

Telekomünikasyon Sektöründe PSO ile Müşteri Bölümlemesi

Mihrimah ÖZMEN¹, Yılmaz DELİCE², Emel KIZILKAYA AYDOĞAN¹

¹Endüstri Mühendisliği, Erciyes Üniversitesi, Kayseri, Türkiye

²Yönetim ve Organizasyon Bölümü, Erciyes Üniversitesi, Kayseri, Türkiye
mihrimahozmen@gmail.com, yilmaz80@gmail.com, ekaydogan@erciyes.edu.tr

(Geliş/Received: 18.12.2017; Kabul/Accepted: 21.04.2018)

DOI: 10.17671/gazibtd.368460

Özet— Rekabetçi piyasa ekonomisi koşullarında, işletmelerin gelişiminde etkili olan en önemli kaynak müşterilerdir. Farklı müşteri gruplarının tercihlerini, alışveriş tutumlarını ve fiyat duyarlılıklarını anlamak; pazarlama faaliyetlerinin yönelimi açısından çok önemlidir. Bu durumda müşteri bölümlemesi, hedef pazardaki uygun müşteri gruplarını seçmek için kullanılmaktadır. Bu çalışmada, Türkiye'nin ilk 100 telekomünikasyon şirketlerinden birine müşteri bölümlemesi uygulanmıştır. Çalışmada yer alan firmanın veri ambarından, müşteri davranışlarıyla ilgili çağrı detayları, fatura bilgisi, müşteri demografik özellikleri gibi çok miktarda veri toplanmıştır. Operasyonel ve stratejik amaçlar için uygulanabilecek değerli bilgileri içeren bu verilerin boyutu, manuel analiz için çok büyüktür. Bu nedenle, verilerden anlamlı bilgi çıkarmak için bu çalışmada, gelişmiş veri madenciliği tekniklerinden yararlanılmıştır. PSO (Parçacık Sürü Optimizasyonu) tabanlı kümeleme tekniği ve DB ((Davies-Bouldin) uygunluk fonksiyonu ile müşteri bölümleri belirlenmiştir.

Anahtar Kelimeler— Müşteri Bölümlemesi, Kümeleme, Meta Sezgisel Optimizasyon, PSO

Customer Segmentation with PSO in Telecommunication Sector

Abstract— Under the competitive market economy conditions, the most important source of the development of businesses is the customers. Understanding the preferences, shopping attitudes and price sensitivities of different customer groups is very important in terms of the direction of marketing activities. In this case, customer segmentation is used to select the appropriate customer groups in the target market. In this study, customer segmentation was applied to one of the Turkey's top 100 telecommunication companies. The company involved in the study collects a lot of data on customer behaviors such as call details, billing information, customer demographics, etc. in the data warehouse. The size of this data is so large that manual analysis is not possible. However, these data contain valuable information that can be applied for operational and strategic purposes. Advanced data mining techniques are required to obtain meaningful information from these data. In this study, customer segments were identified with PSO-based clustering technique and DB(Davies-Bouldin) fitness function.

Keywords— Customer Segmentation, Clustering, Meta Heuristic Optimization, PSO

1. GİRİŞ (INTRODUCTION)

MİY (Müşteri ilişkileri yönetimi), bir işletmenin müşterilerini anlamaya çalışan iş süreçlerini ve teknolojilerini birleştirerek, müşterilerle olan ticari etkileşimleri yönetme çabalarının tamamı olarak tanımlanabilir [1]. MİY, geleceğin önde gelen iş stratejilerinden birisi haline gelmektedir. Şirketler, MİY' in sağladığı müşteri sadakatini artırma, yüksek müşteri karlılığı, müşteri için değer yaratma, ürünlerin ve hizmetlerin özelleştirilmesi, süreçleri kısaltma, daha

yüksek kaliteli ürünler ve hizmetler gibi faydalardan yararlanmaktadırlar [2].

Müşterilerin getirilerini anlamak ve karlı müşterileri korumak, MİY' in faaliyetlerinin temel parçalarındandır. Müşterilerinin kar potansiyellerini geliştirmek için birçok şirket halihazırda müşterilerinin değerlerini ölçmeye, bu değerlere göre müşterilerini bölümlemeye ve karlı müşteri bölümlerini korumak için stratejiler geliştirmeye çalışmaktadır [2].

Müşteri bölümlenme, her bir müşteriyi tanımlanan müşteri gruplarına göre sınıflandırmaktır [3]. Bunun için literatürde iki yaklaşım yer almaktadır. İlki; müşteri değerine dayalı bölümlenme ve diğeri ise; veri madenciliği tekniklerinden özellikle kümeleme yaklaşımı ile müşteri bölümlenmesidir [4]. Kümeleme: KKO (Karınca Koloni Algoritması), GA (Genetik Algoritma), Diferansiyel Gelişim, Yapay Arı Kolonisi, Benzetimli Tavlama; ayrıca Bal Arısı Çiftleştirme ve Bakteriyel Besin Arama optimizasyonları; bununla birlikte Yapay Balık Sürüsü, Guguk Kuşu Arama, Yabani Ot, Yerçekimi Arama, Kara Delik Arama Algoritmaları ve gibi birçok sezgisel yaklaşım temel alınarak literatürde etkili bir şekilde uygulanmıştır [5-20].

Bu çalışmada, Türkiye'nin ilk 100 telekomünikasyon şirketlerinden birinin, müşteri beklentilerini ve davranışlarını daha iyi anlamak ve bunun için stratejiler geliştirmek için müşteri bölümlenme modeli önerilmiştir. İlk olarak, şirketin müşteri bölümlenmesine katkı sağlayacak en etkili nitelikler belirlenmiş ve daha sonra PSO tabanlı kümeleme yaklaşımı uygulanmıştır. Uygulama sonuçları, önerilen modelin telekomünikasyon sektörü müşteri bölümlenmesi için iyi sonuçlar verdiğini göstermektedir. Böylece, telekomünikasyon sektörü müşteri yapısına uygun bir sistem kurarak rekabet gücünü daha da iyi bir hale getirebilmek için sektöre katkıda bulunulmaya çalışılmıştır.

Bu çalışma hem önerilen model hem de uygulama açısından literatüre katkı sağlamaktadır. Literatür araştırmamıza göre, PSO tabanlı müşteri bölümlenme yaklaşımı ilk defa bu çalışmada uygulanmıştır. Aynı zamanda telekomünikasyon sektörüne de uygulaması bu çalışma ile literatüre kazandırılmıştır. Bunun yanı sıra müşteri veri seti literatürdeki benzer çalışmalara nazaran daha detaylı ve geniş nitelikler içermektedir.

2. TELEKOMÜNİKASYON SEKTÖRÜNDE VERİ MADENCİLİĞİ UYGULAMALARI (DATA MINING APPLICATIONS IN THE TELECOMMUNICATION SECTOR)

Telekomünikasyon sektörü özellikle son on yıllık süreçte, çoğunlukla şebeke alt yapısında önemli teknolojik gelişmeler kaydetmiştir. Bu gelişmeler sektörün ekonomideki ağırlığını artırmış ve sektördeki büyüme hızı ülke ekonomilerinin üzerinde seyretmiştir. Bu durum günümüz ekonomi altyapısının temelini oluşturmuştur. Telekomünikasyon sektörü, ekonomik gelişmenin en büyük itici güçlerinden ve göstergelerinden aynı zamanda geleceğin sektörlerinden birisi haline gelmiştir.

İnternet ve iletişim araçlarındaki gelişmelerin ekonomik büyümenin yanı sıra sosyal/kültürel kalkınma üzerindeki önemli etkisi bütün kesimlerce kabul görmektedir. Hem iş hem de günlük hayatın içinde önemli yer tutan telekomünikasyon sektörü stratejik bir sektör haline geldiğini ispatlamıştır. Bu sektörü anlayabilen ve iyi analiz edebilen ülkeler, global ekonomide söz sahibi olabilmektedir. Bu minvalde, son zamanlarda ülkeler

sektördeki kalitelerini üst seviyelere taşımada birbiriyle yarışmaktadır [21]. Sektörün öne çıkması ile araştırmacılar müşteri bölümlenme ile ilgili kapsamlı çalışmalar yapmışlardır [2,4,22 -26].

