

## Makine Öğrenimi ile Uzun Kuyruk Ürünler için İyileştirilmiş Sonraki Öğe Önerisi

Ahmet Zencirli<sup>1\*</sup>, Harun Çetin<sup>2</sup>, Nedim Tuğ<sup>3</sup>, Engin Seven<sup>4</sup>, and Tolga Ensari<sup>5</sup>

<sup>1\*</sup> Ar-Ge Departmanı, MNM Teknoloji, İstanbul, Türkiye (ahmet@mnm.com.tr) (ORCID: 0000-0002-9456-395X)

<sup>2</sup> Ar-Ge Departmanı, MNM Teknoloji, İstanbul, Türkiye (harun@mnm.com.tr) (ORCID: 0000-0001-8681-8919)

<sup>3</sup> Ar-Ge Departmanı, MNM Teknoloji, İstanbul, Türkiye (nedim@mnm.com.tr) (ORCID: 0000-0001-8449-2230)

<sup>4</sup> İstanbul Üniversitesi-Cerrahpaşa, Bilgisayar Mühendisliği, İstanbul, Türkiye (engin.seven@ogr.iuc.edu.tr) (ORCID: 0000-0002-7994-2679)

<sup>5</sup> Arkansas Tech University, Computer and Information Science, Russellville, USA (tensari@atu.edu) (ORCID: 0000-0003-0896-3058)

**Türkçe Özet** – Elektronik ticaret platformlarında birçok farklı ürün türü müşterilerin nerede olduklarından bağımsız olarak satılabilmektedir. Bu platformlarda bulunan öneri sistemi kullanıcılar için ilgi çekici ürünlerin seçilmesi ve görüntülenmesinde kritik rol oynamaktadır. Yapılan bu çalışmada elektronik ticaret platformlarında bulunan müşterilere bir sonraki alacakları ürünlerin en doğru şekilde tavsiye edilmesi için makine öğrenmesi algoritmaları kullanılmış sonuçlar karşılaştırılmıştır. Tekil değer ayrışımı (Singular value decomposition-SVD) yönteminin daha başarılı sonuçlar elde ettiği gösterilmiştir.

**Anahtar Kelimeler** – Uzun kuyruk ürünler, tekil değer ayrışımı, en yakın komşuluk, matris ayrıştırma, sonraki öğe önerisi.

**Atf:** Zencirli, A., Çetin, H., Tuğ N., Seven E., Ensari T. (2022). Makine Öğrenimi ile Uzun Kuyruk Ürünler için İyileştirilmiş Sonraki Öğe Önerisi. International Journal of Multidisciplinary Studies and Innovative Technologies, 6(1): 97-103.

## Improved Next Item Recommendation for Long Tail Products with Machine Learning

**Extended Abstract** – Many different types of products can be sold on electronic commerce platforms. Products can be sold regardless of where customers are. The recommendation system on these platforms plays a critical role in selecting and displaying interesting products for users. In the study, the products to be purchased next to the customers were recommended in the most accurate way. For this, machine learning algorithms were used and the results were compared. The singular value decomposition (SVD) method has achieved more successful results.

**Research Problem/Questions** – To make the most appropriate match between an infinite number of products and many customers in the most accurate way.

**Short Literature Review** – Many algorithms have been developed for the best solution in recommendation systems that try to persuade their customers to sell niche products, and it is currently a subject open to research.

**Methodology** – Singular value decomposition (SVD) and  $k$  nearest neighbor (kNN) algorithms are used. Products in the long queue in electronic commerce have been improved by developing the next item recommendation system.

**Results and Conclusions** – Machine learning algorithms can be used to solve problems in product recommendation systems. The SVD method suggested less erroneous recommendations for large datasets.

**Keywords** – Long tail products, singular value decomposition,  $k$ -nearest neighbor, matrix factorization, next item recommendation

**Citation:** Zencirli, A., Çetin, H., Tuğ N., Seven E., Ensari T. (2022). Improved Next Item Recommendation for Long Tail Products with Machine Learning. International Journal of Multidisciplinary Studies and Innovative Technologies, 6(1): 97-103.

## I. GİRİŞ

[1] numaralı referansta yapılan çalışmada, elektronik pazarlarda satılan ürünler için e-ticaret talebinin uzun kuyruğu gösterilmiştir. Toplanan veriler amazon.com sitesinden alınmış olup 200 den fazla kitap kategorisinde ve 250.000 den fazla kitap ile çalışma yapılmıştır. Talep ve gelir arasındaki ilişki incelenmiştir. PageRank algoritmasının bir türevidi kullanılarak bir ürünün tavsiye sistemi ile olan ilişkisi gözlemlenmiştir. Ayrıca Gini katsayısı kullanılarak gelir ve talep dağılımındaki anormallik ortaya konulmuştur [1].

Öneri sistemlerinde sadece doğruluk oranlarına odaklanmak tavsiye kalitesinin ölçülmesinde eksikliklere yol açmaktadır. Bu nedenle doğruluğun tek ölçüt olarak alınması değil doğruluk ölçütüne ek olarak çeşitlilik, benzerlik ve uzun kuyruk teorisi gibi ölçütlerde eklenmiştir. Yapılan çalışmada bu ek ölçütleri de dikkate alarak Markovian grafiğine dayalı bir öneri sistemi geliştirilmiştir. Minnesota Üniversitesi'ndeki GroupLens Araştırma Projesi tarafından toplanan MovieLens100K veri kümesi ile Last.fm çevrimiçi müzik sisteminden toplanan Last.fm veri kümesi kullanılmıştır [2].

