

KÜMELEME ANALİZİ İLE BANKACILIK SEKTÖRÜNDE SATIŞ STRATEJİLERİNİN BELİRLENMESİ

Aslı ÇALIŞ (*aslicalis@gazi.edu.tr*)

Gazi Üniversitesi Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, Ankara, Türkiye

Kasım BAYNAL (*kbaynal@kocaeli.edu.tr*)

Kocaeli Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, Kocaeli, Türkiye

ÖZET

Veri madenciliği, büyük miktardaki veriden, anlamlı bilgiler ve kurallar keşfetme sürecidir. Günümüzde karar verme sürecine ihtiyaç duyulan birçok alanda olduğu gibi bankacılık alanında da veri madenciliği teknikleri yaygın bir şekilde kullanılmaktadır. Bankacılık sektöründe yapılan bu çalışmada, veri madenciliğinde kümeleme yöntemlerinden k-ortalamlar ile Türkiye’de faaliyet gösteren bir banka şubesine ait iki yüz müşterinin, on iki farklı değişkene kümelmesi ve kümelerdeki müşteri profillerine göre satış stratejilerinin geliştirilmesi amaçlanmıştır. Analizde veri madenciliği yazılımı olarak SPSS Clementine programı kullanılmış ve sonuçlar yorumlanmıştır.

Anahtar Kelimeler: *Bankacılık Sektörü, K-Ortalamlar, Kümeleme Analizi, Veri Madenciliği.*

DETERMINATION OF SALES STRATEGIES IN BANKING SECTOR WITH CLUSTER ANALYSIS

Ash ÇALIŞ (*aslicalis@gazi.edu.tr*)

*Gazi University, Engineering Faculty, Industrial Engineering Department,
Ankara, Turkey*

Kasım BAYNAL (*kbaynal@kocaeli.edu.tr*)

*Kocaeli University, Engineering Faculty, Industrial Engineering Department,
Kocaeli, Türkiye*

ABSTRACT

Data mining is the process of discovering meaningful information and rules from big data. Nowadays, data mining techniques are used extensively in area of banking such as many areas which need decision making process. In this study, conducted in banking sector, it was aimed to cluster to two hundred customers who belong to a bank branch operating in Turkey according to twelve different variables and develop strategies according to customer profiles with k-means which is one of the clustering methods in data mining. In the study, SPSS Clementine was used as a software of data mining and the results were interpreted.

Keywords: *Banking Sector, K-Means, Cluster Analysis, Data Mining.*

1. GİRİŞ

Ham veri kendi başına değersizdir. Veri, bilgisayar sistemleriyle belirli bir amaç doğrultusunda işlenerek bilgiye dönüşmektedir. Bilgisayar teknolojilerindeki gelişmeler, üretilen bilgi miktarlarında ve veri tabanı sistemlerinin hacminde artış meydana getirmiştir. Veri tabanlarında saklı tutulan, yararlı olma potansiyeline sahip verilerin keşfedilerek anlamlı örüntülerin ortaya çıkarılması, veri madenciliği kavramıyla ifade edilmektedir. Bilginin olağanüstü artışıyla birlikte her alanda strateji geliştirme konusunda kurumları ve bireyleri desteklemek amacıyla ileriye dönük tahmin sistemlerine ihtiyaç duyulmuştur. Bu bağlamda veri madenciliği (VM) teknikleri diğer birçok alanda olduğu gibi bankacılık alanında da yaygın bir şekilde kullanılmaktadır.

VM, kredi taleplerinin değerlendirilerek kredi geri ödemelerinin kontrol altında tutulması, müşteri gruplarının ve riskli müşteri tiplerinin belirlenmesi, sadık müşteri portföyünün oluşturulması, müşterilere özgü satış politikalarının oluşturularak çapraz satış ile birim müşteriye yapılan satış miktarının artırılması gibi amaçlardan ötürü bankacılık ve finans sektöründe oldukça yaygın bir biçimde kullanılmaktadır. Bu doğrultuda, çalışmada ülkemizde faaliyet gösteren büyük bir bankanın bir ildeki birinci sınıf şubesinden elde edilen verilerle bir VM uygulaması gerçekleştirilmiştir. Bankaya ait veriler, bireysel kredi müşterilerinin yaş, cinsiyet, medeni hal, öğrenim durumu, aylık gelir, ev, araç, çocuk sahibi olma durumu, eş geliri, ödeme durumu, banka maaş müşterisi olma durumu ve çalışma şekli olmak üzere toplamda on iki farklı değişkenden oluşan kişisel özelliklerini içermektedir. Uygulamada k-ortalamar yöntemi ile KA yapılarak mevcut müşterilerin değerlendirilmesi sağlanmıştır. Bu süreçte, kümeleme algoritmalarını kolaylıkla uygulayarak, kısa sürede verideki gizli örüntülere ulaşılmasını sağlayan bir VM programı olan SPSS Clementine kullanılmıştır.

Literatürde kümeleme analizi ile ülkemizde faaliyet gösteren bir bankanın bireysel kredi müşterilerinin değerlendirilerek, müşteri profillerine göre pazarlama stratejilerinin geliştirildiği başka bir veri madenciliği çalışmasına rastlanılmamıştır. Bu yönüyle çalışmanın literatüre katkı sağlaması hedeflenmiş olup, ülkemizde bankacılık ve finans alanında yapılan veri madenciliği çalışmalarını şu şekilde özetlemek mümkündür: Tosun (2006), kredi kartı müşterilerinin kaybedilme sebeplerini belirleyebilmek için veri madenciliği yöntemlerinden faydalanarak sonuçlara ulaşmayı amaçlamıştır [1]. Aşan (2007), kredi kartı kullanan banka müşterilerini sosyo-ekonomik özelliklerine göre kümeleme analizi ile gruplamaya çalışmıştır [2]. Çakır (2008) çalışmasında VM standart sürecinin tüm aşamalarını bankacılık müşteri veri tabanından rastlantısal olarak seçerek veri kümesi üzerinde uygulamış ve VM' nin sınıflandırma fonksiyonu üzerinde durmuştur [3]. Doğan (2008), Türk Bankacılık Sektörü' nde 1998–2006 dönemi itibariyle faal olan ticaret bankalarına ait finansal oranları temel alarak Kümeleme Analizi (KA) uygulamasına yer vermiştir [4]. Albayrak (2009), yerli ve yabancı olarak önceden grup üyeliği belirlenmiş bankaların sınıflandırmasında yaygın olarak kullanılan VM tekniklerinden diskriminant, lojistik regresyon ve karar ağacı modellerini bankalarla ilgili seçilmiş likidite, gelir-gider, karlılık ve faaliyet oranları kullanarak karşılaştırmıştır [5]. Bilen (2009), banka satış personellerinin performanslarını değerlendirerek, kümeleme yöntemlerinden k ortalama ile personellerin performans başarı düzeylerine göre sınıflandırılmasını sağlamıştır [6]. Çil (2010) çalışmasında, öncelikle bir bankanın mevcut fonlarını alıp satan ve belli bir işlem geçmişinden sonra bankadaki hesabını kapatarak banka yatırım fonu müşterisi olmaktan çıkmış müşterilerin, işlem hareket detayının öğrenilerek, hesabını kapatmış müşterilerin sosyo-demografik karakteristiğinin çıkartılması ve bundan sonra hesabını kapatmaya meyilli müşterilerin tespit edilmesi üzerinde durmuştur [7]. Taşkın ve Emel (2010) çalışmalarında bir perakende işletmenin

