

AKDENİZ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

Özcan ASILKAN

VERİ MADENCİLİĞİ KULLANILARAK
İKİNCİ EL OTOMOBİL PAZARINDA FİYAT TAHMİNİ

Danışman

Prof.Dr. Ayşe KURUÜZÜM

İşletme Anabilim Dalı

Doktora Tezi

Antalya, 2008

Akdeniz Üniversitesi
Sosyal Bilimler Enstitüsü Müdürlüğüne,

Özcan ASILKAN'ın bu çalışması, jürimiz tarafından İşletme Anabilim Dalı Doktora tezi olarak kabul edilmiştir.

Başkan

: Prof. Dr. İbrahim GÜNGÖR

Üye (Danışmanı)

: Prof. Dr. Ayşe KURUŞÖM

Üye

: Doç. Dr. Can Deniz KÖKSAL

Üye

: Yrd. Doç. Dr. Esir KÜŞKÜSİLLE

Üye

: Yrd. Doç. Dr. Selvak HELHEL

Tez Konusu:

Veri Madenciliği Kullanılarak İkinci El Otomobil Pazarlarında Fiyat Tahmini

Onay : Yukarıdaki imzaların, adı geçen öğretim üyelerine ait olduğunu onaylarım.

Tez Savunma Tarihi : 5.1.2008

Mezuniyet Tarihi :/200

Prof. Dr. Burhan VARKIVANÇ
Müdür

İÇİNDEKİLER

| | Sayfa |
|---|--------------|
| ŞEKİLLER LİSTESİ | vi |
| TABLolar LİSTESİ..... | vii |
| KISALTMALAR LİSTESİ | viii |
| ÖZET | ix |
| SUMMARY | x |
| ÖNSÖZ | xi |
| GİRİŞ | 1 |
| 1. VERİTABANLARI VE VERİ AMBARLARI..... | 3 |
| 1.1. Veri ve Bilgi Kavramları | 3 |
| 1.2. Veritabanı..... | 4 |
| 1.2.1. Çeşitleri | 5 |
| 1.2.2. Yararları | 6 |
| 1.2.3. Katmanları..... | 7 |
| 1.3. Veri Ambarı | 8 |
| 1.3.1. Unsurları | 9 |
| 1.3.2. Özellikleri | 11 |
| 1.3.3. Kullanıcıları | 13 |
| 1.3.4. Mimarisi | 13 |
| 1.3.5. Süreçleri | 14 |
| 1.3.6. Yararları | 15 |
| 1.3.7. Dezavantajları | 16 |
| 1.3.8. Veri Ambarı Projelendirmesi | 16 |
| 1.3.9. Veri Depoları | 17 |
| 1.4. Veritabanları ile Veri Ambarları Arasındaki Farklar..... | 18 |
| 1.5. Bilgi Keşfi Kavramı..... | 19 |
| 2. VERİ MADENCİLİĞİ VE TAHMİN YÖNTEMLERİ..... | 21 |
| 2.1. Veri Madenciliği | 21 |
| 2.1.1. Tanımı | 22 |

| | |
|---|----|
| 2.1.2. Gereksinimleri..... | 23 |
| 2.1.3. Amaçları..... | 23 |
| 2.1.4. Karşılaşılan Problemler..... | 24 |
| 2.1.5. Kullanım Alanları | 25 |
| 2.1.6. Etkilendiği Unsurlar..... | 27 |
| 2.1.7. Modelleri..... | 28 |
| 2.1.7.1. Sınıflama ve Regresyon Modelleri..... | 29 |
| 2.1.7.2. Kümeleme Modelleri | 30 |
| 2.1.7.3. Birliktelik Kuralları ve Ardışık Zamanlı Örüntüler | 31 |
| 2.1.8. Yöntemleri | 31 |
| 2.1.9. Uygulama Süreci..... | 32 |
| 2.2. Tahminleme | 35 |
| 2.2.1. Tahmin Kavramı | 35 |
| 2.2.2. Tahmin Yöntemleri..... | 36 |
| 2.2.2.1. Nitel Tahmin Yöntemleri | 37 |
| 2.2.2.2. Nicel Tahmin Yöntemleri | 37 |
| 2.2.3. Tahmin Modellerinin Doğruluklarının Ölçülmesi..... | 38 |
| 2.3. Regresyon Analizi..... | 40 |
| 2.3.1. Tanımı | 40 |
| 2.3.2. Çeşitleri..... | 41 |
| 2.3.3. Basit Doğrusal Regresyon..... | 41 |
| 2.3.4. Çoklu Doğrusal Regresyon | 42 |
| 2.4. Zaman Serileri Analizi..... | 43 |
| 2.4.1. Zaman Serilerinin Tanımı ve Unsurları | 43 |
| 2.4.2. Zaman Serilerinin Temel Özellikleri | 44 |
| 2.4.2.1. İç Bağımlılık..... | 44 |
| 2.4.2.2. Stokastik (Tasadüfi) Süreç Olması..... | 45 |
| 2.4.3. Zaman Serilerinin Sınıflandırılması..... | 45 |
| 2.4.4. Analiz Amaçları ve Uygulama Alanları | 46 |
| 2.4.5. Analiz Aşamaları | 47 |
| 2.4.5.1. Tanımlama..... | 47 |
| 2.4.5.2. Açıklama | 47 |
| 2.4.5.3. Tahminleme..... | 47 |
| 2.4.5.4. Kontrol | 47 |
| 2.4.6. Zaman Serilerinde İleriye Dönük Tahmin Yöntemleri | 48 |

| | |
|--|-----------|
| 2.4.6.1. Trend Analizi | 49 |
| 2.4.6.2. Mekanik Tahmin | 49 |
| 2.4.6.3. Hareketli Ortalamalar | 50 |
| 2.4.6.4. Üstel Düzleştirme | 50 |
| 2.4.6.5. Uyarlayıcı Arındırma | 51 |
| 2.4.6.6. Box-Jenkins Yöntemleri | 52 |
| 2.5. Yapay Sinir Ağları | 57 |
| 2.5.1. Yapay Zekâ Kavramı | 57 |
| 2.5.2. Tanımı | 59 |
| 2.5.3. Yapısı | 59 |
| 2.5.3.1. Yapay Sinir Hücresi | 60 |
| 2.5.3.2. Katmanları | 61 |
| 2.5.4. Süreçleri | 61 |
| 2.5.4.1. İşlem Süreci | 61 |
| 2.5.4.2. Eğitim ve Öğrenim Süreci | 63 |
| 2.5.5. Mimarisi | 67 |
| 2.5.6. Uygulama Alanları | 68 |
| 2.5.7. Üstünlükleri ve Eksikleri | 69 |
| 2.5.8. Modelin Oluşturulması | 70 |
| 2.5.8.1. Verileri Normalleştirme Yöntemi | 71 |
| 2.5.8.2. Eğitim Kümesi ve Test Kümelerinin Oranı | 71 |
| 2.5.8.3. Ağ Mimarisi | 72 |
| 2.5.8.4. Katman Sayısı | 72 |
| 2.5.8.5. Nöron Sayıları | 72 |
| 2.5.8.6. Aktivasyon Fonksiyonu | 73 |
| 2.5.8.7. Öğrenme Algoritması ve Parametreleri | 73 |
| 2.5.8.8. Doğruluk Ölçüleri | 74 |
| 2.6. Yazın Taraması | 74 |
| 3. OTOMOTİV ENDÜSTRİSİ ve İKİNCİ EL OTOMOBİL PAZARI | 79 |
| 3.1. Otomotiv Endüstrisi | 79 |
| 3.1.1. Tanımı ve Kapsamı | 79 |
| 3.1.2. Tarihçesi | 81 |
| 3.1.3. Ürünleri | 83 |
| 3.1.3.1. Otomotiv Ana Sanayi Ürünleri | 83 |

| | |
|---|------------|
| 3.1.3.2. Otomotiv Yan Sanayi Ürünleri | 85 |
| 3.1.4. Ekonomideki Önemi | 85 |
| 3.1.5. Üretim Alanındaki Gelişmeler | 88 |
| 3.1.6. Güncel Durum..... | 89 |
| 3.1.7. Lider Otomobil Üreticileri | 90 |
| 3.1.8. Avrupa'daki Durumu | 94 |
| 3.1.9. Türkiye'deki Durumu | 95 |
| 3.2. İkinci El Otomobil Pazarı | 96 |
| 3.2.1. Tanımı ve Kapsamı | 96 |
| 3.2.2. Önemi..... | 97 |
| 3.2.3. Tüketicilerin İkinci El Otomobiller Hakkındaki Tutumları..... | 98 |
| 3.2.4. Pazardaki Kurumsallaşma..... | 99 |
| 3.2.5. Otomobil Borsası | 101 |
| 3.2.6. İkinci El Otomobil Alırken Dikkat Edilmesi Gerekenler | 101 |
| 3.2.7. İkinci El Otomobil Pazarının Avrupa' daki Durumu..... | 103 |
| 3.2.8. İkinci El Otomobil Pazarının Almanya' daki Durumu | 104 |
| 3.2.9. İkinci El Otomobil Pazarının Türkiye'deki Durumu | 104 |
| 3.2.10. İkinci El Otomobil Pazarında E-Ticaret Yoluyla Satış..... | 106 |
| 3.2.10.1. Avrupa' daki Önemli İkinci El Otomobil E-Ticaret Siteleri..... | 107 |
| 3.2.10.2. Türkiye' deki Önemli İkinci El Otomobil E-Ticaret Siteleri..... | 109 |
| 4. İKİNCİ EL OTOMOBİL FİYATLARININ TAHMİN EDİLMESİ: AVRUPA | |
| ÜLKELERİ UYGULAMASI..... | 111 |
| 4.1. Araştırmanın Amacı, Kapsamı ve Önemi..... | 111 |
| 4.2. Araştırmanın Yöntemi | 113 |
| 4.2.1. Veri Kaynağı..... | 113 |
| 4.2.2. Veri Analizi..... | 114 |
| 4.2.3. Uygulama Araçları..... | 114 |
| 4.2.4. Örneklem..... | 117 |
| 4.3. Uygulama Süreci..... | 118 |
| 4.3.1. İşin Kavranması | 118 |
| 4.3.2. Verilerin Toplanması ve Kavranması | 118 |
| 4.3.3. Verinin hazırlanması | 119 |
| 4.3.4. Modellerin Kurulması ve Değerlendirme | 125 |
| 4.3.4.1. Güncel Fiyatları Tahmin Etmeye Yönelik Modelleme..... | 125 |

| | |
|--|------------|
| 4.3.4.2. Gelecekteki Fiyatları Tahmin Etmeye Yönelik Modelleme | 133 |
| 4.3.5. Uygulamaya Koyma | 140 |
| SONUÇ VE ÖNERİLER..... | 143 |
| KAYNAKÇA..... | 146 |
| ÖZGEÇMİŞ | 157 |

ŞEKİLLER LİSTESİ

| | |
|---|-----|
| Şekil 1.1. Veri ambarının işletme bilgi teknolojileri içindeki yeri | 11 |
| Şekil 1.2. Veri depolarını içeren veri ambarı mimarisi..... | 18 |
| Şekil 1.3. Veritabanlarında bilgi keşfi süreci..... | 20 |
| Şekil 2.1. Veri madenciliği uygulama süreci (CRISP-DM) | 33 |
| Şekil 2.3. Yapay sinir hücresi | 60 |
| Şekil 2.4. Çok katmanlı yapay sinir ağı yapısı | 61 |
| Şekil 4.1. Kayıp analizlerinin özet sayfası..... | 121 |
| Şekil 4.2. Portal tabanlı kayıp analizi | 123 |
| Şekil 4.3. Marka tabanlı kayıp analizi | 124 |
| Şekil 4.4. Veri Kümesi 1'e ait Ocak 2005 – Aralık 2007 arası aylık fiyat değişimi | 135 |
| Şekil 4.5. Veri Kümesi 2' ye ait Ocak 2005 – Aralık 2007 arası aylık fiyat değişimi | 136 |
| Şekil 4.6. Veri Kümesi 3' e ait Ocak 2005 – Aralık 2007 arası aylık fiyat değişimi | 136 |
| Şekil 4.7. İkinci amaca uygun yapay sinir ağları modelinin şematik yapısı..... | 139 |
| Şekil 4.8. Veri kümesi 1 için gerçek ve tahmini fiyatlar | 141 |
| Şekil 4.9. Veri kümesi 2 için gerçek ve tahmini fiyatlar | 142 |
| Şekil 4.10. Veri kümesi 3 için gerçek ve tahmini fiyatlar | 142 |

TABLOLAR LİSTESİ

| | |
|---|-----|
| Tablo 2.1. Yapay zekâ ile doğal zekânın karşılaştırılması | 58 |
| Tablo 2.2. Toplama fonksiyonu işlemleri | 62 |
| Tablo 2.3. Aktivasyon fonksiyonları | 63 |
| Tablo 2.4. Öğrenme algoritmaları..... | 66 |
| Tablo 2.5. Veri madenciliğinin tahmin amaçlı kullanıldığı çalışmalar | 75 |
| Tablo 3.1. Yıllar itibariyle dünya motorlu araç üretimi (Bin adet)..... | 88 |
| Tablo 3.2. Dünya otomobil üretimi (Bin adet) | 91 |
| Tablo 3.3. Otomotiv endüstrisinde firmaların sıralaması 2005 (Bin adet)..... | 93 |
| Tablo 4.1. İlan sayısının çokluğuna göre ilk 10 otomobil markası..... | 117 |
| Tablo 4.2. Veri kümesinin ülkelere göre dağılımı | 126 |
| Tablo 4.3. Güncel fiyatları tahmin etmede kullanılan örneklemin veri yapısı | 126 |
| Tablo 4.4. Regresyon model özeti | 128 |
| Tablo 4.5: Regresyon modeli ANOVA sonuçları..... | 128 |
| Tablo 4.6. Regresyon analizi katsayılar tablosu | 129 |
| Tablo 4.7. İlk amaca uygun yapay sinir ağı modeli..... | 133 |
| Tablo 4.8. Modellerin karşılaştırmasına yönelik doğruluk ölçüleri..... | 133 |
| Tablo 4.9. Gelecekteki fiyatları tahmin etmede kullanılan otomobil modelleri..... | 134 |
| Tablo 4.10: Veri kümeleri..... | 135 |
| Tablo 4.11. Modellerin karşılaştırmasına yönelik doğruluk ölçüleri..... | 137 |
| Tablo 4.12. İkinci amaca uygun yapay sinir ağı modeli..... | 139 |
| Tablo 4.13. Modellerin karşılaştırmasına yönelik doğruluk ölçüleri..... | 140 |
| Tablo 4.14. Veri kümelerinin 2008-2009 tahmin değerleri | 141 |

KISALTMALAR LİSTESİ

| | |
|----------|---|
| AAMA | American Automobile Manufacturers Association |
| AB | Avrupa Birliği |
| ABD | Amerika Birleşik Devletleri |
| AI | Artificial Intelligence |
| ART | Adaptif Rezonans Teorisi |
| CRISP-DM | Cross-Industry Standard Process for Data Mining |
| CRM | Customer Relationships Management |
| DBMS | Database Management Systems |
| ĐİE | Devlet İstatistik Enstitüsü |
| DPT | Devlet Plânlama Teşkilatı |
| ERP | Enterprise Resource Planning |
| GSYİH | Gayri Safi Yurtiçi Hâsılası |
| GB | Geri Beslemeli |
| GY | Geri Yayılımlı |
| GYA | Geri Yayılım Algoritması |
| İB | İleri Beslemeli |
| İGBS | İhracatçı Birlikleri Genel Sekreterliği |
| İMKB | İstanbul Menkul Kıymetler Borsası |
| KBA | Kraftfahrt-Bundesamt |
| KDS | Karar Destek Sistemleri |
| MAE | Mean Absolute Error |
| MAPE | Mean Absolute Percentage Error |
| MPE | Mean Percentage Error |
| MSE | Mean Squared Error |
| OHKK | Ortalama Hata Kareleri Kökü |
| OICA | Organisation Internationale des Constructeurs d'Automobiles |
| OLAP | Online Analytical Processing |
| OMH | Ortalama Mutlak Hata |
| OMYH | Ortalama Mutlak Yüzde Hata |
| OSD | Otomotiv Sanayii Derneği |
| RMSE | Root Mean Squared Error |
| SQL | Structured Query Language |
| TC | Türkiye Cumhuriyeti |
| vb. | Ve başkaları, ve benzerleri, ve bunun gibi |
| VTBK | Veritabanlarında Bilgi Keşfi |
| VTYS | Veritabanı Yönetim Sistemleri |
| YSA | Yapay Sinir Ağları |

ÖZET

Günümüzde ikinci el otomobiller dünyanın hemen her ülkesinde tüketicilerin en çok ilgi gösterdiği dayanıklı ürünler arasında yer almaktadır. Bu ürünlerin pazardaki güncel fiyatları ve gelecekteki fiyatları her zaman merak konusu olmuştur.

Bu çalışmanın temel amacı, ikinci el otomobillerin hem pazardaki güncel fiyatlarını hem de gelecekteki fiyatlarını tahmin etmeye yönelik uygun modeller oluşturmaktır. Bu amacı gerçekleştirmek için, öncelikle internet üzerinden elde edilen ikinci el otomobil ilanlarına ait veriler kapsamlı analizlere tabi tutularak veri madenciliği analizlerine uygun hale getirilmiştir. Bu işlemlerin sonunda sahip olunan 2005–2007 yılları arasını kapsayan 3 yıllık nitelikli veri kümesi üzerinde veri madenciliği analizleri yapılmıştır.

İkinci el otomobillerin güncel pazar fiyatlarını modelleyebilmek için, bu ürünlerin temel, opsiyonel ve zamanla değişen özellikleri ile fiyatları arasındaki ilişki regresyon analizi ve yapay sinir ağları kullanılarak belirlenmiştir. İki yöntemin tahmin doğrulukları karşılaştırıldığında, yapay sinir ağlarının daha başarılı sonuçlara sahip olduğu gözlenmiştir. Gelecekteki fiyatları modelleyebilmek için ise öncelikle pazarda en çok ilgi gören otomobil marka ve modelleri belirlenmiştir. Bu otomobillerin son üç yıllık fiyat verilerine çeşitli zaman serisi tahmin yöntemleri ile yapay sinir ağları uygulanmıştır. Ortaya çıkan tahmin modellerinin karşılaştırılması sonucu yapay sinir ağlarının yine daha başarılı sonuçlar verdiği gözlenmiştir. Böylece, ikinci el otomobillerin gerek güncel gerekse gelecekteki fiyatlarının tahmininde yapay sinir ağlarının başarıyla kullanılabileceği sonucuna ulaşılmıştır.

SUMMARY

FORECASTING THE PRICES IN THE SECOND-HAND CAR MARKET USING DATA MINING

Today, second-hand cars are among the most attractive durable goods in almost all countries of the world. Current and future prices of these products have always been wondered by the consumers.

The main goal of this study is to develop suitable models to forecast both the current and future prices of second-hand cars in the market. In order to carry out this goal, firstly, the data of second-hand cars gathered from the internet has been analyzed in a comprehensive manner and transformed into a suitable form to be used in data mining analysis. As a result of these operations, data mining analysis have been applied on the qualitative data set obtained for the period of 2005 through 2007.

In order to develop a model for the current market prices of second-hand cars, the relation between the prices and various attributes of those cars have been determined by using regression analysis and artificial neural networks. After comparing the forecast accuracy of these methods, it has been observed that artificial neural networks have been more successful. Later, in order to model the future prices of second-hand cars, firstly popular car models have been selected. Then various time series forecasting methods and artificial neural networks have been applied on the past 3-years-prices of those cars. After comparing the forecast accuracy of these methods, it has been observed that artificial neural networks have still been more successful than others. As a result, it has been concluded that artificial neural networks could successfully be used to forecast both the current and future prices of second-hand cars.

ÖNSÖZ

Çalışmamın her aşamasında değerli yardımlarıyla beni yönlendiren, engin hoşgörü sahibi saygıdeğer danışman hocam Prof.Dr. Ayşe KURUÜZÜM'e, tez izleme komitelerinde yer alarak katkılarını esirgemeyen değerli hocalarıma, uygulama kısmındaki görüş ve önerileriyle destek olan sayın Yrd. Doç. Dr. Ecir Küçüksille ile sevgili dostum Arş. Gör. Sezgin Irmak'a ve doktora öğrenimim boyunca yoğun çalışmalarına katlanmak zorunda kalan ailem ile yakın dostlarıma teşekkürlerimi sunarım.

GİRİŞ

Otomotiv endüstrisi, sanayileşmiş ülkelerin hemen hepsinde ekonominin lokomotif konumunda bulunmaktadır. Bu endüstrideki toplam motorlu araç üretiminin yaklaşık % 70'ini otomobillerin oluşturduğu gerçeğinden yola çıkarak, otomobil sektörünün dünya ekonomisinde tartışılmaz bir yere sahip olduğu söylenebilir.

Öte yandan, pek çok ülkede ikinci el otomobil satış hacmi, yeni otomobil satış hacminden çok daha büyük hale gelmiş olup ikinci el otomobil pazarı günümüzde hemen her ülkede giderek büyümekte ve ekonomideki önemli yerini korumaya devam etmektedir.

Dünya ekonomisinde ve insan hayatında böylesine önemli yeri olan bu sektördeki en önemli unsur ikinci el otomobillerin fiyatlarında ortaya çıkan farklılıklar ya da değişimlerdir. İkinci el otomobillerin fiyatları yeni otomobil fiyatlarına göre çok daha değişken olduğu için tahmin edilmesi zorlaşmakta; bu amaca yönelik olarak öncelikle uzun süreli ve nitelikli bir veri toplama işleminin yapılması gerekmektedir. Kapsamlı araştırmalara rağmen, ne yazık ki ikinci el otomobil fiyatlarının tahminine yönelik akademik bir çalışmaya rastlanamamıştır. Sektördeki bu eksiklik, bu çalışmanın konusunun belirlenmesinde önemli bir etken olmuştur.

Tahminleme amacıyla veri madenciliği sürecinin kullanıldığı bu çalışma dört bölümden oluşmaktadır.

Birinci bölümde, veri madenciliğinin temelini oluşturan veritabanı, veri ambarı ve bilgi keşfi konuları, teknik ayrıntılara girilmeden anlatılmaya çalışılmıştır.

İkinci bölüm, veri madenciliği ve tahminleme konularına ayrılmıştır. Bu amaçla öncelikle veri madenciliği ve tahminleme konuları kavramsal düzeyde okuyucuya aktarılmıştır. Ardından tahmin etme amaçlı kullanılan önemli veri madenciliği yöntemleri kendi başlıkları altında ayrıntılı olarak sunulmuştur. Bu yöntemler; Regresyon Analizi, Zaman Serileri Analizi ve Yapay Sinir Ağlarından oluşmaktadır.

Üçüncü bölümde, otomotiv endüstrisi ve ikinci el otomobil pazarı hakkında bilgi verilmiştir. Bu bölümde öncelikle hem otomotiv endüstrisindeki hem de ikinci el otomobil

pazarındaki sektöre özgü genel bilgiler sunulmuştur. Çalışmanın uygulama kısmının Almanya’ da gerçekleştirilmesi, verilerin ise Avrupa’ da yer alan 8 ülkenin ikinci el otomobil pazarına ait olması nedeniyle, ikinci el otomobil pazarının Almanya’ daki, Avrupa’ daki ve ülkemizdeki durumu hakkında bilgiler verilmiştir. Ayrıca veri toplama işleminin internet aracılığıyla gerçekleştirilmesinden dolayı, ikinci el otomobillerin e-ticaret ortamında satışı ile bu alandaki önemli web siteleri hakkında derlenen araştırma sonuçları sunulmuştur.

Uygulamanın anlatıldığı dördüncü ve son bölümde ise, CRISP-DM sürecine uygun olarak; işin ve verinin kavranması, modellerin kurulup değerlendirilmesi ve uygulamaya konulması aşamaları sırasıyla anlatılmıştır. Uygulamada; iki amaç belirlenmiş ve bu amaçları gerçekleştirmek için üç yöntem kullanılmıştır. İlk amaç, ikinci el otomobillerin çeşitli temel ve opsiyonel özellikleri ile fiyatları arasındaki örüntüleri saptadıktan sonra, özellikleri verilen herhangi bir aracın güncel pazar değerini tespit etmektir. Bu amaçla sırasıyla regresyon analizi ve yapay sinir ağları kullanılmıştır. İkinci amaç; ikinci el otomobillerin, geçmiş yıllardaki zamana bağlı fiyat değişimlerinin incelenmesinden hareketle, gelecek yıllardaki fiyatlarını aylar itibariyle tahmin etmek olarak belirlenmiş, bu amaca yönelik olarak da zaman serileri analiz yöntemleri ile yapay sinir ağları kullanılmıştır. Her iki uygulamanın sonunda belirlenen modellerden elde edilen sonuçlar karşılaştırılarak yöntemlerin tahmin doğrulukları ortaya konmuştur.

Birinci Bölüm

1. VERİTABANLARI VE VERİ AMBARLARI

1.1. Veri ve Bilgi Kavramları

“Veri” (data) ve “bilgi” (information, knowledge) kavramları, veritabanlarının ve veri ambarlarının temel unsurlarını oluşturmaktadır. Veri, günlük yaşam döngüsü içerisinde doğal olarak oluşmakta ya da çeşitli cihazların olağan aktiviteleri sonucu meydana gelmektedir. Verinin temel özelliği, henüz işlenmemiş olmasıdır. İşlenerek anlamlı hale getirilmiş veri ise bilgi olarak adlandırılmaktadır.

Veri, bilginin hammaddesi olarak da nitelenebilir (Kaçtıoğlu, 1999, s.1). Alpaydın (2000, s.1)’ a göre; “veri kendi başına değersizdir. Bilgi ise bir amaca yönelik işlenmiş, organize edilmiş veridir. Veriyi bilgiye çevirmeye veri analizi denir. Bir soruya yanıt vermek için veriden çıkardığımız sonuç da bilgi olarak tanımlanabilir. Veri sadece sayılar veya harflerden ibaret değildir; veri, sayı ve harfler ile onların anlamıdır. Veri hakkındaki bu veriye meta veri denir”.

Latince “Informatio” kökünden gelen bilgi kavramı; biçim verme ve bilgi-haber eylemlerini ifade etmektedir. Lucas (1989, s.28) bilgiyi kısaca, “bir durum ya da olay hakkında kesin değilliği, tereddütlerimizi azaltan soyut bir varlık” olarak tanımlamaktadır. Örneğin, yarın havanın bulutsuz ve güneşli olacağını bildiren bir hava tahmin raporu, yarın futbol maçı düzenlenip düzenlenmemesi konusundaki tereddütleri azaltacaktır.

Davis ve Olson (1985, s.6) tarafından öne sürülen bir başka tanıma göre: “Bilgi, algılayan için anlamlı bir hale dönüştürülen ve şu anki veya ilerisi için muhtemel kararlarda gerçek bir değerdir”. Bu tanım, verinin bilgi üretebilmesi için, bir şekilde işlenmesi gerektiği, yani bilginin ham veriden öte bir kavram olduğu gerçeğini vurgulamaktadır.

Yönetim açısından bilgi, “yöneticinin karar almasına yardımcı olan öğeler” dir (İnceler, 1996, s.196). İşletme açısından bilgi ise işletmedeki ve çevresindeki faaliyetleri gösteren,

saklanabilen, üzerinde işlem yapılabilen ve organizasyon içerisinde rapor halinde bildirilen sembollerdir (Kriebel ve Horn, 1971, s.17-18). O halde bilginin başka bir tanımı da şöyle olacaktır: “Bilgi, verilerin veri işleme süreci yardımıyla faydalı ve anlamlı bir şekilde sokulmuş sonuçlarıdır” (Yılmaz, 1988, s. 8).

Hacimsel olarak verilere göre çok daha az yer kaplayan bilgi, kullanım değeri açısından çok daha güçlüdür. Ayrıca İngilizce’de “knowledge” kelimesinin karşılığı olan bir bilgi türü daha vardır ki, Türkçe’de tam karşılığı olmadığı için yine genel olarak “bilgi” ismiyle nitelendirilmekte, bazı yazarlar tarafından ise “özbilgi” olarak adlandırılmaktadır. Özbilgi, bilginin daha öz ve anlamlı bir şekilde dönüştürülmesini ifade eder (Gürsakar vd., 2000, s.371). Bu çalışmada “knowledge” kavramını ifade etmek için “özbilgi” yerine alışıldığı şekliyle yine “bilgi” kelimesi kullanılacaktır.

Bilgi günümüzde üretim unsurları arasına girmiştir. Örgütsel hedeflere ulaşmak için etkin yönetsel eylemler gerekir ki bu da doğru kararlarla mümkündür. Elde bilgi olmadan iyi karar verebilmek neredeyse imkânsızdır (Ülgen, 1990, s.8).

Veritabanı kavramına geçmeden önce iki grup veri tipinin varlığından söz etmek yerinde olacaktır (Gürsakar vd., 2000, s.372):

- Yapılandırılmış veriler: Bu veriler tablolar üzerinde satır ve sütunlar halinde düzenlenmiş verilerdir. Bunlar kağıt üzerinde olabileceği gibi bilgisayarlarda istatistik paket programlarının satır ve sütunlardan oluşan matris şeklindeki yapılarına kaydedilmiş bir şekilde de olabilir. İstatistik, yapılandırılmış verilerle ilgilenmektedir. Yapılandırılmış verilerin sayısı çok fazla olduğunda, bu tür verilerle veri madenciliği dalı ilgilenir.
- Yapılandırılmamış veriler: Bu veriler mektup, doküman, kitap gibi kağıt üzerindeki veya e-mail, web sayfası gibi sanal ortamlardaki metinlerden; fotoğraf gibi durağan veya film gibi hareketli görüntülerden ve seslerden oluşur. Günümüzde verilerin % 90’ ı yapılandırılmamış biçimdedir.

1.2. Veritabanı

Bilgisayar teknolojisiyle birlikte ortaya çıkan “veritabanı”, büyük miktardaki verileri dijital ortamda güvenli ve yapısal bir biçimde çok hızlı bir şekilde saklamaya, sorgulamaya

ve kullanmaya hizmet eden bir yazılım sistemidir. Veritabanlarında verilerin yapısal özellikleri ve birbirleriyle ilişkileri de saklanmaktadır. Bugün veri yoğunluklu olarak kullanılan yazılımların büyük çoğunluğu, kullanmakta olduğu verilerini veritabanlarında saklamaktadır. Günümüzün en güçlü ticarî veritabanı üreticileri arasında Oracle, Informix, Sybase, IBM, Microsoft ve Sun Microsystems sayılabilir.

Veritabanı Yönetim Sistemleri (Database Management Systems) ya da kısaca VTYS (DBMS); veritabanlarına kolayca ulaşmayı, yeni veri girmeyi, sorgulamayı veya verilerle ilgili çeşitli işlemler yapmayı kolaylaştıran, kısaca veritabanının yönetimini sağlayan yazılımlardır. Veritabanlarında doğrudan işlem yapmak için SQL dilini bilmek gerekirken, günümüzde VTYS' ni kullanarak SQL bilme zorunluluğunda olmadan da pek çok işlemi grafik ortamda kolayca yapmak mümkün hale gelmiştir.

1.2.1. Çeşitleri

Veritabanları kısaca beş grupta incelenebilir (Özkan, 2008):

1. İlişkisel olmayan
2. İlişkisel
3. Nesne
4. Nesne-ilişkisel
5. Masaüstü

- **İlişkisel olmayan veritabanları:** Sadece veri veya bilginin depolanmasına hizmet eden bu ilk veritabanlarında tutulan veriler arasında yapısal ilişki bulunmamaktadır. Bu nedenle veriler sadece görüntülenebilmekte, fakat sorgulama yapılamamaktaydı. İlişkisel veritabanlarının ortaya çıkışı bu veritabanlarının kullanımına son vermiştir.
- **İlişkisel veritabanları:** İlişkisel veritabanları, günümüzde en yaygın kullanılan veritabanı tiplerindedir. Bu veritabanları; ERP (Enterprise Resource Planning, Kurumsal Kaynak Plânlaması) Sistemleri, Karar Destek Sistemleri (KDS), CRM (Customer Relationships Management, Müşteri İlişkileri Yönetimi) uygulamaları, veri ambarları, veri madenciliği ve OLAP (Online Analytical Processing,, Çevrimiçi Analitik İşleme) gibi uygulamalarda kullanılmaktadır. İlişkisel veritabanları, veri ya da bilgileri satır ve sütunlardan oluşan yapısal tablolarda

saklarlar. Veriye erişim ve sorgulama, küresel veritabanı sorgulama dili olan SQL (Structured Query Language, Yapısal Sorgulama Dili) ile gerçekleştirilir.

- **Nesne veritabanları:** Nesne veritabanlarının en önemli özelliği, ilişkisel veritabanlarının satır-sütun yapısından kurtularak grafik, resim, ses, video görüntüsü gibi karmaşık veri tiplerini rahatlıkla saklayabilmeleridir.
- **Nesne-ilişkisel veritabanları :** Sektördeki önemli ilişkisel veritabanı üreticileri, kendi veritabanlarının karmaşık veri tipleriyle baş edemediklerini fark edip nesne uzantıları ekleyerek eksikliklerini giderme yoluna gitmişlerdir. Bu şekilde nesne-ilişkisel veritabanları ortaya çıkmıştır.
- **Masaüstü Veritabanları:** Bireylerin, küçük çalışma gruplarının veya proje takımlarının ucuz olarak elde edip herhangi bir sunucu bilgisayara ihtiyaç duymadan kendi kişisel bilgisayarlarında çalıştırabilecekleri veritabanlarıdır. Önemli masaüstü veritabanları arasında Oracle'ın "Personal Edition" 1, Microsoft'un "Access" i, Sybase'in "SQL Anywhere" i veya Sun Microsystems' in "mySQL" i sayılabilir.

1.2.2. Yararları

Veritabanlarının yararları dört ana grupta toplanabilir (Uysal, 2000, s.7; Özkan, 2008, s.15):

1. **Veri tekrarını engellemesi:** Veriler, veritabanlarında yapısal olarak saklanmaktadır ve gerekli kriterler belirtildiğinde bunlara ulaşılması çok kolay ve hızlı gerçekleşmektedir. Veritabanlarının kullanılmadığı klasik dosyalama sistemlerinde bir verinin ihtiyaç duyulduğu her ortamda bir kopyasının bulunması gerekirken, veritabanlarında bu veri tek bir yerde tutulur. Başka ortamlardan ise ilgili referans numaraları ya da kodları belirtilerek bu veriye erişim sağlanmaktadır. Böylece, verinin ortak bir yerde tutulması sağlandığından tekrarı da önlenmiş olur. Bu ise, veriyi tek bir yerde güncelleme kolaylığı sunduğu gibi aynı zamanda önemli bir yer (depolama alanı) tasarrufu sağlar.
2. **Veri bütünlüğünü sağlaması:** Veri bütünlüğü (data integrity), verinin doğruluğunu ve tutarlılığını (çelişkisizliğini) ifade etmektedir. Veritabanlarına konulan bilgi giriş kısıtlamaları hatalı bilgi girişini önler, depolanan bilgilerin tutarlı olmasını sağlar. Örneğin, kişi bilgilerinin girildiği bir tablodaki "Doğum Yeri İl Trafik Kodu" alanına konulan bir veri aralığı kısıtı sayesinde (Türkiye için

1-81 aralığında olması gibi), kullanıcının yanlışlıkla 91 gibi anlamsız bir il kodu girmesi önlenmiş olunur. Başka bir örnek olarak, yine aynı tabloda “Mezun Olduğu Üniversite” alanı, mevcut üniversitelerin tutulduğu “Üniversiteler” tablosundaki ilgili alana bağlanarak, kullanıcının bilgi girişinde sadece bugüne kadar tanımlı olan bir üniversite adı seçebilmesine izin verilerek daha sonra çelişki doğuracak bilgi girişi önlenmiş olur. Şayet kullanıcının mezun olduğu üniversite, “Üniversiteler” tablosunda henüz yer almıyorsa, öncelikle o tabloya bu üniversite adının da eklenmesi sağlanmış olur. Veri bütünlüğü özelliği, belirli bir veride aynı anda farklı kullanıcılar tarafından güncelleme yapılmasını da engeller. Salise farkıyla da olsa ilk güncellemeyi başlatan kişiye öncelik verir, aynı zamanda diğer kişileri bu konuda bilgilendirerek mevcut güncelleme işlemi bitene kadar yeni bir güncelleme yapılmasına izin vermez. Güncelleme bittiği anda verinin son halini diğer kullanıcılara göstererek bu güncel veri üzerinde değişiklik yapmalarına olanak tanır. Bu da verinin eş zamanlı olarak güncellenmesinden doğacak tutarsızlıkları engellemiş olur.

3. **Verinin güvenli saklanması sağlanması:** Veritabanlarındaki verilere erişecek kişilerin yetkilendirilmesi sağlandığı için yetkisiz (gerekli parolaya sahip olmayan) kişilerin bu verilere erişmesi engellenmiş olur. Ayrıca istenirse verilerin içeriği de şifrelenebilir. Böylece, veritabanı yetkisiz kişiler tarafından usulsüz bir şekilde ele geçirilmiş olsa bile, içerdiği bilgilerin anlaşılması engellenmiş olur.
4. **Ortak bir sorgulama dili sunması:** Veritabanlarında işlem yapmak için geliştirilen SQL komutlarını öğrenen bir kullanıcı, bu küresel dili kullanarak her tür veritabanında istediği işlemleri gerçekleştirebilir (yetkisi dâhilinde).

1.2.3. Katmanları

VTYS, üç katmandan oluşmaktadır (Yüksek, 2001, s.10; Özkan, 2008, s.18):

1. **Fiziksel katman:** Veritabanlarının fiziksel olarak disk üzerinde depolanmasıyla ilgili özellikleri içerir.
2. **Kavramsal katman:** Veriler arasındaki mantıksal ilişkileri ve yapısı hakkındaki bilgileri içerir.
3. **Görsel katman:** Kullanıcıların veritabanı üzerinde kolaylıkla işlem yapmalarına olanak sağlayan bir katmandır. Günümüzdeki grafiksel arayüzlü, kullanıcı dostu yazılımlar bu amaca hizmet etmekte, kullanıcıların bilgisayar faresini kullanarak

sürükle-bırak yöntemiyle kolayca veriler arasında ilişkiler oluşturmalarını sağlamaktadır.

1.3. Veri Ambarı

Sürekli olarak birtakım işlemlerin gerçekleştirildiği işlemsel (transactional) veritabanlarında yoğun olarak bilgi girişi, silinmesi ve değiştirilmesi söz konusudur. Bir süpermarketteki anlık satış bilgilerinin girişi, satılan ürünlerin stoktan düşüşü, uçak bileti rezervasyonları işlemsel sistemlere örnek olarak gösterilebilir. Bu tip sistemlerde veritabanları çok yoğun olarak kullanılmaktadır ve bu veritabanları da oldukça büyük boyuttadırlar. İşlemsel veritabanları üzerinde karmaşık sorgular ve analitik işlemler çalıştırmak, veritabanı ve yazılımların performansını büyük ölçüde düşürdüğü için bu tip sistemler karmaşık analizlerin yapılmasına uygun değildir. Bunun yerine analiz işlemleri için “veri ambarları” (datawarehouses) (veya fonksiyonel veri ambarlarını ifade eden “veri depoları (data marts)”) kullanılmaktadır. Çoğunlukla durağan ve geçmişe ait özet ya da birleştirilmiş bilgi içeren veri ambarlarının temelinde de aslında veritabanları bulunmaktadır (Adriaans ve Zantinge, 1998, s.25-26).

Aslında birbirlerinin tamamlayıcısı olan veri ambarları ile veri madenciliği kavramları zaman zaman birbiriyle karıştırılmaktadır. Veri ambarları analize elverişli verinin belirli bir yapıda saklanması görevini üstlenirken, veri madenciliği bu verinin bilgiye dönüştürülmesini sağlamaktadır. Veri madenciliği yöntemleri, ihtiyaç duydukları bilgiyi veritabanları yerine veri ambarlarından almayı tercih etmektedirler. Veri ambarlarının veri madenciliğinin “omurgası” nı oluşturduğu söylenebilir (Harrold, 2000, s.9-10). Dolayısıyla, veri madenciliği uygulamalarına başlamadan önce veri ambarlarının oluşturulması büyük önem taşımaktadır.

Veri ambarı bilgiyi işlemekten ziyade, sorgulamak ve analiz etmek için tasarlanmış bir ilişkisel veritabanıdır. Genellikle bilgi işlemeden elde edilen tarihsel veriyi içermekle birlikte, başka kaynaklardan da veri alabilen veri ambarı, analiz yükünü işlem yükünden ayırarak, kuruluşların çeşitli kaynaklardaki verilerini birleştirmelerine imkân sağlar. Analiz amaçlı kullanılacak, değiştirilmeye ihtiyaç duyulmayan verilerin veri ambarına yerleştirilmesiyle, kullanıcılar performansa karşı duyarlı üretim sistemlerini etkilemeden, herhangi bir zaman kısıtlamasına bağlı kalmadan, veri ve yararlı bilgileri inceleyebilme ve yeni analitik bilgiler ve sonuçlar ortaya koyabilme imkânına sahip olurlar (Oracle 9i Data

Warehousing Guide, 2001, s. 1-2).

Veri ambarları, milyonlarca satırdan oluşan tabloların birleştirilmesini gerektiren sorgulamalara yanıt verecek şekilde tasarlanmıştır. İşletmelerde veriler genellikle değişik birimlerde parça parça birikmiş halde bulunur. Bu nedenle, acil ve önemli karar verme durumlarında tamamen bütünleşik bilgi birikimi kullanılmadığından dolayı, genelde olayı tüm yönleriyle ortaya koyamayan eksik bilgiden yararlanır. Kurumlardaki tüm analizcilerin ve karar vericilerin beklentilerine uygun, bütünleşik, güncel ve tarihi verileri depolayan veri ambarları sayesinde işletmenin iç ve dış kaynaklarından yıllık, aylık, haftalık, günlük, hatta saatlik olarak derlenen veriler bu veritabanlarına kopyalanır. Bu veriler işletmenin yönetim kademelerinde kolayca kullanılabilir şekilde standartlaştırılarak verilerde bütünlük sağlanmış olur (Şahin, 2001, s. 94, Özkan, 2008).

1.3.1. Unsurları

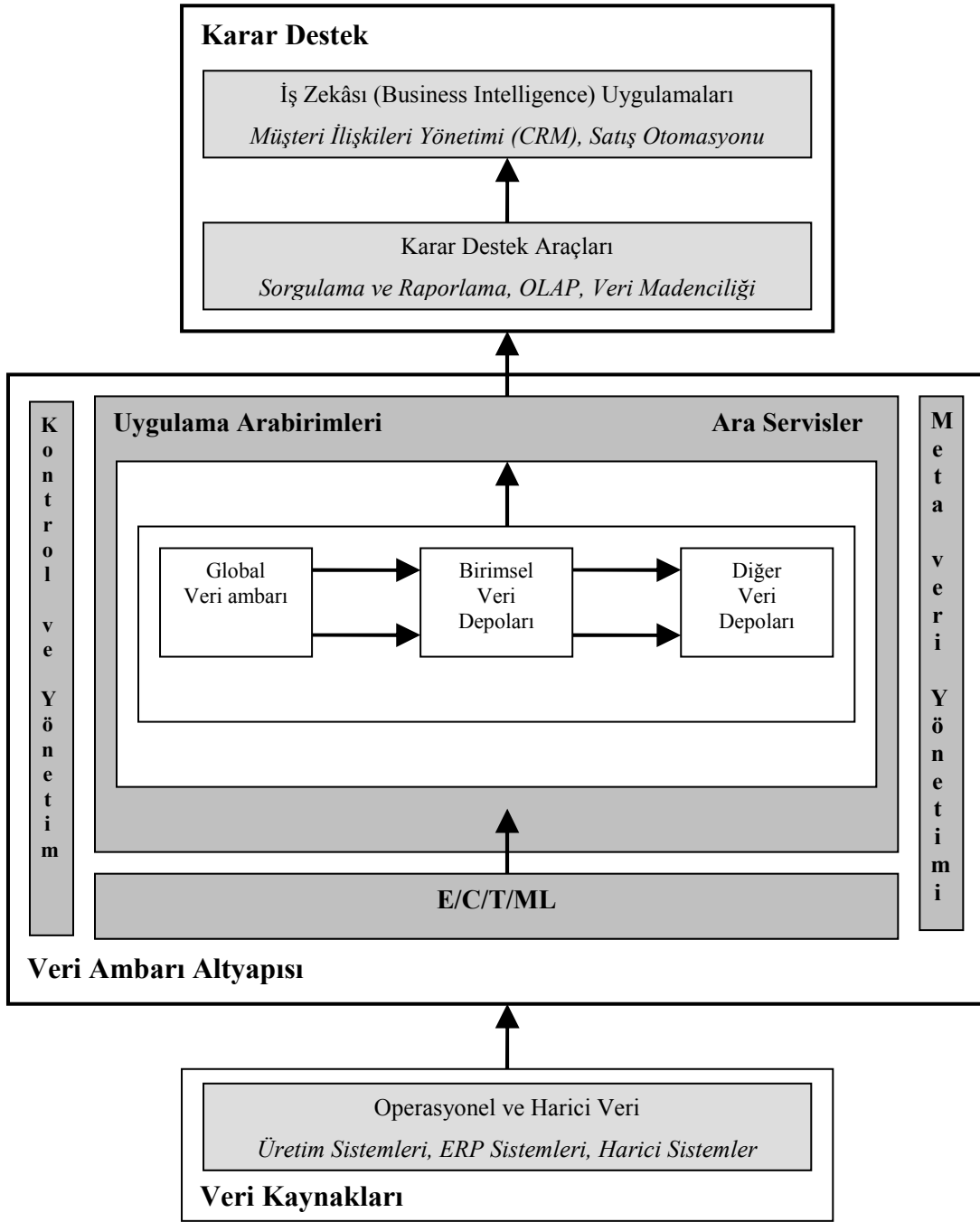
Tipik bir veri ambarının üç temel unsuru bulunmaktadır:

1. **Meta Veri:** Genel anlamıyla “veri hakkındaki veri” olarak da ifade edilebilen meta veri, verilerin biçim ve tanımlarının saklandığı kayıtlardır. Meta veri, son kullanıcının verileri bulup anlamasına yardımcı olan bilgiler olduğu için bir veri ambarı için son derece önemlidir. Meta verinin içermesi gereken özelliklerden başlıcaları; verinin yapısı, özetleme için kullanılan algoritma ve verinin operasyonel ortamdan veri ambarına aktarılma süreci ve şeklidir (Dilly, 1995).
2. **Operasyonel Veri:** Büyük miktarlarda ve anlık olarak değişebilme potansiyeline sahip verileri saklayan operasyonel sistemler, işletmenin geçmişteki bilgilerinden ziyade bugünkü güncel durumunu gösteren verilere ihtiyaç duyarlar. Operasyonel sistemlerin ihtiyaç duyduğu ve kendi veri tabanlarında sakladıkları bu bilgiler “operasyonel veri” olarak adlandırılır. Operasyonel veri, işletmenin günlük rutin işlemlerinin devamlılığını sağlamak için temel teşkil eder, fakat bu veriler başka analizler için uygun değildir. Veri ambarcılığının başlıca odak noktası, operasyonel veriyi işleyip entegre ederek analitik veri haline getirip veri madenciliği uygulamaları için uygun şekilde veri ambarında saklamaktır (Hair vd.,1998, s.676).
3. **Analitik Veri:** Analitik sistemler, bir olay ya da durum hakkında bilgi sağlayan, karşılaştırmalar yapan, ilişki ve trendleri analiz eden sistemlerdir. Bu sistemlerin ihtiyaç duyduğu ve genellikle veri ambarlarında sakladıkları analitik veriler,

operasyonel sistemlerde ortaya çıkan işlenmemiş verilerin belirli bir soruna çözüm sağlamak amacıyla bir araya getirilmiş durumudur. Bir müşterinin satın alma geçmişi veya bir ürünün satış durumu, analitik veri örnekleridir (Hair vd., 1998, s.676).

Veri ambarı ortamı, veritabanlarına ilaveten; elde etme, taşıma, dönüştürme ve yükleme (Extract, Transport, Transform, Load) çözümünü, OLAP motorunu, istemci analiz araçları ile veri toplamayı ve iş kullanıcılarına ulaştırma sürecini yöneten diğer uygulamaları da içerir (Ryan, 1999).

Şekil 1.1'de, veri ambarının işletme içerisindeki diğer sistemler arasındaki yeri görülmektedir.



Şekil 1.1. Veri ambarının işletme bilgi teknolojileri içindeki yeri
(Kaynak: Ryan, 1999)

1.3.2. Özellikleri

Tipik bir veri ambarı şu özellikleri taşımalıdır (Adriaans ve Zantinge, 1998, s.25-26; Oracle 9i Data Warehousing Guide, 2001, s. 1-2):

- Konu yönelimli
- Bütünleşik

- Durağan
- Zaman deęişimli
- **Konu yönelimli:** Her türlü veriyi kaydeden işlemsel veritabanlarından farklı olarak veri ambarları veriyi analiz ederek sadece belirlenmiş konulardaki verileri saklamaktadır. Örneğin, bir şirketin satış verisi hakkında detaylı bilgi elde etmek için, satışlar üzerine yoğunlaşan bir veri ambarı kurulabilir ve bu veri ambarı kullanılarak “geçen sene bu ürünü alan en iyi müşteri kimdi?” şeklindeki sorulara yanıt bulunabilir. Bir veri ambarını, bu olaydaki satışlar gibi, bir konu açısından tanımlama yeteneęi, veri ambarını “konu yönelimli” yapar.
- **Bütünleşik:** Bütünleşiklik, konu yönelimlilik ile yakından ilgilidir. İşletmelerin değerli bilgileri çoğunlukla birden fazla sistem içinde gömülüdür. Etkin bir bilgi yönetimi için, etrafa dağılmış olan bu verilerin kolayca erişilebilen tek bir kaynak olan veri ambarında toplanması gereklidir. Farklı kaynaklardaki verileri tek bir noktada toplayıp birleştirmeden sistem içinde saklı kalmış olan faydalı bilgiler tam olarak ortaya çıkarılamaz. Örneğin; satış, ürün ve müşteri verileri merkezi bir yerde toplanıp, temizlenip analiz edilmeden, iş yapma biçimlerinde deęişime yol açabilecek değerli müşteri bilgilerinin ortaya çıkarılması mümkün olamaz. Veri ambarları, isimlendirme çatışmalarını ve ölçüm birimleri arasındaki tutarsızlıkları da çözümleyerek, farklı kaynaklardaki verileri tutarlı bir şekilde barındırmalıdır. Bu özellikleri de onları bütünleşik yapar.
- **Durağan:** Veri ambarları, meydana gelen olayları kronolojik olarak analiz eder. Durağanlık özellięi, verinin veri ambarına girildikten sonra deęiştirilmesinin uygun olmadığını ifade eder. Veri ambarına alınan verilerde deęişiklik gerektiğinde bu deęişiklik mümkün olduğunca işlemsel veritabanlarında gerçekleştirilmeli, daha sonra da uygun bir zamanda veri ambarları bu deęişikliğe uygun olarak kaynak veritabanlarından güncellenmelidir. Veri ambarında dinamik verinin bulunması, bakım ve yönetim işlevlerini zorlaştırır.
- **Zaman deęişimli:** İşlemsel veritabanı sistemlerinde mümkün olduğunca asıl işe yarayacak güncel bilgiye ihtiyaç duyulur ve bu nedenle de performans artışı için tarihsel veriler, arşive gönderilerek sistemden kaldırılmaya çalışılır. Analistler ise iş dünyasındaki eğilimleri keşfedebilmek için çok büyük miktarda ve tarihsel verilere ihtiyaç duyarlar. Veri ambarları, verilerin kronolojik olarak deęişimine odaklandığından dolayı, zaman deęişimli olmalı, zaman aralığı da olabildiğince

uzun tutulmalıdır. Bir veri ambarındaki faydalı bilgi, yıllar boyunca erişilmek üzere saklanılır. Saklama süresinin beş yıldan fazla olması yaygın bir yaklaşımdır.

1.3.3. Kullanıcıları

Şirket içerisinde veri ambarından yararlanabilecek kullanıcı grupları şunlardır (Goldman, 1995):

- **İşlemsel kullanıcılar**, genelde günlük, hatta anlık verilecek kararlara cevap ararlar.
- **Karar Destek Sistemi kullanıcıları**, daha kapsamlı ve uzun vadeli soruların cevaplarını aramaktadırlar.
- **Temel iş kullanıcıları**, veri ambarına düzenli olarak erişip, sordukları soruların cevabını alırlar. Genellikle aynı tip sorgulamaları kullanırlar.
- **Analistler**, fazla miktarda ve çok detay düzeyde veriler üzerinde çalışırlar. Yaptıkları sorgulamalar genellikle anlıktır ve belli bir standardı yoktur.
- **Yöneticiler**, genelde fazla miktarda veri üzerinden elde edilmiş sonuçlarla ilgilenirler. Düzenli olarak veriye ulaşırlar. Temel iş kullanıcılarının ve analistlerin kendilerine sunduğu raporlar üzerinde çalışırlar.

1.3.4. Mimarisi

Veri ambarı mimarisi 8 temel katmandan oluşur (Cambazoğlu, 2002):

1. **İşlemsel veritabanı/dış veritabanı katmanı:** Kritik işlemsel gereksinimleri desteklemek üzere verilerin işlendiği katmandır. Bunu sağlayabilmek için işlemsel veritabanları oluşturulur. Bu seviye sadece, değerli bilginin elde edilebilmesi için gerekli olan ham materyali sağlar.
2. **Bilgi erişim katmanı:** Son kullanıcının doğrudan olarak kullandığı bu katman, özel olarak, kullanıcının günlük olarak kullandığı ofis uygulamalarını da içerir. Bu katman aynı zamanda, analiz ve sunum amaçlı raporları, hesap tablolarını, grafikleri ve çizelgeleri görüntülemek ve yazıcıdan çıktısını almak için gerekli olan donanım ve yazılımları da içerir. Günümüzde ön plâna çıkan eğilim ise, kullanıcıların bilgiye standart bir arayüz haline gelen Web tarayıcı yardımıyla erişebilmelerini sağlamaktır.

3. **Veri erişim katmanı:** SQL'in değişik üreticiler tarafından kullanılan değişik versiyonları arasındaki farklılıklar, "aracı yazılımlar (middleware)" kullanılarak yumuşatılabilir. Veri erişim katmanı, bilgi erişim katmanının işlemsel katman ile konuşmasını olanaklı hale getiren katmandır.
4. **Veri izin katmanı:** Bu katman, veri dizini veya organizasyon içindeki veri hakkındaki meta veri bilgisini depolar.
5. **Süreç yönetim katmanı:** Veri ambarı ve veri dizini bilgilerini oluşturmak ve devam ettirmek için yapılması gereken çeşitli görevlerin zaman plânını gerçekleştiren bu katman, veri ambarını güncel tutabilmek için oluşması gereken birçok süreçte zaman plânlayıcı veya yüksek seviyeli görev denetleyici olarak düşünülebilir.
6. **Uygulama mesajlaşma katmanı:** Organizasyonun bilgi işlem ağı boyunca iş hareketlerinin, mesajların ya da bilginin toplanıp, belirli bir sürede belirli bir yere gönderilmesi için kullanılan bu katman, uygulamaları her iki uçtaki gerçek veri formatından yalıtımda kullanılabilir.
7. **Veri ambarı (fiziksel) katmanı:** Bu katmanda temel olarak bilişim amaçlı kullanılan gerçek veri oluşturulur. Bunlar, veritabanı yönetim sistemleridir. Bazı hallerde veri ambarı, basitçe, verinin mantıksal veya sanal görünümü olarak ortaya çıkabilir, verinin saklanması içermeyebilir.
8. **Verinin yürütülüp idare edildiği katman:** "Kopya yönetimi" veya "replikasyon (eşleştirme) yönetimi" olarak da adlandırılan bu katman; veri ambarı seçimi, düzeltilmesi, özetlenmesi, birleştirilmesi, yüklenmesi ve işlemsel/dış veritabanlarından bilgiye erişim süreçlerini içerir. Bu katmanda, birçok kaynaktan gelen verilerin veri ambarından erişimini olanaklı kılmak üzere, önemli oranda dönüştürme ve çıkarma işlemleri kullanılır.

