

VERİ MADENCİLİĞİNDE KÜMELEME ANALİZİ VE SAĞLIK SEKTÖRÜNDE BİR UYGULAMASI

Hüdaverdi BİRCAN¹ Selim ÇAM²

ÖZET

Çalışmanın amacı, hastaların verilerinin bulunduğu çok boyutlu bir veri tabanının kümeleme analizi yöntemleriyle incelenmesidir. Veri madenciliği yöntemleri çok boyutlu ve büyük hacimlerdeki veri tabanlarında başarılı sonuçlar üretmektedir.

Bu çalışmada, Cumhuriyet Üniversitesi Hastanesi'ne 2011 yılında başvurmuş olan hastaların 2006-2011 arasındaki kayıtlar, hasta başvuru davranışlarının belirlenmesi amacıyla incelenmiştir. Oluşturulan veri seti yasalar tarafından yetişkin sayılan 18 yaş ile emeklilik sınırı olan 65 yaş arasında bulunan hastalara indirgenmiştir. Böylece veri seti 78.239 hastanın hastane veri tabanından alınan verileri ile oluşturulmuştur.

Kümeleme tekniklerinin veri tipine göre yeterlilikleri göz önüne alınarak çalışmada K-Means ve Yoğunluk Tabanlı Kümeleme tekniği uygulanmıştır. Sayısal verilerin yanı sıra hastaların demografik verileri de veri setine dahil edilerek, hastalar hakkında daha somut yargılara varılmak istenmiştir.

Uygulanan kümeleme analizi tekniklerinden Yoğunluğa Dayalı olan kümeleme yöntemi klasik bir kümeleme yöntemi olan K-Means Kümeleme yöntemine göre daha uygun sonuçlara ulaşılmasını sağlamıştır.

Anahtar Kelimeler: Veri Madenciliği, Yoğunluk Tabanlı Kümeleme Yöntemi, K-Means Kümeleme Yöntemi

CLUSTERING ANALYSIS IN DATA MINING AND AN APPLICATION IN HEALTH SECTOR

ABSTRACT

The aim of the study is examining of a multidimensional database via clustering analysis methods. Data mining methods produce successful results in multidimensional and large databases.

In this study, records of patients who applied to Cumhuriyet University Hospital in 2011 between the years of 2006-2011 were examined for the purpose of determination of patient

¹ Doç. Dr., Cumhuriyet Üniversitesi, İktisadi İdari Bilimler Fakültesi, İşletme Bölümü, Sivas, hbircan@cumhuriyet.edu.tr

² Doktora Öğrencisi, Gaziosmanpaşa Üniversitesi, Sosyal Bilimler Enstitüsü, Tokat, selim-cam@gmail.com

Application behaviours. Data set were reduced to the interval of 18 which is respected as adult age by law and 65 which is the legal limit of retirement. In this way, data set was produced with 78,239 patients' data derived from hospital database.

Regarding clustering techniques according to data types, K-Means and Density Based Clustering methods were used in the study. Integrating demographic data with numerical data to the database, it was intended to acquire more tangible jurisdictions about patients.

Consequently, Density Based Clustering Method provide more favourable results than K-Means Clustering Method which is a classical clustering method.

Key Words: Data Mining, Density Based Clustering Method, K-Means Clustering Method

GİRİŞ

Veri madenciliği, geniş veri yığınları içerisinde, ihtiyaç duyulan noktalarda kullanılabilir ihtimali olan, ortaya çıkmamış bilgilerin ve ilişkilerin anlaşılır ve kullanılabilir bir şekilde araştırılması yöntemlerine verilen genel bir analiz yöntemidir (Nisbet vd., 2009, s.4; Witten, 2011, s.3). Başka bir bakışla: “daha önceden bilinmeyen geçerli ve uygulanabilir bilgilerin geniş veri tabanlarından elde edilmesi ve bu kararların işletme kararları verirken kullanılmasıdır” (Silahtaroglu, 2008, s.10).

