

**T.C.  
MİMAR SİNAN GÜZEL SANATLAR ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**BİRLİKTELİK ANALİZİ VE BİR UYGULAMASI**

**YÜKSEK LİSANS TEZİ**

**Onur BAYRAM**

**İstatistik Anabilim Dalı**

**Tez Danışmanı: Prof. Dr. Gülay BAŞARIR**

**HAZİRAN 2014**

Onur BAYRAM tarafından hazırlanan "Birliklilik Analizi ve Bir Uygulaması" adlı bu tezin Yüksek Lisans Tezi olarak uygun olduğunu onaylarım.

  
Prof. Dr. Gülay BAŞARIR  
Tez Yöneticisi

Bu çalışma, Jürimiz tarafından İstatistik Anabilim Dalında Yüksek Lisans Tezi Olarak kabul edilmiştir.

Başkan : Prof. Dr. Gülay BAŞARIR

Üye : Prof. Dr. Aydın ERAR

Üye : Doç. Dr. Atıf Ahmet EVREN


Bu tez, Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygundur.

# BİRLİKTELİK ANALİZİ VE BİR UYGULAMASI

(Yüksek Lisans Tezi)

Onur BAYRAM

MİMAR SİNAN GÜZEL SANATLAR ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

Haziran 2014

## ÖZET

Günümüzde çoğu şirket ve kurumlar, veri madenciliği yöntemlerini kullanarak, her alanda sürekli olarak artan veri yığınları içerisindeki belirli ilişkileri bulmak ve bu ilişkileri geçerli ve etkin bir şekilde yorumlamak istemektedir. Bu yüzden, veri madenciliğinin en çok kullanılan tekniklerinden biri olan, özel bir uygulama alanı olan Birliktelik Analizi önemli bir yere sahiptir. Birliktelik analizinde ortaya çıkarılan birliktelik kuralları nesnelere arasındaki ilişkilerin keşfedilmesine sağlamaktadır. Keşfedilen bu ilişkiler ile geleceğe yönelik tahminler yapılması kolaylaşır.

Tez çalışmasının amacı, veri madenciliğinin önemli tekniklerinden olan birliktelik analizinin daha kolay anlaşılabilmesini ve uygulanabilmesini sağlamaktır, aynı zamanda yapılan uygulama ile veri setlerindeki ürün gruplarının satışlarının artırılmasını ve ürün gruplarının hedef kitesinin kontrol edilebilmesini kolaylaştırmaktır. Bu amaç doğrultusunda, bu tez çalışmasında öncelikle veri madenciliğinin genel disiplini açıklanmış, veri madenciliğinde kullanılan teknikler ve modeller anlatılmıştır. Veri madenciliğinin önemli tekniklerinden biri olan Birliktelik Analizi, birliktelik kuralları ve algoritmaları ile beraber ayrıntılı olarak incelenmiştir. Bu çalışmanın uygulama bölümünde ise dünya çapında pazar araştırmaları yapan bir şirketin, Türkiye'deki panel bölümünün, geleneksel ve modern kanalındaki marketler kullanılarak elde edilmiş gerçek bir veri seti ile bu marketlerde satın alınan ürün grupları arasındaki ilişkiler, birliktelik kuralları ile sepet analizi, SPSS Clementine 12.0 paket programı ve apriori algoritması yardımıyla belirlenmiş ve yorumlanmıştır.

**Anahtar Kelimeler:** Veri Madenciliği, Birliktelik Analizi, Birliktelik Kuralları, Apriori Algoritması, Market Sepet Analizi, SPSS Clementine 12.0, İstatistik.

**Danışman:** Prof. Dr. Gülay BAŞARIR, Mimar Sinan Güzel Sanatlar Üniversitesi, İstatistik Bölümü

# ASSOCIATION ANALYSIS AND AN APPLICATION

(M.Sc. Thesis)

Onur BAYRAM

MIMAR SINAN FINE ARTS UNIVERSITY  
INSTITUTE OF SCIENCE AND TECHNOLOGY

June 2014

## ABSTRACT

Nowadays, most companies and institutions request, using data mining methods, within a constantly growing piles of data in all areas to find specific relationships and these relationships are valid and effective way to interpret. Therefore, Association Analysis , which is one of the most widely used techniques in data mining , the particular application has an important place in data mining. In association analysis revealed association rules allow fort he discovery relationships between objects. These discovered relationships are easier to forecast for the future.

Aim of the thesis , the association analysis, the important technique of data mining, more easily understandable and implementation is to ensure that, at the same time, with the application, the sales of product groups in data set to increase and the target audience of product groups in data set to be controlled to facilitate. For this purpose, this thesis primarily explained the general discipline of data mining, and described used the techniques and models in data mining. Association Analysis which is one of the important techniques of data mining, with association rules and algorithms are studied in detail together. Application of this study, a company of the worldwide doing market research business in panel section, Turkey, mrkets in the traditional and modern channels, have been achieved using a real data set with it in stores purchased product group relationships, market basket analysis by association rules, with SPSS Clementine 12.0 package program and apriori algorithm determined and interpreted.

**Keywords:** Data Mining, Association Analysis, Association Rules, Apriori Algorithm, Market Basket Analysis, SPSS Clementine 12.0, Statistics

**Advisor:** Prof. Gülay BASARIR, Mimar Sinan Fine Arts University, Department of Statistics

## **TEŐEKKÜR**

Bu alıőmanın meydana gelmesinde yardımcı olan, alıőma süresince görüş ve önerilerini benimle paylaşan danışmanım Sayın Prof. Dr. Gülay BAŐARIR başta olmak üzere, alıőma için gerçek bir veri seti elde etmem de katkıları bulunan Sayın Serkan CERAN'a, özellikle uygulama bölümünde elinden gelen desteęi veren Sayın Yrd. Do. Dr. Elif Özge ÖZDAMAR'a, Mimar Sinan Güzel Sanatlar Üniversitesi ve Eskiőehir Osmangazi Üniversitesi İstatistik Bölümü tüm öğretim üyelerine ve eğitim hayatım boyunca maddi ve manevi olarak beni destekleyen aileme ve arkadaşlarıma teşekkür ederim.

Onur BAYRAM

## İÇİNDEKİLER

<b>ONAY SAYFASI</b>	<b>ii</b>
<b>ÖZET</b>	<b>iii</b>
<b>ABSTRACT</b>	<b>iv</b>
<b>TEŞEKKÜR</b>	<b>v</b>
<b>İÇİNDEKİLER</b>	<b>vi</b>
<b>TABLO LİSTESİ</b>	<b>viii</b>
<b>ŞEKİL LİSTESİ</b>	<b>ix</b>
<b>EKRAN LİSTESİ</b>	<b>x</b>
<b>KISALTMALAR LİSTESİ</b>	<b>xi</b>
<b>1. GİRİŞ</b>	<b>1</b>
1.1. Giriş ve Çalışmanın Amacı	1
<b>2. VERİ MADENCİLİĞİ</b>	<b>3</b>
2.1. Genel Bilgiler	3
2.2. Veri Madenciliğinin Farklı Tanımları	6
2.3. Veri Kaynakları	7
2.4. Veri Ambarları	9
2.5. İlişkisel Veritabanları	10
2.6. İşlemsel Veritabanları	11
2.7. Gelişmiş Veritabanı Sistemleri	11
2.8. Veri Ambarları ile Veritabanı Arasındaki Farklar	12
2.9. Veri Madenciliğinin Uygulama Alanlar	12
2.10. Veri Madenciliğinin Modelleri ve Algoritmaları	14
2.10.1. Sınıflama ve Regresyon Modelleri	16
2.10.2. Kümeleme Modelleri	16
2.10.3. Birliktelik Kuralları ve Ardışık Zamanlı Örüntüler	17
<b>3. BİRLİKTELİK ANALİZİ</b>	<b>19</b>
3.1. Genel Bilgiler	19
3.2. Market Sepet Analizi	20
3.3. Birliktelik Kuralları	21
3.3.1. Birliktelik Kurallarının Temel Kavramları	22
3.3.2. Birliktelik Kurallarının Çeşitleri	24
3.3.2.1. Genelleştirilmiş Birliktelik Kuralları	24
3.3.2.2. Uzaysal Birliktelik Kuralları	25
3.3.2.3. Nicel Birliktelik Kuralları	26
3.3.2.4. Aralık Veri Birliktelik Kuralları	26
3.3.2.5. Çoklu min_destek Birliktelik Kuralları	27
3.3.2.6. Multimedya Birliktelik Kuralları	27

3.3.2.7. Maksimal Birliktelik Kuralları	27
3.3.3. Birliktelik Kurallarının Matematiksel Modeli	27
3.3.3.1. Güven (confidence) ve Destek (support)	29
3.4. Birliktelik Analizinde Kullanılan Algoritmalar	31
3.4.1. AIS Algoritması	31
3.4.2. SETM Algoritması	32
3.4.3. Apriori Algoritması	32
3.4.4. AprioriTID Algoritması	40
3.4.5. FP-Growth Algoritması	41
3.4.6. Diğer Algoritmalar	42
3.4.7. Algoritmaların Karşılaştırılması	45
<b>3. BİRLİKTELİK ANALİZİ</b>	<b>47</b>
4.1. Genel Bilgiler	47
4.2. Verilerin Tanımı	47
4.3. Veriler İçin Kullanılan Yöntem ve Algoritma	50
4.4. Paket Programın Kullanımı ve Analizin Hazırlanması	51
4.5. Bulgular	57
<b>4. SONUÇLAR VE DEĞERLENDİRME</b>	<b>67</b>
<b>KAYNAKLAR</b>	<b>69</b>
<b>EKLER</b>	<b>72</b>
<b>ÖZGEÇMİŞ</b>	<b>82</b>

## TABLO LİSTESİ

### Sayfa No

Tablo 3.1. Aralık Veri Seti.....	26
Tablo 3.2. Ürün Satış Tablosu.....	30
Tablo 3.3. Hareketler ve Ürünler.....	36
Tablo 3.4. Tekli Birlikteliklerin Destek Değerleri.....	36
Tablo 3.5. Minimum Destek Değerini Sağlayan Ürünler.....	37
Tablo 3.6. İkili Birliktelikler ve Destek Değerleri.....	37
Tablo 3.7. İkili Birlikteliklerden Destek Değerini Sağlayan Setler.....	38
Tablo 3.8. Üçlü Birliktelikler ve Destek Değerleri.....	38
Tablo 3.9. Üçlü Birlikteliklerden Destek Değerini Aşan Ürün Setleri.....	39
Tablo 3.10.Üçlü Birlikteliklerden Çıkan Birliktelik Kuralları.....	40
Tablo 4.1. Alışveriş Hareketlerinin Tanımlayıcı İstatistikleri.....	49
Tablo 4.2. Veri Seti İçerisindeki En Çok Satılan Ürün Grupları.....	50
Tablo 4.3. Ürün Gruplarının Alışveriş Hareketlerindeki Dağılımı.....	59

## ŞEKİL LİSTESİ

### Sayfa No

Şekil 2.1. Veri Madenciliği ve Diğer Disiplinler.....	4
Şekil 2.2. VTBK Sürecini Oluşturan Adımlar .....	5
Şekil 2.3. Veri Madenciliğinin Uygulandığı Alanların Dağılımı.....	14
Şekil 2.4. Veri Madenciliğinin Modelleri .....	16
Şekil 3.1. Market Sepet Sınıflandırması.....	25
Şekil 3.2. Apriori Algoritması.....	35
Şekil 3.3. FP-Growth Örnek Veri Seti .....	41
Şekil 3.4. Algoritmaların Karşılaştırılması .....	46
Şekil 4.1. Ürün Gruplarının Alışveriş Hareketlerindeki Dağılımının Grafiği.....	59
Şekil 4.2. SPSS Clementine 12.0 Ürün Gruplarının Web Grafiği.....	64
Şekil 4.3. SPSS Clementine 12.0 Sıklığı Yüksek Ürün Gruplarının Web Grafiği.....	65
Şekil 4.4. SPSS Clementine 12.0 Sıklığı En Yüksek Olan Ürün Gruplarının Son Web Grafiği.....	65

## EKRAN LİSTESİ

### Sayfa No

Ekran 4.1. Veri Setinin Excel Paket Programındaki Görüntüsü .....	48
Ekran 4.2. SPSS Clementine 12.0 Boş Arayüzü, Ana Ekranı.....	51
Ekran 4.3. SPSS Clementine 12.0 Veri Seti İçin Kaynak Formatı Seçim .....	52
Ekran 4.4. SPSS Clementine 12.0 Kaynak Formata Veri Setini Tanımlama.....	52
Ekran 4.5. SPSS Clementine 12.0 Filter ve Type Özelliklerini Ekleme ve Özellikleri Bağlantı Kurma.....	53
Ekran 4.6. SPSS Clementine 12.0 Filter Özelliği .....	54
Ekran 4.7. SPSS Clementine 12.0 Type Özelliği .....	54
Ekran 4.8. SPSS Clementine 12.0 Apriori Modeli.....	55
Ekran 4.9. SPSS Clementine 12.0 Modeli Çalıştırma.....	56
Ekran 4.10. SPSS Clementine 12.0 Analiz Sonuçlarının Arayüze Aktarılışı .....	56
Ekran 4.11. SPSS Clementine 12.0 Birliktelik Analizi İçin Kullanılan Özellikler ve Kurulan Modelin Son Hali.....	57
Ekran 4.12. SPSS Clementine 12.0 Table Özelliği Kullanılarak Veri Setinin Kategorik Görüntüsü.....	58
Ekran 4.13. Ürün Grupları İçin Birliktelik Kuralları (Destek Sıralı).....	60
Ekran 4.14. Ürün Grupları İçin Birliktelik Kuralları (Güven Sıralı).....	61

## KISALTMALAR LİSTESİ

<b>VM</b>	: Veri Madenciliği
<b>VTBK</b>	: Veri Tabanlarında Bilgi Keşfi
<b>DBMS</b>	: Database Management Systems
<b>VT</b>	: Veritabanı Tablosu
<b>c</b>	: confidence
<b>s</b>	: support
<b>VLDB</b>	: Very Large Database Endowment
<b>OCD</b>	: Offline Candidate Determination
<b>CD</b>	: Count Distribution
<b>PDM</b>	: Parallel Data Mining
<b>CCPD</b>	: Common Candidate Partitioned Database
<b>IDD</b>	: Intelligent Data Distribution
<b>HPA</b>	: Hash-based Parallel Mining of Association Rules
<b>PAR</b>	: Parallel Association Rules

# BİRİNCİ BÖLÜM

## 1.GİRİŞ

### 1.1. Çalışmanın Amacı ve İçeriği

Günümüzde sürekli olarak artan veri yığınları oluşmaktadır. Gelişen teknoloji ve bilgisayarlar yardımıyla bu veri yığınları kolay bir şekilde depolanabilmektedir. Bu veri yığınlarından istenilen şekilde ve anlamlı bilgiler elde edebilmek için çeşitli istatistiksel teknikler geliştirilmiştir. Verilerin büyüklüğü, ürün ve müşterilerin sayısı göz önünde bulundurulduğunda analizlerin gözle veya elle işlem yapılamayacak düzeyde olduğu görülmüştür ve bu analizlerin otomatik olarak yapılabileceği anlaşılmıştır. Bu noktada bilgisayar programları yardımıyla veri madenciliği kavramı ortaya çıkmıştır.

Veri madenciliği, veri yığınlarındaki bağıntı ve kuralların ortaya çıkarılmasında büyük bir öneme sahiptir. Veri madenciliğinin önemli tekniklerinden biri olan birliktelik kuralları da özellikle pazar araştırmalarında sıklıkla kullanılmaktadır. Genellikle büyük süpermarketlerde oluşan veri yığınlarını, market sepet analizi altında inceleyen bu teknik, ilk olarak 1993 yılında Agrawal ve diğerleri tarafından birliktelik kuralı problemi olarak ele alınmıştır. Birliktelik analizi, veri madenciliği tekniklerinde yararlı ve ilginç sonuçlara erişimin en önemli parçasıdır.

Veri madenciliğinde birliktelik analizi çalışmaları, çeşitli alanlarda kullanılmaktadır. En yoğun olarak kullanılan alan, market sepet analizidir ve bu analiz market-mağaza verileri üzerine yoğunlaşmaktadır.

Bu çalışmanın amacı, veri madenciliğinin önemli bölgesel desenlerinden biri olan birliktelik analizi ve birliktelik kurallarını incelemek ve bu analiz hakkında örnek bir uygulama yapmaktır.

Bu çalışma beş bölümden oluşmaktadır. Birinci bölümde; çalışmanın konusu, önemi, amacı ve içeriğine ilişkin bilgiler verilmiştir. İkinci bölümde; veri madenciliği, tanımları, veri kaynakları, kullanım alanları, modelleri ve algoritmaları hakkında

genel bilgiler verilmiştir. Üçüncü bölümde; birliktelik analizi, market sepet analizi, birliktelik kuralları, kavramları, çeşitleri, matematiksel modeli ve algoritmaları anlatılmıştır. Dördüncü bölümde; birliktelik analizi üzerine yapılacak örnek uygulama anlatılmıştır. Uygulamanın amacı, verilerin tanımı, elde edilmesi ve düzenlenmesi, verilerin hazırlanması, kullanılan yöntem ve algoritma, birliktelik kurallarının belirlenmesi, analiz sonuçları ve bulgular bu bölümde sunulmuştur. Beşinci bölümde ise; uygulamanın sonucu ve sonuçlar üzerinden fikirler ve öneriler sunulmuştur.

## İKİNCİ BÖLÜM

### 2. VERİ MADENCİLİĞİ

#### 2.1. Genel Bilgiler

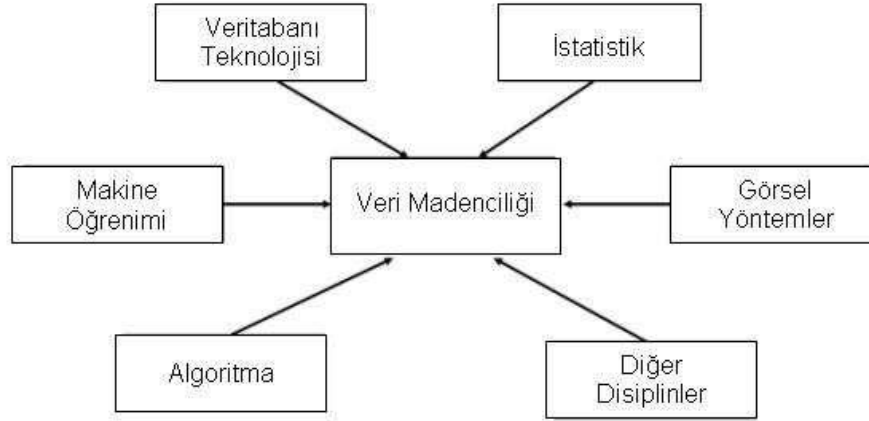
Günümüzde birçok alanda elde edilen veriler sahip olunan teknolojik olanaklar sayesinde depolanabilmekte ve bunun sonucunda oluşan veri yığınları gün geçtikçe artmaktadır. Sürekli artmakta olan bu veri yığınları beraberinde bazı sorunları da getirmektedir.

Çoğu zaman iyi kullanılmamaları durumunda veri tabanlarında tutulan veri, insanlar için problem haline de gelebilmektedir. Toplanan veri miktarı büyüdükçe ve toplanan verilerdeki karmaşıklık arttıkça, daha iyi çözümlene tekniklerine olan gereksinim de artmaktadır. Bu noktada Veri Madenciliği (VM) (Data Mining) ve Veritabanlarında Bilgi Keşfi (VTBK) (Knowledge Discovery in Databases) kavramları ortaya çıkmaktadır (Gürgen, 2008, s.2).

Kayıtlı verilerden anlamlı bilgilere ulaşım süreci Veritabanlarında Bilgi Keşfi olarak nitelendirilmektedir. Veritabanlarında bilgi keşfi, depolanmış veri içerisindeki geçerli, yeni, faydalı ve sonuç olarak anlaşılabilir örüntülerin çıkarılması sürecidir. Bu sürecin ilk adımı, uygulama alanının öğrenilmesi ile başlar. Veritabanlarında bilgi keşfinin son basamağı ise, elde edilen bilginin görüntüleme ve bilgi gösterimi yöntemleri kullanılarak kullanıcıya sunulması şeklindedir. Bazı araştırmacılar veritabanlarında bilgi keşfi ile Veri Madenciliği'ni eş anlamlı olarak kabul etmelerine rağmen, genel görüş veri madenciliği VTBK sürecinin bir aşaması şeklindedir (Şen, 2008, s.3).

VM, gelecekteki kararlara yardımcı olmak için veritabanlarından eğilimler (trends), örüntüler ve ilişkiler bulur. Buna rağmen VM sihir değildir. Bu işlemin bizi iyi sonuçlara götüreceğini hiç kimse garanti edemez. VM, sadece uzmanlara veriyi anlamada ve iyi karar vermede yardımcı olur. VM, araştırma ve çözümlenmenin birden fazla disiplin kullanılarak yapılmasıdır. Makine öğrenimi, istatistik, veritabanı

teknolojisi, uzman sistemler ve verilerin görüntülenmesi (data visualization) gibi yöntemlerin birlikte kullanıldığı bir yöntemdir (Maindonald,J.). Her bir disiplin bu veri keşfine kendi özünü katmaktadır (Dolgun, 2006, s.21).

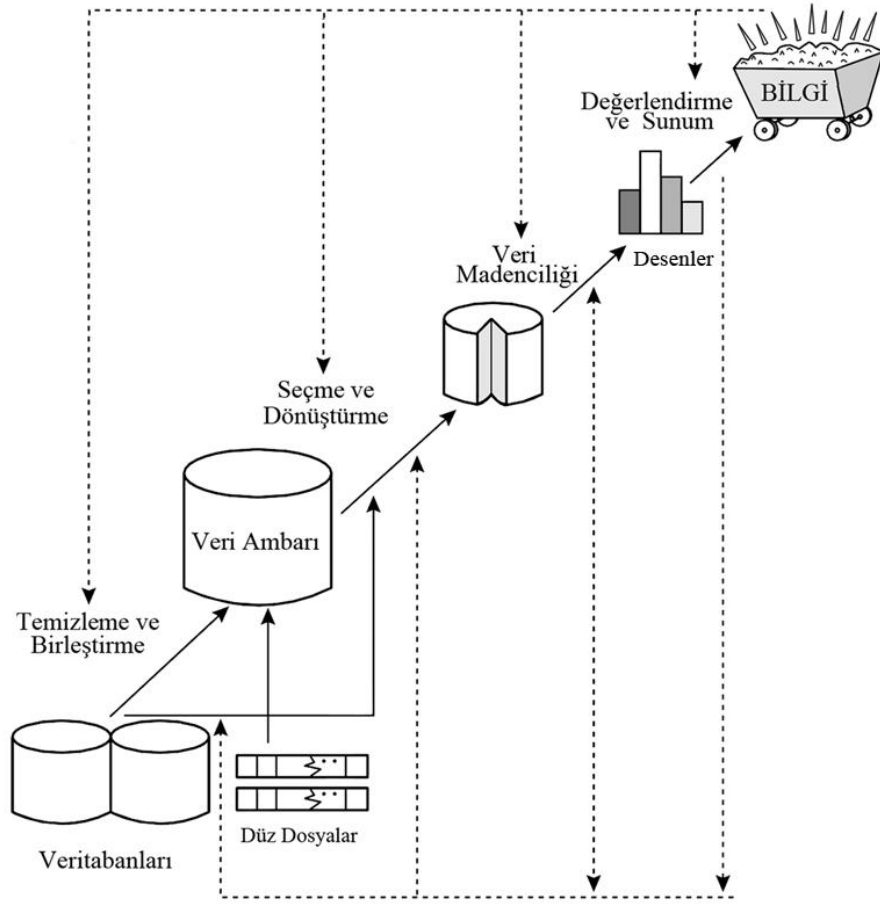


**Şekil 2.1.** Veri Madenciliği ve Diğer Disiplinler

Kaynak: Dolgun, M. Ö., 2006.

Birçok insan veri madenciliği ile eş anlamlı olarak veri tabanlarında bilgi keşfi terimini kullanmaktadır.

- **Veri Temizleme:** Tutarsız ve gürültülü verilerin veri tabanından silinmesidir. Bu aşama, keşfedilecek bilgilerin kalitesini arttıracaktır.
- **Veri Birleştirme:** Birden fazla veri tabanının bulunduğu durumlarda verilerin birleştirilmesidir.
- **Veri Seçimi:** Veri tabanından, konu ile ilgili verilerin bulunup seçilmesidir. Bu adım, birkaç veri kümesini birleştirerek sorguya uygun örneklem kümesini elde etmeyi gerektirir.
- **Veri Madenciliği:** Veriden örüntüler çıkarmak için kullanılan çeşitli yöntemleri içeren en önemli aşamadır.
- **Örüntü/Model Değerlendirme:** Veri tabanından gerçekte ilginç ve doğru olan verilerin tanımlanmasıdır.
- **Bilgi Sunumu:** Keşfedilen ve elde edilen bilgilerin geçerlilik, yenilik, yararlılık ve basitlik kriterlerine göre değerlendirilmesi ve sunulmasıdır (Karabatak, 2008, s.7).



**Şekil 2.2.** VTBK Sürecini Oluşturan Adımlar

Kaynak: J. Han, M. Kamber, 2006.

İlk iki adımda analizin amacı belirlenerek veri temini yapılırken, sonraki dört adım veri madenciliği için gerekli veri ön işleme (veri temizleme, veri birleştirme, veri dönüşümü, veri azaltma işlemleridir) sürecinin farklı şekilleridir. Veri madenciliği adımı kullanıcı veya bilgi tabanı ile etkileşebilir. İlgi çekici desenler kullanıcıya sunulur ve yeni bir bilgi olarak bilgi tabanına aktarılabilir. VTBK süreci, herhangi iki adım arasında dönen ve önem arz eden tekrarlamalara yol açabilir. Bu süreci takip ettiğinizde veri madenciliğinin yalnızca bir adım olduğunu görebilirsiniz. Ancak diğer adımların başarılı bir VTBK sürecini tamamlamak için gerekliliği olsa da değerlendirme için gizlenmiş desenlerin gün yüzüne çıkarılması için en önemli adım veri madenciliğidir (Han ve Kamber, 2006; Brachman ve Anand, 1996).

Temel olarak dört alanda meydana gelen gelişmeler veri kaynaklarını bulma, onları sorgulama, sorgulanan verileri derleme ve verilerden bilgi üretmek için analiz yapma konusunda yeni fırsatlar ve araştırma konuları doğmasına neden olmaktadır:

- 1) Veri işleme ve saklama alanındaki teknolojik gelişmeler gün geçtikçe artmakta ve daha fazla veriyi daha kısa sürelerde işlememize olanak sağlamaktadır.
- 2) Bilgisayar penetrasyonu üçüncü dünya ülkeleri de dahil yıllar içerisinde artmakta ve gün geçtikçe daha fazla kişi daha fazla dijital ortamda çalışmaya başlayarak daha fazla dijital veri üretmektedir.
- 3) İletişim teknolojileri ve internet gibi altyapılar hızla tüm dünyayı sarmakta, yer ve zamandan bağımsız bir yaşam şekli gelişmektedir.
- 4) İnsanlar daha hızlı ve doğru karar almak için veriye dayalı bir araştırma, inceleme ve muhakeme kültürünü benimsemektedir.