1.3.5. Süreçleri

Veri ambarı, altta kısaca bahsedilen beş ayrı süreçten oluşur (Cambazoğlu, 2002):

1. **Veri elde etme:** ERP gibi veri kaynaklarından verilerin elde edilme sürecidir. Dosya aktarımı, veri replikasyonu ve veri yükleme gibi işlemleri içerir.
2. **Veri dönüştürme:** Bu süreç, verinin, veri ambarına girerken standardizasyonu, temizlenmesi, aynı kayıtların belirlenerek tek bir kayıt haline getirilmesi veya farklı kayıtların tek bir yerde toplanılması gibi işlemleri içerir.

3. **Veri düzenleme ve yönetimi:** Etkin bir veri ambarı, meta verinin yoğun bir biçimde kullanım ve yönetimini gerektirir. Bu sayede kullanıcılar aradıkları veri veya faydalı bilgileri rahatlıkla bulabilirler. Meta veriler, depolar (repository) aracılığıyla yönetilir. Birçok üretici kendi depolarını sunmaktadır.
4. **Veri ambarı yönetimi:** Kullanıcıların yönetimi, güvenliğin sağlanması ve hassas performans ayarlarına yönelik araçları içeren süreçtir.
5. **Veri erişimi ve sağlama:** Bu süreç, son kullanıcılara veri aktarımı sağlayan SQL gibi araçları içerir. Şirketin Bilgi İşlem Merkezi birimi, kullanıcıları veri ve veritabanının ayrıntılarından yalıtılmak üzere bir “meta katman” oluşturmalıdır.

1.3.6. Yararları

Veri ambarı yapısının en önemli sonucu, son kullanıcıların ve karar vericilerin doğru kararlar vermesine yardım ederek, kurumlara rekabet avantajı sağlamasıdır. Sistemli bir veri ambarı projesinin genel anlamda işletmeye şu yararları getirdiği söylenebilir (Değirmendereli, 2000):

- Geçmişteki bilgilere sistematik erişim sağlanması,
- Farklı uygulamaların verilerindeki tutarsızlıkların giderilerek tüm bilgilerin birlikte kullanılabilmesine imkân vermesi,
- Doğru, kesin, tutarlı ve güncel bilgileri hızlı bir şekilde sunması,
- Anlık sorulara hızlı cevaplar vermesi,
- Dinamik ve kolay hazırlanan raporlar sunması,
- Raporların kurum içinde etkin dağıtım ve paylaşımını sağlamasının yanında ek bir yatırım yapmadan Web ortamında görülebilmeye imkân vermesi,
- Veri altyapılarının ve süreçlerinin gözden geçirilmesiyle, mevcut bazı aksaklıkların ortaya çıkarılması ve düzeltilmesine fırsat vermesi,
- Etkin kararların alınmasına yardımcı olması.

Veri ambarları, kriz dönemlerinde de büyük avantajlar sağlar. Zira yaşanan ekonomik krizlerin sebepleri, oluşum süreçleri, etkileri gibi faktörlerin verilerinin depolanmasını, raporlanmasını, müşterilerin bilgilendirilip yönlendirilmesini ve doğru stratejik kararlar alınmasını sağlayarak krizden avantajlı çıkmanın yollarını gösterir.

Veri ambarları sayesinde müşterilerinin gelir düzeyleri, tüketim alışkanlıkları,

tercihleri, eğitim durumları gibi önemli verileri analiz edebilen şirketler, bu veriler ışığında müşterilerine çapraz satış yapabilmek, tüketim alışkanlıklarını analiz edip onları istedikleri kârlı ürünlere ya da hizmetlere yönlendirebilmek gibi fırsatlara da sahip olabilmektedirler.

1.3.7. Dezavantajları

Veri ambarları kullanmanın birtakım dezavantajları da olabilir, bunlardan bazıları aşağıda sıralanmıştır (Web-Wikipedia2, 2008 ve Web-Exforsys, 2008).

- Karmaşık bir yapıya sahip olduğu için, uygulamada karşılaşılan sorunlar iş süreçlerinde de aksamalara neden olabilir.
- İşletimi ve bakımı yüksek maliyet gerektirmektedir.
- Veritabanları gibi sürekli güncellenmediği için bir müddet sonra tamamen güncelliğini yitirebilir. Bu haliyle yapılan analizler yanlış sonuçlara ve hatalı kararlara yönlendirebilir.

Veri ambarlarının bu dezavantajlarından dolayı, yeterli maddi ve insan kaynağına sahip olmayan, veri madenciliği çıkarımlarına çok ihtiyaç duymayan küçük ve orta ölçekli firmaların veri ambarı yerine sadece veritabanları kullanmaları önerilebilir.

1.3.8. Veri Ambarı Projelendirmesi

Büyük miktarda verileri bulunan işletmelerde veri ambarlarının oluşturulması için kesin hedeflerin önceden belirlenmesi, etkin bir proje plânlaması yapılması, özverili bir proje ekibinin oluşturulması ve çok etkin bir yöntemin kullanılması gerekir.

Bir veri ambarı projesinde yönetim kademesinin kesinlikle işin içinde olması ve ne istediğini iyi bilmesi gerekir. Yönetimin sisteme inanmış olması, projeyi başarıya götüren en önemli etkenlerdendir (Kirişçi, 1999).

İşlemsel sistemlerde mevcut durumla ilgili veri bulundurulurken, veri ambarında statik olan ve işletmeyle ilgili her çeşit konuda veriyi bulunduran tarihsel veriler bulundurulmalıdır. Başarılı veri ambarı projesinin ekibinde işlemsel sistem ekibinden farklı olarak işletme fonksiyonlarını bilen analistler ve son kullanıcılar en önemli kişiler olarak görülmelidirler (Nakipoğlu, 1999).

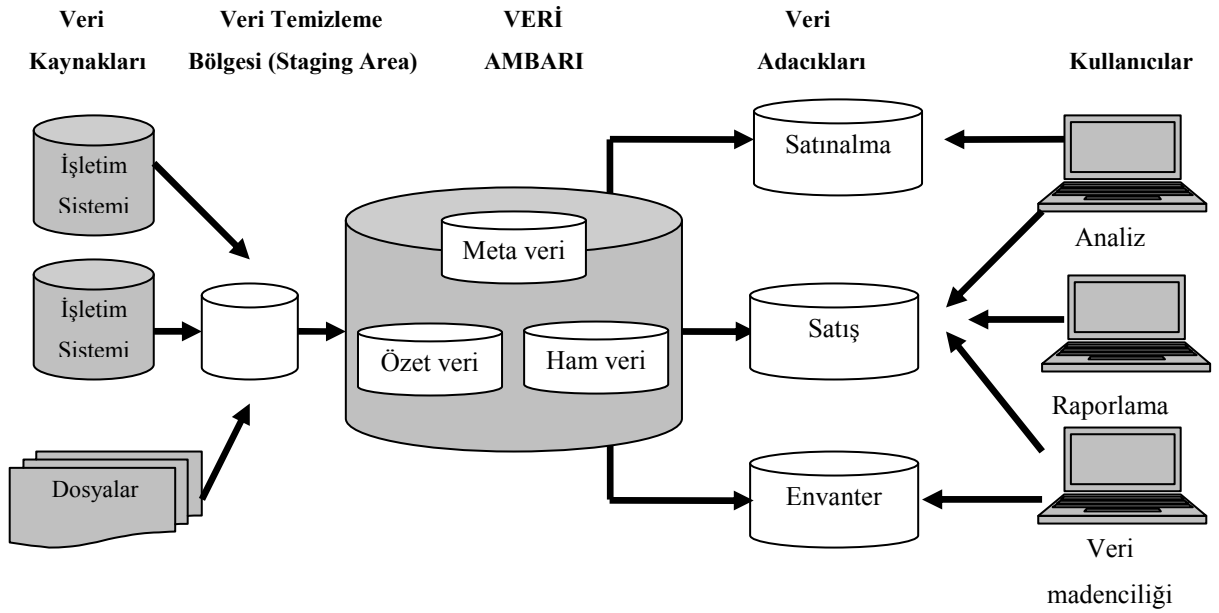
Veri ambarı sürecini bir paket programdan ayıran en önemli etken, veri ambarı uygulamalarının her kuruma göre farklılık göstermesi ve uygulama alanlarının farklı olmasıdır. Veri ambarı projelerinde öncelikler belirlenmeli, aşama aşama bir kurum içindeki karar destek altyapısı oluşturulmalı ve başarısı kanıtlanmış metodolojilerle desteklenmelidir (Kuyucu, 1999). Bir veri ambarı projesine başlamadan önce şirketin kendine özgü iş ve teknik hedefleri çok iyi belirlenmelidir.

Veri ambarı projelerindeki aşamalara üç ana başlık altında bakılabilir: Veri ambarı sürecinin başlangıcı olan “Plânlama” aşaması ile öncelikle firmanın hedefleri ve ihtiyaçları belirlenmelidir. Plânlama sonucunda ortaya çıkan şema “Tasarım ve Uygulama” aşamasında resmen uygulanmaya başlayacaktır. “Destek ve Geliştirme” aşaması olarak adlandırılan, kullanıcıların veri ambarını kullanmaya ve veri ambarından faydalanmaya başlamaları süreci başarılı bir projenin en önemli parçasıdır (NCR, 1999, s.3-12).

1.3.9. Veri Depoları

Bazı durumlarda, veri ambarı mimarisi işletmenin değişik gruplarına özgü olarak özelleştirilmek istenebilir; bunu gerçekleştirmek içinse, özel bir iş kolu için tasarlanmış sistemler olan “Veri depoları (data marts)” kullanılır. Kısa tanımla veri depoları, özel işlevli veri ambarlarıdır. Organizasyon içindeki farklı birimler ya da iş kolları, kendilerine özgü veri depolarına sahip olabilirler ya da farklı coğrafi bölgeler için farklı veri depoları tanımlanabilir. Veri depoları bazı durumlarda kuruluşun veri ambarının uyduları olabilir. Diğer bazı durumlarda ise, ortada bir kurumsal veri ambarı olmayabilir. Çok sayıdaki veri deposunun varlığı, gerçek bir veri ambarı yerine mantıksal bir veri ambarına yol açabilir (Ryan, 1999).

Şekil 1.2’ de, veri depolarını da içeren yeni veri ambarı mimarisi görülmektedir. Şekildeki Veri Temizleme Bölgesi (Staging Area), verilerin ayıklandığı, hatalı olanların düzeltildiği ve gerekli hesaplamaların yapıldığı bölgeyi ifade etmektedir.



Şekil 1.2. Veri depolarını içeren veri ambarı mimarisi
(Kaynak: Ryan, 1999)

1.4. Veritabanları ile Veri Ambarları Arasındaki Farklar

Veritabanlarında genelde belirli bir süre sonunda güncelliğini kaybedecek olan anlık veriler tutulmaktadır. Güncelliğini kaybedince anlamı kalmayacak olan veriler, ya sürekli yenileri ile güncellenirler ya da veritabanından silinirler. Zira veritabanında kayıt sayısının çok fazla olması veritabanını performansını düşürmekte, dolayısıyla veriye erişim çok yavaşlamaktadır. Diğer yandan, veri ambarı içerisindeki veriler genelde yığılarak birikirler ve verilerin geçerliliği çok daha uzun sürmektedir. Veri ambarı içerisinde ne kadar çok kayıt olursa, yapılan incelemelerin sonucu da o kadar doğru olacaktır. Veritabanları ile veri ambarlarının arasındaki bazı önemli farklar aşağıdaki gibi sıralanabilir (Özdemir, 2004, s.14 ve Han ve Kamber, 2001):

| Veritabanları | Veri ambarları |
|---|---|
| Çok küçük boyutlarda veri içerebileceği gibi, GB (GigaByte) boyutlarında büyük veri içerebilir. | Genelde TB (TeraByte) boyutunda çok büyük veri içerir. |
| Güncel veriler tutulur. | Tarihsel arşiv verileri tutulur. |
| Veriler sürekli güncellenir. | Veriler durağandır ve çoğunlukla okuma işlemi gerçekleştirilir. |
| Ham verileri içerir. | Birleştirilmiş ya da özet veriler içerir. |
| Çok geniş bir kullanıcı kitlesine açık olabilir. | Sadece analiz yapanlara ya da karar vericilere açıktır. |
| Aynı anda binlerce kişinin erişimine izin verilebilir. | Eş zamanlı kullanıcı sayısı kısıtlıdır ve genelde yüzü geçmez. |

| | |
|---|--|
| Veriler üzerinde küçük işlemler yapılır. | Genelde veriler üzerinde karmaşık sorgu ve analizler yapılır. |
| İşlemler kısa olduğu için hızlıdır. | İşlemlerin sonuçları günlerce sürebilir. |
| İndeks veya karışık veri erişim işlemleri yapılır. | Çok miktarda veri tarama yöntemi ile işlem yapılır. |
| Veri girişi amaçlı tasarlanmış ve bu yönde optimize edilmiştir. | Bilgi çıkışı amaçlı tasarlanmıştır. Veriler analize uygun şekilde optimize edilmiştir. |

1.5. Bilgi Keşfi Kavramı

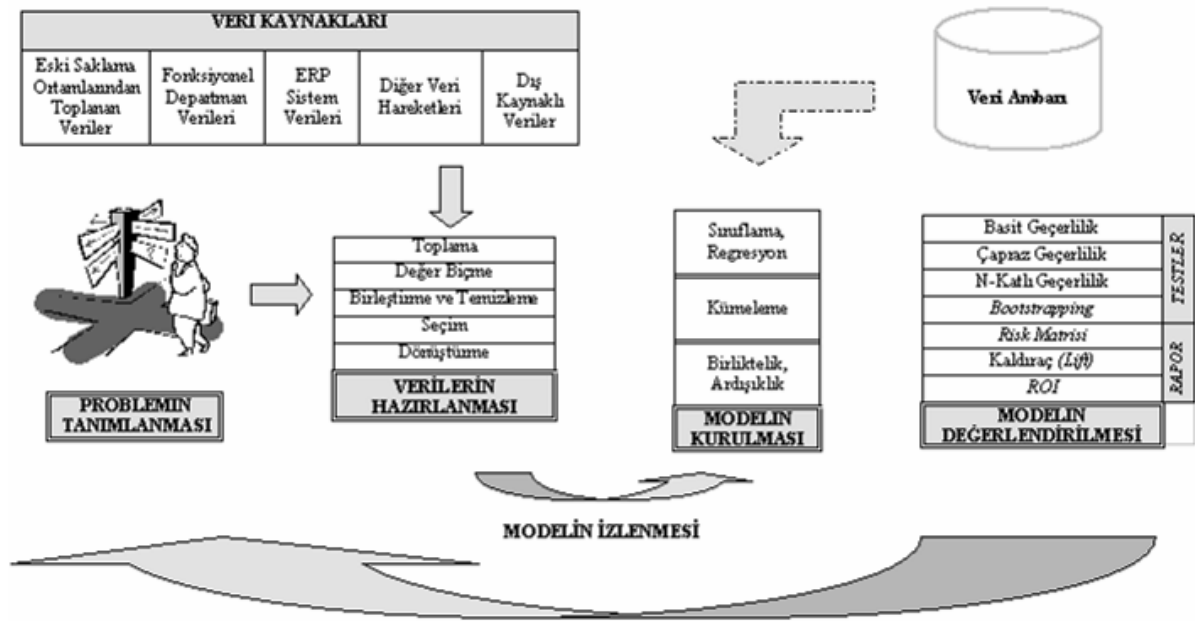
Dünyadaki veri miktarı her yıl yaklaşık iki katına çıkmasına karşılık anlamlı bilgi oranı hızla azalmaktadır (Adriaans vd., 1998, s.2). Büyük verilerden anlamlı verilerin elde edilmesi ihtiyacı “bilgi keşfi” kavramını ortaya çıkarmıştır. Klasik sorgulama araçları (yazılımları) bu çok büyük hacimli verileri işlemekte yetersiz kalınca bu amaca hizmet edecek yeni süreç ve/veya yöntemler kullanılmaya başlanmıştır. Yazında bu süreç ve yöntemler “Veri Tabanlarında Bilgi Keşfi” (kısaca VTBK) ve veri madenciliği kavramları olarak adlandırılmaktadır. VTBK kavramı, büyük miktarlardaki verilerin yeni kuşak araç ve yöntemlerle işlenip anlamlı bilgi çıkartılmasını ifade etmektedir. “Yöntemler”den kastedilen özellikle veri madenciliği teknikleri, “araçlar”dan kastedilen ise bu tekniklerin bilgisayar ortamında etkin bir biçimde kullanılmasını sağlayan yazılımlardır. Bazı kaynaklar VTBK’ni daha geniş bir disiplin olarak ele alıp veri madenciliği terimini VTBK sürecinde yer alan bir adım olarak görmektedir (Fayyad vd, 1996, s.10-11, Adriaans ve Zantinge., 1998, s.37, Berson vd., 1999, s.203; Olaru ve Wehenkel, 1999, s.20-21; Cios vd., 1998; Goebel ve Gruenwald, 1999).

Örneğin Adriaans ve Zantinge (1998) VTBK sürecini aşağıda görülen altı adımda tanımlamakta ve bunların içerisindeki beşinci adımı veri madenciliği olarak görmektedirler.

1. Veri seçimi
2. Veri ayıklaması
3. Veri zenginleştirilmesi
4. Veri kodlaması
5. Veri madenciliği
6. Raporlama

Diğer yazarlar da benzer şekilde VTBK sürecini veri madenciliğini de kapsayan 5-7

adım olarak tanımlamışlardır. Diğer yandan Akpınar (2000) gibi bir çok araştırmacı da ya VTBK ile veri madenciliği terimlerini eş anlamlı olarak kullanmakta ya da VTBK kavramı yerine doğrudan veri madenciliği kavramını kullanmayı tercih etmektedir. Örneğin Akpınar' a göre VTBK süreci, diğer pek çok yazarın tanımladığı veri madenciliği süreçleri ile eşdeğer görülmektedir. Bu süreç Şekil 1.3' de görülmektedir.



Şekil 1.3. Veritabanlarında bilgi keşfi süreci
(Kaynak: Akpınar, 2000)

Bu çalışmada da yazındaki çoğunluğun yaklaşımı doğrultusunda hareket edilerek, VTBK ayrı bir süreç olarak ele alınmamış, veritabanlarından (ya da veri ambarlarından) bilgi keşfi sürecinin tamamının “veri madenciliği” olarak nitelendirilmesi uygun görülmüştür.

İkinci Bölüm

2. VERİ MADENCİLİĞİ VE TAHMİN YÖNTEMLERİ

2.1. Veri Madenciliği

Bilgisayarların ortaya çıkışından sonraki en önemli görevleri, verilerin güvenle saklanmasına ve istenildiğinde hızlıca erişilebilmesine hizmet etmek olmuştur. Bu amaçla, 1960'larda ilkel dosya işlem sistemleri, 1970'lerde ilişkisel veritabanı sistemleri, 1980'lerde uzay ve multimedya verileri gibi büyük boyutlu verileri tutmaya olanak sağlayan nesne-tabanlı ilişkisel veritabanları geliştirilmiştir (Han ve Kamber, 2001).

2000 yılında sadece uydu ve diğer uzay araçlarından elde edilen görüntülerin boyutunun saatte 50 Gigabyte civarında olması, bu inanılmaz veri artışının boyutlarını açık bir şekilde ortaya koymaktadır. 1995 yılında düzenlenen "Knowledge Discovery in Databases" konferansının bildiri kitabı sunusunda, enformasyon teknolojilerinin oluşturduğu veri dağları aşağıdaki cümleler ile vurgulanmaktadır (Akpınar, 2000).

"Dünyadaki enformasyon miktarının her 20 ayda bir ikiye katlandığı tahmin edilmektedir. Bu ham veri seli ile ne yapabiliriz? İnsan gözleri bunun ancak çok küçük bir kısmını görebilecektir. Bilgisayarlar bilgelik pınarı olmayı vaat etmekte, ancak veri sellerine neden olmaktadır."

İnternetin ortaya çıkışı ile veri boyutlarında olağanüstü bir artış başlamış, veritabanı sistemlerinin hacimleri de bir müddet sonra inanılmaz şekilde büyümüştür. Geleneksel sorgulama ve raporlama araçları her ne kadar veritabanlarından verilerin ve veriler arasındaki sıradan ilişkilerin kullanıcılara sunulmasına imkân sağlasa da, veriler arasındaki gizli ilişkilerin ortaya çıkarılmasında yetersiz kalmıştır. Bu yetersizlik ise veri madenciliği gibi kavramların doğmasına neden olmuştur. Bu gelişmelere bağlı olarak 1990'larda Veri Ambarı ve Veritabanlarında Bilgi Keşfi (VTBK) (Knowledge Discovery in Databases) kavramları oluşturulmuştur. Çeşitli aşamalardan oluşan VTBK süreci içerisindeki en önemli aşamayı Veri Madenciliği (Data Mining) oluşturmaktadır. Bu önem, birçok araştırmacı tarafından VTBK ile veri madenciliği kavramlarının eş anlamlı olarak da kullanılmasına neden olmaktadır (Akpınar, 2000). Daha önce de belirtildiği gibi, bu

çalışmada da veri madenciliği kavramı sürecin tamamını ifade etmek için kullanılacaktır.

2.1.1. Tanımı

Veri madenciliği kavramına ait çok sayıda tanım yer almaktadır. Bu tanımlar içerisinde önemli görülenleri aşağıda verilmiştir.

En basit tanımla veri madenciliği, büyük ölçekli veriler arasından bilgiye ulaşma, yani bilgiyi madenleme işidir. Başka bir deyimle, büyük veri yığınları içerisinde gelecek ile ilgili tahminde bulunulmasını sağlayabilecek bağıntıların bilgisayar programı kullanılarak aranmasıdır. Bu kavram ile eş değer olarak şu kavramlar da kullanılmaktadır: “bilgi çıkarımı”, “veri ve örüntü analizi”, “veri arkeolojisi” (Web-Wikipedia3).

Veri madenciliği; kümeleme, veri özetleme, değişikliklerin analizi, sapmaların tespiti gibi çeşitli teknik yaklaşımları kullanarak mevcut verilerden üstü kapalı, çok net olmayan, önceden bilinmeyen ancak potansiyel olarak kullanışlı bilginin çıkarılmasıdır (Frawley vd., 1991).

Veri madenciliği, önceden bilinmeyen, geçerli ve uygulanabilir bilgilerin geniş veritabanlarından elde edilmesi ve bu bilgilerin işletme kararları verilirken kullanılmasıdır (Cabena, 1998, s.12).

Veri madenciliği konusunda önde gelen uzmanlardan birisi olan Piatetsky-Shapiro, bu kavramı “verilerden daha önceden bilinmeyen, gizli ve muhtemelen faydalı bilginin monoton olmayan bir süreçte çıkartılması işlemi” olarak tanımlamaktadır. Bu süreç; kümeleme, veri özetleme, sınıflama kurallarının öğrenimi, bağımlılık ağlarının bulunması, değişikliklerin analizi ve anomali tespiti gibi farklı birçok teknik yaklaşımı kapsamaktadır (Akpınar, 2000; Dilly, 1999).

Larose (2005), veri madenciliğini veriler arasındaki anlamlı ve yararlı ilişkileri tespit ederek veriyi yeni bir şekilde özetlemek şeklinde ifade etmektedir.

Veri kümesi içerisinde keşfedilmemiş örüntüleri bulmayı hedefleyen teknikler koleksiyonunu betimleyen veri madenciliği, geçmiş faaliyetlerin analizini temel alarak gelecekteki davranışların tahminine yönelik karar-verme modelleri oluşturmayı

amaçlamaktadır (Koyuncugil, 2007).

Gartner Group tarafından yapılan tanıma göre ise veri madenciliği, istatistik ve matematik tekniklerle birlikte örüntü tanıma teknolojilerini kullanarak, depolanmış veri yığınlarının elenmesi ile anlamlı yeni korelasyon, örüntü ve eğilimlerin keşfedilmesi sürecidir (Akpınar, 2000, s.1-22).

Bu tanımların ışığı altında veri madenciliği; çok büyük hacimli veriler içinde var olan gizli ya da fark edilmemiş anlamlı ilişkilerin, bilgi teknolojileri yardımıyla bilgiye dönüştürülerek ortaya çıkartılması işlemi olarak ifade edilebilir. Veri madenciliğinde bilgi teknolojilerinin kullanımı şart olmakla birlikte en önemli faktörün insan olduğu unutulmamalıdır.

2.1.2. Gereksinimleri

Veri madenciliğinin temel gereksinimleri şunlardır:

- Bu konuda uzman kişi(ler),
- Yeterli büyüklükte veri,
- Veritabanı ve/veya veri ambarı,
- Yüksek performanslı bilgisayarlar,
- Veri madenciliği yazılımları,
- Etkin algoritmalar.

2.1.3. Amaçları

Veri madenciliğinin dört temel amacı bulunmaktadır (Fayyad vd., 1996):

1. **Sınıflandırma:** Verinin daha önce tanımlanmış belirli kategoriler içerisine yerleştirilmesi.
2. **Tahminleme:** Bilinen özelliklerden yola çıkılarak, bilinmeyenlerin ortaya çıkarılması.
3. **Kümeleme:** Benzerliklere dayalı olarak nesnelerin doğal gruplarının araştırılması.
4. **Örüntü tanıma:** İşlemsel veri içerisinde yer alan yararlı örüntülerin ortaya çıkarılması.

2.1.4. Karşılaşılan Problemler

Veri madenciliğinin temel girdisi veritabanlarında tutulan verilerdir. Bu veriler çeşitli yöntemlerle işlenip veri madenciliği işlemlerine uygun hale getirilerek veri ambarlarına aktarılmaktadır. Bu uygunlaştırma, pek çok problemin çözülmesini gerektirmektedir. Bu problemler ve çözüm yolları aşağıda anlatılmıştır (Quinlan, 1986, Oğuz, 2000, Almuallim ve Dietterich, 1991; Deogun vd., 1995; Kira ve Rendeli, 1992; Pawlak, 1986).

Büyük boyutlu veritabanı: Veritabanlarının hacmi iki boyutlu olarak genişlemektedir:

Yatay Boyut: nesnelerin özellik sayılarındaki artış,

Dikey Boyut: nesnelerin kayıt sayısındaki artış.

Örneklemin büyük olması, örüntülerin gerçekten var olduğunu göstermesi açısından bir avantaj olmakla birlikte elde edilen örüntü sayısının çok artması işlem sürelerinin uzaması dezavantajını da beraberinde getirmektedir. Bu yüzden yüksek veritabanı boyutu veri madenciliği sistemleri için önemli problemlerden biridir. Bu sorunu çözmek için uygulanabilecek yöntemlerden bazıları şunlardır: örneklemin yatay ve dikey boyutta indirgenmesi ve veri madenciliği sisteminin sezgisel/buluşsal bir yaklaşımla arama uzayını taraması (Oğuz, 2000).

Gürültülü veri: Veri girişi ya da toplanması esnasında oluşan sistem dışı hatalara gürültü adı verilmektedir. Quinlan (1986) etiketli öğrenmede etiket üzerindeki gürültünün öğrenme algoritmasının performansını doğrudan etkileyerek düşürdüğünü tespit etmiştir. Şayet veri kümesinde gürültü varsa, sistem bu gürültüden etkilenmiş olan bozuk verileri tanımalı ve ihmal etmelidir.

Artık veri: Örneklem kümesindeki mevcut probleme uygun olmayan gereksiz nitelikler artık veri olarak adlandırılmaktadır. Artık verileri elemek için geliştirilmiş algoritmalar “özellik seçimi” olarak adlandırılır. Bu algoritmalar, hedef bağlamı tanımlamak için yeterli ve gerekli olan niteliklerin küçük bir alt kümesini seçmeye çalışır. Özellik seçimi yalnızca arama uzayını küçültmekle kalmayıp, sınıflama işleminin kalitesini de artırır (Kira ve Rendeli, 1992; Pawlak, 1986).

Dinamik veri: İçeriği sürekli olarak değişen veritabanlarına dinamik veritabanları denir. Bu veritabanlarına örnek olarak kurumsal çevrim-içi veritabanları gösterilebilir. Bir veritabanındaki içeriğin sürekli değişmesi veri madenciliği uygulamalarını zorlaştırmaktadır. Bu veritabanlarında veri madenciliği örüntüleri elde edildikten sonra değişen verilerin örüntülere yansımaları gerekmektedir. Bu işlem, veri madenciliği uygulamasının ürettiği örüntüleri değişen veriye göre, gereken yerlerini güncelleme yeteneğine sahip olmasını gerektirir (Almuallim ve Dietterich, 1991; Deogun vd., 1995).

Eksik veri: Eksik veri, örneklem kümesindeki niteliklerin ya da kayıtların eksik olması ya da bazı kayıtlar için bazı niteliklerin veya nitelik değerlerinin olmamasını ifade etmektedir. Bu eksiklik; hatalı ölçüm araçlarından, veri toplama sürecinde deneyin tasarımında yapılan değişiklikten ya da birbirine benzer ancak özdeş olmayan veri kümelerinin birleştirilmesinden kaynaklanıyor olabilir (Pawlak, 1986).

Null (boş) değerler: Veritabanlarında, niteliğin bilinmeyen veya uygulanamaz bir değere sahip olmasını ifade etmek için Null (boş) değer kavramı kullanılır. Kurulacak bir ilişkide kullanılacak verilerin aynı sayıda niteliğe ve Null olsa bile aynı sayıda nitelik değerine sahip olması gerekir. Veri kümelerinde yer alan Null değerleri için şu çözümler söz konusu olabilir (Quinlan, 1986):

- Null değerli kayıtlar tamamıyla ihmal edilebilir,
- Null değerler yerine olası bir değer atanabilir. Bu değerler o nitelikteki en fazla frekansa sahip bir değer, ortalama bir değer veya varsayılan bir değer olabilir,
- Null değerın kendisine en yakın değeri atanabilir.

2.1.5. Kullanım Alanları

Günümüzde veri madenciliği astronomi, biyoloji, finans, pazarlama, sigorta, tıp ve birçok başka dalda uygulanmaktadır. Son 20 yıldır Amerika Birleşik Devletleri'nde çeşitli veri madenciliği algoritmalarının gizli dinlemeden, vergi kaçakçılıklarının ortaya çıkartılmasına kadar çeşitli uygulamalarda kullanıldığı bilinmektedir (Akpınar, 2000; Dilly, 1999). Bununla birlikte günümüzde veri madenciliği yöntemleri özellikle işletmelerin çeşitli alanlarında başarı ile kullanılmaktadır. Bu uygulamaların başlıcaları ilgili alanlara göre aşağıda gösterilmiştir (Akpınar, 2000; Eker, 2002; Larose, 2005; Yılmaz, 2002; Güvenç, 2001):

- Pazarlama
 - Müşteri gruplandırması
 - Müşterilerin satın alma örüntülerinin belirlenmesi
 - Posta kampanyalarında cevap verme oranının artırılması
 - Mevcut müşterilerin elde tutulması için pazarlama stratejilerinin geliştirilmesi
 - Pazar sepeti analizleri
 - Çapraz satış analizleri
 - Müşteri değerlendirme
 - Müşteri ilişkileri yönetimi
 - Satış tahminleri
- Bankacılık
 - Farklı finansal göstergeler arasındaki gizli korelasyonların bulunması
 - Kredi kartı dolandırıcılıklarının tespiti
 - Kredi kartı harcamalarına göre müşteri gruplarının belirlenmesi
 - Kredi taleplerinin değerlendirilmesi
 - Risk analizleri
- Sigortacılık
 - Yeni poliçe talep edebilecek potansiyel müşterilerin tahmin edilmesi
 - Sigorta dolandırıcılıklarının tespiti
 - Riskli müşterilerin belirlenmesi
- Perakendecilik
 - Satış noktası veri analizleri
 - Alışveriş sepeti analizleri
- Borsa
 - Hisse senedi fiyat tahmini
 - Genel piyasa analizleri
 - Alım-satım stratejilerinin optimizasyonu
- Sağlık
 - Ürün geliştirme
 - Teşhis ve tedavi
 - Test sonuçlarının ya da ilaçların hastalar üzerindeki etkilerinin tahmini
- Eğitim
 - Öğrenci davranışlarının tahmin edilmesi
 - Öğrencilerin performanslarına göre branşlara ayrılması

- Öğrencilerin ders seçme eğilimlerinin belirlenmesi
- Turizm
 - Bir bölgeye gelen turistlerin o bölgeyi tercih sebeplerinin belirlenerek ilişkili reklâm kampanyası düzenlenmesi
 - Gelecek sezonlardaki turist sayısının tahmini ve buna yönelik kapasite plânlanması
- İnternet
 - Kullanıcıların profilini belirleyip ziyaret ettiği sayfalarda onların ilgilerine uygun ürün reklâmlarının gösterilmesi
 - Maillere gelen yanıt oranının artırılması

E-ticaret veri madenciliğinin en hızlı ve kolay gerçekleştirilebileceği alanlardandır, zira bu ticaretin işlemleri sırasındaki her türlü bilgi zaten elektronik ortamda tutulmaktadır. Müşteri ve ürün bilgileri de benzer şekilde veritabanlarında yer almaktadırlar. Hacim ya da boyut ne olursa olsun birbirleriyle ilişkilendirilmiş bu verilerin arasına gizlenmiş olan desenler ya da düzensizlikler, en kolay tespit edilebilen ve göze çabuk çarpan kurallar olarak çıkarılabilir (Vahaplar ve İnceoğlu, 2000).

Veri madenciliği genellikle aşağıdaki karakteristiklere sahip problemlerin çözümünde tercih edilmektedir (Han ve Kamber, 2001):

- Büyük miktarlarda veriye erişilebildiği zaman
- Veri çok sayıda değişkene sahipse
- Veri karmaşık, çok değişkenli ve doğrusal olmayan yapıdaysa
- Çıktıları ya da davranışı tahmin etmek gerekiyorsa
- Henüz anlaşılmayan birliktelik ve ilişkileri bulmak gerekiyorsa

2.1.6. Etkilendiği Unsurlar

Veri madenciliği 5 temel harici unsurdan etkilenmektedir (Vahaplar ve İnceoğlu, 2000):

- a) **Veri:** Son yirmi yılda hızla artan sayısal veri sayısı veri madenciliğindeki gelişmeleri hızlandırmıştır. Veri sayısının çok hızlı oranda artmasına rağmen, bu

verilerle uğraşan bilim adamı sayısının önemli seviyede artmaması nedeniyle verileri analiz etme yöntemleri geliştirilmektedir.

- b) Donanım:** Veri madenciliği, büyük veri kümeleri üzerinde yoğun işlemler yapmayı gerektirir. Gelişen bellek ve işlem hızı kapasitesi sayesinde, birkaç yıl önce madencilik yapılamayan veriler üzerinde çalışmak mümkün hale gelmiştir.
- c) Bilgisayar Ağları:** Hızı ve kapasitesi hızla artan bilgisayar ağları ve internet, dağıtık verileri analiz etmeyi ve farklı algoritmaları kullanmayı mümkün hale getirmektedir.
- d) Bilimsel Hesaplamalar:** Günümüzde bilim adamları, simülasyonu teori ve deneyden sonra bilimin üçüncü yolu olarak görmektedirler. Veri madenciliği ve bilgi keşfi, bu üç metodu birbirine bağlamada önemli rol almaktadır.
- e) Ticari Eğilimler:** Bilgisayar ve internet teknolojilerinin hızlı ilerlemesi sonucu günümüzde ticaret hızı ve hacmi büyük boyutlara ulaşmış, ticaretin önemli bir bölümü elektronik ortamda gerçekleştirilir olmuştur. E-ticaret; minimum maliyet ve en az insan gücünü kullanarak maksimum kârlılığı hedeflemektedir. Bu ticari eğilimler ve hedefler, veri madenciliğinin kullanımını neredeyse zorunlu hale getirmiştir. Zira, veri madenciliği sayesinde müşteri işlemlerinin doğurduğu fırsat ve riskler kolaylıkla ve açıkça görülebilmektedir.

2.1.7. Modelleri

Veri madenciliğinde kullanılan modeller, aşağıda görüldüğü gibi tahmin edici ve tanımlayıcı olmak üzere iki ana başlık altında toplanabilir.

- Tahmin Edici (Predictive) Modeller
 - Sınıflama (Classification) ve Regresyon (Regression)
- Tanımlayıcı (Descriptive) Modeller
 - Kümeleme (Clustering)
 - Birliktelik Kuralları (Association Rules)
 - Ardışık Zamanlı Örüntüler (Sequential Patterns)

Yazında tahmin edici modellerin “denetimsiz modeller”, tanımlayıcı modellerin ise “denetimli modeller” olarak da ifade edildiği görülmektedir (Koyuncugil, 2007; Hastie vd., 2001). Tahmin edici ve tanımlayıcı modeller arasındaki fark kesin sınırlarla ayrılmamıştır. Tahmin edici modeller anlaşılabilir olduğu ölçüde tanımlayıcı model olarak

kullanılabildiği gibi, tanımlayıcı modeller de tahmin edici model olarak kullanılabilirler (Velickov ve Solomatine, 2000).

Tahmin edici modeller, sonuçları veya geçmişteki değerleri bilinen verilerden yola çıkarak bir model geliştirilmesi ve kurulan bu modelden yararlanılarak sonuçları bilinmeyen veri kümelerinin sonuç değerlerinin ya da gelecek değerlerinin tahmin edilmesini hedeflemektedir. Örneğin, borsadaki bir hisse senedinin geçmiş yıllardaki değer hareketlerinden hareket ederek bu senedin gelecek dönemlerde alacağı muhtemel değerlerin tahmin edilmesi bu tip bir modeldir (Akpınar, 2005).

Tanımlayıcı modeller ise mevcut veri kümesinin içindeki ilişkilerin ortaya çıkarılmasını hedeflemektedir. Bir marketten alışveriş yapan müşterilerin özellikle hangi ürünleri aynı anda satın aldıklarının belirlenmesi bu tip modellere bir örnek olarak gösterilebilir. Analizcinin çok geniş veritabanlarındaki bilgileri incelemek, örüntüleri keşfetmek için doğru soruları sorup hipotezler geliştirmesi pratikte zor olduğundan, ilginç örüntüleri keşfetme inisiyatifi veri madenciliği programına bırakılır. Keşfedilen bilginin kalitesi ve zenginliği, uygulamanın kullanılabilirliğini ve gücünü oluşturur (Güvenç, 2001).

2.1.7.1. Sınıflama ve Regresyon Modelleri

Veri madenciliği yöntemleri içerisinde en yaygın kullanıma sahip olan sınıflama ve regresyon modelleri, mevcut verilerden hareket ederek gelecekteki verilerin tahmin edilmesinde kullanılmaktadır. Bu iki model arasındaki temel fark, tahmin edilen bağımlı değişkenin kategorik veya süreklilik gösteren bir değere sahip olmasıdır (Han ve Kamber, 2001).

Sınıflandırma, veri nesnesini daha önceden belirlenen sınıflardan biriyle eşleştirme sürecidir. Verileri ve karşı gelen sınıfları içeren eğitim kümesi ile eğitilen sistem, sonraki aşamalarda sınıf bilgisine sahip olunmayan verilerin ait olduğu sınıfların bulunması için kullanılır. Bu model, müşteri segmentasyonu, kredi analizi, iş modellemesi gibi pek çok alanda sıklıkla kullanılmaktadır (Pryke, 1998).

Regresyon modeli ise sürekli sayısal bir değişkenin, aralarında doğrusal ya da doğrusal olmayan bir ilişki bulunduğu varsayılan diğer değişkenler yardımıyla tahmin edilmesi sürecidir (Bidgoli, 2004). Bu nedenle regresyon yerine tahmin kavramı da

kullanılmaktadır.

Han ve Kamber (2001), sınıflandırma ve regresyon yöntemlerinde aşağıdaki karşılaştırma kriterlerini önermektedir:

- Tahmin doğruluğu: Modelin yeni ya da daha önceden bilinmeyen bir verinin sınıf etiketinin doğru olarak tahmin edilme yeteneğidir.
- Hız: Modelin geliştirilmesindeki hesaplama maliyetine karşılık gelir.
- Sağlamlık: Modelin, gürültülü ya da kayıp veri verildiğinde doğru tahmin yapma yeteneğidir.
- Ölçeklendirilebilirlik: Büyük miktarlarda veri verildiğinde, modeli etkin bir şekilde kurma yeteneğidir.
- Yorumlanabilirlik: Model tarafından sağlanan anlaşılma seviyesidir.

Sınıflandırma ve regresyon modellerinde kullanılan başlıca yöntemler aşağıdaki gibi sıralanabilir (Akpınar, 2005):

- Karar Ağaçları (Decision Trees)
- Yapay Sinir Ağları (Artificial Neural Networks)
- Genetik Algoritmalar (Genetic Algorithms)
- K-En Yakın Komşu (K-Nearest Neighbor)
- Bellek Temelli Nedenleme (Memory Based Reasoning)
- Naive-Bayes
- Çoklu Regresyon (Multiple Regression)
- Lojistik Regresyon (Logistic Regression)
- Faktör ve Ayırma Analizleri

2.1.7.2. Kümeleme Modelleri

Kümeleme, heterojen veri kümelerini veri karakteristikleri bakımından homojen sayılabilecek gruplara bölme işlemidir (Güvenç, 2001). Bu işlem, birbirine benzer nesnelerin aynı kümede, birbirinden farklı nesnelerin ayrı kümelerde yer almasıyla gerçekleştirilir (Chen, 2001). Sınıflamanın aksine, kümeleme önceden tanımlanacak hedef bir değişkene gereksinim duymaz. Başlangıç aşamasında veritabanındaki kayıtların hangi kümelere ayrılacağı veya kümelemenin hangi değişken özelliklerine göre yapılacağı,

konunun uzmanı olan bir kişi tarafından belirlenebilir (Akpınar, 2005).

2.1.7.3. Birliktelik Kuralları ve Ardışık Zamanlı Örüntüler

Birliktelik kuralları, bir arada olan olayların ya da özelliklerin keşfedilmesi sürecidir. Nitelikler arasındaki bütün kombinasyonlar çıkarılarak, her niteliğin bütün değerleri bütün kombinasyonlarda deneyerek örüntüler keşfedilmeye çalışılır (Agrawal vd., 1993). 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. Birliktelik kuralları genellikle “*eğer şu olursa daha sonra bu olur*” şeklinde ifade edilebilir (Akpınar, 2000). Aşağıda buna yönelik örnekler verilmiştir:

- Kahve satın alan müşteriler % 70 ihtimalle çikolata da satın almaktadırlar.
- Kepekli ekmek alan müşteriler, % 90 ihtimalle kola almamaktadırlar.

Ardışık zamanlı örüntüler ise aşağıdaki örneklerde görüldüğü gibi birbirleri ile ilişkisi olan ancak birbirini izleyen dönemlerde gerçekleşen ilişkilerin tanımlanmasında kullanılmaktadır (Akpınar, 2000):

- “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 de satın alacaktır.”

2.1.8. Yöntemleri

Veri madenciliği ile ilgili kullanılan pek çok yöntemin yanına hemen her geçen gün yeni yöntem ve algoritmalar eklenmektedir. Bunlardan bir kısmı uzun yıllardan beri kullanılan, klasik teknikler olarak da ifade edilen istatistiksel yöntemlerdir, örneğin; Regresyon, K - En Yakın Komşuluk ve Kümeleme gibi. Diğer bir kısmı ise genellikle istatistiği temel alan ama daha çok makine öğrenimi ve yapay zekâ destekli yeni nesil yöntemlerdir, örneğin; Karar Ağaçları, Birliktelik Kuralları ve Yapay Sinir Ağları gibi (Koyuncugil, 2007).

Veri madenciliğinde kullanılan başlıca yöntemlerin isimleri, model kategorileri altında önceden verilmişti. Bunların dışında birden fazla tekniği içine alan hibrid yöntemler ve zaman serilerine dayalı yöntemlerden de veri madenciliği yöntemi olarak faydalanılmaktadır. Özet olarak, bilgi keşfine yarayan her yöntem veri madenciliği yöntemi olarak kullanılabilir (Kovalerchuk ve Vityaev, 2002).

Kdnuggets' in Mart 2007' de çevrimiçi gerçekleştirdiği bir anket sonucuna göre en yaygın kullanılan veri madenciliği yöntemleri şunlardır (Web-Kdnuggets1):

- Karar Ağaçları, % 62,6
- Regresyon, % 51,2
- Kümeleme, % 50,2
- Tanımlayıcı İstatistik, % 46,3
- Görselleştirme, % 32,5
- Birliktelik Kuralları, % 26,1
- Yapay Sinir Ağları, % 17,2
- Zaman Serisi Analizi, % 17,2

Çalışmanın kapsamını aşmamak için burada tüm veri madenciliği yöntemlerinin açıklanmasına gerek duyulmamıştır. Bunun yerine, özellikle tahminlemeye yönelik en çok tercih edilen ve çalışmanın uygulamasında da kullanılan Regresyon, YSA ve Zaman Serileri yöntemleri ilerleyen kısımlarda ayrıntılı olarak açıklanmıştır.

2.1.9. Uygulama Süreci

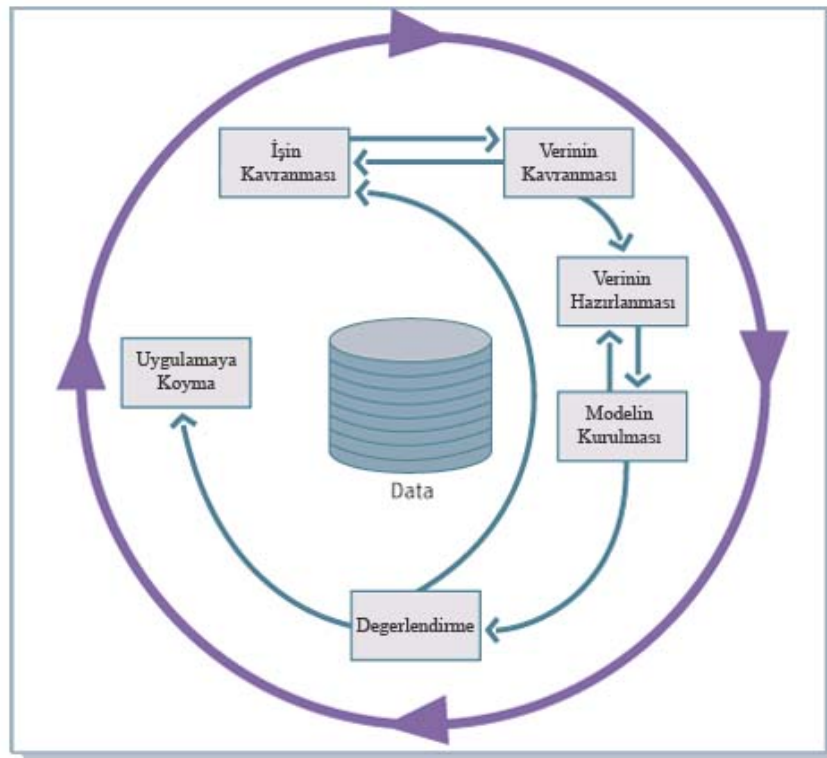
Veri madenciliği uygulaması yapan kurumların bir kısmı bu uygulamanın süreçlerini kendi problemlerine, verilerinin yapısına ve sahip oldukları diğer kaynaklara göre belirlemektedir. Fakat bu süreçlerin belirlenmesinde ve belirlenen süreçteki aşamaların gerçekleştirilmesinde yapılan metodolojik hatalar, uygulamaların başarısız olmasına ya da etkinliğinin azalmasına neden olabilmektedir (Larose, 2005). Bu hataları engelleyerek veri madenciliği projelerinde sistematik bir yaklaşım ve standardizasyon sağlamak amacıyla çeşitli kurumlar farklı sistemler ortaya koymuşlardır. Bunlar arasında en çok tercih edileni; NCR, Daimler Chrysler, SPSS ve OHRA' dan oluşan bir konsorsiyumun geliştirdiği "Çapraz Endüstri Veri Madenciliği Standart Süreci" dir (CRISP-DM, Cross-Industry

Standard Process for Data Mining).

Bu çalışmada da kullanılmasına karar verilen CRISP-DM sürecine göre veri madenciliği süreci altı aşamadan oluşan etkileşimli ve yinelemeli bir süreçtir. Sürecin herhangi bir aşamasında elde edilen sonuçlara göre sonraki aşamaya ya da önceki bir aşamaya geçilip yeni belirlenen problemlere, ilgi alanlarına göre iyileştirmeler ya da farklı işlemler yapılabilir (Larose, 2005).

Veri madenciliği yaşam döngüsü Şekil 2.1'de görülen ve aşağıda açıklanan altı aşamadan oluşmaktadır (Web-CrispDM):

1. İşin Kavranması (Business Understanding)
2. Verinin Kavranması (Data Understanding)
3. Verinin Hazırlanması (Data Preparation)
4. Modelin Kurulması (Modeling)
5. Değerlendirme (Evaluation)
6. Uygulamaya Koyma (Deployment)



Şekil 2.1. Veri madenciliği uygulama süreci (CRISP-DM)
(Kaynak: Web-CrispDM)

1. **İşin Kavranması:** Bu aşamada proje hedeflerinin ve gereksinimlerinin anlaşılmasına odaklanılır. 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.
2. **Verinin Kavranması:** Bu aşamada öncelikle veri toplanmaya başlanır. Veri analizcisi ve alan uzmanları veriyi oluşturan tabloların, satır ve sütunların neyi ifade ettiğini, nasıl bilgiyi barındırdığını belirleyip veriyi tanımaya ve kalitesi hakkında fikir sahibi olmaya çalışır. Ardından bu veriler içinden hangilerinin kullanılacağına karar verilir. Proje hakkında ilk izlenimlere sahip olmak için veriden küçük ilginç örnekler seçilerek hipotezler de oluşturulabilir (Chapman vd., 2000).
3. **Verinin Hazırlanması:** Veri madenciliğinin en önemli aşamalarından bir tanesi olan verinin hazırlanması aşaması analistin toplam zaman ve enerjisinin yaklaşık % 75 ini harcamasına neden olmaktadır. Bu aşamada ham verinin projede kullanılabilir duruma getirilmesi amaçlanır. Firmanın mevcut bilgi sistemleri üzerinde üretilen sayısal bilginin iyi analiz edilmesi, veriler ile mevcut iş problemi arasında ilişki olması gerektiği unutulmamalıdır. Verinin hazırlanması aşaması kendi içerisinde aşağıda anlatılan toplama, birleştirme ve temizleme ile dönüştürme adımlarından meydana gelmektedir (Eker, 2002).
 - a. **Toplama:** Tanımlanan problem için gerekli olduğu düşünülen verilerin ve bu verilerin toplanacağı veri kaynaklarının belirlenmesi adımdır. Verilerin toplanmasında kuruluşun kendi veri kaynaklarının dışında; Merkez Bankası, Devlet İstatistik Enstitüsü, Meteoroloji İşleri gibi dış kaynaklı resmi ya da ticari kurumların veritabanlarından faydalanılabilir.
 - b. **Birleştirme ve Temizleme:** Toplanan verilerde tespit edilen farklılıklar giderilmeye çalışılır. Hatalı veya analizin yanlış yönlenmesine sebep olabilecek veriler temizlenir. Genellikle yanlış veri girişinden veya bir kereye özgü bir olayın gerçekleşmesinden kaynaklanan verilerin, önemli bir uyarıcı enformasyon içerip içermediği kontrol edildikten sonra veri kümesinden atılması tercih edilir Ancak basit yöntemlerle ve baştan savma olarak yapılacak sorun giderme işlemlerinin, ileriki aşamalarda daha büyük

sorunların kaynağı olacağı unutulmamalıdır.

- c. **Dönüştürme:** Kullanılacak model ve algoritma çerçevesinde verilerin tanımlama veya gösterim şeklinde değiştirilmesi gerekebilir. Örneğin kredi riski uygulamasında iş tiplerinin, gelir seviyesi ve yaş gibi değişkenlerin kodlanarak gruplanması gibi.
4. **Modelin Kurulması:** Seçilmiş olan veri ve uygulama aracı ile çeşitli modeller oluşturulur. Genelde, 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 oluşturulduktan sonra kullanılan tekniğin gereksinimlerine uygun olarak veri hazırlanması ya da işin kavranması aşamasına tekrar dönülüp gerekli değişiklikler yapılabilir (Chapman vd., 2000).
5. **Değerlendirme:** Oluşturulan modelin kalitesinin sorgulandığı, bu modelle belirlenen işletme hedeflerine ulaşıp ulaşılamayacağına karar verildiği aşamadır. Bir modelin doğruluğunun test edilmesinde kullanılacak kriterler, geçerlilik testleri veya modelin anlaşılabilirliğidir. Modelin anlaşılabilirliği ve doğruluk oranı gibi konularda da model amaç için yeterli kaliteyi sağlıyorsa uygulama aşamasına geçilir (Chapman vd., 2000; Eker, 2002).
6. **Uygulamaya Koyma:** Kurulan ve geçerliliği kabul edilen model doğrudan bir uygulama olabileceği gibi, bir başka uygulamanın alt parçası olarak da kullanılabilir. 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 (Web-CrispDM).

2.2. Tahminleme

2.2.1. Tahmin Kavramı

Tahmin kavramı genel olarak “bilinmeyi ortaya çıkarma çabası” olarak tanımlanabilir. Bu çabanın temelinde, bilinen verileri inceleyerek bilinmeyen veriler hakkında varsayımlarda bulunmak yatmaktadır. Tahmin, özellikle geleceğe dönük olarak

yapılıyorsa bu amaçla “öngörü” ya da “kestirim” kavramları kullanılabilir. İngilizce’de, mevcut duruma yönelik tahminler için “estimation” ve “prediction” kavramları, geleceğe dönük tahminler için ise “forecasting” ve “prediction” kavramları kullanılmaktadır. “Prediction” kavramına benzer şekilde, Türkçe’de de tahmin kavramı hem mevcut durumlara hem de geleceğe yönelik varsayımlarda genel amaçlı kullanılmaktadır. Bu çalışmada da; tahmin, öngörü ve kestirim gibi birden fazla kavram yerine tüm durumlar için genel olarak “tahmin” kavramının kullanılması tercih edilmiştir.

Tahmin için nitel olarak uzmanların tecrübelerine, yargılarına ve görüşlerine ya da nicel olarak geçmiş verilerin genel yapısına dayanan matematik modellere başvurulabilir. Neredeyse bütün tahminler, tahmin edilecek nesnenin başka bir değişken (örneğin zaman) ya da bir değişken grubu (bir bölgedeki genel ekonomik koşullar gibi) tarafından etkilendiği varsayımına dayanır. Tahmin yöntemlerinin birçoğu on dokuzuncu yüzyılda geliştirilmiş olmasına rağmen, bu alana damgasını vuran gelişmeler bilgisayarların yaygınlaşması ile başlamıştır. Günümüzde, tahmin yöntemlerindeki kendini tekrar eden yordamlardan dolayı bilgisayarların kullanılması bir şart haline gelmiştir. Bu amaçla 90’lı yıllarda çeşitli tahmin yöntemlerini uygulamak amacıyla özel yazılım paketleri geliştirilmiştir (Deniz, 2006).

