



Araştırma Makalesi / Research Article
**CHURN ANALYSIS AND CUSTOMER SEGMENTATION OF A COSMETICS
BRAND USING DATA MINING TECHNIQUES**

Emel KIZILKAYA AYDOĞAN^{*1}, Cevriye GENCER¹, Sinem AKBULUT²

¹Gazi Üniversitesi, Mühendislik Mimarlık Fakültesi, Endüstri Mühendisliği Bölümü, Maltepe-ANKARA

²Türk Telekom A.Ş., ANKARA

Geliş/Received: 20.08.2007 Kabul/Accepted: 24.03.2008

ABSTRACT

Data mining is the process of finding hidden and unknown patterns in huge amounts of data. Data mining has a wide application area such as marketing, banking and finance, medicine and manufacturing. Churn analysis and customer segmentation are two common application areas of data mining. Churn modeling is predicting which customers will leave the company. This allows companies to increase customer loyalty by producing special campaigns. Customer segmentation is the process of dividing customers into mutually exclusive groups. Segmenting customers enables companies to develop customized marketing programs for them. This study aims to determine the customer profile who likely to leave and the customer segments of a cosmetic brand and develop customized campaigns and marketing strategies. Clustering techniques used for segmentation and classification techniques used for determining the churners.

Keywords: Data mining, customer segmentation, classification, clustering.

**VERİ MADENCİLİĞİ TEKNİKLERİ İLE BİR KOZMETİK MARKANIN AYRILAN MÜŞTERİ
ANALİZİ VE MÜŞTERİ BÖLÜMLENMESİ**

ÖZET

Veri madenciliği, büyük veri kümeleri içindeki anlamlı bilgiyi ortaya çıkarma sürecidir. Veri madenciliğinin en yaygın kullanıldığı uygulama alanlarından biri, ayrılma eğilimi gösteren müşteri kesitini belirleme ve müşteri bölümlenmesidir. Ayrılma eğilimi gösteren müşteri kesitini belirleme, şirketlerin bu müşterilere özel pazarlama kampanyalarını geliştirmelerini sağlamaya yöneliktir. Müşteri bölümlenmesi ise benzer özellikler gösteren müşterileri gruplandırarak; gruplara özel, pazarlama programlarının geliştirilmesini sağlar. Yapılan çalışma, bir kozmetik markasının müşteri gruplarını ve ayrılma eğilimi gösteren müşteri kesitini belirleyerek; bu müşterilere özel pazarlama stratejileri geliştirilmesini hedeflemektedir. Bölümlenme için kümeleme teknikleri, ayrılacak müşteri kesitini belirlemek için sınıflama teknikleri kullanılmıştır.

Anahtar Sözcükler: Veri madenciliği, müşteri bölümlenmesi, sınıflandırma, kümeleme.

1. GİRİŞ

Globalleşme ve artan rekabet, müşteri satın alma davranışlarını ve beklentilerini, büyük ölçüde değiştirmiştir. Değişen müşteri davranışlarını, işletme amaç ve hedeflerine uygun olarak yönetebilmek; değişim yönetimi ve müşteri ilişkileri yönetimi kavramlarını anlayabilmek ve

*Sorumlu Yazar/Corresponding Autor: e-mail/e-ileti: ekizilkaya@gazi.edu.tr, tel: (312) 231 70 00 / 2845

doğru bir şekilde yorumlayabilmek ile mümkün olabilecektir. Müşteri ilişkileri yönetiminde pazar payı kavramından, müşteri payı kavramına geçiş söz konusu olmaktadır. Müşteri payı aynı müşteriye birden fazla ürün satabilmeyi ve müşteriye aktif ve sadık bir müşteriye dönüştürmeyi amaçlamaktadır. Bu durum ise tek yönlü pazarlama anlayışından karşılıklı ilişkiye dayalı pazarlamaya geçişi ifade etmektedir. Günümüzde kitlesel pazarlama yöntemleri, giderek müşteriye hitap eden kişisel pazarlama yöntemlerine yönelmektedir.

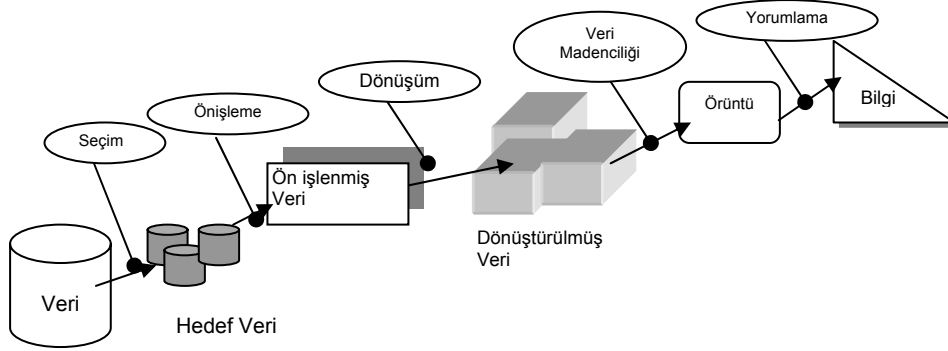
Diğer taraftan bilgisayar teknolojilerindeki gelişmeler, işletmelerin çok miktarda veriyi saklayabilmesini ve işleyerek anlamlı bilgilere dönüştürmesini mümkün hale getirmiştir. Bugün, işletmeler satın alma davranışından demografik özelliklere kadar müşterilerle ilgili pek çok detayı veritabanlarında tutmaktadır. Veri madenciliği tekniklerini kullanarak bu veriler içerisindeki anlamlı ve gizli örüntülerin ortaya çıkarılması mümkün olmaktadır. Veri madenciliği sonuçları müşteri odaklı birçok uygulamaya girdi teşkil etmektedir.

Çalışmanın 2. bölümünde veri madenciliği, müşteri ilişkileri yönetimi ve veri madenciliğinden bahsedilmiş, 3. bölümde yapılan çalışma anlatılmış ve son bölüm olan 4. bölümde sonuç ve önerilerde bulunulmuştur.

2. VERİ MADENCİLİĞİ

2.1. Veri Madenciliğine Genel Bakış

Veri madenciliği, önceden bilinmeyen ilişki ve trendlerin bulunması için bugünün endüstrisinde yaratılan büyük miktarlardaki veriyi analiz eden bir yoldur [1]. Şekil 1'de veri madenciliğinin veri işleme dönüştürme süreci içerisindeki yeri gösterilmektedir [2].



