

Geliş Tarihi:
08.01.2022
Kabul Tarihi:
27.05.2022
Yayımlanma Tarihi:
15.06.2022

Kaynakça Gösterimi: Budak, H., & Gümüştas, E. (2022). Kişiselleştirilmiş ürün öneri sistemi için kullanıcı bazlı işbirlikçi filtreleme ve kümeleme kullanan hibrit bir yaklaşım. *İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi*, 21(43), 253-268. doi: 10.46928/iticusbe.1055162


KİŞİSELLEŞTİRİLMİŞ ÜRÜN ÖNERİ SİSTEMİ İÇİN KULLANICI BAZLI İŞBİRLİKÇİ FİLTRELEME VE KÜMELEME KULLANAN HİBRİT BİR YAKLAŞIM

Araştırma

Hüseyin Budak 

Sorumlu Yazar (Correspondence)

huseyin.budak@hotmail.com.tr

Enis Gümüştas 

Tanı Pazarlama Hizmetleri A.Ş.

enisgumustas@yandex.com

Dr. Hüseyin Budak, istatistik alanında doktora derecesine sahip bir veri bilimcidir. Serbest analitik danışman olarak çalışmaktadır. Başlıca araştırma alanları arasında; uygulamalı istatistik, veri madenciliği, yöneylem araştırması ve makine öğrenmesi bulunmaktadır.

Enis Gümüştas, Mimar Sinan Güzel Sanatlar Üniversitesi'nden İstatistik alanında Lisans ve Yüksek Lisans derecelerine sahiptir. Doktora eğitimine Mimar Sinan Güzel Sanatlar Üniversitesi İstatistik Bölümü'nde devam etmektedir. Halen Tanı Pazarlama'da veri bilimci olarak çalışmaktadır. Araştırma ilgi alanları arasında makine öğrenimi ve öneri sistemleri yer almaktadır.

KİŞİSELLEŞTİRİLMİŞ ÜRÜN ÖNERİ SİSTEMİ İÇİN KULLANICI BAZLI İŞBİRLİKÇİ FİLTRELEME VE KÜMELEME KULLANAN HİBRİT BİR YAKLAŞIM

Hüseyin Budak
huseyin.budak@hotmail.com.tr
Enis Gümüştas
enisgumustas@yandex.com

Özet

Günümüz rekabet koşulları firmaları, özellikle perakende ve e-ticaret firmalarını, müşterilerini daha iyi tanımaya, onların tercihlerini ve davranışlarını anlamaya, ihtiyaçlarını tahmin etmeye; böylelikle, onlara kendilerini özel hissettirecek, teklifler sunmaya zorlamaktadır. Firmaların söz konusu kişiselleştirme ihtiyaçlarını karşılayabilmeleri adına kullandıkları yöntemlerden biri ürün öneri sistemleridir.

Amaç: Çalışmada, kişiselleştirilmiş ürün öneri sistemleri için literatürde ve iş dünyasında sıklıkla kullanılan yöntemlerden biri olan kullanıcı bazlı işbirlikçi filtreleme yöntemini iyileştirmek adına k -means ile kullanıcı bazlı işbirlikçi filtreleme algoritmalarını birlikte kullanan hibrit bir yaklaşım önerilmesi amaçlanmıştır.

Yöntem: Kullanıcı bazlı işbirlikçi filtreleme ve k -means yöntemleri kullanılmıştır.

Bulgular: Mevcut yöntem ve önerilen yöntem iki farklı veri seti için uygulanmıştır. Yöntemlerin karşılaştırılması amacıyla veri setleri %80'i eğitim, %20'si de test seti olmak üzere ikiye ayrılmış ve eğitim verisi üzerinden kurulan modellerin test verisindeki hataları (RMSE) hesaplanmıştır. Yapılan karşılaştırma sonucunda her iki veri setinde de önerilen yöntemle ilişkin hata değeri daha az olduğu görülmüştür.

Özgünlük: Bu çalışma ile sadece kullanıcı-ürün skorları üzerinden çalışan kullanıcı bazlı işbirlikçi filtreleme yöntemine kullanıcılara ilişkin farklı bilgileri de kullanabilen bir yaklaşım önerilmiştir. Ayrıca, önerilen yöntem literatürde sıklıkla kullanılan MovieLens veri setinden uygulanmasının yanı sıra gerçek bir süpermarket verisinde de uygulanmıştır.

Anahtar Kelimeler: Öneri Sistemleri, Kümeleme, Kullanıcı Bazlı İşbirlikçi Filtreleme, K-Means

JEL Sınıflandırması: M31, C10, C80

A HYBRID APPROACH USING USER-BASED COLLABORATIVE FILTERING AND CLUSTERING FOR PERSONALIZED PRODUCT RECOMMENDATION SYSTEM

Abstract

Nowadays competition conditions are forced companies, especially retail and e-commerce companies, to know their customers better, to understand their preferences and behaviours, to predict their needs, in this way, to make offers that feel them special. One of the methods used by companies to meet their personalization needs is product recommendation systems.

Purpose: In the study, it is aimed to propose a hybrid approach that uses k -means and user-based collaborative filtering algorithms together to improve the user-based collaborative filtering method, which is one of the most frequently used methods in the literature and business world for personalized product recommendation systems.

Method: User-based collaborative filtering and k -means methods are used.

Findings: The current method and the proposed method were applied for two different data sets. In order to compare the methods, the data sets were divided into two as 80% training and 20% test set, and the errors (RMSE) of the models built on the training data were calculated. As a result of the comparison, it was seen that the error value for the proposed method was less in both data sets.

Originality: In this study, an approach that can also use different information about users is proposed to the user-based collaborative filtering method, which works only on user-product scores. In addition, the proposed method has been applied to a real supermarket data as well as being applied from the MovieLens dataset, which is frequently used in the literature.