Özel sektörde her ne kadar veri madenciliği kullanılan bir karar verme yöntemi olarak uygulanırsa da kamu kurumlarında bu tarz bir çalışmanın etkili bir şekilde kullanıldığından bahsetmek zordur. Özellikle çalışmanın merkezinde yer alan sağlık sektöründe veri madenciliği yöntemlerinin kilit bir öneme sahip olduğu düşünülmektedir. Uluslararası Hastalık Sınıflandırma Sistemine (ICD-10, International Statistical Classification of Diseases and Related Health Problems) göre 11332 adet (http://hastane.dicle.edu.tr/index.php?option=com_content&view=article&id=125:icd, 11.08.2012) hastalık tanısı vardır. Bu hastalıklara karşılık gelen hastalar ve demografik bilgiler göz önüne alınacak olunursa elde edilen verilerin analizi veri madenciliği yöntemleriyle hesaplama yapmayı zorunlu kılmaktadır. Türkiye'nin nüfus, iklimi v.b. özellikleri bakımından bölgesel olarak hastalıkların çeşidi, kronik olup olmaması değişiklik göstermektedir. Dolayısıyla tedavide yer alacak hekim ve nitelikli personel sayısının, ilaç ve tıbbi malzeme bulundurulmasının etkili bir şekilde planlanmasının yapılması, ayrıca ülke ekonomisine yük getirmeyecek bir düzenlemenin yapılması gerekliliği de önemlidir.

Veri madenciliği uygulamaları sağlık sektöründe farklı yöntemler ve konu başlıkları altında incelenmiştir. Literatürde hastaların teşhis, tedavi süreçleriyle ilgili verileri toplanarak hastalık tahminleri; hastaların hastaneye başvuruları ile hasta davranış tahminleri yapılmıştır. Örneğin Kaur ve Wasan. 2006 yılında

yaptıkları bir çalışmada hastaların yaş, cinsiyet, belirti sıklığı, kan tahlili sonuçları ve vücut kitle endeksi gibi verilerini alarak hastalarda herhangi bir hastalık olup olmadığını araştırmışlardır (Kaur ve Wasan, 2006).

Nagadevara 2004 yılında Uluslararası e-Yönetim Konferansı'ndaki bildirisinde hastaların demografik ve çevresel bilgileri analizi sonuçlarını sunmuştur (Nagadevara, 2004). Benzer şekilde Ertuğrul v.d 2013 yılında Pamukkale Üniversitesi Hastanesi'ne gelen hastaların hastaneye geldikleri zamanı, demografik verileri, poliklinik verilerini alarak hastaneye gelen 2009-2011 dönemindeki hastaların hasta profilini belirlemişlerdir (Ertuğrul vd., 2013).

Mullins vd. (2006), 667.000 hastaya ait verilerle yaptıkları çalışmalarında, tıbbi akademik bir veri madenciliği yaklaşımı olan HealthMiner® adını verdikleri tescilli bir sistem aracılığıyla uyguladıkları CliniMiner® isimli denetimsiz bir veri madenciliği metodu kullanmışlardır (Mullins vd., 2006). Demografik, sosyo-ekonomik ve klinik verilerle, seçilen vakalardaki biyolojik çıktılarını değerlendirdiği çalışmanın sonucunda, farklı klinik hastalıklara yönelik tahmin analizleri aracılığıyla bilgi ve örüntüleri ortaya çıkarmışlardır.

Koh ve Tan (2011), yaş, beden kitle endeksi, bel-kalça ölçüsü, haftada yapılan egzersiz sayısı değişkenlerini kullanarak veri madenciliği yöntemlerinden birisi olan karar ağaçları yönteminden faydalanarak analizler yapmışlardır (Koh ve Tan, 2011). Çalışmanın sonuçları ile geliştirilen müşteri ilişkileri yönetim metodu, elde edilen sınıflara ait farklı hizmet tipleri geliştirilmelerine olanak tanımıştır.

Bu çalışmada Sivas ilindeki uygulama ve araştırma hastanesinin veri tabanlarından alınan hastaların tedavi-demografik verilerinin elde edilmesi ve bu verilere göre hasta profillerinin çıkartılarak, hastanedeki hizmet planlamasına yardımcı olmak amaçlanmıştır.