2.2.2. Tahmin Yöntemleri

Kullanılacak tahmin yöntemi belirlenirken; verilere erişebilme, elde edilen verilerin niteliği ve seyri, kullanılan kaynakların maliyeti, hata payı, anlama ve uygulama kolaylığı, tahminlerin kapsadığı zaman aralığı ve tahminlerin hazırlanması için gerekli olan süre gibi pek çok faktörün dikkate alınması gerekmektedir (Schroeder, 1989, s.71; Klassen ve Flores, 2001, s.172).

Tahmin yöntemleri, tahmin edilen değişken sayısına, tahmin süresinin uzunluğuna ya da yöntemin özelliğine göre çeşitli gruplara ayrılabilirler. Tahmin edilen değişken sayısı bir ise, “tek değişkenli”, birden fazla ise “çok değişkenli” nitelmesi yapılabilir. Geleceği tahmin etmek için genellikle zaman serisi verileri kullanılmaktadır. Bu durumda, tahmin süresi üç aya kadar olan tahminler için “kısa dönemli”, üç ay – iki yıl arası için “orta dönemli” ve iki yıldan daha uzun olanlar için “uzun dönemli” nitelmesi yapılabilir. Geçmiş verilerin süresinin artması veya gelecek sürenin azalması tahminin doğruluğunu artırır (Deniz, 2006).

Tahmin yöntemleri en genel itibariyle “nitel” (kalitatif) ve “nicel” (kantitatif) olarak iki ana grupta ayrıntılı olarak incelenebilir (Gaynor ve Kirkpatrick, 1994).

2.2.2.1. Nitel Tahmin Yöntemleri

“Öznel”, “sübjektif” ya da “yargısal” yöntemler olarak da adlandırılabilen nitel tahmin yöntemleri, tahmin edilmeye çalışılan olgunun anlaşılmasının zor olduğu, bu olguya ilişkin sayısal verilerin elde edilemediği ya da olgunun sayısal olarak ifade edilemeyen bilgiler tarafından etkilendiği, belirsizliğin ve verilerin değişkenliğinin çok fazla olduğu durumlarda kullanılırlar. Bu yöntemler, mevcut durumlara ve gelecekle ilgili plânlara ilişkin bilgisi olması beklenen kişilerin görüşlerinin toplanmasını gerektirmektedir (Monks, 1987, s.268).

Nitel tahmin yöntemleri; esas olarak dört grupta toplanabilir: Delphi Tekniği, Pazar Araştırmaları, Uzman Grup Görüşleri ve Satış Gücü Karması (Demir ve Gümüšoğlu, 2003, s.497).

Nitel yöntemlerin temel üstünlüğü, soyut faktörlerin ve sübjektif tecrübelerin kullanılabilmesine imkân vermesidir. Diğer yandan bu yöntemler altta sıralanan dezavantajları nedeniyle genellikle daha düşük performansa sahip tahminlerle sonuçlanmaktadır (Özdemir, 2006; Deniz, 2006):

- Tahminlerin doğruluğunu ölçebilecek ve geliştirebilecek sistematik yolların olmaması
- Görüşüne başvurulmuş kişilerin önyargı ve eğilimleri nedeniyle yanlış yönlendirmelerine açık olması

2.2.2.2. Nicel Tahmin Yöntemleri

Analizlerde girdi olarak soyut verileri değil, sayısal verileri kullanan yöntemler nicel tahmin yöntemleri olarak adlandırılır. Bu yöntemlerde, verilerin geçmişteki sayısal değerleri kullanılarak elde edilen şablon yardımıyla uygun bir matematiksel model kurulur ve geleceğe yönelik tahminlerde bulunmak için bu modelin tahmin denklemleri kullanılır. Günümüzde nicel tekniklerde yararlanılan sayısal veriler; barkod okuyucuları, çevrimiçi

işlemler veya gözlem sonuçlarından elde edilen verilerin bilgisayar yardımıyla saklandığı veritabanlarından alınmaktadır (Özdemir, 2006; Deniz, 2006):

Nicel tahmin yöntemleri; nedensel teknikler ve zaman serileri olmak üzere iki ana gruba ayrılmaktadır.

Nedensel ya da ilişkisel tahmin modelleri (diğer bir deyişle regresyon analizi), değerleri tahmin edilecek deęişkenle ilişkili olan diğer deęişkenlerin belirlenmesini içermektedir. Bu deęişkenler belirlendikten sonra geliştirilen istatistiksel model, tahmin edilecek deęişken ile diğer deęişkenler arasındaki ilişkiyi tanımlamakta ve ele alınan deęişkene ilişkin tahminler yapılmasında kullanılmaktadır. Bu modellerin özellikle işletmelerde yoğun olarak kullanılmasının nedeni, yönetimin çeşitli alternatif politikaların etkilerini değerlendirmesine imkân tanınmasıdır. Fakat modelin geliştirilmesinin zor olması, tüm deęişkenlere ilişkin geçmiş verilere gereksinim duyulması ve bunun gerektireceęi zaman ve maliyet nedeniyle çeşitli dezavantajlara sahip olduęu unutulmamalıdır (Özdemir, 2006; Bhattacharya, 1997, s. 5).

Zaman serileri analizinde, tahmin edilecek deęişkene ilişkin geçmiş veriler belirli bir veri seyri elde etmek üzere analiz edilmektedir. Bu nedenle tahminleme sadece geçmiş verilerin bu amaçla analiz edilmesine ve yapılacak tahminlerde kullanılmasına dayanmaktadır. Bu özelliğinden dolayı zaman serileri analizi, deęişmeyen koşullar altında daha etkin olmaktadır (Bhattacharya, 1997).

Veri madencilięi yöntemleri arasında yer alan yapay sinir aęları, hem nedensel modellerin hem de zaman serilerine dayanan modellerin kurulmasında başarıyla kullanılabilirlerdir.

İlerleyen kısımlarda regresyon analizi, zaman serileri analizi ve yapay sinir aęları konusunda daha ayrıntılı bilgiler sunulacaktır.

2.2.3. Tahmin Modellerinin Doğruluklarının Ölçülmesi

Bir tahmin modelinin doğruluęu, tahmin deęerlerinin gerçek gözlem deęerlerine ne kadar yakın olduęuna baęlıdır. Bu doğruluęu ölçmek için, ortaya konan model kullanılarak gerçek gözlem deęerleri bilinmiyormuşçasına tekrar hesaplanarak tahmin edilir. Ardından

bu tahmin değerleri ile gerçek gözlem değerleri arasındaki farklar, çeşitli istatistikler yardımıyla karşılaştırılır. Tahmini gözlem değerleri ile gerçek gözlem değerleri arasındaki bu farklara tahmin hatası (kalıntı) adı verilmektedir (Akal, 2004).

Tahmin modellerinin doğruluğunu ölçmek için yaygın olarak şu istatistikler kullanılmaktadır: Ortalama Mutlak Hata (OMH) (Mean Absolute Error, MAE), Ortalama Mutlak Yüzde Hata (OMYH) (Mean Absolute Percentage Error, MAPE), Ortalama Hata Kareleri Kökü (OHKK) (Root Mean Squared Error, RMSE), Hata Kareleri Ortalaması (HKO) (Mean Squared Error, MSE), Ortalama Yüzde Hata (OYH) (Mean Percentage Error, MPE). Yazındaki çalışmalarda bu istatistikler arasında en çok kullanılanların OMH, OMYH ve OHKK olduğu görülmüştür.

Tahmin hatası ile tahmin doğruluğu ters orantılıdır, yani tahmin hatası küçüldükçe modelin doğruluk derecesi artmaktadır. Farklı tahmin modellerini karşılaştırırken en iyi model ideal olarak, kullanılan istatistiklerin tamamında en küçük değeri veren modeldir. Ancak, verilerin yapısına bağlı olarak, özellikle aşırı sapmalar nedeniyle bu ideallik her zaman sağlanamaz ve bazı istatistikler abartılı sonuçlar verebilir. Bu tip durumlarda, hataları yüzde olarak ifade etmesi nedeni ile tek başına da bir anlam taşıyan OMYH istatistiği referans alınıp doğrulukları karşılaştırmada tercih edilebilir. OMYH değeri % 10'un altında olan modeller "çok iyi", % 10 ile % 20 arasında olan modeller "iyi", % 20 ile % 50 arasında olan modeller "kabul edilebilir" ve % 50'nin üzerinde olan modeller ise "yanlış ve hatalı" olarak kabul edilmektedir (Akgül, 2003, s.138; Lewis, 1982, s. 40).

En iyi tahmin modelinin belirlenmesinde, tahmin hatası istatistiklerinin yanında şu kriterlerin de göz önünde bulundurulmasında fayda vardır: Veriye ulaşılabilirlik, veri toplama maliyeti, modelleri çalıştıracak bilgisayar yazılımlarının varlığı ve maliyeti, modelin kurulduğu zaman çerçevesi, istenilen tahmin çeşidi (nokta/aralık) (Gaynor ve Kirkpatrick, 1994).

Çalışmada kullanılan tahmin doğruluğu ölçme istatistiklerine ait formüller aşağıda verilmiştir:

$$\text{OMH} = \frac{\sum_{t=1}^n |e_t|}{n}$$

$$\text{OMYH} = \frac{\sum_{t=1}^n \frac{|e_t|}{y_t}}{n} 100(\%)$$

$$\text{OHKK} = \sqrt{\frac{\sum_{t=1}^n e_t^2}{n}}$$

Bu formüllerde;

n : tahmin edilen dönem sayısını,

y_t : t dönemindeki gerçek gözlem değerini,

\hat{y}_t : t dönemi için hesaplanan tahmin değerini,

e_t : t dönemindeki tahmin hatasını göstermektedir ($e_t = y_t - \hat{y}_t$).

Gelecek tahmini yapabilmek için bilgisayar destekli birçok yöntem geliştirilmiştir. Bilgisayar destekli grafikler ile analiz yöntemleri, uzman sistemler, bulanık mantık, regresyon analizi, zaman serileri analizi ve yapay sinir ağları bu yöntemlerden birkaçıdır.

Bu çalışma kapsamında tercih edilen yöntemler olan regresyon analizi, zaman serileri analiz yöntemleri ve yapay sinir ağları aşağıda ayrıntılı olarak anlatılmıştır.

2.3. Regresyon Analizi

2.3.1. Tanımı

Regresyon analizi bir bağımlı (açıklanan) değişken ile en az bir bağımsız (açıklayıcı) değişken arasındaki ilişkiyi matematiksel bir denkleme dönüştürmek için kullanılan bir istatistiksel analiz yöntemidir. Ortaya çıkarılan matematiksel denklem, regresyon denklemi olarak adlandırılır. Regresyon analizinde amaç her bağımsız değişkenin bağımlı değişkendeki değişime katkısının hesaplanması, yani bağımsız değişkenlerin değerinden hareketle bağımlı değişkenin değerinin tahmin edilmesidir (Xu, 2003). Böylece ilgilenilen değişken üzerine yapılacak plânlarda hangi faktörlerin ne derecede etkisinin olduğunu önceden bilme imkânı elde edilmiş olur.

2.3.2. Çeşitleri

Regresyon analizi, kullanılan bağımsız değişken sayısına göre;

- Basit regresyon analizi (Tek bağımsız değişken),
- Çoklu regresyon analizi (Birden çok bağımsız değişken),

fonksiyon tipine göre;

- Doğrusal regresyon analizi,
- Doğrusal olmayan (eğrisel) regresyon analizi,

verilerin kaynağına göre;

- Anakütle verileriyle regresyon analizi,
- Örnek verileriyle regresyon analizi

olmak üzere üç ana grupta toplanabilir (Orhunbilge, 2002, s.5-12).

Çalışmanın uygulama bölümünde çok değişkenli doğrusal regresyon yöntemi kullanıldığı için aşağıda sadece basit ve çoklu doğrusal regresyon yöntemleri açıklanmıştır

2.3.3. Basit Doğrusal Regresyon

Aralarında doğrusal ilişki bulunan bir bağımlı ve bir bağımsız değişken arasındaki ilişkiyi çözümleyen regresyon analizi türüdür. Basit doğrusal regresyon modeli aşağıdaki denklemde görülmektedir.

$$Y = b_0 + b_1 X + e$$

Bu denklemdeki;

X : Bağımsız değişkeni,

Y : Bağımlı değişkeni,

b_0 : Doğrunun $X=0$ olduğunda y eksenini kestiği noktayı,

b_1 : Doğrunun eğimini, diğer adıyla regresyon katsayısını,

e : Hata değerini, diğer adıyla artıkları (ya da kalıntıları)

göstermektedir. Regresyon katsayısı b_1 , bağımsız değişkendeki bir birimlik değişimin bağımlı değişkende ne miktarda bir değişim oluşturduğunu gösterir.

2.3.4. Çoklu Doğrusal Regresyon

Çoklu regresyon analizi, birden çok sayıdaki bağımsız değişkenden yola çıkarak bağımlı değişkenin tahmin edilmesi ve hangi bağımsız değişkenlerin bağımlı değişken üzerinde ne kadarlık etkiye sahip olduğunu bulunması için kullanılır (Xu, 2003).

Çoklu doğrusal regresyon modeli aşağıdaki denklemde görülmektedir. Bu denklemdeki X_n , n nolu bağımsız değişkeni ve bu değişkenin regresyon katsayısını göstermektedir.

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n + e$$

Regresyon denklemlerindeki regresyon katsayıları en küçük kareler yöntemi kullanılarak hesaplanır. Oluşturulan regresyon modelinin veriyi ne kadar iyi açıkladığının ölçülmesi için 0-1 aralığında yer alan R , R^2 ve düzeltilmiş R^2 değerleri kullanılmaktadır. Bu değerler aşağıda kısaca açıklanmıştır (Web-Sakarya; Xu, 2003):

R : Bağımlı değişkenle bağımsız değişken arasındaki korelasyonu temsil eden bu değer yüksek olması bağımsız değişkenle bağımlı değişken arasında sıkı bir ilişkinin olduğunu veya bağımsız değişkenin bağımlı değişkendeki değişimin önemli bir kısmını açıkladığını göstermektedir.

R^2 : Açıklayıcılık katsayısı olarak da bilinen bu değer, bağımlı değişkendeki değişimin yüzde kaçının bağımsız değişken tarafından açıklandığını ifade etmektedir.

Düzeltilmiş R^2 : Regresyon modeline ilave edilen her yeni bağımsız değişken bağımlı değişkeni açıklama yeteneği olsun veya olmasın R^2 değerinin büyümesine sebep olarak sanki modelin açıklayıcılığının iyileştiği izlenimi vermektedir. Bu sebeple modele görece katkısı az olan veya katkısı olmayan değişkenlerin sebep olduğu R^2 değerindeki yapay artışın dikkate alınarak modelin açıklayıcılığının belirlenmesi gerekmektedir. Düzeltilmiş

R^2 değeri bu düzeltme sonucunda ortaya çıkan bir değer olup, regresyon modelinin gerçek açıklama gücünü göstermektedir. Böylece, daha gerçekçi bir regresyon modelinin oluşturulması ve R^2 değerinin hesaplanması mümkün olmaktadır. Genellikle R^2 değerinden biraz daha küçük değere sahip olan düzeltilmiş R^2 , model uygunluğu ölçümünde göreceli olarak daha iyi bir performansa sahiptir ve birçok istatistik yazılımı tarafından tercih edilmektedir.

F Testi: Regresyon modelinin anlamlı olup olmadığını incelemek için ANOVA (Analysis of Variance) testi uygulanmaktadır. ANOVA testi sonucunda ortaya çıkan F değerine karşılık gelen anlamlılık seviyesi, oluşturulan modelin uygun olup olmadığının kararında yardımcı olmaktadır. F test değerinin anlamlı olması (yani 0,05'ten küçük olması) söz konusu modelin bağımlı değişkeni açıklamada önemli katkı sağladığına işaret etmektedir.

2.4. Zaman Serileri Analizi

2.4.1. Zaman Serilerinin Tanımı ve Unsurları

Bir ya da daha fazla değişkenin belirli zaman aralıklarında yapılan ölçümlerinden elde edilen sıralı veri kümeleri “zaman serileri” olarak adlandırılmaktadır. Zaman serileri yardımıyla geçmiş gözlem dönemlerine ait veriler kullanılarak, gelecek dönemler hakkında tahminde bulunulabilmektedir (Yılmaz, 2002; Kutlar, 2000).

Zaman serileri, kendisine etki eden pek çok faktörün etkisinde dalgalanma göstermektedir. Bu dalgalanmalar; ekonomik, sosyal, demografik, psikolojik vb. çeşitli faktörlerin zaman serilerini farklı yön ve şiddette etkilemesinden kaynaklanmaktadır. Zaman serilerinin en önemli özelliği gözlem değerlerindeki değişmelerin genelde bu unsurların ortak etkisi neticesinde oluşmasıdır. Bu unsurlar, aşağıda açıklanan dört ana grupta toplanabilir (Gürtan, 1977, s. 421; Çömlekçi, 1979, s. 307-308; Serper, 1981, s.203-206; Hamburg, 1970, s.542-544; Chatfield, 1996, s.12-13):

- Trend (Uzun dönemli eğilim) (T)
- Mevsimsel dalgalanma (S)
- Konjonktürel (Devirsel) dalgalanmalar (C)
- Tesadüfi dalgalanmalar (I)

Üstte anılan unsurlara sahip bir zaman serisi genel olarak alttaki denklemlerle gösterilebilir:

$$Y = T + C + S + I$$

Yazında “rassal terim” ya da “rassal gürültü” olarak da ifade edilen tesadüfî dalgalanmalar, I sembolü dışında ϵ sembolü ile de gösterilmektedir.

- *Trend*; bir zaman serisinin uzun dönemde belirli bir yöne doğru gösterdiği gelişmedir. Trendler doğrusal veya eğrisel olabilir, önemli olan istikrarlı olmalarıdır.
- *Mevsimsel dalgalanma*; Serinin bazı dönemlerin etkisiyle kısa zaman aralıklarında kendini tekrarlama eğilimi göstermesidir. Bu dönemler, bir yıldan uzun süreli olamaz. Örneğin tarım ürünlerinin fiyatlarının ya da turist sayılarının mevsime bağlı olarak değişmesi ve her yılın aynı mevsimlerinde benzer iniş-çıkışlar göstermesi gibi.
- *Konjonktürel dalgalanmalar*; Serinin 3-5 yıllık orta zaman aralıklarında kendini tekrarlama eğilimi göstermesidir. Örneğin; üretim, satışlar, gelirler gibi iktisadi olaylarda önce bir süre gelişme görülür, bu yükselmenin maksimuma çıktığı anda bir iktisadi kriz ortaya çıkar ve düşüş başlar, bu düşüşten bir müddet sonra tekrar yukarı doğru hareketlenme başlar ve aynı olaylar benzer şekilde sürekli tekrarlanır. Bu hareketler mevsimsel dalgalanmalara benzer şekilde devri olarak tekrar etseler de, devreler daha uzun, süreler de belirsizdir.
- *Tesadüfî dalgalanmalar*; Serinin, üstte sayılan dalgalanmaların dışında, belirli bir sistematik yapıya sahip olmaksızın rastgele değişim gösteren kısmını ifade etmektedir. “Oto regresif süreç” olarak adlandırılan kargaşa, savaş, deprem, sel, grev gibi bu hareketlerin ne zaman nasıl tekrarlanacağı belirsizdir.

2.4.2. Zaman Serilerinin Temel Özellikleri

2.4.2.1. İç Bağımlılık

Zaman serilerini diğer serilerden ayıran en belirgin özellik, gözlem değerlerinin birbiri ile ilişkili olmasıdır. Birbirine yakın gözlemler genellikle uzak olanlara kıyasla daha çok benzerdir. Yani veriler arasındaki ilişki zaman aralığıyla ters orantılıdır. Bu nedenle,

gözlemlerin birbiriyle bağımsız olduğu varsayımı altında gerçekleştirilen istatistiksel teknikler zaman serileri için geçerli olmayacaktır. Zaman serilerinde var olan “iç bağımlılık” nedeniyle, geçmiş dönem gözlem verilerinden yola çıkarak gelecek dönemlerde oluşacak değerleri tahmin etme imkânı bulunmaktadır (Kutlar, 2000; Göçmençebe, 1976, s. 185).

2.4.2.2. Stokastik (Tesadüfi) Süreç Olması

Zaman serilerinin bir başka önemli özelliği de stokastik süreç olmasıdır. Yazında matematiksel fonksiyon ile kesin bir biçimde belirlenebilen seriler “deterministik” olarak ifade edilirken, gözlem değerleri bir olasılık dağılımı ile açıklanabilen seriler “stokastik” olarak adlandırılmaktadır. Zaman serilerini açıklamak için kullanılan modellerde, olayları etkileyen tüm faktörlerin her zaman modele dâhil edilmesi mümkün olmadığı için modele rassal hareketleri gösterecek bir “tesadüflük” unsurunun eklenmesi gerekmektedir. Bu rassal unsur nedeniyle zaman serileri stokastik olma özelliği taşımaktadır (Chatfield, 1996, s.6).

2.4.3. Zaman Serilerinin Sınıflandırılması

Zaman serilerini; değişken sayısına göre; “tek veya çok değişkenli”, elde edilmiş biçimine göre; “sürekli veya kesikli”, ortalamadan gösterdiği sapmalara göre; “durağan veya durağan olmayan”, devirsel hareketlenmelerine göre; “mevsimsel veya mevsimsel olmayan” şeklinde sınıflandırmak mümkündür.

Tek veya çok değişkenli zaman serileri: Zaman içindeki hareketi izlenen olay sayısı bir olan seriler “tek değişkenli”, birden fazla olan seriler ise “çok değişkenli” olarak adlandırılır.

Sürekli ve Kesikli Zaman Serileri: Şayet incelenen olaya ait gözlem değerleri zaman içinde sürekli olarak elde ediliyorsa, bu serilere “sürekli zaman serisi” denilmektedir. Bu tür seriler genellikle zaman içinde eşit olmayan aralıklarla elde edilen gözlem değerlerinden oluşur. Şayet gözlemler sadece belirli zaman aralıklarıyla yapılıyorsa, bu tip serilere “kesikli zaman serileri” denir. Kesikli zaman serileri genellikle eşit ve birbirini takip eden zaman aralıklarıyla yapılan gözlem değerlerinden oluşur. Özellikle ekonomik büyüklükleri gösteren seriler, belirli zaman aralıklarında ölçüldüğü için, kesikli zaman

serileri sınıflaması altında incelenmektedir (Gürsakal, 2000, s.36). Gözlemlerin sürekli yapıldığı hallerde bile, belirli zaman aralıkları için gözlem değerlerinin ya toplamı veya ortalaması alınarak, ya da örnekleme yoluna gidilerek sürekli seriler kesikli hale dönüştürülebilir (Box ve Jenkins, 1976, s. 23). Uygulamada çoğunlukla kesikli seriler üzerinde çalışılmaktadır. Bu çalışmada kullanılan zaman serileri de, günlük gözlemlenen verilerin aylık ortalamalara dönüştürülmesi nedeniyle kesikli zaman serileri kategorisinde yer almaktadır.

Durağan ve Durağan Olmayan Zaman Serileri: Durağanlık stokastik süreçlerde çok önemli bir kavramdır. Stokastik süreç olarak bir zaman serisinin tüm özellikleri; yani ortalaması, varyansı, kovaryansı ve daha yüksek dereceden momentleri zamana göre değişmiyorsa veya seri periyodik dalgalanmalardan arınmışsa “durağan zaman serisi” olarak adlandırılmaktadır. Gerçek hayatta durağan zaman serisi örneklerine çok az rastlanır, özellikle iktisadi zaman serilerinin çoğu durağan olmayan serilerdir. Bu seriler, zaman serisini meydana getiren trend, mevsimsel dalgalanmalar, konjonktür dalgalanmaları ve tesadüfi dalgalanmalardan birini veya birkaçını birlikte içerirler. Bu nedenle uygulamada en çok karşılaşılan durağan olmayan seriler bir takım dönüşüm yöntemleri kullanılarak durağan hale getirilir, daha sonra analiz edilir. Bu dönüşüm zorunludur, çünkü zaman serileri analizi için geliştirilmiş ve kullanılan olasılık modelleri sadece durağan zaman serilerine uygulanabilir (Wasney A.F., 1976, s.3; Erdoğan, 2006).

Mevsimsel ve Mevsimsel Olmayan Zaman Serileri: Birbirini takip eden yılların aynı dönemlerinde benzer hareketler gösteren zaman serileri; “mevsimsel zaman serisi” olarak adlandırılmaktadır. Bunu tespit edebilmek için zaman serilerinin yeterli sayıda gözlem değerini içermesi gerekir.

2.4.4. Analiz Amaçları ve Uygulama Alanları

Bir zaman serisi genellikle şu amaçlarla analize tabi tutulmaktadır: tanımlama (grafiksel, sayısal, özellik), açıklama (açıklayıcı değişkenler), tahmin (gelecek yıl satış rakamları) ve kontrol (ekonomide imalat yönteminin kalitesi). Zaman serileri zamana göre değişen bir ya da daha çok olayın incelenmesinde kullanılmaktadır. Uygulama alanlarına örnek olarak; ekonomi (enflasyon, faiz oranları), ticaret (satışlar, fiyatlar), sosyoloji (suç oranları, boşanma oranları), meteoroloji (yağış miktarı, sıcaklık), astronomi (güneş hareketleri, göktaşlarının hareketleri), çevrebilim (çevre kirliliği, su kirliliği) gösterilebilir

(Erdoğan, 2006, s.7).

2.4.5. Analiz Aşamaları

Zaman serisi analizi aşağıda açıklanan dört aşamada gerçekleştirilmelidir (Chatfield, 1996, s.7.):

2.4.5.1. Tanımlama

Zaman serisi analizlerinde yapılması gereken ilk işlem, verilerin zamana karşı grafiklerini oluşturarak serinin özelliklerini belirlemeye çalışmaktır. Bu tespitler sonucunda, gerek duyuluyorsa aşağıda sıralanan dönüşümler gerçekleştirilmelidir.

- Varyansı homojen yapmak,
- Mevsimsel etkiyi gidermek,
- Normal dağılıma yaklaştırmak.

2.4.5.2. Açıklama

Bu aşamada uygun bir zaman serisi modeli kurularak, modelin verilere uyumunun test edilmesi ve varyasyon kaynaklarının açıklanması gerekmektedir.

2.4.5.3. Tahminleme

İkinci aşamada kurulmuş olan modele göre, gelecek dönemlere ait değerler olasılık kuralları içerisinde tahminlenmelidir. Gelecek ile ilgili tahminlerde bulunurken elde edilen doğruluk derecesi tahminin gücünü gösterecektir. Tahminin gücü, kullanılan modelin gerçek modele yakınlık derecesi olarak ifade edilebilir.

2.4.5.4. Kontrol

Tahmin sonrasında elde edilen geçmişe yönelik değerler, gerçek değerler ile karşılaştırılarak modelin doğruluk derecesi kontrol edilmelidir. Şayet model istenen doğruluk derecesine sahip değilse geriye dönülerek daha farklı tahmin modellerinin kurulmasına yönelik işlemler tekrar edilmelidir.

Zaman serilerinin analizi esnasında, durağanlığın testi için şu testler kullanılmaktadır:

- Birim Kök Testi
- Dickey & Fuller Birim Kök Testi
- Geliştirilmiş Dickey Fuller Birim Kök Testi
- Normal Dağılım Testi
- Ljung & Box ve Box & Pierce Testi
- Dönüm Noktası Testi
- İşaret Testi
- Trend Testi

2.4.6. Zaman Serilerinde İleriye Dönük Tahmin Yöntemleri

Zaman serilerinde ileriye dönük tahmin amaçlı kullanılan yöntemler, serinin tek ya da çok değişkenli olmasına göre farklılık gösterebilmektedir.

Çok değişkenli zaman serilerinde genelde dinamik regresyon modelleri ve çok değişkenli zaman serileri analizi gibi yöntemler kullanılmaktadır. Çok değişkenli zaman serileri analizi, tahmin sistemiyle ilgili her şeyin bilindiğini varsayımını dikkate almaktadır, bu amaçla çeşitli simülasyon ya da istatistik yöntemleriyle bilinmeyen değişken değerlerinin yapay olarak üretilmesi yoluna gidilir. Bu şekilde hesaplanan verileri kullanarak yapılan tahminler gerçekçi sonuçlar vermekten uzaklaşacağı için güvenilirlikleri düşük olmaktadır. Bu nedenle uygulamada çok değişkenli zaman serilerinin tahmini çok tercih edilmemekte; bunun yerine yaygın olarak tek değişkenli zaman serileri kullanılmaktadır. Bu çalışmada da tek değişkenli zaman serileri kullanıldığı için, ilerleyen kısımlarda sadece tek değişkenli zaman serilerine yönelik tahmin yöntemlerine yer verilmiştir.

Tek değişkenli zaman serileriyle ilgili tahmin yöntemleri, mevcut olan zaman serisi unsurlarının gelecek dönemde de aynı kalacağı varsayımına dayanmaktadır. Bu varsayım nedeniyle geçmiş dönem gözlem değerleri kullanılarak gelecek dönem tahmin değerlerinin elde edilmesi amaçlanır (Çömlekçi, 1979, s. 320).

Zaman serilerinin gelecek değerlerinin tahmin edilmesinde kullanılan yöntemin seçimi;

tahminin amacına, zaman serisinin tipine ve unsurlarına (trendin ve mevsimselliğin olup olmaması gibi), geçmişe ait ne kadar veri elde edilebileceğine ve tahmin döneminin uzunluğuna bağlı olarak değişmektedir (Akgül, 2003, s.53). Günümüzde, gelişmiş bilgisayar yazılımları aracılığıyla en uygun modelin otomatik olarak belirlenmesi mümkün olmaktadır. Bu nedenle, çalışmanın kapsamı dışına çıkılmaması amacıyla aşağıda bu yöntemlerin sadece kavramsal temelleri açıklanmış, alt yapısını oluşturan karmaşık matematiksel fonksiyonların anlatımına girilmemiştir.

2.4.6.1. Trend Analizi

Trend analizi yöntemi, zaman serilerinde ileriye dönük tahmin yapma konusunda geliştirilen en eski yöntemdir. Bu yöntemin temelinde, tahmini yapılacak olan zaman serisi değişkenine ait değerlerin serpilme diyagramındaki dağılımlarına uygun bir trend denklemi belirleyerek gelecek dönemler için rutin tahminlerde bulunmak yatmaktadır. Genelde orta ve uzun dönemli tahmin amacıyla kullanılan bu yöntemin en önemli avantajları; anlaşılmasının ve hesaplanmasının kolay, maliyetinin de düşük olmasıdır. Hesaplamalar bilgisayar kullanma zorunluluğu gerektirmeden de yapılabilmektedir (Charmbers vd., 1971, s.55). Basit olmasından kaynaklanan bu avantajlarına karşılık belli başlı dezavantajları ise şunlardır (Chatfield, 1996; Makridakis ve Wheelwright, 1978, s. 121):

- Sadece tek değişkenli zaman serilerinde kullanılabilmesi,
- Mevsimselliği dikkate almaması,
- En az yedi yıla ait yıllık veriye gereksinim duyması,
- Seriyeye yeni gözlem değerleri eklendiğinde kolay adapte olamaması; hesaplamaların tüm gözlem değerleri kullanılarak yeniden yapılması,
- Belirlenen tahmin denkleminin yeterliliğinin istatistiksel olarak test edilememesi.

2.4.6.2. Mekanik Tahmin

Bu yöntem, zaman serilerinde ileriye dönük tahmin yapma konusunda geliştirilen en basit yöntemdir. Yöntemin temelinde, zaman serisine konu olan olayın son dönemde aldığı değerlerin bir sonraki dönemin tahmini olarak kullanılması yatmaktadır. Mekanik yöntemler, daha önceden yapılmış olan tahminlerin üstünlüğünün testinde de kullanılabilir (Özmucur, 1990, s. 35).

2.4.6.3. Hareketli Ortalamalar

Tahmin doğruluğunu artırmak düşüncesinden yola çıkan hareketli ortalamalar yöntemi; önce zaman serisindeki gözlem değerlerini belirli büyüklükteki kümeler halinde toplar, sonra her küme için aritmetik ortalamayı hesaplar ve bu ortalamaları ait olduğu kümenin en yeni terimini takip eden terimin tahmin değeri olarak kullanır. “Hareketli ortalamalar” adının kullanılmasının sebebi; seriye eklenen her yeni değer ile birlikte yeni bir ortalamanın hesaplanarak tahmin amacıyla kullanılabilmesidir (Makridakis ve Wheelwright, 1989, s. 67).

Gözlem değerlerinin oluşumundaki tesadüflüğün yüksek, buna karşılık ardışık gözlem değerleri arasındaki otokorelasyonun düşük olduğu zaman serilerinde tercih edilen bu yöntemin dezavantajları şöyle sıralanabilir (Şahin, 2001, s. 163; Wheelwright ve Makridakis, 1973, s. 34-35; Chambers vd., 1971, s. 55):

- Mevsimselliği dikkate almaması,
- Sadece hareketli ortalama dönemindeki tarihi verilere eşit ağırlık verip eski dönemleri görmezden gelmesi,
- Çok sayıda gözlem değerine gereksinim duyması,
- Tahmini yapılacak bir olayın temel unsurlarında değişiklik meydana geldiğinde, bu yöntemin hemen adapte olamaması,
- Uzun dönem tahmin değerlerinin doğruluk derecesinin düşük olması nedeniyle genelde sadece kısa dönem tahmin amacıyla kullanılabilmesi.

Hareketli ortalamalar yöntemleri aşağıda görülmektedir. Bu yöntemlerin birbirinden farkı, ortalamaların hesaplanmasında kullandıkları değerlemelerdir.

- Basit Hareketli Ortalama,
- Ağırlıklı Hareketli Ortalama,
- Çift Hareketli Ortalama.

2.4.6.4. Üstel Düzleştirme

Üstel düzleştirme yöntemleri hareketli ortalamalar yöntemine benzemektedir. Temel farklılıkları; geçmiş dönem verilerine eşit değil, eskidikçe üstel olarak azalan ağırlıklar

vermeleridir. Bu nedenle, yakın dönemdeki verilerin geleceğe etkisi, uzak dönemdeki verilerden daha çok olmaktadır. 1958’de Holt tarafından geliştirilen üstel düzleştirme yöntemleri ilk olarak mevsim ve trend unsuru içermeyen basit zaman serilerinde uygulanmıştır. Ardından Brown, yönteme tam açıklık kazandırarak uygulama alanına koymuştur, Winter ise yöntemi mevsimsellik unsuru içeren zaman serilerine de uygulanabilecek hale getirmiştir (Chatfield, 1996, s. 85). Bu yöntemlerin avantajları aşağıdaki gibi sıralanabilir (Orhunbilge, 2002, s. 95):

- Trend ve mevsimsellik unsurlarını dikkate alması,
- Hızlı olması,
- Düşük maliyetli olması,
- Yeni gözlem değerlerinin eklenmesi durumunda kolayca adapte olabilmesi ve yeni dönemlerin tahminlerine hemen katkıda bulunabilmesi.

Günümüze kadar geliştirilmiş çok sayıda üstel düzleştirme yöntemi bulunmaktadır. Geliştirilen her yöntem, kendinden önceki yöntemlerin dezavantajlarını avantaja dönüştürmeyi amaçladığı için üstel düzleştirme yöntemleri arasında; doğruluk derecesi, hesaplama yöntemi, işlem süresi, parametre sayısı gibi çeşitli farklılıklar bulunmaktadır. Zaman serisi değerlerine bağlı olarak bazı yöntemler diğerlerinden daha iyi sonuçlar vermektedir. Aşağıda yazında en çok adı geçen belli başlı üstel düzleştirme yöntemleri görülmektedir:

- (Brown’un) Basit Üstel Düzleştirme Yöntemi
- Brown’un Tek Parametrelili Doğrusal Üstel Düzleştirme
- Holt’un Çift Parametrelili Doğrusal Üstel Düzleştirme Yöntemi
- Winter’ın Mevsimsel Üstel Düzleştirme Yöntemi
- Chow’un Uyarlanabilir Kontrol Yöntemi
- Harrison’un Harmonik Düzleştirme Yöntemi
- Uyarlanabilir Tepki Oranlı Basit Üstel Düzleştirme Yöntemi
- Brown’un Tek Parametrelili Uyarlanabilir Yöntemi

2.4.6.5. Uyarlayıcı Arındırma

1970’lerin ilk yarısında Wheelwright ve Makridakis tarafından geliştirilmiş olan bu yöntem, zaman serisine konu olan olayın tahmin modeli belirlendikten sonra bu olayı

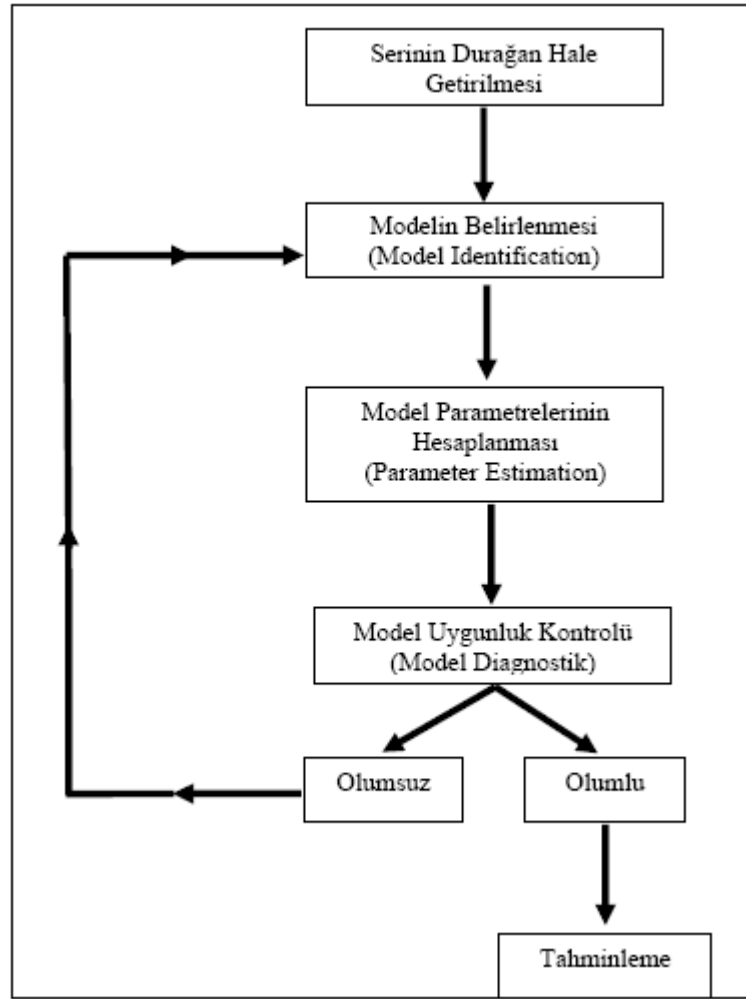
meydana getiren unsurlarda meydana gelebilecek deęişiklikleri yeniden bir tahmin modeli belirlemeye gerek bırakmadan doğrudan tahmin deęerlerine yansıtma imkânı sağlamaktadır. Bu özelięinden dolayı “kendi kendini yenileyen (otoregresif)” modeller olarak da adlandırılmaktadır. Gelecek dönemin tahmin deęeri, hareketli ortalamalar ve üstel yöntemlerde olduęu gibi geçmiş dönem gözlem deęerlerinin toplamları alınarak elde edilir. Bu yöntem, hareketli ortalama ve üstel düzleştirme yöntemleri gibi sadece kısa dönem tahmin amacıyla kullanılmakla birlikte, tahmin sonuçlarının güvenilirlięi her iki yöntemden daha fazladır. Tahmin işlevinde analistin müdahalesini minimum düzeye indiren bu yöntem, az veriyle de kullanılabilmesi, kolay uygulanabilmesi, hesaplama süresinin kısa olması gibi avantajlara sahiptir (Makridakis ve Wheelwright, 1977, s. 426; Ekern, 1976, s. 707).

2.4.6.6. Box-Jenkins Yöntemleri

İncelenen zaman serisinin stokastik, olayların da rassal karakterde olduęu varsayımına dayanarak geliştirilmiş olan Box-Jenkins yöntemleri, rassal deęişkenin zaman içinde aldığı ardışık deęerler arasında mevcut olan iç baęımlılıęı çok etkili bir şekilde dikkate almasından dolayı “doęrusal stokastik modeller” olarak da adlandırılmaktadır (Box ve Jenkins, 1976, s. 7-21).

Box-Jenkins yöntemleri, zaman serisi analizlerinde çok yaygın olarak kullanılmaktadır. Bunun temel sebebi ise yöntemin duraęan olmayan ve mevsimsellik içeren zaman serilerine de kolayca uygulanabilmesinden kaynaklanmaktadır (Sevüktekin ve Nargeleçekenler, 2005).

Zaman serilerinin ARIMA ile modellenmesine, George Box ve Gwilym Jenkins’in 1970 yılında yayınlanan ünlü “Time Series Analysis: Forecasting and Control” kitaplarında ayrıntılarıyla yer verilmiştir. Box ve Jenkins’in bu kitapta açıkladıkları ve daha sonra Box-Jenkins metodolojisi olarak yazına giren metot, temelinde duraęan ya da duraęanlıęı sağlanmış bir zaman serisine ilişkin birçok olası model arasında uygun ARIMA modelinin belirlenmesi, parametrelerin hesaplanması, modelin uygunluęunun deęerlendirilmesi olmak üzere üç aşamadan oluşur. Bulunan modelin uygunluk testlerini geçememesi durumunda, süreç baştan başlatılarak uygunluk kriterlerinden en iyi dereceyi alan model nihai olarak seçilir ve tahminlemede bu model kullanılır (Hyndman, 2001, s.1). Box-Jenkins süreci Şekil 2.2’de gösterilmektedir.



Şekil 2.2. Box-Jenkins süreci
(Kaynak: Hyndman, 2001, s.1).

Genel olarak ARIMA modellerinin temsilinde kullanılan notasyon: $ARIMA(p,d,q)$ şeklindedir. Notasyonda p ve q sırasıyla ilgili modelin standart otoregresif ve standart hareketli ortalama derecelerini gösterirken, d serinin durağanlaştırılabilmesi için kaçınıcı dereceden standart farkının alınması gerektiğini göstermektedir.

Box-Jenkins metodolojisinde, uygun modelin belirlenmesi, ele alınan zaman serisinin örneklem otokorelasyon ve kısmi otokorelasyon katsayılarının artan gecikme uzunluklarında izledikleri seyirin, çeşitli derecelerden teorik $AR(p)$, $MA(q)$ ve $ARMA(p,q)$ modelleriyle karşılaştırılmasına dayanır (Hanke vd., 2001, s.408). Bu modellerin parametrelerinin hesaplanmasında doğrusal olmayan en küçük kareler yöntemi veya maksimum olasılıklar yöntemi kullanılabilir.

Box-Jenkins model belirleme yaklaşımı, salt AR(p) veya MA(q) süreçlerini teşhis etmede kolaylık sağlamakla beraber, ileri düzey AR, MA ve özellikle ARIMA süreçlerinin bu yöntemle teşhisi çok karmaşıklaşabilmektedir. Her ne kadar ekonometri yazılımları, tüm hesaplamaları yapmaktaysa da bulunan katsayıların ve korelogramların yorumlanması, çalışmayı yürüten araştırmacıya kalmaktadır bu bağlamda model belirleme prosedürü araştırmacının bilgi ve tecrübesine bağlı olarak belli ölçüde sübjektiflik taşımaktadır (Quinn vd., 1998, s.16). Fakat model seçimindeki bu sübjektiflik, model belirleme sürecinde Akaike Information Criterion (AIC) veya Schwarz Information Criterion (SIC) gibi genel kabul gören objektif model seçim kriterlerine de başvurmak suretiyle belli ölçüde giderilebilir.

Box-Jenkins yöntemleri, incelenen zaman serilerinin (stokastik süreçlerin) durağan olup olmaması durumuna göre “doğrusal durağan stokastik yöntemler” ve “durağan olmayan doğrusal stokastik yöntemler” olarak iki temel sınıfa ayrılır. Durağan olmayan stokastik süreçlerin modellerle gösterilebilmesi, seride durağanlığın sağlanmasına bağlı olmaktadır. Durağanlık, serinin uygun derecede farkları alınarak elde edilebilir. Durağan olmayan serideki verilere uyumu sağlanmaya çalışılan verilerin farklarının uyumunun söz konusu olduğu durağan modellere “bütünleşik” ya da “entegre (integrated)” modeller denilmektedir (Box ve Jenkins, 1976).

Otoregresif entegre hareketli ortalama (ARIMA) modelleri olarak bilinen durağan olmayan doğrusal stokastik modeller, zaman serilerinin mevsim unsurunu içerip içermemesi durumuna göre “mevsimsel ARIMA” ve “mevsimsel olmayan ARIMA” modelleri olarak da ayrıca sınıflandırılmaktadır.

Mevsimsel Olmayan Modeller

Mevsimsellik içermeyen modeller serinin durağanlığına göre aşağıda gösterilmiştir:

Doğrusal durağan stokastik yöntemler:

- Otoregresif (AutoRegressive) - AR(p)
- Hareketli Ortalama (Moving Average) - MA(q)
- Otoregresif Hareketli Ortalama (AutoRegressive Moving Average) - ARMA(p,q)

Doğrusal durağan olmayan stokastik yöntemler:

- Entegre Hareketli Ortalama (Integrated Moving Average) - IMA(0,d,q)
- Otoregresif Entegre Hareketli Ortalama (AutoRegressive Integrated Moving Average) - ARIMA(p,d,q)

Uygulamalarda en sık kullanılan Box Jenkins modelleri şunlardır:

- ARIMA(0,1,1) veya IMA(0,1,1) Modeli
- ARIMA(0,2,2) veya IMA(0,2,2) Modeli
- ARIMA(1,1,1)

Mevsimsel Modeller

Daha önceden de belirtildiği gibi, zaman serilerinin belirli dönemlerine ait (genelde aylık, üç aylık veya haftalık) gözlem değerlerinin her yılın aynı dönemlerinde benzer hareketler göstermesi eğilimi mevsim dalgalanmalarını ifade etmektedir. Genelde doğal veya sosyal nedenler sonucu ortaya çıkan bu dalgalanmaları içeren seriler “mevsimsel zaman serileri” olarak adlandırılmaktadır.

Uygulamalarda ele alınan birçok zaman serisi genellikle mevsimsel yapıdadır. Örneğin tarım ürünlerinin ya da mevsime özel ürünlerin satış miktarı (dondurma, salep gibi), belirli yörelere gelen turist sayısı, vb. değerlerden oluşan serilerde yılın belirli aylarında artışlar ya da azalışlar görülmektedir.

Mevsimsel zaman serisinin tahmini için belirlenen model belli bir trendle veri düzeyindeki değişmelerin yanında mevsimlerin etkisiyle oluşacak değişmeleri yansıtan parametreleri içermelidir. Durağan olmayan modellere mevsimsel periyodu içeren parametreler eklenerek kurulan mevsimsel model yardımıyla tahminlemeler yapılmalıdır. Mevsimsellik zaman serilerinin durağanlığını bozan unsurlardan birisidir, bu serilerde durağanlığın sağlanması için serinin mevsim etkisinden arındırılması gerekir. Trendin bulunduğu serilerde otokorelasyon fonksiyonundaki gerçek değerler ortaya çıkarılmayacağı için mevsimsel etkiler de durağan serilerdeki gibi açıklıkla görülemeyebilir. Yapılacak işlem durağan olmayan serileri, farkları alınarak durağan hale

getirmektedir. Seri durağan hale dönüştürüldükten sonra mevsimsel etki ortaya çıkar. Fark serisinin otokorelasyonları mevsimsel karakter taşıyorsa serinin mevsimsel zaman serisi olduğuna karar verilir (Kendal vd., 1983, s. 506-507).

Mevsimsel dalgalanmaların dalga uzunluğu s ile gösterilir, aylık gözlem değerlerinden meydana gelen serilerde tekrar periyodu genellikle 1 yıldır ($s=12$). Ancak bazen haftalık, 3 aylık ya da 6 aylık ($s=6$) periyoda sahip mevsimsel dalgalanmalara da rastlanabilir.

Mevsimsel modellerin derecelerinin gösteriminde kullanılan notasyon (p,d,q) X (P,D,Q) s şeklinde olup (p,d,q) mevsimsel olmayan modelin derecelerini, (P,D,Q) ise mevsimsel modelin dereceleri ve s de daha önce belirtildiği gibi periyodu, yani devir süresini göstermektedir. Daha açık ifadeyle, p otoregresif mevsimsel olmayan modelin derecesini, P otoregresif mevsimsel model derecesini, q mevsimsel olmayan hareketli ortalama derecesini, Q mevsimsel hareketli ortalama derecesini, d zaman serisinin belli bir periyoda göre alınmış farklarını, D ise serinin mevsimlere göre alınmış farklarını göstermektedir.

Mevsimsel modellerin belirlenmesi, parametrelerin tahmini, modelin uygunluk testi gibi işlemler mevsimsel olmayan modellerde izlenen aşamalar uygulanarak yapılır.

Uygulamada en sık karşılaşılan ve iktisadi olaylarla ilgili zaman serilerinin tahmininde kullanılan Box-Jenkins modeli, $(0,1,1)X(0,1,1)_{12}$ dir.

Box-Jenkins yöntemlerinin üstünlükleri aşağıdaki gibi sıralanabilir (Newbold ve Granger, 1975, s. 97; Chatfield ve Prothero, 1973, s. 296):

- Zaman serilerinin çoğunda mevcut olan iç bağımlılık en etkin şekilde bu yöntemde kullanılmaktadır.
- En uygun model genellikle eldeki verilerin yapısıyla belirlendiği için, “verilerin kendi kendine konuşması sağlanmış olur”. Bu nedenle diğer yöntemlere göre daha güvenilir sonuçlara ulaşılmaktadır.
- Sürecin her aşamasında modelin analiz edilecek seriye uygunluğunu denetleme imkânı vardır.

Diğer yandan bu yöntemin zayıf yanları da bulunmaktadır (Newbold ve Granger, 1974,

s. 133):

- Yinelemeli hesaplama yöntemleri nedeniyle, modelin üretim süresi uzundur.
- Modelin oluşturulmasında bilgisayar yazılımları kullanılsa dahi, yöntemin karışıklığından dolayı yazılımın kullanımında elle müdahaleye gerek duyulmaktadır. Bu nedenle deneyimli ve bilgili işgücüne gereksinim vardır, bu da maliyeti artırmaktadır.
- En uygun modelin belirlenmesi için çok sayıda (en az 50) gözlem değerine gereksinim vardır.

2.5. Yapay Sinir Ağları

Hem veri madenciliği hem de yapay zekâ yöntemleri içerisinde yer alan yapay sinir ağlarını incelemeden önce aşağıda yapay zekâ hakkında ön bilgi verilmiştir.

2.5.1. Yapay Zekâ Kavramı

Massachusetts Institute of Technology Bilgisayar Bilimleri laboratuvarı yöneticilerinden Edward Fredkin yapay zekânın önemini şu sözleriyle ifade etmektedir (Bilim ve Teknik, Aralık 2001): “Tarihte üç büyük olay vardır. Bunlardan ilki evrenin oluşumu, ikincisi hayatın başlangıcıdır. Bu ikisiyle aynı derecede önemli olan üçüncüsüyse, yapay zekânın ortaya çıkışıdır”

“Yapay zekâ” (Artificial Intelligence, AI), insanın zekâ davranışlarının otomasyonunu araştıran, bilgilerin saklanması ve işlenmesinde veri yapıları, algoritmalar, programlama dilleri ve yöntemleri gibi bilgi işlem yöntemlerini kullanan, bilgisayar biliminin bir alt dalıdır (Akpınar, 1994, s.43).

1950’ lerin ikinci çeyreğinde John McCarthy tarafından ortaya atılan “Yapay Zekâ” kavramı önceleri beyine kıyasla çok yavaş ve basit kalmıştır. Ancak elektronik çağın etkisiyle hızla ilerleyen tıp teknolojisi sayesinde beynin ayrıntılı incelemesinden elde edilen sonuçlar, yine çok hızlı ilerleyen bilgi teknolojisi sayesinde yüksek kapasiteli ve hızlı bilgisayarlarda yapay zekâyı uygulama imkânını doğurmuştur. İdeal olarak yapay zekâ çalışmalarının hedefi insan beyninin çalışmalarını taklit etmek, daha açık bir ifadeyle bir makinenin ya da bilgisayarın insan gibi düşünmesini, karar vermesini, tahminde

bulunmasını ve insanımsı davranışlar sergilemesini sağlayabilmektir. İlk yapay zekâ çalışmaları oyun oynamada ve teorem ispatlamada kullanılmıştır. Satranç gibi insanların zekâsını çalıştıran oyunlarda bilgisayarlar çok sayıda hamle ihtimalini (çözüm yolunu) inceleyerek içlerinden en iyisini seçmek suretiyle çok hızlı bir şekilde çözüme ulaşmaktadır (Macy ve Pandya, 1995; Öztemel, 2003).

Yapay bir sistemin insanı andıran zekice davranışlar sergileyebilmesi için en azından aşağıdaki temel bileşenlere sahip olması gerekmektedir (Öztemel, 2003):

- Algılama
- Kavrama
- Eylem

Yapay ve doğal zekânın özellikleri Tablo 2.1’ de karşılaştırmalı olarak gösterilmiştir.

Tablo 2.1. Yapay zekâ ile doğal zekânın karşılaştırılması
(Kaynak: Nabiyev, 2003; Web-Yapayzeka)

| Özellik | Doğal Zekâ | Yapay Zekâ |
|----------------|--|---|
| Kalıcılık | Kalıcı değildir, bilgiler zamanla unutulmaktadır. | Kalıcıdır. Silinmediği ya da fiziksel hasar görmediği sürece her bilgi aynen korunur. |
| Kopyalanabilme | Bilgilerin başkasına aktarılması zor olmakta ve uzun bir süreç gerektirebilmektedir. | Bilgiler sayısal olduğu için hiç değişmeden tümü kolaylıkla başa yerlere aktarılabilir. |
| Maliyet | Bilgilerin başkasına aktarılması zaman ve maliyet gerektirmektedir. | Sayısal bilginin ilk üretim maliyeti yüksek olsa da kopyalanma maliyeti çok düşüktür. |
| Tutarlılık | Kararlar kişinin o anki psikolojisine bağlı olarak bile değişebilir. | Tutarlıdır, aynı koşullarda sonuç hep aynı olur. |
| Yaratıcılık | Mevcut bilgilerden yararlanarak başka bilgileri keşfedebilme ve yeni fikirler üretebilme ve öğrenebilme yeteneği vardır. | Yoktur. Ne öğrendiyse, parametreler ne ise sadece ona göre yanıt verir. |
| Muhakeme Gücü | Bir konudaki problemi çözmek için başka konulardaki tecrübelerinden de yararlanabilir. | Öğrendiği kadarı ile dar bir alanda çalışır. Muhakeme gücü yoktur. |

Genel olarak en bilinen yapay zekâ teknolojileri şu şekilde sıralanabilir:

- Genetik Algoritmalar (Genetic Algorithms)
- Uzman Sistemler (Expert Systems)
- Bulanık Mantık (Fuzzy Logic)
- Robotik (Robotics)
- Doğal Dil İşleme (Natural Language Processing)
- Yapay Sinir Ağları (Artificial Neural Networks)

2.5.2. Tanımı

Yapay Sinir Ağları konusunda çok sayıda tanım bulunmaktadır. Aşağıda, bu tanımlar içerisinde en önemli görülenlerine yer verilmiştir.