Keywords: Recommendation Systems, Clustering, User-Based Collaborative Filtering, K-Means

JEL Classification: M31, C10, C80

GİRİŞ

Her geçen gün daha çok küreselleşen günümüz dünyasında, teknolojik yeniliklerle elde edilen farklılaşmalar anlamlı ve kalıcı olmadığından, firmaların müşterilerine benzersiz bir satış önerisiyle gidebilmesi neredeyse imkânsız hale gelmektedir (Demir ve Kırdar, 2007). Bu durum, aynı sektörde hizmet veren birkaç firmanın ürün yelpazesini ve fiyatlarını karşılaştırdığımızda rahatlıkla görülmektedir. Üretim tarafında giderek aynılaşan firmalar, günümüz rekabet koşullarında ayakta kalabilmek için, pazarlama tarafında farklılaşmanın yollarını aramaktadır. Özellikle mevcut müşteriler için bu farklılaşmayı sağlayacak en önemli araçlardan biri müşteri ilişkileri yönetimidir.

Müşteri ilişkileri yönetimi ile firmalar doğru ürünle, doğru fiyatla, doğru zamanda ve doğru kanaldan müşterilerin karşısına çıkmayı ve bu sayede rakiplerinden farklılaşarak tercih edilmeyi hedeflemektedir. Elbette söz konusu doğrular her kişi için farklı olabilmektedir. Firmalar bu farklılıkları anlamaya ve böylelikle müşterileri için kişiselleştirilmiş hizmetler vermeye çalışmaktadır. Bu farklılıkları anlamamanın yolu da veriden geçmektedir. Veriyi merkezine almış analitik müşteri ilişkileri yönetimi ile firmaların kişiselleştirilmiş hizmet vermesinin yollarından biri müşterilerine ürün önerileri sunmaktır. Ancak, her ürün öneri yaklaşımı ile kişiselleştirilmiş öneriler üretilmemektedir.

Kişiselleştirme bakış açısıyla ürün önerilerini üçe ayrılabilir: kişiselleştirilmemiş, yarı kişiselleştirilmiş ve kişiselleştirilmiş öneriler. Kişiselleştirilmemiş önerilerde sıklıkla kullanılan yöntemlerden biri kullanıcılara en popüler veya en çok satan ürünleri önermektir. Yarı kişiselleştirilmiş yaklaşımda ise genel eğilime göre değil kullanıcının bulunduğu segmentteki eğilime göre öneri yapılmaktadır. Örneğin, yaş, cinsiyet gibi müşteriyi tanımlayan demografik bilgiler ya da müşterinin satın alma sıklığı, ortalama satın alma tutarı gibi davranışsal bilgileri kullanılarak oluşturulan herhangi bir segmentteki en popüler ürünün o segmentteki tüm kullanıcılara önerilmesidir. Son yaklaşım olan kişiselleştirilmiş öneriler ise kullanıcıların geçmiş alışveriş alışkanlıkları kullanılarak oluşturulan modeller/kurallar ile her kullanıcıya beğenme ihtimali yüksek olan ürünlerin önerilmesidir (Falk, 2019).

Kişiselleştirilmiş önerileri yapabilmek için genellikle ürün öneri sistemleri kullanılmaktadır. Öneri sistemleri, diğer adıyla tavsiye sistemleri, kullanıcılara ait geçmiş verileri kullanarak onların bir sonraki tercihlerini tahmin eden ve bu tahminlere bağlı olarak öneriler sunan sistemlerdir. Kullanıcıların satın alacakları ürün, izleyecekleri film veya dinleyecekleri müzik gibi konularda karar verme sürecini desteklemeyi amaçlamaktadır.

Öneri sistemleri içinde sıklıkla kullanılan yöntemlerden birisi işbirlikçi filtrelemedir. Her geçen gün kullanım alanı genişleyen işbirlikçi filtreleme, günümüzde, e-ticaret (Shen ve diğerleri, 2020; Gupta ve diğerleri, 2021; Kumar ve diğerleri, 2020), perakende sektörü (Castelo-Branco ve diğerleri, 2021; Pratama ve diğerleri, 2020; Rebelo ve diğerleri, 2021), film önerisi (Awan ve diğerleri, 2021; Anwar ve Uma, 2021; Gupta ve diğerleri, 2020), müzik önerisi (Kathavate, 2021; Jitendra ve Radhika, 2021),

sosyal medya (Munuswamy ve diğeri, 2021; Chung ve diğeri, 2021), haber önerisi (Gao ve diğeri, 2020; Han, 2020) ve eğitim sektörü (Murad ve diğeri, 2020; Zhou ve diğeri, 2020) ağırlıklı olmak üzere pek çok alanda kullanılmaktadır.

Çalışmada kişiselleştirilmiş ürün öneri sistemleri için literatürde ve iş dünyasında sıklıkla kullanılan yöntemlerden biri olan kullanıcı bazlı işbirlikçi filtreleme yöntemini iyileştirmek adına k -means ile kullanıcı bazlı işbirlikçi filtreleme algoritmalarını birlikte kullanan hibrit bir yaklaşım önerilmiştir. K -means kullanılarak kullanıcılar değer segmentlerine göre farklı alt kümelere ayrılmıştır. Alt kümelerin her birine kullanıcı bazlı işbirlikçi filtreleme uygulanmıştır. Böylece kullanıcılara buldukları değer segmentine uygun ve daha isabetli öneriler yapılması amaçlanmıştır.