Veri madenciliği bu gelişmelere paralel doğmuş ve veri yığınları içindeki önceden keşfedilmemiş ilişkileri bulmaya odaklanmış sıcak bir araştırma alanıdır (Güven, Bozkurt ve Kalıpsız, 2007, s.2).

Verilerin nitelikli bir şekilde analiz edilmesi ve içerisinde gizlediği bilgilerin ortaya çıkarılması veri madenciliği ile mümkün olabilmektedir. Ayrıca veri madenciliği veriyi pratik bilgiye dönüştürerek, eylem planları oluşturmayı sağlar (Ay ve Çil, 2008, s.15).

## **2.2. Veri Madenciliğinin Farklı Tanımları**

Veri Madenciliği, beklenmeyen veya tahmin edilemeyen ilişkileri tahmin etmek, anlaşılabilir ve kullanışlı bilgileri elde etmek için büyük boyutlardaki verilerin çeşitli yöntemlerle analiz edilmesi ve özetlenmesidir.

Temel olarak veri madenciliği, veri setleri arasındaki desenlerin ya da düzenin, verinin analizi ve yazılım tekniklerinin kullanılması ile ilgidir. Veriler arasındaki ilişkiyi kuralları ve özellikleri belirlemekten bilgisayar sorumludur. Amaç daha önceden fark edilmemiş ve ya tahmin edilemeyen veri desenlerini tespit edebilmektir (Delibaş, 2010, s.4).

Veri madenciliği; veri ambarlarındaki tutulan çok çeşitli verilere dayanarak daha önce keşfedilmemiş bilgileri ortaya çıkarmak, bunları, karar vermek ve eylem planını gerçekleştirmek için kullanma sürecidir.

Veri madenciliği, eldeki verilerden; üstü kapalı, çok net olmayan, önceden bilinmeyen ancak potansiyel olarak kullanışlı bilginin çıkarılmasını sağlayan bir

tekniktir. Veri madenciliği, istatistik, makine bilgisi ve veri tabanları sistemleri ile yakında ilişkilidir (Timor, Ezerçe ve Gürsoy, 2011, s.129).

Veri madenciliği, veri tabanı kullanıcıları için veri tabanlarındaki beklenmeyen ilginç, değerli ilişkilerin bulunmasını hedefleyen süreç olarak tanımlanır (Ay ve Çil, 2008, s.15).

Veri madenciliği, büyük miktarda veri içinden gelecekle ilgili tahmin yapmamızı sağlayacak bağıntı ve kuralların bilgisayar programları kullanarak aranmasıdır. Bir başka ifadeyle büyük miktardaki verinin analiz edilerek anlamlı şablon ve kuralların keşfedilmesine imkan verir (Erpolat, 2012, s.137).

Gartner Grup tarafından yapılan tanımda ise veri madenciliği, istatistik ve matematik tekniklerle birlikte ilişki tanıma teknolojilerini kullanarak, depolama ortamlarında saklanmış bulunan veri yığınlarının elenmesi ile anlamlı yeni ilişki ve eğilimlerin keşfedilmesi sürecidir.

VM, temel olarak bilgisayar destekli bir bilgi çözümleme işlemidir.

VM, ayrı sorgular vererek büyük miktarda olan veriden yararlı bilgi, desenler, ve eğilimler (genelde önceden belli olmayan) çıkarabilmektir.

VM, verinin sahibine anlamlı ve yararlı olacak şekilde veri kümesinin içinde şüphe uyandırmayan ilişkileri bulmak ve veriyi yeni bir şekilde özetlemek için veri kümelerinin incelenmesidir (Kalıkov, 2006, s.8).

### **2.3. Veri Kaynakları**

Veri madenciliği uygulamaları için farklı veri kaynakları söz konusudur. Esas itibariyle düşündüğümüzde veri madenciliği veri akışının söz konusu olduğu, geçici veriler de dahil, herhangi bir veri kaynağı üzerine uygulanabilir. Bu bağlam ele alınacak veri kaynakları, ilişkisel veritabanları, veri ambarları, işlemsel veritabanları ve gelişmiş veritabanı sistemlerini içermektedir. Veri madenciliği teknikleri ve yöntemleri veri kaynağına göre farklılık göstermektedir (Delibaş, 2010, s.7).

Veri yığınlarının niteliklerine göre sınıflandırılması yapısal, yarı yapısal ve yapısal olmayan şeklindedir. İstatistiklere göre tüm verilerin %80'i yarı yapısal verilerdir. Bahsedilen veriler, içerisinde metin, resim, grafik vs. olan dokümanlardır (Güven, Bozkurt ve Kalıpsız, 2007, s.6).

Son zamanlarda bilgi teknolojilerinde dikkati çekecek ölçüde yaşanan gelişmeler ile depolanan veriler çok büyük boyutlara ulaşmaya başlamıştır. Moore yasasına göre, veri büyümesi aynı oranda devam ederse her 18 ayda ikiye katlanacağı düşünülmektedir. Günümüzde megabyte düzeylerinde veri toplama çok olağan iken, gigabyte düzeylerinde veri toplama giderek yaygınlaşmaktadır. Bazı ticari veritabanları world wide web ile terabyte düzeylerine kadar genişlemiştir. Hatta petabyte düzeyinde veri toplamadan yavaş yavaş bahsedilmektedir (Hegland, 2001).

VM aslında, bilgi teknolojilerinin doğal gelişim sürecinin sonucu olarak da değerlendirilebilir. Veritabanı sistemlerinin gelişimi üç ayrı fonksiyon ile değerlendirilmektedir:

- Veri toplama ve veritabanı oluşturma,
- Veri yönetimi (veri depolama, bilgi keşfi),
- Veri çözümlemesi ve veriyi anlama (veri ambarcılığı ve VM gerektirir).

Dünyada 1960'larda veri toplama sistemleri, 1970'lerde ise ilişkisel veritabanları kullanılmaya başlanmıştır. 1980'lerde ise ilişkisel veritabanlarının popüler olmaya başladığını görmekteyiz. 1990 ve 2000'lerde ise bilgisayar sistemlerindeki teknolojik gelişmelere paralel olarak ilişkisel veritabanlarında tutulan veri depoları kullanılmaya başlanmıştır.

VM birçok depolama birimi üzerinde uygulanabilir. Bunlar; ilişkisel veritabanları, veri ambarları, işlemsel veritabanları, gelişmiş veritabanları, düz dosyalar ve web olabilir. Gelişmiş veritabanı sistemleri arasında, nesne-tabanlı veritabanları ve uzaysal veritabanları, zaman serisi veritabanları, metin (text) veritabanları, multimedya veritabanları gibi özel uygulama-tabanlı veritabanları yer almaktadır. VM yöntemlerinin avantajları, üzerinde uygulandığı depolama sistemlerine göre değişiklik gösterebilir (Han ve Kamber, 2001).

Veritabanı analizinde bir bilgi birçok veri kullanılarak elde edilebilir. İş dünyası ve şirketler, etkin yönetimi sağlamak ve kazançlarını ve gelirlerini en üst düzeye çıkarmak için bilgiyi yönetmeye ihtiyaç duyarlar. Birçok fatura ve kağıt parçası içinden yöneticinin sorduğu sorulara cevap vermek zor olsa da, bilgisayarların sevdiği iş olarak bu tekrarlanan görevleri yerine getirmek ve sorulara doğru cevaplar bulmak kolaylaşmaktadır.

Veritabanı (database), sistematik erişim imkanı olan, yönetilebilir, güncellenebilir, taşınabilir, birbirleri arasında tanımlı ilişkiler bulunabilen düzenli bilgiler topluluğudur. Bir başka tanımı ise, bir bilgisayarda sistematik şekilde saklanmış, programlarca istenebilecek veri yığınlarına veritabanı denir. Bir veritabanını oluşturmak, saklamak, çoğaltmak, güncellemek ve yönetmek için kullanılan programlara *Veritabanı Yönetim Sistemleri* (Database Management Systems - DBMS) adı verilir. DBMS özelliklerinin ve yapısının nasıl olması gerektiğini inceleyen alan *Bilgi Bilimi* (Information Science)'dir (Döşlü, 2008, s.4).

#### **2.4. Veri Ambarları**

Veri ambarı, bir işletmenin ya da kurumun çeşitli birimleri tarafından canlı sistemler aracılığı ile toplanan verilerin, ileride değerlendirmeye alınabilecek olanlarının geri planda yer alan bir sistemde birleştirilmesinden oluşan büyük ölçekli bir veri deposudur (Dolgun, 2006, s.9).

Veri Ambarı, kaynak sistemlerinden boyutlandırılmış ve ya standartlaştırılmış bir veri deposuna verileri periyodik olarak getiren ve birleştiren sistemdir. Genellikle uzun yıllara ait verileri tutar ve iş zekası ve ya diğer analitik aktiviteler için sorgulanır. Veri ambarları genellikle her işlemten sonra değil belirli periyotlarda güncellenirler (Delibaş, 2010, s.8).

W. H. Inmon'a göre; Veri ambarı, özne-tabanlı, bütünleşmiş, zaman dilimli ve yöneticinin karar verme işleminde yardımcı olacak biçimde toplanmış olan değişmeyen veriler topluluğudur. Bu kısa ama ayrıntılı açıklama, veri ambarına ilişkin önemli özellikleri belirtmektedir. Özne-tabanlı, bütünleşmiş, zaman dilimli ve değişmeyen kavramları; veri ambarlarını, ilişkisel veritabanı sistemleri, hareket işleme (transaction processing) sistemleri ve dosya sistemleri gibi diğer veri ambarı sistemlerinden ayıran başlıca dört özelliktir. Bu dört özellik aşağıda açıklanmıştır.

·**Özne-tabanlı:** Bir veri ambarı; tüketici, tedarikçi firma, ürün ve satış gibi önemli konular etrafında kurulur. Veri ambarı bir organizasyonun her güne ait işleri ve hareket işleme faaliyetleri üzerinde yoğunlaşmak yerine, karar verecek kimseler için veriye ait modelleme ve çözümlere üzerinde yoğunlaşır. Bundan dolayı, veri ambarları tipik olarak ayrıntılı konu sorunları çevresinde, karar destek sürecinde faydalı olmayan veriyi hariç tutarak basit ve öz bir bakış sağlar.

· **Bütünleşik:** Bir veri ambarı genellikle; ilişkisel veri tabanları, düz dosyalar, ve OLAP kayıtları gibi çeşitli farklı türde (heterojen) kaynakları bütünleştirerek oluşturulur. Veri temizleme ve veri bütünleme yöntemleri; isimlendirmede, şifreleme yapılarında, nitelik ölçütlerinde ve benzeri konularda tutarlılığı garantilemek için uygulanır.

· **Zaman Dilimli:** Veriler, tarihi bir bakış açısından bilgi sağlamak için depolanır (örneğin 5-10 yıllık geçmiş içerisinden). Veri ambarındaki her anahtar yapı, zamanın bir elemanı olarak hem kesinlik hem de açıklık içerir.

· **Değişmeyen:** Bir veri ambarı, her zaman işlemsel (operational) çevre içerisinde bulunan uygulama verisinden dönüştürülen veri kaydından fiziksel olarak ayrıdır. Bu ayrıma bağlı olarak bir veri ambarı hareket işlemeyi, geri almayı, ve rastlantısal kontrol mekanizmalarını gerektirmez. Veriye erişim için sadece iki işlem gerektirir; verinin ilk yüklemesi ve verinin erişimi.

Özetle, veri ambarı stratejik kararları verme konusunda bir kurumun ihtiyacı olan bilgiyi depolayan ve karar destek veri modelinin fiziksel bir sunumu gibi çalışan, anlamsal olarak tutarlı bir veri deposudur. Veri ambarı aynı zamanda, yapısal ve/veya planlanmamış sorgular, analitik raporlar ve karar vermeyi desteklemek için farklı türde kaynaklardan veriyi bütünleştirerek oluşturulan bir mimari olarak da görülür (Han ve Kamber, 2001).

## 2.5. İlişkisel Veritabanları

İlişkisel veritabanlarına gelindiğinde her biri farklı isimlerde, veri özelliğine göre ayrılmış tablolarla karşılaşılmaktadır. Tablolar “alan”lar (nitelikler) kümesinden oluşmakta ve “kayıt” olarak adlandırılan satırlardan meydana gelmektedir. İlişkisel veri tabanları, kendisini oluşturan varlıklar ve onların ilişkilerinden meydana gelen veri modelleri ile tanımlanır.

İlişkisel veritabanlarına veri madenciliği uygulandığında eğilimler ve veri desenleri araştırılarak ileriye doğru adımlar atılabilir. Sayılan tüm özellikleri ile ilişkisel veritabanları, en sık kullanılan, zengin ve veri madenciliği çalışmalarındaki en gözde veri kaynaklarından biridir (Delibaş, 2010, s.7).

İlişkisel veri, SQL gibi yapısal sorgu dilleri ile yazılan sorgular ile ya da grafik kullanıcı ara yüzleri ile erişilebilir veridir. Kullanılan sorgu dili ya da kullanıcı ara

yüzünün sağladığı olanaklar doğrultusunda, veriler ile istenilen soruların yanıtları alınmaktadır (Dolgun, 2006, s.9).

VM, ilişkisel veritabanlarındaki kayıtlara ait eğilimleri (trends) çözümlmek için ya da verideki örüntüleri bulabilmek için kullanılabilir. Veritabanları en sık kullanılan VM uygulama platformlarından birisidir (Han ve Kamber, 2001).

## **2.6. İşlemsel Veritabanları**

Han ve Kamber'e göre genel anlamda işlemsel veritabanları, her kaydının bir işleme denk geldiği dosya yapısından oluşan veritabanlarıdır. Veritabanına her an kayıt yapılması söz konusudur. Bir işlem, tekil bir işlem tanımlama koduna ve işlemi gerçekleştiren birimlerin listesinden oluşur (bir mağazadan satın alınan ürünler gibi).

İşlemsel veritabanı, satışın gerçekleştirildiği şube ve ya kasa, satışı yapan personel bilgileri, müşteri bilgileri vb. bilgileri içeren, işlemle ilgili verilerin tutulduğu başka tabloları da içerebilmektedir. Bu tür veritabanlarına "A müşterisinin satın aldığı tüm ürünlerin listesi" ve ya "C2 kodlu ürüne ait kaç adet satış işlemi gerçekleştirilmiştir?" gibi sorular bir analist tarafından yöneltilebilir. Bu tür sonuçları elde etmek için işlemsel veritabanlarında tarama yapacak basit sorgular yazarak sonuçlar elde edilebilecektir. Daha derine inmek istendiğinde "Hangi ürünler birlikte daha çok satılmıştır?" sorusu sorulabilir. Bu tür işlemlerde de sıradan veri sorgulama işlemleri değil de sık görülen ürün birliktelikleri analizi gibi veri madenciliği sistemleri kullanılmaktadır.

## **2.7. Gelişmiş Veritabanı Sistemleri**

Veritabanı teknolojisinin gelişmesiyle birlikte, çeşitli veri ve bilgi sistemleri ortaya çıkmış ve yeni uygulamalara olan gereksinimi tetiklemiştir. Bu uygulamalar geniş ölçekli verilere hitap edeceğinden etkili ve güçlü veri yapılarına ihtiyaç duymaktadır. Bunlar nesne-ilişkili veritabanı sistemleri, ısısal ve zaman serisi veritabanları, kalıtım veritabanları, veri akışı yönetim sistemleri gibi yukarıda sıralanan veri gruplarının tutulduğu gelişmiş veritabanı sistemleridir.

Bu türden karmaşık ve büyük boyutlardaki veriler bir çok araştırma ve veri madenciliği uygulamasına kaynak niteliği taşımaktadır (Delibaş, 2010, s.12).

Yeni veritabanı uygulamaları, uzaysal veri (haritalar vb.), mühendislik tasarım verileri (bina tasarımları, sistem bileşenleri, devreler), multimedya veriler, zaman

eksenli veriler ve web verileri gibi veriler üzerinde işlem yapmaktadırlar. Bu tür uygulamalar, karmaşık nesne yapıları, değişken boyutlu kayıt yapıları, metin (text) ve multimedya veriler için verilerdeki dinamik değişimler açısından daha etkin veri yapıları gerektirmektedirler. Bu ihtiyaçlara cevap verebilmek için, gelişmiş veritabanı sistemleri geliştirilmiştir (Han ve Kamber, 2001).

## **2.8. Veri Ambarı ile Veritabanı Arasındaki Farklar**

Veritabanı içerisindeki bilgiler genelde anlık bilgilerdir. Yani belirli bir süre sonunda güncelliğini kaybedecek olan bilgilerdir. Ancak veri ambarı içerisinde veriler genelde yığılarak birikirler ve verilerin geçerliliği çok daha uzun sürmektedir. Veri ambarı içerisinde ne kadar çok kayıt olursa, yapılan incelemelerin sonucu da o kadar doğru olacaktır. Oysa veritabanı içerisindeki kayıt sayısının çok fazla olması durumunda, veritabanını kullanan canlı sistemlerin performansları düşecek ve bu da verilere erişimi çok yavaşlayacaktır. Bu durum, canlı sistemlerde en istenmeyen durumdur (Dolgun, 2006, s.14).

## **2.9. Veri Madenciliğinin Uygulama Alanları**

Veri madenciliği teknikleri günümüz iş hayatında bir çok alanda başarı ile kullanılmaktadır (Timor, Ezerçe ve Gürsoy, 2011, s.129).

Veri kaynakları ve buna bağlantılı olarak verinin kendisi hızla artmaktadır. Verilerin analizi ile insan tarafından ortaya çıkarılmış değerli bilgiler olmasına rağmen artık günümüzün veri yığınlarının insanlar tarafından analiz edilebilmesi mümkün değildir. Bu işlemin otomasyonu olarak değerlendirilebilecek veri madenciliği çalışmaları giderek daha fazla alanda kullanım alanı bulmaktadır (Güven, Bozkurt ve Kalıpsız, 2007, s.6).

Veri madenciliğinin hangi alanlarda ve konularda kullanılabileceği konusunda aşağıdaki bir takım başlıklar sıralanmıştır (Delibaş, 2010, s.13).

Pazarlama;

- Müşterilerin satın alma alışkanlıklarının belirlenmesi
- Müşterilerin demografik özellikleri arasındaki bağlantıların ortaya konulması
- Mevcut müşterilerin elde tutulması, yeni müşterilerin kazanılması
- Pazar sepet analizi

- Müşteri ilişkileri yönetimi
- Müşteri değerlendirme
- Satış tahmini
- Seyirci davranışları tahmini
- Taşımacılık sektöründe şube organizasyonu

#### Bankacılık;

- Farklı finansal göstergeler arasında gizli ilişkilerin ortaya konulması
- Kredi kartı dolandırıcılıklarının ve sahtekarlıklarının belirlenmesi
- Kredi kartı harcamalarına göre müşteri gruplarının belirlenmesi
- Kredi taleplerinin değerlendirilmesi

#### Sigortacılık;

- Yeni poliçe talep edecek müşterilerin tahmin edilmesi
- Sigorta dolandırıcılıklarının tespiti
- Riskli müşteri gruplarının belirlenmesi

#### Elektronik Ticaret;

- Saldırıların çözümlenmesi
- e-CRM uygulamalarının yönetimi
- Web sayfalarına yapılan ziyaretlerin çözümü

#### Fizik - Kimya;

- Organik bileşiklerinin analizi
- Toksik risk analizi
- Termik santral optimizasyonu

#### Bu alanların dışında;

- Otomatik özetleme sistemleri
- Emlak gerçek değer belirleme
- Hava tahminleri

Aşağıda yer alan şekil 2.3.'te veri madenciliğinin günümüz iş hayatında uygulandığı alanların yüzdesel dağılımı gösterilmektedir.

Bankacılık (51)	12%
Biyoteknoloji / Genetik (11)	3%
Kredi skorlama (35)	8%
CRM (52)	12%
Doğrudan pazarlama (34)	8%
e-Ticaret (11)	3%
Eğlence/ Müzik (4)	1%
Sahtekarlık tespiti (31)	7%
Şans oyunu (2)	0,01 %
Kamu uygulamaları (12)	3%
Sigortacılık (24)	6%
Yatırım / Hisse senedi (5)	1%
Junk email / Anti-spam (5)	1%
Sağlık/ İK (15)	4%
İmalat (19)	5%
Tıp/ Farmakoloji (12)	3%
Perakende (25)	6%
Bilim (17)	4%
Güvenlik / Anti-terörizm(5)	1%
Telekomünikasyon (23)	5%
Seyahat (8)	2%
Web (9)	2%
Diğer (11)	3%

**Şekil 2.3.** Veri Madenciliğinin Uygulandığı Alanlarının Dağılımı

Kaynak: Şen, F., 2008.

### 2.10. Veri Madenciliğinin Modelleri ve Algoritmaları

Veri madenciliği ilk zamanlar merkezi olarak toplanmış statik verilerin analizinde kullanılmıştır. Beklenin ötesinde yararların elde edilmesi bu tekniklerin akan veriler ve sürekli değişen dinamik veriler üzerinde de kullanılmak istenmesini doğurmuştur. Bu yönde çalışmalar halen sürmektedir.

Veri madenciliği çalışmalarını verinin nerede olduğu, biçimi ve niteliği gibi etkenler de etkilemektedir. Özellikle ilişkisel veritabanlarındaki verilerin analizi için kullanılan teknikler artık web sayfalarının, yarı yapısal tabir edilen dokümanların analizi gibi alanlarda kullanılmakta ve değişik yaklaşımların doğmasını sağlamaktadır (Güven, Bozkurt ve Kalıpsız, 2007, s.5).

Lee ve Siau (2001)'ya göre istatistik veri seçme, örnekleme, veri madenciliği ve ortaya çıkan bilginin değerlendirilmesinin kaçınılmaz bir bileşenidir. İstatistik veri

madenciliğinin sonuçlarının değerlendirilmesinde iyi ile kötünün ayrılmasında kullanılmaktadır. Veri temizleme sürecinde istatistik verideki aykırılıkların tespiti, gerektiğinde veriyi düzeltmek ve gürültüyü kestirebilmek için teknikler sunmaktadır (Delibaş, 2010, s.20).

Kümeleme ve deneylerin tasarımındaki teknikler keşif amaçlı veri analizi için oyuna katılmaktadır. İstatistikte çalışma genel olarak teknik ve modellerin teorik yönünün üzerinde durur. Sonuç olarak veri madenciliğinde önem arz eden araştırma, küçük çapta ele alınır. Buna ek olarak, veri tabanı tarafında, büyük veri kümeleri ile ilgilenen teknikler ve etkili veri yönetimi için kullanılan teknikler veri madenciliği için çok önemli konulardır (Delibaş, 2010, s.21).

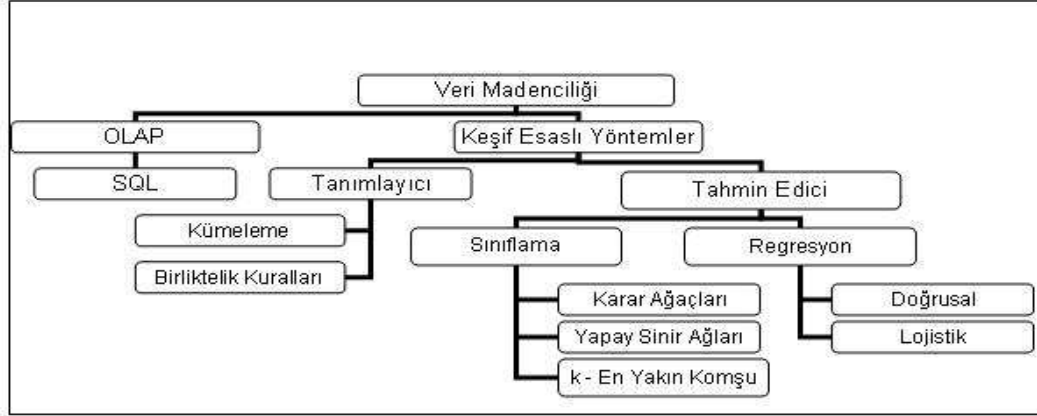
Veri madenciliği teknikleri iki amaç için kullanılır (Güven, Bozkurt ve Kalıpsız, 2007, s.4) :

- **Tanımlama:** İncelenen veri kümesinin tamamı için ya da bu veri kümesindeki bir alt küme için geçerli bir ilişkinin tanımlanmasını sağlayan tekniklerdir ve genelde daha önceden bilinmeyen ilişkiler üzerinde durduklarından önemlidirler.
- **Kestirim Yapma:** Eldeki veri kümesinin incelenmesiyle, veri kümesinin kapsamındaki konuyla ilgili gelecekte neler olabileceğini tahmin etmeye yarayan tekniklerdir.

VM modellerini gördükleri işlemlere göre,

- Sınıflama (Classification) ve Regresyon,
- Kümeleme (Clustering),
- Birliktelik Kuralları (Association Rules) ve Ardışık Zamanlı Örüntüler (Sequential Patterns), olmak üzere üç ana başlık altında incelemek mümkündür. Sınıflama ve regresyon modelleri tahmin edici, kümeleme, birliktelik kuralları ve ardışık zamanlı örüntü modelleri tanımlayıcı modellerdir (Dolgun, 2006, s.25).

Aşağıda yer alan şekil 2.4.'te veri madenciliği modellerinin genel şekli yer almaktadır.



**Şekil 2.4.** Veri Madenciliği Modelleri

Kaynak: Dolgun, M. Ö., 2006.

### 2.10.1. Sınıflama ve Regresyon Modelleri

Mevcut verilerden hareket ederek geleceğin tahmin edilmesinde kullanılan ve veri madenciliği teknikleri içerisinde en yaygın kullanıma sahip olan sınıflama ve regresyon modelleri arasındaki temel fark, tahmin edilen bağımlı değişkenin kategorik veya süreklilik gösteren bir değere sahip olmasıdır. Ancak çok terimli lojistik regresyon (multinomial logistic regression) gibi kategorik değerlerin de tahmin edilmesine olanak sağlayan tekniklerle, her iki model giderek birbirine yaklaşmakta ve bunun bir sonucu olarak aynı tekniklerden yararlanılması mümkün olmaktadır. Sınıflama ve regresyon modellerinde kullanılan başlıca teknikler:

- Karar Ağaçları (Decision Trees),
- Yapay Sinir Ağları (Artificial Neural Networks),
- Genetik Algoritmalar (Genetic Algorithms),
- K-En Yakın Komsu (K-Nearest Neighbor),
- Bellek Temelli Nedenleme (Memory Based Reasoning),
- Naive-Bayes,
- Lojistik Regresyondur (Logistic Regression).