Yapay sinir ağları, karmaşık yapıda içsel ilişkilere sahip ya da arasındaki ilişkiler bilinmeyen problemlere çözüm bulabilmek için geliştirilen bir yapay zekâ tekniğidir (Çakar vd., 1996, s.77).

Haykin (1999), yapay sinir ağları için şu tanımları kullanmaktadır: “Bir sinir ağı, basit işlem birimlerinden oluşan, deneyimsel bilgileri biriktirmeye yönelik doğal bir eğilimi olan ve bunların kullanılmasını sağlayan yoğun bir şekilde paralel dağıtılmış bir işlemcidir. Bu işlemci iki şekilde beyin ile benzerlik göstermektedir: 1. Bilgi, ağ tarafından bir öğrenme süreciyle çevreden elde edilir. 2. Elde edilen bilgileri biriktirmek için sinaptik ağırlıklar olarak da bilinen nöronlar arası bağlantı güçleri kullanılır.”

Yapay sinir ağlarını bilgisayar yazılımları ile bütünleştiren Anderson ve Mc Neill (1992, s.4) ise yapay sinir ağlarını “bir örnekler kümesi yardımıyla parametrelerin uyarlanabilmesini sağlayacak bir matematiksel formül için yazılan bilgisayar yazılımı” olarak tanımlamaktadır.

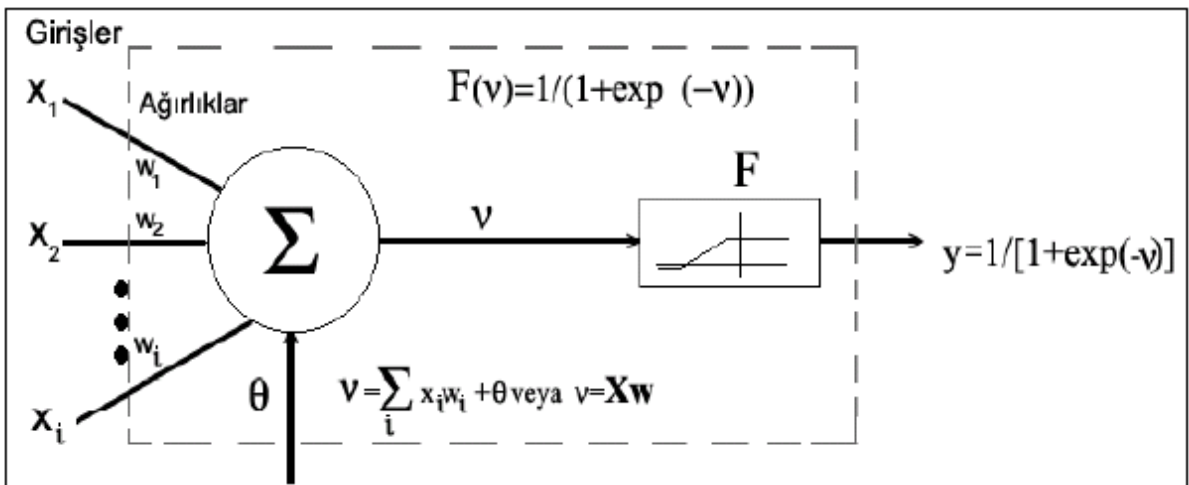
2.5.3. Yapısı

Bir yapay sinir ağı, kendisini oluşturan yapay sinir hücrelerinin birbirleriyle çeşitli katmanlar içerisinde paralel bağlantılar kurarak bir araya gelmelerinden oluşmaktadır. Bu yapıda yer alan yapay sinir hücreleri, katmanlar ve süreçler aşağıda açıklanmaktadır.

2.5.3.1. Yapay Sinir Hücresi

Bilindiği üzere, insan sinir sisteminin temel taşı oluşturulan ve nöron olarak adlandırılan beyin hücrelerinden bir insan beyninde yaklaşık olarak 10^{11} adet bulunmaktadır. Bu nöronlar arasında ise yaklaşık olarak 10^{15} adet bağlantı bulunduğu düşünülmektedir. Bir beyin hücresi dört elemandan oluşur; gövde, çekirdek, dendrit ve akson. Çekirdek, temel işlem elemanıdır, kendisine gelen sinyalleri (girdileri) işleyerek bir çıktıya dönüştürür. Dendrit ve akson, iletişimi sağlayan uzantılardır. Dendrit kısa olup binlerce daldan oluşmaktadır, görevi diğer nöronlardan gelen sinyalleri toplamaktır. Akson ise uzun olup tek parçadan oluşmaktadır, görevi çıktı bilgilerini diğer sinir hücrelerine taşımaktır. Bir nöronun dendritiyle başka bir nöronun birleştiği yer ise “sinaps” olarak adlandırılmaktadır. Biyolojik bir sinir hücresi diğer hücrelerden gelen girdileri toplayıp çeşitli fonksiyonlar aracılığıyla işledikten sonra çıktı değeri üretir (Kantardzic, 2003; Web-Wikipedia5).

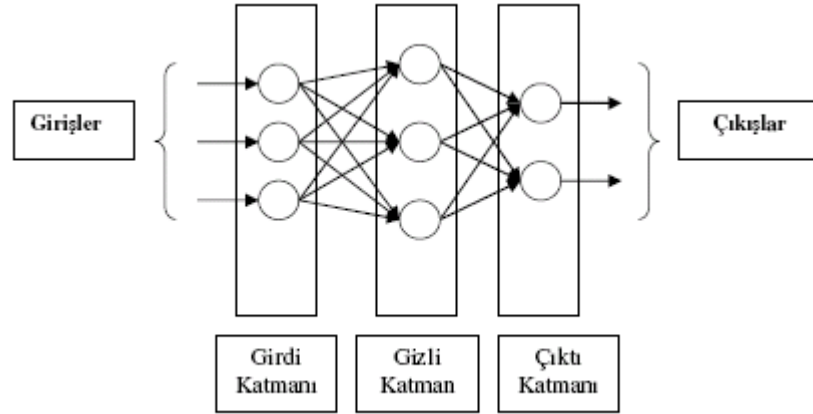
Biyolojik sinir sisteminin bu doğal yapısı, yapay sinir ağlarında matematiksel modellerle oluşturulmaya çalışılmıştır. Bu şekilde modellenen bir yapay sinir hücresi Şekil 2.3’ de görülmektedir (Aydın, 2005). Yapay bir nörona; girdiler (X_1, X_2, \dots, X_n) kendilerine ait ağırlıklarla çarpılarak toplandıktan sonra aktivasyon fonksiyonuna gönderilirler. Girdilerin ağırlık değerlerinin pozitif veya negatif olması, ağa etkilerinin pozitif veya negatif yönde olduğunu gösterir (Öztemel, 2003). Aktivasyon fonksiyonuna gelen sinyal, burada işlenerek çıktıya dönüştürülür.



Şekil 2.3. Yapay sinir hücresi
(Kaynak: Aydın, 2005)

2.5.3.2. Katmanları

Yapay sinir hücreleri, katmanlar içerisinde paralel şekilde bir araya gelerek yapay sinir ağını meydana getirmektedir. Yapay sinir ağlarında üç ana katman bulunmaktadır: girdi katmanı, gizli katman ve çıktı katmanı. Girdilerden oluşan girdi katmanı ile çıktılarından oluşan çıktı katmanı tektir. Gizli katman sayısı ise yapay sinir ağının yapısına göre bir veya birden çok olabilmektedir. Girdi katmanından alınan girişler, girdi katmanı ve gizli katman arasında bulunan bağlantı ağırlıkları ile çarpılıp gizli katmana iletilmektedir. Gizli katmandaki nöronlara gelen girişler toplanarak aynı şekilde gizli katman ile çıktı katmanı arasındaki bağlantı ağırlıkları ile çarpılarak çıktı katmanına iletilir. Çıktı katmanındaki nöronlar da, kendisine gelen bu girişleri toplayarak uygun bir çıkış üretirler (Efe ve Kaynak, 2000, s.8). Bağlantıların ağırlık değerleri öğrenme sürecinde belirlenmektedir. Şekil 2.4’ de bir yapay sinir ağını oluşturan katmanlar görülmektedir.



Şekil 2.4. Çok katmanlı yapay sinir ağı yapısı
(Kaynak: Fırat ve Güngör, 2005, s.61)

2.5.4. Süreçleri

Yapay sinir ağları aşağıda listelenen iki temel süreçten oluşmaktadır (Hosein, 2002):

- İşlem süreci
- Eğitim ve öğrenim süreci

2.5.4.1. İşlem Süreci

Bu süreçte toplama ve aktivasyon adlarında iki fonksiyon yer almaktadır.

Toplama (Birleştirme) Fonksiyonu

Yapay sinir hücresine gelen girdilerin kendilerine ait ağırlıklarla çarpıldıktan sonra birleştirilmesi işlemini gerçekleştiren fonksiyondur. Bu fonksiyon, adından da anlaşılacağı gibi, genelde toplama işlemini kullanmakla birlikte farklı işlemleri de kullanılabilir. Hatta araştırmacının kendi kurduğu işlemi de kullanması mümkündür. Toplama fonksiyonunda kullanılan işlem, genellikle seçilen ağ mimarisine de bağlıdır. Sık kullanılan işlemler Tablo 2.2’ de açıklamalı olarak verilmiştir (Öztemel, 2003).

Tablo 2.2. Toplama fonksiyonu işlemleri

| Fonksiyon | Matematiksel Gösterimi | Açıklaması |
|------------------|-------------------------------|---|
| Toplama | $\sum_i^n X_i W_i$ | Girdiler kendilerine ait ağırlıklar ile çarpılıp daha sonra hepsi toplanır. |
| Çarpım | $\prod_i X_i W_i$ | Girdiler kendilerine ait ağırlıklar ile çarpılıp daha sonra da birbirleri ile çarpılır. |
| Maksimum | $Max(X_i W_i), i = 1 \dots j$ | Tüm girdiler kendilerine ait ağırlıklar ile çarpıldıktan sonra aralarından en büyük değer alınır. |
| Minimum | $Min(X_i W_i), i = 1 \dots j$ | Tüm girdiler kendilerine ait ağırlıklar ile çarpıldıktan sonra aralarından en küçük değer alınır. |
| Çoğunluk | $\sum_i \text{sgn}(X_i W_i)$ | Tüm girdiler kendilerine ait ağırlıklar ile çarpıldıktan sonra pozitif ve negatif olanların sayısı bulunur. Hangisi daha çok ise hücrenin net girdisi o kabul edilir. |
| Kümülatif Toplam | $Net(eski) + \sum_i X_i W_i$ | Hücreye uygulanan tüm girdiler toplanır ve daha önceki toplama eklenir. |

Aktivasyon Fonksiyonu

Yapay sinir hücresinin çıktısının büyüklüğünü sınırlandıran fonksiyondur. Bazı kaynaklarda transfer, eşik veya sıkıştırma fonksiyonu olarak da isimlendirilmektedir (Mandic ve Chambers, 2001). Bir ağdaki tüm hücrelerin aktivasyon fonksiyonu birbirinden farklı olabilir. Aktivasyon fonksiyonunda doğrusal fonksiyonlar genelde tercih edilmez. Zaman serileri için “Sigmoid”, ikili (binary) değişkenler için “Adım” fonksiyonu önerilmektedir (Tebelkis, 1995). En çok kullanılan aktivasyon fonksiyonları aşağıda gösterilmiştir (Demuth ve Beale, 2000; Şen, 2004):

Tablo 2.3. Aktivasyon fonksiyonları

| Fonksiyon | Matematiksel Gösterimi | Açıklaması |
|-------------------------------|---|---|
| Doğrusal Fonksiyon | $F(x) = x$ | Hücreye gelen veriler bir katsayı ile çarpılarak sonuç üretilir. Katsayı 1 ise girdiler olduğu gibi çıkar. |
| Adım Fonksiyonu | $F(x) = 1$ eğer $x >$ eşik değeri 0 diğer durumlarda | Gelen veri belirlenen bir eşik değerinin üstünde ya da altında olmasına göre 1 ya da 0 değerlerini alır. |
| Esik değer Fonksiyonu | $F(x) = 0$ eğer $x \leq$ eşik değeri x eğer $0 < x < 1$ 1 eğer $x \geq 1$ | Gelen veriler belirlenen eşik değerleri arasında ise gelen değer kendisi çıktı olur. Veriler eşik değerinden büyük ya da küçük olmasına göre 1 ya da 0 değerlerini alırlar. |
| Hiperbolik Tanjant Fonksiyonu | $F(x) = (e^x - e^{-x}) / (e^x + e^{-x})$ | Çıktı değeri hücreye gelen verilerin tanjant fonksiyonuna tabi tutulmasıyla hesaplanır. |
| Sigmoid Fonksiyonu | $F(x) = 1 / (1 + e^{-x})$ | Yapay sinir ağları oluşturulurken en çok kullanılan aktivasyon fonksiyonudur. Doğrusal ve doğrusal olmayan davranışlar arasında denge sağlayan sürekli artan bir fonksiyon olarak tanımlanır. |

2.5.4.2. Eğitim ve Öğrenim Süreci

Yapay sinir ağları uygulamalarındaki en önemli süreçlerden birisi olan Eğitim ve Öğrenimin temel amacı girdi ve çıktı arasında bir eşleştirme yapmaktır. Ağ yeterince eğitildiğinde daha önceden görmediği girdilere karşı uygun çıktılar üretir. Bu özellik yapay sinir ağlarının genelleme kapasitesi olarak adlandırılır ve çoğunlukla ağın yapısına ve eğitim için uygulanan yineleme sayısına bağlıdır (Alkan, 2001).

Yapay sinir ağlarının veriyi öğrenebilmeleri (eğitimi) için çeşitli öğrenme algoritmaları geliştirilmiştir. Bu algoritmalar ağ modeline ve problemin yapısına bağlı olarak üç ana gruba ayrılır: danışmansız (denetimsiz/ öğreticisiz), danışmanlı (denetimli/öğreticili) ve takviyeli (destekli) öğrenme.

Danışmansız Öğrenme: Çıktıların istenen değerlerinin ağa tanıtılmadığı öğrenme şeklidir. Ağa tek başına uygulanan girdiler aynı zamanda çıktı görevi görmektedir. Girdi değerlerine uygun bir çıktı üretilinceye kadar bağlantı ağırlıkları değiştirilir. Bunun için bir fonksiyon kullanılmaz. Yapay sinir ağları kendi kendine gerçek verilerle çıktı arasındaki farkı en aza indirmeye çalışır. Öğreticisiz öğrenme algoritmaları daha çok sistemin geçmişte karşılaştığı veri kümesinin içerdiği istatistiksel bilgilerin elde edilmesini amaçlar.

“Adaptif Rezonans Teorisi (ART)” ve “Kohonen Ağı” en çok kullanılanlarıdır. Çok elemanlı veri kümeleri içerisinde deneme-yanılma yoluyla bilgi genelleştirilmesi yapılabilir (Efe ve Kaynak, 2000, s.1-25).

Danışmanlı Öğrenme: Çıktıların istenen değerlerinin ağa tanıtılabildiği öğrenme şeklidir. Ağa hem girdi, hem de çıktı kümeleri verilir. Ağırlıkları değiştiren bir fonksiyon modelin ne üretmesi gerektiğini önceden modele bildirir. Ağ, girdiyi işleyerek kendi çıktısını üretir ve gerçek çıktı ile karşılaştırır. Öğrenme metodu sayesinde var olan hata en aza indirilene kadar ağırlıklar değiştirilir. Bu işlem kabul edilebilir bir hata seviyesine ulaşıncaya kadar devam eder. “Geriye Yayılım” ve “Boltzmann” en yaygın kullanılan danışmanlı ağlardır (Anderson ve Mc Neill, 1992, s. 2-38).

Takviyeli öğrenme: Takviyeli öğrenmede giriş değerlerine karşılık gelecek uygun çıktıların elde edilmesi sırasında ağırlıklar optimize edilmeye çalışılır; fakat çıktının ne olması gerektiği önceden belli değildir. Bunun yerine çıktının ne ölçüde doğru olduğunu gösteren dereceler bildirilir. Öğrenme kuralı giriş işaretlerine ve aktivasyon fonksiyonu tarafından sağlanan değerlere cevap olarak, başlangıçtaki ağırlıkların hepsini veya bazılarını değiştiren bir denklemdir. Bu şekilde ağ, kendisini istenen cevaplara uyarlayabilir ve kendi içinde bilgiyi düzenleyebilir, yani öğrenir. En yaygın kullanılan takviyeli öğrenme ağı “Radyal Tabanlı Ağ” dır.

Öğrenme Algoritmaları

Yapay sinir ağlarının en önemli özelliklerinden birisi, ilgili probleme ait örneklerle veriler arasındaki ilişkiyi öğrenebilmesidir. Yapay sinir ağlarının, veri yapısındaki ilişkiyi öğrenmesi, probleme ait örnekler yardımı ile ağ ağırlıklarının en uygun değerlerinin belirlenmesine dayanır. Herhangi bir ağırlığın yeni değeri, eski değeri ile ağırlık değerlerinin değişim miktarının (ΔW) toplamı olarak belirlenir ($W_{\text{yeni}} = W_{\text{eski}} + \Delta W$). Belli bir kurala göre hesaplanan bu değişim miktarını (ΔW) belirlemek için tanımlanmış kurallara “öğrenme algoritmaları” denir.

Öğrenme prosedüründe iki kademe vardır (Rumelhart vd, 1986):

- Ağın gerçek çıktıları veya aktivasyon seviyelerini hesaplamak (ileri hesaplama)
- Ağın gerçek ve istenen çıktıları arasındaki hatayı geriye doğru yaymak (geriye

hesaplama)

Yapay sinir ağlarının eğitimi için birçok değişik optimizasyon yöntemi bulunmakla birlikte genel bir doğrusal olmayan optimizasyon problemi için uygun bir zamanda genel optimumu garanti eden bir algoritma bulunmamaktadır. Yapay sinir ağlarının eğitiminde hatayı en aza indirmek için genellikle, hata fonksiyonunun yönünü bulmaya ve hata fonksiyonunu azaltmaya çalışan dereceli azaltma (gradient descent) tabanlı algoritmalar kullanılır. Dereceli azaltma algoritmaları içerisinde en bilineni, uygulamalarda en çok tercih edilen “Geri Yayılım Algoritması (GYA)” dır (Auclair, 2004).

GYA’ da üç önemli faktör bulunmaktadır; öğrenme oranı, momentum katsayısı ve bias elemanı (Web-Mekatronikkulubu):

Öğrenme oranı; bağlantıların ağırlık değerlerindeki değişimin miktarını tanımlar, bu nedenle geri yayılım öğrenme algoritmasında çok önemlidir. Küçük öğrenme oranı, değişim miktarını azaltarak öğrenme sürecinin yavaş ilerlemesine sebep olur. Büyük öğrenme oranı ise değişimin miktarını artırarak ağ ağırlıklarının çok fazla salınım göstermesine yol açar.

Momentum katsayısı; ağı daha hızlı toparlanmasına yardım eden bir faktördür. Temel amaç, daha önceki değişimin bir kısmını mevcut değişime ekleyerek öğrenme esnasında ağı salınımını (oscillation) önlemektir. Momentum katsayısı, büyük öğrenme oranıyla eğitim sürecinin hızlanmasını sağlarken, salınma eğilimini de en aza indirmeye yardımcı olur.

Bias elemanı; geriye yayma ağlarında daha iyi öğrenme sağlamak için aktivasyon fonksiyonunun orijinini dengeye getirmeye çalışan bir özel işlem girdi elemanıdır.

YSA’ da öğrenme başlamadan önce öğrenme oranı ve momentum katsayısı gibi bazı parametrelerin değerleri belirlenmelidir. Ağı bağlantılarının başlangıç ağırlıkları genellikle rastgele atanır.

Yapay sinir ağlarının eğitiminde çok sayıda öğrenme algoritması geliştirilmiştir. Tablo 2.4’ de bu algoritmalar ve kullanım yerleri görülmektedir.

Tablo 2.4. Öğrenme algoritmaları
(Kaynak: Anil vd., 1996, s.38)

| Öğrenme Yöntemi | Öğrenme Kuralı | Ağın Mimarisi | Öğrenme Algoritması | Kullanıldığı Yerler |
|-----------------|----------------------------|------------------------------|---|--|
| Danışmanlı | Hata Düzeltme | Tek ya da çok katmanlı | Perceptron Geriye Yayılım Adaline ve Madaline | Örüntü tanıma Tahmin Kontrol |
| | Boltzmann | Geri Dönüştü | Boltzmann öğrenmesi | Örüntü sınıflandırma |
| | Hebbian | Çok katmanlı ileri beslemeli | Lineer diskriminant | Veri analizi Örüntü sınıflandırma |
| | Rekabetli | Rekabetli | Vektör kuantalama öğrenmesi | Veri sıkıştırma Sınıflandırma |
| ART ağı | | ART Haritalaması | Örüntü sınıflandırma Sınıflandırma | |
| Danışmansız | Hata Düzeltme | Çok katmanlı ileri beslemeli | Sammon projeksiyonu | Veri analizi |
| | Hebbian | İleri beslemeli rekabetçi | Principal Component Analizi | Veri analizi Veri sıkıştırma |
| | | Hopfield Ağı | Associative Memory Öğrenmesi | Associative Memory |
| | Rekabetli | Rekabetli | Vektör Kuantalaması | Sınıflandırma Veri sıkıştırma |
| | | Kohonen SOM | Kohonen SOM | Sınıflandırma Veri analizi |
| | | ART ağı | ART1, ART2 | Sınıflandırma |
| Takviyeli | Hata Düzeltme ve Rekabetli | Radyal Tabanlı | Radyal Tabanlı Öğrenme | Sınıflandırma Örüntü tanıma Fonksiyon yaklaşımı Tahmin Kontrol |

Perceptron geri yayılım algoritması ile bunun geliştirilmiş bir versiyonu olan Levenberg- Marquardt algoritması günümüzde en yaygın kullanılan algoritmalar arasında yer almaktadır. Aşağıda bu iki algoritma hakkında özet bilgi sunulmuştur.

Geri Yayılım Algoritması

Danışmanlı öğrenen, ileri beslemeli bir yapay sinir ağı olan Geri Yayılım (Back Propagation) Algoritması, adını hatayı yayma biçiminden alır. Hata azaltma işlemini hatayı tüm ağa yayarak gerçekleştirdiği için bu ağa “Hatayı Geriye Yayma – Error Back Propagation” da denmektedir. Bu algorithmada elde edilen çıktı ile olması gereken çıktı arasındaki fark yani hata, tüm ağırlıklara yansıtılarak dereceli olarak minimum düzeye indirilmeye çalışılır (Aydoğmuş ve Çöteli, 2005, s.239-246). Geri yayılım algoritmasında eğitime rastgele bir ağırlık kümesi ile başlanır, birçok uygulamada ağırlık başlatma ağı atanan ilk ağırlık değerlerinin uygun seçilmesine bağlıdır (Nabiyev, 2003).

GYA, öğrenmesi ve uygulaması kolay bir modeldir. Günümüzdeki uygulamaların yaklaşık % 80’inde kullanıldığı tahmin edilen bu ağ, özellikle mühendislik alanında neredeyse her türlü problemin çözümü için kullanılabilecek kadar güçlüdür (Kaastra ve Boyd, 1996). Finansal zaman serilerinin gelecekteki davranışlarını tahmin etmede de sıklıkla tercih edilmektedir.

GYA, çok güçlü olmasına karşın bazı zayıf yanlara da sahiptir. Yavaş olması ve basit bir problemin çözümünde bile yapay sinir ağının eğitiminin binlerce yineleme gerektirmesi bu algoritmanın en belirgin dezavantajlarıdır.

Levenberg-Marquardt Algoritması

Geleneksel geri yayılım algoritmasının dezavantajlarını gidermek amacıyla bazı varyasyonları ve modifikasyonları önerilmiştir. Bunlar arasında en etkili olanı, ikinci derece yöntemlerden olan Levenberg-Marquardt algoritmasıdır. Bu algorithmada minimum hatayı bulmak için her bir iterasyon adımında hata yüzeyine parabolik şekilde yaklaşılır ve parabolün minimumu o adım için çözümü oluşturur. Hızlı çalışması ve en iyi yerel minimumu bulmadaki başarısı nedeniyle günümüzde çok yaygın kullanılır hale gelmiştir (Zhang vd., 1998, s.35-62).

2.5.5. Mimarisi

Yapay sinir ağı mimarileri için belirlenmiş standart bir sınıflandırma yoktur. Nöronların bağlanma şekillerine, kullanılan öğrenme algoritmasına, zaman gecikmesine ve

işlenilen verinin türüne göre farklı sınıflandırmalar yapılmaktadır (Kim, 2003).

En belirgin olan ve yaygın olarak kullanılan sınıflandırma nöronların bağlanma biçimlerine göre yapılmaktadır. Buna göre yapay sinir ağları “ileri beslemeli” ve “geri beslemeli” olmak üzere ikiye ayrılmaktadır (Slaughter, 2003):

İleri beslemeli ağlar: Verilerin sadece girdi birimlerinden çıktı birimlerine ileri doğru aktığı ağ yapısıdır. Bu yapıda nöronlar katmanlar şeklinde düzenlenir. Bir katmandaki nöronların çıkışları bir sonraki katmana ağırlıklar üzerinden giriş olarak verilir. Aynı katmandaki nöronlar arasında veya bir önceki katmana bağlantı yani geri besleme çevrimi yoktur. Uygulamalarda genellikle ileri beslemeli ağların tercih edildiği görülmektedir.

Geri beslemeli ağlar: Veri akışının sadece ileriye doğru değil geriye doğru da olabileceği ağ yapısıdır. Bu yapıda en az bir tane geri besleme çevrimi bulunur. Geri besleme, aynı katmandaki hücreler arasında olabileceği gibi farklı katmanlardaki nöronlar arasında da olabilir.

2.5.6. Uygulama Alanları

Yapay sinir ağları gerçek hayatta karşılaşılan problemlerde oldukça geniş bir uygulama alanı kazanmıştır. Günümüzde birçok endüstride başarılı şekilde kullanılmaktadırlar. İstatistiksel yöntemlerle optimum çözüme ulaşılamayan veya değişkenler arasındaki ilişkinin doğrusal olmadığı problemlerde özellikle tercih edilmektedirler. Uygulama alanları için bir sınır yoktur; fakat özellikle aşağıdaki alanlarda yaygın olarak kullanıldığı söylenebilir (Özgen, 2007; Yurtoğlu, 2005, s. 9):

- Tahminleme
- Öğrenme
- Genelleme
- İlişkilendirme
- Sınıflandırma
- Özellik belirleme
- Optimizasyon

Yapay sinir ağıları son yıllarda zaman serilerinin analizinde ve gelecekteki değerinin tahmininde sıklıkla kullanılan bir yöntem olmuştur. Çoğunlukla doğrusal bir modelle ifade edilemeyen zaman serileri, yapay sinir ağlarının doğrusal olmayan modellenmesi sayesinde başarıyla tahminlenebilmektedir (Lawrance, 1997).

2.5.7. Üstünlükleri ve Eksikleri

Yapay sinir ağı modelleri biyolojik sinir ağlarının çalışma biçimlerinden esinlenerek ortaya çıkarılmıştır. Yapay sinir ağları özellikle doğrusal olmayan sistemlerde tahminler açısından istatistik tekniklere göre daha kolaylık sağlayan bir çok üstün özelliğe sahiptir. Bu özelliklerden önemli görülenleri aşağıda listelenmiştir (Özgen, 2007).

- **Doğrusal Olmama:** Yapay sinir ağlarında kullanılan aktivasyon fonksiyonunun doğrusal olmaması nedeniyle, doğrusal olmayan problemlerin çözümünde başarıyla kullanılabilir.
- **Görev Paylaşımı:** Yapay sinir ağlarında her bir işlem birimi, çözülecek problemin tümü ile ilgilenmek yerine, sadece problemin gerekli parçası ile ilgilenmektedir. Bu görev paylaşımı sayesinde çok karmaşık problemler rahatlıkla çözülebilmektedir.
- **Hatayı Giderme:** Yapay sinir ağlarının herhangi bir elemanında meydana gelebilecek bir problem çok büyük önem arz etmeyebilir. Zira nöronların paralel çalışması nedeniyle yapay sinir ağlarının bazı bağlantılarının hatta bazı hücrelerinin etkisiz hale gelmesi, ağın doğru bilgi üretmesini önemli ölçüde etkilemez. Bu nedenle, diğer yöntemlere göre hatayı giderme yetenekleri yüksektir.
- **Öğrenebilirlik:** Yapay sinir ağları problemleri önbilgiye gereksinim duymadan sadece aldığı örneklerle çözebilir, yani örneklerden kendisi öğrenme yeteneğine sahiptir.
- **Uyarlanabilirlik:** Yapay sinir ağları ilgilendiği problemdeki değişikliklere göre ağırlıklarını ayarlar. Yani belirli bir problemi çözmek amacıyla eğitilen yapay sinir ağları, problemdeki değişimlere göre tekrar eğitilerek kolaylıkla uyarlanabilir.
- **Hız:** Nöronların paralel dizilişi, bilgilerin çok hızlı bir şekilde öğrenilmesini sağlamaktadır.
- **Veri kaybının olmaması:** Box-Jenkins modellerinde olduğu gibi seriyi durağanlaştırmak amacıyla fark alma gibi bir işleme gerek duyulmadığı için herhangi bir veri kaybı oluşmaz.

Yapay sinir ağlarının zayıf olduğu yanlar ise şunlardır (Özgen, 2007; Çerkez, 2003, s.77-78):

- **Kara kutu olması:** Ağın ağırlıklarının yorumlanması zordur. Doğrusal regresyon modelinde bağımsız değişkenlerin katsayıları, bağımlı değişken üzerindeki etkilerini açık olarak gösterirler. Ancak yapay sinir ağlarında, girdi değişkenlerinin çıktı üzerindeki etkilerini analitik olarak belirlemek çok zordur.
- **Yanlış sonuçlara yönlendirme ihtimali:** Yapay sinir ağları genellikle her eğitim kümesi için bir sonuç üretir, ancak üretilen sonucun doğru bir modeli yansıttığının garantisi yoktur. Yetersiz test kümesi ile elde edilen performans değerlendirmeleri modelin başarılı olduğunu gösterse bile elde edilen sonuçlar hatalı kararlara yönlendirebilir.
- **Genel minimumu bulamaması:** Bütün doğrusal olmayan tahmin yöntemlerinde olduğu gibi yapay sinir ağlarında da genel minimumu bulamama riski vardır. Ancak, genel minimuma yakın yerel minimumlar da oldukça iyi sonuçlar verebilir.
- **Çok sayıda örnek gerektirmesi:** Yapay sinir ağlarının tam olarak genelleştirme yapabilmesi için yeterli miktarda örnekle ağ eğitilmelidir. Aksi durumlarda veri yapısındaki ilişkiyi ortaya çıkartamama ya da yanlış eğitime (ezberleme) durumları ile karşılaşılabilir.
- **Uzun çalışma süresi:** Yapay sinir ağlarının eğitimi ve model üretimi uzun zaman alabilir. Ayrıca girdilerin normalleştirme süreci de ek bir süre gerektirmektedir.

2.5.8. Modelin Oluşturulması

Bir yapay sinir ağları modeli oluşturulurken belirlenmesi gereken önemli kriterler şunlardır:

- Verileri normalleştirme yöntemi
- Eğitim ve test kümelerinin oranı
- Ağ mimarisi
- Katman sayıları
- Nöron sayıları
- Aktivasyon fonksiyonu
- Öğrenim algoritması ve parametreleri
- Doğruluk ölçüleri

Bu kriterlerin belirlenmesi ilgilenilen probleme göre deęişiklik göstermektedir. Optimal aę mimarisinin belirlenmesinde önerilen bazı yöntemler olmasına rağmen, bu yöntemler oldukça karmaşıktırlar ve yerine getirilmeleri zordur. Ayrıca, bu yöntemlerden hiçbirisi gerçek tahmin problemleri için en uygun çözümü garanti etmemektedir. Bu parametrelerin optimal belirlenmesini sağlayacak açık yöntemler olmadığı için genelde sezgisel yaklaşımlar ve deneylere dayalı benzetim çalışmaları kullanılmaktadır. Bu nedenle, yapay sinir aęlarının tasarımı bir bilimden daha çok bir sanattır (Zhang vd., 1998).

2.5.8.1. Verileri Normalleştirme Yöntemi

Önceden de belirtildięi gibi, yapay sinir aęlarında genellikle doğrusal olmayan aktivasyon fonksiyonları tercih edilmektedir. Doğrusal olmayan aktivasyon fonksiyonları, nöronların çıktısını $[0,1]$ ya da $[-1,1]$ aralığına sıkıştırmaktadırlar. Bu nedenle, hesaplama hatalarından kaçınmak için, öncelikle girdilerin ve çıktıların aktivasyon fonksiyonuna baęlı olarak ilgili aralıkta ölçeklendirilmesi, yani normalleştirilmesi gerekmektedir. Bu sayede veriler uç deęerlerden arındırılarak daha iyi bir şekilde modellenabilmektedir. Normalleştirme için yazında deęişik yazarlar tarafından farklı normalleştirme formülleri önerilmiştir. Bunların içerisinde en yaygın kullanılanları şunlardır:

- $X_{norm} = X / X_{maks}$,
- $X_{norm} = (X - X_{min}) / (X_{maks} - X_{min})$,
- $X_{norm} = [2*(X - X_{min}) / (X_{maks} - X_{min})] - 1$ (düzeltilmiş normalleştirme).

Veri normalleştirme işlemi eğitim süreci başlamadan uygulanmalıdır.

2.5.8.2. Eğitim Kümesi ve Test Kümelerinin Oranı

Yapay sinir aęları modellerinin oluşturulmasında eğitim ve test örneklerine ihtiyaç vardır. Eğitim örnekleri, yapay sinir aęları modeli geliştirmek için kullanılırken, test örnekleri, geliştirilen modelin tahmin yeteneğinin deęerlendirilmesinde kullanılmaktadır. Eğitim ve test örneklerinin seçimi, yapay sinir aęlarının performansını etkilemektedir. Eğitim ve test kümelerinin uygunsuz belirlenmesi, optimal yapay sinir aęları yapısının seçimini ve yapay sinir aęlarının tahmin performansının deęerlendirilmesini etkileyecektir. Yazında eğitim ve test kümelerinin belirlenmesinde genelde % 90 - % 10, % 80 - % 20

veya % 70 - % 30 oranlarının tercih edildiği görülmektedir (Zhang vd., 1998).

2.5.8.3. Ağ Mimarisi

Uygulamalarda genellikle ileri beslemeli ağların kullanımı tercih edilmektedir.

2.5.8.4. Katman Sayısı

Daha önceden de belirtildiği gibi yapay sinir ağlarında girdi nöronlarını içeren bir girdi katmanı, çıktı nöronlarını içeren bir çıktı katmanı ve gizli nöronları içeren bir veya daha fazla katman bulunmaktadır. Bu nedenle, uygun gizli katman sayısının belirlenmesi, modelin etkinliği açısından önem taşımaktadır. Veri içerisindeki belirleyici özellikleri ortaya çıkartan ve girdi ile çıktı arasındaki doğrusal olmayan ilişkinin kurulmasına yardımcı olan unsurlar, gizli katman ile bu katmandaki nöronlardır. Yapılan çalışmalar tek gizli katmanın, doğrusal olmayan karmaşık fonksiyon yaklaşımlarında bile başarılı sonuçlar verdiğini göstermiştir. Örneklem sayısının çok fazla olduğu problemlerde iki katman kullanılabilir. İki'den fazla gizli katman kullanımı pratikte bir yarar sağlamamaktadır. Gizli katman sayısının artırılması, hesaplama zamanını artırdığı için ideal sayının 1 ve 2 değerleri kullanılarak deneme-yanılma yoluyla belirlenmesi önerilebilir (Zhang vd., 1998).

2.5.8.5. Nöron Sayıları

Yapay sinir ağlarında üç tip nöron bulunmaktadır; girdi, gizli ve çıktı. Bu nöronların sayısı, aşağıda açıklandığı gibi ağın başarısı için büyük önem taşımaktadır. (Kaastra ve Boyd, 1996, s. 215-236, Zhang vd., 1998)

Girdi nöronu sayısı: Bu sayı, girdi vektöründeki değişken sayısına bağlıdır. Neden-sonuç ilişkisine dayalı tahmin problemlerinde girdi nöron sayısını belirlemek kolaydır. Zaman serisi tahmin problemlerinde ise girdi nöronu sayısı gecikme sayısı ile ilişkilidir. Ancak bu sayıyı belirlemede önerilen kesin bir yol yoktur. Birçok araştırmacı girdi nöronu sayısını belirlemek için deney tasarımı yaparken, bazıları sezgisel yaklaşımlarda bulunmuşlardır.

Gizli nöron sayısı: Bir ağda gizli nöron sayısının belirlenmesi kritik bir karardır.

Ancak önemine rağmen en iyi gizli nöron sayısının belirlenmesinde kullanılan kesin bir formül yoktur. Genelde az sayıda gizli nöron ile çalışma tercih edilir. Az sayıda gizli nöronun genelleştirme yetenekleri daha yüksektir, ancak bu sayının gereğinden az olması öğrenme yeteneğini düşürmektedir. Diğer yandan çok sayıda gizli nöron, ağı genelleştirmeden daha çok ezberleme yeteneği kazanmasına sebep olmaktadır. Ağı eğitim kümesinde iyi, test kümesinde kötü sonuçlar üretmesi eğitim kümesindeki örneklerin ağı tarafından ezberlendiğe işaret etmektedir. Bu nedenle sistematik bir yöntem olarak, küçük bir başlangıç değeri ile başlanması ve ağı performansı iyileşene kadar gizli nöron sayısının artırılması önerilebilir.

Çıktı nöronu sayısı: Çıktı nöronu sayısını çalışılan probleme doğrudan bağlı olduğu için hesaplanması kolaydır. Neden-sonuç ilişkisine dayalı problemlerde bu sayı çıktı (bağımlı değişken) sayısına eşittir. Zaman serisi tahmin problemlerinde ise çıktı nöronu sayısı, tahmin döneminin uzunluğuna eşittir. Tek dönemlik tahminlerde ve Box-Jenkins gibi iteratif yöntemlerin kullanıldığı çok dönemli tahminlerde bu sayı 1'e eşittir. Birden fazla dönemin aynı anda tahmin edildiği çok dönemli tahminlerde ise bu sayı tahmin edilmek istenen dönem sayısına eşittir.

2.5.8.6. Aktivasyon Fonksiyonu

Aktivasyon fonksiyonları içerisinde genelde en çok tercih edilenleri şunlardır:

1. Sigmoid (Lojistik) Fonksiyonu
2. Hiperbolik Tanjant (Tanh) Fonksiyonu
3. Doğrusal Fonksiyon

Bunların arasında en yaygın kullanılanı, Sigmoid transfer fonksiyonudur. Genel olarak, aynı ya da farklı katmanlarındaki nöronlar farklı aktivasyon fonksiyonunu kullanabilirler. Uygulamaların çoğunda ise, aynı katmandaki nöronların aynı aktivasyon fonksiyonunu kullandıkları görülmektedir.

2.5.8.7. Öğrenme Algoritması ve Parametreleri

Popülaritesinden dolayı günümüzde yaygın olarak kullanılan öğrenme algoritmaları, daha önce de belirtildiği gibi GYA ve Levenberg-Marquardt algoritmasıdır. Öğrenme

oranının ve momentum katsayısının eş zamanlı olarak seçiminde sadece birkaç sistematik yol olduğu için, bu öğrenme parametrelerinin en iyi değerleri genellikle deneme yolu ile seçilir. Genellikle 0 ile 1 arasında herhangi bir değer alan öğrenme oranı ve momentum katsayısının en iyi kombinasyonunu bulmak oldukça zordur.

2.5.8.8. Doğruluk Ölçüleri

Yapay sinir ağlarının başarılarını değerlendirmede; doğruluk ölçüleri, modelleme süresi veya eğitim süresi gibi çeşitli başarı kriterleri kullanılabilir. Bu ağların tahmin amaçlı kullanımında ise en etkin olan başarı kriteri doğruluk ölçüleridir (OMH, OMH, OHKK, vd.). İkinci bölümün “tahminleme” başlığı altında anlatılan doğruluk ölçüleri yapay sinir ağları tahminlemelerinde de yaygın olarak kullanılmaktadır.

2.6. Yazın Taraması

Veri madenciliği çok geniş bir kapsama sahip olduğu için hemen her alanda uygulamalarına rastlamak mümkündür. Veri madenciliğinin tahmin amaçlı kullanımı da azımsanmayacak kadar fazladır. Yazın incelendiğinde, çoğunlukla tahmin edilmeye çalışılan değişkenlerin; maliyet, talep miktarı, satış miktarı, enflasyon oranı, döviz kuru değişimi, enerji tüketimi gibi konuların çevresinde yoğunlaştığı görülmektedir. Bu değişkenlerin tahmini için en çok tercih edilen yöntemlerin ise regresyon analizi, zaman serisi yöntemleri ve yapay sinir ağları olduğu anlaşılmaktadır. Yaygın olarak, bağımlı ve bağımsız değişkenler arasındaki ilişkiyi inceleyerek bağımlı değişkenin belirlenmesine yönelik modellerin kurulmasında çok değişkenli doğrusal regresyon ile yapay sinir ağları kullanılırken, zaman serisi verilerini kullanarak gelecek dönemlerdeki tahminler için ARIMA ve ileri beslemeli GY-YSA yöntemlerinden faydalandığı anlaşılmaktadır.

Diğer yandan, yazında ikinci el otomobillerin fiyatlarının tahmin edilmesine yönelik bir çalışmaya rastlanamamıştır. Bu nedenle yazın taraması kapsamında genel olarak veri madenciliğinin tahmin amaçlı kullanıldığı çalışmalara yer verilerek, bu çalışmalarla ilgili önemli değişkenler Tablo 2.5’de sunulmuştur.

Tablo 2.5. Veri madenciliğinin tahmin amaçlı kullanıldığı çalışmalar

| Çalışmanın Yazar(lar)ı | Yayın Yılı | Yapıldığı Ülke | Kullanılan Yöntemler | Veri Kümesi | Tahmin Değişkeni | Bulunan Sonuç |
|------------------------------------|------------|-------------------------|-------------------------------------|--|---|---|
| İşeri A., Karlık B. | 2008 | Türkiye | GY-YSA | 159 adet yeni otomobil verisi | Yeni otomobil fiyatlaması | YSA' nın otomobil fiyatlamada iyi bir model olması |
| Usta A.S. | 2007 | Türkiye | YSA, VAR, ARMA | 1990–2006 yılları arasındaki ÜFE zaman serisi verileri | Türkiye' deki gelecek dönem ÜFE değerleri | YSA' nın diğer yöntemlere göre daha iyi bir tahmin performansına sahip olması |
| Özgen D. | 2007 | Türkiye | YSA | Ocak 2000– Eylül 2006 ayları arasındaki İMKB 30 zaman serisi verileri | İMKB-30 Endeksi | YSA modelinin İMKB-30 Endeksi' ni modelleyebilmesi |
| Bucak S. | 2007 | Türkiye | YSA | Otomobil parçalarına ait; üretim miktarları, genel giderler, fiili üretim süresi bilgileri ve standart maliyet bilgileri | Otomotiv sektöründe maliyet | YSA ile yapılan birim maliyet tahmininin oldukça tutarlı sonuçlar vermesi |
| Yıldız Ö. | 2006 | Türkiye | GY-YSA, Elman YSA | Amerikan Doları' nın 4 Ocak 1999–28 Şubat 2006 tarihleri arası günlük verileri | Amerikan Doları' nın gelecekteki seyri | Zaman serilerinin kullanımında GYA' nın Elman ağlarına göre daha başarılı sonuçlar vermesi |
| Cortez P. | 2006 | Portekiz | Box-Jenkins, Üstel Düzleştirme, YSA | İnternet trafiğinde 5 dakikalık ve 1 saatlik periyotlarda elde edilen veriler | İnternet trafiği | Box-Jenkins ve YSA' nın üstel düzleştirmeye göre daha iyi sonuçlar vermesi; internet gibi gerçek zamanlı verilerde YSA kullanımının önerilmesi |
| Özçınar H. | 2006 | Türkiye | Çoklu Regresyon, YSA | Pamukkale Üniversitesi Sınıf Öğretmenliği bölümüne 1999, 2000 ve 2001 yıllarında kayıt olan öğrencilere ait veriler | KPSS sonuçları | YSA' nın regresyona göre daha doğru sonuçlar vermesi |
| Ayyıldız M., Kalıpsız O., Yavuz S. | 2006 | Türkiye | İB-GY YSA, Elman-YSA | 160 adet veri | Yazılım projeleri maliyeti | Elman YSA' nın, İB-GY-YSA' na göre daha başarılı sonuçlar vermesi |
| Palmer A., Montaña J. J., Sesé A. | 2005 | Balea Adaları (İspanya) | YSA | Üçer aylık zaman serisi verileri | Balea Adaları' nın turizm gelirleri | 8-1-1 mimarisine sahip modelin diğerlerine göre daha yüksek tahmin başarısı sağladığı ve turizm zaman serileri için uygun bir tahmin tekniği olması |
| Subaşı D. | 2005 | Türkiye | ARIMA | 1994–2004 yılları arasındaki aylık enflasyon verileri | Türkiye' nin 2005 yılı aylık enflasyon oranları | Aylık enflasyon serisini en iyi açıklayan modellerin SARIMA(0,1,1)(0,0,1)+sabit ve SARIMA(1,0,1)(0,0,1)+trend olması |

| Çalışmanın Yazar(lar)ı | Yayın Yılı | Yapıldığı Ülke | Kullanılan Yöntemler | Veri Kümesi | Tahmin Değişkeni | Bulunan Sonuç |
|-----------------------------------|------------|-----------------|--|---|--|--|
| Kapucugil A. | 2005 | Türkiye | Regresyon, YSA | 1995–2004 yılları arasındaki hisse senedi halka arz fiyatları | Hisse senetlerinin halka arzlarının ilk gün fiyat performansları | YSA' nın regresyona göre çok yüksek bir performans sergilemesi |
| Çalık İ. | 2005 | Türkiye | İB-YSA, Monte Carlo Benzetimi | 1963–2003 yılları arasındaki Mersin Limanı yük trafiği | Mersin Limanı yük hacmi | İB-YSA yardımıyla Mersin Limanı'nda 2030 yılına kadar oluşacak yük hacminin tahmin edilebileceği ve kapasite kullanımının belirlenebileceği |
| Kişi Ö. | 2005 | Türkiye | Stephens-Stewart modeli, YSA, Çoklu doğrusal regresyon | Günlük sıcaklık verileri | Buharlaşma miktarı | YSA modelinin çalışmada kullanılan diğer yöntemlere göre daha iyi sonuçlar vermesi |
| Partal T., Cıgızoğlu K. | 2005 | Türkiye | YSA, Çoklu doğrusal regresyon | 1977-2001 yılları arasındaki günlük meteorolojik veriler (sıcaklık, rüzgar hızı, nem ve buharlaşma miktarı) | Yağış miktarı | YSA modelinin çalışmada kullanılan diğer yöntemlere göre daha iyi sonuçlar vermesi |
| Doğan E., Işık S. | 2005 | Türkiye | Penman-Monteith modeli, Radyal temelli YSA modeli | 2000, 2002, 2003, 2004 yıllarına ait günlük buharlaşma değerleri | 2001 yılı günlük buharlaşma miktarı | Radyal temelli YSA modelinin çalışmada kullanılan diğer yöntemlere göre daha iyi sonuçlar vermesi |
| BaFail A.Ö. | 2004 | Suudi Arabistan | YSA | 1975–1986 yılları arasındaki uçak yolcusu sayısı | Yerel ve uluslararası havayollarındaki uçak yolcusu sayısı | YSA kullanılarak uçak yolcu sayısının tahmin edilebileceği ve bu tahminde GSYİH, kişi başına düşen milli gelir ve nüfus değişkenlerinin etkin olması |
| HamzaÇebi C., Kutay F. | 2004 | Türkiye | ARIMA, YSA, Regresyon | 1970–2002 yılları arasındaki yıllık zaman serisi verileri | 2010 yılına kadar Türkiye Elektrik Enerjisi Tüketimi | Yapay sinir ağlarının elektrik enerjisi tüketiminde iyi bir tahmin aracı olması |
| Cho V. | 2003 | Hong Kong | Üstel düzleştirme, ARIMA, Elman YSA | Üçer aylık zaman serisi | Hong Kong'a yönelik dış turizm talebi | Turist sayılarının tahmininde Elman yapay sinir ağının diğer yöntemlere göre en iyi sonuçları vermesi |
| Baldemir E., Bahar O. | 2003 | Türkiye | Hareketli ortalamalar, Çoklu regresyon, İB-GY-YSA | 1984–1999 yılları arasındaki yıllık veriler | Beş farklı ülkeden Türkiye'ye gelen turist sayıları | Geri yayımlı yapay sinir ağı modelinin daha doğru sonuçlar vermesi |
| Fritzer F., Moser G., Scharler J. | 2002 | Avusturya | ARIMA, VAR (Vector Auto Regression) | Tüketici fiyat endeksinden hareketle hesaplanan aylık zaman serisi verileri | Avusturya enflasyon oranları | Kısa dönem tahminlerde ARIMA, uzun dönem tahminlerde ise VAR modellerinin daha iyi performans göstermesi |

| Çalışmanın Yazar(lar)ı | Yayın Yılı | Yapıldığı Ülke | Kullanılan Yöntemler | Veri Kümesi | Tahmin Değişkeni | Bulunan Sonuç |
|--|-------------------|-----------------------|--|--|--|--|
| Eğrioğlu E., Aladağ Ç. | 2002 | Türkiye | ARIMA, YSA, Melez yaklaşım | Mart 1994–Kasım 2002 ayları arasındaki aylık zaman serisi verileri | Ankara Temmuz 2002 – Kasım 2002 arası aylık hava kirliliği | YSA modeli, ARIMA ve melez yaklaşıma göre en iyi sonucu vermesi |
| Kermanshah B., Iwamiya H. | 2002 | Japonya | GY-YSA, Jordan GB-YSA | Zaman serisi verileri | 2020 yılına kadar elektrik enerjisi tüketimi | GY-YSA modeli ile uzun vadeli elektrik enerjisi tüketimi tahmini yapılabileceği |
| Montanes E., Quevedo J., Prieto M., Menendez C. | 2002 | İspanya | Box-Jenkins, YSA | Nükleer enerji ünitesinden alınan gerçek veriler | Veri serisindeki ani sistematik değişiklikleri | Box-Jenkins tekniğinin veri serisindeki ani sistematik değişiklikleri öngörememesi |
| Burger C. J. S. C., Dohnal M., Kathrada M., Law R. | 2001 | Güney Afrika | Mekanik tahmin, Regresyon analizi, ARIMA, YSA | Aylık zaman serisi verileri | Durban’a (Güney Afrika) yönelik Amerikan turizm talebi | YSA’ nın diğer modellere göre daha başarılı olduğu ve zaman serilerinde güvenle kullanılabilceği |
| Plummer E. A. | 2000 | ABD | İB-YSA, K-en yakın komşu | Yapay olarak üretilmiş zaman serisi | Zaman serisindeki gelecek değerler | K-en yakın komşu ile yapılan tahmin sonuçlarının İB-YSA’ dan daha iyi sonuçlar vermesi |
| Uysal M. El, Roubi S. | 1999 | ABD | Çoklu regresyon modeli, YSA | Üçer aylık zaman serisi verileri | ABD’ yi ziyaret eden Kanada’lı turistlerin yapmış oldukları harcamalar | Her iki modelin de birbirine yakın sonuçlar vermesi |
| Law R., Au N. | 1999 | Hong Kong | Mekanik tahmin, Çoklu regresyon, Üstel düzleştirme, Hareketli ortalama, İB-YSA | 1967–1996 yılları arasındaki yıllık zaman serisi verileri | Hong Kong’a yönelik Japon turizm talebi | YSA’ nın diğer modellere göre daha iyi tahmin sonuçları vermesi |
| Ahmed S., Cross J. | 1999 | Yeni Zelanda | GB-YSA, Self adaptive YSA | Üçer aylık zaman serisi | Turizm talebi | Self adaptive YSA’ nın diğer modellere göre daha yüksek tahmin performansı göstermesi |
| Al-Saba T., El-Amin L. | 1999 | Suudi Arabistan | YSA, AR, ARMA ve ARIMA | Yük istemini gösteren zaman serileri | 1997-2006 tarihleri arasında en yüksek yük istemi | YSA ile yapılan kestirim sonuçlarının daha iyi olması |
| Choi H. Y., Kim W., An S. Y. | 1997 | Güney Kore | ARIMA, GB-YSA | Aylık zaman serisi verileri | Otelin doluluk oranı | Yazarlar tarafından geliştirilen “Geri beslemeli ayrıştırılmış yapay sinir ağı modelinin (Recurrent and Decomposed Neural Network)”, ARIMA ve diğer yapay sinir ağı modellerine göre daha yüksek tahmin performansı göstermesi |

| Çalışmanın Yazar(lar)ı | Yayın Yılı | Yapıldığı Ülke | Kullanılan Yöntemler | Veri Kümesi | Tahmin Değişkeni | Bulunan Sonuç |
|----------------------------|------------|----------------|----------------------------------|--|--|--|
| Pattie D. C., Snyder J. | 1996 | ABD | Üstel düzleştirme, ARIMA, GY-YSA | Aylık zaman serisi verileri | Milli parkları ziyaret eden ziyaretçi sayısı | Yapay sinir ağının en yüksek doğruluğa sahip tahmin modeli olması |
| Sharp K. P. | 1992 | Kanada | ARIMA, VAR | Tüketici fiyat endeksi ve ücretlere ilişkin aylık zaman serilerinden hareketle hesaplanan aylık enflasyon verileri | Kanada fiyat ve ücret enflasyonu | ARIMA(1,1,0) modelinin Kanada ücret ve fiyat enflasyonunu açıklayan en iyi model olması |
| Cummins D., Griepentrog G. | 1985 | ABD | ARIMA, Ekonometrik modeller | 1974–1983 yılları arasındaki otomobil sigorta firmalarının ödediği hasar tazminleri | Hasar tazminleri | ARIMA' nın ekonometrik modellere göre daha başarılı sonuçlar vermesi |
| Carlson R. L., Umble M. | 1980 | ABD | Çoklu regresyon | Farklı otomobil türleri için; harcanabilir gelir, otomobil türüne göre değişen fiyatlar, benzin fiyatları, benzin kıtlığı verileri | Satınalma talebi | Çoklu regresyon ile belirlenen otomobiller için gelecek 5 yıllık satınalma talebinin tahminlenebilmesi |

Üçüncü Bölüm

3. OTOMOTİV ENDÜSTRİSİ ve İKİNCİ EL OTOMOBİL PAZARI

Günümüzde ikinci el otomobil pazarı, hemen her ülkede giderek büyümekte olan ve ekonomide çok önemli hacim işgal eden bir sektör haline gelmiştir. Dünya otomotiv endüstrisindeki toplam motorlu araç üretiminin yaklaşık % 70'ini ise otomobil üretimi oluşturmaktadır. 2006 yılında Avrupa' da üretilen tüm motorlu araçların % 87' sini otomobiller oluşturmuştur (Web-ACEA). Bu verilerden hareketle, otomobil sektörü için otomotiv endüstrisinin en temel yapı taşı nitelmesi yapılabilir. Otomobil üretimi, güçlü yan sanayisi ile birlikte diğer taşıtların üretimine de destek olmaktadır (Onat, 2007; Web-Kobifinans1).