İşbirlikçi filtre tabanlı öneri sistemlerinde popüler ürünlerin daha fazla önerildiği ancak ürünlerin daha az önerildiği belirtilmiştir. Bu sorunun üstesinden gelmek için grafik tabanlı öneri sistemi geliştirilmiştir. Vuruş süresi, emme süresi ve emme maliyeti algoritmaları kullanılarak sistemin performansı artırılmaya çalışılmıştır. Ayrıca, farklı kullanıcı kalemi derecelendirme çiftlerindeki varyasyonu ayırt etmek için entropi-maliyet modeli önerilmiştir. İki veri kümesi üzerinde testler gerçekleştirilmiştir. Bu veri kümeleri sırasıyla bir Çin Web 2.0 web sitesi, en büyük çevrimiçi Çince kitap, film ve müzik veritabanı ve 50 milyondan fazla kullanıcısı olan Çin'deki en büyük çevrimiçi topluluktan alınan Douban Veri Kümesi, ikinci olarak MovieLens'ten gerçek bir film veri kümesidir [3].

[4] numaralı referansta yapılan çalışmada, pazara geleneksel olarak en çok satan birkaç ürün hakimdir. Ancak internet pazarları, ürünlerin yarattığı satışların payını artırma potansiyeline sahiptir. Ürün mevcudiyetindeki varyasyonları kontrol edip ve talep yönlü faktörlerin İnternet üzerinden niş ürünlerin artan satışlarıyla ilişkili olup olmadığı araştırılmıştır. Yapılan çalışmada internet kanalı ve katalog kanalı ürün satışları üzerine çalışılmıştır [4].

[5] numaralı referansta yapılan çalışmada, çok seviyeli benzerlik özellikleri ve kullanıcı geçmişi kullanılarak içerik tabanlı filtrelemeye dayanan yeni bir anahtarlama yöntemi önerilmiştir. Önerilen hibrit yöntem benzer kullanıcılara dayalı olarak tahmin edilen derecelendirmeyi hesaplayan ortak bir filtreleme bileşeni ve kullanıcının geçmişte derecelendirdiği diğer benzer öğeleri kullanarak tahmin edilen derecelendirmeyi hesaplayan içerik tabanlı filtreleme bileşeninden oluşmaktadır. İki gerçek veri kümesi üzerinde deneyler gerçekleştirilmiş başarılı sonuçlar elde edildiği ifade edilmiştir. Bu veri kümeleri sırasıyla MovieLens 100K ve MovieLens Im'dir.

Doğruluk ve uzun kuyruk tavsiye sistemleri için iki önemli hedef olmakla birlikte aynı zamanda bunlar çelişkilidir ve aynı

anda maksimize edilemezler. Doğruluk ve yeniliği dikkate alan uzun kuyruk önerisi çok amaçlı bir optimizasyon problemi olarak formüle edilmiştir. İkili olarak formüle edilen bu problem için çok amaçlı evrimsel algoritma kullanılmıştır. Evrimsel algoritma aday çözümler üretmek en iyi çözümü kullanıcıya iletmektedir. Üç veri kümesi MovieLens, Jester veri kümesi ve Netflix veri kümesi, önerilen algoritmanın performansını değerlendirilmiştir [6].

[7] numaralı referansta yapılan çalışmada, bir uzun kuyruk değerlendirme kriterleri önerilmiştir ve farklı modeller arasında uzun kuyruk tavsiyesindeki performansı karşılaştırılmıştır. Davranış özellikleri, ürün özellikleri ve zaman özellikleri olarak tasnif edilmiştir. Ayrıntılı uzun kuyruk kriterleri belirlenmiştir. Böylece uzun kuyruk probleminin daha iyi ölçülmesi sağlanmıştır. Veri kümesi, bir e-ticaret platformu olan Tmall'ın 200 günlük kullanıcı ve öğelerin temel bilgilerini ve ayrıca kullanıcı davranışlarını içeren kayıtlardan oluşmaktadır.

Öte yandan, ürünlerin nasıl ele alındığı ve ürünlerin uzun süre neden durgun kaldığı tartışılmaktadır. Denetimli öğrenme algoritmalarında Random Forest algoritması ile soğuk ürün ve uzun kuyruk problemine çözüm aranmıştır [8].

[9] numaralı referansta yapılan çalışmada, popülerliklerinin neden olduğu önyargıyı dengelemek için uzun kuyruklu ürünlere daha yüksek ve popüler ürünlere daha az ağırlık veren bir ağırlıklandırma şeması kullanılmıştır. Uzun kuyruklu katalog kapsamı ile doğruluk arasındaki dengeyi, herhangi bir kullanıcı ve öğe çifti için ağırlıklı bir karma puan hesaplaması olarak modellenmiştir. MovieLens1M ve Epinions isimli iki veri kümesi üzerinde deneyler gerçekleştirilmiştir.

Kullanıcı davranışını temel alan işbirlikçi filtreleme algoritmaları grup açısından yola çıktığı için bireysel yetenekler göz ardı edilmektedir. [10] numaralı referansta yapılan çalışmada, önerilen model iki ana bölüme ayrılmış olup sırasıyla kişisel deneyim gelişimine dayalı puanlama tahmini ve deneyim düzeyine dayalı en iyi öneridir. Bireysel seviyeden kullanıcı beğenilerinin evrimi modellenip daha sonra kullanıcı deneyimi düzeyine ve öğelerin ilgili popüleritesine göre öneri puanlama modelini oluşturulmuştur. Veri kümesi olarak inceleme sitesi RateBeer'in inceleme verileri kullanılmıştır.