Şekil 1. Veri madenciliğinin veri işleme süreci içindeki yeri [2]

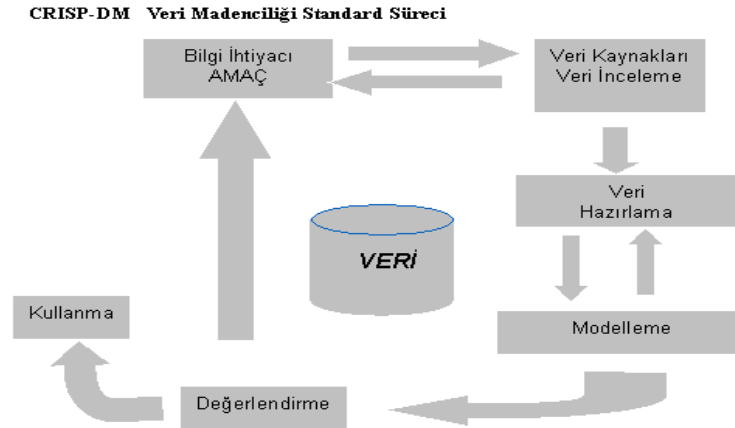
Veri tabanlarından bilgi keşfi, etkileşimli ve tekrarlanan bir süreçtir, aşağıda özetlenen çeşitli adımlardan oluşmaktadır [3]:

- Uygulama tanım kümesini öğrenme: İlişkili ön bilgiyi ve uygulamanın amaçlarını içerir
- Bir hedef veri kümesi oluşturma: Bir veri kümesinin seçimini veya keşfin yapılacağı değişkenlerin veya veri örneklerinin bir alt kümesinde odaklanmayı kapsamaktadır.
- Verilerin temizlenmesi ve ön işleme yapılması: Veri türleri, yöntem, eksik ve bilinmeyen değerlerin eşleştirilmesi gibi veri tabanı yönetim sistemine ait hususların karşılaştırılmasının yanı sıra; uygun ise normal olmayan (gürültülü) veya aykırı değerlerin çıkarılması, gürültüyü modellemek veya açıklamak için gerekli bilginin toplanması, eksik veri alanlarının ele alınması için stratejilerin kararlaştırılması, ardışık zamanlı bilgi ve bilinen değişikliklerin açıklanması gibi temel işlemleri içermektedir.

- Verilerin indirgenmesi: Görevin amacına bağlı olarak verileri temsil etmek için yararlı özelliklerin bulunmasını ve göz önüne alınan değişkenlerin etkin sayısını azaltmak ve veriler için farklı olmayan gösterimler bulmak için boyut indirgeme ve dönüşüm yöntemlerini içerir.
 - Veri madenciliğinin fonksiyonunu seçme: VM algoritması tarafından türetilen modelin amacını (regresyon, kümeleme, sınıflama,özetleme vb.) karşılaştırmayı kapsar.
 - Veri madenciliği algoritmasını seçme: Hangi modellerin ve parametrelerin uygun olabileceğine karar verme gibi verilerdeki örüntüleri arama için kullanılacak yöntemlerin seçilmesini (örneğin kategorik veriler için modeller gerçel sayı vektörlerine dayanan modellerden farklıdır) ve belirli bir VM yönteminin Veri Tabanı Bilgi Keşfi (VTBK) sürecinin bütün kriterleriyle eşleştirilmesini içerir.
 - Veri madenciliği: Sınıflama kuralları veya ağaçları, regresyon, kümeleme, ardışık modelleme, bağımlılık ve doğru analizi gibi belirli bir gösterim biçiminde veya bunların bir kümesinde, ilgilenilen örüntülerin aranmasını kapsar.
 - Yorum: Çıkarılan örüntülerin görselleştirilmesi, gereğinden fazla ve ilişkisiz örüntülerin çıkarılması ve yararlı olanların kullanıcılar tarafından anlaşılabilir ifadelere dönüştürülmesinin yanı sıra, keşfedilen örüntülerin yorumlanması ve muhtemelen önceki adımlardan herhangi birine dönülmesini içerir.
 - Keşfedilen bilginin kullanılması: Bu bilginin daha önceden çıkarılan veya inanılan bilgi ile potansiyel uyumsuzluklarının kontrol edilmesi ve giderilmesine ilaveten, bilginin icra sistemine katılması, bilgiye dayanan eylemlerin gerçekleştirilmesi, basitçe belgelenmesi ve ilgili kişilere rapor edilmesini içerir [3].
- Veri madenciliğinin gereksinimleri ise erişilebilir veri, etkin erişim yöntemleri, açık problem tanımı, etkin algoritmalar, yüksek tabanlı uygulama sunucusu, sonuç oluşturmada esneklik olarak sıralanabilir [3].

2.2. Veri Madenciliği Süreci

Başarılı bir veri madenciliği projelerinde, izlenmesi gereken adımlar; problemin tanımlanması, verilerin hazırlanması, modelin kurulması ve değerlendirilmesi, modelin kullanılması ve modelin izlenmesidir [4]. Veri madenciliği süreci Şekil 2 'de gösterilmektedir.



Şekil 2. Veri madenciliği süreci

Şekil 2'de veri kaynakları veri inceleme adımında problemin tanımlanması yapılır. Projenin hangi işletme amacı için yapılacağını ve elde edilecek sonuçların başarı düzeylerinin

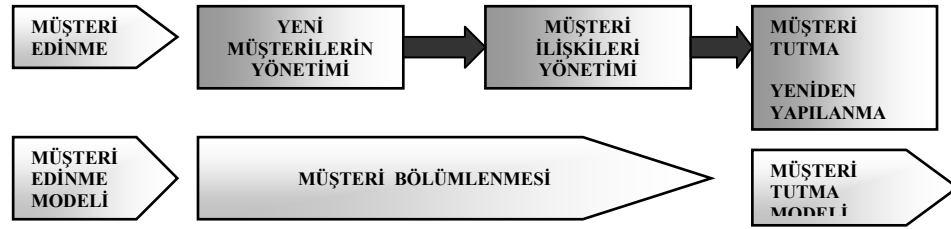
nasıl ölçüleceği tanımlanır. Veri hazırlama adımında toplama, birleştirme ve temizleme, dönüştürme işlemleri yapılarak veri, modelleme adımına uygun hale getirilir. Modelleme adımında, model kuruluş süreci denetimli ve denetimsiz öğrenimin kullanıldığı modellere göre farklılık gösterir. Denetimli öğrenimde sistemin amacı, verilen örneklerden hareket ederek her bir sınıfa ilişkin özelliklerin bulunmasıdır. Denetimsiz öğrenimde, ilgili örneklerin gözlenmesi ve bu örneklerin özellikleri arasındaki benzerliklerden hareket ederek sınıfların tanımlanması amaçlanır. Değerlendirme adımında kurulan modelin doğruluğu test edilir. Kullanma adımında, kurulan ve geçerliliği kabul edilen model doğrudan kullanılabilir veya bir başka uygulamanın alt parçası olabilir. Bütün bu adımlar gerçekleştirildikten sonra kurulan model izlenir.