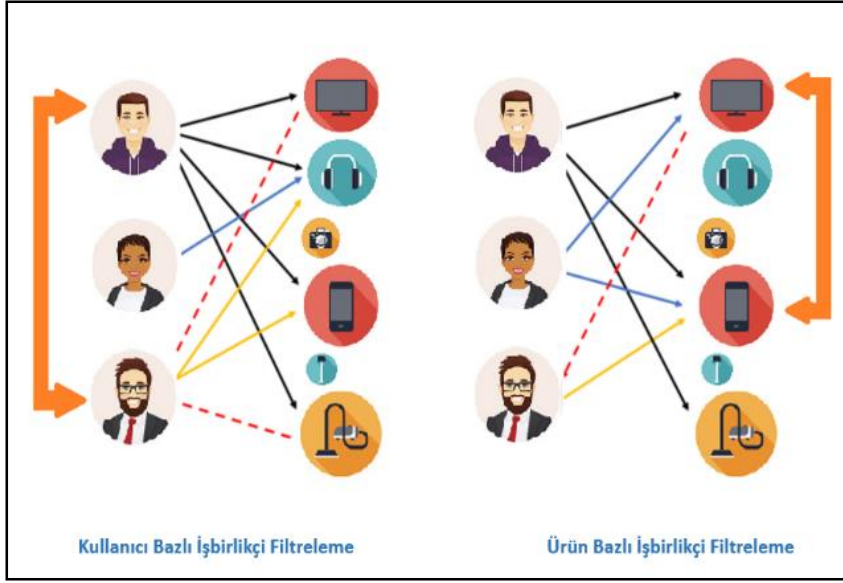
YÖNTEM

İşbirlikçi Filtreleme

İşbirlikçi filtreleme, kullanıcıların mevcut ürünlere verdikleri oyları/beğenileri veya her bir kullanıcı-ürün ilişkisi üzerinden oluşturulan skorları analiz ederek kullanıcıların daha önce deneyimlemedikleri ürünlerdeki beğeni/davranış skorlarını tahmin etmeyi ve bu tahminler üzerinden kullanıcılara önerilerde bulunmayı amaçlamaktadır.

İşbirlikçi filtreleme yöntemleri, model tabanlı (model based) ve bellek tabanlı (memory based) olarak iki ana yaklaşıma ayrılmaktadır. Model tabanlı işbirlikçi filtrelemede kümeleme, regresyon, sınıflandırma, matris faktörizasyonu gibi yöntemler kullanırken bellek tabanlı yöntemler benzerlik ilişkilerini kullanmaktadır (Do ve diğeri, 2010; Su ve Khoshgoftaar, 2009). Model tabanlı yöntemler, geçmiş verilerden kullanıcı-ürün eğilimine ilişkin bir örüntü yakalamaya ve kurulan model ile ürünlere ilişkin kullanıcı eğilimlerini veya beğeni skorlarını tahmin etmeye çalışmaktadır. Bellek tabanlı yöntemler ise komşuluk ilişkisini kullanılarak öncelikle kullanıcılara ya da ürünlere benzer elemanları tespit etmekte ve sonrasında bu elemanlar üzerinden ürünlere ilişkin kullanıcı eğilimlerini veya beğeni skorlarını tahmin etmeye çalışmaktadır.

Bellek tabanlı işbirlikçi filtreleme yöntemleri temel olarak kullanıcı bazlı (User Based Collaborative Filtering - UBCF) ve ürün bazlı (Item Based Collaborative Filtering – IBCF) olmak üzere ikiye ayrılmaktadır (Su ve Khoshgoftaar, 2009). Her iki yöntemde çalışma adımları çok benzerdir. İki yöntem arasındaki temel fark, biri kullanıcı-kullanıcı benzerlikleri üzerinden bir komşuluk oluşturup buna göre tahmin üretirken diğeri ise ürün-ürün benzerliklerini kullanarak tahmin üretmektedir. İki yöntem arasındaki farkı daha anlaşılır kılan karşılaştırma görseli Şekil 1.'deki gibidir. İlerleyen bölümlerde bu çalışmada kullanılan Kullanıcı Bazlı İşbirlikçi Filtreleme yönteminin detayına yer verilecektir.



Şekil 1. Kullanıcı bazlı ve ürün bazlı işbirlikçi filtreleme yöntemleri

Kullanıcı Bazlı İşbirlikçi Filtreleme

İşbirlikçi filtrelemeye dayanan bu yöntem aktif kullanıcıya benzer niteliklere sahip diğer kullanıcıların geçmişte beğenmiş oldukları ürünlerin öneri olarak sunulması temeline dayanmaktadır. Aktif kullanıcının henüz satın almadığı ya da beğenmediği her bir ürün için, benzerlik gösterdiği kullanıcıların yapmış oldukları değerlendirmeler veya işlemler üzerinden bir tahmin değeri hesaplanmaktadır. Tahmin edilen bu değer aktif kullanıcının ilgili ürüne ilişkin beğeni veya davranış skorunu temsil etmektedir. Yöntem, tahmin edilen bu skorlar üzerinden aktif kullanıcıya en uygun ürün veya ürünleri önermeye çalışmaktadır.

Kullanıcı Bazlı İşbirlikçi Filtreleme Adımları

Kullanıcı bazlı işbirlikçi filtreleme algoritması benzerlik hesaplaması, komşuluk seçimi ve tahmin olmak üzere üç temel adımda çalışmaktadır. Yöntem özetle, ilk olarak aktif kullanıcı ile ilgili ürünleri değerlendirmiş diğer kullanıcılar arasında benzerlik hesaplaması yapmaktadır. Sonrasında, hesaplanan benzerlik değerlerine göre aktif kullanıcıya en yakın komşuları seçmektedir. Son olarak, seçilen komşular ve hesaplanan benzerlik değerleri (ağırlık vektörleri) üzerinden aktif kullanıcının ilgili ürünlere ilişkin skorlarını tahmin etmektedir.

İşbirlikçi filtreleme yönteminin ilk adımı benzerlik hesaplamasıdır. Bu aşamada aktif kullanıcı ile diğer kullanıcılar arasında benzerliği temsil eden ağırlıklar hesaplanır. Benzerlik hesaplaması için birçok yöntem bulunmaktadır. Bu yöntemler arasında işbirlikçi filtreleme için en sık kullanılanı Pearson Korelasyon katsayısıdır.

