

Büyükşehir Belediyesi Çağrı Merkezi Verilerinin Kümeleme Analizi ile İncelenmesi

Ezel Özkan¹, Selen Avcı^{*2}, Zerrin Aladağ³

^{*1}Kocaeli Üniversitesi Mühendislik Fakültesi Endüstri Mühendisliği, KOCAELİ

²Kocaeli Üniversitesi Mühendislik Fakültesi Endüstri Mühendisliği, KOCAELİ

³Kocaeli Üniversitesi Mühendislik Fakültesi Endüstri Mühendisliği, KOCAELİ

(Alınış / Received: 18.07.2019, Kabul / Accepted: 26.12.2019, Online Yayınlanma / Published Online: 31.12.2019)

Anahtar Kelimeler

Belediye,
Çağrı Merkezi,
K- ortalamalar Yöntemi,
Kümeleme Analizi

Öz: İstatistiksel veri analizinde en önemli konulardan biri büyük verilerdir. Büyük veriler arasından faydalı bilgiye ulaşma süreci veri madenciliği olarak adlandırılır. Veri madenciliği yöntemlerinden biri olan kümeleme analizinde, çeşitli algoritmalar ile birbirine benzer veriler aynı kümeye, birbirinden farklı olan veriler ise farklı kümelere dâhil edilmek istenir. Hiyerarşik kümeleme ve k-ortalamar yöntemleri kümeleme analizinde en yaygın kullanılan algoritmalarlardır. Bu çalışmada, Türkiye'nin nüfus yoğunluğu oldukça fazla olan bir büyükşehir belediyesinin beş büyük ilçesine ait çağrı merkezi verileri, k-ortalamar yöntemi ile "ilişkili olduğu birim", "başvuru tipi", "başvuru ilçesi", "eğitim durumu", "cinsiyet", "yaş" ve "anlık çözüm" olmak üzere yedi farklı değişkene göre kümelendiği görülmüştür. Analiz için veri madenciliği paket programlarından SPSS Clementine ve WEKA kullanılmış olup elde edilen sonuçlar karşılaştırılarak yorumlanmıştır.

Investigation of Data of a Metropolitan Municipality Call Center by Cluster Analysis

Keywords

Call Center,
Clustering Analysis,
K-means Method,
Municipality

Abstract: One of the most important issues in statistical data analysis is big data. The process of accessing useful information from big data is called data mining. In clustering analysis, which is one of the data mining methods, via different algorithms, it is desirable to include similar data in the same cluster and divergent data in different clusters. Hierarchical clustering and k-means methods are the most commonly used algorithms in clustering analysis. In this study, call center data of five districts in a metropolitan municipality with a high population density in Turkey were examined. Call center data were clustered according to seven different variables by k-means method. These variables are respective department, application type, application district, education level, gender, age and instant solution. SPSS Clementine and WEKA which are the data mining package programs were used for the analysis and the results were compared and interpreted.

Endüstri Mühendisliği Bölümü, Kocaeli Üniversitesi, Kocaeli
^{*}İlgili yazar, email: selenavciem@gmail.com

1. Giriş

Çağrı merkezleri, büyük miktarda bilgiyi telefonla almak ve iletmek amacıyla kullanılan merkezi ofislerdir. Müşteriler, çağrı merkezleri aracılığıyla sorunlarını raporlayabilmekte veya sunulan ürün ya da hizmetler hakkında bilgi isteyebilmektedir. Bu bağlamda çağrı merkezleri, şirketler ile müşteriler arasında ara yüz görevi görür [1]. Çağrı merkezleri günümüzde, kuruluşların müşterilerle doğrudan iletişim kurabildiği bir kanal olduğundan müşteri desteğinin en temel parçalarından biridir. "Telefon bankacılığı", "bilet rezervasyonu", resmi kuruluşların "şikâyet talep merkezleri" çağrı merkezlerine örnek olarak verilebilir [2].

Çalışmada, Türkiye'nin yüksek nüfuslu bir büyükşehir belediyesinin beş ilçesine ait verilerle kümeleme analizi gerçekleştirilmiştir. En yaygın bilinen veri madenciliği uygulamalarından biri olan kümeleme analizi, büyük ölçekli verilerle başa çıkmada önemli bir rol oynamaktadır. Kümeleme analizinde amaç, küme içinde birbirine benzer verileri toplamak, farklılıkları olan veriler için de birbirinden farklı kümeler oluşturmaktır [3].

Çalışmada veriler; çağrı merkezine başvuran kişilerin ilişkili olduğu birim, başvuru ilçesi, başvuru tipi, cinsiyet, soruna çözüm bulunup bulunmadığı, yaş ve eğitim durumu olmak üzere yedi değişkenden oluşmaktadır. Uygulamada, veri madenciliği programları olan SPSS Clementine ve WEKA ile k-ortalamlar analizi yapılmıştır. Takip eden bölümde kümeleme analizi anlatılmıştır. Üçüncü bölümde analizler yapılmış ve son bölümde elde edilen sonuçlar yorumlanmıştır.

Aşağıda sıralanan çalışmalar; literatürde farklı karar ortamları için kümeleme analizinin sıklıkla kullanıldığını göstermektedir. Ancak incelenen çalışmalarda çağrı merkezi verileri ile yapılmış olan bir çalışmaya rastlanılmamıştır. Çelik ve diğ. (2019), Türkiye'de 2016 yılında 81 ile ait 11 tarım değişkenini kullanarak, benzer yapıyı gösteren homojen il gruplarının belirlenmesi ve Türkiye'de benzer iller grubunun ortaya çıkarılması amacıyla "K ortalamlar kümeleme" ve "Hiyerarşik kümeleme" yöntemlerini kullanmışlardır [4]. Akay (2019) yapmış olduğu çalışmada Türkiye'de kütüphane kullanımında benzer özellik gösteren illeri belirlemek ve illerin insani gelişmişlikleri arasındaki ilişkiyi belirlemek için Ward yöntemini kullanmıştır [5]. Ünlü ve Atik (2019), AB'ye aday ülke konumunda olan Türkiye'nin bu ülkeler karşısındaki göreceli Endüstri 4.0 performansını belirlemek için Ward yöntemini kullanmışlardır [6]. Almeida ve diğ. (2018), nehirlerin zarar görmesinden kaynaklanan balık türlerindeki değişimleri analiz etmek için kümeleme analizini kullanmışlardır [7]. Hernandez ve diğ. (2018) yılında yapmış oldukları çalışmada American Thoracic Society (ATS) tarafından önerilen 225 alerjik astımlı hasta için 19 değişken belirleyerek hastaların hiyerarşik kümeleme analizi ile kümelendiğini amaçlamışlardır [8]. Horiuchi ve diğ. (2018) AHF patofizyolojisini geliştirmek, klinik olarak önemli alt grupları tespit etmek ve klinik olarak karar vermeyi kolaylaştırmak için kümeleme analizini kullanmışlardır [9]. Özari ve Eren (2018) Türkiye İstatistik Kurumu (TÜİK) tarafından 2016 yılında "illerde yaşam endeksi" olarak adlandırılan 11 farklı boyut ve yaklaşık 41 alt göstergeden oluşan verileri çok boyutlu ölçekleme yöntemi ve k-ortalamlar yöntemi ile kümelere ayırmışlardır [10]. Gazel ve Akel (2018), Borsa İstanbul'da (BIST) farklı sektör endeksleri içinde yer alan hisse senetlerinin, geçmiş fiyat verilerine göre kümeleme analizine tabi tutulduğunda aynı sektör içerisinde yer alıp almadığını belirlemek için hiyerarşik yığınsal kümeleme analizini kullanmışlardır [11]. Çelik ve Kırıl (2018), Türkiye'de illerin konut satışlarını etkileyen faktörleri dengeli panel veri analizi ve Ward yöntemi ile incelenmişlerdir [12]. Sarıışık ve Öğütü (2018), deneylerden elde ettikleri verilerle doğal taşların kesme kuvveti, spesifik kesme enerjisi ve spesifik enerjilerini K ortalamlar kümeleme yöntemiyle incelemişlerdir [13]. Javadi ve diğ. (2017) İran'da yer altı su tabakalarını, belirledikleri parametreleri kullanarak k-ortalamlar kümeleme analiziyle incelemişlerdir [14]. Ceylan ve diğ. (2017) yılında yapmış oldukları çalışmada Türkiye'de faaliyet gösteren bir emeklilik şirketinin müşteri profilini değerlendirmek için iki aşamalı kümeleme analizi yapmışlardır [15]. Ülgen ve Özalp (2017), refah rejimlerini kümelemek için k-ortalama tekniğini kullanarak, 26 ülke ve 11 farklı değişken açısından analiz etmişlerdir [16]. Çalış ve Baynal (2016), k ortalamlar yöntemi ile Türkiye'de faaliyet gösteren bir banka şubesinin iki yüz müşterisini on iki farklı değişkene göre k-ortalamlar kümeleme analizi ile kümelendiğini amaçlamışlardır [17]. Dong ve diğ. (2016), kümeleme analiziyle sayısal hava tahmini bilgilerini kullanarak yeni bir model önermeyi amaçlamışlardır [18]. Giray (2016), iki aşamalı kümeleme analiziyle bir cezaevinin son 10 yıldaki hükümlü verilerinin temel bilgilerinin analizini yaparak benzer özelliklere sahip hükümlülere ait alt grupları değerlendirmeyi amaçlamıştır [19]. Turan ve diğ. (2016) 1980 ve 2013 yıllarında Ortadoğu ülkeleri ve Türkiye'nin ekonomik yapılarını üç farklı döneme ayırarak benzer olup olmadığını belirlemek için kümeleme analizi yöntemini kullanmışlardır [20]. Canıkalp ve Ünlükaplan (2016), geçiş ekonomisi ülkelerinin 2014 yılı itibarıyla yönetim kalitelerini değerlendirmek için Ward yöntemini kullanmışlardır [21]. Aydın ve Seven (2015), Türkiye'deki İl Nüfus ve Vatandaşlık Müdürlüklerini, iş yoğunluklarına göre hibrid hiyerarşik k-ortalamlar kümeleme analizi ile sınıflandırılmışlardır [22]. Tekin (2015), 2013 yılına ait 16 farklı sağlık göstergesi ile Türkiye'deki 81 ili Ward yöntemiyle kümelemeyi amaçlamıştır [23]. Yukarıda açıklanan çalışmalar, Tablo 1'de özet olarak sunulmuştur.