2.3. Müşteri İlişkileri Yönetimi ve Veri Madenciliği

Günümüzde şirketlerin müşteri ile ilişkileri büyük ölçüde değişmiştir. İşletmelerde başarılı bir müşteri ilişkileri yönetimiyle, küresel bir dünyada ve giderek artan rekabetçi piyasalarda, işletmeler için yaşamsal önem taşıyan, müşteri için değer yaratmak, müşteri sadakati sağlamak ve bu konularda kurumsallaşmayı gerçekleştirmek mümkün olabilecektir [5]. Diğer taraftan yazılım ve donanım teknolojilerindeki gelişmeler, işletmelerin çok miktarda veriyi saklayabilmesini ve işleyerek anlamlı bilgilere dönüştürmesini mümkün hale getirmiştir. Veri madenciliği sonuçları müşteri edinme, müşteri bölümlenmesi, müşteriyi elde tutma, ayrılma eğilimi gösteren müşteri kesitini ortaya koyma, müşteri değerlendirme, kredi derecelendirme ve pazar sepeti analizi gibi pek çok müşteri odaklı uygulamaya girdi teşkil etmektedir [6,7,8].

2.3.1. Müşteri Edinme

Yeni müşteriler edinme, şirketler için mevcut müşterileri elde tutma kadar kritik bir fonksiyondur. Müşteri edinme faaliyetleri ile diğer pazarlama fonksiyonları arasındaki ilişki Şekil 3'de gösterilmektedir.

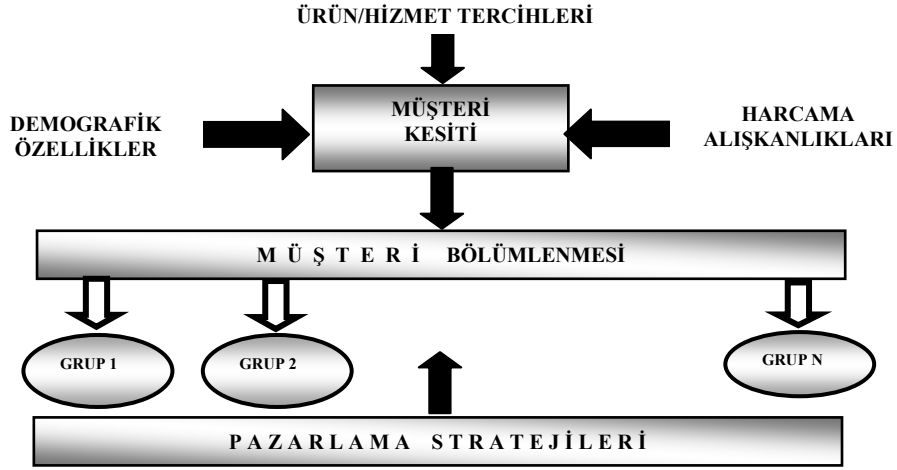


Şekil 3. Müşteri edinme faaliyetleri ve diğer pazarlama fonksiyonları ilişkisi

Şirketler pazar paylarını ancak yeni müşteriler edinerek arttırabilirler. Burada dikkat edilmesi gereken nokta, şirkete kazandırılması gereken müşteri profilinin doğru saptanması ve bu profile uygun eylem planları hazırlanmasıdır.

2.3.2. Müşteri Bölümlenmesi

Müşteri bölümlenmesi ile şirketler geniş ve dağınık pazarları küçük gruplara ayırarak, müşteriye bireysel istek ve ihtiyaçlarına cevap veren ürün ve hizmet sunar, müşteri grupları bazında ürün önerileri oluşturur ve pazarlama programları geliştirirler [8]. Müşteriler; coğrafik, demografik, psikografik ve davranışsal özelliklerine göre gruplandırılırlar [9]. Müşteri bölümlenmesi Şekil 4'de gösterilmiştir.



Şekil 4. Müşteri bölümlemesi

Şekil 4’de görüldüğü gibi müşteri bölümlemesinde ilk adım müşteri profilinin belirlenmesidir. Müşteri profili belirleme sürecinde, müşterilerle ilgili: temel demografik özellikleri nelerdir, harcama alışkanlıkları nasıldır, hangi ürün ya da hizmetleri satın alırlar, müşteri hangi sıklıkla alışveriş yapar, niçin sizin ürün yada hizmetlerinizi tercih ederler, en değerli müşterilerinizle benzer nitelikli müşterileri nasıl elde edersiniz gibi sorulara cevap aranır.

2.3.3. Ayrılan Müşteri Analizi ve Müşteri Tutma

Yeni bir müşteri edinme maliyeti, mevcut müşterileri elde tutma maliyetinin yaklaşık 5 katı, memnun olmayan müşteriyi geri kazanmanın maliyeti mevcut müşteriyi elde tutmanın yaklaşık 10 katıdır [10]. Bu nedenle şirketler, mevcut müşterilerini elde tutmaya yönelik bir takım sadakat programları geliştirirler.

2.3.4. Çapraz Satış

Çapraz satış, genel satın alma eğilimini analiz ederek müşteriye birbiriyle ilişkili ürün veya hizmetleri satma stratejisi olarak tanımlanabilir. Çapraz satışın amacı müşterinin şirketle olan bağımlı güçlendirmek ve rakip şirketlere gidişini engellemektir [11].

2.3.5. Müşteri Yaşam Ömrü Değeri

Müşteri yaşam ömrü değeri, müşterinin bugün, yarın ve ömrü boyunca aynı kaynaktan (marka) yapacağı alışverişler sonucu ortaya çıkacak net değeridir. Müşteri yaşam ömrü değeri analizlerinde dikkate alınan kriterler; müşterinin harcama tutarı ve alışveriş periyodudur [12].

2.3.6. Tepki Modelleme

Pazarlama iletişimi, doğru müşteriye, doğru mesajın, doğru kanallar kullanılarak iletilmesidir [13]. Seçilen iletişim kanalı ile her müşteriye mesaj göndermek yeterli değildir. İletişimin etkisini ölçmek gerekmektedir. Bu ise “tepki analizi” denilen bir yaklaşımla gerçekleştirilmektedir.

