







Sigortacılık Sektöründe Müşteri İlişki Yönetimi İçin Kümeleme Analizi

Using Clustering Analysis for Customer Relationship Management in Insurance Sector

Buket Doğan¹ , Ali Buldu¹ , Önder Demir¹ , Bahar Erol Ceren² 

¹Marmara Üniversitesi Teknoloji Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul, Türkiye

²Intellica - İş Zekası Yazılım ve Danışmanlık, İstanbul, Türkiye

Öz

Teknolojik ilerlemeler insanlar arası ilişkileri etkilediği kadar şirketlerin müşterileri ile olan ilişkilerini, satış ve pazarlama sistemlerini, hatta tüm kurumsal organizasyonlarını etkilemiştir. Müşteri ilişkileri yönetimi kavramı, müşteri odaklı bir yaklaşımla uzun dönemli ilişki kurarak şirketin karlılığını arttırmayı hedefleyen bir yaklaşımdır. Bu kapsamda birbirine benzer nitelikte müşterilerin özelliklerini tanımlamak ve gruplandırmak önemli bir faaliyettir. Bu amaçla farklı kaynaklardan alınan müşterileri verileri bir araya getirilir ve bu müşterilerin karakteristik özelliklerini belirlemek için analizler yapılır. Veri madenciliği (VM) bu karakteristik özelliklerin tespitinde kullanılabilecek veri analizi yöntemlerini içeren bir disiplindir. Bu çalışmada, Türkiye’de faaliyet gösteren bir sigorta şirketinin müşterilerine ait veriler VM’nin en çok kullanılan kümeleme algoritmalarından k-means algoritması ile analiz edilmiştir. Bu analiz ile elde edilen sonuçlar yardımıyla, şirketin benzer müşterilerinin özelliklerini tespit etmesi ve onlara uygun yeni pazarlama stratejileri geliştirebilmesi hedeflenmektedir.

Anahtar Kelimeler: k-means algoritması, Kümeleme, Müşteri ilişki yönetimi

Abstract

Technological advances in relations with customers of the company not only affect relationships between people, but also sales and marketing systems and the entire enterprise organizations. Customer relationship management aims to establish a long-term relationship with a customer-oriented approach for increasing the profitability of the company. In this context, it is important to define and group the characteristics of similar customer attributes. For this purpose, data analysis is done on the combined data of customers from different data sources to determine the similar customer characteristics. Data mining is a discipline that involves data analysis methods used in the determination of these characteristics. In this study, the data of customers of an insurance company operating in Turkey were analyzed by the most widely used algorithms k-means clustering algorithm. The obtained results with this analysis identify the similar feature of the company’s customers and they are useful to develop appropriate and targeted new marketing strategies.

Keywords: k-means algorithm, Clustering, Customer relationship management


1. Giriş


Müşteri ilişki yönetimi (MİY), müşteriyi bir birey olarak tüm özellikleri ile tanıyarak, her müşteriye özelliğine göre farklı davranmak ilkesi üzerine kurulan bir pazarlama yaklaşımıdır. MİY, müşteriye verilen değerinin artışını ve değerli müşterilerin yaşam boyu sadık kalmasını sağlayarak satın almaya yönelmesini sağlayan doğru araçları kullanmayı mümkün kılan bir alt yapıdır. Müşterileri belli segmentlere


aydırdıktan sonra bu segmentlere göre ürün ve hizmet sunmak, istek ve ihtiyaçlarını belirleyerek müşteri memnuniyet ve sadakatini sağlamak MİY’in en önemli hedefidir. Müşterilere, işletmeye sağladığı mevcut ve potansiyel değere uygun olarak farklı davranış sergilenirse, bu durum işletmeyi müşteri gözünde daha farklı bir konuma getirir. Farklı müşterilere farklı davranışta bulunabilmek için, öncelikle bu müşterilere ait verileri kayıt etmek ve sonrasında analizini gerçekleştirmek, yani müşteriyi tanımlamak gerekmektedir. (Demir ve Kırdar 2007, Dyche 2002, Kırım 2001, Özilhan 2010).


İşletmeler müşterilerini tanıyıp, kârlı ve kârsız müşterilerini ayırabildikleri ölçüde müşterilerine bekledikleri pazarlama hizmetini daha iyi bir şekilde verebileceklerdir. MİY

*Sorumlu yazarın e-posta adresi: buketb@marmara.edu.tr

Buket Doğan  orcid.org/0000-0003-1062-2439

Ali Buldu  orcid.org/0000-0002-8508-3065

Önder Demir  orcid.org/0000-0003-4540-663X

Bahar Erol  orcid.org/0000-0003-1129-0149

rakiplerden farklı olabilmek için uygulanması gereken yöntemler bütünü ifade eden pazarlama stratejisidir. MİY müşteri odaklı bir şirket yönetme stratejisidir ve müşterinin bağlılık ve memnuniyetini sağlamaya çalışır. Bu strateji ile müşteri ilişkilerini karlı hale getirmek, işletme verimini artırmak, müşteri taleplerini sağlamak amaçlanmaktadır (Kırım 2001, Özdağoğlu vd. 2008).

Veri Madenciliği ise büyük veri ambarlarında tutulan verilerden modeller sayesinde anlamlı bilgileri, ilişkileri ve davranışları ortaya çıkarma süreci olarak tanımlanmaktadır. Bu süreçte, veri içinde önceden bilinmeyen önemli bilgiler ve örüntüler ortaya çıkarılmaktadır.

VM, verilerin içerisindeki desenlerin, ilişkilerin, değişimlerin, düzensizliklerin, kuralların ve istatistiksel olarak önemli olan yapıların keşfedilmesini sağlar. VM'de amaç daha önceden fark edilmemiş veri desenlerini tespit edebilmektir (Han vd. 2011, Kantardzic 2011).