Tablo 1. 2015-2019 yılları arası kümeleme analizi literatür özeti

Yazar (Yıl)	Hiyerarşik Kümeleme Analizi	Hiyerarşik Olmayan Kümeleme Analizi
Çelik vd. (2019)	Hiyerarşik kümeleme analizi	K-ortalamlar
Akay (2019)	Ward yöntemi	-
Ünlü ve Atik (2019)	Ward yöntemi	-
Almeida vd. (2018)	-	K-ortalamlar
Hernandez vd. (2018)	Hiyerarşik kümeleme analizi	-
Horiuchi vd. (2018)	-	K-ortalamlar
Özari ve Eren (2018)	-	K-ortalamlar
Gazel ve Akel (2018)	Hiyerarşik yığınsal kümeleme analizi	-
Çelik ve Kiral (2018)	Ward yöntemi	-
Sarışık ve Öğütlü (2018)	-	K-ortalamlar
Javadi vd. (2017)	-	K-ortalamlar
Ceylan vd. (2017)	Ward yöntemi	K-ortalamlar
Ülgen ve Özalp (2017)	-	K-ortalamlar
Çalış ve Baynal (2016)	-	K-ortalamlar
Dong vd. (2016)	-	K-ortalamlar
Giray (2016)	Ward yöntemi	K-ortalamlar
Turan vd. (2016)	-	K-ortalamlar
Canıkalp ve Ünlükaplan (2016)	Ward yöntemi	-
Aydın ve Seven (2015)	-	Hibrid hiyerarşik k-ortalamlar kümeleme analizi
Tekin (2015)	Ward yöntemi	-

2. Materyal ve Metot

Kümeleme analizi, elde bulunan verilerin birbirlerine olan benzerliklerine göre “küme” adı verilen gruplara ayrılmasıdır. Oluşturulan kümelere veriler arasındaki uzaklıkların ölçülmesiyle iki ya da daha fazla küme oluşturulabilir. Literatürde incelenen makalelerde en çok kullanılan uzaklık ölçüleri dikkate alınarak, kümelere bulunacak birim ya da birimlerin arasındaki minimum ve maksimum benzerlik uzaklık ölçütlerinden Öklid uzaklığı Denklem 1’de, Minkowski uzaklığı Denklem 2’de, City-Block (Manhattan) uzaklığı Denklem 3’te, Canberra uzaklığı Denklem 4’te, Tchebyshev uzaklığı Denklem 5’te ve Mahalanobis uzaklığı Denklem 6’da verilmiştir [24, 3]. Denklemlerde x_{ik} i. birimin k. değişken değerini, x_{jk} j. birimin k. değişken değerini ve S p x p tipinde örneklem kovaryans matrisini ifade etmektedir.

$$d_{ij} = \sqrt{\sum_{i=1}^p (x_{ik} - x_{jk})^2} \quad (1)$$

$$d_{ij} = \left[\sum_{k=1}^p |x_{ik} - x_{jk}|^q \right]^{1/q} \quad (2)$$

$$d_{ij} = \sum_{k=1}^p |x_{ik} - x_{jk}| \quad (3)$$

$$d_{ij} = \sum_{k=1}^p |x_{ik} - x_{jk}| / x_{ik} + x_{jk} \quad (4)$$

$$d_{ij} = \max_{i:1,2,..,n} |x_{ik} - x_{jk}| \quad (5)$$

$$Md_{ik} = (x_i - x_j)' S^{-1} (x_i - x_j) \quad (6)$$

Kümeleme algoritmaları genel olarak hiyerarşik kümeleme analizi ve hiyerarşik olmayan kümeleme analizi olarak iki gruba ayrılır [25]. Alt başlıklarda bu konulara kısaca değinilmiştir.

2.1. Hiyerarşik kümeleme analizi

Hiyerarşik kümeleme analizinde, küme ağacı hiyerarşik yöntemler ile oluşturulur. Hiyerarşik yöntemlerde, sonuçlar dendogram adı verilen ağaç diyagramları ile gösterilir [26].

Hiyerarşik yöntemler, gruplayıcı ve bölünür kümeleme algoritmaları olarak iki gruba ayrılır. [9]. Başlangıç veri tabanındaki her noktayı küme varsayıp bu kümeleri birleştirerek yeni kümeler oluşturulmasına gruplayıcı kümeleme algoritmaları denir. Başlangıç veri tabanındaki noktaların tamamının tek bir küme olarak varsayılması ile birbirinden farklı olan noktaların atılarak k adet küme oluşturulmasına ise bölünür kümeleme algoritmaları denir [24].

2.2. Hiyerarşik olmayan kümeleme analizi

Küme sayılarının daha önceden belirlenmiş olduğu durumlarda hiyerarşik olmayan kümeleme yöntemleri kullanılmaktadır. Hiyerarşik olmayan kümeleme yöntemleri, hiyerarşik tekniklere göre daha hızlıdır ve birim sayısı büyük ise daha anlamlıdır. Hiyerarşik olmayan kümeleme analizinde en yaygın kullanılan yöntem k-ortalamlar (k-means) yöntemidir [25, 27].

2.2.1. K-ortalamlar (k-means) kümeleme yöntemi

K ortalama kümeleme analizi, ilk olarak MacQueen (1967) tarafından kullanılan denetimsiz bir modelleme yöntemidir. Bu yöntemde birkaç nesne, orta noktaya ya da en yakın merkeze göre mevcut kategoriler veya koşullar temel alınıp bölümlere ayrılarak oluşturulur. Bu yöntem aynı kümedeki veriler arasındaki farklılığı en aza indirir ve farklı kümelerdeki veriler arasındaki farklılığı en üst düzeye çıkarır [28, 29].

K-ortalamlar kümeleme yöntemi adımları aşağıdaki gibidir [26, 30]:

1. Veri seti rassal olarak k adet başlangıç kümesine ayrılır veya merkez noktalar tüm nesnelerin ortalaması alınarak belirlenir.
2. Veri setindeki birimler, merkezi kendisine en yakın olan kümeye atanır.
3. Küme merkezi her atamadan sonra tekrar hesaplanır.
4. Tüm birimler kümelere atanıncaya kadar 2. ve 3. adımlar tekrarlanır.