[11] numaralı referansta yapılan çalışmada, talep tarafı faktörünün (çevrimiçi kullanıcı incelemeleri) ve arz tarafı faktörünün (ürün çeşitliliği) çevrimiçi yazılım indirme bağlamında uzun kuyruk ve süper yıldız fenomeni üzerindeki etkisini incelemektedir. Ürün çeşitliliğinin tüketicilerin çevrimiçi karar verme sürecinde oynadığı rolün daha kapsamlı bir şekilde anlaşılmasını sağlamayı amaç edinilmiştir. Çevrimiçi kullanıcı incelemelerinin ve ürün çeşitliliğinin, farklı popülerliğe sahip yazılım programlarının kullanıcı seçimlerini etkilemek için nasıl etkileşime girdiğini araştırmak için nicel regresyon metodolojisi kullanılmıştır. Bu makalenin amacı, çevrimiçi alışveriş ortamındaki hem talep hem de arz yan faktörlerinin kullanıcı seçimlerini nasıl etkilediğini ve böylece toplu olarak çevrimiçi tüketim kalıplarındaki değişikliklere nasıl katkıda bulunduğunu göstermektir. Veri kümesi 30.000'den fazla ücretsiz veya denemesi ücretsiz

yazılım programının bir koleksiyonu olan CNET Download.com'dan (CNETD) alınmıştır.

[12] numaralı referansta yapılan çalışmada, üçlü grafik öneri sistemi geliştirilmiştir. Geliştirilen model kullanıcı ögesi, kullanıcı türü ve tür ögesi olmak üzere üç katmanı birbirine bağlamaktadır. Verileri daha iyi kullanmak için kullanıcı-tür ilişkisini ve grafikte daha fazla yol oluşturmak için tür-öge ilişkisi kullanılmıştır. Üçlü grafikte daha fazla yol oluşturmak için "temel türler" kavramını kullanılır ve daha düşük dereceli türlere çok fazla ağırlık vermekten kaçınmak için basit bir ortalamayı kullanmaktan Bayes ortalamasını tercih edilmiştir. Ayrıca, uzun kuyruklu öneri problemini çözmek için Gizli Semantik İndekslemede kullanılan Tekil Değer Ayırıştırması benimsenmiştir. Veri kümesi olarak MovieLens üzerinde testler gerçekleştirilmiştir.

[13] numaralı referansta yapılan çalışmada, işbirliğine dayalı filtre ile yapılan tavsiye sistemlerini iyileştirmek için Hybrid Reranking Framework in Collaborative Filtering (HyReCF) isimli yeni bir çerçeve önermişlerdir. Önerilen yaklaşım iki aşamada gerçekleştirilir. İlk kısımda (tahmin aşaması olarak anılır), bir tahmin listesi geleneksel bir işbirlikçi filtreleme tekniği uygulayan her kullanıcı için oluşturulur. Bu yaklaşımın ana fikri, kullanıcılara daha çeşitli ve uzun kuyruklu öğeler sunmak için bu çerçeveyi mevcut işbirlikçi öneri sistemiyle birlikte dağıtmaktır. Tahmin listesi için kullanıcı tabanlı CF, madde tabanlı CF, matris faktoring gibi bilenen yöntemler kullanılmıştır. İkinci kısımda, ögenin popülerliği, kullanıcıların geçmiş derecelendirmelerine benzemeyen öğelerin bulunması ve iyi tahmin edilen öğeler gibi önemli faktörlerin gözlemlenmesi gerekir. Öngörülen öğeler, maksimum çeşitliliği elde etmek ve uzun kuyruk öge önerilerini iyileştirmek için farklı sınıflara ayrılmıştır. MovieLens ve Netflix olmak üzere iki gerçek dünya film derecelendirme veri seti ile test edilmiştir.

[14] numaralı referansta yapılan çalışmada, Hitting Time Clustered Algorithm (HTCL) isimli bir model önerilmiştir. Önerilen model temel olarak vuruş süresi (Hitting Time) ve kümeleme (clustering) yaklaşımlarını temel almaktadır. Vuruş süresi ile öğeler ve kullanıcılar arasındaki yakınlığı hesaplamak amaçlanmıştır. Kümeleme yaklaşımı ile önerileri daha uzun kuyruklu maddelere odaklayacaktır. Veri kümesi olarak MovieLens100k kullanılmıştır.

[15] numaralı referansta yapılan çalışmada, Co-occurrence based Enhanced Representation model (CER) isimli bir model önerilmiştir. Önerilen model hedef öğelerin anlamlarını tamamlamak için geliştirilmiş bir temsil etmeye çalışılmıştır. Burada, hedef ögenin birlikte meydana gelen öğelerine dayalı olarak geliştirilmiş temsili öğrenme önerilmiştir. Gürültü azaltma ve hesaplama karmaşıklığı önlemek için birlikte meydana gelen öge setinden sabit sayıda ögeyi örneklemiştir. Bu örnekleme stratejilerinde iki kriter göz önünde bulundurulur. Birincisi, hedef ögeye yüksek anlamsal benzerliğe sahip öğeleri örnekleme. Bir diğeri, yerleştirmeleri düşük anlamsal belirsizliğe sahip öğeleri örnekleme. Birlikte meydana gelen öğelerin farklı önemini ayırt etmek için dikkat mekanizması kullanılmıştır. AmazonBeauty and Amazon-Books veri kümelerinde deneyler yapılmıştır.

Matris çarpanlara ayırma, öge tabanlı, kullanıcı tabanlı, çeşitli alanlarda yaygın olarak kullanılan teknik olan işbirlikçi filtreleme adı verilen bir öneri sistemi sınıfına aittir. Bu sistemlerin genelde popüler ürünleri önermede ön yargılı olduğu ifade edilmiştir. Önerilen model, yaygın olarak kullanılan ortak filtreleme tekniğini yeni bir grafik tabanlı teknikle birleştirilmiştir. Önerilen yaklaşımdaki öneriler üç aşamada gerçekleştirilir. İlk aşamada, uzun kuyruklu kalemlere odaklanmak için grafik tabanlı yaklaşım kullanılır. Daha sonra öge-öge etkileşim matrisinden yararlanarak her aktif kullanıcı için ağırlıklı bir graf oluşturulur, son olarak, oluşturulan grafın her düğümü (ögesi) için entropi ve ters popülerlik hesaplanır ve ardından, bunlara dayalı öneriler için öğeleri seçilmiştir. MovieLens 100K and MovieLens 1M datasets veri setlerinde deneyler gerçekleştirilmiştir [16].

