

T.C.
RECEP TAYYIP ERDOĐAN ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İŞLETME ANABİLİM DALI

VERİ MADENCİLİĐİ VE MARKET VERİTABANINDA BİRLİKTELİK
KURALLARININ BELİRLENMESİ

Yüksek Lisans Tezi

Tezin Yazarı

Burcu ALAN

Tez Danışmanı

Yrd.Doç.Dr. Erol USTA AHMETOĐLU

RİZE 2016

T.C.
RECEP TAYYIP ERDOĞAN ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İŞLETME ANABİLİM DALI

**VERİ MADENCİLİĞİ VE MARKET VERİTABANINDA BİRLİKTELİK
KURALLARININ BELİRLENMESİ**

Yüksek Lisans Tezi

Tezin Yazarı

Burcu ALAN

Tez Danışmanı

Yrd.Doç.Dr. Erol USTAAHMETOĞLU

Tez Savunma Tarihi

15.07.2016

Tez Jürisi Üyeleri

Adı ve Soyadı

İmza

Başkan:Yrd.Doç.Dr. Erol USTAAHMETOĞLU

Üye:Yrd.Doç.Dr. Özgür DOĞAN

Üye:Yrd.Doç.Dr. Güler YANIK

Enstitü Müdürü
Doç. Dr. Ahmet İshak DEMİR

...../ / 2016
Onay Tarihi

RECEP TAYYİP ERDOĞAN ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Bu tezi bilimsel metotlara ve etik davranış ilkelerine uygun olarak hazırlayıp sunduğumu, tezde bana ait olmayan tüm bilgi, düşünce ve sonuçları belirttiğimi ve kaynağını gösterdiğimi beyan ederim. 15/07/2016

Burcu ALAN

ÖNSÖZ

Bu tez çalışmasında öncelikle veri tabanlarında bilgi keşif süreci ve veri madenciliği hakkında temel bilgiler verilerek, diğer disiplinlerle ilişkileri, faydaları, uygulama alanları, uygulama süreci, veri madenciliği modelleri ve teknikleri, veri madenciliğinde karşılaşılan problemler ve kullanılan yazılımlar konuları üzerinde detaylı olarak durulmuştur. Tezin uygulama kısmında bir perakende firmasına ait alışveriş hareketleri üzerinde birliktelik kuralları uygulanarak birbirleri ilgisi olan ürünler arasındaki ilişkiler tespit edilmiş ve market yöneticilerine yeni raf düzenlemeleri hakkında önerilerde bulunulmuştur.

Tez çalışmam boyunca değerli fikir ve önerileriyle beni yönlendiren tez danışmanım Yrd.Doç.Dr. Erol USTAAHMETOĞLU'na ve her zaman destek ve dualarını yanımda hissettiğim en kıymetli varlığım eşim Kadir ALAN'a ve aileme de teşekkür ederim.

Rize,2016

Burcu ALAN

İÇİNDEKİLER

Sayfa

ÖNSÖZ.....	4
İÇİNDEKİLER.....	5
TABLolar LİSTESİ.....	8
ŞEKİLLER LİSTESİ.....	9
KISALTMALAR LİSTESİ.....	11
GİRİŞ.....	12

BİRİNCİ BÖLÜM

VERİ TABANLARINDA BİLGİ KEŞİF SÜRECİ VE VERİ MADENCİLİĞİ

1. VERİ TABANLARINDA BİLGİ KEŞİF SÜRECİ.....	14
1.1. Bilgi Keşif Süreci	14
1.1.1. Veri Hazırlama	15
1.1.1.1. Veri Temizleme	16
1.1.1.2. Veri Birleştirme	16
1.1.1.3. Veri Dönüştürme	17
1.1.1.4. Veri İndirgeme	18
1.1.2. Veri Madenciliği	21
1.1.3. Yorumlama ve Değerlendirme.....	21
1.2. Veri Madenciliğinin Tanımı	21
1.3. Veri Madenciliği İle İlişkili Bilim Dalları.....	22
1.4. Literatürde Veri Madenciliği.....	23
1.5. Veri Madenciliğinin Faydaları	24
1.6. Veri Madenciliği Uygulama Alanları.....	25
1.7. Veri Madenciliği Uygulama Süreci.....	27
1.7.1. Problemin Tanımlanması	28
1.7.2. Verilerin Hazırlanması	29
1.7.2.1. Toplama.....	30

1.7.2.2.	Değer Biçme	30
1.7.2.3.	Birleştirme ve Temizleme	30
1.7.2.4.	Seçim	31
1.7.2.5.	Dönüştürme	31
1.7.3.	Modelin Kurulması ve Değerlendirilmesi.....	32
1.7.4.	Modelin Kullanılması	35
1.7.5.	Modelin İzlenmesi.....	35
1.8.	Veri Madenciliğinde Kullanılan Modeller ve Teknikler	36
1.8.1.	Veri Madenciliği Modelleri.....	36
1.8.1.1.	Tahmin Edici Modeller	37
1.8.1.1.1.	Sınıflama	37
1.8.1.1.2.	Regresyon.....	39
1.8.1.2.	Tanımlayıcı Modeller	40
1.8.1.2.1.	Kümeleme	41
1.8.1.2.2.	Birliktelik Kuralları	43
1.8.2.	Veri Madenciliği Teknikleri.....	46
1.8.2.1.	Birliktelik Kuralı ve Pazar Sepet Analizi	46
1.8.2.1.1.	Apriori Algoritması.....	49
1.8.2.1.2.	Sepet Analizi	55
1.8.2.2.	Bellek Tabanlı Yöntemler	64
1.8.2.3.	Karar Ağaçları	65
1.8.2.4.	Yapay Sinir Ağları.....	68
1.8.2.5.	Genetik Algoritmalar.....	71
1.9.	Veri Madenciliği Uygulamalarında Karşılaşılan Problemler	73
1.9.1.	Veri Tabanı Boyutu.....	73
1.9.2.	Gürültülü Veri	73
1.9.3.	Boş (Null) Değerler.....	74
1.9.4.	Eksik Veri	74
1.9.5.	Artık Veri	74
1.9.6.	Dinamik Veri.....	75
1.9.7.	Farklı Tipteki Verileri Ele Alma.....	75

1.9.8. Sınırlı Bilgi.....	75
1.10. Veri Madenciliği Alanında Kullanılan Yazılımlar	75
1.10.1. Intelligent Miner.....	76
1.10.2. Darwin.....	76
1.10.3. Enterprise Miner.....	77
1.10.4. SPSS Clementine	77
1.10.5. Weka	78

İKİNCİ BÖLÜM

UYGULAMA

2. MARKET VERİTABANINDA BİRLİKTELİK KURALLARININ BELİRLENMESİ.....	79
2.1. Veri Madenciliği Süreci	79
2.1.1. Verilerin Toplanması ve Verilerin Düzenlenmesi	79
2.1.2. Veri Tabanının Oluşturulması.....	81
2.1.3. SPSS Clementine Programında Verinin Hazırlanması	85
2.1.4. Modelleme ve Bulgular.....	86
2.1.5. Modelin Geçerliliği ve Değerlendirme	96
3. SONUÇ VE ÖNERİLER.....	99
KAYNAKÇA.....	101
ÖZET.....	115
ABSTRACT.....	116
ÖZGEÇMİŞ.....	117

TABLolar LİSTESİ

	<u>Sayfa</u>
Tablo 1: Risk Matrisi	34
Tablo 2: Apriori Algoritmasının Kodları	51
Tablo 3: Market Satış Noktasındaki Alışveriş Kayıtlarını İçeren Veri tabanı	52
Tablo 4: Birliktelik Kuralları.....	55
Tablo 5: Süpermarketten Yapılan Alışveriş Bilgilerini İçeren Veri Tabanı	56
Tablo 6: Ürün Tekrar Sayısı.....	57
Tablo 7: Birliktelik Kuralları.....	58
Tablo 8: Kapalı Sık Tekrar Eden Ürün Kümeleri	63
Tablo 9: Kapalı Sık tekrar Eden Ürün Kümelerinden Yaratılan Birliktelik Kuralları	63
Tablo 10: Veri Madenciliğinde Kullanılan Bazı Yazılımlar ve Özellikleri.....	76
Tablo 11: Ürün Kategorileri.....	81
Tablo 12: Ürün Grupları İçin Elde Edilen Birliktelik Kuralları.....	89
Tablo 13: Ürün Kategorileri İçin Elde Edilen Birliktelik Kuralları	93

ŞEKİLLER LİSTESİ

	<u>Sayfa</u>
Şekil 1: Veri Tabanında Bilgi Keşif Süreci	15
Şekil 2: Veri Hazırlama (Önişleme) Biçimleri	20
Şekil 3: Veri Madenciliği ve Diğer Disiplinler.....	22
Şekil 4: Veri Madenciliği Süreci	28
Şekil 5: Tahmin Edici ve Tanımlayıcı Modeller	36
Şekil 6: Kümeleme	42
Şekil 7: Market Sepet Analizi.....	47
Şekil 8: Süpermarket Yerleşimi.....	48
Şekil 9: Apriori Algoritmasının Uygulaması.....	53
Şekil 10: Karar Ağacı Örneği	67
Şekil 11: Yapay Sinir Ağı.....	70
Şekil 12: Market Alışveriş Kayıtlarına Ait Örnek Veri Seti.....	80
Şekil 13: SQL Server Management Studio’da Oluşturulmuş Olan MVDM Veri Tabanı.....	82
Şekil 14: MVDM Veri Tabanındaki t_market_veri_seti Tablosunun Yapısı.....	82
Şekil 15: MVDMSSIS Projesinde Yer Alan Excel ve SQL Server 2005 Veri Tabanı Bağlantı Nesnelerinin Yapılandırılması Adımı 1.....	83
Şekil 16: MVDMSSIS Projesinde Yer Alan Excel ve SQL Server 2005 Veri Tabanı Bağlantı Nesnelerinin Yapılandırılması Adımı 2.....	84
Şekil 17: pMarketSatis.dtsx Package’ı İçinde Yer Alan Data Flow Task ve Bağlantıları.....	84
Şekil 18: pMarketSatis.dtsx’teki Data Flow Task’ın İçeriği	85
Şekil 19: SQL Veri Tabanının Veri Kaynağı Yöneticisine Eklenmesi	85
Şekil 20: SPSS Clementine’de Kurulan Modelin Görüntüsü.....	86

Şekil 21: Ürün Gruplarının Alışveriş Hareketlerindeki Dağılımı	87
Şekil 22: Alışveriş Hareketlerindeki Tüm Ürün Gruplarının Dağılımı	88
Şekil 23: Ürünlerin Kategorileşmiş Veri Seti	91
Şekil 24: Ürün Kategorilerinin Alışveriş Hareketlerindeki Dağılımı.....	92
Şekil 25: Web Grafiği İle Tüm Ürün Kategorileri Arasındaki Birliktelik Diyagramı (çizgi değeri \geq %0)	95
Şekil 26: Birliktelik Sıklığı Yüksek Olan İlişkiler (çizgi değeri \geq %60)	96
Şekil 27: Önerilen Yerleşim Düzeni.....	98

KISALTMALAR LİSTESİ

I : Ürün Kümesi

I_d : Ürün (Nitelik)

m: Ürün Sayısı

D: İşlem Veritabanı

s: Destek

α : Güven

T: D Veritabanı İçindeki Hareket

X,Y: Ürün Kümeleri

$X \rightarrow Y$: Birliktelik Kuralı

L: Büyük Ürün Kümelerinin Kümesi

l : Büyük Ürün Kümesi

L_k : k Değerindeki Büyük Ürün Kümelerinin Kümesi

l_k : k Değerindeki Büyük Ürün Kümesi

C_k : k Değerindeki Aday Kümeler

p: Bölümlerin Sayısı

GİRİŞ

Günümüzde teknoloji her geçen gün hızla ilerlemekte ve bilgisayar sistemlerinin güçleri artmaktadır. Bununla birlikte işletmeler, marketler ve diğer kuruluşlarda veri tabanlarında kuruluşun amacına ve yapısına bağlı olarak çeşitli türlerde veri toplamaktadır. Gelişen veri tabanı teknolojisi ve veri hacimlerindeki artış, firmaların topladığı veriyi kullanılabilir bilgiye çevirme isteği toplanan bu veriyi işleyerek, verinin içerisindeki faydalı ve ilginç ilişkilerin, birlikteliklerin ortaya çıkarılması ihtiyacını ortaya çıkarmıştır. Bu noktada ihtiyaçları karşılayacak bir disiplin (çözüm) olarak veri madenciliği kavramı ortaya çıkmıştır.

Veri madenciliği, istatistiksel yöntemler ile çeşitli bilgisayar yazılımlarını kullanarak yeni ve kullanışlı bilgiyi veri tabanlarından çıkarma işlemidir. Bu yüzden veri madenciliği, veri tabanından anlamlı ilişkiler ve kurallar elde etmek için geniş uygulama imkânı olan bir alan olarak görülmektedir. Veri madenciliğinin veri tabanları üzerine uygulanmasıyla, veri tabanlarındaki bilgi keşfi süreci kavramı ortaya çıkmıştır. Genelde çok büyük hacimli verilerin çözümleme işlemlerinin yapıldığı bu süreç, veri temizleme, veri birleştirme, veri dönüştürme, veri indirgeme, veri madenciliği uygulaması ve sonuçların değerlendirilmesi adımlarından oluşur. Büyük hacimli veri tabanlarından anlamlı ve gizli ilişkilerin çıkarılması olarak tanımlanan veri madenciliği adımı bu sürecin en önemli kesimini oluşturmaktadır. Bu bağlamda veri madenciliği özellikle pazarlama, bankacılık ve finans, sigortacılık, borsa, eğitim ve elektronik ticaret alanlarındaki uygulamalarda yeni ve temel bir araştırma sahası olarak yaygın bir şekilde kullanılmaktadır.

Çalışmanın ilk bölümünde, veri tabanları, veri tabanlarında bilgi keşif süreci, veri madenciliği tanımı, veri madenciliğinin diğer bilim dalları ile olan ilişkisi, veri madenciliği uygulama alanları incelenmiştir. Bununla birlikte, pazarlama, bankacılık-finans, elektronik ticaret ve perakende sektörlerindeki pratik uygulamalar ve bu uygulamalardan firmaların elde etmiş olduğu bilgiler ve faydalar da anlatılmıştır. Genel olarak veri madenciliği süreçleri, bu süreçte kullanılan model ve teknikler, kullanılan yazılımlar ve uygulamada karşılaşılan problemler de bu bölümde yer almaktadır.

Çalışmanın ikinci bölümünde ise SPSS Clementine 11.1 paket programı kullanılarak birliktelik kurallarının en temel algoritmalarından olan Apriori Algoritması ile perakende sektöründe faaliyet gösteren bir market veri tabanı üzerinde birliktelik analizi yapılmıştır. Birliktelik kuralları sonuçlarına göre birlikte satın alınma eğilimindeki ürünler ve ürün kategorileri hakkında bilgiler verilmiştir. Sonuç olarak; birliktelik kuralları analizleri sonuçlarına göre market yöneticilerine yeni bir raf yerleşim düzeni önerilmiştir.

BİRİNCİ BÖLÜM

VERİ TABANLARINDA BİLGİ KEŞİF SÜRECİ VE VERİ MADENCİLİĞİ

1. VERİ TABANLARINDA BİLGİ KEŞİF SÜRECİ

1.1. Bilgi Keşif Süreci

Günümüzde teknolojinin hızlı gelişimi ile birlikte bilgisayar teknolojileri hayatımızın büyük bir bölümünde yer almaya başlamıştır. Yüksek kapasiteli veri kayıt cihazlarının maliyetlerinin düşük seviyede olması hem bireylere hem de işletmelere veri toplama alışkanlığı kazandırmış durumdadır. Örneğin; hastanelerde, belediyelerde, bankalarda yaptığımız her işlem, marketlerden aldığımız her ürün, gün içerisinde alışveriş merkezlerine giriş çıkış saatlerimiz düzenli olarak işletmeler tarafından kayıt altına alınarak saklanmaktadır. Veri kayıt sistemlerinin artan kullanımı ile birlikte veri hacimlerindeki büyük oranda artış, işletmeleri toplanan bu verilerden nasıl yararlanılabileceği sorunu ile karşı karşıya bırakmıştır. Geleneksel sorgulama ve raporlama araçlarının veri yığınları karşısında yetersiz kalması, Veri Tabanlarında Bilgi Keşfi gibi yeni araç arayışlarına neden olmaktadır.

Veri Tabanlarında Bilgi Keşfi, depolanmış veri içerisindeki geçerli, yeni, faydalı ve sonuç olarak anlaşılabilir bağıntıların çıkarılması sürecidir (Ay,2009:3).

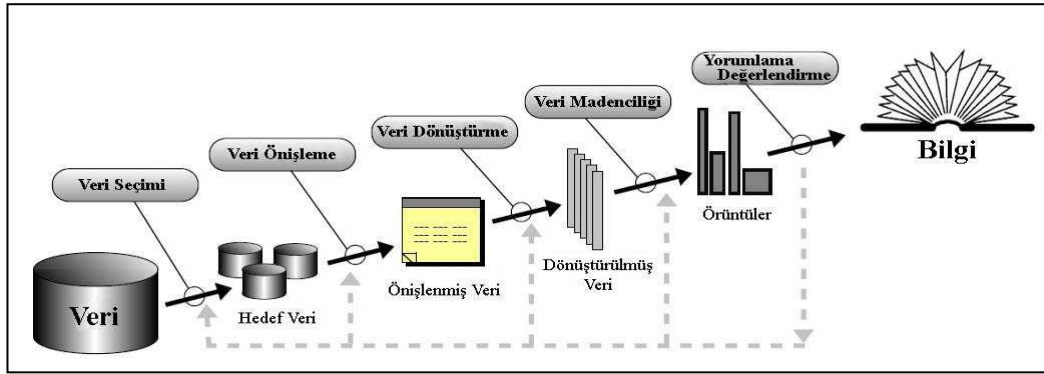
Veri tabanlarında saklı bilgiyi keşfedebilmek amacıyla ihtiyaç duyulan yeni nesil hesaplama teknikleri ve araçları, veri tabanlarında bilgi keşfinin konusunu oluşturmuştur (Fayyad vd,1996:82). Daha basit değerlendirmeyle veri tabanlarında bilgi keşfi, büyük verilerin analizindeki bilgisayar destekli işlemlerdir (Sivanandam vd.,2006:9).

Veri tabanlarında bilgi keşfi birbirini izleyen şu adımlardan oluşmaktadır (Han vd.,2001:7):

1. Veri Temizleme
2. Veri Birleştirme
3. Veri Dönüştürme

4. Veri İndirgeme
5. Veri Madenciliği
6. Yorumlama ve Değerlendirme

Veri Tabanlarında Bilgi Keşif Sürecinde bulunan ilk dört adım veri madenciliği adımına verinin hazırlanması için uygulanan veri önileme aşaması olarak da değerlendirilmektedir.



Şekil 1: Veri Tabanında Bilgi Keşif Süreci

(Kaynak: Jiawei Han, Micheline Kamber, Data Mining Concepts and Techniques, 2.bs., USA, The Morgan Kaufman Publishers,2006)

Veri tabanlarındaki bilgi keşif süreci aşağıda açıklanmıştır.

1.1.1. Veri Hazırlama

Verinin kalitesi, veri madenciliği için temel niteliği taşır. Veri madenciliğinden olumlu, işe yarar doğru sonuçlar alınabilmesi için ham veriler önceden bir takım işlemlerden geçirilmelidir. Veri hazırlama bilgi keşfi sürecindeki önemli adımlardan biridir. Veri madenciliği adımı modelin kurulması sırasında çıkabilecek problemler, yeniden bu adıma dönülmesine neden olabilir (Akpınar,2000:7). Keşif sürecindeki ilk adım verinin toplanmasıdır. İlk olarak elde var olan mevcut veriler gözden geçirilir ve toplanmasına gereksinim duyulan veri belirlenir. İhtiyaç duyulacak verilerden bazıları daha önce hiç toplanmamış olacağından, veri toplama adımı yeni verilerin elde edilmesini de içerebilir (Akçay,2014:24). Fakat bu veriler farklı kaynaklardan toplanacak ise bu verilerin bütünleştirilmesi ve veriler arası tutarsızlıkların bulunup, düzeltilmesi yine bu adımda yapılır. Kullanılacak veri belirlenirken her ne kadar uygulanacak

veri madenciliği metodu bilinmese de, veri madenciliği hedefleri ve kurumun hedefleri göz önünde bulundurulmalıdır (Giovinazzo,2002:331).

Verinin toplanmasıyla başlayan keşif sürecinde, veri hakkında daha fazla bilgi sahibi olmak için yapılan işlemler, veri kalitesiyle ilgili problemlerin belirlenmesi ve veri hakkındaki ilk izlenimi edinmeyle, verinin kavranması sağlanır (Becerra vd.,2002:790). Veri madenciliği analizi düşük kalitedeki veri seti üzerinde yapılırsa, kaliteli bilgiye ulaşmak imkânsız olur. Burada amaç, verinin kalitesini iyileştirmeye çalışmak olmalıdır (Fretias,2002:65).

1.1.1.1 . Veri Temizleme

Veri temizleme ile hatalı, geçersiz, eksik ve boş verilerin temizlenme işlemleri yapılır. Veri temizleme işlemindeki ilk basamak veri tabanındaki uyumsuzlukları tespit etmektir. Uyumsuzluklar çeşitli sebeplerle meydana gelebilir. Örneğin sipariş ile ilgili bir veri tabanında siparişi veren ile ilgili her bilgi bulunmayabilir ya da sipariş veri tabanı ile müşteri veri tabanında uyumsuzluklar olabilir. Bununla birlikte veri girişlerinde insan hataları olmuş olabilir. Veri temizleme işlemi için verinin nitelikleri, niteliklerin alabileceği değerleri ve uzunlukları bilinmelidir (Bilekdemir,2010:10).

1.1.1.2 .Veri Birleştirme

Veri birleştirme adımı, çeşitli veri kaynaklarındaki verilerin birleştirilmesi işlemini içermektedir. Bu adımda farklı veri tabanlarındaki veriler veri ambarlarında birleştirilirler (Bilekdemir,2010:12). Veri birleştirme işlemi yapılırken birçok faktörün dikkate alınması gereklidir. Veri madenciliğinde kullanılacak olan veriler birçok kaynaktan toplanmış olduğundan her kaynaktaki veriler kendi niteliklerine göre depolanmışlardır. Bunların bir araya getirilmesi doğal olarak veri uyumsuzluğuna sebep olacaktır. Bu uyumsuzluklar çoğunlukla verilerin farklı kodlama biçiminde olmasından kaynaklanır. Veri birleştirme adımı, toplanan bütün verilerin aynı biçimde tek bir veri tabanında veya veri ambarında birleştirilmesini kapsamaktadır (Moss,2003:314). Örneğin; bir mağazanın veri tabanının alışveriş kayıtlarında girişler “fatura numarası” biçiminde yapılmışken bir diğerinde “fiş numarası” biçiminde yapılmış olabilir.

Bu iki deęerin aynı biçimde birleřtirilmesi sürecin hatalı ilerlemesine sebep olabilir.

1.1.1.3 . Veri Dönüřtürme

Veri dönüřtürme adımı, verilerin anlam ve içeriklerinin deęiřtirilmeden veri madencilięinde kullanılacak modele uygun hale dönüřtürülmesi iřlemidir. Bazı veri madencilięi modelleri sadece sayısal deęerlerle çalıřırken, bazıları sadece kategorik deęerlerle çalıřabilmektedir. Örneęin bir veri tabanında “evet=1” , “hayır=0” řeklinde sayısal deęerler kullanılmıř, farklı bir veri tabanında ise “evet/hayır” biçiminde kategorik deęerler girilmiř olabilir. Veri tabanı üzerine uygulanacak metod hangisini destekliyorsa bütün girdilerin o formata dönüřtürülmesi daha iyi sonuç alınması için faydalı olacaktır (Akpınar,2000:8).

Veri dönüřtürme iřlemleri řu basamaklardan oluřur (Roiger vd.,2003:156):

- Verilerin düzeltilmesi: Verilerdeki gürültünün temizlenmesi adımıdır.
- Verilerin Birleřtirmesi: Verileri sonuçlara varabilmek için birleřtirme iřlemidir. Örneęin, müřteriye verilecek kredi miktarını belirleyebilmek için müřterinin aylık maařı ve ek gelirleri birleřtirilir.
- Verilerin Genelleřtirilmesi: Düşük seviyedeki ham veriyi daha yüksek seviyedeki deęere dönüřtürme basamaęıdır. Örneęin, mahalle ve sokak gibi gruplandırılmıř özellikler kent ve devlet isimleri altında toplanarak genelleřtirilebilir.
- Verilerin Normalleřtirilmesi: Verilerin daha önceden tespit edilmiř ufak mesafelere göre ölçeklendirilmesidir. Çeřitli řekillerde veri normalleřtirme yöntemleri bulunmaktadır. Bu yöntemlerden bazıları řunlardır (Oęuzlar,2003:73):
 - Min-Max veri normalleřtirmesi yöntemi ile özgün veriler yeni veri aralıęına doęrusal dönüşüm ile dönüřtürülürler.
 - Z Skor veri normalleřtirmesi teknięinde ise deęiřkenin rastgele bir y deęeri, deęiřkenin ortalaması ve standart sapmasına baęlı olarak bilinen Z dönüşümü ile normalleřtirilir.

- Ondalık Ölçekleme ile veri normalleştirmede ele alınan değişkenlerin değerlerinin ondalık kısmı hareket ettirilerek normalleştirme oluşturulur. Hareket edecek olan ondalık nokta sayısı, değişkenin azami mutlak değerine bağlıdır.

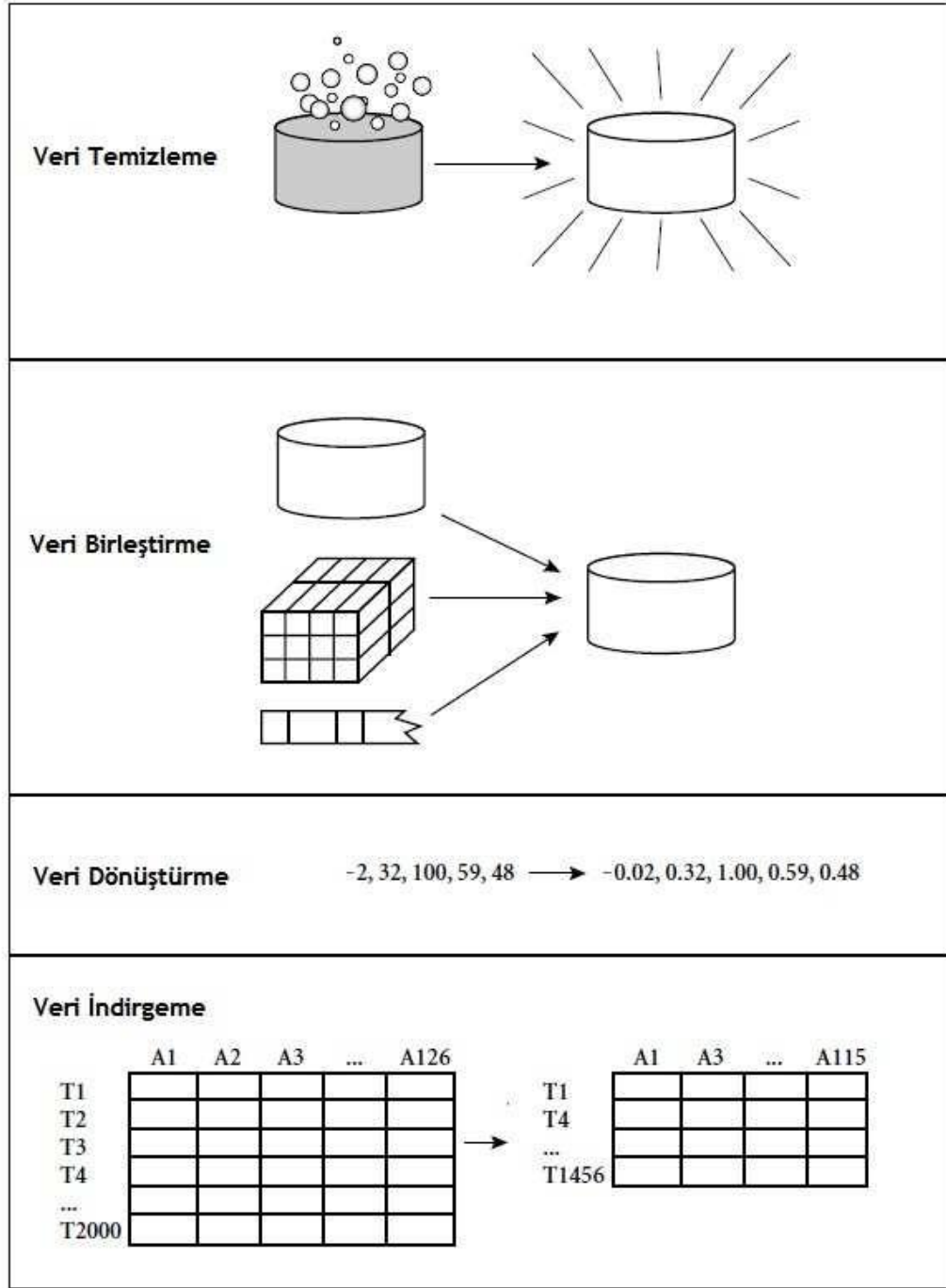
1.1.1.4. Veri İndirgeme

Üzerinde çalışılacak veri setleri çok büyük boyutlarda olduğunda veri madenciliği uygulamalarından olumlu sonuç almak zorlaşmakta ve işlemler çok uzun zaman almaktadır. Veri indirgeme, verilerin öznelikleri korunarak hacimce daha küçük biçime getirilme işlemidir. Bu sayede elde edilen indirgenmiş veri kümesine veri madenciliği teknikleri daha az zamanda uygulanarak daha etkin sonuçlar elde edilebilir (Han vd.,2001:115). Veri indirgeme adımı için bazı teknikler kullanılır. Bunlar şu şekildedir (Oğuzlar,2003:73):

- **Veri Birleştirme veya Veri Küpü:** Veri birleştirme yapılacak 2013-2015 yılları için çeyrek dönemlik ürün satış miktarlarından oluşan bir veri kümesinin bulunduğunu varsayalım. Bu yıllar için yıllık ürün satış miktarlarına ait veriler tek bir tabloda toplandığı zaman veri birleştirilmesi yapılmış olur. Sonuç olarak elde edilen veri kümesinin boyutu daha düşük seviyededir ancak yapılacak analiz için bir bilgi kaybı söz konusu değildir. Veri küpleri ise çok değişkenli birleştirilmiş bilginin saklanmış olduğu küplerdir. Örneğin, bir işletmenin yıllara göre ürün satış miktarları, satışı yapılan ürünler ve işletmenin farklı satış yerleri ile ilgili veriler için aynı küp üzerinde gösterilebilir. Veri küpleri özet bilgiye herhangi bir hesaplama yapmadan hızlı bir şekilde ulaşılmasına olanak sağlarlar(Oğuzlar,2003:73).
- **Boyut Azaltma (Dimension Reduction):** Veri madenciliği yapılacak veri kümesi bazen ilgisiz ve gereksiz değişkenler içerebilir. Örneğin, ürünlerin satışlarına istinaden oluşturulan bir veri yığnında, müşterilerin adresleri lüzumsuz bir değişken olarak yer alabilir. Bu tür lüzumsuz değişkenler elde edilecek ilişkileri niteliksizleştirebileceği gibi veri madenciliği sürecinin uzun zaman almasına da yol açacaktır. Lüzumsuz ve ilgisiz değişkenlerin elenmesi amacıyla ileri veya geri yönlü olarak sezgisel seçimler

yapılabilir. İleri yönlü sezgisel seçimde orijinal değişkenleri en iyi yansıtacak değişkenler belirlenir. Sonrasında ise her bir değişken veya değişkenler grubunun, bu kümeye dâhil edilip edilmeyeceği sezgisel olarak belirlenir. Geri yönlü sezgisel seçimde ise ilk olarak değişkenlerin tüm kümesi ele alınır. Daha sonra ise gereksiz ve ilgisiz bulunan değişkenler kümeden dışlanarak, en iyi değişken kümesi elde edilmeye çalışılır(Oğuzlar,2003:74).

- **Veri Sıkıştırma (Data Compression):** Veri sıkıştırmada ise analiz edilecek veri miktarını azaltmak için aynı cinsteki veriler sıkıştırılır. Bu teknikte verinin orijinal yapısını bozmamak önemlidir ancak bu yöntem bazen veri kayıplarına neden olabilmektedir (Han vd.,2001:115).
- **Kesikli Hale Getirme (Discretization):** Kesikleştirme, bazı veri madenciliği algoritmaları sadece sınıflandırılmış değerler üzerine ele aldığından, sürekli verilerin kesikli değerlere dönüştürülmesini içerir. Böylece sürekli verilerin kesikli değer aralıklarına dönüştürülmesiyle elde edilen kategorik değerler, özgün veri değerlerinin yerine kullanılabilirler. Örneğin, bir veri tabanında kullanıcı yaş değişkeni ayrıntılı biçimde yer alırken, bu değişkenler 1-15, 16-40, 40+ olacak biçimde kategorize edilebilir (Oğuzlar,2003:74).



Şekil 2: Veri Hazırlama (Önişleme) Biçimleri

(Kaynak: Jiawei Han, Micheline Kamber, Data Mining Concepts and Techniques, 2.bs., USA, The Morgan Kaufman Publishers,2006)

1.1.2. Veri Madenciliđi

Veri madenciliđi, veri tabanlarından bilgi keşfi sürecinin en önemli adımıdır. Bazı kaynaklarda eş değer süreci kapsadığı ileri sürülerek veri madenciliđi ve veri tabanlarında bilgi keşif süreci kavramlarının birbirlerinin yerine kullanıldığı görülmektedir (Becker vd.,2005:95). Fakat veri tabanlarında bilgi keşif süreci veriden faydalı bilgiyi keşfetmedeki bütün işlemleri kapsarken, veri madenciliđi bu sürecin önemli ama sadece bir adımıdır (Fayyad,1996:83).