Veri madenciliği MİY uygulamalarında oldukça önemli bir role sahiptir ve küresel ekonomideki değeri her geçen gün yükselmektedir. VM uygulamaları kullanarak büyük şirketlerin veritabanlarında yer alan müşterilere ait verileri anlamlı bilgilere dönüştürülebilmektedir. İşletmelerde, verilerin toplanmasında pazara yönelik stratejik bilgilerin oluşturulmasında, pazarlama planlamasında VM önemli bir araç olarak kullanılmaktadır. Veri madenciliği yöntemleri ile müşterilerin tüketim davranışlarını, benzer müşteri profillerini belirlemek, harcama alışkanlıklarını ortaya çıkarmak mümkün hale gelmiştir (Ngai vd. 2009, Tsipsis ve Chorianopoulos 2011).

2. Literatür Araştırması

Literatürde MİY faaliyetlerinde veri madenciliğinin kullanıldığı çalışmalar kümeleme, sınıflama, birliktelik kuralı analizi gibi farklı yöntemleri içerebilmektedir. Bu başlık altında çalışma konusuyla ilgili önemli çalışmalara yer verilmektedir.

Wang ve Wu (2011) tarafından gerçekleştirilen uygulamada, bankacılık sektöründe müşterilerin finansal ürünler alma sıklığının keşfedilmesi için kredi kartı alışverişleri ile ilgili veri setinde karar ağaçları algoritması kullanılmıştır. Sonuçta müşterilerin demografik özellikleri ile hangi finansal ürünleri tercih ettikleri arasındaki ilişki ortaya konmuştur.

Wei ve arkadaşları (2013) tarafından gerçekleştirilen çalışmada kuaförlük sektörünün hizmet sektöründe giderek daha önemli bir rol oynadığını belirtmiştir. Çalışmalarında SOM (Self Organizing Maps) ve K-means algoritmasını

birleştirerek iki aşamalı kümeleme yöntemi uygulanarak Tayvan'da bir kuaförde müşteri profil analizini gerçekleştirmişlerdir. Modelin analiz sonuçlarına göre kuaförde müşterileri; sadık müşterileri, potansiyel müşteriler, yeni müşteriler ve kayıp müşteriler olarak dört gruba ayırmışlardır. Çalışma yardımı ile sistematik müşteriler profili ve müşteri tipleri oluşturulması, böylece kuaförde müşterilerin türüne uygun benzersiz pazarlama stratejilerinin geliştirilmesi sürecinin daha doğru ilerlemesi sağlanabilmektedir. Bu şekilde değerli müşterileri hedeflemek ve farklı pazarlama stratejileri uygulamak mümkün olmaktadır.

Bayram (2001) ise veri madenciliği tekniklerinden karar ağaçları, kümeleme analizleri ve lojistik regresyon analizlerini kullanarak kablosuz iletişim endüstrisinde müşterilerin segmentasyonunu ve sadık olmayan, şirketten ayrılacak müşterilerin tahmin modelini ortaya koymuştur.

Bahari ve Elayidom değerli müşterileri elde tutmak amacıyla karar alma süreçlerini geliştirerek için müşterilerin davranışlarını tahmin etmek için bir model sunmaktadır. Çalışma bir Portekiz bankasının Mayıs 2008 ve Kasım 2010 tarihleri arasında gerçekleştirdiği 17 kampanyayı içerir. Deneysel veri kümesi, ön işlenmiş veri setinin % 10 kadarını kullanmakta ve 16 giriş değişkeni içermektedir. Ayrıca, iki sınıflandırma modeli müşteri davranışlarını tahmin etmek için kullanılarak doğruluk, duyarlılık ve özgüllük bakımından sınıflandırıcı performansları karşılaştırılmıştır. Sınıflandırma işleminde yapay sinir ağları yönteminin doğruluk değeri Naive Bayes yönteminden daha yüksek çıkmıştır (Bahari ve Elayidom 2015).

Hwang, Jung ve Suh, müşterinin geçmişte yarattığı potansiyel değeri ve rakip şirketi tercih etme olasılığını dikkate alarak, lojistik regresyon, yapay sinir ağları ve karar ağacı tekniklerini kullanmak suretiyle bir müşteri yaşam ömrünü yani elde tutulduğu süreyi bulan bir model geliştirmişlerdir. Model ile müşteri değeri, şimdiki değeri, potansiyel değeri ve müşteri sadakati olarak üç kategoride sınıflandırılmaktadır (Hwang vd. 2004).

3. Gereç ve Yöntem

3.1. K-Means Algoritması

Kümeleme, bir dizi örüntüyü ayrık ve homojen guruplar oluşturacak şekilde gruplandırma işlemidir. Bu işlem örüntülerin benzerlik derecelerine göre sınıflara veya kümeler ayrılmasıyla gerçekleştirilir. Kümeleme işleminin sonunda elde edilen her başarılı kümede yer alan nesnelere arasında maksimum benzerlik ve kümeleme sonucu elde

edilen diğer kümeler arasında maksimum farklılık oluşması sağlanır. Kümeleme analizinin sınıflandırmadan farkı, sınıflandırmada sınıflar önceden belli iken kümelemede sınıfların önceden belli olmamasıdır (Han vd. 2011, Silahtaroglu 2008).

Literatürde kullanılan birçok kümeleme tekniği mevcuttur. Teknikler birbirinden kümelemenin oluşturuluş şekline, kullanılan veri tipine ve yapılacak çalışmanın amacına göre de farklılık gösterirler. Bu çalışmada benzer müşteri kümelerini bulmak amacıyla, bölümlenmeli kümeleme algoritmalarından k-means algoritması kullanılmıştır.

K-means algoritması MacQueen (1967) tarafından geliştirilmiştir. K-means algoritması büyük verilerin kümeleneğinde oldukça etkin bir algoritmadır. Algoritma n boyutlu bir kümedeki tüm veriyi inceleyerek k adet kümeye böler.

Verilen k adet küme için algoritma şu şekilde çalışmaktadır:

- 1) Başlangıç: k adet küme için küme merkezleri rasgele olarak belirlenir
- 2) Her veri noktasının merkezlere olan uzaklıkları kullanılan uzaklık ölçütüne göre hesaplanır ve kendisine en yakın olan kümeye atanır.
- 3) Atamaların sonucunda küme merkezi yeniden hesaplanır.
- 4) Her veri noktasının ait olduğu küme; değişiyorsa 2. Adıma git ve devam et, değişmiyorsa algoritmayı sonlandır.