Telekomünikasyon sektörü, müşterileri hakkında çok büyük hacimlerde bilgiye sahip olan, veri yoğun bir sektör olduğu için müşterilerden elde edilen karı maksimize etmek, müşteriyi elde tutmak ve müşteriyi tanımak için birçok alanda veri madenciliği yaklaşımları uygulamıştır. “Pazarlama ve müşteri ilişkilerini geliştirmek”, dolandırıcılık tespiti, ağ yönetimi olarak; uygulama alanları üç ana başlıkta incelenebilir. Pazarlama ve müşteri ilişkilerini geliştirebilmek için müşteri değerini ölçüp ve özelliklerini çıkarmak önemlidir. Buradan elde edilen bilgiler pazarlama faaliyetlerinde kullanılabilir. Örneğin: karlı müşterileri elde tutmak, yeni müşterileri elde etmek, müşteri kaybı analizi.

Telekomünikasyon sektöründe, pazarlama ve müşterilere yönelik son zamanlarda yapılan birçok veri madenciliği çalışmalarından başlıcaları şunlardır: Ahn vd. [27], müşterilerin daha önce kullanmış olduğu ürünler ve demografik özellikleri gibi verileri kullanarak çapraz satışta kullanılacak bir müşteri sınıflandırma modeli önermişlerdir. Cheng ve Sun [28], ise ilk olarak telekomünikasyon sektörü için değerli müşterileri belirleyen yeni bir model önermişler ve daha sonra müşterileri bölümlenmişlerdir. Lim vd. [29], mobil telekomünikasyon sektörü için Çin’de yeni bir bölgesel kümeleme yaklaşımı önermiştir. Chen vd. [30], sabit hat kullanıcılarının ödeme davranışlarını tahmin etmek için birliktelik kuralları, kümeleme ve karar ağaçları yöntemlerinden yararlanarak bir veri madenciliği stratejisi geliştirmişlerdir. Vidya vd. [31], mobil telefon satıcılarının itibar analizi için Twitter verilerini kullanarak duygu analizi yapmışlardır.

Telekomünikasyon sektöründe veri madenciliği uygulama alanlarından ikincisi dolandırıcılık tespittir. Dolandırıcılık yapan müşterileri tespit etmek, müşterilerin hesaplarına erişen dolandırıcıları denetlemek ve olağandışı kullanım düzenlerini bulmak bu kapsamda yapılan çalışmalar içerisinde yer almaktadır. Dolandırıcılık tespitinde ise sınıfların yanlış sınıflandırma maliyeti birbirinden farklı olduğu için standart veri madenciliği teknikleri yetersiz kalmakta ve maliyet duyarlı teknikler uygulanmaktadır [32].

Literatürde dolandırıcılık tespiti ile ilgili güncel birçok çalışma yer almaktadır. Farvaresh ve Sepeshri [33], çalışmalarında Tahran telekomünikasyon şirketi için önleme, kümeleme ve sınıflandırma aşamalarından oluşan bir hibrid yaklaşım uygulamışlardır. Olszewski [34], iki sınıflı bir problem olan hileli hesapları sınıflandırmak için bir çalışma yapmıştır.

Telekomünikasyon ağları ziyadesiyle karmaşık yazılım ve donanım yapıları içerir. Bu ağları etkin bir şekilde

yönetebilmek için meydana gelmeden önce ağ hataları belirlenmeli ve alarmlar otomatik olarak analiz edilmelidir.

Üçüncü veri madenciliği uygulama alanı olan telekomünikasyon sektörü ağ yönetiminde birçok çalışma yer almaktadır. Bu çalışmalar: “ağ hatalarını belirlemek ve tahmin etmek”, “sistemin iş yükünü ve kaynak kullanımını yönetmek”, “kullanıcı gruplarının davranışlarını belirlemek ve veri trafiğini karşılaştırmak” tır [35]. Bu alanda literatürde yer alan çalışmalar şunlardır: Chao vd. [36], Beijing üniversitesinden elde ettikleri verilerle olağandışı internet kalitesini k-ortalamar kümeleme algoritması ile analiz etmişlerdir. Ren vd. [37], yine k-ortalamar yöntemini kullanarak telekomünikasyon ağları için kümeleme çalışması yapmışlardır. Velmurugan [38] ise çalışmasında sunucu yerleşimleri için k-Ortalamar ve Bulanık C-Ortalamar kümeleme yöntemlerini uygulamıştır.

3. META SEZGİSEL TABANLI KÜMELEME YAKLAŞIMLARI (META HEURISTIC BASED CLUSTER APPROACHES)

Kümeleme, ilk örutü tanıma yaklaşımlarından birisidir ve veri madenciliğinin başlangıcından bu yana ilgi alanı olmuştur. Bu yaklaşımda, benzer özelliklere sahip nesnelerin anlamlı kümeleri oluşturulur ve sınıflandırmadan farklı olarak nesnelere daha önceden tanımlanmış sınıflara atanmazlar. Kümelemede, ilk olarak nesnelere arasındaki benzerliklerin belirlenebilmesi için ölçüler ve değişkenler seçilir. Bu ölçülere göre nesnelere kümelendirilir ve daha sonra belirlenen kümelerin uygunluğu değerlendirilip ve istatistiksel olarak geçerliliği ortaya konur.

Klasik kümeleme yaklaşımları ile istenen sonuçlar elde edilememektedir. Ancak bu yaklaşımlarla optimal olmayan yerel bir çözüme yakınsanabilmekte bazen yerel çözüme dahi yakınsamayı başılamamaktadır. Bu sorunun üstesinden gelebilmek için önerilen çeşitli tekniklerden birisi ise: kümeleme görevini, küme niteliğine dayalı bir nesnel işlevi kullanan optimizasyon problemi olarak modellemek ve daha sonra küresel optimumu bulmaya çalışmaktır. Bunlar NP-Tam problemdir. Meta sezgisel algoritmalar, diğerlerinden farklı olarak optima yakın çözümleri bulmak için kullanılmaktadır [39]. Dolayısıyla, PSO, GA ve KKO gibi meta sezgisel algoritmalar bu tür problemleri optimize etmek için kullanılan yöntemlerdendir. Jain vd. [40] farklı kümeleme tekniklerine ilişkin bir çalışma, Nanda ve Panda [41] ise meta sezgisel algoritmalarla ilişkin bir derleme çalışması yapmışlardır.

Meta sezgisel popülasyona dayalı optimizasyon algoritmaları kümelemede olduğu gibi genelde ya saf olarak ya da çoklu algoritmaları birleştiren melez bir varyant olarak kullanılmaktadır. Tablo 1’ de literatürde yer alan meta sezgisel kümeleme yaklaşımları listelenmiştir.

3.1. PSO Tabanlı Kümeleme Yaklaşımı (PSO Based Clustering Approach)

Populasyon tabanlı bir optimizasyon tekniği olan PSO kuş sürülerinin davranışlarından esinlenilerek 1995 yılında Russell Eberhart ve James Kennedy tarafından ortaya atılmıştır [42]. Çok parametrelili ve çok değişkenli optimizasyon problemlerine hızlı sonuç bulması, az parametre gerektirmesi ve yerel optimumlara takılma riskinin az olması sebebiyle diğer birçok arama algoritması üzerinde üstünlük kurmuştur.