Denklem (1)'de formülü yer alan Pearson korelasyon katsayısı ile iki kullanıcı arasındaki benzerlik değeri hesaplanmaktadır. Pearson korelasyon katsayısı ise -1 ile 1 arasında değer almaktadır. Bu değer 1'e yaklaşması kullanıcılar arasındaki benzerliğin arttığını göstermektedir (Bulut ve Milli, 2016; Hahsler, 2015).

$$sim(x, y) = \frac{\sum_{i \in I} (R_{x,i} - \bar{R}_x) \cdot (R_{y,i} - \bar{R}_y)}{\sqrt{\sum_{i \in I} (R_{x,i} - \bar{R}_x)^2} \cdot \sqrt{\sum_{i \in I} (R_{y,i} - \bar{R}_y)^2}} \quad (1)$$

Denklem (1)'de yer alan x aktif kullanıcıyı, y ise benzerliğine bakılan diğer kullanıcıyı temsil etmektedir. I ise aktif kullanıcı ile diğer kullanıcının değerlendirdiği ortak ürünlerin kümesidir. $R_{x,i}$ aktif kullanıcının i ürününe ait değerlendirmesini, $R_{y,i}$ de diğer kullanıcının i ürününe ait değerlendirmesini göstermektedir. \bar{R}_x x kullanıcısının ürünlere verdiği puanların ortalamasını, \bar{R}_y ise y kullanıcısının ürünlere verdiği puanların ortalamasını göstermektedir (Bulut ve Milli, 2016).

İşbirlikçi filtrelemenin ikinci adımı olan komşuluk seçiminde bir önceki adımda hesaplanan benzerlik değerleri kullanılarak aktif kullanıcıyla benzer örneği gösteren diğer kullanıcılar seçilmektedir. Benzer kullanıcıların seçiminde en sık kullanılan yöntem ise k en yakın komşu algoritmasıdır. Bu yöntem ile aktif kullanıcıya en yüksek benzerliğe sahip olan k adet kullanıcı komşu olarak seçilmektedir.

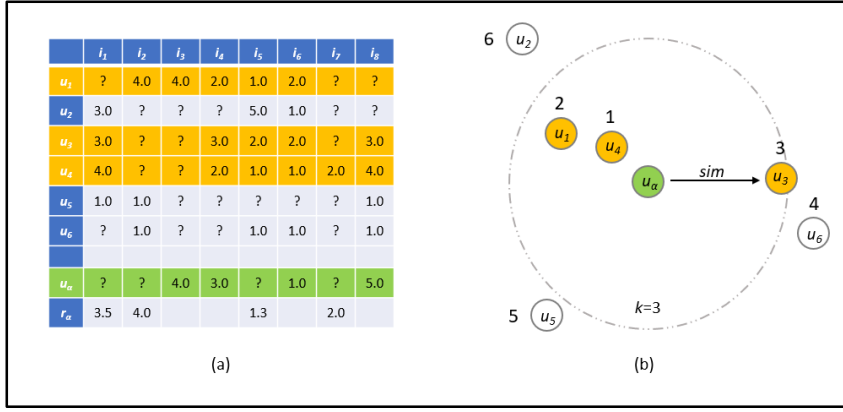
İşbirlikçi filtrelemenin son adımı olan tahmin hesaplamasında ise bir önceki adımda seçilen komşular üzerinden aktif kullanıcının daha önce değerlendirmedeği veya satın almadığı ürünler için davranış skorları hesaplanmaktadır. Tahmin işlemi komşu kullanıcıların ilgili ürüne ilişkin skorlarının ağırlıklı ortalaması alınarak yapılmaktadır. Ağırlıklandırma işlemi ise komşuluk seçiminde kullanılan benzerlik değerleri üzerinden yapılmaktadır. Ayrıca, tahmin işlemi yapılırken kullanıcıların değerlendirme skalasını nasıl algıladıklarını veya genel alışveriş davranışlarını yansıtabilmek adına hesaplama formülüne kullanıcıların genel ortalamaları da eklenmektedir. Tahmin hesaplamasında kullanılan formül denklem (2)'deki gibidir (Bulut ve Milli, 2016; Ekstrand ve diğerleri, 2011).

$$R_{x,i} = \bar{R}_x + \frac{\sum_{y \in U} (R_{y,i} - \bar{R}_y) \cdot sim(x,y)}{\sum_{y \in U} sim(x,y)} \quad (2)$$

Denklem (2)'de yer alan x aktif kullanıcıyı, y ise komşu kullanıcıları temsil etmektedir. U aktif kullanıcının komşularının bulunduğu kümedir. $sim(x,y)$ x kullanıcısı ile y kullanıcısı arasındaki benzerlik değeridir. $R_{x,i}$ aktif kullanıcının i ürününe ait tahmin değerini, $R_{y,i}$ de diğer kullanıcının i ürününe ait değerlendirmesini göstermektedir. \bar{R}_x x kullanıcısının ürünlere verdiği puanların ortalamasını \bar{R}_y ise y kullanıcısının ürünlere verdiği puanların ortalamasını göstermektedir (Bulut ve Milli, 2016).

Tahmin hesaplamasında kullanılan kullanıcı-ürün-skor (*rating matrix*) matrisi, tahmin işlemi ve komşuluk seçimi için örnek görsel Şekil 2.'deki gibidir. Şekil 2.'de yer alan $u\alpha$ aktif kullanıcıyı ve $r\alpha$ aktif kullanıcının daha önce değerlendirmedeği ürünler için beğeni skoru tahminini göstermektedir.

Örnek tahmin işleminde daha kolay anlaşılabilmesi adına aritmetik ortalama kullanılmıştır.



Şekil 2. Kullanıcı-ürün-skor matrisi ve tahmin hesaplaması (a), komşuluk seçimi (b) (Hahsler, 2015)

Kümeleme

Kümeleme, belirli benzerlik kriterleri kullanılarak bir veri setindeki üyelerin birbirlerine olan benzerliklerinin bulunması ve benzer olan üyelerin yan yana getirilmesi ile veri setinin kümeler bölünmesi işlemi olarak tanımlanmaktadır. Kümeleme işleminde küme içindeki üyelerin benzerliğinin fazla ve kümeler arası benzerliğin ise az olması amaçlanmaktadır.

Kümeleme analizinin, ses ve resim tanıma, müşteri ve pazar segmentasyonu, şehir planlama ve coğrafi bilişim sistemleri, örüntü tanıma, web uygulamaları ve arama motorları, DNA analizi gibi farklı birçok uygulama alanı bulunmaktadır.