K-means algoritmasında uzaklık ölçütü olarak genellikle öklit uzaklığı kullanılırken, algoritmanın değerlendirilmesinde en yaygın olarak karesel hata kriteri SSE(sum of squared error) kullanılmaktadır. En düşük SSE değerine sahip kümeleme sonucu en iyi sonucu verir. Nesnelerin buldukları kümenin merkez noktalarına olan uzaklıklarının karelerinin toplamı aşağıdaki formülle hesaplanmaktadır(Kantardzic 2011).

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

x : C_i kümesinde bulunan bir nesne

m_i : C_i kümesinin merkez noktası

Bu kriterler sonucu k tane kümenin olabildiğince yoğun ve birbirinden ayrı sonuçlanması hedeflenmektedir. Algoritma, karesel-hata fonksiyonunu azaltacak k parçayı belirlemektedir (Kantardzic 2011).

3.2. CRISP-DM Modeli

Sigortacılık sektörü üzerine gerçekleştirilen uygulamada farklı sektörlerle ve iş problemlerine uygulanabilen CRISP-DM(Cross-Industry Standard Process for Data Mining)

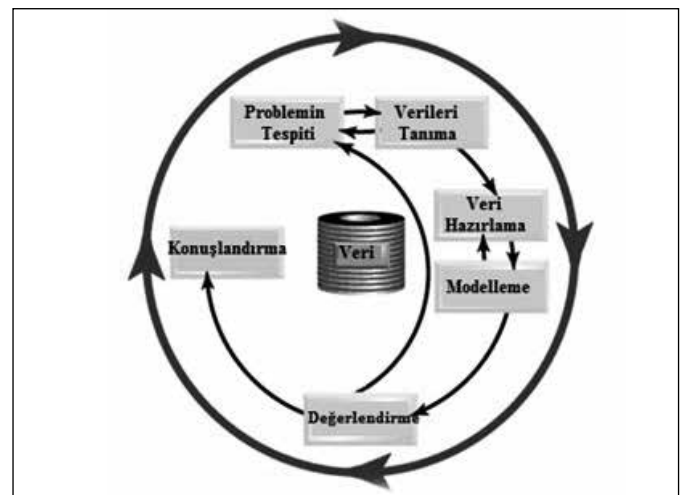
modelindeki adımlar izlenmiştir. CRISP-DM modeli VM sürecinde izlenecek adımları sunmaktadır. CRIPS-DM modeli, Şekil 1'de görüldüğü gibi VM sürecinde işlenecek olan adımların sunulduğu bir çevrimden oluşmaktadır (Chapman vd. 2000).

CRISP-DM referans modelinin fazları altı aşamadan meydana gelmektedir. Bu aşamalar araştırma sürecinde adım adım uygulanarak veri madenciliği işlemi gerçekleştirilmiştir.

Bu yöntemde öncelikle problemin tespit edilmesi, ardından verilerin tanımlanması ve verilerde saklı olabilecek bilgiler için ilk izlenimlerin oluşturulmasını gerektirmektedir. Ardından ham veriden başlayarak son veriye kadar yapılması gereken düzenlemeleri içeren veri hazırlama aşaması ise veri hazırlama, tablo, kayıt, veri dönüşümü ve modelleme araçları için temizleme gibi alt adımları içermektedir. Modelleme aşaması ise, veri madenciliği analiz modelinin seçilmesi ve bu modele ilişkin parametrelerin optimal değerlere ayarlanması gibi çalışmaların tümü yer almaktadır. Uygun modelleme tekniğinin seçimi, test tasarımının üretimi, model geliştirme ve tahmin işlemlerini içermektedir. Değerlendirme aşaması elde edilen modelin iş hedeflerini başarma konusundaki yeterliliğinin karşılaştırıldığı ve veri madenciliği sonuçlarının nasıl kullanılacağına karar verileceği aşamadır. Konuşlandırma aşamasında ise analist tarafından modelin ve sonuçlarının sunulması eylem önerilerinde bulunulması mümkündür.

3.3. Kümeleme Uygulaması

Verileri tanıma aşamasında öncelikli olarak sigorta şirketinden veriler sağlanmıştır ve kullanılan VM yazılımları araştırılmıştır. Uygulamada kullanılan veri



Şekil 1. CRISP-DM Modeli (CRISP-DM; Chapman ve ark., 2000)

kümesi Türkiye'nin önemli sigorta şirketlerinden birinin farklı tarih aralıklarında satışı gerçekleştirilmiş olan bireysel poliçe verilerinden oluşmaktadır. Verilerin saklanması için bir veritabanına ihtiyaç duyulmuş ve fonksiyonel veritabanı yönetim sistemlerinden birisi olan Oracle Veritabanı (Oracle Database Enterprise Edition 11g) kullanımına karar verilmiştir.

Ardından veri hazırlama aşamasında müşteri gizlilik prensiplerine bağlı kalınarak müşteri kimlik bilgileri ve poliçe numaraları belirli algoritmalar kullanılarak değiştirilmiştir. Müşterilerin kişisel bilgilerini saklamak için isim bilgileri gizlenmiş ve müşteriyi tanımlayan bir alan olan "müşteri_id" ile analizler gerçekleştirilmiştir.

"GENELHAYAT" tablosunda Hayat Poliçeleri satın alan müşterilerin demografik özellikleri tutulmaktadır.

