

T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMETRİ ANABİLİM DALI
EKONOMETRİ PROGRAMI
YÜKSEK LİSANS TEZİ

İLAÇ SEKTÖRÜNDE DİJİTAL PAZARLAMA
FAALİYETLERİNE YÖNELİK CRM UYGULAMALARINA
DAYALI KÜMELEME ANALİZİ İLE MÜŞTERİ
SEGMENTASYONU

Mine HIZLI

Danışman
Prof. Dr. Mehmet Aksaraylı

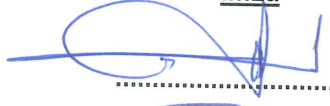
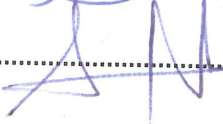
İZMİR – 2019

YÜKSEK LİSANS
TEZ ONAY SAYFASI

Üniversite : Dokuz Eylül Üniversitesi
Enstitü : Sosyal Bilimler Enstitüsü
Adı ve Soyadı : Mine HIZLI
Öğrenci No : 2011800209
Tez Başlığı : İlaç Sektöründe Dijital Pazarlama Faaliyetlerine Yönelik CRM Uygulamalarına Dayalı Kümelem Analizi İle Müşteri Segmentasyonu

Savunma Tarihi : 26/08/2019
Danışmanı : Prof.Dr. Mehmet AKSARAYLI

JÜRİ ÜYELERİ

<u>Ünvanı, Adı, Soyadı</u>	<u>Üniversitesi</u>	<u>İmza</u>
Prof.Dr. Mehmet AKSARAYLI	- Dokuz Eylül Üniversitesi	
Prof.Dr. İpek DEVECİ KOCAKOÇ	- Dokuz Eylül Üniversitesi	
Dr. Öğr. Üyesi Aynur İNCEKIRIK	- Celal Bayar Üniversitesi	

Mine HIZLI tarafından hazırlanmış ve sunulmuş olan bu tez savunmada başarılı bulunarak oy birliği (/ oy çokluğu () ile kabul edilmiştir.

Doç. Dr. Erkan GÖKSU
Müdür V.

YEMİN METNİ

Yüksek Lisans Tezi olarak sunduğum “İlaç Sektöründe Dijital Pazarlama Faaliyetlerine Yönelik CRM Uygulamalarına Dayalı Kümeleme Analizi ile Müşteri Segmentasyonu” adlı çalışmanın, tarafımdan, akademik kurallara ve etik değerlere uygun olarak yazıldığını ve yararlandığım eserlerin kaynakçada gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve bunu onurumla doğrularım.

Tarih

.... /.... /2019

Mine HIZLI

İmza

ÖZET

Yüksek Lisans Tezi

İLAÇ SEKTÖRÜNDE DİJİTAL PAZARLAMA FAALİYETLERİNE YÖNELİK CRM UYGULAMALARINA DAYALI KÜMELEME ANALİZİ İLE MÜŞTERİ SEGMENTASYONU

Mine Hızlı
Dokuz Eylül Üniversitesi
Sosyal Bilimler Enstitüsü
Ekonometri Anabilim Dalı
Ekonometri Programı

Günümüzde, sağlık sektöründe yapılan dijital pazarlama faaliyetleri stratejik önem taşıyan bir araç haline gelmiştir. Bununla birlikte son yıllarda dijitalleşmenin hız kazandığı ilaç sektöründe, firmalar dijital pazarlama pratiklerini çeşitli veri madenciliği algoritmaları ile birlikte kullanarak tüketicilere kolay ve hızlı bir şekilde ulaşmayı hedeflemektedirler. Bu çalışmada, Türkiye’de faaliyet gösteren uluslararası bir ilaç firmasının dijital pazarlama uygulamaları üzerinden veri madenciliği algoritmaları ile müşteri verilerinin analizi ve segmentasyonu amaçlanmıştır.

Çalışmanın ana odağı olan dijital pazarlama ve CRM üzerinde durularak veri madenciliği kavramları ve algoritmalar detaylı olarak incelenmiştir. Yapılan tez çalışmasında kullanılan veri seti Türkiye’nin birçok ilinde çalışan hekimlerin dijital pazarlama kanallarına göre ulaşılabilirlik düzeyini göstermektedir. Hekimlerin dijital pazarlama kanallarını kullanım sayılarına göre segmentasyonunun yapılması için veri madenciliğinin kümeleme analizi algoritmalarından k-ortalamlar algoritması ve iki aşamalı algoritma ile çalışılmıştır. Veri seti SPSS Clementine programında 3 ayrı model için çalıştırılarak elde edilen kümelerin analizi ve hangi müşteri kümesine hangi dijital pazarlama kanalı ile yaklaşım uygulanmasına karar verilmesi hedeflenmiştir. Büyük sayıdaki veri setine analizin uygulanabilirliğini kolaylaştırması ve zaman tasarruf sağlaması sebebiyle ilgili ekip tarafından bu kümeleme çalışmasının sektörel açıdan da oldukça faydalı olduğuna karar

verilmiştir. Analiz sonrası elde edilen çıktılar sayesinde firma, müşteri kümelerine uygulayacağı dijital pazarlama faaliyetlerinin seçiminde kolaylık sağlamıştır.

Gizlilik ilkeleri gereği herhangi bir demografik veri veya kişisel bir bilgi kullanılmamış olup yalnızca pazarlama kanallarına erişim sayıları kullanılmıştır. Uygulamada söz konusu veri setinin bilgisayar yardımıyla analizi için SPSS Clementine programından faydalanılmıştır.

Anahtar Kelimeler: Veri Madenciliği, Kümeleme, Dijital Pazarlama, İlaç Sektöründe Pazarlama, SPSS Clementine, K-means, K-ortalamlar, Müşteri Segmentasyonu.



ABSTRACT
Master's Thesis

**CUSTOMER SEGMENTATION WITH CLUSTER ANALYSIS BASED ON
CRM APPLICATIONS FOR DIGITAL MARKETING ACTIVITIES IN THE
PHARMACEUTICAL INDUSTRY**

Mine Hızlı
Dokuz Eylül University
Graduate School of Social Sciences
Department of Econometrics
Econometrics Program

Nowadays, digital marketing activities in the pharmaceutical industry have become a strategic tool. Besides, in the pharmaceutical industry where digitalization has accelerated in recent years, companies aim to reach out their customers easily and quickly using their digital marketing practices together with various data mining algorithms. This study's aims are, segmentation and analysis of the customers of a leading multinational Pharmaceutical company in Turkey using data mining algorithms.

With emphasis on the main focuses of the study which are, digital marketing and CRM. Concepts of data mining and algorithms are examined in detail. The data set was used in this study shows the accessibility of the healthcare professionals who were based in Turkey according to the digital marketing channels. K-means and two step algorithms were used to be able to segment the healthcare professionals in reference to times of access to the digital marketing channels. The data set was run in SPSS Clementine program for 3 different models and it was aimed to analyze the clusters obtained and to decide which digital marketing channel to approach to which customer set. This clustering study was found to be very useful in the sector as it facilitates the applicability of the analysis to a large number of data sets and saves time. With the outputs obtained from this analysis, the company has benefited from the ease of selection of the digital marketing activities to be applied to the customer clusters.

In terms of the privacy principles, no demographic data or personal information was used to protect confidentiality of the healthcare professionals and only the numbers of access to the marketing channels were used. In the application, SPSS Clementine program was used to analyze the data set.

Keywords: Data Mining, Clustering, Digital Marketing, Clustering in the Pharmaceutical Industry, SPSS Clementine, K-Means, Customer Clustering.



İLAÇ SEKTÖRÜNDE DİJİTAL PAZARLAMA KANALLARINA YÖNELİK CRM UYGULAMALARINA DAYALI KÜMELEME ANALİZİ İLE MÜŞTERİ SEGMENTASYONU

İÇİNDEKİLER

TEZ ONAY SAYFASI	ii
YEMİN METNİ	iii
ÖZET	iv
ABSTRACT	vi
İÇİNDEKİLER	viii
KISALTMALAR	xii
TABLolar LİSTESİ	xiii
ŞEKİLLER LİSTESİ	xiv
EKLER LİSTESİ	xvi
GİRİŞ	1

BİRİNCİ BÖLÜM

DİJİTAL PAZARLAMANIN TANIMI, KAPSAMI VE GELİŞİMİ

1.1.	DİJİTAL PAZARLAMA KAVRAMI	2
1.2.	DİJİTAL PAZARLAMADA KULLANILAN ARAÇLAR	3
1.2.1.	İnternet Sitesi	3
1.2.2.	Arama Motoru Pazarlaması ve Arama Motoru Optimizasyonu	3
1.2.3.	Elektronik Posta ile Pazarlama	4
1.2.4.	İçerik Pazarlaması	4
1.2.5.	Satış Ortaklığı	5
1.2.6.	Sosyal Medya Pazarlaması	5
1.2.7.	Mobil Pazarlama	6
1.3.	DİJİTAL PAZARLAMA VE GELENEKSEL PAZARLAMA KARŞILAŞTIRMASI	7
1.4.	SAĞLIK HİZMETLERİNDE PAZARLAMA KANALLARINDA DİJİTALLEŞME	7
1.4.1.	Sağlık Sektörünün Yapısı, Stratejileri, Günce Pazarlama Yöntemleri ve Yaklaşımlar	7

İKİNCİ BÖLÜM

MÜŞTERİ İLİŞKİLERİ YÖNETİMİ (CRM)

2.1.	MÜŞTERİ İLİŞKİLERİ YÖNETİMİ (CRM) KAVRAMI/TANIMI	13
2.2.	CRM KAVRAMININ ORTAYA ÇIKIŞI	15
2.3.	CRM'IN SAĞLADIĞI FAYDALAR VE KULLANIM AMACI	15
2.4.	CRM SÜRECİ	16
2.4.1.	Müşteri Seçimi	17
2.4.2.	Müşteri Edinme	17
2.4.3.	Müşteri Koruma	17
2.4.4.	Müşteri Derinleştirme	17
2.5.	CRM'DE VERİ MADENCİLİĞİ	18
2.6.	CRM UYGULAMA ALANLARI	19

ÜÇÜNCÜ BÖLÜM

VERİ MADENCİLİĞİ

3.1.	VERİ MADENCİLİĞİ KAVRAMI	21
3.2.	VERİ MADENCİLİĞİ SÜRECİ	21
3.2.1.	Problemin Tanımlanması	22
3.2.2.	Verinin Anlaşılması	22
3.2.3.	Verinin Hazırlanması	23
3.2.4.	Modelleme Aşaması	24
3.2.5.	Değerlendirme	25
3.2.6.	Uygulama	25
3.2.7.	Modelin İzlenmesi	25
3.3.	VERİ MADENCİLİĞİNDE KARŞILAŞILAN PROBLEMLER	25
3.4.	VERİ MADENCİLİĞİ UYGULAMA ALANLARI	26
3.5.	VERİ MADENCİLİĞİNDE KULLANILAN MODELLER	29
3.5.1.	Sınıflama ve Regresyon Modelleri	30
3.5.1.1.	Karar (Sınıflandırma) Ağaçları	31
3.5.1.2.	Yapay Sinir Ağları (YSA)	32

3.5.1.3.Genetik Algoritmalar	35
3.5.1.4.Bellek Tabanlı Yöntemler(K En Yakın Komşu)	36
3.5.2.Kümeleme Analizi	36
3.5.3.Birliktelik Kuralları	36
3.6. TIP VE SAĞLIK HİZMETLERİNDE VERİ MADENCİLİĞİ VE UYGULAMALARI	37

DÖRDÜNCÜ BÖLÜM

KÜMELEME ANALİZİ VE MÜŞTERİ SEGMENTASYONU

4.1. KÜMELEME ANALİZİ TEKNİKLERİ	38
4.1.1. Hiyerarşik Kümeleme Yöntemleri	39
4.1.1.1. Tek Bağlantı Yöntemi	40
4.1.1.2. Tam Bağlantı Yöntemi	41
4.1.1.3. Ortalama Bağlantı Yöntemi	42
4.1.1.4. Merkez Yöntemi	42
4.1.1.5. Ward Yöntemi	43
4.1.1.6. Medyan Yöntemi	44
4.1.2. Hiyerarşik Olmayan Kümeleme Yöntemleri	45
4.1.2.1. K-Ortalamlar (K-Means) Yöntemi	46
4.1.2.2. En Çok Olabilirlik Yöntemi	48
4.1.2.3. İki Aşamalı Kümeleme Yöntemi	49
4.2 KÜMELEME ANALİZİ SÜRECİ	49
4.3. KÜMELEME ANALİZİNİN TIP ALANINDA KULLANIMINA ÖRNEKLER	51
4.4. MÜŞTERİ SEGMENTASYONU	52

BEŞİNCİ BÖLÜM

UYGULAMA

5.1. UYGULAMANIN AMACI	54
5.2. VERİ SETİNİN TANITILMASI	54
5.3. VERİ SETİNİN HAZIRLANMASI VE UYGULAMA AŞAMASI	57
5.4. UYGULAMADA KULLANILAN MODELLER	61

5.5. ANALİZDE ELDE EDİLEN BULGULAR	70
SONUÇ	72
KAYNAKÇA	74
EKLER	



KISALTMALAR

CRM	Customer Relationship Management / Müşteri İlişkileri Yönetimi
VM	Veri Madenciliği
YSA	Yapay Sinir Ağları
CHAID	Chi-Squared Automatic Interaction Detector
SPSS	Statistical Packages for the Social Sciences
SEM	Search Engine Marketing
SEO	Search Engine Optimization
E-Mail	Elektronik Posta



TABLÖLAR LİSTESİ

Tablo 1: Sosyal medyayı en etkin kullanan ilk 10 ilaç firması	s. 11
Tablo 2: CRM'nin amaçları	s. 16
Tablo 3: Biyolojik sinir ağı ile yapay sinir ağının karşılaştırılması	s. 34
Tablo 4: Kümeleme yaklaşımları avantaj ve dezavantajları	s. 46
Tablo 5: Veri setinde kullanılan değişkenler (Dijital pazarlama kanalları)	s. 55
Tablo 6: Veri setinde kullanılan değişkenlere ait özellikler	s. 55
Tablo 7: AHP ağırlıkları	s. 65



ŞEKİLLER LİSTESİ

Şekil 1: 2010-2021 yılları arası dünya çapında sosyal medya kullanıcısı rakamları	s. 5
Şekil 2: 2018 ve 2023 yıllarında ülkelere göre sosyal medya kullanıcısı rakamları	s. 6
Şekil 3: Ortalama yaşam süresi ve 65 yaş üzeri nüfusun toplam nüfusa oranı	s. 8
Şekil 4: Küresel toplam reçeteli ilaç satışları (2018-2022), Milyar USD	s. 8
Şekil 5: İlaç firmaları sosyal medya 2017 paylaşım sayıları grafiği	s. 10
Şekil 6: Müşteri verilerinin kullanım amaçları	s. 19
Şekil 7: Veri madenciliği süreci	s. 22
Şekil 8: Veri hazırlama adımları	s. 23
Şekil 9: Veri madenciliğinin kullanıldığı alanlara göre dağılımı	s. 29
Şekil 10: Tahmin edici ve tanımlayıcı veri madenciliği modelleri	s. 30
Şekil 11: Karar ağaçları	s. 32
Şekil 12: Biyolojik sinir hücresi ve bileşenleri	s. 33
Şekil 13: Yapay sinir hücresi	s. 34
Şekil 14: Kümeleme tekniklerinin genel görünümü	s. 39
Şekil 15: Hiyerarşik kümelemede sıkça kullanılan yöntemler	s. 40
Şekil 16: Tek bağlantı tekniği	s. 40
Şekil 17: Tam bağlantı tekniği	s. 41
Şekil 18: Merkez (centroid) yöntemi	s. 42
Şekil 19: K-ortalamlar kümeleme yöntemi ile birimlerin kümelenmesi, küme merkezlerinin güncellenmesi ve buna göre yeniden atanması	s. 46
Şekil 20: K-ortalamlar algoritması akış şeması	s. 48
Şekil 21: Veri birimleri ve kümeleme	s. 50
Şekil 22: Kümeleme analizi aşamaları	s. 50
Şekil 23: Müşteri segmentasyonu	s. 53
Şekil 24: Dijital pazarlama kanallarının genel olarak yorumlanması	s. 56
Şekil 25: SPSS Clementine programına veri'nin excel yardımıyla yüklenmesi	s. 57
Şekil 26: SPSS Clementine programına veri setinin yüklenmesi	s. 57
Şekil 27: "Types" sekmesinde veri türleri	s. 58
Şekil 28: Veri setinin kontrolü	s. 58

Şekil 29: Veri setinin görünümü	s. 59
Şekil 30: Kullanılan yöntemlerin seçimi	s. 59
Şekil 31: Modelde küme sayısı ve diğer seçimlerin yapılması	s. 60
Şekil 32: Model diyagramı	s. 60
Şekil 33: Model 1 ağırlıklandırılmamış veri segmentasyon sonuçları	s. 62
Şekil 34: Model 1 segmentasyon sonuçları karşılaştırma	s. 62
Şekil 35: SPSS Clementine programı aracılığıyla model 1’de ağırlıklandırılmamış veri setinin sonuçları	s. 63
Şekil 36: Model 1 ağırlıklandırılmış veri segmentasyon sonuçları	s. 63
Şekil 37: Clementine programı aracılığıyla model 1’de ağırlıklandırılmış veri setinin sonuçları	s. 64
Şekil 38: AHP ikili karar matrisi ve AHP ağırlıkları	s. 65
Şekil 39: Model 2 AHP ile ağırlıklandırılmış veri seti’nin segmentasyon Sonuçları	s. 66
Şekil 40: Model 2 AHP ile ağırlıklandırılmış veri seti’nin iletişim kanalı bazında segmentasyon sonuçları	s. 67
Şekil 41: Model 2 ağırlıklandırılmamış veri seti’nin segmentasyon sonuçları	s. 67
Şekil 42: Model 2 ağırlıklandırılmamış veri seti’nin iletişim kanalı bazında segmentasyon sonuçları	s. 68
Şekil 43: Two-Step 4’lü kümeleme sonuçları	s. 69
Şekil 44: Two-Step 4’lü kümelemenin iletişim kanalları bazında sonuçları	s. 69
Şekil 45: K-ortalamlar 5’li kümeleme sonuçları	s. 70
Şekil 46: K-ortalamlar 5’li kümelemenin iletişim kanalları bazında sonuçları	s. 70

EKLER LİSTESİ

EK 1: AHP Ağırlıklandırması için Kullanılan Anket Formu

ek s.1



GİRİŞ

Hızla gelişen teknoloji ve internet kullanımı ile birlikte dijitalleşme kavramı da hayatımızda kilit bir noktaya gelmiştir. İnternet/teknoloji ile yakın ilişki içerisinde olan tüketicilere ulaşmanın en kolay ve hızlı yolu dijital pazarlamadan geçmektedir.

Çalışmada, veri madenciliğinde kullanılan kümeleme algoritmalarından yola çıkılarak yöntemler incelenmiş ve karşılaştırmalar yapılmıştır. Uygulama bölümünde ise sağlık profesyonellerinin dijital pazarlama kanallarına verdikleri tepki dikkate alınarak kümeleme analizi ile davranış biçimlerinin ortaya konulması amaçlanmıştır. Bu çalışma, İlaç sektöründe yaygınlaşan dijital pazarlama tekniklerinin incelenmesini hedeflemektedir. Günümüzde hızla artan dijitalleşme ile birlikte pazarlama tekniklerinde yaşanan gelişmelere değinilecektir.

Tez çalışması beş bölümden oluşmaktadır. İlk iki bölümde çalışmanın ana odağı olan dijital pazarlama ve CRM üzerinde durulmuştur. Birinci bölümde dijital pazarlama kavramı, teknikleri ve uygulama alanları açıklanmıştır. İkinci bölümde ise CRM uygulamalarının etkinliği incelenmiştir.

Üçüncü bölümde veri madenciliği kavramı ile veri madenciliği süreçleri incelenerek günümüzde veri madenciliğinin kullanım alanlarına kısaca değinilmiştir. Daha sonra veri madenciliği disiplininin meydana gelmesinde önemli bir yeri olan veritabanı teknolojisindeki gelişmelerden bahsedilerek veri madenciliği ile diğer disiplinler arasındaki ilişkiler ortaya konulmaya çalışılmıştır. Ayrıca veri madenciliğinin temel amaçları üzerinde durularak veri madenciliği ile veritabanlarında bilgi keşfi süreci arasındaki ilişki açıklanmıştır.

Dördüncü bölümde tezin esas konusu olan müşteri segmentasyonu ve kümeleme teknikleri hakkında ayrıntılı bilgi verilmeye çalışılmıştır. Çok değişkenli istatistiksel bir teknik olan klasik kümeleme analizinden hareketle kümeleme analizinin tanımı ve kümeleme analizi teknikleri açıklanmıştır. Veri madenciliğinde kümeleme analizinin öneminden bahsedilerek kümeleme algoritmalarının özellikleri incelenerek türleri hakkında genel bir bilgi verilmiştir. Yine bu bölüm içerisinde kümeleme teknikleri bölümlere ayrılmış, her bir teknik kendi içerisinde ayrıntılı bir şekilde anlatılmış ve her bir algoritmanın bir diğer algoritmaya olan avantaj ve dezavantajları tartışılmıştır.

Beşinci bölümde Türkiye’de faaliyet gösteren çok uluslu bir ilaç firmasının çalıştığı sağlık profesyonellerinin dijital pazarlama kanallarına verdikleri cevapları içeren 24 aylık veriler alınarak segmentasyon uygulanmıştır. Uygulama sonucunda elde edilen bulgular karşılaştırılarak çalışmada elde edilen bulgular sonuç bölümünde değerlendirilmiş ve yorumlanmıştır.

BİRİNCİ BÖLÜM

DİJİTAL PAZARLAMANIN TANIMI, KAPSAMI VE GELİŞİMİ

1.1. DİJİTAL PAZARLAMA KAVRAMI

Pazarlama, tüketicinin talep ve isteklerine göre günün şartlarına uygun marka yaratmak ve rekabet gücü elde etmek için çeşitli stratejiler geliştirmeyi hedeflemektedir. Dijital pazarlama bu noktada ürünlerin olası müşterilere ulaştırılması ve performans ölçümlerinin daha etkin ve hızlı bir şekilde yapılmasını kolaylaştıran, zaman açısından da sonuç odaklı çözümler ortaya koyan yöntem olarak karşımıza çıkmaktadır (SAS, 2015).

Dijital pazarlama ise elektronik platformların kullanımıyla ürün veya markanın tutundurulma çalışmaları olarak tanımlanabilmektedir. Dijital pazarlama ile geleneksel pazarlamayı birbirinden ayıran nokta, uygulanacak pazarlama stratejilerinin etkinliğinin analiz edilmesini sağlayan kanallar ve yöntemleri içermesidir (SAS, 2015).

Dijital pazarlama, geleneksel pazarlama yöntemlerinden farklı olarak, markanın veya yapılan işin tanıtımını gerçekleştirmek ve tüm pazarlama faaliyetlerine destek vermek sebebiyle interaktif platformlar kullanılarak yapılan pazarlama faaliyetlerini kapsamaktadır. Dijital pazarlama kanallarından en önemlisi de internetin oldukça az maliyetle büyük kitlelere erişilebilmesine ortam yaratması ve pazarlama stratejilerini geleneksel yöntemlere göre tamamıyla değiştirmesidir (Chaffey ve diğerleri, 2013: 102).

Dijital platformları kullanarak tüm pazarlama faaliyetlerinin dijital ortamda gerçekleşebilmesine olanak sağlayan pazarlama yöntemi olarak tanımlanabilen dijital pazarlama, geleneksel pazarlama uygulamalarına oranla günümüzde çok büyük avantaj sağlamaktadır (Dholakia ve Bagozzi, 2001: 168).

Dijital platformların kullanımıyla büyük faydalar elde eden işletmeler, dijital kanallar sayesinde çok daha büyük kitlelere düşük maliyetler ile ulaşarak sundukları ürün veya hizmetleri tanıtmakta, satış sonrası destek olanakları sunmakta ve tüketicilerin rahatlıkla düşüncelerini paylaşabileceği bir ortam yaratmaktadırlar. Teknolojideki hızlı değişimin bir sonucu olarak kullanımları oldukça fazla artan mobil cihazlar insanların yaşam şeklinde yaptığı değişikliklerin yanısıra alışveriş alışkanlıklarını da değiştirmiştir. Teknolojinin kendi kendini pazarlaması ve teknoloji sayesinde ürünlerin pazarlanması ile işletmeler günümüzde bütçelerinde dijital pazarlamaya yer ayırmaktadırlar (Ryan, 2016: 1660).

İşletmeler tarafından sonsuz sayıda üretilen dijital kaynaklar sundukları düşük üretim maliyetleri sayesinde daha az maliyetle pazarlama faaliyetleri yapabilme imkanı sunmaktadırlar. Geleneksel pazarlama faaliyetlerinde ise imkanları sağlayabilmek için işletmeler oldukça fazla elde etme maliyetine katlanmak durumunda kalmaktaydılar. Bilginin dijital ortamlara da girmesi ayrıca işletmelerin kırtasiye maliyetlerinin düşmesini sağlamıştır.

Geçmiş dönemlerde kaynakların kısıtlı olması müşteri taleplerinin karşılanmasını zorlaştırırken, günümüzde internetin oldukça gelişmiş olması talebin farklı alanlara hızlı bir şekilde kaymasına neden olmaktadır.

Bu gibi sebeplerle günümüzde talebi analiz ve takip edebilmek için firmaların fazla sayıda girişim yapmaları gerekmektedir (Aksoy, 2009: 81-88).

Tüketici davranışlarında meydana gelen değişimler, işletmelerin dijital olarak pazarlama stratejilerini gözden geçirmelerini zorunlu hale getirmektedir. İletişimin kritik öneme sahip olduğu dijital pazarlama, interaktif bir biçimde müşterilerden geri bildirimlerin alınmasına, müşteri ile ilişkilerde samimi ortamın oluşmasına, bilgi düzeyinde artışa, iç ve dış çevre ile bağlantıların gelişmesine, karar alma süreçlerinin gelişimine ve verimliliğin artmasına olanak sağlamaktadır (Tiego ve diğerleri, 2014: 703-705).

1.2. DİJİTAL PAZARLAMADA KULLANILAN ARAÇLAR

1.2.1. İnternet Sitesi

İnternet sitesi her şirket, organizasyon, kişi ya da marka için önemli bir pazarlama enstrümanıdır. Çünkü internet siteleri, şirket ya da kişi hakkında bilgileri, ürün veya hizmetler ile ilgili detaylı bilgi içermekte ve barındırdığı görseller veya videolar elektronik broşür olarak hizmet vermektedir. Kişiler bir ürün veya hizmeti satın almadan önce internet sitelerine girerek bilgi almaktadır. Bu da internet sitelerin öneminin altını çizmektedir (Strauss ve Frost, 2009: 303).

İnternet sitelerin birçok avantajı bulunmaktadır. Bunları Strauss ve Frost (2009) aşağıdaki şekilde sıralamıştır;

- İnternet siteleri düşük maliyetli bir dijital alternatiftir.
- Ürün ya da hizmet bilgileri şirket veri tabanları sürekli olarak güncellenmektedir.

Böylece internet siteleri de her zaman güncel bilgileri içermektedir.

1.2.2. Arama Motoru Pazarlaması ve Arama Motoru Optimizasyonu

Tüketici satın alma sürecinde bilindiği gibi ikinci adım “bilgi arayışı” ’dır. İnternet kullanımının yaygınlaşması ile birlikte kişi satın alacağı ürün ya da hizmet hakkında bilgi aramasını internet üzerinden yapmaktadır. Şirketler de müşterinin aramış olduğu bilgiyi en doğru yerde, en doğru şekilde, en doğru zamanda karşısına çıkarması gerekmektedir. Aynı zamanda değişen kuşak özellikleri ile birlikte tüketiciler geçmiş kuşaklara göre daha sabırsızdır ve hızlı hareket etmektedir. Bundan dolayı şirketler müşteri bilgi aradığında ön sıralarda çıkması önem arz etmektedir.

Arama Motoru Pazarlaması (SEM), tüketicilere doğrudan erişebilmek amacıyla arama motorları üzerinde yapılan pazarlama faaliyetleridir. Arama motorları pazarlama faaliyetleri genellikle 3 aktiviteden oluşmaktadır. Bunlar; “Arama Motoru Optimizasyonu”, “ücretli yerleşim” ve “ücretli katılım”’dır. Yine de Arama motoru pazarlaması denildiğinde akla ilk gelen pazarlama faaliyeti: Arama Motoru Optimizasyonudur (SEO) (Yurdakul ve Bat, 2011: 50-52).

Arama Motoru Optimizasyonu (SEO), arama sonuçlarında en iyi yere sahip olmak için yapılan optimizasyon çalışmasıdır (Fenwick ve Wertime, 2008: 99).

İnternet, veri tabanları açısından son derece önemli olan arama ve erişimin ön planındadır. Temel olarak SEO, internette çok yaygın olarak kullanılan arama motorlarının içeriğini iyileştirmenin bir yoludur. Bu sayede, bir internet sitesi kullanıcıların ulaşabileceği ve aradığı özellik ve niteliklerden daha cazip hale getirilebilecektir (Puchkov, 2016: 5).

Anahtar Kelime Pazarlaması ise; internet üzerinde belirli bir anahtar sözcük üzerinden arama yapan kişiye o anahtar kelime ile ilgili doğru sonucu sunmaktadır. Veri toplamanın en kolay yöntemlerinden bir tanesidir. Anahtar kelimeler genellikle arama motorları üzerinde kullanıldığından dolayı ayrı olarak düşünmek çok doğru değildir. Anahtar kelimeler doğru seçildiğinde şirketlerin internet sitelerinin trafikleri artacak ve arama motorlarında üst sıralara taşınacaklardır (Yurdakul ve Bat, 2011: 50-52).

1.2.3. Elektronik Posta ile Pazarlama

İnternetin toplumda çok yaygın olarak kullanımı ile birlikte en sık kullanılan pazarlama tekniklerinden biri de e-posta yoluyla müşterileri bilgilendirmektir. İşletmeler bu yöntemle müşterilerine son derece düşük bir maliyetle hızlı bir biçimde ulaşabilmektedir.

Elektronik posta ile pazarlama, elektronik posta sisteminin müşterinin izni ve kontrolü alınarak diğer internet araçları ile birlikte yürütülüp işletme tarafından geri dönüşleri ölçülebilen ve çeşitli aşamalardan meydana gelen bir pazarlama sürecidir (Haşiloğlu ve Süer, 2010: 62).

Elektronik Posta Pazarlaması, işletmelerin müşteri tatminini sağlamasında ve rekabet edebilmelerinde stratejik bir araç konumuna gelmiştir. Ancak bu aracın doğru olarak kullanılmaması durumunda istenmeyen e-posta engelleyicisine takılmak söz konusu olabilir. İzinsiz bir şekilde, tek veya toplu gönderilen reklam mesajları, aldatmaya yönelik bilgiler, virüs içerikli yazılımlar vb. kullanıcının konforunu bozan birçok paylaşımda elektronik postanın kullanılması spam olarak kabul edilmektedir. Bir işletmenin spam durumuna düşmesi, mevcut müşterilerini kaybetmelerine sebebiyet verebilir (Haşiloğlu ve Süer, 2010: 62-63).

1.2.4. İçerik Pazarlaması

Değişen, gelişen ve ortaya çıkan çok sayıda farklı pazarlama yöntemlerinin geçerli olduğu günümüz dünyasında, işletmeler satış rakamlarına odaklanmanın yanı sıra müşterilerine sundukları ürün ve hizmetleri farklılaştırarak bir değer yaratmak ve müşterileri ile uzun soluklu kaliteli ilişkilerle iletişimde kalmayı hedeflemektedir. Bu bakış açısındaki firmalar hedef kitlelerine ürün ve hizmetlerini tanıtarak erişmeyi tercih etmektedirler.

Günümüzün pazarlama stratejileri arasında içerik pazarlama olarak bilinen bu yöntemler ile işletmeler kendilerini detaylı bir biçimde müşterilerine tanıtmak gayesindedirler. İçerik pazarlamasında, üreticiler doğrudan ürün veya hizmetlerini satmaktan ziyade kendilerini tanıtır ve bu süreçte marka bilinirliğine destek sağlayacak faaliyetlerde bulunurlar (Lieb, 2012: 3).

1.2.5. Satış Ortaklığı

İnternet üzerinden reklam yaparak gelir elde etmek internet kullanıcıları ile işletme sahiplerinin kendi işlemleri için kolay olmasına rağmen, süreç içerisinde koordine bir şekilde hareket etmek ve böylece ortaklık sistemine dayalı faaliyetlerde bulunmak olmazsa olmaz bir hale gelmiştir. Bağlı kuruluş ortaklığı (affiliate marketing), her bir taraf için oldukça karlı bir reklam ve pazarlama işlemi olarak kabul edilir. Buna göre, bağlı şirket pazarlama sisteminde taraflar, satıcı ile ürün veya hizmetlerin satılmasına aracılık eden internet siteleridir. Böylece, satıcılar yalnızca satılan ürün başına daha fazla kar elde eder, daha az reklam maliyeti ile internet siteleri üzerinden çalışırlar (Chachra ve diğerleri, 2015: 41).

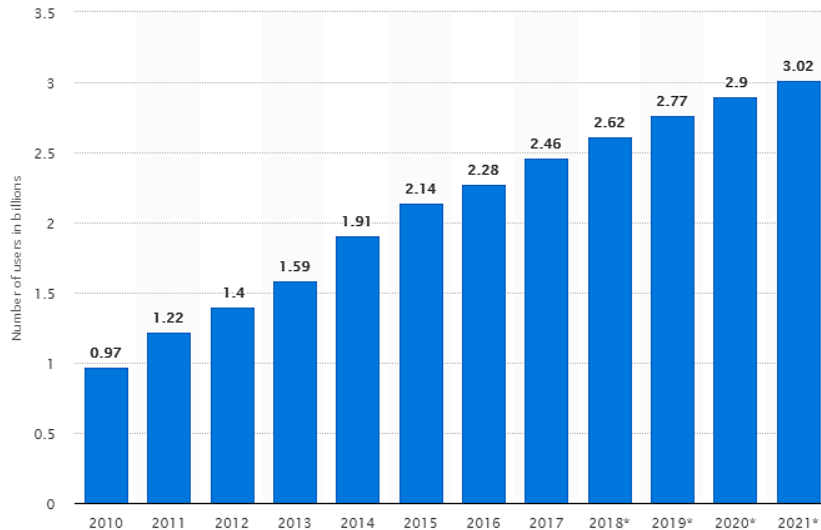
1.2.6. Sosyal Medya Pazarlaması

Sosyal medya; kişilerin birbirleriyle internet üzerinden kolaylıkla iletişime geçmelerini ve bilgi paylaşımı yapmalarını sağlayan, farklı çevrim içi teknoloji araçları için kullanılan genel bir terimdir.

İlk zamanlar kişilerin arkadaş ve aileleri ile bilgi paylaşımı için tasarlanmış olsa da akıllı pazarlamacılar sosyal medya üzerinde fırsatı görmüş ve pazarlama aktivitelerine başlamışlardır (Ontario.ca, 2015).