Yapılan çalışmada güç yasası dağılımı ile birleşen bir işbirlikçi filtreleme öneri algoritması önermektedir. Ayrıca, güç yasası dağılımını yerelliğe duyarlı karma (locality sensitive hashing-LSH) ile birleştirilmiştir. bu yazıda, yaklaşık en yakın komşu araması için verimli bir indeks yapısı olan seyahat verileri ve yerelliğe duyarlı karma (LSH) özelliklerini kullanarak (tourism recommendation systems) TRS'lerin öneri çeşitliliğini ve hesaplama verimliliğini iyileştirmeye çalışılmıştır. Lokasyona duyarlı hash fonksiyonlarını kullanarak, kullanıcıları yüksek boyutlu vektörlerden düşük boyutlu ikili tamsayılar hash eder ve benzer seyahat tercihlerine sahip kullanıcıları aynı paketlere ayırır. Sadece aynı paketteki kullanıcılar arasındaki benzerlikleri hesaplanır ve aynı paketteki kullanıcıların seyahat tercihlerine göre önerilerde bulunulur, bu da yüksek boyutlu seyahat verilerinin hızlı en yakın komşu araması ve aşırı seyreklik sorununa uygulanabilir bir çözüm üretir. Kullanılan veri seti Sina Weibo check-in verileridir [17].

[18] numaralı referansta yapılan çalışmada, içerik tabanlı filtreleme baz alınarak otonom olarak kendisini güncelleyen öneri tavsiye sistemi oluşturulmuştur. Oluşturulan modelde RBM ve ContentKNN, RBM ve SVD++, AutoEncoder ve ContentKNN, AutoEncoder ve SVD++ gibi algoritmalar bir arada kullanılmış ve hibrit bir model tasarlanmıştır. Sonuçlar karşılaştırmalı olarak analiz edilmiştir. Çalışmada, gerçek online ticari satış ortamından elde edilen kullanıcı-ürün veri kümesi kullanılmıştır.

[19] numaralı referansta yapılan çalışmada, önerilen model, ürün popülerlik bilgilerinin kullanılmasına dayanmaktadır. "Uzun kuyruk" önerisinin, her kümenin dengeli öge popüleritesi ile kesin olarak çıkarılabileceği gösterilmiştir. Veri seyrekliği sorununu ele almak için, modellemeyi ve güçlü etkili temsilleri anlamak için SDAE (Yığın Gürültü Giderici Otomatik Kodlayıcı) kullanan derin öğrenme modeli önerilmiştir. Web hizmeti önerisi için veri seyrekliğini ele almak ve performansı artırmak için evrişimli sinir ağları önerilmiştir. Gerçek dünya veri kümesi üzerindeki deneysel sonuçlar, bu tür ortak otomatik kodlayıcı tabanlı temsil ve içerik kullanımı öğrenme çerçevesi ile önerilen algoritmanın en son teknoloji temel çizgilerinden önemli ölçüde daha iyi performans gösterdiğini göstermektedir.

Literatürde, genellikle kullanıcıları ve öğeleri düşük boyutlu uzayda vektörler olarak temsil eden matris ayırıştırma (matrix factorization-MF) dayalı yerleşik bir metodoloji kullanılmaktadır. Bu çalışmada, MF tabanlı kişiselleştirilmiş

öneriler için hızlı ve doğru erişim sağlayan literatürdeki son gelişmeleri ve son teknoloji yaklaşımları araştırılmıştır. Ayrıca farklı boyutlardaki yaklaşımların analitik tartışılarda incelenmiştir. Bu anket, tavsiye alma aşamasına odaklanır ve matris çarpanlara ayırma önerisinin verimli bir şekilde alınması için son gelişmelere genel bir bakış sunmuştur. Her adım için, yüksek öneri doğruluğunu korurken, alma verimliliğini artırmak için ilgili yöntemleri ilgili stratejilerine göre kategorize edilmiştir [20].

Geleneksel öneri algoritmasının doğruluğuna dayalı yaklaşımı yeterli görülmemektedir. Bundan dolayı yenilik, çeşitlilik, doğruluk ve hatırlamadan oluşan dört öneri hedefini aynı anda optimize etmek ve çok amaçlı optimizasyon önerisi için geliştirilmiş bir matris çarpanlara ayırma tabanlı model önerilmiştir. Çift katmanlı öneri modeli tasarlanmıştır. Alt katman, geliştirilmiş bir MF algoritmasını benimser. Alt katmanın (improved matrix factorization-IMF) rolü, bilinmeyen öğelerin derecesini tahmin eden geleneksel öneri yöntemlerine benzer. Önerilen IMF algoritmasının avantajı, elde edilen tahmin derecelendirmesinin çeşitlilik, yenilik ve doğruluk olmak üzere üç hedefi bütünleştirmesi için ek düzenleme öğelerine sahip olmasıdır. Önerilen algoritmayı optimize etmek ve parametre matrisini elde etmek için stokastik gradyan inişi (SGD) kullanılır. En üst katmanda, çok amaçlı evrimsel algoritmaya dayanan temsili algoritmaların biri olan ve Pareto'nun egemenliğine dayanan NSGA-III (klasik bir model for many-objective optimization recommendation), önerilen öneri modelimize uyacak şekilde uygun şekilde düzenlenmişlerdir [21].