Veri setindeki sütunlarda verilerin eksik değerlerinin bulunup bulunmadığının tespiti ve değişkenlerin dağılımının incelenmesi için "Data Audit" aracı kullanılmıştır. Bu araç, veri kalitesinin belirlenmesi ve eksik değerlerin tespitine olanak sağlar. Bu işlemin sonucunda kümeleme analizinde kullanılmak üzere müşterilerin demografik özelliklerini içeren "kümeleme_analizi" adı altında veri seti oluşturulmuştur. Bu veri setinde 3662 satır ve beş alan belirlenerek bu veri ile kümeleme işlemi yapılmasına

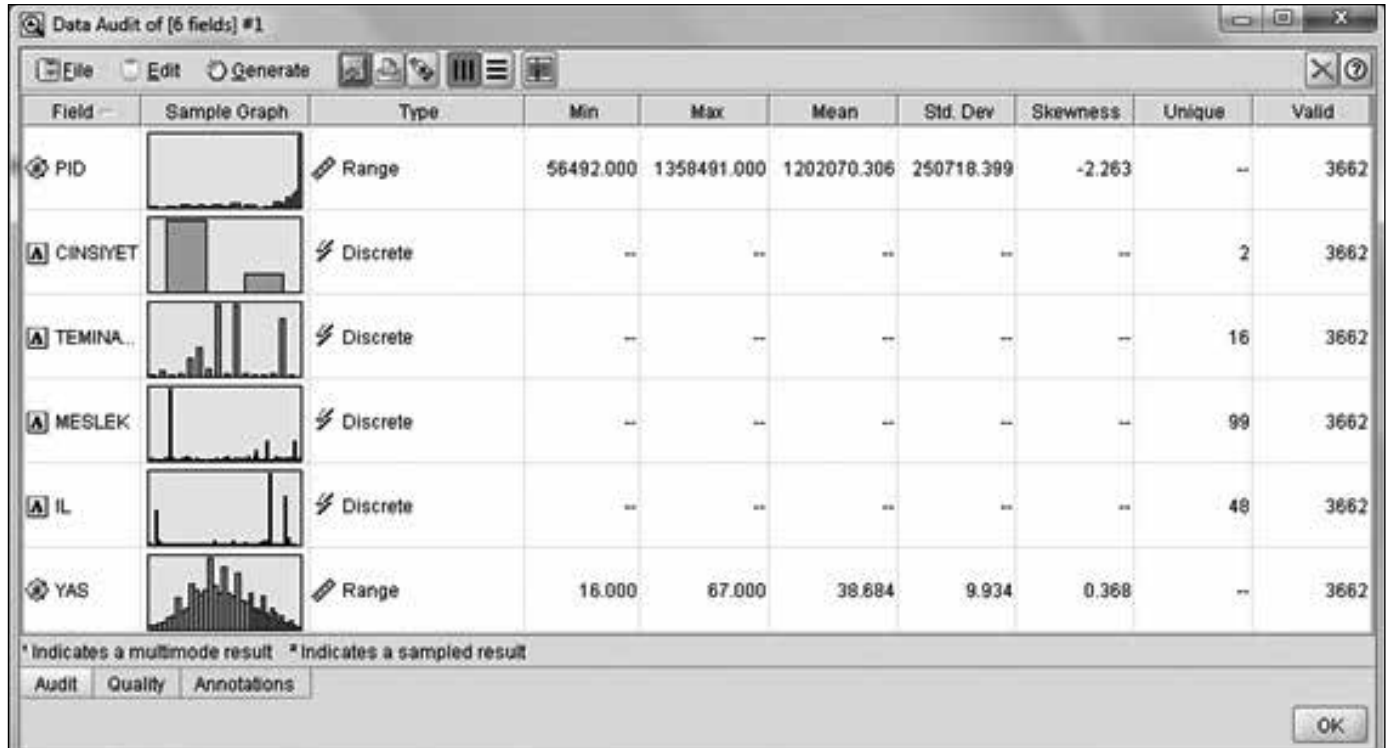
karar verilmiştir. Kümeleme analizinde kullanılan beş alan; müşteri cinsiyeti, müşterinin yaşadığı il, meslek, yaş ve satın alınan teminatlar niteliklerinden oluşmaktadır. Modelin kurulması aşamasında bu alanlar üzerinde sonuca yönelik bilgi aranmıştır.

Verilerin dönüştürülmesi aşamasında Clementine programının "Derive" aracı kullanılarak müşterilerin doğum tarihi alanından yaş değerinin hesaplanması ve kullanılması sağlanmıştır.

Şekil 2'de kümeleme veri setinin Data Audit aracı ile elde edilen analiz sonuçları verilmiştir.

Şekil 2'deki Data Audit ekranında her bir değişkenin adı, türü, ortalaması, standart sapması, en büyük ve en küçük değerleri gibi istatistiksel sonuçlar görülmektedir. PID(Person Id) alanı müşteri id'sini temsil etmektedir. Bu yüzden aşağıda analiz sonuçlarını incelerken müşteri id alanı anlamlı değer ifade etmeyeceğinden dolayı dikkate alınmamıştır. Kümeleme analizi yapılmadan önce SQL sorgusu oluşturulurken kümeleme veri setindeki meslekler alanı alınırken yapılan analizlere göre oranı %1,85'den az olan meslek grupları seçime dâhil edilmemiş "Diğer" olarak isimlendirilmiştir.

Modelleme aşamasında gerçekleştirilen denemeler sonucunda en optimum sonucu veren K-Means algoritması seçilirken küme sayısı (Number of clusters) üç olarak seçilmiştir.



Şekil 2. Kümeleme veri seti analizi.

K-Means algoritmasının sonuçları incelendiğinde önem derecesi %95'den büyük olan değişkenler önemli(important) değişken adı altında görülmektedir. Değişkenin önemli veya önemsiz(unimportant) olmasını gösteren önem değeri, değişkenin kümeler arasındaki olasılığındaki varyasyonuna göre değişiklik gösterir. Önem değeri; 1 değerinden, istatistiksel testin anlamlılık değeri çıkartılarak bulunmaktadır. Anlamlılık değeri, kategorik değerler için ki-kare testi (chi-squares test), sürekli değişkenlerde t-testi değeri ile belirlenmektedir. Şekil 3'de K-Means sonucunda elde edilen sonuçların değişkenler üzerindeki önem değerleri görülmektedir. Bu sonuçlara göre, birinci küme(Cluster-1)'de meslek, müşteri cinsiyeti, müşteri il, satın alınan teminatlar ve yaş değişkenlerinin önemli değişkenler olduğu görülmektedir.

