

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

**MÜŞTERİ İLİŞKİLERİ YÖNETİMİNDE İŞ ZEKASI ve
VERİ MADENCİLİĞİ YÖNTEMLERİ**

**YÜKSEK LİSANS TEZİ
Handeris Seldan ÇERKEZ**

Anabilim Dalı : ENDÜSTRİ MÜHENDİSLİĞİ

Programı : ENDÜSTRİ MÜHENDİSLİĞİ

OCAK 2003

**MÜŞTERİ İLİŞKİLERİ YÖNETİMİNDE İŞ ZEKASI ve
VERİ MADENCİLİĞİ YÖNTEMLERİ**

**YÜKSEK LİSANS TEZİ
Handeris Seldan ÇERKEZ
(507971105)**

**Tezin Enstitüye Verildiği Tarih : 24 Aralık 2002
Tezin Savunulduğu Tarih : 16 Ocak 2003**

**Tez Danışmanı : Doç. Dr. Alpaslan FIĞLALI
Diğer Jüri Üyeleri: Doç. Dr. Ziya ULUKAN
Y.Doç.Dr. Ahmet BEŞKESE**

OCAK 2003

ÖNSÖZ

Müşteri İlişkileri Yönetiminde İş Zekası ve Veri Madenciliği Yöntemleri konusunda araştırma olanağı sağlayan, bu çalışma sırasında göstermiş olduğu ilgi ve destek için tez danışmanım Doç. Dr. Alpaslan FIĞLALI' ya teşekkürü bir borç bilirim.

Aralık 2002

Handeris Seldan ÇERKEZ

İÇİNDEKİLER

TABLO LİSTESİ	vi
ŞEKİL LİSTESİ	vii
ÖZET	viii
SUMMARY	x
1. GİRİŞ	1
1.1. Giriş ve Çalışmanın Amacı	1
2. İŞ ZEKASI	3
2.1. İş Zekası Kavramı	3
2.2. İş Zekası Faydaları	4
2.2.1. Maliyet Tasarrufu	4
2.2.2. Gelir Sağlar	4
2.2.3. Müşteri Memnuniyetinin İyileştirilmesi	5
2.2.4. Kurum İçi İletişiminin İyileştirilmesi	5
2.3. İş Zekası Uygulamaları	5
2.3.1. Kampanya Analizi (Campaign Management)	5
2.3.2. Müşteri Profili Analizi (Customer Profile Analysis)	5
2.3.3. Müşteri Bağlılık Analizi (Loyalty Analysis)	6
2.3.4. Satış Analizi (Sales Analysis)	6
2.3.5. Müşteri İlişkileri Analizi (Customer Care Analysis)	6
2.3.6. İş Performansı Analizi (Business Performance Analysis)	6
2.3.7. Karlılık Analizi (Profitability Analysis)	6
3. VERİNİN YARARLI BİLGİYE DÖNÜŞÜM SÜRECİ	8
3.1. Data (Veri)	8
3.2. Enformasyon (Information)	9
3.3. Analitik (Analytic)	9
3.4. Bilgi (Knowledge)	10
3.5. Yararlı Bilgi (Wisdom)	10
3.6. Verinin Anlam Düzeylerine Göre Kullanılan Teknolojiler	11
4. VERİ MADENCİLİĞİ	13
4.1. Veri Madenciliği Kavramı	13
4.2. Veri Madenciliğin Önemini Arttıran Nedenler	14
4.2.1. Veri Üretimi	14
4.2.2. Verinin Depolanması	15
4.2.3. Bilişimsel Gücün Yeterliliği	15
4.2.4. Artan Rekabet Baskısı	15
4.3. Veri Madenciliği Modelleri	16
4.3.1. Sınıflandırma (Classification)	16
4.3.2. Tahmin (Estimation)	17

4.3.3. Öngörme (Prediction)	17
4.3.4. Birliklilik Kuralları (Association Rules) ve Ardışık Zamanlı Örüntüler (Sequential Patterns)	18
4.3.5. Kümeleme (Clustering)	18
4.3.6. Tanımlama (Description)	19
4.4. Veri Madenciliği Türleri	19
5. VERİ MADENCİLİĞİN ETKİLEŞİMLİ ÇEVİRİMİ	21
5.1. Veri Madenciliği Çevrimi ile Müşteri Kaybı Probleminin Çözümlemesi	23
5.2. Veri Madenciliğin Etkileşimli Çevrimi	25
5.2.1. Problemin Tanımlanması	26
5.2.2. Veri Madenciliği Yöntemleri ile Verinin Anlamlı Enformasyon' a Dönüştürülmesi	27
5.2.3. Enformasyonun Aksiyona Çevrilmesi	28
5.2.4. Sonuçların Ölçülmesi	28
5.3. Modelin Değerlendirilmesinde Kullanılan Yöntemler	30
5.3.1. Basit Geçerlilik Testi	30
5.3.2. Çapraz Geçerlilik Testi	31
5.3.3. N-Katlı Çapraz Geçerlilik Testi	31
6. PAZAR SEPETİ ANALİZİ (Market Basket Analysis)	32
6.1. Pazar Sepeti Analizinin Kuvvetli Yönleri (Strengths)	34
6.2. Pazar Sepeti Analizinin Zayıf Yönleri (Weaknesses)	35
7. BELLEĞE DAYALI MUHAKEME (Memory-Based Reasoning)	36
7.1. Uzaklık Fonksiyonu	36
7.2. Belleğe Dayalı Muhakeme Yönteminin Kuvvetli Yönleri (Strengths)	40
7.2. Belleğe Dayalı Muhakeme Yönteminin Zayıf Yönleri (Weaknesses)	40
8. YAPAY SİNİR AĞLARI (Artificial Neural Networks)	41
8.1. Bir Gayrimenkul Değerinin Hesaplanması Örneği	42
8.2. Doğrudan Veri Madenciliği (Directed Data Mining) için Sinir Ağları	46
8.3. Sinir Ağı Tanımı	47
8.4. Sinir Ağı Biriminin Tanımı	48
8.5. İleri Beslemeli Sinir Ağları	51
8.6. Geri Yayım İle Gerçekleşen Öğrenme Prosesi	53
8.7. Genetik Algoritmaların Kullanımı İle Öğrenme	55
8.8. Öğrenim Kümesinin Seçimi	56
8.8.1. Öğrenim Kümesinin Değer İçeriği	57
8.8.2. Özelliklerin Sayısı	57
8.8.3. Girdilerin Sayısı	57
8.8.4. Çıktıların Sayısı	58
8.9. Verinin Hazırlanması	58
8.9.1. Sürekli (Sabit ve Kayar Noktalı) Değerli Özellikler	59
8.9.2. Sıralı, ayırık (Tamsayı) Değerli Özellikler	61
8.9.3. Kesin Değerli (Kategorik) Özellikler	63
8.10. Sonuçların Yorumlanması	64
8.11. Zaman Serileri İçin Sinir Ağları	67
8.12. İleri Beslemeli, Geri Yayınımlı Ağların Kullanımı İle İlgili Bulgular	69
8.13. Sinir Ağına Anlaşılması	70
8.14. Dolaylı Veri Madenciliğinde Sinir Ağlarının Kullanımı	71

8.14.1. Kendini Düzenleyen Haritalar (SOM)	72
8.14.2. Kendini Düzenleyen Haritalar İle Kümelerin Bulunması Örneği	74
8.15. Yapay Sinir Ağlarının Kuvvetli Yönleri (Strengths)	77
8.16. Yapay Sinir Ağlarının Zayıf Yönleri (Weaknesses)	77
8.17. Sinir Ağları'nın Uygulanabildiği Durumlar	78
9. MÜŞTERİ İLİŞKİLERİ YÖNETİMİ (CRM)	79
9.1. Müşteri (Customer)	79
9.2. İlişki (Relationship)	79
9.3. Yönetim (Management)	80
9.4. İrtibat Merkezleri (Contact Center)	80
9.5. Başlıca İş Akımları	81
9.6. Başarılı CRM Kavramı	82
9.7. Başarılı CRM İçin İş Analizi	84
10. BİREYSEL KREDİ TALEPLERİNİN DEĞERLENDİRİLMESİNDE BİR UYGULAMA ÖNERİSİ	88
10.1. Giriş	88
10.2. Problemin Tanımlanması	90
10.2.1. Proje Konusu : Müşteri Risk Değerleme Sistemi	90
10.2.2. Projenin Amacı	91
10.2.3. Projenin Kapsamı ve İçeriği	91
10.2.4. Kredi Değerlendirme Süreci	92
10.2.5. Kredi Skor Kartı	92
10.2.6. Sistem Gerekleri	95
10.2.7. Davranışsal Risk Kıstasları	96
10.2.8. Proje Planı	97
10.3. Verilerin Hazırlanması	97
10.3.1. Toplama (Collection)	97
10.3.2. Değer Biçme (Assessment)	98
10.3.3. Birleştirme ve Temizleme (Consolidation and Cleaning)	98
10.3.4. Seçim (Selection)	98
10.3.5. Dönüştürme (Transformation)	99
10.4. Modelin Kurulması ve Değerlendirilmesi	99
10.4.1. Müşteri Risk Değerleme Sistemi Tasarımı	100
10.4.1. Sinir Ağı Yöntemi İle Oluşturulacak Modelin Kurulumu	104
10.4.1. Kurulan Modelin Doğruluk Derecesinin Değerlendirilmesi	104
10.5. Modelin Kullanılması	106
10.6. Modelin İzlenmesi	106
10.7. Sonuç	106
11. SONUÇ	109
KAYNAKLAR	110
ÖZGEÇMİŞ	112

KISALTMALAR

TABLO LİSTESİ

	Sayfa No
Tablo 3.1 Veri İşlem Örneği.....	9
Tablo 3.2 Verilerin Enformasyona Dönüşecek Şekilde Toplanması.....	9
Tablo 3.3 Enformasyonun Analiz İçin Ayırıştırılması.....	10
Tablo 3.4 Bir Satın Alma Örüntüsünün Teşhis Edilmesi.....	11
Tablo 3.5 Verilerin Anlam Düzeylerine Karşılık Gelen Teknolojilerin Sınıflandırılması.....	11
Tablo 7.1 Örnek Bilgiler 1.....	38
Tablo 7.2 Örnek Bilgiler 2.....	38
Tablo 7.3 Örnek Bilgiler 3.....	38
Tablo 7.4 Örnek Bilgiler 4.....	39
Tablo 7.5 Örnek Bilgiler 5.....	39
Tablo 7.6 Örnek Bilgiler 6.....	40
Tablo 8.1 Örnek Bilgiler 1.....	43
Tablo 8.2 Öğrenme Kümesi Örneği.....	44
Tablo 8.2 İndirgenmiş Öğrenim Kümesi Örneği.....	45
Tablo 8.3 Farklı Yöntemlerle İndirgenmiş Gelir Değerleri.....	61
Tablo 8.4 Kategorik Değerlerin İndirgenme Örneği.....	63
Tablo 8.5 Kategorik Değer Çıktılarının Güven Payları.....	65
Tablo 8.6 Zaman Serisi Örneği.....	68
Tablo 8.7 Geçmişe Dönük Üç Zamanlı Bir Zaman Serisi Örneği.....	69
Tablo 9.1 İrtibat Merkezinin Evrimi.....	81
Tablo 9.2 Başlıca İş Akımlarının Evrimi.....	82
Tablo 9.3 CRM tutum ve yaklaşım farklılıkları.....	83
Tablo 9.4 CRM Uygulamasının Temel Aşamaları.....	84
Tablo 10.1 Kredi Notu Değerleri.....	101
Tablo 10.2 Başvuru Skor Modelinin Girdileri.....	101
Tablo 10.3 Risk Notu Değerleri.....	102
Tablo 10.4 Parametrik Koşullar 1.....	103
Tablo 10.5 Parametrik Koşullar 2.....	103
Tablo 10.6 Risk Matrisi.....	105
Tablo 11.1 Bilgi Yönetimi ile Bilgi Ekolojisi Akımlarının Kıyaslanması...	109

ŞEKİL LİSTESİ

	Sayfa No
Şekil 3.1 Veri Dönüşüm Süreci Aşamaları	8
Şekil 4.1 Kümeleme Modeli	19
Şekil 7.1 MBR Benzerlik Kavramı (En yakın iki nokta arasındaki uzaklık).....	37
Şekil 8.1 Sinir Ağı Modeli.....	43
Şekil 8.2 İleri Beslemeli Sinir Ağı Örnekleri.....	48
Şekil 8.3 Sinir Ağı Birimi.....	49
Şekil 8.4 Transfer Fonksiyonları.....	50
Şekil 8.5 Sinir Ağının Öğrenme Prosesi.....	52
Şekil 8.6 Birden Fazla Çıktı Üreten Sinir Ağı Örneği.....	53
Şekil 8.7 Ağırlıkların Genetik Kromozomlarda Toplanması.....	56
Şekil 8.8 Gelir Dağılımı Örneği.....	60
Şekil 8.9 Test Kümesi Üzerinde Uygulanan Bir Sinir Ağının Sonuçları..	65
Şekil 8.10 Sinir Ağlarının Zaman Serilerine Uyarlanması.....	67
Şekil 8.11 Kendini Düzenleyen Haritalar (Self-Organizing Maps).....	72
Şekil 8.12 SOMs ile Girdiyi En İyi Tanıyabilecek Çıktının Belirlenmesi..	74
Şekil 8.13 SOM ile Tanımlanan Kümelerin Kıyaslanması.....	76
Şekil 10.1 Bilgi Keşfi Süreci.....	88
Şekil 10.2 Veri Ambarının Temel Bileşenleri.....	89
Şekil 10.3 Müşteri Merkezli Veri Ambarı Tasarımı.....	90
Şekil 10.4 Modelin Oluşumundaki Veri Madenciliği Prosesi.....	100
Şekil 10.5 Kredi Taleplerini Değerlendirme Süreci İş Akış Şeması.....	107
Şekil 10.6 Kredi Taleplerini Otomatik Değerlendirme Süreci İş Akış Şeması.....	108

MÜŞTERİ İLİŞKİLERİ YÖNETİMİNDE İŞ ZEKASI ve VERİ MADENCİLİĞİ YÖNTEMLERİ

ÖZET

İş zekası temelde, organizasyonların operasyonel verilerini yüksek değerde enformasyon içeren bir ambara (veri ambarı) dönüştürmek ve *doğru* enformasyonun, *doğru* şekilde, *doğru* kişiye, *doğru* zamanda dağıtmak ile ilgilidir.

İş zekasının değer zinciri, veri kaynağı ile başlar. Zeki, öğrenen bir organizasyonun bilgi çevresini destekleyecek enformasyon veri kaynağından gelişir. Veri enformasyonun ham maddesi, enformasyon da bilginin ham maddesidir. Bilgi ise, iş stratejilerini destekleyen iş zekasının ham maddesidir.

Son yıllarda şirketlerin müşterileri ile olan etkileşim yöntemleri hızla değişmiştir. Bir müşteri ile uzun süreli iş yapma garantisi yoktur. Sonuç olarak, şirketler müşterilerini daha iyi tanımaları, onların talep ve ihtiyaçlarını hızlı cevaplamaları gerektiğinin farkına varmışlardır.

Müşteri ve aday müşterilerin iletişim beklentileri, şirketin iş yapma biçimini çoklu kriterlere göre değerlendirmesini ve doğru kanaldan, doğru zamanda, doğru kişiye, doğru teklifin sunulmasını otomatik hale getirmesini gerektirir.

Müşteri ilişkileri yönetimi şirketle müşterileri arasındaki etkileşimleri yöneten bir süreçtir, veri madenciliği de bu amaçla gerekli değerli enformasyonu sağlayan güçlü yöntemler kümesidir.

Veri madenciliği, veri tabanındaki girdilerle müşteri davranışını tahmin eden modeller kurmaktadır. Bir model tarafından yapılan tahmin genellikle skor olarak adlandırılmaktadır. Veri tabanındaki her bir kayda bir skor atanır ve bu değer ilgili müşterinin özel bir davranışı sergileme olasılığını belirtir.

Müşteri kümelemesi ve segmentasyonu, veri madenciliğin pazarlama ve müşteri ilişkileri yönetiminde kullanılan, en önemli iki yöntemidir. Müşterinin davranışını

incelemek ve stratejik girişimlerde bulunabilmek için, işlem verilerinden faydalanılmaktadır. Bu veri kullanılarak müşterilerin, karlılıklarına ve bazı risk faktörlerine göre segmentlere ayrılması, müşterinin ömür değerini ve terk etme olasılığını ölçme imkanı verir, ayrıca pazarlama fırsatlarının farkına varılmasını sağlar.

Veri madenciliği, işe rekabet üstünlüğü sağlayacak yararlı enformasyon için, yüksek miktardaki ham veriyi filtre etmektedir. Bu enformasyon, önceden veri içerisinde varolan ancak fark edilmeyen anlamlı örüntü ve trendlerden oluşmaktadır.

Veri madenciliği, şirketin daha iyi kararlar almasını sağlayacak, yeni ve yararlı enformasyonun kazanımı ile sonuçlanmalı.

Veri madenciliği prosesi aslında hedeflenmiş bir konu ile başlar. Projenin yönetilebilirliği açısından, hedeflenen birer konu ile sınırlandırılmalı ve her bir konu için proses tekrarlanmalı. Şirket, veri madenciliğinden ne beklediğini veya hangi iş problemini çözmek istediğini önceden tanımlarsa, daha başarılı sonuçlar elde edilir.

BUSINESS INTELLIGENCE and DATA MINING TECHNIQUES for CUSTOMER RELATIONSHIP MANAGEMENT

SUMMARY

Business intelligence is fundamentally concerned with transforming organization's operational data into an accessible store of high-value information (called a *data warehouse*) and distributing the *right* information in the *right* way to the *right* people at the *right* time.

Business intelligence value chain begins with the data resource. Information is developed from the data resource to support the knowledge environment of an intelligent learning organization. Data is the raw material for information which is the raw material for the knowledge environment. Knowledge is the raw material for business intelligence that supports business strategies.

The way in which companies interact with their customers has changed dramatically over the past few years. A customer's continuing business is no longer guaranteed. As a result, companies have found that they need to understand their customers better, and to quickly respond to their wants and needs.

Customers and prospective customers want to interact on their terms, meaning that companies need to look at multiple criteria when evaluating how to proceed. They will need to automate the *right offer* to the *right person* at the *right time* through the *right channel*.

Customer relationship management is a process that manages the interactions between a company and its customers and data mining is a powerful set of techniques which provide valuable information for this purpose.

Data mining builds models by using inputs from a database to predict customer behavior. The prediction provided by a model is usually called a score. A score is

assigned to each record in the database and indicates the likelihood that the customer whose record has been scored will exhibit a particular behavior.

Customer clustering and segmentation are two of the most important data mining methodologies used in marketing and customer relationship management. They use customer-purchase transaction data to track buying behavior and create strategic business initiatives. Businesses can use this data to divide customers into segments based on such "shareholder value" variables as current customer profitability, some measure of risk, a measure of the lifetime value of a customer, and retention probability. Creating customer segments based on such variables highlights obvious marketing opportunities.

Data mining is simply filtering through large amounts of raw data for useful information that gives businesses a competitive edge. This information is made up of meaningful patterns and trends that are already in the data but were previously unseen.

The end result of data mining should be the acquisition of new and useful information that can help a company make better decisions that improve business.

The overall mining process actually begins with a targeted problem. To keep the project manageable, the business should narrow the scope of the mining process to a single issue, such as increasing repeat business. Data mining is more successful when the company first decides what it wants to get out of the mining or what business problem it wants to solve.

1. GİRİŞ

1.1. Giriş ve Çalışmanın Amacı

Organizasyonlar canlı organizmalara benzer. Bilgiyle beslenir, iletişim ağları üzerinden ilişki kurar, topladığı verileri değerlendirerek bir sonraki adımını planlarlar. Özetle, ham veri işlenerek stratejiye dönüştürülür. Ancak iş verilerinin hızlı artışı karşısında, personel sayısının aynı kalması veya azalması, verinin bilgiye dönüşüm hızını gündeme getirmiştir.

1970'lere gelene kadar üretici ve satıcıların açısından bakıldığında, seri üretimi yapılan birçok ürünün satışında hiç bir sorun yoktu. Genelde ne üretilse satılabilir düşüncesi hakimdi. Daha sonraları ise, gelişen teknoloji sayesinde üretimin ucuzlaması ve bollaşması, malların ve hizmetlerin kolay taklit edilebilir bir hale gelmesi ve ucuzlaması, rekabeti daha farklı bir boyuta taşıdı. Mevcut müşteriye elde tutmak, onu tanıyarak ihtiyaçlarını zamanında karşılamak ve işletme açısından karlılığını arttırmak önem kazanmıştır. Bu doğrultuda gelişen iş modellerin uygulanması, onları destekleyecek teknolojilerin kullanılması ve personelin yeterince eğitilmesi sayesinde, duruma özgü karar verme imkanı sağlanmış ve böylece güçlü bir rekabet avantajı elde edilmiş olunur.

Bu tez çalışmasının amacı, müşteri ilişkileri yönetiminde iş zekası kavramını ve veri madenciliği yöntemlerini incelemektir. Yapılan çalışma Giriş, İş Zekası Kavramı, Verinin Yararlı Bilgiye Dönüşüm Süreci, Veri Madenciliği Kavramı, Veri Madenciliğin Etkileşimli Çevrimi, Pazar Sepeti Analizi, Belleğe Dayalı Muhakeme, Yapay Sinir Ağları, Müşteri İlişkileri Yönetimi, Bireysel Kredi Taleplerinin Değerlendirilmesinde Bir Uygulama Önerisi ve Sonuç olmak üzere on bir bölümden oluşmaktadır.

İkinci bölümde, iş zekası kavramı; maliyet, gelir, müşteri memnuniyeti ve kurum içi iletişim açısından faydaları ve başlıca uygulamaları anlatılmaktadır.

Üçüncü bölümde, verinin dönüşüm süreci ve bu sürecin tüm aşamaları örneklerle açıklanmakta, verinin her anlam düzeyinde kullanılan teknolojiler belirtilmektedir.

Veri madenciliği kavramına; işletme açısından önemini arttıran nedenlere; sınıflandırma, tahmin, öngörü, kümeleme ve tanımlama gibi veri madenciliğın başlıca modellerine, dördüncü bölümde değinilmektedir.

Beşinci bölümde, veri madenciliğın etkileşimli çevrimi, müşteri kaybı probleminin çözümlenmesi örneğı ile anlatılmakta, oluşturulan modelin değerlendirilmesinde kullanılan başlıca yöntemler belirtilmektedir.

Altıncı ve yedinci bölümlerde, veri madenciliğın, sırasıyla Pazar Sepeti Analizi ve Belleğe Dayalı Muhakeme yöntemleri örneklerle incelenmekte, yöntemlerin kuvvetli ve zayıf yönleri belirtilmektedir.

Veri madenciliğın güçlü ve yaygın kullanımı olan Sinir Ağları yöntemi, sekizinci bölümde, bir gayrimenkul değerin hesaplanması örneğı yardımıyla anlatılmaktadır. Sinir ağının ve ağı oluşturan birimlerin tanımı yapılmakta, ileri beslemeli sinir ağının topolojisi açıklanmaktadır. Modelin kurulmasındaki öğrenme prosesi; bu prosesteki geri yayım yönteminin ve genetik algoritmaların kullanımı; modelin oluşumu sırasında, öğrenim kümesinin seçimi, verilerin hazırlanması ve sonuçların yorumlanması konusunda dikkat edilmesi gereken hususlar, yine bu bölümde anlatılmaktadır. Zaman serileri için sinir ağların kullanımı incelenmekte, kendini düzenleyen haritalar (SOM) olarak tanımlanan özel bir sinir ağı, bir örnekle değerlendirilmektedir. Son olarak sinir ağı yönteminin kuvvetli ve zayıf yönleri irdelenmekte ve uygulanabildiğı durumlar belirtilmektedir.

Dokuzuncu bölümde, müşteri ilişkileri yönetimini tanımlayan, müşteri, ilişki ve yönetim bileşenleri tanımlanmakta; bu akımla birlikte irtibat merkezlerinin gelişim aşamaları irdelenmekte; başlıca iş akımların evrimi ve bunların odaklanılan alan ile beklenen sonuç açısından farkları anlatılmakta; başarılı bir müşteri ilişkileri yönetiminin tanımı yapılmakta ve başarılı bir CRM için yapılması gereken iş analizi incelenmektedir.

Onuncu bölümde, bireysel kredi taleplerinin değerlendirilmesi için uygulama önerisinde bulunmaktadır. Müşterinin başvuru ve davranışsal bilgilerinden oluşturulan kredi skor kartına bağılı olarak müşteriye verilebilecek limit belirlenir, böylece şirketin minimum risk ile maksimum kar elde etmesi sağlanır.

On birinci ve son bölüm yapılan inceleme ile varılan sonuçların tartışıldığı bölümdür.

2. İŞ ZEKASI

2.1. İş Zekası Kavramı

Bir organizasyonun her departmanında, her gün yüzlerce kararlar alınmaktadır. Bu kararların çoğu tecrübeye, biriken bilgiye ve usullere dayanır. Tecrübenin, bilginin ve usullerin geliştirilmesi uzun yıllar alır. İş kararlarının kalitesinin iyileştirilmesi, doğrudan maliyet ve kazançta etki eder. Örneğin, bir müşteriye indirim yapmak, ilişkisi süresince, karlılığına bağlı olarak sonucu etkiler veya etkilemez.

Firmalar iki temel hususu büyüteç altına almalıdır.

1. Kararların iyileştirme yöntemleri

- Yöneticiler tüm kritik görevleri üstlenir.
- Yetenekli personel sayısı artırılır.
- Olası tüm kararlar için genel bir yaklaşım oluşturulmaya çalışılır.
- Mevcut personele, daha iyi kararlar alabilmeleri için gerekli araç ve imkanlar sağlanır.

2. İyi bir kararın tanımı

En iyi karar, kuruluşun hedeflerine en yakın karardır. Bir çalışanın, alacağı kararın kurumun amacına hizmet edip edemeyeceğini belirlemek için stratejik bilgiye ihtiyaç vardır.

Örneğin, bir duvar ustasının yaptığı iş çok değişik şekilde tanımlanabilir :

- Taşları yan yana dizerek onları çimento ile birleştirir
- Bir duvar örür
- Bir katedral inşa eder

Üç tanım da doğru, ancak duvar ustası üçüncü tanıma yoğunlaşır ise işinin hedefini daha iyi anlayacak ve yaptığı işi daha ilginç bulacaktır. Yapının diğer kısımlarının nasıl ilerlediği, ne kadar taşın tamamlandığını, geçmişte diğer duvar ustalarının işlerini

ne şekilde yaptıkları gibi bilgilerle donatılan duvar ustasının değişen koşullara daha hızlı cevap verebilmesi sağlanır.

2.2. İş Zekası Faydaları

İş Zekasının sağladığı faydalar dört grupta toplanabilir.

2.2.1. Maliyet tasarrufu

- Self-servis şeklinde bilgiye ulaşımı sağlayarak operasyonel verimlilik arttırılabilir.
- Raporlamanın ihmal edilmesi ve geciktirilmesi önlenebilir.
- Müşteri ve tedarikçilerle daha iyi ilişkiler kurulabilir.
- Asıl sebepler bulunup, çözümlenebilir.



Sonuç : Teslimat probleminin çözülmesi.

2.2.2. Gelir sağlar

- Teklifler farklılaştırılabilir.
- Daha doğru pazar analizleri yapılarak stratejiler iyileştirilebilir.
- Satış gücü kuvvetlendirilebilir.

2.2.3. Müşteri memnuniyetinin iyileştirilmesi

- Kullanıcı isteklerine daha hızlı cevap verilebilmesi

- Kullanıcıların daha iyi kararlar almalarının sağlanması

2.2.4. Kurum içi iletişiminin iyileştirilmesi

- Ortak bir dilin oluşturulması
- Sorumluluğun ve verimliliğin artırılması
- Müşterilerin bütün bir resminin çıkarılabilmesi
- Yaratıcılığın teşvik edilmesi

2.3. İş Zekası Uygulamaları

2.3.1. Kampanya Analizi (Campaign Management)

Kampanya ve promosyonlarda doğru müşterileri hedeflemek ve onların geri dönüşünü analiz etmek, kitlesel pazarlamadan (Mass Marketing) kitle farklılaştırmasına (Mass Customization) geçişte kritik rol oynamaktadır. Kuruluşlar farklı kitle iletişim araçları aracılığıyla, çeşitli ürünler için farklı türden kampanyalar düzenlemektedir.

Bu uygulama, hedef kitle seçiminden, müşterilerinin cevaplama şekline kadar tüm sürecinin kuruluşlar tarafından daha iyi anlaşılmasını mümkün kılar.

Kampanya Analizi, kampanyaların aile ve bireysel bazda cevaplanmasının ölçülebilirliğini sağlar. Bireysel kampanyalarının ve farklı iletişim araçlarının etkinliğini ölçerek, kampanyanın maliyet-fayda analizini yapma imkanı verir.

2.3.2. Müşteri Profili Analizi (Customer Profile Analysis)

Müşteri Profili Analizi, bir kuruluşun müşteri kitlesinin bütünü oluşturarak, en küçük kümelerinin ayırt edilmesini (mikrosegmentasyon) sağlar. Müşteri segmentasyonu (Customer segmentation), özellikle müşteri segmentlerinin farklı ve belirgin özellikler sergiledikleri pazarlarda, pazarlama stratejisinin temelini oluşturmaktadır. Müşteri profilinin belirlenmesi ve segmentlere ayrılması, bire bir pazarlama çağında müşteri ilişkilerinin kurulmasını kolaylaştırmaktadır.

2.3.3. Müşteri Bağlılık Analizi (Loyalty Analysis)

Karlılığın anahtarlarından biri müşteri sadakatidir. Günümüz rekabet koşullarında kritik bir faktör olmasına rağmen henüz çok az sayıda kuruluş, müşteri sadakatini düzenli olarak ölçmekte veya müşteri kaybının nedenlerini araştırmaktadır.

Müşteri bağlılık analizi, müşteri sadakatinin, ilişkinin sürekliliği, alınan ürün ve hizmet dizisi, terk etme eğiliminin demografik ve coğrafi etkenleri gibi, farklı açılardan ölçülmesini sağlar ve müşteriye elde tutma programlarının geliştirilmesini kolaylaştırır.

2.3.4. Satış Analizi (Sales Analysis)

Satış analizi uygulamaları, kanal, satış noktası veya birim bazında; ürün, ürün kategorisi veya grubu bazında; bölge ve dönem bazında olmak üzere, satışların farklı boyutlara göre analizini sağlar.

Bu uygulamalar, satış yöneticilerine, sonuçların bütününe görerek satış verilerinden ortaya çıkan eğilim (trend) ve örüntüleri (patterns) belirleme imkanı verir.

2.3.5. Müşteri İlişkileri Analizi (Customer Care Analysis)

Müşteriler ile kuruluşlar devamlı etkileşim içindeler: bilgi alırlar, geri bildirimde bulunurlar veya öneri sunarlar. Bu enformasyon, müşterilerin davranışları hakkında ve kuruluşların hizmet performansı konusunda değerli ipuçları verir.

Bir müşterinin memnuniyeti veya memnuniyetsizliği, irtibat tarihçesinden (Contact History) tespit edilebilir.

Müşteri irtibatların analizi, müşteri ilişkilerin sürdürülmesinde ve beslenmesindeki temel bileşenlerden biridir ve müşteri sadakatinin korunmasını sağlar.

2.3.6. İş Performansı Analizi (Business Performance Analysis)

İş performansı analizi, sunulan farklı ürünlere karşı gösterilen ilgi de dahil olmak üzere tüm müşteri davranışlarının anlaşılmasını gerektirir.

2.3.7. Karlılık Analizi (Profitability Analysis)

Her kuruluşta, fiyatların belirlenmesi, gerekli kaynağın ayrılması veya stratejinin geliştirilmesi için karlılığın ölçülmesi esastır.

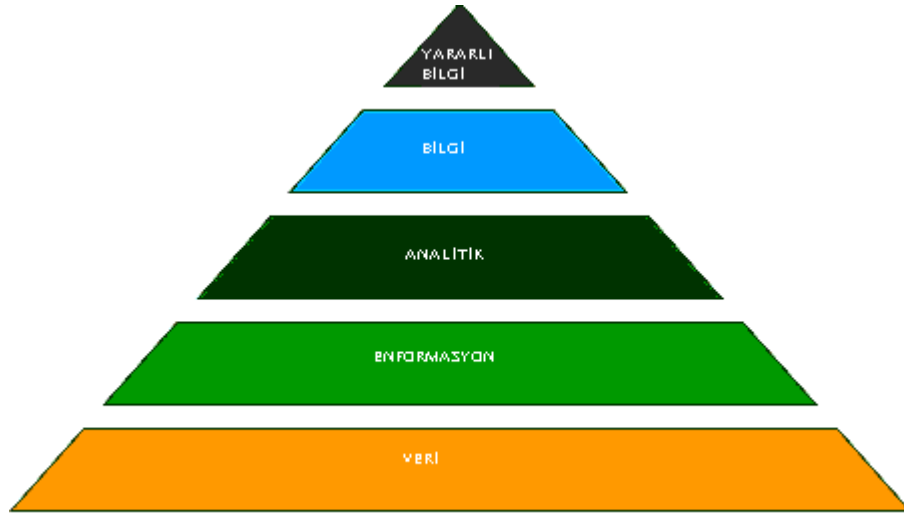
Karlılık çok geniş bir kavram ve kurum, kanal, ürün, ürün grubu, marka, müşteri veya bir müşteri segmenti bazında değerlendirilebilir. Birçok kuruluş ayrıca brüt karı, net karı ve marjı ölçmek isteyecektir.

Hemen her sektörde, müşteri karlılığının ölçülmesi, işin etkili yönetimi açısından önem arz etmektedir. Yapılan işin türü ve hacmi, satın alınan ürün ve hizmet çeşitliliği ve müşterinin otomatik işlem olanaklarından faydalanabilirliği gibi birçok faktör müşteri karlılığını etkilemektedir.

3. VERİNİN YARARLI BİLGİYE DÖNÜŞÜM SÜRECİ

Bir organizasyonun enformasyon sistemlerinde ve veritabanlarında birikmiş verilerde muazzam fırsatlar gizlidir. Bu verileri bilgiye dönüştürebilen organizasyonlar, trendleri belirleyebilir, daha önce fark edilmemiş örüntü ve ilişkileri keşfedebilir ve yeni fırsatlar oluşturarak en önemli rekabet avantajını yaratmış olurlar.

Verinin yararlı bilgiye dönüşüm süreci birkaç aşamadan oluşur. Aşağıdaki şekilde verinin dönüşüm sürecindeki aşamaları resmedilmiştir.