### 2.10.2. Kümeleme Modelleri

Kümeleme modellerinde amaç, küme üyelerinin birbirlerine çok benzediği, ancak özellikleri birbirlerinden çok farklı olan kümelerin bulunması ve veritabanındaki

kayıtların bu farklı kümelere bölünmesidir. Başlangıç aşamasında veritabanındaki kayıtların hangi kümelere ayrılacağı veya kümelemenin hangi değişken özelliklerine göre yapılacağı bilinmemekte, konunun uzmanı olan bir kişi tarafından kümelerin neler olacağı tahmin edilmektedir.

### **2.10.3. Birliktelik Kuralları ve Ardışık Zamanlı Örüntüler**

Bir alışveriş sırasında veya birbirini izleyen alışverişlerde müşterinin hangi mal veya hizmetleri satın almaya eğilimli olduğunun belirlenmesi, müşteriye daha fazla ürünün satılmasını sağlama yollarından biridir. Satın alma eğilimlerinin tanımlanmasını sağlayan birliktelik kuralları ve ardışık zamanlı örüntüler, pazarlama amaçlı olarak pazar sepeti analizi (Market Basket Analysis) adı altında veri madenciliğinde yaygın olarak kullanılmaktadır. Bununla birlikte bu teknikler, tıp, finans ve farklı olayların birbirleri ile ilişkili olduğunun belirlenmesi sonucunda değerli bilgi kazanımının söz konusu olduğu ortamlarda da önem taşımaktadır (Döşlü, 2008, s.24).

Birliktelik kuralları aşağıda sunulan örneklerde görüldüğü gibi eş zamanlı olarak gerçekleşen ilişkilerin tanımlanmasında kullanılır.

- Müşteriler bira satın aldığıında, % 75 ihtimalle patates cipsi de alırlar,
- Düşük yağlı peynir ve yağsız yoğurt alan müşteriler, %85 ihtimalle diyet süt de satın alırlar.
- Ardışık zamanlı örüntüler ise aşağıda sunulan örneklerde görüldüğü gibi birbirleri ile ilişkisi olan ancak birbirini izleyen dönemlerde gerçekleşen ilişkilerin tanımlanmasında kullanılır.
- X ameliyatı yapıldığında, 15 gün içinde % 45 ihtimalle Y enfeksiyonu oluşacaktır,
- İMKB endeksi düşerken A hisse senedinin değeri % 15'den daha fazla artacak olursa, üç iş günü içerisinde B hisse senedinin değeri % 60 ihtimalle artacaktır,
- Çekiç satın alan bir müşteri, ilk üç ay içerisinde % 15, bu dönemi izleyen üç ay içerisinde % 10 ihtimalle çivi satın alacaktır (Akpınar, 2000).

Birliktelik kuralları, işlemlerden oluşan ve her bir işlemin de elemanlarının birlikteliğinden oluştuğu düşünülen bir veri tabanında, bütün birliktelikleri tarayarak, sık tekrarlanan birliktelikleri veritabanından ortaya çıkarmaktadır.

Birliktelik kuralı çıkarmada bazı algoritmalar mevcuttur. Bu algoritmalarda minimum güvenilirlik ve destek değerlerini sağlayan birliktelik kuralı çıkarımı sağlanmaktadır (Gürgen, 2008, s.16).

Birliktelik analizi ve birliktelik kurallarının temel kavramları, çeşitleri, matematiksel modeli, ilgili algoritmaları ve algoritmaların işleyişleri hakkında üçüncü bölüm'de ayrıntılı bilgiler verilecektir.

## ÜÇÜNCÜ BÖLÜM

### 3. BİRLİKTELİK ANALİZİ

#### 3.1. Genel Bilgiler

Veri madenciliği veri sahibi açısından kullanışlı ve anlaşılır olması ilkesiyle beklenmeyen ilişkileri bulmak ve yeni yöntemlerle veriyi özetlemek için çoğunlukla büyük boyutlardaki verinin analizidir. Bir veri madenciliği çalışmasından çıkarılan ilişkiler ve özetler, modeller ve ya desenler olarak adlandırılmaktadır. Doğrusal denklemler, kurallar, kümeler, grafikler, ağaç yapıları ve zaman serilerindeki tekrarlı desenler bunların bazılarıdır. Birliktelik kuralları veri madenciliğindeki bölgesel desenlerin en popüler temsilcilerindendir.

Birliktelik kural keşfi sınıflandırma kuralları ile birçok özelliği paylaşmaktadır. Her ikisi de bir veri kümesi içindeki düzenleri tanımlamak için kurallar kullanır. Fakat bu iki kural keşfi modeli aslında amaçları bakımından ayrılmaktadır. Sınıflandırma kuralı tahmin yapma kapasitesi edinme odaklı iken, birliktelik kural keşfi kullanıcıya sezme veya yapının iç yüzünü görebilme hedefine odaklanmaktadır. Bilhassa veri elementleri arasındaki beklenmeyen ilişkileri tespit etmek ve tanımlamak temeline dayanmaktadır (Delibaş, 2010, s.22).

Günümüzde, birçok alandaki veriler bilgisayarlarda ve veritabanları üzerinde saklanmaktadır. Bu verilerden, istenilen ve kayda değer bilgilere ulaşmak için kullanılan tekniklerden biri de birliktelik kurallarıdır. Birliktelik kuralı, birçok alanda geniş kullanım alanına sahiptir ve nesnelerin veya niteliklerin bir arada olma durumlarını belirlemede kullanılmaktadır. Birliktelik kuralı bulma işlemi, yoğun nesne kümesi hesaplamaya dayalı bir işlem olup büyük veritabanları üzerinde uygulanması oldukça pahalı bir işlemdir. Bu nedenle daha önceden tespit edilen birliktelik kurallarının korunması da oldukça önemli bir konu olmaktadır.

Son zamanlarda, otomatik tanıma ve veri toplama uygulamalarındaki gelişmeler sayesinde firmaların satış noktalarında barkot sistemleri kullanımı yaygınlaşmaya

başlamıştır. Bu gelişmeler ile beraber bir harekete ait satış verilerinin elektronik ortamlara aktarılmasına olanak sağlanmıştır. Genellikle büyük süpermarketlerde oluşan bu tür verilere market sepet verisi adı verilmektedir. Birçok kuruluş, market sepet verilerini kullanarak bu verilerden büyük faydalar sağlamayı amaçlamaktadır (Karabatak, 2008, s.24).

### **3.2. Market Sepet Analizi**

Market sepet verisi üzerinde birliktelik kuralı problemi, ilk olarak 1993 yılında ele alınmıştır. Sepet analizinde amaç, nitelikler (ürün satışları) arasındaki ilişkiyi bulmaktır. Bu ilişkilerin bilinmesi şirketin kârını arttırmak için kullanılabilir. Eğer  $X$  malını alan müşterilerin  $Y$  malını da çok yüksek bir olasılıkla aldıkları biliniyorsa veya bir müşteri  $X$  malını alıyor ama  $Y$  malını almıyorsa o potansiyel bir  $Y$  müşterisidir. Sepet analizi günlük işlemler sonucu elde edilen verilerden anlamlı bağıntılar çıkarmada kullanılır. “Eğer  $A$  malını alıyorsa  $\% x$  ihtimalle  $B$  malını almaya da meyillidirler” şeklinde bir sonuç  $A$  malını satan bir mağaza için çok faydalı bir bilgi olabilmektedir. Sepet analizi uygulamaları; çapraz satış (cross-selling), mağaza raflarının düzenlenmesi (layout), katalog tasarımı ve fiyatlandırma (pricing) gibi alanlarda kullanılmaktadır (Karabatak, 2008, s.24).

Geçmiş tarihli hareketleri çözümlmek karar destek sistemlerinde karar verme aşamasında verilen kararın kalitesini arttırmak için izlenen bir yaklaşımdır. 1990’lı yılların başına değin teknik yetersizlikten dolayı, kurumlara veya müşterilere satış yapıldığı anda değil belirli bir zaman aralığında (günlük, haftalık, aylık, yıllık) gerçekleşen satış hareketlerinin tamamına ilişkin genel veriler elektronik ortamda tutulmaktaydı. Barkod uygulamalarındaki gelişme, bir harekete ait verilerin satış hareketi olduğu anda toplanmasına ve elektronik ortama aktarılmasına olanak sağlamıştır. Genellikle süper marketlerin satış noktalarında bu tür veriler toplandığından, toplanan bu veriye pazar sepeti verisi adı verilmiştir. Pazar sepeti verisinde yer alan bir kayıta, tekil olan hareket numarası, hareket tarihi ve satın alınan ürünlerine ilişkin ürün kodu, miktarı, fiyatı gibi bilgiler yer alır (Agrawal et al., 1993; Han ve Kamber, 2001).

Müşterilerin satın alma davranışlarını tanımlamada kullanılan birliktelik kuralları ve ardışık zamanlı örüntüler, veri madenciliğinde, Pazar Sepet Analizi adı altında yaygın olarak kullanılmaktadır. Alışverişlerde müşterinin hangi mal ve ya hizmetleri

satın almaya eğilimli olduğunun belirlenmesi amacıyla kullanılan bu teknik, bir müşteriye daha fazla ürünün satılması yollarından biridir. Birliktelik kuralları müşteri alışveriş davranışlarını belirlemede etkili bir yöntemdir (Timor, Ezerçe ve Gürsoy, 2011, s.130).

Satın alma eğilimlerinin tanımlanmasını sağlayan birliktelik kuralları ve ardışık zamanlı örüntüler pazarlama amaçlı olarak Pazar Sepet Analizi adı altında veri madenciliğinde sıkça kullanılmaktadır. Pazar Sepet Analizi, müşterilerin alışveriş alışkanlıklarının veritabanındaki bilgiler aracılığıyla ortaya çıkartılması işlemidir (Erpolat, 2012, s.138).

Bu işlem müşterinin sepetine koyduğu farklı ürünler arasındaki birliktelikleri bularak müşterinin satın alma alışkanlıklarını analiz etmektedir. Elde edilen sonuçlar perakendecilere birlikte satılacak ürünlerden daha iyi kazançlar elde etme konusunda pazarlama stratejileri geliştirme imkanı sağlamaktadır. Örneğin süt alan müşterilerin eklemek alma alışkanlıklarının belirlenmesi ile bu iki ürün grubuna ait reyonların düzenlenmesi yapılabilecektir (Delibaş, 2010, s.22).

Buna benzer veri analizleri yaparak her ürün için bir sonraki ayın satış tahminleri çıkarılabilir, birlikte satın alınan ürünler için promosyon uygulaması ve reyon dizilişleri yapılabilir, müşteriler satın aldıkları ürünlere göre gruplandırılabilir, yeni bir ürün için potansiyel müşteriler belirlenebilir.

### **3.3. Birliktelik Kuralları**

Birliktelik kuralları (association rules), veri madenciliği alanında üzerinde çok fazla araştırma ve çalışma yapılmış olan ilgi çekici bir konudur. Birliktelik kuralları, aynı işlem içinde çoğunlukla beraber görülen nesnelere içeren kurallardır (Döşlü, 2008, s.26).

Birliktelik kuralı, geçmiş verilerin analiz edilerek bu veriler içindeki birliktelik davranışlarının tespiti ile geleceğe yönelik çalışmalar yapılmasını destekleyen bir yaklaşımdır. 90 yılların başına kadar saklanan satış verilerinde ürün ve müşteri verisi çok nadir yer alırken, genelde mali açıdan önemli olan tutarsal gelir verilerinin depolaması yapılıyordu. 90 yılların başından itibaren veri toplama uygulamalarındaki gelişmeler doğrultusunda firmaların satış noktalarında yeni teknoloji otomatik ürün veya müşteri tanıma sistemleri (barkod ve manyetik kart okuyucular) yaygınlaşmaya başlamıştır. Bu tip teknolojik gelişmeler, bir satış hareketine ait verilerin satış

esnasında toplanmasına ve elektronik ortamlara aktarılmasına olanak tanımıştır (Şen, 2008, s.24).

Birliktelik kuralında, müşterilerin alışveriş esnasında satın aldıkları ürünler arasındaki birliktelik-ilişki bağlarını bularak, müşterilerin satın alma alışkanlıklarının tespit edilmesi amaçlanmaktadır. Keşfedilen bu birliktelik-ilişki bağıntıları sayesinde satıcılar daha etkin ve kazançlı satışlar yapabileme imkanına sahip olmaktadır. Birliktelik kuralı, bir ilişkide bir niteliğin aldığı değerler arasındaki bağımlılıkları, anahtarda yer almayan diğer niteliklere göre gruplama yapılmış verileri kullanarak bulur. Keşfedilen örüntüler veri kümesinde sıklıkla birlikte geçen nitelik değerleri arasındaki ilişkiyi gösterir.

Birliktelik kurallarının kullanıldığı en tipik örnek market sepeti analizidir. Bu işlem, müşterilerin yaptıkları alışverişlerdeki ürünler arasındaki birliktelikleri bularak müşterilerin satın alma alışkanlıklarını analiz eder. Bu tip birlikteliklerin keşfedilmesi, müşterilerin hangi ürünleri bir arada aldıkları bilgisini ortaya çıkarır ve market yöneticileri de bu bilgi ışığında daha etkin satış stratejileri geliştirebilirler (Gürgen, 2008, s.18).

### **3.3.1. Birliktelik Kurallarının Temel Kavramları**

Bir veritabanı tablosunda (VT), yer alan tüm nesnelerin  $N = \{n_1, n_2, n_3, \dots, n_m\}$  olarak ifade edildiğini düşünelim. Bu durumda veritabanı tablosunda  $VT = \{H_1, H_2, H_3, \dots, H_t\}$  yer alan bir hareket kaydındaki,  $k$  nesne, tüm nesneler kümesinin bir alt kümesi,  $H_j \subseteq N$ ,  $H_j = \{n_{j1}, n_{j2}, n_{j3}, \dots, n_{jk}\}$  olacaktır. Bir birliktelik kuralı,  $X$  ve  $Y$ ,  $N$  tüm nesneler kümesinin bir alt kümesi,  $X \subset N$ ,  $Y \subset N$  ve  $X \cap Y = \emptyset$  olmak üzere  $X \rightarrow Y$  biçiminde bir bağımlılık ifadesidir. Bu ifade ile  $X$  nesnesi/nesneleri,  $Y$  nesnesini/nesnelerini belirler. Başka bir deyişle  $Y$  nesnesi/ nesneleri,  $X$  nesnesine/ nesnelere bağımlıdır. Veritabanı kayıtlarında, nesnelerin gruplandırılması ile elde edilen bağımlılık ilişkilerinin yüzde yüz geçerli olması beklenemez. Ancak, çıkarılması yapılan kuralın, veritabanının önemli bir kısmı tarafından desteklenmesi, yani söz konusu durumun sıkça görülüyor olması gerekir. Bu nedenlerden dolayı, bir  $X \rightarrow Y$  birliktelik kuralı destek ve güven eşik değerlerini sağlayacak biçimde üretilir. Destek değeri,  $X$  ve  $Y$  nesnelerinin birlikte bulunduğu kayıt sayısının, veritabanındaki tüm kayıt sayısına oranı olarak hesaplanır. Güven değeri ise,  $X$  ve  $Y$  nesnelerinin birlikte bulunduğu kayıt sayısının,  $X$  nesnesinin (veya nesnelerinin)

bulunduğu kayıt sayısına oranıdır. Destek ve güven ifadeleri 0 ile 1 arasında değişir ve ne kadar 1'e yakınlarsa nesnelere arasında o kadar güçlü ilişki olduğunu gösterirler. Dolayısıyla, bağıntının önemli olması için her iki değer de olabildiğince yüksek olması gereklidir (Birant, Kut, Ventura, Altınok ve Ihlamur, 2010, s.216).

Yüksek güven ve destek değerine sahip kurallara güçlü (strong) kurallar adı verilir (Agrawal ve diğerleri, 1993). Kullanıcı tarafından minimum destek eşik değeri (min\_destek) ve minimum güven eşik değeri (min\_güven) belirlenir. Bu belirlenen eşik değerlerini aşan birliktelik kuralları dikkate alınır ve ilginç olarak ifade edilir. İlginç bir örüntü, bilgi olarak ifade edilir. Genel olarak bu değerler 0 - 1 aralığı olarak değil, 0% - 100% şeklinde, yüzde olarak ifade edilmektedir (Han ve Kamber, 2000).

Minimum güven ve destek değerlerini sağlayan birliktelik kuralları çıkarım problemi iki adıma bölünmüştür (Agrawal ve Srikant, 1994);

**1.** Yaygın geçen nesne kümeleri bulunur: Kullanıcı tarafından belirlenmiş olan minimum destek eşik değerini sağlayan nesne kümelerine yaygın nesne küme adı verilmektedir. Bu adımda yaygın nesne kümeleri bulan etkili yöntemler kullanılmalıdır.

**2.** Yaygın nesne kümelerden güçlü birliktelik kuralları oluşturulur: Yaygın nesne kümeleri kullanarak minimum güven eşik değerini sağlayan birliktelik kurallarının bulunmasıdır. Bu adımdaki işlem oldukça basittir. Minimum güven eşik değerine göre taranarak bulunan birliktelik kuralları kullanıcının ilgilendiği ve potansiyel olarak önemli bilgiyi içeren kurallardır. Birliktelik kuralı algoritmasının performansını belirleyen adım birinci adımdır. Yaygın nesne kümeleri belirlendikten sonra, birliktelik kurallarının bulunması kolay bir adımdır (Han ve Kamber, 2000).

Sepet analizinin başarılı olduğu noktalar;

- Kolay ve anlaşılır sonuçlar üretir,
- Değişik boyutlardaki veriler üzerinde çalışabilir,
- Her ne kadar kayıtların sayısı ve kombinasyon seçimine göre işlem miktarı artsa da sepet analizi için her adımda gerekli olan hesaplamalar diğer yöntemlere göre (genetik algoritmalar, yapay sinir ağları vb.) çok daha basittir.

Sepet analizinin başarısız olduğu noktalar;

- Sorunun boyutu büyüdükçe, gerekli hesaplamalar üstel olarak artmaktadır,
- Kayıtlarda çok az rastlanan ürünleri yok sayar. Sepet çözümlemesi yönteminin en doğru sonucu, tüm ürünlerin kayıtlar içinde yaklaşık aynı frekansta görüldüğü durumlarda üretmektedir,
- Destek ve güven eşik değerleri üretilen kural sayısında sınırlama getirirler fakat eşik değerlerinin çok düşük belirlendiği durumda kullanıcı gerçekten ilgilendiği kuralları kaybetme tehlikesi ile karşı karşıya kalır (Döşlü, 2008, s.30).

Sepet analizi yöntemi, birliktelik kurallarının uygulama alanındaki türlerinden birisidir.

Birliktelik kurallarının birçok türü vardır. Birliktelik kuralları aşağıdaki kriterleri taşıyan çok değişik yollar ile sınıflandırılabilir.

**-Kuralda kullanılan değerlerin tiplerine göre:** Eğer bir kural nesnelere varlığı ve yokluğu arasındaki birliktelikler ile ilgili ise, bu duruma mantıksal birliktelik kuralıdır denir. Kurallar nicel nesnelere ya da özellikler arasındaki birliktelikleri tanımlıyor ise nicel birliktelik kuralıdır. Bu kurallarda, nesnelere için nicel değerler ya da özellikler aralıklara bölünmüştür.

**-Kuralın içerdiği verinin boyutlarına göre:** Bir birliktelik kuralındaki özellikler (attribute) ya da nesnelere sadece bir boyutu temsil ediyorlarsa, o zaman kurala tek boyutlu birliktelik kuralıdır denir.

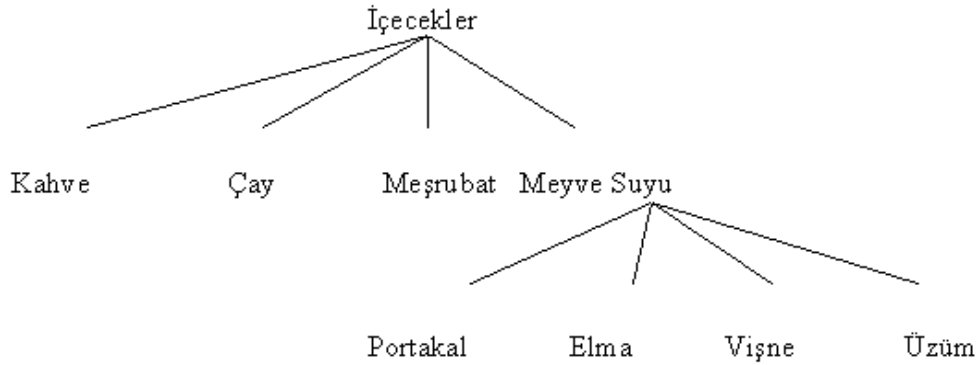
**-Birliktelik kurallarının çeşitli boyutlarına göre:** Birliktelik kuralları çözümlemesi, korelasyon çözümlemesinin genişletilmiş olabilir. Aynı zamanda, “maxpattern” (maksimum örüntü) ve “frequent closed itemset” (sık kapalı veri seti) çözümlemelerinin genişletilmiş de olabilir. Bu iki yöntem, çözümleme sırasında oluşturulan yaygın nesne kümelerinin sayısını azaltmak için kullanılmaktadır (Dolgun, 2006, s.40).

### **3.3.2. Birliktelik Kurallarının Çeşitleri**

#### **3.3.2.1. Genelleştirilmiş Birliktelik Kuralları**

Şekil 3.1’de, market sepet verileri üzerine bir sınıflandırma örneği gösterilmektedir. Burada içecekler, kahve, çay, meşrubat ve meyve suyu olarak gruplara ayrılmıştır. Meyve suyu da portakal, elma, vişne ve üzüm olarak ayrılmıştır. Birliktelik kuralları oluşturulduğunda, onları mevcut hiyerarşik seviyelerden herhangi birinde

oluşturabiliriz. Beklendiği gibi, kurallar sıralandırmada yüksek seviyedeki ürünlerden oluşturulduğunda, hem destek hem de güven artacaktır. Verilen bir işlem veritabanında, farklı ürünler için çoklu sınıflandırmalar ve hatta aynı ürün için çoklu sınıflandırmalar olabilir. Bir  $X \rightarrow Y$  *genelleştirilmiş birliktelik kuralı*, Y'deki herhangi bir ürünün X'dekilerin atası olamaması durumu dışında düzenli bir birliktelik kuralının ifadesiyle tanımlanır. Bir ürünün *atası* ifadesi, herhangi bir sıralamada o ürünün üzerinde olacağı anlamına gelmektedir. Bir süpermarket genel olarak meşrubatlarla ilgili birliktelikleri bulmak veya bu meşrubatları türlerine göre ayırarak tanımlamak (kola gibi) isteyebilir. Genelleştirilmiş birliktelik kuralları bunun gerçekleşmesini çok iyi şekilde sağlar ve tüm birliktelik kurallarının (farklı sınıflandırmalardaki farklı seviyedekiler de dahil) bulunmasını temin eder.



**Şekil 3.1.** Market Sepet Sınıflandırması

Kaynak: M.H. Dunham, Y. Xiao, L. Gruenwald, Z.Hossain, 2001.

Genelleştirilmiş birliktelik kuralı problemi tüm sınıflandırmaların tüm seviyeleri için birliktelik kuralları oluşturur. Bunu yapmak için yaklaşım, her işlemi almak ve herhangi bir hiyerarşide tüm ürünleri her bir ürünün üzerinde tutmak için her ürünü genişletmektir. Bu pek kullanılmayan yaklaşım çok pahalı olmakla beraber bu yaklaşıma ek olarak daha etkin başka algoritmalar önerilmiştir.

### 3.3.2.2. Uzaysal Birliktelik Kuralları

Bir özellik kümesinin sahip olduğu başka bir özellik kümesinde ima ettiği özelliklerin tanımını içerir. Örneğin; yapılan bir çalışmada, Dallas'daki bir ailenin ortalama geliri, White Rock gölü yakınında yaşayan aileler için 100,000\$' dır.

### 3.3.2.3. Nicel Birliktelik Kuralları

Birliktelik kuralları için yapılan çalışmalar önceleri verilerin kategorik olduğunu düşünülerek geliştirilmiştir. Verinin hem nicel hem kategorik olacağı düşüncesinin oluşması ile nicel birliktelik kuralları ortaya çıkmıştır.

Nicel veriler belirli kurallarla aralıklara bölünerek kategoriler oluşturulur ve ardışık tamsayı olarak kodlanır, böylece aralıkların sırası korunur. Veriler bölümlendirilmezse yine her biri ardışık tamsayı numaralarla kodlanır ve sırası korunur.

### 3.3.2.4. Aralık Veri Birliktelik Kuralları

Birliktelik kuralı probleminin formüleştirilmesinde, taramanın karmaşıklığının sadece özelliklerin sayısına değil, aynı zamanda bir özelliğin değerlerinin sayılarına da bağlı olduğu gözlemlenmiştir. Büyük bilgi alanlarına sahip tablolarla ilişkili birliktelik kurallarının veri madenciliği yoluyla çözümlenmesindeki karmaşıklık, verileri bir yerde gruplayarak ve kolektif düşünerek azaltılmıştır. Bir özellik doğrusal olarak sıralanmışsa, o zaman değerler gruplandırılabilir. Örneğin, bir yaş özelliğinin tüm değerlerini hesap etmektense, yaş özelliği sınıflandırılabilir (örneğin: 5 yıllık artışlar halinde). Veri değerlerinin arasındaki ayrımların anlamlı olduğu aralık verisine uygulandığında, az önce önerilen sonuçlar pek de başarı sağlamaz.

Tablo 3.1'de verilen örneği, 20K'lık bir aralığı ele alalım. Görülüyor ki herhangi bir değer taşımayan gereksiz bir sürü aralık mevcut. Bu gereksiz aralıklardan kurtulmak için daha geniş aralıklarla bölümler oluşturulmalıdır. Bu şekilde daha belirgin birliktelik kuralları elde edilebilir.

**Tablo 3.1. Aralık Veri Seti**

Kaynak: M.H. Dunham, Y. Xiao, L. Gruenwald, Z.Hossain, 2001.

Maaş	Aralık 18K-38K	Aralık 38K-58K	Aralık 58K-78K	Aralık 78K-98K
18K	1	0	0	0
30K	1	0	0	0
31K	1	0	0	0
80K	0	0	0	1
81K	0	0	0	1
82K	0	0	0	1

### 3.3.2.5. Çoklu min\_destek Birliktelik Kuralları

İncelenen önceki çalışmalar, tek destekle büyük veritabanlarında birliktelik kurallarının veri madenciliği yoluyla çözümlenmesi üzerine odaklanmıştır. Bazen tek bir min\_destek eşiği kullanımı etkin olmayabilir.

Tüm veritabanı için bir eşik desteği kullanıldığından beri, verideki tüm ürünlerin aynı doğadan ve / ya da aynı sıklıkta olduğu farz edilmiştir. Aslında, bazı ürünler nadiren görünürken bazı ürünler gayet sık olabilir. Ancak ikinci olan, ilk olandan daha kapsamlı bilgiye sahip ve daha ilginç olabilir.