Veri tabanlarında bilgi keşif süreci, veri madenciliđini de barındıran adımlardan oluşan bir süreçtir. Bu süreçte sürekli tekrarlar, adımlar arası ileri geri hareketler, atlamalar olabilir. Günümüzde genellikle sürecin veri madenciliđi adımına eğilim gösterilmektedir, fakat diğer aşamalar sürecin en az veri madenciliđi kadar önemli bir parçasıdır (Han vd.,2001:6).

1.1.3. Yorumlama ve Deđerlendirme

Veri tabanlarında bilgi keşfi sürecinin son adımı olan yorumlama ve deđerlendirme, veri tabanı üzerine uygulanan veri madenciliđi modellerinin izlenmesi ve deđerlendirilmesidir. Bilgi keşfi sürecinin ne kadar verimli olduğunu, işletmenin hedeflerini karşılayıp karşılayamadığını, karar verme sürecine yarar sağlayıp sağlamayacağını deđerlendirme adımıdır. Bununla birlikte sürecin tekrar edilip edilmeyeceđine de karar verilir. Bu adımda çeşitli görselleştirme ve raporlaştırma araçları kullanılarak keşfedilmiş bilgiler ilgili kullanıcılara sunulur (Akçay,2014:31).

1.2. Veri Madenciliđinin Tanımı

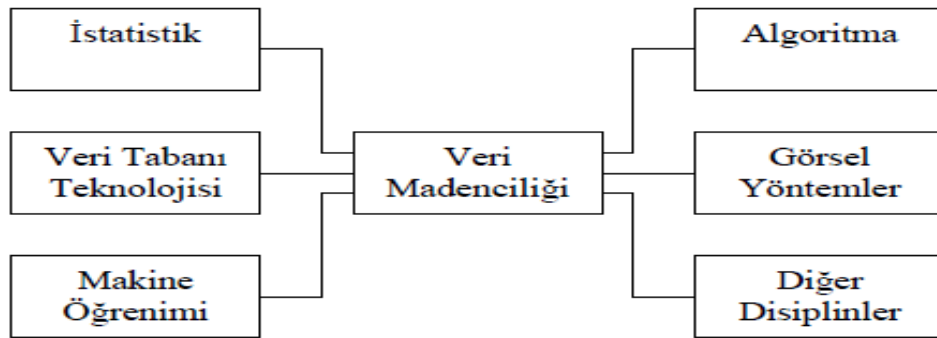
Veri madenciliđi genel anlamda, veri tabanlarında saklanan çok çeşitli verilerden daha önceden keşfedilmemiş, gizli kalmış, deđerli, kullanılabilir bilgilerin açığa çıkarılması biçiminde ifade edilmektedir. Veri madenciliđinin amacı, veri tabanlarında toplanmış olan bilgilerin istatistik yöntemlerle incelenip ilgili kurum ve yönetim destek sistemlerinde kullanılmak üzere deđerlendirilmesidir (Han vd., 2006:5).

Veri madenciliđinin farklı kaynaklarda birçok tanımı yapılmaktadır. Bunlardan en çok kullanılanları aşağıdaki gibidir:

- Veri madenciliği; veri ambarlarında tutulan çok çeşitli verilere dayanarak daha önce keşfedilmemiş bilgileri ortaya çıkarmak, bunları karar vermek ve yönetimde eylem planını gerçekleştirmek için kullanma sürecidir (Özmen,2001:2).
- Veri madenciliği, veri tabanları, veri ambarları veya diğer veri havuzlarında depolanmış büyük veri yığınlarından ilginç bilgilerin keşfi sürecidir (Han vd., 2006:105).
- Veri madenciliği gizli, bilinmeyen, potansiyel ve değerli, ilginç ya da eğitici bilgiye ulaşmada kullanılan araç ve yaklaşımlar olarak ifade edilir (Lu vd., 2008:3536).
- Veri madenciliği, büyük miktarda veri içinden gelecekle ilgili tahmin yapmamızı sağlayacak bağıntı ve kuralların bilgisayar programları kullanılarak aranmasıdır (Gürgen, 2008:2).
- Veri madenciliği, genellikle büyük veri setlerindeki, bazı bilinmeyen veya gizli kalmış kuralların keşfine ve analizine yarayan yöntemler ve teknikler kümesidir. Kısaca veri madenciliği, veriden bilgi çıkarma sanatıdır (Tuffery, 2011:4).

1.3.Verit Madenciliği İle İlişkili Bilim Dalları

Veri madenciliği farklı disiplinlerden yararlanan disiplinler arası bir alandır. Veri madenciliği; veri tabanı teknolojisi, istatistik, makine öğrenimi, görsel yöntemler gibi farklı disiplinler kümesinden oluşmaktadır.



Şekil 3: Veri Madenciliği ve Diğer Disiplinler

(Kaynak: Maindonald, J. Data Mining From A Atatistical Perspective, Preprint. Australian Nat. Univ. Stat Cons Unit, (<http://maths.people.anu.edu.au/~johnm/dm/dmpaper.html> Erişim Tarihi: 12.10.2015)

Veri madenciliği yaklaşımına bağlı olarak, veri madenciliğinde farklı disiplinlerde kullanılan tekniklerden de faydalanılmaktadır. Bu tekniklere; sinir ağları, bulanık küme teorisi, mantık gibi örnekler verilebilir (Şaylan, 2013:4).

1.4.Literatürde Veri Madenciliği

Veri madenciliği 1990'lı yıllarda ortaya çıkmıştır. Bir online veri tabanı olan Science Direct'te 1990 dan günümüze kadar bir taraması yapıldığında veri madenciliği ile ilgili 1950'ye yakın makale olduğu görülmektedir. İlgili yayınlarda veri madenciliğinin özellikle 2001 yılından bu yana büyük bir gelişme gösterdiği göze çarpmaktadır. Aşağıda veri madenciliği alanında farklı konularda gerçekleştirilen uygulama örneklerine yer verilmiştir.

- Caskey, genetik algoritma ve sinir ağları teknikleri ile bir fabrikadaki çalışma koşullarının problemlerini ortaya koymuş ve bu problemleri iyileştirici işletme stratejileri önermiştir (Caskey, 2001).
- Hong, Park ve Jon ve Rho, veri madenciliği tekniklerini kullanarak, piyasada rekabet üstünlüğü sağlamak isteyen işletmelere müşteri ihtiyaçlarını değiştirdiğini dikkate alarak tedarikçi seçiminde bulunmalarını önermiştir (Hong vd., 2005).
- Baker, veri madenciliği yöntemlerini kullanarak eğitim ortamında elde edilen veriler ile keşifler yaparak öğrencileri daha iyi anlamak için stratejiler önermiştir (Baker, 2010).
- Woody ve arkadaşları, veri madenciliği yöntemlerini kullanarak eğitim ortamında elde edilen veriler ile keşifler yaparak öğrencileri daha iyi anlamak için stratejiler önermiştir (Woody vd., 2010).
- Birant ve arkadaşları bir mağaza zincirinden edindikleri verilerle mağazalarda, çeşitli zamanlarda en sık birlikte satılan ürünleri tespit etmeye yönelik bir uygulama yapmışlardır (Birant vd., 2010).
- Cevahir, yaptığı bir sepet analizi çalışmasında perakendeci bir işletmenin veri tabanındaki kayıtlı satış hareketlerini incelemiş ve ayrıntılı bir satış analizi yapmıştır. Çalışma sonucunda en yüksek sıklığın şekerler ile yardımcı ve katkı malzemeleri arasında olduğunu ortaya çıkarmıştır. Bu sonuç ile firmanın reklam stratejilerini belirlemeye ve satış işlemlerini planlamaya yardımcı olabilecek bulgular elde etmiştir (Cevahir, 2011).

- Çınar ve Silahtaroğlu, yaptıkları çalışmada, 301 kişi ile yapılan müşteri memnuniyet anketinin sonuçlarını incelemiş, memnuniyet ya da memnuniyetsizlik öykülerinin altında yatan gizli örüntü ve kuralları ortaya çıkarmışlardır (Çınar vd., 211).
- Boland ve arkadaşları, veri madenciliği tekniklerini kullanarak gelecekteki su ihtiyaç tahminlerini açıklamışlardır (Boland vd.,1981).
- Çalış, yaptığı yüksek lisans tez çalışmasında, veri madenciliği yöntemlerinden kümeleme ve sınıflandırma algoritmalarını kullanarak bireysel kredi müşterilerinin ödeme performanslarına göre profillerini çıkarmış ve gelecekteki kredi müşterileri için kurallar önermiştir (Çalış, 2013).
- Akçay, yaptığı yüksek lisans tez çalışmasında Dünya Değerler Anketi verilerini incelemiş, Türkiye ve seçilmiş ülkelerle ilgili anket sorularını ve bunlara verilen cevapları birliktelik kuralları apriori algoritması ile belirlemiştir. Çalışmanın sonucunda hem Türkiye'nin mevcut durumunu hem de seçilmiş ülkelerdeki yerini ortaya çıkarmıştır (Akçay, 2013).
- Yurtay ve arkadaşları, 2014 yılında yaptıkları çalışmada sınıflandırma, kümeleme ve birliktelik analizi tekniklerini kullanarak Sakarya ilinde çıkan yangınları önleyici stratejiler geliştirmişlerdir (Yurtay vd., 2014).
- Ekmann, veri madenciliği tekniklerini kullanarak maden işletmesinde kömür depolama ve taşıma maliyetlerini azaltıcı çözüm stratejileri önermiştir (Ekmann, 2014).

1.5. Veri Madenciliğinin Faydaları

Veri madenciliği uygulamaları özellikle yoğun rakipliğin yaşandığı pazarlarda bir takım rolleri üstlenerek firmaların konumlarını güçlendirmede faydalar sağlayabilmektedir. Veri madenciliği uygulamalarında ulaşılabilecek olası yararlar bazı örnekler şu şekilde sıralanabilir (Çıngı,2010:3;Şaylan,2013:6):

- Mevcut müşterilerin firma tarafından daha iyi tanınmasına yardımcı olarak müşteri davranışlarının anlaşılabilmesi,
- Müşterileri kazanmak için katlanılan masrafları düşürerek maliyetlerin azaltılması,

- Büyük gelir sağlayacak mevcut alıcıların saptanarak onlara özel pazarlama kampanyalarının oluşturulması,
- Ürün veya hizmette hangi özelliklerin müşteri memnuniyetini etkilediğinin ortaya çıkarabilmesi,
- Sigortacılık, finans, bankacılık sektörlerinde geçmiş veriler ile mevcut müşterilerin performanslarının incelenerek risk yönetim politikalarının oluşturulması,
- Araştırma geliştirme çalışmalarında kolaylık sağlanabilmesi,
- Firmaların satış politikalarının oluşturulması,
- Finans ve borsa kuruluşlarında stok fiyat tahminlerinin yapılabilmesi,
- Firmaların ürün ve hizmetleri hakkında müşterilerini bilgilendirebilmesi,
- Laboratuvar ve bilgisayar ortamında elde edilen yüksek miktarda bilimsel verilerin anlamlandırılması,
- Web sunucusu veya online servis kullanıcılarının erişim desenlerinin analiz ve keşfinin yapılabilmesi.

1.6. Veri Madenciliği Uygulama Alanları

Günümüzde veri madenciliği, verilerin yaygın kullanıldığı birçok farklı alanda uygulanma imkânı bulabilmektedir. Veri madenciliğinin kullanım alanlarından bazıları şöyledir:

- Pazarlama
- Bankacılık ve Finans
- Sigortacılık
- Borsa
- Endüstri
- Sağlık ve İlaç
- Eğitim
- Telekomünikasyon
- Elektronik Ticaret

Veri madenciliğinin genel olarak kullanıldığı sektörler ve uygulama alanları şu şekilde özetlenebilir (Özkan,2008:38; Aydın,2006:392, Ay,2009:14;

Çalışkan, 2006:6; Çankırı vd., 2009:148; Karagöz, 2007:17; Baykasoğlu, 2005:8;Park,2001:205) :

- **Pazarlama**

- Tüketicilerin satın alma eğilimlerinin belirlenmesi,
- Müşterilerin karakteristik nitelikleri arasındaki ilişkilerin belirlenmesi,
- Var olan alıcıların elde tutulması ve olası alıcıların kazanılması için pazarlama yönetim kararlarının belirlenmesi,
- Market sepet analizi,
- Kazanç tahminleri,
- Tüketici ihtiyaçlarının belirlenmesi,
- Çeşitli pazarlama kampanyalarını belirleme,
- Mağaza yerleşimi,
- Bir ürün veya hizmet satın alan müşteriye bir başka ürünü kendisi talep etmeden tavsiye edip satmanın amaçlandığı çapraz satış önerileri,

- **Bankacılık ve Finans**

- Müşterilerin kredi kartı ve kredi ihtiyaçlarının değerlendirilmesi,
- Kredi kartı limitlerini harcama oranlarına göre müşteri profillerinin tespit edilmesi,
- Kredi kartı hilekârlıklarının belirlenmesi,
- Döviz fiyatlarındaki değişikliklerin önceden tahmin edilmesi,
- Risk analizleri ve yönetimi,

- **Sigortacılık**

- Poliçelerini yenileyecek veya yenilemeyecek alıcıların önceden tahmin edilmesi,
- Sigorta hilekârlıklarının belirlenmesi,
- Riskli müşterileri gruplarının tespit edilmesi,

- **Borsa**

- Genel piyasa analizleri,
- Hisse senedi fiyat tahmini,
- Alım-satım stratejilerinin tespit edilmesi,

- **Endüstri**
 - Lojistik uygulamaları,
 - Etkin kaynak kullanımı,
 - Ürün hataları, müşteri memnuniyet oranları gibi çıktı değişkenlerindeki sapmaların nedenlerinin belirlenmesi,
- **Sağlık ve İlaç**
 - Test sonuçlarının tahmini,
 - Yeni ilaç türlerinin keşfi,
 - Tıbbi teşhisler,
 - Tedavi süreçlerinin belirlenmesi,
- **Eğitim**
 - Öğrenci davranışlarının öngörülmesi,
 - Öğrenci başarı ve başarısızlık nedenlerinin tespiti,
 - Öğrencilerin ders seçme eğilimlerinin belirlenmesi,
- **Telekomünikasyon**
 - Kalite ve kontrol analizleri,
 - Telefon ve internet hatlarının kullanılma yoğunluklarının tahminleri,
- **Elektronik Ticaret**
 - Web sayfalarını kullanan ziyaretçilerin sayfa içindeki davranışlarının analiz edilmesi,
 - Ziyaretçilerin profillerinin belirlenmesi,
 - İnternet alışveriş siteleri kullanıcılarının satın alma alışkanlıklarının belirlenmesi.

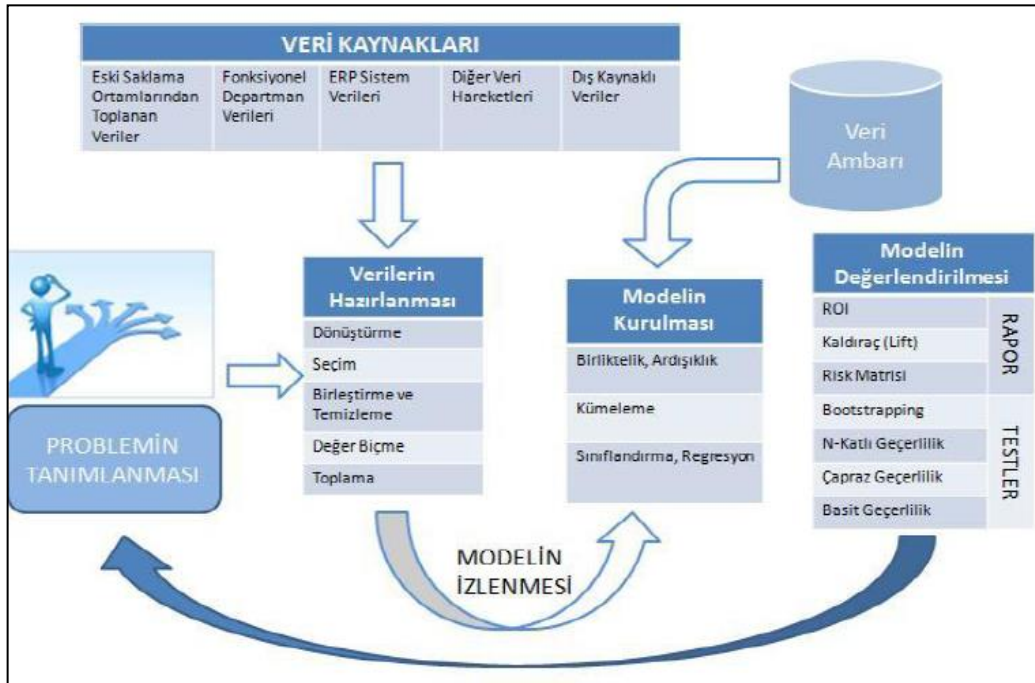
1.7. Veri Madenciliği Uygulama Süreci

Veri madenciliği bir analiz ve araştırma sürecidir. Veri madenciliği süreçlerinden en yaygın olarak kullanılanı, veri madenciliği araçlarını satan firmaların ve veri madenciliği uygulamalarını işletme faaliyetlerine uyarlayan şirketler birliği tarafından geliştirilen Çapraz Endüstri Veri Madenciliği Süreci (CRISP-DM)'dir (Fernandez vd., 2002:787). Bu sürecin geliştiricilerini Daimler-Chrysler (sonradan Daimler-Benz AG, Almanya), NCR Sistem Mühendisliği Kopenhag (ABD-Danimarka), SPSS Inc. (İngiltere) ve OHRA Verzegeeringen en

Bank Groep B.V (Hollanda) isimli firmalar oluşturmaktadır. (Chapman vd.,2004:13).

Veri madenciliği sürecinin temel aşamaları aşağıdaki gibidir (Shearer,2000:13):

- Problemin tanımlanması
- Verilerin hazırlanması
- Modelin kurulması ve değerlendirilmesi
- Modelin kullanılması
- Modelin izlenmesi



Şekil 4: Veri Madenciliği Süreci

(Kaynak: Akpınar, H., "Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği", İstanbul Üniversitesi İşletme Fakültesi Dergisi, 2000, Cilt:29, Sayı: 1, ss. 12)

1.7.1. Problemin Tanımlanması

Problemin tanımlanması aşaması, veri madenciliği sürecinin temel aşamasıdır. Bu aşama, gerçekleşmeden kullanılan algoritma ne kadar etkin olursa olsun sonuçlar istenilen güven seviyesinde gerçekleşmeyecektir. Bu alt yapıdan yoksun bir veri madenciliği süreci verileri doğru hazırlama ve sonuçları doğru değerlendirme konusunda sorunlar içerecektir.

Veri madenciliği çalışmalarında başarılı olabilmek için öncelikle kurumun hedeflerinin belirlenmesi gerekir. Hedef, problem üzerine odaklanmış, açık bir dille ifade edilmiş olmalı, elde edilecek sonuçların başarı düzeyinin nasıl ölçüleceği tanımlanmalıdır (Akpınar,2000:8).

Veri madenciliği yöntemini kullanacak bir perakende firması, yapacağı pazarlama kampanyasına cevap verme olasılığı yüksek müşterileri belirlemek, üreteceği yeni ürün için hedef kitle belirlemek, ürün için yeni pazarlama stratejileri geliştirerek markanın pazar payı oranını arttırmak gibi hedefler belirleyebilir. Belirlenecek hedeflerin, firmanın daha çok ürün satması, daha çok kar etmesi gibi genel geçer hedefler değil, daha belirgin bir problem üzerine yoğunlaşmış olması gerekir. Bir kütüphane, ulusal veri tabanlarından birini satın almak istediğinde hangisini satın alması gerektiğini belirlemek için üyelerinin yapmış olduğu taramaları analiz etmeye karar verdiğinde, bu problemi açıkça tanımlaması gerekir. Yani problem, üyelerin en çok arama yaptıkları konuyu bulmak ve bu konuyla ilgili ulusal veri tabanı satın almaktır, şeklinde açıkça belirlemelidir (Akçay,2014:41).

Temel olarak bu aşamada aşağıdaki faaliyetlerin gerçekleşmesi beklenmektedir (Olson vd.,2008:9):

- İşe yönelik hedeflerin belirlenmesi,
- Mevcut durumun değerlendirilmesi,
- Veri madenciliği hedeflerinin belirlenmesi,
- Proje planının yapılması.

1.7.2.Verilerin Hazırlanması

Veri madenciliği algoritmaları düşük kalitedeki veri seti üzerinde çalışırlarsa yüksek kalitede bilgiye ulaşmakta başarısız olurlar. Veri madenciliğinin en önemli aşamalarından biri olan verinin hazırlanması aşamasında, veri madenciliği uygulamasında kullanılacak olan verinin kalitesinin iyileştirilmesine çalışılmaktadır (Freitas,2002:65). Bu aşama veri madenciliği süreci içerisinde en çok zaman alan (%60-%75) aşama olmasına rağmen başarılı bir veri madenciliği uygulaması için en önemli basamaktır (Arlene,2005:9).

Verinin hazırlanması aşaması, başlangıçtaki ham veri setinden modelleme veya temel analiz çalışmalarında kullanılacak son veri setini oluşturma faaliyetlerinin tümünü içermektedir (Shearer,2000:16).

Verilerin hazırlanması aşaması kendi içerisinde toplama, değer biçme, birleştirme ve temizleme, seçim, dönüştürme adımlarından meydana gelmektedir (Çil,2010:6).

1.7.2.1.Toplama

Veri toplama; tanımlanan problem için gerekli olduğu düşünülen verilerin ve bu verilerin toplanacağı veri kaynaklarının belirlendiği adımdır (Döşlü,2008:18). Veri setlerinin toparlanmasında analist kurumun kendisine ait bilgi kaynaklarını kullanabileceği gibi kendisine ait bilgi kaynağından hariç bankaların riskli müşteri dizelgesi, müşteri profilleri, ağ günlükleri, medya paylaşımları, fotoğraf, bloglar gibi veri tabanlarından veya verileri satışa sunan kurumların veri seti kaynaklarından da faydalanabilmektedir.

1.7.2.2.Değer Biçme

Veri madenciliğinde kullanılacak verilerin farklı kaynaklardan toplanması, veri uyumsuzluklarına neden olacaktır. Bu uyumsuzlukların en önemlileri; farklı zamanlara ait olmaları, kodlama farklılıkları ve farklı ölçü birimleridir (Cevahir,2011:20).

Bu nedenlerle, iyi sonuç alınacak modeller ancak iyi verilerin üzerine kurulabileceği için, toplanan verilerin ne ölçüde uyumlu bu adımda incelerek değerlendirilmelidir (Akpınar,2000:9).

1.7.2.3.Birleştirme ve Temizleme

Bu adım, farklı kaynaklardan toplanan verilerin format, içerik olarak uygun hale getirilmesini, aynı nesne hakkında birden fazla veri kaynağında bulunan verilerin bir veri tabanında birleştirilmesini içermektedir (Berry,2000:52). Bu durum, veri madenciliği sürecinin hızının ve doğruluğunun gelişmesini sağlayarak uygulamanın daha başarılı sonuçlar vermesine katkıda bulunacaktır.

Veri birleřtirme sırasında gerekleřtirilecek olan veri temizleme iřleminin amacı, veriler ierisindeki uygun olmayan veya hatalı girilmiř verileri ayıklamaktır (Tang vd., 2005).

1.7.2.4.Seim

Bu adımda kurulacak modele baėlı olarak veri seimi yapılır. rneėin; tahmin edici bir model iin, bu adım baėımlı ve baėımsız deėiřkenlerin ve modelin eėitiminde kullanılacak veri kumesinin seilmesi anlamını tařımaktadır.

Sıra numarası, kimlik numarası gibi anlamlı olmayan ve diėer deėiřkenlerin modeldeki aėırlıėının azalmasına da neden olabilecek deėiřkenlerin modele girmemesi gerekmektedir. Bazı veri madenciliėi programları konu ile ilgisi olmayan bu tip deėiřkenleri otomatik olarak elese de, pratikte bu iřlemin kullanılan yazılıma bırakılmaması daha akılcı olacaktır. Verilerin grselleřtirilmesine imkan saėlayan grafik aralar ve bunların sunduėu iliřkiler, baėımsız deėiřkenlerin seilmesinde nemli yararlar saėlayabilir (etin,2009:15).

Modelde kullanılan veri tabanının ok byk olması durumunda tesadfiliėi bozmayacak řekilde rnekleme yapılması uygun olabilir. Gnmzde hesaplama olanakları ne kadar geliřmiř olursa olsun, ok byk veri tabanları zerinde ok sayıda modelin denenmesi ok zaman alması nedeni ile mmkn olamamaktadır. Bu sebeple tm veri tabanını kullanarak birkaç model denemek yerine tesadfi olarak rneklenmiř bir veri tabanı parası zerinde birok modelin denenmesi ve bunlar arasında en gvenilir, en gl modelin seilmesi daha uygun olacaktır.

1.7.2.5.Dnřtrme

Veri dnřtrme amacı, zmlenmesi iin kullanılması dřnlen verilere iliřkin deėiřkenleri uygun formlara dnřtrmektir. rneėin; tıbbi veri kmelerinde sıklıkla rastlanan hasta aėırlık ve boy deėerleri veri dnřtrme yoluyla Vcut İndeksi gibi bir orana evrilebilmektedir. Bu oran yardımıyla hastaların kilolu olup olmadıėının tespiti daha iyi yapılabilir (Kantardzic,2003:16).

Kredi tahmini için geliştirilen bir modelde ise, borç/gelir gibi önceden hesaplanmış bir oran yerine, ayrı ayrı borç ve gelir verilerinin kullanılması, veri dönüştürme ile tercih edilebilmektedir. Ayrıca modelde kullanılan algoritma, verilerin gösteriminde önemli rol oynamaktadır. Örnek olarak, bir uygulamada bir yapay sinir ağı algoritmasının kullanılması durumunda kategorik değişken değerlerinin evet/hayır olması; bir karar ağacı algoritmasının kullanılması durumunda ise örneğin gelir değişken değerlerinin yüksek/orta/düşük olarak gruplanmış olması modelin etkinliğini arttıracaktır.

1.7.3. Modelin Kurulması ve Değerlendirilmesi

Tanımlanan problem için en uygun modelin bulunması ve bu modelleme tekniğinin kurulması aşamasıdır. İyi kurulan bir model analiz sonucunda elde edilecek sonuçların kalitesini belirleyecektir. Eğer uygun model kurulamazsa veriler arasında bulunabilecek bağlantılar doğru çözümlenemez ve önemli örüntüler saptanamaz. Böylece veri madenciliğinden başarılı sonuç elde edilemez (Michael vd.,1998:25).

Uygun modelin bulunabilmesi, olabildiğince çok sayıda modelin kurularak denenmesi ile mümkündür. Bu nedenle bu aşama, en uygun olduğu düşünülen modele ulaşıncaya kadar tekrarlanan bir süreçtir. Süreç tekrarlandıkça performans iyileşmekte, sonuçlar daha güvenilir olmaktadır (Arlene vd.,2013).

Model kuruluş süreci denetimli (Supervised) ve denetimsiz (Unsupervised) öğrenimin kullanıldığı modellere göre farklılık göstermektedir (Witten vd.,2000:395).

Örnekten öğrenme olarak da isimlendirilen denetimli öğrenimde, bir denetçi tarafından ilgili sınıflar önceden belirlenen bir kıstaslara göre ayrılarak, her sınıf için çeşitli örnekler verilir. Sistemin amacı verilen örneklerden hareket ederek her bir sınıfa ilişkin özelliklerin bulunması ve bu özelliklerin kural cümleleri ile ifade edilmesidir.

Öğrenme süreci tamamlandığında, tanımlanan kural cümleleri verilen yeni örneklere uygulanır ve yeni örneklerin hangi sınıfa ait olduğu kurulan model tarafından belirlenir.

Denetimsiz öğrenmede, kümeleme analizinde olduğu gibi ilgili örneklerin gözlenmesi ve bu örneklerin özellikleri arasındaki benzerliklerden hareket ederek sınıfların tanımlanması amaçlanmaktadır.

Denetimli öğrenimde seçilen algoritmaya uygun olarak ilgili veriler hazırlandıktan sonra, ilk aşamada verinin bir kısmı modelin öğrenimi, diğer kısmı ise modelin geçerliliğinin test edilmesi için ayrılır. Modelin öğrenimi öğrenim kümesi kullanılarak gerçekleştirildikten sonra, test kümesi ile modelin doğruluk derecesi (Accuracy) belirlenir.

Bir modelin doğruluğunun test edilmesinde kullanılan en basit yöntem basit geçerlilik (simple validation) testidir. Bu yöntemde verilerin bir kısmı test verisi olarak ayrılır, kalan kısım üzerinde modelin öğrenimi gerçekleştirildikten sonra ayrılan kısım üzerinde test işlemi yapılır. “Bir sınıflama modelinde yanlış olarak sınıflanan olay sayısının, tüm olay sayısına bölünmesi ile hata oranı, doğru olarak sınıflanan olay sayısının tüm olay sayısına bölünmesi ile doğruluk oranı hesaplanır (Doğruluk Oranı=1-Hata Oranı) (Akpınar,2000:11).

Sınırlı miktarda veriye sahip olunması durumunda, kullanılacak diğer bir yöntem çapraz geçerlilik (Cross Validation) testidir. Bu yöntemde veri kümesi tesadüfî olarak iki eşit parçaya ayrılır. İlk aşamada x parçası üzerinde model eğitimi ve y parçası üzerinde test işlemi; ikinci aşamada ise y parçası üzerinde model eğitimi ve x parçası üzerinde test işlemi yapılarak elde edilen hata oranlarının ortalaması kullanılır.

Bir kaç bin veya daha az satırdan meydana gelen küçük veri tabanlarında, verilerin n gruba ayrıldığı n katlı çapraz geçerlilik (N-Fold Cross Validation) testi tercih edilebilir. Verilerin örneğin 10 gruba ayrıldığı bu yöntemde, ilk aşamada birinci grup test, diğer gruplar öğrenim için kullanılır. Bu süreç her defasında bir grubun test, diğer grupların öğrenim amaçlı kullanılması ile sürdürülür. Sonuçta elde edilen hata oranının ortalaması, kurulan modelin tahmini hata oranı olacaktır.

Küçük veri kümeleri için modelin hata düzeyinin tahmininde kullanılan bir başka teknik ise Bootstrapping'dir. Çapraz geçerlilikte olduğu gibi model bütün veri kümesi üzerine kurulur. Daha sonra en az 200, bazen binin üzerinde olmak

üzere çok fazla sayıda öğrenim kümesi tekrarlı örneklemelerle veri kümesinden oluşturularak hata oranı hesaplanır.

Model kuruluşu çalışmalarının sonucuna bağlı olarak, aynı teknikle farklı parametrelerin kullanıldığı veya başka algoritma ve araçların denendiği değişik modeller kurulabilir. Model kuruluş çalışmalarına başlamadan önce, imkânsız olmasa da hangi tekniğin en uygun olduğuna karar verebilmek güçtür. Bu nedenle farklı modeller kurarak, doğruluk derecelerine göre en uygun modeli bulmak üzere sayısız deneme yapılmasında yarar bulunmaktadır.

Özellikle sınıflama problemleri için kurulan modellerin doğruluk derecelerinin değerlendirilmesinde basit ancak faydalı bir araç olan risk (yakınsallık) matrisi kullanılmaktadır. Aşağıda bir örneği görülen bu matriste sütunlarda fiili, satırlarda ise tahmini sınıflama değerleri yer almaktadır. Örneğin fiilen B sınıfına ait olması gereken 46 elemanın, kurulan model tarafından 2'sinin A, 38'inin B, 6'sının ise C olarak sınıflandırıldığı matriste kolayca görülebilmektedir (Cevahir,2011:27).

Tablo 1: Risk Matrisi

Tahmin Edilmiş Sınıf	Gerçek sınıf		
	A Sınıfı	B Sınıfı	C Sınıfı
A Sınıfı	45	2	3
B Sınıfı	10	38	2
C Sınıfı	4	6	40

Önemli diğer bir değerlendirme kriteri modelin anlaşılabilirliğidir. Bazı uygulamalarda doğruluk oranlarındaki küçük artışlar çok önemli olsa da, birçok işletme uygulamasında ilgili kararın niçin verildiğinin yorumlanabilmesi çok daha büyük önem taşıyabilir. Seyrek olarak yorumlanamayacak kadar karmaşık olsalar da, genel olarak karar ağacı ve kural temelli sistemler model tahmininin altında yatan nedenleri çok iyi ortaya koyabilmektedir.

Model tarafından önerilen uygulamadan elde edilecek kazancın bu uygulamanın gerçekleştirilmesi için katlanılacak maliyete bölünmesi ile elde edilecek olan yatırımın geri dönüş oranı kurulan modelin değerinin belirlenmesinde kullanılan diğer bir ölçüdür (Karagöz,2007:54).

Kurulan modelin doğruluk derecesi ne denli yüksek olursa olsun, gerçek dünyayı tam anlamı ile modellediğini garanti edebilmek mümkün değildir. Yapılan testler sonucunda geçerli bir modelin doğru olmamasındaki başlıca nedenler, model kuruluşunda kabul edilen varsayımlar ve modelde kullanılan verilerin doğru olmamasıdır. Örneğin modelin kurulması sırasında varsayılan enflasyon oranının zaman içerisinde değişmesi, bireyin satın alma davranışını belirgin olarak etkileyecektir.

1.7.4.Modelin Kullanılması

Kurulan ve geçerliliği kabul edilen modelin kullanılması aşaması, modelin zaman içerisinde izlenip ortaya çıkan değişikliklerin yeniden modele yansıtılması sürecidir (Akçay,2014:44). Kurulan ve geçerliliği kabul edilen model doğrudan bir uygulama olabileceği gibi, bir başka uygulamanın alt parçası olarak kullanılabilir. Kurulan modeller risk analizi, kredi değerlendirme, dolandırıcılık tespiti gibi işletme uygulamalarında doğrudan kullanılabilen gibi, promosyon planlaması simülasyonuna entegre edilebilir veya tahmin edilen envanter düzeyleri yeniden sipariş noktasının altına düştüğünde, otomatik olarak sipariş verilmesini sağlayacak bir uygulamanın içine gömülebilir.