müşterilerinin Kohonen ağları ile kümelenmesini ele almışlardır [8]. Savaş vd. (2012), VM' nin günümüz disiplinleri arasında geldiği noktaya değinerek Türkiye'de VM üzerine yapılan çalışmaları ve gerçekleştirilen uygulamaları incelemişlerdir [9]. Çınar ve Silahtaroglu (2012), müşteri memnuniyetine etki eden gizli nedenlerin keşfi için 301 kişi ile yapılan anketin sonuçlarını inceleyerek, memnuniyet ya da memnuniyetsizliğin altında yatan nedenleri grafiksel analiz, K- means ve C5.0 algoritması kullanarak ortaya çıkarmaya çalışılmışlardır [10]. Yakut ve Elmas (2013), İMKB'de işlem gören 140 sanayi işletmesinin 2005-2008 yılları arasındaki finansal başarısızlıklarını VM ve diskriminant analizi modelleri ile tahmin ederek hangi yöntemin daha iyi sonuç verdiğini tespit etmişlerdir [11]. Dolgun ve Ersel (2014) çalışmalarında karar ağaçları, lojistik regresyon, bayesci ağlar ve destek vektör makineleri gibi veri madenciliği yöntemlerini kullanarak bankacılık sektöründe doğrudan pazarlama kampanyalarının nasıl yönlendirilebileceği üzerinde durmuşlardır [12]. Duran vd. (2014), bir bankanın doğrudan pazarlama kampanyasına katılabilecek hedef müşteri kitlesini belirleyebilmek için yapay sinir ağları, lojistik regresyon ve karar ağaçlarının tahmin güçlerini karşılaştırarak kampanyayı en iyi açıklayan modeli belirlemişlerdir. Çalışma sonucunda, yapay sinir ağları yönteminin kampanyaya katılacak müşteri kitlesinin tahmininde daha iyi sonuç verdiğini gözlemlemişlerdir [13]. Doğan (2015) çalışmasında bir e-ticaret sitesi kullanıcılarının şifre yapılarına ilişkin bazı istatistiksel çıkarımlar ortaya koymuştur. Kullanıcıların şifre ve bazı diğer bilgileri ile oluşturulan veri kümesini, veri madenciliği tekniklerinden birliktelik analizlerine tabi tutarak değişkenler arası kullanışlı kural tabanlı bilgiler elde etmeye çalışmıştır [14]. Yaşar (2016), veri madenciliği yöntemlerinden C5.0, CART ve GRI algoritmalarını kullanarak Borsa İstanbul'da işlem gören sanayi şirketlerinin 2011-2014 dönemi verilerine dayalı olarak, olumlu görüş dışındaki denetim görüşünü tahmine ilişkin kurallar elde

etmiştir [15]. Kaygın vd. (2016), Borsa İstanbul’da 2010-2013 döneminde kesintisiz olarak işlem gören 143 imalat sanayi şirketinin yıllık bilanço ve gelir tablosu verilerinden yararlanarak işletmelerin finansal başarılı ve başarısız olma durumlarını Veri Madenciliği ve Lojistik Regresyon Analizi yöntemleri ile tahmin etmeye çalışılmışlardır. Analiz sonucunda, işletmelerin finansal başarılı ve başarısız olma durumlarını tahmin etmek için oluşturulan tüm modellerde 2012 yılını tahmin gücü en başarılı yıl olarak saptamışlardır [16].

2. VERİ MADENCİLİĞİ

Veri, bilgisayar sistemleriyle belirli bir amaç doğrultusunda işlenerek bilgiye dönüşmektedir. VM, tek başına ham verinin sunmadığı bilgiyi ortaya çıkaran veri analizi sürecidir [17]. Farklı bir ifadeyle VM, büyük miktardaki veriden, anlamlı örüntüler ve kurallar keşfetme sürecidir [18]. VM’ nin ortaya çıkışı veri yığınlarının geniş yer kaplamasına ve büyük miktardaki verilerin yararlı bilgilere dönüştürülmesi ihtiyacına dayanmaktadır [19].

VM, karar destek, pazar stratejisi, finansal tahminler gibi birçok alanda uygulanabilir olması nedeniyle son zamanlarda veritabanı kullanıcıları ve araştırmacıların önemli ölçüde dikkatini çekmektedir. VM, makine öğrenme, istatistik ve veri tabanları alanlarındaki teknikleri birleştirerek, büyük veri tabanlarından faydalı ve değerli bilgiyi çıkarmamıza imkan tanımaktadır [20]. VM, istatistik, sinir ağları, karar ağaçları, genetik algoritma ve görsel teknikler gibi yıllardır geliştirilen çeşitli teknikleri içermektedir. VM, pazarlama, finans, bankacılık, üretim, sağlık, müşteri ilişkileri yönetimi ve organizasyon öğrenme gibi çoğu alanda uygulanmaktadır [21].

VM, kredi kartı dolandırıcılıklarının tespiti, kredi kartı harcamalarına göre müşteri gruplarının belirlenmesi, kredi taleplerinin değerlendirilmesi, kredi geri ödemelerinin kontrol

altında tutulması, riskli müşteri gruplarının ve sadık müşterilerin belirlenerek bu müşterilere özgü satış politikalarının oluşturulması gibi kullanım amaçlarından ötürü bankacılık sektöründeki uygulamalarda tercih edilmektedir.

3. VERİ MADENCİLİĞİNDE KÜMELEME ANALİZİ

3.1. Kümeleme Analizi

Kümeleme modellerinde amaç, küme üyelerinin birbirlerine çok benzediği ancak özellikleri birbirlerinden çok farklı olan kümelerin bulunması ve veri tabanındaki kayıtların bu farklı kümelere bölünmesidir [22]. Kümeleme, VM' nin temel işlemlerinden biridir. Kümeleme, müşteri segmentasyonu ve dolandırıcılık tespiti gibi problemlerin çözümünde yaygın biçimde kullanılır. Kümeleme uygulamalarında üç görevi yerine getirmiş oluruz [20].

1. Veri setlerini kümeler içinde bölümlere ayırma,
2. Kümeleme sonuçlarını doğrulama,
3. Kümeleri yorumlama.

KA, önceden belirlenen seçme kriterlerine göre birbirine çok benzeyen birey ya da nesnelere aynı küme içinde gruplandırmaktadır. Analizin sonucunda bir kümeyi oluşturan birey veya nesnelere birbiriyle benzerken, diğer kümelerin birey veya nesnelere benzeşmeyeceğinden, kümeler kendi içlerinde homojen iken, kümeler arasında heterojenlik söz konusu olmaktadır. Oluşturulan kümeler çok boyutlu uzayda gösterildiğinde, eğer kümeleme başarılı ise aynı küme içinde yer alan birey veya nesnelere birbirlerine oldukça yakın çıkması, bununla birlikte farklı kümelerin de birbirinden fark edilir düzeyde uzak olması beklenmektedir [23].

3.2. Kümeleme Analizi Teknikleri

Birimlerin benzerliklerine göre kümelere dahil edilmesinde kullanılacak çeşitli yaklaşımlar mevcuttur. Bu yaklaşımlardan biri, en çok benzer iki birimi aynı gruba atamakla başlayıp, tüm birimlerin aynı gruba atanması ile biten hiyerarşik bir yaklaşımdır. Bir başka yaklaşım ise tüm verilerin ortalama değerlerine en yakın değerlere sahip birimlerin aynı kümeye atanmasını esas alan yaklaşımdır. En çok kullanılan bu iki yaklaşım dışında diğer yaklaşımlar da mevcuttur. Tüm yaklaşımlarda en önemli ölçüt, kümeler arası farklar ile kümeler içi benzerliklerin maksimum olmasını sağlamaktır. En çok kullanılan kümeleme algoritmaları hiyerarşik ve hiyerarşik olmayan kümeleme adı altında iki kategoride toplanmaktadır [24].

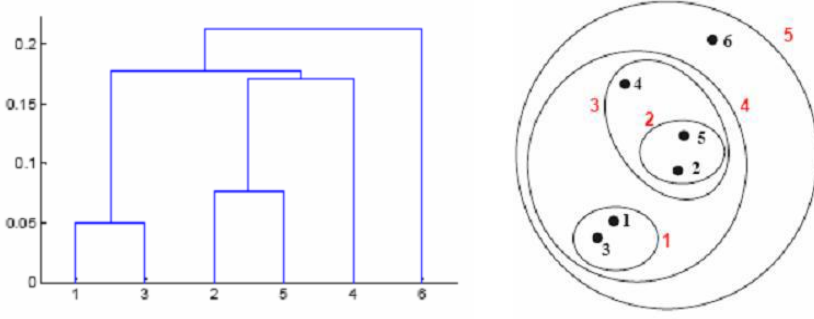
3.2.1. Hiyerarşik kümeleme yöntemleri

Hiyerarşik kümeleme teknikleri, kümeleri peş peşe birleştirme sürecidir ve bir grup, diğeri ile bir kez birleştirildikten sonra, daha sonraki adımlarda kesinlikle ayrılamaz. Hiyerarşik tekniklerin ağaç diyagramları ile gösterilen sonuçlarına dendogram denir [25].

Hiyerarşik kümeleme nesnelerin yakınlık ilişkisine göre oluşturulan kümelerden bir ağaç inşa eder. Hiyerarşik kümeleme aşağıdaki özelliklere sahiptir [7].

- Bir veri tabanını bir kaç kümeye ayrıştırır.
- Bu ayrıştırma dendogram adı verilen bir ağaç sayesinde yapılır.
- Bu ağaç, yapraklardan gövdeye doğru veya gövdeden yapraklara doğru kurulabilir. Dendogram istenen seviyede kesilerek kümeler elde edilir.

Hiyerarşik yöntemle veri kümelemeye ilişkin örnek Şekil 1'de verilmektedir.



