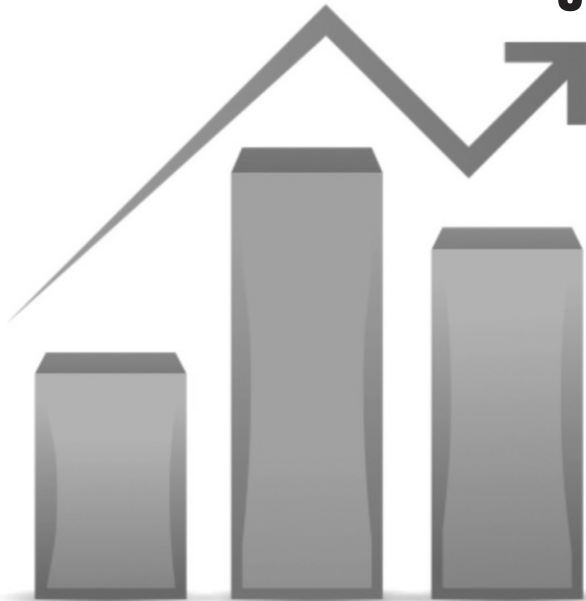


UYGULAMALI VERİ MADENCİLİĞİ SEKTÖREL ANALİZLER

Umman Tuğba Şimşek Gürsoy

3. Baskı





Yrd. Doç. Dr. Umman Tuğba Şimşek Gürsoy

UYGULAMALI VERİ MADENCİLİĞİ SEKTÖREL ANALİZLER

ISBN 978-605-364-040-0

Kitapta yer alan bölümlerin tüm sorumluluğu yazarına aittir.

© 2012, Pegem Akademi

Bu kitabın basım, yayın ve satış hakları Pegem Akademi Yay. Eğt. Dan. Hizm. Tic. Ltd. Şti'ye aittir. Anılan kuruluşun izni alınmadan kitabın tümü ya da bölümleri, kapak tasarımı, mekanik, elektronik, fotokopi, manyetik, kayıt ya da başka yöntemlerle çoğaltılamaz, basılamaz, dağıtılamaz.

Bu kitap T.C. Kültür Bakanlığı bandrolü ile satılmaktadır. Okuyucularımızın bandrolü olmayan kitaplar hakkında yayinevimize bilgi vermesini ve bandrolsüz yayınları satın almamasını diliyoruz.

1. Baskı: Nisan 2010
2. Baskı: Ocak 2011
3. Baskı: Şubat 2012

Dizgi-Grafik Tasarım: Didem Gürleyik
Kapak Tasarımı: Gürsel Avcı
Baskı: Ayrıntı Matbaası
(Ankara-0312-3945590)

Yayıncı Sertifika No: 14749
Matbaa Sertifika No: 13987

YAYINEVİ

adakale sokak 4/1 yenişehir-ankara
tel: +90 312 4306750-51 (pbx)
belgeç: +90 312 4354460
gsm: 0506 3451936 - 0541 9104545 - 0533 2055230
e-ileti: pegem@pegem.net

DAĞITIM

sağlık sokak 17/a yenişehir-ankara
tel: +90 312 4345408 - +90 312 4345424
belgeç: +90 312 4313738
gsm: 0506 3451937 - 0541 4345424 - 0533 2055231
e-ileti: siparis@pegem.net

YAZIŞMA

p.k.175 06442 yenişehir-ankara

internet: **PEGEM.NET**

ÖZGEÇMİŞ

08.10.1978 yılında Fransa'da doğdum. Fransa'da başlayan eğitim hayatım, İstanbul'da devam etmektedir.

1999 yılında İstanbul Üniversitesi İşletme Fakültesi'nden mezun oldum. 2000 yılında aynı fakültede Sayısal Yöntemler Anabilim Dalı'nda Araştırma Görevlisi olarak akademik hayatıma başladım. Yüksek lisans ve doktoramı bu birimde tamamladım. Çok değerli hocam Prof. Dr. Haldun Akpınar'ın doktora sırasında vermekte olduğu "Veri Madenciliği" dersi ile bu konuya ilgim başladı. "Veri Madenciliği ve Müşteri İlişkileri Yönetimi" konulu doktora tezim ile bilgilerimi pekiştirdim.

2007 yılında Yrd. Doç. Dr. unvanını aldım. İ.Ü. İşletme Fakültesi'nde lisans düzeyinde "Sayısal Karar Verme Teknikleri", "Veri Madenciliği ve Bilgi Keşfi"; yüksek lisans düzeyinde "Veri Kaynakları Yönetimi ve İşletmecilik Uygulamaları", "Yöneylem Araştırması"; İ.Ü. Mühendislik Fakültesi'nde ise yüksek lisans düzeyinde "Veri Madenciliği" derslerini vermekteyim.

Veri madenciliği konusuyla ilgili öğretim üyeleriyle ortak çalışmak amacıyla Japonya, Yunanistan, İspanya, İtalya, Dubai ve Slovenya'da çeşitli üniversitelere misafir öğretim üyesi olarak ziyaretlerde bulundum, seminer verdim ve konferanslara katıldım.

*Canım Eşim Yusuf'a
Sevgiyle...*

*Aşk imiş her ne var alemde
İlim bir kıyl ü kal imiş ancak*

Fuzuli

ÖNSÖZ

Bilgi teknolojileri ve veritabanı sistemlerinin gelişmesi ile büyük hacimli verilere ulaşmak kolaylaşmış ve bu verilerin anlamlı bilgiye dönüştürülebilmesi ihtiyacı ortaya çıkmıştır. Veri Madenciliği bu amaçla son yıllarda yaygın olarak kullanılan ve dikkati çeken bir tekniktir. Elde edilen bilgiler, pazar sepet analizi, hilekârlıkların tespiti, müşteri elde etme ve kalite kontrol gibi farklı alanlarda kullanılmaktadır.

“*Veri Madenciliği ve Bilgi Keşfi*” adlı ilk kitabımda, okuyuculara veri madenciliği kavramını ve kullanılan teknikleri genel olarak tanıtmışım. *Tanımlayıcı ve Tahmin Edici Yöntemler* ana başlıkları altında;

- Birliktelik Kuralları Analizi,
- Ardışık Zamanlı Örüntüler,
- Kümeleme Analizi,
- Karar Ağaçları,
- Yapay Sinir Ağları,
- Genetik Algoritmalar,
- Regresyon Analizi,
- Lojistik Regresyon Analizi,
- Web Madenciliği ve
- Müşteri İlişkileri Yönetimi hakkında detaylı teorik bilgiler verilmiştir.

Bu kitabımda ise, farklı alanlardaki uygulamalara ve analizlere yer ayrılmıştır. İtalya’da bulunduğum süre içinde, İtalya’da bulunan büyük bir perakende zincirinin verilerini Sepet Analizi ile analiz etmişim. Bu çalışma ikinci bölümde incelenebilir. Bir üniversitenin yüksek lisans programına başvuran öğrencilerinin profilinin belirlenmesi, televizyon izleyicilerinin bir kanalı izlerken hangi demografik özelliklerinin belirleyici olduğunun tespit edilmesi, banka müşterilerinin kümeleme analizi ile incelenmesi, müşteri ilişkileri yönetimi, web madenciliği, tıp ve gazetecilik alanında gerçekleştirdiğim farklı uygulamalar da bu kitabımda yer almaktadır.

Günümüz bilgi çağında her alanda, gerçekleştirilen her işlem ile çok büyük miktarda veri elde edilmektedir. Artan veri miktarı kullanıcıların veriyi analiz etme ve yorumlaması için yeni stratejiler geliştirmesine sebep olmaktadır. Bunu başarabilen işletmeler rekabet avantajı elde etmekte ve daha başarılı olmaktadır. Bu nedenle kitabımın sadece öğrencilerime değil, özel sektörde veri analizi ile uğraşan çalışanlara da faydalı olmasını diliyorum.