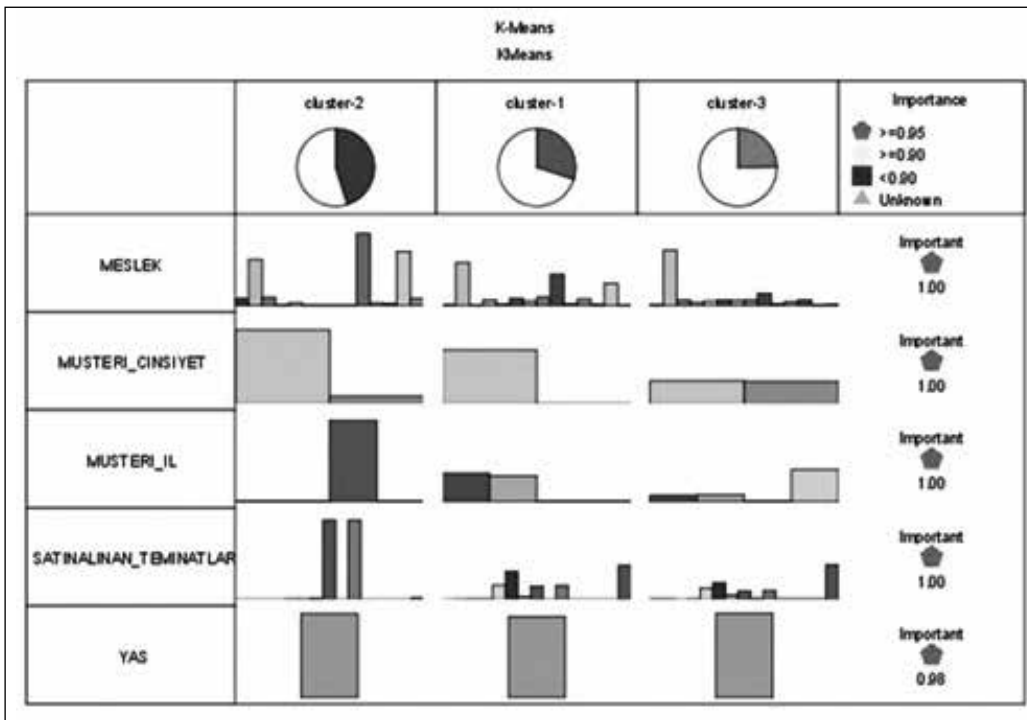
K-means kümeleme analizi ile elde edilen üç küme için özet modeli Şekil 4'deki gibidir.

MUSTERI_CINSIYET değişkeninin birinci küme-1(Cluster-1) üzerine etkisi incelendiğinde erkek müşterilerden oluştuğu gözlenmektedir. MUSTERI_IL değişkeni incelendiğinde Ankara ili çoğunlukta görülmektedir. Satın alınan teminatların birinci küme üzerine etkisi incelendiğinde en fazla satın alınan teminat Vefat Teminatı tercih edenlerden oluşmaktadır. İkinci en çok tercih edilen teminat ise Deprem Dahil Kaza Sonucu Vefat teminatıdır.

İkinci küme (cluster-2)'ye etki eden MUSTERI_IL, MUSTERI_CINSIYETİ, MESLEK ve SATINALI-

NAN_TEMINATLAR değişkenlerinden oluşmaktadır. Şekil 4'de görüldüğü gibi MUSTERI_IL değişkeninin Küme-2(Cluster-2) üzerindeki etkisi incelendiğinde "Trabzon" şehrinde yaşayanlardan oluşmakta olduğu gözlenmektedir. MUSTERI_CINSIYETİ değişkeninin Küme-2(Cluster-2) üzerine etkisi incelendiğinde Erkek(E) müşterilerin çoğunlukta olduğu görülmektedir. MESLEK değişkeni incelenecek olursa, Tüccar'lardan oluştuğu gözlenmektedir. SATINALINAN_TEMINATLAR değişkeninin Küme-2(Cluster-2) üzerine etkisi incelenecek olursa, üzere teminatlarda Kaza Sonucu Maluliyet ve Kaza Sonucu Vefat teminatı ilgi görmektedir.

K-Means yöntemi ile elde edilen üçüncü Küme-3 (cluster-3)'ye etki eden değişkenler incelendiğinde, Şekil 4'de CINSIYET değişkeninin küme-3(cluster-3) üzerine etkisi, erkekler ile kadınların sayıları yaklaşık olarak birbirine eşittir. MUSTERI_IL değişkeni incelendiğinde, İstanbul'da yaşayanlar çoğunluğu oluşturmaktadır. Sonuçlar SATINALINAN_TEMINATLAR değişkeni bazında incelendiğinde, tercih edilen teminat türü Vefat Teminatının çoğunlukta olduğu görülmektedir. MESLEK değişkeninin Küme-3(Cluster-3) üzerine etkisi incelendiğinde En çok DİĞER olan meslek grubu görülmektedir. Diğer meslek grubu, çalışmamızda sayıları az olan farklı mesleklerden oluşmaktadır. Bu yüzden uygulamanın sonucunda anlamlı gibi gözükse de, aslında anlamlı bir ifade değildir. YAS



Şekil 3. K-Means yöntemi ile elde edilen kümeler.

değişkeninin Küme-3(Cluster-3) üzerine etkisi incelendiğinde ortalama 39 yaş olan müşterilerden oluşturduğu gözlenmektedir.