VERİ SETİ

Çalışmada kullanılan veri seti 2011 yılında hastaneye başvurmuş olan hastaların hastaneye her bir başvurularının verisini içermektedir. Veri setinde kullanılan parametreler, hastaların ayakta tedavi olması göz önüne alınarak poliklinik sayısı veya yatarak tedavi gören hastaların kaç defa yatarak tedavi aldığı ve bu yatışlarındaki toplam hastanede kalma süreleri alınmıştır. Hastaların yatarak tedavi olması gerekiyorsa bunun sebebi, hastayı gözetim altına almak mı? Cerrahi müdahalede bulunmak mı? soruları akla gelmektedir. Bu sebeple hastaların oldukları ameliyatların sayıları da çalışmaya dahil edilmiştir. Ayrıca hastalıklarda tedavinin süreci ve şiddetinde önemli faktörler olarak düşünülen yaş ve cinsiyet parametreleri de alınmıştır. Genel olarak hastaneye yapılan başvuruların dar bir aralıkta olup olmadığını anlamak amacıyla hekimler tarafından hastalara verilen teşhisler ve hastaların kaç farklı servise başvurduğu, çalışmada hasta davranışlarını

ve üçüncü basamak bir hastaneyi incelemek için faydalı olacağı düşünülmüştür. Başvuran hastaların Sivas ilinden gelip gelmemesi de önemli bir değişkendir. Ayrıca hasta yapılarını belirlemek amacıyla hangi ilden gelirse gelsin hastaların il merkezinden mi? merkez dışındaki bölgelerden mi? geldikleri de bu çalışmada incelenmiştir.

Söz konusu parametreler toplanırken yaş unsurunun öne çıkarttığı bazı sorunlar olmuştur. Öncelikle, hastanedeki tedavi birimleri 18 yaş altı ve üstü olarak ayrılmaktadır. Çocuk hastalıklarına başvuran hastalara uygulanan işlemler, dolayısıyla tedavi süresi ve tipi farklı olacaktır. Ayrıca ileri yaşlarda bulunan hastaların hastaneye başvuru sıklığı, ameliyat sayıları ve yatarak tedavi edilen gün sayıları, bu parametrelerdeki ortalamayı yukarı çekmektedir. Bu sebeple veri seti yasalar tarafında yetişkin sayılan 18 yaş ile emeklilik sınırı olan 65 yaş arasında bulunan hastalara indirgenmiştir. Böylelikle veri seti 78.239 hastanın veri tabanından alınan verileri ile oluşturulmuştur. Buradan da 18 yaşındaki bir hastanın önceki yıllarda yaptığı hastane başvurularının nasıl kullanılacağı sorunu ortaya çıkmaktadır. Bu yüzden veri tabanından veriler alınırken bu sorunla karşılaşılacak hastaların 18 yaş altındaki başvuruları veri setine dahil edilmemiştir.

YÖNTEMLER

Veri tabanından süzülen veriler uygulamaya hazır hale getirildikten sonra oluşan 78.239 hasta ve 11 parametre kullanılarak öncelikle veri madenciliği kümeleme analizinde sıkça kullanılan yöntemlerden olan K-Means kümeleme yöntemi kullanılmıştır. Kümeleme tekniklerinin veri tipine göre yeterlilikleri göz önüne alınarak bir diğer kümeleme tekniği olan Yoğunluk Tabanlı Kümeleme tekniği uygulanmıştır.

Veri seti farklı iki teknikte kümelere ayrıldıktan sonra kümeler arasındaki farklar hastaların demografik özelliklerine göre sınanmak istenmiştir. Ancak öncelikle karşılaştırma analizinin belirlenebilmesi için verilerin normallik sınavını geçmesi gerekmektedir. Bu amaçla öncelikle verileri Kolmogorov-Smirnov Z testi uygulanmıştır. Verilerin normal dağılıma uygun olmaması sebebiyle kümelerin dağılımlarının farklılığını ölçmek amacıyla Kruskal-Wallis H testi ve küme ortalamalarının farklılığını test edebilmek amacıyla da Mann-Whitney U testi kullanılmıştır. Ayrıca kategorik olan değişkenlerin kümelere göre farklılığını ölçmek amacıyla Ki-Kare testi kullanılmıştır.