TEŐEKKÜR

Beni bugüne getiren çok deęerli aileme,
Anlayışları için eőimin saygıdeęer ailesine,
Sevgili eőim Yusuf 'a,

Akademik hayatımdaki tüm destek ve yönlendirmeleri için hocalarım Sayın Prof.Dr. Öner Esen ve Sevgili Doç.Dr. Mehpare Timor'a,

Başarılarıyla ve yaratıcı fikirleriyle bizleri şaşırtan, kitabımın adı konusunda bana çok yardımcı olan sevgili meslektaşım Doç.Dr. Eyüp Çetin'e,

İsteđimi kırmayarak kitabımın on birinci bölümünü kaleme alan, Yönetim Bilişim Sistemleri Yüksek Lisans programından başarılı öđrencim sevgili Bayram Özşenele,

Bilimsel olarak gelişmeme katkıda bulunan tüm hocalarıma,

Güzel bir çalışma ortamını paylaştığımız İstanbul Üniversitesi İşletme Fakültesi Sayısal Yöntemler Ana Bilim Dalı'ndaki tüm arkadaşlarıma,

Sistematiik ve disiplinli çalışmaları ile birinci kitabımın da basımını gerçekleştiren Pegem Akademi Yayıncılık ekibine, özellikle Sayın Servet Sarıkaya'ya ve iki kitabımın da editörlüğünü titizlikle gerçekleştiren Sevgili Didem Gürleyik'e,

özel teşekkürlerimi sunarım.

Yrd. Doç. Dr. Umman Tuđba GÜR SOY

İÇİNDEKİLER

Giriş	1
-------------	---

BÖLÜM 1

VERİ MADENCİLİĞİ KAVRAMI	3
VERİ MADENCİLİĞİNDE KULLANILAN MODELLER	5
TAHMİN EDİCİ MODELLER	5
YARARLANILAN KAYNAKLAR	6

BÖLÜM 2

SÜPERMARKET VERİLERİNİN SEPET ANALİZİ İLE İNCELENMESİ	7
ANALİZİN AMACI	8
KULLANILAN VERİ SETİ	8
MODELLEME.....	9
ÖZET	14
YARARLANILAN KAYNAKLAR.....	14

BÖLÜM 3

ÖĞRENCİ PROFİLİ BELİRLEMEDE KÜMELEME ANALİZİ.....	15
ANALİZİN AMACI	16
KULLANILAN VERİ SETİ	16
MODELLEME.....	18
ÖZET	27
YARARLANILAN KAYNAKLAR	28

BÖLÜM 4

TELEVİZYON İZLEYİCİLERİNİN KARAR AĞAÇLARI İLE DEĞERLENDİRİLMESİ	29
ANALİZİN AMACI	31
KULLANILAN VERİ SETİ	31
MODELLEME.....	32
ÖZET	34
YARARLANILAN KAYNAKLAR	34

BÖLÜM 5

BANKA MÜŞTERİLERİNİN PROFİLLERİNİN İSTATİSTİK ANALİZLERLE BELİRLENMESİ.....	35
ANALİZİN AMACI	36
KULLANILAN VERİ SETİ	36
MODELLEME.....	37

ÖZET	46
YARARLANILAN KAYNAKLAR	46

BÖLÜM 6

İSTATİSTİK ANALİZLER	47
BASİT DOĞRUSAL REGRESYON ANALİZİ	47
ANALİZİN AMACI	48
KULLANILAN VERİ SETİ	48
MODELLEME.....	49
ÇOKLU DOĞRUSAL REGRESYON ANALİZİ	52
ÖZET	57
YARARLANILAN KAYNAKLAR	57

BÖLÜM 7

MÜŞTERİ İLİŞKİLERİ YÖNETİMİ.....	59
CRM BİLEŞENLERİ	61
CRM'NİN AMAÇLARI	62
CRM SÜRECİ.....	62
1. Müşteri Seçimi:	63
2. Müşteri Edinme:	63
3. Müşteri Koruma:	64
4. Müşteri Derinleştirme:	64
CRM KAPSAMI.....	65
OPERASYONEL CRM	65
ANALİTİK CRM.....	66
İŞBİRLİKÇİ CRM	66
UMMAN FİRMASI.....	66
CRM UYGULAMALARI	68
CRM UYGULAMASININ GETİRDİĞİ AVANTAJLAR.....	68
CRM UYGULAMASINDA KARŞILAŞILAN RİSKLER.....	71
ÖZET	73
YARARLANILAN KAYNAKLAR.....	74

BÖLÜM 8

WEB MADENCİLİĞİ	75
ANALİZİN AMACI	78
KULLANILAN VERİ SETİ	79
VERİNİN ANALİZİ.....	81
MODELLEME.....	85
ÖZET	88
YARARLANILAN KAYNAKLAR	88

BÖLÜM 9

TIP ALANINDA BİR VERİ MADENCİLİĞİ UYGULAMASI	89
ANALİZİN AMACI	90
KULLANILAN VERİ SETİ	90
MODELLEME.....	94
ÖZET	97
YARARLANILAN KAYNAKLAR	97

BÖLÜM 10

GAZETE OKUR PROFİLİNİN VERİ GÖRSELLEŞTİRME TEKNİKLERİ İLE BELİRLENMESİ	99
ÖZET.....	106
YARARLANILAN KAYNAKLAR	106

BÖLÜM 11

OLAP KÜPLERİ.....	107
VERİ ANALİZİNE NEDEN İHTİYAÇ DUYULUR?	107
OLAP (ON LINE ANALYTIC PROCESSING)	107
OLAP ÖZELLİKLERİ	109
OLAP SİSTEMLERİNİN YARARLARI.....	111
SQL SERVER ANALYSIS SERVICES	112
SQL SERVER DATABASE ENGINE.....	112
BUSSINESS INTELLIGENCE DEVELOPMENT STUDIO(BIDS).....	114
Data source view (DSV).....	115
Dimensions Table	117
Veri Küpleri (Cubes)	119
ANALYZER	120
Müşterilere Hammadde Satışları.....	120
Makine Bazında Üretim	121
ÖZET.....	124
KAYNAKÇA	125