Kümeleme, dağınık bir halde bulunan verileri benzerliklerine göre sınıflandırarak daha kolay işlenebilir hale getirebilmesi nedeniyle sıklıkla farklı veri madenciliği yöntemlerinin uygulanmasında bir ön analiz olarak da kullanılmaktadır.

Literatürde farklı birçok küme yöntemi bulunmakta olup, ilerleyen bölümde sadece bu çalışmada kullanılan k -means algoritmasının detayına yer verilecektir.

K-Means Algoritması

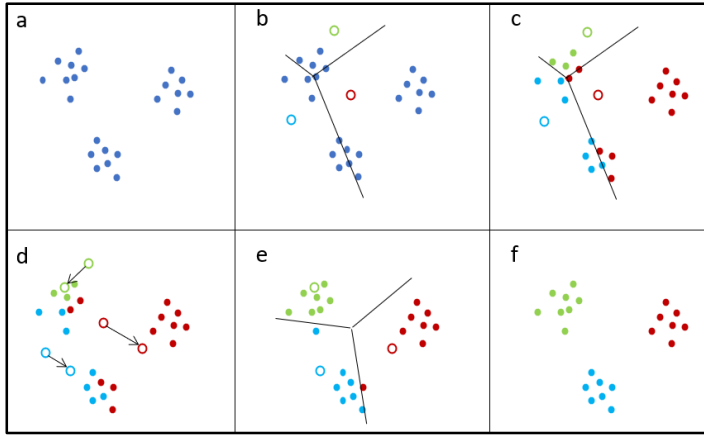
Yaygın olarak kullanılan kümeleme algoritmalarından biri olan k -means, n adet noktadan oluşan bir veri setini kullanıcı tarafından önceden belirlenmiş k adet kümeye ($1 < k < n$) bölmeye çalışmaktadır. Bu bölme işlemi temel olarak, küme merkezlerinin kümeyi temsil ettiği varsayımıyla, veri noktalarının küme merkezlerine olan uzaklıklarına bakarak yapmaktadır. Bölme işlemi sonrasında küme içi benzerliklerinin maksimum ve kümeler arası benzerliklerinin minimum olduğu kümeler elde etmeyi amaçlamaktadır. Benzerlik kriteri olarak Öklid (*Euclid*) uzaklığı veya kareli Öklid uzaklığı kullanılmaktadır (Akpınar, 2014; Sarıman, 2011).

k -means algoritmasının genel işlem adımları aşağıdaki gibidir (Akpınar, 2014; Silahtaroglu, 2013).

1. Rastgele olarak k adet küme merkezi belirlenir.
2. Tüm verilerin belirlenen küme merkezlerine olan uzaklıkları hesaplanır.
3. Her bir veri merkezine en yakın olduğu kümeye atanır.

4. Her kümede yer alan verilerin özellik değerlerine göre aritmetik ortalaması alınarak yeni küme merkezleri hesaplanır.
5. Tüm verilerin yeni küme merkezlerine olan uzaklıklarına göre daha yakın bir kümeye taşınıp taşınmayacağına karar verilir.
6. Küme elemanlarında herhangi bir değişiklik yoksa durulur aksi takdirde 2. adımdan itibaren iterasyona devam edilir.

Yukarıdaki işlem adımlarının, iki boyutlu örnek bir veri seti ve $k = 3$ için, şematik olarak gösterimi de Şekil 3.'deki gibidir.



Şekil 3. İki boyutlu veri seti için k -means algoritmasının şematik bir gösterimi ($k = 3$) (Chen ve Lai, 2016)

Önerilen Yöntem

Ürün öneri sistemleri içinde sıklıkla kullanılan ve başarılı sonuçlar veren yöntemlerden biri olan işbirlikçi filtreleme yönteminin iyileştirme ihtiyacı olan bazı yönleri de bulunmaktadır. Bu yönlerden en temeli, işbirlikçi filtreleme yönteminin kullanıcı-ürün-skor matrisi üzerinden tahmin ve öneri işlemlerini yapması ve kullanıcıların ürünlere verdikleri skorların dışında herhangi bir bilgi kullanmamasıdır. Söz konusu alanda iyileştirme sağlamak ve daha başarılı tahminler yapabilmek adına kullanıcı bazlı işbirlikçi filtreleme yöntemi ile kümeleme yöntemini bir arada kullanıldığı hibrit bir yaklaşım önerilmiştir. Önerilen yöntemde, özetle, tüm kullanıcılar için tek işbirlikçi filtreleme modeli kullanmak yerine öncelikle kullanıcıları kendi içlerinde homojen kümelere ayırıp sonrasında her bir küme için farklı işbirlikçi filtreleme modelleri geliştirilmiştir. Şekil 4.'de işlem akışına ilişkin detaylara yer verilen önerilen yöntem ile temel olarak aşağıdaki faydaların elde edilmesi amaçlanmıştır.

Kullanıcı bazlı işbirlikçi filtreleme yöntemi tahmin işleminde sadece kişilerin ürünlere verdikleri skor veya 0 ve 1 şeklinde satın alma bilgisini kullanmaktadır. Önerilen yöntem ile kümeleme ve işbirlikçi filtreleme iç içe uygulandığında kümelemede kullanılacak (farklı işlem bilgileri, demografi, konum özellikleri gibi) farklı değişkenlerden elde edilecek bilgilerden de ürün öneri modeli faydalanmış olacaktır. Bu sayede daha başarılı tahminler yapılması hedeflenmektedir.

Kullanıcı bazlı işbirlikçi filtreleme yönteminde tahmin başarısı etkileyen en önemli faktör ilgili kullanıcıya benzer diğer kullanıcıların (komşuların) seçilmesi işlemidir. Çok farklı tipte kullanıcıların bulunduğu veri setlerinde tüm veriye işbirlikçi filtreleme modelini uygulamak yerine daha homojen alt gruplar oluşturup bu gruplara işbirlikçi filtreleme modelini uygulamak nispeten daha benzer komşular bulmayı sağlayabilecek ve böylelikle tahmin başarısını arttırabilecektir.