PSO’da parçacıklar muhtemel çözümleri, parçacık topluluğu ise sürüyü ifade etmektedir. Parçacık, arama uzayındaki herhangi bir kuşu sembolize etmektedir. Doğadaki kuşlar yiyecek ararken hem kendi tecrübelerine bağlı hareket ederlerken bir yandan da yiyeceğe en yakın olan kuşu takip etme davranışı da göstermektedirler. Böylece her bir parçacığın değeri sürü hafızasının da etkisiyle daha iyi noktaya ulaşmaktadır. PSO, birçok meta sezgisel algoritmada olduğu gibi rastgele çözümler içeren bir popülasyonla başlar ve en iyi (optimum) çözümü araştırır. Her iterasyonda; parçacık konumları, komşularının ve kendi en iyi koordinatlarının bir birleşimi olarak belirlenir. Sürüdeki her bir parçacığın o ana kadar elde ettiği en iyi çözüm pbest (yerel en iyi) olarak adlandırılırken popülasyonda o ana kadar tüm parçacıklar tarafından elde edilen en iyi çözüm ise gbest (global en iyi) olarak adlandırılır.

Tablo 1. Meta Sezgisel Kümeleme Yaklaşımları
(Meta Heuristic Based Cluster Approaches)

Cowgill vd. 1999 [5]	GA
Murthy ve Chowdhury 1996 [6]	
Krishna and Murty 1999 [7]	
Kwedlo 2011 [8]	Diferansiyel Gelişim
Karaboga ve Ozturk 2011 [9]	Yapay Arı Kolonisi
Zhang vd. 2010 [10]	
Afshar vd. 2007 [11]	Bal Arısı Çiftleştirme
Fathian vd. 2007 [12]	Optimizasyonu
Cheng vd. 2009 [13]	Yapay Balık Sürüsü
Zhu vd. 2012 [14]	Algoritması
Wan vd. 2012 [15]	Bakteriyel Besin Arama Optimizasyonu
Saida vd. 2014 [16]	Guguk Kuşu Arama Algoritması
Chowdhury vd. 2011 [17]	Yabani Ot Algoritması
Hatamlou vd. 2012 [18]	Yerçekimi Arama Algoritması
Hatamlou 2013 [19]	Kara Delik Arama
Niknam vd. 2008 [20]	KKO ve Benzetimli Tavlama

PSO algoritmasındaki her bir parçacık iki önemli parametreye sahiptir. Pozisyon olarak adlandırılan ilk parametre parçacığın o andaki konumunu belirtmektedir.

Hız olarak adlandırılan ikinci parametre ise parçacığın iki iterasyon arasında konumundaki değişim miktarını göstermektedir. Algoritmada, başlangıç pozisyonu ve hız değerlerinin rastgele olarak üretilmesi ile sürü oluşturulur. Algoritma başlatıldıktan sonra ikinci adımda sürü içerisindeki tüm parçacıklara ait uygunluk değerleri hesaplanır. Üçüncü adımda her bir parçacık için o ana kadar elde edilen pbest değerleri belirlenir. Elde edilen pbest değerleri toplam parçacık sayısı kadardır. Dördüncü adımda sürüdeki her bir parçacığın sahip olduğu pbest değerleri içerisinde en iyisi belirlenerek gbest olarak saklanır. Son adımda ise her bir parçacığa ait pozisyon ve hız değerleri aşağıdaki denklemler kullanılarak yenilenir.

$$V_{id}^{t+1} = W * V_{id}^t + c_1 * rand_1 * (P_{id} - X_{id}) + c_2 * rand_2 * (P_{gd} - X_{id}) \quad (1)^{42}$$

$$X_{id}^{t+1} = X_{id}^t + V_{id}^{t+1} \quad (2)^{42}$$

Burada X_{id} pozisyon, V_{id} hız değerlerini gösterirken, $rand_1$ ve $rand_2$ değerleri rastgele üretilmiş sayılardır. W atalet ağırlık değeri ve C_1 , C_2 ise ölçeklendirme faktörleridir. Yukarıda açıklanan adımlar durdurma kriteri sağlanana kadar devam ettirilir.

Literatürde, PSO algoritması verilerin kümelenmesinde birçok kez kullanılmıştır [43-45]. PSO, k-ortalamlar ile hibrid edilerek yakınsama yeteneği geliştirilmeye çalışılmıştır [45-48]. Niknam ve Amiri [49]; PSO, KKO ve k-ortalamları içeren bir algoritma ile kümeleme yaklaşımını uygulamışlardır. Ye ve Chen [50], Gaussian temel alan metriğini kullanan PSO kümeleme algoritmasını uygulamışlardır. Kuo ve Lin [51], GA ve PSO'yu kombine eden bir kümeleme algoritması önermişlerdir. GA, PSO ve k-ortalamlar' dan oluşan bir başka hibrid algoritma ise, küme sayısını belirlemek ve kümelemek için uygulanmıştır [52].

4. TELEKOMÜNİKASYON SEKTÖRÜNDE MÜŞTERİ BÖLÜMLEMESİ (CUSTOMER SEGMENTATION IN TELECOMMUNICATION SECTOR)

Telekomünikasyon şirketleri, binlerce müşteriyle düzenli olarak ilgilenmeye, aynı seviyede hizmet vermeye ve sosyal erişim sağlamaya çalışmaktadır. Ancak bütün müşterileri tek tek ele alarak bu mümkün değildir. Müşteri ihtiyaçlarını daha iyi anlayabilmek ve karşılayabilmek için müşteriler bölümlenmelidir. Bölümleme, büyük bir müşteri popülasyonunu daha küçük, daha yönetilebilir alt popülasyonlara dönüştürme olanağı sağlamaktadır. Ayrıca bölümleme; hedefleri daha gerçekçi, taktiksel bir şekilde izlenebilir ve yürütülebilir hale getirebilmektedir.

Bölümleme modellerinde önemli bir konu ise belirlenen müşteri gruplarının geçerlilik süreleridir. Kısa ömürlü bölümleme modelleri yaygın olarak pazarlama kampanyaları için kullanılabilir. Belirli pazarlama faaliyetleri için en iyi müşteri grupları ve bu gruplardaki kişilere hitap edecek en iyi reklam mesajları belirlenmeye çalışılmaktadır. Bölümler tanımlandıktan ve kampanya

uygulandıktan sonra, genellikle bu bölümleme tekrar kullanılmamaktadır.

Geçici olmayan bir bölümleme modeli hazırlanırken yaşam döngüsü dikkate alınmalıdır. Aksi takdirde, belirli bir süre sonra anlamına yitirecek olan bu bölümleme modellerinin kullanılmaları mantıklı olmayacaktır. Durağan bir model yerine periyodik olarak yenilenecek ve düzenlenecek yenilenebilir bölümleme modeli geliştirmek uzun vadeli ilerlemelerin izlenmesini ve organizasyonel değerlendirme raporlarının doğruluğunu ve uygunluğunu önemli ölçüde artıracaktır.

Bölümleme yalnızca müşterinin bir özelliğiyle sınırlanabilir (örneğin: müşterileri yalnızca yaş, cinsiyet veya coğrafi bölgelere göre bölümlere ayırmak) ancak, piyasa koşullarında bu bölümleme etkili olmayacaktır. Müşteriler, coğrafi (fiziksel konumları), davranışsal (satın alma, alışveriş ve ürün kullanım davranışları), demografik ve psikolojik (fiziksel, sosyal, psikolojik veya tutumları) özelliklerine göre bölümlenebilmektedir [53].