Şekiller

Şekil 1.1 Veri Madenciliği Süreci	3
Şekil 2.1 Sepet Analizi İçin Yapılan Tanımlamalar	9
Şekil 2.2 Web Grafiği	10
Şekil 2.3 Sıklık Grafiği	11
Şekil 2.4 Birliktelik Kuralları Analizi İçin Modelin Kurulması	12
Şekil 2.5 Birliktelik Kuralları Analizi Sonuçları	12
Şekil 3.1 Verilere İlişkin Serpilme Grafiği	15
Şekil 3.2 Cinsiyet Dağılımı	18
Şekil 3.3 Doğum Yerleri Dağılımı	18
Şekil 3.4 Medeni Durum Dağılımı	19
Şekil 3.5 Mezun Olunan Üniversite Dağılımı	19
Şekil 3.6 Mezun Olunan Bölüm Dağılımı	20
Şekil 3.7 Yüksek Lisansa Girme Nedenleri Dağılımı	20
Şekil 3.8 Tercih Edilen Anabilim Dalları Dağılımı	21
Şekil 3.9 Öğrencilerin Yükseköğretim Sistemini Yeterli Bulup Bulmadıkları Dağılımı	21
Şekil 3.10 Öğrencilerin Türkiye'nin Avrupa Birliği'ne Üyeliği ile İlgili Görüşleri Dağılımı	22
Şekil 3.11 Öğrencilerin Yaş Dağılımı	22
Şekil 3.12 Öğrencilerin Mezun Oldukları Liselere Göre Dağılımları	23
Şekil 3.13 Öğrencilerin Ayda Kaç Kez Sinemaya Gittiklerine İlişkin Dağılım.....	23
Şekil 3.14 Öğrencilerin Ayda Kaç Kez Tiyatroya Gittiklerine İlişkin Dağılım.....	24
Şekil 3.15 Öğrencilerin Ayda Kaç Kez Konsere Gittiklerine İlişkin Dağılım	24
Şekil 3.16 Öğrencilerin Ayda Kaç Kez Sergiye Gittiklerine İlişkin Dağılım.....	24
Şekil 3.17 Öğrencilerin Günde Kaç Saat TV İzlediklerine İlişkin Dağılım.....	25
Şekil 3.18 Öğrencilerin Tercih Ettikleri Program Türüne Göre Dağılımları.....	25
Şekil 3.19 Öğrencilerin Okudukları Kitap Sayısına Göre Dağılımları.....	25
Şekil 3.20 Öğrencilerin Düzenli Dergi Okuma Durumuna Göre Dağılımları.....	26
Şekil 3.21 Kümeleme Analizi Sonuçları	26
Şekil 4.1 Örnek Karar Ağacı	30
Şekil 4.2 Karar Ağacı Analizi Sonuçları	32
Şekil 5.1 Veri Madenciliğinin İlişkide Olduğu Disiplinler	35
Şekil 5.2 Cinsiyet Dağılımları	37
Şekil 5.3 Meslek Dağılımları	38
Şekil 5.4 Medeni Durum Dağılımları.....	39
Şekil 5.5 Eğitim Seviyesi Dağılımları	39
Şekil 5.6 Aylık Gelir Aralığı Dağılımları	39

Şekil 5.7 Müşteri Tipi Dağılımları	40
Şekil 5.8 Nüfusa Kayıtlı Olduğu İl Dağılımları	40
Şekil 5.9 Yaş Değişkenine İlişkin Histogram.....	41
Şekil 5.10 Birinci Kümenin Profili	43
Şekil 5.11 İkinci Kümenin Profili	44
Şekil 5.12 Üçüncü Kümenin Profili	45
Şekil 5.13 Dördüncü Kümenin Profili	45
Şekil 6.1 Basit Regresyon Analizi İçin Kurulan Model.....	49
Şekil 6.2 Regresyon Analizi Bağımlı ve Bağımsız Değişkenlerin Tanımlanması	50
Şekil 6.3 Regresyon Denklemi Katsayıları.....	50
Şekil 6.4 Basit Regresyon Analizi Çıktısı	51
Şekil 6.5 Çoklu Regresyon Analizi İçin Kurulan Model	53
Şekil 6.6 Regresyon Denklemi Katsayıları.....	53
Şekil 6.7 Çoklu Regresyon Analizi Çıktısı- 1.....	54
Şekil 6.8 Çoklu Regresyon Analizi Çıktısı- 2.....	55
Şekil 6.9 Çoklu Regresyon Analizi Çıktısı- 3.....	56
Şekil 8.1 Web Madenciliği.....	76
Şekil 8.2 Web Madenciliği Aşamaları.....	78
Şekil 8.3 Web Grafiği	86
Şekil 9.1 Kurulan Model	91
Şekil 9.2 İlaç Dağılımları.....	92
Şekil 9.3 Sodyum ve Potasyum Oranlarına Göre İlaç Dağılımları.....	93
Şekil 9.4 Sodyum/ Potasyum Oranına Göre İlaç Dağılımları	94
Şekil 9.5 Kümeleme Analizi ve Karar Ağaçları İçin Veri Setinin Hazırlanması.....	94
Şekil 9.6 Kümeleme Analizi	95
Şekil 9.7 Kümeleme Analizi Sonuçları.....	96
Şekil 9.8 Karar Ağaçları Analizi Sonuçları.....	96
Şekil 10.1 Okur Sayısı Grafiği.....	100
Şekil 10.2 Erkek Okur Sayısı Grafiği	100
Şekil 10.3 Kadın Okur Sayısı Grafiği	101
Şekil 10.4 Okur Sayısı Grafiği.....	101
Şekil 10.5 25- 34 Yaş Arası Okur Sayısı Grafiği	102
Şekil 10.6 Gazete Okuma Sıklığı	102
Şekil 10.7 Düzenli Okur-Düzensiz Okur Dağılımı	102
Şekil 10.8 Bölgelere Göre Dağılım	103
Şekil 10.9 Cinsiyete Göre Sayılar ile Dağılım	103
Şekil 10.10 Cinsiyete Göre Oranlar ile Dağılım	103

Şekil 10.11 21- 44 Yaş Okurların Cinsiyete Göre Dağılımı.....	104
Şekil 10.12 Okurların Yaşadıkları Bölgelere Göre Dağılımı	104
Şekil 10.13 Okurların Yaşadıkları Bölgelere Göre Oransal Dağılımı	104
Şekil 10.14 Araç Sahibi Olma Dağılımı.....	105
Şekil 10.15 Gazeteyi Okuma Süresi Dağılımı	105
Şekil 10.16 En Çok Okunan Sayfaların Dağılımı	106
Şekil 11.1 OLAP Çalışma Ağı.....	108
Şekil 11.2 Veri Analiz Süreci	110
Şekil 11.3 Molap, Rolap, Holap.....	112
Şekil 11.4 Database Engine View Örneği.....	113
Şekil 11.5 Design View Örneği.....	113
Şekil 11.6 View Sorgulama Örneği.....	114
Şekil 11.7 BIDS Analysis Services Bağlantısı.....	114
Şekil 11.8 Data Source View.....	115
Şekil 11.9 Ana Tablo ve Boyutlar	116
Şekil 11.10 Data Source View, Küp Yapısı	117
Şekil 11.11 Üç Boyutlu Küp Yapısı	118
Şekil 11.12 Dört Boyutlu Bir Küp Yapısı	118
Şekil 11.13 Dimensions (Boyut) Yapısı.....	119
Şekil 11.14 Yeni Küp	119
Şekil 11.15 Küp Yapısı	120
Şekil 11.16 Histogram Görünümü.....	123
Şekil 11.17 Pasta Görünümü.....	123
Şekil 11.18 Çizgi Grafik	124

Tablolar

Tablo 2.1 Veri Seti.....	8
Tablo 2.2 Birliktelik Kuralları Analizi Detaylı Sonuçları	13
Tablo 3.1 Veri Seti	15
Tablo 3.2 Veri Seti.....	17
Tablo 3.3 Veri Setinde Yer Alan Bazı Değişkenler.....	17
Tablo 4.1 Veri Seti	30
Tablo 4.2 Veri Seti	31
Tablo 4.3 İzleyici Profili.....	33
Tablo 5.1 Veri Seti	36
Tablo 5.2 Yaş Değişkenine İlişkin İstatistikler	41
Tablo 5.3 Meslek ve Cinsiyete İlişkin Çapraz Tablo	42

Tablo 6.1 <i>Veri Seti</i>	48
Tablo 8.1 <i>Veri Seti</i>	79
Tablo 8.2 <i>Ziyaretçiler Tablosu</i>	79
Tablo 8.3 <i>Tıklama Sayısına Göre Dağılım</i>	81
Tablo 8.4 <i>Gezinme Süresine Göre Dağılım</i>	81
Tablo 8.5 <i>Başlama Süresine Göre Dağılım</i>	82
Tablo 8.6 <i>Frekans Tablosu</i>	83
Tablo 8.7 <i>Satın Alma Davranışına Göre Karşılaştırma Tablosu</i>	84
Tablo 8.8 <i>Kümeleme Analizi Sonuçları</i>	84
Tablo 8.9 <i>Birliktelik Kuralları Analizi Sonuçları</i>	87
Tablo 9.1 <i>Veri Seti</i>	90
Tablo 9.2 <i>Veri Seti</i>	91
Tablo 9.3 <i>Hesaplanan Na/K Değerleri</i>	92
Tablo 11.1 <i>Hammadde Bazında Aylık Müşteri Kg Satışları</i>	121
Tablo 11.2 <i>Makine Bazında Üretimde Kullanılan Hammadde Tonajları</i>	122
Tablo 11.3 <i>Raporda Kullanılan Boyutlar</i>	122

GİRİŞ

Veri madenciliği tekniđi büyük veri yığınları arasından anlamlı, önceden bilinmeyen ve tahmin edilemeyen yapıların keşfedilmesi için son yıllarda yaygın olarak kullanılan ve popüler olan bir konudur.