I-K-MEANS KÜMELEME YÖNTEMİ

Analiz için, n nesneden oluştuğu ve her bir kümenin en az bir nesne içerdiği varsayılır (Kanungo vd., 2002). Böylelikle k adet kümeden oluşan analiz gerçekleştirilirken kümeleme işlemi her bir nesnenin ayrı ayrı hesaplanması

yapılarak küme aidiyetlikleri belirlenir (Han vd., 2012, s.327).K-Means algoritmasıyla n adet veriden oluşan bir veri setinden merkez noktası c ile temsil edilen k adet küme oluşturulmaktadır (Işık ve Çamurcu, 2007; Vattani, 2011). K-Means algoritmasına Lloyd's algoritması da denilmektedir (Frahling ve Sohler, 2005). Küme merkezleri değişmeyecek hale gelene kadar algoritma kendisini tekrar eder. Merkezin değişmeyecek hale gelmesi, hata kareler toplamının minimum olmasıyla anlaşılır. Ayrıca K-Means kümeleme algoritması öklidyen uzaklık hesaplamalarını kullandığı için bu yöntem nominal ve ordinal ölçekli verilerde uygun sonuç vermeyecektir (Han vd., 2012, s.451).

II-YOĞUNLUK TABANLI KÜMLEME YÖNTEMİ

Bu tekniklerde uzaklığa dayalı küme seçiminden ziyade verilerin yoğunluğuna göre bir kümeleme işlemi yapılmaktadır. Yoğunluğa dayalı kümeleme tekniklerinde, kümeler veri tabanındaki daha yüksek yoğunluklu alanlar olarak tanımlanmaktadır (Ester vd., 1996). Küme yoğunluklarının seyrek olduğu alanlarda ise ya gürültülü veriler ya da küme sınırını oluşturan veriler bulunmaktadır (Jiang vd., 2003). Bu algoritmayı büyük veri tabanları ve gürültülü verisi çok olan yapılarda kullanmak oldukça uygundur (Ester vd., 1996; Kriegel vd., 2011). Farklı büyük ve şekillerdeki kümelerin oluşturulmasında da bu algoritma kullanılabilir bir yöntemdir. Algoritmanın çalışmasını daha iyi anlamak amacıyla algoritmanın bazı özelliklerinin belirtilmesi gerekmektedir (Ye vd., 2003).

- Herhangi bir p noktasının Eps değeri MinPts değerinden daha fazla veri içeriyorsa kümenin yeni merkezi bu p değeri olacaktır.
- Eğer bir veri merkez nokta değilse sınır değeri olacaktır. Bu sınır değeri başka bir merkez noktası için yoğunluğa katılabilir niteliktedir.
- Bir p noktası q noktasının yoğunluğuna katılabilirse; bu durum $p_1=q$ ve $p_n=p$ koşullarında ve p_1, \dots, p_n nesnelere zincirleme ise p_{i+1} nesnesi de yoğunluğa katılmış olacaktır (Borah ve Bhattacharyya, 2004).
- Eğer yukarıdaki maddelere uygun herhangi bir veri yoksa; yani kendi başına kalmış bir veri söz konusuysa, bu veri gürültü olarak nitelendirilmektedir

Çalışmada kullanılan analizler SPSS 14.0 ve WEKA (<http://www.cs.waikato.ac.nz/ml/weka/downloading.html>) paket programları aracılığıyla yapılmış olup sonuçlar aşağıda belirtilmiştir.

BULGULAR

Çalışmaya katılan 78.239 hastanın, birinci bölümde K-Means kümeleme analizi ile küme merkezleri, parametre ortalaması ve standart sapmaları

hesaplanmıştır. Benzer şekilde, aynı hastalara Yoğunluk Temelli kümeleme yöntemi uygulanarak, bu analize göre oluşan kümelerin merkezleri ve kümelerdeki parametrelerin ortalama ve standart sapmaları hesaplanmıştır. Ayrıca çalışmaya eklenen demografik verilerin oluşan kümeler ile nasıl bir etki gösterdiği araştırılmıştır.