3. YAPILAN ÇALIŞMA

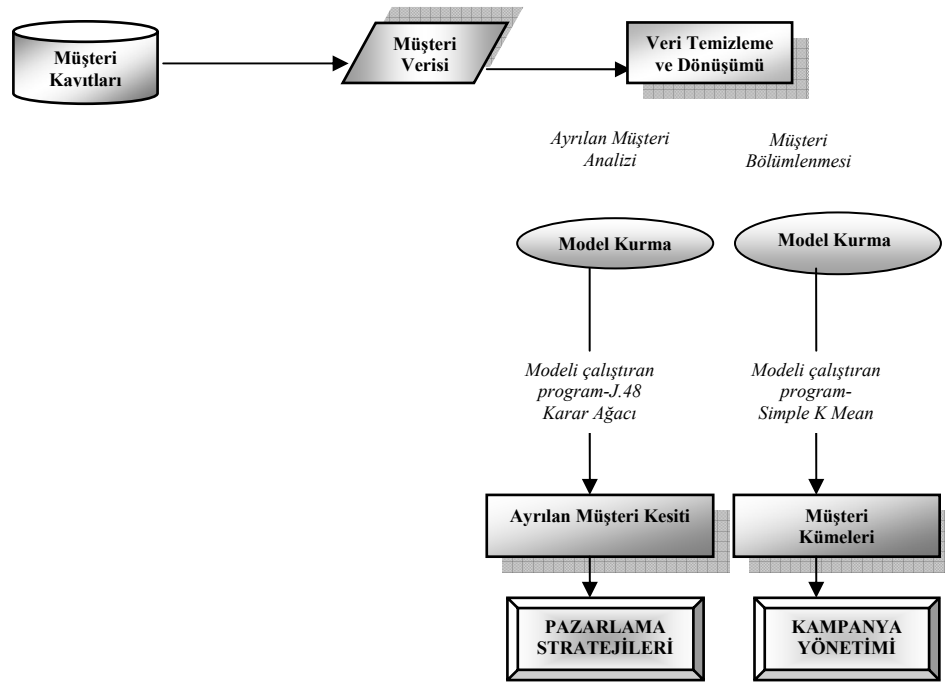
3.1. Problem Tanımı

Çalışmanın amacı, İstanbul yolu üzerinde bulunan bir kozmetik mağazasının sattığı markalardan birinin müşterilerini gruplandırarak, her grubun özelliklerine uygun pazarlama stratejileri geliştirmek ve karlı müşterileri elde tutmaya yönelik bir eylem planı hazırlamaktır.

Mağaza bünyesindeki tüm markalar mevcut müşterilerini elde tutmak için; ücretsiz cilt bakımı yapmakta, belirli bir tutarın üzerinde alışveriş yapan veya belli sayıda ürün satın alan müşterilerine hediye ürün vermekte, indirim günleri düzenlemektedir. Kozmetik mağazası, reklâm ve promosyonlarını yazılı (gazete, dergi) ve görsel (TV) basın aracılığı ile duyurmaktadır. Mağazanın, müşterilerini ücretsiz bakım günleri, kampanya ve promosyonlar hakkında bilgilendirmek için kullandığı iletişim kanalı telefondur.

3.2. Yapılan Çalışmada Veri Madenciliği Süreci

Kozmetik markasının veri madenciliği sürecinin aşamaları Şekil 5’de özetlenmiştir.



Şekil 5. Kozmetik markası için veri madenciliği metodolojisi

3.2.1. Veri temizleme

Veri temizleme, ikinci bölümde de ifade edildiği gibi veri madenciliği sürecinin en kritik ve zaman alıcı adımıdır. Müşteri demografik verisi genellikle yüksek oranlarda kayıp değer içermektedir. Bu tür kategorik verilerdeki kayıp değer problemi çözmek için, *kayıp değeri bir değerle kodlama yöntemi* kullanılır. Bu yöntemde göre, kayıp değerli alana “u” gibi bir değer

atanır. Eğer bir niteliğe ait verilerin büyük bir çoğunluğu eksikse bu nitelik veri tabanından çıkarılmalıdır. Bir niteliği veri tabanından çıkarma kararı alırken sadece kayıp değerlerin toplam içindeki büyüklüğünü değil, aynı zamanda bu niteliğin kayıp değer içerme nedenini de dikkate almak gerekir. Nümerik nitelikler için kayıp değer sorununu çözmek biraz daha zordur. Çünkü çözmeye çalışmak bu niteliğe ait istatistikleri ve veri dağılımını değiştirebilir. Nümerik veri alanlarındaki kayıp değer sorununu çözmek için çeşitli yöntemler söz konusudur:

- Kayıp değerli alanlara, o niteliğe ait diğer değerlerin ortalaması atanır. Bu yöntem, basit olmakla birlikte verinin dağılımı üzerinde azımsanmayacak bir etkiye sahiptir. Bu nedenle, sadece verinin dağılımını minimal şekilde etkileyen durumlarda kullanılmalıdır.
- Mevcut değerler kullanılarak verinin dağılımı elde edilir ve kayıp değerli alanlara bu dağılıma uygun olarak değer atanır. Bu yöntem, verinin dağılımını çok fazla değiştirmez. Fakat kurulan modelde değer atanan değişken çok önemli ise veri madenciliği sonuçlarını etkileyecektir.

Müşteri verilerinde kategorik alanlar için kayıp değer söz konusu değilken; HARCAMA TUTARI adlı nümerik alanda kayıp değerler yer almaktadır. Kayıp değer problemini çözmek için “Structured Query Language” (SQL) komutları kullanılmıştır [14]. Nümerik kayıp değerler için yazılan SQL sorgusuyla bilinmeyen değerlerin yerine bu niteliğe ait bilinen değerlerin ortalaması atanmıştır.

```
update veritabanı HARCAMA TUT 1=ORT_HRC1 where HARCAMA TUTARI 1 is null;  
update veritabanı HARCAMA TUT n= ORT_HRC2 where HARCAMA TUTARI n is null;  
update veritabanı HARCAMA TUT n= ORT_HRCn where HARCAMA TUTARI n is null;
```

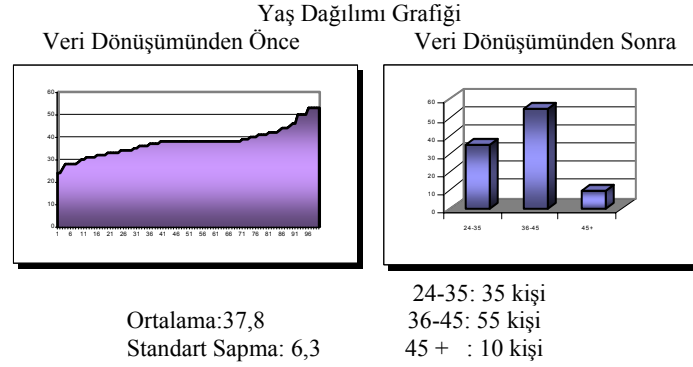
3.2.2. Veri Dönüştürme

Veri temizlemeden sonraki adım veri dönüştürmedir. Bu veri dönüştürme işleminde uzman görüşü olarak nitelendirebileceğimiz mağaza personelinden satılan ürünlerin hangi yaş aralıklarına göre hangi gruba dahil edilebileceği, hangi semtlerin yakınlık olarak bir bütün olarak düşünülebileceği, ortalama alışveriş periyotları bilgisi alınmış buna göre orijinal veri, kayıp değerlerden arındırıldıktan sonra aşağıdaki dönüşümler yapılmıştır.

1. SQL komutları ile doğum tarihi alanı yaş aralıklarına dönüştürülmüştür.

- *update veritabanı set YAS=datepart(year,getdate()) datepart(year,DOG.TAR.)*
- *update veritabanı YAS= '24-35' where YAS<=35 and YAS >24*
- *update veritabanı YAS= '36-45' where YAS<=45 and YAS >35*
- *update veritabanı YAS= '45+' where YAS>45*

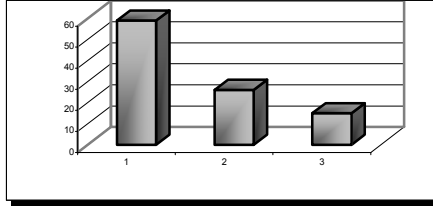
Şekil 6’da veri dönüşümünden önce ve sonraki yaş dağılımı grafikleri gösterilmektedir.