Çalışmanın bu bölümünde, ikinci el otomobil pazarının önemini vurgulamak için öncelikle otomotiv endüstrisi hakkında bilgi verilmiş, ardından ikinci el otomobil pazarındaki araştırma sonuçları sunulmuştur.

3.1. Otomotiv Endüstrisi

3.1.1. Tanımı ve Kapsamı

Otomotiv endüstrisi, motorlu taşıt aracı üreten bir sanayidir. Bu endüstri gelişmiş ve gelişmekte olan ülkelerin ekonomisinde lokomotif olarak görülmektedir. Petro kimya, demir çelik, lastik, savunma sanayi gibi çok sayıda önemli sanayi dallarıyla yakın ilişki içerisinde olan bu dev endüstri kolu, önemli hacimde bir istihdam da sağlamaktadır.

Uluslararası düzeyde otomotiv endüstrisindeki en üst seviyedeki kuruluş, Türkiye' nin de üyesi olduğu Uluslararası Motorlu Taşıt Üreticileri Derneği (The International Organization of Motor Vehicle Manufacturers) dir. Bu dernek kısaca OICA (Organisation Internationale des Constructeurs d'Automobiles) olarak da bilinir. 1919' da Paris' de kurulan OICA, temel olarak ülkelerin ulusal otomobil derneklerinin bağlantısını sağlamakta, araç üreticilerinin, montajcılarının ve ithal edicilerinin çıkarlarını korumaya yönelik tedbirlerin alınmasını organize etmektedir. Bunun yanında uluslararası araç fuarlarını koordine etmekte, endüstri istatistiklerini yayınlamakta, araçlarla ilgili

uluslararası mevzuatların geliştirilmesine katkı sağlamaktadır (Onat, 2007; Web-Kobifinans1).

Otomotiv sektörü genel olarak “karayolu taşıt araçları (binek otomobili, otobüs, midibüs, minibüs, çekici, kamyon ve traktör) ile bu araçların üretiminde kullanılan parçaları üreten bir sanayi dalı” olarak tanımlanmaktadır (Türkiye Vakıflar Bankası Yayınları, 2004, s.4)

Dumanlı (1987, s.23) otomotiv endüstrisini, “bir ülkenin yük ve yolcu taşıma ihtiyaçlarını karşılayacak karayolu taşıtlarını üreten bir sanayi kolu” olarak tanımlamaktadır.

Devlet Plânlama Teşkilatı (DPT) ise, otomotiv endüstrisi için “Raya Bağlı Olmayan Motorlu Taşıtları” tanımını kullanmış ve kamyon, kamyonet, minibüs, otobüs, binek ve arazi taşıtı ve diğerlerinin imalâtı, montajı, onarımı ve tamirâtı ile sınırlamıştır (İMKB Yayınları, 1995).

Avrupa Komisyonu’nun 1994 yılı tebliğindeki bir ifadede ise otomotiv endüstrisi; “Karayolları Taşıt Araçları İmalât Sanayi” olarak tanımlanmaktadır. Bu tanım, Makine Mühendisleri Odası ve Devlet İstatistik Enstitüsü tarafından da kullanılmaktadır. Avrupa Komisyonu, otomotiv endüstrisinin kapsamına giren ürünleri şu şekilde sıralamaktadır (DİE Yayınları, 1990, s.15):

1. Motorlu Kara Taşıtları

- Otomobil
- Arazi Aracı
- Otobüs
- Minibüs
- Midibüs
- Kamyon
- Kamyonet

2. Traktörler

3. Motosikletler ve Bisikletler

4. Motorlu/Motorsuz Diğer Kara Taşıtları

- Dozer

- Kepçe
- Forklift
- Diğer İş Makineleri

Otomotiv Sanayii Derneği (OSD) ise otomotiv endüstrisini; alttaki görüldüğü gibi iki bölüme ayırmış, genel itibariyle de çekici, kamyon, kamyonet, otomobil, otobüs, minibüs, minibüs imalatı, montajı ve tamirâtı ile sınırlandırmıştır (Yeltin, 1999, s.3).

1) Ana sanayi

a. Otomobiller

b. Ticari araçlar

- Hafif ticari araçlar
- Ağır ticari araçlar

2) Yan sanayi (aksam ve parçalar)

Bu temel sınıflandırma küresel olarak kabul görmekle birlikte, alt sınıflandırmalar ülkelere göre farklılıklar gösterebilmektedir. Türkiye Cumhuriyeti'nde bir motorlu aracın hangi sınıfa dâhil olduğu Sanayi ve Ticaret Bakanlığı tarafından belirlenmektedir.

3.1.2. Tarihçesi

Yüz yılı aşkın bir tarihi geçmişe sahip olan otomotiv endüstrisi faaliyetleri, başlangıçta otomobil üretimi ile başlamış ve Birinci Dünya Savaşı yıllarında ticari araç üretimi de gerçekleştirilerek, toplam üretim içerisinde otomobil ağırlıklı olmak üzere sürekli bir gelişim ve değişim içerisinde olmuştur. 1769 yılında buhar gücüyle çalışan ilk aracın keşfiyle temelleri atılan otomotiv endüstrisi, başta Almanya ve Fransa olmak üzere Avrupa'da filizlenmiş, binek otomobillerin üretimi ile Amerika Birleşik Devletleri'nde gelişimini sürdürmüş ve Birinci Dünya Savaşı yıllarının ardından otomobil ağırlıklı olmak üzere sürekli bir gelişim ve değişim içerisinde olmuştur (Bedir, 1999, s.7).

Bazı yazılı kaynaklar otomotiv endüstrisinin ilk başlangıç noktası olarak, Newton'un 1680'de tepkili bir araç projesine dayanarak geliştirdiği taşıtı gösterse de, kendi kendine hareket eden bir aracın üzerine oturup kontrolünü eline almış ilk kişinin Fransız yüzbaşı mühendis Nicholas Joseph Cugnot olduğu konusunda görüş birliği bulunmaktadır. Cugnot'un 1769 yılında yaptığı ve saatte 4 km hıza sahip buhar gücüyle çalışan ilk araç

“Otomobilin Atası” olarak kabul edilmektedir. Cugnot, buhar gücüyle çalışan bu ilk aracı, üç tekerlekli olarak ve silahları çekmede kullanılmak amacıyla üretmiştir. Ancak, bu aracın aşırı yavaşlığı ve yeterli işlevde olmayışı beklenen kullanımını engellemiştir (Azcanlı, 1995, s.11).

Daha sonra, 1801 yılında İngiliz Richard Trevithick tarafından bir başka buharla çalışan araç üretilmiş ve bunun da oldukça ağır olması güçlük teşkil etmiştir. Bu çalışmalar, 1805 yılında Amerikalı Oliver Evans’ın yine buharla çalışan ancak daha çok limanda kullanılmak üzere düşünülmüş ve trafikte de kullanılan bir başka araç üretimiyle sürdürülmüştür. Zaman içerisinde bu çalışmalar geliştirilmiş ve örneğin, 1829 yılında Sir Goldswort Guynes isimli bir İngiliz saatte 25 km hız yapabilecek buharla çalışan aracı icat etmiştir (Uludağ İGBS, 2006, s.2). İçten yanmalı motorun ilk keşfi ise 1860 yılında Etienne Lenoir tarafından Paris’te gerçekleştirilmiştir. Bu tarihten dört yıl sonra da Köln’deki Gasmotorenfabrik Deutz AG fabrikasında içten yanmalı sabit motorların üretimine başlanmıştır. Bu fabrikanın kurucularından olan Otto, 1876 yılında ilk olarak dört silindri içten yanmalı benzinli motorun üretimini gerçekleştirmiştir (Bedir, 2002, s.2).

Ancak, içten yanmalı motorlu, bugünkü anlamda modern bir otomobilin ilk olarak üretimi 1886 yılında Karl Benz ve Gottlieb Daimler tarafından gerçekleştirilmiştir (OSD, 2005, s.7). Otomobil kullanımı ve üretimi bu yıllardan sonra hızlı bir şekilde Avrupa’da yayılmıştır. Ayrıca, 1893 yılında Amerika’da da içten yanmalı motorlu otomobil üretimi başlamış ve gittikçe de üretim ve kullanımı artmıştır. Dünya genelinde otomobil marka sayısı, 1880’de 8 adetten, 1885’te 50 ye, 1890’da ise 500 adede ulaşmıştır (Bloomfield, 1978, s.17). “Bu araçların üretiminin küçük atölyelerde, basit işleme aletleriyle, standart dışı ve emek yoğun olarak yapıldığı göz önüne alınırsa, başlangıç yıllarında hızlı ve yoğun bir gelişme gösterdiği anlaşılmaktadır” (Bedir, 2002, s.2).

Motorlu araç üretiminin ilk yıllarında güç kaynağı olarak buharlı, elektrikli ve içten yanmalı motorların hepsinden faydalanılmıştır. Özellikle başlangıç yıllarında buharlı ve elektrikli motorlar kullanılmakta ise de belirli bir süre içerisinde içten yanmalı motorlar piyasanın tamamına yakını oluşturmuştur. Buhar gücünden faydalanarak çalışan araçlarda, yanmayı sağlamak için devamlı olarak yüksek miktarda yakıt ve su gerekmesi ve bu araçların çok ağır oluşları zamanla kullanımı azaltmıştır. Elektrikli motorların ise temiz, sessiz ve basit idareli oluşu önemli bir üstünlük olarak görülmesine rağmen, akünün büyük ve ağır oluşu ve sıkça doldurulması gerekliliği, bu motorların da kullanımını

sınırlamıştır. İçten yanmalı motorların üretimi ile buhar gücüyle çalışan motorların ve elektrikli motorların yukarıda bahsedilen olumsuzluklarından kurtulunmuş; hafif, yakıt tasarruflu ve işlevsel bir güç kaynağına kavuşulmuştur (Bedir, 1999, s.7). Örneğin ABD’de, 1900 yılında üretimi gerçekleştirilen 4.100 adet aracın % 22,4’ü içten yanmalı motorla, % 37,5’i elektrikli motorla, % 40,1’i buhar gücüyle çalışmakta iken, 1909 yılına gelindiğinde üretilen 126.700 adet aracın, % 95,1’i gibi tamamına yakın bir kısmını içten yanmalı motorlu araçlar teşkil etmiş olup; içten yanmalı motorların büyük bir çoğunluğu da benzinli motorlardır (Bloomfield, 1978, s.17).

Günümüzde içten yanmalı motorlu araçlarda görülen önemli özelliklerin büyük bir bölümünün (çoklu silindirler, V silindir, elektrikli ateşleme, karbüratör vb.) üretimi Birinci Dünya Savaşı yıllarına kadar gerçekleştirilmiş ve araç üzerine uyarlanması sağlanmıştı. Standart ölçülerde ve büyük miktarlarda otomobil üretimine Henry Ford’un Model T üretimi ile başlanmıştır. Model T üretiminde uygulanan seri üretim tekniği, yalnızca otomotiv endüstrisinde değil, diğer tüm endüstrilerde de bir devrim niteliği taşımaktadır. Geniş pazar imkânları iyi değerlendirilerek yapılan büyük miktarlarda araç üretimiyle, düşük maliyette üretim sağlanmıştır. Seri üretim tekniğiyle üretilen bu otomobiller, 1920’de ABD’deki araçların % 65-70’ini, dünyada ise % 50’ sini oluşturmuşlardır. Ticari araç üretimi ise otomobil üretimine göre daha sonraki yıllarda meydana gelmiştir. Örneğin, karavan (ev araba) ve otobüs üretimine 1912 yılında, kamyon üretimine ise Birinci Dünya Savaşı yıllarında başlanmıştır. Dizel motorunun icadı 1892 yılında olmasına rağmen ticari araçlarda kullanımı ilk olarak 1908 yılında İsveç Saurer firmasınınca gerçekleştirilmiş, ancak yaygın olarak kullanımı özellikle 1930’lardan sonra olmuştur. Otomobilde ilk dizel motor kullanımına ise 1952 yılında Mercedes firmasınınca başlanmıştır (Bedir, 2002; Onat, 2007).

3.1.3. Ürünleri

Otomotiv endüstrisi ürünleri iki temel kategoride toplanmaktadır: Otomotiv Ana Sanayi Ürünleri ve Otomotiv Yan Sanayi Ürünleri.

3.1.3.1. Otomotiv Ana Sanayi Ürünleri

Otomotiv ana sanayi üreticileri tarafından üretilen, son mamul niteliğindeki ürünler “Otomotiv Ana Sanayi Ürünleri” olarak adlandırılmaktadır. İşlevlerine göre otomotiv ana sanayi ürünleri şunlardır (Çildir, 2006, s.14-15):

“A. Yolcu Araçları: Yolcu taşımak için imal edilmiş otomobil, minibüs, otobüs ve motosiklet gibi motorlu araçlardır.

1. Otomobil: Yapısı itibariyle sürücüsünden başka en çok yedi oturma yeri olan ve insan taşımak için üretilmiş motorlu araçtır. Bunlardan taksimetre veya tarife ile yolcu taşıyanlara “Taksi Otomobil”, kişi başına ücretle yolcu taşıyanlara “Dolmuş Otomobil”, bunların dışında kalan ve 237 sayılı taşıt kanununa tabi olmayanlara da “Özel Otomobil” denir.
2. Minibüs: Yapısı itibariyle, sürücüsünden başka 8 ile 14 oturma yeri olan ve insan taşımak için imal edilmiş bulunan motorlu araçtır.
3. Midibüs-Otobüs: Yapısı itibariyle, sürücüsünden başka en az 15 oturma yeri olan ve insan taşımak için imal edilmiş olan motorlu araçlardır.
4. Motosiklet-Moped: 2 veya 3 tekerlekli, sepetli veya sepetsiz motorlu araçlardır. Bunlardan karoseri yük taşıyabilecek şekilde sandıklı veya özel biçimde yapılmış olan ve yolcu taşımalarında kullanılmayanlarına “Yük Motosikleti” ya da “Triportör” denir.

B. Yük Araçları: Yük taşımak için imal edilmiş motosiklet, kamyonet ve kamyon gibi motorlu araçlardır.

1. Çekici: Römork ve yarı römorkları çekebilmek için kullanılan ve yük taşımayan motorlu araçtır.
2. Kamyon: İzin verilen azami yüklü ağırlığı 3500 kg’dan fazla olan ve yük taşımak için üretilmiş motorlu araçtır.
3. Kamyonet: İzin verilen azami yüklü ağırlığı 3500 kg’ı geçmeyen ve yük taşımak için üretilmiş motorlu araçtır.

C. Diğer Taşıtlar: Çeşitli amaçlar için imal edilmiş özel amaçlı taşıtlar ile yol ve iş makineleridir.

1. Arazi Aracı: Karayolunda yolcu ve yük taşıyabilecek şekilde üretilmiş olmakla beraber bütün tekerlekleri motordan güç alan veya alabilen motorlu araçtır.
2. Özel Amaçlı Araçlar: Özel amaçla, insan veya eşya taşımak için üretilmiş olan ve İtfaiye, Cankurtaran, Cenaze, Radyo, Sinema, Televizyon, Kütüphane, Araştırma

araçları ile bozuk veya hasara uğramış araçları çekmek veya taşımak, kaldırmak gibi özel işlerde kullanılan motorlu araçlardır.

3. İş Makineleri: Paletli veya madeni tekerlekli Traktör, Biçer-Döver, Yol ve İş Makineleri ile benzeri tarım ve sanayi, bayındırlık, milli savunma ile çeşitli kuruluşların iş ve hizmetlerinde kullanılan, karayolunda insan, hayvan ve yük taşımada kullanılan motorlu araçlardır.”

3.1.3.2. Otomotiv Yan Sanayi Ürünleri

Otomotiv ana sanayi ürünlerinin üretiminde kullanılmasına karşın, genellikle otomotiv fabrikaları dışında üretilen mamul ve yarı mamuller “Otomotiv Yan Sanayi Ürünleri” olarak adlandırılmaktadır. Niteliklerine göre otomotiv yan sanayi ürünleri iki ana grupta incelenir (Orhan, 1997, s.81):

1. Orijinal Yan Sanayi Ürünleri: Araç üreticisinin markası adı altında üretilen ve genelde araç üreticisinin dağıtım kanalları kullanılarak pazarlanan ürünlerdir.
2. Orijinal Olmayan Yan Sanayi Ürünleri: Bağımsız üreticiler tarafından üretilen ve genelde markasız veya üretici firmanın markasından bağımsız kanallarca dağıtımı gerçekleştirilen ürünlerdir.

3.1.4. Ekonomideki Önemi

Otomotiv endüstrisi, tüm sanayileşmiş ülkelerde ekonominin lokomotifini olarak kabul edilmektedir. Bunun en önemli nedeni ise diğer sanayi dalları ile olan çok yakın ilişkisidir. Otomotiv endüstrisi demir-çelik, petro-kimya, lastik gibi temel sanayi dallarında başlıca alıcı ve bu sektörlerdeki teknolojik gelişmenin de sürükleyicisidir. Turizm, altyapı ve inşaat ile ulaştırma ve tarım sektörlerinin gerek duyduğu her çeşit motorlu araçlar sektör ürünleri ile sağlanmaktadır. Bu sektördeki değişimler, ekonominin tümünü yakından etkilemektedir (Web-Kobifinans2).

Dünya genelinde sahip olduğu güçlü tüketici talebiyle dinamik bir yapı arz eden otomobil sektörü; kullanılan teknoloji, üretim miktarı, istihdam hacmi, ürün çeşitliliği ve ticaret hacmi bakımından global etkiye sahip bir konuma gelmiştir. Ancak, gerek üretim miktarı ve kalitesi, gerekse ticaret ve yatırımda piyasa şartları ve tüketici esneklerinin değerlendirilmesinde görülen farklılıklar, sektörün henüz dünya ölçeğinde homojen bir

yapıya ulaşmadığını göstermektedir (Bilgin, 1999, s.3).

Sanayileşmiş ülkelerin tamamına yakınında otomotiv endüstrisi firmaları, ülkelerin büyük şirketleri sıralamasında ilk sıralarda yer almakta; otomotiv sektörü, üretim büyüklüğü ve yarattığı katma değer itibariyle ise toplam imalat sanayii üretimi içerisinde yine ilk sıralarda bulunmaktadır. Örneğin, 1914 yılında ABD imalat sanayii üretimi içerisinde 14. sırada bulunan otomotiv endüstrisi, yüzyılın ilk çeyreğinden sonra birinci sıraya yükselmiş ve bu durumunu uzun yıllar sürdürmüştür (Bloomfield, 1978, s.26).

Otomotivin temel özelliklerine baktığımızda şu özelliklerin ön plâna çıktığı görülmektedir (Web-Kobifinans1, 2006; Özerman, 2004):

- Uzay-havacılık endüstrisinden sonraki en karmaşık teknolojiyi gerektiren otomotiv endüstrisi, başlıca önemli mühendislik alanlarını içeren çok disiplinli bir teknoloji gerektirmektedir.
- Motorlu taşıt aracı; niteliği, malzeme yapısı, prosesi, teknolojisi ve üretim yeri farklı olan 5 bin dolayında parçanın, ortak kalite yönetimi ve verimlilik anlayışı ile üretimi ve bir araya getirilmesi ile ortaya çıkmaktadır.
- Bir motorlu aracın üretimi ve trafiğe çıkabilmesi için güvenlik, trafik ve çevre ile ilgili 50 dolayında küresel teknik mevzuata uyumlu olması ve bunun belgelendirilmesi zorunludur. Bunun yanı sıra, isteğe bağlı olarak uygulanabilen 10 dolayında diğer uluslararası mevzuat bulunmaktadır. Ayrıca söz konusu kapsamlı mevzuat, teknolojideki gelişmelere bağlı olarak sürekli yenilenmektedir.
- Pazardaki yoğun rekabet nedeni ile müşteri tatmini ancak teknolojik gelişme ile sağlanmaktadır. Bu nedenle sektörde, yoğun Ar-Ge ve sürekli gelişme esastır.
- Otomotiv sektörü kendisi dışında, ham madde ve yan sanayi ile otomotiv ürünlerinin tüketiciye ulaşmasını sağlayan ve bunu destekleyen pazarlama, bayi, servis, akaryakıt, finans ve sigorta sektörlerinde geniş bir iş hacmi ve istihdam oluşturmaktadır.
- Sektör, savunma sanayinin gelişmesinde ve teknolojik düzeyin yükselmesinde temel oluşturmaktadır.

Bu özellikleri nedeni ile otomotiv endüstrisi, stratejik bir sanayi olarak hükümetlerin yakın ilgisini çekmekte ve bu sektör için özel bir plânlama yapılmaktadır. Özellikle hızla küreselleşmekte olan bu sektörde rekabet çok yoğun bir şekilde hissedilmektedir. Bu

nedenle otomobil üreticisi ülkeler arasında ve AB, Nafta gibi ekonomik birlikler içerisinde bu sektörün korunması ve rekabet gücünün geliştirilmesi için özel politikalar uygulanmaktadır (Alkan, 1996, s.6; Web-Kobifinans1).

Otomotiv endüstrisi yeni üretim tekniklerinin ilk defa kullanılmaya başlandığı sektörlerin başında yer almaktadır ve ortaya koyduğu katma değeri ile ülke ekonomilerine çok önemli katkılar sağlamaktadır. Otomotiv ana sanayisinde çalışan bir kişinin ana ve yan sanayide 5 kişiye istihdam oluşturduğu genel kabul görmektedir. Ayrıca, akaryakıt istasyonları, otomotiv ürünlerine yönelik reklam ve sigorta hizmetleri, oto kiralama şirketleri, karayolu taşımacılığı ve otopark hizmetleri, yedek parça, satış ve satış sonrası hizmetleri gibi araç kullanımına paralel hizmetler de düşünüldüğünde ortaya çıkan dolaylı istihdamın büyüklüğü bu endüstrinin önemini açıkça ortaya koymaktadır. En büyük üç otomotiv endüstrisindeki bazı veriler de, bu endüstrinin önemini destekler niteliktedir (Bedir, 1999; Web-OICA):

ABD Otomotiv Endüstrisi

- Üretim büyüklüğü itibariyle ülkenin en büyük sektörü olup, özel sektörün ürettiği yurtiçi katma değer % 5'inden fazlası,
- ABD'nin tarım dışı ürünleri ihracatının % 12'si,
- Ana sanayide 620.000 kişinin istihdamı, otomotiv endüstrisinde sağlanmaktadır. Ayrıca sektörün 1998 yılı Ar-Ge yatırımı 18,4 milyar dolar olup, Ar-Ge ye en fazla kaynak aktarılan sektör konumundadır.

AB Otomotiv Endüstrisi

- Doğrudan yaratılan istihdam 1.904.000 kişi olup, imalat sanayii istihdamının % 8,2'si,
- İmalat sanayii katma değerinin % 9,3'ü,
- AB içerisinde yaratılan toplam katma değer % 1,61'i, otomotiv endüstrisinde oluşturulmaktadır.

Japonya Otomotiv Endüstrisi

- Geçmiş yıllarda imalat sanayii üretiminin % 10'undan fazlası,
- 1998 yılında 77 milyar \$ ihracat seviyesiyle toplam ihracatın % 20'si,
- 750.000 kişilik doğrudan istihdam, otomotiv sektöründe sağlanmaktadır.

3.1.5. Üretim Alanındaki Gelişmeler

20. yüzyılın ilk yıllarında dünya otomotiv endüstrisinin çok hızlı bir büyüme gösterdiği söylenebilir. Tablo 3.1' de görüldüğü gibi 1900 yılında Fransa ve ABD ağırlıklı olmak üzere toplam 9.500 adet olan üretim, daha sonraki yıllarda seri üretimin de sağladığı düşük fiyat avantajıyla oluşan talep artışıyla, 1915 yılında 1.000.000 adetin üzerinde olmuştur. 1900-1915 yılları arasında yıllık ortalama % 37'lik bir üretim artışı sağlanmıştır (Bedir, 2002, s.3). Tablo 3.1' deki verilerin orantılara dönüştürülmesi sonucunda; Birinci ve İkinci Dünya Savaşı yılları ağırlıklı olmak üzere 1950 yılına kadar dünya motorlu araç üretiminin % 80'inden fazlasının ABD önderliğinde gerçekleştirildiği anlaşılmaktadır (Bedir, 1999; Web-OICA):.

Tablo 3.1. Yıllar itibariyle dünya motorlu araç üretimi (Bin adet)
(Kaynak: Web-OICA ve Web-AAMA)

| | Almanya | Fransa | İtalya | İngiltere | ABD | Japonya | Diğer | Toplam |
|------|---------|--------|--------|-----------|--------|---------|--------|--------|
| 1900 | 2 | 3 | 0 | 0 | 4 | 0 | 0 | 9 |
| 1905 | 16 | 22 | 0 | 0 | 25 | 0 | 0 | 63 |
| 1910 | 13 | 38 | 0 | 14 | 187 | 0 | 3 | 255 |
| 1915 | 0 | 0 | 15 | 0 | 970 | 0 | 30 | 1.015 |
| 1920 | 0 | 40 | 21 | 0 | 2.227 | 0 | 94 | 2.383 |
| 1930 | 71 | 230 | 46 | 237 | 3.363 | 1 | 186 | 4.133 |
| 1940 | 72 | 0 | 0 | 134 | 4.513 | 51 | 172 | 4.942 |
| 1950 | 306 | 358 | 128 | 784 | 8.006 | 82 | 914 | 10.577 |
| 1960 | 2.055 | 1.370 | 645 | 1.811 | 7.905 | 814 | 1.889 | 16.488 |
| 1970 | 3.842 | 2.750 | 1.854 | 2.099 | 8.284 | 5.289 | 5.301 | 29.419 |
| 1980 | 3.879 | 3.378 | 1.612 | 1.313 | 8.010 | 11.043 | 9.330 | 38.565 |
| 1990 | 4.977 | 3.769 | 2.121 | 1.566 | 9.783 | 13.487 | 12.852 | 48.554 |
| 1995 | 4.667 | 3.475 | 1.667 | 1.765 | 11.986 | 10.196 | 16.227 | 49.983 |
| 2000 | 5.198 | 3.351 | 1.738 | 1.817 | 12.810 | 10.145 | 22.479 | 57.539 |
| 2001 | 5.692 | 3.629 | 1.580 | 1.685 | 11.425 | 9.777 | 22.602 | 56.390 |
| 2002 | 5.470 | 3.602 | 1.428 | 1.824 | 12.280 | 10.258 | 24.132 | 58.994 |
| 2003 | 5.507 | 3.621 | 1.322 | 1.847 | 12.115 | 10.287 | 25.965 | 60.664 |
| 2004 | 5.670 | 3.666 | 1.143 | 1.857 | 11.990 | 10.512 | 29.659 | 64.497 |
| 2005 | 5.758 | 3.550 | 1.039 | 1.804 | 11.947 | 10.800 | 31.585 | 66.483 |
| 2006 | 5.820 | 3.170 | 1.212 | 1.649 | 11.264 | 11.485 | 34.658 | 69.258 |
| 2007 | 6.214 | 3.020 | 1.285 | 1.751 | 10.781 | 11.597 | 38.454 | 73.102 |

1950'li yıllara gelindiğinde Avrupa ülkeleri motorlu araç üretiminde kendilerini ciddi olarak hissettirmişlerdir. 1960 yılında, ABD'nin toplam üretimdeki payı % 47,9'a düşmüş, buna karşılık Almanya'nın payı % 12,5'e çıkmış, İngiltere'ninki % 11'e, Fransa'nınki ise % 8,3'e ulaşmıştır. Japonya, 1960 yılından sonra otomotiv sanayinde çok hızlı bir gelişme

göstermiş ve 1960 yılında % 4,9 olan dünya üretimi içerisindeki payı 1980 yılında 11 milyon adetlik üretim miktarıyla % 28,6'lık bir paya ulaşmış ve motorlu araç üreticisi ülkeler içerisinde birinci sıraya yükselmiştir. Japonya'nın bu başarısında, 1970'li yıllardan sonra dünyanın en büyük motorlu araç ihracatçısı ülke olma konumu etkili olmuştur. 1980'li yılların ortalarından sonra ise, Japonya'ya benzer bir şekilde, ihracata dayalı bir büyüme başarısı Güney Kore'de gözlenmiş olup bu ülke dünya otomotiv endüstrisi içerisinde önemli bir konuma ulaşmıştır. 1980'den sonra ve özellikle 1990'lı yıllarda otomotiv endüstrisi üretiminin ABD, Japonya ve Avrupa'nın sanayileşmiş ülkelerinden gelişmekte olan diğer ülkelere doğru yayıldığı göze çarpmaktadır. Bu yayılımda etkili olan temel faktörler şunlardır (Bedir, 2002, s.4-5; Onat, 2007, s.8-9):

- Sanayileşmiş ülkelerde iç pazarın doyum noktasına ulaşması,
- Sanayileşen ülkelerde otomobil endüstrisinin kurulması için özel teşvik sistemleri uygulanarak, araçların komple ithalatı yerine ithal ikamesine dayalı montaj sanayinin kurulmasının desteklenmesi,
- Yüksek koruma ve ucuz işçilik gibi yerel olanaklar,
- Teknik mevzuatın uyumlaştırılması ve küreselleşmesi.

3.1.6. Güncel Durum

21. yüzyılın başlarında petrol ve enerji fiyatlarında meydana gelen yükselişler nedeniyle otomotiv endüstrisinde beklenen oranda artışlar sağlanamamıştır. Özellikle toplam araç satışında en büyük pazar kabul edilen ABD'de düşüşler yaşanmıştır.

2005 yılı araç satışında tüm dünyada ortalama % 3,4'lük artış kaydedilmiştir. 61 milyon adetlik taşıt aracı satış rakamının 51,5 milyon adedi binek otomobil satışından, 9,5 milyon adedi ise ticari araç satışlarından oluşmaktadır. Aynı yıl, Güneydoğu Asya bölgesinde talep patlaması yaşandığı görülmektedir. Malezya satışlarını % 15, Endonezya % 13, Tayland % 12, Filipinler % 10 oranında yükseltmiştir (OSD, 2006, s.18). Satışlara paralel olarak üretim miktarını da yükselten Asya ülkeleri dünya çapında üretimin % 9,7'sini gerçekleştirmektedir. Dünyanın en önemli yatırım ülkelerinden biri konumunda olan Çin Halk Cumhuriyeti'nde taşıt araçları üretimi 5,3 milyon adede ulaşmıştır. Avrupa Birliği (AB) ülkelerinde binek otomobil satışlarında 2005 yılında % 10'luk düşüş gözlemlenirken, ticari araç satışlarında % 10'luk artış kaydedilmiştir. Japonya taşıt araçları ihracatını % 2 artırırken, sektörde belirgin sıçramalar kaydeden Brezilya ve Arjantin'in

ihracattaki artış oranı sırasıyla % 26 ve % 24 olarak gerçekleşmiştir. (Onat, 2007, s.5).

AB istatistiklerine göre, Avrupa'da 2,2 milyon kişi otomotiv sektöründe istihdam edilmektedir. Toplam istihdamın % 6'sına tekabül eden bu rakamda düşüşler yaşandığı görülmüştür. Emek ücretinin yüksekliği ve gelişmiş sendikal haklar nedeniyle üreticiler maliyetin daha düşük olduğu doğu ülkelerine doğru yönelmektedirler. Avrupa markalarının satışları 2005 yılında % 3 gerilemiştir. Yükselen maliyetler nedeniyle ABD'de de önemli miktarlarda iş kayıpları görülmeye başlanmıştır. Ayrıca, ABD'nin en büyük otomobil tedarikçisi Delphi Corp'un iflası da sektörün kayıplar yaşamasında etkili olmuştur. En büyük üç firma olan General Motors, Ford Motor ve Daimler Chrysler'in toplam pazar payı üç yılda % 61,7'den, % 56,9'a gerilemiştir. ABD'nin üç devinin toplam otomobil satışları 2005 yılında % 2 oranında düşmüştür (Onat, 2007, s.5).

Japon otomotiv üreticileri, üretim tesislerini yeni pazarlara taşımakta ve uluslararası ortaklıklar kurmaktadır. Japonya birçok ülke ile sermaye, teknoloji ve üretim dallarında işbirliğini sürdürmekte ve yeni ortaklıklara yönelmektedir. Japon Otomotiv Üreticileri Derneği otomotiv sektöründe kaydettikleri başarının ardında, otomotiv firmalarının uluslararası ortaklık antlaşmalarını göstermektedir. Japon firmalarının uyguladığı yalın üretim sistemi ile stok maliyetleri düşürülmüş ve üretimde kalite yükselmiştir. Asya markaları satışlarının % 7 artması ve Japon firmaların sektör paylarını geçen üç yıl içerisinde % 34,6'dan, % 36,5'e yükseltmesi uygulanan yöntemlerin başarısını ortaya koymaktadır (Bloomfield, 1978, s.11).

Petrol fiyatlarındaki ani ve şiddetli dalgalanmalar nedeniyle tüketiciler dizel yakıt tipine yönelmiştir. 2004 yılından, 2005 yılına dizel motorlu araç satışı % 11 artarak 1,3 milyon adede ulaşmıştır (Onat, 2007, s.5).

3.1.7. Lider Otomobil Üreticileri

Otomobil pazarı günümüzde üç büyük ticari bölgenin hakimiyeti altındadır; AB, ABD ve Japonya. Avrupalı firmalar % 80'i aşan bir pazar payı ile AB piyasasına hakim bulunmakta, buna karşın Japonya ve ABD pazarlarında çok fazla varlık gösterememektedir. Amerikalı üreticiler, Amerikan piyasasının % 70'ine sahip olup, Japon piyasasında çok düşük seviyede bir pazar payı yakalayabilmişlerdir. Öte yandan Amerika'nın, AB'ye otomobil ihracatı çok düşük düzeyde kalmasına rağmen Amerikalı

üreticiler uzun süreden beri AB sınırları içerisinde üretim yapmaktadırlar. Nitekim bu firmalardan Ford ve General Motors, AB piyasasının 1/4'ünü ellerinde tutmaktadır. Japon üreticiler ise, Japon piyasalarında % 90'ı aşan pazar payına hakim olmakla birlikte, gerek ABD gerekse AB piyasalarında kayda değer pazar paylarına sahiptir. Bu üç büyük ticari bölge dışında özellikle 1990'lardan itibaren önemli bir gelişme kaydeden yeni piyasalar içinde Asya-Pasifik ülkeleri olarak adlandırılan Güney Kore, Çin, Tayvan ve Latin Amerika ülkeleri ile doğu Avrupa ülkelerinden oluşan grup da dördüncü büyük grup olarak kabul edilmektedir. Güney Kore firmaları, hükümetin özel ve istikrarlı plânları ile kısa sürede gelişmiş, 1994 yılından sonra özellikle eski doğu bloğu ülkelerindeki tesisleri satın almak suretiyle AB pazarı çevresinde yaklaşık 500 bin adet/yıl kapasiteli otomobil üretimi gerçekleştirmek üzere yatırımlara başlamışlardır (Onat, 2007, s.5 ve OSD, 2006). Ülke bazında bakıldığında, dünya otomotiv üretiminin 2002-2005 yılları itibariyle 10 ülkede yoğunlaştığı görülmektedir: Japonya, ABD, Almanya, Fransa, Güney Kore, İspanya, İngiltere, İtalya, Çin ve Brezilya. Bu ülkelerin oluşturduğu grup, toplam üretimin % 80'ine yakın bir üretim gerçekleştirmiştir (Türkiye Vakıflar Bankası Yayınları, 2004).

Özellikle son yıllardaki çalışmalarını sayesinde Japonya dünya otomobil üretiminde % 21' lik bir paya sahip olarak en büyük güç konumundadır. Japonya'nın ardından ise Almanya, Fransa ve İspanya gelmektedir. Bunun yanı sıra son 10 yılda yaptığı ataklar ile Güney Kore de önemli bir seviyede otomotiv ihracatı yapmaya başlamış ve bu büyük ülkeler arasında yerini almıştır. Tablo 3.2'de görüldüğü gibi, son yıllarda dünya genelinde bir üretim yükselişi olmakta ve bu yükselişte özellikle Asya ülkelerinden Japonya ve Güney Kore'nin payının gün geçtikçe arttığı görülmektedir (Dalyanoğulları, 2007, s.83).

Tablo 3.2. Dünya otomobil üretimi (Bin adet)
(Kaynak: Web-OICA)

| Ülke | 2002 | 2003 | 2004 | 2005 |
|-------------------|-------|-------|-------|-------|
| Almanya | 5.123 | 5.145 | 5.192 | 5.350 |
| Fransa | 3.293 | 3.220 | 3.227 | 3.113 |
| İspanya | 2.267 | 2.399 | 2.403 | 2.098 |
| ABD | 5.019 | 4.510 | 4.230 | 4.321 |
| Türkiye | 204 | 294 | 447 | 454 |
| Japonya | 8.619 | 8.478 | 8.720 | 9.016 |
| Güney Kore | 2.651 | 2.768 | 3.123 | 3.357 |
| İngiltere | 1.630 | 1.658 | 1.647 | 1.596 |
| Rusya | 981 | 1.010 | 1.110 | 1.068 |
| Kanada | 1.369 | 1.340 | 1.335 | 1.356 |

Dünya otomotiv üretimi giderek belirli ülkelerde olduğu gibi belirli firmalarda da yoğunlaşmakta, motorlu taşıt aracı üreten firmaların sayısı azalmakta ancak firma ölçekleri büyümektedir.

Sektördeki üretici firma sayısı son 15 yılda birleşme ve satın alma yolu ile 25 civarına inmiştir. Birleşmelerin başlıca nedenleri arasında; sınırlı ve giderek düşen kâr oranları, aşırı kapasite, küreselleşme zorunluluğu ve yeni teknolojilerin geliştirilebilmesi için giderek artan maliyetler sayılmaktadır. Gelişen rekabetçi ortamda ekonomik ölçek hayati önem taşımakta ve bu birleşmeleri hızlandırmaktadır. Bugün dünya otomotiv pazarında üretimin 25 firma tarafından yürütülmesi otomobil sektörünün global nitelikli bir sanayi dalı olduğunu göstermektedir (Uludağ İGBS, 2006, s. 4).

Tablo 3.3' de görüldüğü gibi, dünya araç üretiminde en büyük 10 üretici firma; General Motors, Toyota, Ford (Jaguar-Volvo), Volkswagen Gruppe, Daimler-Chrysler, PSA (Peugeot-Citroen), Honda, Nissan, Hyundai-Kia, Renault-Dacia-Samsung olarak sıralanmaktadır (Bloomfield, 1978, s. 9). Bu firmalar toplam motorlu araç üretiminin % 77'sini gerçekleştirmektedirler. 2005 yılı verilerine bakıldığında taşıt araçlarının % 71'inin ihraç edildiği görülmektedir (Maybek, 2006).

Tablo 3.3. Otomotiv endüstrisinde firmaların sıralaması 2005 (Bin adet)
(Kaynak: OSD, 2006, s. 2)

| Firma | Otomobil | Toplam |
|-----------------------|-----------------|---------------|
| General Motors | 5.348 | 8.314 |
| Toyota | 5.095 | 7.506 |
| Ford (Jaguar-Volvo) | 5.150 | 5.950 |
| Volkswagen | 4.960 | 5.410 |
| Daimler-Chrysler | 2.183 | 4.860 |
| PSA (Peugeot-Citroen) | 2.580 | 2.977 |
| Honda | 2.150 | 2.853 |
| Nissan | 2.290 | 2.795 |
| Hyundai-Kia | 2.218 | 2.611 |
| Renault-Dacia-Samsung | 2.105 | 2.598 |
| Ara Toplam | 34.079 | 45.874 |
| Diğer Firmalar | 9.841 | 14.016 |
| Genel Toplam | 43.920 | 59.890 |

Bu ilk 10 firmanın toplam otomobil üretimindeki payı % 77, ilk 20 firmanın payı ise % 98'dir. İlk 10 firma 59,89 milyon adetlik üretimin, 34,07 milyon adedini karşılamaktadır. İlk 10 firmanın menşesine bakıldığında ise;

- 3 ABD firması: General Motors, Ford (Jaguar-Volvo), Daimler-Chrysler
- 3 Japon firması: Toyota, Honda, Nissan
- 2 Fransız firması: PSA (Peugeot-Citroen), Renault-Dacia-Samsung
- 1 Alman firması: Volkswagen
- 1 Güney Kore firması: Hyundai-Kia

olduğu görülmektedir. General Motors, otomobil üretiminin % 12'sini, toplam üretimin de % 13,8'ini karşılayarak ilk sırada yer almaktadır. Otomobil üretiminde Ford (Jaguar-Volvo) toplam üretimin % 11,7'sini karşılayarak ikinci sırada yer alırken, Toyota % 11,6'lık üretim payıyla üçüncü sırada yer almaktadır. Toplam üretim oranlarına bakıldığında; Toyota'nın % 12,5 ile ikinci sırada, Ford (Jaguar-Volvo)'un ise % 9,9'luk üretim oranı ile üçüncü sırada yer aldığı görülmektedir (Onat, s.11-12).

3.1.8. Avrupa'daki Durumu

Otomotiv endüstrisinin Avrupa' daki durumu hakkında elde edilen bilgiler aşağıda sunulmuştur (Web-Kobifinans3, Web-Foreigntrade ve Web-ACEA).

Dünya otomotiv pazarının % 35'ine sahip olan Avrupa Birliği ülkelerinde bu endüstri dalında istihdam edilen kişi sayısının 1 milyon civarında olduğu tahmin edilmektedir.

Avrupa, kayıtlı otomobil sayısı konusunda dünya genelinin % 32'lik payına sahiptir. Avrupa'daki araç parkı 175-180 milyon adet olup en yüksek oranın Lüksemburg'a (bin kişide 548 adet) ait olduğu ve Lüksemburg'u İtalya ile Almanya'nın izlediği tespit edilmiştir. Yunanistan, İrlanda ve Portekiz ise Avrupa'da en düşük orana sahip ülkeler olarak dikkat çekmektedir. ABD'de ise bu rakamın her bin kişide 565 olduğu ifade edilmektedir.

Avrupa'da otomobil üretimi birkaç ülkede merkezileşmiştir; Almanya, Fransa, İngiltere, İspanya ve İtalya. Bunların dışında İsveç'te, Hollanda'da ve Belçika'da da üretim yapılmaktadır. Parça ve donanım üretimi de bu üretim alanları yakınındadır.

Tüketici tercihleri Avrupalıların refahının bir yansıması olarak artmaktadır. Trendler daha büyük ve daha lüks arabalara yönelmektedir. Tüketiciler daha güvenli, konforlu, ekonomik ve çevre kirliliğini azaltan ya da önleyen teçhizat içeren araçları tercih etmektedirler. Modern tüketici, kaliteli parça ve donanım talep etmektedir. Opsiyonel olarak kabul edilen birçok özellik pazarın üst kesiminde standart hale gelmektedir. Bunlar arasında da klima, airbag, abs ve elektronik benzin enjeksiyonu sayılmaktadır.

AB'de ortak para birimi olan Avro' ya geçiş ile kur farklılıkları ortadan kalktığından dolayı sınır ötesi ticaret sırasında ortaya çıkan masraflar kaybolmakta, tedarik fiyatları düşmekte, fiyatların şeffaflaşması sonucu olarak da kaynakların daha etkin dağılımı söz konusu olmaktadır.

AB genelinde trendler bir birine yakın gibi görünse de ulusal farklılıklar, gelir düzeyleri ve tüketici tercihleri otomotiv ve yedek parça pazarını etkilemektedir. Eski Doğu Bloğu ülkeleri bu alanda farklılık göstermekle birlikte önemli fırsatlar sunmaktadır.

Avrupa’ daki otomobillerin ortalama trafikte kalış süresi yaklaşık 8 yıldır. Araçların yaklaşık % 70’ i 10 yaşın altındadır. Araçların yıllık kat ettiği mesafe ortalaması yaklaşık 15.000 km’ dir. Araçların % 30’ u dizel motora sahiptir (Web-ACEA).

3.1.9. Türkiye’deki Durumu

Türk otomotiv endüstrisi, gerek yarattığı katma değer, doğrudan ve dolaylı olarak istihdama katkısı ve ithalata yönelik döviz tasarrufu gücü oranında dış ödemeler dengesi üzerinde sağladığı pozitif katkılar, gerekse sektörün içinde yer alan üretici işletmelerin ödedikleri ve sektör ürünlerin satışlarından elde edilen vergi gelirleriyle, devlet bütçesi içinde önemli bir yer tutmaktadır. Otomotiv endüstrisi bugün ulaştığı noktada, gıda ve tekstil sektöründen sonra, ülkenin üçüncü büyük endüstri dalı olarak kabul edilmektedir (Web-Kobifinans2 ve Web-Exi26)

Türkiye’nin otomobil ile ilk tanışması ve üretime başlanması yirminci yüzyılın başlarında gerçekleşmiştir. Birinci Dünya Savaşı’ndan sonra İstanbul’un işgal altında olduğu yıllarda American Foreign Trade şirketi vasıtası ile ABD’nin Ford ve Chevrolet marka otomobil ve kamyonları Türkiye piyasasına girmiştir. Yine aynı yıllarda Torino’ya bağlı olarak İstanbul’da faaliyet gösteren özel bir büro vasıtası ile İtalya’nın Fiat marka otomobilleri piyasaya girmiştir. Ford Motor Company 1929 yılında İstanbul’da ilk montaj denemesine başlamıştır. Otomobil, kamyon ve traktör üretmek üzere kurulan ilk montaj fabrikasında yapılacak üretimin bir kısmının Sovyetler Birliği’ne ihraç edilmesi öngörülmüştür. 450 işçi çalıştırılan bu fabrika, modern sayılabilecek teknolojik imkânlarla sahipti. Fabrikada günlük 48 adet otomobil ve kamyon yapım kapasitesine ulaşılmıştır. Ancak, 1930’larda yaşanan dünya ekonomi krizinin olumsuz etkileri nedeniyle hedeflenen ihracat gerçekleştirilememiştir. Bu ve diğer sebeplerden dolayı 1934 yılında fabrikada üretim durdurulmuş ve böylece ilk montaj üretim denemesi başarısızlıkla sonuçlanmıştır (Onat, 2007, s.12).

İkinci Dünya Savaşı’nın sona ermesinden sonra ülkede kentleşme hareketiyle başlayan yapısal değişimler, alt yapının geliştirilmesi ve karayollarına dönük bir ulaştırma sisteminin seçilmesi, iktidarın da otomotiv sektörünün kurulmasına destek olmasını beraberinde getirmiştir. % 100 yerli otomobil yapılması yönündeki düşünceye dayanan ve 1960 sonrası iktidarın desteğiyle de üretimine başlanmış olan “Devrim Otomobili” yapılması girişimi, üretilen ilk araçların denemesinin başarısızlıkla sonuçlanması nedeniyle

ilerleyemeden sonlandırılmıştır (Bedir, 1999).

Karayolu taşımacılığının ön plâna geçmesiyle ulaşım ağı da gittikçe genişlemiş, buna bağlı olarak da karayolu taşımacılığının yolcu taşımacılığındaki ağırlığı giderek artmıştır. 1990'lı yılların başından itibaren, gümrük korumalarının azaltılması ile yeni modellerin ülkemize giriş imkânı artmış ve buna paralel olarak yerli endüstri de yeni model üretimine yönelerek, hayli aktif olan iç pazar talebini ithal otomobiller yerine kendi ürünlerine çekmeyi amaçlamışlardır. Yine aynı dönemde, otomobil endüstri kalkınma plânlarında özel önem taşıyan sektörler arasına dâhil edilerek özellikle yatırım teşvik sistemi içerisinde özel bir konuma getirilmiştir (Onat, 2007, s.3).

Otomotiv sektörü için bin kişi başına düşen otomobil sayısı baz alındığında; Türkiye’de yüksek potansiyel talep görülmesine rağmen, efektif talebin kişi başına düşen gayri safi milli hasılaya bağlı olarak yetersiz kalması ve sektör üzerindeki vergi yükünün özellikle Avrupa Birliği ülkeleri ile karşılaştırıldığında ağır olması rekabetçi ortama uyum sağlamaya çalışan Türk otomobil endüstrisini olumsuz yönde etkilemektedir. Özellikle yabancı otomobil üreticilerinin 2000’li yıllardan sonra Türkiye’ye yaptıkları doğrudan yatırımlar ve bu yatırımlara paralel olarak gerçekleştirilen otomobil ihracatındaki artış, sektörü ülke ihracatı açısından da lokomotif konuma getirmiştir. 2004 yılında sektör, ülke ihracatında ilk sırada yer alırken, 2005 yılında az bir farkla hazır giyim ve konfeksiyon sektörünün ardından ikinci sırada yer almıştır (Bedir, 1999; Onat, 2007).

3.2. İkinci El Otomobil Pazarı

3.2.1. Tanımı ve Kapsamı

Otomotiv endüstrisinde “ikinci el (second hand)” kavramı yerine bazı ülkelerde, hatta bir ülkenin farklı bölgelerinde “kullanılmış (used)” kavramı da kullanılmaktadır (örneğin ABD). Bu çalışmada, Türkiye’ de daha çok tercih edilen “ikinci el” kavramı kullanılmıştır.

Genel bir kural olarak, bir aracın ikinci el olup olmaması, onun tescil durumuna bağlıdır. Daha önce tescilli yapılmış bir araç (hiç kullanılmamış olsa bile) ikinci el statüsünde değerlendirilir. Almanya’ daki Motorlu Araçlar Trafik Federal Dairesi (Kraftfahrt-Bundesamt, KBA), aylık olarak “yeni” ve “ikinci el” araç tescil istatistiklerini yayınlamaktadır. Bu bilgiler, yaklaşık 460 yerel makamlarca toplanan tescil verilerinden

elde edilmektedir. Bu tescillerde kullanılan formlar, KBA' ya rapor edilecek ilk tescil yılının bildirilmesini isteyen bir alan içermektedir (Ilgaz, 2005, s.45-46 ve Web-KBA, 2004).

AB Birleşik Terminolojisi (EU Combined Nomenclature), yeni araç için “daha önce hiç tescili yapılmamış araç”, ikinci el araç için ise “en azından bir kere tescil ettirilmiş araç” tanımını kullanmaktadır (Ilgaz, 2005, s.46).

Wikipedia' da ise “ikinci el otomobil” için şu tanımlama yer almaktadır: “Önceden sahip olunmuş” ya da “ikinci el” olarak da bilinen “kullanılmış araç”, daha önce bir ya da fazla kişinin sahip olduğu araçtır (Web-Wikipedia4, 2008).

İkinci el araç pazarlarının kapsamına pek çok ülkede otomobillerin yanında zaman zaman ticari araçlar da dahil edilmekle birlikte, otomobiller pazarın çok büyük bir hacmini işgal etmektedir. Otomobillerin ve ticari araçların fiyat farklılıklarının ya da değişimlerinin farklı eğilimlerde olacağı varsayımından hareketle bu çalışma kapsamında incelenen araçlar sadece otomobillerden oluşmaktadır.

3.2.2. Önemi

Pek çok ülkede ikinci el otomobil satış hacmi, yeni otomobil satış hacminden çok daha büyük hale gelmiş durumdadır. Bu da ikinci el araç pazarlarının önemini giderek artırmaktadır. Örneğin ABD ekonomisinde en büyük perakende sektörünü oluşturan ikinci el otomobil pazarının hacmi, yeni otomobil pazar hacminin iki katından daha fazladır ve sürekli olarak büyümeye devam etmektedir (Lee, 2006, s. 120). 2005 yılında ABD' de satılan ikinci el otomobil sayısı 44 milyon, satış değeri ise 370 milyar dolardır (Web-Wikipedia4, 2008).

Türkiye, AB ile yapılan gümrük birliği anlaşmasında her türlü ikinci el motorlu aracın Türkiye'ye ithalini sınırsız olarak durdurmuştur. AB, bu kararı zaman veya alan bakımından sınırlandırmak istemişse de Türkiye AB' ye girene kadar bu kararından vazgeçmeyeceğini bildirmiştir. Türkiye gibi dünyanın pek çok ülkesinde de halen ikinci el araçların ithalinde çeşitli kısıtlamalar, ya da yüksek vergiler uygulanmaktadır. Bu uygulamalar, ikinci el araçların ülke ekonomileri için önemini göstermektedir.

Ayrıca günümüzde ikinci el otomobil pazarının potansiyelini fark eden neredeyse tüm büyük otomotiv üreticisi markaları, bu pazara girmişler ya da girmeye çalışmaktadırlar. Bu da ticari açıdan ikinci el otomobil pazarının cazibesini ortaya koymaktadır.

3.2.3. Tüketicilerin İkinci El Otomobiller Hakkındaki Tutumları

Genel olarak otomobil sahibi olma isteği kentli yaşam tarzının benimsenmesiyle birlikte yükselmektedir. İkinci el otomobil pazarına olan talebi artıran en önemli faktör ise satın alma gücü parametresidir. Satın alma gücü düştükçe ikinci ele olan talep artmaktadır, diğer bir deyişle, yeni araç almaya gücü yetmeyen tüketiciler kullanılmışını tercih etmektedir (Web-Capital1).

Diğer yandan, yeterli satın alma gücüne sahip tüketicilerin bir kısmı da sıfır araçlardaki ilk alım vergilerinin getirdiği yükten kurtulmak amacıyla ikinci el araçlara talep göstermektedir. Bu tüketiciler, genellikle bir yaşın altında olan çok az kullanılmış araçları tercih etmektedirler. Böylece, aracı tekrar satmaya kalktıklarında, ilk alım vergilerini zaten ödemedikleri için zarar etmemekte ya da yeni araç alan tüketicilere göre çok daha az zararla araçlarını değiştirebilmektedirler. Ayrıca, otomobil kullanım vergilerinin aracın yaşıyla orantılı olarak düşmesi de özellikle satın alma gücü düşük tüketicileri, yaşı büyük olan ikinci el araçlara yöneltmektedir.

Uzmanlar, ikinci el satışlarının; yeni otomobil satışı, otomobil parkı ve içinde yaşadığımız ekonomik gelişmelerle doğru orantılı olduğunu ifade etmektedir. Makro-ekonomik dengelerde ortaya çıkan ani değişiklikler ise talebin geçici bir süre için canlanmasına veya durağanlaşmasına neden olabilmektedir. Örneğin, faiz ve vergi oranlarının yükselmesi, özellikle otomobil gibi dayanıklı tüketim malları talebinin ertelenmesine yol açmaktadır. Tüketiciler faizden kazanabilecekleri getiriden mahrum kalmamak için alım kararlarını ertelemeyi tercih edebilmektedirler. Ülkedeki gelir dağılımının düzelmesi, sadece ekonomide istikrarlı bir büyüme dönemine girilmesini sağlamakla kalmayıp talebi de artırmaktadır. Ülke içindeki gelir dağılımının bozulması lüks ürünlere (örneğin yazlıklar, villalar) talebi artırırken, gelir dağılımının düzelmesiyle birlikte orta ve ortanın üstü gelir gruplarının toplam gelirden aldığı pay artmakta ve otomobil gibi dayanıklı tüketim mallarının satışları yükselmektedir (Web-Capital1).