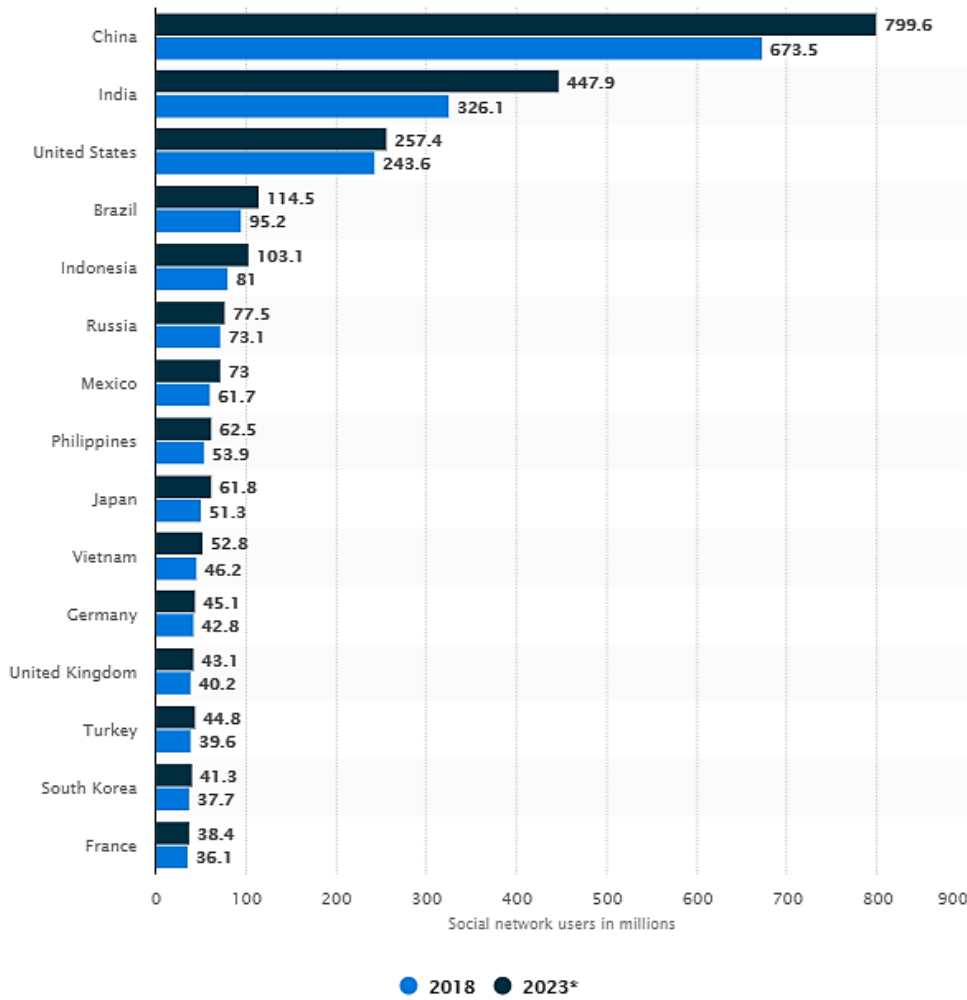
Statista.com'da yer alan Küresel sosyal medya kullanıcılarına ilişkin paylaşımda 2018 yılında toplam 2.62 milyar olan sosyal medya kullanıcısı sayısının 2021 yılında 3 milyarın üzerine çıkacağı tahmin edilmektedir. Yine aynı websitesinde yer alan istatistiklere göre 2018 yılında 39.6 milyar olan sosyal medya kullanıcısı rakamının 2023 yılında 44.8 milyara çıkması beklenmektedir.

Şekil 1: 2010-2021 yılları arası dünya çapında sosyal medya kullanıcısı rakamları



Kaynak: <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>

Şekil 2: 2018 ve 2023 yıllarında ülkelere göre sosyal medya kullanıcısı rakamları



Kaynak: <https://www.statista.com/statistics/278341/number-of-social-network-users-in-selected-countries/>

1.2.7. Mobil Pazarlama

Mobil Pazarlama, işletmelerin mobil telefonlar aracılığıyla mallarını, hizmetlerini ve fikirlerini sektörde mevcut ve olası müşterine iletme amacıyla yararlandıkları kablosuz etkileşimli (interaktif) bir pazarlama aracı olarak tanımlanabilir (Scharl ve diğerleri, 2005: 159-173).

Mobil telefonlar, önemli bir mobil pazarlama aracı olmaları sebebiyle satış, reklam, satış geliştirme, müşteri etkileşimi ve doğrudan pazarlama faaliyetlerinin etkin bir şekilde yerine getirilmesinde büyük rol oynamaktadır. Örneğin, bir mağaza, işyeri veya lokantanın önünden geçerken hedef müşterilerin mobil telefonuna gelen kampanya bilgileri, o mesajı alan müşterilerin satın alma eylemine geçmesine neden olabilir (Barutçu, 2008: 264-265).

1.3. DİJİTAL PAZARLAMA VE GELENEKSEL PAZARLAMA KARŞILAŞTIRMASI

Dijital pazarlama geleneksel pazarlama yöntemlerine dayalı olup teknolojiyi de kullanarak, çağın hızına ayak uyduran bir pazarlama dalı olarak karşımıza çıkmaktadır. Geleneksel pazarlama yöntemlerinden farklı olarak dijital ortamda var olan bir pazarlama dalıdır. Dijital pazarlamanın geleneksel pazarlamadan ayrılan özellikleri aşağıdaki gibi verilebilmektedir (Wind ve Mahajan, 2001: 39).

- Pazarlama faaliyetlerinin kişiselleştirilebilmesine olanak sağlayarak, müşteriler ile pazarlama ve müşteri hizmetlerinin uyumunu sağlar.
- İnternet küresel pazara yönelik bir bakış açısı geliştirerek müşteri ihtiyaçlarını, beklentilerini, davranış biçimlerini ve rekabeti küresel bir açıdan görme olanağı da sunar.
- Hızlı değişen pazarlama stratejilerine bağlı olarak hızlı karar alma gerekliliğinde kilit rol oynayarak, zamandan tasarruf edilmesinde etkin çözümler sunar.
- Aynı anda birden fazla kişiye erişebilme, onları dinleme ve cevap verme imkanı ile birlikte, zaman ve mekân sınırlaması olmaksızın hedef kitleye tam katılım olanağı sunar.
- Uzun vadeli planlamaya olanak sağlayan yapısı ile beraber, mevcut ve olası durumlara göre uygulanan stratejilerin kapsam ve uygulamalarında değişiklik yapabilme konusunda esneklik sağlar.

1.4. SAĞLIK HİZMETLERİNDE PAZARLAMA KANALLARINDA DİJİTALLEŞME

Sağlık sektöründe özellikle son yıllarda dijital pazarlama uygulamaları sıklıkla kullanılmaya başlanmıştır. Sağlığın geliştirilmesine yönelik tanıtıcı bilgilendirmeler yaygın olarak kullanılmaya başlanmıştır. Türkiye’de ilaç sektörü ve hastaneler tarafından sosyal medya araçlarının kullanımı sürekli olarak daha yaygın bir hal almaktadır (Tosyalı ve Sütçü, 2016: 3-22).

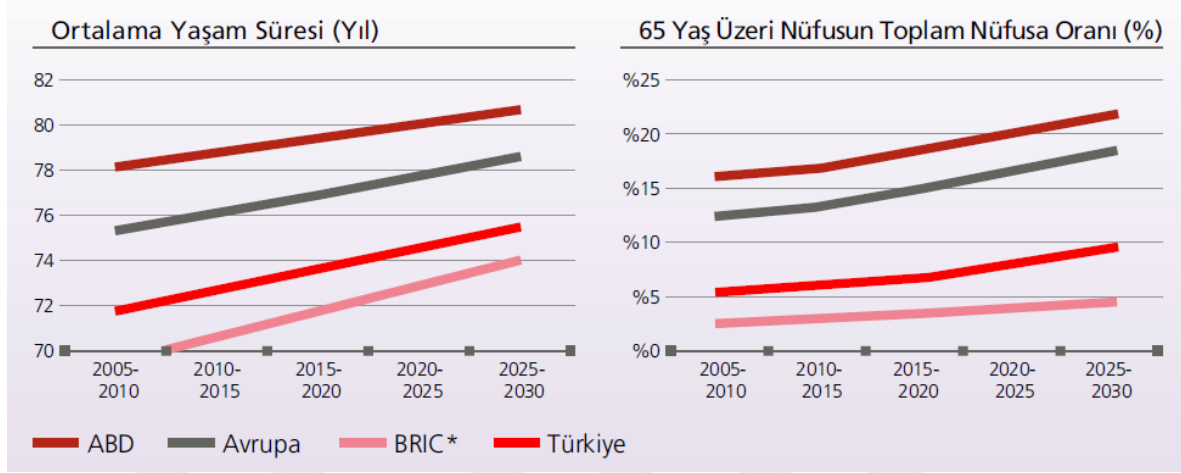
1.4.1. Sağlık Sektörünün Yapısı, Stratejileri, Güncel Pazarlama Yöntemleri ve Yaklaşımlar

Dünya nüfusunun artmasına paralel olarak toplumların refah düzeyi, demografik özellikleri ve yaşam koşullarının değişmesinden etkilenen ekosistem ile insanların sağlık sorunları da giderek çeşitlenip artmaya devam etmektedir. Bu durumda ilaç sektörü, insanın hayat standardını artırmada ve ömrünün uzamasında önemli rol oynayan, küresel ölçekte oldukça kritik bir öneme sahip bir sektördür.

Tedavisi bulunamamış hastalıklar ile kronik eğilimler, ilaç sektörünün toplumdaki önemini artırırken aynı zamanda bu eğilimlere karşılık araştırma, geliştirme ve üretim ihtiyacı sektöre yapılan yatırımları artırmaktadır.

Hızla artan dünya nüfusu ve bu toplamın içerisindeki yaşlı nüfus oranı ile birlikte sağlık hizmetlerine olan talep giderek artmakta ve yeni ihtiyaçlar ortaya çıkmaktadır (PwC Pharma, 2012: 23).

Şekil 3: Ortalama yaşam süresi ve 65 yaş üzeri nüfusun toplam nüfusa oranı



*BRIC: Brezilya, Rusya, Hindistan ve Çin

Kaynak: Dünya Sağlık örgütü, OECD, Earthtrend

2005'te 6,5 milyar olarak kaydedilen dünya nüfusunun 2020 itibarıyla 7,6 milyara çıkması, 65 yaş ve üzeri insanların sayısının ise 242 milyon artarak toplam nüfus içerisindeki oranının %9,4'e çıkması beklenmektedir. Yapılan araştırmalarda yaşlı insanların genç insanlara göre daha çok ilaç kullandığı bulgusuna varılmıştır. 75 yaş üzeri insanların %80'i en az 1 reçeteli, %36'sı ise en az 4 reçeteli ilaç kullanmaktadır (PwC Pharma, 2012: 23).

Dünya reçeteli ilaç satışları ve kategorileri incelendiğinde, yılda 5,6 oranında ortalama büyüme ile 2022'de rakamın toplam 1 trilyon dolar seviyesine ulaşacağı öngörülmektedir (KPMG, 2018: 6).

Şekil 4: Küresel toplam reçeteli ilaç satışları (2018-2022), Milyar USD



Kaynak: Evaluate, Mayıs 2017

Mevcut rakamlara göre Dünya’da sađlık sektörünün son yıllarda hızlı bir büyüme içerisinde olduđu görülmektedir. Dünya nüfusundaki demografik deđişim, teknolojideki gelişme, ortalama yaşam sürelerindeki artış ve sektörel rekabet gibi faktörler pazarın kendi iç dinamiklerine uygun stratejiler ile hızlı deđişme ayak uydurması gerekliliđini ortaya koymuştur.

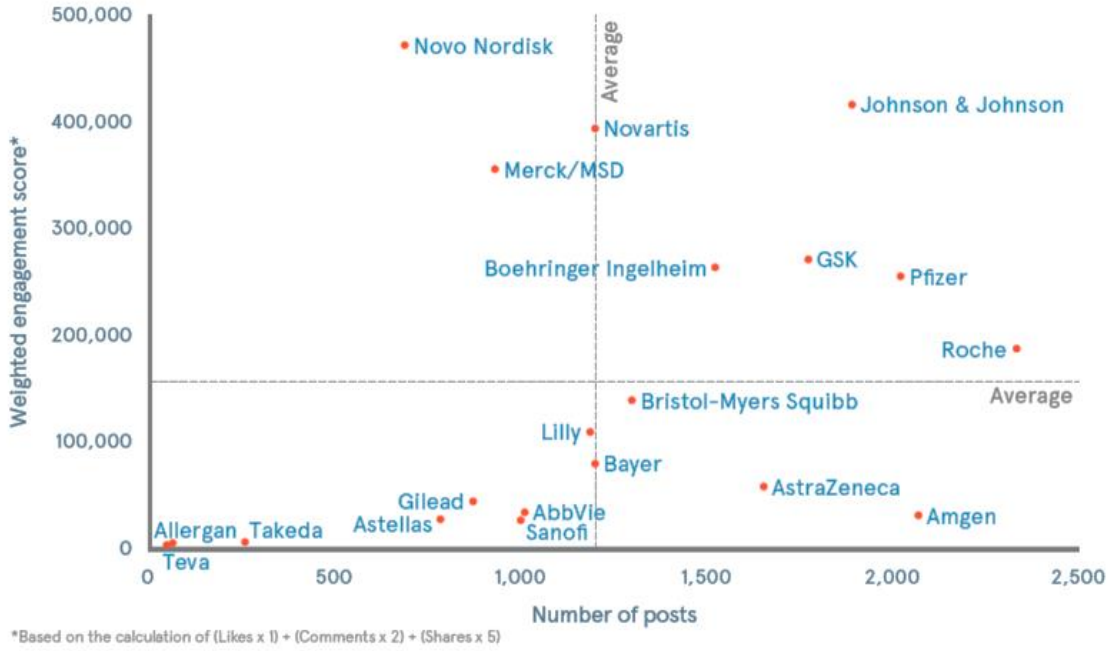
Sađlık hizmetlerinin temelini oluşturan ilaç endüstrisi teşhis, tedavi ve önleyici nitelikte ürünler geliştirmekte olup hem orijinal hem de jenerik ilaçlar piyasada yer almaktadır.

İlaç sektörünün küresel önemi, her geçen gün daha da artmaktadır. İlaç kullanımı doğum öncesinden başlayarak insanın tüm yaşamı boyunca önemli bir yer tutmaktadır. Tedavisi henüz bulunamayan hastalıklar ve kronik eğilimler, ilaç sektörünün, büyük bir sorumlulukla, kendisini, ürünlerini ve geleceğin formüllerini geliştirmesini zorunlu hale getirmiştir. Bu zorunluluğun etkisiyle dünyada ilaç sektörü en yoğun Ar-Ge yatırımının yapıldığı ve en hızlı büyüyen sektörlerin başında gelmektedir. Sađlık sistemindeki pek çok yenilik ilaç sektörünü de ilgilendirmektedir. Bugün robotik sistemlerle yapılan ve olumlu sonuçlar veren ameliyatlar, sađlık sektöründeki ilaç kullanım dinamiklerini etkilemekle birlikte, tedavi etmenin yanında koruyucu tıp alanında da önemli adımlar arasında sayılmaktadır. Sektör bu çalışmaların da merkezinde yer alır. Ancak sađlık sektöründe ilaç firmalarının sorumluluđu Ar-Ge ve üretimle sınırlı deđildir. Başta sađlık profesyonelleri olmak üzere tüm sađlık sektörü, ilaç firmalarının yönlendirmelerinden yararlanmaktadırlar. İlaç sektörü, toplam sađlık sektörünün kanaat önderi olarak daha fazla bilgi üretmeyi de hedefleri arasında görmelidir (KPMG, 2018: 3). Bu sebeplerle ilaç sektörünün temel pazarlama faaliyeti olan sađlık profesyonellerine yapılan tanıtım çalışmalarında teknolojideki gelişmelerin ve İnternetin dünyada hızla yayılması ile bir etki kaynağı haline gelen sosyal medyanın yakından takip edilmesi önemli rol oynamaktadır.

Sosyal medya, sosyal bir varlık olan insanın hem kendisini hem de medyayı sosyalleştirerek evrilmesi sonucunda ortaya çıkmıştır (Sütçü, 2012: 87). Son yıllarda en çok kullanılan sosyal medya araçlarının başında Facebook, Twitter, Youtube ve Instagram gelmektedir. 2017 yılı sosyal medya kullanım istatistiklerine göre Facebook 2.13 milyar kullanıcı (Facebook, 2017) ile lider konumunda yer almaktaydı. 2018 yılı istatistiklerine bakıldığında ise Facebook 2.32 milyar kullanıcı (Facebook, 2018) ile yine birinciliğini korumaktadır. En çok kullanılan ikinci sosyal medya aracı olan Youtube’u üçüncü olarak Instagram takip etmektedir.

Ogilvy Healthworld’ün 2017 yılında en son yayınlanan sosyal medya raporu ilaç firmalarının sosyal medya kullanımlarının sosyal mecralardaki deneyim ve güvenlerinin artışına bađlı olarak olgunlaştığını göstermektedir. Yapılan araştırma, ilk 20 küresel ilaç firmasına bakıldığında sosyal medyada içerik, hedef ve stratejilerinin daha da geliştiđini ortaya koymaktadır.

Şekil 5: İlaç firmaları sosyal medya 2017 paylaşım sayıları grafiği



Kaynak: Ogilvy Healthworld

1.4.2. İlaç Sektöründe Dijital Pazarlama Trendleri

Dünya Sağlık Örgütüne göre sağlık; “sadece hastalık ve sakatlığın olmayışı değil, bedence, ruha ve sosyal yönden tam bir iyi olma hali” şeklinde tanımlanmaktadır. Tanımda yer alan hastalık kavramı doktorlar ile doktor olmayanlar tarafından farklı algılanmaktadır. Hastalık doktorlar tarafından “Doku ve hücrelerde normal dışı yapısal ve işlevsel değişikliklerin doğurduğu her türlü durum” şeklinde tanımlanırken doktor olmayanlar için hastalık durumu, kişinin bilgi ve kültür düzeyine göre değişiklik göstermektedir (Fişek, 1982). Bu sebeple, toplumun sağlık konusunda bilinçlendirilmesi oldukça fazla önem taşımaktadır. Bu bilinçlendirmenin yanında aynı zamanda bireysel davranışlar ile alınan kararlarda olumlu yönde etki oluşturulması, sağlık hizmetlerinde kalitenin artırılması ve toplumsal sağlığın geliştirilmesi sağlık iletişimi kapsamında ele alınabilecek diğer önemli konulardandır (Tosyalı, 2016: 11).

Sağlık sektöründe sosyal medyanın reklam ve tanıtım aracı olarak kullanılması özellikle hastane ve ilaç firmaları için neredeyse bir zorunluluk haline gelmiştir. Sağlık problemi yaşayan bir birey geçmişte aile büyüklerinden veya arkadaş çevresinden ilaç ve tedavi önerileri alırken günümüzde ise karar vermeden önce internet ve sosyal medyada araştırma yapmaktadır. Bu sebeple diğer ticari kuruluşlarda olduğu gibi sağlık kuruluşları tarafından da sosyal medya araçları pazarlama amacıyla aktif bir şekilde kullanılmaktadır. Griffis ve diğerleri (2014), Amerika’da faaliyet gösteren 3.371 hastaneden; 3.351’inin (%99,41) Facebook ve Foursquare, 1.713’ünün (%50,82) Twitter hesabının bulunduğunu, bu hastanelerden 1.699’unun (%50,40) her dört sosyal medya aracını da kullanırken, 42’sinin (%1,25) sadece bir veya iki sosyal medya aracını kullandığını ve bu kuruluşlardan; Büyük ölçekli olanlar, şehir merkezlerinde faaliyet gösterenler ve kâr amacı gütmeyenler ile eğitim ve araştırma hastanelerinin sosyal medya araçlarını daha aktif kullandıklarını belirtmektedir.

Aynı çalışmada, hastaların e-posta ve telefon gibi araçların yanı sıra yeni iletişim kanallarına yöneldikleri, bu sebeple hastanelerin, klinik uzmanlık alanlarıyla ilgili farkındalık oluşturmak, karşılaşılan vakalarla ilgili toplumu bilgilendirmek, önemli sağlık ipuçlarını topluma yaymak ve elde ettikleri ödül ve başarıları takipçileriyle paylaşmak amacıyla sosyal medya stratejileri geliştirdikleri belirtilmektedir (Tosyalı, 2016: 12).

IMS Health tarafından 3 sosyal medya kanalının (Facebook, YouTube ve Twitter) erişim, ilgililik ve katılım endeksleri (beğeniler, paylaşımlar ve re-tweetler kullanarak) baz alınarak yapılan çalışmada 10 büyük ilaç firmasının sıralaması aşağıda verilmiştir.

Tablo 1: Sosyal Medyayı en etkin kullanan ilk 10 ilaç firması

Sıralama	Şirket	Puanlama
1	Johnson & Johnson	70
2	GlaxoSmithKline	25
3	Novo Nordisk	23
4	Pfizer	20
5	Novartis	18
6	Boehringer Ingelheim	18
7	Bayer	16
8	Merck & Co	13
9	AstraZeneca	10
10	UCB	9

Kaynak: IMS Health

İlaç firmaları arasında son dönemde sosyal medyada aktif olarak etkileşimde bulunmak gelişen pazarlama stratejilerine ayak uydurabilmeleri açısından kaçınılmaz hale gelmiştir. Sektöre uygulanan düzenleme ve yaptırımlar sebebiyle sosyal medyayı diğer sektörlere göre daha geç kullanmaya başlayan ilaç firmaları sosyal medyanın yanı sıra dijital pazarlama tekniklerini de aktif olarak sektöre uyarlamaya başlamışlardır. Firmalar, devamlı olarak teknolojinin gelişimine paralel olarak pazarlama stratejilerini yenilemektedirler. İlaç firmalarında dijital pazarlama, son on yılda gelişen dijitalleşme ile bugün oldukça önemli hale gelmiş durumdadır. Sağlık sektöründe faaliyet gösteren kuruluşlar, diğer sektörlere kıyasla pek çok hukuki yaptırıma takıldığından dijital pazarlama faaliyetlerinden bütünüyle faydalanamamışlardır, dijital pazarlama bir çok sağlık profesyoneline daha hızlı, kolay ve az maliyetli erişim imkanı sunmaktadır.

Sağlık profesyonelleri ve hastalar internet ve sosyal medyayı kullanarak tedavi yöntemleri ve hastalıklar ile ilgili pek çok bilgiye hızlı, az maliyetli ve kolay bir şekilde erişebilmektedirler. Ulaşılan medikal bilgilerin doğruluğu halk sağlığı açısından büyük önem taşımaktadır.

İnteraktif olarak medikal bilgi paylaşımında bulunan bazı ilaç firmalarının dijital pazarlama örneklerine bakılacak olursa;

GlaxoSmithKline Türkiye'nin desteklediği asidunyasi.com projesi ile sunulan uygulama ve internet sitesi, ebeveynlere ve ailelere aşı ile önlenabilir hastalıklar hakkında bilgi vermek, aşılardan ve aşılamanın değerini aktarmak ve hangi aşılardan hangi durumlarda, ne zaman ve nasıl yapılmalı konusunda bilgi vermek amacıyla oluşturulmuştur. İnternet sitesi ve uygulama sayesinde bireylerin aşılardan kolayca takip edebilmeleri ve aşı takvimi oluşturup, aşılama tarihlerini bildirimler ile takip edebilmeleri sağlanmaktadır.

Sanofi tarafından diyabet hastaları için geliştirilen GoMeals adlı uygulamada yenilen yiyeceklerin besin değerlerini hesaplayarak sağlıklı yemek tercihleri yapılabilmektedir.

Turuncu Hat, GlaxoSmithKline tarafından 2015 yılında hayata geçirilmiş, eczacı ve eczanelere özel bir platformdur. Turuncu Hat üzerinden GSK ürünleri ile ilgili detaylı bilgi alınabilir. Ayrıca, mesleki gelişimi destekleyen canlı yayınlar, yazılar ve videolar ile güncel içerikler sunulmaktadır.

Lundbeck, Selincro için yalnızca doktorların ve onların hastalarının kullanımına açık olan Sigara, Alkol ve Madde Bağımlılığı Tedavi Programı'nı (SAMBA) kullanmaktadır. Birçok farklı oyun ve eğitim modülü içeren bu programın amacı; alkol ve madde konusunda kişinin bilgilendirilmesi, motivasyon sağlamak, yinelenmeyi engelleyecek beceriler kazandırmak, öfke ve stresele başa çıkma becerisini öğretmektir.

Roche, geliştirdiği Medikaynak adlı online platform aracılığıyla sağlık profesyonellerine yönelik medikal bilgi, alanında uzman doktorlar ile canlı yayınlar, tedavi alanlarına göre güncel video ve makaleler sunmaktadır. İnternet sitesinin bir diğer özelliği de "Benim Vaktim" uygulamasıdır. Bu uygulama sağlık profesyonelinin siteye ayırabileceği süre için ihtiyaçlara yönelik içerik sunmaktadır.

Pfizer'in sunduğu aile hekimlerinin kullanımına yönelik Ufakbirara adlı platformda aile hekimlerinin ihtiyaçları ve ilgi alanlarına uygun olarak bir çok medikal ve pratik bilgiler yer almaktadır. Aile hekimleri internet sitesinde medikal, mesleki ve sosyal konularda makaleler, videolar, görseller ve indirilebilir içeriklere erişebilmektedirler.

İKİNCİ BÖLÜM

MÜŞTERİ İLİŞKİLERİ YÖNETİMİ

Günümüz modern işletmeleri, yatırımlarını müşteri odaklı politikalar üzerine kurmaya başlamışlardır. Rekabet şartlarının oldukça zorlaştığı bir ortamda işletmeler satış ve satış sonrası hizmetleri ile tüm müşterileri olan iletişim noktalarını güçlendirmek için çalışmaktadır. Bu sebeple işletmelerin iş süreçleri ile bilgi yönetimi teknolojilerini bir araya getiren sistemler oluşturması bir gereklilik haline gelmiştir. Rekabetçi ortamda işletmeler hizmet kaliteli ve sürdürülebilir olarak müşterilerine ulaştırma zorunluluğundadırlar. İşletmelerin müşteri odaklı yönetim stratejisi ve olan, “Müşteri İlişkileri Yönetimi” (Customer Relationship Management, CRM) oldukça önemli bir yapıdır (Taşpınar, 2005: 13).

CRM altyapısı iyi kurulan işletmeler, müşterisine sadece satış yapmak yerine müşterisini iyi tanıyıp, beklentilerini iyi belirleyen ve ihtiyaçlarına göre hizmet sunan bir yapı sahibi olmaktadır.

Yeni müşteri edinme maliyeti, mevcut müşterinin elde tutulmasının maliyetine göre oldukça fazladır. Müşteri İlişkileri Yönetimi, sadece satışların etkinliğini artırmaya yarayan bir teknik olarak kalmayıp, aynı zamanda mevcut müşterilerin elde tutulmasını da sağlayan bir tekniktir (Lipscomb, 2004).

Müşterilerin talep ve tercihleri ilgili ürün ve hizmetler ile doğru bir şekilde karşılayabilmek müşteri ilişkilerini yönetmenin en temel noktasıdır (Stiehl, 2004).

Günümüz rekabet ortamında CRM stratejilerinin, hızla gelişen ve değişen dünya dinamiklerine uygun ve güncel kalması gerekmektedir. Önceki dönemlerde pazarlama tanımları verilirken müşteri yerine ürün odaklı tanımlar yapılmakta iken CRM ile yürütülen bu ürün odaklı yaklaşım son bularak müşteri odaklılık önem kazanmıştır. CRM ile firmaların sadece satış-pazarlama stratejileri değişmemiş aynı zamanda iç yapıları da değişerek güç kazanmıştır. Bu da CRM uygulayan işletmeler için olumsuz rekabet şartlarıyla daha kolay mücadele etmek anlamına gelmektedir (www.teknoturk.org, 2005).

2.1. MÜŞTERİ İLİŞKİLERİ YÖNETİMİ (CRM) KAVRAMI

Günümüzde dinamik ve hızla değişen, gelişen bir müşteri yapısı ile karşılaşilmektedir. Bu durum müşterilerin davranışlarını takip etmeyi ve beklentilerini karşılamayı zorlaştıran bir etki sunmaktadır. Bugün dünyada pek çok firmanın karşılaştığı temel sorunlardan biri , müşteri sadakatının ciddi boyutlarda azalmış olması durumudur. Bu durumun temel sebebi, günümüzde hızla değişen teknoloji trendleri ve artan rekabetin sonucu, müşterilere sunulan daha uygun fiyatlı alternatif ürünlerdir. Müşterinin sadakatının azalması kar marjlarında da düşüşe sebebiyet vermiştir. Bu anlamda CRM, müşterilerin sadakatını arttırmayı hedefleyerek kar marjlarının da artmasını sağlamaya çalışan bir teknik olarak karşılanmaktadır (Kırım, 2004: 46-47).

CRM; etkin pazarlama, satış ve satış sonrası destek sağlayabilmek adına müşteri odaklı bir işletme yapısı ve kültürü gerektirmektedir. CRM ile işletmeler için en kıymetli müşteriler seçilip yönetilebilmektedir. CRM uygulamaları ile birlikte müşteri ilişkileri yönetiminde etkinlik sağlanmaktadır (Thompson, 2004).

Etkili CRM stratejileri sayesinde firmalar, müşterileri hakkında detaylı bilgi toplayarak, müşterilerini küçük ayrıntılarla kümeleyebilmekte (mikro-segmentasyon) ve bu kümeleri de karlılık oranlarına göre ayırıştırıp, kar getirisi yüksek müşterilere yapılabilecek yatırımın boyutunu belirleyip her müşterinin kendisine özel pazarlama stratejisi uygulayabilmektedir (Kırım, 2004:60).

Yeni CRM stratejisini destekleyen uygulamalar ile işletmeler ürün ya da hizmet odaklılıktan müşteri odaklılığa geçiş yapmıştır. İşletmeler satış ve pazarlama politikalarını oluşturabilmekte ve iş yapış biçimlerini ve süreçlerini iyileştirebilmektedirler (Yalçın, 2008: 7).

Günümüzde zaman ilerledikçe artan rekabet ortamında, hızla artış gösteren müşteri sayıları sebebiyle CRM stratejilerinin gün geçtikçe önemi artmaktadır. İnteraktif bir altyapı ile bilgi kullanımını da istenilen yönde geliştirip analiz edebilen firmalar doğru bilgilere daha kolay ulaşarak bir adım öne geçmektedirler. Günümüzün pazarlama yapısına uygun iyi bir CRM sistemine sahip firmalar müşterilerini daha kolay bir şekilde ellerinde tutabilmektedirler (Bergeron, 2002: 15-19).

CRM'in işletmelere pek çok farklı yararı bulunmaktadır. Bunları şu şekilde özetlemek mümkündür (Swift, 2001: 28):

- Doğru kurulmuş CRM stratejileri ile uygulanan pazarlama politikaları sayesinde müşteriye ulaşma, takibini yapma, pazarlama ve iletişim gibi faaliyetlerde tasarruf sağlanmaktadır. İşletmelerin daha düşük maliyetlerle müşteri edinmeleri mümkün olabilmektedir.
- Müşterilerle uzun ve kaliteli ilişkileri hedefleyen CRM ile artan sadık müşteri sayısı, yeni müşteri kazanılmasına dair ihtiyacı azaltmaktadır
- Mevcut müşterilerin genellikle satış hizmetlerine daha fazla yanıt verici olmaları sebebiyle, satış faaliyetlerine bağlı maliyetler de azalmaktadır. Bununla birlikte, teknoloji destekli modern müşteri etkileşim kanallarının kullanılması, müşterilerle ilişkileri etkin bir hale getirmekle birlikte pazarlama faaliyetlerinin daha düşük maliyetlerle gerçekleşmesini mümkün kılmaktadır.
- Piyasada yer alan ürünlerin benzer olması nedeniyle farklılık yaratabilmek ve rekabet üstünlüğü sağlayabilmenin yolu, müşterilerini iyi tanıyarak bu müşterilere özel stratejiler hazırlamaktan geçmektedir.
- Müşteri işlemlerinden sağlanan veriler bir araya getirilerek depolanabilmektedir. Ancak artan pazarlama kanallarına bağlı olarak, müşterilerle iletişime geçilen diğer kanallardan elde edilen yeni verilerin doğrudan iletişim ile elde edilen verilerle birleştirilmesi gerekir. Geleneksel satış kanalları ile yeni satış kanallarından sağlanan veriler bir araya getirilip en detaylı müşteri bilgilerine yönelik fırsatlar oluşturulabilmektedir.

CRM, firmaların pazarlama stratejilerine sağladığı faydalar sebebi ile günümüzde oldukça fazla ilgi görmektedir. Uygulamaya geçildiğinde alınan doğru olmayan bazı kararlar sebebiyle uygulanan başarısız CRM stratejilerine bağlı olarak fayda sağlayan işletme sayısında azalış olduğu görülebilmektedir. CRM stratejilerinde firmaların başarısızlık oranları %55 ile %75 aralığında değişim gösterebilmektedir. Bunun esas sebebi, CRM stratejilerinin işletmelere oldukça fazla fayda vaadinde bulunmasıdır. Bu vaatlere göre, CRM'in net bir tanımı yapılmadan, işletmenin kendisine uygun bir strateji geliştirilmeden, özgün iş süreçleri ve uygulamalar meydana getirilmeden yapılan uygulamalar sonucunda, aslında son derece önemli bir fırsat olan CRM'in başarısız örnekleri ile karşılaşmaktadır. CRM'in işletmeye faydalarının bilinirliğinin artması ile, firma sahiplerinin işletmeye has çalışmalar olmaksızın uygulamaya geçme düşüncesiyle bazı yazılım paketlerine yatırımlar yapıp sistem değişikliklerine gidilmekte ve bu aksiyonların ne amaçla yapıldığına önem verilmeden, CRM uygulanmaya başlanmaktadır. Sonucunda ise, birçok başarısız sonuç elde edilmektedir (Kotorov, 2003: 566-568).

2.2. CRM KAVRAMININ ORTAYA ÇIKIŞI

CRM kavramı, geleneksel pazarlama stratejilerinde uygulanmayan politikaların pazarlama alanında uygulanmaya başlanması ile yeni bir gelişmeyi ifade eder. CRM kısaltması Amerikan medyasında 1989 yılında birkaç defa kullanılmasına rağmen 2000 yılına gelindiğinde bu rakamın 14 binlere ulaştığı tespit edilmiştir. Hangi dönemde olursa olsun, ileri görüşe sahip yöneticilerin çoğu müşterilerini anlamanın stratejilerini geliştirmek için ne kadar değerli olduğunun farkında olmuştur. 50 sektörü kapsayan bir araştırmaya göre, 1923 yılında kendi sektöründe lider olan işletmelerin 43'ü günümüzde hala o sektörlerin liderleri olarak kabul görmektedir. Bu işletmelerin stratejileri incelendiğinde, çoğunda müşteri odaklı bir işletmecilik felsefesine sahip oldukları ve müşterilerini anlamak için çeşitli yatırımlar yaptıkları görülmektedir (Shahnam, 2009).

2.3. CRM'İN SAĞLADIĞI FAYDALAR VE KULLANIM AMACI

Müşteri İlişkileri Yönetimi uygulamaları;

- Tahmin ve planlama için gelişmiş araçlar sağlar
- Ticaret süreçlerinde işbirlikçi bir anlayışı destekler
- Ticari işlemlerin yönetilmesinde etkinlik sağlar
- Satış döngüsü sürecinin kısaltılmasıyla nakit artışına katkı sağlayabilir
- Daha iyi hizmetler sunulmasını sağlayarak müşteri memnuniyetinde artışa olanak sağlayabilir
- Müşteriye erişimde kullanılan teknolojik pazarlama kanallarının doğru bir biçimde kullanımına imkan sağlar
- Dağıtım kanallarının üçüncü tarflar ile ilişkisinde kontrolünü sağlayabilir (Yereli, 2001: 10).

Müşteri ilişkileri teorisinde firma ve müşteri arasında meydana gelen tüm alışveriş hareketleri kaydedilmektedir. Firmalar, müşterileri bilgilerinin yer aldığı bir veri tabanları yaratarak bu veri tabanı ile mevcut müşterilerine en kaliteli hizmeti sunmaktadır. Veri tabanında yer alan veriler maliyetlerin düşürülmesini sağlarken müşteri beklentilerini de yükseltmek amacıyla da kullanılır.