İşbirliğine dayalı öneri sistemlerinin çoğu, kullanıcıları ve öğeleri ortaklaşa temsil etmek için yerleştirmeleri (embedding) öğrenerek çalışmaktadır. Yapılan çalışmada Matrix Factorization (MF) tarafından öğrenilen yerleştirmelerin güvenilirliğini ne derece doğru olduğu tartışılmıştır. Gizli faktörlere verilen ilk değerleri değiştirmek suretiyle aynı MF yönteminin çok farklı öğeler ve kullanıcılarda oluşturduğu etkiyi ölçmeye çalışılmıştır. Bu etkinin daha az popüler olan öğeler için ne derecede etkili olduğu incelenmiştir. Bu problemi çözebilmek için öğeler ve kullanıcılar hakkındaki bilgileri komşularına yayan En Yakın Komşular Matrisi Çarpanlarına ayırma (NNMF) adı verilen bir MF genellemesi tasarlanmıştır. Tavsiyeleri ve temsilleri destekleyen bilgi miktarını genişletmek için yaygın MF yaklaşımının NNMF varyantlarını tanımlanmıştır. Beş farklı veri kümesi üzerinde kapsamlı deneyler gerçekleştirilmiştir [22].

[23] numaralı referansta yapılan çalışmada, öneri sistemlerinde uzun kuyruk problemlerinin sebeplerinden birinin birçok öğenin sadece birkaç derecelendirilmesi ile temsil edilmesinden kaynaklandığı belirtilmiştir. Bu problemin üstesinden gelmek için tüm öğe kümesini baş ve kuyruk parçalarına bölerek sadece kuyruk öğelerini kümelemektedir. Kuyruk da kümelenebilir öğeler için öneriler aynı kümedeki derecelendirmelere dayanmaktadır. Bölümleme ve kümeleme en uygun şekilde yapılırsa iyi bir hesaplama performansını korur, kuyruk öğeleri için öneri hata oranlarını azalttığı ifade edilmiştir.

[24] numaralı referansta yapılan çalışmada, tavsiye sistemleri için genel bir çerçeve oluşturulmuştur. Etkileşimin yaratıcı öneriler için bir dizi fırsat sunduğu bir örnek olarak eBay'i kullanılmıştır. Bir pazar platformu için tavsiye sistemleri

oluşturmadaki zorlukları, fırsatları ve yaklaşımları tanımlanmaya çalışılmıştır. Tavsiye sistemlerindeki sorunların, sıklıkla görüldüğü gibi, sadece çapraz satışın ötesinde olduğu ve bu ruhla, Kim, Ne, Neden, Nerede, Ne Zaman ve Nasıl tavsiyelerini içeren tavsiye sisteminin 5 W ve H'sini tartışılmıştır. Ayrıca, öneriler yalnızca ürünler ve öğeler için değil, öğe kategorileri, ilgili sorgu önerileri ve markalar için de geçerli olduğu belirtilmiştir [24].

[25] numaralı referansta yapılan çalışmada, daha önceki hiçbir araştırmanın, çeşitli satış stratejileri bağlamında tavsiyenin sağlandığı perspektife ve stratejilerin müşteriler, ürünler ve USP ile nasıl bağlantı kurduğuna dikkat etmemiş olması bakımından önemli olduğu vurgulanmıştır. Daha doğru önerilerde bulunmak için kontrollü değişkenleri ve satış stratejilerini içermektedir. Bu çalışmada üç hipotez (H1, H2 ve H3) önerilmektedir. H1 ve H2 hipotezleri birincil verilerle test edilir ve doğrulanır. Hipotez H3, mevcut literatüre karşı test edilmiş ve doğrulanmıştır. H1, İşletme kişiselleştirilmiş dijital pazarlamayı ne kadar çok gerçekleştirirse, gelir artış oranı o kadar yüksek olur. Kişiselleştirilmiş dijital pazarlama ve gelir yönetimi arasındaki potansiyel sinerji, pazarlama hunisinin iki farklı bileşeninin kontrolünü sağlar. Daha yüksek bir değer, şirketin sürdürülebilir büyümesi anlamına gelir. H2, işletme kişiselleştirilmiş dijital pazarlamayı ne kadar çok gerçekleştirirse, AOV (Average Order Value-kullanıcı tarafından bir siparişi tamamlamanın potansiyel değerine ilişkin bir tahmin sağlar ve yüksek bir değer performansı ve satışları artırması muhtemeldir.) oranı o kadar yüksek olduğu ifade edilmiştir. Daha yüksek bir değer, müşterinin satın alma alışkanlıklarında bir artış anlamına gelmektedir. H3, İşletme kişiselleştirilmiş dijital pazarlamayı ne kadar çok gerçekleştirirse, halka arz oranı o kadar yüksek olur. Daha yüksek bir değer, toplam gelirden daha fazla kazanılan, yani daha fazla yeni müşteri çekmek anlamına gelmektedir. Önerilen model e-pazarın tasarımı, kişiselleştirilmiş tavsiye modeli ve tavsiye süreci olmak üzere üç katmanlıdır. Modelin gücü, öğeleri e-pazarı tasarımına uygun olarak önermek ve önerilen kişiselleştirilmiş öneri modelinin benimsenmesiyle önerilen öneri sürecini takip etmek olarak belirtilmiştir [25].

[26] numaralı referansta yapılan çalışmada, ilk olarak geri bildirim puanı olarak kullanıcının satın alma sayısına dayalı bir kullanıcı-öge matrisi oluşturularak bu sorunu ele almaktadır. İkinci olarak genelde kullanılan İşbirlikçi Filtreleme yaklaşımını belirli algoritmalar altında incelemektedir. Hedef müşteri tarafından verilen derecelendirmeyi tahmin etmek için Bellek tabanlı ve Model tabanlı İşbirlikçi Filtreleme kullanarak daha iyi bir yaklaşım bulmak için testler gerçekleştirilmiştir. Sonuç olarak, k-NN'li Bellek tabanlı CF'nin SVD aracılığıyla Model tabanlı CF'den daha iyi performans gösterdiği vurgulanmıştır [26].