### 3.3.2.6. Multimedya Birliktelik Kuralları

Multimedya veritabanları, veritabanı toplulukları arasında en çok umut veren çalışmalardan biri haline gelse de, multimedya veritabanlarında birliktelik kurallarını bulmak o kadar da dikkat çekici olmamıştır. Birçok ilişkili ve nesne yönlendirmeli veritabanı, geçmiş yıllara göre günümüzde multimedya nesnelere daha sık kullanır hale gelmiştir. Bu multimedya nesnelere fotoğraf, video, müzik, vs. gibi şeyleri içermektedir. Global internetin aşırı kullanımı, multimedya nesnelere talebinde artışa neden olmuştur. Bu verileri inceleyip aralarındaki ilişkiyi bulmak gerekli bir hale gelmiştir.

### 3.3.2.7. Maksimal Birliktelik Kuralları

Maksimal birliktelik kuralları, düzenli birliktelik kurallarının kolaylaştırılmasından çok, vasıfların kümeleri arasındaki ilişkinin daha güçlü ifade edilmesine izin vermektedir (Gürgen, 2008, s.34).

### 3.3.3. Birliktelik Kurallarının Matematiksel Modeli

Birliktelik kuralının matematiksel modeli 1993 yılında Agrawal, Imielinski ve Swami tarafından ifade edilmiştir. Bu modele göre;  $I = (i_1, i_2, i_3, \dots, i_m)$  nesnelere kümesi ve  $D$  işlemler kümesi olarak ifade edilir. Her  $i$ , bir nesne (ürün) olarak adlandırılır.  $D$  veritabanında her işlem  $T$ ,  $T \subseteq I$  olacak şekilde tanımlanan nesnelere kümesi olsun. Her işlem bir tanımlayıcı alan olan  $TID$  ile temsil edilir.  $A$  ve  $B$  nesnelere kümeleri olsun. Bir  $T$  işlemler kümesi ancak ve ancak  $A \subseteq T$  ise yani  $A$ ,  $T$ 'nin alt kümesi ise  $A$ 'yı kapsıyor denir. Bir birliktelik kuralı  $A \Rightarrow B$  formunda ifade edilir.  $A$  önce ve  $B$  sonuç olarak adlandırılır. Burada,  $A \subseteq I$ ,  $B \subseteq I$  ve  $A \cap B = \emptyset$  dir.

Hareket numaraları gruplandırılarak elde edilen ürünler arasındaki bağımlılık ilişkisinin %100 doğru olması beklenemez. Benzer şekilde, çıkarsama yapılan kuralın eldeki hareketler kümesinin önemli bir kısmı tarafından desteklenmesi istenir. Bu nedenlerden dolayı,  $A \Rightarrow B$  eşleştirme kuralı kullanıcı tarafından minimum değeri belirlenmiş güven (c: confidence) ve destek (s: support) eşik değerlerini sağlayacak biçimde üretilir.  $A \Rightarrow B$  eşleştirme kuralına, c güven ölçütü ve s destek ölçütü iliştilir ve biçimsel olarak birliktelik-ilişki  $\theta(D) = (A \Rightarrow B, c, s)$  ile gösterilir. Burada D örnekleme;  $A \Rightarrow B$  kuralını; c eşik değeri, ilgili kuralın minimum güvenilirliğini (A ürünlerini içeren hareketlerin en az %c oranında B içeren hareketler kümesinde yer aldığı); s ilgili kuralın, minimum desteğini (A ve B ürünlerini içeren hareket tutanaklarının toplam hareket tutanakları içinde en az % s oranında var olduğunu) gösterir.

Ürünler kümesi ailesini I ile gösterelim ve A ve B'nin her ikisi de I üzerinde değişebilen iki rastgele değişken olsun.  $\Pr(A)$ , A kümesi içinde yer alan tüm ürünlerin herhangi bir sepet varlığında bulunma olasılığını;  $\Pr(A \cap B)$ , A ve B rastgele değişkenlerince paylaşılan ortak ürünlerin herhangi bir sepet varlığında bulunma olasılığını; ve  $\Pr(A \cup B)$ , A ve B rastgele değişkenlerinin birleşiminde yer alan ürünlerin herhangi bir sepet varlığında bulunma olasılığını gösterebilir. O zaman, güven eşiği  $\Pr(B/A) = \Pr(A \cap B) / \Pr(A)$  ile, destek eşiği ise  $\Pr(A \cup B)$  ile ifade edilir. Güven metriği, eşleştirme kuralının doğruluk derecesini, destek metriği ise kuralda yer alan öğelerin (ürünlerin) geçiş sıklığını gösterir. Yüksek güvenilirlik ve destek değerine sahip kurallara güçlü kurallar adı verilir.

Birliktelik-ilişki kuralı formüsel olarak şu şekilde tanımlanabilir;

$$A_1, A_2, \dots, A_n \Rightarrow B_1, B_2, \dots, B_m$$

Buradaki,  $A_i$  ve  $B_j$  yapılan iş veya nesnelere. Bu kural genellikle  $A_1, A_2, \dots, A_n$  meydana geldiğinde, sık olarak  $B_1, B_2, \dots, B_m$  aynı olay veya hareket içinde yer aldığı anlamına gelmektedir.

Örneklendirmek gerekirse; aşağıdaki kural bir dijital ürün satış mağazasının satış hareketlerinden gelmektedir.

$$\text{Ürün}(X, \text{"dijital fotoğraf makinesi"}) \Rightarrow \text{Ürün}(X, \text{"bellek kartı"})$$

Burada X bir hareketteki deęişkeni simgelemektedir. Bu kural da, dijital fotoğraf makinesi alan müşterinin aynı zamanda ayrıca bellek kartı almaya yöneldiđi anlamı çıkarılmaktadır.

Başka bir örnek; aşağıdaki kural üç boyutlu bir veri ambarından gelmektedir: Yaş, Meslek ve Ürün.

Yaş (X, “12 – 17”), Meslek(X, “öğrenci”)  $\Rightarrow$  Ürün(X, ”oyun konsolu - playstation”)

Bu kural ile, “12-17 yaşları arasındaki öğrenci en çok “oyun konsolu (playstation) almaktadır” anlamı elde edilmektedir.

Yaş(X, “30...39”)  $\wedge$  gelir(X, “60K...69K”)  $\Rightarrow$  alış(X, “Plazma TV”)

Meslek(X, “öğrenci”)  $\wedge$  yaş (X, “15...20”)  $\Rightarrow$  alış(X, “Oyun Konsolu”)

Yukarıdaki ilk kuralda, otuzlu yaşlarındaki, yıllık gelirleri 60K-69K arasında olan müşterilerin Plazma TV satın almış olduğunu gösterir . Bir sonraki kural ise, yirmi yaş altı öğrenci olan müşterilerin oyun konsolu satın almış olduğunu ifade etmektedir.

### 3.3.3.1. Güven (confidence) ve Destek (support)

Kuralın destek ve güven deęerleri, kuralın ilginçliğini ve ilgililiğini ifade eden iki ölçüdür. Bu deęerler sırasıyla keşfedilen kuralların yararlılığını (kullanışlılığını) ve kesinliğini (doğruluğunu) ifade eder.

(A  $\Rightarrow$  B) güveni aşağıdaki gibi hesaplanır:

güven (A  $\Rightarrow$  B) = (A ve B'nin bulunduğu satır sayısı) / (A'nın bulunduğu satır sayısı)

(A  $\Rightarrow$  B) desteđi ise şu şekilde hesaplanır:

destek (A  $\Rightarrow$  B) = (A ve B'nin bulunduğu satır sayısı) / (toplam satır sayısı)

Başka bir ifade ile, D işlemler kümesi olarak belirlendiğinde, destek ve güven deęerleri aşağıdaki gibi hesaplanır:

Destek (A) =  $|A| / |D|$

Destek (A  $\Rightarrow$  B) =  $|A.B| / |D|$

Güven (A  $\Rightarrow$  B) = Destek (A  $\Rightarrow$  B) / Destek (A)

Güven ve destek değerlerinin örnek bir formülü şu şekildedir:

$$A \Rightarrow B [\text{destek} = \% 2, \text{güven} = \% 60]$$

Güven değerinin % 60 olduğu yukarıdaki örnekten çıkan sonuç ; A ürünü satın alanların % 60'ı B ürününü de almışlardır. Güven değerinin % 100 olması demek A ürünün alan her kişi B ürünün de almıştır anlamına gelir ve böyle kurallara kesin kural adı verilir.

Destek değeri % 2 olan yukarıdaki örnekten çıkan sonuç ise; Satılan tüm satışların % 2'sinde A ve B birlikte bulunmaktadır.

**Tablo 3.2. Ürün Satış Tablosu**

TID	Ürünler
1	Su , Ekmek, Kek, Süt
2	Su, Kek, Ekmek, Balık
3	Bira, Ekmek, Kek, Süt
4	Ekmek, Kek, Süt
5	Su, Bira, Kek, Süt

Tablo 3.2'den yola çıkarak toplam alış hareketlerine göre {Kek, Süt} ile Su arasındaki ilişki şu şekilde açıklanabilir:

$$(\text{Kek, Süt, Su}) \quad 2$$

$$\text{Destek} = \frac{\text{-----}}{\text{-----}} = \text{---} = 0.4$$

$$\text{Toplam hareket} \quad 5$$

$$(\text{Kek, Süt, Su}) \quad 2$$

$$\text{Güven} = \frac{\text{-----}}{\text{-----}} = \text{---} = 0.5$$

$$(\text{Kek, Süt}) \quad 4$$

Bu eşitliklerden de anlaşılacağı gibi, {Kek, Süt}  $\Rightarrow$  Su kuralı %40 destek, %50 güven ölçülerine sahiptir (Şen, 2008, s.25).

### **3.4. Birliktelik Analizinde Kullanılan Algoritmalar**

Agrawal, Imielinski and Swami (1993) tarafından birliktelik kurallarının keşfi problemi ilk olarak ortaya atıldıktan sonra bu işlemlerin gerçekleştirilebilmesi için çeşitli algoritmalar geliştirilmiştir. Aşağıda bu algoritmalarından bazılarına değinilecektir. Bu algoritmalarından en ilkeli AIS (Agrawal et al., 1993) ve en çok bilineni Apriori (Agrawal and Srikant, 1994) algoritmasıdır.

#### **3.4.1. AIS Algoritması**

AIS algoritması Agrawal, Imielinski ve Swami tarafından 1993 yılında veritabanındaki tüm yaygın ürün kümelerini oluşturmak için geliştirilmiş ve yayınlanmış ilk algoritmadır. Karar destek sorgulamaları yapmak için veri tabanlarının işlevlerini artırmaya odaklanmıştır. Veritabanındaki ürün isimlerinin A' dan Z' ye sıralanması kısıdını taşır (Delibaş, 2010, s.26).

AIS algoritması veritabanını birçok kez tarar ve her tarama esnasında tüm işlemleri okur. İlk tarama esnasında veritabanındaki ürünleri, teker teker sayarak hangilerinin yaygın ürünler olduğunu belirler. Bunlardan yaygın olanlar aday ürün kümeleri olarak işaretlenirler. Bu işlem tarandıktan sonra, bir önceki taramada yaygın oldukları belirlenen ürün kümeleriyle, o işlemin ürünleri arasındaki ortak ürün kümeleri belirlenir. Belirlenen bu ortak ürün kümeleri işlemde mevcut olan diğer ürünlerle birleştirilerek yeni aday kümeler oluşturulur. Herhangi bir *I* ürün kümesi, bir işlemdeki ürünlerle birleşip aday kümelere birini oluşturabilmesi için, birleşeceği kümenin hem yaygın olması hem de harf sırası açısından ürün kümesi üçündeki tüm ürünlerden sonra geliyor olması gerekmektedir (Silahtaroglu, 2008).

AIS algoritması bu adımı gerçekleştirebilmek için bir budama tekniği kullanır. Budama tekniğinin özünde, aday kümeler içindeki gereksiz kümelerin silinmesi vardır. Bu adımdan sonra, her aday kümesinin desteği hesaplanır. Destek seviyeleri minimum destek seviyesine eşit veya bu seviyeden büyük çıkanlar, yaygın ürün kümesi olarak işaretlenir. Bir sonraki taramada bu yaygın işareti taşıyan kümeler, yukarıda anlatıldığı gibi bir sonraki aday kümelerin belirlenmesi için kullanılır (Agrawal, Imielinski ve Swami, 1993).

### 3.4.2. SETM Algoritması

SETM algoritması 1995 yılında Houtsma ve Swami (1995) tarafından ortaya atılmış olup yaygın ürün kümelerinin hesaplanmasında SQL kullanılmasını baz almaktadır. Bu algoritmada yaygın nesne kümelerinin her bir üyesi  $L_k$ , TID (transaction ID) birincil anahtar olmak üzere  $\langle \text{TID}, \text{ürün kümesi} \rangle$  biçimindedir. Benzer şekilde aday kümelerinin her bir üyesi,  $C_k$  da  $\langle \text{TID}, \text{ürün kümesi} \rangle$  biçimindedir (Delibaş, 2010, s.26).

AIS algoritmasına benzer olarak, SETM algoritması da veritabanı üzerinden çoklu taramalar yapar. İlk taramada, ayrı ayrı her bir ürünün destek sayısını sayar ve veritabanında hangilerinin büyük veya sık olduğunu bulur. Daha sonra, bir önceki taramada geniş olarak işaretlenmiş ürün kümelerini kullanarak aday kümeleri belirler. Farklı olarak, SETM algoritması aday kümelerle birlikte üzerinde çalışılan işlemlerin TID bilgisini de tutar. Bundan sonra aday ürün kümeleri ürün ismine göre sıraya dizilir ve küçük nesne kümeleri silinir. Eğer veritabanı TID numarasına göre sıralanmışsa, bir sonraki tarama esnasında herhangi bir işlemdeki yaygın ürün kümeleri  $L_k$ 'nin TID numarasına göre sıralanmasıyla elde edilir. Bu şekilde veritabanı birkaç kez taranır. Artık başka herhangi bir yaygın ürün kümesi bulunamadığında algoritma sonlandırılır (Silahtaroglu, 2008).

### 3.4.3. Apriori Algoritması

Agrawal ve diğerleri tarafından 1994 yılında geliştirilen Apriori algoritması, veri madenciliği tarihinde birliktelik kurallarının çıkarılması konusunda elde edilmiş büyük bir başarıdır. Birliktelik kuralları çıkarımında en çok bilinen algoritma olmuştur. Algoritmanın ismi, yaygın nesnelerin önsel bilgilerini kullanmasından yani bilgileri bir önceki adımdan almasından “önceki (prior)” anlamında aprioridir (Döşlü, 2008, s.34).

Apriori algoritması birliktelik analizinin yapılarak birliktelik kurallarının çıkarılmasında en çok bilinen ve kullanılan algoritmadır. Tan, Steinbach ve Kumar (2005)' a göre temel prensip, “bir ürün kümesi sık görülüyorsa, onun tüm alt kümeleri de sık görülmelidir” şeklinde açıklanmıştır (Delibaş, 2010, s.27).

Şu ana kadar bilinen en iyi birliktelik kuralı algoritmasıdır. Bu yöntem, büyük bir ürün kümesinin herhangi bir alt kümesinin de yine büyük bir ürün kümesi olması gerektiği özelliğini kullanır. Aynı zamanda, bir ürün kümesindeki ürünlerin sözlük

sırasına göre sıralandığı bir yöntemdir. Bu algoritmanın, AIS ve SETM algoritmalarından temel farklılıkları aday ürün kümelerini oluşturma yolu ve aday ürün kümelerinin sayım için seçilme şeklidir. Daha önce de belirtildiği gibi, AIS ve SETM algoritmalarında, bir önceki döngüde elde edilen büyük ürün kümeleri ile işlemde kullanılan ürünler arasındaki ortak ürün kümeleri elde edilir. Bu ortak ürün kümeleri, aday ürün kümelerini oluşturmak için işlemde kullanılan başka ayrı ürünlerle arttırılır. Ancak, bu ayrı ürünler büyük ürün kümeleri olmayabilir. Bildiğimiz gibi bir büyük ürün kümesinin üst kümesi ve bir küçük ürün kümesi, küçük bir ürün kümesi olarak sonuçlanacaktır. Bu yöntemler, sonradan küçük ürün kümeleri olarak sonuçlanacak çok sayıda aday ürün kümesi oluşturacaktır. Apriori Algoritması bu önemli noktaya hitap etmektedir. Apriori, aday ürün kümelerini, bir önceki döngüde elde edilen büyük ürün kümelerini birleştirerek ve veritabanındaki işlemlere bakmaksızın bir önceki döngüde küçük ürün kümeleri olarak elde edilen alt kümelerin silinmesiyle oluşturur. Sadece bir önceki döngüde elde edilen büyük ürün kümelerini ele alarak, aday ürün kümelerinin sayısı kayda değer bir azalma gösterir. İlk döngüde, tek ürünlü ürün kümelerinin sayımı yapılır. İlk döngüde bulunan büyük ürün kümeleri ile ikinci döngünün aday ürün kümelerini oluşturmak apriori\_gen fonksiyonu kullanılır. Aday ürün kümeleri bulunduğunda, veritabanı taranarak ikinci en büyük ürün kümelerini bulmak için aday ürün kümelerinin desteklerinin sayımı yapılır. Üçüncü döngüde, ikinci döngüde bulunan büyük ürün kümeleri, bu döngünün büyük ürün kümelerini bulmak için aday ürün kümeleri olarak kullanılır. Bu iteratif işlem, daha fazla büyük ürün kümesi bulunamayınca kadar sürer ve sonra sona erer. Algoritmanın her  $i$  döngüsü veritabanını bir kez tarar ve  $i$  büyüklüğündeki büyük ürün kümelerini belirler.  $L_i$ ,  $i$  büyüklüğündeki büyük ürün kümelerini,  $C_i$  ise  $i$  büyüklüğündeki aday ürün kümelerini ifade eder.

Agrawal ve Srikant tarafından geliştirilen Apriori algoritması 1994 yılında 20. Very Large Database Endowment (VLDB) konferansında sunulmuştur. Bu bildiriye, Agrawal ve Srikant algoritmanın çalışma ayrıntılarını ve algoritmanın kaba kodunu şu şekilde sunmuştur (Gürgen, 2008, s.39):

\* Verilerin ilk taranması esnasında, geniş ürün kümelerinin tespiti için, tüm ürünler sayılır.

\* Bir sonraki tarama,  $k$  ıncı tarama olsun, iki aşamadan oluşur;

\* Apriori-gen fonksiyonu kullanılarak,  $(k-1)$ inci taramada elde edilen,  $L_{k-1}$  ürün kümeleriyle,  $C_k$  aday ürün kümeleri oluşturulur,

\* Sonra veritabanı taranarak,  $C_k$  daki adayların desteği sayılır.

\* Hızlı bir sayım için, verilen bir  $L$  işlemindeki,  $C_k$ 'yı oluşturan adayların çok iyi belirlenmesi gerekir.

Apriori algoritmasını benzer bir şekilde başka ifadelerle anlatmak gerekirse; Apriori algoritması ise aday ürünleri üretirken veritabanındaki işlemleri hiç işin içine sokmadan, yalnızca bir önceki taramada yaygın olduğu tespit edilmiş ürün kümelerini kullanarak oluşturur. Apriori algoritmasının yaygın bir ürün kümesinin herhangi bir alt kümesinin de yaygın olacağı kabulüne dayandığından bahsetmiştik. Böylece  $k$  adet üründen oluşmuş bir ürün kümesi,  $k-1$  adet ürüne sahip yaygın ürün kümelerinin birleştirilmesi ve alt kümeleri yaygın olmayanların silinmesiyle elde edilebilir. Bu birleşme ve silme işlemi sonunda daha az sayıda aday ürün kümeleri oluşacaktır.

Apriori, bir hareket veya ürün kümesi içindeki ürünlerin alfabetik şekilde sıralandığını varsaymaktadır.  $F_k$ ,  $k$  boyutlu sık görülen bir ürün kümesi ve  $C_k$  da onların adayları olsun. Apriori ilk önce veritabanını tarar ve her ürünün sayısını artırarak ve onları gereken minimum destek değeri sağlanıncaya kadar toplayarak boyutu 1 olan sık görülen ürün kümelerini arar. Daha sonra aşağıdaki üç adımı tekrar ederek bütün sık görülen ürün kümelerini bulur.

- $k$  boyutlu sık görülen ürün kümelerinden,  $C_{k+1}$  adet,  $k+1$  boyutlu aday sık görülen ürün kümesi üret.

- Veritabanını tara ve sık görülen ürün kümelerinin her adayı için destek değerini hesapla.

- Gereken minimum destek seviyesi sağlanıncaya kadar bu ürün kümelerini  $F_{k+1}$ 'e ekle.

Apriori algoritması aşağıdaki Şekil 3.2.'de gösterilmiştir. 3. satırdaki apriori-gen fonksiyonu aşağıdaki iki adımlık süreci uygulayarak  $F_k$ 'dan  $C_{k+1}$  üretir.

- **Birleşme adımı:**  $k$  boyutlu iki sık görülen, ilk  $k-1$  elemana sahip  $P_k$  ve  $Q_k$  ürün kümelerini birleştirerek  $k+1$  boyutlu sık görülen ürün kümelerinin başlangıç adayları  $R_{k+1}$  i belirle.

• **Budama adımı:**  $RK+I$ 'deki  $k$  boyutlu tüm ürün kümelerinin sık görülen olup olmadığını kontrol et ve bu gerekliliği sağlamayanları  $RK+I$ 'den çıkararak  $CK+I$ 'i üret. Çünkü  $CK+I$ 'in sık görülen olmayan  $k$  boyutlu hiçbir alt kümesi  $k+1$  boyutlu sık görülen ürün kümesinin alt kümesi olamaz.

5. satırdaki alt küme fonksiyonu  $t$  hareketindeki sık görülen ürün kümelerinin tüm adaylarını belirler. Daha sonra Apriori veritabanını tarayarak sadece bu yolla üretilen adayların frekanslarını hesaplar.

Apriori' nin sık görülen ürün kümelerinin maksimum boyutu  $kmax$  iken veritabanını en fazla  $kmax+1$  kere tarar.

Apriori aday kümelerin boyutunu indirgemede iyi performans gösterir. Bununla birlikte, çok fazla sık görülen nesne kümesi, büyük nesne kümeleri ya da minimum destek değerinin çok düşük olduğu durumlarda, çok fazla sayıda aday küme üretme ve çok sayıda aday ürün kümesini kontrol etmek için veritabanını tekrar tekrar kontrol etme gibi sorunlarla karşı karşıya kalır (Delibaş, 2010, s.30).

#### Apriori Algoritması

```

 $F_1$ =(1 elemanlı sık görülen ürün kümeleri );
for( $k = 1$ ;  $F_k \neq \phi$ ;  $k + +$ ) do begin
     $C_{k+1}$  = apriori-gen( $F_k$ ); //Yeni adaylar
    for all işlemler  $t \in$  Database do begin
         $C'_t$  = altküme ( $C_{k+1}, t$ ); // $t$  içindeki adaylar
        for all adaylar  $c \in C'_t$  do
             $c.count + +$ ;
        end
         $F_{k+1} = \{C \in C_{k+1} \mid c.count \geq \text{minimum support} \}$ 
    end
end
Answer  $\cup_k F_k$ ;

```

### Şekil 3.2. Apriori Algoritması

Kaynak: Agrawal, R., Srikant, R., 1994.

Bu algoritmayla ilgili örnek bir veri seti Tablo 3.3'te verilmiştir. Bu örnek tablo üzerinde Apriori algoritmasını çalıştırarak en çok sık geçen ürünleri bulmaya çalışalım.

**Tablo 3.3. Hareketler ve Ürünler**

Müşteri Numarası	Aldığı Ürünler
101	Elma , Şeker , Çay , Domates
102	Ekmek , Domates , Un , Şeker
103	Elma , Domates , Ekmek
104	Şeker , Çay , Domates ,Peynir,
105	Elma , Domates , Un ,Çay
106	Makarna , Domates , Çay
107	Elma , Zeytin ,Domates
108	Un , Üzüm , Çay
109	Üzüm , Şeker ,Çay
110	Çay , Makarna , Elma , Un , Domates

Birliktelik kuralları, item-setler arasındaki eğilimi ya da ilişkiyi bulur. Item set, itemlerin kümesini oluşturur. Her bir hareket, item set olarak adlandırılır. Örneğin, Tablo'daki örnekte 108 numaralı müşterinin yapmış olduğu alışverişteki “Un-Üzüm-Çay” bir item settir.

Adım 1: Minimum destek sayısı ve minimum güven değerinin belirlenmesi

Minimum Destek: 2

Minimum Güven : %70 olarak seçilmiştir.

Adım 2 : İtem setler içerisindeki her bir itemin destek değerinin bulunması(Her ürünün hareket listesindeki geçiş sayısı yani C<sub>1</sub> tablosu).

**Tablo 3.4. Tekli Birlikteliklerin Destek Değerleri**

Ürün Seti	Destek Değeri
Elma	5
Şeker	4
Çay	7
Un	4

**Tablo 3.4. (Devamı) Tekli Birlikteliklerin Destek Deęerleri**

Ürün Seti	Destek Deęeri
Ekmek	2
Domates	8
Peynir	1
Makarna	2
Üzüm	2
Zeytin	1

Adım 3 : Minimum destek deęerinden daha düşük desteęe sahip olan itemlerin devre dışı bırakılması (Destek deęeri 2 den küçük olan ürünlerin çıkarılması yani L1 tablosu)

**Tablo 3.5. Minimum Destek Deęerini Sağlayan Ürünler**

Ürün Seti	Destek Deęeri
Elma	5
Şeker	4
Çay	7
Un	4
Domates	8

Adım 4 : Elde edilen tekli birliktelikler dikkate alınarak ikili birlikteliklerin oluşturulması (C2 tablosu)

**Tablo 3.6. İkili Birliktelikler ve Destek Deęerleri**

Ürün Seti	Destek Deęeri
Elma , Şeker	1
Elma , Çay	3
Elma , Un	2
Elma , Domates	5
Şeker , Elma	1
Şeker , Çay	3
Şeker , Un	2
Şeker , Domates	3
Çay , Elma	2
Çay , Şeker	3
Çay , Un	3
Çay , Domates	5
Un , Elma	2
Un , Şeker	2

**Tablo 3.6. (Devamı) İkili Birliktelikler ve Destek Değerleri**

Ürün Seti	Destek Değeri
Un , Çay	3
Un, domates	3
Domates , Elma	5
Domates , Çay	5
Domates , Un	3
Domates , Şeker	3

Bu aşamaların her birinde, oluşturulan birlikteliklerin destek değerleri göz önüne alınarak minimum destek değeri belirlenir. Burada  $A \Rightarrow B$  ile  $B \Rightarrow A$  ikililerinin biri dikkate alınmamaktadır.