I-K-MEANS KÜMELEME YÖNTEMİ

Yapılan çalışmada alınan sonuçlara göre, küme sayısı arttıkça benzemezlik değeri düşmekte ve benzer şekilde küme içi hata kareler değeri de düşmektedir. Ancak küme sayısının artması birçok parametre için ayırt ediciliğin azalarak yok olmasına neden olmaktadır. Bu sebepten çalışmada küme sayısı üç olarak belirlenmiştir. Küme sayısının üç olduğu çözümde benzemezlik değeri 31,14 ve küme içi hata kareler toplamı da 2621,34 olarak hesaplanmıştır.

Tablo 1.K-Means Sonuçlarına Göre Küme Merkezleri ve Standart Sapmaları

Parametre	1.Küme		2.Küme		3.Küme	
	Merkez	Std	Merkez	Std	Merkez	Std
Yaş	39,9	±4,7	56,4	±5,0	24,8	±4,1
Poliklinik Tedavi	10,0	±13,3	12,8	±16,7	6,3	±7,9
Yatarak Tedavi	0,9	±1,9	1,5	±2,6	0,5	±1,5
Farklı Servis	4,2	±3,4	4,8	±3,8	3,2	±2,5
Ameliyat	0,9	±2,4	1,3	±3,4	0,5	±1,7
Farklı Teşhis	5,7	±6,1	6,9	±6,9	4,1	±3,9
Yatılan Gün Sayısı	6,4	±19,9	12,6	±30,6	3,1	±12,4
Tedavi Ücreti	1939,2	±5339,4	3680,6	±8435,9	927,9	±2757,3

Analiz sonrası oluşan kümelerin parametrelere göre merkez noktaları Tablo 1'de gösterilmektedir. Oluşan kümelere göre hastaların, %29'u (n=23.044) birinci küme, %26'sı (n=20.685) ikinci küme ve % 44'ü (34.510) üçüncü küme ayrıştırılmıştır. Kümelerin bir birlerinden farklı olup olmadığını, yani küme ayrımlarının düzgün olup olmadığını test etmek amacıyla Mann-Whitney U testi ile karşılaştırmalar yapılmıştır. Bu testi kullanılmasının amacı, veri miktarı yeterli olsa da Kolmogorov-Smirnov Z testi sonucunda verilerin normal dağılıma uygun olmadığı sonucuna ulaşılmasıdır. Mann-Whitney testi sonucunda da kümelerin 0,05 ve 0,01 anlamlılık düzeylerinde istatistiksel olarak farklı olduğu tespit edilmiştir.

Hastaların cinsiyetlerine göre yapılan incelemede, oransal olarak birinci ve ikinci kümedeki cinsiyet dağılımının eşit ve üçüncü kümede diğer iki kümeye oranla daha fazla kadın olduğu (%44,8) sonucu görülmektedir. Ayrıca ikinci kümedeki erkek sayısı (%29,9) birinci kümeye göre (%26,8) daha fazla olduğu sonucu ortaya çıkmıştır. Hastaların çoğunluğunun il merkezinden geldiği görülmektedir (%76,5). Ayrıca merkez dışından gelen hastaların çoğunluğu üçüncü kümede yoğunlaşmıştır (%46,5). Sivas ili dışından gelme oranı tüm hastalar içinde %86,1'lik bir kısmı oluşturmaktadır. Sivas ili dışından gelen hastaların çoğunluğu üçüncü kümeye atanmıştır (%18,2).

II-YOĞUNLUK TABANLI KÜMLEME YÖNTEMİ

Analiz sonucunda veri seti iki kümeye ayrılmış olup, algoritma işlemi 14 iterasyonda sonlanmıştır. Bu analize göre ortaya çıkan küme merkezleri ve standart sapma değerleri Tablo 2'de verilmektedir.