Şekil 1. Hiyerarşik yöntemle veri kümeleme örneği

3.2.2. Hiyerarşik olmayan kümeleme yöntemleri

Birimlerin kendi içinde homojen ve kendi aralarında heterojen olan kümelerle ayrılmasını hedefleyen ve prototip kümeler aracılığı ile alt popülasyonların parametre tahminlerini yapmayı amaçlayan yöntemlerdir. Hiyerarşik olmayan kümeleme yöntemlerinde birimlerin uygun oldukları kümelerde toplanmaları ve n birimin k kümeye parçalanması hedeflenmektedir [26].

Hiyerarşik olmayan kümeleme yöntemleri başlığı altında birçok teknikten söz etmek mümkündür ancak bunlardan en sık kullanılanı k-ortalamlar yöntemidir.

3.2.2.1. K-ortalamlar (K-means) yöntemi

K-means algoritması 1967 yılında J.B. MacQueen tarafından geliştirilmiştir [27]. En yaygın kullanılan denetimsiz öğrenme yöntemlerinden biri olan k-means'ın amacı, başlangıçta belli k sayıdaki küme için, küme içi kareler toplamını minimize ederek n nesneyi k kümeye bölmektir.

K-ortalamlar yöntemi, verideki kümeleri bulmaya yarayan en açık ve etkili kümeleme algoritmasıdır [28]. K-means kümeleme yönteminin değerlendirilmesinde en yaygın olarak karesel hata kriteri olan SSE kullanılmaktadır. En düşük SSE değerine sahip olan kümeleme sonucu en iyi sonucu vermektedir. Nesnelerin

buldukları kümenin merkez noktalarına olan uzaklıklarının karelerinin toplamı (1) nolu eşitlikteki gibi hesaplanmaktadır [29].

$$SSE = \sum_{i=1}^K \sum_{x \in C_i} dist^2(m_i, x) \quad (1)$$

Burada, “dist” iki nesne arasındaki standart Öklid Uzaklığı, “x” değeri C_i kümesinde bulunan bir nesne, “ m_i ” değeri C_i kümesinin merkez noktasıdır. Yukarıda açıklanan k-means algoritması, iki boyutlu veriler üzerinde Öklid Uzaklık ölçütüne göre çalışmaktadır ve hiçbir nesne kümesini terk etmeyene kadar ötelenmektedir [30].

K-means algoritmasının adımlarını aşağıdaki gibi özetlemek mümkündür [31]:

1. Veri seti rassal olarak k adet başlangıç kümesine ayrılır.
2. Veri setinde yer alan örnekler; merkezi kendisine en yakın olan kümeye atanır.
3. Her atamanın sonunda küme merkezi (ortalama) yeniden hesaplanır.
4. Veri setindeki tüm örneklerin ataması yapılmaya kadar 2. ve 3. adımlar tekrarlanır.

Bu süreç tüm gözlemler gruplara atanıncaya kadar devam eder. Tüm gözlemler gruplara atandıktan sonra atandıkları küme ortalamasından daha yakın küme ortalaması varsa, gözlemlerin yerleri değiştirilmektedir. Amaç diğer kümeleme yöntemlerinde olduğu gibi, gerçekleştirilen kümeleme işlemi sonucunda elde edilen kümelerin, küme içi benzerliklerinin maksimum, kümeler arası benzerliklerinin ise minimum olmasını sağlamaktır [24]. Bu yöntem çok sayıda birimden elde edilmiş olan sürekli değişkenli veri setlerini küme içi kareler toplamını minimize edecek biçimde k kümeye ayırmayı amaçlamaktadır. Birimlerin az sayıda kümeye yerleştirilmesi iteratif bir biçimde yapılmakta olup, her iterasyonda farklı kümelere atanarak en uygun çözüm permutasyonel bir

yaklaşım ile belirlenir. K- ortalamalar yöntemi, birimlerin incelenmesi ile $k=2$ 'den başlayarak küme sayılarını her defasında birer arttırarak deneysel olarak en uygun kümelemeyi bulmak şeklinde uygulanabilir. Böyle bir yaklaşımda toplam küme içi varyans matrisinin izi minimize edilir [26].

4. UYGULAMA

4.1. Uygulamaya Genel Bakış

Çalışmada özel bir bankanın birinci sınıf şubelerinden birine ait bireysel kredi müşterilerinin kredi geri ödeme performanslarının değerlendirilmesine yönelik bir VM uygulaması gerçekleştirilmiştir. Uygulamada kullanılan veriler, Türkiye'nin finans sektöründe faaliyet gösteren en büyük bankalarından birinin birinci sınıf kategorisindeki bir şubesinin bireysel kredi müşterilerine aittir.

Uygulamada veri madenciliği yöntemlerinden kümeleme ele alınmıştır. Kümeleme analizi ile ilgili şubedeki iki yüz müşterinin on iki farklı değişkene göre kümelenmesi ve kümelerdeki müşteri profillerine göre satış stratejilerinin belirlenmesi amaçlanmıştır. Analizde VM uygulamaları için geliştirilmiş bütünsel bir görsel modelleme gereci olan SPSS Clementine programı kullanılmıştır.

4.2. Verilerin Hazırlanması

4.2.1. Veri toplama: Bu aşamada bankanın genel müdürlük bünyesindeki departmanlarından müşteri numaralarına göre son altı aylık döneme ait takipli ve normal ödemeli toplam 200 adet müşterinin kredi geri ödemelerine ilişkin veriler elde edilmiştir.

4.2.2. Veri birleştirme ve temizleme: Müşteri numaralarına göre müşterilerin cinsiyet, medeni hal, yaş, aylık gelir, öğrenim durumu, ev ve araç sahibi olma, çocuk sahibi olma, banka maaş müşterisi olma durumu, çalışma şekli ile ilgili bilgilerine ise bankadaki mevcut sistem üzerinden ulaşılarak veriler birleştirilmiştir. Eksik ve hatalı

veriler temizlenerek, toplamda 200 adet müşteriye ait on iki farklı değişkenden oluşan 200x12’lik bir veri kümesi elde edilmiştir.

4.2.3. Veri dönüştürme: Bu aşamada, bankadaki yetkililerden uzman görüşü alınarak değişkenler aşağıdaki gibi kategorik hale getirilmiştir.

- *Cinsiyet:* Bayan ve erkek müşteriler sırasıyla “B” ve “E” olarak tanımlanmıştır.
- *Medeni Hal:* Medeni hal değişkeni de “EVLI” ve “BEKAR” olmak üzere iki kategoride ele alınmıştır.
- *Yaş:* Müşterilerin doğum tarihleri dikkate alınarak yaşları hesaplanmıştır. Yaş değişkenine ait tanımlamalar Tablo 1’de gösterilmektedir.
- *Aylık gelir:* Müşterilerin aylık gelirleri dikkate alınarak, veriler Tablo 2’de gösterildiği gibi anlamlı gruplarda kategorize edilmiştir.

Tablo 1. Yaş değişkenine ait tanımlamalar

YAŞ ARALIĞI	TANIMLAMA
24-30	24-30 YAS
31-37	31-37 YAS
38-44	38-44 YAS
45-51	45-51 YAS
52-58	52-58 YAS
59-65	59-65 YAS
66 yaş ve üzeri	66 YASVEUSTU

Tablo 2. Aylık gelir değişkenine ait tanımlamalar

AYLIK GELİR	TANIMLAMA
≤750	750TLVEALTI
751-1400	751-1400TL
1401-2050	1401-2050
2051-2700	2051-2700
2751-3350	2751-3350
3351-4000	3351-4000
≥4001	4001TLVEUSTU

- *Eş Geliri:* Müşterilerin eş gelirlerinin mevcut olup olmama durumu sırasıyla “VAR” ve “YOK” olarak tanımlanmıştır.
- *Ev:* Müşterilerin ev sahibi olma ya da olmama durumları “VAR” ve “YOK” olarak iki kategoride toplanmıştır.
- *Araç:* Müşterilerin araç sahibi olma ve olmama durumları sırasıyla “VAR” ve “YOK” olarak gruplandırılmıştır.
- *Çocuk:* Çocuklu müşterileri, çocuk sahibi olmayan müşterilerden ayırmak için yine “VAR” ve “YOK” şeklinde tanımlamalar yapılmıştır.
- *Banka maaş müşterisi:* Kredi kullanmış olduğu bankadan aynı zamanda maaş alan müşteriler için “EVET”, sadece kredi müşterisi olup maaşını farklı bankalardan alanlar içinse “HAYIR” tanımlamaları yapılmıştır.
- *Çalışma şekli:* Çalışma şekline göre müşteriler üç farklı gruba ayrılmıştır. Kamuda çalışan müşteriler için “KAMU”, özel sektör çalışanları için “OZEL”, emekli müşteriler içinse “EMEKLİ” olarak tanımlama yapılmıştır.