Şekil 3.1 Veri Dönüşüm Süreci Aşamaları

(Kaynak : DM Review)

3.1. Data (Veri)

Veritabanı yönetim sistemlerinin keşfedilmesi ve veri saklama teknolojilerindeki ilerleme ile birlikte organizasyonlar, kolay analiz edilebilecek kişi, yer, işlem, kavram ve olaylar ile ilgili muazzam miktarlarda veri toplamaya, işlemeye ve saklamaya başlamışlardır. Verinin büyük bir kısmı organizasyonun fonksiyonel prosesleri ile ilişkilidir.[12] Örneğin, bir market, müşterisinin satın almış olduğu ürünleri kasadan geçirerek, ilgili verileri kaydetmektedir. Kasada ürünler sisteme

tanıtıldığında, sistem tarafından her bir ürünün fiyatı belirlenir ve ödenecek toplam tutar hesaplanır. Bu işlemle, ürün, miktar, fiyat, tarih, kayıt numarası, satışı gerçekleştiren personel ve müşterinin mağaza kartı bulunuyorsa, kart bilgisi verileri sistem tarafından kaydedilir. Aşağıdaki tabloda bir işleme ait örnek veriler gösterilmektedir.

Tablo 3.1 Veri İşlem Örneği

Ürün	Miktar	Fiyat	Tarih	Kayıt #	Kullanıcı ID	Club Card ID
Çocuk bezi	1	4.99	11/1/00	001	213	1209

3.2. Enformasyon (Information)

İşlem sayıları arttıkça toplanan veri miktarı da hızla artmaktadır. Her bir veri elemanı, işlemin bir bileşeni olduğundan, tek başına bir anlam ifade etmez. Verinin anlamlı şekilde biriktirilmesi enformasyonu sağlar. İş zekası uygulamalarının, sorgu ve raporlama yetenekleri sayesinde, veritabanındaki verinin çekilerek enformasyona dönüşümü sağlanır. [12] Örneğin, ürün, miktar ve fiyat toplamaları, satılan ürünlerle bunların miktar ve hacimleri enformasyonunu sağlar. Aşağıdaki tabloda veriler, enformasyona dönüştürülecek şekilde özetlenmiştir.

Tablo 3.2 Verilerin Enformasyona Dönüşecek Şekilde Toplanması

Ürün	Miktar	Fiyat	Satış Tutarı
Bira	265	6.85	1,815.25
Hububat	430	3.90	1,677.00
Ekmek	850	1.59	1,351.50
Süt	1100	1.20	1,320.00
Çocuk bezi	200	4.99	998.00

3.3. Analitik (Analytic)

Enformasyonu yaratırken veri ile anlamını birleştirmek çok faydalı iken, enformasyonu ayrıştırıp veya yeniden gruplamak onun değerini artırır. OLAP (online analytical processing) teknolojisini içeren uygulamalar sayesinde, ilişki,

örüntü, eğilim ve istisnaların belirlenebilmesi için enformasyonun analiz edilmesi mümkün.[12] Market örneğinde, bir önceki adımda elde edilen enformasyon, aşağıdaki şekilde dönemlere ayrıştırılarak analiz edilebilir.

Tablo 3.3 Enformasyonun Analiz İçin Ayrıştırılması

Ürün	Dönem 1	Dönem 2	Dönem 3	Dönem 4	Toplam Miktar	Fiyat	Satış Tutarı
Bira	35	75	100	55	265	6.85	1,815.25
Hububat	110	110	100	110	430	3.90	1,677.00
Ekmek	200	215	235	200	850	1.59	1,351.50
Süt	200	300	300	300	1100	1.20	1,320.00
Çocuk bezi	10	20	50	120	200	4.99	998.00

Dönem bazında satılan ürün miktarlarını gösteren tablodan, bira ve çocuk bezi satışlarının dönemden etkilendiklerini, ancak hububat, ekmek ve süt satışlarının döneme bağlı olarak değişkenlik arz etmedikleri sonucuna varılabilir.

3.4. Bilgi (Knowledge)

Önceki aşamalardan elde edilebilirliği veya mevcut bilginin mantıksal çıkarımları sonucu oluşabilirliği nedeniyle bilgi, veri ve enformasyondan farklıdır. Veri madenciliği (data mining) teknolojisini içeren uygulamalar sayesinde, veri içerisindeki gizli eğilim ve örüntüler belirlenebilir. [12] Market örneğinde veri madenciliği teknikleri kullanılarak çocuk bezi alan müşterilerin %50'si bira satın aldıkları bilgisi elde edilmiştir. Çocuk bezi alan müşterilerin mama da satın alacağını veya bira satın alanların cips de alacağı tahmin edebilir, ancak sadece otomatik bir analiz bütün olasılıkları göz önüne alır ve kolay düşünölemeyecek, örneğin çocuk bezi ve bira arasındaki bağıntıları da bulur.

3.5. Yararlı Bilgi (Wisdom)

Yararlı bilgi, toplanan bilginin yarar sağlayacak şekilde kullanılmasıdır. [12] Market örneğinde, alışılmışın dışındaki bira ve çocuk bezi ile ilgili örüntünün keşfedilmesi

neticesinde alınabilecek aksiyonların tespiti için, bu sonuç analitik veri kümesinde analiz edilmelidir.

Tablo 3.4 Bir Satın Alma Örüntüsünün Teşhis Edilmesi

Ürün	Dönem 1	Dönem 2	Dönem 3	Dönem 4	Toplam Miktar
Bira	35	75	100	55	265
Çocuk bezi	10	20	50	120	200
İlişkilendirilmiş bira satışları	5	15	25	55	100

1, 2 ve 3. dönemlerinde, *çocuk bezi alan müşterilerin %50'si bira satın alır* kuralına karşılık gelen miktarının üzerinde bira satışı gerçekleşmiştir. 4. dönemde ise bu satış kuralın öngördüğü miktarda gerçekleşmiştir. Bu sonuçlar neticesinde 4. ve 3. dönemlerdeki bira ile ilgili pazarlama kampanyalarının stratejileri kıyaslanarak, 3.dönemdeki etki ile 4.dönemdeki değişim analiz edilmelidir. Ayrıca diğer dönemlerdeki satışlar da bütçe altına alınarak, öngörülen miktarın üzerinde gerçekleşen bira satışlarına katkıda bulunan diğer olaylar araştırılmalıdır. Bilginin bu şekilde kullanımı, kendisini oluşturan verinin daha iyi anlaşılmasını sağlar.

Operasyonel sistemlerdeki verileri biriktiren organizasyonlar, kendilerine değer katacak verideki potansiyeli anlama fırsatına sahip olurlar.

3.6. Verinin Anlam Düzeylerine Göre Kullanılan Teknolojiler

Aşağıdaki tabloda verilerin anlam düzeyleri ve bunlara karşılık gelen teknolojiler sınıflandırılmıştır. [12]

Tablo 3.5 Verilerin Anlam Düzeylerine Karşılık Gelen Teknolojilerin Sınıflandırılması

Anlam düzeyi	Teknoloji
Veri	OLTP (Online transaction processing) sistemleri
Enformasyon	Sorgu ve raporlama uygulamaları
Analitik	OLAP (Online analytical processing) uygulamaları
Bilgi	Veri madenciliği uygulamaları
Yararlı bilgi	İnsan zekası

Yapay zeka, insanın düşünce işleyişini taklit etmeyi denediği halde, hiçbir teknoloji insan beynin yerini alamaz. Organizasyonların çoğu veri proses sürecinin analitik aşamasındadırlar. Ancak sadece verinin ve teknolojilerin değerini anlayan organizasyonlar, rekabet avantajı sağlayacak diğer aşamalara geçmiştir.

4. VERİ MADENCİLİĞİ

4.1. Veri Madenciliği Kavramı

Küçük ölçekli işletmeler, müşterileri hakkında sahip oldukları bilgiye dayanarak onların sadakatini kazanmaya çalışırlar. Uzun yıllar alışveriş yaptığınız bakkal sahibi, ürün, marka ve fiyat tercihlerinizi bilir. Bir ürün konusunda kendisine danışırsanız, hakkınızda sahip olduğu bilgiler ışığında size öneride bulunur. Köşede yeni bir bakkal açılabilir ancak sizi tanıması uzun zaman alacağından eskisine olan bağlılığınız devam eder.

Her ölçekteki işletme, hizmete yönelik küçük işletmelerin her zaman başarıyla yürüttükleri, müşterisi ile bire bir ilişkiler yaratmayı öğrenmelidir. Tüm sektörlerde, geleceği görebilen kuruluşlar her bir müşterisini anlamaya çalışır ve öğrenilen bilgiyi, kendisini rakibine tercih etmesini sağlayacak şekilde kullanır.

Küçük bir işletme, müşterisi ile olan ilişkilerini kurarken, ihtiyaçlarının *farkına varır*, tercihlerini *hatırlar* ve geçmişteki etkileşimlerinden, kendisine gelecekte nasıl daha iyi hizmet verebileceğini *öğrenir*. Müşterilerinin birçoğu, çalışanları ile bir kez bile yüz yüze etkileşimde bulunamamış büyük işletmeler, benzer bir ilişkiyi kurmaları önceleri imkansızdı.

Hatta bir etkileşim söz konusu olsa dahi, müşteri her defasında büyük ihtimalle farklı bir satış temsilcisi ya da bir çağrı merkezi (call-center) çalışanı ile görüşecektir. Bu etkileşimler sırasında her çalışan, müşteri ile ilgili farkına varma, hatırlama ve öğrenme eylemlerini başarması güçtür. Müşteriyi isminden, yüzünden ve sesinden tanıyan, alışkanlıklarını ve tercihlerini hatırlayan, tek mal sahibinin yaratıcı sezgilerinin yeri doldurulamasa da, enformasyon teknolojilerinin başarılı uygulamaları aracılığıyla şaşırtıcı derecede yaklaşılabılır.

Öğrenmenin gerçekleşebilmesi için, birçok kaynaktan gelen, başvuru ve işlem bilgileri, sisteme aktarılmış her türlü belge, irtibat kayıtları vb. gibi verilerin tümü tutarlı ve faydalı şekilde bir arada toplanmalı ve enformasyon yönetiminde bilinen ismiyle **Veri Ambarı** oluşturulmalıdır. [3]

Veri Ambarı tüm işletmenin hafızası konumundadır, ancak zekice kullanılmadığı takdirde bir anlam ifade etmez. Zeka sayesinde hafızalarımızı yoklayarak örüntüleri tespit edebilir, yeni fikirler üreterek ve gelecek hakkında tahmin yürüterek kurallar tasarlayabiliriz. Veri madenciliği, hızla artan veri hacminden değer üretmeyi sağlayacak araçların başında gelir.

Veri Madenciliğin Tanımı

Konunun önde gelen uzmanlarından Piatetsky-Shapiro veri madenciliğini, verilerden daha önceden bilinmeyen, zımnî, muhtemelen faydalı enformasyonun monoton olmayan bir süreçte çıkartılması işlemi olarak tanımlamaktadır. Bu süreç kümeleme (*Clustering*), veri özetleme (*Data Summarization*), sınıflama kurallarının (*Classification Rules*) öğrenilmesi, bağımlılık ağlarının (*Dependency Networks*) bulunması, değişikliklerin analizi (*Analysing Changes*) ve anomali tespiti (*Detecting Anomaly*) gibi farklı bir çok teknik yaklaşımı kapsamaktadır.[1]

Gartner Group tarafından yapılan bir diğer tanımda ise veri madenciliği, istatistik ve matematik tekniklerle birlikte örüntü tanıma (*Pattern Recognition*) teknolojilerini kullanarak, depolama ortamlarında saklanmış bulunan veri yığınlarının elenmesi ile anlamlı yeni korelasyon,örüntü ve eğilimlerin keşfedilmesi sürecidir.

4.2. Veri Madenciliğin Önemini Arttıran Nedenler

Veri madenciliğinde kullanılan birçok teknik, aslında yıllardır, hatta yüzyıllardır varolan algoritmalardır, ancak son birkaç yıl içinde bu akım sayesinde daha yaygın kullanımları gündeme geldi. Nedenleri aşağıdaki şekilde sıralanabilir :

- Veri üretimi
- Verinin depolanması
- Bilişimsel gücün yeterliliği
- Artan rekabet baskısı

4.2.1. Veri Üretimi

Veri madenciliği ancak yüksek hacimli veriler için anlam kazanır. Veri madenciliği işlevlerinde kullanılabilmesi amacıyla oluşturulacak ve test edilecek modeller için çok sayıda veriye ihtiyaç duyulmaktadır.

Yakın zamana kadar, ilk örnekleri iletişim ve bankacılık olmak üzere, çok az sektörde müşteri ile olan etkileşimler otomatikleştirilmekte ve tüm işlemler kaydedilmekteydi, ancak son zamanlarda günlük hayattaki otomatik süreçler hızla yaygınlaşıyor. Bugün, POS, ATM teknolojilerinin ve bununla birlikte kredi ve elektron kart kullanımının hızla artması, Internet'ten yapılan alışverişler, elektronik fon transferleri, otomatik ödeme talimatları gibi uygulamaların yaygınlaşması sonucunda inanılmaz oranlarda veri üretilmektedir.

4.2.2. Verinin depolanması

Veri ambarları, müşteri ile olan etkileşimlerden üretilen veriler ile sınırlı kalmamakta, taleplerin veya şikayetlerin değerlendirme süreçlerinde üretilen ve bunun gibi çok farklı kaynaklardan elde edilen tüm verileri bir arada toplamaktadır.

Veri madenciliği, sadece işlem performansını sağlayacak şekilde tasarlanmış ve karar destek sistemleri için uygun olmayan operasyonel veritabanlarında yapılmamalıdır. Veri ambarları ise özellikle karar destek sistemlerinin performansını ve kalitesini arttıracak şekilde tasarlanır.

4.2.3. Bilişimsel gücün yeterliliği

Veri madenciliği algoritmaları genellikle muazzam veri miktarları üzerinden defalarca işlem yapmayı gerektirmektedir. Çoğunda yoğun hesaplamalar yapılmaktadır. Disk, hafıza ve işlem gücü fiyatların düşüşü, önceden maliyeti nedeniyle az yerde kullanılan yöntemlerin yaygınlaşmasını sağladı.

4.2.4. Artan rekabet baskısı

Enformasyon bakımından zengin, iletişim, sigorta ve finans en başta olmak üzere, birçok sektörde kıyasıya rekabet yaşanmaktadır. Birçok trend enformasyonun rekabetteki önemini arttırmaktadır :

- Hizmete dayalı bir ekonominin gelişmesi
- Kitle farklılaştırmasının (Mass customization) baş göstermesi
- Enformasyonun bir ürün olarak öneminin artması

4.3. Veri Madenciliği Modelleri

Veri madenciliğinde kullanılan modeller, tahmin edici (*Predictive*) ve tanımlayıcı (*Descriptive*) olmak üzere iki ana başlık altında incelenmektedir. [1]

Tahmin edici modellerde, sonuçları bilinen verilerden hareket edilerek bir model geliştirilmesi ve kurulan bu modelden yararlanılarak, sonuçları bilinmeyen veri kümeleri için sonuç değerlerin 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.

Tanımlayıcı modellerde ise karar vermeye rehberlik etmede kullanılabilecek mevcut verilerdeki örüntülerin tanımlanması sağlanmaktadır. X/Y aralığında geliri ve iki veya daha fazla arabası olan çocuklu aileler ile, çocuğu olmayan ve geliri X/Y aralığından düşük olan ailelerin satın alma örüntülerinin birbirlerine benzerlik gösterdiğinin belirlenmesi tanımlayıcı modellere bir örnektir.

Veri madenciliği modelleri, gördükleri işlevlere göre aşağıdaki şekilde sıralanabilir :

- Sınıflandırma (*Classification*)
- Tahmin (*Estimation*)
- Öngörme (*Prediction*)
- Birlikte Kuralları (*Association Rules*) ve Ardışık Zamanlı Örüntüler (*Sequential Patterns*),
- Kümeleme (*Clustering*)
- Tanımlama (*Description*)

4.3.1. Sınıflandırma (Classification)

Sınıflandırma, veri madenciliği işlevleri arasında en yaygın olanıdır. Dünyayı algılayabilmek ve onunla iletişim kurabilmek için onu oluşturan her şeyi devamlı sınıflandırırız. Örneğin yaşayan tüm canlıları türlerine ve cinslerine göre, maddeleri elementlerine göre ayırırız.

Sınıflandırma, yeni bir nesnenin özellikleri araştırılarak, önceden tanımlanmış sınıflar kümesindeki uygun olanına dahil edilmesidir. Veri madenciliğinde işleme tabi nesneler, veritabanındaki kayıtlar ile temsil edilmekte ve sınıflandırma işlemi neticesine her bir kayda belirli bir sınıf kodu atanmaktadır.

Tanımı iyi yapılmış sınıflar ve önceden sınıflandırılmış örnekler kümesi, sınıflandırmanın karakteristik özelliklerindendir. Genel amaç, verilerin sınıflandırılmasında uygulanabilecek bir modelin kurulmasıdır.

Sınıflandırma işlevinin pratikte kullanımı ile ilgili bazı örnekler :

- *Kredi başvurularının risk derecelerine göre sınıflandırılması*

4.3.2. Tahmin (Estimation)

Sınıflandırma işlevinde bağımlı değişkenler kategorik bir değere sahip iken, tahmin işlevinde söz konusu değerler süreklilik göstermektedir. Girdi olarak kullanılan birkaç veriden tahmin işlemi sonucunda gelir, boy veya kredi kartı bakiyesi gibi bilinmeyen ancak süreklilik arz eden değişkenler için değer üretilir.

Tahmin işlevinin pratikte kullanımı ile ilgili bazı örnekler :

- *Bir ailedeki çocuk sayısının tahmini*
- *Bir ailedeki toplam gelirin tahmini*
- *Bir müşteri ile devam eden ilişkinin ömür değeri tahmini*
- *Bakiye transferi ile ilgili teklifinizin cevaplanma olasılığının tahmini*

4.3.3. Öngörme (Prediction)

Öngörme işlevini sınıflandırma ve tahmin işlevlerinden ayıran en önemli özelliği, kayıtların, mevcuttan öte, ileride öngörülen davranış ve değerler ışında sınıflandırılmasıdır.

Öngörme işlevinin pratikte kullanımı ile ilgili bazı örnekler :

- *Bakiye transferi ile ilgili teklifinizin olumlu karşılanması durumunda transfer edilecek miktarının öngörülmesi*
- *İlk altı ayda sizi terk etme olasılığı yüksek olan müşterilerin öngörülmesi*
- *Yeni ürünü talep edebilecek müşterilerin öngörülmesi*

4.3.4. Birliktelik Kuralları (*Association Rules*) ve Ardışık Zamanlı Örüntüler (*Sequential Patterns*)

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.

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

- *Müşteriler bira satın aldığı anda, % 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.

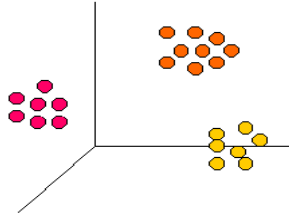
- *X ameliyatı yapıldığında, 15 gün içinde % 45 ihtimalle Y enfeksiyonu oluşacaktır,*
- *İMKB endeksi düşerken A hisse senedinin değeri % 15'den daha fazla artacak olursa, üç iş günü içerisinde B hisse senedinin değeri % 60 ihtimalle artacaktır,*
- *Çekiç satın alan bir müşteri, ilk üç ay içerisinde % 15, bu dönemi izleyen üç ay içerisinde % 10 ihtimalle çivi satın alacaktır.*

4.3.5. Kümeleme (*Clustering*)

Kümeleme modellerinde amaç, küme üyelerinin birbirlerine çok benzediği, ancak özellikleri birbirlerinden çok farklı olan kümelerin bulunması ve veri tabanındaki kayıtların bu farklı kümelerle bölünmesidir. Kümeleme işlemi sınıflandırma işleminden ayıran temel özelliği, önceden tanımlanmış sınıflarla ilgili olmamasıdır.

Sınıflandırma işleminde, nüfusun bölümlenmesi, her bir ögesinin veya kaydının, önceden sınıflandırılmış örnekler doğrultusunda oluşturulan modele göre, belirli bir sınıfa atanmasıyla gerçekleştirilir. Başlangıç aşamasında veri tabanındaki kayıtların hangi kümelere ayrılacağı veya kümelemenin hangi değişken özelliklerine göre yapılacağı bilinmemekte. Kümeleme neticesinde oluşan sonuçların anlamlı olup olmadığının kararı, konunun uzmanına bırakılmıştır. Semptom kümeleri farklı hastalıkları, yaprak ve çekirdek özelliklerinden oluşan kümeler farklı mısır türlerini işaret edebilir.

Kümeleme işlemi, çoğunlukla veri madenciliğinin diğer yöntemleri veya modelleme çeşitleri için bir başlangıç niteliğinde kullanılır. Örneğin Pazar Segmentasyonu çalışmasında kümeleme ilk adım olabilir : “Müşteriler en çok hangi promosyona ilgi gösterebilirler?” sorusu yerine, müşterileri benzer alışveriş alışkanlıklarına göre kümelere ayırdıktan sonra, “Her bir müşteri kümesi için en uygun promosyon hangisidir?” sorusu sorulabilir.



Şekil 4.1 Kümeleme Modeli

4.3.6. Tanımlama (Description)

Veri madenciliği ayrıca, karmaşık bir veritabanında olup bitenleri tanımlayarak, verileri oluşturan müşteri, ürün ve prosesleri daha iyi anlamamıza yardımcı olur. Bir davranış ne kadar iyi tanımlanırsa, o kadar doğru açıklanabilir. İyi bir tanımlama en azından nereden başlanabileceği konusunda fikir verir.

4.4. Veri Madenciliği Türleri

Temelde iki çeşit veri madenciliği vardır:

1. **Hipotez testi** : Önyargılı düşünceleri kanıtlamaya veya çürütmeye çalışan yukarıdan aşağı bir yaklaşımdır.

2. **Yararlı bilgi keşfi** : Verilerden başlayarak daha önce bilinmeyenleri açığa çıkarmaya çalışan aşağıdan yukarı bir yaklaşımdır. Yararlı bilgi keşfi, *doğrudan* (directed) veya *dolaylı* (undirected) olabilir. Doğrudan bilgi keşfi, belirli alan değerlerinin, diğerleri cinsinden ifade edilmesidir. Dolaylı bilgi keşfinde ise hedeflenen bir alan yoktur. Veri içerisindeki ilişkilerin tanınmasında *dolaylı*, önceden bulunan ilişkilerin açıklanmasında ise, *doğrudan* bilgi keşfi kullanılır.

İki yaklaşım kullanılarak, veri madenciliği problemlerinin iki yönden çözülmesine çalışılır. Öncelikle, gözlemlenen davranışlar doğrultusunda oluşan hipotezler, verilerle analiz edilir, ardından verilerin yeni hipotezler önermesi sağlanır. [2]

5. VERİ MADENCİLİĞİN ETKİLEŞİMLİ ÇEVİRİMİ

Verilerden elde edilen enformasyon ile, müşteri segmentleri tanımlanır ve pazarlama faaliyetlerine odaklanılır. Bu çerçevede, müşterilerin gerçek ihtiyaçlarını karşılayacak ürün tasarımları belirlenebilir ve talepleri tahmin edilerek en uygun kaynak tahsisi yapılabilir.

Veriler, birçok şirketin iş proseslerinin merkezinde yer almaktadır. Veri madenciliği, gerek sektöre özgü işlemlerden, gerekse dış kaynaklardan elde edilen tüm bu verilerin içinde gizli örüntüleri keşfetmeyi vaat etmektedir. Ancak bu kadarı yeterli değil, asıl önemlisi, bulunan bu örüntülerin cevaplanabilmesi ve aksiyona dönüştürülebilmesidir. Özetlemek gerekirse veriyi enformasyona, enformasyonu aksiyona, aksiyonu değere dönüştürmek veri madenciliği çevriminin özünü oluşturmaktadır.

Vaat ettiklerini başarabilmek için veri madenciliğini, pazarlama, satış, müşteri ilişkileri, ürün tasarımı ve stok yönetimi ile birlikte temel bir iş prosesi haline getirilmeli ve keşfetme sürecinin kendisine değil, keşfe dayalı aksiyona odaklanılmalıdır.

Algoritmaların önemi yadsınamaz, ancak veri madenciliği çözümleri güçlü tekniklerin ötesinde bir çaba gerektirmektedir. Bu teknikler, doğru alanlarda, doğru verilere uygulanmalıdır.

Veri madenciliği uzakta bir ada değildir; müşteri, pazar, ürün ve rakiplerin daha iyi anlaşılma çabasıyla kazanılan yararlı bilginin iç proseslere uyarlanması sürecinin içinde yer alır.

Büyük bir finansal kuruluşun pazarlama grubu, karlılığını arttırması gerekiyor, ancak kuruluş yeni müşteri kazanmaktan çok müşteri kaybetmekte ve yeni kazanılan müşteriler de eskilerine oranla daha az kar getirmektedir. Zayıf iyi müşterilerin kaybından kaynaklanmaktadır. Pazar payını koruyabilmek için yeni müşteriler aranmalı, ancak kayıplar nedeniyle yeni müşterinin kazanılabilmesi de çok maliyetlidir. Veri madenciliğin sunabileceği çözümden önce, işletmelerin genelde

verilerden elde edebilecekleri faydaları göz ardı ederek izledikleri yöntemleri irdeleyelim.

Yöntemlerden biri, işletmenin tümüyle daha rekabetçi duruma getirilmesi: mevduat faiz oranlarını yükseltmek, minimum bakiye sınırını düşürmek ve kredi faiz oranlarını düşürmek. Bu yaklaşım oldukça çekici görünmekte; birçok yeni müşteri kazandıracak ve ayrılanların sayısını azaltacak. İşletme pazar payını arttırmakta, ancak bu programların uygulanması oldukça maliyetlidir. Mevcut müşteriler tahminen verilen hizmetten memnun ve birçoğunun ilişkilerini devam ettirmeleri için bu tür teşviklere gerek olmayabilir. Daha da kötüsü bu yöntem işletmeye en sadakatsiz müşterileri kazandırabilir. Biraz daha iyi oranlar sunan bir rakip her an bu müşterileri kapabilir. Sonuç olarak önerilen program sadakatsiz müşteriler kazandırmakta ve zayıf problemini çözmemektedir.

İkinci olarak, işletme, kendisine para kaybettiren hizmetleri durdurabilir. İyi bir yaklaşım gibi görünse de, müşteriden çok işletme içerisindeki kar merkezlerine odaklı bir çözümdür. Kendisine para kaybettirdiği gerekçesiyle, fatura ödeme hizmetlerinden bazılarının durdurulma kararı alındı, ancak çok geçmeden en sadık ve en karlı müşterilerinin bu hizmetlerden faydalandıkları fark edildi. Bu hizmetin durdurulması söz konusu müşterileri kaybetme riskini doğurabileceğinden, hizmetin durdurulmasından vazgeçildi ve böylece çok daha maliyetli bir hata önlenmiş oldu. Değer katan bu tür hizmetlerin durdurulması, en değerli müşterilerinin yeni arayışlara yönelmesine neden olabilir ve neticede karlılık sorunu daha kötü bir duruma gelebilir.

Muhtemelen müşteriler işletmenin hizmetlerinden memnun değiller. İşletme, bankamatik sayısını arttırabilir ve bekleme sürelerini azaltmak amacıyla, çağrı merkezi ve kalabalık şubelerinde daha fazla personel görevlendirebilir. Karlılığını arttırmak için, hizmetlerine yenilerini ekleyebilir ve ürün paketlerini tüm müşterilerine sunabilir. Daha iyi hizmet anlayışı doğru bir yaklaşım, ancak tüm müşteriler için her şeyi birden iyileştirmek oldukça zahmetli ve maliyetlidir. Halbuki öncelikle işletmeyi terk etme olasılığı bulunan müşteriler hedeflenirse zayıf problemi aşılmış olur.

Bir analiz uzmanı verileri derinlemesine inceleyebilir ve örneğin, tüm hesapların bakiye sorgulaması gibi belirli bir işlemin, işletme ile ilişkisini kesme olasılığı

bulunan bir müşteriye önceden haber verebileceğini öğrenir. Bu enformasyonu keşfeden analiz uzmanı, onunla ne yapacağını bilemediğinden önemsemeyecektir. Daha da ötesi, merkezde çalışan bir kişinin, kilometrelerce uzaklıkta bulunan bir şubedeki müşteri için hemen aksiyona geçmesi de olanaksız.

Müşteri kaybı ortak bir sorun ve bu amaca uygun çözümler geliştirilmeli. Örneğimizde, karlılık probleminin müşteri kaybı ile ilişkili olduğunu farkına varılması, işletmeye oldukça yol aldırıştır. İzleyen adımlarda müşterilerin anlaşılması ve elde edilen bilginin amaca en uygun şekilde kullanılması üzerine odaklanılmalı.

5.1. Veri Madenciliği Çevrimi ile Müşteri Kaybı Probleminin Çözümlemesi

Müşteri kaybı probleminin çözümü ile ilgili tahminlerde bulunmak yerine, müşteriler hakkında sahip olunan veriler, önce enformasyona sonra da aksiyona dönüştürülerek, çok iyi değerlendirilebilir.

Örnek olarak, işletmenin, müşterilerini tutmaya yardımcı olacak bir çözümü nasıl keşfedebileceği ve bu çözümü nasıl uygulayabileceği irdelenecek. Çözümün bir kısmı veri madenciliğine, ancak büyük bir kısmı ayrıca işletmenin işle ilgili deneyim ve bilgisine dayanır.

Öncelikli adım, fırsatın belirlenmesidir. Örneğimizde, en önemli şeyin karlı müşterilerin kaybı olduğu belirlenmiş, böylece kayıp azaltılırsa, kazanç fırsatı elde edilmiş olacaktır. İşletmenin kaybedilebilir durumdaki müşterilerini tespit edebileceği en iyi yaklaşımlardan biri, önceden kaybedilen müşterileri inceleyerek, nedenlerini saptamaya çalışmaktır. Bu durumla karşılaşan organizasyonlar çoğunlukla eski müşterilerle anketler yaparak nedenleri çıkarmaya çalışırlar. Bu anketler firmanın müşteri servisindeki bir grup tarafından veya bağımsız bir başka şirket tarafından yapılabilir. Her iki durumda da anketler aşağıdaki nedenlerden dolayı muhtemelen doğru sonuçlar üretmeyecektir:

- Anketleri cevaplayan kişiler, eski müşterilerinizi temsil etmeyebilir.
- Eski müşterilerinizin artık işletmenizle bir ilişkisi kalmadığına göre size yardımcı olmaları veya bu konuda dürüst olmaları konusunda zorlayamazsınız.

- Eski müşterilerin ayrılma nedenleri birden fazla olabilir: evlerine yakın şube kapanmış, diğer şubeler oldukça uzak ve bankamatiklerde çok beklemek zorunda kalıyorlardı. Siz bu nedenlerden sadece birini, büyük bir ihtimalle daha belirgin olan ilkini duyabilirsiniz.

Sonuç olarak işletmenin mevcut ve eski müşterilerine ait tüm verilerine bakmak daha doğru bir yaklaşım olacaktır. Eski müşterilere nazaran, veriler gerçekleri yansıtır ve kendilerine ulaşılması daha kolaydır.

Sonraki adımda, işletme, verilerini analiz ederek, ilk adımda belirlenen fırsatın gerçekleştirilebilmesi için en uygun yaklaşımı bulmalıdır. Bunu veri madenciliği sayesinde başarabilir.

Örneklenen amaç doğrultusunda, analiz için uygun veri, çağrı merkezi kayıtlarından sağlandı. Analizler neticesinde ilginç kümeler elde edildi.

Kümelerden biri, yoğunluğu artık işletmeyle ilişkileri bulunmayan müşterilerden oluşmaktadır. Bu kümedeki kişiler, ortalama yaşın oldukça üzerindeler ve çok az ihtimalle bir ipoteğe veya kredi kartına sahipler. İleriki analizler, bu kişilerin vefat etmiş olabileceği ihtimalinin yüksek olduğunu göstermiştir. Dolayısıyla bu küme müşteri kaybının önlenmesi konusunda bir fayda sağlamayacaktır.

Bir başka küme aşağıdaki özelliklere sahip müşterilerden oluşmaktadır :

Birden fazla hesapları bulunmakta ve genellikle iş saatleri dışında aramaktadırlar. Çağrı merkezini aradıklarında ise, beklemek durumunda kalıyorlar. İleriki analizler, söz konusu müşterilerin herhangi bir şubeyi asla ziyaret etmedikleri ve sıklıkla başka işletme bankamatikleri kullandıklarını göstermiştir. Hizmet maliyeti az olan bu müşteri kümesi, çözüm vaat etmektedir.

Bu noktada, veri madenciliği, uygun bir kümenin seçimi ile ilgili görevini yerine getirmiş oldu. Ancak kümenin tanımlanmış olması yeterli değil yine de. İşletme, bundan sonraki adımda, veri madenciliği analizinden elde edilen sonuçları aksiyona dönüştürmelidir. İş fırsatı ve küme hakkındaki enformasyon doğrultusunda, işletme birkaç aksiyon belirledi :

- Bir şey yapılmayacak. Muhtemelen tanımlanan müşteriler çok kar getirmemektedir. Daha iyi sonuçlar elde edilinceye kadar veri analizine devam edilecektir.

- İş saatleri dışındaki bekleme süreleri azaltmak için çağrı merkezi personel sayısı arttırılacak. Bekleme süresi azalacak, ancak çağrı merkezi takviyesi oldukça maliyetlidir.
- Bu kümede bulunan müşteriler için, hizmet önceliği bulunan, bir başka telefon numarası sağlanacaktır.
- Çağrı merkezine, gelen çağrıların telefon numaralarını tespit edebilme yeteneği kazandırılacak. Telefon numarasının ait olduğu müşteriye göre çağrı önceliği belirlenecek.

Bu örnekte, işletme üçüncü alternatifi uyguladı. Söz konusu kümede bulunan müşterilere, önceliği bulunan bir telefon numarası vererek, daha iyi bir hizmet sağladı. Diğerlerine göre, bu çözüm hem çağrı merkezine daha kolay uygulanabildi, hem daha az maliyet gerektirdi. Bununla birlikte, seçilen alternatifin uygulanması biraz da ustalık ister çünkü, verilen yeni numaranın sadece belirlenen kümedeki müşterilere değil de, ait oldukları ev halkına da gittiğinden emin olunmalıdır.

Uygulamadan sonra, tanımlanan kümedeki müşterilerin ihtiyaçları daha hızlı cevaplanabildi ve bu nedenle de hizmet ile ilgili şikayetleri azaldı.