“Veri Madenciliđi ve Bilgi Keşfi” adlı ilk kitabımda, veri madenciliđi kavramını oluřturan teknikler detaylı olarak ele alınmıřtı.

Okuyucular, Veri Madenciliđi teknikleri ile ilgili teorik bilgiye ulařabilecekleri yerli ve yabancı pek çok kaynađa eriřebilirler. Bu kitapta, diđer kitaplardan farklı olarak veri madenciliđi tekniklerinin çeřitli sektörlere ait uygulama örneklerine yer verilmiřtir. Böylece konunun somutlařması ve daha kolay anlaşılabilir olması hedeflenmiřtir.

Veri madenciliđi teknikleri pazarlama, tıp, finans gibi farklı alanlarda farklı uygulamalarda kullanılabilir. Bu kitabımda perakendecilik, eğitim, medya, bankacılık, müşteri ilişkileri yönetimi, internet ve tıp alanında gerçekleřtirdiđim uygulamalar yer almaktadır.

Kitapta onbir bölüm bulunmaktadır.

Birinci bölümde Veri Madenciliđi kavramına ve Veri Madenciliđi tekniklerine genel olarak değinilmiřtir.

İkinci bölümde büyük bir perakende firmasının satıř verileri ele alınmıř ve Sepet Analizi ile bir uygulama gerçekleřtirilmiřtir. Analizde ürün satıřlarına ait birlikteliklerin belirlenmesi hedeflenmektedir. Bu durum çapraz satıřların artmasına olanak sađlamaktadır.

Üçüncü bölümde, öğrenci profili belirlemede kümeleme analizinden yararlanılmıřtır. Bir fakültenin iřletme bölümüne bařvuran yüksek lisans adayı öğrencilerin verileri kullanılarak tanımlayıcı analizler gerçekleřtirilmiřtir.

Dördüncü bölümde bir televizyon kanalının izleyici verileri analiz edilmiřtir. Amaç kanalın yayın akıřında yer alan bir haber programının izleyici sayısını maksimize edebilmek için, izleyicilerin profilini karar ađaçları analizi ile belirlemektir. Belirlenen profil dikkate alınarak kanal, izleyicilerine yönelik stratejiler geliřtirebilir.

Beřinci bölümde, bir bankanın müşteri verileri dikkate alınarak, müşteri demografik verileri istatistik tekniklerle analiz edilmiř, daha sonra kümeleme analizi uygulanarak müşteri profilleri belirlenmiřtir.

Kitabın altıncı bölümünde İstatistik Tekniklerden Basit ve Çoklu Regresyon Analizlerine iliřkin uygulamalara yer verilmiřtir.



Müşteri İlişkileri Yönetimi (Customer Relationship Management-CRM) son yılların gözde kavramlarından. CRM projesi ile yapılan şirketler, müşterilerini iyi tanıyıp onların beklentilerini iyi belirleyen ve onların ihtiyacına göre hizmet üreten bir yapı oluşturmaktadır. Yedinci bölümde CRM kavramı kısaca ele alınmış; bir işletmedeki CRM uygulamalarına yer verilmiştir.

Web madenciliği tekniği, web üzerinden alınan verilerin analizinde yaygın olarak kullanılmaktadır. Kitabın sekizinci bölümünde bir e-ticaret sitesini ziyaret eden kullanıcıların tıkladıkları sayfalar izlenerek, bir web madenciliği uygulaması gerçekleştirilmiştir.

Kitabın dokuzuncu bölümü, tıp alanında gerçekleştirilen bir uygulamayı içermektedir. Hasta verileri ve hastalara verilen ilaç verileri dikkate alınarak öncelikle veri setinin genel yapısı okuyucuya tanıtılmış, daha sonra kümeleme analizi ve karar ağaçları analizi ile bir uygulama gerçekleştirilmiştir.

Kitabın onuncu bölümünde bir gazetenin okuyucu profili veri madenciliği görselleştirme araçları ile incelenmiş ve sonuçlar detaylarıyla sunulmuştur.

Kitabımın onbirinci bölümü, Yönetim Bilişim Sistemleri programında yüksek lisans öğrencim olan sevgili Bayram Özşenel tarafından ele alınmıştır. Bu bölümde OLAP küpleri ile ilgili bir uygulama gerçekleştirilmiştir.

Kitabımın okuyuculara faydalı olmasını diliyorum. Çok titiz davranmama karşın kitapta bazı eksiklik ve hatalar bulunabilir. Okuyuculardan gelecek eleştirilerin bundan sonraki çalışmalarında yararlı olacağını düşünüyorum.

*Umman Tuğba Gürsoy
İstanbul, Nisan 2010
tugbasim@istanbul.edu.tr*

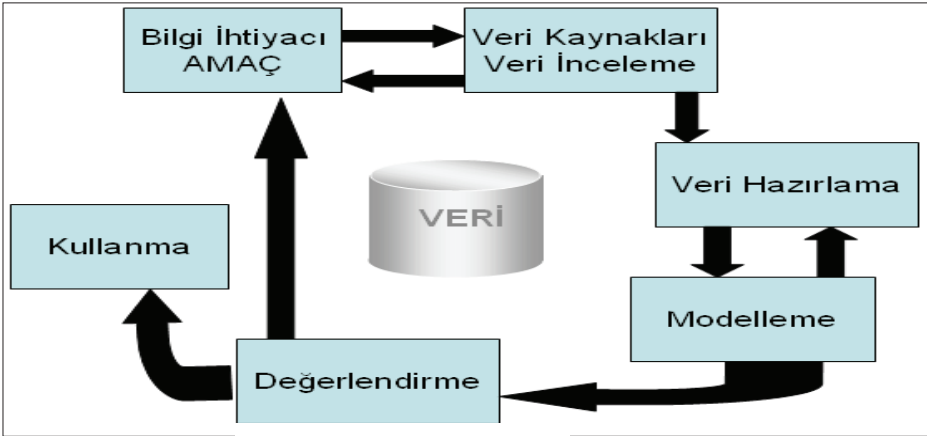
VERİ MADENCİLİĞİ KAVRAMI

Veri madenciliği tekniği “*Büyük veri tabanlarından gizli bilgilerin, önceden tahmin edilemeyen örüntülerin (pattern¹) ve yeni kuralların keşfedilmesi*” olarak tanımlanabilir. Veri Madenciliği, Veri Tabanlarında Bilgi Keşfi (Knowledge Discovery in Databases- KDD) sürecinin en önemli aşamasıdır. [1]

Bilgi ve teknolojinin yoğun olarak kullanıldığı günümüzde bilgisayarların ve sayısal teknolojilerin gelişmesiyle birlikte çok büyük miktarda veri birikimi meydana gelmiş, mevcut veri yığınları arasından anlamlı ilişkilerin, yapıların ve eğilimlerin ortaya çıkarılması ihtiyacı doğmuş, gelecek adına sağlıklı tahminlerin yapılması önem kazanmıştır. İşte bu tür ihtiyaçların doğduğu noktada, istatistiksel analiz ve modellemeler ile yapay zeka tekniklerini kullanan veri madenciliği uygulamaları devreye girmiştir.