Adım 5 : Minimum destek 3 olarak seçilirse ve bu değerden düşük olan ürün setleri çıkartılırsa liste aşağıdaki Tablo'daki gibi olur. (L2 tablosu)

**Tablo 3.7. İkili Birlikteliklerden Destek Değerini Sağlayan Setler**

Ürün seti	Destek Değeri
Elma , Çay	3
Elma , Domates	5
Şeker , Çay	3
Şeker , Domates	3
Çay , Un	3
Çay , Domates	5

Adım 6 : Üçlü birlikteliklerin oluşturulması (C<sub>3</sub> tablosu). Genelde ikili birliktelikler göz önüne alınsa da veritabanındaki itemlerin birbirleri ile olan ilişkileri dikkate alınarak üçlü ve dördü veya daha fazla birliktelikler oluşturulabilir. Burada ele alınan market-basket verisine göre üçlü birliktelikler oluşturulabilir.

**Tablo 3.8. Üçlü Birliktelikler ve Destek Değerleri**

Ürün Seti	Destek Değeri
Elma , Çay , Şeker	1
Elma , Çay , Domates	3
Elma , Domates , Şeker	2
Elma , Domates , Çay	3
Elma , Domates , ekmek	1
Elma , Domates , Un	2

**Tablo 3.8. (Devamı) Üçlü Birliktelikler ve Destek Değerleri**

Ürün Seti	Destek Değeri
Elma , Domates ,Makarna	1
Şeker , Çay ,elma	1
Şeker , Çay ,Domates	2
Şeker , Çay ,Peynir	1
Şeker , Çay , üzüm	1
Şeker , Domates , Elma	1
Şeker , Domates ,Çay	2
Şeker , Domates ,un	1
Şeker , Domates ,ekmek	1
Şeker , Domates ,Peynir	1
Çay , Un , Domates	2
Çay , Un ,Elma	2
Çay , Un , Üzüm	1
Çay , Un ,Makarna	1
Çay , Domates , Şeker	2
Çay , Domates , Elma	3
Çay , Domates , Peynir	1
Çay , Domates , Un	2
Çay , Domates ,Makarna	2

Adım 7 : Üçlü birlikteliklerden minimum destek değeri olan 3 değerini geçenlerin dışındakilerin çıkarılması (L3 Tablosu)

**Tablo 3.9. Üçlü Birlikteliklerden Destek Değerini Aşan Ürün Setleri**

Ürün Seti	Destek Değeri
Elma , Çay , Domates	3

Tabloda oluşan üçlü ürün setinin ikili alt kümelerinden herhangi birisi Adım 5'teki L2 tablosunda yer almasaydı bu ürün seti de silinmiş olacaktı. Fakat (Elma, Çay), (Elma, Domates) ve (Çay, Domates) ürün setlerinin her biri L2 sık geçen öğeler tablosunda yer aldığı için (Elma, Çay, Domates) ürün seti kabul edilir.

Adım 8 : Üçlü birlikteliklerden birliktelik kurallarının çıkarılması

**Tablo 3.10. Üçlü Birlikteliklerden Çıkan Birliktelik Kuralları**

Birliktelik	Açıklama	Güven
Elma & Çay $\Rightarrow$ Domates	Elma ve Çayın bulunduğu sette domatesin olma olasılığı	$3/3 = \%100$
Elma & Domates $\Rightarrow$ Çay	Elma ve domatesin bulunduğu sette çayın olma olasılığı	$3/5 = \%60$
Çay & Domates $\Rightarrow$ Elma	Çay ve Domatesin bulunduğu sette elmanın olma olasılığı	$3/5 = \%60$
Çay $\Rightarrow$ Elma & Domates	Çayın bulunduğu sette elma ve domatesin olma olasılığı	$3/7 = \%42$
Domates $\Rightarrow$ Elma & Çay	Domatesin bulunduğu sette çay ve elmanın olma olasılığı	$3/8 = \%38$
Elma $\Rightarrow$ Çay & Domates	Elmanın bulunduğu sette çay ve domatesin olma olasılığı	$3/5 = \%60$

Tablo'dan çıkan sonuçlara göre minimum güvenilirlik değeri olan %70 barajını geçen (Elma & Çay  $\Rightarrow$  Domates) kesin kural olarak çıkmıştır. Burada dikkat edilecek husus, her (Elma & Çay ) grubunu alan kişilerin Domates de aldığı fakat her (Elma & Domates) grubunu alan kişilerin Çay almadığı sonucudur (Şen, 2008, s.42).

Bu çalışmanın uygulama bölümünde apriori algoritması üzerinden birliktelik analizi gerçekleştirileceğinden; apriori algoritması geniş olarak, ayrıntılı bir örnekle ele alınmıştır. Bu bölümde yer alan diğer algoritmalara bilgi verme amaçlı değinilmiştir.

#### **3.4.4. AprioriTID Algoritması**

Bu algoritma Apriori algoritmasına benzer bir yapıya sahiptir. Apriori\_TID, her tekrarda var olan nesne kümelerinin kodlayarak depolayan bir yapı kullanmaktadır. Bu nedenle, bir sonraki adımlar için gerekli olan nesne kümelerine ait destek değerleri veri tabanı kullanılarak elde edilmek yerine bir önceki aday kümelerine ait destek değerlerinden elde dilmektedir. Böylece, ilerleyen adımlarda nesne kümelerine ait destek değeri hesaplama işlemleri daha verimli hale dönüşmektedir. Ancak Apriori algoritması, geniş veri kümeleri üzerinde daha verimli çalışmaktadır (Karabatak, 2008, s.34).

Apriori\_TID algoritması da taramadan önce aday nesne kümelerini belirlemek için apriori-gen fonksiyonunu kullanır. Apriori'den en büyük farkı ilk geçişten sonra veritabanının, destek seviyesini bulmak için taranmamasıdır. Bu iş için  $Ck$  kullanılır.

SETM algoritmasında olduğu gibi  $C_k$ 'nin her elemanı  $\langle TID, \{X_k\} \rangle$  formundadır. Burada  $X_k$ , TID numaralı işlemde bulunan potansiyel geniş  $k$ -nesne kümesidir,  $k=1$  iken  $CI$  veritabanına karşılık gelir. Bununla beraber her nesne, nesne kümesiyle yer değiştirir,  $k>1$  olduğu durumlarda  $C_k$  algoritmanın onuncu adımında olduğu gibi üretilir,  $t$  işlemindeki  $C_k$  bir elemanı  $\langle TID, c \rangle$  şeklindedir. Burada  $c$ ,  $t$  işlemindeki  $C_k$ 'ya ait bir aday elemanıdır.  $\{c \in C_k / c\}$ . Eğer bir işlemin, herhangi bir  $k$  nesne kümesi adayı yoksa bu durumda  $C_k$ 'nin bu işlem için herhangi bir girdisi, elemanı olmayacaktır. Daha doğrusu bu işlemin TID numarasını taşıyor olacaktır. Böylece  $C_k$ 'daki girdi sayısı, özellikle bu  $k$  değerleri için, veritabanındaki işlem sayısından daha küçük olabilir. Bunun dışında yine büyük  $k$  değerleri için her girdi kendisine karşılık gelen işlemde daha küçük olabilir. Çünkü o işlemde çok az sayıda aday barınıyor olabilir. Ancak, küçük  $k$  değerleri için bunun tersi olacaktır; yani girdiler kendilerine karşılık gelen işlemlerden daha büyük olabileceklerdir (Agrawal ve Srikant, 1994).

### 3.4.5. FP-Growth Algoritması

FP-Growth algoritması, sık görülen nesne kümelerini bulmak için kullanılan bir birliktelik algoritmasıdır. Bu algoritmanın önceki çoğu algoritmadan daha etkili bir şekilde çalışarak maliyeti azalttığı görülmüştür. Bunun en büyük nedeni, tüm veritabanını daha küçük ve daha yoğun bir veri yapısı, sık desen ağacı (FP-Tree), içinde tutmasıdır (Delibaş, 2010, s.31).

Öncelikle veritabanı, Apriori'deki gibi bir kez taranarak nesne kümeler bulunur. Yaygın nesnelere destek sayılarına göre büyükten küçüğe sıralanırlar (F-list = f, c, a, b, m, p). Ardından, veritabanı bir kez daha taranarak FP-Tree oluşturulur. FP-Tree, yaygın nesnelere bulmak için gerekli tüm bilgiyi barındırır. Yaygın olmayan nesnelere ağaçta bulunmaz ve destek sayısı daha büyük olan nesnelere köke daha yakındır. Ayrıca, FP-Tree asıl veri kümesinden daha büyük değildir (Döşlü, 2008, s.45).

<i>TID</i>	<i>Nesneler</i>	<i>Yaygın nesnelere (sıralı)</i>
100	{f, a, c, d, g, i, m, p}	{f, c, a, m, p}
200	{a, b, c, f, l, m, o}	{f, c, a, b, m}
300	{b, f, h, j, o, w}	{f, b}
400	{b, c, k, s, p}	{c, b, p}
500	{a, f, c, e, l, p, m, n}	{f, c, a, m, p}

*minsup = 3*

Şekil 3.3. FP-Growth Örnek Veri Seti

Kaynak: <http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

Örüntüleri bulmak için yaygın nesnelere, f-listesine göre altkümelere bölünür. Örnek veri seti için f-list = f, c, a, b, m, p 'dir. p nesnesi bulunan örüntüler, m nesnesi bulunan ancak p nesnesi bulunmayan örüntüler, ..., c nesnesi bulunan ancak a, b, m, p nesnelere bulunmayan örüntüler, f nesnesi bulunan örüntüler gibi tarama yapılarak örüntüler bulunur (Döşlü, 2008, s.46).

FP-Growth metodu üzerinde yapılan performans çalışmaları göstermiştir ki, algoritma büyük yaygın nesne kümelerinin modellenmesinde etkili ve ölçeklenebilir bir yapıdadır ve Apriori algoritmasından daha hızlı olduğu tespit edilmiştir (Han ve Kamber, 2000).

### 3.4.6. Diğer Algoritmalar

Bu konuda bugüne kadar birçok algoritma geliştirilmiştir. Bazı algoritmalar eş zamanlı olarak birbirinden bağımsız ve habersiz birden fazla grup tarafından da geliştirilmiştir. Literatürdeki diğer birliktelik kuralları algoritmaları önceki algoritmalara benzer mantık yürütmektedirler. Aşağıda bu algoritmaların kısa bir özeti verilmiştir (Silahtaroglu, 2008):

- Bu algoritmalarından bir tanesi Apriori ve Apriori\_TID algoritmalarının bir karışımı olan Apriori-Hybrid algoritmasıdır (Agrawal ve Srikant, 1994).

Apriori ve Apriori\_TID algoritmalarının ikisini beraber kullanan melez bir algoritmadır. Algoritma çalışması esnasında her tekrarda aynı algoritma kullanılmak yerine Apriori algoritmasının avantajlı olduğu durumlarda Apriori algoritmasını, Apriori\_TID algoritmasının avantajlı olduğu durumlarda da ise Apriori\_TID algoritma yapısını kullanarak çalışan bir yapıya sahiptir. Apriori-Hybrid algoritması birçok durumda Apriori algoritmasından daha iyi sonuçlar vermektedir (Karabatak, 2008, s.34).

- Geniş nesne kümelerini belirlemek için veritabanından alınmış küçük örneklerin çok iyi sonuçlar verebileceği fikrine dayanan Offline Candidate Determination – Sıra dışı Aday Belirleme (OCD) algoritması (Mannila, Toivonen, Verkamo, 1994).

OCD yöntemi, gereksiz aday kümelerini yok etmek için önceki döngülerden elde edilen bilgilerin birleşik analiz sonuçlarını kullanır. Bir  $Y \subseteq I$  altkümesinin seyrek olup olmadığını bilmemiz için, işlemlerin en az  $(1-s)$  kadarının taranması gerekmektedir. Burada  $s$  destek eşliğidir. Bu yüzden  $s$ ' nin düşük değerleri için nerdeyse bütün ilişkili veriler okunmalıdır. Açıkça görülüyor ki eğer veritabanı çok

büyükse, veri üzerinden mümkün olduğunca az döngü yapılması önem taşımaktadır (Gürgen, 2008, s.46).

- Veritabanını küçük parçalara bölerek, bellekte işgal edilen yeri azaltıp daha hızlı sonuca ulaşma sağlayan bölümlenme (partitioning) tekniği (Toivonen, 1996).

Partition algoritması, bilinen diğer algoritmalarından biraz farklı olup veri tabanı üzerindeki en iyi tarama özelliğine sahip algoritmadır. Veri tabanını üzerinde en fazla iki tarama işlemi gerçekleştirerek veri tabanını n adet parçaya ayırmakta ve her parça için yoğun nesne kümelerini hesaplamaktadır. Algoritma, daha sonra tüm parçaların yoğun nesne kümelerini birleştirilmekte ve süper bir yoğun nesne kümesi hesaplamaktadır. Ancak tüm veri tabanında yoğun olan nesne kümelerinin, her parça setinde de yoğun olması gerekmektedir. Partition algoritmasının temel avantajı Giriş/Çıkış maliyetini azaltmak ve yoğun kümeleri sayarken ana hafızayı kullanmaktır (Karabatak, 2008, s.35).

- 1996'da Toivonen tarafından ortaya atılan ve veritabanındaki tarama sayısını azaltan örnekleme (sampling) (Savasere, Omiecinski, Navathe, 1995) tekniği.

Sampling, en iyi durumda tarama sayısını bire düşürür ve en kötü durumda da bu sayı iki olur. Ana belleğe sığacak bir örnek öncelikle veritabanından alınır. Örnekteki büyük ürün kümelerinin kümesi, Apriori gibi yüksek seviye bir algoritma metodu kullanılarak bu örnekten bulunur (Gürgen, 2008, s.49).

- Kullanıcıya her taramadan sonra oluşan kuralları gösterip, minimum destek ve güven seviyelerini değiştirme olanağı veren CARMA (Continuous Association Rule Mining Algorithm - Sürekli Bağlantı Kuralı Madenciliği) (Hidber, 1999)

Carma algoritması, birliktelik kurallarının çevrimiçi olarak hesaplanması için önerilen bir algoritma olup veri tabanı üzerinden tam iki kez tarama gerçekleştirmektedir. Veri tabanının ilk taranması esnasında potansiyel yoğun nesne kümeleri ile ilgili bir bölüm oluşturulur ve istenirse destek değeri bu aşamada değiştirilebilmektedir. Veri tabanının ikinci taranmasında ise algoritma her bölümdeki nesne kümesinin destek değerini hesaplamakta ve yoğun olmayan nesne kümelerini temizlemektedir. Bölümler oluşturulurken her nesne kümesi ile ilgili alt ve üst sınır değerlerine göre yeni adaylar bu bölümlere eklenir veya çıkarılır. Ayrıca, sayma işlemi ikinci taramada yer almaktadır (Karabatak, 2008, s.37).

- Veri paralelliğine dayanan Count Distribution – Sayım Dağılımı (CD) (Agrawal ve Shafer, 1996).

- Parallel Data Mining - Paralel Veri Madenciliği (PDM) (Park, 1995), DMA (Distributed Mining Algorithm - Dağıtılmış Madencilik Algoritması) (Cheung vd. 1996).

PDM, doğrudan hashing tekniğini içerecek şekilde CD'nin değiştirilmesiyle oluşturulmuştur (Park, 1995). Hash tekniği sonraki geçişteki bazı adayları budamak için kullanılır.

- Common Candidate Partitioned Database - Ortak Aday Bölünmüş Veritabanı (CCPD) (Parthasarathy, Zaki, Ogihara ve Li, 2001).

1996 yılında Zaki'nin önerdiği CCPD, bazı iyileştirmelerle SGI Power Challenge paylaşımli bir bellek üzerinde CD'nin yürütülmesidir. Paylaşımli bir bellek ortamında etkili bir şekilde aday üretilmesi ve sayılması için teknikler önermektedir. Yaygın nesne kümelerini ortak öneklere (genellikle ilk nesne) göre eşdeğer sınıflar şeklinde gruplandırır ve her bir eşdeğer sınıftan adayları oluşturur. Yaygın nesne kümelerinin gruplandırılması aday sayısını azaltmazken aday üretim süresini düşürür. Her bir harekette adayların etkili bir şekilde sayılması için kısa-döngülü (short-circuited) bir altküme kontrol metodunu öne sürmektedir (Döşlü, 2008, s.52).

- Intelligent Data Distribution (IDD) (Han, Karypis ve Kumar, 1997).

Aday ürün kümelerinin, aday ürün kümelerindeki ilk ürüne dayanarak işlemciler arasında bölümlere ayırır. Bu sayede aynı ilk ürüne sahip aday ürün kümeleri aynı bölümde bulunurlar. Bu yüzden, her işlemcinin sadece işlemciye atanmış ürünlerden biriyle başlayan alt kümeleri kontrol etmesi gerekir (Gürgen, 2008, s.52).

- Hash-based Parallel Mining of Association Rules – Bağlantı Kurallarının Çırpı Temelli Paralel Madenciliği (HPA) (Masaru ve Kitsuregawa, 1996).

HPA, adayları farklı işlemcilere dağıtmak için bir hashing tekniği kullanır (Shintani, 1996). Örneğin; her bir işlemci, kendisine dağıtılmış olan adayları hesaplamak için aynı hash fonksiyonunu kullanır. Hesaplama, işlemciler arasında bölünmüş veritabanlarında hareket etmek yerine, aynı hash tekniğiyle hareketlerin alt nesne kümelerini hedef işlemcilere hareket ettirirler. Böylelikle, bir hareketteki alt nesne kümesi  $n$  adet işlemci yerine sadece bir işlemciye gider. HPA, çarpık isleme tarzı

(skew handling) tekniğiyle daha da geliştirilmiştir (Shintani, 1996). Çarpık işleme tarzı teknik, her bir işlemcide kullanılabilecek ana hafıza olması durumunda her bir işlemciye düşen isin daha orantılı olması için bazı adayların kopyalanmasıdır.

• Parallel Association Rules - Paralel Bağlantı Kuralları (PAR) (Zaki, 1997)

PAR, farklı adayları bölme ve saymada kullanılan bir grup algoritmadan oluşur (Zaki, 1997). Doğal yatay veritabanı bölmelerine (hareket listeleri) karşın dikey veritabanı bölmeleriyle (her bir nesnenin TID listesi) çalışırlar. Veritabanları için dikey organizasyonu kullanarak nesne kümesindeki nesnelerin TID listelerinin kesişimi ile bir nesne kümesinin hesaplanması daha kolay yapılmaktadır. Buna rağmen, veritabanı yatay olarak düzenlenmişse, dikey bölmelere dönüşümü gerekmektedir (Döşlü, 2008, s.54).

### **3.4.7. Algoritmaların Karşılaştırılması**

Ticari niteliğe sahip birçok kullanılabilir uygulama, birliktelik kuralları oluşturmak için Apriori tekniğini baz alırlar.

Bazı algoritmalar belirli durumlar altında kullanmak için son derece uygundur. AIS, veritabanındaki nesnelerin çok büyük sayıda olması durumunda iyi çalışmaz. Bundan dolayı AIS, küçük işlemsel veritabanları için uygundur. Daha önce de bahsedildiği üzere, Apriori ilk geçişlerde Apriori-TID'e oranla küçük çalışma sürelerine ihtiyaç duyar. Diğer yandan, Apriori-TID sonraki geçişlerde Apriori'ye göre daha iyidir. Bundan dolayı uygun geçişle, Apriori-Hybrid en iyi performansı göstermektedir. Buna rağmen, Apriori'den Apriori-TID'e geçiş çok önemli ve ek yük getiren bir işittir. OCD, tahmini teknik olmasına rağmen, düşük destek eşik değerlerindeki yaygın nesne kümelerini bulmakta çok etkindir. CARMA, kullanıcı etkileşimli, geri dönüşüm tabanlı online bir tekniktir ve hareket dizilerinin bir ağdan okunması sırasında etkilidir.

Şekil 3.4'te, bahsedilen çeşitli algoritmaların kısaca karşılaştırılmasına ışık tutmakta ve özetlemektedir. Bu şekil, azami tarama sayısını, önerilen veri yapılarını ve bazı açıklamaları içermektedir (Döşlü, 2008, s.58).

Algoritma	Tarama Sayısı	Veri Yapısı	Açıklamalar
AIS	m+1	Belirtilmemiş	Belli başlı, az veri içeren seyrek hareketsetel veritabanları için uygundur.
SETM	m+1	Belirtilmemiş	SQL uyumlu.
Apriori	m+1	$L_{k-1}$ : Hash tablosu $C_k$ : Hash ağacı	Belli başlı, orta düzey veri içeren hareketsetel veritabanları için uygundur. AIS ve SETM'e oranla daha iyidir. Paralel algoritmaların baz aldığı algoritmadır.
Apriori-TID	m+1	$L_{k-1}$ : Hash tablosu $C_k$ : TID ile dizinlenmiş dizi $\overline{C_k}$ : Sıralı mimari ID: bitmap	$\overline{C_k}$ 'nin büyük olması durumunda çok yavaştır. Küçük boyutlu $\overline{C_k}$ ile Apriori'den daha iyidir.
Apriori-Hybrid	m+1	$L_{k-1}$ : Hash tablosu <u>İlk safhada:</u> $C_k$ : Hash ağacı <u>İkinci safhada:</u> $C_k$ : TID ile dizinlenmiş dizi $\overline{C_k}$ : Sıralı mimari ID: bitmap	Apriori'den daha iyidir. Buna rağmen, Apriori'den Apriori-TID'e geçiş ek yük gerektirir. Geçiş noktasını saptamak çok önemlidir.
OCD	2	Belirtilmemiş	Düşük destek eşik değerleri ve çok büyük veritabanları için uygundur.
PARTITION	2	Hash tablosu	Büyük veritabanları için uygundur. Homojen veri dağılımını destekler.
Sampling	2	Belirtilmemiş	Düşük destek eşik değerleri ve çok büyük veritabanları için uygundur.
DIC	Aralık boyutuna bağlıdır	Ağaç	Veritabanı, hareketsetel aralıklar şeklindedir. Yüksek boyutlu adaylar, bir aralığın sonunda oluşturulur.
CARMA	2	Hash tablosu	Hareket dizilerinin bir ağdan okunmasına uygundur. Online. Kullanıcılardan sürekli geri-besleme gelir ve destek ve/veya güven değerlerini işlem sırasında herhangi bir anda değiştirebilirler.
FP-Growth	2	FP-Tree	Aday üretimsiz bir tekniktir. Apriori ve türevlerinden daha iyi performansa sahiptir.
CD	m+1	Hash tablosu ve ağacı	Veri paralelleştirme.
PDM	m+1	Hash tablosu ve ağacı	Veri paralelleştirme, erken aday budaması.
DMA	m+1	Hash tablosu ve ağacı	Veri paralelleştirme, aday budaması.
CCPD	m+1	Hash tablosu ve ağacı	Veri paralelleştirme, paylaşımlı-bellek içeren makinelerde kullanılır.

**Şekil 3.4.** Algoritmaların Karşılaştırılması

Kaynak: M.H. Dunham, Y. Xiao, L. Gruenwald, Z.Hossain, 2001.

Çalışmanın dördüncü bölümünde, çeşitli alışveriş hareketlerinden oluşan gerçek bir veri seti üzerinden apriori algoritması ve SPSS Clementine 12.0 paket programı kullanılarak birliktelik kuralları belirlenecek ve yorumlanacak, örnek bir birliktelik analizi uygulaması yapılacaktır.

## DÖRDÜNCÜ BÖLÜM

### 4. BİRLİKTELİK ANALİZİ UYGULAMASI

#### 4.1. Genel Bilgiler

Veri madenciliği alanında birliktelik kuralları ile ilgili çeşitli uygulamalar bulunmaktadır. Birliktelik Analizi ve Market Sepet Analizi konularına ilişkin yapılan çalışmalarda genellikle bir marketteki alışveriş hareketlerinden oluşan veri setlerine birliktelik kuralları uygulanarak, ürün grupları arasındaki yapılar ve ürünler arasındaki ilişkiler tespit edilmeye çalışılmıştır. İlgili algoritmalar kullanılarak verilerin analizi gerçekleştirilmiş ve ürünler arasındaki birliktelik kuralları ortaya çıkarılmıştır. Market Sepet Analizi'nde ilgili algoritmalar genellikle Apriori algoritması tercih edilmektedir. Bu çalışmada da SPSS Clementine 12.0 paket programı yardımı ile gerçek bir veri seti üzerinden Apriori algoritması aracılığıyla birliktelik kuralları belirlenecektir.

#### 4.2. Verilerin Tanımı

Bu uygulamada kullanılan veri setinde gizlilik ilkelerine bağlı kalınarak, müşteri, fatura, mağaza ismi ve yeri belirtilmemiştir. Veri seti, dünya çapında pazar araştırmaları yapan lider bir şirketin, Türkiye'deki panel bölümünün, geleneksel ve modern kanalındaki marketler kullanılarak elde edilmiştir. Uygulamada kullanılan veri setinde, Türkiye'de bulunan bu marketlerden alım yapan 1395 müşterinin alışveriş hareketleri, ürün grupları bazında elde edilmiştir. İfade edilen 1395 müşterinin, 25140 alışveriş hareketi bulunmaktadır. Veri setinde 36 ürün grubu bulunmaktadır. Bu veri setinin Excel paket programındaki görüntüsü Ekran 4.1'de gösterilmektedir.