Veri madenciliği çevriminin son adımı, sonuçların ölçülmesidir. Öncelikli müşteri hizmetleri numarasının uygulamaya alınmasından sonra, söz konusu kümedeki müşteri kaybının azalıp azalmadığı, bu müşterileri elde tutma çabasının, karlılığı olumlu etkileyip etkilemediği ölçülmelidir.

Bulunan sonuçlar, analiz için daha fazla veri sağladığından, başka hangi müşterilerin işletme ile olan ilişkilerini koparmak üzere oldukları araştırılmalı. Genelde, bir iş fırsatı için bulunan bir çözüm, yeni analizler için de yeni fırsatlar doğurur. Bu ise, veri madenciliğin etkileşimli çevrimini özetlemektedir: bulunan bir çözüm, daha iyi sonuçlar üreten daha fazla enformasyona yol açar.

5.2. Veri Madenciliğin Etkileşimli Çevrimi

Çevrim, birbirinin sonuçlarını doğrudan etkileyen dört adımdan oluşmaktadır :

1. Problemin tanımlanması
2. Veri madenciliği yöntemleri ile verinin anlamlı enformasyon' a dönüştürülmesi

3. Enformasyonun aksiyona çevrilmesi
4. Sonuçların ölçülmesi

Sonuç olarak, veri madenciliğinin iş proseslerine dahil edilmesi, başarının anahtarıdır. Tüm adımlar birbirine bağlıdır. Bir adımın sonuçları onu takip eden adımın girdisidir. Sonuçlar üzerine kurulu bir yaklaşımdır, diğer bir deyişle de, her bir adım öncekinin sonuçlarına bağlıdır.[2]

5.2.1. Problemin Tanımlanması

İş fırsatının tanımlanması, organizasyonun tamamında biriken enformasyonun, insanlara işlerini daha iyi yapabilmelerini sağladığı sürece gerçekleşen bir adımdır. Amacı, değer üretebilecek veri alanlarını tanımlamaktır. Bu alanlar da veri madenciliğinin girdisi olacaklardır. Bu adımla ilgili birçok farklı yaklaşım bulunmakta, ancak hiçbirisi ‘doğru’ değildir. Asıl amaç, içerisinde değer katabilecek örüntülerin bulunduğu veri alanlarını belirlemektir.

Bazı iş proseslerin işlemleri, birtakım veri analizlerine dayandırılmakta ve bunlar veri madenciliğinin etkileşimli çevrimindeki *Tanımlama* adımı olarak kullanılabilir :

- Yeni bir ürün için pazarlama planının hazırlanması
- Mevcut ürün ve hizmetlerin fiyatlandırılması
- Pazarlama hedeflerinin belirlenmesi
- Müşteri kaybının anlaşılması
- Benzer diğer prosesler

Bu örneklerde, iş fırsatının çok iyi anlaşılmış, veri madenciliği de bu proseslerin ayrılmaz bir parçası olmalıdır.

Çoğu kez firma içindeki çeşitli düzeylerden gelen soru ve gözlemler, veri madenciliğine kılavuzluk eder. Genellikle yönetim tarafından sorulan ve tümüyle gözleme dayalı bu türden birkaç örnek şöyledir:

- California’daki satışların, güneydoğudaki satışların gerisinde kalmasının nedeni
- Bankamatiklerde uzun bekleyişlerin müşteri kaybına etkisi
- Uzak mesafelerdeki kullanımının mevsimsel örüntüleri (seasonal patterns)

- Müşteri desteğine daha fazla yatırım ile harcamaların biraz azaltılması çözümlerinin olası neticelerinin karşılaştırılması
- Çamaşır suyu ile birlikte promosyonu yapılabilecek başka ürünlerin belirlenmesi

Veriye ulaşılmadan bu gözlemlerden birçoğu asla yapılamazdı. Etkin bir veri madenciliği grubu olmaksızın da, bu soruların birçoğu asla cevaplanamazdı. Bu grubun teknik yetenekleri arttıkça da, enformasyona dayalı yeni sorgulamalar gündeme gelir.

Diğer bir yaklaşım ise, değer alanlarının organizasyon içindeki kilit kişilerin görüşleri doğrultusunda tanımlanmasıdır. Veri madenciliğin ilk zamanlarında, yumurta tavuk problemi nedeniyle, bu yaklaşım çok değerlidir. İş prosesinin içerisinde yer alan bir kişinin şu ana kadar enformasyonun avantajlarından faydalanamadığından onu ne şekilde kullanacağını da anlayamaz. Bir organizasyona veri madenciliğin değeri anlatılırsa, iki yönlü iletişim sağlanır ve bunun neticesinde her düzeydeki iş fırsatları belirlenebilir.

Konu uzmanlarının, belirledikleri iş fırsatı ile ilgili, veri madenciliğinden elde edilecek sonuçların ne şekilde ölçülebileceği ve bunların ne şekilde değerlendirebileceği konusunda fikir sahibi olması, sağlanacak avantajlar açısından son derece önemlidir.

5.2.2. Veri Madenciliği Yöntemleri ile Verinin Anlamlı Enformasyon' a Dönüştürülmesi

Tanımlanan iş fırsatlarından ve biriktirilen verilerden, aksiyona dönüştürülebilecek sonuçlar üretilir. Başarılı bir çevrim için ne tür sonuçların üretilmesi gerektiği iyice anlaşılmalıdır. Bir önceki adımda elde edilen sonuçlar kullanıldığında, karşılaşılabilecek pek çok tehlikeli durumdan bazıları şunlardır :

- *Tutarsız veri biçimleri*; örneğin, bir otomatik mail uygulamasında gerekli posta kodu dokuz haneli iken verilerde beş haneli olması
- *Veri alanlarının karıştırılması*; örneğin, teslim tarihi, bir sistemde planlanan teslim tarihi olarak kullanılırken bir diğer sistemde gerçekleşen teslim tarihi olarak kullanılabilir.

- *İşlevsel eksiklik*; örneğin, müşteri bazına indirgenmemiş bir çağrı merkezi uygulamasının kayıtları
- *Hukuki boyutlar*; örneğin bir krediyi reddederken mutlaka bir hukuki nedenin gösterilmesi
- *Organizasyonel faktörler*; bazı gruplar, özellikle teşvik edilmemiş iseler, işlem yapma biçimlerini değiştirmek konusunda isteksiz davranırlar.
- *Zamanlama*; örneğin bir ay sonra gelen sonuçlar artık aksiyona dönüştürülemeyebilir.

5.2.3. Enformasyonun Aksiyona Çevrilmesi

Bu adımda, veri madenciliğinden elde edilen sonuçlar ışında eyleme geçilir ve oluşan neticeler, bir sonraki adımı besler. Burada, iş prosesiyle enformasyonun nasıl birleştirilebileceği sorulmalıdır.

İş proseslerinden sorumlu farklı gruplar, enformasyonun gerekliliği konusunda duyarlı olmalıdırlar:

- Sadece bir ürünü lanse etmek yeterli değil. Yeni bir ürün lanse edildiğinde, esas müşteri tabanı ile ilgili enformasyon toplanabilir ve sonuçları gelecekteki pazarlama faaliyetleri için kullanılabilir. Müşteri tabanı genişledikçe de, pazarlama çabalarının etkileri takip edilmeli ki gelecekteki çabalara ışık tutabilsin.
- Sadece müşteriden gelen hizmet taleplerini cevaplamak yeterli değil. Müşteriden toplanan verilerin önceden tanımlanan müşteri profiline uygunluğu kontrol edilmelidir.

5.2.4. Sonuçların Ölçülmesi

Ölçüm daha iyi sonuçların devamlılığını sağlar. Buradaki ölçüm, ortalama ve standart sapmanın dışında, cevaplama oranı ve maliyetin ötesine geçen, işin değer ölçümüdür. Ölçümün ve devamlı iyileşmenin getireceği artılar bilinse de, genellikle bu konuya yeterince ilgi gösterilmemektedir.

Bir organizasyonda yapılan ölçümler, genellikle aksiyona geçilmesinden aylar sonrası oluşan standart rapor şeklinde gerçekleşmektedir. Problem, raporlardaki bilginin içeriği, geçerliliği ve en önemlisi kaynağının ne olduğudur. Bilginin,

fırsatları görebilecek ve bunları analiz ederek, aksiyona dönüştürebilecek kişilere ulaşım hızı önem arz eder. Çoğu organizasyonda bu değer aylarla ifade edilir. Anında veriye erişim ve analiz imkanı tanıyan raporlama araçları, bilginin zamanında doğru kişilere ulaşımını sağlamaktadır.

Her bir veri madenciliği çabasını küçük bir iş olayı olarak düşünmek gerekir. Beklentiler ile gerçekleşenlerin kıyaslanması sayesinde, çoğu zaman bir sonraki çevrimde kullanılmak üzere, olası fırsatlar tanımlanabilir. Her bir veri madenciliği çabasının başarılı veya başarısız sonuçları, gelecekteki çabalar için ders niteliğindedir. Gelecekteki kullanım için en iyi girdilerin sağlanması amacıyla, neyin ölçülmesi ve ne tür bir yaklaşım izlenmesi gerektiği sorulmalıdır.

Örnek olarak, hedefe dayalı bir pazarlama kampanyası için neyin ölçüleceği ile başlanabilir. Ölçüm olarak genellikle cevaplama oranı dikkate alınır. Kampanya için hedeflenen müşterilerden hangi oranda geri dönüş gerçekleşti sorusuyla sınırlı kalan bir ölçüm, birtakım önemli bilgilerin kaybına neden olmaktadır.

Bir pazarlama kampanyasında, geleceğe değer katmak amacıyla ölçülmesi gereken birkaç husus :

- *Kampanya sırasında kazanılan müşterilerin sağladıkları kazanç.* Bu ölçüm ancak karlı müşteri modelini oluşturan organizasyonlar tarafından yapılabilir. Bununla birlikte, bir ay, altı ay ve bir yıl sonraki değeri gibi sorular da deneysel ölçümler ile cevaplanabilir.
- *Müşteri sadakati.* Bir kampanyanın başarısı uzun dönemde sağlayacağı fayda ile ölçülmelidir. Uzun süreli müşteri ilişkileri, işin değeri açısından çok önem arz etmektedir.
- *Kampanya ile ulaşılan ve sadakati en yüksek müşterilerin demografik bilgileri.* Bilinen müşterilerin demografik bilgileri olası yeni müşterilerin kazanımına ışık tutabilir.
- *Müşterilerin başka ürünleri de tercih etme olasılığı.* Müşterilerin satın aldıkları ürün çeşitliliği, organizasyon içerisindeki farklı sistemler tarafından tespit edilebilirliğine bağlı olarak ölçülebilir.

- Telefona karşılık mail ile pazarlama veya indirim karşılık kupon gibi çoklu bir kampanyada farklı yöntemlerle kazanılan müşterilerin değer hesaplamalarının belirlenmesi

Tüm bu ölçümler, gerçekte, mevcut çabanın sonuçlarının, gelecekte ne şekilde kullanılabileceği sorusunun cevabını aramaktadır. Eğer bir telefon ile pazarlama neticesinde iyi sonuçlar elde edilmişse, bir sonraki veri madenciliği çevriminde daha önceki benzer kampanyalarda kullanılan pazarlama metinlerinin kıyaslanması istenilebilir.

Kampanya ile ulaşılan müşterilerin, uzun dönemde kazanç getirip getirmediikleri veya sadece kampanyanın fırsatlarından yararlanıp, bir sonraki kampanyayı bekleme eğiliminde olup olmadıkları bilgisi kampanya sonucunda elde edilen verilerin içinde gizlidir.

Bir başka örnek, müşteriye elde tutmak amacıyla düzenlenen bir kampanyadır :

- Karlı müşterilerin bağlılıkları ile diğerlerine oranı
- Kampanyanın, hedeflenen kitlenin müşteri ömrü değerine etkisi
- Yeniden kazanılan müşterilerin teşvikler sonrasında da bağlılıklarının devamı
- Tahminlerin geçerlilik süresi

Bu sorular, bilinen maliyet, gelir ve kar gibi özet ölçümlerin ötesinde, işin değer ölçümüne ve ürün, müşteri ve pazar gibi müdahale edilebilir atomik birimlerin ölçümüne odaklanmak gerektiğini işaret eder.

Sonuçların ölçülmesi adımı, önceki adımdan sağlanan enformasyon son derece önemlidir. Bu nedenle ölçüm için doğru enformasyonun sağlanabilirliği, soruların çok önceden doğru tespit edilmesine bağlıdır.

5.3. Modelin Değerlendirilmesinde Kullanılan Yöntemler [1]

5.3.1. Basit Geçerlilik Testi

Bir modelin doğruluğunun test edilmesinde kullanılan en basit yöntem basit geçerlilik (*Simple Validation*) testidir. Bu yöntemde tipik olarak verilerin % 5 ile % 33 arasındaki bir kısmı test verileri olarak ayrılır ve kalan kısım üzerinde modelin öğrenimi gerçekleştirildikten sonra, bu veriler üzerinde test işlemi yapılır. Bir

sınıflama modelinde yanlış olarak sınıflanan olay sayısının, tüm olay sayısına bölünmesi ile hata oranı, doğru olarak sınıflanan olay sayısının tüm olay sayısına bölünmesi ile ise doğruluk oranı hesaplanır. (*Doğruluk Oranı = 1 - Hata Oranı*)

5.3.2. Çapraz Geçerlilik Testi

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

5.3.3. N-Katlı Çapraz Geçerlilik Testi

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

6. PAZAR SEPETİ ANALİZİ (Market Basket Analysis)

Pazar Sepeti Analizinin bir resmini canlandırmak için, markette alışveriş yapan bir kişinin satın almış olduğu çeşitli ürünlerden oluşan bir sepeti düşünelim. Bu sepetin içindekiler, bir müşterinin alışveriş sırasında alabileceği, meyve suyu, muz, soda, cam temizleyici ve deterjan gibi ürün çeşitleri hakkında bize bilgi vermektedir. Bir sepet, bir müşteri hakkında bilgi vermekte, ancak tüm müşteriler tarafından yapılan alışverişler daha fazla enformasyon içermektedir. Her bir müşteri farklı ürün kombinasyonlarını, farklı miktarlarda, hafta boyunca farklı zamanlarda satın almaktadır.

Pazar Sepeti Analizi, müşteriler tarafından *hangi* ürünlerin satın alındığı bilgisinden, (enformasyon) müşterilerimizin *kimler* olduğu ve *neden* bazı alışverişlerin, bu müşteriler tarafından yapılmasının kaçınılmaz olduğunu anlamamamıza ışık tutar. Ayrıca, ürünlerden *hangilerinin* birlikte satılabileceği ve yapılacak promosyonlarda bunların kullanılabilirliği konusunda fikir verir.

Kökene POS (Point-Of-Sale) işlemlerinin analizi olmasına rağmen, Pazar Sepeti Analizi, bir müşterinin belirli bir zaman diliminde, bir arada aldığı ürün ve hizmetler veya art arda sergilediği davranışlar söz konusu olduğu sürece, perakende dışında da birçok sektörde uygulanabilir :

- Aynı kredi kartı üzerinden yapılan araba kiralama ve otel rezervasyonu gibi işlemler, müşterilerin bir sonraki adımda alabilecekleri ürün veya hizmet konusunda ipucu verir;
- Bireysel müşteriler tarafından kullanılan portföy hesapları, yatırım hizmetleri, araç kredileri v.b bankacılık hizmetleri, yeni hizmetlere ilgi gösterebilecek müşterileri belirleyebilir;
- Hasta vakalarından, bir arada yapılan tedavilerin kaçılmaz yan etkileri belirlenebilir; v.b.

Alışveriş sepeti örneğine dönersek, araştırılması gereken konulardan birkaçı şunlardır:

- Olabilecek en fazla satışın gerçekleşebilmesi için deterjanların yeri belirlenmeli
- Deterjan ile meyve suyu birlikte alındığında cam temizleyici ürününün satılma olasılığı
- Soda ile muz satışlarının korelasyonu ve farklı soda markalarının etkisi
- Marketin bulunduğu çevrenin demografik özelliklerin, müşterilerin alışveriş alışkanlıkları üzerindeki etkisi

Çoğu kez Pazar Sepeti Analizi, eldeki işlem verilerinden araştırılabilecek örüntülerin (patterns) belirlenemediği durumlarında, çıkış noktası olarak kullanılır. Bu özelliğine bağlı olarak, **dolaylı veri madenciliğine** örnek gösterilmesinin yanında, Pazar Sepeti Analizi **doğrudan veri madenciliğine** de uygundur.

Sepet analizinde amaç, alanlar arasındaki ilişkileri bulmaktır. Bu ilişkiler biliniyorsa, şirketin kârını arttırmak için kullanılabilir. Eğer X malını alanların Y malını da çok yüksek olasılıkla aldıklarını biliniyorsa ve eğer bir müşteri X malını alıyor ama Y malını almıyorsa o potansiyel bir Y müşterisidir.

Örneğin Internet üzerinden kitap satan Amazon şirketi *BookMatcher* adlı programıyla müşterilerine okudukları ve sevdikleri kitaplara göre satın almaları için kitap tavsiye etmektedir.

Eğer eldeki veride mallar için sadece satın alındı/alınmadı bilgisi varsa, sepet analizinde mallar arasındaki bağıntı, destek ve güven kıstasları aracılığıyla hesaplanır. İki mal, X ve Y, için destek (*support*) ve güven (*confidence*) tanımları şöyledir:

Destek (Support) :

$$P(X \text{ ve } Y) = X \text{ ve } Y \text{ mallarını satın almış müşteri sayısı} / \text{Toplam müşteri sayısı} \quad (6.1)$$

Güven (Confidence) :

$$P(X|Y)=P(X \text{ ve } Y)/P(Y) = X \text{ ve } Y \text{ mallarını satın almış müşteri sayısı} / Y \text{ malını satın almış müşteri sayısı} \quad (6.2)$$

Destek, veride bu bağıntının ne kadar sık olduğunu, *güven* de, Y malını almış bir kişinin hangi olasılıkla X malını alacağını söyler. Bağıntının önemli olması için her iki değerin de olabildiğince büyük olması gerekir.

Eğer eldeki malların müşteri tarafından ne kadar tüketildiği, ne kadar beğenildiği ile ilgili bilgi varsa o zaman bağıntı daha iyi hesaplanabilir. Örneğin markette müşterinin aylık toplam X malını kullanım miktarı hesaplanabilir.

Amazon'un *BookMatcher* programı okuyuculara okudukları her kitap için 1 ile 5 arasında bir beğeni notu vermelerini ister. Bu durumda X ve Y nümerik veriler olduğundan X ile Y'nin korelasyonu hesaplanabilir:

$$\text{Corr}(X,Y)=\text{Cov}(X,Y)/(\text{Std}(X)*\text{Std}(Y)) \quad (6.3)$$

X ile Y'nin kovaryansı, birbirlerine göre doğrusal olarak nasıl değer aldıklarını belirtir:

$$\text{Cov}(X,Y)=E[(X-m_X)(Y-m_Y)] \quad (6.4)$$

m_X X'lerin ortalaması, $\text{std}(X)$ 'de standart sapmasıdır. Örneğimizde m_X X malının ortalama olarak ne kadar beğenildiğini, $\text{std}(X)$ de beğenilerin bu ortalama etrafında ne kadar değişken olduğunu gösterir.

Eğer X'i sevenler genelde Y'yi de sevdiyse hem X, hem de Y değeri ortalamadan daha yüksek olacak ve $\text{Cov}(X,Y)>0$ olacaktır. Aynı şekilde X ve Y beraber beğenilmiyorsa her iki değer de ortalamadan küçük olacak ve yine $\text{Cov}(X,Y)>0$ olacaktır. Eğer X'i beğenenler Y'yi beğenmediyse (veya aksi takdirde) değerlerden biri ortalamadan yüksek, diğeri ortalamadan düşük olacak ve $\text{Cov}(X,Y)<0$ olacaktır. $\text{Corr}(X,Y)$ 'de $\text{Cov}(X,Y)$ 'nin -1 ile 1 arasında standart sapmalara göre normalize edilmiş halidir. $\text{Corr}(X,Y)$ değerinin 0 olması X ile Y arasında (doğrusal) bağlantı olmadığını, negatif değer ters, pozitif değer de doğrudan bağıntı olduğunu gösterir.

Bu şekilde olası bütün mallar arasında korelasyon bilgileri varsa X'i kullanan ve seven kişiye tavsiye edilecek, Y müşterinin kullanmadığı diğer bütün mallar arasında X ile korelasyonu en fazla ve olabildiğince 1 'e yakın olan mal olmalıdır.

6.1. Pazar Sepeti Analizinin kuvvetli yönleri (Strengths)

- Açık ve anlaşılabilir sonuçlar üretir.

- Dolaylı veri madenciliğini destekler
- Değişken uzunluktaki veri üzerinde çalışabilmektedir.
- Kullanılan hesaplama basit ve anlaşılır.

6.2. Pazar Sepeti Analizinin zayıf yönleri (Weaknesses)

- Problemin büyüklüğü arttıkça, bilişimsel çaba üstselsel olarak artmaktadır.
- Veri özniteliklerini (attributes) kısıtlı ölçüde desteklemektedir.
- Gerçek öge sayısını tespit etmek güçtür.
- Nadir öğeler ihmal edilmektedir.

Ayrıca, temel algoritma, belirli bir öğeyi içeren kuralları dikkate alacak şekilde indirgendiğinde, örneğin yeni bir ürünün satılabilirliği hakkında bilgi verebilir.

Bu yöntemlerin uygulanabileceği bir diğer alan, Zaman Serileri problemleridir. Verilerin basit bir dönüşümü ile birçok zaman serisi problemi Pazar Sepeti Analizi için uygun hale getirilebilir.

7. BELLEĞE DAYALI MUHAKEME (Memory-Based Reasoning)

İnsanlar eski deneyimlerine dayanarak karar verebilirler. Kalabalığın içinde tanıdık yüzler aradığınızda, bütün yüzleri, tanıdığınız yüzlerle kıyaslırsınız. Hekimler bir hastalığın teşhisini koyarken geçmişte karşılaştıkları benzer hasta ve belirtilerden faydalanırlar. Her iki durumda da kullanılan yöntem hemen hemen aynı : İlk adımda, önceden yaşanmış benzer durumlar *tanımlanır*, ardından elde edilen enformasyon söz konusu duruma *uygulanır*. Bu ise doğrudan veri madenciliği yöntemlerinden biri olan Belleğe Dayalı Muhakeme (MBR)' nin temelini oluşturmaktadır. Bilinen kayıtlardan oluşan bir veritabanında, MBR yöntemi, yeni kayıtların benzerlerini tespit ederek bunların sınıflandırılmasını ve öngörülmesini sağlar.

MBR yönteminde iki işlemin varlığı önem taşımaktadır : herhangi iki veri arasındaki uzaklığı tayin eden *uzaklık fonksiyonu* ve cevaplanması gereken duruma en yakın veri sonuçlarını bir araya getiren *birleştirme fonksiyonu*. [2]

7.1. Uzaklık Fonksiyonu

Uzaklık fonksiyonu, birleştirme fonksiyonu ve en yakın nokta sayısı yönteminin ürettiği sonuçların doğruluk derecesini belirlemektedir. Bu kriterlere bağlı olarak, mevcut kayıtlardan oluşan bir veri kümesi, çok doğru olduğu kadar, bir o kadar da uzak bir öngöründe bulunabilir.

Uzaklık Fonksiyonu tanımı

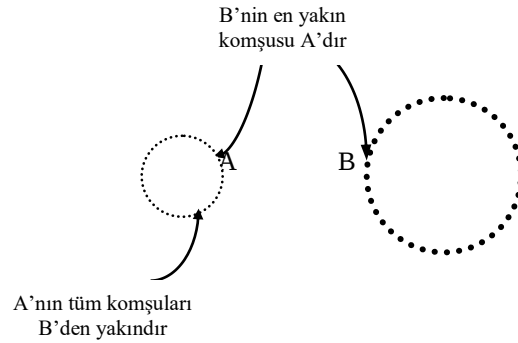
Uzaklık, MBR yönteminin benzerlik ölçüm şeklidir. A noktasından B noktasına olan uzaklık $d(A,B)$ ile gösterilir ve aşağıdaki dört temel özelliğe sahiptir :

1. *İyi tanımlanmış* : İki nokta arasında uzaklık her zaman tanımlıdır ve değeri negatif olmayan bir gerçek sayıdır. $d(A,B) \geq 0$
2. *Özdeş* : Bir noktanın kendisine olan uzaklığı her zaman sıfırdır. $d(A,B) = 0$

3. *Değişim özelliği* : Özellikle akış diyagramlarında A' dan B' ye olan uzaklık, B' den A' ya olan uzaklığa eşit : $d(A,B)=d(B,A)$ özelliği çoğu zaman geçerli olmamaktadır.
4. *Üçgen eşitsizliği* : A' dan B' ye giderken aradaki C' noktasından geçmek, A' dan B' ye olan mesafeyi asla kısaltamaz. $d(A,B) \geq d(A,C) + d(C,B)$

MBR için noktalar veritabanındaki kayıtlara karşılık gelmektedir. Uzaklığın biçimsel tanımı, benzerliğin ölçüm temelini oluşturmaktadır. Uzaklığın iyi tanımlanmış olması, her bir kaydın veritabanında bir benzerinin bulunduğu işaret etmektedir. Özdeşlik özelliği ise, bir kayıt için bulunan bir benzerinin kendisinden başkası olup olmadığının tespit edilmesine yardımcı olur. Değişim özelliği ve üçgen eşitsizliği kullanılarak, benzerlikleri en fazla olan, en yakın noktalar bulunur. Veritabanına yeni bir kaydın daha eklenmesi mevcut kaydın yakınlığını değiştirmeyecektir. Bir başka değişle, ancak iki kayıt arasındaki benzerlikten söz edilebilir. [2]

En yakın iki noktanın bulunmasında uzaklık kullanılıyor olsa da en yakın noktalar kümesinin kendine özgü birtakım özelliklere sahip olabileceği unutulmamalıdır. Örneğin, B kaydına en yakın kayıt A olmasına rağmen, A' ya B den daha yakın kayıtlar bulunabilir.



Şekil 7.1 MBR Benzerlik Kavramı (En yakın iki nokta arasındaki uzaklık)

Farklı türden alanlar için uzaklık fonksiyonunun oluşturulması :

Beş müşteriden oluşan bir Pazarlama Veritabanı düşünelim.

Tablo 7.1 Örnek Bilgiler 1

Kayıt no	Cinsiyet	Yaş	Gelir (\$)
1	Bayan	27	19.000
2	Bay	51	64.000
3	Bay	52	105.000
4	Bayan	33	55.000
5	Bay	45	45.000

Kayıtlar, iki nümerik ve bir de nümerik olmayan alanlardan oluşmaktadır. Bu örnekte, her bir alan için tanımlanan uzaklık fonksiyonları, iki kayıt arasındaki uzaklığı ölçen tek bir uzaklık fonksiyonunda birleştirilecektir.

Nümerik alanlar için en yaygın üç uzaklık fonksiyonu aşağıdaki şekildedir :

- Farkın mutlak değeri $|A-B|$
- Farkın karesi $(A-B)^2$
- Normalleştirilmiş mutlak değer : $|A-B|/(\text{Farkların en büyüğü})$

Normalleştirilmiş mutlak değerın avantajı, her zaman 0 ile 1 arasında değerler almasıdır. Bu örnekte, yaş ile ilgili değerler, gelire göre çok küçük olduğundan, normalleştirilmiş mutlak değer her ikisi için de en uygun çözüm olacaktır. Böylece alanlardan herhangi birinin, kayıtlar arasındaki uzaklık fonksiyonuna hakimiyeti bulunmayacaktır. [2] Aşağıdaki tabloda, müşterilerin yaşlarına göre birbirlerine olan uzaklık matrisi oluşturulmuştur.

Tablo 7.2 Örnek Bilgiler 2

	27	51	52	33	45
27	0,00	0,96	1,00	0,24	0,72
51	0,96	0,00	0,04	0,72	0,24
52	1,00	0,04	0,00	0,76	0,28
33	0,24	0,72	0,76	0,00	0,48
45	0,72	0,24	0,28	0,48	0,00

Müşterilerin gelire göre birbirlerine olan uzaklık matrisi aşağıdaki tabloda gösterilmiştir.

Tablo 7.3 Örnek Bilgiler 3

	19.000	64.000	105.000	55.000	45.000
19.000	0,00	0,52	1,00	0,42	0,30
64.000	0,52	0,00	0,48	0,10	0,22
105.000	1,00	0,48	0,00	0,58	0,70
55.000	0,42	0,10	0,58	0,00	0,12
45.000	0,30	0,22	0,70	0,12	0,00

Cinsiyet nümerik olmayan bir değer ve uzaklık fonksiyonu, en basit olarak, cinsiyetlerin aynı olduğu durumlarda 1, diğer durumlarda 0 değeri alacak şekilde tanımlanır.

$$d_{\text{cinsiyet}}(\text{bayan}, \text{bayan}) = 1$$

$$d_{\text{cinsiyet}}(\text{bayan}, \text{bay}) = 0$$

$$d_{\text{cinsiyet}}(\text{bay}, \text{bayan}) = 0$$

$$d_{\text{cinsiyet}}(\text{bay}, \text{bay}) = 1$$

Üç alana göre bulunan uzaklık fonksiyonları, kayıtlar arasındaki uzaklığın ölçümü için, yeni bir fonksiyonda birleştirilmelidir. Bu amaçla kullanılan üç yaygın yöntem bulunmaktadır :

$$1. \text{ Toplam : } d_{\text{toplam}}(A,B) = d_{\text{cinsiyet}}(A,B) + d_{\text{yaş}}(A,B) + d_{\text{gelir}}(A,B) \quad (7.1)$$

$$2. \text{ Normalleştirilmiş Toplam : } d_{\text{norm}}(A,B) = d_{\text{toplam}}(A,B) / \max(d_{\text{toplam}}) \quad (7.2)$$

$$3. \text{ Euclid denklemi : } d_{\text{euclid}}(A,B) = [d_{\text{cinsiyet}}(A,B)^2 + d_{\text{yaş}}(A,B)^2 + d_{\text{gelir}}(A,B)^2]^{1/2} \quad (7.3)$$

Her bir noktanın bu yöntemlere göre bulunan en yakın noktaları aşağıdaki tabloda gösterilmiştir.

Tablo 7.4 Örnek Bilgiler 4

	d_{toplam}	d_{norm}	d_{euclid}
1	1, 4, 5, 2, 3	1, 4, 5, 2, 3	1, 4, 5, 2, 3
2	2, 5, 3, 4, 1	2, 5, 3, 4, 1	2, 5, 3, 4, 1
3	3, 2, 5, 4, 1	3, 2, 5, 4, 1	3, 2, 5, 4, 1
4	4, 1, 5, 2, 3	4, 1, 5, 2, 3	4, 1, 5, 2, 3
5	5, 2, 3, 4, 1	5, 2, 3, 4, 1	5, 2, 3, 4, 1

Üç yöntemle de sonuçların aynı çıkması, seçilen beş kaydın, biri düşük gelirli, genç bayanlar, diğeri ise yüksek gelirli, orta yaş üzeri beylerden oluşan ve birbirinden tümüyle bağımsız kümeler oluşturmalarından kaynaklanmaktadır.

Örnek veritabanımıza, mukayese için kullanılacak yeni bir kaydın girildiğini varsayalım.

Tablo 7.5 Örnek Bilgiler 5

Kayıt no	Cinsiyet	Yaş	Gelir (\$)
Yeni	Bayan	45	100.000

Yeni kayda en yakın noktaların üç yöntemle göre sonuçları aşağıdaki tabloda gösterilmiştir.

Tablo 7.6 Örnek Bilgiler 6

	1	2	3	4	5	Sonuç
d_{toplam}	1,662	1,659	1,338	1,003	1,640	4, 3, 5, 2, 1
d_{norm}	0,554	0,553	0,446	0,334	0,547	4, 3, 5, 2, 1
d_{euclid}	0,781	1,052	1,251	0,494	1,000	4, 1, 5, 2, 3

Çıkan sonuçların neticesinde, yeni kaydı yukarıda bahsi geçen iki kümeden herhangi birine dahil edemeyiz. Euclid denklemine göre en uzak bulunan 3 numaralı kayıt, diğer iki yöntemle göre en yakın ikinci kayıt olarak bulunmuştur. Yeni kaydı 3 ile kıyasladığımızda cinsiyetlerinin farklı olduğunu ve bu alan için tanımlanan uzaklık fonksiyonunun alabileceği en yüksek değeri aldığını görmekteyiz. Bu nedenle, alanlar için bulunan uzaklık fonksiyonlarının hangi yöntemle birleştirileceğinin iyi belirlenmesi, çıkacak sonuç açısından son derece önemlidir. Ayrıca alanların ağırlık dereceleri tanımlanarak, söz konusu alanların uzaklık ölçümündeki etkisi belirlenebilir.

7.2. Belleğe Dayalı Muhakeme yönteminin kuvvetli yönleri (Strengths)

- Kolaylıkla anlaşılabilir sonuçlar üretir.
- Her tür veri tipine hatta ilişkisel olmayan veriye dahi uygulanabilir.
- Alan sayısına bağlı olmaksızın verimli çalışmaktadır.

7.3. Belleğe Dayalı Muhakeme yönteminin zayıf yönleri (Weaknesses)

- Sınıflandırma ve tahmin için maliyeti yüksek.
- Öğrenim kümesi için büyük miktarda veriye ihtiyaç duyulmaktadır.
- Sonuçlar, uzaklık ve kombinasyon fonksiyonlarının seçimine ve komşu noktaların sayısına bağlı olabilir.

8. YAPAY SİNİR AĞLARI (Artificial Neural Networks)

Yapay Sinir Ağları, birçok veri madenciliği ve karar destek uygulamalarındaki kanıtlanmış performansı nedeniyle, yaygın olarak kullanılmaktadır. Sinir ağları, öngörü, sınıflandırma ve kümeleme modellerine doğrudan uygulanabilen çok güçlü bir yöntemdir.

Mali serilerin tahmininden, sağlık durumların teşhisine, değerli müşterilerin tanımlanmasından, hileli kredi kartı işlemlerin tespit edilmesine, çekler üzerindeki rakamların tanınmasından, motorların arızalanma oranının öngörülmesine kadar birçok alanda uygulanmaktadır. [3]

Sinir ağı modellerinin temel çıkış noktası insan sinir fizyolojisidir. Sinir sisteminin bir parçası olan ve ortalama 1.5 kilogram ağırlığındaki insan beyninde, tahminen 10^{11} sayısında sinir hücresi bulunmaktadır. Öğrenme, hatırlama, düşünme, algılama gibi tüm bilişsel davranışları da içeren, her türlü insan davranışının temelinde nöron hücreleri bulunmaktadır. Tüm sinir hücrelerine doğuştan sahip olan bir insanın, sinir hücreleri yaşamı içerisinde yenilenmemektedir. Beynin gelişmesi ve ağırlık kazanması, sinir hücrelerinin büyümesi ve aralarında yeni bağlantıların kurulmasından kaynaklanmaktadır.