Veritabanı sistemlerinin gelişmesiyle birlikte perakende, tıp, bankacılık, finans, ekonomi, otomotiv, güvenlik ve savunma gibi pek çok alanda veri madenciliği teknikleri geliştirilmeye başlanmış, tahmin analizleri, modellemeler ve yapay zeka teknikleri alanlarında ciddi ilerleme kaydedilmiştir. Yapılan uygulamalarda amaç maliyetleri düşürmek, satışları artırmak ve araştırma-geliştirme çalışmalarını daha etkin kılmaktır.

Farklı sektörlerde gerçekleştirilebilecek uygulamalar için toplanan verilerin analizi Şekil 1.1’de görülen veri madenciliği standart süreci dikkate alınarak yapılmaktadır. Bu süreç CRISP-DM (Cross Industry Standard Process for Data Mining) olarak adlandırılmaktadır.



Şekil 1.1 Veri Madenciliği Süreci [5]

1. Sürekli devam eden ve tekrar eden yapılara **Örüntü (Pattern)** adı verilmektedir.



Şekil 1.1 incelenirse altı aşamadan oluşan sürecin **Bilgi İhtiyacının Belirlenmesi** ile başladığı görülmektedir. Müşteri sadakatinin veya müşteri karlılığının artırılması, tedarik sürecinin etkinliğini ve verimliliğini arttırmak, e-ticaret sitesine giren ve alışveriş yapan ziyaretçi sayısını arttırmak gibi genel amaçlar belirlenebilir.

İkinci aşamada **Verilerin Hangi Kaynaklardan Elde Edileceğine** karar verilmelidir. Bilginin keşfi için gerekli veriler farklı kaynaklarda bulunabilir. Toplama işlemi; verinin farklı kaynaklardan alınarak bir kaynaktan birleştirilmesidir. İhtiyaç duyulan veriler firma içi kaynaklardan, veri ambarlarından veya hazır veri sunan sitelerden elde edilebilir. Örneğin hisse senetleri ile ilgili analizler yapılacaksa, İstanbul Menkul Kıymetler Borsası sitesi güvenilir bir kaynak olabilir.

Veri madenciliğinin üçüncü aşaması **Veri Hazırlama** aşamasıdır. Bu aşamada hangi değişkenlerin ve hangi kayıtların analizde kullanılacağına karar verilmektedir. Gözlem sayısının çok fazla olması durumunda örnekleme yapılarak zaman ve maliyetlerden tasarruf edilebilir.

Veri setinde eksik veya gürültülü veri bulunup bulunmadığı kontrol edilmelidir. **Eksik veriler**, veri girişi sırasında elde edilememiş olan değerlerdir. Örneğin veri setinde kişilerin yaşları veya meslekleri eksik olabilir. Bu durumda eksik verileri zamanla tamamlama yoluna gidilebilir, kayıt tamamen silinebilir veya eksik değer tahmini bir değerle doldurulabilir.

Gürültülü veriler ise veri girişinin hatalı olması durumunda ortaya çıkmaktadır. Örneğin kişinin yaşı 300 veya çocuk sayısı 90 olarak girilmiş olabilir. Bu özel durumlar dikkate alınarak veri setinin düzenlenmesi gerekmektedir.

Veri kodlama işleminin standart olup olmadığına da ayrıca dikkat edilmelidir. Veri kodlama aşamasında cinsiyetler için (1, 0) veya (Erkek, Kadın) değerleri girilmiş olabilir. Standardizasyon işlemi, analiz sırasında sorunlar yaşanmasını engellemektedir.

Modelleme aşamasında veri setinin yapısına ve analiz amaçlarına uygun olan *Tanımlayıcı veya Tahmin Edici Yöntemlere* karar verilmektedir. Modelin varsayımları dikkate alınarak, uygun algoritmalar seçilmektedir. Örneğin kümeleme analizi uygulanıyorsa, k-ortalama algoritması tercih edilebilir.

Başlangıçta belirlenen problemin amacı ile uygulanan model sonucunda elde edilen çıktılarının uygun olup olmadığı **Değerlendirme** aşamasında incelenmektedir.

Analiz sonucu elde edilen sonuçlar anlamlı ise **Modelin Kullanılması** aşamasına geçilmektedir.

Sürecin bazı aşamalarında akışın çift yönlü olduğu dikkati çekebilir. Örneğin **bilgi ihtiyacının belirlenmesi** ve **verilerin hangi kaynaklardan elde edileceğine** karar verilmesi aşamalarında iki yönlü akış dikkati çekmektedir. Bu durum, sürecin bir aşamasında problem yaşanırsa, bir diğer adıma geri dönmek gerektiğini göstermektedir.



VERİ MADENCİLİĞİNDE KULLANILAN MODELLER

Veri madenciliğinde kullanılan modeller “Tanımlayıcı ve Tahmin Edici Modeller” olmak üzere ikiye ayrılmaktadır. [3]

TANIMLAYICI MODELLER

Tanımlayıcı modellerde (Descriptive Models) amaç, karar vermeye rehberlik etmede kullanılacak mevcut verilerdeki örüntülerin tanımlanmasını sağlamaktır. Geliri 3.000 ve 5.000 para birimi aralığında olan ve iki veya daha fazla arabası olan çocuklu aileler ile çocuğu olmayan ve geliri 1.500 ve 3.000 para birimi aralığından düşük olan ailelerin satın alma örüntülerinin benzerlik gösterip göstermediğinin belirlenmesi tanımlayıcı modelele bir örnektir. [2]

Tanımlayıcı modeller,

- İlişki analizi ve
- Kümeleme analizi olmak üzere iki grupta incelenmektedir.

Birliktelik Kuralları (Association Rules) ile Ardışık Zamanlı Örüntüler (Sequential Patterns Analysis) ilişki analizi kapsamında yer almaktadır.

TAHMİN EDİCİ MODELLER

Tahmin edici modellerde (Predictive Models) sonuçları bilinen verilerden hareket edilerek bir model geliştirilmesi ve kurulan bu modelden yararlanılarak, sonuçları bilinmeyen veri kümeleri için sonuç değerlerinin tahmin edilmesi amaçlanmaktadır. Örneğin bir banka önceki dönemlerde vermiş olduğu kredilere ilişkin gerekli tüm verilere sahip olabilir. Bu verilerde bağımsız değişkenler kredi alan müşterinin özellikleri, bağımlı değişken değeri ise kredinin geri ödenip ödenmediğidir. Bu verilere uygun olarak kurulan model, daha sonraki kredi taleplerinde müşteri özelliklerine göre verilecek olan kredinin geri ödenip ödenmeyeceğinin tahmininde kullanılmaktadır. [4]

Tahmin edici modeller,

- Sınıflandırma ve
- İstatistiksel Tahmin Modelleri olmak üzere iki ana başlık altında incelenmektedir.

Karar Ağaçları, Yapay Sinir Ağları ve Genetik Algoritmalar en yaygın olarak kullanılan Sınıflandırma teknikleridir. İstatistiksel Tahmin Modelleri ise Regresyon Analizi, Diskriminant (Ayrırma) Analizi ve Lojistik Regresyon Analizini kapsamaktadır. İstatistik literatüründe kullanılan pek çok teknik bulunmaktadır. Ancak, kitabımda bu üç tekniğe ait uygulamalara yer verilmiştir.



Veri madenciliği büyük veri setleriyle çalışan tüm organizasyonlarda kullanılacak olan bir tekniktir. “**Veri Madenciliği ve Bilgi Keşfi**” adlı ilk kitabımda tanımlayıcı ve tahmin edici modeller detaylarıyla ele alınmıştır. Bu nedenle bu kitabımda veri madenciliği tekniklerinin farklı sektörlerdeki uygulamalarına ağırlık verilmiştir.