- **Öğrenim durumu:** Öğrenim durumuna göre üç farklı tanımlama yapılmıştır. İlköğretim ve daha düşük seviyede öğrenim durumuna sahip müşterilere ait veriler “ILKOĞRETİM”, lise mezunu müşterilere ilişkin veriler “LİSE”, lisans ve üstünde öğrenim durumuna sahip müşterilere ait veriler ise “UNİVERSİTE” şeklinde kategorize edilmiştir.
- **Ödeme durumu:** Kredi taksitlerini üç dönem boyunca ödemeyen müşteriler kanuni takibe girmektedir. Kanuni takipteki müşterilere ait veriler “KANUNI_TAKIP”, geri ödemede herhangi bir sorun yaşamayan müşterilere ait veriler ise “NORMAL_ODEME” başlığı altında toplanmıştır. Ödeme durumuna ilişkin veriler son altı aylık döneme aittir.

Veri dönüştürme işlemi tamamlandıktan sonra elde edilen veri setinin bir kısmı Tablo 4’te gösterilmektedir.

Tablo 4. Veri Dönüştürme İşlemi Sonrası Veri Setinin Bir Bölümü

	cinsiyet	medeni_hal	yas	aylik_gelir	es_geliri	ev	arac	cocuk	banka_maas_musterisi	calisma_sekli	ogrenim_durumu	odeme_durumu
81	E	EVLI	52-58	751-1400TL	YOK	VAR	YOK	VAR	EVET	EMEKLİ	LİSE	KANUNI_TAKIP
82	E	EVLI	68YASVEUSTU	751-1400TL	YOK	VAR	YOK	VAR	HAYIR	EMEKLİ	LİSE	KANUNI_TAKIP
83	E	EVLI	45-51	1401-2050TL	YOK	VAR	YOK	VAR	EVET	KAMU	LİSE	KANUNI_TAKIP
84	E	EVLI	59-65	751-1400TL	YOK	VAR	YOK	VAR	HAYIR	OZEL	ILKOĞRETİM	KANUNI_TAKIP
85	E	EVLI	45-51	2051-2700TL	YOK	YOK	YOK	VAR	HAYIR	KAMU	ILKOĞRETİM	KANUNI_TAKIP
86	E	EVLI	45-51	1401-2050TL	YOK	YOK	VAR	HAYIR	OZEL	LİSE	KANUNI_TAKIP	
87	B	BEKAR	45-51	1401-2050TL	YOK	YOK	YOK	YOK	HAYIR	OZEL	LİSE	KANUNI_TAKIP
88	B	EVLI	68YASVEUSTU	750TLVEALTI	VAR	VAR	YOK	VAR	HAYIR	EMEKLİ	ILKOĞRETİM	KANUNI_TAKIP
89	E	EVLI	68YASVEUSTU	751-1400TL	YOK	VAR	YOK	VAR	EVET	EMEKLİ	ILKOĞRETİM	KANUNI_TAKIP
90	E	EVLI	38-44	2051-2700TL	YOK	YOK	YOK	VAR	EVET	KAMU	ILKOĞRETİM	KANUNI_TAKIP
91	E	EVLI	31-37	751-1400TL	YOK	YOK	VAR	HAYIR	OZEL	LİSE	KANUNI_TAKIP	
92	E	EVLI	45-51	750TLVEALTI	YOK	VAR	YOK	VAR	HAYIR	EMEKLİ	ILKOĞRETİM	KANUNI_TAKIP
93	E	EVLI	52-58	751-1400TL	YOK	YOK	VAR	EVET	EMEKLİ	LİSE	KANUNI_TAKIP	
94	B	BEKAR	24-30	751-1400TL	YOK	YOK	YOK	HAYIR	OZEL	LİSE	KANUNI_TAKIP	
95	B	BEKAR	45-51	2051-2700TL	YOK	YOK	YOK	HAYIR	KAMU	LİSE	KANUNI_TAKIP	
96	E	EVLI	45-51	1401-2050TL	YOK	VAR	YOK	VAR	EVET	EMEKLİ	ILKOĞRETİM	KANUNI_TAKIP
97	E	EVLI	38-44	751-1400TL	YOK	YOK	VAR	HAYIR	OZEL	LİSE	KANUNI_TAKIP	
98	E	EVLI	45-51	1401-2050TL	YOK	YOK	VAR	HAYIR	KAMU	LİSE	KANUNI_TAKIP	
99	E	BEKAR	38-44	751-1400TL	YOK	YOK	YOK	HAYIR	OZEL	ILKOĞRETİM	KANUNI_TAKIP	
100	E	EVLI	31-37	751-1400TL	YOK	YOK	VAR	HAYIR	OZEL	LİSE	KANUNI_TAKIP	
101	E	EVLI	45-51	2051-2700TL	VAR	VAR	VAR	EVET	KAMU	UNİVERSİTE	NORMAL_ODEME	
102	B	BEKAR	24-30	4001TLVEUSTU	YOK	YOK	YOK	HAYIR	KAMU	UNİVERSİTE	NORMAL_ODEME	
103	E	EVLI	52-58	1401-2050TL	VAR	VAR	YOK	EVET	EMEKLİ	ILKOĞRETİM	NORMAL_ODEME	
104	E	EVLI	52-58	1401-2050TL	YOK	VAR	VAR	EVET	EMEKLİ	ILKOĞRETİM	NORMAL_ODEME	
105	E	EVLI	38-44	1401-2050TL	VAR	YOK	VAR	EVET	KAMU	LİSE	NORMAL_ODEME	
106	E	EVLI	59-65	751-1400TL	YOK	VAR	YOK	VAR	EVET	EMEKLİ	UNİVERSİTE	NORMAL_ODEME
107	B	EVLI	38-44	1401-2050TL	VAR	YOK	VAR	EVET	KAMU	LİSE	NORMAL_ODEME	
108	E	EVLI	38-44	1401-2050TL	VAR	YOK	VAR	EVET	KAMU	LİSE	NORMAL_ODEME	
109	E	EVLI	38-44	2751-3350TL	VAR	VAR	VAR	EVET	KAMU	UNİVERSİTE	NORMAL_ODEME	
110	E	BEKAR	24-30	751-1400TL	YOK	YOK	YOK	EVET	OZEL	LİSE	NORMAL_ODEME	
111	B	BEKAR	31-37	1401-2050TL	YOK	YOK	YOK	HAYIR	KAMU	UNİVERSİTE	NORMAL_ODEME	
112	E	EVLI	38-44	2051-2700TL	VAR	YOK	VAR	HAYIR	KAMU	LİSE	NORMAL_ODEME	
113	B	EVLI	31-37	2051-2700TL	VAR	VAR	VAR	HAYIR	KAMU	UNİVERSİTE	NORMAL_ODEME	
114	E	BEKAR	24-30	2051-2700TL	YOK	YOK	VAR	EVET	KAMU	UNİVERSİTE	NORMAL_ODEME	
115	E	EVLI	52-58	1401-2050TL	YOK	VAR	YOK	VAR	EVET	EMEKLİ	ILKOĞRETİM	NORMAL_ODEME
116	E	EVLI	45-51	3351-4000TL	VAR	VAR	VAR	EVET	KAMU	UNİVERSİTE	NORMAL_ODEME	
117	B	BEKAR	31-37	3351-4000TL	YOK	VAR	YOK	HAYIR	OZEL	UNİVERSİTE	NORMAL_ODEME	
118	E	EVLI	38-44	1401-2050TL	VAR	YOK	VAR	EVET	KAMU	LİSE	NORMAL_ODEME	
119	E	BEKAR	31-37	1401-2050TL	YOK	YOK	YOK	EVET	OZEL	LİSE	NORMAL_ODEME	
120	E	EVLI	59-65	751-1400TL	VAR	VAR	YOK	EVET	EMEKLİ	ILKOĞRETİM	NORMAL_ODEME	

4.3. K-Means Yöntemi ile Kümeleme Analizi (Cluster Analysis with K-Means Method)

Uygulamada KA için hiyerarşik olmayan kümeleme yöntemlerinden k-means algoritması kullanılmıştır. K-means algoritmasında “k” değeri probleme göre belirlenebilir veya belirlenemez. Hata kareler ölçütü gibi bir kümeleme ölçütünün olması gerekir. K-ortalamalar algoritması “k” kümelerini, her bir kümeyi temsil edecek bir nesnenin keyfi seçimiyle başlatır. Kalan her nesne bir kümeye atanır ve kümeleme kriteri, küme ortalamasını hesaplayabilmek için kullanılır. Bu ortalamalar yeni küme noktaları olarak kullanılır ve her bir nesne kendisine en benzer olan kümeye yeniden atanır. Bu kümeler yeniden hesaplanır ve kümelere hiçbir değişimin gözlenmediği duruma ve değişim istenen hata düzeyinin altına düşürülünceye kadar bu döngü devam ettirilir [6].