İnsanların tecrübelerine dayanarak, genelleme yapabilme konusundaki yeteneği karşısında, bilgisayarlar bilgileri defalarca işleyebilme gücüne sahiptirler. Sinir ağlarının cazibesi, insan beynindeki sinir bağlantılarını, dijital bir bilgisayar üzerinde modellemesi ile, bu iki özelliği birleştirmesinden ileri gelmektedir. Doğru alanlarda kullanıldığında, benzer verilerden öğrenme ve genelleme yaparak, insanlardaki, tecrübelerinden öğrenme konusundaki başarıyı gösterirler. Bu yeteneği sayesinde sinir ağları, veri madenciliğindeki yaygın kullanımı ile birlikte, gelecekte yeni ve daha iyi sonuçlar vaat eden bir araştırma alanına dönüşmüştür.

Bütün bunlara rağmen bir dezavantajı bulunmaktadır. Bir sinir ağının oluşumundan doğan sonuçlar, ağ içerisinde dağılmış (weights) ağırlıklardır. Bu ağırlıklar çözümün geçerlilik *nedeni* ile ilgili fazla bilgi sağlamadığından, verilen kararın doğruluğu

konusunda uzman kişilere danışılmalıdır. Ağırlıklar kolayca anlaşılamadığı halde, sinir ağlarını sondalayabilen ve gittikçe karmaşıklaşan teknikler ile çözümlerin açıklanabilirliği olanaklı hale gelmektedir.

Sinir ağları, insan bilincinin kaynağı kadar gizemli, içinde gizli işleme tarzlarının bulunduğu siyah kutulara benzemektedir.

8.1. Bir Gayrimenkul Değerinin Hesaplanması Örneği

Değer biçmenin otomatikleştirilmesi ile emlak acenteleri, olası alıcıları ile olası evleri daha iyi eşleştirebilir ve böylece henüz istenilen deneyimi kazanmamış acente temsilcilerinin de verimliliklerini arttırırlar. Ayrıca olası alıcıların, aradıkları evleri tanımlayabilecekleri ve fiyatı konusunda bilgi alabilecekleri kiosk veya web sayfaları sağlayabilirler.

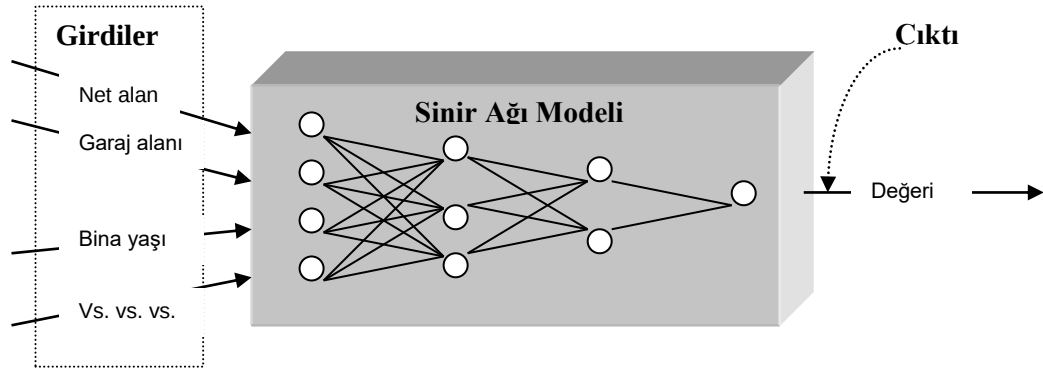
Mülkün özelliklerine göre pazar değerini belirleyen uzmanın yerine, sinir ağı kullanılarak değer biçilecektir. Değeri biçen kişi, şehrin bir tarafındaki evlerin diğerlerine nazaran daha kıymetli olduğunu bilir. Ayrıca yatak odaları, geniş garajı, evin tarzı ve arazinin büyüklüğü, hesabını etkileyen diğer faktörlerdir. Değerlemesini, birtakım formüller kullanarak yapmaz, bunun yerine tecrübelerine dayanarak ve benzer evlerin satış fiyat bilgilerini tartarak belirler. Evlerin fiyatı statik olmadığından, bölgedeki en son satış fiyatlarından haberdar olur ve zaman içindeki fiyat değişimlerine bağlı olarak değerlemesini son veri üzerinden günceller.

Eksper veya emlak acentesi temsilcisi, alanında uzmanlaşmış kişi için iyi bir örnek teşkil etmektedir. Evler, standart özellikleri dikkate alınarak, eksper tarafından değere dönüştürülen, sabit bir küme olarak tanımlanmaktadır. 1992 de IBM araştırmacıları, bu prosesin sinir ağları için iyi bir örnek olabileceğinin farkına varmışlar.

Bir sinir ağı, belirli girdileri (inputs) alarak, bunları belirli bir çıktıya (output) dönüştürür. Girdiler listesi önceden tanımlı olmalı, farklı kaynaklardan gelen listeler nedeniyle de standartlaştırılmalı. Beklenen çıktı da aynı şekilde önceden tanımlı olmalı. Ayrıca bir evin değerini nasıl hesaplayacağını öğretecek, önceki satışlar konusunda zengin bir tecrübeye ihtiyaç duyulmaktadır.

Sinir ağları öngörü problemleri için uygundur. Sinir ağları ile çözülebilecek bir problemin üç temel özelliği :

- *Girdiler anlaşılır olmalı.* Verinin hangi özelliklerinin önem arz edeceği bilinir, ancak bunların nasıl birleştirileceğinin bilinmesi zorunlu değil.
- *Çıktılar anlaşılır olmalı.* Neyin öngörüsü yapılacağı bilinir
- *Tecrübe bulunmalı.* Geçmişten hem girdilerin hem çıktıların bilindiği bol örnekler bulunmaktadır. Bu tecrübe, ağ modelinin kurulmasında kullanılacaktır.



Şekil 8.1 Sinir Ağı Modeli

(Kaynak : Berry, M. & Linoff,G.)

Bir evi tanımlayan ortak özellikler aşağıdaki tabloda gösterilmiştir.

Tablo 8.1 Örnek Bilgiler 1

Özellik	Açıklama	Değer
Daire sayısı	Apartmandaki daire sayısı	1 – 3
İnşa yılı	İnşa edildiği yıl	1850-1986
Sihhi boru	Yapının sihhi tesisatını oluşturan borular ve boru bağlama parçaları	5 – 17
Isınma tipi	Isıtma sisteminin tipi	A ve B şeklinde kodlanmış
Alt garaj	Bodrum garajı (araba sayısı)	0 – 2
Garaj	Bina çevresindeki garaj alanı (m ²)	0 – 228
Net alan	Net alan (m ²)	714 – 4185
Açık alan	Bahçe/Açık alan (m ²)	0 – 738
Balkon alanı	Balkon alanı (m ²)	0 – 452
Oturma salonu alanı	Oturma salonu alanı (m ²)	0 – 672
Bodrum alanı	Bodrum alanı (m ²)	0 – 810

Bir evin değerini hesaplayabilmek için fiyatını etkileyebilecek özelliklerin bilinmesi gerekmektedir. Tabloda verilen özellikler ile aynı bölgedeki evlerin değeri ölçülebilir. Farklı bölgelerdeki evlerin değerini etkileyebilecek, bölgenin demografik

bilgileri, ulaşım imkanları, yaşam standardı ile ilgili göstergeler gibi, daha pek çok özellik bulunmaktadır.

Öngörü için kullanılacak sinir ağına, önceki satış örnekleri ile ilgili bilgi verilmelidir. Literatürde öğrenme kümesi olarak adlandırılan bu veri setinde, daha önce bir evi tanımlayan özelliklere ek, satış fiyatı ve söz konusu satışın ne zaman gerçekleştiği bilgileri yer almalıdır.

Aşağıdaki tabloda öğrenme kümesi örneği verilmektedir.

Tablo 8.2 Öğrenme Kümesi Örneği

Özellik	Değer
Satış Fiyatı	\$ 171,000
Ay öncesi	4
Daire sayısı	1
İnşa yılı	1923
Sıhhi boru	9
Isınma tipi	A
Alt garaj	0
Garaj	120
Net alan	1,614
Açık alan	0
Balkon alanı	210
Oturma salonu alanı	0
Bodrum alanı	175

Tüm girdi ve çıktı değerleri 0 ile 1 arasında olduğunda, sinir ağları en iyi şekilde çalışır. Bu nedenle tüm değer aralıklarını ve kesin değerleri 0 ile 1 arasına bir değere indirgemek gerekir. Satış fiyatı, daire sayısı gibi özellikler, bilinen iki değer arasında değişkenlik gösterir. Örnek veride sadece ısınma tipi, A veya B gibi ayrık değerler alır. Farklı uygulamalarda, medeni durum, cinsiyet, hesap durumu, ürün kodu, vs. bu tür değerlere örnek olabilir.

Aralık olarak tanımlı bir özelliğin değeri, kendisinden aralığın alt limiti düşülerek elde edilen sonucun, aralık büyüklüğüne bölünmesi ile indirgenmiş olur. Örnek verideki 1923 inşa yılına, bu yöntemle göre karşılık gelen değer 0.5328 dir.

$$((1923 - 1850) / (1986 - 1850 + 1)) = 73 / 137 \cong 0.5328)$$

Ayrık değerler için ise 0 ile 1 arasında kesir değerler atanır. Örnekte B için 1, A için 0 değeri atanmıştır. Üç seçenekli bir örnek olsaydı her birine sırasıyla 0, 0.5 ve 1 değerleri atanırdı.

Öğrenim kümesindeki tüm değerler indirgendiğinde, ağ tarafından işlenmeye hazırdır. Veri madenciliği proseslerinde, öğrenme fazı olarak adlandırılan bu aşamada, sinir ağı örnek kümedeki veriler üzerinden öngördüğü çıktı değerini, gerçeği ile karşılaştırıp tüm iç ağırlıklarını yeniden ayarlamaktadır. En doğru değeri döndürecek optimum ağırlık setine ulaşınca kadar işlemleri defalarca tekrarlamaktadır. Ağırlıkların değişkenliği iyice azaldığı veya öğrenim veri setindeki örneklerin tekrarlanma sayısı, maksimum bir değere ulaştığında, mevcut veriden elde edilebilecek optimum çözüme ulaştığı kabul edilir.

Sonraki aşamada, öğrenme kümesinden elde edilen sonuçlar, bu kez daha önceki kümede yer almayan örneklerden oluşan bir test kümesi üzerinde test edilir. Test kümesi üzerinden de beklenen performans elde edildiğinde model geçerli hale gelir.

Modelin öngördüğü çıktı değeri de indirgenmiş bir değer olduğu ve bunun gerçek değere dönüşümü için daha önceki işlemlerin, bu kez tersine yapılması gerektiği unutulmamalıdır. Örnekte oluşan çıktının 0.75 gibi bir değer olduğu varsayılırsa, bu değer aralık büyüklüğü ile çarpımına, aralığın alt limit değeri eklenirse \$ 213,250 gerçek değeri elde edilir.

$$(0.75 * (\$ 250,000 - \$ 103,000)) + \$ 103,000 = \$ 213,250$$

İndirgenmiş öğrenim kümesi aşağıdaki tabloda gösterilmiştir.

Tablo 8.3 İndirgenmiş Öğrenim Kümesi Örneği

Özellik	Değer Aralığı	Gerçek Değer	İndirgenmiş Değer
Satış Fiyatı	\$103,000 - \$250,000	\$171,000	0.4626
Ay öncesi	0 – 23	4	0.1739
Daire sayısı	1 – 3	1	0.0000
İnşa yılı	1850 – 1986	1923	0.5328
Sıhhi boru	5 – 17	9	0.3333
Isınma tipi	A ve B şeklinde kodlanmış	A	1.0000
Alt garaj	0 – 2	0	0.0000
Garaj	0 – 228	120	0.5263
Net alan	714 – 4185	1,614	0.2593
Açık alan	0 – 738	0	0.0000
Balkon alanı	0 – 452	210	0.4646
Oturma salonu alanı	0 – 672	0	0.0000
Bodrum alanı	0 – 810	175	0.2160

8.2. Doğrudan Veri Madenciliği (Directed Data Mining) için Sinir Ağları

Önceki örnekte, sınıflandırma veya öngörü için bir model kurularak, sinir ağlarının en yaygın kullanımını resmedilmiştir. [5]

Bu prosesin adımları özetle şöyledir :

1. Girdi ve çıktı özelliklerin belirlenmesi
2. Girdi ve çıktı değerlerin 0 ile 1 arasında bir değere dönüştürülmesi
3. Benzer bir topolojideki ağının kurulması
4. Öğrenim kümesi örnekleri üzerinden ağının optimize edilmesi
5. Oluşan optimum ağının, öğrenim kümesinden tümüyle bağımsız bir test kümesi üzerinde test edilmesi. Gerekğinde, öğrenim kümesi, ağ topolojisi ve parametreler yeniden ayarlanarak adımların tekrarlanması
6. Ağ tarafından geliştirilen modelin, bilinmeyen girdilerin çıktı tahminlerinde uygulanması

İç işlemlerin detaylı bilinmesi zorunlu olmasa da, ağının kullanımında birtakım başarı anahtarları bulunmaktadır. İlki, doğru öğrenim kümesinin seçimidir. İkincisi, ağının veri içerisindeki örüntüleri tanıma yeteneğini maksimize edecek şekilde bilgilerin doğru betimlenmesidir. Üçüncüsü, ağ tarafından üretilen sonuçların yorumlanmasıdır. Son olarak, ağ içerisindeki topoloji ve öğrenmeyi kontrol eden parametreler ile ilgili birtakım detayların bilinmesi daha iyi sonuçların elde edilmesini sağlar.

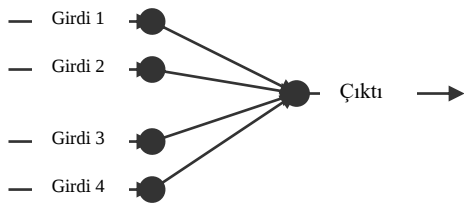
Öngörü veya sınıflandırma amaçlı kurulan herhangi bir modeldeki, geçerliliğini yitirme tehlikesi, sinir ağlarında da bulunmaktadır. Gayrimenkul değerinin hesaplanması örneğinde, öğrenim seti içeriğinde yer alan, geriye dönük örüntülerin sağladığı ölçüde öngörü yapılabilmektedir. Öğrenim kümesinin oluşturulduğu, en güncel pazar koşullarının, geçen hafta, geçen ay veya altı ay öncesi ile aynı olduğu garanti edilemez. Her gün alınan ve satılan evler, daha önce öğrenim kümesinde yer almayan pazar güçleri yaratmaktadır. Talepteki bir artış veya düşüş, ya da bir enflasyon yükselişi mülkün değerini hızla değiştirebilir. Sinir ağı modelinin güncel tutulmasını güçleştiren iki faktör bulunmaktadır. İlki, model hazır kurallar şeklinde ifade edilmediğinden, güncelliğini yitirdiği açık şekilde görülmeyebilir. İkincisi, sinir

ağları oldukça sağlam. Bu nedenle performansındaki azalışı fark edilmeyebilir. Özetle, modelin zaman aşımına uğraması yavaş olmakta ve güncelleme gerekliliği çok net tespit edilememektedir.

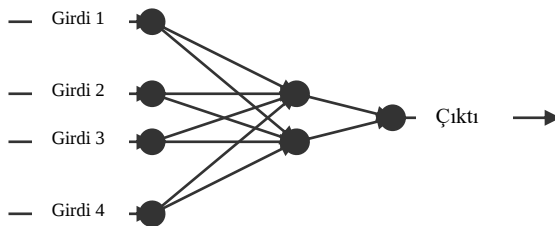
Çözümü sinir ağına daha güncel verileri dahil etmektir. Yöntemlerden biri aynı sinir ağını, yeni değerlerin eklenmesi ile oluşan bir öğrenim kümesi üzerinden yeniden optimize etmektir. Bu yaklaşım, ağının ürettiği sonuçların, doğruya oldukça yakın olduğu ve doğruluk yüzdesini artırmak için daha güncel örnekler ile beslenmesinin yeterli olduğu durumlarda kullanılır. Bir diğer yöntem, yeni oluşturulan öğrenim kümesi üzerinden, belki de farklı topolojideki yeni bir ağın optimize edilmesidir. Bu yaklaşım, pazar koşullarının aniden değiştiği ve önceki öğrenim kümesindeki örüntülerin geçerliliğini yitirdiği durumlarda uygulanır. Sonuç olarak, bir sinir ağının performansı, tümüyle gelişiminde kullanılan öğrenim kümesine bağlıdır.

8.3. Sinir Ağı Tanımı

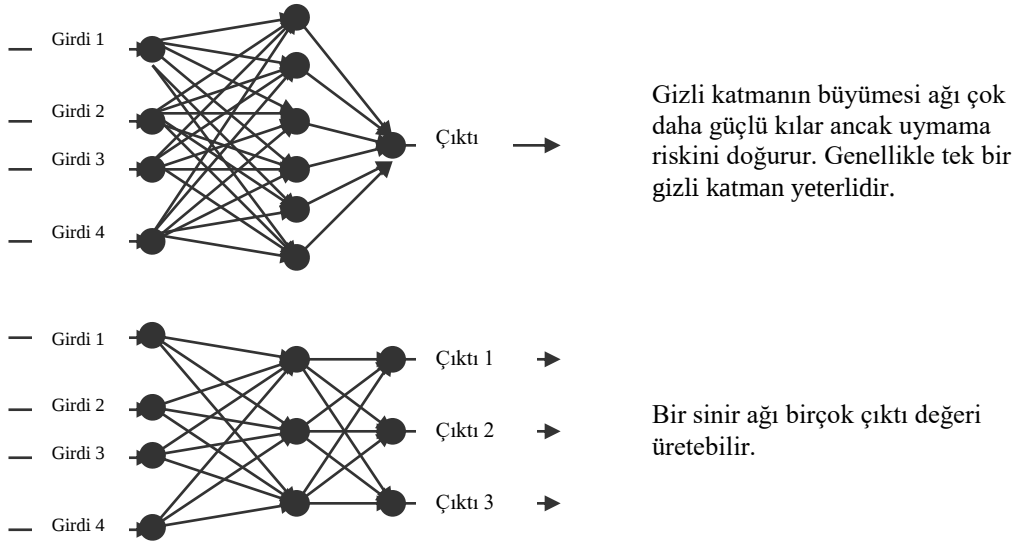
Sinir ağları, biyolojik nöronlar şeklinde modellenmiş temel birimlerden oluşurlar. Her bir birimin birçok girdisi bulunmakta ve bu girdiler, bir çıktı değeri oluşturacak şekilde birleştirilmektedir. Birimler, şekildeki gibi birbirlerine bağlanırlar ve bazı birimlerin çıktıları, başka birimlerin girdisi olabilmektedir. Şekildeki tüm örnekler ileri beslemeli sinir ağlarına örnek olup, girdiden çıktıya doğru tek yönlü akış arz ederler ve herhangi bir çevrim içermezler. [2]



Dört girdi alan ve tek bir çıktı üreten basit bir sinir ağı. Oluşan ağın sonucu, istatistikten bilinen lojistik regresyon ile eşdeğerdir.



Ağ, *gizli katman* olarak adlandırılan bir orta katman içermektedir. Gizli katman, daha fazla örüntünün tanınmasını mümkün kılması ile ağı daha güçlü kılar.



Şekil 8.2 İleri Beslemeli Sinir Ağı Örnekleri

(Kaynak : Berry, M. & Linoff,G.)

İleri beslemeli ağlar en basit ve kullanımı en yaygın olan ağ tipidir. Bu ağlarla ilgili sorulabilecek üç temel soru :

1. Birimler nedir ve nasıl davranırlar? Aktivasyon fonksiyonun tanımını verir.
2. Birimler birbirleriyle nasıl birleşirler? Bir ağın topolojisini tanımlar.
3. Ağ örüntüleri tanımayı nasıl öğrenir? Geri yayımı tanımlar.

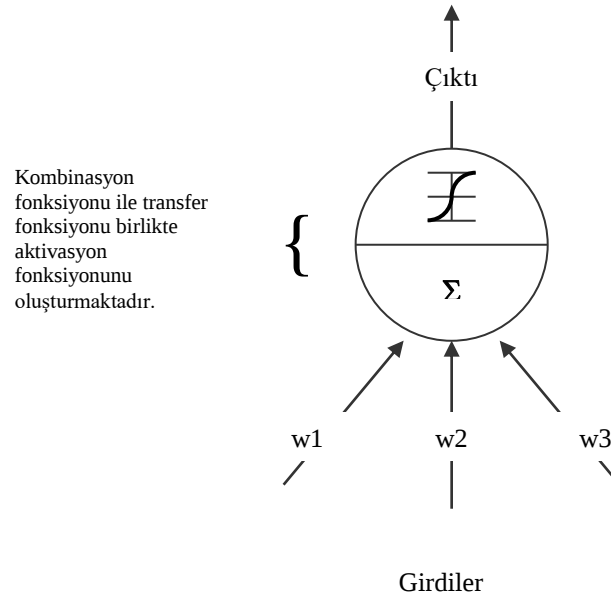
Bu soruların cevapları sinir ağlarının temelini oluştururlar ve bu güçlü veri madenciliği yönteminden en iyi sonuçları elde etme açısından fayda sağlayacaklardır.[5]

8.4. Sinir Ağı Biriminin Tanımı

Yukarıda da ifade edildiği gibi, sinir ağları, biyolojik nöronların davranışını modelleyecek şekilde tasarlanmış birimlerden oluşmaktadır. Bir birim, girdilerini, bir çıktıya dönüştürecek şekilde birleştirmektedir. Bu birleşim, birimin *Aktivasyon Fonksiyonu* olarak isimlendirilmektedir. En yaygın aktivasyon fonksiyonları, biyolojik modellere dayanır. Birleşen girdiler eşik değerine ulaşmaya kadar, çıktı değeri çok düşük kalmaktadır. Eşik değere ulaşıldığında, birim aktive edilir ve çıktı değeri yüksek olur.

Sinir ağındaki birimin girdilerindeki küçük değişimlerin (birleşen girdiler eşik değere yakın olduklarında), çıktı üzerindeki etkileri büyük, girdilerdeki büyük değişimlerin (birleşen girdiler eşik değere uzak olduklarında), çıktı üzerindeki etkileri de az olabilir. Küçük değişikliklerin bazen çok önem arz ettiği, bazen de etki etmediği durumlar *doğrusal olmayan davranış* olarak isimlendirilir.

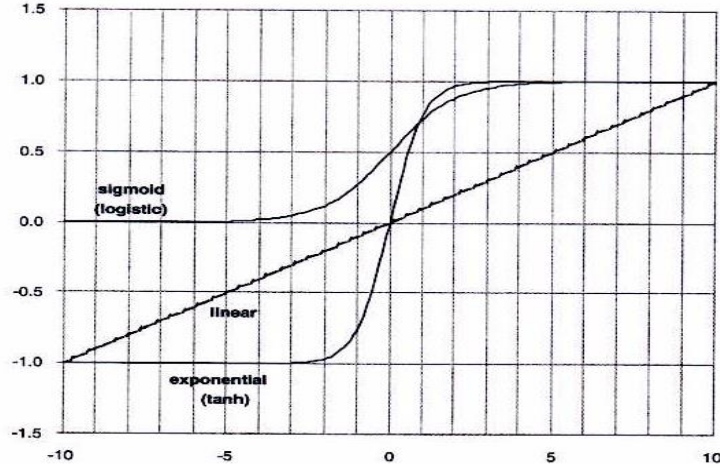
Aktivasyon fonksiyonu iki bölümden oluşmaktadır. İlk kısım tüm girdileri tek bir değerde birleştiren *kombinasyon fonksiyonudur*. Birimdeki her bir girdinin kendi ağırlığı vardır. En yaygın kullanılan kombinasyon fonksiyonu, her bir girdinin kendi ağırlığıyla çarpımının toplamlarından oluşan ağırlıklı toplamdır. Bazen de ağırlıklı girdilerin maksimumu, minimumu veya AND, OR gibi farklı kombinasyon fonksiyonları da kullanılmaktadır. Her ne kadar kombinasyon fonksiyonların seçiminde esneklik söz konusu olsa da, standart ağırlıklı toplam, birçok durum için doğru sonuç üretir. [2]



Şekil 8.3 Sinir Ağı Birimi

(Kaynak : Berry, M. & Linoff,G.)

Aktivasyon fonksiyonun ikinci kısmı, kombinasyon fonksiyonun değerini birim çıktısına dönüştüren *transfer fonksiyonudur*.



Şekil 8.4 Transfer Fonksiyonları

(Kaynak : Berry, M. & Linoff,G.)

Sigmoid, lineer ve hiperbolik tanjant olmak üzere üç tipik transfer fonksiyonu bulunmaktadır. Transfer fonksiyonun kendisi, almış olduğu belirli değerlerden daha önemlidir. Doğrusal transfer fonksiyonun uygun değerleri sınırlıdır. Sadece doğrusal transfer fonksiyonu içeren birimlerden oluşan ileri beslemeli bir sinir ağı, bir doğrusal regresyon' dan farklı sonuç üretmez. Sigmoid ve hiperbolik tanjant, doğrusal olmayan davranış sergileyen non-lineer fonksiyonlardır. İki fonksiyon arasındaki asıl farklılık ürettikleri çıktıların değer aralığıdır. Sigmoid fonksiyonun çıktı değerleri 0 ile 1 arasında değişirken, hiperbolik tanjant fonksiyonun ürettiği çıktı değerleri -1 il 1 arasındadır.

S şeklindeki sigmoid fonksiyonu diğerlerine göre daha yaygın kullanılmaktadır. Doğrusal olmadığı halde, sigmoid fonksiyonun davranışı istatistik uzmanlarının ilgisini çekmektedir. Tüm girdilerin ağırlıkları küçük olduğunda, kombinasyon fonksiyonun sonucu da -1 ile 1 arasında küçük bir değer olur. Bu aralık için sigmoid fonksiyonu hemen hemen doğrusaldır ve birim (veya tüm sinir ağı) yaklaşık doğrusal bir davranış sergiler. İstatistik uzmanları genellikle doğrusal sistemleri tercih etmekte ve yaklaşık doğrusal sistemler de aynı derecede kabul görür. Ağırlıklar büyüdükçe, sigmoid fonksiyonu -1 ya da 1 değerlerinde doyum noktasına ulaşır. Bu davranış, girdinin doğrusal bir modelden yavaş yavaş doğrusal olmayan bir model dönüşümüne uymaktadır. Özetle, sinir ağları, *doğrusal* (linear), *doğrusal olmayan* (non-linear) ve *yaklaşık doğrusal* (near-linear) olmak üzere üç tip problem için uygun öngöründe bulunabilir. [2]

Bir ağ farklı transfer fonksiyonları içeren birimlerden oluşabilir, ancak birçok durumda sigmoid fonksiyonu, varsayılan transfer fonksiyonudur. Sinir ağları için tasarlanan daha karmaşık programlar bazen farklı kombinasyon ve transfer fonksiyonlarının denenmesini mümkün kılar.

Sigmoid Fonksiyonu

Toplam fonksiyonun olabilecek bütün çıktıları için Sigmoid fonksiyonu 0 ile 1 arası değerler üretir.

Sigmoid Fonksiyonu

$$\text{Sigmoid}(x) = (1 + e^{-x})^{-1} \quad (8.1)$$

Sinir ağlarında kullanıldığında, x değişkeni, genellikle bir birimdeki girdilerin ağırlıklı toplamından oluşan kombinasyon fonksiyonunun sonucudur.

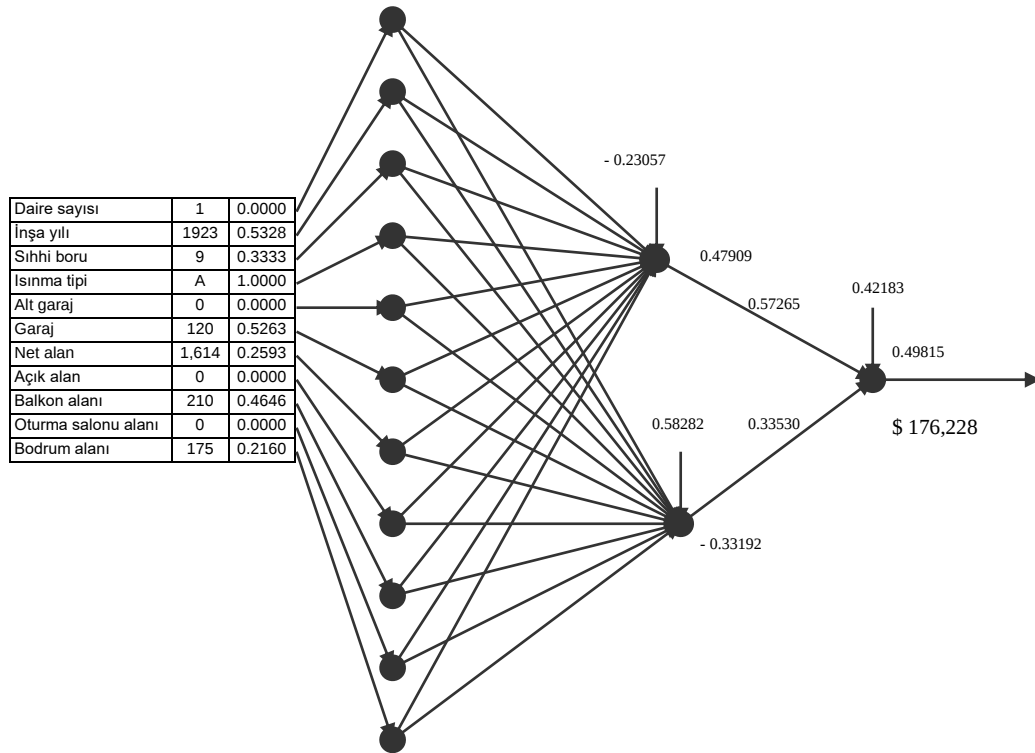
8.5. İleri Beslemeli Sinir Ağları

Şekil 8.5'te ileri beslemeli bir sinir ağının, girdi değerlerinden çıktı değerini nasıl hesapladığı gösterilmektedir. Bu ağın topolojisi veya yapısı, öngörü ve sınıflandırma için kullanılan ağlara özgüdür. Birimler üç katman şeklinde düzenlenmiştir. İlk katman, 0 ile 1 aralığındaki değerlere indirgenmiş girdilere bağlı. Bu birimler ağının *girdi katmanı*dır. Girdi katmanındaki her birim tek bir kaynağa bağlı.

İkinci katman, ağının ne girdilerine ne de çıktısına bağlı olmaması nedeniyle *gizli katman* olarak isimlendirilmektedir. Gizli katmanın her birimi, girdi katmanındaki tüm birimlere tam bağlıdır. Örnekteki ağ standart birimlerden oluştuğuna göre, gizli katmandaki birimler, çıktılarını, girdi değerlerin kendi ağırlıklarıyla çarpımının toplamını sigmoid fonksiyonu ile dönüştürerek hesaplamaktadır. Bir sinir ağı arzu edildiği kadar gizli katmandan oluşabilir, ancak genellikle bir gizli katman yeterlidir. Katmanı oluşturan birim sayısı arttıkça, ağının örüntüleri tanıma kapasitesi de artmaktadır. Ancak çok yüksek kapasitenin de bir dezavantajı bulunmaktadır. Sinir ağı, hafızaya almış olduğu öğrenim kümesindeki örneklerle bağlı olarak örüntüleri tanıyabilmektedir. Amaç, öğrenim kümesindeki örneklerin ezberlenmesi değil de, bunların genellenmesi olduğundan, gizli katmanın optimum sayıda birim içermesi idealdir.

Aşağıdaki örnekte, ara katmandaki her bir birim ek olarak, dışarıdan bir girdi daha almaktadır. Kombinasyon fonksiyonuna dahil olan bu girdi sabit ve değeri 1'dir. Ayrıca diğer girdiler gibi bir ağırlığı vardır. Sabit girdi, global bir dengeleyici rolünde, ağırlığın örüntüleri daha iyi tanınmasına yardımcı olmaktadır. Öğrenme aşamasında, ağ içerisindeki diğer ağırlıklarla birlikte bu tür girdilere ait ağırlıklar da ayarlanmaktadır.

Son katman, *çıktı katmanı*dır. Gizli katmanındaki tüm birimlere bağlıdır. Çoğunlukla, sinir ağı tek bir değer hesaplar, bu nedenle bu katman tek birimden oluşur, ürettiği değer de 0 ile 1 aralığında kalır. Çıktının anlaşılabilirliği için bu değer yeniden dönüştürülmesi gerekir. Örnekteki 0.49815 çıktısı, \$103,000 ile \$250,000 arasında bir değere dönüştürülmelidir. Karşılığı \$176,228 bu olan çıktı, evin gerçek değerine oldukça yakındır.



Şekil 8.5 Sinir Ağı'nın Öğrenme Prosesi

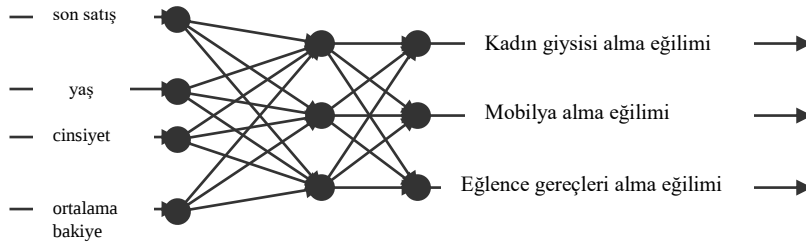
(Kaynak : Berry, M. & Linoff,G.)

Bazen çıktı katmanı birden fazla birimden oluşabilir. Örneğin, bir mağazalar zinciri, müşterilerinin, kadın giysisi, mobilya ve eğlence gereçleri gibi farklı mağazalardan alışveriş etme olasılığını öngörmek istemektedir. Üç mağaza, bu bilgiyi promosyon ve doğrudan hedefe yönelik mail kampanyaların planlanmasında kullanacaktır. Bu

öngörüyü yapabilmek için Şekil 8.6'daki sinir ağı kuruldu. Her bir departman için toplam üç çıktı bulunmaktadır. Çıktılar, girdilerde tanımlanan müşterinin ilişkili mağazalardan gelecekte yapacağı alışveriş eğilimini gösterir. Bir müşteriye ait girdiler verildiğinde, ağ tarafından üç değer hesaplanır. Elde edilen tüm bu çıktılarından, mağazaların en uygun promosyonlarını belirleyebilecek yaygın yöntemlerden birkaçı aşağıdaki şekildedir :

- En yüksek değerdeki birime tekabül eden mağaza seçilir;
- En yüksek üç değerdeki birimlere tekabül eden mağazalar seçilir;
- Belirli değerleri aşan birimlere tekabül eden mağazalar seçilir; veya
- En yüksek değerdeki birimin belirli bir yüzdesi kadar olan birimlere tekabül eden mağazalar seçilir.