1.7.5.Modelin İzlenmesi

Veri madenciliği sürecinin son aşaması; geçerliliği onaylanan ve uygulanan modelin izlenmesidir. Bu aşamada uygulanan modelin kurumun hedeflerine ne kadar ulaştığı, belirlenen probleme ne kadar yanıt verdiği ölçülür. Sürecin gözden geçirilmesiyle modelin doğru kurulup kurulmadığı, modelleme tekniklerinden ne ölçüde verim alındığı, gelecekte yapılacak analizlerde bunların kullanılabilirliği değerlendirilir. Değerlendirme sonuçlarına göre çalışmanın nasıl devam edeceğine, sürecin tekrar edilip edilmeyeceğine bu aşamada karar verilir (Chapman,2010:31). Model sonuçlarının izlenmesinde, modelin kurulması

öncesinde tahmin edilen ve sonrasında gözlemlenen değerler arasındaki farklılığı gösteren grafiklerin kullanılması yararlı bir yöntemdir (Akpınar,2000:14).

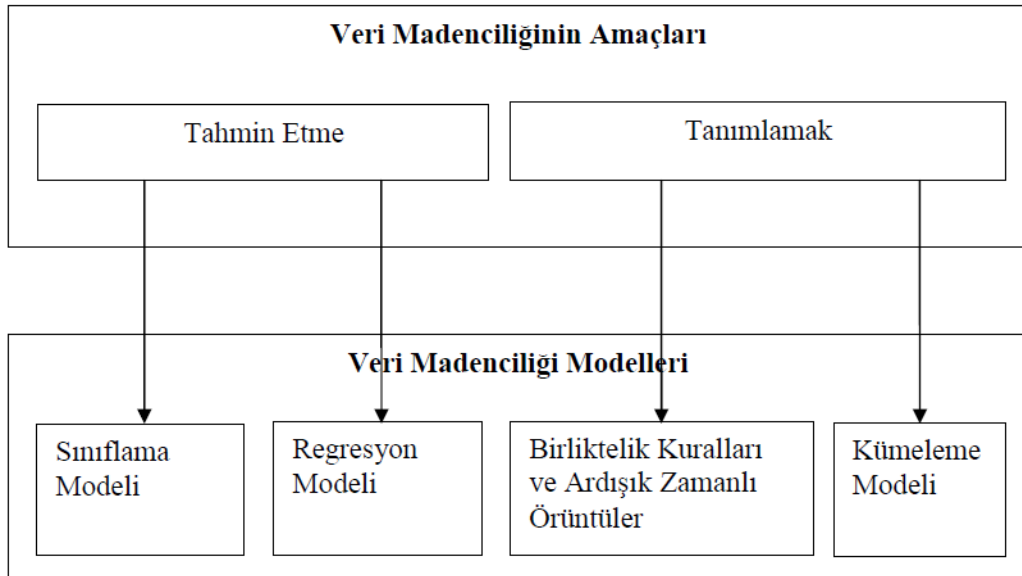
Veri madenciliği süreci başarılı olursa, keşfedilen yeni bilgiler kurumun karar verme sürecini iyileştirmede kullanılabilir (Freitas,2002:1).

1.8.Verit Madenciliğinde Kullanılan Modeller ve Teknikler

1.8.1.Verit Madenciliği Modelleri

Verit madenciliği temel olarak, büyük ölçekli verilerin yazılım teknikleri kullanılarak analiz edilmesini sağlayan süreçtir. Bu analiz sürecinde, kullanılabilir pek çok veri madenciliği modeli ve teknikleri söz konusudur. Kullanılacak olan model veriye ve araştırmanın ulaşmak istediği sonuca göre değişiklik gösterebilir.

Verit madenciliğinde yapılan analizlerin amacı genellikle, bir sonucu tahmin etmek ya da belirli bir sonucu tanımlamaktır. Bu nedenle veri madenciliği modelleri, tahmin edici ve tanımlayıcı modeller olmak üzere iki ana başlık altında incelenmektedir (Çetin,2009:22).



Şekil 5: Tahmin Edici ve Tanımlayıcı Modeller

(Kaynak: Çetin, M., 2009, "Bir Üretim İşletmesinde Verit Madenciliği Uygulaması, Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, s.22)

1.8.1.1.Tahmin Edici Modeller

Modellerde, sonuçları bilinen geçmiş verilerden hareket edilerek bir model geliştirilmesi ve kurulan bu modelden yararlanılarak sonuçları bilinmeyen veri kümeleri için sonuç değerlerin tahmin edilmesi amaçlanmaktadır (Kantardzic,2003:1). Bu modelde öngörme, yargıya ulaşma, sınıflandırma gibi işlevler ön plandadır. Örneğin; bir üniversite kütüphanesi, kendisinden daha önce herhangi bir materyal almış bir üyesiyle ilgili bütün verilere sahiptir. Burada üyenin özellikleri, daha önce kaç tane materyal, kitap, dergi aldığı ve aldıklarını ne zaman getirdiği ile ilgili veriler bilinenlerdir. Bununla birlikte, yeni alacağı materyali ne kadar sürede getireceği ise bilinmeyen veridir. Bu verilere uygun olarak kurulan yöntemle üyenin almış olduğu ödünç materyali ne zaman getireceği tahmin edilebilir.

Bankalar bu modeli kredilendirme sistemlerinde sıklıkla kullanırlar. Bir banka, müşterisine önceki dönemlerde vermiş olduğu kredinin geri ödenip ödenmediğini, veri tabanındaki verilerden yararlanarak bulabilir ve buna göre yeni kredinin tutarı veya vadesi hakkında tahminde bulunabilir (Özekes,2003:67).

Tahmin edici modellerde kullanılan başlıca yöntemler; sınıflama ve regresyon yöntemleridir (Çıngı,2010:10).

1.8.1.1.1.Sınıflama

Veri tabanlarında işlemler değişken tanımlanarak yapılır. Değişken, bir niceliği veya ölçülebilir niteliği ifade etmek için kullanılan sembolik değerlerdir. Veri tabanlarında, belli bir veri tipine ait değerleri saklarlar. Bir sistemin işletilmesinde veya bir deneyin durumu doğrultusunda değişebilecek değerlerin yerlerini tutan varsayımlardır. Örneğin, bir kütüphane veri tabanında üyenin adı, soyadı şeklinde girilen veriler, sistem tabanında “üye” gibi bir değişkenle tanımlanır. Bu değişken, kütüphaneyi kullanan herhangi bir üye olarak değerlendirilebilir. Değişkenler; bağımlı değişkenler ve bağımsız değişkenler olarak iki şekilde sınıflandırılmaktadır. Bağımsız değişken, araştırmacının atadığı değişkendir. Bağımlı değişken ise araştırmacının atadığı değişkene, yani bağımsız değişkene bağlı olarak değişen değişkendir. Bir analizde bağımsız değişkenin bağımlı değişken üzerindeki etkisi incelenir (Akçay,2014:64). Örneğin; “gelir” bir

bağımsız değişkendir, “satın alma” ise bağımlı değişkendir ve aldığı değer “gelir”in aldığı değere göre değişir. “gelir” açıklayıcı, tahmin edici, “satın alma” ise açıklanan, tahmin edilen değişkendir.

Veri madenciliği uygulamalarında sıklıkla kullanılan bir yöntem olan sınıflandırma (Park,2001:206), eldeki mevcut verileri, önceden belirlenen bir özelliğe göre sınıflara ayırmak ve yeni eklenecek verilerin hangi sınıfa dahil olacağını tayin etme işlemidir (Gürgen,2008:8).

Sınıflandırma, bir ürün veya hizmetin ve alıcı özelliklerinin eşlenmesi için kullanılabilir. Bu durumda alıcılar için en uygun hizmet, hizmet içinde en uygun alıcı nitelikleri çıkartılabilir. Örneğin; bir GSM operatörü belli bir gün içerisinde belirlenmiş bazı saatler arasında indirimli konuşma veya ücretsiz mesajlaşma kampanyaları başlatmak istediğinde, kampanyayı uygulayacağı saatler arasında en çok konuşan müşterilerini tespit etmek isteyebilir. Veri tabanında geçmişte müşterilerin konuşma ve mesajlaşma verilerinin mevcut olduğunu varsayalım; o müşterilerin ne kadar süre telefonla konuştuklarını ve mesajlaştıklarını gösteren alan bağımlı değişken olacaktır. Müşterilerin diğer nitelikleri ise (yaş, cinsiyet, gelir düzeyi, eğitim düzeyi gibi) bağımlı değişken üzerinde etkisi araştırılan bağımsız değişken olacaktır (Akbulut,2006:28). Söz konusu firmanın veri madenciliği uygulamaları sonucunda 20-24 yaş aralığındaki gençlerin o saatlerde daha fazla telefonla konuştuğunu ve mesajlaştığı bulgularını tespit ettiğini varsayalım. Firma başlattığı bu kampanyanın tutundurma çalışmalarını bu kitle üzerinde yoğunlaştırarak bu kampanyadan daha hızlı ve olumlu sonuçlar alacaktır.

Bu yöntemin en önemli işlevi, sınıflandırma sonrasında her kategoride yer alan kayıtların, alanların, nesnelere, kişilerin, kurumların özelliklerini ortaya çıkarmaktır. Bu sınıflar çeşitli müşteri grupları için oluşturulmuş ise, burada öncelikli amaç her bir grubun özelliğini ortaya çıkarmak, profilini belirlemek olmalıdır. Örneğin, e-ticarette internet müşterileri yaptıkları alışveriş sıklıklarına göre, “seyrek kullanıcı”, “orta sıklıkta kullanıcı” ve “sık kullanıcı” olarak sınıflandırılabilir. Müşteriler bu şekilde gruplandıktan sonra amaç ise her bir grubun özelliklerini analiz etmek, profilini ortaya çıkarmak ve grupların tutum ve davranışlarını değiştirecek yöntemler geliştirmektir.

Bir halk eğitim merkezi, merkeze ilgi çekebilmek ve merkezin daha fazla kullanılmasını sağlamak için kursiyerlerine merkezde yeni açılan kurslarla ilgili mesajlar gönderebilir. Ancak her açılan kursun duyurusunun bütün kursiyerlere gönderilmesi ilgi çekmekten çok caydırıcı olabilir. Ancak kursiyerler, geçmiş kurs kayıtlarından oluşan verilerin analiziyle belirli sınıflara ayrılırsa, sadece o sınıfla ilgili kategorideki kursların duyurusunun mesaj olarak gönderilmesi kursiyerlerin dikkatini çekebilir. Halk eğitim merkezlerinin kullanımının gittikçe azaldığı günümüzde kursiyerler, bu tür yöntemler kullanılarak merkezlere çekilebilir.

1.8.1.1.2.Regresyon

Regresyon analizi, geçmişteki değerleri temel alarak gelecekteki değerleri tahmin etmek için kullanılan tahmin edici modeller sınıfında yer alan yöntemdir. Başka bir anlatımla, bağımlı bir değişkenin, bir veya birden fazla bağımsız değişkenle kurduğu ilişkinin matematiksel bir fonksiyon biçiminde yazılması ve bu fonksiyon yardımıyla bağımlı değişkenin ulaşabileceği değerlerin tahmin edilmesidir (Orhunbilge,1996:9).

Regresyon analizi, iki değişken arasındaki ilişkiyi bulmak, ilişki varsa bu ilişkinin şiddetini belirlemek, değişkenler arasındaki ilişkinin türünü belirlemek, ileriye dönük değerleri tahmin etmek gibi konularda kullanılmaktadır (Argüden vd.,2008:50). Örneğin; bir bankaya kredi başvurusunda banka müşterilerinden “kredi kartı borçlarını zamanında ödeyen, ev sahibi olan, işinde 3 yıldır fazladır çalışan, evli bir kişinin kredi skoru 925 iken, kredi kartı borçlarının geç ödemesi bir ayı geçmemiş, işe yeni başlamış bir kişinin ise kredi skoru 750’dir” sonucu bir regresyon ilişkisidir. Bu tür bir analizle müşterinin kredi ödemelerini zamanında ödeyip ödemeyeceği ilişkisi tahmin edilmeye çalışılır.

Mevcut verilerden hareket ederek geleceğin tahmin edilmesinde faydalanılan ve veri madenciliği yöntemleri içerisinde en yaygın kullanıma sahip olan sınıflama ve regresyon modelleri arasındaki temel fark, tahmin edilen bağımlı değişkenin kategorik (belli özelliklere göre gruplanmış) veya süreklilik (sayısal) gösteren bir değere sahip olmasıdır. Sınıflandırma, belli özelliklere göre gruplanmış kategorik değerleri tahmin ederken, regresyon süreklilik (sayısal) gösteren değerlerin tahmin edilmesinde kullanılır. Örneğin; sınıflandırma modeli

banka kredi uygulamalarının güvenli veya riskli olmalarını sınıflamak amacıyla kurulurken, regresyon modeli gelir ve mesleği verilen potansiyel müşterilerin elektronik ürünleri alırken yapacakları harcamaları tahmin etmek için kurulabilir (Özekes vd.,2002:17).

Sınıflama ve regresyon modellerinde kullanılan başlıca teknikler;

- Karar Ağaçları,
- Yapay Sinir Ağları,
- Genetik Algoritmalar,
- K-En Yakın Komşu,
- Bellek Tabanlı Yöntemler,
- Nalve-Bayes,
- Lojistik Regresyon

olarak sıralanabilir (Akpınar,2000:22).

1.8.1.2.Tanımlayıcı Modeller

Tanımlayıcı modellerin amacı, belirli bir hedefi tahmin etmek değildir. Bu modellerin amacı, karar vermeye yardım edebilecek mevcut veri setinde yer alan veriler arasındaki ilişkileri bulmaktır. Var olan verileri yorumlayarak davranış biçimleri ile ilgili tespitler yapmayı ve bu davranış biçimini gösteren alt veri setlerinin niteliklerini tanımlamayı hedefler (Argüden vd.,2008:39).

Perakende sektöründe, hangi ürünlerin birlikte satıldığı, hangi müşteri grubunun hangi zaman dilimlerinde bir hizmeti kullandığını ve buna göre en iyi müşteri grubunun hangisi olduğunun belirlenmesinde bu yöntem kullanılır (Moskovich vd.,2002:303).

Örneğin, A/B aralığında geliri ve arabası olan çocuklu aileler ile geliri A/B aralığından düşük çocuğu olmayan ailelerin satın alma davranışlarının birbirine benzerlik gösterdiğinin belirlenmesi tanımlayıcı modelle gerçekleşir. Bununla birlikte, kredi ödemeleri üzerinde araştırma yapan bir bankacı, bekar kişiler ile evli kişiler üzerinde yapılan kredi ödeme performanslarını gösteren bir analize yine tanımlayıcı modeller yöntemiyle ulaşabilir. Bu analizlere geleneksel istatistiksel verilerle ulaşılması olanaklı değildir.

Tanımlayıcı modellerde kullanılan başlıca yöntemler; kümeleme ve birliktelik kurallarıdır (Çingir,2010:13).

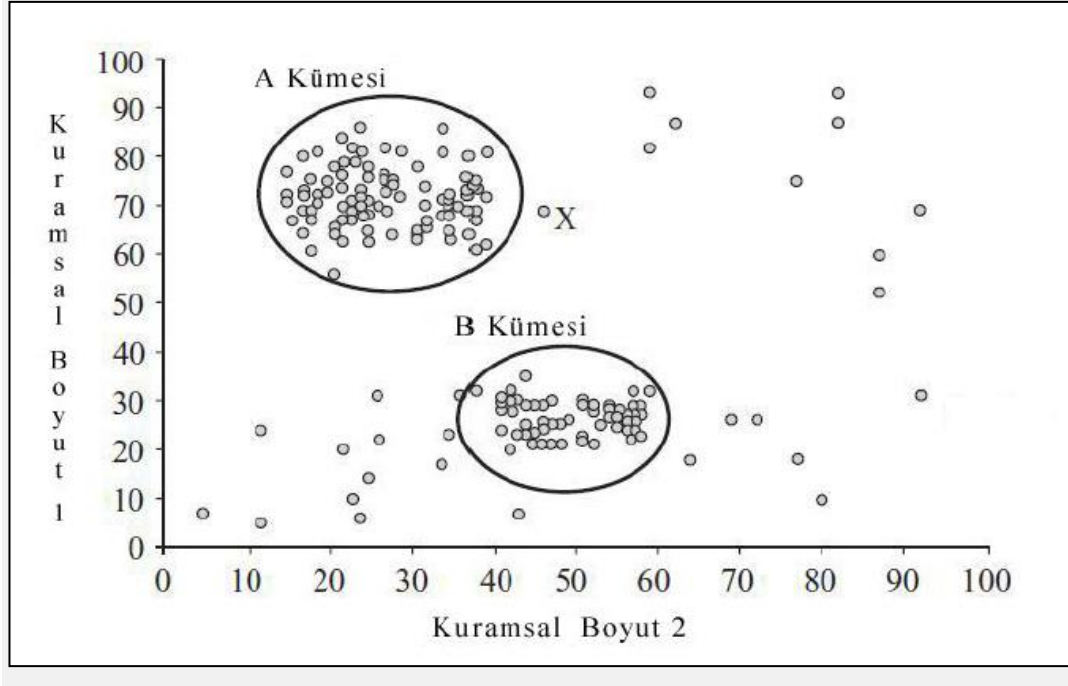
1.8.1.2.1.Kümeleme

Veri tabanlarında veriler çok büyük boyutlarda olabileceği gibi karmaşık bir yapıya da sahip olabilir. Bunlar üzerinde en iyi şekilde uygulanabilecek veri madenciliği yöntemlerinin bile başarılı olması, bu karmaşık verilerden anlamlı sonuçlar çıkarması zordur. Bu tür sorunları çözmekte izlenilecek yöntem genellikle verileri parçalara ayırmak, alt bölümlere ayrıştırmak ve öncelikle alt bölümlerden başlayarak çözüm üretmeye başlamaktır. Ancak bazı durumlarda veriler öyle dağılmışlardır ki verilerin nereden bölüneceğini ve hangi şekilde alt gruplara ayrılabilceğini kestirmek mümkün değildir. Bu yüzden dağınık verilerin nasıl bölüneceğini ve gruplara ayrılacağını belirlemek için kümeleme yöntemleri geliştirilmiştir.

Kümeleme yöntemi, veriyi birbirlerine benzeyen elemanlardan oluşan sınıflara (kümelere) ayırarak, heterojen bir veri grubundan, homojen alt veri grupları elde edilmesi işlemidir (Argüden vd.,2008:39). Diğer bir tanımla; büyük veri kümelerinin anlamlı şekilde parçalanarak alt kümelerine ayrıldığı ve homojen grupların bir arada ele alındığı bir süreçtir (Ohsuga,2005:121).

Kümeleme modellerinde amaç, küme üyelerinin birbirine çok benzediği, ancak özellikleri birbirinden çok farklı olan kümelerin bulunması ve veri tabanındaki kayıtların bu farklı kümelere bölünmesini sağlamaktır (Çetin,2009:27).

Örneğin; Türkiye'deki şehirlerin profilini çıkarmak üzere bir araştırma yapmak istiyoruz. Sanayiye dayalı bir gelir sistemi olan bir şehri, geliri tarıma dayalı olan şehirlerle mukayese etmenin ne kadar güvenilir sonuçlar verebileceği tartışılır. Aynı şekilde nüfusu yüz binlerle ölçülebilen şehirleri nüfusu milyonlar olan şehirlerle mukayese etmek de hatalıdır. Belirlediğimiz ölçütlere göre benzer özellik gösteren şehirler bir gruba toplanır ve kendi aralarında analiz yapılır. Söz gelimi Şırnak'ı İstanbul ile mukayese etmek yerine profili benzer olabilecek Bitlis, Siirt, Batman vb. şehirlerle mukayese etmek çok daha güvenilir sonuçlar elde etmemizi sağlayacaktır (Çingir,2010:13).



Şekil 6: Kümeleme

(Kaynak: Glenn J. Myatt, Making Sense of Data, US, John Wiley & Sons Publication, 2007, s. 110)

Birbirine benzer gibi gözükten sınıflama ile kümelemeyi birbirinden ayıran en önemli fark, kümeleme işleminin sınıflama işleminde olduğu gibi önceden belirlenmiş bir takım sınıflara göre bölme yapmamasıdır. Sınıflama yönteminde her bir veri, önceden sınıflandırılmış bir takım sınıflar üzerinde yapılan bir eğitim neticesinde ortaya çıkan bir modele göre önceden belirlenmiş olan bir sınıfa atanmaktadır. Kümeleme yönteminde ise, önceden tanımlanmış sınıflar ya da örnek sınıflar bulunmamaktadır. Verilerin kümelenmesi işlemi, analiz sırasında yer alan verilerin birbirlerine olan benzerliklerine göre yapılmaktadır. Veri tabanındaki verilerin hangi kümelere ayrılacağı, oluşan kümelerin hangi anlamı taşıdığı veya kümelemenin hangi değişkenin özelliklerine göre yapılacağı tamamen analizi yapan kişiye bağlıdır (Dolgun,2006:27).

Kümeleme analizi üç adımdan oluşan bir çözümleme sürecidir. Birinci aşamada veri girişi yapılır. Doğal olarak oluşan sınıflamalarla ilgili net bilgilerin bulunmadığı verilerin, incelenen değişkene ilişkin gözlem sonuç değerleri elde edilir. Böylelikle veri şeması oluşturulmuş olur. İkinci aşamada ise uygun kümeleme tekniği seçilir ve uygulanır. Kümeleme tekniğinin uygulanmasıyla

veriler kümelere ayrıştırılmış olur. Kümeleme sonuçlarının anlamlılığının yorumlandığı aşama ise analizin son aşamasını oluşturur (Mainly,1994:280).

Veri madenciliğinde kümeleme modellerinin uygulama alanları çok geniştir. Marketlerde farklı müşteri gruplarının keşfedilmesi ve bu grupların alışveriş birlikteliklerinin ortaya çıkarılması, biyolojide bitki ve hayvanlarda benzer genlerin sınıflandırılması, şehir planlanmasında evlerin tiplerine, değerlerine ve coğrafik konumlarına göre gruplara ayrılması, eğitimde kültürel ve eğitsel alt yapılaraya göre eğitim programlarının geliştirilmesi ve örnek öğrenme kalıpları geliştirilmesi gibi uygulamalar tipik kümeleme uygulamalarıdır. Kümeleme aynı zamanda web üzerinde e-ticaret bilgi keşfi için dokümanların sınıflanması için de kullanılabilir.

1.8.1.2.2.Birliktelik Kuralları

Veri madenciliğinin en yaygın kullanılan yöntemlerinden biri olan birliktelik kuralı, geçmiş verilerin analiz edilerek bu veriler içindeki birliktelik örüntülerinin tespiti ile geleceğe yönelik çalışmalar (pazarlama, iş yönetimi vb.) yapılmasını destekleyen bir yaklaşımdır (Özçakır vd.,2007:37). Örneğin; bir alışveriş sırasında veya birbirini izleyen alışverişlerde müşterinin hangi mal ve hizmetleri satın almaya eğilimli olduğunun belirlenmesi, o müşteriye daha fazla ürünün satılarak şirketin kar oranının artırılmasında önemli rol oynayabilir (Han vd.,2001:25).

Birliktelik kuralında, veriler arasındaki ilişkileri açıklamak için eğer-sonra ifadeleri kullanılır ve eğer <bazı şartlar sağlanıyorsa> sonra <bazı niteliklerin değerlerini tahmin et> şeklinde belirtilebilir. Eğer bölümü ile ilişkili durumlar öncül, sonra bölümü ile ilişkili durumlar sonuç olarak isimlendirilir. Kuralda öncül ve sonuç durumları a ve b simgeleri olarak ele alınırsa aralarındaki ilişki $a \Rightarrow b$ şeklinde sembolize edilebilir (Emel vd.,2005:32). Buna kot pantolon \Rightarrow kazak, makarna \Rightarrow ketçap, çay \Rightarrow şeker biçiminde örnekler verilebilir.

Birliktelik kuralları yöntemi; <x nesnesini alan bir kişinin, muhtemelen y nesnesini de alması> tipindeki kuralların tanımlanmasını amaçlamaktadır (Brin vd.,1997:265).

Örneğin; “bir müşteri market alışverişinde un satın alıyorsa aynı alışverişte unun yanında kabartma tozu satın alma olasılığı nedir?” sorusunun yanıtına ulaşan market yöneticileri, bu bilgi ışığında market raflarını düzenleyerek satışlarını arttırabilir. Eğer bu markette un ve kabartma tozunun beraber alınma oranı yüksekse, market yöneticileri un ve kabartma tozu ürünlerini yakın raflara koyarak, insörtlerde birlikte satılan ürünlerin birlikte görülmesini sağlayarak veya müşteriler için ideal ürün paketleri oluşturarak müşterilerin alışveriş eğilimlerini yönlendirebilir.

Birliktelik kuralı ile müşterilerin alışveriş davranışları keşfedilmeye çalışılır. Bir ürün satın alındığında eş zamanlı olarak gerçekleşen veya birbirini izleyen alışverişlerde müşterinin hangi mal veya hizmeti satın alma eğiliminde olduğu saptanarak, müşteriye daha fazla ürün ve hizmet sağlanılmaya çalışılır (Yen vd., 2006:650). Eğer müşteriye sunulacak herhangi bir ürün veya hizmet önerisinin müşteri tarafından kabul edilip edilmediğinin sonucu, müşterinin demografik özellikleriyle veya geçmişteki tutum ve davranışlarıyla ilişkilendirilebilirse pazarlama faaliyetleri bakımından önemli bir bilgi birikimine ulaşmış olur.

Örneğin, bankadan hizmet alan müşterilerin maaş hesabına sahip olmalarıyla kredi kartına sahip olmaları arasında kuvvetli bir örüntü bulunmuşsa, maaş hesabı bulunan diğer müşterilere kredi kartı konusunda bir öneri sunulabilir.

Birliktelik kuralı ile önceden tahmin edilebilir bir sonuca varılabileceği gibi daha önce tahmin edilmesi mümkün olmayan bir sonuca da ulaşılabilir.

Örneğin;

- Light kola ve lightbisküvi alan müşteriler, yüzde 80 olasılıkla light çikolatada satın alırlar.
- Hazır yemek alan kadın müşterilerin yüzde 70’i bakım ürünlerini de satın almaktadır.

Veri tabanının büyüklüğü ile birliktelik kuralının olumlu sonuç vermesi arasında doğru orantı vardır. Bu yüzden büyük veri tabanlarında birliktelik kurallarını bulmak için algoritma geliştirmek çok zor değildir. Buradaki zorluk bu tür algoritmaların yakın veya uzak ilişkili birçok birliktelik kuralını da meydana

çıkarmasıdır. Bulunabilecek olası birliktelik kuralı sayısı sonsuzdur. Birliktelik kurallarıyla ilgili ortaya çıkabilecek problem, birliktelik kurallarını bulmada bir eşik değeri (herhangi bir olayın gerçekleşmesi için gereken en küçük şey) bulmaktır. Önemsiz gürültülü, kayıp, eksik veriden değerli bilgiyi ayırabilmek ve bu eşik değerini bulabilmek çok zordur. Bu bakımdan ilişkileri tanımlamak ve ilginç birliktelik kurallarından ilginç olmayanları ayırt edebilmek için bazı ölçütlerin belirlenmesi gerekir. Bu ölçütler, matematiksel kavram olan destek ve güven değerleridir (Adriaans vd.,1996:25;Agrawal vd.,1993:207).

Destek, kural cümlesini (<bazı şartlar sağlanıyorsa> sonra <bazı niteliklerin değerlerini tahmin et>) sağlayan olayların birlikte gerçekleştiği durumların tüm durumlara oranı iken, güven, kural cümlesini sağlayan olayların birlikte gerçekleştiği durumların öncül bölümde gerçekleşen olaya oranıdır (Taşkın vd.,2005:33).

Örneğin, bir mağazadan satın alınan ürünlerin hangilerinin birlikte alındığını araştırdığımızı varsayalım. Kişi bir alışverişinde, etek, gömlek, hırka ürünlerini, diğer bir alışverişinde pantolon, kazak ve gömlek ürünlerini, başka bir alışverişinde etek, çorap ürünlerini son olarak da etek, gömlek ve ceket ürünlerini aldığını düşünelim. Toplam dört kez alışveriş yapan bu kişi 2 defasında etek ve gömlek ürünlerini beraber almıştır. Bu durumda etek ve gömlek ürünlerinin beraber alındığı olay sayısı iki, tüm olay sayısı dört olduğundan destek $2/4$, etek ve gömleğin beraber alındığı olay sayısı olan 2'nin öncül bölüm olarak eteğin bulunduğu olay sayısı üçe bölümü olan güven de $2/3$ olarak bulunacaktır. Bu durumda kişi, tüm işlemlerinin %50'sinde etek ve gömlek ürününü birlikte alırken, etek ürününü aldığı işlemlerin %33'ünde de gömlek ürününü aldığını söyleyebiliriz.

Birliktelik kuralları kullanışlı ve anlaşılması basit olduğundan finans, telekomünikasyon, pazarlama, perakendecilik, e-ticaret, mühendislik, fen ve sağlık gibi endüstriyel alanlarda uygulanmaktadır.

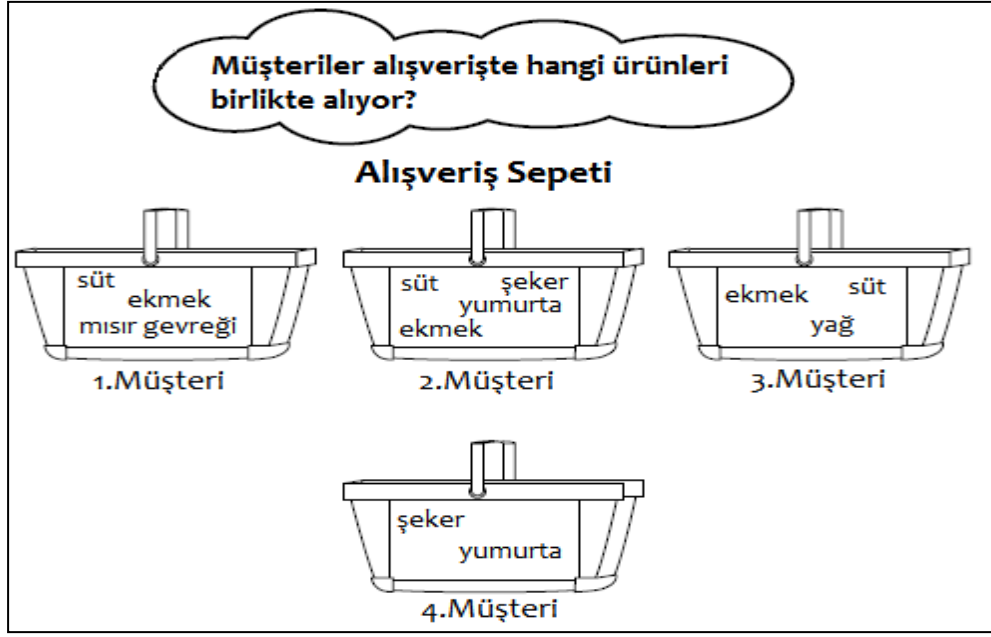
1.8.2. Veri Madenciliği Teknikleri

Veri madenciliği teknikleri, verinin hazırlanmasından sonra, model oluşturma ya da bağıntı bulmak için veri tabanlarında kullanılan algoritmalarıdır. Bunlar;

- Birliktelik Kuralı ve Pazar Sepet Analizi,
- Apriori Algoritması,
- Bellek Tabanlı Yöntemler,
- Karar Ağaçları,
- Yapay Sinir Ağları
- Genetik Algoritmalar

1.8.2.1. Birliktelik Kuralı ve Pazar Sepet Analizi

Birliktelik kurallarının uygulandığı en bilinen örnek market sepet analizi uygulamasıdır. Bu analiz, müşterilerin yaptıkları alışverişlerdeki ürünler arasındaki birliktelikleri veya yakınlıkları bularak müşterilerin satın alma alışkanlıklarını ve alışveriş eğilimlerini analiz eder. Bununla birlikte sepet analizi tekniği, çok basit olarak hangi ürünlerin birlikte alındığını, hangi ürünlerin veya hizmetlerin kampanyaya girmesi gerektiğini, tüketim davranışlarını ortaya çıkarır ve ürün kombinasyonlarının birbirleri ile yakınlıklarını belirlemek için sepet verilerinin analizini yapar. Analiz sonucunda elde edilen bu bilgiler yöneticiler tarafından daha etkin satış stratejisi, promosyon ve müşteriye özel cazip teklifler geliştirmek için kullanılabilir.



Şekil 7:Market Sepet Analizi

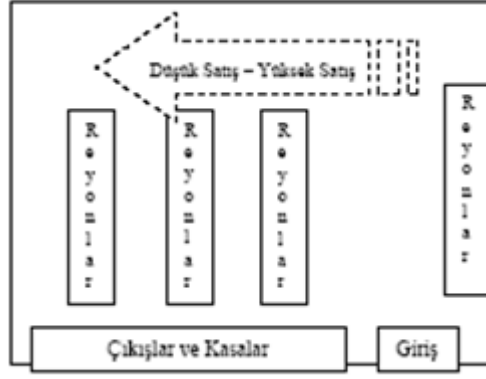
(Kaynak: Akçay, A., 2014, “Bilgi ve Belge Yönetiminde Veri Madenciliği”, İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, s.73)

Sepet analizindeki amaç, satışlar (değişkenler) arasındaki örüntüleri (ilişkileri) bulmak ve buna bağlı kuralları ortaya çıkarmaktır. Eğer “kahve” ürünlerini alan müşterilerin genelde “çikolata” ürünlerini de çok yüksek olasılıkla aldıkları biliniyorsa ve bir müşteri “kahve” ürünlerini alıyor ve “çikolata” ürünlerini almıyorsa, bu müşteri potansiyel bir “çikolata” ürün müşterisidir denilebilir. Bu müşterilere “kahve alanlara çikolata fiyatında indirim yapmak” gibi kampanyalar düzenlenebilir ve “çikolata” ürününün reklamı yapılabilir.

Buna benzer veri analizleri yaparak her ürün veya hizmet için bir sonraki yılın satış tahminleri çıkarılabilir, birlikte satılan ürünler için promosyon uygulaması ve raf düzenlemeleri yapılabilir, yeni bir ürün için muhtemel müşteriler belirlenebilir (Alpaydın,2000:7).