KA’ nin en kritik konusu küme sayısına karar vermektir. Araştırmacının küme sayısına karar vermede özneliği minimize etmesi gerekmektedir. Ancak günümüzde yayınlanan birçok makalede bu konuda kesin bulunmuş sonuçlar yoktur. İlk önerilen yaklaşımlardan en çok bilinen eşitlik:

$$k = (n/2)^{1/2} \quad (2)$$

biçiminde hesaplanmaktadır. Burada “k” küme sayısını, “n” birim (nesne) sayısını göstermektedir. Bu eşitliğin küçük örneklemlerle araştırmalarda kullanılması tavsiye edilir. Büyük örneklemlerle araştırmalarda kullanılması durumunda sağlıklı sonuçlara ulaşılması zorlaşır [24].

Uygulamada küme sayısını belirlemek için iki farklı yöntem kullanılmıştır. İlk olarak küme sayısı:

$k = (200/2)^{1/2}$ ’den “10” olarak bulunmuştur. Burada “200” değeri müşteri sayısını göstermektedir.

İkinci aşamada, $k=2$ ’den $k=10$ ’a kadar küme sayısı birer arttırılmış ve her değer için hata kareleri toplamı bulunmuştur. Yukarıdaki

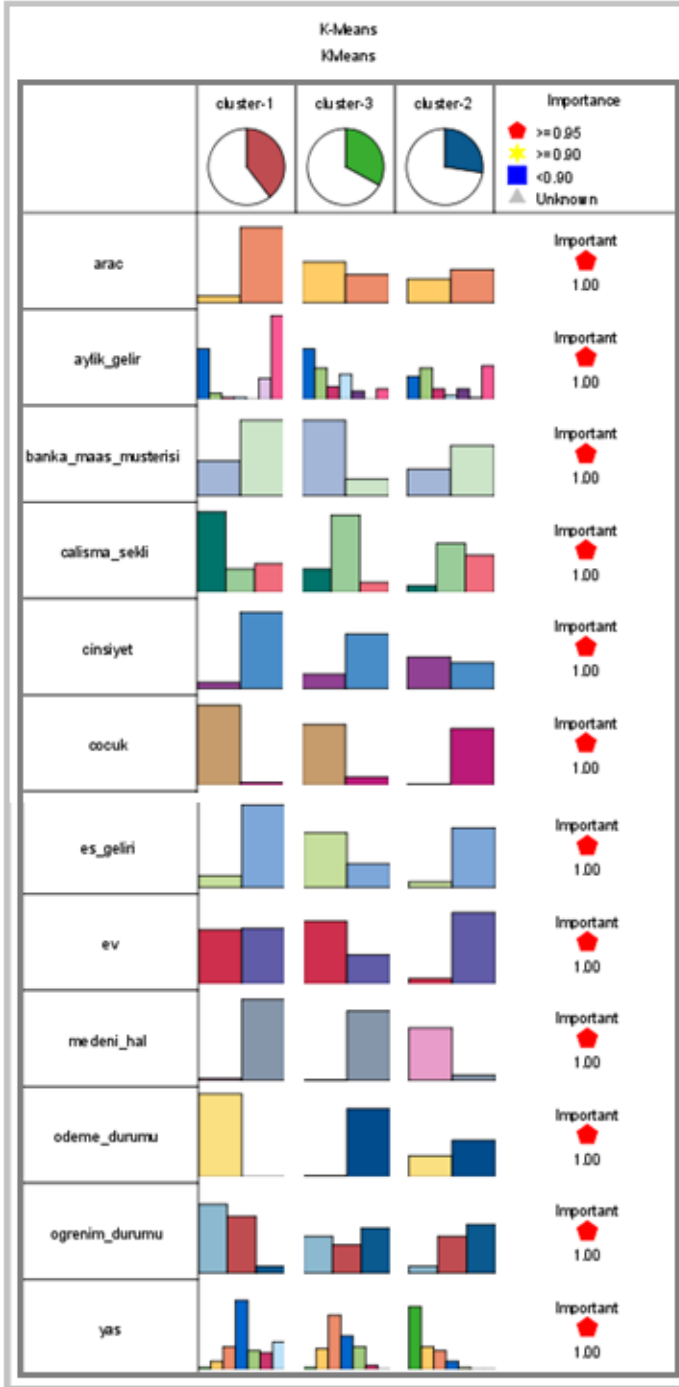
formüle göre bulunmuş olan $k=10$ değeri ile diğer küme sayılarına ilişkin hata kareleri karşılaştırılmış ve hata kareleri toplamı en küçük olan değer küme sayısı olarak kabul edilmiştir [32].

Tablo 5’te her küme sayısı için hata kareleri toplamına ilişkin değerler görülmektedir. Hata kareleri toplamı en az olan küme sayısı 3 olarak bulunmuş ve bir sonraki aşamada $k=3$ için değişkenlerin kümeler üzerindeki önem dereceleri belirlenmiştir.

Tablo 5. K-Ortalamalar için Küme sayısı ve Hata Kareleri Toplamı [32]

Küme Sayısı	2	3	4	5	6	7	8	9	10
Hata Kareleri Toplamı	2,22372	1,73550	1,78784	1,90777	2,11905	1,80539	1,98876	2,07131	1,92548

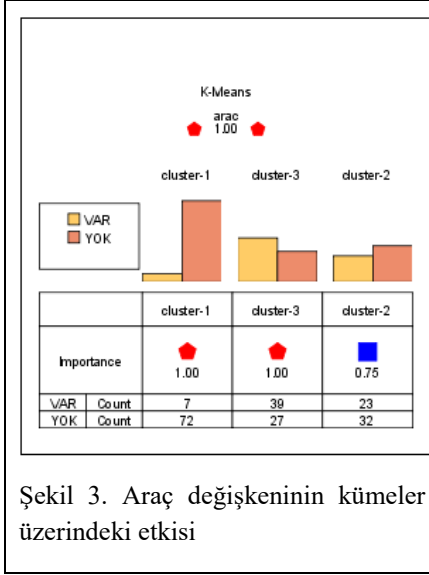
K-ortalamalar yöntemi ile elde edilen kümeler ve değişkenlerin kümeler üzerindeki etkileri Şekil 2’de gösterilmektedir. Programda önem dereceleri 0.90’ın altında olan değişkenlerin kümeler üzerindeki etkilerinin önemli düzeyde olmadığı kabul edilmektedir. Şekil 2 incelendiğinde, değişkenlerin tümünün üç küme üzerindeki etkilerinin önemli ölçüde olduğu sonucuna ulaşılabilir.



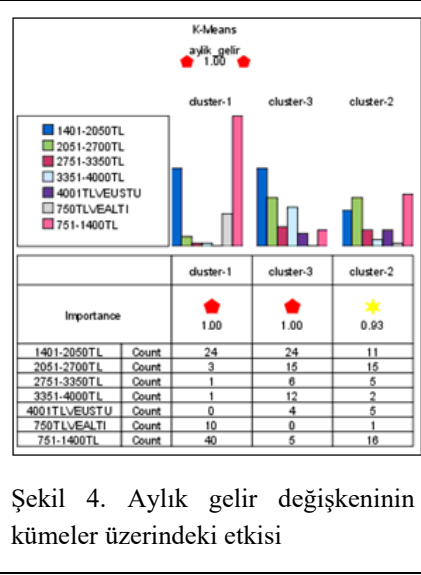
Şekil 2. K-Ortalamalar yöntemi ile elde edilen kümeler

Bundan sonraki aşamada, değişkenlerin her birinin kümeler üzerindeki etkileri ayrı ayrı incelenmiştir. İlk olarak araç değişkeninin kümeler üzerindeki etkisi dikkate alındığında, bu değişkenin 1. ve 3. kümeler için önemli iken, 2. küme için ayırt edici bir niteliğe sahip olmadığı Şekil 3'te görülmektedir. Buna göre birinci kümedeki müşterilerin %91,14'ü araç sahibi değilken, üçüncü kümedeki müşterilerin %59,09'unda araç mevcuttur. Birinci kümede toplam 79, ikinci kümede 55 ve üçüncü kümede 66 kişi bulunmaktadır.

Aylık gelir değişkeni açısından kümeler incelendiğinde, bu değişkenin her üç küme için de önemli olduğu Şekil 4'te görülmektedir. Buna göre birinci kümedeki müşterilerin çoğunun aylık geliri 751- 1400 TL arasındadır. Bu gelir aralığına sahip kişilerin birinci kümedeki oranı %50,63'tür. İkinci kümenin aylık gelir oranları dikkate alındığında, müşterilerin %29,09'unun aylık gelirinin 751- 1400 TL aralığında, %27,27'sinin ise 2051- 2700 TL arasında olduğu görülmektedir. Üçüncü kümedeki müşterilerin aylık gelirleri %36,36'lık oranla 1401- 2050 TL arasındadır. Birinci kümede aylık geliri 4001 TL ve üstünde olan müşteri, üçüncü kümede ise aylık geliri 750 TL ve altında olan müşteri bulunmamaktadır.