Şekil 6. Yaş dağılımı (Age distribution)

2. Müşterilerinin ikamet ettikleri 11 değişik semt, programın efektif çalışmasını ve doğru sınıflandırmalar yapmasını sağlamak amacıyla 3 temel grupta toplanmış, bunun için aşağıdaki SQL komutları kullanılmıştır.

- *update veritabanı SEMT= '1' where SEMT='Batıkent' OR SEMT='Yenimahalle' OR SEMT='Ostim' OR SEMT='Demetevler'*
 - *update veritabanı SEMT= '2' where SEMT='Eryaman' OR SEMT='Etimesgut' OR SEMT='Sincan'*
 - *update veritabanı SEMT= '3' where SEMT='Kolej' OR SEMT='Emek' OR SEMT='Kavaklıdere' OR SEMT='Bahçelievler'*
- Müşterilerin semtlere göre dağılımı Şekil 7'de gösterilmektedir.



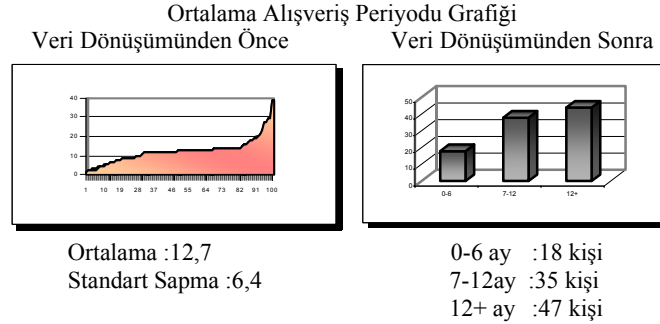
Semtler
 1: Batıkent, Yenimahalle, Demetevler, Ostim (59 kişi)
 2: Eryaman, Etimesgut, Sincan (26 kişi)
 3: Kızılay, Emek, Kavaklıdere, Bahçelievler(15 kişi)

Şekil 7. Müşterilerin semtlere göre dağılımı

3. Alışveriş tarihi arasında geçen süreler hesaplanmak suretiyle ortalama alışveriş periyotları elde edilmiş; bu periyotlar SQL komutlarıyla aşağıdaki şekilde kategorize edilmiştir:

- *update veritabanı PERİYOT= '0-6' where PERİYOT<=6 and PERİYOT >0*
- *update veritabanı PERİYOT= '7-12' where PERİYOT<=12 and PERİYOT >6*
- *update veritabanı PERİYOT= '12+' where PERİYOT>12*

Şekil 8'de veri dönüşümünden önce ve sonraki ortalama alışveriş periyodu grafikleri gösterilmektedir.



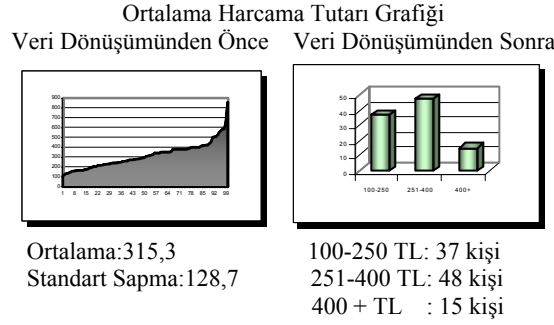
Şekil 8. Ortalama alışveriş periyodu

4. Müşterinin her alışverişte yaptığı harcama tutarlarının ortalaması alınarak, ortalama harcama tutarına ulaşılmıştır. Bu değer de SQL komutlarıyla aşağıdaki şekilde gruplandırılmıştır:

- *update veritabanı HRC. = '100-250' where HRC. <=250 and HRC. >100*
- *update veritabanı HRC. = '251-400' where HRC. <=400 and HRC. >250*
- *update veritabanı HRC. = '400+' where HRC. >400*

Churn Analysis and Customer Segmentation of a ...

Şekil 9'da veri dönüşümünden önce ve sonraki ortalama harcama tutarı grafikleri gösterilmektedir.



Şekil 9. Ortalama harcama tutarı

Çizelge 1'de veri dönüşümünden önce ve sonraki nitelik alanları; Çizelge 2'de ise kozmetik markasının müşteri verilerine ait istatistikleri gösterilmektedir.

Çizelge 1. Nitelik isimleri

NİTELİKLER	
VERİ DÖNÜŞÜMÜNDEN ÖNCE	VERİ DÖNÜŞÜMÜNDEN SONRA
<ul style="list-style-type: none">Doğum TarihiAlışveriş TarihiÇalışma DurumuBakım DurumuCilt Tipi (Karma, Kuru, Yağlı)Ürün Tipi (Bakım, Bakım/Makyaj, Makyaj, Bakım/Güneş, Tem/Nem, Bakım/Parfüm)Semt (Batıkent, Ostim, Yenimahalle, Demetevler, Eryaman, Etimesgut, Sincan, Bahçelievler, Emek, Kızılay, Kavaklıdere)Harcama Tutarı	<ul style="list-style-type: none">Yaş (24-35, 36-45, 45+)Alışveriş Periyodu (0-6 ay, 7-12 ay, 12+ ay)Çalışma Durumu (Evet, Hayır)Bakım Durumu (Evet, Hayır, Bazen)Cilt Tipi (1,2,3)Ürün Tipi (1,2,3,4,5,6)Semt (1,2,3)Ortalama Harcama Tutarı (100-250, 251-400, 400+)Ayrılan Müşteri (Churn) (Evet, Hayır)

3.2.3. Modelleme

Veri temizleme ve dönüşümünden sonraki adım modelleme adımdır. İkinci bölümde de ifade edildiği gibi bu adımda farklı modeller veri kümesi üzerinde denenerek doğruluğu en yüksek olan model seçilir.

Bu uygulamada model kurma aşamasında WEKA paket programı kullanılmıştır. WEKA, Java'da yazılmış, Windows, Linux ve Masintosh gibi farklı işletim sistemleri üzerinde çalışabilen bir programdır [15].

Çizelge 2. Veritabanı istatistikleri

NİTELİKLER	ÖZELLİĞİ/GRUBU	KİŞİ SAYISI
Yaş	24-35	35
	36-45	55
	45+	10
Cilt Tipi	1:Karma	70
	2:Kuru	26
	3:Yağlı	4
Harcama Tutarı (TL)	100-250	37
	251-400	48
	400+	15
Bakım Yaptırma	E:Evet	21
	H:Hayır	62
	B:Bazen	17
Çalışma Durumu	Evet	83
	Hayır	17
Ürün Tipi	1: Bakım	56
	2: Bakım/Makyaj	16
	3: Makyaj	7
	4: Bakım/Güneş	10
	5: Temizleme/Nemlendirici	8
	6: Bakım/Parfüm	3
Ayrılan Müşteri	E:Evet	25
	H:Hayır	75
Alış Veriş Periyodu (Ay)	0-6	18
	7-12	35
	12+	47
Semt	1: Batıkent, Yenimahalle, Ostim, Demetevler	59
	2: Eryaman, Etimesgut, Sincan	26
	3: Kavaklıdere, Kızılay, Bahçelievler, Emek	15

Modellemenin ilk aşamasında; kozmetik markası ürünlerini almaktan vazgeçerek rakipleri tercih eden müşterileri tespit etmek amacıyla ayrılan müşteri analizi, ikinci aşamasında; farklı müşteri grupları için taşıdıkları karakteristiklere uygun pazarlama politikaları geliştirmek amacıyla müşteri bölümlenmesi gerçekleştirilmiştir.