YARARLANILAN KAYNAKLAR

- [1] Adrians Pieter, Zantinge Dolf, **Data Mining**, Addison- Wesley, England, 1997.
- [2] Akpınar Haldun, **Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği**, İ.Ü.İşletme Fakültesi Dergisi, İstanbul, Sayı:1, Nisan 2000.
- [3] Bounsaythip Catherine, Rinta-Runsala Esa, **Overview of Data Mining for Customer Behaviour Modeling**, VTT Information Technology Research Report TTEI, 2001.
- [4] Eker Hakan, **Veri Madenciliği ve Bilgi Keşfi**, <http://www.bilgiyonetimi.org/cm/pages/>, 09.12.2005.
- [5] Özmen Şule, **Bilginin Değeri/ Verinin Önemi**, Marmara Üniversitesi İ.İ.B.F. Sunumu, 2002.

SÜPERMARKET VERİLERİNİN SEPET ANALİZİ İLE İNCELENMESİ

Bu bölümde süpermarket müşterilerinin davranışlarını Sepet Analizi (Market Basket Analysis) ile belirlemek amaçlanmıştır. Sepet analizi ile aynı alışverişte birlikte satın alınan ürünler veya ürün grupları tespit edilmektedir. Analiz sonucunda elde edilen bilgiler çok değerlidir. Bu bilgiler kullanılarak raf düzenlemeleri yapılabilir, sıklıkla birlikte satılan ürünlerin birbirine yakın yerlere konması satışların artmasını sağlamaktadır. Analiz sonuçları, düzenlenen promosyonların daha etkin olmasını da sağlamaktadır.

Sepet analizinde kullanılacak veri seti, genellikle belirli bir satış yerindeki (örneğin süpermarket) belirli bir döneme ait (örneğin 1 yıl) alışveriş kayıtlarından oluşmaktadır. Müşteriler veri setinde birden fazla kez yer alabilir. Müşteriler satış yerinde yaptıkları her bir işlem için veri setinde kayıt olarak yer alırlar. Sepet analizinin amacı müşterilerin en sık satın aldığı ürün kombinasyonlarını belirlemektir. Müşterilerin ürünleri satın alma sıklığı kadar, kaç kayıta bu birlikteliklerin tekrarlandığı da önemlidir. Bu durum bizi kurallar elde etmeye yöneltmektedir. Bu nedenle Sepet analizi, Birliktelik kuralları analizi (Association Rules Analysis) olarak da adlandırılmaktadır. Birliktelik kuralları analizi sonuçları kolaylıkla yorumlanabilmektedir. [2]

Birliktelik kuralları, eş zamanlı olarak gerçekleşen ilişkilerin tanımlanmasında kullanılır. Örneğin;

- Bir giyim mağazasında ceket ve kemer alan müşterilerin %89'u pantolon da satın almaktadır.
- Bir restaurantta havyar siparişi veren bir kişi %95 olasılıkla votka da isteyecektir. [1]
- Tatile gitmek üzere bir otelde yer ayırtan, ayrıca kredi kartı ile uçakla rezervasyon yaptıran çocuklu bir aile; %85 olasılıkla gideceği yerde araba da kiralayacaktır.



+



→





Bu bilgiye sahip olan bir seyahat acentası, araba kiralama şirketleriyle bağlantıya geçerek; aileye daha tatile gitmeden araba kiralamada indirim sunabilir.

ANALİZİN AMACI

Sepet analizi ile süpermarketlerde müşterilerin birlikte satın aldığı ürünlerin belirlenmesi hedeflenmektedir. Bu bölümde yer alan uygulamada İspanya'da bulunan büyük bir perakende firmasının verileri kullanılarak, alışveriş birlikteliklerinin tespiti hedeflenmiştir.

KULLANILAN VERİ SETİ

Analizde kullanılan veri seti İspanyol perakende firmasına ait Mayıs 2007 – Kasım 2007 dönemine ait 7 ayı kapsayan 529.298 kayıt ve 31 üründen oluşmaktadır. Veri seti Fatura numarası ve ürünlerin satın alınma-alınmama bilgilerini içermektedir. Müşterilerin satın aldığı ürünler “1” ile, satın almadıkları ürünler ise “0” ile kodlanmıştır. Veri seti Tablo 2.1’de görülmektedir.

Tablo 2.1 *Veri Seti*

Fatura No	Kahvaltılık	Kahve	Yağ	Aperitif	Konserve	Şeker	Yoğurt
100	0	1	0	0	0	0	0
120	0	0	1	0	0	0	0
140	0	0	0	0	1	1	0
160	0	0	0	1	1	1	0
180	0	0	0	0	1	1	0
200	1	0	0	0	0	0	0
220	0	0	0	0	0	0	1
240	1	0	0	0	1	0	0
260	0	0	0	0	0	0	0
280	0	1	0	0	0	0	0
300	0	0	1	0	0	0	0
320	0	0	0	0	1	0	0
340	0	0	0	0	0	0	1
360	1	0	0	0	0	0	1
380	0	0	0	0	0	1	0
400	1	1	0	0	0	0	0

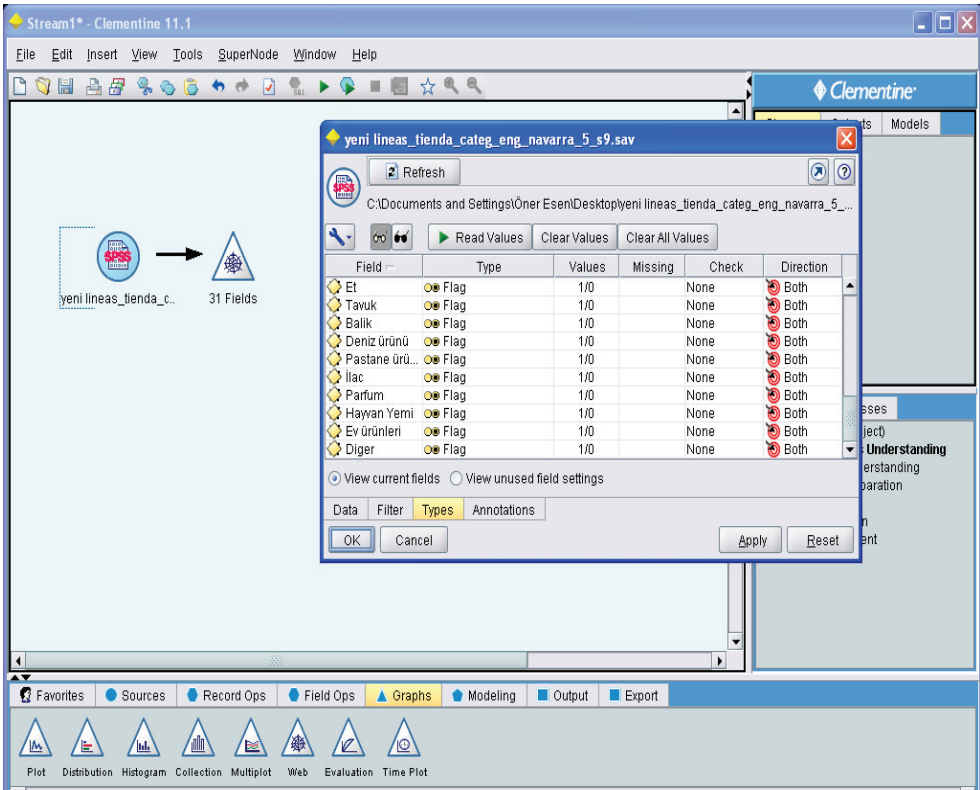


MODELLEME

Süpermarket verileri ile Sepet Analizi gerçekleştirilirken, sıklıkla birlikte satın alınan ürünlerin belirlenmesi hedeflenmektedir. SPSS Clementine programında Web grafiği ile bu durum görsel olarak elde edilebilir, Birlikte Kuralları Analizi ile görsel olarak belirlenen birlikte kuralları kural seti olarak ifade edilebilir.