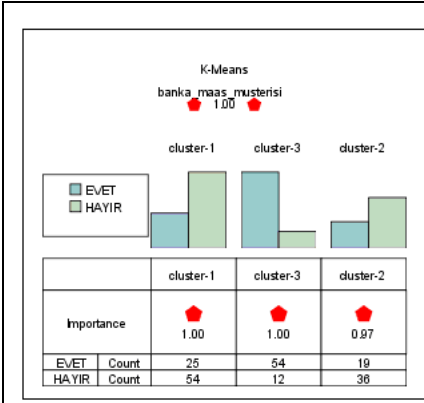
Şekil 3. Araç değişkeninin kümeler üzerindeki etkisi



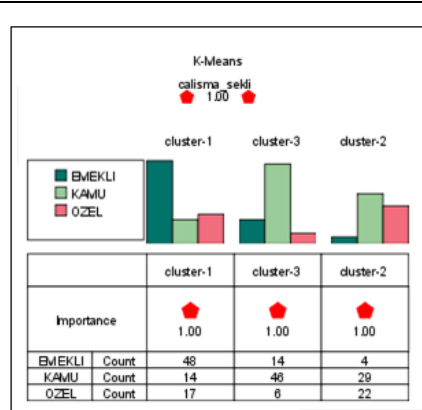
Şekil 4. Aylık gelir değişkeninin kümeler üzerindeki etkisi

Banka maaş müşterisi olup olmaması açısından kümeler incelendiğinde, bu değişkenin üç küme üzerinde de önemli derecede etkili olduğu görülmektedir. Şekil 5'ten birinci ve ikinci kümelerin sırasıyla, %68,35 ve %65,45'lik oranla farklı bankalardan maaş alan kredi müşterilerinden, üçüncü kümenin ise %81,82'lik oranla kredi kullandıkları bankadan maaş alan müşterilerden oluştuğu görülmektedir.

Çalışma şekli değişkeninin kümeler üzerindeki etkisi incelendiğinde, bu değişkenin üç küme açısından da önem arz ettiği görülmektedir. Birinci kümenin %60,76 oranla emekli müşteri ağırlıklı olduğu, ikinci ve üçüncü kümelerin ise sırasıyla, %52,73 ve %69,7 oranında kamu çalışanlarından oluştuğu Şekil 6'da görülmektedir.



Şekil 5. Banka maaş müşterisi olma değişkeninin kümeler üzerindeki etkisi



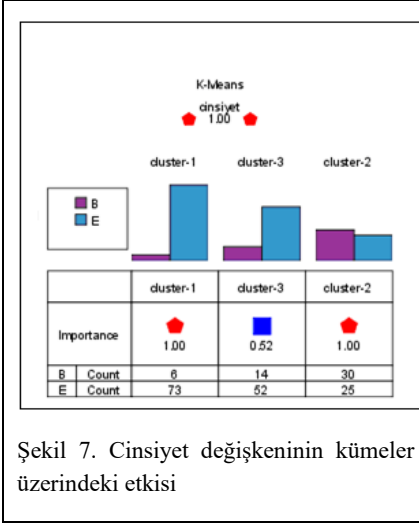
Şekil 6. Çalışma şekline ait değişkenin kümeler üzerindeki etkisi

Şekil 7’de görüldüğü gibi cinsiyet değişkeni birinci ve ikinci kümeler için ayırt edici niteliğe sahipken, üçüncü kümeyi diğer kümelerden ayıracak bir etkisi olmadığından, bu küme için önemli bir unsur değildir. Birinci küme %92,41 oranla erkek, ikinci küme ise %54,55 oranla bayan müşterilerden oluşmaktadır.

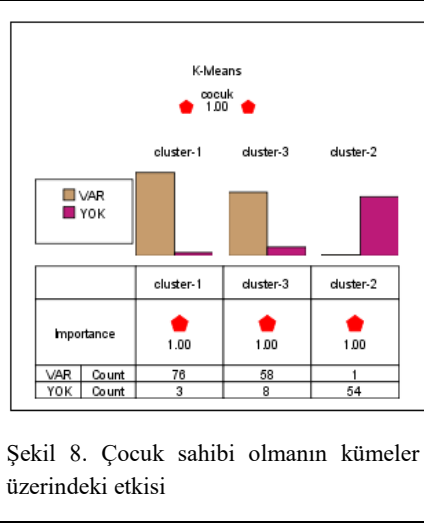
Çocuk sahibi olma durumunun üç küme için de önemli olduğu Şekil 8’de görülmektedir. Birinci ve üçüncü kümelerdeki müşterilerin sırasıyla %96,2 ve %87,88’i çocuk sahibi iken, ikinci kümedeki müşterilerin %98,18’u çocuk sahibi değildir.

Eş geliri üç kümeyi de etkileyen önemli bir değişkendir. Şekil 9’a göre birinci ve ikinci kümelerdeki müşterilerin sırasıyla %87,34 ve %90,91’lik kesiminde eş geliri mevcut değilken, üçüncü kümedeki müşterilerin %69,7’si eş gelirine sahiptir.

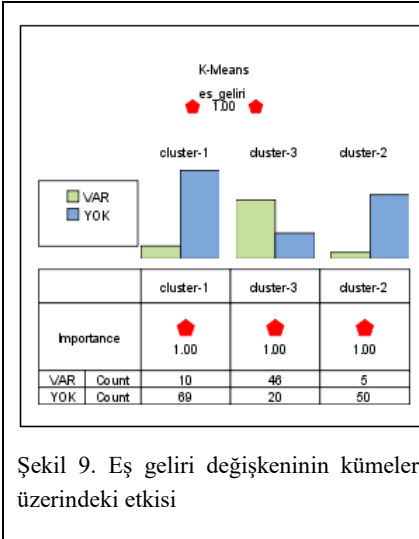
Ev sahibi olma değişkeni ikinci ve üçüncü kümeler için önemli iken, birinci küme için önem arz etmemektedir. Şekil 10 incelendiğinde birinci, ikinci ve üçüncü kümelerdeki müşterilerin sırasıyla %49,37, %7,27 ve %68,18’inin kendilerine ait evlerinin mevcut olduğu görülmektedir.



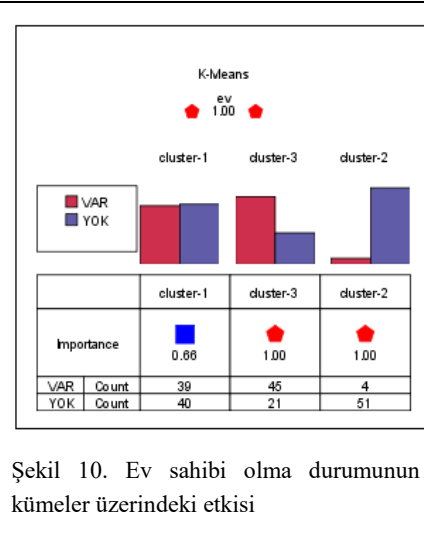
Şekil 7. Cinsiyet değişkeninin kümeler üzerindeki etkisi



Şekil 8. Çocuk sahibi olmanın kümeler üzerindeki etkisi



Şekil 9. Eş geliri değişkeninin kümeler üzerindeki etkisi



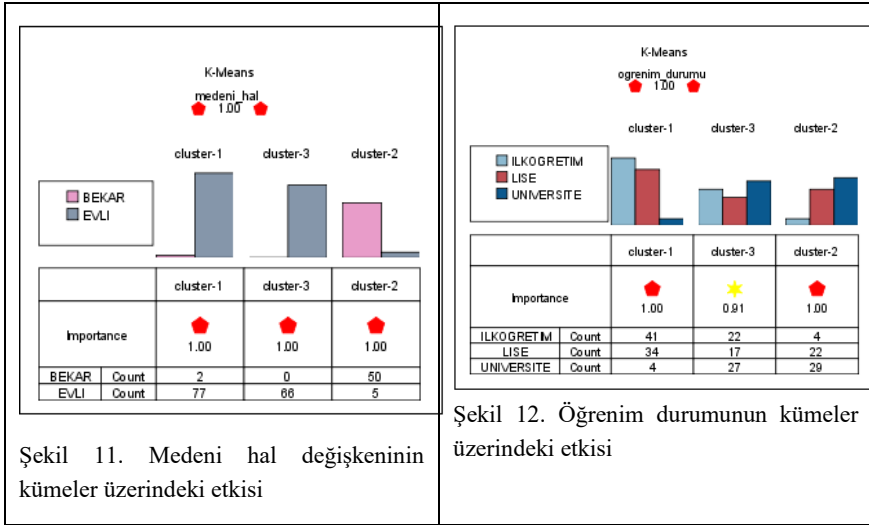
Şekil 10. Ev sahibi olma durumunun kümeler üzerindeki etkisi

Şekil 11'e göre medeni hal değişkeninin her üç küme üzerindeki etkisi önemli düzeydedir. Birinci kümede %97,47 oranla evli, ikinci kümede ise %90,91 oranla bekar müşteriler çoğunluktadır. Üçüncü kümenin ise tamamı evli müşterilerden oluşmaktadır.