Günümüzde firmalar hem kendilerine sadık müşteriler hem de müşterileri ile kaliteli ilişkiler kurmak istemektedirler. Öte yandan imajlarının da olumlu ve kaliteli olmasını beklemektedirler. Yapılan araştırmalar ise piyasadaki müşteri tatmininin düşük olduğunu göstermektedir. Bunun sebebi de artan müşteri beklentileri, günümüz müşterilerinin sabırsız davranması ve maliyetlerin azaltılması için firmaların takip ettikleri politikalardır.

CRM’de firmaların iki ayrı hedefi olabilmektedir. Bu hedeflerin ilki firmanın ulaşabileceği maksimum sayıda kar oranı yüksek müşteriye ulaşabilmesi, bir diğeri ise daha az sayıda karlı müşteriye ulaşmasıdır. İkinci amaç firmaların kaliteli müşteriler ile devamlı olarak karlarını maksimize etmek istedikleri durumda ortaya çıkar. Dell Bilgisayar, Chevrolet ve Docker’s gibi firmalar ulaşabildikleri maksimum sayıda karlı müşteriye ulaşma amacını benimserken; Apple Bilgisayar, Porsche ve Armani gibi firmalar daha az sayıda karlı müşteriye ulaşma amacını benimsemektedir (Bergeron, 2002: 1-6).

CRM’ nin daha birçok farklı amacı bulunmaktadır. Yapılan bir araştırmanın sonuçları aşağıda verilen tabloda özetlenmiştir. Araştırmaya göre etkili bir şekilde uygulanmış CRM uygulaması başta müşteri bağımlılığında yaptığı artış olmak üzere; yeni müşterilere erişim, karlı müşterileri sayısında artış ve maliyetlerin azaltılması konularında da yardımcı olmaktadır.

Tablo 2: CRM’nin Amaçları

Birincil Amaç	Yüzde
Müşteri bağımlılığında artış	%42
Yeni müşterilere erişim	%21
Karlı müşterileri sayısında artış	%18
Maliyetlerin azaltılması	%9
Diğer	%10

CRM mevcut müşteriye elde tutmak için uygulanabilecek bir yöntemdir. CRM aynı zamanda firmaların uzun vadeli yatırımlarının geri dönüş oranlarını maksimize edilmesi veya firma kaynaklarının yönetimi için de kullanılabilir (Şimşek, 2006: 98).

2.4. CRM SÜRECİ

CRM müşteri ile firma arasındaki ilişkilerin sürdürülebilir bir şekilde yürütüldüğü bir süreçtir. Standart bir üretim süreci ya da bir algoritma gibi, belli başlı adımları takip eden ve önceden belirlenen bir sonuca erişmeyi hedefleyen bir teknolojidir. Sürecin takibi tutarlı ve ölçülebilir sonuçlar elde edilmesinde olasılığı artırır (Bergeron, 2002: 3).

2.4.1. Müşteri Seçimi

Müşteri seçimi evresinin ana amacı, işletmenin En karlı müşterisinin kim olduğu sorusunun cevabının bulunmasıdır. Bu amaçla yapılan çalışmalar şu şekilde özetlenebilmektedir (Alabay, 2007: 115):

- Hedef müşteri kitlesinin belirlenmesi
- Müşterilerin kümelenmesi (Segmentasyon)
- Müşteri konumlandırma
- Müşterilere özel kampanya stratejileri
- Yeni ürünlerin tanıtılması

2.4.2. Müşteri Edinme

Müşteri edinme aşamasında, belirli bir müşteriye en etkili yoldan satışın gerçekleştirilmesi için çalışmalar yapılmaktadır. İşletmeler ile olarak potansiyel müşterilerini belirlemektedirler. Sonrasında elde edilen verilerle potansiyel müşteriler farklı kümelerle ayrıştırılmaktadır. Bir sonraki adımda ise müşterilerin beklenti ve istekleri açık bir şekilde belirlenmektedir. Tüm bunların yanı sıra, müşteri edinme aşamasında satış elemanlarının etkinliğini yükseltmek amacıyla işletmeler çalışanlarına yönelik bazı özel eğitim programları uygulamaktadır. Öte yandan işletme, zamanının büyük bir kısmını müşteri süreçlerine ve müşteri ile ilişkiler için yapılan araştırmalara ve kıyaslama çalışmalarına harcamaktadır. Bu aşamada işletmeler kendilerini müşterilerine cazip kılacak bazı ihtiyaç analizleri ve müşteriler için teklif ve kampanya stratejileri oluşturma gibi çalışmalar yapmaktadırlar (Barnes, 2000: 19).

2.4.3. Müşteri Koruma

CRM uygulamalarında başarı sağlamak isteyen işletmeler en temelde sadık müşteri kazanabilmelidirler. CRM stratejileri işletmelerin yeni müşteriler kazanabilmelerinin yanında, mevcut müşterilerinin elde tutulabilmesini de hedeflemektedir. Geleneksel Pazarlama anlayışının aksine, işletmelerin mevzuat müşterileri ile uzun süreli ve artan bir satış ilişkisi kurmaları çok önemlidir. Mevcut müşteri sayısının korunması amaçlanan bu aşamada müşterinin ne kadar süreyle elde tutulabileceği konusunda çalışmalar yapılmaktadır. İşletmelerin bu aşamadaki başlıca hedefi, müşterilerini işletmelerine bağlama ve karşılıklı ilişkilerinin sürekli ve müşteri sadakatinin devamını sağlamaktır. İşletmeler bu hedeflerini gerçekleştirmek adına siparişlerin yönetilmesi, talep tahmini ve problem yönetimi gibi çabalar içerisinde olmalıdırlar (Yalçın, 2008: 13).

2.4.4. Müşteri Derinleştirme

Mevcut müşterinin, sadakatinin uzun dönem korunabilmesi, kârlılığının sürekli bir hale getirilebilmesi ve müşteri harcamalarındaki payının artırılabilmesi için gerekli olan faaliyetler müşteri derinleştirme evresinde gerçekleştirilmektedir. Harcamaların payı müşterinin sadakatinin sonucu olarak müşterinin harcamaları içerisinde işletmenin ürün ya da hizmetlerinin payının ya da satın alma oranının artırılmasıdır (Zengin, 2004: 52).

Müşterinin iş hacminde önemli rol oynaması stratejisinin üstünlükleriye şunlardır (Doyle, 2003: 173) :

- Böyle müşteriler eski müşteriler olduğundan, sadakat katsayısı, stratejik değer ve karlılık gibi kriterlere sahiptirler.
- Birlikte bir iş geçmişi bulunduğundan risk unsuru daha düşüktür, onları müşteri olarak kazanma çalışması yapılmayacağından daha az maliyetle hizmet verebilirler.
- Müşteri firma tarafından tanındığı için ihtiyaçları ve alışverişteki davranışları bilinir.
- Güvenli ilişkiler daha öncesinde kurulduğundan, müşteri performansının artırılması konusunda yapılan çalışmalar daha çabuk sonuç verir.

İşletmelerin çok yönlü düşünüp çalışmaları ve müşteri sadakati ile kârlılığını koruyarak müşteri harcamalarındaki payın artırılmasını sağlamaları gerekmektedir (Yurdakul, 2002: 5).

2.5. CRM'DE VERİ MADENCİLİĞİ

CRM günümüzde oldukça yaygın ve aktif bir biçimde faydalanılan bir sistemdir. Etkin bir CRM stratejisi müşteriyi daha iyi tanıyabilmeyi ve müşterilerin alışveriş konusundaki alışkanlıklarını bilmeyi ve iyi bir şekilde analiz etmeyi gerektirmektedir. CRM, müşterilerin istek ve beklentilerinin belirlendikten sonra, en uygun stratejinin geliştirilmesini de kapsamaktadır.

Veri madenciliği tekniği, CRM uygulamalarında oldukça önemli bir role sahiptir. Ancak veri madenciliği uygulamaları ile, büyük şirketlerin veri tabanlarında yer alan kayıtlar anlamlı bilgilere dönüştürülebilmektedir (Berry ve Linoff, 2000: 14).

Veri madenciliği, CRM stratejisi kurabilmek için hayati önem taşıyan, mevcut müşteri anlayışını oluşturabilmek için gerekmektedir. Şirketlerin müşteri odaklı olmalarına yardımcı olan bir yöntemdir (Rygielski ve diğerleri, 2002: 483-502).

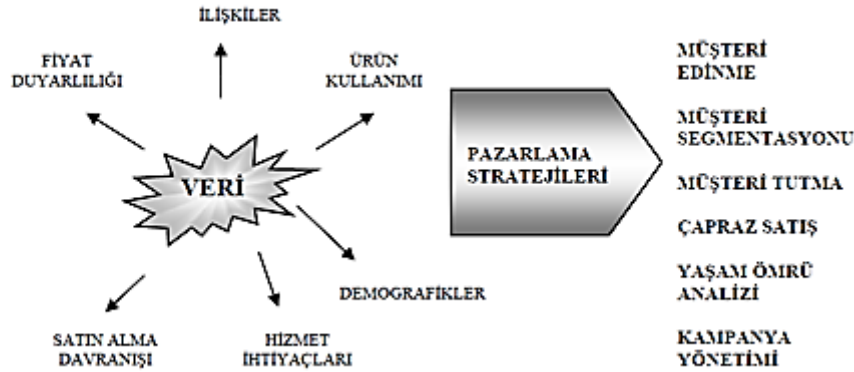
Müşteri ilişkileri yönetiminde, veri madenciliğinin temel fikri şudur:“Geçmişteki veriler gelecekte faydalı olacak bilgileri içermektedir”Bu fikir işe yaramaktadır çünkü kurumsal verilerde yakalanan müşteri davranışları rastgele değil, müşterilerin farklı ihtiyaçları, tercihleri ve eğilimlerini yansıtmaktadır (Berry ve Linoff, 2000: 14).

İşletmeler açısından müşterileri ile kurdukları uzun soluklu bağlılık, günümüz rekabet ortamında önemli bir değişimi de ortaya çıkarmıştır. Müşteri sadakatinin sağlanması ve sürekli kılması için bir çok yöntem ve tekniğin kullanıldığı görülmektedir. Bu yöntemlerden en önemlisi, müşteri merkezli ve müşteri memnuniyetinin sağlanmasını amaçlayan müşteri ilişkileri yönetimi yaklaşımıdır. Bu yaklaşım ile aynı zamanda yeni pazarlama anlayışı da ortaya çıkmaktadır (Gümüş ve Korkmaz, 2014: 2).

Günümüzde işletmeler, müşterilerin satın alma davranışından demografik özelliklerine kadar pek çok detayı veritabanlarında tutmaktadır. Veri madenciliği tekniklerini ile bu veriler içerisindeki anlamlı bağlantıların kurulması ve ortaya çıkarılması da mümkün olmaktadır (Aydogan ve diğerleri, 2008: 2).

Şirketlerin veritabanlarında tuttıkları müşteri verilerinin hangi amaçlarla kullanıldığı Şekil 6'da gösterilmektedir.

Şekil 6: Müşteri verilerinin kullanım amaçları



Kaynak: Aydoğan ve diğerleri, 2008: 45.

Veri madenciliği, yeni müşteri elde etme, müşteri kümelenmesi/bölümlendirmesi, mevcut müşterinin elde tutulması, ayrılma eğilimi gösterebilecek müşteri grubunun ortaya koyulması, müşterilerin değerlendirilmesi, kredi derecelendirme ve pazar sepet analizi gibi pek çok müşteri odaklı uygulamaya girdi sağlamaktadır (Aydoğan ve diğerleri, 2008: 45).

2.6. CRM UYGULAMA ALANLARI

Günümüzde veri madenciliğinin çok farklı kullanım alanları mevcuttur. Örneğin, bankacılık, pazarlama, e-ticaret, sigortacılık, telekomünikasyon ve sağlık araştırmaları gibi alanlarda yaygın olarak kullanılmaktadır.

Kullanım alanlarına göre veri madenciliği örnekleri aşağıdaki gibi sınıflandırılabilir.

Bankacılık:

- ✓ Kredi kartı sahtekarlıklarının tespiti
- ✓ Kredi kartı kullanımı analizi ile müşterilerin gruplandırma
- ✓ Müşteri kredi notunun değerlendirilmesi

Pazarlama:

- ✓ Müşterilerin satın alma davranışlarının takibi
- ✓ Müşterilerin demografik özelliklerindeki bağlantıların araştırılması
- ✓ Yeni müşterilerin kazanımı, mevcut müşterilerinin elde tutulması
- ✓ Müşteri ilişkilerinin yönetilmesi
- ✓ Müşterileri önem derecesine göre gruplandırma
- ✓ Satış planlama ve talep tahminlerinde bulunma

E-Ticaret:

- ✓ Sanal ortam saldırılarının çözülmesi
- ✓ E-müşteri ilişkileri uygulamalarının etkinlik yönetimi
- ✓ Kullanıcı davranışlarının analizine göre internet sitesinin güncellenmesi
- ✓ İnternet sayfalarına yapılan ziyaretlerin takip ve analizi

Sigortacılık:

- ✓ Poliçe talebinde bulunabilecek müşterilerin tahmin edilmesi
- ✓ Risk içeren müşterilerin belirlenmesi
- ✓ Sigorta dolandırıcılıklarının tespiti

Telekomünikasyon:

- ✓ İletişim ağındaki sorunlu bölgelerin belirlenip onarılması
- ✓ Müşteri davranış analizine göre yeni hizmet politikası oluşturulması
- ✓ Kullanıcı alışkanlıklarının tespiti

Sağlık araştırmaları:

- ✓ DNA içindeki gen diziliminin tespiti
- ✓ Protein analizlerinin gerçekleştirilmesi
- ✓ Hastalıklara tanı konması
- ✓ Hastalık haritasının çıkarılması
- ✓ Sağlık politikalarının oluşturulması

Bu alanların dışında da veri madenciliğinden faydalanılabilecek Perakende, Bilim, Finans, Reklam, Enerji vb bir çok alan bulunmaktadır.

ÜÇÜNCÜ BÖLÜM

VERİ MADENCİLİĞİ

Bu bölümde, büyük miktarda veri içerisinde kullanılabilir bilgiye ulaşılmasına imkan veren Veri madenciliği kavramı incelenerek veri madenciliği metotları, süreci ile uygulama alanları anlatılıp, önemli noktalar üzerinde durulacaktır.

3.1. VERİ MADENCİLİĞİ KAVRAMI

Veri madenciliği, büyük miktardaki veri içeren veri tabanlarında bulunan veriler arasında var olduğu bilinmeyen, klasik yöntemlerle tespit edilemeyen ve sıradan olmayan ilişki yapısını veya eğilimleri tespit edebilmek amacıyla istatistik, matematik, makine öğrenimi ve bilgisayar uygulamaları alanlarının birleşimi ile gerekli tekniklerin kullanılıp analiz edilmesi ve anlamlı sonuçların özetlenmesi ve görselleştirilmesi sürecidir (Irmak, 2009: 5).

Veri madenciliği, büyük miktarda verinin geniş ölçüde ulaşılabilir olması ve bu verilerin faydalı bilgiye dönüştürülmesi gereği nedeniyle son yıllarda bilgi endüstrisi ve bilgi toplumunda büyük ilgi görmektedir (Han ve Kamber, 2006: 10).

Veri madenciliği, veri yığınlarındaki geçerli ve uygulanabilir bilginin veri yığınlarından dinamik bir süreç ile elde edilmesi olarak tanımlanabilir (Baykal, 2006: 96).

Veri Madenciliği klasik istatistiksel çözümlemeden farklı olarak, daha fazla değişken ve daha fazla gözleme sahip veri kümelerini analiz etmek için kullanılmaktadır. Veri Madenciliği sahip olduğu algoritmaları kullanarak karmaşık veri kümelerini analiz edip, faydalı bilgiler ve modeller keşfetmeye yarayan bir süreçten oluşur. Bu süreçte Veri Tabanı Sistemleri, Yüksek Performanslı Programlama, Veri Görselleştirme, Görüntü ve Sinyal İşleme, İstatistik, Yapay Öğrenme ve Örüntü Tanıma gibi disiplinler arası kavramlar kullanılır (Ledolter, 2013: 2; Zaki ve diğerleri, 2014: 25).

Veri madenciliği işlemleri veri tabanlarında yer alan gizli bilginin ortaya çıkarılması için gerekli veri madenciliği işlemi, istatistik, matematik, modelleme teknikleri, veri tabanı teknolojisi ve çeşitli bilgisayar programları kullanılarak yapılır. Veri madenciliği, analiz aşamasında veriler arasındaki korelasyonu bulma konusunda fayda sağlamaktadır (Baykal, 2006: 96).

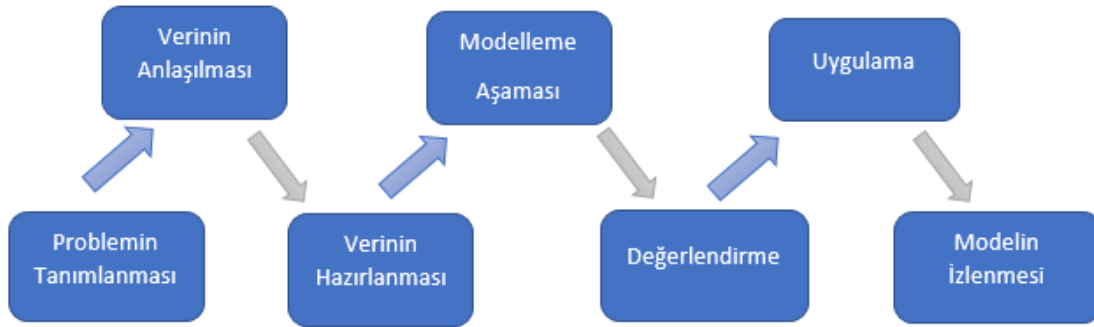
3.2. VERİ MADENCİLİĞİ SÜRECİ

İşletmeler ve devlet kurumları değişen ve gelişen teknolojik dünyada veri tabanı sistemlerine yatırım yapmaya ve daha fazla oranda veriyi bu sistemlerde depolamaya başlamışlardır. Büyük boyutlardaki veri birçok işletme veya kurumda anlamlı ve verimli bir biçimde işlenememektedir. Bu büyük boyuttaki verilerden faydalanmak için, veriler üzerinde çeşitli yöntemler uygulanarak değerli bilginin ortaya çıkarılmasına gerek duyulmuştur. Bu süreç, veritabanlarında bilgi keşfi - VTBK (Knowledge Discovery in Databases – KDD) adı verilmektedir (Fayyad ve diğerleri, 1996: 27)

Veri tabanlarında depolanan verinin incelenmesi, analizi ve yorumlanması veri tabanı bilgi keşfi süreci içerisinde yürütülür. Veri madenciliği tekniklerini uygularken ilk adım olarak problemin tanımlanması ile başlanmalıdır. Bir sonrak aşama olarak ise verilerin anlaşılıp hazırlanması gelir. Süreçte ilerlerken, sırasıyla model kurulur, değerlendirilir, uygulama hazırlanır ve izlenir. Süreç, model doğrulanarak tamamlanır (Dick ve diğerleri, 2004: 91).

Veri madenciliği, büyük ölçekli veri tabanlarında gizli olarak yer alan anlamlı bilginin ortaya çıkarılıp keşfedilmesi sürecidir (Han ve Kamber, 2006: 10).

Şekil 7: Veri madenciliği süreci



3.2.1. Problemin Tanımlanması

Başlangıç aşamasında işletmenin bakış açısına uygun olarak projenin amaç ve gereksinimlerinin anlaşılmasına odaklanılmaktadır. Ardından amaç ve gereksinimler bir veri madenciliği problemi olarak tanımlanmakta ve amaçlara erişebilmek üzere öncü bir plan hazırlanmaktadır (Tsiptsis ve Chorianopoulos, 2009).

Başarılı veri madenciliği çalışmalarının ilk şartı, yapılan uygulamanın amacının açık ve net bir biçimde tanımlanmasıdır. İşletmenin karşılaştığı problem ile işletmenin amacına odaklanmış ve açıkça ifade edilmiş olmalıdır. Çalışmalar sonucunda alınacak çıktıların başarı düzeylerinin ölçümü için seçilecek yöntemler de belirlenmelidir. Yanlış tahminlerde işletmelerin katlanması gerekecek maliyetlere ve doğru tahminlerde ise kazanacakları faydalara ilişkin tahminlere de bu aşamada yer verilmelidir (Ayık ve diğerleri, 2007: 442).

3.2.2. Verinin Anlaşılması

Veriyi anlama ile işi anlama aşaması birbiriyle iç içe geçmiş süreçler olarak değerlendirilir. Veriler incelendikçe işle ilgili farklı bakış açıları kazanılır ve işle ilgili farklı bakış açıları geliştikçe farklı verilere bakılır ve bu döngü kendi içerisinde devam ettiği sürece, çalışma süresince kullanılacak olan veriler daha da netlik kazanır. Bu süreci; (Yıldırım, 2016: 6).

- Başlangıç verilerinin bir araya getirilmesi
- Elde edilen verinin tanımlanarak ihtiyaçları hangi oranda karşılayabileceğinin değerlendirilmesi

- Veri tabanındaki eksikliklerin tespit edilerek üzerinde çalışılacak verinin analiz edilmesi
- Verinin kalitesinin belirlenmesi olarak tanımlanabilir

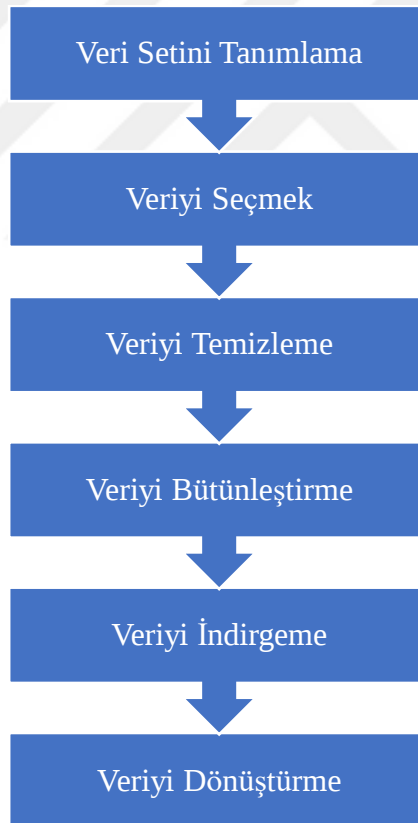
3.2.3. Verinin Hazırlanması

Veri hazırlama aşaması ham veriden başlayarak son veri seti elde edilene kadar izlenmesi gereken gereken tüm yolları kapsamaktadır. Veri hazırlama işlemleri tablo, kayıt, öznitelik seçimi ve modelleme araçları için verinin temizlenmesi ve dönüştürülmesini içermektedir (Tsiptsis ve Chorianopoulos, 2009).

Veri madenciliğinin önemli aşamalarının başında olan verinin hazırlanması aşaması, çalışmayı gerçekleştiren kişinin toplam zaman ve enerjisinin %50-%85 ini harcamasına neden olmaktadır (Piramuthu, 1998: 294)

Verinin hazırlanma aşaması, doğru sonuçlar elde edebilmek için oldukça dikkatli bir biçimde gerçekleştirilmesi gereken bir aşamadır. Şekil 8'deki sırayla gerekli yolların izlenmesi ile elde edilen veri analiz için hazır hale getirilmiş olur (Larose, 2005: 7).

Şekil 8: Veri hazırlama adımları



Veriyi hazırlama aşamasında elde edilen veriler işlenerek veri madenciliği sürecinde kullanılabilir hale getirilir. Bu aşama, veri madenciliği sürecinde en çok zaman alan ve işgücü

gerektiren adım olarak görülür. Araştırmacı bu aşamada; verinin toplanması, temizlenmesi, bütünleştirilmesi, dönüştürülmesi ve indirgenmesi gibi işlemleri yürütmektedir (Can, 2017: 16).

Veri Dönüştürme sürecinde, veri setlerine ayırıklaştırma ve normalizasyon işlemleri uygulanabilmektedir. Normalizasyon ayırık verilerin belli bir aralığa indirilmesi işlemidir. Verilerin iyi bir şekilde öğrenme algoritmalarına sokulması için sürekli ya da ayırık veri durumlarının göz önünde bulundurulması da gerekmektedir. Normalizasyon; veri setleri üzerinde sürekliliği değişmiş, veri kalitesinin ve kod yapısının bozulmuş olduğu durumlarda başvurulan bir işlemdir. Bu veri setlerine örnek olarak öğrenci notları, bilgi sistemlerinde tutulan kişilik verileri, maaş, tutar gibi mali verilerin yanı sıra insan kaynakları ve mali verilerin sistemde tutulmasını örnek gösterilebilir. Bu sebeple verinin normalleştirilmesi tekniklerinden sık kullanılanlar aşağıdaki şekilde sıralanabilir (Roiger ve Geatz, 2003: 156).

1. Min-Max
2. Z-Skor
3. Ondalık Ölçekleme

Min-max normalizasyonu sayesinde orijinal veriler yeni veri aralığına doğrusal dönüşüm ile dönüştürülürler. Bu veri aralığı 0-1 aralığıdır.

$$X^* = (X - X_{\min}) / (X_{\max} - X_{\min})$$

Z Skor normalizasyonunda ise değişkenin herhangi bir y değeri, değişkenin ortalaması ve standart sapmasına bağlı olarak bilinen Z dönüşümü ile normalleştirilir.

$$z = \frac{(x - \bar{x})}{s}$$

Ondalık ölçekleme normalizasyonunda, ele alınan değişkenin değerlerinin ondalık kısmı hareket ettirilerek normalleştirme gerçekleştirilir. Hareket edecek ondalık nokta sayısı, değişkenin maksimum mutlak değerine bağlıdır. Ondalık ölçeklemenin formülü aşağıdaki şu örnekle: 900 maksimum değer ise, n=3 olacağından 900 sayısı 0,9 olarak normalleştirilir.

3.2.4. Modelleme Aşaması

Modelleme aşamasında, verilerden bilgi çekmek için ileri istatistiksel çözümleme yöntemleri kullanıldığından uygun modelleme tekniğinin seçimi, test tasarımının üretimi, model geliştirme ve tahmin işlemlerini içermektedir. Uygun modellerin seçilip uygulanmasıyla beraber parametreler en uygun değişkenlere dönüştürülebilmektedir. Veri madenciliği, her problem tipi için farklı yöntemler içermektedir. Bazı yöntemler, veri tipi için uygun değildir veya özel tanımlamalar gerektirmektedir. Bu nedenle gerekli olduğunda 3. aşama olan veri hazırlama aşamasına geri dönebilir (Larose, 2005: 7).

3.2.5. Değerlendirme

Değerlendirme aşamasında, uygun model veya modeller oluşturulduktan sonra, veri madenciliği sonuçlarının araştırma probleminin amaçlarını gerçekleştirip gerçekleştirmediği değerlendirilir. Bu aşama sonuçların değerlendirilmesi, veri madenciliği sürecinin gözden geçirilmesi ve sonraki adımların ne olacağı hususlarını içermektedir. Bu aşamanın sonunda veri madenciliği sonuçlarının kullanımı üzerindeki karara varılmaktadır (Larose, 2005: 7).

3.2.6. Uygulama

Veri hazırlanmasının son aşaması olan uygulama aşamasında veri madenciliği süreciyle üretilen bilgiler, pratik işletme problemlerinin çözümünde kullanılabilir. Bu aşamada elde edilen bilgilerin uygulanabilmesine yönelik hazırlanan planlar, gözden geçirme ve bakım faaliyetlerini içerir. Ayrıca bu aşamada son araştırma raporunun yazılması ve projenin gözden geçirilmesi işlemleri yer almaktadır (Albayrak ve Yılmaz, 2009).

Veri madenciliğinin bu aşaması, geçerliliği kabul edilen modellerin uygulandığı aşamadır. Model doğrudan kullanılabileceği gibi başka bir modelin alt parçası olarak da kullanılabilir. Veri madenciliği süreciyle üretilmiş bilgiler, pratik işletme problemlerinin çözümünde kullanılmaktadır. Veri madenciliğinde temin edilen bilgilerin uygulanabilmesi için; planlanması, hazırlanması, gözden geçirilmesi, bakım faaliyetleri, rapor edilmesi, projenin gözden geçirilmesi gibi bütün işlemlerden geçmesi gereklidir (Sarmanova, 2013: 22).

3.2.7. Modelin İzlenmesi

Zaman içerisinde bütün sistemlerin özelliklerinde veya üretilen verilerde meydana gelebilecek değişimlerden dolayı elde edilen model sürekli olarak izlenmeli ve gerekli görüldüğünde düzeltilmelidir (Savaş ve diğerleri, 2012: 8).

Veri madenciliği sürecinde, belirlenen probleme uygun olarak kurulan, test edilen ve geçerliliği kabul edilip kullanılan modeller, ileriki zamanlarda problemin farklılaşması sebebiyle de geçerliliğini kaybedeceğinden dolayı sürekli izlenmeli ve tekrardan düzeltilmelidir. Dolayısıyla, tahmin edilen ve gözlenen değişkenler arasında farklılıklar incelenip model sonuçları izlenebilir. Bu izlenimlere bakılarak modelin düzeltilip düzeltilmemesi kararına da varılır (Özbay, 2015: 52).

3.3. VERİ MADENCİLİĞİNDE KARŞILAŞILAN PROBLEMLER

Veri madenciliğinde girdi olarak kullanılan ham veri veritabanlarından alınır. Veritabanlarının eksiksiz, dinamik ve net veri içermediği durumlarda sorunlar ile karşılaşılmasına sebep olur (Aydoğan, 2003: 12-16).

Veri Tabanının Boyutu: Gün geçtikçe veri tabanlarında bulunan veriler hızla artış göstermektedir. Veri madenciliği yöntemleri büyük veri tabanında kullanılırken analiz aşamasında aşırı dikkat gerektirmektedir. Veri tabanının büyüklüğü veri madenciliğinde karşılaşılan en önemli problemlerden biridir.

Artık Veri: Veri madenciliğinde kullanılacak verilerin arasında çözülecek probleme uygun olmayan ya da problemin çözümüne katkısı olmayan veriler artık veri olarak değerlendirilir. Bu gereksiz verilerle problemin çözümünün birçok aşamasında karşılaşmak mümkündür.

Dinamik Veri Tabanı: Çevrimiçi veri içeriği sürekli değiştiğinden bu veri tabanları dinamiktir. Veri tabanlarının sürekli değiştiği göz önünde bulundurularak, gerekli kısımlar güncellenebilir hale getirilmelidir.

Gürültü: Gürültü, verilerin toplanma ya da verilerin giriş aşamasında oluşan sistem dışı hatalardır. Hatalı veriler, gerçek dünya veri tabanlarında ciddi problemler meydana getirmektedir. Günümüzde kullanılmakta olan ileri teknoloji veri tabanları, sisteme veri girişi sırasında oluşabilecek hataları otomatik olarak gidermeye yönelik çalışmaktadır.

Boş Değerler: Boş değer, veri tabanında kendisi de dahil olmak üzere başka hiç bir değere eşit olmayan değerdir. Bir veri setinde bir nitelik değeri boş ise bilinmeyen ve uygulanamaz bir değere sahiptir. Bu durum ilişkisel veri tabanlarında sıklıkla görülmektedir (Uzar, 2013: 51).

3.4. VERİ MADENCİLİĞİ UYGULAMA ALANLARI

Bilgi işleme maliyetinin azalması ile birlikte, verinin toplanması ve saklanması sağlanan kolaylık, veri tabanı yönetim sistemi teknolojilerindeki ilerlemeler, kullanılabilecek analitik araçların artmasıyla birlikte veri madenciliği uygulamalarına olan ilgi artmıştır (Park ve diğerleri, 2001: 205).

Veri madenciliği günümüzde verinin toplandığı saklandığı tüm alanlarda uygulama imkanı bulabilmektedir. Veri madenciliğinin kullanılmasında sektör farkı gözetilmemekle birlikte, geniş veri ambarlarının oluşturulmasına olanak veren tüm işletmelerde çeşitli alanlarda uygulanmaktadır.

Veri madenciliği finans, pazarlama, tıp, astronomi, biyoloji, sigortacılık gibi bir çok dalda uygulanmaktadır. Son 20 yıldır ABD’de çeşitli veri madenciliği algoritmalarının gizli dinlemeden, vergi kaçakçılarının ortaya çıkarılmasına kadar pek çok çeşitli uygulamalarda kullanıldığı bilinmektedir (Akpınar, 2000: 4).

Veri madenciliğinin yaygın olarak kullanıldığı alanların bazıları, aşağıda özetlenmiştir.

Pazarlama ve Perakende:

Günümüzde değişik zevk ve tercihlere sahip olan müşterileri kendileri ile aynı özelliklere sahip büyük homojen gruplara ayırmak, pazarlama stratejileri oluşturmak açısından mümkün değildir. Tüm müşteriler kendi bireysel ve eşsiz ihtiyaçlarına yönelik olarak kendisine hizmet verilmesini beklemektedir. Veri madenciliğine dayalı pazarlama bu gereksinime karşılık vermekte ve müşteriler hakkında saklı kalmış bilgilerin ortaya çıkarıp pazarlama yönetiminde kullanılmasını mümkün kılmıştır (Tan ve diğerleri, 2005: 32).

Veri madenciliği uygulamaları yaygın olarak müşteri profillerinin ortaya çıkarılmasında kullanılmaktadır. Müşteri profili, müşterilerinin ihtiyaçlarını tatmin etmek amacıyla pazarlama yöneticisinin stratejilerini ve taktik kararlarını dayandırdığı müşterinin bir modelidir (Baragoin ve diğerleri, 2002: 52).

Pazarlama alanında, müşterilerin yaşı, cinsiyeti, gelir seviyesi, eğitimi ve yaşadığı yer gibi müşterilerin demografik özellikleri ile satın aldıkları ürünler, hizmetler arasında ilişki bulunup bulunmadığının ortaya çıkarılması hedeflenmektedir (Silahtaroglu, 2004: 6).

Müşterilerin birden fazla ürün grubu satın aldıkları durumda alınan ürünler arasında bir bağıntı ya da genelleme olup olmadığı da yine cevap arayan sorular arasında yer almaktadır. Buna ek olarak bir ürün ya da hizmetin bir kez alındıktan sonra bir daha alınıp alınmadığı, alınıyorsa alışlar arasında geçen süre ne kadar gibi soruların cevapları veri madenciliği uygulamaları ile etkin bir şekilde belirlenerek uygulanmaktadır (Kiremitçi, 2005: 31)

Satış kampanyaları gibi pazarlama faaliyetleri yönetilirken demografik özellikleri ve geçmiş satın alma davranışları benzerlikler gösteren müşteri bölümleri veri madenciliğinin kümeleme ve bölümlendirme metotları kullanılarak belirli bir nitelik bakımından homojen veya benzer özellikler gösteren müşteri grupları ortaya çıkarılmaktadır. Bu sayede bölümlerin her biri için farklı stratejiler uygulanabilmektedir (Park ve diğerleri, 2001: 205).

Sepet analizi yaklaşımı pazarlama ve perakende alanında sıkça kullanılan yöntemlerin başında gelmektedir. Bu yaklaşımda temel amaç; müşterilerin satın aldıkları ürünlerin arasındaki ilişkileri kurup bu ilişkiler yardımıyla firmanın satış grafiği ile karlılığını arttırmaktır (Küçükşille, 2009: 53).

Finans:

Veri madenciliği uygulamaları bankacılık sektöründe, kredi ve kredi kartı dolandırıcılığını tespit edebilmek, kredi riskini değerlendirmek, müşteri eğilimlerini analiz etmek, kar analizlerini gerçekleştirmek için kullanılabilir. Borsa ve finansal kuruluşlarda, hisse senedi fiyatlarının tahmini, portföy yönetiminin gerçekleştirilmesi gibi alanlarda veri madenciliği uygulamaları oldukça yaygın olarak kullanılmaktadır (Baykasoğlu, 2005: 9).