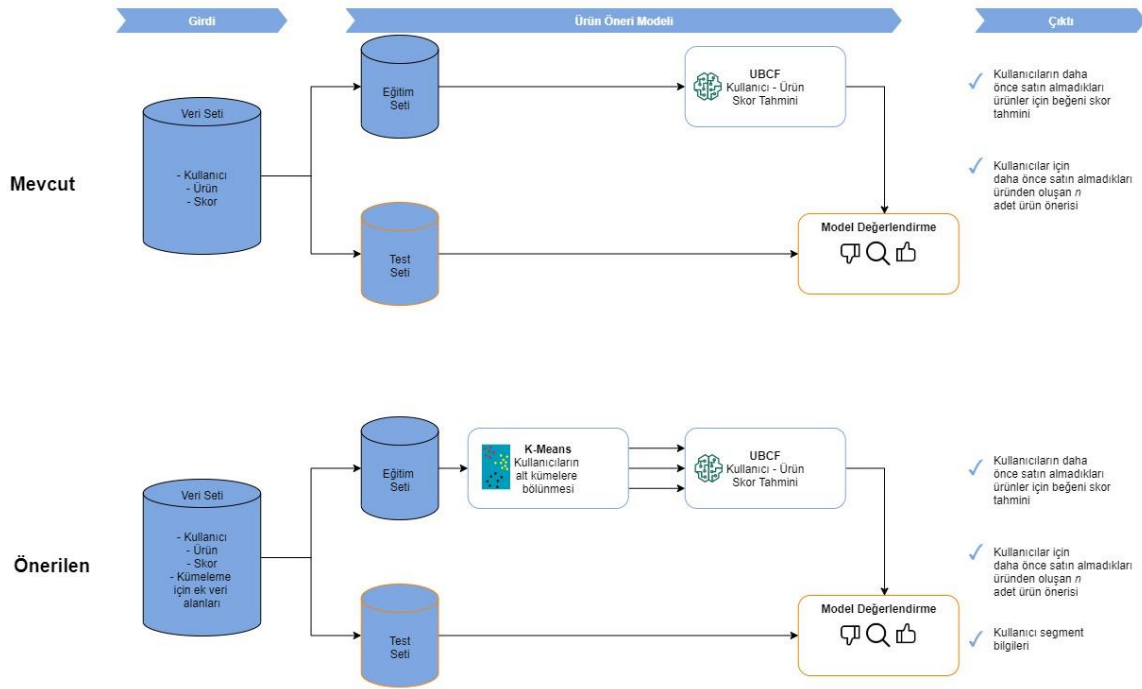
Uygulama Adımları ve Değerlendirme Kriteri

Mevcut ve önerilen ürün öneri yöntemlerinin karşılaştırılması için bir süpermarket müşterilerinin alışveriş işlemleri üzerinden oluşturulmuş örnek bir veri seti kullanılmıştır. Veri setinde temel olarak kullanıcıların toplam ve ürün bazlı alışveriş bilgileri bulunmaktadır. 12,284 kullanıcı ve 128 ürün bulunan veri setinde her kullanıcının en az 5 farklı üründe işlemi bulunmaktadır. Gerek perakende sektörünün ürün öneri sistemlerinin sahada en çok tercih edilen sektörlerinden biri olması gerekse çalışmayı gerçek veri ile yapma motivasyonumuz nedenleriyle uygulama için ilgili veri seti tercih edilmiştir. Bununla birlikte, yapılan akademik çalışmanın literatürce kabul edilebilir ve sunulan bulguların ispatlanabilir olması açısından uygulama literatürde sıklıkla kullanılan MovieLens 1 milyon (ML-1M) veri seti (Harper ve Konstan, 2015) için de tekrarlanmış ve elde edilen sonuçlar bulgular kısmında ayrıca paylaşılmıştır.

Uygulama aşamasındaki veri hazırlığı, kümeleme, model sonuçlarının birleştirilmesi ve değerlendirilmesi gibi işlemler için KNIME Analitik Platformu ve işbirlikçi filtreleme yönteminin uygulanması için R programlama dili kullanılmıştır. Kullanıcı bazlı işbirlikçi filtreleme yönteminin uygulanmasında R programlama diline ait recommenderlab paketi tercih edilmiştir.

Uygulama aşamasında öncelikle işbirlikçi filtreleme yönteminde kullanılan kullanıcı-ürün-skor matrisi için gerekli olan skorlar oluşturulmuştur. Söz konusu skor için kullanıcıların ürünlere ilişkin beğeni/tercih skorunu temsil eden RFM (*Recency, Frequency, Monetary*) skoru kullanılmıştır. Her kullanıcı için ürün bazlı oluşturulan RFM skorunun hesaplanmasında öncelikle kullanıcıların ilgili ürünlerdeki alışveriş yakınlığı (*recency*), işlem sıklığı (*frequency*) ve işlem tutarı (*monetary*) değerleri 1 ile 9 arasında derecelendirilmiştir. Sonrasında yakınlık için 100, sıklık için 10 ve tutar için 1 ağırlıkları kullanılarak ilgili RFM skorları elde edilmiştir.

Her kullanıcının geçmişte satın aldıkları ürünleri için skorlar hesaplandıktan sonra uygulamaya geçilmiştir. Şekil 4.'de detaylarına yer verilen yaklaşım ile gerçekleştirilen uygulamanın adımları temel olarak aşağıdaki gibidir.



Şekil 4. Mevcut yöntem ve önerilen yöntemin karşılaştırılmalı gösterimi

Mevcut yöntemde ilk olarak veri seti eğitim ve test seti olmak üzere ikiye ayrılır. Eğitim seti kullanılarak geliştirilen kullanıcı bazlı işbirlikçi filtreleme modeli ile kullanıcıların diğer ürünlerdeki skorları tahmin edilir. Son olarak ise test setindeki gerçek kullanıcı-ürün skorları ile tahmin skorlar arasındaki hata hesaplanır.