[27] numaralı referansta yapılan çalışmada, derecelendirme sayısına, derecelendirmelerin ortalama değerine veya tavsiye sıklığına dayalı olarak bir öneri veri setinden uzun kuyruk öğelerini tahmin etmek için belirli stratejiler önerilmiştir. Her uzun kuyruk öğesi için bir uygunluk modeli oluşturan olasılıksal bir işbirlikçi filtreleme algoritması olan (item-based Relevance Model) IRM2'yi tasarlanmıştır. Alaka Tabanlı Dil Modelleri (Relevance-based Language Models) kullanılmıştır (Sıklıkla Uygunluk Modelleri veya RM olarak kısaltılır). Bu yöntem, sözde alaka düzeyi geri bildirim yoluyla otomatik sorgu genişletme gerçekleştirmek

için en gelişmiş tekniktir. İlgili Modellerini de kullanarak işbirlikçi filtrelemeye farklı bir yaklaşım tasarlanmıştır. Her uzun kuyruklu ürün için bir Uygunluk Modeli oluşturulmasını önerilmiştir. Benzer öğelerden gelen bilgileri kullanarak öğe profillerini ilgili kullanıcılarla genişleterek modelin performansı iyileştirilmiştir.

[28] numaralı referansta yapılan çalışmada, işbirlikçi filtreleme için sağlam bir negatif olmayan matris çarpanlara ayırma modeli oluşturulmuştur. Ek olarak modelin kararlılığının bir analizini yapılmıştır. Tahmin edilen derecelendirmelerin doğru ve negatif olmadığını garanti eden, negatif olmayan matris çarpanlarına ayırmanın yinelemeli güncelleme algoritmasına dayanan özellik matrislerinin yinelemeli bir optimizasyon yöntemi önerilmiştir. Önerilen yöntemde matris çarpanlarına ayırma sürecinde, öznel vektörlerinin girişlerindeki negatif olmayan kısıtlama çok önemlidir çünkü öğrenilen öznel özelliklerin CF sistemlerinde kullanıcı ilgileri veya tercihleri gibi gerçek dünya anlamlarını temsil etmesini sağlayabilmektedir. NMF yöntemleri, CF sistemlerinde kullanıcı derecelendirmelerinin doğal olarak negatif olmamasını garanti eden gizli özellikler üzerinde negatif olmayan kısıtlamalar getirebilmektedir. NMF modeli, kullanıcı öğesi derecelendirme matrisi  $R^i$ 'yi iki düşük dereceli negatif olmayan faktör matrisi  $W$  ve  $H^i$ 'ye ayırır ve belirli bir öğe için belirli kullanıcının tercihlerini tahmin etmek için bunların doğrusal kombinasyonunu kullanmaktadır. Model MovieLens 10M ve Netflix veri kümesi üzerinde test edilmiştir.

Tüketiciler, çevrimiçi incelemeler yaparak bir ürünle ilgili deneyimlerini paylaşmaya giderek daha fazla katılmaktadır. Web içinde kullanıcı tarafından oluşturulan içeriğin ortaya çıkması ve etkisi, klasik WOM'u İnternet ve çevrimiçi iletişim tarafından etkinleştirilen (Electronic word of mouth) eWOM'a doğru hareket ettirmiştir. Güç yasası ile ürün kategorilerinin dağılımının modellenmesi tasarlanmıştır. Güç yasası dağılımına uyması için içerik değil, inceleme sayısı kullanılmıştır. Bu nedenle, kısa başlıkta yüksek frekanslı olaylar mı yoksa kuyrukta düşük fakat daha büyük olaylar mı olduğunu ayırt etmek için ürün kategorilerinin dağılımını güç yasası ile modellemek uygun olacaktır. Araştırmalar, çevrimiçi incelemelerin mesaj içeriğinin, kullanıcılar bir satın alma kararı vermek üzereyken çok sayıda incelemeye baktıklarında daha az etkiye sahip olduğunu göstermiştir. Veri toplama ve analiz için seçilen eWOM topluluğu, 7 milyondan fazla yorum yazan 1,3 milyondan fazla kullanıcının yaklaşık 1,4 milyon ürün hakkında yazdığı popüler bir web sitesi olan Ciao UK kullanılmıştır [29].

## II. MALZEME VE YÖNTEMLER

Bu çalışmada, makine öğrenmesi yöntemlerinden tekil değer ayrışımı (Singular value decomposition-SVD), SVD++ ve en yakın komşuluk (k nearest neighbor-kNN) algoritmaları kullanılarak, uzun kuyruklu ürünler kullanılarak, özellikle elektronik ticarete sonraki öğe öneri sistemi geliştirilerek iyileştirilmiştir.

Tekil değer ayrışımı algoritması (Singular value decomposition-SVD), ortonormal bir matris, köşegen bir matris ve ortonormal bir matris olmak üzere üçlü bir çarpıma ayrıştıran bir algoritmadır [30]. K en yakın komşuluk algoritması ise (k-nearest neighbour-kNN) çeşitli metrik tanımlarını kullanarak en yakın eleman değerine göre veri

noktasının sınıflandırma tahmininin yapıldığı eğitimci öğrenme algoritmalarından biridir.

Bu bölümde, deneysel çalışmalar için veri toplama, model oluşturma ve deneysel çalışmalar gerçekleştirilmiştir. Veri kümesini toplamak için JS ve Java programlama diliyle bir veri toplayıcı modül yazılmıştır. Kullanıcıların, ürün görüntüleme, tıklama, sepete atma ve satın alma verilerini kaydedilmiştir. Kullanıcı-ürün etkileşimleri (view, click, basket, order) dış değerlere (rating) çevrilmiştir (0-5 arası). Ön işleme işlemleri gerçekleştirilmiş ve anonimleştirilmiştir. Bu bağlamda, kullanıcı (user), ürün (item), değerlendirme (rating) bazında bir veri kümesi oluşturduk. Veri kümesinin büyüklüğü 1.1 milyon (1.1M) adettir. Ayrıca, bu veri kümesinden yola çıkarak 100.000 (100K) adetlik daha küçük boyutlu bir alt veri kümesi daha oluşturulmuştur.