Tablo 2. Yoğunluk Tabanlı Kümeleme Analizi Sonuçlarına Göre Küme Merkezleri ve Standart Sapmaları

Parametre	1.Küme		2.Küme	
	Merkez	Std	Merkez	Std
Yaş	27,9	±6,3	52,2	±7,3
Poliklinik Tedavi	7,1	±9,3	12,2	±16,0
Yatarak Tedavi	0,6	±1,6	1,3	±2,5
Farklı Servis	3,4	±2,7	4,7	±3,7
Ameliyat	0,6	±1,9	1,1	±3,1
Farklı Teşhis	4,5	±4,6	6,6	±6,8
Yatılan Gün Sayısı	3,8	±14,7	10,7	±27,7
Tedavi Ücreti	1143,3	±3.414,4	3.187,6	±7.733,8

Analiz sonucuna göre hastaların, %62'si (n=48.624) birinci kümeye, %38'i (n=29.615) ikinci kümeye ayrılmıştır. Oluşan kümelerdeki parametrelerin dağılımı Kolmogorov-Smirnov Z testi sonucunda göre normallik sınavından geçememiştir. Bu sebepten kümelerin doğru bir şekilde ayrılıp ayrılmaması Mann-Whitney U testi ile incelenmiştir. Bu testin sonucunda da kümelerdeki parametrelerin 0,05 ve 0,01 anlamlılık düzeylerinde istatistiksel olarak farklı olduğu tespit edilmiştir.

Elde edilen kümelerle hastaların cinsiyetleri incelenmiş olup, kadın ve erkek dağılımlarının neredeyse eşit olması sebebiyle {Küme 1 (Kadın, Erkek)= (%54,2; %45,8): Küme 2 (Kadın, Erkek)= (%54,7;%45,3)} Ki-Kare testi sonucunda kümelere göre cinsiyetin istatistiksel olarak anlamlı bir farklılık oluşturduğu söylenememektedir. Ayrıca merkez dışından gelen hastaların çoğunluğu ikinci kümede yoğunlaşmıştır (%61,5). Ek olarak Sivas ili dışından gelen hastaların çoğunluğu birinci kümeye atanmıştır (%71,7).

SONUÇLAR

K-Means kümeleme analizine göre, hastaneden en çok hizmet alan grubun ikinci grup olduğu gözükmemektedir. Bu durumda, analize katılan tüm parametrelerde ikinci kümedeki ortalamaların değerleri diğer kümelere göre daha yüksek olduğunu göstermektedir. Ortalama yaşta ikinci kümenin ortalaması daha yüksek olup neredeyse 65 yaş olan üst sınırdadır. Sadece üçüncü kümedeki, yani yaş olarak diğer kümedeki hastalardan daha genç olan hastalar kullanılan parametrelerin tamamında genel ortalamanın aşağısında kalmıştır. Birinci kümede sadece yatılan gün sayısı ve ücret parametreleri genel ortalamanın aşağısındadır. Bunlardan ücret parametresi genel ortalamanın sadece 14,34 TL aşağısında olması önemli bir fark olarak görülmemektedir. Ücret parametresindeyse, birinci kümedeki hastaların genel ücret ortalamasının yaklaşık %2,8 daha azı maliyete katlanmakta olduğu sonucu ortaya çıkmıştır, buradan yola çıkarak hastaların üçüncü kümede % 52,5 daha az ve ikinci kümede %88,4 daha fazla sağlık ücretiyle ödedikleri yorumu söylenebilir. Kümeler kendi arasında incelendiğinde ikinci kümedeki hastaların en az bir defa yatarak tedavi olduğu, yine en az bir defa ameliyat oldukları belirlenmiştir. Ayrıca ikinci gruptaki bir hastanın bir seferlik yatarak tedavisinin yaklaşık 8,5 gün sürdüğü buna karşılık birinci grupta ki bir hasta 6,9 gün ve üçüncü gruptaki bir hasta da 5,6 gün yatarak tedavi olduğu anlaşılmıştır. Genel ortalamaya bakıldığında bu sürenin 7,2 gün olduğu hesaplanmıştır. İkinci kümedeki bir hastanın polikliniğe gelişinde farklı bir teşhis konmasının olasılığı %54,2, birinci kümede aynı durumun olasılığı %57,3 ve üçüncü kümede %65,5 olarak hesaplanmıştır. Demografik verilere göre hastaların il dışından gelip gelmemeleri, il merkezinde yaşayıp yaşamamaları da hastaların aldığı sağlık hizmeti ve bu sağlık hizmetinin ödenmesi noktalarında etkin olduğu sonucuna varılmıştır.