4. Sonuçlar

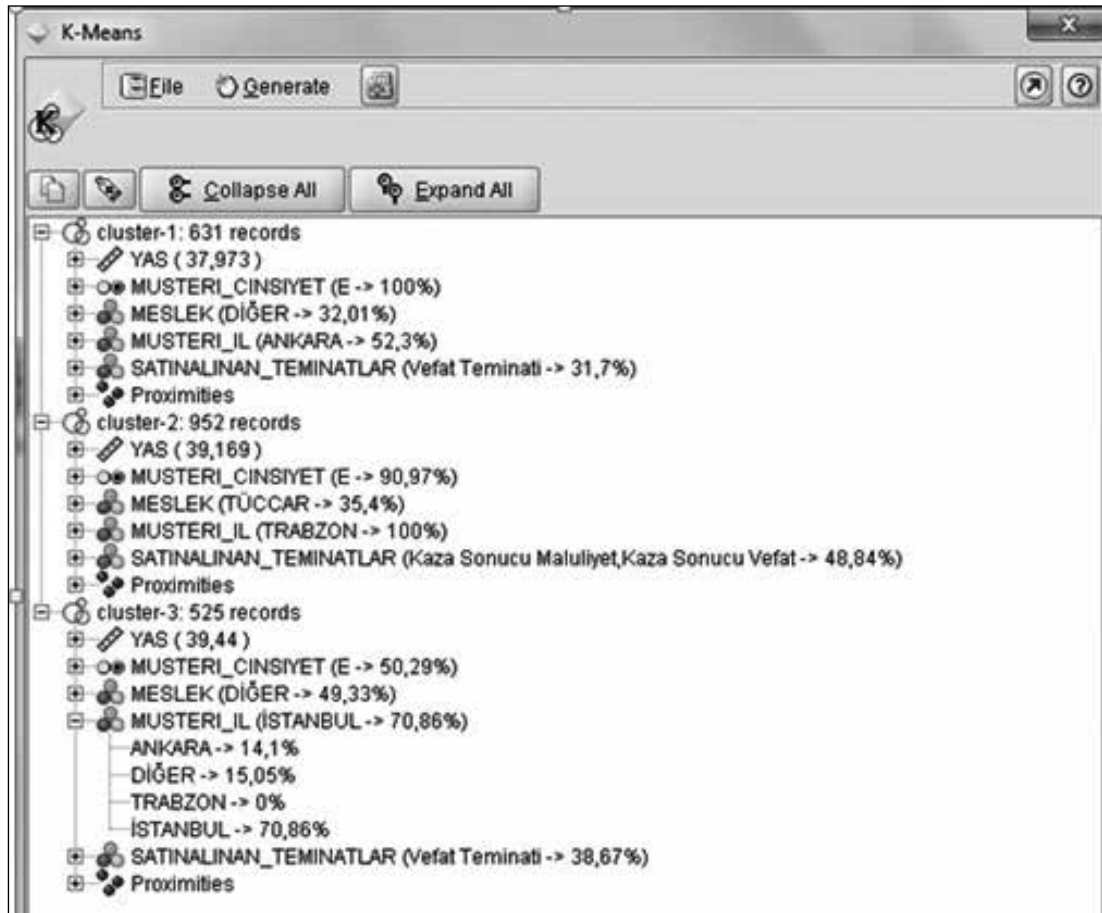
Elde edilen verilerin K-means algoritması ile kümelmesi sonucunda birbirine benzer müşteri gruplarının oluşturulması sağlanmıştır. Böylece sigorta şirketinin müşterileri daha yakından tanınması ve verileri anlamlandırması sağlanmıştır. Değişkenlerin kendi içinde kümeler üzerine etkisi incelendiğinde aşağıda belirtilen sonuçlara ulaşılmıştır.

Elde edilen üç küme içinde müşteri cinsiyeti önemli değişkendir. Şekil 5(a)'da ve Şekil 5(b)'de görülen count değerleri ilgili kümeden o değişken değerine sahip kaç adet kayıt olduğunu göstermektedir. Şekil 5(a)'da görüldüğü gibi her üç kümede de Erkek sayısı çoğunlukta olmakla birlikte, birinci kümenin tamamı erkeklerden oluşmaktadır. Üçüncü kümede ise Erkekler ile Kadınların sayıları birbirine yakındır. Cinsiyet göz önüne alınarak bir satış kampanyası yapılacağı zaman ikinci kümedeki müşterilerin demografik özelliklerinden faydalanılabilir.

Şekil 5(b)'de daha detaylı bir şekilde görüldüğü gibi, il her üç küme içinde önemli bir değişkendir. Birinci kümenin çoğunluğu Ankara ilinde yaşayanlardan, İkinci kümenin tamamı Trabzon ilinde yaşayanlardan, Üçüncü kümenin çoğunluğu ise İstanbul'da yaşayanlardan oluşmaktadır. Bu sonuçlara istinaden şehir bazında kampanya düzenlenmek istendiğinde gruplarındaki müşterilerin özellikleri göz önüne alınarak daha çok müşteri çekilmesi sağlanabilmektedir.

Şekil 6(a)'da görüldüğü gibi Satın alınan teminatların kümeler üzerine etkisi incelendiğinde, birinci kümede en çok tercih edilen teminatların başında Vefat Teminatı ve Deprem Dahil Kaza Sonucu Vefat Teminatı gelmektedir. İkinci kümede çoğunluk Kaza Sonucu Maluliyet ve Kaza Sonucu Vefat Teminatlarını satın alan müşterilerden oluşmaktadır. Üçüncü Kümede ise Vefat Teminatı en çok tercih edilen üründür.

Şekil 6(b)'de Meslek değişkeninin kümeler üzerindeki etkisi incelendiğinde birinci küme-1(cluster-1)'de Diğer adı altına toplanan meslek grupları sonucu çıkmıştır. Fakat diğer meslek grupları az olan, birden fazla meslekten meydana geldiği için değerli gibi gözükse de aslında değersiz bir ifadedir. Küme-