Finansal tablolardaki bilgiler arasındaki gizli ilişkilerin tespit edilmesi, kredi kartı harcamalarına göre müşteri gruplandırma işlemlerinde, talep edilen kredilerin değerlendirilmesi ve kredi derecelendirme işlemlerinde, sigortacılık sektöründe ise poliçesini yenilemek isteyen müşterinin tahmin edilmesi, uygun sigorta paketlerinin teklif edilmesi, sigorta dolandırıcılıklarının tespiti ve tahmin edilmesi, riskli müşterilerin analiz edilerek önceden tespit edilebilmesi gibi alanlarda da sıkça veri madenciliği uygulamalarından faydalanılmaktadır (Akpınar, 2000: 4-5).

İnteraktif Uygulamalar:

Günümüz teknolojik dünyasında internet üzerinde yer alan veriler hızla artarak gelişim göstermektedir. İnternetteki faydalı bilginin keşfedilebilmesi için gerçekleştirilen uygulamalar web madenciliği olarak adlandırılmaktadır. Web madenciliği uygulamaları sayesinde birçok web sunucusu veya çevrimiçi web sayfasından kullanıcıların erişim aktivitelerinin elde edilmesi ve analizi gerçekleştirilebilir hale gelmiştir (Baykasoğlu, 2005: 9).

Tıp ve Sağlık:

Gelişen teknoloji ile beraber veri madenciliği uygulamaları tıp ve sağlık alanında da yoğun bir şekilde kullanılmaya devam etmektedir. Özellikle tarama testlerinden elde edilen çıktıların yorumlanması ile bireylerin karşılaşabileceği çeşitli hastalıklarda ön teşhisin yapılması, risk gruplarının belirlenmesi, acil servis hastalarının tedavi sırasında risk ve önceliklerinin tespit edilmesi konularında oldukça geniş bir uygulama alanına sahip olduğunu söyleyebiliriz (Baykasoğlu, 2005: 8).

İş Süreçleri:

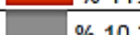
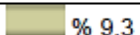
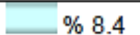
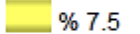
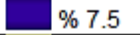
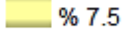
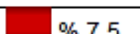
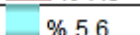
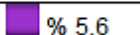
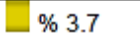
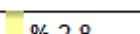
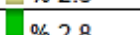
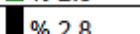
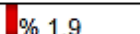
İş süreçlerinde üretilen çeşitli ve fazla sayıdaki veriler, yönetim seviyesinde karar verme aşamalarına kullanılmaktadır. Karar verme sürecinde faydalanılacak olan bu veriler ile çalışanların performans ölçümü, işe alım süreçlerinde yapılan iyileştirmeler, her müşteriye uygun reklam ve promosyonların tespit edilmesi gibi birçok faydalı bilgiye erişilebilmektedir (Baykasoğlu, 2005: 8).

Yukarıda açıklanan alanların yanı sıra veri madenciliğinin fayda sağladığı alanlar arasında aşağıdaki başlıklar da özetlenebilir (Baykasoğlu, 2005: 9):

- Taşımacılık ve ulaşım
- Bankacılık
- Eğitim
- Turizm
- Eğlence
- Telekomünikasyon
- Belediyeler

Veri madenciliği uygulama alanları burada sayılmayan daha başka birçok alanda da uygulanabilmektedir. www.kdnuggets.com sitesinin veri madenciliğinin uygulama alanlarını ve oranlarını belirlemek için, web sitesi üzerinden analistlere yönelik olarak yaptığı anket sonuçları Şekil 9'da verilmiştir (Akman, 2010: 11).

Şekil 9: Veri madenciliğinin kullanıldığı alanlara göre dağılımı

CRM/ Müşteri Analizleri (41)		% 38.3
Bankacılık (34)		% 31.8
Dolandırıcılık Tespiti (21)		% 19.6
Finans (18)		% 16.8
Doğrudan Pazarlama (15)		% 14.0
Diğer (14)		% 13.1
Yatırım/Borsa kararları (14)		% 13.1
Kredi kartı Skorum (14)		% 13.1
Telekomünikasyon(13)		% 12.1
Perekandecilik(13)		% 12.1
Reklam (13)		% 12.1
Biyoteknoloji/Genetik (12)		% 11.2
Bilim (11)		% 10.3
Sigortacılık (11)		% 10.3
Sağlık (10)		% 9.3
İmalat (9)		% 8.4
E-ticaret (8)		% 7.5
Web Kullanım Madenciliği(8)		% 7.5
Sosyal Politikalar/Anket Analizi(8)		% 7.5
Tıp/Farmakoloji (8)		% 7.5
Güvenlik/Anti-terör (6)		% 5.6
Web içerik madenciliği (6)		% 5.6
Kamu/Askeri uygulamalar (4)		% 3.7
Seyahat (3)		% 2.8
Junk e-posta / Anti-spam tespiti (3)		% 2.8
Eğlence /Müzik (3)		% 2.8
Sosyal Ağlar(2)		% 1.9

Kaynak: Akman, 2010: 11.

3.5. VERİ MADENCİLİĞİNDE KULLANILAN MODELLER

Modelleme aşamasında elde edilen verilerden bilgiye ulaşmak için çeşitli çözümleme yöntemlerinden faydalanılmaktadır. Üzerinde çalışılacak veri setine uygun olan modelin belirlenip uygulanması sonrasında parametreler de uygun değişkenlere dönüştürülmelidir (Albayrak ve Yılmaz, 2009: 37).

Veri madenciliği mevcut veri içerisinde var olan geçerli, potansiyel kullanışlı, kolayca anlaşılabilen korelasyonlar ve örüntüler bularak amacına ulaşmaktadır.

Veri madenciliğinin bu amacı tanımlayıcı ya da tahmin edici modeller yaratarak sağlanmaktadır (Sondwale, 2015: 262-265).

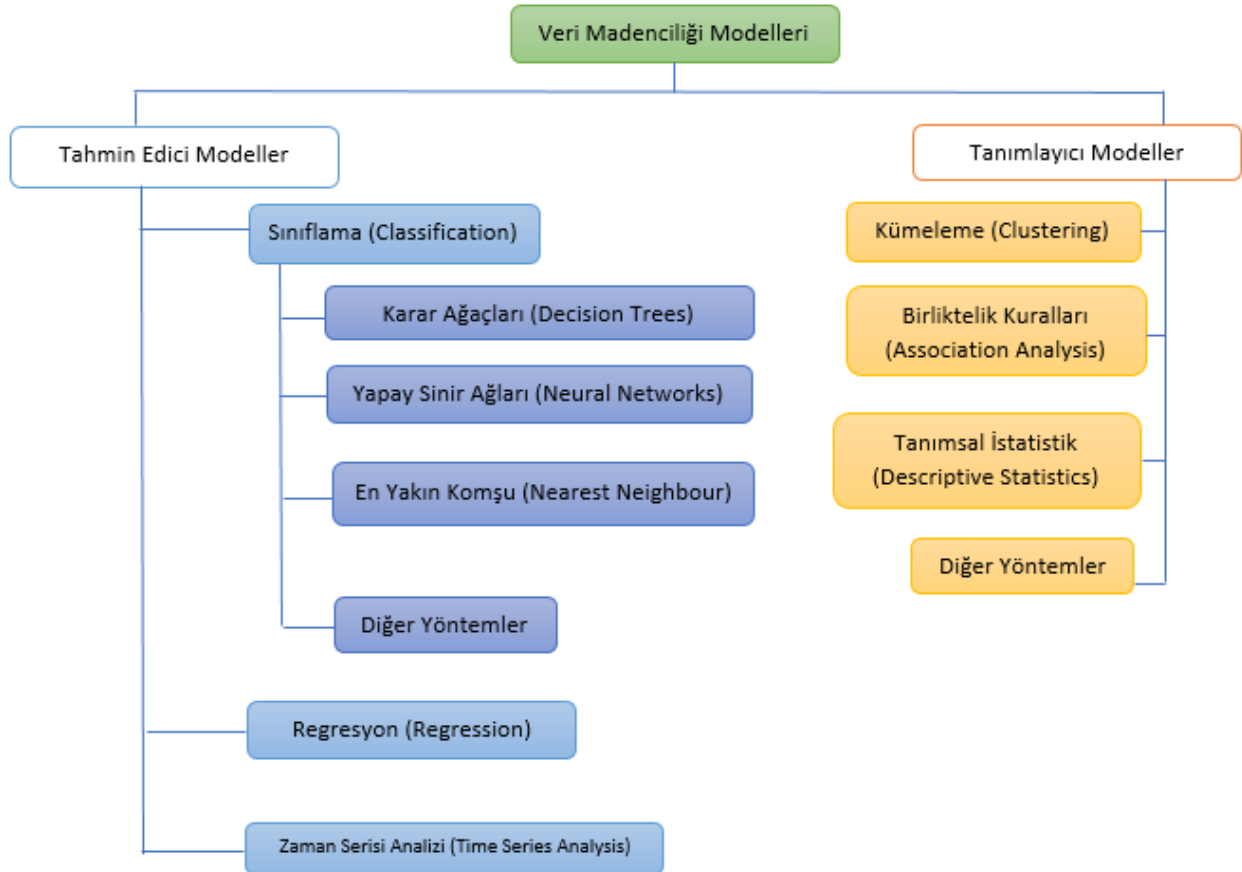
Tanımlayıcı modellerde, karar vermeye rehberlik etmede kullanılabilecek mevcut verilerdeki örüntülerin tanımlanması sağlanmaktadır. X/Y aralığında geliri ve iki veya daha fazla arabası olan çocuklu aileler ile, çocuğu olmayan ve geliri X/Y aralığından düşük olan

ailelerin satın alma örüntülerinin birbirlerine benzerlik gösterdiğinin belirlenmesi tanımlayıcı modellere bir örnektir (Akpınar, 2000: 5).

Tahmin edici bir model, veritabanındaki bazı değişkenleri kullanarak, bilinmeyen ya da ilgili diğer değişkenlerin gelecek değerlerini tahmin etmemizi içermektedir (Fayyad ve diğerleri, 1996: 37-39). Tahmin edici model de, modeli "öğrenmek" için önceden bilinen bir veri kümesine ihtiyaç duymaktadır (Kotu ve Deshpande, 2015). Bu modellerde, sonuçları önceden bilinen veriler üzerinden çalışılarak yeni bir geliştirilecek modellerden faydalanılarak sonuçları bilinmeyen veri setleri için çıktı değerlerinin elde edilebilmesi hedeflenmektedir (Karagöz, 2007).

Sınıflandırma, regresyon modelleri ve zaman serisi analizleri tahmin edici modellerdir. Veri madenciliği modellerinin tahmin edici ve tanımlayıcı olarak sınıflandırılmasına Şekil 10'da yer verilmiştir.

Şekil 10: Tahmin edici ve tanımlayıcı veri madenciliği modelleri



3.5.1. Sınıflama ve Regresyon Modelleri

Sınıflama ve regresyon, önemli veri gruplarını ortaya koyan ya da gelecek veri eğilimlerini öngören modelleri kurabilen iki veri analiz tekniğidir. Sınıflama kategorik değerleri tahmin ederken, regresyon süreklilik gösteren değerlerin tahmin edilmesinde kullanılır (Han and Kamber, 2006: 10).

Sınıflama yöntemi, verinin önceden belirlenen çıktılarına uygun bir şekilde ayrıştırılmasını sağlayan bir yöntemdir. Çıktılar, analiz öncesinde bilindiği için sınıflama, veri kümesini denetimli (supervised) bir şekilde öğrenir (Giudici, 2003: 85).

Mevcut verilerden hareketle gelecekte elde edilecek çıktının tahmin edilmesinde yararlanılan ve veri madenciliği teknikleri içerisinde oldukça fazla bir kullanıma sahip olan sınıflandırma ve regresyon modelleri arasındaki en temel fark, tahmin edilen bağımlı değişkenin kategorik veya süreklilik gösteren bir değer almasıdır. Ancak çok terimli lojistik regresyon gibi kategorik değerlerin de tahmin edilmesine olanak sağlayan tekniklerle, her iki model giderek birbirine yaklaşmakta ve bunun bir sonucu olarak aynı tekniklerden yararlanılması mümkün olmaktadır (Akpınar, 2000: 6).

Sınıflandırma ve regresyon modellerinde kullanılan başlıca algoritmalar şöyle verilebilir (Kulluk, 2009: 30):

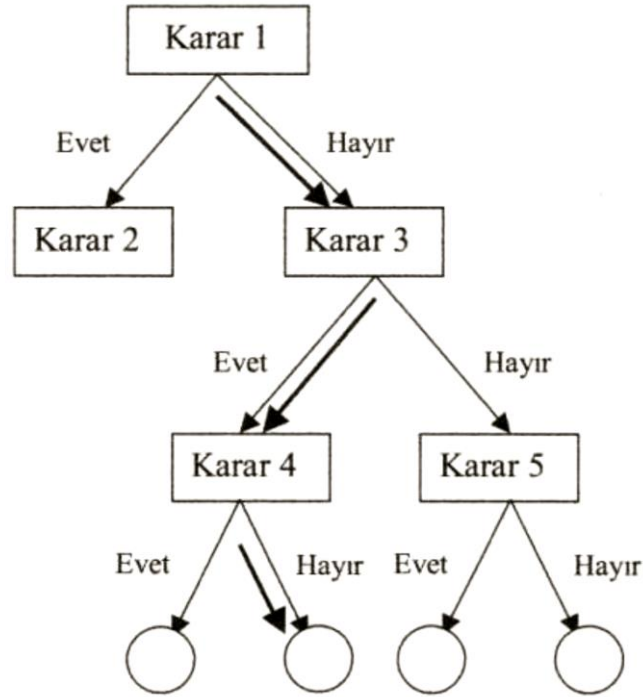
- Karar ağaçları,
- Yapay sinir ağları
- K-en yakın komşu
- Bayes sınıflandırıcılar
- Doğrusal regresyon
- Doğrusal olmayan regresyon
- Lojistik regresyon
- Sürü zekası teknikleri
- Durum tabanlı nedenleme
- Kaba küme yaklaşımı
- Genetik algoritmalar
- Bulanık küme yaklaşımı

3.5.1.1. Karar (Sınıflandırma) Ağaçları

Karar ağaçları, sınıfları önceden tespit edilmiş örnek veriden tümevarım yöntemiyle öğrenilen ağaç şekilli bir karar yapısı tekniğidir. Bir karar ağacı, basit karar verme adımları uygulanarak, büyük miktarlardaki kayıtları, çok küçük kayıt gruplarına bölerek kullanılan bir yapıdır. Her başarılı bölme işlemiyle, sonuç gruplarının üyeleri bir diğeriyle çok daha benzer hale gelmektedir. Büyük veri tabanlarında kullanılan birçok sınıflandırma problemi ile hata içeren bilgilerde karar ağaçları oldukça faydalı çözümler sunmaktadır (Albayrak ve Yılmaz, 2009: 31-52)

Karar ağaçları, adında da belirtildiği üzere ağaç şeklinde, tahmin edici bir yöntemdir. Ağaç yapısı ile, kolay anlaşılabilen kurallar ortaya koyan, bilgi teknolojileri işlemleri ile entegrasyonu kolaylıkla sağlanabilen en kullanışlı sınıflama tekniklerindendir (Serper, 1997: 202).

Şekil 11: Karar ağaçları



Karar ağaçları tekniğinde verilen bir başlangıç noktasından itibaren daha küçük dilimlere bölünür.

Karar ağaçları tekniğinde tümevarım metodu uygulanır. Veriler başlangıç noktasından itibaren adım adım küçük alt kümelerle bölünür. Her bir bölüm için aynı seviyedeki en iyi değişken tespit edilir. Karar verildikten sonra sistem geriye dönük olarak diğer bölümlerin daha iyi sonuç üretilip üretmeyeceği konusunda sorgulama yapmaz. Verilerin kümelere ayrıştırılması bir noktada durur. Bu nokta maksimum derinliğe ulaşıldığı durumdadır (Baunsaythip ve diğerleri, 2001: 121)

Karar ağaçları baz alınarak geliştirilmiş birçok algoritma bulunmaktadır. Bu algoritmalar birbirlerinden kök, düğüm ve dallanma gibi kriter seçimleri sürecinde izlenen yol açısından ayırım gösterirler (Tezcanlar, 2007: 103).

En sık uygulanan karar ağacı algoritmaları şu şekilde verilebilir;

- Chi-Squared Automatic Interaction Detector (CHAID, Kass 1980),
- Classification and Regression Trees, Breiman ve Friedman (C&RT, 1984),
- Induction of Decision Trees, Quinlan (ID3, 1986),
- Quinlan (C4.5, 1993).

3.5.1.2. Yapay Sinir Ağları (YSA)

Yapay sinir ağları, insan beyninin işleyişinden yola çıkılarak; öğrenme yoluyla yeni bilgiler üretebilme, keşfetme ve oluşturma gibi vasıfları herhangi bir yardım olmadan otomatik olarak gerçekleştirebilmek amacıyla geliştirilmiş bilgisayar sistemleridir. Geleneksel programlama yöntemleri ile bu özelliklerin gerçekleştirilmesi oldukça çok zordur. Bu sebeple, yapay sinir ağları pek olası olmayan olaylar ya da programlanması oldukça zor olan olaylar için geliştirilen bilgi işlemeye dayalı bir bilgisayar bilim dalı olarak tanımlanabilir.

Yapay sinir ağlarının eksik, hatalı ve karmaşık bilgileri işlemede oldukça başarılı sonuçlar vermesi sebebiyle pek çok problem çözümünde uygulandığı görülmektedir. Yapılarındaki paralellikten dolayı çok hızlı çalışabilmeleri sebebiyle özellikle gerçek zamanlı olaylarda kolaylıkla çalışmalarını sağlamaktadır.

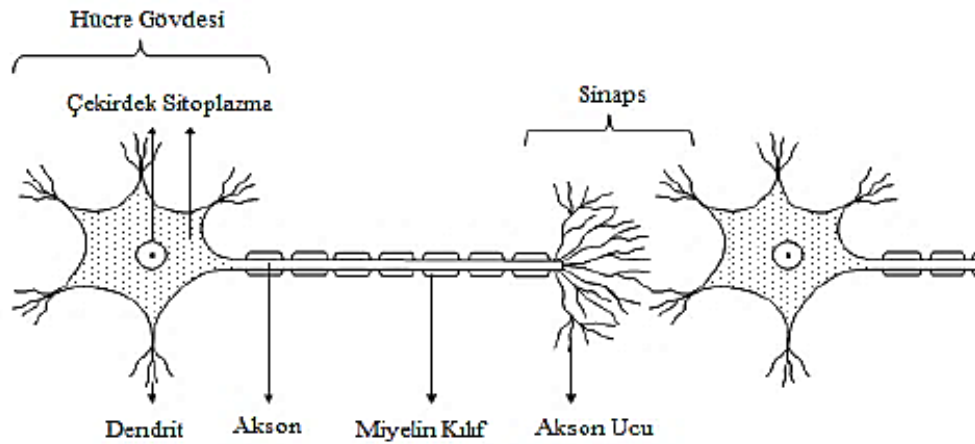
İnsan beyninin fonksiyonel özelliklerinden,

- ✓ Öğrenme,
- ✓ İlişki kurma ve tahminleme
- ✓ Sınıflandırma ve modelleme,
- ✓ Özellik belirtme,
- ✓ Optimizasyon,

konularında başarılı bir şekilde uygulandığı örneklere sıkça rastlanmaktadır (Öztemel, 2003: 29).

Biyolojik sinir ağları beynimizde bulunan birçok sayıda sinir hücresinin bir parçasıdır. İnsan beyni birçok sinir hücresinden oluşup, bu hücreler birbiriyle ilişkilidir. Yapay sinir ağlarının çalışma sistemi biyolojik sinir ağlarının karakteristik özelliklerine benzemektedir (Gan, 2012: 340-349). İnsan beyninin en temel parçası, hatırlama, düşünme, her harekette daha önceki deneyimlere başvurma yeteneğini sağlayan kendine özgü sinir hücreleridir. Bir sinir ağı milyarlarca sinir hücresinin veya bağlantının bir araya gelmesiyle meydana gelmektedir. Biyolojik sinir sistemi yaklaşık 10^{12} adet sinir hücresi ve bu sinir hücrelerinin 6×10^{13} kadar bağlantılarının oluşturduğu; bilgiyi veya girdiyi alan, yorumlayan, değerlendiren ve uygun bir karar üreten bir sistemdir (Öztemel, 2003: 45).

Şekil 12: Biyolojik sinir hücresi ve bileşenleri



Kaynak: Öztemel, 2003: 45.

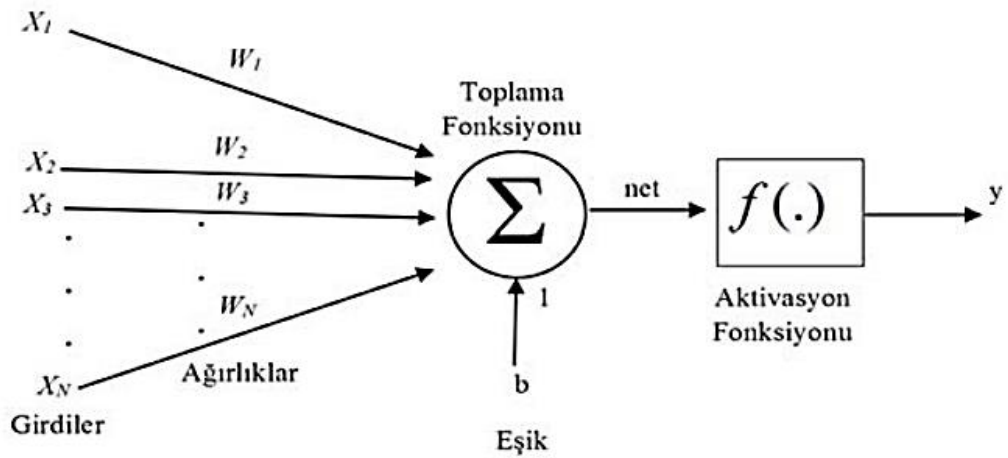
Bir yapay nöron temel olarak girişler, ağırlıklar, toplama fonksiyonu, aktivasyon (etkinlik) fonksiyonu ve çıkış olmak üzere beş kısımdan oluşmaktadır. Yapay sinir ağları, insan beyninin çalışma prensibinden esinlenerek geliştirilmiş olup aralarında yapısal açıdan benzerlikler görülmektedir. Bu benzerlikler Tablo 3'te görülebilir.

Tablo 3: Yapay sinir ağı ve biyolojik sinir ağının karşılaştırılması

Biyolojik Sinir Ağı	Yapay Sinir Ağı
Sinir Sistemi	İşlem elemanı
Sinir	Toplama fonksiyonu
Sinapslar	Sinirler arası bağlantı ağırlıkları
Dendrit	Toplama fonksiyonu
Aksonlar	Yapay sinir çıkışı
Hücre Gövdesi	Transfer fonksiyonu (Etkinlik İşlevi)

Kaynak: Bose ve Liang, 1996: 3-31.

Şekil 13: Yapay sinir hücresi



Yapay sinir ağları kullanılarak yapılan analizlerde, elde edilebilecek birçok avantajdan bazıları aşağıda verilmiştir. Yapay sinir ağlarının diğer sistemlerle karşılaştırılmasında bazı üstün özellikleri vardır (Symeonidis, 2000: 17-18).

- Şekil tanıma, ilişkilendirme, sınıflandırma ve genelleme için kuvvetli bir tekniktir.
- YSA'lar, geleneksel hesaplama yöntemlerine oranla problem çözümünde zamanı daha etkin kullanırlar.
- Yapay sinir ağları değişen ortamlarda esneklik sahibidir.
- Yapay sinir ağları, programlama yerine öğrenme yoluyla kendini geliştirir.
- Yapay sinir ağların performansı klasik (geleneksel) istatistiksel yöntemler kadar iyidir. Sinir ağları daha az zamanda veri yapısını açık şekilde yansıtan modeller oluşturur.
- Yapay sinir ağları doğrusal olmayan verilerle ve matematiksel olarak modellenemeyen problemler için kolay uygulama sağlar.
- Yapay sinir ağlarının uygulama maliyeti daha ucuzdur bu da tekniğin daha çok tercih edilmesine sebep olabilir.

Yapay sinir ağlarının yukarıda belirtilen birçok avantajına karşılık bazı dezavantajları da vardır (Elmas, 2003; Öztemel, 2003). Bunlar aşağıdaki gibi özetlenebilir:

- Yapay sinir ağının probleme uygun ağ yapısının belirlenmesi deneme yanılma yöntemiyle gerçekleştirir.
- Sinir ağlarının donanıma bağımlı olarak çalışmaları önemli bir sorun teşkil eder.
- Ağın eğitime ne zaman son verileceğine karar vermek için kesin bir yöntem teknik yoktur.
- Bazı ağlarda, ağın parametre değerlerinin öğrenme katsayısı, katman sayısı vb. belirlenmesinde bir kural bulunmaması diğer bir sorundur.
- Ağın davranışlarının açıklanamaması.
- İstatistiksel çözümlemelerin beraberinde sorun alanına ilişkin anlaşılabilir ve yorumlamaya olanak veren parametreler üretmesine rağmen, bu parametrelerin henüz yorumlama imkanı bulunmamasıdır.

YSA, motorlu bir taşıtın yönlendirilmesi, karakterize edilmemiş DNA dizilerindeki genlerin tanınması, döviz kurlarının öngörülmesi gibi ilginç görevlerin yerine getirilmesini sağlayacak modellerin öğrenilmesini yerine getirebilecek modeller için çok çeşitli alanlarda uygulanmaktadır (Craven ve Shaylik, 1997: 211-229).

YSA'lar eksik veri ile çalışabilme ve normal olmayan verileri işleyebilme yeteneklerinden dolayı günümüzde geliştirilmiş en mükemmel örüntü tanıyıcı ve sınıflandırıcılardan sayılabilirler (Öztemel, 2003: 35).

Günümüzde YSA uygulamalarını; Tahmin (hava durumu tahmini, pazar payı tahmini vb.), sınıflandırma (karakter tanıma, hastalık teşhis etme vb.), veri ilişkilendirme (karakterleri sınıflandırmanın yanında karakterleri ilişkilendirerek kelimeleri kontrol etme gibi), veri yorumlama (benzer verileri gruplandırılması, kalite kontrolü vb.) ve veri filtreleme (gürültü ayıklama, ses temizleme, görüntü netleştirme vb.) olmak üzere beş sınıfa ayırmak mümkündür (Sağıroğlu ve diğerleri, 2003).

3.5.1.3. Genetik Algoritmalar

Genetik algoritmalar evrimsel hesaplama (evolutionary computing) metotlarının örnekleridir ve optimizasyon tipli algoritmalar (Dunham, 2003).

Karmaşık çok boyutlu uzayda en iyinin hayatta kalması felsefesine göre en iyi çözümü aramaktadırlar. Genetik algoritmalar problemlere sadece tek bir çözüm üretmez, farklı birçok çözümün oluştuğu bir çözüm kümesi üretir. Bu sayede arama uzayında aynı anda pek çok nokta değerlendirilebilir ve sonuçta bütünsel çözüme ulaşılma olasılığı da artar.

Genetik algoritmalar problemlerin çözümünde evrimsel süreci bilgisayar ortamına uyarlar. Diğer en iyileme yöntemlerinde de olduğu gibi çözüm için tek bir yapının geliştirilmesi yerine, böyle yapılardan meydana gelen bir küme oluştururlar. Problem için olası pek çok çözümü temsil eden bu küme genetik algoritma terminolojisinde nüfus adını alır. Nüfuslar vektör veya birey adı verilen sayı dizilerinden oluşur. Birey içindeki her bir elemana gen adı verilir. Nüfustaki bireyler evrimsel süreç içinde genetik algoritma işlemcileri tarafından belirlenir (Ayık ve diğerleri, 2007: 446).

3.5.1.4. Bellek Tabanlı Yöntemler (K En Yakın Komşu)

Bellek Tabanlı Yöntemler, günümüzde bilgisayarların ucuzlaması ve kapasitelerinin artmasıyla, özellikle de çok işlemcili sistemlerin yaygınlaşmasıyla, kullanılabilir hale gelmiştir (Han ve Kamber, 2006).

Bu yönteme en iyi örnek en yakın k komşu algoritmasıdır (k-nearest neighbour) (Alpaydın, 2000: 1-5). Bu algoritma ile sınıflandırılacak olan nesne, öz değerlerine göre kendisine en yakın komşu veya komşular sınıfına atanmaktadır. Performansı en yakın komşuları tanımlamak için kullanılan mesafe metriğine göre belirlenmektedir (Weinberger ve Saul, 2009: 207-244).

3.5.2. Kümeleme Analizi (Clustering)

Kümeleme, verilerin kümelerle gruplandırılması işlemidir, böylece, ancak diğer kümelerdeki nesnelere kıyasla, bir kümedeki nesneler birbirleri ile yüksek benzerlik göstermektedir. Amaç, gerçekleştirilen bölümlendirme işlemi sonunda elde edilen kümelerin, küme içi benzerliklerinin maksimum ve kümeler arası benzerliklerinin ise minimum olmasını sağlamaktır. Genellikle mesafe ölçütleri kullanılır (Han ve Kamber, 2006). Sınıflandırma yöntemleri öngöründe bulunurken, kümeleme yöntemleri tanımlayıcıdır (Maimon ve Rokach, 2005:1-17).

3.5.3. Birliktelik Kuralları (Association Rules)

Olayların birlikte gerçekleşme durumlarını çözümleyen veri madenciliği yöntemlerine birliktelik kuralları (Association Rules) adı verilmektedir. Bu yöntemler, birlikte gerçekleşme kurallarını belirli olasılıklarla ortaya koymaktadır (Özkan, 2008: 157). Genellikle perakende sektöründe “Pazar sepet analizi” olarak adlandırılır (Bland, 2002: 45).

Bir alışveriş sırasında ya da art arda yapılan alışverişlerde müşterinin hangi ürün ya da hizmetleri satın alma eğiliminde olduğunun belirlenmesi, müşteriye daha çok ürünün satılmasını da sağlama yollarından biridir. Satın alma eğilimlerinin tanımlanmasına olanak sağlayan birliktelik kuralları ile ardışık zamanlı örüntüler, pazarlama amacı taşıyan pazar sepeti analizi adı ile veri madenciliğinde sıkça kullanılmaktadır. Bunun yanı sıra bu teknikler, tıp, finans, pazarlama ve farklı olayların birbiri ile ilişkili olduğunun belirlenmesi sayesinde değerli bilgi kazanımının olduğu ortamlarda da fazla öneme sahiptir (Akpınar, 2000: 7).

Birliktelik kuralları aşağıda verilen örneklerde görüleceği gibi eş zamanlı olarak gerçekleşen ilişkilerin tespit ve tanımının yapılmasında kullanılır.

- Müşteriler gazoz satın aldıklarında, %75 ihtimalle cips de alırlar,
- Az yağlı peynir ve yağsız süt alan müşteriler, %85 ihtimalle az yağlı yoğurt da satın alırlar.

Ardışık zamanlı örüntüler ise aşağıda verilen örneklerde görüleceği gibi birbirleri ile ilişkisi olan fakat birbirini izleyen dönemlerde gerçekleşen ilişkilerin tanımında kullanılır (Akpınar, 2000: 7).

- X ameliyatı yapıldığında, 15 gün içinde %45 ihtimalle Y enfeksiyonu oluşacaktır,

- Çekiş satın alan bir müşteri, ilk üç ay içerisinde %15, bu dönemi izleyen üç ay içerisinde %10 ihtimalle çivi satın alacaktır.

Birliktelik kurallarının örnek uygulamaları aşağıdaki gibidir (Mitra ve Acharya, 2003: 267-290);

- Market - sepet analizi (market –basket analysis),
- Direkt satışta başka bir ürün önermek,
- Kredi kartı sahtekarlığını ortaya çıkarmak,
- Standların nasıl dizilmesi gerektiğini ortaya koymak

Birliktelik kurallarının matematiksel modeli 1993 yılında Agrawal, Imielinski ve Swami tarafından sunulmuştur (Agrawal ve diğerleri, 1993: 207-216).

3.6. TIP VE SAĞLIK HİZMETLERİNDE VERİ MADENCİLİĞİ VE UYGULAMALARI

Sağlık ve Tıp sektörü bilginin en hızlı değiştiği alanlar arasında yer alır. Sağlık hizmetlerinin en hızlı, en güvenli, en yüksek kalitede ve ihtiyaca yönelik bir şekilde sunulabilmesi için sağlık profesyonellerinin en güncel ve doğru bilgiye ulaşması ayrıca bu bilgiyi doğru bir şekilde kullanması gerekmektedir. Veri madenciliği, Sağlık sektöründe yeni ilaç geliştirilmesi, hastalıkların teşhis edilmesi, tedavi süreçlerinin belirlenmesi gibi alanlarda yaygın olarak kullanılmaktadır.

Veri Madenciliği fazla miktardaki veri içerisinde, gizli, değerli ve kullanılabilir bilgileri ortaya çıkarmak ve stratejik karar desteği sağlamak amacıyla kullanılan; verinin analizini temel alıp karar verme modelleri yaratan bir yöntemdir. Bu sebeple sağlık hizmetlerinin sunumu, sağlık kurumlarının yönetilmesi ve sağlık politikalarının geliştirilmesinde bir karar destek aracı olarak Veri Madenciliği'nin kullanılması sağlık profesyonellerinin optimum kararı almasına yardımcı olacaktır (Koyuncugil ve Özgürbaş, 2009: 27-28).

DÖRDÜNCÜ BÖLÜM

KÜMELEME ANALİZİ VE MÜŞTERİ SEGMENTASYONU

İstatistik çalışmalar çok sayıda veriye ihtiyaç duymaktadır. Ancak verilerin çokluğu birçok istatistik tekniğin kullanımında sorunlar yaratmaktadır. Böylece tekniklerin kullanılabilirliğini sağlamak için verilerin gruplandırılması gereği ortaya çıkmaktadır.

Küme, birbirine benzer nesnelerin çok boyutlu uzayda oluşturdukları bütünler olarak tanımlanabilirken, kümeleme analizi ise; bu kümeleri oluşturma sürecidir. Kümeleme analizi için yapılabilecek bir diğer tanım ise; küme adı verilen homojen grupların sınıflandırılması için kullanılan tekniklerdir (Akın, 2008: 5). Böylelikle küme içerisindeki nesneler, örneklendikleri kitlenin özelliklerini en iyi şekilde yansıtan bir temsil gücüne sahip olmaktadır. Kümeleme, sınıflamanın tersi şekilde yeniden tanımlanmış sınıflara dayalı olmaz. Kümeleme, bir denetimsiz öğrenme (unsupervised learning) tekniğidir (Aydoğan, 2003: 88). Kümelemede amaç, küme içerisindeki elemanların birbirine çok benzediği, ancak özellikleri bakımından birbirinden oldukça farklı olan kümelerin oluşturulması ve kayıtların bu farklı kümelere (gruplara) bölünmesidir (Durmuş ve İplikçi, 2007).

Kümeleme analizi (clustering) veri matrisinde yer almakta olan ancak doğal grupları kesin olarak bilinmeyen X birimi, değişkeni birbirleri ile benzer olan alt kümelere, grup veya sınıflara ayırmaya yardımcı olan yöntemler topluluğudur. Kümeleme analizi, birimleri değişkenler arası benzerlik ya da uzaklıklara dayalı olarak hesaplanan bazı ölçülerden faydalanılarak homojen gruplar oluşturmaya çalışır (Özdamar, 2004: 279)

Kümeleme analizinde bağımlı ve bağımsız değişkenler arasındaki fark veya üstünlük gözetilmeyip, birbirine bağımlı tüm ilişkiler incelenir. Kümeleme analizi ile hedefi belirlenen değişkenler grubuna dayalı homojen grupların içerisine nesnelerin sınıflandırılması yapılmaktadır. Aynı grup içerisinde yer alan nesneler; değişkenler bakımından birbirlerine benzerlik gösterebilirken diğer gruptaki nesnelerden farklıdırlar (Malhotra, 1996: 670-672).