Yoğunluk tabanlı kümeleme algoritması sonucunda kümeleri sağlık hizmetine daha fazla ihtiyaç duyan ya da talep eden grup, birinci kümeyi; sağlık hizmetlerine daha az ihtiyaç duyan ya da talep eden grup, ikinci kümeyi temsil etmektedir. Birinci kümeye dahil olan hastaların ortalama olarak 7,1 defa ayaktan tedavi aldıkları buna rağmen yeni bir başvuru olması durumunda farklı bir teşhis ile karşılaşmaları olasılığı yaklaşık %63 olarak hesaplanmıştır. Aynı durum ikinci kümede incelenirse, hastaların ortalama olarak 12,2 defa ayaktan tedavi aldıkları ve

buna karşılık yeni bir başvuruda bulduklarında farklı bir teşhis ile karşılaşma olasılıkları %54,6 olarak karşımıza çıkmaktadır. Buradan da yaşı ilerlemiş hastalarda hastalıkların önceden belirlenmiş olduğu, başvuruların kronik hastalıkların yol açtığı sağlık sorunları olduğu düşünülmektedir. Yatarak tedavi edilen hastalar birinci kümede ortalama olarak 6 gün ve ikinci kümede de ortalama olarak 8,2 yatarak tedavi olmuşlardır. Bu yöntemde de hastaların il dışından gelip gelmemeleri, il merkezinde yaşayıp yaşamamaları da hastaların aldığı sağlık hizmeti ve bu sağlık hizmetinin ödenmesi noktalarında etkin olduğu sonucuna varılmıştır.

Analizler sonucu ortaya çıkan kümelerin ayrımları Tablo 3'te gösterilmektedir

Tablo3.Kümelere Göre Hasta Yapıları

K-Means Kümeleme Sonuçlarına Göre Hasta Yapıları		Yoğunluk Tabanlı Kümeleme Algoritması Sonuçlarına Göre Hasta Yapıları	
1.Küme	<ul style="list-style-type: none"> • Orta Yaşlı • Ortalama Sağlık Hizmet Alan • İl Merkezinde Yaşayan • Sivas'ta Yaşayan 	1.Küme	<ul style="list-style-type: none"> ▪ Genç ▪ Ortalama Altı Sağlık Hizmeti Alan ▪ İl Merkezinde Yaşayan ▪ Sivas Dışında Yaşayan
2.Küme	<ul style="list-style-type: none"> • Yaşlı • Ortalama Üstü Sağlık Hizmeti Alan 	2.Küme	<ul style="list-style-type: none"> ▪ Yaşlı ▪ Ortalama Üstü Sağlık Hizmeti Alan ▪ İl Merkezi Dışında Yaşayan ▪ Sivas'ta Yaşayan
3.Küme	<ul style="list-style-type: none"> • Genç • Ortalama Altı Sağlık Hizmeti Alan • Kadın • İl Merkezi Dışında Yaşayan • Sivas Dışından Gelen 		

Analiz için kullanılan iki yöntemde de sonuçlar incelenmiş olup, K-Means algoritmasında küme sayısının araştırmacı tarafından belirlenmesi eksiklik olarak düşünülmektedir. Ayrıca bu yöntemde verilerin kanonik bir şekle uyması analiz için daha uygun olacağı ve gürültülü değerlerden etkilenmesi hususları da analizin doğruluğunu etkilemektedir. Ancak yoğunluk tabanlı kümeleme algoritmasının verilerin şeklinden ya da veri setindeki gürültülü verilerden etkilenmemesi

sonuçların doğruluğu konusunda araştırmanın güvenilirliğini arttıran bir niteliktedir.