Literatürde telekomünikasyon sektöründe müşteri bölümleme hakkında birçok çalışma yer almaktadır. Hung vd. [22]; karar ağaçları, yapay sinir ağları ve k-ortalamlar teknikleri ile müşteri bölümlemesi gerçekleştirmişlerdir. Hwang vd. [23], bir kablosuz iletişim şirketi için tüketici ömür değeri modeli önermişlerdir. Bu modelde, bir müşteriden geçmişte elde edilen kazanç, müşterinin potansiyel faydası ve ayrılma ihtimali göz önüne alınmıştır. Müşteriler; güncel değerleri, potansiyel değerleri ve sadakatlerine göre bölümlere ayrılmıştır. Kim vd. [2], bir kablosuz telekomünikasyon şirketinin müşterilerini değerlerine göre bölümlere ayırmışlardır. Han vd. [4], müşteri yaşam döngüsüne dayalı ve müşterinin güncel, geçmiş ve uzun vadeli değer tahminini, kredibilitelerini ve sadakatini içeren yeni bir müşteri bölümleme yöntemi önermişlerdir. Önerilen model, Çin ülkesindeki bir telekomünikasyon operatörü müşterilerine uygulanıp bu modelin etkinliği gösterilmiştir. Hamka vd. [24], şebeke operatörleri, telefon üreticileri ve uygulama geliştiricileri perspektifinden; mobil telekomünikasyon sektöründe müşteri bölümlemesi geliştirmişlerdir. Araştırmaları, yenilikçi olmakla birlikte müşteri bölümlemesinin akıllı telefon ölçüm verileriyle alakalı olduğunu da ortaya koymuştur. Ye vd. [25], Jiangsu bölgesindeki Changzhou telekomünikasyon şirketi için KXEN veri madenciliği programını kullanarak k-ortalamlar kümeleme yöntemi ile müşteri bölümlemesi gerçekleştirmiştir. Zhao vd. [26], küme merkezini başlatmak için k-ortalamlar kümeleme algoritması için yeni bir yaklaşım önermişlerdir. Önerilen yaklaşım, telekomünikasyon müşterilerini bölümlemek için uygulanmıştır. Elde edilen sonuçlar, yeni algoritmanın verimlilik ve doğruluk bakımından orijinal algoritmaya göre daha iyi sonuçlar verdiğini göstermiştir.

5. UYGULAMA (CASE STUDY)

Telekomünikasyon sektöründe, piyasa şartları gereği müşteri bölümlenmesi giderek önem kazanan bir problem haline gelmektedir. Bu nedenle hazırlanan çalışmada telekomünikasyon sektörünün bu problemine yönelik olarak müşteri veri karakteristiğine uygun PSO tabanlı müşteri bölümlenme yaklaşımı önerilmiştir. Bu yaklaşımda, PSO bir kümeleme algoritması olarak kullanılmış ve doğru küme sayısının yanında en uygun kümeleme seçeneği de bulunmaya çalışılmıştır. Uygulamada uygunluk fonksiyonu DB indeks olarak belirlenmiştir. Ayrıca önerilen modelden mümkün olduğunca gerçekçi sonuçlar elde edebilmek için, çok sayıda nitelik içeren müşteri veri seti derlenmiştir. Bölümlemede kullanılan veriyi temsil eden örnek veri kümesi ekte verilmiştir.

Önerilen model; veri toplama, ön işleme ve kümeleme olmak üzere 3 ana aşamadan oluşmaktadır ve temel adımları aşağıda verilmiştir:

ÖNERİLEN MÜŞTERİ BÖLÜMLEME MODELİ

Aşama 1: Veri Toplama

Müşteri bölümlenmesi için uygun niteliklerin belirlenmesi

Veri setinin, ETL araçları kullanılarak şirket veri ambarlarından çekilmesi

Aşama 2: Veri Temizleme ve Önleme

Eksik değerlerin doldurulması ve hatalı verilerin düzeltilmesi

Aşama 3: PSO tabanlı Kümeleme

Her bir parçacık, rastgele seçilmiş küme merkezlerini içerecek N_c şekilde başlatılır.

For $t=1$ to t_{max} (en büyük iterasyon sayısı) do

Her bir parçacık i için DB değeri hesaplanır

Global ve lokal en iyi pozisyonlar

güncellenir

Küme Merkezleri güncellenir

Aşama 1: Veri Toplama

Müşteri bölümlenmesini mümkün olduğunca gerçeğe yakın gerçekleştirebilmek için müşteri verilerini şirket veri ambarlarından çekmeden önce veri setinde yer alması gereken niteliklerin eksiksiz ve kapsayıcı olarak belirlenmesi önemlidir. Literatürde, veri kümesindeki niteliklerin kümelemede önemini göz önünde bulunduran birçok çalışma yer almaktadır. Genellikle, veri madenciliği uygulamalarının tamamında olduğu gibi bilgi toplanırken, her bir niteliğin uygulamadaki önemi dikkate alınmadan mümkün olduğunca fazla bilgi toplanmaya çalışılır. Ancak veri kümesindeki niteliklerin kapsayıcı olması sağlanmaya çalışırken alakasız niteliklerin veri kümesine eklenmesi, kümeleme kalitesini düşürebilmektedir. Kümeleme açısından; ilgisiz niteliklerin çıkarılması, gerekli depolama ve hesaplama sürelerini azaltılmasını sağlarken aynı zamanda anlamsız kümelemelerin üretilmesini de önüne geçilebilir. Gereksiz nitelikler çıkarılırken bir yandan da

gerekli olan niteliklerin de gözden kaçırılmaması kümelemenin doğruluğu açısından çok önemlidir [58]. Bu nedenle nitelikler belirlenirken literatürde yer alan çalışmalar incelenmiş daha sonra da sektörde uzmanların ve akademisyenlerin görüşleri alınarak Şekil 1' de ana başlıkları verilen uygun nitelikler belirlenmiştir. Belirlenen niteliklerle veri kümesi, ETL (Çıkarma-Dönüştürme-Yükleme) araçları kullanılarak şirket veri ambarlarından çekilmiştir. Modelde uygulanan ETL işlemleri şunlardır: ilgili müşteri bilgilerinin kaynağa tanımlanması, bu bilgilerin çıkarılması, çeşitli kaynaklardan gelen verilerin entegre edilmesi, veri kümesinin temizlenmesi ve verilerin veri ambarına yüklenmesi.

Aşama 2: Veri Temizleme ve Önleme

Veri madenciliğinde, iyi sonuçlar elde edebilmek için önemli bir aşama olan veri temizleme ve hazırlama, toplam veri mühendisliği çabasının yaklaşık %80' ini oluşturmaktadır. Veri setlerinde eksik veri oranının % 1'den az olması, genellikle önemsiz olarak kabul edilir ve %1-5 oranında eksik veri içeren veri setleri ise yönetilebilmektedir. Ancak %5-15 oranında eksik veri içeren veri setleri ile başa çıkabilmek için gelişmiş yöntemlerin uygulanması gerekmektedir. %15'den fazla oranda eksik veri içeren veri setleri ise veri madenciliği uygulamalarında ciddi şekilde sonuçları etkileyebilmektedir [55]. Bu çalışmada uygulanan müşteri veri setinde eksik veri oranı %8.9' dur ve bu nedenle bölümlenmede başarılı bir netice elde edebilmek için veri temizleme ve ön işleme aşamasına ayrıca önem verilmiştir. Veri setindeki kayıp verilerle başa çıkabilmek için; eksik verileri veri setinden silmek yerine, eksik veriler tamamlanmıştır. Böylece bilgi kaybının önüne geçilmiştir. Veri setindeki eksik değer içeren nitelik süreklilye ortalama değerle, nitelik kategorik veya ikili ise en sık kullanılan değer ile tamamlanmıştır. Veri kümesindeki gürültü ve yanlış değerlerle başa çıkabilmek içinse; hatalı değerler, düzeltilmiş ve aykırı değerler çıkarılmıştır. Veri kümesi düzenlendikten sonra, müşteri bölümlenmesinde kullanılmak üzere veri tabanına yüklenmiştir.

Aşama 3: PSO tabanlı Kümeleme

Kümeleme problemlerinde, her bir parçacık küme merkez vektörü N_c 'yi (N_c küme merkezlerinin yani oluşturulacak kümelerin sayısını belirtir) temsil eder ve her bir parçacık x_i aşağıdaki gibi gösterilir:

$$x_i = (m_{i1}, \dots, m_{ij}, \dots, m_{iN_c}) \quad (3)$$

m_{ij} , C_{ij} kümesindeki i inci parçacığın j küme ağırlık merkezi vektörünü belirtir. Bu durumda, parçacık sürüsü, mevcut veri vektörleri için aday kümeleri temsil eder.