Kümeleme işlemi, heterojen bir yapıya sahip kitlenin daha homojen birkaç alt gruba ya da kümeye bölünmesi işlemidir. Sınıflandırma ve kümelemenin birbirinden ayrılan en önemli farkı, kümeleme işleminde sınıflandırmada olduğu gibi önceden belirlenen sınıflara göre bölümlendirme yapılmamasıdır. Sınıflandırmada her bir veri, önceden sınıflandırılmış veya belirlenmiş sınıflar üzerinde yapılan bir eğitim sonucunda elde edilen bir modele göre, önceden belirlenmiş olan bir sınıfa atanmaktadır. Kümeleme işleminde ise önceden belirlenmiş sınıflar veya örnek sınıflar yer almamaktadır. Verilerin kümelenebilmesi, birbirlerine olan benzerliklerine dayalı olarak yapılmaktadır. Ortaya çıkan sınıfların hangi anlamları taşıdığının belirlenmesi ise tamamıyla çözümlemeyi yapan kişiye kalmaktadır (Vatansever, 2008: 53).

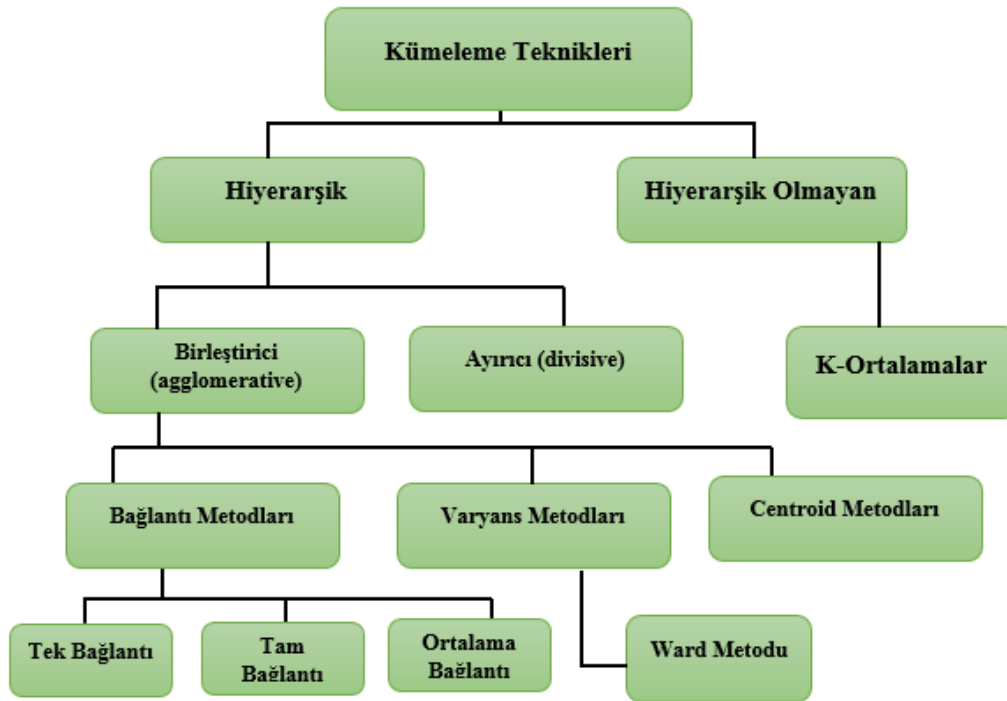
4.1. KÜMELEME ANALİZİ TEKNİKLERİ

Kümeleme teknikleri; uzaklık matrisinin kullanılarak nesnelerin ya da değişkenlerin kendi içinde homojen, kendi aralarında ise heterojen gruplar oluşturmaya olanak vermektedir. Kümeleme analizi için pek çok algoritma öne sürülmektedir. Ancak literatürde bu

algoritmalar iki başlık altında toplanabilir: Hiyerarşik kümeleme teknikleri ile Hiyerarşik olmayan kümeleme tekniği (Ketchen ve diğerleri, 1996: 444). Her iki teknikte de karşımıza çıkan temel amaç kümelerin arasındaki farklılıklar ile küme içi benzerliklerini en yüksek seviyeye çıkarmaktır. Hangi tekniğin kullanılacağı ise küme sayısına bağlı değişmekle beraber her iki tekniğin birlikte kullanılması daha faydalı sonuçlar verecektir. Bu sayede sonuçların ve de iki tekniğin verdiği sonuçların uygunluğunu karşılaştırmak mümkün olabilmektedir (Akın, 2008: 8).

Kümeleme tekniklerine ait genel yapı Şekil 14’te verilmiştir

Şekil 14: Kümeleme tekniklerinin genel yapısı



4.1.1. Hiyerarşik Kümeleme Yöntemleri

Hiyerarşik kümeleme yöntemi, kümelerden bir elemanı silme ya da eklemeye bir ağaca benzeyen yapı gösteren aşamalar grubudur (Ketchen and Shook, 1996: 444). Hiyerarşik kümeleme yöntemleri veri setindeki birimlerin birbirlerine göre uzaklık veya benzerliklerini dikkate alarak birimleri birbirleriyle değişik aşamalarda bir araya getirerek ardışık biçimde kümeler belirlemeye ve bu kümelere girecek elemanların hangi uzaklık veya benzerlik düzeyinde küme elemanlarının olduğunu belirlemeye yönelik yöntemlerdir (Özdamar, 2004: 293).

Hiyerarşik kümeleme yöntemleri, ayrıcı ve birleştirici olmak üzere iki kategoriye ayrılmaktadır. Birleştirici hiyerarşik kümeleme yöntemlerinde, öncelikle her gözlem bir küme olarak kabul edilmekte ve daha sonra birbirine yakın olan kümeler birleştirilerek küme sayıları her aşamada azaltılacak şekilde birleştirme işlemi yapılmaktadır. Ayrıcı hiyerarşik kümeleme yöntemlerinde ise, işlemler tüm gözlemlerin tek bir kümede yer almasıyla başlamakta ve küme içerisinde benzer olmayan gözlemlerin ayrılmasıyla daha küçük kümeler oluşturularak devam etmektedir (Timm, 2002: 523).

Hiyerarşik kümeleme yöntemleri, küme sayısının bilinemediği durumlarda kullanılır. Hiyerarşik kümeleme yöntemleri, gözlemleri veya değişkenleri kümelerken uzaklık ve benzerlik ölçülerini dikkate almalarının yanı sıra kümelene yapısının daha kolay anlaşılabilmesi için dendrogram (ağaç diyagramı) olarak bilinen grafikleri kullanılmaktadır (Alpar, 2011: 314).

Hiyerarşik yöntemlerde kümeleme analizi kümelerin elde edilmesinde hızlı bir yöntem olmasına karşın büyük örneklemelerin olduğu ya da çok fazla değişken içeren verilerde kolaylık sağlamamaktadır (Hair ve diğerleri, 2014: 445).

Uygulamalarda en sık karşımıza çıkan birleştirici hiyerarşik kümeleme yöntemleri; tek bağlantı yöntemi, tam bağlantı yöntemi, ortalama bağlantı yöntemi, merkez bağlantı yöntemi ile Ward yöntemidir.

Şekil 15: Hiyerarşik kümelemede sıklıkla kullanılan yöntemler

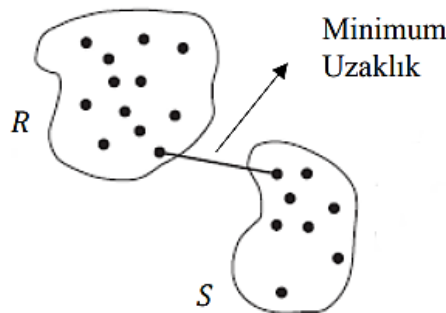


Kaynak: Hair ve diğerleri, 2014: 445.

4.1.1.1. Tek Bağlantı Yöntemi

Tek bağlantı yöntemi en basit kümeleme yöntemi olup, en yakın komşuluk yöntemi olarak da bilinmektedir (Anderberg, 1973: 137). Tek bağlantı yönteminde, kümeleme sürecinin başında hesaplanan uzaklık matrisinde en küçük uzaklığa sahip olan değerler alınmaktadır (Mooi ve Sarstedt, 2011: 250). Eğer benzerlik matrisi ile ilgileniliyorsa en büyük benzerlik değerine sahip olan değerler alınmaktadır.

Şekil 16: Tek bağlantı tekniği



Kaynak: Hair ve diğerleri, 2014: 442.

En küçük uzaklığa sahip iki gözlem bir kümeye yerleştirildikten sonra oluşturulan bu kümeye, en yakın uzaklıkta (ya da en büyük benzerlikte) olan yeni bir gözlem veya yeni bir küme eklenmektedir. Bu işlem tek bir küme elde edilinceye kadar devam etmektedir.

Şekil 16'da verildiği gibi R ve S iki küme olsun. x_i , R kümesinin i . elemanını ve x_j de S kümesinin j . elemanını gösterebilir. Bu durumda d_{RS} , R ve S kümeleri arasındaki uzaklığı göstermek üzere;

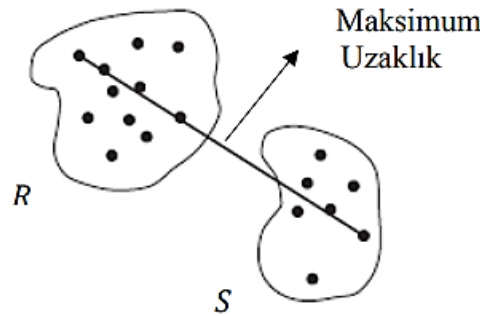
$$d_{RS} = \min[d(x_i, x_j), x_i \in R \text{ ve } x_j \in S]$$

olarak elde edilmektedir (Timm, 2002: 524).

4.1.1.2. Tam Bağlantı Yöntemi

Tam bağlantı yöntemi, en uzak komşuluk ya da çap yöntemi olarak da adlandırılmaktadır (Hair ve diğerleri, 2014: 441). Bu yöntem, tek bağlantı yöntemi ile çok benzer bir süreçle ilerlemektedir. İki yöntemin birbirinden ayrılan yönü; tek bağlantı yöntemi minimum uzaklıkları kullanırken, tam bağlantı yöntemi iki gözlem arasındaki maksimum uzaklığı (ya da minimum benzerliği) kullanmaktadır (Johnson ve Wichern, 1998: 744).

Şekil 17: Tam bağlantı tekniği



Kaynak: Hair ve diğerleri, 2014: 442.

Şekil 17'de verildiği gibi R ve S iki küme olsun. x_i , R kümesinin i . elemanını ve x_j de S kümesinin j . elemanını gösterebilir. Bu durumda d_{RS} , R ve S kümeleri arasındaki uzaklığı göstermek üzere;

$$d_{RS} = \max[d(x_i, x_j), x_i \in R \text{ ve } x_j \in S]$$

olarak elde edilmektedir (Timm, 2002: 526).

4.1.1.3. Ortalama Bağlantı Yöntemi

Tam ve tek bağlantının uzak ya da yakın değerlerle, yani küme içindeki aşırı değerlerle hesaplanması bazen dezavantaj oluşturmaktadır. Bu gibi durumlarda ortalama bağlantı yöntemi tercih edilecek bir tekniktir.

Ortalama bağlantı yönteminde, iki ayrı kümede yer alan gözlem çiftleri arasında yer alan ortalama uzaklıklardan yola çıkılarak kümeler arası uzaklıklar hesaplanabilmektedir. Ayrıca bu yöntem, küme içi değişkenliği daha küçük olan kümeler oluşturma eğilimindedir. Başka yöntemlere göre aşırı değerlere karşı hassasiyeti daha az olan bir yöntemdir (Hair ve diğerleri, 2014: 441).

R ve S iki küme olmak üzere; x_i , R kümesinin i . elemanını ve x_j de S kümesinin j . elemanını gösterebilir. Bu durumda d_{RS} , R ve S kümeleri arasındaki uzaklığı göstermek üzere;

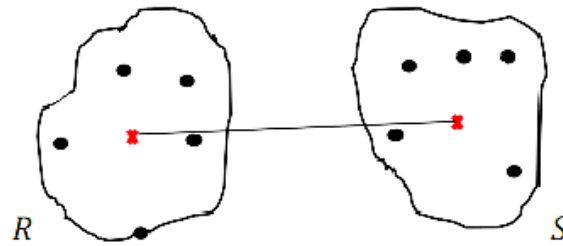
$$d_{RS} = \frac{1}{n_R n_S} \sum_{x_i \in R} \sum_{x_j \in S} d(x_i, x_j)$$

olarak elde edilmektedir (Khattree ve Naik, 2000: 365).

4.1.1.4. Merkez Yöntemi

Merkez yöntemi, Centroid yöntemi olarak da bilinmektedir. Bu yöntemde her kümenin geometrik merkezleri hesaplanmakta ve iki küme arasındaki uzaklık kümelerin merkezleri arasındaki uzaklığa eşit olmaktadır (Mooi ve Sarstedt, 2011: 250).

Şekil 18: Merkez (centroid) yöntemi



Bu yöntemde kümeler arasındaki uzaklık hesaplanırken genellikle Öklid uzaklığı ya da Kareli Öklid uzaklığı kullanılmaktadır (Hair ve diğerleri, 2000: 180). Uzaklıklar, her gözlem gruplandırıldıktan sonra tekrar ortalamaları alınarak elde edilmektedir.

Merkez yönteminde küme uzaklıkları;

$$d_{RS} = d(\bar{x}_R, \bar{x}_S)$$

olarak elde edilmektedir. Burada $\bar{x}_r$; R 'inci kümeye ait gözlem değerlerinin ortalamasını, $\bar{x}_s$ ise S 'inci kümeye ait gözlem değerlerinin ortalamasını göstermektedir. Yeni oluşturulan küme merkezleri ise;

$$\frac{n_R \bar{x}_r + n_S \bar{x}_s}{n_R + n_S}$$

formülü ile hesaplanmaktadır. Burada n_R ; R 'inci kümede yer alan gözlem sayısını, n_S ise S 'inci kümede yer alan gözlem sayısını ifade etmektedir (Timm, 2002: 528). Merkez yönteminin diğer hiyerarşik yöntemlere göre avantajı, küme merkezlerinden hareketle uzaklıkları hesaplayıp, küme büyüklükleri ile ağırlıklandırma işlemi uyguladığı için uç değerlerden daha az etkilenmesidir (Hair ve diğerleri, 2014: 442). Ancak küme büyüklükleri ile ağırlıklandırma yapılırken, küme büyüklükleri arasında çok fark olması durumunda yeni oluşturulan kümenin merkez noktası, daha büyük olan kümeye yakın olma eğilimi gösterecektir (Khattree ve Naik, 2000: 364).

4.1.1.5. Ward Yöntemi

Ward tarafından (1963) ortaya atılan bu teknik, minimum varyans yöntemi şeklinde de adlandırılmaktadır. Ward yöntemi; diğer yöntemlerde olduğu gibi kümeler arası uzaklık bağlantılarını dikkate almaktan çok, kümeler için kare toplamlarıyla ilgilenmektedir. Ward yönteminin amacı, elde edilen kare toplamlarını kullanarak kümeler içi homojenliği maksimum yapmaktır (Sharma, 1996: 193).

Ward yöntemine ilk olarak, her bir kümenin içinde tek gözlem biriminin bulunduğu n tane küme ile başlanır. Her gözlemin bir kümede olmasından dolayı yöntemin ilk adımında kareler toplamı sıfır olmaktadır (Chatfield ve Collins, 1980: 224).

Ward yönteminde kümeler birleştirilirken bilgi kaybının minimum olması gerekmektedir. Bu yöntem uygulanırken, bilgi kaybının minimum olması için hata kareler toplamında minimum artışı sağlayan iki küme birleştirilmektedir (Johnson ve Wichern, 1998: 751). Yöntemin her adımında $n(n - 1)/2$ çift küme elde edilmekte ve hata kareler toplamını en az yapan kümeler birleştirilmektedir. Bu işlem tüm kümeler birleştirilinceye kadar devam etmektedir (Timm, 2002: 529).

$\bar{x}_r$ ve $\bar{x}_s$ sırasıyla R ve S kümelerinin ortalama vektörlerini ve $\bar{x}_t$ 'de iki kümenin birleşiminden elde edilen yeni kümenin ortalamasını göstermek üzere;

$$\bar{x}_t = \frac{n_R \bar{x}_r + n_S \bar{x}_s}{n_R + n_S}$$

olarak elde edilir. Burada T kümesi R ve S kümelerinin birleşiminden oluşmaktadır. R ve S kümelerine ait hata kareler toplamı;

$$SSE_R = \sum_{r \in R} \|x_r - \bar{x}_r\|^2$$

$$SSE_S = \sum_{s \in S} \|x_s - \bar{x}_s\|^2$$

olmak üzere T kümesinin hata kareler toplamı,

$$SSE_T = \sum_{t \in T} \|x_t - \bar{x}_t\|^2$$

olarak ifade edilmektedir. SSE_T , SSE_R ve SSE_S cinsinden,

$$SSE_T = SSE_R + SSE_S + n_R \|\bar{x}_r - \bar{x}_t\|^2 + n_S \|\bar{x}_s - \bar{x}_t\|^2$$

olarak ifade edilebilmektedir. R ve S kümeleri birbirine ne kadar benzerse $\bar{x}_r$ ve $\bar{x}_s$ değerleri birbirine o kadar yakın olacaktır. Dolayısıyla yukarıda verilen formüldeki son iki terim, yani $SSE_T - (SSE_R + SSE_S)$, küme uzaklığının bir ölçüsü olarak alınabilir. Bu ölçü Ward yönteminin temelini oluşturmaktadır. Bu uzaklık ölçüsü;

$$d_{RS} = \frac{n_R n_S}{n_R + n_S} \|\bar{x}_r - \bar{x}_s\|^2$$

ile ifade edilmektedir (Khattree ve Naik, 2000: 367).

4.1.1.6. Medyan Yöntemi

İlk olarak J.C. Gower (1967) tarafından merkez yönteminin ağırlıklandırmadan kaynaklanan olumsuzluklarını gidermeye yönelik olarak ortaya çıkarılmıştır. Medyan yönteminde, oluşturulan yeni kümenin merkezi bu kümeyi oluşturan diğer kümelerin büyüklüklerinden bağımsızdır (Gan, Ma, ve Wu, 2007: 130).

Medyan yöntemi, yeni oluşturulacak kümeler için, birleştirilebilecek kümeler arasında medyanın değerini alarak kümelerin uzaklıklarını hesaplamaktadır. Bu yöntemin ilk adımında, her biri bir gözlemden oluşan n küme olduğunda, bu gözlemlerin tümü medyan olarak değerlendirilerek bir sonraki adıma geçilmektedir.

m_r ve m_s sırasıyla R ve S kümelerinin medyan değerleri olmak üzere yeni oluşturulan kümenin medyan değeri;

$$m_{yeni} = \frac{m_r + m_s}{2}$$

olarak hesaplanmaktadır (Khattree ve Naik, 2000: 364).

k . kümenin R . ve S . kümelerin birleştirilmesiyle elde edilen yeni kümeye uzaklığı;

$$d_{k(rs)} = \frac{1}{2} d_{kr} + \frac{1}{2} d_{ks} - \frac{1}{4} d_{ij}$$

şeklinde ifade edilmektedir (Gan, Ma, ve Wu, 2007: 131).

4.1.2. Hiyerarşik Olmayan Kümeleme

Hiyerarşik olmayan kümeleme teknikleri değişkenlerden çok, gözlem birimlerini gruplandırmak amacıyla kullanılmaktadır. Küme sayısı (k) önceden bilinmekte ya da kümeleme sürecinin bir parçası olarak saptanabilmektedir. Hiyerarşik olmayan yöntemlerde, uzaklık ve benzerlik ölçüleri kullanılmadığından hiyerarşik kümeleme yöntemlerine göre daha çabuk sonuç verebilmektedir (Johnson ve Wichern, 1998: 754).

Hiyerarşik olmayan kümeleme yöntemlerinin, çok büyük veri setlerine uygulanması ve aykırı değerlere karşı az duyarlı olması bakımından çeşitli avantajlar vardır (Alpar, 2011: 333).

Hiyerarşik olmayan kümeleme yöntemlerinde de amaç, birbirine benzeyen birimleri bir araya getirerek kendi içinde homojen, birbirlerine göre heterojen kümeler elde etmektir (Lattin ve diğerleri, 2003: 288).

Hiyerarşik olmayan yöntemler genellikle aşağıdaki adımları izlemektedir (Sharma, 1996: 202):

1. Başlangıç küme merkezi veya orta noktanın seçimi yapılır.
2. Her gözlem en yakında olduğu kümeye atanır.
3. Belirlenen bir durdurma kuralına göre her bir gözlem k tane kümeden herhangi birine yeniden atanabilir.
4. Yeniden atama yapılabilecek gözlem kalmayınca veya başlangıçta belirlenen durdurma kuralı gerçekleşince algoritma durdurulur. Aksi halde ikinci adıma geçilir.

Bu algoritmada orta noktalar birimler arasından tesadüfi olarak seçilmekte ve birimler başlangıçta tesadüfi olarak kümelere atanmaktadır (Johnson ve Wichern, 1998: 755).

İki teknikte de ortak amaç kümelerin arasındaki farklılık ve kümeler içi benzerliklerin en yüksek düzeye çıkarılmasıdır. Yani, küme içi homojenlik artırılırken kümeler arası homojenlik ise azaltılmaktadır.

Tablo 4: Kümeleme yaklaşımları avantaj ve dezavantajları

Kümeleme Yaklaşımları	Avantajlar	Dezavantajlar
Hiyerarşik Kümeleme	Küme sayısını belirleyebilmesi	Büyük veri seti üzerinde çalışma zorluğu, Aykırı değerlerden kolayca etkilenmesi, geri alınamama
Hiyerarşik Olmayan Kümeleme	Başlangıç noktası ve küme sayısı verildiğinde daha yüksek doğruluk, büyük hacimli veriler üzerinde çalışabilme	Küme sayısını belirleyememe, Başlangıç noktasının ve küme sayısının rastgele seçilmesi, aynı gruba ait iki merkezin seçilebilmesi

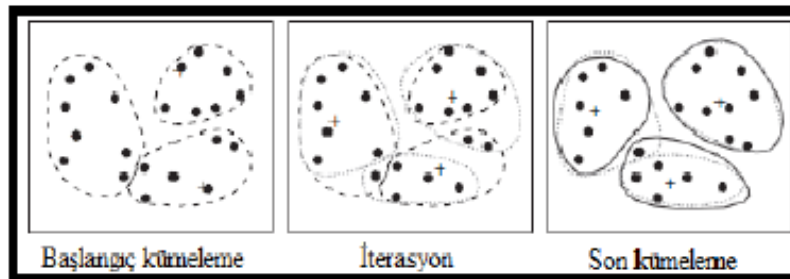
Hiyerarşik olmayan kümeleme yöntemlerinden en sık kullanılan algoritma K-ortalamlar algoritmasıdır. Hiyerarşik olmayan kümeleme yöntemleri, hiyerarşik kümeleme yöntemlerine göre daha büyük veri dosyalarına uygulanabilir.

4.1.2.1. K-ortalamlar (K-means) Yöntemi

K-ortalamlar (k-means) tekniği 1967 yılında Mac Queen tarafından ortaya atılmıştır. Yıllar içerisinde birçok uygulama alanında da oldukça fazla bir biçimde kullanılan bir kümeleme algoritmasıdır. Bu algoritma içerisinde k sayıda küme oluşmakta ve her küme içerisinde bulunan verilerin ağırlıklı ortalamaları sonucu bir değer elde edilmektedir. Küme içerisinde bu değere en yakın olan nokta değeri küme merkezi olarak alınmaktadır (Berkhin, 2002). Başlangıç noktalarına karşı duyarlı olup iyi bir sonuca ulaşmak için farklı başlangıç noktaları ile yöntemi yeniden denemek gerekmektedir (Lattin ve diğerleri, 2003: 288).

Şekil 19'da k-ortalamlar kümeleme yöntemi ile birimlerin kümelenmesi aşamaları verilmiştir.

Şekil 19: K-ortalamlar kümeleme yöntemi ile birimlerin kümelenmesi, küme merkezlerinin güncellenmesi ve buna göre yeniden atanması (her küme ortalaması + ile gösterilmiştir)



Kaynak: Han ve diğerleri, 2006.

K-ortalamalar metodu, pek çok gözlemden elde edilen sürekli p değişkenden oluşmuş veri setinin, küme içi kareler toplamını minimum yapacak şekilde k kümeye ayırmayı hedeflemektedir. Küme sayısı ile ilgili bir bilgi yoksa, uygun küme sayısını deneysel olarak belirleyebilmek mümkündür. Küme sayısı;

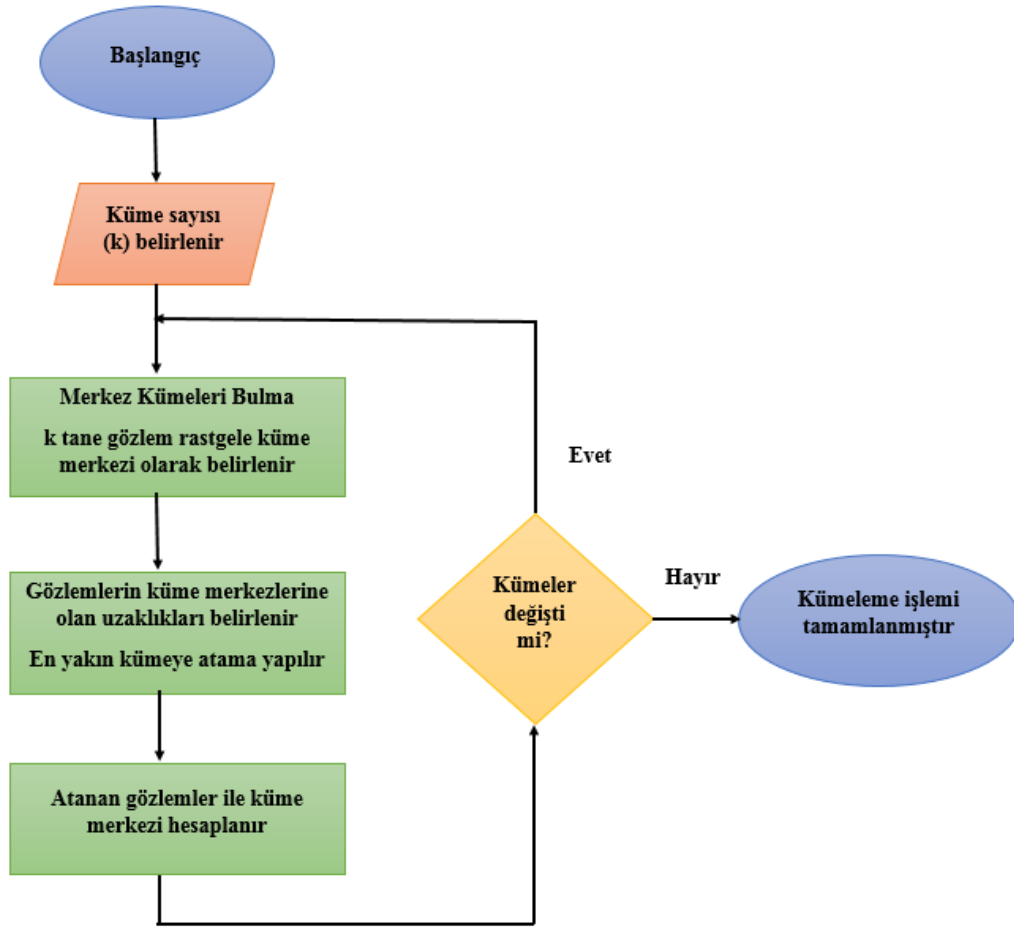
- Hiyerarşik kümeleme yöntemlerinden elde edilen dendogramlar kullanılarak,
- Olasılıklı olarak başlangıç noktalarının tesadüfi seçimi ile,
- Küme sayısını birer birer artırarak en yüksek önemliliğe sahip Wilk's Lambda değerinin elde edildiği küme sayısını alarak belirlenebilir (Özdamar, 2013: 298).

En yaygın kullanılan denetimsiz öğrenme yöntemlerinden biridir. K-ortalamalar algoritması her gözlemin sadece bir kümeye ait olabilmesine izin verir. K-ortalamalar algoritmasının genel mantığı n adet gözlemden oluşan bir veri kümesini, algoritma uygulanmaya başlamadan önce belirlenen k adet kümeye ayırmaktır. Amaç, algoritma ile elde edilen kümeleme işlemi sonunda elde edilen kümelerin, küme içi benzerliklerinin maksimum ve kümeler arası benzerliklerinin minimum olmasını sağlamaktır (Jain, 2010: 651-666).

K-ortalamalar algoritması ilk olarak rastgele seçilen k adet merkez noktayla başlar. Veri kümesindeki her nokta kendisine en yakın küme merkezine ait olan kümeye dahil edilir. Küme merkezlerinin değeri o kümede bulunan noktaların aritmetik ortalaması alınarak hesaplanır. Bu işlem merkezlerin değerleri değişmeyinceye kadar devam eder (Kanungo ve diğerleri, 2002: 881-892).

K-ortalamlar algoritmasının akış şeması Şekil 20’de verilmiştir.

Şekil 20: K-ortalamlar algoritması akış şeması



K-ortalamlar algoritmasının gerçekleştirilmesi Şekil 20’deki 5 adımı takip ederek elde edilir (Ray ve diğerleri, 1999: 137-143).

1. Önceden belirlenmiş olan küme sayısı olan k değeri seçilir.
2. Veri kümesinden k adet gözlem rastgele seçilir ve küme merkezleri olarak kabul edilir.
3. Kalan gözlemlerin en yakın olduğu küme merkezleri belirlenir ve ilgili gözlem o kümeye atanır.
4. Her küme için atanan gözlemler ile küme merkezi hesaplanır.
5. Eğer kümelerin yeni merkez noktaları bir önceki merkez noktaları ile aynı ise kümeleme tamamlanır. Değil ise hesaplanan yeni küme merkezleri ile 3. adımdan itibaren işlemler tekrarlanır.

4.1.2.2. En Çok Olabilirlik Yöntemi

Kümeleme analizinin dışında Diskriminant analizinde de kullanılan bir yöntemdir. Her bir x_i gözlem vektörünün, birbirinden bağımsız ve aynı dağılıma sahip olduğunu varsayan En Çok Olabilir Yönteminde, her bir gözlem olabilirlik değerini en büyük yapacak biçimde daha önceden belirlenen kümelere atanmaktadır. İşlemsel olarak kolaylık sağlamaması nedeni ile kuramsal olarak güçlü bir yöntem olmakla birlikte uygulamada çok tercih edilmemektedir (Tatlıdil, 2002: 339).

4.1.2.3. İki Aşamalı Kümeleme Yöntemi

İki aşamalı kümeleme analizi, küme sayısının önceden belirlenmediği veya bilinmediği durumlarda kullanılabilmektedir. İki aşamalı teknik hem kategorik ölçme düzeyindeki veriler için hem de sürekli veriler için kümeleme analizini uygulamayı mümkün kılmaktadır. Bununla birlikte geleneksel kümeleme yöntemlerinin büyük veri setinde uygulanmasında karşılaşılan sorunları da iki aşamalı teknik sayesinde düzeltebilmek mümkün olmuştur. Geleneksel yöntemler, büyük verilerin ancak küçük gruplara bölünerek analiz edilmesi ile teorik altyapıya uygun sonuçlar verebilmektedir. Bunu yanısıra iki aşamalı teknik, veri setinin büyüklüğünden etkilenmeden, tüm gözlemlerin aynı anda analizini mümkün kılmaktadır (Giray, 2016: 16).

Küme Sayısının Belirlenmesi

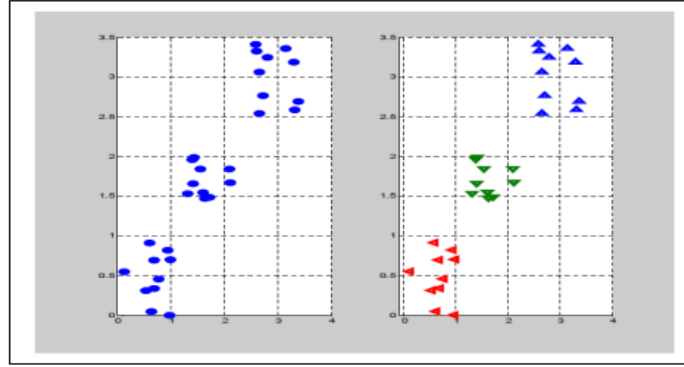
Kümeleme analizinin en önemli aşamalarından biri uygun küme sayısının belirlenmesidir. Bazen aykırı değerlerden dolayı bir kümeye tek bir gözlem biriminin atanması gibi sorunlar ortaya çıkabilmektedir. Bu durumda, bir kümede en az kaç birim olacağı, tek birimden oluşan kümelerin kabul edilip edilmeyeceği, eğer edilecekse nasıl yorumlanacağı küme sayısının belirlenmesi açısından önem taşımaktadır (Bülbül, 2001: 60).

Hiyerarşik kümeleme yöntemlerinde küme sayısı bilinmemekte ve analiz sonuçlarına göre küme sayısına karar verilmektedir. Hiyerarşik olmayan kümeleme yöntemlerinde ise küme sayısı araştırmacı tarafından analizin başında belirlenmektedir. Veri setinin yapısına göre n birimlik bir seride küme sayısı 1 ile n arasında herhangi bir değer alabilir. Kümeleme yöntemlerinde küme sayısının kaç olması gerektiği konusunda net bir teknik olmamakla birlikte, araştırmacının deneyimi ve araştırma konusu ile alakalı önsel bilgisi genellikle en iyi yaklaşım olarak kabul görmektedir (Alpar, 2011: 312).

4.2. KÜMELEME ANALİZİ SÜRECİ

Kümeleme analizi; birkaç adımı takip eden bir çözüm sürecidir. Analizin ilk aşaması veri girişidir. İlk olarak doğal sınıflamaları hakkında net bilgilerin elde edilemediği ana kütlelerden alınan n sayıda birimin incelenen p sayıda değişkene ait gözlem sonucu değerleri elde edilir. Böylece veri matrisi oluşmuş olur. Sonrasında verinin ölçüm tipine uygun bir benzerlik ölçüsü ile nesnelerin uzaklıklar matrisi elde edilir. Uygun kümeleme tekniği seçilip uygulanır. Tekniğin uygulanması sonucunda nesneler kümelere ayrılmış olur. Kümeleme sonuçlarının anlamlılığının yorumlanması aşaması ise analizin son aşamasıdır. Kümeleme işlemi başarılı olduysa, bir geometrik çizim yapıldığı zaman birimler küme içerisinde birbirlerine oldukça yakın, kümeler ise birbirlerinden hayli uzak olacaktır. Şekil 21’de basit bir kümeleme işlemi verilmiştir (Vatansever, 2008: 84) Kümeleme analizi sürecinin aşamaları problemin tanımlanmasından küme geçerliliğinin incelenmesine kadar Şekil 22’de detaylandırılmıştır.