Her duruma uygulanabilecek tek bir yöntem yok. Tümünün farklı durumlarda güçlü ve zayıf yönleri bulunmakta ve duruma göre her biri doğru sonuçlar üretmektedir. Pratikte, olası tüm yöntemler test kümesi üzerinde denenerek duruma en uygun olanı tespit edilir.



Şekil 8.6 Birden Fazla Çıktı Üreten Sinir Ağı Örneği

8.6. Geri Yayım ile Gerçekleşen Öğrenme Prosesi

Sinir ağlarının öğrenme prosesi, her bir biriminin girdilerine en uygun ağırlıkların atanmasıdır. Amaç, öğrenim kümesi kullanılarak, yine bu kümedeki mümkün olduğunca fazla örneğinin, beklenen çıktısına olabildiğince yakın çıktı oluşturacak ağırlıkların üretilmesidir. Bunu yapabilecek en yaygın yöntem John Hopfield tarafından geliştirilen *geri yayım* yöntemidir. [2]

Geri yayım yöntemi aşağıdaki üç adımdan oluşmaktadır :

1. Ağ öğrenim kümesindeki örnekleri alır ve mevcut ağırlıklarla çıktılarını hesaplar.
2. Geri yayım, 1.adımda elde edilen değer ile beklenen (gerçek değer) arasındaki farkı alarak hatayı hesaplar.
3. Geri bildirilen bu hatayı minimize edilecek şekilde ağırlıklar yeniden ayarlanır.

Tek bir birimdeki ağırlıkların ayarlanabilmesi için, hatanın ölçümü bu algoritmanın kritik bir parçasıdır. Her birim, hata üzerinde belirli bir sorumluluk taşır. Örneğin, çıktı katmanındaki bir birim hatanın tümünden sorumludur. Bu birim, hata ile ilgili sorumluluğunun bir kısmını gizli katmandan gelen girdilerine atar. Birden fazla katmanın olması durumunda sorumluluğunun paylaşımı bu şekilde geriye doğru devam eder. İşleyişi ile ilgili olarak, transfer fonksiyonun kısmi türevlerini gerektiren karmaşık matematiksel bir yöntem olduğunu söylemek yeterlidir.

Bir hata söz konusu olduğunda, bir birimin ağırlıklarını ayarlaması özetle şöyle gerçekleşir: Çıktının, birimin her bir girdisine ne derece duyarlı olduğunun ölçülmesi ile başlanır. Her bir girdideki ağırlığın değişimi hatayı artan yönde mi yoksa azalan yönde mi etkilediği değerlendirilir. Ardından birim her bir ağırlığını, hatayı minimize edecek ancak tümüyle yok etmeyecek şekilde ayarlar. Öğrenim kümesindeki her bir örnek için yapılan ayarlamalar, ağırlıkları optimum değerlerine biraz daha yaklaştırır. Daha önce de ifade edildiği gibi amaç, öğrenim kümesinin bire bir benzerini yaratmak yerine girdilerdeki örüntülerin belirlenmesi ve genellenmesidir. Değerlendirilen örnekler belirli bir sayıya ulaştığında, ağırlıkların değişimi ve hata değerinin azalışı gittikçe yavaşlar. Bu noktada ağın öğrenme süreci tamamlanır.

Ağırlıkların ayarlanması için kullanılan bu yöntem, *genelleştirilmiş delta kuralı* olarak isimlendirilir. Genelleştirilmiş delta kuralının kullanımı ile ilgili olarak iki parametre önem taşımaktadır. Bunlardan ilki, her birimdeki ağırlıkların artış veya azalış yönündeki değişim eğilimi ile ilgili *momentum* dur. Momentum her ağırlığın mevcut değişim yönünü aynı doğrultuda devam ettirmeye çalışır. Yüksek momentumlu bir ağ, ağırlıkları ters yönde değiştirecek yeni örneklerle oldukça yavaş cevap verir. Bu özellik, öğrenim kümesindeki örneklerin, benzerliklerine göre sıralı oldukları durumlarda fayda sağlar. Eğer momentum düşük ise ağırlıklar daha rahat kararsızlık gösterebilirler. [5]

Öğrenme oranı ağırlıkların hangi sıklıkta değiştiğini kontrol eder. Öğrenme oranı için en iyi yaklaşım büyük değerle başlayıp, ağın öğrenimine bağlı olarak da yavaş yavaş azalmasıdır. Başlangıçta ağırlıklar rasgele dağıldığından, büyük değişimler, uygun değerlere yaklaşımı sağlar. Ancak optimum çözüme yaklaşıldıkça öğrenme oranının azalması gerekir, böylece ağı en uygun ağırlıklarını ayarlayabilir.

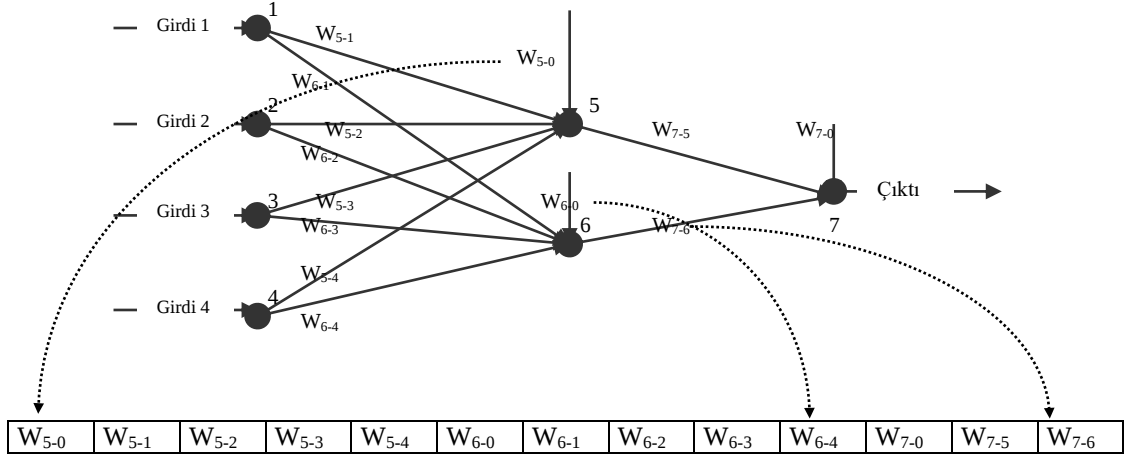
Araştırmacılar genelleştirilmiş delta kuralının yüzlerce varyasyonunu yarattılar. Her yaklaşımın avantajları ve dezavantajları bulunmaktadır. Her durumda, ağı hızlıca optimum çözüme ulaştıracak yöntem araştırılır. Bazı sinir ağları paketleri, kullanıcılarına problemlerine en uygun çözümü deneyebilmek için birkaç alternatif öğrenme yöntemi sunar.

Yerel optimum olarak adlandırılan ve tüm öğrenme yöntemlerinde karşılaşılabilecek bir tehlike bulunmaktadır. Bu olay, ağının öğrenim kümesi için geçerli tüm doğru sonuçları ürettiği ve ağırlıkların ayarlanması, performansı etkilemediği durumlarda meydana gelir. Bununla birlikte, daha iyi sonuç verebilecek ve ağı içerisindekinden oldukça farklı ağırlık kombinasyonları vardır. Öğrenme oranı ve momentumun kontrolünün birlikte ele alınması, en iyi global çözümün bulunmasını sağlar.

8.7. Genetik Algoritmaların Kullanımı ile Öğrenme

Sinir ağlarının öğrenme sürecinde genetik algoritmaların kullanımı hızla yaygınlaşıyor. Genetik algoritmaların bu amaca uyarlanması oldukça basit.

Genetik algoritmaların kullanımında ilk adım, bir kromozomun tanımlanmasıdır. Bir ağı içerisindeki tüm ağırlıklar, bir kromozomu oluşturacak şekilde, bir dizi halinde bir araya getirilir. Sonraki adımda kromozom için uygunluk fonksiyonu belirlenir. Verilen herhangi bir örneğin kromozomu değerlendirilebilir ve hatası belirlenebilir. (Hata, hesaplanan ve gerçek değer arasındaki farkın karesi veya mutlak değeridir.) Amaç, öğrenim kümesindeki hatanın minimize edilmesi olduğundan, kromozom ağırlıklarından oluşan ağı, tüm öğrenim kümesi üzerinde değerlendirilerek, tüm hatalar toplanır. O halde, uygunluk fonksiyonu, öğrenim kümesindeki tüm örnekler için oluşan hataların toplamıdır. Bunun üzerine, genetik algoritması, seçme, çaprazlama ve mutasyon yöntemleri ile uygunluk fonksiyonunu minimize edebilir.



Şekil 8.7 Ağırlıkların Genetik Kromozomlarda Toplanması

(Kaynak : Berry, M. & Linoff,G.)

Genetik algoritmaları, aynı zamanda ağ topolojisinin belirlenmesinde de rol oynayabilirler. Örneğin, ağırlıkların kodlanması yerine, gizli birim sayısı, momentum ve öğrenme oranı kodlanabilir. Bu senaryodaki ağ, kromozom içerisindeki parametreler kullanılarak, test kümesinin bir kısmı üzerinden yapılandırılacaktır. Yapılanma sonrasında, değerlendirme kümesi üzerinden değerlendirilerek, ağın uygunluk fonksiyonu hesaplanacaktır. Bu durumda yeni nesillerin yaratılması zaman kaybettiren bir prosestir, ancak sonuç, olabilecek en uygun ileri beslemeli ağa yakın olmalıdır.

Genetik algoritmaların sonuçları umut verici ve sinir ağları paketlerinde yerlerini almış durumdadır. Genetik algoritmaların kullanımı, sinir ağlarının hızlı ve etkili öğrenmesini sağlayan birçok farklı yöntemden biridir.

8.8. Öğrenim Kümesinin Seçimi

Öğrenim kümesi, öngörü veya sınıflandırma değerleri hazır bilinen kayıtlardan oluşmaktadır. İyi bir öğrenim kümesinin seçimi kritiktir. Zayıf bir seçim, ağın oluşumu için harcanan çaba ne olursa olsun kötü bir sonuca neden olur. Doğru bir kümenin seçiminde ise sadece birkaç noktaya dikkat edilmesi yeterlidir.

8.8.1. Öğrenim Kümesinin Değer İçeriği

Öğrenim kümesinin, karşılaşılabilecek özelliklerin tüm değer aralığını kapsamaması en önemli faktörlerden biridir. Gayrimenkul örneğinde, en yüksek veya en düşük fiyatta, büyük veya küçük, garajı olan veya olmayan evlerin tümünü içeriyor olması anlamına gelmektedir. Genellikle, öğrenim kümesinde, kesin ve sürekli değerli özellikler için, sırasıyla her bir seçenekten veya değer aralığından birkaç örneğin bulunması tercih edilir.

Sinir ağlarının girdi değerleri 0 ile 1 aralığında olmalıdır. Buna göre, tüm değerler bu aralıktaki bir değere indirgenmelidir. Bazı durumlarda, en düşük değer 0 ve en yüksek değerin 1'e tekabül etmeyebilir. Gayrimenkul örneğinde evlerin büyüklüğü 714 m² ile 4,185 m² arasında değişmekteydi. 714 değerinin 0 ve 4,185 değerinin de 1 ile eşleştirilmesi yerine uç noktaların civarında bir pay bırakılması daha iyi bir yaklaşımdır. En küçük değer 0.1 ve en büyük değer 0.9 ile eşleştiği durumda, gayrimenkul örneği için 280 m² ile 4,619 m² arasında, az daha küçük veya büyük evler de ağ tarafından değerlendirilebilir.

8.8.2. Özelliklerin Sayısı

Bir sinir ağının oluşumu için gerekli süre, doğrudan ağ tarafından kullanılacak özellik girdi sayısına bağlıdır. Özellik sayısı arttıkça, ağın optimum çözüme yaklaşma süresi uzar. Aslında, özellik sayısı arttıkça, ağın çözümden uzaklaşması gibi daha büyük bir problemle karşılaşılabilir. Öngöründe etki olasılığı bulunmayan özelliklerin çıkarılması, ağın öngörü gücünü belirgin şekilde artırır.

Öngöründe kullanılacak en önemli özelliklerin belirlenmesi için farklı yöntemler vardır. İstatistik korelasyonlar önem önceliklerin belirlenmesinde kullanılabilir. Bir diğer yöntem karar ağaçlarının kullanımıdır. Ağacın üst seviyelerindeki özelliklerin öngörü gücü daha yüksektir. Bu özellikler daha sonra sinir ağları için girdi olarak kullanılabilir.

8.8.3. Girdilerin Sayısı

Ağ içerisindeki özellik sayısının artması, verideki örüntülerin belirlenmesi için kullanılacak örneklerin de arttırılmasını gerektirir. Ancak, özellik sayısı ile öğrenim kümesinin büyüklüğü arasındaki ilişkiyi ifade edecek basit bir kural ne yazık ki yoktur.

8.8.4. Çıktıların sayısı

Öğrenme kümelerinde, girdi sayısı çıktı sayısından genellikle fazladır. Girdilerin doğru belirlenmesi, doğru bir çıktının elde edilmesini sağlar. Bununla birlikte, ağın olası tüm çıktıları için yeterli sayıda örneğin bulunması önemlidir. Ayrıca, her bir olası çıktıya ait örnek sayısı hemem hemen aynı olmalıdır.

Örneğin, bir diesel motorun arızalanma oranı, hileli kredi kartı işlemleri gibi nadir ancak maliyeti yüksek bir durumun sinir ağı tarafından tespit edilmesi isteniyorsa, öğrenim kümesinin söz konusu nadir durum için yeterli sayıda örnek bulundurduğundan emin olunmalıdır. Mevcut veriden tesadüfi örnekleme yapmak, nadir durumların, sık rastlananların içerisinde kaybolma ihtimalini arttırır. Bunu aşmak için, öğrenim kümesini daha fazla nadir durum örnekleri ile güçlendirmek gerekir. Bu tür problemlerde, 10,000 “iyi” ve 10,000 “kötü” örnekten oluşan bir öğrenim kümesi, rasgele seçilmiş 100,000 iyi ve 1,000 kötü örnekten oluşandan daha iyi sonuçlar verir. Bununla birlikte, tesadüfi bir örnekleme seti kullanıldığında, girdi ne olursa olsun sinir ağı büyük olasılıkla “iyi” sonucu üretecektir ve %99 oranında doğru işleyecektir.

Özet olarak, bir sinir ağının öğrenim kümesi, tüm özelliklerin alabileceği değerleri kapsayacak kadar büyüklükte olmalıdır. Her bir girdi özelliğine ait en azından birkaç örneğin bulunması istenir. Ağ çıktılarının, değerlerin düzgün bir dağılımını yansıttığından emin olunmalıdır. Bazı durumlarda, öğrenim kümesini yeni örneklerle besleyerek daha iyi sonuçların elde edilmesi sağlanır, ancak “kötü” durumların tespit edilmesi isteniyorsa, “iyi” örneklerin aşırılığına dikkat edilmelidir. Öğrenim kümesinin büyüklüğü ayrıca modeli çalıştıracak makinenin gücüne de bağlıdır. Bir sinir ağının öğrenme ve optimum çözüme ulaşma süresi, öğrenim kümesinin büyüklüğüne bağlıdır. Muhtemelen, bu sürenin farklı özelliklerin, indirgeme fonksiyonların veya ağ parametrelerin denenmesinde kullanılması daha iyi olabilir.

8.9. Verinin Hazırlanması

Çoğu kez girdi verilerin hazırlanması, sinir ağlarının kullanımındaki en karmaşık kısımdır. Veri madenciliği çalışması için gerekli doğru veri ve örneklerin seçimi karmaşıklığın bir parçasıdır. Diğerleri ise verilerin 0 ile 1 arasındaki bir değere indirgenmesidir. Veri biçimi ağı performansını doğrudan etkiler.

8.9.1. Sürekli (sabit ve kayar noktalı) Değerli Özellikler

Bazı özellikler, genellikle bilinen minimum ve maksimum sınırları arasında değişen sürekli değerler alırlar. Bu türden özelliklere örnek :

- Tutar (satış fiyatı, aylık bakiye, haftalık satışlar, gelir, vb.)
- Ortalama (ortalama aylık bakiye, ortalama satış hacmi, vb.)
- Oran (borç/gelir, maliyet/kar, vs.)
- Fiziki ölçüler (alan, sıcaklık, vb.)

Gayrimenkul örneğindeki sürekli değerli özelliklerin değerlendirilmesinde kullanılan yöntem en yaygın olanıdır. Bu değerler, önceden tanımlanmış minimum ve maksimum değer aralığında iseler, 0 ile 1 arasında bir değere indirgenebilirler.

$$\text{İndirgenmiş değer} = (\text{gerçek değer} - \text{minimum}) / (\text{maksimum} - \text{minimum})$$

Bununla birlikte pratikte, örneğin maksimum ve minimum değerlerin bilinmediği durumlarda, birtakım ek kabullere ihtiyaç duyulabilir. Gayrimenkul örneğinde, civarda 5,000 m² büyüklüğünde bir ev inşa edilebilir ve bu durumda önceden oluşturulan ağ kullanılamaz hale gelir.

Birkaç yöntem izlenebilir :

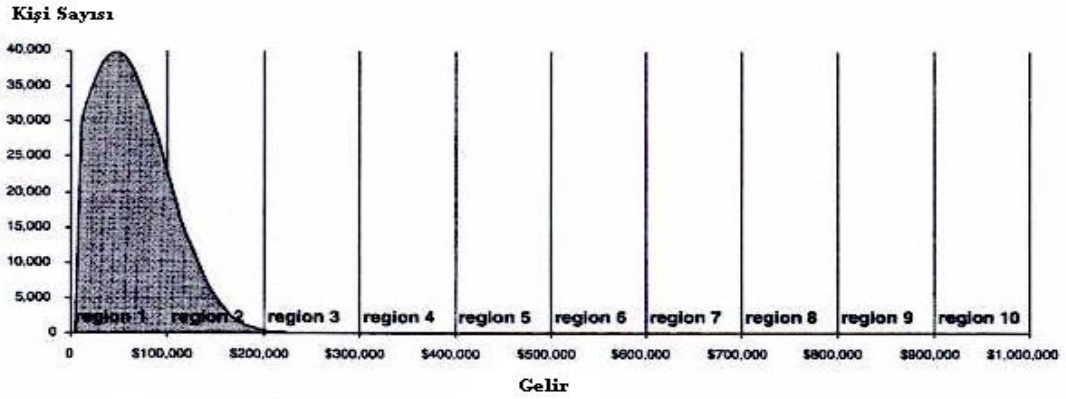
- Daha geniş bir aralık için planlama yapılabilir. Öğrenim kümesindeki net alan değerleri 714 m² ile 4,185 m² aralığında değişmekteydi. Aralığın alt ve üst sınırları için bu değerleri kullanmak yerine sırasıyla 500 ve 5,000 alınabilir.
- Aralığın dışındaki değerler dikkate alınmayabilir. Öğrenim kümesindeki aralık değerlerin ötesinde değer biçilmeye başlandı mı, sonuçların güvenilirliği azalır. Ağ, yalnızca önceden tanımlanmış değer aralıklarındaki girdiler için kullanılır. Bu, özellikle üretim sürecinin kontrolünde kullanılan bir ağ söz konusu ise, tümüyle hatalı sonuçların, büyük felaketlere yol açabileceğinden, önemlidir.
- Minimumdan düşük değerler minimuma, maksimumdan büyük değerler maksimuma sabitlenebilir. Böylece 4,000 m² büyük evlerin tümü aynı şekilde değerlendirilir. Bu yöntem birçok durumda doğru sonuç verir. Bununla birlikte, evin fiyatı ile büyüklüğü arasındaki korelasyon yüksek ise, diğer

koşullar aynı olmak şartıyla önceden belirlenen maksimum büyüklükteki evden %20 daha büyük bir evin fiyatı %20 daha fazla olmalıdır.

- Minimum ve maksimum değerleri, sırasıyla 0 ve 1 yerine, yine sırasıyla 0.1 ve 0.9 ile eşleştirilebilir.

Bazen değerlerin düzenli dağılmaması problemiyle karşılaşılabilir. Verilerin hemen hepsi 100,000 \$ altında, ancak değer aralığı 10,000 \$' den 1,000,000 \$' e kadardır. Değerlerin önerilen şekilde indirgenmesi sonucu 30,000 \$'lık bir gelir 0.0303'e ve 65,000 \$'lık bir gelir de 0.0657'e karşılık gelir. Gerçek değerler arasındaki farklılık, özellikle bir pazarlama uygulamasında çok önem arz etmesine rağmen, indirgenmiş değerler arasında hemen hemen bir fark olmadığı görülmektedir. Öte yandan, 250,000 \$ ve 800,000 \$ sırasıyla 0.2525 ve 0.8081' e indirgenmiş olur. Aradaki büyük farka rağmen, gelirdeki bu farklılık çok belirleyici olmayabilir. Gelirlerin çoğunun düşük tarafta yoğunlaşması, sinir ağının gelir alanından avantaj sağlamasını zorlaştırabilir. Bu tür dağılımlar çok önemli alanların etkin kullanımını engelleyebilir.

Çözüm olabilecek birkaç yöntem bulunmaktadır. En yaygın olanı, gelir gibi bir özelliği aralıklara ayırmaktır.



Şekil 8.8 Gelir Dağılımı Örneği

(Kaynak : Berry, M. & Linoff,G.)

Şekildeki gibi gelirin 10 eşit aralığa bölünmüş olması fayda sağlamıyor. Tüm fiili değerlerin ilk iki aralığa düşmesi nedeniyle, aralıkların aşağıdaki gibi tanımlanması daha doğru bir seçimdir.

10,000 \$ - 17,999 \$	Çok düşük
18,000 \$ - 31,999 \$	Düşük
32,000 \$ - 63,999 \$	Orta
64,000 \$ - 99,999 \$	Yüksek
100,000 \$ ve üzeri	Çok yüksek

Bu şekilde yapılan bir dönüşümün bilgi kaybına da neden olacağı kabul edilmedi. 65,000 \$'lık gelirli bir ev halkı, 98,000 \$'lık geliri olan bir ev halkında farksız olmaktadır. Diğer yandan, aralıklar sinir ağı tarafından daha kolay işlenebilmektedir. Diğer bir olasılık, dağılımlarını iyileştirecek şekilde verilerin filtre edilmesidir. Gelirlerin filtre edilmesinde logaritma fonksiyon kullanılabilir.

Tablo 8.3 Farklı Yöntemlerle İndirgenmiş Gelir Değerleri

Gelir (\$)	İndirgenmiş Gelir	Log(Gelir)	İndirgenmiş log
10,000	0.0101	4.0000	0.0000
18,000	0.0182	4.2553	0.1276
32,000	0.0323	4.5051	0.2526
63,000	0.0636	4.7993	0.3997
100,000	0.1010	5.0000	0.5000
250,000	0.2525	5.3979	0.6990
800,000	0.8081	5.9031	0.9515
1,000,000	1.0101	6.0000	1.0000

İki farklı şekilde indirgenmiş değerler kıyaslandığında, logaritma fonksiyonun kullanımı ile oluşan sonuçların gerçeğe daha yakın olduğu görülmektedir.

Son olarak, bazen mevcut özelliklerden yeni özellikler türetilmek istenebilir. Örneğin, menkul kıymetler borsasındaki fiyatların tahmin edilmesinde sinir ağıları kullanılırsa, sadece fiyatlandırmanın dikkate alınması ile yapılacak tahmindeki performansı hayal kırıklığı yaratabilir. Ağ, borsanın yükselme veya düşme trendinde olup olmadığını belirlemek için kullanılır ve geçmişteki fiyat değişimleri ile beslenirse daha gerçekçi bir tahmin yapılabilir. Bu durumda, fiyatlar yerine farkları girdi olarak kullanılmalıdır.

8.9.2. Sıralı, ayırık (tam sayı) Değerli Özellikler

Önceki bölümde sürekli değerlerin, ayırık değerlere dönüştürülebileceği görüldü. Diğer sıralı değerlere örnek olarak aşağıdakiler verilebilir :

- Adet (çocuk sayısı, satın alınan ürün sayısı, vb.)
- Yaş
- Sıralama kategorileri (düşük, orta, yüksek)

Sürekli özelliklerde olduğu gibi, bu tür özellikler de bir minimum ve bir maksimum değere sahiptirler. Örneğin, yaş genellikle 0 ile 100 aralığındadır, ancak gerçek aralık kullanılan veriye bağlı olarak değişebilir. Çocuk sayısı için, 4 üzerindeki değerlerin tümü 4 kabul edilerek, 0 ile 4 aralığı tanımlanır. Bu tür değerlerin indirgenmesi oldukça basittir. Tüm farklı değerlerin sayısı bulunur. Çocuk sayısı örneğinde, 0, 1, 2, 3, 4 olmak üzere toplam 5 farklı değer bulunmaktadır. Bu değerler 0 ile 1 arasında eşit olarak aşağıdaki şekilde dağıtılır.

$$0 \rightarrow 0.00, \quad 1 \rightarrow 0.25 \quad 2 \rightarrow 0.50 \quad 3 \rightarrow 0.75 \quad 4 \rightarrow 1.00$$

Birim aralığındaki değerlerle yapılan eşleştirme esnasında, gerçek sıralamanın korunması gerektiğine dikkat edilmelidir.

Ayrık, sıralı değerler için kullanılan bir diğer yöntem aşağıdaki şekildedir :

$$0 \rightarrow 10000 = 0.5000$$

$$1 \rightarrow 11000 = 0.7500$$

$$2 \rightarrow 11100 = 0.8750$$

$$3 \rightarrow 11110 = 0.9375$$

1' ler bir taraftan başlayıp ardışık şekilde, değeri kadar artmaktadır. Örnekte, 1 ile 16 arasındaki değerlerin ikili sistemdeki karşılıkları kodlanmıştır. Bu yöntem, akademik derecelendirme ve tahvil değerlendirme gibi, çizelgenin bir tarafındaki farkın, diğer taraftakine göre daha az önem arz etmesi durumlarında fayda sağlar. Örneğin, D ile F arasındaki farkın, A ile B arasındakinden oldukça önemli ise, yukarıdaki kodlama şekliyle, A, B, C, D ve F dereceleri için sırasıyla 0.9375, 0.8750, 0.7500, 0.5000 ve 0.0000 değerleri atanacaktır. Ancak 8 den daha yüksek sayıda farklı değere sahip bir özelliğe uygulandığında, son değerlerin 1'e çok fazla yaklaşması nedeniyle yöntemin etkinliği azalmaktadır.

Bu yöntem, önceki bilginin kodlama sistemine dahil edilmesini sağlar. Kodların yakınlığı, ilgili değerler arasındaki ilişkiyi tanımlar.

8.9.3. Kesin Değerli (Kategorik) Özellikler

Bir kategoriye ait özelliklerin değerleri sıralı değildir. Değerleri bir sıralamaya tabi tutmak, gerçekçi olmayabilir. Bu türden değerlere örnek olarak, en çok rastlananlar aşağıdaki şekildedir :

- Cinsiyet, medeni durum, vb.
- Durum kodları
- Ürün kodları
- Posta kodları, vb.

Posta kodları, sayısal değerler olmalarına rağmen, gerçekte farklı coğrafi bölgeleri temsil ederler ve bu bölgeler hakkında çok az bilgi verirler. Bu nedenle, 10014 posta kodunun 02116' ya, 95407'ye olduğundan daha yakındır sonucuna varılamaz. Sıralamanın belirleyici bir unsur olmaması nedeniyle, posta kodları, sıralı, ayırık değerler grubunda değil de, bu grupta değerlendirilmeli.

Kategorik değerlerin işlenmesi ile ilgili iki temel yöntem bulunmaktadır. Birincisi, seçenekleri ayırık, sıralı değerler gibi değerlendirerek, bunlara önceki bölümde anlatılan yöntemlerle, değerlerin atanmasıdır. Sinir ağları değerlerin sıralı olduğunu kabul etmektedir. Medeni durum için “bekar”, “boşanmış”, “evli”, “dul”, ve “bilinmeyen” seçeneklerine sırasıyla 0.00, 0.25, 0.50, 0.75 ve 1.00 değerleri atanmış olsun. Ağ açısından “bekar” ile “bilinmeyen” birbirlerine çok uzak iken, “boşanmış” ile “evli” seçenekleri oldukça yakındır. Birçok girdi alanı için bu türden bir sıralamanın fazla etkisi olmayacaktır. Değerlerin birbirleriyle ilişkileri bulunduğu durumlarda ise, bu yöntem, ağı yanıltabilir.

İkinci yöntem, bir kategorinin her bir değeri ayrı bir özellik olacak şekilde ayrıştırmaktır. Cinsiyet için “bay”, “bayan” ve “bilinmeyen” olmak üzere üç değer olsun. Bunlar, 0.00 veya 1.00 değerini alabilecek üç ayrı özellik şeklinde tanımlanabilir.

Tablo 8.4 Kategorik Değerlerin İndirgenme Örneği

Cinsiyet	Cinsiyet Bay	Cinsiyet Bayan	Cinsiyet Bilinmeyen
Bay	1.00	0.00	0.00
Bayan	0.00	1.00	0.00
Bilinmeyen	0.00	0.00	1.00

Bu yöntemle ağın girdi değişkenleri artmaktadır. Bu durum sinir ağlarının performansı açısından tercih edilmediğinden, “bay” veya “bayan” olmayanlar “bilinmeyen” şeklinde tanımlanarak özellik sayısı azaltılabilir.

Yukarıda bahsi geçen türlerin dışında da özellikler bulunabilir. Tarih bilgisi, bu tür özelliklere uygun bir örnektir. Herhangi bir tarih, gün veya saniye cinsinden ifade edilerek, ağı besleyecek şekilde indirgenebilir değerlere dönüştürülebilir. Bununla birlikte, bir işlem tarihi söz konusu ise, haftanın veya ayın kaçınıcı günü olduğu, gerçek tarihten daha fazla önem taşımaktadır. Örneğin, ay bilgisi, veri içerisindeki mevsimsel trendin belirlenmesi açısından önemlidir.

8.10. Sonuçların Yorumlanması

Çıktıların yorumlanabilmesi için, girdilerin indirgenmesinde kullanılan yöntemlerin tersi kullanılır. Örneğin, bir evin değerini hesaplayacak bir ağ ve çıktıları, 103,000 \$'lık bir değer 0.1'e, 250,000 \$'lık bir değer de 0.9' a tekabül eden bir öğrenim kümesi olsun. Model, yeni bir evin değerinin hesaplanmasında kullanıldığında, üretilen çıktı değeri 0.5 ise buna karşılık gelen evin gerçek değeri 176,500 \$ dır. Bu türden bir ters dönüşüm, sinir ağlarının sürekli değerlerin tahminindeki kullanımını oldukça basitleştirmektedir.

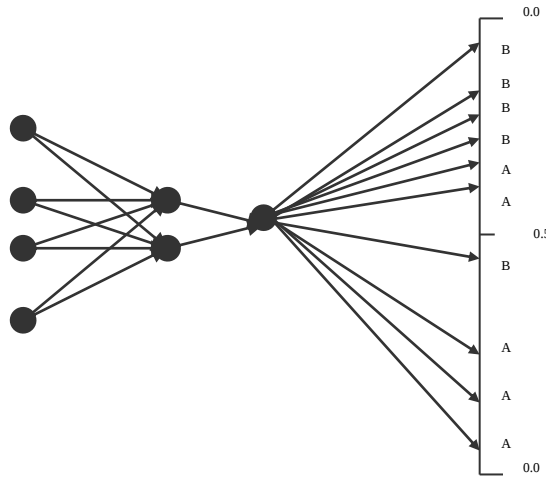
Diğer yandan, bir müşterinin kampanyaya dahil edilip edilmeyeceği veya hisselerin satılıp satılmayacağı gibi kesin değerlerin yorumlanması oldukça güçtür. Ağın sürekli değerler üretmesi nedeniyle, bu amaçla kullanıldığında, çıktıların değer aralıkları belirlenmelidir. İdeali, ağın düşük değerler için 0, yüksek değer için ise 1 üretmesidir. Ancak bazı girdiler için ağın orta değerler üretmesi, çıktının yorumlanmasını güçleştirmektedir. Basit olarak, 0.5 altındaki çıktıların bir değere, 0.5 üzerindeki çıktıların diğer değere (0.5 herhangi birine) tekabül edeceği kabul edilebilir. Uygulanması ve anlaşılması basit olmasına rağmen, en iyi çözüm değildir. İkinci bir yaklaşım, 0.33 altındaki çıktıların bir değere, 0.67 üzerindeki çıktıların diğer değere karşılık gelmesi ve aradaki çıktıların ise “bilinmeyen” olarak tanımlanması şeklinde, aralığı üçe bölmektir.

Bir diğer yaklaşım, değerler arası güven derecesinin atanmasıdır.

Tablo 8.5. Kategorik Değer Çıktılarının Güven Payları

Çıktı Değeri	Kategori	Güven payı
0.0	A	100%
0.2	A	80%
0.49	A	51%
0.51	B	51%
0.8	B	80%
1.0	B	100%

En uygun yaklaşımın belirlenmesi amacıyla, öğrenme kümesi ile birlikte, sonuçların yorumlanması için, çalışmaya test kümesi dahil edilmektedir. Test kümesi üzerinde uygulanan bir sinir ağının tipik sonuçları Şekil 8.9’da gösterilmiştir. Veriler A ve B olarak sınıflandırılmıştır. A kümesi elemanlarının çoğu alt kısımda ve B kümesi elemanlarının çoğu üst kısımda yer almaktadır. Bununla birlikte istisnaların bulunduğu dikkat edilmelidir. Şekilden, 0.62’den küçük değerlerin A kümesinde ve 0.64’ten büyük değerlerin B kümesinde olması gerektiği ve 0.63 değerinin sınır olarak seçilebileceği sonucu çıkar.



Şekil 8.9 Test Kümesi Üzerinde Uygulanan Bir Sinir Ağının Sonuçları

(Kaynak : Berry, M. & Linoff,G.)