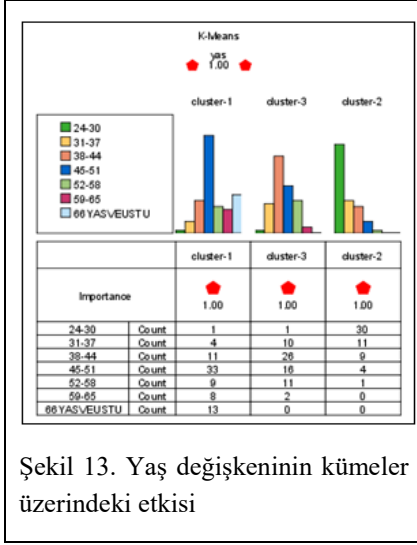
Öğrenim durumu değişkeninin üç küme için de önemli olduğu Şekil 12'de görülmektedir. Birinci kümede %51,9'luk oranla ilköğretim

mezunu müşteriler, ikinci ve üçüncü kümelerde ise %52,73 ve %40,91’lik oranla üniversite mezunu müşteriler çoğunluktadır.

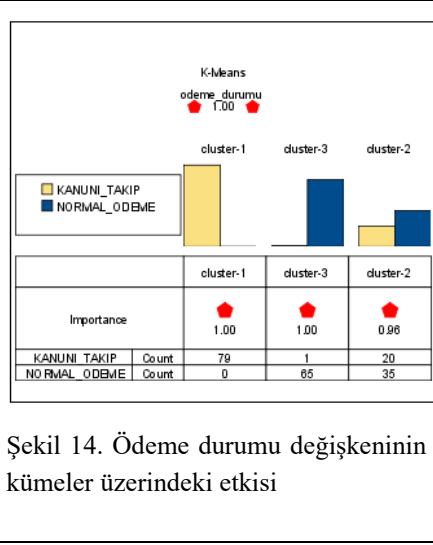
Şekil 13’te yaş değişkeninin kümeler üzerindeki etkisi incelendiğinde, bu değişkenin üç küme için de ayırt edici bir özelliğe sahip olduğu görülmektedir. Birinci kümeyi oluşturan müşterilerin %41,77’lik çoğunluğu 45-51 yaş aralığındadır. İkinci kümede %54,55 oranla 24-30 yaş aralığındaki müşteriler çoğunluktadır. Bu kümede 59-65 yaş aralığı ile 66 yaş ve üstünde müşteri bulunmamaktadır. Üçüncü kümede ise %39,39’luk oranla 38-44 yaş aralığındaki müşterilerin çoğunlukta olduğu görülmektedir. Bu kümede de 66 yaş ve üzerinde müşteri mevcut değildir.



Son olarak ödeme durumu değişkeni açısından kümeler incelendiğinde, bu değişkenin üç küme için de önem arz ettiğini söylemek mümkündür. Şekil 14’te görüldüğü gibi, birinci kümenin tamamı kredi ödemelerinde sorun yaşamış ve kanuni takibe düşmüş müşterilerden oluşurken, ikinci ve üçüncü kümeler sırasıyla %63,64 ve %98,48’lik çoğunlukla kredi geri ödemelerinde sorun yaşamamış müşterilerden oluşmaktadır.



Şekil 13. Yaş değişkeninin kümeler üzerindeki etkisi



Şekil 14. Ödeme durumu değişkeninin kümeler üzerindeki etkisi

Değişkenlerin kümeler üzerindeki etkileri incelendiğinde; çalışma şekli, çocuk sahibi olma durumu, eş geliri, medeni hal ve yaş değişkenlerinin önem derecelerinin üç küme için de “1.00” olduğu, yani bu değişkenlerin kümeleri ayırt etmede %100 önemli olduğu görülmektedir

5. SONUÇLAR VE ÖNERİLER

Bir karar verici için verilen kararın doğruluğu, karar vericinin yeteneklerine, deneyimine ve bilgi birikimine olduğu kadar veri kümesinin yeterliliğine de bağlıdır. Diğer bir ifadeyle kararın başarısında, verilerin doğru depolanması, doğru sınıflanması, doğru ayıklanması, doğru işlenmesi ve doğru yorumlanması çok önemli bir rol oynamaktadır. Günümüzde, karar süreçlerinin karmaşıklaşması sayısal olarak daha fazla veriye gereksinimi ortaya çıkarmış, bu durum ise veri depolarının manuel olarak kontrol edilemeyecek kadar büyük boyutlara ulaşmasına neden olmuştur [33]. Veri yığınlarının geniş yer kaplaması ve büyük miktardaki verilerin yararlı bilgilere dönüştürülmesi ihtiyacı veri madenciliği kavramını doğurmuştur [19]. VM, genellikle büyük ölçüdeki veri setlerindeki

bilinmeyen veya gizli kalmış kuralların keşfine ve analizine yarayan yöntemler ve teknikler kümesidir. Kısaca VM, veriden bilgi çıkarma sanatıdır [34]. Bu yönü VM’ni diğer karar verme yöntemlerinden ayırmaktadır.

Günümüzün tüketici odaklı pazarlarında işletmeler süreklilik arz eden yoğun bir rekabetin içindedirler. İşletmelerin bu rekabet şartlarında başarılı olabilmeleri için etkin ve düşük maliyetli pazarlama stratejileri uygulamaları gerekmektedir. Etkin pazarlama stratejilerinin oluşturulabilmesi için doğru bilgilere, doğru bilgilerin elde edilebilmesi için ise verileri çok boyutlu analiz edebilen VM gibi araçlara gereksinim duyulmaktadır [35]. Günümüzde karar verme sürecine ihtiyaç duyulan birçok alanda olduğu gibi bankacılık sektöründe de VM uygulamaları yaygın olarak kullanılmaktadır.

Çalışmada VM tekniklerinden KA ile mevcut müşterilerin farklı kümelere ayrılarak değerlendirilmesi amaçlanmıştır. Elde edilen veriler ışığında kümelerdeki müşteri profilleri incelendiğinde, birinci kümenin çoğunlukla 45-51 yaş aralığında, kendilerine ait evleri, araçları olmayan, aylık gelirleri 751-1400 TL aralığında olan maaşını farklı bankalardan alan ve ilköğretim mezunu emekli erkek müşterilerden oluştuğu görülmektedir. Bu kümedeki müşterilerin tamamı kredi ödemelerini aksatarak kanuni takibe düşmüştür. Bu kişilerin tekrar kredi kullanabilmeleri için merkez bankası kayıtlarındaki olumsuzlukların giderilmesi gerekmektedir. Bunun için de ortalama beş yıllık bir süreye ihtiyaç vardır. Birinci kümedeki müşterilerin, bu süre sonunda tekrar kredi talep etmeleri durumunda başvuruları olumsuz değerlendirilebilir ya da risk oranını azaltmak için ipotek veya kefil talep edilebilir.

İkinci küme genellikle 24-30 yaş aralığındaki kamu ve özel sektör çalışanı bekâr müşterilerden oluşmaktadır. Diğerlerinin aksine bu kümede bayanların sayısı erkeklere göre daha fazladır. Dikkat edilecek diğer bir husus da bu kümedeki müşterilerin %92,73’lük kısmının kendilerine ait evlerinin olmaması ve %63,64’lük

çoğunluğun normal ödeme durumuna sahip olmasıdır. Bu durumda taksitlerini geciktirmeyen müşterilerin konut kredisi kullanmaları teşvik edilebilir.