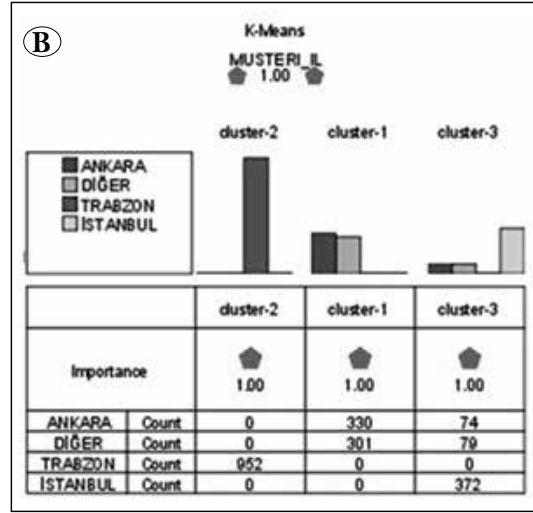
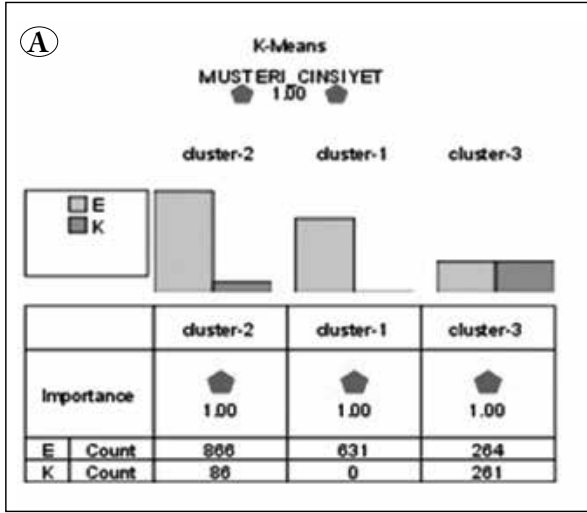
Şekil 4. K-Means analizi model özeti.

1'de yer alan ikinci en büyük meslek grubunda ise, 147 değeri ile Sigortacılık mesleği yer almaktadır. İkinci küme-2(cluster-2)'de ise Tüccarlar çoğunlukta olup, son olarak üçüncü küme(cluster-3)'de Diğer meslek grubu ön plana çıkmıştır. Bu sonuçlara oranla Birinci küme(cluster-1)'de bulunan en çok sayıda meslek grubu Sigortacılık Sektöründe çalışanlardan, ikinci küme(cluster-2)'de ise Tüccar meslek grubunda olanlar anlamlı bir sonuç ifade etmektedir.

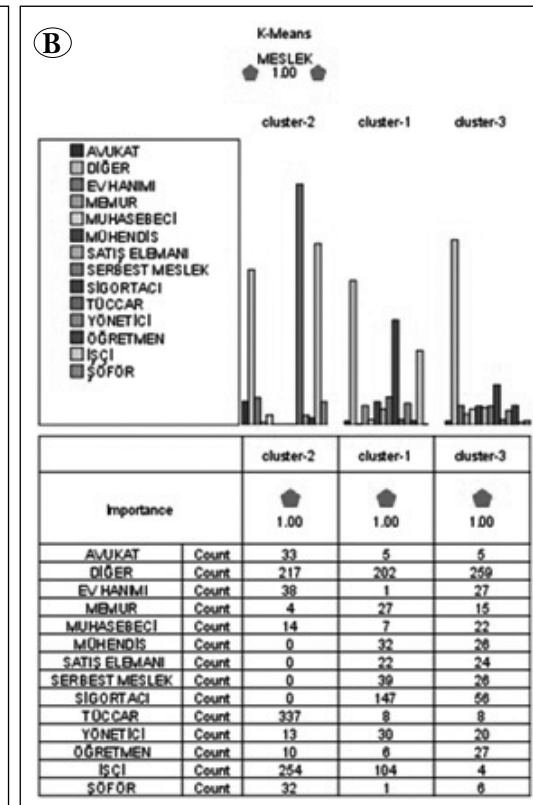
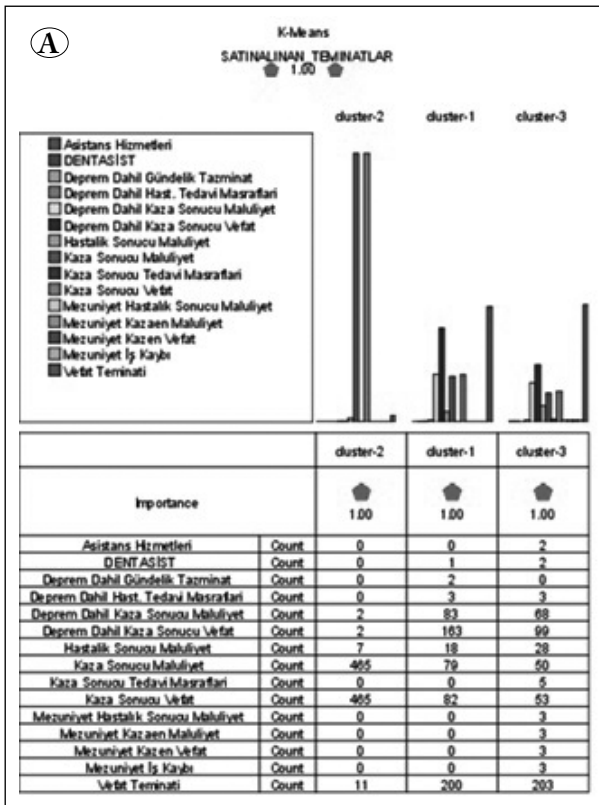
Şekil 7'de görüldüğü gibi yaş değişkeni önem değeri 0.92 ile sadece üçüncü küme için önemli ayırt edici özelliği olan bir değişkendir. Yaş değişkeni birinci ve ikinci küme için düşük önem değerine sahip ayırt edici özelliği düşük olan bir değişkendir.

5. Değerlendirme

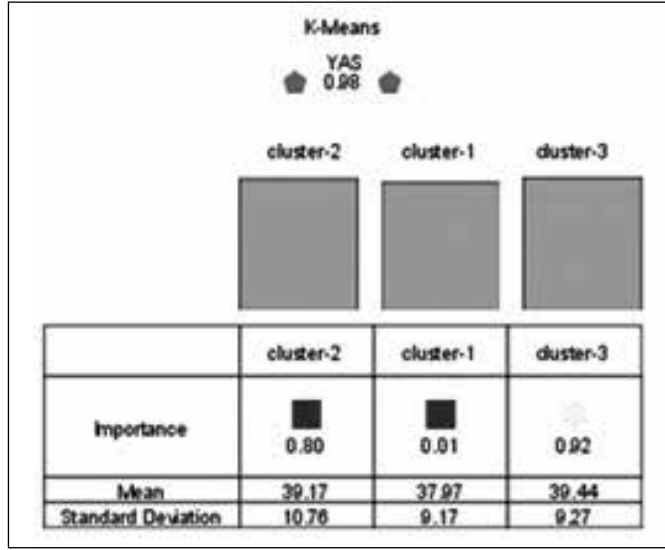
Elde edilen sonuçlara göre sigortacılık sektöründe veri madenciliği analizleri veri yapısına göre incelenerek uygu-



Şekil 5. A) Müşteri cinsiyeti değişkeninin kümeler üzerindeki etkisi B) Müşteri ili değişkeninin kümeler üzerindeki etkisi.