İkili değerler için bir başka yaklaşım, biri A sınıfı, diğeri B sınıfı için iki çıktı üreten bir ağın kullanılmasıdır. Öğrenim kümesindeki A sınıfı öğeleri için, A sınıfını belirleyen çıktının 1, B sınıfını belirleyen çıktının 0 değerini, B sınıfı öğeleri için ise bunların tersini üretecek şekilde bir ağ oluşturulur. Sonuçların yorumlanması sırasında, her bir değer güven seviyesi olarak alınabilir. Buna göre, 0.8 ve 0.2 çıktı değerlerini taşıyan bir örneğin, %80 ihtimalle A sınıfı, %20 ihtimalle de B sınıfına ait

olabileceği söylenebilir. A ve B sınıflarını belirleyen çıktıları sırasıyla 0.1 ve 0.3 olan bir örnek aynı şekilde, %30 ihtimalle B sınıfına dahil edilebilir. B sınıfında olma ihtimali, A'da olma ihtimalinin üç katı olması nedeniyle güven seviyesi %75'e yükseltilebilir. Optimum sınır değerleri, ağır test kümesi üzerinde test edilmesi ile elde edilir.

İkiden fazla seçeneğin incelenmesi durumunda da yaklaşım benzerdir. Örneğin, bir uzun mesafe nakliye şirketi, hedefe uygun aşağıdaki üç hizmet teklifi ile yeni müşteriler kazanmayı hedeflemektedir.

- Tüm uluslararası taleplerde indirim
- Uluslararası olmayan tüm uzun mesafe taleplerinde indirim
- Önceden belirlenen müşterilerin taleplerinde indirim

Nakliye şirketi, müşterilere üç paket için cazip teklifler sunmayı planlamaktadır. Tekliflerin maliyeti yüksek olması nedeniyle, kampanyadan kar elde edilebilmesi için doğru müşteriye doğru hizmetin önerilmesi gerekmektedir. Üç ürünü tüm müşterilere sunmak çok maliyetli, daha kötüsü ise müşterilerin doğru belirlenmemesi durumunda geri dönüş oranının düşmesidir.

Nakliye şirketi, ürünlerin pazar araştırmasını, üç ürün teklifini alan ancak sadece bir tanesini cevaplayan, küçük bir müşteri kümesi üzerinde yapmaktadır. Elde edilen sonuç doğrultusunda, her bir teklifin cevaplanma oranının tahmininde kullanılacak bir modelin kurulması amaçlanmaktadır. Test amaçlı yapılan pazarlama kampanyasından oluşturulan öğrenim kümesi, aşağıdaki kodlamaları kullanmaktadır:

cevap yok → 0.00; 1.teklif → 0.33; 2.teklif → 0.67; 3.teklif → 1.00

Müşteri bilgilerinden oluşturulan bir sinir ağı sonrasında nakliye şirketi modeli uygulamaya başlamaktadır. Ancak modelin uygulanmasından beklenildiği kadar iyi sonuç alınamadı. Birçok müşteri, ağı oluşumunda kullanılan dört değer etrafında kümelenmesine rağmen, bazı durumlarda ağ tarafından 0.5 ve 0.83 gibi ara değerler üretilmiştir.

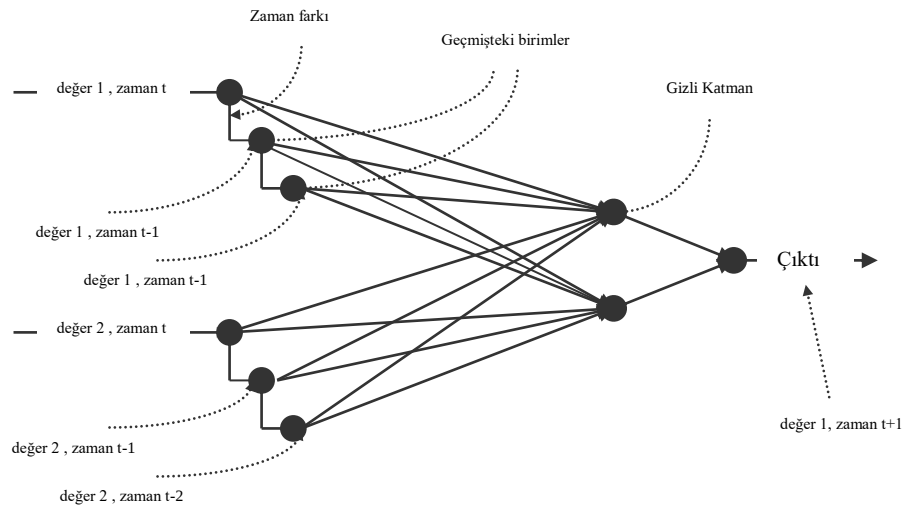
Bu durumda nakliye şirketi bir test kümesi yardımıyla çıktı değerlerini araştırmalıdır. Çıktı değerlerin yorumlanması sonucu doğru aralıklar tespit edilerek, bu çıktıların pazarlama segmentlerine dönüşümü sağlanır.

Örnekle ilgili bir diğer tespit, ağı, müşterilerin her bir kampanyayı cevaplaması gibi aslında üç farklı konuda tahminde bulunduğudır. Bu nedenle, ağın her bir kampanya için ayrı bir çıktı üretecek şekilde yapılandırılması önerilmektedir. Test kümesi de cevaplamayanlar sınırının belirlenmesinde kullanılmalıdır.

8.11. Zaman Serileri İçin Sinir Ağları

Birçok veri madenciliği problemlerinde veri, doğal olarak bir zaman serisinde sıralanmaktadır. Bu tür serilere örnek olarak, günlük IBM hisse senedinin kapanış fiyatı veya CHF/USD döviz kuru verilebilir. Bir sonraki değeri ya da serinin yükselme veya düşme eğiliminde olduğunu tahmin edebilen bir yatırımcı, çok büyük bir avantaj yakalamış olmaktadır.

Çoğunlukla mali serilerde kullanılması ile birlikte, bir müşterinin terk etme eğilimin tespiti amacıyla tüm işlem kayıtlarının analiz edilmesi veya bir motorun arızalanma ihtimalinin öngörülmesi amacıyla yapılan fiziksel ölçümler sırasında yine zaman serileri kullanılmaktadır. Sinir ağları zaman serilerine kolayca uyarlanabilir. Proses Şekil 8.10'da resmedilmiştir. Ağ, en eski noktadan başlayarak, zaman serisi verileri üzerinde eğitilir. Öğrenme ikinci en eski noktaya geçer, ilk en eski nokta ise sonraki birim kümesindeki girdi katmanında yerini alır ve bu şekilde devam eder. Ağ her adımda, serideki bir sonraki değeri tahmin etmeye çalışan ileri beslemeli, geri yayınlı bir ağ şeklinde eğitilir. [2]



Şekil 8.10 Sinir Ağlarının Zaman Serilerine Uyarlanması

(Kaynak : Berry, M. & Linoff,G.)

Zaman serileri ağı, sadece bir zaman serisindeki verilerle sınırlı kalmayıp, birden fazla girdi alabilir. Örneğin CHF/USD döviz kurunun tahmin edilmesi sırasında, önceki gün işlem hacmi, USD/DEM döviz kuru, menkul kıymetler borsasının kapanışı ve haftanın günü gibi başka zaman serileri bilgisi de dahil edilebilir. Ayrıca, incelenen döneme ait enflasyon oranları gibi zaman serisi oluşturmeyen veriler de eklenebilir.

Geçmişteki birim sayısı, ağın tanıyabileceği örüntü büyüklüğünü kontrol eder. Örneğin, bir hisse senedinin kapanış değerini tahmin edecek bir ağın içerisine geçmişteki 10 birim alındığında, iki haftalık zaman aralıklarında gerçekleşen örüntülerin tanınması sağlanır. Bu ağ, hisse senedinin ilerideki üç ay sonrası değeri hakkında büyük olasılıkla doğru bir tahminde bulunamaz.

Gerçekten, düzgün ileri beslemeli, geri yayınlı bir ağının kullanılması ve giriş verilerinin değiştirilmesi ile zamana bağlı bir ağın etkisi elde edilebilir. Aşağıdaki tabloda görülen ve haftanın günleri ile kapanış fiyatlarını konu alan 10 veriden oluşan zaman serileri olsun.

Tablo 8.6 Zaman Serisi Örneği

Veri ögesi	Haftanın günü	Kapanış fiyatı
1	1	\$ 40.25
2	2	\$ 41.00
3	3	\$ 39.25
4	4	\$ 39.75
5	5	\$ 40.50
6	1	\$ 40.50
7	2	\$ 40.75
8	3	\$ 41.25
9	4	\$ 42.00
10	5	\$ 41.50

Geçmişe dönük üç zamanlı bir zaman serisini oluşturmak için, önceki değerlerden oluşan yeni özellikler eklemek yeterlidir. Aşağıdaki şekilde oluşan veriler, zaman serileri için özgü bir işleme gerek kalmaksızın, ileri beslemeli, geri yayınlı bir ağın girdisi olabilmektedir.

Tablo 8.7 Geçmişe Dönük Üç Zamanlı Bir Zaman Serisi Örneği

Veri ögesi	Haftanın günü	Kapanış fiyatı	Önceki kapanış fiyatı	Daha önceki kapanış fiyatı
1	1	\$ 40.25		
2	2	\$ 41.00	\$ 40.25	
3	3	\$ 39.25	\$ 41.00	\$ 40.25
4	4	\$ 39.75	\$ 39.25	\$ 41.00
5	5	\$ 40.50	\$ 39.75	\$ 39.25
6	1	\$ 40.50	\$ 40.50	\$ 39.75
7	2	\$ 40.75	\$ 40.50	\$ 40.50
8	3	\$ 41.25	\$ 40.75	\$ 40.50
9	4	\$ 42.00	\$ 41.25	\$ 40.75
10	5	\$ 41.50	\$ 42.00	\$ 41.25

8.12. İleri Beslemeli, Geri Yayınımlı Ağların Kullanımı İle İlgili Bulgular

Karmaşık sinir ağları paketlerinden dahi, iyi sonuçlar alınabilmesi, belirli bir çaba gerektirir. Bu bölüm, iyi sonuçlar üretecek bir ağın kurulumundaki bazı bulguları içermektedir.

En önemli karar, gizli katmandaki birim sayısının belirlenmesidir. Birim sayısı arttıkça, ağ tarafından tanınabilecek örüntü sayısı da artmaktadır. Ancak, ağın öğrenim kümesi üzerinde genelleme yapmak yerine, kümenin tamamını ezberleme sakıncası bulunmaktadır. Bu durumda, birim sayısı fazlalığı istenilen aksine iyi sonuç üretememektedir. Bununla birlikte ağın gereğinden fazla eğitildiği durumların tespiti mümkün. Eğer bir ağ, öğrenim kümesi üzerinde iyi, test kümesi üzerinde ise kötü sonuçlar üretiyorsa, öğrenim kümesindeki örneklerin ağ tarafından ezberlendiğine işaret eder.

Gizli katmanın büyüklüğü kritik bir faktördür. Pratik olarak, asla girdi katmanının iki katından fazla olmamalıdır. Girdi katmanı büyüklüğünde bir gizli katman ile başlanmalıdır. Eğer ağ gereğinden fazla eğitilmiş ise, katmandaki birim sayısı azaltılır, ağın eğitimi tamamlanmamış ise birim sayısı arttırılır. Sınıflandırma amaçlı kullanılan bir ağ için, her bir sınıfa ait bir birim içeren, bir gizli katman ile başlanmalıdır.

Bir başka önemli karar, öğrenim kümesinin büyüklüğüdür. Öğrenim kümesi, her bir özelliğin alabileceği girdi değer aralıklarını içerecek büyüklükte olmalıdır. Ayrıca, ağ içerisindeki her bir ağırlık için birkaç örnek bulunabilir. S girdi birimi, h gizli birimi ve 1 çıktısı bulunan bir ağın, $n*(s+1) + h + 1$ adet ağırlığı olur. Örneğin, 15 girdi özelliği, 10 gizli birimi bulunan bir ağın ağırlık sayısı 162'dir. Ağın her bir ağırlığı

için, en azından 5 ile 10 arasında örnek isteniyorsa, öğrenim kümesindeki toplam örnek sayısı minimum 810 olmalıdır.

Son olarak, ağ tarafından doğru sonuçların üretilebilmesi açısından, öğrenme oranı ve momentum parametreleri önem taşımaktadır. Başlangıçta, ağın ağırlıklarındaki büyük ayarlamalar nedeniyle öğrenme oranı yüksek olmalıdır. Öğrenme prosesi ilerledikçe, en uygun ağırlık değerlerine ulaşılması nedeniyle bu oran düşmelidir. Momentum parametresi ise, etkisi az olan ağırlıklardaki karasızlıkları önleyerek, ağın daha hızlı çözüme yaklaşmasını sağlar.

8.13. Sinir Ağının Anlaşılması

Sinir ağları saydam değildir. Ağ içerisindeki tüm birimlerin ağırlıkları bilinse dahi, ağın çalışması konusunda fazla bilgi vermez. İnsan düşüncelerinin, beyindeki nöronlardan ne şekilde geliştiği açıklanamadığı gibi, ağın anlaşılmasının da biraz felsefi bir çekiciliği bulunmaktadır. Sinir ağının saydam olmayışı, ürettiği sonuçların anlaşılmasını güçleştirmektedir.

Ağın kararlarında etkili olan kurallar sorgulanabilir, ancak birimlere güç kazandıran, doğrusal olmayan özellikleri, aynı zamanda bu birimleri basit kurallar üretmez hale getirir. Kural çıkarımı konusundaki araştırmalar kaçınılmaz iyi sonuçlar getirecektir. Her şeyden önce ağın kendisi bir kuraldır ve içerisinde olup biteni anlamak için başka yöntemlere ihtiyaç vardır.

Duyarlılık analizi tekniği, saydam olmayan modellerin yorumlanmasında kullanılabilir. Duyarlılık analizi, kuralları açık şekilde sağlamıyor, ancak girdilerin sonuçlar üzerindeki izafi önemini işaret etmektedir. Duyarlılık analizinde, her bir girdinin ağ üzerindeki duyarlılığının tespiti için test kümesi kullanılmaktadır.

Temel adımlar şöyledir :

1. Her bir girdinin ortalama değeri bulunur. Test kümesindeki değerler düzgün bir dağılım sergilemediğinden, ortalama değer genellikle tam 0.5'e karşılık gelmeyecektir. Bu ortalama değer test kümesinin merkezi gibi düşünebilir.
2. Tüm girdilerin ortalama değerleri ile oluşan çıktı ölçülür.
3. Her seferinde bir girdinin değerini değiştirmek üzere, her birinin minimum ve maksimum değerleri için oluşan çıktı ölçülür.

Bazı girdilerin minimum, ortalama ve maksimum deęerleri iin ıktı ok az deęiřkenlik gsterir. Bu durumda aę, sz konusu girdilere karřı duyarlı deęildir. Bazı girdiler ise aęın rettięi ıktı zerinde ok etkilidirler. Aę bu tr girdilere karřı duyarlıdır. ıktının deęiřim miktarı, aęın her bir girdiye olan duyarlılıęın ltdr. Tm girdiler iin bu ltler kullanılarak, her bir zellięin nemi llebilir.

Bu iřleyiř deęiřebilir. İki veya  zellięin deęeri aynı anda deęiřtirilerek, bu kombinasyonun zel bir neminin olup olmadıęı izlenebilir. Bazen, test kmesinin merkezi yerine, belirli bir noktadan bařlamak daha faydalı olabilir. rneęin, aęın sınır deęerlere karřı olan duyarlılıęını izleyebilmek amacıyla, zelliklerin maksimum ve minimum deęerleri iin analiz tekrarlanır. Eęer duyarlılık analizi  durum iin olduka farklı sonular retiyorsa, zelliklerin kombinasyonundan avantajlar elde eden aęın ierisinde daha fazla sıralama etkileri grlr.

İleri beslemeli, geri yayınımlı bir aę kullanıldığında, her bir zellięin baęımsız řekilde test edilmesi yerine, genellikle ęrenme fazında kullanılan hata lmleri sırasında duyarlılık analizinden avantaj saęlanabilir. Aęın test kmesi zerinden rettięi ıktı ile beklenen ıktı kıyaslanarak hata hesaplanır. Aę hatayı yeniden birimler arasında daęıtmaktadır. Ama herhangi bir aęırlıęın ayarlanması deęil, her bir girdinin duyarlılıęını izleyebilmektir. Hata, her bir girdinin, ıktıyı ne derece etkiledięini belirleyerek, duyarlılık iin vekalet etmektedir. Test kmesi zerinden elde edilen tm duyarlılık sonuları bir araya getirilerek, ıktıyı en fazla etkileyen girdiler belirlenir.

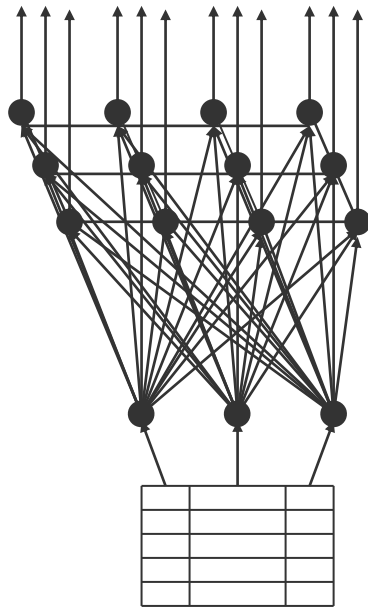
8.14. Dolaylı Veri Madencilięinde Sinir Aęlarını Kullanımı

Sinir aęları ayrıca dolaylı veri madencilięi alıřmalarında kullanılabilir. Kendini dzenleyen haritalar (SOMs : Self-organizing maps) arařtırmacı Dr. Tuevo Kohonen tarafından icat edilmiřtir. Asıl resim ve ses iin kullanılmalarına raęmen bu aęlar veri ierisinde gizlenmiř kmeleri tanıyabilirler. İleri beslemeli, geri yayınımlı aęlar ile benzer temel birimlere dayanmasına raęmen, kendini dzenleyen haritalar iki ynden olduka farklıdır. Topolojileri farklıdır ve ęrenmenin geri yayılım yntemi uygulanamıyor. [2]

8.14.1. Kendini Düzenleyen Haritalar (SOM)

Kendini düzenleyen haritalar, veri içerisindeki bilinmeyen örüntüleri tanıyabilen, kümelerin belirlenmesinde kullanılabilen sinir ağlarının özel bir türüdür. Tüm diğer ağlar gibi, kendini düzenleyen haritalar' da temel olarak bir girdi katmanı, bir de çıktı katmanına sahiptir. Girdi katmanındaki her birim bir kaynağa bağlıdır. Ayrıca her bir birim, her bir girdi bağlantısı ile eşleşen, bağımsız bir ağırlığa sahiptir. Kendini düzenleyen haritalar ile ileri beslemeli, geri yayınlımlı ağlar arasındaki benzerlikler sadece bu kadar.

Çıktı katmanı, diğerlerin aksine birçok birimden oluşmaktadır. Çıktı katmanındaki her bir birim, girdi katmanındaki tüm birimlere bağlıdır. Çıktı katmanı grid şeklinde düzenlenmiştir. Bu katmandaki birimler birbirlerine bağlı olmasalar da, gride benzer yapı, bu ağların eğitilmesi sırasında önemli rol oynamaktadır.



Ağın çıktısı için tüm çıktı birimleri birbirleriyle rekabet halindedir.

Çıktı katmanı grid şeklinde tasarlanmıştır. Bu katmandaki birimlerin birbirleriyle bağı bulunmamakta ancak, her biri tüm girdi birimlerine bağlıdır

Girdi katmanı girdilere bağlıdır.

Şekil 8.11 Kendini Düzenleyen Haritalar (Self-Organizing Maps)

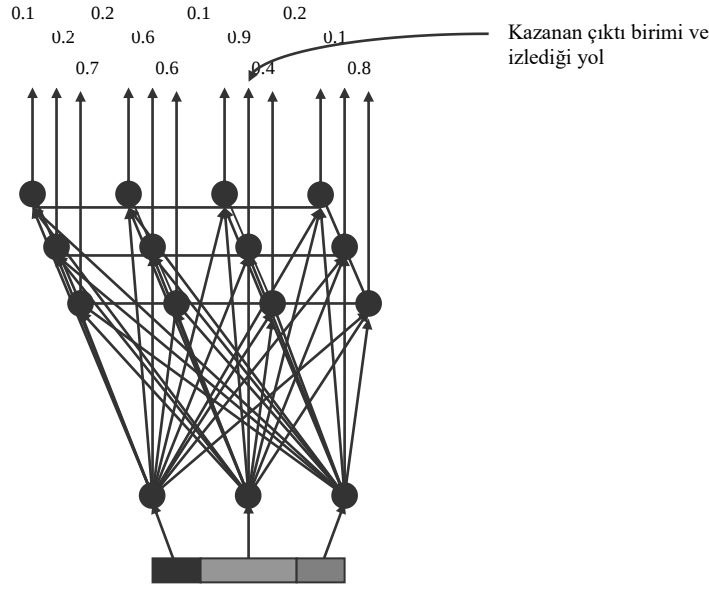
(Kaynak : Berry, M. & Linoff, G.)

Kendini düzenleyen haritaların, örüntüleri ne şekilde tanımlayabildiği bir örnekle açıklanacaktır. Bir eğlence standındaki deliklerle dolu bir duvara fırlatılan topları canlandırın. Eğer top deliklerden birine isabet ederse, ödül seçimine hak kazanırsınız. Kendini düzenleyen haritanın eğitilmesi de, önceki oyunda gözü bağlı olmaya ve başlangıçta duvarda tek bir deliğin bulunmamasına benzer. Aynı benzetme, büyük

veri tabanlarındaki örüntüleri araştırmaya başlanıldığında, nereden başlanılacağını bilinmediği durumlar için de geçerlidir. Her top atışında duvarda ufak bir çukur oluşmaktadır. Aynı civara isabet eden toplar çoğaldıkça bir delik açılır. Bu konuma yeni bir top isabet ettiğinde deliğin arasından geçer. Bu durumda tanımlanabilir bir kümenin varolduğu sonucuna varılır.

Şekil 8.12’ de, bunun basit bir kendini düzenleyen harita ile nasıl çalıştığı gösterilmektedir. Öğrenim kümesinin bir ögesi ağa sunulduğunda, değerler çıktı katmanındaki birimlere doğru ilerlemektedir. Çıktı katmanındaki birimler birbirleriyle rekabet etmekte ve en yüksek değerli olanı “galip” gelmektedir. Mükafatı, girdi örüntülerinin, kazanan birim tarafından cevaplanmasını kuvvetlendirecek şekilde ağırlıkların ayarlanmasıdır. Bu olay ağ içerisinde ufak bir girintinin oluşturulmasına benzerdir.

Ağ eğitiminin bir yönü daha vardır. Sadece kazanan birimin ağırlıkları ayarlanmamakta, ayrıca en yakınındaki birimlerin ağırlıkları da ayarlanarak girdileri cevaplamaları güçlendirilmektedir. Ayarlamalar, civar büyüklüğünü ve ayar miktarını kontrol eden, komşuluk parametresi tarafından denetlenmektedir. Başlangıçta civar büyüklüğü ve ayarlamalar oldukça fazladır. Öğrenme süreci ilerledikçe civar büyüklüğü ve ayarlamalar azalmaktadır. Komşuluk parametresinin aslında birçok etkileri bulunmaktadır. Bunlardan bir tanesi, birimlerin birbirlerine doğrudan bağlı olmamalarına rağmen, çıktı katmanı birleşmiş bir yapıya dönüşmektedir. Benzer kümeler, benzer olmayanlara nazaran, birbirlerine daha yakın olmalıdırlar. Daha da önemlisi, komşuluk parametresi, bir birim grubunun tek bir kümeyi temsil etmelerine imkan sağlamaktadır. Komşuluk parametresi olmadan, ağ çıktı katmanındaki birim sayısı kadar küme bulma eğiliminde olacaktır.



Şekil 8.12 SOMs ile Girdiyi En İyi Tanıyabilecek Çıktının Belirlenmesi

(Kaynak : Berry, M. & Linoff,G.)

Genellikle, kendini düzenleyen haritalar sahip oldukları çıktı birimlerinden daha az sayıda kümeyi tanımlar. Yeni kayıtların kümelerle atanması sırasında söz konusu girdilerin kullanılmayan çıktı birimlerine ilerlediği takdirde, bu durum verimsizliğe neden olur. Aktif birimlerin tespiti için kendini düzenleyen harita, test kümesi üzerinde test edilmelidir. Her bir test kümesi ögesi için, ağırlık katmanındaki kazanan birim izlenir. İsabet oranı çok düşük çıktı birimleri atılır. Bu birimlerin elenmesi, yeni durumlar için gerekli hesaplamaları azaltarak, ağırlık çalışma performansını artırır.

Sadece belirli kümeleri tanımlayan çıktı birimlerinden oluşacak şekilde sonuçlandırılan ağ, yeni durumlara uygulanabilir. Bilinmeyen bir örnek, en büyük ağırlıklı çıktı birimindeki kümeye atanır.

8.14.2. Kendini Düzenleyen Haritalar ile Kümelerin Bulunması Örneği

Büyük bir banka verdiği ev kredilerinin sayısını arttırmayı amaçlamaktadır. Pazar payını arttıracak en doğru stratejinin belirlenmesi için mevcut ev kredisi kullanmış müşterilerin tanınması gerektiği kararlaştırılıyor. Prosesi başlatmak amacıyla ev kredisi kullanmış 5,000 ve kullanmamış 5,000 olmak üzere toplam 10,000 müşteriye ait demografik bilgiler toplandı.

Toplanan veriler aşağıdaki alanları içermektedir :

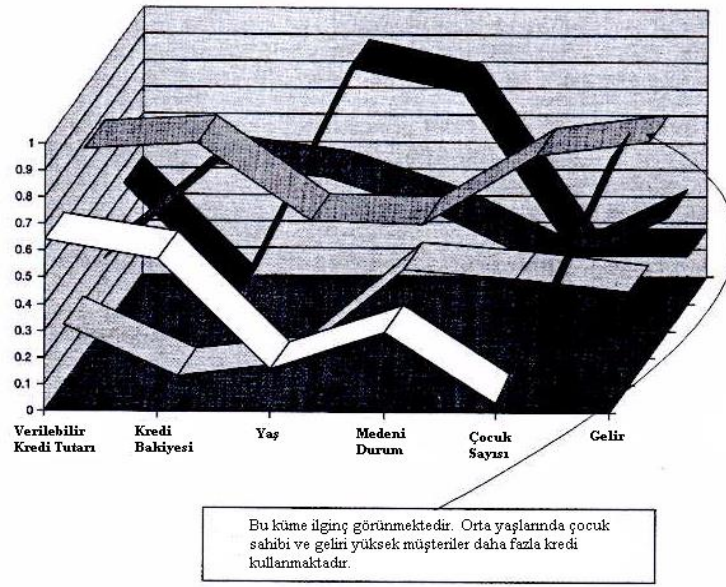
- Evin değeri
- Gerekli kredi miktarı
- Verilebilecek kredi miktarı
- Yaş
- Medeni durum
- Çocuk sayısı
- Toplam gelir

Bu veriler kümelemede kullanılacak iyi bir öğrenim kümesi oluşturmaktadır. Girdi değerleri 0 ile 1 arasında değişen değerlere indirgendi ve kendini düzenleyen harita ağının eğitiminde kullanıldı. Ağ tarafından veri içerisinde beş küme belirlendi, ancak bu kümeler hakkında bir bilgi bulunmamaktadır.

Farklı kümelerin kıyaslanabileceği bir yönteme ihtiyaç vardır. Sinir ağları ile oldukça iyi sonuçlar veren ve yaygın olarak kullanılan bir yöntem, *ortalama öge* yöntemidir. Amaç, her bir kümenin merkezini bulmaktır. Duyarlılık analizi yaklaşımına benzerdir. Küme içerisindeki her bir özelliğin ortalama değeri bulunur.

Örneğin, bir kümeye ait elemanların yarısı bay, yarısı bayan olsun. Baylar 0.0'a, bayanlar 1.0'a tekabül etmektedir. Bu kümeye ait ortalama ögenin, söz konusu özellik için değeri 0.5 tir. Bir başka küme her bir bay' a karşılık dokuz bayandan oluşur ise bu kümeye ait ortalama ögenin, söz konusu özellik için değeri 0.9 dur.

Kümeyi diğerlerinden farklılaştıran özelliğin tespiti için, oluşan tüm kümelere ait her bir özelliğin ortalama değerleri, aynı grafik üzerinde bir araya getirilir. Banka örneğinde, kırk yaşlarında evli ve çocuk sahibi müşterilerden oluşan bir küme oldukça ilgiye değer bulundu. Biraz daha araştırıldığında bu müşterilerin son on yıl içerisinde çocuk sahibi oldukları belirlendi.



Şekil 8.13 SOM ile Tanımlanan Kümelerin Kıyaslanması

(Kaynak : Berry, M. & Linoff,G.)

Bankanın pazarlama departmanı bu kümedeki kişilerin, çocuklarının kolej taksitlerini ödemek amacıyla kredi kullandıkları kararına vardı. Departman bu amaca özel bir pazarlama programı tasarladı. Ancak kampanya sonuçları hayal kırıklığı yarattı.

Pazarlama programının başarısız olması nedeniyle, kendini düzenleyen harita ağının, vaat edildiği kadar başarılı kümeler üretmediği sonucu çıkarıldı. Aslında sorun başkaydı. Banka başlangıçta müşterilerin sadece genel bilgilerini kullanmıştı. Müşterilerine sunduğu diğer hizmet bilgilerini birleştirmemişti. Banka, bu kez mevduat, kredi kartı gibi diğer hizmet bilgilerini dahil ederek müşterilerini belirleme problemini yeniden ele almıştır.

Temel yöntem değişmedi. Eklenen yeni verilerle, kolej çağında çocukları olan müşterilerin varolduğunu, ancak bununla birlikte gözden kaçan bir hususun varlığı tespit edildi. Bu kümedeki müşterilerin tasarruf mevduatı kadar, ticari mevduata da sahip oldukları belirlendi. Bu doğrultuda, çocukları koleje başlayan ailelerin yeni bir işe başlama fırsatı yakaladıkları ve krediyi bu yönde kullandıkları sonucuna varıldı.

Banka bu kümedeki müşterilere özel, yeni bir pazarlama programı hazırladı ve beklenen performans elde edildi. Bu örnekten çıkarılması gereken sonuç, sınır ağlarının, küme tespitinde güçlü olmalarına rağmen, başarıları tümüyle ele alınan verilerin niteliğine bağlı olduğudur.

8.15. Yapay Sinir Ağlarının Kuvvetli Yönleri (Strengths)

- *Geniş bir yelpazedeki problemlerde kullanılabilir.*

(Ağın ürettiği çıktı, bir evin değeri gibi sürekli bir değer ise öngörü, ayrık bir değerse sınıflandırma işlevi yerine getirilmiş olur. Nöronların basit bir düzenlenmesi ile, ağ küme tespitine uygun hale getirilebilir.)

- *Karmaşık alanlarda dahi iyi sonuçlar üretir.*

(Sinir ağlarının başarıları birçok sektördeki çok sayıdaki uygulamada defalarca kanıtlanmıştır. Bu sonuçlar, diğer yöntemlerin çözemeyeceği, zaman serilerinin analizi ve sahtekarlık tespiti gibi oldukça karmaşık alanlardaki başarılarından ileri gelmektedir. Uygulamadaki en büyük sinir ağı AT&T tarafından çek numaralarının okunmasında kullanılan ve toplam yedi katmanda dağılmış yüz binlerce birimden oluşan ağdır.)

- *Hem sürekli hem kategorik değişkenleri işleyebilir.*

(Verilerin indirgenmiş olmaları ile birlikte, gerek girdiler gerek çıktılar için bu değerler kategorik ve sürekli verilerden oluşmaktadır.)

8.16. Yapay Sinir Ağlarının Zayıf Yönleri (Weaknesses)

- *Girdi değerleri 0 ile 1 arasında olmalıdır.*

(Sinir ağının girdileri genellikle 0 ile 1 arasındaki bir değere indirgenmelidir. Bu ek bir dönüşüm ve çalışma ile birlikte zaman, CPU gücü ve disk alanı gerektirmektedir. Ayrıca seçilen dönüşüm yöntemi ağın sonucunu etkileyebilir.)

- *Sonuçları hakkında bir yorumda bulunamıyor.*

(Sonuçların yorumlanamaması sinir ağlarına yöneltile en başlıca eleştiridir. Kuralların yorumlanması kritik bir faktör olan alanlarda tercih edilen bir yöntem değildir. Ancak sinir ağları anlaşılır kurallar üretemese de, duyarlılık analizi sayesinde sonuca diğerlerinden daha etkili olan girdileri belirleyebilmektedir.)

- *Fayda getirmeyecek bir sonuca yönlendirme ihtimali vardır.*

(Sinir ağıları genellikle verilen herhangi bir öğrenim kümesi için sonuçlar üretir. Ancak üretilen sonucun verinin doğru bir modelini yansıttığı garantisi yoktur. Modelin performansı test kümesi verileri üzerinde test edilmelidir.)

8.17. Sinir Ağları'nın Uygulanabildiği Durumlar

Sinir ağıları, modelin nasıl çalıştığından öte, ürettiği sonuçların önem arz ettiği durumlarda, sınıflandırma ve öngörü işlevlerinde özellikle, doğru bir tercihtir. Saydam olmaması nedeniyle kurallarını çıkartmak zordur.

Sinir ağıları, ayrıca kümeleme gibi dolaylı veri madenciliği işlevlerinde kullanılabilir. Birbirine benzer kayıtların kümelerini belirler, ancak benzerlikleri konusunda bilgi vermez. Bu bilgi, kümelerin merkezlerini grafiksel olarak kıyaslayabilen başka yöntemlerin aracılığı ile sağlanır.

Çok fazla sayıdaki girdi özelliğinin bulunması, örüntülerin keşfedilmesini güçleştirir, öğrenim fazını uzatır ve büyük ihtimalle doğru sonuca ulaştırmayabilir. En önemli değişkenlerin belirlenmesi amacıyla karar ağaçları yöntemi kullanılabilir.

9. MÜŞTERİ İLİŞKİLERİ YÖNETİMİ (CRM)

2000’li yıllarda, CRM yazılımlarına dünyada 10 trilyon \$ ve yaklaşık iki katı kadar da CRM hizmetlerine harcanmıştır. Araştırmacıların bu konudaki 2003 tahminleri, miktarın iki katından fazla olacağı yönündendir.

Yatırımların çoğu ise beklenen geri dönüşü sağlayamayacaktır. Planlama eksikliği, tanımlanmamış iş hedefleri ve ölçüm eksikliği başarısızlığın başlıca nedenleridir. İşletmelerin, CRM uygulamasından ne bekledikleri, daha da ötesi CRM’ in ne olduğunu bilmemelerinden kaynaklanmaktadır.

CRM bir teknoloji veya teknolojiler grubu değildir, içe dönük geleneksel iş modelinin değişimini gerektiren, devamlı gelişen bir süreçtir.