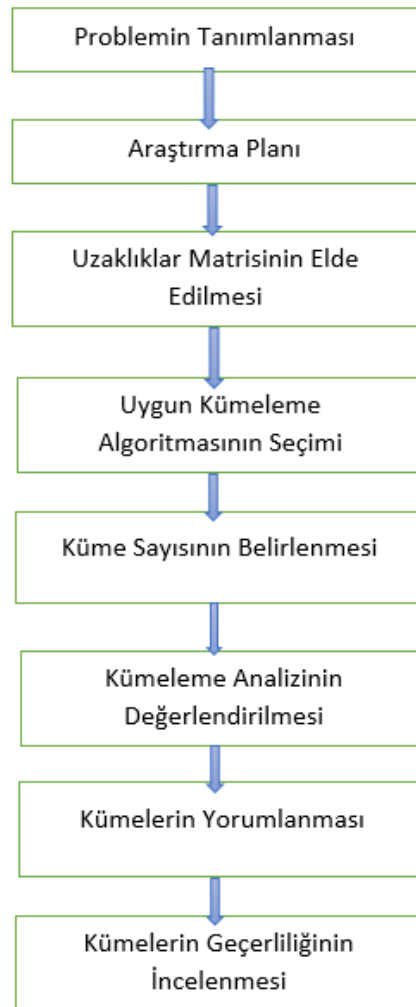
Şekil 21: Veri birimleri ve kümeleme



Kaynak: Vatansever, 2008: 84.

Kümeleme analizinin aşamaları aşağıdaki gibi sıralanabilir:

Şekil 22: Kümeleme analizi aşamaları



Küme Geçerliliğinin Ölçülmesi (Cluster Validation)

Kümeleme algoritmalarından sonuçlar elde edildikten sonra, kümeleme yapılarının uygunluğunun doğruluk derecesi belirlenmelidir. Uygun küme sayısının belirlenmesi konusunda kullanılan yöntemler eskidir ve çok güvenilir değildir. Birçok algoritma, küme sayısının önceden bilinmesi durumunda sağlıklı sonuçlar vermektedir. Ancak uygulamalarda karşılaşılan küme sayısı hakkındaki önsel bilgi eksikliği, küme sayısının gerçekte mevcut olan küme sayısı ile karşılaştırılmasına olanak vermemektedir. Optimal (en uygun) küme sayısının belirlenmesi için yapılan işlemler küme geçerliliği (cluster validity) olarak adlandırılmaktadır. Bu işlemler, algoritmalar sonucunda oluşturulan kümelerin uygunluğunu değerlendirmeye imkan sağlamaktadır. Verilerin iki boyutlu uzayda olduğu durumlarda grafiksel yöntemlerden yararlanılırken, çok boyutlu uzayda aynı yöntem yetersiz kalmaktadır (Höppner ve diğerleri, 1999: 185).

Kümelerin geçerliliğini sınamak için;

1. Kümeler için hipotez testi,
2. Model seçim tekniği,
3. Küme geçerlilik indekslerinin kullanımı olmak üzere üç yöntem mevcuttur.

İlk iki yöntem daha çok matematiksel istatistik alanı içinde yer almakta ve olasılık dağılımı varsayımına dayanmaktadır. Kümelerin uygululuğunu da olasılık dağılımı çerçevesinde incelemektedir. Hipotez testi yönteminde tek bir model varsayımı varken; model seçim tekniğinde en iyi modelin yer aldığı olasılık dağılım ailesi varsayımı vardır. Ancak bulanık kümeleme yönteminde ilk iki yöntem faydalı sonuçlar çıkarmada etkili değildir. Bu nedenle, küme geçerlilik indekslerinin kullanımı uygun olmaktadır (Miyamoto ve diğerleri, 2008: 108).

4.3. KÜMELEME ANALİZİNİN TIP ALANINDA KULLANIMINA ÖRNEKLER

Birçok alanda olduğu gibi kümeleme analizinin tıp alanında da yoğun bir biçimde kullanıldığı görülmektedir. Özellikle son yıllarda klinik olaylardaki artış ve karmaşık vakaların incelenmesine bağlı bir biçimde etkin ve yorumlanması daha kolay bir yöntem olan kümeleme analizine çoğunlukla başvurulmaktadır. Tıp alanında bu analizin kullanıldığı çalışmalara ilişkin örnekler aşağıdaki gibi özetlenebilir.

Kanser hastalığına yakalanmış bireylerin bu hastalıkla nasıl mücadele ettikleri araştırılmıştır. Kanserle mücadele çok boyutlu olduğundan, konuyu daha net araştırmak sebebiyle kümeleme analizinden faydalanılmıştır. Kanserle mücadele yolları yapılan kümeleme analizi sonucunda 4 farklı gruba ayrılmıştır. Kümeleme analizi, kanser tedavisi sırasında ve sonrasında hastanın yaşam kalitesine etkilerini daha kolay incelemek için gruplar elde edilmesini sağlamıştır (Shapiro ve diğerleri, 1994: 151-159).

Kalp seslerinin farklı morfolojilerini belirleyip bunların fizyolojik durumlarının sınıflandırılması için kümeleme analizi yöntemlerinden hiyerarşik yöntem kullanılmıştır (Amit ve diğerleri, 2009: 26-36).

4.4. MÜŞTERİ SEGMENTASYONU

Müşteri segmentasyonu, müşterilerin talep ve istekleri ile davranışları esas alınarak pazarın benzer özellikler taşıyan müşteri gruplarına ayrılmasıdır. Müşteri segmentasyonu ile firmalar geniş ve heterojen pazarları daha küçük gruplara ayırıp, müşterilerinin bireysel istek ve ihtiyaçlarına yanıt veren ürün veya hizmet sunarak müşteri grupları bazında ürün önerileri tespit eder ve pazarlama programları geliştirebilirler (Hwang ve diğerleri, 2004: 181-188).

Müşteri segmentasyonu, işletme açısından müşterilerin değerlerini grupları itibarıyla ortaya çıkararak, pazarlama faaliyetlerinin uygulamasında en değerli müşterilere öncelikli hizmet verilmesini sağlar ve böylece mevcut müşterilerden elde edilecek geliri maksimize eder. Ayrıca, doğru gruba doğru mesajın gönderilmesini sağlayarak da kampanya yönetim maliyetlerini azaltır (Kotler, 1976: 141-158).

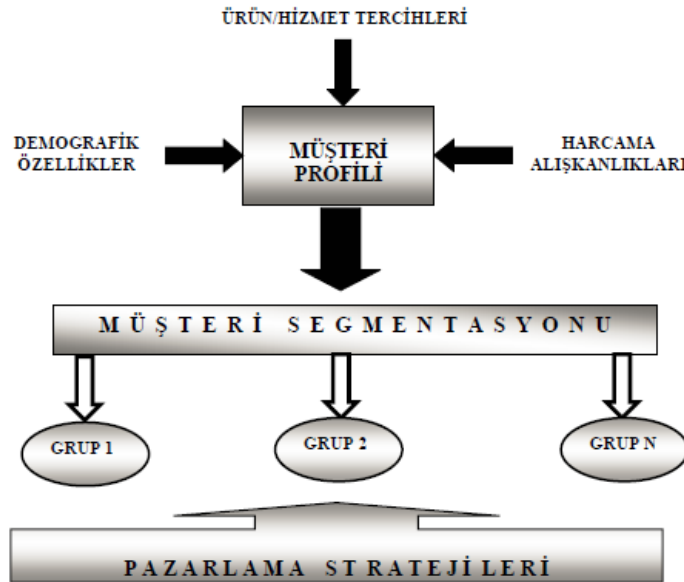
Müşteriler; demografik, psikografik ve davranışa dayalı özelliklerine göre ayrıştırılabilirler (Kotler ve Armstrong, 2001: 245-256).

Demografik segmentasyon; müşterileri cinsiyet, yaş, aile büyüklükleri, gelir, meslek, öğrenim durumu gibi kriterleri temel alarak gruplandırır.

Psikografik segmentasyon; müşterileri yaşam tarzı, sosyal sınıf ve kişilik özelliklerine dayalı olarak gruplandırır.

Davranışa dayalı segmentasyon ise müşterinin ürünü kullanma, kendisine sunulan öneriye cevap verme, satın alma miktarı ile satın alma sıklığı gibi kriterlere dayalı olarak gruplandırmaktadır.

Şekil 23: Müşteri segmentasyonu



Kaynak: Akbulut, 2006: 32.

Şekil 23’te görülebileceği gibi müşteri segmentasyonunda ilk aşama müşteri profilinin oluşturulmasıdır. Müşteri profili belirleme aşamasında, çoğunlukla müşterilerle ilgili aşağıda verilen sorulara yanıt aranmaktadır (Akbulut, 2006: 32):

- Temel demografik özellikleri
- Müşterinin harcama alışkanlıkları
- Satın aldıkları ürün veya hizmetler
- Müşterinin alışveriş sıklığı
- Müşterinin belirli ürün veya hizmetleri tercih sebepleri



BEŞİNCİ BÖLÜM

UYGULAMA

5.1. UYGULAMANIN AMACI

Çalışmanın amacı, 50 yılı aşkın bir süredir ilaç sektöründe faaliyet gösteren Türkiye ilaç sektörünün lider firmaları arasında yer alan uluslararası bir ilaç firmasının dijital pazarlama faaliyetlerinde kullandığı iletişim kanallarının yaklaşık 24 aylık verileri kullanılarak müşterilerinin “stratejik”, “orta”, “düşük” düzeyde dijital yatkınlıkları açısından segmentasyonunu belirlemektir. Veri setinin büyüklüğü sebebiyle veri madenciliğinden yararlanılacaktır.

Çalışma sonucunda ortaya çıkan segmentasyon baz alınarak hekimlerin dijital yatkınlığına göre firmanın dijital pazarlama stratejilerini uygulayabilmesi hedefi yorumlanacaktır.

Çalışmada söz konusu veri setinin bilgisayar yardımıyla analizi için SPSS Clementine paket programından faydalanılmıştır.

5.2. VERİ SETİNİN TANITILMASI

Çalışmada, İlaç sektöründe faaliyet gösteren uluslararası bir ilaç firmasının, ülke genelinde çalıştığı hekimlerin dijital pazarlama kanallarını kullanımlarının yer aldığı veri tabanından elde edilen yaklaşık 24 (Mart 2016-Eylül 2018) aylık bir veri seti üzerinden uygulanmış ve veri madenciliğinin kümeleme analizi algoritmalarından k-ortalamlar algoritması ve iki aşamalı algoritma ile çalışılmıştır. K-ortalamlar algoritmasının seçilmesinin sebebi segmentasyon sonucunda elde edilmek istenen küme sayısının önceden belirlenmiş olmasıdır. Bu çalışma sonrasında stratejik”, “orta”, “düşük” düzeyde dijital yatkınlık açısından 3 küme elde edilmesi hedeflenmektedir. Ayrıca çalışmanın etkinliğini ölçmek ve küme sayısının validasyonu için uygulama iki aşamalı yöntemler algoritması da uygulanacak ve SPSS Clementine paket program aracılığıyla küme sayısı belirlenmeden de tekrarlanacaktır.

Temin edilen veri seti üzerinde etkin bir çalışma yapabilmek bu verileri iyi bir şekilde tanımakla mümkündür. Veri setinin özelliklerinin iyi bir şekilde kavranması ve yapılacak işlemlerin de bu özellikler dikkate alınarak yapılması gerekmektedir. Bu aşamada veri setinin yorumlanmasında sektör uzmanlarının önemli bir rolü bulunmaktadır.

Hangi hekim segmentine hangi bilimsel hizmetlerin sunulması, hangi tanıtım faaliyetleriyle ulaşılması gerektiği ilaç sektöründe önemli bir konu olarak ortaya çıkmaktadır. Hekimleri bölmek için datası elde edilebilecek ve sayısallaştırılabilecek değişkenler belirlenmiştir. Bu aşamada firma yetkililerinden 4 üst düzey müşteri ilişkileri yöneticisi ve veri analistinin sektör tecrübesinden faydalanılmıştır. Hekimlerin dijital yatkınlıklarının ölçümü hedeflendiği için firmanın kullandığı Pazarlama kanalları içerisinden seçim yapılarak, hekim segmentasyonu için kullanılacak pazarlama kanalları Tablo 5’deki gibi belirlenmiştir.

Tablo 5: Veri setinde kullanılan değişkenler (Dijital pazarlama kanalları)

Mass E-Mail Opened/Sent
Mass E-Mail Clicked/Sent
Webinar Attendance
CoBrowse
1:1 Email Opened/Sent
1:1: Email Clicked/Sent
Portal Visits

Veri setinde kullanılan pazarlama kanallarına ait özellikler Tablo 6’da açıklanmıştır.

Tablo 6: Veri setinde kullanılan değişkenlere ait özellikler

Değişkenler		Açıklama
Mass E-Mail Opened/Sent	Toplu e-posta gönderimi	Merkez ofisten tüm hekimlere aynı şekilde gönderilen e-postaların açılma oranı
Mass E-Mail Clicked/Sent	Toplu e-posta açılımı	Merkez ofisten tüm hekimlere aynı şekilde gönderilen e-postaların içerisindeki linkin tıklanma oranı
Webinar Attendance	Canlı yayın katılımı	Yıl içerisindeki çeşitli konularda ve çeşitli uzman hekimler ile düzenlenen canlı yayınlara katılım/izlenme oranı
CoBrowse	Uzaktan tanıtım	Uzaktan tanıtım
1:1 Email Opened/Sent	Birebir e-posta gönderimi	Hekim bazlı gönderilen e-posta açılma oranı
1:1: Email Clicked/Sent	Birebir e-posta açılımı	Hekim bazlı gönderilen e-postadaki linkin tıklanma oranı
Portal Visits	İnternet sitesi	Hekimin firma internet sitesini kendi kullanıcı adı ve şifresi ile bağlanarak ziyaret etmesi

Uygulamada kullanılacak olan veri seti uluslararası bir firmanın müşterilerine ait olduğundan hem müşterilerin kişisel bilgileri hem de firmanın ticari bir değeri olması sebebiyle gizlilik içermektedir. Dolayısıyla çalışmada firma ve müşterilerine ilişkin demografik veri veya kişisel bir bilgi kullanılmamış olup yalnızca pazarlama kanallarına erişim sayıları kullanılmıştır.

İlk olarak veri setinde yer alan tüm pazarlama kanalları içerisinde dijital pazarlama tekniklerinin kullanıldığı Tablo 5’te yer alan 7 pazarlama kanalı seçilmiş olup bu kanallara

ilişkin erişim verileri kullanılmıştır. Toplam 47.447 hekimin verilerini içeren veri seti incelenerek tüm kanallarda erişim sayıları sıfır olan gözlemler uygulamadan çıkarılarak 29.171 hekimin dijital pazarlama kanallarına erişim verisi üzerinden uygulama yapılmıştır.

Veri setinin ağırlıklandırması konusunda en uygun metodun seçilebilmesi için aşağıda açıklanacak yöntemler denenmiştir. Uygulamada hem ham veri seti üzerinden hem de ağırlıklandırılmış veri seti ile modellemeler yapılmıştır. Ağırlıklandırma için iki yöntem izlenmiş olup ilk olarak firma yetkililerinin sektöre yönelik bilgi ve tecrübelerine dayanarak oluşturdukları ağırlıklandırma elde edilmiştir. Aynı zamanda firma yetkililerinden 4 üst düzey müşteri ilişkileri yöneticisi ve veri analistinin yönlendirmeleri doğrultusunda AHP yöntemi ile kriterlerin ağırlıklandırılması için ikili karar matrisi oluşturularak Expert Choice programı ile ağırlıklandırma yapılmıştır. Uygulamada kullanılan tüm modeller ilerleyen kısımlarda açıklanmıştır.

Veri setindeki gözlemler aynı türde numerik veri olsalar da değer aralıkları olarak birbirlerinden farklıdır. Bir gözlemin toplam gerçekleştirilme sayısı diğer bir gözlemden farklılık göstermektedir. Örneğin toplu e-posta gönderimi merkez ofisten tüm hekimlere toplu bir şekilde aynı zamanlarda yapılırken birebir e-posta gönderimi satış temsilcisinin görüşüne ve iş planına göre hekim bazlı olarak gönderilmektedir. Yıl içerisinde düzenlenen canlı yayın sayısı da diğer gözlemler ile karşılaştırılabilecek bir gözlem değildir. Şirket internet Sitesi ziyaretleri ise tamamen hekimlerin kendi isteğine bağlı olarak gerçekleştirdikleri bir davranıştır. Çalışmada kullanılan dijital pazarlama kanallarını aktif olan tarafa göre Şekil 24'deki şekilde ayırabiliriz.

Şekil 24: Dijital pazarlama kanallarının genel olarak yorumlanması

Firma Aktif	Hekim Aktif	Karşılıklı
• Toplu e-posta gönderimi • Birebir e-posta gönderimi	• Toplu e-posta açılma/tıklanma • Birebir e-posta açılma/tıklanma • İnternet sitesi ziyareti • Canlı yayın katılımı	• Uzaktan tanıtım

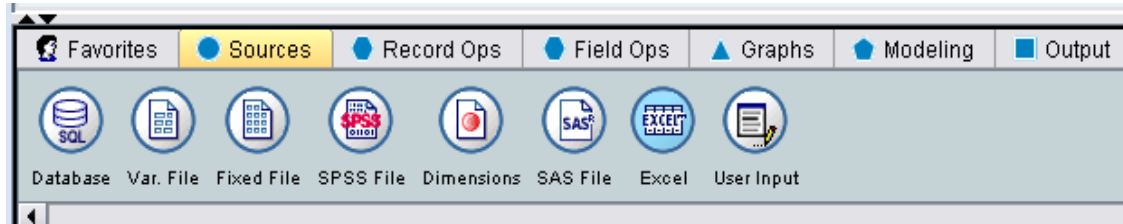
Veri setindeki değişkenlerin aldığı değer aralıklarının birbirinden farklı olduğu bu durumda normalizasyon yöntemlerinden yararlanma kararı alınmıştır. Çünkü değişkenlerin ortalama ve varyanslarının birbirlerinden önemli ölçüde farklı olması durumu büyük ortalama ve varyansa sahip değişkenlerin diğer değişkenler üzerindeki baskısını artırarak doğruluk ve performansı etkilediği izlenmiştir. Min-max ve z-skor normalizasyon teknikleri veri setine uygulanmış olup son aşamada yorumlama kolaylığı sağlaması açısından Excel yardımıyla Min-Maks normalizasyonu uygulanarak verilerin (0,1) aralığında olması sağlanmıştır.

Firma yetkililerinden 4 üst düzey müşteri ilişkileri yöneticisi ve veri analisti ile yapılan görüşmeler sonrasında segmentasyon işlemi yapılırken küme sayısı 3 seçilerek atama işlemleri yapılması planlanmıştır. Ayrıca küme sayısı modele bırakılarak da kümeleme işlemi uygulanmıştır. Bu kümeler hekimlerin dijital pazarlamaya yatkınlıklarına göre “stratejik”, “orta”, “düşük” olarak değerlendirilmiştir.

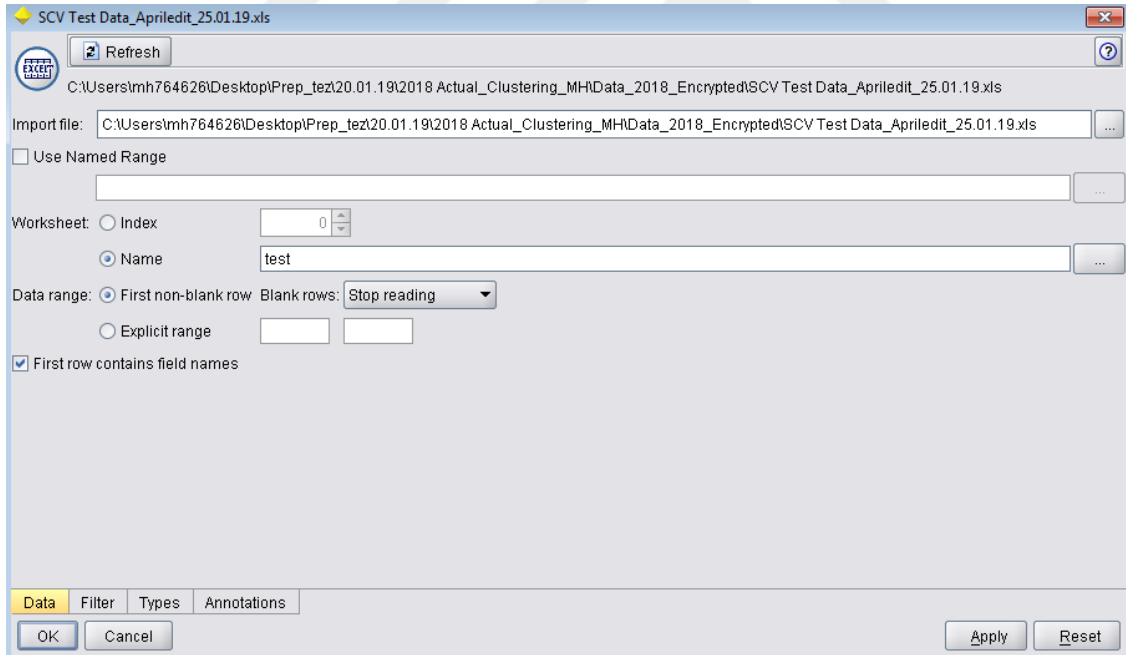
5.3. VERİ SETİNİN HAZIRLANMASI VE UYGULAMA AŞAMASI

Veri'nin SPSS Clementine programına aktarımı Şekil 25'te gösterilmiştir. “Sources” dan “Excel” seçilmiş, sonrasında Şekil 26'da gösterilen ekranda çıkan tablodan “Import File” seçilerek kullanılacak veri seti programa yüklenmiştir.

Şekil 25: SPSS Clementine programına veri'nin excel yardımıyla yüklenmesi

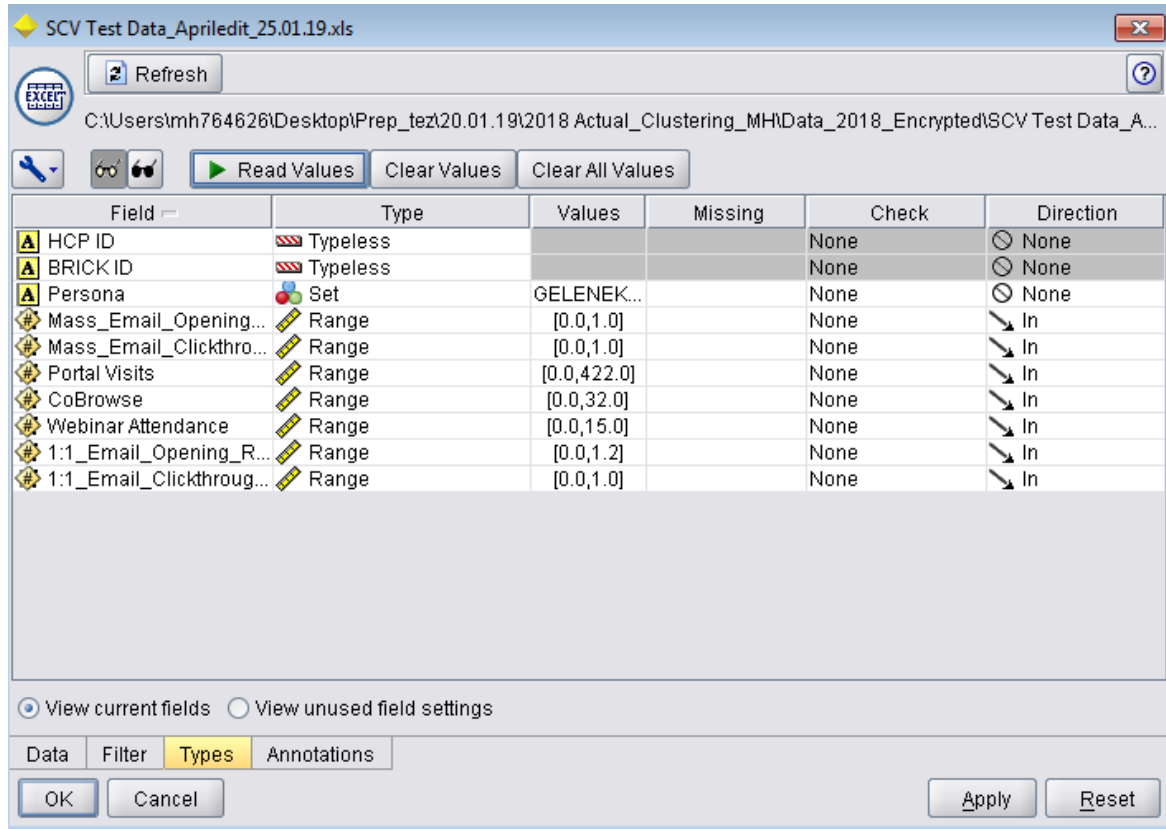


Şekil 26: SPSS Clementine programına veri setinin yüklenmesi



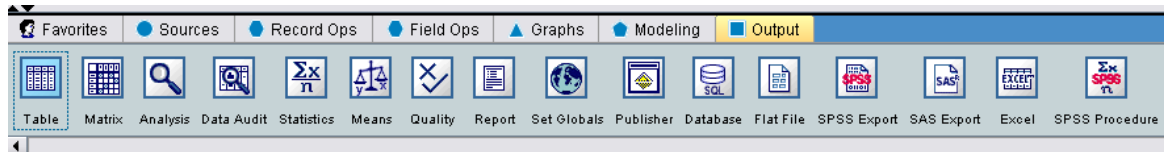
Bu aşamada değişkenler SPSS Clementine programına aktarılırken “Types” sekmesinden veri türleri “Range” olarak seçilmiştir. Seçilen veri türleri Şekil 27’de gösterilmiştir.

Şekil 27: “Types” sekmesinde veri türleri

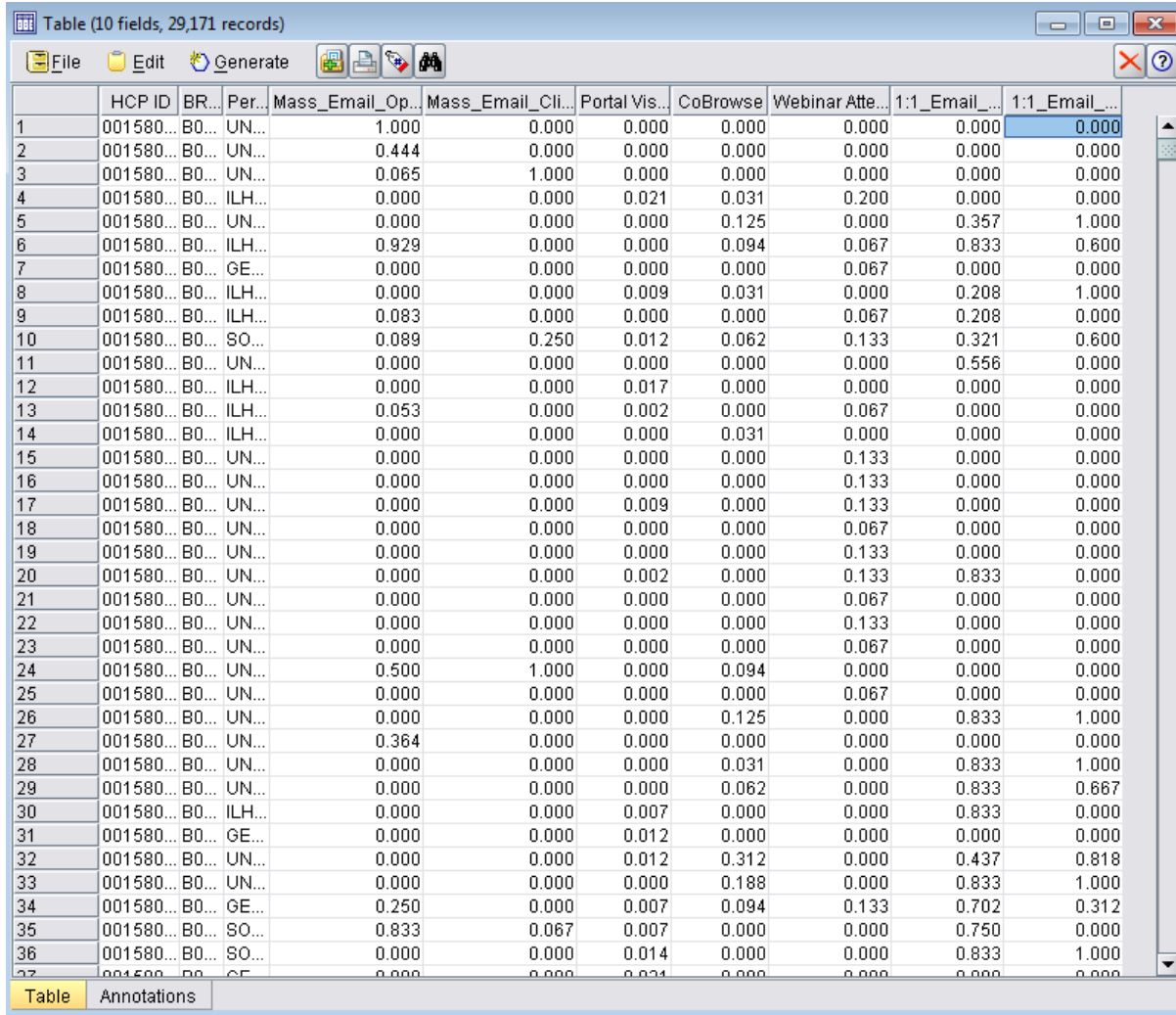


Programa aktarılan veri setinin kontrolü için “Table” düğümü yürütülerek yapılmıştır. Şekil 28’deki gibi Veri setine erişim sağlanarak kontrolü tamamlanır. Şekil 29’da SPSS Clementine programına yüklenen veri setinden bir kesit görülmektedir. Satırlarda yer alan her bir kayıt, bir hekimin ilgili sütunlardaki davranışlara karşılık gelen değerlerinin birleşimini ifade etmektedir. Sütunlarda yer alan her bir davranış ise tüm hekimlerin aynı ölçüme karşılık gelen değerlerinin topluluğunu ifade etmektedir.

Şekil 28: Veri Setinin kontrolü



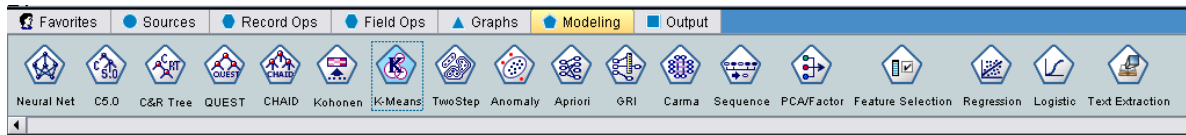
Şekil 29: Veri setinin görünümü



	HCP ID	BR...	Per...	Mass_Email_Op...	Mass_Email_Cli...	Portal Vis...	CoBrowse	Webinar Atte...	1:1_Email_...	1:1_Email_...
1	001580...	BO...	UN...	1.000	0.000	0.000	0.000	0.000	0.000	0.000
2	001580...	BO...	UN...	0.444	0.000	0.000	0.000	0.000	0.000	0.000
3	001580...	BO...	UN...	0.065	1.000	0.000	0.000	0.000	0.000	0.000
4	001580...	BO...	ILH...	0.000	0.000	0.021	0.031	0.200	0.000	0.000
5	001580...	BO...	UN...	0.000	0.000	0.000	0.125	0.000	0.357	1.000
6	001580...	BO...	ILH...	0.929	0.000	0.000	0.094	0.067	0.833	0.600
7	001580...	BO...	GE...	0.000	0.000	0.000	0.000	0.067	0.000	0.000
8	001580...	BO...	ILH...	0.000	0.000	0.009	0.031	0.000	0.208	1.000
9	001580...	BO...	ILH...	0.083	0.000	0.000	0.000	0.067	0.208	0.000
10	001580...	BO...	SO...	0.089	0.250	0.012	0.062	0.133	0.321	0.600
11	001580...	BO...	UN...	0.000	0.000	0.000	0.000	0.000	0.556	0.000
12	001580...	BO...	ILH...	0.000	0.000	0.017	0.000	0.000	0.000	0.000
13	001580...	BO...	ILH...	0.053	0.000	0.002	0.000	0.067	0.000	0.000
14	001580...	BO...	ILH...	0.000	0.000	0.000	0.031	0.000	0.000	0.000
15	001580...	BO...	UN...	0.000	0.000	0.000	0.000	0.133	0.000	0.000
16	001580...	BO...	UN...	0.000	0.000	0.000	0.000	0.133	0.000	0.000
17	001580...	BO...	UN...	0.000	0.000	0.009	0.000	0.133	0.000	0.000
18	001580...	BO...	UN...	0.000	0.000	0.000	0.000	0.067	0.000	0.000
19	001580...	BO...	UN...	0.000	0.000	0.000	0.000	0.133	0.000	0.000
20	001580...	BO...	UN...	0.000	0.000	0.002	0.000	0.133	0.833	0.000
21	001580...	BO...	UN...	0.000	0.000	0.000	0.000	0.067	0.000	0.000
22	001580...	BO...	UN...	0.000	0.000	0.000	0.000	0.133	0.000	0.000
23	001580...	BO...	UN...	0.000	0.000	0.000	0.000	0.067	0.000	0.000
24	001580...	BO...	UN...	0.500	1.000	0.000	0.094	0.000	0.000	0.000
25	001580...	BO...	UN...	0.000	0.000	0.000	0.000	0.067	0.000	0.000
26	001580...	BO...	UN...	0.000	0.000	0.000	0.125	0.000	0.833	1.000
27	001580...	BO...	UN...	0.364	0.000	0.000	0.000	0.000	0.000	0.000
28	001580...	BO...	UN...	0.000	0.000	0.000	0.031	0.000	0.833	1.000
29	001580...	BO...	UN...	0.000	0.000	0.000	0.062	0.000	0.833	0.667
30	001580...	BO...	ILH...	0.000	0.000	0.007	0.000	0.000	0.833	0.000
31	001580...	BO...	GE...	0.000	0.000	0.012	0.000	0.000	0.000	0.000
32	001580...	BO...	UN...	0.000	0.000	0.012	0.312	0.000	0.437	0.818
33	001580...	BO...	UN...	0.000	0.000	0.000	0.188	0.000	0.833	1.000
34	001580...	BO...	GE...	0.250	0.000	0.007	0.094	0.133	0.702	0.312
35	001580...	BO...	SO...	0.833	0.067	0.007	0.000	0.000	0.750	0.000
36	001580...	BO...	SO...	0.000	0.000	0.014	0.000	0.000	0.833	1.000
37	001580...	BO...	GE...	0.000	0.000	0.031	0.000	0.000	0.000	0.000

Veri seti amaca uygun olarak düzenlendikten sonra toplam 29.171 hekim ile veri madenciliği tekniklerinden k-ortalamlar algoritması ve iki aşamalı algoritma (two steps) tekniğinden yararlanılmıştır.

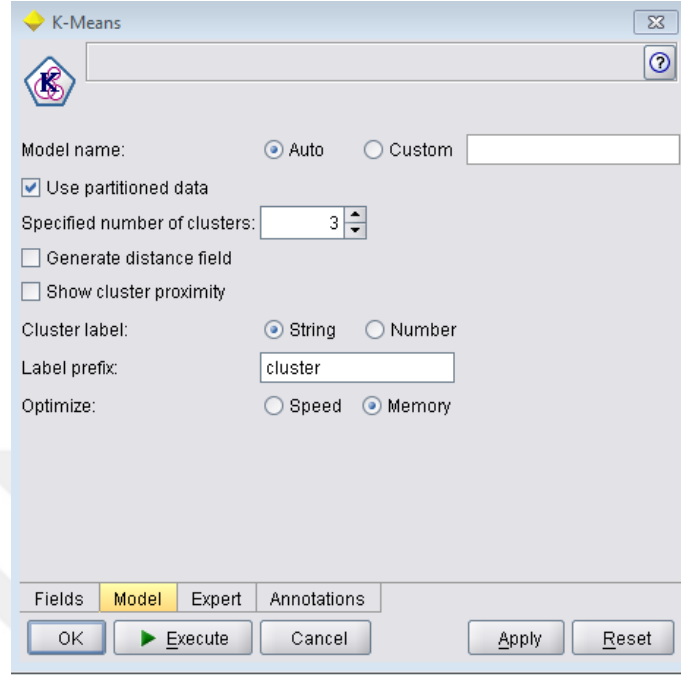
Şekil 30: Kullanılan yöntemlerin seçimi



“Modelling” sekmesi altından yöntem seçimi yapıldıktan sonra işlem yürütülerek küme sayısı, değişkenler ve diğer seçimler yapılarak devam edilir. Bu aşama k-ortalamlar algoritması için örnek olarak Şekil 30’da gösterilmiştir.