K-ortalamlar yönteminde en az küme sayısı (2 küme) ile başlanarak küme sayısının birer artırılması sonucunda en uygun küme sayısı belirlenmeye çalışılır [31]. Küme sayısını belirlemek için çeşitli yöntemler vardır. Bunlardan en çok bilinen eşitlik yöntemi Denklem 7'de, Silhouette indeksi Denklem 8'de, Mariott ölçütü Denklem 9'da, Wilk's Lamda ölçütü Denklem 10'da verilmiştir [32].

$$k = \left(\frac{n}{2}\right)^{1/2} \quad n: \text{birim sayısı} \quad (7)$$

$$sil(i) = \frac{b(i) - a(i)}{\max(a(i) - b(i))} \quad (8)$$

$a(i)$: i.birimin kendi kümesindeki tüm noktalara olan ortalama uzaklıkları

$b(i)$: i.birimin diğer kümelerdeki tüm noktalara olan ortalama uzaklıkların minimumu

$$M = k^2|W|$$

W: Grup içi kareler toplamı matrisi

(9)

M değerini minimum yapan k değeri küme sayısını gösterir.

$$\Delta = \frac{|W|}{|W + B|} = \frac{|W|}{|T|}$$

T: Çarpımlar ve kareler toplam matrisi

(10)

B: Gruplar arası çarpımlar ve kareler toplam matrisi

W: Grup içi çarpımlar ve kareler toplam matrisi

Yukarıda verilen yöntemlere ek olarak Lewis ve Thomas yaklaşımına göre, kümeler toplam varyasyon %80'ini açıklamalı ve varyansta %5' e kadarlık bir artış durumunda yeni bir küme ilave edilmelidir. Elbow (Dirsek) metoduna göre ise kümeler, küme içi kareler toplamına göre belirlenir.

3. Yöntem ve Bulgular

Çalışmada bir büyükşehir belediyesinin çağrı merkezini arayan kişilerin değerlendirilmesine yönelik veri madenciliği uygulaması gerçekleştirilmiştir. Uygulamada veri madenciliği yöntemlerinden kümeleme analizi ele alınmıştır. Kümeleme analizi ile çağrı Merkezinin son beş aylık verilerinin yedi değişkene göre kümeleneceği amaçlanmıştır.

Uygulamada kullanılan 3860 çağrı merkezi verisi, önemli nüfus yoğunluğuna sahip bir büyükşehir belediyesinin beş büyük ilçesine aittir. Analizde veri madenciliği uygulamaları için geliştirilmiş bütünsel bir görsel modelleme gereği olan SPSS Clementine ve WEKA programları kullanılmıştır.

3.1. Verilerin hazırlanması

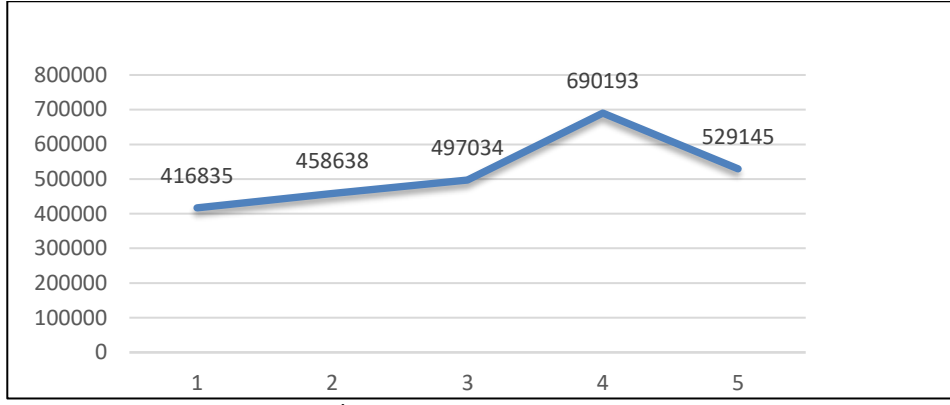
Veri hazırlama aşamasında, beş aylık dönemde büyükşehir belediyesi çağrı merkezine başvuran toplam 58.735 kişiye ait veriler elde edilmiştir. Verilerin çok fazla olmasından dolayı analiz için, nüfusu 400.000'den fazla, konum olarak birbirine yakın ve ilçede ikamet eden bireylerin ortalama eğitim düzeyinin yüksek olduğu beş ilçe seçilmiştir.

Veri birleştirme ve temizleme aşamasında arayan kişilerin ilişkili olduğu birim, başvuru ilçesi, başvuru tipi, cinsiyet, çözüm olup olmadığı, yaş ve eğitim durumu bilgileri baz alınarak veriler birleştirilmiştir. Eksik ve hatalı verilerin temizlenmesi sonucunda toplam olarak çağrı merkezine gelen 3860 adet çağrı için yedi değişkenden oluşan 3860*7' lik bir veri kümesi oluşturulmuştur.

Üçüncü ve son aşamada veri dönüştürme yapılmıştır. Veriler aşağıdaki gibi kategorik hale getirilmiştir.

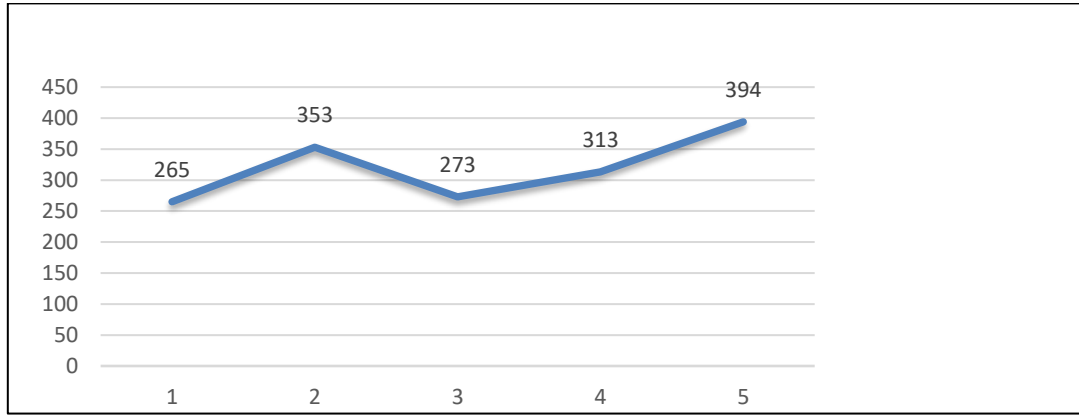
- İlişkili olduğu birim: Bu değişken, bütün birimler dikkate alındığında "bakım ve onarım faaliyetleri", "çevre", "denetim", "sağlık" ve "sosyal hizmetler, spor ve kültürel hizmetler" ve "ulaşım" olmak üzere altı farklı kategoriye ayrılmıştır.
- Başvuru tipi: Başvuru tipi değişkeni "bilgi istemi", "talep" ve "şikâyet" olmak üzere üç gruba ayrılmıştır.
- Başvuru ilçesi: İkamet eden bireylerin ortalama eğitim düzeyinin en yüksek olduğu beş ilçe ele alınmıştır.
- Eğitim Durumu: Çağrı merkezine başvuran kişilerin eğitim durumu "ilkokul", "ortaokul", "lise", "ön lisans", "lisans", "yüksek lisans" ve "doktora" olmak üzere yedi gruba ayrılmıştır.
- Cinsiyet: Arayan kişiler kadın ve erkek olarak sırasıyla "1" ve "0" olarak tanımlanmıştır.
- Yaş: Çağrı merkezine başvuran kişilerin yaşları dikkate alınarak veriler altı anlamlı grupta kategorize edilmiştir. Bu gruplar; "16-25 yaş arası", "26-34 yaş arası", "35-44 yaş arası", "45-54 yaş arası", "55-64 yaş arası" ve "65 yaş ve üstü" olarak oluşturulmuştur.
- Anlık çözüm: Veriler "anlık çözüm var" ve "anlık çözüm yok" olmak üzere iki gruba ayrılmıştır.

Şekil 1'de ilçelerin nüfus dağılımları verilmiştir [33]. Nüfus yoğunluğu bakımından 400.000 in üzerindeki ilçeler için çağrı merkezine gelen başvurular analiz kapsamına alınmıştır.