Kümeleme aşamasında uygunluk fonksiyonu olarak DB indeksi uygulanmıştır. DB indeksi, küme dağılımı içindeki toplamın küme ayırımı arasındaki oranı bulmaktadır [56]. Bu metrik, küme ayırılmaları tahmin etmek için kullanılmaktadır. Daha küçük indeks değeri, daha iyi

kümelemeyi göstermektedir. K küme sayısını göstermekte iken DB metriği aşağıdaki şekilde tanımlanmıştır [56]:

$$DB = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} \{R_{ij}\}, \quad (4)^{56}$$

R_{ij} , c_i ve c_j kümeleri arasındaki benzerlik ölçüsüdür.

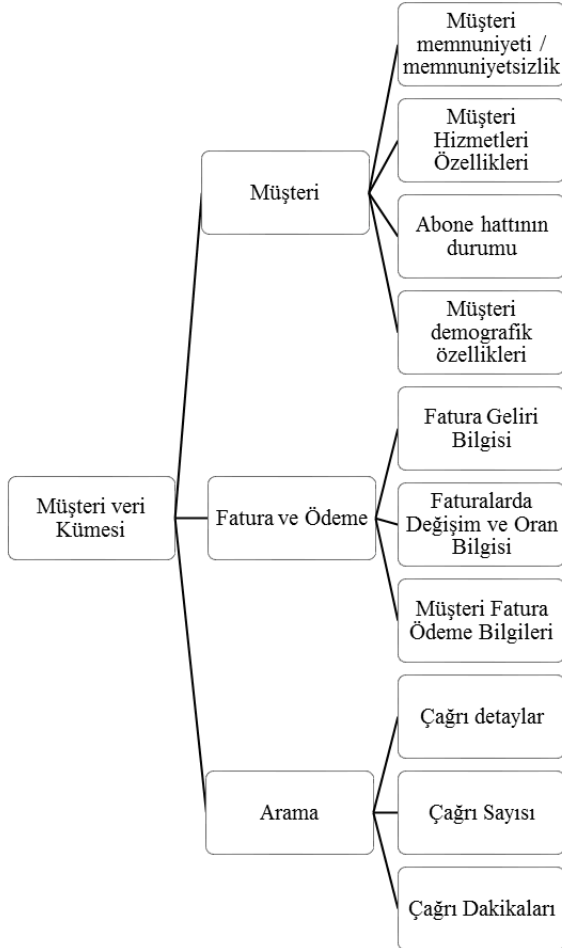
$$R_{ij} = \frac{S_i + S_j}{M_{ij}}, \quad (5)^{56}$$

$|c_i|$ i kümesi eleman sayısı iken S_i ve S_j , i ve j . kümelerin sırasıyla dağılımlarını göstermektedir

$$S_i = \frac{1}{|c_i|} \sum_{x \in c_i} \|x - z_i\|, \quad (6)^{56}$$

M_{ij} ise bu iki değer arasındaki uzaklığı göstermektedir.

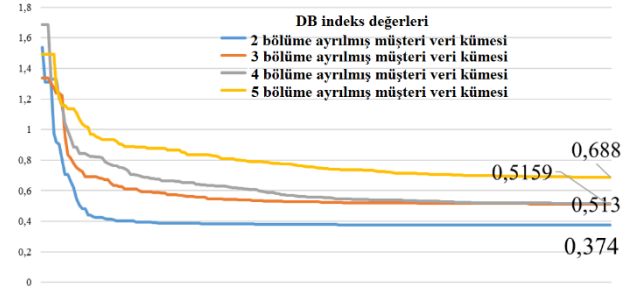
$$M_{ij} = \|z_i - z_j\|. \quad (7)^{56}$$



Şekil 1. Müşteri Veri Kümesi [54]
(Customer Dataset)

5.1. Deneysel Sonuçları (Experimental Results)

PSO tabanlı bölümlenme yaklaşımı ile müşteriler 2,3,4, ve 5 bölüme ayrılarak 4 ayrı bölümlenme sonucu elde edilmiştir. Şekil 2' de PSO bölümlenme yaklaşımının 4 ayrı bölümlenme sonuçlarının iterasyonel ilerlemesi gösterilmiştir. Belirli iterasyon sayısından sonra DB indeks değerinde değişim olmaması PSO' nun yeterli iterasyon sayısında çalıştırıldığını göstermektedir.



Şekil 2. Bölümlenmiş müşteri veri kümelerine ait DB indeks değerlerinin iterasyon sayısına göre değişimi
(The DB index values of the segmented customer data sets according to the number of iterations)

Ayrıca yine sonuçlar göstermiştir ki bölüm sayısı arttıkça DB indeks uygunluk değeri giderek kötüleşmektedir. 2 bölüme ayrılan müşteri bölümlenme sonucu 0.374 DB indeks değeri ile en iyi bölümlenmeyi verirken, giderek artan bölüm sayısı ile doğru orantılı olarak DB indeks değeri daha da kötüleşerek 5 bölüme ayrılmış müşteri bölümlenme sonucu 0.688 DB indeks sonucunu vermiştir. Bu nedenle deneyler müşterileri en çok 5 bölüme ayrılarak sonlandırılmıştır. Bu 4 farklı bölümlenmeden elde edilen sonuçlar aşağıda açıklanmıştır.

Müşteri bölümleri belirlenirken birçok nitelik arasından daha çok aşağıdaki 10 niteliğin öne çıktığı belirlenmiştir:

- K1: müşterinin ortalama fatura tutarı
- K2: müşteriden elde edilen toplam gelir
- K3: müşterinin son 6 ayda ortalama çağrı adedi
- K4: müşterinin ortalama arama uzunluğu (sn)
- K5: müşterinin ortalama paket dışı arama uzunluğu (sn)
- K6: müşterinin şikâyet sayısı
- K7: müşterinin taahhüdünün olup olmaması (taahhtülü ise 1, değilse 2 değerini almaktadır)
- K8: müşterinin kontratlı olduğu ay sayısı
- K9: müşterinin ödenmemiş aylık fatura adedi
- K10: müşterinin şimdiye kadar müşteri hizmetlerini arama sayısı

Tablo 2. 2 Bölüme Ayrılmış Müşteri Veri Kümesinin
Deney Sonuçları
(Experimental Results of 2 Segmented Customer Dataset)

Nitelikler	Müşteri Bölümlerine Ait Ortalama Nitelik Değerleri	
	B-1	B-2
K1	29.691	51.365
K2	51.768	149.662
K3	38.543	205.571
K4	343.145	1635.786
K5	1187.688	5061.5
K6	9.844	9.429
K7	1.056	1
K8	18.339	17.071
K9	0.615	1.5
K10	12.774	24.5
DB indeks Değeri	0.374	

2,3,4 ve 5 bölüme ayrılmış müşteri bölümlenmeye ait sonuçlar Tablo 2-5'te özetlenmiştir. Bu tabloların her birinde, ait olduğu bölümlenmedeki her bir bölümün yukarıda belirtilen 10 niteliğe ait ortalama değeri ve bölümlenmenin DB indeks değeri yer almaktadır. Buna göre; 2 bölüme ayrılmış müşteri veri kümesinin DB indeks değeri 0.374, 3 bölüme ayrılmış müşteri veri kümesinin DB indeks değeri 0.513, 4 bölüme ayrılmış müşteri veri kümesinin DB indeks değeri 0.5159 ve 5 bölüme ayrılmış müşteri veri kümesinin DB indeks değeri 0.688' dir. DB indeks değerlerine göre, en küçük 0.374 değeri ile en iyi bölümlenme sonucunu 2 bölüme ayrılmış müşteri veri seti vermektedir.