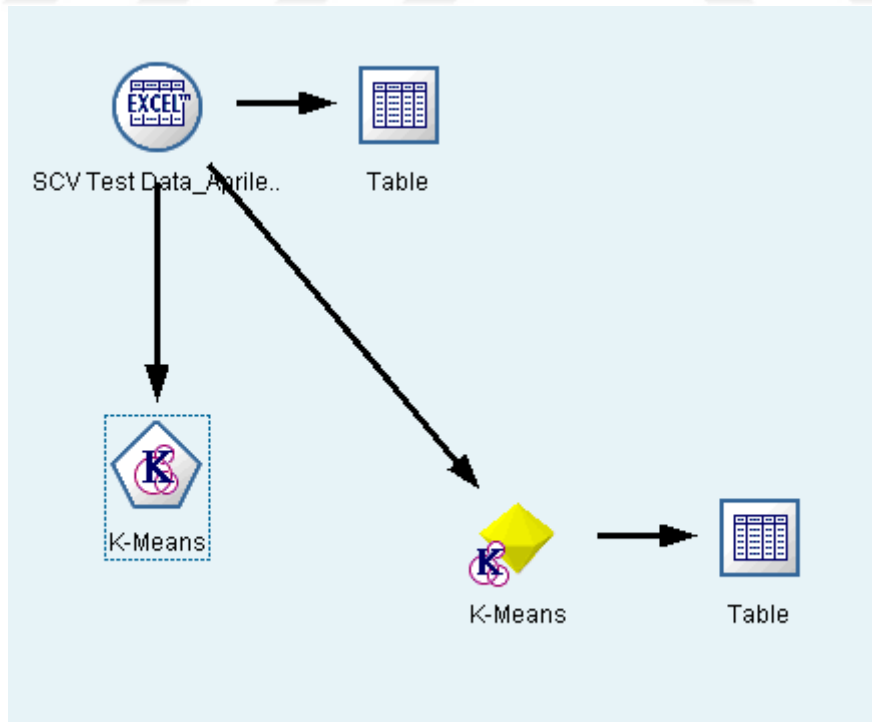
Kullanılan modellerde küme sayısı 3 seçilerek atama işlemleri yapılması planlanmıştır. Ayrıca küme sayısı modele bırakılarak da kümeleme işlemi uygulanmıştır.

Şekil 31: Modelde küme sayısı ve diğer seçimlerin yapılması



Kullanılan modelin diyagramı ise Şekil 32’de görülebilir.

Şekil 32: Model diyagramı



5.4. UYGULAMADA KULLANILAN MODELLER

Çalışma sonucu elde edilen bulgular Expert Choice ve SPSS Clementine paket programları yardımıyla elde edilmiş olup sonuçlar aşağıda özetlenmiştir. Çalışmada kullanılan her modelin sonucuna göre sağlanan ilerleme ile veri seti için en uygun modele erişilmeye çalışılmıştır.

Modeller aşağıdaki gibi özetlenebilir:

Model 1: Firma yetkililerinden alınan kriterler ile ağırlıklandırılmış veri ve ham veri seti kullanılarak k-ortamalar yöntemi ile küme sayısı 3 seçilerek segmentasyon yapılmıştır. Bu aşamada veri setine normalizasyon yöntemi uygulanmamıştır.

Model 2: Model 1'deki veri setinin normalize edilmesi gerekliliği anlaşılmış olup bu modelde kullanılan veri setine normalizasyon uygulanmıştır. Ayrıca firma yetkilileri ile oluşturulan İkili karar matrisinden elde edilen AHP kriterleri ile ağırlıklandırılan veri de kullanılmıştır. Bu modelde de k-ortalamlar algoritması ile devam edilmiştir.

Model 3: Model 2 sonucunda ağırlıklandırmanın bu veri setinde etkin olmadığı anlaşılmıştır. Ağırlıklandırma işlemine gerek duyulmadan normalize edilmiş veri seti ile k-ortalamlar algoritması kullanılarak küme sayısı seçilmeyerek deneme tekrarlanmıştır.

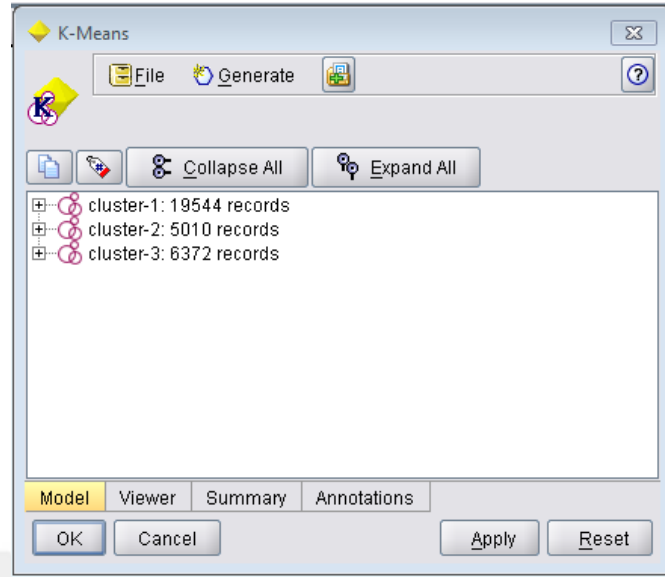
İlk modelden itibaren her modelin sonuçları dikkatlice incelenip firma yetkililerinden 4 üst düzey müşteri ilişkileri yöneticisi ve veri analisti ile de takip edilerek veri seti için uygun olacağı düşünülen çeşitli modeller üzerinde uygulamalar tekrarlanmıştır.

5.4.1. Modellere Uygulanan Kümeleme Analizi Yöntemleri ve İncelemesi

Model 1: Normalize Edilmemiş Veri Seti Üzerinden K-ortalamlar Algoritması ile Segmentasyon

Model 1'de Yaklaşık 24 aylık temin edilen ham veri seti incelenerek tüm kanallarda erişim sayıları sıfır "0" olan gözlemler uygulamadan çıkarılarak 30.926 gözlem üzerinden SPSS Clementine programı yardımıyla küme sayısı 3 olarak belirlenerek K-means algoritması ile segmentasyon işlemi yapılmıştır.

Şekil 33: Model 1 ağırlıklandırılmamış veri segmentasyon sonuçları

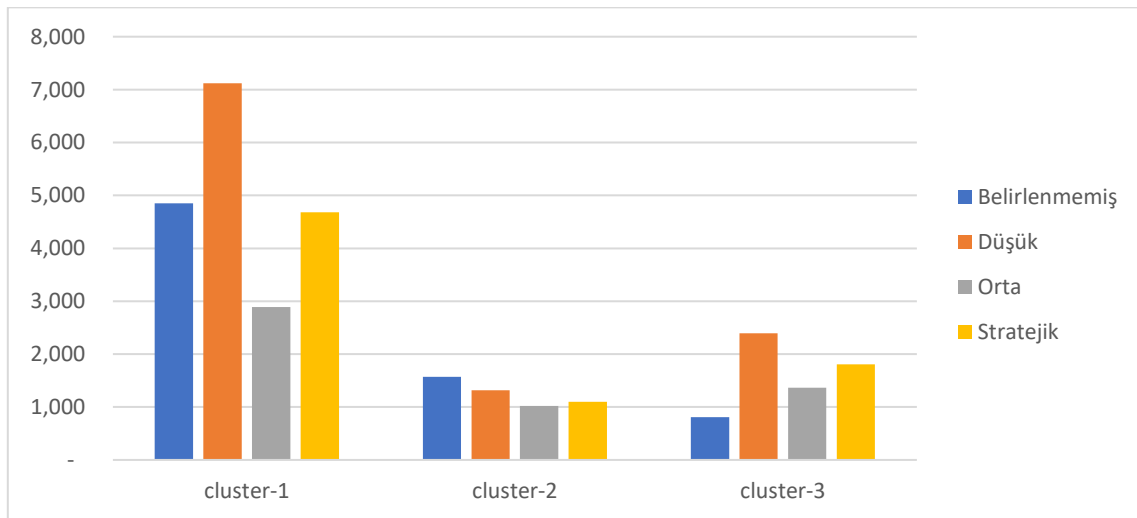


İlk olarak ağırlıklandırılmamış ve normalize edilmemiş veri seti üzerinden gerçekleştirilen uygulamada elde edilen 3 küme'ye ait sonuçlar Şekil 33'te görülebilir.

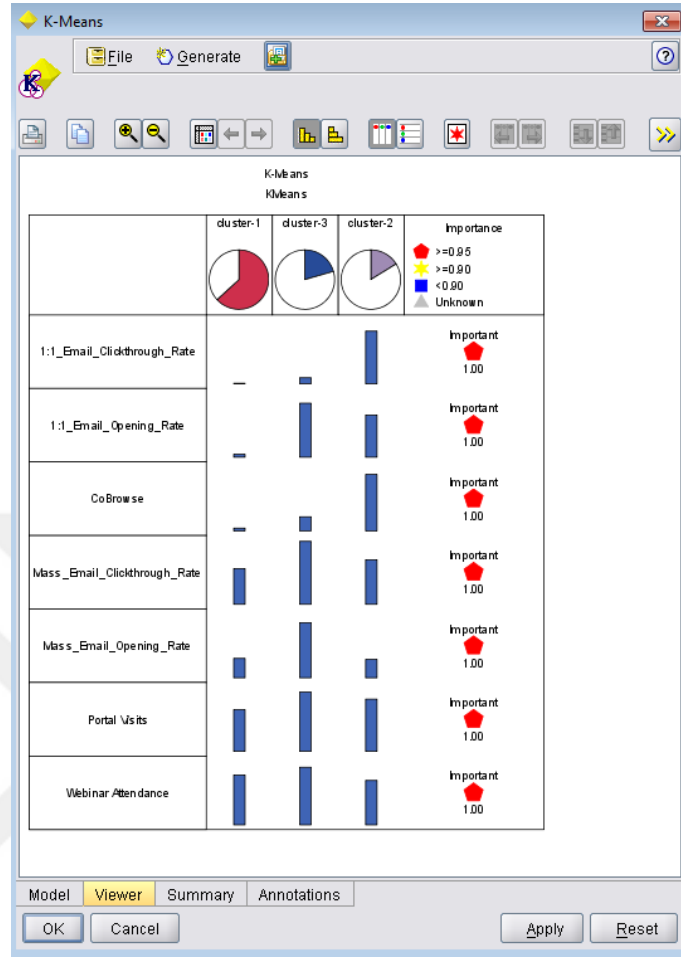
Hekimlerin %63'ü birinci kümeye (19.544), %16 ikinci kümeye, %21'i ise üçüncü kümeye atanmıştır.

SPSS Clementine aracılığıyla oluşturulan kümelerin firmanın anket yoluyla gerçekleştirdiği segmentasyon sonucu oluşan kümelere göre karşılaştırması da Şekil 34'te verilmiştir. "Cluster 1, 2 ve 3" olarak ifade edilenler uygulama sonucu elde edilen kümeleri, sütunlarda yer alan "düşük", "orta", "stratejik" ve "belirlenmemiş" olarak ifade edilenler firmaya ait kümeleri göstermektedir. 1. Küme (Cluster 1) en fazla gözlemin yer aldığı kümedir. Aynı zamanda veri setinde yer alan çoğunluk da dijital yatkınlığı düşük olan kümede yer almaktadır. Bu aşamada elde edilen sonuçlardan yapılan karşılaştırmada farklı bir sonuç gözlemlenmemiştir.

Şekil 34: Model 1 segmentasyon sonuçları karşılaştırma

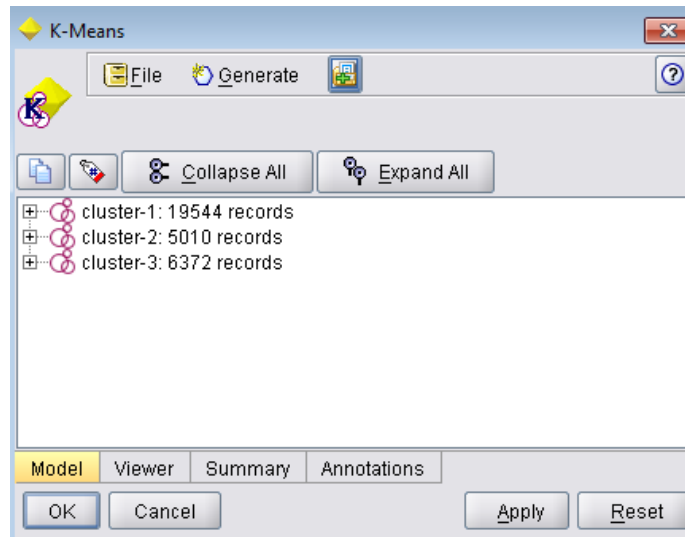


Şekil 35: SPSS Clementine programı aracılığıyla Model 1’de ağırlıklandırılmamış veri setinin sonuçları



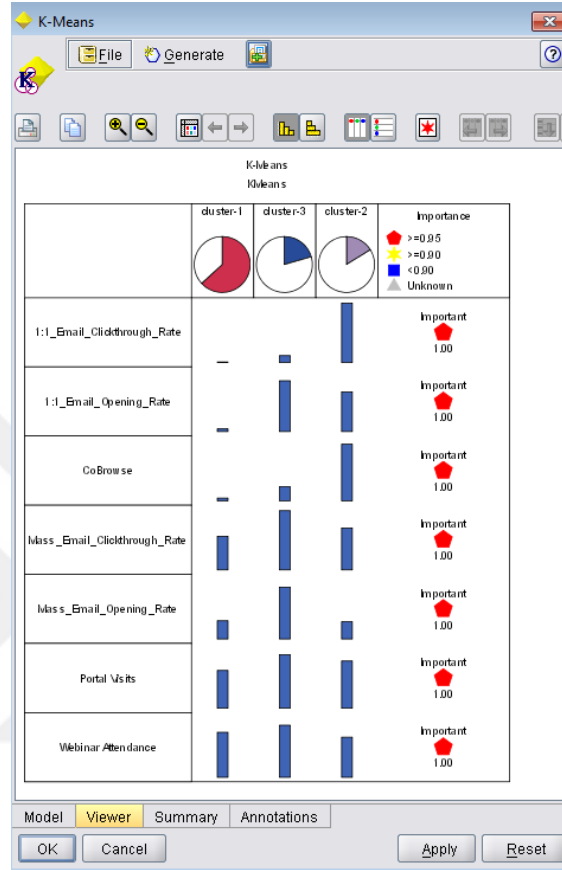
Model 1’de ikinci olarak firma yetkililerinden temin edilmiş ağırlıklar kullanılarak ağırlıklandırılmış ancak normalize edilmemiş veri seti üzerinden gerçekleştirilen uygulamada elde edilen 3 kümeye ait veriler de Şekil 36’da verilmiştir. Ağırlıklandırılmış ve ağırlıklandırılmamış veri setinden aynı kümeler elde edildiği gözlemlenmiştir.

Şekil 36: Model 1 ağırlıklandırılmış veri segmentasyon sonuçları



Model 1’de ilk gerçekleştirilen uygulama sonuçları ile aynı olarak 2. Uygulamada da Hekimlerin %63’ü birinci kümeye (19.544), %16 ikinci kümeye, %21’i ise üçüncü kümeye atanmıştır.

Şekil 37: SPSS Clementine programı aracılığıyla Model 1’de ağırlıklandırılmış veri setinin sonuçları



Model 1’in İzlenmesi, Sonuçlar ve Görüşler

Model 1 uygulaması sonucunda çıkan tabloya göre ağırlıklandırmanın veri seti üzerinde belirleyici bir etkisi olmadığı anlaşılmış olup model 2’de ayrıca AHP ile elde edilen ağırlıklar ile de uygulama tekrarlanıp ağırlıklandırmanın etkisi ölçülmeye çalışılmıştır. İletişim kanallarının yorumlamasına gelindiğinde internet sitesi ziyareti (portal visit) ve canlı yayın (webinar) Katılımı gözlemlerinin tüm data içerisinde aldıkları değerler bakımından büyük varyansa sahip olduğu anlaşılmıştır. Yukarıdaki tablo ve şekil’lerde görülebileceği gibi tüm kümelerde bu gözlemlerin baskın olup sonuç üzerinde ağırlıklı olduğu izlenmiştir.

Veri setindeki gözlemler aynı türde numerik veri olsalar da değer aralıkları olarak birbirlerinden farklıdır. Bir gözlemin toplam gerçekleştirilme sayısı diğer bir gözlemden farklılık göstermektedir. Örneğin toplu e-posta gönderimi merkez ofisten toplu bir şekilde yapılırken birebir (1:1) e-posta gönderimi satış temsilcisinin görüşüne ve iş planına göre hekim bazlı olarak gönderilmektedir. Yıl içerisinde düzenlenen canlı yayın (webinar) sayısı da diğer gözlemler ile karşılaştırılabilecek bir bilgi değildir. İnternet sitesi ziyareti ise tamamen hekimlerin kendi isteğine bağlı olarak gerçekleştirdikleri bir davranıştır.

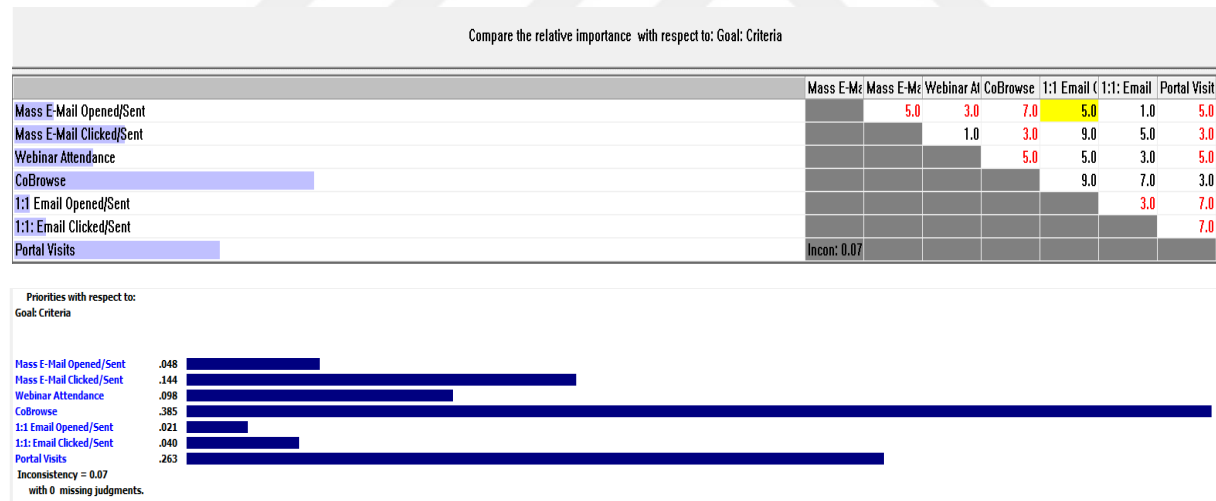
Veri setindeki değişkenlerin aldığı değer aralıklarının birbirinden farklı olduğu bu durumda normalizasyon yöntemlerinden yararlanma kararı alınmıştır. Çünkü değişkenlerin ortalama ve varyanslarının birbirlerinden önemli ölçüde farklı olması durumu büyük ortalama ve varyansa sahip değişkenlerin diğer değişkenler üzerindeki baskısını artırarak doğruluk ve performansı etkilediği izlenmiştir.

Model 2: AHP Tabanlı Ağırlıklandırma Kullanılmış ve Normalize Edilmiş Veri Setinin K-ortalamalar Algoritması ile Kümelemesi

Model 2’de elde edilen sonuçlara göre veri setinin normalize edilmesi gerekliliği ortaya çıkmıştır. Bu sebeple ilk olarak veri setine excel yardımıyla min-maks normalizasyonu uygulanmıştır. Firma yetkilileri ile oluşturulan İkili karar matrisinden elde edilen AHP kriterleri ile ağırlıklandırılan veri de kullanılarak bir kez daha ağırlıklandırmanın veri seti üzerindeki etkisi görülmeye çalışılmıştır. Bu modelde de k-ortalamalar algoritması ile devam edilmiştir.

Normalize edilen veri setinden tüm kanal etkileşimi sıfır olan gözlemler çıkarılarak 29.171 gözlem üzerinden SPSS Clementine programı yardımıyla küme sayısı 3 olarak belirlenerek K-means algoritması ile segmentasyon işlemi yapılmıştır. Firma yetkililerinden 4 üst düzey müşteri ilişkileri yöneticisi ve veri analisti ile AHP anket formu doldurularak Şekil 38’de yer alan ikili karar matrisi Expert Choice programı kullanılarak elde edilmiştir. Anket Formu EK 1’de görülebilir.

Şekil 38: AHP ikili karar matrisi ve AHP ağırlıkları

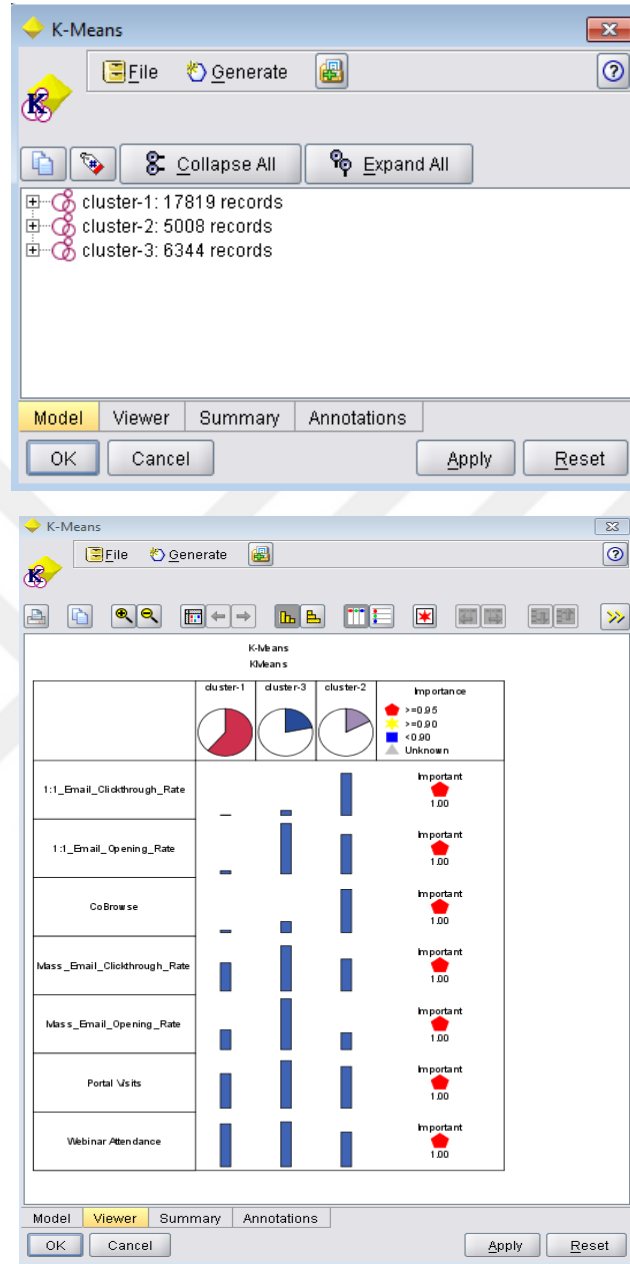


Tablo 7: AHP ağırlıkları

Değişkenler	AHP ağırlıkları
Toplu e-posta açılma oranı	0,048
Toplu e-posta tıklanma oranı	0,144
Canlı yayın katılımı	0,098
Uzaktan tanıtım	0,385
Birebir e-posta açılma oranı	0,021
Birebir e-posta tıklanma oranı	0,040
İnternet sitesi ziyareti	0,263

Expert Choice programı kullanılarak elde edilen AHP ağırlıkları ile yapılan uygulama sonuçları Şekil 39’da verilmiştir.

Şekil 39: Model 2 AHP ile ağırlıklandırılmış veri seti’nin segmentasyon sonuçları



İlk olarak normalize edilmiş ve AHP ağırlıkları kullanılmış veri seti üzerinden gerçekleştirilen uygulamada elde edilen 3 küme’ye ait sonuçlar Şekil 39’da görülebilir.

Hekimlerin %61’i birinci küme’ye (17.819), %17’si ikinci küme’ye, %22’si ise üçüncü küme’ye atanmıştır.

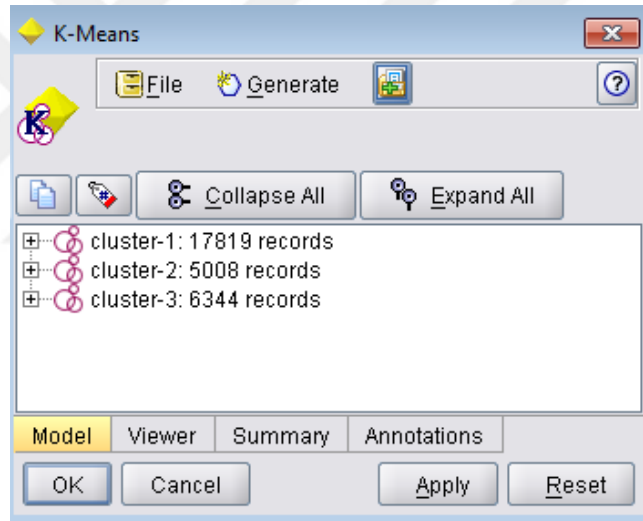
SPSS Clementine aracılığıyla oluşturulan kümelerin iletişim kanalı bazında kümelere dağılımı Şekil 40’da verilmiştir. 1. Küme (Cluster 1) en fazla gözlemin yer aldığı kümedir. Küme 1’de iletişim kanalı bazında canlı yayın (webinar) katılımı ile internet sitesi ziyaretinin ağırlıklı olduğu görülmektedir. Küme 2’de ağırlıklı olarak sanal tanıtımın tercih edilmiş olduğu göze çarpmaktadır. Küme 3 ise tüm kümelerin bir karışımı olarak karşımıza

çıkmaktadır. Bu aşamada elde edilen sonuçlar ağırlıklandırma yapılmadan tekrar denenerek çıktılar gözlemlenecektir.

Şekil 40: Model 2 AHP ile ağırlıklandırılmış veri seti'nin iletişim kanalı bazında segmentasyon sonuçları

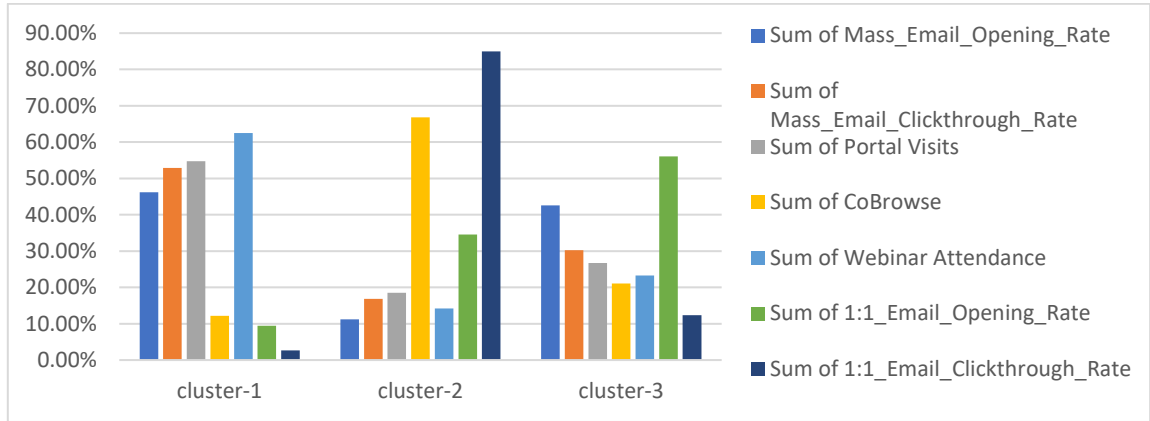


Şekil 41: Model 2 ağırlıklandırılmamış veri seti'nin segmentasyon sonuçları



Model 2 de uygulanan iki farklı yoruma göre kümelere dağılımın Şekil 39 ve Şekil 41'deki sonuçlarına bakıldığında tamamen aynı sayıda bir dağılım olduğu gözlemlenmiştir. Şekil 39' daki veri seti AHP kriterleri ile ağırlıklandırılmış iken Şekil 41'deki veri seti ağırlıklandırılmamıştır. Şekil 42'de son uygulamaya ait sonuçlar görülebilir. İlk 2 uygulama sonucunda elde edilen çıktılardan veri setinin iletişim kanalları bazında sayısal olarak aynı küme sayılarını verdiği Şekil 39 ve 41'deki program çıktılarından görülüp, her bir müşteri için tüm listeler karşılaştırılarak da kontrol edilmiştir. Şekil 39 ve 41'de 1. Küme (Cluster 1) için 17.819, 2. Küme (Cluster 2) için 5.008, 3. Küme (Cluster 3) için 6.344 aynı küme sayıları elde edilmiştir.

Şekil 42: Model 2 ağırlıklandırılmamış veri seti'nin iletişim kanalı bazında segmentasyon sonuçları



Model 2'nin İzlenmesi, Sonuçlar ve Görüşler

Model 2 uygulaması sonucunda çıkan tabloya göre ağırlıklandırmanın veri seti üzerinde belirleyici bir etkisi olmadığı teyit edilmiş olup AHP kriterleri ile ağırlıklandırılmış veri seti ve ağırlıklandırılmamış veri setinden tamamen aynı kümeler elde edildiği görülmüştür. Ortaya çıkan kümelerin iletişim kanallarının yorumlamasına gelindiğinde Küme 1'de iletişim kanalı bazında canlı yayın (webinar) katılımı ile internet sitesi ziyaretinin ağırlıklı olduğu görülmektedir. Küme 2'de ağırlıklı olarak sanal tanıtımın tercih edilmiş olduğu göze çarpmaktadır. Küme 3 ise tüm kümelerin bir karışımı olarak karşımıza çıkmaktadır. Bu aşamada elde edilen sonuçlar firma yetkililerinden 4 üst düzey müşteri ilişkileri yöneticisi ve veri analisti ile de görüşülmüş olup mantıklı sonuçlar elde edildiği anlaşılmış ve Model 3'de modele küme sayısı belirtmeden de kümeleme işlemi tekrarlanmak istenmiştir.

Veri setindeki gözlemler aynı türde numerik veri olsalar da değer aralıkları olarak birbirlerinden farklı olduklarından bahsetmiş bu sebeple veri setine normalizasyon uygulanması gerektiğinden bahsetmiştir. Model 2'de gerçekleştirilen tüm denemelerde min-max normalizasyonu ile normalize edilmiş veri seti kullanılmıştır.

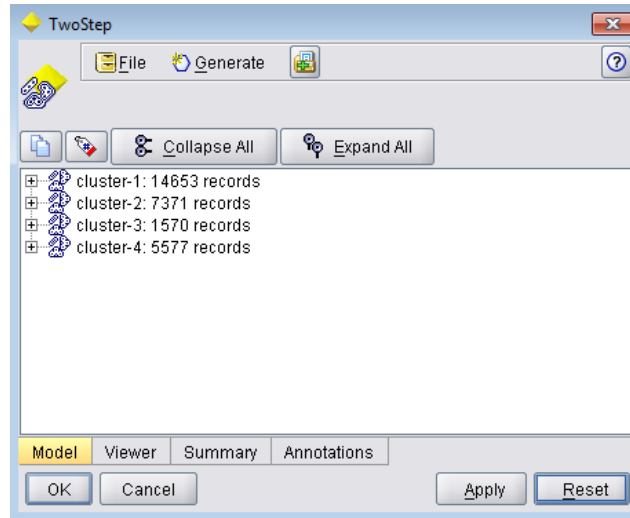
Model 3: K-ortalamlar algoritması ile değişik küme sayıları için Kümeleme

Model 2 sonucunda ağırlıklandırmanın bu veri setinde etkin olmadığı anlaşılmıştır. Ağırlıklandırma işlemine gerek duyulmadan normalize edilmiş veri seti ile k-ortalamlar algoritması kullanılarak küme sayısı seçilmeyerek deneme tekrarlanmıştır.

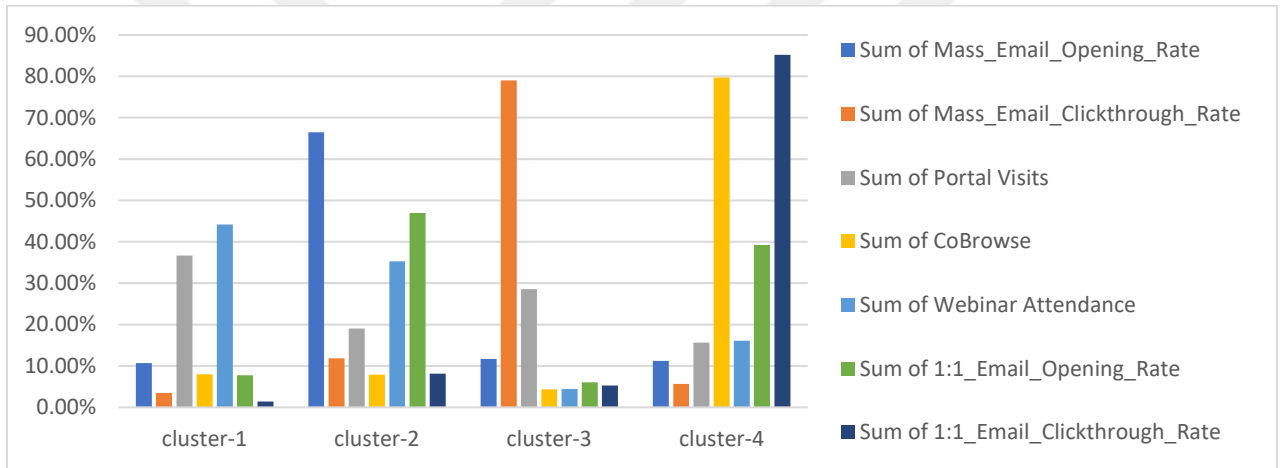
Model 3: Two-Step 4'lü Kümeleme sonuçları

SPSS Clementine programı yardımıyla seçilen iki aşamalı algoritmaya küme sayısı verilmeden normalize edilmiş veri seti ile kümeleme yapılmıştır. Bu kümeleme sonucunda elde edilen 4 küme Şekil 43'de verilmiştir.

Şekil 43: Two-step 4'lü kümeleme sonuçları



Şekil 44: Two-step 4'lü kümelemenin iletişim kanalları bazında sonuçları

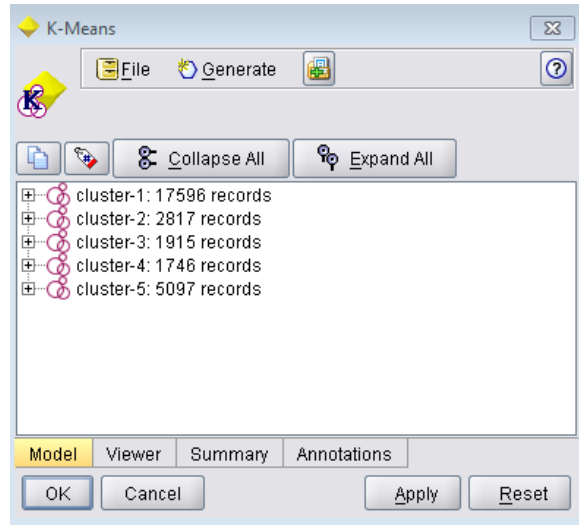


İki aşamalı algoritma ile elde edilen 4 küme iletişim kanallarına göre incelendiğinde aslında hemen hemen Model 2 de olduğu gibi Küme 1 ve Küme 2’de canlı yayın (webinar) katılımı ve internet sitesi ziyaretinin ağırlıklı olduğu görülmektedir. Küme 4’te ise oldukça yüksek oranda sanal tanıtıma yakınlık olduğu görülmektedir.