Bununla birlikte birliktelik kuralı ve sepet analizi sonuçları;

- Katalog (insört) tasarımı,
 - Müşterilerin satın aldıkları ürünlere göre gruplandırılması,
 - Mağaza ürün yerleşim planı gibi birçok uygulama alanında kullanılabilir.
- Buna bir marketin yerleşim planı örnek olarak verilebilir (Çingı,2010:16);



Şekil 8: Süpermarket Yerleşimi

(Kaynak: Çıngı, H., 2010, Hacettepe Üniversitesi Ders Notları 6.Bölüm)

Sepet analizi, her ne kadar yoğunlukla pazarlama ve satış alanında kullanılsa da bunun dışında birçok alanda daha kullanılmaktadır:

- İnternet üzerinden yapılan alışverişlerin işlenmesi ve müşterilerin yapacakları potansiyel harcama kalemlerinin bulunması,
- Sigortacılıkta ortaya çıkan değişik yapıdaki suçların dolandırılma olasılığının saptanması ve yapılacak soruşturmaya bilgi sağlanması,
- Hastanedeki hastaların sağlık kayıtlarından, yeni hastaların tedavisi için kullanılacak yöntemin belirlenmesi ve önerilen çeşitli tedavi birleşimlerinden doğan yan etkilerin tahmin edilmesi,
- Telekomünikasyon şirketlerinin isteğe bağlı hizmetlerinin (GPRS,WAP, telesekreter vb.) müşteriler tarafından tercih oranlarına göre karı arttırmak için hangi ürünlerin birlikte kampanyaya dahil edilmesi gerektiğinin belirlenmesi.

Sepet analizi çoğunlukla ticari anlam içeren verilerin var olduğu ancak bu veri üzerinde hangi bağlantıların aranılacağı bilinmediği durumlarda bir başlangıç noktası olarak kullanılabilir. Böylece veri içerisindeki bazı kalıplar sayesinde kazancı arttırmak üzere çeşitli kampanyalara gidilebilir (Karagöz,2006:26).

Örneğin; yurtdışında yapılan market sepet analize uygulamasına göre perşembe günleri bira ve çocuk bezi ürünlerinin birlikte satışlarının çok fazla sayıda olduğu görülmüştür (Linoff vd.,2004:55). Bunun temel sebebi olarak ise evli çiftlerin hafta sonunu evde geçirmek istemeleri ve bu sürede de keyiflerini

bozmamak için gerekli olması muhtemel bira ve çocuk bezi ürünlerini hafta sonu gelmeden satın almak istemeleri olarak gösterilmiştir (Karagöz,2006:26).

1.8.2.1.1.Apriori Algoritması

Veri madenciliğinin en önemli fonksiyonlarından biri kural (sık geçen nesne kümelerin bulunması, güçlü birliktelik kurallarının oluşturulması) çıkarmaktır. Bu maksatla farklı yöntemler geliştirilmiştir. Bu yöntemlerden en çok kullanılanı Apriori algoritmasıdır (Akçay,2014:74).

Apriori Algoritması, sık geçen örüntülerin yakalanmasında kullanılan temel bir algoritmadır. Bu algoritmanın temeli yinelemeli bir yaklaşım sunar (Sever vd.,2002:39). Örüntüleri ya da diğer bir deyişle sık geçen nesne kümelerini bulmak için veri tabanını birçok kez taramak gerekir. İlk taramada bir elemanlı minimum destek eşik değerini (ölçüsünü) sağlayan sık geçen nesne kümeleri bulunur. Destek ölçüsü, nesne kümesinin veri tabanındaki kaç satır tarafından ya da veri tabanındaki toplam satır sayısının yüzde kaçını tarafından doğrulandığını ortaya koyar. Takip eden taramalarda bir önceki taramada bulunan sık geçen nesne kümeleri aday kümeler adı verilen yeni potansiyel sık geçen nesne kümelerini üretmek için kullanılır. Aday kümelerin destek değerleri tarama sırasında hesaplanır ve aday kümelerinden minimum destek eşik değerini (ölçüsünü) sağlayan kümeler o geçişte üretilen sık geçen nesne kümeleri olur. Sık geçen nesne kümeleri bir sonraki geçiş için aday küme olurlar. Bu süreç yeni bir sık geçen nesne kümesi bulunmaya kadar devam eder (Han vd.,2000:40).

Apriori Algoritmasının geleneksel kullanım alanı süpermarketlerde ürünler arası bağlantıları (ilişkileri) tanımlamaktır. Buna benzer şekilde hızlı tüketim mamulleri üreticisi şirketlerde depo sistemlerinin optimizasyonunda da kullanılabilir. Genel olarak bir arada nakledilen ürünlerin yakın raflara yerleştirilmesi depo içindeki hareketi ve taşıma miktarını azaltıcı sonuçlar sağlayacaktır. Algoritma restoranlarda servis hızının artırılması için de çözüm önerileri sağlayabilir. Müşterilerin sipariş verme olasılığı olan ürünleri önceden tahmin ederek hazırlamak veya bağlantılı ürünlerden mönüler oluşturmak gibi çözümler üretilebilir (Argüden vd.,2008:56).

Büyük boyuttaki veri tabanları için birliktelik kuralları bulunurken iki işlem basamağı izlenir. Birinci basamakta işlem hareketlerinde sık tekrarlanan öğeler bulunur. Bu basamakta önemli olan ele alınacak her bir öğenin en az eşik değeri (minimum destek sayısı) kadar tekrarlanmış olmasıdır. Eşik değeri, veri tabanının büyüklüğüne ve karar mekanizmasına göre değişebilir. Bir ilişkinin üç kere tekrarlanması da yüz kere tekrarlanması da eşik değeri olarak nitelendirilebilir. İkinci basamakta ise sık tekrarlanan öğelerden güçlü birliktelik kuralları oluşturulur. Bu kurallar destek ve güven değerleri göz önünde bulundurularak, minimum destek ve güven değerlerini karşılayacak şekilde oluşturulmalıdır. Destek değeri, bir ilişkinin hangi oranda tekrar ettiğini belirlerken, güven değeri bir ilişkinin doğru olarak sonuçlanma olasılığını ortaya koymaktadır (Erpolat,2012:137).

Agrawal ve Srikant tarafından geliştirilen Apriori algoritması 1994 yılında “20. Very Large Database Endowment” (Amerika Birleşik Devletleri’nde veri tabanı ile ilgili alanlarda bilimsel çalışmalar yapmak amacıyla kurulmuş kuruluştur.) konferansında sunulmuştur. Bu bildiride, Agrawal ve Srikant algoritmasının işleyiş ayrıntıları ve algoritmanın kaba kodu şu şekilde sunulmuştur (Agrawal vd.,1994:487):

- Verilerin ilk taranması sırasında, geniş ürün kümelerinin tespiti için, tüm ürünler sayılır.
- Bir sonraki tarama, k 'inci tarama olsun, iki adımdan oluşur.
- Birinci adımda, Apriori-gen fonksiyonu kullanılarak $(k-1)$ 'inci taramada ulaşılan, L_{k-1} ürün kümeleriyle, C_k aday ürün kümeleri oluşturulur.
- İkinci adımda ise veri tabanı taranarak, C_k daki adayların desteği sayılır.
- Hızlı bir sayım için, verilen bir L işlemindeki, C_k 'yi oluşturan adayların çok iyi tayin edilmesi gerekir.

Tablo 2: Apriori Algoritmasının Kodları

<p>Algoritma: Apriori. Sık geçen nesne kümeleri tekrarlı seviye mantığı yaklaşımı kullanarak aday oluşturma</p> <p>Girdi: Bir D veri tabanı, minimum destek esik değeri</p> <p>Output: L, sık nesne kümeler</p> <p>Metod:</p> <ol style="list-style-type: none">1) L1= bul {large 1-nesnekümeler};2) for { k=2; Lk-1 \neq \emptyset; k++ } do begin3) Ck = apriori-gen (Lk-1); // yeni adaylar4) forall işlemler t \in D do begin // D destek sayıları için taranır5) Ct = alt küme (Ck, t); // Aday olan t' lerin alt kümeleri elde edilir6) forall adaylar c \in Ct do7) c.sayısı ++;8) end9) Lk = { c \in Ck \square c.sayısı \geq min_sup }10) end11) return = U_kL_k; <p>Apriori-gen prosedürü (L k-1: sık geçen (k-1)-nesne kümeler)</p> <ol style="list-style-type: none">1) forall nesneküme l1 \in L k-12) forall nesneküme l2 \in L k-13) if (l1[1] = l2 [2]) \cup (l1[2] = l2[2]) \cup ... \cup (l1[k-2] = l2[k-2]) \cup (l1[k-1] < l2[k-1])4) then {5) c = l1 \cup l2 ; // birleştirme adımı: aday oluşturma6) if sık olmayan alt kümeler(c, L k-1) varsa then7) sil c; // budama adımı: gereksiz adaylar çıkarılır8) else c' yi Ck' ya ekle;9) return Ck; <p>Sık olmayan alt kümeler prosedürü (c: aday k-nesneküme;</p> <p>L k-1: sık geçen (k-1)-nesnekümeler); // prior bilgisini kullan</p> <ol style="list-style-type: none">1) forall (k-1)-alt küme c' nin s' i2) if s \in L k-1 then3) return TRUE;4) return FALSE;
--

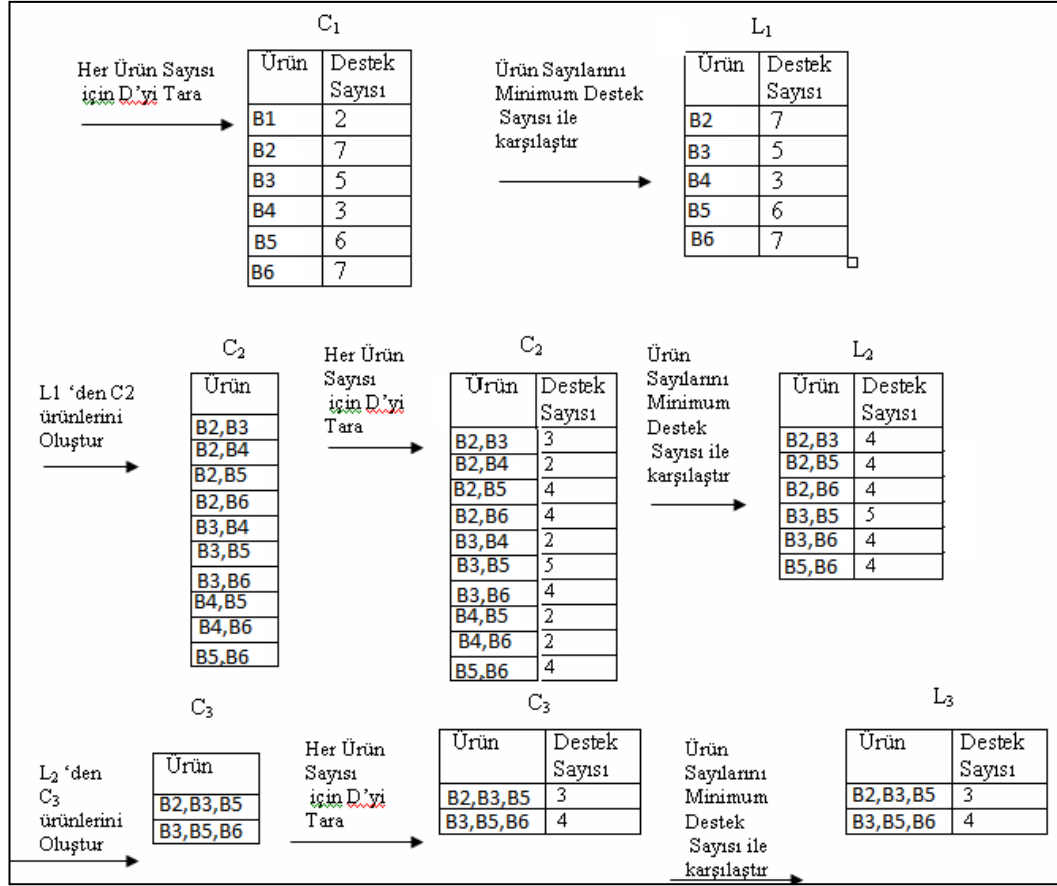
(Kaynak: Ay, D., 2009, “Veri Madenciliği ve Apriori Algoritması ile Süpermarket Analizi”, Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, s.46)

Apriori algoritması ile sık geçen nesne kümelerinin elde edilmesi için işleyiş ayrıntıları yukarıdaki maddelerde anlatılmıştır. Tablo 2’ de kodları verilen birleştirme adımları aşağıdaki Tablo 3’de verilen örnek veri tabanına göre maddeler halinde anlatılmış ve Şekil 9’da gösterilerek özetlenmiştir. (Han vd.,2001:232)

Tablo 3: Market Satış Noktasındaki Alışveriş Kayıtlarını İçeren Veri tabanı

Fatura No	Satın Alınan Ürün No
K1	B1, B2, B6
K2	B2, B3, B5
K3	B2, B6
K4	B2, B4
K5	B3, B4, B5, B6
K6	B3, B5, B6
K7	B2, B5
K8	B2, B3, B4, B5, B6
K9	B1, B6
K10	B2, B3, B5, B6

Tablo 3’de bir marketten yapılan alışveriş kayıtlarını içeren D veri tabanı görülmektedir. Bu veri tabanında yapılan alışverişlerin numaraları Fatura No sütununda görülmektedir. Her alışverişte satın alınan ürünler de Satın alınan Ürünler No sütununda görülmektedir. Han ve Kamber Apriori algoritmasında takip edilen basamaklar şu şekilde (Şekil 9) gösterilmektedir (Han vd.,2001:232):



Şekil 9: Apriori Algoritmasının Uygulaması

(Kaynak: Agrawal, R., ve Srikant, R. Fast Algorithms For Mining Association Rules. In Proceedings of the 20th International Conference on Very Large Databases. 1994. San Francisco: Morgan Kaufmann. s. 487)

1. Algoritmanın ilk basamağında, her ürün tek başına bulunduğu C_1 kümesinin elemanıdır. Algoritma, her ürünün sayısını bulmak için tüm alışverişleri tarar ve elde edilen sonuçlar Şekil 9'da Destek Sayısı (o ürünü içeren alışveriş sayısı) sütununda görülmektedir. Tablo 2'de görülebileceği gibi D veri tabanında B1 ürününden 2 adet, B2 ürününden 7 adet, B3 ürününden 5 adet, B4 ürününden 3 adet, B5 ürününden 6 adet ve B6 ürününden de 7 adet satıldığı görülmektedir.

2. Minimum alışveriş destek sayısının (o ürünü içeren alışveriş sayısı) 3 olduğu kabul edilirse, tek başlarına sık tekrarlanan ürünler L_1 kümesinde görülmektedir. C_1 kümesindeki B1 ürünü hariç tüm ürünlerin destek sayısı,

minimum destek eşik değeri olan 3'ten fazla olduğu için C_1 tüm B1 haricindeki ürünler sık tekrarlanan ürün olarak değerlendirilir ve L_1 kümesine aktarılır.

3-. Hangi ürünlerin ikili olarak sık tekrarlandığını tayin etmek için L_1 kümesindeki ürünlerin ikili kombinasyonları bulunarak C_2 kümesi oluşturulur.

4. C_2 kümesindeki ürünlerin destek sayılarını bulmak amacıyla D veri tabanı taranır ve bulunan değerler destek sayısı sütununda belirtilir.

5. C_2 kümesindeki ürünlerden minimum destek eşik değerini aşan ürünler L_2 kümesine devredilir.

6. Hangi ürünlerin üçlü olarak sık tekrarlandığını belirlemek için L_2 kümesindeki ürünlerin üçlü kombinasyonları bulunarak C_3 kümesi oluşturulur. Bu durumda $C_3 = \{\{B2,B3,B5\}, \{B2,B3,B6\}, \{B2,B5,B6\}, \{B3,B5,B6\}\}$ olması beklenir. Ancak Apriori algoritmasına göre, sık tekrarlanan nesnelerin alt kümeleri de sık tekrarlanan nesne olması gerekmektedir. Buna göre yukarıdaki C_3 kümesindeki elemanlar sık tekrarlanan olmadığı için, yeni C_3 kümesi $C_3 = \{\{B2,B3,B5\}, \{B3,B5,B6\}\}$ olur.

7. C_3 kümesindeki ürünlerin destek sayılarını tespit etmek amacıyla D veri tabanı taranır ve bulunan değerler destek sayısı sütununda belirtilir.

8. C_3 kümesindeki ürünlerden minimum destek eşik değerini aşan ürünler L_3 kümesine devredilir.

9. Hangi ürünlerin dördü olarak sık tekrarlandığını belirlemek için L_3 kümesindeki ürünlerin dördü tek kombinasyonu $\{B2, B3, B5, B6\}$ olarak belirlenir. Ancak bu kümenin kendisi ve alt kümelerinin tamamı sık tekrarlanan nesne olmadığı için C_4 kümesi boş küme olur ve Apriori tüm sık tekrarlanan öğeleri bularak sonlanmış olur (Han vd.,2001:233).

Sık tekrarlanan nesnelere bulduktan sonra, sıra birliktelik kurallarını oluşturmaya gelir. Örneğin sık tekrarlanan bir nesne için $\{B2, B3, B5\}$, boş olmayan tüm alt kümeler $\{B2, B3\}$, $\{B2, B5\}$, $\{B3, B5\}$, $\{B2\}$, $\{B3\}$, $\{B5\}$ alt

kümeleridir. Bu durumda Tablo 3'deki veri tabanına bakılarak şu birliktelik kuralları çıkartılabilir (Han vd.,2001:234).

Tablo 4: Birliktelik Kuralları

1	$B2 \wedge B3 \Rightarrow B5$	Güven (Olasılık) = $3 / 3 = \% 100$
2	$B2 \wedge B5 \Rightarrow B3$	Güven (Olasılık) = $3 / 4 = \% 75$
3	$B3 \wedge B5 \Rightarrow B2$	Güven (Olasılık) = $3 / 5 = \% 60$
4	$B2 \Rightarrow B3 \wedge B5$	Güven (Olasılık) = $3 / 7 = \% 43$
5	$B3 \Rightarrow B2 \wedge B5$	Güven (Olasılık) = $3 / 5 = \% 60$
6	$B5 \Rightarrow B2 \wedge B3$	Güven (Olasılık) = $3 / 6 = \% 50$

Eğer minimum güven eşik değeri %70 olarak belirlenmişse, birinci ve ikinci kurallar dikkate alınır çünkü diğer kurallar eşik değerini aşmamış olurlar (Han vd.,2001:234).

Uygulamış olduğu veri madenciliği yöntemiyle Tablo 3'deki sonuçlara ulaşan market yöneticisi, ürün satışlarını arttırmak için çeşitli pazarlama stratejileri geliştirebilir. Örneğin, güven oranı yüksek çıkan B2,B3 ve B5 ürünlerini reyonda aynı rafa veya yakın raflara dizerek müşterisine farkındalık oluşturabilir. Bu üç üründen birisini almayı unutan müşterinin diğer ürünleri almak için rafa yaklaştığında unutmuş olduğu ürünü görerek hatırlamasını sağlayabilir.

1.8.2.1.2.Sepet Analizi

Sepet analizinin mantığında yer alan teknikler istatistik ve olasılık kavramlarından oluşmaktadır. Ticari anlam taşıyan veriler üzerinde belirli bir ürün kombinasyonunun kaç kez geçtiğinin bulunması işlemi tek başına yeterli değildir. Bu kombinasyonu firma açısından anlamlı hale getirecek asıl nokta, bu kombinasyonu meydana getiren kuralı bulmaktır (Karagöz,2007:33).

Kural tanımı iki adımdan oluşur: koşul adımı ve sonuç adımı (Han vd.,2001:240);

Eğer Koşul doğru ise, Sonuç da doğrudur.

Firma açısından pratikte uygulamaya dönüştürülebilecek kuralların sadece bir adet sonuç kısmı vardır. Yani; Eğer cips ve cumartesi ise kola satılabilir kuralı.

Eğer cumartesi ise cips ve kola kuralından daha çok faydalıdır. Çünkü sadece günün cumartesi olmasından dolayı müşteriye cips ve kola satmaya çabalamak anlamsız olacaktır. Tam tersine eğer günlerden cumartesi ise ve müşteri cips almış ise bu müşterinin kola alma olasılığı çok yüksek demektir. Bunun için firma kola satışlarını arttırmak için cumartesi günleri cips ürünleri ile kolaları beraber satmak üzere promosyona girebilir. Böylelikle ikinci kural firma açısından çok daha anlamlı ve uygulamaya dönüştürülebilecek bir dokuya sahiptir.

Sepet analizi tekniğinin işleyiş kurallarına örnek olarak bir süpermarketin veri tabanında yer alan alışveriş bilgileri aşağıdaki örnekte (Tablo 5) incelenmiştir (Karagöz,2007:34).

Tablo 5:Süpermarketten Yapılan Alışveriş Bilgilerini İçeren Veri Tabanı

Müşteri No	Satın Alınan Ürünler
1	süt, peynir, yumurta
2	zeytin, ekmek
3	ekmek
4	siyah çay, kahve, yumurta, sucuk, ekmek
5	siyah çay, su, peynir, yumurta, ekmek
6	ayran, su, zeytin, sucuk
7	salam, peynir, yumurta
8	su, peynir, kahve, yumurta, ekmek
9	siyah çay, meyve suyu, süt, kahve, yumurta, reçel, ekmek
10	siyah çay, sucuk

Tablo 5 detaylı olarak incelendiğinde peynir içeren dört alım işleminin yumurta da içerdiği ve altı alım işleminin içinde yer alan yumurtanın dört alım işleminde peynirle birlikte olduğu görülmektedir. Birliktelik kuralı bu tür benzerlikleri nitelendirmeye ve karakterize etmeye çalışır. Her bir alım işlemi bir

ürün seti olarak nitelendirilir. Hatta alım işleminde görünmeyen farklı kombinasyonlar da ürün seti olarak kabul edilir.

Sık tekrarlanan ürün seti stratejisine göre tespit edilen minimum destek ve güven değerleri araştırılan veri yığınının değerlendirilen nadir veri için yönetilebilir biçime dönüşmesini sağlar. Sık tekrarlanan ürün seti üretimi minimum destek değeri=0,25 için Tablo 6’te özetlenmiştir.

Tablo 6: Ürün Tekrar Sayısı

siyah çay [Destek=0,4]
kahve [Destek=0,3]
ekmek [Destek=0,6]
peynir [Destek=0,4]
su [Destek=0,3]
sucuk [Destek=0,3]
yumurta [Destek=0,6]
siyah çay, ekmek [Destek=0,3]
siyah çay, yumurta [Destek=0,3]
kahve, ekmek [Destek=0,3]
kahve, yumurta [Destek=0,3]
ekmek, yumurta [Destek=0,4]
peynir, yumurta [Destek=0,4]
siyah çay, ekmek, yumurta [Destek=0,3]
kahve, ekmek, yumurta [Destek=0,3]

Tablo 6’dan de minimum güven değeri = 0,5 için, Tablo 7’deki 41 birliktelik kuralı çıkarılabilir. Her bir kuralın önce geleni oktan önce, sonuç değeri ise oktan sonra gösterilmiştir.

Tablo 7:Birliktelik Kuralları

→ ekmek [Destek=0,6, Güven=0,60]
→ yumurta [Destek=0,6, Güven=0,60]
siyah çay → [Destek=0,4 , Güven=1,00]
siyah çay → ekmek [Destek=0,3 , Güven=0,75]
siyah çay → ekmek, yumurta [Destek=0,3 , Güven=0,75]
siyah çay → yumurta [Destek=0,3, Güven=0,75]
siyah çay, ekmek → [Destek=0,3, Güven=1,00]
siyah çay, ekmek, yumurta → [Destek=0,3, Güven=1,00]
siyah çay, yumurta → [Destek=0,3, Güven=1,00]
siyah çay, ekmek → yumurta [Destek=0,3, Güven=1,00]
siyah çay, yumurta → ekmek [Destek=0,3, Güven=1,00]
kahve → [Destek=0,3, Güven=1,00]
kahve → ekmek [Destek=0,3, Güven=1,00]
kahve → yumurta, ekmek [Destek=0,3, Güven=1,00]
kahve → yumurta [Destek=0,3, Güven=1,00]
kahve, ekmek, yumurta → [Destek=0,3, Güven=1,00]
kahve, ekmek → [Destek=0,3, Güven=1,00]
kahve, yumurta → [Destek=0,3, Güven=1,00]
kahve, ekmek → yumurta [Destek=0,3, Güven=1,00]
ekmek → [Destek=0,6, Güven=1,00]
ekmek → siyah çay, yumurta [Destek=0,3, Güven=0,50]
ekmek → siyah çay [Destek=0,3, Güven=0,50]
ekmek → kahve [Destek=0,3, Güven=0,50]
ekmek → yumurta, kahve [Destek=0,3, Güven=0,50]
ekmek → yumurta [Destek=0,4, Güven=0,67]

ekmek, yumurta → [Destek=0,4, Güven=1,00]
peynir → [Destek=0,4, Güven=1,00]
peynir → yumurta [Destek=0,4, Güven=1,00]
peynir, yumurta → [Destek=0,4, Güven=1,00]
su → [Destek=0,3, Güven=1,00]
sucuk → [Destek=0,3, Güven=1,00]
yumurta → [Destek=0,6, Güven=1,00]
yumurta → siyah çay, ekmek [Destek=0,3, Güven=0,50]
yumurta → siyah çay [Destek=0,3, Güven=0,50]
yumurta → kahve, ekmek [Destek=0,3, Güven=0,50]
yumurta → kahve [Destek=0,3, Güven=0,50]
yumurta → ekmek [Destek=0,4, Güven=0,67]
yumurta → peynir [Destek=0,4, Güven=0,67]
yumurta, ekmek → siyah çay [Destek=0,3, Güven=0,75]
yumurta, ekmek → kahve [Destek=0,3, Güven=0,75]
yumurta, kahve → ekmek [Destek=0,3, Güven=1,00]
kahve, yumurta → [Destek=0,3, Güven=1,00]
kahve, ekmek → yumurta [Destek=0,3, Güven=1,00]

Başlangıçtaki kullanıcının destek ve güven değerlerine göre birliktelik kuralları çoğunlukla binlerce kuralın ortaya çıkmasına neden olmaktadır. Bu konumda istenen, kural sayısının muhtemel olduğunca azaltılması ve en ilginç kuralların tanımlanmasıdır. Çoğunlukla ilginçlik kuralı kuralın destek değeri ile önce gelenin destek değeri ve sonucun destek değeri çarpımının farkını ilişkilendirir. Önce gelen ve sonuç birbirinden bağımsız ise kuralın destek değeri takribi olarak önce gelenin ve sonucun destek değerleri çarpımına eşittir. Önce gelen ve sonuç bağımsız ise kural güven değerinin yüksekliğiyle alakalı değildir

(Shapiro,1991).İlginçliğin nesnel ve sübjektif ölçüleri vardır ve bunların ayırt edilmesinde fayda vardır(Liu vd.,2000:47).

Objektif ölçüler özel uygulamaya bağımsız olarak uygulanan güven ve destek değerleri gibi ölçüleri içerir. Sübjektif ölçüler ise özel içerikte kullanıcının ihtiyaç duyduğu özel bilgi ihtiyaçlarıyla kuralın ilginçliğini özdeşirir. En fazla objektif ölçü de yükseltmedir(Lift)(Karagöz,2007:37).

$$\text{Yükseltme}(A \rightarrow C) = \text{Güven}(A \rightarrow C) / \text{Destek}(C)$$

Yükseltme değerinin 1'den yüksek olması sonucun önce geleni içeren işlemlerde önce geleni içermeyenlere göre daha fazla görüldüğünü gösterir. Örneğin, {yumurta} → {peynir} birlikteliği düşünülecek olursa, Destek({peynir})=0,4, Güven({yumurta} → {peynir})=0,67. Bu nedenle, yükseltme({yumurta} → {peynir})= 0,67/0,4 = 1,675 olur. Tam tersine, aynı güven değeri için {yumurta} → {ekmek} birliktelik kuralı düşünüldüğünde; destek({ekmek})=0,6 , Güven({yumurta} → {ekmek})=0,67. Bu nedenle, yükseltme({yumurta} → {ekmek})=0,67/0,6=1,117 olur. Yükseltmenin bu göreceli değerleri yumurtanın ekmeğin sıklık değerine göre peynir sıklığı üzerinde daha fazla etkiye sahip olduğunu gösterir.

Düşük tekrar sayılı ve yüksek yükseltme değerli bir birliktelik daha düşük yükseltme değerli ve daha yüksek tekrar sayılı seçenekleri olan kuraldan daha az dikkat çekici olabilir. Zira sonraki daha bireyseldir ve önce gelen ve sonuç arasındaki ilişkiden kaynaklanan birimlerin sayılarındaki toplam artış daha büyüktür. Bu konumda tek bir değerdeki hem gücü hem de hacmi içeren ölçü dürtü (leverage)'dür (Shapiro,1991:39).

$$\text{Dürtü}(A \rightarrow C) = \text{destek}(A \rightarrow C) - \text{destek}(A) \times \text{destek}(C),$$

Örneğin, {kahve} → {yumurta} ve {peynir} → {yumurta} birlikteliklerini düşünersek; her ikisi de güven =1,0 ve yükseltme =1,667 değerlerine sahiptir. Bununla beraber, ikincisi daha fazla müşteriye uygulandığından daha büyük ilginçliğe sahip olabilir. Destek({kahve} → {yumurta}) = 0,3. Destek({kahve}) = 0,3. Destek({yumurta}) = 0,6. Bu nedenle, dürtü({kahve} → {yumurta}) = 0,3 - 0,3×0,6 = 0,3 - 0,18 = 0,12. Destek({peynir} → {yumurta}) = 0,4. Destek({peynir}) = 0,4. Destek({yumurta}) = 0,6. Bu yüzden, dürtü({kahve} → {yumurta}) = 0,4 -

$0,4 \times 0,6 = 0,4 - 0,24 = 0,16$. Sonrakinin birlikteliğe mutlak etkisi ise öncekinden daha da büyüktür.

Daha küçük değerli birlikteliklerin önemsenmediği minimum değer konularak tespit edilen birliktelik kümesini daha da kısıtlamak için yükseltme veya dürtü ölçüleri kullanılır.

Tablo 7'yi yakından incelendiğinde, benzer birliktelik kurallarının değişik şekiller oluşturduğu görülür. Örneğin son iki birliktelik kuralı aynı ürünleri lakin farklı önce gelen ve sonuç kombinasyonlarını içerirler. Herhangi bir ürün seti için, ürün seti parçaları arasındaki tüm birliktelik kuralları bire bir destek fakat farklı güven, yükseltme ve dürtü değerlerine sahip olabilir. Örneğin verilen iki örnekte de 0,3 destek değerine ama farklı güven değerine sahiptir. Her ikisinde de önce gelen ve sonuç değerleri farklı {kahve, ekme, yumurta} kümesinin farklı parçalarından oluşur. Bazı uygulamalar için ürün setlerinin birbirleri ile olan ilginçlik benzerliği ve ürünlerin önce gelen ve sonuç olarak bölünmesi yarar sağlamayabilir. Örneğin, markete hangi ürünün hangi ürüne yakın yerleştirileceği düşünüldüğünde, bir kural yapısı tarafından gösterilen rastlantılı ilişkiye göre kategorilendirmektense ilgili olanları yan yana kategorilendirmek daha mantıklıdır. Daha önceden tanımlanmış ürün setinin kendi aralarında net bir düzenleme olmaksızın ürünlerin benzerliğini tanımlar. İlginçlik ölçüleri, ilginç ürün setleri kendiliğinden keşfedilecekse gereklidir. Birliktelik kuralının emsali olan yükseltme ve dürtü aşağıdaki gibi tanımlanır(Karagöz,2007:38):

$$\text{ürün seti-yükseltmesi (I)} = \text{destek(I)} / \Pi \text{ destek(\{i\})}$$

$$\text{ürün seti-dürtüsü (I)} = \text{destek(I)} - \Pi \text{ destek(\{i\})}$$

Ürün küme yükseltme oranı gözlenen desteğin ürünler arası korelasyon olmaması durumunda gelebilecek desteğe olan oranıdır. {siyah çay, ekme, yumurta} için ürün küme yükseltme, bu küme için geçerli destek değeri (0,3) 'ün ürün setindeki her bir ürünün destek değerlerinin çarpımına ($0,4 \times 0,6 \times 0,6 = 0,144$) bölünmesiyle 2,08 olarak bulunur. Ürün seti dürtüsü izlenen destek ile ürünler arası korelasyon olmaması konumunda beklenecek destek arasındaki farktır.

Birliktelik kuralı buluş arařtırmalarının geneli sık tekrar eden ürün seti keşfinin alt problemlerine yoğunlaşmıştır.Kısıtlamaları destek ve güven değerlerine göre gerçekleřtiren birliktelikleri arařtırırken sık tekrar eden ürün seti bulununca birliktelik kuralı oluřturmak çok kolaydır (Han vd.,2000:8).

Daha önce tanımlanan Apriori Algoritması ürünlerin birbirleri ile sık tekrarlanmadığı ve minimum destek değerinin yüksek olduđu konumlara karşılık gelen nadir veri kümeleri için çok etkilidir. Ayrıca ürün küme sayısı daha az ve sebebi ile daha az işlem gerektiği için oldukça seridir. Fakat Apriori algoritması, veri yoğun olduğunda tekrar eden ürün kümeleri fazla olduğundan ve bu nedenden işlem sayısını ve hesaplama süresini arttıracığından daha az etkilidir (Mao vd.,2000:24).

Bunun azaltılabilmesi için yapılması gereken işlem sık tekrar eden ürün küme sayısını azaltmaktır. Örneğın sık tekrar eden kapalı ürün kümeleri işlemi yalnızca kapalı ürün kümelerini kaydeder (Pei vd., 2001:445).

Bir ürün seti D 'nin destek değeri ((destek(V)) olan ve kendisinin destek değerine ((Destek (D)) eşit olan bir V üst kümesi mevcut değilse D ürün kümesi kapalıdır. Eğer V üst kümesi destek değeri, D 'nin destek değerine eşit değerde ise her iki küme de özdeş işlem içerisinde birlikte görünürler demektir. Kapalı sık tekrar eden ürün kümesinden ve onun destek değerlerinden tüm sık tekrar eden ürün kümelerini ve destek değerlerini tayin etmek mümkündür. Tablo 7'de sık tekrar eden kapalı ürün kümesi görülmekte ve bu kümenin Tablo 6' daki 15 ürün kümesiyle karşılaştırıldığında 9 kapalı sık tekrar eden ürün kümesine düřtüğü görülmektedir (Karagöz,2007:39).