Fam Code	ALKOLLU İÇECEKLER	BAHARATLAR	BAKLIYAT	BEYAZETLER	BİRALAR	BİSKUVİLER	CAYLAR	CEREZLER	ÇİKOLATA KAPLAMALAR	ÇİKOLATALAR	ET MAMULLERİ	ET VE TAVUK SULARI	GAZLİ MESRUBATLAR	HAZİR CORBALAR
10000	1	0	0	1	0	1	0	0	1	1	1	0	0	0
10001	0	0	1	1	0	1	1	0	0	0	1	0	1	0
10002	0	1	1	0	0	1	0	1	1	1	0	0	0	0
10003	0	1	0	1	0	1	0	1	1	0	1	0	1	0
10004	0	0	1	0	0	1	0	1	1	1	1	0	1	0
10005	0	1	1	1	0	1	1	0	1	1	1	0	0	0
10006	0	0	1	1	0	1	0	1	1	0	1	0	1	1
10007	0	1	1	1	0	1	1	1	1	1	0	0	0	1
10008	0	0	0	1	0	1	1	1	1	0	1	0	1	0
10009	0	0	0	1	0	1	0	0	0	1	1	0	0	0
10010	0	1	1	1	0	1	0	1	1	1	1	0	0	0
10011	0	1	1	1	0	1	1	1	1	0	1	0	1	0
10012	0	0	0	1	0	1	0	1	0	1	0	0	1	0
10013	0	1	1	1	0	1	0	0	1	1	0	0	0	0
10014	0	1	1	1	0	1	1	1	0	0	0	0	0	0
10015	0	1	1	1	0	1	1	1	1	0	1	0	1	0
10016	0	0	1	1	0	1	0	0	1	0	1	0	1	0
10017	0	1	1	1	0	1	1	1	1	1	1	1	0	1
10018	0	1	0	1	0	1	0	0	1	1	1	1	0	0
10019	0	1	1	0	0	1	1	1	0	1	1	0	0	0
10020	0	1	0	1	0	0	1	0	0	0	0	0	0	0
10021	0	0	1	0	0	1	0	1	1	0	1	0	1	0
10022	0	1	0	1	0	1	0	0	1	0	1	0	1	1
10023	0	1	0	1	0	1	1	1	0	1	1	0	1	0
10024	1	0	1	1	0	1	0	0	1	1	0	0	0	0
10025	0	1	1	1	0	0	1	0	0	0	1	0	1	0
10026	0	0	1	1	0	0	0	1	1	1	1	0	0	0
10027	0	1	1	0	0	1	1	1	1	1	1	1	0	0
10028	0	0	1	1	0	1	1	1	1	1	1	1	1	0
10029	0	1	1	0	0	1	0	1	1	1	1	0	1	0
10030	0	0	1	0	1	1	1	1	0	0	1	0	1	0
10031	0	0	1	1	1	1	1	1	1	1	1	1	1	0
10032	0	0	1	0	1	1	1	1	0	1	0	0	1	0

**Ekran 4.1.** Veri Setinin Excel Paket Programındaki Görüntüsü

Veri setinde 36 ürün grubu bulunduğu ifade edilmiştir. Bu ürün grupları şunlardır; “ALKOLLU İÇECEKLER”, “BAHARATLAR”, “BAKLIYAT”, “BEYAZETLER”, “BİRALAR”, “BİSKUVİLER”, “CAYLAR”, “CEREZLER”, “ÇİKOLATA KAPLAMALAR”, “ÇİKOLATALAR”, “ET MAMULLERİ”, “ET VE TAVUK SULARI”, “GAZLİ MESRUBATLAR”, “HAZİR CORBALAR”, “HELVALAR”, “KETCHAPLAR”, “KREMCİKOLATALAR”, “KURUYEMİSLER”, “MADENSULARI”, “MAKARNALAR”, “MARGARINLER”, “MAYONEZLER”, “MEYVESULARI”, “MISIRGEVREKLERİ”, “PEYNİRLER”, “SALCALAR”, “SIVI YAĞLAR”, “SUTLER”, “TEREYAGLAR”, “TOZKESMESEKERLER”, “TÜRKKAHVELERİ”, “TUZ”, “UNLAR”, “YOGURTLAR”, “YUMURTALAR”, “ZEYTİNLER”.

Bu ürün grupları, veri kaynağı tarafından belirlenmiştir. Ürün grupları içerisindeki ürünler ayrıntılı olarak belirtilmemiştir. Bunun yanında, özel içeriğe sahip bazı ürünlerin özel olarak gruplandırıldığı görülmektedir. Örneğin; Biralara, Çikolata Kaplamalar, Krem Çikolatalar, Çerezler bu şekilde düşünülerek ayrı gruplara konulmuş olabilir.

Ekran 4.1’de gösterilmiş olan veri setinin birliktelik kuralları analizinin yapılabilmesi için verinin kategori bazında kodlanmış olması gerekmektedir. Yani tüm veri seti 0-1 sayıları ile kodlanmalıdır. Bu veri setinde müşterilerin satın aldığı ürün grupları “1” ile, satın almadıkları ürün grupları “0” ile kodlanmıştır. Bu ikili kodlamalar müşterinin ilgili marketten o ürün grubunu alıp almadıklarını göstermektedir. Ayrıca Ekran 4.1’de görülen “Famcode” sütunu müşterileri ifade etmektedir. “Famcode” sütunundaki sayılara ise ilgili müşterinin temsili sırası, fiş numaraları denilebilir.

Veri madenciliğinde veri kümesinin büyüklüğünden kaynaklanan en fazla zaman alıcı aşama, verilerin ön işlemden geçirilmesi aşamasıdır. Bu veri seti, analize uygun olacak şekilde ön işlemden geçirilmiş, kategori bazında kodlanmış olarak elde edilmiştir.

Analizde kullanılacak bu veri seti ışığında, bir fişteki en az hareket sayısı 1, bir fişteki en çok hareket sayısı 32 olarak hesaplanmıştır. Yani, en az 1 ürün grubundan alışveriş yapmıştır, en çok 32 ürün grubundan alışveriş yapmıştır. 25140 fiş hareketi göz önünde bulundurulduğunda bir fişteki ortalama hareket sayısı 18, medyan yani ortanca hareket sayısı 18 ve mod (tepe değeri) yani en çok tekrar eden hareket sayısı 20 olarak hesaplanmıştır. Tablo 4.1.’de veri setinin tanımlayıcı istatistikleri gösterilmektedir.

**Tablo 4.1. Alışveriş Hareketlerinin Tanımlayıcı İstatistikleri**

	Fişlerdeki Hareket Sayıları
En Az Hareket Sayısı	1
En Çok Hareket Sayısı	32
Ortalama	18
Medyan	18
Mod	20

Veri setindeki en çok satın alınan ürün grupları ve ürün gruplarının satın alınma sayıları ve yüzdeleri Tablo 4.2.’de gösterilmektedir.

**Tablo 4.2. Veri Seti İçerisindeki En Çok Satılan Ürün Grupları**

En Çok Satılan Ürün Grupları	Satılma Sayısı	Yüzdeler %
Sütler	1329	5,29
Peynirler	1296	5,16
Bisküviler	1232	4,90
Yumurtalar	1143	4,55
Yoğurtlar	1138	4,53
Çikolata Kaplamalar	1005	4,00
Makarnalar	1004	3,99
Bakliyat	984	3,91
Kuruyemişler	958	3,81
Gazlı Meşrubatlar	950	3,78

### **4.3. Veriler İçin Kullanılan Yöntem ve Algoritma**

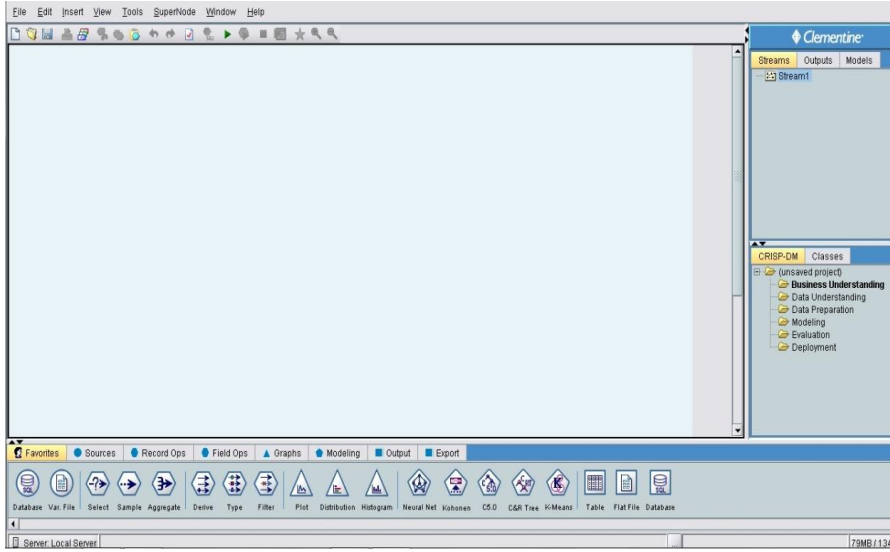
Analizde SPSS Clementine 12.0 paket programı kullanılacaktır. SPSS Clementine 12.0 paket programındaki “Modelling” özelliğinde bulunan “Apriori” algoritması modeli, birliktelik analizini gerçekleştirecektir.

SPSS Clementine paket programı uygulamacıya, kolay kullanabileceği bir arayüz ve kısa sürede sonuca ulaşabileceği görsel bir çalışma ortamı sunmaktadır. Araştırmacının kullanmak isteyeceği tüm fonksiyonlar arayüzün alt tarafında bulunan bir araç çubuğunda ikonlar halinde yer almaktadır. Yapılması gereken, gerekli ikonların analiz sayfasına taşınarak birbirlerine bağlanması ve çalıştırılmasıdır.

Veri madenciliği ve bilgi keşfi konusunda yayın ve araştırma yapan KDnuggets dergisi, “En çok tercih edilen veri madenciliği çözümü” konulu bir araştırma yapmış, üçüncüsü düzenlenen uluslar arası geleneksel anket çalışmalarının tümünde uzman sektörel kullanıcılar, kendilerine yöneltilen sorulara verdikleri yanıtlarda SPSS Clementine’i birinci sıraya taşımışlardır.

Yaygın kullanılması ve tercih edilen bir program olması, kullanım kolaylığı, farklı kaynaklardan elde edilen veri setlerinin kolay entegre edilebilmesi özelliklerinden dolayı bu tezin uygulama bölümünde analizlerin SPSS Clementine paket programı ile gerçekleştirilmesine karar verilmiştir (Gürgen, 2008, s.62).

Ekran 4.2’de, SPSS Clementine 12.0 paket programının ana ekranı görülmektedir. Analizin, “Apriori” algoritması modeli ile birliktelik kuralları bu ekran üzerinden hareketle belirlenecektir.

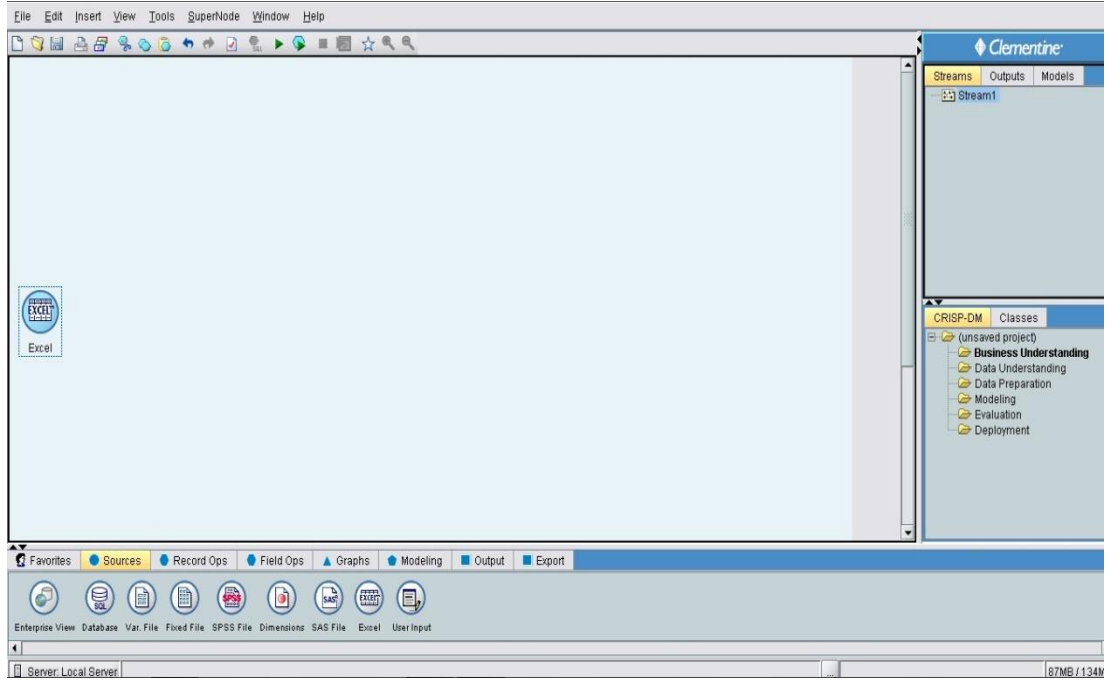


**Ekran 4.2.** SPSS Clementine 12.0 Boş Arayüzü, Ana Ekranı

#### **4.4. Paket Programın Kullanımı ve Analizin Hazırlanması**

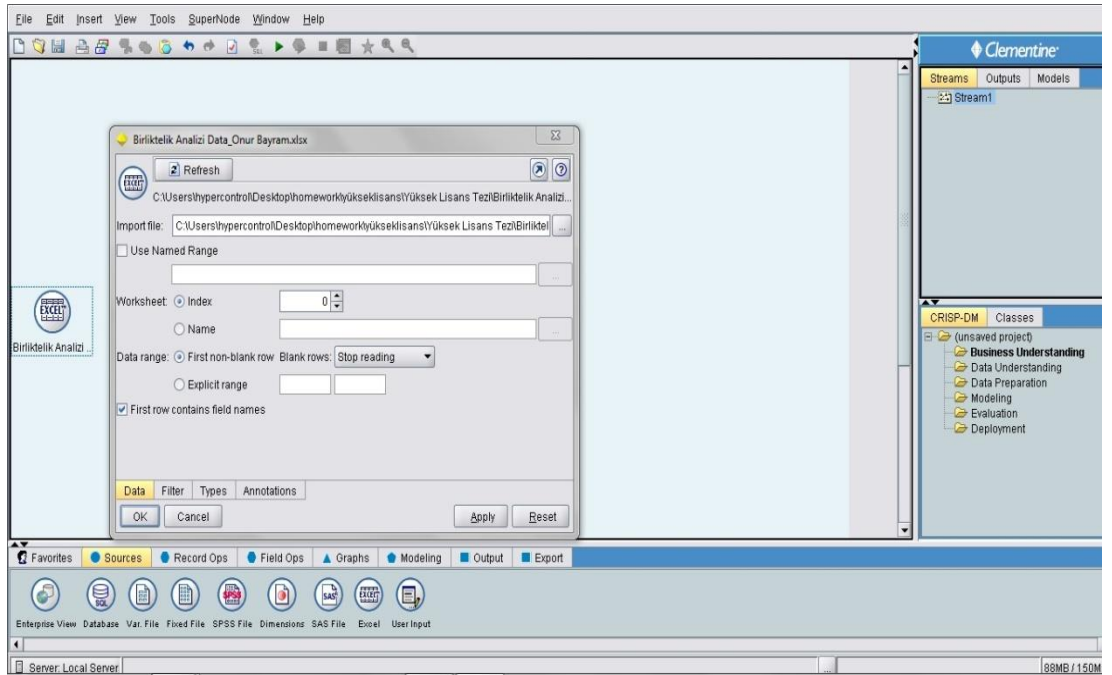
SPSS Clementine 12.0 paket programı uygun bir şekilde bilgisayara kurulduktan sonra Şekil 4.2.’de gösterildiği gibi boş bir arayüz, “Stream” ana ekranı şeklinde açılır. Burada öncelikle birliktelik analizine başlamadan önce programın nasıl kullanıldığı ve analiz için gerekli olan “Apriori” algoritması modelinin nasıl kurulduğu ekran görüntüleri yardımıyla anlatılacaktır.

İlk olarak Ekran 4.2’deki ana ekrana, elde edilen veri seti hangi formatta ise aşağıda bulunan “Sources” özelliğinden gerekli format (veri setimiz Excel olduğundan burada “Excel” formatı seçilmiştir) sürükleyip atılır. İlgili işlem aşağıdaki ekranda görülmektedir.



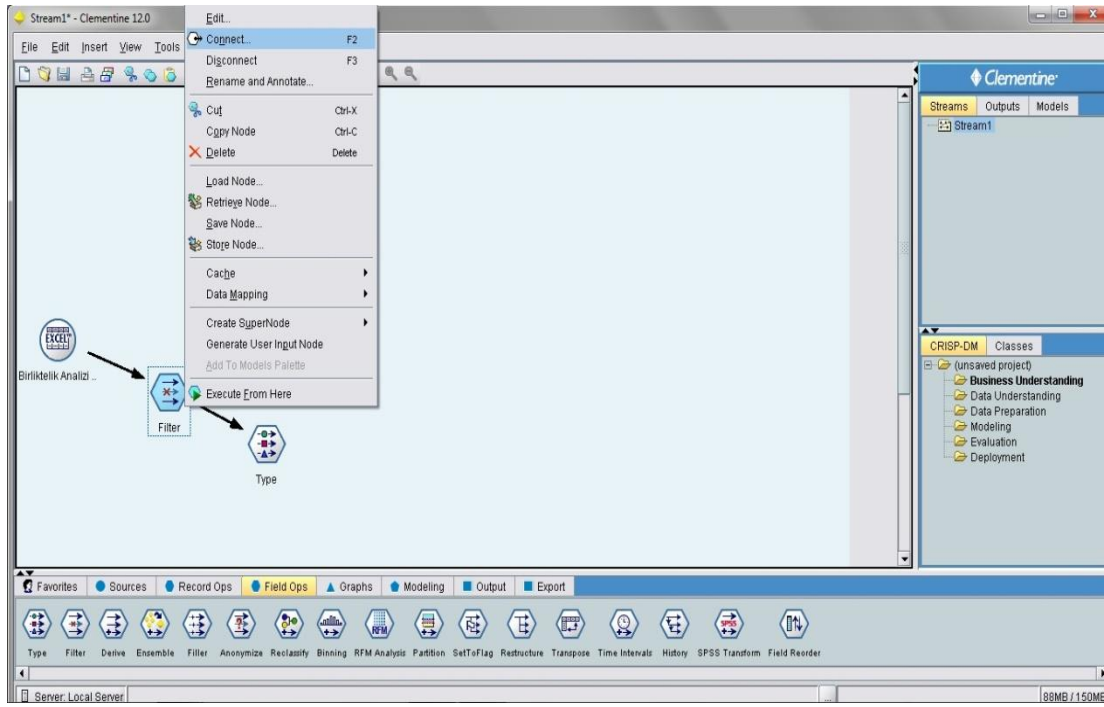
**Ekran 4.3.** SPSS Clementine 12.0 Veri Seti İçin Kaynak Formatı Seçim

Daha sonra arayüz üzerine taşınan “Excel” formatına sağ tıklanarak “Edit” kısmı açılır ve burada “Import File” kısmına kullanılacak veri seti eklenip, “OK” kısmına basılarak veri seti paket programa tanımlanır. İlgili işlem aşağıdaki ekranda görülmektedir.



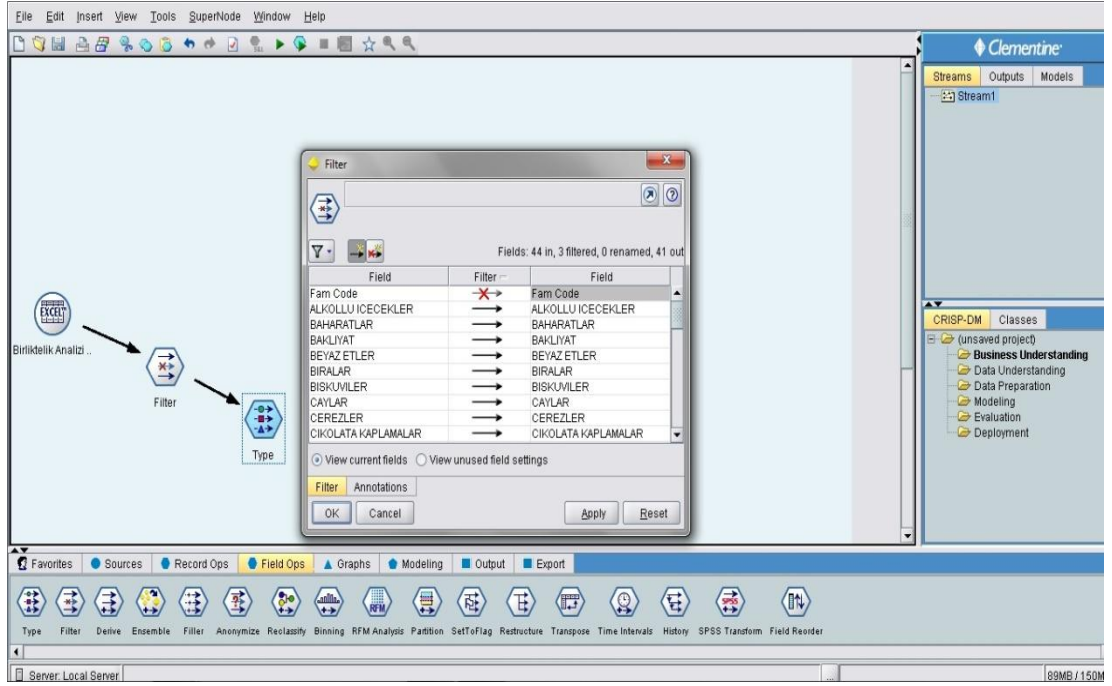
**Ekran 4.4.** SPSS Clementine 12.0 Kaynak Formata Veri Setini Tanımlama

Bundan sonraki aşamada, “Field Ops” kısmından “Filter” ve “Type” özellikleri ana ekrana taşınır. Ana ekranda bulunan bu üç özelliğin, modeli oluşturabilmek adına her birinin üzerine sağ tıklanarak, “Connect” kısmına basarak birbirleri ile bağlantı kurulması sağlanır. Sağ tıklanıp “Connect” denilen özelliği bir sonraki aşamada hangi özellik ile bağlantı kurulacaksa o özellik seçilir, bu sayede en sonuncu bağlantıda her özellik birbiri ile bağlantılı olmuş olacaktır. İlgili işlem aşağıdaki ekranda görülmektedir.



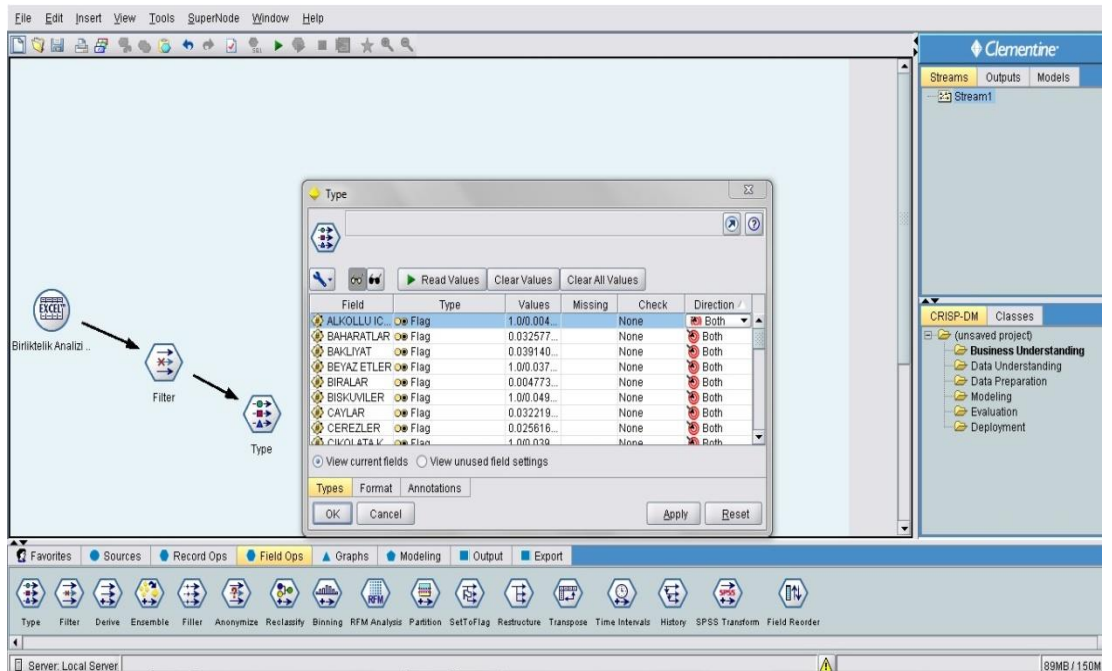
**Ekran 4.5.** SPSS Clementine 12.0 Filter ve Type Özelliklerini Ekleme ve Özellikleri Bağlantı Kurma

Bir sonraki aşamada, “Filter” özelliğinden veri seti içerisinde analizde kullanılmayacak olan veri girişlerini filtreleyerek, “OK” kısmına basılarak modelin kurulmasına devam edilir. Birliktelik kurallarının yaratılması için sadece analize girecek ürün gruplarının modelde olması gerekmektedir. Aşağıdaki ekranda bu filtreleme işlemi görülmektedir.



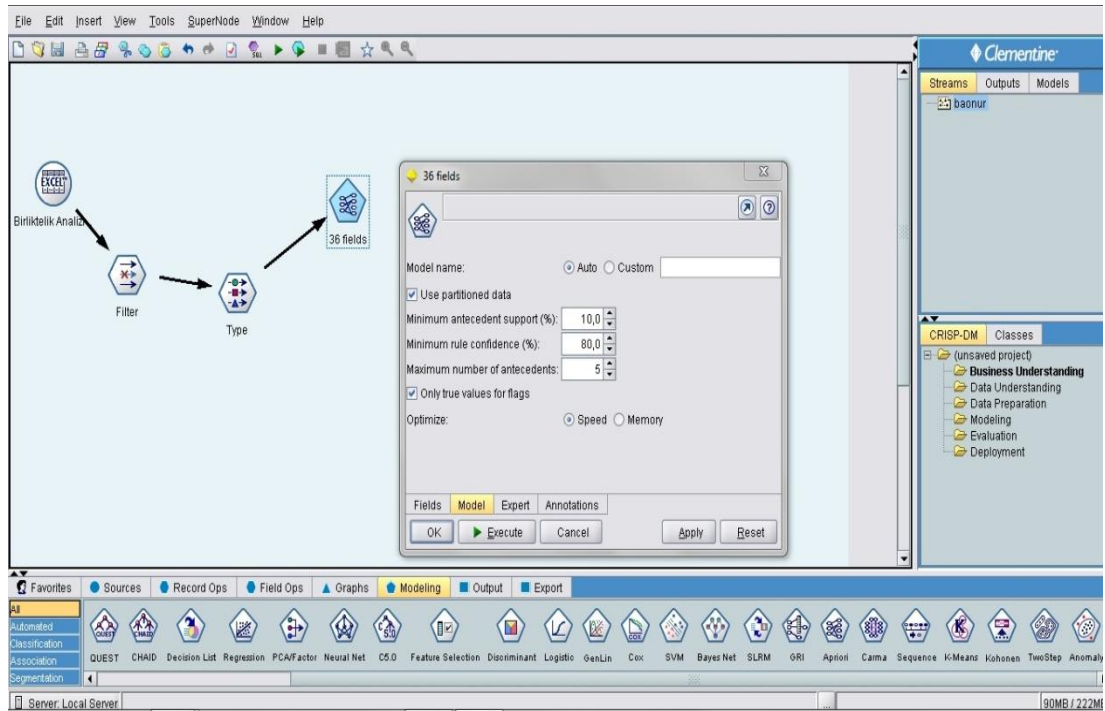
**Ekran 4.6. SPSS Clementine 12.0 Filter Özelliği**

Bundan sonraki aşamada ise, “Type” özelliği tıklanarak model için gerekli düzenlemeler yapılır. Açılan özellik ekranının tüm ürün gruplarının “type” kısımları “Flag” olarak düzenlenir, “direction” kısımları da “Both” düzenlenir ve bunlardan sonra “Values” kısmı için yukarıda bulunan “Read Values” kısmına tıklanıp verilerin okunması sağlanmalıdır, bu işlemlerden sonra “OK” kısmına basılarak model kurulmasına devam edilir. İlgili işlem aşağıdaki ekranda görülmektedir.