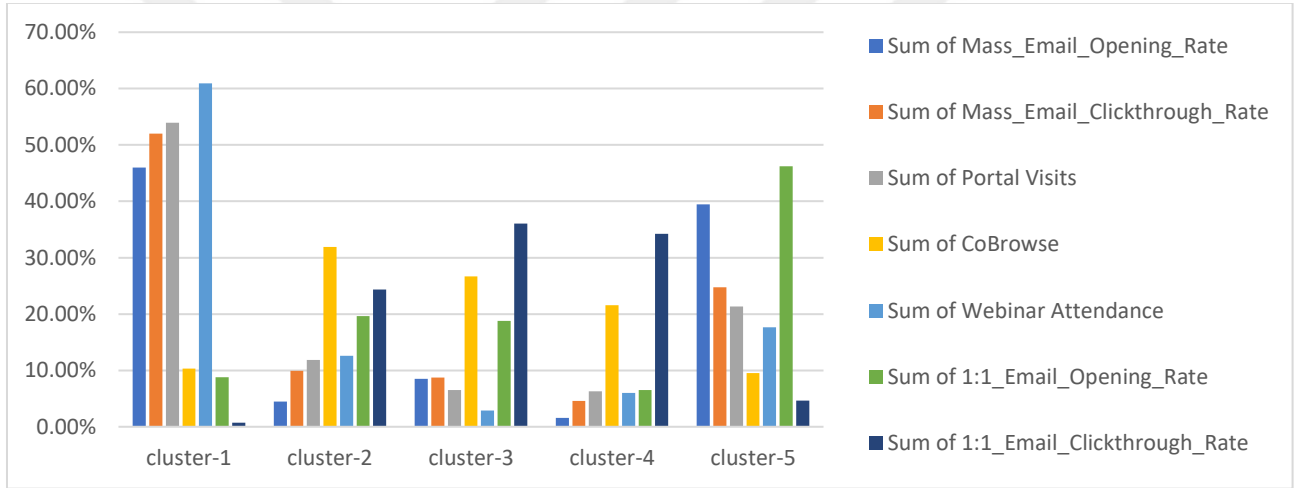
Model 3: K-ortalamlar 5’li Kümeleme Sonuçları

SPSS Clementine programı yardımıyla seçilen k-ortalamlar algoritmasına küme sayısı 5 verilerek normalize edilmiş veri seti ile kümeleme yapılmıştır. Bu kümeleme sonucunda elde edilen 5 küme Şekil 45’de verilmiştir.

Şekil 45: K-ortalamlar 5’li kümeleme sonuçları



Şekil 46: K-ortalamlar 5’li kümelemenin iletişim kanalları bazında sonuçları



K-ortalamlar algoritması ile elde edilen 5 küme iletişim kanallarına göre incelendiğinde Küme 1 ve Küme 5’de canlı yayın (webinar) katılımı ve internet sitesi ziyaretinin ağırlıklı olduğu görülmektedir. Küme 2, 3 ve 4’te ise oldukça yüksek oranda sanal tanıtıma yatkınlık olduğu görülmektedir.

Model 3’ün İzlenmesi, Sonuçlar ve Görüşler

Model 3’te k-ortalamlar ve iki aşamalı ile gerçekleştirilmiş 4 ve 5 li kümeleme çalışmaları sonucunda alınan çıktılar yorumlandığından hemen hemen aynı cevaba ulaşıldığı görülmüştür.

5.5. UYGULAMADA ELDE EDİLEN BULGULAR

Uygulamada çalışılan tüm Modeller sonucunda gerçek hayata da uyarlanabileceği düşünülen 3 küme elde edilmişti. Bu kümelerin özellikleri aşağıdaki gibi sıralanabilir:

- ✓ Küme 1: İnternet sitesi ve canlı yayın (webinar) katılımı gösteren hekimler
- ✓ Küme 2: Sanal tanıtıma’a yatkın hekimler
- ✓ Küme 3: Etkin katılım göstermeyen hekimler

Hekimlerin gerekleřtirdiđi aktivitelere gre, İnternet sitesi ile webinar katılımı sađlayan grup ve sanal tanıtıma yatkın grup ile ayrıca etkin katılım sađlamayan grup olmak zere 3 kmeye ayrıldıkları gzlenmiř. İř sahiplerinin sanal tanıtıma yatkın kiřilerin webinar katılımı da gsterir dřncesinin hekimlerin gerek davranıřlarına bakıldığında tam olarak dođruyu yansıtmadığını sylemek elde edilen sonular ile mmkndr.



SONUÇ

Çalışmada, k-ortalamalar ve iki aşamalı kümeleme yöntemleri uygulanarak 50 yılı aşkın bir süredir ilaç sektöründe faaliyet gösteren Türkiye ilaç sektörünün lider firmaları arasında yer alan uluslararası bir ilaç firmasının müşterilerinden en uygun sayıda kümenin oluşturulması ve ortaya çıkan kümelerden dijital pazarlama faaliyetleri açısından anlamlı sonuçlar elde edilmesi amaçlanmıştır. Çalışma sonucunda ortaya çıkan segmentasyon baz alınarak hekimlerin dijital yatkınlığına göre firmanın dijital pazarlama faaliyetlerini etkin bir şekilde uygulayabilmesi için elde edilen kümelerden faydalanılabileceği gözlemlenmiştir.

Temin edilen veri seti üzerinde etkin bir çalışma yapabilmek bu verileri iyi bir şekilde tanımakla mümkündür. Veri setinin özelliklerinin iyi bir şekilde kavranması ve yapılacak işlemlerin de bu özellikler dikkate alınarak yapılması gerekmektedir. Bu aşamada veri setinin yorumlanmasında sektör uzmanlarının önemli bir rolü bulunmaktadır. Ayrıca hangi hekim segmentine hangi bilimsel hizmetlerin sunulması, hangi tanıtım faaliyetleriyle ulaşılması gerektiği ilaç sektöründe önemli bir konu olduğu için dört üst düzey müşteri ilişkileri yöneticisi ve veri analisti ile çalışmada kullanılmak üzere hekimleri bölümlenmek için datası elde edilebilecek ve sayısallaştırılabilecek değişkenler ile dijital pazarlama kanalları belirlenerek firma müşterilerinin bu kanalları kullanımlarının yer aldığı veri tabanından yaklaşık 24 aylık bir veri seti temin edilip kontrolü gerçekleştirilmiştir. Veri setinin büyüklüğü sebebiyle verinin analizinde veri madenciliği algortimalarından yararlanılmıştır. Veri setinin excel yardımıyla min-maks normalizasyonu kullanılarak normalize edilip, SPSS Clementine programında 3 ayrı model için çalıştırılarak elde edilen kümelerin firmanın kendi yaptığı kümeleme sonuçları ile karşılaştırılması sağlanmış, hangi müşteri kümesine hangi dijital pazarlama kanalı ile yaklaşım uygulanmasına karar verilmesi hedeflenmiştir. Firmanın elinde bulunan kümeleme sonuçları satış temsilcilerinin, müşterilerinin dijital pazarlama kanallarına yatkınlıklarını yorumlamaları yoluyla elde edilmiş olup, bu çalışmadan elde edilen kümeleme sonuçları ile de karşılaştırılarak müşteri ilişkileri bölüm yetkilileri ile de incelenmiştir. Büyük sayıdaki veri setine analizin uygulanabilirliğini kolaylaştırması ve zaman tasarruf sağlaması sebebiyle ilgili ekip tarafından bu kümeleme çalışmasının sektörel açıdan da oldukça faydalı olduğuna karar verilmiştir. Analiz sonrası elde edilen çıktılar sayesinde firma, müşteri kümelerine uygulayacağı dijital pazarlama faaliyetlerinin seçiminde kolaylık sağlamıştır.

Çalışılan veri setine uygulanan ağırlıklandırma yöntemlerinin hepsinde aynı sonuçlara ulaşıldığı görülmüştür. Ağırlıklandırmanın bu veri seti üzerinde ayırıcı bir etkisi olmadığı sonucuna erişilmiştir. Sektörel uygulanabilirliği sebebiyle 3 olarak tercih edilen küme sayısı, aynı zamanda kümeleme algoritmalarına da bırakılarak ortaya çıkan küme sayılarının uygulanabilirliği gözlemlenmiş ve elde edilen küme sayılarının analiz edilmesi ile en efektif küme sayısının 3 olacağı sonucuna varılmıştır.

Sağlık ve ilaç sektöründe faaliyet gösteren firmaların, diğer sektörlerle oranla fazla sayıda hukuki düzenlemeye tabi olsalar da sektörde hızla yaygınlaşan sosyal medya kullanımı ve dijital pazarlama faaliyetleri ile gelişen pazarlama stratejilerine ayak uydurabilmeleri için pazarlama faaliyetlerini hızlı ve kolay bir şekilde yürütmeleri oldukça önemli hale gelmiştir. Teknoloji ve dijital pazarlama faaliyetlerinin değişimi ve gelişimiyle paralel olarak sıklıkla pazarlama faaliyetlerini yenilemeleri gerekmektedir. Veri madenciliği, büyük veri setlerine uygulanabilirliği ve kısa zamanda çıktılara erişilebilme imkanı vermesi açısından dijital

pazarlama faaliyetlerinin analizi için firmalara oldukça kolaylık sağlamaktadır. Veri madenciliğinin kümeleme metotları kullanılarak oluşturulabilecek müşteri gruplarına, uygun stratejilerin hızlı bir şekilde uygulanabilmesi firmaların rekabet ortamında satış politikalarını düzenlemesi için yararlı çıktılar elde edilmesini sağlamaktadır. Dijital pazarlama oldukça hızlı değişen ve gelişen bir konu olduğundan, zaman içerisinde veri setinin değişim ve gelişimine bağlı olarak yeni modeller geliştirilebilir. Yapılan segmentasyon çalışması belli aralıklarla yinelenerek elde edilen çıktıların pazarlama faaliyetlerine ne yönde katkıda bulunduğu gözlemlenebilir. Bu yönde yeni çalışmalar ortaya konabilir.



KAYNAKÇA

- Agrawal, R., Imielinski, T., Swami, A. (1993). *Mining Association Rules Between Sets of Items in Large Databases*. In Proceedings of the ACM SIGMOD International Conference on Management of Data. Washington, USA. 207-216.
- Akbulut, S. (2006). *Veri Madenciliği Teknikleri ile Bir Kozmetik Markanın Ayrılan Müşteri Analizi ve Müşteri Segmentasyonu*. Yüksek Lisans Tezi. Gazi Üniversitesi.
- Akın, Y.K. (2008). *Veri Madenciliğinde Kümeleme Algoritmaları ve Kümeleme Analizi*. Yayınlanmamış Doktora Tezi. İstanbul: Marmara Üniversitesi. Sosyal Bilimler Enstitüsü.
- Akman, M. (2010). *Veri Madenciliğine Genel Bakış ve Random Forests Yönteminin İncelenmesi: Sağlık Alanında Bir Uygulama*. Yüksek Lisans Tezi. Ankara Üni. Sağlık Bil. Ens.
- Akpınar, H. (2000). *Veritabanlarında bilgi keşfi ve veri madenciliği*. İstanbul Üniversitesi İşletme Fakültesi Dergisi. 29: 1-22.
- Aksoy, R. (2009). *İnternet Ortamında Pazarlama*, Seçkin Yayıncılık, Ankara
- Alabay, M., N. (2007). “*Müşteri İlişkileri Yönetimi (MİY) Turizm Sektörü Üzerinde Bir Uygulama*”, Gazi Üniversitesi SBE İşletme Anabilim Dalı Doktora Tezi, Ankara, s.115.
- Albayrak, A. Yılmaz S.K. (2009). *Veri Madenciliği: Karar Ağacı Algoritmaları ve İMKB Verileri Üzerine Bir Uygulama*. Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi. 14(1): 31-52.
- Alpar, R. (2011). *Uygulamalı Çok Değişkenli İstatistiksel Yöntemler*. Ankara: Detay Yayıncılık.
- Alpaydın, E. (2000). *Zeki Veri Madenciliği: Ham Veriden Altın Bilgiye Ulaşma Yöntemleri*. Bilişim 2000. 1-5.
- Amit G., Gavriely, N. (2009). *Intrator, Cluster Analysis and Classification of Heart Sounds. Biomedical Signal Processing and Control*. 2009; 4:26-36
- Anderberg, M. R. (1973). *Cluster Analysis for Applications*. New York: Academic Press.
- Aydoğan, E.K. Gencer, C. Akbulut, S. (2008). “*Veri Madenciliği Teknikleri ile Bir Kozmetik Markanın Ayrılan Müşteri Analizi ve Müşteri Bölümlenmesi*”. Journal of Engineering and Natural Sciences, Mühendislik ve Fen Bilimleri Dergisi. Vol./Cilt 26 Issue/Sayı 1.
- Aydoğan, F. (2003). *E-ticarete veri madenciliği yaklaşımlarıyla müşteriye hizmet sunan akıllı modüllerin tasarımı ve gerçekleştirimi*. Yüksek Lisans Tezi. Hacettepe Üniversitesi Fen Bilimleri Enstitüsü. Ankara. 12-16.
- Ayık, Z.Y. Özdemir, A.. (2007). *Lise Türü Ve Lise Mezuniyet Başarısının, Kazanılan Fakülte İle İlişkisinin Veri Madenciliği Tekniği İle Analizi*. Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi. 10(2). 443-447.
- Barnes, J.J. (2000). *Secrets Of Customer Relationship Management: It's All About How You Make Them Feel*. New York. McGraw-Hill. s.19.

- Baunsaythip, Catherine Runsal, Esa Rinta. (2001). “*Overview of Data Mining for Customer Behaviour Modeling*”. VTT Information Technology Research Report TTEI. s.121
- Baragoin, C. Gottschalk, H. Verhees, J. Pereira, P. Chan, R. (2002). *Enhance Your Business Applications: Simple Integration of Advance Data Mining Functions*. IBM. p.52
- Barutçu, S. (2008). “*Mobil Pazarlama*”. Ed: İnci Varinli, Kahraman Çatı, Güncel Pazarlama Yaklaşımlarından Seçmeler. Ankara: Detay Yayıncılık.
- Baykal, A. (2006). *Veri Madenciliği Uygulama Alanları*. D.Ü.Ziya Gökalp Eğitim Fakültesi Dergisi 7, 95-107
- Baykasoğlu, A. (2005). *Veri Madenciliği ve Çimento Sektöründe Bir Uygulama*. 7. Akademik Bilişim. Gaziantep.
- Bergeron, Bryan. (2002). *Essentials of CRM*. U.S.A.Wiley.
- Berkhin, P. (2002). *Survey of Clustering Data Mining Techniques*. California. U.S.A.
- Bland, C. (2002). *The Discovery of Multiple Level Profile Association Rules*. (Doctoral Thesis). Missisipi: Graduate School of Computer and Information Science of University of Missisipi.
- Bose, Nirmal, K. Liang, Ping. (1996). *Neural Network Fundamentals with Graphs, Algorithms and Applications*. United States. McGraw-Hill.
- Bülbül, Ş. (2001). *Çok Değişkenli İstatistiksel Yöntemler İle Bankaların Mali Yapılarının Analizi*. İstanbul: Marmara Üniversitesi İstatistik ve Ekonometri Araştırma ve Uygulama Merkezi
- Can, O. (2017). *Türkiye Sağlık Araştırmasının Veri Madenciliği Teknikleri ile İncelenmesi*. Kafkas Üniversitesi. Sosyal Bilimler Enstitüsü. Yüksek Lisans Tezi.
- Chachra, N., Savage, S., Voelker, G. M. (2015). *Affiliate Crookies: Characterizing Affiliate Marketing Abuse*. IMC '15 Proceedings of the 2015 ACM Conference on Internet Measurement Conference. Tokyo. 41-47.
- Chaffey, D., Smith, P. R., & Smith, P. R. (2013). *eMarketing eXcellence: Planning and optimizing your digital marketing*. Routledge.
- Chatfield, C., & Collins, A. J. (1980). *Introduction to Multivariate Analysis*. London: Chapman and Hall.
- Chung, H. M., Gray, P. (1999). *Special Section: Data Mining*. Journal of Management Information Systems. s:16/1
- Craven, M. W. and Shavlik, J. W. (1997). ‘*Using neural networks for data mining*’. Future Generation Computer Systems. 13(2–3). pp. 211–229. doi: 10.1016/S0167-739X(97)00022-8.
- Dick S. Meeks, A. Last, M. Bunke, H. and Kandel, A. (2004). *Data Mining In Software Metrics Databases*. Fuzzy Sets and Systems. 145/1.
- Dholakia, U., & Bagozzi, R. P. (2001). *Consumer behavior in digital environments*. Digital marketing, 163-200.
- Doyle, P. (2003). *Değer Temelli Pazarlama*. MediaCat Kitapları. İstanbul.

- Dunham, M.H. (2003). *Data Mining Introductory and Advanced Topics*. Pearson Education Inc. New Jersey.
- Durmuş, M.S. ve İplikçi, S. (2007). “*Veri Kümeleme Algoritmalarının Performansları Üzerine Karşılaştırmalı Bir Çalışma*”. Akademik Bilişim 2007. 31 Ocak - 02 Şubat. Kütahya: Dumlupınar Üniversitesi.
- Elmas, Ç. (2003). *Yapay Sinir Ağları*. Ankara. Seçkin Yayıncılık. 1.Baskı.
- Facebook. (2016, Ocak 27) Facebook Reports Fourth Quarter and Full Year 2015 Results. (Çevrimiçi). *Facebook Investor Relations*.
<https://investor.fb.com/investor-news/default.aspx>, [Erişim Tarihi: 09/04/2019]
- <https://investor.fb.com/investor-news/press-release-details/2018/Facebook-Reports-Fourth-Quarter-and-Full-Year-2017-Results/default.aspx> [Erişim Tarihi: 09/04/2019]
- <https://investor.fb.com/investor-news/press-release-details/2019/Facebook-Reports-Fourth-Quarter-and-Full-Year-2018-Results/default.aspx>
- Fayyad, U., Piatetsky-Shapiro, G. ve Smyth, P. (1996). “*The KDD Process for Extracting Useful Knowledge From Volumes of Data*”. Communications of ACM. 39, 11, 27-34.
- Fişek, Nusret. (1982). *Prof. Dr. Nusret Fişek'in Kitaplaşmamış Yazıları-I: Sağlık Yönetimi*. (Çevrimiçi). *Türk Tabipleri Birliği*. http://www.ttb.org.tr/n_fisek/kitap_1/33.html (Erişim Tarihi: 19/03/2015).
- Gan, G., Ma, C., & Wu, J. (2007). *Data Clustering Theory, Algorithms, and Applications*. Philadelphia: Siam.
- Giray, S. (2016). *İki Aşamalı Kümeleme Analizi İle Hükümlü Verilerinin İncelenmesi*. Ekonometri ve İstatistik e-dergisi. s.16.
- Giudici, P. (2003). *Applied Data Mining: Statistical Methods for Business and Industry*. John Wiley & Sons Inc. Chichester. 85-100.
- Gümüş S., Korkmaz M. (2014). *Hizmet Kalitesinin Davranış Tabanlı Müşteri İlişkileri Yönetimiyle İlişkisi: Bir Banka Uygulaması*. Akademik Bakış Dergisi.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate Data Analysis*. Edinburg: Pearson.
- Hair Jr., J. F., & Black, W. (2000). *Cluster Analysis*. L. G. Grimm, & P. R. Yarnold, Reading and Understanding More Multivariate Statistics (s. 147-206). Washington: American Psychological Association
- Han, J., Kamber, M. (2006). *Data Mining Concepts & Techniques*. San Francisco. Morgan Kauffmann Publishers. s. 10.
- Haşiloğlu, S. B., Süer, İ. (2010). *Elektronik Posta İle Pazarlama Üzerine Bir Araştırma*. IUUD. Sayı: 1. 61-74.
- Höppner, F., Klawonn, F., Kruse, R., & Runkler, T. (1999). *Fuzzy Cluster Analysis, Methods for Classification, Data Analysis and Image Recognition*. New-York: John Wiley&Sons.

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 System with Applications. 26: 181-188.

IMS Health Report (2014). *Engaging Patients Through Social Media*.

Jain, A. K. (2010). *Data clustering: 50 years beyond K-means*. Pattern Recognition Letters. 31(8): 651-666.

Johnson, R. A., & Wichern, D. W. (1998). *Applied Multivariate Statistical Analysis*. New jersey: Prentice Hall.

Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation. IEEE transactions on pattern analysis and machine intelligence, 24(7): 881-892.

Karagöz, N. (2007). *Market Veri Tabanında Veri Madenciliği Uygulaması*. Yüksek Lisans. İstanbul Ticaret Üniversitesi Sosyal Bilimler Enstitüsü.

Ketchen, D.Jr. ve Shook, C.L. (1996). “*The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique*”. Strategic Management Journal. 17(6). pp. 441-458.

Kiremitçi, B. (2005). *Veri Ambarlarında Veri Madenciliği ve Ulaştırma-Lojistik Sektöründe Bir Uygulama*. Yüksek Lisans Tezi. _İstanbul Üniversitesi. Sayısal Yöntemler. s.31.

Khattree, R., & Naik, D. N. (2000). *Multivariate Data Reduction and Discrimination with SAS Software*. North Carolina: SAS Institute Inc.

Kırım, Arman. (2004). *Strateji ve Bire-Bir Pazarlama CRM*. Sistem Yayıncılık. İstanbul. 46-47.

Kotler P., Armstrong G. (2001). “*Principles of Marketing*”. Prentice Hall. New Jersey. 245-256.

Kotler P. (1976). “*Marketing Management, Analysis, Planning and Control*”. Prentice Hall. New Jersey. 141-158.

Kotorov, R. (2003). “*Customer relationship management: strategic lessons and future directions*”. Business Process Management Journal. Vol.9, No:5, pp. 566-568.

Kotu, V. Deshpande, B. (2015). *Predictive analytics and data mining Concepts and Practice with RapidMiner*. Elseiver, ISBN: 978-0-12-801460-8.

Koyuncugil, A.S. ve Özgülbaş, N. (2009). *Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları*. BİLİŞİM TEKNOLOJİLERİ DERGİSİ. CİLT: 2. SAYI: 2.

KPMG. (2018). *Sektörel Bakış İlaç*. 2018:6.

Kulluk, S. (2009). *Karınca Koloni Optimizasyonu ile Yapay Sinir Ağlarından Kural Çıkarımı*. Doktora Tezi. Erciyes Üniversitesi. Fen Bilimleri Enstitüsü. Kayseri.

Küçükşille, E. (2009). *Veri Madenciliği Süreci Kullanılarak Portföy Performansının Değerlendirilmesi Ve İmkb Hisse Senetleri Piyasasında Bir Uygulama*. (Doktora Tezi). Isparta: Süleyman Demirel Üniversitesi Sosyal Bilimler Enstitüsü.

Larose, D. T. (2005). *Discovering Knowledge in Data, An Introduction to Data Mining*. USA: John Wiley & Sons, Inc.

Lattin, J. M., Carroll, J. D., & Green, P. E. (2003). *Analyzing Multivariate Data*. Toronto: Thomson.

Ledolter, J. (2013). *Data mining and business analytics with R*. John Wiley & Sons. New Jersey.

Lieb, R. (2012). *Content Marketing: Think Like a Publisher—How to Use Content to Market Online and in Social Media*. Que Publishing. Indianapolis

Lipscomb, Darrin. (2004) “*Making the Case for Customer Relationship Management*”.

Maimon, O. and Rokach, L. (2005). ‘*Introduction to knowledge discovery in database*’. Data Mining and Knowledge Discovery Handbook. pp. 1–17. doi: 10.1007/0-387- 25465-X_1.

Malhotra K. Naresh. (1996). *Marketing Research: An Applied Orientation*. Second Edition. New Jersey: Prentice Hall Inc.. s.670-672.

Michael J.A. Berry, Gordon S. Linoff. (2000). *Mastering Data Mining The Art Science of CRM*. U.S.A. Wiley. s.14.

Mitra, S., Acharya, T. (2003). *Data Mining :Multiple Soft Computing and Bioinformatics*. John Wiley & Sons Publisher.

Miyamoto, S., Ichihashi, H., & Honda, K. (2008). *Algorithms for Fuzzy Clustering Methods in c-Merans Clustering with Application*. Heidelberg: Springer.

Mooi, E., & Sarstedt, M. (2011). *A Concise Guide Market Research: The Process, Data and Methods Using IBM SPSS Statistics*. Berlin Heidelberg: Springer.

“*Müşteri İlişkileri Yönetimi, CRM Alt Yapısının Oluşturulması*”, (Çevrimiçi) <http://www.teknoturk.org>, 05.08.2005

Ogilvy Common Health Worldwide (2017). *Annual Research Report*.

Ontario.ca (2015). “*Social media for small business*”. <https://www.ontario.ca/page/social-media-small-business> (Erişim Tarihi: 16.03.2019)

Özbay, Ö. (2015). *Öğretim Yönetim Sistemi Üzerinde Üniversite (Lisans) Düzeyindeki Öğrenci Hareketliliğinin Veri Madenciliği Yöntemleriyle Analizi*. Başkent Üniversitesi. Eğitim Bilimleri Enstitüsü. Yüksek Lisans. Ankara.

Özdamar, K. (2013). *Paket Programlar ile İstatistiksel Veri Analizi 2*. Ankara: Nisan Kitabevi.

Özdamar, K. (2004). *Paket Programlar İle İstatistiksel Veri Analizi 2*. Eskişehir: Kaan Kitabevi.

Özkan, Y. (2008). *Veri Madenciliği Yöntemleri*. Papatya Yayınları. İstanbul

- Öztemel, E.(2003). *Yapay Sinir Ağları*. İstanbul. Papatya Yayıncılık. 1.Baskı.
- Park, S. Piramuthu, S. Shaw, M. (2001). “*Dynamic Rule Refinement in Knowledge Based Data Mining Systems*”. Decision Support Systems. No:3. p.205.
- Piramuthu, S. (1998). “*Evaluating Feature Selection Methods For Learning in Data Mining Applications*”. Thirty- First Annual Hawai International Conference on System Sciences. IEEE Computer Society. 6-9 January. Kohala Coast Hawaii USA. 294.
- Puchkov, S. (2016). *Internet Marketing: Top 10 Most Effective Strategies*, CreateSpace Independent Publishing Platform. South Carolina
- PwC Pharma. (2023). *Vizyon 2023 Küresel İlaç Raporu*. 2012:23.
- Ray, S. Turi, R. H. (1999). *Determination of number of clusters in k-means clustering and application in colour image segmentation*. In Proceedings of the 4th international conference on advances in pattern recognition and digital techniques (pp. 137-143).
- Roiger, R. J. Geatz M. W. (2003). *Data Mining A Tutorial-Based Primer*. Addison Wesley. USA. 350p.
- Ryan, D. (2016). *Understanding digital marketing: marketing strategies for engaging the digital generation*. Kogan Page Publishers.
- Rygielski, C., Wang, J.-C. and Yen, D. C. (2002) ‘*Data mining techniques for customer relationship management*’, Technology in Society, 24(4), pp. 483–502. doi: 10.1016/S0160-791X(02)00038-6.
- Sağıroğlu, Ş. Beşdok, E. Erler, M. (2003) *Mühendislikte Yapay Zeka Uygulamaları-I: Yapay Sinir Ağları*. Ufuk Kitap Kırtasiye Yayıncılık. 1. Baskı. Kayseri.
- Sarmanova, A. (2013). *Veri Madenciliğindeki Sınıf Dengesizliği Sorununun Giderilmesi*. Yıldız Teknik Üniversitesi. Fen Bilimleri Enstitüsü. Yüksek Lisans Tezi. İstanbul.
- SAS (2015). “*Digital Marketing*”. http://www.sas.com/en_us/insights/marketing/digital-marketing.html. (Erişim Tarihi: 16.03.2019)
- Savaş, S. Topaloglu, N. Yılmaz, M. (2012). *Veri Madenciliği ve Türkiye’deki Uygulama Örnekleri*. İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi. s. 21, ss. 1-23.
- Serper, Ö. (1997). *Uygulamalı İstatistik I*. 1.Baskı. Marmara Kitabevi. Bursa. s.202.
- Shahnam, Liz. (2009). CRM. http://www.microsoft.com/turkiye/dynamics/crm/crm_nedir.msp.
- Shapiro, D.E., Rodricue, J.R., Boggs, S.R., et al. (1994). *Cluster Analysis of The Medical Coping Modes Questionnaire: Evidence for Coping with Cancer Styles*. Journal of Pschosomeric Research. 1994; 38(2):151-159
- Sharma, S. (1996). *Applied Multivariate Techniques*. New York: John Wiley & Sons, Inc.
- Silahtaroglu, G. (2004). *Veri Madenciliğinde Kümeleme Analizi ve Öğretim Başarısının Değerlendirilmesine İlişkin Bir Uygulama*. Doktora Tezi, İstanbul Üniversitesi. Sayısal Yöntemler. ss.4-5

- Scharl A., Dickinger, A., Murphy, J. (2005). “*Success Factors and Industry Diffusion of Mobile Marketing*”. In *Electronic Commerce Research and Applications*. vol.4(1). pp.159–173.
- Sondwale, P. P. (2015) ‘*Overview of Predictive and Descriptive Data Mining Techniques*’. *International Journal of Advanced Research in Computer Science and Software Engineering*. 5(4), pp. 262–265.
- Soyer, A. (1999). *İlaç Endüstrisinde Küreselleşme*. Doğa Basın Yayın. İstanbul. 1999.
- Stiehl C. (2011). “*Find Out What the Customer Wants, First*”.
- Strauss, J., Frost, R. (2009). *E-Marketing*. 6th Ed. Upper Saddle River, NJ: Prentice Hall. Pearson.
- Sütçü, Cem Sefa. (2012). *Sosyal Medyaya Girmeden Önce Bilinmesi Gerekenler*. İstanbul: Anahtar Kitaplar.
- Swift, R. S. (2001). “*Accelerating Customer Relationships Using CRM and Relationship Technologies*”. Prentice Hall PTR. Upper Saddle River, NJ.
- Symeonidis, Klimis. (2000). “*Hand Gesture Recognition Using Neural Networks*”. School of Electronic and Electrical Engineering. Master of Science in Multimedia Signal Processing Communications.
- Şimşek, T. (2006). *Veri Madenciliği ve Müşteri İlişkileri Yönetiminde (CRM) bir Uygulama*. Doktora Tezi. İstanbul Üniversitesi. Sosyal Bilimler Enstitüsü. İstanbul.
- Tan, P. Michael, S. Vipin, K. (2005) *Introduction to Data Mining*. Addison Wesley. p.32.
- Taşpınar, Hasan. *Bilişim Altyapısıyla CRM, Müşteri İlişkileri Yönetimi*. Seçkin Yayınevi. Ankara. 2005.
- Tatlıdil, H. (2002). *Uygulamalı Çok Değişkenli İstatistiksel Analiz*. Ankara: Akademi Matbaası.
- Tezcanlar, P. (2007). *Müşteri İlişkileri Yönetimi Veri Madenciliği ve Bir Uygulama*. Yüksek Lisans Tezi. İstanbul.
- Timm, N. H. (2002). *Applied Multivariate Analysis*. New York: Springer.
- Thompson, Bob. (2004) “*What is CRM?*”, (Çevrimiçi) <http://www.crmguru.com/>
- Tiego, Maria. Teresa Pinheiro. Melo Borges, Jose’ Manuel Cristo’va’o Veri’ssimo. 2014, “*Digital Marketing and Social Media: Why Bother?*”, *Business Horizons*, Vol.57, Num.6, pp. 703-708.
- Tosyalı, H. ve Sütçü, C.S. (2016). *Sağlık İletişiminde Sosyal Medya Kullanımının Bireyler Üzerindeki Etkileri*. Maltepe Üniversitesi İletişim Fakültesi Dergisi. 3(2), 3-22.
- Tsiptsis, K. Chorianopoulos, A. (2009). *Data Mining Techniques in CRM: Inside Customer Segmentation*. Chichester. West Sussex: John Wiley & Sons.

Uzar, C. (2013). *Finansal Bilgi Sisteminde Veri Madenciliği Teknolojisinin Kullanılması: Borsa İstanbul Üzerine Bir Uygulama*. Dokuz Eylül Üniversitesi. Sosyal Bilimler Enstitüsü. Doktora Tezi. İzmir.

Vatansever, M. (2008). *Görsel Veri Madenciliği Tekniklerinin Kümeleme Analizlerinde Kullanımı ve Uygulanması*. Yayınlanmamış Yüksek Lisans Tezi. İstanbul: Yıldız Teknik Üniversitesi Fen Bilimler Enstitüsü.

Weinberger, K. Q. and Saul, L. K. (2009) ‘Distance metric learning for large margin nearest neighbor classification’, The Journal of Machine Learning Research(JMLR), 10, pp. 207–244. Available at: <http://portal.acm.org/citation.cfm?id=1577078>.

Wertime, K., Fenwick, I. (2008). *The essential Guide to New Media and Digital Marketing*.

Wind, Jerry. Vijay Mahajan. (2001). *Digital Marketing: Global Strategies From the World's Leading Experts*.
https://books.google.com.tr/books?hl=tr&lr=&id=Esrc5ZfQj3wC&oi=fnd&pg=PR9&dq=properties+of+digital+marketing&ots=gabLQIDaZk&sig=HyUXMzFzlzo013k7qTiX_xCsc2U&Retir_esc=y#v=onepage&q=properties%20of%20digital%20marketing&f=false. (Erişim Tarihi: 19.04.2019)

Yalçın, D. (2008). *CRM ve CRM Algısı*. (Beykent Üniversitesi, Sosyal Bilimler Enstitüsü, İşletme Yönetimi, Anabilim Dalı, Yayınlanmamış Yüksek Lisans Tezi), İstanbul.

Yereli, A., N. (2001). “Yönetim ve Ekonomi”. Celal Bayar Üniversitesi İ.İ.B.F. Dergisi, Yıl: 2001, Cilt:7,Sayı 1, s. 10.

Yıldırım, M. (2016). *İldeki Kurumlar Arası Çalışma Performansının Arttırılmasında Veri Madenciliği Tekniklerinin Kullanılması*. Fırat Üniversitesi Fen Bilimleri Enstitüsü. Yüksek Lisans Tezi. Elazığ.

Yurdakul, N. B., Bat, M. (2011). *Şirketler İçin Rekabette Sanal Farkındalık: Arama Motoru Pazarlaması*. Gümüşhane Üniversitesi İletişim Fakültesi Elektronik Dergisi.

Yurdakul, M. (2002). *Yeni Bir Pazarlama Stratejisi Olarak Müşteri İlişkileri Yönetimi (CRM) 'nin Sektörel Bazda Uygulanabilirliği*, <http://dergipark.gov.tr/dpusbe/issue/4749/65230>, Erişim Tarihi: 22.04.19.

Zaki, M. J. and Meira Jr, W. (2014). *Data mining and analysis: fundamental concepts and algorithms*, Cambridge University Press. New York.

Zengin, H. (2004). *Türkiye’de Müşteri İlişkileri Yönetimi Uygulamaları ve Yönetici Algılarının İncelenmesi*. Sakarya Üniversitesi. SBE. Doktora Tezi. Sakarya. s.52.