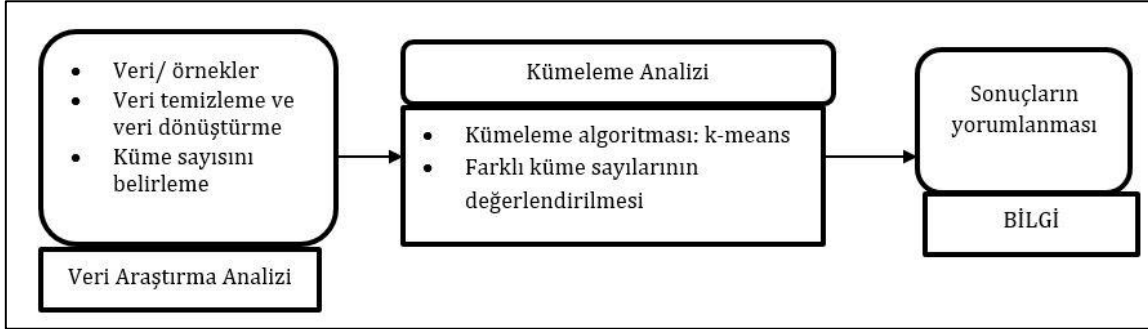
Şekil 1. İlçelerin 2018 yılı nüfus dağılımları

Şekil 2’de çağrı merkezine başvuran kişilerin lisans ve lisansüstü eğitim durumları gösterilmiştir. Başvurularda 5 numaralı ilçenin lisans ve lisansüstü eğitim durumuna göre en yüksek ilçe olduğu belirlenmiştir.



Şekil 2. İlçelere göre lisans ve üzeri eğitim düzeyinde başvuran kişi sayısı

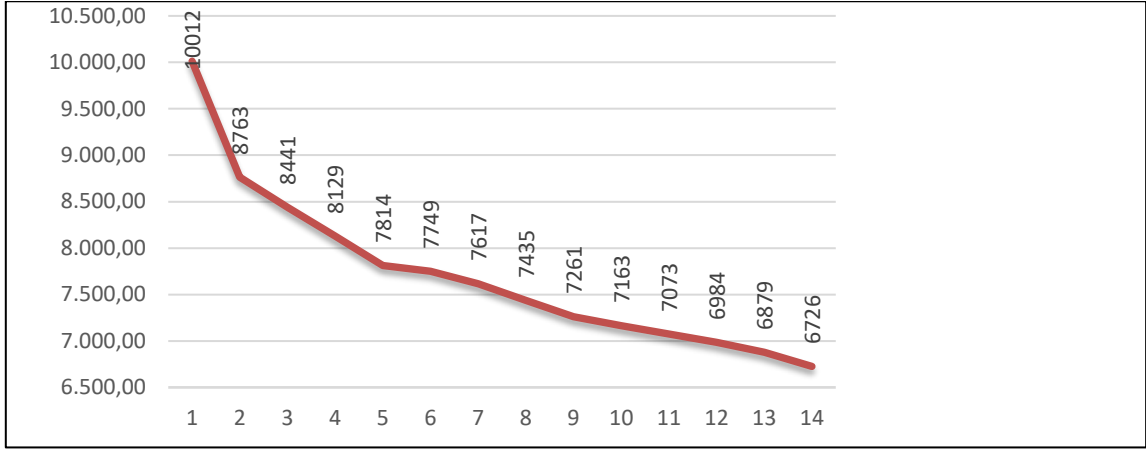
Şekil 3’te ise analiz için kullanılan modelin kavramsal yapısı verilmektedir.



Şekil 3. Analiz için kullanılan modelin kavramsal yapısı

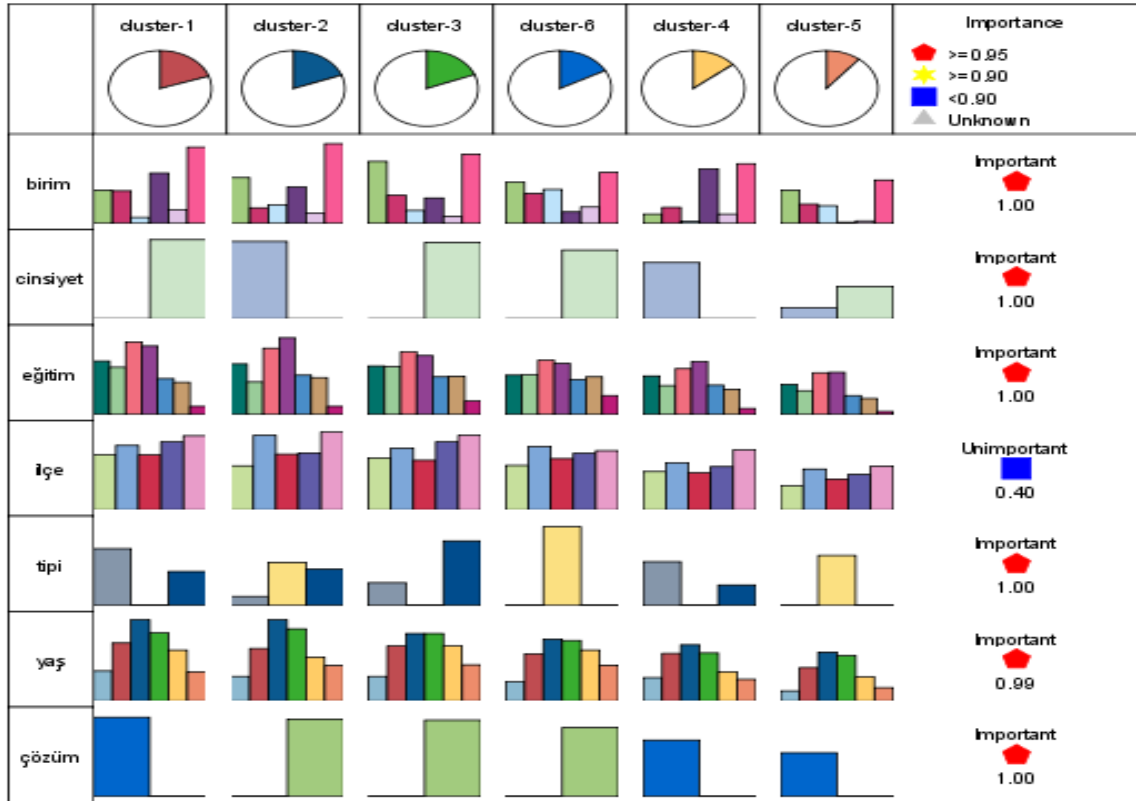
3.2. K - ortalamalar yöntemi ile kümeleme analizi

Çalışmada, çağrı merkezi verilerine ait küme sayısını belirlemek için iki yöntem kullanılmıştır. Analiz kapsamındaki 3860 adet kayıt için; küme sayısı; $k = \left(\frac{n}{2}\right)^{1/2}$ eşitliği kullanıldığında; $k = [2, 43]$ aralığında bulunmuştur. Ancak verilerin ve küme sayısının oldukça fazla olması sebebiyle; veri kümesinde bir dizi k değeri için kümeleme yapılarak her k değeri için küme içi kareler toplamını bulan Elbow Metodu tercih edilmiştir [34]. Küme sayıları-hata kareler toplamını gösteren Şekil 4 incelendiğinde, küme sayısının 6 değeri ve sonrası için grafiğin yatay olarak seyrettiği görülmektedir. Grafiğe göre en uygun küme sayısının altı veya dokuz olduğu sonucuna varılmıştır. Dokuz küme sayısı ile kümeleme analizi yapıldığında değişkenlerin önem dereceleri değişmediğinden altı küme sayısı uygun görülmüştür.