Web grafiği (Web Graph) tüm ürünler arasındaki ilişkileri göstermektedir. SPSS Clementine programında web grafiği elde etmek için analize girecek olan ürünlerin “Both”, analize girmeyecek diğer verilerin ise “None” olarak tanımlanması gerekmektedir.

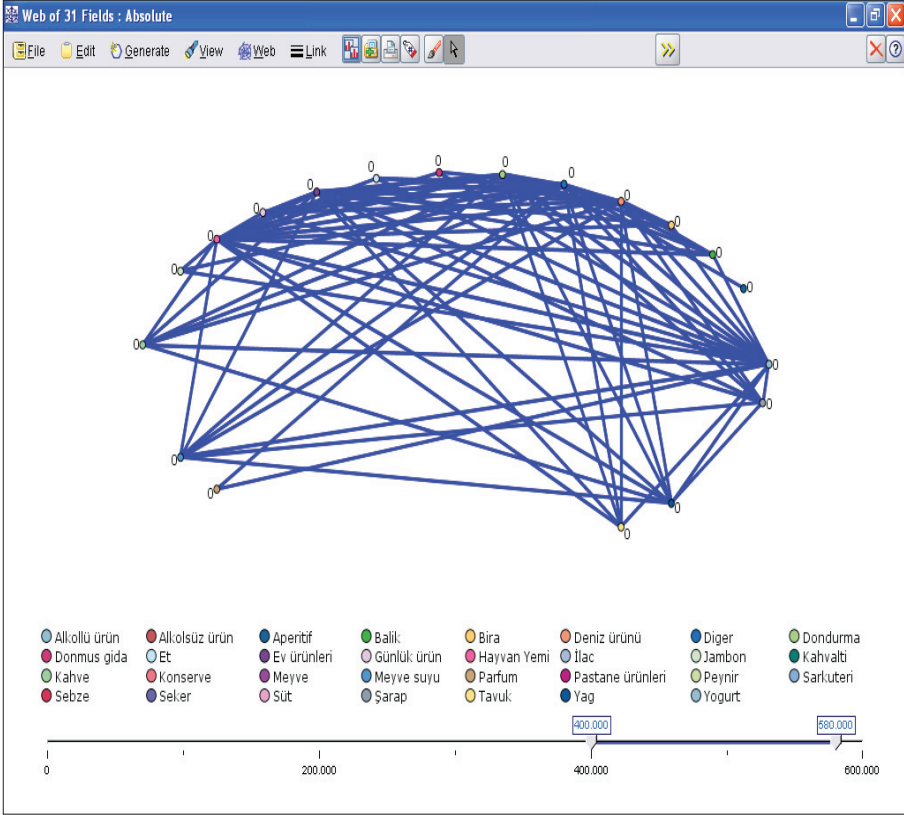
Veri setinde 31 ürün bulunmaktadır. Yapılan tanımlamalar Şekil 2.1’de görülmektedir. “Type” sütununda görülen “Flag” tanımlaması, veri setindeki değişkenlerin iki şıka sahip olması durumunda kullanılmaktadır. Bu analizde müşterilerin bir ürünü satın alması durumunu belirtmek için “1”, satın almaması durumu için “0” kodu kullanılmıştır.



Şekil 2.1 Sepet Analizi İçin Yapılan Tanımlamalar



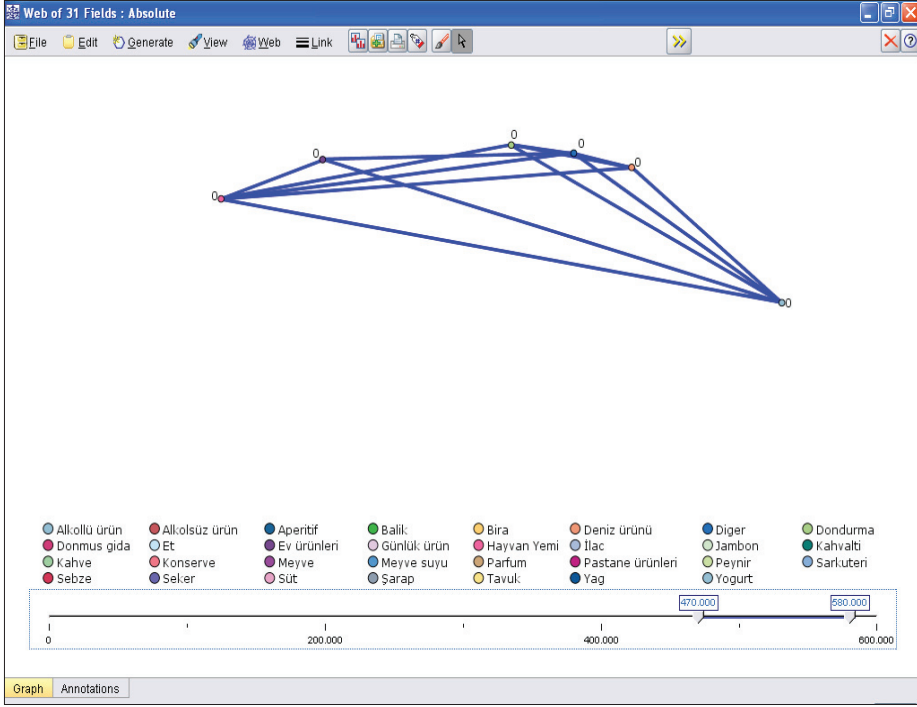
31 ürün grubu ile yapılan analiz sonucunda elde edilen web grafiği Şekil 2.2’de görülmektedir.



Şekil 2.2 *Web Grafiği*

Web grafiği veri setindeki tüm ilişkileri göstermektedir. Oysa uzman kişiler “Sıklıkla satın alınan ürünler nelerdir?” sorusuna cevap aramaktadırlar. Bu soruya cevap verebilmek için, “Frequency Graph- Sıklık Grafiği” elde edilmelidir.

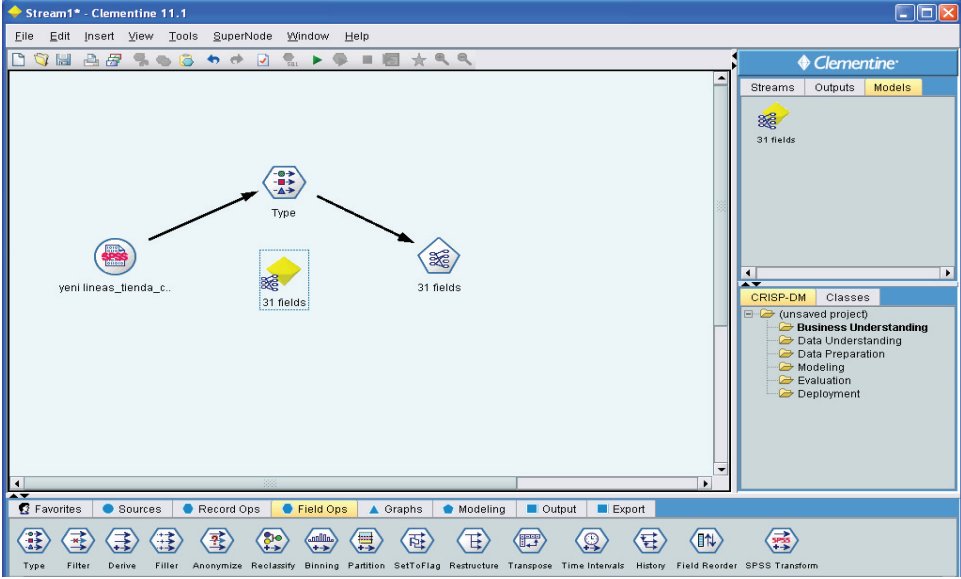
Şekil 2.3’ te alt tarafta görülen çubuk üzerindeki buton sağa kaydırılarak, en fazla birlikte satın alınan ürünler belirginleştirilebilir.