Tablo 8:Kapalı Sık Tekrar Eden Ürün Kümeleri

siyah çay [Destek=0,4]
ekmek [Destek=0,6]
su [Destek=0,3]
sucuk [Destek=0,3]
yumurta [Destek=0,6]
ekmek, yumurta [Destek=0,4]
peynir, yumurta [Destek=0,4]
siyah çay, ekmek, yumurta [Destek=0,3]
kahve, ekmek, yumurta [Destek=0,3]

İşlenmesi gereken sık tekrarlanan ürün kümeleri sayısını azaltmaktan kaynaklanan hızlı hesaplama yararlarına ilaveten, kapalı sık tekrarlanan ürün setini destekleyen yaklaşıma göre, $A \rightarrow C$ kuralına göre oluşturulan birliktelik kurallarını da A ve $A \cup C$ 'yi de kapalı sık tekrarlanan ürün kümesi olarak kabul eder ve birliktelik kuralı sayısını kısıtlar. Sonuçta Tablo 9'da görüldüğü gibi 41 yerine 26 birliktelik kuralı kalmış olur(Karagöz,2007:40).

Tablo 9: Kapalı Sık tekrar Eden Ürün Kümelerinden Yaratılan Birliktelik Kuralları

→ siyah çay, ekmek, yumurta [Destek=0,3, Güven=0,30]
→ siyah çay [Destek=0,4, Güven=0,40]
→ ekmek, yumurta [Destek=0,4, Güven=0,40]
→ ekmek [Destek=0,6, Güven=0,60]
→ peynir, yumurta [Destek=0,4, Güven=0,40]
→ su [Destek=0,3, Güven=0,30]
→ sucuk [Destek=0,3, Güven=0,30]
→ yumurta [Destek=0,6, Güven=0,60]
siyah çay → [Destek=0,4, Güven=1,00]

siyah çay → ekmek, yumurta [Destek=0,3, Güven=0,75]
siyah çay, ekmek, yumurta → [Destek=0,3, Güven=1,00]
kahve, ekmek, yumurta → [Destek=0,3, Güven=1,00]
ekmek → [Destek=0,6, Güven=1,00]
ekmek → siyah çay, yumurta [Destek=0,3, Güven=0,50]
ekmek → yumurta, kahve [Destek=0,3, Güven=0,50]
ekmek → yumurta [Destek=0,4, Güven=0,67]
ekmek, yumurta → [Destek=0,4, Güven=1,00]
peynir, yumurta → [Destek=0,4, Güven=1,00]
su → [Destek=0,3, Güven=1,00]
sucuk → [Destek=0,3, Güven=1,00]
yumurta → [Destek=0,6, Güven=1,00]
yumurta → siyah çay, ekmek [Destek=0,3, Güven=0,50]
yumurta → kahve, ekmek [Destek=0,3, Güven=0,50]
yumurta → ekmek [Destek=0,4, Güven=0,67]
yumurta → peynir [Destek=0,4, Güven=0,67]
yumurta, ekmek → siyah çay [Destek=0,3, Güven=0,75]
yumurta, ekmek → kahve [Destek=0,3, Güven=0,75]

1.8.2.2. Bellek Tabanlı Yöntemler

Bellek tabanlı yöntemler, verilen bir problemi doğrudan doğruya geçmiş deneyimleri ve çözümleri kullanarak çözmeye çalışan bir tekniktir. Bu yöntemlerde, yeni bir problem verildiğinde, öncelikle karşılaşıp saklanmış durumlar incelenir ve benzer durumlar var mı diye bakılır. Eğer benzer durumlar bulunmuşsa, bunların çözümü yeni probleme uygulanır ve problem gelecekte yeni problemlere referans olmak üzere saklanır. En yakın k komşu algoritması bellek tabanlı yöntemler için kullanılabilir en uygun yöntemdir (Gürgen,2008:13).

Bellek tabanlı yöntemlerin en dikkat çeken özelliği veriyi olduğu gibi kullanabilme yeteneğidir. Diğer veri madenciliği yöntemlerinin tersine bellek tabanlı yöntemler, kayıtların formatı (şekli) yerine sadece iki işlemin varlığı ile ilgilenir. Bu işlemler, iki kayıt arasındaki uzaklığı belirleyen bir uzaklık fonksiyonu ve komşu kayıtları işleyerek bir sonuç üreten kombinasyon (birleşim) fonksiyonudur. Bellek tabanlı yöntemler; müşteri tepkisi tahmini sahtekarlık tespiti, klinik işlemler gibi alanlarda kullanılmaktadır (Dolgun,2006:33).

Bellek tabanlı yöntemlerin güçlü olduğu noktalar şunlardır (Tantuğ,2002:45);

- Basitçe anlaşılabilir sonuçlar üretir,
- Gelişigüzel seçilen, hatta birbiri ile alakasız olabilen verilere bile uygulanabilir,
- Analiz alanlarının çok olduğu durumlarda dahi etkili olarak çalışabilir,
- Eğitim öbeklerinin oluşturulması basittir.

Bellek tabanlı yöntemlerin zayıf olduğu noktalar da şunlardır (Tantuğ,2002:45);

- Sınıflama ve tahmin işlemleri için kullanıldığı zaman işlem maliyeti yüksektir,
- Eğitim öbekleri (kümesi) için büyük boyutlarda alana ihtiyaç vardır,
- Üretilen sonuçlar seçilen uzaklık fonksiyonuna, kombinasyon fonksiyonuna ve komşu sayısına direkt olarak bağlıdır.

1.8.2.3.Karar Ağaçları

Karar alma, işletmelerin işleyişleri sırasında çoğunlukla karşılaştıkları bir durumdur. Dışarıdan bakıldığında sıradanmış gibi gözükse de birçok işlem esasen bir dizi karar alma işlemini içermektedir (Decrop,2006:1). Karar verici, kaç adet seçeneğin gerçekleşeceğini meçhul olduğu bir problemle ilgili en doğru kararı verebilmek için bir takım yöntemlere ihtiyaç duyar. Karar ağacı tekniği, basitçe anlaşılabilir kurallar üretmesi, kuralların görselleştirilebilmesinin, kurgulanmasının, yorumlanmasının ve veri tabanları ile uyumunun kolaylığı avantajları nedeniyle

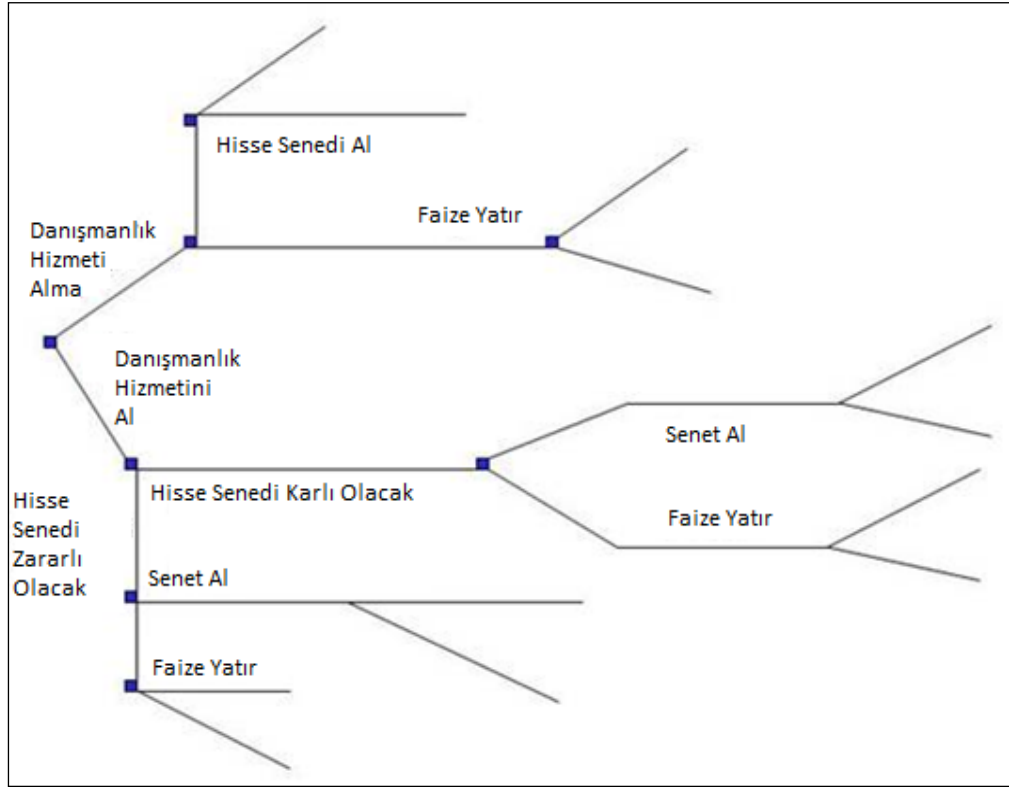
sınıflandırma ve kümeleme yöntemlerinde sıkça kullanılan veri madenciliği tekniklerindedir. (Atlas vd.,2012:4).

Karar ağacı tekniği, özellikle seçenek sayısının fazla olduğu, birden fazla kararın art arda verilmesi gibi daha karmaşık bir problemin çözüm analizinde kullanılan bir tekniktir (Sezen,2004:5). Genellikle sınıflama amacıyla kullanılan karar ağaçlarının hedefi bağımlı değişkendeki farklılıkları maksimize edecek şekilde verileri yukarıdan aşağıya, genelden özele hiyerarşik olarak alt sınıflara (farklı gruplara) ayırmaktadır (Akpınar,2000:12).

Karar ağacı, adında da belirtildiği gibi akış diyagramına benzer bir ağaç görünümünde bir tekniktir. Bu teknikte her bir dal, bir testin sonucunu, yaprak düğümleri ise sınıfları temsil eder. Problemden bilinmeyen bir örneklemini sınıflamak için örneklemin nitelik değerleri karar ağacı karşısında test edilir. Kökten, o örneklem için sınıf tahminini barındıran yaprak düğümüne kadar bir yol izlenir ve karar ağaçları basitçe sınıflama kurallarına dönüştürülebilir(Toktaş vd.,2004:102). Diğer bir anlatımla; karar ağacının her düğümünde değişkenler test edilir. Karar ağacı algoritması, ağacın kökünde hangi değişken ile test edilmesi gerektiği sorusu ile işleme başlayarak yukarıdan aşağıya, genelden özele doğru ağacı oluşturur. Bu işlemi her örnek değişken, eğitim öğelerinin sınıflandırmasına karar vermek için istatistiksel test kullanılarak değerlendirilir. Değişkenler arasından en iyi değişken seçilir ve ağacın kök düğümünde test için kullanılır. Her bir düğüm için oluşturulacak dalların sayısı, test sonucunda kabul görmüş olan değişkenin alabileceği değer sayısına göre farklılaşmaktadır (Kök vd., 2005:39).

Karar ağaçları ile ilgili örnek verecek olursak; Siz bir işletme yöneticisisiniz ve elinizde işletmeye dair yüklü bir miktarda nakit para var. Bu parayı sizden en yüksek kazancı sağlayacak şekilde faiz veya senet olarak değerlendirmeniz talep ediliyor. İsterseniz bir finans danışmanından yardım alabilir isterseniz kendiniz işletme yöneticisi olarak karar verebilirsiniz. İhtimalleri çıkartacak olursak; ilk olarak finans danışmanına başvuralım. Finans danışmanı size senet al veya faize yatır alternatiflerini sunacaktır. Bu alternatiflerde kendi aralarında olumlu (başarılı) veya olumsuz (başarısız) olarak ikiye ayrılacaktır. İlk aşamada finans ile ilgili gerekli bilgiye sahip olmadığımızı

düşünülerek finans danışmanına başvurmak mantıklı gelecektir. Fakat bu alternatif sonucunda danışmana da bir miktar ödeme yapmamız gerekecektir. Diğer alternatif de ise danışmanlık hizmeti almazsınız ve kendiniz işletme yöneticisi olarak karar verirsiniz. Hisse senedi karlı kazanç sağlayacak (karlı olmasına karşın faize yatır veya senet al) veya hisse senedi zararlı olacaktır. Bahsi geçen bu olayların karar ağacı aşağıdaki Şekil 10' daki gibi olacaktır (Çıngı,2010:11).



Şekil 10: Karar Ağacı Örneği

(Kaynak: Çıngı, H., 2010, Hacettepe Üniversitesi Ders Notları 6.Bölüm)

Karar ağaçları tekniğinin sağladığı avantajlar yanında dezavantajları da vardır. Çoğunlukla karar vermek bu yaklaşım için önemli bir sorun oluşturur. Sorun şundan kaynaklanır: Ağaç, ayrılan bölümleriyle genişledikçe, yapılan sınıflandırma veriyi çok fazla sayıda alt bölümlere böler. Bu alt bölümlerin içeriği küçülmeye başlar. İncelenmesi gereken bölümlerde farklı durumların sayısı arttıkça, kümelerinin her biri daha da küçülür. Bu durum karar ağacını yönetmeyi zorlaştırır ve uygulamada bazı sorunlara (hatalı sonuçlara) yol açabilir (Altaş,2012:5).

Karar ağacı temelli analizlerin yaygın olarak kullanıldığı sahalarda arasında;

- Belirli bir sınıfın olası üyesi olacak elemanların belirlenmesi (segmentation),
- Çeşitli vakaların yüksek, orta, düşük risk grupları gibi biçiminde çeşitli kategorilere ayrılması (stratification),
- Gelecekte gerçekleşmesi muhtemel olayların tahmin edilebilmesi için kurallar oluşturulması,
- Parametrik modellerin kurulmasında kullanılmak üzere çok sayıdaki değişken ve veri kümesinden faydalı olanlarının seçilmesi,
- Yalnızca belirli alt gruplara özgü olan ilişkilerin tanımlanması,
- Kategorilerin birleştirilmesi ve sürekli değişkenlerin kesikli değişkenlere dönüştürülmesi sayılabilir (Akpınar,2000:19).

Karar ağacı tekniklerinin farklı alanlarda gerçekleştirilen uygulama örnekleri ise şu şekilde sıralanabilir;

- Hangi demografik grupların mektupla yapılan pazarlama uygulamalarında yüksek yanıt oranına sahip olduğunun belirlenmesi (Direct Mail),
- Kişilerin kredi geçmişlerini kullanarak kredi kararlarının verilmesi (Credit Scoring),
- Geçmişte işletmeye en faydalı olan kişilerin özelliklerini kullanarak iş alma süreçlerinin belirlenmesi,
- Tıbbi gözlem verilerinden yararlanarak en aktif kararların verilmesi,
- Hangi değişkenlerin satışları etkilediğinin tespit edilmesi,
- Üretim verilerini inceleyerek ürün hatalarına yol açan değişkenlerin tespit edilmesidir (Akpınar,2000:19).

1.8.2.4.Yapay Sinir Ağları

Yapay Sinir Ağları tekniği, veri madenciliği yöntemlerinde en çok bilinen ancak en az anlaşılan tekniktir (Michael,2000:122). Yapay sinir ağları tekniği değişik yapılarda tanımlanmaktadır. Bu tanımlardan bazılarında yapay sinir ağları, insan beyninin yapısından yola çıkarak tasarlanan, ağırlıklı bağlantılar yardımıyla birbirine bağlanan ve her biri kendi belleğine sahip işlem elemanlarından meydana gelen paralel ve dağıtılmış bilgi işleme yapıları olarak belirtilmiştir

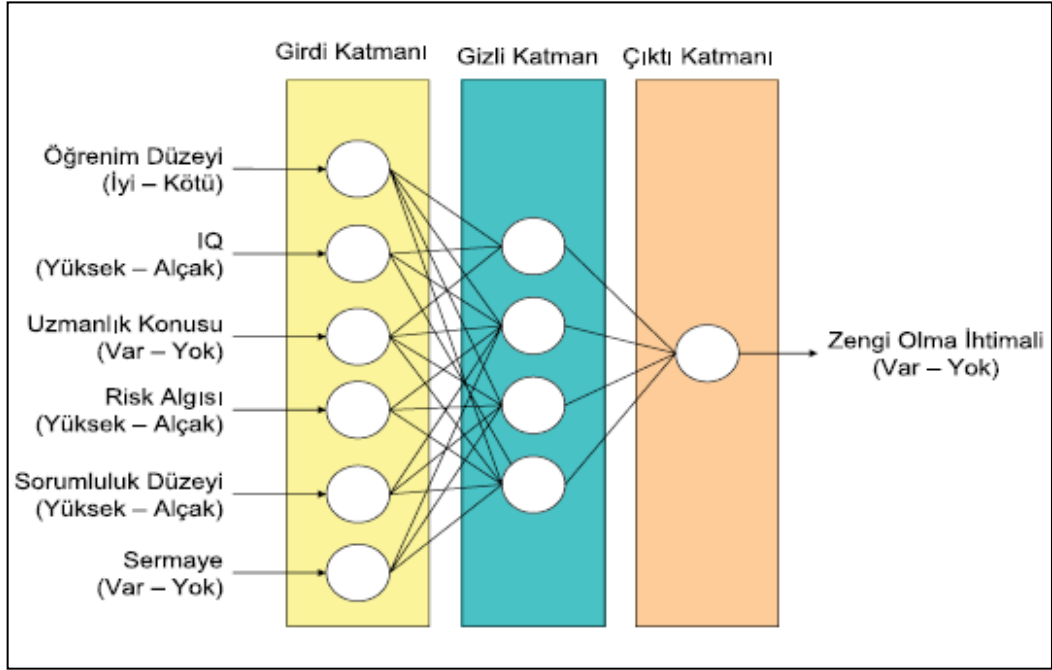
(Kiremitçi,2005:58). Başka bir tanımda ise yapay sinir ağları, insan beyninin özelliklerinden olan öğrenme yolu ile yeni bilgiler üretebilme veya keşfedebilme gibi yeteneklerden esinlenerek geliştirilmiş, bu yetenekleri herhangi bir yardım almaksızın otomatik olarak gerçekleştirmeye çalışan bilgisayar programlarıdır (Öztemel,2003:29).

Yapay sinir ağları, bilgisayar programcısının geleneksel yeteneklerine gereksinim duymayan kendi kendine öğrenme yetenekleri olan düzeneklerdir. Bu ağlar, öğrenme ile birlikte ezberleme ve bilgiler arasında örüntüler oluşturabilme yeteneğine de sahiptirler. Yapılan tanımların ortak birkaç önemli noktası bulunmaktadır. Bunlardan en önemlisi, yapay sinir ağlarının birbirine hiyerarşik olarak bağlantılı ve paralel olarak çalışan yapay hücrelerden meydana gelmesidir (Kiremitçi,2005:59).

Yapay sinir ağları, insanların tecrübelerinden bir takım bilgiler çıkartması gibi kendisine verilen örneklerden de bir takım bilgiler ortaya çıkartma yeteneğine sahiptir. Yapay sinir ağları öncelikle neticesi bilinen belli bir veri kümesi üzerinde öğrenme algoritmaları çalıştırılarak eğitilir. Bu eğitim sonucunda yapay sinir ağının içerisindeki ağırlıklar tespit edilir. Bu ağırlıklar kullanılarak yeni gelen veriler işlenerek bir sonuç elde edilir. Yapay sinir ağlarının en olumsuz hususu ise bu ağırlıkların neden ilgili değeri aldığının bilinmemesi ya da çıkan sonuçun neden geçerli bir sonuç olduğunu açıklayamamasıdır. Yapay sinir ağlarını kullanmak için en doğru yaklaşım ise onları içi bilinmeyen bir yapıda çalışan kara kutular olarak düşünmek olacaktır (Dolgun,2006:31).

Yapay sinir ağlarının yapısı nöronlar arasındaki bağlantılar ve bağlantıların ağırlıkları (öğrenme mekanizması ile geliştirilen) üzerine kuruludur. Modelin karmaşıklığı bu bağlantı şekline bağlıdır. Yapıda nöronların bir araya geldiği alanlara katman denir (Argüden vd.,2008:62). Basit bir yapay sinir ağı genellikle giriş katmanı, gizli katman ve çıkış katmanından oluşmaktadır. Girdi katmanındaki her bir yapı bir değişkeni ifade etmektedir (Çelik,2009:32). Modelde öncelikle her bir değişken bir bağlantı ağırlığı ile çarpılır. Daha sonra nöronlar giriş bilgilerini ağırlıklandırdıktan sonra toplayarak doğrusal ya da doğrusal olmayan bir fonksiyonda işler ve çıktı bilgisine dönüştürürler

(Yazıcı,2007:17). Oluşturulan bilgi, diğer nöronlar için girdi bilgisi olarak kullanılır ve model kurulduktan sonra veriler sürekli olarak modele girilir ve elde edilen sonuçlar gerçek sonuçlar ile karşılaştırılarak modelde iyileştirmeler (ağırlıklarda değişiklikler) yapılır. Gereksiz bilgiler elenip minimum kabul edilebilir hata seviyesine ulaşıldığında yapay sinir ağı modeli tamamlanmış olur (Argüden vd.,2008:62). Şekil 11’de bir yapay sinir ağı çalışma modeli görülmektedir.



Şekil 11:Yapay Sinir Ağı

(Kaynak: Argüden, Y., Erşahin, B., 2008, “Veri Madenciliği”, Arge Danışmanlık Yayınları, s.62)

Günümüzde yapay sinir ağlarının genel kullanım alanlarından bazıları şunlardır (Elmas,2003:25):

- Denetim
- Sistem modelleme
- Ses tanıma
- El yazısı ve parmak izi tanıma
- Meteorolojik tahminler ve değerlendirmeler
- Otomatik araç kontrolü

- Fizyolojik işaretleri (kan basıncı, kalp atışı, tansiyon) izleme, tanıma ve değerlendirme.

Bunlarla beraber yapay sinir ağları, veri madenciliğinde hem tanımlayıcı hem de tahmin edici yöntemlerde birçok amaç için kullanılmaktadır (Giudici,2003:107). Bankalarda kredi isteyenlerin kamu kurumlarında ise ihaleye girenlerin risklerinin değerlendirilmesi (Huang,2004:544), hisse senedi piyasaları tahmini, belirtilere göre hastalık tahmini (Argüden vd.,2008:62), pazarda ürünün performansını öngörme, kredi kartları sahtekarlıklarını saptama, optik karakter tanıma ve çek okuma (Öztemel,2003:36) gibi işlemlerde kullanılabilir.

Birçok farklı alanlarda kullanılan yapay sinir ağlarının veri madenciliği açısından kuvvetli yönleri şunlardır (Tantuğ, 2002:47);

- Çok geniş açıdaki sorunların çözümünde kullanılabilirler,
- Çok karmaşık durumlarda bile iyi sonuçlar üretmektedirler,
- Sayısal ve kategorik veriler üzerinde işlem yapabilirler.

Yapay sinir ağlarının veri madenciliği açısından olumsuz yönleri de vardır (Tantuğ, 2002:47);

- 0 ile 1 arasında giriş verileri olması zorunludur,
- Ürettikleri sonuçların açıklamasını yapamazlar,
- Varılan sonucun olası en iyi sonuç olduğunun teminatı yoktur.

1.8.2.5.Genetik Algoritmalar

Genetik algoritmalar, biyoloji biliminden faydalanarak geliştirilmiş önemli makine öğrenimi yöntemlerinden birisidir (Adriaans vd.,1996:159). Genetik algoritmalar iki konuda doğadan ilham almıştır. Bunlardan birincisi, çok sayıda canlı türlerinden çevreye uyum sağlayanın varlığını sürdürmesi ve uyum sağlayamayanların yok olmasıdır. Bu yaklaşım, problem çözümlerine ilham kaynağı olmuştur. İkinci konu ise, canlıların DNA yapısıdır. Bu yaklaşım ise, kodlama problemleri için ilham kaynağı olmuştur (Yalçıntaş, 2003:32).

Genetik algoritmalar, etkili sonuçlara varmaya çalışan rastgele arama teknikleridir (Saruhan,2004:105). Bu teknikte, bir problemi çözebilmek amacıyla ilk olarak rastgele başlangıç çözümleri tespit edilmektedir. Daha sonra ise bu

çözümler birbirleri ile eşleştirilerek performansı en yüksek çözümler meydana getirilmektedir. Bu şekilde sürekli çözümler birleştirilerek yeni (tanınmayan) çözümler aranmaktadır. Bu arama en iyi sonuç elde edilinceye kadar devam etmektedir. Bu aramayı yapmaktaki amaç, en iyi çözümlerin bulunduğu çözüm kümesini daha iyi hale getirmektir (Gürgen,2008:13).

Genetik algoritmalar, geleneksel yöntemlerle çözümü zor veya olanaksız olan problemlerin çözümünde kullanılmaktadır (Kurt vd.,2011:2). Farklı sonuçları aynı zamanda değerlendirerek arama eğilimini daha iyi çözüm alanlarına yönlendirir. Çoğunlukla bir çözüme odaklanılan problemin en etkili sonucuna varmaya çalışır (Kim vd.,2004:216).

Genetik algoritmaların özellikleri şu şekilde sıralanabilir (Tatei vd.,1994:74):

- İsaletli çözümler için birden çok popülasyon (varlık) olabilir.
- Önceden bilinen çözümlerin niteliklerini bir araya getirerek veya bu özellikleri tesadüfi olarak değiştirerek yeni uygun çözümler üretebilir.
- Popülasyon içinden öncelik vererek, farklı çözümleri seçebilir ve bazı çözümleri gözardı edebilir.

Finans, ekonomi, muhasebe, yöneylem araştırması, ulaştırma, dağıtım, tıp ve siyaset bilimi gibi birçok alanda uygulama imkânı ve uygulamaları olan genetik algoritmaların işleyiş adımları aşağıdaki gibidir (Chambers,1998:31):

- Tüm mümkün çözümler dizi olarak tanımlanır.
- Tesadüfi bir çözüm kümesi seçilir ve başlangıç popülasyonu olarak değerlendirilir.
- Belirlenen çözümler için uygunluk fonksiyonu tanımlanarak, uygunluk değeri hesaplanır ve bu uygunluk fonksiyonlarına göre bireyler seçilir. Seçim işleminde uygun iyi olmayan bireyler elenir ve dizilerin çözüm kalitesi artırılır.
- Çaprazlama ve mutasyon işlemleri ile yeni nesiller oluşturulur.
- Süreç belirlenen nesil sayısına ulaşıncaya kadar devam ettirilir.

1.9. Veri Madenciliği Uygulamalarında Karşılaşılan Problemler

Veri madenciliğinde girdi olarak kullanılacak ham veri veritabanlarından alınır. Bu da küçük veri kümelerinde hızlı ve doğru bir biçimde çalışan bir sistemin, çok boyutlu veri tabanlarına uygulandığında sorun çıkmasına neden olabilir. Bir veri madenciliği sistemi ayıklanmış veri üzerinde mükemmel çalışırken, aynı veriye gürültü eklendiğinde net olmayan sonuçlar oluşabilir. Veri madenciliği sistemlerinde karşılaşılan problem aşağıdaki maddelerde açıklanmıştır (Yurtsever,2002:23).

1.9.1. Veri Tabanı Boyutu

Veri tabanı boyutunun çok büyük olması, veri madenciliği sistemlerinin karşı karşıya olduğu en önemli sorunlardan birisidir. Küçük örneklemi ele alabilecek bir biçimde geliştirilmiş bir algoritmanın, çok büyük örneklerde kullanabilmesi azami dikkat gerektirmektedir. Dolayısıyla veri madenciliği yöntemleri ya sezgisel bir yaklaşımla arama uzayını taramalıdır ya da örnekleme en aza indirilmelidir. Belli bir niteliğin alan değerleri önceden sıra düzensel olarak kategorize edilir. Sonrasında ise, ilgili niteliğin değerleri yatay/dikey olarak aşağıdan yukarıya doğru seviye seviye güncellenir. Yani tekrarlı çokluklar çıkarılarak tekil hale getirilir. Oldukça sağlam bir örneklem kuramı kullanılarak çok büyük boyutlu veri öyle bir boyuta indirgenir ki, hem kaynak veri belirli bir güven aralığında temsil edilir hem de indirgenen veri kümesinin boyutu kullanılan algoritma tarafından işlenebilir hale gelir (Karagöz,2007:18).

1.9.2. Gürültülü Veri

Büyük veri tabanlarındaki değerlerin birçoğu doğru olmayabilir. Bu hata, veri girişi sırasında yapılan insan hataları veya girilen değerlerin hatalı ölçülmesinden kaynaklanır. Veri tabanlarına, veri girişi veya veri toplanması sırasında oluşan sistem dışı hatalara gürültü adı verilir. Hatalı veri gerçek dünya veri tabanlarında ciddi problemler oluşturması veri madenciliği yönteminin kullanılan veri kümesinde bulunan gürültülü verilere karşı daha az duyarlı olmasını gerektirir. Gürültülü verinin sebep olduğu problemler tümevarımsal karar ağaçlarında uygulanan metotlar bağlamında kapsamlı bir biçimde

araştırılmıştır. Eğer veri kümesi gürültülü ise sistem bozuk veriyi tanımalı ve yok saymalıdır (Ay,2009:23).

1.9.3.Boş (Null) Değerler

Boş (null) değer, kendisi de dâhil olmak üzere hiçbir değere eşit olmayan değerdir. Diğer bir anlatımla bu kavram verinin içeriğinin bilinmemesi anlamını taşımaktadır. Eğer çoklu veri üzerinde bir nitelik değeri boş (null) ise o nitelik bilinmeyen ve uygulanamaz bir değere sahiptir. Bu durumla ilişkisel veri tabanlarında çoğu kez karşılaşılmaktadır. Bir ilişkide bulunan tüm çokluklar eş değer sayıda niteliğe, niteliğin değeri boş (null) olsa bile sahip olmalıdır. Örneğin; masaüstü bilgisayarların özelliklerini tutan bir ilişkide bazı model bilgisayarlar için ses kartı modeli niteliğinin değeri boş (null) olabilir. Boş değerli nitelikler veri kümesinde bulunuyorsa ya boş değerli veriler yok sayılarak, algoritma içinde ihmal edilmeli ya da olası bir değerle değiştirilmelidir (Ay,2009:24).

1.9.4.Eksik Veri

Eksik veriler, yapılacak olan istatistiksel analizlerde ciddi problemler yaratmaktadır. Çünkü istatistiksel analizler ve bu analizlerin yapılmasına olanak sağlayan ilgili paket programlar, verilerin tümünün var olduğu durumlar için geliştirilmiştir. Bu analizlerin eksik veri içeren veri setlerine uygulanması istatistiklerin geçerliliğini düşürmektedir (Akçay,2013:17).

1.9.5.Artık Veri

Verilen veri kümesi, eldeki mevcut probleme uygun olmayan veya işe yaramayan artık nitelikler içerebilir. Bu durum işlem sırasında sıkça karşımıza çıkabilir. Örneğin, eldeki mevcut problem ile ilgili veriyi elde etmek için iki ilişkiyi ortak nitelikler üzerinden birleştirirsek, sonuç ilişkide kullanıcının farkında olmadığı artık nitelikler bulunur. Artık nitelikleri elemek için geliştirilmiş algoritmalar özellik seçimi olarak tanılandırılır. (Sever vd., 2002:173).

1.9.6.Dinamik Veri

Kurumsal çevrim içi veri tabanlarının dinamik (içeriğinin sürekli olarak değişmesi) olması, bilgi keşfi metotları için ciddi sakıncalar doğurmaktadır. Sadece okuma yapan ve uzun süre çalışan bilgi keşfi metodu bir veri tabanı uygulaması olarak eldeki veri tabanı ile beraber çalıştırıldığında mevcut uygulamanın da performansı önemli ölçüde düşer. Veri tabanında bulunan verilerin kalıcı olduğu kabul edilip, çevrimdışı veri üzerinde bilgi keşif metodu çalıştırıldığında, değişen verinin elde edilen örüntülere yansımaları gerektiğinden bu işlem, bilgi keşfi metodunun ürettiği örüntüleri zaman içinde değişen veriye göre sadece ilgili örüntüleri yığılmalı olarak güncelleme yeteneğine sahip olmasını gerektirir (Karagöz,2007:19).

1.9.7.Farklı Tipteki Verileri Ele Alma

Gerçek hayattaki uygulamalar makine öğrenmesinde olduğu gibi, sadece sembolik veya kategorik veri türleri değil aynı zamanda tamsayı, kesirli sayı, çoklu ortam verisi, coğrafi bilgi içeren veri gibi farklı tipteki veriler üzerinde işlem yapılmasını gerektirir. Bununla beraber veri çeşitliliğinin fazla olması bir veri madenciliği algoritmasının tüm veri tiplerini ele alabilmesini olanaksız hale getirmektedir. Bu yüzden veri tipine mahsus, veri madenciliği algoritmaları geliştirilmektedir (Akçay,2013:17).

1.9.8.Sınırlı Bilgi

Veri tabanları genel olarak veri madenciliği dışındaki hedefler için tasarlanmışlardır. Bu sebeple, öğrenme görevini kolaylaştıracak bazı özellikler bulunmayabilir (Akbulut,2006:12).

1.10.Verit Madenciliği Alanında Kullanılan Yazılımlar

Veri madenciliği uygulamalarını gerçekleştirmek için bilgisayar programı kullanmak gereklidir. Bu kapsamda farklı algoritmaları bünyesinde barındıran ve farklı işletim sistemleri üzerinde çalışabilen birçok veri madenciliği yazılımı geliştirilmiştir. Tablo 9'da, bu yazılımlardan bazıları, üzerinde çalıştıkları platform ve fonksiyonel nitelikleri ile birlikte gösterilmiştir (Cevahir,2011:12).