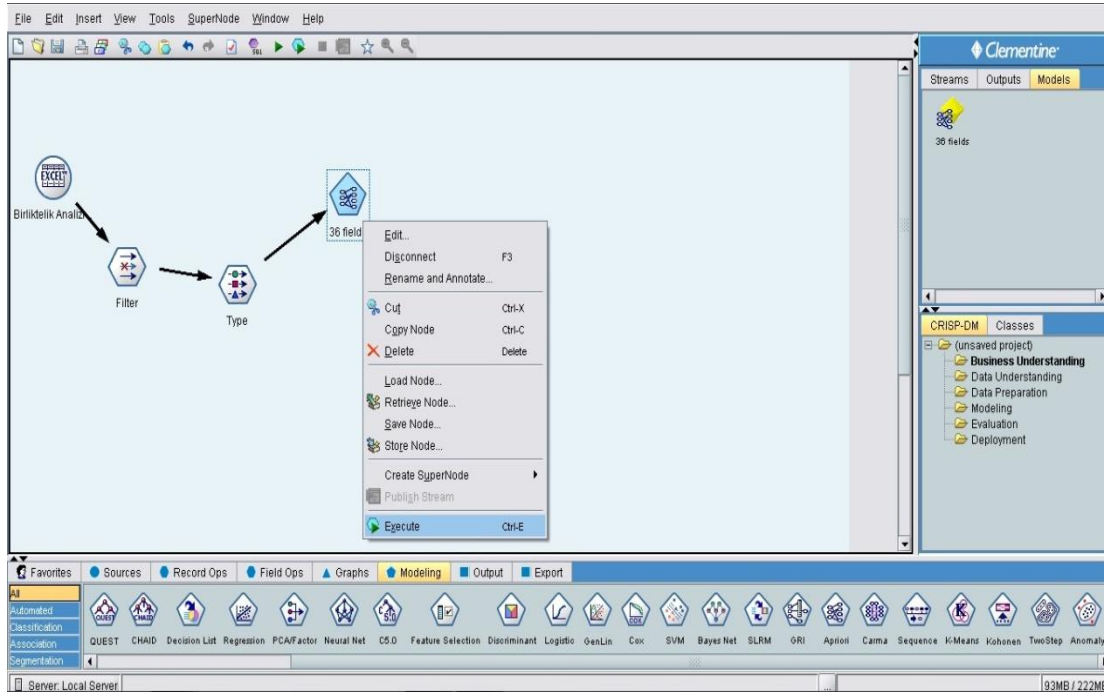
**Ekran 4.7. SPSS Clementine 12.0 Type Özelliği**

Modelin kurulması için son ve en önemli aşama buradaki aşama olacak. Öncelikle, alt kısımda bulunan “Modelling” özelliğinden “Apriori” modeli ekrana sürüklenerek, ona “Type” özelliği ile bağlantı kurulur, daha sonra “Apriori” modelinde kaç ürün grubu ile işlem yapılıyorsa, üzerinde isim olarak otomatik belirecektir. Daha sonra, “Apriori” modeline tıklanıp özellik ekranının altındaki “Model” kısmına tıklanıp oradaki belirli kısımlara minimum destek ve güven değerlerini, maksimum öncül sayısı değeri girilir. Bu değerler analizin amacı için geçerli olacak düzeylerde olmalıdır, bu değerler uygulayıcı tarafından belirlenebilir. Bu analizde minimum destek değeri %10, minimum güven değeri %80, maksimum öncül sayısı 5 olarak belirlenmiştir. “OK” kısmına basılarak model çalıştırma aşamasına gelinmiş olur. İlgili işlem aşağıdaki ekranda görülmektedir.

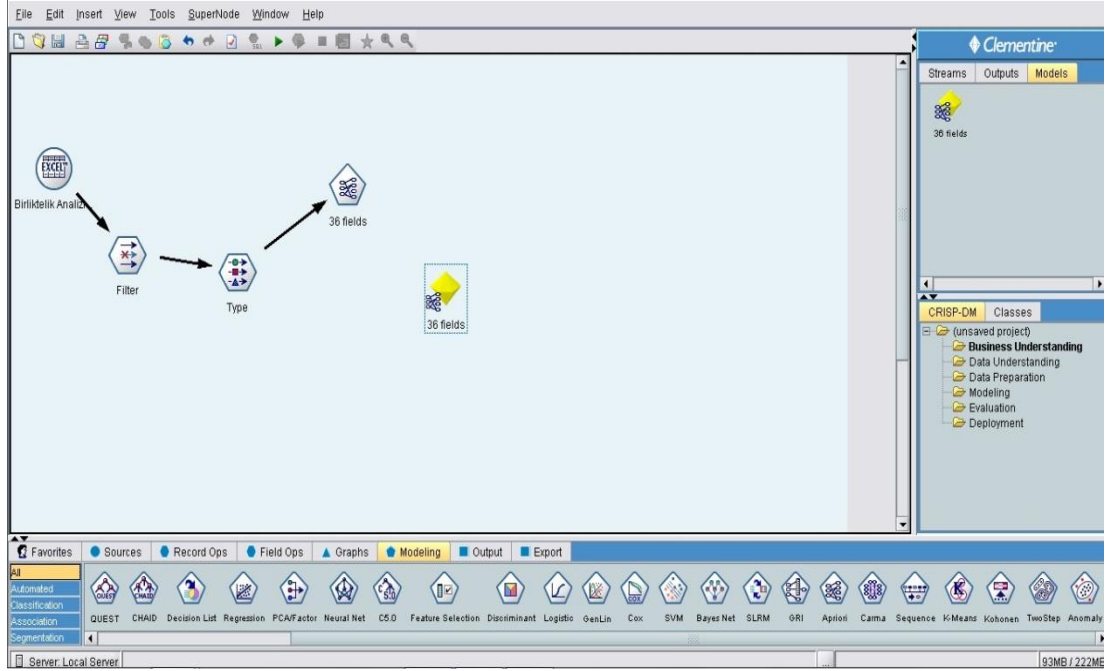


**Ekran 4.8.** SPSS Clementine 12.0 Apriori Modeli

Son olarak, Ekran 4.9’da bulunan “Apriori” modeline (burada 36 fields, 36 ürün grubu ile işlem yapıldığından) sağ tıklanarak “Execute” kısmına basılarak analiz gerçekleştirilir. Daha sonra üst kısımda bulunan “Streams” bölümünde analizin sonuçları ortaya çıkacaktır. O özellik ana ekrana sürüklenerek ve üzerine tıklanarak analizin sonuçları ayrıntılı olarak incelenebilir. İlgili işlem aşağıdaki ekranlarda görülmektedir.



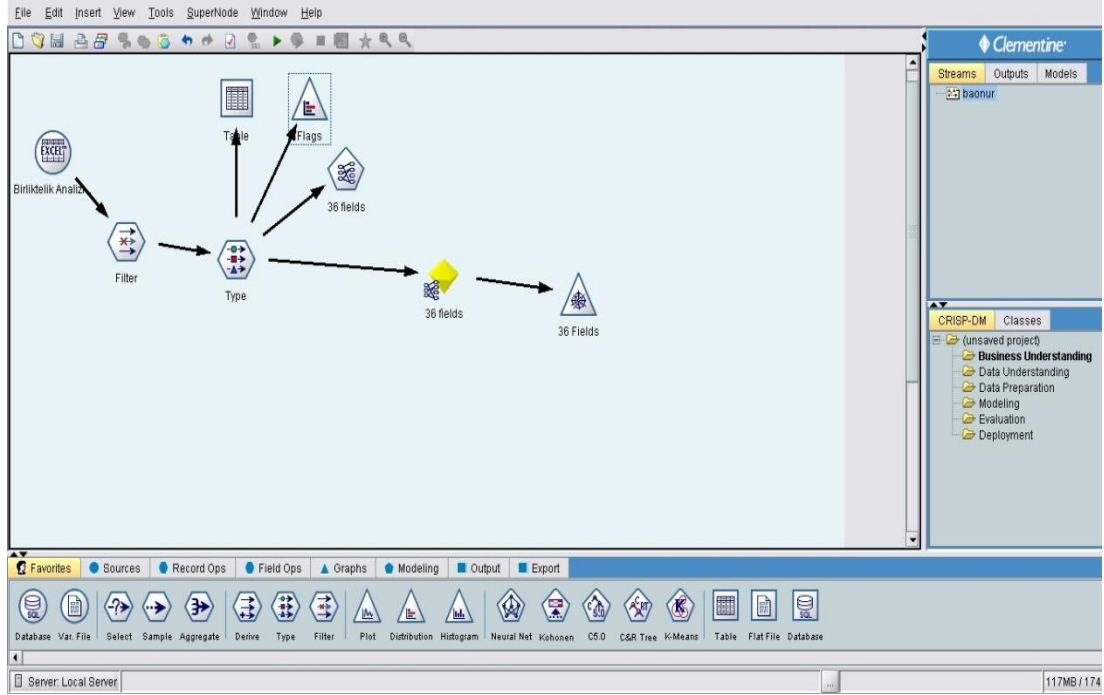
**Ekran 4.9.** SPSS Clementine 12.0 Modeli Çalıştırma



**Ekran 4.10.** SPSS Clementine 12.0 Analiz Sonuçlarının Arayüze Aktarılışı

Aşağıdaki Ekran 4.11.'de bu analizde kullanılan tüm işlem ve model özelliklerinin SPSS Clementine 12.0 paket programındaki son görüntüsü bulunmaktadır. Ekran

4.11.'de bulunan değinilmeyen özellikler, yani tablolar, web grafiği ve analiz sonuçlarının çıktıları, bulgular bölümünde ayrıntılı olarak anlatılacaktır.



**Ekran 4.11.** SPSS Clementine 12.0 Birliktelik Analizi İçin Kullanılan Özellikler ve Kurulan Modelin Son Hali

#### 4.5. Bulgular

Öncelikle Ekran 4.11.'deki “Table” özelliğine sağ tıklanıp “Execute” kısmına basıldığında veri setinin kategorik olarak görüntüsü elde edilecektir. Analizde kullanılan 36 ürün grubu, 0-1 şeklinde kategorik olarak görülmektedir.

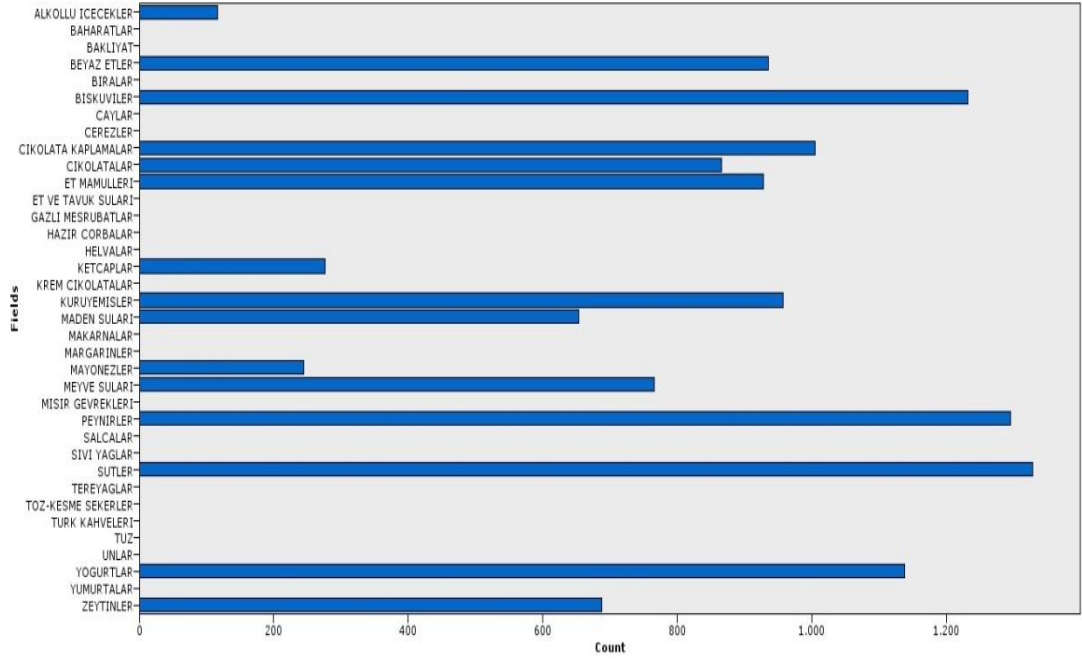
	ALKOLLÜ İÇECEKLER	BAHARATLAR	BAKLIYAT	BEYAZ ETLER	BİRALAR	BİSKUVİLER	ÇAYLAR	CEREZLER	ÇİKOLATA KAPLAMALARI	ÇİKOLATALAR	ET MAMULLERİ	ET VE TAVUK SULARI	GAZLI MESRUBATLAR	HAZIR COR
1	1.000	0.000	0.000	1.000	0.000	1.000	0.000	0.000	1.000	1.000	1.000	0.000	0.000	0.000
2	0.000	0.000	1.000	1.000	0.000	1.000	1.000	0.000	0.000	0.000	1.000	0.000	1.000	1.000
3	0.000	1.000	1.000	0.000	0.000	1.000	0.000	1.000	1.000	1.000	0.000	0.000	0.000	0.000
4	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	1.000	0.000	1.000	0.000	1.000	1.000
5	0.000	0.000	1.000	0.000	0.000	1.000	0.000	1.000	1.000	1.000	1.000	0.000	1.000	1.000
6	0.000	1.000	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000	1.000	1.000	0.000	0.000
7	0.000	0.000	1.000	1.000	0.000	1.000	0.000	1.000	1.000	0.000	1.000	0.000	1.000	1.000
8	0.000	1.000	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	1.000	0.000	1.000	1.000	1.000	1.000	0.000	1.000	0.000	1.000	1.000
10	0.000	0.000	0.000	1.000	0.000	1.000	0.000	0.000	0.000	1.000	1.000	0.000	0.000	0.000
11	0.000	1.000	1.000	1.000	0.000	1.000	0.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000
12	0.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	1.000	0.000	1.000	0.000	1.000	1.000
13	0.000	0.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	0.000	0.000	1.000
14	0.000	1.000	1.000	1.000	0.000	1.000	0.000	0.000	1.000	1.000	0.000	0.000	0.000	0.000
15	0.000	1.000	1.000	1.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
16	0.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	1.000	0.000	1.000	0.000	1.000	1.000
17	0.000	0.000	1.000	1.000	0.000	1.000	0.000	0.000	1.000	0.000	1.000	0.000	1.000	1.000
18	0.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000
19	0.000	1.000	0.000	1.000	0.000	1.000	0.000	0.000	1.000	1.000	1.000	0.000	0.000	0.000
20	0.000	1.000	1.000	0.000	0.000	1.000	1.000	1.000	0.000	1.000	1.000	0.000	0.000	0.000
21	0.000	1.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
22	0.000	0.000	1.000	0.000	0.000	1.000	0.000	1.000	1.000	0.000	1.000	0.000	1.000	1.000
23	0.000	1.000	0.000	1.000	0.000	1.000	0.000	0.000	1.000	0.000	1.000	0.000	1.000	1.000
24	0.000	1.000	0.000	1.000	0.000	1.000	1.000	1.000	0.000	1.000	1.000	0.000	0.000	1.000
25	1.000	0.000	1.000	1.000	0.000	1.000	0.000	0.000	1.000	1.000	0.000	0.000	0.000	0.000
26	0.000	1.000	1.000	1.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	1.000	1.000
27	0.000	0.000	1.000	1.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000
28	0.000	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000
29	0.000	0.000	1.000	1.000	0.000	1.000	0.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000
30	0.000	1.000	1.000	0.000	0.000	1.000	0.000	1.000	1.000	1.000	1.000	0.000	1.000	1.000
31	0.000	0.000	1.000	0.000	1.000	1.000	1.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
32	0.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
33	0.000	0.000	1.000	0.000	1.000	1.000	1.000	1.000	0.000	1.000	0.000	0.000	0.000	1.000
34	0.000	1.000	1.000	1.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000
35	0.000	1.000	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000	0.000	0.000
36	0.000	1.000	1.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	1.000	0.000	0.000	1.000
37	0.000	0.000	0.000	1.000	0.000	1.000	0.000	0.000	1.000	0.000	1.000	0.000	0.000	0.000

**Ekran 4.12. SPSS Clementine 12.0 Table Özelliği Kullanılarak Veri Setinin Kategorik Görüntüsü**

Ürün grubu kategorilerinin dağılımları, satın alınma oranları ve satın alınma frekansları ise, paket programın arayüzünde bulunan “Distribution” özelliği (burada veriler tanımladığı için Flags yazıyor) çalıştırılarak elde edilir. Ekran 4.11.’deki “Flags” özelliğine sağ tıklanıp “Execute” kısmına basıldığında veri setindeki ürün grubu kategorilerinin dağılımları elde edilecektir. Aşağıda bu dağılımlara ilişkin tablo ve şekil verilmiştir.

**Tablo 4.3. Ürün Gruplarının Alışveriş Hareketlerindeki Dağılımı**

Field /	Proportion True	%	Count
ALKOLLU İÇECEKLER		8,29	116
BAHARATLAR		0,07	1
BAKLİYAT		0,07	1
BEYAZ ETLER		66,9	936
BİRALAR		0,07	1
BİSKÜVİLER		88,06	1232
CAYLAR		0,07	1
CEREZLER		0,07	1
ÇİKOLATA KAPLAMALAR		71,84	1005
ÇİKOLATALAR		61,9	866
ET MAMULLERİ		66,33	928
ET VE TAVUK SULARI		0,07	1
GAZLI MESRUBATLAR		0,07	1
HAZIR CORBALAR		0,07	1
HELVALAR		0,07	1
KETÇAPLAR		19,9	277
KREM ÇİKOLATALAR		0,07	1
KURUYEMİŞLER		68,48	959
MADEN SULARI		46,75	654
MAKARNALAR		0,07	1
MARGARİNLER		0,07	1
MAYONEZLER		17,51	245
MEYVE SULARI		54,75	766
MISIR GEVREKLERİ		0,07	1
PEYNİRLER		92,64	1296
SALÇALAR		0,07	1
SIVI YAĞLAR		0,07	1
SÜTLER		95,0	1329
TEREYAĞLAR		0,07	1
TOZ-KESME SEKERLER		0,07	1
TURK KAHVELERİ		0,07	1
TUZ		0,07	1
UNLAR		0,07	1
YOĞURTLAR		81,34	1138
YUMURTALAR		0,07	1
ZEYTİNLER		49,18	688



**Şekil 4.1. Ürün Gruplarının Alışveriş Hareketlerindeki Dağılımının Grafiği**

Tablo 4.3. incelendiğinde en sık tekrarlanan ürün grupları Sütler, Peynirler, Bisküviler ve Yoğurtlar olarak gözlemlenmektedir. Ayrıca % 95 oranı ile Sütler ürün grubundan alışveriş yapıldığı söylenebilir. Yine benzer şekilde, % 92,64 oranı ile Peynirler, oranı ile % 88,06 oranı ile Bisküviler, % 81,34 oranı ile Yoğurtlar ürün gruplarından alışveriş yapıldığı söylenebilir.

Bu veri seti ve SPSS Clementine 12.0 paket programı ile gerçekleştirilen birliktelik analizinde, model çözümlemesinde 6384 birliktelik kuralı elde edilmiştir. Bu birliktelik kuralları “Consequent” (Takip Eden) ve “Antecedent” (Önce Gelen) şeklinde belirtilmiştir. Yani, Takip Eden ürün grupları ile Önce Gelen ürün grupları arasındaki birliktelik kuralları ortaya çıkarılmıştır.

Consequent	Antecedent	Support %	Confidence %
PEYNIRLER	SUTLER	94,996	93,228
BISKUVILER	SUTLER	94,996	89,24
YOGURLAR	SUTLER	94,996	82,092
SUTLER	PEYNIRLER	92,638	95,602
BISKUVILER	PEYNIRLER	92,638	88,272
YOGURLAR	PEYNIRLER	92,638	82,639
BISKUVILER	PEYNIRLER SUTLER	88,563	89,023
YOGURLAR	PEYNIRLER SUTLER	88,563	82,97
SUTLER	BISKUVILER	88,063	96,266
PEYNIRLER	BISKUVILER	88,063	92,857
YOGURLAR	BISKUVILER	88,063	82,792
PEYNIRLER	BISKUVILER SUTLER	84,775	93,002
YOGURLAR	BISKUVILER SUTLER	84,775	82,968
SUTLER	BISKUVILER PEYNIRLER	81,773	96,416
YOGURLAR	BISKUVILER PEYNIRLER	81,773	83,741
SUTLER	YOGURLAR	81,344	95,87
PEYNIRLER	YOGURLAR	81,344	94,112
BISKUVILER	YOGURLAR	81,344	89,631
YOGURLAR	BISKUVILER PEYNIRLER SUTLER	78,842	83,862
PEYNIRLER	YOGURLAR	77,884	84,885

**Ekran 4.13.** Ürün Grupları İçin Birliktelik Kuralları (Destek Sıralı)

36 fields

File Generate

Sort by: Confidence % 6384 of 6384

Consequent	Antecedent	Support %	Confidence %
SUTLER	KETCAPLAR MEYVE SULARI PEYNIRLER	10,936	100,0
SUTLER	KETCAPLAR MEYVE SULARI BISKUVILER PEYNIRLER	10,15	100,0
SUTLER	KETCAPLAR CIKOLATALAR YOGURTLAR BISKUVILER	10,15	100,0
SUTLER	MADEN SULARI ZEYTINLER YOGURTLAR BISKUVILER PEYNIRLER	18,013	99,603
SUTLER	ZEYTINLER MEYVE SULARI BEYAZ ETLER CIKOLATA KAPLAMA... BISKUVILER	17,155	99,583
SUTLER	ZEYTINLER MEYVE SULARI CIKOLATALAR YOGURTLAR BISKUVILER	17,012	99,58
SUTLER	ZEYTINLER MEYVE SULARI	10,000	99,570

Model Settings Summary Annotations

OK Cancel Apply Reset

**Ekran 4.14.** Ürün Grupları İçin Birliktelik Kuralları ( Güven Sıralı)

Destek sıralı ürün grupları için ortaya çıkarılan birliktelik kurallarından bazıları aşağıdaki gibi yorumlanabilir.

*Peynirler* → *Sütler* ürün grupları birliktelik kuralı için;

Peynirler ve Sütler ürün gruplarının toplam alışveriş hareketlerinde birlikte görülme olasılıkları % 94,996'dır. Ayrıca Peynirler ürün grubundan alan müşterilerin % 93,228 olasılıkla Sütler ürün grubundan da alışveriş yaptığı söylenebilir.

*Bisküviler* → *Peynirler* ürün gurupları birliktelik kuralı için;

Bisküviler ve Peynirler ürün gruplarının toplam alışveriş hareketlerinde birlikte görülme olasılıkları % 92,638'dir. Ayrıca Bisküviler ürün grubundan alan müşterilerin % 83,272 olasılıkla Peynirler ürün grubundan da alışveriş yaptığı söylenebilir.

*Yoğurtlar* → *Peynirler - Sütler* ürün grupları birliktelik kuralı için;

Yoğurtlar ve Peynirler – Sütler ürün gruplarının toplam alışveriş hareketlerinde birlikte görülme olasılıkları % 88,563'dür. Ayrıca Yoğurtlar ürün grubundan alan müşterilerin % 89,023 olasılıkla Peynirler – Sütler ürün gruplarından da alışveriş yaptığı söylenebilir.

*Yoğurtlar* → *Bisküviler* ürün grupları birliktelik kuralı için;

Yoğurtlar ve Bisküviler ürün gruplarının toplam alışveriş hareketlerinde birlikte görülme olasılıkları % 88,063'dür. Ayrıca Yoğurtlar ürün grubundan alan müşterilerin % 82,792 olasılıkla Bisküviler ürün grubundan da alışveriş yaptığı söylenebilir.

*Yoğurtlar* → *Bisküviler – Peynirler – Sütler* ürün grupları birliktelik kuralı için;

Yoğurtlar ve Bisküviler – Peynirler – Sütler ürün gruplarının toplam alışveriş hareketlerinde birlikte görülme olasılıkları % 78,842'dir. Ayrıca Yoğurtlar ürün grubundan alan müşterilerin % 83,062 olasılıkla Bisküviler- Peynirler – Sütler ürün gruplarından da alışveriş yaptığı söylenebilir.

*Çikolata Kaplamalar* → *Çikolatalar – Peynirler – Sütler* ürün grupları birliktelik kuralı için;

Çikolata Kaplamalar ve Çikolatalar – Peynirler – Sütler ürün gruplarının toplam alışveriş hareketlerinde birlikte görülme olasılıkları % 55,611'dir. Ayrıca Çikolata Kaplamalar ürün grubundan alan müşterilerin % 82,648 olasılıkla Çikolatalar – Peynirler – Sütler ürün gruplarından da alışveriş yaptığı söylenebilir.

*Kuruyemişler* → *Maden Suları – Çikolatalar – Beyaz Etlere – Bisküviler* ürün grupları birliktelik kuralı için;

Kuruyemişler ve Maden Suları – Çikolatalar – Beyaz Etlere – Bisküviler ürün gruplarının toplam alışveriş hareketlerinde birlikte görülme olasılıkları % 21,225'dir.

Ayrıca Kuruyemişler ürün grubundan alan müşterilerin % 81,481 olasılıkla Maden Suları – Çikolatalar – Beyaz Etler – Bisküviler ürün gruplarından da alışveriş yaptığı söylenebilir.

Güven sıralı ürün grupları için ortaya çıkarılan birliktelik kurallarından bazıları aşağıdaki gibi yorumlanabilir.

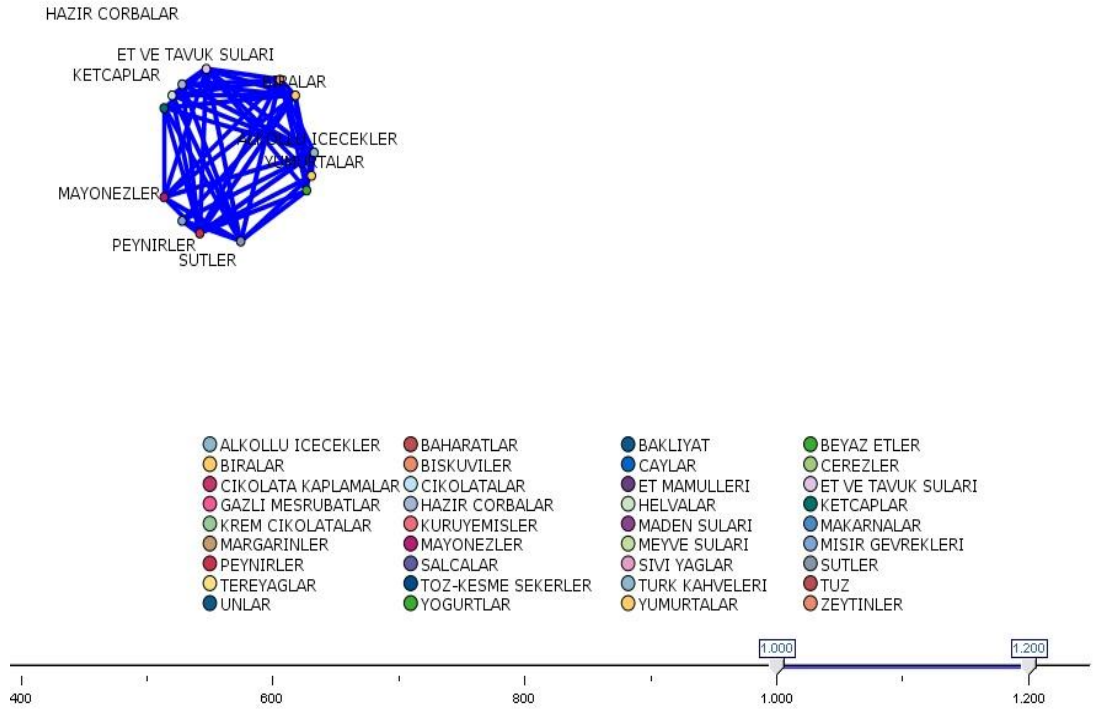
*Sütler* → *Ketçaplar* – *Meyve Suları* – *Peynirler* ürün grupları birliktelik kuralı için;  
Sütler ve Ketçaplar – Meyve Suları – Peynirler ürün gruplarının toplam alışveriş hareketlerinde birlikte görülme olasılıkları % 10,939'dür. Ayrıca Sütler ürün grubundan alan müşterilerin % 100 olasılıkla Ketçaplar – Meyve Suları – Peynirler ürün gruplarından da alışveriş yaptıkları söylenebilir.

