı ve bu kriterlerin kullanıcı tercihlerinde daha az baskın ancak yine de anlamlı bir rol oynadığını göstermektedir. Açıklık ve Sorumluluk kriterleri, kullanıcıların kararlarını belirlemede en belirgin faktörlerdir. Bu da muhtemelen kullanıcıların bu kriterlerdeki farklılıkları daha net algıladığını ve tercihlerine daha fazla yansıtıldığını göstermektedir. Diğer kriterlerin göreceli önemlerinin daha düşük ve dengeli olması, modelin bu kriterleri destekleyici ancak belirleyici olmayan faktörler olarak kullandığını düşündürmektedir.

Her bir kritere göre uygulanan yöntem doğrultusunda performans sonuçları aşağıdaki tabloda yer almaktadır.

Table 1. Performans sonuçları

	nDCG	GAP	APLT	LTC
Açıklık	0.8052	-0.8934	0.4919	0.0034
Uyumluluk	0.6067	-0.9455	0.6739	0.0047
Duygusalılık	0.6155	-0.9171	0.602	0.0042
Sorumluluk	0.4381	-0.9479	0.7092	0.0049
Dışadönüklük	0.3573	-0.9543	0.8233	0.0057
Genel	0.5646	-0.7153	0.6601	0.0046

Tablo 1'de her bir kişilik özellikleri ve genel skor için nDCG, GAP, APLT ve LTC performans metrikleri raporlanmıştır. Bulgular, modelin farklı kişilik özellikleri üzerindeki tahmin başarımı hakkında önemli bilgiler sunmaktadır.

nDCG Sonuçları: nDCG, sıralama kalitesini ölçen bir metriktir. Açıklık değişkeni için elde edilen nDCG değeri (0.8052), diğer tüm kişilik boyutlarına kıyasla en yüksek düzeyde bulunmuştur. Bu sonuç, modelin Açıklık boyutundaki sıralamaları diğer kişilik boyutlarına göre daha başarılı bir şekilde gerçekleştirdiğini göstermektedir. Uyumluluk (0.6067) ve Duygusalılık (0.6155) için orta düzeyde nDCG değerleri elde edilmişken, Sorumluluk (0.4381) ve Dışadönüklük (0.3573) için nDCG değerleri daha düşük bulunmuştur. Genel nDCG skoru ise 0.5646 olup modelin tüm kişilik boyutlarında orta düzeyde bir sıralama performansı sergilediğini işaret etmektedir.

GAP Sonuçları: GAP değerleri tüm kişilik boyutlarında negatif olarak bulunmuştur. En düşük GAP değeri Dışadönüklük değişkeni için (-0.9543) gözlemlenmiştir; bu, modelin bu değişken için sıralama performansında ciddi sorunlar yaşadığını göstermektedir. Açıklık değişkeninin GAP değeri (-0.8934) diğerlerine göre daha az negatif olmakla birlikte, tüm boyutlarda GAP değerlerinin negatif olması genel sıralama başarısının düşük olduğunu ve modelin, özellikle doğru sıralama yapma konusunda zorluk yaşadığını göstermektedir. Genel GAP skoru ise -0.7153 ile nispeten daha iyi bir düzeydedir, ancak yine de mükemmel sıralamadan uzak bir performans sergilenmiştir.

APLT Sonuçları: APLT metriği, modelin sıralamanın son aşamalarındaki doğruluk düzeyini ölçmektedir. Buradaki sonuçlar, en yüksek APLT skorunun Dışadönüklük için (0.8233) elde edildiğini göstermektedir. Bu durum, Dışadönüklük değişkeni için genel sıralama kötü olmasına rağmen, sıralamanın son kısımlarında modelin daha iyi bir performans sergilediğini göstermektedir. Benzer şekilde, Sorumluluk (0.7092) ve Uyumluluk (0.6739) değişkenleri için de nispeten yüksek APLT değerleri elde edilmiştir. Açıklık (0.4919) ve Duygusalılık (0.6020) için ise APLT değerleri orta seviyelerde bulunmuştur. Genel APLT skoru 0.6601'dir ve bu değer, modelin sıralamanın son aşamalarında genel olarak daha başarılı olduğunu göstermektedir.

LTC Sonuçları: LTC, modelin doğru şekilde tahmin ettiği etiketlerin kapsamını ölçmektedir. Elde edilen LTC değerleri tüm kişilik boyutları için oldukça düşük bulunmuştur (0.0034 ile 0.0057 arasında). Özellikle Dışadönüklük için en yüksek LTC değeri (0.0057) elde edilirken, Açıklık için en düşük LTC değeri (0.0034) gözlemlenmiştir. Genel LTC skoru ise 0.0046 olarak hesaplanmıştır. Bu sonuçlar, modelin kişilik özelliklerini doğru şekilde kapsamada sınırlı bir başarı gösterdiğini ve tahmin kapsamının dar olduğunu ortaya koymaktadır.