Özellikle birinci el otomobillerde başlatılan çeşitli kampanyaların, düşük faizli kredilerin ve özel indirimlerin ikinci el otomobil pazarına olan talebi beklenmedik bir şekilde aniden düşürebildiği görülmektedir. Bunun yanında, ikinci el piyasasında özellikle fiyatlandırma konusunda yaşanan önemli dalgalanmalar, ikinci el otomobil almak isteyenlerin fikirlerini olumsuz yönde etkilemektedir. Sektör temsilcilerine göre ikinci el otomobillerin eskisi gibi prim yapmaması, tüketicilerin sadece ihtiyaç duydukları zamanlarda otomobil almalarına ve ihtiyaçları geçer geçmez de elden çıkarmalarına neden olmaktadır. Tüm bunların sonucunda da insanlar, kullanılmış otomobilleri bir yatırım aracı olarak görmemeye başlamakta ve ilgi göstermemektedir. Uzmanlar, buna rağmen özellikle son dönemde yakalanan ekonomik istikrar ile fiyat dalgalanmalarının eskisi kadar büyük olmadığını belirtmektedirler (Web-Kobifinans4).

İkinci el otomobil konusunda tüketicilerin taleplerini azaltan önemli bir etmen de pazardaki satıcılara duyulan güvensizliktir. Satın alınan otomobilin, tüketiciden gizlenen önemli bir kusurunun olması bir süre sonra tüketicide hayal kırıklığına neden olmaktadır. Bu konuda mağdur olan tüketiciler, ya yeni araçlara yönelmekte ya da otomobil sevdasından tamamıyla vazgeçmektedir. Bu önemli sorunu ortadan kaldıracılabilecek en önemli adım ise, ABD ve Avrupa’ da çok yaygın olan, ülkemizde ise yeni filizlenen “kurumsallaşma”dır (Web-Capital2).

3.2.4. Pazardaki Kurumsallaşma

İkinci el otomobil pazarındaki kurumsal firmalar tüketicilere daha detaylı ve kaliteli hizmet sunmayı hedeflemektedir. Bu hizmetin maliyeti, diğer satıcılarınkine göre çok daha yüksek olmakla birlikte, tüketiciler tarafından daha fazla tercih ediliyor olmak bu dezavantajı ortadan kaldırmaktadır.

Kurumsal firmaların bir başka amacı ise ikinci el araba satışlarında her zaman gündemde olan hukuki riskleri tespit etmektir. Örneğin, satın alan müşteriyi suçlu duruma düşürecek şekilde, üzerinde rehin bulunan araçlara, hatta çalıntı araçlara rastlanabilmektedir. Kurumsal firmalar, kendi araç satışları için bu tespitleri yaptığı gibi, bağımsız olarak üçüncü şahıslara da belirli bir ücret karşılığında bu hizmeti sunabilmektedirler (Web-Hurriyet1).

Kurumsal firmalar ikinci el otomobilleri bünyesine katmadan önce çeşitli sağlık ve

güvenlik testleri uygulayıp bu testleri geçemeyen araçları satın almamaktadırlar. Aynı şekilde stoklarında bulunan araçları da müşterilerine satmadan önce benzer testleri uygulayıp, araç testi geçebilmişse satışı yapmaktadırlar. Örneğin General Motors' un ikinci el otomobil satış markası olan G2, web sitesindeki açıklamasında araçların satın alma öncesindeki kontrollerine ek olarak teşhirden önce tam 50 noktadan kontrol ve yenileme işlemine tabi tutulduğunu, ayrıca satış sonrası belirli bir süre mekanik garantisi de sunduklarını belirtmektedir. Böylece müşteri; aracın belirli kriterlerdeki güvenliğinden emin olmakta, araçta kendisinden gizlenen önemli bir kusur bulunabileceği tereddüdünü taşımamakta ve kullanım esnasında ortaya çıkabilecek beklenmedik hasar risklerinden kurtulmuş olmaktadır.

Kurumsal satıcıların sunduğu önemli bir avantaj da takas imkânıdır. Artık müşteriler otomobil ihtiyaçlarını tek bir çatı altında karşılamak, bir sorun yaşadıklarında ya da araç değiştirmeye karar verdiklerinde karşılarında güven duyacakları tek bir muhatap görmek istemektedirler. Örneğin aracını değiştirmek isteyen bir müşteri, oto pazarlarına çıkıp uzun süre aracını satmakla ve sattıktan sonra yeni bir araç almakla uğraşmak yerine, kurumsal satıcılara gidip uygun gördüğü bir araçla kendi aracını değiştirebilmektedir. Sadece iki aracın fiyatı arasındaki fark kadar ödeme yaparak zamandan tasarruf etmekte ve iki defa alım-satım değer kaybından kurtulmuş olmaktadır.

Kurumsal satıcılar, bankalarla yaptıkları özel anlaşmalar sayesinde müşterilerine düşük faizli otomobil kredisi kullandırtabilmektedirler.

Bunların dışında, kurumsal satıcıların bir bölümü, satılan araçlar için temin edilen kullanım kılavuzu, servis el kitabı, garanti belgesi gibi tüm önemli belgeleri içeren bir “doküman çantası”, garanti süresince yol yardımı (acil servis, çekici, kurtarma, vb.), kiralık araç, konaklama ve tıbbi yardım gibi bir takım hizmetler de sunabilmektedir.

İkinci el otomobil pazarındaki gelişimin yolu kurumsal hizmetten geçmektedir. Gelişmiş ülkelerde ikinci el otomobil pazarının potansiyelini ve yeni araç satışlarına etkisini gören büyük otomobil üreticisi markalar artık ikinci el pazarında da yer almaya başlamışlardır.

3.2.5. Otomobil Borsası

İkinci el otomobil pazarında son yıllarda ilgi çeken bir satış platformu oluştu: Otomobil borsası. Bu borsa, bireysel müşterilerden ziyade kurumsal firmalar arasında işlemektedir. Galeriler, yetkili satıcılar ve filo sahibi kuruluşlar arasında özel artırma sistemiyle ikinci el otomobil ticareti gerçekleştirilmektedir. Ayrıca mevcut hukuki ve ticari düzenlemeler çerçevesinde alıcı ve satıcıları güvenli bir platformda buluşturarak, ikinci el motorlu araç piyasasının kurumsallaştırılması sürecine katkıda bulunmaktadır (Web-BTinsan).

3.2.6. İkinci El Otomobil Alırken Dikkat Edilmesi Gerekenler

İkinci el otomobil pazarlarında, kurumsal satıcılardan alınmayan araçlarda müşteriden saklanan ve daha sonra sorun çıkartan pek çok kusur olabilmektedir. Maalesef ülkemizde bu konuda çok fazla suistimal yaşandığı bilinen bir gerçektir. Bu nedenle, ikinci el otomobil alımında aşağıdaki uzman önerilerini dikkate almakta fayda bulunmaktadır (Web-Basgaza ve Web-Mynet):

- Özellikle kaportadaki kusurların anlaşılabilmesi için mutlaka gün ışığı gerekmektedir. Akşam saatlerinde ya da suni ışık altında yapılan gözlemler yanıltıcı olabilir. Bu nedenle, satın alınacak otomobili mutlaka gündüz incelemek gerekmektedir.
- Mümkünse satın almaya bir kaporta ve motor uzmanı ile gidilmesi önemle tavsiye edilmektedir. Zira araçlardaki pek çok kusur, normal bir kullanıcı tarafından fark edilemeyebilir.
- Karar verilen marka ve model otomobil için çok yönlü bir fiyat araştırması yapılmasında fayda bulunmaktadır. Bu amaçla; ikinci el otomobil e-ticaret siteleri, oto galericileri, gazete ilanları ve açık otomobil pazarları araştırmaya tabi tutulabilir.
- Otomobilin çalıntı olup olmadığını anlamak için motor ve şase numaralarını ruhsattaki motor ve şase numaraları ile karşılaştırmak gerekir. İthal otomobil alırken gümrük kâğıdının veya ithal belgesinin mutlaka kontrol edilmesi gereklidir, çünkü ruhsat sahte olabilmektedir. Ayrıca, otomobilin hacizli olup olmadığını ve trafik ile vergi borçlarını da mutlaka araştırmak gerekir. Çalıntı, hacizli ya da

borçlu otomobiller, tüketicinin parasını kaybetmesine neden olacağı gibi, hukuki olarak da başına dertler açabilmektedir.

- Araçla viraj, yokuş gibi değişik yol tiplerinde test sürüşü yapmak gerekir. Aracın güzel fiziksel görünüşüne karşın önemli mekanik kusurları olabilir.
- Son yıllarda kötü niyetli kişiler, araç kiralama firmalarından kiraladıkları araçları pazarda satarak haksız kazanç elde etmeye çalışmaktadır. Bu otomobilleri alan kişiler mağdur durumda kalmaktadırlar. Bu nedenle araç çalıntı görünmese dahi, sicilini mutlaka araştırmak gereklidir. Satıcı kişi ruhsat sahibi değilse, ruhsat üzerinde adı geçen kişiyle irtibata geçip aracın geçmişi hakkında bilgi almakta yarar vardır.
- Kaportadaki en büyük kusurlar; darbe, çürüme ve sonradan boyamadır. Darbeli araçların, piyasa fiyatı normallerden daha düşük, satılması zor olmaktadır. Kaportada gözle görülür bir darbe olmasa da, önceden meydana gelmiş bir darbenin sonucunda kaportanın düzeltilerek ya da darbeli parçanın değiştirilerek boyanması bile sorun olabilmektedir. Aracın kaportasındaki hataların uzman kişiler tarafından düzeltilip boyanması durumunda bu sorun sıradan tüketiciler tarafından çoğunlukla anlaşılammaktadır. Bu nedenle özel cihazlara sahip tamirciler ya da uzmanlardan yardım almakta fayda vardır.
- Aracın motoru mutlaka kontrolden geçirilmelidir. Bunun için motor kapağı açılıp yağ akıtması kontrol edilmeli ve contaların yağ sızdırıp sızdırmadığına bakılmalıdır. Ayrıca motoru çalıştırıp sesini dinlemek gerekir. Motor, soğukken çalıştırıldığında rölantisi düzgünse ve uğultu yapmıyorsa bu iyiye işarettir.
- Önemli bir parça olan radyatörün ve motor hortumlarının su kaçırmaya başlamadığına da bakılması gerekir. Eğer radyatör suyu paslı ve yağlıysa sorun oluşturabilir.
- Akünün kontrolü esnasında, akü başlarının okside olup yağlandığı fark edilmesi akünün ömrünün azaldığı anlamına gelmektedir.
- Aracın yaptığı kilometre, otomobilin yıpranmasıyla doğru orantılı olan en belirgin değişkendir. O nedenle, kilometresi düşük olan araçları tercih etmek gerekir. Fakat özellikle eski tip model arabaların kilometre göstergesinde kolaylıkla oynama yapılabilmektedir. Bunu anlamak için öncelikle servis/bakım kitapçığı incelenerek geçmişteki bakımlar kontrol edilmeli, gerekirse bir uzmana ya da yetkili servise danışılmalıdır.

Aracı satın aldıktan sonra, derhal yetkili ya da uzman bir servise götürüp komple bir bakımdan geçirilmesi gereklidir.

3.2.7. İkinci El Otomobil Pazarının Avrupa' daki Durumu

İkinci el otomobil pazarı, hemen her ülkede hacmi ve değeri giderek yükselen bir potansiyele sahiptir. Avrupa ikinci el otomobil pazarında 2007 yılı itibariyle yapılmış araştırma sonuçlarından derlenen bilgiler aşağıda sunulmuştur (Datamonitor, 2008).

Avrupa ikinci el otomobil pazarı 2003-2006 yılları arasında yavaş fakat istikrarlı bir büyüme göstermiştir. Bu gidişatın sonraki beş yılda da devam edeceği beklenmektedir.

Pazar değeri, 2003-2006 yılları arasındaki yaklaşık % 2 civarındaki bileşik yıllık büyüme oranıyla 2006 sonunda 263 milyar dolar değerine ulaşmıştır. 2007 – 2012 arasında % 11,3 büyüyerek 298,5 milyar dolar değerine ulaşacağı tahmin edilmektedir.

Pazar hacmi, 2003-2006 yılları arasındaki yaklaşık % 1,3 civarındaki bileşik yıllık artış oranıyla 2006 sonunda yaklaşık 28,3 milyon birim hacme ulaşmıştır. 2007 – 2012 arasında % 7,9 büyüyerek 31 milyon birim hacme ulaşacağı tahmin edilmektedir.

Dağıtım açısından, 2006 sonu itibariyle, özel satışların payı % 38,4, franchise ile çalışan satıcıların payı % 37,3, bağımsız satıcıların payı % 24,3 civarındadır.

Pazar segmentasyonunda özel satışlar (alıcı ve satıcının özel şahıslar olduğu), % 56,8 oranına karşılık gelen 16,3 milyon birim hacim ile en geniş hacme sahip segmentasyonu oluşturmaktadır.

Ülkeler bazında Almanya, bölgesel pazar değerinin % 26,3 oranındaki değeri ile en geniş pazar konumundadır. Almanya' yı sırasıyla % 23,7 ile İngiltere, % 19,6 ile Fransa ve % 16,5 ile İtalya takip etmektedir.

Makroekonomik göstergelerden döviz kurunun seyri (Dolar/Euro oranıyla) 2003' de 1,12, 2004' de ve 2005' de 1,24, 2006' da 1,25' dir.

3.2.8. İkinci El Otomobil Pazarının Almanya’ daki Durumu

Avrupa otomobil pazarının lideri konumundaki Almanya’ nın ikinci el otomobil pazarında 2007 yılı itibariyle yapılmış araştırma sonuçlarından derlenen bilgiler aşağıda sunulmuştur (Datamonitor, 2008).

Pazar değeri, 2003-2006 yılları arasındaki yaklaşık % 0,6 civarındaki bileşik yıllık büyüme oranıyla 2006 sonunda yaklaşık 70,2 milyar dolar değerine ulaşmıştır. 2007 – 2012 arasında % 0,1 büyüyerek 71 milyar dolar değerine ulaşacağı tahmin edilmektedir.

Pazar hacmi, 2003-2006 yılları arasındaki yaklaşık % 0,2 civarındaki bileşik yıllık artış oranıyla 2006 sonunda yaklaşık 6,7 milyon birim hacme ulaşmıştır. 2007 – 2012 arasında % 1,3 büyüyerek 7,3 milyon birim hacme ulaşacağı tahmin edilmektedir.

Dağıtım açısından, 2006 sonu itibariyle, özel satışların payı % 41,1, franchise ile çalışan satıcıların payı % 40,3, bağımsız satıcıların payı % 18,6 civarındadır.

Pazar segmentasyonunda özel satışlar (alıcı ve satıcının özel bireylerden oluştuğu), % 72,1 oranına karşılık gelen 4,9 milyon birim hacim ile en geniş hacme sahip segmentasyonu oluşturmaktadır.

Makroekonomik göstergelerden döviz kurunun seyri (Dolar/Euro oranıyla) 2003’ de 1,12, 2004’ de ve 2005’ de 1,24, 2006’ da 1,25’ dir. Almanya nüfusu, 2003-2006 yılları arasında çok fazla değişmeden yaklaşık 82,4 milyon kişi olarak kalmıştır. Almanya Gayri Safi Yurtiçi Hâsılası (GSYİH) 2003-2006 yılları arasında sırayla, 1920, 1951, 1970 ve 1983’ dür. Almanya enflasyonu ise yine aynı yıllar arasında sırayla, “1,1”, “1,6”, “2,0”, “1,1” dir.

3.2.9. İkinci El Otomobil Pazarının Türkiye’deki Durumu

Türkiye’de gelir seviyesi çok yüksek olmadığından ve yeni araçlardaki vergi yüksekliğinden dolayı sıfır kilometre araç sahibi olmanın kolay olmadığı bilinmektedir. Bununla beraber mutlaka bir otomobil ihtiyacı olup da yenisine ekonomik gücü yetmeyenler ikinci eli tercih etmektedir. Türkiye’ de ikinci el oto piyasası denilince, terkedilmiş bir arsada rastgele dizili otomobiller, üzerlerine acemice “satılık” yazılarak

otomobil camlarına yapıştırılmış kâğıtlar ve kendi aracını kendi pazarlamaya çalışan satıcılar akla gelmektedir. Biraz daha düzgün olanlarında ise çevrelenmiş bir satış alanı ve organizasyonu yapan bir oto galerisi devrededir. Maalesef Türkiye’deki ikinci el pazarı, şimdilik profesyonellikten oldukça uzak görünmektedir (Web-Hürriyet1).

Buna bağlı olarak, Türkiye’deki ikinci el otomobil satışları da gelişmiş ülkelere göre çok yetersiz seviyede seyir etmektedir. Özellikle ABD ve Avrupa pazarları verileri ile karşılaştırıldığında bu yetersizlik çok daha açık biçimde ortaya çıkmaktadır. Örneğin ABD’de bir sıfır kilometre otomobile karşılık 3.4 ikinci el araç, Avrupa’da ise 2.3 araç dolaşıma girerken Türkiye’de bu oran 0.75 civarındadır.

Ülkemizde ikinci el otomobil pazarında fiyat ve satış şartları hususunda referans alınacak bir otoritenin bulunmaması önemli bir eksiklik teşkil etmektedir. Bu nedenle pazardaki fiyatlar tamamen ikili ilişkilerle, yani satıcının satış yeteneği ve alıcının pazarlık gücü ile belirlenmekte olduğu için, sistem etkin bir biçimde işlememektedir. Türkiye’deki ikinci el otomobil pazarı büyük çoğunlukla oto galericilerinin üzerinden işlemekte, fiyatlar serbest piyasa koşullarında kendiliğinden belirlenmektedir. Pazardaki fiyat dalgalanmaları oldukça fazla olup bu durum sistemi suistimallere açık hale getirmektedir. Özellikle kayıt dışılık konusunda önemli açıkların bulunması mümkündür.

Sistemin etkin işleyememesinin yanında, büyük oyuncular henüz ikinci el pazarının önemini yeterince kavrayamamış görünmektedir. İkinci el otomobil pazarı, finansal getirisinin yanında "müşteri sadakati"ni artırmada önemli bir rol oynayabilmektedir. Yeni otomobil almaya gücü yetmeyen ya da otomobile çok fazla para harcamak istemeyen bir tüketici ikinci el bir araba aldıktan sonra, şayet bu araçtan memnun kaldıysa, aracı ileride değiştirmek istediğinde kendi aracını satarak, aynı markanın yeni modelini almak isteyebilmektedir. Ayrıca, ikinci el piyasası yüksek olan markaların tüketici gözündeki itibarı nedeniyle yeni araçlarının da değeri göreceli olarak yükselmekte ve satışları hız kazanmaktadır. Ancak Türkiye’deki sektörde ise bu durumun çok fazla dikkate alınmadığı görülmektedir. Otomobil pazarındaki büyük üreticiler, asıl getiriye kurulan servis ağlarında verilen satış sonrası hizmet ve yedek parçadan elde etmekte, kendilerine yeterince para kazandırmayan ikinci el pazarına çok önem vermemektedirler (Web-Businessweek).

Bu olumsuz tabloya rağmen, son yıllarda kurumsallaşma konusunda artık Türkiye’de de önemli adımların atılmaya başlandığı söylenebilir. Bugün için Türkiye’deki ikinci el

otomobillerin yaklaşık % 80' i pazarlarda ve galerilerde satılmaktadır. Geriye kalan kurumsal yapılara güvenen müşteri sayısı ise her geçen gün artış göstermektedir. İnsanlar önceki yaşadıkları hoş olmayan araç satın alma tecrübeleri sayesinde daha hızlı bilinçlenmektedir. Türkiye'deki belli başlı kurumsal satıcılar şunlardır:

- Doğu Oto Değerlendirme
- Renault2
- Borusan
- General Motors G2
- Toyotasa
- Kosifler Oto

İkinci el otomobil pazarına büyük firmaların yatırım yapması ve kurumsallaşmalarının artması ile pazarın kısa zamanda büyüyerek arzu edilen seviyelere gelmesi mümkün görünmektedir.

3.2.10. İkinci El Otomobil Pazarında E-Ticaret Yoluyla Satış

21. yüzyıl ile birlikte çağ atlayan ve artık neredeyse her ürünün pazarlandığı internet, ikinci el otomobil pazarını da bünyesine katmıştır. Yakın zamana kadar sadece otomobil pazarlarında alım satımı yapılan ikinci el araçlar, artık internet üzerinden de müşterilerin beğenisine sunulmaya başlanmıştır.

Otomobilini satmak isteyen bireyler ya da firmalar, araçlarla ilgili akla gelebilecek ve müşterinin bilmek istediği tüm özelliklerinin yanında çeşitli fotoğraflarını da web sitelerine ekleyerek küresel pazara açılmaktadırlar. İkinci el otomobil satın almak isteyen müşteriler ise, araç özelliklerine göre çok detaylı sorgulamalar yapıp, binlerce araç içerisinde sadece ilgi duydukları ya da bütçesine uygun olanları görüntüleyebilmekte ve satıcılarla irtibata geçebilmektedirler.

Türkiye'deki en büyük ikinci el otomobil e-ticaret sitelerinden "araba.com.tr" sitesi genel müdürü Selma Elmacı Bağcı, ikinci el araç almak isteyenlerin artık araba pazarları yerine internet sitelerini tercih ettiğini belirtip konunun önemini şu sözleriyle ifade etmektedir: "Eskiden müşteri gittiği pazarda bir kaç araçla sınırlı kalıyordu. Seçenek az olduğu için belki de beğenmediği bir aracı almak zorunda kalıyordu. Ancak araba satışı

yapan siteler bu olumsuzluğu giderdi. Bünyesinde yüzlerce galeri ve binlerce araç bulunduran siteler, müşteriye çok değişik seçenekler sunuyor. Sitenin, 'araç arama' bölümüne giren bir kişi, renk, özellik ve fiyatta binlerce alternatifle karşılaşır ve bütçesine en uygun aracı alabiliyor.” İnternet üzerinden hizmet veren sitelerin genellikle galericiyle müşteri arasında köprü kurduğunu belirten Bağcı, her yıl binlerce aracın bu yolla satıldığını belirtmektedir. Kuruldukları 2003 yılında bir kaç yüz kullanıcıya hizmet verdiklerini ifade eden Bağcı, 5 yıl içinde yaklaşık 452 bin üyeye ulaştıklarını ifade etmektedir (Web-Hürriyet2).

Dünyada olduğu gibi Türkiye’de de son yıllarda sayıları iyice artan ikinci el otomobil web sitelerinin en önemli katkısı, alıcı ve satıcıların geniş kitlelere ve otomobil portföyüne ulaşmalarına fırsat vererek zamandan tasarruf etmelerini sağlamaktır.

Avrupa’da ikinci el otomobil e-ticaret siteleri Türkiye’dekilere göre çok daha yoğun kullanılmaktadır. Bunda bilgisayar ve internet kullanım oranı en büyük faktörü oluşturmaktadır.

Avrupa’ da ve Türkiye’ de en çok ziyaret edilen ve en fazla araç portföyüne sahip ikinci el otomobil web siteleri incelenerek bu sitelerin profili çıkarılmış ve aşağıda sırayla sunulmuştur.

3.2.10.1. Avrupa’ daki Önemli İkinci El Otomobil E-Ticaret Siteleri

1. Autoscout24.de

Almanya’ nın en büyük ikinci el araç sitesidir. Veritabanında 136 marka ve 1.700.000 üzerinde araç bulunmaktadır. İlandaki araçlara ait teknik detay sayısı 49’ dur. Temel sorgulama kriterleri şu alanları içermektedir: Marka, araç tipi, temel ve detay model, model yılı aralığı, kilometre aralığı, renk, yakıt, vites tipi, yetkili satıcı, garanti durumu, en yüksek ve en düşük fiyat aralığı, donanım bilgileri (ABS, ESP, klima, vb.) ve koltuk kaplaması şekli. Ayrıca aracın daha önce kaza geçirip geçirmediği hakkında da sorgulama yapılabilmektedir.

2. Mobiliti.de

Almanya'nın ikinci sıradaki en büyük ikinci el araç sitesidir. Avrupa'dan yaklaşık 90.000 müşterisi bulunmaktadır ve bunların 30.000'i Almanya'dandır. Veritabanında toplam 132 marka yer almaktadır. İlandaki araçlara ait teknik detay sayısı 32' dir. Temel sorgulama kriterleri şu alanları içermektedir: Marka, araç tipi, temel ve detay model, model yılı aralığı, kilometre aralığı, renk, yakıt, vites tipi, yetkili satıcı, garanti durumu, kaza geçmişi, en yüksek ve en düşük fiyat aralığı ve donanım bilgileri.

3. Paruvendu.fr

Fransa'nın en büyük ikinci el araç sitesidir. Veritabanında 206.000 üzeri listelenen araç bulunmaktadır. Temel sorgulama kriterleri şu alanları içermektedir: Marka, araç tipi, temel ve detay model, model yılı aralığı, kilometre aralığı, renk, yakıt, vites tipi, yetkili satıcı, garanti durumu, v.b. İlandaki araçlara ait teknik detay sayısı 24' dür.

4. Viaturasonline.pt

Veritabanında 115 farklı marka ve 36.000 üzeri listelenen araçla Portekiz' in en büyük ikinci el araç sitesidir. Temel sorgulama kriterleri şu alanları içermektedir: Marka, araç tipi, temel ve detay model, model yılı aralığı, kilometre aralığı, renk, yakıt, vites tipi, yetkili satıcı, garanti durumu, kapı sayısı, v.b. İlandaki araçlara ait teknik detay sayısı 29' dur.

5. Segundamano.es

Veritabanında 916.000 üzeri listelenen araçla İspanya' nın en büyük ikinci el otomobil sitesidir. Temel sorgulama kriterleri şu alanları içermektedir: Yıl aralığı, kilometre aralığı, ve bulunduğu şehir. İlandaki araçlara ait teknik detay sayısı 5-10 arasında değişmektedir.

6. DerStandard.at

Avusturya' nın en büyük ikinci el otomobil sitesidir. Veritabanında 95 farklı marka ve 32.000 üzeri listelenen araç yer almaktadır. Temel sorgulama kriterleri şu alanları içermektedir: Marka, model, yıl aralığı, kilometre aralığı ve donanım detayları. İlandaki

araçlara ait teknik detay sayısı 30 civarındadır.

7. Autoscout24.ch

İsviçre' nin en büyük ikinci el araç sitesidir. Veritabanında 136 marka ve 114.000 üzerinde araç bulunmaktadır. Temel sorgulama kriterleri şu alanları içermektedir: Marka, araç tipi, temel ve detay model, model yılı aralığı, kilometre aralığı, renk, yakıt, vites tipi, yetkili satıcı, garanti durumu, kaza durumu, koltukların kaplaması, en yüksek ve en düşük fiyat aralığı ve donanım bilgileri. İlandaki araçlara ait teknik detay sayısı 49' dur.

3.2.10.2. Türkiye' deki Önemli İkinci El Otomobil E-Ticaret Siteleri

1. Araba.com.tr

Sitede toplam 115 marka, 50.000 çeşit araç, 456.377 otomobil ilanı, 262.126 resimli ilan, 604.716 bireysel üye ve 5.213 oto galeri üyesi vardır. temel sorgulama kriterleri şu alanları içermektedir: Marka, araç tipi, temel ve detay model, model yılı aralığı, kilometre aralığı, renk, yakıt, vites tipi, silindir hacmi (cm³ olarak), en yüksek ve en düşük fiyat aralığı, donanım bilgileri (ABS, ESP, klima, vb.). İlandaki araçlara ait teknik detay sayısı 26' dır.

2. Otogazete.com.tr

Sıfır ve ikinci el otomobil ilanlarıyla, araba alıcı ve satıcılarını bir araya getirmek üzere yayın hayatına başlayan site, aynı zamanda otomobil kullanıcılarının ihtiyacı olabilecek yedek parça, servis, sigorta, vergi vb. hizmetlerini veren kurum ve kuruluşlar hakkında da içerik sunmaktadır. Sitede toplam 63 marka, 6.300 çeşit araç bulunmaktadır. Temel sorgulama kriterleri şu alanları içermektedir: Marka, araç tipi, temel ve detay model, silindir hacmi (cm³), model yılı aralığı, fiyat aralığı, kilometre aralığı, renk, yakıt, vites tipi, satıcının üyelik durumu (üye olan ve üye olmayan arasında). İlandaki araçlara ait teknik detay sayısı 23' dür.

3. Otomax.com

Borusan A.Ş. nin bir kuruluşu olan Otomax.com sitesinde toplam 108 marka, 6.200

çeşit araç bulunmaktadır. Temel sorgulama kriterleri şu alanları içermektedir: Marka, araç tipi, temel ve detay model, model yılı aralığı fiyat aralığı, kilometre aralığı, renk, yakıt, vites tipi, satıcının üyelik durumu. İlandaki araçlara ait teknik detay sayısı 27' dir.

4. Sahibinden.com

Sahibinden.com, Türkiye' nin en büyük genel alışveriş siteleri arasında yer almaktadır. Otomobil dışında emlak da dahil olmak üzere, hemen her çeşit ürünün alım satımı yapılmaktadır. Sitede ikinci el alanında toplam 90 marka, 118.000 ilan bulunmaktadır. Temel sorgulama kriterleri şu alanları içermektedir: Marka, araç tipi, temel ve detay model, model yılı aralığı fiyat aralığı, kilometre aralığı, renk, yakıt, vites tipi, satıcının üyelik durumu. İlandaki araçlara ait teknik detay sayısı 27' dir.

5. Dod.com.tr

Doğuş Otomotiv' e ait olan bu sitedeki temel sorgulama kriterleri şu alanları içermektedir: Marka, araç tipi, temel ve detay model, model yılı aralığı, kilometre aralığı, renk, yakıt, vites tipi, yetkili satıcı, garanti durumu, en yüksek ve en düşük fiyat aralığı ve donanım bilgileri. İlandaki araçlara ait teknik detay sayısı 31' dir.

Dördüncü Bölüm

4. İKİNCİ EL OTOMOBİL FİYATLARININ TAHMİN EDİLMESİ: AVRUPA ÜLKELERİ UYGULAMASI

4.1. Araştırmanın Amacı, Kapsamı ve Önemi

İkinci el otomobil pazarları ile yeni otomobil pazarları arasında çok yakın bir ilişki olduğu bilinmektedir. Her ne kadar ikinci el otomobil pazarları yeni otomobil pazarlarına belirli bir oranda bağımlı olsa da, yine de kendilerine özgü birtakım özellikleri mevcuttur. Genel olarak, ikinci el otomobillerin tedarikinin esnek olmamasına ve pazar dışı faktörlerden etkilenmesine rağmen talep edilmeleri en az yeni araçlardaki kadar esneklerdir. Bu nedenle de, ikinci el otomobillerin fiyatlarının, yeni otomobil fiyatları ile karşılaştırıldığında çok daha değişken olduğu görülmekte, buna bağlı olarak da tahmin edilmesi zorlaşmaktadır (Lee, 2006, s. 120-121). Bu tahminleri yapabilmek için her şeyden önce uzun süreli ve nitelikli bir veri toplama işlemi gerçekleştirilmelidir. İnternet, bu konuda veri toplamayı kolaylaştırmış olsa da her gün milyonlarca ikinci el otomobilin sergilendiği yüzlerce web sitesinin (portal) sayfalarındaki ham verileri elde etmenin ve yapısal hale getirmenin zorluğu, belki de bu konuda yapılması düşünülen çalışmalarda en önemli engeli oluşturmaktadır. Bu çalışma ise, sözü edilen zorlukları aşarak web sitelerinden toplanan verileri işleyip veri madenciliğinde kullanmayı başarmıştır.

Çalışma kapsamında en genel ifadesiyle; veri madenciliği yöntemleri kullanılarak ekonomide çok önemli yeri olan ikinci el otomobillerin fiyatlarının tahmin edilmesi amaçlanmıştır. Bu genel amacın alt amaçlarını şu şekilde ayırmak mümkündür:

1. İkinci el otomobillerin çeşitli temel ve opsiyonel özellikleri ile fiyatları arasındaki örüntüleri saptadıktan sonra, özellikleri verilen herhangi bir aracın güncel pazar değerini tespit etmek.
2. İkinci el otomobillerin, geçmiş yıllardaki zamana bağlı fiyat değişimlerinin incelenmesinden hareketle, gelecek yıllardaki fiyatlarını aylar itibarıyla tahmin etmek. Daha açık ifadeyle; zaman serilerinde tahmin amaçlı kullanılan geleneksel analiz yöntemleri ile yapay sinir ağı modellerinin ikinci el otomobil fiyatlarına ait

tahmin doğruluklarını karşılaştırarak en yüksek doğruluğu sağlayan modeli belirleyip gelecekte oluşacak fiyatları tahmin etmek.

Çalışmanın ikinci bölümünde de belirtildiği gibi, hem pazardaki güncel fiyatları tespit etmek (estimation), hem de gelecekteki fiyatları öngörmek (ya da kestirmek) (forecasting veya prediction) kavramları yerine Türkçe’ mizde bunların tümünü kapsayan “tahmin etmek” terimi kullanılmıştır.

Çalışmanın üçüncü bölümünde, gerek otomotiv endüstrisinin gerekse de ikinci el otomobil pazarlarının günümüzde ulaştığı hacimden ve ekonomideki öneminden bahsedilmişti. Yazın taraması bölümünde de belirtildiği gibi, ikinci el otomobil fiyatlarına yönelik akademik araştırmaya rastlanamamış olması bu çalışmanın özgünlüğünü ortaya koymaktadır. Bu nedenle çalışmanın bu konuda yapılacak ileriki çalışmalara da katkıda bulunması hedeflenmektedir.

Amaçların gerçekleştirilmesi sonucunda elde edilecek modeller, uygun bir web ara yüzü yazılımı geliştirilerek internet ortamında yararlanıcıların hizmetine sunulabilir. Böylece;

- Özellikle ikinci el otomobil satın almayı ya da elindeki otomobili satmayı düşünen tüketicilere ilgili otomobilin pazar fiyatını kolaylıkla ve hızlıca öğrenebilme imkânı sağlanmış olacaktır. Ayrıca, bu tüketicilerin hangi otomobillerin hangi dönemlerde değer kazandığını ya da kaybettiğini görerek mevcut otomobillerini satma ve/veya yeni otomobil alma konusunda isabetli karar vermeleri mümkün olabilecektir.
- Otomobil kredisi veren finans kuruluşlarının, otomobil kiralama firmalarının, otomobil galericilerinin ve sigorta şirketlerinin otomobil fiyatlarında meydana gelen değişimleri görerek risk durumlarını tespit etmeleri ve buna uygun yatırım yapmaları mümkün olabilecektir.
- Özellikle Avrupa’ daki otomobil üreticileri, kendi marka ve modellerinin ikinci el pazarındaki tercih edilebilirliğini yakından takip ederek buradan elde ettikleri çıkarımları yeni otomobil üretimlerinde (örneğin motor hacminin ya da çeşitli opsiyonların belirlenmesinde) kullanmaktadırlar. Çalışmanın sonuçları bu üreticilerin de ikinci el otomobil pazarını etkin olarak izlemelerine imkân verebilecektir.
- Ayrıca, çalışmanın insan hayatında önemli yer teşkil eden diğer ikinci el dayanaklı

ürünlerin (emlak gibi) fiyatlarının tahmin edilmesine yönelik ileride yapılacak çalışmalara da referans oluşturması hedeflenmektedir.

4.2. Araştırmanın Yöntemi

4.2.1. Veri Kaynağı

Yazın taramasından da anlaşıldığı gibi, veri madenciliği konusunda yapılan çalışmaların tamamına yakın bölümünde hazır verilerin kullanılmış olduğu görülmektedir. Örneğin, talep tahmini, döviz kuru, enflasyon oranı, hava durumu gibi bilgilerin tahminine yönelik çalışmalardaki veriler hazır olarak ilgili kaynaklardan tedarik edilmektedir. Özellikle resmi kuruluşlar tarafından uzun yıllar boyunca ölçülüp veritabanlarında saklanan bu tip veriler, herhangi bir değişikliğe tabi tutulmaksızın doğrudan kullanılabilen ve çok hızlı sonuçlar alınabilmektedir, zira veri madenciliği projelerindeki çalışma süresinin yaklaşık % 75'lik bölümünü verinin hazırlanması aşaması oluşturmaktadır. Bu süreç; verilerin temizlenmesi, birleştirilmesi, dönüştürülmesi gibi çok uzun zaman gerektiren aşamaları içermektedir.

Bu çalışmanın uygulamasında ise tamamen hazır verileri kullanmak yerine, hem geçmişte ham olarak toplanmış verilerin veri madenciliğinde kullanılacak aşamaya getirilmesine katkıda bulunulması, hem de yeni verilerin nitel ve nicel olarak daha etkin toplanması hedeflenmiştir. Bu sayede, özellikle otomobillerin güncel pazar fiyatlarını tahmin etmeye yönelik ilk uygulamada, amaca uygun olarak toplanmış kaliteli ve güncel verilerin kullanılması mümkün olmuştur.

Uygulamalara başlamadan önce yapılan araştırmalar neticesinde, Almanya' nın Hamburg şehri yakınlarında faaliyet gösteren bir firma ile görüşmelere başlanmıştır. Yaklaşık 3-4 yıl boyunca Avrupa' daki ikinci el otomobil ilanı yayınlayan önemli web sitelerindeki ilan verilerini toplamakta olan bu firma ile yapılan karşılıklı görüşmelerin olumlu sonuçlanmasının ardından tez yazarının liderliğinde bir veri madenciliği projesi oluşturularak uygulamalara başlanmıştır.

Verilerin toplanması ve hazırlanması sürecinin de uygulamaya katılması, doğal olarak çalışmanın beklenenden çok daha uzun sürmesine neden olmuştur.

4.2.2. Veri Analizi

Çalışmanın ilk amacını gerçekleştirmek için; ikinci el otomobil özellikleri ile fiyatları arasındaki yapısal ilişkinin modellenmesinde “regresyon analizi” ve “yapay sinir ağları” kullanılmıştır. İkinci amaca yönelik olarak, geçmiş dönem ikinci el otomobil fiyatlarından oluşan zaman serilerinin, gelecek dönemlerdeki tahminlemede “Üstel Düzleştirme”, “ARIMA (Box-Jenkins)” ve “yapay sinir ağı modelleri” kullanılmıştır.

Oluşturulan çeşitli modellerin tahmin doğruluklarının karşılaştırılmasında; OMH (ortalama mutlak hata), OMH (ortalama mutlak yüzde hata) ve OHKK (ortalama hata kareler kökü) doğruluk ölçüleri kullanılarak amaçları gerçekleyen en uygun modeller ortaya konmuştur.

4.2.3. Uygulama Araçları

Çalışmada mümkün olduğunca açık kaynak kodlu yazılımların kullanılmasına özen gösterilmiştir. Kullanılan yazılımların tümü aşağıda listelenmiş, bunlar içerisinde açık kaynak kodlu ve ücretsiz olanları yanlarındaki bir (*) simgesi ile belirtilmiştir.

- İşletim Sistemi: İstemci olarak MS-Windows XP, sunucu olarak Linux*
- Kelime İşlemci ve Hesaplama Yazılımı: OpenOffice*
- Veritabanı: mySQL*
- Veritabanı Komutları Çalıştırma Yazılımı: HeidiSQL ve mySQL Browser*
- Java Kodları Geliştirme Yazılımı: Eclipse*
- Veri Madenciliği Analizi Yazılımları: SPSS 16.0, Weka* 3.4, R*2.7.1, Matlab R2007b, Clementine 11.1

Regresyon analizlerinde SPSS ve Weka, zaman serisi yöntemlerinde SPSS ve R, yapay sinir ağları modellemelerinde ise SPSS, Weka, Clementine ve Matlab yazılımlarından yararlanılmıştır. Yazılımların kendine özgü özellikleri ve algoritmalarda kullandıkları parametre çeşitliliğindeki farklılıkları nedeniyle aynı veri kümelerinde küçük de olsa farklı sonuçlar verebilmektedir. Ayrıca işlem hızlarının da önemli farklılıklar gösterdiği gözlenmiştir. Bu nedenle her veri kümesi yapılacak analize uygun olan yazılımların tümünde denenerek en iyi değerleri veren yazılımların sonuçları kullanılmıştır.

Her yazılım genelde kendine özgü bir dosya biçimini kullanmakta olsa da veri transferi için hemen hepsi CSV dosya tipini desteklemektedir. Bu nedenle çalışmalarda verilerin veritabanlarından yazılımlara transferinde CSV dosya tipi tercih edilmiştir.

Kdnuggets' in Mayıs 2008' de çevrimiçi gerçekleştirdiği bir anket sonucuna göre en yaygın kullanılan veri madenciliği yazılımları şunlardır (Web-Kdnuggets2):

- SPSS Clementine % 21,3
- RapidMiner % 20,7
- Salford CART, MARS, TreeNet, RF % 20,7
- SPSS % 19,6
- Excel % 17,6
- SAS % 15,9
- Diğer Yazılımlar % 15,3
- Kendi Yazılımları % 14,4
- R % 11,2
- Weka % 10,4
- KXEN % 9,2
- KNIME % 8,6
- SAS Enterprise Miner % 6,9
- MATLAB % 6,3
- SQL Server % 5,8
- Angoss % 2,3
- C4.5/C5.0 % 2,3
- Oracle DM % 2,0

Çalışmada tercih edilen veri madenciliği yazılımları ile ilgili olarak aşağıda özet bilgi sunulmuştur.

R

Açık kaynak kodlu bir istatistiksel analiz aracı olan R, istatistiksel hesaplama ve grafik işleme için geliştirilmiş bir dil ve programdır. Komut tabanlı çalışan bu yazılım, dünyanın önde gelen üniversiteleri, araştırma enstitüleri ve kurumları tarafından yoğun olarak kullanılmaktadır. İstatistiksel amaçla kullanılan çeşitli pahalı ticari yazılımların içerisinde özgür ve ücretsiz olarak kullanıma açık olan R; özellikle istatistiksel hesaplama amacıyla tasarlanmış olması, güçlü programlama dili ve Unix türevi işletim sistemlerinde meta-programlamaya müsait altyapısı sayesinde akademisyenlerin ve araştırmacıların ilgisini çekmektedir. CRAN adı verilen arşiv ağında (Comprehensive R Archive Network) yüzlerce paket bulunmaktadır. Ülkemizde yeterince tanınmamakla birlikte bazı üniversitelerde ders materyali olarak kullanılmaya başlandığı bilinmektedir (Web-Fazlamesai).

Weka

Yeni Zelanda'daki University of Waikato tarafından Java programlama dili kullanılarak geliştirilmiş olan Weka; Linux, Mac OS ve Windows işletim sistemleri altında çalışabilen, içerisinde pek çok makine öğrenmesi algoritması barındıran bir veri madenciliği yazılımıdır. Weka'nın asıl alanı, makine öğrenmesi algoritmasının uygulandığı sınıflandırma problemleri olmasına rağmen bünyesinde çeşitli kümeleme, birliktelik kuralları, regresyon analizi ve yapay sinir ağları algoritmaları da barındırmaktadır.

SPSS

İlk defa 1968 yılında piyasaya sürülmüş olan SPSS, istatistiksel analize yönelik bir bilgisayar yazılımıdır. Özellikle sosyal bilimlerde, pazar araştırmalarında, sağlık araştırmalarında başta anket şirketleri, hükümetler ve eğitim kurumları olmak üzere pek çok kurum tarafından kullanılmaktadır. Grafikselleştirilmiş ve kullanıcı dostu bir arayüze sahip olan yazılım, makro dilleri yardımıyla kullanıcının kendi amaçları doğrultusunda programı yönlendirebilmesine imkân sağlamaktadır.

Matlab

“Matlab” yüksek seviyeli bir teknik programlama dili olmasının yanında algoritma geliştirme, verilerin görselleştirilmesi, veri analizi ve sayısal hesaplamalar için kullanılan etkileşimli bir yazılım paketidir. Yazılımın birçok alanda uygulamaları vardır. İçerdiği “toolbox” adı verilen paketler aracılığıyla yapay sinir ağları, bulanık mantık, sayısal işaret işleme, finansal modelleme ve analiz gibi pek çok alanda kullanılabilir. Yazılımın büyük boyuttaki verilerle çok hızlı çalışabilme avantajının yanında çok pahalı olması dezavantajı bulunmaktadır.

Clementine

SPSS firması tarafından özellikle veri madenciliği analizlerinde profesyonel olarak kullanılması amacıyla üretilmiş bir yazılımdır. SPSS içerisinde yer alan YSA modülüne göre çok daha gelişmiş veri okuma, modelleme ve görüntüleme seçenekleri mevcuttur. İş akışları CRISP-DM sürecine uygun olarak tasarlanmıştır. Veri madenciliğinde kullanılan tüm nesnelerin (tablolar, ilişkiler, grafikler, hata ölçüleri, vb.) grafikselleştirilmiş ortamda akış

diyagramları şeklinde gösterilebilmesi, kullanım ve anlama kolaylığı sunması açısından Clementine' ı diğer veri madenciliği yazılımlardan ayıran en belirgin özelliğidir.

4.2.4. Örneklem

Araştırmanın örneklemini, uygulamanın yapıldığı firmada elde edilen yaklaşık 40 adet ikinci el otomobil web sitesindeki (portal) veriler oluşturmaktadır. Bu portallerde yer alan otomobil verileri şu 8 ülkeden elde edilmiştir: Almanya, Fransa, İtalya, Portekiz, İspanya, Avusturya, Hollanda, İsviçre.

Örneklemdaki markalar, verilerin elde edildiği tarihteki ilan sayısına göre sıralanarak ilk 10 sırada yer alanları Tablo 4.1' de listelenmiştir.

Tablo 4.1. İlan sayısının çokluğuna göre ilk 10 otomobil markası

| No | Marka | İlan sayısı |
|----|----------|-------------|
| 1 | VW | 182689 |
| 2 | BMW | 126810 |
| 3 | MERCEDES | 126371 |
| 4 | AUDI | 102758 |
| 5 | OPEL | 94277 |
| 6 | RENAULT | 81744 |
| 7 | PEUGEOT | 67304 |
| 8 | FORD | 64474 |
| 9 | CITROEN | 34082 |
| 10 | TOYOTA | 28973 |

İlk amacı sağlamaya yönelik uygulamada, en güncel ikinci el otomobil bilgilerinin yer aldığı, Aralık 2007 tarihinde elde edilen otomobil verileri kullanılmıştır. Bu veriler, otomobillerin fiyatını ve ileride ayrıntılı olarak anlatılan özelliklerini kapsamaktadır.

İkinci amacı gerçeklemeye yönelik uygulamada ise piyasada en çok işlem gören, diğer bir ifadeyle en popüler konumda olan 3 otomobil modelinin Ocak 2005 – Aralık 2007 tarihleri arasındaki 3 yıllık günlük fiyat verilerinin aylık ortalamalarından oluşan zaman serileri kullanılmıştır. Günlük veriler yerine aylık ortalamalarının kullanılmasının iki temel sebebi vardır:

1. Otomobil fiyatlarının dikkate değer ölçüde günlük değişimler göstermemesi,

2. Zaman zaman yaşanan teknik sorunlar veya bazı büyük hacimli web sitelerinin tamamının taranmasının birkaç gün sürebilmesi nedenleriyle sürekli olarak günlük veri elde edilememesi.

4.3. Uygulama Süreci

Çalışmanın uygulama sürecinde, ikinci bölümde açıklanan CRISP-DM süreci referans alınmıştır. Bu sürecin aşamalarında gerçekleştirilen işlemler aşağıda anlatılmaktadır.

4.3.1. İşin Kavranması

Bu aşamada öncelikle çalışmanın yapıldığı işletmedeki iş süreçlerinin analizi yapılmıştır. Bu analizler, çalışmanın ileriki aşamalarında ihtiyaç duyulabilecek bilgilerin önceden kavranmasını sağlamıştır. Bu bilgiler arasında işletmenin kimliği, çalışma prensipleri, iş akışları, müşteri ilişkileri, bilişim güvenliği, verilerin elde edilme şekli ve veri tabanı yapıları yer almaktadır.

Ardından proje hedefleri ve gereksinimleri belirlenmeye çalışılmıştır. Projenin hedefleri ortaya konarak bu hedeflere ulaşabilmek için gereken insan kaynağı ve yazılım araçları tespit edilmiştir. Ayrıca ön analizler sonucunda, işletmede üretilen sayısal verilerin kalitesinin ve boyutlarının veri madenciliği için yeterli düzeyde olmadığı anlaşıldığı için bu konuda da gerekli ön çalışmalar plânlanmıştır.

4.3.2. Verilerin Toplanması ve Kavranması

Veriler, ikinci el otomobil ilanlarının yayınlandığı portallerin “web-crawling” yöntemi ile elektronik ortamda sorgulanmaları sonucu elde edilmiştir. Web-crawling, bugün itibariyle Türkçe’ de yer almayan bir kavramdır. Wikipedia’ da bu konuda şu bilgiye rastlanmıştır: “Web-crawling (aynı zamanda spidering olarak da bilinir) world wide web’ in, metodolojik olarak ve otomatik tarzda, bir program ya da otomatikleştirilmiş kod aracılığıyla taranması işlemidir” (Web-Wikipedia1). Bu işlem özellikle Google, Yahoo gibi çeşitli arama motorları tarafından sıklıkla kullanılmaktadır.

Verilerin “web-crawling” yöntemi ile toplanması, işletme içerisinde önceden geliştirilmiş olan Java kodu aracılığıyla otomatik şekilde gerçekleştirilmiştir. Toplanan

verilerde zaman zaman hatalar ve/veya eksikler görüldüğünden dolayı, bu hataları gidermeye yönelik analizlerden elde edilen bilgilerle Java kodunda güncellemeler yapılmış, böylece verilerin niteliğinde ve niceliğinde sürekli olarak artış sağlanmıştır.

Veri kümelerinin yapıları, ilerleyen kısımlarda ayrıntılı olarak açıklanacaktır.

4.3.3. Verinin hazırlanması

Süreçteki toplam zaman ve enerjinin yaklaşık % 75' inin harcandığı bu aşamada; veri toplama, birleştirme, temizleme ve dönüştürme işlemleri neticesinde yeni toplanan ve geçmişte toplanmış olan verilerin analiz ve modelleme çalışmasına uygun hale getirilerek bir veri ambarında toplanması sağlanmıştır.

Web sayfalarından toplanan veriler düz metin halde bulunduğu için, bunları öncelikle düzgün yapıda veritabanlarına aktarabilmek önemli boyutta zaman ve emek gerektirmiştir. Bu amaçla geliştirilen SQL kodları ile ilanlarda yer alan veriler analiz edilerek tablolardaki ilgili alanlara aktarılmıştır. Örneğin, rastlanan bir sözcük herhangi bir araç markası ile eşleşiyor ise bu sözcük ilgili tablonun BRAND (marka) alanına aktarılmış, şayet sayısal bir değer olup yanında Euro sözcüğü ya da € simgesi yer alıyorsa, PRICE (fiyat) alanına aktarılmıştır.

Verilerin tablolara aktarılması sonrasında elde edilen kayıtlarda çok sayıda hata ve/veya eksik olduğu gözlenmiş ve bu sorunlar yine SQL ile geliştirilen kodlarla otomatik olarak giderilmeye çalışılmıştır. Aşağıda bu sorunlardan bazıları ve çözüm için gerçekleştirilmiş olan işlemler kısaca özetlenmiştir:

- **Tekrar eden kayıt:** Bu konuda temelde iki tip sorun görülmüştür:
 - Bazen aynı sitede aynı ilanın birden fazla kopyası bulunduğu gibi, bazı ilanlar pek çok sitede eş zamanlı yayınlanabilmektedir. Sitelerden toplanan verilerin birleştirilmesi esnasında aynı ilandan birden fazla sayıda tespit edildiğinde ilanlardan bir tanesi veritabanında tutulmuş, diğerleri silinmiştir.
 - Herhangi bir sitedeki bir ilan aylarca yayınlanmaya devam etmektedir. Her sorgulamada tekrar toplanan bu ilanlar, tekrar eden kayıt oluşturduğu için sadece bir tanesi tutulmuş, diğerleri silinmiştir.

- **Yazım hatalı veri:** Özellikle marka ve modellerde bu tip hatalara çok sık rastlanmıştır. Örneğin, “Mercedes” markası yerine “Mersedec” yazılması gibi. Verileri giren kullanıcılardan kaynaklanan bu tip hatalar için “sık rastlanan hatalı kelimeler” belirlenmiş ve geliştirilen kod yardımıyla otomatik olarak düzeltilmesi sağlanmıştır. Metodolojik olarak düzeltilemeyecek durumdaki yazım hatalarına rastlandığında ise, ilanın hatalı olduğu kanaatine varılarak veritabanından silinmiştir.
- **Mantıksal hatalı veri:** Bazı otomobil ilanlarında mantıksal olarak hatalı bilgiler yer almaktadır (örneğin fiyatının 10 Avro, üretim tarihinin 1300 girilmesi, vb.) Bu tip ilanların veritabanında yer almasını engellemek için, araçların temel özellikleri (fiyat, üretim tarihi, motor hacmi, vb.) için uygun veri aralıkları ikinci el otomobillerin ayrıntılı özelliklerini barındıran bir “model bankası” ile karşılaştırılmış, bu aralıkların dışında kalan ilanlara ait kayıtlar silinerek veritabanından çıkarılmıştır.
- **Eksik veri:** İlanların bir kısmında, veri girenlerin ihmalkârlığından kaynaklanan eksik verilere rastlanmıştır. Örneğin araç modelinin ya da yaşının belirtilmemesi gibi. Bu tip veriler de analizleri olumsuz etkileyeceği için veritabanından çıkarılmıştır.
- **Birim farkı:** Özellikle motor güçlerinde bu tip hatalara sık rastlanmıştır. Günümüzde motorlu araçların güçlerini belirtmek için halen ülkeye göre değişen farklı güç birimleri kullanılabilir: “hp”, “PS”, “CV”, “pk”, “ch” gibi. Standardizasyonu sağlamak için güç birimi olarak “hp” temel alınmış, diğer birimler ise uygun dönüşüm formülleriyle “hp” birimine dönüştürülerek veritabanlarına kaydedilmiştir.

Verilerin kalitesini ve miktarını artırmak amacıyla sürekli olarak “Kayıp Analizleri (Loss Analysis)” yapılmıştır. Bu amaçla, daha önce anlatılan hatalı veya eksik veri içermesi nedeni ile veritabanlarından silinen her kayıt, silinme sebebini gösteren bir kod (Delete Indicator) ile ayrı bir tabloya aktarılmıştır. Belirli periyotlarla silinen kayıtların toplandığı tablolar analize tabi tutulmuştur. Böylece, silinmelerin en çok hangi sebeplerden kaynaklandığı istatistiksel olarak gözlemlenmiş ve en sık rastlanan silinme kodlu kayıtlar ayrıntılı olarak incelemeye tabi tutularak bu hataların önlenmesine yönelik SQL kodları geliştirilmiştir. Örneğin, “hatalı marka ismi” nedeniyle silinen kayıtlar incelendiğinde bu hatalı marka isminin çoğunlukla “Peugout” olduğu görüldüyse, ilanı giren kullanıcıların bu marka isminde sıkça yazım hatası yaptığı varsayılarak, geliştirilen bir SQL kodu ile

bundan sonra “Peugout” olarak alınan tüm markalar otomatik olarak “Peugeout” ya dönüştürülmüştür. Böylece bu hata nedeniyle oluşan veri kayıpları engellenmiştir.