Tekil değer ayrıştırma, tekil değer ayrıştırma++ ve en yakın komşuluk algoritması yöntemleriyle kök ortalama karesel hata (root mean square error-RMSE) ve hatanın mutlak ortalaması (mean absolute error-MAE) hata değerleri ölçülmüştür. Aşağıdaki tabloda (Tablo I) 100K adet veri kümesi için SVD, SVD++ ve kNN algoritmalarının sınıflandırma sonuçları verilmiştir. Bu sınıflandırmada kök ortalama karesel hata (root mean square error-RMSE) ve ortalama mutlak hata (mean absolute error-MAE) hata değerleri ölçülerek sonuçlar Tablo I'de paylaşılmıştır. En az hata değerine sahip (en başarılı) yöntemler, köyü renkli olarak gösterilmiştir. Buna göre en başarılı yöntemin SVD olduğu gözlenmiştir.

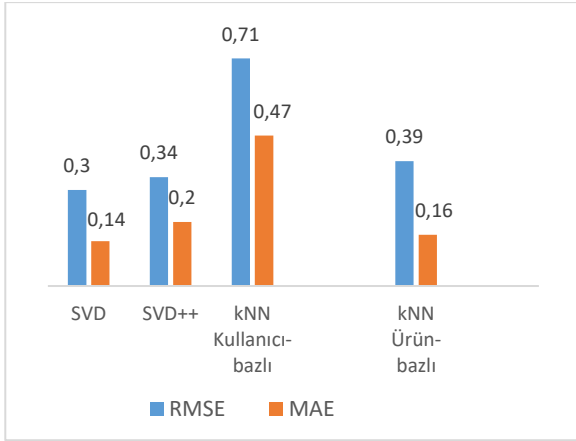
Tablo I: 100K boyutundaki veri kümesi için sınıflandırma sonuçları (Hata değerleri)

Yöntem (Algoritma)	RMSE	MAE
<b>SVD</b>	<b>0,30</b>	<b>0,14</b>
SVD++	0,34	0,20
kNN Kullanıcı-bazlı (User-based)	0,71	0,47
kNN Ürün-bazlı (Item-based)	0,39	0,16

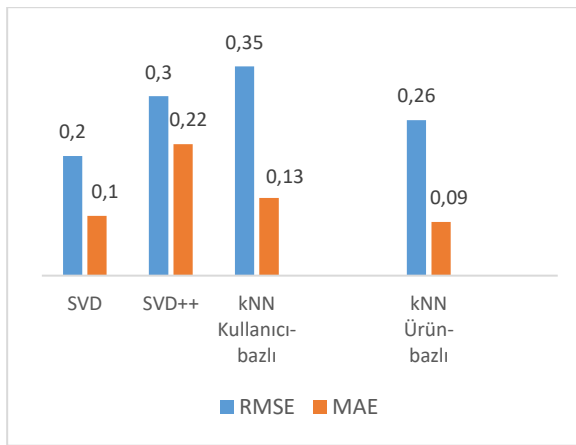
Öte yandan, 1.1M boyutlu veri kümesi için de aynı makine öğrenmesi yöntemleri kullanılmış ve sonuçlar Tablo II'de paylaşılmıştır. Buna göre, RMSE olarak en iyi yöntemin SVD, MAE olarak ise en iyi yöntemin, kNN Ürün-bazlı (Item-based) yönteminin olduğu belirlenmiştir.

Tablo II. 1.1M boyutundaki veri kümesi için sınıflandırma sonuçları (Hata değerleri)

Yöntem (Algoritma)	RMSE	MAE
<b>SVD</b>	<b>0,20</b>	0,10
SVD++	0,30	0,22
kNN Kullanıcı-bazlı (User-based)	0,35	0,13
kNN Ürün-bazlı (Item-based)	0,26	<b>0,09</b>



Şekil 1. 100K veri kümesi sınıflandırma sonuçlarının görsel halı



Şekil 2. 1.1M veri kümesi sınıflandırma sonuçlarının görsel halı

### III. SONUÇLAR

Ürün tavsiye sistemleri elektronik ticaret sitelerinin kalitesini gösteren ve müşterilere en iyi hizmeti sunmak için oluşturulmuş algoritmalarıdır. Bir tavsiye sistemi müşterilere en iyi önerileri yapabilmek için onları iyi tanımalıdır. Ayrıca ürünler hakkında bilgileri toplayarak en uygun tavsiyeyi vermeli böylece kullanıcıları ürünleri satın almaya ikna etmelidir. 1.1M veri seti üzerinde tekil değer ayrıştırma, tekil değer ayrıştırma++ ve en yakın komşuluk algoritması yöntemleri test edilmiştir. Kök ortalama karesel hata (root mean square error-RMSE) ve hatanın mutlak ortalaması (mean absolute error-MAE) hata değerleri ölçülmüş sonuçlar karşılaştırılmıştır. Ürün tavsiyesinde SVD yönteminin daha başarılı sonuçlar elde ettiği gözlenmiştir.

### BİLGİLENDİRME

Bu çalışma, 7190894 proje numarasıyla Tübitak tarafından desteklenmiştir. Tübitak'a vermiş olduğu destekten dolayı teşekkürlerimizi sunarız.

### Authors' Contributions

The authors' contributions to the paper are equal.

### Statement of Conflicts of Interest

There is no conflict of interest between the authors.

### Statement of Research and Publication Ethics

The authors declare that this study complies with Research and Publication Ethics.