*Et Mamulleri* → *Zeytinler* – *Meyve Suları* – *Çikolatalar* – *Kuruyemişler* ürün grupları birliktelik kuralı için;

Et Mamulleri ve Zeytinler – Meyve Suları – Çikolatalar – Kuruyemişler ürün gruplarının toplam alışveriş hareketlerinde birlikte görülme olasılıkları % 16,154'dür. Ayrıca Et Mamulleri ürün grubundan alan müşterilerin % 83,186 olasılıkla Zeytinler – Meyve Suları – Çikolatalar – Kuruyemişler ürün gruplarından da alışveriş yaptığı söylenebilir.

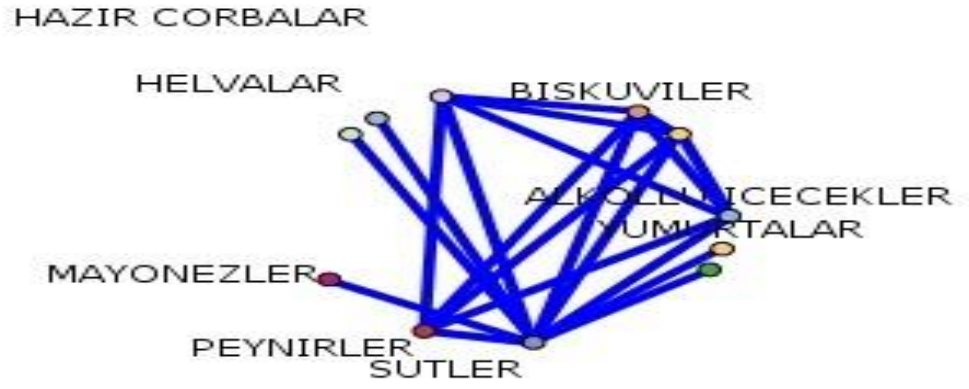
Yukarıdaki şekillerde görülen ve ortaya çıkan 6384 birliktelik kuralından geriye kalan tüm kurallar benzer şekillerde yorumlanabilir.

Birliktelik kurallarının daha net bir şekilde görsel olarak görülebilmesi için Web Grafiği yardımı ile ürün grupları kategorileri arası ilişkiler incelenmiştir. Web Grafiği özelliği, paket programın arayüzü üzerinde analiz sonuçlarına bağlantı kurarak oluşturulabilir. Bu işlem Ekran 4.11.'de görülebilir.



**Şekil 4.2.** SPSS Clementine 12.0 Ürün Gruplarının Web Grafiği

Ürün grupları kategorilerinin birbiri ile ilişkileri bu ağda çok yoğun olarak görülebilmektedir. Bu yoğunluk değeri değiştirilebilir, Şekil 4.2.'nin alt tarafında bulunan yoğunluk birimleri hareket ettirilerek, grafiğin yoğunluk değeri değiştirilip kategoriler arası ilişkiler farklı yoğunluklarda gözlemlenebilir. Aşağıda, sıklığı daha yüksek olan ilişkilerin bulunduğu iki şekil ile bu veri seti için Web Grafiği'nin önemi daha iyi anlaşılabilir.



**Şekil 4.3.** SPSS Clementine 12.0 Sıklığı Yüksek Ürün Gruplarının Web Grafiği



**Şekil 4.4.** SPSS Clementine 12.0 Sıklığı En Yüksek Olan Ürün Gruplarının Son Web Grafiği

Yukarıda görüldüğü gibi, Peynirler ve Sütler ürün grupları birlikte çok sık alınmaktadır. Bir önceki şekilde ise, Peynirler ve Sütler ürün gruplarına ek olarak sıklığı daha az olan Helvalar, Bisküviler, Yoğurtlar, Yumurtalar, Mayonezler gibi ürün grupları arasındaki ilişkiler de gözlenebilmektedir.

## BEŞİNCİ BÖLÜM

### 5. SONUÇ VE ÖNERİLER

Bu çalışmada veri madenciliğinin en bilinen uygulamalarından biri olan birliktelik analizinin açıklanması ve bir uygulaması yapılması amaçlanmıştır.

Gerçek bir veri kümesi ile SPSS Clementine 12.0 paket programı kullanılarak birliktelik kuralları uygulaması yapılmış ve birliktelik analizi gerçekleştirilmiştir. Bu analizden elde edilen bulgular açıklanmış ve yorumlanmıştır.

Bulgulardan elde ettiğimiz sonuçlara göre Sütler, Peynirler, Yoğurtlar, Bisküviler ürün gruplarının alım açısından diğer ürün gruplarına göre daha fazla olduğu öne sürülebilir. Ayrıca bu ürün grupları arasında daha yoğun bir birliktelik ilişkisi gözle çarpılmaktadır. Fakat bu sonuçlarda çeşitli kısıtlar göz önünde bulundurulmalıdır; ilk olarak bu analiz sadece araştırmanın yapıldığı marketlerde yer alan ürün grupları ile yapılmıştır, belirli bir dönemde veya farklı dönemlerde toplanan alışveriş hareketleri ile daha farklı sonuçlar elde edilebilir, yapılacak kampanyalar, marketlerin bulunduğu bölgeler ve hedef kitlelerin alacağı ürünler, hedef kitledeki insan sayısı analiz sonuçlarını doğrudan etkileyebilir.

Çalışma sonunda elde edilen bilgiler ile, daha düzenli, etkin raf düzenlemeleri, ürün kampanyaları yapılabilir, stoktaki ürün miktarlarının uygun düzeyde tutulması ve daha kolay kontrol edilebilmesi sağlanabilir.

Bu çalışmayı ileriye götürebilecek olan araştırma önerileri ise; ürün gruplarının yanında yaş, cinsiyet, gelir düzeyi gibi müşteri bilgileri de bulunan müşteri profillerini içeren bir veri seti üzerinden çalışılabilir. Ürün grupları, birliktelik kuralları ile müşteri profillerine göre gruplandırılabilir ve çeşitli reklam ve kampanya stratejileri farklı ürünlerin alımlarını ve ulaşılabilirliklerini arttırabilir.

Diğer yandan kullanılan veri setinin periyodunun mümkün olduğunca geniş tutulması, çalışmanın daha net sonuçlar vermesi açısından önemlidir. Ayrıca bu anlamda, birliktelik kurallarının mevsimsel olarak incelenmesi ve toplam olarak karşılaştırılması kolaylaşacaktır. Yine bu şekilde gerçekleştirilecek araştırmalar,

gerekli kampanyaların ve düzenlemelerin aylara hatta günlere göre yapılmasına olanak sağlayacaktır.

Son olarak, bundan sonraki yapılacak çalışmalar için veri madenciliği literatüründe bulunan istatistiksel tekniklerin bütünleştirilmesi, birliktelik analizi çalışmalarına zenginlik katabilir.

## KAYNAKLAR

**Agrawal, R., Imielinski, T., Swami, A.,** 1993, Mining Association Rules between Sets of Items in Large Databases, Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington D.C..

**Agrawal, R., Srikant, R.,** 1994, Fast Algorithms for Mining Association Rules, Proceedings of the 20 th VLDB Conference Santiago, Chile.

**Ay, D., Çil, İ.,** 2010, Migros Türk A.Ş.de Birliktelik Kurallarının Yerleşim Düzeni Planlamada Kullanılması, Sakarya Üniversitesi Endüstri Mühendisliği Bölümü, Endüstri Mühendisliği Dergisi, Cilt 21, Sayı 2, Adapazarı.

**Birant, D., Kut, A., Ventura, M., Altınok, H., Antınok, B., Altınok, E., Ihlamur, M.,** 2010, İş Zekası Çözümleri İçin Çok Boyutlu Birliktelik Kuralları Analizi, Dokuz Eylül Üniversitesi Bilgisayar Mühendisliği Bölümü, Akademik Bilişim Konferansları Bildirileri, 10-12 Şubat 2010 Muğla Üniversitesi.

**Brachman, R., Anand, T.,** 1996, The Process Of Knowledge Discovery In Databases: A Human Centered Approach. In U. Fayyad, G. Piatetsky- Shapiro, P. Smyth, and R. Uthurusamy,(Edt.), Advances in Knowledge Discovery And Data Mining (pp 37-58). Menlo Park, California: AAAI Press.

**Delibaş, E.,** 2010, Birliktelik Analizi İle Reçeteli İlaç Satışları Üzerinde Bir Uygulama, Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, Sivas.

**Dolgun, M. Ö.,** 2006, Büyük Alışveriş Merkezleri için Veri Madenciliği Uygulamaları, Yüksek Lisans Tezi, Hacettepe Üniversitesi İstatistik Anabilim Dalı, Ankara.

**Döşlü, A.,** 2008, Veri Madenciliğinde Market Sepet Analizi ve Birliktelik Kurallarının Belirlenmesi. Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmamış Yüksek Lisans Tezi.

**Erpolat, S.**, 2012, Otomobil Yetkili Servislerinde Birliktelik Kurallarının Belirlenmesinde Apriori ve FP-Growth Algoritmalarının Karşılaştırılması, Mimar Sinan Güzel Sanatlar Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü, Anadolu Üniversitesi Sosyal Bilimler Dergisi.

**Gürgen, G.**, 2008, Birliktelik Kuralları Sepet Analizi ve Uygulaması. Marmara Üniversitesi Sosyal Bilimler Enstitüsü Yayınlanmamış Yüksek Lisans Tezi.

**Güven, A., Bozkurt, Ö., Kalıpsız, O.**, 2007, Veri Madenciliğinin Geleceği, Yıldız Teknik Üniversitesi, Bilgisayar Mühendisliği Bölümü, Üniversite Dergisi İçin Makale, Yıldız, İstanbul.

**Han, J., Kamber, M.**, 2001, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, San Francisco.

**Han, J., Kamber, M.**, 2006, Data Mining Concepts And Tecniques Second Edition. San Francisco:Morgan Kaufman.

**Hegland, M.**, 2001, Data mining techniques, Acta Numerica.

**Hidber, C.**, 1999, Online association rule mining, proceedings ACM SIGMOD International Conference on Management of Data, Philadelphia, Pennsylvania.

**Kalkov, A.**, 2006, Veri Madenciliği ve Bir E-Ticaret Uygulaması, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Ankara.

**Karabatak, M.**, 2008, Özellik Seçimi, Sınıflama ve Öngörü Uygulamalarına Yönelik Birliktelik Kuralı Çıkarımı ve Yazılım Geliştirilmesi, Fırat Üniversitesi Fen Bilimleri Enstitüsü, Doktora Tezi, Elazığ.

**Lee, S. J., Siau, K.**, 2001, A Review Of Data Mining Tecniques. Industrial Management & Data Systems. V:101.

**M.H. Dunham, Y. Xiao, L. Gruenwald, Z.Hossain**, “A Survey Of Association Rules”, Southern Methodist University, University of Oklahoma, 2001.

**Manila, H., Toivonen, H., Verkamo, A.I.**, 1994, Efficient algorithms for discovering association rules. In Proceedings of AAAI'94 Workshop on Knowledge Discovery in Databases (KDD'94), Seattle, Washington, USA.

**Silahtaroglu, G.**, “Kavram ve Algoritmalarıyla Temel Veri Madenciliği”, 1. Basım, İstanbul, Papatya Yayıncılık Eğitim, 2008.

**Şen, F.**, 2008, Veri Madenciliği İle Birlikte Kurallarının Bulunması. Sakarya Üniversitesi Fen Bilimleri Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi.

**Timor, M., Ezerçe, A., Güsoy, T.**, 2011, Müşteri Profili ve Alışveriş Davranışlarını Belirlemede Kümeleme ve Birlikte Kuralları Analizi: Perakende Sektöründe Bir Uygulama, İstanbul Üniversitesi, İşletme Fakültesi, Sayısal Yöntemler Anabilim Dalı, Üniversite Dergisi için Makale, İstanbul.


**Toivonen, H.**, 1996, Sampling large databases for association rules, 22. International Conference on Very Large Databases, Mumbai, Hindistan.

**URL - 1**, <http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

**URL - 2**, [www.sbd.anadolu.edu.tr](http://www.sbd.anadolu.edu.tr)

## EKLER

### Ek-1: Ürün Grupları İçin Diğer Birliktelik Kurallarından Bazıları (Destek Sıralı)



36 fields

File Generate

Sort by: Support % 6384 of 6384

Consequent	Antecedent	Support %	Confidence %
PEYNIRLER	KURUYEMISLER	68,477	93,946
BISKUVILER	KURUYEMISLER	68,477	91,023
YOGURTLAR	KURUYEMISLER	68,477	83,507
SUTLER	CIKOLATA KAPLAMA... BISKUVILER	67,191	96,809
PEYNIRLER	CIKOLATA KAPLAMA... BISKUVILER	67,191	93,085
YOGURTLAR	CIKOLATA KAPLAMA... BISKUVILER	67,191	83,617
SUTLER	BEYAZ ETLER	66,905	96,474
PEYNIRLER	BEYAZ ETLER	66,905	95,192
BISKUVILER	BEYAZ ETLER	66,905	90,385
YOGURTLAR	BEYAZ ETLER	66,905	84,829
SUTLER	CIKOLATA KAPLAMA... PEYNIRLER	66,69	96,356
BISKUVILER	CIKOLATA KAPLAMA... PEYNIRLER	66,69	93,783
YOGURTLAR	CIKOLATA KAPLAMA... PEYNIRLER	66,69	84,352
SUTLER	ET MAMULLERI	66,333	96,013
PEYNIRLER	ET MAMULLERI	66,333	95,151
BISKUVILER	ET MAMULLERI	66,333	90,841
YOGURTLAR	ET MAMULLERI	66,333	84,591
PEYNIRLER	KURUYEMISLER SUTLER	65,833	94,245
BISKUVILER	KURUYEMISLER SUTLER	65,833	91,422

Model Settings Summary Annotations

OK Cancel Apply Reset

**Ek-1 (Devamı): Ürün Grupları İçin Diğer Birliktelik Kurallarından Bazıları  
(Destek Sıralı)**

Consequent	Antecedent	Support %	Confidence %
SUTLER	CIKOLATALAR BISKUVILER PEYNIRLER	54,325	96,316
BISKUVILER	CIKOLATA KAPLAMA... YOGURLAR PEYNIRLER SUTLER	54,325	94,737
BISKUVILER	BEYAZ ETLER YOGURLAR PEYNIRLER	54,325	90,921
YOGURLAR	CIKOLATALAR BISKUVILER PEYNIRLER	54,325	85,395
CIKOLATA KAPLAMA...	CIKOLATALAR BISKUVILER PEYNIRLER	54,325	84,342
SUTLER	KURUYEMISLER YOGURLAR PEYNIRLER	54,182	96,966
BISKUVILER	KURUYEMISLER YOGURLAR PEYNIRLER	54,182	92,084
PEYNIRLER	ET MAMULLERI YOGURLAR SUTLER	54,11	96,037
BISKUVILER	ET MAMULLERI YOGURLAR SUTLER	54,11	92,602

**Ek-1 (Devamı): Ürün Grupları İçin Diğer Birliktelik Kurallarından Bazıları  
(Destek Sıralı)**

36 fields

File Generate

Sort by: Support % 6384 of 6384

Consequent	Antecedent	Support %	Confidence %
CIKOLATA KAPLAMA...	MEYVE SULARI BISKUVILER	50,465	82,011
PEYNIRLER	BEYAZ ETLER YOGURTLAR BISKUVILER SUTLER	50,25	95,59
SUTLER	BEYAZ ETLER CIKOLATA KAPLAMA...	50,179	97,009
PEYNIRLER	BEYAZ ETLER CIKOLATA KAPLAMA...	50,179	95,157
BISKUVILER	BEYAZ ETLER CIKOLATA KAPLAMA...	50,179	94,16
YOGURTLAR	BEYAZ ETLER CIKOLATA KAPLAMA...	50,179	85,613
PEYNIRLER	ET MAMULLERI YOGURTLAR BISKUVILER SUTLER	50,107	95,72
SUTLER	ET MAMULLERI CIKOLATA KAPLAMA...	49,964	96,423
PEYNIRLER	ET MAMULLERI CIKOLATA KAPLAMA...	49,964	94,993
BISKUVILER	ET MAMULLERI CIKOLATA KAPLAMA...	49,964	94,564
YOGURTLAR	ET MAMULLERI CIKOLATA KAPLAMA...	49,964	85,837
SUTLER	KURUYEMISLER YOGURTLAR		

Model Settings Summary Annotations

OK Cancel Apply Reset

**Ek-1 (Devamı): Ürün Grupları İçin Diğer Birliktelik Kurallarından Bazıları  
(Destek Sıralı)**

36 fields

File Generate

Sort by: Support % 6384 of 6384

Consequent	Antecedent	Support %	Confidence %
BISKUVILER	KURUYEMISLER CIKOLATA KAPLAMA... YOGURLAR PEYNIRLER	41,887	95,222
BISKUVILER	MEYVE SULARI CIKOLATA KAPLAMA... SUTLER	41,816	96,923
PEYNIRLER	BEYAZ ETLER CIKOLATA KAPLAMA... YOGURLAR SUTLER	41,816	96,068
BISKUVILER	BEYAZ ETLER CIKOLATA KAPLAMA... YOGURLAR SUTLER	41,816	95,385
PEYNIRLER	MEYVE SULARI CIKOLATA KAPLAMA... SUTLER	41,816	94,359
YOGURLAR	MEYVE SULARI CIKOLATA KAPLAMA... SUTLER	41,816	85,641
SUTLER	ET MAMULLERI BEYAZ ETLER YOGURLAR	41,744	97,603
SUTLER	ET MAMULLERI KURUYEMISLER YOGURLAR	41,744	97,089
SUTLER	ZEYTINLER	41,744	96,747

Model Settings Summary Annotations

OK Cancel Apply Reset

**Ek-1 (Devamı): Ürün Grupları İçin Diğer Birliktelik Kurallarından Bazıları  
(Destek Sıralı)**

36 fields

File Generate

Sort by: Support % 6384 of 6384

Consequent	Antecedent	Support %	Confidence %
YOGURTLAR	MADEN SULARI ZEYTINLER MEYVE SULARI ET MAMULLERI	12,009	91,667
CIKOLATA KAPLAMA...	MADEN SULARI ZEYTINLER ET MAMULLERI BEYAZ ETLER KURUYEMISLER	12,009	85,119
KURUYEMISLER	MADEN SULARI ZEYTINLER MEYVE SULARI ET MAMULLERI	12,009	83,929
ET MAMULLERI	MADEN SULARI ZEYTINLER MEYVE SULARI KURUYEMISLER	12,009	83,929
CIKOLATA KAPLAMA...	MADEN SULARI ZEYTINLER MEYVE SULARI KURUYEMISLER	12,009	83,929
BEYAZ ETLER	MADEN SULARI ZEYTINLER MEYVE SULARI ET MAMULLERI	12,009	82,143
BEYAZ ETLER	MADEN SULARI ZEYTINLER MEYVE SULARI	12,009	81,548

Model Settings Summary Annotations

OK Cancel Apply Reset

**Ek-2: Ürün Grupları İçin Diğer Birliktelik Kurallarından Bazıları (Güven Sıralı)**

36 fields

File Generate

Sort by: Confidence % 6384 of 6384

Consequent	Antecedent	Support %	Confidence %
SUTLER	CIKOLATALAR BEYAZ ETLER KURUYEMISLER YOGURLAR PEYNIRLER	27,734	97,165
SUTLER	BEYAZ ETLER BISKUVILER	60,472	97,163
SUTLER	ZEYTINLER ET MAMULLERI YOGURLAR PEYNIRLER	30,236	97,163
PEYNIRLER	MADEN SULARI ZEYTINLER MEYVE SULARI KURUYEMISLER CIKOLATA KAPLAMA...	10,079	97,163
SUTLER	CIKOLATALAR KURUYEMISLER YOGURLAR BISKUVILER PEYNIRLER	35,239	97,16
BISKUVILER	MEYVE SULARI KURUYEMISLER CIKOLATA KAPLAMA... PEYNIRLER	30,164	97,156
PEYNIRLER	CIKOLATALAR BEYAZ ETLER KURUYEMISLER	30,164	97,156

Model Settings Summary Annotations

OK Cancel Apply Reset

**Ek-2 (Devamı): Ürün Grupları İçin Diğer Birliktelik Kurallarından Bazıları  
(Güven Sıralı)**

36 fields

File Generate

Sort by: Confidence % 6384 of 6384

Consequent	Antecedent	Support %	Confidence %
PEYNIRLER	ET MAMULLERI KURUYEMISLER BISKUVILER	44,603	95,673
BISKUVILER	MADEN SULARI MEYVE SULARI BEYAZ ETLER KURUYEMISLER YOGURLAR	14,868	95,673
PEYNIRLER	MADEN SULARI ET MAMULLERI	33,024	95,671
BISKUVILER	ZEYTINLER MEYVE SULARI CIKOLATALAR	19,8	95,668
PEYNIRLER	MADEN SULARI MEYVE SULARI CIKOLATA KAPLAMA... YOGURLAR BISKUVILER	19,8	95,668
BISKUVILER	MADEN SULARI CIKOLATALAR ET MAMULLERI SUTLER	23,088	95,666
PEYNIRLER	CIKOLATALAR BEYAZ ETLER CIKOLATA KAPLAMA...	36,24	95,661
PEYNIRLER	MADEN SULARI BEYAZ ETLER	32,881	95,652

Model Settings Summary Annotations

OK Cancel Apply Reset

**Ek-2 (Devamı): Ürün Grupları İçin Diğer Birliktelik Kurallarından Bazıları  
(Güven Sıralı)**

36 fields

File Generate

Sort by: Confidence % 6384 of 6384

Consequent	Antecedent	Support %	Confidence %
YOGURTLAR	ZEYTINLER MEYVE SULARI CIKOLATALAR PEYNIRLER	19,085	90,262
YOGURTLAR	MADEN SULARI ET MAMULLERI KURUYEMISLER CIKOLATA KAPLAMA... BISKUVILER	19,8	90,253
BISKUVILER	BEYAZ ETLER PEYNIRLER	63,688	90,236
YOGURTLAR	MADEN SULARI CIKOLATALAR ET MAMULLERI KURUYEMISLER PEYNIRLER	18,299	90,234
YOGURTLAR	MADEN SULARI ET MAMULLERI KURUYEMISLER BISKUVILER SUTLER	23,374	90,214
BISKUVILER	MAYONEZLER BEYAZ ETLER YOGURTLAR SUTLER	10,222	90,21
YOGURTLAR	MADEN SULARI ZEYTINLER ET MAMULLERI	13,867	90,206

Model Settings Summary Annotations

OK Cancel Apply Reset

**Ek-2 (Devamı): Ürün Grupları İçin Diğer Birliktelik Kurallarından Bazıları  
(Güven Sıralı)**

36 fields

File Generate

Sort by: Confidence % 6384 of 6384

Consequent	Antecedent	Support %	Confidence %
YOGURTLAR	CIKOLATA KAPLAMA... BISKUVILER PEYNIRLER	62,545	84,8
YOGURTLAR	KURUYEMISLER BISKUVILER SUTLER	60,186	84,798
YOGURTLAR	KETCAPLAR CIKOLATALAR PEYNIRLER SUTLER	12,223	84,795
KURUYEMISLER	MADEN SULARI ZEYTINLER CIKOLATALAR PEYNIRLER	15,511	84,793
CIKOLATA KAPLAMA...	CIKOLATALAR BEYAZ ETLER PEYNIRLER	40,886	84,79
CIKOLATA KAPLAMA...	ZEYTINLER CIKOLATALAR ET MAMULLERI SUTLER	23,016	84,783
YOGURTLAR	MADEN SULARI PEYNIRLER	43,674	84,779
BEYAZ ETLER	ZEYTINLER MEYVE SULARI CIKOLATALAR ET MAMULLERI CIKOLATA KAPLAMA...	14,081	84,772

Model Settings Summary Annotations

OK Cancel Apply Reset

**Ek-2 (Devamı): Ürün Grupları İçin Diğer Birliktelik Kurallarından Bazıları  
(Güven Sıralı)**

36 fields

File Generate

Sort by: Confidence % 6384 of 6384

Consequent	Antecedent	Support %	Confidence %
CIKOLATA KAPLAMA...	MADEN SULARI ET MAMULLERI YOGURTLAR	28,863	80,05
KURUYEMISLER	ZEYTINLER CIKOLATALAR BISKUVILER	29,021	80,049
BEYAZ ETLER	ZEYTINLER ET MAMULLERI YOGURTLAR PEYNIRLER SUTLER	29,378	80,049
CIKOLATALAR	MEYVE SULARI CIKOLATA KAPLAMA... YOGURTLAR BISKUVILER SUTLER	34,739	80,041
CIKOLATA KAPLAMA...	MEYVE SULARI KURUYEMISLER	39,385	80,036
KURUYEMISLER	ZEYTINLER ET MAMULLERI YOGURTLAR SUTLER	30,021	80,0
KURUYEMISLER	ZEYTINLER BEYAZ ETLER CIKOLATA KAPLAMA... PEYNIRLER SUTLER	26,447	80,0
CIKOLATA KAPLAMA...	ZEYTINLER		

Model Settings Summary Annotations

OK Cancel Apply Reset

## ÖZGEÇMİŞ

### KİŞİSEL BİLGİLER

**Adı Soyadı** :Onur BAYRAM  
**Doğum Tarihi ve Yeri** :04.10.1988 / BURSA  
**Yabancı Dili** :İngilizce, Almanca  
**E-Posta** :onurbayram3007@gmail.com  
**Telefon** :0506 785 74 93

### ÖĞRENİM DURUMU

Derece	Alan	Okul / Üniversite	Mezuniyet
Yüksek Lisans	İstatistik	Mimar Sinan Güzel Sanatlar Üniversitesi	2014
Lisans	İstatistik	Eskişehir Osmangazi Üniversitesi	2011
Lise	Fen Bilimleri	Yalova Fen Lisesi	2006