Şekil 4.1’ de, örnek bir hata analizi özet sayfası görülmektedir (Firma bilgilerindeki gizliliği korumak amacıyla bazı alanlar ve içerikleri “XXX” ile gösterilmiştir). Bu sayfanın en solundaki tablo, belirli bir günde en sık rastlanan silinme kodlarını ve açıklamalarını göstermektedir. Ortadaki tabloda toplam ilan sayısı, silinen ilan sayısı, veritabanlarına aktarılan ilan sayısı ve silinme sayısına oranı gibi bilgiler yer almaktadır. Şekildeki örnekte görüldüğü gibi, 15.09.2007’ de alınan 895 bin araçtan 779 bini, yani yaklaşık % 87’ si silinmekte, sadece % 13’ e yakın bir bölümü veritabanlarında tutulabilmektedir. Bu da kayıp analizlerinin önemini göstermektedir. Şeklin sağında yer alan tabloda ise silinme kodları, silinme sayısına göre çoktan aza doğru sıralı olarak görülmektedir. Bu tablo aracılığıyla da en çok hangi nedenlerden dolayı silinmelerin gerçekleştiği tespit edilmekte, bu nedenle silinen kayıtlar ayrıntılı analize tabi tutulmakta ve tekrar veri toplanması esnasında silinmelerini önleyecek şekilde SQL kodları geliştirilmeye ya da güncellenmeye çalışılmaktadır.

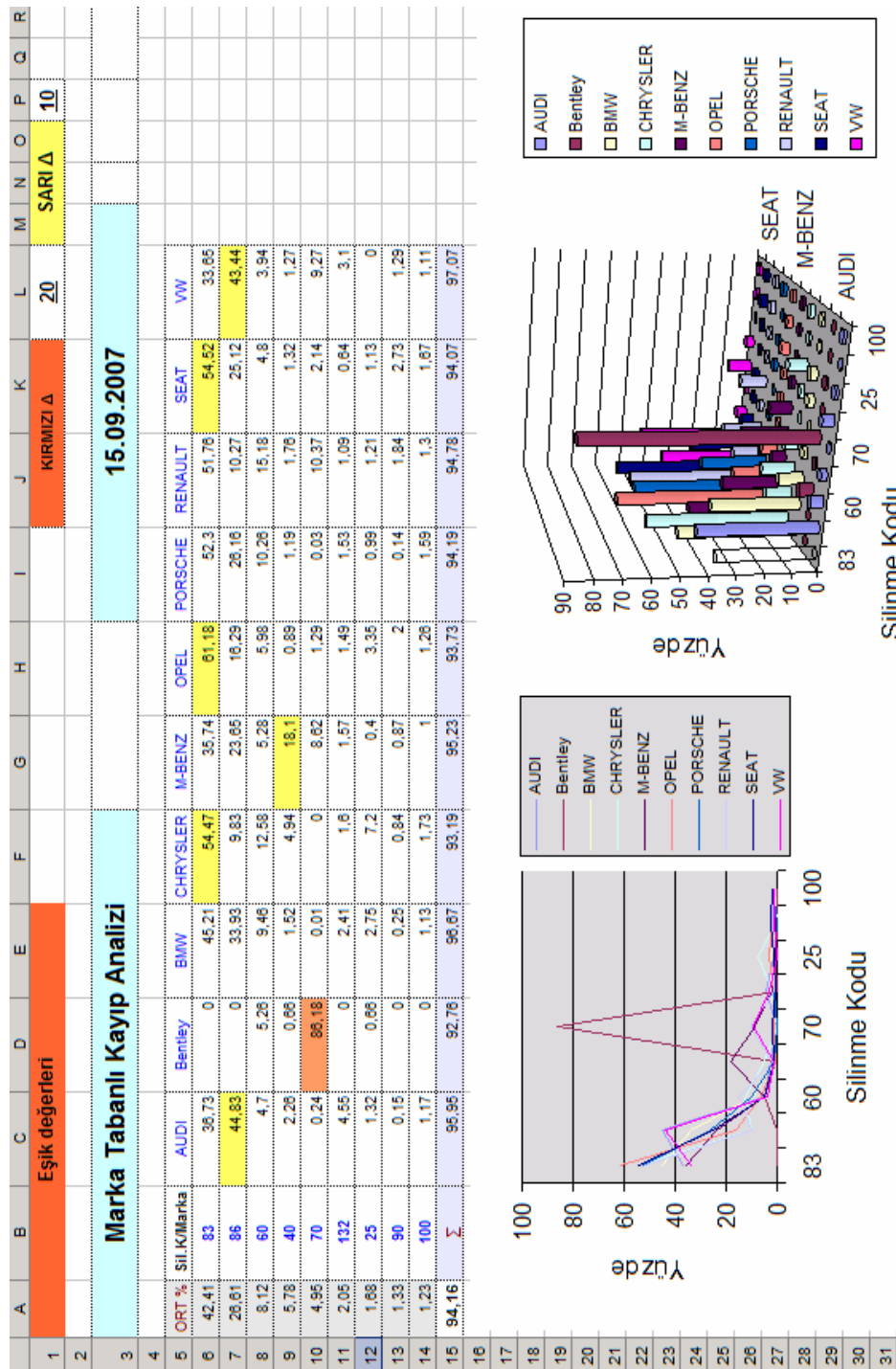
| Kayıp Analizi | | | Özet Sayfası | | | | 15.09.2007 | | |
|---------------|---------------------------|-------------|--------------|---------|-------|--------------|------------|-------|--|
| Silinme Kodu | Açıklama | Toplam İlan | Silinen | Tutulan | % | Silinme Kodu | S. Sayısı | % | |
| 10 | KM çok küçük | 895.668 | 779.664 | 116.004 | 12,95 | 83 | 330.650 | 42,41 | |
| 25 | Hatalı ZIP kodu | XXX | XXX | | | 86 | 207.483 | 26,61 | |
| 40 | Hatalı Model Adı | 1.209.775 | 277.533 | | | 60 | 63.308 | 8,12 | |
| 60 | Teknik özellikler uyumsuz | XXX | | | | 40 | 45.072 | 5,78 | |
| 70 | Tekrar eden kayıt | 13.828 | | | | 70 | 38.559 | 4,95 | |
| 83 | XXX | XXX | XXX | | | 132 | 15.989 | 2,05 | |
| 85 | XXX | 59.531 | 91.758 | | | 25 | 13.122 | 1,68 | |
| 88 | XXX | | | | | 90 | 10.393 | 1,33 | |
| 90 | XXX | | | | | 100 | 9.621 | 1,23 | |
| 100 | XXX | | | | | 6 | 7.645 | 0,98 | |
| 110 | XXX | | | | | 130 | 7.432 | 0,95 | |
| 120 | XXX | | | | | 10 | 6.158 | 0,79 | |
| 120 | XXX | | | | | 89 | 5.260 | 0,67 | |
| 130 | XXX | | | | | 120 | 4.954 | 0,64 | |
| 132 | XXX | | | | | | | | |

Şekil 4.1. Kayıp analizlerinin özet sayfası

Bu kayıp analizleri, portal ve marka bazında da ayrıntılı olarak gerçekleştirilmektedir. Şekil 4.2’ de görülen portal tabanlı kayıp analizi, özellikle hangi portallerde ne tip kayıpların olduğunu göstermektedir. Şeklin üstünde yer alan tablonun üstündeki iki karakterli kodlar (2C, 3A, vb.) verilerin çekildiği portalleri ifade etmektedir. Sol tarafta ise silinme kodları ile bu silinmelerin yüzdesel ortalamaları yer almaktadır. Tablo içinde yer alan hücre değerleri, hangi portalden hangi nedenlerle yüzde kaçlık bir silinme olduğunu

belirtmektedir. Alt toplamların % 100 görünmemesi, tabloda sadece önemli silinme kodlarına yer verilmiş olmasından kaynaklanmaktadır. Bu işleme özgü geliştirilen kod yardımıyla, oranların büyüklüğü ortalama değerler ile karşılaştırılmakta, aradaki farkın tablo üstünde görülen ilgili eşik değeri sınırları içerisinde olması durumunda hücreler otomatik olarak kırmızı ya da sarıya dönüştürülmektedir (Örnek olarak C6 hücresi için kullanılan kod, şeklin alt kısmında gösterilmiştir). Kırmızı renkteki değerler, en önemli olan ve acil olarak detaylı analize tabi tutulması gereken portalleri göstermektedir. Sarı renkteki oranlar ise kayıp (silinme) açısından ikincil derece öneme sahip portallere işaret etmektedir. Benzer şekilde, sayıları görselleştirmek amacıyla hazırlanmış olan 3 boyutlu grafik yardımıyla da, hızlı olarak hangi portallerde çok fazla silinme olduğunu görmek mümkün olmaktadır. Kodlarla otomatik olarak renklendirilmiş tablo ya da görselleştirilmiş grafik yardımıyla önemli görülen sorunların belirlenmesinin ardından, veritabanlarına dönülerek ilgili portaldeki önemli silinmelere ait kayıtlar ayrıntılı olarak analiz edilmekte, gerekirse ilgili web sitesindeki orijinal ilan ile karşılaştırılmakta ve bu silinmeleri azaltabilecek tedbirlerin alınmasına çalışılmaktadır.

Şekil 4.3' de görülen marka tabanlı kayıp analizi de Şekil 4.2' deki portal tabanlı kayıp analizine benzer işlev görmektedir. Farklı olarak, buradaki tablo ve çeşitlendirilmiş grafikler aracılığıyla, portaller yerine bu defa hangi otomobil markalarında ne tip kayıpların olduğu görülebilmektedir. Tablodan anlaşılacağı gibi, üstte yer alan isimler otomobil markalarını göstermektedir.



Şekil 4.3. Marka tabanlı kayıp analizi

Verilerin kalitesinin ve miktarının artırılmasında çok büyük katkısı olan bu kayıp analizleri, periyodik olarak gerçekleştirilmekte ve elde edilen sonuçlara göre sürekli geliştirilmektedir.

Veri toplanması ve hazırlanması aşamasının sonunda, Aralık 2007 tarihinde yaklaşık 1.253.000 adet otomobil ilan sayısı ile her farklı araç modeli için Ocak 2005-Aralık 2007 ayları arasındaki 36'şar aylık veriye ulaşılarak analizlere başlanmıştır. Özellikle kayıp analizleri neticesinde son 6 ayda toplanan veri miktarında ve kalitesinde önemli bir artış olmuştur. Bu sayede, özellikle ilk uygulama bölümündeki güncel fiyatların tahmin edilebilmesine yönelik olarak nitel ve nicel açıdan tatmin edici güncel verilerin kullanımı mümkün olmuştur.

4.3.4. Modellerin Kurulması ve Değerlendirme

Çalışmanın ilk uygulama bölümünde ikinci el otomobillerin güncel fiyatlarının tahminine yönelik modeller, ikinci uygulama bölümünde ise gelecekteki fiyatlarının tahminine yönelik modeller kurulmuştur. Aşağıda bu modellemeler sırasıyla açıklanmıştır.

4.3.4.1. Güncel Fiyatları Tahmin Etmeye Yönelik Modelleme

Bu aşamada; regresyon analizi ve yapay sinir ağları kullanılarak, araçların güncel fiyatlarını tahmin etmeye yönelik tahmin modelleri kurularak doğrulukları karşılaştırılmıştır.

Veri Kümesi

Bu modellemenin veri kümesini Aralık 2007' de elde edilmiş olan ikinci el otomobillerin fiyatları ve özellikleri oluşturmaktadır. Veriler SQL komutlarıyla veri madenciliğine uygun hale getirildikten sonra 1.253.000 kayıtlık veri kümesi elde edilmiştir. Bu büyüklükte veri kümesini işlemede, eldeki donanım ve/veya yazılımların yetersiz kalması üzerine, her ülkeden 1000 adet kayıt alınarak toplamda 8000 kayıttan oluşan bir veri kümesi kullanılmıştır. Tablo 4.2' de bu veri kümesinde yer alan ilan sayısının ülkelere göre dağılımı görülmektedir.

Tablo 4.2. Veri kümesinin ülkelere göre dağılımı

| Ülke | İlan Sayısı |
|-----------|-------------|
| Almanya | 822446 |
| Fransa | 221208 |
| İspanya | 47228 |
| İtalya | 26886 |
| Avusturya | 20647 |
| Portekiz | 13194 |
| İsviçre | 7989 |
| Hollanda | 5917 |

Çalışmada 48 adet otomobil markasına ait toplam 20.000 otomobil modelinden oluşan bir “model bankası” referans amacıyla kullanılmıştır. Bu referans modeli, otomobil üreticisi firmalar tarafından belirli periyotlarla güncellenerek yayınlanan bilgilerin toplamından elde edilmiş olup bu üreticilere ait tüm otomobil modellerini ve ayrıntılı özelliklerini içermektedir. İlanlardan elde edilen özellikler arasında bir modeli benzersiz olarak tanımlayan 6 adet temel özellik (Marka, Temel Model, Detay Model, Kasa Tipi, Kapı Sayısı, Motor Gücü) kullanılarak ilanın model bankasındaki karşılığı tespit edilmekte ve ilanlarda bildirilen özellikler ile birleştirilerek veri madenciliği analizleri için oluşturulan veri ambarına aktarılmaktadır (Veri ambarının oluşturulması uygulamanın temel hedefleri dışında kaldığı için buna yönelik uygulama ayrıntılarının anlatılmasına gerek duyulmamıştır). Bu veri ambarında, her otomobile ait 100’ ün üzerinde özellik tutulmaktadır. Ancak bu özelliklerin büyük çoğunluğu otomobil fiyatı ile doğrudan ilişkili olmayıp bilgi amaçlı saklanmaktadır (Örneğin araç modellerinin hangi tarihte üretime başladığı bilgisi gibi). Bu nedenle, otomobil fiyatları ile doğrudan ilgisi olacağı düşünülen özellikler sezgisel olarak belirlenip analizlerin yapıldığı tablolara aktarılmıştır. 17 adet temel, 32 adet opsiyonel özellikten oluşan bu yapı Tablo 4.3’ de gösterilmiştir.

Tablo 4.3. Güncel fiyatları tahmin etmede kullanılan örneklemin veri yapısı

| No | Orijinal Alan Adı | Açıklama | Veri Türü | Min.-Maks. |
|----|----------------------|----------------------|-----------|------------|
| 1 | Brand_No | Marka Numarası | Sayısal | 1-69 |
| 2 | Brand | Marka | Metin | |
| 3 | Model_Name_Rough_No | Temel Model Numarası | Sayısal | 1-772 |
| 4 | Model_Name_Rough | Temel Model | Metin | |
| 5 | Model_Name_Detail_No | Detay Model Numarası | Sayısal | 1-15685 |
| 6 | Model_Name_Detail | Detay Model | Metin | |
| 7 | Body_Type_No | Kasa Tipi | Kategorik | 0-14 |
| 8 | Doors_No | Kapı Sayısı | Kategorik | 2-5 |
| 9 | Engine_Power | Motor Gücü | Sayısal | 29-450 |
| 10 | Cylinder_Capacity | Silindir Kapasitesi | Sayısal | 599-7011 |

| No | Orijinal Alan Adı | Açıklama | Veri Türü | Min.-Maks. |
|----|--------------------------------|----------------------------|-----------|-------------|
| 11 | Cylinders_No | Silindir Sayısı | Sayısal | 2-12 |
| 12 | Fuel_Type_No | Yakıt Tipi | Sayısal | 1-9 |
| 13 | Mileage | Katedilen Mesafe | Sayısal | 1-350.000 |
| 14 | Age | Yaş (Ay Olarak) | Sayısal | 1-176 |
| 15 | Color | Renk | Kategorik | 0-10 |
| 16 | Country | Ülke | Kategorik | 1-8 |
| 17 | Zip_Code | Posta Kodu | Kategorik | 0-99990 |
| 18 | Automatic_Gears | Otomatik Vites | Kategorik | 0-1 |
| 19 | Leather_Seats | Deri Koltuk | Kategorik | 0-1 |
| 20 | Navigation | Navigasyon Sistemi | Kategorik | 0-1 |
| 21 | Air_Conditioner | Klima | Kategorik | 0-1 |
| 22 | Xenon_Lights | Xenon Önfarlar | Kategorik | 0-1 |
| 23 | Sunroof | Sunroof (Üst Pencere) | Kategorik | 0-1 |
| 24 | Heated_Seats | Isıtılmalı Koltuklar | Kategorik | 0-1 |
| 25 | Board_Computer | Yol Bilgisayarı | Kategorik | 0-1 |
| 26 | Parking_System | Park Sensörü | Kategorik | 0-1 |
| 27 | Cruise_Control | Hız Kontrol Düzeni | Kategorik | 0-1 |
| 28 | Elbow_Rest | Kolçak | Kategorik | 0-1 |
| 29 | Alloy_Rims | Alüminyum Jantlar | Kategorik | 0-1 |
| 30 | Electric_Seats | Elektrikli Koltuk | Kategorik | 0-1 |
| 31 | Electric_Windows | Elektrikli Pencere | Kategorik | 0-1 |
| 32 | Metallic_Lacquering | Metalik Boya | Kategorik | 0-1 |
| 33 | ESP | Esp | Kategorik | 0-1 |
| 34 | Power_Steering | Hidrolik Direksiyon | Kategorik | 0-1 |
| 35 | Central_Door_Locking | Merkezi Kilit Sistemi | Kategorik | 0-1 |
| 36 | ABS | Abs | Kategorik | 0-1 |
| 37 | Airbags | Hava Yastıkları | Kategorik | 0-1 |
| 38 | Alarm_Device | Alarm Cihazı | Kategorik | 0-1 |
| 39 | Cd_Changer | CD Değiştirici | Kategorik | 0-1 |
| 40 | Immobiliser | Immobiliser | Kategorik | 0-1 |
| 41 | Electric_Outside_Mirrors | Elektrikli Yan Aynalar | Kategorik | 0-1 |
| 42 | Cooling_Box | Buzdolabı | Kategorik | 0-1 |
| 43 | Multifunctional_Steering_Wheel | Çok Fonksiyonlu Direksiyon | Kategorik | 0-1 |
| 44 | Fog_Lamp | Sis Lambası | Kategorik | 0-1 |
| 45 | Night_Vision | Gece Görüşü | Kategorik | 0-1 |
| 46 | Cd_Player | CD Çalar | Kategorik | 0-1 |
| 47 | Rain_Sensor | Yağmur Algılayıcısı | Kategorik | 0-1 |
| 48 | TV | Televizyon | Kategorik | 0-1 |
| 49 | Price | Fiyat | Sayısal | 190-266.350 |

Regresyon Modellemesi

Optimum regresyon modelinin oluşturulması için öncelikle modele anlamlı katkısı olan değişkenlerin belirlenmesine çalışılmıştır. Bunun için SPSS ile gerçekleştirilen ilk regresyon analizi sonucu oluşan modelin özeti Tablo 4.4' de görülmektedir. Bu tablo, otomobil fiyatlarının % 80' e yakın bölümünün bağımsız değişkenler tarafından açıklanabildiğini göstermektedir. Bu yüksek oran, ikinci el otomobil fiyatlarının tahminine yönelik olarak gerçekleştirilen veri toplama ve hazırlama işlemlerinin başarısını

göstermektedir. Aynı zamanda, tüm veriler içerisinde fiyatlarla doğrudan ilişkisi olduğu düşünülen değişkenlerin sezgisel olarak belirlenmesinin de oldukça isabetli yapılmış olduğu söylenebilir.

Tablo 4.4. Regresyon model özeti

| Model | R | R ² | Düzeltilmiş R ² | Standart Hata |
|-------|-------|----------------|----------------------------|---------------|
| 1 | 0,894 | 0,799 | 0,798 | 5760,504 |

Bu regresyon modelinin anlamlı olup olmadığını anlamak için incelenen ANOVA testi sonuçları Tablo 4.5’ de yer almaktadır. Bu testin sonucunda ortaya çıkan F değerine karşılık gelen p anlamlılık seviyesinin 0.05’ten küçük olması nedeniyle, söz konusu regresyon modelinin açıklayıcılığının istatistiksel açıdan önemli olduğunu ifade edebiliriz.

Tablo 4.5: Regresyon modeli ANOVA sonuçları

| Model | Kareler Toplamı | df | Ortalama Kare | F | p | |
|-------|-----------------|----------|---------------|----------|---------|-------|
| 1 | Regresyon | 1,048E12 | 43 | 2,437E10 | 734,283 | 0,000 |
| | Kalıntı | 2,640E11 | 7956 | 3,318E7 | | |
| | Toplam | 1,312E12 | 7999 | | | |

Ardından, mevcut modelin değişkenlerinin modele olan katkılarını belirlemek ve anlamlı katkısı olmayan değişkenleri modelden çıkarmak amacıyla regresyon analizinden elde edilen ve Tablo 4.6’ da görülmekte olan katsayılar tablosu incelenmiştir. Tabloda bazı değişkenlerin β katsayılarının negatif işaretli olduğu görülmektedir. Fiyatı ters yönde etkilemesi beklenen yaş (age), kat edilen mesafe (mileage) gibi önemli değişkenlerin negatif işaretli olması Regresyon modelinin anlamlılığının diğer bir göstergesidir (Hair vd., 1998).

Tablo 4.6. Regresyon analizi katsayılar tablosu

| Değişkenler | Katsayılar | | Standart Katsayılar | t | p |
|------------------------------|------------|-----------|---------------------|---------|------|
| | B | Std. Hata | β | | |
| (Sabit) | 14645,299 | 1112,507 | | 13,164 | ,000 |
| Model Name Detail No | -,038 | ,018 | -,014 | -2,171 | ,030 |
| Body Type No | 149,355 | 25,995 | ,033 | 5,746 | ,000 |
| Doors No | -515,158 | 68,687 | -,044 | -7,500 | ,000 |
| Engine Power | 205,650 | 3,844 | ,761 | 53,499 | ,000 |
| Cylinder Capacity | -1,593 | ,288 | -,090 | -5,528 | ,000 |
| Cylinders No | 234,547 | 144,907 | ,019 | 1,619 | ,106 |
| Fuel Type No | 2868,032 | 139,998 | ,127 | 20,486 | ,000 |
| Mileage | -,034 | ,002 | -,149 | -17,433 | ,000 |
| Age | -93,070 | 3,409 | -,258 | -27,300 | ,000 |
| Color | -3,349 | 21,490 | ,000 | -,156 | ,876 |
| Country | 111,529 | 30,002 | ,020 | 3,717 | ,000 |
| Zip Code | -,002 | ,002 | -,005 | -,938 | ,348 |
| Automatic Gears | 809,096 | 211,783 | ,025 | 3,820 | ,000 |
| Leather Seats | -643,941 | 159,921 | -,024 | -4,027 | ,000 |
| Navigation | 1203,424 | 215,241 | ,032 | 5,591 | ,000 |
| Air Conditioner | -1678,572 | 238,852 | -,042 | -7,028 | ,000 |
| Xenon Lights | 582,914 | 181,421 | ,020 | 3,213 | ,001 |
| Sunroof | 651,912 | 221,524 | ,015 | 2,943 | ,003 |
| Heated Seats | 176,831 | 193,319 | ,005 | ,915 | ,360 |
| Board Computer | -530,207 | 158,482 | -,021 | -3,346 | ,001 |
| Parking System | 2361,238 | 229,040 | ,060 | 10,309 | ,000 |
| Cruise Control | -612,561 | 183,776 | -,021 | -3,333 | ,001 |
| Elbow Rest | -819,543 | 162,398 | -,029 | -5,047 | ,000 |
| Alloy Rims | -26,496 | 172,225 | ,000 | -,154 | ,878 |
| Electric Seats | 1492,592 | 209,339 | ,048 | 7,130 | ,000 |
| Electric Windows | -400,442 | 365,837 | -,007 | -1,095 | ,274 |
| Metallic Lacquering | -463,308 | 138,615 | -,018 | -3,342 | ,001 |
| ESP | -691,948 | 162,678 | -,027 | -4,253 | ,000 |
| Power Steering | -466,902 | 445,958 | -,006 | -1,047 | ,295 |
| Central Door Locking | 1284,772 | 363,269 | ,021 | 3,537 | ,000 |
| ABS | -3926,817 | 421,354 | -,055 | -9,320 | ,000 |
| Airbags | -11158,111 | 896,186 | -,071 | -12,451 | ,000 |
| Alarm Device | -129,281 | 178,208 | -,004 | -,725 | ,468 |
| Cd Changer | -156,584 | 287,357 | -,005 | -,545 | ,586 |
| Immobiliser | -649,183 | 220,971 | -,017 | -2,938 | ,003 |
| Electric Outside Mirrors | 662,420 | 145,705 | ,026 | 4,546 | ,000 |
| Cooling Box | 1414,574 | 3339,703 | ,002 | ,424 | ,672 |
| Multifunctional Steering Whe | 1753,681 | 218,651 | ,051 | 8,020 | ,000 |
| Fog Lamp | -864,690 | 164,788 | -,032 | -5,247 | ,000 |
| Night Vision | 1130,495 | 2599,807 | ,002 | ,435 | ,664 |
| Cd Player | 1421,471 | 288,742 | ,042 | 4,923 | ,000 |
| Rain Sensor | 697,914 | 207,374 | ,020 | 3,365 | ,001 |
| TV | 2998,244 | 705,214 | ,022 | 4,252 | ,000 |

Katsayılar tablosunun incelenmesinden sonra, şu değişkenlerin modele anlamlı katkısının olmadığı anlaşılmıştır: Cylinders_No (Silindir Sayısı), Color (Renk), Zip_Code (Posta Kodu), Heated_Seats (Isıtmalı Koltuklar), Alloy_Rims (Alüminyum Jantlar), Electric_Windows (Elektrikli Pencereleler), Power_Steering (Hidrolik Direksiyon), Alarm_Device (Alarm Cihazı), Cd_Changer (CD Değiştirici), Cooling_Box (Buzdolabı),

Night_Vision (Gece Görüşü). Daha duyarlı bir regresyon denklemi elde edebilmek amacıyla, bu değişkenler elimine edilerek yeniden regresyon analizi gerçekleştirilmiştir. Bu analiz sonucu elde edilen model özeti; otomobil fiyatlarının % 80' e yakınının bağımsız değişkenler tarafından açıklanabildiğini gösterirken, ANOVA testi sonuçları da doğal olarak yine modelin istatistiksel geçerliliğini ortaya koyuştur. Yeni modeldeki değişkenlere ait katsayılar tablosu kullanılarak oluşturulan regresyon denklemi aşağıda görülmektedir:

$$\begin{aligned}
 \text{PRICE (FİYAT)} &= \\
 14987.813 * 1 \text{ (sabit)} &+ \\
 -0.041 * \text{Model_Name_Detail_No} &+ \\
 143.990 * \text{Body_Type_No} &+ \\
 -519.552 * \text{Doors_No} &+ \\
 205.472 * \text{Engine_Power} &+ \\
 -1.315 * \text{Cylinder_Capacity} &+ \\
 2807.328 * \text{Fuel_Type_No} &+ \\
 -0.034 * \text{Mileage} &+ \\
 -93.128 * \text{Age} &+ \\
 108.567 * \text{Country} &+ \\
 882.573 * \text{Automatic_Gears} &+ \\
 -627.482 * \text{Leather_Seats} &+ \\
 1209.559 * \text{Navigation} &+ \\
 -1741.538 * \text{Air_Conditioner} &+ \\
 575.335 * \text{Xenon_Lights} &+ \\
 668.277 * \text{Sunroof} &+ \\
 -542.407 * \text{Board_Computer} &+ \\
 2399.647 * \text{Parking_System} &+ \\
 -633.713 * \text{Cruise_Control} &+ \\
 -808.711 * \text{Elbow_Rest} &+ \\
 1513.273 * \text{Electric_Seats} &+ \\
 -476.520 * \text{Metallic_Lacquering} &+ \\
 -697.958 * \text{ESP} &+ \\
 1122.684 * \text{Central_Door_Locking} &+ \\
 -4059.457 * \text{ABS} &+ \\
 -11362.311 * \text{Airbags} &+ \\
 -713.992 * \text{Immoboliser} &+ \\
 627.288 * \text{Electric_Outside_Mirrors} &+ \\
 1788.594 * \text{Multifunctional_Steering_Wheel} &+ \\
 -886.279 * \text{Fog_Lamp} &+ \\
 1307.844 * \text{Cd_Player} &+ \\
 703.422 * \text{Rain_Sensor} &+ \\
 3058.928 * \text{TV} &+
 \end{aligned}$$

Yapay Sinir Ağları Modellemesi

Yapay sinir ağı modelleri, üçüncü bölümde detaylı olarak açıklanan Çok Katmanlı Geri

Yayımlı yapay sinir ağı yapısı kullanılarak oluşturulmuştur. Bu yöntem, nicel verilerle tahmin amaçlı çalışmalarda en çok kullanılan ağ mimarisi olması nedeniyle bu çalışmada da tercih edilen metot olmuştur. Hem doğrusal hem de doğrusal olmayan yapılarda oldukça başarılı sonuçlar vermesi, diğer mimarilere göre çok daha kullanışlı bir yöntem olması ve tahmin amaçlı çalışmalarda sıkça kullanılması bu çalışmada da tercih edilmesinin en temel sebeplerini oluşturmaktadır.

Normalleştirme

İkinci bölümde belirtildiği gibi, yapay sinir ağları uygulamalarında eğitim süreci başlamadan önce girdi ve çıktılarının normalleştirilmesi yani $[0,1]$ ya da $[-1,1]$ aralığına getirilmesi gerekmektedir. Çalışma esnasında, altta gösterilen yaygın formüllerin hepsi denenmiş ve eldeki verilerde en iyi sonuçların “düzeltilmiş normalleştirme” formülüyle elde edildiği gözlenmiştir:

- $X_{norm} = X / X_{maks}$,
- $X_{norm} = (X - X_{min}) / (X_{maks} - X_{min})$,
- $X_{norm} = [2*(X - X_{min}) / (X_{maks} - X_{min})] - 1$ (düzeltilmiş normalleştirme).

Eğitim, Test ve Doğrulama Kümeleri

Yapay sinir ağı modelini geliştirmek için örneklem büyüklüğünün % 80’ i eğitim kümesi, % 20’ si de test kümesi olarak belirlenmiştir.

Parametrelerin Belirlenmesi

Başarılı bir yapay sinir ağları tahminlemesi için aşağıdaki temel parametrelerin optimum olarak belirlenmesi gerekmektedir:

- Girdi nöronu sayısı
- Gizli katman sayısı
- Gizli nöron sayısı
- Çıktı nöronu sayısı

Girdi nöronu sayısı

Neden-sonuç ilişkisine dayanan tahmin problemlerinde girdi nöronu sayısı girdi vektöründeki değişken sayısına eşit olduğu için bu parametre kolaylıkla belirlenmiştir. Veri kümesindeki analizlerde kullanılan değişken sayısı 43 olduğu için bu değer girdi nöronu sayısı olarak alınmıştır.

Gizli katman sayısı

Yazındaki çalışmalarda genellikle tek gizli katmanın ya da en fazla iki katmanın kullanılarak başarılı sonuçlar elde edildiği görülmüştür. Bu sayının artırılmasının, hesaplama zamanını da artırmasının yanında ağır öğrenme yerine ezberlemesine neden olduğu bilinmektedir. Bu nedenle, bu uygulamada da verilerin çokluğu göz önüne alınarak iki gizli katman kullanılmasına karar verilmiştir.

Gizli nöron sayısı

Yine yazındaki çalışmalarda gizli nöron sayısının tayininde çok belirgin bir formülün bulunmadığı, bu sayının sezgisel olarak deneme-yanılma yöntemiyle belirlendiği, genelde az sayıda nöronun tercih edildiği ve başlangıç değeri olarak rastgele bir değer, girdi sayısı ya da girdi ve çıktı sayısı toplamının yarısının alındığı görülmüştür. Bu bilgilere istinaden bu çalışmada da deneme-yanılma yöntemiyle optimum sayının bulunmasına çalışılmıştır.

Çıktı nöronu sayısı

Çıktı nöronu sayısı neden sonuç ilişkisine dayalı problemlerde çıktı sayısına eşit olarak alındığı için bu parametre de kolaylıkla belirlenerek 1 olarak alınmıştır.

Aktivasyon Fonksiyonu ve Öğrenme Algoritması

Yazındaki yaygın kullanımlar dikkate alınarak, aktivasyon fonksiyonu olarak Sigmoid (Lojistik) fonksiyonu, öğrenme algoritması olarak geri yayılım (back-propagation) algoritması kullanılmıştır.

Geri yayılım algoritmalarında ağırlıkların değişim büyüklüğünü belirleyen öğrenme oranı (katsayısı) ile eğitim sürecini hızlandırmayı sağlayan momentum katsayısının belirlenmesi için yazında yine belirli bir formül bulunmadığı ve deneme-yanılma yönteminin kullanıldığı görülmüştür. Bu çalışmada da benzer yöntemle optimum parametrelerin bulunması yoluna gidilmiştir.

Bu değerlendirilmelerin sonucunda en uygun yapay sinir ağı yapısı Tablo 4.7' de özetlenmiştir.

Tablo 4.7. İlk amaca uygun yapay sinir ağı yapısı modeli

| Girdi nöronu sayısı | Gizli katman sayısı | Gizli nöron sayısı | Çıktı nöronu sayısı: | Aktivasyon Fonksiyonu |
|---------------------|---------------------|--------------------|----------------------|-----------------------|
| 43 | 2 | 2198 | 1 | Sigmoid |

Değerlendirme

Modellerin tahmin doğruluğunun karşılaştırılması için kullanılan OMH, OMYH ve OHKK değerleri Tablo 4.8' de gösterilmiştir.

Tablo 4.8. Modellerin karşılaştırmasına yönelik doğruluk ölçüleri

| Model | OMH | OMYH | OHKK |
|--------------------|-----------------|--------------|-----------------|
| Regresyon Analizi | 3826,016 | 0,462 | 6222,049 |
| Yapay Sinir Ağları | 2785,218 | 0,336 | 4893,494 |

İkinci bölümde bahsedildiği gibi, tahmin doğruluğunun belirlenmesinde ideal olarak doğruluk ölçülerinin tümünün birden düşük olması arzu edilse de, bu her zaman mümkün olmayabilmektedir. Ancak Tablo 4.8 incelendiğinde tüm doğruluk ölçülerinde en küçük değerlerin yapay sinir ağlarında elde edildiği görülmektedir. Bu sonuçtan hareketle, uygun parametrelerin seçildiği yapay sinir ağı yapısının ikinci el otomobil araçlarının güncel fiyatlarını tahmin etmede başarıyla kullanılabileceği görüşü savunulabilir.

4.3.4.2. Gelecekteki Fiyatları Tahmin Etmeye Yönelik Modelleme

Bu bölümde geleneksel zaman serileri analiz yöntemleri ve yapay sinir ağı yapısı kullanılarak, araçların gelecek fiyatlarını tahmin etmeye yönelik modeller kurularak tahmin

doğrulukları karşılaştırılmıştır.

Veri Kümesi

Bu kısımdaki uygulamanın veri kümesini Tablo 4.9’ da gösterilmiş olan 3 popüler otomobil modelinin, 1 yaşında (12 aylık) ve 20.000 km’ de Ocak 2005 – Aralık 2007 tarihleri arasındaki internet ilanlarında yer alan günlük fiyat verilerinin aylık ortalamalarından oluşan zaman serileri oluşturmaktadır. Bu otomobil modelleri Tablo 4.9’ da görülmektedir:

Tablo 4.9. Gelecekteki fiyatları tahmin etmede kullanılan otomobil modelleri

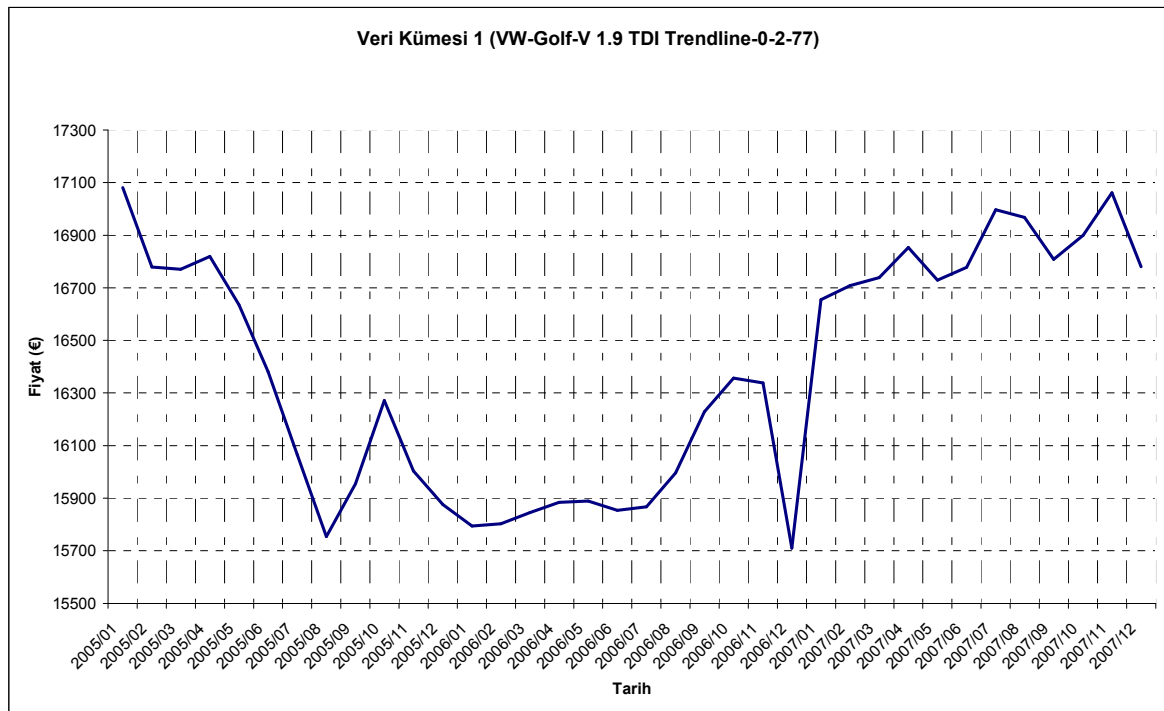
| Marka | Temel Model | Detay Model | Kasa Tipi | Kapı Sayısı | Motor Gücü |
|----------|-------------|---------------------|-----------|-------------|------------|
| VW | Golf | V 1.9 TDI Trendline | 0 | 2 | 77 |
| BMW | 525 | D DPF | 0 | 4 | 130 |
| MERCEDES | A150 | Classic | 0 | 3 | 70 |

Veri kümesi, belirli bir otomobilin zaman içerisinde kullanıma bağlı olarak değişen fiyatlarını değil, tamamen aynı temel ve opsiyonel özelliklere sahip olan otomobillerin sabit bir yaş ve kat edilen mesafe (km) ile zaman içerisinde aldığı fiyatları göstermektedir. Örnekle açıklamak gerekirse; herhangi bir veri kümesindeki ilk değer, bu veri kümesine konu olan otomobil modelinin 1 yaşında ve 20.000 km göstergesine sahip olması halinde Ocak 2005’ de pazardaki satış (ilan) fiyatını gösterirken, son değer ise yine aynı model bir otomobilin 1 yaşında ve 20.000 km göstergesine sahipken Aralık 2007’ deki satış (ilan) fiyatını göstermektedir.

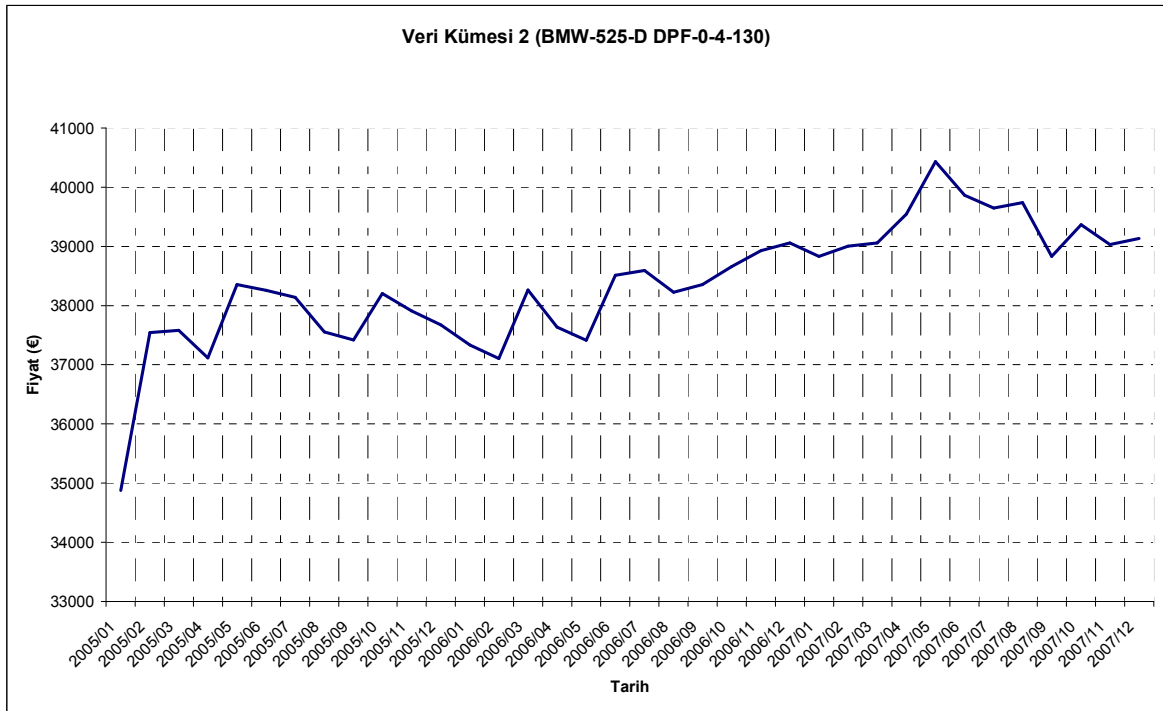
Elde edilen veri kümeleri Tablo 4.10’ da, grafikler ise Şekil 4.4, Şekil 4.5 ve Şekil 4.6’ da verilmiştir.

Tablo 4.10: Veri kümeleri

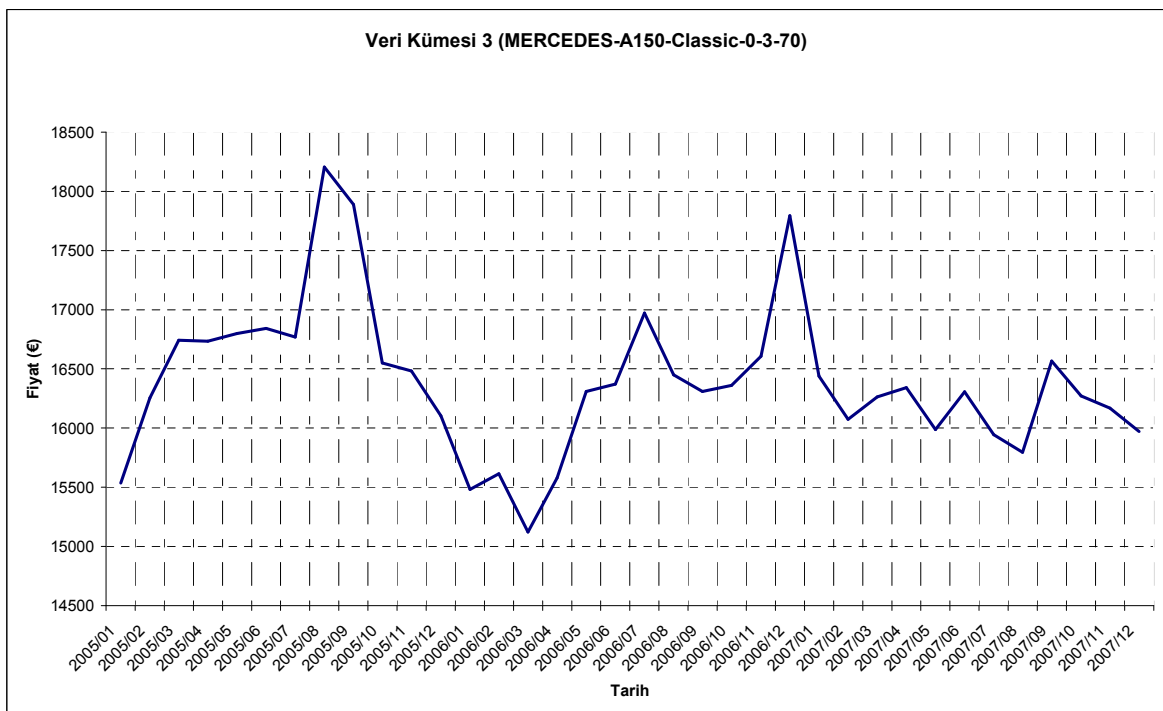
| | Veri Kümesi 1 | | | Veri Kümesi 2 | | | Veri Kümesi 3 | | |
|----------------|---------------|-------|-------|---------------|-------|-------|---------------|-------|-------|
| | 2005 | 2006 | 2007 | 2005 | 2006 | 2007 | 2005 | 2006 | 2007 |
| Ocak | 17081 | 15794 | 16655 | 34875 | 37334 | 38833 | 15535 | 15481 | 16438 |
| Şubat | 16779 | 15802 | 16709 | 37543 | 37108 | 39001 | 16255 | 15613 | 16073 |
| Mart | 16771 | 15845 | 16739 | 37582 | 38261 | 39059 | 16741 | 15122 | 16262 |
| Nisan | 16819 | 15884 | 16853 | 37113 | 37637 | 39545 | 16733 | 15579 | 16341 |
| Mayıs | 16635 | 15890 | 16729 | 38352 | 37416 | 40432 | 16798 | 16311 | 15985 |
| Haziran | 16381 | 15854 | 16778 | 38258 | 38512 | 39862 | 16843 | 16372 | 16306 |
| Temmuz | 16066 | 15868 | 16997 | 38139 | 38593 | 39646 | 16769 | 16970 | 15944 |
| Ağustos | 15754 | 15996 | 16967 | 37550 | 38227 | 39739 | 18204 | 16449 | 15794 |
| Eylül | 15955 | 16229 | 16808 | 37417 | 38354 | 38830 | 17888 | 16311 | 16565 |
| Ekim | 16271 | 16357 | 16900 | 38205 | 38655 | 39364 | 16548 | 16361 | 16271 |
| Kasım | 16004 | 16338 | 17061 | 37907 | 38931 | 39032 | 16482 | 16607 | 16167 |
| Aralık | 15876 | 15710 | 16780 | 37672 | 39060 | 39135 | 16105 | 17795 | 15971 |



Şekil 4.4. Veri Kümesi 1'e ait Ocak 2005 – Aralık 2007 arası aylık fiyat değişimi



Şekil 4.5. Veri Kümesi 2' ye ait Ocak 2005 – Aralık 2007 arası aylık fiyat değişimi



Şekil 4.6. Veri Kümesi 3' e ait Ocak 2005 – Aralık 2007 arası aylık fiyat değişimi

Zaman Serisi Yöntemleri

Zaman serisi yöntemlerine yönelik işlemlerde R ve SPSS yazılımları kullanılarak optimum parametre tespitleri yapılmaya çalışılmıştır. Zaman serisi yöntemlerini kullanarak tahmin çalışmalarında veri kümesini oluşturan serilerin öncelikle durağanlaştırılması, yani varsa trendden arındırılması ve mevsimsel ayrıştırılmasının yapılması gerekmektedir. Günümüzde bu işlemler, gerek üstte adı geçen yazılımlar, gerekse gelişmiş diğer birtakım yazılımlar kullanılarak kullanıcının yönlendirmesi yardımıyla otomatik olarak yapılmaktadır. Bu nedenle çalışmanın kapsamını gereksiz yere genişletmemek için bu işlemlerin ayrıntılarına girilmeyip sadece kullanılan yazılımlardan elde edilen optimum sonuçlar sunulacaktır

Bu aşamada kullanılan yöntemler iki ana kategoride gösterilebilir:

- “Üstel Düzleştirme” yöntemleri için; mevsimsellik içermeyen serilerde: Basit trend, Holt’ un doğrusal trendi, Brown’ un doğrusal trendi, Damped trendi; mevsimsellik içeren serilerde; Basit mevsimsel model, Winter’ ın toplamsal modeli ve Winter’ ın çarpımsal modeli.
- ARIMA (Box-Jenkins) yöntemleri için; AR(p) (Autoregressive), Integration (d) ve MA(q) (Moving Average). Bu yöntemler, mevsimselliğe göre de periyod=12 alınarak ayrıca denenmiştir.

Zaman serileri yöntemlerinin uygulaması sonucunda elde edilen sonuçlar Tablo 4.11’ de gösterilmiştir. Her veri kümesindeki en düşük hata değerleri kalın renkle belirtilmiştir. Buna göre, optimum zaman serisi yöntemleri; 1., 2. ve 3. veri kümeleri için sırasıyla ARIMA(0,1,0)(0,0,0), Winter’ın Toplamsal ve ARIMA(1,0,0)(0,0,0) olarak bulunmuştur.

Tablo 4.11. Modellerin karşılaştırmasına yönelik doğruluk ölçüleri

| Veri Kümesi | Doğruluk ölçüsü | Üstel Düzleştirme | ARIMA |
|---------------|-----------------|----------------------------|----------------------------|
| Veri Kümesi 1 | | Basit Mevsimsel | ARIMA(0,1,0)(0,0,0) |
| | OMH | 142,873 | 134,314 |
| | OMYH | 0.875 | 0,821 |
| | OHKK | 192.694 | 170,455 |
| Veri Kümesi 2 | | Winter’ın Toplamsal | ARIMA(0,1,0)(1,1,0) |

| | | | |
|---------------|------|-----------------|---------------------|
| | OMH | 381.887 | 562,943 |
| | OMYH | 1,001 | 1,457 |
| | OHKK | 501,491 | 746,873 |
| Veri Kümesi 3 | | Basit Mevsimsel | ARIMA(1,0,0)(0,0,0) |
| | OMH | 376,699 | 259,547 |
| | OMYH | 2,278 | 1,605 |
| | OHKK | 485,931 | 356,927 |

Yapay Sinir Ağları Modellemesi

İlk amacı gerçeklemeye yönelik yapay sinir ağları modellemesinde zaten anlatılmış olan ağ ve parametre belirleme işlemleri bu kısımda da kullanıldığı için bu işlemlerin ayrıntıları tekrar edilmeden özet olarak anlatılacaktır.

Bu uygulamada yine Çok Katmanlı Geri Yayılımlı Yapay Sinir Ağı kullanılmıştır.

Verilerin normalleştirilmesi için, ilk uygulamada belirtilen formüller deneme-yanılma yoluyla test edilmiş ve sonuçta “düzeltilmiş normalleştirme” formülünün kullanılmasına karar verilmiştir. Böylece tüm veriler $[-1,1]$ aralığına getirilmiştir.

Eğitim, Test ve Doğrulama Kümeleri

Eğitim ve test kümeleri için sırasıyla %70-%30, %80-%20 ve %90-%10 oranları denenmiş ve en başarılı sonuçların %90-%10 oranlarında elde edildiği gözlenmiştir. Eğitim kümesi için böylesine yüksek oranın ortaya çıkmasının, örneklem sayısının azlığından kaynaklandığı düşünülebilir.

Parametrelerin Belirlenmesi

Girdi nöronu sayısı: Zaman serisi verilerinin tahminine yönelik problemlerde girdi nöronu sayısı, neden-sonuç ilişkisinde olduğu gibi kolayca girdi değişken sayısı olarak belirlenmemektedir. Bu sayı yazındaki pek çok çalışmada olduğu gibi bu çalışmada da deneme-yanılma yöntemiyle belirlenmiştir.

Gizli katman sayısı ve Gizli nöron sayısı: deneme-yanılma yöntemiyle tespit

edilmiştir.

Çıktı nöronu sayısı: Zaman serisi tahmin problemlerinde çıktı nöronu sayısı tahmin döneminin uzunluğuna eşittir. Bu çalışmada tek dönemlik tahmin yapıldığı için çıktı nöron sayısı yine kolaylıkla belirlenerek 1 olarak alınmıştır.

Aktivasyon Fonksiyonu ve Öğrenme Algoritması

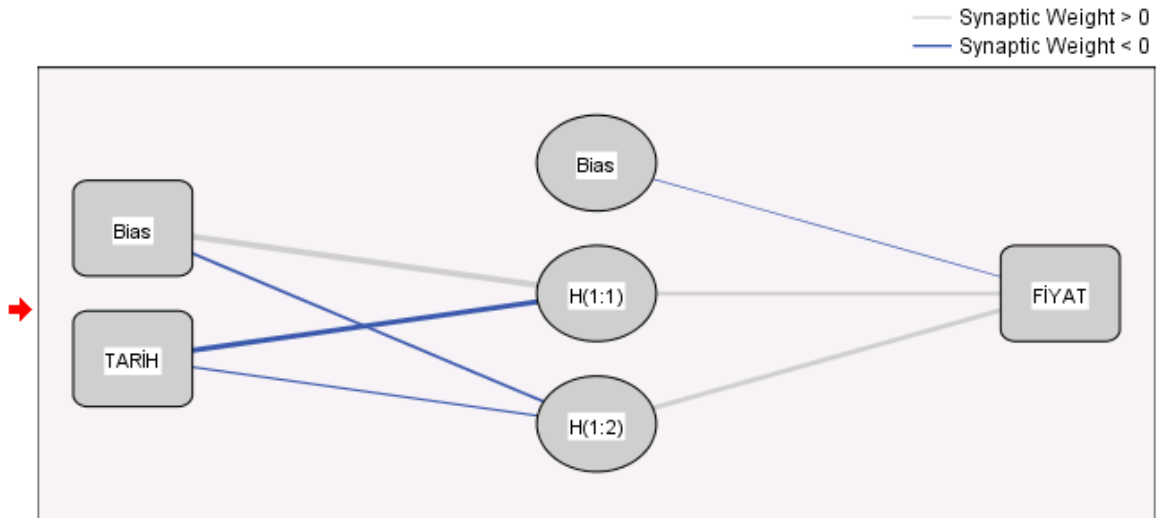
Bu bölümde kullanılan veri sayısının az olması nedeniyle (her veri kümesinde 36' şar adet) aktivasyon fonksiyonu için Sigmoid ve Hiperbolik Tanjant fonksiyonları sırasıyla denenmiştir. Öğrenme algoritması olarak geri yayılım (back-propagation) algoritması tercih edilmiş, öğrenme oranı (katsayısı) ve momentum katsayılarının belirlenmesi için ise deneme-yanılma yöntemleri kullanılmıştır.

Bu değerlendirilmelerin sonucunda ikinci amaca en uygun yapay sinir ağı modeli tüm veri kümelerinde aynı çıkmış ve Tablo 4.12' de özetlenmiştir.

Tablo 4.12. İkinci amaca uygun yapay sinir ağı modeli

| Girdi nöronu sayısı | Gizli katman sayısı | Gizli nöron sayıları | Çıktı nöronu sayısı | Aktivasyon Fonksiyonu |
|---------------------|---------------------|----------------------|---------------------|-----------------------|
| 1 | 1 | 2 | 1 | Hiperbolik Tanjant |

Ağ yapısı şematik olarak Şekil 4.7' de görülmektedir.



Şekil 4.7. İkinci amaca uygun yapay sinir ağı modelinin şematik yapısı

Değerlendirme

Zaman serileri yöntemlerinin uygulaması sonucunda elde edilen ve Tablo 4.11’ de gösterilmiş olan sonuçların içerisinde en yüksek doğruluğu veren yöntemler ile yapay sinir ağlarının uygulaması sonucu elde edilen değerler birleştirilerek Tablo 4.13’ de gösterilmiştir. Tablonun incelenmesiyle de anlaşılacağı üzere, tüm veri kümelerinde en küçük hata değerlerinin yapay sinir ağlarında elde edildiği görülmektedir.