Ayrılan müşteri analizi

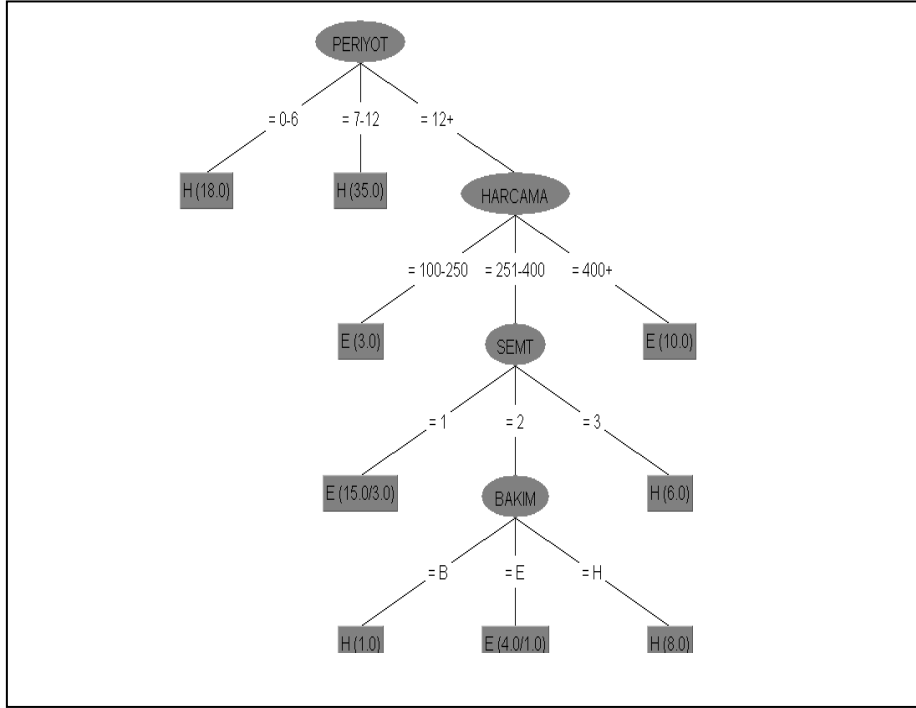
Bu analizin amacı ayrılma eğilimi gösteren müşteri karakteristiklerini tanımlamak ve bu müşterileri elde tutmaya yönelik programlar ve pazarlama politikaları geliştirmektir. Ayrılma eğilimi taşıyan müşteri kesitini belirlemek için sınıflandırma teknikleri kullanılmıştır. “churn” niteliği, hedef değişken (target variable) olarak belirlenmiştir. Farklı sınıflandırma algoritmaları kullanılarak, kurulan modellerin güvenilirlikleri karşılaştırılmıştır. Modelin doğruluğunun test edilmesinde ise 100 kişilik veri kümesi üzerinde 10 katlı çapraz geçerlilik (10-fold cross validation) testi uygulanmıştır. 10 katlı çapraz geçerlilik testinde veriler 10 gruba ayrılır ilk aşamada birinci grup test, diğer gruplar öğrenim için kullanılır. Bu süreç her defasında bir grubun test diğer grupların öğrenim amaçlı kullanılması ile sürdürülür. Sonuçta elde edilen hata oranının ortalaması, kurulan modelin tahmini hata oranı olacaktır.

Kullanılan algoritmalar ve bu algoritmaların doğruluklarının karşılaştırılması Çizelge 3’de gösterilmiştir.

Çizelge 3. Sınıflama algoritmaları ve doğrulukları

ALGORİTMA	NAİVE BAYES	LOJİSTİK REGRES.	ID.3	J.48	JRIP	PART	SİNİR AĞL.
Doğru olarak sınıflandırılan Örnek Yüzdesi	78	88	91	93	91	92	91
Yanlış olarak sınıflandırılan Örnek Yüzdesi	22	12	7	7	9	8	9
Ortalama Mutlak Hata	0.21	0.14	0.09	0.09	0.11	0.11	0.11
Ortalama Hata Karekök	0.37	0.34	0.27	0.24	0.25	0.26	0.28
Görelî Mutlak Hata	%52.7	%35.5	%23.7	%23.08	%27.8	%27.8	%28.6
Kappa İstatistiği	0.46	0.72	0.82	0.83	0.79	0.80	0.78
Sınıf E (Evet) / Keskinlik	0.62	0.73	0.83	0.89	0.77	0.89	0.80
Sınıf H (Hayır) / Keskinlik	0.83	0.95	0.97	0.974	0.98	0.93	0.95
Sınıf E (Evet) / Anma	0.64	0.89	0.92	0.89	0.96	0.89	0.89
Sınıf H (Hayır) / Anma	0.83	0.87	0.93	0.94	0.89	0.93	0.92
Sınıf E (Evet) / F-Ölçütü	0.62	0.81	0.88	0.88	0.86	0.86	0.85
Sınıf H (Hayır) / F-Ölçütü	0.84	0.91	0.95	0.95	0.93	0.94	0.94

Çizelge 3'deki değerler WEKA paket programı yardımıyla elde edilmiştir. WEKA paket programında aynı veri kümemiz için sırasıyla Naive Bayes, Lojistik Regresyon, ID3, J.48, JRIP, PART, Sinir Ağları algoritmaları seçilerek program çalıştırılmış ve elde edilen sonuçlarla Çizelge 3 hazırlanmıştır. Çizelge 3'den de görüldüğü gibi doğruluğu en yüksek olan sınıflandırma algoritması J.48 olduğu için uygulamanın ayrılan müşteri analizi bölümünde bu algoritma esas alınacaktır. J.48 algoritması; C4.5 karar ağacı algoritmasının WEKA'ya uyarlanmış versiyonudur. Her satır, ağaçtaki bir düğümü, alt satırlar, ilk satırın çocuk düğümlerini; düğümlerde parantezin içindeki ilk sayı veri kümesindeki kaç vakanın bu düğüm için doğru olarak sınıflandırıldığını; eğer varsa; parantezin içindeki ikinci sayı, düğüm tarafından yanlış olarak sınıflandırılan vakaların sayısını gösterir.



Şekil 10. J.48 Karar ağacı

Şekil 10'daki kademeler en üstten aşağıya doğru, ay kademesi, harcama kademesi, yerleşim (semt) kademesi ve müşteri kademesidir.