Tablo 3. 3 Bölüme Ayrılmış Müşteri Veri Kümesinin
Deney Sonuçları
(Experimental Results of 3 Segmented Customer Dataset)

Nitelikler	Müşteri Bölümlerine Ait Ortalama Nitelik Değerleri		
	B-1	B-2	B-3
K1	27.838	29.745	61.412
K2	81.756	51.272	176.668
K3	125.19	36.972	275.5
K4	796.341	335.176	2294.833
K5	2416.54	1167.128	6228.833
K6	9.913	9.842	9.5
K7	1	1.057	1
K8	17.254	18.359	16.5
K9	1.167	0.605	1.5
K10	16.643	12.701	36.5
DB indeks Değeri	0.513		

Tablo 4. 4 Bölüme Ayrılmış Müşteri Veri Kümesinin
Deney Sonuçları
(Experimental Results of 4 Segmented Customer Dataset)

Nitelikler	Müşteri Bölümlerine Ait Ortalama Nitelik Değerleri			
	B-1	B-2	B-3	B-4
K1	8.62	27.43	29.099	29.754
K2	25.23	314.32	87.589	51.234
K3	688	50	125.05	36.892
K4	499	363	847.482	334.781
K5	4647	3451	2450.504	1168.016
K6	11	8	9.95	9.84
K7	1	1	1	1.057
K8	24	25	17.312	18.359
K9	1	2	1.184	0.605
K10	9	6	17.794	12.688
DB indeks Değeri	0.5159			

Tablo 2-5'deki niteliklerin alt üst aralıkları birbirinden çok farklıdır. Nitelik değerlerinin bölümlere göre değişimini şekil üzerinde daha net gösterebilmek için normalize değerler tercih edilmiştir. Şekil 3-6'da Tablo 2-5'te yer alan nitelik değerlerinin değişimi denklem (8)' deki doğrusal normalizasyon tekniği ile normalize edildikten sonra grafik üzerinde gösterilmiştir [57].

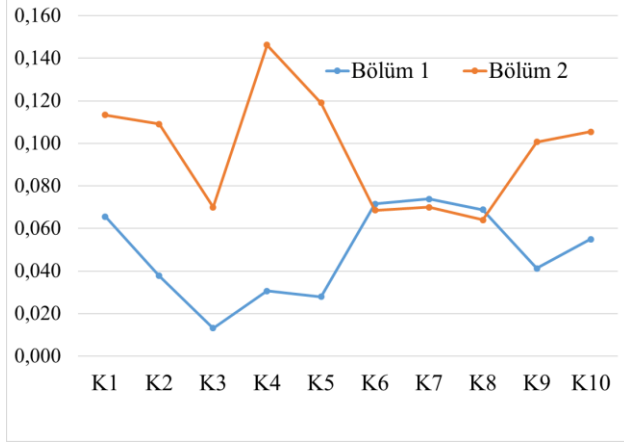
$$r_{ij} = \frac{x_{ij}}{\sum_{i=1}^m x_{ij}}, i = 1, 2, \dots, m; j = 1, 2, \dots, n \quad (8)^{57}$$

Tablo 5. 5 Bölüme Ayrılmış Müşteri Veri Kümesinin
Deney Sonuçları
(Experimental Results of 5 Segmented Customer Dataset)

Nitelikler	Müşteri Bölümlerine Ait Ortalama Nitelik Değerleri				
	B-1	B-2	B-3	B-4	B-5
K1	61.412	28.519	29.71	29.76	8.62
K2	176.668	76.93	51.375	51.872	25.23
K3	275.5	313.933	37.722	37.975	688
K4	2294.833	260.933	337.004	344.392	499
K5	6228.833	1475.867	1173.258	1191.672	4647
K6	9.5	10	9.853	9.852	11
K7	1	1	1.056	1.056	1
K8	16.5	17	18.341	18.317	24
K9	1.5	1	0.61	0.617	1
K10	36.5	12.6	12.736	12.734	9
DB indeks Değeri	0.688				

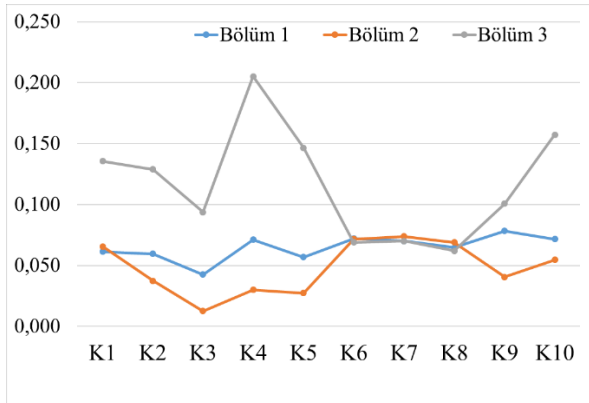
Şekil 3-6 incelendiğinde 2 bölüme ayrılmış müşteri veri setinde bölümlerin almış olduğu ortalama nitelik değerlerinin birbirinden belirgin şekilde farklı olduğu net

olarak görülmektedir. 5 bölüme ayrılmış müşteri veri setinde, 3. ve 4. bölüm müşteri verileri birbirine çok yakın olduğu için grafikte bu bölümlerin müşteri veri değerleri üst üste gelmiş sadece 4 bölümün değerleri şekilde görülebilmektedir.

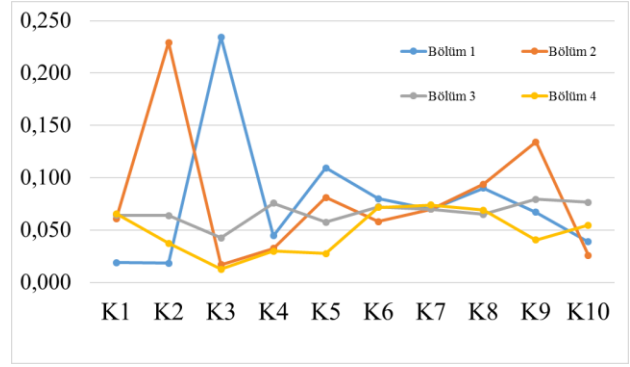


Şekil 3. 2 Bölüme Ayrılmış Müşteri Veri Kümesinde Her Bölüme ait Ortalama Nitelik Değerleri
(The average criteria values for each segment in 2 segmented customer dataset)

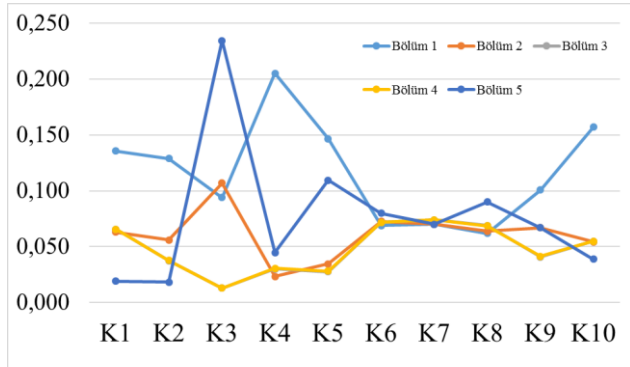
DB indeks değerlerinin de doğruladığı gibi 0.374 ile 2 bölüme ayrılmış müşteri veri kümesi daha anlamlı bir bölümlenme sonucu vermiştir. Bu bölümlenmedeki müşteri grupları; ortalama fatura tutarı, elde edilen toplam gelir, son 6 ayda ortalama çağrı adedi, ortalama arama uzunluğu (sn), ortalama paket dışı arama uzunluğu (sn) ve şimdye kadar müşteri hizmetlerini arama sayısı nitelikleri açısından birbirinden belirgin şekilde ayrılmaktadır. Şikâyet sayısı, taahhüdünün olup olmaması, kontratlı olduğu ay sayısı ve ödenmemiş aylık fatura adedi nitelikleri açısından diğer niteliklere nazaran benzer özellikler göstermektedir. Buna göre müşteri bölümlerinin, telekomünikasyon ürünlerini sık kullanıp kullanmamasına göre bölümlere ayrılmış olduğu görülmektedir.