Tablo 10:Veri Madenciliğinde Kullanılan Bazı Yazılımlar ve Özellikleri

Ürün Adı	IBM Intelligent Miner	Oracle Darwin	SAS Enterprise Miner	SPSS Clementine	WEKA
Platform	AIX 4.1. NVS AS/400 Windows	Windows	Windows Macintosh Unix	Windows Macintosh Unix	Windows Macintosh Unix
Karar Ağacı	x	x	x	x	x
Sinir Ağları	x	x	x	x	x
Zaman Serileri	x		x	x	
Tahmin	x	x	x	x	x
Kümeleme	x		x	x	x
Birliktelik	x		x	x	x
Görselleştirme	x	x	x	x	x

(Kaynak: Cevahir, F., 2011, “Bir Perakende Firmasına Ait Veriler Üzerinden Veri Madenciliği Uygulaması”, Mimar Sinan Güzel Sanatlar Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, s.12)

1.10.1.Intelligent Miner

Eşleştirme, sınıflama, tahmin modelleme, gruplama, sıralı desen analizi, regresyon analizi gibi fonksiyonları barındıran IBM şirketinin geliştirmiş olduğu bir yazılımdır. Bununla birlikte, yapay sinir ağları algoritmaları, istatistik metotları, veri hazırlama ve görsel gösterim araçları gibi ek özellikler içermektedir (Cevahir,2011:12).

1.10.2.Darwin

Darwin Oracle şirketinin geliştirmiş olduğu bir veri madenciliği aracıdır. Darwin regresyon ağaçları, karar ağaçları, kümeleme, yapay sinir ağları, Bayesian öğrenme, k-en yakın komşu gibi pek çok algoritmayı destekleyen paralel

sunucular için geliştirilmiş olan veri madenciliği yazılımıdır. Darwin kullanımı basit bir ara yüze sahiptir. Darwin veri madenciliği algoritmalarından CART karar ağacı, StarTree, StarNet ve StarMatch'i için kullanılmaktadır (Ay,2009:26).

1.10.3.Enterprise Miner

SAS şirketinin geliştirmiş olduğu veri madenciliği aracıdır. İşletmelerin çok boyutlu veri yığınlarından kritik bilgileri elde etmelerini sağlayan veri madenciliği çözümlerinde dünyada önemli bir yere sahip olan SAS şirketi, veri üzerinde değil, bilgi üzerinde düşünme ve strateji geliştirme avantajını bir adım öne taşıyarak SAS Enterprise Miner yazılımını geliştirmiştir. Enterprise Miner karar ağaçları, yapay sinir ağları, regresyon, kümeleme, zaman serileri, sınıflama, birliktelik kurallarının bulunması gibi veri madenciliği sorgularını ele alabilmektedir. Grafikselleştirilmiş ara yüzü sayesinde uygulamada kullanım kolaylığı sağlarken kullanıcılar uygulamanın karmaşıklığından habersiz bir şekilde sadece girdi ve çıktılara yoğunlaşabilirler Örneğin; Garanti Bankası müşterilerine sunmuş olduğu hizmetleri daha iyi bir noktaya taşımak amacıyla SAS'ın veri madenciliği ürününü tercih etmiştir. Müşterilerin finansal davranış modelleri ile ilgili detaylı bilgi elde etmeyi amaçlayan Garanti Bankası, böylelikle daha etkin müşteri ilişkileri yönetimi çalışmaları yapabilecektir (Cevahir,2011:13;Çingir,2010:10).

1.10.4.SPSS Clementine

Clementine SPSS şirketinin geliştirmiş olduğu istatistiksel analize yönelik veri madenciliği aracıdır. İngiltere'de geliştirilen SPSS Clementine, 1980 yılından beri kullanılmakta ve araçları giderek iyileştirilmektedir. SPSS Clementine yapay sinir ağları, karar ağaçları, zaman serileri, kümeleme, regresyon, birliktelik kuralları algoritmaları mevcuttur. SPSS Clementine pazar araştırmaları, sağlık araştırmaları, müşteri ilişkileri yönetimi, kimya sektöründe maddelerin aşındırıcılık tahmini, bankacılık alanında kredi kartı dolandırıcılıkları ve anket çalışmaları gibi konularda kendine uygulama alanı bulmuştur (Ay,2009:25).

SPSS Clementine yazılımının avantajları aşağıdaki gibidir (Dolgun,2006:51);

- Veri madenciliği algoritmaları çok çeşitlidir ve birçok veri madenciliği algoritmasının içine yerleştirilmiş oldukça güçlü, en iyi parametreyi araştıran programlar bulunmaktadır,
- Veri kalitesinin kontrolü ve kayıp gözlem tahmini için araçlar bulunmaktadır,
- Karmaşık değişken süreçleri için güçlü dâhili yazı dili bulunmaktadır,
- Ortalama kullanım kolaylığı mevcuttur.

SPSS Clementine yazılımının dezavantajları aşağıdaki gibidir (Yurtsever,2002:52);

- Diğer programlara göre az sayıda betimsel istatistik ve parametrik çözümler barındırır,
- Diğer programlara göre grafik özelliği ve betimsel istatistik açısından zayıftır.

1.10.5.Weka

Weka; Yeni Zelanda Waikito Üniversitesinde bir proje olarak başlayıp bugün dünya üzerinde birçok insan tarafından kullanılmaya başlanan bir veri madenciliği uygulaması programıdır. Weka, obje yönelimli programlama dillerinden biri olan Java platformu üzerinde geliştirilmiş ve halen yeni sürümleri geliştirilmeye devam eden açık kodlu bir programdır. Weka yazılımının Java ile geliştirilmiş olması, Linux, Unix, Windows ve Macintosh gibi işletim sistemleri ile kullanılabilirliğini sağlamıştır (Cevahir,2011:13). Weka, içerisinde veri işleme, veri sınıflandırma, veri kümeleme, veri ilişkilendirme özelliklerini barındırmaktadır(Akçay,2013:22).

İKİNCİ BÖLÜM

UYGULAMA

2. MARKET VERİTABANINDA BİRLİKTELİK KURALLARININ BELİRLENMESİ

Marketlerde satılan ürünlerin birbirleri ile olan ilişkilerini yani hangi ürünün hangi ürünle birlikte satılma eğiliminde olduğunu belirlemek günümüzde yöneticilerin market stratejilerini belirlemesinde oldukça önemli rol oynamaktadır. Bu amaçla market yöneticileri birliktelik analizi yapılan ürünlerin raflara yerleşimini yan yana ya da birliktelikleri tespit edilen ürünlerin arasına görülmesi veya fark edilmesi istenen başkaca ürünler yerleştirerek ürünlerin satışlarını arttırabilirler.

Bu uygulamanın amacı, bir marketteki alışveriş hareketlerinden oluşan kayıtları kullanarak hangi ürünlerin hangi ürünle birlikte satın alınabileceği bilgisini vermek ve bu bilgiler ışığında market yöneticisine yarar sağlayacak yeni bir raf yerleşim düzeni önerisi geliştirmektir.

Uygulama için SPSS firmasının veri madenciliği çözümü için geliştirmiş olduğu SPSS Clementine 11.1 paket programının Apriori Algoritması, veri tabanı olarakta SQL Server 2005 kullanılacaktır.

2.1. Veri Madenciliği Süreci

2.1.1. Verilerin Toplanması ve Verilerin Düzenlenmesi

Bu uygulamada kullanılan veri seti Türkiye’de perakende sektöründe faaliyet gösteren market zincirlerinden bir tanesinin, 2016 Ocak-Şubat ayı içerisindeki 7 günlük (28.Ocak.2016-03.Şubat.2016) zaman diliminde toplanan alışveriş hareketlerinden oluşmaktadır. Gizlilik prensipleri sebebi ile firma ismi çalışmada belirtilmeyecektir. Bu veriye ait örnek görüntüler Şekil 12’de görülmektedir.

Alışveriş verisinin her satırında işlem tarihi, işlem yapılan kasa numarası, günlük işlem numarası, işlem saati, alınan ürün kodu, alınan ürün markası, alınan

ürün adı, ürün adet/kg, ürün tutarı ve işlem numarasında yapılmış toplam harcama miktarı bilgileri bulunmaktadır.

	TARİH	KASA NO	SAAT	İŞLEM NO	ÜRÜN NO	MARKA	İÇERİK	ADET/KG	TUTAR	HARCAMA
3	28.01.2016	2	13:09	62	8697997000292	Malatya Pazarı	Kuru Kayısı	1	13,95	167,74
4	28.01.2016	2	13:09	62	8690914003529	Bumaş	Bulgur	1	2,15	167,74
5	28.01.2016	2	13:09	62	8690637049903	Sana	Margarin	1	6,40	167,74
6	28.01.2016	2	13:09	62	8690637501705	Calve	Mayonez	1	6,10	167,74
7	28.01.2016	2	13:09	62	8694274100041	Günlük	Yufka	1	3,50	167,74
8	28.01.2016	2	13:09	62	8694274100041	Günlük	Yufka	1	3,50	167,74
9	28.01.2016	2	13:09	62	8690565100530	Pınar	Süt	1	2,25	167,74
10	28.01.2016	2	13:09	62	8690530222502	Selpak	Kağıt Havlu	1	5,50	167,74
11	28.01.2016	2	13:09	62	8690347030901	Kaanlar	Zeytin	1	11,95	167,74
12	28.01.2016	2	13:09	62	8690565100530	Pınar	Süt	1	2,25	167,74
13	28.01.2016	2	13:09	62	8690576029189	Ankara	Makarna	1	0,99	167,74
14	28.01.2016	2	13:09	62	2223782		Lahana	4	4,00	167,74
15	28.01.2016	2	13:09	62	2223785		Maydanoz	1	0,99	167,74
16	28.01.2016	2	13:09	62	2223786		Soğan	1	1,65	167,74
17	28.01.2016	2	13:09	62	8690829229405	Dr.Oetker	Krem Şanti	1	5,10	167,74
18	28.01.2016	2	13:09	62	2923590		Domates	1,252	5,57	167,74
19	28.01.2016	2	13:09	62	2900603		Patates	1	7,50	167,74
20	28.01.2016	2	13:09	62	8690637548024	Cif	Deterjan	1	6,45	167,74
21	28.01.2016	2	13:09	62	8690570502685	Finish	Bulaşık Deterjanı	1	17,95	167,74
22	28.01.2016	2	13:09	62	8690637610950	Cif	Deterjan	1	10,90	167,74
23	28.01.2016	2	13:09	62	8690637616426	Domestos	Çamaşır Suyu	1	9,95	167,74

Şekil 12: Market Alışveriş Kayıtlarına Ait Örnek Veri Seti

Şekil 12’de gösterilen veri setine ait başlıca özellikler şunlardır:

- 19344 adet satırdan oluşmaktadır.
- 28.Ocak.2016 Perşembe – 03.02.2016 Çarşamba tarihleri arasındaki alışveriş bilgilerini içermektedir.
- 1493 adet hareketten (birlikte alınma işlemi) oluşmaktadır.
- 288 adet farklı üründen oluşmaktadır.
- En erken yapılan alışveriş saat 08.06’ da (02.Şubat.2016), en geç yapılan alışveriş saat 21.54’ de (02.Şubat.2016) gerçekleşmiştir.
- 214 adet farklı işlem numarası 1- 292 arasında değişmektedir.

Veri setinden birliktelik kuralları elde edileceğinden dolayı, toplanan 1493 fiş içerisinde uygun elemeler yapılarak fiş sayısı 511’e düşürülmüştür. 511 fişe gözlemlenen 132 ürün grubu bulunmaktadır. Daha sonra bu ürün grupları

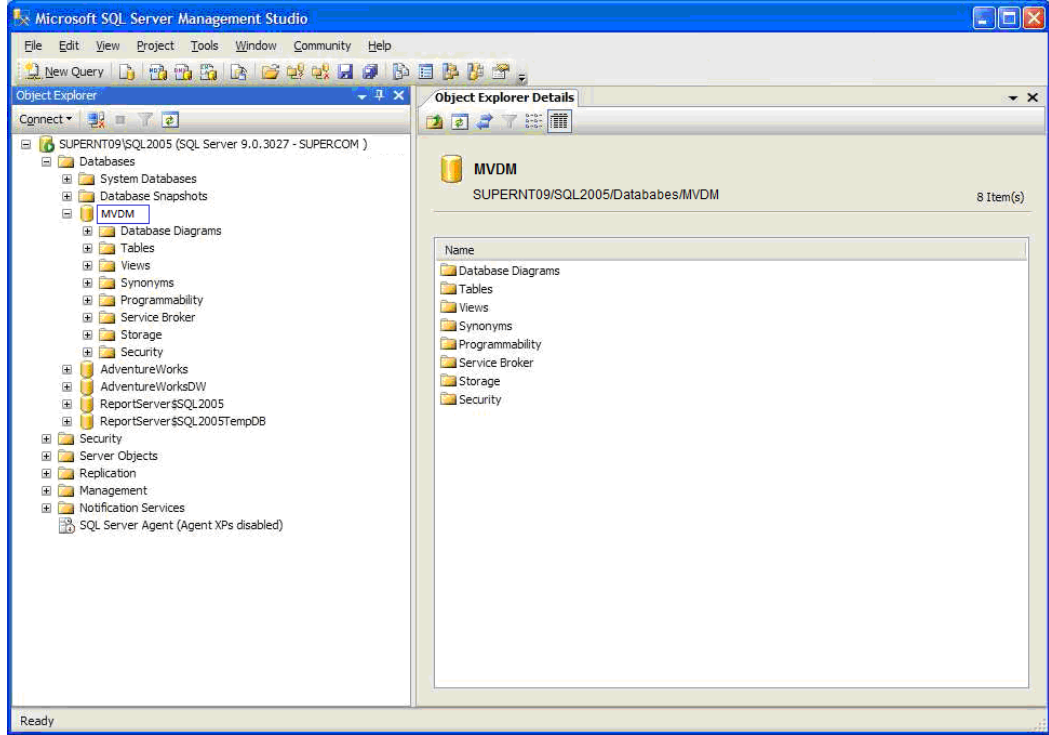
birleştirilerek, birliktelikleri daha iyi tanımlayabilmek için 20 ürün kategorisi oluşturulmuştur. Oluşturulan ürün kategorileri aşağıdaki tabloda gösterilmiştir.

Tablo 11: Ürün Kategorileri

Ürün Kategorileri
Atıştırmalık Ürünler
Bebek Ürünleri
Çay-Şeker-Kahve
Et Ürünleri
Gıda ve Temizlik Dışı Ürünler
Hazır-Konserve ve Dondurulmuş Ürünler
İçecekler
Kahvaltılık Ürünler
Kozmetik Ürünleri
Kuru Gıda Ürünleri
Makarna ve Soslar
Meyve
Sebze
Sıvıyağ ve Margarin
Süt
Şekerli Ürünler
Temizlik Ürünleri
Tuzlu Ürünler
Unlu Mamuller
Yoğurt ve Çeşitleri

2.1.2. Veri Tabanının Oluşturulması

Excel formatındaki “xlsx” uzantılı market alışveriş veri setinin üzerinde işlem ve sorgulamalar yapıp, Transact SQL dili ile yazılmış algoritmalarla sonuçlar alabilmek için Microsoft SQL Server 2005’e ait SQL Server Management Studio (SSMS) üzerinde “MVDM” (Market Veriseti Data Mining) veri tabanı oluşturulmuştur.



Şekil 13: SQL Server Management Studio’da Oluşturulmuş Olan MVDM Veri Tabanı

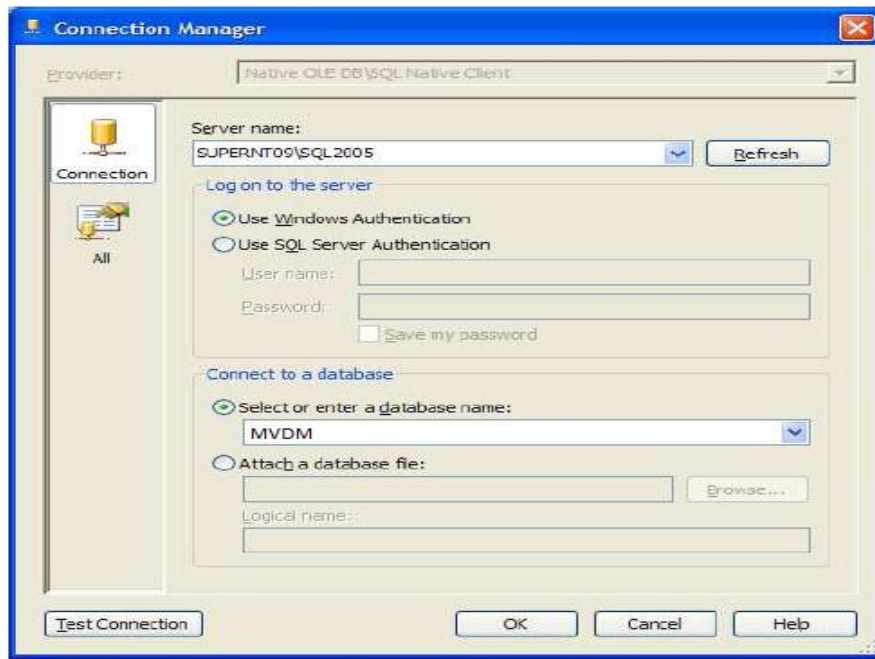
Excel formatındaki Market alışveriş veri setini MVDM veri tabanına alabilmek için öncelikle, Excel’deki sütun sırasını ve isimlendirmelerinin benzerini barındıracak yapıda t_market_veri_seti isimli bir tablo oluşturulmuştur. t_market_veri_seti tablosunda Tarih, Kasa_No, Islem_No, Saat, Urun_No, Urun_Marka, Urun_Adi, Urun_Adet_Kg, Urun_Tutar, Harcama isimli sütunlar yer almaktadır. Tabloların yapısı Şekil 14’de görülmektedir.

	Column Name	Data Type	Allow Nulls
▶	Hesap No	decimal (10,0)	<input type="checkbox"/>
	Tarih	datetime	<input type="checkbox"/>
	Kasa_No	tinyint	<input type="checkbox"/>
	Islem_No	int	<input type="checkbox"/>
	Saat	decimal (4,0)	<input type="checkbox"/>
	Urun_No	decimal (10,0)	<input type="checkbox"/>
	Urun_Marka	nvarchar (50)	<input type="checkbox"/>
	Urun_Adi	nvarchar (50)	<input type="checkbox"/>
	Urun_Adet_Kg	decimal (10,3)	<input type="checkbox"/>
	Urun_Tutar	decimal (10,2)	<input type="checkbox"/>
	Harcama	decimal (10,2)	<input type="checkbox"/>

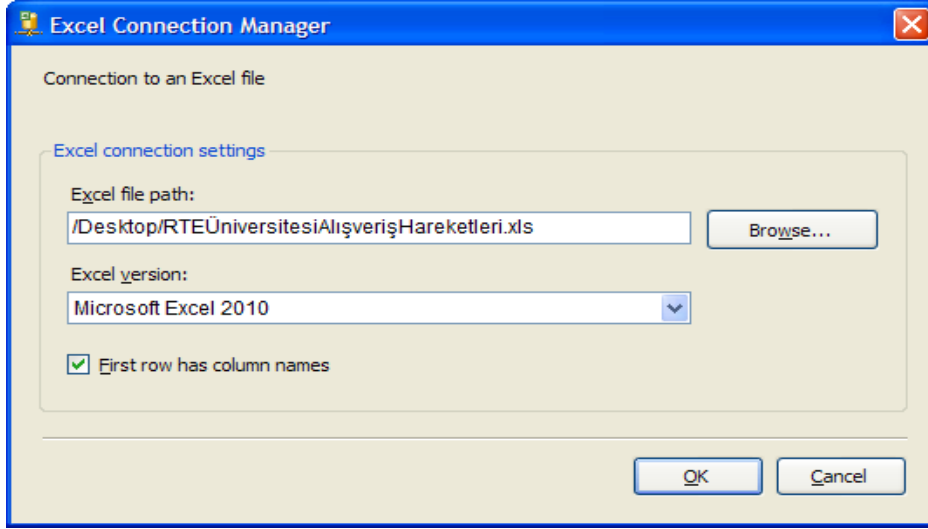
Şekil 14: MVDM Veri Tabanındaki t_market_veri_seti Tablosunun Yapısı

Oluşturulan t_market_veri_seti tablosuna Excel formatındaki veri setinin atılabilmesi için, farklı kaynaklar arasında veri aktarımına olanak sağlayan Microsoft SQL Server Integration (SSIS) kullanılmıştır.

Gerçekleştirilecek aktarım işlemini tasarlamak için SQL Server Business Intelligence Development Studio'da MVDMSIS isimli bir Integration Services Project oluşturulmuştur. Proje içinde SSIS Packages altında oluşturulan pMarketSatis.dtsx'te, Excel ve SQL Server 2005 veri tabanı bağlantılarını sağlamak için bağlantı nesneleri oluşturulmuştur.

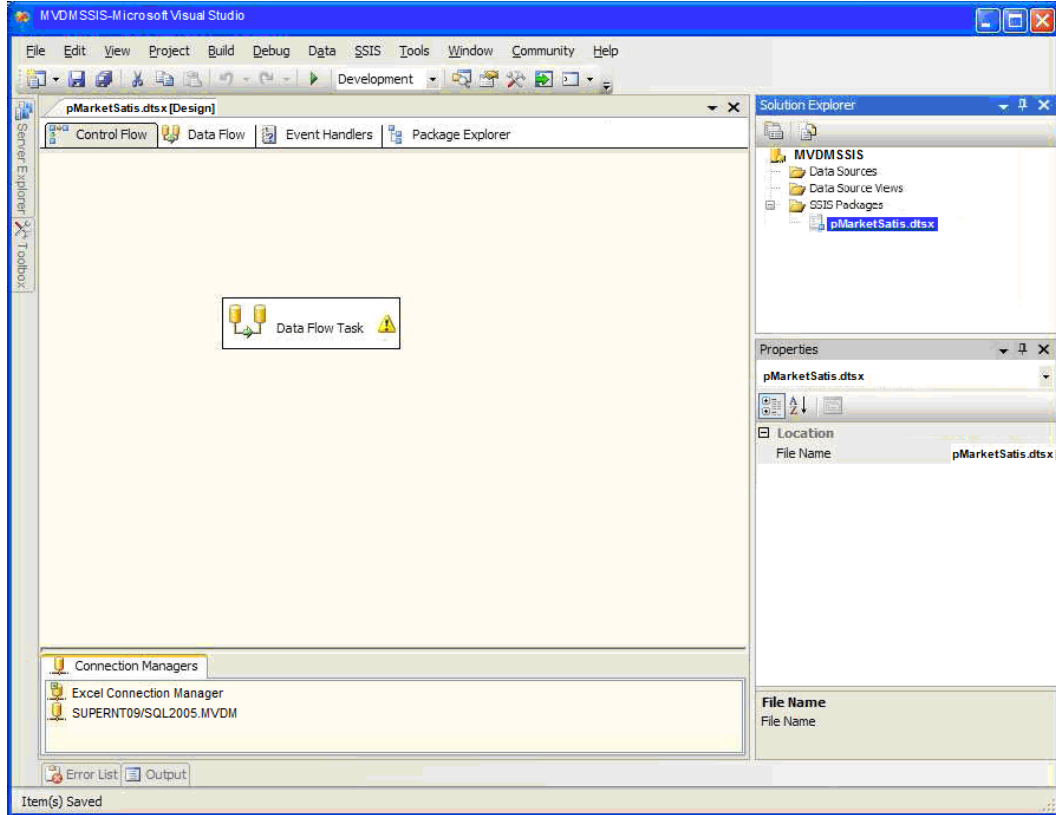


Şekil 15: MVDMSIS Projesinde Yer Alan Excel ve SQL Server 2005 Veri Tabanı Bağlantı Nesnelерinin Yapılandırılması Adımı 1

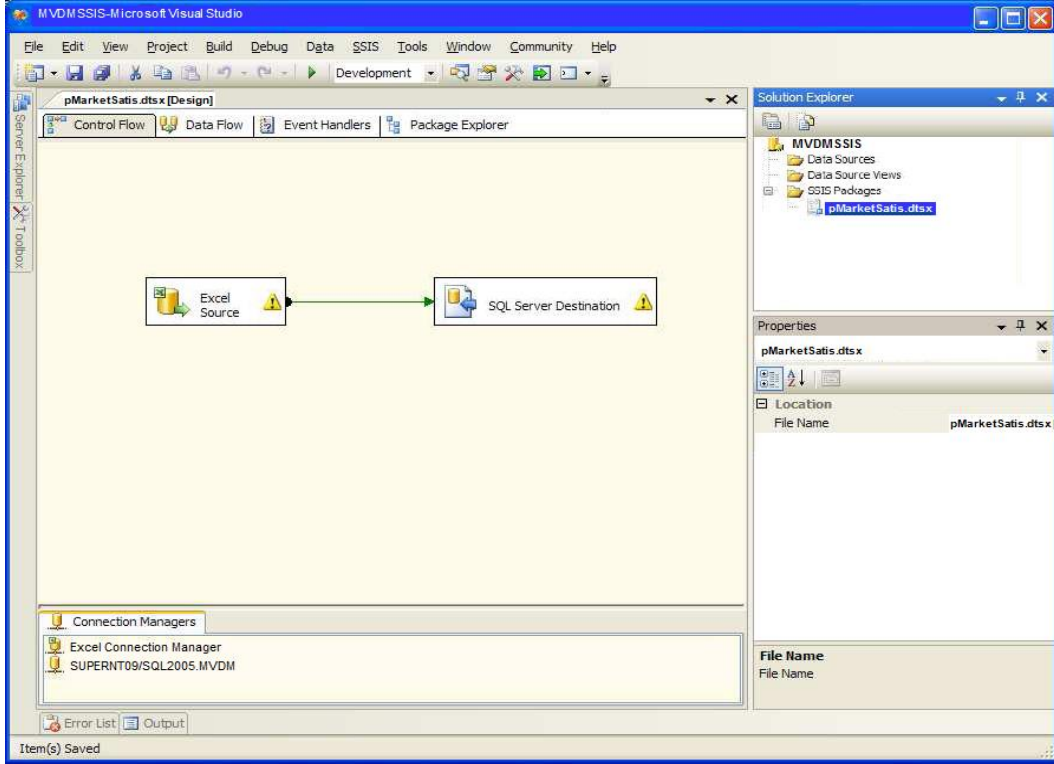


Şekil 16: MVDMSIS Projesinde Yer Alan Excel ve SQL Server 2005 Veri Tabanı Bağlantı Nesnelerinin Yapılandırılması Adımı 2

Sonrasında, bu bağlantı nesnelerini kullanarak Excel'deki Market Alışveriş Hareketleri örnek veri setinin bulunduğu dosyanın t_market_veri_seti tablosuna atılması için Data Flow Task ve içerikleri tasarlanmıştır. pMarketSatis.dtsx package'ının yapısı Şekil de verilmiştir.



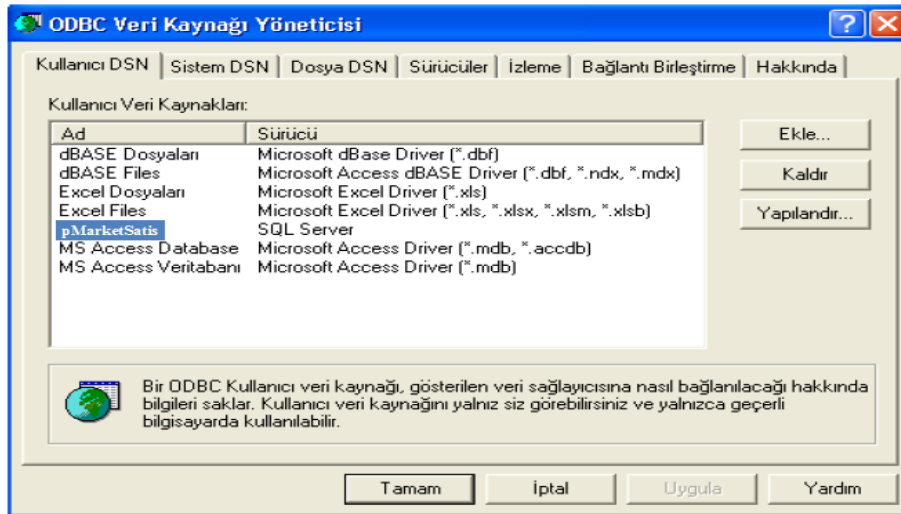
Şekil 17: pMarketSatis.dtsx Package'ı İçinde Yer Alan Data Flow Task ve Bağlantıları



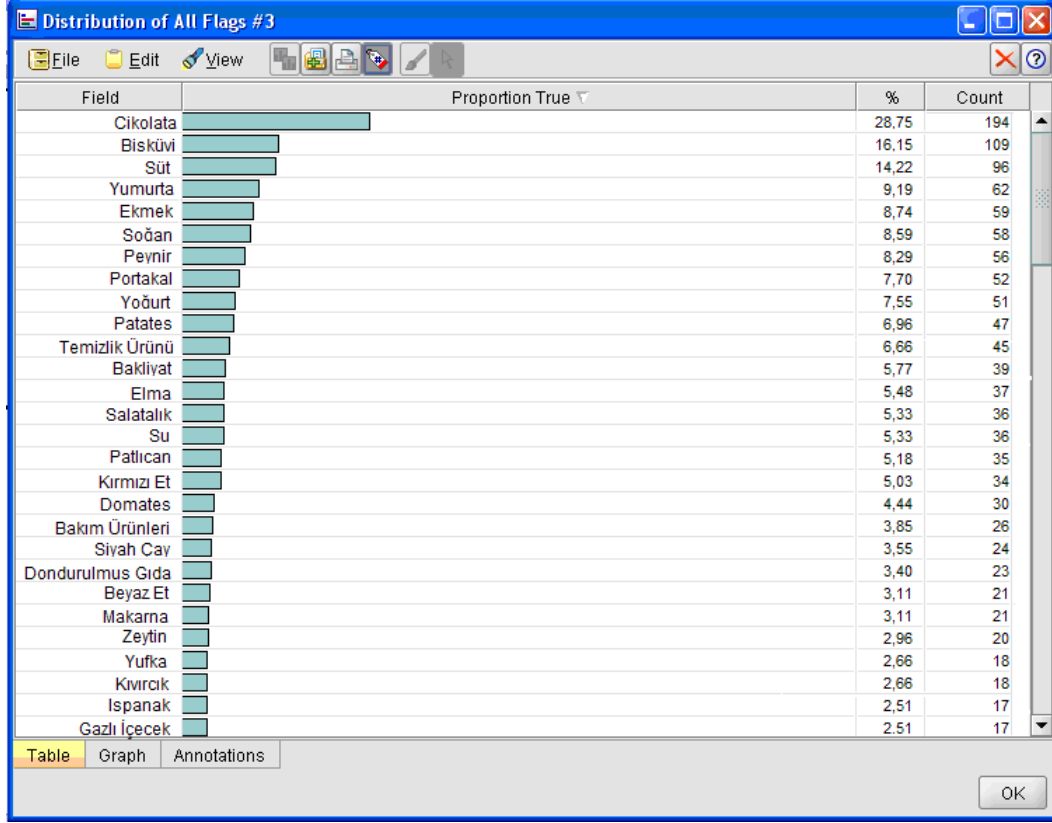
Şekil 18: pMarketSatis.dtsx'teki Data Flow Task'ın İçeriği

2.1.3.SPSS Clementine Programında Verinin Hazırlanması

Clementine' da veri kaynağına erişim sağlanması için öncelikle sırasıyla denetim masası-yönetimsel araçlar-veri kaynakları (ODBC) menüsü izlenerek SQL' de kullanılmış olan veri tabanının eklenmesi gerekmektedir. Şekil 19'da pMarketSatis veri tabanının eklenmiş hali verilmiştir.

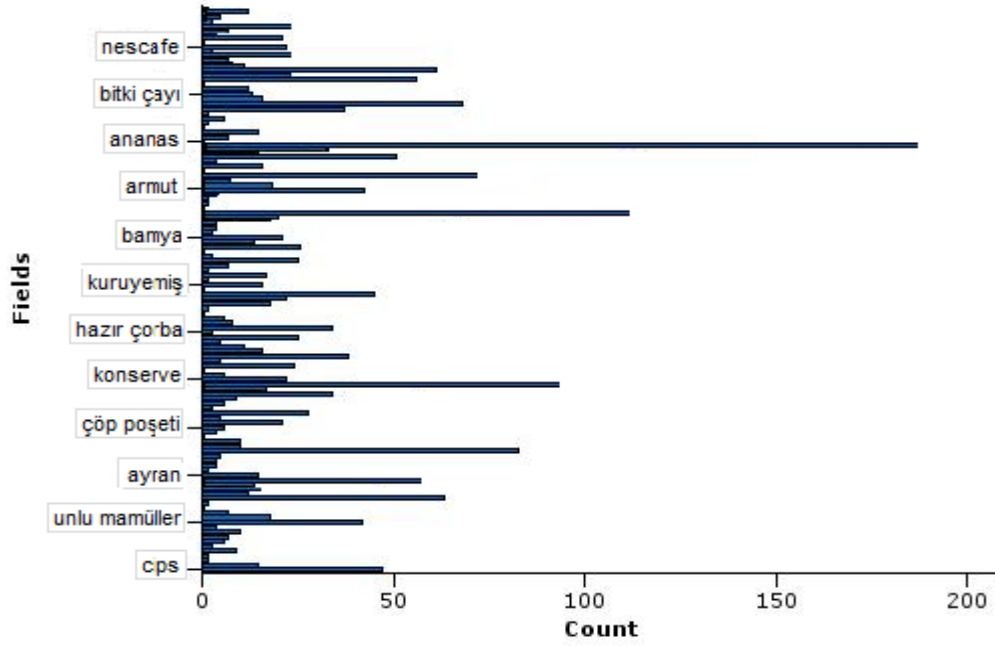


Şekil 19: SQL Veri Tabanının Veri Kaynağı Yöneticisine Eklenmesi



Şekil 21: Ürün Gruplarının Alışveriş Hareketlerindeki Dağılımı

Şekil 21’den de görüldüğü gibi alışveriş fişlerinde en sık tekrarlanan ürünler çikolata (%28,75), bisküvi (%16,15), süt (14,22), yumurta (%9,19), ekmek (%8,74), soğan (%8,59), peynir (%8,29), portakal (%7,70) olarak sıralanabilir. Bununla birlikte tüm ürün gruplarının dağılımı ayrı bir şekil üzerinde Şekil 22’de gösterilmektedir.