J.48 karar ağacı algoritmasının sonuçlarını aşağıdaki şekilde değerlendirmek mümkündür:

1. İlk dallanmada geliş periyodu 12 + ay olan müşteri grubunun ayrılma eğilimi taşıdığı görülmektedir.
2. İkinci dallanmada, harcama tutarı 100-250 TL aralığındaki müşteriler ile harcama tutarı 400 TL'nin üzerindeki müşteriler ayrılma eğilimi taşıyan müşterilerdir. Bu müşterilerin ayrılmasını ve rakip mağazaları tercih etmelerini önlemek amacıyla geliştirilecek kampanya ve stratejiler aşağıda sıralanmıştır:
 - Müşteriye doğum günü, yılbaşı gibi özel günlerde yaşına, cilt tipine ve genel alışveriş alışkanlıklarına uygun kozmetik ürünlerinden oluşan hediye paketlerinin hazırlanması ve sunulması.
 - Bu müşterilerin alışveriş tarihleri ve satın aldıkları ürünün özellikleri (türü, kaç mililitre olduğu vb.) bilindiği için ürünün bitme zamanına yakın bir tarihte müşteriye telefon ve e-posta ile ulaşılarak mağazaya davet edilmesi.
 - Müşterilerin her alışverişte yaptıkları harcama tutarına paralel olarak puan toplamaları ve bu puanları sonraki alışverişlerinde indirim ya da ücretsiz ürün almak üzere kullanmaları.
3. Üçüncü dallanmada, ortalama harcama tutarı 250 ile 400 TL arasında değişen müşterilerin rakip mağazayı tercih etmelerinde ikamet ettikleri semtin etkili olduğu görülmektedir. Batıkent

Churn Analysis and Customer Segmentation of a ...

hattı üzerinde ("1" ile kodlanan semtlerde) ikamet eden müşterileri elde tutmak daha zordur. Bu müşterileri elde tutmak için geliştirilecek stratejiler şunlar olabilir:

- Reklam ve tanıtımlarda özellikle Batıkent hattı üzerinde olan billboardları ve metrodaki ilan panolarını kullanmak.
- Batıkent ve civarında ikamet eden müşterileri diğer semtlerde ikamet edenlerden ayırarak sosyodemografik özellikleri saptayarak bu grubun karakteristiklerine uygun indirim ve promosyonlar düzenlemek.
- Batıkent ve çevresi lokasyonlarda müşteriyi mağazaya çekmek yerine ürünü müşteriye götürmek.

4. Son dallanma, "2" ile kodlanan semtlerde ikamet eden müşterilerin rakibi tercih etmesindeki etkili faktörün ücretsiz cilt bakımı yaptırıp yaptırmadıkları olduğu görülmektedir. Burada ortaya çıkan sonuç son derece ilginçtir. Cilt bakımı yaptırmayan müşterilerin ayrılma eğilimi daha düşüktür. Bu durumda, mağazaya cilt bakım programlarını yeniden gözden geçirmeleri ve ortaya çıkan bu sonucun nedenlerini araştırmaları önerilir.

Müşteri bölümlenmesi

Müşteri bölümlenmesi için WEKA'daki kümeleme algoritmalarından Simple K Mean kullanılmıştır. Kümeleme algoritmaları, sınıflama algoritmalarının tersine modeli denetimsiz olarak öğrenirler. Sınıflama algoritmalarında veriler önceden belirlenen sınıflara atandıkları (denetimli öğrenme) için kurulan modelin doğruluğu çeşitli metriklerle ölçülürken; kümeleme algoritmalarında böyle bir değerlendirme söz konusu değildir.

Bu uygulamanın müşteri bölümlenmesi bölümünde kullanılan Simple K Mean (en yakın komşu) algoritmasında k, tanımlanan küme sayısını gösterir. k ortalamalar algoritması, nümerik değerlerle; k modlar algoritması ise kategorik değerlerle çalışan algoritmalar. WEKA'da Simple K Mean algoritması her iki algoritmayı da kullanarak hem nümerik hem de kategorik değerlerle çalışan bir algoritmadır [16]. Bu algoritma başlangıç koşullarına göre farklı demetler üreten bir yöntemdir. Bu yüzden farklı demet sayılarında farklı sonuçlar ürettiği için bunların hata karelerinin toplamı (Sum of Squared Error-SSE) en az olanı seçmek gerekir. SSE değerleri Çizelge 4'dedir. Bu algoritma ile elde edilen 6 müşteri grubu Çizelge 5'de gösterilmektedir.

Çizelge 4. Hata Karelerinin Toplamı

Simple K Mean	k=2	k=3	k=4	k=5	k=6
SSE	261.0	243.0	220.0	215.0	176.0

Çizelge 5. Müşteri grupları

NİTELİKLER	Grup1	Grup2	Grup3	Grup4	Grup5	Grup6
Yaş Aralığı	24-35	45+	24-35	45+	36-45	36-45
Semt	1	2	1	3	2	1
Çalışma Durumu	Hayır	Evet	Evet	Hayır	Evet	Evet
Ort. Alışveriş Periyodu	0-6	12+	7-12	12+	0-6	12+
Cilt Tipi	Karma	Kuru	Kuru	Yağlı	Yağlı	Karma
Bakım Durumu	Hayır	Hayır	Hayır	Evet	Evet	Hayır
Ürün Tipi	1	2	1	2	1	2
Ort. Harcama Tutarı	100-250	400+	251-400	100-250	100-250	400+
Toplam Müşteriler İçindeki Oranı	%21	%12	%17	%6	%14	%30
Toplam Gelir İçindeki Payı	%15	%17	%18	%10	%14	%26

Simple K Mean algoritmasının sonuçlarını aşağıdaki şekilde değerlendirmek mümkündür:

1. Mağaza, gelirlerinin %15'ini birinci grup müşterilerinden, %18'sini üçüncü grup müşterilerinden elde etmektedir. Bu müşterilerin mağazaya geliş periyotları kısa ancak ortalama harcama tutarları çok yüksek değildir. Bu müşterilere daha yüksek tutarlarda alışveriş yaptırmak için uygulanacak stratejiler şunlar olabilir:

- Bu grup müşterilerin yaş aralığı 24-35'dir. Belirli bir tutarın üzerinde harcama yapanlara, konser ve sinema bileti vermenin, bu genç grubun harcama tutarını artırmak için etkili bir uygulama olacağı düşünülmektedir.
- 1. ve 3. grup müşterilerin harcama tutarını artırmaya yönelik diğer bir öneri çapraz satıştır. Çapraz satışta amaç, birbiriyle ilişkili ürünleri müşteriye bir indirim paketi içinde sunarak daha yüksek tutarlarda alışveriş yapmasını sağlamaktır. Bakım ürünlerini tercih eden 1. grup müşterilerine örneğin; bakım ve cilt temizleme ürünlerinin belirli bir indirim paketi içinde sunulması, müşterilerin daha fazla harcama yapmalarını sağlayacaktır.
- 2. Mağaza, gelirlerinin %17'ünü ikinci grup, %10'unu dördüncü grup müşterilerinden sağlamaktadır. Bu gruptaki müşterilerin yaş aralığı 45+'dır. Bu gruptaki müşteriler için bakım ve yaşlandırmayı geciktirici ürünlerde kampanya yapmak faydalı olacaktır. Bu gruptaki müşterilere anne kız paketleri hazırlamak ve indirimli olarak kombine satmakta satışları arttıracaktır.
- 3. Mağaza gelirlerinin büyük bir bölümünü 6. grup müşterilerden sağladığı için asıl odaklanılması gereken grup; 6. gruptur. 6. gruptaki müşteriler mağazaya daha uzun periyotlarla gelmekte, fakat geldiği zaman da daha büyük tutarlarda alışveriş yapmaktadır. O halde, veri madenciliği çıktılarını kampanya yönetiminde kullanılırken üzerinde durulması gereken nokta; 6. grup müşterilerini daha sık mağazaya çekmeyi sağlayacak stratejiler geliştirmektir.
- Bu grubun yaş aralığı 36-45, cilt tipleri karma ve satın aldıkları ürünler bakım/makyaj kategorisi olduğu için kampanyalarda müşteriye sunulan ürünler orta yaş grubuna hitap eden, yaşlanmayı önleyici ve geciktirici, karma cilt tipine uygun bakım/makyaj kombinasyonları olmalıdır. Tanıtım ve kampanyalarda orta yaşlı bayanların okuduğu bir dergiyi kullanmak etkili bir yöntem olabilir.
- Bu gruptaki müşterilerin çalıştığı düşünüldüğünde ve her birinin aynı zamanda Internet kullanıcısı olduğu varsayıldığında ürün, indirim ve promosyonlar hakkında bilgi vermek için sıfır maliyetli e-postayı kullanmak da mümkündür. Farklı gruplar için farklı ürün paketlerini içeren online kataloglar belirli periyotlarla e-posta aracılığı ile müşteriye sunulabilir.

4. SONUÇ VE ÖNERİLER

Yapılan çalışmada, veri madenciliği teknikleri kullanılarak bir kozmetik markasının müşterilerini gruplandıran ve mağazadan ayrılarak rakip mağazaların müşterisi olmayı tercih eden müşteri davranışlarını ortaya koyan modeller geliştirilerek; müşteri tutmaya yönelik çeşitli kampanyalar ve pazarlama stratejileri önerilmiştir.

SQL komutları kullanarak verinin temizlenmesi ve istenilen formata dönüştürülmesi işlemlerinin ardından; WEKA'daki sınıflandırma ve kümeleme teknikleri kullanılarak veri modellenmiştir. Sınıflama tekniklerinden Naive Bayes, lojistik regresyon, J.48 ve ID.3 karar ağaçları, JRIP ve PART kuralları ile sinir ağları kullanılmıştır. Bu algoritmaların doğruluk dereceleri karşılaştırılarak sınıflandırma için en uygun algoritmanın J.48 karar ağacı olduğu saptanmıştır. J.48 karar ağacı ile müşterinin ayrılarak rakip şirkete tercih etme (churn) davranışı modellenmiştir.

Yaptığımız çalışmayla veri madenciliği yaklaşımı Türkiye'de ilk defa bir kozmetik firmasına uygulanmış ve sonuçlar sunulmuştur. Üretim mühendislerinin büyük mühendislik veri kümelerinde gizli olan bilgiyi çıkarmasını, analiz etmesini ve yorumlamasını sağlayan veri madenciliği, çeşitli algoritmalar geliştirmelerine ışık tutmada, yazılım uygulamaları kullanarak veri madenciliği, algoritmalar ve teorilerini değerlendirmektedir. Böylece mühendislerin üretim sürecini etkin bir şekilde kontrol etmesini sağlamada yol gösterici bir çalışma olması bakımından

Churn Analysis and Customer Segmentation of a ...

önemlidir. İlerideki çalışmalarda, önerilen stratejilerin mağaza tarafından uygulamaya alınmasının ardından; satışları ve müşterilerin ayrılma oranlarını nasıl etkilediğine dayalı olarak yatırımın geri dönüş oranını (ROI) saptamak yoluyla çalışmanın etkinliği ölçülebilecektir.

KAYNAKLAR

- [1] Hudaury H., “Data Mining and Decision Making Support In The Governmental Sector”, Master Thesis, Louisville University, 2004.
- [2] Han J., Kamber M., “Data Mining Concepts and Techniques”, Morgan Kaufmann Publishers, U.S.A, 2001.
- [3] Jacobs P., “Data Mining: What General Managers Need to Know”, Harvard Management Update, Cilt 4, No 10, 8, 1999.
- [4] Akpınar H., “Veritabanlarında Bilgi Keşfi ve Veri Madenciliği”, İÜ İşletme Fakültesi Dergisi, Cilt 29, 1-22, 2000.
- [5] Rud O., “Data Mining Cookbook: Modelling Data for Marketing, Risk and Customer Relationship Management”, John Wiley & Sons, Canada, U.S.A, 2001
- [6] Berry M, Linoff G., “Data Mining Techniques For Marketing, Sales And Customer Support”, John Wiley & Sons , New Jersey, U.S.A, 1997.
- [7] Rygielski C., Wang J., Yen D., “Data Mining Techniques for Customer Relationship Management”, Technology in Society, 24:483-502, 2002.
- [8] Hwang H., Jung T., Suh E., “An LTV model And Customer Segmentation Based On Customer Value: A Case Study On The Wireless Telecommunication Industry”, Expert System with Application, 26: 181-188, 2004.
- [9] Kotler P., Armstrong G., Principles of Marketing, Prentice Hall, New Jersey, U.S.A, 2001.
- [10] Massey P., Mitzi M., “ Reengineering CRM: Leveraging Knowledge Assets In IBM”, Decision Support Systems, Cilt 32, No 2, 155-170, 2001.
- [11] Nath V., Data Warehousing And Mining In The Wireless Industry, Master Thesis, Florida Atlantic University, 2003.
- [12] Jonker J., Piersma N., Poel D.,”Joint Optimization Of Customer Segmentation And Marketing Policy To Maximize Long Term Profitability”, Working Paper of Faculty of Economics and Business Administration Ghent University, Cilt 3, 214-239, 2003.
- [13] Mucuk İ., Pazarlama İlkeleri, Türkmen Kitabevi, İstanbul, 233-255, 2004.
- [14] Allen C., Chatwin S., Creary C.A., Introduction to Relational Databases and SQL Programming, USA, 2004.
- [15] Witten I., Frank E., Data Mining: Practical Machine Learning Tools And Techniques With Java Implementations, Morgan Kaufmann Publishers, USA, 267-277, 2000.
- [16] İnternet: K Means Clustering in WEKA, <http://maya.cs.depaul.edu/~Classes/Ect584/Weka/k-means.html>, 2006.