CRM, insan kaynağı ve teknoloji yatırımları ile iş süreçleri tarafından desteklenen, müşteriye odaklı bir yaklaşımdır.

CRM kavramı tüm işletmelerin başarısı için gerekli tüm bileşenleri içerir :

9.1. Müşteri (Customer)

Tüm işletmeler en başta müşteri ihtiyaçlarına odaklanmalıdır. İşletmeler, büyüdükçe ve işler daha karmaşık hale geldikçe, müşterilerinden çok iç problemlerin çözümüne yönelirler. CRM’in hedeflerinden biri, işletme açısından da maliyeti makul bir seviyede tutacak şekilde, müşterinin yeniden önem kazanmasını sağlamaktır.

9.2. İlişki (Relationship)

Yakın bir geçmişe kadar büyük işletmelerin milyonlarca müşterisi ile ilişki kurmaları imkansızdı. Teknoloji sayesinde, işletmelerin müşterileri ile devamlı bir ilişki kurmaları, onların sadakatini kazanmaları ve işletmeye kattıkları değeri arttırmaları olanaklı hale gelmiştir.

9.3. Yönetim (Management)

Sadakat, kar ile eşdeğerdir ve bundan hem müşteri hem de işletme kazançlı çıkmaktadır. CRM in yönetim bileşeni, doğru enformasyonu doğru zamanda sağlayarak, müşteriye memnun edecek doğru fiyatlandırmayı sunarak ve müşterinin ihtiyaç duyabileceği başka bir ürün veya hizmeti ve nedenlerini önceden tahmin ederek, müşteri ile olan ilişkiyi kontrol eder.

CRM ile amaçlanan, işletmenin müşterilerini tanıyarak, onlara iyi hizmet ederek ve ihtiyaçlarını karşılayarak karını arttırmasıdır.

CRM kontrol listesi :

- İhtiyaç duyan herkese, organizasyon çapında müşteri ile ilgili tek bir bakış açısının sağlanması
- Organizasyondaki her çalışanın, müşterinin ihtiyaçlarını aynı şekilde cevaplayabilmesi
- Müşterilerin her kanaldan aynı standartta hizmet alabilmesi
- Pazarlama maliyetinden de tasarruf ederek, müşterilere ihtiyaç duyabilecekleri ürün ve hizmetler hakkında önceden bilgi verilebilmesi
- En karlı müşterilerin kim olduklarının bilinmesi
- Bu müşterileri elde tutacak strateji ve taktiklerin bulunması

9.4. İrtibat Merkezleri (Contact Center)

İrtibat Merkezi, müşteri ilişkileri yönetiminin geleceğine yön vermektedir. Başlıca hizmet kanalı telefon veya İnternet olan işletmelerde, irtibat merkezleri işletimsel gelişmelerin ve CRM stratejilerin odağında yer almalıdır.

Aşağıdaki tabloda irtibat merkezlerinin, maliyet merkezi modelinden, kar merkezi ve CRM stratejisinin anahtar bileşenine dönüşüm aşamaları gösterilmektedir.

Tablo 9.1 İrtibat Merkezinin Evrimi

	1.Aşama	2.Aşama	3. Aşama	4.Aşama (Vizyon)
İşletmenin İrtibat Merkezine bakış açısı	Zorunlu maliyet merkezi	Potansiyel kar merkezi	Müşteri ilişkileri açısından stratejik	İrtibat merkezi işletmenin kendisi
İrtibat Merkezinin Odağı	Maliyetleri düşürür Verimliliği artırır.	Maliyet merkezinden kar merkezine döndürür.	Müşterilerle karlılığı arttıracak devamlı ilişkiler kurar.	İrtibat merkezinin ötesindeki işlemler kurumun tamamına taşınır.
İş ihtiyaçları/amaç	Verimlilik	Gelir kuşağı	Müşteri sadakati	İş iletişiminin optimum hale getirilmesi

(Kaynak : Contact Babel)

9.5. Başlıca İş Akımları

CRM aslında, 1980’li yıllarda küçülme ile başlayan, başlıca ticari ve IT girişimleri serisinin mantıksal bir adımıdır. Önceki çoğu girişimde, çalışanlar, çalışma yöntemleri veya teknoloji gibi iç bileşenlere ağırlık verilmiş, maliyet azaltmaya odaklanılmıştır.

Kazanç artışı, maliyet tasarrufu ile sağlanıyordu.

- **Küçülme (Downsizing)** – 1980’li ve 1990 yılların başında en basit maliyet azaltma yöntemi
- **İş Süreçlerinin Yenilenmesi (Business Process Re-engineering)** - 1980’lerin sonlarında, 1990’ların başında kişi sayısının azaltılması sonrasında, verimliliğin artırımı, iş akış biçimlerinin değişimini ve otomasyonun yaygınlaşması ile sağlandı.
- **Kurumsal Kaynak Planlama (Enterprise Resource Planning)**

Finanstan, satış ve dağıtımdan, personele kadar her şeyin entegrasyonu sağlanarak, işlemlerin gerçek zamanlı izlenebilmesi ve kontrol edilmesi hedeflenmekteydi.

Tüm bu girişimler, maliyelerin azaltılması ve verimliliğin artırılması temeline dayanmaktaydı.

Tablo 9.2 Başlıca İş Akımlarının Evrimi

Odaklanan alan	İNSAN	Küçülme	Yetkilendirme
	İŞ	İş süreçlerin yenilenmesi	İş Odaklı CRM
	TEKNOLOJİ	Kurumsal Kaynak Planlama (ERP)	Teknoloji Odaklı CRM
	Düşük Maliyet		Yüksek Gelir
	Beklenen sonuç		

(Kaynak : Contact Babel)

Ancak bir zaman sonra, işletmenin fonksiyonlarını engellemeden maliyetleri azaltmak zorlaşır. Bu aşamada CRM devreye girmektedir.

9.6. Başarılı CRM Kavramı

Başarılı bir CRM, teknolojiye bağlı olduğu kadar, tutuma da bağlıdır. CRM, işletmelerin büyümesini sağlar ve iş akımların evrimi bu paradigmayı destekler.

- Teknoloji Odaklı CRM : 1990'ların sonundan günümüze, ilk CRM uygulamalarında çoğunlukla teknoloji çözümlerine odaklanıldı. İş süreçleri ve çalışanlar, IT liderliğindeki CRM den etkilenmedi. Bu ise, gerçek bir CRM uygulamasından öte, nokta çözümler serisi olabilir ancak.
- İş Odaklı CRM : Mevcut ve aday müşterilerin kurum ile olan ilişkilerini büyüteç altına alan, iş liderliğindeki CRM, ticari problemlerin araştırılması ve çözümüne, müşterilere daha etkili ve karlı hizmet verebilecek, kurum kültürünün değişimine odaklanmaktadır. İşe odaklı CRM, müşteri değerinin kurum tarafından anlaşılmasına ve karlılığı sağlayacak bağımlılıklarını kazandıracak etkileşimlerin özelleştirilmesine imkan tanır.

İki yaklaşım arasındaki farklılıkların çoğu izlenilen tutumla ilgilidir.

Tablo 9.3 CRM tutum ve yaklaşım farklılıkları

	Bugün IT odaklı CRM yaklaşımı	Öngörülen İş odaklı CRM yaklaşımı
İrtibat Merkezine yüksek sayıda çağrı gelmektedir.	E-mail ve web tabanlı self-servisleri sunarak bu müşterileri Internet kanalına yönlendirmeye çalışır.	Ne tür çağrılar geldiği ve kimden geldiğini anlamaya çalışır : Aynı müşteri defalarca arıyor mu? En iyi müşteriler telefonda çok bekletiliyor mu? Her zaman aynı sorular mı soruluyor? Müşteriler Internet'i kullanabiliyor mu?
Satışların artırılması gerekmektedir.	Çapraz satış amaçlı, bir veri ambarı kurup, müşteri tipleri ile ürünleri eşleştirir	Bugüne kadar yapılan satışların farkına varmaya çalışır : İrtibat noktalarındaki temsilciler, müşterilerle yapılan telefon görüşmelerin kısa tutulması baskısı altında mıdır? Müşteri başına düşen satış miktarı doyum noktasına ulaştı mı? Ürün yelpazesi mevcut müşterilere hitap ediyor mu?

(Kaynak : Contact Babel)

Her iki yaklaşım aynı çözümle sonuçlanabilir, ancak problemin anlaşılması önemlidir.

Yukarıdaki ilk örnekte, irtibat merkezini arayan ve ürün satışında ikna olabilen müşterileri, Internet kullanımına yatkın olmadıkları halde bu kanala yönlendirmek, verimsizliğe neden olur. Bu durumda, basit olarak satış temsilcilerin sayısının artırılması verimlilik açısından daha doğru bir karar olur. Teknolojinin faydalarından uzak bir çözüm olabilir, ancak işe odaklı bir CRM yaklaşımı, mevcut kurumun yapısına yeni teknolojileri katmayı amaçlamıyor.

İşin ve bireylerinin doğru tutumundan oluşan sağlıklı bir karışım, iş yeteneği ve teknik bilgi, başarılı bir CRM için kritik bileşenlerdir.

Başarılı CRM = İşin yeniden yapılanması + teknolojik yeterlilik + yetkili personel

CRM öncesi, personel ile ilgili çözülmesi gereken başlıca sorunlardan bazıları şunlardır :

- Yetersiz eğitim ve ulaşılamayan veriler nedeniyle müşteri taleplerin tümünün aynı temsilci tarafından karşılanamaması
- Eksik müşteri ve ürün bilgisi
- Yönetimin, müşteri memnuniyetinden (gelirin artışı) öte konuşma süresinin kısaltılmasına (maliyetin azalması) odaklanması
- Tek düze, tekrarlanan işler ve kısıtlı kariyer imkanları nedeniyle personelin motivasyon eksikliği

Müşterilere her bir kanaldan üstün hizmet sağlamak, müşteri bazında karı arttırmak, şirketin pazar payını büyütmek gibi CRM' in temel hedefleri, tecrübeli ve yetkili takımlar olmaksızın gerçekleştirilemez.

9.7. Başarılı CRM İçin İş Analizi

Aşağıdaki tablo bir CRM uygulamasının anahtar aşamalarını göstermektedir. Tam detayı işin tipine, maliyetine, zamana ve becerilere bağlı olarak belirlenir. Teknik projelerde kullanılan Analiz-Tasarım-Uygulama yaklaşımı, CRM vasıtası ile iyileştirilen iş proseslerinde de uygulanabilir.

Tablo 9.4 CRM Uygulamasının Temel Aşamaları

Başka projeleri etkileyebilecek yenilikler konusundaki izlenimler	İş Analizi	CRM uygulamasından çıkarılan dersler ve etkileri konusundaki izlenimler
	Tasarım	
	Uygulama	
	İzleme ve ayarlama	

(Kaynak : Contact Babel)

İlk bakışta CRM uygulamaları, birçok projede olduğu gibi, ihtiyaçların analizini takiben detaylı tasarım şablonu ile benzerlik göstermektedir. Detaylar olduğu kadar, izleme aşamasının kullanımı da CRM' i diğer projelerden farklı kılmaktadır. Birçok projede izleme, her şey tamamlandıktan sonra başlıyor, ancak CRM de proje sürecinin temel bir aşamasıdır.

CRM analizinde dikkat edilmesi gereken hususlardan bazıları aşağıdaki şekildedir :

Yönetim grubunu bir araya getirmek

- Bu grup IT, satış, pazarlama, müşteri hizmetleri ve iş analizi yöneticilerini içermelidir.
- Grupta, çalışmaların tümü ile ilgilenecek bir proje sahibinin bulunması gerekir.

Mevcut işi gözden geçirmek

- Projenin başarısını ölçmek için performans metriklerin toplanması
- Müşteri geri dönüşlerin toplanması
- İş birimlerin uygulamaları kıyaslanarak, tespit edilen en iyilerin devam ettirilmesi ve paylaşılması

Müşteri deneyimlerini anlamak

- Şirketin gerçek müşterisiymiş izlenimini veren testler yapılmalı
- Ürünlerin, faaliyet bölgenin müşterisi açısından uygunluğunun anlaşılması
- Kanallar arası servis kalitesindeki farklılıkların ve her bir kanalı tercih eden müşteri tiplerinin anlaşılması

İç girdileri toplamak

- IT (mevcut sistemin ayarlanabilirliği, entegrasyonu, altyapı sorunları, veri transferi, vb.)
- İrtibat merkezi temsilcileri (tamamlanması gereken iş ve teknoloji açıkları)
- Satış ve pazarlama (müşteri kaybetme oranı, müşteri hakkındaki mevcut bilgiler, satın alma davranış örnekleri, vb.)
- Strateji ve üst düzey yönetimi (şirketin yönetimi, büyüme hedefleri, ürün çeşitliliği, vb.)

Ölçülebilir hedefleri kararlaştırmak

- Ne kadar ve ne zaman?
 - Sadakat

- Gelir
- Maliyet azalışı
- Kar
- Müşteri başına ürün satışı
- Müşteri memnuniyeti
- Personel değişim sıklığı
- Verimlilik
- Vb...

İş analizi aşaması tümüyle tamamlanmadan tasarım aşamasına geçilmemelidir. Analiz aşamasında yeterli beceriye ve etki gücüne sahip, doğru kişilerin bulunması, daha sonra çıkabilecek olumsuzlukların önüne geçer.

Bu ise, iş analizi sonucunun daha sonra değiştirilemeyeceği anlamına gelmez. Tasarım ve uygulama aşamalarında edinilen tecrübeler ile şirketin yönetim ve operasyonunda ihtiyaç duyulan değişiklikler tespit edilir. Ancak CRM projesinin başında, analizin mümkün olduğunca kapsamlı yapılması amaçlanmalıdır.

Kritik Başarı Faktörleri :

- CRM girişimini yönetecek, hem ticari hem kültürel açıdan yaklaşabilen en azından bir proje lideri seçilmelidir.
- CRM den etkilenen tüm iş birimlerinin yöneticilerinden oluşan bir komitenin oluşturulması
- Yeni bir teknolojiyi uygulamadan önce şirketin tüm operasyonları kıyaslanmalı
- İş yapma modelinin değişimi konusunda müşteri beklentilerine kulak verilmeli
- İşin en çok geliştirilmesi istenen yönleri ile ilgili ölçülebilir göstergeler belirlenmeli
- Desteği kanıtlanmış ve öngörülen gelecekte de güven vaat eden tedarikçilerle çalışılmalı

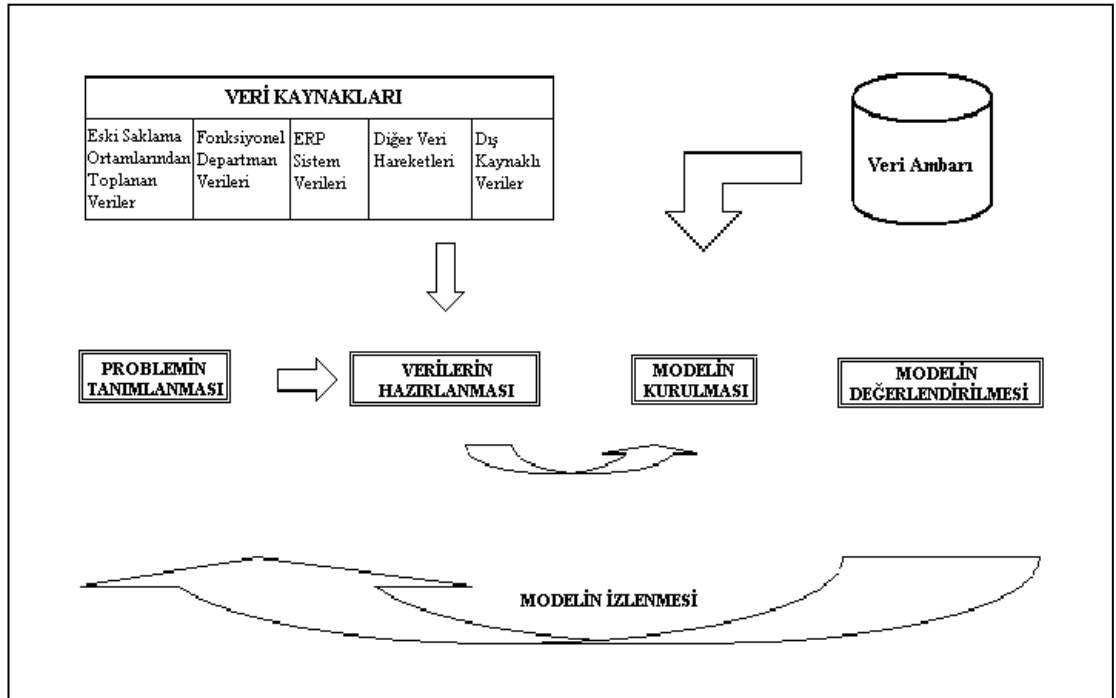
- Her bir alt projenin etkileri ölçülmeli ve sonuçları CRM projesinin bütünü ile ilgili analiz ve bağılı tasarım aşamalarına yansıtılmalı
- Müşteri odaklı organizasyon kültürünü oluşturmak için, sürecin her aşamasında çalışanlara danışılmalı, onları bilgilendirmeli ve gerekli eğitimler sağlanmalı.

10. BİREYSEL KREDİ TALEPLERİNİN DEĞERLENDİRİLMESİNDE BİR UYGULAMA ÖNERİSİ

10.1. Giriş

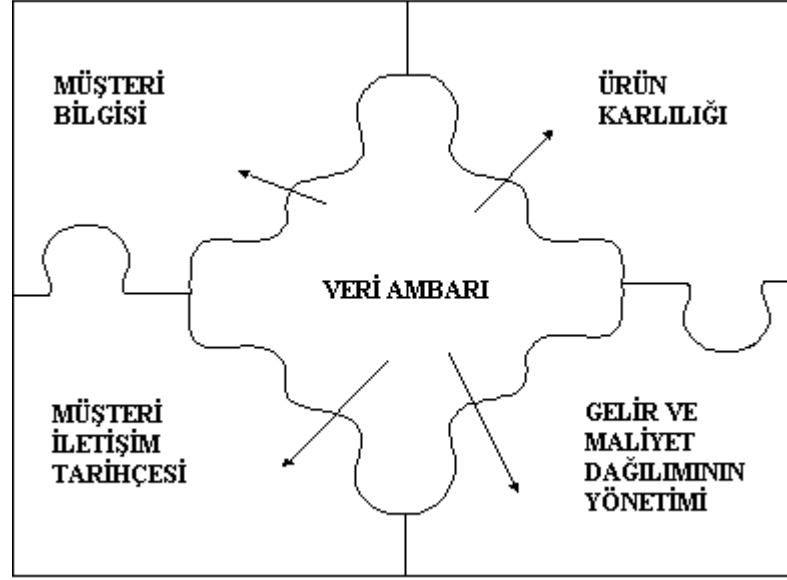
Çeşitli karar verme aşamalarında, veri madenciliği yöntemlerinin kullanımı kuramsal açıdan mümkündür. Veri madenciliğinde amaç, çok büyük miktardaki ham veriden değerli bilginin çıkarılmasıdır. Çok miktarda güvenilir (geçerli, hatasız ve eksiksiz) veri ön şarttır, çünkü elde edilen sonucun kalitesi öncelikle verinin kalitesine bağlıdır.

Şekil 10.1 de görüldüğü gibi çeşitli veri kaynaklarından verilerin toplanması ile başlayan bilgi keşfi süreci, toplanan verilerin analiz için uygun hale getirilmesi aşaması ile devam etmektedir. Ancak veri ambarına (*Data Warehouse*) sahip olan kuruluşlarda, gerekli verilerin *Data Mart* olarak isimlendirilen işleve özel veri tabanlarına aktarılması ile doğrudan veri madenciliği işlemlerine başlanabilmesi de mümkündür.



Şekil 10.1 Bilgi Keşfi Süreci

En doğru kar modellerinin elde edilebilirliği, kurumun bütün resmini yansıtacak ve aşağıdaki şekilde gösterildiği gibi temel bileşenlerini içerecek bir veri ambarının oluşturulması ile mümkündür.



Şekil 10.2 Veri Ambarının Temel Bileşenleri

(Kaynak : Geiger. Jonathan G)

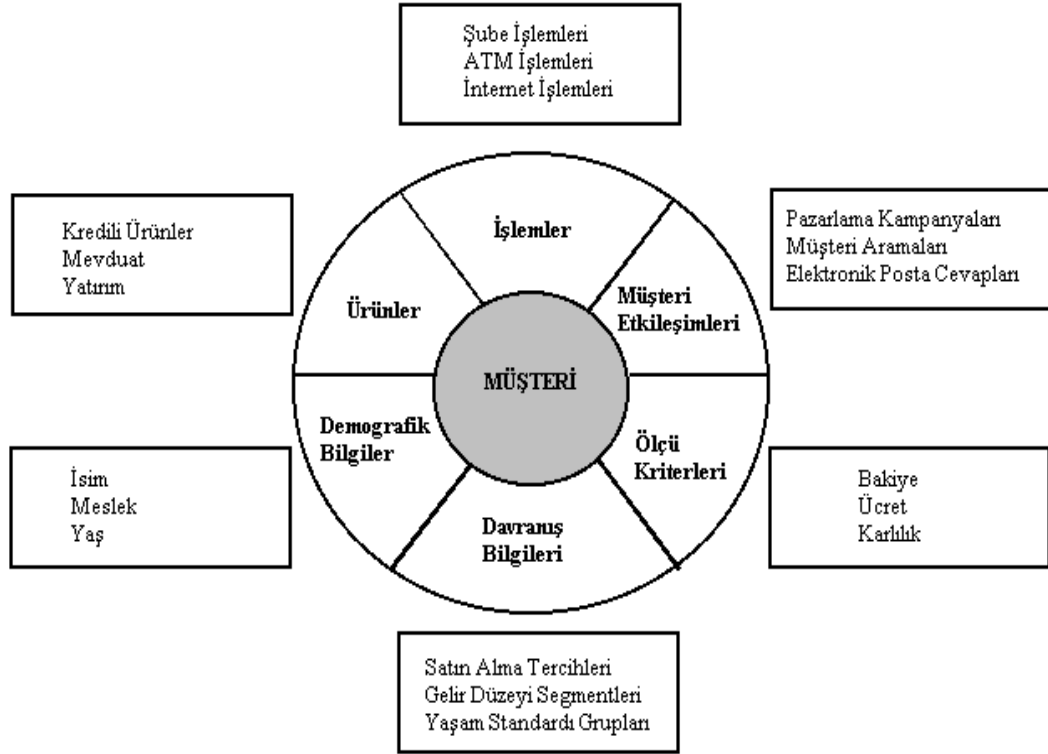
Şirketin rekabet gücü, stratejisinin Müşteri İlişkileri Yönetimi bazlı bir pazarlamayı içermesine bağlıdır. Bu genel strateji ise alışlageldiğimiz *ürün merkezli* bir strateji değil, tamamen *müşteri merkezli* bir strateji olmak durumunda.

Strateji, ya da iş stratejisi ile vurgulanmak istenen şirketi rakiplerden farklılaştıracak yöntemler bütünüdür. Bugüne dek hakim olan düşünce, üretilen ürünün performansını iyileştirmek üzerine odaklanıyordu. Ancak teknolojinin kolay elde edilebilirliği, herhangi bir yeni ürünün hızlı şekilde taklit edilebilirliğini olanaklı hale getirmiştir. Buna bağlı olarak, üretilen ürün diğer sunumlara benzer ise, fiyatlar, temel maliyetleri karşılama noktasına kadar geriler ve karlılık sıfıra yaklaşır.

Farklılaştırma çabalarına ürünü iyileştirmekten başlamak yerine, *müşteri isteklerinden* başlamak, stratejik açıdan çok farklı açılımlar getirebilir.

Bu doğrultuda, müşteri ile ilgili tüm demografik bilgileri, ürün ve hizmet tercihleri, yapmış olduğu işlemleri, kurumla olan ilişkileri, yaşam standardı detay bilgilerini

içeren, aşağıdaki şekilde gösterildiği gibi müşteri merkezli bir veri ambarı tasarımı yapılmalıdır.



Şekil 10.3 Müşteri Merkezli Veri Ambarı Tasarımı

(Kaynak : Geiger. Jonathan G)

10.2. Problemin Tanımlanması

Veri madenciliği çalışmalarında başarılı olmanın ilk şartı, uygulamanın hangi işletme amacı için yapılacağını açık bir şekilde tanımlanmasıdır. İlgili işletme amacı işletme problemi üzerine odaklanmış ve açık bir dille ifade edilmiş olmalı, elde edilecek sonuçların başarı düzeylerinin nasıl ölçüleceği tanımlanmalıdır. Ayrıca yanlış tahminlerde katlanılacak olan maliyetlere ve doğru tahminlerde kazanılacak faydalara ilişkin tahminlere de bu aşamada yer verilmelidir.

10.2.1. Proje Konusu : Müşteri Risk Değerleme Sistemi

Projenin konusu; bireysel müşterilerin risk doğuran ürün ve hizmetlerin kullanımına ilişkin taleplerini, müşterilerin riske etkiyen müşteri bilgilerini dikkate alarak

değerleyen, parametrik bir sistem kurmak, müşterilere risk açısından doğru ürün ve/veya hizmeti en kısa sürede sunabilmektir.

10.2.2. Projenin Amacı

- müşterinin riskini doğru bir şekilde belirlemek,
- riski minimize ederek, karı maksimize etmek,
- müşterinin riskini azaltmak, riskli ürünlere ilişkin sahteciliği önlemek
- riskli ürünlerdeki kaynak kayıplarını bertaraf etmek,
- müşteriye taşıdığı risk faktörlerine göre en uygun ürün ve/veya hizmeti öneren bir yapı oluşturmak,
- otomasyonlu bir müşteri değerlendirme sistemi kurarak müşteri değerlendirme için ayrılan iş gücü kaynağından tasarruf sağlamak,
- müşteri değerlendirme sürecini hızlandırmak,
- müşteri başvurularının yanıtlanması sürecini çabuklaştırmak,
- güvenilir ve tutarlı bir müşteri değerlendirme sistemi kurarak manuel değerlemenin göreceli ve tartışmalı neticelerini minimize etmek ve böylelikle risk değerlendirme politikasını kurumsallaştırmak,
- müşteri değerlendirme hususunda görev yapan çalışma gruplarına karar vermede yardımcı olmak,
- müşteri değerlemeyi olabildiğince objektif kıstaslara dayandırarak inisiyatif kullanımını azaltmak.

10.2.3. Projenin Kapsamı ve İçeriği

Müşteri Risk Değerleme' ye Konu Olabilecek Ürünler :

- Kredi Kartı
 - İlk Başvuru
 - Limit Artışı Talebi
- Bireysel Kredi
 - Başvuru

Müşteri Risk Değerleme Sistemi'nin Hedef Kitlesi :

- Mevcut Bireysel Müşteri
 - Kredi Kartı Sahibi
 - Kredi Kullanan Bireysel Müşteri
- Aday Bireysel Müşteri
 - Kredi Kartı / Kredi Talebi Olan Bireysel Müşteri

10.2.4. Kredi Değerlendirme Süreci

Kredili bir ürün talebinde bulunan bir müşteriye verilecek kredi miktarı, alınacak teminatlar ve uygulanacak faiz oranları, ilgili işletmenin kredi analistleri ve/veya kredi değerlendirme grupları tarafından, müşterinin nicel ve nitel verileri incelenerek verilmektedir. (Bkz. Şekil 10.5)

Çalışmanın bu bölümünde kredi değerlendirme sürecini otomatikleştirmek için veri madenciliği yöntemleri ile müşterinin risk seviyesini belirleyebilecek bir model önerilmektedir. Müşteri bilgilerinden oluşacak bir kredi skor kartı, müşteriye verilebilecek maksimum kredi limiti belirlenecektir.

10.2.5. Kredi Skor Kartı

Kredi, bir finansal kurum tarafından bir müşteriye ödünç verilen ve faiz eklendikten sonra genelde düzenli aralıklı taksitler halinde geri ödenmesi gereken paradır. Bir kredi başvurusunda müşterinin krediyi geriye ödeyememesi olasılığını (*propensity to default*) hesaplamaya kredi skora (*credit scoring*) denir. Buna, davranış veya performans skorlamadan ayırmak için başvuru skora da denir. Başvuru skorlamada (*application scoring*) kredi başvurusu sırasında müşterinin kredi talebi için not verilir. Davranış veya performans skora (*behavior/performance scoring*) ise kredi almış ve taksitlerini ödemekte olan bir müşteriye değerlendirmekte kullanılır; örneğin kredi kartının limitini değiştirmek, yenilemede yeni kartın süresi, olası sorunların tahmini, geç ödeme durumunda alınacak eylem bu şekilde belirlenebilir.

Skorlama yaparak yüksek riskli müşterilere kredi vermeyi reddetmek finansal kurumun olası zararını azaltacak, düşük riskli müşterilere kredi vererek kârını

arttıracak, üstelik müşterilerin ödeyemeyecekleri kredilerden dolayı rahatsızlığını azaltacaktır.

Skor kartı (*score-card*) olarak isimlendirilen model, müşterinin başvuru formundaki bilgilere ve diğer (örneğin kredi izleme bürosundan gelen) bilgilere dayanarak kredinin geriye ödenememesi olasılığını hesaplar. Bu değer uygun bir eşik değeri ile karşılaştırılarak kredi talebi kabul veya red edilir.

Skor kartı, geçmiş müşterilerin verilerinden oluşturulur ve genelde basit bir ağırlıklı toplamadır. Tipik olarak kullanılan alanlar şunlardır:

- İkamet adresinde oturduğu zaman: 0-1, 1-2, 3-4, 5+ yıl
- Ev durumu: Sahip, kiracı, diğer
- Posta kodu: Kodlanmış
- Telefon: Evet, hayır
- Yıllık gelir: Kodlanmış
- Kredi kartı: Evet, hayır
- Yaş: 18-25, 26-40, 41-55, 55+ yıl
- Meslek: Memur, işçi, serbest, işsiz, ...
- Medeni hali: Evli, bekar, diğer
- Bankanın müşterisi olduğu zaman: yıl
- Çalıştığı kurumda çalışma zamanı: yıl

Müşteri ile ilgili her bilginin, iyi müşteriye kötüden ayırmadaki etkisine göre bir ağırlığı vardır. Kullanılan alanlar ve ağırlıkları veriden otomatik olarak hesaplanır.

Kredi talep kararlarının manuel değil de otomatik olarak verilmesinin birçok yararı vardır:

- Daha doğru karar vermeyi sağlar; kârı artırır.
- Bir skor kartı standardizasyonu ve nesnel karar getirir; kişilerin değerlendirmesi öznel olabilir. Böylece kurumun bütün şubelerinde tutarlı olarak aynı kararın verilmesi sağlanır.

- Hızlı ve ucuzdur; talebin yoğun olduğu durumlarda kredi başvuruları konusundaki kararların manuel verilebilmesi olası değildir.
- Kullanılan tek skor kartının üzerinde kurumun kontrolü vardır ve örneğin eşik değeri ile oynayarak müşterilerin kalitesi denetlenebilir.
- Müşterilerin (veya genel ekonominin) değişmesi durumunda (*population drift*), yeni skor kartı otomatik olarak kolay ve hızlı bir şekilde hesaplanabilir.

Skor kartının oluşturulabilmesi için geçmiş müşterilerin iyi risk ve kötü risk olarak gruplanması gerekir. İyi risk, örneğin hiç geç ödemesi olmayan müşteri, kötü risk de, örneğin üç veya daha fazla arka arkaya geç ödeme yapmış müşteriler olabilir.

Kötü riski tanımlamak kolay değildir; belki geç ödemeler daha yüksek faiz nedeniyle kurum için kârlı olabilir. Gerçekte tanımlamak istediğimiz zarara neden olan müşteridir. Yapmak istediğimiz kârlı ve zararlı müşterileri birbirinden ayırabilmektir.

Bu şekilde kârlı ve zararlı müşterilerin bilgileri iki grup olarak verildikten sonra genelde doğrusal regresyon veya doğrusal ayırıcı (*linear discriminant*) kullanılarak alanların ağırlıkları hesaplanır. Bu ağırlıklar hesaplandıktan sonra kabul/red eşiği hesaplanır. Bunun için her iyi müşterinin kuruma kaç birim kârlı, her kötü müşterinin kuruma kaç birim zararlı olduğunun verilebilmesi gerekir. Hangi değer üstünde beklenen toplam kâr beklenen toplam zararı aşarsa o değer eşik olarak kabul edilir.

Kredi skoru kredi talebini kabul/red kararı dışında da kullanılabilir: Kredi verirken riske bağlı fiyatlandırma yapılabilir. Örneğin yüksek riskli bir müşteriye reddetmek yerine daha yüksek bir faizle veya daha küçük bir miktarla kabul edebilir. İyi müşteriler için başka ürünler (*cross-selling*), daha yüksek kredi limitleri gibi kararlar verebilir. Davranış skorlamada kredi skoruna göre eylemimiz değişir: Geç ödeme durumunda iyi bir müşteriye hiç rahatsız etmeyiz, başka bir müşteriye telefon açarız, daha kötü (skoru daha düşük) bir müşteri için kanuni takibe gideriz.

Sınıflandırma başarısı bir skor kartı sisteminin tek performans kıstası değildir. Bundan başka skor hesap hızı, gerektiğinde skor kartının güncelleştirilme hızı, skor kartının anlaşılma kolaylığı ve verilen kararın kolay açıklanabilmesi önemlidir.

Bir müşteri için kabul kararı verince zaman içinde o müşterinin kârlı mı, zararlı mı olduğu görülebilir ama reddedilen bir müşteri için gerçek sonuç bilinemez. Bu da örnek kümenin gittikçe daralması anlamına gelir. Buna neden olmamak ve skor

kartın sınıflandırma başarısını iyileştirebilmek için normalde red edilecek müşterilerle ilgili ek bilgi almaya çalışmak yararlıdır. Bunun için örneğin kredi izleme bürosundan o kişiye kredi vermiş başka finans kurumlarının o kişi ile ilgili bilgisine başvurulabilir; ya da reddetmek yerine daha küçük bir miktarla müşteri olarak kabul edilebilir veya böyle kişilerden çok azı denemek ve bilgi kazanmak için kabul edilebilir.

10.2.6. Sistem Gereklere

Esneklik : Sistem esnek bir yapıyı temel almalı. Yeni çıkan riskli bir ürünü içerebilmeli ve/veya risk değerlendirme açısından önem taşıyan yeni bir sorunsal değerlendirme kısıtlarına kolaylıkla dahil edebilmelidir.

Sonuçların Kaydedilmesi : Değerleme sonuçları, “tarih” ve “ürün tercihi” bazında kaydedilmelidir.