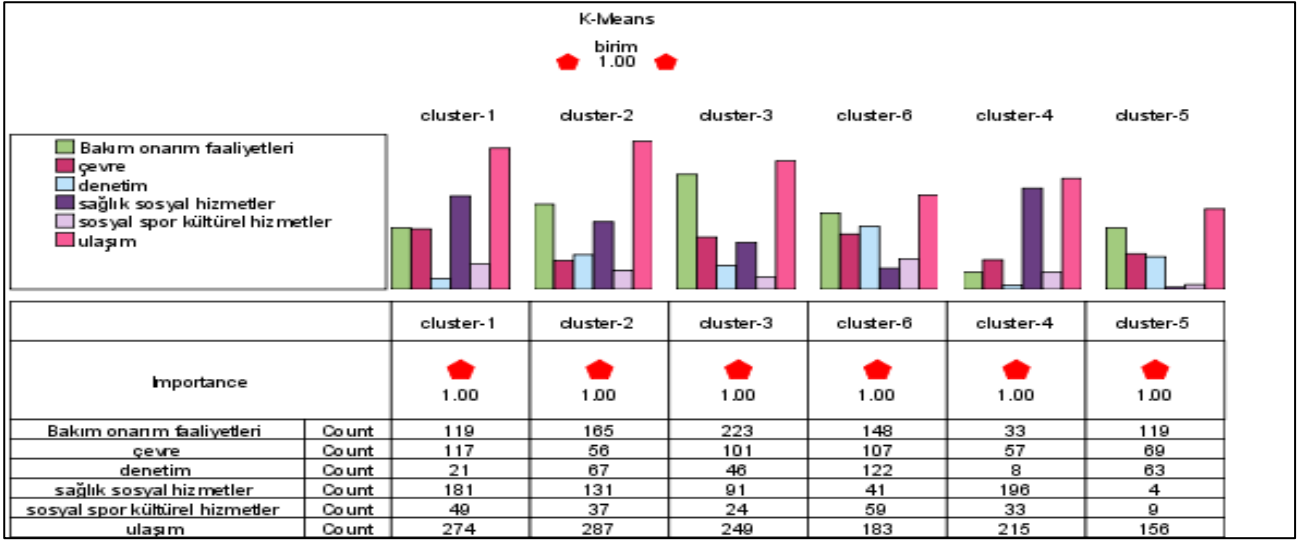
Şekil 4. Küme sayılarına göre hata kareleri toplamları

K-ortalamlar yöntemiyle oluşturulan kümeler ve kümelerin değişkenler üzerindeki etkileri Şekil 5’ te gösterilmiştir. Clementine’de önem dereceleri 0.90’ ın üzerinde olan değişkenlerin kümelere etkilerinin önemli olduğu vurgulanmaktadır [17]. Şekil 5’e göre; birim, cinsiyet, eğitim, başvuru tipi, yaş ve çözüm değişkenlerinin altı küme üzerindeki etkilerinin önemli olduğu sonucu ortaya çıkmıştır. İlçe değişkeninin ise önemsiz olduğu görülmektedir. 3860 çağrının kümelere dağılımı; 761, 743, 734, 542, 420 ve 660 şeklindedir.



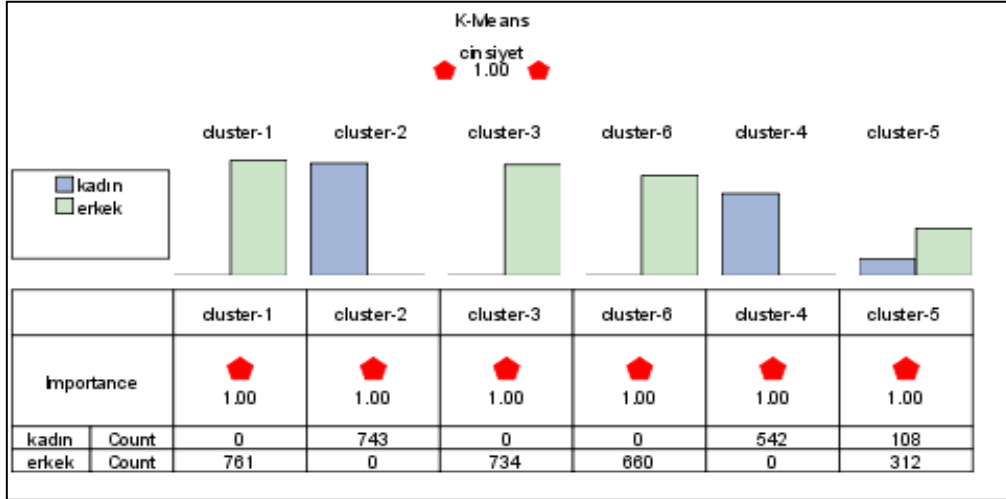
Şekil 5. K- Ortalamalar yöntemi ile elde edilen kümeler

İlk olarak “başvuru” birim değişkeninin kümeler üzerindeki etkisi dikkate alındığında, bu değişkenin Şekil 6’da görüldüğü üzere altı küme için önemli olduğu sonucuna ulaşılmıştır. Başvurular incelendiğinde en çok başvurunun ulaşım birimine yapıldığı ve bunu bakım ve onarım faaliyetlerinin izlediği görülmüştür. Buna göre ikinci kümedeki kişilerin % 22,20’si bakım ve onarım faaliyetleri birimine başvuruda bulunmuşken, dördüncü kümedeki kişilerin % 39,66’sı ulaşım birimine başvuru yapmıştır.



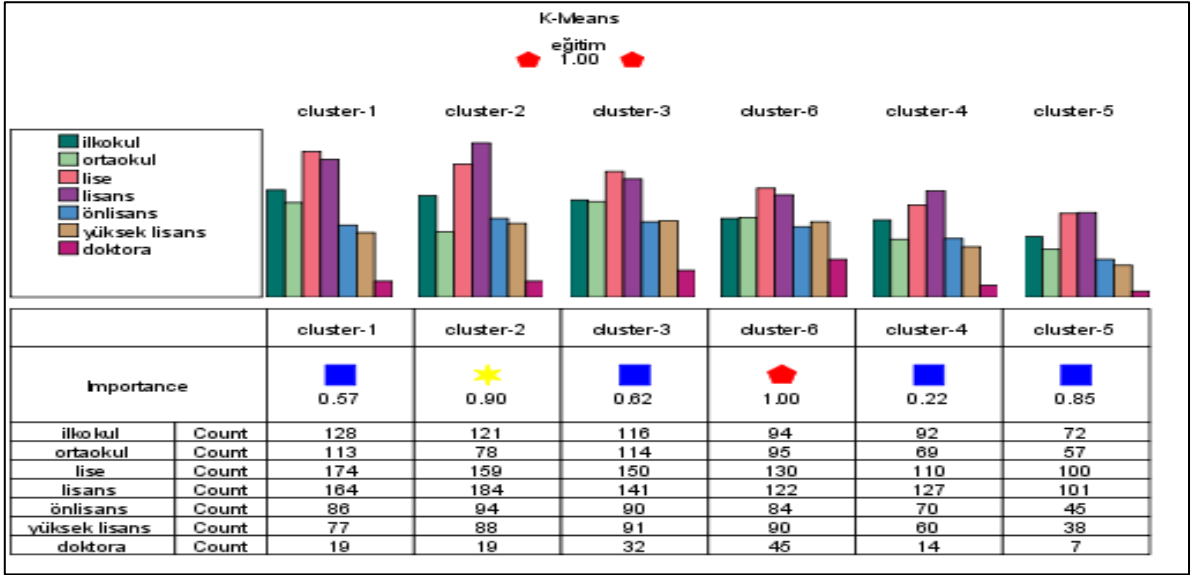
Şekil 6. Başvuru birim değişkeninin kümeler üzerindeki etkisi

Şekil 7’de cinsiyet değişkeninin altı küme için önemli olduğu görülmektedir. Birinci, üçüncü ve altıncı kümelerdeki başvuruların tamamı erkek bireylerden, ikinci ve dördüncü kümelerde ise başvuruların tamamının kadın bireylerden geldiği görülmektedir. Beşinci kümede başvuruların % 25,71’ini kadınlar oluşturmaktadır.



Şekil 7. Cinsiyet değişkeninin kümeler üzerindeki etkisi

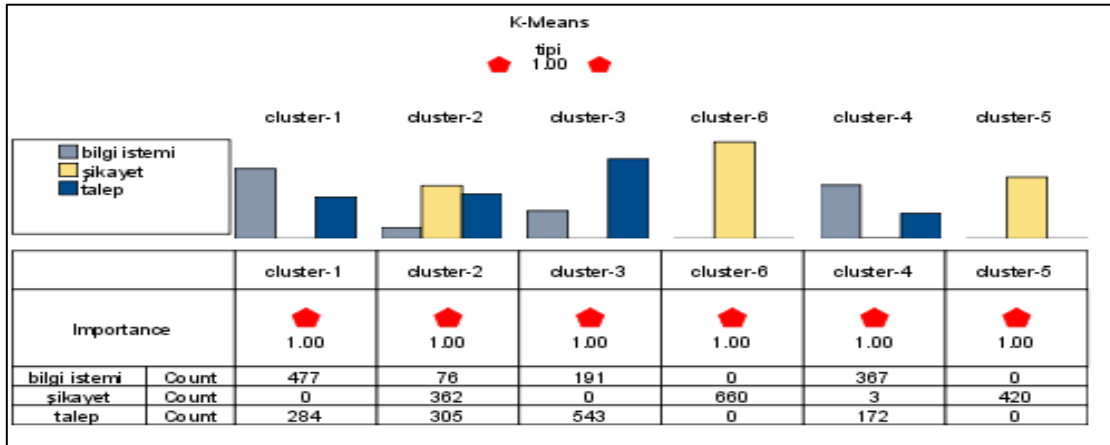
Kümeler eğitim değişkeni açısından incelendiğinde, bu değişkenin ikinci ve altıncı kümeler için önemli olduğu Şekil 8’de görülmektedir. Üçüncü kümedeki kişilerin %19,20’si lisans mezunu iken altıncı kümenin %19,69’u lise mezunu bireylerden oluşmuştur.