Üçüncü küme ele alındığında, bu kümenin çoğunlukla aylık gelirleri 1401-2050 TL aralığında olan, maaşlarını kredi kullandıkları bankadan alan, 38-44 yaş aralığındaki ev ve araç sahibi, kamu çalışanı, emekli erkek müşterilerden oluştuğu görülmektedir. Bu kümedeki müşterilerin büyük bir kısmının eş geliri mevcuttur ve %98.48'i ödemelerini düzenli bir şekilde gerçekleştirmektedir. Bu kümedeki müşteriler özel müşteri olarak değerlendirilebilir ve bu müşterilere internet bankacılığı, hesap işletim ücreti olmayan yatırım hesabı, döviz hesabı, herhangi bir kart aidatı olmayan kredi kartları, HGS cihazları ile kaza, deprem, yangın gibi durumları içeren sigorta ürünlerinin çapraz satışları yapılabilir. Ayrıca, yeniden kredi talep etmeleri durumunda özel faiz indirim politikaları uygulanabilir. Bu şekilde müşterilerin ilgili bankaya olan bağlılıklarını da korumak mümkündür. Bu durum bankanın karlılık oranını artırmada önemli bir etkidir. Çünkü yeni bir müşteri kazanmak, eski müşterileri elde tutmaktan çok daha maliyetlidir.

KAYNAKLAR

- [1] Tosun T., 2006. “Veri Madenciliği Teknikleriyle Kredi Kartlarında Müşteri Kaybetme Analizi”, *Yüksek Lisans Tezi*, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü.
- [2] Aşan Z., 2007. “Kredi Kartı Kullanan Müşterilerin Sosyo Ekonomik Özelliklerinin Kümeleme Analizi ile İncelenmesi”, *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 17, 256-267.
- [3] Çakır Ö., 2008. “Veri Madenciliğinde Sınıflandırma Yöntemlerinin Karşılaştırılması: Bankacılık Müşteri Veri Tabanı Üzerinde Bir Uygulama”, *Doktora Tezi*, Marmara Üniversitesi, Sosyal Bilimler Enstitüsü.
- [4] Doğan B., 2008. “Bankaların Gözetiminde Bir Araç Olarak Kümeleme Analizi: Türk Bankacılık Sektörü için Bir Uygulama”, *Doktora Tezi*, Kadir Has Üniversitesi, Sosyal Bilimler Enstitüsü.
- [5] Albayrak A.S., 2009. “Türkiye’de Yerli ve Yabancı Ticaret Bankalarının Finansal Etkinliğe Göre Sınıflandırılması: Karar Ağacı, Lojistik Regresyon ve Diskriminant Analizi Modellerinin Bir Karşılaştırması”, *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14(2) 113-139.
- [6] Bilen H., 2009. “Bankacılık Sektöründe Personel Seçimi ve Performans Değerlendirilmesine İlişkin Veri Madenciliği Uygulaması”, *Yüksek Lisans Tezi*, Gazi Üniversitesi, Fen Bilimleri Enstitüsü.
- [7] Çil F., 2010. “Banka Yatırım Fonu Müşteri Hareketlerinin Belirlenmesine Yönelik Bir Veri Madenciliği Uygulaması”, *Yüksek Lisans Tezi*, Gazi Üniversitesi, Fen Bilimleri Enstitüsü.
- [8] Taşkın Ç., Emel G. G., 2010. “Veri Madenciliğinde Kümeleme Yaklaşımları ve Khonen Ağları ile Perakendecilik Sektöründe Bir Uygulama”, *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 15(3), 395-409.

- [9] Savaş S., Topaloğlu N., Yılmaz M., 2012. “Veri Madenciliği ve Türkiye’deki Uygulama Örnekleri”, *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 11(21), 1-23.
- [10] Çınar A., Silahtaroğlu G., 2012. “Veri Madenciliği Teknikleri ile Müşteri Memnuniyetine Etki Eden Gizli Nedenlerin Keşfi”, *Marmara Üniversitesi İ.İ.B. Dergisi*, 33(2), 309-330.
- [11] Yakut E., Elmas B., 2013. “İşletmelerin Finansal Başarısızlığının Veri Madenciliği ve Diskriminant Analizi Modelleri ile Tahmin Edilmesi”, *Afyon Kocatepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 15(1), 261-280.
- [12] Dolgun M. Ö., Ersel D., 2014. “Doğrudan Pazarlama Stratejilerinin Belirlenmesinde Veri Madenciliği Yöntemlerinin Kullanımı”, *İstatistikçiler Dergisi: İstatistik&Aktüerya*, 7, 1-13.
- [13] Duran A. E., Pamukcu A., Bozkurt H., 2014. “Comparison of Data Mining Techniques for Direct Marketing Campaigns”, *Sigma Journal of Engineering and Natural Sciences*, 32, 142-152.
- [14] Doğan O., 2015. “Bir E-Ticaret Sitesi Kullanıcı Hesaplarında Şifre Yapılarının Birliktelik Kuralları ile İncelenmesi”, *İnternet Uygulamaları ve Yönetimi Dergisi*, 6(2), 49-61.
- [15] Yaşar A., 2016. “Olumlu Görüş Dışındaki Denetim Görüşlerinin Veri Madenciliği Yöntemleriyle Tahminine İlişkin Karar ve Birliktelik Kuralları”, *Mali Çözüm Dergisi*, 133, 81-109.
- [16] Kaygın Y. C., Tazegül A., Yazarkan H., 2016. “İşletmelerin Finansal Başarılı ve Başarısız Olma Durumlarının Veri Madenciliği ve Lojistik Regresyon Analizi İle Tahmin Edilebilirliği”, *Ege Akademik Bakış*, 16(1), 147-159.
- [17] Jacobs P., 1999. “Data Mining: What General Managers Need to Know”, *Harvard Management Update*, 4(10), 8-9.
- [18] Linoff G.S., Berry M.J.A., 2011. “Data Mining Techniques for Marketing, Sales and Customer Relationship Management”, *Wiley*, 3rd.ed. , Canada.

- [19] Han J., Kamber M., 2006. “Data Mining: Concepts and Techniques”, *Morgan Kaufmann*, 2nd ed., USA.
- [20] Ching W.K., Ng K. M., 2002. “Advances in Data Mining and Modeling”, *World Scientific*, 1st ed., Hong Kong, China.
- [21] Chien C.-F., Chen L.-F., 2008. “Data Mining to Improve Personnel Selection and Enhance Human Capital: A Case Study in High-Technology Industry”, *Expert Systems with Applications*, 34(1), 280-290.
- [22] Akpınar H., 2000. “Veritabanlarında Bilgi Keşfi ve Veri Madenciliği”, *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, cilt 29(1), 1-22.
- [23] Suner A., Çelikoğlu C.C., 2010. “Toplum Tabanlı Bir Çalışmada Çoklu Uygunluk Analizi ve Kümeleme Analizi ile Sağlık Kurumu Seçimi”, *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 25, 43-55.
- [24] Atbaş A.C., 2008. “Kümeleme Analizinde Küme Sayısının Belirlenmesi Üzerine Bir Çalışma”, *Yüksek Lisans Tezi*, Ankara Üniversitesi, Fen Bilimleri Enstitüsü.
- [25] Çakmak, Z., Uzgören, N., Keçek, G., 2005. “Kümeleme Analizi Teknikleri ile İllerin Kültürel Yapılarına Göre Sınıflandırılması ve Değişimlerinin İncelenmesi”, *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 12, 15-36.
- [26] Özdamar K., 2004. “Paket Programlar ile İstatistiksel Veri Analizi II”, *Kaan Kitabevi*, 5. Baskı, Eskişehir.
- [27] MacQueen, J. B., 1967. “Some Methods for Classification and Analysis of Multivariate Observations”, *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1, pp. 281-297.
- [28] Larose D.T., 2005. “Discovering Knowledge in Data: An Introduction in Data Mining”, *Wiley*, 1st ed., USA.
- [29] Pang-Ning Tan, P. N., Steinbach, M., Kumar, V., 2006. “Introduction to Data Mining”, *Pearson Addison Wesley*.

- [30] Işık M., Çamurcu A. Y., 2010. “K-Means ve Aşırı Küresel C-Means Algoritmaları ile Belge Madenciliği”, *Marmara Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 22, 1-18.
- [31] Akbulut S., 2006. “Veri Madenciliği Teknikleri ile Bir Kozmetik Markanın Ayrılan Müşteri Analizi ve Müşteri Segmentasyonu”, *Yüksek Lisans Tezi*, Gazi Üniversitesi, Fen Bilimleri Enstitüsü.
- [32] Çalış A., Boyacı A., Baynal K., 2015. “Data Mining Application in Banking Sector with Clustering and Classification Methods” *International Conference on Industrial Engineering and Operations Management (IEOM), Dubai, UAE, pp. 1-8.*
- [33] Yaralıoğlu K., 2005. “Veri Madenciliği”,
http://www.deu.edu.tr/userweb/k.yaralioglu/dosyalar/ver_mad.doc,
(Erişim Tarihi: 20.05.2016).
- [34] Tuffery S., 2011. “Data Mining and Statistics for Decision Making”, *Wiley*, 1st ed., USA.
- [35] Emel G. G., Taşkın Ç., 2005. “Veri Madenciliğinde Karar Ağaçları ve Bir Satış Analizi Uygulaması”, *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, 6(2), 221-239.