Şekil 22: Alışveriş Hareketlerindeki Tüm Ürün Gruplarının Dağılımı

132 ürün grubu için Apriori algoritması kullanılarak aşağıdaki gibi birliktelik kuralları elde edilmiştir. Analizde minimum destek değeri %5 ve minimum güven değeri %50 olarak alınmıştır. Analizde yalnızca müşterinin satın alınan ürünler arasındaki birliktelik kuralları incelenmiş ve satın alınıp-alınmayan ürünler arasındaki örüntü analizinin yapılmasına çalışmanın amacı doğrultusunda gerek duyulmamıştır. Bu doğrultuda yapılan analiz sonucunda, veri seti içerisinde 21 adet birliktelik kuralı elde edilmiş ve bu kurallar Tablo 12’de verilmiştir.

Tablo 12: Ürün Grupları İçin Elde Edilen Birliktelik Kuralları

Kural	Önce Gelen (Antecedent)	Takip Eden (Consequent)	Destek % 5 (Support)	Güven % 50 (Confidence)
1	Meyve Suyu	Bisküvi-Çikolata	21,912	79,910
2	Yumurta	Peynir-Zeytin	13,208	70,456
3	Meyve Suyu	Yumurta ve Yoğurt	11,007	68,637
4	Yoğurt	Yumurta ve Meyve Suyu	12,157	65,141
5	Meyve Suyu	Peynir-Zeytin	12,607	61,922
6	Yoğurt	Peynir-Zeytin ve Yumurta	9,306	60,878
7	Meyve Suyu	Yumurta	21,112	59,984
8	Meyve Suyu	Su	12,557	58,973
9	Meyve Suyu	Yoğurt	24,613	57,489
10	Yumurta	Yoğurt ve Meyve Suyu	13,658	57,312
11	Meyve Suyu	Temizlik Ürünleri	10,706	56,574
12	Meyve Suyu	Süt Ürünleri	10,106	56,476
13	Meyve Suyu	Bakliyat	12,607	56,382
14	Yoğurt	Yumurta	21,112	56,134
15	Meyve Suyu	Unlu Mamuller	11,807	55,120
16	Yoğurt	Peynir-Zeytin	13,208	54,137
17	Peynir-Zeytin	Yumurta ve Yoğurt	11,007	52,456
18	Ekmek	Temizlik Ürünleri	6,975	51,953
19	Ekmek	Yoğurt	7,894	51,536
20	Bakliyat	Ekmek	5,658	51,119
21	Süt	Ekmek	10,844	51,089

Bu veriler doğrultusunda ürünler için oluşturulan bazı birliktelik kuralları aşağıdaki gibidir;

Meyve Suyu → Bisküvi-Çikolata kuralı için;

Meyve Suyu ve Bisküvi-Çikolata ürünlerinin toplam fiş hareketlerinde birlikte görülme olasılıkları %21,92'dir. Meyve Suyu alan müşterilerin %79,91 olasılıkla Bisküvi-Çikolata da aldığı söylenebilir.

Yumurta →Peynir-Zeytin kuralı için;

Yumurta ve Peynir-Zeytin ürünlerinin toplam fiş hareketlerinde birlikte görülme olasılıkları %13,21'dir. Yumurta alan müşterilerin %70,46 olasılıkla Peynir-Zeytin de aldığı söylenebilir.

Meyve Suyu →Bakliyat kuralı için;

Meyve Suyu ve Bakliyat ürünlerinin toplam fiş hareketlerinde birlikte görülme olasılıkları %12,61'dir. Meyve Suyu alan müşterilerin %56,39 olasılıkla Bakliyatta aldığı söylenebilir.

Yoğurt →Yumurta kuralı için;

Yoğurt ve Yumurta ürünlerinin toplam fiş hareketlerinde birlikte görülme olasılıkları %21,12'dir. Yoğurt alan müşterilerin %56,14 olasılıkla Yumurta da aldığı söylenebilir.

Süt →Ekmek kuralı için;

Süt ve Ekmek ürünlerinin toplam fiş hareketlerinde birlikte görülme olasılıkları %10,85'dir. Süt alan müşterilerin %51,09 olasılıkla Ekmek de aldığı söylenebilir.

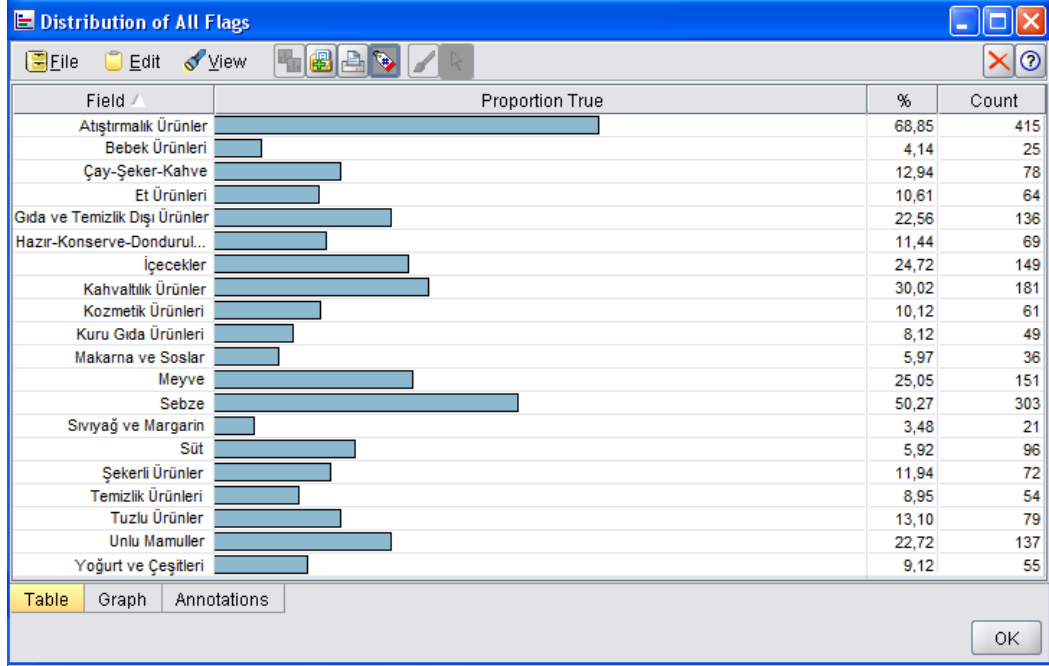
Ürün bazında yapılan bu analiz yerine oluşturulan ürün kategorileri ile yapılacak bir analiz daha net birliktelik kuralları elde etmemizi sağlayabilir. Bu sebeple 132 ürün grubu 20 kategoride toplanmış ve tekrar birliktelik analizine tabi tutulmuştur.

	Hareket..	Gıda_ve..	Sebze	Meyve	Margarin_	Süt	Kuru_Gi..	Kahvaltılı	Makarna	Temizlik..
1	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
2	2.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
3	3.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
4	4.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	5.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
6	6.00	1.00	0.00	1.00	1.00	0.00	1.00	0.00	0.00	1.00
7	7.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
8	8.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9	9.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	10.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
11	11.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
12	12.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
13	13.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
14	14.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
15	15.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
16	16.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
17	17.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
18	18.00	0.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00
19	19.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
20	20.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00

Şekil 23: Ürünlerin Kategorileşmiş Veri Seti

Ürün grupları birleştirilerek müşterilerin satın almadığı ürünleri değerleri “0” (boş değer), satın aldığı ürünlerin değerleri “1” olarak verilmek suretiyle kategorileştirilmiş veri seti oluşturulmuştur. Oluşturulan veri seti Şekil 23’te görülmektedir.

Ürün kategorilerinin satın alınma oranları ve satın alınma frekanslarını gösteren dağılım listesi Şekil 24’de gösterilmektedir.



Şekil 24: Ürün Kategorilerinin Alışveriş Hareketlerindeki Dağılımı

Şekil 24 incelendiğinde alışveriş fişlerinde en sık satın alınan ürün kategorilerinin Atıştırılmalık Ürünler (%68,85), Sebze (%50,27), Kahvaltılık Ürünler (%30,02), Meyve (%25,05) ve İçecekler (%24,72) olduğu görülmektedir.

Ürün kategorisi için birliktelik kuralları Apriori Algoritması kullanılarak elde edilmiştir. Ürün kategorilerinde olabildiğinde fazla kural görebilmek için minimum destek değeri %3 ve minimum güven değeri %25 olarak belirlenmiş analiz bu kısıtlar içerisinde gerçekleştirilmiştir. Bunun sonucunda Tablo 13'deki gibi destek ve güven değerlerini gösteren 35 adet birliktelik kuralları elde edilmiştir.

Tablo 13: Ürün Kategorileri İçin Elde Edilen Birliktelik Kuralları

Kural	Önce Gelen (Antecedent)	Takip Eden (Consequent)	Destek % 3 (Support)	Güven % 25 (Confidence)
1	Kahvaltılık Ürünler	Süt	47,850	61,889
2	Meyve	Sebze ve Unlu Mamuller	43,758	60,855
3	Et Ürünleri	Kuru Gıda Ürünleri	46,963	59,773
4	Makarna ve Soslar	Et Ürünleri	4,408	58,000
5	Meyve	Unlu Mamuller	29,003	52,801
6	Meyve	Sebze	26,507	50,869
7	Meyve	Kahvaltılık Ürünler	11,748	46,032
8	Atıştırmalık Ürünler	Çay-Şeker-Kahve	9,985	45,823
9	Meyve	Hazır-Konserve ve Dondurulmuş Ürünler	6,781	42,555
10	Temizlik Ürünleri	Kozmetik Ürünler	7,826	41,355
11	Süt	Unlu Mamuller	12,451	40,702
12	Meyve	İçecekler ve Unlu Mamuller	7,590	38,998
13	Süt	İçecekler	10,089	36,000
14	Meyve	Atıştırmalık Ürünler	9,619	36,963
15	Unlu Mamuller	Meyve	46,344	35,658
16	İçecekler	Tuzlu Ürünler	24,873	34,002
17	Unlu Mamuller	Kozmetik Ürünleri	7,912	33,331
18	Tuzlu Ürünler	İçecekler	25,102	32,872
19	Yoğurt ve Çeşitleri	Atıştırmalık Ürünler	9,752	31,073
20	Sebze	Hazır-Konserve ve Dondurulmuş Ürünler	7,258	30,112
21	Sıvıyağ ve Margarin	Hazır-Konserve ve Dondurulmuş Ürünler	7,259	30,111
22	Tuzlu Ürünler	Atıştırmalık Ürünler	9,574	30,035
23	İçecekler	Atıştırmalık Ürünler	8,147	30,028

24	Tuzlu Ürünler	Süt	11,048	29,999
25	Yoğurt ve Çeşitleri	Süt	10,042	29,855
26	Unlu Mamuller	Atıştırmalık Ürünler	9,619	29,540
27	Tuzlu Ürünler	Hazır-Konserve ve Dondurulmuş Ürünler	8,652	29,115
28	Meyve	Kozmetik Ürünleri	6,983	28,571
29	Sebze	Meyve	56,879	27,453
30	Sebze	Atıştırmalık Ürünler	9,804	27,205
31	Bebek Ürünleri	Süt	10,202	27,058
32	Sebze	Süt	10,208	26,993
33	Unlu Mamuller	Sebze ve Meyve	13,102	26,288
34	Tuzlu Ürünler	Kozmetik Ürünleri	5,825	26,258
35	Atıştırmalık Ürünler	Kozmetik Ürünleri	6,569	26,108

Bu veriler doğrultusunda ürünler için oluşturulan bazı birliktelik kuralları aşağıdaki gibidir;

Kahvaltılık Ürünler → Süt kuralı için;

Kahvaltılık Ürünler ve Süt ürünlerinin toplam fiş hareketlerinde birlikte görülme olasılıkları %47,85'dir. Kahvaltılık Ürünler alan müşterilerin %61,89 olasılıkla Süt de aldığı söylenebilir.

Meyve → Sebze ve Unlu Mamuller kuralı için;

Meyve ve Sebze ve Unlu Mamuller ürünlerinin toplam fiş hareketlerinde birlikte görülme olasılıkları %43,76'dır. Makarna ve Soslar ürünlerini alan müşterilerin %60,86 olasılıkla Et Ürünleri de aldığı söylenebilir.

Meyve → Hazır-Konserve ve Dondurulmuş Ürünler kuralı için;

Meyve ve Hazır-Konserve ve Dondurulmuş Ürünlerin toplam fiş hareketlerinde birlikte görülme olasılıkları %6,78'dir. Meyve alan müşterilerin %42,55 olasılıkla Hazır-Konserve ve Dondurulmuş Ürünler de aldığı söylenebilir.

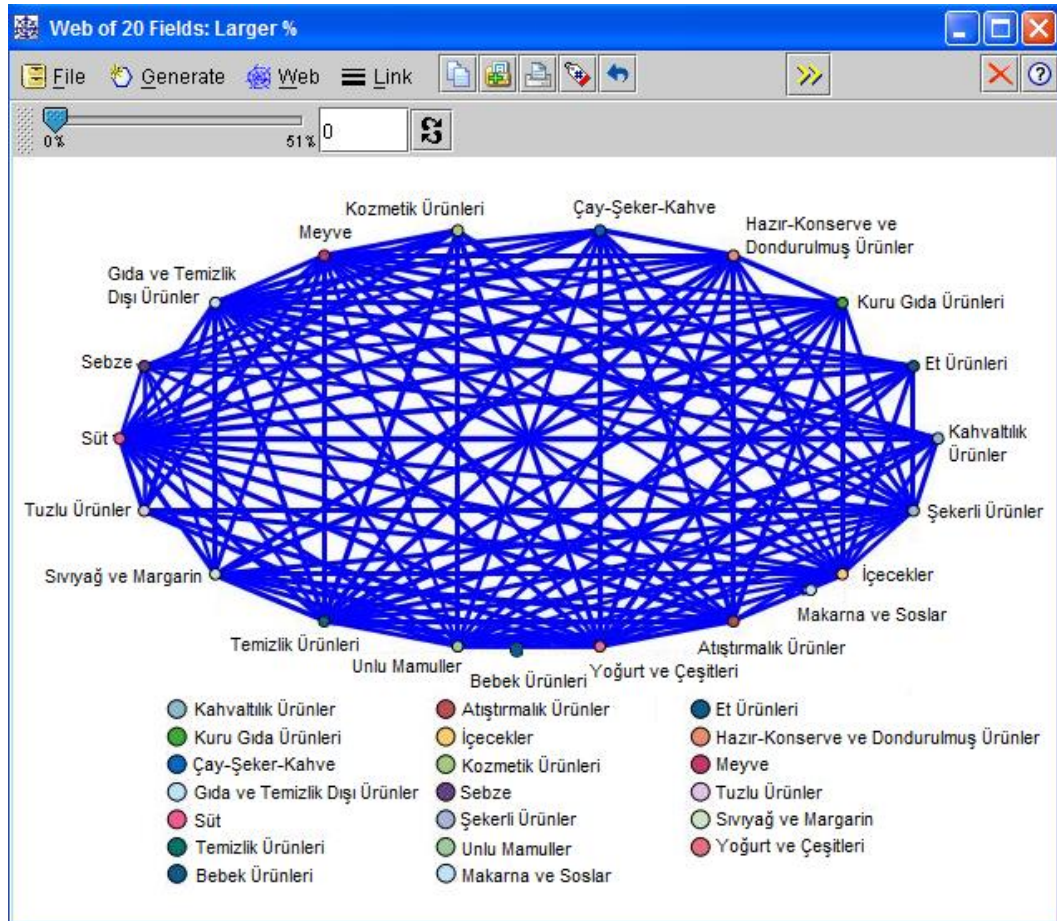
İçecekler → Atıştırmalık Ürünler kuralı için;

İçecekler ve Atıştırmalık Ürünlerinin toplam fiş hareketlerinde birlikte görülme olasılıkları %8,15'dir. İçecek alan müşterilerin %30,03 olasılıkla Atıştırmalık Ürünlerde aldığı söylenebilir.

Bebek Ürünleri→Süt kuralı için;

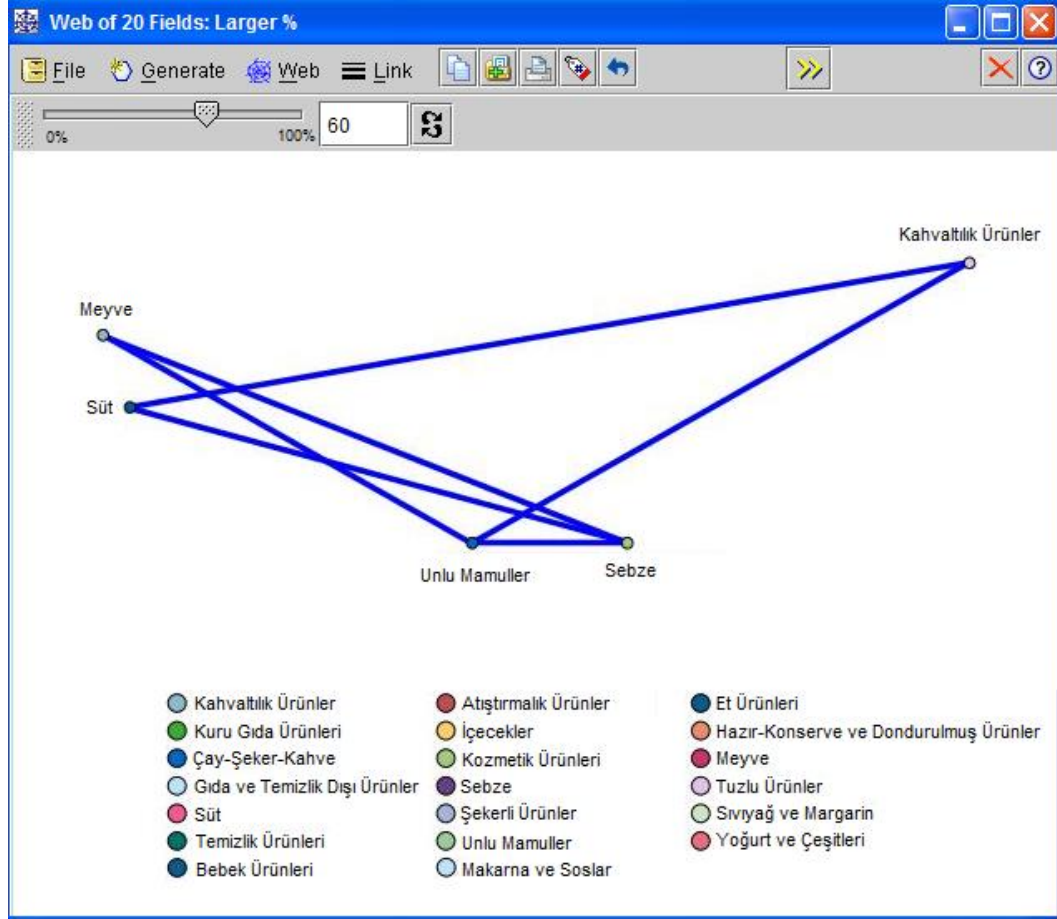
Bebek Ürünleri ve Süt ürünlerinin toplam fiş hareketlerinde birlikte görülme olasılıkları %10,20'dir. Bebek Ürünleri alan müşterilerin %27,06 olasılıkla Süt de aldığı söylenebilir.

Birliktelik kurallarının ürün kategorileri arasındaki ilişkileri görebilmek için SPSS Clementine'de web grafiğinden yararlanılmıştır. Web grafiğindeki çizgi kalınlığı ilişkinin yoğunluğunu göstermektedir. Çizgi kalınlığı arttıkça ilişkinin şiddeti de artmaktadır. Şekil 25'de çizgi yoğunluk değeri %0 olarak ayarlanmıştır. Bu doğrultuda tüm ürün kategorileri arasındaki ilişkinin şiddeti görülebilmektedir.



Şekil 25: Web Grafiği İle Tüm Ürün Kategorileri Arasındaki Birliktelik Diyagramı (çizgi değeri>=%0)

Ürün kategorilerinin birbirleri ile olan ilişkileri Şekil 25'deki birliktelik diyagramında çok yoğun olarak görülmektedir. Çizgi yoğunluk değeri değiştirildikçe web grafiğinde gösterilen sıklıklarda değişmektedir. Birliktelik sıklığı daha yüksek olan ilişkilerin görüntülenmesini sağlamak amacıyla web grafiğinde çizgi yoğunluk değeri %60 olarak ayarlanmış ve Şekil 26'da gösterilmiştir.



Şekil 26: Birliktelik Sıklığı Yüksek Olan İlişkiler(çizgi değeri \geq %60)

Şekil 26'daki web grafiğinde Unlu Mamuller ile Kahvaltılık Ürünler ve Meyve ürün kategorileri birlikte sık alındığı görülmektedir. Bununla birlikte Sebze ile Unlu Mamuller, Süt ve Meyve ürün kategorilerinin birlikte alım sıklığı da yüksektir.

2.1.5. Modelin Geçerliliği ve Değerlendirme

Aprori Algoritması kullanılarak oluşturulan birliktelik kuralları sonucunda ürün kategorileri arasındaki ilişki diyagramı (Şekil:25) ve en baskın birlikteliklere

sahip ürünler tablosuna (Tablo:13) bakıldığında Kahvaltılık Ürünler, Süt, Meyve, Sebze, Unlu Mamuller, Et Ürünleri, Kuru Gıda ürün kategorileri arasındaki ilişkilerin en güçlü birliktelikler olduğu görülmektedir. Elde edilen birliktelik kuralları doğrultusunda müşterinin satın alma davranışını etkileyecek ve satın alımları arttırmayı hedefleyen market yöneticisine yarar sağlayacak yeni raf yerleşim düzeni önerileri devam eden maddelerde anlatıldığı gibidir.

- Alışveriş fişlerinde en sık satın alınan ürün kategorileri arasında meyve ve sebze ürünleri yer almaktadır. Bu sebeple mağaza içi yerleşimine ilk olarak meyve sebze reyonu ile başlanabilir.
- Meyve ve sebze ürün kategorileri ile arasında en güçlü birlikteliğe sahip ürün kategorisi olan unlu mamuller, meyve ve sebze reyonunun yanında yer almalıdır.
- Unlu mamuller ürün kategorisinden sonra aralarında güçlü ilişki bulunan kahvaltılık ürünler ve süt ürün kategorileri yerleştirilebilir. Bu ürün kategorilerinin satış sıklığı yüksek olduğundan müşterilerin bu ürünleri market içerisinde arama ihtimali yüksektir. Bu sebeple bu ürün kategorilerinin mağazanın arka tarafına doğru yerleştirilmesiyle müşterinin mağaza içerisinde daha fazla dolaşması ve zaman harcaması sağlanabilir.
- Kahvaltılık ürünler buzdolabında teşhir edilen ürün kategorisi olduğundan bu ürünlerin devamında yine buzdolabında teşhir edilmesi gereken et ürünleri, yoğurt ve çeşitleri ürün kategorileri yerleştirilebilir.
- Kahvaltılık ürünler, et ürünleri, yoğurt ve çeşitleri ürünlerinin arasına birliktelikleri güçlü olan hazır-konserve ve dondurulmuş ürünler, kuru gıda ürünleri, makarna ve soslar, sıvıyağ ve margarin, çay-şeker-kahve ürünleri yerleştirilebilir.
- Marketin orta sıralarındaki reyonlarına birbirleri arasındaki güçlü birliktelik ilişkilerine göre içecekler, atıştırmalık ürünler, şekerli ürünler, tuzlu ürün kategorileri yerleştirilebilir. Bununla birlikte orta sıralardaki diğer reyonlara temizlik ürünleri, kozmetik ürünleri ve bebek ürün kategorileri yerleştirilebilir.

3. SONUÇ VE ÖNERİLER

Bilginin firmalar için hayati bir önem taşıdığı günümüzde, gelişen teknoloji ile birlikte veri tabanlarına çok büyük miktarlarda veriler toplanmaktadır. Toplanan verilerin çözümlenmesi, değerlendirilmesi, karar vermede kullanılması veri madenciliği kullanımının geniş bir alana yayılmasını sağlamaktadır.

Veri madenciliğindeki amaç, veri tabanlarında saklanan çok büyük miktardaki verilerden daha önce keşfedilmemiş değerli bilgilerin elde edilmesidir. Değerli bilgilerin elde edilmesi için farklı birçok veri madenciliği tekniği bulunmaktadır. Bu çalışmada günümüzde en yaygın olarak kullanılan veri madenciliği tekniklerinden biri olan birliktelik kuralları araştırmanın konusunu oluşturmaktadır.

Birliktelik kuralları, aynı işlem içerisinde çoğunlukla beraber görülen nesneleri içeren kuralları keşfetmeyi amaçlar. Başka bir ifadeyle birliktelik kuralları, büyük miktardaki veri kümeleri arasından birliktelikleri bulan kurallar bütünüdür.

Birliktelik kuralları kullanışlı ve anlaşılması kolay olduğundan pazarlama, perakendecilik, ticaret, mühendislik, bankacılık gibi endüstriyel alanlarda geniş bir kullanım alanına sahiptir. Bu nedenle birliktelik kuralları gerek ticari gerekse akademik alandaki çalışmalar hız kazanmış ve son yıllarda en çok araştırılan konulardan biri haline gelmiştir.

Çalışmada öncelikle veri tabanlarında bilgi keşif süreci, veri madenciliği, uygulama alanları ve uygulama süreci hakkında temel bilgiler verilmiştir. Çalışmanın temelini oluşturan veri madenciliği tekniklerinden birliktelik kurallarının temel kavramları, sık geçen veri kümelerinin bulunmasını sağlayan Apriori algoritması detaylı olarak anlatılmıştır.

Çalışmanın uygulama bölümünde, birliktelik kuralı bir perakende işletmesinin belirli bir dönem boyunca toplanan alışveriş fişleri üzerinde uygulanmıştır. Çalışmanın amacı birliktelik kurallarını kullanarak alışveriş hareketlerindeki veri topluluğu içerisinde birbirleri ile ilgisi olan ürünler arasındaki ilişkileri bularak, satış işleminde bu ürünlere uygulanacak kampanya ve

promosyonlarda bu ilişkilerden yararlanmak, raf düzenlemelerinin ürün ilişkisine göre yapılmasını sağlamak, dolayısıyla müşterilerin mağaza dolaşım zamanını ve satış gelirini arttırmaya çalışmaktır.

Marketten 7 gün boyunca toplanan alışveriş fişleri birliktelik kurallarına uygun olarak bilgisayar ortamına aktarılmış ve birliktelik analiz tekniği SPSS Clementine 11.1 programı Apriori Algoritması kullanılarak belirli güven ve destek değerlerine göre uygulanmıştır.

Birliktelik kuralları analizi çalışmalarında ilk olarak ürün grupları daha sonra ise ürün kategorileri için gerçekleştirilen işlemler incelenmiş ve birlikte satılma eğilimi gösteren ürünler ve ürün kategorileri arasındaki ilişkiler hakkında bilgiler verilmiştir. Çalışmalar sonucunda elde edilen bilgiler ışığında firma için daha etkin bir raf yerleşim düzenlemesi önerilmiştir. Böylece çalışmada amaçlandığı üzere önerilen raf yerleşim düzeniyle müşterilerin satın alma davranışlarının etkilenmesi sağlanabilecektir.

Yapılan bu çalışmanın ışığında bu çalışmayı daha ileriye götürecek olan araştırma önerileri devam eden maddelerde listelenmiştir.

- Zaman serilerini de içinde barındıran birliktelik kuralları analizleri gerçekleştirilerek haftanın hangi günlerinde hangi ürünlerin daha ilişkili oldukları üzerine bir çalışma yapılarak zaman bazında kampanyalar yapılabilir.
- Müşteri bilgilerinin de bulunduğu veriler ile birliktelik analizi yapılarak müşteri nezdinde kampanyalar oluşturulabilir. Bununla birlikte ürünler için elde edilen birliktelik kuralları müşteri profilini de içeren bir veri seti üzerine uygulanarak müşterilerin farklı ürünleri de satın almaları sağlanabilir.

KAYNAKÇA

- Adriaans, P. ve Zantinge, D. (1996). "Data Mining, Addison Wesley Longman Limited". Harlow, p.159.
- Agrawal, R., Imielinski, T. ve Swami, A. (1993). "Mining Association Rules Between Sets Of Items In Large Databases". ACM SIGMOD Conference on Management of Data. Washington, DC: ACM Pres, pp. 207-216.
- Agrawal, R. ve Srikant, R. (1994). Fast Algorithms For Mining Association Rules. In Proceedings of the 20th International Conference on Very Large Databases. San Francisco: Morgan Kaufmann, pp. 487-499.
- Akarşlan, E. ve Hocaoğlu, F. O. (2014). "Deformation classification of cutting discs using artificial neural networks", 2014 22nd Signal Processing and Communications Applications Conference (SIU).
- Akbulut, S. (2006). *Veri Madenciliği Teknikleri ile Bir Kozmetik Markanın Ayrılan Müşteri Analizi ve Müşteri Segmentasyonu* (Yüksek Lisans Tezi), Gazi Üniversitesi Fen Bilimleri Enstitüsü.
- Akçay, A. (2014). *Bilgi ve Belge Yönetiminde Veri Madenciliği* (Yüksek Lisans Tezi), İstanbul Üniversitesi Sosyal Bilimler Enstitüsü.
- Akçay, M. A. (2013). *Dünya Değerler Anket Verilerinin Veri Madenciliği Yöntemiyle İncelenmesi* (Yüksek Lisans Tezi), İstanbul Aydın Üniversitesi Fen Bilimleri Enstitüsü.
- Akgöbek, Ö. ve Çakır, F. (2010). "Veri Madenciliğinde Bir Uzman Sistem Tasarımı". <http://fstd.blogcu.com/veri-madenciliginde-bir-uzman-sistem-tasarimi/4784269> [11.11.2015]
- Akpınar, H. (2000). "Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği", *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, c.29, sy:1, ss.1-22.
- Aksoy, A. ve Öztürk, N. (2012). "The Fundamentals of Global Outsourcing for Manufacturers", Manufacturing System. <http://www.intechopen.com/books/manufacturing-system/thefundamentals-of-global-outsourcing-formanufacturers> [19.03.2016]
- Alagöz. A., Öge, S., Ortakarpuz M. (2014). "Bir Kurumsal Zekâ Teknolojisi Olarak Veri Madenciliği İle Muhasebe Bilgi Sistemi İlişkisi". *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, Dr. Mehmet YILDIZ Özel Sayısı, ss.1-21.
- Alpaydın, E. (2000). "Zeki Veri Madenciliği: Ham Veriden Altın Bilgiye Ulaşma Yöntemleri", *Bilişim Eğitim Semineri*, ss.1-10.

- Altaş, D. ve Gülpınar, V. (2012). “Karar Ağaçları Ve Yapay Sinir Ağlarının Sınıflandırma Performanslarının Karşılaştırılması: Avrupa Birliği Örneği”, *Trakya Üniversitesi Sosyal Bilimler Dergisi*, c.XIV, sy:1, ss.4.
- Argüden Y. ve Erşahin, B.(2008). *Veri Madenciliği*, İstanbul:ARGE Danışmanlık.
- Arlene Z. ve James K. (2013). “A Data Mining Primer for the Data Warehouse Professional”. <http://www.tdan.com/view-articles/5827> [24.11.2015]
- Ataseven, B. (2013). “Yapay Sinir Ağları İle Öngörü Modellemesi”. http://uvt.ulakbim.gov.tr/uvt/index.php?cwid=9&vtadi=TSOS&c=iparadigms&ano=156365_ed67f8f8629b99ecd5b69cb57a3d9930 [08.04.2016]
- Ay, D. (2009). *Veri Madenciliği ve Apriori Algoritması ile Süpermarket Analizi* (Yüksek Lisans Tezi), Sakarya Üniversitesi Fen Bilimleri Enstitüsü.
- Ay, D. ve Çil, İ. (2010). “Migros Türk A.Ş.’de Birliktelik Kurallarının Yerleşim Düzeni Planlamada Kullanılması”. *Endüstri Mühendisliği Dergisi*, c.21, sy.2, ss.14-29.
- Aydın, M. (2013). "Assamese Character Recognition With Artificial Neural Networks", 2013 21st Signal Processing and Communications Applications Conference (SIU).
- Aydın, Y.S., Akgöbek Ö., Öztemel E. ve Aksoy, M.S. (2006). “A New Algorithm For Automatic Knowledge Acquisition in Inductive Learning”. *Knowledge-Based Systems*, Vol. 19, pp. 388-395.
- Ayık, Y. Z., Özdemir, A. ve Yavuz, U. (2007). “Lise Türü ve Lise Mezuniyet Başarısının, Kazanılan Fakülte İle İlişkisinin Veri Madenciliği Tekniği İle Analizi”. http://uvt.ulakbim.gov.tr/uvt/index.php?cwid=9&vtadi=TSOS&c=iparadigms&ano=89547_9f1ee0db0ecc84cc909fc9b88f81685a [02.05.2016]
- Aytaç, M. B. (2013). *Doğrudan Pazarlama Aracı Olarak Tele Pazarlama İçin Veri Madenciliği Çözümleri: Banka Müşterileri Üzerine Bir Uygulama* (Yüksek Lisans Tezi), Gazi Üniversitesi Bilişim Enstitüsü.
- Baker, R. S. J. (2010). “Data Mining International Encyclopedia of Education”. *Third Edition*, pp.112-118.
- Baralis, E. (2004). "Essential Classification Rule Sets", *ACM Transactions On Database Systems*, vol.29, no.4, pp.635-674.
- Basso D. M., Aires, B. Escalada, R. (2014). “Propuesta de Metricas para Proyectos de Explotacion de Informacion”. *Revista Latinoamericana de Ingeniería de Software*, volume.2, sy.4, pp.157-218.