Şekil 8. Eğitim değişkeninin kümeler üzerindeki etkisi

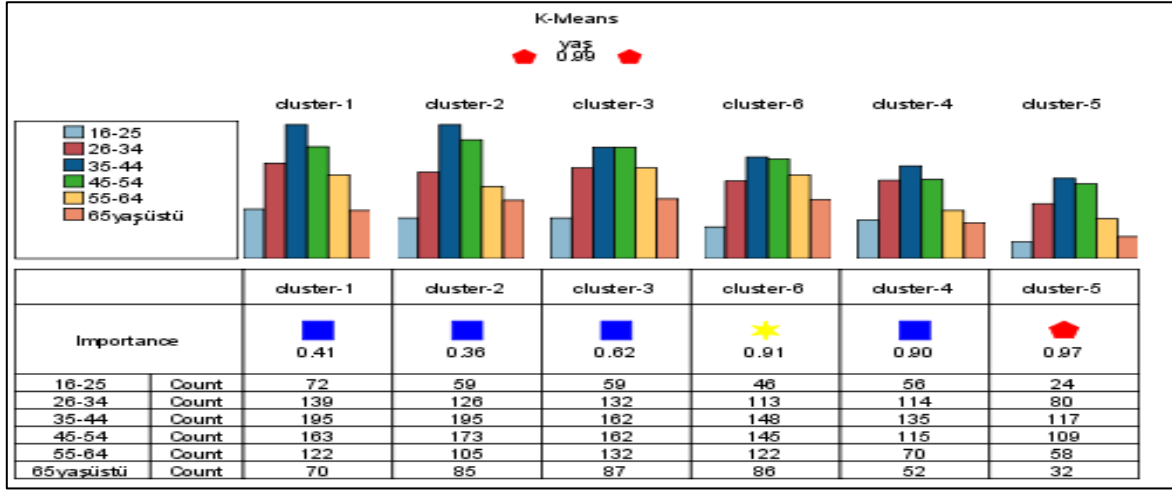
İlçe değişkeninin kümeler üzerindeki etkisi dikkate alındığında, bu değişkenin altı küme için ayırt edici bir niteliğe sahip olmadığı sonucuna ulaşılmıştır.

İstenilen hizmet tipi değişkeninin altı küme için de önemli olduğu Şekil 9'da görülmektedir. Altıncı ve beşinci kümenin tamamı şikâyet için başvuruda bulunurken, dördüncü kümenin %67,71'i bilgi istemi için başvuruda bulunmuştur. Üçüncü kümenin %73,97'sini yeni talepler oluşturmaktadır.



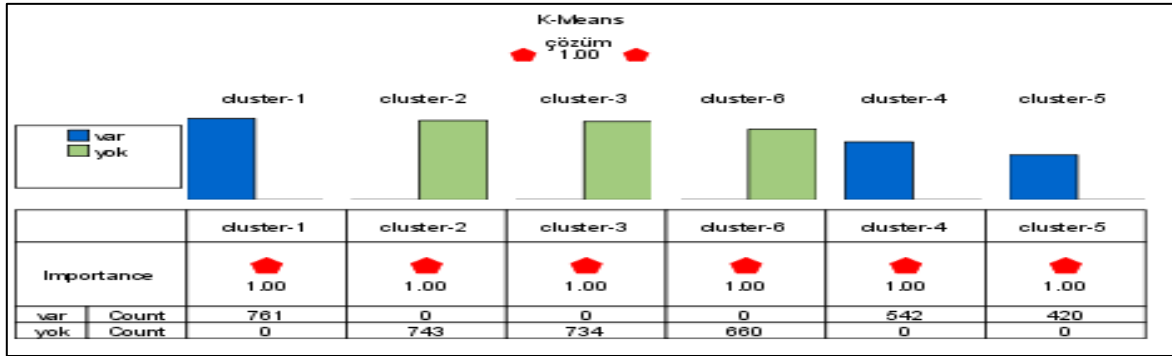
Şekil 9. Başvuru tipi değişkeninin kümeler üzerindeki etkisi

Yaş değişkeninin, beşinci ve altıncı kümeler içinde önemli olduğu Şekil 10'da görülmektedir. İkinci kümenin %26,24'ünün 35-44 yaş aralığında olduğu, beşinci kümenin ise %25,95'inin 45- 54 yaş aralığında olduğu sonucu ortaya çıkmıştır.



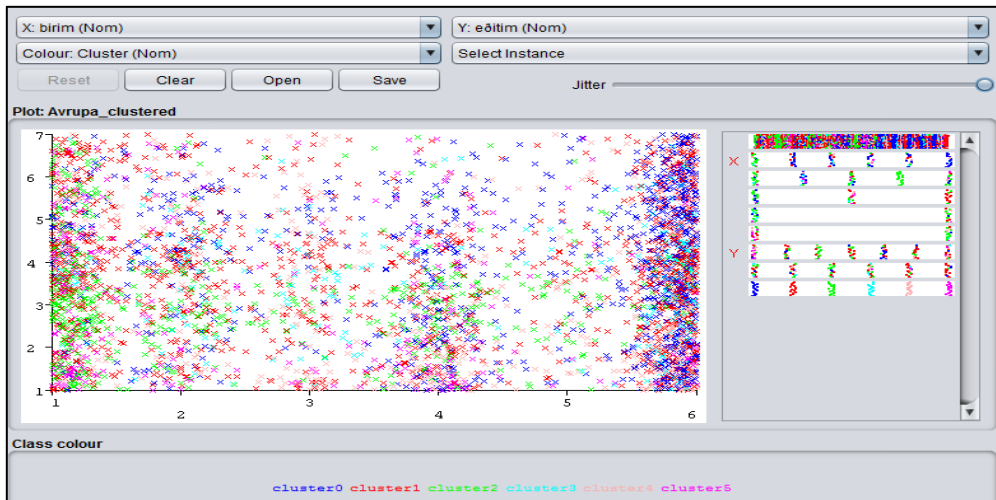
Şekil 10. Yaş değişkeninin kümeler üzerindeki etkisi

Anlık çözüm değişkeninin ise tüm kümeler için önemli olduğu Şekil 11’de görülmektedir. İkinci, üçüncü ve altıncı kümelerde anlık çözümün olmadığı, birinci, dördüncü ve beşinci kümenin tamamında ise anlık çözümün olduğu sonucuna ulaşılmıştır.