Şekil 4. 3 Bölüme Ayrılmış Müşteri Veri Kümesinde Her Bölüme ait Ortalama Nitelik Değerleri
(The average criteria values for each segment in 3 segmented customer dataset)



Şekil 5. 4 Bölüme Ayrılmış Müşteri Veri Kümesinde Her Bölüme ait Ortalama Nitelik Değerleri
(The average criteria values for each segment in 4 segmented customer dataset)



Şekil 6. 5 Bölüme Ayrılmış Müşteri Veri Kümesinde Her Bölüme ait Ortalama Nitelik Değerleri
(The average criteria values for each segment in 5 segmented customer dataset)

6. SONUÇ (CONCLUSION)

Müşteri ilişkileri yönetiminde müşteri bölümlenme, birçok sektörde olduğu gibi telekomünikasyon sektörü için de vazgeçilmez yönetsel bir araçtır. Bu çalışmada, telekomünikasyon sektörüne özgü PSO temelli bir müşteri bölümlenme model önerilmiştir. Önerilen modelde, şirketin müşteri bölümlenmesine katkı sağlayacak en etkili nitelikler belirlenmiştir. Bu model Türkiye'de bir telekomünikasyon şirketinin müşteri verilerine uygulanmıştır. Uygulama sonuçları, önerilen modelin müşteri bölümlenmesini etkili bir şekilde uygulayabildiğini göstermiştir.

İleriki çalışmalarda, gerçekleştirilen uygulama müşteriyi daha iyi tanımak ve ihtiyaçlarını karşılamak için nitelik indirgeme ile entegre edilerek iyileştirilebilir. Ayrıca, bu çalışmada telekomünikasyon sektörüne uygulanan PSO tabanlı müşteri bölümlenme sonuçları, ileride farklı meta sezgisel algoritmalar ile elde edilen sonuçlar ile karşılaştırılabilir.

REFERANSLAR (REFERENCES)

- [1] Kim, J., Suh, E., & Hwang, H., "A model for evaluating the effectiveness of CRM using the balanced scorecard", *Journal of Interactive Marketing*, 17(2), 5-19, 2003.
- [2] Kim, S. Y., Jung, T. S., Suh, E. H., & Hwang, H. S., "Customer segmentation and strategy development based on customer lifetime value: A case study", *Expert systems with applications*, 31(1), 101-107, 2006.
- [3] Jansen, S. M. H., **Customer segmentation and customer profiling for a mobile telecommunications company based on usage behavior**, A Vodafone Case Study, 2007.
- [4] Han, S. H., Lu, S. X., & Leung, S. C., "Segmentation of telecom customers based on customer value by decision tree model", *Expert Systems with Applications*, 39(4), 3964-3973, 2012.
- [5] Cowgill, M. C., Harvey, R. J., & Watson, L. T., "A genetic algorithm approach to cluster analysis", *Computers & Mathematics with Applications*, 37(7), 99-108, 1999.
- [6] Murthy, C. A., & Chowdhury, N., "In search of optimal clusters using genetic algorithms", *Pattern Recognition Letters*, 17(8), 825-832, 1996.
- [7] Krishna, K., & Murty, M. N., "Genetic K-means algorithm", *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 29(3), 433-439, 1999.
- [8] Kwedlo, W., "A clustering method combining differential evolution with the K-means algorithm", *Pattern Recognition Letters*, 32(12), 1613-1621, 2011.
- [9] Karaboga, D., & Ozturk, C., "A novel clustering approach: Artificial Bee Colony (ABC) algorithm", *Applied soft computing*, 11(1), 652-657, 2011.
- [10] Zhang, C., Ouyang, D., & Ning, J., "An artificial bee colony approach for clustering", *Expert Systems with Applications*, 37(7), 4761-4767, 2010.
- [11] Afshar, A., Haddad, O. B., Mariño, M. A., & Adams, B. J., "Honey-bee mating optimization (HBMO) algorithm for optimal reservoir operation", *Journal of the Franklin Institute*, 344(5), 452-462, 2007.
- [12] Fathian, M., Amiri, B., & Maroosi, A., "Application of honey-bee mating optimization algorithm on clustering", *Applied Mathematics and Computation*, 190(2), 1502-1513, 2007.
- [13] Cheng, Y., Jiang, M., & Yuan, D., "Novel clustering algorithms based on improved artificial fish swarm algorithm", *IEEE Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, Tianjin, Çin, 141-145, 14-16 Ağustos 2009.
- [14] Zhu, W., Jiang, J., Song, C., & Bao, L., "Clustering algorithm based on fuzzy C-means and artificial fish swarm", *Procedia Engineering*, 29, 3307-3311, 2012.
- [15] Wan, M., Li, L., Xiao, J., Wang, C., & Yang, Y., "Data clustering using bacterial foraging optimization", *Journal of Intelligent Information Systems*, 38(2), 321-341, 2012.
- [16] Saida, I. B., Nadjet, K., & Omar, B., "A new algorithm for data clustering based on cuckoo search optimization", **In Genetic and Evolutionary Computing**, Springer, International Publishing, 55-64, 2014.
- [17] Chowdhury, A., Bose, S., & Das, S., "Automatic clustering based on invasive weed optimization algorithm", **International Conference on Swarm, Evolutionary, and Memetic Computing**, Visakhapatnam, Hindistan, 105-112, 19-21 Aralık 2011.
- [18] Hatamlou, A., Abdullah, S., & Nezamabadi-Pour, H., "A combined approach for clustering based on K-means and gravitational search algorithms", *Swarm and Evolutionary Computation*, 6, 47-52, 2012.
- [19] Hatamlou, A., "Black hole: A new heuristic optimization approach for data clustering", *Information sciences*, 222, 175-184, 2013.
- [20] Olamaei, J., Mazinan, A. H., Arefi, A., & Niknam, T., "A hybrid evolutionary algorithm based on ACO and SA for distribution feeder reconfiguration", **IEEE 2nd International Conference on Computer and Automation Engineering (ICCAE)**, Singapur, 265-269, 26-28 Şubat 2010.
- [21] Diskaya, F., Emir, S., & Orhan, N., "Measuring the technical efficiency of telecommunication sector within global crisis: comparison of G8 countries and Turkey", *Procedia-Social and Behavioral Sciences*, 24, 206-218, 2011.
- [22] Hung, S. Y., Yen, D. C., & Wang, H. Y., "Applying data mining to telecom churn management", *Expert Systems with Applications*, 31(3), 515-524, 2006.
- [23] Hwang, H., Jung, T., & Suh, E., "An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry", *Expert systems with applications*, 26(2), 181-188, 2004.
- [24] Hamka, F., Bouwman, H., De Reuver, M., & Kroesen, M., "Mobile customer segmentation based on smartphone measurement", *Telematics and Informatics*, 31(2), 220-227, 2014.
- [25] Ye, L., Qiuru, C., Haixu, X., Yijun, L., & Guangping, Z., "Customer segmentation for telecom with the k-means clustering method", *Information Technology Journal*, 12(3), 409-413, 2013.
- [26] Zhao, J., Zhang, W., & Liu, Y., "Improved K-means cluster algorithm in telecommunications enterprises customer segmentation", **2010 IEEE International Conference on Information Theory and Information Security (ICITIS)**, Beijing, Çin, 167-169, 17-19 Aralık 2010.
- [27] Ahn, J. H., Han, S. P., & Lee, Y. S., "Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry", *Telecommunications Policy*, 30(10), 552-568, 2006.
- [28] Cheng, L. C., & Sun, L. M., "Exploring consumer adoption of new services by analyzing the behavior of 3G subscribers: An empirical case study", *Electronic Commerce Research and Applications*, 11(2), 89-100, 2012.
- [29] Lim, J., Nam, C., Kim, S., Lee, E., & Lee, H., "A new regional clustering approach for mobile telecommunications policy in China", *Telecommunications Policy*, 39(3), 296-304, 2015.
- [30] Chen, C. H., Chiang, R. D., Wu, T. F., & Chu, H. C., "A combined mining-based framework for predicting telecommunications customer payment behaviors", *Expert Systems with Applications*, 40(16), 6561-6569, 2013.
- [31] Vidya, N. A., Fanany, M. I., & Budi, I., "Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers", *Procedia Computer Science*, 72, 519-526, 2015.
- [32] Weiss, G., & Hirsh, H., "Learning to predict rare events in event sequences", **Fourth International Conference on Knowledge Discovery and Data Mining**, New York, A.B.D., 359-363, 27-31 Ağustos 1998.
- [33] Farvaresh, H., & Sepehri, M. M., "A data mining framework for detecting subscription fraud in telecommunication", *Engineering Applications of Artificial Intelligence*, 24(1), 182-194, 2011.
- [34] Olszewski, D., "A probabilistic approach to fraud detection in telecommunications", *Knowledge-Based Systems*, 26, 246-258, 2012.
- [35] Joseph, M. V., "Data mining and business intelligence applications in telecommunication industry", *International Journal of Engineering and Advanced Technology (IJEAT)*, 2(3), 525-528, 2013.
- [36] Chao, D. O. N. G., LEI, Z. M., & Feng, L. I. U., "Internet quality abnormal analysis with k-means clustering", *The Journal of China Universities of Posts and Telecommunications*, 18, 94-100, 2011.