- Baykasoğlu, A. (2005). "Veri Madenciliği ve Çimento Sektöründe Bir Uygulama", *Gaziantep Üniversitesi Akademik Bilişim Konferansı*, ss.174.
- Becerra, F. I., Stelios H. Z. ve Steven W. (2002). "Knowledge Discovery Techniques for Predicting Country Investment Risk". *Computers and Industrial Engineering*, c:4, no:43, pp.787-800.
- Berry, M. J. A., Gordon S. L. (2000). "Mastering Data Mining: The Art and Science of Customer Relationship Management". Wiley Computer Publishing, USA.
- Becker, K. ve Ghedini, C. (2005). "A Documentation Infrastructure for the Management of Data Mining Projects". *Information and Software Technology*, c.XLVII, no:2, pp.95.
- Bilekdemir, G. (2010). *Veri Madenciliği Tekniklerini Kullanarak Üretim Süresi Tahmini ve Bir Uygulama* (Yüksek Lisans Tezi), Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü.
- Birant, D., Kut, A., Ventura, M., Altınok, H., Altınok, B., Altınok, E. ve Ihlamur, M., (2010). "İş Zekası Çözümleri için Çok Boyutlu Birliktelik Kuralları Analizi". *Akademik Bilişim '10*, Muğla, ss.256.
- Boland, J.J., Baumann, D.D. ve Dziegielewska B. (1981). "An Assessment of Municipal and Industrial Water Use Forecasting Approaches". *Contract Report 81-C05, USACE Institute For Water Resources*, Fort Belvoir, Virginia, May.
- Brin, S., Motwani, R. ve Silverstein, C. (1997). "Beyond Market Baskets: Generalizing Association Rules to Correlations". *Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data*, New York, USA, pp.265-276.
- Budak Çoban, E. (2013). "Üniversite Öğrencileri İçin Bilgisayar Okuryazarlığını Etkileyen Faktörlerin Etkisinin Veri Madenciliği İle Analizi", *Online Academic Journal of Information Technology*, c.4, sy.11, ss.57-70. http://www.ajit-e.org/?menu=pages&p=details_of_article&id=65 [12.05.2016].
- Caskey, K. (2001). "A Manufacturing Problem Solving Environment Combining Evaluation, Search and Generalization Methods", *Computers in Industry*, c.2, no:44, pp.175-187.
- Ceran, G. (2006). *Esnek Akış Tipi Çizelgeleme Problemlerinin Veri Madenciliği ve Genetik Algoritma Kullanılarak Çözülmesi* (Yüksek Lisans Tezi), Selçuk Üniversitesi Fen Bilimleri Enstitüsü.

- Cevahir, F. (2011). *Bir Perakende Firmasına Ait Veriler Üzerinde Veri Madenciliği Uygulaması* (Yüksek Lisans Tezi), Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü.
- Chambers, L. D. (1998). "Practical Handbook of Genetic Algorithms Complex Coding Systems". Volume 3, CRC, pp.31,32.
- Chapman, P. v.d., (2013). "CRISP-DM 1.0 Step-by-Step Data Mining Guide,"<http://www.crisp-dm.org/CRISPWP-0800.pdf>[16.11.2015].
- Chen, M.C. (2007). "Ranking Discovered Rules From Data Mining With Multiple Criteria By Data Envelopment Analysis", *Expert Systems With Applications*, vol.33, no.4, pp.1110-1116.
- Çalış, A. (2013). *Veri Madenciliği Yaklaşımı İle Bireysel Müşterilerin Kredi Ödeme Performanslarının Değerlendirilmesi* (Yüksek Lisans Tezi), Kocaeli Üniversitesi Fen Bilimleri Enstitüsü.
- Çalış, A., Kayapınar, S. ve Çetinyokuş, T. (2014). "Veri Madenciliğinde Karar Ağacı Algoritmaları İle Bilgisayar ve İnternet Güvenliği Üzerine Bir Uygulama". *Endüstri Mühendisliği Dergisi*, c.25, sy.3-4, ss.2-19.
- Çalışkan, B. (2006). "Veri Madenciliği ve Müşteri İlişkileri Yönetimi (CRM)". http://behlulcaliskan.com/wpcontent/uploads/2010/05/veri_madenciligi_ve_crm.pdf [22.12.2015]
- Çalışkan, S. K. ve Soğukpınar, İ. (2008). "KXKNN: K-Means ve K En Yakın Komşu Yöntemleri ile Ağlarda Nüfuz Tespiti", 2. *Ağ ve Bilgi Güvenliği Sempozyumu Girne*, ss.120-124.
- Çankırı, S., Kartal E., Yıldırım, K., ve Gülseçen, S. (2009). "Organizasyonlarda Bilgi Yönetimi Sürecinde Veri Madenciliği Yaklaşımı". ÜNAK'09 İstanbul, ss.148-167.
- Çelik, M. (2009). *Veri Madenciliğinde Kullanılan Sınıflandırma Yöntemleri ve Bir Uygulama* (Yüksek Lisans Tezi), İstanbul Üniversitesi Sosyal Bilimler Enstitüsü.
- Çetin, E. (2003). *Yapay Sinir Ağları (Kuram, Mimari, Eğitim, Uygulama)*, Ankara: SeçkinYayıncılık.
- Çetin, M. (2009). *Bir Üretim İşletmesinde Veri Madenciliği Uygulaması* (Yüksek Lisans Tezi), Sakarya Üniversitesi Fen Bilimleri Enstitüsü.
- Çınar, A. ve Silahtaroğlu, G. (2012). " Veri Madenciliği Teknikleri ile Müşteri Memnuniyetine Etki Eden Gizli Nedenlerin Keşfi". *Marmara Üniversitesi İ.İ.B. Dergisi*, c. XXXIII, sy:2, ss.309-330.

- Çingı, H., (2010). Hacettepe Üniversitesi Ders Notları 6.Bölüm, <http://yunus.hacettepe.edu.tr/~umutal/lesson/marketingschedule-bby2010.htm> [15.10.2015]
- Çil, F. (2010). *Banka Yatırım Fonu Müşteri Hareketlerinin Belirlenmesine Yönelik Bir Veri Madenciliği Uygulaması* (Yüksek Lisans Tezi), Gazi Üniversitesi Fen Bilimleri Enstitüsü.
- Çuhadar, M. (2013). “Türkiye’ye Yönelik Dış Turizm Talebinin MLP, RBF ve TDNN Yapay Sinir Ağı Mimarileri İle Modellenmesi ve Tahmini: Karşılaştırmalı Bir Analiz”. *Journal of Yasar University*, c.8, sy.31, ss.5274-5295.
- Decrop, A. (2006). *Vacation Decision Making*, Londra, CABI Publishing, p.1.
- Dolgun, M. (2006). *Büyük Alışveriş Merkezleri İçin Veri Madenciliği Uygulamaları*(Yüksek Lisans Tezi), Hacettepe Üniversitesi Fen Bilimleri Enstitüsü.
- Döşlü, A. (2008). *Veri Madenciliğinde Market Sepet Analizi ve Birliktelik Kurallarının Belirlenmesi* (Yüksek Lisans Tezi), Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü.
- Duman, S., Döşoğlu, M. K., Öztürk, A. ve Erdoğan P. (2010). “Türkiye’deki Güç Sisteminde Tavlama Benzetimi, Genetik Algoritma ve Tabu Araştırma Algoritmaları Kullanılarak Ekonomik Dağıtım”. *E-Journal of New World Sciences Academy Engineering Sciences*, volume.5, no.1, pp.64-78.
- Dursun, O. Ö. (2005). *Meteorolojik Verilerin Veri Madenciliği İle Değerlendirilmesi* (Yüksek Lisans Tezi), Fırat Üniversitesi Fen Bilimleri Enstitüsü.
- Eker, M. E., Oktaş, R. (2015). *Apriori Algoritması ve Türkiye’deki Örnek Uygulamaları*, <http://ab.org.tr/ab16/bildiri/46.pdf> [12.02.2015]
- Ekmann, J. M. (2014). “Coal Storage and Transportation”, *Reference Module in Earth Systems and Environmental Sciences*.
- Emel, G. G. ve Taşkın, Ç. (2002). “Genetik Algoritmalar ve Uygulama Alanları”. *Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, c.XXI, sy.1, ss.129-152.
- Emel, G. G. ve Taşkın, Ç. (2005). “Veri Madenciliğinde Karar Ağaçları ve Bir Satış Analizi Uygulaması”, *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, c.6 sy.2, ss.221-239.
- Emel, G.G., Taşkın, Ç. ve Tok, A. (2005). “Pazarlama Stratejilerinin Oluşturulmasında Bir Karar Destek Ağacı, Birliktelik Kuralı Madenciliği”.*Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, c.7,sy.3,ss.30-59.

- Eraldemir, S. G. ve Yildinm, E. (2014). "Classification Of Simple Text Reading and Mathematical Tasks From EEG", 22nd Signal Processing and Communications Applications Conference (SIU).
- Eriş, M.H. ve Sancak, M.F. (2011). "Bankaların İnternet Bankacılık Üzerindeki Hukuki Sorumluluğunun Değerlendirilmesinde Veri Madenciliği", *İzmir 2. Uluslararası Bilişim Hukuku Kurultayı Bildiriler Kitabı*, ss.136-149.
- Erişti, H., Yıldırım, Ö., Erişti, B. ve Demir Y. (2012). "Classification Of Power Quality Event Using Wavelet Transform and Associaton Rules", 2012 20th Signal Processing and Communications Applications Conference (SIU).
- Erpolat, S. (2012). "Otomobil Yetkili Servislerinde Birliktelik KurallarınınBelirlenmesinde Apriori ve FP-Growth AlgoritmalarınınKarşılaştırılması".*Anadolu Üniversitesi Sosyal BilimlerDergisi*, c.XII, sy:2, ss.137-146.
- Ertuğrul, E. (2008). *Ürün Kategorileri Arasındaki Satış İlişkisinin Birliktelik Kuralları ve Kümeleme Analizi İle Belirlenmesi ve Perakende Sektöründe Bir Uygulama* (Doktora Tezi), Afyon Kocatepe Üniversitesi Sostal Bilimler Enstitüsü.
- Ertuğrul, İ., Organ, A. ve Şavlı, A. (2013). "Veri Madenciliği Uygulamasına İlişkin PAÜ Hastanesinde Hasta Profiline Belirlenmesi". *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, c.19, sy.2, ss.97-103.
- Fayyad, U., Gregory, P. S. ve Padhraic, S. (1996). "Knowledge Discovery and Data Mining: Towards a Unifying Framework", *Proceedings of the SecondInternational Conference on Knowledge Discovery and Data Mining (KDD-96)*, c.XXXII, AAAI Press, pp.82-83.
- Freitas, A. A. (2002). *Data Mining and Knowledge Discovery with Evolutionary Algorithms*, Almanyaya, Springer-Verlag.
- Gemici, B. (2012). *Veri Madenciliği ve Bir Uygulama* (Yüksek Lisans Tezi), Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü.
- Genç, S. (2011). *Veri Madenciliğine Genel Bir Bakış*, <http://www.sevdanurgenc.com/archives/826> [05.02.2016]
- Giovinazzo, W. A. (2002). *Internet Enabled Business Intelligence*, U.S.A., Prentice Hall PTR, pp.331.
- Glenn, J. M. (2007). *Making Sense of Data*, US, John Wiley & Sons Publication, p. 110.
- Gomes, B.M.V. (2011). "PrevisAo de Churnem Companhias de Seguros (Master's Thesis)" Universidade do Minho Escola de Engenharia.

- Gökgöz, K. B. (2015). *Veri Madenciliği İle Tıbbi Cihaz Bakım Karar Modeli*(Yüksek Lisans Tezi), Başkent Üniversitesi Sosyal Bilimler Enstitüsü.
- Gürbüz, F., Öbakır, L. ve Yapıcı, H. (2009). "Türkiye’de Bir Havayolu İşletmesine Ait Parça Söküm Raporlarına İlişkin Veri Madenciliği Uygulaması". *Gazi Üniversitesi Mimarlık Mühendislik Fakültesi Dergisi*, c.24, sy.1, ss.73-78.
- Gürgen, G. (2008). *Birliktelik Kuralları ile Sepet Analizi ve Uygulaması* (Yüksek Lisans Tezi), Marmara Üniversitesi Sosyal Bilimler Enstitüsü.
- Halfawy, M. R. ve Hengmeechai, J. (2014). "Integrated Vision-Based System for Automated Defect Detection in Sewer Closed Circuit Television Inspection Videos". *Journal Of Computing in Civil Engineering*.
- Hatlaş, A., Alkan, A. ve Karabulut, M. (2015). "Metin Sınıflandırmada Sezgisel Arama Algoritmalarının Performans Analizi". *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, c.30, sy.417-427.
- Han, J. ve Kamber, M. (2001). "Data Mining : Concepts and Techniques".*Academic Pres*.
- Han, J. ve Kamber, M. (2006). "Data Mining Concepts and Techniques". 2.bs., USA, The Morgan Kaufman Publishers.
- Han, J., Pei, J., ve Yin, Y.(2000). "Mining Frequent Patterns Without Candidate Generation". *In Proceedings of the ACM SIGMOD International Conference on Management of Data SIGMOD*, New York: ACM Press. pp. 1–12.
- Ho, G.T.S., Lau, H.C.W., Chan, T.M., ve Tang, C. (2010). "An Online Analytical Processing Based Predictive System For Better Process Quality in the Supply Chain Network", *International Journal of Services Technology and Management*.
- Hong, G., Park, S., Jang, D. ve Rho, H. (2005). "An Effective Supplier Selection Method for Constructing a Competitive Supply Relationship".*Expert Systems with Applications*, c.4, no:28, pp.629-639.
- Huang, W. (2008). "Analysis Of The User Behavior and Opinion Classification Based On The BBS", *Applied Mathematics and Computation*.
- Huang, Z., v.d. (2004). "Credit Rating Analysis with Support Vector Machines and Neural Networks: A Market Comparative Study".*Decision Support Systems*, no: 37, pp. 543-558.
- Kantardic, M. (2002). "General Data Analysis and Data Mining". Wiley-IEEE Pres,p.1.

- Kantardzic, M. (2003). "Chapter 9:Artificial Nevral Networks Chapter 1-1.14", Data Mining: Concepts, Models, Methods, and Algorithms. 2nd Edition.
- Karagöz, N. E. (2007). *Market Veri Tabanında Veri Madenciliği Uygulaması* (Yüksek Lisans Tezi), İstanbul Ticaret Üniversitesi Sosyal Bilimler Enstitüsü.
- Kartal, E. (2015). *Sınıflandırmaya Dayalı Makine Öğrenmesi Teknikleri ve Kardiyolojik Risk Değerlendirmesine İlişkin Bir Uygulama* (Doktora Tezi), İstanbul Üniversitesi Fen Bilimleri Enstitüsü.
- Kavurkacı, Ş. ve Aydın, Z.G. (2011). "Büyük Ölçekli Veri Tabanlarında Bilgi Keşfi". <http://ab.org.tr/ab11/bildiri/76.pdf> [22.02.2016]
- Kim, Y. ve Street, W. N. (2004). "An Intelligen System for Customer Targeting: A Data Mining Approach". *Decision Support Systems*, No:37, p. 215-228.
- Kiremetçi, B. (2005). *Veri Ambarlarında Veri Madenciliği ve Ulaştırma- Lojistik Sektöründe Bir Uygulama* (Yüksek Lisans Tezi), İstanbul Üniversitesi Sosyal Bilimler Enstitüsü
- Kumdereli, Ü. C. (2012). *Tıp Bilişimi Ve Veri Madenciliği Uygulamaları: EEG Sinyallerindeki Epileptiform Aktiviteye Veri Madenciliği Yöntemlerinin Uygulanması* (Yüksek Lisans Tezi), Trakya Üniversitesi Fen Bilimleri Enstitüsü.
- Koltan Yılmaz, Ş. ve Patır, S. (2011). "Kümeleme Analizi ve Pazarlamada Kullanımı", *Akademik Yaklaşımlar Dergisi*, c.2, sy.1, ss.91-113.
- Koyuncugil, A. S. ve Özgülbaş, N. (2009). "Veri Madenciliği:Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları". *Bilişim Teknolojileri Dergisi*, c.2, sy.2, ss.21-32.
- Kök, B.V. ve Kuloğlu, N. (2005). Sollama Esnasında Taşıt ve Yol ile İlgili Faktörlerin Karar Ağacı Yöntemi ile İrdelenmesi, *Erciyes Üniversitesi Fen Bilimleri Enstitüsü Dergisi*.
- Kurgan, L.A. ve Musilek, P. (2006). "A Survey of Knowledge Discovery and Data Mining Process Models", *The Knowledge Engineering Review*, vol.21, no.1, pp.1-24.
- Kurt M. ve Semetay, C. (2011). *TMMOB Makine Mühendisleri Odası Mühendis ve Makine Dergisi*, Ankara, no: 501, ss.2.
- Küçükşille, E. (2009). *Veri Madenciliği Süreci Kullanılarak Portföy Performansının Değerlendirilmesi ve İMKB Hisse Senetleri Piyasasında Bir Uygulama* (Doktora Tezi), Süleyman Demirel Üniversitesi Sosyal Bilimler Enstitüsü.

- Leventhal, B. (2015). "All Models Are Wrong But Some Are Useful: The Use Of Predictive Analytics in Direct Marketing", *Quality Technology & Quantitative Management*.
- Lezki, Ş. ve Er, F. (2010). "Tatil Yeri Kararının Verilmesinde Karar Ağacı ve Etki Diyagramının Uygulanması", *Anatolia: Turizm Araştırmaları Dergisi*, c.21, sy.2, ss.233-242.
- Linoff, G. ve Berry, M.J.A. (2004). "Data Mining techniques For Marketing Sales and Customer Relationship Management". New York: Wiley Publishing.
- Liu, B., Hsu, W., Chen, S. ve Ma, Y. (2000). "Analyzing The Subjective Interestingness of Association Rules", *IEEE Intelligent Systems*, No: 15, pp 47-55.
- Liao, C.W. (2008). "Data Mining For Occupational Injuries in the Taiwan Construction Industry", *Safety Science*, vol.46, no.7, pp.1091-1102.
- Lu, C. V. ve Chen, C. (2008). "A Study Of Applying Data Mining Approach To the Information Disclosure For Taiwan". *Stock Market Investors, Expert Systems With Applications*, c.36, no.2, pp.3536.
- Maimon, O. Z. ve Rokach, L. (2005). "Data Mining and Knowledge Discovery Handbook". *Sprenger science Business Media*, USA.
- Maindonald, J. (2007). "Data Mining From A Atatistical Perspective, Preprint". Australian Nat. Univ. Stat Cons Unit, (<http://maths.people.anu.edu.au/~johnm/dm/dmpaper.html> [12.10.2015])
- Mainly, F.J.B. (1994). "Multivariate Statistical Methods". 2. bs, Londra: Chapman Hall, p.280.
- Malik, M. ve Sharma, A. K. (2011). "3DCCOM Polygon Reduction Algorithm in Presence of Obstacles, Facilators and Constrains". *International Journal of Computer Applications*, volume 30, no.7, pp.6-12.
- Men-Hsieu, H. (2004). "Knowledge Discovery With SOM Networks in Financial Investment Strategy", Fourth International Conference on Hybrid Intelligent Systems (HIS 04).
- Michael, J. A. B. ve Gordon, L. (1998). "Data Mining Solutions". *Wiley Computer Publishing, U.S.A.*, p.25.
- Michael, J. A. B. ve Gordon, S.L. (2000). "Mastering Data Mining: The Art and Science of Customer Relationship Management". *Wiley Computer Publishing, USA*.

- Moshkovich, M.H., Alexander, I. M. ve Olson, D. L. (2002). "Rule Induction in Data Mining: Effect of Ordinal Scales". *Expert Systems with Applications*, c.XXII, no:4, pp.303.
- Moss, T. L. (2003). *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision- Support Applications*, Almanya, Addison Wesley.
- Narlı, S., Aksoy, E., Ercire, Y. E. (2014). *Veri Madenciliği ile İlköğretim Matematik Öğretmen Adaylarının Öğrenme Stillерinin ve Aralarındaki İlişkilerin İncelenmesi*. *International Journal of Educational Studies in Mathematics*, c.1, sy.1, ss.37-57.
http://www.ijesim.com/upload/dosya/narli_investigation-of-prospective-elementary-mathematics-teachers-learning-styles-and-relationships-between-them-using-data-mining_153a82060e2e6f.pdf [14.03.2016]
- Oğuzlar, A. (2003). "Veri Ön İşleme". *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, sy: 21, ss. 67-76.
- Ohsuga, L. ve Liao H. (2005). *Foundations and Novel Approaches in Data Mining*, Warsaw, Springer, pp. 121.
- Olson, D. L. ve Delen, D. (2008). "Advanced Data Mining Techniques". Springer, Berlin Heidelberg.
- Orhunbilge, N. (1996). *Uygulamalı Regresyon ve Korelasyon Analizi*, İstanbul: İstanbul Üniversitesi İşletme Fakültesi Yayınları.
- Özçakır F. C. ve Çamurcu, A.Y. (2007). "Birliktelik Kuralı Yöntemi İçin Veri Madenciliği Yazılımı Tasarımı ve Uygulaması", *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, sy:12, ss:21-37.
- Özekes, S. (2003). "Veri Madenciliği Modelleri Ve Uygulama Alanları". *İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi*, c.III, sy:3, ss.67.
- Özekes, S. ve Çamurcu, A.Y. (2002). *Veri Madenciliğinde Sınıflama ve Kestirim Uygulaması*, *Marmara Üniversitesi Fen Bilimleri Dergisi*, sy:18, ss. 1-17.
- Özkan, Y. (2008). *Veri Madenciliği Yöntemleri*. İstanbul: Papatya Yayıncılık.
- Özmen, Ş. (2001). "İş Hayatı Veri Madenciliği ile İstatistik Uygulamalarını Keşfediyor". *Çukurova Üniversitesi V. Ulusal Ekonometri ve İstatistik Sempozyumu*, ss.2-15. <http://mimoza.marmara.edu.tr/sozmen/teblig-sunumlar/3is-hayati-verimadenciliği-istatistik.pdf> [11.10.2015]
- Öztemel, E. (2003). *Yapay Sinir Ağları*, İstanbul: Papatya Yayıncılık.
- Paolo, G. (2003). *Applied Data Mining Statistical Methods for Business and Industry*, İngiltere, Wiley.

- Park, S. C., Selwyn P. ve Michael J. S. (2001). "Dynamic Rule Refinement in Knowledge-based Data Mining Systems". *Decision Support Systems*, no:31, ss. 205-222.
- Pei, J., Han, J., Lu, H., Nishio, S., Tang, S., ve Yang, D. (2001). "H-MINE: Hyper-Structure Mining Of Frequent Patterns in Large Databases". *In Proceedings of the First IEEE International Conference on Data Mining*. New York: IEEE Pres, pp. 441-448.
- Pei, J., Han, J., ve Mao, R. (2000). "CLOSET: An Efficient Algorithm for Mining Frequent Closed Itemsets". *In Proceedings of the ACM-SIGMOD International Workshop on Data Mining and Knowledge Discovery*. New York: ACM Press. ss. 21-30.
- Pollo-Cattaneo, F., Amatriain, H., Rodriguez, D., Pytel, P., Ciccolella, E., Vegega, C., Dearriba, M., Rodriguez Aubert, M., Bose, F., Giordano, L., Britos, P. ve Garcia-Martinez, R. (2010). "Ingenieria De Proyectos De Explotacion De Informacion". *Proceedings XII Workshop de Investigadores en Ciencias de la Computación Área Bases de Datos y Minería de Datos*. Pag. Pp.172-176.
- Roiger, R. J. ve Michael, W. G. (2003). "Data Mining A Tutorial-Based Primer". USA: Addison Wesley.
- Sarıgül, H. (2014). "Kümeleme Analizi İle İllerin Bankacılık Hizmetlerine Erişim ve Kullanım Göstergelerinin Karşılaştırılması". *Bankacılar Dergisi*, sy.89, ss.41-62.
- Saruhan, H. (2004). "Genetic Algorithms: An Optimization Technique", *Teknoloji Dergisi*, c.VII, no:1, ss.105-114.
- Sever, H. ve Buket, O. (2002). *Veritabanlarında Bilgi Keşfine Formal Bir Yaklaşım Kısımlı: Eşleştirme Sorguları ve Algoritmalar*, <http://web.cs.hacettepe.edu.tr/~sever/bilgi-p02-03.pdf> [22.01.2015]
- Sezen, H.K. (2004). *Yöneylem Araştırması*, Bursa: Ekin Kitapevi.
- Shapiro, P.G. (1991). "Discovery, Analysis, and Presentation of Strong Rules in Knowledge Discovery in Databases". Menlo Park, CA: AAAI/MIT Press.
- Shearer, C. (2000). "THE CRISP-DM Model: The New Blueprint For Data Mining". *Journal of Data Warehousing*, c.4, no:5, pp.13-23.
- Siraj, F. ve Mansour A. (2011). "Mining Enrollment Data Using Descriptive and Predictive Approaches", *Knowledge-Oriented Applications in Data Mining*.
- Sivanandam, S. N. ve Sumathi, S. (2006). "Introduction to Data Mining and its Applications", New York, Springer.

- Şaylan, A. Ç. (2013). *Böbrek Nakli Geçirmiş Hastalarda Akıllı Yöntem Tabanlı Yeni Öznitelik Seçme Algoritması Geliştirilmesi* (Yüksek Lisans Tezi), Kadir Has Üniversitesi Fen Bilimleri Enstitüsü.
- Tang, Z. ve Maclennan, J. (2005). "Data Mining With Sql Server 2005". Wiley.
- Tantuğ, A.C. (2002). *Veri Madenciliği ve Demetleme* (Yüksek Lisans Tezi), İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü.
- Taş, M., Adak, M. F. ve Yurtay, N. (2013). "Öğrencilerin Staj Verileri Üzerine Uygulanan Apriori Algoritması ile Birliktelik Kurallarının Çıkarılması ve Staj Eğiliminin Belirlenmesi", *Akademik Platform*, ss.1086-1095. <http://docplayer.biz.tr/1623038-Ogrencilerin-staj-verileri-uzerine-uygulanan-apriori-algoritmasi-ile-birliktelik-kurallarinin-cikarilmasi-ve-staj-egiliminin-belirlenmesi.html> [13.03.2016]
- Tatei, D. M. ve Smith, A.E. (1994). "A Genetic Approach to the Quadratic Assignment Problem". *Computers & Operations Research*, c.XXII, no:1, pp.73-83.
- Timor, M. ve Şimşek, U. T. (2008). "Veri Madenciliğinde Sepet Analizi İle Tüketici Davranışı Modellemesi", c.19, sy.59, ss.3-10. <http://www.arastirmax.com/system/files/dergiler/323/makaleler/19/59/arastirmax-veri-madenciliginde-sepet-analizi-ile-tuketici-davranisi-modellemesi.pdf> [24.05.2016]
- Toktaş, P. ve Demirhan, M.B. (2004). "Risk Analizinde Veri Madenciliği Uygulamaları," *Yöneylem Araştırması/Endüstri Mühendisliği–XXIV Ulusal Kongre Bildirisi*.
- Topaloğlu, S. (2013). "Sosyal Güvenlik Suiistimallerinin Tespiti ve Önlenmesi İçin Risk Analizi ve Sürekli Denetim Yöntemleri", Türkiye İşveren Sendikaları Konfederasyonu.
- Tosunoğlu, N. G. ve Benli, Y. K. (2012). "Morgan Stanley Capital International Türkiye Endeksinin Yapay Sınır Ağları ile Öngörüsü". *Ege Akademik Bakış*, c.12, sy.4, ss.541-547.
- Tuffery, S. (2011). "Data Mining and Statistics For Decision Making". 1st ed., Wiley, USA.
- Tüzüntürk, S. (2010) . "Veri Madenciliği ve İstatistik". *Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, c.XXIX, sy.1, ss.65-90.
- Valsamidis, S., Kazanidis, L., Kontogiannis, S. ve Karakos, A. (2012). "An Approach For LMS Assessment", *International Journal of Technology Enhanced Learning*, vol.4, no:3/4.

- Yalçıntaş, G. (2003). *Veri Madenciliği* (Yüksek Lisans Tezi), Gazi Üniversitesi Fen Bilimleri Enstitüsü.
- Yavuz, S. (2009). “Hataları Ardışık Bağımlı (Otokorelasyonlu) Olan Regresyon Modellerinin Tahmin Edilmesi”. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, c.23, sy.3, ss.123-140.
- Yazıcı, A. (2007), “Yapay Sinir Ağlarına Genel Bakış”, *Tıp Bilimleri Dergisi*, http://209.85.229.132/search?q=cache:nwaMpL4GkEJ:tipbilimleri.turkiyelinkleri.com/download_pdf [13.03.2015]
- Ye, N. (2003). *The Handbook Of Data Mining*, London:Lawrence Erlbaum Associates.
- Yen, S. J. ve Lee, Y. S. (2006). “An Efficient Data Mining Approach for Discovering Interesting Knowledge from Customer Transactions”. *Expert Systems with Applications*, c.XXX, no:4, pp.650-657.
- Yıldızeli, A., Arıkan, A. ve Çakmak, T. (2011). *Bilgi Çağında Varoluş: “Fırsatlar ve Tehditler” Sempozyumu Bildiriler Kitabı*, İstanbul:Yeditepe Üniversitesi.
- Yoones A. S. (2010). "Mining Important Association Rules Based On The RFMD Technique", *International Journal of Data Analysis Techniques and Strategies*, vol.2, no.1, pp.1-21.
- Yurtay, Y., Yurtay, N., Çelebi, N., Bacınoğlu, N. Z. ve Ak, G. (2013). “Sakarya İline Ait Yangın Verilerinin Veri Madenciliği Yöntemleriyle Değerlendirilmesi”, *Akademik Platform*, ss.930-939. <http://www.isites.info/pastconferences/isites2014/isites2014/papers/B6-ISITES2014ID149.pdf> [14.10.2015]
- Yurtsever, U. (2002), *Veri Madenciliği Ve Uygulaması* (Yüksek Lisans Tezi), Sakarya Üniversitesi Sosyal Bilimler Enstitüsü.
- Wasan, P.S., Uttamchandani, M., Moochhala, S., Yap, V.B., ve Yap, P.H. (2013). “Application Of Statistics and Machine Learning For Risk Stratification Of Heritable Cardiac Arrhythmias”, *Expert Systems with Applications*, volume.40, pp.2476-2486.
- Witten, I. ve Frank, E. (2000). “Data Mining Practical Machine Learning Tools and Techniques With Java Implementations”, *San Fransisco:Morgan Kaufmann Publishers*, pp.395.
- Woody, W.D., Danie, D. B. ve Baker, C.A. (2010). “E-books or textbooks: Students prefer textbooks, *Computers & Education*” no: 55, pp.945–948.

Wu, Y. (1998). "Synthesis and Characterization Of AProtected Amino Alcohol Containing Ortho-Carborane", Polyhedron, Volume.17, Issue.19, pp.3247-3476.

ÖZET

VERİ MADENCİLİĞİ VE MARKET VERİTABANINDA BİRLİKTELİK KURALLARININ BELİRLENMESİ

Günümüzde bilgi sistemleri ve teknolojinin gelişmesi sonucunda çok büyük miktardaki veriler firmalar tarafından toplanıp saklanabilmektedir. Uygun bilgisayar programlarının gelişimi ve firmaların topladığı verileri kullanılabilir bilgiye çevirme isteği toplanan bu veriyi analiz ederek içerisindeki anlamlı, gelecekle ilgili tahmin yapılmasını sağlayan bağıntı ve kuralların ortaya çıkarılmasını gerekli hale getirmiştir. Veri madenciliğinin veri tabanlarında saklı olan bilgilere otomatik bilgi keşfetme teknikleriyle ulaşması bu gerekliliği karşılayacak bir disiplin olarak ortaya çıkmasını sağlamıştır.

Bu çalışmada veri tabanlarında bilgi keşif süreci ve veri madenciliği hakkında temel bilgiler verilerek, diğer disiplinlerle ilişkileri, faydaları, uygulama alanları, uygulama süreci, veri madenciliği modelleri ve teknikleri, veri madenciliğinde karşılaşılan problemler ve kullanılan yazılımlar konuları üzerinde detaylı olarak durulmuştur.

Çalışmanın uygulama bölümünde SPSS Clementine 11.1 programı aracılığıyla, Apriori Algoritması kullanılarak perakende sektöründe hizmet veren bir firmaya ait veriler üzerinde veri madenciliği birliktelik analizi yapılmıştır. Yapılan bu çalışma sonucunda birlikte satılma eğilimi gösteren ürünler ve ürün kategorileri hakkında bilgiler verilerek, yeni bir market raf yerleşim düzeni önerilmiştir.

Anahtar Kelimeler: Veri Madenciliği, Veri Madenciliği Modelleri, Veri Madenciliği Teknikleri, Birliktelik Kuralları, Apriori Algoritması.

ABSTRACT

DATA MINING AND DETERMINATION OF THE RULES OF TOGETHERNESS IN GROCERY DATABASE

Nowadays, huge amount of data is able to be collected and stored by companies with the help of advancing technology and information system. Moreover, determination of certain rules and relations enabling analysis of data to have required knowledge and idea about future became necessity with the enhancement of suitable software, and desire of companies converting collected data into knowledge. Thus, data mining came up with automatic information detection techniques to access information hidden in database as a discipline to meet the requirements.

In this thesis, by presenting the process of discovery of knowledge along databases, and fundamental information regarding to data mining; benefits, application areas, application process, and models as well as techniques of data mining along with encountered problems were committed.

In the part of thesis work, the data belonging to a private retail store was compiled and realized the togetherness of data mining analysis by applying Apriori Algorithm via SPSS Clementine 11.1 software. As a result of this evaluation, having more information about products which tend to be on sale together with another products and their categories, new positions of products on the shelf were suggested and reorganized.

Key Words: Data Mining, The Models of Data Mining, The Techniques of Data Mining, Apriori Algorithm

ÖZGEÇMİŞ

Adı Soyadı : Burcu ALAN

Doğum Yeri: Büyükçekmece

Doğum Tarihi: 22.01.1989

Eğitim Durumu

Lise

-Fatih Ticaret Meslek Lisesi (2005)

Ön Lisans

- Anadolu Üniversitesi, Açıköğretim Fakültesi, Kamu Yönetimi (2012)

Lisans

- Gazi Üniversitesi, Ticaret ve Turizm Eğitim Fakültesi, Pazarlama Öğretmenliği (2012)

- Anadolu Üniversitesi, İşletme Fakültesi, İşletme (2015)

Yüksek Lisans

- Recep Tayyip Erdoğan Üniversitesi, Sosyal Bilimler Enstitüsü, İşletme (2016)

Çalıştığı Kurumlar

2005-2008: Pirgos Tekstil Ltd. Şti.(Muhasebe)

2013-2015:Sakarya Akyazı Ticaret Meslek Lisesi (Pazarlama ve Perakende Öğretmeni)

2015-Devam Ediyor: İHKİB Kağıthane Mesleki ve Teknik Anadolu Lisesi (Pazarlama ve Perakende Öğretmeni)