Raporlar, MIS Raporları :

- Müşteri Risk Değerleme Sistemi’nin Genel Performans Raporları
- Müşteri Risk Değerleme Sistemi sonuçlarının risk grubu, dönem, ürün ve kanal bazında dağılım raporları

Farklı Değerleme Modelleri : Sistemde sadece tek tip bir risk değerlendirme modeli olmamalıdır. Farklı müşteri segmentlerine göre risk kriterlerinin değişebildiği pluralist bir risk değerlendirme modeli olmalıdır. Farklı risk modelleri rahatlıkla aynı anda çalışabilmelidir.

Yeni müşteri için oluşturulan değerlendirme modeli sadece başvuru formundaki bilgilerle sınırlı iken, mevcut müşterilerde, ürün varlıkları, karlılıkları ve ödeme alışkanlıkları gibi başka faktörleri de içeren daha kapsamlı bir değerlendirme modeli kurulmalıdır.

Geriye Dönük Taramalar : Risk değerlendirme yaklaşımında değişiklik olduğunda ve/veya riskli ürünlerin risk içeriği değiştiği takdirde (kefil istenen bir ürünün kefilsiz hale gelişi gibi) talep halinde zaman zaman sistem “yüksek risk” ve “komite tarafından reddedilmiş orta risk” grubu içinde yer verdiği müşteriler için yeni risk değerlendirme yaklaşımını esas alarak tekrar bir değerlendirme yapmalıdır. Böylelikle başvurusu kabul olmamış kişiler için yeni yaklaşımda kabul edilme olasılığının varlığı araştırılabilmelidir.

İşletme Dışı Risk Bilgileri İle Entegrasyon : Sistem banka dışı risk bilgileri içeren veritabanları (TCMB Risk Listesi, Haciz, İflas Bilgileri, Kara Liste Kontrolleri gibi) ile entegre olmalıdır. Bu türden risk bilgileri değerlendirme süreci içinde dikkate alınmalıdır.

Başvuru aşamalarının izlenmesi : Başvurusunun akıbetini araştıran müşteriye, hangi aşamada olduğu ile ilgili bilginin, her kanaldan hızlı verilebilmesi sağlanmalıdır.

10.2.7. Davranışsal Risk Kıstasları

- Limit kullanım oranı
- Müşterinin kişisel bilgilerindeki değişimin korelasyonu
- Sık iş değiştirme
- Gelir/Gider değişimleri
- Müşterinin riskli olmayan ürünleri/hizmetleri kullanma biçimi
- Otomatik veya düzenli ödemeleri için zamanında fon ayırıp ayırmadığı
- Tasarruf etme oranı
- Hesaplarını aktif tutma eğilimi
- Yatırım araçlarındaki risk eğilimi (risk sever, muhafazakar, vb.)
- Çek karnesi kullanma alışkanlığı
- Müşteri kredi borcunu zamanında ödüyor mu?
- Dönem içinde gecikme sayısı
- Dönem içinde gecikme gün sayısı toplamı
- Dönem içinde temerrüde düşen borcun asıl borca oranı
- Dönem içinde müşterinin borcunu taksitlendirme yüzdesi
- Dönem içinde müşterinin kart kaybetme, çaldırma eğilimi
- Vb.

10.2.8. Proje Planı

Söz konusu Proje Çalışması'na göre öncelikli olarak “başvuru risk değerlendirme” sürecinin otomasyona geçirilmesi esastır. “Başvuru Risk Değerleme”nin ardından oluşturulan yapı “Davranışsal Risk Değerleme” yapısı ile entegre edilmelidir.

Proje planında aşağıdaki adımlar izlenecektir :

1. Proje Grubu'nun Oluşturulması
2. İş Kararlarının Çıkarılması ve Modelin Kurulması
3. Analiz ve İş Akışı Fonksiyonların Belirlenmesi
4. Müşteri Değerleme Sisteminin Programlanması
5. Müşteri Değerleme Sisteminin Test Edilmesi
6. Başvuruların İş Akışında Modelin Entegre Edilmesi
7. Başvuru Sürecinin Tamamının Test Edilmesi
8. Uygulamaya Alınması ve Performansının Ölçülmesi

10.3. Verilerin Hazırlanması

Modelin kurulması aşamasında ortaya çıkacak sorunlar, bu aşamaya sık sık geri dönülmesine ve verilerin yeniden düzenlenmesine neden olacaktır. Bu durum verilerin hazırlanması ve modelin kurulması aşamaları için, veri keşfi sürecinin toplamı içerisinde enerji ve zamanın % 50 - % 85'inin harcanmasına neden olmaktadır.

Verilerin hazırlanması aşaması kendi içerisinde, veri ambarının oluşum sürecine benzer şekilde, toplama, değer biçme, birleştirme ve temizleme, seçme ve dönüştürme adımlarından meydana gelmektedir. Kurumun, veri madenciliği uygulamalarını destekleyebilecek, iyi tasarlanmış bir veri ambarı bulunuyorsa, uygun veriye erişim kolaylaşır, böylece bu adımda harcanacak zamandan tasarruf ederek, sonuca kısa sürede ulaşılır.

10.3.1. Toplama (Collection)

Tanımlanan problem için gerekli olduğu düşünülen verilerin ve bu verilerin toplanacağı veri kaynaklarının belirlenmesi adımdır. Verilerin toplanmasında

kuruluşun kendi veri kaynaklarının dışında, merkez bankası kara listesi gibi veri tabanlarından faydalanılabilir.

10.3.2. Değer Biçme (Assessment)

Veri madenciliğinde kullanılacak verilerin farklı kaynaklardan toplanması, doğal olarak veri uyumsuzluklarına neden olacaktır. Bu uyumsuzlukların başlıcaları farklı zamanlara ait olmaları, kodlama farklılıkları (örneğin bir veri tabanında cinsiyet özelliğinin e/k, diğer bir veri tabanında 0/1 olarak kodlanması), farklı ölçü birimleridir. Ayrıca verilerin nasıl, nerede ve hangi koşullar altında toplandığı da önem taşımaktadır. Bu nedenlerle, iyi sonuç alınacak modeller ancak iyi verilerin üzerine kurulabileceği için, toplanan verilerin ne ölçüde uyumlu oldukları bu adımda incelenerek değerlendirilmelidir.

10.3.3. Birleştirme ve Temizleme (Consolidation and Cleaning)

Bu adımda farklı kaynaklardan toplanan verilerde bulunan ve bir önceki adımda belirlenen sorunlar mümkün olduğu ölçüde giderilerek veriler tek bir veri tabanında toplanır. Ancak basit yöntemlerle ve baştan savma olarak yapılacak sorun giderme işlemlerinin, ileriki aşamalarda daha büyük sorunların kaynağı olacağı unutulmamalıdır.

10.3.4. Seçim (Selection)

Bu adımda kurulacak modele bağlı olarak veri seçimi yapılır. Örneğin tahmin edici bir model için, bu adım bağımlı ve bağımsız değişkenlerin ve modelin eğitiminde kullanılacak veri kümesinin seçilmesi anlamını taşımaktadır.

- Sıra numarası, kimlik numarası gibi anlamlı olmayan ve diğer değişkenlerin modeldeki ağırlığının azalmasına da neden olabilecek değişkenlerin modele girmemesi gerekmektedir. Bazı veri madenciliği algoritmaları konu ile ilgisi olmayan bu tip değişkenleri otomatik olarak elese de, pratikte bu işlemin kullanılan yazılıma bırakılmaması daha akılcı olacaktır.
- Verilerin görselleştirilmesine olanak sağlayan grafik araçlar ve bunların sunduğu ilişkiler, bağımsız değişkenlerin seçilmesinde önemli yararlar sağlayabilir.

- Genellikle yanlış veri girişinden veya bir kereye özgü bir olayın gerçekleşmesinden kaynaklanan verilerin (*Outlier*), önemli bir uyarıcı enformasyon içerip içermediği kontrol edildikten sonra veri kümesinden atılması tercih edilir.
- Modelde kullanılan veri tabanının çok büyük olması durumunda tesadüfiliği bozmayacak şekilde örnekleme yapılması uygun olabilir. Günümüzde hesaplama olanakları ne kadar gelişmiş olursa olsun, çok büyük veri tabanları üzerinde çok sayıda modelin denenmesi zaman kısıtı nedeni ile mümkün olamamaktadır. Bu nedenle tüm veri tabanını kullanarak bir kaç model denemek yerine, tesadüfi olarak örneklenmiş bir veri tabanı parçası üzerinde bir çok modelin denenmesi ve bunlar arasından en güvenilir ve güçlü modelin seçilmesi daha uygun olabilir.

10.3.5. Dönüştürme (Transformation)

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

Hem modelin oluşumunda, hem kullanımındaki başarı, verinin tamlığına bağlı olması ve gerçek hayatta çoğu müşteri bilgisinin noksan olması dikkate alındığında, bilgilerin tamamlanması yönünde, pazarlama gruplarına hedefler verilmelidir.

10.4. Modelin Kurulması ve Değerlendirilmesi

Tanımlanan problem için en uygun modelin bulunabilmesi, olabildiğince çok sayıda modelin kurularak denenmesi ile mümkündür. Bu nedenle veri hazırlama ve model kurma aşamaları, en iyi olduğu düşünülen modele varılıncaya kadar yinelenen bir süreçtir.

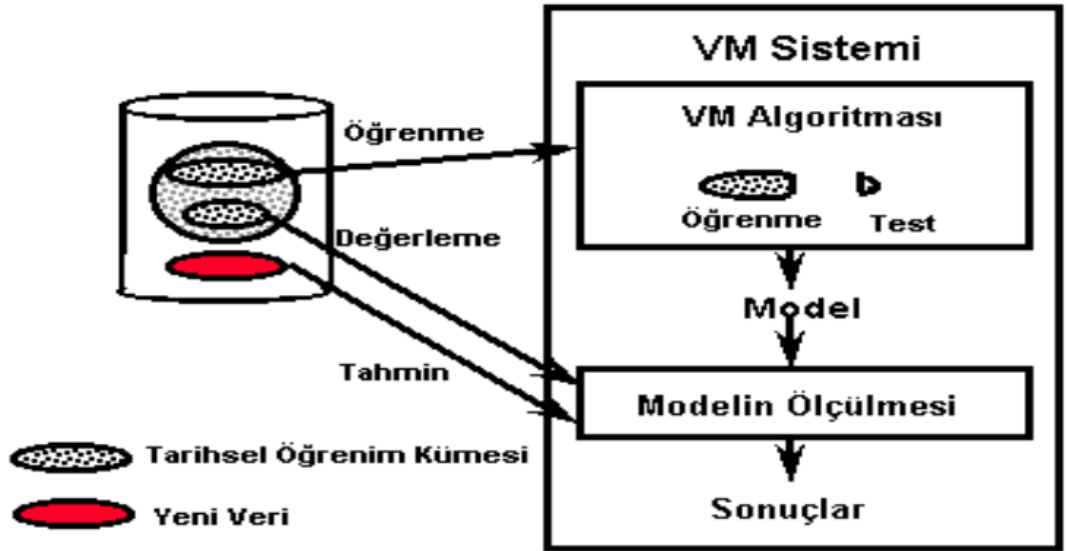
Model kuruluş süreci denetimli (*Supervised*) ve denetimsiz (*Unsupervised*) öğrenimin kullanıldığı modellere göre farklılık göstermektedir.

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

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

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

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



Şekil 10.4 Modelin Oluşumundaki Veri Madenciliği Prosesi

(Kaynak : Thearling, Kurt)

10.4.1. Müşteri Risk Değerleme Sistemi Tasarımı

Müşteri Risk Değerleme Sistemi'ne göre müşterinin başvuru formundan alınan verilerle **kredi notu** tespit olunur.

Davranış bilgilerinden elde edilen sonuç ile **risk notu** belirlenir.

Kredi notu, risk değerlendirme açısından “güvenlik ve istikrar unsurları” esas alındığında bir müşterinin risk değerini gösterir.

Tablo 10.1 Kredi Notu Değerleri

Değer	Puan
A	Çok Güvenilir, Kefilsiz ve Teminatsız Çalışılabilir
B	Güvenilir, Gerekğinde Kefil ve Teminat İstenebilir
C	Güvenilir, Kefil ve Teminat Zorunludur
D	Az Güvenilir
E	Çok Risklidir

Kredi notunun tespitinde dikkate alınan risk faktörleri tablodaki gibi sıralanabilir:

Tablo 10.2 Başvuru Skor Modelinin Girdileri

Referans Tablosu	Değer Grubu	Özellik
1	Demografik Değer	Cinsiyet
2	Varlık Gücü Değeri	Evin Mülkiyeti
3	İstikrar Değeri	Adreste Oturma Süresi
4	Demografik Değer	Uyruk
5	Demografik Değer	Doğum Tarihi
6	Demografik Değer	Medeni Durum
7	Ödeme Gücü Değeri	Bakmakla Yükümlü Kişi Sayısı
8	Demografik Değer	Eğitim Durumu
9	Ödeme Gücü Değeri	Çalışma Şekli
10	Ödeme Gücü Değeri	Sosyal Güvenlik Kurumu
11	Demografik Değer	Meslek
12	Ödeme Gücü Değeri	Kamu/Özel

13	İstikrar Değeri	İşyerinde Çalışma Süresi
14	Varlık Gücü Değeri	İşyerinin Mülkiyeti
15	Ödeme Gücü Değeri	İşyerinin Faaliyet Konusu
16	İstikrar Değeri	Kuruluş Yılı
17	Demografik Değer	Hisse Oranı
18	Ödeme Gücü Değeri	Çalışan Kişi Sayısı
19	Ödeme Gücü Değeri	Aylık Gelir
20	Ödeme Gücü Değeri	Aylık Gideri
21	Varlık Gücü Değeri	Gayrimenkul Bilgisi
22	Varlık Gücü Değeri	Otomobil Bilgisi

Risk durumu, talep edilen ürün/hizmet açısından müşterinin risk kriterlerinin yeterlilik düzeyini gösterir.

Tablo 10.3 Risk Notu Değerleri

Değer	Puan
DÜŞÜK	Komitesiz Kredi Verilebilir.
ORTA	Kredi Tahsisi Komite Kararına Bağlıdır
YÜKSEK	Kredi Verilemez

Yüksek Risk : Ürün ve/veya hizmet talebi “riskli” bulunan müşteri grubu

Orta Risk : Ürün ve/veya hizmet talebi “kısmen riskli” bulunan müşteri grubu olup ilgili ürün ve/veya hizmet talebi hakkında Müşteri Risk Değerleme Komitesi’nin karar vermesinin uygun olacağı müşteri grubu

Düşük Risk : Ürün ve/veya hizmet talebi açısından görece “riskli” bulunmayan müşteri grubu

Parametrik Koşullar :

Müşterinin “Çalışma Şekli” beyanına göre bazı risk faktörleri sadece belli çalışma şekli için değerlemeye katılacaktır. Şöyle ki;

Tablo 10.4 Parametrik Koşullar 1

	Serbest	Ücretli
Meslek		X
İşyerinde Çalışma Süresi		X
Sosyal Güvenlik Kurumu		X
Kamu / Özel		X
İşyerinin Mülkiyeti	X	
İşyerinin Faaliyet Konusu	X	
Kuruluş Yılı	X	
Hisse Oranı	X	
Çalışan Kişi Sayısı	X	

Eğer müşteri, gelirleri hakkında bir beyanda bulunur ve bakmakla yükümlü olduğu kişi sayısı hakkında da beyanda bulunmuşsa veya oturduğu evin mülkiyeti hakkında beyanda bulunmuşsa ve giderleri hakkına bir beyanda bulunmamışsa, gider kalemi için aşağıdaki tablo dikkate alınır:

Tablo 10.5 Parametrik Koşullar 2

	Kira /Aidat Gideri	Eğitim Gideri
Bakmakla Yükümlü Olduğu Kişi Sayısı		Her kişi sayısı için ; Kişi sayısı x Toplam gelirinin 1/10'u
Kiracı	Toplam Gelirinin 1/4'ü	
Kendi Evi /Lojman	Toplam Gelirinin 1/20'si	

Başvuru bilgilerinden hesaplanan kredi notu ile müşterinin önceki işlemlerinden hesaplanan risk notu ile birlikte kara liste sonucunun da dikkate alınması neticesinde müşteriye verilebilecek limitin tayin edilmesi aşağıdaki şekilde modellenmiştir.

Adı	Yaş	Gelir	Gider	Vb.	Kredi Notu	Limit Kul. (%)	Geç Öd.	Vb.	Risk Notu	Kara liste	Limit (\$)	Sonuç
A	20	1000	500		B	75	0		Orta	Evet	250	Red
B	25	300	100		C	100	2		Yüksek	Hayır	50	Red
C	45	1500	500		A	50	0		Düşük	Hayır	400	Kabul
D	50	1000	200		B	25	1		Düşük	Hayır	250	Kabul

Model kuruluşu çalışmalarının sonucuna bağlı olarak, aynı teknikle farklı parametrelerin kullanıldığı veya başka algoritma ve araçların denendiği değişik

modeller kurulabilir. Model kuruluş çalışmalarına başlamazdan önce, imkansız olmasa da hangi tekniğin en uygun olduğuna karar verebilmek güçtür. Bu nedenle farklı modeller kurarak, doğruluk derecelerine göre en uygun modeli bulmak üzere sayısız deneme yapılmasında yarar bulunmaktadır.

10.4.2. Sinir Ağı Yöntemi ile Oluşturulacak Modelin Kurulumu

Ağın girdi ve çıktı katmanları haricinde, gizli katman sayısı belirlenerek ağın topolojisi belirlenir. Kredi riski modellerinde, genellikle gizli katman içermeyen basit sinir ağları kullanılır.

Öğrenme fazında sinir ağı örnek kümedeki veriler üzerinden öngördüğü çıktı değerini, gerçeği ile karşılaştırıp tüm iç ağırlıklarını yeniden ayarlamaktadır. En doğru değeri döndürecek optimum ağırlık setine ulaşınca kadar işlemleri defalarca tekrarlamaktadır.

Öğrenim Kümesi Seçiminde dikkat edilmesi gereken hususlar :

- Veriler güncel olmalıdır
- Tüm özelliklerin değerlerini kapsamalı
- Sonuca etki etmeyecek özellikler çıkarılmalı (Öngöründe kullanılacak en önemli özelliklerin belirlenmesi için farklı yöntemler vardır. İstatistik korelasyonlar önem önceliklerin belirlenmesinde kullanılabilir. Bir diğer yöntem karar ağaçlarının kullanımıdır. Ağacın üst seviyelerindeki özelliklerin öngörü gücü daha yüksektir. Bu özellikler daha sonra sinir ağları için girdi olarak kullanılabilir.)
- Olası tüm çıktılar için örnek içermeli
- Girdi ve çıktı değerleri 0 ile 1 arasında bir değere dönüştürülmelidir.

10.4.3. Kurulan Modelin Doğruluk Derecesinin Değerlendirilmesi

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

Tablo 10.6 Risk Matrisi

	<i>Fiili</i>		
<i>Tahmini</i>	<i>A Sınıfı</i>	<i>B Sınıfı</i>	<i>C Sınıfı</i>
<i>A Sınıfı</i>	45	2	3
<i>B Sınıfı</i>	10	38	2
<i>C Sınıfı</i>	4	6	40

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

Kaldıraç (*Lift*) oranı ve grafiği, bir modelin sağladığı faydanın değerlendirilmesinde kullanılan önemli bir yardımcıdır. Örneğin kredi kartını muhtemelen iade edecek müşterilerin belirlenmesinde kullanılan modelin belirlediği 100 kişinin 35'i gerçekten bir süre sonra kredi kartını iade ediyorsa ve tesadüfi olarak seçilen 100 müşterinin aynı zaman diliminde sadece 5'i kredi kartını iade ediyorsa kaldıraç oranı 7 olarak bulunacaktır.

Kurulan modelin değerinin belirlenmesinde kullanılan diğer bir ölçü, model tarafından önerilen uygulamadan elde edilecek kazancın bu uygulamanın gerçekleştirilmesi için katlanılacak maliyete bölünmesi ile edilecek olan yatırımın geri dönüş (*Return On Investment*) oranıdır.

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

10.5. Modelin Kullanılması

Kurulan ve geçerliliği kabul edilen model doğrudan bir uygulama olabileceği gibi, bir başka uygulamanın alt parçası olarak kullanılabilir. Kurulan modeller risk analizi, kredi değerlendirme, dolandırıcılık tespiti gibi işletme uygulamalarında doğrudan kullanılabilir. Promosyon planlaması simülasyonuna entegre edilebilir.

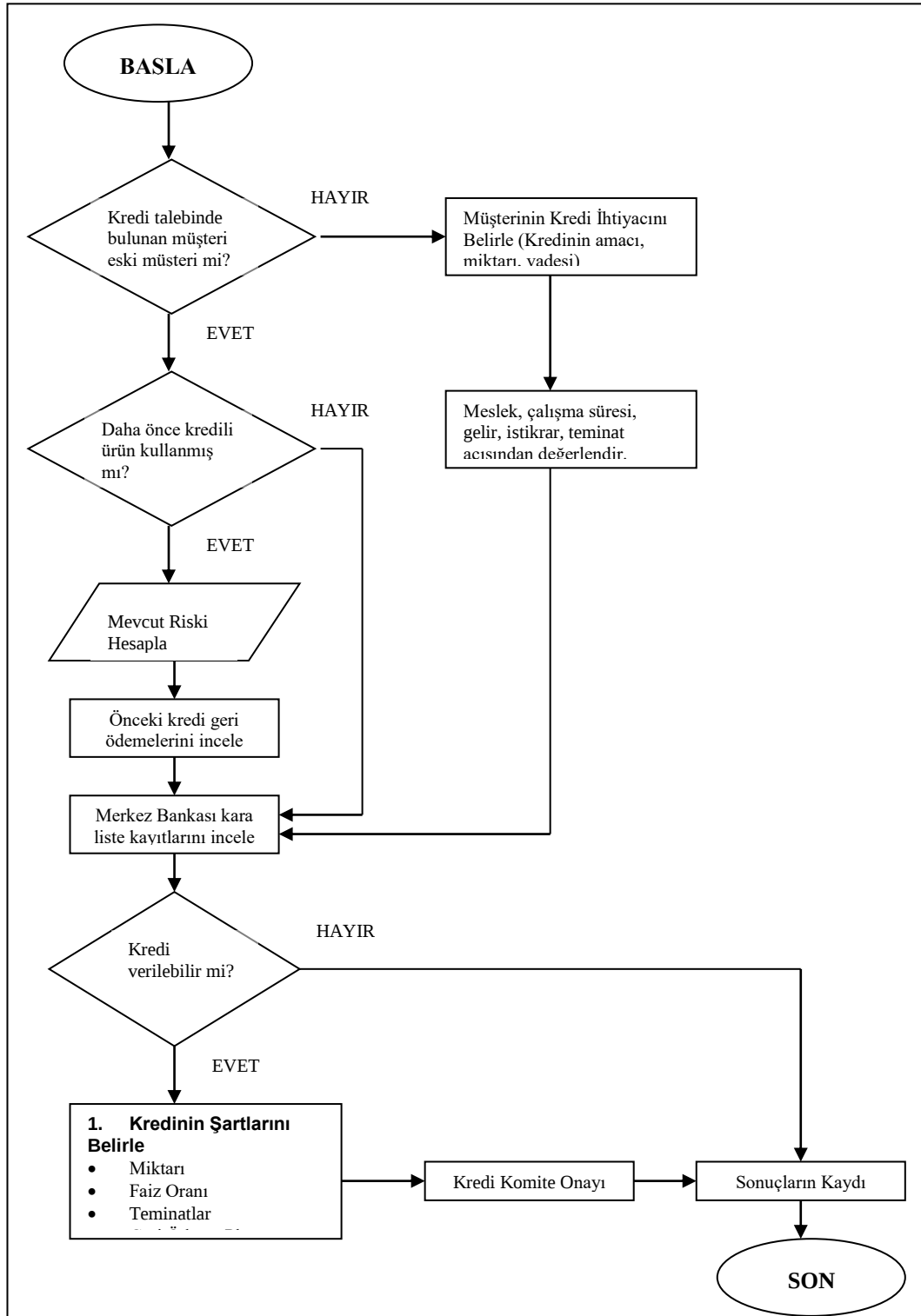
10.6. Modelin İzlenmesi

Zaman içerisinde bütün sistemlerin özelliklerinde ve dolayısıyla ürettikleri verilerde ortaya çıkan değişiklikler, kurulan modellerin sürekli olarak izlenmesini ve gerekiyorsa yeniden düzenlenmesini gerektirecektir. Tahmin edilen ve gözlenen değişkenler arasındaki farklılığı gösteren grafikler model sonuçlarının izlenmesinde kullanılan yararlı bir yöntemdir.

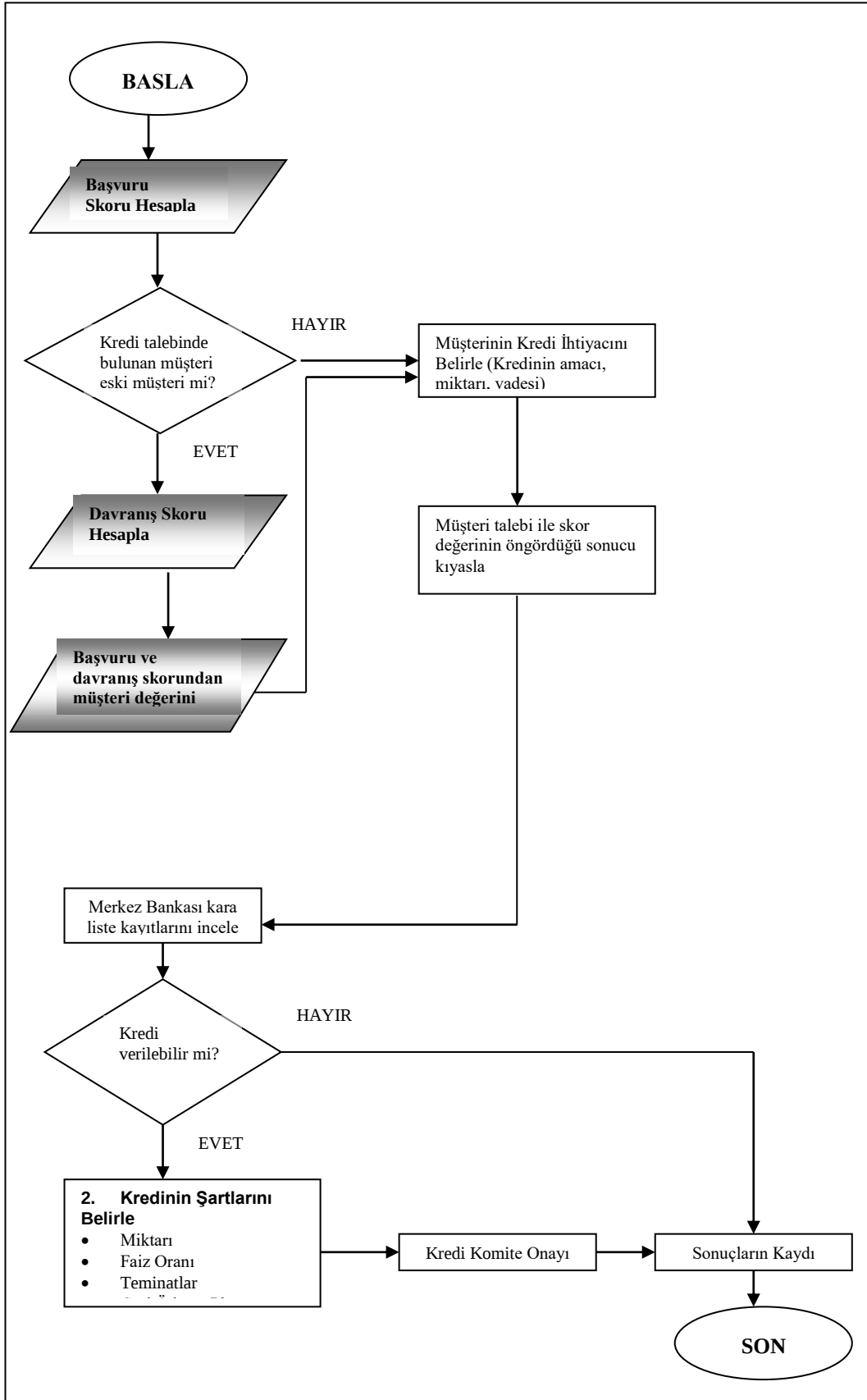
10.7. Sonuç

İşletmenin üstlenebileceği riski belirleyecek bir model olması ile birlikte, sonucun red olacağı tahmin edilen müşteriye kredi pazarlamayarak sebepsiz bir memnuniyetsizliğin önüne geçilmiş olunur.

Benzer bir yöntemle, müşterinin yatırım ürünlerinde üstlenebileceği riski tahmin ederek en uygun ürün portföyü belirlenebilir.



Şekil 10.5 Kredi Taleplerini Değerlendirme Süreci İş Akış Şeması



Şekil 10.6 Kredi Taleplerini Otomatik Değerlendirme Süreci İş Akış Şeması

11. SONUÇ

Günümüzün yoğunlaşan rekabet koşullarında firmaların karlılıklarını arttırmaları, maliyetlerini düşürmeleri, yeni pazarlar yaratmaları ve rakiplerine üstünlük sağlamaları için stratejik kararlara ihtiyaç duyarlar. Doğru kararların alınabilmesi, öncelikle doğru karar tanımına ve bu kararı sağlayacak verinin tutarlı olmasına bağlıdır. Özellikle müşteriye sınırsız hizmet imkanı tanıyan çağrı merkezleri ve Internet gibi yeni teknolojilerin gelişimi, önceleri ürün bazlı yönetim modellerindeki zamanında üretime alternatif olarak, zamanında yönetim kavramını gündeme getirmektedir. Alınan kararlar doğru zamanda aksiyona dönüştürülemiyorsa değerini hızla yitirmektedir. Kararların güncelliği, onları kullanacak kişilerin yeteneği, başarı açısından önem arz etmektedir. Bunların sağlanması ancak bir “Bilgi Ekolojisinin” kurulması ile mümkündür. 1990’lı yıllarda başlayan Bilgi Yönetimi akımı 2000’li yıllar gelişen Bilgi Ekolojisi akımı ile devam etmektedir. Aşağıdaki tabloda iki akımı karşılaştırılması yapılmıştır. (Kaynak : Intelligence Labs)

Tablo 11.1 Bilgi Yönetimi ile Bilgi Ekolojisi Akımlarının Kıyaslanması

Bilgi Yönetimi	Bilgi Ekolojisi
Aksiyona dönüştürülebilecek enformasyon ve fırsatlar sunar.	Enformasyonun kullanımı, fırsatın farkına varılması ve bunların bilgi ve aksiyona dönüştürülmesi için gerekli zaman, yer, durum gibi tüm detayları, sinerji ve güveni de sağlar
Vurgu, bilgi konuları, entelektüel değerlerin korunması ve teşvik edilmesi, enformasyon mimarisi, bilgi akışı ve kurallarının denetlenmesi ve iyileştirilmesi üzerinedir.	Vurgu, kültür, bilginin geliştirilmesi, kullanımı kolay sistemler, örüntü tanıma, prototip oluşturma, bilginin yaratılması ve kullanımı üzerinedir.
Tabana yönelik; Bilgideki fırsat ve tehlikeleri görerek, bunları değerlendirme, organize etme, tanımlama ve kazanca dönüştürme imkanı sağlar.	Kurumla ilişkisi bulunan herkese yönelik; Çıkabilecek bilgiden, ilişkileri sürdürmek ve geliştirmek için hangilerinin kullanılabileceğini görme imkanı sağlar.
Bilginin dağıtımı, ulaşım politikaları ve uygun olarak temin edilmesi üzerine odaklanır.	Tüm çalışanların yorumlar konusunda hemfikir olmalarının sağlanması, ortak bir anlayışın oluşturulması, paylaşılan içerik anlamının incelenmesi ile ilgili politikalara odaklanır. Bir düzen geliştirmekle birlikte, kontrolünde belirli düzeyde esneklik tanınır.
Entelektüel “madde” ile ilgilidir. “Parçacık” veya “nesne” ye odaklanır : kurallar, bilgi transferinin planlanması, en iyi uygulamalar, patent, doküman koleksiyonu, en sık sorulan sorular, metrikler.	Entelektüel “enerji” ile ilgilidir. “Dalga” veya “ilişki” ye odaklanır: güven, bilginin yaratılması, anlam, inanç, diyalog, görüş, yenilik, yaratıcılık.

KAYNAKLAR

- [1] **Akpınar, Prof. Dr. Haldun**, 2000. Veri tabanlarında Bilgi Keşfi ve Veri Madenciliği. İşletme Fakültesi Dergisi. C.29 Nisan 2000
- [2] **Berry, Michael J. A. & Linoff, Gordon**, 1997. Data Mining Techniques For Marketing, Sales and Customer Support, John Wiley&Sons
- [3] **Berson, M.A. and Smith, S.**, 1997. Data Warehousing, Data Mining and OLAP McGraw-Hill, New York
- [4] **Berson, Alex. Smith, Stephen and Thearling, Kurt**, 1998. Building Data Mining Applications for CRM
- [5] **Bigus, Joseph P.** 1996. Data Mining with Neural Networks : Solving Business Problems – From Application Development to Decision Support, New York : McGraw-Hill.
- [6] **Brackett, Michael H**, 1994. Data Sharing Using Common Data Architecture, John Wiley
- [7] **Brackett, Michael H.**, 1996. Open OLAP, DBMS, April 1998, The Data Warehouse Challenge : Taming Data Chaos. John Wiley
- [8] **Geiger, Jonathan G.** 1997. Data Warehousing : Supporting Business Intelligence
- [9] **Peppers, Don & Martha Rogers**, 1997. Enterprise One to One : Tools for Competing in The Information Age, New York : Doubleday.
- [10] **Piatetsky-Shapiro, G. and Frawley, W.J.**, 1991. Knowledge Discovery in Databases., MIT Press.
- [11] **Reichheld, Frederick F.**, 1996. The Loyalty Effect : The Hidden Force Behind Growth, Profits, and Lasting Value, Boston : Harvard Business School Press

[12] Smyth, P. and Uthurusamy R., 1996. Advances in Knowledge Discovery and Data Mining, MIT Press

[13] Wu, Johnathan, 2000. Business Intelligence : The Transition of Data into Wisdom. DM Review. November 2000

ÖZGEÇMİŞ

Handeris Seldan ÇERKEZ, 1974 Romanya doğumludur. 1997 yılında İstanbul Teknik Üniversitesi Fen-Edebiyat Fakültesi Matematik Mühendisliği Bölümü'nü bitirmiştir. 1999 yılından bu yana özel bir bankanın Bilgi Sistemleri Bölümü'nde Uygulama Geliştirme Uzmanı olarak çalışmaktadır.