F. SONUÇLAR VE TARTIŞMALAR

Bu çalışmada, kişilik boyutlarına yönelik tahmin performansı dört farklı metrik (nDCG, GAP, APLT ve LTC) kullanılarak değerlendirilmiştir. Sonuçlar, Açıklık değişkeninin diğer kişilik boyutlarına kıyasla daha yüksek bir sıralama başarısına sahip olduğunu göstermiştir. Özellikle nDCG skorunun Açıklık için yüksek olması, modelin bu boyutta daha başarılı tahminlerde bulunduğunu ortaya koymuştur. Buna karşın, GAP değerlerinin tüm değişkenler için negatif olması, sıralama doğruluğu açısından önemli zorluklar yaşandığını göstermektedir. APLT sonuçları, modelin sıralamanın son aşamalarında nispeten daha iyi performans sergilediğini belirtmektedir. Ancak, LTC değerlerinin genel olarak düşük olması, kişilik boyutlarının doğru kapsamında modelin sınırlı bir başarıya sahip olduğunu ortaya koymuştur. Bu bulgular, modelin belirli kişilik boyutlarında (özellikle Açıklık) güçlü tahminler yapabilmesine rağmen, genel doğruluk ve kapsam açısından geliştirilmesi gerektiğini işaret etmektedir.

Elde edilen bulgular doğrultusunda, gelecekteki çalışmalar için birkaç öneri sunulabilir. Öncelikle, modelin sıralama doğruluğunu artırmak amacıyla daha karmaşık mimariler (örneğin derin öğrenme tabanlı sıralama modelleri) veya sıralama odaklı kayıp fonksiyonları (ranking loss) kullanılabilir. Ayrıca, GAP skorlarını iyileştirmek için özellikle az temsil edilen sınıflar üzerinde veri dengesizliğini gidermeye yönelik stratejiler uygulanabilir. LTC değerlerinin artırılması amacıyla, daha zengin özellik mühendisliği (feature engineering) yapılması ve modelin kişilik özelliklerinin nüanslarını daha iyi yakalayabilmesi için bireysel farklılıkları dikkate alan modellerin geliştirilmesi önerilmektedir. Son olarak, Duygusalık ve Dışadönüklük gibi modelin düşük performans gösterdiği boyutlarda, farklı veri kaynakları ve çok boyutlu yaklaşımlar (örneğin çoklu görev öğrenme, multi-task learning) ile performans artırımı hedeflenebilir.

REFERENCES

- [1] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, pp. 734-749, 2005.
- [2] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," *arXiv preprint arXiv:1301.7363*, 2013.
- [3] H. Abdollahpouri, "Popularity Bias in Ranking and Recommendation," in *Proc. AAAI/ACM Conf. on AI, Ethics, and Society (AI/ES'19)*, Honolulu, HI, USA, Jan. 2019.
- [4] E. Brynjolfsson, Y. J. Hu, and M. D. Smith, "From niches to riches: Anatomy of the long tail," *Sloan Management Review*, vol. 47, no. 4, pp. 67-71, 2006.
- [5] T. T. Nguyen, P.-M. Hui, F. M. Harper, L. Terveen, and J. A. Konstan, "Exploring the filter bubble: the effect of using recommender systems on content diversity," in *Proc. 23rd Int. Conf. on World Wide Web*, ACM, pp. 677-686, 2014.
- [6] P. Resnick, R. K. Garrett, T. Kriplean, S. A. Munson, and N. J. Stroud, "Bursting your (filter) bubble: strategies for promoting diverse exposure," in *Proc. 2013 Conf. on Computer Supported Cooperative Work Companion*, ACM, pp. 95-100, 2013.
- [7] Ö. Celma and P. Cano, "From hits to niches?: or how popular artists can bias music recommendation and discovery," in *Proc. 2nd KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition*, ACM, p. 5, 2008.
- [8] S. M. Lundberg and S. Lee, "A unified approach to interpreting model predictions," in *Proc. 31st Int. Conf. on Neural Information Processing Systems*, pp. 4768-4777, 2017.
- [9] "L. S. Shapley, "A value for n-person games," RAND Corporation, Paper P-295, 1953.
- [10] R. Sanders, "The Pareto principle: its use and abuse," *Journal of Services Marketing*, vol. 1, no. 2, pp. 37-40, 1987.

- [11] T. Chamorro-Premuzic, *Personality and Individual Differences*, John Wiley & Sons, 2016.
- [12] E. Yalcin and A. Bilge, "Evaluating unfairness of popularity bias in recommender systems: A comprehensive user-centric analysis," *Information Processing & Management*, vol. 59, no. 6, 2022, doi: <https://doi.org/10.1016/j.ipm.2022.103100>.
- [13] H. Abdollahpouri, M. Mansoury, R. Burke, and B. Mobasher, "The unfairness of popularity bias in recommendation," *arXiv preprint arXiv:1907.13286*, 2019.
- [14] H. Abdollahpouri, "Popularity Bias in Recommendation: A Multi-stakeholder Perspective," *arXiv preprint arXiv:2008.08551*, 2020.
- [15] H. Abdollahpouri, M. Mansoury, R. Burke, and B. Mobasher, "The connection between popularity bias, calibration, and fairness in recommendation," in *Proc. 14th ACM Conf. on Recommender Systems (RecSys'20)*, pp. 726-731, 2020.