- [37] Ren, D. Q., Zheng, D., Huang, G., Zhang, S., & Wei, Z., "Parallel Set Determination and K-Means Clustering for Data Mining on Telecommunication Networks", **In High Performance Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing (HPCC_EUC), 2013 IEEE 10th International Conference on, Zhangjiajie, Çin, 1553-1557, 13-15 Kasım 2013.**
- [38] Velmurugan, T., "Performance based analysis between k-Means and Fuzzy C-Means clustering algorithms for connection oriented telecommunication data", *Applied Soft Computing*, 19, 134-146, 2014.
- [39] Pakrashi, A., & Chaudhuri, B. B. "A Kalman filtering induced heuristic optimization based partitioned data clustering", *Information Sciences*, 369, 704-717, 2016.
- [40] Jain, A. K., Murty, M. N., & Flynn, P. J., "Data clustering: a review", *ACM computing surveys (CSUR)*, 31(3), 264-323, 1999.
- [41] Nanda, S. J., & Panda, G., "A survey on nature inspired metaheuristic algorithms for partitioned clustering", *Swarm and Evolutionary computation*, 16, 1-18, 2014.
- [42] Eberhart, R., & Kennedy, J., "A new optimizer using particle swarm theory", **IEEE Sixth International Symposium on Micro Machine and Human Science**, Nagoya, Japonya, 39-43, 4-6 Ekim 1995.
- [43] Kao, Y. T., Zahara, E., & Kao, I. W., "A hybridized approach to data clustering", *Expert Systems with Applications*, 34(3), 1754-1762, 2008.
- [44] Niknam, T., Firouzi, B. B., & Nayeripour, M., "An efficient hybrid evolutionary algorithm for cluster analysis", *In World Applied Sciences Journal*, 2008.
- [45] Rana, S., Jasola, S., & Kumar, R., "A boundary restricted adaptive particle swarm optimization for data clustering", *International journal of machine learning and cybernetics*, 4(4), 391-400, 2013.
- [46] Ahmadyfard, A., & Modares, H., "Combining PSO and k-means to enhance data clustering", **International Symposium on Telecommunications (IST 2008)**, Tahrán, İran, 688-691, 27-28 Ağustos 2008.
- [47] X. Cui, T.E. Potok, "Document clustering analysis based on hybrid PSO+ K-means algorithm", *Journal of Computer Sciences*, (special issue), 27-33, 2005.
- [48] Van der Merwe, D. W., & Engelbrecht, A. P., "Data clustering using particle swarm optimization", **IEEE The 2003 Congress on Evolutionary Computation (CEC'03)**, Canberra, Avustralya, 215-220, 8-12 Aralık 2003.
- [49] Niknam, T., & Amiri, B., "An efficient hybrid approach based on PSO, ACO and k-means for cluster analysis", *Applied Soft Computing*, 10(1), 183-197, 2010.
- [50] Ye, F., & Chen, C. Y., "Alternative KPSO-clustering algorithm", *Tamkang J. Sci. Eng.*, 8(2), 165-174, 2005.
- [51] Kuo, R. J., & Lin, L. M., "Application of a hybrid of genetic algorithm and particle swarm optimization algorithm for order clustering", *Decision Support Systems*, 49(4), 451-462, 2010.
- [52] Kuo, R. J., Syu, Y. J., Chen, Z. Y., & Tien, F. C., "Integration of particle swarm optimization and genetic algorithm for dynamic clustering", *Information Sciences*, 195, 124-140, 2012.
- [53] Mattison, R., **The telco churn management handbook**. Lulu.Com, 2006
- [54] Özmen, M., **Telekomünikasyon Sektöründe Müşteri Kaybı Yönetimi İçin Meta Sezgisel Tabanlı Karar Destek Sistemi**, Doktora Tezi, Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, 2017
- [55] Acuna, E., & Rodriguez, C., "The treatment of missing values and its effect on classifier accuracy", *Classification, clustering, and data mining applications*, 639-647, 2004.
- [56] Davies, D. L., & Bouldin, D. W., "A cluster separation measure", *IEEE transactions on pattern analysis and machine intelligence*, (2), 224-227, 1979.
- [57] Özdağoğlu, A., "Çok ölçütlü karar verme modellerinde normalizasyon tekniklerinin sonuçlara etkisi: COPRAS örneği", *Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 8(2), 229-255, 2013.
- [58] Alelyani, S., Tang, J., & Liu, H., "Feature Selection for Clustering: A Review", *Data Clustering: Algorithms and Applications*, 29, 110-121, 2013.

EK

Tablo 6 Örnek Müşteri Veri Kümesi
(Sample Customer Dataset)

Müşteri	K1: müşterinin ortalama fatura tutarı	K2: müşteriden elde edilen toplam geliri	K3: müşterinin son 6 ayda ortalama çağrı adedi	K4: müşterinin ortalama arama uzunluğu (sn)	K5: müşterinin ortalama paket dışı arama uzunluğu (sn)	K6: müşterinin şikâyet sayısı	K7: müşterinin taahhüdünün olup olmaması (taahhütü ise 1, değilse 2 değerini almaktadır)	K8: müşterinin kontratlı olduğu ay sayısı	K9: müşterinin ödenmemiş aylık fatura adedi	K10: müşterinin şimdiye kadar müşteri hizmetlerini arama sayısı
1	2.27	27.01	3	82	415	12	1	19	1	23
2	49.64	52	39	346	1196	13	1	15	1	7
3	30	52	39	346	1196	2	1	18	0	3
4	30	52	39	346	1196	5	1	13	0	1
5	14.45	8.44	6	1338	43	20	1	7	2	25
6	51.78	52	39	346	1196	13	1	4	2	20
7	10.54	14.53	71	93	157	11	1	20	2	17
8	30	52	39	346	1196	1	1	17	0	3
9	6.52	5.55	28	190	166	18	1	13	1	13
10	39.84	52	39	346	1196	9	2	27	1	7
11	8.37	91.74	7	59	1135	8	2	29	1	37
12	17.22	93.8	6	352	1009	23	2	28	1	21
13	4.27	3.12	2	346	198	12	2	28	1	30
14	25.65	316.83	229	155	3835	13	2	10	1	18
15	13.36	115.72	46	192	2685	8	2	8	1	10