Şekil 11. Anlık çözüm değişkeninin kümeler üzerindeki etkisi

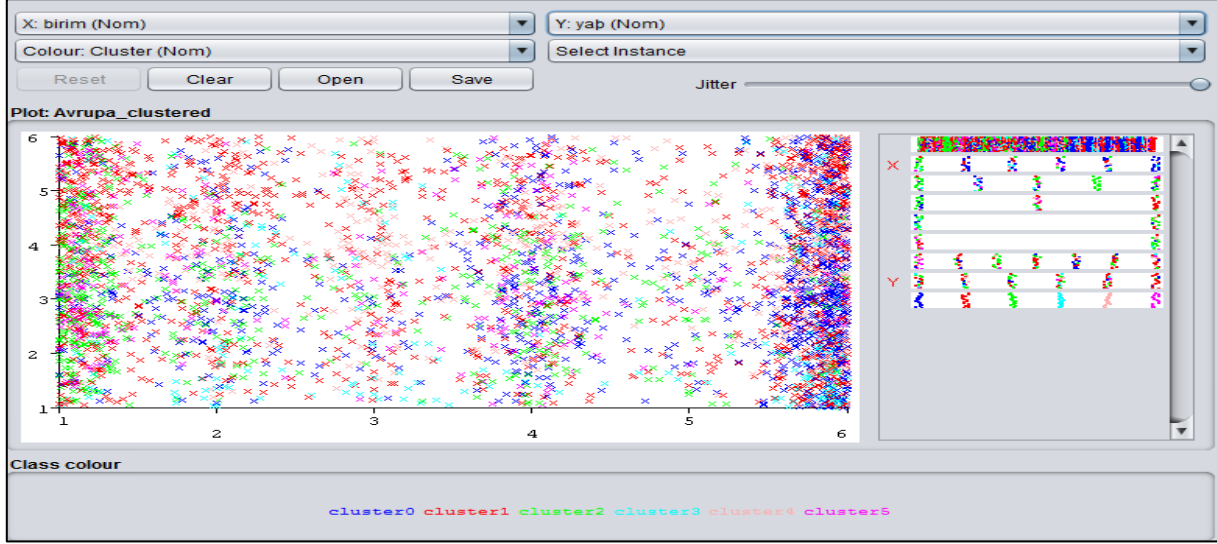
WEKA yazılımı desteğiyle yapılan analizde ise 3860 çağrının altı kümeye dağılımı sırasıyla 973, 1174, 557, 270, 606, 280 şeklinde olmuştur. WEKA programı ile de incelenen birim değişkeni aynı sayıdadır. Değişkenler bu kez yazılım gereklerinden dolayı bakım ve onarım faaliyetleri (1), çevre (2), denetim (3), sağlık ve sosyal hizmetler (4), spor ve kültürel hizmetler (5) ve ulaşım (6) olarak kodlanmıştır. Yaş değişkeni altı anlamlı grupta kategorize edilmiştir. Bu gruplar; 16-25 yaş arası (1), 26-34 yaş arası (2), 35-44 yaş arası (3), 45-54 yaş arası (4), 55-64 yaş arası (5) ve 65 yaş ve üstü (6) olarak oluşturulmuştur. Çağrı merkezine başvuran kişilerin eğitim durumu ilkököl (1), ortaokul (2), lise (3), lisans (4), önlisans (5), yüksek lisans (6) ve doktora (7) olmak üzere yedi gruba ayrılmıştır.



Şekil 12. Başvuru birim değişkeninin eğitim durumuyla ilişkisi

Şekil 12'de başvuru birim değişkenlerinin eğitim durumuyla olan ilişkisi görülmektedir. Kümeler incelendiğinde Clementine sonuçlarıyla uyumlu olduğu anlaşılmaktadır. En çok başvuru ulaşım birimine yapılmıştır. Üçüncü kümenin büyük çoğunluğunun eğitim düzeyinin lise seviyesinde olduğu sonucuna varılabilir.

Şekil 13'te başvuru birim değişkenlerinin yaş ile ilişkisi gösterilmiştir. Birinci kümenin büyük çoğunluğunun 35-44 yaş arasında olduğu sonucuna ulaşılmıştır.



Şekil 13. Başvuru birim değişkeninin yaş ile ilişkisi

4. Tartışma ve Sonuç

Bu çalışmada çağrı merkezi verilerinin kümelmesi amaçlanmıştır. Analiz aşamasında hiyerarşik olmayan kümeleme yöntemlerinden k-ortalamlar yöntemi kullanılmış ve küme sayısının altı olmasının uygun olduğu belirlenmiştir. Başvurular incelendiğinde, en çok tekrarın ulaşım birimi çağrıları için olduğu tespit edilmiştir. Analiz kapsamına alınan ilçelerin nüfus yoğunluğu ve konum olarak yakınlıkları ulaşım ağı sorunları bakımından beklenen sonuçtur. Ayrıca eğitim düzeyi bakımından lise ve üzeri bireylerin yaşadığı ilçelerde, çalışan nüfusun çoğunlukta olduğu varsayımı da başvuru tekrar sayısını açıklar niteliktedir. Clementine desteği ile belirlenen altı kümenin küme yapıları aşağıdaki gibi özetlenebilir:

Birinci kümede tamamı erkek 761 kişi yer almıştır. Bireylerin %36,01'i ulaşım birimine çağrıda bulunmuştur. Ayrıca bu kümede yer alan bireylerin %22,86'sının eğitim durumunun lise, %62,86'sının başvuru tipinin bilgi istemi, %25,62'sinin 35-44 yaş aralığında olduğu tespit edilmiş ve başvuruların tamamında anlık çözüme ulaşılmıştır.

İkinci kümede bulunan 743 kişinin tamamının kadın olduğu, %38,63'ünün ulaşım birimine çağrıda bulunduğu, %24,76'sının eğitim durumunun lisans, %48,72'sinin başvuru tipinin şikâyet, %26,24'ünün 35-44 yaş aralığında olduğu tespit edilmiş ve başvuruların hiçbirinde anlık çözüm olmamıştır.

Üçüncü kümede 734 kişinin çağrıları yer almıştır. Bu kümenin tamamının erkek olduğu, %33,92'sinin ulaşım birimine çağrıda bulunduğu, %20,44'ünün eğitim durumunun lise, %73,98'inin başvuru tipinin talep, %22,07'sinin 45-54 yaş aralığında olduğu analiz edilmiş ve başvuruların hiçbirinde anlık çözüm olmadığı sonucuna ulaşılmıştır.

Dördüncü kümede 542 kişi yer almıştır. Bu kümenin tamamı kadın bireylerden oluşmuştur. Bireylerin %39,67'si ulaşım birimine çağrıda bulunmuştur. Çağrıda bulunanların %23,43'ünün eğitim durumunun lisans, %67,71'sinin başvuru tipinin bilgi istemi, %24,91'sinin 35-44 yaş aralığında olduğu tespit edilmiş ve başvuruların tamamında anlık çözüm olduğu sonucuna ulaşılmıştır.

Beşinci kümede bulunan 420 kişinin tamamının başvuru tipinin şikâyet ve %74,29'unun erkek olduğu, %37,14'ünün ulaşım birimine çağrıda bulunduğu, %24,05'inin eğitim durumunun lisans, %27,86'sının 35-44 yaş aralığında olduğu ve başvuruların tamamında anlık çözüm olduğu sonucuna ulaşılmıştır.

Altıncı kümede bulunan 660 kişinin tamamının erkek ve başvuru tipinin şikâyet olduğu, %27,73'ünün ulaşım birimine çağrıda bulunduğu, %19,7'sinin eğitim durumunun lise, %22,42'sinin 35-44 yaş aralığında olduğu ve başvuruların hiçbirinde anlık çözüm olmadığı sonucuna ulaşılmıştır.

I. ve IV. Kümelerde bilgi istemi türünde gelen çağrılar oranı yüksektir. Bu tür çağrılar için anlık çözümün %100 oranında sağlanmış olması sebebiyle çağrı merkezinde hizmet verenlerin bilgi verme yetkinliğine sahip olduğu sonucuna varılabilir.

II. ,V. ve VI. kümelerde şikâyet türünde gelen çağrılardan V. Kümenin şikâyetlerinin anlık çözüme ulaştırıldığı ancak II ve VI. Kümelerde anlık çözüm sağlanamadığı anlaşılmaktadır. Çağrı merkezi hizmet performansını arttırmak üzere şikâyet kırılımları ile ilgili eğitim ve yetkin hizmet birimleri tanımlarının netleştirilmesi gereklidir.

III. kümede talep niteliğindeki çağrılar oranı yüksektir ve anlık çözüm sağlanamamıştır. Bu sonuç da şikâyet niteliğindeki çağrılar için yapılabilecek iyileştirmelerle daha yüksek müşteri memnuniyetine ulaşılabileceğini göstermektedir.

Tüm kümelerde ulaşım niteliği taşıyan çağrılar aynı düzeyde seyretmektedir. Anlık çözümün %100 sağlandığı I, IV, V. kümelerde ulaşım sorununa üretilen anlık çözümün payının ne oranda olduğuna dair daha ayrıntılı analizlere ihtiyaç vardır.

WEKA programı ile yapılan analize göre ise birinci kümede 973, ikinci kümede 1174, üçüncü kümede 557, dördüncü kümede 270, beşinci kümede 606, altıncı kümede 280 kişi bulunmaktadır. Programda uzaklıklar değiştirildiğinde kümelerin birim değerlerinin değişmediği fakat küme içi hata kareleri toplamlarının değiştiği belirlenmiştir.