Önerilen yöntemde ise mevcut yöntemde olduğu gibi ilk olarak veri seti eğitim ve test seti olmak üzere ikiye ayrılır. Eğitim setindeki diğer kullanıcı bazlı değişkenler üzerinden *k*-means algoritması kullanılarak kullanıcılar segmente edilir ve alt kümeler bölünür. Oluşturulan alt kümeler kullanılarak farklı kullanıcı bazlı işbirlikçi filtreleme modelleri ile kullanıcıların diğer ürünlerdeki skorları tahmin edilir. Son adımda ise test setindeki gerçek kullanıcı-ürün skorları ile tahmin skorlar arasındaki hata hesaplanır.

Değerlendirme kriteri olarak kullanıcıların gerçek ürün skoru ile tahmin skoru arasındaki hata kullanılmıştır. Söz konusu hata ölçümü için denklem (3)'de verilen RMSE (*Root Mean Square Error* – Hata Kareleri Ortalamasının Karekökü) metriği tercih edilmiştir.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (R_{u,i} - R'_{u,i})^2} \quad (3)$$

Denklem (3)'de yer alan *n* test verisindeki örnek sayısını, $R_{u,i}$ *u* kullanıcısının *i* ürünün gerçek skoru ve $R'_{u,i}$ *u* kullanıcısının *i* ürünün tahmini skorunu göstermektedir (Su ve Khoshgoftaar, 2009; Bulut ve Milli, 2016).

BULGULAR

Uygulama için veri seti %80'i eğitim %20'si de test seti olmak üzere ikiye ayrılmıştır. Önerilen yöntemin uygulanabilmesi için öncelikle eğitimi verisi üzerinden k -means algoritması ile kümeleme analizi yapılarak kullanıcılar alt kümelere ayrılmıştır. Değer segmentasyonu bakış açısıyla kullanıcıların toplam işlem ve ciro değerleri kullanılarak yapılan kümeleme analizi sonucunda 4 alt küme elde edilmiştir. Kümeleme analizi sonrasında, her küme için geliştirilen farklı kullanıcı bazlı işbirlikçi filtreleme modeli ile kullanıcıların diğer ürünlerdeki skorları tahmin edilmiş ve bu tahminler üzerinden önerilen yönteme ilişkin hata değeri hesaplanmıştır ve sonuçlar Tablo 1'de verilmiştir.

Tablo 1. Önerilen ve mevcut yöntemin karşılaştırılması (süpermarket).

Yöntem	RMSE
Mevcut Yöntem	188.45
Önerilen Yöntem	183.86

Uygulamanın ikinci aşamasında yukarıda bahsedilen süreç MovieLens-1M veri seti kullanılarak tekrarlanmıştır. MovieLens-1M veri seti, kullanıcıların izledikleri filmler için verdikleri 1 ile 5 arasında değere sahip olan oyları içermektedir. Yaklaşık 1 milyon adet oy bulunan veri setinde 6040 tekil kullanıcı ve 3706 farklı film yer almaktadır. Veri seti en az 20 oy kullanmış olan kullanıcılardan oluşmaktadır.

Uygulama aşamasında, öncelikle kümeleme için kullanıcıların geçmiş oyları incelenmiş ve kümelemede kullanılmak üzere her bir kullanıcıya ait değişkenler hesaplanmıştır. Kullanıcıların daha önce izledikleri filmlerin film türlerine göre izleme sayıları (aksiyon, macera gibi 18 tür), filmlere verdikleri oy ortalamaları, toplam oy sayısı gibi değişkenler kullanılarak kümeleme yapılmış ve kullanıcılar 4 kümeye ayrılmıştır.

Kümeleme analizi sonrasında, her küme için geliştirilen farklı kullanıcı bazlı işbirlikçi filtreleme modeli ile kullanıcıların diğer filmlere verecekleri oylar tahmin edilmiş ve bu tahminler üzerinden önerilen yönteme ilişkin hata değeri hesaplanmıştır ve sonuçlar Tablo 2'de verilmiştir.

Tablo 2. Önerilen ve mevcut yöntemin karşılaştırılması (ML-1M).

Yöntem	RMSE
Mevcut Yöntem	1.27
Önerilen Yöntem	1.03

SONUÇ

Bu çalışmada, kişiselleştirilmiş ürün öneri sistemleri için literatürde ve iş dünyasında sıklıkla kullanılan yöntemlerden biri olan kullanıcı bazlı işbirlikçi filtreleme yöntemi incelenmiş ve yöntemi iyileştirmek adına k -means ile kullanıcı bazlı işbirlikçi filtreleme algoritmalarını birlikte kullanan hibrit bir yaklaşım önerilmiştir. Önerilen ve mevcut yöntemlerin uygulanması için bir süpermarket müşterilerinin alışverişler verileri ve MovieLens-1M veri seti kullanılmıştır. Uygulama sonucunda her iki yöntem ile de ilgili kullanıcılar için önerilen ürün/film skorları tahmin edilmiştir. Yöntemlerin karşılaştırılması için değerlendirme kriteri olarak kullanıcıların gerçek ürün/film skoru ile tahmin skoru arasındaki hatayı ölçen RMSE metriği kullanılmıştır. Yapılan karşılaştırma sonucunda her iki veri setinde de önerilen yönteme ilişkin hatanın daha az olduğu görülmüştür. Tahmin hatasındaki azalma incelendiğinde, süpermarket veri setindeki %2.7 azalma olduğu, MovieLens-1M veri setinde ise %18.9 azalma olduğu görülmüştür.

Sonuç olarak, önerilen yaklaşımın; gerek kullanıcılara ilişkin farklı bilgiler içeren değişkenleri kullanabilme kabiliyeti gerekse ürün skor tahminlerindeki hatayı azaltabileceği yönündeki bulgular dikkate alındığında ürün öneri sisteminin modelleme aşamasındaki alternatif yöntemler arasında kullanılabilmesi söylenebilir.