Tablo 4.13. Modellerin karşılaştırmasına yönelik doğruluk ölçüleri

| Veri Kümesi | Doğruluk ölçüsü | Zaman Serisi Yöntemleri | Yapay Sinir Ağları |
|---------------|-----------------|-------------------------|--------------------|
| Veri Kümesi 1 | | ARIMA(0,1,0)(0,0,0) | |
| | OMH | 134,314 | 106,659 |
| | OMYH | 0,821 | 0,653 |
| | OHKK | 170,455 | 151,127 |
| Veri Kümesi 2 | | Winter’in Toplamsal | |
| | OMH | 381.887 | 297,795 |
| | OMYH | 1,001 | 0,778 |
| | OHKK | 501,491 | 369,775 |
| Veri Kümesi 3 | | ARIMA(1,0,0)(0,0,0) | |
| | OMH | 259,547 | 249,879 |
| | OMYH | 1,605 | 1,502 |
| | OHKK | 356,927 | 315,837 |

Bu sonuçlardan hareketle, ikinci el otomobil fiyatlarından oluşan zaman serilerinin ileriye yönelik tahminlerinde yapay sinir ağlarının geleneksel zaman serisi yöntemlerine göre daha güvenilir sonuçlar verdiği iddia edilebilir.

4.3.5. Uygulamaya Koyma

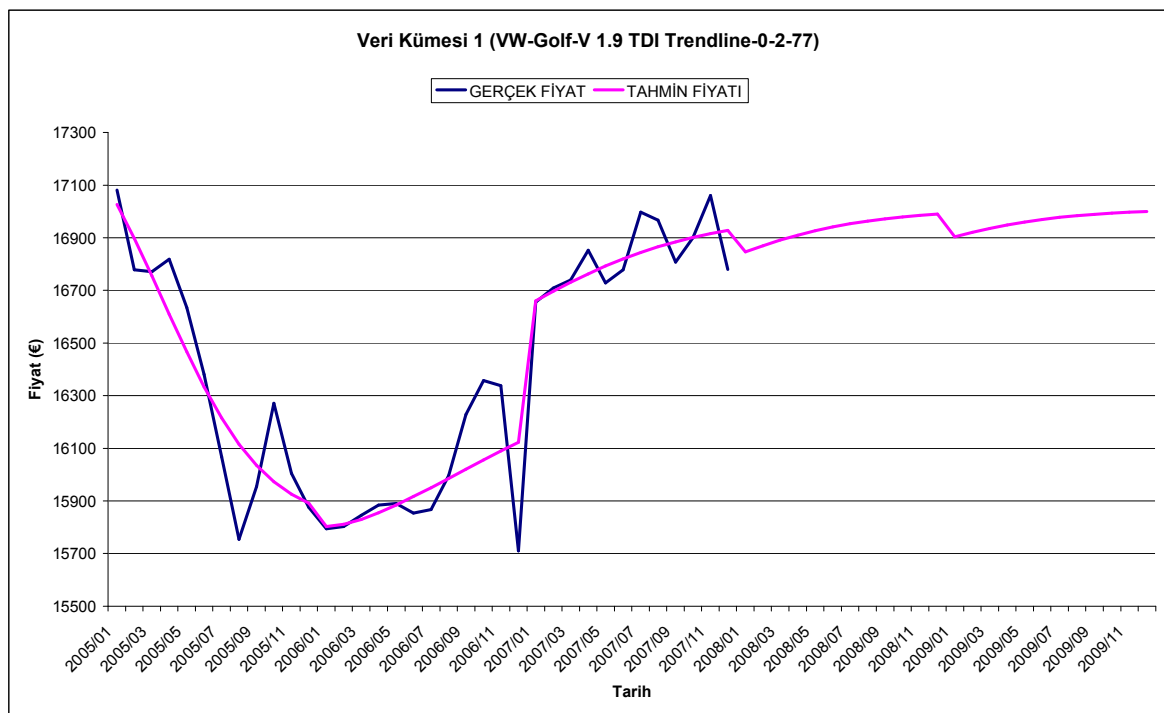
Değerlendirmeler sonucunda uygun bulunan modeller çalışmanın yapıldığı işletmenin müşterilerine hizmet veren mevcut web sitesinin içerisine gömülerek uygulamaya geçirilmiştir. Bu amaçla işletmenin web sitesinin yazıldığı Java kodlarında gerekli güncellemeler yapılmıştır. Ticari faaliyet sürdüren ilgili işletmede geliştirilen SQL ve Java kodları ile web ara yüzleri gizlilik anlaşması gereği tez metni içerisine yerleştirilmemiştir. Ancak çalışmada ortaya konan modeller ve sonuçlar, tez yazarını bilgilendirmek kaydıyla,

ilgili kişi ya da kurumların kendi geliştirecekleri web arayüzleri ile de ticari ya da bilgilendirme amaçlı olarak yararlanıcıların hizmetine sunulabilir.

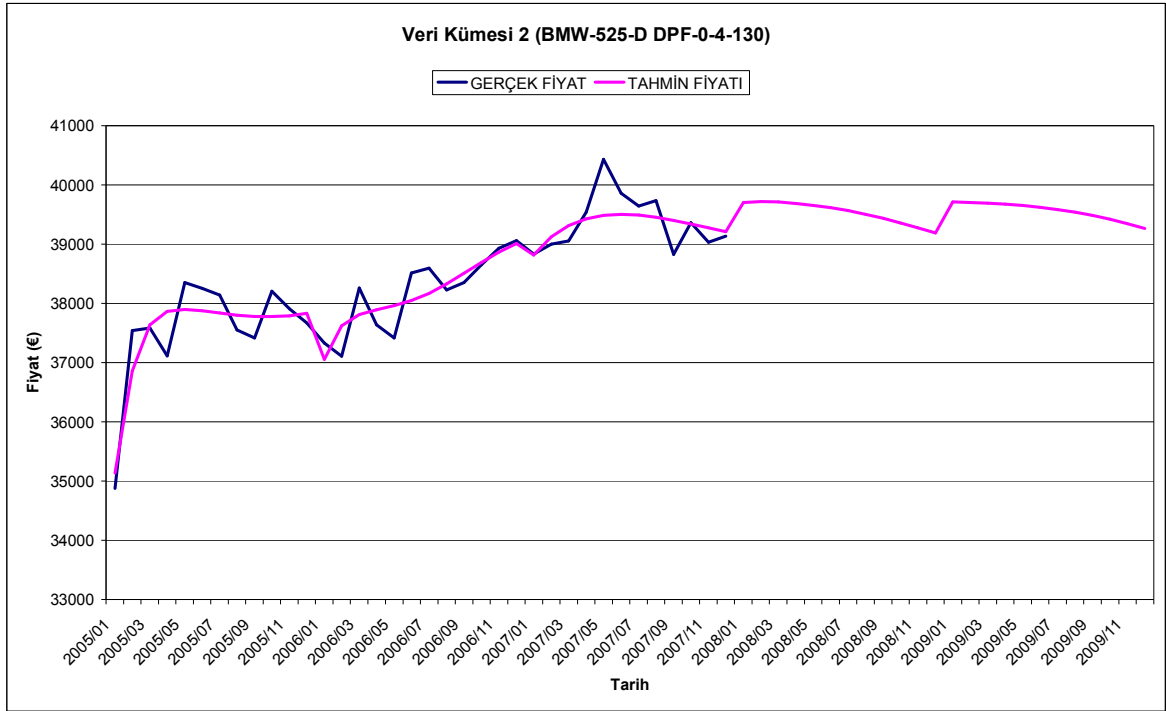
Sonuçları görselleştirmeye örnek olarak, çalışmanın ikinci uygulamasında yapay sinir ağları ile elde edilen ileriye dönük iki yıllık tahmin değerleri, üç veri kümesi birleştirilerek Tablo 4.14’ de; gerçek ve tahmin değerlerini birlikte gösteren grafikler ise her veri kümesi için sırasıyla Şekil 4.8, Şekil 4.9 ve Şekil 4.10’ da sunulmuştur.

Tablo 4.14. Veri kümelerinin 2008-2009 tahmin değerleri

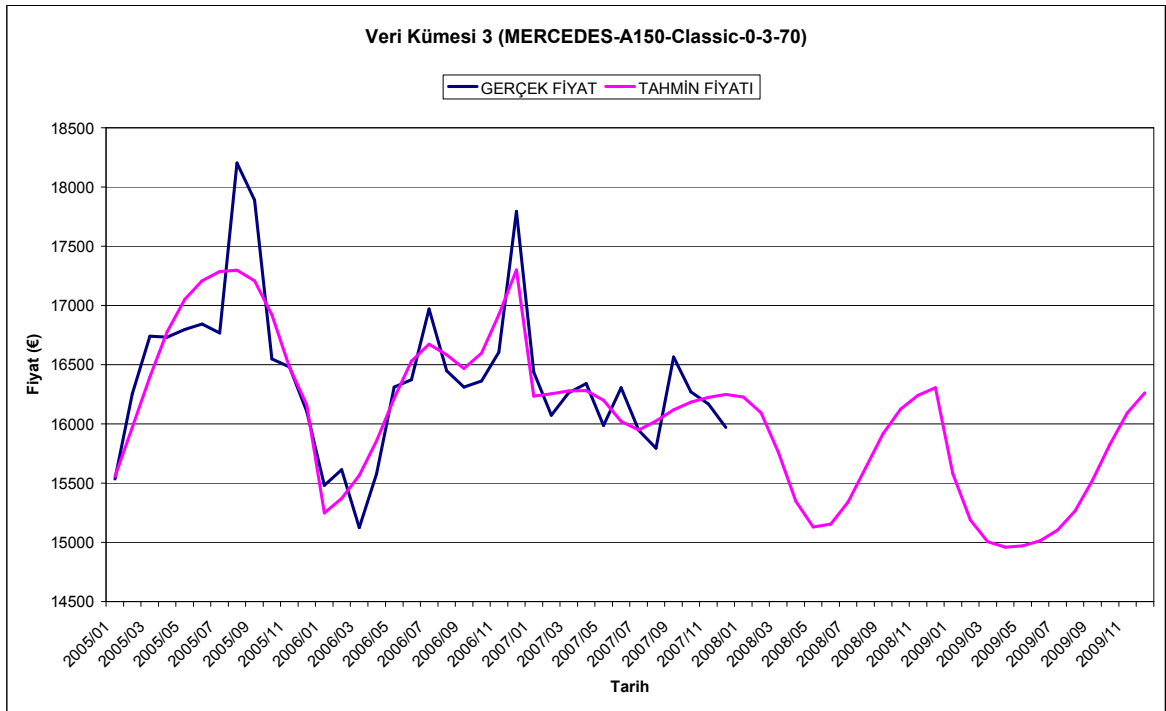
| | Veri Kümesi 1 | | Veri Kümesi 2 | | Veri Kümesi 3 | |
|----------------|---------------|-------|---------------|-------|---------------|-------|
| | 2008 | 2009 | 2008 | 2009 | 2008 | 2009 |
| Ocak | 16847 | 16904 | 39707 | 39712 | 16227 | 15582 |
| Şubat | 16870 | 16921 | 39721 | 39705 | 16094 | 15187 |
| Mart | 16891 | 16936 | 39712 | 39693 | 15764 | 15004 |
| Nisan | 16910 | 16949 | 39690 | 39675 | 15345 | 14958 |
| Mayıs | 16927 | 16960 | 39659 | 39653 | 15129 | 14968 |
| Haziran | 16941 | 16970 | 39618 | 39624 | 15153 | 15012 |
| Temmuz | 16953 | 16978 | 39567 | 39588 | 15341 | 15101 |
| Ağustos | 16964 | 16984 | 39507 | 39543 | 15629 | 15265 |
| Eylül | 16972 | 16989 | 39437 | 39488 | 15916 | 15521 |
| Ekim | 16980 | 16994 | 39359 | 39424 | 16123 | 15828 |
| Kasım | 16986 | 16997 | 39276 | 39350 | 16242 | 16091 |
| Aralık | 16991 | 17000 | 39192 | 39269 | 16305 | 16262 |



Şekil 4.8. Veri kümesi 1 için gerçek ve tahmini fiyatlar



Şekil 4.9. Veri kümesi 2 için gerçek ve tahmini fiyatlar



Şekil 4.10. Veri kümesi 3 için gerçek ve tahmini fiyatlar

SONUÇ VE ÖNERİLER

Bu çalışmada ikinci el otomobillerin pazardaki güncel fiyatları ile gelecekteki fiyatları tahminlenmeye çalışılmıştır. Bu amaçla ilk olarak internet üzerinden elde edilmiş olan veriler analiz edilerek hem bu verilerin veri madenciliği analizlerine uygun hale getirilmesine, hem de verilerdeki nitel ve nicel sorunların düzeltilerek daha uygun verilerin elde edilmesine çalışılmıştır.

Gerekli çalışmaların sonucunda analizlere elverişli hale getirilen veriler kullanılarak iki tahminleme uygulaması gerçekleştirilmiştir. İlk uygulamada ikinci el otomobillerin pazar fiyatları modellenmeye çalışılmıştır. Bu amaçla otomobillerin temel özelliklerini, opsiyonel özelliklerini ve kullanıma bağlı olarak değişen özelliklerini (yaş, km gibi) içeren veriler bir veri ambarında toplanarak regresyon analizi ile yapay sinir ağları uygulamalarına tabi tutulmuştur. Elde edilen sonuçlar karşılaştırıldığında yapay sinir ağlarının, regresyon analizine göre daha başarılı olduğu gözlenmiştir.

İkinci uygulama olarak ikinci el otomobillerin gelecekteki fiyatları tahmin edilmeye çalışılmıştır. Veri ambarında çok sayıda otomobil bulunmasından dolayı, veri kümesi olarak internette en çok ilanı yayınlanan, diğer bir deyişle en popüler olan üç otomobil modeli tespit edilmiştir. Bu modellere ait geçmiş üç yıllık veriler üzerinde çeşitli zaman serisi yöntemleri ile yapay sinir ağları uygulanmıştır. Sonuçların karşılaştırılmasıyla, yapay sinir ağlarının zaman serileri yöntemlerine göre de daha başarılı olduğu anlaşılmıştır.

Uygulamalarda kullanılan tüm yöntemlerde OMYH değerinin % 10' un altında çıkması, oluşturulan tüm modellerin “çok iyi” olduğunu göstermektedir. Bu nedenle, çalışmada oluşturulan tüm modeller ikinci el otomobil fiyatlarının tahmin edilmesinde başarıyla kullanılabilirler. Diğer yandan, kullanılan yöntemler arasında gözlenen hata değerleri arasındaki farklar çok yüksek olmasa da, yapay sinir ağlarının tüm veri kümelerinde daha başarılı olması, bu yöntemin klasik istatistiksel yöntemlere alternatif oluşturabileceğini göstermiştir. Sonuç olarak, uygun parametrelerin seçildiği yapay sinir ağlarının ikinci el otomobillerin güncel ve gelecek fiyatlarını tahmin etmede başarıyla kullanılabileceği ifade edilebilir.

Çalışma sonucunda ortaya çıkan modeller, uygun bir web yazılımı ile internet ortamında yararlanıcıların hizmetine sunabilir. Bu yararlanıcılar arasında; ikinci el

otomobil pazarı ile ilişkisi olan tüketiciler, otomobil üreticileri, finans kuruluşları, otomobil kiralama firmaları, otomobil galericileri ve sigorta şirketleri sayılabilir.

İleriki çalışmalara ışık tutması beklentisiyle, çalışmada yaşanan bazı kısıtlamalar ve öneriler aşağıda özetlenmeye çalışılmıştır.

İkinci el otomobillerin bir ülkenin farklı bölge ya da şehirlerinde farklı fiyatlara sahip olduğu bilinen bir gerçektir. Çalışmada; internette yer alan ilanların çoğunda şehir adının değil yerel alan kodunun (Zip_code) bulunduğu gözlemlenmiştir. Bu kodlar; mahalle, semt, cadde gibi yerel alanları gösterdiği için doğal olarak analizlerde önemli bir etkisi görülememiştir. Örneğin aynı şehirdeki farklı mahalle ya da semtlerde bulunan otomobillerin fiyat ilişkileri ortak bir şehir kodu altında incelenememiştir. Verilerin toplandığı ülkelerdeki alan kodu farklılıkları nedeniyle bu kodlardan benzersiz (unique) şehir ya da bölge kodlarının kolayca elde edilemeyeceği ve çözümü için ek bir çalışma yapılması gerektiği anlaşıldığı için, çalışma kapsamında zaman kısıtı nedeniyle bu ayrıştırma işlemine girilmemiştir. İleriki çalışmalarda bu işlemin gerçekleştirilmesi, modelin açıklayıcılığını artıracak gibi şehirlerin ya da bölgelerin fiyatlar üzerindeki etkisini görme fırsatı da sunacaktır.

Box-Jenkins yöntemlerinde etkin bir tahmin yapabilmek için en azından 50 veri bulunması önerilmesine rağmen (Chatfield, 1996) çalışmamızda farklı araç modellerini ifade eden her veri kümesi için sadece 36 adet (3 yıllık) veri kullanılabilmiş ve her veri kümesi diğer veri kümelerinden bağımsız olarak analize tabi tutulmuştur. Çalışmada kullanılan Box-Jenkins yönteminin yapay sinir ağlarına kıyasla daha az başarılı olmasının sebeplerinden birisi bu olabilir. Yapay sinir ağlarının, az sayıda veriye rağmen güvenilir sonuçlar vermesi bu yöntemin güçlülüğünü ortaya koymaktadır. İleriki çalışmalarda, hem zaman serisi analiz yöntemlerinden hem de yapay sinir ağlarından daha iyi sonuçlar elde etmek için mümkünse öncelikle tarihsel veri sayısının artırılması önerilebilir. Ayrıca, ortak trendlerden etkilenen veri kümelerinin birlikte analiz edilmesini sağlayacak yöntemlerin araştırılıp uygulanması da araştırmacılara daha iyi sonuçlar verebilir.

Bunlara ilaveten, ikinci el otomobil fiyatlarının tahminlenmesine yönelik veri madenciliği çalışmalarında tahmin doğruluğunu artırmak amacıyla farklı yöntemlerin birlikte kullanıldığı melez (hybrid) yöntemlerin kullanılması önerilebilir. Örneğin güncel pazar fiyatlarının tahmin edilmesine yönelik uygulamalarda regresyon analizi ile modele

anlamli katkısı olan deęişkenler belirlenip yapay sinir aęlarının girdisi olarak sadece bu deęişkenler kullanılabilir. Dięer yandan, mevsimsellięin yüksek olduęu zaman serilerinde gelecekteki fiyatları tahmin etmek için öncelikle zaman serisi yöntemleri, ardından yapay sinir aęları uygulanarak her iki yöntemden elde edilen deęerlerin ortalamaları alınabilir. Yapay sinir aęının eğitim sürecini hızlandırmak, aęın ezberlemesini önlemek ve yerel minimum noktalarına takılmasından kurtarmak için ise “yapay-bulanık” (neuro-fuzzy) melez yöntemleri kullanılabilir.

Çalışmanın metodolojisi referans alınarak farklı sektörlerde de benzer veri madencilięi uygulamaları yapılabilir. Yazın taramasında özellikle otomotiv ve emlak endüstrilerinin bu konuda çok bakir olduęu görülmüştür. Hâlbuki her iki sektörün de insan hayatında çok önemli yeri bulunmaktadır. İkinci el otomobil pazarında gerçekleştirilen bu uygulamaya paralel olarak ikinci el emlak sektöründe de geliştirilebilecek benzer uygulamalar ilgili kiři ve kurumlara önemli yararlar sağlayacaktır. Ancak gözden kaçırılmaması gereken nokta; bu çalışmaların başarısının ardında sistematik olarak uzun süreli veri toplama ve hazırlama işlemleri bulunmaktadır. Günümüzde ikinci el konutların da internet ortamında yoğun bir şekilde satışa sunulduęu göz önüne alınarak bu alanda da bir an önce adı geçen işlemlerin başlatılması tavsiye edilebilir.

Çalışmada kullanılan veri madencilięi yazılımlarının kendilerine özgü pek çok avantaj ve dezavantaj içerdięi görülmüştür. Bu yazılımların veri madencilięi yöntemlerini kullanım kolaylıkları, analizlerde sunduęu ayrıntılı seçenekleri ve analiz hızları açısından karşılaştırılmasına yönelik yapılacak çalışmalar araştırmacılara yazılım seçimi ve kullanımı aşamalarında önemli katkılar sağlayabilir.

KAYNAKÇA

- Adriaans P., Zantinge D., Data Mining, Addison Wesley Longman, USA, 1998.
- Agrawal R., Imielinski T., Swami A., “Mining Association Rules Between Sets of Items in Large Databases”, In Proc. of The ACM Sigmod Conference On Management of Data, Washington, 1993.
- Ahmed S., Cross J., “A Tourist Growth Model To Predict Accommodation Nights Spent in Australian Hotel Industry”, 11th Annual Colloquium of The Spatial Information Centre, University of Otago, New Zealand, 1999, 1-9.
- Akgül, I., Zaman Serilerinin Analizi ve ARIMA Modelleri, Der Yayınları, İstanbul, 2003.
- Akpınar H., “Veritabanlarında Bilgi Keşfi ve Veri Madenciliği”, İ.Ü. İşletme Fakültesi Dergisi, c. 29, (Nisan 2000), 1-22.
- Akpınar H., “Yapay Sinir Ağları Gelişim ve Yapılarının İncelenmesi”, İ.Ü. İşletme Fakültesi Dergisi c. 23, (Nisan 1994), 41-77.
- Alkan A., Predictive Data Mining with Neural Networks and Genetic Algorithms, Ph.D. Thesis, İTÜ, İstanbul, 2001, 51.
- Alkan, S., Otomotiv Sektörü, Türkiye Vakıflar Bankası Yayınları, Ankara 1996, 6.
- Almuallim H., Dietterich T.G., “Learning with Many Irrelevant Features”, 9th National Conference on Artificial Intelligence, USA, 1991, 547–552.
- Alpaydın E., “Zeki Veri Madenciliği”, Bilişim 2000 Eğitim Semineri Notları, Boğaziçi Üniversitesi, 2000.
- Al-Saba T., El-Amin L., “Artificial Neural Networks As Applied To Long-term Demand Forecasting”, Artificial Intelligence in Engineering, No. 13, (1999), 189-197.
- Anderson, D., Mc Neill, G., Artificial Neural Networks Technology, Kaman Sciences Corporation, New York, 1992, 2-38.
- Anil J. K., Jianchang M., “Artificial Neural Networks: A Tutorial”, IEEE Computational Science & Engineering, Vol.29, (March 1996), 38.
- Auclair A., Feed-Forward Neural Networks Applied To The Estimation of Magnetic Distributions, M.S. Thesis, McGill University, Montreal, Canada, 2004, 128.
- Aydın Ö.,Yapay Sinir Ağlarını Kullanarak Bir Ses Tanıma Sistemi Geliştirilmesi, Yüksek Lisans Tezi, Trakya Üniversitesi, Edirne, 2005, 74.
- Aydoğmuş Z., Çöteli R., “Yapay Sinir Ağları Yardımıyla İzolatör Yüzeyinde Potansiyel Tahmini”, F.Ü. Fen ve Müh.Bil.Dergisi, 17/2, (2005), 239-246.
- Ayyıldız M., Kalıpsız O., Yavuz S., “YEEM: Yazılım Projeleri Maliyet Tahminleme Ölçev Seti ve Modeli, Eleco 2006 Bildirisi”, Bursa, 2006.
- Azcanlı A., Türk Otomotiv Sanayii'nin Tarihsel Gelişimi, OSD Yayınları, İstanbul, 1995, 11.

- BaFail A.Ö., “Applying Data Mining Techniques To Forecast Number of Airline Passengers in Saudi Arabia”, *Journal of Transportation*, 9-1, (2004), 100.
- Baldemir E., Bahar O., “Türkiye’ye Yönelik Turizm Talebinin Neural (Sinir) Ağları Modelini Kullanarak Analizi”, *Gazi Üniversitesi Ticaret ve Turizm Eğitimi Fakültesi Dergisi*, Sayı 2, (2003).
- Bedir A. , Türkiye’de Otomotiv Sanayii Gelişme Perspektifi, DPT Yayınları, 2002.
- Bedir A., Gelişmiş Otomotiv Sanayilerinde Ana-Yan Sanayi İlişkiler ve Türkiye’de Otomotiv Yan Sanayinin Geleceği, DPT Uzmanlık Tezleri, Kasım 1999.
- Berson A., Smith S., Thearling K., *Building Data Mining Applications for CRM*, Mcgraw Hill, 510, USA, 1999.
- Bhattacharya S., *A Comparative Study of Different Methods of Predicting Time Series*, Yüksek Lisans Tezi, Concordia University, Canada, 1997.
- Bilgin G., Küresel Pazarlama Kapsamında Dünya Otomotiv Sanayi ve Türkiye Otomotiv Sanayinin Durumu, T.C. Başbakanlık Dış Ticaret Müsteşarlığı Yayınları, 1999.
- Bilim ve Teknik Dergisi*, Sayı. 409, İstanbul, Aralık 2001, 26.
- Bloomfield, G., *The World Automotive Industry*, David & Charles Inc., Vermont, 1978, 17.
- Box G., Jenkins G., *Time Series Analysis: Forecasting and Control*, Holden Day, San Francisco, 1976.
- Bucak S., Otomotiv Sektöründe Yapay Sinir Ağı Kullanarak Maliyet Tahmini, Yüksek Lisans Tezi, Sakarya Üniversitesi, 2007.
- Burger, C.J.S.C., Dohnal M., Kathrada M., Law R., “A Practitioners Guide To Time-Series Methods for Tourism Demand Forecasting—A Case Study of Durban”, *Tourism Management*, No. 22, South Africa, (2001).
- Cabena P., Hadjinian P., Stadler R., Verhees J, Zanasi, A., *Discovering Data Mining: From Concept to Implementation*, Prentice Hall, NJ, (1998), 517.
- Çakar T., Türker A. K., Toraman A., “İmalat Sistemlerinin Tasarlanmasında Yapay Sinirsel Ağların Kullanılması”, *Birinci Ulusal Zeki İmalat Sistemleri Sempozyumu Zis’96 Bildirisi*, Sakarya, (Mayıs 1996), 10.
- Çalık İ., Yapay Sinir Ağları ve Monte Carlo Benzeşimi ile Mersin Limanı’nın Yük Trafığı ve Gelir Tahmini, Yüksek Lisans Tezi, Gazi Üniversitesi, 2005.
- Carlson R. L., Umble M. M., “Statistical Demand Functions for Automobiles and Their Use for Forecasting in An Energy Crisis”, *The Journal of Business*, 2, (1980), 53.
- Çerkez, H. S., Müşteri İlişkileri Yönetiminde İş Zekası ve Veri Madenciliği Yöntemleri, Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, 2003.
- Chapman P., Clinton J., Kerber R., Khabaza T., Reinartz T., *Step By Step Data Mining Guide*, SPSS Inc., USA, 2000.
- Charmbers J. C., Mullick S. K., Smith D. D., “How To Choose The Right Forecasting Technique” *Harvard Business Review*, (July-August 1971), 55.
- Chatfield C., Prothero, D.L., “Box-Jenkins Seasonal Forecasting: Problems in A Case-

Study”, Journal of The Royal Statistical Society, C. 136, 1973.

Chatfield C., The Analysis of Time Series, Chapman & Hall, 5th Ed., New York, 1996.

Chen, Z., Data Mining and Uncertain Reasoning: An Integrated Approach, John Wiley & Sons, Inc., 370, Canada, 2001.

Cho, V., “A Comparison of Three Different Approaches To Tourist Arrival Forecasting”, Tourism Management, Volume: 24, No: 3, (2003), 323-330.

Choi H. Y., Kim W., An S. Y., “Recurrent and Decomposed Neural Network – Based Hotel Occupancy Rate”, The New Review of Applied Expert Systems, No: 4, (1997), 121-136.

Çildir İ., Otomotiv Sanayi Ürünlerinin İnternet Üzerinden Pazarlanması ve Üniversite Öğretim Üyelerinin on-Line Otomobil Satın Alma Tercihleri Üzerine Bir Uygulama, Yüksek Lisans Tezi, Anadolu Üniversitesi, 2006.

Cios K., Pedrycz W., Swiniarski R., Data Mining Methods for Knowledge Discovery, Kluwer, 1998.

Çömlekçi N., İstatistik, Kalite Matbaası, Ankara, 1979, 307-308.

Cortez P., Data Mining with Multi Layer Perceptrons: Intensive Care Medicine and Internet Traffic Forecasting, University of Minho, Portekiz, 2006.

Cummins J. D., Griepentrog G.L., “Forecasting Automobile Insurance Paid Claim Costs Using Econometric and ARIMA Models”, International Journal of Forecasting, Vol. 3, 1985, 203-216.

Dalyanoğulları M., Küreselleşmenin Otomotiv Ana ve Yan Sanayi Sektörüne Etkileri, Eskişehir, 2007.

Davis G. B., Olson M., “Management Information Systems: Conceptual Foundations, Structure & Development”, 2nd Ed., McGraw-Hill, New York, (1985), 6.

Değirmendereli, G., TBD 17. Bilişim Kurultayı, Bildiriler, 2000.

Demir H., Gümüşoğlu Ş., Üretim Yönetimi-İşlemler Yönetimi, Beta Basım Yayım, 6. Baskı, İstanbul, 2003.

Demuth H., Beale M., Neural Network Toolbox for Use with Matlab, MathWorks Inc., 2000.

Deniz E., Bulanık Mantık Tabanlı Tahmin Modeli ve Uygulaması, Yüksek Lisans Tezi, Muğla Üniversitesi, 2006.

Deogun J. S., Raghavan V. V., Sever H., “Exploiting Upper Approximations In The Rough Set Ethodology”, The First International Conference on Knowledge Discovery and Data Mining, Canada, 1995, 69-74.

DİE Yayınları: İstatistik Pozisyonlarına Bölünmüş Giriş Tarifeleri Cetveli, Ankara, 1990, 15.

Doğan E., Işık S., “Sapanca Gölü Günlük Buharlaşıma Miktarının Radyal Temelli Yapay Sinir Ağı Modeli Kullanarak Tahmin Edilmesi”, BMYS' 2005 Bildirisi, Kocaeli, Kasım 2005.

- Dumanlı R., Türkiye’de Otomotiv Sanayii’nin Durumu, Yapısal Özellikleri ve Sorunları, DPT Yayınları, Ankara, 1987.
- Efe, Ö., Kaynak O., Yapay Sinir Ağları ve Uygulamaları, Boğaziçi Üniversitesi Yayınları, Basım No: 696, İstanbul, 2000, 148.
- Eğrioğlu Ç., Aladağ H., “YSA ve ARIMA Modellerinin Melez Yaklaşımı ile Zaman Serilerinde Öngörü”, VII. Ulusal Ekonometri ve İstatistik Sempozyumu Bildirisi, İstanbul, 2002.
- Ekern S., “Forecasting with Adaptive Filtering: A Critical Reexamination”, Operational Research Quarterly, Vol. 27, No: 3, (1976), 707.
- Erdoğan, G., Yapay Sinir Ağları ile İktisadi ve Finansal Zaman Dizilerinin Kestirimi, Yüksek Lisans Tezi, Gazi Üniversitesi, Ankara, 2006.
- Fayyad U., Piatetsky-Shapiro G., Smyth P., “From Data Mining To Knowledge Discovery in Databases”, AI Magazine, (Fall 1996), 37-54.
- Fırat M., Güngör M., “Askı Madde Konsantrasyonu ve Miktarının Yapay Sinir Ağları ile Belirlenmesi”, İMO Teknik Dergi, (2004), 3267-3282.
- Frawley W. J., Piatetsky-Shapiro G., Matheus C. J., “Knowledge Discovery Databases: An Overview”, Knowledge Discovery In Databases, Cambridge, (1991), 1-27.
- Fritzer F., Moser G., Scharler J., “Forecasting Austrian HICP and its Components Using Var and ARIMA Models”, Austrian Central Bank, Working Papers 73, 2002.
- Gaynor P.E., Kirkpatrick R.C, Introduction To Time Series Modeling and Forecasting in Business and Economics, Mcgraw-Hill, Inc., New York, 1994, 625.
- Göçmençelebi K., İstatistik Metodları, Ogun Kardeşler Matbaacılık Sanayii, Ankara, 1976, 185.
- Goebel M., Gruenwald L., “A Survey of Data Mining and Knowledge Discovery Software Tools”, SIGKDD Explorations, Vol.1, USA, (1999), 20-33.
- Goldman,S., Agile Competitors and Virtual Organizations: Strategies for Enriching the Customer, Kenneth Preiss, 1995.
- Gürsakal N., Oğuzlar A., Şentürk A., “Değişen Veri Kavramı ve Yeni Alanlar”, İstatistik Araştırma Sempozyumu, (Kasım 2000).
- Gürtan K, İstatistik ve Araştırma Metodları, Fatih Yayınevi Matbaası, İstanbul, 1977, 421.
- Güvenç E., Student Performance Assesment in Higher Education Using Data Mining, Yüksek Lisans Tezi, Boğaziçi Üniversitesi, İstanbul, 2001.
- Hair J.F., Anderson R.E., Tahtam R.L., Black W.C., Multivariate Data Analysis, Pearson Education, New Jersey, 1998.
- Hamburg, M., Statistical Analysis for Decision Making, Harcourt, Brace and World Inc., New York, 1970.
- Hamzaçebi C, Kutay F., “Yapay Sinir Ağları ile Türkiye Elektrik Enerjisi Tüketiminin 2010 Yılına Kadar Tahmini”, Gazi Üniv. Müh.Mim. Fak. Der., Cilt 19, No:3, Ankara, (2004).

- Han J., Kamber M., "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, Tokyo, 2001, 30-33.
- Hanke J. E., Reitsch A. G., Wichern D. W., "Business Forecasting", Prentice Hall, 2001.
- Harrold, D., "What's Your Data Telling You ?", Control Engineering, 47(3), (2000), 9-10.
- Hastie, T., Tibshirani, Friedman, J., The Elements of Statistical Learning; Data Mining, Inference and Prediction, Springer Series in Statistics, USA, 2001.
- Haykin S., Neural Networks: A Comprehensive Foundation, Prentice Hall International Inc., 1999.
- Hosein P., "Stock Price Prediction By Artificial Neural Networks: A Study of Tehran's Stock Exchange", Management Studies Quarterly Journal of Management & Accounting School, No.31&32, (2002).
- Hyndman R. J., "Box-Jenkins Modelling", Informed Student Guide to Management Science, London, (2002), 1-2.
- Ilgaz V.T., Trade Restrictions in Motor Vehicle Industry in Turkey: European Union Demands To Eliminate Trade Restrictions On Used Motor Vehicles, Yüksek Lisans Tezi, İstanbul, 2005.
- İMKB-Sektör Araştırmaları Serisi: Otomotiv Sektörü, No:1,İMKB Yayınları, İstanbul, 1995, 3.
- İnceler H., "Yönetim Bilişim Sistemleri", Byte Dergisi, İstanbul, (Temmuz 1996), 196.
- İşeri A., Karlık B., "An Artificial Neural Networks Approach on Automobile Pricing", Expert Systems with Applications, (January 2008).
- Kaastra I., Boyd M., "Designing A Neural Network for Forecasting Financial and Econometric Time Series", Neurocomputing, Vol. 10, (1996), 215-236.
- Kaçtıoğlu S., Özen Ü., Yavuz U., Bilgisayara Giriş ve Güncel Yazılımlar-I, Aktif Yayınevi, Erzurum, 1999.
- Kantardzic M., "General Data Analysis and Data Mining", Wiley-IEEE Press, (2002), 1-18.
- Kapucugil A., Halka Arzların İlk Gün Fiyat Performanslarının Tahmininde YSA Kullanımı, Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi, İzmir, 2005.
- Kendal M., Stuart A., Ord J.K., The Advanced Theory of Statistics, Charles Friffin and Company, Belfast, 1983, 506-507.
- Kermanshahi B., Iwamiya H., "Up to Year 2020 Load Forecasting Using Neural Nets", Electrical Power and Energy Systems, Vol. 24, (2002), 789-797.
- Kim, J. Y., ANN Wave Prediction Model For Winter Storms and Hurricanes, Ph. D. Thesis, The School of Marine Science, The College of William and Marry, Virginia, 2003.
- Kira K., Rendeli L., "The feature selection problem; Traditional methods and a new algorithm", In Proceedings of AAAI 92, AAAI Press, 1992, 129-134.
- Kirişçi, Ö., "Veri Ambarı Projesinin Başarısı İçin Gerekli Faktörler", BT/Haber Gazetesi, Sayı : 219, 1999.

- Kişi Ö., Yapay Sinir Ağları ile Meteorolojik Verileri Kullanarak Buharlaştırmanın Modellenmesi, BMYS' 2005 Sempozyumu Bildirisi, Kocaeli, Kasım 2005.
- Klassen R. D., Flores B.E., “Forecasting Practices of Canadian Firms: Survey Results and Comparisons”, *International Journal of Production Economics*, (2001), 70.
- Kovalerchuk, B., Vityaev E., “Data Mining in Finance: Advances in Relational and Hybrid Methods”, Kluwer Academic Publishers, 308, USA, (2002).
- Koyuncugil A. S., Veri Madenciliği ve Sermaye Piyasalarına Uygulanması, Sermaye Piyasası Kurulu Araştırma Raporu, 2007.
- Kriebel H., Horn R. L., *Management Information Systems*, Cornaige-Melton University, Pittsburg, Penn, 1971, 17-18.
- Kutlar A., *Ekonometrik Zaman Serileri*, Gazi Kitabevi, Ankara, 2000.
- Kuyucu, A., “Veri Ambarı Uygulamaları ve Sonuçları”, *BT/Haber Gazetesi*, Sayı : 219, 1999.
- Larose D.T., *Discovering Knowledge in Data: An Introduction To Data Mining*, John Wiley & Sons Inc., 2005, 42-70.
- Law Rob., Au N., “A Neural Network Model To Forecast Japanese Demand for Travel To Hong Kong”, *Tourism Management*, Vol. 20, (1999), 89-97.
- Lawrence R., *Using Neural Networks To Forecast Stock Market Prices*, University of Manitoba, December 1997.
- Lawrence S., Giles C. L., Fong S., “Natural Language Grammatical Inference with Recurrent Neural Networks”, *IEEE Transactions On Knowledge and Data Engineering*, Vol. 12, No. 1, (2000), 126–140.
- Lee J., *Empirical Analysis of Wholesale Used Car Auctions*, A Dissertation in University of California, 2006.
- Lewis. C. D., *Industrial and Business Forecasting Methods*, Butterworths Publishing, London, 1982, 40.
- Lucas H. C., *Management Information Systems*, Mcgraw-Hill, New York, 1989, 28.
- Macy B. R, Pandya S. A: *Pattern Recognition with Neural Networks in C++* , CRC Press, 1995.
- Makridakis S., Wheelwright S., “Adaptive Filtering: An Integrated Autoregressive/Moving Average Filter for Time Series Forecasting”, *Operational Research Quarterly*, Vol. 28, No: 2, (1977), 426.
- Makridakis S., Wheelwright S., *Forecasting Methods for Management*, John Wiley & Sons Inc, New York, 1989.
- Makridakis S., Wheelwright S., *Interactive Forecasting Univariate and Multivariate Methods*, Holden-Day Inc., San Francisco, 1978.
- Mandic D. P., Chambers J. A., *Recurrent Neural Networks for Prediction - Learning Algorithms Architectures and Stability*, John Wiley & Sons Ltd, 2001.
- Maybek İ., “2005 Yılı Taşıt Araçları ve Yan Sanayi Sektör Raporu”, Uludağ İBGS

Raporu, Ocak 2006.

Monks J. G., Operations Management, Mcgraw-Hill International Editions, Singapore, 1987.

Montanes E., Quevedo J., Prieto M., Menendez C., “Forecasting Time Series Combining Machine Learning and Box-Jenkins Time Series”, Artificial Intelligence Center, Viesques, (2002).

Nabiyev V.V.,Yapay Zeka, Seçkin Yayıncılık, Ankara, 2003.

Nakipoğlu, A., “Veri Ambarı İşletme Sorunlarının Çözümüne Odaklanır”, BT/Haber Gazetesi, Sayı : 219, 1999.

NCR, “Bilgi Elinizden Akıp Gitmesin”, BT/Haber Gazetesi, 1999, 3-12.

Newbold P., Granger C.W.J., “The Principles of The Box-Jenkins Approach”, Operational Research Quarterly, Vol. 26. No: 2, (1975).

Oğuz B., Eşleştirme Haznelemesinin Biçimsel Kavram Analizi ile Modellenmesi, Yüksek Lisans Tezi, Hacettepe Üniversitesi, Ankara, 2000, 21-26.

Olaru C., Wehenkel L. “Data Mining”, IEEE Computer Applications in Power, (July 1999), 19-25.

Onat M. G., Otomotiv Sektöründe Oranlar Yöntemi Aracılığı ile Finansal Analiz, Yüksek Lisans Tezi, Marmara Üniversitesi, İstanbul, 2007

Oracle 9i Data Warehousing Guide, Oracle Inc, 2001, 1-2.

Orhan O. Z., Gümrük Birliği Sürecinde Türk Otomotiv Sanayii'nin ve Otomotiv Yan Sanayii'nin Rekabet Gücü, İstanbul Ticaret Odası Yayınları, No: 54, İstanbul, 1997, 11.

Orhunbilge N., Uygulamalı Regresyon ve Korelasyon Analizi, İstanbul Üniversitesi İşletme Fakültesi Yayını, No: 281, İstanbul, 2002, 5-12.

OSD-Dünya Otomotiv Sanayi Genel İstatistikî Bilgiler, Yıllık Bülten, İstanbul, 2006, 18.

OSD-Globalleşen Dünyada Otomobil, Yıllık Bülten, İstanbul, Nisan 2005, 7.

Özçınar H., KPSS Sonuçlarının Veri Madenciliği Yöntemleriyle Tahmin Edilmesi, Yüksek Lisans Tezi, Pamukkale Üniversitesi, Denizli, 2006.

Özdemir A., Veritabanlarında Bilgi Keşfi ve Veri Madenciliği, Doktora Tezi, Erzurum 2004.

Özdemir Ali, Özdemir Aslı, “Talep Tahminlemesinde Kullanılan Yöntemlerin Karşılaştırılması: Seramik Ürün Grubu Firma Uygulaması”, Ege Academic Review, Vol. 6, (2006), 105-114.

Özerman M. E., “2003 Yılı Otomotiv ve Yan Sanayi İhracatımız ve 2004 Yılına İlişkin Beklentiler”, İge Dergisi, Sayı: 26, (Ocak–Nisan 2004).

Özgen D., Yapay Sınır Ağları Analizi ve Türk Finans Piyasaları: İMKB 30 Endeksi Uygulaması, Marmara Üniversitesi, Yüksek Lisans Tezi, İstanbul, 2007.

Özkan Y., Veri Madenciliği Yöntemleri, Papatya Yayıncılık, İstanbul, Mayıs 2008.

- Özmucur S., Geleceği Tahmin Yöntemleri, İstanbul Sanayi Odası Araştırma Dairesi Yayını, No: 1990/2, İstanbul, (1990).
- Öztemel E., Yapay Sinir Ağları, Papatya Yayıncılık, Ağustos 2003.
- Palmer A., Montaño, J. J., Sesé A., “Designing An Artificial Neural Network for Forecasting Tourism Time Series”, *Tourism Management*, Vol. 26, No: 1, (2005), 1-10.
- Partal T., Cıgızoğlu K., “Yapay Sinir Ağları ile Meteorolojik Verileri Kullanarak Yağış Tahmini,” *BMYS' 2005 Sempozyumu Bildirisi*, Kocaeli, Kasım 2005.
- Pattie D. C., Snyder J., “Using A Neural Network To Forecast Visitor Behavior”, *Annals of Tourism Research*, Volume: 23, No: 1, (1996), 151-164.
- Pawlak Z., Slowinski, K., and Slowinski, R., “Rough Classification of Patients After Highly Selective Vagotomy for Duodenal Ulcer”, *International Journal of Man-Machine Studies*, Vol. 24, (1986), 413-433.
- Plummer E. A., Time Series Forecasting with Feed-Forward Neural Networks: Guidelines and Limitations, Master Thesis, University of Wyoming, USA, 2000.
- Pryke A. N., Data Mining Using Genetic Algorithms and Interactive Visualization, Ph. D. Thesis, University of Birmingham, 1998.
- Quinlan J. R., “Induction of Decision Trees”, *Machine Learning*, vol.1, (1986), 81-106.
- Quinn T., Kenny G., Meyler A., “Forecasting Irish Inflation Using Arima Models”, *Central Bank of Ireland*, 3/Rt, December 1998.
- Rumelhart D. E., Hinton G. E., Williams R., “Learning Representations By Back Propagating Errors”, *Nature*, 323, (1986), 533-536.
- Ryan J., Building and Deploying An Enterprise Data Warehouse, White Paper, 1999.
- Şahin M., Yönetim Bilgi Sistemi, Anadolu Üniversitesi Yayınları, Eskişehir, 2001.
- Schroeder R. G., Operations Management: Decision Making in The Operations Function, Mcgraw-Hill Book Co., Singapore, 1989.
- Şen Z., Yapay Sinir Ağları İlkeleri, Su Vakfı, İstanbul, 2004.
- Serper Ö., İstatistik, Fikir Kitabevi, İstanbul, 1981, 23-206.
- Sevüktekin M., Nargeleşkenler M., Zaman Serileri Analizi, Nobel Basımevi, Ankara, 2005, 336.
- Sharp K. P., “Modeling Canadian Price and Wage Inflation”, *Canadian Institute of Actuaries Publications*, (October 1992).
- Slaughter G. E. F., Artificial Neural Network for Temporal Impedance Recognition of Neurotoxins, M.S. Thesis, Virginia Commonwealth University, Virginia, 2003, 115.
- Sönmez H., R Yazılımı ile İstatistiksel Analiz, Kaan Kitabevi, Eskişehir, 2006.
- Subaşı D.B., Enflasyonun Arıma Modelleri ile Tahminlenmesi: 1994–2005 Türkiye Uygulaması, Yüksek Lisans Tezi, Dumlupınar Üniversitesi, Kütahya, 2005.
- Tebelkis J., Speech Recognition Using Neural Networks, Ph.D. Thesis, Carnegie Mellon

University, Pennsylvania, Mayıs 1995.

Türkiye Vakıflar Bankası Yayınları, Plânlama ve İktisadi Araştırmalar Grup Yönetmeliği, Sektör Araştırmaları Serisi/No:28, “Otomotiv Sektörü”, 2003, 4.

Ülgen H., İşletme Yönetiminde Bilgisayarlar, İstanbul Üniversitesi İşletme Fakültesi Yayını, No: 5, Yön Ajans Basımevi, İstanbul, 1990, 4.

Uludağ İBGS Yayını, 2005 Yılı Taşıt Araçları ve Yan Sanayi Sektör Raporu, Bursa, 2006, 2.

Usta A.S., YSA Uygulaması Kullanılarak Üretici Fiyat Endeksi (ÜFE) Değerlerinin Öngörü Modellemesi ve Analizi, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi, İstanbul, 2007.

Uysal M., El Roubi S., “Artificial Neural Network Versus Multiple Regression in Tourism Demand Analysis”, Journal of Travel Research, Vol. 38, No: 2, (1999), 111-118.

Uysal M., SQL Veri Tabanı Sorgulama Dili, Beta Basım Yayım Dağıtım, İstanbul, Temmuz 2000.

Vahaplar A., İnceoğlu M., “Veri Madenciliği ve Elektronik Ticaret”, İnet 2000 Bildirisi, İstanbul, 2000.

Velickov S., Solomatine D., “Predictive Data Mining: Practical Examples”, Artificial Intelligence In Civil Engineering, Germany, (2000), 1-17.

Wasney A. F., “Introduction To Statistical Time Series”, John-Wiley and Sons Inc., USA, (1976), 3.

Whellwright S.C., Makridakis S., Forecasting Methods for Management, John Wiley & Sons, Inc., New York, 1973, 123-139.

Xu Y., Using Data Mining In Educational Research: A Comparison of Bayesian Network with Multiple Regression in Prediction, Department of Educational Psychology, The University of Arizona, Arizona, 2003, 242.

Yeltin L.T., Gümrük Birliği Çerçevesinde Avrupa Birliği ve Türkiye’de Otomotiv Sektörü, İstanbul, 1999.

Yıldız Ö., Döviz Kuru Tahmininde Yapay Sinir Ağlarının Kullanımı, Yüksek Lisans Tezi, Osmangazi Üniversitesi, Eskişehir, 2006.

Yılmaz C., Bilgi İşlem ve Yönetim Bilgi Sistemi, Erciyes Üniversitesi, Kayseri, 1988.

Yılmaz L., A Decision Support Systems Using Data Mining, Yüksek Lisans Tezi, Yeditepe Üniversitesi, İstanbul, 2002, 16-22.

Yüksek A. G., Bilgisayar Sistemlerinde Veri Tabanı Yönetimi ve Bir Uygulama, Yüksek Lisans Tezi, Cumhuriyet Üniversitesi, Sivas, 2001.

Yurtoğlu H., Yapay Sinir Ağları Metodolojisi ile Öngörü Modellemesi: Bazı Makroekonomik Değişkenler İçin Türkiye Örneği, DPT Uzmanlık Tezi, 2005.

Zhang G., Patuwo B.E., Hu M.Y., “Forecasting with Artificial Neural Networks: The State of The Art”, International Journal of Forecasting, Vol. 14, No.1, (1998), 35-62.

(Elektronik Kaynaklar)

Akal M., “Optimum Öngörü Tekniğinin Seçimi”, Adapazarı, 2004,
www.sakarya.edu.tr/~hgurak/yazilar/makale/ONGOM.doc, (01.04.2008).

Cambazoğlu T., “Kurumlarda Yararlı Bilginin Yönetimi ve İlihtili Teknolojiler”, 2002,
www.bilisimrehber.com.tr/arastirma/tr_arastirma_intraextra14.phtml, (01.04.2008).

Datamonitor, Used Cars Industry Profile, Proquest Elektronik Veritabanı, 2007,
proquest.umi.com, (01.05.2008).

Dilly R., Data Mining: An Introduction, 1999,
www.pcc.qub.ac.uk/tec/courses/datamining/ohp/dm-OHP-final_1.html, (01.12.2007).

Eker H., Veri Madenciliği veya Bilgi Keşfi, 2002,
www.bilgiyonetimi.org/cm/pages/mkl_gos.php?nt=538, (01.12.2007).

SAS Institute Inc., The Data Mining Challenge: Turning Raw Data Into Business Gold,
www.sas.com/software/data_mining, (15/09/2007).

Web-ACEA, Association Des Connstructeurs Europeens D’automobiles, Economic Report, Acea’s Position On Motor Vehicle Distribution in The European Union, 2001,
www.acea.be, (15/09/2007).

Web-Basgaza, www.basgaza.com/fkonu_goster.asp?id=6532, (01.06.2008).

Web-Btinsan, www.btinsan.com/510/31.asp, 2003, (01.06.2008).

Web-Capital1, www.capital.com.tr/haber.aspx?HBR_KOD=1778, (01.06.2008).

Web-Capital2, www.capital.com.tr/haber.aspx?HBR_KOD=657, (01.07.2008).

Web-CrispDM, www.crisp-dm.org, (01.06.2008).

Web-Exforsys, www.exforsys.com/tutorials/data-warehousing/data-warehouse-disadvantages.html, (01.06.2008).

Web-Exi26, www.exi26.com, (01.06.2008).

Web-Fazlamesai, www.fazlamesai.net/index.php?a=article&sid=2884, (01.06.2008).

Web-Foreigntrade, www.foreigntrade.gov.tr, (01.06.2008).

Web-Hürriyet1, arama.hurriyet.com.tr/arsivnews.aspx?id=231202, (15.06.2008).

Web-Hürriyet2, arama.hurriyet.com.tr/arsivnews.aspx?id=8000581, (15.06.2008).

Web-KBA, Official Website of German Federal Authority for Automotive Transport,
www.kba.de, (2004), (15.06.2007).

Web-Kdnuggets1, Mart 2007, www.kdnuggets.com/polls/2007/data_mining_methods.htm,
(15.06.2008).

Web-Kdnuggets2, Mayıs 2008, www.kdnuggets.com/polls/2008/data-mining-software-tools-used.htm, (15.06.2008).

Web-Kobifinans1, www.kobifinans.com.tr/tr/sector/011001/12455, (01.06.2008).

Web-Kobifinans2, www.kobifinans.com.tr/tr/sektor/011002/13652, (01.06.2008).

Web-Kobifinans3, www.kobifinans.com.tr/tr/sektor/011001/13544, (01.06.2008).

Web-Kobifinans4, www.kobifinans.com.tr/tr/sektor/011002/18020, (01.06.2008).

Web-Mekatronikkulubu,
www.mekatronikkulubu.org/yazi/cok_katmanli_yapay_sinir_glari, (01.07.2008).

Web-Mynet, otomobil.mynet.com/haberler/157-kullanilmis-arac-alirken-I, (01.06.2008).

Web-OICA, www.oica.net, (01.06.2007).

Web-OSD, www.osd.org.tr, (01.12.2007).

Web-Sakarya, www.sakarya.edu.tr/~altunr/site/ders/pa, (01.08.2008).

Web-Scholar, scholar.google.com.tr, (01.07.2008).

Web-Wikipedia1, en.wikipedia.org/wiki/Web_crawling, (11.06.2008).

Web-Wikipedia2, en.wikipedia.org/wiki/Data_warehouse, (11.06.2008).

Web-Wikipedia3, tr.wikipedia.org/wiki/Veri_madencili%C4%9Fi, (11.06.2008).

Web-Wikipedia4, en.wikipedia.org/wiki/Used_car, (11.06.2008).

Web-Wikipedia5, tr.wikipedia.org/wiki/Yapay_sinir_aglari, (15.07.2008).

Web-Yapayzeka, www.yapay-zeka.org, (15.07.2008).

ÖZGEÇMİŞ

Adı ve SOYADI : Özcan ASILKAN
Doğum Tarihi ve Yeri : 01.04.1971
Medeni Durumu : Evli

Eğitim Durumu

Mezun Olduğu Lise : Kütahya Lisesi
Lisans Diploması : Marmara Üniversitesi, Müh.F., Bilgisayar Mühendisliği, 1995
Yüksek Lisans-1 Diploması : Dumlupınar Üniversitesi, SBE, İşletme Anabilim Dalı, 1998
Tez-1 Konusu : İşletmelerde Yönetim Bilişim Sistemi Geliştirme ve Bir Şirketler Grubunda Uygulama
Yüksek Lisans-2 Diploması : Sakarya Üniversitesi, FBE, Bilgisayar ve Bilişim Mühendisliği Anabilim Dalı, 2002.
Tez-2 Konusu : Kurumsal Müşteri İlişkileri Yönetimi Projelerinde Yazılım Otomasyonu
Yabancı Diller : İngilizce ve Almanca

Bilimsel Faaliyetler

Makaleler

- Asilkan Ö., “CRM Uygulamalarının Verimliliğini Artırmak İçin Kurumsal Veri Ambarlarının Kullanılması”, Sakarya Üniversitesi FBE Dergisi, Temmuz 2002.
- Kuruüzüm A., Asilkan Ö., Çizel B.R., “Student Participation in Higher Education Institutions in Turkey”, Higher Education in Europe, Vol.30, No.3-4, 2005.

İş Deneyimi

Stajlar

- Deutsche Post World Net, Bonn/Almanya, Temmuz-Ağustos 2002.
- Hamburg Üniversitesi, Doktora Tez Çalışması, Hamburg/Almanya, Mayıs-Ağustos 2002.

Çalıştığı Kurumlar

- Akdeniz Üniversitesi, Antalya, 2003-2007.
- Güral Şirketler Grubu, Kütahya, 1995-2002.
- AC, Lüneburg/Almanya, 2007-2008.

Adres : Liman Mh, Mimoza Sitesi, F1/14, Antalya.
Tel. no : +90 (505) 331 97 59
Email : ozcan@asilkan.net