Kaynakça

- [1] Kurniali, S. 2015. Customer Service Information System For A Call Center. *Procedia Computer Science*, 59(2015), 298-304.
- [2] Phung-Duc, T., Kawanishi, K. I. 2014. Performance Analysis of Call Centers with Abandonment, Retrial and After-Call Work. *Performance Evaluation*, 80(2014), 43-62.
- [3] Yakut, E. 2012. Veri madenciliği tekniklerinden C5. 0 algoritması, destek vektör makineleri ile yapay sinir ağlarının sınıflandırma başarılarının karşılaştırılması: İmalat sektöründe bir uygulama. Atatürk Üniversitesi Sosyal Bilimleri Enstitüsü, Doktora Tezi, Erzurum.
- [4] Çelik, Ş, Şengül, T, Şengül, Ö, İnci, H. 2019. Türkiye'de İllere Göre Hayvansal ve Bitkisel Ürünlerin Kümeleme Analizi İle İncelenmesi. *Journal of Awareness*, 3(5), 385-398.
- [5] Akay, Ö. 2019. Türkiye'de Halk Kütüphanesi Kullanımının Panel Veri Kümeleme Analizi İle İncelenmesi. *Opus Uluslararası Toplum Araştırmaları Dergisi*, 10(17), 1076-1099.
- [6] Ünlü, F., Atik, H. 2019. Türkiye'deki İşletmelerin Endüstri 4.0'a Geçiş Performansı: Avrupa Birliği Ülkeleri ile Karşılaştırmalı Ampirik Analiz. *Ankara Avrupa Çalışmaları Dergisi*, 17(2), 431-463.
- [7] Almeida, R., Steiner, M. T. A., Dos Santos Coelho, L., Francisco, C. A. C., Neto, P. J. S. 2018. A Case Study on Environmental Sustainability: A Study of The Trophic Changes in Fish Species as a Result of the Damming of Rivers Through Clustering Analysis. *Computers and Industrial Engineering*.
- [8] Sendín-Hernández, M. P., Ávila-Zarza, C., Sanz, C., García-Sánchez, A., Marcos-Vadillo, E., Muñoz-Bellido, F. J., ... ,Dávila, I. 2018. Cluster Analysis Identifies 3 Phenotypes Within Allergic Asthma. *The Journal of Allergy and Clinical Immunology: In Practice*, 6(3), 955-961.

- [9] Horiuchi, Y., Tanimoto, S., Latif, A. M., Urayama, K. Y., Aoki, J., Yahagi, K., ..., Komiyama, K. 2018. Identifying Novel Phenotypes of Acute Heart Failure Using Cluster Analysis of Clinical Variables. *International Journal of Cardiology*, 262, 57-63.
- [10] Özari, Ç., Eren, Ö. 2018. İllerin Yaşam Endeksi Göstergelerinin Çok Boyutlu Ölçekleme Ve K-Ortalamalar Kümeleme Yöntemi İle Analizi. *Afyon Kocatepe Üniversitesi Sosyal Bilimler Dergisi*, 20(2), 303-313.
- [11] Gazel, S., Akel, V. 2018. Borsa İstanbul'da Sektör Sınıflandırmasının Kümeleme Analizi İle Belirlenmesi. *Muhasebe ve Finansman Dergisi*, 77, 147-164.
- [12] Kıral, G, Çelik, C. 2018. Panel Veri Analizi ve Kümeleme Yöntemi ile Türkiye'de Konut Talebinin İncelenmesi. *Atatürk Üniversitesi İktisadi Ve İdari Bilimler Dergisi*, 32 (4), 1009-1026.
- [13] Sarıışık, G, Öğütlü, A. 2018. İstatistiksel K-Ortalamalar Kümeleme Analizi ile Doğal Taşların Yeni İşlenebilirlik İndeksi. *Harran Üniversitesi Mühendislik Dergisi*, 3 (3), 156-165.
- [14] Javadi, S., Hashemy, S. M., Mohammadi, K., Howard, K. W. F., Neshat, A. 2017. Classification of Aquifer Vulnerability Using K-Means Cluster Analysis. *Journal of Hydrology*, 549, 27-37.
- [15] Ceylan, Z, Gürsev, S, Bulkan, S. 2017. İki Aşamalı Kümeleme Analizi ile Bireysel Emeklilik Sektöründe Müşteri Profilinin Değerlendirilmesi. *Bilişim Teknolojileri Dergisi*, 10(4), 475-485.
- [16] Ülgen, G, Arda Özalp, L. 2017. Refah Rejimleri Sınıflandırma Çalışmaları: Cinsiyet Boyutları. *Marmara Üniversitesi İktisadi Ve İdari Bilimler Dergisi*, 39(2), 639-658.
- [17] Baynal, K., Çalış, A. (2016). Veri Madenciliğinde Kümeleme Analizi ile Bankacılık Sektöründe Bir Uygulama. *Beykent Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi*, 9(1), 13-41.
- [18] Dong, L., Wang, L., Khahro, S. F., Gao, S., Liao, X. 2016. Wind Power Day-Ahead Prediction with Cluster Analysis of Nwp. *Renewable And Sustainable Energy Reviews*, 60, 1206-1212.
- [19] Giray, S. 2016. İki Aşamalı Kümeleme Analizi ile Hükümlü Verilerinin İncelenmesi. *Ekonometri ve İstatistik E-Dergisi*, 25, 1-31.
- [20] Turan, K, Özari, Ç, Demir, E. 2016. Kümeleme Analizi ile Türkiye ve Ortadoğu Ülkelerinin Ekonomik Göstergeler Açısından Karşılaştırılması. *İstanbul Aydın Üniversitesi Dergisi*, 8 (29), 143-165.
- [21] Cankalp, E, Ünlükaplan, İ. 2011. Geçiş Ekonomilerinde Yönetişim Kalitesinin Kümeleme Analizi İle Belirlenmesi. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 25(3), 1-14.
- [22] Aydın, N., Seven, A. 2015. İl Nüfus ve Vatandaşlık Müdürlüklerinin İş Yoğunluğuna Göre Hibrid Kümeleme ile Sınıflandırılması. *Yönetim ve Ekonomi Araştırmaları Dergisi*, 13(2), 181-201.
- [23] Tekin, B. 2015. Temel Sağlık Göstergeleri Açısından Türkiye'deki İllerin Gruplandırılması: Bir Kümeleme Analizi Uygulaması. *Çankırı Karatekin Üniversitesi İİBF Dergisi*, 5(2), 389-416.
- [24] Silahtaroğlu, G. 2008. Veri Madenciliği – Kavram ve Algoritmaları. 2 basım. Papatya Yayınları, İstanbul.
- [25] Atbaş, A. 2008. Kümeleme analizinde küme sayısının belirlenmesi üzerine bir çalışma. Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Ankara.
- [26] Hubert, L. 1974. Approximate Evaluation Techniques for the Single-Link and Complete-Link Hierarchical Clustering Procedures. *Journal of the American Statistical Association*, 69, 698-704.
- [27] Akın, Y. K. 2008. Veri madenciliğinde kümeleme algoritmaları ve kümeleme analizi. Marmara Üniversitesi Sosyal Bilimler Enstitüsü, Doktora Tezi, İstanbul.
- [28] Al Kindhi, B., Sardjono, T. A., Purnomo, M. H., Verkerke, G. J. 2019. Hybrid K-Means, Fuzzy C-Means, and Hierarchical Clustering for DNA Hepatitis C Virus Trend Mutation Analysis. *Expert Systems with Applications*, 121, 373-381.
- [29] Bárcena, J. F., Camus, P., García, A., Álvarez, C. 2015. Selecting Model Scenarios of Real Hydrodynamic Forcings on Mesotidal and Macrotidal Estuaries Influenced by River Discharges Using K-Means Clustering. *Environmental Modelling and Software*, 68, 70-82.

- [30] Erdoęmu, P., olak, B., Durdaę, Z. 2016. K-Means Algoritması ile Otomatik Kmeleme. El-Cezeri Fen ve Mhendislik Dergisi, 3(2), 315-323.
- [31] zdamar, K. 2004. Paket Programlar ile İstatistiksel Veri Analizi (ok Deęikenli Analizler). Kaan Kitabevi, Eskiehir.
- [32] Cengiz, D., ztrk, F. 2012. Trkiye'de İllerin Eęitim Dzeylerine Gre Kmeleme Analizi ile İncelenmesi. Trakya University Journal of Social Science, 14(1), 69-84.
- [33] İle Nfus Daęılımları, <http://www.tuik.gov.tr> (Eriim Tarihi: 09.04.2019).
- [34] stnel, M. 2018. K ortalamalar algoritmasına dayalı kmeleme analizi sistemi ve perakendecilik sektrnde uygulanması. Yıldız Teknik niversitesi, Fen Bilimleri Enstits, Yksek Lisans Tezi, İstanbul.