Şekil 6. A) Satın alınan teminatların kümeler üzerindeki etkisi B) Meslek değişkeninin kümeler üzerindeki etkisi.



Şekil 7. Yaş değişkeninin kümeler üzerine etkisi.

lanabilmektedir. Bunun yanı sıra şirketlerin veritabanlarında tutulan verilerden yola çıkılarak, bilgi keşfi yapılmış olup, veritabanında gizli kalmış verilerden bilgi açığa çıkartılmaktadır. Bunun sonucunda değerli ve değersiz müşterilerin farkına varılmaktadır.

Bu yolla, müşterilerin kuruma karşı sadakati sağlanırken değerli müşterileri özellikleri de keşfedilerek hangi özelliklere sahip bireylere satış kampanyalarının düzenlenmesi gerektiği hakkında bilgi sahibi olunmuştur. Örneğin elde edilen birinci küme sonuçlarını inceleyecek olursak; müşterilerin cinsiyeti erkekler olan, Ankara'da yaşayanların çoğunlukta olduğu ayrıca en çok satın alınan teminatları Vefat Teminatı ve Deprem Dahili Kaza Sonucu Maluliyet teminatlarının tercih edildiği görülmektedir. Bu sonuçlara göre örneğin, Ankara şehrindeki acenteliğinde daha çok müşteri çekmek amacıyla en çok tercih edilen ürünlerde kampanyalar düzenlenerek daha çok müşteriye ulaşılması ve satışların artırılması sağlanabilir. Aynı şekilde diğer kümeler için de hangi müşterilerin ne tür poliçe tercih ettiği sonucundan yola çıkılarak, o tür müşterilere odaklanılarak daha çok müşteri elde edilmesi sağlanabilir. Bölgesel kampanyaların kapsamı, daha çok hangi özelliklere sahip müşterilere uygun olarak yapılması konusunda da kümeleme analizi yol gösterici olmaktadır. Örneklerden de görüldüğü gibi daha çok müşteri çekmek amacıyla elde edilen bilgilerden yararlanarak çeşitli satış kampanyaları düzenlenebilir. Böylece elde edilen müşterilerin devamlılığını sağlamanın yanı sıra, daha çok müşteriye ulaşılması sağlanmış olur.

Ayrıca farklı kümeleme yöntemleri ile çalışmada kullanılan verilerin analiz edilmesi ve k-means kümeleme yöntemiyle

elde edilen sonuçlarının karşılaştırılması gelecek dönem çalışması olarak planlanmaktadır.

4. Kaynaklar

- Bahari, TF., Elayidom, MS. 2015.** An efficient CRM-data mining framework for the prediction of customer behaviour. *Procedia Comput. Sci.*46: 725-731.
- Bayram, E. 2001.** Customer Segmentation and Churn Modeling In Wireless Communications. *Yüksek Lisans Tezi, Boğaziçi Üniversitesi Fen Bilimleri Enstitüsü, İstanbul:* 1-5.
- Chapman, P., Clinton, J., Keber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. 2000.** CRISP-DM 1.0 Step by step DM guide. Edited by SPSS.
- Demir, FO., Kırdar, Y. 2007.** Müşteri ilişkileri yönetimi: CRM. *Review of Socia., Econ. & Bus. Std.* 8: 293-308.
- Dyche, J. 2002.** The CRM handbook: a business guide to customer relationship management: Addison-Wesley Professional.
- Han, J., Pei, J., Kamber, M. 2011.** Data mining: concepts and techniques: Elsevier.
- Hwang, H., Jung, T., Suh, E. 2004.** An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. *Expert sys. with app.* 26: 181-188.
- Kantardzic, M. 2011.** Data mining: concepts, models, methods, and algorithms: John Wiley & Sons.
- Kırım, A. 2001.** Strateji ve Birebir Pazarlama CRM. *Sistem Yayıncılık, İstanbul.*
- MacQueen, J. 1967.** Some methods for classification and analysis of multivariate observations. In *Proc. of the fifth Berkeley symposium on mathematical statistics and probability*, Vol. 1, pp 281-297.
- Ngai, EW., Xiu, L., Chau, DC. 2009.** Application of data mining techniques in customer relationship management: A literature review and classification. *Expert sys. with app.* 36: 2592-2602.
- Özdağoğlu, A., Özdağoğlu, G., Öz, E. 2008.** Müşteri Sadakatinin Sağlanmasında Müşteri İlişkileri Yönetiminin Önemi: İzmir'de Bir Hipermarket Araştırması. *Atatürk Üniversitesi İkt.ve İdari Bilim. Der.* 22: 367-383.
- Özilhan, D. 2010.** Müşteri İlişkileri Yönetimi (MİY) Uygulamalarının İşletme Performansına Etkileri. *Gümüşhane Üniversitesi Sos. Bilim. Enst. Der.* 1:18:30.
- Tsipsis, KK., Chorionopoulos, A. 2011.** Data mining techniques in CRM: inside customer segmentation: John Wiley & Sons.
- Wang, Y., Wu, DS. 2011.** Research of the Bank's CRM Based on Data Mining Technology. *Com. Inf. Sci. Man. Eng.* 1.
- Wei, JT., Lee, MC., Chen, HK., Wu, HH. 2013.** Customer relationship management in the hairdressing industry: An application of data mining techniques. *Expert sys. with app.* 40: 7513-7518.