Sonraki çalışmalarda, farklı veri setleri ile yapılacak uygulamalar ile tahmin başarısının istatistiksel olarak anlamlı olup olmadığı test edilebilir. Ayrıca, önerilen yöntemin daha az hatalı tahminler üretebilmesinin arkasında uygulanan kümeleme yönteminin başarısı yatmaktadır. Bu nedenle, sonraki çalışmalarda, farklı kümeleme yöntemlerinin uygulanması ile kümeleme işlemindeki parametrelerin ürün öneri tahmin başarısına olan etkisi incelenebilir.

KAYNAKÇA

- Akpınar, H. (2014). *Data veri madenciliği veri analizi*. İstanbul: Papatya Yayıncılık Eğitim.
- Anwar, T., & Uma, V. (2021). Comparative study of recommender system approaches and movie recommendation using collaborative filtering. *International Journal of System Assurance Engineering and Management*, 12(3), 426-436.
- Awan, M. J., Khan, R. A., Nobanee, H., Yasin, A., Anwar, S. M., Naseem, U., & Singh, V. P. (2021). A Recommendation engine for predicting movie ratings using a big data approach. *Electronics*, 10(10), 1215.
- Bulut, H., & Milli, M. (2016). İşbirlikçi filtreleme için yeni tahminleme yöntemleri. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 22(2), 123-128.
- Castelo-Branco, F., Reis, J. L., Vieira, J. C., & dos Santos, J. P. M. (2021). Business intelligence in sports retail: data mining application, *16th Iberian Conference on Information Systems and Technologies (CISTI)*, 23-26 June 2021.
- Chen, Y. Z., & Lai, Y. C. (2016). Universal structural estimator and dynamics approximator for complex networks. *ArXiv E-prints*, arXiv-1611.
- Chung, K. R., Park, K. R., & Park, S. H. (2021). Design and implementation of a music recommendation model through social media analytics. *Journal of Convergence for Information Technology*, 11(9), 214-220.
- Demir, F. O., & Kırdar, Y. (2007). Müşteri ilişkileri yönetimi: crm. *Review of Social, Economic & Business Studies*, 8, 293-308.
- Do, M. P. T., Nguyen, D. V., & Nguyen, L. (2010). Model-based approach for collaborative filtering. *The 6th International Conference on Information Technology for Education*, 217-228, 18-20 August 2010.
- Ekstrand, M. D., Riedl, J. T., & Konstan, J. A. (2011). Collaborative filtering recommender systems. *Foundations and Trends in Human-Computer Interaction*, 4(2), 81-173.
- Falk, K. (2019). *Practical recommender systems*. New York: Manning Publications.
- Gao, Y., Zhao, H., Zhou, Q., Qiu, M., & Liu, M. (2020). An improved news recommendation algorithm based on text similarity. *3rd International Conference on Smart BlockChain (SmartBlock)*, 132-136, 23-25 October 2020.
- Gupta, M., Thakkar, A., Gupta, V., & Rathore, D. P. S. (2020). Movie recommender system using collaborative filtering. *International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 415-420, 2-4 July 2020.

- Gupta, S., & Dave, M. (2021). A hybrid recommendation system for e-commerce. *In Proceedings of International Conference on Communication and Computational Technologies*, 229-236, Singapore: Springer.
- Hahsler, M. (2015). Recommenderlab: a framework for developing and testing recommendation algorithms. 13.03.2022 tarihinde <https://cran.r-project.org/web/packages/recommenderlab/vignettes/recommenderlab.pdf> adresinden alındı.
- Han, K. (2020). Personalized news recommendation and simulation based on improved collaborative filtering algorithm. *Complexity*, 2020, 1-12.
- Harper, F. M., & Konstan, J. A. (2015). The movielens datasets: history and context. *Acm Transactions on Interactive Intelligent Systems (tiis)*, 5(4), 1-19.
- Jitendra, M., & Radhika, Y. (2021). An automated music recommendation system based on listener preferences. *In Recent Trends in Intensive Computing*, 80-87. Amsterdam: IOS Press.
- Kathavate, S. (2021). Music recommendation system using content and collaborative filtering methods. *International Journal of Engineering Research & Technology (IJERT)*, 10(02), 167-171.
- Kumar, P. S. (2020). Recommendation system for e-commerce by memory based and model based collaborative filtering. *In Proceedings of the 11th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2019)*, (1182), 123, Springer Nature.
- Munuswamy, S., Saranya, M. S., Ganapathy, S., Muthurajkumar, S., & Kannan, A. (2021). Sentiment analysis techniques for social media-based recommendation systems. *National Academy Science Letters*, 44, 281–287.
- Murad, D. F., Heryadi, Y., Isa, S. M., & Budiharto, W. (2020). Personalization of study material based on predicted final grades using multi-criteria user-collaborative filtering recommender system. *Education and Information Technologies*, 25(6), 5655-5668.
- Pratama, B. Y., Budi, I., & Yuliawati, A. (2020). Product recommendation in offline retail industry by using collaborative filtering. *International Journal of Advanced Computer Science and Applications*, 11(9), 635-643.
- Rebelo, M. Â., Coelho, D., Pereira, I., & Fernandes, F. (2021). A new cascade-hybrid recommender system approach for the retail market. *International Conference on Innovations in Bio-Inspired Computing and Applications*, 371-380, 16-18 December 2021.
- Sariman, G. (2011). Veri madenciliğinde kümeleme teknikleri üzerine bir çalışma: k-means ve k-medoids kümeleme algoritmalarının karşılaştırılması. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 15(3), 192-202.

- Shen, J., Zhou, T., & Chen, L. (2020). Collaborative filtering-based recommendation system for big data. *International Journal of Computational Science and Engineering*, 21(2), 219-225.
- Silahtaroglu, G. (2013). *Veri madenciliği kavram ve algoritmaları*. İstanbul: Papatya Yayıncılık Eğitim.
- Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, 2009, 1-19.
- Zhou, L., Zhang, F., Zhang, S., & Xu, M. (2021). Study on the personalized learning model of learner-learning resource matching. *International Journal of Information and Education Technology*, 11(3).