Şekil 2.3 Sıklık Grafiği

Birlikte sıklıkla satın alınan ürünler peynir, kahvaltılık ürünler, süt, şarküteri ürünleri, meyve, sebze olarak görülmektedir. Şekildeki dairelerin üzerine mouse götürüldüğünde, ürün isimleri okunabilmektedir.

Görsel olarak elde edilen ilişkiler, Birliktelik Kuralları analizi ile kural seti olarak ifade edilebilirler. Öncelikle modelin kurulması gereklidir. Analiz için sepet analizinde en yaygın olarak kullanılan **APRIORI** algoritması seçilmiştir.



Şekil 2.4 Birliktelik Kuralları Analizi İçin Modelin Kurulması

Modelleme sonucunda Şekil 2.5' teki sonuçlar elde edilmiştir:

Consequent	Antecedent	Support %	Confidence %
Sebze	Meyve	25,134	59,331
Sarkuteri	Jambon	15,892	59,277
Kahvalti	Aperitif	18,119	56,561
Kahvalti	Seker	20,242	55,689
Kahvalti	Peynir	23,752	55,196
Süt	Alkolsüz ürün	15,797	54,239
Peynir	Jambon	15,892	53,863
Kahvalti	Yogurt	29,191	53,581
Kahvalti	Konserve	23,958	53,521
Kahvalti	Sarkuteri	25,239	53,418
Alkolsüz ürün	Aperitif	18,119	53,215
Kahvalti	Jambon	15,892	53,175
Kahvalti	Donmus gida	16,076	53,094
Alkolsüz ürün	Yogurt	15,641	52,628
Kahvalti	Kahvalti		
Süt	Yogurt	15,641	52,275
Kahvalti	Kahvalti		
Yogurt	Alkolsüz ürün	15,797	52,108
Kahvalti	Parfum	17,34	52,083
Kahvalti	Süt	28,057	52,043
Sebze	Et	15,479	51,915
Sarkuteri	Peynir	23,752	51,598
Alkolsüz ürün	Süt	28,057	51,298
Kahvalti	Ilac	25,317	49,304
Sebze	Yogurt	15,641	48,682
Kahvalti	Kahvalti		
Peynir	Sarkuteri	25,239	48,559
Alkolsüz ürün	Alkolsüz ürün	23,958	48,311

Şekil 2.5 Birliktelik Kuralları Analizi Sonuçları



Tablo 2.2 BirlikteLik Kuralları Analizi Detaylı Sonuçları

Consequent	Antecedent	Support %	Confidence %
Sebze	Meyve	25,134	59,331
Sarkuteri	Jambon	15,892	59,277
Kahvaltı	Aperitif	18,119	56,561
Kahvaltı	Seker	20,242	55,689
Kahvaltı	Peynir	23,752	55,196
Süt	Alkolsüz ürün	15,797	54,239
	Kahvaltı		
Peynir	Jambon	15,892	53,863
Kahvaltı	Yogurt	29,191	53,581
Kahvaltı	Konserve	23,958	53,521
Kahvaltı	Sarkuteri	25,239	53,418
Alkolsüz ürün	Aperitif	18,119	53,215
Kahvaltı	Jambon	15,892	53,175
Kahvaltı	Donmus gıda	16,076	53,094
Alkolsüz ürün	Yogurt	15,641	52,628
	Kahvaltı		
Süt	Yogurt	15,641	52,275
	Kahvaltı		
Yogurt	Alkolsüz ürün	15,797	52,108
	Kahvaltı		

Eğer elde olan veride ürünler için sadece satın alındı/alınmadı bilgisi varsa birlikteLik kuralları analizinde ürünler arasındaki bağıntı, Destek (Support) ve Güven (Confidence) Oranları aracılığı ile hesaplanır.

Destek Oranı: $P(X \Rightarrow Y) = X$ ve Y mallarını satın almış müşteri sayısı / Toplam müşteri sayısı

Güven Oranı: $P(X \Rightarrow Y) = P(X \text{ ve } Y) / P(Y) = X$ ve Y mallarını satın almış müşteri sayısı / Y malını satın almış müşteri sayısı

Destek kriteri veride bu bağıntının ne kadar sık olduğunu, güven kriteri ise Y malını almış bir kişinin hangi olasılıkla X malını alacağını belirtmektedir. İki ürünün satın alınmasındaki bağıntının önemli olması için her iki değerin de olabildiğince yüksek olması gerekmektedir. [3]

Tablo 2.2'de elde edilen kurallar aşağıdaki gibi yorumlanmaktadır:

Birinci satıra bakıldığında;

“Meyve satın alan müşterilerin %59,331'i sebze de satın almaktadır. Bu birlikteLik tüm alışveriş kayıtlarının %25,134'ünde görülmektedir.”

kuralı elde edilmiştir.

**Altıncı satır:**

“Alkolsüz ürün ve kahvaltılık yiyecek satın alan müşterilerin %54,239’u süt de almaktadır. Bu birliktelik tüm alışveriş kayıtlarının %15,797’sinde görülmektedir.”

kuralını belirtmektedir.

ÖZET

Sepet analizinde alışverişler sırasındaki birlikteliklerin tespit edilmesi amaçlanmaktadır. Bu analiz temel amacı çapraz satışla ürün satışlarını arttırmaya yönelik olan tüm problemlere uygulanabilir. Süpermarket verisine uyguladığımız analiz; bir bankanın veya seyahat acentasının veri tabanına da uygulanabilir. Genel olarak bir üründen veya servisten fazlasını sunan her organizasyonda sepet analizi kullanılabilir.

Uyguladığımız analizdeki amaç en önemli alışveriş örüntülerini tespit etmektir. Burada kullanılan örüntü kavramı, birlikte satın alınan ürün gruplarını belirtmektedir. Analiz sonucunun değerlendirilmesinde Destek ve Güven oranları kullanılmaktadır.

Veri seti İspanya’da bulunan büyük bir perakende firmasına aittir. 2007 yılına ait 7 aylık dönem içerisinde, perakende zincirine ait süpermarketlerde yapılan her bir işlem kaydedilerek veri tabanı oluşturulmuştur. Veri setinde her satır bir müşteriyi göstermektedir. Sütunlarda ise ürün gruplarının satın alınıp alınmadığını belirten binary- ikili kodlamalar bulunmaktadır.

Analizde okuyucuların sonuçları kolaylıkla yorumlayabilmesi için web grafiği kullanılmıştır. Birliktelik kuralları ile elde edilen sonuçlar ise kural seti olarak ifade edilmiştir.

Analiz sonucu elde edilen kurallar oldukça anlamlı görülmektedir. Süpermarket yöneticileri bu kuralları dikkate alarak, raf düzenlemeleri yapabilirler. Ayrıca market yöneticileri, satışını arttırmak istedikleri ürünler ile ilgili promosyonlar düzenleyebilirler. Ürünlerin birbirine yakın yerlere konulması, yapılan promosyonlar ve müşteriye dağıtılan kuponlar satışları arttırmayı sağlayacaktır.

YARARLANILAN KAYNAKLAR

- [1] Berry Michael J.A., Linoff Gordon, **The Art and Science of Customer Relationship Management**, Wiley Computer Publishing, U.S.A., 2000.
- [2] Bounsaythip Catherine, Rinta-Runsala Esa, **Overview of Data Mining for Customer Behaviour Modeling**, VTT Information Technology Research Report TTEI, 2001.
- [3] Giudici Paolo, **Applied Data Mining: Statistical Methods for Business and Industry**, John Wiley & Sons Ltd., England, 2003.