KAYNAKÇA

- Borah, B. ve Bhattacharyya, D. K. (2004). *An Improved Sampling-Based DBSCAN for Large Spatial Databases*. 8th International Conference on Spoken Language Processing, Jeju Island, South Korea, 4-8 October.
- Ertuğrul, İ., Organ, A., ve Şavlı, A. (2013). Veri Madenciliği Uygulamasına İlişkin PAÜ Hastanesinde Hasta Profilinin Belirlenmesi. *Pamukkale University Journal of Engineering Sciences*, 19(2),97-103
- Ester, M.,Kriegel, H. P., Sander, J. ve Xu, X. (1996). *A Density-Based Algorithm for Discovering Cluster in Large Spatial Databases with Noise*. 2nd International Conference on Knowledge Discovery and Data Mining, Portland, OR, 2-4 August.
- Frahling, G.ve Sohler, C. (2005). Coresets in Dynamic Geometric Data Streams. 37. Annual ACM Symposium on Theory of Computing. New York, USA, 22-24 May
- Han, J., Kamber, M. ve Pei, J. (2012). *Data Mining Concepts and Techniques (3. Baskı)*. Boston, USA, Elsevier.
- http://hastane.dicle.edu.tr/index.php?option=com_content&view=article&id=125:icd, 11.08.2012
- Işık, M. ve Çamurcu, A. Y. (2007). K-Means, K-Medoids ve Bulanık C-Means Algoritmalarının Uygulamalı Olarak Performanslarının Tespiti. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*. 11(1), 31-45
- Jiang, D.,Pei, J. ve Zhang A. (2003). *DHC: A Density-Based Hierarchical Clustering Method for Time Series Gene Expression Data*. Third IEEE Symposium on Bio Informatics and Bio Engineering, Washington, DC, USA, 12 March.
- Kanungo, T.,Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R. ve Wu, A. Y. (2002). An Efficient K-Means Clustering Algorithm: Analysis and Implementation. *Pattern Analysis and Machine Intelligence*, 24(7), 881-892
- Kaur, H., ve Wasan, S. K. (2006). Empirical Study on Applications of Data Mining Techniques in Healthcare. *Journal of Computer Science*, 2(2),194-200
- Koh, H. C.,& Tan, G. (2011). Data mining applications in healthcare. *Journal of Healthcare Information Management*, Vol, 19(2), 65-75.
- Kriegel, H. P.,Kröger, P., Sander, J. ve Zimek, A. (2011). Density-Based Clustering. *Wire's Data Mining and Knowledge Discovery*, 3, 231-240.
- Mullins, I. M.,Siadaty, M. S., Lyman, J., Scully, K., Garrett, C. T., Greg Miller, W., Muller, R., Robson, B., Apte, C., Weiss, S., Rigoutsos, I., Platt, D.,

- Cohen, S. ve Knaus, W. A. (2006). Data mining and clinical data repositories: Insights from a 667,000 patient data set. *Computers in biology and medicine*, 36(12), 1351-1377.
- Nagadevara, V. (2004). Application of neural prediction models in healthcare. In *Proceedings of the 2nd International conference on e-Governance ICEG* (pp. 139-149).
- Nisbet, R., Elder, J. ve Miner G. (2009). *Handbook of Statistical Analysis Data Mining Applications*. Toronto, Canada, Elsevier.
- Silahtaroglu, G. (2008). *Kavram ve Algoritmalarıyla Temel Veri Madenciliği*. İstanbul, Türkiye, Papatya Yayıncılık.
- Vattani, A. (2011). K-Means Requires Exponential Many Iterations Even in the Plane. *Discrete Computer Geom.* 45, 596-616
- Witten, I. H., Frank, E. ve Hall, M. A. (2011). *Data Mining: Machine Learning Tools and Techniques*. New York, USA, Morgan Kaufmann Publisher.
- Ye, Q., Gao, W. ve Zeng, W. (2003). *Color Image Segmentation Using Density-Based Clustering*. IEEE International Conference on Acoustics, Speech and Signal Processing, Hong Kong, Hong Kong, 6-10 April.