### REFERANSLAR

- [1] Oestreicher-Singer, G., and Sundararajan, A. (2012). Recommendation networks and the long tail of electronic commerce. *Mis quarterly*, pp. 65-83.
- [2] Shi, L. (2013). Trading-off among accuracy, similarity, diversity, and long-tail: a graph-based recommendation approach. In *Proceedings of the ACM Conference on Recommender Systems* (pp. 57-64).
- [3] Yin, H., Cui, B., Li, J., Yao, J., and Chen, C. (2012). Challenging the long tail commendation. *arXiv preprint arXiv:1205.6700*.
- [4] Brynjolfsson, E., Hu, Y., and Simester, D. (2011). Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Science*, 57(8), 1373- 1386.
- [5] Alshammari, G., Jorro-Aragoneses, J. L., Polatidis, N., Kapetanakis, S., Pimenidis, E., and Petridis, M. (2019). A switching multi-level method for the long tail recommendation problem. *Journal of Intelligent and Fuzzy Systems*, 37(6), pp. 7189-7198.
- [6] Wang, S., Gong, M., Li, H., and Yang, J. (2016). Multi-objective optimization for long tail recommendation. *Knowledge-Based Systems*, 104, pp. 145-155.
- [7] Hu, X., Zhang, C., Wu, M., and Zeng, Y. (2017). Research on long tail recommendation algorithm. In *IOP Conference Series: Materials Science and Engineering*, vol. 261, no. 1.
- [8] Pandey, A. K., and Ankararkanni, B. (2020). Recommending e-commerce products on cold start and long tail using transaction data. *International Conference on Trends in Electronics and Informatics*, pp. 661-663.
- [9] Abdollahpouri, H., Burke, R., & Mobasher, B. (2018). Popularity-aware item weighting for long-tail recommendation. *arXiv preprint :1802.05382*.
- [10] Wang, Y., Wang, J., & Li, L. (2018). Enhancing Long Tail Recommendation Based on User's Experience Evolution. *Int. Conference on Computer Supported Cooperative Work in Design*, pp. 25-30.
- [11] Zhou, W., and Duan, W. (2012). Online user reviews, product variety, and the long tail: An empirical investigation on online software downloads. *Electronic Commerce Research and Applications*, 11 (3), pp. 275-289.
- [12] Luke, A., Johnson, J., and Ng, Y. K. (2018, November). Recommending long-tail items using extended tripartite graphs. *International Conference on Big Knowledge*, pp. 123-130.
- [13] Agarwal, P., Sreepada, R. S., and Patra, B. K. (2019). A hybrid framework for improving diversity and long tail items in recommendations. *International Conference on Pattern Recognition and Machine Intelligence*, pp. 285-293.
- [14] De Sousa Silva, D. V., De Oliveira, A. C., Almeida, F., and Durão, F. A. (2020). Exploiting Graph Similarities with Clustering to Improve Long Tail Items Recommendations. *Brazilian Symposium on Multimedia and the Web*, pp. 193-200.
- [15] Cai, Y., Cui, Z., Wu, S., Lei, Z., and Ma, X. (2021). Represent Items by Items: An Enhanced Representation of the Target Item for Recommendation. *ArXiv preprint arXiv:2104.12483*.
- [16] Achary, N. S., and Patra, B. K. (2021). Graph Based Hybrid Approach for Long-Tail Item Recommendation in Collaborative Filtering. *ACM IKDD CODS and COMAD* (pp. 426-426).
- [17] Chen, X., Pan, Y., and Luo, B. (2020). Research on power-law distribution of long-tail data and its application to tourism recommendation. *Industrial Management and Data Systems*.
- [18] Zencirli, A., Çetin, H., Tuğ, N., and Ensari, T. (2021). Deep Learning Classification of Location Oriented Recommendation System for Low-Sale Products. *International Congress on Human-Computer Interaction, Optimization and Robotic Applications*, pp. 1-4.
- [19] Meenakshi, M. (2019). A Novel Approach Web Services Based Long Tail Web Services Using Deep Neural Network. *Inte. Conference on Issues and Challenges in Intelligent Computing Techniques*, v1, pp 1-9.
- [20] Le, D. D., and Lauw, H. (2021). Efficient Retrieval of Matrix Factorization-Based Top-k Recommendations: A Survey of Recent Approaches. *Journal of Artificial Intelligence Research*, 70, pp. 1441-1479.
- [21] Cui, Z., Zhao, P., Hu, Z., Cai, X., Zhang, W., and Chen, J. (2021). An

- improved matrix factorization based model for many-objective optimization recommendation. *Information Sciences*, 579, pp. 1-14.
- [22] Gabbolini, G., D'Amico, E., Bernardis, C., and Cremonesi, P. (2021). On the instability of embeddings for recommender systems: the case of Matrix Factorization. In *Proceedings of the 36th Annual ACM Symposium on Applied Computing* (pp. 1363-1370).
- [23] Park, Y. J., and Tuzhilin, A. (2008). The long tail of recommender systems and how to leverage it. *ACM conference on Recommender systems*, pp. 11-18.
- [24] Sundaresan, N. (2011). Recommender systems at the long tail. *ACM conference on Recommender systems*, pp. 1-6.
- [25] Behera, R. K., Gunasekaran, A., Gupta, S., Kamboj, S., and Bala, P. K. (2020). Personalized digital marketing recommender engine. *Journal of Retailing and Consumer Services*, 53.
- [26] Pratama, B. Y., Budi, I., and Yuliawati, A. (2020). Product Recommendation in Offline Retail Industry by using Collaborative Filtering. *International Journal of Advanced Computer Science and Applications*, 11(9), pp. 635-643.
- [27] Valcarce, D., Parapar, J., and Barreiro, A. (2016). Item-based relevance modelling of recommendations for getting rid of long tail products. *Knowledge-Based Systems*, 103, pp. 41-51.
- [28] Zhang, F., Lu, Y., Chen, J., Liu, S., and Ling, Z. (2017). Robust collaborative filtering based on non-negative matrix factorization and R1-norm. *Knowledge-based systems*, 118, pp. 177-190.
- [29] Olmedilla, M., Martínez-Torres, M. R., and Toral, S. L. (2019). The superhit effect and long tail phenomenon in the context of electronic word of mouth. *Decision Support Systems*, 125, pp. 113-120.
- [30] L. Zhaoyang, *Matris Ayrışımı*, Yüksek Lisans tezi, 2006.